SENTIMENT CLASSIFICATION AND OPINION MINING ON AIRLINE REVIEWS

Developed by: Naineni Sravanthi, Bairi Aravind, Mohammad Abdul Khaliq, Bingi Chandra Vamshi

Smart Bridge-Remote Summer Internship Program

1.INTRODUCTION

As twitter gains great popularity as an online social networking service in recent years, tweets on twitter Become a valuable source of information for companies to get feedbacks from their customers. However, Extracting such information from seemingly random comments and reviews is not an easy task due to the Scale of data size and the difficulty in semantic analysis. For example, someone tweets: "@Virgin America Hey first time flyer next week – excited! But I'm having a hard time getting my flights added to my Elevate Account. Help?" Is she happy with the flight experience or not? If not, what makes her unhappy? Having Machines answer these types of questions and learn the sentiment of each tweet automatically will be Extremely helpful for service providers to understand the strength and weakness of their products and to Improve them in the future.

In this paper, we focus on reviews on twitter for major U.S. airlines and try to extract sentiment and opinions From these reviews. We explored four learning techniques: Naive Bayes, Support Vector Machines (SVM), Lexicon-based methods and Convolutional Neural Networks (CNN), to predict the sentiment of these Tweets – positive, neutral or negative (referred as sentiment task). In addition, for negative reviews, we also Applied the same techniques to classify the reasons of bad feedbacks to see whether it is due to late flight, Cancelled flight, flight booking problems, or customer service issue, etc.

Opinion mining on Twitter data for airline services is a valuable application which assists User in collecting messages expressing the positive opinion or the negative comment. The objective of The application is to analyze messages about airline services on Twitter data by sentiment analysis.

1.1 Overview

Opinion Mining (OM) or Sentiment Analysis (SA) can be defined as the task of detecting, extracting and classifying opinions on something. It is a type of the processing of the natural language (NLP) to track the public mood to a certain law, policy, or marketing, etc. It involves a way that development for the collection and examination of comments and opinions about legislation, laws, policies, etc., which are posted on the social media. The process of information extraction is very important because it is a very useful technique but also a challenging task. That mean, to extract sentiment from an object in the web-wide, need to automate opinion-mining systems to do it. The existing techniques for sentiment analysis include machine learning (supervised and unsupervised), and lexical-based approaches.

1.2 Purpose

The main aim of this project presents a survey of sentiment analysis (SA) and opinion mining (OM) approaches, various techniques used that related in this field.

2.LITERATURE SURVEY

Sentiment analysis has aroused the interest of many researchers in recent Years, since subjective texts are useful for many applications. In Particular, sentiment analysis on online reviews has become a hot research Field. Studies on sentiment analysis mainly focus on framework and Lexicon construction, feature extraction, and polarity determination. This Paper presents a survey on the latest development in sentiment analysis, And makes an in-depth introduction of its research and application in Business and Blog sphere. The methods used in current research are Especially emphasized and the existing problems of those studies are Discussed. Finally, some possible future directions of research are pointed Out.

2.1 Existing problem

Sentimental Analysis in airways system is methodically done with the help of feedback forms or Online questionnaires, in their respective websites. The procedure is quite simple on an overview but Demands much of a complex nature when one tries breaking it down. Collecting feedback forms from A mass public and then analysing each and every form is a difficult task, requiring manpower as well As cost. In case of online pooling maintaining site regulations and keeping a database while performing Computations on the database is also a complex way of approach. As for existing algorithms for Sentimental analysis one such being Maximum Entropy (MaxEnt) Classifier coverts labelled feature Sets to vector using encoding.

$$P(fs|label) = \frac{dotprod(weights, encode(fs, label))}{sum(dotprod(weights, encode(fs, l)) for l in labels)}$$

