**Fake News Detection**

****

*Developed by:*

**Abdul Wahab** 131-FBAS/BSIT4/F16

*Supervised by:*

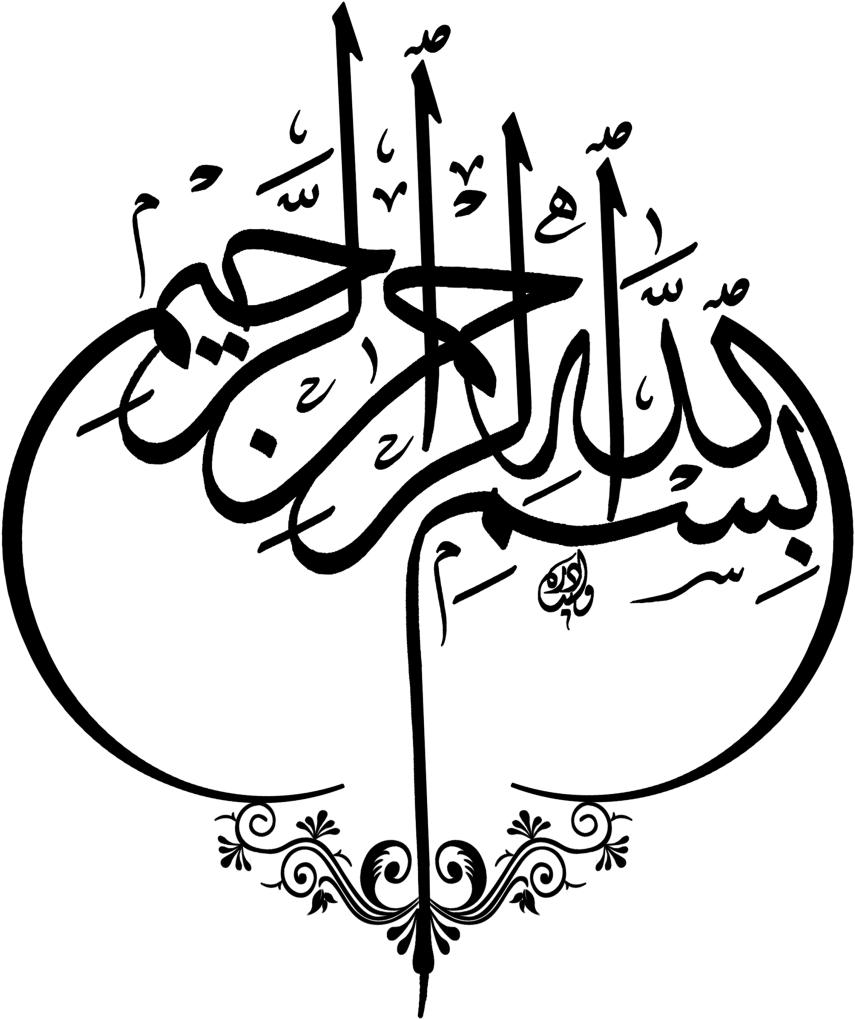
**Dr. Syed Musharaf Ali**

**Department of Computer Science and Software Engineering**

**Faculty of Basic and Applied Sciences**

**International Islamic University, Islamabad**

**2020**

****

**"In the Name of ALLAH, the Most Beneficent, the Most Merciful"**

\*

Department of Computer Science &Software Engineering

**INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD**

**Final Approval**

**Date:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

It is certified that I have checked the documentation titled “(Fake News Detection)” submitted by **Abdul Wahab (131-FBAS/BSIT4/F16)**. It is my discernment that this project is of enough standard to permit its acceptance by the International Islamic University, Islamabad for bachelor’s degree in Information Technology.

**Committee:**

**External Examiner:**

**Mr.**

Department of Computer Sciences and Software Engineering

International Islamic University Islamabad

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Internal Examiner:**

**Mr.**

Assistant Professor

Department of Computer Sciences and Software Engineering

International Islamic University Islamabad

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor:**

**Dr. Syed Musharaf Ali**

Department of Computer Sciences and Software Engineering International Islamic University Islamabad

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Dissertation*

A critique submitted to

Department of Computer Science & Software Engineering,

International Islamic University, Islamabad

As partial accomplishment of the requirements

For award of the degree of

Bachelor’s in Information Technology.

i | P a g e

*Dedication*

**Dedication**

To my grandfather,

The reason for what I am today.

Thanks for his encouragement, endless love and support.

To my teachers,

Gratitude for their endless help, supervision and hard work.

ii | P a g e

*Declaration*

**Declaration**

I declare that this project and documentation completely based on my own efforts made under the supervision of my supervisor Dr. Syed Musharaf Ali. This work contains nothing from other sources, institutes or universities. I declare that this project and documentation are submitted as requirements for the degree of Bachelor of Science in Information Technology.

**Abdul Wahab**

(131-FBAS/BSIT4/F16)

iii | P a g e

*Acknowledgement*

**Acknowledgement**

First of all I would like to thank my project supervisor **Dr. Syed Musharaf Ali** for helping and his continuous guidance and support. Some other people that deserve greatly to be acknowledged that are Dr. Jamal Abdul Nasir, Mr. M. Imran Saeed, Dr. Qamar Abbas,

Mr. Majid Bashir, Mr. Muhammad Nadeem and Dr. Muhammad Adeel are the best teachers who guided me in different situations and gives me good advices all the time.

The credit goes to my class mates and the members of department of Computer Science who taught me in last four years. All the teachers were friendly and supportive. Now I am able to submit final year project because of their hard work, support and teachings. I had a vital company of the best teachers I ever had. Their supervision maintained my progress in track. I also would like to acknowledge the guidance from teachers as well as the panel particularly in my final year project presentation that helped me to improve my presentation skills by their comments and tips.

iv | P a g e

*Project in Brief*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Project in Brief** | |
|  |  |  |  |  |  |
|  | ***Project Title:*** |  | Fake News Detection | | |
|  |  |  |
|  |  |  |  |  |  |
|  | ***Undertaken By:*** |  | Abdul Wahab | | |
|  |  |  |  | | |
|  |  |  |  | | |
|  |  |  |  |  |  |
|  | ***Supervised By:*** |  | Dr. Syed Musharaf Ali | | |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  | ***Date Started:*** |  | 04-02-2019 | |  |
|  |  |  |  |  |  |
|  | ***Date Completed:*** |  | 31-6-2020 |  |  |
|  |  |  |  |  |  |
|  | ***Tools,*** |  | Anaconda,Python, Scikit learn, Machine learning | | |
|  | ***Technologies and*** |  | Jupyter Lab , Natural Language Toolkit | | |
|  | ***language Used:*** |  |
|  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  | ***System Used:*** |  | Core m3, RAM 8 GB | | |
|  |  |  |  |  |  |

v | P a g e

*Preface*

**Preface**

This documentation is divided in chapters for sake of understanding. Each chapter explained methodologies that are required for understanding of the project.

This report contains the following chapters.

Chapter 1: It presents overview and introduction of system.

Chapter 2: It describes analysis of the system.

Chapter 3: It gives design process of the system.

Chapter 4: It explains system components and design.

