

**Department of Computer Engineering**

**Graduation Project**

**Project Title will be Written Here**

**Student-Name Surname**

**Student Number**

**Advisor:**

**May 2020**

#### Document Information

|  |  |
| --- | --- |
| **File Name** |  |
| **Total Page Number** | 35 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Authors** | **Comment** | **Date** |
| First Draft |  |  |  |
| Last Draft |  |  |  |
| Version 1 |  |  |  |
| Version 2 |  |  |  |

**Publication Approval**

**Name-Surname**

Thesis Supervisor

**Name-Surname**

Thesis Coordinator

**Name-Surname**

Thesis Coordinator Assistant

#### PREPARATION AND REVIEW APPROVAL

**Name/Surname St 1 Name/Surname St 2**

**CONTENTS**

1. [Introduction 1](#_TOC_250013)
   1. [Purpose of the Project 1](#_TOC_250012)
   2. [Project Scope, Outcomes and Schedule 1](#_TOC_250011)
   3. [Objectives and Success Criteria of the Project 3](#_TOC_250010)
   4. [Project Members and Their Responsibilities and Tasks 3](#_TOC_250009)
2. [Related Work 4](#_TOC_250008)
   1. [Beautiful Soup 6](#_TOC_250006)
   2. [Neural Network 6](#_TOC_250006)
   3. [Bayes Naive 6](#_TOC_250006)
   4. [Dictionary 6](#_TOC_250006)
3. Proposed System 7
   1. Overview 7
   2. Analysis and Visualization Tools 8
   3. The Proposed System Architecture 8
   4. Functional Requirements 8
   5. Nonfunctional Requirements 9
      1. Quality Requirements 9
      2. Performance Requirements 10
   6. Pseudo Requirements 10
   7. System Models 11
      1. Scenarios 11
      2. Use Case Model 11
      3. Static Models 12
      4. Dynamic Models 15
   8. User Interface 19
4. Conclusion 29
5. References 30
6. Glossary 31
   1. Definitions 31
   2. Abbreviations 31

# Introduction

This project examines to explain the basics of collecting, storing, and analyzing the sentimental content of the given brand. The main issues in this project are in the area of, Artificial Intelligence, Machine Learning, Deep Learning and Sentiment Analysis.

Main goal of this project is to develop and provide an efficient tool for finding sentimental information from the collected data for the end users.

The first half of this project will study collection and storage of sentimental data from the websites. It will start by explaining how to collect sentimental data. The only option that possible for this problem is scraping the sentimental data by using required libraries. The library deemed appropriate for this job is BeautifulSoup. Necessary information about this library will be explained in the next sections. After that, it goes to understand to how to store this data for later use.

The second half of this project is focused on sentimental analysis of the collected and stored sentimental data. Here, it was focused on common measures and algorithms that are used to extract the sentimental information from the collected data. The measures and algorithms are determined as Neural Network, Naïve Bayes, Dictionary and Word Cloud. Necessary explanations about these methods also will be made in next sections.

And the Project ends with the showing sentiment analysis by visual graphs. With these graphs, end users will be able to understand the sentiments of the collected data through intense visualizations.

## Purpose of the Project

The main purpose of this project is to give a brief sentimental information to end users about society’s thoughts for a consumable, service or brand.

Customer satisfaction is pretty important factor for a brands’ selling. If consumers have positive thoughts about for a product, service or brand, this refers to the image of that brand is quite good. And if the brand’s public opinion is good, that means this company can achieve great successes and they can profit more.

Nowadays, people can collectively share their thoughts and ideas. And web is the key source of information for that kind of sentimental data. Any opinion or idea can reach millions of people in a very short time. If these institutions want to profit more and care about what society thinks about their service, product or image, they should also consider to collect and analyze the sentimental data.

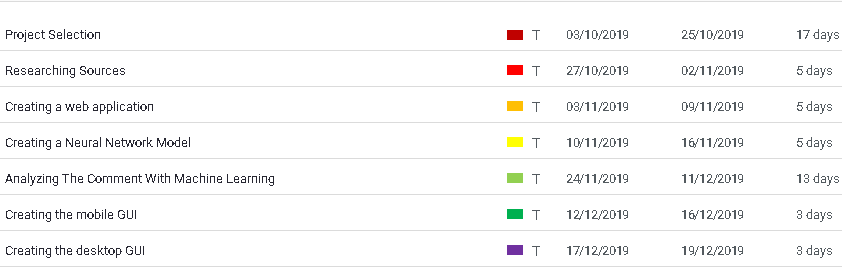
There are some systems in business world for analyzing thoughts on specific company or product for such problem that mentioned above. But in this project, it was aimed to provide a system that collects data from all kinds of websites and it was designed a program that analyzes this data using different sentiment analyze methods.

## Project Scope, Outcomes and Schedule

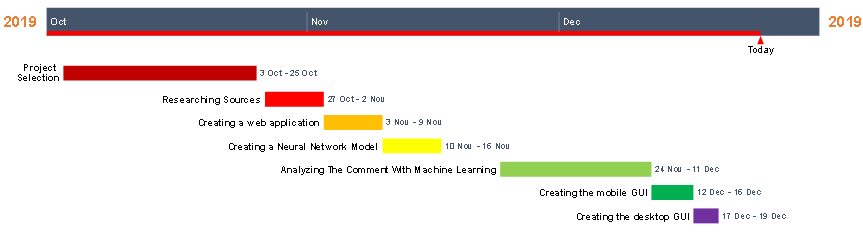
The scope of the project is building software which consists of a web scraper that collects sentimental information from the websites and building an analyzer that outputs meaningful sentimental graphs from extracted data to display important graphs for users.

This project is going to produce a software program for end users and a report for a side component. The software program is going to be a scraper which has the capability of crawling recent comments from all over the world with a given brand name. Then, it will be able to analyze data and show the results with informative visualizations.

The project started on 25.10.2019 with the announce of the project titles. In the previous version of this project report, it was focused on research. And in the second part of the project, it was focused on implementing the project.



