

ANÁLISE DE DADOS DE INFLUENCIADORES NO INSTAGRAM

DESCRIÇÃO DO PROJETO

Este projeto tem como objetivo analisar dados de influenciadores no Instagram para explorar padrões e correlações relevantes e implementar um modelo kNN (k-Nearest Neighbors) para prever a influência dos usuários. A análise inclui a visualização da distribuição de seguidores, curtidas, taxas de engajamento, e a construção de um modelo preditivo baseado em variáveis selecionadas.

PRE-REQUISITOS

Certifique-se de que possui um ambiente com Python instalado. Você pode usar qualquer IDE, como:

- **PyCharm** (usado no desenvolvimento do projeto);
- Visual Studio Code;
- Jupyter Notebook.

INSTALAÇÃO

Instale as seguintes bibliotecas: `pip install pandas numpy matplotlib seaborn scikit-learn`.

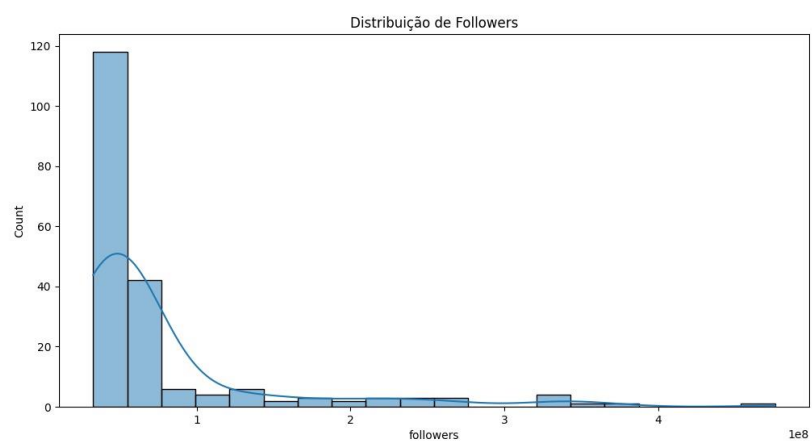
Faça o download do projeto via GitHub:

- `git clone https://github.com/silass4ntos/Algoritimo_Knn.git`
`cd Algoritimo_Knn`.

COMO EXECUTAR

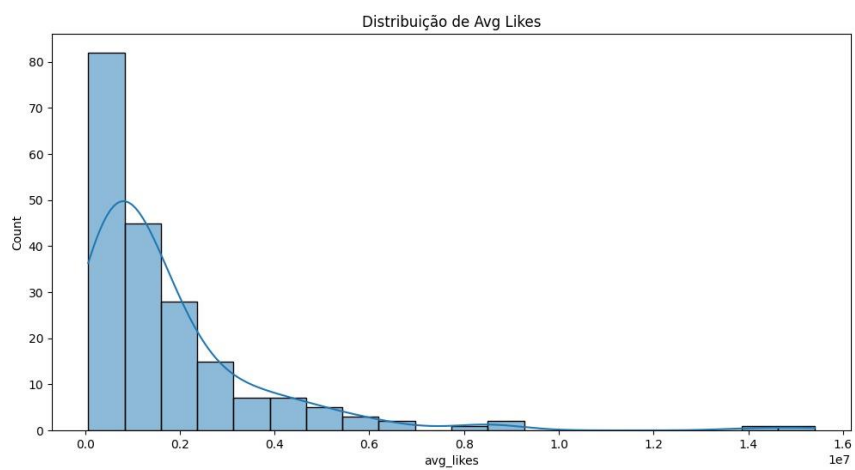
1. Certifique-se de que os dados de entrada estão na pasta **Knn/**.
2. O arquivo de dados necessário para análise já está incluído no repositório.
3. Para executar a análise e os modelos, utilize:
4. `python Knn/main.py`

