

COMPONENTE CURRICULAR:	Projeto aplicado II
NOME COMPLETO DO ALUNO:	Mariana Simões Rubio; Patrícia Corrêa França; Valdiney Atílio Pedro.
RA:	10424388; 10423533; 10424616.

## APLICANDO CONHECIMENTO A3

Link Github: <https://github.com/valdineyatilio/ProjetoAplicado-II>

### 1. Método analítico aplicado à base de dados escolhida:

Os métodos analíticos foram aplicados à base de dados escolhida, utilizando os seguintes algoritmos:

Exploração e visualização de dados: Histogramas, gráficos de dispersão, gráficos de barras etc.

Tratamento de dados: Limpeza, transformação, e normalização de dados.

Modelos de predição: Redes Neurais Artificiais (MLPRegressor) e RandomForestClassifier.

Análise estatística: ANOVA para comparação de categorias.

Os algoritmos se encontram na pasta da Aula-03.

### 2. Medidas de acurácia:

▪ Utilizamos as seguintes métricas para avaliar a acurácia dos modelos:

- Erro Quadrático Médio (MSE): Avaliado para o modelo de regressão.
- Precisão, Recall, F1-Score: Avaliados para o modelo de classificação.
- AUC-ROC: Calculado para o modelo de classificação (quando aplicável).

O algoritmo se encontra na pasta da Aula-03.

3. Descrição dos resultados preliminares, apresentando um produto gerado e rascunhando um possível modelo de negócios:

▪ Descrição dos Resultados Preliminares dos Algoritmos

a. Análise Exploratória dos Dados

Objetivo: Entender a estrutura e características da base de dados do YouTube.

Resultados:

Visualizações Gráficas: Histogramas, gráficos de dispersão e gráficos de barras mostraram a distribuição das variáveis numéricas e a relação entre visualizações de vídeo e outras variáveis.

Verificação de Dados: Identificação de valores ausentes e outliers, proporcionando uma base para tratamento de dados posterior.

Mapa de Calor: Análise de correlação identificou relações significativas entre variáveis, fornecendo insights para modelos preditivos.

b. Redes Neurais Artificiais (RNA) com Regressor MLP

Objetivo: Prever os ganhos anuais mais altos com base em assinantes, visualizações e uploads.

Resultados:

Erro Quadrático Médio (MSE): O modelo MLPRegressor apresentou um MSE relativamente baixo, indicando uma boa capacidade de predição dos ganhos anuais.

Treinamento e Teste: Divisão dos dados em conjuntos de treinamento e teste garantiu a robustez do modelo. O modelo foi capaz de generalizar bem os dados de teste.

c. ANOVA (Análise de Variância)

Objetivo: Comparar as visualizações de vídeos entre diferentes categorias do YouTube.

#### Resultados:

Estatística F e Valor p: ANOVA indicou diferenças significativas nas visualizações médias entre categorias. Um valor p baixo sugere que pelo menos uma categoria de vídeos tem um número significativamente diferente de visualizações.

#### d. Algoritmo Apriori para Regras de Associação

Objetivo: Encontrar associações entre variáveis no conjunto de dados, como entre país de origem e ganhos anuais.

#### Resultados:

Regras de Associação: Descobertas regras significativas indicando que determinados países estão associados a altos ganhos no YouTube.

Métricas de Avaliação: Suporte, confiança e lift foram utilizados para avaliar a relevância das regras geradas. Regras com alto suporte e confiança são particularmente úteis para insights comerciais.

#### e. RandomForestClassifier

Objetivo: Classificar e prever variáveis categóricas com base em diferentes características do canal do YouTube.

#### Resultados:

Acurácia, Precisão, Recall, F1-Score: O modelo apresentou bons valores para essas métricas, indicando um desempenho robusto na classificação.

AUC-ROC: Em situações com mais de uma classe, o AUC-ROC mostrou a capacidade do modelo de distinguir entre diferentes classes de forma eficaz.

#### ▪ Produto Gerado

Os algoritmos aplicados à base de dados do YouTube fornecem insights valiosos sobre o desempenho dos canais, previsão de ganhos e associações entre variáveis. Esses resultados preliminares formam a base para o desenvolvimento de um modelo de negócios focado em consultoria para criadores de conteúdo, ajudando-os a otimizar seus canais e maximizar seus ganhos.

#### ▪ Modelo de Negócios

##### Consultoria para Criadores de Conteúdo

##### Descrição:

A empresa pode oferecer serviços de consultoria para criadores de conteúdo do YouTube, utilizando modelos preditivos para otimizar a performance dos canais. Os serviços incluirão análise de dados detalhada, insights preditivos, recomendações personalizadas e estratégias de crescimento baseadas nos dados históricos e previsões futuras.

##### Justificativa:

Os algoritmos utilizados, como regressão com redes neurais e análise de dados exploratória, fornecem uma base sólida para entender e prever o comportamento dos canais do YouTube. Com essas ferramentas, é possível identificar padrões de crescimento, prever ganhos anuais e mensais, além de oferecer recomendações baseadas em dados para aumentar o engajamento e os ganhos dos criadores.

##### Benefícios:

**Análise de Desempenho:** Avaliação detalhada do desempenho do canal e dos vídeos, utilizando métricas como visualizações, inscritos e ganhos.

**Previsões de Ganhos:** Modelos preditivos para prever ganhos futuros com base em variáveis como número de inscritos, visualizações e uploads.

**Recomendações Personalizadas:** Estratégias baseadas em dados para otimizar conteúdo, horários de postagem, e engajamento com o público.

**Monitoramento Contínuo:** Ferramentas para monitorar o desempenho do canal em tempo real e ajustar estratégias conforme necessário.

Este modelo de negócios alavanca as capacidades analíticas dos algoritmos aplicados e oferece um serviço de alto valor para criadores de conteúdo que buscam maximizar seu impacto e ganhos no YouTube.

#### 4. Esboço do storytelling

Arquivo DataStorytelling-Aula3.pdf está no link do GitHub que foi fornecido neste documento.