Twitter Sentiment Analysis: An NLP Application

Nikhil Kappagantula, Angela Ploegman, William Schmidt, Bailey Scoville

Executive Summary

With the rise of social media as a means to communicate, spread news, and share information, sentiment analysis is increasingly important. We will examine a slice of Twitter (now X) tweet data tagging six major US airlines to **predict tweet sentiment**.

Benefits of sentiment analysis include:

- Increasing the airlines' ability to quickly respond to customer service issues
- Gauging public sentiment on a major social media platform
- Identifying recurring customer service topics

Research Objective

Build four machine learning models to predict text sentiment using Twitter data.

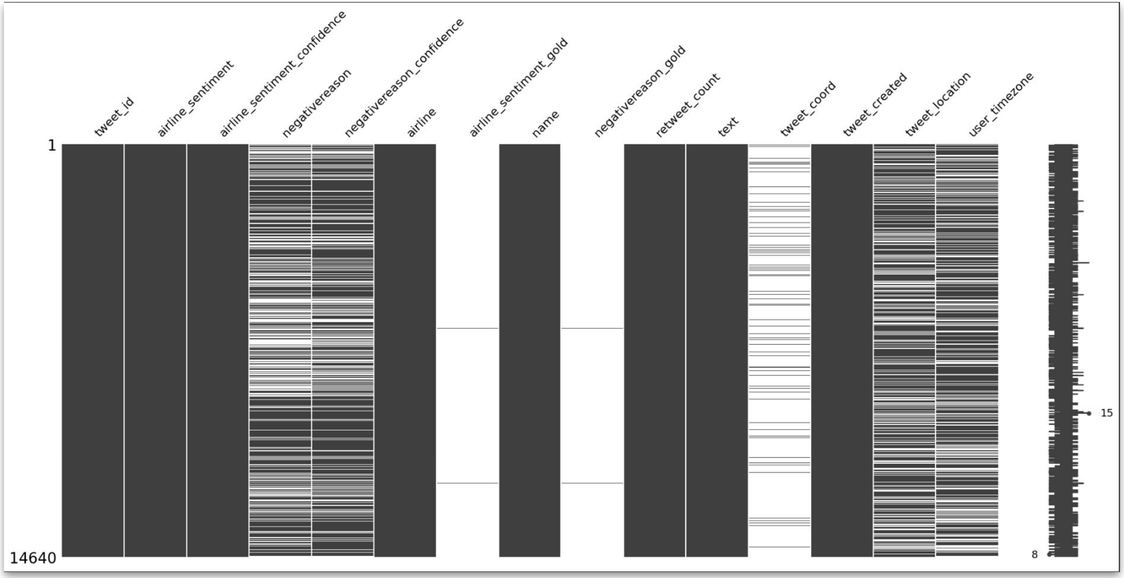
Guiding Questions:

- How can tweet text be used to predict sentiment?
- Do the models improve with text and additional features (e.g. timezone)?
- Which model type (statistical, decision tree, deep learning) best predicts tweet sentiment?