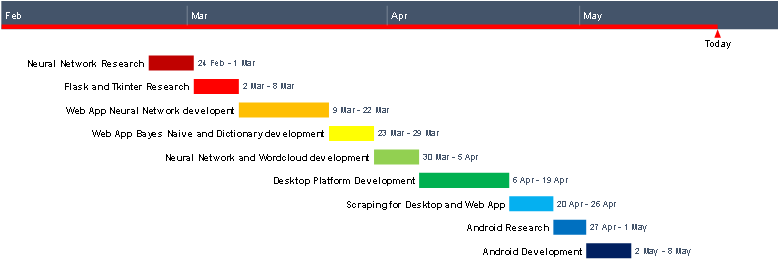
**Figure 1.2. 1-**Timeline of Project Part I.



**Figure 1.2.2-** Timeline of Project Part I.



**Figure 1.2. 3-** Timeline of Project Part II



**Figure 1.2.4-** Timeline of Project Part II.

The timelines for the development of the project are mentioned above. (**Figure 1.2.1), (Figure 1.2.2), (Figure 1.2.3), (Figure 1.2.4).** The first two figures show completed jobs in the previous part. The other two graphs show the things that done.

Both Windows and Linux operating system are used in this project. The main programming language is determined as Python because of its useful libraries that created for machine learning. Flask framework was used as a web application development tool, Tkinter used for desktop GUI. And Android Studio is used for developing mobile application. Analysis and visualizations are implemented with Matplotlib library.

Nowadays, many social networks do not easily provide their API keys. And many other API keys have restrictions. Therefore, the required sentimental data were obtained using python libraries. BeautifoulSoup library is used for extracting sentimental information from the websites. The main drawback is, this library can only extract static content from the websites. It cannot extract dynamically loaded content.

## Objectives and Success Criteria of the Project

The objectives and success criteria of the project are to:

Provide a web and desktop application which capable to:

* + - Extract data from the social network.
    - Make the sentimental analysis of selected data.
    - Show sentimental graphs to the end user.

Provide a mobile application which capable to:

* + - Make the sentimental analysis of selected data.
    - Show sentimental graphs to the end user.

## Project Members and Their Responsibilities and Tasks

**Table 1.4.1** Project members and their responsibilities and tasks

|  |  |
| --- | --- |
| **Project Member** | **Responsibilities and Tasks** |
| Doğacan Biçer | Research, gathering essential data for sentiment analysis, Analyzing the gathered data, visualizing the results, creating a web, mobile and desktop application. |

# Related Work

To complete the project, two things must be accomplished. One is scraping the sentimental data. Other is analyzing the crawled sentimental data. Two approaches were considered to collect the sentimental data. First method is to use the API’s provided by social media sites. Application Programming Interface also known as “API” refers to principles which social systems permitting other applications and websites to take their social media information and work coordinated with the user’s websites or applications. Since social network is itself giving these principles, so users of the APIs are able clearly expect that Application Programming Interface is a fair manikin. They are giving restricted attributes, Call or processes that could be performed by API user and they require software developer to produce an API Key before utilizing the Application Programming Interface.[1]

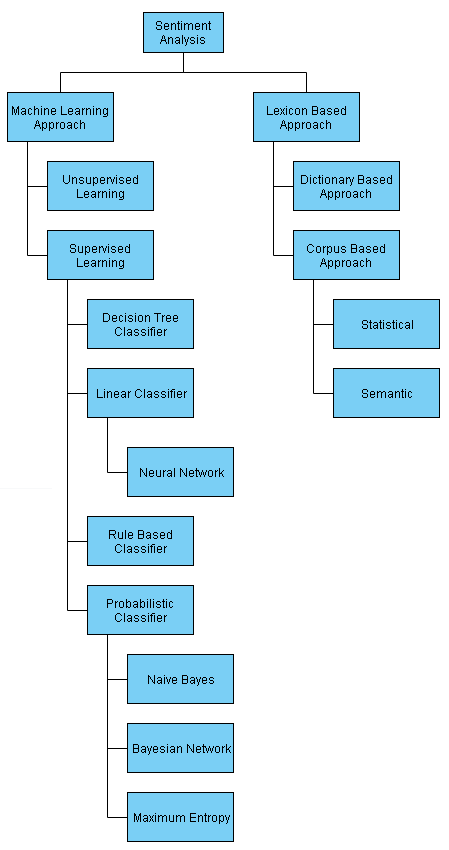
Second method was using scraping libraries. Web scraping, moreover known as web information extraction, is the method of recovering or scraping the contents from an online source. Not at all like the ordinary and wearisome handle of physically extricating information, scraping employments shrewdly robotization to capture millions, or billions of information from the worldwide web’s apparently perpetual wilderness [2]. This option is used in this project because of restrictions of API’s and scraping library is chosen as BeautifulSoup.

Other thing that must be accomplished is Analyzing the collected data. Basically, there are two main methods when analyzing the sentiment. One approach is Machine Learning approach. Other one is Lexicon- based approach. These also have subbranches.

Machine Learning Approach utilizes machine-learning strategies and differing attributes to build a classifier that can recognize content. These days, deep-learning strategies are well known since they fit on information learning representations like Recurrent Neural Networks.[3]

Lexicon-Based strategy employments an assortment of words explained by extremity score, to choose the common appraisal score of a given substance. The most useful resource of this procedure is that it does not require any preparing information, whereas its drawback is that an expansive number of words and expressions are not included in assumption vocabularies.[4]

Simply, sentiment analysis can be expressed with the graph below. **(Figure 2.1)**



**Figure 2.1-** Types of Sentiment Analysis.

* 1. **Beautiful Soup**

BeautifulSoup is Python package which is programmed to parse information, i.e., to extricate information from HTML or XML files.

Since BeautifulSoup can is used for parsing the information and can’t get the html and xml files themselves, it is regularly utilized with the library named Request. In circumstances like these, Request library is going to make the HTTP to ask the site to get the internet page, and once the page has been returned, BeautifulSoup package can be utilized to parse the target information from the HTML page.

