Sentiment Analysis
on Airline Tweets Dataset

Problem Statement / Opportunity

In the airline industry, traditional customer feedback forms are tedious and time consuming. Tweeter data serves as a good source to gather customer feedback. Sentiment analysis helps monitor social media mentions and proactively manage negative comments, so as to understand customers preference.

Sentiment Analysis using Data Science
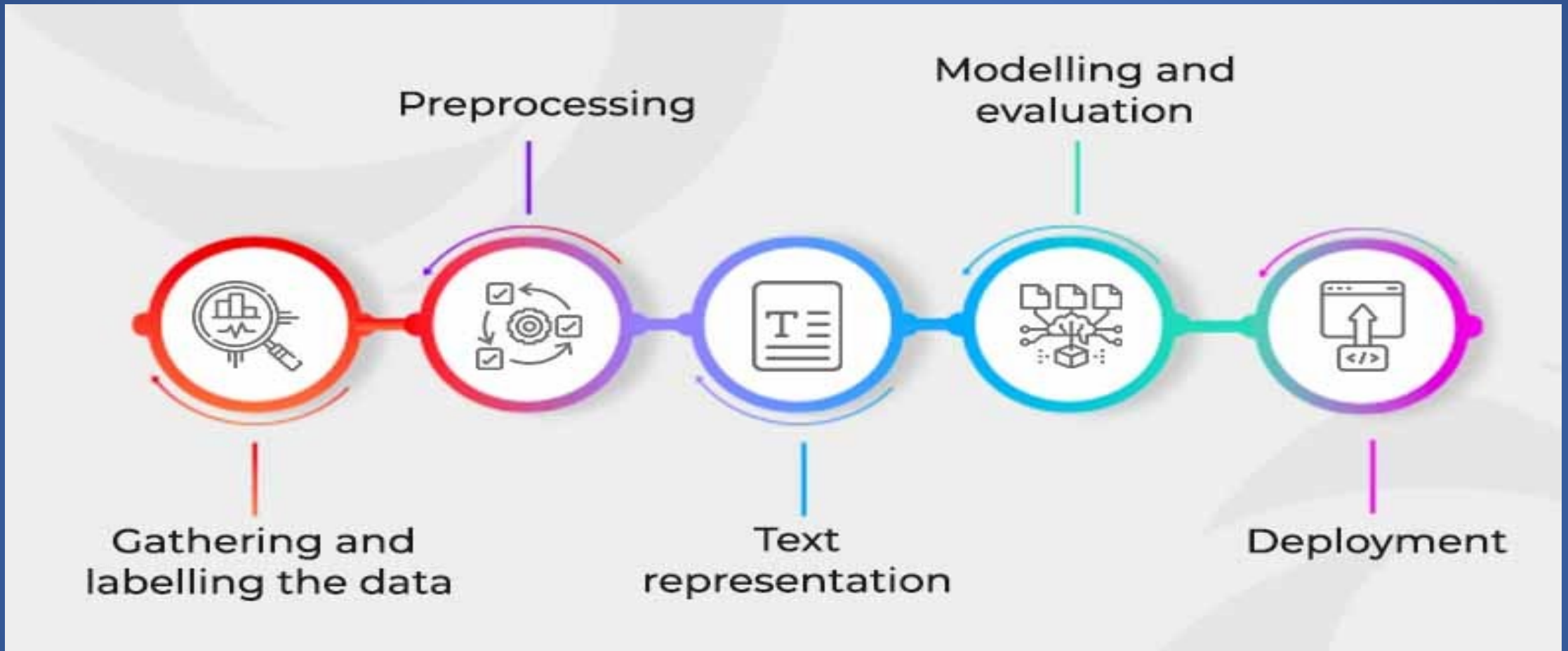
Dataset and Pre-processing