Research Methods

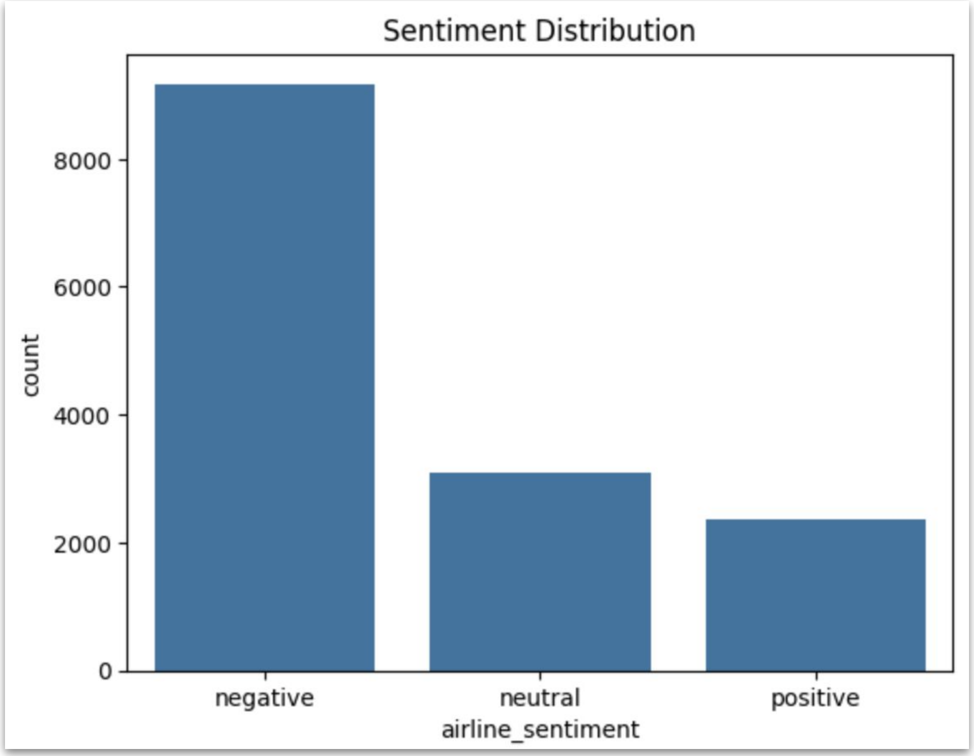
Exploratory Data Analysis

- Dataset includes **14640 tweets** mentioning six major US airlines
- Thousands of missing values in 7 of 15 columns, but no missing values in columns needed for primary analysis (text, airline sentiment)

