Fake News Detection

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Computer Engineering

by

Mr. Mann Arish Bajaria (08)

Mr. Parth Bhadresh Bhatt (15)

Mr. Shubham Divyeshkumar Bhatt (16)

Supervisor:

Mrs. Kanchan Dabre

Department of Computer Engineering
VIDYA VIKAS EDUCATION TRUST'S
UNIVERSAL COLLEGE OF ENGINEERING
KAMAN, VASAI-401202
UNIVERSITY OF MUMBAI

2019-2020

Vidya Vikas Education Trust's

Vidya Vikas Education Trust's Universal College of Engineering,Vasai (E) Department of Computer Engineering

CERTIFICATE

This is to certify that, **Mr.Mann Arish Bajaria (08), Mr.Shubham Divyeshkumar Bhatt (16) and Mr.Parth Bhadresh Bhatt (15)** has successfully completed their project on the topic **"Fake News Detection"** at Universal College of Engineering, Vasai, Mumbai towards the partial fulfillment of the Graduate Degree course in B.E. Computer Engineering at the department of Computer Engineering, in the academic Year 2019-2020, Semester – VII as prescribed by the Mumbai University.

Prof. Kanchan Dabre

Internal Examiner **External Examiner**

Dr.Jitendra Satturwar Dr. J.B. Patil

Head of Department **Principal**

Project Report Approval for B.E.

This project report entitled **Fake News Detection** by **Parth Bhadresh Bhatt, Shubham Divyeshkumar Bhatt, Mann Arish Bajaria** is approved for the degree of **Bachelor of Computer Engineering**.

Examiner:

1.- -

2..- -

Date:

Place: Vasai

ABSTRACT

The expansion of deceiving data into regularly viewing news sources, such as internet-based life networks, news websites, and online journals, has made it possible to discern reliable news sources, thus increasing the need for computer equipment ready to give bits of information into the unwavering nature of online content. Right now, the proposed framework center around the programmed recognizable proof of false substance in online news. The research incorporates studies with various algorithms of machine learning such as Naive Bayes, SVM, Logistic Regression, Random Forest, SGD with varying sets of features. The task is isolated into two sections. First, by presenting a dataset for the undertaking of phony news discovery, the dataset is demographically blended composed of the U.S. and the Indian region. The Indian region is just considering its training space in the education sector and present many exploratory analyses to recognize linguistic discrepancies in the content of false and legitimate news. Second, the project contribution is to provide the probability of different algorithms for the given news headline. It also offers input in the form of image.

Keywords: Support Vector Machine (SVM); Machine Learning (ML); Stochastic Gradient Descent (SGD); Natural Language Processing (NLP), Term frequency–Inverse Document Frequency (TF-IDF),Optical Character Recognition (OCR).

TABLE OF CONTENT

List of Figure

List of Tables

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Mann Arish Bajaria (CS 08)

Parth Bhadresh Bhatt (CS 15)

Shubham Divyeshkumar Bhatt (CS 16)

Date:

Chapter 1
Introduction

Fake news detection has recently attracted a growing interest from the public and researchers as the circulation of miss information on-line increases, particularly in media outlets such as social media feeds, news blogs, and online news-papers. For example, ongoing report by the Jump-shot Tech Blog1 found that Facebook referrals represented half of the absolute traffic to counterfeit news destinations and 20% all out traffic grown-ups 62% gets news via web-based networking media (Jeffrey and Elisa,2016), having the option to distinguish phony substance in online sources is a squeezing need. To date, computational methodologies for counterfeit news location has depended on ironical news sources, for example, “The Onion” and truth checking sites, for example, “PolitiFact” and “Snopes". How-ever, the utilization of these sources represent a few difficulties and potential disadvantages. For instance, using mocking substance as a hotspot for phony substance can carry basic jumbling factors into the examination, for example, silliness and absurdity. This is especially the situation for sarcastic news from “The Onion ”, which has been utilized in the past to investigate other content properties, for example, humor (Mihalcea and Strapparava,2005) and irony(Wallace,2015). On the other hand, fact-checking websites are usually constrained to a particular domain of interest, such as politics, and require human expertise, thus making it difficult to obtain dataset that provide some degree of generalization over several domains. In this paper, the project is developed from computational resources and

models for the task of fake news detection. Proposed model present the construction of two novel dataset covering seven different domains. One of the datasets is collected using a combination of a manual and crowd sourced annotation efforts, while the second is collected directly from the web. Using this dataset, the proposed model conduct several exploratory analyses to identify linguistic properties that are predominantly present in fake content, and the proposed model is relying on linguistic features that achieve accuracy of up to 78%. To place the results in perspective, comparison is done between the accuracy of proposed model with an empirical human baseline accuracy.

1.1 Project Overview

Fake news or false stories which spread as a rumor sometimes tends to cause an uproar by creating psychological or emotional trauma,political disturbance. The fake news detectors are humans themselves and take plenty of time to detect whether the article is hoax or true. The project aims to provide fully automated detection procedure with more efficiency then the traditional approach.The general rule is to gain something sacrificing something else is necessary and the same has happened with the project it will efficiently detect whether the news is fake or fact at the cost of speed. The goal of the project is to create social awareness about the fake news which are trending on that particular time span.

Chapter 2
Review of literature

2.1 Survey Existing System

In the existing model Conroy, Rubin, and Chen outline several approaches that seem promising toward the aim of correctly classifying misleading articles. They note that simple content-related ngrams and shallow part-of- speech (POS) tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora.

Paper Name	Year Of Publication	Author	Publication	Proposed Work
Evaluating Machine learning Algorithms For Fake News Detection.	2017	Shlok Gilda	IEEE	Detectin Fake News using Naïve bayes classifier
Detecting Fake News With Machine Learning Method	2018	Supanya Aphiwongsophon ,Prabhas Chongstit vatana	IEEE	Detecting Fake News using NN and Support Vector machine classifier.
Fake News Detection	2018	Akshay Jain Amey Kasbe	IEEE	Detecting Fake News using Naïve bayes classifier

Table 2.1 Literature survey

The above table provides the literature survey done for improving the existing model and creating a new proposed, the data provided in the table serves as the important inputs for the features which are necessary for improving the proposed system and including various machine learning algorithms which is useful for encorporating these algorithms for improving the efficiency of the proposed system.

2.2 Limitation in Existing system or research gap

Parameters	Existing system	Limitation
Methodology used to detect fake news	The concept of PCFG i.e probabilistic context free grammar with deep syntax analysis was used to classify the news.	Although it seemed promising at the beginning, it was proven insufficient to classify just by using simple content related programs and shallow parts of speech (POS).
Accuracy in classification	The existing system has accuracy in the range of 60-70%.	There is still a gap of 30-40% which can be critical and needs to be solved
Self-learning	The existing system cannot learn from new experiences to increase efficiency.	Thus unlike machine learning algorithms, the existing system can work in fixed context only

Table 2.2 Literature survey Limitations

The limitations in the existing system are the first step for making improvement in new system. The parameters help to undestand the specific features which has room for improvement which is helpful in making the necessary changes.

