TWITTER SENTIMENT ANALYSIS REPORT

PROJECT OVERVIEW

Background Information

Sentiment analysis is an NLP technique that helps determine the emotional tone of text, categorizing it as positive, negative, or neutral (Pang & Lee, 2008). It's commonly used to analyze social media posts and product reviews, providing businesses with insights into customer feelings.

Twitter is a key platform where people share their opinions on products and services. By analyzing tweets, companies like Google and Apple can track public sentiment in real time, using this data to improve their products and boost customer satisfaction (Suh et al., 2010).

Challenges in Sentiment Analysis

- Tweets are full of slang, and abbreviations, making them hard to interpret (Riloff et al., 2013).
- Sarcasm can make it harder to classify sentiments accurately (Davidov et al., 2010).
- Irrelevant or spam tweets can add noise and affect the analysis (Ravi & Ravi, 2015).

Stakeholders'

- Tech Companies**:** Google and Apple, as they benefit from insights into public perception of their products.
- Product Teams: Use sentiment analysis to identify areas for product improvement.
- Marketers: Leverage sentiment trends to craft targeted marketing campaigns..

Proposed Solution

Analyzing 9,000 tweets related to Google and Apple products, classification of the sentiments as positive, negative, or neutral. This will allow us to compare user perceptions of both brands and identify which aspects of their products are what people love or criticize.

Projected Conclusion

This analysis will help companies understand public opinion, identify areas for improvement, and discover what users value most, enabling better product development and marketing strategies.

Problem Statement

Developing a robust sentiment analysis model that accurately identifies both positive and negative customer sentiments about Apple and Google products from Twitter data, with particular emphasis on improving the detection of negative feedback..

Objective

Primary Objective

- To create a machine learning model capable of accurately determining the sentiment of a tweet (positive, negative, or neutral) based on its content.

Secondary Objectives

- **Improve Negative Sentiment Detection:** Focus on enhancing recall for better detection of negative feedback.
- **Actionable Insights for Stakeholders:** Provide clear, interpretable results to marketing, customer service, and leadership for timely decision-making.
- **Regular Tracking:** Monitor sentiment trends and market share of voice for Apple and Google, providing concise brand comparison reports to track changes in consumer perception.

Metrics of Success

- **Accuracy:** Ensure the model correctly identifies the sentiment (positive, negative, or neutral) in at least above 90% of tweets, showing how reliable it is overall.
- **Precision:** Make sure that when the model labels a tweet as positive or negative, it's correct over 85% of the time, avoiding false positives.
- **Recall:** Focus on capturing as many real negative tweets as possible, reducing the risk of missing critical negative feedback.
- **F1 Score:** Measure how well the model balances accuracy and completeness. A high F1 score means the model is good at correctly identifying sentiments while catching most relevant tweets.
- **Sentiment Class Balance:** Ensure the model performs equally well across all sentiment types (positive, negative, and neutral), without favoring one over the others.

DATA UNDERSTANDING

The dataset is sourced from Crowd Flower via Data.World and it contains 9093 Tweets that have been meticulously annotated by human raters with sentiment labels: positive, negative, or neutral. The data was collected during March, around the time of SXSW event and tweets were filtered by using specific hash tags or keywords. The dataset has 3 columns:

tweet_text: The actual text of the tweet, which provides insights into what users are saying about Apple and Google products.

emotion_in_tweet_is_directed_at: The specific product or brand mentioned in the tweet (e.g., iPhone, iPad, Google).

is_there_an_emotion_directed_at_a_brand_or_product: Indicates whether the tweet expresses: No emotion toward brand or product, Positive emotion, Negative emotion, I can't tell.

