Airline Sentiment Analysis

Kierra Dangerfield

Table of Contents

Introduction

As a business changes, so do their customer interests and sentiments. Businesses can use sentiment analysis to track customer sentiment when launching a new product or changing the price of their products. Tracking sentiment analysis can help businesses improve products and services. Twitter is a great way to gather sentiment from customers because the platform was created to communicate and stay connected through the exchange of quick, frequent messages. Using Twitter is a great way to keep track of customer sentiment about your product, service, or business on social media. It also can help detect angry customers or negative comments. A business can use this for marketing. Using Twitter for your sentiment analysis can provide valuable insights that drive business decisions. What do customers like about your business? What do customers hate or love about your products? A business can also use this for customer support. When a model can predict if a tweet is positive or negative, a bot can respond with an appropriate message to help a customer. My goal is to gather the sentiment analysis of a variety of airlines. I will get the predictive words that will determine a positive or negative sentiment.

DATA

The data was collected from Kaggle and based on tweets that are classified as positive, negative, or neutral. It originally had 14,640 rows and 15 columns. **"airline_sentiment_gold" , "negativereason_gold", "tweet_coord"** columns were removed because over 90% of the rows were null. The Twitter data was scraped from February 16, 2015 and February 24, 2015. I will mainly focus on just the positive and negative tweets. Below are the main columns I am working with.

- **Airline_sentiment**
    - sentiment of the airline, has 3 different sentiments: positive, neutral and negative
    - object data type
- **Negativereason** -
    - reason for the negative sentiment
    - Object data type
- **Airline**
    - name of the airline
    - Object data type

- **Text**
    - The tweet
    - Object

ANALYSIS

I. Distribution of Airline Sentiment

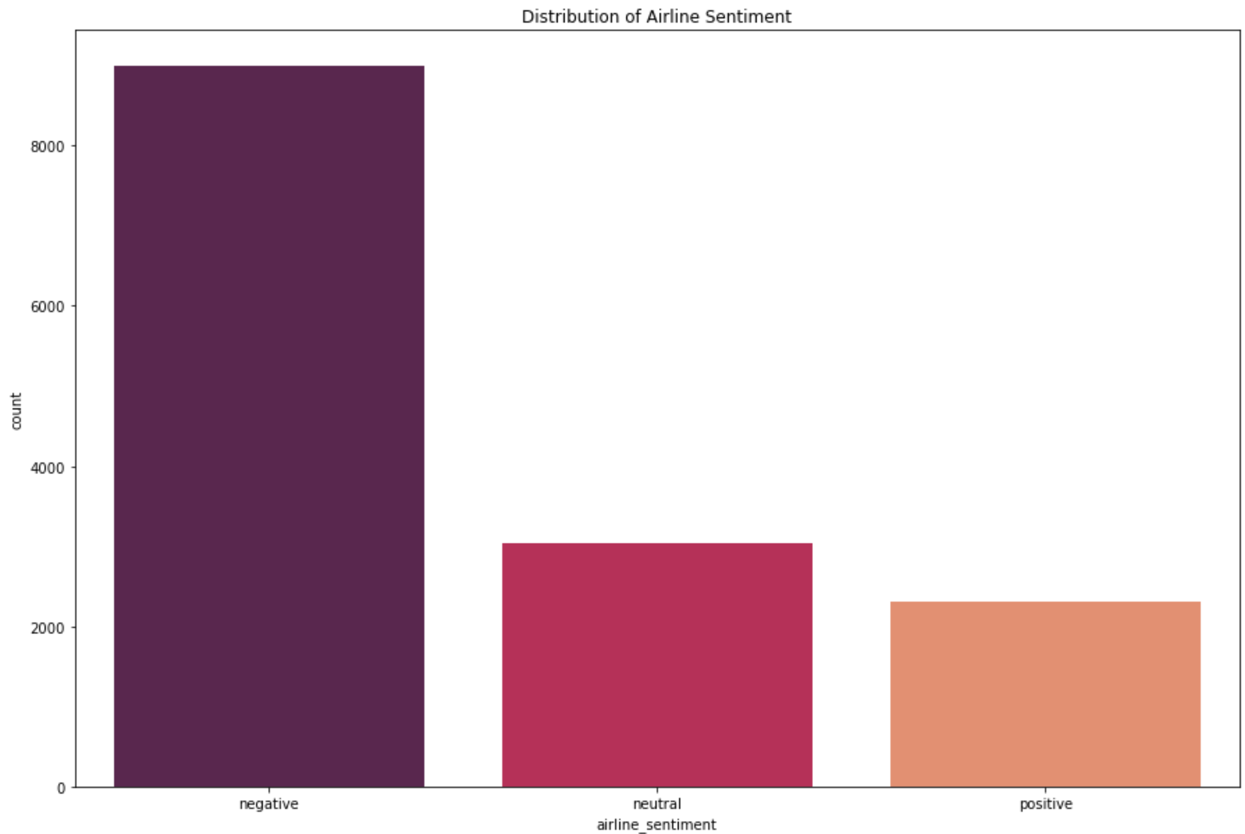

Figure 1 Distribution of Airline Sentiment

Figure 1 shows the distribution of airline sentiment. Most of the sentiments are negative. By looking at this distribution, I would say that there is an imbalance classification in this dataset.

