Imam Mohammad Bin Saud Islamic University
College of Computer and Information Sciences

Restaurants Trend Analysis for End-Users

By:

NAME	ID
Abdulrhman Saad Alburaidi	437017654
Ibrahim Abdullah Alhammad	437014473
Mohammed Abdulrhman AlBkhet	437018118

Instructor: Dr. Qaisar Abbas Muhammad Abbas

Supervisor: Dr. Abdulaziz Ibrahim Almohimeed

Supervisor approval: Signature

- # Declaration

Abdulrhman Saad Alburaidi, Ibrahim Abdullah Alhammad, and Mohammed Abdulrhman AlBkhet are the main members who developed this research project entitled “Restaurants Trend Analysis for End-Users” to complete the Bachelor degree in Computer Science. It has been declared that this project and report along with the website are all our efforts to fulfill the requirements of the final year project. There is not any information, which so every has been copied from any source. Moreover, all information from these other sources has been duly referenced and acknowledged in accordance with the University Policy on Plagiarism.

Furthermore, we declare that in completing the project, the individual group members had the following responsibilities and contributed in the following proportions to the final outcomes of the project

Student ID	Responsibility	% Contributed
437017654	Website - database(MandoDB) - Design	33%
437014473	Data Analyst,-working with data to gain insights. - Writing Report	33%
437018118	Data Analyst - Gather data - Writing Report	34%

- # Acknowledgment

An effort was made in this project, but it was not possible without the support and assistance of many doctors and people, we would like to express our deep thanks to all of them. , Specifically, Dr. Aseel Addawood Our first supporter and Dr. Ahmed Khorsi would like to thank (College of Computer and Information Sciences) for their guidance and constant supervision. We would like to express our gratitude towards my parents & supervisor (Dr. Abdulaziz Ibrahim Almohimeed) for his kind co-operation and encouragement and giving us such attention and time, which help us to complete this project. Our thanks and gratitude also go to my colleague Computer and Information Sciences to develop our skills and knowledge.

- # Abstract

There is a huge amount of information in social media and it’s increasingly growing throughout decades. We have many restaurants that provide products and the customers are always confused between the many choices they are faced with. There is an unfair evaluation that exists in Food delivery applications because the evaluation is based on their customers. The people depend to choose a restaurant on their social relationships. social media famous users only make an announcement to those who pay and this makes the good restaurants unknown to the public, friends, and families have a close circle so not everyone will know about the restaurant and the people who have not to try all restaurants. Our project extracts the opinion of people about restaurants from two resources Twitter and foursquare to help other people to choose a restaurant. There is a different way to classification text by using lexicon or machine learning algorithms or deep learning we use in our project support vector machine (SVM), Decision Tree(DT), logistic regression and CNN deep learning algorithm. The difference between our project and another project that we use deep learning and compare it with machine learning all project we saw use machine learning, We use more than one algorithm that we can take the best accuracy, we use support vector machine (SVM) which take data and classified them into two categories, Decision Tree(DT) and it's work like going from top to down from a Tree and every branch means new decision, logistic regression is used to predict result from specific values, CNN is an image take image and process by going around it many times then classified.

- **Abstract (in Arabic)**

توجد كمية كبيره من البيانات في وسائل التواصل الاجتماعي وتزداد مع مرور الوقت، يوجد الكثير من المطاعم والتي تقدم العديد من المنتجات والزبائن يتيهون من كثرة الخيارات الموجودة، كما يوجد هناك تقييم غير عادل في تطبيقات توصيل المطاعم لان التقييم يعتمد على عملائهم فقط، الناس تعتمد غالبا على علاقاتهم في اختيار المطاعم مشاهير مواقع التواصل الاجتماعي يعلنون فقط لمن يدفع لهم وهذا يجعل المطاعم الجيدة التي لا تستطيع دفع قيمة, الإعلان غير معروفه للعامة ,الأصدقاء والعائلة لديهم دائرة صغيره من المعرفة حول كل المطاعم الموجودة كما ان الناس لا تستطيع تجربة جميع المطاعم ,مشروعنا هو معرفة اراء الناس حول المطاعم من خلال تطبيقيين تويتر و فورسكوير لمساعدة الناس في اختيار المطاعم ,توجد طرق لتصنيف النصوص من خوارزميات تعلم الاله و المعاجم و خوارزمية التعلم العميق والخوارزميات التي استخدمنا في مشروعنا هي :

support vector machine (SVM), Decion Tree (DT), logistic regression and CNN deep learning algorithm.

الفرق بيننا و المشاريع الاخرى هو اننا اضفنا خوارزمية التعلم العميق الى المقارنه مع خوارزميات تعلم الالة , وقد استخدمنا اكثر من خوارزمية لكي نقوم باختيار الاكثر دقة من بينهم

support vector machine وهي تقوم بأخذ البيانات وتقسيمها إلى فئتين ,

وهي تعمل مثل النزول عن الشجرة و الأغصان تكون قرارات جديدة وينتهي التقسيم بالوصول الى الأوراق , Decision Tree

logistic regression وهي تستخدم لتوقع النتائج من خلال فئات محددة ,

CNN وهي تاخذ الصورة كمدخلة و تقوم بمعالجتها بالمرور عليها مرات عديدة ومن ثم تقسيما

Keywords

analysis, restaurant, web application, twitter, foursquare, sentiment.

List of abbreviation

SVM: support vector machine

DT : Decion Tree

CNN : Convolutional Neural Networks

API : application program interface

MongoDB : mongo data base

CSS: Cascading Style Sheets

HTML: HyperText Markup Language

SVD : Singular Value Decomposition

PCA : Principal component analysis

FP-growth : Frequent Pattern Growth

KNN : k nearest neighbor

BPNN : Back Propagation Neural Network

NB : Naïve Bayes

IEEE : Institute of Electrical and Electronics Engineers

ME : Maximum Entropy

TP : True Positive

TN : True Negative

FN : False Negative

FP: False Positive

SQL : Structured Query Language.

- # Table of Contents

Contents

- # List of Tables

- # List of Equation

- # List of Figures

CHAPTER 1: INTRODUCTION

1.1 Introduction

There is a huge amount of information in social media and it's increasingly growing throughout decades. we have many restaurants that provide products and the customers confused between the number of choices. for example: when friends go together, they spend a lot of time to select a restaurant, the restaurants get an unfair assessment from local food delivery companies because the evaluation based just on their customers. For our project, the target audience is local food delivery companies as HungerStation, Carriage, MRSOOL and Jahez and Individuals who are search about restaurants to get eat.

1.2 Aims and Objectives

The aim of this project is to:

- Analyze Arabic and English tweets based on different factors
- Cover more than one category of restaurants
- Show the ratings to end-user in dashboard

Objectives:

1. Help customers keep track of restaurants updates and reviews.
2. Assist local food delivery companies in most popular restaurants.
3. Assist customers in choosing the best restaurant that comply with their needs.
4. Provide real-time analysis for stakeholders.

1.3 Methodology

Our methodology consists of 4 phases, represented below:

Figure 1: Methodology

As seen in Figure 1 that all steps of the methodology. In this project, we will be using different tools, for code editing, we will be using Jupiter notebook editor. Also, we will be using MongoDB as database manipulation and Tableau software to represent the data. Our first phase consists of collecting data from different sources, mainly Tweets from Twitter API and restaurant information from Foursquare API. Our second phase consists of storing all the data in MongoDB[1]. Our third phase consists of analyzing data and integration using Jupyter notebook[3] for code editing using python. In our fourth phase, we will represent and visualize the data using Tableau[2]. Some many barriers or pitfalls could happen in our project, the importance of them is determining which restaurants we gather data about it. Moreover, learning more about this software and how can we use them. Lastly, knowing how can we analyze Arabic words in different dialects. Our project focus on rating restaurants based on two famous application in Saudi Arabia and this will show different results than other applications.

1.5 Team Qualifications

Name	Qualifications
Abdulrhman Saad Alburaidi	I learned java and python. I will try to learn MongoDB.
Ibrahim Abdullah Alhammad	I am data Analyst, working with data to gain insights.
Mohammed Abdulrhman AlBkhet	I learned java. I will try to learn python and JavaScript.