DATA PREPARATION

Data Cleaning

To prepare the dataset for NLP, a thorough cleaning process was conducted:

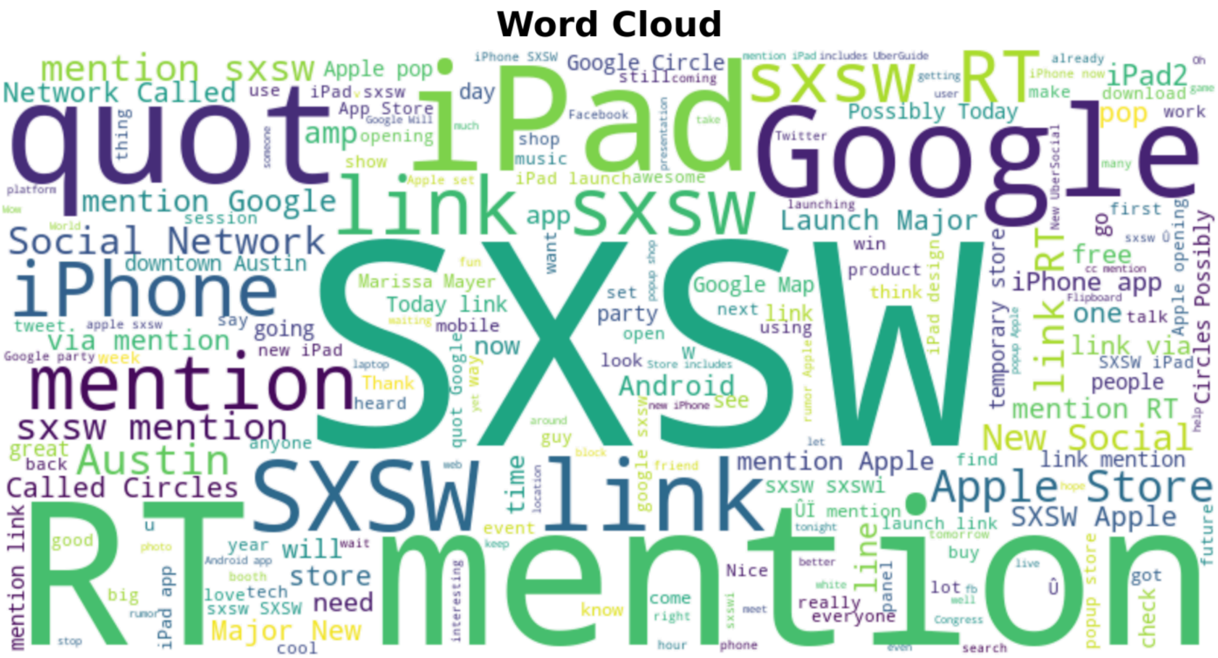
- **Column Renaming**: Columns were renamed for easier readability.
- **Handling Missing/Null Values**:
  - A single NaN value in the text column was removed to maintain data quality.
  - The 'Item' column, with over 60% missing data, was addressed by replacing missing values with 'Unknown,' ensuring context retention.
- **Category Merging**: Rows labeled as "I can't tell" in the emotion column were merged into the 'No Emotion' category, as they lacked clear emotional responses.
- **Duplicate Removal**: Duplicate rows were removed to eliminate redundancy and maintain dataset accuracy.
- Words in the text column were replaced with corresponding brand names (e.g., Google, Apple) to preserve context while filling gaps.
- **Text Preprocessing**:
  - Converted all text to lowercase for consistency.
  - Removed special characters, numbers, and non-alphabetic content.
  - Tokenized text into individual words.
  - Removed stopwords using the NLTK library.
  - Applied lemmatization to reduce words to their root forms.
  - Joined the cleaned tokens back into cohesive strings.

These steps helped create a clean and reliable dataset for NLP tasks.