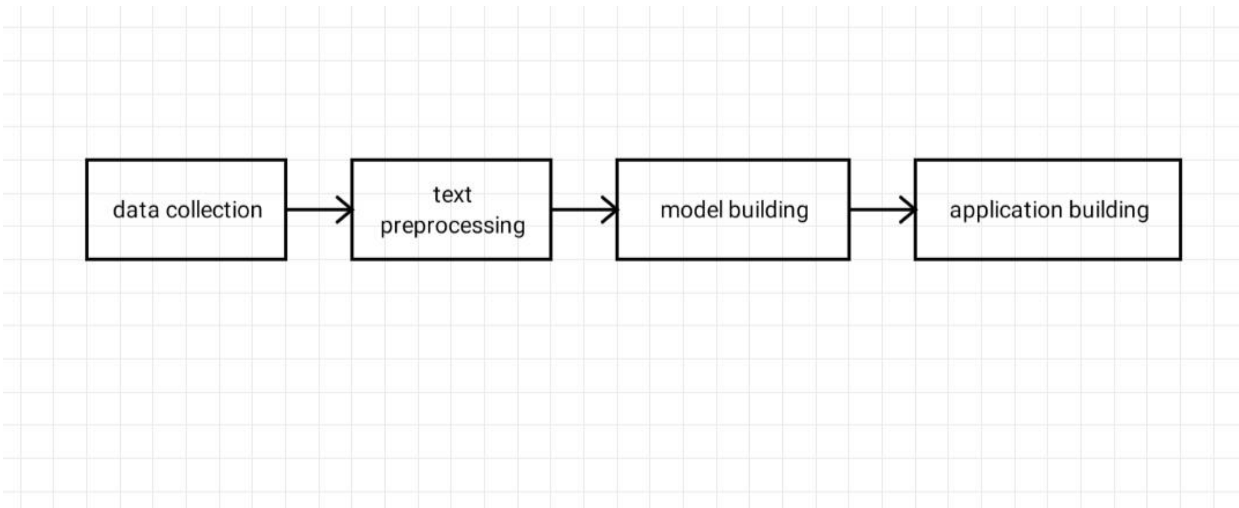
The problem faced with this type of approach is that it works best with dependant features, Meaning one event is related with another. But, doing sentimental analysis two events must be Uniquely identified so as, we can differentiate between the mass tweets.

