Analyzing Tweets to Predict Customer Sentiment

Final Report

Zubair Lakhia; *Zlakhia3*
Eric Ung; *ericung72*
Jose Perez; *joperez08*
Mautasim Hussain; *mautasim_hussain*

Georgia Tech/EdX
GTx MGT6203x: Data Analytics for Business
Fall 2023

Table of Contents

Project Introduction:

Background and Context

In today's business landscape, understanding customer sentiments is essential for companies aiming to optimize operations and meet consumer demands. Brand perception acts as a key indicator of operational effectiveness and market receptiveness to products and services. Social media platforms, like Twitter (now known as X), offer a unique opportunity for businesses to access real-time data, enabling them to promptly forecast emerging sentiment trends. This project focuses on leveraging data from Twitter to create a predictive model capable of anticipating customer sentiments for businesses.

Customer sentiment provides invaluable insights, presenting a snapshot of operational effectiveness and demand for goods and services. Utilizing social media for real-time sentiment trend forecasting is crucial, particularly in the service-oriented sector. Social media platforms, especially Twitter, serve as sources of immediate customer feedback, offering real-time observations on consumer sentiments.

The primary objective of this project is to use Twitter data to build a predictive model for real-time monitoring of post sentiments. This systematic methodology aims to provide companies with the means to interpret and respond to customer sentiments, fostering an environment of adaptability and customer-centric operations.

Business Objectives and Goals

A predictive model for sentiment analysis holds significant benefits for businesses across various sectors. By accurately gauging customer sentiments, companies can gain valuable insights into how their products, services, and overall brand are perceived in the market. Understanding customer sentiment in real-time allows businesses to promptly address concerns, capitalize on positive feedback, and tailor their strategies to align with consumer expectations. This proactive approach enhances customer relations, fosters brand loyalty, and positions the company to make informed decisions for continuous improvement. Moreover, sentiment analysis models provide a quantitative measure of customer satisfaction, enabling businesses to track performance over time and adjust their operations to meet evolving market sentiments. In essence, a well-developed sentiment prediction model becomes a strategic tool, empowering businesses to stay responsive, adaptive, and customer-focused in today's dynamic business environment.

Identification of the Problem Statement

Our study attempts to perform sentiment analysis for businesses by using data obtained from Twitter. While we specifically focus on airlines, this approach can be used for any industry and business. Similarly, even though we source our training data from Twitter, it’s essential to recognize that sentiment monitoring can be performed across multiple social media platforms. We need to develop a flexible approach for extracting crucial features from extensive text data, emphasizing the need for a strong feature selection process. The other goal is building an optimized predictive model that accurately categorizes sentiments (negative, neutral, or positive) from tweet text.

Approach

Our project is centered around the Twitter US Airline Sentiment dataset on Kaggle, comprising tweets from 2015 mentioning major airline handles. The dataset is preclassified as positive, negative, or neutral, with additional sentiment confidence ratings. To build a predictive model, we initiated data cleaning by removing usernames, emojis, hashtags, and common stopwords, creating a condensed, meaningful version. Tokenization indexed each word as a token for modeling, and term frequency-inverse document frequency (TF-IDF) calculation further extracted features indicating term importance. Four machine learning algorithms—Naive Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest—are explored for sentiment analysis. Naive Bayes uses probabilistic modeling, Logistic Regression estimates sentiment probability, SVM excels in complex sentiment patterns, and Random Forest leverages ensemble learning for accuracy.

Following the identification of the best-performing model, we then apply it to recent tweets. Our objective is to create airline profiles based on sentiment analysis, identifying the ratio between negative, positive and neutral sentiment. Gathering tweets over time will enable the creation of a time series dataset for tracking sentiment changes. In summary, our approach encompasses comprehensive model exploration, data preparation, and feature extraction. Continuous optimization aims to enhance accuracy and robustness in sentiment analysis, ultimately utilizing a fully developed model for predicting sentiment from recent tweets and gaining insights into current airline sentiments.

Explaining TF-IDF

To construct predictive models from text-based features, such as those extracted from tweets, the initial step involves tokenizing words to make them accessible for model analysis. Text Sentiment Analysis, a prevalent natural language processing (NLP) technique, plays a vital role in automating the assessment of subjective information within textual content. It aims to discern expressed sentiments, spanning from positive satisfaction to negative discontent, providing valuable insights for applications like marketing, social media monitoring, customer feedback analysis, and financial prediction (Taboada, 2016).

