Spring 2021

Analyzing customer reactions towards popular Airlines with Aspect-based Sentiment Analysis

Kavita Jain

Creative Component

Iowa State University Capstone

Analyzing customer reactions towards popular Airlines with Aspect-based Sentiment Analysis

Kavita Jain

Spring 2021

MASTER OF SCIENCE IN INFORMATION SYSTEMS

Advisory Committee: Dr. Anthony Townsend

Ivy College of Business

Gerdin Business Bldg, 1200, 2167 Union Dr

Ames, IA 50011

Phone: (515) 294-8300

ACKNOWLEDGEMENT

It gives me an immense sense of accomplishment on finishing my graduation, which I must acknowledge was impossible without the encouragement of few people in my life. I would like to thank my advisor Dr. Townsend Anthony; he has been proactive in guiding us with the process and sharing his knowledge. This journey would not have been achievable without all the knowledge, experience, and guidance gathered from the Iowa State faculties we came across throughout the course.

First and foremost, I would like to thank my husband, Anant Kumar Jain, who has always been my inspiration. He motivated me to take up this opportunity to start studying again after a gap in my career. My pursuit of knowledge would not have been possible without the support he provided financially and emotionally. My family is the reason for what I am as a person. My father, Kamlesh Jain, played a significant role in providing all the support needed to move ahead in my career. My mother, Mamta Jain, my role model, has always been there and takes pride in whatever path I choose. I am eternally grateful to my parents for their perseverance and their efforts in shaping our lives and career.

ABSTRACT

With the increasing Social Media presence, people share thoughts and reviews on the Internet of their experiences with food, places, brands, travel, or any services, which has made the Internet a place full of relevant and irrelevant information. Therefore, it is impossible for businesses to manually keep track of every customer feedback present on the Internet. Sentiment Analysis is one such panacea to this problem. When it comes to sentiment analysis, the preferable platform is Twitter, allowing the extraction of public opinions and performing any possible analysis. My research is about understanding the people's sentiments for popular airlines Delta, American, Southwest, and United Airlines with similar intentions. The study aims to provide decision support for the customer to select the best fit Airlines and for the business to understand the customer feelings, analyzing their feedback, and maneuver to improve customer experience. The comprehensive study of Airline reviews with Aspect based sentiment analysis techniques extracts the top five aspects and classifies the views positive or negative for each Aspect. The relative evaluation of the four Airlines gives an overall perspective about each of them strengths and weaknesses and the unexpected discovery of business improvement opinions.

TABLE OF CONTENTS

INTRODUCTION

Traditionally, customer feedback or reviews were fetched by asking customers to fill the survey and questionnaires to get the perceptions about products or services. The process was not an accurate and efficient way to gather insights as customers were not always comfortable or serious about their reviews. In today's world, with the massive social media presence and awareness, customers do not hesitate to express their frustrations or honest opinions about the services or products they experience.

In a study by Li and Liu (2014), 81% of Internet users have searched related comments before buying a commodity at least once. The search rates for related words before using restaurants, hotels, and different services reports from 73% to 87%. Hence, the conclusion is that there is a significant impact on the customer's decisions of the online inspections. The social media platform demonstrates a place where people express their thoughts freely. The online platform wields an enormous influence in determining the opinion of other consumers. Consumer voices can influence brand awareness, brand loyalty, and brand encouragement. So, mining user opinions can help large brands and industries to strategize their next moves. The exponential and progressive increase of internet usage and public opinion exchange has given rise to the need for research in Sentiment Analysis.

This paper aims to unveil the following questions by performing research on Aspect-based Sentiment Analysis techniques.

1. What are the positive and negative sentiments for each Airline?
2. What are the frequent topics mentioned by consumers while talking about these Airlines?
3. What are their sentiments towards those topics?

SENTIMENT ANALYSIS

There are several definitions for Sentiment Analysis, also known as opinion mining, across the Web world. One of them is mentioned by Liu in his book as “Sentiment analysis is the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.” (Liu 2012).

Another significant definition proposed by Saleh et al. (2011) is “The automatic processing of documents to detect opinion expressed therein, as a unitary body of research.”

As shown in the below figure, there are four sentiment analysis levels, i.e., document level, sentence level, aspect level, and concept level. The usual sentiment analysis works on the document level, where the sentiment summarizes as positive or negative for the whole document. The results of document-level sentiment analysis are inaccurate as it the most abstract level of study. The sentence-level of opinion mining aims to classify sentiments for each sentence. Although the polarity of views for sentence-level is more accurate than document-level sentiment analysis, it might not be efficient when there is more than one Aspect in a sentence. And both the aspects don't need to have the same polarity. For example, in the sentence “I love the food of this restaurant, but I wish they had good staff to serve,” there are two aspects discussed and have two different views for each of the Aspect. The aspect level sentiment analysis aims to identify and extract the elements irrespective of the levels and then specify the sentiments. Therefore, this level provides more accurate results. The concept-level sentiment analysis requires a deep understanding and focuses on semantic analysis of text and analyzes the concepts which do not explicitly express any emotion. (Poria et al., 2014). This paper is focusing on the Aspect level of Sentiment Analysis for the Twitter data about Airlines.