One of the enormous focal points of utilizing BeautifulSoup is its effortlessness and capacity to computerize a few of the repeating parts of parsing information while scraping. While using small number of codes, the user will be able to design BeautifulSoup to explore a whole parsed archive and discover all occasions of the information that they need or naturally identify encodings such as extraordinary characters. [5]

* 1. **Recurrent Neural Network**

A Recurrent Neural Network also known as RNN is a type of neural network where associations between lstm nodes shape a coordinated chart along a worldly grouping [6].

Recurrent Neural Networks are named as recurrent because they are able to do the same assignment for each component of a arrangement, with the result being related to the past results. In Another way, we can assume that Recurrent Neural Networks have a short memory which captures data approximately what has been calculated yet. In hypothesis, Recurrent Neural Networks can make utilize of data in self-assertively long arrangements, but in reality, they are not able to look back many steps. [7]

In feedforward running structures, incoming information is processed only forward. In this kind of structures, the input data is passed through the network to obtain an output value. The resulting output value is compared with the correct values to obtain an error. The weight values on the network are changed depending on the error, with this way created model can produce the most accurate result.

When training a feedforward network, the error must be reduced sufficiently. Thus, the weights to the neurons are renewed and a structure is formed to give the appropriate output to the input.

Recurrent structures are separated from feedforward structures because they use their outputs as input in the next process. In other words, the inputs in these networks produce output by combining current and previous information.

* 1. **Naïve Bayes Classifier**

Naive Bayes Classifiers is pretty useful when solving text classification problems. The Text classifications aims is to classify documents into one or many classes. Some applications of usage of Naïve Bayes are Recommendation Systems, Spam Filtering, Multiclass or Real time Prediction.

Another case of Naïve Bayes classifier is Sentiment Analysis, whose objective is to decide the writer’s point of see approximately a specific subject, item, benefit, etc. [8]

* 1. **Dictionary Based Method**

Dictionary Based method works like this: First of all, positive and negative words are collected and stored. Then each word in collected data compared with this bag of words. If collected data has more positive words than negative words, this means that this comment is positive. Likewise, If comment has more negative word than positive words, This comment become negative.

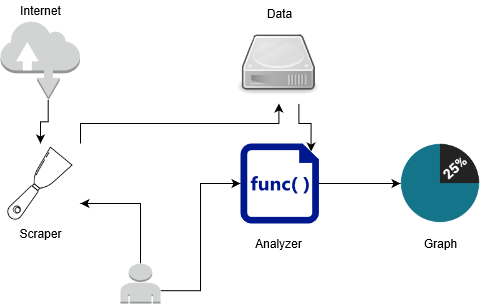
When using this method, the negative suffix is ​​also taken into consideration. For example, “don’t like” considered as negative word even belong to the positive words. Because previous word is negative.

# PROPOSED SYSTEM

## Overview

For web application and desktop platforms, the system has 4 main processes. The processes are Scraping, Storing, Analyzing and Visualizing the collected data. A scraper runs in the background to download web posts. After downloading the data, user can insert this data sheet to the analyzers. Finally, the analyzer applies various of techniques to the data and displays the resulting images to the user. The system architecture is shown at below.

For mobile application part, user can only analyze the data with dictionary method. The system architecture for this platform is shown at below.



**Figure 3.1-** The proposed system architecture

The first step is the gathering required data for analyzers. The required data will be the posts that published in the websites.

Second step is the storing the data. The user has one option, he/she can store it in a local drive.

The final step is the analyzing the data. For analyzing the data, Naïve-Bayes, Neural Network, dictionary and Word Cloud techniques will be used. And finally, on the desktop and web platforms, the result will be shown to the user with the help of matplotlib and on the mobile platform result will be shown with the MPAndroidChart graphical library.

* 1. **Analysis and Visualization Tools**

Flask framework is used when developing the web application. Flask is a micro WSGI web framework written in Python. Flask is developed to make starting process rapid and clear. With using Flask framework, the users also can create complicated programs [9].

On the desktop side, Tkinter was used as the graphical interface. Tkinter is a graphical user interface written in Python programming language. Tkinter can be used in Unix systems, Windows or Linux [10]. The main purpose of using Tkinter in this project is because of its cross-platform feature. It runs both windows and Linux operating systems.

On the mobile side, java will be used as a programming language and android studio will be used as a development tool. It is planned that all platforms will share the same database. With this way, A data sheet which uploaded from the web application to the database can be analyzed on the mobile platform. In web and desktop application, python’s matplotlib library will be used to display graphics to the user. And MPAndroidChart library determined as a graph view library on the mobile side.

This app will serve institutions that want to learn what the community thinks of their brand. Although the main purpose of the program is to show the brands' images in the community, it can also provide sentimental information about the products, people and events according to the entered keyword.

* 1. **The Proposed System Architecture**

For the web application and Desktop platforms, In the Scrap part, users can scrap the text in a website. This part uses BeautifulSoup package. In the website link part, Users will enter the link of the website. Then they will enter the tag which has the interested text. And they can write the class of the tag If class exists. If they write nothing to tag and class part, the program assume that they write “p” in the tag part and there is no class of this tag. In the backend, BeautifulSoup and Request Libraries downloads the texts.

But this section does not scrap the dynamically loaded content. Output is going to appear as csv file. And program automatically remove punctuations, numbers and lowers the words.

Users also can create a neural network model. When they select a csv and entered the neural network model information, Keras library creates a neural network model. It uses TensorFlow library in the backend. The csv must have texts and their sentiments like “positive”, “negative” or “neutral”. And when the model is finished, User can download or save the model. Model is going to save as h5 file. Now, Users can use that model anytime.

In the Analyzing part, User can select one sentiment analyzing method among the others. If they want to analyze the data using Neural Network, they must have a model. The neural network model can be created in “Create Model” section. Once the model is selected, now they have to select a csv which has the comments or texts. The csv must consist the texts. And program analyzes the csv and shows the resulting graph to the user.

Users also may want to analyze the texts using Naïve Bayes Approach, for this method, there is no need a model File. But users need a csv file that contains comments and their sentimental information like “positive” or “negative”. The method is going to use this file as training file. And they need another csv that consist text. And program analyzes the csv and shows the resulting graph to the user.

