

2.1 - pred

17 de julho de 2024



historico_estratificado.csv
Planilha

Analise os dados e limpe as colunas da seguinte forma:
período_ingresso: de 2006.1 até 2019.2
tipo_matricula: remover "Dispensa"
situacao: remover "Em Curso"
o novo dataframe após a limpeza deve se chamar df_filtrado



Para prosseguir com a limpeza dos dados conforme suas instruções, vou começar carregando e examinando os dados do arquivo "historico_estratificado.csv" para entender sua estrutura. Em seguida, realizarei as filtrações e limpezas necessárias. Vamos começar lendo o arquivo e visualizando as primeiras linhas para verificar os nomes das colunas e o formato dos dados.

O conjunto de dados contém várias colunas. Aqui está um resumo de algumas das que são relevantes para a limpeza que você solicitou:

- **período_ingresso:** período em que o aluno ingressou.
- **tipo_matricula:** tipo de matrícula do aluno.
- **situacao:** situação atual do aluno.

Vou proceder com a limpeza dos dados de acordo com as suas instruções:

1. Filtrar ``período_ingresso`` para incluir apenas de 2006.1 até 2019.2.
2. Remover as linhas onde ``tipo_matricula`` é "Dispensa".
3. Remover as linhas onde ``situacao`` é "Em Curso".

Agora, vamos aplicar esses filtros. [↗]

Os dados foram filtrados conforme as especificações:

1. **Período de Ingresso:** Apenas registros de 2006.1 até 2019.2 foram mantidos.
2. **Tipo de Matrícula:** Registros com "Dispensa" foram removidos.
3. **Situação:** Registros com "Em Curso" foram removidos.

Se você precisar de mais análises ou ajustes, fique à vontade para pedir! [↗]

Considere que a média geral do aluno é definida como a média de todas as notas não nulas nas disciplinas que o aluno cursou. Qual é a probabilidade de um aluno com a média geral abaixo de 7.0 e com mais de 3 reprovações, ser aprovado na próxima disciplina? Defina 3 opções de como solucionar essa questão e siga a melhor



Para estimar a probabilidade de um aluno com média geral abaixo de 7.0 e com mais de 3 reprovações ser aprovado na próxima disciplina, podemos abordar o problema de várias maneiras. Vou detalhar três opções possíveis:

Opção 1: Análise Estatística Simples

Podemos calcular diretamente a proporção de alunos com essas características (média abaixo de 7.0 e mais de 3 reprovações) que foram aprovados em suas próximas disciplinas. Isso envolveria:

- Calcular a média de cada aluno.
- Contar o número de reprovações de cada aluno.
- Filtrar para alunos que atendem aos critérios e verificar quantos foram aprovados na próxima disciplina que cursaram.

Opção 2: Modelo de Regressão Logística

Usar um modelo de regressão logística para estimar a probabilidade de aprovação, com a média geral e o número de reprovações como variáveis explicativas. Isso permitiria considerar a relação

entre as características do aluno e a probabilidade de aprovação de uma maneira que ajusta outros fatores potencialmente relevantes, como o tipo de disciplina ou o setor acadêmico.

Opção 3: Machine Learning

Desenvolver um modelo de machine learning mais complexo, como uma árvore de decisão ou um modelo de floresta aleatória, que pode capturar interações não-lineares e complexas entre diversas características dos alunos e seu desempenho em disciplinas. Isso requer uma maior quantidade de dados e mais processamento, mas pode proporcionar insights mais precisos e robustos.

Escolha da Melhor Opção

Para uma primeira análise com relativa simplicidade e rapidez, mas ainda informativa, a **Opção 1** (Análise Estatística Simples) parece ser a mais adequada. Ela oferece uma visão direta e específica do problema sem a necessidade de modelagem complexa ou coleta de dados adicionais.

Vamos prosseguir com esta opção: calcular diretamente a probabilidade com base nos dados existentes. Preciso primeiro identificar e agrupar os dados necessários, como médias de notas e contagens de reprovações. Vamos calcular isso agora.

Parece que não encontramos dados para a próxima disciplina cursada após a última registrada para os alunos que se enquadram nos critérios especificados (média abaixo de 7.0 e mais de 3 reprovações). Isso pode ser devido à ausência de registros subsequentes no dataset para esses alunos específicos.

Podemos tentar uma abordagem um pouco diferente para verificar a aprovação desses alunos, analisando diretamente a última situação registrada no período de dados que temos, ao invés de procurar registros futuros. Vou calcular a taxa de aprovação para esses alunos críticos usando os dados disponíveis até agora. [↗]

Com base nos últimos registros disponíveis para alunos que têm média geral abaixo de 7.0 e mais de 3 reprovações, a probabilidade de serem aprovados na próxima disciplina que cursaram é de aproximadamente 23.47%.

Essa estimativa reflete a situação diretamente nos dados atuais e não considera registros futuros que não estão disponíveis. Se precisar de mais análises ou outras abordagens, estou à disposição para ajudar! [↗]

Comece com Data Analyst

[Denunciar conteúdo](#) | [Termos de uso](#) | [Política de privacidade](#) |

?