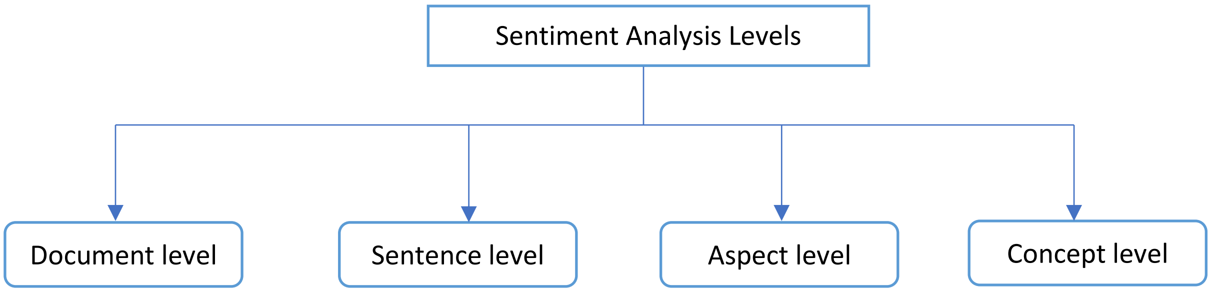

Figure 1. Different levels of Sentiment Analysis

Since Aspect extraction is one of the critical steps for Aspect based sentiment analysis, there are four related techniques discussed in the book (Liu 2012). They are:

- Extraction based on the frequent noun phrases and nouns (Jeyapriya and Selvi 2015, Hu and Liu 2004; Li et al. 2015)
- Extraction based on exploiting opinion and aspect relations (Qiu et al. 2011; Wu et al. 2009)
- Extraction based on supervised learning (Jin et al. 2009; Yu et al. 2011)
- Extraction based on topic modeling (Vuli’c et al. 2015; Mukherjee and Liu 2012)

This paper's research for Aspect-based sentiment analysis on Twitter data helps understand the Airline industry's customer sentiment. It performs using the Topic modeling technique mentioned above for aspect extraction.

LITERATURE REVIEW

With advancements in World Wide Web, an enormous data source is currently available and generated every minute. The Internet has given a platform for exchanging ideas, sharing opinions, and learning. Popular social networking platforms like Twitter, Facebook, Google+, and similar others have given people a place to share their thoughts and freely express their views on different brands, industries, or communities.

A few related works discussed in this paper discuss different industries and techniques for sentiment analysis. In the case study on Automotive Industry (Shukri, Yaghi, Aljarah, Alsawalqah 2015), sentiment analysis models are applied to the top three automotive industry companies – Audi, BMW, and Mercedes to understand the emotions of their customers. This information is helpful for companies to develop marketing strategies and address their customers' unseen interests. The sentiment analysis algorithm used to classify the polarity and emotions is the Naïve Bayes classification. The results of their analysis concluded that Audi's polarity is higher than other companies. This research can help the customers with their buying decisions by learning the three leading automotive companies' comparisons.

In another related case study for Apparel Brands (Abdur Rasool *et al.* 2019 *J. Phys.: Conf. Ser.* 1176 022015), sentimental analysis is on the top two international brands – Adidas and Nike. The Naïve Bayes algorithm for sentiment classification uses to extract the brands' polarity, customer opinion, and emotions. The results concluded that Adidas's positive polarity is 2.7% higher than Nike's.

The authors of the case study (Guoning, Bhargava, Fuhrmann, Ellinger, and Nemanja, 2017) performed via sentiment analysis algorithms on Twitter opinions about different industries, consumer brands, and relevant topics. The approaches of Deep learning-based classifies the sentiments as positive, negative, and neutral. Supervised learning applied to the text with Long Short Term Memory (LSTM). The authors of the paper found some exciting results during the analysis of 19M unique users with 62 industries and 12898 consumer brands. Based on the research, the industries with the most negative opinions are Airlines, Mail and Shipping, and Telecommunications, which tend to provide customer services. The industries with the most favorable views are the manufacturers and selling consumer goods like Household appliances. And the most expected polarized industries include Politics and Sports. One reason to choose this paper's research topic is to analyze the Airline Sentiment Analysis based on the article's statistics (Guoning, Bhargava, Fuhrmann, Ellinger and Nemanja, 2017). There is a curiosity to know about the Airline service industry's consumers' topics and polarize the sentiments for different aspects.

The previous works discussed here have either used Naïve Bayes classifier for sentiment analysis or supervised learning algorithms. This paper explores the Twitter comments by customers on the famous four Airlines using the unsupervised learning method. The aspect-based sentiment analysis will detail aspect categorization and polarizing sentiments as positive and negative for each Aspect. The information captured with this process will help the business to understand consumer expectations and strategize future goals.

METHODOLOGY

Introduction

In the Web world, sentiment analysis (also called opinion mining) plays a vital role in discovering the knowledge from the electronic media's expressed comments. With the advent of new programming technologies, the possibilities of exchanging opinions, sharing thoughts, and feedback have become prominent, which has increased the opportunities for techniques like sentiment analysis.

The main objective of opinion mining is to discover all sentiments in the documents (Saleh et al., 2011); in other words, it also uncovers the writer's attitude towards different aspects of the problem. The following figure depicts the process of Aspect-based sentiment analysis.

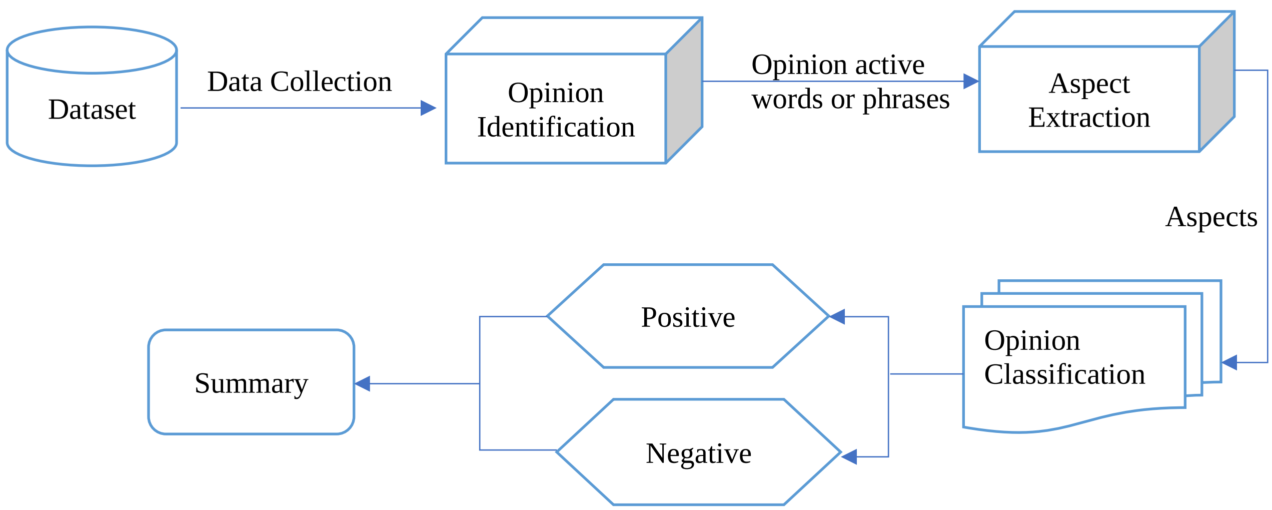

Figure 1 - Opinion mining process ref. (Hemmatian, F., Sohrabi, M.K., 2019)