At the core of sentiment analysis lies the crucial concept of TF-IDF (Term Frequency-Inverse Document Frequency), a statistical measure evaluating the significance of words in documents by considering their frequency within a document (TF) and rarity across a population (IDF). TF-IDF effectively addresses the challenge of term relevance, highlighting meaningful words while diminishing the importance of common ones. Machine learning, particularly probabilistic classifiers like Naïve Bayes and linear classifiers such as Support Vector Machines (SVM), leverages TF-IDF insights to construct predictive models for sentiment analysis. Naïve Bayes employs mixture models to calculate class posterior probabilities based on word distributions, while SVM establishes optimal linear separators using TF-IDF data, enhancing the precision of sentiment classification (Ramos, 2003; Medhat, Hassan, & Korashy, 2014).

In conclusion, Text Sentiment Analysis stands as a crucial tool for unraveling emotional content within textual data, finding application in various domains. The incorporation of TF-IDF, along with machine learning techniques like Naïve Bayes and SVM, not only advances accuracy but also enhances the efficiency of sentiment analysis, contributing to comprehensive understanding and interpretation of textual sentiments.

Data Overview:

Source of Data and Information

Our project revolves around the Twitter US Airline Sentiment dataset, available on Kaggle, containing tweets from airline passengers in 2015, with a focus on major airline handles like Delta, United, Southwest, American, Virgin, and US Airways. Each tweet is preclassified as positive, negative, or neutral, with associated negative classifications specifying reasons such as 'Bad Flight' or 'Customer Service Issue.' The dataset, comprising 14,640 rows, features essential variables such as tweet_id, airline_sentiment, airline_sentiment_confidence, negative reason, negativereason_confidence, airline, retweet_count, and text.

Obtained from Kaggle's Data for Everyone Library, the dataset serves to explore sentiments expressed by travelers in February 2015. The dataset is well-organized, available in both SQLite and CSV formats, with our project opting for the latter. Notably, we made thoughtful decisions in data preprocessing, excluding columns like 'negativereason_gold,' 'airline_sentiment_gold,' and 'tweet_coord' due to limited useful information. Additionally, we omitted 'tweet_location,' 'tweet_created,' 'user_timezone,' and 'name' to streamline our dataset for a more focused and meaningful analysis. The dataset's structured design, encompassing crucial features, aligns with our project's goal of sentiment analysis and predictive modeling.

Data Cleaning Processes

The data cleaning process for the tweets was executed using Python within a Jupyter Notebook. This step played a pivotal role in preparing the tweets for analysis, addressing the informal nature of tweet data, often laden with imperfections that could act as noise detrimental to model performance. The tweets underwent a normalization process, ensuring the accurate analysis of their core content.

The initial phase of the cleaning procedure involved the removal of usernames, identified by the "@" symbol. Although usernames are integral for social interaction, they contribute minimally to the analytical aspects of the tweets. Additionally, non-unicode characters were systematically eliminated to enhance data quality. Subsequently, punctuation marks, devoid of substantive meaning, were targeted for removal to streamline the text for more in-depth analysis. Simultaneously, links within the tweets were eliminated, and the text was standardized to lowercase, guided by regular expressions, further refining the data.

The concluding step of the cleaning process centered on the elimination of stopwords, words carrying limited semantic weight. Common words such as 'the,' 'a,' and 'to' primarily function as linguistic fillers and contribute minimally to the text's meaning. The Natural Language Toolkit (nltk) library provided a list of these stopwords (157 in total), which were systematically purged from the tweets. To further refine the data, we explored methods to map emojis and emoticons to words, recognizing their potential to convey sentiment.. We also used stemming techniques to reduce words to their root form, thereby potentially enhancing the model's performance by presenting a more unified representation of words, facilitating easier detection of sentiment and patterns.

This meticulous procedure led to the establishment of a pipeline that extracted crucial features, laying the groundwork for the development of predictive models. The outcomes of the feature extraction process are summarized in Table 1.

