Software Requirements Specification

Political Tweets Classification

Prepared By

Furkan Mete

Aziz Doğay Ekinci

Berkay Osman Yıldırım

16/12/2021

Content

Change History

Name	Date	Reason for Change	Version
Aziz Doğay Ekinci	22.11.2021	The libraries to be used have been defined. The path of the dataset in the computer is defined.	0.1
Furkan Mete	24.11.2021	The null values in the dataset were checked and the code to clear the unnecessary data in the dataset was added.	0.2
Berkay Osman Yıldırım	26.11.2021	Comments have been added to the code so that the code can be easily understood by the developers.	0.3
Berkay Osman Yıldırım	27.11.2021	Access to these columns was obtained by obtaining the unique items in the columns. Checked the first five rows in the dataframe.	0.4
Aziz Doğay Ekinci	29.11.2021	Factorize function was used. A bar graph was made to understand the number of political and non-political tweets.	0.5
Furkan Mete	03.12.2021	Links were made for political and non-political tweets.	0.6
Aziz Doğay Ekinci	06.12.2021	Some machine learning applications were made and their precision was shown as a box plot. Based on the graph, an algorithm was selected and confusion matrix was created.	0.7
Berkay Osman Yıldırım	08.12.2021	Misclassifications were checked. Final classification reports printed.	0.8
Furkan Mete	13.12.2021	Comments have been added where necessary so that the code can be understood and developed more easily for developers.	0.9
Aziz Doğay Ekinci, Furkan Mete, Berkay Osman Yıldırım	16.12.2021	Final checks were made and improvements were made where necessary. The software is ready.	1.0

1. Introduction

1.1 Purpose

A software prepared for the Political Tweets Classification Big Data Algorithm course project. The Project aims to include all 5000 political tweets written in 2015 in the USA in a classification within the dataset. Classification is for information purposes, announcement of media image, attack on another island, etc. will be divided into groups such as It is also to find the relationship between the attitudes of different political groups and to predict the political groups and hashtags of the tweets. It displays graphics for easy analysis after classification. The software requirements analysis document was prepared for the first version of Political Tweets Classification.

1.2 Document Standards

Bold fonts are headings in the document. Subheadings have the same priority as the main sections.

1.3 Target Audience and Reading Recommendations

This document is aimed at programmers, data analysts, Tweet analysts and analysts interested in politics. After the introduction in the first section, the second section contains a brief summary of the product's technical features and structure. In the third chapter, the interfaces and external requirements of the program are explained. In the fourth chapter, the features that should be in the system are explained in detail. Those who are not interested in the technical part of the program should read the fourth and fifth chapters.

1.4 Product Scope

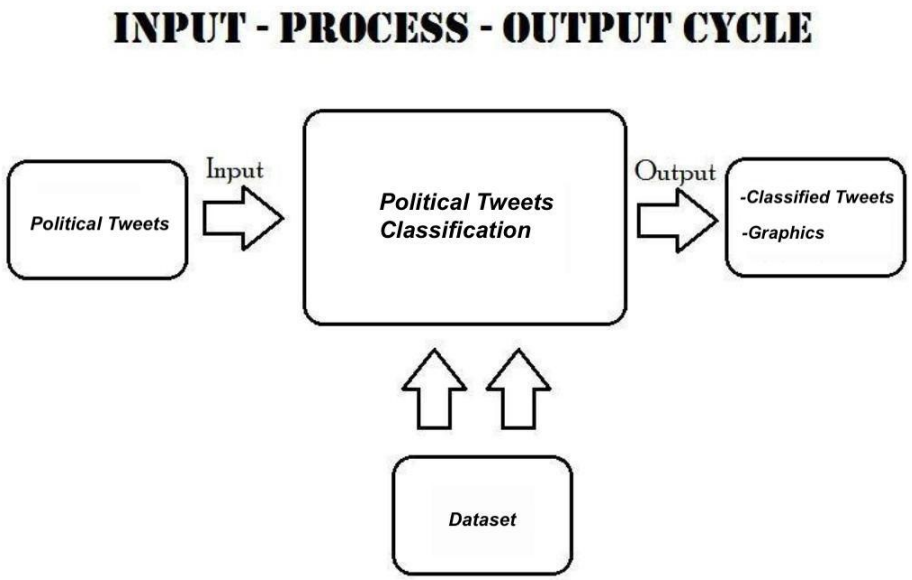
Political Tweets Classification project aims to classify the tweets that have been sent and to make analysis on the tweets as easy as possible. It was written for those who will divide the tweets sent by the users into groups and analyze the data more easily and clearly. It also shows the distribution of tweets as graphs after classification. Political Tweets Classification project has a low cost structure.