DATA ANALYSIS

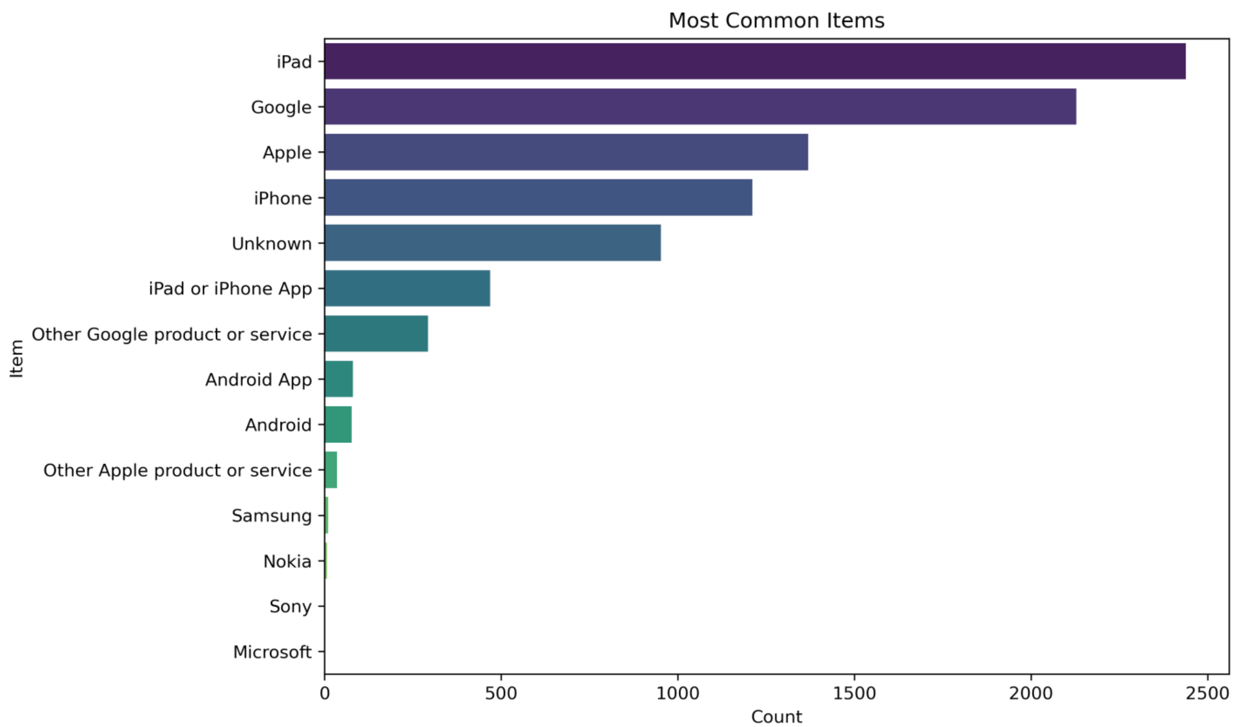
UNIVARIATE ANALYSIS

- **Most Frequent Words**



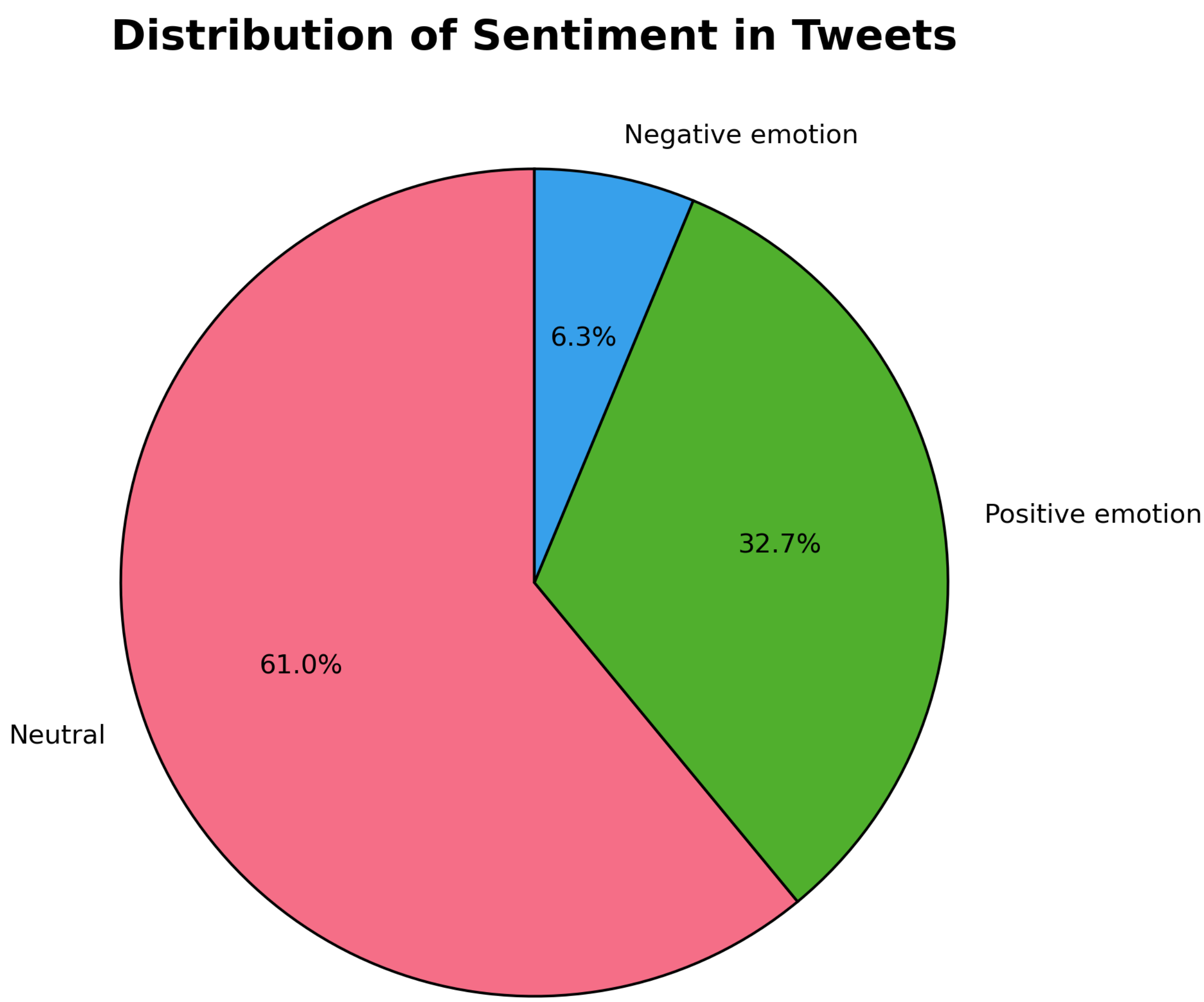
The word cloud graph indicates the most frequently discussed topics related to the SXSW event on Twitter, highlighting key terms such as “SXSW”, ”Google”, and “iPhone”. This suggests a focus on technology and social media , which can be further analyzed for sentiment trends and public perception during the event.

