Kingdom of Saudi Arabia
Ministry of Education

King Faisal University
College of Computer Sciences & Information Technology

SENTIMENTAL ANALYSIS FOR AIRLINE PASSENGERS USING TWITTER API

by

Ahmed Ali Bahossain	(214110857)
Tammam ghalep altamimi	(214110858)
Ibrahim Alrashdi	(213112680)

Supervised by

Mr. Mutiullah Khan.

Committee Member Names

Dr. Nauman A. Qureshi

Mr. Conrado Padua Vizcarra

December 2018

ACKNOWLEDGEMENT

First, we thank Allah for everything and then it is our pleasure to be indebted to various people, who deserve our greatest gratitude, who directly or indirectly contributed in the development of this project and who influenced our thinking, behavior, and work during our senior design.

We are thankful to Mr. Mutiullah Khan for his support, cooperation, and motivation to us for accomplishing the project in a perfect manner who introduced us to the methodology of work, whose passion for the “underlying structures” which had lasting effects and who provided his valuable suggestions and precious time in accomplishing senior design project. Ultimately, we would also like to expand our deepest gratitude to our committee members, who have made valuable suggestions on this project which gave us an inspiration to improve our assignment.

UNDERTAKING

This work has been done by us (students below.) for the subject Project Proposal by Computer Science Department, College of Computer Sciences and Information Technology, King Faisal University. Development, Technical writing have been accomplished by the undersigned. Moreover, this project has not been submitted to any other college or university.

Student 1: Ahmed Ali Bahossain. Signature: __________________

Student 2: Ibrahim Alrashdi. Signature: __________________

Student 3: Tammam ghalep altamimi. Signature: __________________

ABSTRACT

Selecting the right airline is very important decision and due to increase in number of airlines difficulty of passengers have also increased. On the other hand, usage of social media has grown drastically. We can use the social data to analysis the sentiments of airline passengers traveling from/to Gulf Cooperation Council (GCC). Twitter is one of the most popular social media and the twitter posts of passengers can be used as the data set. Using twitter post our system can analyze the satisfaction of passengers towards the airline with which they have travelled. The system will be performing sentiment analysis (Positive, Negative and Neutral) on stream data of almost 7 days and generate a statistical report for 10 different airlines operating in GCC. For sentiment analysis we will be using Twitter API's with few data mining techniques and one of the algorithms such as Social Sentiment analysis, naive Bayes classifiers and Support Vector Machine (SVM). The final product will be a web application and it will give statistical data presented as graphs and visual components. This can help the user to decide which airline is good for them through results, real experiences and feedback of others.

Table of Contents

LIST OF TABLES

LIST OF Figure

Abbreviations and Acronyms

Table 1: Abbreviations and Acronyms

Term	Abbreviation
SVM	Support Vector Machines
API	Application Programming Interface
CTM	Correlated Topics Model
VEM	Variationally Expectation Maximization
URL	Uniform Resource Locator
UC	Use Case Diagram
DB	Database
SVC	Support Vector Classification
AQR	Airline Quality Rating
CTM	Correlated Topics Models
HTML	Hypertext Markup Language
JS	Java Script
CSS	Cascading Style Sheet
PHP	Hypertext Preprocessor
SMOTE	Synthetic Minority Over-Sampling Technique
ERD	Entity Relationship Diagram
XAMPP	Cross-platform Apache MariaDB PHP PERL
RF	Random Forest
STR	Sentiment Topic Recognition

1. Introduction

Sentiment analysis is needed for its usefulness in acquiring evaluation of information and any feedback for product you need through social media. Sentimental analysis will target the product of real users they have use and give you the feedback. This need can be met by creating a software which enables people to have feedback regarding the quality and service offered by the airline companies. More specifically, this software collects the feedback given from ordinary people in Twitter and certain analysis is made to provide the user with a real impression regarding the desired company. This software can be thought as a remarkable invention. The software solves effectively and efficiently the difficulty in deciding the best airline company. This is done by analyzing past tweets that mention the specified airline and categorize it. By using twitter API analyze collected data of ten airlines of last seven days. In addition, hashtags of airlines are helpful to collect posts and analysis their opinion and feedback.

The system should be able to request live tweets from twitter API. Web-based system course explains web API, and how it works. The system shall analysis twitter data using SVM algorithm. Before data analysis, preprocessing is required. The Data mining course state number of preprocessing techniques for text like natural language processing (NLP). The SVM is coded in python since python is power full in machine learning filed and has a huge set of built-in libraries to carry-out sentimental analysis process. The analysis process result is statistical report which is stored in DB. A Data-base concept and design describes how to create, manage, store and query DB. Then system shall show the result as figure, table and word cloud in web application using XAMPP as a local server. Development of web-based system and configuration of XAMPP server were discussed in web-based system course. The system will be submitted with a report. The report proper content, style, and format were taught in technical report course.

In the next section we discuss the project background, motivation, problem statement, scope and analysis of related work. The following section are devoted to project requirements, alternative solution and justification of propose solution. Followed by the system design and database design in section 3. Later the project implementation tools and techniques, metadata and algorithm with result are addressed in section 4. The last part of the report includes the conclusion and work plan.

1.1 Background

Nowadays, air travel has gained popularity in GCC which was not thought a few years ago. This popularity explains the dramatic surge in number of people traveling using planes. Due to this increase, many companies compete and provide different offers in this field to financially grow. However, as the number of airline companies increases dramatically, this indeed provides many alternatives and creates a difficulty in choosing the best company to travel with. In fact, this rises the need to have a way which helps people to choose the best company among others.

1.1.1 Airline Selection:

Below is the list of airlines operating in Saudi Arabia and they are selected based on some criteria mentioned.

S.No.	Airlines
1.	Singapore Airlines
2.	Japan Airlines
3.	Emirates Airlines
4.	Cathay Pacific
5.	EVA Air
6.	Etihad Airways
7.	Lufthansa Airline
8.	Oman Air
9.	Saudi Arabian Airlines
10.	Royal Air Maroc

1.1.2 Standards for Evaluation Airlines Companies

- **Safety**:

Airline companies are considered safe according to the report on how much accident, if any, happened in the last five years of operation. In addition, the ability of the airline to provide to its customers the needed medical aid during the flight. [1]

- **Number of flights:**

Increased number of flights give customers more options to choose from, which can be in their convenience, especially, in case of emergencies. This criterion can have a big impact on custom preference. In fact, Saudi Airlines has been able to reach 655 daily flights, with few airlines flying around the world. An example is shown if figure 1, where the data of 2016 number of flight is obtained. [2]

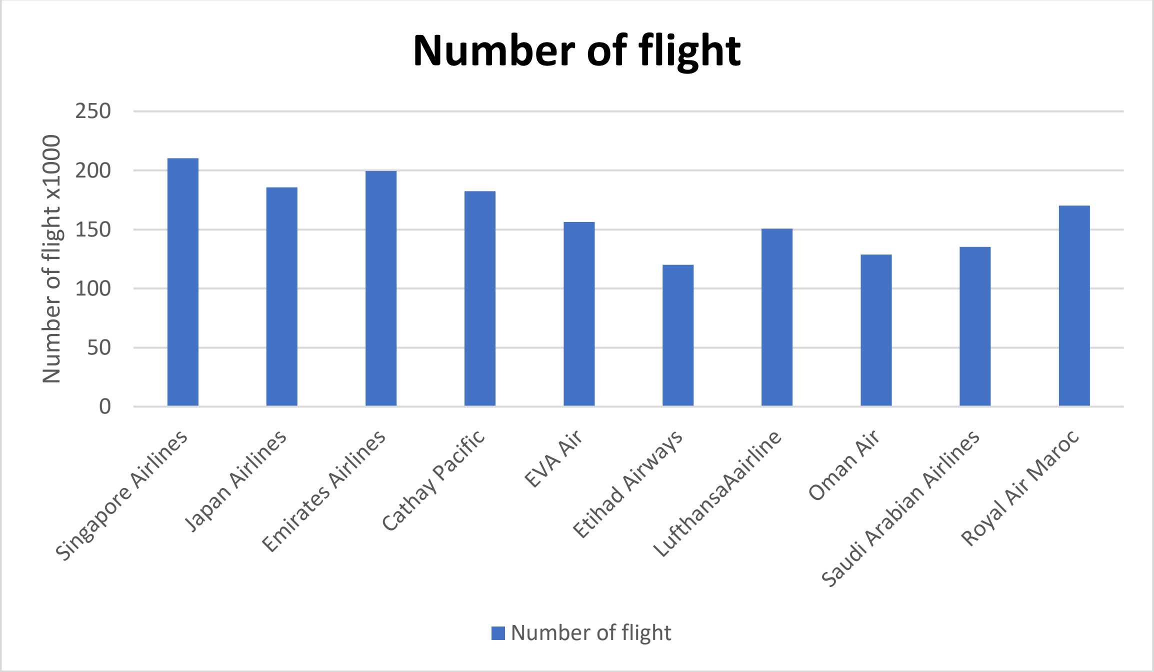


Figure 1:Air companies number of flight. [3][4][5][6][7]

- **Cost:**

Customer usually tends to appreciate a low-cost travelling. These criteria are dominant, after safety, in evaluation of airline companies. [8]

- **The less complaints from customer:**

The less complaints from customer about the company, the more good reputation it gains and vice versa. For example, the source of complaints is generated from luggage loss, flights delay and bad service inside the airplane. [9]

- **Entertainment and services provided:**

Comfortable seats and having space and leg room are important factors in customer comfort. In addition to other services and entertainment such as enabling cellphone calls and providing variety of videos for adult and children… etc. All these appeal to potential customers and leave a good impression on its customers. [10]

1.2 Motivation

The increasing number of airlines and the diversity of their services, especially those operating within GCC countries, have made it difficult to choose the right company for each person according to their needs. Moreover, since Twitter is one of the most interactive social networking programs, there is tremendous data on the responses of people and their opinions about the products and services of the airlines that they use. Therefore, our site will provide the possibility to study and analyze the reactions of the ten most used companies in GCC.

1.3 Problem statement

Knowing which airline is best for you can be a difficult problem that we face nowadays. It need a lot of data "feedback of customers' experience" and that will be sharing their experience. Here comes the problem of how you collect their feedback and how you are going tell which airline is the best for a user. It is difficult task for people to decide between many alternatives, especially with increasing numbers of alternatives.

This table show raw data received from twitter API. Each tweet has so many attributes, but table below show most useful of them.

Table 2: show row data received from twitter API

Tweet Text	Tweet publisher	Date and time	Hashtags	No. of Likes	No. of retweets
It's all about having fun and smiling! Flying crew	Cabincrew Tradeunio	Thu Feb 15 07:16:37 +0000 2018	#Gulfair #airlinescrew #gulfaircabincrew	**0**	**0**
Let's try this with @GulfAir as they are refusing to issue me a refund for the inconvenience their staff caused by giving false information of the 10 hour delay	Jason4PM2022	Mon Feb 12 04:41:04 +0000 2018	#GulfAir	77	43
From Georgia with Love !!	Cabincrew Tradeunion	Sun Feb 11 11:00:11 +0000 2018	#Gulfair #airlines#tiblisi #georgia #gulfaircabincrew #crewfie	4	0

Table 2 depicts twitter data after filtering process. Each row contains one tweet object. Tweet object has many attributes, but important attributes require by the system are tweet text, date, hashtags and number of likes as shown in table 3 below.

Table 3: present tweets attributes after fileting a related data

Tweet Text	Date	hashtag	Likes No.
#gulfair GF506 to Mumbai why no seating as per allocated seats. It's like a bus commute everyone is seating anywhere. A certain male staff is rude too	Thu Feb 08 19:56:47 +0000 2018	# gulfair	1
Air miles @GulfAir are useless, they will refuse ur upgrade request using ur miles & next day they will send u an email asking you to "BID" for an upgrade! Are u serious #GulfAir ? & by the way once u r onboard u'll see business class is almost empty.	Sun Feb 11 09:24:06 +0000 2018	#GulfAir	0
From Georgia with Love !!	Sun Feb 11 11:00:11 +0000 2018	#Gulfair	4
Let's try this with @GulfAir as they are refusing to issue me a refund for the inconvenience their staff caused by giving false information of the 10 hour delay @BahrainAirport on New Year's Day and not offering a hotel to half of the passengers of the same flight	Mon Feb 12 04:41:04 +0000 2018	# Gulfair	77
It's all about having fun and smiling! Flying crew	Thu Feb 15 07:16:37 +0000 2018	# Gulfair	0

1.4 Scope

This project will attempt to support the decision of customers while selecting the right airlines according to their requirements and satisfaction. The project will perform sentimental analysis on twitter dataset of 10 airlines operate in GCC gathered using twitter API. Gathered tweets will be of last 7 days. The project will implement data mining techniques using one of sentimental analysis algorithm. Its outcome can be expected as the statistical reports of all 10 airlines. Statistical results will help stakeholders to take right decisions to fly using suitable airline. The constraints in this project might be the availability of twitter data according to the airline hashtags and data must be only in English language. The proposed system will not be analyzing data other than the specified 10 airlines.

1.5 Comprehensive analysis of related work

The great spread of social media sites among most people facilitated people on sharing their opinions and experiences freely on a variety of topics, including their feedback about the products they tried, or services provided to them. Mostly those opinions are either positive, natural or negative. The most prominent of these sites is Twitter, where it has an advantage which is known as hash-tag. Those hash-tag include feedback of people on certain events or products. Therefore, one can create a website that analyzes these data in measuring the satisfaction of people around certain product or service, as well as to makes it easier for site visitors to increase their awareness and improve decision-making process. In addition, the twitter sentiment analysis offers organizations ability to monitor public feeling towards the products and events related to them in real time.

In airline service industry, it is difficult to collect data about customers' feedback by questionnaires, but Twitter provides a sound data source for them to do customer sentiment analysis. However, little research has been done in the domain of twitter sentiment classification about airline services. In our project, we will mention four specific studies. two concerning the algorithm and the last two on airline evaluation according to the Twitter data analysis.

a. Sentiment Analysis of Tweet by SVM:

Research in the last few years has begun to increase about emotional analysis and what are the best algorithms that are more efficient and effective than others. According to research by Suman Rani, Jaswinder Singh, tweets were analyzed which were compiled from the Twitter API related to people's opinions about the three Indian politicians. Researchers have shown that Two types of Support Vector Machine, linear and kernel SVM, are the most effective classification method that use as technique in the supervised learning approach. In addition, the researchers found that the supervised learning approach is higher in performance accuracy by up to 80%, while the unsuccessful learning approach reaches 78.6%. They also proved that linear SVM provides better performance than kernel SVM on all the measures, i.e., accuracy, precision, recall, and f-measure. According to the comparison, the researchers found that Linear SVM is better than SVM Kernel in all the measurements with an overall average of variables' performance of more than 90%. [11]

b. Sentimental Analysis of tweets 2013 FIFA Confederations Cup:

Researchers analyzing people's tweets and interpreting them in Portuguese, as can be applied to any other language, in the FIFA Confederations Cup in 2013 in Brazil. The researchers compared the two methods of classification (SVM and Naive-Bayes), which are better in terms of accuracy in performance using the most recognize international standards (i.e. accuracy, precision, recall, and f-measure). After testing and comparing two types of databases, SVM was found to be better than Naive-Bayes with an average range of more than 80% in performance for all previous measurements, while Naive-Bayes performance was 72.7%, which is about 8% less than SVM. [12]

c. Analysis of tweets for Airline Quality Rating:

Concerning evaluating airline companies, the first case of sentimental analysis was used to detect people’s opinions. For instance, Esi Adeborna and Keng Siau have used sentiment analysis in the case of airline quality rating. In this study, the approach was by using topic recognition model which was based on Correlated-Topics-Model (CTM) with algorithm of Variationally-Expectation-Maximization (VEM). In the end of their research had encouraging results with a very high accuracy which was around 85% on average [13]. Nonetheless, the number of data analyzed were relatively low and their method was incompatible with detecting figurative expression like jokes and irony. Therefore, a solution to this limitation is required.

d. Analysis of twitter data for airline services analysis:

Moreover, the second case study, done by Dr. Qigang Ga and Yun Wan, used an innovative approach in analyzing three American airline companies. This study merged between classification methods including Naive Bayes, SVM, Bayesian Network, Decision Tree and Random Forest algorithms that is called ensemble sentiment classification strategy. It was found that the use of methods of individual classification alone reduces the accuracy of the correct results and increases the rate of errors, yet the ensemble approach improved the overall accuracy in twitter sentiment classification by up to 91.7% and the error rate for the ensemble classifier is still the lowest which was 8.3%. [14].

Table 4 show comparison between different projects in sentimental analysis and the proposed system in this report. The system compares to others based on some characteristics like type of application, topics analyzed by each system, type of used data weather is it live or not, how many classifications it has, and does it use graphical elements or not.

Table 4: Comparison for different Sentimental analysis systems

Project Name	Application Type	Topics to Compare and analysis	Dataset	Classification	Output Visuals
Comparison of SVM versus naive-Bayes techniques for sentiment analysis in tweets	Desktop application	2013 FIFA Confederations Cup	Dataset	positive, negative, and neutral.	Yes
Sentiment analysis of tweets using support vector machine	Desktop application	3-Indian politicians	Live data	Positive and negative.	Yes
An approach to sentiment analysis – the case of airline quality rating	Desktop application	Airline Quality Rating of 3- airlines	Live data	positive, neutral or negative.	Yes
An ensemble sentiment classification system of twitter data for airline services analysis	Desktop application	Thirteen airlines in North America	Dataset	positive, neutral, negative and irrelevant.	Yes
Sentimental Analysis for airline passengers of using Twitter API	Web application	10 airlines operate in Saudi Arabia	Live data	positive, neutral or negative.	Yes

The table 4 shows four different projects similar to this system. It addresses key feature for each project. It compares this system with others, what is similar and what is difference. As you can see from table 4, most of them are desktop application where this project is web-based application. Everyone can access and use it. The data collected is a live data to have last update. Half of projects doesn’t use live data, because live data preprocessing is hard. Most of these systems have sentimental classification in three class. Common feature among them is visual output. All systems are presented their output using visual elements.

2. Detailed Project Requirements

2.1 Functional Requirement:

1. **Get data from twitter using twitter API**: Enable system to communicate to twitter by handling all required petameters for authentication process. The system will use twitter go get tweets about listed airlines. Live tweets will be a data set to perform sentimental process in.
2. **Preprocessing of data**: Prepare tweets for sentimental analysis process by remove dummy words, substitute emotions, quotations marks, URLs, usernames and perform Tokenization. These steps are essential to perform analysis process. The preprocessing creates an abstract model of each tweet. This tweet model will be used to get good classification result.
3. **Fit the algorithm with training data**: divide knowing label data set into two parts one training data take 60% from data and second test data take 40% from data. The training data will be used to train classification model to be able to predict sentimental value of new unknown tweets. Where, testing data will be used to measure accuracy to get idea about the model performance.
4. **Preform sentiment analysis using SVM:** Take input as bag of words and check sentimental value for each tweet from data which result in three reports good, bad and neural. In this step, the system will predict label for tweet which is the main task of the system.
5. **Store the result into database:** store result which are three reports into DB. Since the system is developed in two languages. DB is required to store result from python script to be further used by PHP to present in webpages
6. **Show the result using graphical components**: Show result of sentimental analysis in useful form using graphical representors. The visual output helps end user to get information easily.