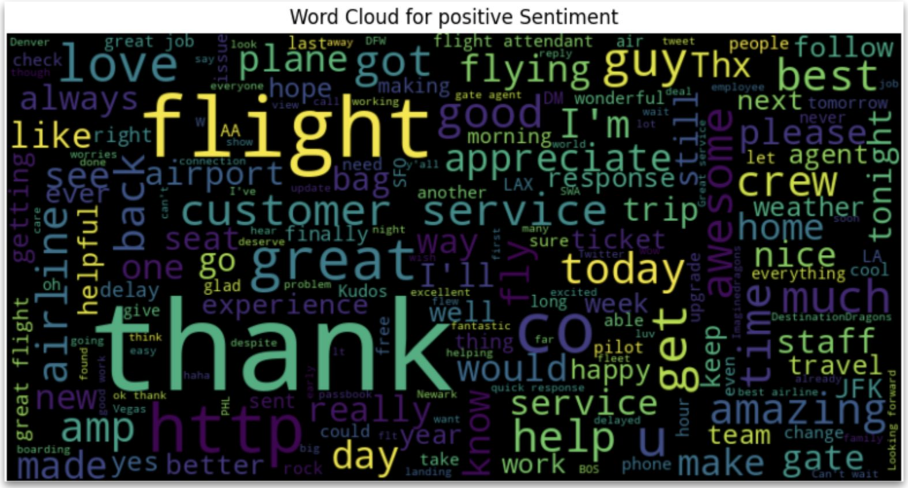


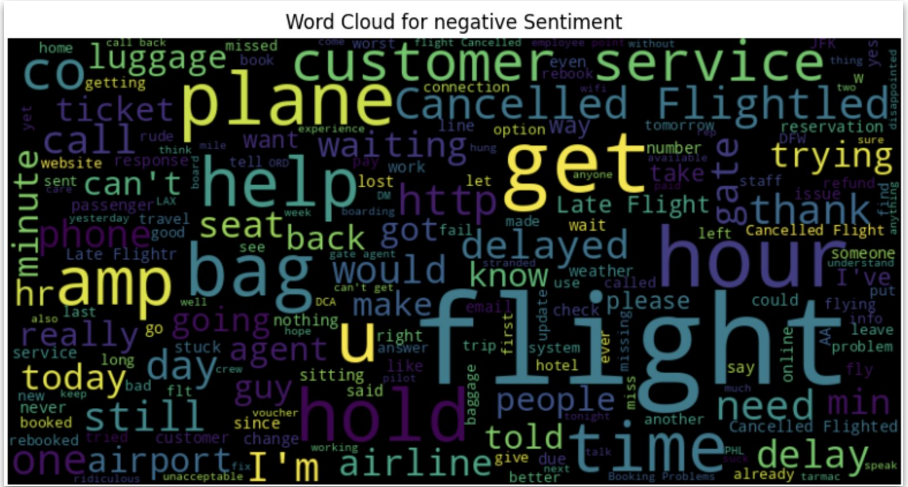
Exploratory Data Analysis - Sentiment

- Barplots show tweet sentiment in dataset skews negative
- Word clouds show most commonly-used words in positive and negative sentiment tweets
  - Word clouds on next slide



Exploratory Data Analysis - Word Clouds

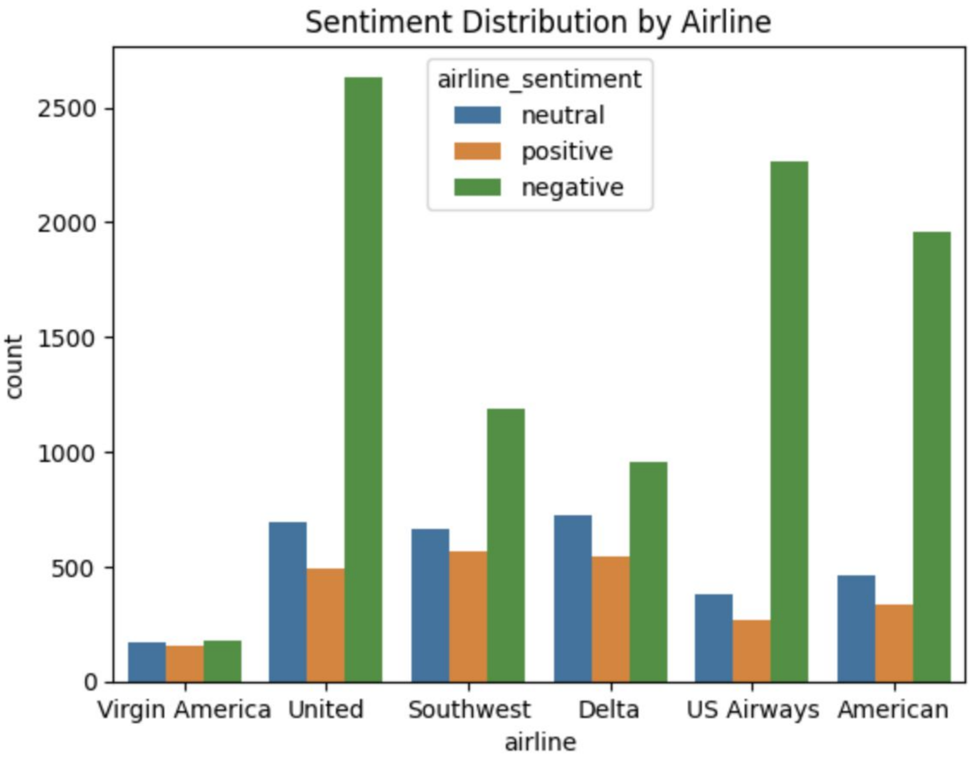




Data Preparation - Feature Engineering

New features:

- Clean timezone
- Tweet day
- Tweet time of day
- Numeric sentiment
  - Positive = 1
  - Neutral = 0
  - Negative = -1



Data Preparation - Tweet Cleaning

- Defined functions to remove:
  - Stopwords
  - Punctuation
  - Twitter usernames
  - Hyperlinks
  - Emojis
- Reformatted remaining text:
  - Lower case
  - Stemmed words

Data Preparation - Text Preprocessing

- Bag of Words (pictured)
  - Turn words into 0s and 1s for analysis
- Tokenization
  - Breaks words into small bits of text
- Encoding
  - Creates numerical representations of categories
    - E.g. Virgin America = 5

```
['plu youv ad commerci experi tacki',
 'didnt today must mean need take anoth trip',
 'realli aggress blast obnoxi entertain guest face amp littl recours',
 'realli big bad thing']
```

real	realiz	realli	reason	rebook	receipt
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0