Chapter 5: It explains System Testing.

vi | P a g e

*Table of Contents*

**Contents**

**1. SYSTEM INTRODUCTION ------------------------------------------------------------** **1**

1.1. OVERVIEW ----------------------------------------------------------------------------------------------- 1

1.2. Fake News Detection ---------------------------------------------------------------------------------------- 1

1.3. The Idea------------------------------------------------------------------------------------- 1

1.4. Problem Statement----------------------------------------------------------------------------------- 2

1.4.1. Human Effort------------------------------------------------------------------------------------ 2

1.5. Proposed Solution ----------------------------------------------------------------------------------- 2

1.6. Project Scope **------------------------------------------------------------------------------** 2

1.7. Machine Learning------------------------------------------------------------------------- 2

1.7.1. Natural Language Processing ------------------------------------------------------------------- 3

1.8. Main Modules -------------------------------------------------------------------------------------------- 3

1.8.1. Dataset ------------------------------------------------------------------------------------- 3

1.8.2. Pre-Processing----------------------------------------------------------------------------- 3

1.8.3 Filtering ------------------------------------------------------------------------------------- 3

1.8.4. Classifier ------------------------------------------------------------------------------------ 3

1.9. Literature Review---------------------------------------------------------------------------------- 4

**2. SYSTEM ANALYSIS ---------------------------------------------------------- 5**

**2.1 SYSTEM ANALYSIS ----------------------------------------------------------------------- 6**

2.1 Use Cases --------------------------------------------------------------------------------------- 6

2.2 Actors -------------------------------------------------------------------------------------------- 6

2.3 USE CASES Diagram---------------------------------------------------------------------------- 6

2.4 Use Case description --------------------------------------------------------------------------- 7

2.4.1 Dataset -------------------------------------------------------------------------------------- 7

2.4.2 Pre Processing ------------------------------------------------------------------------------ 8

2.4.3 Filtering - ------------------------------------------------------------------------------------ 8

*Table of Contents*

2.4.4 Classifier ------------------------------------------------------------------------------------- 8

2.5 Detailed Use Cases ------------------------------------------------------------------------------- 9

2.5.1 Dataset---------------------------------------------------------------------------------------- 9

2.5.2 Pre Processing ------------------------------------------------------------------------------- 9

2.5.3 Filtering --------------------------------------------------------------------------------------- 10

2.5.4. Classification --------------------------------------------------------------------------------- 10

2.6. Sequence Diagram --------------------------------------------------------------------------------- 11

2.6.1 Pre Processing --------------------------------------------------------------------------------- 11

2.6.2 Filtering ---------------------------------------------------------------------------------------- 12

2.6.3. Classifier -------------------------------------------------------------------------------------- 12

2.7. Operation Contracts ------------------------------------------------------------------------------- 13

2.7.1. Dataset ---------------------------------------------------------------------------------------- 13

2.7.1.1. Datasetfilename -------------------------------------------------------------------- 13

2.7.2. Pre-Processing ------------------------------------------------------------------------------- 13

2.7.2.1. Pre-Processing ---------------------------------------------------------------------- 13

2.7.3. Filtering -------------------------------------------------------------------------------------- 14

2.7.3.1 Filtering ---------------------------------------------------------------------------- 14

2.7.5. Classification --------------------------------------------------------------------------------- 14

2.7.5.1 Classifiers ---------------------------------------------------------------------- 14

2.7.5.3. DrawGraph ----------------------------------------------------------------------------- 14

2.8. Domain Model----------------------------------------------------------------------------------------- 15

2.9. Activity Diagram-------------------------------------------------------------------------------------- 15

*Table of Contents*

**3. METHODOLOGY AND SYSTEM DIAGRAMS --------------------------------------- 16**

**3.1. Methodology --------------------------------------------------------------------------------------- 17**

3.1.1. Dataset ----------------------------------------------------------------------------------------- 17

3.1.2. Pre Processing ---------------------------------------------------------------------------------17

3.1.3. Filtering ---------------------------------------------------------------------------------------- 18

3.1.4. Classification ---------------------------------------------------------------------------------- 18

3.1.5. Graph Plotting -------------------------------------------------------------------------------- 18

**3.2 System Design ------------------------------------------------------------------------------------- 19**

3.2.1 Sequence Diagram ------------------------------------------------------------------------ 20

3.2.1.1 Pre Processing ----------------------------------------------------------------- 20

3.2.1.2 Filtering ------------------------------------------------------------------------ 20

3.2.1.3 Classifier ----------------------------------------------------------------------- 21

3.2.2 Class Diagram------------------------------------------------------------------------------ 21

**4. EXPERIMENTAL RESULTS & IMPLEMENTATION -------------------------------- 22**

**4.1 Experimental Results ---------------------------------------------------------------------------------** **23**

4.1.1. Dataset------------------------------------------------------------------------------------ 23

4.1.2. Pre Processing ------------------------------------------------------------------------- 23

4.1.3. Filtering of Pre-processed data ------------------------------------------------------- 23

4.1.4. Classifier ---------------------------------------------------------------------------------- 23

**5. SYSTEM TESTING ----------------------------------------------------------------------------- 24**

**5.1. Introduction-------------------------------------------------------------------------------------------** **25**

5.1.1 Unit Testing -------------------------------------------------------------------------------- 25

5.1.2 Module Testing ---------------------------------------------------------------------------- 25

5.1.3 Sub System Testing ----------------------------------------------------------------------- 25

*Table of Contents*

5.1.4 System Testing ---------------------------------------------------------------------------- 26

5.1.5 Black Box Testing------------------------------------------------------------------------- 26

5.1.6 White Box Testing ------------------------------------------------------------------------ 27

5.1.7 Acceptance Testing ----------------------------------------------------------------------- 27

**5.2. SOFTWARE TESTING ----------------------------------------------------------------------------------** **27**

5.2.1 Test Cases ---------------------------------------------------------------------------------- 27

5.2.1.1 Dataset------------------------------------------------------------------------- 28

5.2.1.2 Pre-Processing -----------------------------------------------------------------28

5.2.1.3 Filtering of Data -----------------------------------------------------------29

5.2.1.4 Classifier ------------------------------------------------------------------------30

5.2.1.4.1 Naïve Bayes ---------------------------------------------------------------31

5.2.1.4.2 Logistic Regression ------------------------------------------------------32

5.2.1.4.3 Random Forest ------------------------------------------------------------33

5.2.1.4.4 Decision Tree -------------------------------------------------------------34

5.2.1.4.5 KNN ------------------------------------------------------------------------35

5.2.1.4.6 Support Vector Machine ------------------------------------------------36

5.2.1.4.7 Neural Network -----------------------------------------------------------37

**6. CONCLUSION ---------------------------------------------------------------------------** **38**

**6. CONCLUSION---------------------------------------------------------------------------------------** **39**

6.1. Future Enhancements --------------------------------------------------------------------- 39

**6. REFERENCES ----------------------------------------------------------------------------** **41**

**CHAPTER: 1**

**INTRODUCTION**

*Chapter 1* *Introduction*

**1. System Introduction**

I will use publicly available data from Kaggle for fake news detection this data is in text form and it will used to detect only text type of news. I perform natural language processing (NLP) on it in order to predict real and fake news classification based on the text entered by user. I take the approach of identifying a large set of influential news or Dataset and analysing that with a variety of NLP techniques. Both sentence-level and word-level language features are used, and the NLP pipeline is built to consider fake news detection.

My system used a variety of techniques to detect fake news through text entered by user. The volume and news data from this system I review a number of sampling approaches and methods of modelling fake news detection.

All these studies showcased that training data for fake news as a valuable source and a powerful tool for conducting studies and making detection of any type of fake news. I organized my system as follows. Section 1 describes the related works and Section 2 discussed the data portion demonstrating the data collection and pre-processing part. In Section 3 I discuss the features extraction part in my work followed by Section 4 which examines the correlation part for fake news detection. In Section 5 I present the result, accuracy and precision of my classification models and applying some algorithms followed by the accuracy of algorithms. In Section 6 I will present my conclusions and Section 7 deals with my future work plan and taking good decision in best way.

**1.1. Overview**

My system takes news as an input from user and classify that news into categories of real and fake and predict these categories of news.

**1.2. Fake News Detection**

Fake news detection is a System through which I can classify a text news whether the news is real or fake.