There is also another method, which is called Dictionary method. In this approach, there is positive and negative csv. The program compares each word in the selected csv and scores that sentence. It compares all the words in the csv and shows the resulting graph. In doing so, it also takes into account the negative suffix before the word.

## Functional Requirements

The project should be able to achieve below requirements by the time.

* + - The project builds a scraper that collects data from websites.
* Comments made in websites.
  + - It then analyzes this data to be able to reach a result from posted texts in websites.
* A Neural Network
* Naïve Bayes Text Classification
* Dictionary Method
* Word Cloud
  + - Finally, it gives a visual representation of analysis for human by analyzing from different angles.

## Nonfunctional Requirements

## 3.5.1. Quality Requirements

## 3.5.1.1 Special Quality Requirements

## a) Reliability

Since web application platform runs in the server, reliability of this platform depends on the availability of the server. Also scraping the comments depends on the availability of internet connection.

The program is tested lots of times and proven that system outputs reliable results.

## b) Usability

## These programs are not very user-friendly. To use this software, user must know something. For example, to scrap the comments, user must see the source-code of the website and determine the tag and class of the text. But user interface is not complicated. Each analysis method is divided into separate sections. And all methods are available on the main page.

## Also, there is a sequence on the desktop platform. For example, first user must select two files, then he/she click analyze and save.

## Mobile platform is quite user-friendly. All users have to do is choose the CSV. The program automatically displays graph to the user. So, a user guide will be needed on the web application and Desktop platforms.

## c) Security

As with many applications running on the web, vulnerabilities may be found on the web application platform. All noticed vulnerabilities have been removed.

## d) Availability

If web application runs on the server, Availability of the web application is depending on the availability of the server. And, to scrap the comments, an internet connection is required for Desktop and Web Application Platform. For other sections and for mobile platform, Internet connection is not required.

## e) Maintainability

For better understanding and better organization of the program, the program is divided into functions and classes. That is object-oriented programming is utilized. With this way, if something needs to be change, It can be easily changed.

## 3.5.2. Performance Requirements

## a) Timing

If number of comments put on the analyzers is high, the sentiment ratio calculation will take longer. The number of words in the comments also affects time.

## b) Speed

Speed of the program is depending on the user’s hardware. Especially, for creating the neural network, the user’s hardware is pretty important. The better the user’s hardware, the faster the program will run.

## c) Volume

The program will not take up much space. All three platforms need space that less than 100mb.

## d) Throughput

The program is going to display sentiment graphs to the user. Graphs will show the positive, negative and neutral ratio of the comments. Also, program will be able to scrap comments as csv and create a neural network model as h5 file.

## Constraints

## 3.6.1 Physical

Mobile version of the system needs an android mobile device to operate. For desktop side, System needs platform which can operate Linux or Windows operating system. On the web platform, any device which capable of running any web browser will be enough.

## 3.6.2 Legal

Analyzing and selling the data, which is obtained by scraping method, may be a legal problem. Also robots.txt file should be respected when extracting data from sites. For example, any data should not be retrieved from websites that prohibit scraping. They mentioned this on their robots.txt file.

## 3.6.3 Cultural

It will be unethical to make inferences based on people's gender, race and religion. For cultural ethics, in this project, the gender, religion and race of the commentators will not be extracted the commentator will not be known.

## 3.6.4 Environment

On the mobile platform, the program will only run on the android operating system. In desktop and notebooks, the program will require Windows or Linux operating system to operate.

## 3.6.5 Design and Implementation

The System is designed developed using Python and Java environments. The system is developed using libraries that belongs to these languages.

## 3.6.6 Interface

Python and Java programming languages is used in this project. So, web and desktop platforms need a Graphical User Interface which is compatible with python. Tkinter is used for this job. Flask framework is used on the web application platform. All kind of web development libraries can be used with flask. Bootstrap framework is used when developing the user interface of the web-application platform. On the mobile side, the project UI is dependent on android studio development environment.

## System Models

### Scenarios

### Scenario: Analyzing the sentiment scores of a brand for web application and desktop Platforms.

### User enters the website information.

### User saves this data.

### User analyzes this data with analyzer tools.

### User downloads the analyzed graphs.

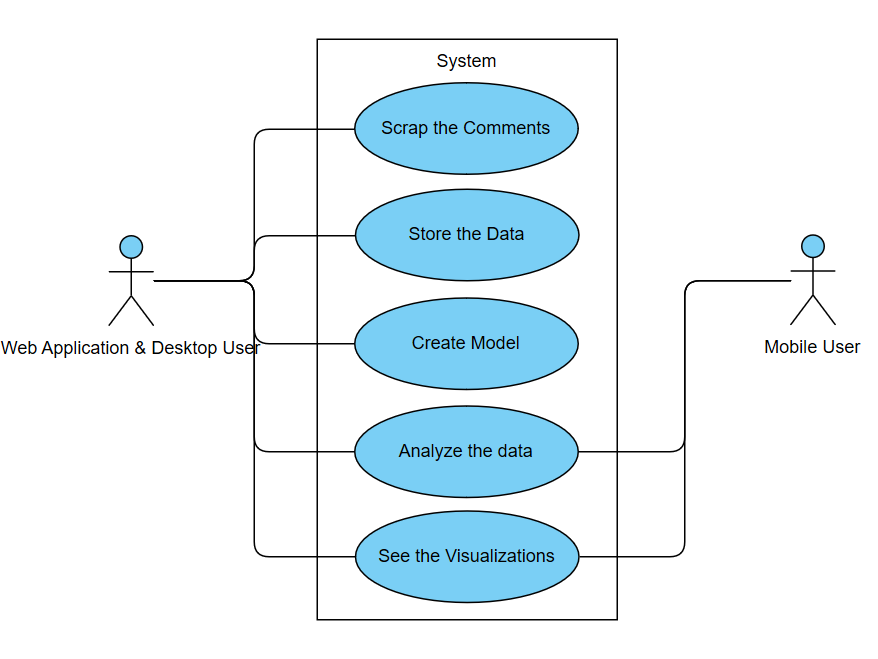
### Scenario: Analyzing the sentiment scores of a brand for web application and desktop Platforms.