2.3 Problem Statement and Objective

Fake news or false stories which spread as a rumor sometimes tends to cause an uproar by creating psychological or emotional trauma, political disturbance. The fake news detectors are humans themselves and take plenty of time to detect whether the article is hoax or true. The project aims to provide fully automated detection procedure with more efficiency then the traditional approach. The general rule is to gain something sacrificing something else is necessary and the same has happened with the project it will efficiently detect whether the news is fake or fact at the cost of speed. The goal of the project is to create social awareness about the fake news which are trending on that particular time span.

2.4 Scope

The scope of the project will be to efficiently and smoothly incorporate machine learning algorithm to easily and accurately classify news as legit or fake

Chapter 3

PROPOSED SYSTEM

This chapter includes a brief description of the proposed system and explores the different modules involved along with the various models through which this system is understood and represented.

3.1 Algorithm

The algorithm used in proposed system is Naïve Bayes.Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem . It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. everypair of features beingclassified is independent of each other. It is one of the most commonly used algorithm for text classification. It is simple to use but it is quite effective.

3.2 Details of Hardware and Software

Certain hardware components or other software resources to be present on the system so that it can be used efficiently. These prerequisites are known as system requirement and are often used as a guideline as opposed to an absolute rule.The Software that are required for the project are shown in the above Table 3.1 it shows what requirements are used while developing the project.

3.2.1 Software Requirments

To use system efficiently following hardware requirements has to be maintained. The proposed systems software requirement are shown in the Table 3.1

Operating System	Windows 7,8,10
Core Language	Python.
Software Used	Spyder,Mongodb.

Table 3.1 Software Requirements

3.2.2 Hardware Requirments

Device	Laptop or Personal Computer
Harddisk	Minimum 160Gb HDD.
Processor	Intel (R) Core(TM) i3
Ram	4GB

Table 3.2 Hardware Requirment

The proposed system's hardware requirement are shown in the Table 3.1. To be used efficiently, all computer software needs certain hardware components to be present on a computer the minimum hardware requirements to run the system consists of atleast 4gb of ram,160gb of hard drive etc as shown in table below.

3.3 Design Details

A system architecture or systems architecture is the conceptual model that denes the structure,behavior,and more views of a system. An architecture

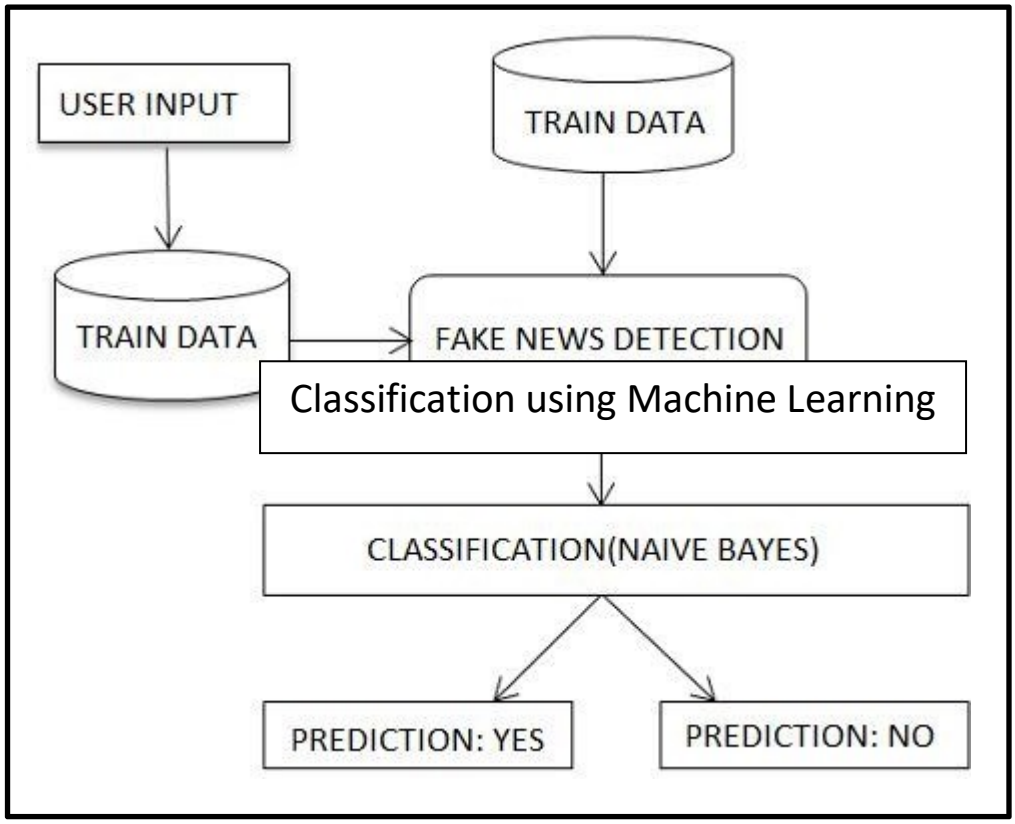

Figure 3.1 System Architecture

organized in a way that supports reasoning about the structures and behaviors of the system. Figure 3.1 elaborates the system architecture of the proposed system.The fundamental organization of a system, embodied in its components, their relationships to each other and to the environment, and the principles governing its design and evolution A formal description of a system, or a detailed plan of the system at component level to guide its implementation.The structure of components, the interrelationships, and the principles and guidelines governing their design and evolution over time.

3.4 Methodology

Methodology of the proposed system is the form of representation of working of the model. It divides the flow of the proposed model into three parts.The first phase being Input data,the second phase being classifier and the third phase being result.

Figure 3.2 Three phase detail

- In First Phase, the proposed model accept the input from the user from the GUI.
- In Second Phase,the proposed model will apply the machine learning algorithm to classify into fake or true data.
- In Third Phase,the proposed model display the output based on the classification.

3.4.1 Fundamental Model

Fundamental modeling concepts (FMC) provide a framework to describe software-intensive systems.It strongly emphasizes the communication about software-intensive systems by using a semi-formal graphical notation that can easily be understood.

3.4.2 Data Flow Model

Data Flow Diagram (DFD) provides a visual representation of the flow of information (i.e. data)with in a system as shown below. DFD is a graphical representation of the flow of data through an information system,modelling its process aspects. By drawing a Data Flow Diagram,it can be said the information provided by and delivered to someone who takes part in system processes,the information needed in order to complete the processes and the information needed to be stored and accessed. It is a preliminary step to create an overview of the

system. DFD is one of the three essential components of the structured systems analysis and design method. The Figure 3.3,Figure 3.4 and Figure 3.5 shows the levels of DFD in the proposed system.