1.5 References

The document you are currently reading is the latest version of this document. To reach the final version of the project, it is sufficient to open the zip file sent to you. For this document template, “yazılım gereksinim analizi bilet takip” is referenced.

2. General Description

2.1 Software Perspective – IPO (Input-Process-OutPut) Diagram



A dataset containing political tweets is obtained.

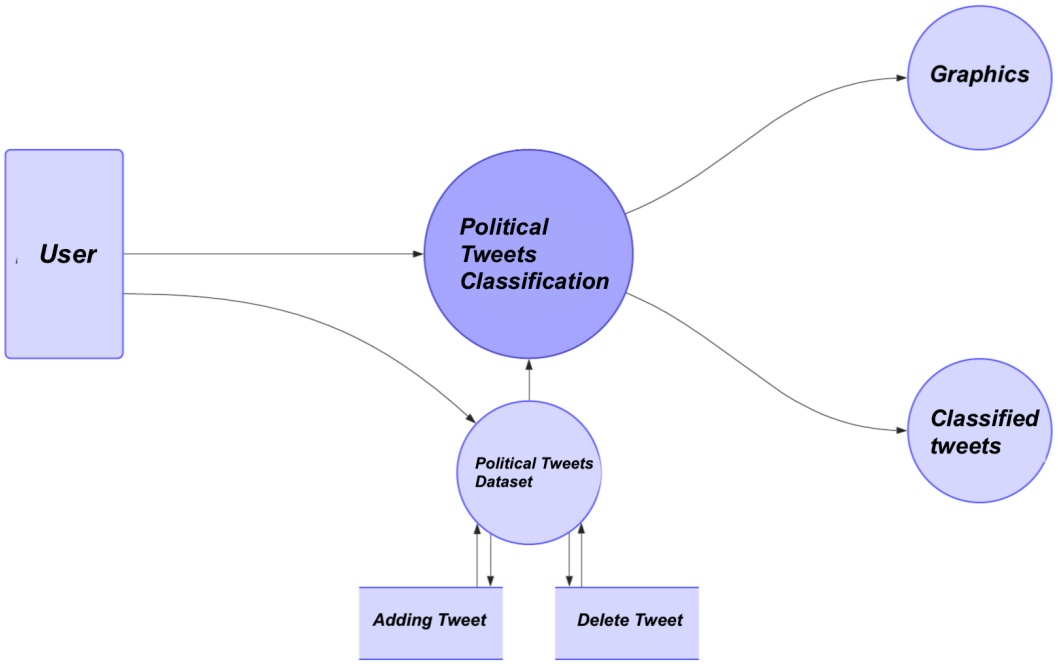
The path of the dataset is entered into the program.

Political Tweets Classification is run.

The resulting classification and graphs are ready for analysis.

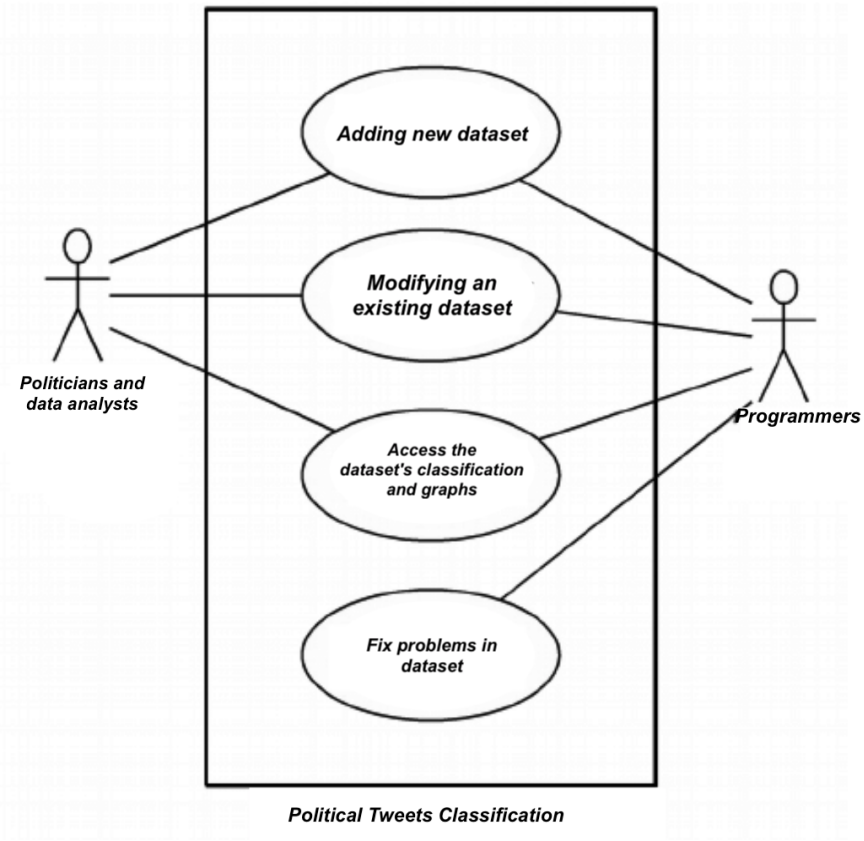
Classification of Political Tweets is software that will save a lot of time for self-employed data analysts and politicians. It is not a sub-product of any software. It aims to meet the needs of analysts for individual employees and politicians who want to analyze.