2.2 Non-Functional Requirements:

- **Accessibility**: normal people without any disabilities shall be able to use the system. Also, people who know how to access and use websites.
- **Availability**: As it is web-application, it shall be available 24/7 all the time. A user can use a system at any time as needed to support the decision-making process.
- **Efficiency**: The system shall have less amount of time to handle user request of web-pages as well as analysis process. If it isn't efficient, users maybe stop using it due to waste their time.
- **Integrity**: The system data shall be live data. It will be fetch from twitter for each analysis process.

- **Usability**: The system shall be very easy to use. It is usable by a large group of people how are interested in listed airlines. System user shall be familiar with the system from the second-time use. If the user found that, the system is difficult, the user will never use it again.
- **Flexibility**: The system web-pages shall have responsive pages. It can be used in a variety of mobile devices and computers with different screen resolutions.
- **Interoperability**: The system shall be able to interface with HTML browser to handle interactions between users and system. It is required to interface with Twitter API to fetch tweets.

2.3 Software Requirement:

Table 5: show the software requirement

Operating System	Supported Browser List
Windows Workstation Browsers	
Windows 8.1	Chrome Firefox Internet Explorer 11
Windows 10	Chrome Edge Firefox Internet Explorer 11
Mac Browsers	
Mac OS X 10	Chrome Firefox Safari 9

The table shows type of browser and operating system that you need to use to access web application.

2.4 Hardware Requirement:

Table 6: show the hardware requirement

Operating system	Windows* Server, Linux*, or any operating system that can run as a webserver, capable of delivering HTML5 content, including JSON and MP4.
Processor	Intel® Celeron® Processor 847, 1.10 GHz, or equivalent.
Storage	Between 1.3 GB - 2.3 GB depending on the language version.
RAM	Minimum of 512 MB. The recommended amount can vary depending on number of users connected, number of websites hosted on the device, and other factors.
Hard Disk	3 GB of available hard-disk space for installation; additional free space is required during installation. You cannot install on removable flash storage devices.

The table 6 shows requirements to achieve sentimental analysis process.

2.5 Use case Diagrams:

Usecase-1: Sentiment Analysis for twitter data.

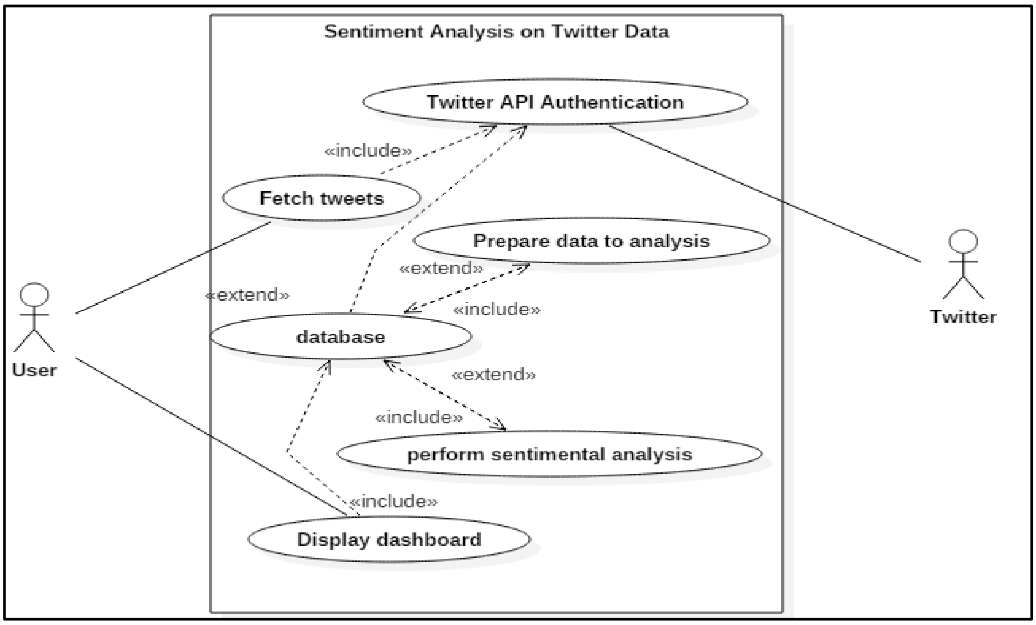


Figure 2: Sentiment Analysis on Twitter Data Use Case Diagram

Table 7: UC-01 Sentiment analysis on twitter data

Sentiment Analysis on Twitter Data		
Description	Perform sentiment analysis on twitter data and compare between different airlines in the list.	
Actors	User, Twitter	
Main flow	Request	Response
	1. select required airlines form list. 2. press analysis button. 3. request possible hashtags from DB. 5. use twitter API to request tweets. 7. start preparing data to analysis process. 8. perform sentimental analysis on data 9. display representation of result to user.	4. replay from DB with data required. 6. check authentication and replay with tweets.

The above table depicts how the system allows the user to choose required airlines (from the list of airlines provided), to perform a sentimental analysis. On click analysis button. System will start sentimental analysis process.

Usecase-2: for fetch tweets.

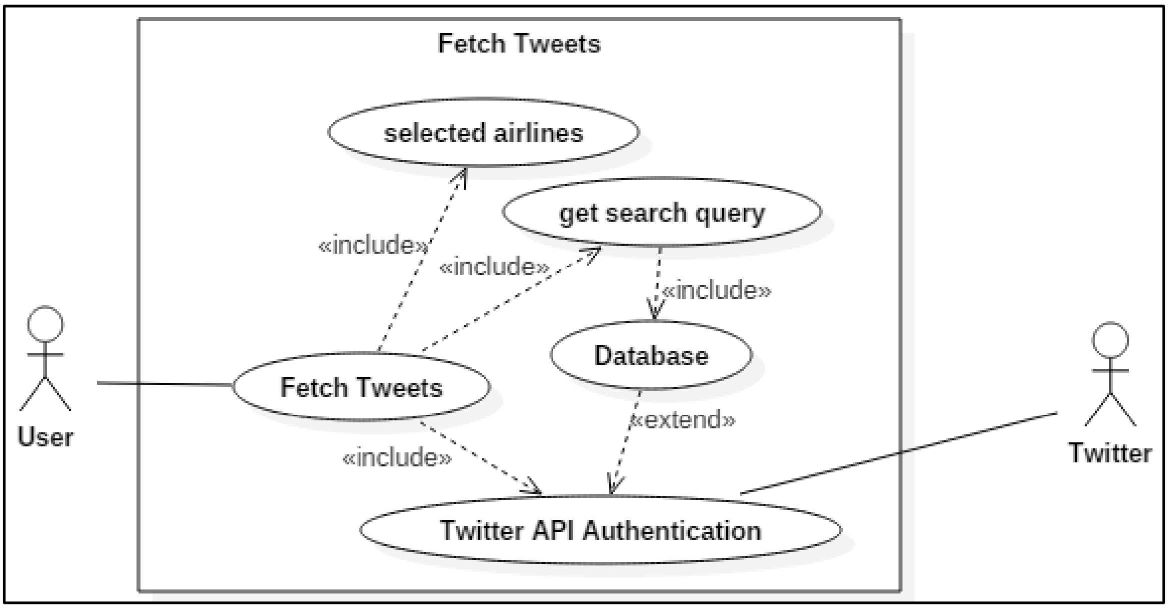


Figure 3: Fetch Tweets Use Case Diagram

Table 8: UC-02: Fetch Tweets

Fetch Tweets		
Description	Search for tweets related to selected airlines based on search query using twitter API	
Actors	User, Twitter	
Main flow	Request	Response
	1. Select needed airlines form list. 2. Press analysis button. 3. Get query form DB. 5. Generate request and send using twitter API authentication library. 7. Store data into DB.	5. responded with required data 6. Process request and return tweets as response.

The above table depicts how the system selected airline the data start gathering from twitter API by search query and store into databased.

Usecase-3: for twitter API authentication

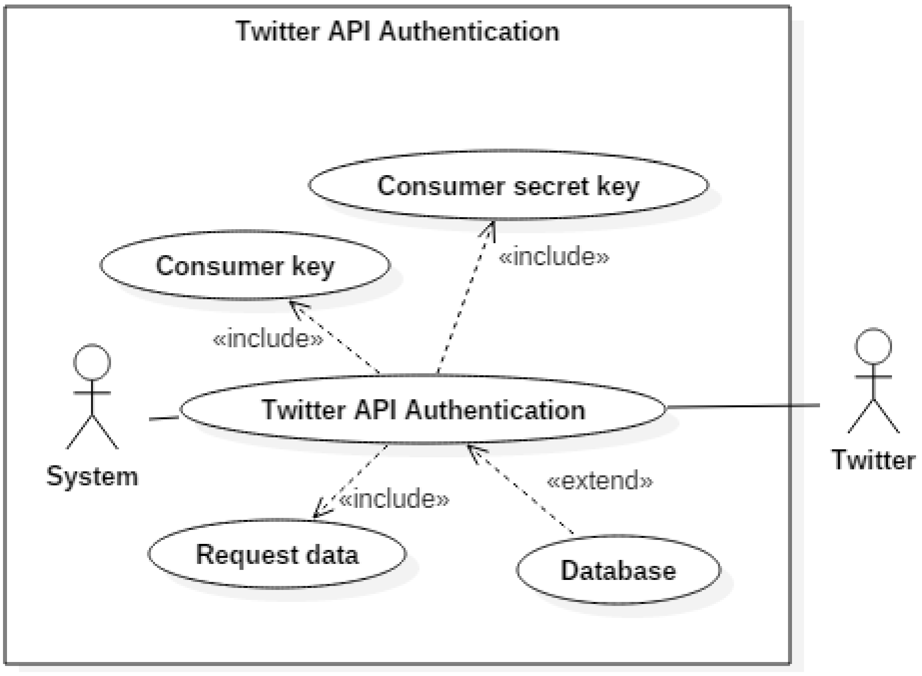


Figure 4: Twitter API Authentication Use Case Diagram

Table 9: UC-03: Twitter API Authentication

<table>
<tr><th colspan="2">Twitter API Authentication</th></tr>
<tr><td>Description</td><td>Enable system to communicate to twitter by handling all required petameters for authentication process.</td></tr>
<tr><td>Actors</td><td>System, Twitter</td></tr>
<tr><td>Main flow</td><td><table>
<tr><td>Request</td><td>Response</td></tr>
<tr><td>1. Get Consumer key and secret key of twitter app.
2. Request Data to be send to twitter API.
3. Send request.
6. submit data to requested function</td><td>4. Check authentication keys.
5. Response with data if authentication process successfully complete.</td></tr>
</table></td></tr>
</table>

The above table depicts how the system at authentication communicate with twitter by getting all required parameters to process stating with the consumer key, secret key twitter will check then request data.

Usecase-4: for prepare data to analysis.

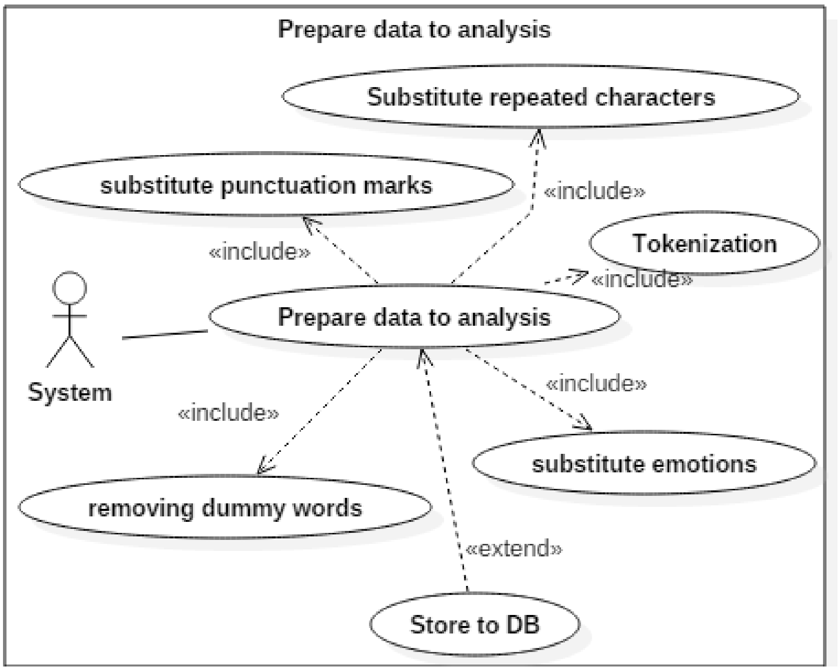


Figure 5: Prepare data to analysis Use Case Diagram

Table 10: UC-04: Prepare data to analysis

Prepare data to analysis		
Description	Prepare tweets for sentimental analysis process by remove dummy words, substitute emotions, quotations marks, URLs, usernames and perform Tokenization.	
Actors	System, DB	
Main flow	Request	Response
	1. get data from DB. 3. removing sop words, punctuations marks, URLs and usernames. 4. substitute repeated characters and emotions. 5. perform tokenization. 6. send to DB.	2. DB respond with data. 7. stored in DB.

The above table depicts how the system starts Prepare tweets for sentimental analysis process by remove stop words, links, quotations marks and substitute emotions then stored into DB.

Usecase-5: for perform sentimental analysis.

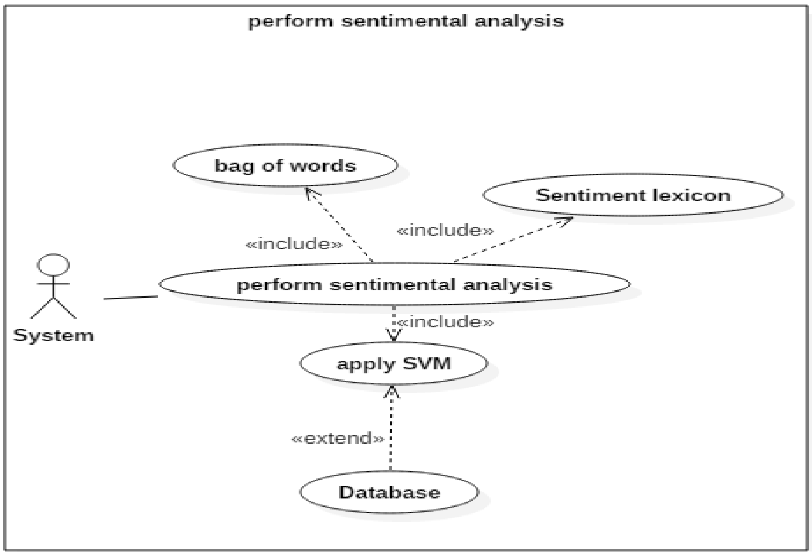


Figure 6: perform sentimental analysis Use Case Diagram

Table 11: UC-05: perform sentimental analysis

<table>
<tr><th colspan="2">Perform sentimental analysis</th></tr>
<tr><td>Description</td><td>Take input as bag of words and check sentimental value for each word from lexicon which result in three reports good, bad and neural.</td></tr>
<tr><td>Actors</td><td>System, DB</td></tr>
<tr><td>Main flow</td><td><table>
<tr><td>Request</td><td>Response</td></tr>
<tr><td>1. get data from DB which is bag of words.
3. check sentimental value for each word using SVM Algorithm.
4. send result which are three reports into DB.</td><td>2. respond with data.
5. stored in DB.</td></tr>
</table></td></tr>
</table>

The above table depicts how the system Take the input as bag of words and check sentimental value for each word from training data then apply SVM, the result in three reports positive, negative and neural.

Usecase-6: for display dashboard.

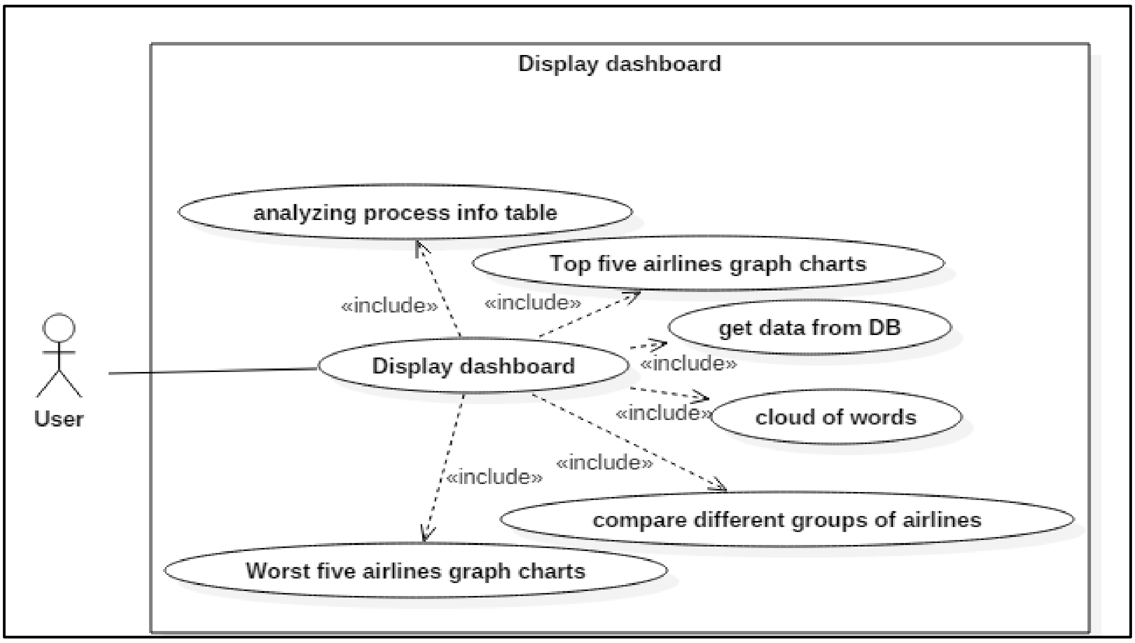

Display dashboard
analyzing process info table
Top five airlines graph charts
get data from DB
cloud of words
compare different groups of airlines
Worst five airlines graph charts
Display dashboard
User
«include»
«include»
«include»
«include»
«include»
«include»

Figure 7: Display dashboard Use Case Diagram

Table 12: UC-06: Display Dashboard

Display dashboard		
Description	Show result of sentimental analysis in useful form using graphical representors.	
Actors	System, DB, User	
Main flow	Request	Response
	1. get data from DB which are three statistical reports. 3. show information about analyzing process. 4. table to show top five airlines. 5. table to show worst five airlines. 6. cloud of words to show frequently repeated words. 7. Table to compare different group of airlines.	2. responded with data.