**1.3. The Idea**

The Idea behind this project is to provide a platform using Machine Learning and Natural Language Processing techniques to detect the fake news from text.

**Fake News Detection** 1 | P a g e

*Chapter 1* *Introduction*

**1.4. Problem Statement**

Currently there is no such system to classify and detect fake news that is circulating around us.

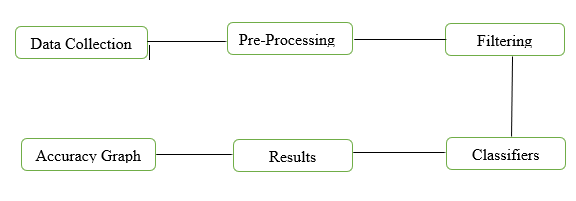
**1.4.1. Human Effort:**

Even today most of the detection of such is done manually. It consumes a lot of

human efforts to detect fake news.

**1.5. Proposed Solution**

The suggested system is to get news from user and classify that news into real or fake by using ML , NLP and predict the results. It is faster, cheaper.



Classifiers

Filtering

Pre-Processing

Accuracy Graph

Results

Data Collection

**1.6. Project Scope**

My proposed design is meant to solve the problem related to text types of news posts available on media websites. However it is currently a detection model only but it can lead to detection and deletion of these post in near future.

**1.7 Machine learning**

Machine learning is a core sub-area of artificial intelligence; it enables computers to get into a mode of self-learning without being explicitly programmed. When exposed to new data, these computer programs are enabled to learn, grow, change, and develop by themselves. SAS, a North Carolina-based analytics software developer, uses this definition; “Machine learning is a method of data analysis that automates analytical model building.”

**Fake News Detection** 2 | P a g e

*Chapter 1* *Introduction*

In other words, it allows computers to find insightful information without being programmed where to look for a particular piece of information; instead, it does this by using algorithms that iteratively learn from data. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction/recommendation-making through the use of computers. It has strong tie

to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

**1.7.1 Natural Language Processing**

In Natural Language processing is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of data. In natural language processing I pre-process, filter the data and apply algorithm in order to predict.

**1.8. MAIN MODULES**

There are four main modules:

**1.8.1. Dataset**

First, I collect dataset about fake news detection in text form which contain in csv file.

**1.8.2. Pre-Processing:**

Then I pre-processed the dataset, the main purpose of pre-processing is to remove hashtags, extra white spaces, convert news into lower-case, and remove punctuations etc.

**1.8.3. Filtering:**

Then I filter the pre-processed data. To filter the data I use stop word dictionary and remove all the stop words from text and get the filtered text. Now the data I get is ready for analysis.

**1.8.4. Classifier:**

Classifier is used to check the sense of news (real, fake) and check the accuracy of my results.

**Fake News Detection** 3 | P a g e

*Chapter 1* *Introduction*

**1.9 Literature Review**

One of the latest Research paper named **Fake news detection using naïve Bayes classifier.** This article describes a simple method for fake news detection based on one of the artificial intelligence algorithms – naive Bayes classifier. The goal of research is to examine how this particular method works for this particular problem given a manually labeled news dataset and to support the idea of using artificial intelligence for fake news detection. The difference between these article and articles on the similar topics is that in this paper naive Bayes classifier was specifically used for fake news detection. [1]

The second research paper that named **Fake News Pattern Recognition using Linguistic Analysis.** In this paper, they propose a framework that can be adopted in future elections worldwide to augment humans in making better decisions when it comes to recognizing fake news and identifying hidden bias of the author. They constructed a dataset of 200 tweets on “Hilary Clinton”, while performing veracity assessment. They initially perform “text normalization” on tweets, explore techniques for feature extraction to classify news into categories, perform a comprehensive linguistic analysis on tweets, extract bag-of-words to find noticeable pattern, and finally apply k-nearest neighbor algorithm for classifying polarized news from credible. They later turn to quantify the success rate of their framework.[2]

The Third research paper that named **Fake News Detection.** This paper describes a method for fake news detection based on one of the machine learning algorithms – naïve Bayes classifier. The dataset was produced by GitHub, containing 11000 news article. Every news article labeled as "REAL" and "FAKE" were tagged as 1 and 0 respectively. The dataset was divided into two unequal subsets: training dataset, test dataset. The training dataset contains 75% while rest 25% was allotted to test dataset. While iterating through the words of the news articles that are being classified, a corner case is possible: some specific words might not be present in the training dataset at all. In such cases, the current implementation just ignores these words stating them as neither Real nor Fake.[3]

**Fake News Detection** 4 | P a g e

**CHAPTER 2**

**SYSTEM ANALYSIS**

*Chapter 2* *System Analysis*

**2. System Analysis**

The process of analyzing a procedure in order find the purpose and build systems and procedures that will achieve in accurate manner. It is also called problem solving technique.

It contains the implementation of final year project, defines the methods that are used for source code.

System Analysis contains following:

* Use Cases
* Actor requirements
* Problem Domain
* Requirements

**2.1 Use Cases:**

For the sake of understanding each component of the system, Use Case Diagram is used to explain the user interaction with the system.

1. Dataset
2. Pre-Processing
3. Filtering
4. Classifier

**2.2 Actors:**

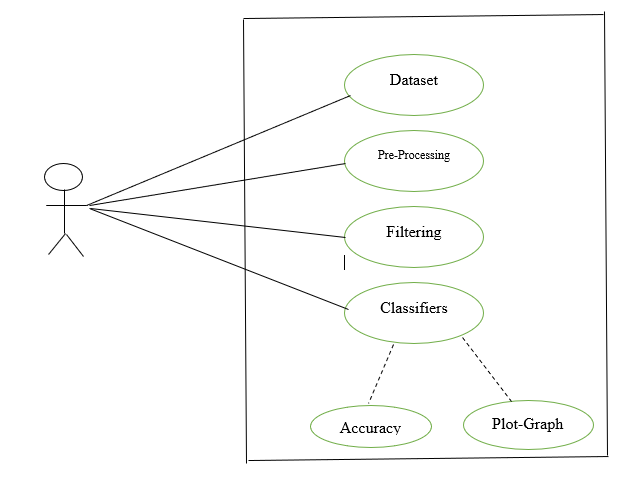
1. Dataset
2. Host

**2.3 Use Cases Diagram**

It is a graphical methodology that is used in a system analysis to find, manage and organize the project requirements. Use case diagram represents graphical explanation of the interactivity between components of the project.

**Fake News Detection** 6 | P a g e

*Chapter 2* *System Analysis*



**Figure 2. 1 Use case Diagram**

***Use Case Description***

In software engineering **Use Cases** consists of actions that define the interactions between different role and system to accomplish the goals. The actors can be an external system or computer user.