Table 1: Team Qualifications

1.6 Conclusion

Our project focuses on rating restaurants based on two famous social media applications used regularly in Saudi Arabia. We started with data collection and then analyzing and integrating the data. The final step is to show the results in a visual dashboard to make it easier for the user to make better decisions.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Our project is about rating restaurants using sentiment analysis, our data resources will be from Twitter and foursquare. After reading papers that solve similar problems, we found that multiple algorithms are used to classify tweets. All of these papers show different results. In this section we will be talking about how the problem begins.

2.2 Background

Our project is based mainly on two parts, part one concerning understanding the ratings of restaurants and their sentiments. However, first let's talk about how the problem that we are trying to solve begins. These days, the number of restaurants has increased, which was coupled with an increase in the number of mobile applications. These applications allowed people to contribute in the process of evaluating these restaurants. There are three types of applications that we are focusing on in our project. Social Media applications such as Twitter. Food delivery applications such as HungerStation, Carriage, MRSOOL and Jahez. Restaurants evaluation Applications such as Foursquare. Restaurants rating depends on food application, social media famous users, friends and families, but this has some disadvantages. Food applications like HungerStation ask their customers only to rate the restaurants. social media famous users only make an announcement to who pay and this makes the good restaurants unknown to the public, friends and families have a close circle so not everyone will know about the restaurant and The people who have not to try all restaurants. In this project, we are going to use sentiment analysis to know the people opinion about restaurants , Some of these applications we can integration between them to get most of the data that we need to get the most benefit from them For example integration between comment of people about restaurants in Foursquare and their talk about these restaurants in twitter that gives us more data about these restaurants and more accuracy evaluation .

2.3 Related Work

Sentiment analysis lets talking what is the sentiment analysis. sentiment analysis is Sentiment Analysis and Opinion Mining is currently an active research area. It can be defined as the classification of text based on the overall sentiments expressed by opinion holders This classification is usually done in positive, negative and neutral.Sentiment analysis is a research space in which a lot of difficult problems are to be tackled. A number of different approaches have been taken to sentiment analysis. In general, these approaches have relied on one of two techniques: either supervised or unsupervised machine learning. These techniques for sentiment analysis have been principally focused on Indo-European languages, especially English. However, to date, there are few approaches to sentiment analysis on social media's texts that have focused on Arabic language [7].

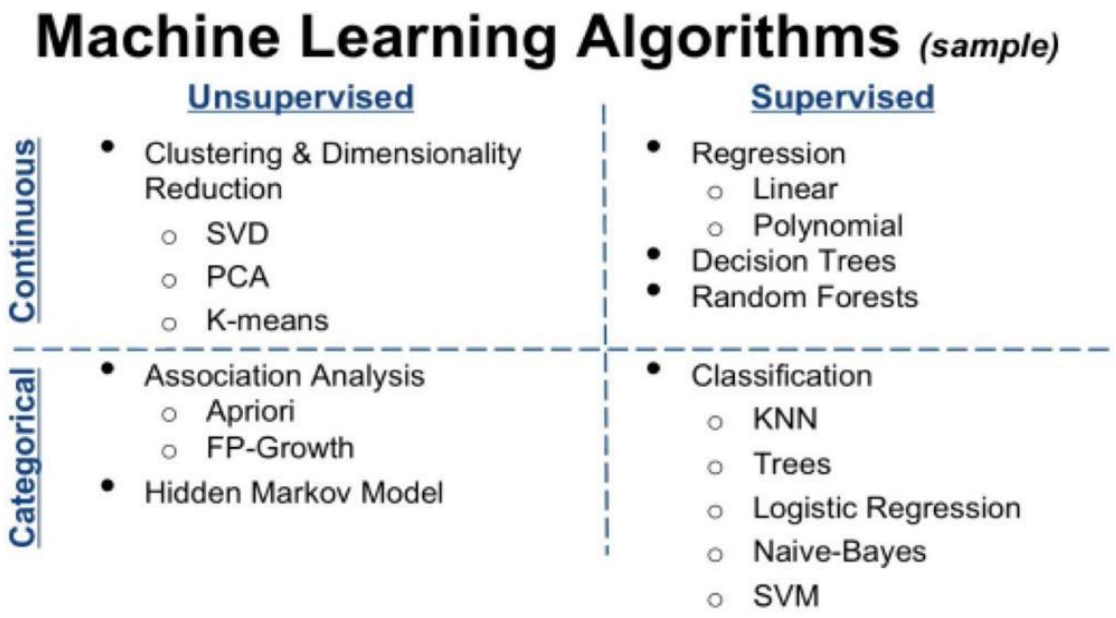

Figure 2 : Supervised vs Unsupervised Learning

Figure 2 shows there are two types of data continusous and categorical, we use in our project is supervised learning categorical data to the classification.in previous work, paper ([7] M. Hammad and M. Al-awadi, "Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning," in Information Technology: New Generations, ed: Springer, 2016, pp. 131-139.) has shown that using multiple algorithms to classification such as SVM, BPNN, NB and DT in the data set. The data set is a combination of Arabic reviews and comments from Facebook, Twitter, and YouTube resulted in SVM outperformed the Naïve Bayes classifier in

analysis and provided more consistent, reliable results in SVM achieved the highest accuracy, which is 96.06%, followed by Naïve Bayes with an average accuracy of 88.38%.

Classifier	Precision	Recall	F-measure
SVM	95.80%	96.40%	96.06%
BPNN	72.14%	67.61%	69.77%
Naïve Bayes	92.62%	84.99%	88.38%
Decision Tree	83.98%	87.99%	85.82%

Table 2: Performance results for each classifier [7].

paper([4] Snehal Kale, Vijaya Padmadas, "Sentiment Analysis of Tweets Using Semantic Analysis", 2017 IEEE), has shown that using multiple algorithms to classification such as NB and Maximum Entropy(ME) , the dataset is tweets from twitter resulted in NB achieved highest accuracy 63.9% then comes ME with 27.8% .

Methods	Accuracy
Naïve Bayes	63.9
Maximum Entropy	27.8

Table 3: Accuracy Comparison [4].

paper([3] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, Sentiment Analysis of Tweets using Machine Learning Approach, Vol. 8, 2018 IEEE), government and companies need to know what kind of speech content can be shared with audience paper has shown that using multiple algorithms to classification such as Support Vector Machine (SVM) , Decision Tree(DT) and Adaboosted D-Tree , the dataset is combination of tweets , movies reviews resulted in DT achieved the highest accuracy 84% then comes SVM 82% followed by Adaboosted D-Tree with 67%.

Sequence no	F-measure and Accuracy		
	ALGORITHM USED	*F-measure*	*Accuracy*
1	SVM	82%	82%
2	Adaboosted D-Tree	67%	67%
3	Decision Tree	84%	84%

Table 4: Value of F-measure and Accuracy [3].

paper([8] Ankita Gupta1, Jyotika Pruthi2, Neha Sahu3 Ankita Gupta et al, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 444-458.) , In this present work, the SVM and KNN based hybrid classification model is presented to process the tweet features and to identify the hidden sentiments from these tweets. this SVM and KNN based hybrid present to improve the classification accuracy. And the figure 3 show SVM and KNN based hybrid achieved 76%.

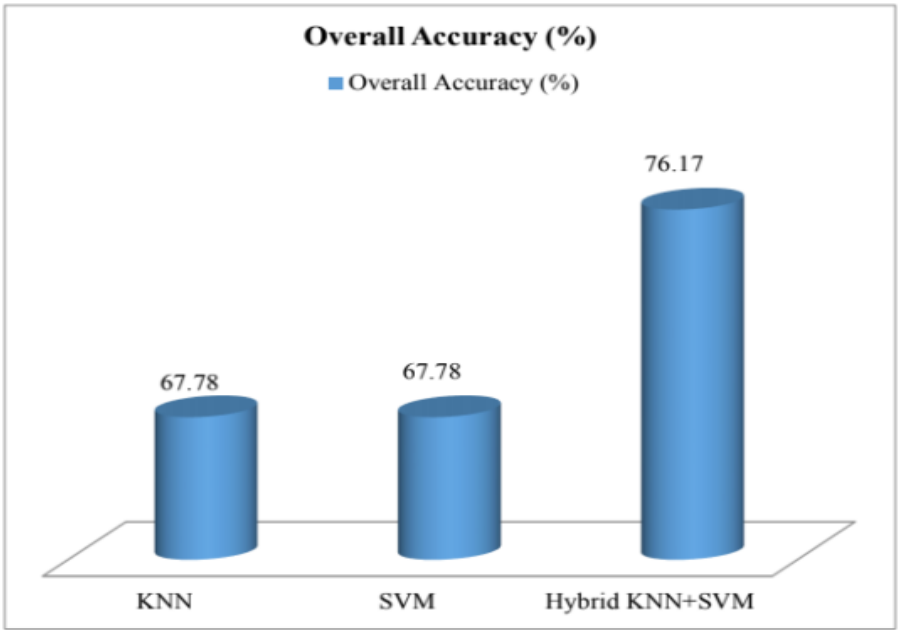

Figure 3: showing overall accuracy for 3 approaches [8].