II. Proportion of Airline Sentiment

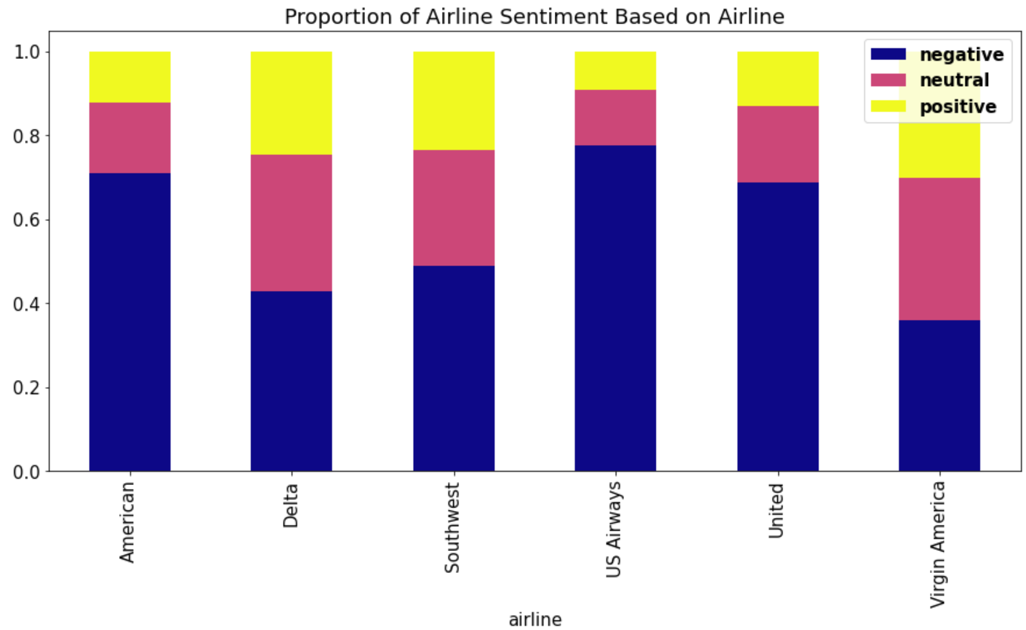

Figure 2 Proportion of Airline Sentiment

Figure 2 shows the proportion of different sentiments based on the airline. Most of the sentiments are negative. US Airways has the most percentage amount of negative sentiment. Virgin America has the least amount of positive sentiment, but it only accounts for 3.4% of the data .

III. Negative Reason for Negative Sentiment

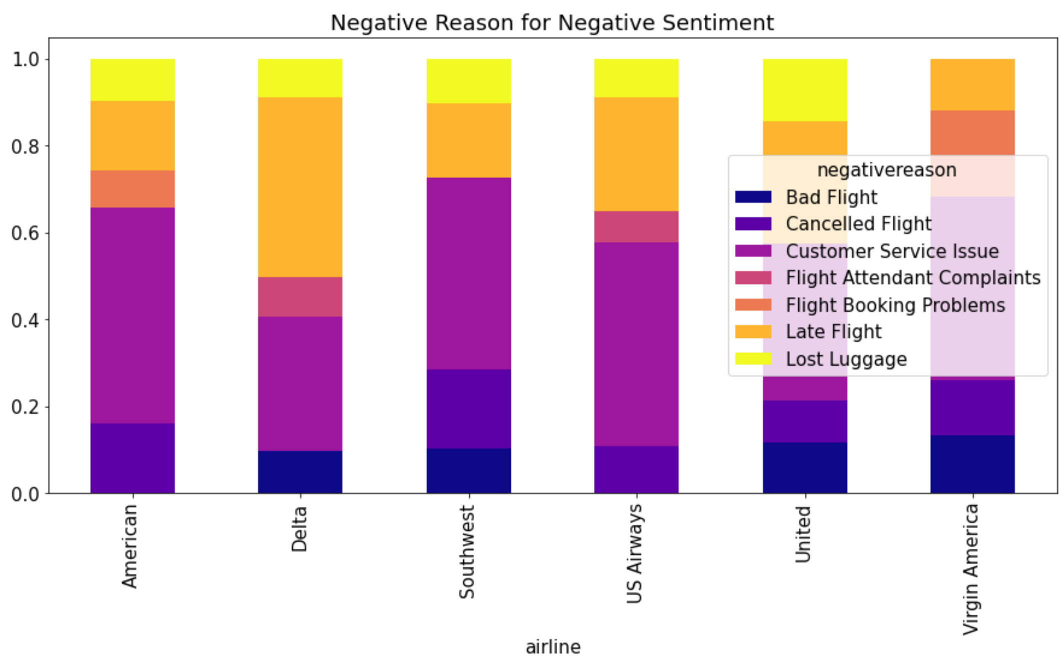

Figure 3 Top Negative Reasons for each airlines

Figure 3 shows the top reasons for a negative sentiment for each airline. Reasons like “Can’t Tell” and NaN values were removed. “Can’t Tell” reason is when someone does not specify the reason. Customer service issues and late/canceled flights are some of the top reasons customers give for the negative sentiment.

All of the airlines except for Delta have “Customer Service Issue” as the top negative reason for a negative sentiment. The customer service issues are a various range between issues outside and inside the airport. Customers have complained about calling and emailing customer support with no response, the website being down , and not being able to rebook a flight. Customers also have complained about not being able to check in and late flights without any communication.

Delta, United, and Virgin America had “Bad Flight” as one of the top reasons for negative sentiment. The bad flight complaints are mainly about their discomfort during their flight like the seats are uncomfortable, having to move seats without notice, and more.

Flight booking problems are another issue airlines are facing. Customers have claimed that their website was down for a period of time. Virgin America seems to have this complaint more compared to the other airlines. Overall customer service issues, late flights, and canceled flights are some of the top complaints across all of the airlines.

IV. Proportion of Word Count

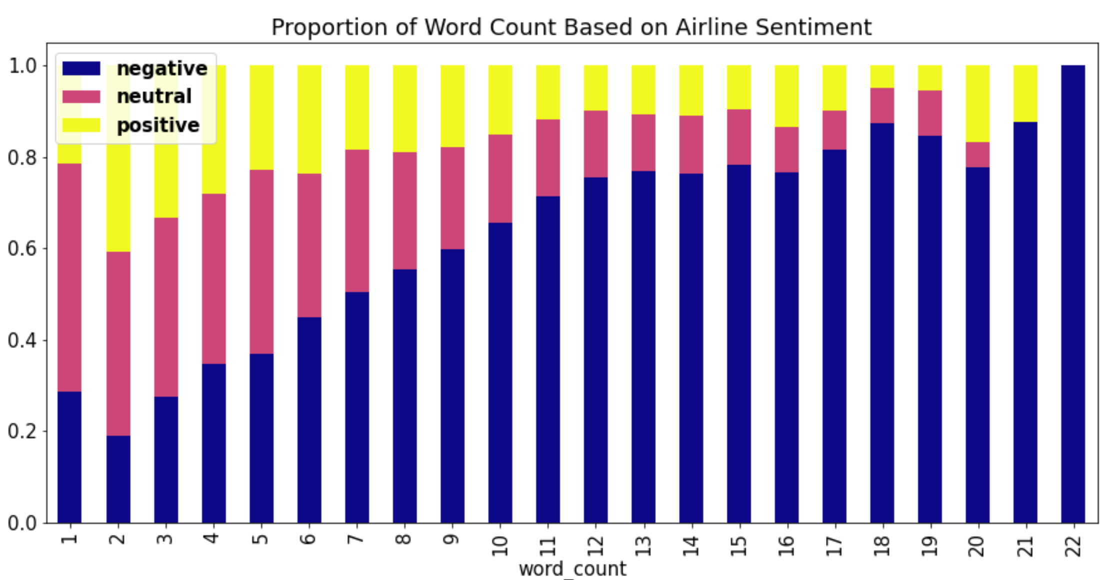

Figure 4 Distribution of Word Count

Figure 4 displays the proportion of word count in each tweet. The negative sentiment tweets have more words in their text. The tweets with a positive or neutral sentiment have the least amount of words per tweet.