**2.4 Use case descriptions**

**2.4.1 Dataset**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Dataset** |  |  |
|  |  |  |
| **Use case ID:** | 1 |  |
|  |  |  |
| **Actor:** | Host |  |
|  |  |  |
| **Type:** | Primary |  |
|  |  |  |
| **Description** | Dataset about News in Text form. |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **Fake News Detection**  *Chapter 2*  **2.4.2 Pre Processing** | 7 | P a g e |  |
|  |  |  |
| **Pre-Processing** |  |  |
|  |  |  |
| **Use case ID:** | 2 |  |
|  |  |  |
| **Actor:** | Host |  |
|  |  |  |
| **Type:** | Primary |  |
|  |  |  |
| **Description** | In this section I remove hash tags, user name, extra white spaces, |  |
|  | Convert uppercase into lowercase etc |  |
|  |  |  |
| **2.4.3 Filtering** |  |  |
|  |  |  |
| **Filtering** |  |  |
|  |  |  |
| **Use case ID:** | 3 |  |
|  |  |  |
| **Actor:** | Host |  |
|  |  |  |
| **Type:** | Primary |  |
|  |  |  |
| **Description** | In filtering process filter raw words which are not important (an, on, |  |
|  | the, are, or, my) etc. |  |
|  |  |  |
|  |  |  |
| **2.4.4 Classifier** |  |  |
|  |  |  |
| **Classifier** |  |  |
|  |  |  |
| **Use case ID:** | 4 |  |
|  |  |  |
| **Actor:** | Host |  |
|  |  |  |
| **Type:** | Primary |  |
|  |  |  |
| **Description** | Check the news whether it is real or fake and find the |  |
|  | Accuracy |  |
|  |  |  |

**Fake News Detection** 8 | P a g e

*Chapter 2 System Analysis*

**2.5. Detailed Use Cases**

**2.5.1. Dataset**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Dataset** | |  |  |  |  |
|  |  | |  |  |  |  |
|  | **Use case ID:** | | 1 | |  |  |
|  |  | |  |  |  |  |
|  | **Actor:** | |  | Host |  |  |
|  |  | |  |  |  |  |
|  | **Purpose** | |  | Dataset that contain text related to news |  |  |
|  |  | |  |  |  |  |
|  | **Overview** | |  | I use dataset that contain news and saved as csv file |  |  |
|  |  | |  |  |  |  |
|  | **Type:** | |  | Primary |  |  |
|  |  | |  |  |  |  |
|  | **Pre-Conditions:** | |  | Need dataset |  |  |
|  |  | |  |  |  |  |
|  | **Typical course of Events** | |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  | **Actor Actions** |  | **System Response** |  |  |
|  |  |  |  |  |  |  |
|  | 1) | User runs the project |  | 2) System done the pre-processing |  |  |
|  |  |  |  |  |  |  |
|  | 3) | User give input |  | 4) Save the model into a file |  |  |
|  |  | |  |  |  |  |

**2.5.2. Pre-Processing**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  |  |  |
|  | **Pre-Processing** | |  |  |  |  |
|  |  | |  |  |  |  |
|  | **Use case ID:** | |  | 2 |  |  |
|  |  | |  |  | |  |
|  | **Actor:** | |  | Host | |  |
|  |  | |  |  | |  |
|  | **Purpose** | |  | Pre Process data of news articles | |  |
|  |  | |  |  | |  |
|  | **Overview** | |  | In pre-processing remove hash tags, username, extra white | |  |
|  |  |  |  | spaces, Convert uppercase into lowercase etc | |  |
|  |  | |  |  | |  |
|  | **Type:** | |  | Primary | |  |
|  |  | |  |  | |  |
|  | **Pre-Conditions:** | |  | Need Input from user | |  |
|  |  | |  |  |  |  |
|  | **Typical course of Events** | |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Actor Actions** |  | **System Response** |  |
|  |  |  |  |  |
| 1) | Open the csv file | 2) | Send these data to Pre-processed Function |  |
|  |  |  |  |  |
|  |  | 3) | One by One that function Pre-Process the data |  |
|  |  |  | |  |
| **Fake News Detection** 9 | P a g e  *Chapter 2*  **2.5.3. Filtering** | |  |  |  |
|  | |  |  |  |
| **Filtering** | |  |  |  |
|  | |  |  |  |
| **Use case ID:** | | 3 |  |  |
|  | |  | |  |
| **Actor:** | | Host | |  |
|  | |  | |  |
| **Purpose** | | Pre-Processed data is now filter in this section | |  |
|  | |  | |  |
| **Overview** | | I use Stop word dictionary to filter the pre-processed | |  |
|  |  | News articles data | |  |
|  | |  | |  |
| **Type:** | | Primary | |  |
|  | |  | |  |
| **Pre-Conditions:** | | Pre-Processing must be done | |  |
|  | |  |  |  |
| **Typical course of Events** | |  |  |  |
|  |  |  |  |  |
|  | **Actor Actions** |  | **System Response** |  |
| 1) | Pre-Processed Data | 2) | Send the data into Feature Vector Function |  |
|  |  |  |  |  |
|  |  | 3) | Compare the data with stop word dictionary one by one |  |
|  |  |  |  |  |

**2.5.4. Classification**

**Classification**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Use case ID:** | 4 |  |
|  |  |  |  |
|  | **Actor:** | Host , User Interface |  |
|  |  |  |  |
|  | **Purpose** | Classification and accuracy of data |  |
|  |  |  |  |
|  | **Overview** | Detect the fake news and check accuracy of the results |  |
|  |  |  |  |
|  | **Type:** | Primary |  |
|  |  |  |  |
|  | **Pre-Conditions:** | Filtering of data must be done |  |
|  |  |  |  |

**Typical course of Events**

|  |  |  |
| --- | --- | --- |
| **Actor Actions** |  | **System Response** |
| 1) Filtered Data | 2) | Apply the classification Algorithm |
|  |  |  |
|  | 3) | Classify and check whether the news is Real or fake. |
|  |  | |
|  | 4)After That check the accuracy of the result. | |
|  |  |  |
|  | 5) | Plot the graph which contain true and false areas |
|  |  |  |

**Fake News Detection** 10 | P a g e

*Chapter 2 System Analysis*

**2.1. System Sequence Diagrams**

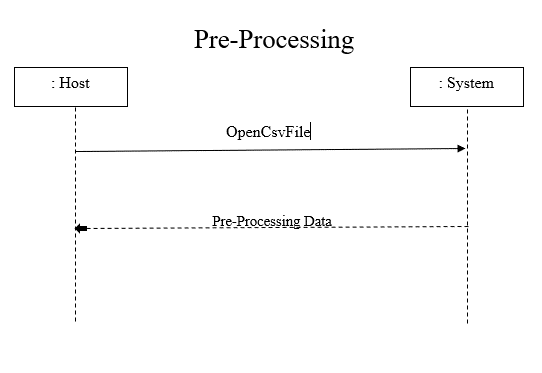
System sequence diagram is used in software engineering that displays scheme of

use cases, events, orders of events and internal system events. It also optical summaries

of every use cases. The system is behaved as a black box. The diagrams emphasize on

events that meet system boundary from actors to systems.

**2.6.1. Pre-Processing:**

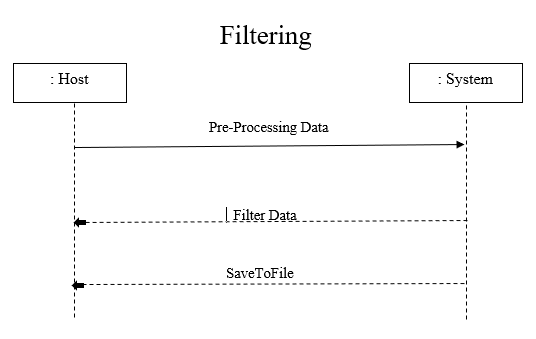


**Figure 2.2: Pre-processing System Sequence Diagram**

**Fake News Detection** 11 | P a g e

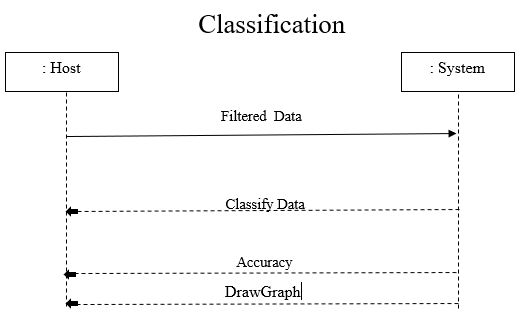
*Chapter 2* *System Analysis*

**2.6.2. Filtering**



**Figure 2.3: Filtering System Sequence Diagram**

**2.6.3.** **Classification**

****

**Figure 2.4: Classification System Sequence Diagram**

**Fake News Detection** 12 | P a g e

*Chapter 2* *System Analysis*

**2.7. Operation Contracts:**

It identifies the changes in system state when operation contrast is happening. Effectively, it will define what system operation does. The operations taken by system sequence diagram is a single event of that diagram.

