



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Tratamento de riscos relacionados à alocação de  
projetos: proposição de módulo para uma plataforma  
de aprendizagem ativa**

Everaldo Silva Júnior

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Orientadora

Prof.a Dr.a Simone Borges Simão Monteiro

Brasília  
2021



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# **Tratamento de riscos relacionados à alocação de projetos: proposição de módulo para uma plataforma de aprendizagem ativa**

Everaldo Silva Júnior

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Prof.a Dr.a Simone Borges Simão Monteiro (Orientadora)  
PPCA - UnB

Prof. Dr. Patricio Esteban Ramírez-Correa      Prof. Dr. Ari Melo Mariano  
Membro Externo / Universidade Católica do Norte - Chile      Membro Interno / PPCA - UnB

Prof. Dr. Marcelo Ladeira  
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 06 de outubro de 2021

# Dedicatória

Dedico este trabalho a Deus, que sempre me deu muito mais do que eu julgava necessário. Sem ele nada seria possível.

Dedico este projeto também a todos que buscam, de alguma maneira, fazer a diferença em seus ambientes de trabalho.

# Agradecimentos

À minha orientadora, Professora Simone Borges Simão Monteiro, por todos os seus ensinamentos, paciência e confiança durante toda a pesquisa, bem como todo incentivo e motivação que tornaram possível o alcance dos resultados apresentados neste documento.

A todos os meus amigos, que, de algum modo, contribuíram para o cumprimento desta etapa fundamental da minha vida profissional.

Aos professores Ari Melo Mariano e Patricio Esteban Ramírez-Correa pela disponibilidade em avaliar este estudo.

# Resumo

As universidades estão cada vez mais interessadas em cooperar com agentes externos por meio de projetos, a fim de permitir que seus alunos adquiram habilidades e conhecimentos para resolver problemas reais. Dessa forma, os métodos de aprendizagem ativa, como o Project-Based Learning (PBL), são utilizados em disciplinas para criar um perfil de aluno alinhado às demandas da prática profissional. No entanto, existem riscos associados aos processos que são executados para implementar o PBL nas disciplinas, por exemplo, o processo de alocação de projetos. Este estudo tem como objetivo propor uma ferramenta computacional para tratar os riscos da alocação de projetos em disciplinas que utilizam a metodologia PBL, com o intuito de dar suporte a essa aprendizagem ativa no curso de graduação em Engenharia de Produção da Universidade de Brasília (UnB). Portanto, a intenção é alocar projetos para disciplinas que aplicam conhecimentos necessários para resolução de problemas que foram submetidos na Plataforma Unificada de Metodologia Ativa (PUMA), onde será implementada a ferramenta computacional proposta. Esta pesquisa possui abordagem quantitativa e qualitativa e é classificada como aplicada e exploratória, a qual apresenta como estratégia um estudo de caso. A estrutura da pesquisa possui etapas de gerenciamento de riscos e de mineração de dados que foram baseadas na ISO 31000:2018 e no CRISP-DM, respectivamente. Deste modo, foram propostas ações de tratamento dos riscos de alocação inadequada de projetos para disciplinas de Projeto de Sistemas de Produção (PSP), no âmbito do Curso de Engenharia de Produção da UnB. Além disso, um classificador automático de artigos baseado em mineração de texto foi desenvolvido. Dentre os métodos utilizados, destaca-se o algoritmo K-NN, que permitiu, com o uso de  $K = 7$  e a Distância Euclidiana, obter uma precisão preditiva de 86% em um conjunto de dados composto por quatro temáticas da Engenharia de Produção, as quais são equivalentes a quatro disciplinas de PSP. Como resultado final, foi proposto o Módulo de Alocação de Projetos para a plataforma PUMA a partir das ações de tratamento levantadas e dos resultados do classificador desenvolvido.

**Palavras-chave:** Alocação de Projetos, Aprendizagem Baseada em Projetos, Gestão de Riscos, Mineração de Dados

# Abstract

Universities are increasingly looking forward to cooperating with stakeholders through projects, in order to enable their students to acquire skills and knowledge to solve real problems. In that way, active learning methods, such as the Project-Based Learning (PBL), are used in courses to create a student profile aligned with the demands of professional practice. However, there are risks associated with the processes that are executed to implement PBL in courses, for example, the project assignment process. This study aims to propose a computational tool to treat the project assignment risks in courses that use the PBL methodology, in order to support this active learning in the Production Engineering undergraduate program of the University of Brasília (UnB). Therefore, the intention is to allocate projects in courses that can apply knowledge necessary to solve the problems submitted at Platform for Unifying Methodologies of Active learning (PUMA), where the computational tool proposed will be implemented. This research has a quantitative and qualitative approach and is classified as applied and exploratory, which presents a case study as its strategy. The research structure has risk management and data mining steps that are based on ISO 31000:2018 and CRISP-DM, respectively. Thus, actions were proposed to treat the risks of inadequate allocation of projects to courses (PSPs), within the scope of the Production Engineering Course at UnB. In addition, an automatic article classifier based on text mining was developed. Among the methods used, the K-NN algorithm stands out, which allowed, with the use of  $K = 7$  and the Euclidean Distance, to obtain a predictive accuracy of 86% in a data set composed of four themes of Production Engineering, which are equivalent to four PSP courses. As a final result, the Project Allocation Module was proposed for the PUMA, based on the risk treatment actions identified and the results of the developed classifier.

**Keywords:** Project Allocation, Project-Based Learning (PBL), Risk Management, Data Mining

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização do Problema . . . . .	1
1.1.1	Aplicação da metodologia PBL no Curso de Engenharia de Produção da UnB . . . . .	2
1.1.2	Plataforma Unificada de Metodologia Ativa (PUMA) . . . . .	3
1.1.3	Definição do Problema . . . . .	4
1.2	Justificativa do Tema . . . . .	5
1.3	Objetivos . . . . .	5
1.3.1	Objetivo geral . . . . .	6
1.3.2	Objetivos específicos . . . . .	6
1.4	Contribuições Esperadas . . . . .	6
<b>2</b>	<b>Referencial Teórico</b>	<b>8</b>
2.1	Aprendizagem ativa . . . . .	8
2.1.1	PBL . . . . .	9
2.2	Gestão de riscos . . . . .	10
2.2.1	ABNT NBR ISO 31000 . . . . .	11
2.2.2	Técnicas para execução do processo de avaliação de riscos . . . . .	13
2.2.3	Riscos relacionados à alocação de propostas de projetos no contexto PBL . . . . .	14
2.3	Mineração de dados e textos . . . . .	17
2.3.1	CRISP-DM . . . . .	18
2.4	Trabalhos relacionados . . . . .	20
2.4.1	Modelos utilizados . . . . .	20
2.4.2	Técnicas aplicadas para preparação de dados . . . . .	35
2.4.3	Medidas e métodos aplicados na modelagem . . . . .	37
2.4.4	Estratégias para avaliação de modelos . . . . .	41

<b>3</b>	<b>Metodologia da Pesquisa</b>	<b>43</b>
3.1	Classificação da Pesquisa . . . . .	43
3.2	Estrutura da Pesquisa . . . . .	44
3.2.1	Etapa 1 - Entendimento do contexto e das questões de negócio . . .	46
3.2.2	Etapa 2 - Processo de avaliação dos riscos e dados . . . . .	49
3.2.3	Etapa 3 - Construção do modelo de alocação de projetos com base no tratamento dos riscos . . . . .	54
3.2.4	Etapa 4 - Proposta de Implantação do Módulo de Alocação de Pro- jetos na plataforma PUMA . . . . .	58
3.2.5	Fases contínuas . . . . .	59
<b>4</b>	<b>Resultados e Discussões</b>	<b>61</b>
4.1	Etapa 1 - Entendimento do contexto e das questões de negócio . . . . .	61
4.1.1	Estabelecimento do contexto . . . . .	61
4.1.2	Compreensão do negócio . . . . .	64
4.2	Etapa 2 - Processo de avaliação dos riscos e dados . . . . .	71
4.2.1	Identificação dos riscos . . . . .	71
4.2.2	Análise dos riscos . . . . .	75
4.2.3	Avaliação dos riscos . . . . .	81
4.2.4	Entendimento dos dados . . . . .	83
4.3	Etapa 3 - Construção do modelo de alocação de projetos com base no tratamento dos riscos . . . . .	89
4.3.1	Tratamento dos riscos . . . . .	89
4.3.2	Preparação dos dados . . . . .	92
4.3.3	Modelagem . . . . .	93
4.3.4	Avaliação . . . . .	96
4.4	Etapa 4 - Proposta de Implantação do Módulo de Alocação de Projetos na plataforma PUMA . . . . .	97
<b>5</b>	<b>Conclusão</b>	<b>101</b>
5.1	Considerações Finais . . . . .	101
5.2	Sugestões de Trabalhos Futuros . . . . .	103
	<b>Referências</b>	<b>105</b>
	<b>Apêndice</b>	<b>112</b>
<b>A</b>	<b>Tabelas com os valores de priorização das causas raízes dos riscos</b>	<b>113</b>



B	Ações de tratamento dos riscos consideradas de médio e longo prazo	115
C	Código desenvolvido no Google Colab	117

# Lista de Figuras

2.1	Processo de gestão de riscos da ISO 31000. . . . .	12
2.2	Fases do CRISP-DM. . . . .	18
2.3	Arquitetura do modelo de seleção e atribuição de projetos. . . . .	20
2.4	Proposta de sistema para gerenciamento de projetos finais de graduação. . . . .	21
2.5	Arquitetura de recomendação automática de projetos do GitHub. . . . .	22
2.6	Abordagem personalizada de recomendação de projetos do GitHub. . . . .	23
2.7	Visão geral do sistema de recomendação de objetos de aprendizagem. . . . .	24
2.8	Esquema do sistema para recomendação de professores de disciplinas. . . . .	25
2.9	Estrutura do sistema para recomendação de projetos de pesquisa. . . . .	26
2.10	Sistema de apoio à decisão para gerenciamento de projetos de último ano. . . . .	27
2.11	Recomendador de recursos de suporte ao desenvolvimento de projetos. . . . .	28
2.12	Arquitetura do sistema de mineração de projetos de conclusão de graduação. . . . .	29
2.13	Sistema de atribuição de avaliadores para projetos finais de graduação. . . . .	30
2.14	Fluxo de informações do <i>plugin</i> YODA. . . . .	32
2.15	Módulos principais da proposta de recomendação de consultores <i>ad-hoc</i> . . . . .	33
2.16	Visualização do VSM. . . . .	37
2.17	Pseudocódigo do algoritmo K-NN. . . . .	39
2.18	Matriz de confusão para duas classes. . . . .	41
3.1	Classificação da pesquisa. . . . .	43
3.2	Etapas da pesquisa. . . . .	45
3.3	Esquema geral da Análise de Causa Raiz. . . . .	51
3.4	Escala Likert de 5 pontos utilizada na priorização das causas raízes dos riscos. . . . .	52
3.5	Enquete para priorização das causas raízes dos riscos. . . . .	52
3.6	Enquete para priorização das ações de tratamento das causas raízes. . . . .	54
4.1	Riscos de utilização inadequada do Módulo de Alocação de Projetos. . . . .	72
4.2	Riscos de execução inadequada do processo de mineração de dados. . . . .	73
4.3	Riscos de alocação inadequada de projetos de graduação. . . . .	73
4.4	Riscos de alocação inadequada de projetos para as disciplinas de PSP. . . . .	74

4.5	Riscos de alocação inadequada de projetos interdisciplinares dos PSPs. . .	74
4.6	Análise de Causa Raiz dos riscos de utilização inadequada do Módulo de Alocação de Projetos. . . . .	76
4.7	Análise de Causa Raiz dos riscos de execução inadequada do processo de mineração de dados. . . . .	77
4.8	Análise de Causa Raiz dos riscos de alocação inadequada de PGs. . . . .	78
4.9	Análise de Causa Raiz dos riscos de alocação inadequada de projetos para as disciplinas de PSP. . . . .	79
4.10	Análise de Causa Raiz dos riscos de alocação de projetos interdisciplinares para as disciplinas de PSP. . . . .	80
4.11	Distribuição dos documentos por ano. . . . .	87
4.12	Gráfico de documentos por temática. . . . .	88
4.13	Nuvem de palavras que compõem o conjunto de dados. . . . .	88
4.14	Técnicas aplicadas para preparação dos dados. . . . .	92
4.15	Diagrama de blocos representando o modelo proposto de alocação de projetos. . . . .	94
4.16	Gráfico de acurácia por valores de K e métricas de distância. . . . .	94
4.17	Matriz de Confusão - K-NN. . . . .	95
4.18	Interface para avaliação do modelo final de alocação de projetos. . . . .	96
4.19	Tela de login do protótipo do módulo. . . . .	97
4.20	Tela de cadastro protótipo do módulo. . . . .	97
4.21	Tela de recuperação de senha protótipo do módulo. . . . .	98
4.22	Tela inicial dos agentes externos protótipo do módulo. . . . .	98
4.23	Tela de submissão de projetos do protótipo do módulo. . . . .	98
4.24	Tela inicial dos professores no protótipo do módulo. . . . .	99
4.25	Tela de criação de disciplina do protótipo do módulo. . . . .	99
4.26	Tela das disciplinas do protótipo do módulo. . . . .	100
4.27	Tela de avaliação de propostas de projeto do protótipo do módulo. . . . .	100

# Lista de Tabelas

2.1	Ferramentas e técnicas adotadas pelos métodos de aprendizagem ativa. . . . .	9
2.2	Riscos relacionados à alocação de propostas de projetos. . . . .	15
2.3	Riscos relacionados ao módulo de alocação de projetos. . . . .	16
2.4	Objetivos gerais e principais dados dos modelos apresentados. . . . .	34
3.1	Reuniões para estabelecimento do contexto. . . . .	47
3.2	Reuniões para compreensão do negócio. . . . .	48
3.3	Sessões de <i>brainstorming</i> para identificação dos riscos. . . . .	50
3.4	Sessões de <i>brainstorming</i> para análise dos riscos. . . . .	51
3.5	Perguntas para avaliação do modelo. . . . .	57
4.1	Lista de recursos disponíveis para a pesquisa. . . . .	65
4.2	Requisitos, premissas e restrições da pesquisa. . . . .	66
4.3	Escopo da pesquisa. . . . .	66
4.4	Critérios para a pesquisa. . . . .	67
4.5	Técnicas selecionadas para o processo de gestão de riscos. . . . .	68
4.6	Avaliação de artigos para seleção de técnicas de mineração de dados. . . . .	69
4.7	Técnicas selecionadas para o processo de mineração de dados. . . . .	70
4.8	Riscos inicialmente priorizados para tratamento. . . . .	82
4.9	Análise preliminar dos dados. . . . .	83
4.10	Artigos do ENEGEP selecionados para a fase de modelagem. . . . .	85
4.11	Análise descritiva básica da tabela de dados. . . . .	86
4.12	Ações de tratamento priorizadas. . . . .	90
4.13	Etapas da pesquisa em que serão implementadas as ações de tratamento. . . . .	91
4.14	Relatório de Classificação - K-NN . . . . .	95
4.15	Comparativo dos resultados da avaliação do Módulo de Alocação de Projetos . . . . .	96
A.1	Priorização das causas raízes dos riscos de utilização inadequada do Módulo. . . . .	113
A.2	Priorização das causas raízes dos riscos de inadequada mineração de dados. . . . .	114

A.3	Priorização das causas raízes dos riscos de alocação inadequada de projetos para as disciplinas de PSP. . . . .	114
A.4	Priorização das causas raízes dos riscos de alocação de projetos interdisciplinares para as disciplinas de PSP. . . . .	114
B.1	Priorização das ações de tratamento de curto prazo. . . . .	115
B.2	Priorização das ações de tratamento de médio prazo. . . . .	116

# Lista de Abreviaturas e Siglas

**BDM** Biblioteca Digital da Produção Intelectual Discente da Universidade de Brasília.

**BDTD** Biblioteca Digital Brasileira de Teses e Dissertações.

**COSO** *Committee of Sponsoring Organizations of the Treadway Commission.*

**CRISP-DM** *Cross Industry Standard Process for Data Mining.*

**EPR** Departamento de Engenharia de Produção.

**ESPRIT** *Ecole Supérieure Privée d'Ingénierie et de Technologies.*

**FT** Faculdade de Tecnologia.

**ISO** *International Organization for Standardization.*

**K-NN** *K-Nearest Neighbors.*

**KDD** *Knowledge Discovery in Databases.*

**PBL** *Project-Based Learning.*

**PG** Projeto de Graduação.

**PSP** Projeto de Sistemas de Produção.

**PUMA** Plataforma Unificada de Metodologia Ativa.

**TF-IDF** *Term Frequency-Inverse Document Frequency.*

**TIC** Tecnologia da Informação e Comunicação.

**UnB** Universidade de Brasília.

**VSM** *Vector Space Model.*

**WoS** *Web of Science.*

# Capítulo 1

## Introdução

Neste capítulo, são apresentadas as informações principais que contextualizam a problemática a ser tratada na pesquisa e justificam o seu tema. Além disso, é delimitado o escopo da pesquisa por meio da definição dos seus objetivos.

### 1.1 Contextualização do Problema

As universidades e as organizações empresariais têm buscado cooperar na formação dos estudantes, pretendendo criar um perfil de ensino e aprendizagem alinhado com a demanda da prática profissional. Este fato é sustentado pela adoção crescente de um modelo de educação focado na disseminação do conhecimento e no desenvolvimento de competências, de modo que os professores e as empresas tornam-se responsáveis pela formação dos futuros profissionais [1, 2, 3].

Neste sentido, diversos cursos apresentam, em sua grade curricular, disciplinas e projetos finais de graduação que adotam a metodologia *Project Based Learning* (PBL), com o objetivo de fazer com que os estudantes apliquem conhecimentos teóricos na resolução de problemas reais [3, 4, 5, 6, 7].

Os problemas reais das empresas, sejam elas privadas, públicas, grandes ou pequenas, fornecem o insumo para as temáticas a serem trabalhadas nos projetos desenvolvidos pelos alunos [8]. Esses projetos se diferem no que tange às suas formas de execução e avaliação, às suas áreas do conhecimento e aos seus contextos de desenvolvimento. Isto porque, a depender do que o professor julgar importante à temática envolvida ou à disciplina base, busca-se desenvolver competências técnicas e transversais específicas [5, 9].

Assim sendo, a qualidade da orientação dos projetos pelos professores é importante, pois impacta diretamente nos processos e ambientes de aprendizagem, influenciando, assim, o desenvolvimento acadêmico dos alunos, seja em projetos finais de graduação ou em disciplinas que utilizam a metodologia PBL [3, 10, 11, 12].

É dado destaque aos papéis do professor orientador e das empresas no desenvolvimento de projetos, pois os orientadores estarão com os alunos em todos os estágios e as necessidades das empresas, devidamente alinhadas e atualizadas, que garantirão as futuras entradas dos estudantes no mercado de trabalho [2, 3].

Diante deste contexto, um exemplo prático de aplicação da metodologia PBL na estrutura curricular de um curso pode ser encontrado no Departamento de Engenharia de Produção (EPR) da Universidade de Brasília (UnB), conforme é apresentado na Seção 1.1.1.

### **1.1.1 Aplicação da metodologia PBL no Curso de Engenharia de Produção da UnB**

Dentre as diversas motivações para a criação do Curso de Engenharia de Produção da UnB, destaca-se o fato de o setor de serviços ocupar uma posição de destaque no Distrito Federal, tanto na área privada quanto no setor público, sendo uma excelente perspectiva a atuação de engenheiros com um perfil profissional que lhes permita operar na identificação e resolução de problemas [13, 14].

Por conta desse panorama, o Curso de Engenharia de Produção da UnB tem oito disciplinas que adotam a aprendizagem ativa, mais especificamente a metodologia PBL, do quarto ao décimo semestre, denominadas Projeto de Sistemas de Produção (PSP) [15]. Essas disciplinas são focadas em diversas áreas do conhecimento, por exemplo, Probabilidade e Estatística; Sistema de Informação aplicado à Engenharia de Produção; Planejamento e Controle da Produção; Gestão da Qualidade; Engenharia do Produto; e Gestão Estratégica [15].

As disciplinas de PSP têm como objetivo fazer com que os alunos apliquem os conhecimentos teóricos da Engenharia de Produção na resolução de problemas reais, garantindo uma visão articulada entre as características da atuação profissional e as diferentes áreas do conhecimento, de forma que os alunos compreendam a diversidade de aspectos determinantes envolvidos na solução de problemas [5, 15].

Além das disciplinas de PSP, os alunos resolvem problemas de engenharia de forma objetiva e multidisciplinar no Projeto de Graduação, que corresponde ao trabalho de final de curso, sintetizando, em dois semestres, o conhecimento adquirido ao longo da graduação e tomando contato com desenvolvimentos científicos e tecnológicos recentes [13].

Os problemas reais tratados tanto nas disciplinas de PSP quanto nos projetos de graduação podem ser oriundos de agentes externos (*stakeholders*), como empresas, ou de ideias de temas dos alunos de graduação, mestrandos, doutorandos ou professores orientadores.



O Departamento de Engenharia de Produção (EPR) da Faculdade de Tecnologia (FT) é o órgão criado pela UnB para manter as atividades ensino, pesquisa e extensão do Curso [16]. Muitas dessas atividades de ensino envolvem a aplicação da metodologia PBL.

Desta forma, visando ajudar na aplicação e melhoria contínua da metodologia PBL no âmbito do Curso de Engenharia de Produção da UnB, está em desenvolvimento a Plataforma Unificada de Metodologia Ativa PUMA, que é apresentada na Seção 1.1.2.

### 1.1.2 Plataforma Unificada de Metodologia Ativa (PUMA)

Embora a plataforma PUMA possa ser utilizada em outros cursos, universidades e países, ela teve seus primeiros requisitos oriundos das necessidades do Curso de Engenharia de Produção da UnB [8].

É uma plataforma voltada ao incentivo acadêmico para o uso de Tecnologia da Informação e Comunicação (TIC) como instrumento útil para medir a eficácia da metodologia PBL, de forma a fornecer *feedbacks* e informações substanciais e seguras para o redirecionamento da aprendizagem ativa aplicada ao longo dos anos nos cursos de ensino superior [17].

Essa plataforma integra os insumos e os resultados dos projetos desenvolvidos ao longo da graduação, sendo a porta de entrada das demandas dos *stakeholders* a serem trabalhadas pelos estudantes na resolução de problemas reais [8, 17]. Desta maneira, ela possibilita que os cursos acompanhem as exigências do mercado, fazendo com que eles estejam sempre alinhados com as expectativas de todos os seus *stakeholders*.

Atualmente, a plataforma PUMA está sendo desenvolvida por meio de um projeto multidisciplinar que envolve alunos e professores de cursos de graduação e de mestrado de dois países, a saber: Brasil, com integrantes de cursos da UnB, mais especificamente da Engenharia de Produção e do Mestrado Profissional em Computação Aplicada (MPCA) do Campus Darcy Ribeiro e da Engenharia de Software da Faculdade UnB Gama (FGA); e Tunísia, em parceria com o Curso de Ciência da Computação da *Ecole Supérieure Privée d'Ingénierie et de Technologies* (ESPRIT) [18, 19].

Tanto os alunos brasileiros quanto os tunisianos são divididos em equipes, de forma que disciplinas de diferentes semestres e cursos são integradas para o desenvolvimento de módulos da plataforma PUMA [20]. Cada disciplina e curso tem um papel delimitado pelo seu escopo de atuação, que vai desde o levantamento de requisitos até a programação do módulo, passando por equipes que verificam a qualidade das entregas realizadas ao longo de cada semestre. Há também, em uma das disciplinas da UnB, uma equipe responsável pela integração de todos os integrantes do projeto PUMA.

A importância dessa multidisciplinaridade dos integrantes está na eficácia do desenvolvimento da Plataforma, pois propicia a junção de áreas do conhecimento de distintas

disciplinas, cursos e países para o intercâmbio de ideias de funcionalidades e para o compartilhamento de tarefas adequadas aos campos de atuação dos alunos e professores [21].

As funcionalidades da plataforma PUMA são divididas em módulos integrados que buscam automatizar desde a captação de projetos até a avaliação e evolução das competências transversais e técnicas dos alunos ao longo do curso [17].

Neste sentido, os módulos da plataforma PUMA buscam o tratamento de diversos riscos associados aos processos manuais que pretendem-se automatizar, de forma que este estudo visa contribuir para a redução do impacto negativo do problema que é apresentado na Seção 1.1.3.

### 1.1.3 Definição do Problema

Um dos grandes desafios que a plataforma PUMA enfrenta é a alocação das propostas de projetos para as disciplinas de PSP. Isso porque os projetos precisam ser direcionados para as disciplinas que tenham áreas do conhecimento relacionadas às propostas, de forma que as equipes de alunos sejam orientadas por professores com conhecimentos e experiências acerca da resolução de cada problema.

Além disso, a depender das características das propostas, elas podem ser destinadas para os projetos de graduação, visto que, nesta etapa acadêmica importante, espera-se que o Curso de Engenharia de Produção da UnB ofereça aos alunos oportunidades de aplicar conhecimentos em contextos práticos.

Sabe-se também que a plataforma PUMA está em desenvolvimento para centralizar as propostas de projetos em seu banco de dados e que o Departamento de Engenharia de Produção da UnB possibilita que, em média, 90 projetos sejam desenvolvidos em disciplinas de PSP e em projetos de graduação ao longo de cada semestre. Entretanto, não há um processo rápido e assertivo de alocação de projetos dentro ou fora da plataforma PUMA, de modo que os riscos atrelados a essa alocação não são tratados.

Ademais, tem-se a dificuldade de leitura completa de todas as propostas de projetos submetidas na plataforma PUMA, pois não há um professor ou servidor da Engenharia de Produção da UnB com habilidades e disponibilidade de tempo suficientes para esta atividade que possibilitaria uma alocação manual de projetos nas disciplinas do Curso.

Ressalta-se também que, embora os agentes externos possam assinalar temas correlatos a suas propostas de projetos ao submeterem elas ao Curso de Engenharia de Produção da UnB, não é possível garantir que eles tenham conhecimentos suficientes acerca das áreas do conhecimento que são trabalhadas nas disciplinas de PSP.

Diante disso, este estudo concentra-se na alocação das propostas de projetos submetidas na plataforma PUMA no contexto do Curso de Engenharia de Produção da UnB, visando a destinação correta delas para as disciplinas que utilizam a metodologia PBL e

para os professores que podem dar uma melhor orientação aos alunos acerca das temáticas abordadas em cada projeto.

Essa pesquisa busca responder a seguinte pergunta: como atribuir de forma mais assertiva e rápida as propostas de projetos para as disciplinas de PSP, de modo que os riscos relacionados a este processo de alocação sejam tratados? A justificativa do tema a ser trabalhado nesta pesquisa e a relevância dessa questão são apresentadas na Seção 1.2.

## 1.2 Justificativa do Tema

A alocação de projetos pode impactar no desenvolvimento deles e, conseqüentemente, na qualidade dos seus resultados, pois esse processo de alocação engloba diversos riscos. Dentre esses riscos, pode-se exemplificar a escolha de disciplinas ou professores que não têm interesse nas áreas do conhecimento necessárias para a resolução dos problemas envolvidos nas propostas de projetos.

Caso os riscos relacionados à alocação de projetos aconteçam, eles podem levar a insatisfação dos *stakeholders*. Assim, torna-se necessário o desenvolvimento de um módulo para destinação eficaz de propostas de projetos dentro da plataforma PUMA, levando em conta o tratamento dos riscos de alocação inadequada de projetos, no contexto do Curso de Engenharia de Produção da UnB.

A relevância da questão de pesquisa investigada neste estudo é evidenciada não apenas por conta dos riscos relacionados ao processo em questão de alocação de projetos, mas também ao aumento da abrangência do uso da plataforma PUMA. Isso porque espera-se que o volume de propostas submetidas na plataforma aumente à medida que houver uma maior divulgação dos resultados dos projetos.

Essa visibilidade poderá impactar no volume de propostas de projetos que necessitarão de uma alocação adequada, tornando cada vez mais difícil que ela seja suficientemente rápida e assertiva.

Desta forma, na Seção 1.3, é sintetizado o que se pretende alcançar com a pesquisa, por meio da inferência dos objetivos. Essa inferência foi feita a partir do problema de pesquisa e seu respectivo contexto apresentado.

## 1.3 Objetivos

Esta seção contém o objetivo geral e os específicos da pesquisa que delimitam o seu escopo.

### 1.3.1 Objetivo geral

Propor uma ferramenta computacional para tratar os riscos de alocação inadequada de projetos em disciplinas que utilizam a aprendizagem ativa, no âmbito do Curso de Engenharia de Produção da UnB.

### 1.3.2 Objetivos específicos

- a) Entender o contexto e as questões de negócio que permeiam a alocação de propostas de projetos no Curso de Engenharia de Produção da UnB;
- b) Aplicar o processo de avaliação de riscos à alocação de projetos, no âmbito da metodologia PBL adotada no Curso;
- c) Sugerir um modelo de alocação de projetos, considerando o tratamento dos riscos avaliados;
- d) Validar o modelo proposto quanto à recomendação de disciplinas mais adequadas para os projetos;
- e) Verificar se o modelo proposto atende às necessidades das partes interessadas e aos objetivos de negócio do Curso;
- f) Sugerir uma forma de implementação do modelo proposto como um módulo da plataforma PUMA para a alocação de projetos.

Na Seção 1.4, são apresentadas as contribuições esperadas com o alcance dos objetivos desta pesquisa.

## 1.4 Contribuições Esperadas

Uma das contribuições esperadas desta pesquisa é que ela colabore para o tratamento dos riscos de alocação inadequada de projetos, de forma que, no início de cada semestre do Curso de Engenharia de Produção da UnB, as propostas de projetos submetidas na plataforma PUMA sejam destinadas de um modo facilitado para as disciplinas e professores orientadores mais adequados.

Assim sendo, essa pesquisa pode auxiliar na melhora da aplicação da metodologia PBL no Curso de Engenharia de Produção da UnB, visto que esta aprendizagem ativa recebe as propostas de projetos como insumos.

Ademais, uma alocação mais assertiva e rápida de projetos poderá possibilitar a ampliação do volume de propostas submetidas na plataforma PUMA, permitindo uma maior rede de parcerias e interligação de conhecimentos.

Outra contribuição está pautada na possibilidade de aumento da qualidade na formação dada aos alunos, pois a alocação eficaz de projetos pode ajudar o Curso de Engenharia de Produção da UnB a adaptar-se às exigências do mercado, recebendo e destinando as novas necessidades dos *stakeholders* para os alunos por meio de projetos.

Para isso, o trabalho está estruturado da seguinte forma: no Capítulo 2, é apresentado o Referencial Teórico, contendo conceitos, estruturas e técnicas importantes para fundamentar a proposta desta pesquisa; no Capítulo 3, é evidenciada a metodologia a ser utilizada; no Capítulo 4, são apresentados os resultados alcançados; e no Capítulo 5, são abordadas as considerações finais e as propostas de trabalhos futuros.

# Capítulo 2

## Referencial Teórico

Neste capítulo, são apresentados conceitos importantes para fundamentar a proposta desta pesquisa, além das estruturas e técnicas utilizadas em seu desenvolvimento.

### 2.1 Aprendizagem ativa

Há uma necessidade crescente de contextos de ensino e aprendizagem que utilizem abordagens capazes de combinar competências técnicas, que são conhecimentos, habilidades e atitudes focadas nos conceitos teóricos e práticos específicos das áreas estudadas; e competências transversais, que, por sua vez, envolvem capacidades de avaliação crítica, comunicação, tomada de decisão, trabalho em equipe, negociação, resolução de problemas, autogestão, entre outros fatores [3, 22].

O motivo dessa necessidade é que as organizações estão demandando cada vez mais profissionais que possuem competências técnicas e transversais de forma integrada, fazendo com que os sistemas educacionais passem a adotar métodos que são capazes de impulsioná-las [2, 5]. Dentre os métodos existentes, destacam-se os de aprendizagem ativa, pois eles trazem ferramentas e técnicas para ensinar e envolver os alunos de forma mais dinâmica nos seus próprios processos de aprendizagem, de maneira que esses últimos passem a ter o papel central na execução das atividades [23].

Diante disso, Silva et al. [24] destacam os métodos de aprendizagem ativa existentes que vêm sendo utilizados nos últimos anos, aplicados especificamente em disciplinas voltadas às áreas do conhecimento da Engenharia de Produção. Os achados desses autores são apresentados na Tabela 2.1.

Tabela 2.1: Ferramentas e técnicas adotadas pelos métodos de aprendizagem ativa.

Método	Ferramentas e técnicas
<i>Flipped Classroom</i>	Videoaulas de conteúdo; controlar as variáveis demográficas; atividades com pequenos grupos; e instrutor apresenta o problema e alunos buscam a solução.
<i>Problem-Based Learning</i>	Grupos de discussão; instruções diretas; problemas reais resolvidos por meio de projetos ( <i>Project-Based Learning</i> ); e estudos de caso.
<i>Cooperative Learning</i>	Designs alternativos para as salas de aula; trabalhos em grupo; e avaliação individual dos alunos.
<i>Blended Classroom</i>	Videoaula de conteúdo; videoaula de atividades; controle de variáveis demográficas; tecnologias de aprendizagem (computadores e telas interativas); recursos online; designs alternativos para as salas de aula; e atividades com pequenos grupos.
<i>Game-Based Learning</i>	Sala de aula com poucos alunos; instrução detalhada; tecnologias de aprendizagem (computadores e telas interativas); recursos online; e sistema de criação de cenários de jogos.
<i>Inquiry-Based Learning</i>	Sala de aula com poucos alunos; instrução detalhada dos problemas projetados; e recursos online.

Fonte: [24].

Observa-se que, dentre as técnicas de aprendizagem ativa utilizadas pelas universidades, há a resolução de problemas reais por meio de projetos, conhecida como *Project-Based Learning* (PBL). Essa técnica é melhor explorada no próximo tópico, pois ela envolve a utilização de propostas de projetos para a sua aplicação.

### 2.1.1 PBL

A metodologia de ensino-aprendizagem *Project-Based Learning* tem sido cada vez mais utilizada em diversas instituições de ensino ao redor do mundo [25]. Ela coloca o aluno no centro da sua aprendizagem durante a execução de cada projeto, que, por sua vez, pode ser definido como uma tarefa complexa, uma oportunidade ou um problema real que envolve atividades de pesquisa, planejamento, levantamento de necessidades, entre outras etapas de desenvolvimento [26].

Segundo Song e Dow [27], o PBL trabalha para integrar e aplicar os seguintes tipos de conhecimento:

- (i) Novos conhecimentos estruturados abordados em disciplinas;
- (ii) Conhecimentos adquiridos em outras disciplinas;
- (iii) Conhecimentos anteriores baseados na experiência de vida;
- (iv) Novos conhecimentos adquiridos pela autoaprendizagem.

Neste sentido, dentre os diversos meios de aplicação do PBL, tem-se as disciplinas e os projetos finais dos cursos de graduação [3, 5, 6, 28, 29, 30]. Durante a aplicação do

PBL em disciplinas, os alunos são divididos em times para resolver problemas reais de clientes em formato de projetos, nos quais os conhecimentos e as habilidades aplicados e adquiridos geralmente estão associados às áreas do conhecimento de cada disciplina [5, 6].

No caso de projetos finais de graduação que aplicam o PBL, são também utilizados problemas reais de clientes que fornecem os requisitos necessários para os projetos, de modo que os alunos devem aplicar as teorias e as habilidades que foram adquiridas ao longo dos diferentes anos de permanência na universidade, sintetizando diversos conhecimentos [3, 28, 29, 30].

Sendo assim, as empresas e os professores desempenham papéis importantes no desenvolvimento dos projetos desenvolvidos pelos alunos, pois constroem e testam as competências adquiridas durante e depois da educação e treinamento deles para se tornarem profissionais do mercado [2, 3, 30].

Entretanto, a alocação das propostas de projetos, bem como outros processos das universidades, possuem muitos riscos que precisam ser geridos, conforme é apresentado na Seção 2.2.

## 2.2 Gestão de riscos

Todas as atividades, as operações, os processos e os projetos de uma organização envolvem riscos que precisam ser gerenciados [31]. Segundo o ABNT ISO Guia 73 [32], que apresenta um vocabulário relacionado à gestão de risco, um risco pode ser considerado como o efeito da incerteza nos objetivos, de maneira que um efeito é um desvio em relação ao esperado, podendo abordar, criar ou resultar em oportunidades e ameaças.

Assim, o processo de gestão de riscos leva em consideração as incertezas e as possibilidades de circunstâncias ou eventos futuros, sejam eles intencionais ou não, e seus efeitos sobre os objetivos estabelecidos [31]. Esse processo tem o propósito de criação e proteção de valor, de maneira que busca melhorar o desempenho, encorajar a inovação e apoiar o alcance de objetivos [33].

Organizações de todos os tipos e tamanhos enfrentam influências e fatores externos e internos que tornam incerto o alcance de seus objetivos [33]. Neste sentido, Perera et al. [34] trazem que também as instituições de ensino superior sofrem impactos de riscos, pois elas estão expostas à incertezas como outras organizações.

Pavlova e Shaposhnikov [35] reforçam o papel da gestão de riscos para aumentar a competitividade das universidades, pois não apenas as organizações financeiras ou industriais estão sujeitas a riscos, mas também as instituições de ensino superior. Isto porque elas estão sob crescente pressão do estado, da sociedade e de outras partes interessadas, sendo importante que haja uma gestão de riscos integrada a todos os seus processos [35].



Entre os benefícios percebidos ao aplicar a gestão de riscos nas instituições de ensino superior, destaca-se que a identificação, mitigação e atenção contínua aos riscos podem ajudar elas a sobreviverem em um ambiente volátil, a reduzirem a vulnerabilidade delas e a construírem meios para o contínuo sucesso de suas atividades [36].

Entretanto, há diversas abordagens, normas, modelos e *frameworks* com diretrizes consistentes e compatíveis para as organizações implementarem processos de gestão de riscos empresariais eficazes [34]. Dentre essas estruturas, pode-se destacar a ABNT NBR ISO 31000 e o *framework* COSO *Enterprise Risk Management*, que são padrões amplamente utilizados na gestão de riscos [37].

Diante disso, muitos autores trazem que a utilização da ISO 31000 é uma boa prática para gestão de riscos corporativos das universidades [35, 37, 38]. Dentre as vantagens de utilizar a ISO 31000 como padrão para o processo de gestão de risco, está o fato dela poder ser facilmente integrada com outros padrões da ISO, por exemplo, a norma de gestão da qualidade ISO 91001 [37, 39].

Segundo Priyarsono, Widhiani e Sari [37], essa vantagem de integração com outras normas ISO pode ser um dos principais motivos para muitas universidades adotarem a ISO 31000 como padrão para implementar o gerenciamento de riscos que afetam as suas atividades.

Diante desse contexto, ressalta-se, ainda, que ao desenvolver projetos de mineração de dados para instituições de ensino superior, como para qualquer outra organização, é necessário identificar os riscos que podem impactar os mesmos [40]. Além disso, é requerida a descrição das consequências desses riscos e a indicação de quais ações podem ser tomadas para minimizar os efeitos das suas ocorrências [40].

Andrade [41] mostra que é possível associar o processo de mineração de dados, em especial o que segue modelo de referência Cross Industry Standard Process for Data Mining (CRISP-DM), com o processo de gestão de riscos apresentado na ISO 31000, pois, embora cada um deles tenha suas particularidades e etapas, suas semelhanças permitem que eles possam acontecer de maneira complementar e paralela. Ambos os padrões são aprofundados neste trabalho, o CRISP-DM na Seção 2.3.1 e a ISO 31000 na próxima seção.

### **2.2.1 ABNT NBR ISO 31000**

Os padrões da ISO são revisados continuamente para garantir que os princípios e as diretrizes sejam pertinentes e atualizados para atender as necessidades do mercado [34].

Neste sentido, a última versão da ISO 31000 foi publicada em 2018 [33], cujo processo de gestão de riscos é ilustrado na Figura 2.1.

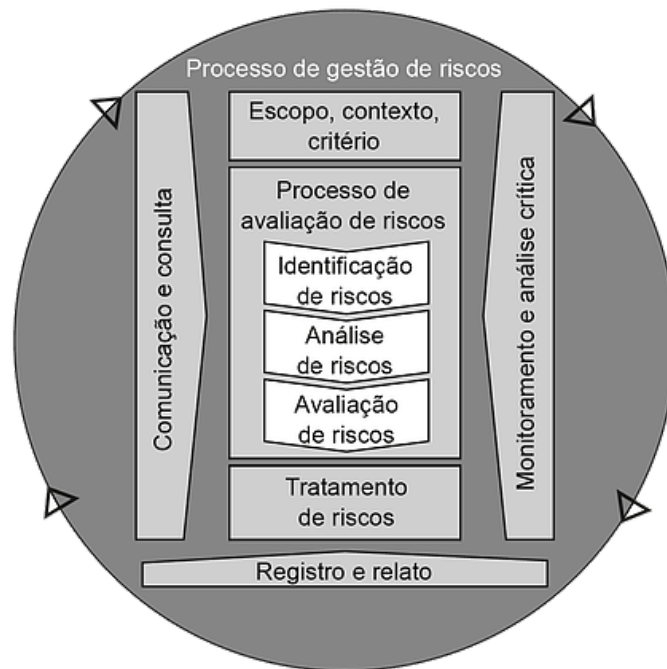


Figura 2.1: Processo de gestão de riscos da ISO 31000.  
 Fonte: [33].

