

**Desempenho Acadêmico dos Alunos**

DOCUMENTAÇÃO DO PROJETO

**Brasília**

09/2023

EQUIPE TÉCNICA

**Elaboração, distribuição e informações:**

**Ministério Da Educação**

Secretaria Executiva

Subsecretaria de Tecnologia da Informação e Comunicação - STIC

Coordenação-Geral de Governança de Dados - CGGD

Esplanada dos Ministérios Edifício Sede e Anexos, Bloco L

Brasília - DF, 70047-900

**Coordenação e organização:**

Jane Adriana de Souza

**Revisão técnica:**

Francisco Aislan da Silva Freitas

**Colaboração:**

Geovana Correa Garcia

SUMÁRIO

[EQUIPE TÉCNICA 2](#_Toc145923155)

[VERSIONAMENTO E HISTÓRICO DE REVISÕES 3](#_Toc145923156)

[SUMÁRIO 4](#_Toc145923157)

[Introdução 6](#_Toc145923158)

[1. DADOS DO PROJETO 7](#_Toc145923159)

[1.1. LEVANTAMENTO INICIAL 8](#_Toc145923160)

[1.2. NÍVEL DE ACESSO 8](#_Toc145923161)

[1.3. OBJETIVOS DE NEGÓCIO 8](#_Toc145923162)

[2. SOBRE O PROJETO 8](#_Toc145923163)

[2.1. HISTÓRICO DO DOCUMENTO 8](#_Toc145923164)

[2.2. DADOS DO REQUISITANTE 9](#_Toc145923165)

[9](#_Toc145923166)

[2.3. EQUIPE TÉCNICA 9](#_Toc145923167)

[3. METODOLOGIA 10](#_Toc145923168)

[4. FONTE DE DADOS 10](#_Toc145923169)

[5. ENTENDIMENTO DE NEGÓCIO 10](#_Toc145923170)

[5.1. MARCOS NEGOCIAIS 10](#_Toc145923171)

[5.2. AVALIAR A SITUAÇÃO 10](#_Toc145923172)

[5.2.1. RECURSOS 11](#_Toc145923173)

[5.2.2. REQUISITOS DO PROJETO 11](#_Toc145923174)

[5.2.3. RISCOS 11](#_Toc145923175)

[5.2.4. ANÁLISE CUSTO-BENEFÍCIO 12](#_Toc145923176)

[5.3. MINERAÇÃO DE DADOS 12](#_Toc145923177)

[6. RESULTADOS 12](#_Toc145923178)

[6.1. ANÁLISE EXPLORATÓRIA 12](#_Toc145923179)

[6.2. PREPARAÇÃO DOS DADOS 12](#_Toc145923180)

[6.3. MODELAGEM 12](#_Toc145923181)

[6.4. AVALIAÇÃO DA MODELAGEM 12](#_Toc145923182)

[6.5. IMPLEMENTAÇÃO 12](#_Toc145923183)

[7. STATUS DO PROJETO 12](#_Toc145923184)

[CONCLUSÃO 13](#_Toc145923185)

[Referências (Lembre-se de deixar em ordem alfabética) 14](#_Toc145923186)

[Glossário (Lembre-se de deixar em ordem alfabética) 15](#_Toc145923187)

­­

Introdução

Nesse estudo iremos realizar a análise exploratória do conjunto de dados de Desempenho Acadêmico dos alunos. Este é um conjunto de dados educacionais que é coletado do sistema de gerenciamento de aprendizagem (LMS) chamado Kalboard 360 que é um LMS multiagente, que foi projetado para facilitar o aprendizado através do uso de tecnologia de ponta. Esse sistema fornece aos usuários um acesso síncrono a recursos educacionais a partir de qualquer dispositivo com conexão à Internet.