V. Uppercase vs Airline Sentiment

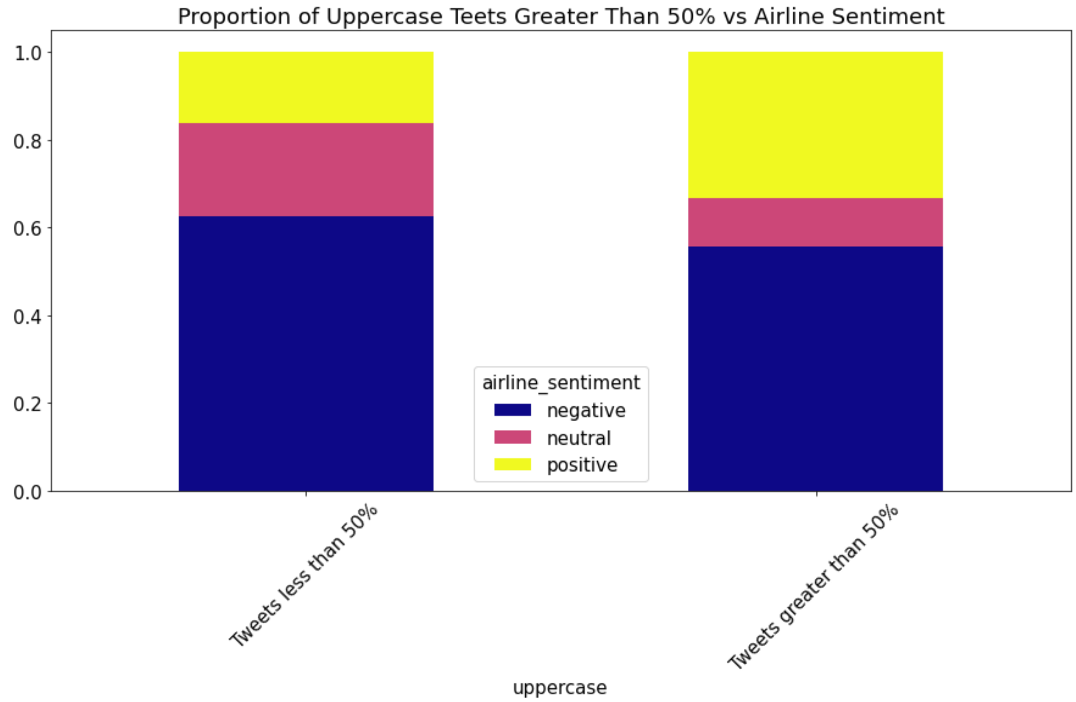

Figure 5 Proportion of Uppercase Text vs Airline Sentiment

Figure 5 is the proportion of airline sentiment grouped into whether the tweet is over 50% uppercase (True) or less than 50% (False). Between the two groups, the negative sentiment is roughly the same. The positive sentiment of the tweets that are greater than 50% uppercase has a higher proportion compared to the tweets that are less than 50% uppercase. I would say that there is not much of a correlation between uppercase tweets and negative airline sentiment based on this dataset.

VI. Emoji vs Tweet

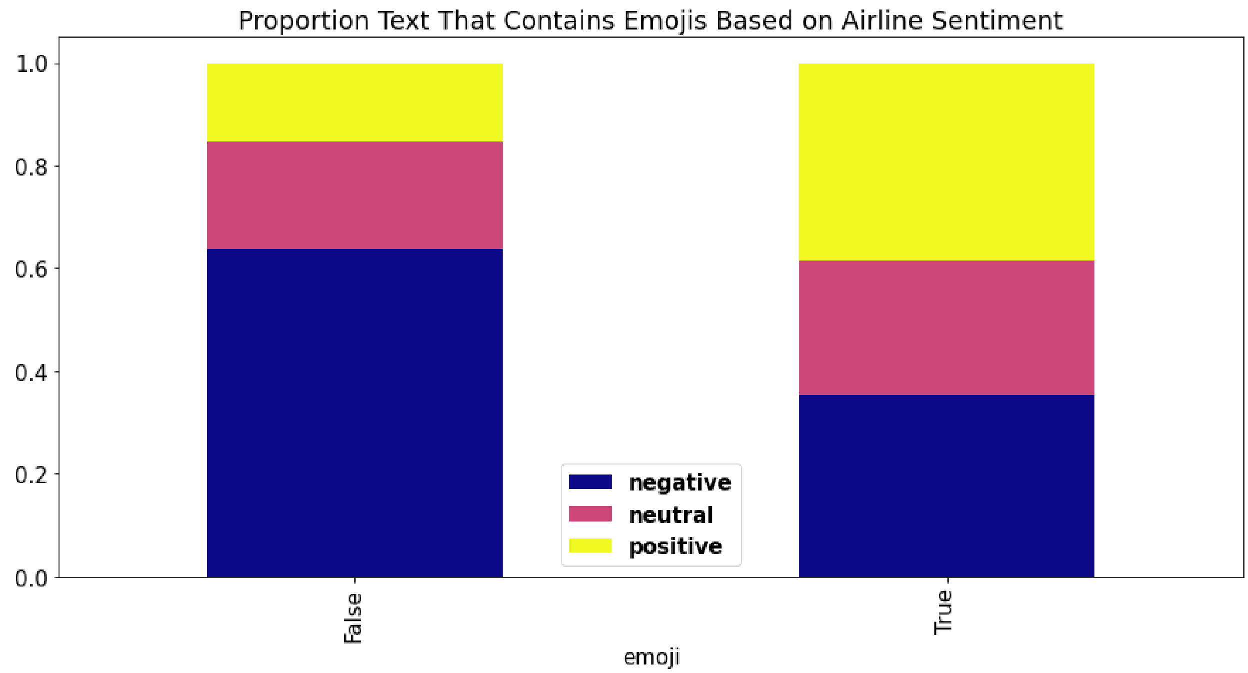

Figure 6 Text vs Emoji

Figure 6 shows the proportion of tweets that have and don't have emojis. Of the tweets that do not have an emoji in them, most of the sentiments are negative. Of the tweets that do have an emoji in them, most of the sentiments are positive. I would assume that customers that have had a negative experience are trying to quickly tweet to the airlines and would not care to add the extra emojis.