2.4 Conclusion

There is a difference in accuracy in each paper, there is no agreement on one more accurate classifier, and there are others there are many classifiers such as: logistic regression and the CNN deep learning algorithm you did not use.

CHAPTER 3: SYSTEM ANALYSIS

3.1 Introduction

We see in the previous chapter apply Sentiment analysis in a different way and different data. And there is more than one technique to classification. We will use to implement our project structured approach because of two reasons first easier to maintain, read and understand second in our opinion the program does not need to apply the object-oriented approach. In this chapter will go more deeply about our users and system.

3.2 Software Requirements Specification

3.2.1 User Characteristics

In general, our users of all ages. Should of them know how to deal with the internet and can read, The user does not need special knowledge or experience. So our program can use by common people. However, we expect most of our users of Young people. And at the corporate level our user is Food delivery applications

3.2.2 Specific Requirements

Software show the special knowledge restaurants rating from twitter and Foursquare applications on a web application. Food will be spread by the type user want from software to be easy to use, clear in presentation, good looking, fast in response the constraints on the system will be must-have internet and on web

3.2.2.1 User Requirements

We observed Our project is in the web application field, there are many web applications in the same field. We search the internet to observe how the web application in restaurant rating is working. We find many web applications like HungerStation and Jahez. The gathered information supports our understanding of the requirements.

Functional requirements:

- User Need to use a computer or mobile phone
- User can access to the internet
- User has a web browser
- Short response time.

Nonfunctional requirement:

- Efficiency
- Available
- Reliability
- Usability

Constraints:

- Work on the web through a web browser
- Collecting data from people with a different dialect
- Access to the server

3.2.2.2 System Requirement

-Functional requirements

- The application should Be Collecting the data.
- The application Store it in the database.
- The application should Analyze data in the correct way.
- The application should Visualize results.

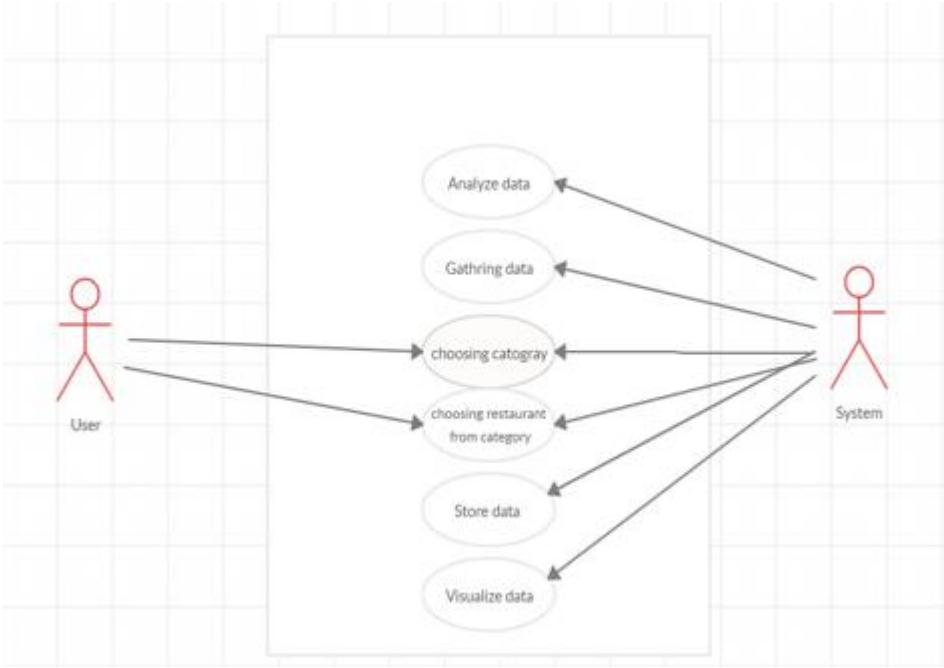

Figure 4: Use diagrams figure

- Non-Functional requirements

1-Efficience: Our analyze should have high accuracy by a good choice of a way to classification and show the result in fast time

2-available: Web applications available to users at any time.

3-Reliability: Information saved in the database whatever happens. It has a backup and prevents the analyze give wrong information

4-Usability: The application interfaces easy to used and learn for common people.

3.3 Project Management Plan

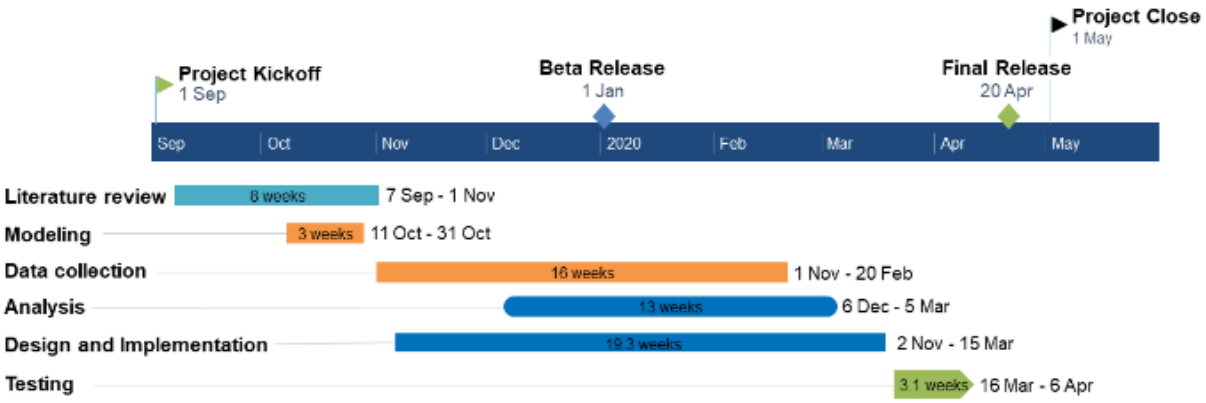

Figure 5: Timeline project

3.4 Conclusion

As seen in this chapter we faced challenges In data collection and access to the server. know what user needs and write some method to Achieve his satisfaction. We explained the interaction point between a system and user. We need to be careful about choosing the algorithm that classifies text to get the highest accuracy.

CHAPTER 4: Design And Development

4.1 Introduction

In this chapter, we will design the system architecture of the project and the user interface design, then we will explain our approach in the System and the methodology. Moreover, we show all the steps that we took to build our system.

4.2 System Architecture

System architecture conceptual model that defines the structure, behavior, and more views of a system.