**2.7.1.** **Dataset:**

A dataset that contains news data in text form.

**2.7.1.1.** **Datasetfilename**

|  |  |
| --- | --- |
| Name | **Dataset filename** |
|  |  |
| Responsibility |  |
|  |  |
| Type | System |
|  |  |
| Cross Reference | Use Case: Dataset |
|  |  |
| Precondition | Dataset file name must be correct and contain news in text form |
|  |  |
| Post Condition | Opened the correct Dataset file must be done |
|  |  |

**2.7.2. Pre-Processing:**

Pre-Processing in which I initially filtering the data (remove username, links, hashtags, unneeded languages and numerical values etc.).

**2.7.2.1.** **Pre-Processing(news):**

|  |  |
| --- | --- |
| Name | **Dataset filename** |
|  |  |
| Responsibility |  |
|  |  |
| Type | System |
|  |  |
| Cross Reference | Use Case: Pre-processing |
|  |  |
| Precondition | Dataset must be available |
|  |  |
| Post Condition | Pre-processing of Dataset must be done |
|  |  |

**Fake News Detection** 13 | P a g e

|  |  |  |  |
| --- | --- | --- | --- |
| *Chapter 2* |  | *System Analysis* |  |

**2.7.3. Filtering:**

Filtering in which I filtering the data by using stop words and break the whole sentences into single separated words.

|  |  |  |
| --- | --- | --- |
| **2.7.3.1.** | **Filtering:** | |
|  |  |  |
| Name |  | Filtering |
|  | |  |
| Responsibility | | Must be done filtering of pre processed data |
|  |  |  |
| Type |  | System |
|  | |  |
| Cross Reference | | Use Case: filtering |
|  | |  |
| Precondition | | Pre-processed data must be available. |
|  | |  |
| Post Condition | | filtering of pre-processed must be done a |
|  |  |  |

**2.7.5.** **Classification:**

Classification in which I use seven types of classifier (Naïve Bayes, Logistic Regression,SVM, Decision Tree, K-NN, Random Forest and Neural Network) to classify the news weather its real or fake and also gives us accuracy.

**2.7.5.1.** **Classifiers:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | Name |  | Classifiers (classifier , train\_set) |  |
|  |  | |  |  |
|  | Responsibility | | Classify the results |  |
|  |  |  |  |  |
|  | Type |  | System |  |
|  |  | |  |  |
|  | Cross Reference | | Use Case: classification |  |
|  |  | |  |  |
|  | Precondition | | Filtered data must be available |  |
|  |  | |  |  |
|  | Post Condition | | Accuracy of result must be done |  |
|  |  |  |  |  |

**2.7.5.3** DrawGraph

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | Name |  | drawGraph(size , explode, label , color) |  |
|  |  | |  |  |
|  | Responsibility | | Plot the graph according given values |  |
|  |  |  |  |  |
|  | Type |  | System |  |
|  |  | |  |  |
|  | Cross Reference | | Use Case: classification |  |
|  |  | |  |  |
|  | Precondition | | Requirement must be clear for plotting the graph |  |
|  |  |  |  |  |
|  | Post Condition |  | Graph must be displayed according to requirement |  |

**Fake News Detection** 14 | P a g e

*Chapter 2* *System Analysis*

**2.8 Domain model:**

A picture containing room, white, clock

Description automatically generated

**Figure 2.5 Domain Model Diagram**

**2.9. Activity Diagram**

It is a behavioral diagram in Unified Modeling Language which describes dynamic aspects of system. Essentially, It is like a flowchart, that represents flow from one activity to another activity. These diagrams used in Business Process Modeling. It also describe steps involved in a use case diagram. It can be sequential or concurrent flow is drawn from one operation to another operation.

A screenshot of a cell phone

Description automatically generated

**Figure 2.6 Activity Diagram**

**Fake News Detection** 15 | P a g e

**CHAPTER 3**

**METHODOLOGY and SYSTEM DESIGN**

*Chapter 3* *System Design*

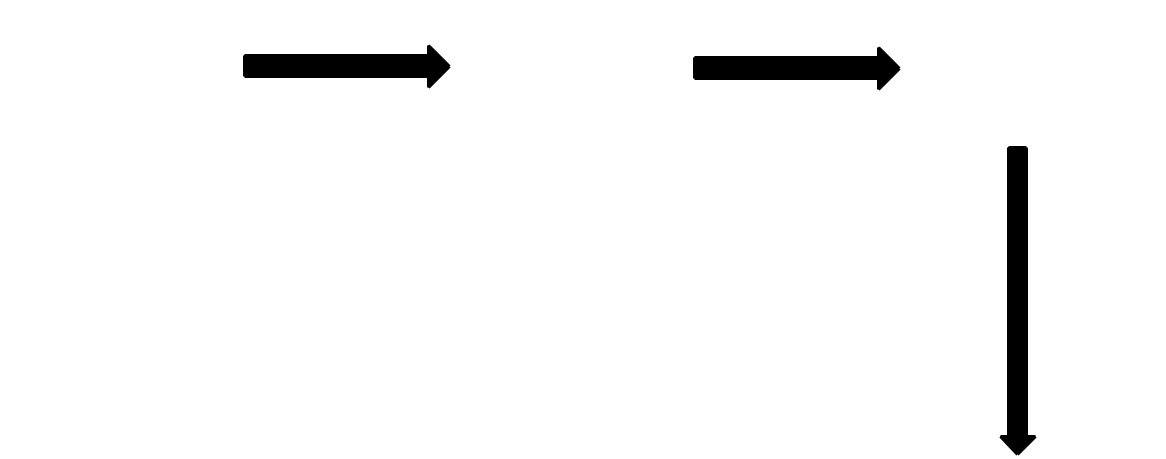
**3.1 Methodology**

I developed the following methodology for my final year project. It contains

6 different phases that is shown in Figure 1. i.e. Dataset, Pre-processing, Filtering,

Classification , Classify News, Accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **Dataset** |  | **Pre-** |  | **Filtering** |
|  | **Processing** |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | **Classification** |
|  |  | **Classify** |  |
| **Accuracy** |  |  |  |
|  | **News** |  |  |
|  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



**Figure 3.1: Methodology Diagram**

**3.1.1 Dataset:**

First I use fake news data in the form of text saved as CSV file. In my case I use **“**Articles**”.** Then save these articles.