DFD LEVEL 0

DFD Level 0 is also called as context diagram.

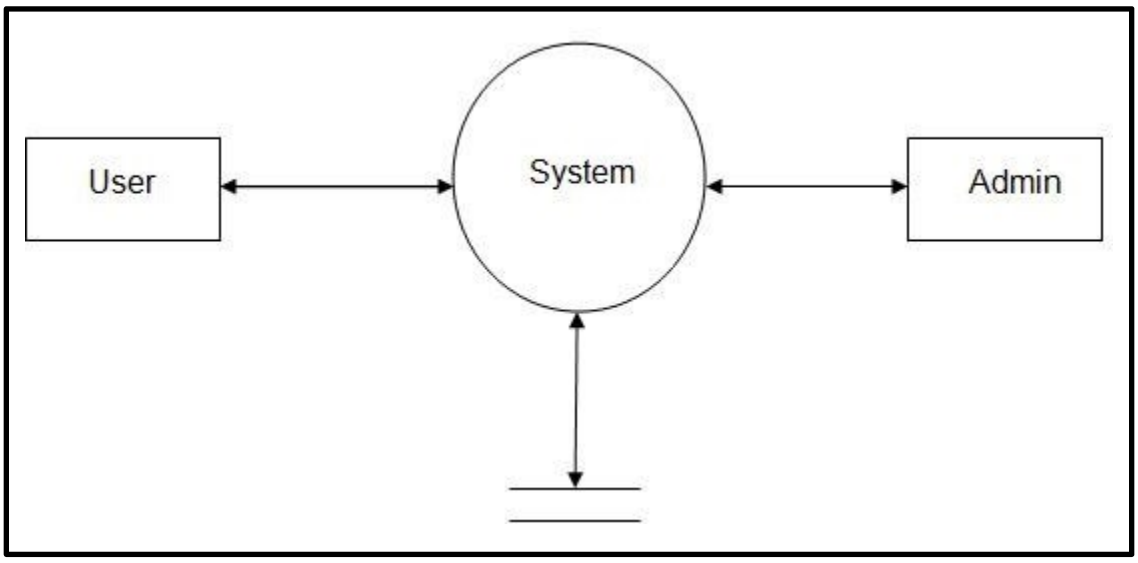

Figure 3.3 DFD Level 0

From the above Figure, in DFD Level 0 both the user and Admin can simultaneously access the system .User access the System to check the status news where as Admin access the system to enter the datasets and checks the history of the news.

A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. It does not show information about the timing of process or information about whether processes will operate in sequence or in parallel unlike a flowchart which also shows this information. DFD level 0 states that both user and admin can access the system simultaneously.

DFD LEVEL 1

The DFD level 1 is derived from the context diagram which is DFD 0.Here the models not only basic working of the model which is referred from DFD

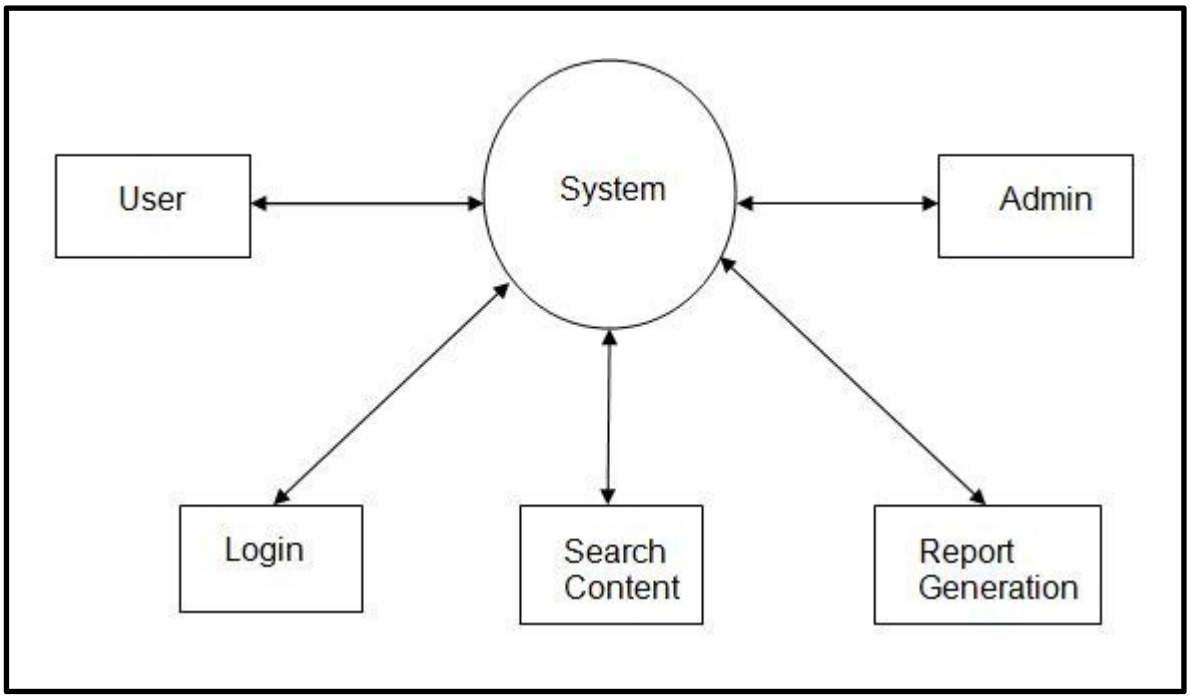

Figure 3.4 DFD Level 1

When constructing a Level 1 DFD, it must start by examining the Context Level DFD. The break up of the single process into its sub-processes. It must then pick out the data stores from the text given and include them in DFD. Like the Context Level DFD's, allentities, data stores and processes must be labelled. It must also state any assumptions made from the text. The next stage is to create the Level 1 Data Flow Diagram. This highlights the main functions carried out by the system produce a Level 1 DFD that shows some of the detail of the system being modeled.

The Level 1 DFD shows how the system is divided into subsystems(processes), like user can login, search content and report content each of which deals with one or more of the data flows to or from an external agent, and which together provide all of the functionality of the system as a whole. The proposed model consist of a user, system, admin which will take the key part in execution. The user will interact with system for login, searching content and

sharing the report. The admin can login and validate the report generated by the system.

DFD LEVEL 2

In DFD level 2 more emphasis is given to the overall process of the system. Detailed procedure is done and each and every entity is counted like to Authenticate user, to check database save report etc.More detailed view is provided to the user with all the procedures being involved where a user logs into the system and how the admin grants access to the required update is shown.The DFD level 2 shows the overall design of the system by combining the level 0 and level 1 diagram it is easiest diagram to design it shows the structure same as level 1 but only some processes are added to it.detailed process of designing of system is done in the level 2 diagram. This same process can be applied to each process appearing within a level 1 DFD. A DFD that represents a decomposed level 1 DFD is called a level 2 DFD. There can be a level 2 DFD for each process that appears in the level 1 DFD.

