

PHASE 4 PROJECT

TWEETS SENTIMENTS ANALYSIS:Natural Language Processing (NLP)

Project Overview

- Our aim is to build a Natural Language Processing (NLP) model that can analyze the sentiment expressed in Tweets about Apple and Google products.
- The main objective is to build a model that can automatically classify a Tweet based on its sentiment, enabling businesses, marketers, or product managers to quickly analyze customer feedback and sentiment about products and services.

Business Problem - Monitoring Social Media Sentiment

- **In the Digital Age, Companies Face a Challenge in Monitoring Customer Sentiment Across Social Media.**
- Businesses like Apple and Google are inundated with user-generated content in the form of:
- **Tweets**
- **Reviews**
- **Comments**

Why It Matters:

- Understanding and analyzing this content is key for businesses to:

1. Track Brand Health

Assess positive vs. negative customer feelings about products or services.

2. Measure Customer Satisfaction

Identify satisfied or dissatisfied customers and take proactive actions.

3. Support Marketing Strategies

Align campaigns, advertisements, and promotions with consumer expectations.

4. Gain Product Development Insights

Use feedback to improve products, address flaws, and meet user needs.

Project Objectives

- **Binary Classification Model**

Classify tweets as **positive** or **negative**.

- **Expand to Multiclass Classification**

Add **neutral** sentiment and create a **multiclass classifier** (positive, negative, neutral).

- **Build an NLP Model**

Accurately classify sentiments, identify **brands/products**, and handle **ambiguity** in tweets.

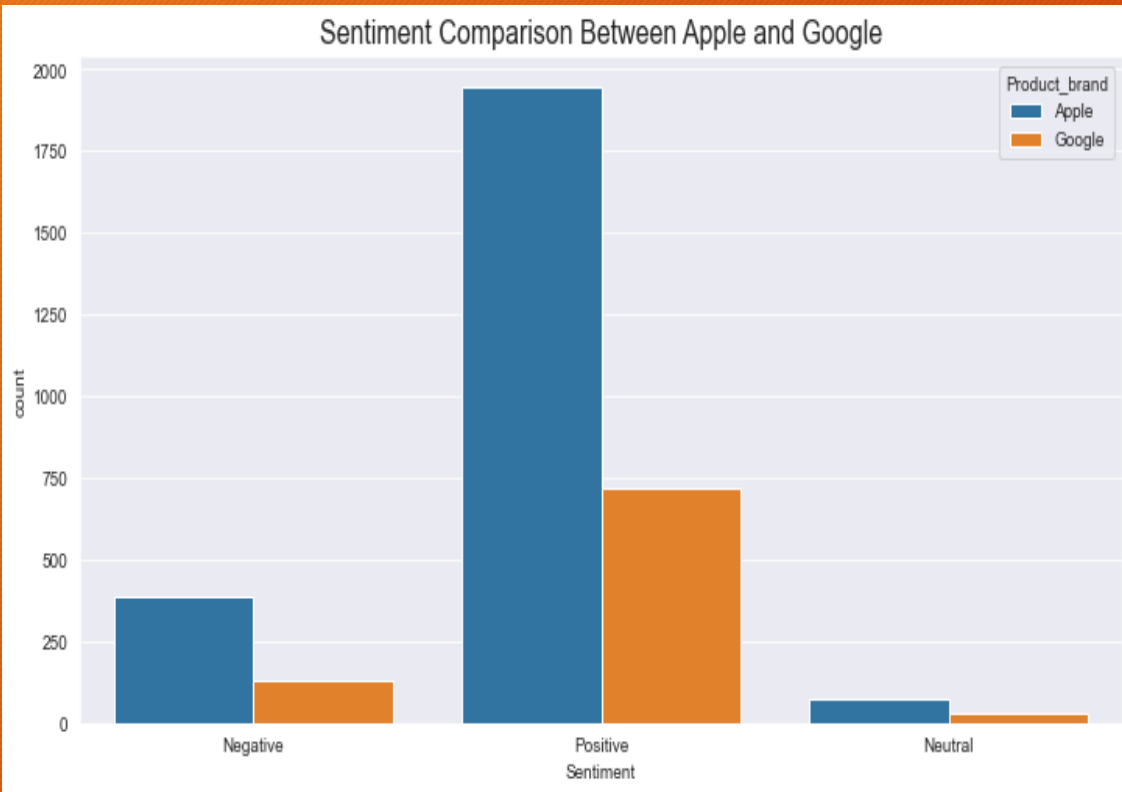
Data Understanding

- **Source:** CrowdFlower via Data.worldApprox. 9,000 tweets expressing sentiments about Apple and Google products.
- **Columns:**
- **tweet_text:** Contains the tweet content.
- **emotion_in_tweet_is_directed_at:** Identifies if emotion is directed at a brand/product.
- **is_there_an_emotion_directed_at_a_brand_or_product:** Flags whether a brand/product is the focus of the emotion

