U.S. Airline Sentiment Analysis Using Tweets

Arthik Alexander

Reha Patel

Siddhesh Koparde

Vinit Nagap

Our Team

Arthik Alexander

Reha Patel

Siddhesh Koparde

Vinit Nagap

Problem

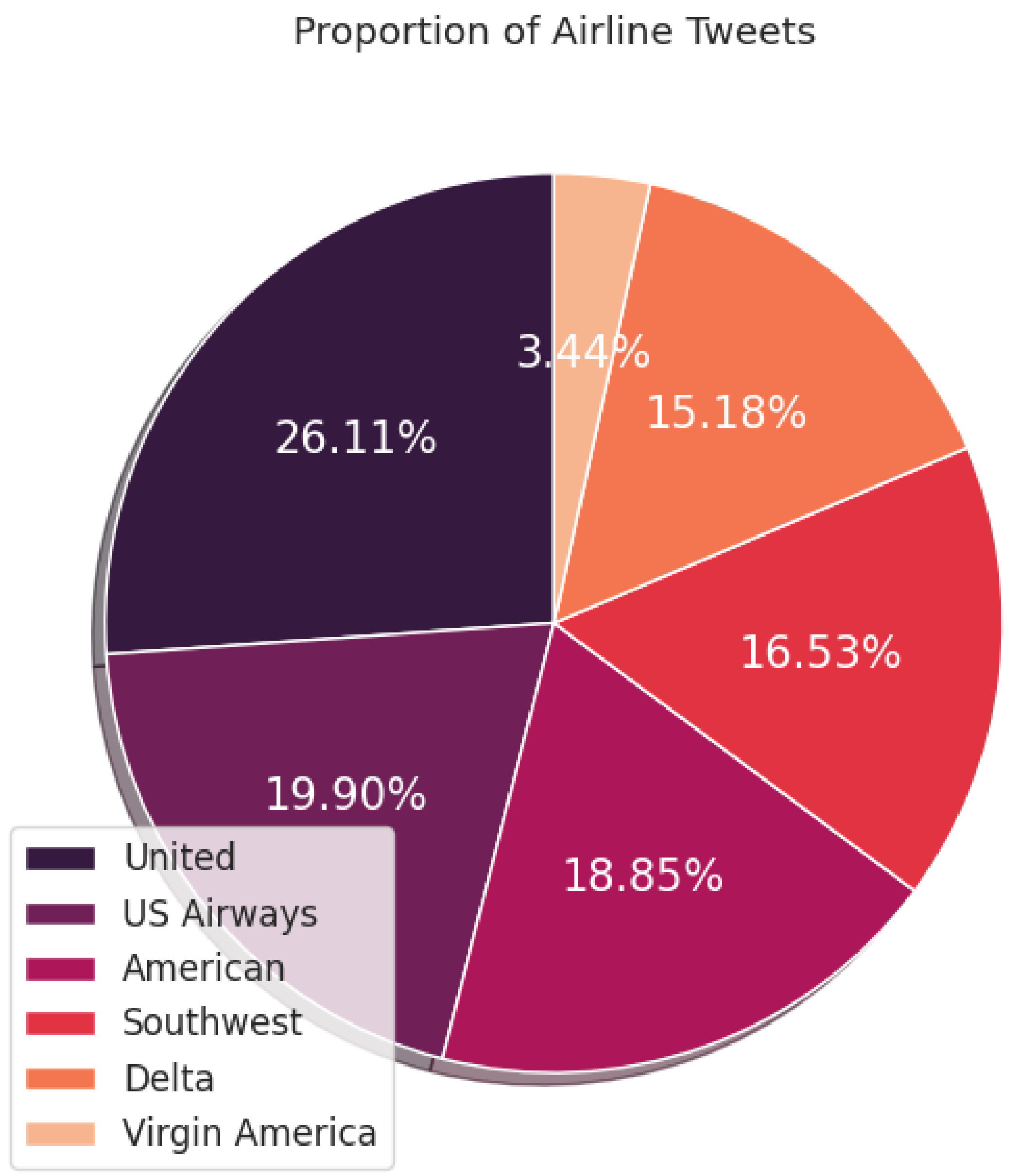
- Social media is a driving factor of attracting and deterring customers.
- Companies in industries like the airlines industry need to track how customers are reacting towards them, with special emphasis on negative reactions.

