Phase 4 project

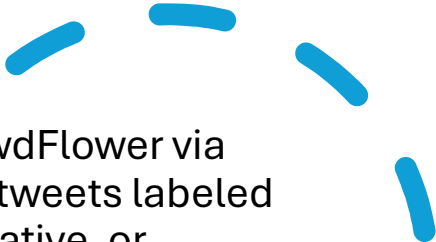Presented by Faith Karaya

Table of contents

- Problem statement
- Data understanding
- Business understanding
- Exploratory Data Analysis
- Conclusion

Problem statement

- Financial markets react strongly to public sentiment, which is influenced by economic trends, market conditions, and unexpected events, yet traditional sentiment analysis often struggles to capture these rapid shifts. Social media platforms like Twitter provide a vast and real-time data source, allowing Machine Learning to efficiently analyze public opinion on companies like Google and Apple without the biases of manual analysis. This project aims to develop and deploy Machine Learning models to extract insights from customer feedback, helping investors, traders, and businesses make informed, data-driven decisions

Data Understanding

- The dataset, obtained from CrowdFlower via data.world, contains over 9,000 tweets labeled by human raters as positive, negative, or neutral. Collected during the 2013 South by Southwest conference, these tweets primarily discuss Google and Apple products, capturing real-time consumer sentiment in a tech-focused environment. To analyze the impact of positive sentiment on sales and investment returns, the target variable was simplified into two categories: tweets expressing positive sentiment and those that do not (including neutral, negative, or unclear sentiments).

Business understanding

- Twitter is a key platform for users to share opinions and experiences about products, making it a valuable source for sentiment analysis. Understanding public sentiment helps businesses strengthen their brand image and identify areas for improvement. In investment and trading, incorporating sentiment analysis alongside financial metrics offers unique insights, as positive public perception can boost investor confidence and enhance stock performance.

