

Equipe Pantanálise

Integrantes da equipe:

- Allan Menchik da Cunha
- João Pedro Santos Vareiro
- Thiago Lütz Dias
- Vinicius Espindola

Descrição

O trabalho de aprendizado de máquina proposto visa explorar a relação entre a polaridade dos tweets (positiva, negativa ou neutra) e o engajamento gerado por eles no Twitter. Para isso, será feito o scraping de dados do Twitter usando técnicas de web scraping para coletar uma grande quantidade de tweets.

Os tweets serão processados usando técnicas de análise de sentimento para determinar a polaridade de cada um deles. Em seguida, serão analisados os possíveis níveis de engajamento gerados por esses tweets, como número de curtidas, retweets e respostas.

O objetivo é testar a hipótese de que os tweets com polaridade mais negativa geram um engajamento maior do que os tweets com polaridade mais positiva. Para isso, serão usados algoritmos de aprendizado de máquina, como redes neurais, para treinar um modelo capaz de prever o nível de engajamento gerado por um tweet com base em sua polaridade.

Os resultados desse estudo podem ter aplicações práticas no marketing e publicidade digital, ajudando a entender como as empresas podem gerar mais engajamento nas redes sociais com base na polaridade dos seus posts. Além disso, o trabalho contribui para o desenvolvimento de novas técnicas de análise de sentimento e aprendizado de máquina aplicados a dados de redes sociais.

Entregável

O entregável proposto é uma aplicação web capaz de prever o nível de engajamento de um tweet com base em sua polaridade. Através da inserção de um possível tweet na plataforma, o usuário poderá obter uma estimativa do potencial de engajamento da mensagem.

A aplicação utiliza técnicas de análise de sentimento para determinar a polaridade do tweet, classificando-o como positivo, negativo ou neutro. Em seguida, o algoritmo de aprendizado de máquina treinado com base em dados de tweets coletados em redes sociais é utilizado para prever o nível de engajamento gerado por essa mensagem.

Utilidade da Ferramenta

O usuário poderá utilizar a aplicação para obter insights sobre como melhorar a efetividade de sua presença nas redes sociais. Através da previsão do engajamento gerado por um tweet, ele pode ajustar sua estratégia de comunicação, adaptando a linguagem e o conteúdo de suas mensagens para gerar um maior impacto junto ao seu público.

A aplicação também pode ser útil para empresas que desejam maximizar o engajamento em suas campanhas de marketing digital. Ao utilizar a plataforma para testar diferentes variações de uma mesma mensagem, a equipe de marketing pode selecionar a opção com maior potencial de engajamento antes de lançá-la ao público em geral.

Por fim, é importante destacar que a aplicação é baseada em algoritmos de aprendizado de máquina, que são alimentados com dados coletados em redes sociais. Por esse motivo, os resultados podem variar de acordo com as particularidades do público e do contexto em que o tweet é inserido.

Ferramentas

- Python
- Pytorch
- Pytest
- FastAPI
- Django
- React
- HTML
- Git
- [Github Organização](#)

Fluxo de trabalho

Para facilitar a colaboração e o gerenciamento das versões deste projeto, será utilizado o modelo de fluxo de trabalho baseado no Git Flow. Este modelo é baseado em um conjunto de práticas e convenções que buscam tornar o desenvolvimento mais organizado e estruturado.

As branches a serem utilizadas serão:

- Branch main: branch principal do projeto que contém apenas código estável e pronto para produção.
- Branch develop: branch que contém o código em desenvolvimento, com funcionalidades novas, mas que ainda não estão prontas para produção.
- Branches feature: normalmente usadas para o desenvolvimento de novas funcionalidades, mas também podem ser utilizadas para correções de erros simples, desde que o erro seja encontrado durante o desenvolvimento de uma nova funcionalidade. Quando a funcionalidade ou correção estiver concluída, a branch feature será mesclada de volta na branch develop por meio de pull requests. A nomeação das branches feature será feita de acordo com o kanban anexado ao projeto, seguindo as convenções do modelo de fluxo de trabalho adotado no projeto.
- Branches Hotfix: são branches criadas para a solução de problemas que devem ser corrigidos rapidamente. Quando o hotfix estiver concluído, ele será mesclado na branch develop.

Commits

Neste projeto, o modelo de commits será baseado no texto "Semantic Commit Messages" disponível em <https://gist.github.com/qoomon/5dfcdf8eec66a051ecd85625518cfd13>. Esse modelo de commits segue uma convenção padronizada que inclui uma descrição clara e concisa do tipo de mudança realizada no código e uma breve descrição do que foi alterado.

Além disso, essa abordagem permite que os commits sejam facilmente rastreados e entendidos por outros membros da equipe de desenvolvimento. A adoção desse modelo de commits é mais uma medida para tornar o desenvolvimento mais organizado e estruturado, garantindo maior clareza e eficiência ao processo.

Cronograma

Data	Semana	Conteúdo	Entregável
01/04/2023	1	Regressão Linear, Regressão Logística, Gradiente estocástico em lote; (prática) treinamento de regressores lineares, logísticos, comparativo de estocástico versus lote;	definição do projeto, scrapy de textos
15/04/2023	2	Backpropagation, Perceptron e MLP (prática); treinamento e construção de uma MLP from scratch;	Rotulação dos dados.
29/04/2023	3	WordPiece, positional encoding, self attention, cross attention, masked attention; (prática) loop treino pytorch, implementação e treinamento do BERT/GPT	Modelo para classificação dos textos.
06/05/2023	4	Apresentação dos projetos com discussões de melhorias	Projeto com front-end da aplicação
13/05/2023	5	apresentação dos projetos com discussões de melhorias	projeto com front-end da aplicação.
20/05/2023	6	Demo Day	-