O processo de gestão de riscos apresentado na ISO 31000 é composto pelas seguintes etapas:

- (i) Estabelecimento do escopo, contexto e critérios: tem como objetivo personalizar o processo de gestão de riscos, permitindo que ele seja eficaz e apropriado ao seu escopo definido; aos contextos externo e interno; e aos critérios definidos para avaliar a significância de cada risco e para apoiar os processos de tomada de decisão.
- (ii) Avaliação de riscos: é conduzida de forma sistemática, iterativa e colaborativa, com base no conhecimento e nos pontos de vista das partes interessadas. Esta etapa é feita por meio das seguintes subetapas:
  - Identificação de riscos, que visa encontrar, reconhecer e descrever riscos que possam ajudar ou impedir que uma organização alcance seus objetivos. Nesta fase, informações pertinentes, apropriadas e atualizadas são importantes;
  - Análise de riscos, que foca em compreender a natureza deles e suas características por meio da consideração detalhada de incertezas, fontes, consequências, probabilidades, eventos, cenários e controles, considerando sua eficácia;
  - Avaliação de riscos, que tem como propósito apoiar decisões, envolvendo a comparação dos resultados da análise de riscos com os critérios de risco estabelecidos para determinar onde é necessária ação adicional.

- (iii) Tratamento de riscos: tem o propósito de selecionar e implementar opções para abordar riscos. Além da implementação, é necessário avaliar a eficácia deste tratamento; decidir se o risco remanescente é aceitável; e, se não for aceitável, realizar tratamento adicional.
- (iv) Comunicação e consulta: busca reunir diferentes áreas de especialização para cada etapa do processo de gestão de riscos; e fornecer informações suficientes para facilitar a supervisão dos riscos e a tomada de decisão. Recomenda-se que esta etapa ocorra em todo o processo de gestão de riscos, tendo como partes interessadas externas e internas apropriadas.
- (v) Monitoramento e análise crítica: objetiva assegurar e melhorar a qualidade e eficácia da concepção, da implementação e dos resultados do processo de gestão de riscos. Assim como a etapa de Comunicação e consulta, convém que o monitoramento seja contínuo e a análise crítica periódica em todas as etapas desse processo, além de que seus resultados sejam uma parte planejada, com responsabilidades claramente estabelecidas.
- (vi) Registro e relato: visa que todas as etapas do processo de gestão de riscos e seus resultados sejam documentados e relatados por meio de mecanismos apropriados. Com esta etapa, pretende-se melhorar a qualidade do diálogo com as partes interessadas e apoiar os responsáveis pelo processo de gestão de riscos na tomada de decisão e no cumprimento de suas responsabilidades.

Nas atividades que compõem o processo de avaliação de riscos, há uma gama de ferramentas e técnicas que podem ser utilizadas, por isso a ABNT NBR ISO/IEC 31010 [31] destina-se a refletir as boas práticas atuais na escolha destes elementos. As técnicas mais apropriadas para execução do processo de avaliação de riscos na pesquisa são apresentadas na Seção 2.2.2.

### **2.2.2 Técnicas para execução do processo de avaliação de riscos**

Após a leitura dos trabalhos relacionados que foram encontrados na literatura científica [42, 43, 44] e o estudo da norma ISO 31010 [31], o *Brainstorming* e a Análise de Causa Raiz foram consideradas as técnicas mais apropriadas para o processo de avaliação de riscos aplicado na presente pesquisa.

#### ***Brainstorming***

Por meio do *Brainstorming*, é possível estimular e incentivar o livre fluxo de conversação entre um grupo de pessoas conhecedoras de um determinado tema, possibilitando, assim,

a identificação de riscos junto às suas opções de tratamento [31]. Esta técnica é particularmente útil ao identificar os riscos de novas tecnologias, em que não existem dados ou são demandadas soluções inovadoras para os problemas [31].

Sendo assim, as saídas dessa técnica dependem do estágio do processo de avaliação de riscos em que ela é aplicada [31], de modo que, para potencializar o processo de avaliação de riscos, o *Brainstorming* pode ser utilizado em conjunto com outras técnicas e ferramentas, como a Análise de Causa Raiz [42].

### **Análise de Causa Raiz**

A Análise de Causa Raiz busca identificar as raízes ou causas originais dos riscos ao invés de lidar somente com as suas consequências, podendo ser aplicada em contextos que envolvam processos de negócio [31]. Essa técnica tem vários benefícios, por exemplo, ela é adequada em situações em que é necessário projetar riscos em vez de analisar apenas os já ocorridos [42].

A estrutura da Análise de Causa Raiz pode ser construída por meio da técnica dos “5 porquês”, em que é feita repetidamente a pergunta “Por quê?” para gerar camadas de causas e subcausas [31]. Além disso, essa estrutura pode ser aprimorada por meio da sua combinação com outros modelos, por exemplo, uma adaptação do processo de priorização e tratamento de risco utilizado pela *Commercial Aviation Safety Team* (CAST) [44].

Para a execução desta técnica, é necessária a definição de uma equipe de participantes, a qual é formada com base no conhecimento específico necessário para analisar o problema [31]. Como saída, tem-se uma estrutura de causa e efeito das causas percebidas do problema principal analisado [43].

Ao aplicar de forma conjunta o *Brainstorming* e a Análise de Causa Raiz para execução do processo de avaliação de riscos, é preciso levar em conta que a equipe de participantes pode não ter especialistas necessários por conta da falta de disponibilidade dos mesmos; bem como não ter a sua disposição tempo suficiente para a avaliação completa do problema [31]. Além disso, como essas técnicas dependem do fluxo de pensamento dos participantes, torna-se difícil garantir que todos os riscos potenciais foram identificados [31].

Na Seção 2.2.3, são apresentados os riscos inferidos de trabalhos relacionados ao problema de pesquisa.

### **2.2.3 Riscos relacionados à alocação de propostas de projetos no contexto PBL**

Na perspectiva dos projetos finais de graduação desenvolvidos em contexto PBL, riscos comuns associados ao desenvolvimento deles podem ocorrer [3, 9, 45, 28, 46]. Um desses

riscos é a seleção equivocada ou isolada, por parte dos alunos, do professor orientador ou da empresa cujo problema ou oportunidade real será o tema do projeto [3, 46].

Outro risco é inferido do artigo de Liu et al. [45], que é de os alunos não terem conhecimentos e experiências acadêmicas suficientes para entender os escopos dos projetos de pesquisa e nem qual a área de cada possível orientador, fazendo com que seja um desafio para eles a escolha de um projeto.

Desta forma, Afolabi et al. [28] trazem que os alunos do último ano de graduação se deparam com o problema de selecionar um projeto apropriado, ocorrendo, em alguns casos, o risco de escolha de projetos que fogem da área de interesse e da capacidade deles, sendo um desafio a atribuição aluno-supervisor. Neste sentido, desse artigo também é possível inferir outro risco, que é de a alocação de projetos não considerar uma possível diferença entre os interesses dos alunos e professores, pois nem sempre eles têm os mesmos objetivos.

Além disso, Núñez e Ramos [9] destacam que a alocação de projetos finais é um desafio, pois, em muitos casos, há interdisciplinaridade de áreas presentes nestes documentos, no qual é um risco não considerar esse fato.

Sabe-se também que as propostas de projetos têm o risco de não serem alocadas para disciplinas que tratam das áreas necessárias para o desenvolvimento delas ou para professores que possuem conhecimentos suficientes para orientar os alunos quanto aos temas demandados por elas. Esse risco é baseado no trabalho de Pesántez-Avilés et al. [10], no qual infere-se que os professores tendem a buscar atualizações acerca dos assuntos relacionados às disciplinas as quais eles têm experiências ou que focaram em seus currículos, tornando-se, assim, especialistas em algumas áreas.

Todos esses riscos do processo de alocação de propostas de projetos foram compilados na Tabela 2.2.

Tabela 2.2: Riscos relacionados à alocação de propostas de projetos.

Riscos	Referências
Seleção equivocada ou isolada, por parte dos alunos, do professor orientador ou da empresa cujo problema ou oportunidade será o tema do projeto	[3, 46]
Falta de conhecimentos e experiências acadêmicas suficientes dos alunos para entender os escopos das propostas de projetos e as áreas dos possíveis orientadores	[45]
Escolha de projetos que fogem da área de interesse e da capacidade dos alunos	[28]
Não levar em conta os interesses dos alunos e dos orientadores, os quais nem sempre são os mesmos	
Não levar em conta a interdisciplinaridade de áreas presentes nos documentos dos projetos	[9]
As propostas de projetos não serem alocadas para as disciplinas que tratam das áreas necessárias para o desenvolvimento delas ou para professores que não possuem conhecimentos suficientes para orientar os alunos quanto aos temas demandados por elas	[10]

Observa-se, por meio da Tabela 2.2, que métodos tradicionais de raciocínio dedutivo e seleção aleatória de propostas de projetos, sem recursos adequados e inferência, podem levar à falta de eficiência na alocação de projetos, pois têm que levar em consideração vários fatores, que muitas vezes são conflitantes [28].

Caso essa alocação aconteça por meio de uma plataforma, outros riscos podem acontecer, principalmente se durante a alocação forem recomendados itens, sejam eles disciplinas, orientadores ou propostas de projetos [47, 48, 49].

Um desses riscos é que os usuários nem sempre fornecem seus perfis de forma completa e atualizada, podendo ter como consequência perfis altamente imprecisos ao longo do tempo [47, 48].

Outros riscos que devem ser considerados durante a implementação de um módulo que recomende alocações para os projetos são apontados por Porter [49], a saber: não ter um bom planejamento para manter e extrair as informações dos bancos de dados; não ocorrer as manutenções necessárias para garantir o funcionamento do sistema; acontecer equívocos quanto aos seus resultados, seja por que eles são errôneos ou simplesmente por que os usuários não ficam satisfeitos com eles; e ter usuários dispostos a manipular ou brincar com os resultados do módulo para o seu próprio benefício.

Desta forma, são apresentados na Tabela 2.3 esses riscos que podem ocorrer caso a alocação de propostas de projetos seja realizada por meio do módulo de uma plataforma.

Tabela 2.3: Riscos relacionados ao módulo de alocação de projetos.

Riscos	Referências
Os usuários não fornecerem seus perfis de forma completa e atualizada	[47, 48]
Dificuldades na alimentação e extração dos bancos de dados necessários	[49]
Negligências na manutenção do módulo	
Resultados equivocados ou que não agradam os usuários	
Ter usuários dispostos a manipular ou brincar com os resultados do módulo para o seu próprio benefício	

Diante desses riscos apresentados nesta seção, observa-se que a tomada de decisão quanto à alocação de projetos precisa ser feita a partir de um processo objetivo, que leva em consideração critérios pré-estabelecidos que procurem garantir uma escolha equitativa [9].

Neste sentido, a utilização de técnicas de mineração de dados e textos, com a ajuda de técnicas de aprendizagem e análise estatística, é um campo de pesquisa emergente para melhorar dos ambientes de aprendizagem, podem, assim, contribuir com a proposição de um processo de alocação de propostas de projetos [9, 10, 26, 45, 28, 29]. Por isso, tanto a mineração de dados quanto a de textos são exploradas na Seção 2.3.

## 2.3 Mineração de dados e textos

O *Knowledge Discovery in Databases* (KDD) refere-se ao processo de encontrar conhecimentos em base de dados, no qual a Mineração de Dados pode ser considerada como uma das suas fases [50]. Essa fase procura encontrar e descrever padrões estruturais em dados, ajudando a explicar eles e a fazer previsões a partir deles, de modo que pode ser automática ou semiautomática [51].

Entretanto, sabe-se que muitos projetos estão em formato de texto, que são dados não estruturados ou semiestruturados, sendo necessária a aplicação de técnicas adicionais de mineração de texto [29]. A mineração de texto é bastante útil quando se quer apresentar informações adequadas em contextos em que há dificuldade de leitura completa de um alto volume de texto, por exemplo, por conta de uma escassez de tempo [51].

Diante do exposto, destaca-se que o processo de mineração de texto se difere do processo de mineração de dados, principalmente no que tange à fase de pré-processamento dos dados [9]. Apesar disso, ambos têm o mesmo foco, o qual é processar grandes volumes de dados ou textos para extrair informações ocultas que podem representar padrões, tendências, relacionamentos, regras e/ou associações, com o intuito de apoiar a tomada de decisão e a resolução de problemas [9].

Assim, é possível destacar as tarefas básicas ou os problemas de mineração de dados, a saber: classificação, regressão, clusterização, sumarização, modelagem de dependências, detecção de desvios e associação [52]. Dentre os problemas existentes de mineração de texto, tem-se a sumarização, classificação e clusterização de documentos [51]. Essas tarefas básicas ou problemas utilizam técnicas de mineração específicas [51, 52].

A fim de possibilitar a aplicação dessas técnicas de mineração da melhor maneira possível, sejam elas de dados ou de textos, alguns modelos de referência são utilizados, por exemplo, o *Cross Industry Standard Process for Data Mining* (CRISP-DM) e o *Sample, Explore, Modify, Model, Assess* (SEMMA) que são populares, estando presentes em muitos estudos [53, 54, 55].

Azevedo e Santos [53] apontam que o SEMMA e o CRISP-DM podem ser vistos como uma implementação do processo KDD. Entretanto, o CRISP-DM, diferente do SEMMA, possui fases explícitas de entendimento dos aspectos de negócio, embora seja possível inferir essa fase do SEMMA [53].

O estudo de Azevedo e Santos [53] mostra ainda que o CRISP-DM é extremamente completo e documentado, de maneira que suas etapas são devidamente organizadas, estruturadas e definidas, no qual possibilita que uma proposta de mineração seja facilmente compreendida ou revisada. Corroborando com essa ideia de completude, o estudo de Shafique e Qaiser [56] aponta que o CRISP-DM é mais completo em comparação com o SEMMA. Neste sentido, o CRISP-DM é melhor detalhado na Seção 2.3.1.

### 2.3.1 CRISP-DM

O modelo de referência CRISP-DM foi desenvolvido em 1996 a partir dos esforços de um consórcio composto por empresas pioneiras na mineração de dados, com o objetivo de fazer com que este processo pudesse ter padrões práticos, bem sucedidos e amplamente adotados [40]. De fato, ele é conhecido pela sua poderosa praticidade, flexibilidade e aplicabilidade ao usar análises para resolver problemas complexos de negócio [54].

Ele é disponível sem custos e não proprietário, de modo que possui o Guia CRISP-DM 1.0 [40] com um passo a passo para aplicá-lo. Esse guia apresenta descrições das fases do modelo, as quais são acompanhadas por explicações detalhadas de tarefas genéricas e específicas necessárias e suas saídas. A visão macro do CRISP-DM é apresentada na Figura 2.2.

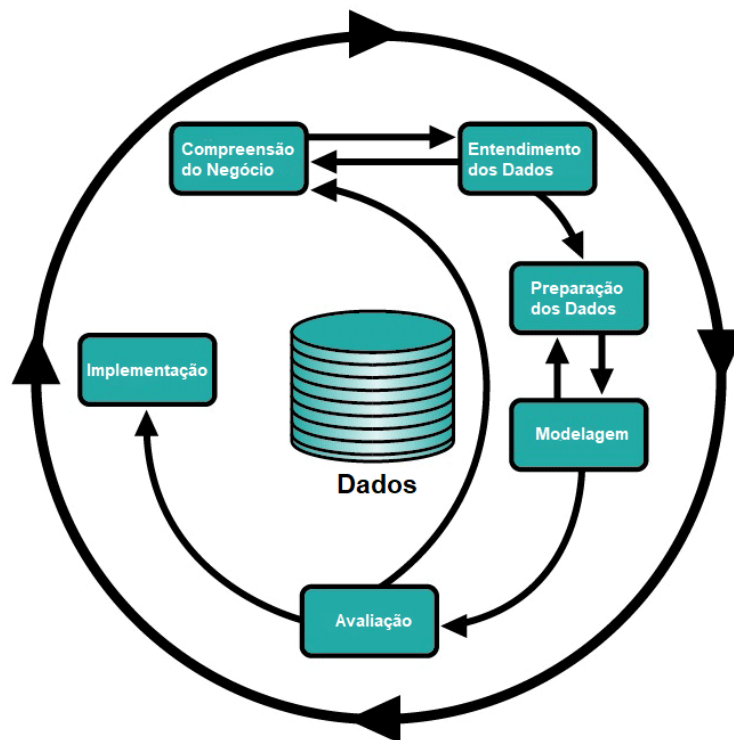


Figura 2.2: Fases do CRISP-DM.  
Fonte: [40].

Conforme pode ser observado na Figura 2.2, o processo de mineração de dados tem natureza cíclica, sendo necessária revisão constante. Além disso, o Guia CRISP-DM 1.0 [40] traz que a sequência entre as fases não é rígida, sendo sempre requerido ir e voltar entre elas. Essas fases são descritas a seguir:

- (i) **Compreensão do Negócio:** concentra-se em compreender os objetivos e os requisitos da pesquisa por uma perspectiva de negócio. O intuito é utilizar este conhecimento



para definir o problema de mineração de dados e traçar um plano preliminar para atingimento dos objetivos.

- (ii) Entendimento dos Dados: inicia-se com uma coleta inicial de dados para uma familiarização com eles. Além disso, essa fase tem o objetivo de identificar problemas de qualidade dos dados; descobrir os primeiros *insights* sobre eles; ou detectar subconjuntos interessantes para formulação de hipóteses relacionadas às informações ocultas.
- (iii) Preparação dos Dados: contempla todas as atividades para construção do conjunto final de dados que será inserido na fase de modelagem. As atividades de preparação dos dados podem ser executadas várias vezes e sem ter uma ordem padrão, as quais incluem a seleção dos dados a partir dos dados brutos iniciais, bem como a transformação e a limpeza deles.
- (iv) Modelagem: nesta fase, várias técnicas são selecionadas e aplicadas, juntamente com a calibração de seus parâmetros para valores ótimos. Ressalta-se que existem várias técnicas que podem ser aplicadas para o mesmo tipo de problema de mineração de dados, de forma que algumas delas possuem requisitos específicos na forma dos dados. Assim sendo, voltar para a fase de preparação de dados pode ser necessário.
- (v) Avaliação: é verificado se o modelo proposto atinge os objetivos de negócio de maneira adequada. Assim, é necessário avaliar detalhadamente o modelo, revisar as etapas executadas para sua construção e verificar se há alguma questão comercial importante que não foi suficientemente considerada. No final dessa fase, é necessário decidir acerca da utilização dos resultados da mineração de dados.
- (vi) Implementação: o modelo é consolidado e incorporado de forma que os *stakeholders* possam utilizá-lo. Esta fase pode ser tão simples como gerar um relatório final com o desenrolar da proposição do modelo ou tão complexo como a implementação dele de forma repetitiva na organização. Nesse último caso sendo importante a elaboração de um plano de controle e manutenção constante a fim de que a organização entenda de antemão quais ações precisam ser realizadas para utilização real do modelo

Ressalta-se que, inicialmente, há diversas técnicas de mineração de dados que podem ser aplicadas para mitigar os riscos relacionados à alocação de propostas de projetos no contexto da presente pesquisa. Isto porque o Guia CRISP-DM 1.0 [40] deixa claro que a aplicação das técnicas de mineração de dados depende das características do problema e da forma dos dados.

Assim, visando dar suporte à seleção de técnicas mais adequadas para a solução a ser proposta pela pesquisa, são apresentados trabalhos relacionados na Seção 2.4.

## 2.4 Trabalhos relacionados

Nesta seção, são apresentados os trabalhos similares ao tema investigado nesta pesquisa que foram encontrados por meio da revisão bibliográfica.

### 2.4.1 Modelos utilizados

É possível encontrar na literatura diversos modelos propostos relacionados com o tema da pesquisa, por exemplo, o trabalho de Alfaro et al. [26]. Esses autores propuseram um módulo de seleção e atribuição de projetos, com o objetivo de dar suporte à metodologia PBL em cursos. Esse modelo foi aplicado em um estudo de caso realizado com um grupo de estudantes da Universidade Nacional de San Agustín de Arequipa, no Peru. Na Figura 2.3, é apresentada a arquitetura do módulo proposto por eles.

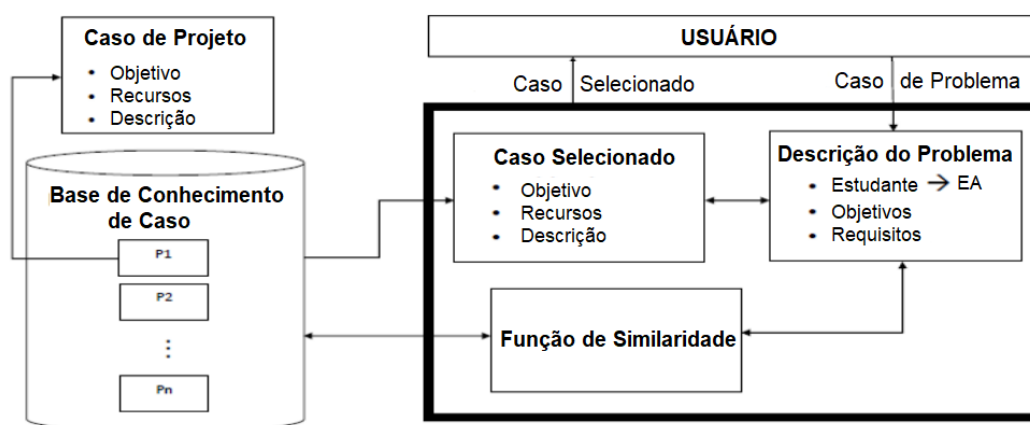


Figura 2.3: Arquitetura do modelo de seleção e atribuição de projetos.

Fonte: Adaptada de [26].

A abordagem apresentada na Figura 2.3 utiliza uma função de similaridade e pesos para comparar os principais dados apresentados a seguir:

- (i) Casos, que podem ser projetos provenientes da aplicação do PBL em anos anteriores, cursos similares, entre outros;
- (ii) Estilos de Aprendizagem (EA) dos alunos, que são identificados pelas interações online deles em uma plataforma *e-Learning*;
- (iii) Descrição do problema de projeto, por exemplo, o objetivo e os requisitos necessários.

O grande benefício da proposta de Alfaro et al. [26] é que ela faz a reutilização de experiências passadas, onde o usuário, tentando resolver um novo problema, reconhece semelhanças com problemas resolvidos anteriormente. Ela foi desenvolvida como uma

extensão do Moodle, que é um sistema de gestão de aprendizagem de código aberto. Os autores também utilizaram a ferramenta de código aberto J-Colibri e o software Colibri Studio.

Não apenas os projetos de cursos que aplicam PBL são objetos de estudo para o desenvolvimento de processos automatizados de alocação, mas também projetos finais de graduação.

Buhari, Valloo e Hashim [3] destacam que, em faculdades particulares na Malásia, o projeto final de cada aluno é composto pelos projetos 1 e 2, onde são exigidas muitas atividades e documentações manuais, além de tomadas de decisão, por exemplo, a definição do orientador para cada projeto. Portanto, os autores propõem a arquitetura de um sistema que automatize as atividades de gestão dos projetos finais de graduação, desde a seleção até a finalização deles. A abordagem utilizada e a arquitetura proposta para esse sistema são apresentadas na Figura 2.4.

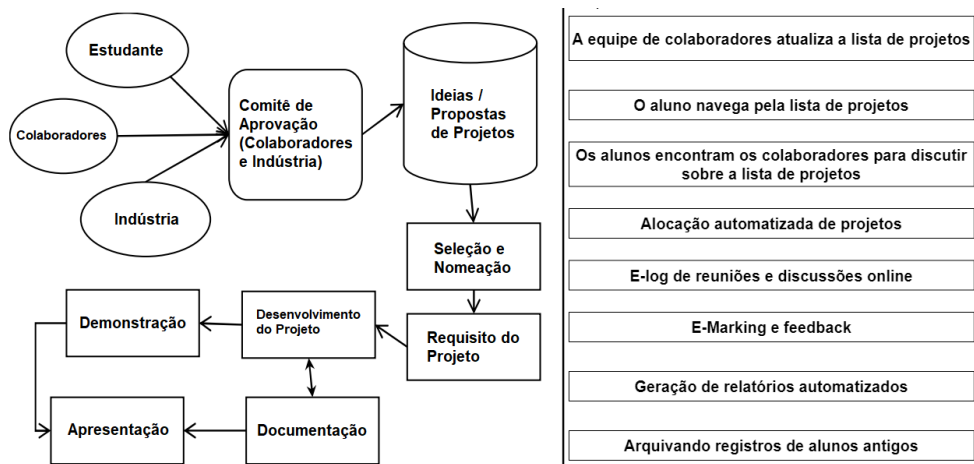


Figura 2.4: Proposta de sistema para gerenciamento de projetos finais de graduação.  
Fonte: Adaptada de [3].

O sistema apresentado na Figura 2.4 foi baseado em três macroetapas, a saber:

- (i) Engajamento com a indústria, onde o corpo docente recomenda a lista de empresas.
- (ii) Seleção do Projeto, que inclui-se também a nomeação do orientador.
- (iii) Execução do Projeto Final de Graduação, que é dividida em cinco partes:
  - Requisito do Projeto, que é a proposta completa do projeto;
  - Desenvolvimento do Projeto, que é o desenvolvimento da solução;
  - Documentação, que é a redação do relatório de desenvolvimento do projeto;
  - Demonstração, que é a aplicação da proposta de solução em vários cenários;

- Apresentação, que abarca a divulgação e apresentação do projeto.

Conforme o estudo de Buhari, Valloo e Hashim [3], o sistema ainda precisa ser implementado com vários tamanhos e cenários de amostra. Após a implantação, os possíveis benefícios elencados pelos autores são a eliminação de trabalhos manuais em papel e a diminuição da presença física para comunicação entre os atores.

Diante da perspectiva apresentada até então, projetos de software também podem ser alocados por processos automatizados. Sun et al. [57] projetaram um sistema de recomendação automática de projetos do GitHub, que é uma plataforma de hospedagem de projetos de software com código-fonte e arquivos com controle de versão. Esse sistema tem o objetivo de fazer com que os desenvolvedores encontrem projetos que atendam aos seus interesses. Esses projetos recomendados podem ser utilizados, completamente ou parcialmente, no desenvolvimento de novos projetos, empregando, assim, menos tempo e esforço. Uma visão geral da arquitetura dessa proposta é apresentada na Figura 2.5.

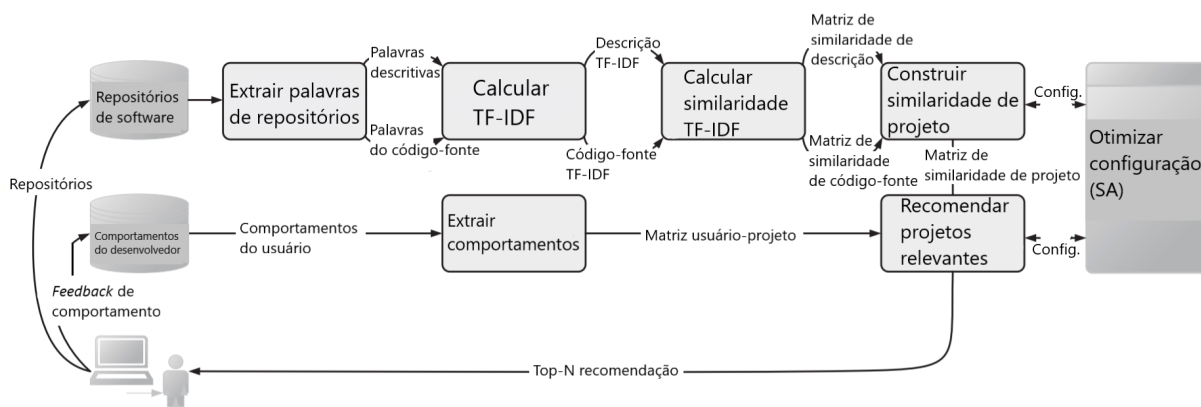


Figura 2.5: Arquitetura de recomendação automática de projetos do GitHub.

Fonte: Adaptada de [57].

A arquitetura apresentada na Figura 2.5 tem os seguintes passos:

- Primeiramente, são extraídas as palavras dos documentos de descrição e do código-fonte dos projetos;
- De posse dos resultados dessa extração, é necessário medir a importância de cada palavra e identificar a similaridade entre elas por meio do *term frequency-inverse document frequency* (TF-IDF);
- Após isso, são formadas duas matrizes de similaridade: descrição e código-fonte. Essas duas matrizes são utilizadas para a criação da matriz de similaridade de projeto;

- (iv) Paralelamente, os comportamentos dos desenvolvedores são extraídos, por exemplo, os projetos que foram marcados com estrela por eles. Esses comportamentos são utilizados para construção da matriz usuário-projeto;
- (v) Por fim, são comparadas as matrizes de similaridade de projeto e usuário-projeto por meio de parâmetros ideias ajustados automaticamente pelo *simulated annealing* (SA), com o intuito de produzir as principais recomendações de projetos de software.

Desta forma, a abordagem proposta por Sun et al. [57] recomenda os N principais projetos mais relevantes com base no comportamento dos desenvolvedores e nos projetos contidos no repositório GitHub. Essa recomendação auxilia para que os projetos progridam mais rapidamente, pois os desenvolvedores interessados poderão descobrir eles por meio da recomendação automática. Ao final de cada recomendação, a abordagem proposta também considera os *feedbacks* dos usuários para melhora desse sistema.

Zhang et al. [58] corroboram com a necessidade de uma recomendação personalizada de projetos para o GitHub. Entretanto, esses autores ressaltam que, na estrutura proposta por Sun et al. [57], foram usados poucos dados e a escalabilidade de seu modelo não foi validada para conjuntos de dados em grande escala. Sendo assim, Zhang et al. [58] propõem a abordagem personalizada de recomendação que é apresentada na Figura 2.6.

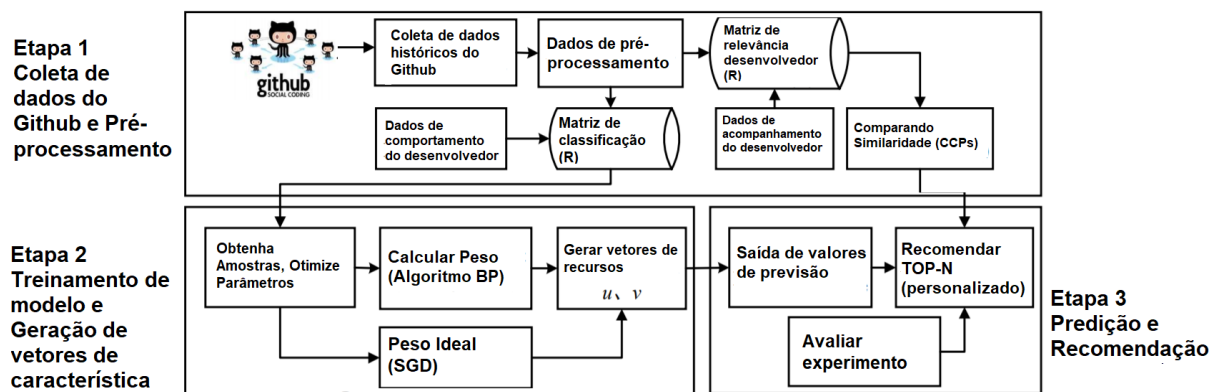


Figura 2.6: Abordagem personalizada de recomendação de projetos do GitHub.  
Fonte: Adaptada de [58].

As três etapas da abordagem da Figura 2.6 são:

- (i) Coleta e pré-processamento de dados, em que são extraídos os dados e criadas as matrizes de relevância (R). O Coeficiente de Correlação de Pearson (CCP) é utilizado para obter a semelhança de interesse entre todos os desenvolvedores;
- (ii) Treinamento de modelo e geração de vetores de recursos, em que os parâmetros do codificador automático profundo são aprendidos para minimizar o erro de reconstrução. Os vetores de recursos desse codificador são aprendidos por *Stochastic*

*Gradient Descent* (SGD), que é calculado pelo algoritmo *Back Propagation* (BP), para ajustar-se à matriz original de pontuação de relevância do projeto do desenvolvedor;

- (iii) Previsão e recomendação, em que são recomendados os Top-N projetos que possam ser de interesse dos desenvolvedores por meio de uma fórmula de recomendação. Essa fórmula é definida com base na matriz de pontuação preditiva e na correlação do desenvolvedor obtida pelo modelo.

Além da questão do volume de dados para processamento, o tempo de execução do projeto também foi considerado na abordagem proposta por Zhang et al. [58]. Isso porque o tempo de execução do projeto pode influenciar nos resultados da recomendação, pois é necessário verificar se os desenvolvedores têm tempo suficiente para cumprir todas as atividades. Desta forma, uma recusa de uma recomendação de projeto não indica necessariamente que não há interesse dos usuários ou que os resultados das recomendações não são precisos, visto que o tempo pode influenciar na não aceitação do projeto.

Entretanto, não apenas as informações do repositório GitHub são utilizadas em recomendações. Sergis e Sampson [47] propõem um sistema de recomendação para repositórios que armazenam objetos de aprendizagem (OA). Essa recomendação é feita por meio dos perfis de competência em Tecnologia da Informação e Comunicação (PC-TIC) dos professores, sendo eles atualizados dinamicamente. O sistema proposto é apresentado na Figura 2.7.

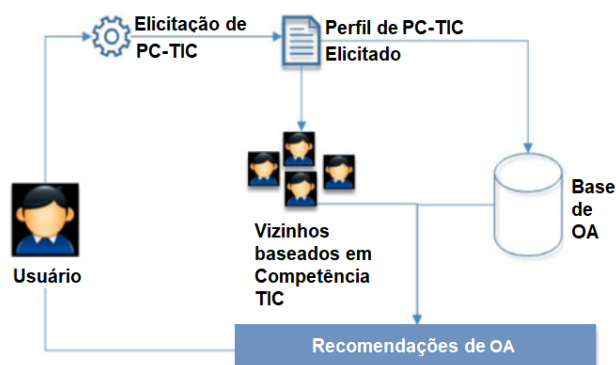


Figura 2.7: Visão geral do sistema de recomendação de objetos de aprendizagem.  
Fonte: Adaptada de [47].

Conforme a Figura 2.7, o sistema é composto por duas camadas principais:

- (i) Camada de Elicitação de Perfil de Competência em TIC, que refere-se à obtenção das competências em TIC do professor ativo no sistema por meio de seus dados de *feedback* ou utilização dos objetos de aprendizagem;

(ii) Camada de Geração de Recomendação de Objeto de Aprendizagem baseada em Competência em TIC, que é executada utilizando os dados de saída gerados da primeira camada, tendo dois intuitos:

- A seleção dos vizinhos do professor ativo no sistema em termos de semelhança nas competências;
- A ponderação de cada objeto de aprendizagem em termos de sua adequação ao perfil do professor ativo no sistema.

O trabalho de Sergis e Sampson [47] destaca que o perfil dos professores é um elemento importante a ser considerado nos sistemas de recomendação ligados à educação. Isso é um diferencial, pois geralmente o foco fica apenas no perfil dos alunos. Esses autores reforçam que, embora os alunos sejam realmente o foco principal dos processos de aprendizagem, outros atores, por exemplo, os professores, também são importantes para a implementação bem-sucedida de procedimentos educacionais.

Embora não tratem de alocação de projetos, Pesántez-Avilés et al. [10] propõem um sistema de revisão e análise de perfis de professores em relação às características das disciplinas dos cursos, visando dar suporte à alocação de professores. Esse sistema funciona por meio de critérios de seleção que são utilizados como diretrizes de alocação. A estrutura geral do sistema proposto pelos autores é apresentada na Figura 2.8.

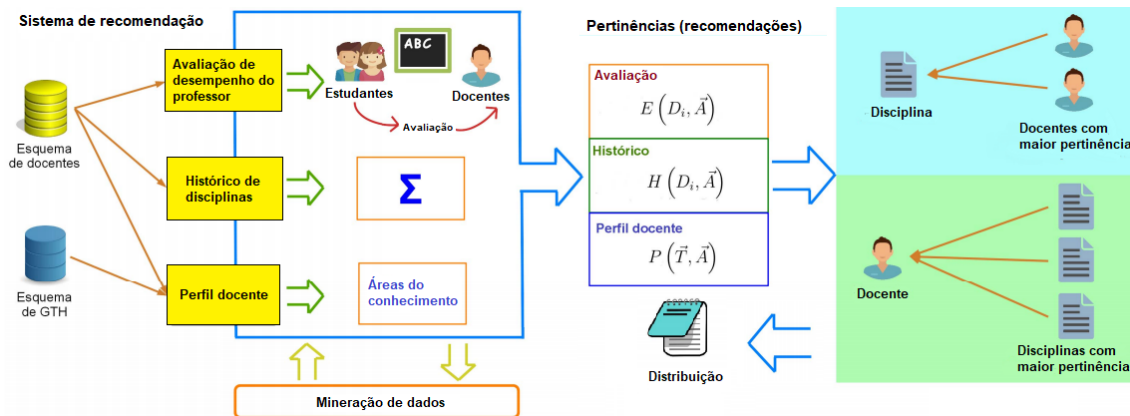


Figura 2.8: Esquema do sistema para recomendação de professores de disciplinas.

Fonte: Adaptada de [10].

Como pode ser visto na Figura 2.8, o sistema é organizado em vários módulos, sendo os mais importantes:

- (i) Gerenciamento de dados, que é responsável por extrair e padronizar as informações das fontes de dados, por exemplo, dos esquemas de docentes e de Gestão de Talento Humano (GTH), além de definir os atributos necessários para analisar a relevância entre os pares professor-disciplina;

- (ii) Recomendador, que é responsável pelo tratamento e processamento das informações fornecidas pelo módulo de Gerenciamento de Dados;
- (iii) Módulo da Web, que é a interface onde o usuário pode pesquisar e inserir as disciplinas ou professores de interesse para obter recomendações do sistema;
- (iv) Relatórios, que fornece informações, muitas delas gráficas, para dar suporte a tomada de decisão acerca da distribuição dos professores nas disciplinas.

O diferencial do sistema proposto por Pesántez-Avilés et al. [10] é que ele utiliza informações acerca da afinidade entre as áreas do conhecimento que vinculam o professor e a disciplina a ele atribuída; da experiência dos professores no exercício em sala de aula; das avaliações de desempenho recebidas pelos professores; e das titulações de cada professor.

A interação entre as informações que vinculam os professores e alunos para alocação de projetos também é explorada por Liu et al. [45]. Esses autores implementaram e testaram um sistema de recomendação com 126 estudantes de graduação para destinar projetos de pesquisa. Esse sistema utiliza dados do ScholarMate, uma rede de colaboração acadêmica. A estrutura do sistema é mostrada na Figura 2.9.

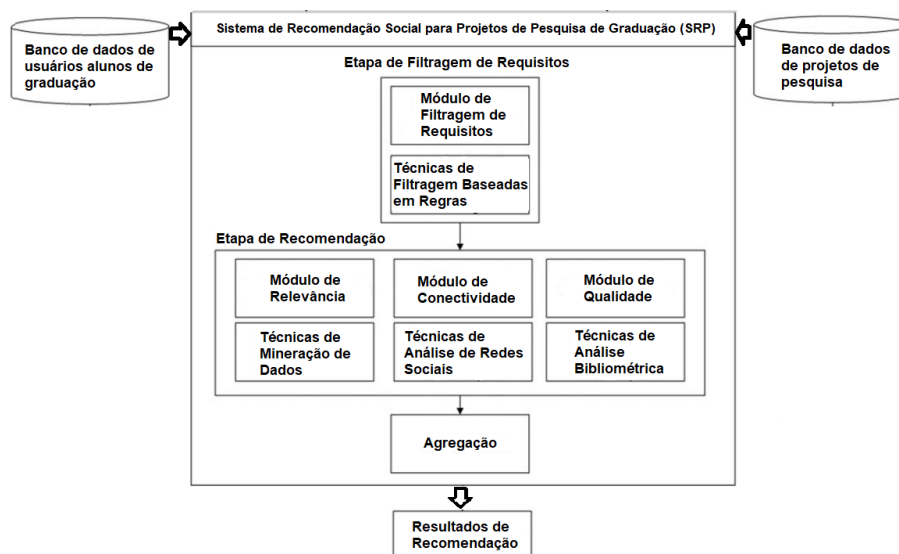


Figura 2.9: Estrutura do sistema para recomendação de projetos de pesquisa.

Fonte: Adaptada de [45].

A estrutura da Figura 2.9 segue as seguintes etapas:

- (i) As informações dos alunos de graduação e dos projetos de pesquisa são coletadas primeiro. Dos alunos, são retirados os interesses de pesquisa, as conexões sociais e as qualificações educacionais. Já as informações do projeto de pesquisa incluem o



conteúdo do projeto, os requisitos para o participante e as informações do orientador. Estas últimas são relacionadas à especialização, colaboração e avaliação da qualidade do orientador em pesquisas.