VII. Predictive Words

I found the top negative predictive words to predict the sentiment for each airline. In order to find the predictive words for each airline, I did the following:

1. Preprocessed the text.
2. Created a term-document matrix with CountVectorizer.
3. Trained a predictive model (Logistic Regression) on the matrix.
4. Created an identity matrix of the words which is a list of documents with the same vocabulary as the CountVectorizer
5. Used the trained classifier to make predictions on this matrix, and get the probabilities of each word.
6. Sorted the rows by predicted probabilities, and picked the top and bottom 10 rows.

A. United

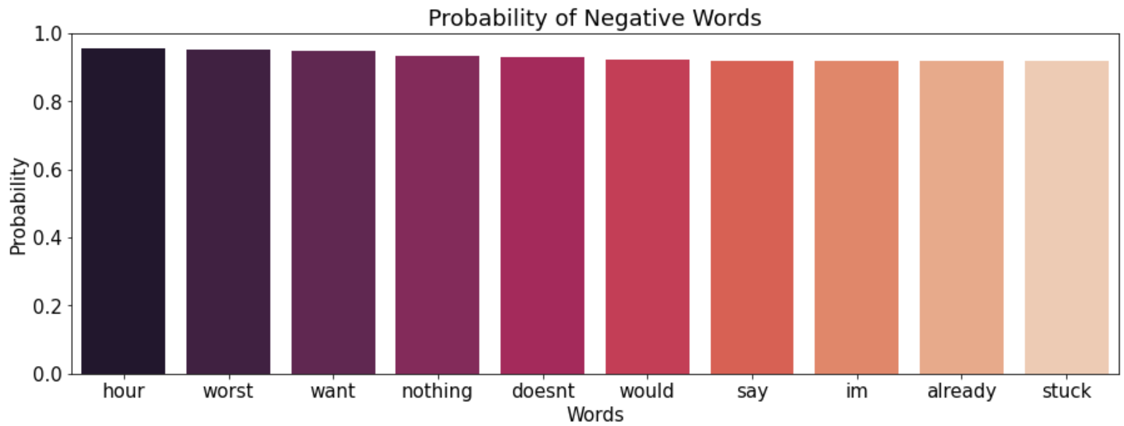

Figure 7.1 Negative Words Probability for United

Several United customers are experiencing late and delayed flights. They also were not able to reach out to customer service to resolve the issue. Some customers were on hold for hours and nothing was resolved. In some incidents customers lost their luggage and had a difficult time recovering them.

One unique issue United had was that several customers complained that they were stuck and were not able to reach their proper destination in a timely manner. This was because of the late and delayed flights. It caused several customers to miss their connecting flight as well.

Below are some tweets from customers who have had a negative experience.

1. “@united I tried but no one was available in bogota and everyone was rude in Houston. I was **stuck** for 35 hours because of you guys”
2. “@united I sure did. I had to drive a total of 3 hours to get my own bag. I'd like to explain that debacle but no one **wants** to talk to me.”
3. “@united Calls to 800# resulted in 2hrs of hold time & 2day wait to check suspect code share fare. **Nothing** investigated—my time wasted (2/2”

B. American

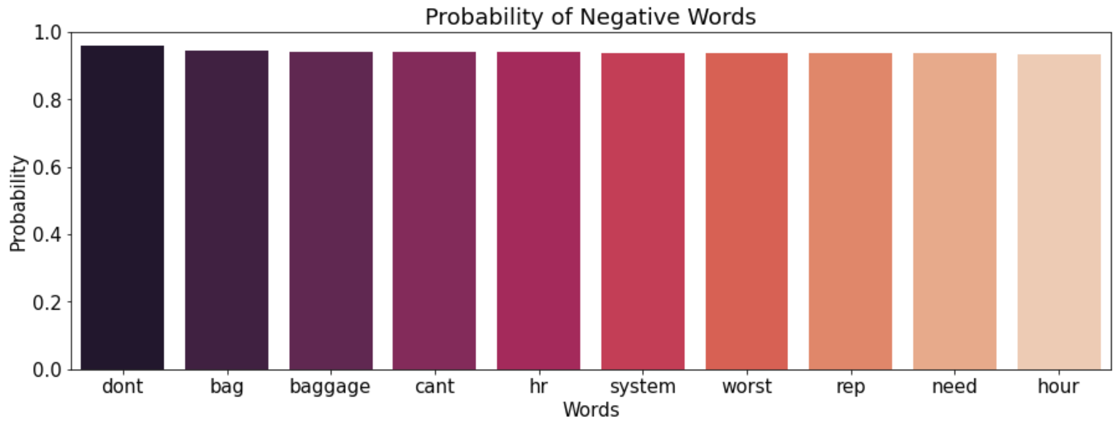

Figure 7.2 Negative Words Probability for American

American customers are experiencing a myriad of late flights and customer service issues. Several customers have complained about late flights with a lack of communication from the airline. Some had issues booking and rescheduling their flight. Customers have also experienced rude flight attendants.

A common problem for the airline is that customers are losing their luggage and having a hard time recovering them. One customer even complained about items being stolen from their luggage. This is one area that American Airlines can improve on.

Below are some tweets from customers who have had a negative experience.

1. "@AmericanAir extremely upset that your **baggage** handlers decide to go in my **luggage** and take my belongings"
2. "@AmericanAir but, what I can always rely on when I fly USAir or American is that employees will be **rude** and unhappy."
3. "@AmericanAir im tryin to book a flight but cant get ahold of anyone!"
4. "@AmericanAir Right. But more than two **hours** Late Flight, and it seems **due** to poor communication, which sounded like it was annoying on-plane staff"