In this research, the dataset used is already available with tweets from Twitter for the Airlines. The Opinion Identification step identifies the phrases representing individual emotions about the Airlines. Then in Aspect extraction, aspects are identified and extracted with the Topic Modelling technique. Following aspect extraction, respective positive or negative attitude classifies for each Aspect. The final summary can be in any form, either text or charts.

Data

The research is on Twitter tweets talking about the selected four Airlines. One of the ways is by extracting tweets with their respective handles, as described below in the Table. The limitation of this method is that Twitter has a limit on the tweets per day. Also, it counts retweets as Tweets. It was challenging to get enough tweets for each Airline unique and provide helpful information with these limitations.

Popular Airlines	Twitter Handles
American Airlines	@AmericanAir
Delta	@Delta
United	@united
Southwest Airlines	@SouthwestAir

Table 1 – Airlines with Twitter handles

Another option is to use the dataset available on CrowdFlower named Airline Twitter Sentiment (Data. World, 2021, March 26). The dataset provides enough information for each Airline to perform Aspect-based sentiment analysis. It contains information about sentiments, but the research is more than that. It is about extracting aspect categories and finding out the feelings for each of them, respectively.

There are 20 fields included in the dataset, as shown in the following table. The research will only use the “airline” and “text” field columns for sentiment analysis.

#	Data Field	Data Type
1	unit_id	Integer
2	golden	Boolean
3	unit_state	String
4	trusted_judgements	Integer
5	last_judgment_at	Datetime
6	airline_sentiment	String
7	airline_sentiment_confidence	Decimal
8	negative reason	String

9	negativereason_confidence	decimal
10	airline	string
11	airline_sentiment_gold	string
12	name	string
13	negativereason_gold	string
14	retweet_count	integer
15	text	string
16	tweet_coord	string
17	tweet_created	DateTime
18	tweet_id	decimal
19	tweet_location	string
20	user_timezone	string

Table 2- Crowd flower dataset information

The dataset provides relevant information for six Airlines, but since my focus is on popular Airlines, the research analyzes tweets for Delta, Southwest Airlines, United Airlines, and American Airlines. The below visualization is to shows the total records for each Airline.

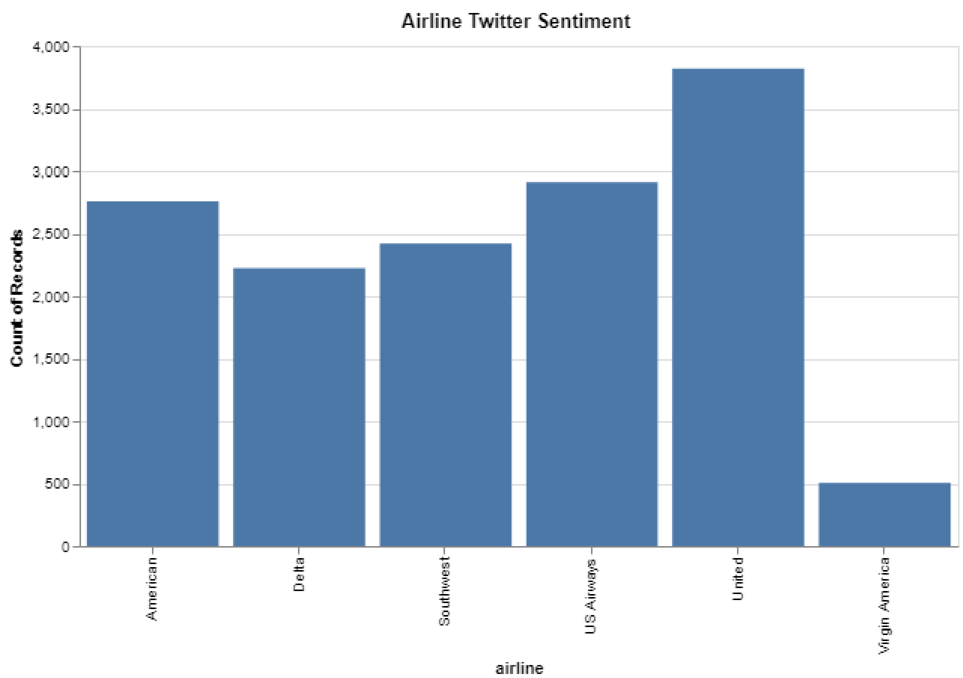

Figure 2 – Airlines vs No. of Tweets

Data Preprocessing

As a standard process for unstructured data, I have followed the traditional pre-processing techniques in Natural Language Processing (NLP). The steps are listed below. Apart from the ordinary automatic pre-processing done using libraries or functions, I manually changed few texts that do not significantly contribute to the tweet's meaning. For example, the tweet text says, “@united still no refund or word via DM. Please resolve this issue as your Cancelled ***Flightled*** flight was useless to my assistant’s trip.” OR “@united Cancelled ***Flighted*** our flight, didn’t rebook us on an added flight, now have to drive from a Denver to KC....thanks!” The words “flighted” or “flightled” is not making sense in the sentence. On inputting this type of text in the Topic Modelling Algorithm, it is categorizing as two different Aspect. Hence, eliminating these words from the ruling makes it easy for the algorithm to classify the aspects.