- (ii) A Etapa de Filtragem de Requisitos pode fazer com que a recomendação de alguns projetos não seja feita, pois há casos em que as informações dos graduandos não têm similaridade com os requisitos das pesquisas, mesmo que eles tenham interesse nelas.
- (iii) Na Etapa de Recomendação, o objetivo é recomendar projetos de pesquisa que tenham qualidade, conexão social e relevância para os alunos de graduação. Assim, essa etapa é composta pelos módulos de relevância, conectividade e qualidade, os quais utilizam, respectivamente, técnicas de mineração de dados, de análise de redes sociais e de análise bibliométrica. Após isso, os resultados desses três módulos são agregados para gerar a lista de recomendação final. Com a lista resultante do sistema, os alunos de graduação podem tomar suas próprias decisões.

O grande diferencial do sistema proposto por Liu et al. [45] é que ele considera os interesses de pesquisa dos estudantes; a possibilidade de que eles se conectem com os supervisores e as universidades; e a qualidade de pesquisa dos supervisores. Entretanto, a abordagem proposta não é dinâmica, de forma que os resultados não são atualizados automaticamente de acordo com os *feedbacks* dos alunos de graduação.

Conforme visto, a recomendação de projetos pode funcionar como apoio à tomada de decisão. Assim, Afolabi et al. [28] apresentam um módulo de suporte à decisão baseado na web que automatiza o gerenciamento de projetos de último ano e atua como ferramenta de apoio ao usuário, pois ele oferece sugestões acerca de tópicos e supervisores de projetos. Na Figura 2.10, é apresentada a visão geral desse módulo.

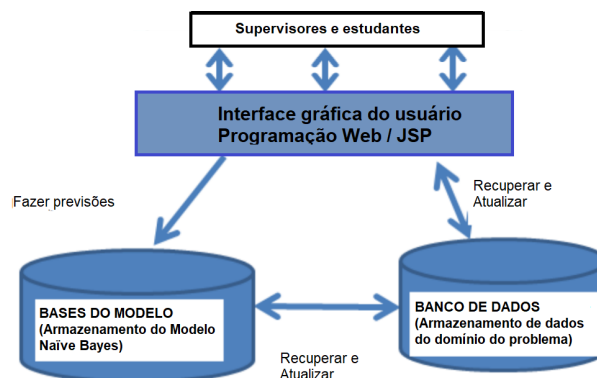


Figura 2.10: Sistema de apoio à decisão para gerenciamento de projetos de último ano. Fonte: Adaptada de [28].

O módulo apresentado na Figura 2.10 tem várias interfaces, a saber:

- (i) Na primeira interface, os usuários podem enviar mensagens, pesquisar por projetos disponíveis com base em um critério de pesquisa específico e carregar novas informações de projetos;
- (ii) Há outra interface que recebe os detalhes dos alunos, por exemplo, o número de matrícula e as notas em disciplinas. Ao clicar no botão enviar, esses dados e outros acerca dos temas e áreas de especialização dos orientadores de projetos são processados pelo algoritmo Naïve Bayes, que é baseado no teorema de Bayes;
- (iii) A última interface é a página de saída da previsão que, por sua vez, apresenta os dados das inferências feitas para o usuário. Nesse caso, as sugestões são o tópico de projeto e o orientador para ele.

Segundo Afolabi et al. [28], o módulo proposto é capaz maximizar o potencial dos alunos e professores para resolver problemas relacionados à indústria e sociedade. Além disso, ele está inserido em uma plataforma rica em protótipos de projetos, ajudando na automação do gerenciamento deles e nas tomadas de decisão no último ano de graduação. Este módulo foi implementado usando Java Server page (NetBeans IDE) em 1,33 GHz CPU, 1024 MB de memória de sistema e 250 GB de espaço de disco.

O suporte aos alunos no desenvolvimento de projetos não se limita à recomendação de itens para tomada de decisão quanto aos temas e orientadores, englobando também sugestões de recursos que podem ser utilizados no desenvolvimento deles, como o trabalho de Diaz-Mosquera et al. [6].

Diaz-Mosquera et al. [6] propõem um mecanismo de recomendação de contribuição colaborativa, de modo que sugere recursos de suporte ao desenvolvimento de projetos PBL em uma disciplina do curso de Engenharia de Software. Com a Figura 2.11, é possível ter uma visão geral da arquitetura desse mecanismo.

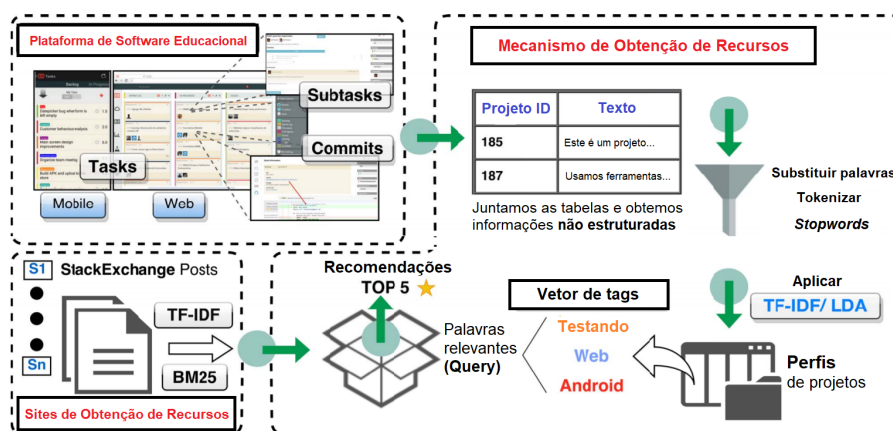


Figura 2.11: Recomendador de recursos de suporte ao desenvolvimento de projetos.  
Fonte: Adaptada de [6].

A arquitetura do mecanismo apresentado na Figura 2.11 é composta por três etapas:

- (i) Primeiramente, existe um procedimento de extração das informações da Plataforma de Software Educacional, onde alunos da Universidade Católica do Chile armazenam todos os dados relacionados aos seus projetos para manter eles sob controle. Essa extração faz a junção dos dados relacionados à *tasks*, *subtasks* e *commits* associados aos identificadores (IDs) dos projetos;
- (ii) A segunda etapa é a de processamento dos dados por meio de técnicas de Linguagem Natural, por exemplo, substituição de palavras, tokenização e eliminação de *stopwords*. Além disso, é gerado um vetor de *tag* usando *term frequency-inverse document frequency* (TF-IDF) ou *Latent Dirichlet Allocation* (LDA). O resultado desse processamento representa o perfil do projeto;
- (iii) Na última etapa, é aplicado o TF-IDF ou BM25 em um conjunto de dados de postagens no StackExchange, que é uma rede de sites de perguntas e respostas sobre tópicos em diversos campos. Com a aplicação dessas técnicas de processamento e o perfil de projeto criado, é gerado um ranking das top 5 relevantes postagens.

Uma das justificativas de Diaz-Mosquera et al. [6] para a proposição desse mecanismo é o fato de que, em muitas ocasiões de aplicação da técnica PBL, os projetos precisam ser desenvolvidos usando ferramentas e recursos que os alunos não têm familiaridade. Com consequência disso, os alunos precisam procurar postagens na web para sanar suas dúvidas, demandando, assim, um direcionamento daquelas mais adequadas para os seus projetos, visto que são inúmeras as existentes.

Além de recomendações, a mineração de projetos pode ter diversos outros resultados, como os apresentados no trabalho de Goh, Gottipati e Shankararaman [29]. Esses autores descrevem um sistema que tem como entrada os dados de projetos de conclusão de graduação em páginas *Wiki*, cuja arquitetura é apresentada na Figura 2.12.

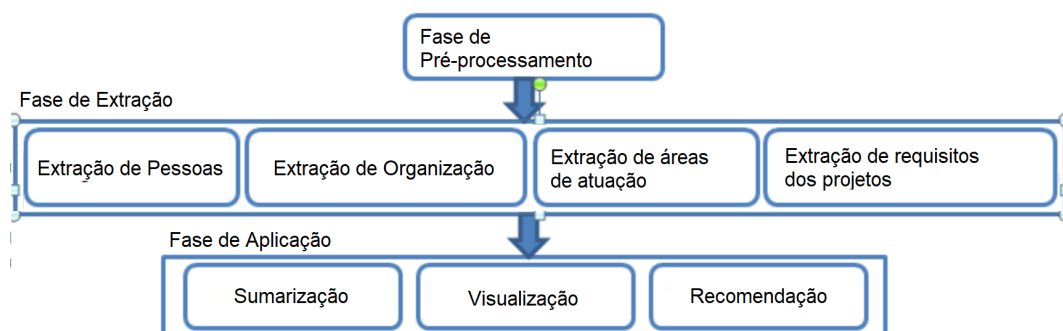


Figura 2.12: Arquitetura do sistema de mineração de projetos de conclusão de graduação. Fonte: Adaptada de [29].

O sistema da Figura 2.12 é composto por três fases:

- (i) Na Fase de Pré-processamento, os dados brutos de texto são preparados para a Fase de Extração, por meio de algumas etapas de processamento de linguagem natural;
- (ii) Na Fase de Extração, são obtidos, via API e técnica de reconhecimento de entidade, os nomes de pessoas, as organizações com as suas áreas de atuação; e os requisitos dos projetos;
- (iii) A Fase de Aplicação resulta na Sumarização, em que são apresentados os dados extraídos da Fase de Extração em formato tabular; na Visualização, onde os dados extraídos são traduzidos em um conjunto de gráficos interativos; e na Recomendação, em que uma lista de projetos antigos semelhantes e os supervisores correspondentes são recomendados pelo sistema para um determinado novo projeto.

Ainda segundo Goh, Gottipati e Shankararaman [29], dentre as vantagens da mineração de projetos finais de graduação, tem-se a identificação de orientadores que têm experiência em semelhantes projetos desenvolvidos para acompanhar as atividades dos alunos, com o intuito de que eles tenham um maior sucesso nesta importante etapa acadêmica. Para a criação do modelo proposto por esses autores, foram utilizados 322 projetos desenvolvidos ao longo de 8 anos em um curso de Sistemas de Informação.

Durante a mineração de projetos para criação de sistemas de alocação, podem ser consideradas as suas áreas profissionais, como no trabalho de Núñez e Ramos [9]. Esse trabalho descreve um processo de mineração de texto para a construção de uma atribuição automática de avaliadores para projetos finais de graduação, denominados Trabalhos Especiais de Graduação (TEG). A arquitetura do sistema oriundo desse processo é apresentada na Figura 2.13.

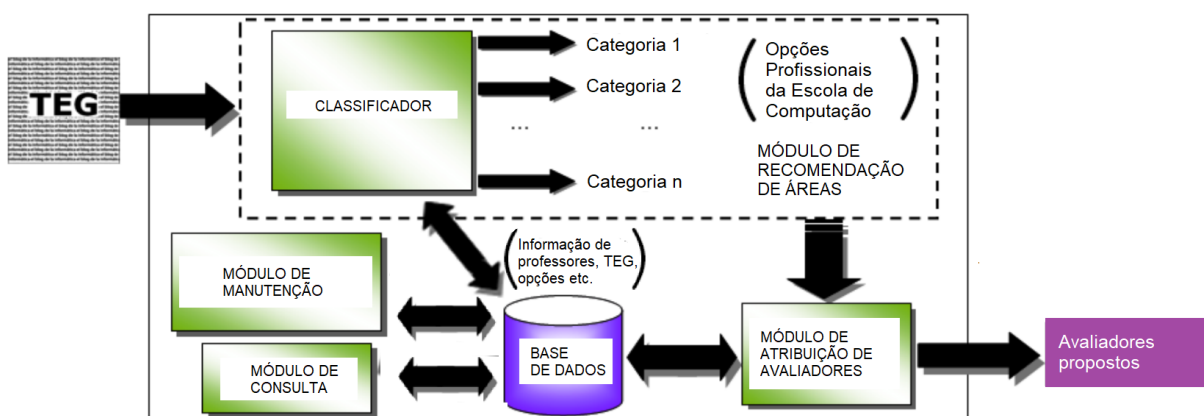


Figura 2.13: Sistema de atribuição de avaliadores para projetos finais de graduação.  
Fonte: Adaptada de [9].

Como pode ser observado na Figura 2.13, o sistema é composto por quatro módulos, sendo eles:

- (i) Módulo de recomendação de áreas: é responsável pelo pré-processamento e classificação dos documentos quanto às áreas profissionais. Ele recebe como entrada o título, resumo (ou introdução) e palavras-chave do projeto a ser classificado. Uma vez que o documento é pré-processado, ele é classificado considerando as áreas profissionais dos cinco vizinhos mais próximos;
- (ii) Módulo de atribuição do avaliador: após a classificação dos documentos, esse módulo seleciona, conforme especificado por um conjunto de regras de eleição, os professores pertencentes às áreas identificadas dos projetos;
- (iii) Módulo de manutenção: é disponível apenas para o administrador da aplicação, pois permite configurar parâmetros e realizar ações que afetam diretamente o desempenho do sistema. Como as atribuições dos projetos para os avaliadores são feitas semestralmente, a principal utilidade do módulo de manutenção é a inicialização ou configuração da aplicação. Além disso, oferece a possibilidade de alteração ou exclusão de informações acerca de professores, usuários e áreas profissionais.
- (iv) Módulo de consulta: permite realizar pesquisas personalizadas sobre o histórico de atribuição de professores como avaliadores dos projetos.

De acordo com os resultados alcançados pelo sistema proposto por Núñez e Ramos [9], observa-se que a recomendação de duas áreas profissionais é considerada eficaz quando há empate nas classificações dos documentos dos projetos, pois nem sempre eles estão relacionados a apenas uma única área. Foram utilizados 198 documentos de projetos para teste e avaliação desse sistema, oriundos de dois anos do curso de Ciência da Computação da Universidade Central da Venezuela (UCV).

No artigo de Canfora et al [59], é apresentado o *Young and newcOmer Developer Assistant* (YODA), um Eclipse *plugin* que recomenda mentores para iniciantes que ingressam em um projeto software, mas que precisam de algum suporte para o desenvolvimento dos seus trabalhos. Além disso, esse *plugin* pode ser utilizado na perspectiva dos gerentes de projeto, pois, em muitos casos, eles desejam ajudar os novatos nos projetos por meio da recomendação de mentores apropriados. O fluxo de informações do plugin YODA é apresentado na Figura 2.14.

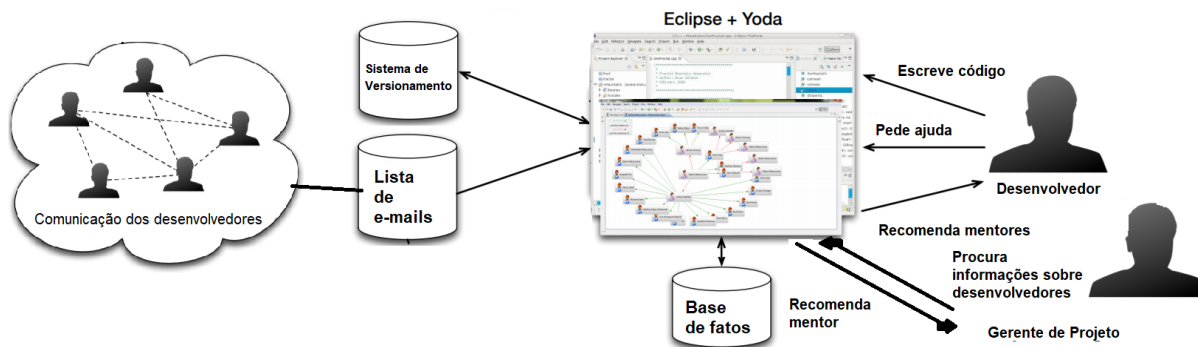


Figura 2.14: Fluxo de informações do *plugin* YODA.  
 Fonte: Adaptada de [59].

O fluxo de informações do YODA, representado na Figura 2.14, é composto pelas seguintes ações:

- (i) Primeiramente, o *plugin* extrai informações de comunicação dos projetos por meio da análise de listas de e-mail; e de alterações realizadas por desenvolvedores mediante a recuperação de *logs* de confirmação do sistema de controle de versão;
- (ii) Depois disso, essas informações são armazenadas em um banco de dados de fatos, pois elas serão utilizadas para identificar candidatos a mentores a serem recomendados aos desenvolvedores iniciantes ou gerentes de projeto;
- (iii) A partir deste ponto, o *plugin* pode ser utilizado quando um desenvolvedor precisa solicitar uma ajuda ou um gerente de projeto precisa procurar informações acerca dos desenvolvedores. O resultado é a recomendação de mentores apropriados.

Um dos diferenciais do *plugin* proposto por Canfora et al. [59] é que ele possibilita que os desenvolvedores novatos dos projetos de software encontrem mentores por meio de uma consulta implícita, clicando com o botão direito nos arquivos de código-fonte que eles estão interessados em trabalhar; ou por meio de uma consulta explícita, escrevendo uma frase em linguagem natural expressando a necessidade de ajuda acerca de uma determinada etapa do projeto. Além disso, o *plugin* retorna, em formato de pontuação, o nível de adequação de cada mentor, apresentando um *ranking* das recomendações.

Bastos [60] propõe um sistema para recomendação de consultores *ad-hoc*, com o intuito deles avaliarem as propostas de projetos de pesquisa submetidas ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esse sistema identifica a similaridade entre os perfis dos consultores, os proponentes e as propostas de projetos. Assim, dentre os insumos desse sistema, estão os currículos da Plataforma Lattes e as propostas de projetos, conforme a Figura 2.15.

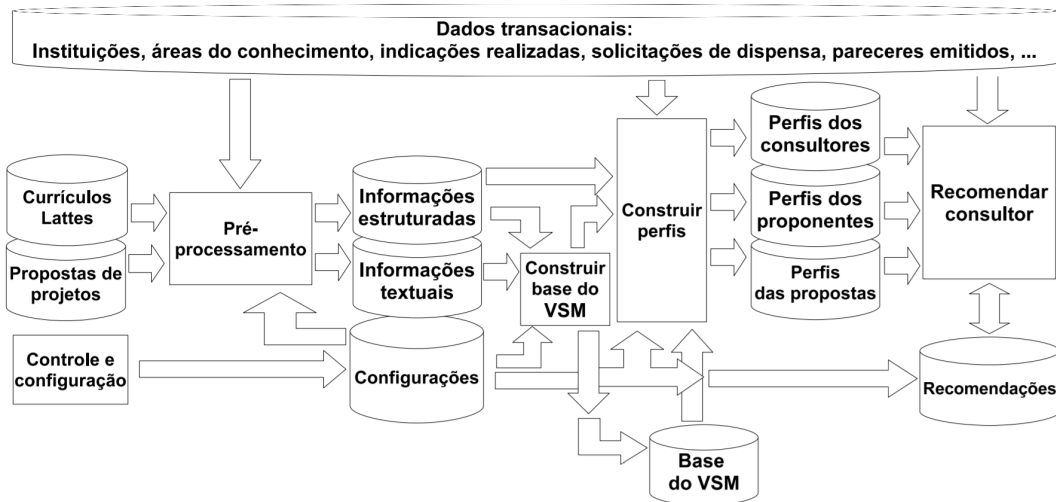


Figura 2.15: Módulos principais da proposta de recomendação de consultores *ad-hoc*.  
Fonte: [60].

A estrutura apresentada na Figura 2.15 possui quatro principais módulos, a saber:

- (i) O módulo de pré-processamento é responsável pela extração e tratamento inicial dos dados que serão utilizados como descritores estruturados, textuais e semiestruturados. Nesses dados são realizadas conversões de formato; padronizações; substituições, remoções e lematização de termos; e construção de estruturas intermediárias.
- (ii) O módulo de construção do *Vector Space Model* (VSM) é o responsável pela criação, redução da dimensão e aplicação das regras de atribuição de pesos da base do VSM; normalização dos vetores resultantes; e construção das matrizes de similaridades consultor-consultor, consultor-proponente e consultor-proposta;
- (iii) O módulo de construção dos perfis é responsável por combinar as informações estruturadas e os vetores VSM em uma representação da constituição dos perfis, com a intenção de cálculo das similaridades;
- (iv) O módulo de recomendação de consultor é responsável por recuperar e recomendar os  $N$  consultores; manter um histórico das recomendações e das ações dos técnicos e dos consultores; e responder aos questionamentos do usuário relativos às recomendações.

Ressalta-se que o sistema proposto por Bastos [60] possui critérios de recomendação divididos em três grupos: positivos, negativos ou excludentes. Os critérios negativos aumentam a probabilidade de recomendação. Em contrapartida, os negativos reduzem essa probabilidade. Já no caso dos excludentes, a ocorrência deles impede a recomendação.

Em resumo, os objetivos gerais e os principais dados dos modelos apresentados nesta seção foram inferidos, por meio da leitura do artigos, e compilados na Tabela 2.4.

Tabela 2.4: Objetivos gerais e principais dados dos modelos apresentados.

Ref.	Objetivos gerais	Principais dados utilizados
[26]	Atribuir projetos que melhor se adéquem às características de cada aluno	Casos de projetos desenvolvidos
		Interação dos usuários com os recursos do Moodle
		Descrição do problema de projeto
[3]	Automatizar as atividades de gestão dos projetos finais de graduação	Lista de empresas e projetos
		Priorização dos estudantes em relação aos projetos inseridos no sistema
[57]	Fazer com que os usuários do GitHub encontrem projetos que atendam aos seus interesses de desenvolvimento de software	Documentos de descrição dos projetos
		Código-fonte dos projetos
		Comportamento dos desenvolvedores quanto aos projetos contidos na plataforma
[58]	Recomendar projetos personalizados para desenvolvedores que utilizam o GitHub	Atributos dos desenvolvedores
		Atributos dos projetos
		Relações sociais entre os desenvolvedores
[47]	Recomendar objetos de aprendizagem armazenados em repositórios	Informações de <i>feedbacks</i> ou de utilização dos objetos de aprendizagem
[10]	Dar suporte à alocação de professores por meio da revisão e análise dos perfis deles em relação às características das disciplinas dos cursos	Avaliações de desempenho e titulações dos professores
		Histórico quantitativo das disciplinas ministradas pelos professores
		Áreas do conhecimento que fazem parte da formação dos professores e do escopo das disciplinas
[45]	Recomendar projetos de pesquisa com base em informações acerca dos alunos de graduação, projetos de pesquisa e orientadores	Interesses de pesquisa, conexões sociais e qualificações dos alunos
		Descrição dos projetos, requisitos do participante e informações do orientador
		Especialização, colaboração em pesquisa e avaliação da qualidade das pesquisas dos orientadores
[28]	Oferecer sugestões de tópicos e orientadores para projetos finais de graduação, atuando como suporte à decisão	Notas de desempenho dos alunos em disciplinas e suas áreas de interesse
		Áreas de especialização dos orientadores
		Temas dos projetos
[6]	Recomendar recursos de suporte ao desenvolvimento de projetos em uma disciplina da Engenharia de Software	Informações relacionadas às atividades dos projetos que são armazenadas em uma plataforma
		Postagem de perguntas e respostas sobre tópicos em diversos campos do conhecimento
[29]	Minerar projetos finais de graduação para descobrir conhecimentos e tendências neles; além de recomendar os que são semelhantes para novos projetos	Informações dos projetos finais de graduação em páginas <i>Wiki</i>
[9]	Atribuir avaliadores para projetos finais de graduação	Título, resumo (ou introdução) e palavras-chave dos projetos finais de graduação
		Áreas profissionais dos professores
[59]	Recomendar mentores para iniciantes que precisam de suporte no desenvolvimento de projetos de software	Comunicação dos desenvolvedores em projetos
		Alterações em projetos por desenvolvedores
		Experiências dos desenvolvedores em mentorias
		Solicitação de ajuda em linguagem natural
		Arquivos de código-fonte
[60]	Recomendar consultores <i>ad-hoc</i> para avaliação de propostas de projetos	Currículos Lattes dos consultores e proponentes
		Propostas de projetos de pesquisa



Observa-se pela Tabela 2.4 que há diferentes meios para elaboração de um processo automatizado de alocação de propostas de projetos, a depender dos objetivos, das bases de dados e do contexto no qual esse processo será aplicado. Neste sentido, há também uma gama de técnicas de mineração de texto que foram utilizadas nos trabalhos relacionados.

Dentre essas técnicas, aquelas aplicadas na elaboração da proposta desta pesquisa são apresentadas na Seção 2.4.2.

## 2.4.2 Técnicas aplicadas para preparação de dados

Os trabalhos relacionados apresentam diversas técnicas que foram aplicadas em tarefas de preparação de dados, com o intuito de atender os requisitos dos modelos propostos. Essas tarefas podem envolver a definição de espaço de termos; a atribuição de pesos aos termos; e a representação dos dados textuais.

### Definição de espaço de termos

A definição de espaço de termos busca uma diminuição no volume de dados, sem prejudicar o desempenho dos modelos. Deste modo, algumas tarefas de pré-processamento de dados foram encontradas na literatura, a saber:

- **Tokenization**, que envolve a quebra dos dados textuais em suas palavras constituintes [29, 51], de modo que o documento é tratado como *strings* que são particionadas em uma lista de *tokens* [61].
- **Remoção de stopwords**, que são palavras irrelevantes por serem extremamente comuns, sendo de pouco valor para diferenciação dos documentos, por exemplo, artigos, pronomes, preposições e conjunções [29, 61, 62]. As *stopwords* estão contidas em listas específicas, denominadas *stoplists*, que podem ser diferentes a depender dos seus idiomas [51, 60].
- **Normalização de termos**, que pode envolver a remoção de sinais de pontuação, caracteres especiais e acentos; a substituição de “ç” por “c”; a remoção de excessos brancos; e a conversão para letras maiúsculas [9, 60].
- **Stemming**, que busca remover sufixos e agrupar os lexemas (grupos de termos com os mesmos significados conceituais) [9, 60]. Para derivação dos termos até os seus radicais, pode ser utilizado o algoritmo de Porter [9, 29, 62, 63].

Após apresentarem formas para definir os primeiros termos que representam os documentos, os trabalhos relacionados trazem métodos de determinação dos pesos desses termos.

## Atribuição de pesos aos termos

A determinação de pesos dos termos pode ser feita por meio de diferentes métodos estatísticos. O método mais utilizado nos trabalhos relacionados foi o *Term Frequency-Inverse Document Frequency* (TF-IDF) [60, 6, 57].

- **TF-IDF** é uma medida amplamente utilizada na recuperação de informações, de modo que a relevância de um termo depende do número de suas ocorrências no documento e do número de documentos no *corpus* para os quais o termo ocorre pelo menos uma vez [6, 51, 57, 60, 62, 64]. Essa medida é normalmente definida pela Equação 2.1.

$$w_{ij} = tf_{ij} * \log \left( \frac{N}{df_i} \right) \quad (2.1)$$

Onde:

- $w_{ij}$  é o peso do termo  $i$  no documento  $j$ ;
- $tf_{ij}$  é a frequência do termo  $i$  no documento  $j$ ;
- $N$  é o número de documentos no *corpus*;
- $df_i$  é o número de documentos no *corpus* onde o termo  $i$  aparece.

Assim, os termos usados em todos ou em quase todos os documentos são inúteis como discriminadores, mesmo que um documento seja basicamente caracterizado pelos termos que aparecem com frequência nele [51]. Deste modo, os resultados do TF-IDF podem auxiliar na identificação dos termos mais importantes em um documento [57].

Os pesos dos termos determinados pelo TF-IDF são utilizados para compor os vetores do espaço vetorial que representam os documentos.

## Representação dos dados textuais

A representação de documentos mais comumente utilizada nos trabalhos relacionados foi realizada por meio do *Vector Space Model* (VSM) [9, 29, 60].

- **VSM** é uma forma de representação dos documentos em vetores de termos, de modo que para cada termo é atribuído um valor que equivale ao seu peso no conjunto de dados [9, 62, 64]. Na Figura 2.16, é possível ver a ilustração de vetores de três documentos: “carro rápido”, “carro vermelho” e “carro vermelho rápido”.

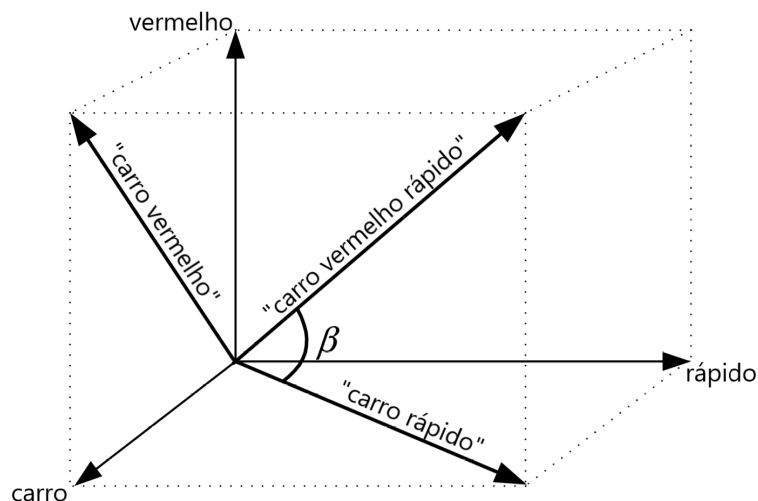


Figura 2.16: Visualização do VSM.  
Fonte: [65].

Observa-se que o ângulo  $\beta$  indica a similaridade entre os documentos representados por vetores n-dimensionais de termos significativos extraídos deles, fazendo com que cada documento seja sua coordenada neste espaço vetorial [60].

Após a representação textual na estrutura do VSM, os trabalhos relacionados buscam definir as similaridades entre os dados textuais por meio de diferentes medidas e métodos, conforme é apresentado na Seção 2.4.3.

### 2.4.3 Medidas e métodos aplicados na modelagem

É possível inferir da literatura diversas medidas e métodos que podem ser utilizados na modelagem, sendo parte deles aplicados para classificar e recomendar itens por meio de informações de projetos.

#### Medidas

Em relação a definição das similaridades entre os dados textuais, duas medidas ganham destaque nos trabalhos relacionados:

- **Similaridade de cossenos**, que é comumente utilizada para calcular a semelhança entre vetores formados por dados textuais [6, 29, 57, 60, 66, 67]. Essa similaridade pode ser dada pela Equação 2.2.

$$sim_{cossenos}(\vec{d}_i, \vec{d}_j) = \frac{\sum_{k=1}^n w_{ki} * w_{kj}}{\sqrt{\sum_{k=1}^n w_{ki}^2} * \sqrt{\sum_{k=1}^n w_{kj}^2}} \quad (2.2)$$

Onde:

- $w_{ki}$  e  $w_{kj}$  são, respectivamente, os pesos do termo  $k$  nos documentos  $d_i$  e  $d_j$ ;
- $n$  é o número de termos nos documentos  $d_i$  e  $d_j$ .

A Equação 2.2 é a definição padrão para similaridade de cossenos, sendo o produto interno normalizado [45]. Esta medida é indicada para espaços de alta dimensionalidade e esparsos, em que muitos atributos têm valores zero [67].

- **Distância Euclidiana**, que também pode ser utilizada para calcular a similaridade entre dados textuais [9, 67], conforme a Equação 2.3.

$$dist_{euclidiana}(\vec{d}_i, \vec{d}_j) = \sqrt{\sum_{k=1}^n (w_{ki} - w_{kj})^2} \quad (2.3)$$

Onde:

- $w_{ki}$  e  $w_{kj}$  são, respectivamente, os pesos do termo  $k$  nos documentos  $d_i$  e  $d_j$ ;
- $n$  é o número de termos nos documentos  $d_i$  e  $d_j$ .

Essa medida não é considerada adequada para calcular a similaridade entre vetores de documentos que possuem muitos atributos de valor zero (esparsos) [67].

Além dessas medidas apresentadas, é possível encontrar na literatura diversos métodos de aprendizado supervisionado que podem ser aplicados para alocação ou recomendação de itens.

## Métodos

Diversos métodos de aprendizado supervisionado foram utilizados nos trabalhos relacionados, sendo o *K-Nearest Neighbors* (K-NN) e o Classificador Naïve Bayes aqueles presentes nos artigos com maior aderência aos objetivos desta pesquisa.

- **K-NN** é um algoritmo de aprendizagem baseada em instância, de modo que faz a previsão ou generalização de uma classe para um novo dado levando em consideração as classes associadas às suas instâncias mais próximas no espaço de entrada, os denominados vizinhos mais próximos [9]. Essa proximidade é definida em termos de uma medida de distância, por exemplo, a Distância Euclidiana ou a Similaridade de Cossenos [9, 62].

Ressalta-se que pode ser utilizado mais de um vizinho mais próximo (valor de  $k$ ), em que este valor é geralmente determinado experimentalmente, a depender do conjunto

de dados para treinamento do modelo, influenciando, assim, na sua qualidade [51, 9]. Se o conjunto de treinamento é grande, espera-se o uso de mais de um vizinho mais próximo, mas, claramente, isso é perigoso no caso oposto [51].

Uma vez que os K vizinhos estejam determinados, é necessário utilizar algum critério de decisão para selecionar a classe a ser atribuída ao novo dado, podendo ser, por exemplo, a classe mais frequente no conjunto de vizinhos; a classe que receber um número de votos maior que um limite pré-estabelecido; a classe com menor distância média; entre outros critérios [9].

O pseudocódigo do algoritmo K-NN é apresentado na Figura 2.17.

```

{Entrada: D (conjunto de treinamento), K (número de vizinhos)}
Para cada exemplo  $z_i$ 
    Para  $j = 1$  até N % N = número de exemplos de treinamento
         $d_i(z_i, x_j)$  = distância entre  $z_i$  e o exemplo de treinamento  $x_j$ 
    Fim_Para
     $D_z$  = conjunto dos K exemplos mais próximos a  $z_i$  (lista de vizinhos)
     $y_i = \underset{v}{\operatorname{argmax}} \sum_{(x_k, y_k) \in D_z} I(v = y_k)$  % Classifica o exemplo de acordo com a
    % classe majoritária de seus vizinhos
Fim_Para
{Saída: conjunto de exemplos de treinamento classificados}

```

Figura 2.17: Pseudocódigo do algoritmo K-NN.

Fonte: [9].

Embora de implementação simples, o algoritmo K-NN pode ser computacionalmente intensivo, especialmente quando o tamanho do conjunto de treinamento cresce [9, 61]. Além disso, o desempenho do algoritmo é severamente degradado pela presença de ruídos ou recursos irrelevantes [61].

- **Classificador Naïve Bayes** é um algoritmo que se baseia no teorema de Bayes, no qual é assumida a hipótese de que os atributos são condicionalmente independentes dos demais, tal como as suas ocorrências e posições no documento [28, 51, 52, 62, 68]. Neste sentido, a Equação 2.4 pode ser utilizada para estimar a probabilidade de cada classe  $i$  do conjunto de treinamento, dado o documento  $D$ .

$$P(c_j | D) = \frac{P(c_j) P(D | c_j)}{P(D)} \quad (2.4)$$

Onde:

- $c_j$  representa uma classe  $j$ ;

- $D$  é um documento representado por um vetor  $n$ -dimensional de termos ( $D = t_1, t_2, \dots, t_n$ ).

Ressalta-se que denominador da Equação 2.4 não se diferencia entre as classes, podendo ser deixado de fora, e, conforme explanado, a parte ingênua de tal método supõe que os valores dos atributos são condicionalmente independentes uns dos outros, dado uma classe [28, 62]. Matematicamente, isso é representado pela Equação 2.5.

$$P(c_j | D) = P(c_j) \prod_{i=1}^n P(t_i | c_j) \quad (2.5)$$

Em que uma estimativa de  $P(c_j)$  pode ser dada pela Equação 2.6.

$$\hat{P}(c_j) = \frac{N_j}{N} \quad (2.6)$$

Onde:

- $N_j$  é o número de documentos associados à classe  $j$ ;
- $N$  é o total de documentos utilizados no treinamento do modelo.

E em que uma estimativa de  $P(t_i | c_j)$  pode ser dada pela Equação 2.7.

$$\hat{P}(t_i | c_j) = \frac{1 + f_{ij}}{V + \sum_{k=1}^V f_{kj}} \quad (2.7)$$

Onde:

- $f_{ij}$  é o número de vezes que o termo  $i$  ocorreu dentro de documentos da classe  $c_j$  no conjunto de treinamento;
- $V$  representa o vocabulário;
- $f_{kj}$  é o número de vezes que todos os termos  $k$  do vocabulário  $V$  ocorreram dentro de documentos da classe  $c_j$  no conjunto de treinamento.

Apesar do fato de que a suposição de independência condicional de termos nos documentos, geralmente, não ser verdadeira, o Classificador Naïve Bayes é surpreendentemente adequado [62, 68].

Para avaliar o desempenho de um modelo, é necessário o uso de uma estratégia de avaliação, conforme é apresentado na Seção 2.4.4.

## 2.4.4 Estratégias para avaliação de modelos

Por meio da revisão da literatura, foi possível encontrar o método *10-fold cross-validation*, que é considerado um padrão para avaliar o desempenho de modelos [9, 51].

- **10-fold cross-validation** é implementado da seguinte maneira: o conjunto de dados é dividido aleatoriamente em 10 partições de tamanhos semelhantes (*fold*s), de modo que cada partição é utilizada para avaliar o modelo treinado com o resto das partições [9]. Como resultado, tem-se índices de desempenho que são obtidos com a média dos valores calculados para as diferentes iterações de avaliação do modelo [9, 52].

Esses índices de desempenho podem ser calculados a partir da matriz de confusão [9, 51, 52]. Um exemplo dessa matriz para duas classes (“Positiva” e “Negativa”) é apresentado na Figura 2.18.

	Classe real	
Classe predita	Positiva	Negativa
Positiva	VP (Verdadeiros positivos)	FP (Falsos positivos)
Negativa	FN (Falsos negativos)	VN (Verdadeiros negativos)

Figura 2.18: Matriz de confusão para duas classes.  
Fonte: [9].

Observa-se que cada campo da matriz de confusão apresenta parte do número de casos resultantes do modelo, os quais são divididos entre as classes preditas e reais que formam, respectivamente, as linhas e as colunas dessa matriz. Além disso, é possível notar que o melhor cenário de resultados é composto por valores altos na diagonal principal e valores nulos no restante da matriz de confusão [52].

A partir da matriz de confusão, podem ser calculados os índices de desempenho, sendo os seguintes comumente utilizados nos trabalhos relacionados: *Accuracy*, *Precision*, *Recall* e *F-measure* [6, 9, 29, 45, 57, 58, 60].

- **Accuracy**, que mede a taxa de acerto de um modelo, sendo uma métrica padrão para descrever o seu desempenho na solução de um problema por meio de um aprendizado supervisionado [9, 64]. Esta métrica pode ser calculada pela Equação 2.8.

$$Accuracy = \frac{\text{Número de predições corretas}}{\text{Número total de predições}} = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.8)$$

- **Precision**, que permite que o desempenho de cada classe seja medido de forma independente, de modo que é calculada, por meio da Equação 2.9, a taxa de informações recuperadas que são corretas de uma dada classe, dentre o total apresentado ao usuário [9, 51].

$$Precision = \frac{\text{Número de predições corretas de uma classe}}{\text{Número total de predições de uma classe}} = \frac{VP}{VP + FP} \quad (2.9)$$

- **Recall**, que calcula a porcentagem de informações corretas que foram recuperadas de um total de itens pertencentes a uma classe [9, 51]. A Equação 2.10 pode ser utilizada para calcular esta medida.

$$Recall = \frac{\text{Número de predições corretas de uma classe}}{\text{Número total itens reais de uma classe}} = \frac{VP}{VP + FN} \quad (2.10)$$

- **F-measure**, cuja fórmula é apresentada na Equação 2.11, é uma medida de síntese que combina as métricas *Precision* e *Recall* [45, 60].

$$F - measure = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (2.11)$$

Ao avaliar os modelos, pode também se ter como métrica o tempo de resposta, a quantidade de espaço utilizado para armazenamento dos dados e a percepção dos usuários acerca dos resultados [60].

A aplicação de enquetes pode ser uma das formas de coleta de dados das percepções dos usuários e especialistas. Essa enquete pode conter, por exemplo, perguntas sobre a relevância, a transparência, a coerência e o desempenho dos modelos [6, 10].

Ademais, com base nos conceitos teóricos e trabalhos relacionados apresentados, foi elaborada a metodologia da presente pesquisa que é detalhada no Capítulo 3.



# Capítulo 3

## Metodologia da Pesquisa

Neste capítulo, a classificação e as etapas da pesquisa com os procedimentos técnicos são descritas, de forma que se possa entender como os objetivos serão atingidos e os resultados alcançados.

### 3.1 Classificação da Pesquisa

Segundo Gil [69], uma pesquisa pode ser definida como um procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são alvos dela. Esse autor afirma ainda que ela é desenvolvida por um processo constituído de várias etapas que utilizam técnicas, ferramentas e/ou outros métodos científicos.