Methodology - Tools

- Google Colab
- Python
  - Pandas
  - Numpy
  - Nltk
  - String
  - Re
  - Scikit-Learn
  - Keras
  - Tensorflow
- Power BI

Model Selection

Model 1 - Naive Bayes

- Naive Bayes uses Bayes' Theorem to compute the probability a tweet expresses a certain sentiment
- Trained on tweet text and tweet characteristics
- Coming in third place for accuracy, our Naive Bayes classifier achieved an overall accuracy of approximately 88.87%
- We see it performed well in correctly identifying instances for sentiment classes 0, 1, and 2, with only a small number of misclassifications observed

Model 2 - Random Forest

- Random Forest constructs many decision trees and returns the most frequent label to evaluate sentiment
- Trained on tweet text and tweet characteristics
- The random forest classifier achieved an overall accuracy of 95.01%, indicating a solid ability to accurately classify tweet sentiment
- However, some misclassifications were observed, particularly between sentiment categories 1 and 2, suggesting potential areas for refinement to enhance accuracy further

Model 3 - Long Short-Term Memory Model (LSTM)

- LSTM recurrent neural networks identify sentiment by taking in sequential inputs and maintaining an internal memory state
- Trained on tweet text
- The LSTM model achieved a moderate level of accuracy, scoring about 75%, meaning it performed worse than our other models
- The loss value of 1.31 suggests that the model's performance could potentially be improved by reducing the discrepancy between predicted and actual sentiments during training

Model 4 - Gated Recurrent Unit (GRU)

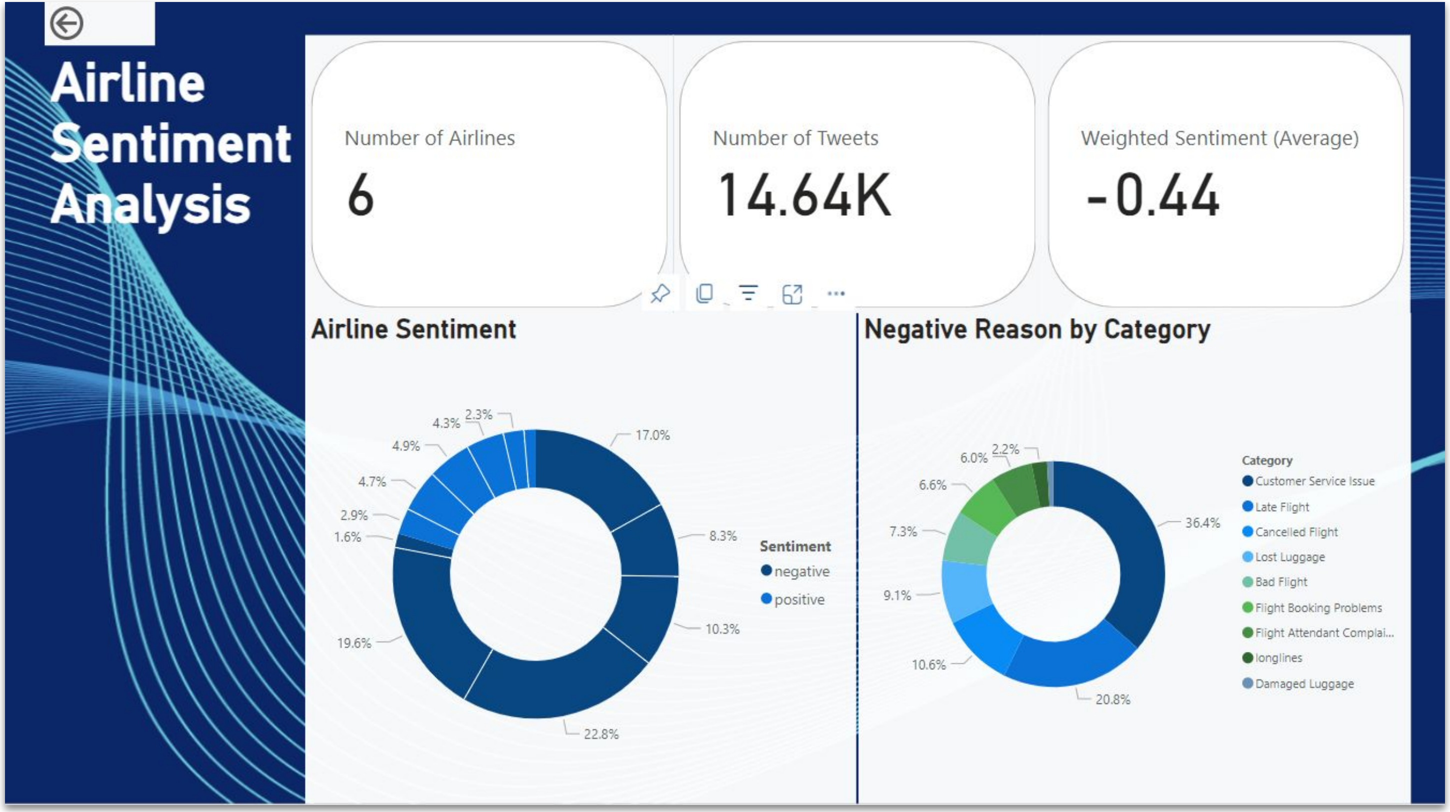
- A GRU recurrent neural network can capture long-range dependencies in text data while mitigating the vanishing gradient problem
- Trained on tweet text
- The model achieved an overall accuracy of 98%, indicating its high capability in correctly predicting the sentiment of tweets
- Precision, recall, and F1-score for all sentiment classes ranged from 0.98 to 1.00, indicating that the model can identify each sentiment category, which is crucial for decision-making