Expanding Contractions

Contractions are words or combinations of shortened words by dropping letters or replacing them with an apostrophe. In social media like Facebook, Twitter, Instagram, people prefer using short forms of words, primarily by dropping vowels or using apostrophes. For example, “I'll be there in a minute," "I'm not gng out tonight," "I’d like to meet me in the evening.” It is essential to remove contractions, especially when performing sentiment analysis on Twitter data.

Install the Contractions library using pip install (GeeksforGeeks. (2020, October 25)).

```
!pip install contractions
```

```
#Expanding Contractions
import contractions

df['pretweets'] = df['Tweets'].apply(lambda x: [contractions.fix(word) for word in x.split()])
df['tweets1'] = [' '.join(map(str, l)) for l in df['pretweets']]
df.head()
```

Removing hyperlinks and ‘&amp.’

The tweet's text contains hyperlinks and ‘&amp,’ which does not contribute to the sentence's meaning. So, it must be removed by the replace function has described in the below screenshots.

```
#Remove '&amp' and links

df['remove_amp'] = df['tweets1'].str.replace(r'&amp;', '&')
df['remove_links'] = df['remove_amp'].str.replace(r'(https|http)?:\/(\w|\.|\/|\?|\=|\&|\%)*\b','')
```

Converting the text to lowercase, removing punctuations, underscore, whitespaces

The replace function removes the extra whitespaces, underscore, and punctuations. The lower() process helps to convert to lowercase.

```
#Convert the string to lowercase, remove punctuation, remove underscore, remove extra whitespace in string and on both sides of string


df['remove_lower_punct'] = df['remove_links'].str.lower().str.replace("'", '')
.str.replace('[^\w\s]', ' ').str.replace(" \d+", " ").str.replace(' +', ' ')
.str.replace("_", " ").str.strip()

display(df.head(10))
```

Tokenization

In this step, the text converts into tokens before transforming into vectors. Here, the text tokenizes into words. For example, the text “united thanks” is converted to [united, thanks] on tokenizing.

```
import nltk
nltk.download('punkt')

# tokenise string

df['tokenise'] = df.apply(lambda row: nltk.word_tokenize(row[5]), axis=1)

display(df.head(10))
```

Stopwords Removal

Stop words are frequently occurring phrases that contain no sentiment and do not provide any deeper meaning to the phrase.

```python
import nltk
nltk.download('stopwords')

# initiate stopwords from nltk

stop_words = stopwords.words('english')

# add additional missing terms

stop_words.extend(stop_words_list)

# remove stopwords

df['remove_stopwords'] = df['tokenise'].apply(lambda x: [item for item in x if item not in stop_words])

display(df.head(10))
```

Lemmatization

Lemmatization is the algorithmic process of grouping together the word's different modulated forms as a single item. It refers to as morphological analysis of words, which helps return the base or dictionary form of a lemma term. (Rungta, K. (2021, March 25))

```python
import nltk
nltk.download('wordnet')
# initiate nltk lemmatiser

wordnet_lemmatizer = WordNetLemmatizer()

# lemmatise words

df['lemmatise'] = df['remove_stopwords'].apply(lambda x: [wordnet_lemmatizer.lemmatize(y) for y in x])

display(df.head(10))
```

Vectorization

This step uses to convert the data into a numerical form for the model to understand. The bag-of-words approach is used, which generates the word, frequency counts. The CounterVectorizer function from Sklearn converts the collection of text to the matrix of word counts, which will be feed into the topic model.

```python
# initialise the count vectorizer

vectorizer = CountVectorizer(analyzer = 'word', ngram_range = (2, 2))

# join the processed data to be vectorised

vectors = []

for index, row in df.iterrows():
    vectors.append(", ".join(row[10]))

vectorised = vectorizer.fit_transform(vectors)

print(vectorised)
```

Topic Modelling – Latent Dirichlet Allocation Technique

Topic Modelling is a form of text mining employing an unsupervised machine learning approach to identify patterns in a corpus or a large amount of unstructured text. The techniques of topic modeling use to extract and categorize aspects from a large corpus of information. It provides methods to coordinate, recognize, and summarize extensive collections of textual data. One of the most used Topics modeling techniques is Latent Dirichlet Allocation (LDA) which uses to discover topics from the Airline reviews on Twitter.

LDA is an iterative algorithm with two main steps:

- In the initial stage, each word appoints to an arbitrary topic
- Iteratively, it goes through each word and relocates the word to a topic taking into consideration:
    - The possibility of the word belonging to a topic
    - The likelihood of the document to be generated by a topic

On applying LDA modeling techniques for each Airline Twitter tweet, the model discovered few exciting aspects. The visuals for each analysis illustrate the top five aspects frequently discussed in the reviews. The sentiments – positive, negative, and neutral for each topic presents the following section.

EVALUATION & RESULTS

The Topic Modelling algorithm is applied to each Airline's reviews separately. The following sections summarize Five Dominant topics with sentiment-wise counts and their visualization as a bar chart. The first bar chart in each cell shows the counts for positive, negative, and neutral sentiment. In the later section, the word cloud for both Positive and Negative tweets displays the most frequent words, which indicates the size of the word.