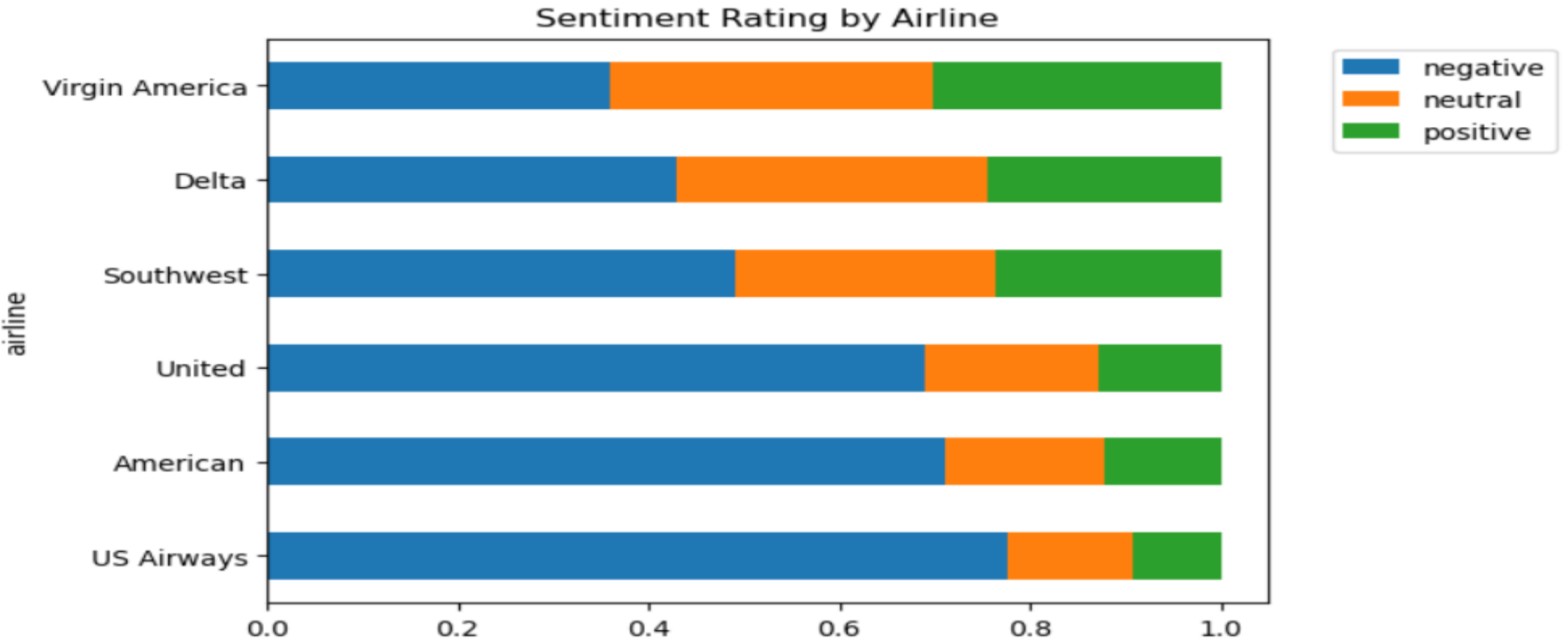- ✓ This dataset contains tweets for 6 US airlines, 14,640 observations and 15 features.

- ✓ We will focus on "Text" and "airline_sentiment"

- ✓ Some features "Negative Reason" and "Airline" are useful for model evaluation

