



**RESIDÊNCIA EM SOFTWARE 36 - BAHIA**

**TRILHA CIÊNCIA DE DADOS**

**GABRYELLE BEATRIZ DUARTE MORAES**

**JOÃO PEDRO DE JESUS SILVA**

**RELATÓRIO TÉCNICO**

**IMPLEMENTAÇÃO E ANÁLISE DO ALGORITMO DE REGRESSÃO LINEAR**

Vitória da Conquista

2024

Gabryelle Beatriz Duarte Moraes

João Pedro de Jesus Silva

## **RELATÓRIO TÉCNICO**

Implementação e Análise do Algoritmo de Regressão Linear

Relatório apresentado à Trilha de Ciência de dados do ResTIC 36 como requisito integral para aprovação na unidade 9.

Vitória da Conquista

2024

## **RESUMO**

Este relatório apresenta, a partir de uma análise exploratoria do dataset '*Top Instagram Influencers Data (Cleaned)*' , disponível no *Kaggle* , o desenvolvimento de um modelo preditivo utilizando o algoritmo de Regressão Linear que estuda como a nacionalidade do influencer impacta na taxa de engajamento.

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>3</b>
<b>2 METODOLOGIA.....</b>	<b>4</b>
<b>3 RESULTADOS.....</b>	<b>8</b>
<b>4 DISCUSSÃO.....</b>	<b>9</b>
<b>5 CONCLUSÃO E TRABALHOS FUTUROS.....</b>	<b>10</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>13</b>

## 1 INTRODUÇÃO

A regressão linear é uma técnica estatística utilizada na análise de dados para entender a relação entre variáveis. Seu objetivo é modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. A aplicação de linhas de regressão em um estudo exige conhecimento de fundamentos estatísticos básicos e interpretação correta dos resultados obtidos.

Uma das principais características da regressão linear é a simplicidade. Draper e Smith (1998) enfatizaram que este modelo é particularmente útil na determinação da magnitude e direção dos efeitos das variáveis de base nas variáveis dependentes. Esse artifício torna os algoritmos uma ótima ferramenta para analisar dados e explicar resultados com clareza.

No contexto das redes sociais, a análise dos fatores que influenciam o engajamento e popularidade de um indivíduo se tornaram uma área relevante de pesquisa. Investigadores como Kaplan e Haenlein (2010) argumentam que as diferenças culturais e regionais (por exemplo, país) podem ter um impacto significativo nas atividades online. A utilização da regressão linear neste contexto permite a análise dessas relações servindo como base para a otimização de estratégias digitais.

A partir do exposto, este relatório tem como finalidade apresentar o desenvolvimento de um modelo preditivo usando o algoritmo de Regressão Linear para determinar como a nacionalidade do influencer impacta na taxa de engajamento na rede social Instagram.

## 2 METODOLOGIA

Etapa 1: Análise inicial do conjunto de dados:

O Dataset *Top Instagram Influencers* retirado da plataforma Kaggle foi importado para o Google Collab, onde foi realizado a manipulação e análise. O material tem 201 registros com as seguintes categorias:

**Rank:** Posição no ranking;

**Channel\_info:** Nome do canal ou influenciador;

**Influence\_score:** Pontuação de influência;

**Posts:** Número de publicações;

**Followers:** Número de seguidores;

**Avg\_likes:** Média de curtidas por publicação;

**60\_day\_eng\_rate:** Taxa de engajamento em 60 dias (em porcentagem);

**New\_post\_avg\_like:** Curtidas médias em novas publicações;

**Total\_likes:** Total de curtidas no perfil;

**Country:** Nacionalidade.

## Etapa 2: Tratamento

Em seguida as colunas formatadas com valores string com 'm' e 'k' bem como a apresentada em porcentagem foram convertidas em numéricas. Os valores nulos da coluna de nacionalidade foram tratados. Houve a distribuição da taxa de engajamento por país e realizou-se a correlação entre essas variáveis.

## 3 RESULTADOS

```
data['total_likes'] = data['total_likes'].str.replace('b', '').replace(',', '.').astype(float) * 1_000_000_000
data['country'] = data['country'].fillna('Unknown') # Preencher nulos com 'Unknown'
data[['total_likes', 'country']].head()
```

## 4 DISCUSSÃO

O projeto revelou que a maioria dos influenciadores mais populares no Instagram são estrangeiros, com destaque para americanos. Isso reflete o domínio econômico e cultural dos Estados Unidos, que influenciam globalmente através de suas plataformas digitais e marcas, ampliando a visibilidade de influenciadores de língua inglesa. Essa predominância levanta questões sobre a falta de diversidade e representatividade nas redes sociais, já que influenciadores de outros países enfrentam barreiras linguísticas e menor alcance.

## 5 CONCLUSÃO E TRABALHOS FUTUROS

O projeto indicou que a maioria dos influenciadores populares são estrangeiros, especialmente dos Estados Unidos, devido à sua influência cultural e econômica nas redes sociais, refletindo uma falta de diversidade e representatividade de influenciadores de outros países.

Para aprimorar o projeto, seria necessário realizar uma avaliação mais detalhada de outros fatores que influenciam a taxa de engajamento, como características demográficas dos seguidores, tipo de conteúdo produzido e horário das postagens. Analisar esses dados em conjunto ajudaria a identificar variáveis ocultas e proporcionar uma visão mais precisa e inclusiva sobre o impacto e a relevância dos influenciadores de diferentes regiões.

## REFERÊNCIAS

Draper, N. R., & Smith, H. (1998). **Applied Regression Analysis**. Wiley.

Garza, Kristian, Carole Goble, John Brooke e Caroline Jay. “**Enquadrando a Interface do Sistema de Dados Comunitários.**” Anais da Conferência HCI Britânica de 2015. ACM, 13 de julho de 2015.

Kaplan, A. M., & Haenlein, M. (2010). **Users of the world, unite! The challenges and opportunities of Social Media**. Business Horizons.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). **Introduction to Linear Regression Analysis**. Wiley.