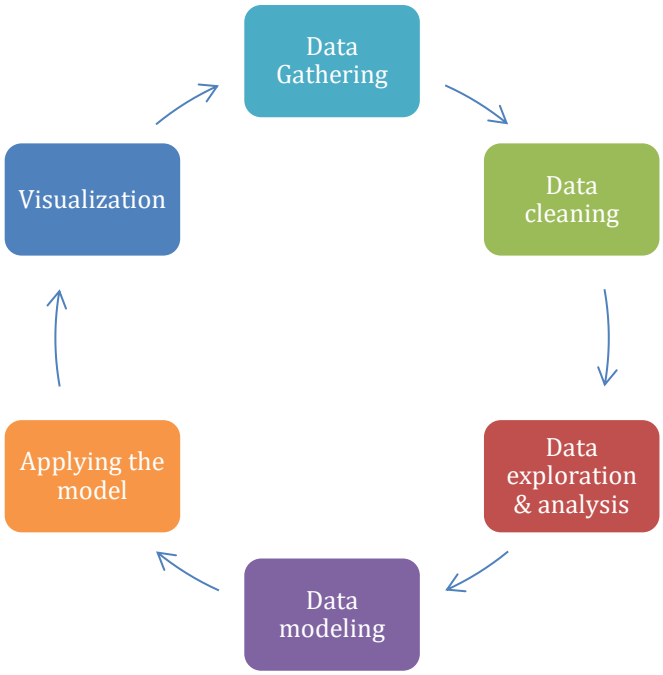

Figure 6: System architecture for restaurants trend analysis

4.3 User interface

Our interface is simple and easy for users that's not expert in websites

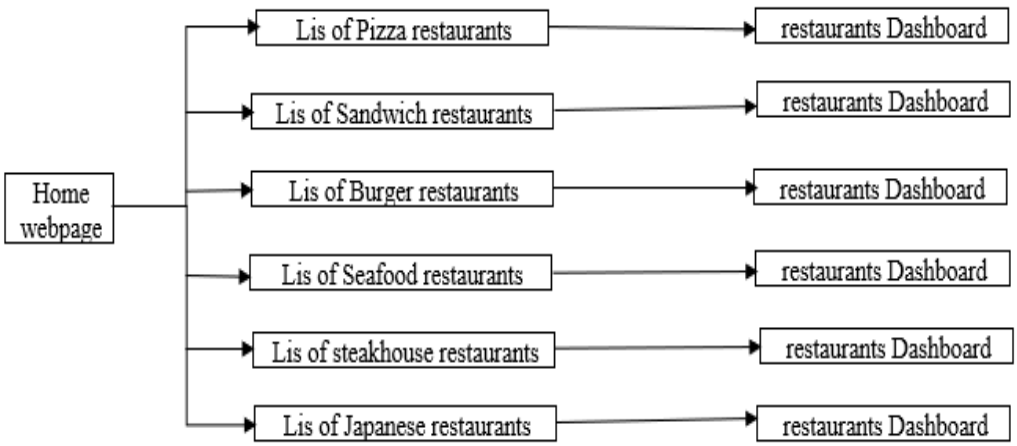

Figure 7: The user interface

As seen in figure 7, that is user interface that structer we us it in out project.

4.4 Development

Our approach to the system is to help customers choose restaurants, and our system works on seven steps, which we will present as follows :

4.4.1 Data Gathering

To collect data we used two main social media sources where people share their opinions and judgments of restaurants in Saudi Arabia. These two sources are twitter and foursquare.

Twitter

We used twitter because it is an application that people Express their opinions on it. And it is very popular in Saudi Arabia. To collect data from Twitter we used Twitter API, which allows us to access features of Twitter without having to go through the website interface. Also, we used other data gathering tools to have access to historical data. We collected more than 140,000 tweets.

Foursquare

We used Foursquare because it is more specific about the field of nutrition. And it is very popular in Saudi Arabia. To collect that data we were not able to extract enough information from the API, for that, we used a different technique, we used web scraping. We collected 30,000 comments.

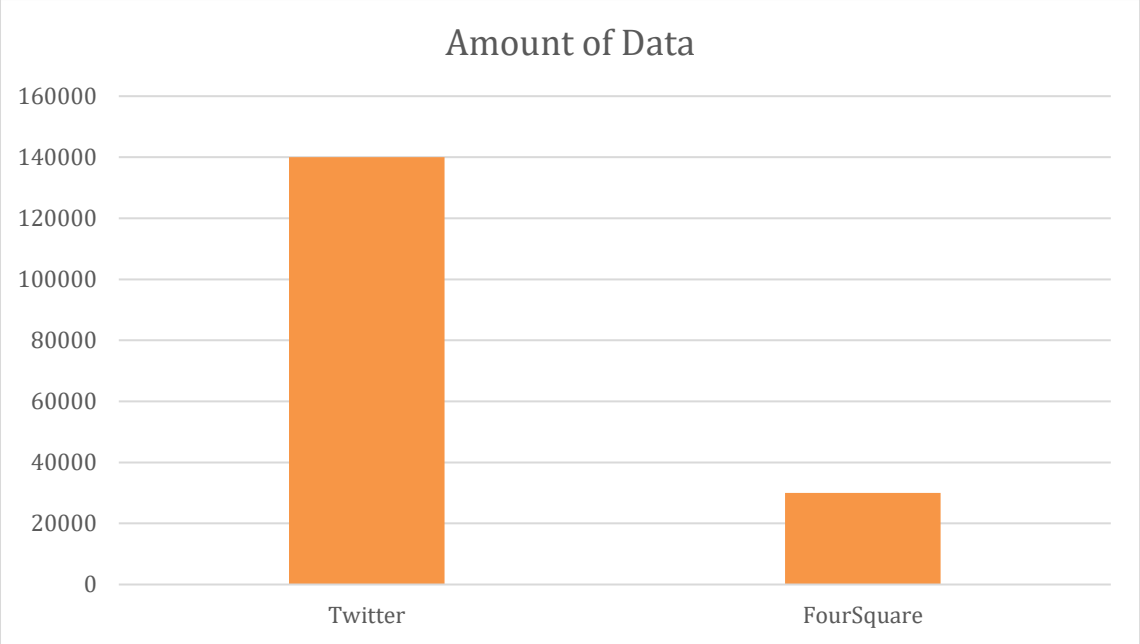

Figure 8: The amount of data for Twitter and Foursquare

4.4.2 Data cleaning

This is impossible to get clean data you most clean it, we remove the null value and there is tweets intersection between the tools we use it and tweet we get it from API resulting from it duplicate tweet we must remove it, some tweet contained multiple restaurants we duplicate the number restaurants what it has. And normalize all Arabic text as following

1- remove repeat letter
2- remove punctuations
3- replace from "[إأآا]" to"ا"
4- replace from “ى" to”ي"
5- replace from "[إأآا]" to"ا"
6- replace from ”ة” to "ه"
7- Convert www.* or https?://* to " "
8- Replace #word with word

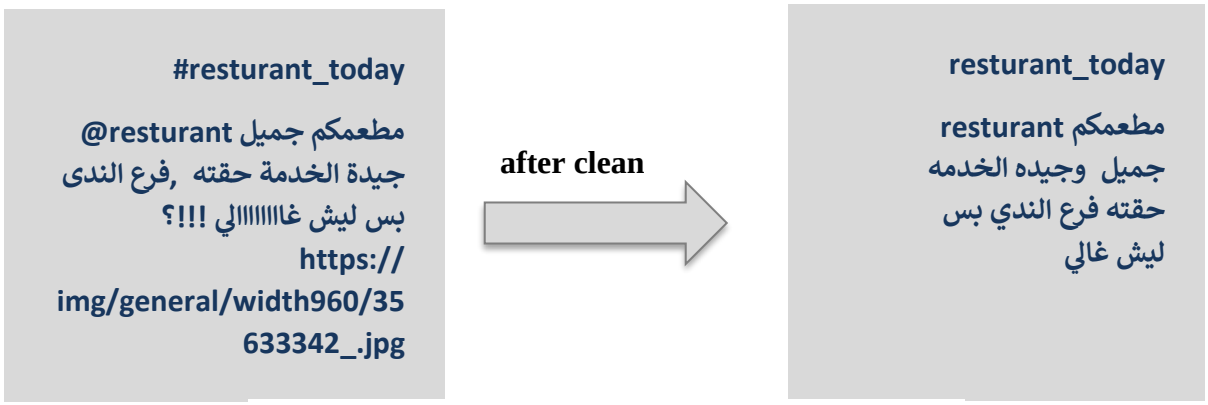

Figure 9: Example of cleaning data

as seen in Figure 7, That our process to standardize all opinions.

4.4.3 Data exploration & analysis

Exploratory data analysis is about exploring and then augmenting your data to maximize the potential of your analyses and visualizations. Visualizing relationships in your data and just building intuition about what you're working with. After that, we added to our data the gender type by collecting Arabic names. this helps us with indicating whether the writer is male or female.

4.4.4 Data modeling

That is an important part of our project, we use many machine learning algorithms to classify the text such as SVM, logistic regression, Tree Decision, and predefined CNN deep learning algorithms. First, we take **1904** textual data, i.e tweets, and foursquare comments, and label it as a training set. each positive opinion has a value of 1 and each negative opinion has a value of 0. after that, we split the data set to training and testing data. Train to train our model and test to validate the model.

