

NLP_TwitterSentiment Analysis

Uncovering insights on social media engagement.

Agenda

- Introduction
- Data Sources
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Sentiment Analysis Methodology
- Results - Sentiment Distribution
- Keyword Analysis
- Topic Modeling
- Brand Competitor Analysis
- Customer Support Feedback
- Recommendations
- Conclusion
- Future Work
- Q&A
- Thank You
- References
- Contact Information



● Introduction

This repository focuses on conducting sentiment analysis on user-generated content related to Android and Apple products. Dive into the world of consumer opinions, feedback, and reviews to gain valuable insights into the perception and sentiment surrounding these two tech giants using Machine learning models.

Objectives

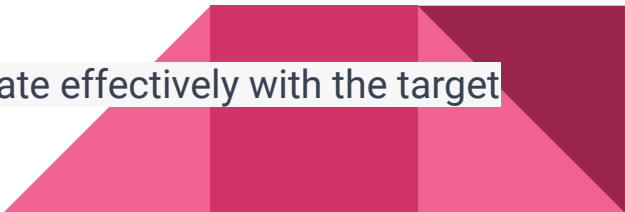
Competitor Analysis – Compare sentiment and customer perceptions against competitors.

Product Launch Evaluation – Assess the success of recent product launches.

Customer Support Enhancement – Improve customer support through feedback analysis.

Influencer Identification – Identify potential brand influencers.

Brand Impact Maximization – Maximize the brand's impact and resonate effectively with the target audience.

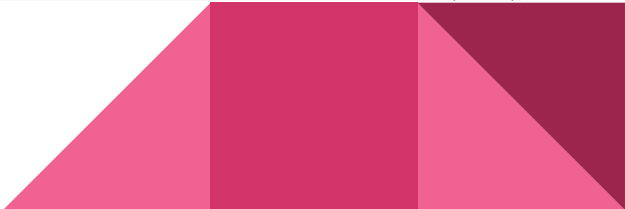


- Data Sources

The data was sourced from a page called Zillow and is stored in a dataset named 'judge-1377884607_tweet_product_company.' This dataset contains Twitter data related to product companies. The structure of the data consist of various columns, including 'tweet_text,' which contains the text of the tweets, 'product' or 'company' columns indicating the product or company mentioned in the tweet, and potentially other columns like 'sentiment' or 'timestamp.' These columns collectively provide information about Twitter users' opinions and discussions regarding specific products and companies.

● Data Preprocessing & Exploratory Data Analysis (EDA)

In the realm of data analysis, the process of data preparation and text preprocessing plays a pivotal role in ensuring that the data is primed and ready for in-depth analysis. This vital step involves several crucial tasks:

- Handling Missing Data: We begin by addressing missing data points, which are gaps in the dataset. This may involve filling in missing values or making informed decisions on how to handle them.
 - Removing Duplicates: Duplicate entries can skew analysis results. Identifying and eliminating duplicate records ensures that our data is accurate and unbiased.
 - Data Transformation: This step involves converting data into a format that is suitable for analysis. It often includes data type conversions and reformatting.
 - Tokenization: In the realm of text data, tokenization comes into play. It involves breaking down text into individual words, phrases, or tokens. This step is fundamental for further text analysis.
 - Stopword Removal: Stopwords are common words (e.g., "the," "and," "is") that don't carry significant meaning in text analysis. Removing them helps in focusing on the meaningful content.
 - Lemmatization: Rather than aggressive word reduction as in stemming, lemmatization involves reducing words to their base or root form. This preserves more accurate word meanings and context.
 - Text Preprocessing for NLP: These text preprocessing techniques are particularly vital for Natural Language Processing (NLP) tasks like sentiment analysis, topic modeling, and language understanding.
- 

● Sentiment Analysis Methodology

- We employed a range of machine learning models for sentiment analysis, catering to diverse complexities.
- Baseline models like Logistic Regression and Random Forest provided initial insights.
- Advanced models such as Recurrent Neural Networks (RNN) and BERT-based models enhanced our analysis, considering nuances in language and context.
- The performance of our models was rigorously evaluated using various metrics, including accuracy, precision, recall, F1-score, and confusion matrices.
- This step ensured the reliability and effectiveness of our sentiment analysis models.
- The culmination of our analysis yields actionable insights into consumer sentiment and perceptions regarding Android and Apple products.
- We saw Apple had more engagement than Android.



Topic Modelling

It's an unsupervised learning technique, meaning it doesn't require prior labeling of documents. Instead, it clusters or groups documents into topics based on similarities in their content.

I used Latent Dirichlet Allocation(LDA) which assumes that documents are mixtures of topics.

- Topics Discovered: Five distinct topics emerged, reflecting the most discussed themes.

Topic 1: A blend of mentions about SXSW, Apple products like iPad and iPhone, Google, free apps, and more.

Topic 2: Focuses on SXSW and Google, discussing links, quotes, social launches, and new circles.

Topic 3: Revolves around SXSW, Apple, and Austin, with mentions of the Apple Store, iPad, pop culture, and iPhone.