Entretanto, a metodologia de pesquisa vai além da descrição dos procedimentos empregados nela, pois indica também as abordagens teóricas escolhidas pelo pesquisador para tratar o objeto de estudo [70]. Essas abordagens teóricas influenciam a classificação dos tipos de pesquisa, que varia de acordo com o enfoque dado pelo pesquisador [71].

Assim sendo, Silva e Menezes [72] apresentam os diferentes pontos de vista em que uma pesquisa pode ser classificada, sendo eles: quanto à sua natureza; à forma de abordagem do problema; aos seus objetivos; e à sua estratégia. Neste sentido, a classificação desta pesquisa é apresentada na Figura 3.1.

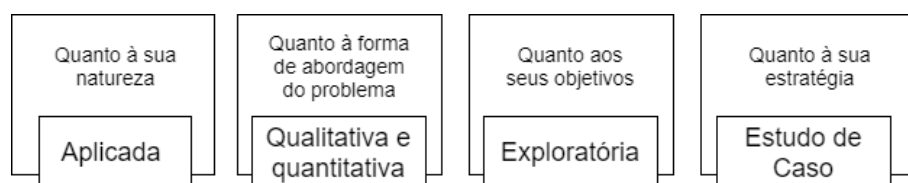


Figura 3.1: Classificação da pesquisa.

A justificativa para a classificação da pesquisa nos diferentes pontos de vista é apresentada a seguir:

- (i) Quanto à sua natureza, ela é aplicada, pois objetiva gerar conhecimentos para o desenvolvimento de um módulo de alocação de projetos dentro da plataforma PUMA, com foco no tratamento dos riscos relacionados a este processo. Esse processo impacta as disciplinas PBL do Curso de Engenharia de Produção da UnB. Deste modo, envolve as necessidades dos professores e alunos do Curso, bem como dos seus agentes externos (*stakeholders*), que são as empresas cujos problemas reais são traduzidos em propostas de projeto;
- (ii) Quanto à forma de abordagem do problema, ela é tanto qualitativa quanto quantitativa, pois ao mesmo tempo que considera que há um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido quantitativamente, busca transpor, por meio de técnicas estatísticas, algumas opiniões e informações em números para classificá-las e analisá-las;
- (iii) Quanto aos seus objetivos, ela é exploratória, uma vez que proporciona maior familiaridade com o problema relacionado à alocação de projetos no contexto da aprendizagem ativa em questão, visando torná-lo explícito. Para isso, a pesquisa envolve levantamento bibliográfico; análise de documentos internos do Departamento EPR, por exemplo, o seu Regimento Interno e o Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB; entrevistas com professores, alunos e agentes externos que têm experiências práticas com o problema; e análise de exemplos que estimulem a compreensão e o desenvolvimento de uma proposta de solução;
- (iv) Quanto à sua estratégia, ela pode ser considerada um estudo de caso, pois envolve a análise profunda da alocação de projetos no Curso de Engenharia de Produção da UnB, de modo que seja possível melhorar o desempenho deste processo por meio do tratamento dos riscos que o afetam.

De posse da classificação da pesquisa no que se refere aos diferentes pontos de vista, definiu-se uma estrutura de pesquisa para o alcance dos objetivos propostos. Essa estrutura é apresentada na Seção 3.2.

## 3.2 Estrutura da Pesquisa

A pesquisa foi estruturada em sete etapas sob o enfoque da norma ISO31000 [33], do modelo de referência CRISP-DM [40] e do trabalho de Andrade [41], de modo que foram realizadas as devidas adaptações para o contexto do estudo. Na Figura 3.2, é apresentada essa estrutura.

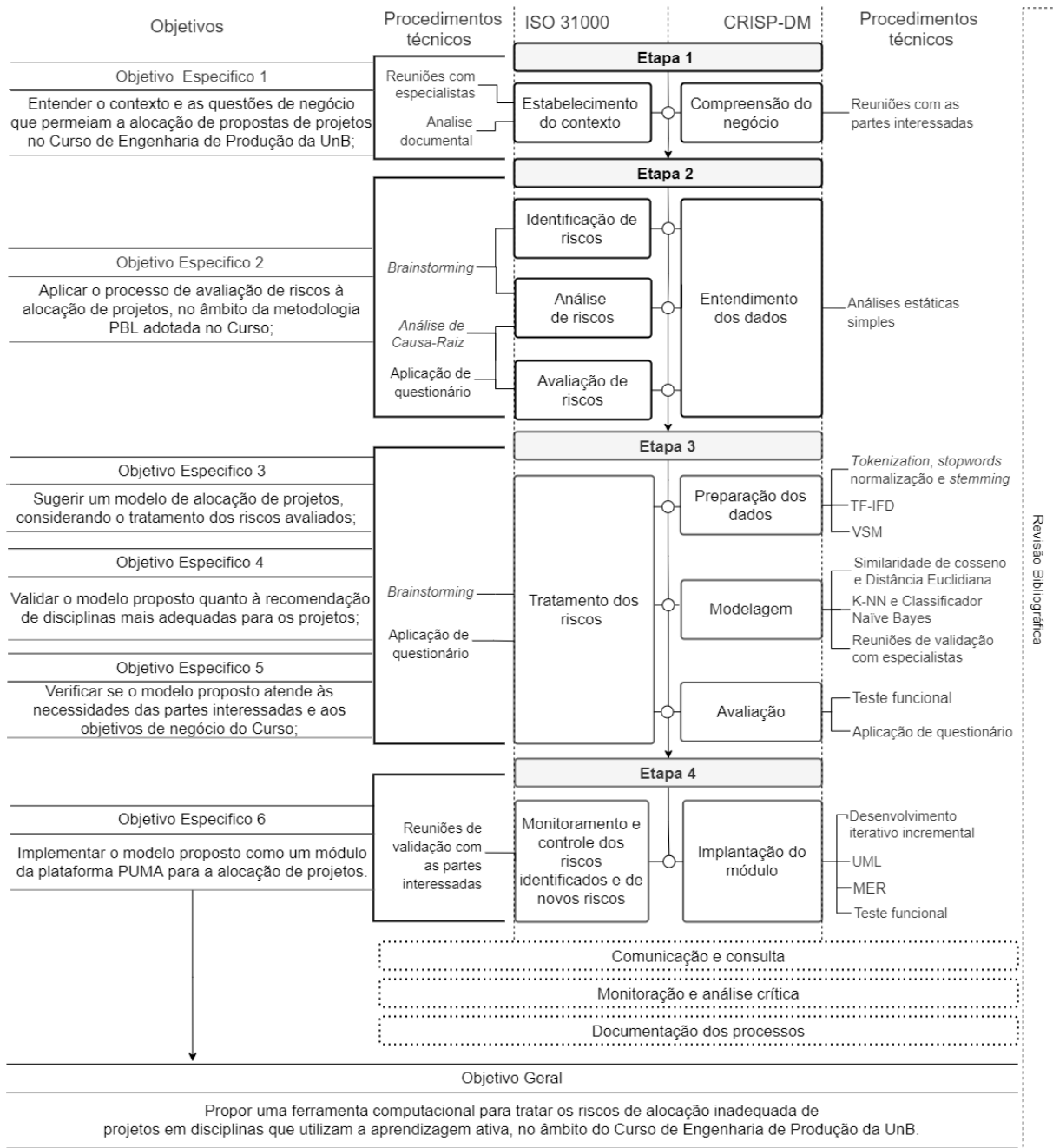


Figura 3.2: Etapas da pesquisa.

Os procedimentos técnicos apresentados na Figura 3.2 foram definidos a partir da revisão bibliográfica realizada. O método utilizado para essa revisão foi baseado na Teoria do Enfoque Meta Analítico Consolidado (TEMAC), cujos autores foram Mariano e Rocha [73]. Isso porque, dentre as possibilidades de uso deste modelo na ciência, tem-se a identificação de metodologias aplicadas com maior frequência em um contexto, além de ser possível descobrir áreas inexploradas ou que têm deficit de trabalhos.

Optou-se por filtrar as buscas dos artigos pelas áreas do conhecimento relacionadas às

Engenharias e Ciências Exatas, bem como pelo espaço temporal delimitado entre os anos 2010 e 2020, objetivando que os resultados fossem os mais assertivos e recentes possíveis. Esses filtros foram aplicados nas bases de dados científicas Scopus, Web of Science (WoS), IEEE Xplore e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), pois elas apresentam artigos de áreas do conhecimento e de contextos similares aos focados neste estudo, além de serem bem-conceituadas nas diversas comunidades acadêmicas [73, 74].

Como palavras-chave, foram utilizados sinônimos e termos similares aos assuntos alocação de projetos; automação; PBL; riscos; e perfis de professores e disciplinas, retornando, assim, um total 308 artigos, sendo 149 da Scopus, 76 da WoS, 60 da IEEE Xplore e 23 BDTD. Desse total, selecionou-se os artigos mais pertinentes aos objetivos da pesquisa por meio da leitura dos seus resumos, de modo que, ao final, 13 trabalhos relacionados foram analisados detalhadamente para um melhor entendimento do assunto e inferências de etapas, funcionalidades e técnicas utilizadas.

Ressalta-se que a sequência das etapas da pesquisa não é rígida, sendo possível ir e voltar entre elas. Essa sequência foi construída de forma que os processos de gestão de riscos e de mineração de dados fossem aplicados paralelamente ao longo da pesquisa, conforme o detalhamento executado nos subtópicos a seguir, que visa apresentar um melhor entendimento do passo a passo a ser desenvolvido em cada uma das etapas.

### **3.2.1 Etapa 1 - Entendimento do contexto e das questões de negócio**

Com intuito de atender ao primeiro objetivo específico da pesquisa, busca-se o entendimento acerca do contexto e das questões de negócio que permeiam a alocação de propostas de projetos do Curso de Engenharia de Produção da UnB. Deste modo, é possível traçar os objetivos, escopos e critérios dos processos de gestão de riscos e de mineração de dados que serão aplicados.

#### **Estabelecimento do contexto**

Para que os processos de gestão de riscos e de mineração sejam eficazes ao aplicá-los conjuntamente na pesquisa, é necessária a compreensão dos contextos interno e externo da alocação de projetos do Curso de Engenharia de Produção da UnB. Deste modo, foram levantadas as seguintes informações: os fatores internos e externos; os direcionadores-chave; as tendências; os relacionamentos; e as percepções, valores, necessidades e expectativas das partes interessadas.

A técnica de coleta de dados utilizada nesta fase foi a análise documental, de modo que sua aplicação ocorreu nos documentos internos do Departamento EPR, como o seu

Regimento Interno e o Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB; e nos artigos publicados acerca da plataforma PUMA e de resultados da metodologia PBL empregada no Curso.

Já para validação de dados, foram utilizadas reuniões com especialistas, conforme é apresentado na Tabela 3.1.

Tabela 3.1: Reuniões para estabelecimento do contexto.

Quantidade	Datas	Amostra	Objetivo
4 reuniões	Entre os primeiros semestres de 2019 e de 2021	2 especialistas em gestão de riscos que ministram disciplinas de PSP, dentre os 8 professores adjuntos do Departamento de EPR que possuem este mesmo papel	Validar as informações acerca dos contextos interno e externo do Curso

As reuniões com especialistas tiveram como intuito a validação de informações acerca dos contextos interno e externo do Curso. Ao longo dos primeiros semestres de 2019 e de 2021, ocorreram quatro reuniões que, em momentos distintos, contaram com a participação de dois especialistas em gestão de riscos. Esses dois especialistas também estão entre os oito professores adjuntos do Departamento de Engenharia de Produção da UnB que ministram as disciplinas de PSP.

Os resultados desta fase contribuem para a compreensão do negócio, conforme a descrição a seguir.

### **Compreensão do negócio**

O foco deste ponto da metodologia é o levantamento de informações de negócio que podem impactar os processos de gestão de riscos e de mineração de dados que serão aplicados na alocação de projetos, tendo em vista o contexto do Curso de Engenharia de Produção da UnB.

Primeiramente, realizou-se a avaliação da situação atual, de modo que foram definidos os recursos humanos, de hardware e de software disponíveis; e os requisitos, premissas e restrições que influenciam a pesquisa.

Após a avaliação da situação atual, para que os processos de gestão de riscos e de mineração de dados sejam eficazes ao aplicá-los na alocação de projetos, foi necessária a delimitação dos seus escopos. Assim sendo, foram inferidos os seguintes pontos: os objetivos, tanto de negócio quanto os gerais para os processos; as decisões que precisam ser tomadas; e os resultados esperados.

Além do estabelecimento dos escopos, tem-se a definição de critérios para ambos os processos. Em relação ao processo de gestão de riscos, a definição de critérios tem como intuito avaliar a significância dos riscos e apoiar a tomada de decisão. Já os critérios de

sucesso para o processo de mineração de dados foram definidos tanto em termos técnicos, relacionados ao nível de precisão das recomendações realizadas pelo modelo proposto e à comparação de resultados das técnicas aplicadas; quanto em termos subjetivos, caso em que foram definidas as partes interessadas que julgarão os resultados.

Já no planejamento inicial, foram selecionadas previamente algumas ferramentas e técnicas que podem ser aplicadas ao problema em questão por meio dos processos de gestão de riscos e de mineração de dados. Além disso, foi definido o cronograma para cumprimento das etapas da pesquisa.

Em relação a seleção das técnicas apropriadas para o processo de gestão de riscos, ela teve como ponto de partida aquelas aplicadas nos artigos que também constam no Referencial Teórico (Capítulo 2) e que tratam de avaliação de riscos de projetos desenvolvidos por estudantes. Além disso, buscou-se na literatura técnicas que pudessem apresentar resultados quantitativos.

Para a seleção de ferramentas e técnicas do processo de mineração de dados, foi realizada a avaliação dos trabalhos relacionados que constam no Referencial Teórico (Capítulo 2). Essa avaliação se deu pelo levantamento das ferramentas e técnicas utilizadas nos trabalhos relacionados que são mais aderentes aos objetivos, ao problema e aos dados da pesquisa, com foco no suporte às demais etapas do processo de mineração de dados.

Ressalta-se que a compreensão do negócio teve também como insumo algumas reuniões de alinhamento da plataforma PUMA, conforme é apresentado na Tabela 3.2.

Tabela 3.2: Reuniões para compreensão do negócio.

Quantidade	Datas	Amostra	Objetivo
7 reuniões	Entre os dias 15/01/2020 e 18/12/2020	3 professores adjuntos e 6 alunos de diferentes disciplinas de PSP	Coleta de informações de negócio relacionadas à alocação de projetos, incluindo o possível uso de ferramentas e técnicas mineração de dados para melhora desse processo de alocação
		2 alunos pertencentes a coordenação do desenvolvimento da plataforma PUMA	
		1 agente externo que já teve problemas resolvidos pelos alunos o Curso de Engenharia de Produção da UnB	
		1 professor e 8 alunos do Departamento de Ciência da Computação da Universidade ESPRIT, localizada na Tunísia	

Assim, entre os dias 15/01/2020 e 18/12/2020, sete dessas reuniões envolveram, entre outros assuntos relacionados ao acompanhamento e planejamento do desenvolvimento de funcionalidade, a coleta de informações de negócio relacionadas à alocação de projetos, incluindo o possível uso de ferramentas e técnicas mineração de dados para melhora desse processo de alocação.

Essas reuniões contaram, em diferentes momentos, com as partes interessadas do Curso de Engenharia de Produção da UnB e, conseqüentemente, do processo de alocação de

projetos, sendo eles: três professores adjuntos e seis alunos de diferentes disciplinas de PSP; dois alunos pertencentes a coordenação do desenvolvimento da plataforma PUMA; e um agente externo que já teve problemas resolvidos pelos alunos.

Além dos participantes do Curso, as reuniões tiveram um professor e oito alunos do Departamento de Ciência da Computação da Universidade ESPRIT, localizada na Tunísia, pois os mesmos fazem parte da equipe de desenvolvimento da plataforma PUMA. Essa equipe se reúne, por videoconferência, com professores e alunos brasileiros da UnB a cada duas semanas durante o período letivo.

A amostra de participantes dessas reuniões que pertencem ao Departamento de Engenharia de Produção da UnB foi retirada da população de oito disciplinas de projeto que fazem parte da grade curricular do Curso. Essas disciplinas são ministradas por oito professores adjuntos, de modo que cada PSP conta, em média, com 30 matriculados e oito clientes externos por semestre. Já a equipe de desenvolvimento da plataforma PUMA, pertencente ao Departamento de Ciência da Computação da Universidade ESPRIT, conta com cerca de dois grupos de quatro alunos supervisionados por um professor da universidade.

Assim sendo, na Etapa 1, buscou-se obter as diferentes perspectivas das partes interessadas relacionadas à alocação de propostas de projetos, de modo que os entendimentos e as definições resultantes norteiam a etapa de avaliação dos riscos e dos dados que é descrita na Seção 3.2.2.

### **3.2.2 Etapa 2 - Processo de avaliação dos riscos e dados**

O desenvolvimento da Etapa 2 busca cumprir o segundo objetivo específico da pesquisa, que tem relação com a aplicação do processo de avaliação de riscos à alocação de projetos, no contexto da metodologia PBL adotada no Curso de Engenharia de Produção. Paralelamente a execução desse processo, é realizado o entendimento dos dados, com o intuito de dar suporte à identificação, análise e avaliação dos riscos.

#### **Identificação de riscos**

O foco da identificação de riscos é encontrar, reconhecer e descrever riscos frente ao processo de alocação de projetos, no âmbito do Departamento EPR da UnB. Com base nisso, as categorias dos riscos foram consideradas.

A identificação dos riscos teve como ponto de partida a revisão bibliográfica realizada na pesquisa. Depois disso, esses riscos previamente identificados foram validados em duas sessões de *brainstorming*, as quais são detalhadas na Tabela 3.3.

Tabela 3.3: Sessões de *brainstorming* para identificação dos riscos.

Quantidade	Datas	Amostra	Objetivos
2 sessões de <i>brainstorming</i> de 1 hora e 30 minutos	10 e 24/03 de 2021	1 professora adjunta que ministra duas disciplinas de PSP e é especialista em gestão de riscos	os riscos previamente identificados por meio da revisão bibliográfica
		1 ex-aluno do Curso de Engenharia de Produção da UnB	
		1 aluno coordenador do desenvolvimento da plataforma PUMA	Identificar novos riscos frente ao processo de alocação de projetos, no âmbito do Departamento EPR da UnB
		1 agente externo que trabalha em uma instituição financeira e também é professora voluntária do Departamento EPR	

Nas sessões de *brainstorming* para identificação dos riscos, participaram uma professora adjunta que ministra duas disciplinas de PSP e é especialista em gestão de riscos; um ex-aluno do Curso; um aluno coordenador do desenvolvimento da plataforma PUMA; e um agente externo que trabalha em uma instituição financeira e também é professora voluntária do Departamento EPR.

Este grupo relativamente pequeno de participantes foi escolhido para que as discussões e ideias ocorressem de uma maneira mais efetiva e rápida, mas possibilitando o envolvimento de diferentes partes interessadas para que pontos de vista distintos pudessem ser ouvidos.

Conforme apresentado na Tabela 3.3, o procedimento técnico de *brainstorming* com as partes interessadas elencadas também foi utilizado para identificação de novos riscos. Assim, para possibilitar que as sessões acontecessem de modo remoto, foram utilizadas as ferramentas virtuais Microsoft Team [75] e IdeaBoardz [76]. Destaca-se que as sessões ocorreram de modo remoto devido a pandemia de COVID-19.

Além disso, foram utilizadas as seguintes perguntas como estrutura básica das sessões de *brainstorming*:

- (i) Todos os riscos levantados na literatura podem ocorrer no âmbito do Curso de Engenharia de Produção da UnB?
- (ii) Quais outros riscos de alocação inadequada de projetos em disciplinas e projetos de graduação não foram identificados por meio da Revisão da Bibliográfica?

Após a identificação de riscos, foi realizada a análise deles, conforme procedimento detalhado a seguir.

### Análise de riscos

Para a análise dos riscos, foi aplicada a Análise de Causa Raiz ou, em inglês, *Root Cause Analysis* (RCA), em duas sessões de *brainstorming* que são apresentadas na Tabela 3.4.



Tabela 3.4: Sessões de *brainstorming* para análise dos riscos.

Quantidade	Datas	Amostra	Objetivo
2 sessões de <i>brainstorming</i> de 1 hora e 30 minutos	31/03 e 07/04 de 2021	1 professora adjunta que ministra duas disciplinas de PSP e é especialista em gestão de riscos	Identificar as causas dos riscos até as suas originais por meio da construção de árvores de causa raiz
		1 ex-aluno do Curso de Engenharia de Produção da UnB	
		1 aluno coordenador do desenvolvimento da plataforma PUMA	
		1 agente externo que trabalha em uma instituição financeira e também é professora voluntária do Departamento EPR	

Deste modo, foram construídas árvores de causas a partir de duas sessões de *brainstorming*, as quais tiveram como participantes as mesmas partes interessadas que contribuíram para a identificação de riscos. A árvore resultante da Análise de Causa Raiz foi desenhada conforme os passos retirados da ISO 31010 [31] e dos artigos de Vanhanen e Lehtinen [43] e Straub [42].

Além disso, as causas raízes dos riscos foram elencadas por meio da técnica dos “5 porquês”, de modo que foi feita a pergunta “Por quê?” para cada risco, gerando, assim, camadas de causas e subcausas. O esquema geral da Análise de Causa Raiz aplicada na pesquisa é apresentado na Figura 3.3.

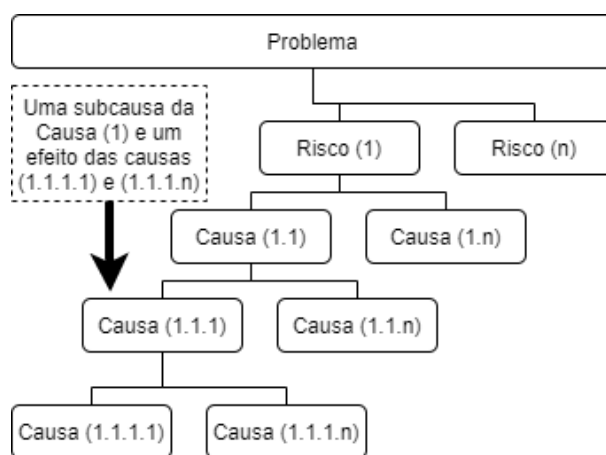


Figura 3.3: Esquema geral da Análise de Causa Raiz.

Fonte: Adaptada de [43].

É possível observar pela Figura 3.3 que os riscos identificados foram associados ao problema objeto da pesquisa, sendo ele a alocação inadequada de projetos. Assim, primeiramente, foi definido o problema, de modo que foram identificados riscos a partir dele. Após isso, por meio de cada risco identificado, foram investigadas as causas.

Desta maneira, as causas ligadas à parte inferior de outra causa são respostas da pergunta “Por quê?” para esta outra causa que está na parte superior. À vista disso, cada causa é um efeito de suas subcausas e cada causa pode ter subcausas em vários níveis. Algumas causas podem ter vários efeitos ou consequências (causas e riscos), sendo a consequência principal o problema de alocação inadequada de projetos.

De posse das árvores de causas elaboradas, foi realizada uma priorização das causas raízes com base na importância em causar o problema atual (Poder 1) e em contribuir para a ocorrência de riscos futuros de alocação inadequada de projetos (Aplicabilidade). O artigo de Pham et al. [44] foi utilizado como base dessa priorização de causas raízes.

As pontuações para essa priorização foram elencadas por meio da Escala Likert de 5 pontos apresentada na Figura 3.4.



Figura 3.4: Escala Likert de 5 pontos utilizada na priorização das causas raízes dos riscos.

Para a coleta dos dados, foi aplicada uma enquete elaborada no Google Sheets [77], a qual é exemplificada na Figura 3.5.

Causas raízes dos riscos de alocação inadequada de projetos	Importância	Importância
	<b>EM CAUSAR O PROBLEMA ATUAL</b> de alocação inadequada de projetos (Poder 1)	<b>EM CONTRIBUIR PARA OCORRÊNCIA DE RISCOS FUTUROS</b> de alocação inadequada de projetos (Aplicabilidade)
Causa Raiz 1	5. Extremamente importante	5. Extremamente importante
Causa Raiz 2	1. Sem importância	4. Muito importante
Causa Raiz 3	2. Pouco importante	5. Extremamente importante
Causa Raiz 4	3. Razoavelmente importante	2. Pouco importante
Causa Raiz 5	3. Razoavelmente importante	4. Muito importante
Causa Raiz 6	4. Muito importante	4. Muito importante
Causa Raiz n	5. Extremamente importante	4. Muito importante

Figura 3.5: Enquete para priorização das causas raízes dos riscos.  
Fonte: Adaptada de [44].

Após o preenchimento da enquete pelas quatro partes interessadas que participaram das duas sessões de brainstorming de análise de riscos, foram multiplicadas as médias das pontuações elencadas para gerar o índice de priorização de causas raízes (Poder 1 x Aplicabilidade).

Os resultados da análise de riscos entrarão como insumos para a avaliação de riscos, que envolve decisões acerca de quais riscos de alocação inadequada de projetos serão considerados prioritários para o tratamento.

## **Avaliação de riscos**

Nesta fase, foi realizada a avaliação dos riscos de alocação inadequada de projetos, visando a escolha daqueles que precisam de algum tratamento adicional para sua eliminação, mitigação ou controle.

Os riscos prioritários foram os que possuem causas raízes com os maiores índices de priorização (Poder 1 x Aplicabilidade), sendo esses últimos resultantes da análise de riscos. Destaca-se que os riscos que também apresentaram causas raízes comuns foram considerados prioritários, com o intuito de otimizar o processo de tratamento. Isso porque, caso as causas raízes desse tipo sejam tratadas, elas podem eliminar, mitigar ou controlar mais de um risco de alocação inadequada de projetos ao mesmo tempo.

Ressalta-se que essa tomada de decisão também foi apoiada pela comparação dos resultados da análise de riscos com os outros critérios de risco que foram estabelecidos na Etapa 1 da pesquisa.

Assim sendo, como saída final da avaliação de riscos, tem-se a lista de riscos prioritários que será apresentada e validada com as partes interessadas para levantamento de ações de tratamento.

Paralelamente a identificação, análise e avaliação dos riscos de alocação inadequada de projetos, será realizada a aplicação de técnicas estatísticas simples nos dados disponíveis para a pesquisa, com o intuito de obter uma maior compreensão deles e descobrir eventuais novos riscos e suas causas.

## **Entendimento dos dados**

O entendimento dos dados será iniciado com o acesso aos dados listados na Etapa 1, como recursos disponíveis para a pesquisa, e o seu carregamento na ferramenta de código aberto Orange 3 [78], que possui componentes para visualização e análise exploratória de dados. Após isso, serão descritos esses dados, incluindo as suas origens, os formatos, os eventuais problemas de extração; a quantidade de registros e campos; as identidades dos campos; e os atributos-chave.

A exploração desses dados também dar-se-á por meio de análises estáticas simples executadas por meio da ferramenta Orange 3, com o intuito de identificar a distribuição e os relacionamentos entre pares dos atributos; os agrupamentos ou subconjuntos existentes; e o status atual de balanceamento dos dados.

A tarefa final consiste na verificação da qualidade dos dados, de forma a entender se eles estão completos e sem erros ou valores ausentes.

Alguns resultados do entendimento dos dados serão plotados em gráficos para indicar as suas características que sugerem ações nas atividades de preparação dos dados, que,

por sua vez, são detalhadas na Etapa 3. Nessa etapa, também é realizado o tratamento dos riscos e a modelagem da proposta.

### 3.2.3 Etapa 3 - Construção do modelo de alocação de projetos com base no tratamento dos riscos

Com a execução desta etapa, pretende-se atender ao terceiro, quarto e quinto objetivos específicos, cujos escopos se concentram na proposição de um modelo de alocação de propostas de projetos considerando o tratamento dos riscos avaliados; sua validação quanto à recomendação de disciplinas e orientadores mais adequadas para os projetos; e, por fim, a verificação do modelo proposto no que se refere ao atendimento às necessidades das partes interessadas e aos objetivos de negócio do Curso. Destaca-se que os dados de entrada desse modelo de alocação de projetos também serão preparados nesta etapa.

#### Tratamento dos riscos

Para o tratamento dos riscos que foram priorizados na Etapa 2 da pesquisa, as partes interessadas, que participaram da avaliação dos riscos, serão instigadas em novas sessões de *brainstorming* a propor ações para tratamento das causas raízes dos riscos de alocação inadequada de projetos.

Essas ações propostas serão avaliadas quanto quão bem elas podem contribuir para a eliminação, mitigação ou controle das causas raízes desses riscos (Poder 2); e a facilidade e o prazo necessário para implementação de cada ação de tratamento. O artigo de Pham et al. [44] também será utilizado com base para aplicação desta estratégia de avaliação das ações de tratamento das causas raízes.

Para a coleta dessa pontuação, será aplicada uma enquete elaborada no Google Sheets, a qual é exemplificada na Figura 3.6.

Causas raízes dos riscos de alocação inadequada de projetos	Ações para tratamento das causas raízes dos riscos de alocação inadequada de projetos	Esta ação irá contribuir para o tratamento da causa raiz (Poder 2)	Esta ação é de fácil implementação (Facilidade)	Classifique esta ação de acordo com o tempo necessário para sua implementação
Causa Raiz 1	Ação 1	5. Concordo totalmente	5. Concordo totalmente	Curto prazo (3 meses)
Causa Raiz 2	Ação 2	1. Discordo totalmente	4. Concordo	Curto prazo (3 meses)
Causa Raiz 3	Ação 3	2. Discordo	5. Concordo totalmente	Médio prazo (6 meses)
Causa Raiz 4	Ação 4	3. Não concordo e nem discordo	2. Discordo	Longo prazo (1 ano)
Causa Raiz 5	Ação 5	4. Concordo	4. Concordo	Longo prazo (1 ano)
Causa Raiz 6	Ação 6	4. Concordo	4. Concordo	Longo prazo (1 ano)
Causa Raiz n	Ação n	5. Concordo totalmente	4. Concordo	Curto prazo (3 meses)

Figura 3.6: Enquete para priorização das ações de tratamento das causas raízes.

Fonte: Adaptada de [44].

As respostas da enquete apresentada na Figura 3.6 serão elencadas por meio de uma Escala Likert de 5 pontos, seguindo a mesma estratégia utilizada na priorização das causas raízes dos riscos que foi executada na Etapa 2.

Após isso, as pontuações médias inseridas na enquete, pelas quatro partes interessadas que participaram das sessões de *brainstorming* de tratamento de riscos, serão multiplicadas para gerar o índice de priorização das ações de tratamento (Poder 2 x Facilidade). Além do uso desse índice, serão priorizadas as ações que forem de curto prazo, pois o período para tratamento dos riscos é de 3 meses, conforme o cronograma elaborado na Etapa 1.

Com a lista priorizada, as ações de tratamento serão implementadas e terão os resultados detalhados ainda neste ponto da pesquisa. Entretanto, parte dessas ações serão colocadas em prática em outras partes da pesquisa, por exemplo, na preparação dos dados, cujo passo a passo é detalhado a seguir.

## Preparação dos dados

A primeira atividade a ser executada na preparação dos dados é a seleção dos atributos e registros que serão utilizados na modelagem da proposta, levando em conta a relevância, a qualidade e as suas restrições técnicas. Ao final dessa tarefa, um quadro é apresentado com os dados que serão incluídos e excluídos no modelo e as justificativas para esta decisão.

A preparação dos dados também inclui a busca pelo aumento da qualidade dos dados, mediante a execução de eventuais ações de melhoria traçadas a partir do entendimento dos dados executado na Etapa 2. Além disso, caso sejam necessários, serão construídos novos registros acerca dos projetos.

Outra tarefa importante a ser executada é a definição do espaço de termos dos dados, fazendo com que seu volume seja menor, mas produzindo resultados idênticos ou similares. Para isso, algumas técnicas de pré-processamento serão aplicadas, sendo elas:

- (i) *Tokenization*, que envolve a quebra dos dados textuais em suas palavras constituintes [29];
- (ii) Remoção de *stopwords*, que são palavras comuns de pouco valor para diferenciar documentos, a partir de uma lista contendo termos em português [9, 29, 60];
- (iii) Normalização de termos, que envolve a remoção ou substituição de caracteres especiais, acentos, cedilhas e excessos em branco; além da conversão para letras maiúsculas [9, 60];
- (iv) *Stemming*, que faz a redução das palavras aos seus radicais [9, 29, 60].

Após isso, serão atribuídos pesos aos termos mediante o emprego do *Term Frequency-Inverse Document Frequency* (TF-IDF) [60]. Os termos e os seus pesos serão utilizados

para a representação dos documentos no modelo *Vector Space Model* (VSM), que resulta vetores de dados textuais [9, 29, 60].

Entretanto, como os textos podem possuir comprimentos diferentes, poderá ser necessário empregar a normalização das coordenadas para obter vetores de comprimento igual a unidade [60, 79]. Isso será apropriado caso haja uma discrepância muito grande entre os tamanhos dos documentos dos projetos já desenvolvidos e as propostas de novos projetos.

Outra estratégia que poderá ser empregada é a eliminação de termos com base na frequência nos dados textuais. Essa eliminação visa a redução de dimensionalidade dos dados, uma vez que seja necessário acelerar o tempo de resposta do modelo de alocação de projetos [60].

Além disso, os dados podem demandar um balanceamento da quantidade de registros das classes, com o intuito de evitar que o modelo de alocação de projetos proposto seja tendencioso para a classe majoritária. Para isso, poderão ser utilizadas as técnicas *undersampling*; *oversampling*; e *oversampling* seguida de *undersampling* [80].

Ao final, serão formatados os arquivos dos dados para atender os requisitos da ferramenta Google Colab [81], que é o ambiente Jupyter Notebook em Python a ser utilizado para a modelagem.

## Modelagem

A modelagem da proposta levará em conta as técnicas selecionadas na Etapa 1; as ações traçadas na fase de tratamento dos riscos; e os resultados da preparação dos dados.

O método *K-Nearest Neighbors* (K-NN) [9] e O Classificador Naïve Bayes [28] serão utilizados para classificação dos projetos, assim como as medidas Similaridade de Cosseno [60] e Distância Euclidiana [9]. Neste sentido, serão realizadas diversas tentativas de construção e execução do modelo a fim de comparar os resultados das técnicas aplicadas e dos parâmetros configurados.

Os resultados serão revisados a cada execução do modelo, que pode ser reconstruído até o atingimento dos objetivos e critérios de sucesso estabelecidos na Etapa 1; e a implantação das ações de tratamento de riscos relacionadas com a modelagem. Além disso, pode ser necessário refazer algumas tarefas de preparação de dados para adaptá-los às tentativas de modelagem.

Conforme dito anteriormente, a modelagem será executada no ambiente Jupyter Notebook em Python 3 por meio do uso da ferramenta Google Colab. Entretanto, as tentativas de construção do modelo também serão executadas na máquina virtual Ubuntu Server 18.04 da plataforma PUMA, a qual é um recurso disponível para a pesquisa. A motivação para execução do modelo também com a tecnologia da plataforma PUMA é medir a sua

velocidade de resposta, visto que ela deve ser adequada para viabilizar a proposta da pesquisa.

Além da velocidade de resposta, o desempenho do modelo será avaliado de acordo com a aplicação do método *10-fold cross validation*; pelos resultados das taxas de sucesso e de erro; e por meio dos valores das medidas *Precision*, *Recall* e *F-measure* [9, 60].

Para aumentar a eficácia da avaliação do modelo sob o olhar mais técnico, um especialista em mineração dos dados e outro em gestão de riscos serão consultados, em reuniões de validação, acerca do cumprimento dos critérios de sucesso estabelecidos.

Após essa avaliação técnica, outras partes interessadas do processo de alocação de projetos do Curso de Engenharia de Produção da UnB também serão consultadas, visando avaliar o modelo sob a perspectiva do negócio, conforme detalhamento a seguir.

## Avaliação

Neste ponto da pesquisa, busca-se avaliar o modelo de alocação de projetos quanto ao atendimento às necessidades das partes interessadas e aos objetivos de negócio do Curso.

Para possibilitar essa avaliação, uma interface dentro da plataforma PUMA será criada para que professores, alunos e agentes externos possam utilizar o modelo para teste funcional. Ao final desse teste, será aplicada uma enquete estruturada por meio de uma Escala Likert de 5 pontos, que terá perguntas baseadas nos trabalhos de Diaz-Mosquera et al. [6] e Pesantez-Aviles [10], cujos indicadores de qualidade norteadores e conteúdos são apresentados na Tabela 3.5.

Tabela 3.5: Perguntas para avaliação do modelo.

Indicadores de qualidade	Perguntas
Relevância	As recomendações foram relevantes para os resultados dos projetos?
Transparência	A origem das recomendações foi clara?
Coerência	As recomendações recebidas foram coerentes?
Desempenho	O desempenho do módulo foi adequado?

Fonte: [6, 10]

Além da avaliação da proposta por meio da enquete, os processos de gestão de riscos e de mineração de dados serão revisados em um quadro comparativo, onde serão diferenciadas as atividades que foram cumpridas e aquelas que deveriam ser repetidas ou acrescentadas em trabalhos futuros. Essa comparação será realizada por meio do passo a passo da norma ISO 31000 [33] e do Guia CRISP-DM 1.0 [40].

Ao final desta avaliação, é decidido acerca da implantação do modelo na plataforma PUMA, de modo a ser um módulo disponibilizado para o Curso de Engenharia de Produção da Universidade de Brasília. Essa implantação seguirá os procedimentos apresentados na Etapa 4.

### 3.2.4 Etapa 4 - Proposta de Implantação do Módulo de Alocação de Projetos na plataforma PUMA

Para atingir o sexto objetivo específico da pesquisa, é necessário implementar o modelo proposto como um módulo da plataforma PUMA para alocação de projetos, de modo a ser utilizado pelas partes interessadas do Curso de Engenharia de Produção da UnB. Ademais, paralelamente a essa implementação, será realizado o monitoramento e controle dos riscos, visando a avaliação da implementação das ações de tratamento prioritizadas e a identificação de novos riscos.

#### Implantação do módulo

Para implantar o módulo de alocação de projetos, será utilizado o processo básico de desenvolvimento iterativo incremental de software [82], pois ele é executado pelas equipes de desenvolvimento da plataforma PUMA.

A implementação do Módulo de Alocação de Projetos se dará a partir do MVP, que é a sigla do Mínimo Produto Viável, em inglês, *Minimum Viable Product*, de modo que serão feitos pequenos incrementos a esse módulo até o limite de tempo de duração da pesquisa. A sequência de atividades que serão executadas para implementação do módulo é:

- (i) Levantamento de Requisitos, que consistirá na coleta das necessidades dos *stakeholders* a partir de entrevistas não estruturadas e da revisão da bibliográfica realizada para a pesquisa. Essas necessidades serão traduzidas em Histórias de Usuário com seus critérios de aceitação, além de serem elaborados protótipos de baixa e alta fidelidade;
- (ii) Análise de Design, que englobará a definição da arquitetura, a escolha das tecnologias e a definição do Modelo Entidade Relacionamento (MER);
- (iii) Construção, que envolverá o desenvolvido do módulo por meio da aplicação dos requisitos, da arquitetura, das tecnologias e do MER;
- (iv) Teste, onde será avaliado o módulo por meio de testes funcionais realizados por alunos, professores e agentes externos da Engenharia de Produção da UnB;
- (v) Produção, que incluirá a liberação do Módulo de Alocação de Projetos para utilização, de modo que ele será disponibilizado, via web, dentro da plataforma PUMA.

Após a implantação do módulo, será elaborado um plano detalhado acerca da sua estratégia de monitoramento e manutenção, incluindo as tarefas necessárias e como executá-las a fim de manter seu funcionamento a longo prazo.



Ademais, para fazer com que a implantação do módulo seja efetiva, alcançando os resultados esperados, será realizado o monitoramento e controle dos riscos.

### **Monitoramento e controle dos riscos identificados e de novos riscos**

O monitoramento e controle dos riscos tem o intuito de avaliar se as ações de tratamento traçadas foram implementadas. Para isso, será elaborado um quadro comparativo com os status e resultados das ações de tratamento de riscos.

Além disso, será realizada uma avaliação quanto à eficácia da eliminação, mitigação e controle dos riscos de alocação inadequada de projetos após a implantação das ações de tratamento. Essa avaliação acontecerá em reuniões de validação com os mesmos participantes das sessões de *brainstorming* de tratamento dos riscos, as quais foram realizadas na Etapa 3.

Como o tratamento de riscos também pode introduzir novos riscos que precisam ser gerenciados, será realizada uma nova identificação de riscos nas reuniões de validação da eficácia das ações de tratamento. Desta forma, ao final da pesquisa, os riscos e as ações de tratamento remanescentes serão documentados para serem submetidos a trabalhos futuros.

Além das etapas apresentadas até então, ocorrerão fases que são consideradas contínuas porque estarão presentes ao longo de toda a pesquisa, conforme é detalhado na Seção 3.2.5.