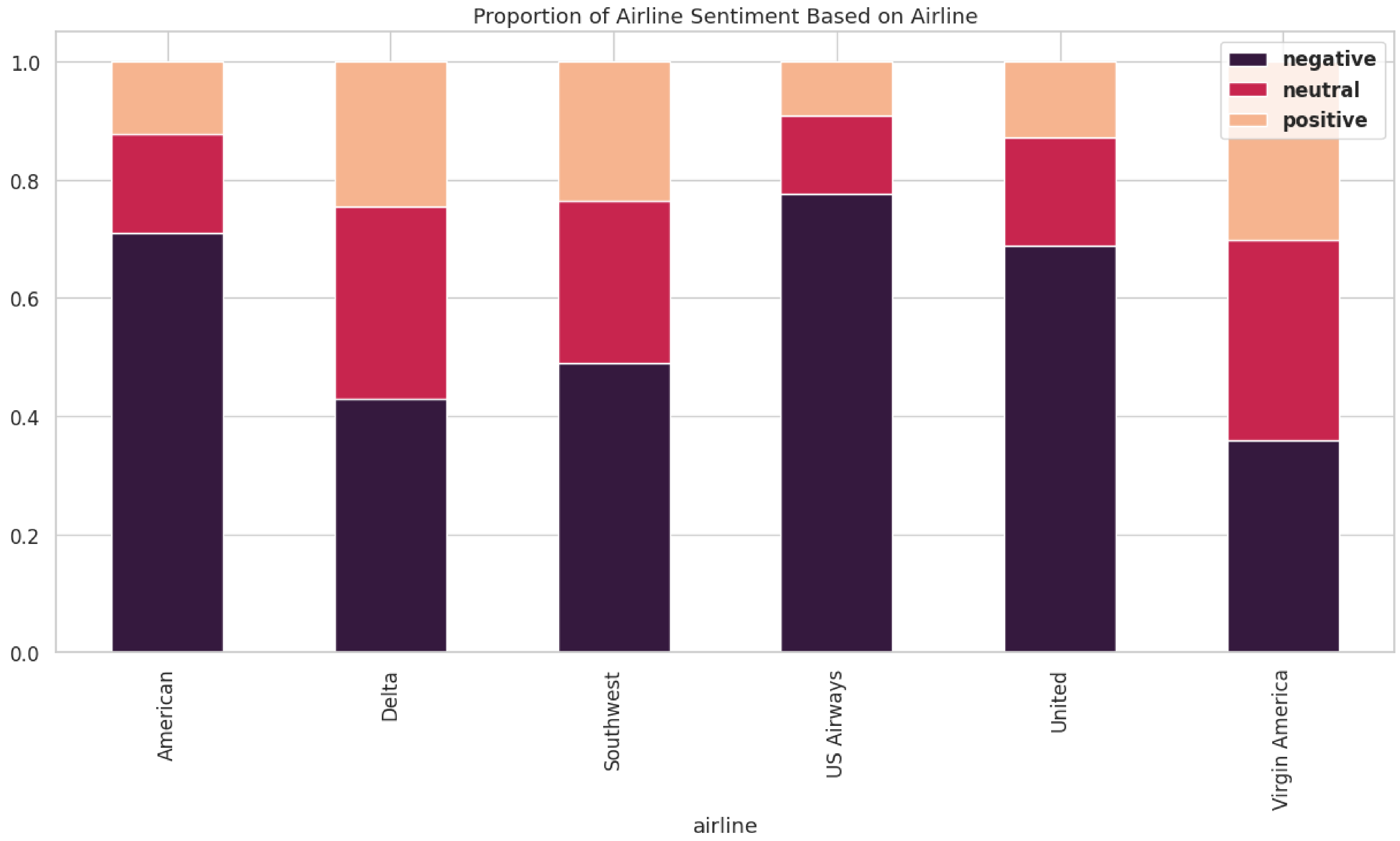
Data Overview

- Dataset is sourced from Kaggle: Twitter US Airline Sentiment.
- The data was scrapped from Twitter/X and classified as negative, positive, neutral.
- Initial dataset had 14640 observations and 16 columns.
- Key columns we focused on: "airline", "airline_sentiment, "negativereason", "text".

