

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 1/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

Projeto da Disciplina Ciência de Dados

Fase de Entendimento do Negócio

Descrição de Objetivos e Metas

Carlos Eduardo Nascimento Cajado

Análise e Predição de Alunos com Risco de

Baixo Desempenho Acadêmico

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 2/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

Histórico de Revisões do Documento

Revisão	Descrição	Modificado por	Status	Data
1.0	Projeto Inicial	Carlos Eduardo Cajado	Em desenvolvimento	17/09/2024
2.0	Ajustes de versão	Carlos Eduardo Cajado	Em desenvolvimento	02/12/2024
3.0	Ajustes de versão 2	Carlos Eduardo Cajado	Aprovado	11/12/2024

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 3/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

1 Área de Negócio

1.1 Contexto

A emergência da Indústria 4.0 traz consigo novos paradigmas que transformam a forma como as empresas operam, especialmente com a integração das ciências de dados e da inteligência artificial. Essa transformação não se limita ao setor industrial; ela se estende também às instituições acadêmicas, onde a adoção dessas tecnologias se torna essencial para a otimização dos processos educacionais.

Diante desse contexto, torna-se imprescindível o desenvolvimento de uma ferramenta que auxilie os gestores de instituições de ensino fundamental e médio na identificação precoce de alunos com risco de desempenho acadêmico inferior ou abaixo do esperado. Um sistema que forneça relatórios detalhados e antecipados pode ser um recurso valioso, permitindo que as instituições tomem decisões informadas e direcionadas para melhorar o desempenho dos estudantes. Essa abordagem não apenas visa reduzir taxas de reprovação e evasão escolar, mas também tem um impacto no futuro dos alunos. Ao identificar dificuldades e problemáticas que refletem no aprendizado, precocemente, as instituições podem implementar intervenções personalizadas, proporcionando o suporte necessário para que cada estudante alcance seu potencial.

Atualmente, o projeto visa ser integrado como um módulo adicional a uma aplicação web de um sistema acadêmico que já atende mais de 100 escolas em quatro estados do Brasil: Maranhão, Piauí, Bahia e Ceará. Neste contexto, não há, até o momento, uma solução disponível para a problemática em questão.

A implementação dessa ferramenta não apenas representa um avanço no projeto de inovação, mas também proporciona uma maneira eficaz de monitorar e melhorar o desempenho dos alunos. Além disso, ao integrar essa solução ao sistema já existente, espera-se facilitar a adoção e maximizar a eficácia do uso de dados na tomada de decisões educacionais, contribuindo, assim, para a melhoria contínua da qualidade do ensino nas escolas atendidas.

1.2 Objetivo do Negócio

Os Principais objetivos é criar relatórios anuais que identifiquem tendências de futuros desempenho acadêmico ruins. A eficácia desses relatórios será medida pela taxa de alunos identificados como em risco, visando identificar alunos não somente os alunos com ótimos desempenho acadêmico para oferecer tipos de descontos diferentes, mas também os alunos com risco de reprovações para acompanhamento especial da equipe acadêmica responsável.

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 4/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

1.3 Critérios de Sucesso

Para garantir o sucesso comercial do projeto, é fundamental estabelecer critérios claros e mensuráveis. A aceitação da ferramenta também será avaliada, buscando uma taxa de adoção de 75% entre os gestores das escolas que utilizam o sistema, assegurando que a solução esteja integrada às práticas de início letivo das instituições.

Além disso, o feedback dos clientes será crucial para o aprimoramento contínuo da ferramenta. O objetivo é alcançar uma avaliação média de satisfação superior a 4,0 em uma escala de 1 a 5, mediante questionários de satisfação aplicados após a implementação. Os dados gerados pela ferramenta devem ser acessíveis e úteis, permitindo que os gestores realizem autoatendimentos na análise de desempenho. Os resultados serão avaliados por gestores acadêmicos, diretores de escolas e a equipe de desenvolvimento de produtos.

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 5/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

2 Situação Atual

Este capítulo apresenta uma visão geral da situação atual, abordando os recursos disponíveis (Seção 2.1), os requisitos, suposições e restrições do projeto (Seção 2.2), os principais riscos e contingências envolvidos (Seção 2.3), a terminologia utilizada (Seção 2.4), e uma análise dos custos e benefícios associados à sua execução (Seção 2.5).