### **3.2.5 Fases contínuas**

Diferentemente da ISO 31000 [33], o CRISP-DM [40] não tem, na sua visão geral, as fases paralelas de Comunicação e consulta; Monitoramento e análise crítica; e Registro e relato. Entretanto, parte dessas fases ocorrem nas tarefas do processo de mineração de dados apresentado neste modelo de referência. Dito isso, a seguir é descrito como essas fases paralelas serão executadas ao longo da presente pesquisa.

#### **Comunicação e consulta**

Nesta fase, busca-se envolver as partes interessadas na pesquisa, com o intuito de promover a conscientização para coleta de informações que auxiliarão na tomada de decisão. Assim, pontos de vista diferentes serão considerados ao se construírem os resultados da pesquisa, por exemplo, as opiniões de professores, alunos, agentes externos e especialistas de gestão de riscos e de mineração de dados.

### **Monitoramento e análise crítica**

Esta fase contribuirá para que os resultados da pesquisa sejam avaliados pelas partes interessadas do processo de alocação de projetos. Essa avaliação acontecerá, principalmente, com o retorno que será dado nas reuniões de validação e nas respostas das enquetes.

### **Registro e relato**

Os processos de gestão de riscos e de mineração de dados serão documentados e relatados nesta dissertação de mestrado, com foco na continuidade futura da pesquisa; nas diferentes necessidades específicas de informação das partes interessadas; e na pertinência da informação para atestar o alcance dos objetivos traçados para ambos os processos. Além disso, esta fase permitirá que ambos os processos sejam revisados, pois incluirá todas as atividades executadas de forma organizada, além de destacar experiências importantes.

Finalizada a apresentação da metodologia da pesquisa, os resultados são detalhados no Capítulo 4.

# Capítulo 4

## Resultados e Discussões

Neste capítulo, são apresentados os resultados obtidos por meio das etapas definidas para o desenvolvimento da pesquisa. Essas etapas foram concebidas mediante o emprego paralelo de processos padronizados de gestão de riscos e mineração de dados, os quais são descritos, respectivamente, na ISO 31000 [33] e no CRISP-DM [40].

### 4.1 Etapa 1 - Entendimento do contexto e das questões de negócio

Ao iniciar a pesquisa, foi necessário entender o contexto e as questões de negócio que permeiam o problema de alocação inadequada de projetos, cujos resultados são apresentados nas seções 4.1.1 e 4.1.2.

#### 4.1.1 Estabelecimento do contexto

Com o estabelecimento do contexto, buscou-se alinhar os objetivos e o escopo da pesquisa com as necessidades do Curso Engenharia de Produção da UnB. Para isso, os resultados foram divididos em relação ao tipo de contexto, o qual pode ser externo ou interno.

##### Contexto Externo

Ao longo dos seus 58 anos de existência, a Universidade de Brasília (UnB) contribuiu diretamente para o desenvolvimento do Distrito Federal (DF) e do Brasil, formando profissionais e desenvolvendo pesquisas de qualidade, tornando-se, assim, uma instituição pública federal de excelência [83].

Essa excelência foi buscada pela UnB por meio de várias medidas, por exemplo, o Conselho Universitário (CONSUNI) aprovou, na 339ª reunião ocorrida no dia 04/07/2008, a “Proposta da Universidade de Brasília para o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI)” [84].

Dentre as ações da UnB presentes nessa proposta, cuja base foi o Decreto nº 6.096 de 24/04/2007 [85], estava incluída a criação de novos cursos e a expansão dos já existentes, sobretudo no período noturno, como forma de melhorar o aproveitamento do seu espaço físico. Assim, a Faculdade de Tecnologia (FT) da UnB propôs incluir, no Programa REUNI, a criação do curso noturno de Engenharia de Produção, com o intuito de melhorar o aproveitamento da sua infra-estrutura física, quase totalmente ociosa no período noturno [13].

A formação de profissionais de Engenharia de Produção contribui para a região que a Universidade de Brasília está inserida, pois no Distrito Federal há uma perspectiva excelente de atuação de engenheiros com um perfil profissional capaz de atuar criativamente na identificação e resolução de problemas, em especial, no setor de serviços [13, 14].

Segundo o Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB [13], as razões básicas que levaram a FT a propor a criação do curso noturno de Engenharia de Produção na UnB foram:

- (i) a importância das atividades de Engenharia de Produção para a região em que a universidade está inserida;
- (ii) a quase completa ociosidade das facilidades da FT no período noturno;
- (iii) e o pequeno número de cursos de graduação em Engenharia de Produção e a inexistência de cursos de pós-graduação na área avaliados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) no Centro-Oeste.

Assim sendo, todos os pontos elencados, como justificativas para a criação do Curso, influenciaram os seus objetivos e a sua rede de relacionamentos, compromissos e dependências externas.

Dentre os diversos relacionamentos do Curso, destacam-se aqueles com os agentes externos (*stakeholders*), pois eles fornecem insumos para as disciplinas de Projeto de Sistemas de Produção (PSP) e Projetos de Graduação (PG). Esses insumos são os problemas reais das empresas que podem ser privadas ou públicas, grandes ou pequenas, de manufatura ou de serviços [8].

Esses relacionamentos externos com as empresas nasceram da concepção metodológica do currículo do Curso, que teve como base a metodologia *Project Based Learning* (PBL) [13]. Essa concepção deu destaque à atividade de projeto ao longo de suas disciplinas.

As oito disciplinas de projetos, denominadas PSPs, são consideradas a “espinha dorsal” do Curso, estando presentes do quarto ao décimo semestre, de modo que buscam desenvolver no aluno competências transversais, tais como liderança, gerenciamento e proatividade, além das competências técnicas, por exemplo, utilização de ferramental matemático e estatístico para modelar sistemas de produção [5].

Os alunos também solucionam problemas no Projeto de Graduação, que deve favorecer a iniciação científica e a aquisição de experiência em pesquisa bibliográfica, na confecção de relatórios técnicos e na exposição oral, bem como demonstrar a capacidade de trabalho, tanto em equipe como individual [13]. Ainda segundo o Projeto Político Pedagógico do Curso [13], no Projeto de Graduação, espera-se que o aluno concilie o seu tema de projeto com as atividades exercidas durante o Estágio Supervisionado.

Desta forma, a estratégia pedagógica do Curso de Engenharia de Produção da UnB procura garantir uma visão articulada entre as características de atuação profissional e as diferentes áreas do conhecimento, permitindo compreender a diversidade de aspectos determinantes envolvidos na solução de problemas [5].

Mostra-se, assim, a complexidade da rede de relacionamentos e as dependências externas do Curso quanto a aplicação da metodologia PBL. A seguir, será abordado o Contexto Interno da organização em questão, destacando-se os fatores que permeiam a execução dos projetos pelos estudantes.

### **Contexto Interno**

O Departamento EPR é o órgão da Faculdade de Tecnologia (FT) organizado com a finalidade de realizar atividades de ensino, pesquisa e extensão no âmbito do Curso de Engenharia de Produção oferecido pela UnB [16].

Dentre os professores que compõem o Departamento EPR em 2021, há cerca de 16 que pertencem ao quadro efetivo da UnB; e 9 voluntários que foram admitidos na forma da lei e das normas vigentes da universidade.

Além dos professores, o Departamento EPR é composto por 5 servidores, designados para atividades de apoio técnico-administrativo à área de Engenharia de Produção.

Em relação ao quantitativo de alunos do Curso, estima-se que haja cerca de 600 distribuídos nas disciplinas obrigatórias, optativas e de módulo livre oferecidas em sua grade curricular.

Segundo o Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB [13], as bases legais que moldaram a sua formação foram:

- (i) A Lei No 9.394 de 20/12/1996, que estabelece as Diretrizes e Bases da Educação Nacional, de modo que foi utilizada para a elaboração do Projeto Político Pedagógico do Curso;

- (ii) A Resolução CNE/CES 11/2002, que institui as Diretrizes Curriculares Nacionais dos Cursos de Graduação em Engenharia, sendo a mesma utilizada na organização curricular do Curso;
- (iii) A Resolução CNE/CES 02/2007, que serviu de base para o estabelecimento das cargas horárias mínimas para o Curso.

A Engenharia de Produção da UnB também segue às diretrizes gerais produzidas pela Associação Brasileira de Engenharia de Produção (ABEPRO) para os cursos desta área. Além disso, a grade curricular do Curso, por meio da aplicação da metodologia (PBL), promove a geração de soluções para os problemas reais de agentes externos e, às vezes, ideias inovadoras, proporcionam, assim, uma experiência completa de gerenciamento de projetos [5].

O bom andamento da aplicação da metodologia PBL não depende apenas dos problemas oriundos, principalmente, dos agentes externos, mas também da interligação de disciplinas de projetos; de conteúdo técnico; e de outras que contemplam áreas do conhecimento específicas que são importantes para o desenvolvimento das soluções [8].

Neste sentido, durante a execução dos projetos no Curso, podem ser gerados muitos dados e ocorrer muitos riscos que, atualmente, não são gerenciados. Essa necessidade de gestão de riscos é espelho do que acontece na UnB, onde este processo foi formalizado e implantado recentemente [83].

É pretendido que a plataforma PUMA seja utilizada para centralizar esses dados, por exemplo, as propostas e as versões finais dos 90 projetos que, em média, são desenvolvidos a cada semestre nas disciplinas de PSP e nos projetos de graduação. Deste modo, é demandado que o Curso de Engenharia de Produção da UnB tenha uma forma rápida e assertiva de alocação de projetos, com o intuito de diminuir os riscos atrelados a este processo.

A partir do entendimento dos contextos externo e interno do Curso, foi possível projetar parte da aplicação dos processos de gestão de riscos e de mineração de dados na pesquisa. Entretanto, antes de aplicá-los, buscou-se também uma compreensão do negócio, conforme os resultados apresentados na Seção 4.1.2.

### **4.1.2 Compreensão do negócio**

Para que a aplicação dos processos de gestão de riscos e de mineração de dados fosse efetiva, foi necessário coletar as expectativas das partes interessadas a partir de uma perspectiva de negócio. Deste modo, realizou-se a avaliação da situação atual do ambiente e a definição dos escopos e critérios desses processos. Além disso, foi feita a seleção inicial de técnicas e ferramentas e a construção do cronograma para a pesquisa.

## Avaliação da situação atual para desenvolvimento da pesquisa

A primeira tarefa da avaliação da situação atual foi o levantamento dos recursos disponíveis para a pesquisa, cujos resultados foram consolidados na Tabela 4.1.

Tabela 4.1: Lista de recursos disponíveis para a pesquisa.

Tipo	Recursos disponíveis
Pessoas	1 mestrando em computação aplicada (responsável pela pesquisa)
	1 especialista em gestão de riscos (orientadora da pesquisa)
	1 especialista em mineração de dados
	1 equipe de gestão do desenvolvimento da plataforma PUMA com 3 alunos
	1 equipe de <i>benchmarking</i> para busca na literatura científica com 3 alunos
	1 equipe de levantamento de requisitos com as partes interessadas com 4 alunos
	1 equipe de validação da qualidade dos requisitos com 4 alunos
	1 equipe de medição da qualidade e teste de funcionalidades com 4 alunos
	Suporte técnico da instituição parceira que fornece tecnologias de hardware e software para a plataforma PUMA
	2 equipes de desenvolvimento do Curso de Engenharia de Software da UnB
Dados	Os projetos de graduação armazenados na Biblioteca Digital da Produção Intelectual Discente da Universidade de Brasília (BDM), os quais também são controlados por uma planilha alimentada pelo Departamento EPR
	Os dados curriculares de professores contidos na Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
	Os projetos realizados nas disciplinas de PSP que estão sendo inseridos em um repositório na plataforma PUMA
	As ementas das disciplinas de PSP que podem ser acessadas no Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA)
Hardware e Software	Plataforma de código aberto Orange 3 para entendimento dos dados
	Serviço gratuito Google Colab para modelagem
	Máquina Virtual com as seguintes configurações: - vCPU: 2 - Memória RAM: 2 GB - Espaço em Disco: 100 GB - Sistema Operacional: Ubuntu Server 18.04

Nota-se que os dados curriculares dos alunos na Plataforma Lattes não foram considerados, pois nem todos os alunos utilizam ou têm o costume de manter as suas informações atualizadas nessa plataforma de currículos.

Os recursos disponíveis para a pesquisa foram separados por tipo (pessoas; dados; e hardware e software), visando um maior nível de detalhamento. Em relação aos custos financeiros, a aplicação da pesquisa não terá, visto que é uma pesquisa desenvolvida no âmbito do Mestrado Profissional em Computação Aplicada (MPCA) da UnB e que tem recursos gratuitos disponíveis para academia ou que são oriundos de parcerias.

Além dos recursos, foram levantados requisitos, premissas e restrições da pesquisa, conforme apresentado na Tabela 4.2.

Tabela 4.2: Requisitos, premissas e restrições da pesquisa.

Tipo	Itens
Requisitos	Os processos de gestão de riscos e de mineração de dados precisam ser aplicados no limite de tempo de duração da pesquisa
	Os dados utilizados em ambos os processos precisam ser públicos ou, quando privados, descaracterizados, respeitando as questões legais de uso
Premissas	Que as bases externas públicas terão mecanismos de extração de dados, por exemplo, APIs e protocolos
	Que os dados dos projetos sejam acessíveis, mesmo que não estruturados em apenas um lugar de armazenamento
	Que as partes interessadas estejam disponíveis para eventuais reuniões, aplicações de enquete, entrevistas e sessões de <i>brainstorming</i>
	Que os recursos de hardware e software estejam disponíveis para uso durante a execução da pesquisa e, caso seja necessária a obtenção de novas tecnologias, será possível solicitá-las à instituição parceira que fornece essas tecnologias para a plataforma PUMA
Restrições	Os recursos de hardware e software utilizados na pesquisa serão de código aberto ou limitados às tecnologias disponibilizadas para a plataforma PUMA

Durante o levantamento dos requisitos, premissas e restrições apresentados na Tabela 4.2, certificou-se acerca da permissão de acesso e utilização dos dados para o desenvolvimento da pesquisa, bem como a disponibilidades das partes interessadas para eventuais interações durante este período. Os itens levantados por meio da avaliação da situação atual serviram como insumos para a definição do escopo da pesquisa.

### Estabelecimento do escopo da pesquisa

O escopo da pesquisa foi estabelecido a partir da avaliação da situação atual e dos aspectos norteadores dos contextos externo e interno em que o Curso de Engenharia de Produção opera, em especial, os relacionados à alocação de projetos para as disciplinas de PSP e Projetos de Graduação. Na Tabela 4.3, é mostrado o esquema geral do escopo da pesquisa.

Tabela 4.3: Escopo da pesquisa.

Item	Descrição
Objetivo de negócio	Propor uma ferramenta computacional para tratar os riscos de alocação inadequada de projetos em disciplinas que utilizam a aprendizagem ativa, no âmbito do Curso de Engenharia de Produção da UnB
Objetivo geral do processo de gestão de riscos	Tratar os riscos relacionados à alocação de projetos do Curso
Objetivo geral do processo de mineração de dados	Recomendar disciplinas mais adequadas para os projetos
Decisões que precisam ser tomadas	Implementação ou não do modelo proposto como um módulo da plataforma PUMA para a alocação de projetos
Resultados esperados	Tratamento dos riscos por meio de um modelo de mineração de dados, o qual será implementado como um módulo da plataforma PUMA para a alocação de projetos



Além dos elementos apresentados na Tabela 4.3, o tipo de problema de mineração de dados foi especificado, sendo ele considerado como de classificação. Neste sentido, as classes serão definidas antecipadamente, de modo a se ter um aprendizado supervisionado.

Após o estabelecimento do escopo da pesquisa, foi necessária a definição de critérios que pudessem nortear a execução das suas etapas.

### Estabelecimento dos critérios para execução da etapas

De forma a garantir resultados úteis e bem-sucedidos da pesquisa, do ponto de vista técnico e de negócio, foram definidos os critérios compilados na Tabela 4.4.

Tabela 4.4: Critérios para a pesquisa.

Processo	Critérios	Julgamento
Gestão de Riscos	Ao avaliar os riscos de alocação inadequada de projetos, é necessário considerar aqueles que podem surgir caso seja uma alocação tanto manual quanto automática.	Participantes das sessões de brainstorming
	Devem ser avaliados os riscos de alocação inadequada de projetos para as disciplinas de PSP e projetos de graduação.	
	As causas dos riscos precisam ser destrinchadas até aquelas que são raízes.	
	Caso as técnicas escolhidas para análise dos riscos apresentem apenas resultados qualitativos, será necessário, ao final, aplicar um processo de quantificação.	
	Devem ser tratados com prioritários os riscos que tiverem causas raízes com maior importância em causar o problema atual e em contribuir para a ocorrência de riscos futuros de alocação inadequada de projetos. Além disso, os riscos que apresentarem causas raízes comuns devem receber uma atenção maior, visando otimizar o tratamento deles.	
	As ações de tratamento dos riscos devem ser priorizadas de acordo com quão bem elas podem contribuir para a eliminação, mitigação ou controle das causas raízes, bem como a facilidade e o prazo necessário para implementá-las.	
Mineração de Dados	Deve ser avaliado o desempenho do modelo por meio da aplicação de diferentes técnicas e parâmetros.	Comparação dos resultados das diferentes técnicas e parâmetros
	O modelo proposto deve apresentar um nível de precisão preditiva igual ou maior do que os níveis encontrados nos trabalhos relacionados.	Comparação dos resultados do modelo com os dos trabalhos relacionados
	O modelo proposto deve ter velocidade de resposta e desempenho adequado para viabilizar o seu uso	Especialista em mineração de dados
	O modelo proposto deve contribuir para tratamento dos riscos associados à alocação inadequada de projetos, no âmbito do Curso de Engenharia de Produção da UnB.	Especialista em gestão de riscos
	O modelo proposto deve fazer a alocação de projetos de modo relevante, transparente, facilitado e coerente.	Professores, alunos e agentes externos

Nota-se que alguns dos critérios definidos para a pesquisa serão medidos de modo objetivo, por exemplo, os relacionados ao nível de precisão preditiva. Entretanto, outros

têm características mais subjetivas, caso em que especialistas e partes interessadas farão o julgamento.

Embora os critérios tenham sido estabelecidos no início da pesquisa, eles são dinâmicos, de modo que serão analisados continuamente e, se necessário, alterados ao longo das etapas.

Os critérios auxiliam na elaboração do plano pretendido para atingir os objetivos da pesquisa, o qual teve como tarefa inicial a seleção de técnicas e ferramentas.

### Seleção inicial de técnicas e ferramentas

O primeiro foco do planejamento inicial da pesquisa foi a avaliação preliminar de técnicas e ferramentas para os processos de gestão de riscos e de mineração de dados.

Para a seleção de técnicas apropriadas para o processo de gestão de riscos, buscou-se aquelas que pudessem facilitar a identificação de diferentes categorias de risco por meio de cenários e perspectivas, além de, ao final da avaliação dos riscos, fosse possível fornecer insumos para o tratamento dos riscos identificados.

As primeiras técnicas foram retiradas dos artigos que constam no Referencial Teórico (Capítulo 2) e que tratam da gestão de riscos de projetos desenvolvidos por estudantes, sendo eles escritos por Vanhanen e Lehtinen [43] e Straub [42].

Outro parâmetro utilizado na seleção das técnicas foi um dos critérios estabelecidos para o processo de gestão de riscos, o qual estabeleceu que, caso as técnicas escolhidas para análise dos riscos apresentem apenas resultados qualitativos, será necessário, ao final, aplicar um processo de quantificação. Para atender a esse critério, foi selecionado como base o trabalho de Pham et al. [44].

Os trabalhos utilizados para a seleção das técnicas têm em comum o fato de, ao aplicar o processo de gestão de riscos, terem buscado gerar ações para tratá-los. Desta forma, na Tabela 4.5, foram compiladas as técnicas selecionadas para a identificação, análise e avaliação dos riscos de alocação inadequada de projetos.

Tabela 4.5: Técnicas selecionadas para o processo de gestão de riscos.

Etapas do processo	Técnicas selecionadas	Artigos
Identificação de riscos	Brainstorming	[42]
Análise de riscos	Análise de Causa Raiz	[43, 42, 44]
Avaliação de riscos		

As técnicas apresentadas na Tabela 4.5 são também detalhadas na ISO 31010 [31], que serviu de base teórica para a implementação delas. Ressalta-se que, mesmo que outras técnicas pudessem ser utilizadas, estas foram as consideradas mais adequadas para a pesquisa e que tiveram o melhor detalhamento do passo a passo de aplicação nos trabalhos encontrados na literatura.

Já em relação a seleção das técnicas e ferramentas para o processo de mineração de dados, também avaliou-se os trabalhos relacionados que constam no Referencial Teórico (Capítulo 2). Essa avaliação teve como ferramenta a Tabela 4.6, em que são comparadas as informações inferidas da leitura dos trabalhos relacionados com os dados, o objetivo e o problema de mineração de dados da pesquisa.

Tabela 4.6: Avaliação de artigos para seleção de técnicas de mineração de dados.

<b>Artigos:</b>	[26]	[3]	[57]	[58]	[47]	[10]	[45]	[28]	[6]	[29]	[9]	[59]	[60]
<b>Dados utilizados</b>													
Projetos de Graduação								x		x	x		
Currículos / áreas dos professores						x	x	x			x		x
Projetos desenvolvidos em disciplinas	x												
Ementas / áreas das disciplinas						x							
<b>Objetivo do processo de mineração de dados</b>													
Recomendar disciplinas e/ou pessoas (orientadores e avaliadores) para os projetos								x		x	x		x
<b>Problema de mineração de dados</b>													
Classificação de projetos											x		

Para que a avaliação apresentada Tabela 4.6 fosse possível, foi necessário generalizar a descrição dos itens de comparação (os dados, o objetivo e o problema de mineração de dados), visando englobar o escopo dos trabalhos relacionados. Um exemplo disso é a adaptação do objetivo do processo de mineração de dados, de modo que, ao recomendar pessoas para os projetos, independente de serem orientadores, o trabalho relacionado pudesse ser considerado como de forte relação com a presente pesquisa.

É possível justificar a existência de alguns artigos que, apesar de terem sido considerados trabalhos relacionados, não têm nenhum item fortemente comum a esta pesquisa. As justificativas para isso são: eles não apresentam resultados práticos de implementação real de uma proposta [3]; focaram nas atividades executadas nos projetos e não diretamente nos seus conteúdos [6]; ou utilizaram como bases de dados principais alguns repositórios de projetos de software ou de objetos de aprendizagem [47, 57, 58, 59].

Há também alguns trabalhos relacionados que, embora tenham utilizados dados comuns aos disponíveis para a pesquisa, não possuem relação direta com o objetivo ou problema do processo de mineração de dados a ser aplicado [10, 26, 45].

Esses trabalhos que apresentaram nenhum ou poucos itens sinalizados, na comparação realizada por meio da Tabela 4.6, fornecem bases teóricas e ideias de funcionalidades para o desenvolvimento da proposta de ferramenta computacional a ser elaborada neste estudo.

Dentro os artigos apresentados na Tabela 4.6, o artigo de Núñez e Ramos [9] foi o que apresentou mais itens assinalados como de forte relação com a presente pesquisa. Isso porque, embora alguns dos outros trabalhos trazem conceitos acerca do problema de classificação, apenas este tem explicitamente em seu escopo a classificação de projetos para recomendação de avaliadores de projetos finais de graduação. Entretanto, os autores Núñez e Ramos [9] utilizaram apenas projetos finais da área de Ciência da Computação distribuídos em quatro campos do conhecimento.

Ademais, foi constatado que os trabalhos de Afolabi et al. [28]; Goh, Gottipati e Shankararaman [29]; Núñez e Ramos [9]; e Bastos [60] possuem forte relação com o objetivo e o problema de mineração de dados desta pesquisa, além de terem utilizados dados similares aos disponíveis para a mesma. Assim, desses quatro trabalhos, foram selecionadas as técnicas que são apresentadas na Tabela 4.7.

Tabela 4.7: Técnicas selecionadas para o processo de mineração de dados.

Preparação dos dados	Definição do espaço de termos	<i>Tokenization</i>	[29]
		Remoção de <i>stopwords</i>	[9, 29, 60]
		Remoção / substituição de sinais de pontuação, caracteres especiais e acentos	[9, 60]
		Substituição de “ç” por “c”	[60]
		Remoção de excessos em branco	
		Conversão para letras maiúsculas	
		<i>Stemming</i>	[9, 29, 60]
Pesos dos termos	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	[60]	
Representação	<i>Vector Space Model (VSM)</i>	[9, 29, 60]	
Modelagem	Medida	Distância Euclidiana	[9]
	Métodos	<i>K-Nearest Neighbors (K-NN)</i>	
			Classificador Naïve Bayes
Avaliação	Estratégia / Métricas	<i>10-fold cross validation</i>	[9]
		Taxas de sucesso e de erro	[9, 60]
		<i>Precision</i>	
		<i>Recall</i>	
	<i>F-measure</i>	[60]	

As técnicas apresentadas na Tabela 4.7 foram frequentemente encontradas nos trabalhos relacionados, sendo descartadas aquelas aplicadas por meio de softwares ou kits de ferramentas externos que não pudessem ser integrados a plataforma PUMA.

Além disso, como é esperado o uso de resultados reais de classes para treinamento do modelo de alocação de projetos, foram selecionadas técnicas adequadas para aprendizado supervisionado.

Caso os documentos dos projetos já desenvolvidos e as propostas de novos projetos possuam comprimentos diferentes, será necessário empregar a normalização das coordenadas para obter vetores de comprimento igual a unidade [60, 79].

A eliminação de termos com base na frequência nos dados textuais poderá ser empregada, com intuito de reduzir a dimensionalidade dos dados para acelerar o tempo de resposta do modelo de alocação de projetos [60].

Em relação ao balanceamento dos dados, os trabalhos relacionados não apresentaram procedimentos, mas, caso seja necessário, serão utilizadas as técnicas de *undersampling*; *oversampling*; e *oversampling* seguida de *undersampling* [80].

Ademais, para que as técnicas apresentadas na Tabela 4.7 sejam aplicadas, será utilizada a ferramenta Google Colab [81], que é o ambiente Jupyter Notebook em Python a ser utilizado para modelagem. Essa ferramenta foi escolhida porque suporta as técnicas selecionadas e permite que sejam executados os diferentes estágios do processo de mineração de dados.

As tentativas de construção do modelo de alocação de projetos, a partir da aplicação das técnicas selecionadas, também serão executadas na máquina virtual Ubuntu Server 18.04 da plataforma PUMA. Isto porque será necessário medir a velocidade de resposta desse modelo no ambiente de produção, de modo a viabilizar a proposta da pesquisa.

Por conseguinte, o trabalho de Witten, Frank e Hall [51] será utilizado como uma das referências teóricas, pois possui instruções práticas para aplicação das técnicas do processo de mineração de dados.

Tendo em vista que o processo de mineração de dados pode sofrer o impacto de riscos durante a sua execução, é necessário traçar ações de tratamento caso eles ocorram. Deste modo, os riscos inerentes a esse processo que estão relacionados à alocação inadequada de projetos foram avaliados na Etapa 2 (Seção 4.2).

## 4.2 Etapa 2 - Processo de avaliação dos riscos e dados

Frente aos eventuais riscos que podem impactar a pesquisa e causar o problema de alocação inadequada de projetos, foi necessário avaliá-los, bem como entender os dados que podem ser utilizados para a elaboração da ferramenta computacional objeto deste estudo. Os resultados dessas atividades são apresentados nas próximas seções.

### 4.2.1 Identificação dos riscos

Durante a identificação dos riscos, buscou-se descrever aqueles que pudessem levar a uma alocação inadequada de projetos, no contexto do Curso de Engenharia de Produção da UnB.

As informações referentes aos riscos foram levantadas em duas sessões de *brainstorming*, nas quais tiveram como participantes uma professora adjunta que ministra duas disciplinas de PSP e é especialista em gestão de riscos; um ex-aluno do Curso; um aluno

coordenador do desenvolvimento da plataforma PUMA; e um cliente externo que trabalha em uma instituição financeira e também é professora voluntária do Departamento EPR.

Observa-se que os participantes dessas sessões possuem conhecimentos práticos do Curso de Engenharia de Produção, da Plataforma PUMA e do processo de alocação de projetos, possibilitando que fossem considerados diferentes tipos de risco.

As sessões ocorreram de forma online nos dias 10 e 24/03 de 2021, por meio da plataforma Microsoft Team, e cada uma delas teve duração de 1 hora e 30 minutos, de maneira que seus participantes foram estimulados a identificar os riscos. A ferramenta IdeaBoardz foi utilizada para fomentar a imaginação dos participantes, possibilitando que post-its virtuais pudessem ser utilizados durante a identificação dos riscos.

Ressalta-se que, antes do início das sessões, alguns riscos previamente identificados na revisão bibliográfica foram enviados para os participantes, visando fomentar uma preparação prévia que auxiliasse na validação da ocorrência ou não dos riscos identificados da literatura no Curso.

No início das sessões, o facilitador instruiu os participantes quanto ao uso da ferramenta IdeaBoardz e aos objetivos da sessão, que, no presente caso, era a identificação de riscos relacionados à alocação inadequada de projetos.

Durante os 15 minutos iniciais da primeira sessão, os participantes colocaram post-its com os possíveis riscos no quadro virtual, sem que, neste momento, houvesse discussões se os itens eram ou não realmente riscos, visando a não inibição do fluxo do pensamento deles. Após esse primeiro momento, todas as ideias iniciais foram revistas até a eliminação da duplicação de itens e a obtenção dos riscos aceitos por todos como reais.

Ressalta-se que foram identificados riscos que pudessem afetar os resultados caso seja uma alocação tanto manual quanto automática, sendo essa última por meio da ferramenta computacional a ser desenvolvida na pesquisa.

Neste sentido, os primeiros riscos identificados têm relação com o módulo de alocação de projetos, visando identificar aqueles que podem impactar a sua utilização. Esses riscos são apresentados na Figura 4.1.

Os usuários não fornecerem as informações de forma completa, correta e atualizada	Desuso do Módulo de Alocação de Projetos por parte dos usuários
Resultados equivocados ou que não agradam os usuários	Ter usuários dispostos a manipular os resultados para benefício próprio

Figura 4.1: Riscos de utilização inadequada do Módulo de Alocação de Projetos.

Observa-se que os quatro riscos levantados podem influenciar a forma de utilização e de recebimento dos resultados do Módulo de Alocação de Projetos, o qual terá como usuários os professores, alunos e agentes externos do Curso de Engenharia de Produção da UnB. O desuso desse módulo também foi considerado no levantamento dos riscos.

Outra categoria de risco identificada foi a relacionada à pesquisa, mais especificamente, no que tange a execução inadequada do processo de mineração de dados a ser utilizado para alocação de projetos, conforme a Figura 4.2.

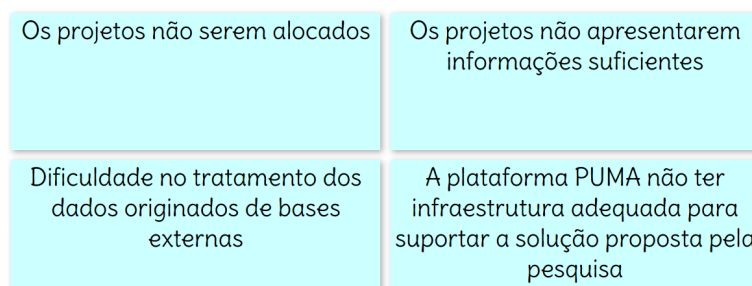


Figura 4.2: Riscos de execução inadequada do processo de mineração de dados.

Embora diversos outros riscos inerentes a qualquer aplicação de um processo de mineração de dados pudessem ser identificados, os apresentados na Figura 4.2 foram considerados os mais importantes no contexto da pesquisa, com base na opinião dos participantes dos *brainstormings*.

Para o processo de mineração de dados que será aplicado, foram identificados riscos relacionados à infraestrutura e às bases externas disponíveis para a pesquisa, além do fato desse processo poder não resultar na alocação de projetos ou, mesmo que eles sejam alocados, as informações contidas neles não serem suficientes para que esse processo seja adequado.

Os riscos apresentados até então são independentes dos tipos de projeto, mas buscou-se também a identificação daqueles específicos da alocação de projetos para as disciplinas de PSP e projetos de graduação (PG).

Sendo assim, os riscos identificados de alocação inadequada de projetos de graduação (PGs) são apresentados na Figura 4.3.

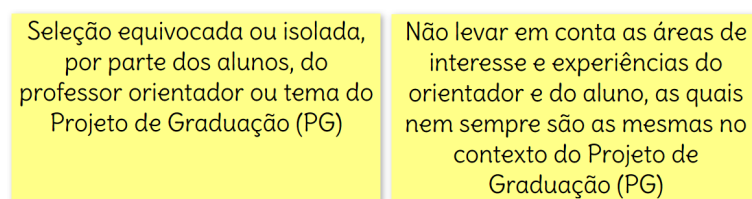
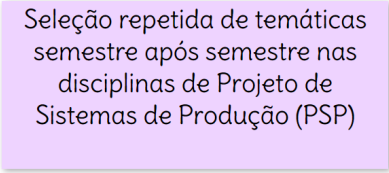


Figura 4.3: Riscos de alocação inadequada de projetos de graduação.

Para alocação de projetos de graduação (PGs), foram identificados dois riscos relacionados à tomada de decisão quanto aos professores orientadores ou temas dos projetos e às preferências e experiências que podem ser diferentes entre os alunos e professores.

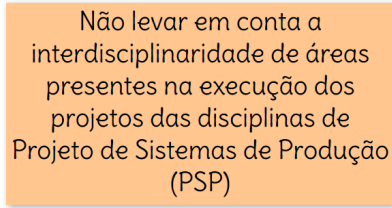
Os riscos identificados de alocação inadequada de projetos para as disciplinas de PSP são expostos na Figura 4.4.



Seleção repetida de temáticas semestre após semestre nas disciplinas de Projeto de Sistemas de Produção (PSP)

Figura 4.4: Riscos de alocação inadequada de projetos para as disciplinas de PSP.

Conforme pode ser observado, a não diversificação de temas nos projetos desenvolvidos foi elencada como um risco referente à alocação de projetos para disciplinas de PSP. Entretanto, os participantes das sessões de *brainstorming* indicaram outro risco que tem relação com a alocação para disciplinas de PSP, mas de projetos do tipo interdisciplinar, conforme a Figura 4.5.



Não levar em conta a interdisciplinaridade de áreas presentes na execução dos projetos das disciplinas de Projeto de Sistemas de Produção (PSP)

Figura 4.5: Riscos de alocação inadequada de projetos interdisciplinares dos PSPs.

Os projetos interdisciplinares estão cada vez mais presentes nas disciplinas de PSP, principalmente por causa do fomento a projetos multidisciplinares e internacionais no Curso. Deste modo, o risco apresentado na Figura 4.5 considera que a solução de um problema pode necessitar de conhecimentos de diferentes áreas, que estão presentes em distintas disciplinas.

Um exemplo de projeto interdisciplinar é a pesquisa em questão, pois envolve alunos e professores de cursos da UnB, de modo específico da Engenharia de Produção e do Mestrado Profissional em Computação Aplicada (MPCA) do Campus Darcy Ribeiro e da Engenharia de Software da Faculdade UnB Gama (FGA); bem como do Curso de Ciência da Computação da Universidade ESPRIT, localizada na Tunísia.

É importante salientar que, embora um Projeto de Graduação (PG) possa demandar conhecimentos interdisciplinares para o seu desenvolvimento, não foram considerados riscos relacionados a este cenário. Isso porque é comum ser demandado ao aluno de final



de curso a aplicação individual de conhecimentos de diversas áreas. Nestes casos, o foco poderá estar na área principal da proposta de cada PG.

Em síntese, como resultado do *brainstorming* realizado para a identificação de riscos, obteve-se a divisão deles em quatro categorias, a saber: de utilização inadequada do Módulo de Alocação de Projetos; de execução inadequada do processo de mineração de dados; de alocação inadequada de projetos de graduação; e de alocação inadequada de projetos para as disciplinas de PSP. Na categoria de riscos das disciplinas de PSP, foi identificada a subcategoria de riscos de alocação inadequada de projetos interdisciplinares.

De posse dos riscos, foi realizada a análise de suas causas, cujos resultados são apresentados na Seção 4.2.4.

## 4.2.2 Análise dos riscos

A análise dos riscos envolveu a consideração detalhada das causas dos riscos que podem ocasionar o problema de alocação inadequada de projetos, afetando, assim, os objetivos de negócio do Curso de Engenharia de Produção da UnB.

Ela foi realizada por meio da ferramenta de Análise de Causa Raiz, a qual foi aplicada em duas sessões de *brainstorming* nos dias 31/03 e 07/04 de 2021, cuja duração de cada uma delas foi de 1 hora e 30 minutos, tendo como entrada os riscos apresentados na Seção 4.2.1.

Os participantes dos *brainstormings* para análise dos riscos foram os mesmos da identificação dos riscos, de modo que foram instigados a identificar as causas dos riscos até as suas raízes.

Como forma de mapeamento das causas raízes dos riscos, foram realizados desenhos de ramos em forma de árvore por meio da utilização da técnica dos “5 porquês”, onde repetidamente perguntou-se o “por quê?” de cada risco para criar camadas de causa e sub-causa até aquelas que fossem as raízes.

O primeiro ramo foi formado por meio da Análise de Causa Raiz dos riscos de utilização inadequada do Módulo de Alocação de Projetos, conforme é apresentado na Figura 4.6.

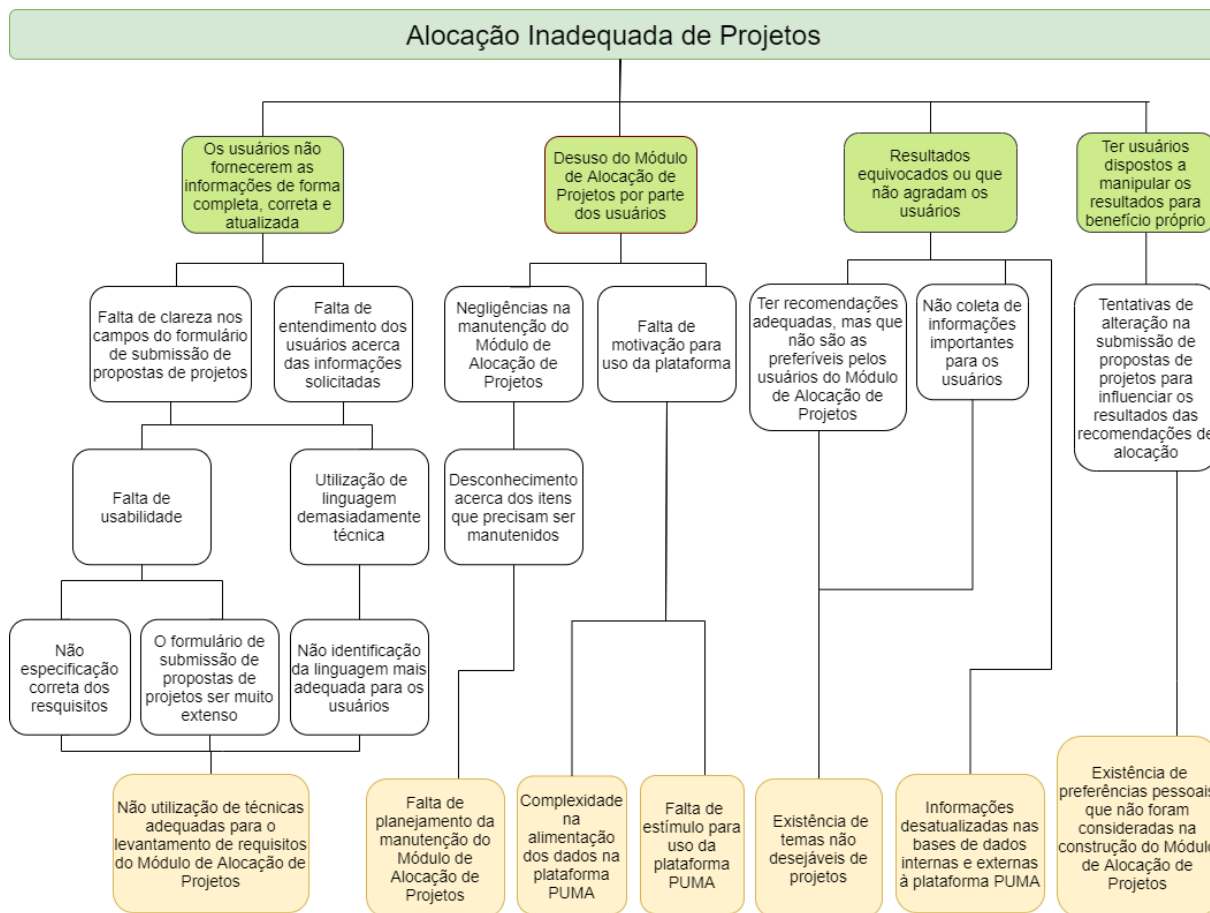


Figura 4.6: Análise de Causa Raiz dos riscos de utilização inadequada do Módulo de Alocação de Projetos.