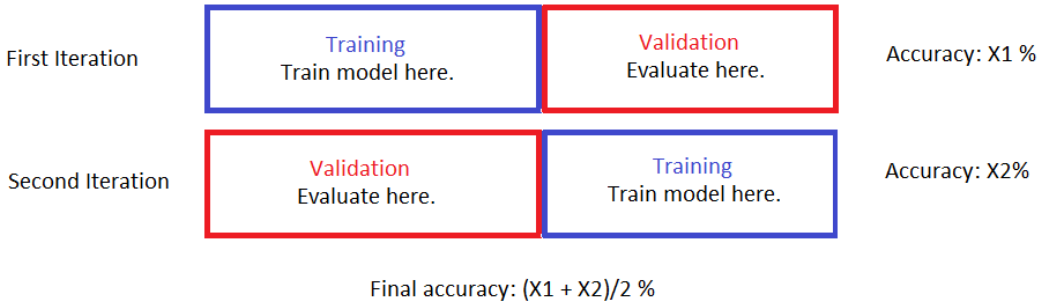

Figure 10: 2 fold cross-validation

As seen in Figure 7, we used two-fold cross-validation to split the data set to 0.5 as training and testing and then input it to the algorithm. And after that take the testing set as the training set and then input to the algorithm, if the accuracy is close the data set is correct.

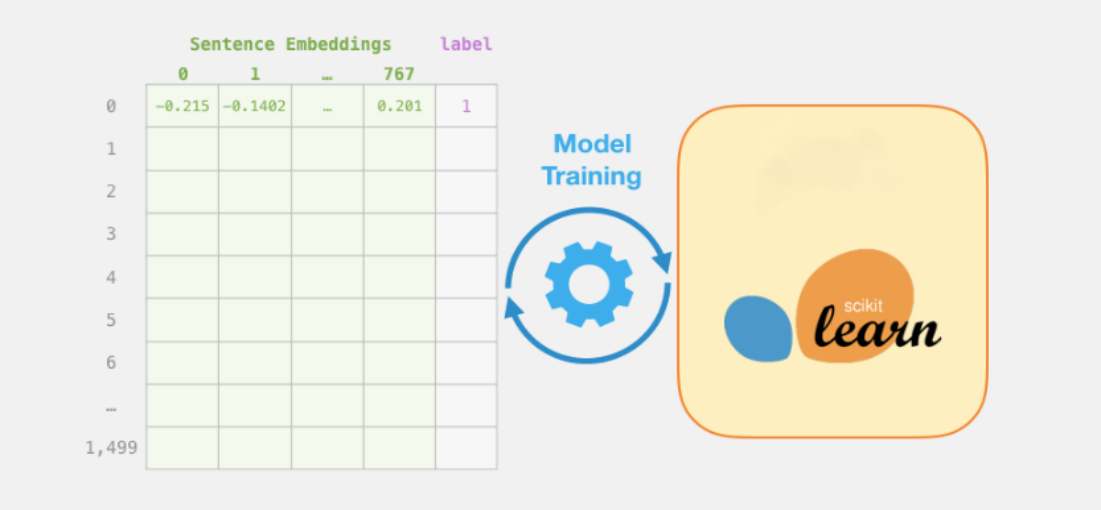

Figure 11: Model training.

As seen in Figure 8, we input our model to the machine learning algorithms to learning from the data set that we label it.

4.4.4.1 SVM

A support vector machine (SVM) is a supervised machine learning model. It uses to classification two-type of data. There is much application of SVM such as classification, regression, outlier detection, and clustering for our project we use SVM type of kernel linear to classify opinions.

4.4.4.2 logistic regression

Logistic regression also called sigmoid function which found by statisticians, is a classification function which used for classifying two types of data by dividing them into 0 and 1. The selection depends on the properties of the data that input to the algorithm if it has more then half it will reference that data to type that belong to.

4.4.4.3 Decision Tree (DT)

Decision Tree (DT) is a supervised machine learning algorithm. The data input from the root and go down into leaves which has the output on another word the root of the tree is the input of the algorithm and the leaves are the output when data input its start split on each parents node until leaves.

4.4.4.4 Accuracy methods used

First, we explain the confusion matrix, It contains information about actual and predicted classifications. And determine the performance of the classifier. The below table shows the confusion matrix. where the TP(True Positive), TN (True Negative), FN(False Negative) and FP(False Positive) and should the numbers of diagonal left more than diagonal right to know the model work Correctly.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Table 5: Confusion Matrix

Accuracy: The percentage of the correct classified objects by used that equation :

$$A = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$$

Equation 1: The accuracy equation.

Precision: Determine the percentage of TP from all predict positive for instance, is calculated as follows:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

Equation 2: The precision equation

Recall: Determine the percentage of TP from all actual positive for instance, is calculated as follows:

$$R = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

Equation 3: The recall equation

F-Measure: Determine the average for both the precision and recall obtained. F- Measure is calculated as follows:

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Equation 4: The F-measure equation

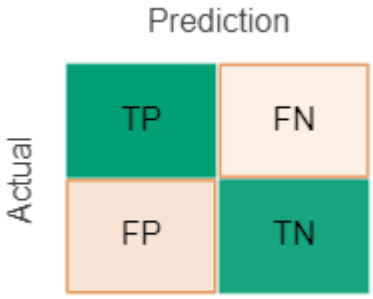

Figure 12: SVM confusion matrix

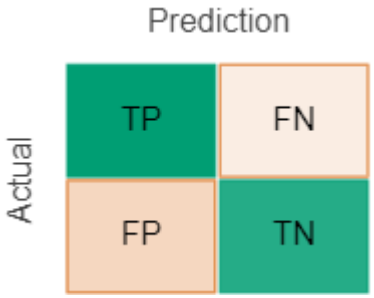

Figure 13: Logistic regression confusion matrix

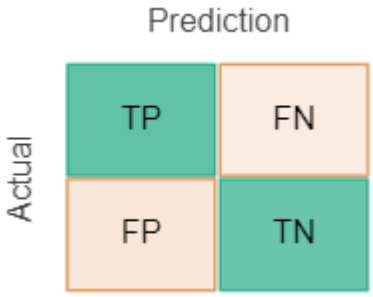

Figure 14: Decision Tree confusion matrix

	SVM	**Logistic regression**	**Decision Tree (DT)**
Precision	0.85	0.80	0.81
Recall	0.91	0.90	0.84
F1	**0.88**	0.84	0.82

Table 6: comparison between algorithms

4.4.4.5 CNN deep learning algorithm

CNN is a deep learning algorithm. Its work by taking the input goes around it much time and extract the properties every time after then gives the output of data. We use word2vec with CNN to do text classification. word2vec is working by transfer text to the neural network that CNN can work with.

	SVM	CNN
Accurcy	**0.88**	0.83

Table 7: comparison between algorithms

4.4.5 Applying the model

Based on the previous results we choose to use SVM. We extract the features and saved them as a pickle file. And after that, we loaded the pickle file to TfidfVectorizer vocabulary and fit data. And then predict all the data that we have.

4.4.6 Visualization

We used Tableau to construct all of the visualizations. We have three charts on each restaurant page. The first chart is the bar chart to represent. The most frequent phrases used, the number of opinions per day, the second chart is the line chart to represent the number of spread opinions per month, the third chart is a donut chart to represent the number of males and females per restaurant. That as the show you follow bellow.

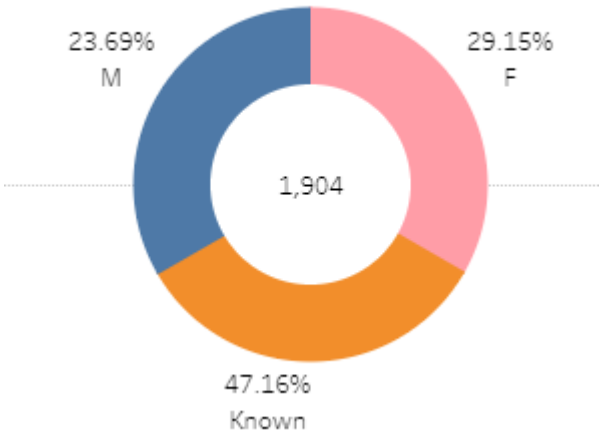

Figure 15: Donut chart about determine gender

4.5 summery

In this chapter, we explained the system Architecture and user interface design. We explained about each classifier we use.

CHAPTER 5: Implementation

5.1 Introduction

In this chapter, we will mention a description of the implementation, first of all, we will mention hardware and software tool that has helped us to build and operate the System, and after that, we will describe the implementation tool. We will include some of the implementation code with details.

