**Trabalho Final: RandomForest & DecisionThree**

A intenção do trabalho é ao se utilizar os modelos de random forest e decision three iremos tentar prever com maior acurácia ao analisar alguns dados sendo: gênero, raça, nota de matemática, nota de escrita e nota de leitura iremos tentar prever se esse aluno fez ou não algum curso preparatório.

O motivo da escolha dos modelos se deu pelo fato da similaridade entre os dois modelos, então iremos tentar ver qual modelo será melhor para essa função.

Descrição dos Atributos:

gender: genero do aluno (masculino; feminino) -

race: raça do aluno (grupo A; grupo B; grupo C) -

math score: nota de matematica -

reading score: nota de escrita -

writing score: nota de leitura -

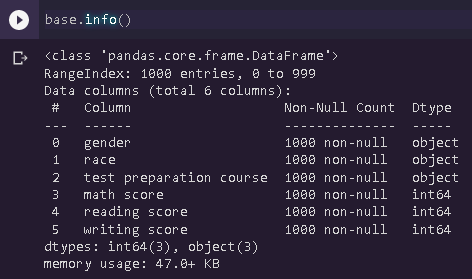
parental level of education: nivel escolar dos pais (removido)

Descrição da classe:

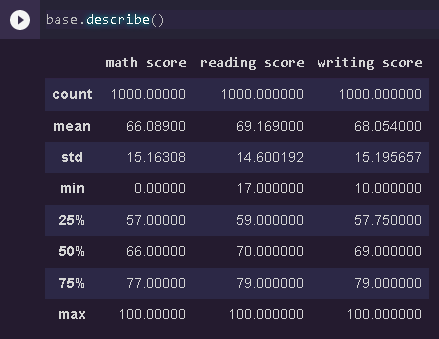
test preparation course: se o aluno fez algum curso preparatório (none; completed)

Dados retirados do site kaggle: <https://www.kaggle.com/spscientist/students-performance-in-exams>

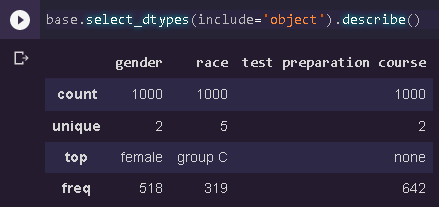
**Análise dos dados:**

****

.

**Relatórios dos algoritmos:**

Podemos ver que não há nenhum valor acima de 100 então não há necessidade de manutenção no mesmo .

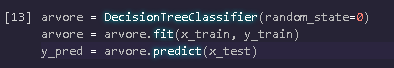


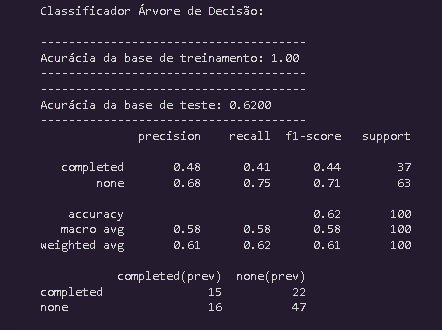
Podemos observar que o gênero dominante é o feminino, o grupo dominante é o C e a maioria dos alunos não fizeram nenhum curso preparatório:

**Resultado das análises:**

* **Decision Three:**

**Primeira arvore:**

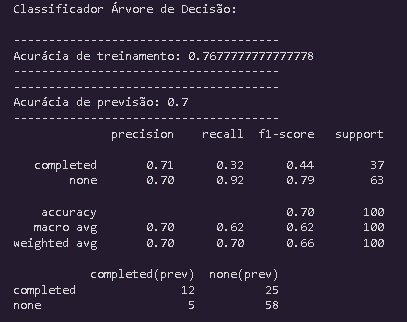
****

**Resultado:**

**Segunda arvore:**

****

**Resultado:**

****

**Considerações finais:**

O que podemos observar nessa análise é que inicialmente tínhamos a primeira arvore de decisão sem nenhum parâmetro definido e tínhamos um overfitting, logo ajustamos isso na segunda arvore fazendo assim diminuir a acurácia, porém evitar o overfitting.

* **RandomForest:**

**Primeira floresta:**

**Texto

Descrição gerada automaticamente**

**Resultado:**

**Texto

Descrição gerada automaticamente**

**Segunda floresta:**

****

**Resultado:**

**Texto

Descrição gerada automaticamente com confiança média**

**Considerações finais:**

O que podemos observar nessa análise é que inicialmente tínhamos a primeira floresta sem nenhum parâmetro definido e tínhamos um overfitting, logo ajustamos isso na segunda floresta fazendo assim diminuir a acurácia, porém evitar o overfitting.

**Resultados:**

Devido as similaridades podemos ver resultados bem parecidos entre os modelos, porém no fim podemos ver que o modelo de RandomForest se sobressaiu entre os dois modelos, sendo assim em casos parecidos com esse na nossa opinião é melhor se utilizar o modelo de RandomForest.