Exploratory Data Analysis

Sentiment by Rating

Tweet Volume Distribution For Negative Sentim
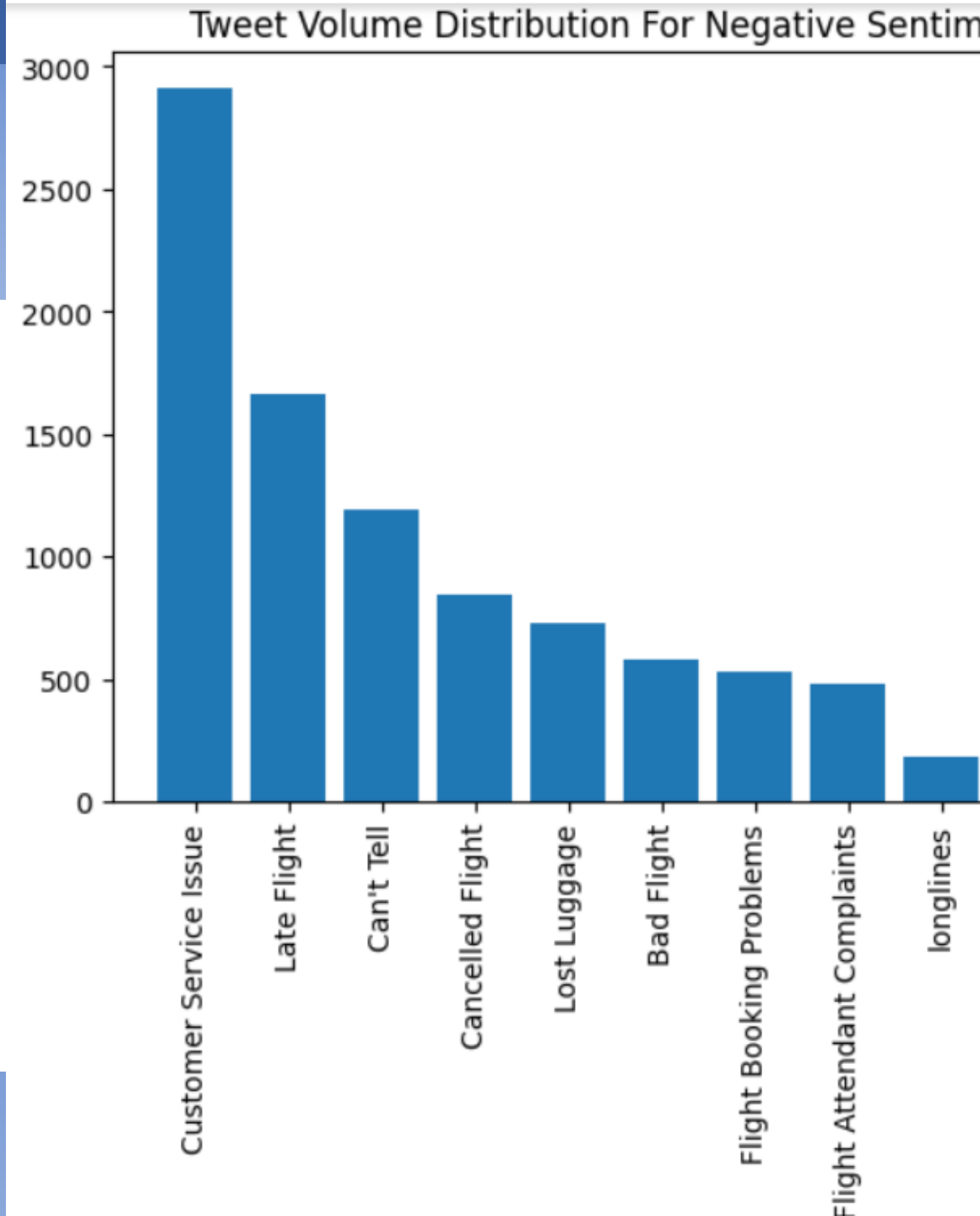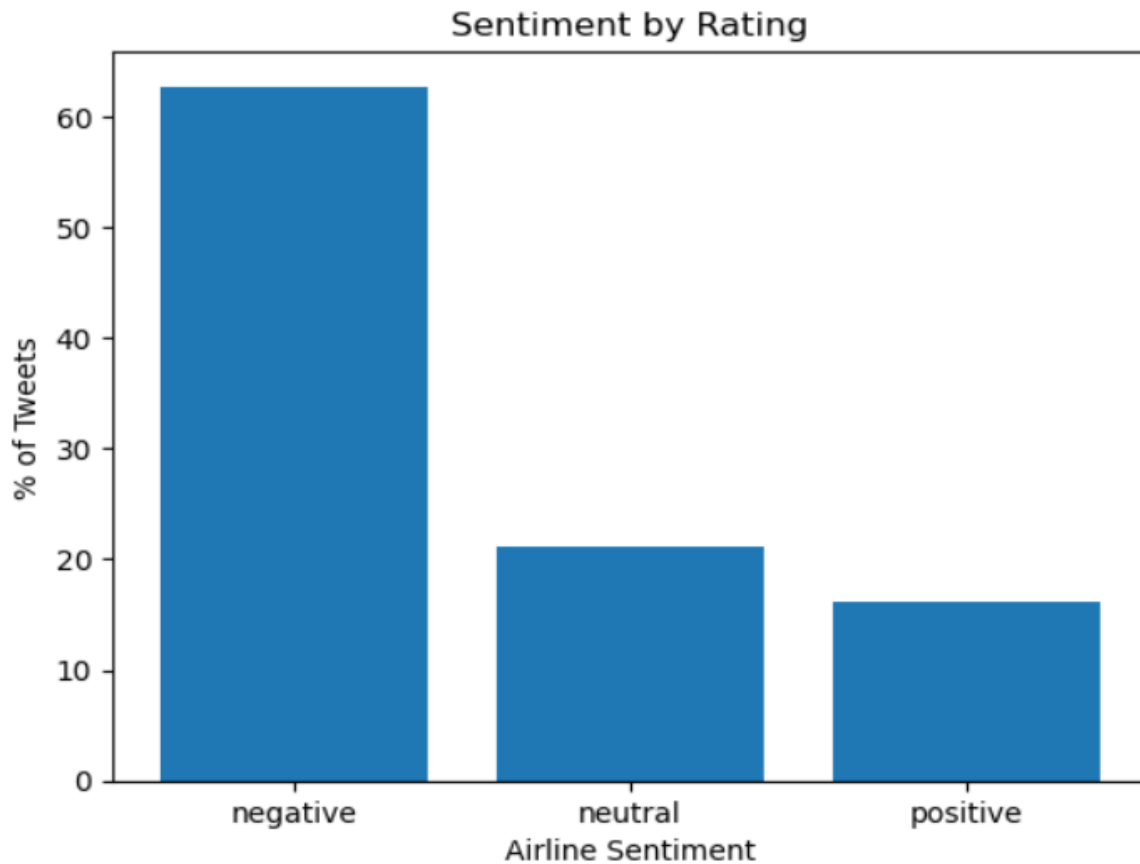
Negative review: 63%. Neural review: 21%. Positive review: 16%.