Topic 4: Explores SXSW, links, mentions of Android and iPhone, Google, apps, and getting something.

Topic 5: Highlights SXSW and Google, emphasizing quotes, Marissa Mayer, maps, locations, and mobile.



Topic Modelling Results

Topic 1: Diverse Discussions

- This topic is a blend of mentions about SXSW (South by Southwest), Apple products such as iPad and iPhone, Google, discussions about free apps, and more. It encompasses a wide range of tech-related discussions.

Topic 2: SXSW and Google Focus

- This topic primarily focuses on SXSW and Google. It includes discussions about links, quotes, social launches, and new circles, indicating conversations related to the event and Google's involvement.

Topic 3: SXSW, Apple, and Austin

- Revolving around SXSW, Apple, and the city of Austin, this topic mentions the Apple Store, iPad, pop culture, and iPhone. It highlights the convergence of these elements during the event.

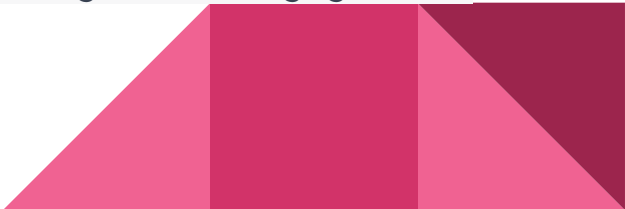
Topic 4: Exploration of Tech

- This topic explores various aspects of technology at SXSW. It includes mentions of SXSW, links, discussions about Android and iPhone, Google, apps, and references to "getting something" related to tech.

Topic 5: Google and SXSW Highlights

- Topic 5 emphasizes SXSW and Google. It highlights quotes, mentions of Marissa Mayer, discussions about maps, locations, and mobile technology. This topic seems to center on Google's activities during the event.

● Results

- **Brand Competitor Analysis** : This section evaluates sentiment and customer perceptions of our brand compared to competitors. It identifies strengths and weaknesses to gain a competitive edge. We found that there was negative emotion towards Apple products due to its weaknesses such as low battery life.
 - **Customer Support Feedback** : Analyzing negative sentiments in customer support-related tweets helps improve services and address pain points. Although there were more positive emotion than negative we found out that Apple has more negative sentiments than Android due to the weaknesses in their products as highlighted above.
 - **Influencer identification** : Through out analysis we uncovered accounts such as @Wesley98 and @Marissa Mayer as potential brand ambassadors due to their high tweet engagement.
- 

● Recommendations

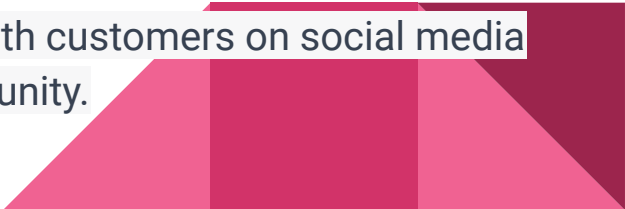
Address Battery Life Issues: Prioritize improving areas where Apple products have received negative sentiment, such as battery life. Investing in product enhancements can lead to a more positive perception.

Enhance Customer Support: While there are more positive sentiments, it's essential to continue focusing on customer support. Resolve issues promptly, offer solutions, and ensure customer satisfaction to maintain positive sentiment.

Competitive Benchmarking: Continuously monitor and benchmark against competitors to identify areas where our brand can outperform. Leverage strengths to gain a competitive edge.

Product Innovation: Invest in research and development to introduce innovative features and improvements that resonate positively with customers.

Feedback Loop: Establish a feedback loop by actively engaging with customers on social media to address concerns, gather insights, and foster a sense of community.



Conclusion:

In conclusion, our comprehensive sentiment analysis has shed light on the intricate landscape of consumer sentiment regarding Android and Apple products. Through this analysis, we have unearthed valuable insights into the strengths and weaknesses of both brands. While Android enjoys praise for its versatility and affordability, Apple exhibits strengths in design and user experience. Nevertheless, it is equally crucial to acknowledge areas for improvement. Negative sentiment towards Apple's battery life and Android's fragmentation serves as a clarion call for both brands to address these concerns. Moreover, the competitive analysis underscores the importance of seizing opportunities to gain a competitive edge in an ever-evolving market. Our findings not only offer actionable strategies for brand enhancement but also emphasize the significance of attentive listening to the voice of the consumer in shaping the future of technology.



Future Work:

Future endeavors will focus on refining our sentiment analysis models, delving into real-time analysis, and extending our reach to other social media platforms. The enhancement of customer support services based on invaluable feedback and continued product differentiation remains pivotal in our pursuit of excellence.

Q&A: We eagerly anticipate your questions and insights, encouraging an open dialogue to further explore the nuances of our analysis and its implications.



Contact Information:

For inquiries, collaboration opportunities, or further discussions, please do not hesitate to reach out to us via researchinstitutekenya@gmail.com. We look forward to connecting with you and continuing our journey towards innovation and consumer-centric technology.

Thank you.