The DFD level 2 Shows that Users as well as Admin can Simultaneously access the system. The System contains 3 important components and 3 sub components.3 Components are Admin,Search content and report Generation which are in DFD level 1 .3 Sub component in DFD level 2 are Authenticate User. This component Authenticates Registered user from the existing database. Database and Report generation are another two sub-components. These components simply Checks the database for contents to be searched and saves the searched history in a report which can be accessed by the User or Admins.

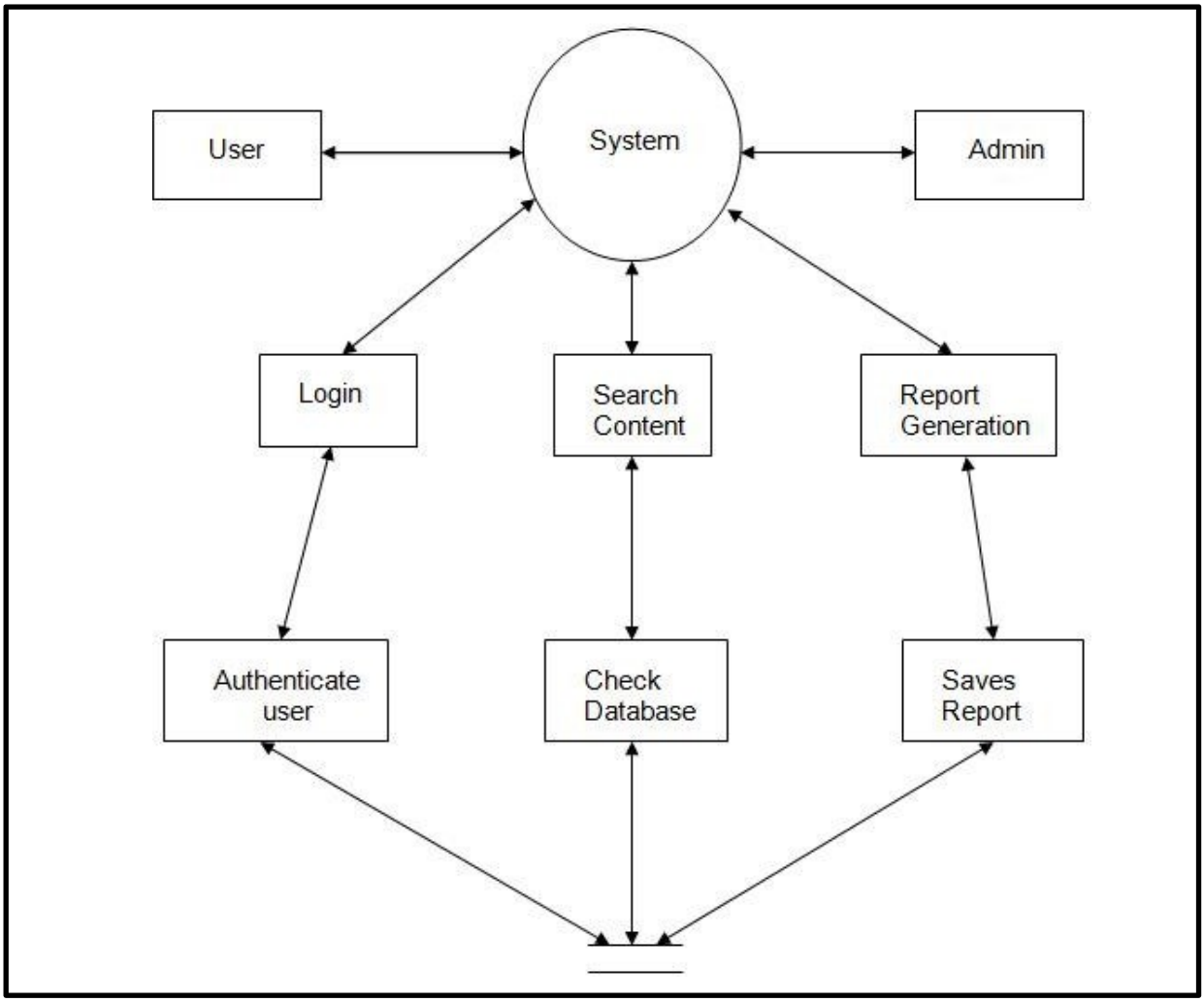

Figure 3.5 DFD 2

3.5 Data Model and Description

A data model is an abstract model that organizes elements of data and standardizes how they relate to one another and to properties of the real world entities. The term data model is used in two distinct but closely related senses. Sometimes it refers to an abstract formalization of the objects and relationships found in a particular application domain, for example the customers, products, and orders found in a manufacturing organization. At other times it refers to a set of concepts used in defining such formalizations: for example concepts such as entities, attributes,

relations, or tables. So the "data model" of a banking application may be defined using the entity-relationship "data model".A data model explicitly determines the structure of data. Data models are specified in a data modeling notation, which is often graphical in form.A data model can sometimes be referred to as a data structure, especially in the context of programming languages. Data models are often complemented by function models, especially in the context of enterprise models.

3.5.1 Entity Relationship Model

An ER model is commonly formed to represent things a business needs to model becomes an abstract data model, that defines a data or information structure which can be implemented in a database, typically a relational database.An ER model is typically implemented as a database. In a simple relational database implementation, each row of a table represents one instance of an entity type, and each field in a table represents an attribute type. In a relational database a relationship between entities is implemented by storing the primary key of one entity as a pointer or "foreign key" in the table of another entit