C. Delta

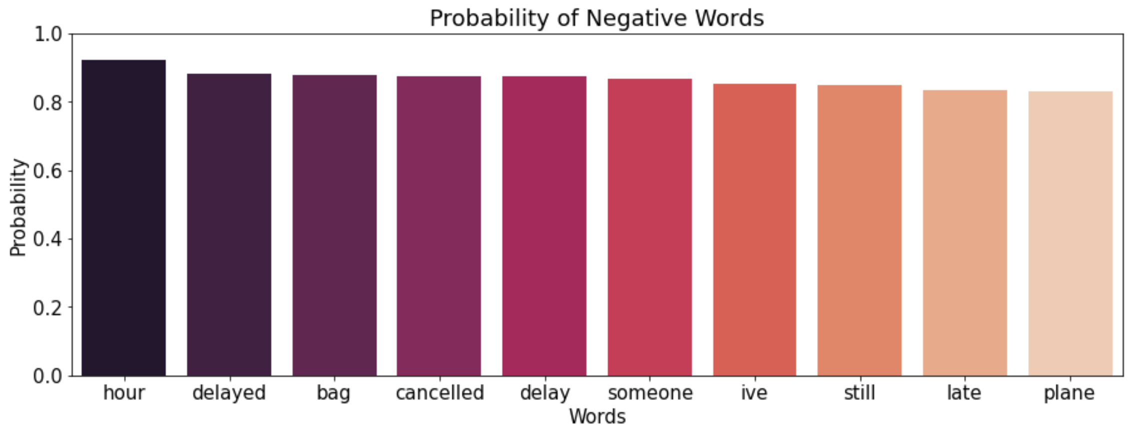

Figure 7.3 Negative Words Probability for Delta

Delta customers have mainly experienced delayed and canceled flights and customer service issues. Customers had trouble contacting customer service to get their issues resolved. Some had to call several times. Customers' flights were delayed causing issues with their connecting flight. Customers also had lost luggage that they had trouble recovering.

All of the other airlines had customer service issues as their top negative reason for a negative sentiment. Delta's top negative reason is late flights. This is something that Delta should look into to improve their customer sentiment.

Below are some tweets from customers who have had a negative experience.

1. "@JetBlue I had to call back five times to get **someone** on the phone who knew what they were doing. By that time my getaway went up by $200."
2. "@JetBlue upset with the lack of communication we've received for our "on time" flight 1170 out of MCO"
3. "@JetBlue what is the deal with flt 460 today? Departure keeps changing. When is it going why is it so **Late** Flight?"
4. "@JetBlue great job getting flight 28 in 10 minutes early. Too bad we're at 50 minutes and counting waiting for our **bags**."

D. Southwest

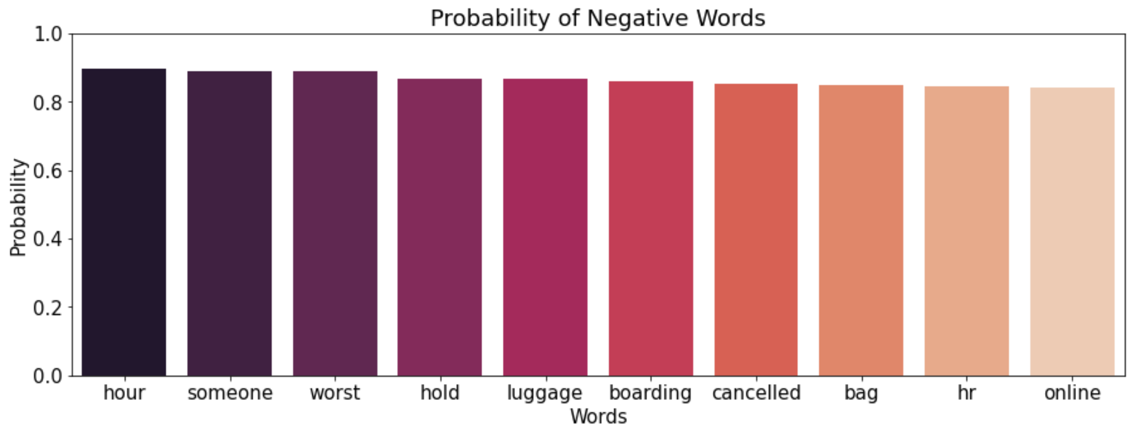

Figure 7.4 Negative Words Probability for Southwest

Southwest customers are experiencing several customer service issues. Many customers are reaching out to customer service without connecting with someone. Some are being put on hold for several minutes and not being able to connect with a representative in a timely manner. Customers are also stating that if they have to change their reservations, get boarding passes, etc, they have to call customer service instead of doing it online. Several customers have experienced delayed and late flights. Customers are also experiencing lost and damaged luggage.

Because Southwest customers can't change their reservations online, they have to call a customer service representative. It makes it difficult for customers to receive the help they need. Southwest can explore efficient customer service systems.

Below are some tweets from customers who have had a negative experience.

1. "@SouthwestAir Why can we no longer change trips with a companion **online**? Been doing it for years, now get message can't be done **online**?"
2. "@SouthwestAir can you have **someone** call me back? I have been on **hold** two times today for over 20 min and still haven't gotten through"
3. "@SouthwestAir and now our arrival is delayed nearly 5**hours**. Yeah an entire day in an airport is not my idea of a stress less vacation #fail"

E. US Airways

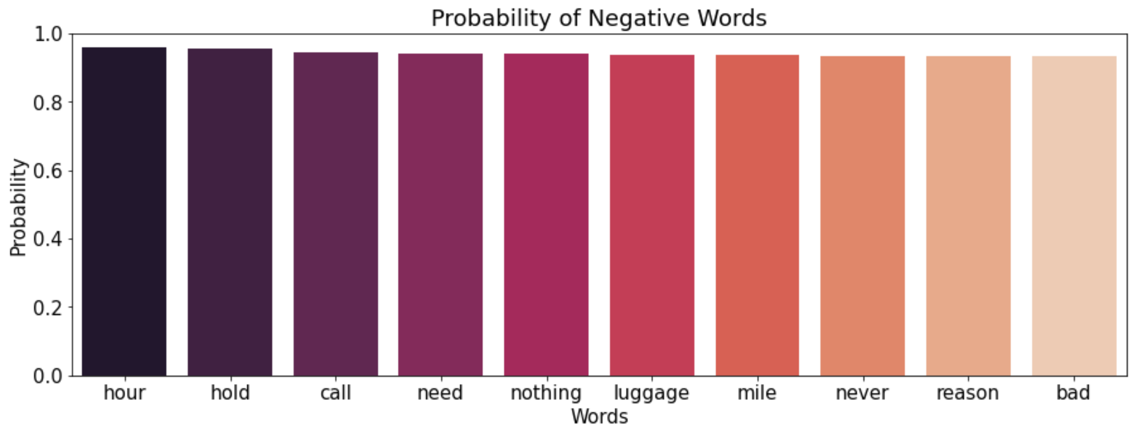

Figure 7.5 Negative Words Probability for US Airways

The cause of negative experiences for US Airways' customers is mainly customer service issues. Customers have complained about not being able to get in contact with customer service and have been on hold for hours. Some customers are also experiencing being hung up on. Customers are needing to communicate with someone for a variety of needs and they are not able to reach a customer service representative. Customers are also experiencing lost luggage with US Airways.

