



Data Warriors

Student Academic Performance

Introdução

- Indicadores de desempenho escolar podem ser ferramentas importantes para a construção de um reforço para a melhoria tanto de alunos quanto de professores.
- A quantidade de faltas e quantidade de vezes que os alunos interagem na aula podem contribuir para o desempenho dos alunos.

Introdução

- Possuir mais de 7 faltas é um fator predominante entre os alunos com desempenho ruim;
- Alunos que pesquisam mais sobre os assuntos em estudo tem um melhor desempenho e as mulheres costumam pesquisar mais;
- Participar da aula fazendo perguntas e de grupos de estudo ajudam a melhorar o desempenho do aluno.

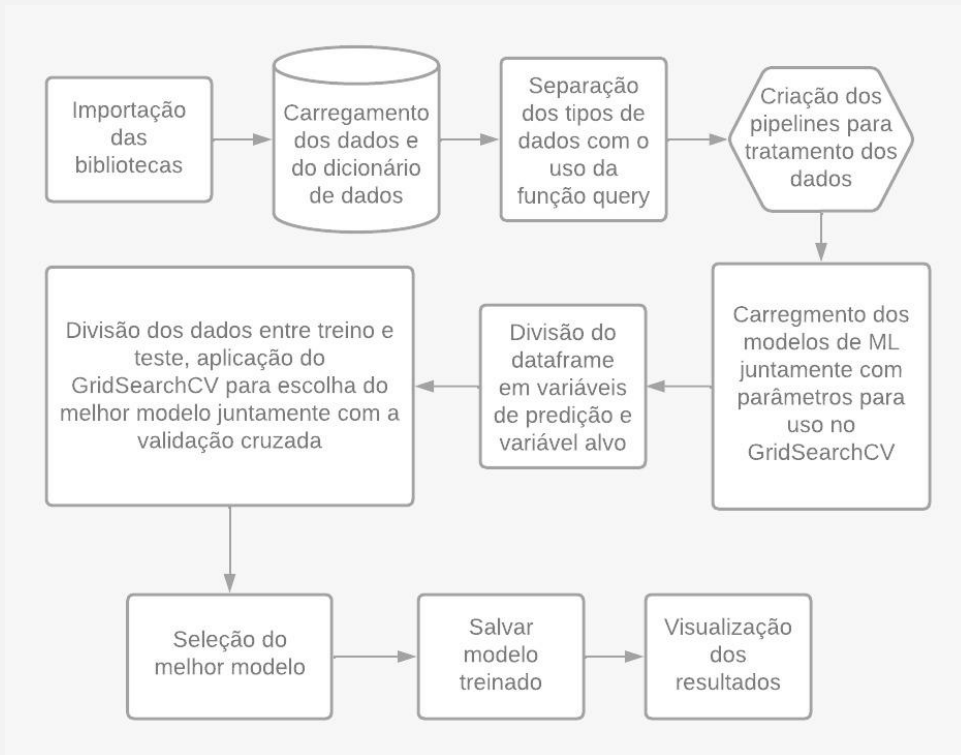
Introdução

- O objetivo principal deste trabalho é desenvolver um modelo de inteligência artificial que prediz o nível escolar de um aluno de acordo com seus indicadores de desempenho.