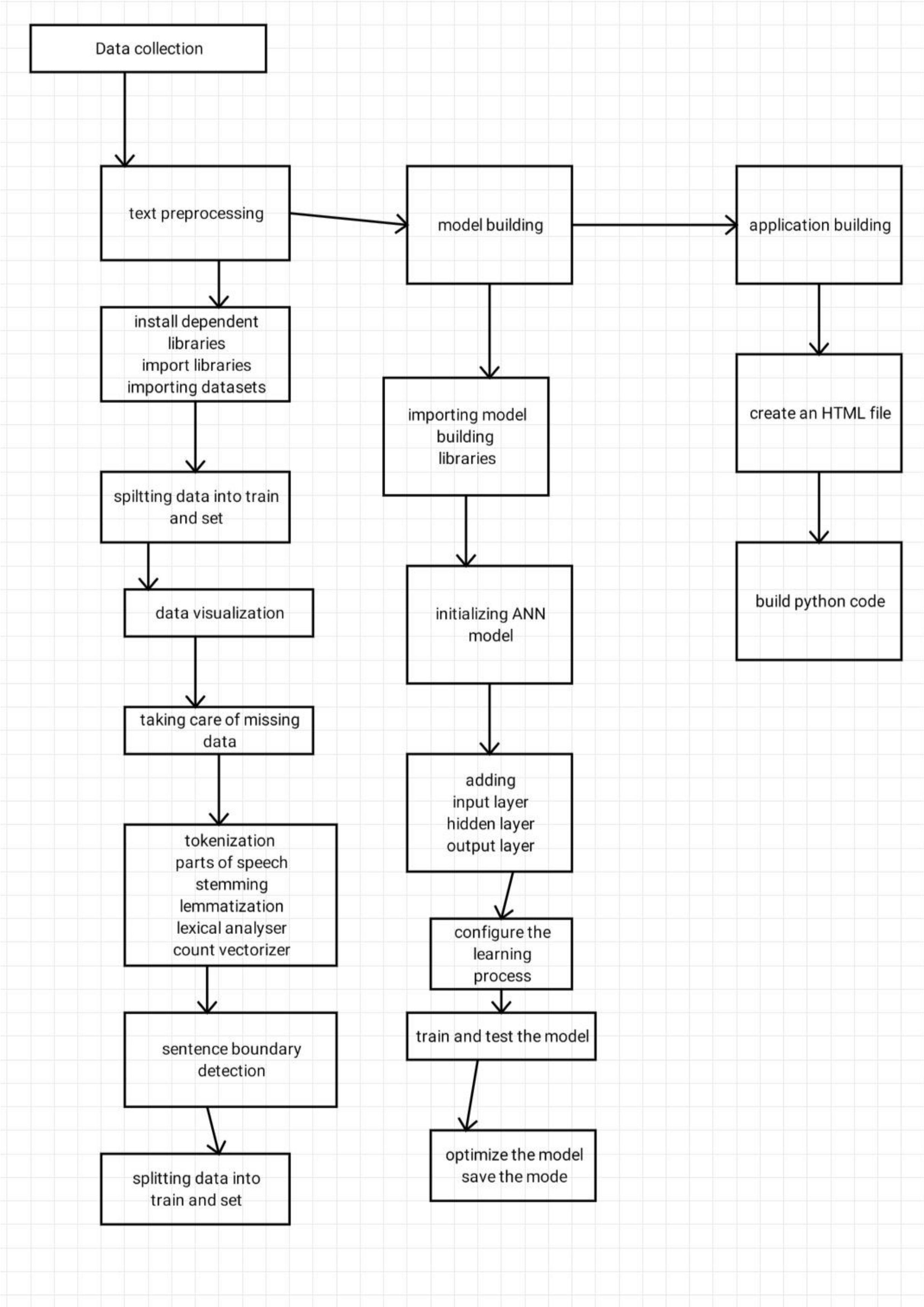
2.2 Proposed Solution

These days, sentiment analysis is gaining importance in the research study of text mining and natural language processing (NLP). There has been a rise in accessibility of online applications and a surge in social platforms for opinion sharing, online review websites, and personal blogs, which have captured the attention of stakeholders such as customers, organizations, and governments to analyze and explore these opinions. Therefore, the major role of sentiment classification is to analyze an online document such as a blog, comment, review and new items as a comprehensive sentiment and categories it as positive, negative, or neutral.

3.THEORETICAL ANALYSIS

3.1BLOCK DIAGRAM




Data collection
text preprocessing
model building
application building
install dependent libraries
import libraries
importing datasets
importing model building libraries
create an HTML file
spiltting data into train and set
data visualization
initializing ANN model
build python code
taking care of missing data
adding
input layer
hidden layer
output layer
tokenization
parts of speech
stemming
lemmatization
lexical analyser
count vectorizer
configure the learning process
train and test the model
sentence boundary detection
optimize the model
save the mode
splitting data into train and set