US Airways customers are having trouble connecting with a customer service representative like Southwest. However their customers are experiencing being disconnected from the customer service line. Improving the communication issues can improve their customer sentiment.

Below are some tweets from customers who have had a negative experience.

1. "@USAirways but I've been trying to **call** them since yesterday and I keep getting hung up on? Can you get me through to them??"
2. "@USAirways I've been on **hold** to change a date on a ticket for over 3 **hours**. Can someone please assist me? Unacceptable."
3. "@USAirways @Beamske how about a real live person talk to the person whose **luggage** was lost for 4 days and vacation wrecked . @yorkshire2002"

F. Virgin America

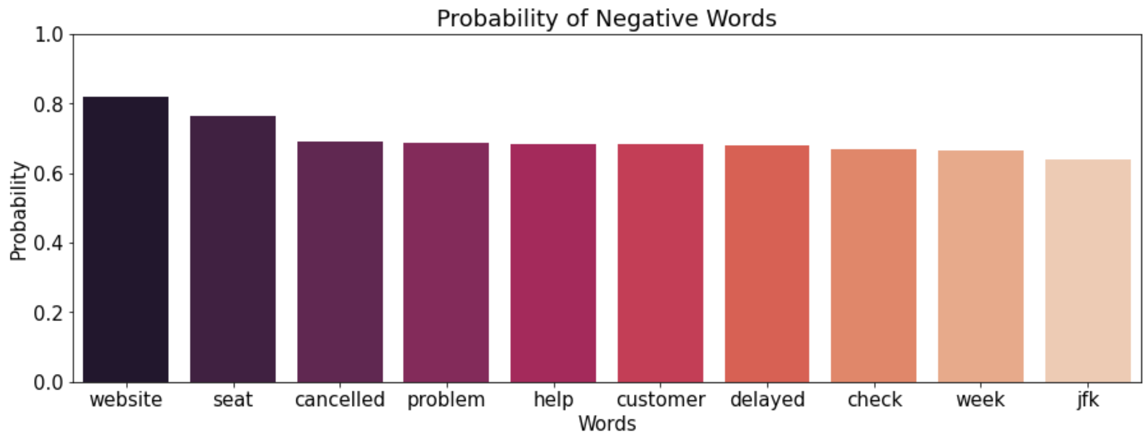

Figure 7.6 Negative Words Probability for Virgin America

A major issue Virgin America customers are experiencing are flight booking problems and canceled flights. With the flight booking problems, customers are trying to reschedule a flight and are not able to do it online. Several customers have complained about the website being down. Customers are also experiencing canceled flights. With customers' flights being canceled, it is hard for them to reschedule their flight when Virgin America's website is down or or has a poor interface.

Virgin America is the only airline that customers complained about the website being down. This is something that Virgin America can fix either by increasing their bandwidth, making frequent updates to their website, add new features or a better user experience, etc. Doing this can make it easier for customers to book and reschedule their flights and can improve their customer sentiment.

Below are some tweets from customers who have had a negative experience.

1. "@VirginAmerica Is it me, or is your **website** down? BTW, your new website isn't a great user experience. Time for another redesign."
2. "@VirginAmerica How do I reschedule my **Cancelled** Flightled flights online? The change button is greyed out!"
3. "@VirginAmerica Hey, first time flyer next **week** - excited! But I'm having a hard time getting my flights added to my Elevate account. Help?"

MODELING

For this model, I am predicting the probability that a tweet has a negative or positive sentiment. The goal is to make a model that airlines are able to use across business scenarios. After going through text preprocessing of lowercasing words, removing punctuation, whitespace, and special characters of the tweets, I tested both a CountVectorizer and Tf-idfVectorizer to vectorize the tweets. I used a Logistic Regression model for the base model.

Vectorizer	ROC-AUC Score	Accuracy
CountVectorizer	0.9577	92.18%
TfidfVectorizer	0.9583	90.94%

Table 1 Vectorizer

The TfidfVectorizer performed the best based on the ROC-AUC score and accuracy. This is the one I chose to vectorize with. ROC-AUC score is a measurement of the area under that curve and tells how much the model is capable of distinguishing between classes across different business scenarios as the ROC curve itself is threshold independent.

Model	Best Parameters	ROC-AUC Score
Logistic Regression 1	**C**: 1, **max_iter**: 1000	0.9583
Logistic Regression 2 - CountVect with n-grams	**C**: 1, **max_iter**: 1000	0.9570
Logistic Regression 3 - Tf-idf with n-grams	**C**: 1, **max_iter**: 1000	0.9517
Naive Bayes	n/a	0.9512
Random Forest	**max_depth:** None, **max_features:** auto, **n_estimators:** 400	0.9373

Table 2 Model Selection

Table 2 shows the different models I have chosen to test. The best parameters column is the column of best parameters based on the GridSearchCV. Based on the ROC-AUC score, the Logistic Regression 1 model performs the best.