Analysis: Distribution of Airline Sentiment

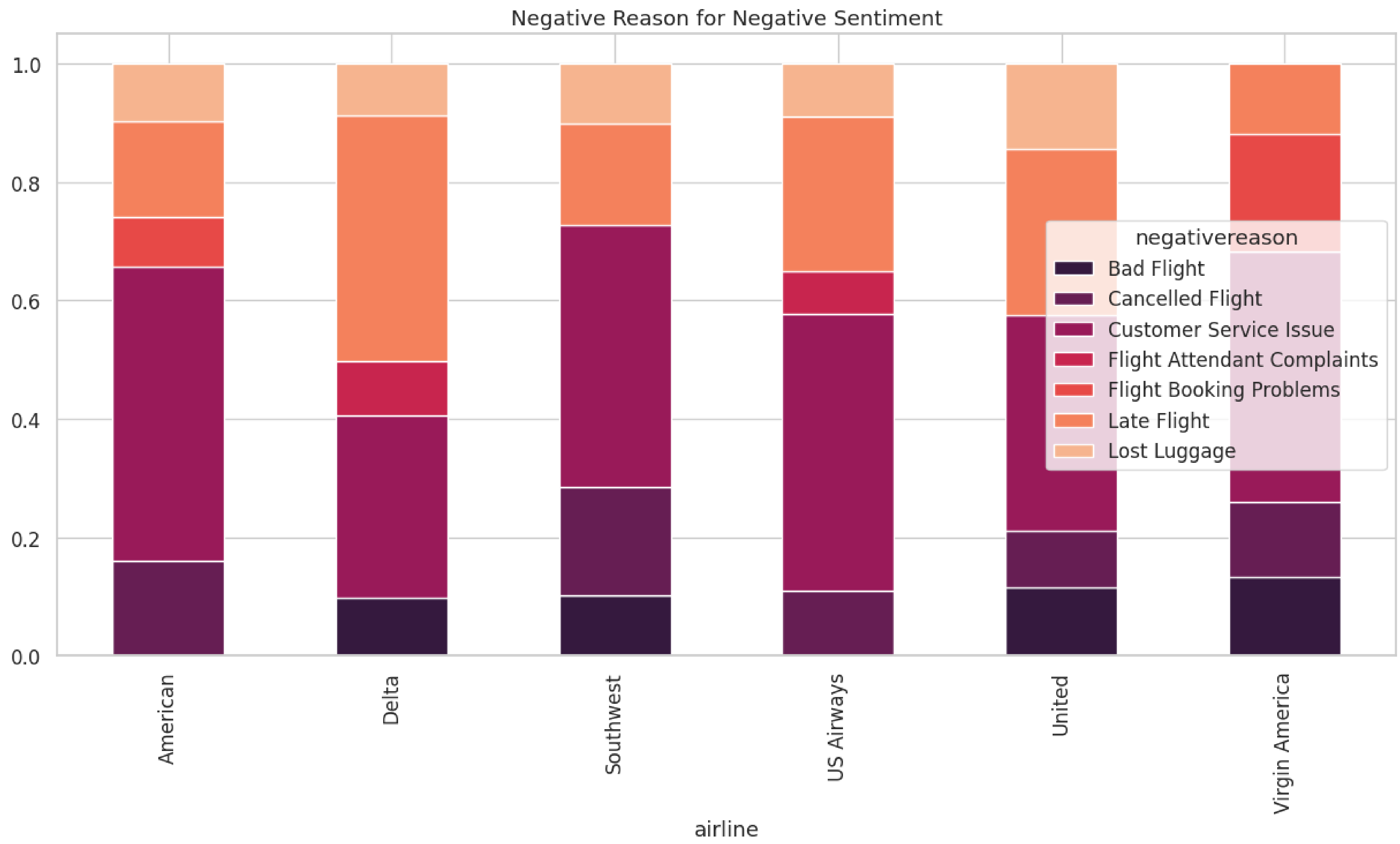
- Customers are primarily sharing negative tweets.
- United Airlines had the most number of tweets mentioning them and Virgin America had the fewest.