**3.1.2 Pre-Processing:**

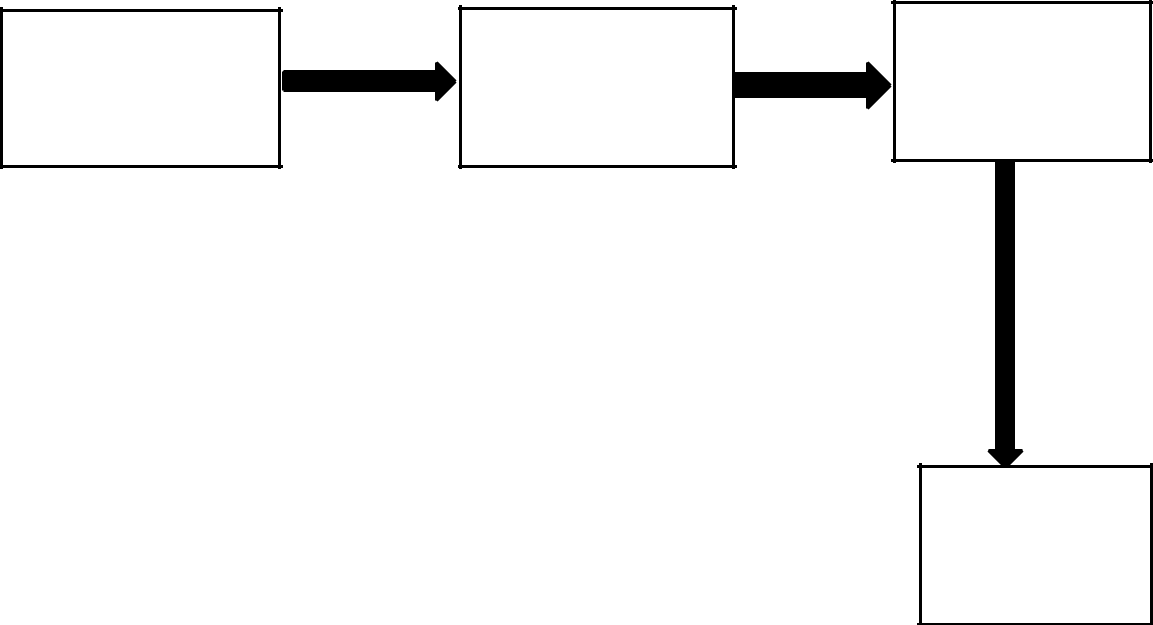
The process of converting data to something a computer can understand is referred to as pre-processing. For Pre-processing I open the csv file. That data contain no. of unnecessary things like user names, hashtags, extra white spaces, special character. I do not need these things, so I remove all these things in pre-processing.

**Fake News Detection** 17 | P a g e

*Chapter 3* *System Design*

**3.1.3 Filtering**

One of the major forms of pre-processing is to filter out useless data. In this section I filter the data. I use a stop word dictionary. And get the filtered data.



**Corpus**

**Pre-Processing** **Extract Features**

**Dictionary**

s

**Saved**

**Figure 3.2: Filtering Diagram**

**3.1.4 Classification**

In this section I classify the data. Classifier is used to analyse the data I can also check the accuracy through classifier. I use seven classifier Algorithms Naïve Bayes, Support Vector Machine, Random Forest, KNN, Logistic Regression, Decision Tree and Neural Network.

**3.1.5 Graph Plotting**

In this Section I plot the graph According to the results and accuracy got by classifiers.

**Fake News Detection** 18 | P a g e

*Chapter 3* *System Design*

**3.2 System Design**

It is a process of defining elements in a systems such that the architecture, modules and components, the different interfaces these component and date that passed to the system. It is needed to meet the specific needs and requirements of a business or organization by methods of a running and coherent system.

Systematic approach is required for good running systems. Designer used the modelling language to explain information and also the knowledge in structure of systems that is designed by consistent set of rules and definitions.

One can understand whole systems by understanding the design of that system. All internal as well as external functionality of a system design tells that how the project is going on. Each and every step is explained. Design ask users the requirement, quality of systems. It is basically a pictorial approach of a system. It help developers for understanding of work easily after seeing the design of systems.

Software architecture, describing the sub\_system decompositions in terms of sub\_system responsibilities, relationship among sub\_systems, major decisions such as data storage, access control and control flow.

It is transformation of analysis model to the System design model. That is designed to achieve objectives of a project. These are distributed in smaller components called the subsystems. Developer select strategies for creating systems, such as hardware and software platform on which systems will run, persistent data management strategy, global control flow, access control policy and handling of boundary conditions.

These diagrams are used here to illustrate the flow of the systems:

* Sequence diagram
* Class diagram

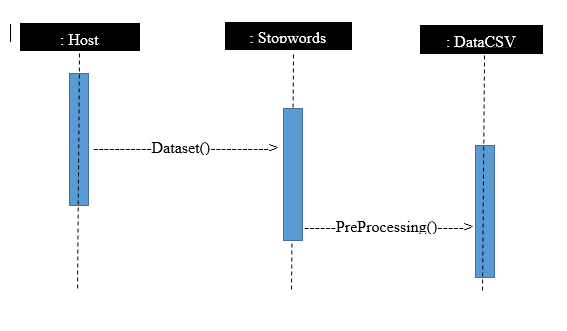
**Fake News Detection** 19 | P a g e

*Chapter 3* *System Design*

**3.2.1 Sequence diagram**

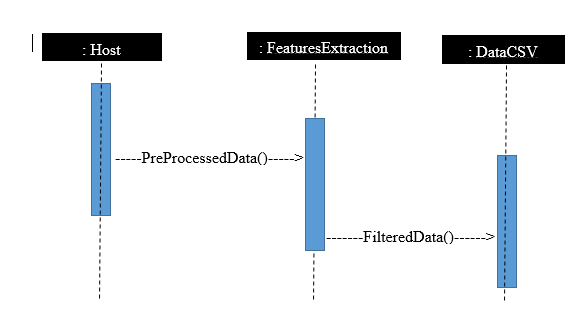
It shows object interactions arranged in time sequence. It depicts the object and class involved in the scenario and sequence of messages exchanged between object needed to carry out capabilityof the scenario.

**3.2.1.1 Pre Processing:**



**Figure 3.3: Sequence Diagram**

**3.2.1.2 Filtering:**

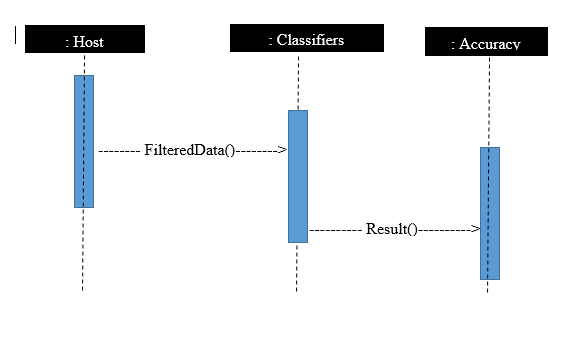


**Figure 3.4: Sequence Diagram**

**Fake News Detection** 20 | P a g e

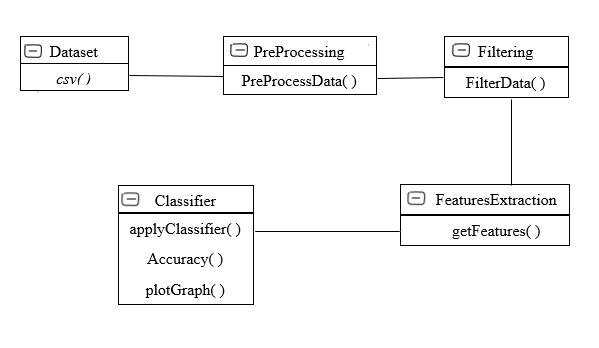
*Chapter 3* *System Design*

**3.2.1.3 Classifier:**



**Figure 3.5: Sequence Diagram**

**3.2.2 Class Diagram**

****

**Figure 3.6: Design Class Diagram**

**Fake News Detection** 21 | P a g e

**CHAPTER 4**

**EXPERIMENTAL RESULT & IMPLEMENTATION**

*Chapter 4* *Implementation*

**4.1 Experimental Results**

I detect the fake news based on existing labeled dataset about fake news. I appllied classifiers in order to predict the popularity. This can be done in different steps.

1. Data-set about fake news
2. Pre-Processing of data
3. Filtering of data
4. Accuracy can be checked by classifier

**4.1.1. Dataset:**

Dataset about fake news in text form.

**4.1.2 Pre-Processing:**

In this Section I pre-processed the data. That data contain no. of unnecessary things like user names, hashtags, extra white spaces, special character. I do not need these things, so I remove all these things in pre-processing.

**4.1.3 Filtering of Pre-processed data:**

After Pre-Processing the data contain like “I am a student and I like learning ”. so in this case I do not need “I”, “a” and “am ” words that is unnecessary words.