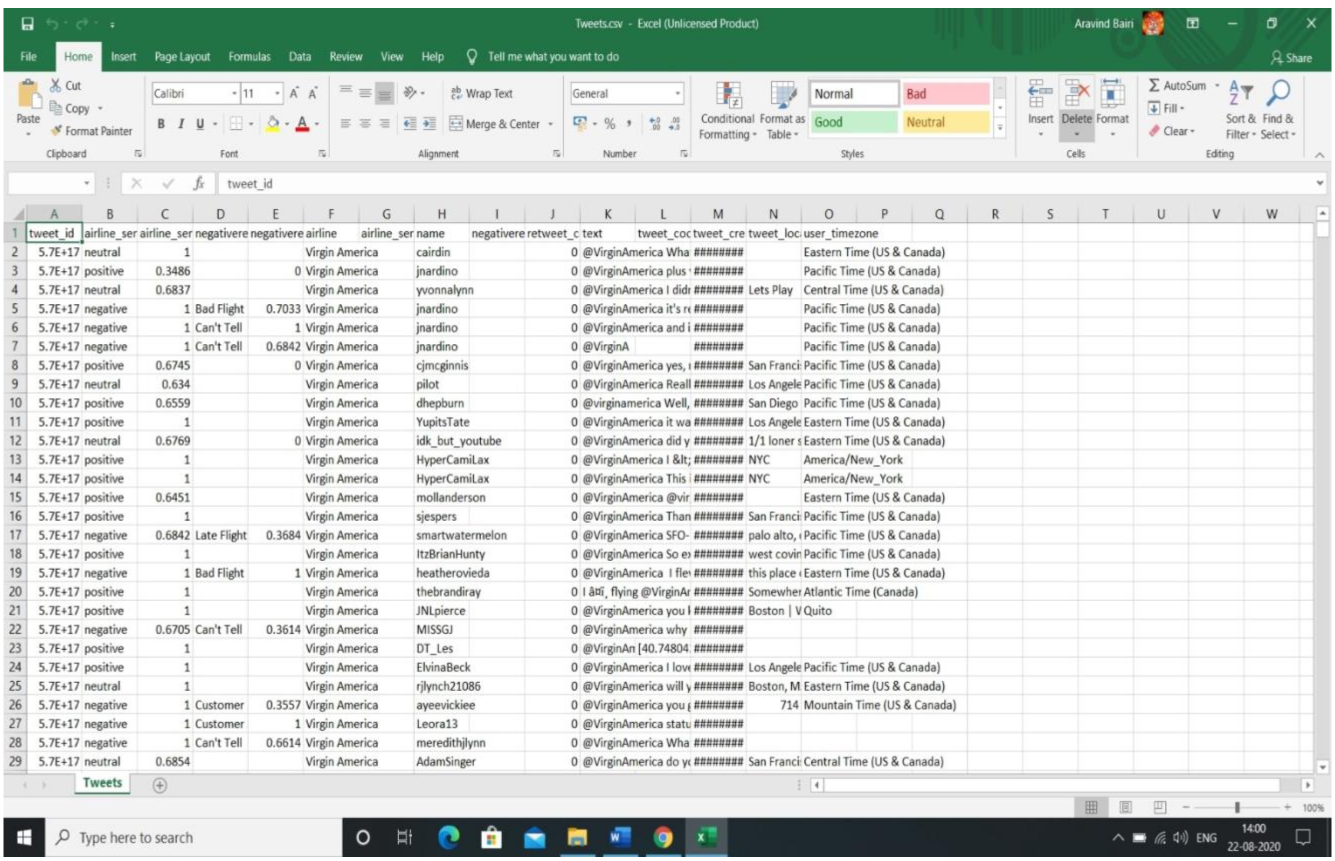
3.2 Software Designing

- Jupiter Notebook Environment
- Spyder Ide
- Natural Language processing
- Python (pandas, numpy, matplotlib, seaborn, sklearn)
- HTML
- Flask

We developed this Sentiment classification and opinion mining on airline reviews by using the Python language which is a Interpreted and high level programming language and using the Natural Language processing. For coding we used the Jupyter Notebook environment of the Anaconda distributions and theSpyder, it is an integrated scientific programming in the python language. For creating an user interface we used the Flask. It is a micro web Framework written in Python. It is classified as a micro framework because it does not require Particular tools or libraries. It has no database abstraction layer, form validation, or any other Components where pre-existing third-party libraries provide common functions, and a scripting Language to create a webpage is HTML by creating the templates to use in the functions of the Flask and HTML