2.2 Product Functions, Data Flow Diagrams (DFD)



Political Tweets Classification is a single-user, easy-to-use software. Notable features; high success rate, effective graphics and free software. You can find the details of each feature included in the project in the fourth section. Easily adaptable to meet the needs of all analysts and politicians.

2.3 User Classes and Behaviors Use Case Diagrams



2.3.1 Politicians and Data Analysts

The main target audience of the project are users, politicians and data analysts. They will use the program many times to save time and do better analysis. The importance of ease of use for this group is at the highest level. It should be such that it can change the content of the political Tweet dataset at will. They will be able to use the dataset they want and change its content. Consistent and successful operation of the whole system is important for this user group. Politicians and data analysts are the group of users who should particularly like the project.

2.3.2 Programmers

In addition to politicians and data analysts, they are people who can interfere with the software in case of an error. In addition to changing the dataset, when there is a problem in sending the dataset to the software, they quickly fix the problem.

2.4 Environment, Technology and Hardware

Political Tweets Classification will work on laptops and desktops. Although users have different computer characteristics, it can be run online even through any browser. The operating system must be standalone. It should work on all platforms, Windows, MacOS, Linux and other similar operating systems. Classification of Political Tweets needs a dataset with itself. It should work fine with at least one dataset server or dataset file on the computer.

2.5 Design and Implementation Constraints

The project is free software, it still needs to work with free software so that it can be easily distributed and used. These should be considered when choosing a database and external application. Since it will also be used in low-capacity computers, the system should not use much memory and processor, and should do as much work with as little load as possible. When used within the network or on the Internet, the response time should be acceptable. While developing the project, free software and technologies that will run on the Linux operating system should be selected for external participation. Since it is licensed under the GNU/GPL, there are no additional applications to be overwritten and there are no obstacles such as distribution and licensing. Python programming language and Jupyter development platform were found suitable for development.

2.6 User Documentation Required Features

Along with the project, there should be help files, user documents and a developer document on the web. Help files will be prepared in detail for each section and will be included in the program. Within the project, a detailed user documentation should accompany the product. This document should be written in such a way that a beginner user can understand both the content and usage of the program. The developer document on the website will be prepared mostly for programmers to access information on which they can add and add new codes. The document containing the design and database of the system should be submitted to the designers' information.

2.7 Acceptances and Interactions

Political Tweets Classification can only work with one dataset at a time. Although the dataset has almost no limit, it is not possible to work with more than one dataset. Since it is written in the Python programming language, the development of the software should be done by programmers who know the python programming language.

3. External Interface Requirements

3.1 User Interfaces

Political Tweets Classification software *user interface in the may vary depending on the program it uses. The user will first enter the file path of the dataset into the software. Then it will be enough to click on the run part of the program you are using. The software first shows the classification of political tweets. Then, the numbers of groups in this classification are shown as a bar graph. The output of the interface is the screen, and the inputs are the mouse and keyboard.*

3.2 Hardware Interfaces

System runs on any hardware that supports Python. Mouse and keyboard required for use. The hardware in the PCs is sufficient for the program to work. If the software is to be used via an online compiler by connecting to the server over the network, a wired or wireless internet connection is required.

3.3 Software Interfaces

Political Tweets Classification to work, the Python programming language must support machine learning. One of the supported datasets to be used in the software must be installed on the computer. The libraries used in the software will come ready-made in the program you use.

3.4 Communication Interfaces

Software will not be used online and the dataset is already present on the computer, the software can be used without the need for any protocol. If the software is to be run through an online program, the default protocol, HTTP, will be used. According to the confidentiality of the datasets, the protocol selections may vary by the user.