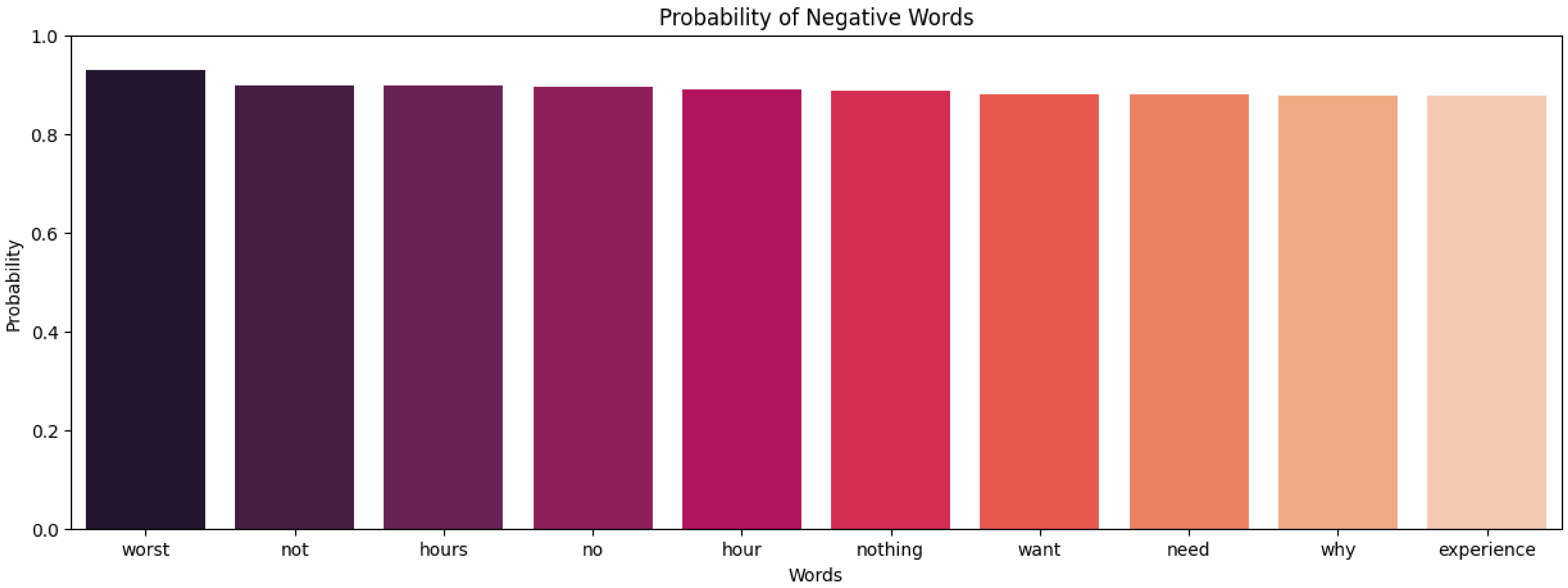
Analysis: Reason for Negative Sentiment

- Top 5 reasons for negative sentiment for each airline are shown.
- Customer service issues and late flights are the top issues across airlines.
- Delta top negative reason is late flights.
- The other airlines were customer service issues.



Predictive Words: United

- A common problem this airline had is related to customer sevice issues as their top negative reason for a negative sentiment.
- This was because of not giving timely response to the question.



"Could you update me on the suitcase please? The online and phone tracking told me nothing. I was told I'd have it back yesterday!"

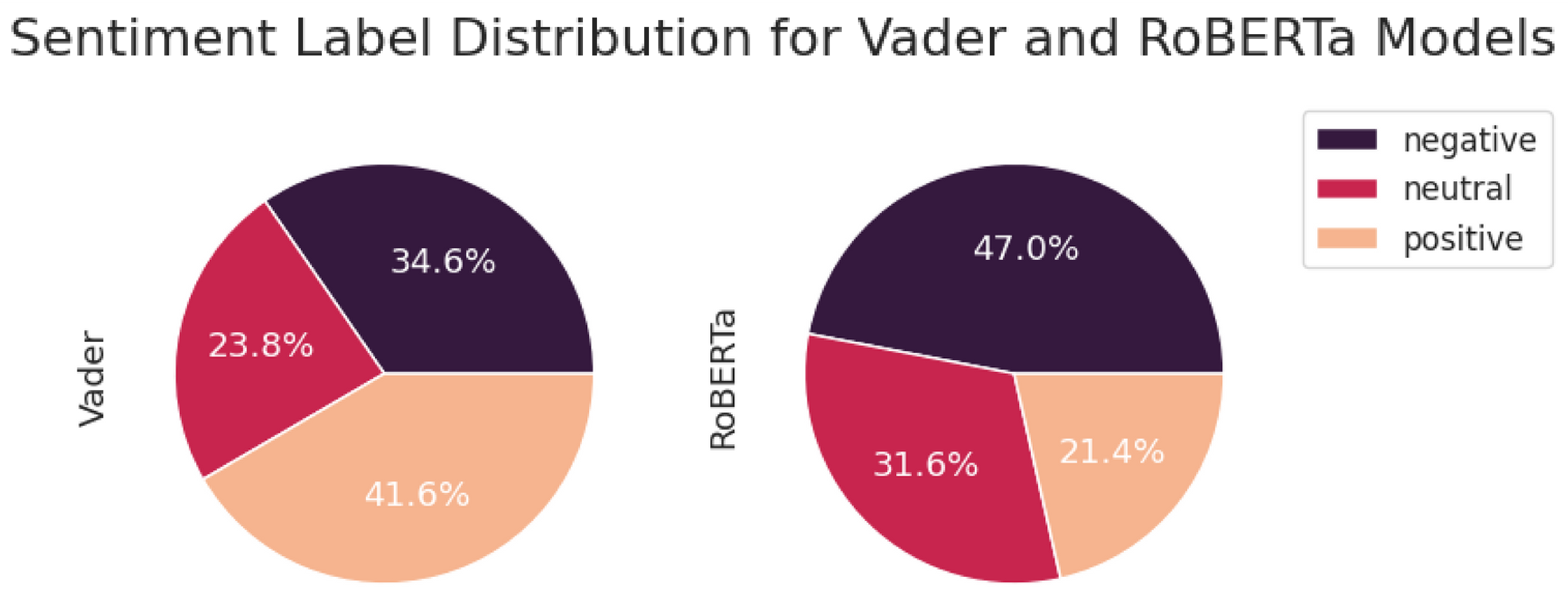
"I'm really glad I just waited on the phone for over an hour to be sent to a voicemail. Your customer service sucks."