1. DADOS DO PROJETO

Foi alinhado com a equipe técnica que serão extraídas informações do conjunto de dados que permitem monitorar o progresso de aprendizagem e ação de cada aluno segundo essas características:

* Composto por 480 registros de alunos e 16 características
* Consiste em 305 homens e 175 mulheres
* Formação acadêmica, como estágio educacional, série e seção, características comportamentais como mão levantada na aula, abertura de recursos, resposta a pesquisa dos pais e satisfação escola
* Coleta por meio de 2 semestres letivos
* Alunos de diferentes origens  
  1. LEVANTAMENTO INICIAL

|  |  |
| --- | --- |
| Tipo do projeto | |
| x | Análise exploratória |
|  | Modelo preditivo |
|  | Modelo classificação |
|  | Modelo agrupamento |
|  | Detecção de anomalias |
|  | Sistema de recomendação |

* 1. NÍVEL DE ACESSO

|  |  |
| --- | --- |
| Confidencialidade | |
|  | Público |
| x | Privado |
|  | Apenas setor de interesse |

* 1. OBJETIVOS DE NEGÓCIO

Usaremos os recursos fornecidos pela base de dados para compreender a distribuição e relações entre as características dos alunos e seu desempenho acadêmico, identificando insights e padrões que possam ser úteis para a tomada de decisões educacionais e estratégias.

1. SOBRE O PROJETO
   1. HISTÓRICO DO DOCUMENTO

|  |  |  |  |
| --- | --- | --- | --- |
| Versão | Data | Responsável/Autor | Alterações |
| 1.0 | 20/09/2023 | Geovana Magalhães | Estrutura inicial do documento |
| 1.1 | 21/09/2023 | Geovana Magalhães | Inserção dos marcos negociais |
|  |  |  |  |

* 1. DADOS DO REQUISITANTE

Tela de vídeo game

Descrição gerada automaticamente

|  |  |  |
| --- | --- | --- |
| Nome | Cargo/Função | Email |
| Jane Adriana de Souza | Especialista de negócio | joaojunior@mec.gov.br |

* 1. EQUIPE TÉCNICA

|  |  |  |
| --- | --- | --- |
| Nome | Cargo/Função | Email |
| Geovana Correa Garcia Magalhães | Estagiário(a) | geovanamagalhaes@mec.gov.br |

1. METODOLOGIA

O Projeto será desenvolvido segundo a metodologia de ciência de dados da CGDA- Coordenação Geral de Data Analytics.

## **3.1 CRISP-DM**

Cross-industry standard process for data mining (Processo padrão inter-industriais para mineração de dados, ou **Crisp-DM**) é uma metodologia, desenvolvida em 1999 para padronizar processos industriais de mineração de dados, que fornece uma abordagem clara e bem estruturada para processos que envolvem projetos de ciência de dados em geral (mineração, análise, aprendizado de máquina, etc.).

Mesmo que não mencionada explicitamente, a maioria das equipes de ciência de dados utilizam alguma metodologia similar a CRISP-DM ou combinação dela com outras metodologias de desenvolvimento de software.

### **3.1.1 FASES DA CRISP-DM**

A CRISP -DM é composta por um conjunto de passos que devem ser seguidos no ciclo de vida de um projeto de ciência de dados. Tais passos (ou etapas) ajudam a planejar, organizar e implementar o projeto de ciência de dados. Além disso, cada um desses passos ajuda a encontrar respostas das seguintes questões: - Entendimento do negócio:

o que o negócio precisa? - Entendimento dos dados: que dados temos? quais dados precisamos? os dados estão "limpos"? - Preparação dos dados: como organizamos os dados para a etapa de modelagem? - Modelagem: quais técnicas de modelagem podemos aplicar ao nosso problema? E quais devemos utilizar? - Avaliação: qual modelo atende melhor aos objetivos de negócios? - Implantação: como as partes interessadas acessem os resultados do projeto?

Uma visão geral da CRISP-DM é apresentada na Figura abaixo.