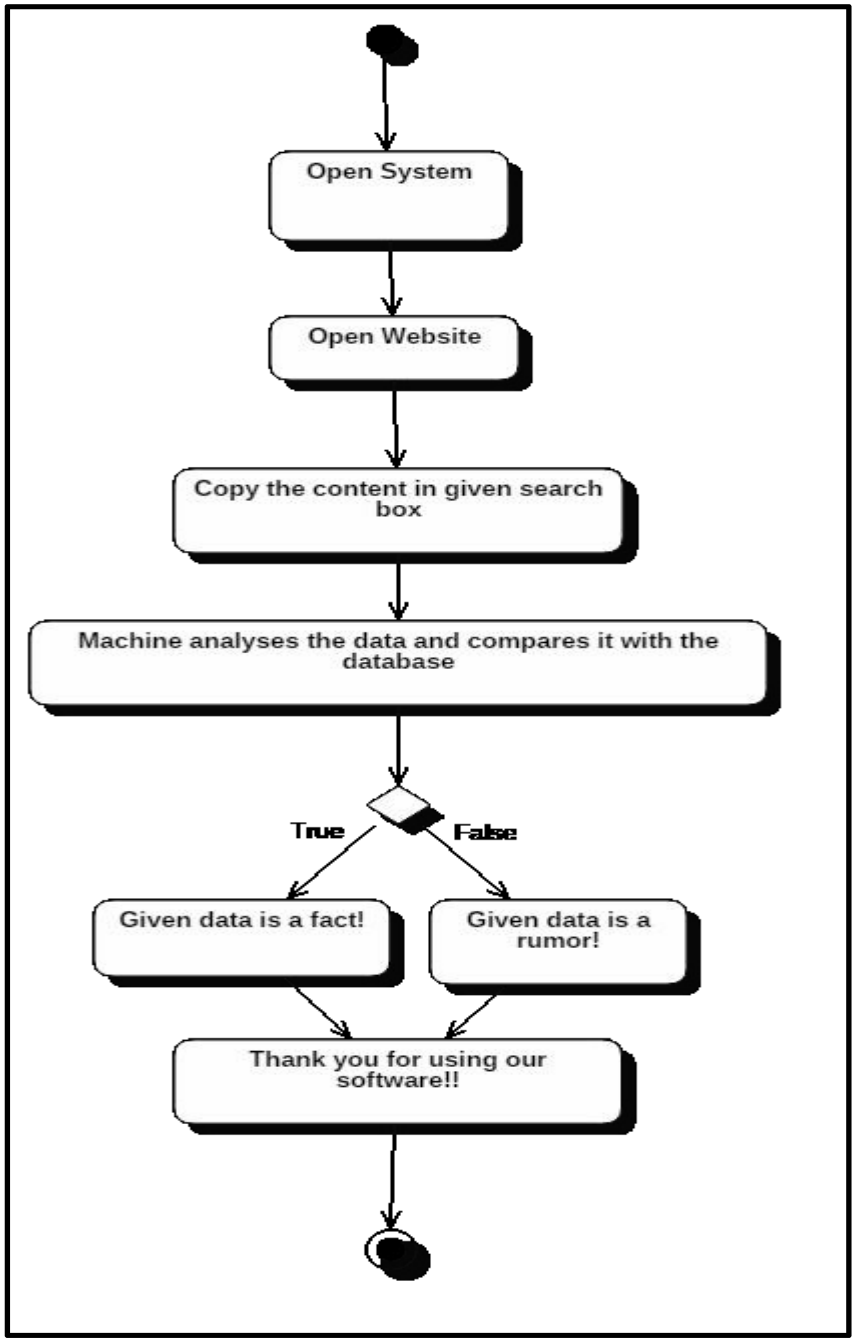

Figure 3.6 ER Model

3.6 Use Case Diagram

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases.

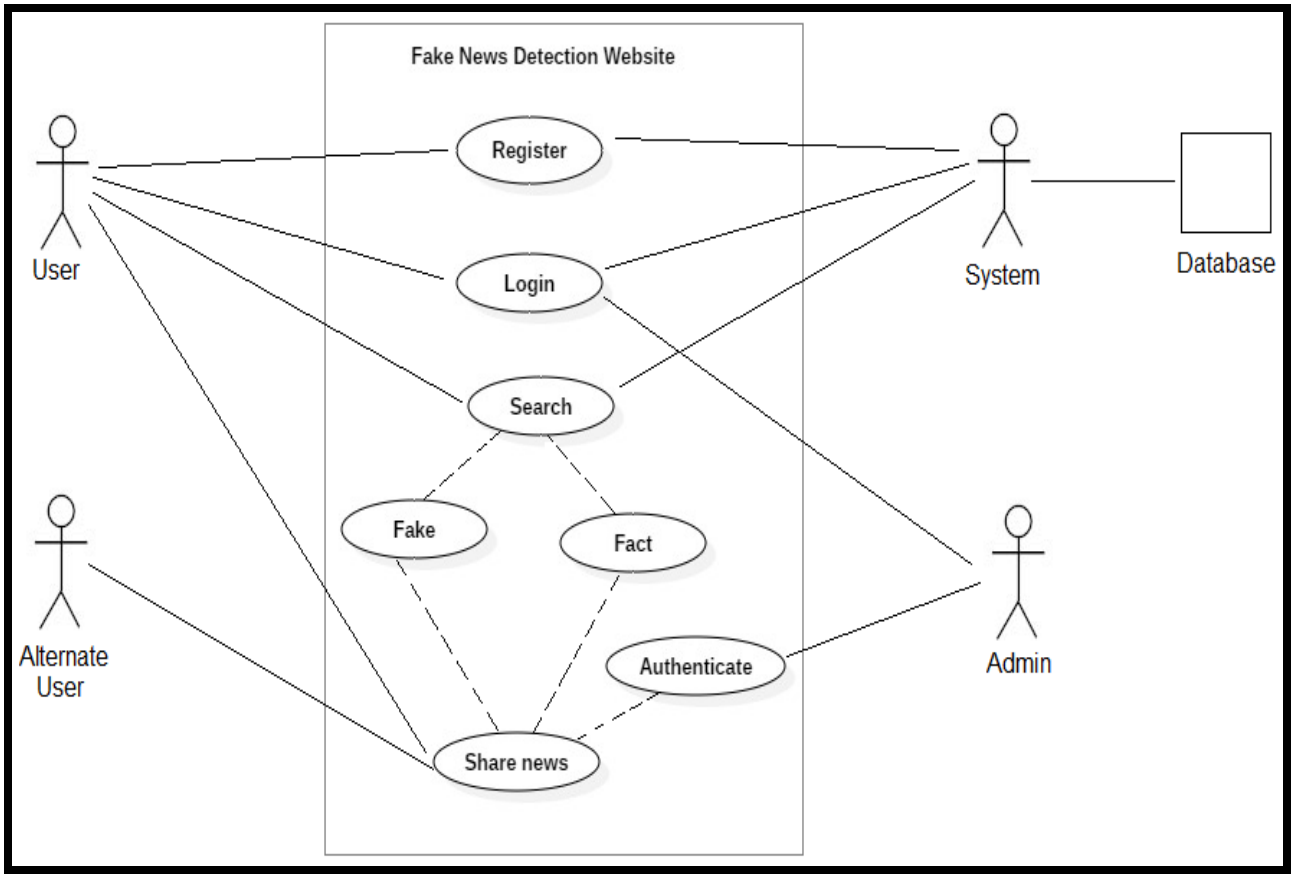

Figure 3.7 Use Case Diagram

In this proposed system,Figure 3.7 shows that actors include user and administrator. Others enclosed in the ovals are use cases.Use case diagrams are usually referred to as behaviour diagrams used to describe a set of actions(use cases) that some system or systems(subject) should or can perform in collaboration diagram with one or more external users of the system(actors).The use case diagram acts in user level and admin level.The user acts and does the process of what user wants and then it goes to admin for particular results.