Para o risco de os usuários não fornecerem as informações de forma completa, correta e atualizada, foi identificado que a não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projeto pode ser uma causa raiz.

O risco de desuso do Módulo de Alocação de Projetos por parte dos usuários pode ter como causas raízes a falta de planejamento da sua manutenção; a sua complexidade de alimentação dos dados; e a falta de estímulo para utilização da plataforma PUMA.

Foi analisado também o risco de o Módulo de Alocação de Projetos ter resultados equivocados ou que não agradam os usuários, podendo ter como causas raízes a existência de temas não desejáveis de projetos e de informações desatualizadas nas bases de dados internas e externas à plataforma PUMA.

O fato de poder ter usuários dispostos a manipular os resultados para benefício próprio também foi considerado um risco, tendo como possível causa raiz a existência de preferências pessoais que não foram consideradas na construção do Módulo de Alocação de Projetos.

A Análise de Causa Raiz dos riscos de execução inadequada do processo de mineração de dados resultou no segundo ramo, o qual é apresentado na Figura 4.7.

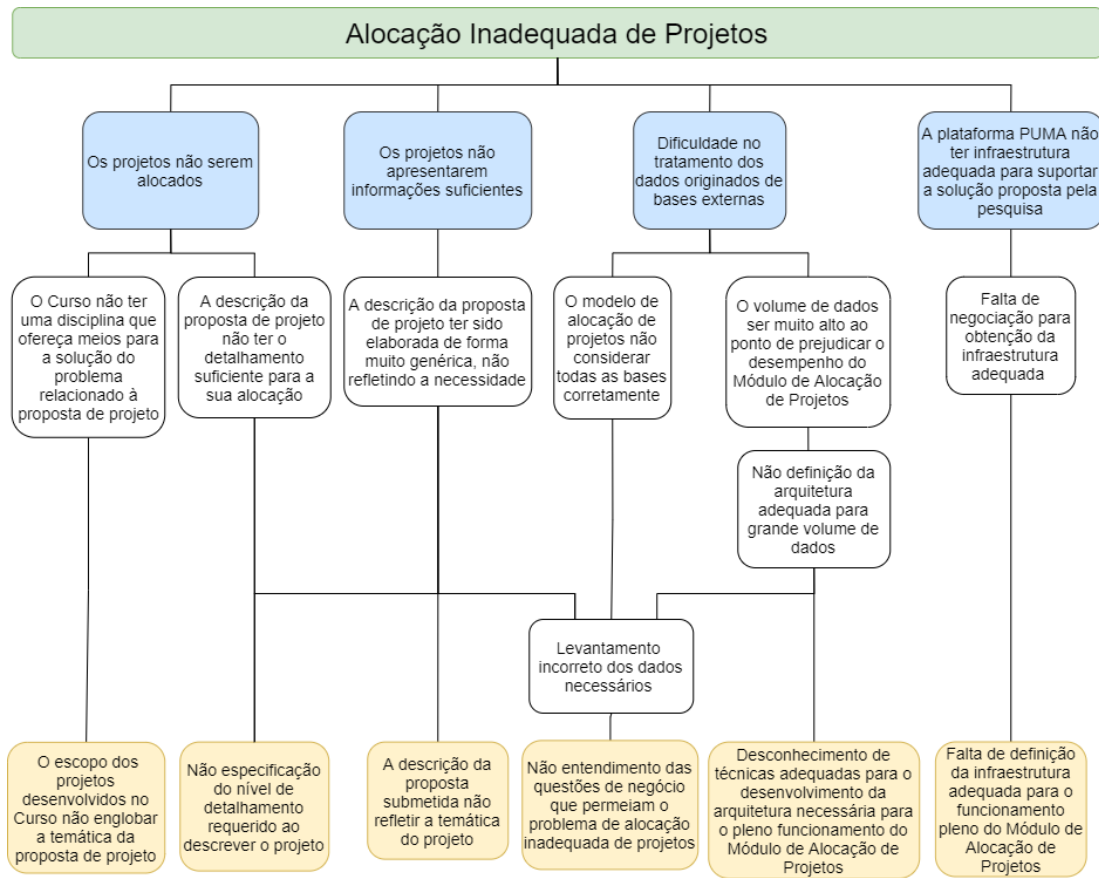


Figura 4.7: Análise de Causa Raiz dos riscos de execução inadequada do processo de mineração de dados.

O primeiro risco analisado foi de os projetos não serem alocados, podendo ter como causas raízes: o escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto; a não especificação do nível de detalhamento requerido ao descrever o projeto; a descrição da proposta submetida não refletir a temática do projeto; e o não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos.

Nota-se que essas últimas três causas raízes, relativas ao primeiro risco, também foram consideradas raízes do risco de os projetos não apresentarem informações suficientes para que aplicação do processo de mineração de dados leve a uma alocação adequada.

A primeira causa raiz do terceiro risco, denominado como dificuldade no tratamento dos dados originados de bases externas, foi o não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos, tendo, assim, semelhança com os dois primeiros riscos. Já a segunda causa raiz para esse terceiro risco foi o desco-

nhecimento de técnicas adequadas para o desenvolvimento da arquitetura necessária para o pleno funcionamento do Módulo de Alocação de Projetos.

A falta de definição de uma infraestrutura adequada, que é demandada para funcionamento pleno do Módulo de Alocação de Projetos, foi considerada a causa raiz do risco de a plataforma PUMA não ter infraestrutura adequada para suportar a solução proposta pela pesquisa, sendo este o último dos quatro riscos identificados que tem relação com a execução inadequada do processo de mineração de dados.

Estes dois primeiros ramos apresentados da árvore de causa raiz dos riscos foram construídos sob uma perspectiva que é independente do tipo de projeto. Entretanto, também foram analisadas as causas raízes dos riscos de alocação inadequada de projetos para as disciplinas de PSP e projetos de graduação.

Neste sentido, o ramo elaborado por meio da Análise de Causa Raiz dos riscos de alocação inadequada de projetos de graduação é apresentado na Figura 4.8.

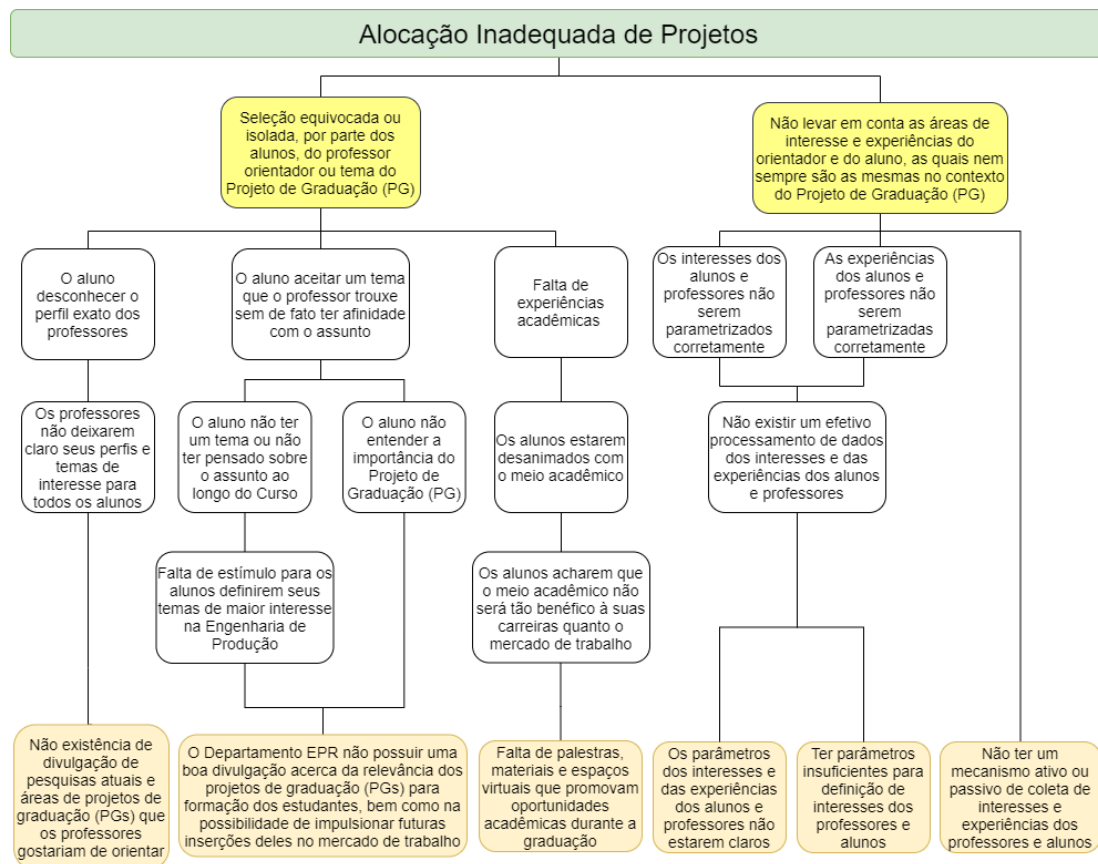


Figura 4.8: Análise de Causa Raiz dos riscos de alocação inadequada de PGs.

O risco de ocorrer uma seleção equivocada ou isolada, por parte dos alunos, do professor orientador ou tema do PG teve três possíveis causas raízes, a saber: a não existência de divulgação de pesquisas atuais e áreas de PGs que os professores gostariam de orientar; o Departamento EPR não possui uma boa divulgação acerca da relevância dos

PGs para formação dos estudantes, bem como na possibilidade de impulsionar futuras inserções deles no mercado de trabalho; e a falta de palestras, materiais e espaços virtuais que promovam oportunidades acadêmicas durante a graduação.

Ao realizar a alocação do projetos de graduação, também tem-se o risco de não levar em conta as áreas de interesse e experiências do orientador e do aluno, as quais nem sempre são as mesmas, de modo que as causas raízes podem ser: os parâmetros dos interesses e das experiências dos alunos e professores não estarem claros; ter parâmetros insuficientes para definição de interesses dos professores e alunos; e não ter um mecanismo ativo ou passivo de coleta de interesses e experiências dos professores e alunos.

Conforme dito anteriormente, a Análise de Causa Raiz também foi aplicada para os riscos de alocação inadequada de projetos para as disciplinas de PSP, a qual é apresentada na Figura 4.9.

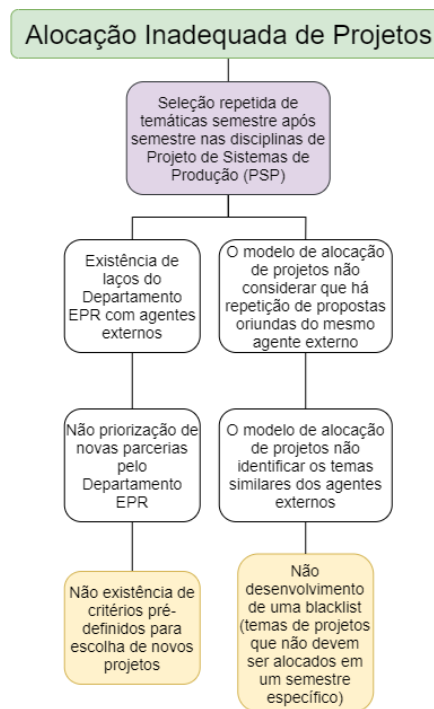


Figura 4.9: Análise de Causa Raiz dos riscos de alocação inadequada de projetos para as disciplinas de PSP.

Um dos riscos analisados no contexto das disciplinas de PSP foi o de seleção repetida de temáticas semestre após semestre, tendo como causa a não existência de critérios pré-definidos para escolha de novos projetos e o não desenvolvimento de uma *blacklist*, que, por sua vez, é formada por temas de projetos que não devem ser alocados em um semestre específico.

Ainda ao analisar os riscos a partir da perspectiva dos PSPs, tem-se a Análise de Causa Raiz do risco vinculado aos projetos interdisciplinares dessas disciplinas, conforme é apresentado na Figura 4.10.

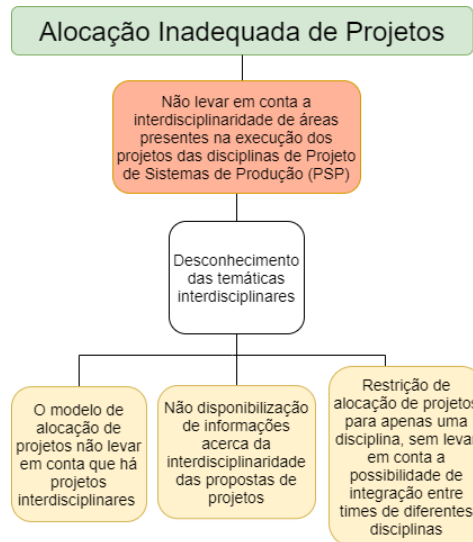


Figura 4.10: Análise de Causa Raiz dos riscos de alocação de projetos interdisciplinares para as disciplinas de PSP.

O risco de não levar em conta a interdisciplinaridade de áreas presentes na execução dos projetos das disciplinas de PSP pode ser originado pelas seguintes causas raízes: o modelo de alocação de projetos não levar em conta que há projetos interdisciplinares; a não disponibilização de informações acerca da interdisciplinaridade das propostas de projetos; e a restrição de alocação de projetos para apenas uma disciplina, sem levar em conta a possibilidade de integração entre times de diferentes disciplinas.

Portanto, ao aplicar a técnica de Análise de Causa Raiz, foi possível considerar a complexidade das causas originais ou raízes dos riscos, ao invés de apenas lidar com suas consequências imediatas no problema de alocação inadequada de projeto. Além disso, a aplicação dessa ferramenta permitiu o envolvimento de diferentes partes interessadas do processo de alocação de projetos em um ambiente de equipe, de modo a ser executar uma análise estruturada.

Com a Análise de Causa Raiz desenhada, foi necessário priorizar as causas raízes mais impactantes e prováveis para o problema de alocação inadequada de projetos. Isso porque a pesquisa não tem o intuito de apenas identificar fatores causais mas também tratar os riscos desse problema.

Para essa priorização, realizou-se uma adaptação da metodologia proposta por Pham et al. [44], na qual utiliza conceitos da técnica Análise de Causa Raiz juntamente com um modelo de priorização e redução de risco semelhante ao usado na aviação pela *Commercial Aviation Safety Team (CAST)*.

Sendo assim, cada participante do *brainstorming* de análise de riscos preencheu individualmente uma enquete para priorização das causas raízes. Essa enquete foi elaborada de modo a ter duas colunas, na qual a primeira coluna foi preenchida com a importância das causas raízes para o problema atual de alocação inadequada de projeto, cuja denominação é Poder 1. Já na segunda coluna, chamada de Aplicabilidade, o foco foi na importância das causas raízes em contribuir para ocorrência de riscos futuros de alocação inadequada de projetos.

Para preenchimento da enquete para priorização das causas raízes dos riscos, foi utilizado um sistema de 5 pontos da Escala Likert na ferramenta Google Sheets, sendo 1 (Sem importância), 2 (Pouco importante), 3 (Razoavelmente importante), 4 (Muito importante) e 5 (Extremamente importante).

Ao final do preenchimento individual da enquete pelos quatro participantes, foram multiplicadas as médias das pontuações elencadas (Poder 1 X Aplicabilidade), resultando no Índice de Priorização. Os resultados dessa enquete são apresentados no Apêndice A.

Assim, como última saída da análise de riscos, tem-se a priorização das causas raízes com base na importância em causar o problema atual e em contribuir para a ocorrência de riscos futuros de alocação inadequada de projetos.

Por fim, sabe-se que a análise de riscos pode ser influenciada por qualquer divergência de opiniões, vieses, percepções e julgamentos, de modo que os resultados apresentados poderiam ter sido diferentes, sendo influenciados pelas escolhas dos participantes do *brainstorming* e técnicas empregadas.

A análise de riscos fornece uma entrada para a avaliação de riscos que tem o enfoque nas decisões dos riscos que precisam ser tratados, bem como na estratégia mais apropriada para tratá-los.

### **4.2.3 Avaliação dos riscos**

A avaliação dos riscos envolveu a comparação dos resultados da análise dos riscos com os critérios de risco estabelecidos na Etapa 1, com o intuito de determinar onde será necessário aplicar ações de tratamento.

Para definição da significância do risco, foi utilizado o índice de priorização resultante da análise dos riscos, que foi calculado por meio da multiplicação dos valores médios elencados para representar a importância de cada causa raiz em causar o problema atual e em contribuir para a ocorrência de riscos futuros de alocação inadequada de projeto.

Deste modo, os riscos originados pelas causas raízes que tiveram índice de priorização maior do que 15 (vide Apêndice A), que é a mediana dos valores, foram inicialmente considerados prioritários para tratamento, os quais são apresentados na Tabela 4.8.

Tabela 4.8: Riscos inicialmente priorizados para tratamento.

Categoria	Riscos	Causas raízes
Utilização inadequada do Módulo	Os usuários não fornecerem as informações de forma completa, correta e atualizada	Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos
	Desuso do Módulo de Alocação de Projetos por parte dos usuários	Complexidade na alimentação dos dados na plataforma PUMA
Execução inadequada do processo de mineração de dados	Os projetos não serem alocados	Não especificação do nível de detalhamento requerido ao descrever o projeto
	Os projetos não apresentarem informações suficientes	
	Dificuldade no tratamento dos dados originados de bases externas	Desconhecimento de técnicas adequadas para o desenvolvimento da arquitetura necessária para o pleno funcionamento do Módulo de Alocação de Projetos
	Os projetos não serem alocados	A descrição da proposta submetida não refletir a temática do projeto
	Os projetos não apresentarem informações suficientes	
	Os projetos não serem alocados	Não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos
	projetos não apresentarem informações suficientes	
	Dificuldade no tratamento dos dados originados de bases externas	
	Os projetos não serem alocados	O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto
A plataforma PUMA não ter infraestrutura adequada para suportar a solução proposta pela pesquisa	Falta de definição da infraestrutura adequada para o funcionamento pleno do Módulo de Alocação de Projetos	
Alocação inadequada de projetos para os PSPs	Seleção repetida de temáticas semestre após semestre nas disciplinas de Projeto de Sistemas de Produção (PSP)	Não existência de critérios pré-definidos para escolha de novos projetos
		Não desenvolvimento de uma <i>blacklist</i> (temas de projetos que não devem ser alocados em um semestre específico)
Alocação inadequada de projetos interdisciplinares dos PSPs	Não levar em conta a interdisciplinaridade de áreas presentes na execução dos projetos dos PSPs	O modelo de alocação de projetos não levar em conta que há projetos interdisciplinares nos PSPs
		Não disponibilização de informações acerca da interdisciplinaridade das propostas de projetos

Por meio da Tabela 4.8, é possível observar que os riscos priorizados para tratamento envolvem todas as categorias identificadas para o problema de alocação inadequada de projetos. Alguns desses riscos prioritários também apresentaram causas raízes comuns.

Além disso, avaliação dos riscos considerou aqueles que pudessem afetar os resultados caso seja uma alocação tanto manual quanto automática; e que tivessem relação com alocação inadequada de projetos para disciplinas de PSP e projetos de graduação; além do que, para priorização, foi aplicado um processo de quantificação.

Deste modo, todos os critérios definidos na Etapa 1 para avaliação dos riscos foram cumpridos, fazendo com que, dos 12 riscos e 25 causas raízes identificados, fossem prioriza-



dos 8 riscos e 12 causas raízes para tratamento. Apenas o critério destinado ao tratamento dos riscos será atendido na Etapa 3 da pesquisa. Isso porque, após a avaliação dos riscos, serão projetadas e implementadas ações que podem eliminar, mitigar ou controlar os riscos identificados.

Paralelamente ao processo de avaliação dos riscos, foi realizada a fase de entendimento dos dados, cujo foco esteve na coleta e compreensão dos primeiros dados. Os resultados desta fase são apresentados na Seção 4.2.4.

#### 4.2.4 Entendimento dos dados

Inicialmente, foi feita uma análise preliminar dos dados listados na Etapa 1 como recursos disponíveis para a pesquisa, sendo eles os projetos de graduação (PGs); os currículos dos professores; os projetos desenvolvidos nas disciplinas de PSP; as ementas dessas disciplinas; e os artigos de congressos da ABEPRO, conforme é apresentado na Tabela 4.9.

Tabela 4.9: Análise preliminar dos dados.

Possíveis dados	Base de dados	Filtros relevantes	Resumo da análise preliminar dos dados
Projetos de Graduação (PGs)	Biblioteca Digital da Produção Intelectual Discente da UnB (BDM)	Orientadores, coorientadores e assuntos dos projetos	Os PGs possuem diversos metadados e formas de extração, entretanto eles não estão catalogados conforme as mesmas áreas do conhecimento abrangidas pelas disciplinas de PSP. Atualmente, a base é composta por 256 projetos da Engenharia de Produção.
Currículos dos professores das disciplinas de PSP	Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)	Nomes dos professores	Os currículos têm metadados que podem ser extraídos por diversos meios, como o Lattes Extrator para instituições e o Número Identificador dos currículos. Entretanto, não é possível filtrar os dados curriculares dos professores para obter informações apenas das áreas do conhecimento relacionadas às disciplinas de PSP.
Projetos realizados nas disciplinas de PSP	Não há uma base de dados estruturada de projetos de PSP	Disciplinas e áreas do conhecimento dos projetos	Atualmente, o armazenamento dos projetos desenvolvidos nas disciplinas de PSP é desestruturado e fica a cargo de cada professor, variando, assim, quanto a forma e o local. Além disso, alguns dos projetos não estão mais acessíveis pelos professores, seja por exclusão ou perda de acesso.
Ementas das disciplinas de PSP	Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA)	Disciplina de PSP referente a cada ementa	Algumas disciplinas de PSP não possuem ementas cadastradas no SIGAA, além de algumas delas terem ementas que não retratam as suas principais características, como conteúdo e áreas do conhecimento envolvidas.
Artigos de congressos da ABEPRO	Portal da Associação Brasileira de Engenharia de Produção (ABEBRO)	Áreas e subáreas do conhecimento dos artigos	A ABEPRO disponibiliza diversas publicações em anais de eventos, revistas e relatos de experiências. Essas publicações estão organizadas de acordo com a natureza de cada meio de divulgação, de modo que a forma de organização dos dados dos Anais Eletrônicos do Encontro Nacional de Engenharia de Produção (ENEGEP) é a mais similar aos PSPs.

A análise preliminar dos dados, apresentada na Tabela 4.9, foi feita sob a ótica da possibilidade de utilizar seus conjuntos ou subconjuntos para alocar os projetos nas disciplinas de PSP. Além disso, foi considerado o uso de um aprendizado supervisionado, em que as classes dos dados são definidas antecipadamente, visto que esse foi o tipo de aprendizagem mais comum nos trabalhos relacionados que possuem forte relação com o objetivo e o problema de mineração de dados desta pesquisa.

Observa-se que embora os PGs possam ser desenvolvidos nas mesmas áreas do conhecimento das disciplinas de PSP, eles não possuem esse tipo de classificação na base de dados BDM. Assim, seria necessário, antes da utilização dos dados, classificar eles com base nas mesmas áreas de conhecimento das disciplinas de PSP e depois validar essa classificação com especialistas, não sendo possível o seu uso no tempo hábil da presente pesquisa.

Os currículos dos professores podem refletir informações dos PSPs, pois espera-se que eles sejam especialistas nas áreas do conhecimento das disciplinas que ministram, mesmo que eles possam possuir experiências em outras áreas não vinculadas às suas disciplinas de PSP. Entretanto, não foi encontrada uma forma adequada de fazer o filtro das informações das áreas do conhecimento dos PSP dentro dos currículos dos professores, com o intuito de utilizá-las na alocação de projetos.

Foi feita também a análise preliminar das ementas das disciplinas, porém constatou-se que nem todas elas possuem dados armazenados no SIGAA e as armazenadas nesse sistema não possuem dados suficientemente representativos dos PSPs. Deste modo, o uso das ementas das disciplinas de PSP na alocação dos projetos dependeria do ajuste e da escrita de algumas delas, inviabilizando, assim, a utilização dos seus dados durante o tempo da pesquisa.

Ressalta-se que foram analisadas as ementas das disciplinas de PSP contidas no Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB [13], mas o nível de detalhamento delas continuou sendo insuficiente para utilizá-las na construção do modelo de alocação de projetos.

Na Tabela 4.9, também é destacado o fato de os projetos já realizados nas disciplinas de PSP não estarem ainda armazenados em uma base de dados estruturada. Sendo assim, mesmo que fosse possível obter alguns desses projetos, a quantidade deles não seria adequada para construção do modelo de alocação de projetos.

Neste sentido, espera-se que, por meio do desenvolvimento da plataforma PUMA, seja construído um repositório estruturado para armazenamento de projetos já realizados nas disciplinas de PSP, possibilitando, assim, que os seus dados sejam utilizados na construção dos próximos modelos de alocação de projetos.

Durante a análise preliminar dos dados, uma das bases disponíveis no Portal da ABE-PRO foi considerada a mais adequada para construção do modelo de alocação de projetos,

mais especificamente os anais do congresso ENEGEP.

Na base de dados do ENEGEP, os artigos são divididos da mesma forma que as disciplinas de PSP, por área do conhecimento da ABEPRO. Essa similaridade entre áreas do conhecimento do ENEGEP e das disciplinas de PSP pode ser justificada pelo Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB [13], pois o mesmo segue às diretrizes gerais para engenharia de produção produzidas pela Comissão de Diretrizes Curriculares da ABEPRO [86].

O Projeto Político Pedagógico do Curso de Engenharia de Produção da UnB [13] destaca que as disciplinas de PSP estão inseridas no Núcleo de Conteúdos Profissionalizantes, cuja organização de seus co-requisitos segue a nomenclatura das áreas do conhecimento propostas pela ABEPRO.

Assim, para compor a amostra de dados utilizada na fase de modelagem, foram selecionados 422 artigos do ENEGEP pertencentes às seguintes temáticas: Planejamento e Controle da Produção (PCP), Sistemas de Informação, Engenharia do Produto e Gestão da Qualidade. Essas temáticas foram compostas por áreas e subáreas do conhecimento da ABEPRO, as quais tiveram a sua nomenclatura e forma de organização aperfeiçoadas ao longo das edições do ENEGEP, conforme é apresentado na Tabela 4.10.

Tabela 4.10: Artigos do ENEGEP selecionados para a fase de modelagem.

Temáticas	Áreas	Subáreas	Quant.	Edições
Planejamento e Controle da Produção (PCP)	Engenharia de Operações e Processos da Produção	Planejamento, Programação e Controle da Produção	110	2019-2021
Sistemas de Informação	Engenharia Organizacional	Gestão da Informação	30	2019-2021
	Gestão do Conhecimento Organizacional	Gestão da Informação de Produção e Operações	78	2010-2018
Engenharia do Produto	Engenharia do Produto	Gestão do Desenvolvimento de Produto	24	2019-2021
		Processo de Desenvolvimento do Produto	45	2019-2021
		Planejamento e Projeto do Produto	19	2019-2021
	Gestão do Produto	Engenharia de Produto	14	2018
Gestão da Qualidade	Engenharia da Qualidade	Gestão de Sistemas da Qualidade	53	2021
		Planejamento e Controle da Qualidade	35	2021
		Normalização, Auditoria e Certificação para a Qualidade	5	2021
		Organização Metrológica da Qualidade	3	2021
		Confiabilidade de Processos e Produtos	6	2021

Destaca-se que os nomes das temáticas são os mesmos de 4 disciplinas de co-requisito de PSP, que, por sua vez, são disciplinas âncoras que apresentam os conteúdos técnicos dos PSP vinculados às áreas e subáreas do conhecimento propostas pela ABEPRO [13].

As 4 disciplinas de PSP que possuem como co-requisito essas temáticas escolhidas são: PSP2, vinculado a Sistemas de Informação em Engenharia de Produção (SIEP); PSP4, cujo tema principal é Planejamento e Controle da Produção (PCP); PSP5, que aplica conceitos de Gestão da Qualidade; e PSP6, cujo foco principal está na Engenharia do Produto. Todas essas 4 disciplinas de PSP são obrigatórias para os alunos de Engenharia de Produção da UnB, em que são envolvidas diretamente a resolução de problemas reais de agentes externos (stakeholders) conforme as respectivas temáticas apresentadas.

Deste modo, a presente pesquisa não englobou no modelo de alocação de projeto as seguintes disciplinas de PSP: as optativas para os alunos (PSP3 e PSP8); e as que, embora obrigatórias, possuem projetos cuja classificação em uma área do conhecimento específica tende a ser mais difícil, devido a natureza de seus conteúdos técnicos (PSP1 e PSP7).

A extração dos dados foi realizada de modo manual, em consequência de não ter sido encontrada disponível uma forma de realizar automaticamente essa ação no Portal da ABEPRO. Essa extração manual não foi considerada um empecilho para a pesquisa porque, conforme explicitado, espera-se utilizar no futuro uma base de projetos já realizados dentro da plataforma PUMA para atualização do modelo de alocação de projetos.

Desta maneira, a tabela construída para organizar o conjunto de dados extraídos da base do ENEGEP foi elaborada sob o olhar de sua posterior construção de um banco de dados de projetos na plataforma PUMA, em que espera-se ter as colunas e tipos de variável semelhantes. Essa tabela de dados foi construída sob o formato de arquivo CSV, que significa Comma Separated Values, em português, Valores Separados por Vírgulas.

Para facilitar o entendimento da dados, a tabela de dados com os 422 registros foi inserida no ambiente Google Colab para visualização e análise exploratória de dados, tendo como primeiro resultado a análise descritiva apresentada na Tabela Tabela 4.11.

Tabela 4.11: Análise descritiva básica da tabela de dados.

Colunas	Tipo de variável	Valores únicos	Valores nulos
ID_documento	Numérica	422	0
tematica	Categórica	4	0
area	Categórica	6	0
subarea	Categórica	12	0
titulo	Textual	422	0
palavras-chave	Textual	422	0
resumo	Textual	422	0
semestre	Categórica	1	0
ano	Numérica	12	0
disciplina	Categórica	4	0
tipo	Categórica	1	0

Observa-se que as colunas do conjunto de dados não possuem valores nulos, devido ao modo manual de extração e ao descarte de 21 documentos com base na qualidade e formato deles. Esses documentos foram descartados porque não possuíam algum metadado relevante (título, resumo ou palavras-chave); não estavam em português; ou tinham erros de codificação de caracteres. Além disso, foi realizada também uma análise da duplicidade dos registros, cujo resultado não apresentou a ocorrência de linhas duplicas na tabela de dados.

A Tabela 4.11 destaca também que os dados possuem mais de uma variável do tipo categórica. Entretanto, a mais importante são as 4 variáveis armazenadas na coluna temática, visto que serão utilizadas como classes no treinamento e teste do modelo de classificação. Já o corpus dos documentos será formado pela junção dos dados textuais, sendo eles: título, palavras-chave e resumo.

O conjunto de dados utilizado na pesquisa possui 9 colunas, sendo elas: ID\_documento, que armazena o número de identificação único de cada documento; temática, que possui a classificação de cada registro nas 4 temáticas selecionadas para alocação; area, subarea, título, palavras-chave e resumo, que armazenam os metadados principais retirados da base do ENEGEP; semestre e ano; que possui os registros dos períodos de origem de cada documento; disciplina; que armazena os PSP correspondentes às temáticas de cada documento; e tipo, que indica se o documento é artigo, PG ou projeto de PSP.

As colunas semestre e tipo receberam apenas um valor, os quais foram, respectivamente, 2 semestre e artigo, pois todos os documentos foram retirados de congresso realizados no segundo semestre de cada ano e eles são todos do tipo artigo. Já a coluna ano apresentou 12 valores únicos, de modo que a distribuição dos documentos por ano é apresentada na Figura 4.11.

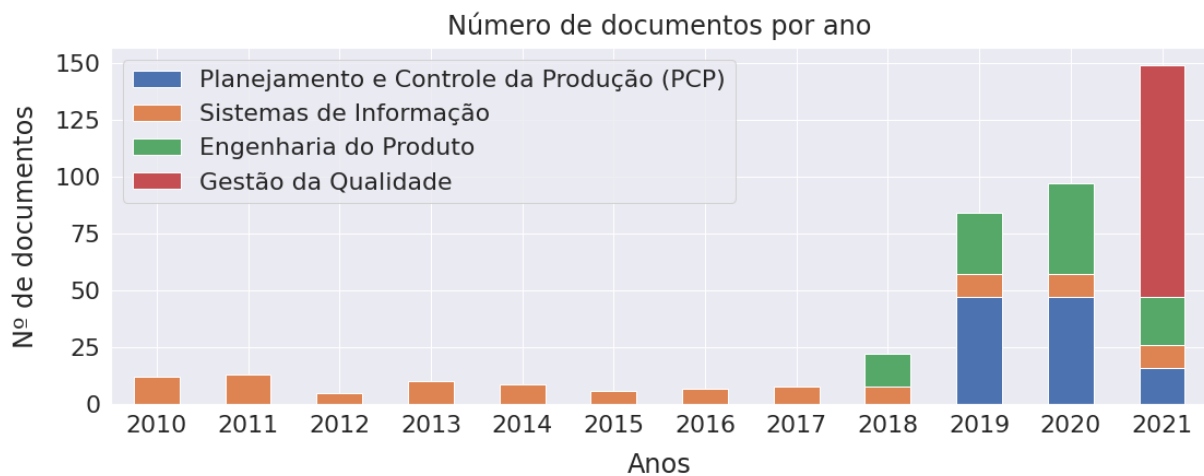


Figura 4.11: Distribuição dos documentos por ano.  
Fonte: [87].

Foi necessária a extração de artigos de mais uma edição do ENEGEP porque eram esperados ao menos 100 documentos de cada temática, mas o número de publicações por edição varia entre as temáticas.

Essa seleção de no mínimo 100 artigos por cada uma das 4 temáticas é justificada pelo fato de ser o dobro da quantidade utilizada no trabalho de Núñez e Ramos [9], que é um dos trabalhos relacionados mais aderentes aos objetivos, ao problema e aos dados da pesquisa. Núñez e Ramos [9] utilizaram 50 documentos de cada uma das 4 áreas profissionais trabalhadas em sua pesquisa. Ao total, foram extraídos 422 documentos na corrente pesquisa, cuja distribuição por temática é apresentada na Figura 4.12.

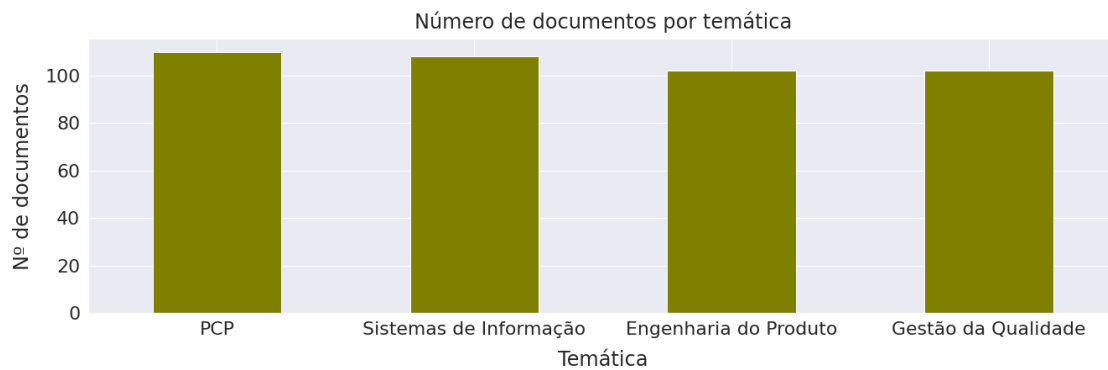


Figura 4.12: Gráfico de documentos por temática.  
Fonte: [87].

Os 422 documentos estão distribuídos em 110 de PCP, 108 de Sistemas de Informação, 102 Engenharia do Produto e 102 de Gestão da Qualidade, as quais são temáticas que representam, respectivamente, as disciplinas e PSP4, PSP2, PSP6 e PSP5. Esse equilíbrio no número de documentos entre as temáticas indica a não necessidade de aplicação de técnicas de balanceamento de dados.

As palavras mais recorrentes nos documentos que compõem o conjunto de dados foram analisadas, conforme a nuvem de palavras apresentada na Figura 4.13.

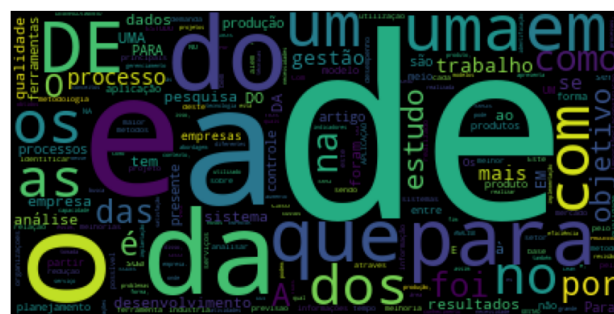


Figura 4.13: Nuvem de palavras que compõem o conjunto de dados.  
Fonte: [87].

Essa nuvem de palavras possibilitou a análise das palavras mais frequentes dentre as 84.867 existentes na junção das variáveis textuais (título, palavras-chave e resumo), sendo 13.551 delas palavras únicas. As palavras mais frequentes foram artigos, pronomes, preposições e conjunções, as quais são palavras consideradas irrelevantes para diferenciação dos documentos, pois são extremamente comuns. É possível observar também que há diferenças entre letras maiúsculas e minúsculas em palavras com o mesmo significado. Assim, é necessário preprocesar esses dados antes de fazer uso deles na fase de modelagem.

Deste modo, os resultados encontrados na Etapa 2 indicam a necessidade de implementação de ações para tratar os riscos de alocação inadequada de projetos, dentre as quais envolveu a construção de um modelo de alocação de projetos, conforme é apresentado na Etapa 3 (Seção 4.3).

### **4.3 Etapa 3 - Construção do modelo de alocação de projetos com base no tratamento dos riscos**

A partir dos riscos de alocação inadequada de projetos priorizados para tratamento, foram propostas ações para eliminação, mitigação ou controle das suas causas raízes. Parte dessas ações de tratamento propostas envolveu a preparação dos dados, modelagem e avaliação de um modelo de alocação de projetos para disciplinas de PSP, cujos resultados são apresentados a seguir.

#### **4.3.1 Tratamento dos riscos**

Para tratamentos dos riscos de alocação inadequada de projetos, os mesmos 4 *stakeholders* que participaram do processo de avaliação dos riscos foram instigados a preencher uma enquete online. Nessa enquete, cada uma das 12 causas raízes dos riscos considerados prioritários foram colocadas junto a um campo de preenchimento de sugestões de ações de tratamento. Além disso, os participantes foram instruídos quanto à possibilidade de colocar, mais de uma ação de tratamento para cada causa; e, caso não tivesse nenhuma sugestão de ação de tratamento para uma dada causa, foi possível deixar o campo de preenchimento em branco.

As ações de tratamento sugeridas pelos quatro participantes foram compiladas e avaliadas quanto a quanto bem elas poderiam contribuir para a eliminação, mitigação ou controle das causas raízes dos riscos (Poder 2); e a facilidade e o prazo necessário para implementação de cada uma delas. Desta maneira, as que receberam maiores valores ao multiplicar Poder 2 x Facilidade e que foram assinaladas como de curto prazo de implementação (3 meses), são apresentadas na Tabela 4.12

Tabela 4.12: Ações de tratamento prioritizadas.

Riscos prioritizados	Causas raízes prioritizadas	Ações de tratamento prioritizadas
Os projetos não serem alocados	Não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos	Fazer a descrição do contexto do processo de metodologias ativas no curso de Engenharia de Produção da UnB para que seja contemplado na alocação dos projetos.
Os projetos não apresentarem informações suficientes		Investigar as questões de negócio relacionadas à alocação inadequada de projetos, visando integrá-las ao desenvolvimento do Modelo de Alocação de Projetos.
Dificuldade no tratamento dos dados originados de bases externas		Realizar benchmarking quanto a problemas e soluções similares.
Os usuários não fornecerem as informações de forma completa, correta e atualizada	Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos	Estabelecer metodologia robusta que envolva técnicas mais aderentes ao levantamento de requisitos.
Desuso do Módulo de Alocação de Projetos por parte dos usuários	Complexidade na alimentação dos dados na plataforma PUMA	Padronizar os campos de alimentação dos dados a fim de unificar inputs de dados e facilitar a interpretação dos mesmos.
Os projetos não serem alocados	Não especificação do nível de detalhamento requerido ao descrever o projeto	Garantir que todas as informações necessárias para posterior alocação estarão contidas na proposta de projeto.
Os projetos não apresentarem informações suficientes		
Dificuldade no tratamento dos dados originados de bases externas	Desconhecimento de técnicas adequadas para o desenvolvimento da arquitetura necessária para o pleno funcionamento do Módulo de Alocação de Projetos	Desenvolver um plano de arquitetura do Módulo de Alocação de Projetos por meio do suporte de uma equipe / parceria externa composta por recursos humanos qualificados, levando em conta uma infraestrutura AWS e containers para escala.
		Implementar a arquitetura do Módulo de Alocação de Projetos por meio de uma parceria externa.
Os projetos não serem alocados	O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto	Inserir no algoritmo de alocação de projetos a verificação de uma possível adequação a alguma disciplina de PSP, mesmo que contemple apenas parte da proposta de projeto, a depender da similaridade.
A plataforma PUMA não ter infraestrutura adequada para suportar a solução proposta pela pesquisa	Falta de definição da infraestrutura adequada para o funcionamento pleno do Módulo de Alocação de Projetos	Descrever do passo-a-passo para implementação da infraestrutura adequada para hospedar o Módulo de Alocação de Projetos
		Garantir a infraestrutura adequada para hospedar o módulo de alocação de projetos (exemplo: parcerias com fornecedores de tecnologia ou uso de infraestrutura sem custo)
Os usuários não fornecerem as informações de forma completa, correta e atualizada	Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos	Seguir a metodologia estabelecida para construção de um MVP do Módulo de Alocação de Projetos, cujo os requisitos devem ser levantados com base em uma ampla pesquisa sobre requisitos de software e métodos para alocação de projetos.
Os projetos não serem alocados	Não especificação do nível de detalhamento requerido ao descrever o projeto	Fazer com que a descrição da proposta de projeto seja feita por meio de perguntas separadas que contemplem todas as informações necessárias para posterior alocação.
Os projetos não apresentarem informações suficientes		
Os projetos não serem alocados	O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto	Criar processo / mecanismo para comunicar o proponente sobre a inadequação / rejeição de sua proposta.