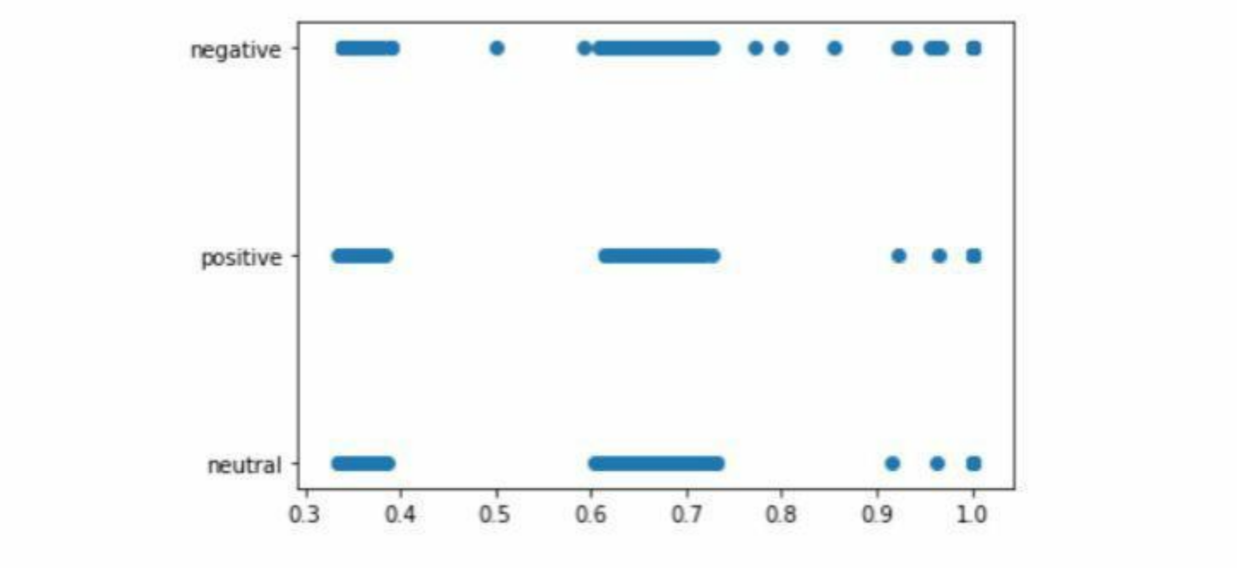
4.EXPERIMENTAL INVESTIGATION

In this paper, we have used the twitter dataset for airline reviews from kaggle.com they are shown below in the screenshot. Those reviews are taken in the form of positive, negative and neutral. It contains more than 5000 reviews of airline travellers with 15 attributes. These reviews are analysed and accurate results are generated by using natural language processing.



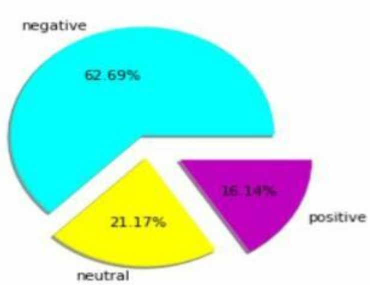
VISUALIZATION OF DATASET

Visualization between tweet id and airline_sentiment Here scatter plot is used to visualize the sentiment



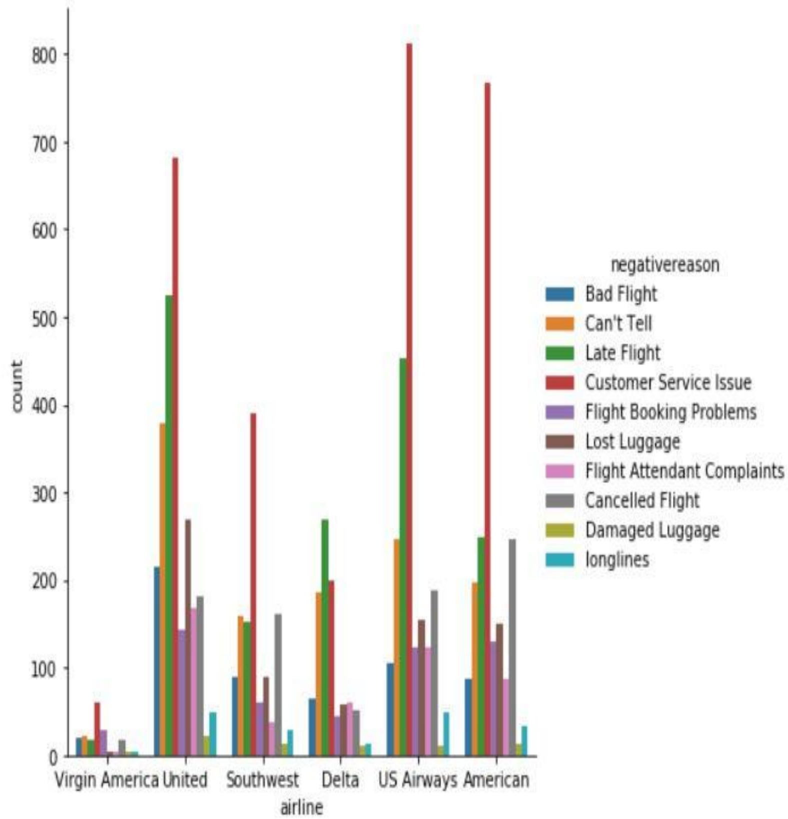
Bar graph

```
In [22]: #Using PieChart

         labels=['negative','neutral','positive']
         size=[9178,3099,2363]
         colors = ["cyan", "yellow", "m"]
         explode = (0.1,0.12,0.28)
         plt.pie(size,labels=labels,explode=explode,colors=colors,autopct='%1.2f%%',shadow=True)

Out[22]: ([<matplotlib.patches.Wedge at 0x184f6c8df48>,
           <matplotlib.patches.Wedge at 0x184f6c81188>,
           <matplotlib.patches.Wedge at 0x184f6f725c8>],
          [Text(-0.46587324595946716, 1.1058761769290402, 'negative'),
           Text(-0.13194990000624415, -1.2128434457457165, 'neutral'),
           Text(1.206352618660121, -0.6701592045565506, 'positive')],
          [Text(-0.2717593934763558, 0.6450944365419401, '62.69%'),
           Text(-0.0778720721348326, -0.7157764597843571, '21.17%'),
           Text(0.7692683365368889, -0.4273478985578003, '16.14%')])
```