Modeling: Data Text Preprocessing

- Initially remove columns where >90% was null
- Impute ~4000 Y values manually
- Made text lowercase and remove punctuations, URLs, special characters, digits.
- Lemmatized text
- Removed stop words ("this", "is", etc.)
- Tokenized text
- Splitting data 70/30 into train and test sets with a random state 1

Vader & RoBERTa Models

- Valence Aware Dictionary for Sentiment Reasoning is an NLTK module for sentiment scoring based on word usage.
- It's a rule-based analyzer categorizing terms as positive or negative in sentiment.
- RoBERT is a transformer-based language model specialized in sentiment analysis.
- Trained on social media text like tweets, it predicts sentiment labels (positive, negative, neutral) by capturing unique language patterns



Vader & RoBERTa Models - Evaluation

- The RoBERTa model outperforms the Vader model in terms of accuracy, macro average F1-score, and weighted average F1-score.
- RoBERTa's higher precision, recall, and F1-scores for all sentiment categories indicate that it provides better overall performance in classifying sentiments compared to the Vader model.

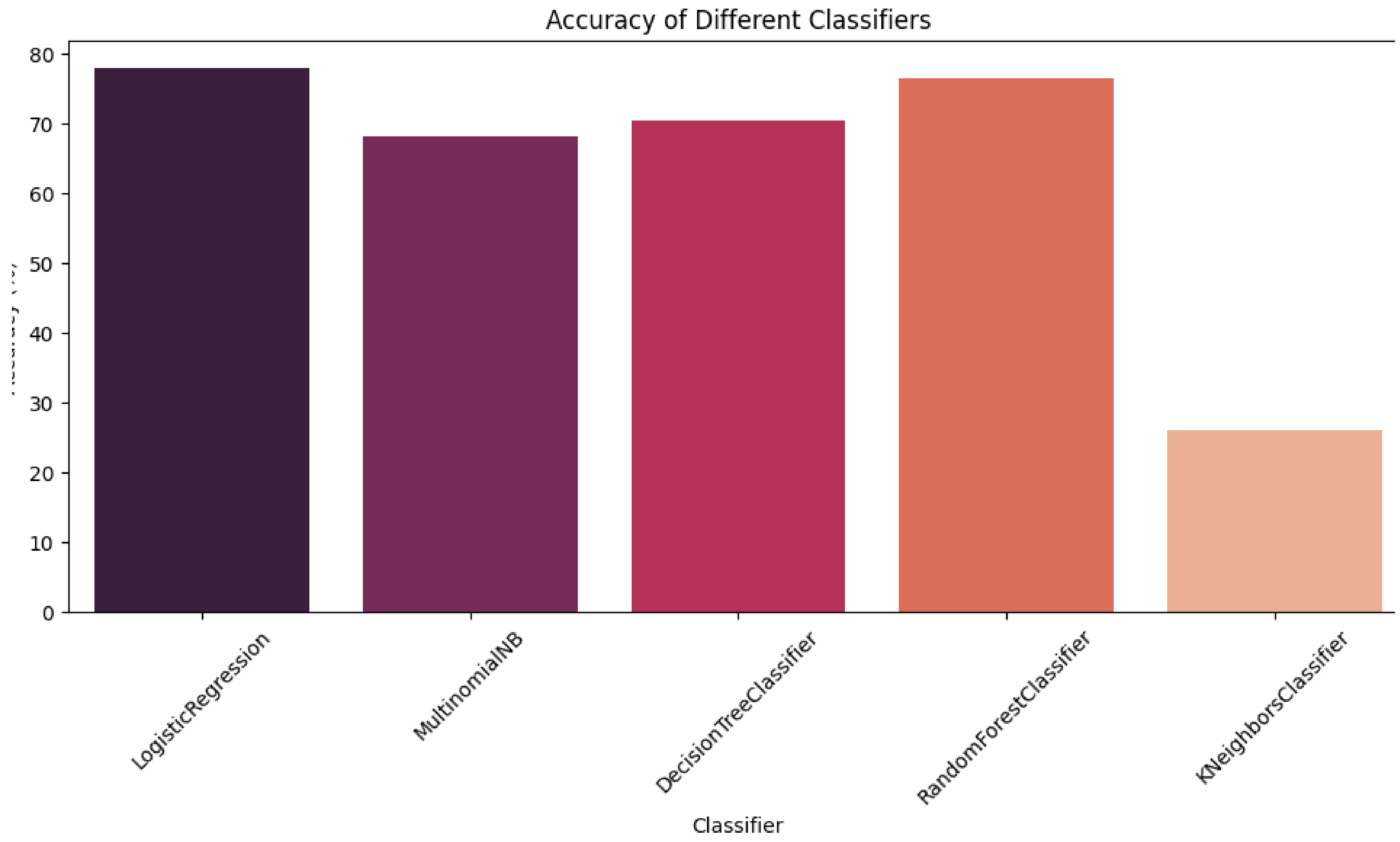
```
Vader Model Metrics:
Accuracy: 0.54
Classification Report:
              precision    recall  f1-score   support

    negative       0.89      0.49      0.63      2349
     neutral       0.43      0.43      0.43       941
    positive       0.34      0.84      0.48       710

    accuracy                           0.54      4000
   macro avg       0.55      0.59      0.52      4000
weighted avg       0.68      0.54      0.56      4000


RoBERTa Model Metrics:
Accuracy: 0.73125
Classification Report:
              precision    recall  f1-score   support

    negative       0.93      0.70      0.80      2349
     neutral       0.50      0.68      0.58       941
    positive       0.67      0.89      0.76       710

    accuracy                           0.73      4000
   macro avg       0.70      0.76      0.71      4000
weighted avg       0.78      0.73      0.74      4000
```