Exploratory Data Analysis

Sentiment Distribution by Brand

Number of Tweets per Sentiment
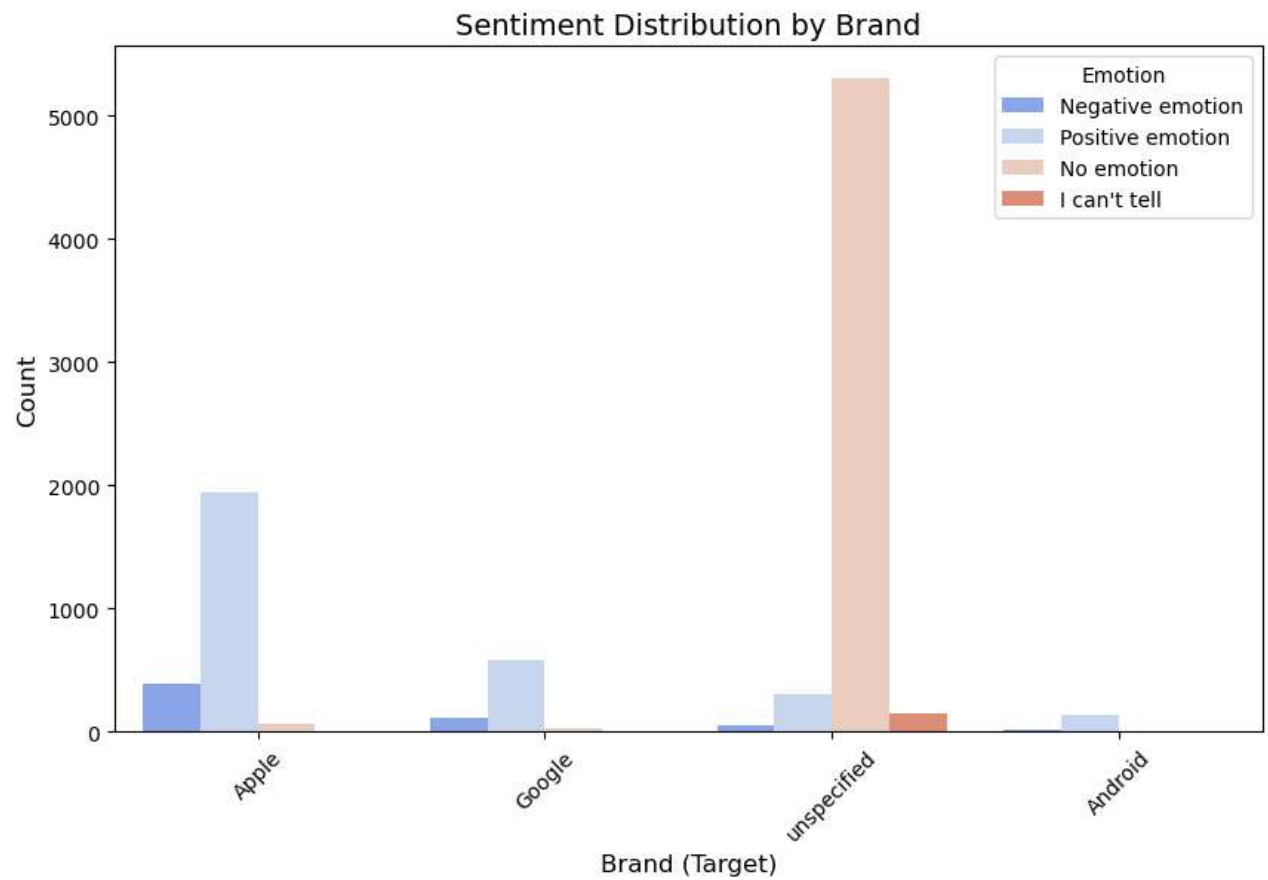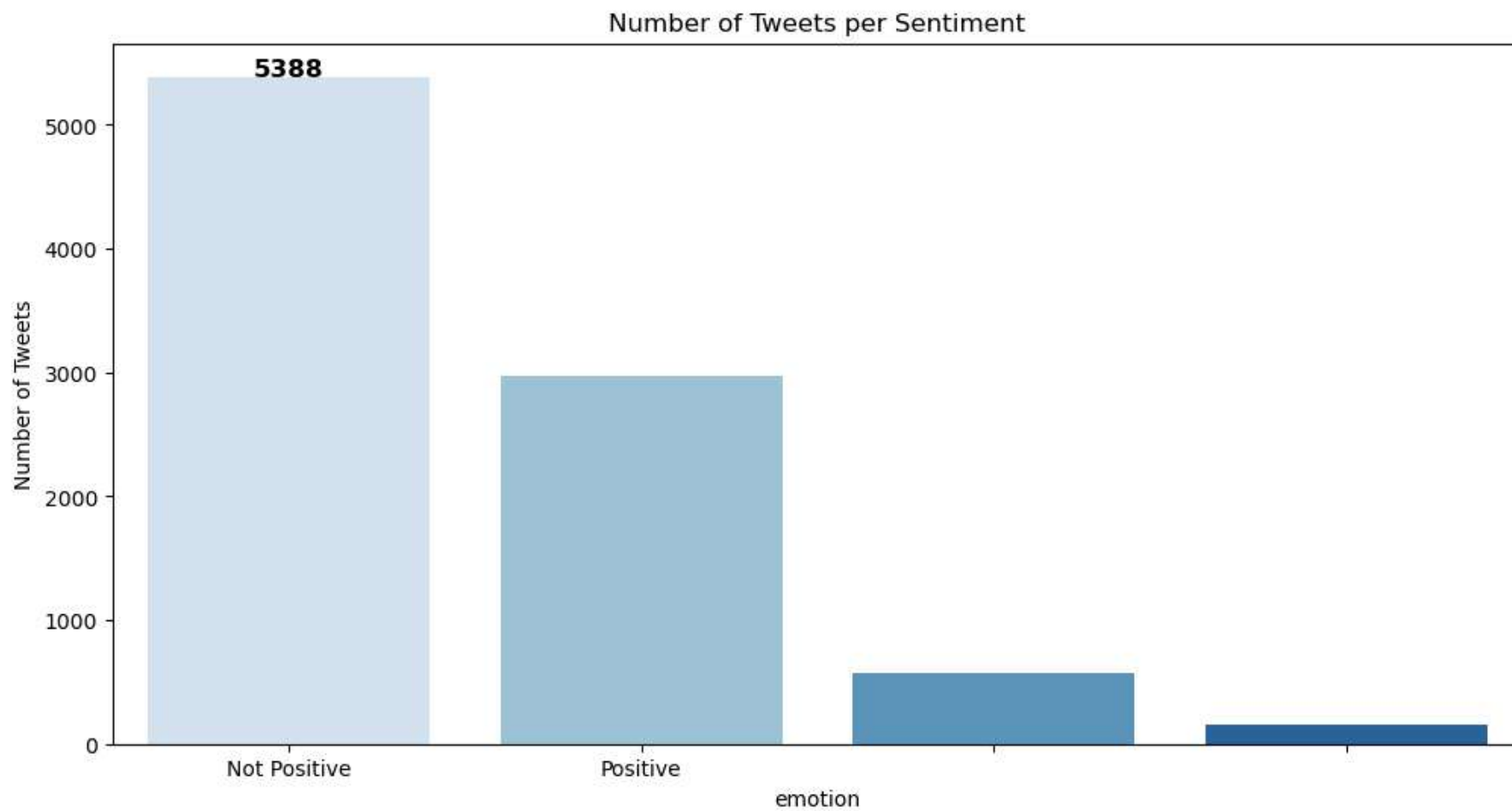
Conclusion

- Random Forest classifier helped improve the performance of the model ,achieving a n accuracy score of 66% on the training set where the positive emotion was able to get a precision of 64.