This is a bar plot used to visualize the percentages of negative ,positive and neutral tweets

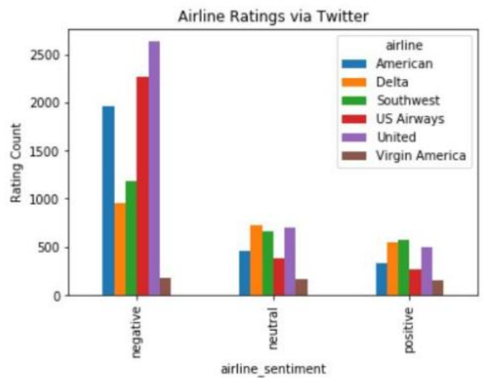
```
In [21]: sns.factorplot(x = 'airline',data = dataset,kind = 'count',hue='negativereason',size=6,aspect=.9)
         plt.show()
```



By this we are visualizing the negative tweets in airlines of different countries We used seaborn library and factorplot to visualize the graph

In [17]:

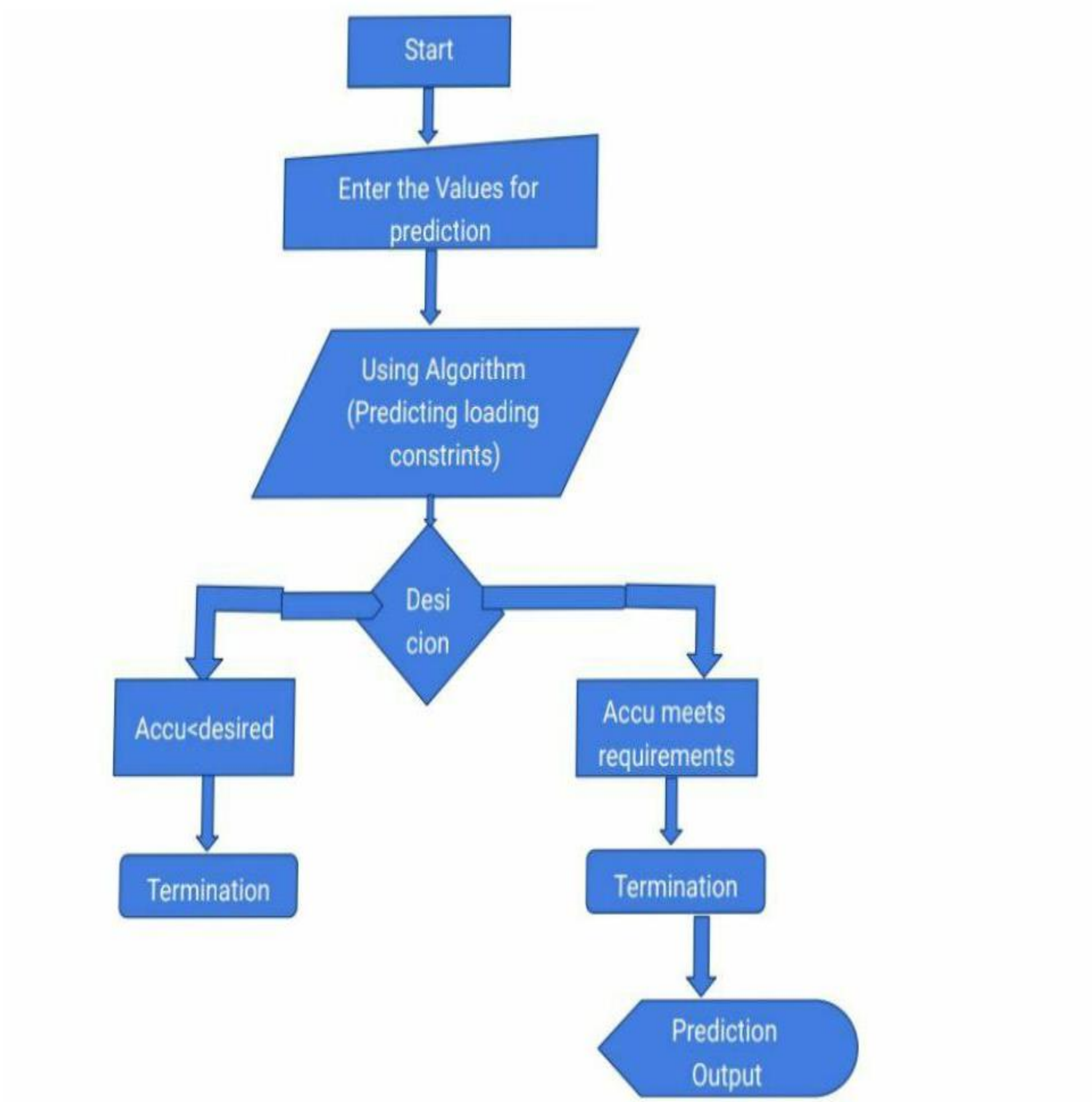
```
#Rating Based On Airline_sentiment

b = dataset.groupby(['airline','airline_sentiment']).count().iloc[:,0].unstack(0).plot(kind = 'bar',
                                                                    title = 'Airline Ratings via Twitter')
b.set_xlabel('airline_sentiment')
b.set_ylabel('Rating Count')
plt.show()
```