	Raw Tweet Text	Extracted Text
1	@AmericanAir any delays on tonight's flight from DFW to SAT? I have family members on board.	delays tonights flight dfw sat family members board
2	@SouthwestAir is there a way to know who checked my bag on the curb? She was awesome!!! And want to be sure she gets a high five!	way know checked bag curb awesome want sure gets high five
3	@united I'm on standby for my connecting flight. 1st on the list. NudgeNudge I'm told there are four slots open. Save it for me. Please.	standby connecting flight st list nudgenudge told four slots open save me please
4	@SouthwestAir What an awesome flight Dallas 2 NY. Virgin America refused bc of my child's peanut allergy but u guys didn't. Thanks! 👍😊	awesome flight dallas ny virgin america refused bc childs peanut allergy u guys didnt thanks thumbs up smiling face smiling eyes

Table 1. Examples of Feature Extraction in Tweet Cleaning and Word Extraction

Exploratory Analysis and Data Visualization

After excluding data with a confidence ratio of less than 1, we obtained a dataset comprising 10,446 rows, which should suffice for training prediction models. To gain a deeper understanding of this dataset, we conducted an initial exploratory data analysis, uncovering several noteworthy insights.

In **Figure 1**, we observe the distribution of sentiments assigned to the tweets. Negative tweets significantly outnumber neutral and positive sentiments, with nearly five times as many negative tweets as the other two combined.

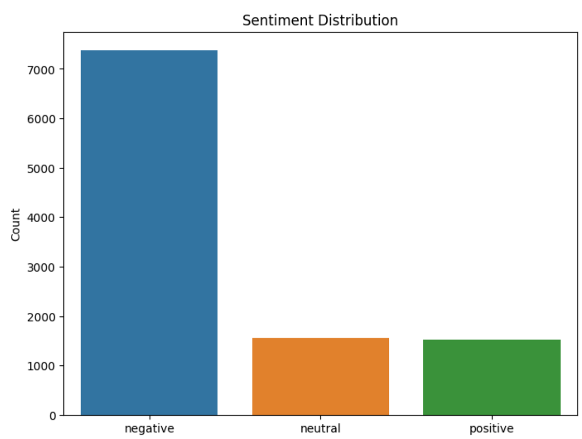

Figure 1. Sentiment Distribution of all cleaned tweets

Figure 2 illustrates the distribution of negative reasons, where 'Customer Service Issue' emerges as the most prevalent reason, accounting for approximately 2,500 tweets. 'Late Flight' and 'Bad Flight' follow as the second and third most common reasons, with around 1,500 and 750 tweets, respectively.

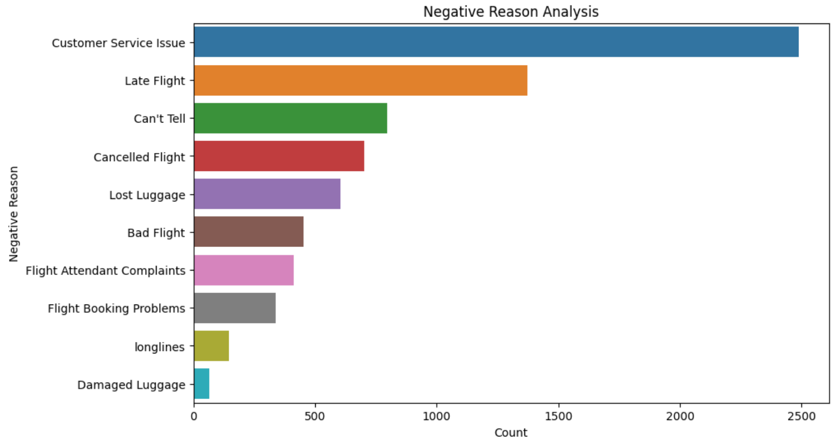

Figure 2. Count of Negative Reason

Figure 3 presents the Text Sentiment counts for each airline. United Airlines has the highest number of tweets and the most negative tweets. Southwest Airlines boasts the highest number of positive tweets. Interestingly, while Virgin America has the fewest tweets, the Normalized sentiment count relative to the total number of tweets within an airline (**Figure 4**) reveals that Virgin America has the highest ratio of positive tweets and the lowest ratio of negative tweets, whereas US Airways has the highest ratio of negative tweets.