### User selects data.

### User analyzes this data with using dictionary method.

### Use Case Model

The main use cases of the program are shown in the below.



**Figure 3.7.1-** Use Case Model of User.

### Static Models

### Class Diagram

### The class diagram of the program is shown in the below. Below diagrams represent the class diagrams of desktop and mobile application. For web Application platform there is no classes.

### 

**Figure 3.7.2-** Class Diagram of Project in Desktop Platform

### 

**Figure 3.7.3 -** Class Diagram of Project in Mobile Platform

### Package Diagram

### 

**Figure 3.7.4-** Package Diagram of Project

### Component Diagram

### The component diagram of the program is shown in the below.

### 

**Figure 3.7.5-** Component Diagram of Project

### Deployment Diagram

### The Deployment diagram of the program is shown in the below. Local drive may differ from platform to platform. Basically, it represents the where the comments are stored in the system.

### 

### Figure 3.7.6- Deployment Diagram of Project

### Dynamic Models

### Activity Diagram

### The activity diagram of the program is shown in the below. Below diagram represents activity diagram for desktop and web application platform. There are only Home Page, Analyze, Dictionary and Show the graph activities are in the mobile platform.

### 

**Figure 3.7.7-** Activity Diagram of Project

### Collaboration Diagram

### The collaboration diagram of the program is shown in the below diagram. Below diagram shows the collaboration diagram of the Desktop and web application platform. For mobile application, there are only Home Page, Analyze and data collaborations.

### 

**Figure 3.7.8-** Collaboration Diagram of Project

### Sequence Diagram

### The sequence diagram of the program is shown in the below diagram. Below diagram shows the sequence diagram of the Desktop and web application platform. For mobile application, there is no scraper sequence.

### 

**Figure 3.7.9-** Sequence Diagram of Project

### State Chart Diagram

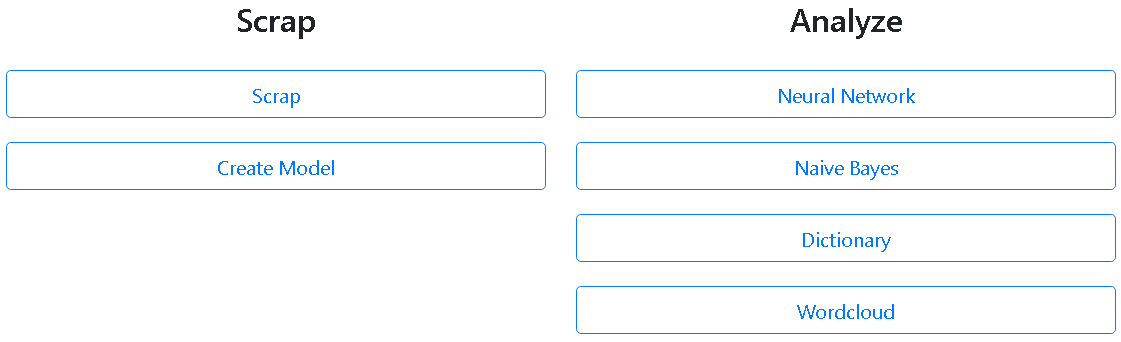
### The state-chart diagram of the program is shown in the below diagram. Below diagram shows the state-chart diagram of the Desktop and web application platform. For mobile application, there are only Home Page and analyze states.

### 

**Figure 3.7.10**- State Chart Diagram of Project

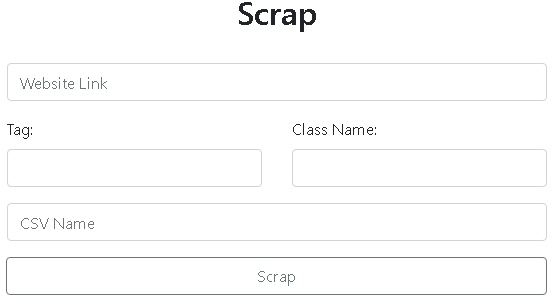
### User Interface

There are two parts on the main page. At the scrap part, the user can scrap comments in the website, or he/she can create a neural network model. The users can save the produced data and they can use these data for later use. At the right side, there is analyze part. In this part, users can use Neural Network, Naïve Bayes or Dictionary Methods to analyze their data. Or they can draw a word cloud.



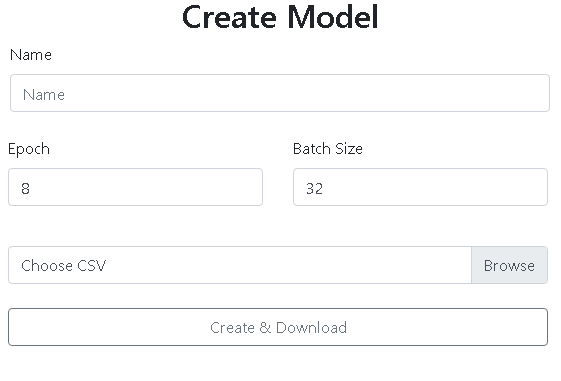
**Figure 3.8.1**-Web Application User Interface

### In the below part, users can scrap the text in a website. (Figure 3.8.2) They need to enter website URL, tag and class of the interested html document and name of the output csv.