LR, MNB, Decision Trees, k-NN, Random Forests

- Text processing using TF-IDF vectorization and Evaluation
- **'TfidfVectorizer'** used to transform text into numerical features
- Utilized Logistic Regression, Multinomial Naive Bayes, Decision Trees, Random Forests, and k-Nearest Neighbors.
- Trained classifiers on TF-IDF transformed training data.
- Evaluated accuracy and generated classification reports.

LR, MNB, Decision Trees, k-NN, Random Forests - Accuracy

Logistic Regression = 78%
Multinomial Naive Bayes = 68%
Decision Tree classifier = 71%
Random Forest classifier = 76%
k-NN = 26%

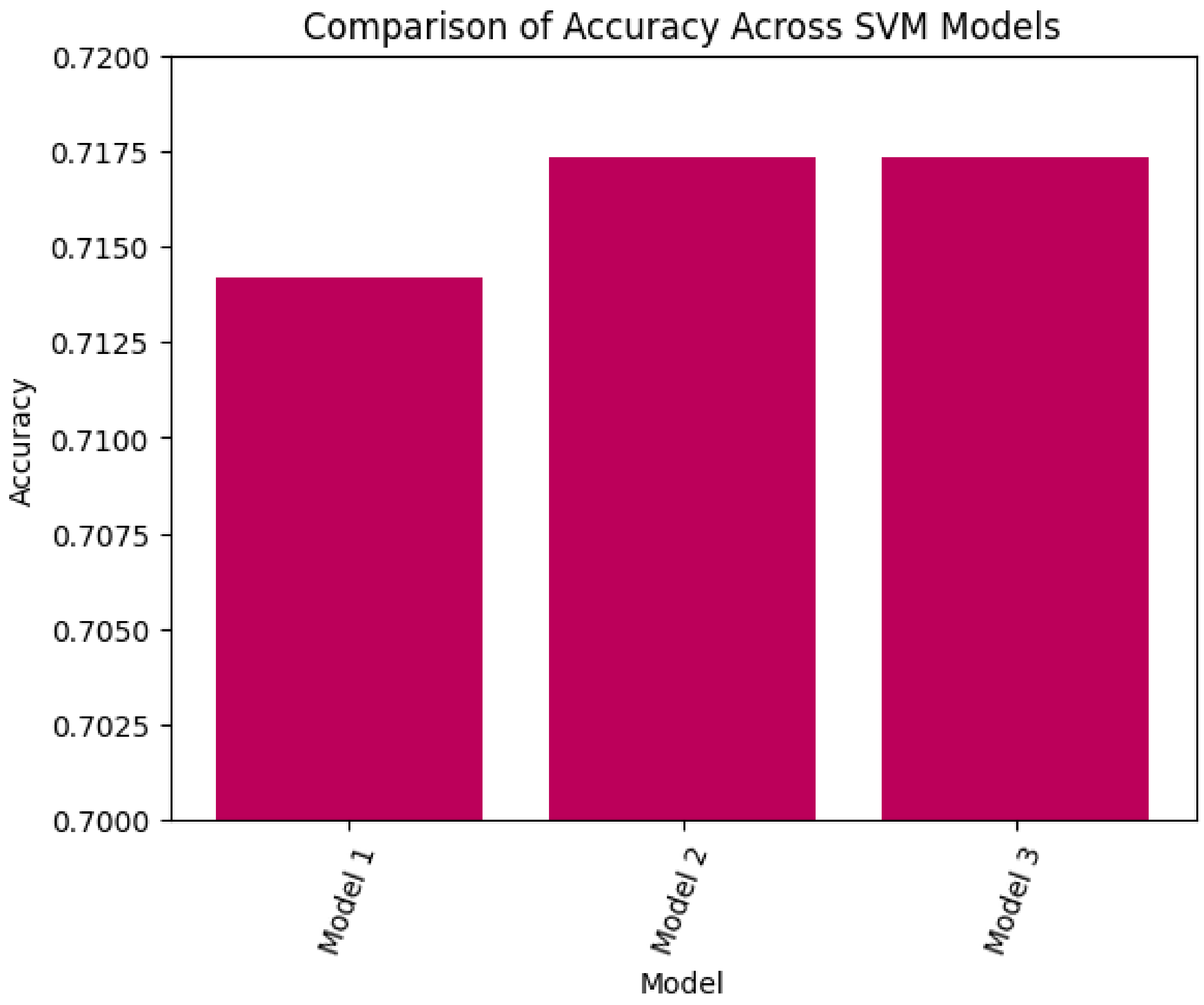


Support Vector Machine

- Created 3 models using SVM
    - **Model 1**: Using CountVectorizer() with no parameters
    - **Model 2**: Using TfidfVectorizer() with parameters from research
    - **Model 3**: Using CountVectorizer() with hypertuned parameters
        - C = 10, Gamma = 0.1, kernel = rbf