I need “student” and “learning”. So in filtering I remove that kind of unnecessary words.

I use stop words dictionary through which I make extract feature List and that List contain all the filtered data.

**4.1.4 Classifier:**

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known classifier. For example I use naïve Bayes classifier this classifier use Bayes theorem. First naïve Bayes classifier train the data set in my case real and fake news data. After training the dataset I find the accuracy on that training dataset.

**Fake News Detection** 23 | P a g e

**CHAPTER 5**

**SYSTEM TESTING**

*Chapter 5* *Software Testing*

**5.1.** **Introduction**

It is basically investigating that the project meets customers requirements. It helps to identify errors of project code which has to be fixed before delivering project to the consumers.

Testing consists of the following:

Unit Testing

Module Testing

Sub\_System Testing System Testing

Black Box Testing White Box Testing

**5.1.1. Unit Testing**

It is small\_scale testing by this software tester identifies syntax and logical errors which has to be removed by developers of that project. In this testing software tester test each unit separately.

**5.1.2. Module Testing**

A module consists of group of objects or classes. In this testing software tester test each module and identify errors if present in that module. At the completion module testing the identified errors and defects must be fixed.

**5.1.3. Sub – System Testing**

It means sub\_system also called a group of modules. This kind of testing done by the tester who tests group of modules of the project. The software tester performs sub\_system testing when the errors is in software due to interaction between components. Sub\_system testing should be done immediately as usable version of project is available.

**Fake News Detection** 25 | P a g e

*Chapter 5* *Software Testing*

**5.1.4. System Testing**

The Sub-Systems are merged together to create system. After the completion of software application it will be test here and known as system testing. Testing is done by tester to identify errors from interconnection in sub-systems. The tester insures that the application should meet the functionality and requirements in all aspects. After this tester check the validation and verification of the application software. In this testing all the components are reviewed. Occasionally, tester need the involvement of other application to test the software. Testing when to integrate components of other applications is called integration testing. Basically, tester done testing to check that developed application is like what consumer needs. There are two ways to perform system testing:

Black Box Testing White Box Testing

**5.1.5. Black Box Testing**

It is simple and initial testing in which software tester can test the application without knowing about the application. Tester did not require much knowledge to test application in black box testing. His goal is to check the outputs against to the inputs which he will enters to the application. In this testing tester is allowed only to interact with interface of the application and is not allowed to review the source code of application.

**Advantages**

* + Mostly used in large system testing.
  + Programming knowledge not required for tester.
  + With less knowledge about application One can perform testing.

**Disadvantages**

* + Most parts of system will be leaved without testing.
  + Test cases are difficult to design.
  + Not a complete testing.

**Fake News Detection** 26 | P a g e

*Chapter 5* *Software Testing*

**5.1.6. White Box Testing**

This testing is done when software tester done testing on the application internally. Some times, it named as inter testing and also called glass box testing. In glass box testing tester should have complete knowledge about the application such as programming knowledge its functionality and other things to test the application. Here software tester allowed to see the source code of the project and to review it. This testing includes to test each class, module, sub\_system and the whole code of the application.

**5.1.7. Acceptance Testing**

After the completion of project the final version of testing is done by tester to review the application and this is called acceptance testing. This testing is done with customer according to the requirements which help to test project accordingly. After this testing the project will be delivered to its consumer.

**5.2. Software Testing**

For software testing I done my work step by step. First step I take dataset of 20800 articles of news. Then I apply pre-processing on that file then filter the files. After that I applied classification for fake news detection.

**5.2.1 Test Cases**

Test Cases in which I test my system from beginning to end. It includes all the output results from all files.

**Fake News Detection** 27 | P a g e

*Chapter 5* *Software Testing*

**5.2.1.1 Dataset**



**Figure 5.1: News Data CSV FILE Diagram**

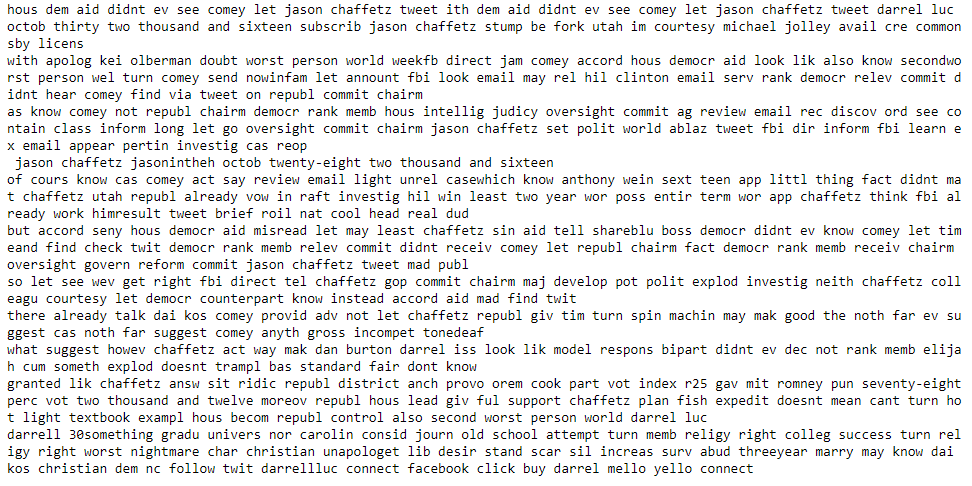
**5.2.1.2 Pre-Processing:**

Pre-Processing in which I initially filtering the data (remove username, links, hashtags, stop words etc.) that I have gets from twitter.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Case ID:** | TC-01 | **Test Case Name:** | | Pre-Processing |  |
|  |  |  | |  |  |
| **Date:28-04-2020** | | **Test by:** | | Abdul Wahab | |
|  | |  | |  |  |
| **Objective:** | | Filtering the data | | (remove username, links, hashtags, unneeded | |
|  |  | languages etc). | |  |  |
|  | |  | | |  |
| **Prerequisite:** | | Should be accessible dataset csv file. | | | |
|  | |  | |  |  |
| **Input:** | | **Expected response** | |  | **Response** |
|  | |  |  | |  |
| User should have csv | | 1. | Filtering the data. | | Pre-Process data |
| file. | | 2. | Remove username, links, | |  |
|  |  |  | hashtags, unneeded languages. | |  |
|  | |  | |  |  |
| **Result:** | | Pass | |  |  |
|  |  |  |  |  |  |

**Fake News Detection** 28 | P a g e

*Chapter 5* *Software Testing*



**Figure 5.2: Pre-Processed News Diagram**

**5.2.1.3 Filtering of the Data:**

Filtering in which I again filtering the data by using stop words and break the whole sentences into single separated words.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID:** | TC-02 | **Test Case Name:** | Filtering |  |
|  |  |  |  |  |
| **Date: 28-04-2020** | | **Test by:** | Abdul Wahab | |
|  | |  |  |  |
| **Objective:** | | Filter the data | and break the whole article into | |
|  |  | single separated words. | |  |
|  | |  | |  |
| **Prerequisite:** | | Admin should be accessible Pre-Processing files. | | |
|  | |  |  |  |
| **Input:** | | **Expected response** |  | **Response** |
|  | |  | |  |
| User should have Pre- | | 1. Filtering the data by using | | Finally clean the |
| Processing file. | | stop words files. |  | data and break the |
|  |  |  |  | whole article into |
|  |  |  |  | single separated |
|  |  |  |  | words and save it. |
|  |  |  |  |  |
|  | |  |  |  |
| **Result:** | | Pass |  |  |
|  |  |  |  |  |

**Fake News Detection** 29 | P a g e

*Chapter 5* *Software Testing*

**5.2.1.4 Classifier:**

Classification in which I use seven classifier Algorithms Naïve Bayes, Support Vector Machine, Random Forest, KNN, Logistic Regression, Decision Tree and Neural Network.