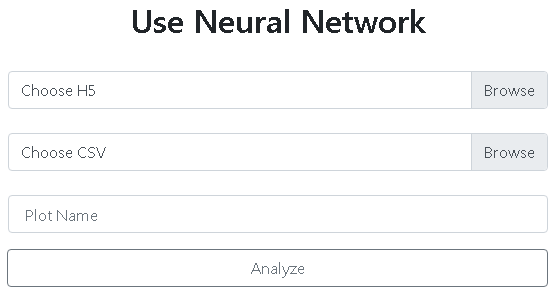
**Figure 3.8.2**-Scrap part of web application

In create model page users can automatically create and download the neural network model. (**Figure 3.8.3**)



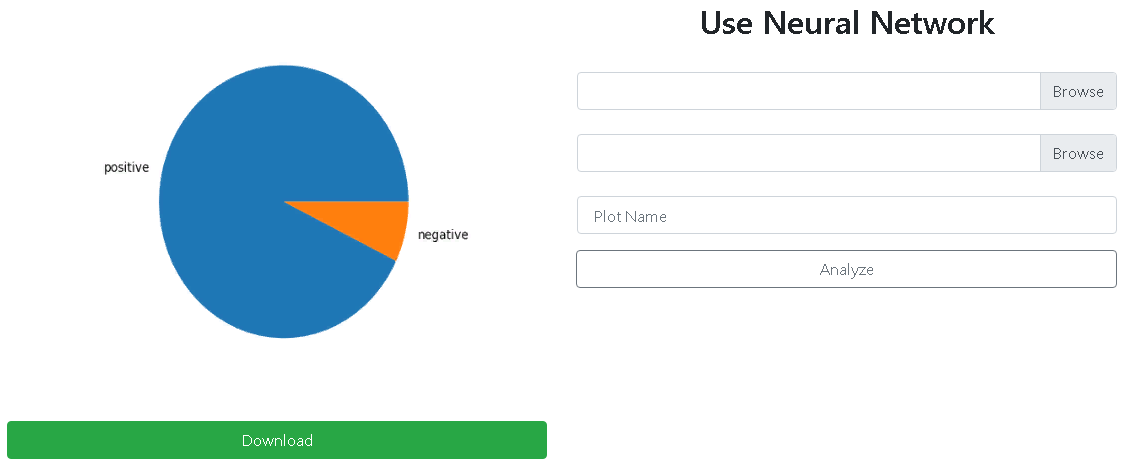
**Figure 3.8.3**- Create Model page

In use neural network page, users can analyze the comment csv using a model file. **(Figure 3.8.4)**

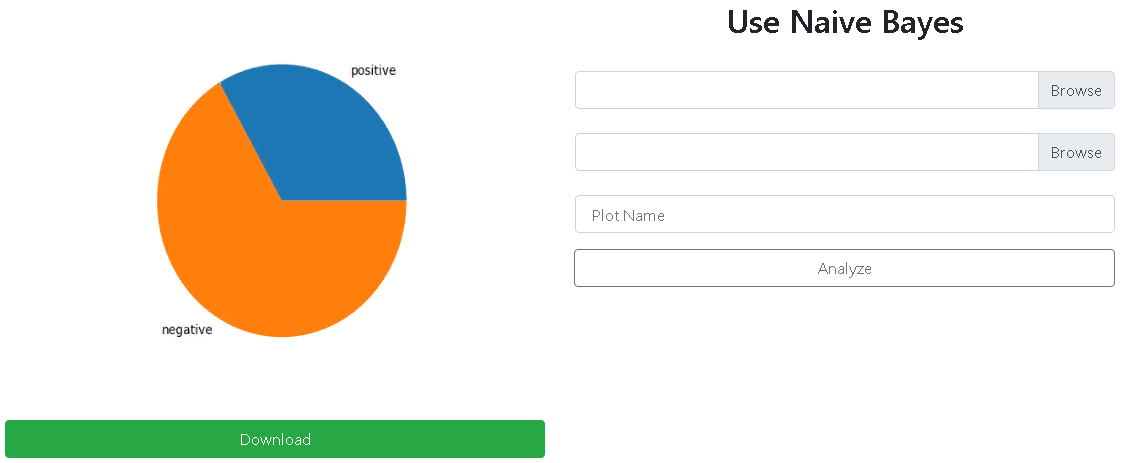


**Figure 3.8.4**-Use neural network page

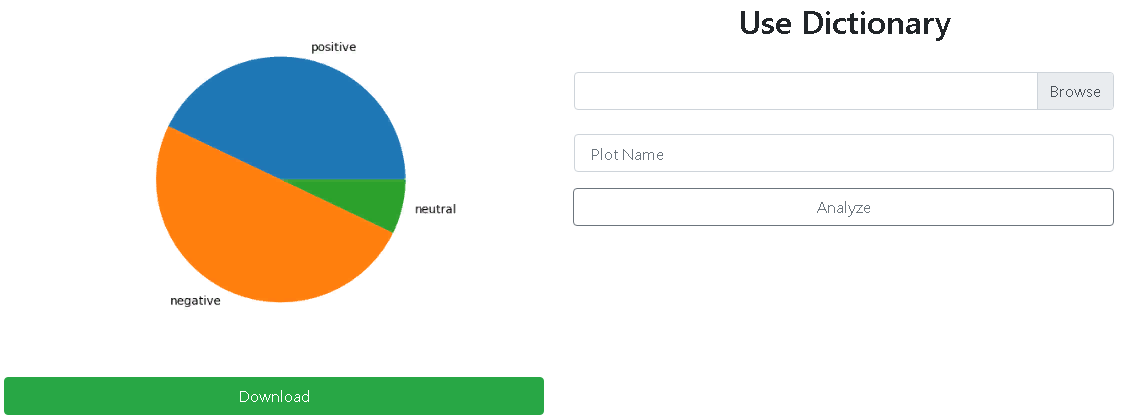
For all analyzers, when analyzing is finished, the user can see and download the resulting image. **(Figure 3.8.5)**



**Figure 3.8.5**-Page when analyzing the comments using neural network is finished.



**Figure 3.8.6**-Page when analyzing the comments using Naïve Bayes is finished.



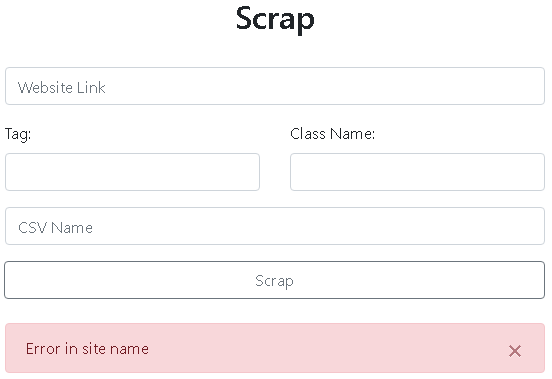
**Figure 3.8.7**-Page when analyzing the comments using dictionary method is finished.