The above table depicts how the system show the result as graphical representation a and table of top/worst five airlines also compression of different group of airlines and cloud of words.

2.6 Identification of alternative solutions and justification of selecting a solution

2.6.1 Alternative Solutions.

a “**an Approach to Sentiment Analysis the Case of Airline Quality rating”** [21]:

In this work, the system using airline data from Twitter. based on sentiments for three major airlines (AirTran Airways, Frontier and SkyWest Airlines) from tweets. The live tweets used in this experiment contain 452 tweets on AirTran, 499 on Frontier Airlines and 195 on SkyWest Airlines, they test their data tested on a limited number of tweets and the results are very good accuracy but with big data most of the time give bad accuracy. The system used R language and the algorithm is Naïve Bayes Algorithm.

b “Online Social Media-based Sentiment Analysis for U.S Airline companies” [22]:

In this work, the dataset contains various tweets on different airline companies operate in the US. The “Twitter Airline Sentiment” dataset was obtained from Kaggle contains tweets covering six U.S. airline companies with a total number of (14,640) tweets, each of which is labelled according to sentiment polarity as: positive, negative, and neutral. (i.e. AdaBoost, Decision Tree, Linear SVM, Naïve Bayes, Random Forest, K-NN, and Kernel SVM). these algorithms implement in Python language. Some of the applied classifiers have a good outperformed than others.

c “**Sentiment Analysis Applied to Airline Feedback to Boost Customers’ Endearment”** [23]:

In this work, used dataset of passenger’s review collected from airline forum. sentiment analysis engine that uses different pre-processing strategies and machine learning approach to determine the polarity of passenger’s review for four airline companies. They applied Linear SVM and Naïve Bayes algorithm. The classifier was trained using 1217 positive reviews and 955 negative reviews. The trained classifier was made to predict the polarity for 868 reviews.

Table 13: Summary of case study

System	No. Of Airlines	Data Set	Algorithm	Result
Case 1	3	Live data	Naïve Bayes	Graphs
Case 2	6	Dataset	AdaBoost, Decision Tree, Linear SVM, Naïve Bayes, Random Forest, K-NN, and Kernel SVM	Graph for all algorithm
Case 3	4	Dataset	Naïve Bayes and LSVM	plotted as bar graph
Our system	10	Live data	SVM	Graph and word cloud

2.6.2 Justification of proposed solution:

The system will be gathering the data form Twitter API as live data, so we make sure that our data is updated so that give trust with analysis also twitter have more user active. In other way SVM is algorithm that give a good accuracy most of the time and not depend for size of data also we have seen Random Forest algorithm have better accuracy than SVM in related work, so we must compare both in the same data and the result as show in the table below:

Table 14: comparison between SVM&RF[25],[26].

RF/SVM	Accuracy (case1)	Accuracy (case 2)
Random Forest	94.50%	86.6%
SVM	95.25%	88.0%

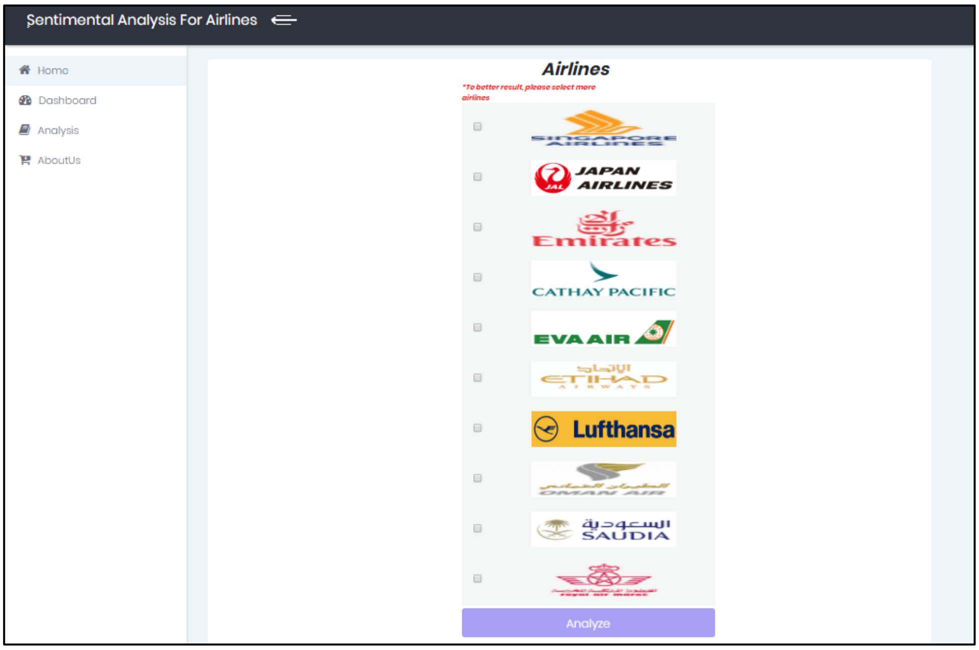
but this is not always true. it is depending in the data but from this point we chose SVM because there is no complexity in processing like RF, in addition, the prediction process using random forests is time-consuming than other algorithms [24]. The number of airlines is having more options for user. Finally give the result for user in graph way makes understanding easily.

3. System Design

3.1 User Interface Design:

a. Home page:

The first screen appear to a user is a home page, which contains best ten aircraft operate in GCC. The user can choose one or more airlines, and then press the analysis button as shown below in Fig. No. 8.



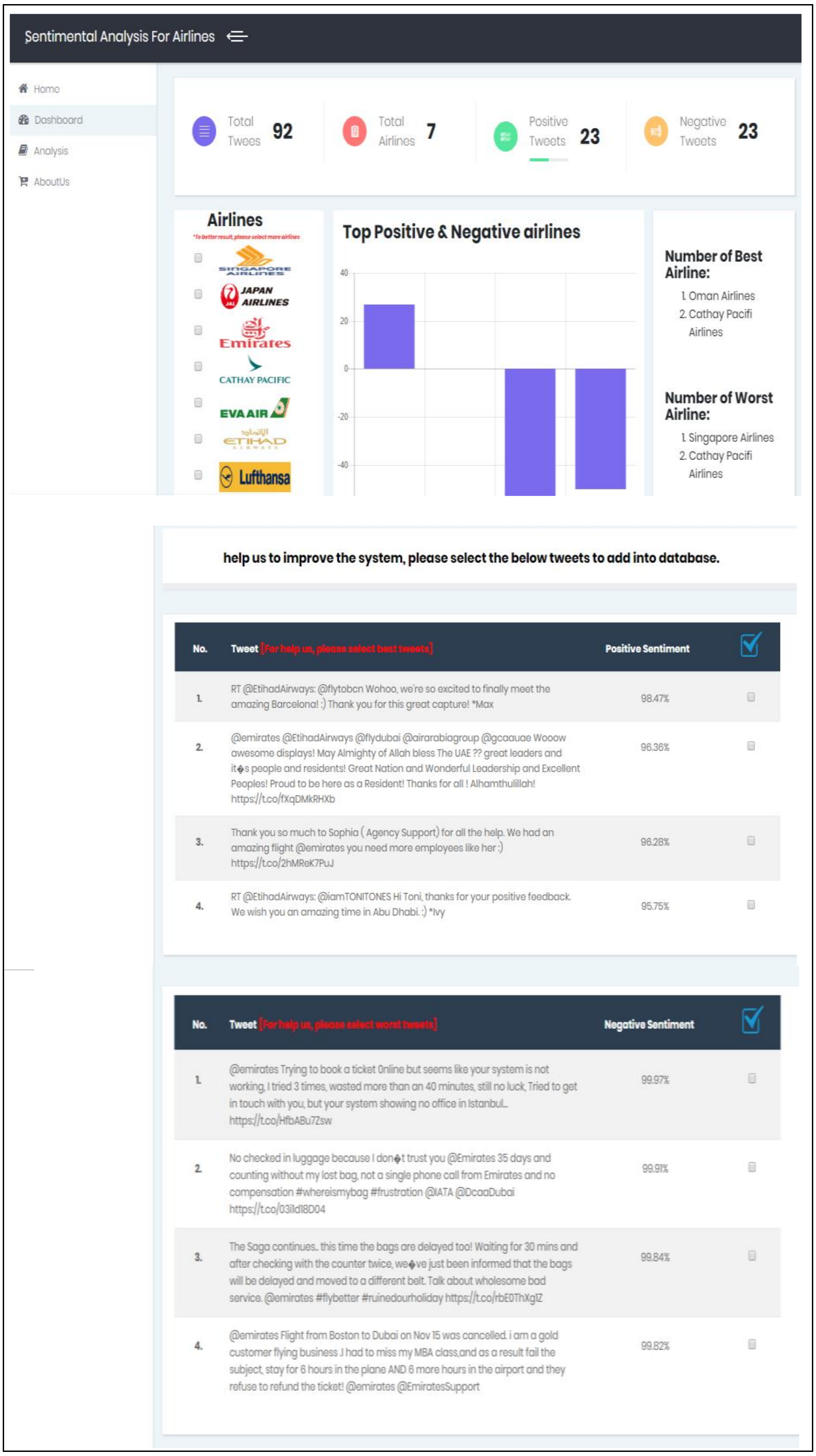
*To better result, please select more airlines

Figure 8: Home webpage

When the user selects the flights and presses the Analysis button, the system automatically analyzes live tweets for the selected flight and then moves to the Dashboard page

b. Dashboard Page:

A dashboard page contains overall statistical data about sentimental analysis process outcome. a first section shows a total number of gathered tweets, the number of selected flights and the total number of positive and negative tweets. The second section contains a selection form as in the home page to execute analysis process with different parameters easily. A bar chart to show best and worst three airlines. The third section has, A pie chart to show distribution of tweet dataset for each airline. A word cloud to represent frequent words in live dataset. The least section handle learning mechanizes. It contains table for each class (positive, neutral, and negative), each table has top five classified tweets of its class. A user can select tweet from these tables to insert into training dataset to enhance learning of the system. The Fig. No. 9 below show dashboard webpage.



Total Twees 92

Total Airlines 7

Positive Tweets 23

Negative Tweets 23

Number of Best Airline:

1. Oman Airlines
2. Cathay Pacifi Airlines

Number of Worst Airline:

1. Singapore Airlines
2. Cathay Pacifi Airlines

help us to improve the system, please select the below tweets to add into database.

No.	Tweet [For help us, please select best tweets]	Positive Sentiment	☑
1.	RT @EtihadAirways: @flytobcn Wohoo, we're so excited to finally meet the amazing Barcelona! :) Thank you for this great capture! *Max	98.47%	□
2.	@emirates @EtihadAirways @flydubai @airarabiagroup @gcaauae Wooow awesome displays! May Almighty of Allah bless The UAE ?? great leaders and it�s people and residents! Great Nation and Wonderful Leadership and Excellent Peoples! Proud to be here as a Resident! Thanks for all ! Alhamthulillah! https://t.co/fXqDMkRHXb	96.36%	□
3.	Thank you so much to Sophia (Agency Support) for all the help. We had an amazing flight @emirates you need more employees like her :) https://t.co/2hMReK7PuJ	96.28%	□
4.	RT @EtihadAirways: @iamTONITONES Hi Toni, thanks for your positive feedback. We wish you an amazing time in Abu Dhabi. :) *Ivy	95.75%	□

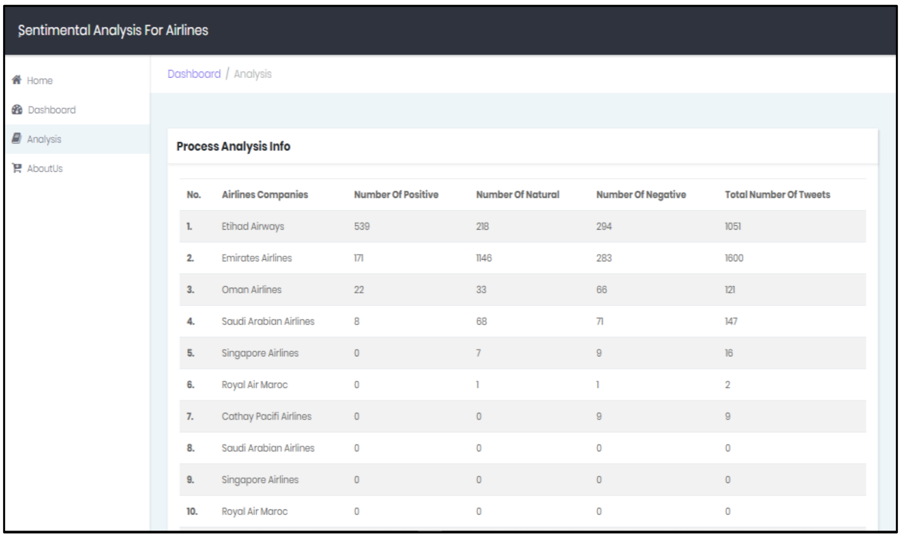
No.	Tweet [For help us, please select worst tweets]	Negative Sentiment	☑
1.	@emirates Trying to book a ticket Online but seems like your system is not working, I tried 3 times, wasted more than an 40 minutes, still no luck, Tried to get in touch with you, but your system showing no office in Istanbul… https://t.co/HfbABu7Zsw	99.97%	□
2.	No checked in luggage because I don�t trust you @Emirates 35 days and counting without my lost bag, not a single phone call from Emirates and no compensation #whereismybag #frustration @IATA @DcaaDubai https://t.co/03iId18D04	99.91%	□
3.	The Saga continues.. this time the bags are delayed too! Waiting for 30 mins and after checking with the counter twice, we�ve just been informed that the bags will be delayed and moved to a different belt. Talk about wholesome bad service. @emirates #flybetter #ruinedourholiday https://t.co/rbE0ThXgIZ	99.84%	□
4.	@emirates Flight from Boston to Dubai on Nov 15 was cancelled. i am a gold customer flying business .I had to miss my MBA class,and as a result fail the subject, stay for 6 hours in the plane AND 6 more hours in the airport and they refuse to refund the ticket! @emirates @EmiratesSupport	99.82%	□

Figure 9: Dashboard webpage

When a user in dashboard, the user can update selected airlines without return to home page. By click analysis button user can update dashboard with new data for new selected airlines. The other function user can perform in dashboard is, helping to enhance system learning. The user can check specific tweet checkbox to insert into training dataset. Update training dataset will be used to train the model in next creation of the SVM model.

c. analysis Page

An analysis page present detailed statistical data about each airline. It contains a table, where each row represents one airline and columns are analysis information related to that airline. The columns are airline name, number of positive, neutral, negative and total number of tweets for that airline. The Fig. No. 10.



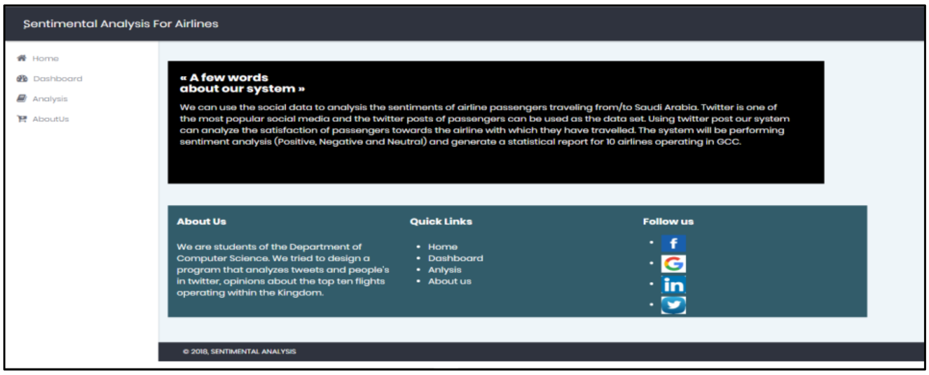
Process Analysis Info

No.	Airlines Companies	Number Of Positive	Number Of Natural	Number Of Negative	Total Number Of Tweets
1.	Etihad Airways	539	218	294	1051
2.	Emirates Airlines	171	1146	283	1600
3.	Oman Airlines	22	33	66	121
4.	Saudi Arabian Airlines	8	68	71	147
5.	Singapore Airlines	0	7	9	16
6.	Royal Air Maroc	0	1	1	2
7.	Cathay Pacifi Airlines	0	0	9	9
8.	Saudi Arabian Airlines	0	0	0	0
9.	Singapore Airlines	0	0	0	0
10.	Royal Air Maroc	0	0	0	0

Figure 10: analysis webpage

d. About us Page

A about us page contains a description of the system in general, in addition to the description of the developers and navigation between Home, Dashboard, analysis and About us webpages as well as social networking accounts. Fig No. 11: About us.



« A few words
about our system »

We can use the social data to analysis the sentiments of airline passengers traveling from/to Saudi Arabia. Twitter is one of the most popular social media and the twitter posts of passengers can be used as the data set. Using twitter post our system can analyze the satisfaction of passengers towards the airline with which they have travelled. The system will be performing sentiment analysis (Positive, Negative and Neutral) and generate a statistical report for 10 airlines operating in GCC.

About Us

We are students of the Department of Computer Science. We tried to design a program that analyzes tweets and people's in twitter, opinions about the top ten flights operating within the Kingdom.