3.6.1 Sequence Diagram

A Sequence diagram is an interaction diagram that shows how objects operate with one another and in what order.It is a construct of a message sequence chart.A sequence diagram shows object interactions arranged in time sequence

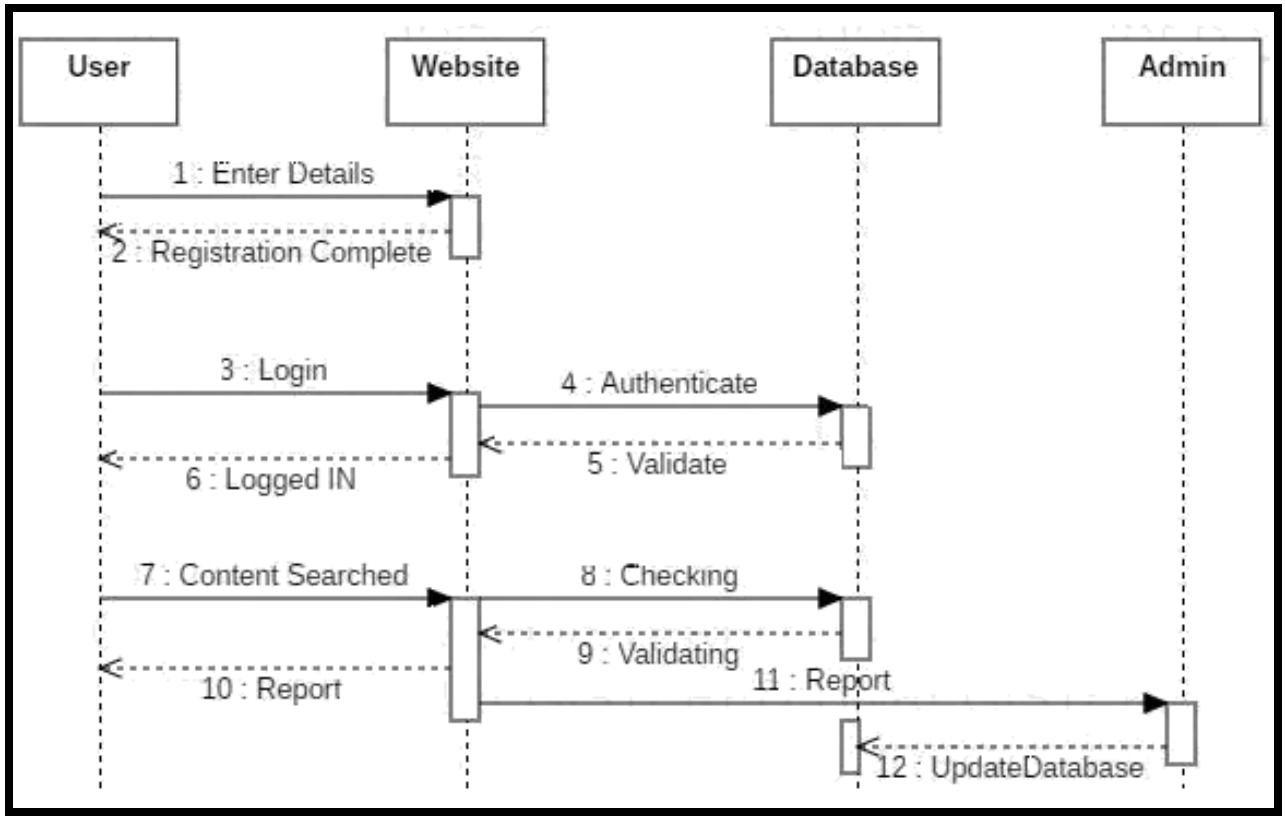

Figure 3.8 Sequence Diagram

It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur.The Figure 3.8 shows the sequence diagram which is good diagram to use to document a system's requirements and to flush out a system's design. The reason the sequence diagram is so useful is because it shows the interaction logic between the objects in the system in the time order that the interactions take place.

Chapter 4
Result

4.1 Algorithm.

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem .It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.It is one of the most commonly used algorithm for text classification. It is simple to use but it is quite effective.Along with Naive Bayes we are incorporating random search algorithm, Support Vector Machine, Logistic Regression, Stochastic Gradient Descent and Natural language Processing.

4.2 Steps in algorithm.

The proposed model will work on 5 major steps:-

1. Accept the input.
2. Create Embedding Matrix.
3. Check the existing data sets.
4. Evaluate the matrix using Naïve bayes algorithm and other methods.
5. Check the result.

4.2.1. ACCEPT THE INPUT

User needs to create an account and login to the website.After logging in user can give their input in three forms:-

1}Text:- Accept the input from the user in the form of text, paragraphs, links, etc. in the search box.

2}Audio:- The input can also be in the form of voice shown in Figure 3 which is converted to text using text to speech one of the fundamentals of NLP.The Fig 4.1 shows the propose model working using the website where the users input a headline of the article

Figure 4.1 Accept Input

3}Visual:- The Figure 4.2 depicts the upload step to take input in the form of visual representation. The input is in the form of an image that needs to be uploaded from the local storage of device and submitted for next step as shown in Figure 4.2