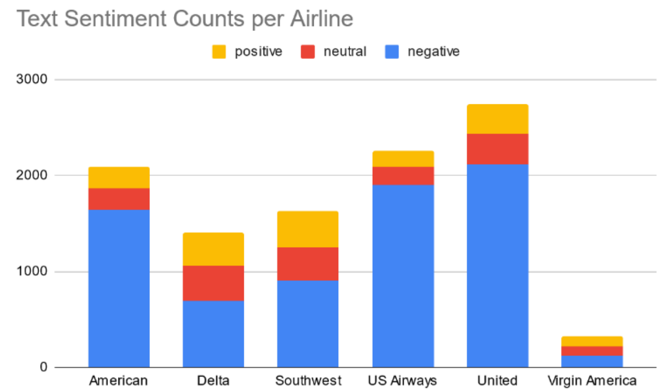

Figure 3. Text Sentiment Counts per Airline

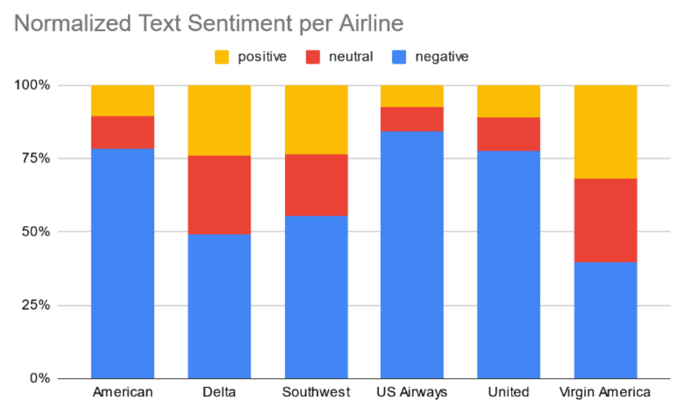

Figure 4. Normalized Text Sentiment Counts per Airline

Figure 5 showcases the normalized sentiment distribution across the top 7 time zones with the most tweets. Upon visual inspection, it appears that there is no significant difference in the proportion of sentiment-related tweets among different time zones. Overall, negative tweets appear to dominate the sentiment distribution

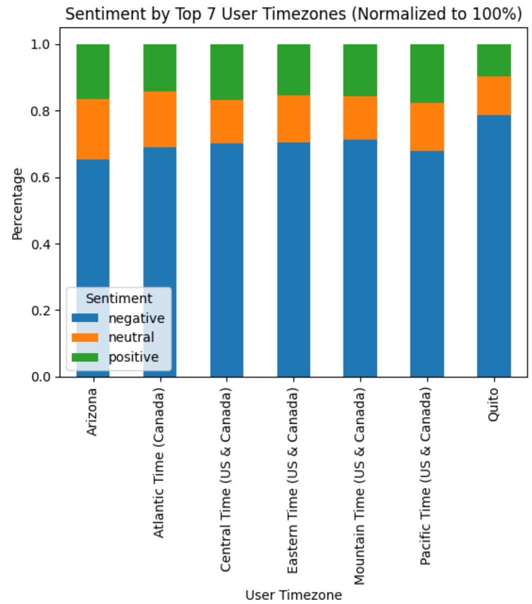

Figure 5. Normalized Sentiment by User Time zone

Lastly, in **Figure 6**, we observe the distribution of text length by sentiment. It's noteworthy that negative tweets tend to be the longest, with neutral and positive texts being shorter and of similar length. Negative tweets typically range from 90 to 125 words in length, suggesting that they tend to be more detailed or elaborate compared to neutral and positive tweets.

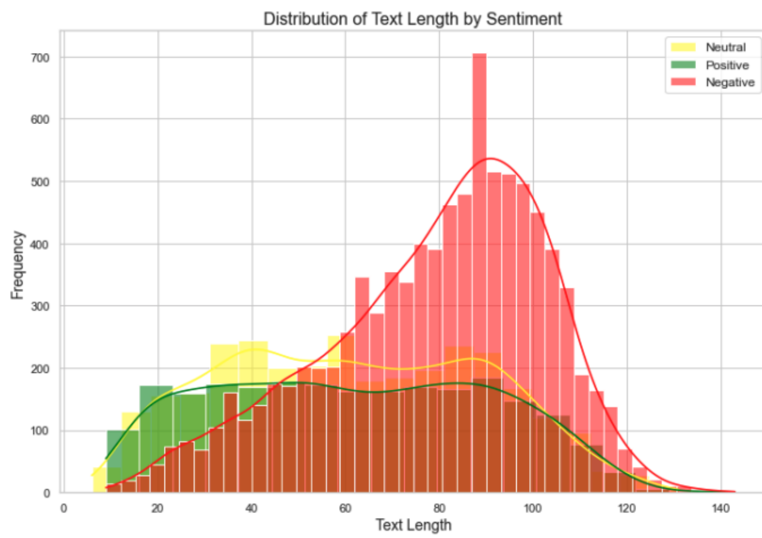

Figure 6. Distribution of Text Length by Sentiment

Model Development:

Screening Models

To determine the best model for sentiment analysis of tweets, we trained four models against the same prepared dataset. Our initial step involved taking the condensed tweets resulting from the data cleanup process and filtering the dataset to retain only those entries with a sentiment confidence score of 1. We then addressed missing data by replacing it with empty strings. The dataset was divided into two primary components: the tweet content and their corresponding sentiment labels. To equip the models for analysis, TF-IDF vectorizer was initiated to transform the textual data into numerical features, and it was fitted to the training data to decode the text data into a format amenable to machine learning. Following this, we proceeded to train four distinct machine learning models for sentiment analysis: Logistic Regression, Random Forest, Support Vector Machine (SVM), and Multinomial Naive Bayes. To facilitate model evaluation and training, we further divided the dataset into training and testing subsets, allocating 20% for testing. The accuracy of each model was ascertained by comparing their predicted sentiment labels to the actual sentiment labels and measuring the proportion of correct predictions using the 'accuracy_score' function from scikit-learn's metrics module. In addition to accuracy, we generated comprehensive classification reports for each model, shedding light on precision, recall, and F1-score for each sentiment category.

The results indicate that the Support Vector Machine model achieved the highest accuracy (84.9%), followed by Logistic Regression (84.2%) and Random Forest (83.3%). The Naive Bayes model had the lowest accuracy (75.1%). In terms of classification reports, SVM and Logistic Regression performed well, with favorable precision and recall values, while Random Forest showed slightly lower performance, especially in the "neutral" category. Naive Bayes exhibited lower precision and recall values, particularly in the "neutral" and "positive" categories, indicating its limitations in correctly classifying those sentiment labels.

Model	**Precision**		
	Negative	**Neutral**	**Positive**
Logistic Regression	0.842	0.755	0.900
Random Forest	0.846	0.741	0.809
SVM	0.864	0.710	0.849
Naive Bayes	0.740	0.850	0.960
Model	**Recall**		
	Negative	**Neutral**	**Positive**
Logistic Regression	0.977	0.369	0.662
Random Forest	0.960	0.388	0.666
SVM	0.959	0.427	0.736
Naive Bayes	1.000	0.070	0.220
Model	**F1-Score**		
	Negative	**Neutral**	**Positive**
Logistic Regression	0.905	0.496	0.763
Random Forest	0.899	0.510	0.730
SVM	0.909	0.533	0.789
Naive Bayes	0.850	0.140	0.360
Model	**Accuracy**		
Logistic Regression	0.842		
Random Forest	0.833		
SVM	0.849		
Naive Bayes	0.750		

Table 2: Overview of Performance for Four Statistical Models on a Common Dataset

Optimization of SVM

A Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. Fundamentally, the SVM works by finding the optimal hyperplane in a high-dimensional space that best separates data points of different classes. In the context of sentiment analysis, an SVM aims to delineate the decision boundary between positive and negative sentiments, with the goal of accurately classifying new, unseen data. The Radial Basis Function (RBF) kernel, also known as the Gaussian kernel, is a pivotal component of SVMs. The RBF kernel measures the similarity between data points in a transformed feature space, mapping input data into a higher-dimensional domain where non-linear relationships can be effectively captured. This transformation is achieved by assigning a weight to each data point based on its distance from a selected reference point, or "landmark."

In the optimization process detailed above, the chosen SVM parameters (‘C': 10, 'gamma': 1, 'kernel': 'rbf') hold specific roles in fine-tuning the model's performance. The regularization parameter 'C' influences the trade-off between achieving a smooth decision boundary and accurately classifying training data, with larger values of 'C' favoring a more precise fit. The gamma parameter ('gamma': 1) determines the reach of influence each data point has, with smaller values resulting in a more extensive influence. The 'rbf' kernel signifies the use of the Radial Basis Function kernel, allowing the SVM model to effectively capture non-linear relationships within the data. In essence, this configuration reflects a well-balanced SVM model with optimized parameters, poised to deliver accurate sentiment predictions in the realm of Twitter text analysis.