Observa-se que ações de tratamento prioritizadas têm o horizonte de implementação dentro do tempo de execução desta pesquisa, o qual é de 3 meses. As demais ações que foram consideradas de médio e longo prazo são apresentadas no Apêndice B.

Na Tabela 4.13, são apresentadas a ordem e as etapas em que ações de tratamento de curto prazo foram implementadas integralmente ou parcialmente na pesquisa.

Tabela 4.13: Etapas da pesquisa em que serão implementadas as ações de tratamento.

Ordem	Ações de tratamento prioritizadas	Etapas de implementação na Pesquisa
1	Realizar benchmarking quanto a problemas e soluções similares.	Revisão Bibliográfica (Trabalhos relacionados)
2	Fazer a descrição do contexto do processo de metodologias ativas no curso de Engenharia de Produção da UnB para que seja contemplado na alocação dos projetos.	Etapas 1 (Estabelecimento do contexto)
3	Investigar as questões de negócio relacionadas à alocação inadequada de projetos, visando integrá-las ao desenvolvimento do Modelo de Alocação de Projetos.	Etapas 1 (Compreensão do Negócio)
4	Garantir que todas as informações necessárias para posterior alocação estarão contidas na proposta de projeto.	Etapas 3 (Preparação dos Dados e Modelagem)
5	Padronizar os campos de alimentação dos dados a fim de unificar inputs de dados e facilitar a interpretação dos mesmos.	
6	Inserir no algoritmo de alocação de projetos a verificação de uma possível adequação a alguma disciplina de PSP, mesmo que contemple apenas parte da proposta de projeto, a depender da similaridade.	
7	Estabelecer metodologia robusta que envolva técnicas mais aderentes ao levantamento de requisitos.	Etapas 4 (Implantação do Módulo)
8	Desenvolver um plano de arquitetura do Módulo de Alocação de Projetos por meio do suporte de uma equipe / parceria externa composta por recursos humanos qualificados, levando em conta uma infraestrutura AWS e containers para escala.	
9	Implementar a arquitetura do Módulo de Alocação de Projetos por meio de uma parceria externa.	
10	Garantir a infraestrutura adequada para hospedar o módulo de alocação de projetos (exemplo: parcerias com fornecedores de tecnologia ou uso de infraestrutura sem custo)	
11	Descrever do passo-a-passo para implementação da infraestrutura adequada para hospedar o Módulo de Alocação de Projetos	
12	Fazer com que a descrição da proposta de projeto seja feita por meio de perguntas separadas que contemplem todas as informações necessárias para posterior alocação.	
13	Criar processo / mecanismo para comunicar o proponente sobre a inadequação / rejeição de sua proposta.	
14	Seguir a metodologia estabelecida para construção de um MVP do Módulo de Alocação de Projetos, cujo os requisitos devem ser levantados com base em uma ampla pesquisa sobre requisitos de software e métodos para alocação de projetos.	

Destaca-se que três ações de tratamento foram implementadas por meio da execução das fases de preparação dos dados e modelagem que são descritas nas seções 4.3.2 e 4.3.3.

### 4.3.2 Preparação dos dados

O entendimento dos dados realizado na Etapa 2 indicou a necessidade da aplicação de técnicas de pré-processamento de dados, visando a limpeza e preparação final do conjunto de dados selecionado para treinamento e teste do modelo de alocação de projetos.

Essas técnicas permitem a definição do espaço de termos dos dados, fazendo com que seu volume seja menor, mas produzindo resultados idênticos ou similares. Por conseguinte foram aplicadas as técnicas que são apresentadas na Figura 4.14.

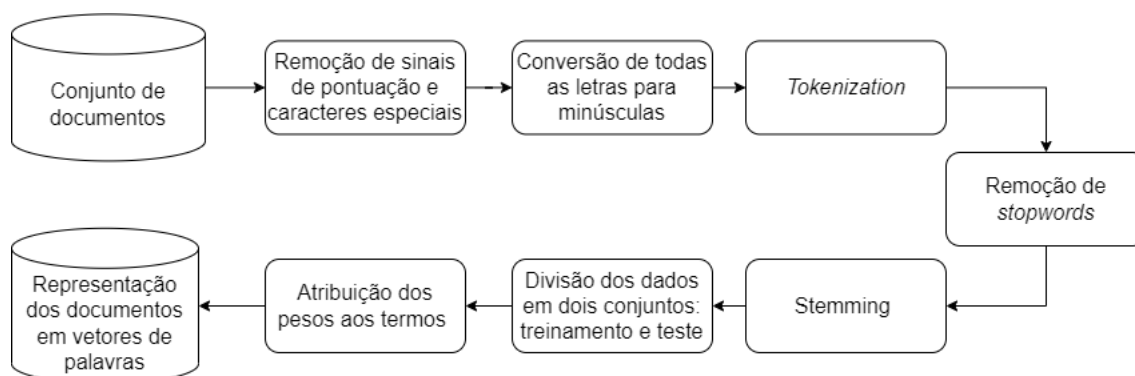


Figura 4.14: Técnicas aplicadas para preparação dos dados.

O processo apresentado na Figura 4.14 apresenta os seguintes tratamentos realizados nos dados:

- (i) Remoção de sinais de pontuação e caracteres especiais:

"!\$"%'()\*+,-./:;<=>?@[|

- (ii) Conversão de todas as letras para minúsculas;

- (iii) *Tokenization* para quebra dos dados textuais em suas palavras constituintes, de modo foram criadas strings;

- (iv) Remoção de *stopwords* a partir da seguinte lista em português da biblioteca *Natural Language ToolKit* (NLTK):

['de', 'a', 'o', 'que', 'e', 'é', 'do', 'da', 'em', 'um', 'para', 'com', 'não', 'uma', 'os', 'no', 'se', 'na', 'por', 'mais', 'as', 'dos', 'como', 'mas', 'ao', 'ele', 'das', 'à', 'seu', 'sua', 'ou', 'quando', 'muito', 'nos', 'já', 'eu', 'também', 'só', 'pelo', 'pela', 'até', 'isso', 'ela', 'entre', 'depois', 'sem', 'mesmo', 'aos', 'seus', 'quem', 'nas', 'me', 'esse', 'eles', 'você', 'essa', 'num', 'nem', 'suas', 'meu', 'às', 'minha', 'numa', 'pelos', 'elas', 'qual', 'nós', 'lhe', 'deles', 'essas', 'esses', 'pelas', 'este', 'dele', 'tu', 'te', 'vocês', 'vos', 'lhes', 'meus', 'minhas', 'teu', 'tua', 'teus', 'tuas', 'nosso', 'nossa', 'nossos', 'nossas', 'dela', 'delas', 'esta', 'estes', 'estas', 'aquele', 'aquela', 'aqueles', 'aquelas', 'isto', 'aquilo',

'estou', 'está', 'estamos', 'estão', 'estive', 'esteve', 'estivemos', 'estiveram', 'estava', 'estávamos', 'estavam', 'estivera', 'estivéramos', 'esteja', 'estejamos', 'estejam', 'estivesse', 'estivéssemos', 'estivessem', 'estiver', 'estivermos', 'estiverem', 'hei', 'há', 'havemos', 'hão', 'houve', 'houvemos', 'houveram', 'houvera', 'houvéramos', 'haja', 'hajamos', 'hajam', 'houvesse', 'houvéssemos', 'houvessem', 'houver', 'houvermos', 'houverem', 'houverei', 'houverá', 'houveremos', 'houverão', 'houveria', 'houveríamos', 'houveriam', 'sou', 'somos', 'são', 'era', 'éramos', 'eram', 'fui', 'foi', 'fomos', 'foram', 'fora', 'fôramos', 'seja', 'sejamos', 'sejam', 'fosse', 'fôssemos', 'fossem', 'for', 'formos', 'forem', 'serei', 'será', 'seremos', 'serão', 'seria', 'seríamos', 'seriam', 'tenho', 'tem', 'temos', 'tém', 'tinha', 'tínhamos', 'tinham', 'tive', 'teve', 'tivemos', 'tiveram', 'tivera', 'tivéramos', 'tenha', 'tenhamos', 'tenham', 'tivesse', 'tivéssemos', 'tivessem', 'tiver', 'tivermos', 'tiverem', 'terei', 'terá', 'teremos', 'terão', 'teria', 'teríamos', 'teriam']

- (v) *Stemming* para derivação dos termos até os seus radicais a fim de agrupá-los pelos mesmos significados conceituais [9, 60]. Para retirada dos sufixos, foi utilizado o Removedor de Sufixos da Língua Portuguesa (RSLP Stemmer) que está disponível na biblioteca NLTK.
- (vi) Divisão dos dados em dois conjuntos, sendo 80% para treinamento do modelo e o restante, 20%, para seu teste.
- (vii) Atribuição dos pesos aos termos dos conjuntos de treinamento e de teste por meio da aplicação do método TF-IDF.

Em suma, a saída principal da fase de preparação de dados foi a representação dos documentos em vetores de termos no *Vector Space Mode* (VSM), visto que, após limpeza dos dados, para cada termo foi atribuído um valor que equivale ao seu peso no conjunto de dados. Esta estrutura de vetores foi utilizada na modelagem (Seção 4.3.3).

### 4.3.3 Modelagem

Para construção do modelo de alocação de projetos, foram aplicadas os métodos de aprendizado supervisionados que foram utilizados nos trabalhos relacionados com maior aderência aos objetivos, ao problema e aos dados da presente pesquisa, a saber: o *K-Nearest Neighbors* (K-NN), Classificador Naïve Bayes e *Support Vector Machine* (SVM). O código desenvolvido é apresentado no Apêndice C.

Por meio desses métodos, foi treinado e testado o modelo proposto de alocação de projetos para disciplinas de PSP que é apresentado na Figura 4.15.

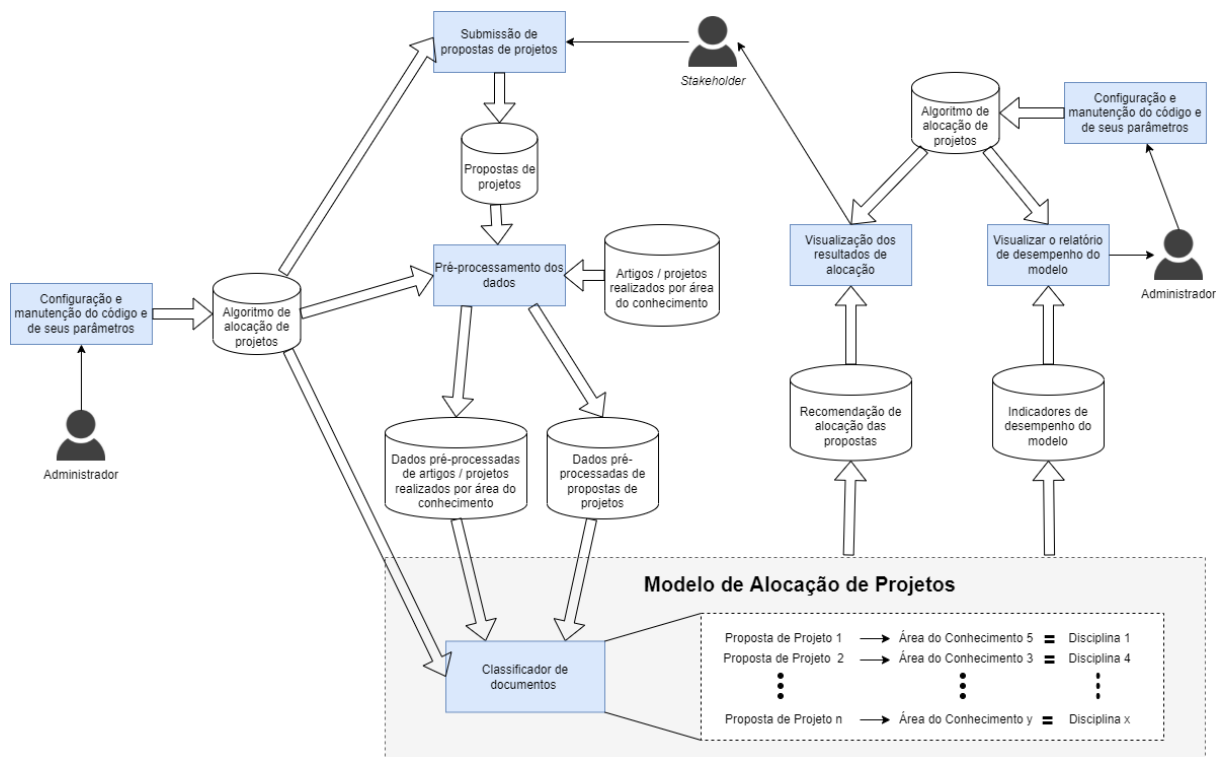


Figura 4.15: Diagrama de blocos representando o modelo proposto de alocação de projetos.

Dentre os métodos utilizados, o K-NN foi o que apresentou o melhor desempenho, de modo que os parâmetros utilizados e seus resultados são apresentados a seguir.

## K-NN

No algoritmo K-NN, a qualidade das classificações depende principalmente de dois parâmetros: o valor de K e a métrica de distância empregada. Portanto, diferentes valores para esses parâmetros foram testados a fim de se obter os melhores resultados de desempenho, em especial o maior valor de acurácia, de acordo com o que é apresentado na Figura 4.16

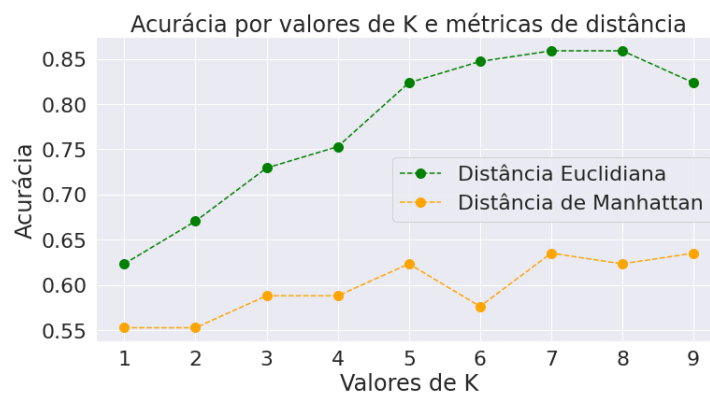


Figura 4.16: Gráfico de acurácia por valores de K e métricas de distância.

Observa-se que foram testados os valores de 1 a 9 para K e as distâncias Euclidiana e de Manhattan. Assim, o maior valor de acurácia encontrado foi de 86% para os valores 7 e 8 para K e a Distância Euclidiana, optando-se por K=7 porque demanda menos recurso computacional ao rodar o modelo de alocação de projetos.

Para se entender melhor os resultados obtidos do modelo, foi gerada a matriz de confusão apresentada na Figura 4.17.

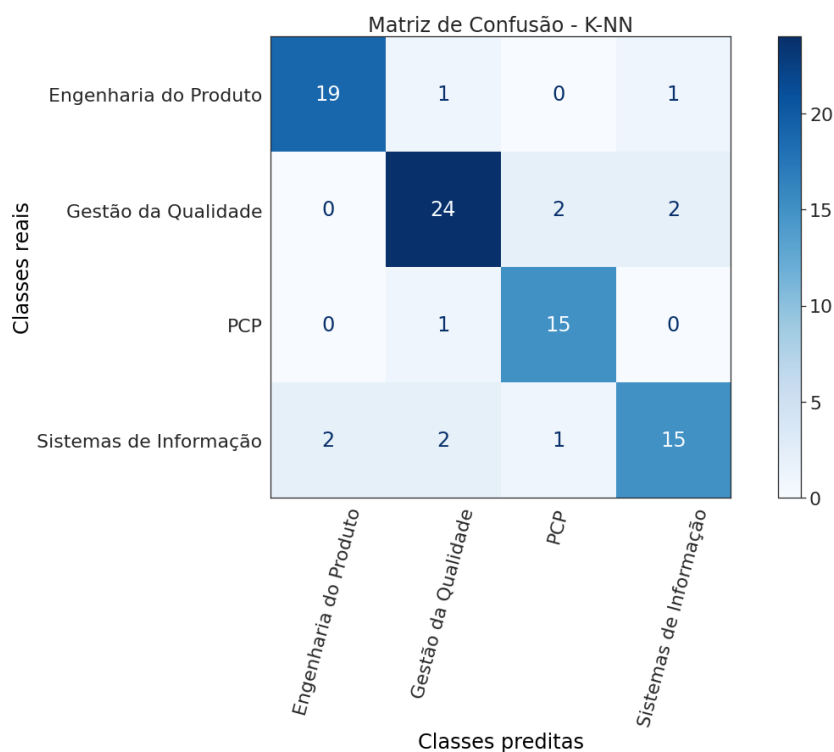


Figura 4.17: Matriz de Confusão - K-NN.

A Matriz de Confusão apresenta o número de casos avaliados, em que as classes reais estão na linha e as classes previstas pelo classificador estão na coluna. Dito isso, é possível notar que os valores mais altos estão na diagonal principal, tendo valores de 0 a 2 no restante da matriz de confusão. Isso indica que os resultados do classificador apresentam uma boa taxa de predição. A partir da matriz de confusão, foi gerado o relatório de classificação apresentado na Tabela 4.14.

Tabela 4.14: Relatório de Classificação - K-NN

Temáticas	precision	recall	f1-score	suporte
Engenharia do Produto	0.90	0.90	0.90	21
Gestão da Qualidade	0.86	0.86	0.86	28
Planejamento e Controle da Produção (PCP)	0.83	0.94	0.88	16
Sistemas de Informação	0.83	0.75	0.79	20
accuracy	0.86			

Após essa avaliação técnica do modelo, todos os dados foram utilizados na construção do modelo final de alocação de projetos para as disciplinas de PSP, o qual também foi avaliado a partir da perspectiva do negócio, conforme é apresentado na Seção 4.3.4

#### 4.3.4 Avaliação

O modelo de alocação de projetos foi avaliado com o uso de 20 propostas de projetos que foram coletadas após a sua construção final, visando a sua avaliação em um contexto que simulasse a sua utilização real. Para isso, foi criada a interface apresentada na Figura 4.18.

**Submissão de Proposta de Projeto**

**Título:** " Aplicação da ISO 9001:2015 no contexto de empresa do terceiro setor "

**Descrição:** " O objetivo deste projeto é a avaliação do Sistema de Gestão da Qualidade (SGQ) "

↳ A área do conhecimento da proposta é ['ENGENHARIA DA QUALIDADE']

Figura 4.18: Interface para avaliação do modelo final de alocação de projetos.

A comparação entre os resultados da alocação realizada pelo modelo de alocação de projetos e pelos especialistas são apresentados na Tabela 4.15.

Tabela 4.15: Comparativo dos resultados da avaliação do Módulo de Alocação de Projetos

ID da proposta de projeto	Temática / disciplina recomendada pelos especialistas	Temática / disciplina recomendada pelo Módulo de Alocação de Projetos
1	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
2	PCP - PSP4	PCP - PSP4
3	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
4	Gestão da Qualidade - PSP5	PCP - PSP4
5	Engenharia do Produto - PSP6	Engenharia do Produto - PSP6
6	Sistemas de Informação - PSP2	Sistemas de Informação - PSP2
7	Sistemas de Informação - PSP2	Gestão da Qualidade - PSP5
8	Engenharia do Produto - PSP6	Engenharia do Produto - PSP6
9	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
10	PCP - PSP4	PCP - PSP4
11	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
12	PCP - PSP4	PCP - PSP4
13	PCP - PSP4	Engenharia do Produto - PSP6
14	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
15	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
16	Engenharia do Produto - PSP6	Gestão da Qualidade - PSP5
17	Engenharia do Produto - PSP6	Sistemas de Informação - PSP2
18	Engenharia do Produto - PSP6	Engenharia do Produto - PSP6
19	Gestão da Qualidade - PSP5	Gestão da Qualidade - PSP5
20	PCP - PSP4	PCP - PSP4

Com base nos resultados da Etapa 3, foi decidido acerca da implantação do modelo de alocação de projetos na plataforma PUMA. Essa implantação seguirá a proposta apresentada na Etapa 4 (Seção 4.4).

## 4.4 Etapa 4 - Proposta de Implantação do Módulo de Alocação de Projetos na plataforma PUMA

Para implementação do modelo elaborado na Etapa 3, foi necessário a proposição de um Módulo de Alocação de Projetos para a plataforma PUMA. Os requisitos deste módulo foram levantados por meio da integração entre disciplinas da Engenharia de Produção do Campus Darcy Ribeiro e da Engenharia de Software da Faculdade UnB Gama (FGA).

Essa integração possibilitou a construção do protótipo de alta fidelidade do MVP do Módulo de Alocação de Projetos, conforme a primeira tela apresentada na Figura 4.19.

A imagem mostra a tela de login do protótipo do módulo PUMA. O cabeçalho da tela é azul escuro com o texto "PUMA" em branco. Abaixo, há um formulário branco com um cabeçalho amarelo que diz "Bem-vindo(a) ao PUMA". O formulário contém dois campos de entrada: "E-mail" e "Senha". Abaixo dos campos, há o texto "Não possui uma conta? Cadastre-se" e "Esqueci minha senha". No final do formulário, há um botão amarelo com o texto "Entrar".

Figura 4.19: Tela de login do protótipo do módulo.

Na tela de login, os usuários, sejam eles agentes externos ou professor, poderão ter acesso a plataforma PUMA, desde que já tenha realizado seu cadastro (Figura 4.20).

A imagem mostra a tela de cadastro do protótipo do módulo PUMA. O cabeçalho da tela é azul escuro com o texto "PUMA" em branco. Abaixo, há um formulário branco com um cabeçalho amarelo que diz "CADASTRO DE USUÁRIO". O formulário contém cinco campos de entrada: "Nome Completo", "E-mail", "Senha" (com ícone de olho para alternar visibilidade), "Repita sua senha" e "Que tipo de usuário você é?" (menu suspenso). No final do formulário, há um botão amarelo com o texto "Cadastrar".

Figura 4.20: Tela de cadastro protótipo do módulo.

Os usuários terão a opção de escolher os seus perfis, podendo ser professor, aluno ou agente externo. Após realizar o seu cadastro, o usuário poderá recuperar sua senha caso esqueça, conforme é apresentado na Figura 4.21.

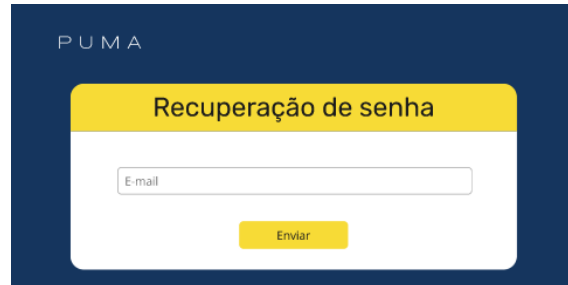


Figura 4.21: Tela de recuperação de senha protótipo do módulo.

Observa-se que será necessário que o usuário o seu e-mail de cadastro para receber o lembrete de seus dados. Caso o usuário cadastrado seja um agente externo, ao entrar terá acesso a tela apresentada mostrada na Figura 4.22.

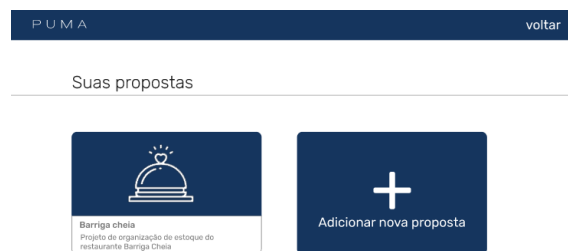


Figura 4.22: Tela inicial dos agentes externos protótipo do módulo.

Os agentes externos terão a opção de acompanhar o status de suas propostas ou de submeter um nova proposta de projeto baseada em seus problemas reais, em que a tela é apresentada na Figura 4.23.

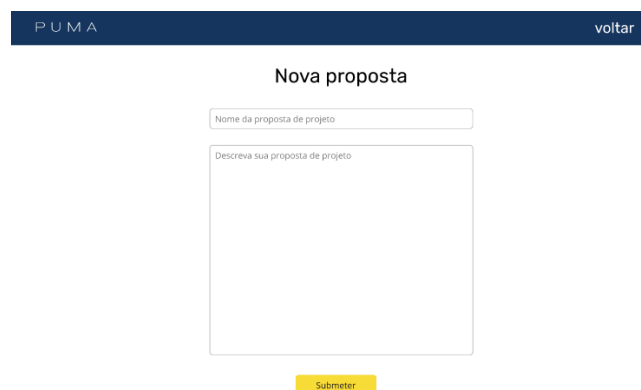


Figura 4.23: Tela de submissão de projetos do protótipo do módulo.



Na tela de submissão de projetos, os agentes externos precisarão descrever o problema real que precisa ser resolvido pelos alunos, com um nível de detalhe suficiente para o entendimento do escopo dos projetos propostos. Após a submissão das propostas de projetos, elas serão automaticamente alocadas para as disciplinas de PSP criadas na Plataforma PUMA, com base no modelo de alocação de projetos elaborado nesta pesquisa. A tela inicial dos professores das disciplinas de PSP que possibilitará a criação delas é apresentada na Figura 4.24.

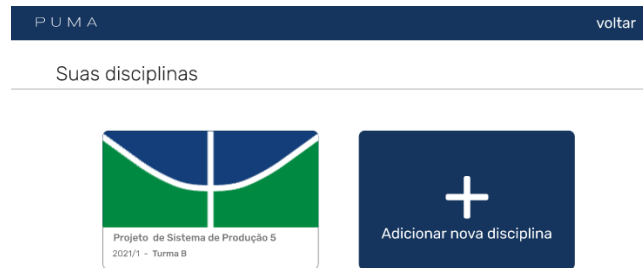


Figura 4.24: Tela inicial dos professores no protótipo do módulo.

A princípio, os professores terão duas opções em sua tela inicial: acessar uma disciplina já existente ou criar uma nova. Ao desejar criar uma disciplina, os usuários com perfil de professor serão direcionados para a tela apresentada na Figura 4.25.

A interface de usuário para a criação de uma nova disciplina. No topo, há uma barra de navegação azul escura com o texto "PUMA" à esquerda e um ícone de menu (três linhas horizontais) à direita. Abaixo, o título "Nova disciplina" é seguido por um formulário. O formulário contém: um campo de texto para "Nome da disciplina"; três campos de texto para "Turma", "Ano letivo" e "Semestre"; um campo de texto grande para "Descreva a disciplina dentro de 1000 caracteres."; e um campo de texto para "Senha de acesso (opcional)". No final do formulário, há um botão amarelo com o texto "Submeter".

Figura 4.25: Tela de criação de disciplina do protótipo do módulo.

A criação das disciplinas é necessária para alocação dos projetos, visto que neste momento são vinculadas suas as áreas e subáreas do conhecimento. Caso o professor deseje acessar uma disciplina criada anteriormente, ele terá acesso a tela apresentada na Figura 4.26.

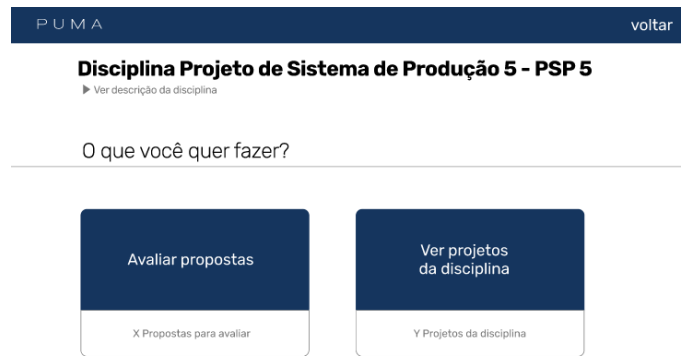


Figura 4.26: Tela das disciplinas do protótipo do módulo.

Em cada disciplina, os professores podem avaliar as propostas que foram alocadas para ela ou visualizar projetos que já foram avaliados. Assim, a tela de avaliação de propostas de projeto é exibida na Figura 4.27.



Figura 4.27: Tela de avaliação de propostas de projeto do protótipo do módulo.

Por meio da tela de avaliação de propostas de projeto, os professores podem optar por confirmar ou não que a alocação automática de projeto foi assertiva. Assim, caso ocorra algum problema de classificação pelo algoritmo, é dada a opção de um ajuste manual de alocação dos projetos.

Deste modo, construiu-se um protótipo de alta fidelidade do MVP do Módulo de Alocação de Projetos, com base na revisão da literatura realizada, que resultou em 13 trabalhos relacionados; no treinamento e teste do modelo de classificação de documentos quanto a quatro temáticas da Engenharia de Produção; e na execução das demais etapas da metodologia de pesquisa.

# Capítulo 5

## Conclusão

### 5.1 Considerações Finais

A presente pesquisa teve como objetivo geral propor uma ferramenta computacional para tratar os riscos de alocação inadequada de projetos em disciplinas que utilizam a aprendizagem ativa, no âmbito do Curso de Engenharia de Produção da UnB. Esse objetivo geral foi alcançado, uma vez que atingiu-se os seis objetivos específicos definidos a partir dele.

O primeiro objetivo específico foi alcançado por meio do entendimento do contexto e das questões de negócio que permeiam a alocação de propostas de projetos no Curso de Engenharia de Produção da UnB, o qual teve como principais resultados o estabelecimento dos contextos interno e externo; a avaliação da situação atual para desenvolvimento da pesquisa; a delimitação do escopo de estudo; a definição de critérios para as etapas executadas; e a seleção de técnicas e ferramentas aplicadas.

Em relação ao segundo objetivo específico, foi aplicado o processo de avaliação de riscos à alocação de projetos, no âmbito da metodologia PBL adotada no Curso. Esse processo teve foco nos eventuais riscos que poderiam impactar a pesquisa ou que teriam como consequência o problema de alocação inadequada de projetos. Como resultado, os riscos foram identificados, analisados e avaliados, de modo que, ao final, obteve-se uma lista com os riscos prioritários para tratamento.

A partir dos riscos priorizados, foi possível definir ações de tratamento, cujo foco esteve eliminação, mitigação ou controle de suas causas-raízes. De modo que, parte dessas ações envolveram a proposição do modelo de alocação de projetos. Deste modo, atingiu-se o terceiro objetivo específico da pesquisa, cujo escopo foi sugerir um modelo de alocação de projetos, considerando o tratamento dos riscos avaliados.

Para atingimento do quarto objetivo específico, validou-se o modelo proposto quanto à recomendação de disciplinas mais adequadas para os projetos. Essa verificação foi feita

a partir da mensuração do desempenho do modelo, a qual teve com base a aplicação do método 10- *fold cross validation*; e os valores das medidas *Accuracy*, *Precision*, *Recall* e *F-measure*. De maneira geral, o algoritmo teve um desempenho satisfatório, com uma taxa de acerto de 86%, a qual ficou acima dos valores encontrados nos trabalhos relacionados.

Além dessa validação técnica, foi verificado se o modelo proposto atende às necessidades das partes interessadas e aos objetivos de negócio do Curso, conforme estabelecido no quinto objetivo específico da pesquisa. É possível atestar o cumprimento desse objetivo devido ao fato do modelo proposto ter contribuído para tratamento dos riscos associados à alocação inadequada de projetos, bem como ter a capacidade de alocar as propostas de projetos nas disciplinas de PSP, conforme a validação feita, sob a perspectiva de negócio, por quatro *stakeholders* do Curso de Engenharia de Produção da UnB.

O sexto objetivo específico também foi alcançado, uma vez que foi sugerida uma forma de implementação do modelo proposto como um módulo da plataforma PUMA para a alocação de projetos, que envolveu a definição da sua arquitetura; a escolha das suas tecnologias de desenvolvimento; e a construção do seu protótipo de alta fidelidade baseado no seu Mínimo Produto Viável.

Destaca-se que, além do cumprimento dos objetivos geral e específicos, este estudo possibilitou dar uma resposta a questão de pesquisa, em razão ter proposto uma forma assertiva e rápida de atribuir as propostas de projetos para as disciplinas de PSP, de modo que parte dos riscos relacionados a este processo de alocação foram tratados.

A revisão de literatura serviu de base para toda a pesquisa, desde o entendimento do contexto e das questões de negócio até a proposição de módulo de alocação de projetos para a plataforma PUMA. Sem o esclarecimento dos conceitos de gestão de riscos e de mineração de dados e o levantamento dos trabalhos relacionados, não seria possível o alcance dos objetivos dessa pesquisa. Os trabalhos relacionados foram essenciais na identificação e tratamento dos riscos; na seleção das técnicas aplicadas; e no desenvolvimento das etapas.

Vale ainda ressaltar que, ao aplicar a técnica de Análise de Causa Raiz, foi possível considerar a complexidade das causas originais ou raízes dos riscos, ao invés de apenas lidar com suas consequências imediatas no problema de alocação inadequada de projeto. Sendo assim, a aplicação dessa ferramenta permitiu o envolvimento de diferentes partes interessadas do processo de alocação de projetos em um ambiente de equipe, de modo a ser executar uma análise estruturada.

A utilização dos processos baseados no CRISP-DM e na ISO 31000 foi essencial ao longo de toda a pesquisa, pois auxiliou no seu planejamento e no desenrolar das suas etapas. Portanto, pode-se inferir a importância da utilização de modelos de referência e normas em projetos de mestrado, pois contribuiu para uma boa estruturação das atividades.

O desenvolvimento desta pesquisa contribui para publicação de três artigos indexados na base científica Scopus [21, 18, 19]. Esses artigos são resultantes, principalmente, das parcerias nacionais e internacionais que foram citadas neste documento, as quais envolveram alunos e professores de graduação e de mestrado de dois países: o Brasil, com integrantes de cursos da UnB, mais especificamente da Engenharia de Produção e do Mestrado Profissional em Computação Aplicada (MPCA) do Campus Darcy Ribeiro e da Engenharia de Software da Faculdade UnB Gama (FGA); e a Tunísia, em parceria com o Curso de Ciência da Computação da Universidade ESPRIT.

Os resultados positivos dessa colaboração entre países e cursos diferentes na pesquisa não foram atestados apenas pela publicação dos artigos citados, mas também pelas práticas de compartilhamento de ideias entre os alunos e professores, dado que elas possibilitaram reuniões multidisciplinares de coleta de informações de negócio e a proposição uma forma de implementar o módulo de alocação de projetos na plataforma PUMA.

Ademais, a pesquisa teve fatores limitantes que geraram as sugestões de trabalhos futuros apresentadas na Seção 5.2.

## 5.2 Sugestões de Trabalhos Futuros

Devido a constatação de que a proposta de módulo de alocação projetos desenvolvida nesta pesquisa pode ser uma forma assertiva e rápida de atribuir as propostas de projetos para as disciplinas de PSP, tem-se a necessidade de implementá-la durante as próximas etapas de desenvolvimento da plataforma PUMA, tendo como ponto de partida a arquitetura definida; as tecnologias de desenvolvimento escolhidas; e o protótipo de alta fidelidade construído.

Devido à limitação de acesso aos dados de projetos já realizados nas disciplinas de PSP, posto que eles ainda não estão armazenados em uma base estruturada, não foi possível utilizar seus dados no treinamento e teste do modelo de alocação de projetos. Dessa forma, deixa-se também como sugestão de trabalho futuro a construção de um repositório estruturado, dentro da plataforma PUMA, para armazenamento de projetos já realizados nas disciplinas de PSP, possibilitando, assim, que os seus dados sejam utilizados na construção dos próximos modelos de alocação de projetos.

Os dados utilizados na pesquisa limitaram-se aos artigos retirados de anais do congresso ENEGEP, podendo ser utilizados outras bases de dados em trabalhos futuros, por exemplo, as ementas das disciplinas, desde que sejam feitos os devidos ajustes nelas a fim de as tornarem representativas dos PSPs.

Devido à limitação de tempo da pesquisa, não foram implementadas integralmente todas as ações de tratamento estabelecidas na Etapa 3. Deste modo, recomenda-se que,

primeiramente, seja finalizada a implementação das ações consideradas prioritárias, para, após isso, serem implementadas as demais ações agrupadas no Apêndice B.

Após implementação do módulo de alocação de projetos dentro da plataforma PUMA, sugere-se medir a velocidade de resposta do mesmo, visto que as tentativas de modelagem durante a pesquisa foram executadas apenas no ambiente Google Colab. Além disso, recomenda-se que sejam realizados testes funcionais no Módulo de Alocação de projetos em trabalhos futuros, os quais podem contar com alunos, professores e agentes externos para a execução deles.

Como outra limitação da pesquisa, tem-se que os metadados textuais utilizados (título, palavras-chave e resumo) possuíram o mesmo peso no modelo de alocação de projetos. Portanto, recomenda-se que sejam utilizados pesos diferentes para os metadados, visto que eles podem ter níveis de importância diferentes ao classificar as propostas de projetos para posterior alocação em disciplinas de PSP.

O uso de quatro temáticas correspondentes a quatro disciplinas obrigatórias de PSP (PSP2, PSP4, PSP5 e PSP6) também pode ser considerado uma limitação da pesquisa, sendo uma proposta de trabalho futuro a busca por englobar as disciplinas de PSP optativas (PSP3 e PSP8); e as que, embora obrigatórias, possuem projetos cuja classificação em uma área do conhecimento específica tende a ser mais difícil, devido a natureza de seus conteúdos técnicos (PSP1 e PSP7).

Ademais, tendo em vista a constatação do potencial de uso de técnicas de mineração de dados para melhorar processos de aprendizagem, sugere-se a aplicação dessas técnicas, juntamente com aprendizagem de máquina, em outros processos, por exemplo, na criação de times de alunos para resolução de problemas reais dos *stakeholders*; na recomendação de membros de bancas de avaliação de projetos de graduação; e na escolha de supervisores de estágios obrigatórios.

Neste sentido, sugere-se que seja realizada uma pesquisa futura relacionada à área de Sistemas de Informação, a qual envolve Computação e Engenharia de Produção, a fim de servir de insumo para o Desenvolvimento Web da Plataforma PUMA, em especial, para a construção do seu banco de dados. A plataforma PUMA precisa ser construída sob perspectiva da necessidade de análise de dados educacionais para impulsionar a aplicação da metodologia PBL em cursos de graduação.

