APPLE TWITTER SENTIMENT ANALYSIS

PROJECT OVERVIEW

- This project uses tweets about Apple and Google products rated by humans for sentiments to train and evaluate the NLP model.

- The company use Sentiment analysis to analyze customer feedback and make data driven decision.

BUSINESS PROBLEM/DATA UNDERSTANDING

- It is crucial for companies to understand customer sentiments and enhance their satisfaction.

- Apple and Google company is interested in understanding the customer sentiments so as to improve their products and enhance customer satisfaction .

- The dataset was sourced from CrowdFlower via data.world.

- The link for Kaggle is https://www.kaggle.com/datasets/slythe/apple-twitter-sentiment-crowdflower

- There are 3886 Records and 12 Features.

CONT'D

- The key Features are :

- 1. sentiment - sentiment labels

- 2. sentiment:confidence - confidence score for sentiment

- 3. date- when tweet was posted

- 4. query- search query used

- 5. text- actual tweet

# EXPLORATORY DATA ANALYSIS

Distribution of Sentiment classes
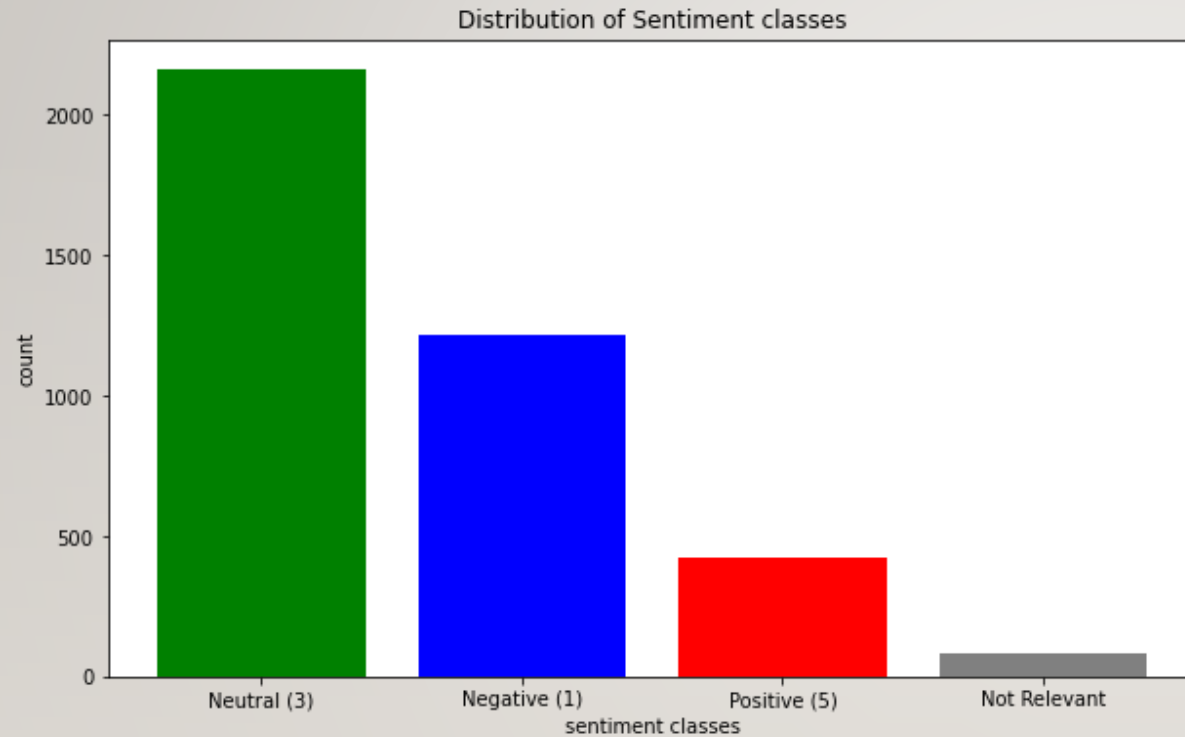
- Neutral(3) had the highest sentiment,followed by Negative sentiment(1),then Positive(5) and the non_relevant which had the smallest.