Support Vector Machine

- Accuracy only increased by 1% between model with no parameters and hypertuned model
- F1 Score for negative tweets also only increased by about 1%



Neural Network

- Leveraged Keras Sequential model for architecture design.
- Utilized an Embedding layer to convert tokenized words into dense vectors.
- Integrated LSTM layer to capture sequence dependencies in the data.
- Final Dense layer with softmax activation for sentiment classification.
- GridSearchCV is used to perform hyperparameter tuning.

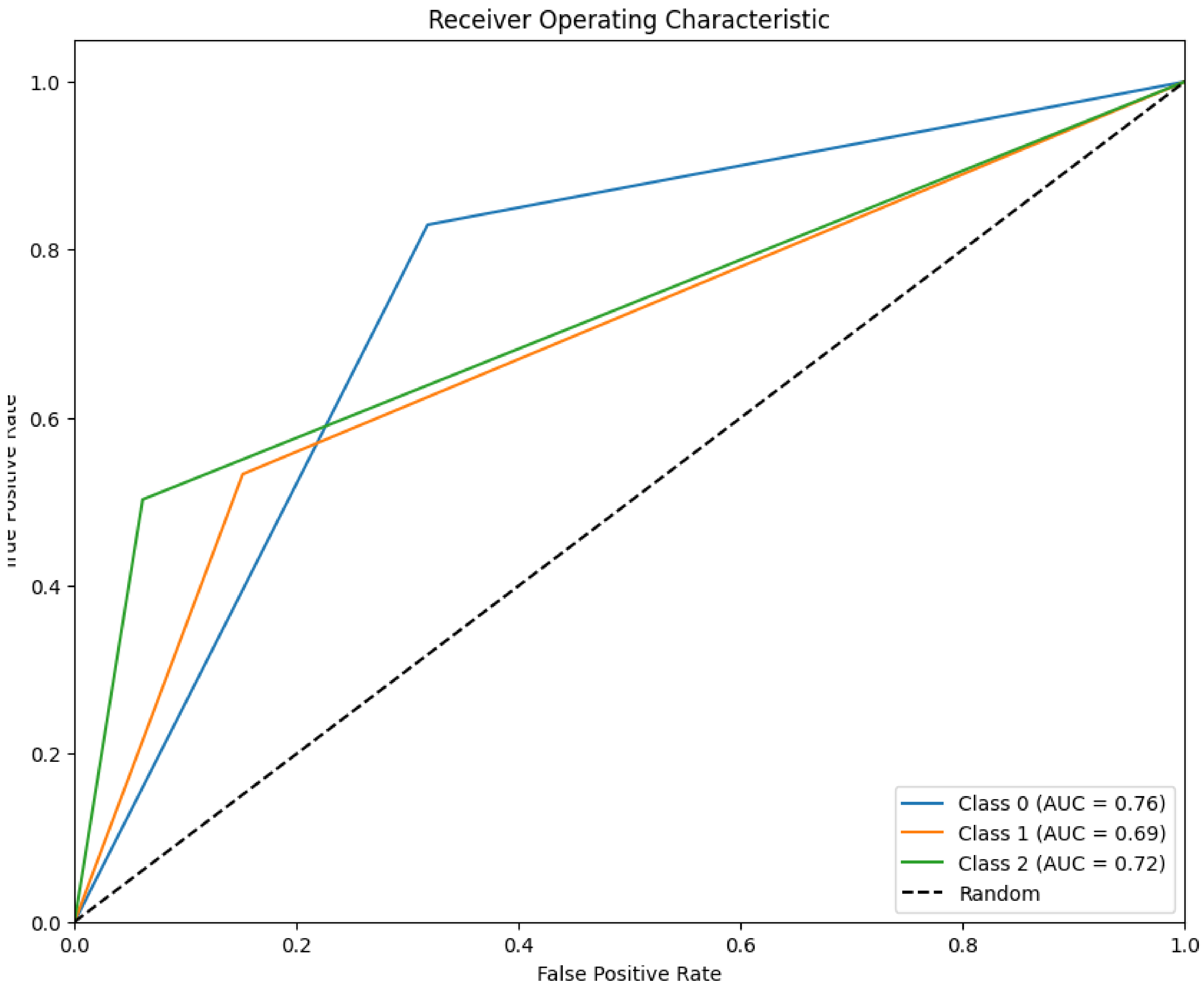
Neural Network

Model Report

	precision	recall	f1-score	support
negative	0.80	0.83	0.81	722
neutral	0.53	0.53	0.53	289
positive	0.61	0.50	0.55	189
accuracy			0.71	1200
macro avg	0.64	0.62	0.63	1200
weighted avg	0.70	0.71	0.70	1200