# Referências

- [1] Markopoulos, Angelos P, Anastasios Fragkou, Petros D Kasidiaris e J Paulo Davim: *Gamification in engineering education and professional training*. International Journal of Mechanical Engineering Education, 43(2):118–131, 2015. 1
- [2] Lima, Rui Manuel, Diana Mesquita, Carla Rocha e Mauro Rabelo: *Defining the Industrial and Engineering Management Professional Profile: a longitudinal study based on job advertisements*. Production, 27, 2017, ISSN 0103-6513. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-65132017000200317&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132017000200317&nrm=iso). 1, 2, 8, 10
- [3] Buhari, A., S. Valloo e H. Hashim: *A streamlined approach to enhance the capacity of undergraduate it students to deliver high quality and demand-driven final year project: A conceptual framework on collaboration between industry and university*. Em *2017 7th World Engineering Education Forum (WEEF)*, páginas 910–914, 2017. 1, 2, 8, 9, 10, 14, 15, 21, 22, 34, 69
- [4] De Los Ríos-Carmenado, I., F. R. López e C. P. García: *Promoting professional project management skills in engineering higher education: Project-based learning (pbl) strategy*. International Journal of Engineering Education, 31(1):184–198, 2015. [www.scopus.com](http://www.scopus.com), Cited By :43. 1
- [5] Monteiro, Simone Borges Simão, Ana Carla Bittencourt Reis, João Mello da Silva e João Carlos Felix Souza: *A Project-based Learning curricular approach in a Production Engineering Program*. Production, 27, 2017, ISSN 0103-6513. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-65132017000200324&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132017000200324&nrm=iso). 1, 2, 8, 9, 10, 63, 64
- [6] Diaz-Mosquera, J. D., P. Sanabria, A. Neyem, D. Parra e J. Navon: *Enriching capstone project-based learning experiences using a crowdsourcing recommender engine*. Em *2017 IEEE/ACM 4th International Workshop on CrowdSourcing in Software Engineering (CSI-SE)*, páginas 25–29, 2017. 1, 9, 10, 28, 29, 34, 36, 37, 41, 42, 57, 69
- [7] Naik, V. e S. Girase: *Project based learning methodology: An effective way of learning software engineering through database design and web technology project*. Journal of Engineering Education Transformations, 34(Special Issue):375–379, 2021. [www.scopus.com](http://www.scopus.com). 1
- [8] Monteiro, S. B. S., M. R. M. Campos, A. C. F. Lima e A. M. Mariano: *Evaluating direct and indirect results of the active methodology in learning: Proposal of*

- an integrative design in 360° via unified platform*. Em *International Symposium on Project Approaches in Engineering Education*, volume 8, páginas 768–775, 2018. [www.scopus.com](http://www.scopus.com), Cited By :3. 1, 3, 62, 64
- [9] Núñez, H. e E. Ramos: *Automatic classification of academic documents using text mining techniques*. Em *2012 XXXVIII Conferencia Latinoamericana En Informatica (CLEI)*, páginas 1–7, Oct 2012. 1, 14, 15, 16, 17, 30, 31, 34, 35, 36, 38, 39, 41, 42, 55, 56, 57, 69, 70, 88
- [10] Pesántez-Avilés, F., D. Calle-López, V. Robles-Bykbaev, M. Rodas-Tobar e C. Vásquez-Vásquez: *A recommender system based on data mining techniques to support the automatic assignment of courses to teachers in higher education*. Em *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, páginas 231–236, 2017. 1, 15, 16, 25, 26, 34, 42, 57, 69
- [11] Datta, S, GS Beriha, B Patnaik e SS Mahapatra: *Use of compromise ranking method for supervisor selection: A multi-criteria decision making (mcdm) approach*. *International Journal of Vocational and Technical Education*, 1(1):007–013, 2009. 1
- [12] Hancock, Mary P e Susan H Russell: *Research experiences for undergraduates (reu) in the directorate for engineering (eng): 2003-2006 participant survey*. SRI International Rep. prepared for The National Science Foundation, Directorate for Engineering, 2008. 1
- [13] Silva, João Mello da Silva e Balthazar, José Carlos: *Projeto político pedagógico do curso de engenharia de produção*, 2010. <https://sig.unb.br/sigaa/verProducao?idProducao=2029779&&key=4b019ab5789b0cde43fe99a209ac6943>, [Online; accessed 13-Fevereiro-2021]. 2, 62, 63, 84, 85, 86
- [14] Distrito Federal Codeplan, Companhia de Planejamento do: *Texto para discussão - aspectos econômicos do distrito federal*, 2018. [http://www.codeplan.df.gov.br/wp-content/uploads/2018/02/TD\\_37-Aspectos-Econ%C3%B4micos-do-Distrito-Federal.pdf](http://www.codeplan.df.gov.br/wp-content/uploads/2018/02/TD_37-Aspectos-Econ%C3%B4micos-do-Distrito-Federal.pdf), [Online; accessed 13-Fevereiro-2021]. 2, 62
- [15] Balthazar, José C. e João Mello da Silva: *A aprendizagem baseada em projeto no curso de engenharia de produção da universidade de Brasília*. Em *Second Ibero-American Symposium on Project Approaches in Engineering Education*, páginas 141–144, 2010. 2
- [16] UnB, Departamento de Engenharia de Produção da: *Regimento interno do departamento de engenharia de produção da faculdade de tecnologia*, 2014. [http://www.epr.unb.br/images/pastaepr/EPR\\_Regimento\\_17Mai2014.pdf](http://www.epr.unb.br/images/pastaepr/EPR_Regimento_17Mai2014.pdf), [Online; accessed 06-Janeiro-2019]. 3, 63
- [17] Júnior, E. S., S. B. Simão Monteiro, A. C. Fernandes Lima, A. M. Mariano e V. B. Neves Da Silva: *Unified platform of active methodology (puma): Development steps of the module divulgation and request projects*. Em *International Symposium on Project Approaches in Engineering Education*, volume 9, páginas 357–365, 2019. [www.scopus.com](http://www.scopus.com), Cited By :2. 3, 4



- [18] Monteiro, S. B. S., K. B. Hafaiedh, J. M. da Silva, A. C. F. Lima, E. S. Júnior, M. H. Torres, M. Boufaied, A. B. Zarb, M. A. B. Rekaya, I. Zouaghi e D. M. Viana: *Peer-assessment for holistic student development (pahsd): Implementing a digital application on a pbl platform*. Em *International Symposium on Project Approaches in Engineering Education*, volume 10, páginas 253–261, 2020. [www.scopus.com](http://www.scopus.com). 3, 103
- [19] Hafaiedh, K. B., S. B. S. Monteiro, J. M. da Silva, E. S. Júnior, M. H. Torres, M. Nizar, N. Abdellaoui, Z. Kouki, A. Ayari, A. M. Mariano e P. C. Reis: *Team building through student's preferences and competences (tbspc): Implementation on a pbl platform*. Em *International Symposium on Project Approaches in Engineering Education*, volume 10, páginas 262–270, 2020. [www.scopus.com](http://www.scopus.com). 3, 103
- [20] Torres, Mateus Halbe: *Espaço Virtual para Evolução de Megatemas e Plataformas aplicado ao New EPIC + SDG Challenge, no contexto do Continuum Realidade-Virtualidade*. Monografia (Engenharia de Produção), UnB (Universidade de Brasília), Brasília, Brasil, 2020. 3
- [21] Monteiro, S. B. S., A. C. F. Lima, A. M. Mariano e E. Silva Júnior: *Unified active methodology platform (puma): A multidisciplinary project*. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, 2020(E28):766–778, 2020. [www.scopus.com](http://www.scopus.com), Cited By :2. 4, 103
- [22] Patacsil, Frederick e Christine Lourrine S. Tablatin: *Exploring the importance of soft and hard skills as perceived by it internship students and industry: A gap analysis*. Journal of Technology and Science Education, 7(3):347–368, 2017, ISSN 2013-6374. 8
- [23] Menekse, Muhsin, Glenda S. Stump, Stephen Krause e Michelene T. H. Chi: *Differentiated overt learning activities for effective instruction in engineering classrooms*. Journal of Engineering Education, 102(3):346–374, 2013. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jee.20021>. 8
- [24] Silva, V. B. da N, S. B. S. Monteiro, R. C. Amoras, A. M. Mariano e A. C. F. Lima: *Best practices for active learning: A literature study using bibliometrics*. Em *International Symposium on Project Approaches in Engineering Education*, volume 9, páginas 602–611, 2019. 8, 9
- [25] Condliffe, Barbara: *Project-based learning: A literature review. working paper*. MDRC, 2017. 9
- [26] Alfaro, Luis, Erick Apaza, Jorge Luna-Urquizo e Claudia Rivera: *Identification of learning styles and automatic assignment of projects in an adaptive e-learning environment using project based learning*. International Journal of Advanced Computer Science and Applications, 10(11), 2019. <http://dx.doi.org/10.14569/IJACSA.2019.0101191>. 9, 16, 20, 34, 69
- [27] Song, J e DE Dow: *Project-based learning for electrical engineering lower-level courses*. Em *Proceedings of the 123rd ASEE Annual Conference and Exposition, New Orleans. Retrieved in*, volume 10, 2016. 9

- [28] Afolabi, Ibukun T, Ayodele A Adebisi, Excel G Chukwurah, Chioma P Igbokwe *et al.*: *Decision support system for final year project management*. Em *The World Congress on Engineering and Computer Science (WCECS 2019)*, 2019. 9, 10, 14, 15, 16, 27, 28, 34, 39, 40, 56, 69, 70
- [29] GOH, Melvrick, GOTTIPATI Swapna e Venky SHANKARARAMAN: *Capstone projects mining system for insights and recommendations*. 2015. 9, 10, 16, 17, 29, 30, 34, 35, 36, 37, 41, 55, 56, 69, 70
- [30] Ahmad, Z, MZ Abu Bakar, KT Lee, MH Uzir, K Abdul Karim, MA Ahmad, BS Ooi, Derek Chan, S Ismail, JK Lim *et al.*: *A first attempt at project based learning (prbl) on final year design project chemical engineering undergraduates student at universiti sains malaysia (usm)*. Em *Proceedings of the 6th International Conference on Process Systems Engineering (PSE ASIA)*, volume 25, página 27, 2013. 9, 10
- [31] Normas Técnicas (ABNT), Associação Brasileira de: *Abnt nbr iso/iec 31010*. Gestão de riscos: Técnicas para o processo de avaliação de riscos, 2012. 10, 13, 14, 51, 68
- [32] Normas Técnicas (ABNT), Associação Brasileira de: *Abnt iso guia 73*. Gestão de riscos - Vocabulário, 2009. 10
- [33] Normas Técnicas (ABNT), Associação Brasileira de: *Abnt nbr iso 31000*. Gestão de riscos — Diretrizes, 2018. 10, 11, 12, 44, 57, 59, 61
- [34] Perera, A. A. S, Abdul Khabir Rahmat, Ali Khatibi e S. M. Ferdous Azam: *Review of literature: Implementation of enterprise risk management into higher education*. 10, 11
- [35] Pavlova, X. L. e S. O. Shaposhnikov: *Risk management for university competitiveness assurance*. Em *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, páginas 1440–1443, Jan 2019. 10, 11
- [36] Governing Boards, Association of, United Educators (AGB e UE): *A wake-up call: Enterprise risk management at colleges and universities today*, 2014. 11
- [37] Priyarsono, D S, A P Widhiani e D L Sari: *Starting the implementation of risk management in a higher education institution: The case of IPB university*. IOP Conference Series: Materials Science and Engineering, 598:012107, sep 2019. <https://doi.org/10.1088/1757-899x/598/1/012107>. 11
- [38] Abaoud, Duaa: *The relationship between knowledge management practices and enterprise risk management in higher education: an action research*. 2019. 11
- [39] Normas Técnicas (ABNT), Associação Brasileira de: *Abnt nbr iso 91001*. Sistemas de Gestão da Qualidade, 2015. 11
- [40] SPSS Inc., NCR Systems Engineering Copenhagen DaimlerChrysler AG: *Crisp-dm 1.0 – step-by-step data mining guide*. SPSS inc e CRISP-DM consortium, 1999. 11, 18, 19, 44, 57, 59, 61

- [41] Andrade, Vinnícius Matheus Madeira de: *Mineração de processos como ferramenta de gestão de riscos*. Tese de Mestrado, Universidade de Brasília, Brasil, 2017. 11, 44
- [42] Straub, Jeremy: *Extending the student qualitative undertaking involvement risk model*. Journal of Aerospace Technology and Management, 6:333–352, 2014. 13, 14, 51, 68
- [43] Vanhanen, Jari e Timo OA Lehtinen: *Software engineering problems encountered by capstone project teams*. International Journal of Engineering Education, 30(6):1461–1475, 2014. 13, 14, 51, 68
- [44] Pham, Julius Cuong, George R Kim, Jeffrey P Natterman, Renee M Cover, Christine A Goeschel, Albert W Wu e Peter J Pronovost: *Recasting the rca: an improved model for performing root cause analyses*. American Journal of Medical Quality, 25(3):186–191, 2010. 13, 14, 52, 54, 68, 80
- [45] Liu, Yang, Chen Yang, Jian Ma, Wei Xu e Zhongsheng Hua: *A social recommendation system for academic collaboration in undergraduate research*. Expert Systems, 36(2):e12365, 2019. <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12365>, e12365 EXSY-Mar-18-080.R2. 14, 15, 16, 26, 27, 34, 38, 41, 42, 69
- [46] I'Anson, Rachel A, Karen A Smith *et al.*: *Undergraduate research projects and dissertations: issues of topic selection, access and data collection amongst tourism management students*. Journal of Hospitality, Leisure, Sport & Tourism Education, 3(1):19–32, 2004. 14, 15
- [47] Sergis, S. e D. G. Sampson: *Learning object recommendations for teachers based on elicited ict competence profiles*. IEEE Transactions on Learning Technologies, 9(1):67–80, 2016. 16, 24, 25, 34, 69
- [48] Belk, Marios, Efi Papatheocharous, Panagiotis Germanakos e George Samaras: *Modeling users on the world wide web based on cognitive factors, navigation behavior and clustering techniques*. Journal of Systems and Software, 86(12):2995 – 3012, 2013, ISSN 0164-1212. <http://www.sciencedirect.com/science/article/pii/S0164121213001040>. 16
- [49] Porter, Joshua: *Watch and learn: How recommendation systems are redefining the web*. [Online]. Retrieved from the Internet:,(Dec. 13, 2006), 5, 2006. 16
- [50] Fayyad, Usama, Gregory Piatetsky-Shapiro e Padhraic Smyth: *From data mining to knowledge discovery in databases*. AI Magazine, 17(3):37, Mar. 1996. <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>. 17
- [51] Witten, Ian H., Eibe Frank e Mark A. Hall: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Amsterdam, 3ª edição, 2011, ISBN 978-0-12-374856-0. <http://www.sciencedirect.com/science/book/9780123748560>. 17, 35, 36, 39, 41, 42, 71

- [52] Ladeira, Marcelo: *Construção e avaliação de modelos com o unbminer, ferramenta aberta para mineração de dados*. Em *Slides do I Congresso de Computação do Sul do Mato Grosso Rondonópolis/MT*, 2005. <https://slideplayer.com.br/slide/3117082/>, acesso em 05 de dezembro de 2020. 17, 39, 41
- [53] Azevedo, Ana e Manuel Filipe Santos: *Kdd, semma and crisp-dm: a parallel overview*. IADS-DM, 2008. 17
- [54] *CRISP-DM by Smart Vision Europe*. <http://crisp-dm.eu/>, acesso em 10 de dezembro de 2020. 17, 18
- [55] Kurniawan, S, W Gata, D A Puspitawati, I K S Parthama, H Setiawan e S Hartini: *Text mining pre-processing using gata framework and RapidMiner for indonesian sentiment analysis*. IOP Conference Series: Materials Science and Engineering, 835:012057, may 2020. <https://doi.org/10.1088/1757-899x/835/1/012057>. 17
- [56] Shafique, Umair e Haseeb Qaiser: *A comparative study of data mining process models (kdd, crisp-dm and semma)*. International Journal of Innovation and Scientific Research, 12(1):217–222, 2014. 17
- [57] Sun, Xiaobing, Wenyuan Xu, Xin Xia, Xiang Chen e Bin Li: *Personalized project recommendation on github*. Science China Information Sciences, 61(5):050106, 2018. 22, 23, 34, 36, 37, 41, 69
- [58] Zhang, P., F. Xiong, H. K. N. Leung e W. Song: *Funkr-pdae: Personalized project recommendation using deep learning*. IEEE Transactions on Emerging Topics in Computing, páginas 1–1, 2018, ISSN 2168-6750. 23, 24, 34, 41, 69
- [59] Canfora, G., M. Di Penta, S. Giannantonio, R. Oliveto e S. Panichella: *Yoda: Young and newcomer developer assistant*. Em *2013 35th International Conference on Software Engineering (ICSE)*, páginas 1331–1334, May 2013. 31, 32, 34, 69
- [60] Bastos, Weliton Moreira: *Metodologia para recomendação de consultores ad-hoc baseada na extração de perfis do currículo lattes*. Tese de Mestrado, Universidade de Brasília, Brasil, 2009. 32, 33, 34, 35, 36, 37, 41, 42, 55, 56, 57, 69, 70, 71
- [61] Khan, Aurangzeb, Baharum Baharudin, Lam Hong Lee e Khairullah Khan: *A review of machine learning algorithms for text-documents classification*. Journal of advances in information technology, 1(1):4–20, 2010. 35, 39
- [62] Aas, Kjersti e Line Eikvil: *Text categorisation: A survey*. Relatório Técnico, Citeseer, 1999. 35, 36, 38, 39, 40
- [63] Porter, Martin F: *An algorithm for suffix stripping*. Program, 1980. 35
- [64] Murugesan, Manoj Kumar: *Comparative analysis of machine learning algorithms using nlp techniques in automatic detection of fake news on social media platforms*. Tese de Mestrado, Dublin, National College of Ireland, December 2019. <http://norma.ncirl.ie/4157/>. 36, 41

- [65] Polyvyanyy, Artem e Dominik Kuropka: *A quantitative evaluation of the enhanced topic-based vector space model*. 2009, ISBN 978-3-939469-95-7. 37
- [66] Liu, Alexander Yun chung: *The effect of oversampling and undersampling on classifying imbalanced text datasets*. Tese de Doutorado, Citeseer, 2004. 37
- [67] Salazar, Frizzi Alejandra San Roman: *Um estudo sobre o papel de medidas de similaridade em visualização de coleções de documentos*, 2012. 37, 38
- [68] Magabo, Fernanda de Camargo: *DESCOBERTA DE CONHECIMENTO EM REDES SOCIAIS E BASES DE DADOS PÚBLICAS*. Monografia, Universidade de São Paulo, São Paulo, Brasil, 2016. 39, 40
- [69] Gil, Antonio Carlos *et al.*: *Como elaborar projetos de pesquisa*, volume 4. Atlas São Paulo, 2002. 43
- [70] Gerhardt, Tatiana Engel e Denise Tolfo Silveira: *Métodos de pesquisa*. Plageder, 2009. 43
- [71] MARCONI, M de A and LAKATOS, Eva Maria: *Técnicas de pesquisa*. São Paulo, Atlas, 2002. 43
- [72] Silva, Edna Lúcia da e Estera Muszkat Menezes: *Metodologia da pesquisa e elaboração de dissertação*. 2005. 43
- [73] Mariano, Ari Melo e Maíra Santos Rocha: *Revisão da literatura: apresentação de uma abordagem integradora*. Em *AEDEM International Conference*, volume 18, 2017. 45, 46
- [74] Vieira, Elizabeth e José Gomes: *A comparison of scopus and web of science for a typical university*. *Scientometrics*, 81(2):587–600, 2009. 46
- [75] Microsoft: *Microsoft teams: Aplicativo de chat interno / trabalho remoto*. Disponível em: <https://www.microsoft.com/pt-br/microsoft-teams/group-chat-software>. 50
- [76] *Ideaboardz - brainstorm, retrospect, collaborate*. Disponível em: <https://ideaboardz.com/>. 50
- [77] Google: *Google sheets: free online spreadsheets for personal use*. Disponível em: <https://www.google.com/sheets/about/>. 52
- [78] Bioinformatics Laboratory, University of Ljubljana: *Orange data mining*. Disponível em: <https://orangedatamining.com/>. 53
- [79] Soares, Victor Hugo Andrade Soares: *Combinações de similaridade semântica e frequência de termos para agrupamento de textos*. 2017. 56, 70
- [80] Han, Hui, Wen Yuan Wang e Bing Huan Mao: *Borderline-smote: a new over-sampling method in imbalanced data sets learning*. Em *International conference on intelligent computing*, páginas 878–887. Springer, 2005. 56, 71

- [81] Google: *Google colabory*. Disponível em: <https://colab.research.google.com/notebooks/welcome.ipynb?hl=pt-BR>. 56, 71
- [82] Alshamrani, Adel e Abdullah Bahattab: *A comparison between three sdlc models waterfall model, spiral model, and incremental/iterative model*. International Journal of Computer Science Issues (IJCSI), 12(1):106, 2015. 58
- [83] *Relatório de gestão - universidade de brasília - 2019*. Disponível em: [http://www.dpo.unb.br/images/dpl/PUB\\_RG2019\\_310820\\_v1\\_aprovadoCAD.pdf](http://www.dpo.unb.br/images/dpl/PUB_RG2019_310820_v1_aprovadoCAD.pdf). 61, 64
- [84] Brasília UnB, Universidade de: *Projeto político-pedagógico institucional da universidade de brasília*, 2018. 62
- [85] Brasil: *Decreto nº 6.096 de 24 de abril de 2007 - institui o programa de apoio a planos de reestruturação e expansão das universidades federais - reuni*. 2007. 62
- [86] *Referências curriculares da engenharia de produção*. Disponível em: <http://www.abepro.org.br/arquivos/websites/1/DocStaBarbara2003.pdf>. 85
- [87] *Anais do encontro nacional de engenharia de produção - enegep*. Disponível em: <http://www.abepro.org.br/publicacoes/>. 87, 88

# Apêndice A

## Tabelas com os valores de priorização das causas raízes dos riscos

Tabela A.1: Priorização das causas raízes dos riscos de utilização inadequada do Módulo.

Causas raízes dos riscos de alocação inadequada de projetos ( <b>Utilização inadequada do Módulo</b> )	Poder 1	Aplicabilidade	Índice	Prioridade
Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos	4,0	5,0	20	1
Complexidade na alimentação dos dados na plataforma PUMA	3,5	4,5	16	1
Falta de planejamento da manutenção do Módulo de Alocação de Projetos	2,8	4,3	12	2
Existência de temas não desejáveis de projetos	3,3	3,5	11	2
Existência de preferências pessoais que não foram consideradas na construção do Módulo de Alocação de Projetos	3,5	3,3	11	2
Falta de estímulo para uso da plataforma PUMA	2,8	4,0	11	2
Informações desatualizadas nas bases de dados internas e externas à plataforma PUMA	2,8	4,0	11	2

Tabela A.2: Priorização das causas raízes dos riscos de inadequada mineração de dados.

Causas raízes dos riscos de alocação inadequada de projetos ( <b>Execução inadequada do processo de mineração de dados</b> )	Poder 1	Aplicabilidade	Índice	Prioridade
Não especificação do nível de detalhamento requerido ao descrever o projeto	4,0	4,5	18	1
Desconhecimento de técnicas adequadas para o desenvolvimento da arquitetura necessária para o pleno funcionamento do Módulo de Alocação de Projetos	4,0	4,5	18	1
A descrição da proposta submetida não refletir a temática do projeto	4	4,5	18	1
Não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos	4,0	4,0	16	1
O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto	3,5	4,3	15	1
Falta de definição da infraestrutura adequada para o funcionamento pleno do Módulo de Alocação de Projetos	3,5	4,3	15	1

Tabela A.3: Priorização das causas raízes dos riscos de alocação inadequada de projetos para as disciplinas de PSP.

Causas raízes dos riscos de alocação inadequada ( <b>Alocação de projetos das disciplinas de PSP</b> )	Poder 1	Aplicabilidade	Índice	Prioridade
Não existência de critérios pré-definidos para escolha de novos projetos	4,5	4,5	20	1
Não desenvolvimento de uma <i>blacklist</i> (temas de projetos que não devem ser alocados em um semestre específico)	3,8	4,0	15	1

Tabela A.4: Priorização das causas raízes dos riscos de alocação de projetos interdisciplinares para as disciplinas de PSP.

Causas raízes dos riscos de alocação inadequada ( <b>Alocação de projetos interdisciplinares das disciplinas de PSP</b> )	Poder 1	Aplicabilidade	Índice	Prioridade
O modelo de alocação de projetos não levar em conta que há projetos interdisciplinares	4,8	4,8	23	1
Não disponibilização de informações acerca da interdisciplinaridade das propostas de projetos	4,5	4,5	20	1
Restrição de alocação de projetos para apenas uma disciplina, sem levar em conta a possibilidade de integração entre times de diferentes disciplinas	3,5	4,0	14	2



# Apêndice B

## Ações de tratamento dos riscos consideradas de médio e longo prazo

Tabela B.1: Priorização das ações de tratamento de curto prazo.

Causas raízes dos riscos de alocação inadequada de projetos	Ações para tratamento das causas raízes dos riscos de alocação inadequada de projetos.	Poder 2	Facilidade	Índice	Horizonte
Não entendimento das questões de negócio que permeiam o problema de alocação inadequada de projetos.	Fazer a descrição do contexto do processo de metodologias ativas no curso de Engenharia de Produção da UnB para que seja contemplado na alocação dos projetos.	5	4	20	Curto prazo (3 meses)
	Investigar as questões de negócio relacionadas à alocação inadequada de projetos, visando integrá-las ao desenvolvimento do Modelo de Alocação de Projetos.	5	4	20	Curto prazo (3 meses)
	Realizar benchmarking quanto a problemas e soluções similares.	5	4	20	Curto prazo (3 meses)
Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos.	Estabelecer metodologia robusta que envolva técnicas mais aderentes ao levantamento de requisitos.	5	3	15	Curto prazo (3 meses)
Complexidade na alimentação dos dados na plataforma PUMA.	Padronizar os campos de alimentação dos dados a fim de unificar inputs de dados e facilitar a interpretação dos mesmos.	5	3	15	Curto prazo (3 meses)
Não especificação do nível de detalhamento requerido ao descrever a proposta de projeto.	Garantir que todas as informações necessárias para posterior alocação estarão contidas na proposta de projeto.	5	3	15	Curto prazo (3 meses)
Desconhecimento de técnicas adequadas de desenvolvimento de arquitetura necessária para o pleno funcionamento do Módulo de Alocação de Projetos.	Desenvolver um plano de arquitetura do Módulo de Alocação de Projetos por meio do suporte de uma equipe / parceria externa composta por recursos humanos qualificados, levando em conta uma infraestrutura AWS e containers para escala.	5	3	15	Curto prazo (3 meses)
	Terceirizar o desenvolvimento da arquitetura do Módulo de Alocação de Projetos.	5	3	15	Curto prazo (3 meses)
O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto.	Inserir no algoritmo de alocação de projetos a verificação de uma possível adequação a alguma disciplina de PSP, mesmo que contemple apenas parte da proposta de projeto, a depender da similaridade.	5	3	15	Curto prazo (3 meses)
Falta de definição da infraestrutura adequada para o funcionamento pleno do Módulo de Alocação de Projetos.	Descrever do passo-a-passo para implementação da infraestrutura adequada para hospedar o Módulo de Alocação de Projetos	5	3	15	Curto prazo (3 meses)
	Garantir a infraestrutura adequada para hospedar o módulo de alocação de projetos (exemplo: parcerias com fornecedores de tecnologia ou uso de infraestrutura sem custo)	5	3	15	Curto prazo (3 meses)
Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos.	Seguir a metodologia estabelecida para construção de um MVP do Módulo de Alocação de Projetos, cujo os requisitos devem ser levantados com base em uma ampla pesquisa sobre requisitos de software e métodos para alocação de projetos.	4	3	12	Curto prazo (3 meses)
Não especificação do nível de detalhamento requerido ao descrever a proposta de projeto.	Fazer com que a descrição da proposta de projeto seja feita por meio de perguntas separadas que contemplem todas as informações necessárias para posterior alocação.	3	4	12	Curto prazo (3 meses)
O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto.	Criar processo / mecanismo para comunicar o proponente sobre a inadequação / rejeição de sua proposta.	3	4	12	Curto prazo (3 meses)

Tabela B.2: Priorização das ações de tratamento de médio prazo.

Causas raízes dos riscos de alocação inadequada de projetos	Ações para tratamento das causas raízes dos riscos de alocação inadequada de projetos.	Poder 2	Facilidade	Índice	Horizonte
Não utilização de técnicas adequadas para o levantamento de requisitos do Módulo de Alocação de Projetos.	Introduzir uma fase de avaliação da qualidade dos requisitos levantados para o Módulo de Alocação de Projetos.	4	3	12	Médio prazo (6 meses)
Complexidade na alimentação dos dados na plataforma PUMA.	Solicitar os dados de maneira objetiva (com instruções claras), de modo a evitar que os atores tenham dúvida quanto à alimentação dos dados.	4	3	12	Médio prazo (6 meses)
	Estabelecer plano de alimentação e manutenção do Módulo de Alocação de Projetos a ser incluído na documentação do PUMA.	4	3	12	Médio prazo (6 meses)
A descrição da proposta submetida não refletir a temática do projeto.	Fazer uma lista de possíveis temas de projeto incluindo a opção "Outros" ou "Não sei Informar".	3	4	12	Médio prazo (6 meses)
O escopo dos projetos desenvolvidos no Curso não englobar a temática da proposta de projeto.	Definir um valor mínimo de similaridade, de modo que, caso uma proposta tenha similaridade com as disciplinas de PSP abaixo desse valor, possa ser realizada a sua rejeição de forma automática.	4	3	12	Médio prazo (6 meses)
Não existência de critérios pré-definidos para escolha de novos projetos.	Pesquisar na literatura de possíveis critérios que podem ser utilizados na escolha de novos projetos.	4	3	12	Médio prazo (6 meses)
	Implementar no algoritmo de alocação os critérios para escolha dos projetos em cada PSP, levando em conta uma possível opção manual de escolha de projetos ou de mudança de critérios de alocação dos mesmos.	4	3	12	Médio prazo (6 meses)
O modelo de alocação de projetos não levar em conta que há projetos interdisciplinares nos PSPs.	Fazer com que o algoritmo de alocação de projetos envie automaticamente as propostas interdisciplinares para mais de uma disciplina ao mesmo tempo, caso identifique fortes indícios (exemplo: similaridades próximas) de categorização para mais área do conhecimento de diferentes PSPs (Exemplo: Gestão da Qualidade e Sistemas de Informação).	4	3	12	Médio prazo (6 meses)
Não especificação do nível de detalhamento requerido ao descrever a proposta de projeto.	Fazer testes de usabilidade no Módulo de Alocação de Projetos a fim de garantir que leitura dos requisitos de detalhamento esteja adequada.	5	2	10	Médio prazo (6 meses)
Não desenvolvimento de uma blacklist (temas de projetos que não devem ser alocados em um semestre específico).	Realizar uma pré-aprovação das temáticas que serão tratadas durante o semestre, as demais temáticas podem ser utilizadas em semestres futuros.	5	2	10	Médio prazo (6 meses)
Não especificação do nível de detalhamento requerido ao descrever a proposta de projeto.	Disponibilizar modelos/exemplos de detalhamento de proposta de projeto.	3	3	9	Médio prazo (6 meses)
	Incluir hints (dicas de tela) explicativos na interface de submissão das propostas de projeto.	3	3	9	Médio prazo (6 meses)
A descrição da proposta submetida não refletir a temática do projeto.	Incluir hints (dicas de tela) explicativos na interface de submissão das propostas de projeto.	3	3	9	Médio prazo (6 meses)
Não existência de critérios pré-definidos para escolha de novos projetos.	Definir critérios de escolha universais (fixos no tempo) e critérios de escolha mutantes (que podem variar de um semestre para outro)	3	3	9	Médio prazo (6 meses)
A descrição da proposta submetida não refletir a temática do projeto.	Propor a associação da proposta de projeto com áreas e subáreas de Engenharia de Produção.	2	4	8	Médio prazo (6 meses)

## Apêndice C

Código desenvolvido no Google  
Colab

---

## Modelo de Alocação de Projetos

### ▼ Back End

### ▼ Importação dos pacotes

```
import pandas as pd
import string
import re
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
nltk.download('rslp')
from nltk.stem import RSLPStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
```

### ▼ Acesso às bases de dados

```
dados_alocacao = pd.read_csv("/content/drive/MyDrive/1 - Dados para alocação de projetos - PUMA/base_ENEGEP (22).csv", sep=";")
```

```
dados_alocacao
```

### ▼ Entendimento dos dados

```
dados_alocacao.info()
```

```
dados_alocacao.describe(include='all')
```

```
dados_alocacao.tematica.value_counts()
```

```
dados_alocacao.ano.value_counts()
```

```
dados_alocacao.isnull()
```

```
dados_alocacao.isnull().sum()
```

```
dados_alocacao.duplicated()
```

```
dados_alocacao[dados_alocacao.duplicated()]
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
sns.set(font_scale=2)
```

```
target_count.plot(kind='bar', figsize=(22, 6), rot=0, color='olive')
plt.xlabel("Temática", labelpad=14)
plt.ylabel("Nº de documentos", labelpad=14)
plt.title("Número de documentos por temática", y=1.02);
```

```
target_count_7 = dados_alocacao.groupby('tematica').get_group('PCP').ano.value_counts()
target_count_8 = dados_alocacao.groupby('tematica').get_group('Sistemas de Informação').ano.value_counts()
```

```
target_count_8 = dados_alocacao.groupby('tematica').get_group('Sistemas de Informação').ano.value_counts()
target_count_9 = dados_alocacao.groupby('tematica').get_group('Engenharia do Produto').ano.value_counts()
target_count_10 = dados_alocacao.groupby('tematica').get_group('Gestão da Qualidade').ano.value_counts()
```

```
df_3 = pd.DataFrame({'Planejamento e Controle da Produção (PCP)': target_count_7, 'Sistemas de Informação': target_count_8, 'Engenharia do Produto': target_count_9, 'Gestão da Qualidade': target_count_10})
```

```
sns.set(font_scale=2)
```

```
df_3.plot(kind='bar', figsize=(18, 6), rot=0, stacked=True)
plt.xlabel("Anos", labelpad=14)
plt.ylabel("Nº de documentos", labelpad=14)
plt.title("Número de documentos por ano", y=1.02);
```

```
!pip install wordcloud
```

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

```
descricoes = dados_alocacao['titulo'] + " " + dados_alocacao['resumo'] + " " + dados_alocacao['palavras-chave']
```

```
descricoes
```

```
todas_descricoes = " ".join(word for word in descricoes)
```

```
todas_descricoes
```

```
numero_de_palavras = len(todas_descricoes.split())
```

```
numero_de_palavras
```

```
from nltk import FreqDist
```

```
todas_palavras = todas_descricoes.split()
```

```
fdist = FreqDist(todas_palavras)
```

```
fdist.most_common(20)
```

```
from collections import Counter
len(Counter(todas_palavras))
len(set(todas_palavras))
```

```
13551
```

```
mywordcloud = WordCloud(background_color= 'black').generate_from_frequencies(fdist)
```

```
plt.figure(figsize = (16,7))
plt.imshow(mywordcloud)
plt.axis('off')
plt.show()
mywordcloud.to_file("nuvemdeplavras.png")
```

## ▼ Preparação dos dados

### ▼ Combinação das colunas que compõem a descrição principal dos documentos

```
dados_alocacao["descricao"] = dados_alocacao['titulo'] + " " + dados_alocacao['resumo'] + " " + dados_alocacao['palavras-chave']
```

```
dados_alocacao
```

### ▼ Remoção / substituição de sinais de pontuação e caracteres especiais

```
pontuacao_caracteresespeciais = string.punctuation
```

```
pontuacao_caracteresespeciais
```

```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
def remove_punctuation(txt):  
    txt_nopunct = "".join([c for c in txt if c not in pontuacao_caracteresespeciais])  
    return txt_nopunct
```

```
dados_alocacao['descricao_sem_pontuacao'] = dados_alocacao['descricao'].apply(lambda x: remove_punctuation(x))
```

```
print(dados_alocacao[['descricao', 'descricao_sem_pontuacao']])
```

### ▼ Conversão para letras minúsculas

```
def lowercase(txt):  
    txt_lower = txt.lower()  
    return txt_lower
```

```
dados_alocacao['descricao_em_minusculo'] = dados_alocacao['descricao_sem_pontuacao'].apply(lambda x: lowercase(x))
```

```
print(dados_alocacao[['descricao_sem_pontuacao', 'descricao_em_minusculo']])
```

### ▼ Remoção de excessos em branco

```
def duplicateblanks(txt):  
    excessblanks = " ".join(txt.split())  
    return excessblanks
```

```
dados_alocacao['descricao_sem_espacosexcedentes'] = dados_alocacao['descricao_em_minusculo'].apply(lambda x: duplicateblanks(x))
```

```
print(dados_alocacao[['descricao_em_minusculo', 'descricao_sem_espacosexcedentes']])
```

### ▼ Tokenization

```
def tokenize(txt):  
    tokens = re.split('\W+', txt)  
    return tokens
```

```
dados_alocacao['descricao_em_tokens'] = dados_alocacao['descricao_sem_espacosexcedentes'].apply(lambda x: tokenize(x))
```

```
print(dados_alocacao[['descricao_sem_espacosexcedentes', 'descricao_em_tokens']])
```

### ▼ Remoção de stopwords

```
stopwords_corpus = stopwords.words('portuguese')
```

```
stopwords_corpus
```

```
def remove_stopwords(txt):  
    txt_clean = [word for word in txt if word not in stopwords_corpus]  
    return txt_clean
```

```
dados_alocacao['descricao_sem_stopwords'] = dados_alocacao['descricao_em_tokens'].apply(lambda x: remove_stopwords(x))

print(dados_alocacao[['descricao_em_tokens', 'descricao_sem_stopwords']])
```

### ▼ Stemming

```
stemmer = RSLPStemmer()

def stemming(txt):
    text_stemming = [stemmer.stem(word) for word in txt]
    return text_stemming

dados_alocacao['descricao_stemmed'] = dados_alocacao['descricao_sem_stopwords'].apply(lambda x: stemming(x))

print(dados_alocacao[['descricao_sem_stopwords', 'descricao_stemmed']])
```

### ▼ Divisão dos dados

```
from sklearn.model_selection import train_test_split

x = dados_alocacao['descricao_stemmed']
y = dados_alocacao['tematica']

x.head(4)

y.head(4)

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.20, train_size = 0.80, random_state = 42)

x_train.head(4)

y_train.head(4)

x_test.head(4)

y_test.head(4)
```

### ▼ Outras tarefas de preparação dos dados

```
#Função que irei utilizar para adicionar outras atividades de limpeza dos dados
def clean(txt_novo):
    text_clean_novo = txt_novo
    return text_clean_novo
```

### ▼ Term Frequency-Inverse Document Frequency (TF-IDF)

```
vectorizer = TfidfVectorizer(analyzer= clean)

x_train_vectors = vectorizer.fit_transform(x_train)

print(x_train_vectors.shape)

df = pd.DataFrame(x_train_vectors.toarray(), columns=vectorizer.get_feature_names())
df.head(20)
```

## ▼ Modelagem

## ▼ Treinamento

## ▼ K-Nearest Neighbors (K-NN)

```
from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier(n_neighbors = 7, metric='euclidean')
```

```
modelo_alocacao_knn = knn.fit(x_train_vectors, y_train)
```

## ▼ Teste

## ▼ Predição com a base de teste

```
x_test_vectors = vectorizer.transform(x_test)
```

```
print(x_test_vectors.shape)
```

```
df = pd.DataFrame(x_test_vectors.toarray(), columns=vectorizer.get_feature_names())  
df.head(20)
```

```
predicao_alocacao_knn = modelo_alocacao_knn.predict(x_test_vectors)
```

```
predicao_alocacao_knn
```

## ▼ Matriz de Confusão

```
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
```

```
confusion_matrix(y_test, predicao_alocacao_knn)
```

```
sns.set(font_scale=2, style='white')
```

```
cm = confusion_matrix(y_test, predicao_alocacao_knn, labels=modelo_alocacao_knn.classes_)  
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=modelo_alocacao_knn.classes_)  
fig, ax = plt.subplots(figsize=(20, 10))  
disp.plot(cmap='Blues', xticks_rotation=75, ax = ax)  
plt.xlabel('Classes preditas', color='black')  
plt.ylabel('Classes reais', color='black')  
plt.title('Matriz de Confusão - K-NN')  
plt.show()
```

## ▼ Accuracy, Precision, Recall e F-measure

```
from sklearn.metrics import classification_report
```

```
from sklearn.metrics import accuracy_score
```

```
print(classification_report(y_test, predicao_alocacao_knn))
```

```
report = classification_report(y_test, predicao_alocacao_knn, output_dict=True)
```



```

report_df = pd.DataFrame(report).transpose()

report_df.to_excel ('report.xls')

print(accuracy_score(y_test, predicacao_alocacao_knn))

import numpy as np

acurracy = []
acurracy_2 = []

for i in range(1, 10):
    knn = KNeighborsClassifier(n_neighbors=i, metric='euclidean')
    knn_2 = KNeighborsClassifier(n_neighbors=i, metric='manhattan')
    knn.fit(x_train_vectors, y_train)
    knn_2.fit(x_train_vectors, y_train)
    pred_i = knn.predict(x_test_vectors)
    pred_i_2 = knn_2.predict(x_test_vectors)
    acurracy.append(accuracy_score(y_test, pred_i))
    acurracy_2.append(accuracy_score(y_test, pred_i_2))

sns.set(font_scale=2)
plt.figure(figsize=(12, 6))
plt.plot(range(1, 10), acurracy, color='green', linestyle='dashed', marker='o',
         markerfacecolor='green', markersize=10, label = 'Distância Euclidiana')
plt.plot(range(1, 10), acurracy_2, color='orange', linestyle='dashed', marker='o',
         markerfacecolor='orange', markersize=10, label = 'Distância de Manhattan')
plt.legend()
plt.title('Acurácia por valores de K e métricas de distância')
plt.xlabel('Valores de K')
plt.ylabel('Acurácia')
plt.show()

```

## ▼ Front End

### Submissão de Proposta de Projeto

**Título:** "Aplicação da ISO 9001:2015 no contexto de empresa do terceiro setor"

**Descrição:** "O objetivo deste projeto é a avaliação do Sistema de Gestão da Qualidade (SGQ) da Softex com o foco na ISO 9001:2015 e a"

[Mostrar código](#)

☞ A área do conhecimento da proposta é ['Gestão da Qualidade']