© 2018, SENTIMENTAL ANALYSIS

Figure 11: About us webpage

3.2 Database Design:

a. Database Conceptual Design (ERD):

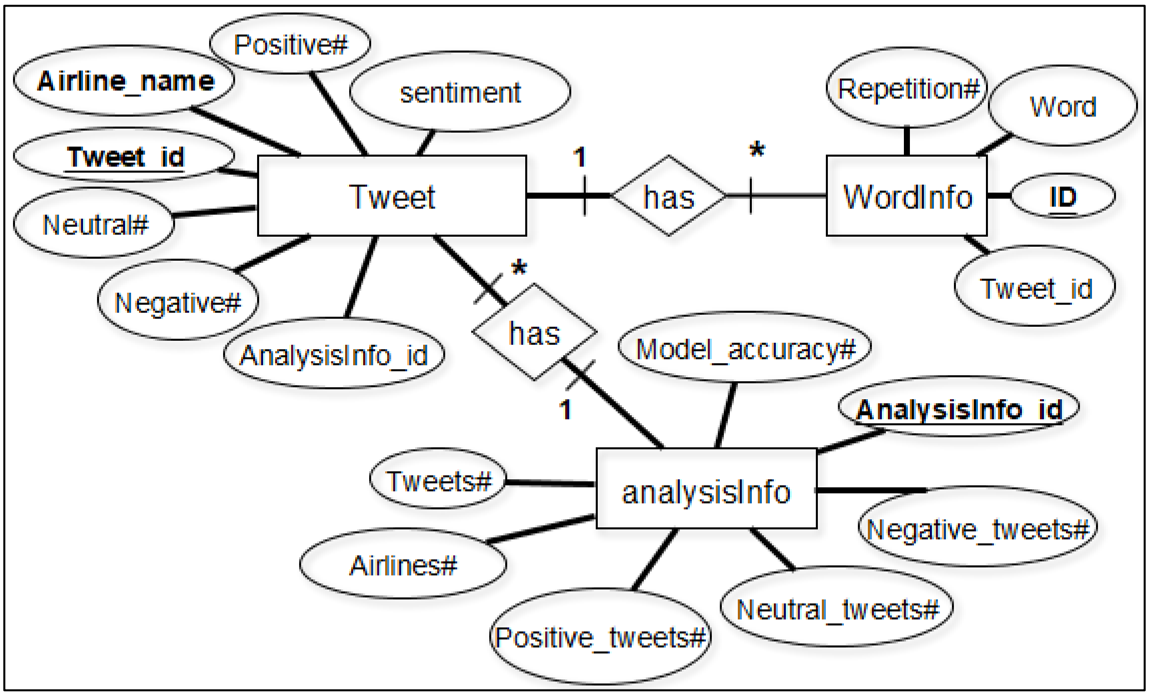
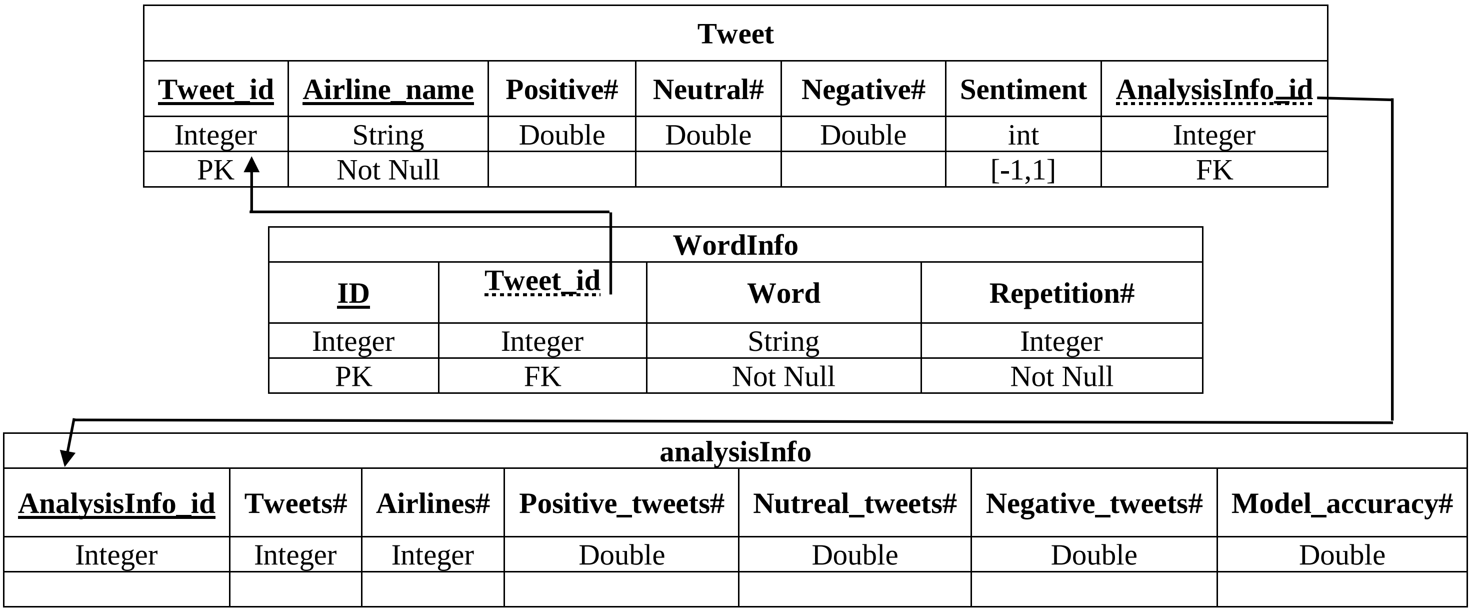


Figure 10: Database ERD

The Fig No. 10 shows database design. The system database consists of three tables which are analysisInfo, tweet, and wordInfo. The analysis table used to store overall information about an analysis process like model accuracy, total number of tweets, airlines, positive tweets, neutral tweets, negative tweets and analysisInfo id which is the primary key. The analysisInfo table has relationship with a tweet table. Each row in the analysisInfo table connect with many rows in the tweet table. The tweet table used to store information about each tweet. The targeted attributes for tweet are, airline name this tweet belong-to, statistical probability of three class, predicted sentimental value, and tweet id as primary key. The tweet table also has relationship with wordInfo table. This relation is one to many. Each row of tweet table relates to many rows of wordInfo. The wordInfo table has word-based information like which words are belong to which tweet, reputation number of individual word and ID as primary key.

b. Database Schema:



Tweet						
Tweet_id	Airline_name	Positive#	Neutral#	Negative#	Sentiment	AnalysisInfo_id
Integer	String	Double	Double	Double	int	Integer
PK	Not Null				[-1,1]	FK

WordInfo			
ID	Tweet_id	Word	Repetition#
Integer	Integer	String	Integer
PK	FK	Not Null	Not Null

analysisInfo						
AnalysisInfo_id	Tweets#	Airlines#	Positive_tweets#	Nutreal_tweets#	Negative_tweets#	Model_accuracy#
Integer	Integer	Integer	Double	Double	Double	Double

Figure 11: database Schema

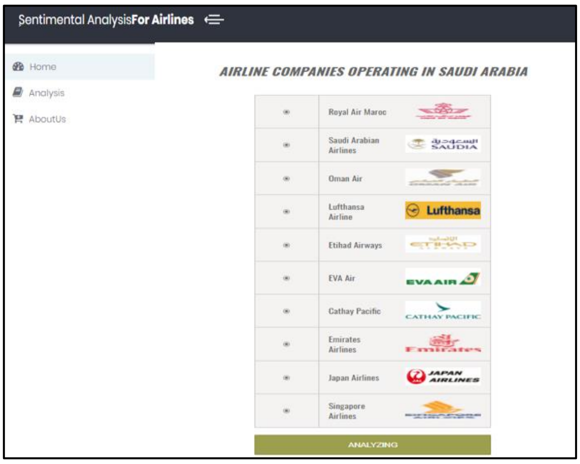
The Fig No. shows database schema. The database schema describes table attributes in details. It states data type for an individual attribute and constrains in input value. It represents relationships between tables clearly. As you can see in Fig. 11 above, analysisInfo primary key is connected to analysisInfo_id foreign key in table tweet. The tweet_id in the tweet table is connect with foreign key tweet_id in wordInfo table to form one to many relationship.

4. Details of project implementation conforming to the project proposal

Table 15:show partial implementation

TASKES	FINSHED	PROCCESSING	REMARK
User can choose airlines	✓		
Request twitter data	✓		
Filter tweet by hashtag/Keywords	✓		
Filter tweet by location	✓		Few tweets without will also be considered based on keywords and hashtags.
Get dataset	✓		
Get training data	✓		Our data is always inspiring each time running "adding new tweet to data"
Remove stop word	✓		Remove in nursery word like verb "is, are, do" the word that not affect the meaning of statement
Preform sentimental analysis by using SVM	✓		
Process linear kernel	✓		
Add new tweet to last dataset ("increase the accuracy by adding Saudi slang")	✓		Since we don't have Saudi airlines data, so we build our data by inspiring it.
Get top 5 positive tweet	✓		
Get top 5 negative tweet	✓		
Get top 5 neutral tweet	✓		
Store result	✓		
Show table comparison between airlines	✓		
Show cloud of word	✓		By storing the result as jpg type in database then show it in home page
Home page with sample GUI components	✓		
Get data from database for GUI components	✓		
Analysis page with table of analysis	✓		
Get data from database for analysis table	✓		

Table 16: design of the proposal phase



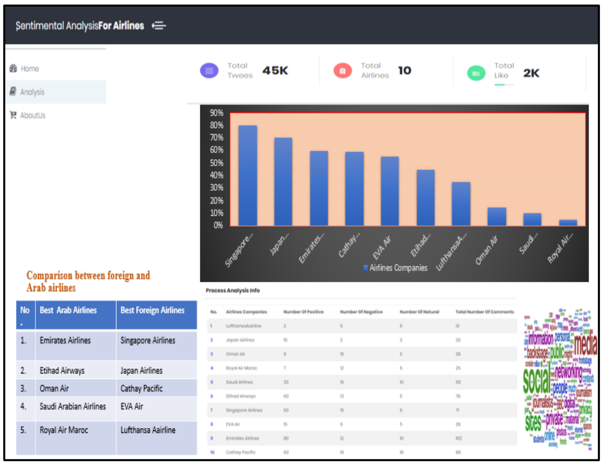
Sentimental AnalysisFor Airlines

- Home
- Analysis
- AboutUs

AIRLINE COMPANIES OPERATING IN SAUDI ARABIA

○	Royal Air Maroc
○	Saudi Arabian Airlines
○	Oman Air
○	Lufthansa Airline
○	Etihad Airways
○	EVA Air
○	Cathay Pacific
○	Emirates Airlines
○	Japan Airlines
○	Singapore Airlines

ANALYZING



Sentimental AnalysisFor Airlines

- Home
- Analysis
- AboutUs

Total Twees 45K

Total Airlines 10

Total Like 2K

Comparison between foreign and Arab airlines

No.	Best Arab Airlines	Best Foreign Airlines
1.	Emirates Airlines	Singapore Airlines
2.	Etihad Airways	Japan Airlines
3.	Oman Air	Cathay Pacific
4.	Saudi Arabian Airlines	EVA Air
5.	Royal Air Maroc	Lufthansa Aairline

Process Analysis Info

No.	Airlines Companies	Number Of Positive	Number Of Negative	Number Of Natural	Total Number Of Comments
1	LufthansaAairline	2	5	6	13
2	Japan Airlines	15	2	3	20
3	Oman Air	6	15	5	26
4	Royal Air Maroc	7	12	6	25
5	Saudi Airlines	35	15	10	55
6	Etihad Airways	60	13	5	78
7	Singapore Airlines	50	15	6	71
8	EVA Air	15	6	5	26
9	Emirates Airlines	80	12	10	102
10	Cathay Pacific	60	10	16	86

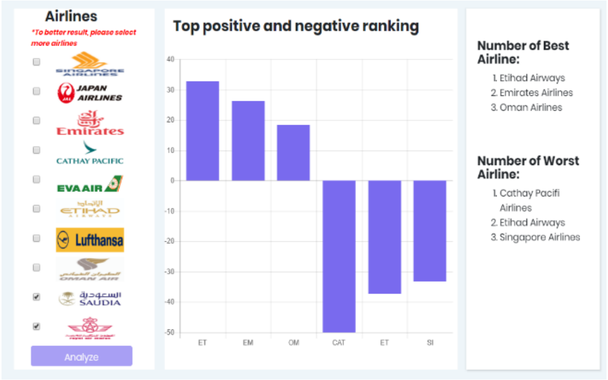
Tweet						
Tweet_id	**Airline_name**	**Positive#**	**Neutral#**	**Negative#**	**Sentiment**	**AnalysisInfo_id**
Integer	String	Double	Double	Double	int	Integer
PK	Not Null				[-1,1]	FK

WordInfo			
ID	**Tweet_id**	**Word**	**Repetition#**
Integer	Integer	String	Integer
PK	FK	Not Null	Not Null

analysisInfo						
AnalysisInfo_id	**Tweets#**	**Airlines#**	**Positive_tweets#**	**Nutreal_tweets#**	**Negative_tweets#**	**Model_accuracy#**
Integer	Integer	Integer	Double	Double	Double	Double

The table 15 above displays system parts during designing phase. It has simple of webpages as well as data base scheme.

Table 17: show design of implementation phase

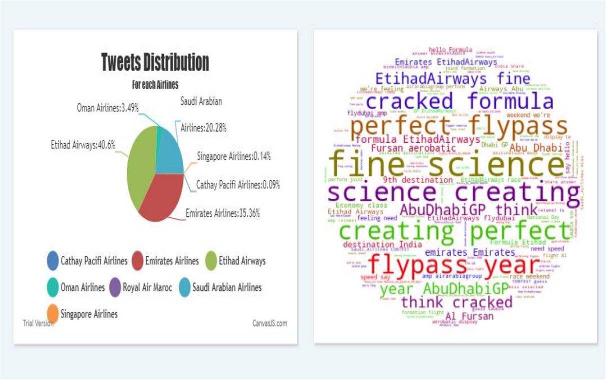


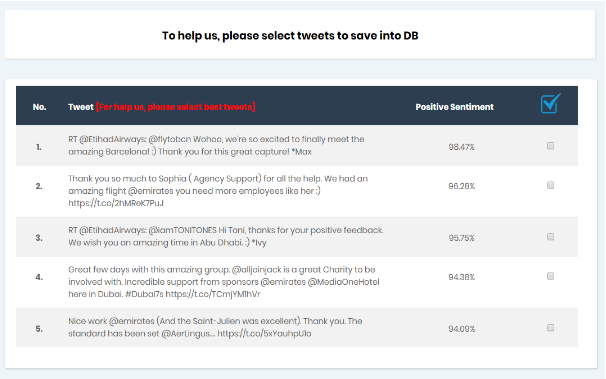
Number of Best Airline:

1. Etihad Airways
2. Emirates Airlines
3. Oman Airlines

Number of Worst Airline:

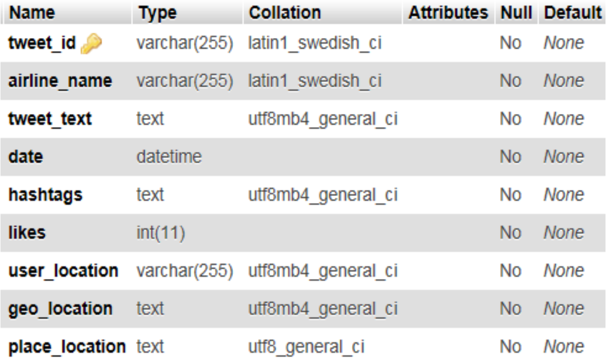
1. Cathay Pacifi Airlines
2. Etihad Airways
3. Singapore Airlines



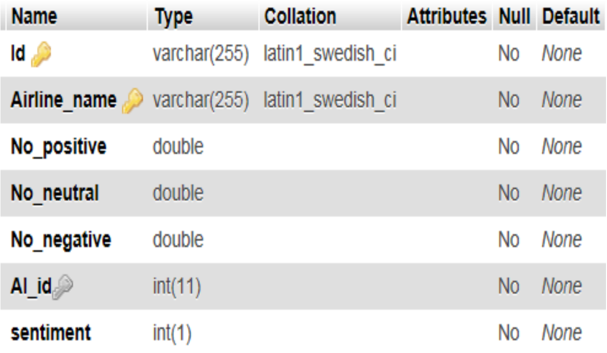


To help us, please select tweets to save into DB

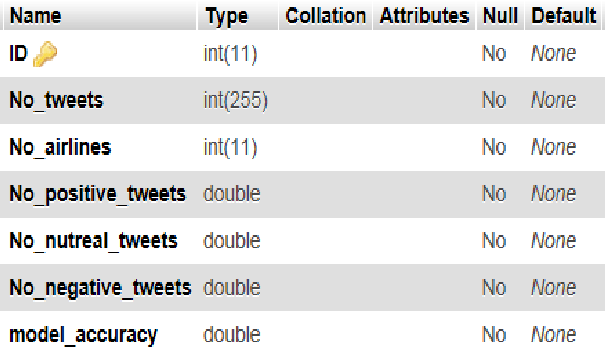
No.	Tweet [For help us, please select best tweets]	Positive Sentiment	☑
1.	RT @EtihadAirways: @flytobcn Wohoo, we're so excited to finally meet the amazing Barcelona! :) Thank you for this great capture! *Max	98.47%	□
2.	Thank you so much to Sophia (Agency Support) for all the help. We had an amazing flight @emirates you need more employees like her :) https://t.co/2hMReK7PuJ	96.28%	□
3.	RT @EtihadAirways: @iamTONITONES Hi Toni, thanks for your positive feedback. We wish you an amazing time in Abu Dhabi. :) *Ivy	95.75%	□
4.	Great few days with this amazing group. @alljoinjack is a great Charity to be involved with. Incredible support from sponsors @emirates @MediaOneHotel here in Dubai. #Dubai7s https://t.co/TCmjYM1hVr	94.38%	□
5.	Nice work @emirates (And the Saint-Julien was excellent). Thank you. The standard has been set @AerLingus.... https://t.co/5xYauhpU1o	94.09%	□



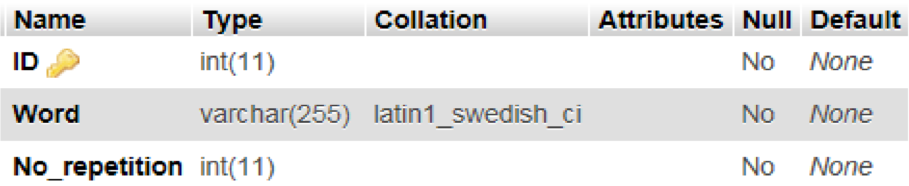
Name	Type	Collation	Attributes	Null	Default
tweet_id 🔑	varchar(255)	latin1_swedish_ci		No	None
airline_name	varchar(255)	latin1_swedish_ci		No	None
tweet_text	text	utf8mb4_general_ci		No	None
date	datetime			No	None
hashtags	text	utf8mb4_general_ci		No	None
likes	int(11)			No	None
user_location	varchar(255)	utf8mb4_general_ci		No	None
geo_location	text	utf8mb4_general_ci		No	None
place_location	text	utf8_general_ci		No	None



Name	Type	Collation	Attributes	Null	Default
Id 🔑	varchar(255)	latin1_swedish_ci		No	None
Airline_name 🔑	varchar(255)	latin1_swedish_ci		No	None
No_positive	double			No	None
No_neutral	double			No	None
No_negative	double			No	None
Al_id 🔑	int(11)			No	None
sentiment	int(1)			No	None



Name	Type	Collation	Attributes	Null	Default
ID 🔑	int(11)			No	None
No_tweets	int(255)			No	None
No_airlines	int(11)			No	None
No_positive_tweets	double			No	None
No_nutreal_tweets	double			No	None
No_negative_tweets	double			No	None
model_accuracy	double			No	None



Name	Type	Collation	Attributes	Null	Default
ID 🔑	int(11)			No	None
Word	varchar(255)	latin1_swedish_ci		No	None
No_repetition	int(11)			No	None

The table 16 shows finished parts of the system. These parts which are done in system implementation phase. The final version of the system

4.1 Mastery of tools and techniques being used in project implementation

- **Python:**

Python is one of famous and high-level programing language. Wring or reading python script is easy. Python formatting style is similar to human spiking language. It is an open source and used for different purposes. [27]

The system analysis script is implanted in python. A twitter API authentication and verification done using tweepy a python library. A preprocessing of received tweets performed in python as well. A sentiment analysis algorithm (Support Vector Machine) import form already built-in library which is sklearn. So, python implement the core part of this system.