Visualization between airline_segment and rating count.

By this We can visualize the sentiment of tweets in different countries

5.FLOWCHART


Start
Enter the Values for prediction
Using Algorithm (Predicting loading constrints)
Desi cion
Accu<desired
Accu meets requirements
Termination
Termination
Prediction Output

6.RESULT

- Development of this project is very much useful for customers and airline companies. It gives very accurate and better results .
- Natural language processing plays a major role in analysing the customer reviews on their travel experience on certain airlines.
- We got 80% accuracy by analyzing 4000+ reviews from Twitter dataset.
- Sentiment analysis is one of the application of natural language processing.

7.ADVANTAGES AND DIS ADVANTAGES

Advantages

- Easy and simple User Interface.
- It is widely used for providing better customer experience in airlines industry.
- It is composed using the HTML and Python for the web usage in real time.
- It can work in real time
- It can give accurate and fast results
- It is the best way to detect customer experience.

Dis Advantages

- Complex Query Language- the system may not be able to provide the correct answer it the question that is poorly worded or ambiguous.
- The system is built for a single and specific task only; it is unable to adapt to new domains and problems because of limited functions

8.APPLICATIONS

- There are various applications of natural language processing in analysing reviews in various businesses like restaurants, hotels, airline industry and various other sectors.
- This NLP and sentiment analysis or opinion mining plays a major role in analysing and detecting the customer reviews based on their travel experience with some airlines.
- This detects good and bad reviews of the customers
- These reviews are mainly taken in the form of positive(+ve) Negative(-ve) ,neutral.
- Most of the companies use Sentiment analysis, to identify opinion’s and sentiment online to help them understand what customers think about their products and services

- Beyond determining simple polarity, sentiment analysis understands sentiment in context to help you better understand what’s behind an expressed opinion.

- Automatic summarization, question answering and text classification are also the major applications of NLP

9.CONCLUSION

At the present time, a large volume of information is available over Social Network like Twitter. Most Of contents may be opinions or comments, except those are the information or knowledge. In addition, Opinions or comments are valuable in the decision making. Therefore, the application analyzing data About interesting services such as airline business on Twitter will be an actual useful assistance to Airline service customers and providers. For the previous reason, this research proposed opinion mining using subjective words' Information from the modified subjectivity lexicon and applied Naïve Bayes classifier to leaning the Sentiment. The outcome of the proposed mechanism is the application that can analyze many contents On Twitter messages about airline services easier and faster.

10 . FUTURE SCOPE

This application can help Customers to make a decision about airline service selection of different airline brands. Furthermore, The airline service providers can gain information to improve the service qualities and set the Marketing plans. In future we will develop a good strategy to give complete accurate results.

11.BIBLIOGRAPHY

- B. Liu: Sentiment Analysis and Subjectivity in Handbook of Natural Language Processing 2nd Edition, edited by N. Indurkhya and F.J. Damerau, Chapman & Hall/CRC Press, Taylor & Francis Group (2010).
- B. Liu: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data 2nd Edition,Springer (2011).
- B. Liu and L. Zhang, “A Survey of Opinion Mining and Sentiment Analysis,” in Mining Text Data, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer US, 2012, pp. 415-463.
- B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Found. Trends Inf. Retr., Vol. 2, no. 1-2, pp. 1-135, 2008.
- S. O. Orimaye, S. M. Alhashmi, and E.-G. Siew, “Performance and trends in recent opinion Retrieval techniques,” The Knowledge Engineering Review, vol. 30, no. 1, pp. 76-105, 2015/001/001 2015.
- J. Liu, S.Seneff, and V.Zue, “Dialogue-oriented review summary generation for spoken Dialogue recommendation systems,” presented at the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, 2010.