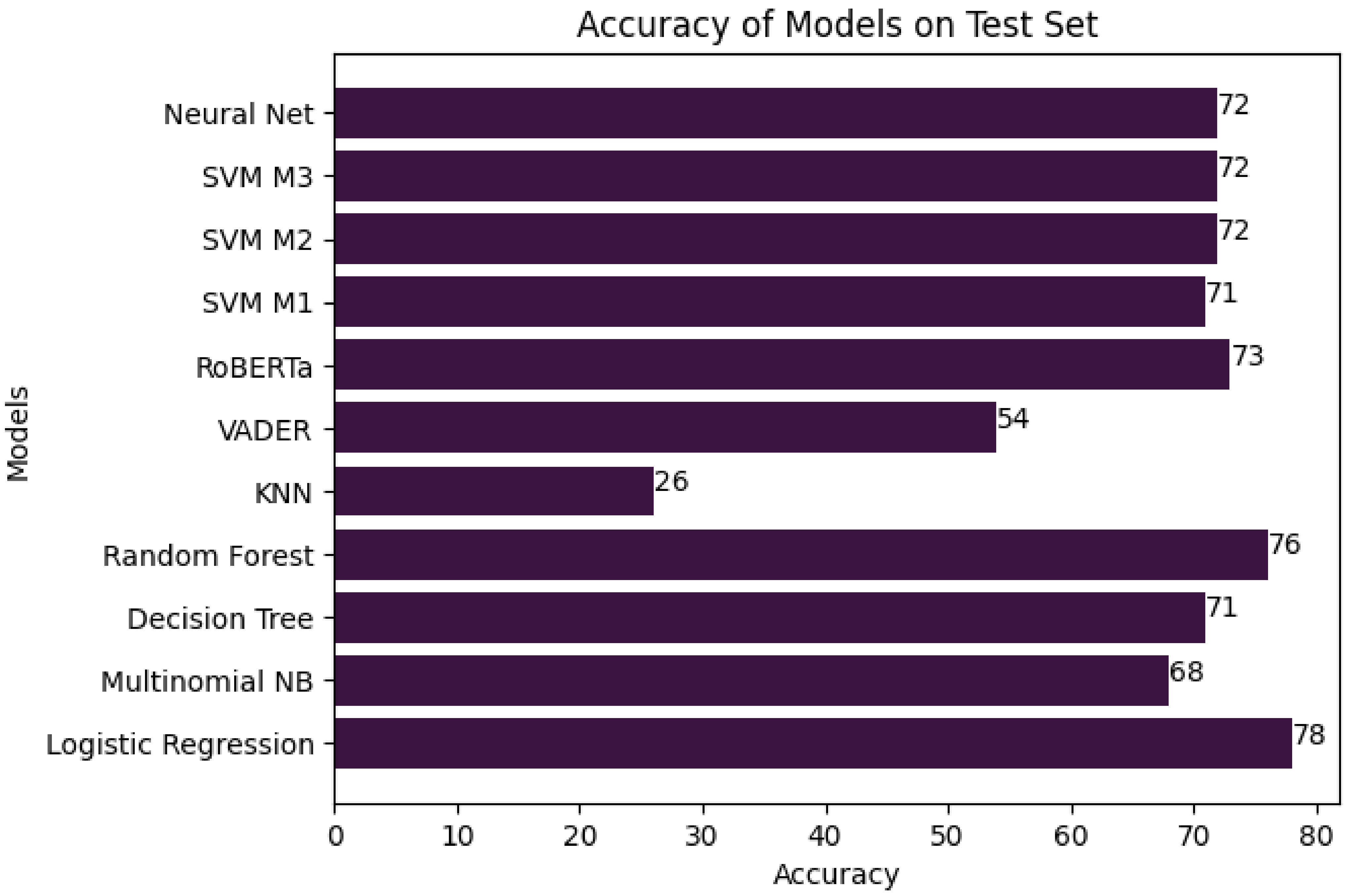
Neural Network

ROC Curve



Overall Results

- Accurate sentiment analysis is critical for various applications.
- Logistic Regression and Random Forest strike a balance between interpretability, but Logistic Regression has higher accuracy.
- Model choice should align with project requirements and data characteristics.
- Neural Network has more room for further work.



Project Impact

- As social media usage grows, it'll be important for companies to have effective ways of combing through customer responses.
- Allowing individuals and companies to gauge consumer response to them and their products.

"It's a dialogue, not a monologue, and some people don't understand that. Social media is more like a telephone than a television."
— Amy Jo Martin

Next Steps

Neural Networks

- Increase epochs
- more hyperparameter tuning
- handle class imbalance
- using pre-trained word embeddings

SVM

- OneVSRestClassifier()
- Balance dataset between positive/negative/neutral

Breaking down cause of negative tweet into specific categories for targeted usage

Thank You