Delta Airlines

The Topic modeling LDA technique applied on the Twitter tweets for Delta results with five Dominant aspects. They classify into positive, negative, and neutral sentiments. Based on the available information after applying the modeling techniques, people are talking more about the booking problems, customer service, Employee, timing of flights, and the flawless Delta fleet. Since the bar chart in Fig 4 shows that most of the tweets are positive, it is clear that each Dominant topic's results will have more positive reviews than negative. It is because the data for negative reviews are less than favorable. From Fig 3, out of the five topics, 'fleet fleek' has a most negative opinion than the other four topics. However, there is a minor disparity among the topic 'booking problem' and 'fleet fleek' in the negative thoughts.

sentiment topic_name	Negative	Neutral	Positive
booking problem	169	134	238
customer service	135	168	277
delta guy	124	125	243
delta time	120	117	229
fleet fleek	177	211	259

Figure 3 – Five Dominant Topics classified sentiment wise for Delta Airlines

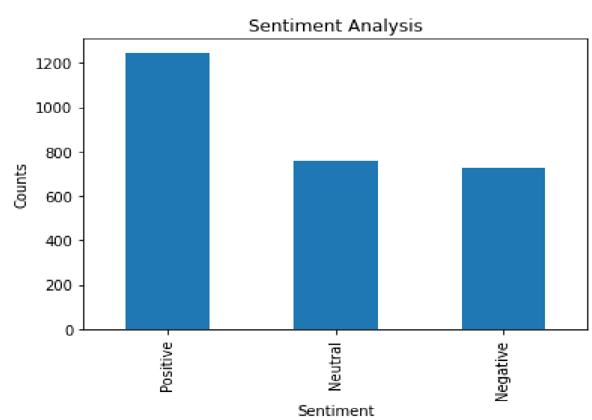

Figure 4 – Sentiment vs Count for Delta Airlines

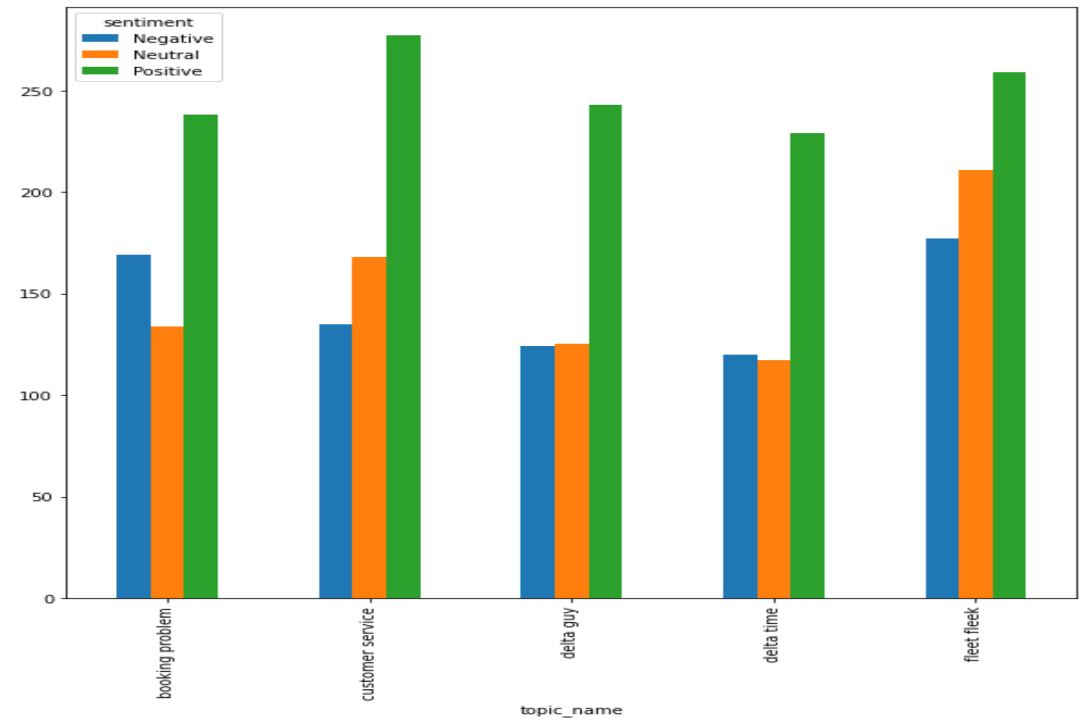

Figure 5 – Sentiment for each Dominant Topic

Word cloud of Positive Tweets

Figure 6 – Word Cloud for positive Tweets

Word cloud of Negative Tweets

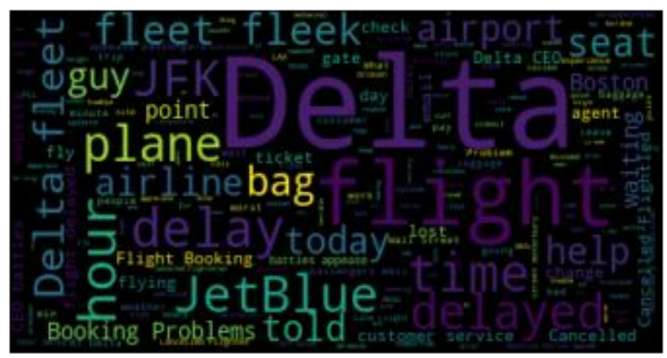

Figure 7 – Word Cloud for negative Tweets

Southwest Airlines

For Southwest Airlines, the scenario remains the same. There are more positive sentiments compared to negative. The dominant topics spoken for Southwest Airlines are 'canceled flight,' 'companion pas,' 'customer service,' 'flight canceled,' and 'follow dm.' Out of the five topics, the maximum positive and negative sentiment is for 'canceled flight.' There is a limitation for the algorithm to understand that 'flight canceled' and 'canceled flight' have the same meaning. It is considering as two different aspects.

sentiment topic_name	Negative	Neutral	Positive
cancelled flight	160	165	337
companion pas	128	97	227
customer service	115	105	209
flight cancelled	139	90	221
follow dm	126	78	210