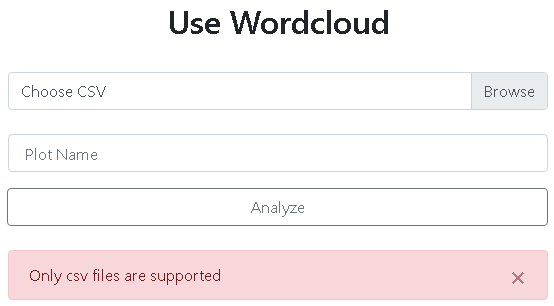
The users can also see the word cloud of comments. **(Figure 3.8.8)**



**Figure 3.8.8**-Page when analyzing the comments using word cloud is finished

The application also shows error when something is missing or wrong. **(Figure 3.8.9)**

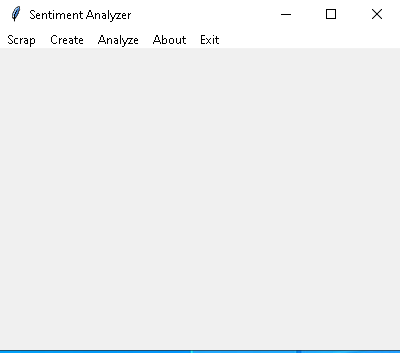




**Figure 3.8.9-**Some error pages of the web application

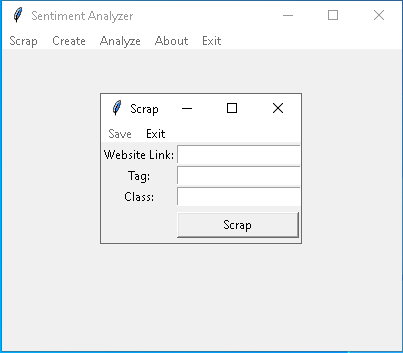
The main page of the desktop application is shown below. In the menu, there is Scrap, Create, Analyze,

About and Exit parts. **(Figure 3.8.10)**



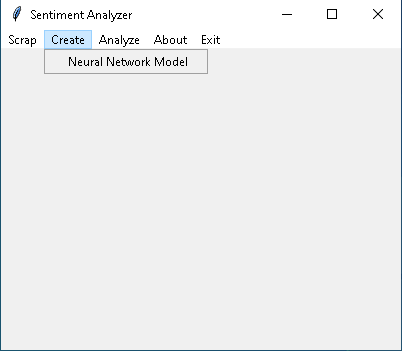
**Figure 3.8.10-** Main window of the application

In the scrap page, users can scrape the comments in the website. Save option is not active. When scrap is finished, the user can save the csv in his/her local drive. **(Figure 3.8.11)**



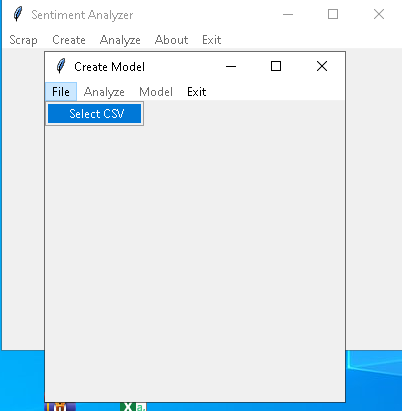
**Figure 3.8.11-**Scrap window of the application

In the create neural network model window, user can create a neural network model. **(Figure 3.8.12)**



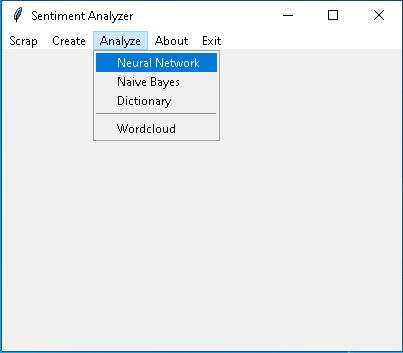
**Figure 3.8.12-** Create option of the application

In the new window, user selects a csv and if everything is okay, analyze option becomes accessible. And when analyze is clicked and analyzing process is finished, the users can click model and save. **(Figure 3.8.13)**



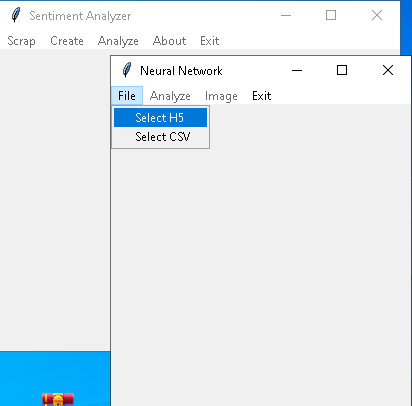
**Figure 3.8.13-** Create neural network model window of the application

In the new window, user selects a csv and if everything is okay, analyze option becomes accessible. And when analyze is clicked and analyzing process is finished, the users can click model and save. **(Figure 3.8.14)**



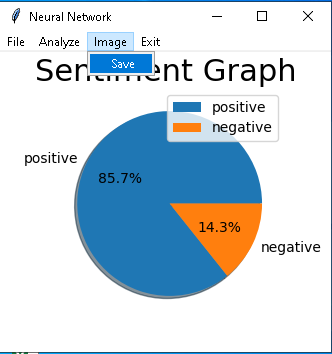
**Figure 3.8.14-** Analyze button of the application

In the neural network part, user can select h5 file and a csv file which has the comments. After that, analyze option becomes available. **(Figure 3.8.15)**



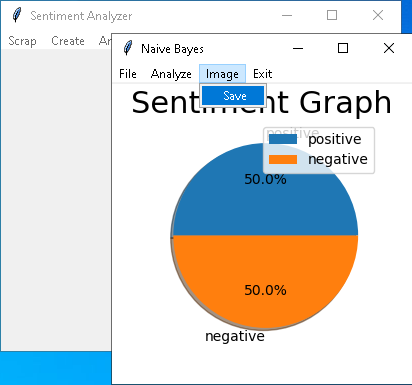
**Figure 3.8.15-** File button of the Neural network window