Data Understanding and Preparation

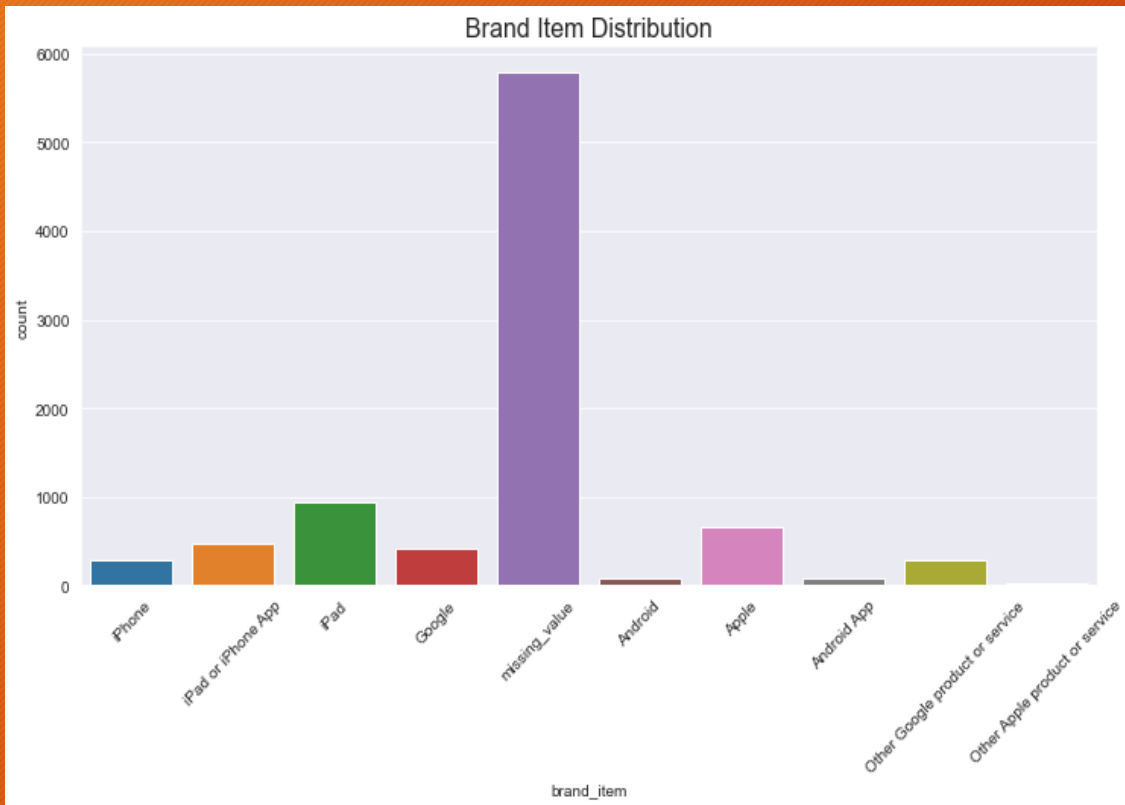
- The dataset used for the analysis includes various columns such as tweet text, and it is crucial to clean and preprocess this data. Preprocessing steps incorporate converting text to lower case, removing URLs and special characters, tokenizing the text into words, and lemmatization to reduce words to their root forms. The document emphasizes the importance of feature engineering, particularly using techniques like TF-IDF vectorization to represent the textual data numerically.

Visualizations : Sentiment Comparison



- Positive Sentiment: Apple has a significantly higher count than Google, showing a stronger positive reaction to Apple products.
- Negative Sentiment: More balanced between both brands, but Apple still has a slightly higher count than Google.
- Neutral Sentiment: Both brands have low counts in neutral sentiment. Apple has slightly more neutral mentions than Google.
- Conclusion: Apple products generate more engagement, especially in positive sentiment, compared to Google products.

Visualization: Brand Distribution



- The bar chart shows a clear trend in user sentiment, where iPad takes the lead as the most mentioned product in tweets. This is followed by other Apple products, such as the iPhone and general Apple mentions. On the other hand, Google products also receive a fair amount of attention, though Android-related products are mentioned less frequently. This pattern emphasizes the strong presence and dominance of Apple products in user conversations, suggesting that Apple products are central to user sentiment and discussions on social platforms.

Modelling

- Different machine learning models, including Logistic Regression and Random Forest, are trained to classify sentiments. Each model's performance is evaluated based on metrics such as accuracy, precision, recall, and F1-score. The initial results show a general accuracy of around 67%, with random forest slightly outperforming logistic regression. Efforts are made to improve performance by using techniques like class weighting and SMOTE (Synthetic Minority Over-sampling Technique) to tackle class imbalances in sentiment distribution.

Results and Visualizations

- Word clouds are created to visualize words associated with different sentiment categories, revealing that terms like “iPhone” and “Google” frequently appear in both positive and negative contexts. The analysis reveals that Apple products generally receive more favorable sentiments compared to Google. Additionally, visualizations help compare sentiment distributions across the two brands, showcasing that Apple generates higher engagement levels among users.

Conclusion

- This encapsulates the necessity of sentiment analysis in contemporary digital marketing, demonstrating how machine learning can automate understanding customer feedback derived from social media platforms, thus enhancing business decision-making processes.