5.2 Software and hardware requirement

a) Software Requirement

The following table shows the computer software requirements to run the system:

Operating system	Any
Programming language	Python3
Platform	Jupyter notebook , Tableau Visual Studio Code , No sql Booster for MongoDB, firebase server

b) Hardware Requirement

The following table shows the computer hardware requirements to run the system:

CPU	Intel Core i5 3.10GHz
Screen Resolution	1280 X 960
Network	Ethernet Network Driver
Hard Disk	250 GB
server	EC2 type of service, Size: 8 GB

5.3 Implementation tools

a) Jupyter notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code. Uses include transformation, data cleaning, machine learning, and statistical modeling. Ending with the ".ipynb" extension.

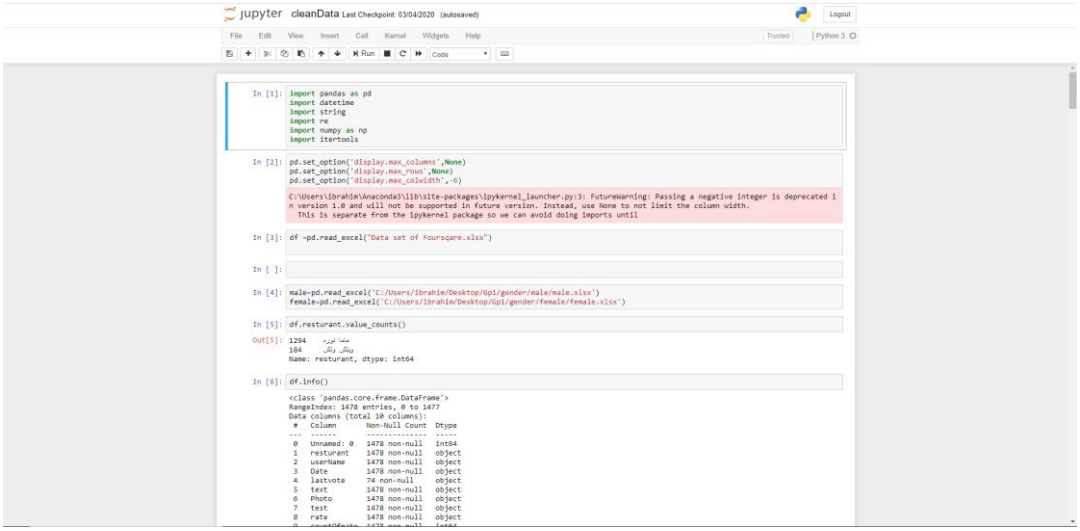

Figure 16: Jupyter notebook interface

b) Tableau

Tableau is an analytics platform and business intelligence (BI) created to help people see, understand and make decisions with data. Tableau Software creat it to allowing a broad population of users to interact with their data, ask questions, solve problems and create value.

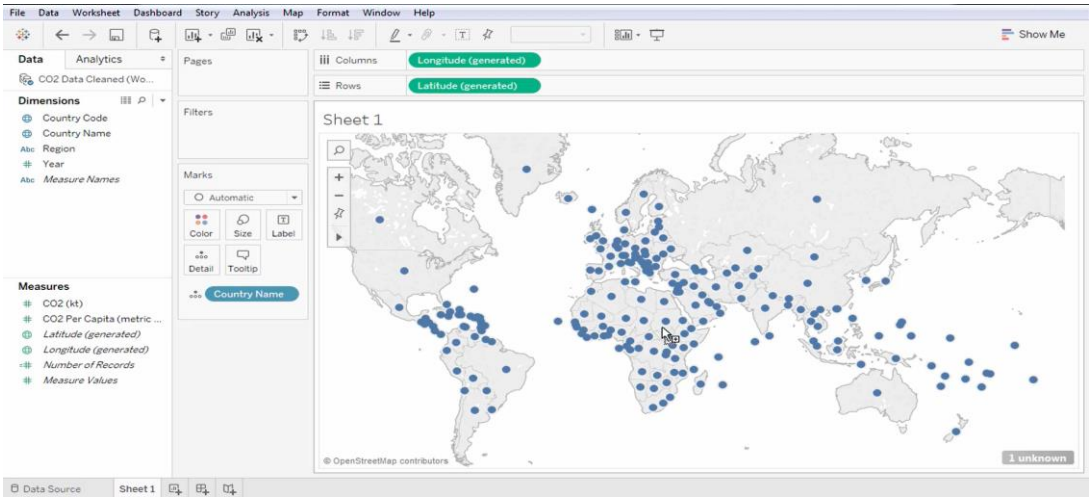

Figure 17: Tableau interface

c) Amazon EC2 server

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides security. It is designed to make web-scale cloud computing Simpler for developers.

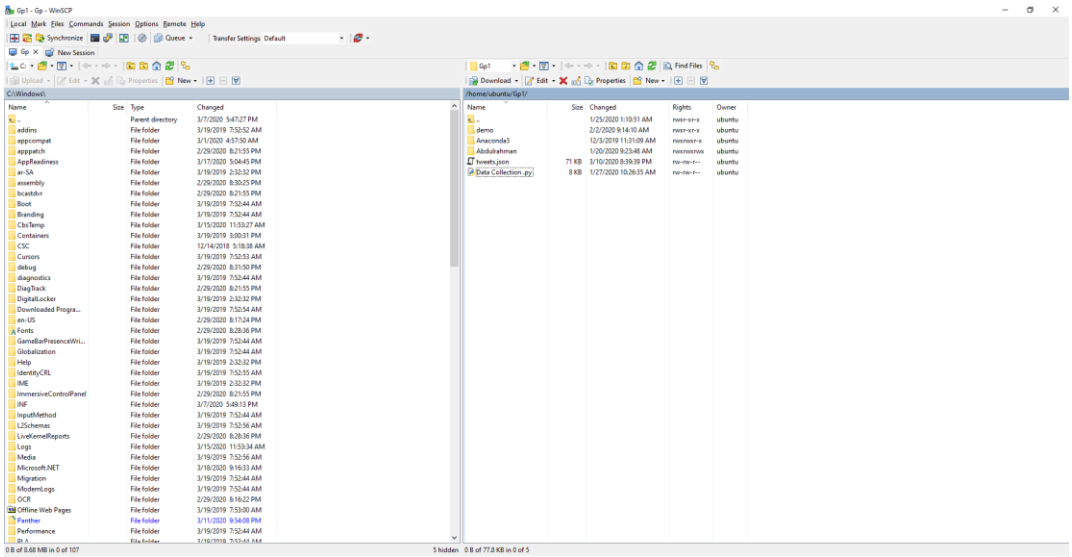

Figure 18: server interface

d) Visual Studio Code:

Is a sourc-code editor that helps the developer to coding easily and published by Microsoft

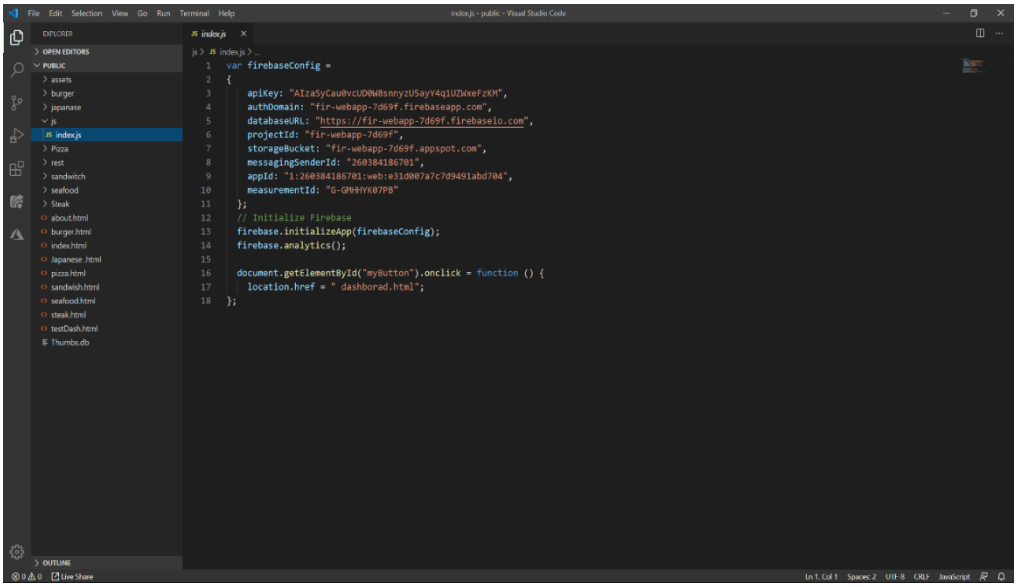

Figure 19: Visual Studio Code

e) Database

We use MongoDB to save data, MongoDB is NoSql database which means its store data as JSON, we use NoSQL Booster for MongoDB and it's act like an interface for MongoDB that makes you interact with MongoDB easier, that show following :