When analyzing is finished, user can click Image and save. **(Figure 3.8.16)**



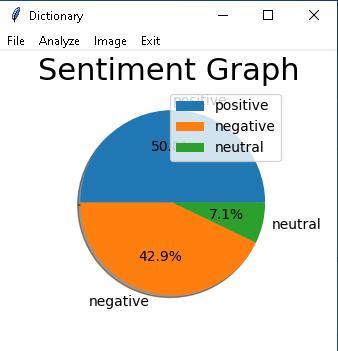
**Figure 3.8.16-** Save image button of the Neural network window

While analyzing the comments using the Naïve Bayes, the user must select training and testing csv files. Then he/she can go Analyze, image and save. **(Figure 3.8.17)**



**Figure 3.8.17-** Save image button of the Naïve Bayes window

For dictionary and word cloud parts, user can select a csv file which comments are exists. Then he/she can go Analyze, image and save. But dictionary part only supports English contents. **(Figure 3.8.18), (Figure 3.8.19)**



**Figure 3.8.18-** Window when analyzing the comment is finished using dictionary method

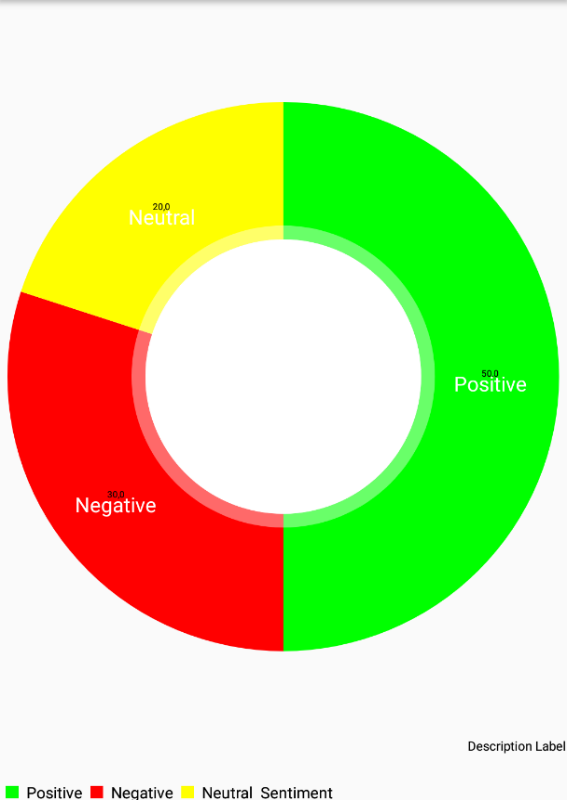


**Figure 3.8.19-** Window when analyzing the comment is finished using word cloud method

The user will also be able to see the graphics in the mobile application. The graphs will appear as follows. The example pie chart shows the positive, negative and neutral ratio of the csv comments. **(Figure 3.8.20)**



**Figure 3.8.20-**The users have to select a csv to analyze



**Figure 3.8.21-** Resulting pie chart on the mobile platform

# Conclusion

As conclusion, Sentiment Analysis is quite useful and important. Nowadays, people can collectively share their thoughts and ideas about any topic. And web is the key source of that kind of information. Based on the posts in the web, it can be determined what society thinks about an issue.

A scraper and analyzer were made in this project. A scraper can easily collect the post in a website and store it. On the other hand, many kinds of analyzers have been added to analyze this collected data. And in addition, a Neural Network model creator is also added to the project. So, with that way, user can create a neural network model and use it after.

The analyzers are selected from different approaches. As known, sentiment analysis is divided into two segments. One is Machine Learning approach and other one is Lexicon Based approach. Machine Learning Approach has sub-branches like Supervised and unsupervised. Also supervised approach has sub-branches. Neural network approach is belonging to the Linear classifier. And Naïve Bayes is belonging to the probabilistic classifier. And Dictionary based approach is subbranch of Lexicon Based approach. In addition to these, the user can also draw a word cloud. The word cloud is useful and easy way to see most used words in a text.

For web application and desktop part, Both Windows and Linux operating systems are supported. Web application can run Linux and windows server. Desktop platform also can be used in both platforms. Mobile application is developed for the Android devices.

Unlike other systems, these programs can scrape any text in any websites. Also, except dictionary method, this program is multilingual and can analyze the texts that in any language. For example, If user train the model in French words, he/she can analyze the French texts.

# References

1. Social Network API, GeeksforGeeks. <https://www.geeksforgeeks.org/social-network-api/>.
2. What is web scraping, Scrapinghub: <https://scrapinghub.com/what-is-web-scraping>
3. Machine-Learnig approach, Kdnuggets: <https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>
4. Lexicon-Based approach, Kdnuggets: <https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>
5. BeautifulSoup, Scrapinghub: <https://info.scrapinghub.com/web-scraping-guide/python-web-scraping-libraries-and-frameworks>
6. Recurrent Neural Networks, Wikipedia,

<https://www.wikizeroo.org/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvUmVjdXJyZW50X25ldXJhbF9uZXR3b3Jr>

1. Introduction to RNNs, wildml, <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>
2. Naïve Bayes, Medium, <https://medium.com/@martinpella/naive-bayes-for-sentiment-analysis-49b37db18bf8>
3. Flask, Github, <https://github.com/pallets/flask>
4. Tkinter, Python, <https://docs.python.org/2/library/tkinter.html>
5. Bulut, Faruk. "Low dynamic range histogram equalization (LDR-HE) via quantized Haar wavelet transform." *The Visual Computer* 38, no. 6 (2022): 2239-2255.

# Glossary

## Definitions

## *Framework -* is a platform for developing applications. It provides an environment on which developers can build software for a specific platform.

## *Lexicon -* all the words utilized in a specific dialect or subject, or a word reference.

* 1. **Abbreviations**

|  |  |
| --- | --- |
|  |  |
| API | Application programming interface |
| GUI | Graphical user interface |
| HTML | Hypertext Markup Language |
| XML | eXtensible Markup Language |
| LSTM | Long short-term memory |
| UI | User Interface |