- **Most Frequent Item**



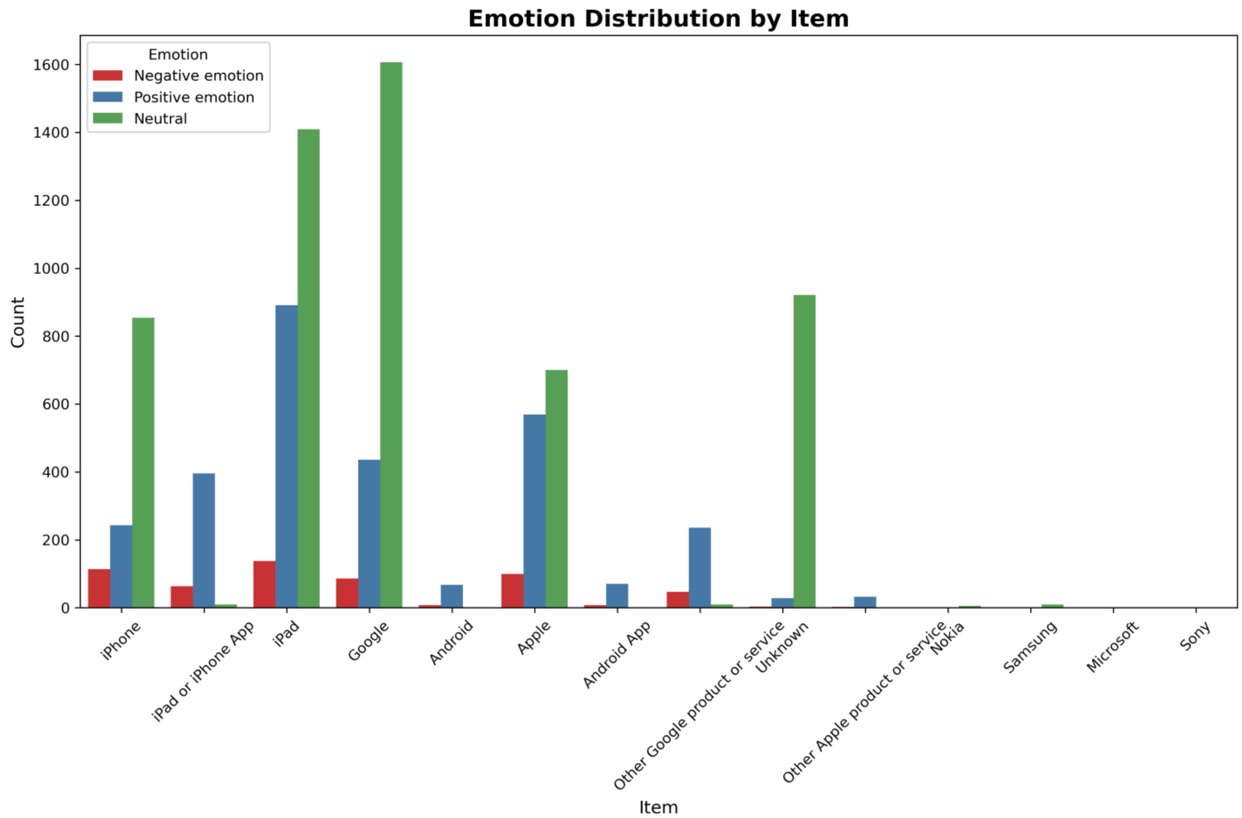
The frequency distribution shows that **iPad** and **Apple** dominate mentions in the 9000-tweet sentiment analysis dataset, with 900 and 650 occurrences respectively, followed by **iPad/iPhone Apps** (~450) and **Google** (~420). Apple products significantly outpace mentions of Google and Android, highlighting their **prevalence** in discussions.

- **Distribution of Sentiments**



The sentiment column is highly imbalanced with neutral emotion being the highest followed by positive then negative .This will require further preprocessing in order to increase model performance.