DATA CLEANING AND TEXT PREPROCESSING

- Dropped 2 columns which had missing values and also irrelevant columns and was left with sentiment and text.

- Text preprocessing to remove URLs,Hashtags,mentions,special characters which helped to reduce noise and focus on the actual content of the tweet .

- Converting text to lowercase to reduce vocabulary size and improve word frequency analysis.

- Stopwords were removed and Lemmatization together with stemming applied to reduce word to their base form.
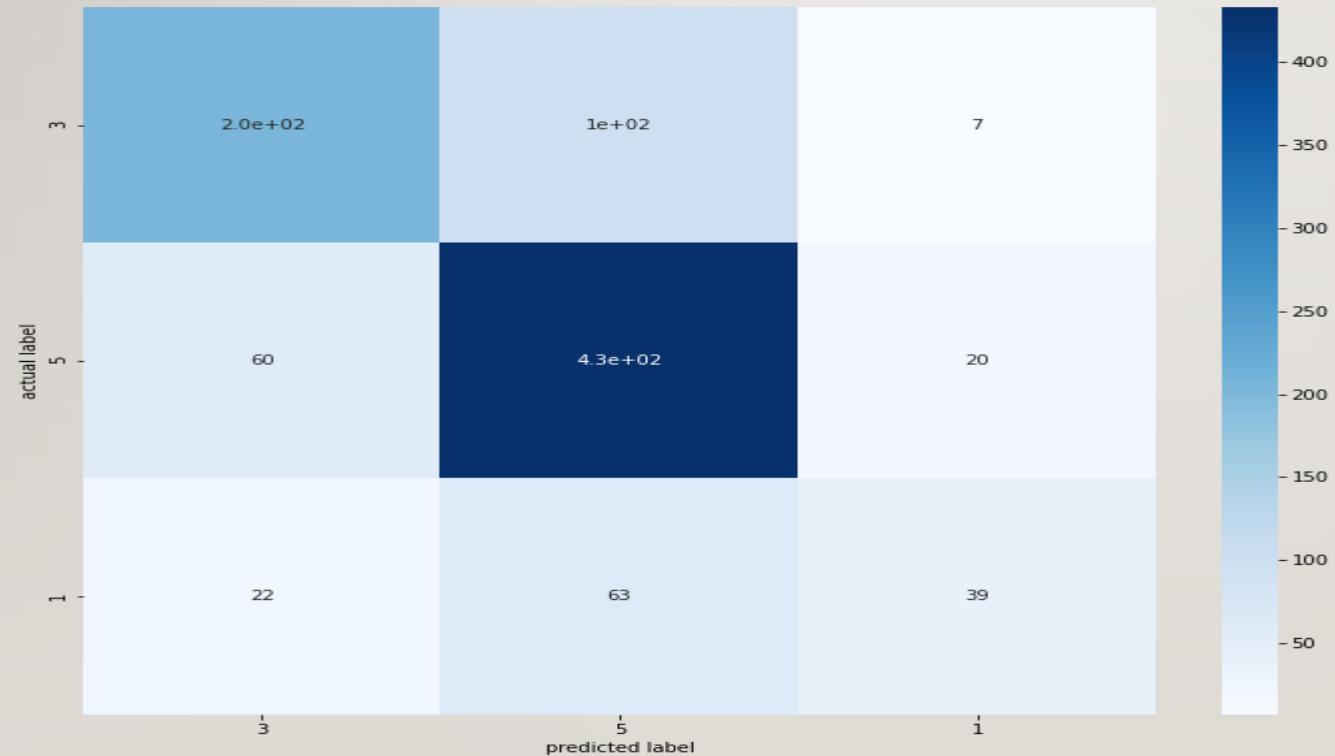
MODELING
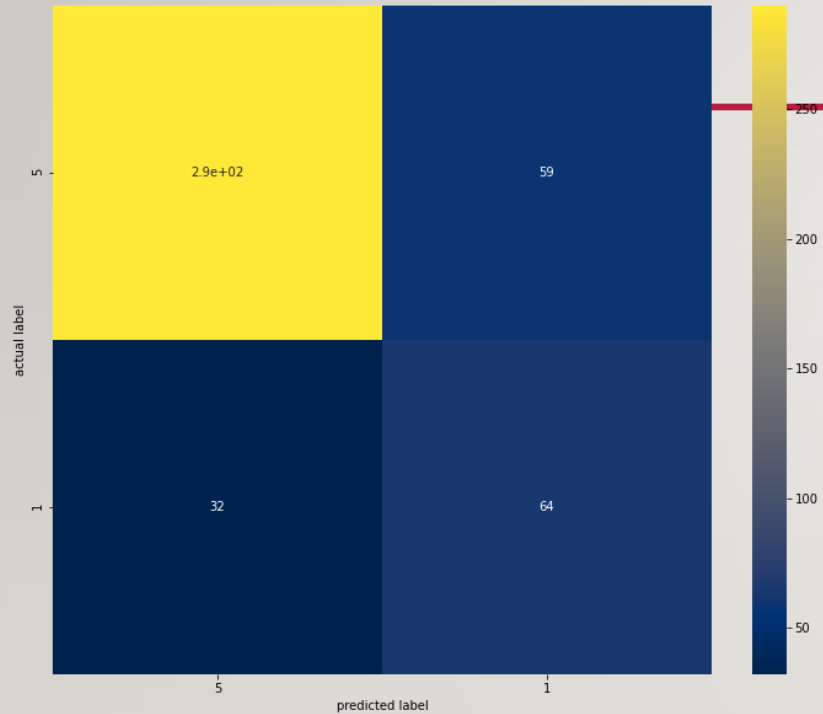
- Performed Binary classification first using Logistic Regression.
- Expanded to multiclass classification
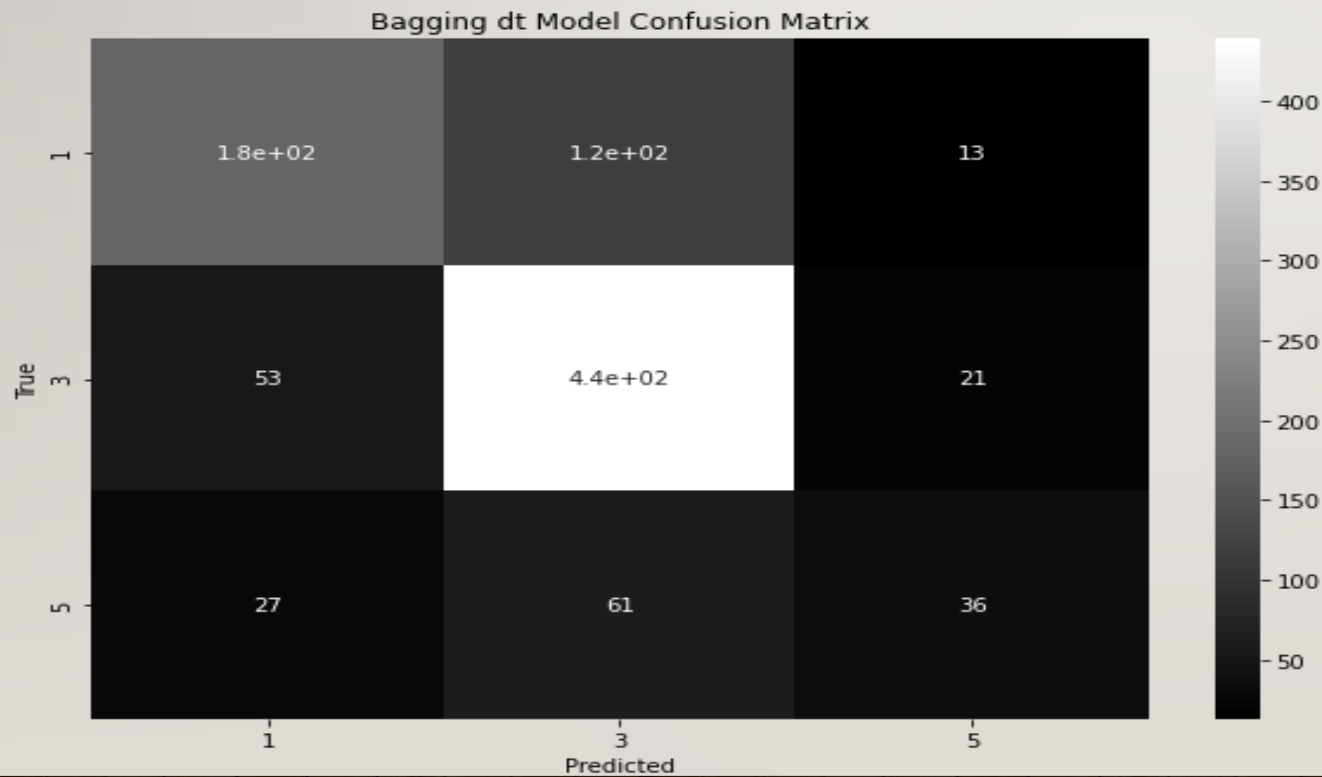- Binary classification had an accuracy of 80% and precision of 90%.