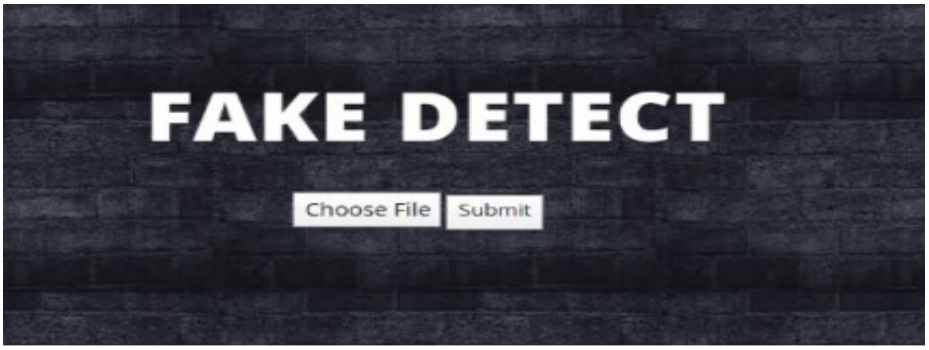

Figure 4.2 Input Acceptance via Image

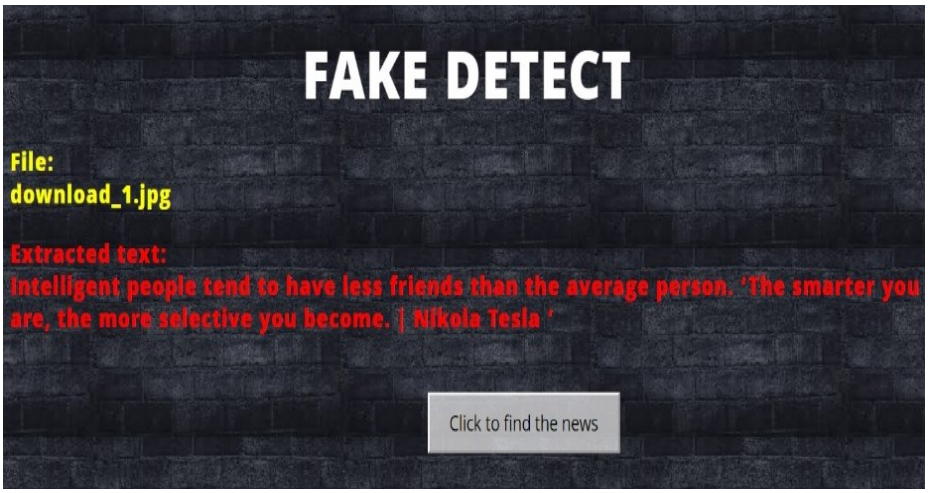

Figure 4.3 Image Result

The input image is converted to text using Pytesseract OCR as shown in Figure 4.3. Here, the image consisted of the famous quote from Nikola Tesla stating “Intelligent people tend to have less friends than the average person. The smarter you are, the more selective you become.” After the text conversion system proceeds to creation of the embedded matrix

4.2.2. Create an Embedding matrix:

Embedding Matrix is used to convert words in to mathematical format.This is done using natural language processing. For example: Let us consider a following string which is converted into Embedding matrix**.**

In the figure 4.4 Embedding Matrix the words of sentence are converted to numeric values which help in the further processing of the model.

$$X = \begin{array}{c} \\ I \\ like \\ enjoy \\ deep \\ learning \\ NLP \\ flying \\ . \end{array} \begin{array}{c} \begin{array}{cccccccc} I & like & enjoy & deep & learning & NLP & flying & . \end{array} \\ \begin{bmatrix} 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Figure 4.4 Embedding Matrix

4.2.3. Check the existing data sets.

class	videos	shots	frames
aero	13	1097	71327
bird	16	205	27532
boat	17	606	74501
car	9	208	14129
cat	21	220	42785

Figure 4.5 Data sets

Data set(dataset) is a collection of data. A data set is organized into some type of data structure. Most commonly a data set corresponds to the contents of a single database table.

4.2.4 Evaluate Matrix Using Various Algorithms.

The proposed model will use Naive Bayes, SVM, SGD, Logistic Regression, Random Forest classifier algorithm to classify the news into fake or true and provide the probabilities.

All the below graphs show the learning curve of average training and test scores for various Classifiers where the color Bright Turquoise indicates training and Violet indicates the Cross-Validation score. Here, only 2,555 datasets were included for plotting purposes but as more datasets are included the accuracy of the model increases significantly.

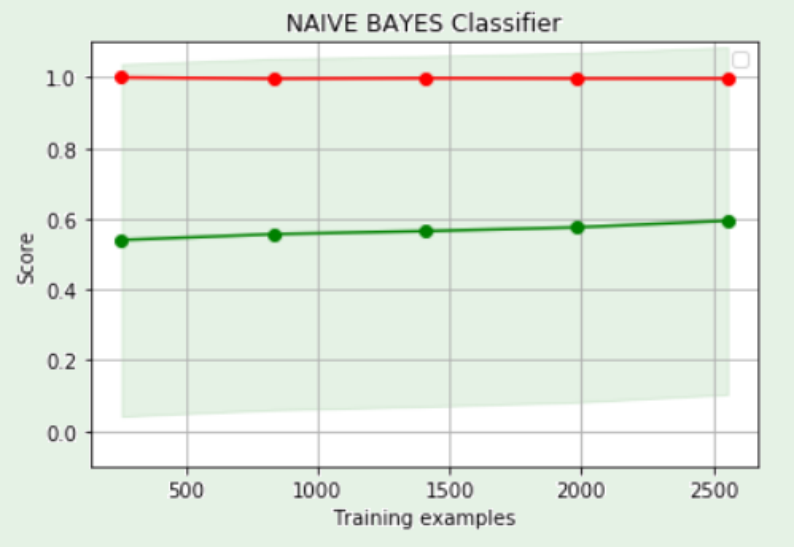

Figure 4.6 Naïve Bayes Classifier

The proposed model will use the Naive Bayes classifier algorithm to classify the news into fake or true since the application of this algorithm is for text classification and spam filtration itself. Figure 4.6 shows the learning curve for Naive Bayes classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for Naive Bayes is 0.669.

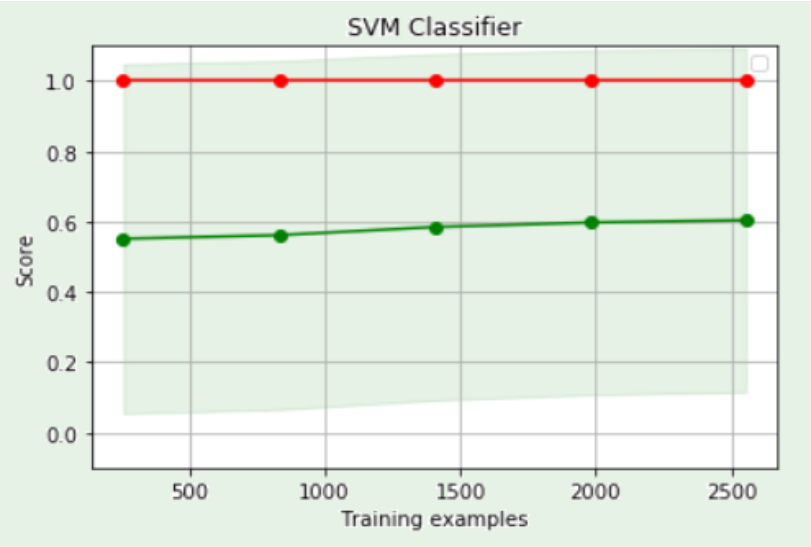

Figure 4.7 SVM Classifier

The proposed system uses inputs provided by the user to process a hyperplane by using the test and training data to compute Support Vector which is then used to distinguish between fake and authentic news. Figure 4.7 shows the learning curve for the SVM classifier where the average score increases from a range of 0.5 to 0.6 as datasets are increased and reaches the 0.6 value at 2,000. The F1 Score for SVM is 0.679.