2.1 Recursos Disponíveis

Pessoal

Nome	Organização	Papel/Função	Contato
Carlos Eduardo Cajado	Primário	Desenvolvedor	(99) 9 8408-5181

Bases

Localização	Formato	Descrição	Contato
Servidor Banco de dados acadêmico da instituição “A” (SQL Server)	<u>CSV</u>	Dados como: Ocorrências acadêmicas, Histórico acadêmico, faltas e desempenhos em anos anteriores.	(99) 9 8408-5181

Software

- *Python*
- *Jupyter Notebook*
- *Visual Studio Code*
- *Dbeaver*

2.2 Requisitos, Suposições e Restrições

Os requisitos do projeto são fundamentais para garantir sua eficácia e sucesso. Portanto, tem-se:

1. A base de dados deve ser carregada uma única vez, utilizando o período de 2019 a 2023, incluindo as características: como número de faltas anuais e ocorrências acadêmicas, sexo, idade e todo histórico de períodos letivos anteriores.

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 6/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

2. Transformações necessária como: normalização, limpeza e remoção de registros, visam lidar com dados ausentes ou ruins, como, por exemplo, alunos novatos ou com históricos incompletos ou inconsistentes.
3. As instâncias, com base nas características extraídas dos dados, serão classificadas em cinco categorias numeradas, a fim de facilitar a análise do desempenho acadêmico dos alunos. Tem-se a seguinte classificação:
 - **(10) Aprovado Acima da Média 2:** Alunos que alcançaram notas muito superiores à média esperada, demonstrando um desempenho excepcional.
 - **(8) Aprovado Acima da Média 1:** Alunos que alcançaram notas significativamente superior à média esperada, demonstrando um desempenho excepcional.
 - **(7) Aprovado:** Alunos que cumpriram os requisitos necessários para a aprovação, alcançando notas dentro da média.
 - **(5) Aprovado com Ressalvas:** Alunos que foram aprovados, mas apresentam algumas áreas de preocupação, indicando a necessidade de acompanhamento adicional.
 - **(1) Reprovado:** Alunos que não atingiram os critérios de aprovação e apresentam um desempenho significativamente abaixo da média, necessitando de suporte adicional.
4. A base de dados será dividida em três partes: 75% para treinamento, 25% para teste.
5. Utilizando aprendizado de máquina supervisionado, será desenvolvido um modelo de redes neurais e outro modelo utilizando árvore de decisão com o objetivo de resolver o problema apresentado.
6. Medidas de segurança robustas devem ser implementadas para proteger os dados dos alunos, garantindo que o uso das informações esteja em conformidade com a LGPD (Lei Geral de Proteção de Dados). Como o mascaramento e a exclusão de qualquer dado sensível.
7. Os relatórios gerados devem ser objetivos para os gestores acadêmicos, utilizando uma linguagem acessível e visualizações intuitivas.
8. O cronograma de conclusão está definido para um prazo de até 3 meses com marcos semanais para avaliar o progresso e realizar ajustes conforme necessário.

Ademais, às suposições feitas durante a mineração de dados incluem a continuidade das escolas participantes em fornecer dados e feedback ao longo do projeto, bem como a disponibilidade da equipe de TI para resolver problemas técnicos e oferecer suporte durante a implementação. Por fim, as restrições do projeto devem ser cuidadosamente consideradas. A disponibilidade de recursos, tanto humanos quanto tecnológicos, pode limitar o progresso. A integração com o sistema existente requer tempo e uma mineração de dados refinada. É crucial que todas as questões legais sobre o uso dos dados sejam tratadas, garantindo conformidade com a legislação vigente. É importante ressaltar,

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 7/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

O orçamento do projeto cobre todos os custos associados, incluindo desenvolvimento, implementação e treinamento para as instituições participantes, conforme os termos estabelecidos para o primeiro ano. Um valor adicional será cobrado para o uso do novo módulo em períodos futuros.

2.3 Riscos e Contingência

- **Atrasos na Coleta e Preparação dos Dados:** Elaborar um cronograma alternativo e explorar fontes adicionais, ou, se necessário, reduzir o escopo das características a serem analisadas.
- **Resistência à Adoção do Plano:** Conduzir apresentações e reuniões informativas para destacar os benefícios do projeto e engajar os envolvidos.
- **Falta de Recursos Humanos:** Avaliar a possibilidade de adicionar mais membros à equipe do projeto para garantir a cobertura adequada das atividades.