Grafico 1 - Distribuição de folowers



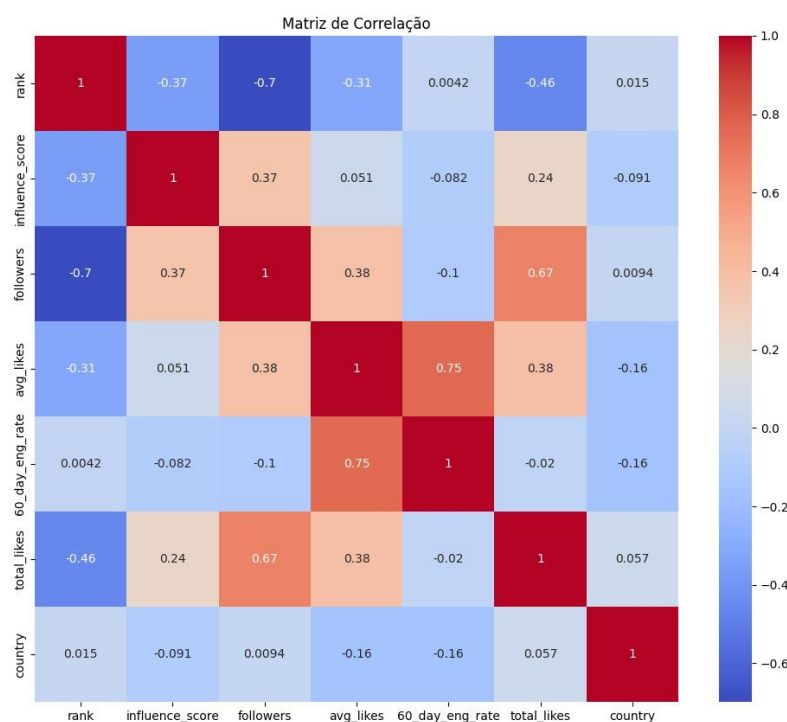
Fonte: Elaboração própria.

Grafico 2 - Distribuição de Avg Likes



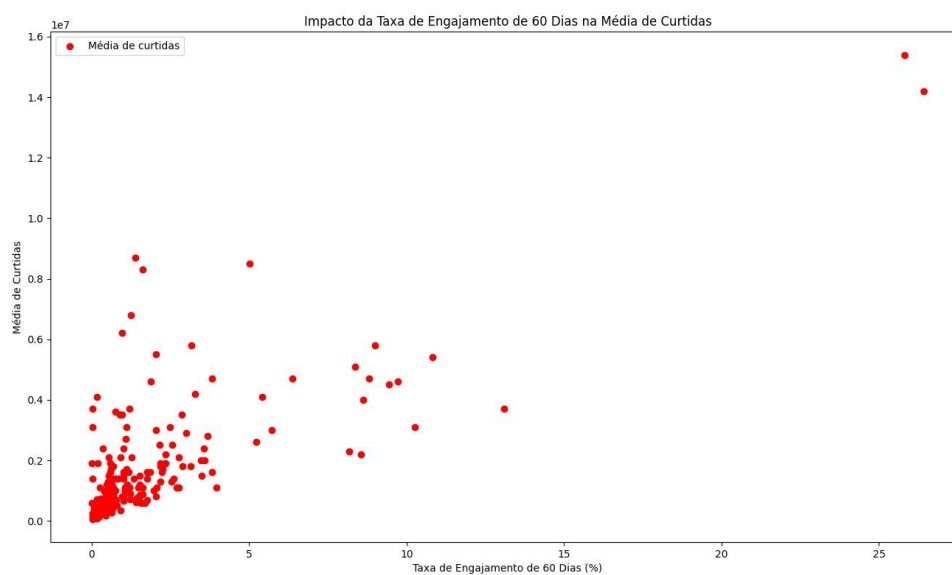
Fonte: Elaboração própria.

Grafico 3 – Matriz de correlação



Fonte: Elaboração própria.