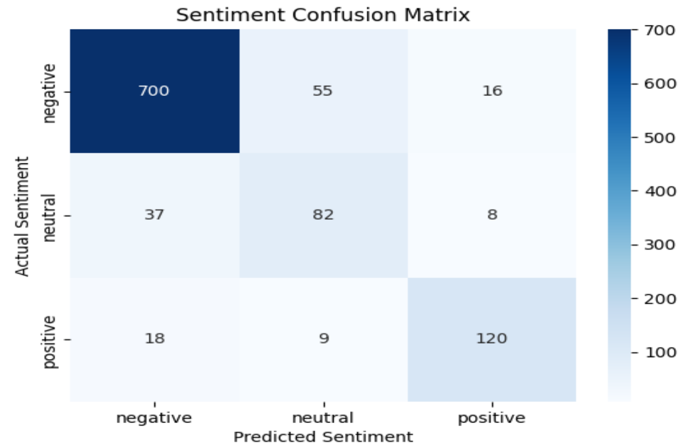

Figure 7: Sentiment Confusion Matrix of optimized SVM model

In the pursuit of optimizing a Support Vector Machine (SVM) model for sentiment analysis, an exhaustive approach was undertaken using GridSearchCV in conjunction with a for loop. This process involved varying the number of features (max_features) for the TF-IDF vectorizer, ranging from 3000 to 5100 in increments of 25. The vectorized text data was then utilized to train the SVM model, and a comprehensive parameter grid was defined, exploring different combinations of regularization parameter (C), kernel coefficient (gamma), and kernel type (linear, poly, rbf, sigmoid). The GridSearchCV algorithm systematically traversed this parameter space, employing 10-fold cross-validation to evaluate model performance for each combination. After running for over a day, the optimal combination of parameters was identified, achieving a maximum k-fold accuracy of 0.8622. Notably, this accuracy was associated with a specific number of features, indicating that the optimal balance between feature richness and model complexity was achieved when utilizing a TF-IDF vectorizer with 4425 features.

In conclusion, the code demonstrated a rigorous exploration of hyperparameter space to identify the configuration that maximizes the SVM model's predictive performance. The chosen parameters, {'C': 10, 'gamma': 1, 'kernel': 'rbf'}, reflect a regularization strength of 10, a kernel coefficient of 1, and the radial basis function (RBF) kernel type. These parameters collectively contribute to an SVM model that strikes a balance between bias and variance, ensuring robust sentiment analysis on the given dataset. The significance of these findings lies in the tuning process, highlighting the importance of parameter optimization in enhancing the effectiveness of machine learning models for sentiment analysis tasks.

Assessment of Model Performance

The Support Vector Machine (SVM) model's performance was evaluated using various metrics, providing a comprehensive understanding of its effectiveness in sentiment analysis. The model exhibited an overall accuracy of 88.04%, showcasing its ability to correctly classify sentiments across three categories: negative, neutral, and positive. The precision, recall, and F1-score for each sentiment category further elucidate the model's discriminative power. Notably, the SVM model demonstrated high precision and recall for negative sentiments (0.89 and 0.97, respectively), indicating a strong ability to accurately identify and distinguish negative sentiment instances. However, the model exhibited slightly lower precision and recall for neutral sentiments (0.81 and 0.49) and positive sentiments (0.88 and 0.78), suggesting some challenges in precisely identifying these sentiments.

The confusion matrix, visualized as a heatmap, provides a clear illustration of the model's performance across sentiment categories. Each cell in the matrix represents the number of instances predicted for a given sentiment compared to the actual sentiment. The diagonal elements indicate correct predictions, while off-diagonal elements reveal misclassifications. The heatmap visually emphasizes the model's proficiency in accurately predicting negative sentiments, with fewer misclassifications in comparison to neutral and positive sentiments. This evaluation process underscores the SVM model's overall effectiveness in sentiment analysis, while the detailed metrics contribute valuable insights into its specific strengths and areas for potential improvement.

