**UNIFACISA – Campina Grande/PB**

Pós-Graduação em Ciência de Dados

Disciplina: introdução a Ciência de Dados

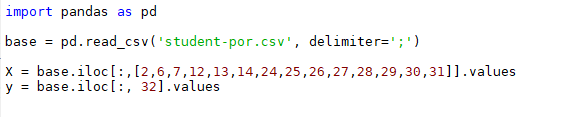
Professor (a); Jones Granatyr.

Aluno (a): Edvaldo Bezerra de Azevedo Filho.

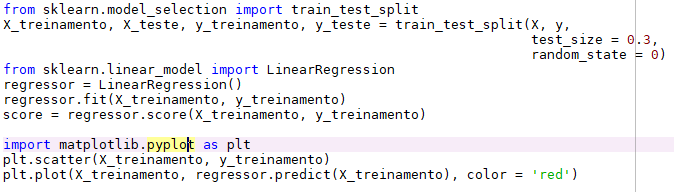
Descrição base de dados:

* O dataset selecionado para desenvolvimento desta atividade foi uma base referente ao desempenho escolar de alunos de duas escolas portuguesas:
* Link do repositório: <https://archive.ics.uci.edu/ml/datasets/Student+Performance>
* O dataset é composto por 649 registros e 33 atributos (colunas)

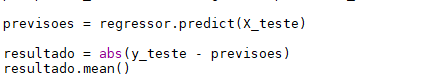
Desenvolvimento:



* Neste passo da figura acima foi realizado o import da lib pandas, dando sequência pela leitura da base de dados, utilizando delimitador “; ”.
* Para os atributos previsores foram selecionados apenas colunas com valores numéricos e para previsão foi escolhida a última coluna da base que é referente a nota final do aluno. Sendo o objetivo da atividade, prever qual seria a nota do aluno dada as características selecionadas.
* Para tal, foi utilizada a técnica de regressão linear.



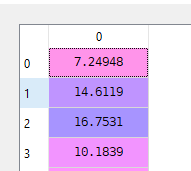
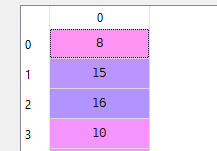
Nesta etapa foi realizado a divisão de um percentual da base para que o modelo fosse treinado e outra parte para avaliação do modelo.



Após o treinamento do modelo, foi chamada a função para prever com a base de teste.

Uma observação que deve ser levada em consideração é que a média final (target) é um valor número de 0 a 20.

Previsões modelo Valor Real

Em resumo a média de acertos de previsão do modelo foi de aproximadamente 86%.