Banco de dados

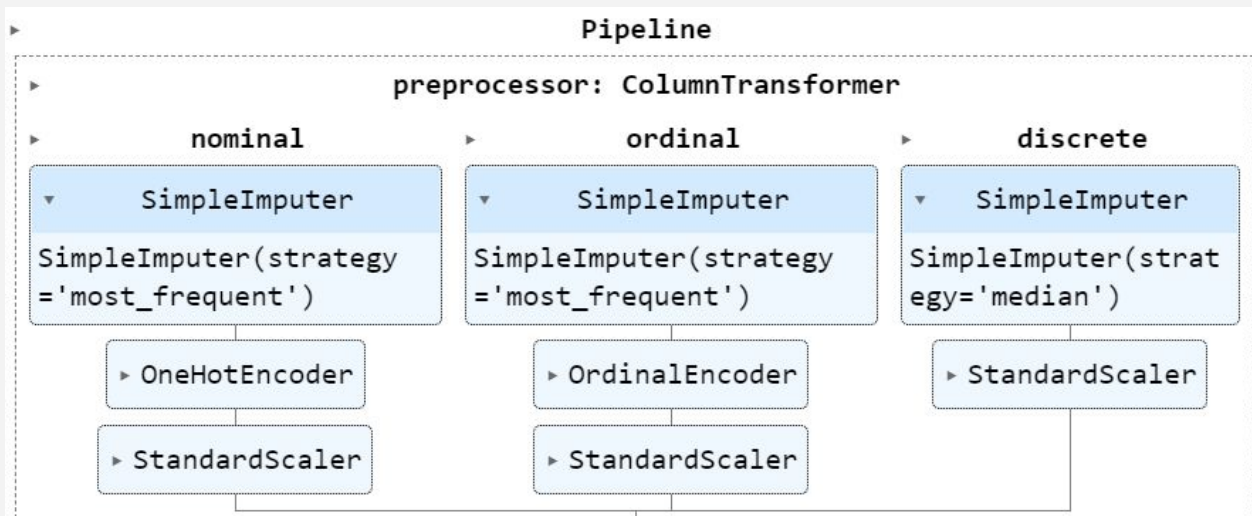
- Base com dados de estudantes.
- 16 campos de indicadores de desempenho.
- 480 dados.

Pipeline



Pré-Processamento

Para realizar o pré-processamento dividimos os dados em nominal, ordinal e discreto para assim aplicar os devidos ajustes em cada tipo de variável.



Modelos

- Para treinamento do conjunto de dados foi utilizado os seguintes modelos:
 - Logistic Regression (LR)
 - K-Nearest-Neighbors (KNN)
 - Random Forest (RF)
 - BernoulliNB (NB)

Treinamento

- Para conseguir extrair o melhor resultado de cada modelo e poder compará-los utilizamos o método de validação cruzada ShuffleSplit e para identificar os melhores parâmetros de treinamento o GridSearchCV.
- Foram utilizadas as métricas de avaliação Acurácia, Precisão e Recall.

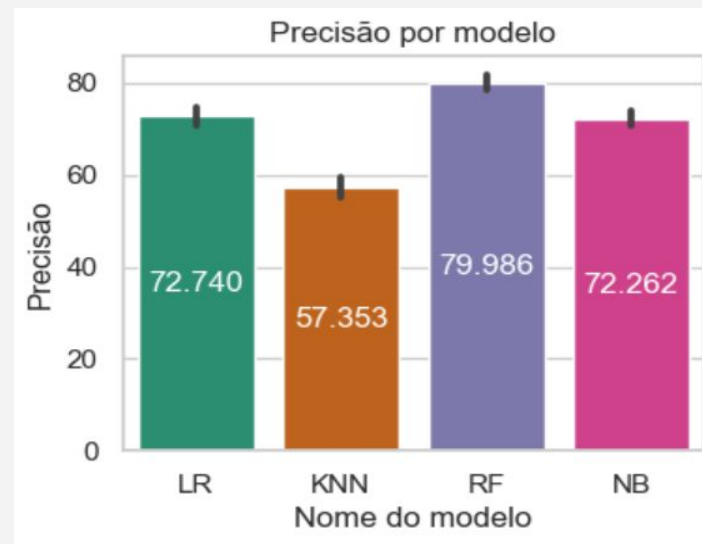
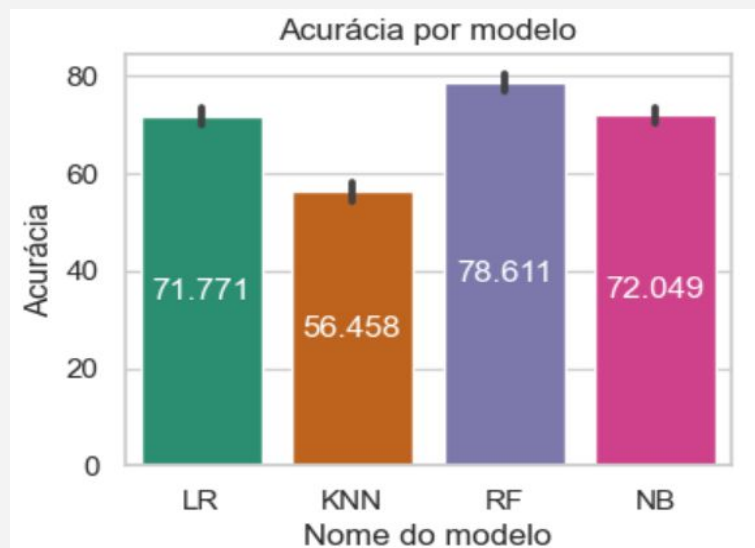
Resultados

- O GridSearch foi utilizado para encontrar o melhor classificador e os melhores parâmetros para serem utilizados.