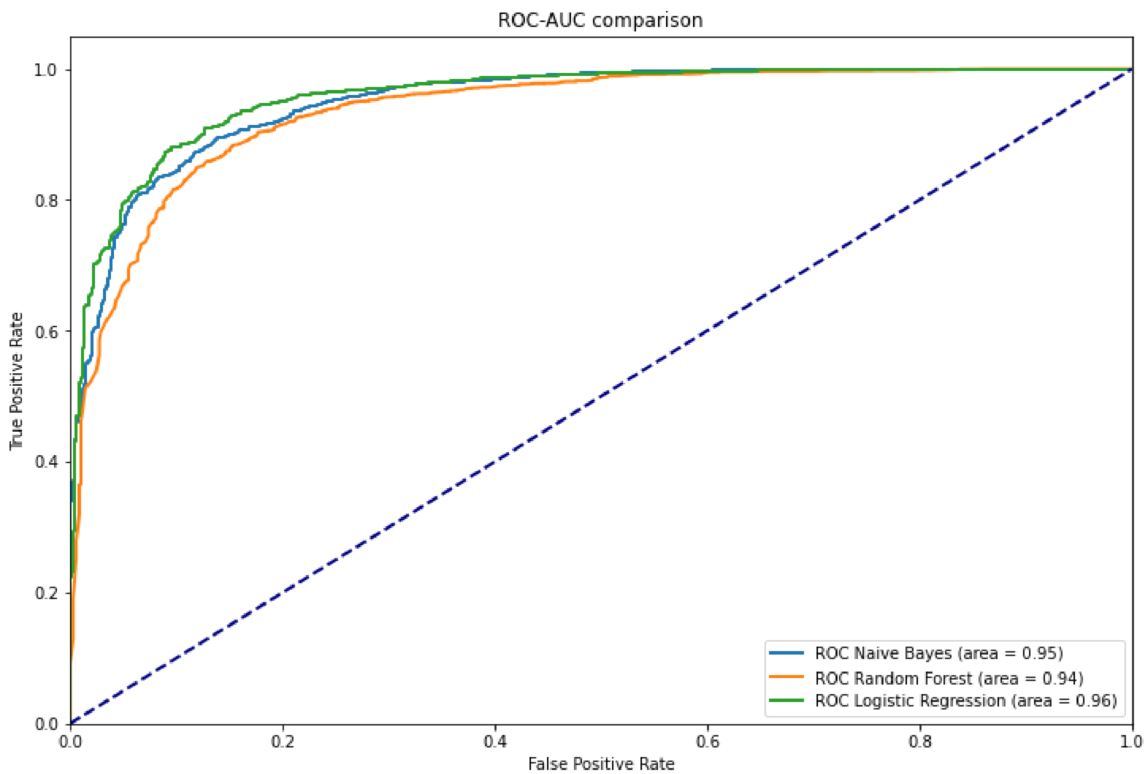

Figure 8 ROC-AUC Comparisons

Figure 8 shows the comparisons of the ROC-AUC between the Naive Bayes, Random Forest, and Logistic Regression 1 models. The Logistic regression model performed the best.

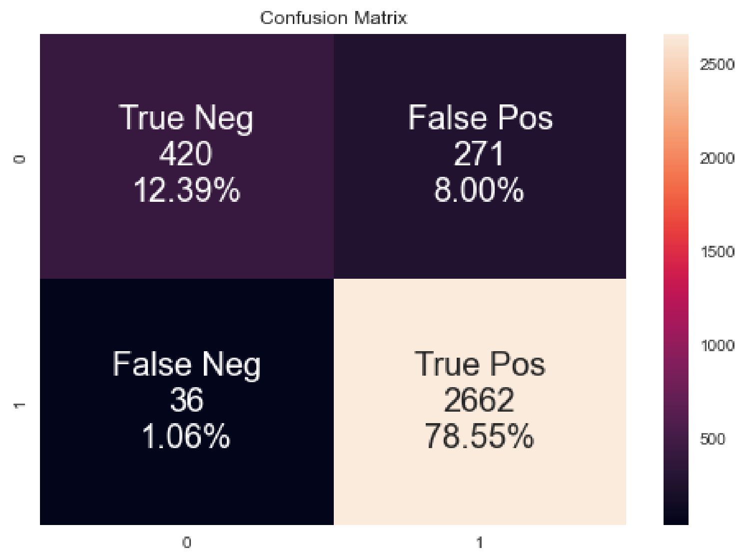

Figure 9 Confusion Matrix

Figure 9 shows the current confusion matrix of the current model with a threshold of 0.5. The recall score is 0.99. The precision score is 0.91. The accuracy score is 0.91. The F1 score is 0.95. Thresholds can be moved to better improve the model.

Depending on the business needs will determine what metric is most important to optimize and, therefore, which threshold would be most appropriate. The next step I took was to explore how the model might be thresholded appropriately for different business scenarios.

Business Case 1: Track Sentiment Analysis

Airlines can use this model to track overall customer sentiment on Twitter. This can help them understand how their company is perceived overall, if there are any upticks in general negative sentiment against the airline that they may need to address, and could help them respond more quickly to such situations. In this scenario, there will be an emphasis on balanced accuracy because, since we are trying to get an overall accurate representation of sentiment, all classes have equal importance. The managers then can analyze the tweets and be able to make an assessment on what needs to be improved for the company so they improve customer relations and not lose money.

The default threshold is 0.5, but changing the threshold can make a better balanced accuracy score. Having a threshold of 0.70 allows the model to have higher balanced accuracy. Balanced accuracy is a good metric to use for this scenario because the data is imbalanced, and we care about the positive and negative classes.

Threshold	Recall	Precision	Accuracy	F1-Score	Balanced Accuracy
Default (0.5)	0.99	0.91	0.91	0.95	0.80
Optimal Balanced Accuracy (0.70)	0.95	0.95	0.92	0.95	0.88

Table 3 Default Threshold vs Optimal Balanced Accuracy Threshold

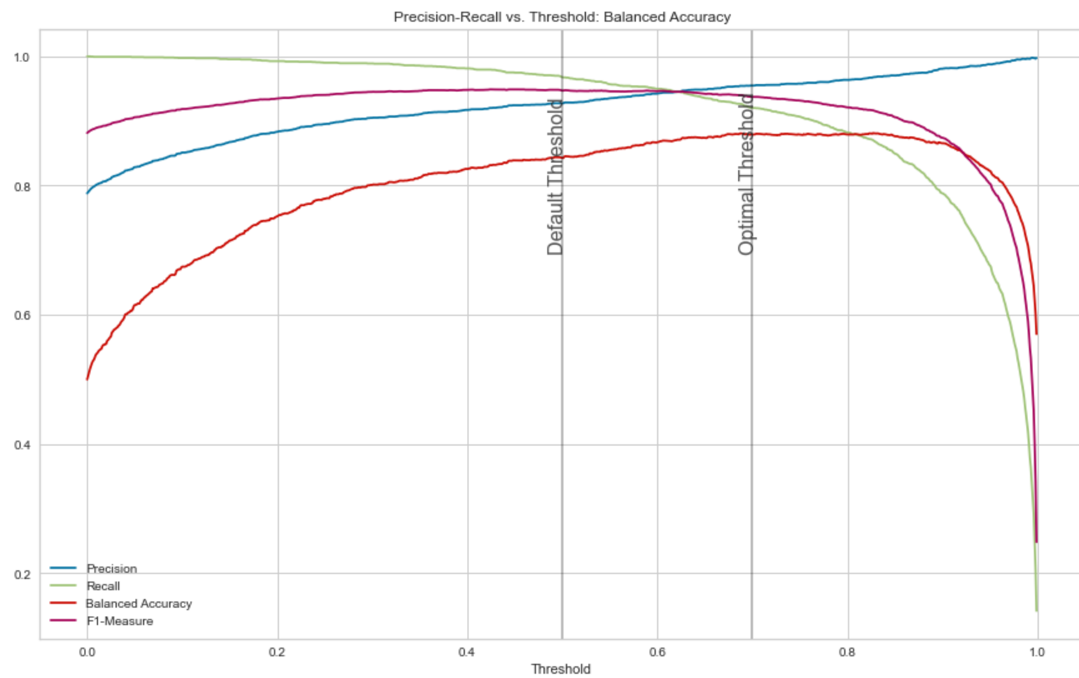

Figure 10.1 Precision-Recall vs. Threshold: Balanced Accuracy

Business Case 2: Customer Service

Airlines can also use this model to identify customers that are having a negative experience and direct tweets with negative sentiment towards the proper channels. The customer then can be helped in a timely manner which may improve the overall customer experience with the airline and the perception of the airline.