Interpretation of Model Results

In our exploration of sentiment analysis on airline-related tweets, we delved into the underlying features that significantly influence sentiment predictions within an SVM model. Employing a linear kernel and the TF-IDF vectorizer, we transformed the raw textual data into a format suitable for training the model. The focus was on extracting coefficients associated with each feature (word) to unravel their impact on sentiment classification across three categories: negative, neutral, and positive. Through a sorting process based on these coefficients, we pinpointed the most crucial words indicative of each sentiment class. Words like "cancelled" and "delayed" dominated the negative sentiment class, capturing common grievances related to flight experiences. On the positive end, expressions of appreciation such as "thank" and "great" stood out. This feature centric analysis offers a deeper understanding of language dynamics, providing valuable insights into the subtle factors shaping sentiment predictions in the realm of airline-related tweets. Table 2 summarizes the top 10 features for each sentiment category.

Sentiment Class	**Top 10 Features**
negative class	hours, cancelled, call, delayed, luggage, hold, worst, hour, hrs, flightled
neutral class	fleets, dm, tomorrow, fleek, hi, cold, chance, question, destinationdragons, photo
positive class	thank, thanks, great, love, awesome, best, amazing, good, always, appreciate

Table 3: A Summary of the top 10 features for each Sentiment Class

Model Demonstration:

To demonstrate the practical implementation of our developed model, we aimed to analyze recent datasets of tweets mentioning airlines, specifically Delta Airlines and American Airlines, with the overarching goal of showcasing how our model can effectively track sentiment over time. This endeavor is crucial for companies to gauge how their services are perceived by customers and make informed, timely adjustments. However, a challenge arose as the recent Twitter policy updates prohibited the free use of its API for scraping tweets. To circumvent this, we employed the Python library, Tweepy, which enables the downloading of tweets from the most recent searches.

Using Tweepy, we successfully obtained essential tweet information such as ID, date, author, and text, scraping data daily from November 14 to November 28, 2023, encompassing the Thanksgiving holiday period, in the hopes that the high travel period will result in an increase of negative sentiment tweets. Following data collection, we cleaned the dataset by filtering for unique tweet IDs. To ensure the exclusion of posts from bot or company accounts, we further filtered out tweets from users with more than 5 posts in the dataset. Subsequently, we applied the same script used in the initial model development to extract valuable features from the tweet texts and eliminate nonessential words. This process yielded a refined database comprising approximately 990 entries for Delta Airlines and 1100 entries for American Airlines.

With the cleaned dataset, we then applied the previously developed SVM model, trained on the training data obtained from Kaggle, to predict whether the scraped tweets were positive, negative, or neutral. The tweets were subsequently grouped by date, and the sentiment ratio for each day was calculated. This information was then visualized over time through plotting, shown in figure 8 and figure 9, providing a clear representation of how sentiments evolve across different time periods.

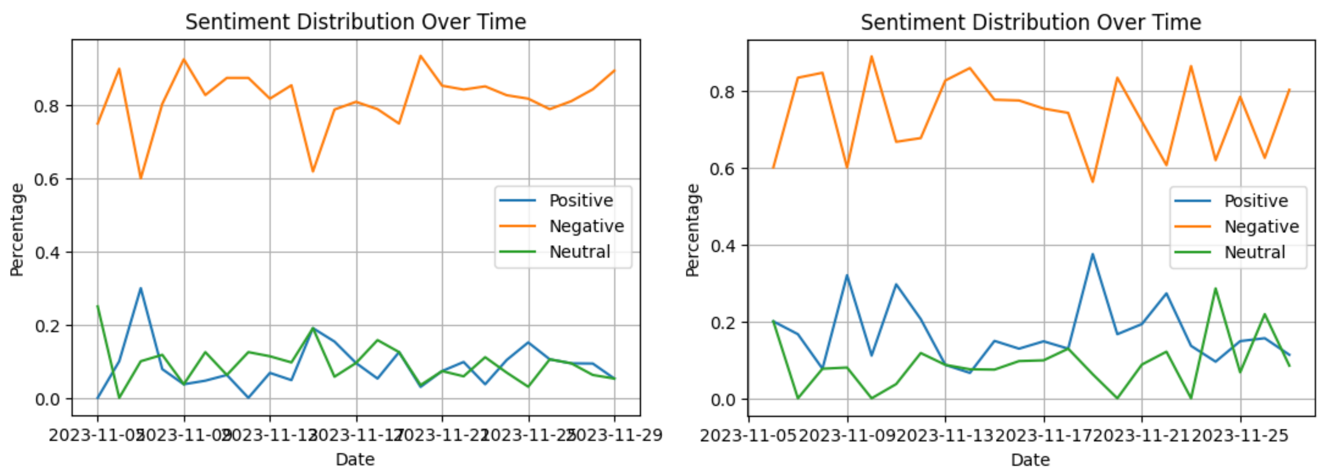

Figure 8: Sentiment ratio over a two week period over Delta Airline and American Airline Respectively