Exploratory Data Analysis

Sentiment Rating by Airline

Model Performance Comparison

| Text Representation	Model	Accuracy
Bag of Words	Logistic Regression	77%
	Random Forest	74%
	Adaboost Classifier/Gradient Boosting	73%
TF-IDF	Decision Tree	71%
	Logistic Regression	80%
	Adaboost Classifier/Gradient Boosting	71% / 75%
Word2Vec	Neural Network	88%
Text_Vectorization /Embedding layer	RNN	75%

Neural Network based on Word2Vec
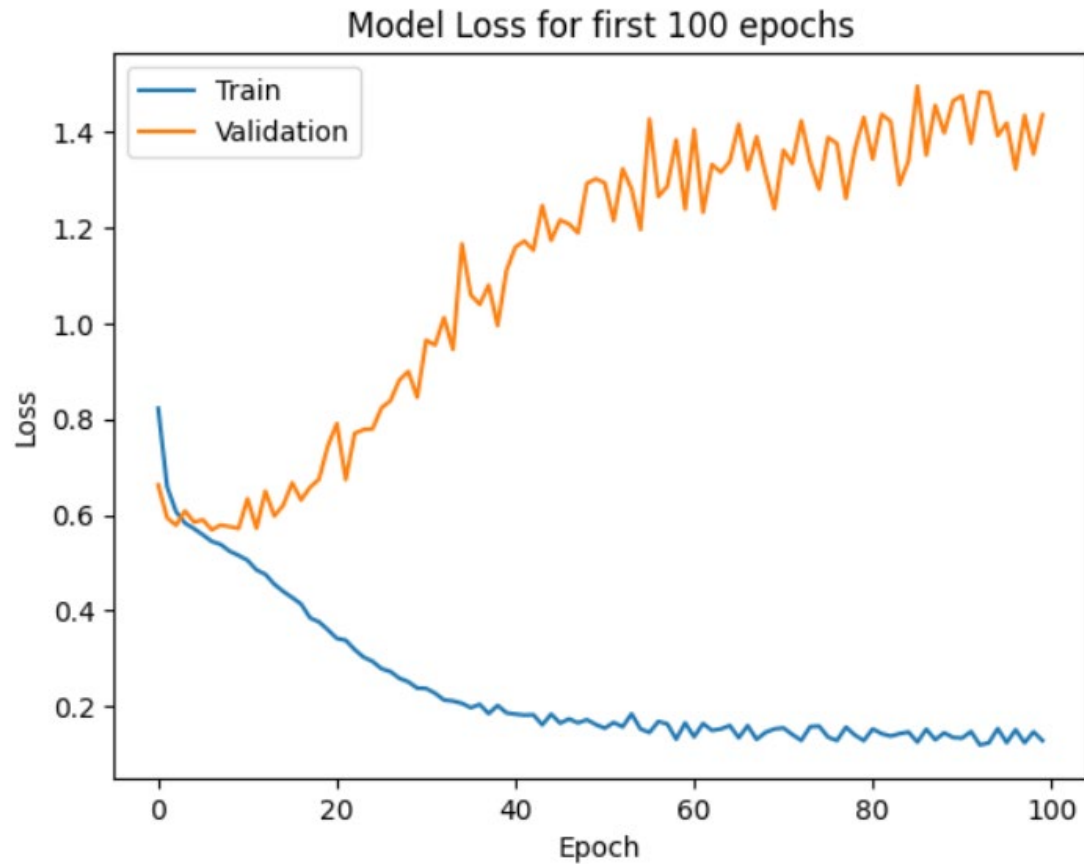
	precision	recall	f1-score	support
0	0.94	0.91	0.92	6093
1	0.84	0.76	0.80	2115
2	0.74	0.91	0.82	1600
accuracy			0.88	9808
macro avg	0.84	0.86	0.85	9808
weighted avg	0.89	0.88	0.88	9808