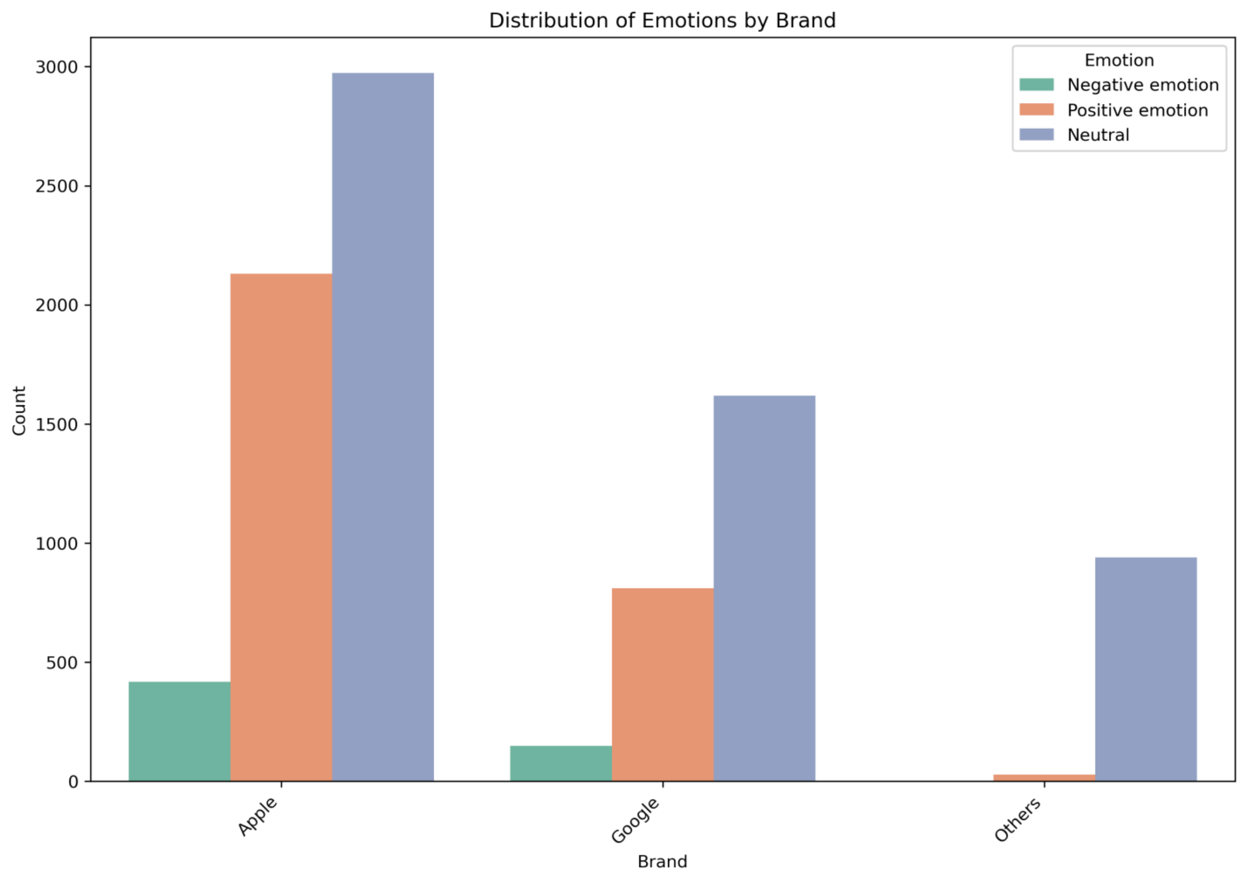
BIVARIATE ANALYSIS



The **iPad** generated the most **positive sentiment**, likely due to the excitement surrounding its launch at the **SXSW event**. Negative sentiment was mostly about design issues. For **Google**, most tweets were **neutral**, focused on **Google Circle**, their new social media app. Some people were excited, but many were unsure, which explains the neutral tone. This sentiment distribution reflects how product events and launches shaped opinions during that time.

To streamline our analysis, we grouped Apple products and Google products together under their respective brand categories. This approach aligns with our focus on evaluating the overall performance and sentiment surrounding the Apple and Google brands.

- **Distribution of Sentiments Based on Apple,Google and Others**



The Emotion Distribution by Item graph supports our insights: the **iPad** leads with **positive sentiment**, likely due to the excitement surrounding the **SXSW event**, while it also has the most **neutral sentiment** because of event mentions. **Apple** overall shows strong positive sentiment, influenced by the iPad's success. For **Google**, most tweets are **neutral**, focusing on **Google Circle** with mixed reactions — some excitement, but many uncertain, which explains the neutral sentiment. Other brands, grouped under "others," show less clear sentiment trends.