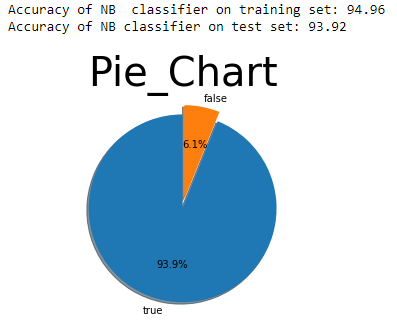
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID:** | TC-03 | **Test Case Name:** | Classifier |  |
|  |  |  |  |  |
| **Date: 28-04-2020** | | **Test by:** | Abdul Wahab | |
|  | |  |  |  |
| **Objective:** | | It check news articles | and give results and plot graph. | |
|  |  | It gives me accuracy. |  |  |
|  | |  | |  |
| **Prerequisite:** | | Admin should be accessible Filtering files. | | |
|  | |  |  |  |
| **Input:** | | **Expected response** |  | **Response** |
|  | |  | |  |
| User should have access | | 1. Checking the news articles if that is real or fake and find the accuracy. | | News articles gives |
| the filtering files. | |  | | Result whatever the |
|  |  |  | | Real or fake and |
|  |  |  |  |
|  |  |  |  | show the results |
|  |  |  |  | in graph form and |
|  |  |  |  | gives us accuracy of |
|  |  |  |  | the result. |
|  |  |  |  |  |
|  | |  |  |  |
| **Result:** | | Pass |  |  |
|  |  |  |  |  |

**Fake News Detection** 30 | P a g e

*Chapter 5* *Software Testing*

**5.2.1.5.1 Naïve Bayes:**

Primarily, it is Naïve Bayes classifiers because it makes guess that a feature is present or absent that corresponds to be correct.



**Figure 5.3: Naïve bayes Classifier Diagram**

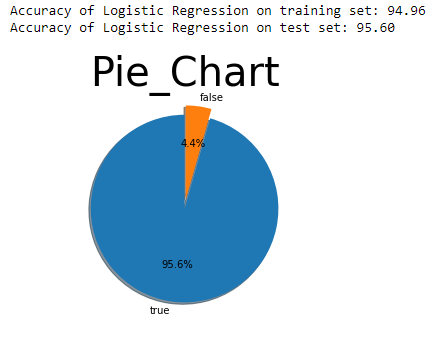
**Fake News Detection** 31 | P a g e

*Chapter 5* *Software Testing*

**5.2.1.5.2 Logistic Regression**

Logistic regression is a statistical technique used to predict probability of binary response based on one or more independent variables. It means that, given a certain factors, Logistic regression is used to predict an outcome which has two values such as 0 or 1,

Pass or fail, yes or no etc.

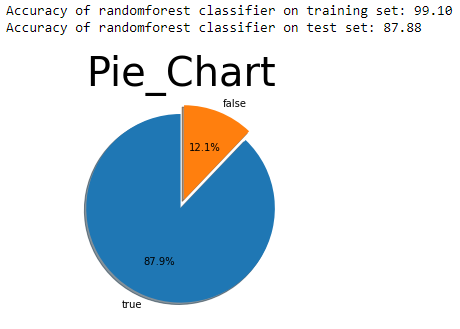


**Figure 5.4: Logistic Regression Diagram**

**Fake News Detection** 32 | P a g e

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Chapter 5* |  | *Software Testing* | | |
| **5.2.1.5.3 Random Forest:** |  |  |  |  |

Random forest consists of numerous dicision trees that uses feature randomness for building individual tree. It created uncorrelated forest of trees whose prediction is accurate.



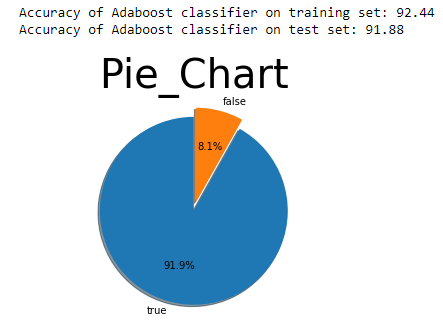
**Figure 5.5: Random Forest Classifier Diagram**

**Fake News Detection** 33 | P a g e

*Chapter 5* *Software Testing*

**5.2.1.5.4 Decision Tree**

Decision tree learning is a supervised machine learning technique for including a decision tree from training data. A decision tree (also referred to as a classification tree) is a predictive model which is a mapping from observation about its target value.

****

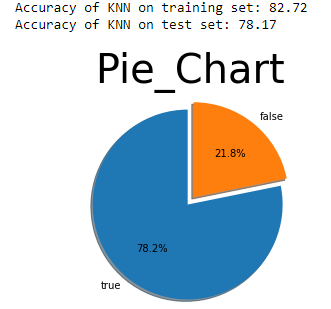
**Figure 5.6: Decision Tree Diagram**

**Fake News Detection** 34 | P a g e

*Chapter 5 Software Testing*

**5.2.1.5.5 K-Nearest Neighbor**

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.



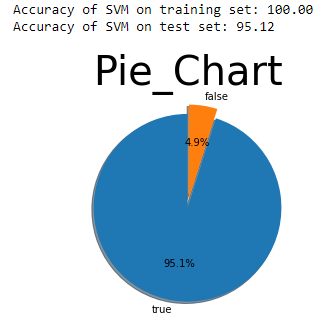
**Figure 5.7: KNN Classifier Diagram**

**Fake News Detection** 35 | P a g e

*Chapter 5 Software Testing*

**5.2.1.5.6 Support Vector Machine(SVM):**

The objective of SVM classifier is to find a line in N-dimentional space that classifies data.

**

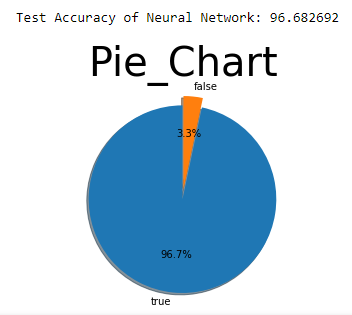
**Figure 5.8: SVM Classifier Diagram**

**Fake News Detection** 36 | P a g e

*Chapter 5 Software Testing*

**5.2.1.5.7 Neural Networks**

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. Thus a neural network is either a biological neural network , made up of real biological neurons, or an artificial neural network, for solving Artificial Intelligence problems.

**

**Figure 5.9: Neural Network Diagram**

**Fake News Detection** 37 | P a g e

*Chapter 6* *Conclusion*

**CHAPTER 06**

**CONCLUSION**

*Chapter 6* *Conclusion*

**6. Conclusion**

I implemented classification models for text related data about fake news detection. Data about fake news is passed to different classifiers and check about fake news or real news. The classifiers I used in this project are Naïve Bayes, Logistic Regression, Sport Vector-Machine, Random Forest, Decision Tree, k-nearest Neighbor and Neural Network and through these classifiers I get accuracy. In this way I detect fake news.

**6.1 Future Enhancements:**

1. Accuracy increase by using different approach.
2. This is only a detection model for fake news it can be enhanced to detect and delete such kind of news on social media.

**Fake News Detection** 39 | P a g e

**REFERENCES**

*Chapter 7 References*

**1. Fake news detection using naïve Bayes classifier**

Link: <https://ieeexplore.ieee.org/document/8100379>

1. **Fake News Pattern Recognition using Linguistic Analysis**

Link: <https://ieeexplore.ieee.org/document/8641018>

**3. Fake News Detection**

Link: <https://ieeexplore.ieee.org/document/8546944>

**Fake News Detection** 41 | P a g e