One_vs_the_Rest ROC Curve and AUC
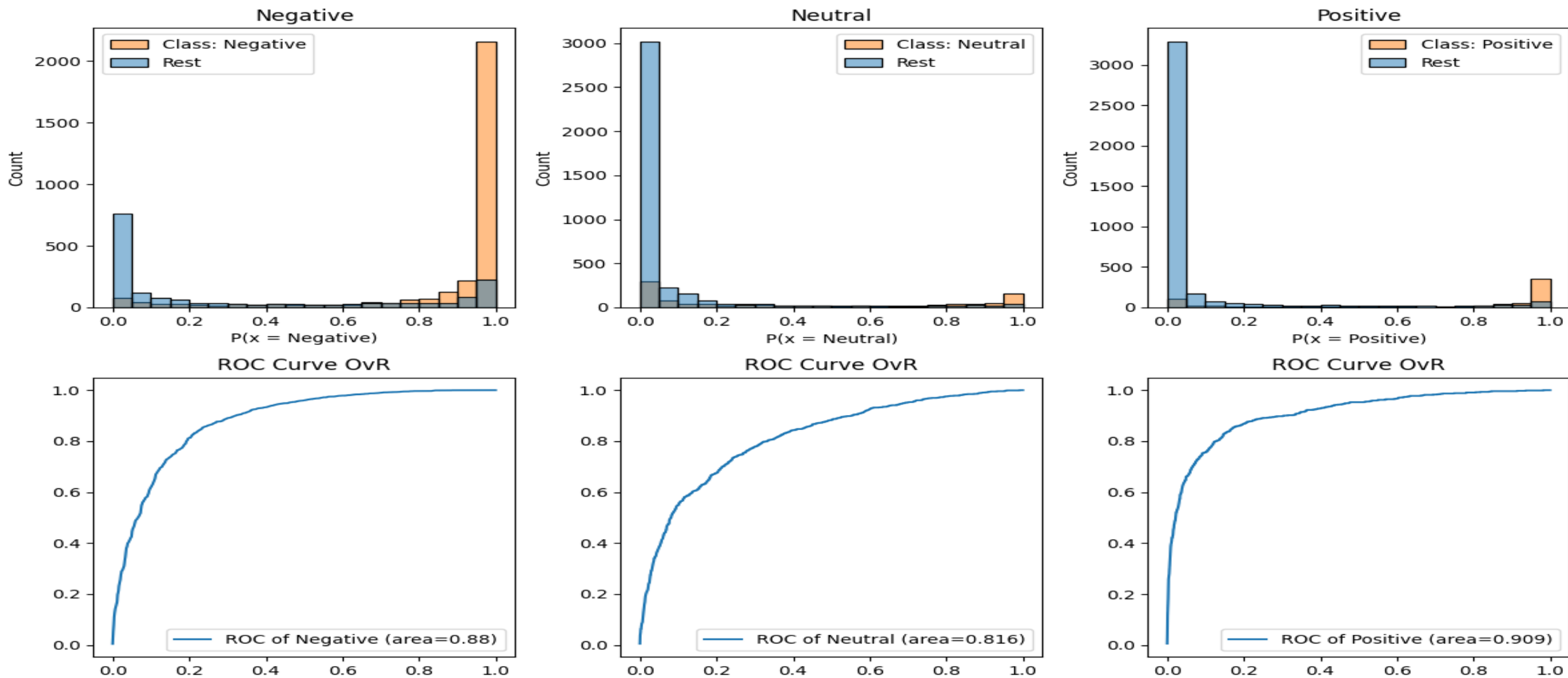
Conclusion

- Neural Network based on Word2vec performs the best at the accuracy of 88%.
- By random guessing, the accuracy of identifying negative reviews is 63%. With the help this sentiment analysis, we can achieve 91%. (28% improvement).

Thanks for listening.

Steve Song   MBA

Linkedin: linkedin.com/in/yinjingsong
Website: github.com/Steve-Yinjing-Song
Email: stevesong325@gmail.com