MODELLING

Baseline Model

In this section, usage of the TF-IDF vectorizer found within the Sci-kit Learn package to prepare the data for modeling, a predefined model known as text blob was used for sentiment analysis evaluation of the performance of several machine learning models to predict the sentiment of tweets. The following models were selected for the task:

1. **Naive Bayes**: A simple probabilistic classifier commonly used for text data.
2. **Logistic Regression**: A strong baseline model for binary and multi-class classification tasks.

3. **Support Vector Machine (SVM)**: A powerful model that works well in high-dimensional spaces, making it suitable for text classification.
4. **Random Forest**: An ensemble learning model that improves prediction accuracy by combining the results of multiple decision trees.

The code iterates over these models, training each one on the training data , and then evaluating them on the test data. For each model, we calculate the **accuracy** and generate a **classification report**. The classification report provides detailed metrics, including precision, recall, and F1-score for each sentiment class (positive, negative, neutral).By comparing the results of the models, SVM performed better. With an;

SVM Accuracy: 0.93
Classification Report for SVM:

	precision	recall	f1-score	support
Negative	0.90	0.61	0.72	170
Neutral	0.89	0.98	0.94	770
Positive	0.96	0.94	0.95	874
accuracy			0.93	1814
macro avg	0.92	0.84	0.87	1814
weighted avg	0.93	0.93	0.92	1814

The SVM model has an overall accuracy of 93%. It performs well on the neutral class (0), with high recall (98%) and precision (89%). However, it struggles with the negative class (-1), showing poor recall (61%) and precision (90%). The positive class (1) is moderately predicted, with recall (94%) and precision (96%). The weighted averages reflect better performance due to the dominance of the neutral class. To improve, focus on better handling of the negative class.

Dealing with Class Imbalance

In order to address the imbalanced dataset, the SVM model has a parameter called class weight, which can be set to 'balanced' to automatically adjust the weights for each class.

Hyper parameter Tuning for SVM:

To optimize the SVM model, hyper parameter tuning was performed using **GridSearchCV** with 5-fold cross-validation. The parameters tuned include:

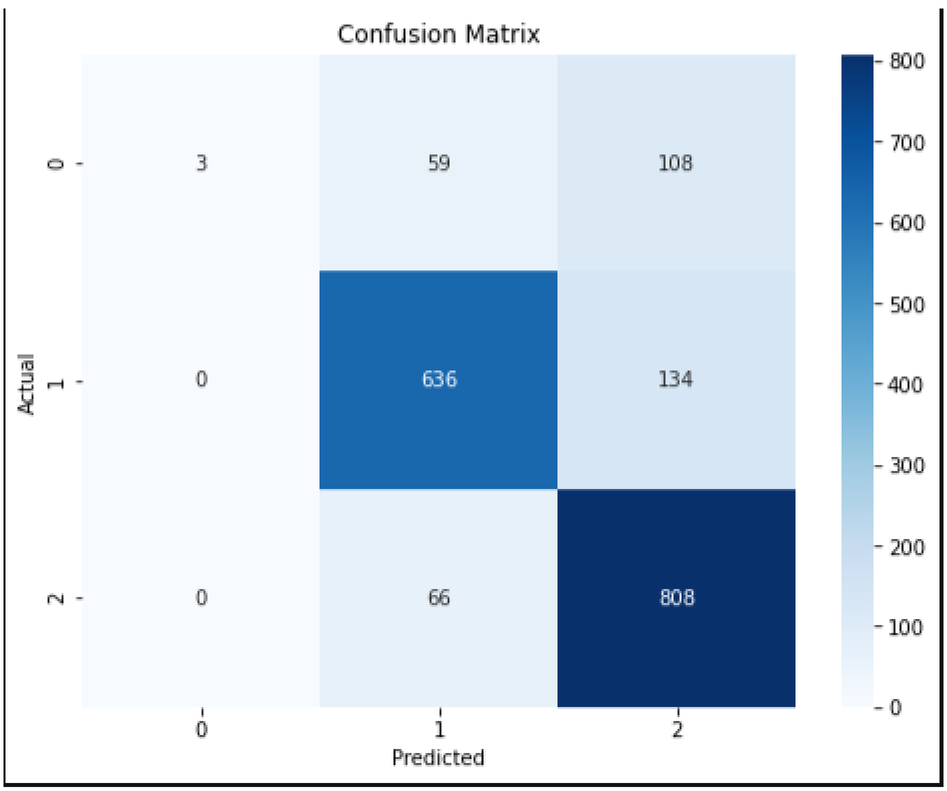
- **C** (Regularization parameter)
- **Kernel** (linear and RBF)
- **Gamma** (scale and auto for the RBF kernel)