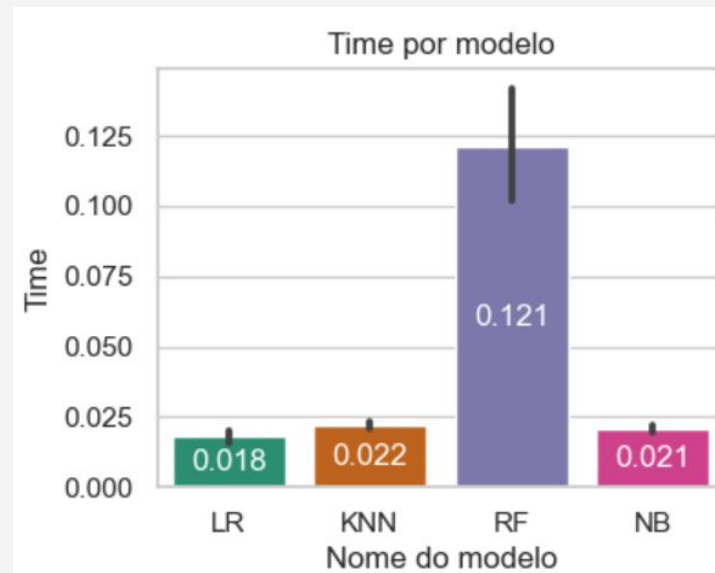
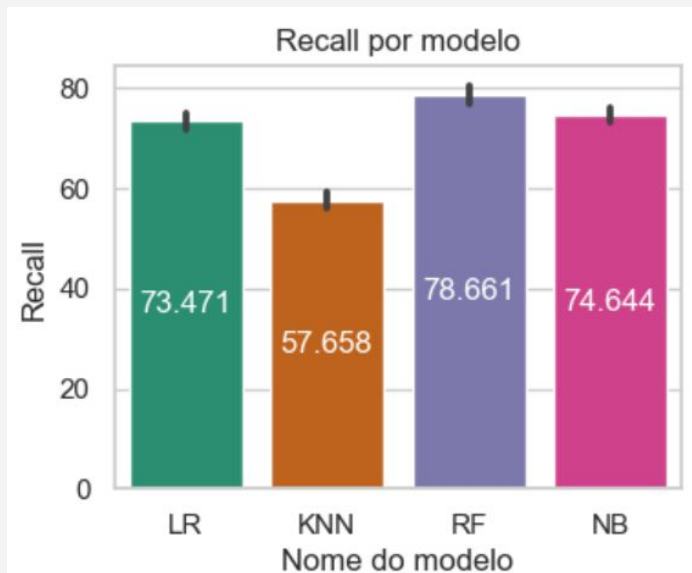
score	KNN	LR	NB	RF
fit_time	0.494 ± 0.193	94.574 ± 94.177	0.155 ± 0.038	245.772 ± 3.843
score_time	0.022 ± 0.003	0.018 ± 0.006	0.021 ± 0.004	0.121 ± 0.057
test_acuracia	0.565 ± 0.057	0.718 ± 0.049	0.720 ± 0.045	0.786 ± 0.050
test_precisão	0.574 ± 0.058	0.727 ± 0.049	0.723 ± 0.045	0.800 ± 0.046
test_recall	0.577 ± 0.050	0.735 ± 0.047	0.746 ± 0.038	0.787 ± 0.049

O melhor modelo é o : **RF**

Resultados



Resultados



Conclusão

- É de suma importância a utilização de métodos que possam comprovar que o modelo desenvolvido possui a maior capacidade de classificação possível.
- O melhor modelo experimentado foi o Random Forest, possuindo o maior valor de acurácia, precisão e recall.
- O KNN obteve os piores resultados de classificação.

Conclusão

- O Random Forest obteve o maior tempo de treinamento.
- O Naive Bayes obteve o menor tempo de treinamento.
- Para trabalhos futuros, fica previsto a realização de experimentos com outros algoritmos de classificação.