Figure 8 – Five Dominant Topics classified sentiment wise for Southwest Airlines

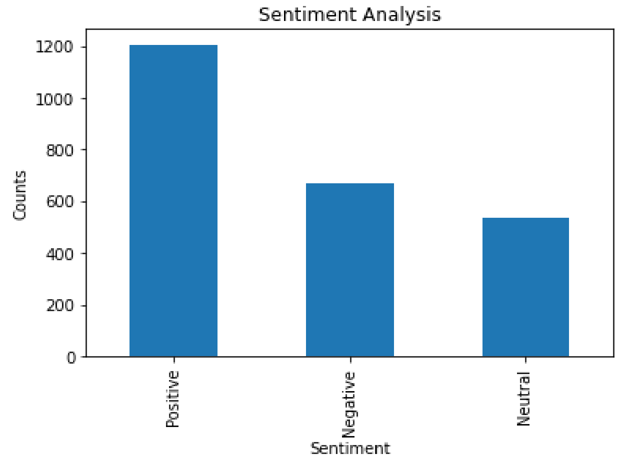

Figure 9 – Sentiment vs Count for Southwest Airlines

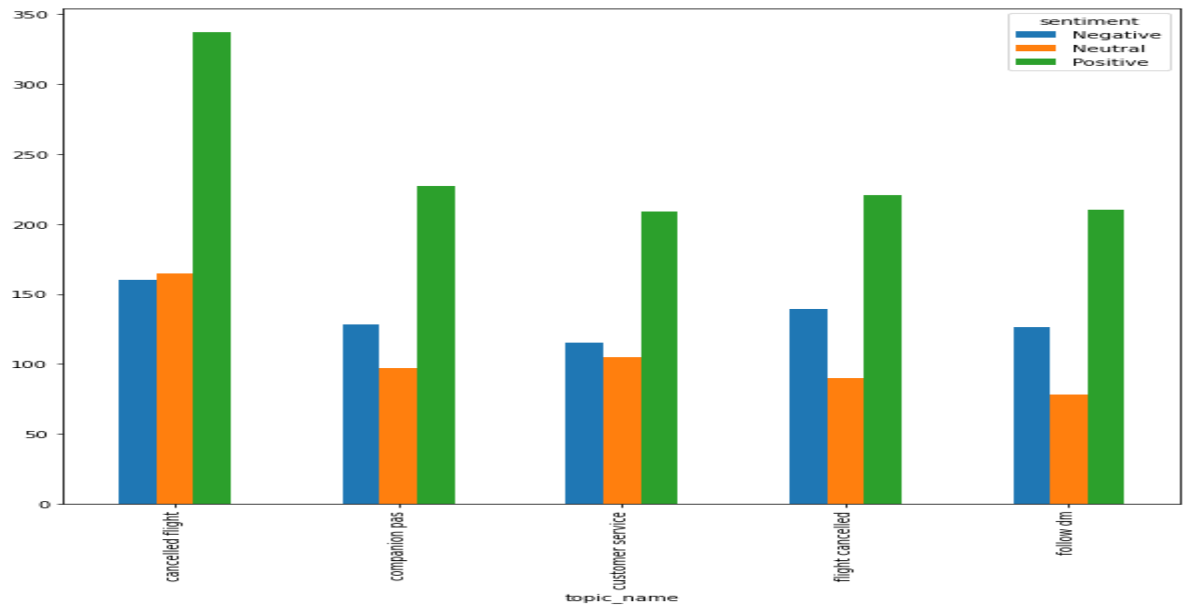

Figure 10 – Sentiment for each Dominant Topic

Positive Word Cloud

Figure 11

Negative Word Cloud

Figure 12

United Airlines

Unlike Southwest Airlines' case, there are five unique aspects: ‘wait time,’ ‘customer service,’ ‘canceled flight,’ ‘booking problem’ and ‘late flight.’ The negative sentiments for all the five topics are at marginal difference with ‘canceled flight’ with the highest count.

sentiment topic_name	Negative	Neutral	Positive
booking problem	140	19	553
cancelled flight	144	36	537
customer service	139	30	551
late flight	142	28	782
wait time	138	35	546