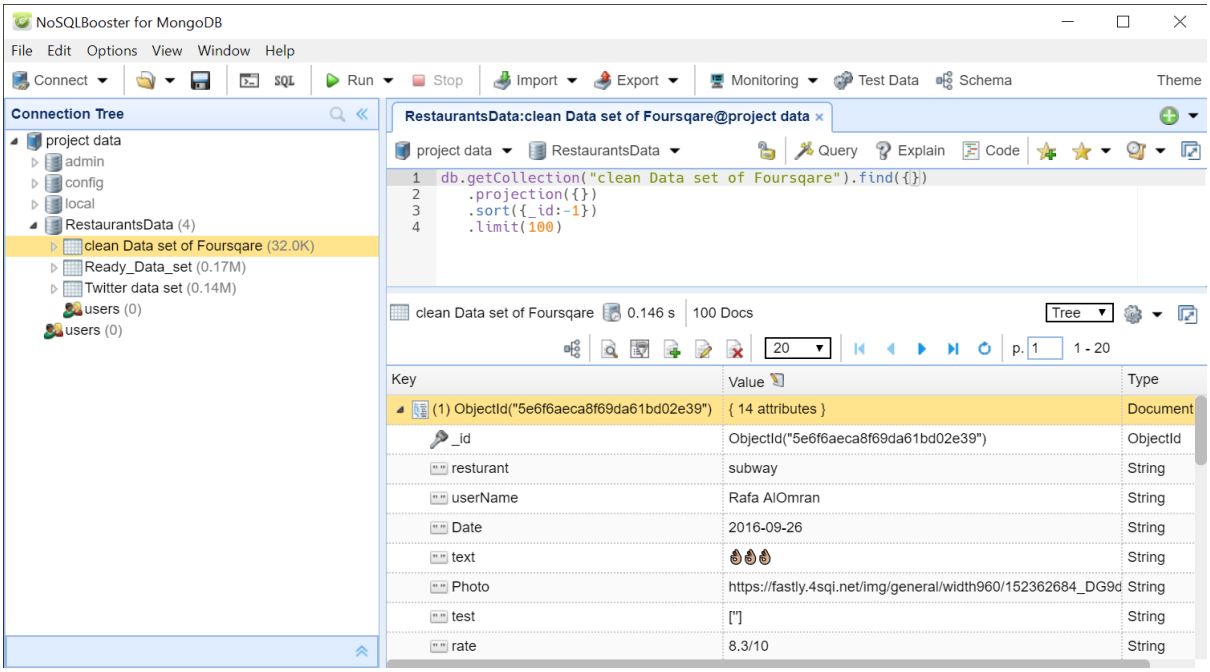

Figure 20: No sql Booster for MongoDB

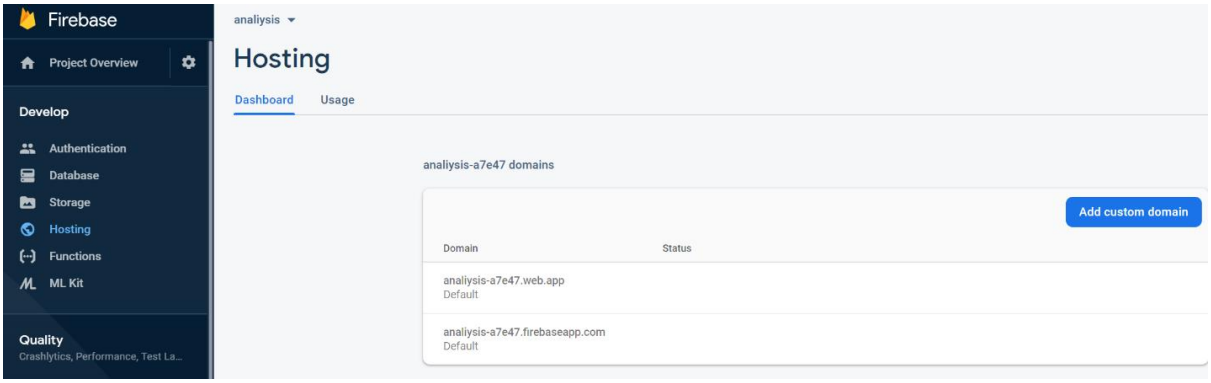

Figure 21: Firebase hosting

5.4 Implementation details

```
import requests
from bs4 import BeautifulSoup
import re
import time
import pandas as pd
```

Figure 22: Set of libraries that we used.

As seen in figure 20, we will explain the most important of them BeautifulSoup we use it for web scraping and pandas for data analysis and manipulation tool.

```
In [11]: # Collect and parse first page

count=0
for url in restlist:
    name=url[0] #Get name of resturant

    time.sleep(7*57)
    for num in range(1,100):

        if(num == 1):
            response = requests.get(url[1]) #Get url of resturant
            print(str(num)+"-"+str(response.status_code))
            if(response.status_code !=200):
                break
        else:
            time.sleep(57*4)
            response = requests.get(url[1].replace("?","?tipsPage="+str(num)+"&")) # Move to next page of comments
            print(str(num)+"-"+str(response.status_code))
            if(response.status_code !=200):
                break

        soup = BeautifulSoup(response.content, 'html.parser')



        print(soup)
        comment = soup.find(class_='tipsSection') #Get element with class="comment"

        rate = soup.find(class_='venueScore positive') #Get element with class="venueScore positive"
        rate=rate.text # extract the rate
        countofrate = soup.find(class_='numRatingsBlock') #Get element with class="numRatingsBlock"

        numberofrate=re.findall(r'\d+',countofrate.text )
        countRate = ''.join(filter(lambda i: i.isdigit(), numberofrate)) #Extract the count of rate

        for item in comment.find_all(class_="tip tipWithLogging useTipUpvotes "):

            restruant.append(name) #Name the resturant
            userName.append(item.find(class_="userName").text)# Extract the name that wrote the comment
            Date.append(item.find(class_="tipDate").text) # Extract the Date time
            text.append(item.find(class_="tipText").text) # Extract the text
            rates.append(rate)
            countsOfrate.append(countRate)
            testIncomment=[]
            for item2 in item.find_all(class_="entity tip_taste_match"):#Extract the important words
                testIncomment.append(item2.text)
            test.append(testIncomment)
            try:
                image= item.find(class_="tipPhoto").get('src')
                image=image.replace('558x200',"width960") # That two show the compelete image (if there)
                tipPhoto.append(image)

            except:
                tipPhoto.append("None")
            try:

                lastvote.append(item.find(class_="lastVoteTime").text) # Extract the last vote of comments (if there)
            except:
                lastvote.append("None")
```

Figure 23: wep scraping code

As seen in figure 21, to extract the comments of all pages of the restaurant.

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
import json
import pandas as pd
```

Figure 24: Set of libraries that we used.

```
l = FileWriteListener()
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

stream = Stream(auth, l)
```

```
stream.filter(track=[u'keyword'])
```

Figure 25: Gather data from twitter

As seen in figures 22 and 23, this handles twitter authentification and the connection to twitter streaming API and line filter twitter streams to capture data by keywords.

```
import pandas as pd
import numpy as np
import re
import string
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score, cross_val_predict, KFold
from sklearn.metrics  import confusion_matrix, classification_report
from sklearn import metrics
from sklearn import tree
import pickle
```

Figure 26: Set of libraries that we used.

```
X_train, X_test, y_train, y_test = train_test_split(X, target, test_size=0.5, random_state=42)
```

```
# create the classifer and fit the training data and lables
classifier_svm = svm.SVC(kernel='linear', C=1)
classifier_svm.fit(X_train,y_train)
print("SVM accuracy: %.2f"%classifier_svm.score(X_test, y_test))

#do a 2 fold cross-validation
results_svm = cross_val_score(classifier_svm, X,target, cv=2)
print("\n2-fold cross-validation:")
print(results_svm)

print("The average accuracy of the SVM classifier is : %.2f" % np.mean(results_svm))

print("\nConfusion matrix of the SVM classifier:")
predicted_svm = classifier_svm.predict(X_test)
print(confusion_matrix(y_test,predicted_svm))


print("\nClassification_report of SVM classifier:")
print(classification_report(y_test,predicted_svm))
print("--------------------------------------------------------------------------------")
```

Figure 27: Machine learning code

As seen in Figures 24 and 25, implement of machine learning code and we use two-fold cross-validation.