LOGISTIC REGRESSION WITH GRIDSEARCH

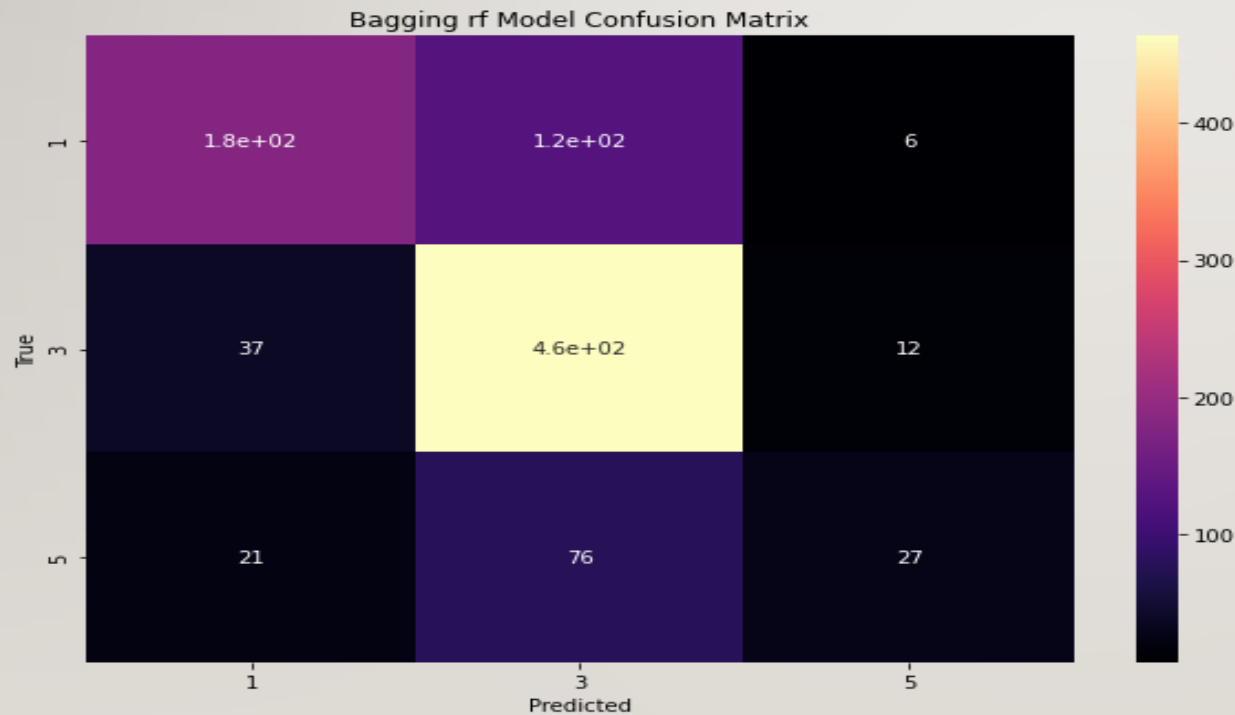- Attained an accuracy of 71% and Precision of 68%.