Figure 13 – Five Dominant Topics classified sentiment wise for United Airlines

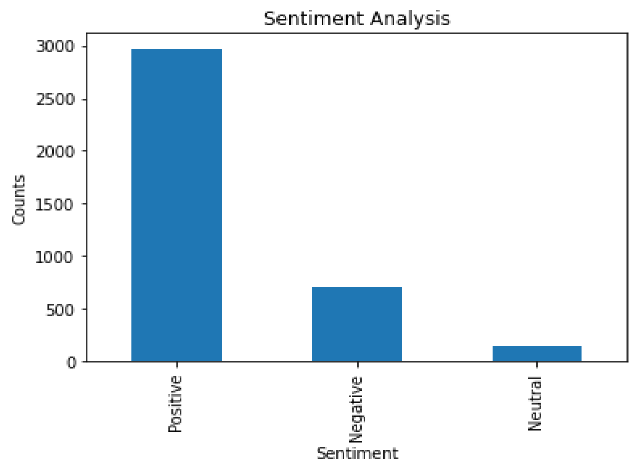

Figure 14 – Sentiment vs Count for United Airlines

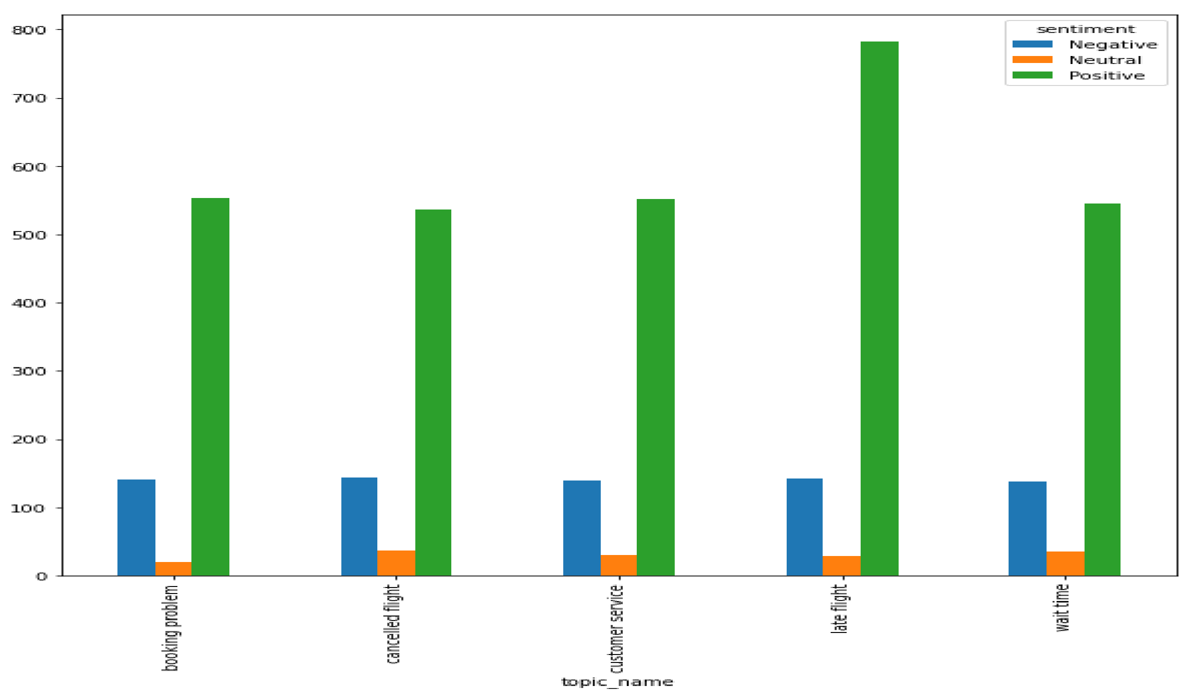

Figure 15 – Sentiment for each Dominant Topic

Positive Word Cloud | **Negative Word Cloud**

Figure 16

Figure 17

American Airlines

In American Airlines' case, the data for positive and negative sentiment is enough for a good comparison. The five dominant topics discovered are ‘canceled flight,’ ‘customer service,’ ‘flight canceled,’ ‘gate agent,’ and ‘hold hour.’ As seen in Southwest Airlines' case, the algorithm again considers ‘canceled flight’ and ‘flight canceled’ as two different Aspects when we know they mean the same. The highest negative sentiment is for the topic ‘flight canceled’ from Fig 18.

sentiment	Negative	Neutral	Positive
topic_name			
cancelled flight	227	96	177
customer service	182	110	203
flight cancelled	200	177	239
gate agent	214	75	205
hold hour	190	98	209