4. System Features

The features that should be in the system are listed below without grouping them. You can participate in the project development and make suggestions by sending an e-mail to metefurkan@gmail.com or dogay_ekinci@hotmail.com about the missing parts and changes.

4.1 Ease of Use

4.1.1 Description and Priority

Ease of use in the project is highly important. Since there will be people who will use the Political Tweets Classification software at the computer all day, attention should be paid to ease of use, including even the smallest details, to speed up their work.

4.1.2 Warning/Response Sequence With

This feature software can be used by simply changing the content of the dataset. Thus, the desired dataset can be easily used by the software.

4.1.3 Functionality Requirements

The dataset file on the computer must be replaceable. It is the presence of a program on the computer that can open the data set necessary for the exchange of the data set. This program may vary depending on the dataset used.

4.2 Different Database Support

4.2.1 Description and Priority

Datasets used by users may differ. Thanks to the different database support, a flexible program that can adapt to every platform occurs. This need can be eliminated if a database that supports macOS, Linux and Windows platforms is selected. Moderately important. It provides convenience for future adaptations.

4.2.2 Warning/Response Sequence

Before using the software, it should be adapted to the software by paying attention to the extension of the dataset. When using different dataset extensions, the file path in the software must be updated.

4.2.3 Functionality Requirements

The dataset must be loaded and accessible in the system. Its path inside the computer must be clear. Otherwise, the software cannot use the dataset. The dataset path of the users must be entered into the software beforehand.

4.3 Multiple Platform Operation

4.3.1 Description and Priority

The project needs to support these platforms, as Linux and the operating system become widespread, Windows dominates the market, and some users use MacOS. Especially Linux and Windows support is highly important. The program should provide the same response and display on different platforms.

4.3.2 Warning/Response Sequence

The user should be able to see the classification results and graphics in the project, no matter what operating system he/she uses.

4.3.3 Functionality Requirements

It can be used online from the browser via the compiler on computers with Internet access. For computers without internet access, a program that can compile python must be installed. Otherwise, the software will not be usable.

4.4 Data Security

4.4.1 Description and Priority

Information records of people who write Tweets are highly confidential. Political Tweets Classification from dangers arising from the connecting to the Internet or using a system open to the Internet must be avoided. It is highly important that the system gives confidence and that the data cannot be viewed by outsiders.

4.4.2 Warning/Response Sequence

Users should be able to turn off the entire Internet connection of the program if they wish.

4.4.3 Functionality Requirements

Internet-related features of the system should be offered optionally. A program that can compile the Python programming language currently installed on the computer must be installed.

5. Other Non-Functional Requirements

5.1 Performance Requirements

System should run acceptably fast. The dataset connection should not be such that the person working at the computer will notice, regardless of the speed of the incoming results, the connection capacity on the network. For this, the approval of consultants and users, and test results on different platforms should be checked. The system can only accept one dataset at a time.

5.2 Health Requirements

Within the scope of this project, no health needs directly related to the project were specified.

5.3 Security Requirements

Data protection and security can be of high importance in Political Tweets Classification software. Data loss can cause great damage as tweet records and personal information are kept. First of all, Transaction properties will be used in the database for the integrity of the data. In this way, data integrity is ensured in case the process is damaged while adding or removing. While processing the data set, it is necessary to work in a stable environment for system integrity in cases such as computer crashes or power cuts. Otherwise, the most recent transaction will be lost if a record has not been retrieved manually.

5.4 Software Attributes

The most important features of the software, which are mentioned in detail in the fourth chapter, are ease of use, support for different datasets and platform independence. It will be preferred because it works on all kinds of operating systems and provides convenience to users. Because the dataset is used, it will run powerfully and with high stability.

6. Other Requirements

Within the scope of this project, there are no other requirements related to the project.

7. Future Plans

Software will be developed to work with more than one dataset at the same time.
The software will be made to produce more graphics for users so that users' analysis will become easier and more diverse.

Appendix A: Terms

Linux: One of the core software that is the most basic part of computer operating systems. It is a free software project released under the GNU General Public License version 2 and developed under the umbrella of the Linux Foundation.

Python: An object-oriented, interpretative, modular and interactive high-level programming language.

Dataset: It is a working object that is disconnected from the database. Thanks to its pure XML-based structure, it provides seamless integration with web forms and DotNet components.

macOS: Macintosh a proprietary graphical operating system developed and marketed by Apple Inc. since 2001. It is the primary operating system for Apple's Mac computers.