- **Jupyter:**

The Jupyter Notebook software is a server-client application that allows editing and executing python programs in a web browser. It can be executed on a local desktop requiring no internet access or can be installed on a remote server and accessed through the internet. [34]

The Jupyter app is used to write and edit python script of this project. It has a good feature. It executes some parts of a code and print result independently.

- **PHP:**

(PHP: Hypertext Preprocessor) A scripting language that is widely used to create dynamic Web pages, and it executes on the server, while a comparable alternative, JavaScript, executes on the client, then the Web server calls PHP to interpret and perform the operations called for in the PHP script, It is a freely licensed programming language that can be used by anyone in a completely free way. [28] [29]

The system is web-based system. Request for dynamic webpages is handled in PHP. The PHP script excited and return a requested webpage with dynamic data update over time.

- **XAMPP:**

XAMPP: stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes, and it is free and open-source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. [30]

The system is web application which mean it requires a server to run it. The XAMMP is local web server used to run web-based application in regular personal computer. The XAMMP will handle requests for webpages and control communication between client and web app.

- **HTML, JS, and CSS:**

These three-client-side markup and scripting languages used in implementation to develop the web-pages of the system. HTML, or Hyper Text Markup Language, is used to create the basic structure and content of a webpage. CSS, or Cascading Style Sheets, is used for the design of a webpage – where everything is placed and how it looks, and it includes colors, layout, and fonts. It allows one to adapt the presentation to different types of devices, such as large screens, small screens, or printers. [31] [32]

JavaScript is used to define the interactive elements of a webpage that help to engage users, it was originally developed by Netscape to add dynamic and interactive elements to websites. [32]

The system webpages are developed using HTML markup language. The webpage style, format, colors and layout are done by CSS. The JavaScript helps in validation at client side. One function of JavaScript is disable analysis button, if user doesn’t select any airline from list.

- **Brackets:**

Brackets is a desktop application. It is an open source project focused on web standards and built with web technologies and is currently maintained on GitHub by Adobe and other open-sourced. [33]

The brackets software used as text editor to write HTML, CSS, JavaScript and PHP codes.

4.2 Metadata and Algorithm

4.2.1 Hashtag

Hashtag is a type of metadata tag. It consists of hash symbol followed by some keywords without space. It is used by social media sites to categories contents. It helps users of those sites to easily search and access need material based on hashtag. The system will use airline name as hashtag to search for tweets related to selected airlines. Nonetheless, the problem is that, each airline name can have multiple hashtags with small difference like underscore or concatenated by uppercase first letter of second word. To solve this problem, system should use most possible hashtags and at sign (@) airline account on twitter. This solution can handle the problem and get maximum number of tweets related to airline name.

To make efficient use of twitter social media, System should get the maximum number of tweets related to selected airlines. A most important parameter to increase the number of tweets in response is search query. Twitter API filter tweets base in the search query. A good query that explores the maximum number of possible hashtags for each search topic. The table below shows possible hashtags for all airlines list.

Table 18: airline hashtag metadata

S.No.	Airlines	Possible Hash Tags
1.	Emirates Airlines	#Emirates_Airlines #EmiratesAirlines @emirates
2.	Singapore Airlines	#Singapore_Airlines #SingaporeAirlines @SingaporeAir @EtihadAirwaysAR
3.	Japan Airlines	#Japan_Airlines #JapanAirlines @JAL_Official_jp @JALFlightInfo_e
4.	Oman Air	#Oman_Air #OmanAir @omanair
5.	Saudi Arabian Airlines	#Saudi_Arabian_Airlines #SaudiArabianAirlines #Saudi_Airlines #SaudiAirlines @Saudi_Airlines
6.	Cathay Pacific	#Cathay_Pacific #CathayPacific #Cathay and #Pacific @cathaypacific
7.	EVA Air	#EVA_Air #EVAAir @EVAAirUS
8.	Lufthansa Airline	#Lufthansa_Airline # LufthansaAirline #Airlines @LufthansaFlyer
9.	Etihad Airways	#Etihad_Airways #EtihadAirways @EtihadAirways @EtihadAirwaysAR
10.	Royal Air Maroc	#Royal_Air_Maroc #Royal_Maroc_Airlines #RoyalAirMaroc #RoyalMarocAirlines @RAM_Maroc

for Hashtags: each airline name can have multiple hashtags It helps us for easily search and access tweet based on hashtag the System should get the maximum number of tweets related to selected airlines. If the hashtags are empty this affect the gathering data for that reason we used many hashtags that possible for one airlines this is not all hashtag this is mostly the possible.

4.2.2 Algorithm description:

There are three algorithms below, out of which, one will be incorporated in this system.

a. Support Vector Machine.

b. Naive Bayes classifiers.

c. Social sentimental algorithm.

a. Support Vector Machine:

Support Vector Machine is one of machine learning algorithm which use classification. This algorithm each item describes as a point in n-dimensional (where n is number of features one has) with the value of each feature being the value of a coordinate, then fined best hyper-plan between the classifications to differentiable among it. Then each data must be labeled each data. This is called training data so that the data can be tested by comparing the result between SVM and training data. The difference will be the accuracy of your analysis. Below are pros and cons associated with SVM:

Table 19: Pros and Cons for SVM

Pros	Cons
It works well with clear margin of separation	It doesn't perform well, when we have large data set because the required training time is higher
It is effective in high dimensional spaces.	It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping
It is effective in cases where number of dimensions is greater than the number of samples.	SVM doesn't directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. It is related support vector classification (SVC) method of Python sickie-learn library.
It uses a subset of training points in the decision function (support vectors), It is memory efficient.	

b. Naïve Bayes classifiers:

This algorithm based on Bayes Theorem. It is a classification technique with assumption of independence among predictors.

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\, p(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \text{ x } P(x_2 \mid c) \text{ x } \ldots\ldots. \text{ x } P(x_n \mid c) \text{ x } P(c)$$

Where,

- P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
- P(c) is the prior probability of class.
- P(x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

Table 20: Pros and Cons for Naïve Bayes Classifiers

Pros	Cons
It is easy and fast to predict class of test data set. It also performs well in multi class prediction	If categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as "Zero Frequency". To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.	On the other side naive Bayes is also known as a bad estimator, so the probability outputs from predictor is not to be taken too seriously.
It performs well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).	Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

c. social sentimental algorithm:

This algorithm gives positive, negative, and natural of any English sentence also known as (opinion mining) which refers to using natural language processing. The output returns 4 type which is positive, negative, natural and compound. First three scales form 0 to 1.and the fourth scales from -1 to 1. [17]

Choice of best suitable algorithm:

First, it depends of available resource that are going to be used and before that sentimental analysis require some processing as the following:

- Noise Removal.
- Classification.
- Named Entity recognition.
- Subjectivity Classification.
- Feature Selection.
- Sentiment Extraction.

Now according to Anderson and others [24] who had tested their data on SVM and naïve Bayes algorithm the results obtained by the SVM sentiments classier indicated an F-Measure of 0.873 and an accuracy of 80.0% for detection of sentiment polarity. The Naive-Bayes sentiment classier presented a F-measure of 0.791 and accuracy of 72.7%.It is evident that SVM classifier increases the accuracy in 8%. Still, it not enough to say whether SVM is always better than naïve Bayes classifier, so there is a need to test with different database to reach a generalization.

However, most of the time SVM is better because:

- Doesn't depend on the size of the data but it will be hard to do the train data if its large.
- The accuracy of data is usually better in SVM.
- Naïve Bayes classifier has different type kind of complex.

Therefore, SVM is chosen to be the algorithm for this project.

Table 21: Accuracy and error rate of classification methods

Classification Methods	Error Rate	Accuracy
Lexicon-based	35%	67.7%
Naïve Bayesian	9.2%	87.4%
Bayesian Network	12.6%	86.3%
SVM	%8	88.9%
C4.5 Decision Tree	14%	85.3%

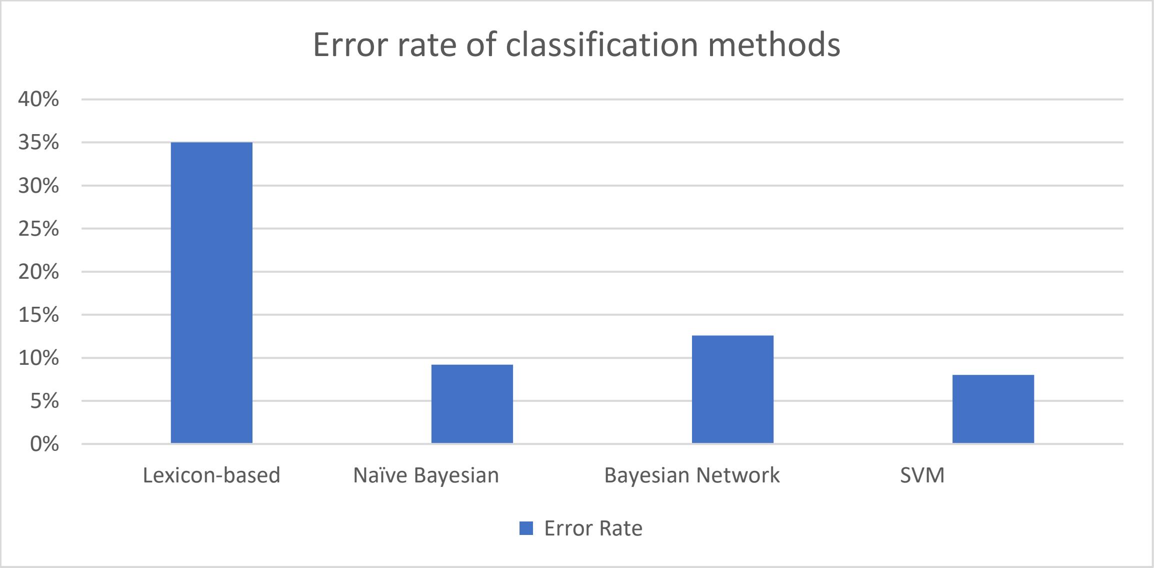


Figure 12: Three-class error rate by F-measure

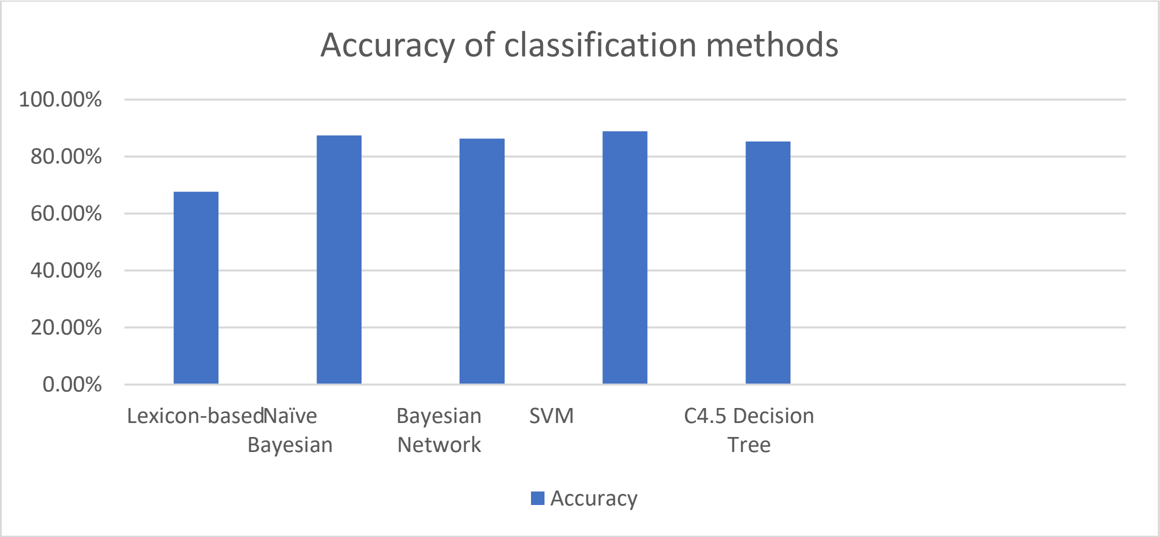
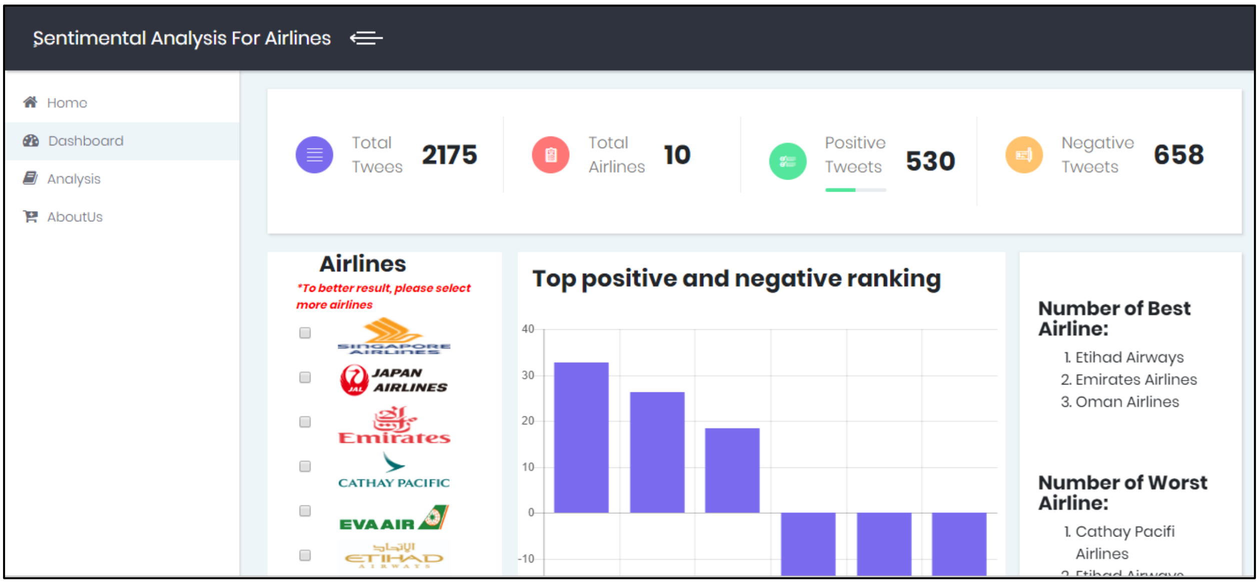


Figure 13: Accuracy of classification methods

4.3 Overall project outcome/achievements

AS Fig. No. 14-19 below shown dashboard total number of tweets, airlines, positive and negative. Also lists of airlines with checkbox and analysis bouton then Pio-chart of positive and negative ranking, in the right said number of best/worst airline. After that cloud of words and tweet description then table of top five positive, negative and neutral.



Sentimental Analysis For Airlines

Total Twees 2175

Total Airlines 10

Positive Tweets 530

Negative Tweets 658

Airlines

*To better result, please select more airlines

☐ SINGAPORE AIRLINES
☐ JAPAN AIRLINES
☐ Emirates
☐ CATHAY PACIFIC
☐ EVA AIR
☐ ETIHAD AIRWAYS

Top positive and negative ranking

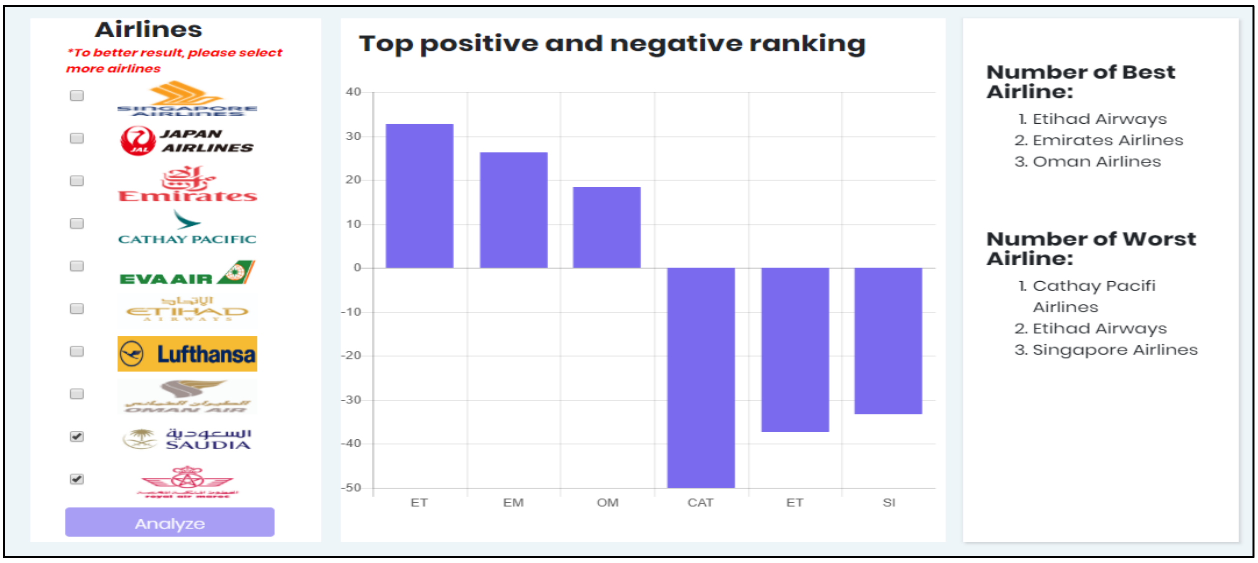
Number of Best Airline:

1. Etihad Airways
2. Emirates Airlines
3. Oman Airlines

Number of Worst Airline:

1. Cathay Pacifi Airlines
2. Etihad Airways

Figure 12: dashboard web page part 1



Airlines

*To better result, please select more airlines

☐ SINGAPORE AIRLINES
☐ JAPAN AIRLINES
☐ Emirates
☐ CATHAY PACIFIC
☐ EVA AIR
☐ ETIHAD AIRWAYS
☐ Lufthansa
☐ OMAN AIR
☑ SAUDIA
☑ royal air maroc

Analyze

Top positive and negative ranking

Number of Best Airline:

1. Etihad Airways
2. Emirates Airlines
3. Oman Airlines

Number of Worst Airline:

1. Cathay Pacifi Airlines
2. Etihad Airways
3. Singapore Airlines

Figure 13: dashboard web page part 2

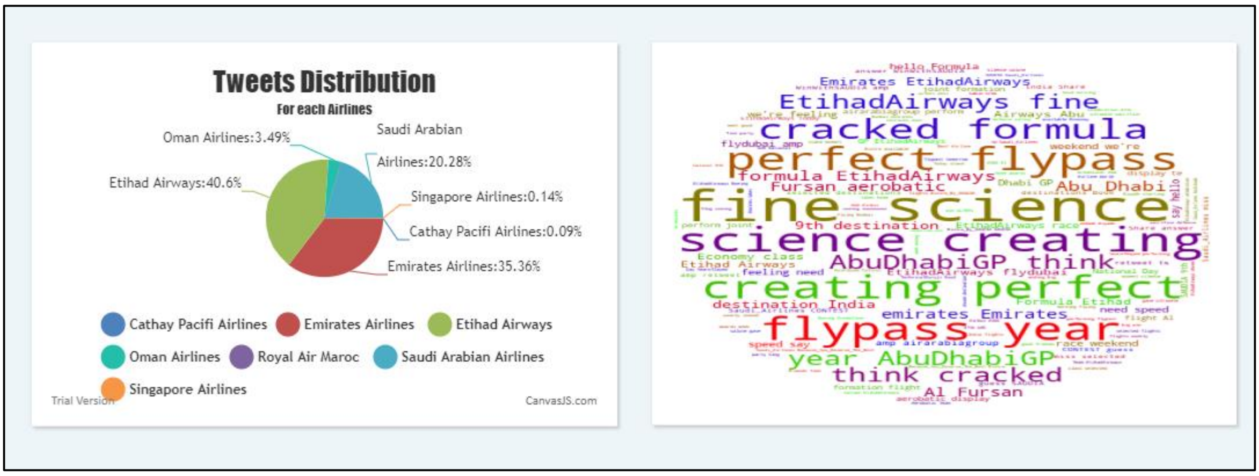
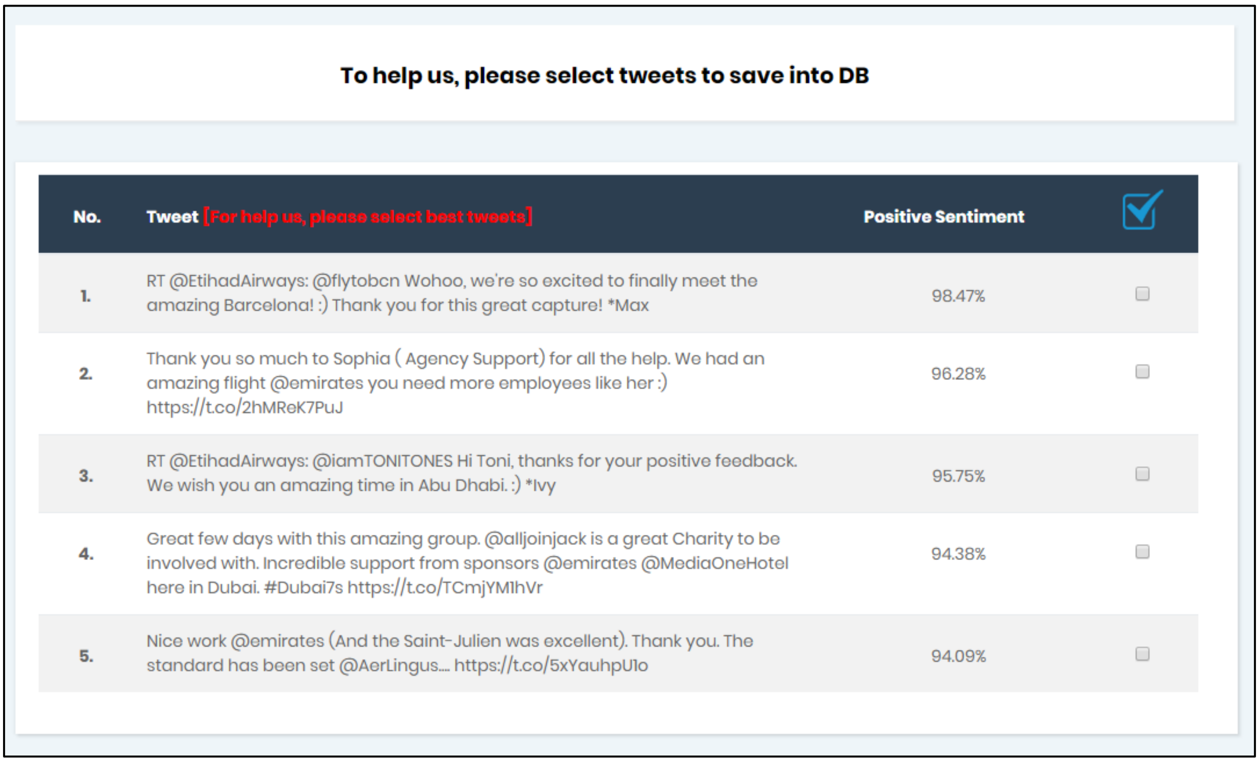


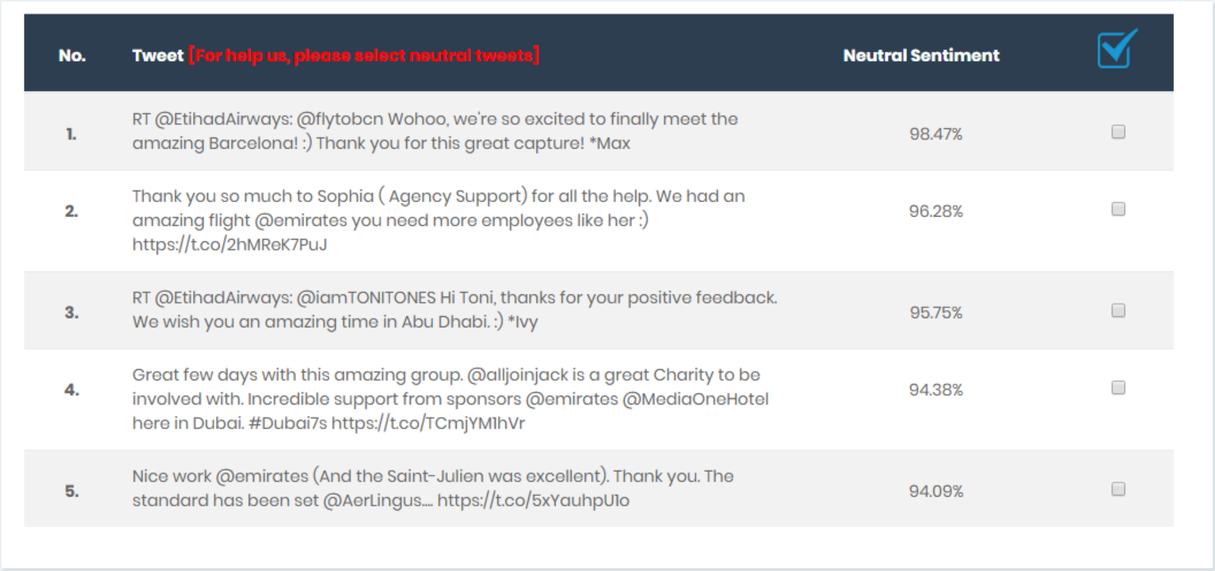
Figure 14: dashboard web page part 3



To help us, please select tweets to save into DB

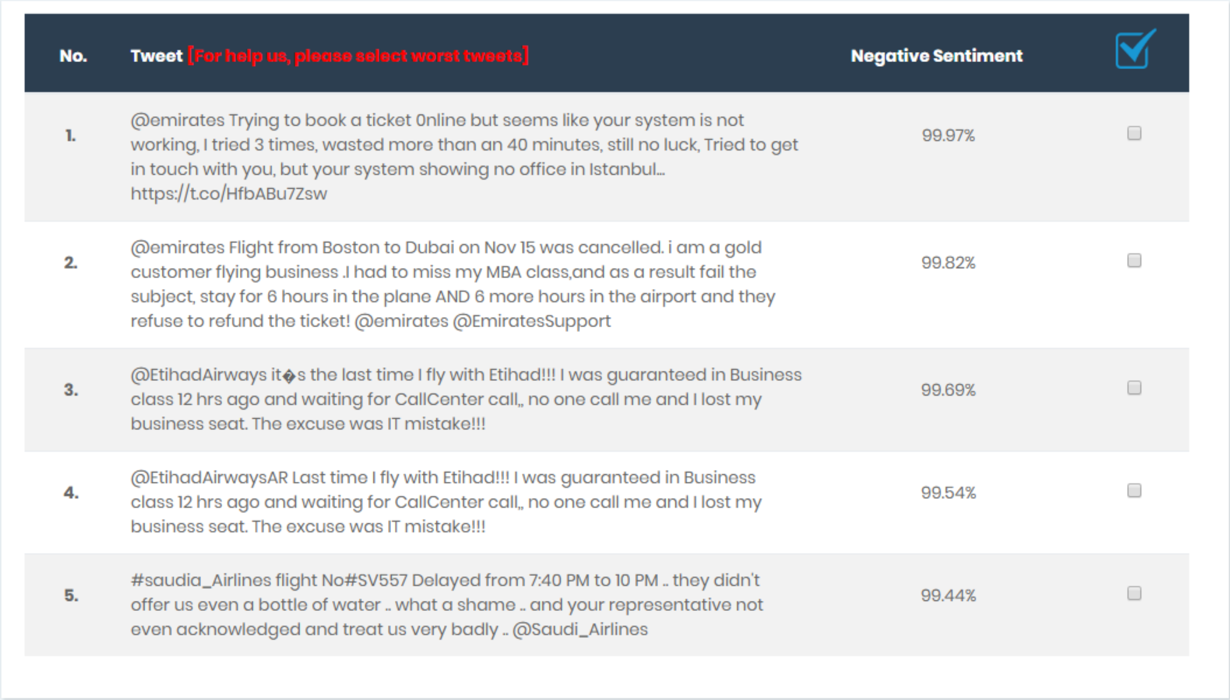
No.	Tweet [For help us, please select best tweets]	Positive Sentiment	☑
1.	RT @EtihadAirways: @flytobcn Wohoo, we're so excited to finally meet the amazing Barcelona! :) Thank you for this great capture! *Max	98.47%	☐
2.	Thank you so much to Sophia (Agency Support) for all the help. We had an amazing flight @emirates you need more employees like her :) https://t.co/2hMReK7PuJ	96.28%	☐
3.	RT @EtihadAirways: @iamTONITONES Hi Toni, thanks for your positive feedback. We wish you an amazing time in Abu Dhabi. :) *Ivy	95.75%	☐
4.	Great few days with this amazing group. @alljoinjack is a great Charity to be involved with. Incredible support from sponsors @emirates @MediaOneHotel here in Dubai. #Dubai7s https://t.co/TCmjYM1hVr	94.38%	☐
5.	Nice work @emirates (And the Saint-Julien was excellent). Thank you. The standard has been set @AerLingus... https://t.co/5xYauhpU1o	94.09%	☐

Figure 15: dashboard web page part 4



No.	Tweet [For help us, please select neutral tweets]	Neutral Sentiment	☑
1.	RT @EtihadAirways: @flytobcn Wohoo, we're so excited to finally meet the amazing Barcelona! :) Thank you for this great capture! *Max	98.47%	□
2.	Thank you so much to Sophia (Agency Support) for all the help. We had an amazing flight @emirates you need more employees like her :) https://t.co/2hMReK7PuJ	96.28%	□
3.	RT @EtihadAirways: @iamTONITONES Hi Toni, thanks for your positive feedback. We wish you an amazing time in Abu Dhabi. :) *Ivy	95.75%	□
4.	Great few days with this amazing group. @alljoinjack is a great Charity to be involved with. Incredible support from sponsors @emirates @MediaOneHotel here in Dubai. #Dubai7s https://t.co/TCmjYM1hVr	94.38%	□
5.	Nice work @emirates (And the Saint-Julien was excellent). Thank you. The standard has been set @AerLingus.… https://t.co/5xYauhpU1o	94.09%	□

Figure 16: dashboard web page part 5



No.	Tweet [For help us, please select worst tweets]	Negative Sentiment	☑
1.	@emirates Trying to book a ticket 0nline but seems like your system is not working, I tried 3 times, wasted more than an 40 minutes, still no luck, Tried to get in touch with you, but your system showing no office in Istanbul… https://t.co/HfbABu7Zsw	99.97%	□
2.	@emirates Flight from Boston to Dubai on Nov 15 was cancelled. i am a gold customer flying business .I had to miss my MBA class,and as a result fail the subject, stay for 6 hours in the plane AND 6 more hours in the airport and they refuse to refund the ticket! @emirates @EmiratesSupport	99.82%	□
3.	@EtihadAirways it�s the last time I fly with Etihad!!! I was guaranteed in Business class 12 hrs ago and waiting for CallCenter call,, no one call me and I lost my business seat. The excuse was IT mistake!!!	99.69%	□
4.	@EtihadAirwaysAR Last time I fly with Etihad!!! I was guaranteed in Business class 12 hrs ago and waiting for CallCenter call,, no one call me and I lost my business seat. The excuse was IT mistake!!!	99.54%	□
5.	#saudia_Airlines flight No#SV557 Delayed from 7:40 PM to 10 PM .. they didn't offer us even a bottle of water .. what a shame .. and your representative not even acknowledged and treat us very badly .. @Saudi_Airlines	99.44%	□

Figure 17: dashboard web page part 6

4.4 Analysis of overall result through comparison, validation or verification

Table 22: Analysis of preliminary result

Tasks	Analysis	Solution	Remark
Filter tweet by location	The tweet that doesn't have location still it will be gathering if there is a hashtag, but the location might me different from what we select or filter	Tweets that doesn't contain GEO location will be filtered by user location	The problem is this data is public source "twitter" that we can't control
Gathering data by hashtags	Some hashtags either not active or no tweet for it	That why we have used many hashtags for one airline so that we make sure that is the right data for it	
Training data	Since we don't have local slang dataset, so that may affect the accuracy.	We need to enhance the dataset by adding new tweet to it so that make our Saudi's airlines data and improve the accuracy.	

4.4.1 Test Case

Table 23: Test case 1

Project Name: Sentimental Analysis for Airlines System	
Test Case # 1	
Test Case ID: Home-01	**Test Designed by:** Ahmed Bahossain
Test Priority (Low/Medium/High): Med	**Test Designed date:** 28/11/2018
Module Name: Home screen/ airlines form	**Test Executed by: Ahmed Bahossain**
Test Title: Analysis button state	**Test Execution date:** 2/12/2018
Description: Test JavaScript functionality of analysis button to handle unselected airlines	
Pre-conditions: Website is accessible	
Dependencies:	

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Navigate to Home webpage		Access Home webpage	User is navigated to home webpage	Pass	JavaScript function handle button as expected
2	Doesn’t select any airline		No airlines selected	No airlines selected	Pass	
3	Start analysis process	Click analysis button	User should not be able to click analysis button	Analysis button unclickable	Pass	

Post-conditions:
User can access home webpage but can’t click analysis button until select some airlines from the list.

Table 24: Test case 2

Project Name: Sentimental Analysis for Airlines System	
Test Case # 2	
Test Case ID: Home-02	**Test Designed by:** Ahmed Bahossain
Test Priority (Low/Medium/High): High	**Test Designed date:** 28/11/2018
Module Name: Home screen/ airlines form	**Test Executed by: Ahmed Bahossain**
Test Title: Perform sentiment analysis process	**Test Execution date:** 2/12/2018
Description: User follow correct steps to perform sentimental analysis process for all airlines in list	
Pre-conditions: Website is accessible	
Dependencies:	

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Navigate to Home webpage		Access Home webpage	User is navigated to home webpage	Pass	System capable to perform sentimental analysis process
2	select all airlines	Check checkboxes for all airlines in list	All airlines selected	All airlines selected	Pass	
3	Start analysis process	Click analysis button	click analysis button	clickable Analysis button	Pass	
4	Redirect User to dashboard		Dashboard webpage appear to	Dashboard webpage appear to user	Pass	

Post-conditions:
Dashboard webpage with statistical data about selected airlines shown to a user.

Table 25: Test case 3

Project Name: Sentimental Analysis for Airlines System	
Test Case # 3	
Test Case ID: Home-03	**Test Designed by:** Ahmed Bahossain
Test Priority (Low/Medium/High): low	**Test Designed date:** 30/11/2018
Module Name: Home screen	**Test Executed by: Ahmed Bahossain**
Test Title: First time access redirected to Home webpage	**Test Execution date:** 1/12/2018
Description: User who access website for first time will be redirected to home webpage	
Pre-conditions: Website is accessible	
Dependencies:	

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Navigate to dashboard webpage		Redirected to Home webpage	User in home webpage	Pass	

Post-conditions: Home webpage is active.

Table 26: Test case 4

Project Name: Sentimental Analysis for Airlines System	
Test Case # 4	
Test Case ID: Home-04	**Test Designed by:** Ahmed Bahossain
Test Priority (Low/Medium/High): Low	**Test Designed date:** 31/11/2018
Module Name: Home screen/ left side menu	**Test Executed by: Ahmed Bahossain**
Test Title: Inactive webpage links for first time access	**Test Execution date:** 3/12/2018
Description: webpage links in left side menu shall be not active only for users who visit website for first time.	
Pre-conditions: Website is accessible	
Dependencies:	

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Navigate to Home webpage		Access Home webpage	User is navigated to home webpage	Pass	
2	select dashboard form left side menu	Click dashboard link	Not clickable	Not clickable	Pass	
3	select analysis form left side menu	Click analysis link	Not clickable	Not clickable	Pass	
4	select about us form left side menu	Click about us link	Not clickable	Not clickable	Pass	

Post-conditions:

User in home webpage where no option in left side menu can be clicked except the home itself.

Table 27: Test case 5

Project Name: Sentimental Analysis for Airlines System	
Test Case # 5	
Test Case ID: Dashboard-01	**Test Designed by:** Ahmed Bahossain
Test Priority (Low/Medium/High): High	**Test Designed date:** 28/11/2018
Module Name: Dashboard screen/ airlines form	**Test Executed by: Ahmed Bahossain**
Test Title: Perform sentiment analysis process from dashboard page	**Test Execution date:** 2/12/2018
Description: User should select the airlines from the dashboard page to perform sentiment analysis again.	
Pre-conditions: Website is accessible and user in dashboard webpage	
Dependencies:	

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Select Emirate, EVT, Oman, and Singapore airlines	Check checkboxes of select Emirate, EVT, Oman, and Singapore.	Airlines selected	Airlines selected	Pass	
2	Perform sentimental analysis process	Click analysis button	Button clicked, and sentimental analysis process start	Button clicked, and sentimental analysis process start	Pass	
3	Dashboard webpage refreshed		new statistical data for selected airlines	with new statistical data for selected airlines	Pass	

Post-conditions: Dashboard webpage with statistical data about new selected airlines shown to a user.

Conclusion and future work

Finally, usage of social media will be helpful in sentiment analysis. It is where many users share their opinion of products, brands, or life experiences. These feedbacks are useful for business information and to know if there are avoidable mistakes. This is truly very beneficial, especially for users who cannot make their minds in choosing which airline to select. Therefore, based on people tweets, airline companies will be rated and analyzed enabling users to get insight for what company to pick for their travels.