Results

Upon analyzing the produced data it was observed that the ratios of negative, positive, and neutral sentiment tweets exhibit little variation over the two-week period. This finding contradicts our initial hypothesis, which assumed an increase in negative sentiment during the Thanksgiving holiday. Unfortunately, our data fails to demonstrate a noticeable shift in sentiment during this time frame. We theorize that the lack of change in sentiment might be attributed to two potential factors.

Firstly, it is plausible that the quantity of tweets analyzed within the specified time period was insufficient to capture significant fluctuations in sentiment. The dataset's size may have influenced the ability to detect subtle changes, thus limiting the conclusiveness of our findings. Secondly, there is a possibility of inherent bias in the user-generated content, wherein individuals may be more inclined to share negative sentiments rather than positive ones. If this bias is prevalent, it could skew the overall sentiment analysis, leading to a more dominant presence of negative tweets.

Conclusions

While our endeavor to showcase the predictive capabilities of a statistical model for anticipating sentinel changes in real-time Twitter data may not have yielded the expected results, it successfully illustrates our capacity to measure sentiment from live Twitter feeds and monitor sentiment ratios dynamically. The ability to gauge sentiment in real-time provides valuable insights into the immediate reactions and perceptions of users, allowing for an agile response to emerging trends or issues. Despite the lack of a pronounced shift in sentiment during the observed period, this approach remains practical for businesses seeking to stay attuned to public sentiment.

Concluding Remarks:

Implications for Business Application

The practical applications of real-time sentiment analysis allows businesses to leverage models for ongoing brand monitoring, reputation management, and customer engagement. By continuously tracking sentiment ratios, companies gain a proactive understanding of how their products, services, or brand are perceived by the public. This allows for timely adjustments to marketing strategies, customer service approaches, or product offerings based on real-time feedback. Additionally, the insights derived from sentiment analysis can aid in identifying potential issues before they escalate, contributing to a more responsive and customer-centric business approach. Ultimately, the ability to measure sentiment in real time serves as a strategic advantage for businesses aiming to enhance customer relations and adapt swiftly to evolving market dynamics.

Recommendations for Future Work

In considering future work, it is recommended to expand the scope of data collection and analysis by incorporating a more extensive data collection. A corporation will have significant access to social media information that can be used to further develop accurate models to predict sentiment. A broader range of topics, time frames, and user demographics could provide a richer understanding of sentiment dynamics. Additionally, refining the model's features and parameters may enhance its predictive accuracy. Exploring advanced natural language processing techniques and deep learning approaches could further elevate the model's ability to discern nuanced sentiments. Investigating the impact of external events or contextual factors on sentiment changes could offer valuable insights for businesses. Finally, conducting a comparative analysis with alternative sentiment analysis models and incorporating real-time feedback mechanisms would contribute to refining and validating the effectiveness of the proposed statistical modeling approach for predicting sentiment from tweets in a business context.

Work Cited

1. Ayat, N.E., Cheriet, M., & Suen, C.Y. (2005). Automatic model selection for the optimization of SVM kernels. *Pattern Recognition, 38*(10), 1733-1745. https://doi.org/10.1016/j.patcog.2005.03.011
2. Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics, 2*(1), 325-347. https://doi.org/10.1146/annurev-linguistics-011415-040518
3. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal, 5*(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011
4. Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries.
5. Tao, K. M. (1993). A closer look at the radial basis function (RBF) networks. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA,* pp. 401-405 vol.1, doi: 10.1109/ACSSC.1993.342544.
6. Salinca, A. (2015). Business Reviews Classification Using Sentiment Analysis. *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania,* pp. 247-250, doi: 10.1109/SYNASC.2015.46.