```
|: pickle.dump(tf_vec.vocabulary_, open("feature.pickel","wb"))

|: transformer = TfidfVectorizer(vocabulary=pickle.load(open("feature.pickel","rb")))
```

Figure 28: Extract feature of the training set

As seen in Figures 26, extract the features of the training set from TF IDF to apply through the machine learning algorithm on all remaining data set.

5.4.1 Website

We use Predefined template from OS TEMPLATES [1] then make a change on it to make it fit with our project website

[1] www.os-templates.com - date of review 07-02-2020

```
<!-- W3S -->
<link rel="stylesheet" href="https://www.w3schools.com/w3css/4/w3.css">

<!-- CSS Files -->
<link rel="stylesheet" href="assets/css/animate-3.7.0.css">
<link rel="stylesheet" href="assets/css/font-awesome-4.7.0.min.css">
<link rel="stylesheet" href="assets/css/bootstrap-4.1.3.min.css">
<link rel="stylesheet" href="assets/css/owl-carousel.min.css">
<link rel="stylesheet" href="assets/css/jquery.datetimepicker.min.css">
<link rel="stylesheet" href="assets/css/style.css">
```

Figure 29: CSS Files call

```
var firebaseConfig =
{
    apiKey: "AIzaSyCau0vcUD0W8snnyzU5ayY4q1UZWxeFzKM",
    authDomain: "fir-webapp-7d69f.firebaseapp.com",
    databaseURL: "https://fir-webapp-7d69f.firebaseio.com
    projectId: "fir-webapp-7d69f",
    storageBucket: "fir-webapp-7d69f.appspot.com",
    messagingSenderId: "260384186701",
    appId: "1:260384186701:web:e31d007a7c7d9491abd704",
    measurementId: "G-GMHHYK07PB"
  };
  // Initialize Firebase
  firebase.initializeApp(firebaseConfig);
  firebase.analytics();
```

Figure 30: Connect with the firebase server

As seen in Figures 27 and 28, we link the HTML file with the external style sheet and connect with the firebase server.

5.5 summery

In this chapter, we explained the hardware and software requirements and tool, then we specified the implementation tool. We give concise information about every tool after we mention in detail every implementation tool with their code.

CHAPTER 6: Testing

Introduction

in this chapter we will test what we build, first, we will test the home page than about page, category page, and dashboard page.

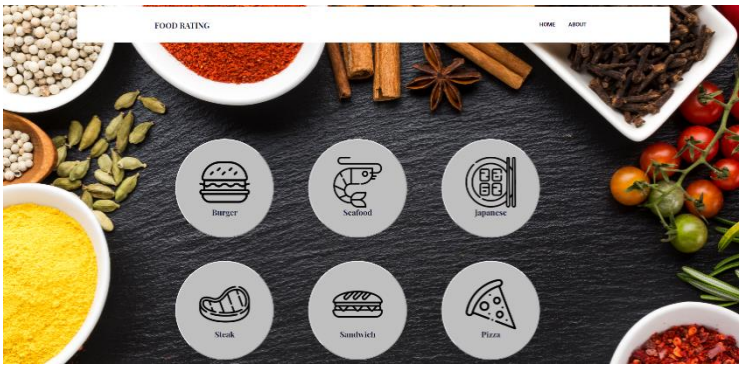

Figure 31: Home page

As seen in Figures 29, On the home page we show the categories can a user select one of them or user can go to about.

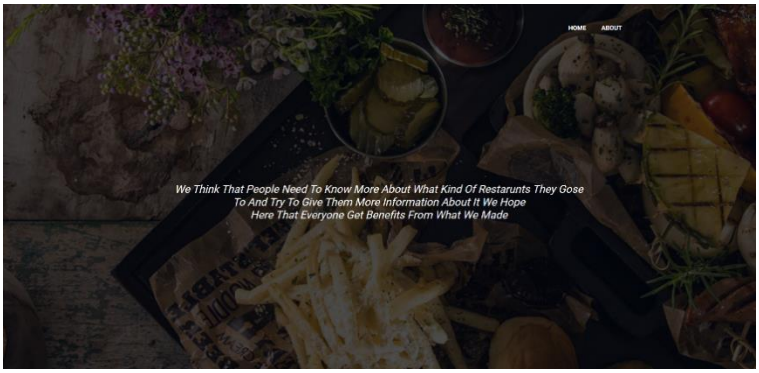

Figure 32: About page

As seen in Figures 30, In about page, we tell the user what our website presents to him.

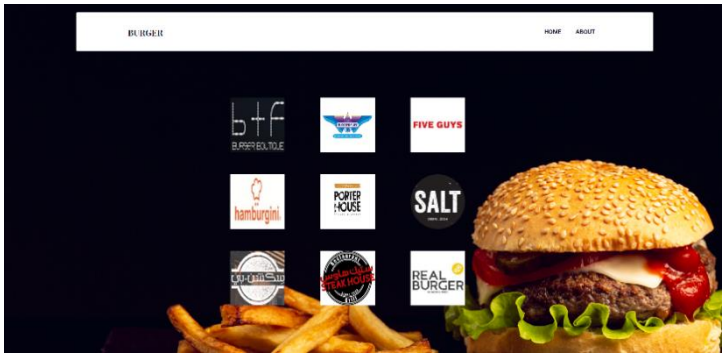

Figure 33: category page

As seen in Figures 31, in category page we show to user restaurants he can select.

Figure 34: Dashboard page

In the dashboard page, we show to user dashboard contain the analysis of the selected restaurant.

1-Rate of restaurant

2- Most words said about the restaurant.

3- How people fell about the restaurant.

4-Crowded in the week.

5- Gender of people visit the restaurant.

Chapter 7: Conclusion and future work

7.1 Conclusion

We build a website in our project to help people to select a restaurant and see more information about the restaurant they interested in it**.**

7.2 Future work

We think in the future to do more analysis and add more information to a user and more restaurants.

References

[1] Sana Alowaidi, Mustafa Saleh, Osama Abulnaja, Semantic Sentiment Analysis of Arabic Texts, Vol. 8, No. 2, 2017.

[2] Malmaz Roshanaei and Shivakant Mishra, An Analysis of Positivity and Negativity Attributes of Users on Twitter, ASONAM 2014.

[3] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, Sentiment Analysis of Tweets using Machine Learning Approach, Vol. 8, 2018 IEEE.

[4] Snehal Kale, Vijaya Padmadas, Sentiment Analysis of Tweets Using Semantic Analysis, 2017 IEEE

[5] melanie Bradley , Stay on the Cutting-Edge, 2019

[6] Aseel Addawood and Amirah Alshamrani and Amal Alqahtani and Jana Diesner and David Broniatowski, Women's Driving in Saudi Arabia – Analyzing the Discussion of a Controversial Topic on Twitter,2018.

[7] M. Hammad and M. Al-awadi, "Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning," in Information Technology: New Generations, ed: Springer, 2016, pp. 131-139.

[8] Ankita Gupta1, Jyotika Pruthi2, Neha Sahu3 Ankita Gupta et al, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 444-458.

[19] R. M. Duwairi, "Sentiment analysis for dialectical Arabic," in Information and Communication Systems (ICICS), 2015 6th International Conference on, 2015, pp. 166-170.

[10] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in The Semantic Web–ISWC 2012, ed: Springer, 2012, pp. 508-524.

[11] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in Contemporary Computing (IC3), 2014 Seventh International Conference on, 2014, pp. 437-442.

[12] E. Cambria, et al., ―SenticNet 4: A Semantic Resource for Sentiment Analysis based on Conceptual Primitives‖, COLING, pp. 2666-2677, 2016.

[13] K. Dashtipour, et al., ―Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques Cognitive Computation‖, vol. 8, no. 4, pp. 757-771, 2016.

[14] B. Pang, et al., ―Thumbs up? Sentiment Classification using Machine Learning Techniques‖, Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79-86, 2002.

[15] Al-Twairesh, et al., ―Subjectivity and Sentiment Analysis of Arabic: Trends and Challenges‖, 11th International Conference on Computer Systems and Applications (AICCSA), IEEE/ACS, pp. 148-155, 2014.