- B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdhury, “Micro-blogging as online word of Mouth branding,” presented at the CHI ’09 Extended Abstracts on Human Factors in Computing Systems, Boston, MA, USA, 2009.
- P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised Classification of reviews,” presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, 2002.
- B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine Learning techniques,” presented at the Proceedings of the ACL-02 conference on Empirical Methods in natural language processing – Volume 10
- Y. Kim, “Convolutional Neural Networks for Sentence Classification,” CoRR, vol. Abs/1408.5882, 2014.

APPENDIX

HTML CODE

```
<!DOCTYPE html>
<html lang=”en”>
<head>
<meta charset=”UTF-8”>
<meta name=”viewport” content=”width=device-width, initial-scale=1.0”>
<title>airline review</title>
<style>
.water {
Position: absolute;
Right: 0px;
Top: 0px;
Height: 100%;
Width: 100%;
z-index: -1;
 background: linear-gradient(to right, #86fde8, #acb6e5);
</style>
```

```
</head>

<body>

<center><h1 style = “position:10%; color:blue;”>SENTIMENTAL CLASSIFICATION AND OPINION MINING ON AIRLINE REVIEWS<h1></center>

<br><br>

<img class=”water” src=”E:\MINIPROJECT\FLASKAPP\Templates\Static\gettyimages-862026202-612x612.jpg” alt=”” >

<p style=”font-size: 40px;text-align:center;”><!—margin-top:5%;→Please Provide Your Valuable Feedback</p><br>

  <form action = “/login” method = “POST”>

  <center>

   <textarea name=”review” rows=”5” cols=”100” placeholder=”Enter your text here!” style = “background-color:white;opacity:0.85;”></textarea>

  <br><br>

  <input type=”submit” style=”font-size: 20px;”>

  </center>

</form>

<center><b style=”font-size:2px;”>{{label}}</b></center>

<center><p>
```

```
 {% if label == “Positive Tweet” %}
 <p style =”color:green;font-size:50px;”>Positive Tweet</p>
{% else %}
{% if label == “Negative Tweet” %}
<p style = “color:red;font-size:50px;”>Negative Tweet</p>
{% else %}
{% if label == “Neutral Tweet” %}
 <p style = “color:black;font-size:50px;”>Neutral Tweet</p>
 {% end if %}
{% end if %}
{% end if %}
</p></center>
</body>
</html>
<!—linear-gradient(rgba(0,0,0,0.3), rgba(0,0,0,0.3)),→
```

FLASK CODE

```
From flask import render_template, Flask, request
From keras.models import load_model
```

```
Import pickle
Import re
Import tensorflow as tf
Global graph
Graph = tf.get_default_graph()
Import nltk
Nltk. Download(‘stopwords’)
From nltk. Corpus import stopwords
From nltk.stem.porter import PorterStemmer
Ps=PorterStemmer()
With open(‘tweet.pkl’,’rb’) as file:
Cv=pickle. Load(file)
Cla = load_model(‘airline_predictions.h5’)
Cla. Compile(optimizer=’adam’,loss=’binary_crossentropy’)
App = Flask(__name__)
@app.route(‘/’)
Def index():
Return render_template(‘index.html’)
```

```
@app.route('/login', methods = ['GET','POST'])
Def page2():
If request.method == 'POST':
Topic = request.form['review']
Topic=re.sub('[^a-zA-Z]',' ', topic)
Topic=topic.lower()
Topic=topic.split()
Topic=[ps.stem(word) for word in topic if not word in set(stopwords.words('english'))]
Topic = ' '.join(topic)
With graph.as_default():
Pred = cla.predict_classes(cv.transform([topic]))
If(pred==0):
Topic = "Negative Tweet"
Elif(pred==1):
Topic = "Neutral Tweet"
Else:
Topic = "Positive Tweet"
Return render_template('index.html',label = topic)
```

If name == ‘__main__’:

App.run(host = ‘localhost’, debug = True , threaded = False)