One way this might be accomplished would be to set up a bot to alert customers with a message if it is a positive sentiment like "Thank you for flying with us. We look forward to seeing you on your next trip." If it is a negative sentiment, a bot can send a similar tweet saying "Sorry for the unsatisfactory experience. Can you please dm us." The bot can then send an alert to a customer service representative, and a discount or refund can be offered to the customer.

In this scenario, a false negative would mean that the model predicted a tweet to be positive, but it was actually negative. The down side to this is that the customer will not receive the proper customer service help and may never fly with a particular airline again. This will ultimately cause the airline company to lose money. A false positive in this scenario would mean that the model predicted a negative sentiment tweet, but it was actually a positive sentiment tweet. The down side to this is that a bot would respond to a customer that had a positive experience and offer a discount or refund and cause a customer service representative to reach out when there isn't a reason to. This will cause the company to lose money and waste time.

In this scenario, since both low recall and low precision have significant downsides, we could use the F1-score. We want to find as much as possible of the negative sentiment tweets. We also don't need bots and customer service representatives wasting time and money on responding to customers and offering discounts if the tweet is a positive sentiment. This metric might change based on an individual airline's particular situation in which either finding more of the customers with negative sentiment or saving more money would take on more importance.

The default threshold and the optimized F1-score have similar metrics. We would need a cost analysis of the false positives and compare them to the false negatives to determine the next best step for the airline.

Threshold	Recall	Precision	Accuracy	F1-Score	Balanced Accuracy
Default (0.5)	0.97	0.93	0.91	0.95	0.84
F1-Measure (0.664)	0.96	0.95	0.92	0.95	0.87

Table 4 Default Threshold vs Optimal F1-Score Threshold

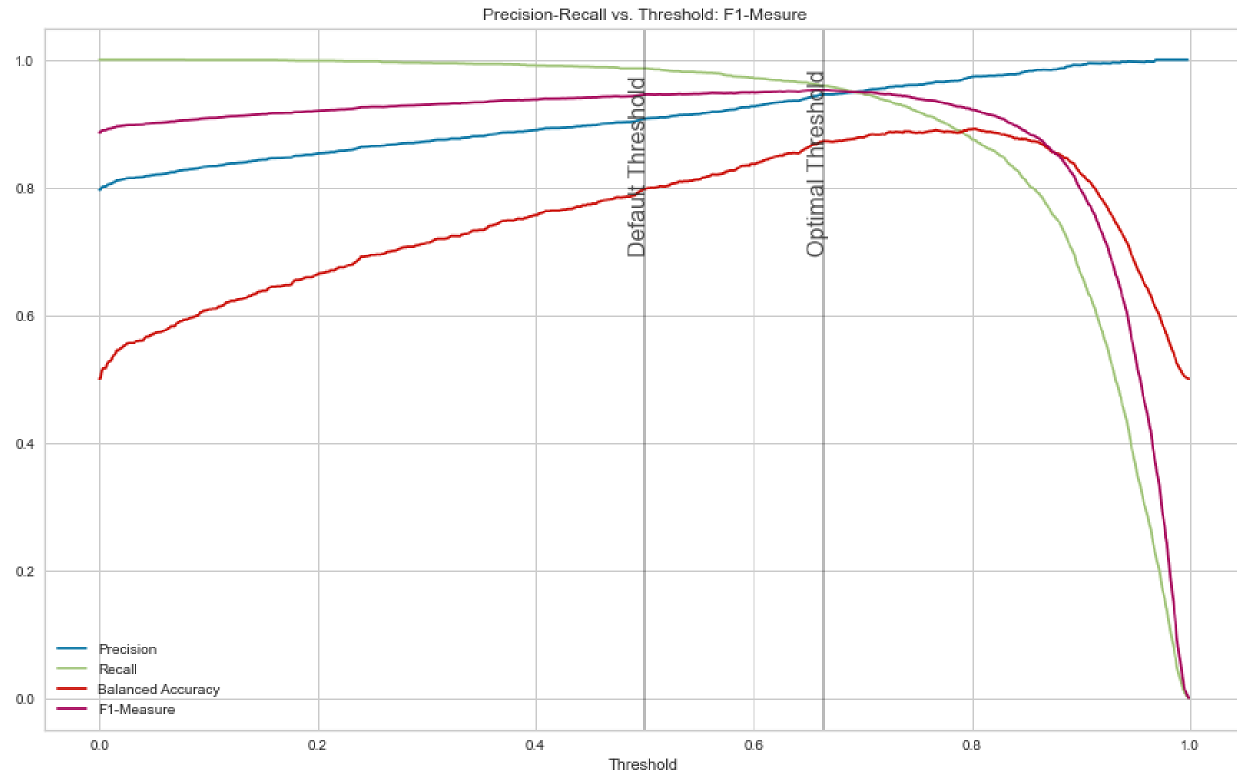

Figure 10.2 Precision-Recall vs. Threshold: F1-Measure

CONCLUSION

Tracking customer sentiment can help businesses improve products and services. In this project, I was able to build models to deal with two separate business scenarios in which predicting sentiment could be useful to airline companies. For tracking customer sentiment, I was able to get an accuracy of 92% and a balanced accuracy score of 88% by optimizing the balanced accuracy score to determine the best threshold for the model. For identifying customers with negative experiences, I was able to get an F1-score of 95%. The next steps would include looking deeper at the causes of misclassification to improve the model. One thing I noticed is that a tweet can have positive language, but the user can be using sarcasm which can throw the model off. With more time, I would look deeper into this. An issue with the dataset is that all the tweets collected are from Feb 16, 2015 and Feb 24, 2015. It's likely that collecting more tweets would yield a better and more generalizable model. I would be interested in what a whole year of customer sentiment would look like. In particular this might help with anomalous periods such as holidays and summers where there may be more customers and higher negative sentiment.