

Bootcamp: Arquiteto(a) Big Data

Trabalho Prático

Módulo 1: Fundamentos de Big Data

Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Coleta de dados.
2. Analisar e realizar tratamento de dados.
3. Criar visualização de dados.
4. Implementar algoritmo de Machine Learning.
5. Analisar resultados obtidos.
6. Conhecimento teórico ministrado nas videoaulas.

Enunciado



Um docente do ensino superior apresentou um desafio, ele consistia em elaborar um protótipo preditivo que empregasse o tempo dedicado pelos estudantes aos estudos como base para antecipar o desempenho na avaliação final da matéria.

Para tal propósito, o professor coletou informações sobre as horas de estudo investidas e os resultados finais das provas de uma amostra de alunos de períodos anteriores. Consequentemente, o docente constatou que a abordagem mais adequada para esse cenário seria a criação de um algoritmo

de Aprendizado de Máquina, especificamente uma regressão linear, como solução para essa questão.

Portanto, o modelo será capacitado com os dados históricos já reunidos e estará apto a prever o possível resultado obtido com base nas horas de estudo de um novo estudante.

Atividades

Para essa atividade, os alunos deverão criar um algoritmo de regressão linear para prever a nota obtida na avaliação final do aluno, baseado no número horas dedicada aos estudos.

1. Criar um projeto no Google Drive;
2. Coletar e inserir o arquivo horas_estudo.csv na plataforma;
3. Analisar os dados coletados;
4. Avaliar a relação entre as variáveis;
5. Criar algoritmo de regressão linear;
6. Responder as questões teóricas e práticas do trabalho;

Dicas do professor:

1. Analisem com cuidado os dados através da representação gráfica.
2. Analisem bem o gráfico gerado e a disponibilização dos dados.
3. Antes de enviar as respostas verifiquem se o gabarito está correto.
4. Tenham atenção no que pede cada questão.
5. Os dados disponibilizados no dataset são fictícios. Ou seja, não tem relação com o mundo real.
6. O dataset utilizado no trabalho pode ser obtido no link:
 - a. <https://github.com/ProfLeandroLessa/classroom-datasets/tree/master/FDA/TP>