**Título do Projeto**: Análise de Sentimento em Redes Sociais sobre Produtos de Tecnologia

Olinda, 2024

**1. Introdução**

Este projeto visa realizar uma análise de sentimento em postagens de redes sociais para identificar a percepção dos usuários sobre produtos de tecnologia, como smartphones, laptops e tablets. Usamos técnicas de processamento de linguagem natural (NLP) e ferramentas de Big Data para processar e analisar um grande volume de dados em tempo real.

**2. Objetivos**

* Coletar dados de postagens em redes sociais relacionados a produtos de tecnologia.
* Aplicar técnicas de processamento de linguagem natural (NLP) para análise de sentimentos.
* Determinar o sentimento geral dos consumidores em relação aos produtos analisados.
* Visualizar os resultados da análise de sentimento ao longo do tempo para detectar tendências.

**3. Arquitetura do Sistema**

A arquitetura do projeto segue o padrão típico de pipelines de Big Data:

* **Coleta de Dados**: Dados de postagens em redes sociais foram coletados utilizando a API do Twitter e armazenados em um Data Lake no Amazon S3.
* **Processamento de Dados**: Utilizamos o Apache Spark para processar e filtrar as postagens, aplicando algoritmos de NLP para determinar o sentimento de cada postagem (positivo, negativo ou neutro).
* **Armazenamento**: Os resultados da análise foram armazenados em um banco de dados NoSQL (Cassandra) para permitir consultas rápidas e escaláveis.
* **Visualização**: Os resultados foram visualizados em tempo real usando o Tableau para mostrar as tendências de sentimento.

**4. Tecnologias Utilizadas**

* **Linguagem de Programação**: Python (Pandas, PySpark, NLTK para NLP)
* **Big Data Framework**: Apache Spark para processamento distribuído.
* **Armazenamento**: Amazon S3 para Data Lake, Cassandra para banco de dados.
* **Coleta de Dados**: Twitter API.
* **Visualização de Dados**: Tableau.
* **Ambiente de Computação**: AWS (EC2 e EMR para Spark).

**5. Coleta de Dados**

* **Fonte de Dados**: Utilizamos a API pública do Twitter para buscar tweets contendo palavras-chave relacionadas a produtos de tecnologia (ex.: "iPhone", "Samsung Galaxy", "MacBook", "Dell Laptop").
* **Período de Coleta**: Coletamos dados ao longo de um período de 2 meses (de junho a agosto de 2024).
* **Volume de Dados**: Aproximadamente 500.000 tweets foram coletados.

**6. Processamento de Dados**

* **Limpeza dos Dados**: Removemos postagens duplicadas, spam e tweets irrelevantes.

**7. Armazenamento de Dados**

Os dados processados foram armazenados em um banco de dados **Cassandra** para permitir consultas rápidas. Escolhemos Cassandra pela sua escalabilidade e desempenho em cenários de grandes volumes de dados.

**Estrutura do Banco de Dados:**

* **Tabela:** 
  + tweet\_id: ID único do tweet.
  + sentimento: Resultado da análise de sentimento (positivo, negativo, neutro).
  + data: Data do tweet.
  + produto: Produto mencionado no tweet.

**8. Visualização dos Resultados**

Utilizamos o Tableau para criar dashboards interativos que mostram o sentimento geral por produto ao longo do tempo. As visualizações incluem:

* Gráficos de linhas mostrando a variação do sentimento (positivo/negativo/neutro) para cada produto.
* Gráficos de barras comparando os produtos mais comentados e com maior número de sentimentos negativos/positivos.

**9. Resultados**

* **Produtos mais mencionados**: iPhone, MacBook, Samsung Galaxy.
* **predominante**: Os produtos da Apple (iPhone e MacBook) apresentaram predominantemente sentimentos positivos, enquanto a linha de produtos Samsung teve uma distribuição mais equilibrada entre sentimentos positivos e negativos.
* **Análise temporal**: Notamos picos de sentimentos negativos associados a problemas técnicos, como falhas de software ou hardware.

**10. Desafios**

* **Volume de Dados**: Processar 500.000 tweets de forma eficiente exigiu otimizações no pipeline de Spark.
* **Qualidade dos Dados**: Muitos tweets irrelevantes e spam precisaram ser filtrados manualmente.

**11. Conclusão**

Este projeto demonstrou a aplicação de técnicas de Big Data e NLP na análise de sentimento em redes sociais. Os resultados fornecem insights valiosos sobre a percepção dos consumidores em relação a produtos de tecnologia, destacando tendências de mercado e áreas para melhorias.

12. **Referências**