Figure 18 – Five Dominant Topics classified sentiment wise for American Airlines

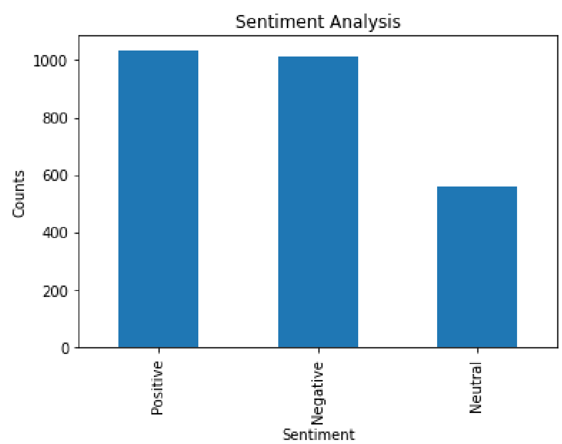

Figure 19 – Sentiment vs Count for American Airlines

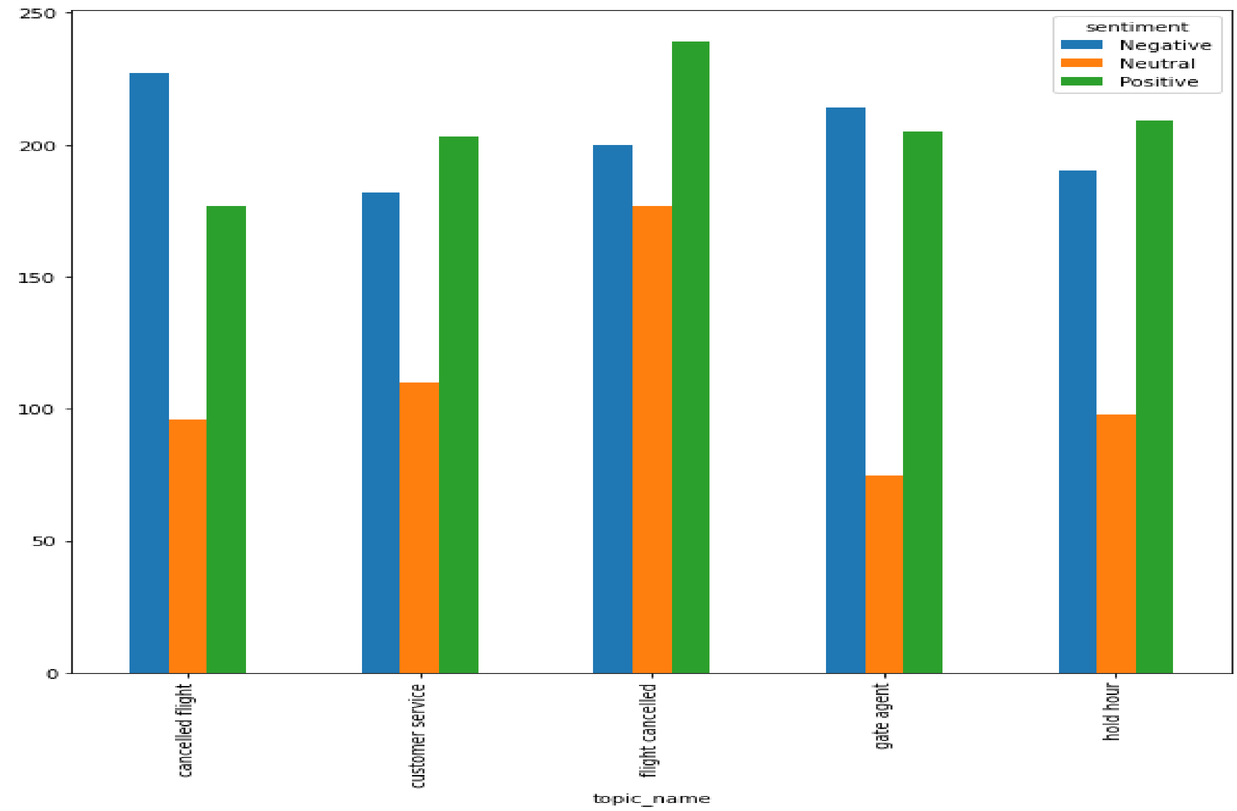

Figure 20 – Sentiment for each Dominant Topic

Positive Word Cloud

Figure 21

Negative Word Cloud

Figure 22

CONCLUSION

The paper's beginning mentions the research questions that answers after analyzing the Twitter tweets for four popular Airlines. Using the VADER sentiment analysis library of NLTK, the positive, negative, and neutral opinions automatically fetch for each Airline. The frequent topics discovery for each Airline using the LDA Topic Modelling technique and the respective sentiment for each topic talks about in the above section. Based on aspect extraction and the sentiment analysis of aspects for each Airline, it is clear that customer service is the most discussed topic and the highest negative opinion it is for American Airlines. The analysis concludes that people are more dissatisfied with American Airlines' customer service compared to others. The second most discussed topic is the 'flight canceled' or 'canceled flight,' and the maximum negative views is for American Airlines.

To conclude, American Airlines is not suitable for customer service and has many issues with flight cancellation. Airlines can be an excellent way to understand their services and improvise on their strategies to have better customer satisfaction.

LIMITATIONS

There were more positive sentiments versus negative for each Airline from the visualizations in Fig 4,9,14 and 19. It means for each Aspect, and there is more favorable data. In this case, the comparison is among the five aspects, which have the highest negative reviews. The VADER (Valence Aware Dictionary for sEntiment Reasoning) method is the most common rule-based sentiment analyzer and optimized for social media data. The main downside with the rule-based approach is that the method only cares about individual words and completely ignores the context in which it uses. It sometimes misinterpreted sarcasm and irony. In case of grammatical mistakes, it overlooks essential words. There is a chance of misunderstanding or missing the discriminating jargons, memes, or expressions. From Fig 8 and 18, the LDA Topic Modelling technique cannot understand that ‘flight canceled’ and ‘canceled flight’ are synonymous. It considers both the topics as two different aspects.

REFERENCES

Airline Twitter Sentiment - dataset by crowdflower. (2021, March 26). Data.World. https://data.world/crowdflower/airline-twitter-sentiment

Analyzing users' sentiment towards popular consumer industries and brands on Twitter Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic Lithium Technologies | Klout, San Francisco, CA https://www.researchgate.net/publication/319977416_Analyzing_users'_sentiment_towards_popular_consumer_industries_and_brands_on_Twitter

D. Arora, K. F. Li, and S. W. Neville, "Consumers' Sentiment Analysis of Popular Phone Brands and Operating System Preference Using Twitter Data: A Feasibility Study," 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, Gwangju, Korea (South), 2015, pp. 680-686, DOI: 10.1109/AINA.2015.253. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7098038

GeeksforGeeks. (2020, October 25). NLP - Expand contractions in Text Processing. https://www.geeksforgeeks.org/nlp-expand-contractions-in-text-processing/

Hemmatian, F., Sohrabi, M.K. A survey on classification techniques for opinion mining and sentiment analysis. Artif Intell Rev 52, 1495–1545 (2019). https://doi.org/10.1007/s10462-017-9599-6

I. (2020, March 9). Aspect-based Sentiment Analysis — Everything You Wanted to Know! Medium. https://intellica-ai.medium.com/aspect-based-sentiment-analysis-everything-you-wanted-to-know-1be41572e238

Jeyapriya A, Selvi K (2015) Extracting aspects and mining opinions in product reviews using a supervised learning algorithm. In: 2015 2nd international conference on electronics and communication systems (ICECS). IEEE, pp 548–552

JinW, Ho HH, Srihari RK (2009) OpinionMiner: a novel machine learning system for web opinion mining and extraction. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1195–1204

Liu B (2012) Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. Morgan & Calypool Publishers, pp 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Li G, Liu F (2014) Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. Appl Intell 40(3):441–452

Poria S, Cambria E, Winterstein G, HuangGB(2014) Sentic patterns: dependency-based rules for concept-level sentiment analysis. Knowl Based Syst 69:45–63

Rungta, K. (2021, March 25). Stemming and Lemmatization in Python NLTK with Examples. https://www.guru99.com/stemming-lemmatization-python-nltk.html

Saleh MR,Martín-Valdivia MT,Montejo-Ráez A, Ureña-López LA (2011) Experiments with SVM to classify opinions in different domains. Expert Syst Appl 38(12):14799–14804

Selvaraj, N. (2020, September 12). A Beginner's Guide to Sentiment Analysis with Python. Medium. https://towardsdatascience.com/a-beginners-guide-to-sentiment-analysis-in-python-95e354ea84f6

S. E. Shukri, R. I. Yaghi, I. Aljarah, and H. Alsawalqah, "Twitter sentiment analysis: A case study in the automotive industry," 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 2015, pp. 1-5, DOI: 10.1109/AEECT.2015.7360594.

Twitter Sentiment Analysis: A Case Study for Apparel Brands Abdur Rasool1, Ran Tao1, Kamyab Marjan1, Tayyab Naveed https://iopscience.iop.org/article/10.1088/1742-6596/1176/2/022015

Understanding Customer Frustrations in the Airline Industry with Aspect-based Sentiment Analysis - AYLIEN News API. (2016). https://aylien.com/blog/understanding-customer-frustrations-in-the-airline-industry-with-aspect-based-sentiment-analysis

Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput Linguist 37(1):9–27

Vuli´c I, De Smet W, Tang J, Moens MF (2015) Probabilistic topic modeling in multilingual settings: an overview of its methodology and applications. Inf Process Manag 51(1):111–147

Williams, L. (2021, January 26). Sentiment Analysis: Aspect-Based Opinion Mining - Towards Data Science. Medium. https://towardsdatascience.com/%EF%B8%8F-sentiment-analysis-aspect-based-opinion-mining-72a75e8c8a6d