We plan to expand airlines list to be more benefit for large group of users not only the GCC countries passengers. The analysis can be process in both language English and Arabic, so the number of data is increase and have more opinion of people.

Work plan

a. Project Plan for Proposal:

Table 27: work plan

Tasks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Identify the topic	🟥														
Study Twitter API & Report templet		🟥	🟥												
Collect Data set and select airlines		🟥	🟥												
Detailed System Analysis			🟥	🟥											
Determine the Scope					🟥										
Analysis of Related Work						🟥	🟥								
Specifying Functional and Non-Functional Requirements								🟥	🟥						
Finding Alternative Solutions									🟥	🟥					
Specify Tools Used and Analysis											🟥	🟥			
Appropriate analysis												🟥	🟥		
Description of tools and techniques to be used for the implementation														🟥	🟥
Identified tasks and a realistic work plan for project implementation															🟥

b. Project Plan for Implementation:

Table 28: Identified tasks and a realistic work plan for project implementation

TASK	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Setup work environment	🟥														
Learning needed skills		🟥	🟥	🟥											
Create interface of the webpage					🟥	🟥									
Create the database							🟥								
Presentation millstone 4								🟥							
Implementing the basic component of the project									🟥	🟥	🟥				
Complete the web application												🟥	🟥		
Documentation														🟥	
Presentation millstone 5															🟥

References

[1] Provision of Aircraft Charter Services by Commercial Operators Technical and Operational Evaluation Criteria, TOEC. [ONLINE] Available at: www.un.org.Depts.ptd/files/files/attachment/page/2014/July%202014/TOEC_Rev% 2003.1.pdf [Accessed 1 February 2018].

[2] Saudi Arabia Airline Flights. [ONLINE] Available at: http://www.alarabiya.net/ar/aswaq/travel-and-tourism.html. [Accessed 1 February 2018].

[3] The Emirates Group Annual report 2016. [ONLINE] Available at: https://cdn.ek.aero/downloads/ek/pdfs/report/annual_report_2016.pdf. [Accessed 5February 2018].

[4] The Fast-Facts-Figures-February-2017. [ONLINE] Available at: http://www.etihad.com/Documents/PDFs/Corporate%20profile/Fast%20facts/Fast- Facts-Figures-February-2017-en.pdf.[Accessed 16February 2018].

[5] OmanAir_annualreport2016. [ONLINE] Available at: https://www.omanair.com/sites/default/files/content/about_us/pdf/2016annualreport_ eng/index.html. [Accessed 16February 2018].

[6] Middle East Airlines_ ANNUAL REPORT 2016. [ONLINE] Available at: http://www.saudiacatering.com/Admin/Content/2016-Annual-Report- PDF204201774655.pdf. [Accessed 16February 2018].

[7] Saudi Airlines_ BOARDOFDIRECTORSREPORT2016. [ONLINE] Available at:https://www.mea.com.lb/Library/Assets/BOARDOFDIRECTORSREPORT2016-042055.pdf. [Accessed 16February 2018].

[8] Miranda, M., Baltazar, M. E., & Silva, J. (2016). Airlines Performance and Efficiency Evaluation using a MCDA Methodology. The Case for Low Cost Carriers vs Legacy Carriers. Open Engineering, 6(1). Lack of customer complaints.

[9] Woodruff, R. B. (1997). Customer value: the next source for competitive advantage. Journal of the academy of marketing science, 25(2), 139.

[10] Gilbert, D., & Wong, R. K. (2003). Passenger expectations and airline services: a Hong Kong based study. Tourism Management, 24(5), 519-532.

[11] Rani, S., & Singh, J. (2017). SENTIMENT ANALYSIS OF TWEETS USING SUPPORT VECTOR MACHINE.

[12] Firmino Alves, A. L., Baptista, C. D. S., Firmino, A. A., Oliveira, M. G. D., & Paiva, A. C. D. (2014, November). A Comparison of SVM versus naive-bayes techniques for sentiment analysis in tweets: a case study with the 2013 FIFA confederations cup. In Proceedings of the 20th Brazilian Symposium on Multimedia and the Web (pp. 123-130). ACM.

[13] Adeborna, E., & Siau, K. (2014, July). An Approach to Sentiment Analysis-the Case of Airline Quality Rating. In PACIS (p. 363).

[14] Wan, Y., & Gao, Q. (2015, November). An ensemble sentiment classification system of twitter data for airline services analysis. In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on (pp. 1318-1325). IEEE.

[15] analyticsVidhya of SVM. [ONLINE] Available at: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/ [Accessed 1 March 2018].

[16] social sentimental algorithm [ONLINE] Available at: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ [Accessed 1 March 2018].

[17] Naïve Bayes classifiers [ONLINE] Available at: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ [Accessed 1 March 2018].

[18] sentimental analysis by SVM [ONLINE] Available at: Https://dl.acm.org/citation.cfm?doid=2664551.2664561 [Accessed 1 March 2018].

[19] customer-frustrations-in-the-airline-industry-with-aspect-based-sentiment-analysis [ONLINE] Available at: http://blog.aylien.com/understanding-customer-frustrations-in-the-airline-industry-with-aspect-based-sentiment-analysis/ [Accessed 1 March 2018].

[20] Online_Social_Media-based_Sentiment_Analysis_for_US_Airline_companies [ONLINE] Available at :https://www.researchgate.net/publication/315643035_Online_Social_Media-based_Sentiment_Analysis_for_US_Airline_companies [Accessed 1 March 2018].

[21] AN_APPROACH_TO_SENTIMENT_ANALYSIS_THE_CASE_OF AIRLINE_QUALITY_RATING [ONLINE] Available at http://aisel.aisnet.org/pacis2014/363/ [Accessed 13 April 2018].

[22] Online_Social_Media-based_Sentiment_Analysis_for_US_Airline_companies [ONLINE] Available at :https://www.researchgate.net/publication/315643035_Online_Social_Media-based_Sentiment_Analysis_for_US_Airline_companies [Accessed 13 April 2018].

[23] SENTIMENT ANALYSIS APPLIED TO AIRLINE FEEDBACK TO BOOST CUSTOMERS' ENDEARMENT [ONLINE] Available at: http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJAPS-50004-2.pdf [Accessed 13 April 2018].

[24] Learn Random Forest using Excel Machine Learning Algorithm [ONLINE] Available at: https://www.newtechdojo.com/learn-random-forest-using-excel/ [Accessed 13 April 2018].

[25] Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins [ONLINE] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5796274/ [Accessed 13 April 2018].

[26] COMPARISON OF MACHINE LEARNING ALGORITHMS RANDOM FOREST, ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE TO MAXIMUM LIKELIHOOD FOR SUPERVISED CROP TYPE CLASSIFICATION [ONLINE] Available at: http://mtc-m16c.sid.inpe.br/col/sid.inpe.br/mtc-m18/2012/05.15.13.21/doc/015.pdf [Accessed 13 April 2018].

[27] What is Python? - Definition from Techopedia:Techopedia.com: https://www.techopedia.com/definition/3533/python

[28] Definition of PHP-Hypertext-Preprocessor[ONLINE] Available at: https://whatis.techtarget.com/definition/PHP-Hypertext-Preprocessor. [Accessed 6 December 2018].

[29] php-hypertext-preprocessor-definition[ONLINE] Available at: https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203597-php-hypertext-preprocessor-definition/. [Accessed 6 December 2018].

[30] Xampp term [ONLINE] Available at: https://findwords.info/term/xampp. [Accessed 6 December 2018].

[31] Web Design and Applications [ONLINE] Available at: https://www.w3.org/standards/webdesign/htmlcss/. [Accessed 6 December 2018].

[32] The difference between HTML, CSS and JavaScript [ONLINE] Available at: https://www.getsmarter.com/blog/career-advice/difference-html-css-javascript/ [Accessed 6 December 2018].

[33] JavaScript Definition [ONLINE] Available at: https://techterms.com/definition/javascript. [Accessed 6 December 2018].

[34] What is the Jupyter Notebook? [ONLINE] Available at: https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html

Appendix

- **Python Code.**

```
import cgitb
import sys
import os
#to handel import error when run this script from php
cgitb.enable()
os.environ['APPDATA'] = os.environ['APPDATA'] if 'APDATA' in os.environ else 'C:\\Users\\-
\\AppData\\Roaming'

import tweepy
import csv
import pandas as pd
import pymysql
import re , nltk
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
import collections
import sys
import numpy as np
from scipy.sparse import hstack
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.multiclass import OneVsRestClassifier
from wordcloud import WordCloud
from PIL import Image
from sklearn.externals import joblib

#initialize connection to database
conn = pymysql.connect(host='127.0.0.1', user = 'pma', password = '', db = 'gp')
db_conn = conn.cursor()

####input your credentials here to authorize with twitter API
consumer_key = 'j9mFgv8fLA8ttEplIV9ZmEWFC'
consumer_secret = 'TA85GOBbsFoq5HJEacVtb5e1z6hffhSXQviff5MPKeh2dMMf02'
access_token = '957905431929618432-nv0dgXdrsUI1mHV8bTBdXiTTWta0kzH'
access_token_secret = '2GPHmEBvBKk9meYXVeDyZd3Eaztpl9ueYNzdVa5K24o83'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth,wait_on_rate_limit=True)
```

```
# Open livetweets.csv file to store tweets retrieved from twitter.
header = ["tweet_id", "airline_name", "tweet_text", "date", "hashtags", "likes"]
csvFile = open('LiveTweets.csv', 'w', newline='', encoding="utf-8")
#Use csv Writer
csvWriter = csv.writer(csvFile, delimiter=',')
csvWriter.writerow(header)

#hashtags used to retrieve tweets.
tag_list = [["Singapore Airlines_SI", "#Singapore_Airlines OR #SingaporeAirlines OR
@SingaporeAir"],
            ["Japan Airlines_JA", "#Japan_Airlines OR #JapanAirlines OR @JAL_Official_jp OR
@JALFlightInfo_e"],
            ["Emirates Airlines_EM", "#Emirates_Airlines OR #EmiratesAirlines OR @emirates"],
            ["Cathay Pacifi Airlines_CAT", "#Cathay_Pacific OR #CathayPacific OR
#Cathay_Pacifi_Airlines OR #CathayPacifiAirlines OR @cathaypacific"],
            ["EVA Airlines_EVA", "#EVA_Air OR #EVAAir OR #EVA_Airlines OR #EVAAirlines OR
@EVAAirUS"],
            ["Etihad Airways_ET", "#Etihad_Airways OR #EtihadAirways OR @EtihadAirways OR
@EtihadAirwaysAR"],
            ["Lufthansa Airline_LU", "#Lufthansa_Airline OR #LufthansaAirline OR
@LufthansaFlyer"],
            ["Oman Airlines_OM", "#Oman_Air OR #OmanAir OR #Oman_Airlines OR
#OmanAirlines OR @omanair"],
            ["Saudi Arabian Airlines_SAU", "#Saudi_Arabian_Airlines OR #SaudiArabianAirlines
OR #Saudi_Airlines OR #SaudiAirlines OR @Saudi_Airlines"],
            ["Royal Air Maroc_MAR", "#Royal_Air_Maroc OR #Royal_Maroc_Airlines OR
#RoyalAirMaroc OR #RoyalMarocAirlines OR @RAM_Maroc"]
            ]

#Filter tweets received from twitter by location of user
def filter_location(tweet, airline_name):
    ksa_reg = ['saudi','SA', 'السعودية', 'riyadh', 'makkah', 'kharj', 'dawasir', 'mecca', 'jeddah',
'taif', 'medina', 'madinah', 'qassim',
               'dammam', 'ahsa', 'batin', 'jubail', 'khobar', 'khafji', 'asir', 'jawf', 'bahah', 'najran',
'jizan', 'hail', 'tabuk',
               'SA','jeedah', 'dmm', 'saudia', 'sa', 'jed', 'alkharj',
               'الخرج', 'الشرقية', 'تنومة', 'الخفجي', 'القصيم', 'بريدة', 'الطائف', 'الاحساء', 'الثقبة', 'الباحة', 'ksa', '
نجران',
               'الدمام', 'السعوديه', 'المدينة', 'حايل', 'السعودي', 'عنيزة', 'جازان', 'الجبيل', 'مكة', 'جدة', 'السعود', '
الرياض','الخبر']
    if tweet.geo:
        return True
    elif tweet.place:
        if str(tweet.place).split('\',')[5][15:] == 'SA':
            return True
        elif str(tweet.place).split('\',')[5][15:] == 'KW':
            return True
        elif str(tweet.place).split('\',')[5][15:] == 'AE':
            return True
        elif str(tweet.place).split('\',')[5][15:] == 'QA':
            return True
```

```
      elif str(tweet.place).split('\',')[5][15:] == 'BH':
        return True
      elif str(tweet.place).split('\',')[5][15:] == 'OM':
        return True
    elif tweet.user.location:
        if set(ksa_reg).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ',
tweet.user.location).split()])):
          return True
        elif set(['kuwait', 'kw', 'الكويت']).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ',
tweet.user.location).split()])):
          return True
        elif set(['emirates', 'emiratos', 'ae', 'abudhabi', 'dhabi', 'dubai', 'دبي', 'أبوظبي', 'uae',
'الامارات']).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ', tweet.user.location).split()])):
          return True
        elif set(['qatar','qa', 'doha', 'قطر']).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ',
tweet.user.location).split()])):
          return True
        elif set(['bahrain','bh', 'البحرين']).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ',
tweet.user.location).split()])):
          return True
        elif set(['oman','om', 'عمان']).intersection(set([s.lower() for s in re.sub('[^a-zA-Zأ-ي]', ' ',
tweet.user.location).split()])):
          return True
    elif airline_name in ['Emirates Airlines_EM', 'Oman Airlines_OM', 'Saudi Arabian Airlines_SAU',
'Etihad Airways_ET']:
        return True
    return False

#read tweets from twitter based on user selection and store them in liveTweets.csv file
for index in list(map(int, sys.argv[1:])):
    for tweet in tweepy.Cursor(api.search,q= tag_list[index][1],
                           tweet_mode='extended',
                           count=10000,
                           lang="en",
                           geocode="24.488370,45.922187,1200km",
                           since="2006-04-03").items():

      hashtags = [h['text'] for h in tweet.entities['hashtags']]
      if filter_location(tweet, tag_list[index][0]):
        csvWriter.writerow([tweet.id_str, tag_list[index][0], tweet.full_text, tweet.created_at,
                       ', '.join(hashtags), tweet.favorite_count])
      sql_query="INSERT INTO `KSA_Dataset` VALUES (%s,%s,%s,%s,%s,%s,%s,%s,%s);"
      sql_data = (tweet.id_str, tag_list[index][0], tweet.full_text, tweet.created_at, ',
'.join(hashtags), tweet.favorite_count, str(tweet.user.location), str(tweet.geo), str(tweet.place))
      try:
        db_conn.execute(sql_query, sql_data)
      except pymysql.IntegrityError as e:
        print("Tweet doesn't inserted to Database.",e.args)
      conn.commit()
csvFile.close()
```

```
#read training dataset form tweets.csv file and perform preprocessing operations
tweets = pd.read_csv("Tweets.csv")
file = open("stopwords.txt")
stop_words = set(stopwords.words('english')).union(set(file.read().split("\n")))
wordnet_lemmatizer = WordNetLemmatizer()

#normalizing function
def normalizer(tweet):
    letters = re.sub("[^a-zA-Z]"," ",tweet)
    tokens = nltk.word_tokenize(letters)
    lower_case = [l.lower() for l in tokens]
    filtered_tweet = list(filter(lambda l:l not in stop_words, lower_case))
    lemmas = [wordnet_lemmatizer.lemmatize(t) for t in filtered_tweet]
    return lemmas

#apply normalizing function on tweets
tweets['normalized_tweet'] = tweets.text.apply(normalizer)

# function to create ngrams of features extracted from tweets
def ngrams(input_list):
    bigrams = [' '.join(t) for t in list(zip(input_list, input_list[1:]))]
    trigrams = [' '.join(t) for t in list(zip(input_list, input_list[1:], input_list[2:]))]
    return bigrams+trigrams

#apply ngrams function on tweets
tweets['grams'] = tweets.normalized_tweet.apply(ngrams)

#count repeat words function
def count_words(input):
    counter = collections.Counter()
    for row in input:
        for word in row:
            counter[word] += 1
    return counter

#apply count_words function
count_vectorizer = CountVectorizer(ngram_range=(1, 2))

#represent tweets as matrix based on feature vector
vectorized_data = count_vectorizer.fit_transform(tweets.text)
indexed_data = hstack((np.array(range(0, vectorized_data.shape[0]))[ : , None], vectorized_data))

#map sentiment classes into 1, 0, and -1
def sentiment2target(sentiment):
    return {
        "negative" : -1,
        "neutral" : 0,
        "positive" : 1
    }[sentiment]
targets = tweets.airline_sentiment.apply(sentiment2target)
```

```
#split dataset into testing and training sets
data_train, data_test, targets_train, targets_test = train_test_split(indexed_data, targets,
test_size=0.4, random_state=0)
data_train_index = data_train[:,0]
data_train = data_train[:, 1:]
data_test_index = data_test[:,0]
data_test = data_test[:, 1:]

#check if the model is stored or not if store import the model no need to train it again.
try:
    clf = joblib.load('gp.joblib')
except (KeyError, FileNotFoundError) as e:
    clf = False
    print(e.args)

if not clf:
    clf = OneVsRestClassifier(svm.SVC(gamma=0.01, C=100., probability=True,
class_weight='balanced', kernel="linear"))
    clf_output = clf.fit(data_train, targets_train)
    joblib.dump(clf, 'gp.joblib')

#check model accuracy.
print("accuracy of linear kernel is: ")
model_accuracy = clf.score(data_test, targets_test)
print(model_accuracy)

# read live tweets to predict their sentiment value.
liveTweets = pd.read_csv("LiveTweets.csv", encoding="utf-8")
sent = count_vectorizer.transform(liveTweets.tweet_text)
liveTweets['sentiment'] = clf.predict(sent)
liveTweets_proba = clf.predict_proba(sent)
liveTweets['negative'], liveTweets['nutral'], liveTweets['postive'] = liveTweets_proba[:,[0]],
liveTweets_proba[:,[1]], liveTweets_proba[:,[2]],
liveTweets[['tweet_text','sentiment','negative', 'nutral', 'postive']].head(20)

#insert statistical data to database
liveTweets['normalized_tweet'] = liveTweets.tweet_text.apply(normalizer)
liveTweets['grams'] = liveTweets.normalized_tweet.apply(ngrams)
liveTweets[['tweet_text','normalized_tweet', 'grams']].head()
print("predect sentimental values for new tweets.")

#create word cloud image
cloud_image = "img/circle_mask.png"
mask1 = np.array(Image.open(cloud_image))
text = " ".join(tw for tw in liveTweets["tweet_text"])
wordcloud = WordCloud(max_font_size=50, max_words=1000, stopwords=stop_words,
background_color="white", mask=mask1, colormap='brg_r').generate(text)
wordcloud.to_file("img/first_review.png")
print("finsh word cloud.")
```

```
#inset statistical data into analysis, tweet, and word tables in DB
statis = liveTweets.groupby(['sentiment']).size()
no_positive = statis[1] if 1 in statis else 0
no_negative = statis[-1] if -1 in statis else 0
no_nutreal = statis[0] if 0 in statis else 0
no_tweets = liveTweets["tweet_id"].count()
db_conn.execute("DELETE FROM `tweet`;")
conn.commit()
db_conn.execute("DELETE FROM `analysisinfo`;")
conn.commit()
sql_query = 'INSERT INTO `analysisinfo` (`id`, `no_tweets`, `No_airlines`, `no_positive_tweets`,
`no_nutreal_tweets`, `no_Negative_tweets`, `model_accuracy`) VALUES(%s,%s,%s,%s,%s,%s,%s);'
sql_data = (1, int(no_tweets), len(sys.argv[1:]), float(no_positive), float(no_nutreal),
float(no_negative), float(model_accuracy))
try:
    db_conn.execute(sql_query, sql_data)
except pymysql.IntegrityError as e:
    print("Tweet doesn't inserted to Database.",e.args)
conn.commit()

sql_query = "INSERT INTO `tweet` VALUES(%s, %s, %s, %s, %s, %s, %s, %s);"
for live_tweet in liveTweets.itertuples():
    sql_data = (live_tweet.tweet_id, live_tweet.airline_name, live_tweet.tweet_text,
float(live_tweet.postive), float(live_tweet.nutral), float(live_tweet.negative), 1,
int(live_tweet.sentiment))
    try:
        db_conn.execute(sql_query, sql_data)
    except pymysql.IntegrityError as e:
        print("Tweet doesn't inserted to Database.",e.args)
    conn.commit()

db_conn.execute("DELETE FROM `wordinfo`;")
conn.commit()
sql_query = 'INSERT INTO `wordinfo` (`id`, `Word`, `No_Repetition`) VALUES(%s,%s,%s);'
count = 1
bag_of_words=()
for data in liveTweets[['grams']].apply(count_words)['grams'].most_common(30):
    sql_data = (count, data[0], data[1])
    count +=1
    try:
        db_conn.execute(sql_query, sql_data)
    except pymysql.IntegrityError as e:
        print("Tweet doesn't inserted to Database.",e.args)
conn.commit()
```