2.4 Terminologia

Glossário de Terminologia Empresarial

- **Projeto de Inovação no Sistema:** Desenvolvimento de uma ferramenta que integra ciência de dados e inteligência artificial para otimizar a gestão acadêmica e a análise de desempenho dos alunos.
- **Gestores Acadêmicos:** Profissionais responsáveis pela administração e supervisão de instituições de ensino, focando em estratégias para melhorar a performance dos alunos.
- **Tipos de Reprovações e Impacto:** Classificação das reprovações em diferentes categorias, cada uma com suas implicações sobre o aprendizado do aluno e a necessidade de intervenções específicas.
- **Evasão Escolar:** O abandono prematuro dos estudos pelos alunos, que compromete a continuidade educacional e gera impactos negativos nas instituições, como perda de receita e reputação.

Glossário de Terminologia de Mineração de Dados

- **Aprendizado de Máquina:** Ramo da inteligência artificial que permite que sistemas aprendam a partir de dados, sem serem explicitamente programados.
- **Rede Neural:** Um modelo computacional que simula o funcionamento do cérebro humano, projetado para identificar padrões complexos e realizar previsões com base em características extraídas de instâncias de dados.
- **Classificação:** O processo de categorizar dados em grupos distintos com base em características específicas. Exemplo: classificar alunos como "Aprovado", "Reprovado", etc.
- **Validação de Dados:** O processo de garantir que os dados sejam precisos e consistentes,

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 8/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

essencial para a eficácia de análises e modelos preditivos

2.5 Custos e Benefícios

Análise de Custo-Benefício

- **Custos do Projeto:** O projeto, que se insere no setor de inovação, será executado com custos de criação e implementação cobertos pela própria equipe de funcionários do setor, garantindo eficiência e aproveitamento dos recursos internos.

Benefícios Potenciais:

- **Redução da Evasão:** Espera-se uma diminuição nas taxas de evasão escolar, contribuindo para a continuidade dos estudos dos alunos.
- **Redução da Taxa de Reprovações:** Espera-se uma diminuição nas taxas de reprovação ao final do período letivo.
- **Melhoria no Desempenho Acadêmico:** O aumento nas taxas de aprovação poderá facilitar a formação de parcerias estratégicas, além de aprimorar a reputação da instituição no mercado educacional e fortalecer a imagem da empresa fornecedora do serviço.
- **Satisfação do Cliente:** O incremento na satisfação dos clientes poderá levar a uma maior retenção de instituições parceiras e atração de novas, resultando em um impacto positivo na receita da organização.

Cenários de Risco

- **Cronograma:** Se o projeto ultrapassar os três meses planejados, custos adicionais podem surgir, como horas extras de trabalho. Um cronograma alternativo deve ser elaborado para minimizar atrasos.
- **Dados:** Se os dados forem de baixa qualidade, será necessário implementar um processo de limpeza e validação, aumentando a demanda de trabalho e prolongando o cronograma. A avaliação inicial da qualidade dos dados deve ser uma prioridade.
- **Resultados:** Se os resultados iniciais não forem tão positivos quanto esperado, será fundamental realizar uma análise detalhada e ajustar as intervenções para garantir a eficácia do projeto.

	Documentação do Projeto	Versão do Modelo: 1.1	
		Emissão: 11/12/2024	Página: 9/9
	Projeto: Análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico		

3 **Objetivos do Projeto**

O presente projeto visa, principalmente, a aplicação de técnicas de aprendizado de máquina para -se não apenas compreender os fatores que influenciam o desempenho, mas também oferecer subsídios para intervenções eficazes.

3.1 **Objetivos Técnicos**

O objetivo técnico deste relatório foi desenvolver um modelo computacional para análise e Predição de Alunos com Risco de Baixo Desempenho Acadêmico. Utiliza-se duas técnicas de aprendizado de máquina: redes neurais e árvores de decisão, visando identificar o modelo mais eficiente para o problema apresentado.

3.2 **Critérios de Sucesso**

Para avaliar o sucesso do projeto, serão considerados os seguintes critérios:

- Extrair, preparar e manipular os dados para uma análise preliminar, identificando informações como correlações e associações relevantes.
- Desenvolver um modelo eficiente para o problema.
- Demonstrar que o modelo mantém bom desempenho em um conjunto de dados de teste separado, com métricas semelhantes ao conjunto de treinamento.
- Validar o modelo em pelo menos três cenários reais diferentes e obter resultados consistentes.