Diagrama

Descrição gerada automaticamente

1. FONTE DE DADOS

Kaggle: [Student Academic Performance | Kaggle](https://www.kaggle.com/competitions/student-academic-performance/overview)

1. ENTENDIMENTO DE NEGÓCIO

A etapa de entendimento de negócio se destina a compreensão dos objetivos do projeto:

* Identificação de Tendências de Desempenho: Compreender como os alunos se saem em diferentes disciplinas, tópicos ou áreas de estudo.
* Coletar e analisar dados sobre a satisfação dos alunos e de seus pais com relação à qualidade da educação e dos serviços oferecidos.
* Fatores de Influência: identificar fatores que podem influenciar o desempenho dos alunos, como frequência às aulas, uso de recursos de ensino, participação em atividades.
  1. MARCOS NEGOCIAIS
* Criar repositório no git
* Extrair conjunto de dados do site
* Criar documentação
* Utilizar ferramenta google colab
* Gerar conjunto de dados
* Realizar a descrição dos dados
* Realizar análise exploratória
* Produzir relatório de entendimento dos dados
  1. AVALIAR A SITUAÇÃO

Concentra em avaliar quais as etapas do projeto já foram concluídas e qual será o passo a seguir.

* Revisar o projeto: Todo o projeto foi concluído? Existem erros ou falhas? Todos as etapas foram preenchidas?
* Determinar próxima etapa: Determinar qual será o próximo passo e quais recursos serão utilizados. Se necessário um novo projeto ou se é possível prosseguir.
  + 1. REQUISITOS DO PROJETO

Escopo: Definição do escopo do projeto incluindo os dados que serão coletados e os procedimentos a seguir;

Armazenamento dos dados: Definir onde e como os dados serão armazenados;

* + 1. RISCOS

Os projetos de ciência de dados apresentam alguns riscos que devem ser considerados e gerenciados para garantir o sucesso do projeto. Alguns dos riscos mais comuns incluem:

* Qualidade dos dados: Dados faltantes ou enviesados os quais podem comprometer parte da análise.
* Segurança: Privacidade e segurança dos dados, como manuseio inadequado de dados pessoais dos alunos causando violação legal de privacidade;
* Tecnologia: Problemas tecnológicos no hardware ou no software, sendo capaz de interromper a análise.
  + 1. ANÁLISE CUSTO-BENEFÍCIO
* Projetos e melhorias na educação;
* Aprimorar o desempenho acadêmico e aumentar a satisfação dos alunos;
* Identificar sinais e auxiliar na redução de abandono escolar;
* Melhorias no desempenho acadêmico onde alunos estão com dificuldades.
  1. MINERAÇÃO DE DADOS

Diagrama

Descrição gerada automaticamente

Os dados serão extraídos do site Kaggle, onde será realizado o download no arquivo e armazenado no repositório criado no Gitlab para manuseio e armazenamento.

1. RESULTADOS
   1. Gráfico, Gráfico de barras

      Descrição gerada automaticamenteUma imagem contendo Gráfico

      Descrição gerada automaticamenteANÁLISE EXPLORATÓRIA

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Gráfico, Gráfico de barras, Histograma

Descrição gerada automaticamenteGráfico, Gráfico de barras

Descrição gerada automaticamenteGráfico

Descrição gerada automaticamente

Gráfico, Gráfico de barras

Descrição gerada automaticamenteTela de computador com texto preto sobre fundo branco

Descrição gerada automaticamenteGráfico, Gráfico de barras

Descrição gerada automaticamenteGráfico, Gráfico de barras

Descrição gerada automaticamente

1. STATUS DO PROJETO

|  |  |
| --- | --- |
| Monitoramento dos itens | |
| x | Criar repositório no git |
| x | Extrair conjunto de dados do site |
| x | Criar documentação |
| x | Utilizar ferramenta google colab |
| x | Realizar a descrição dos dados |
| x | Realizar análise exploratória |
| x | Gerar conjunto de dados |
| x | Produzir relatório de entendimento dos dados |

CONCLUSÃO

Será interessante a Implementação de sistemas de identificação precoce para alunos com baixo desempenho, incluindo testes diagnósticos, monitoramento de notas e avaliações regulares, a fim de proporcionar apoio personalizado e prevenir o agravamento dos problemas acadêmicos. Também Investigar as razões por trás das críticas e elogios à matéria de TI, com base na satisfação dos pais. Realizar avaliações de currículo, treinamento de professores ou ajustes no conteúdo do curso para melhorar o desempenho dos alunos nessa matéria.

Sendo assim Implementar programas de apoio para alunos durante a transição do primeiro para o segundo semestre, a fim de garantir uma melhoria contínua no desempenho.