- **PHP and HTML code of Dashboard page.**

```
<?php
// check if user first time access web set redirect to home page
if(!isset($_COOKIE['USER']))
        header("location: index.php");
$page="Dashboard";
include("header.php");
include("db.php");

// retrieve statistical data from DB
$query="SELECT No_tweets, No_airlines, No_positive_tweets, No_negative_tweets FROM
`analysisinfo`;";
$analysis_row=mysqli_fetch_assoc(mysqli_query($con,$query));

$query="SELECT id,airline_name,sentiment, COUNT(Id) as tweet_number, SUM(IF (sentiment = -
1, 1, 0))*-1 as negative, SUM(IF (sentiment = 1, 1, 0)) as postive  FROM `tweet` GROUP BY
Airline_name;";
$tweet_table = mysqli_fetch_all(mysqli_query($con,$query));

$query="SELECT id,tweet_text, No_positive FROM `tweet` ORDER BY No_positive DESC LIMIT 5;";
$postive_tweets_5 = mysqli_fetch_all(mysqli_query($con,$query));
$query="SELECT id,tweet_text, No_Negative FROM `tweet` ORDER BY No_negative DESC LIMIT
5;";
$negative_tweets_5 = mysqli_fetch_all(mysqli_query($con,$query));
for($i = 0;$i < sizeof($tweet_table);$i++)
{
        $tweet_table[$i][4] = round($tweet_table[$i][4]/$tweet_table[$i][3]*100,2);
        $tweet_table[$i][5] = round($tweet_table[$i][5]/$tweet_table[$i][3]*100,2);
}
?>
<!—HTML code -- >
<!-- Dashboard Counts Section-->
     <section class="dashboard-counts no-padding-bottom"style="padding-top: 20px">
      <div class="container-fluid">
       <div class="row bg-white has-shadow">
        <!-- Item -->
        <div class="col-xl-3 col-sm-6">
         <div class="item d-flex align-items-center">
          <div class="icon bg-violet"><i class="icon-bars"></i></div>
          <div class="title"><span>Total<br>Twees</span>
          </div>
          <div class="number"><strong><?php echo
$analysis_row['No_tweets'];?></strong></div>
         </div>
        </div>
        <!-- Item -->
        <div class="col-xl-3 col-sm-6">
         <div class="item d-flex align-items-center">
          <div class="icon bg-red"><i class="icon-padnote"></i></div>
          <div class="title"><span>Total<br>Airlines</span>
          </div>
          <div class="number"><strong><?php echo
$analysis_row['No_airlines'];?></strong></div>
         </div>
        </div>
        <!-- Item -->
        <div class="col-xl-3 col-sm-6">
         <div class="item d-flex align-items-center">
          <div class="icon bg-green"><i class="icon-list-1"></i></div>
          <div class="title"><span>Positive<br>Tweets</span>
           <div class="progress">
            <div role="progressbar" style="width: 50%; height: 4px;" class="progress-bar bg-
green"></div>
           </div>
          </div>
          <div class="number"><strong><?php echo
$analysis_row['No_positive_tweets'];?></strong></div>
         </div>
        </div>
        <!-- Item -->
        <div class="col-xl-3 col-sm-6">
         <div class="item d-flex align-items-center">
          <div class="icon bg-orange"><i class="icon-check"></i></div>
          <div class="title"><span>Negative<br>Tweets</span>
          </div>
         <div class="number"><strong><?php echo
$analysis_row['No_negative_tweets'];?></strong></div>
         </div>
        </div>
       </div>
      </div>
     </section>

     <!-- Dashboard airlines selection list     -->
     <section class="dashboard-header"style="padding-top: 20px">
      <div class="container-fluid">
       <div class="row">
        <!-- Statistics  ""-->
        <div class="statistics col-lg-3 col-12" style="margin-right: none;" >
        <div class="" style="background-color: white;width: 250px;height:590px;margin-right: px;">
        <div style="margin-left:30px; ">
        <h1 style="margin-left:40px;;height: 30px;padding-
top:10px"><strong>Airlines</strong></h1>
          <P style="margin-left:50px;color:#EC7063;margin-left: 0px;color:red;width: 200px;font-
size: 8px;margin-bottom:0px;"><strong><em>*To better result, please select more
airlines</em></strong></P>
      <form name="selected_airlines" onsubmit="return noo()" method="post">
          <table cellpadding="4px"  >
            <tr>
              <th >
```

```
                        <input  type="checkbox" class="airline" style="margin-right: 20px;" value="0"
name="Singapore" onclick="return validate_selections()" <?php if(in_array('0',
explode(',',$_COOKIE['USER'])))echo 'checked';?>/>
                       </th> td>
                          <img src="img/air/s11.jpg" width="120 px" height="40px">
                       </td></tr><tr> <th>
                         <input type="checkbox" class="airline" name="Japan" onclick="return
validate_selections()" value="1" <?php if(in_array('1', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th> <td>
                          <img src="img/air/9.jpg" width="120 px" height="40px">
                       </td></tr> <tr> <th>
                          <input type="checkbox" class="airline" name="Emirates" onclick="return
validate_selections()" value = "2" <?php if(in_array('2', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th> <td>
                          <img src="img/air/8.png" width="120 px" height="40px">
                       </td></tr> <tr> <th>
                          <input type="checkbox" class="airline" name="Cathay" onclick="return
validate_selections()" value = "3" <?php if(in_array('3', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th> <td>
                          <img src="img/air/7.png" width="120 px" height="40px">
                       </td></tr> <tr> <th>
                          <input type="checkbox" class="airline" name="EVA Air" onclick="return
validate_selections()" value="4" <?php if(in_array('4', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th> <td>
                          <img src="img/air/6.jpg" width="120 px" height="40px">
                       </td></tr> <tr> <th>
                          <input type="checkbox" class="airline" name="Etihad" onclick="return
validate_selections()" value="5" <?php if(in_array('5', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th>
                        <td>
                          <img src="img/air/5.jpg" width="120 px" height="40px">
                       </td></tr><tr><th>
                        <input type="checkbox" class="airline" name="Lufthansa" onclick="return
validate_selections()" value="6" <?php if(in_array('6', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                       </th><td>
                          <img src="img/air/4.png" width="120 px" height="40px">
                       </td></tr><tr><th>
                        <input type="checkbox" class="airline" name="Oman" value="7" onclick="return
validate_selections()" <?php if(in_array('7', explode(',',$_COOKIE['USER'])))echo 'checked';?>/>
                       </th><td>
                          <img src="img/air/oman333.jpg" width="120 px" height="40px">
                       </td></tr><tr><th>
                          <input type="checkbox" class="airline" name="Saudi" onclick="return
validate_selections()" value="8" <?php if(in_array('8', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
```

```
                    </th><td>
                      <img src="img/air/2.jpg" width="120 px" height="40px">
                    </td></tr><tr><th>
                      <input type="checkbox" class="airline" name="Royal" onclick="return
validate_selections()" value="9" <?php if(in_array('9', explode(',',$_COOKIE['USER'])))echo
'checked';?>/>
                    </th><td>
                      <img src="img/air/1morro.png" width="120 px" height="40px">
                    </td></tr> </table>
<div style="margin-top:5px;  ">
            <input type="submit" id="submitbtn" class="btn btn-primary" disabled=""
onclick="load_page()" style="width: 180px; height:33px;" value="Analyze"></div>
</form></div>
</div>
          </div>

          <!-- Bar Chart for best and worst three airlines   -->
          <div class="chart col-lg-6 col-12"  >
            <div class="" style="background: white;width: 520px;height: 590px;margin-right:
20px;">
              <div class="title" style="font-size: 1.3rem;"><h2><strong>Top Positive & Negative
airlines</strong></h2></div>
             <canvas id="barChartHome" style="width: 520px;height: 480px;display: block;margin-
right: 0px;padding-left: 0px;margin-top: 35px;"></canvas>
            </div>
          </div>
          <div class="chart col-lg-3 col-12" style="margin-bottom: 0px;" >
            <div class="work-amount card" style="
  margin-bottom: 0px;
">
              <div class="card-body" style=" height: 590px;" style="
  margin-bottom: 0px;
">

              <h2 style="margin-top: 30px; font-size: 1.3rem; "><strong>Number of Best
Airline:</strong></h2>
              <nav>
                <ol>
                  <?php
$avg = sizeof($tweet_table)/2;
$fit = $tweet_table;
 ($i = 0 ;$i < $avg && $i < 3;$i++)
{
        $max = -1;
        $index = 0;
        for($j = 0 ;$j < sizeof($fit);$j++)
        {
                if($max < $fit[$j][5])
                {
                        $max = $fit[$j][5];
                        $index = $j;
```

```
                                        }
                            }
            echo '<li>',explode("_",$fit[$index][1])[0],'</li>';
            $fit[$index][5] = -10000;
            }
?>
                      </ol>


                  </nav><br><br>
                  <h2 font-size: 1.3rem;><strong>Number of Worst Airline:</strong></h2>
                  <nav>
                      <ol>
                          <?php
                         $avg = sizeof($tweet_table)/2;
                         $fit = $tweet_table;
                         for($i = 0 ;$i < $avg && $i < 3;$i++)
                         {
                                     $min = 1;
                                     $index = 0;
                                     for($j = 0 ;$j < sizeof($fit);$j++)
                                     {
                                                 if($min > $fit[$j][4])
                                     {
                                     $min = $fit[$j][4];
                                     $index = $j;
                                     }}
            echo '<li>',explode("_",$fit[$index][1])[0],'</li>';
            $fit[$index][4] = 10000;
                         }?>
                      </ol>
                  </nav>
                  <div class="chart text-center">
                    <canvas id="pieChart"></canvas>
                  </div>
                </div>
               </div>
             </div>
           </div>
         </div>
        </section>
<!--page footer-->
        <?php
                                     include("footer.php");
                                     ?>
   <!-- Javascript files-->
   <script src="https://code.jquery.com/jquery-3.2.1.min.js"></script>
   <script src="vendor/popper.js/umd/popper.min.js"> </script>
   <script src="vendor/bootstrap/js/bootstrap.min.js"></script>
 </html
```

- **PHP and HTML code of Analysis page.**

```
<?php
//check if the user first time access the web set if user redirect to the home page.
        if(!isset($_COOKIE['USER']))
        header("location: index.php");
  $page="Analysis";
  include("header.php");
        include"db.php";

//retrieve statistical data from database.
        $query = "SELECT Airline_name, sentiment, COUNT(Id) as
tweet_number,COUNT(No_positive) as totale_sentiment FROM `tweet` GROUP BY Airline_name,
sentiment;";
        $tweet_table = mysqli_fetch_all(mysqli_query($con,$query));
        $airlines = ['Emirates Airlines_EM', 'Singapore Airlines_SI', 'Japan Airlines_JA', 'Oman
Airlines_OM', 'Saudi Arabian Airlines_SAU', 'Cathay Pacifi Airlines_CAT', 'EVA Airlines_EVA',
'Lufthansa Airline_LU', 'Etihad Airways_ET', 'Royal Air Maroc_MAR'];
        $analysis = [];
        for ($i = 0; $i < sizeof($airlines) ; $i++)
        {
                $analysis[$airlines[$i]] = array(0, 0, 0, 0);
        }

        for($i = 0;$i < sizeof($tweet_table);)
        {
                $pos = 0;
                $neu = 0;
                $neg = 0;
                if(isset($tweet_table[$i][1]) && $tweet_table[$i][1] == -1) {
                        $neg = $tweet_table[$i][2];
                        $i++;}
                else {
                        $neg = 0; }
                if(isset($tweet_table[$i][1]) && $tweet_table[$i][1] == 0) {
                        $neu = $tweet_table[$i][2];
                        $i++;}
                else {
                        $neu = 0; }
                if(isset($tweet_table[$i][1]) && $tweet_table[$i][1] == 1) {
                        $pos = $tweet_table[$i][2];
                        $i++;}
                else {
                        $pos = 0; }
                $analysis[$tweet_table[($i-1)][0]] = array($pos, $neu, $neg, $pos+$neu+$neg);
        }
ksort($analysis);
arsort($analysis);
?>
```

```
     <!—Breadcrumb HTML code-->
     <div class="breadcrumb-holder container-fluid">
      <ul class="breadcrumb">
       <li class="breadcrumb-item"><a href="dashboard.php">Dashboard</a></li>
       <li class="breadcrumb-item active">Analysis</li>
      </ul>
     </div>
     <section class="tables">
      <div class="container-fluid">
       <div class="row">
        <div class="col-lg-12">
         <div class="card">
          <div class="card-header d-flex align-items-center">
           <h3 class="h4">Process Analysis Info</h3>
          </div>
          <div class="card-body">
<!-- Analysis table -->
           <table class="table table-striped">
            <thead>
             <tr>
              <th>No.</th>
              <th>Airlines Companies</th>
              <th>Number Of Positive</th>
              <th>Number Of Natural</th>
              <th>Number Of Negative</th>
              <th>Total Number Of Tweets</th>
             </tr>
            </thead>
            <tbody>
              <?php $i=1; foreach($analysis as $k=>$airline){?>
              <tr>
              <th scope="row"><?php echo $i++;?>.</th>
              <td><?php echo explode('_',$k)[0];?></td>
              <td><?php echo $airline[0];?></td>
              <td><?php echo $airline[1];?></td>
              <td><?php echo $airline[2];?></td>
              <td><?php echo $airline[3];?></td>
                                                 </tr>
                                                 <?php }?>
          </tbody>
           </table>
          </div>
         </div>
        </div>
       </div>
      </div>
     </section>
     <!-- Page Footer-->
 <?php include("footer.php") ?>
 </body>
</html>
```

- **JavaScript code.**

```
// when user click analysis button this function handles submit data to server and show process
circle
function load_page() {
            loadProgress();
            var airlines_selected = document.getElementsByClassName("airline");
             var formdata = new FormData();
             var data = [];
              for(var i = 0 ; i < airlines_selected.length; i++)
              {
              if(airlines_selected[i].checked)
              {
                          data.push(airlines_selected[i].value);
              }
              }
            var xhttp = new XMLHttpRequest();
            xhttp.onreadystatechange = function() {
             if (this.readyState == 3 && this.status == 200) {
             if(this.responseText == ' finsh')
            document.getElementsByClassName("ani-text")[0].innerHTML = 'Performing sentiment
analysis process....';
               }
             if (this.readyState == 4 && this.status == 200) {
                         if(this.responseText == ' finsh done'){
                         var dd=new Date();
                        document.cookie = 'USER' + "=" + data.toString() + ";" + "expires="+
                        (dd.getTime()+24*60*60)+ ";path=/";
                         var urlpath = document.URL.split('/');
                         window.location.replace('http://'+urlpath[2]+'/'+urlpath[3]+'/dashboard.php');
                         }
                         else
                        {
                                    document.getElementsByClassName('preloader')[0].style.display=
                                    'none';
                                    document.getElementsByClassName("ani-text")[0].innerHTML = 'Sorry,
                                    There was problem. Try later.'
                                    }   } };
                        var dd=new Date();
                        document.cookie = 'sel' + "=" + data.toString() + ";" + "expires="+
                        (dd.getTime()+10*60)+ ";path=/";
                        xhttp.open("GET", "pro.php", true);
                        xhttp.send();}

//this function is used to disable webpage and mouse pointer then show progress circle
function loadProgress()
{
            var page = document.getElementsByClassName('home-page')[0];
            page.style.pointerEvents = 'none';
            page.style.opacity = 0.2;
            document.getElementsByClassName('wrapper')[0].style.display = 'unset';
}
```

```
// this function used to handle save selected tweets from feedback tables in dashboard page
function getFeedback(row)
{
        if (row.cells[4].firstChild.checked){
        var formdata = new FormData();
        formdata.append('op',"insert");
        formdata.append('id',row.cells[1].innerHTML);
        formdata.append('tweet',row.cells[2].innerHTML);
        formdata.append('class',row.parentNode.className);

        var xhttp = new XMLHttpRequest();
        xhttp.onreadystatechange = function() {
                // if (this.readyState == 4 && this.status == 200) alert('responsss
'+this.responseText);
        }

        xhttp.open("POST", "insertdataset.py", true);
        xhttp.send(formdata);

        }else
                {
                        var formdata = new FormData();
                        formdata.append('op',"not");
                        formdata.append('id',row.cells[1].innerHTML);
                        var xhttp = new XMLHttpRequest();
                        xhttp.onreadystatechange = function() {
                                // if (this.readyState == 4 && this.status == 200) alert('responsss
'+this.responseText);
                        }

                        xhttp.open("POST", "insertdataset.py", true);
                        xhttp.send(formdata);
                }
}
```