Applications to Management

Automation

- A Power BI dashboard with drill down analytics was developed to visualize the sentiment data
- The Power Automate functionality can take in a streaming data set from the Twitter API and add rows to the existing dataset
- The selected model or models can be applied in Power BI and provide near real time data to the organization to shorten decision making times

Power BI Dashboard



Findings & Conclusions

- **GRU Model**: Achieved the highest accuracy of 98% with precise classification across sentiment categories
- **Random Forest**: Showed robust accuracy of 95.01%, although slight misclassifications were observed between categories 1 and 2
- **Naive Bayes**: Achieved an overall accuracy of approximately 88.87%, performing well with minimal misclassifications
- **LSTM Model**: Obtained the lowest accuracy level of about 75%, indicating potential for improvement

Recommendations

- **Leverage GRU Model use**: Given 98% accuracy, prioritize further investment in GRU models for sentiment analysis tasks
- **Address Misclassifications in Random Forest**: Focus on refining the Random Forest classifier to reduce misclassifications, between categories 1 and 2
- **Improve LSTM Model Performance**: Enhance LSTM accuracy by optimizing training, adjusting architecture, or incorporating additional features
- Further develop code to automate the assignment of negative sentiment reason from tweet text

Model Applications

This research identifies the most promising model to continue tuning before deployment. Possible applications include:

- Improving customer service workflows by expediting communication and complaint resolution on Twitter (now X)
- Directing operational focus by identifying commonly occurring complaint topics
- Directing customer base research initiatives by monitoring public feedback and complaints/praise topics

References

Chowdhury, Kuldeep. 2021. “10 Hyperparameters to Keep an Eye on for Your LSTM Model - and Other TIps.” Medium. https://medium.com/geekculture/10-hyperparameters-to-keep-an-eye-on-for-your-lstm-model-and-other-tips-f0ff5b63fcd4.

Géron, Aurélien. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2nd ed. Sebastopol, CA: O'Reilly.

Skillcate AI. 2022. “Sentiment Analysis Project - with Traditional ML & NLP.” Medium. https://medium.com/@skillcate/sentiment-analysis-project-with-traditional-ml-nlp-349185bf98dd.