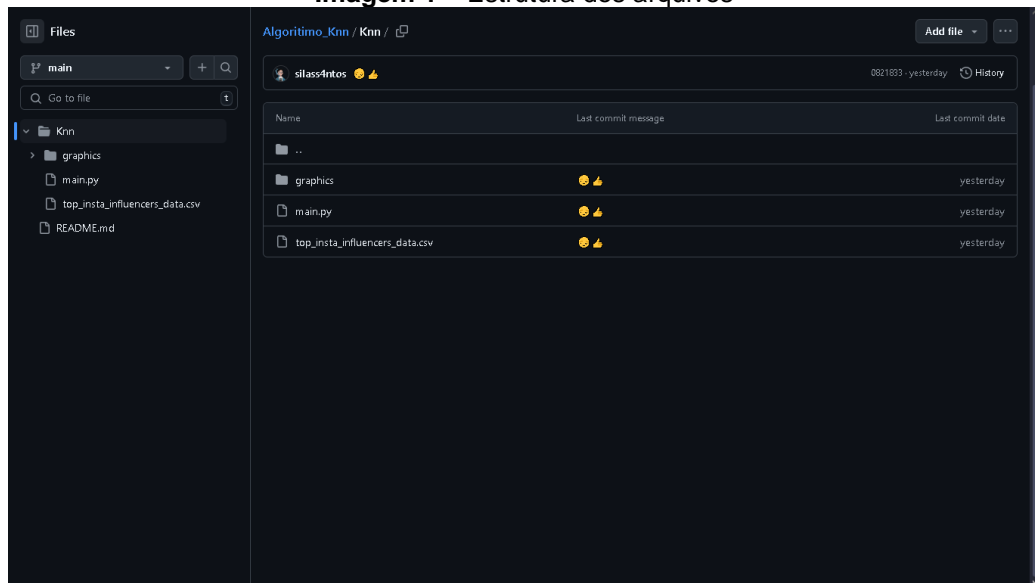
Grafico 4 – Impacto da taxa de engajamento de 60 dias na média de curtidas



Fonte: Elaboração própria.

ESTRUTURA DOS ARQUIVOS

Imagem 1 – Estrutura dos arquivos



Fonte: GITHUB

- Knn/ pasta main com os arquivos do código;
- main.py - script;
- top_insta_influencers_data.csv - arquivo com os dados a serem analisados;
- graphics/ - com os gráficos que foram gerados;
- README.md/

TECNOLOGIAS UTILIZADAS

- **Python:** Linguagem principal usada no projeto.
- **Pandas:** Para manipulação e análise de dados.
- **Seaborn:** Para criação de visualizações estatísticas.
- **Scikit-Learn:** Para implementação do modelo kNN e outras técnicas de aprendizado de máquina.
- **Matplotlib:** Para criação de gráficos e visualizações.

AUTORES E COLABORADORES

Grupo 50 – autores - Luiz Carlos Abreu Neves dos Santos e Silas Santos Moura

Luiz Carlos contribuiu para a organização do projeto e o planejamento da estrutura geral. Ele esteve envolvido no levantamento inicial das necessidades, no tratamento de dados faltantes, e colaborou na análise exploratória, incluindo a interpretação dos gráficos de distribuição e das correlações entre variáveis. Além disso, Luiz também ajudou na documentação do projeto, garantindo a clareza e organização do material.

Silas foi o principal contribuinte do projeto, sendo responsável pela implementação do modelo kNN e pela otimização do mesmo utilizando GridSearchCV. Ele trabalhou diretamente na transformação dos dados, incluindo a conversão de valores como 'k', 'M', e 'b' para valores numéricos. Silas liderou o desenvolvimento do código, implementou a normalização dos dados e estruturou o pipeline de treinamento e avaliação do modelo. Ele também foi o responsável pela criação dos gráficos de dispersão e da matriz de correlação, além da análise detalhada dos resultados.

AGRADECIMENTOS

Quebramos a cabeça por horas resolvendo bugs, e construindo esta pequena pesquisa. Venho por meio deste, agradecer a equipe CEPEDI, no projeto Restic36, pela oportunidade, e também pelo conhecimento adquirido nesta jornada.

REFERENCIAS

Python Software Foundation. ***Python Documentation***. Disponível em: <https://docs.python.org/3/>. Acesso em: 13 nov. 2024.

PANDAS DEVELOPMENT TEAM. ***Pandas Documentation***. Disponível em: <https://pandas.pydata.org/pandas-docs/stable/>. Acesso em: 13 nov. 2024.

WASKOM, Michael L.; SEABORN DEVELOPMENT TEAM. ***Seaborn Documentation***. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 13 nov. 2024.

PEDREGOSA, Fabian et al. ***Scikit-Learn Documentation***. Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 13 nov. 2024.

HUNTER, John D. ***Matplotlib Documentation***. Disponível em: <https://matplotlib.org/>. Acesso em: 13 nov. 2024.