BAGGING WITH DECISION TREE CLASSIFIER

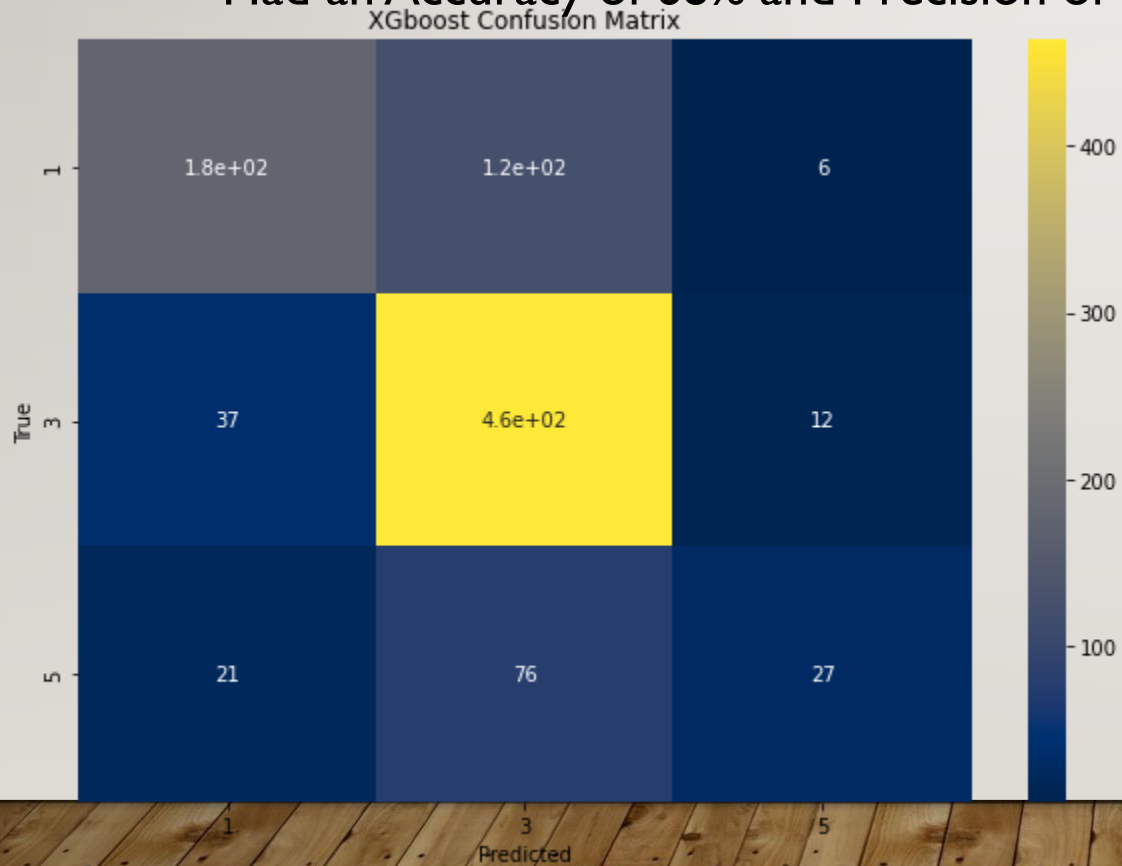- Had an accuracy of 69% and Precision of 64%.

BAGGING WITH RANDOM FOREST CLASSIFIER

- Had an Accuracy of 71% and Precision of 69%.

XGBOOST

- Had an Accuracy of 68% and Precision of 63%.

CONCLUSIONS

- Both Logistic Regression with grid search and Bagged Random Forest performed the best with Accuracy of 71%.

- The Logistic Regression with gridsearch achieved an accuracy of 0.713 and precision of 0.677 while Bagged Random Forest with accuracy of 0.708 and precision of 0.685.

ACHIEVEMENT OF OBJECTIVE

- Insights gained can help business better understand customer sentiment and make Data driven decision to enhance products and service.

RECOMMENDATIONS

By doing the following,the model performance will improve:

➢ Experimenting with advanced feature engineering,

➢ Experiment further with different other models

➢ Implement further ensemble methods

➢ Perform extensive Hyper parameter Tuning

NEXT STEPS

- Collecting additional data to improve the model.

Thank you

Any questions?
© Presentation-Process.com