The best hyperparameters found were:

C = 1, **Kernel** = 'linear', **Gamma** = 'scale', yielding the optimal performance for predicting class labels in our imbalanced dataset. This improved the model accuracy score. SVM was chosen because of its nature of handling non-linearity.

	precision	recall	f1-score	support
Negative	0.81	0.75	0.78	170
Neutral	0.93	0.98	0.95	770
Positive	0.97	0.94	0.95	874
accuracy			0.94	1814
macro avg	0.90	0.89	0.89	1814
weighted avg	0.94	0.94	0.94	1814

The tuned model achieved 94% accuracy, outperforming the baseline SVM model (93%). While the baseline SVM had higher precision for the Negative class (0.90 vs. 0.81), its recall was significantly lower (0.61 vs. 0.75), leading to a reduced F1-score. Both models performed similarly well for Neutral and Positive classes. The tuned model demonstrated better overall performance, especially in handling true negatives."



After performing stratified K-fold, the model correctly classified 636 instances, but 134 instances were misclassified as Class 2.

The model also correctly classified 808 instances, with only 66 misclassified as Class 1.

Recommendation

- Google's Neutral Sentiment: A significant proportion of neutral sentiment toward Google products suggests room to build excitement through more engaging marketing strategies.
- Recurring Issues for Google: Neutral or negative sentiment tweets for Google often mention usability issues. Develop targeted support initiatives or FAQs addressing these concerns to preemptively mitigate dissatisfaction.
- Apple's Opportunities: While Apple has strong positive sentiment, monitor any emerging concerns to maintain its leading market position.

Conclusion

Our analysis highlights valuable opportunities for both Apple and Google. Apple can build on its strong positive sentiment to reinforce its market leadership while staying ahead of emerging concerns. Meanwhile, Google has the chance to close sentiment gaps—especially around neutral and usability-related feedback—by implementing more engaging marketing and targeted support initiatives. The SVM model has demonstrated strong performance in detecting neutral and positive sentiments, making it the most reliable choice for immediate application to drive meaningful improvements.

Next step/ Model Deployment

Deploy the model with a robust monitoring system to continuously track key metrics, such as accuracy and recall for the Negative class, ensuring sustained performance. Implement alerts to quickly identify and address any significant drifts or anomalies in predictions.

Reference

- Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). *Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network*. Proceedings of the 2010 International Conference on Weblogs and Social Media, 177-184.
- Riloff, E., et al. (2013). *Hotel Review Dataset: A Benchmark for Sentiment Analysis*. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1109-1119.
- Davidov, D., Tsur, O., & Rappoport, A. (2010). *Enhanced Sentiment Learning Using Twitter Hashtags and Emoticons*. Proceedings of the 23rd International Conference on Computational Linguistics, 241-249.
- Ravi, K., & Ravi, V. (2015). *A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches, and Applications*. Knowledge-Based Systems, 89, 14-46.
- Gao, H., Tang, J., & Liu, H. (2015). *Modeling Content and Connections in Bipartite Networks*. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 90-98.