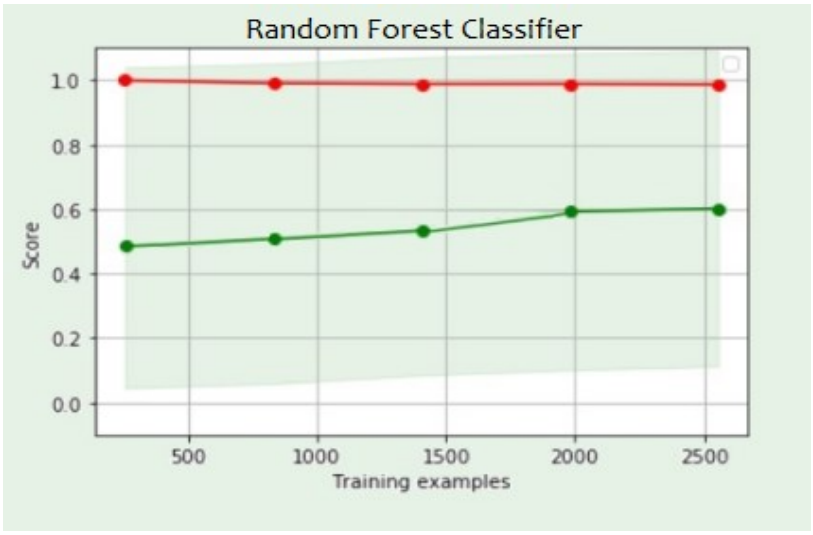

Figure 4.8 Random Forest Classifier

Random Forest is specifically used for classification and it is useful for text

classification where it consists of various decision trees correlated to each other using the training data. The feature selection is done randomly for making subsamples of data set in a form of multiple decision trees. Figure 4.8 shows the learning curve for the Random Forest classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value before 2,000. The F1 Score for Random Forest is 0.703.

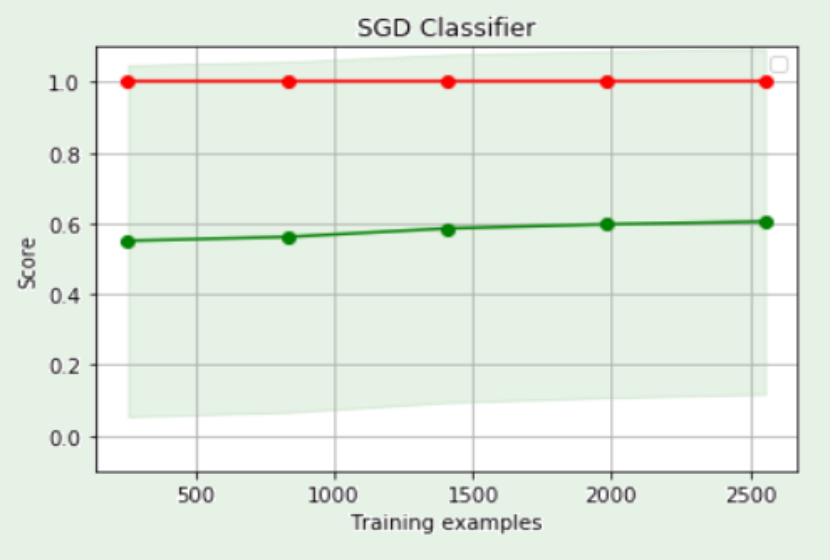

Figure 4.9 SGD Classifier

SGD classifier was chosen for text classification due to the fast computing speed provided by the algorithm which cause the decrease in the accuracy shown in the F1 Score. Figure 4.9 shows the learning curve for the SGD classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for SGD is 0.641.

Figure 4.10 Logistic Regression Classifier

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there would be only two possible classes that are genuine and hoax. Figure 4.10 shows the learning curve for the Logistic Regression classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for Logistic Regression is 0.701.

4.2.5 Results.

Check the result is the last step which is done after all the processing from the algorithm is finished and used to give the final output regarding the news. The proposed model will give the pop up output whether the given statement is true or fake. For example, The probability for Narendra Modi is prime minister of India for Naive Bayes is 0.729, SGD is 0.815, Random Forest is 0.720 while that of Logistic Regression is 0.694 and SVM calculates the distance and not probability so it is not specified in the result. The result also depicts the true and false value for the news for each algorithm.

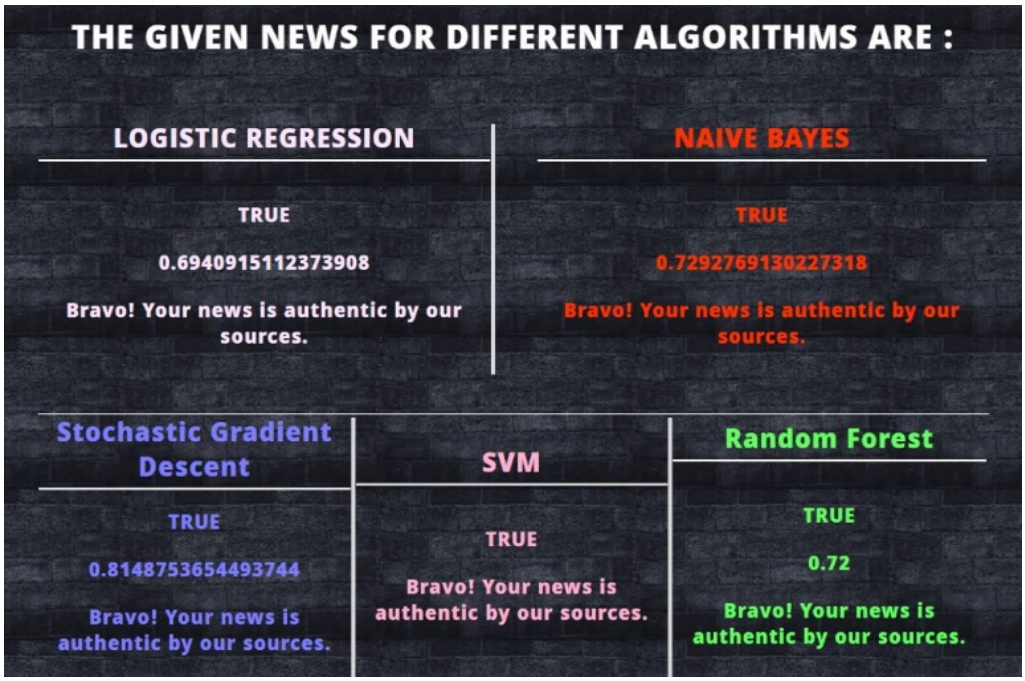

Figure 4.11 Results

Chapter 5
Conclusion

Fake news is the difficult problem because it is the rumors which it is too hard to identify the fact in contents. However Fake news can be accurately identified using machine learning methods. Naive Bayes, Neural Network, Support vector machine, are very good at detecting Fake news with high confidence.

Of course, it may not represent the whole spectrum of News in the real-world. However, there is enough evidence that Fake news is not too difficult to detect, at least in some selected domain. It is also difficult to say with confidence how much the result of this experiment can be applied to real-world news.The scope for data collection was broadened and applied the method in a more general way.

Appendix

1) StarUML

StarUML is an open source UML(Unified Modelling Language) modeling application. StarUML supports most of the diagrams like use case diagrams, activity diagrams, the sequence diagrams, class diagrams, etc. It is rich in feature set and formatting options. The digrams can be exported in jpg, jpeg and also in png formats. It provides various advancements than other modeling applications.

2) Anaconda software

The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 11 million users worldwide, it is the industry standard for developing, testing, and the training on a single machine, enabling individual data scientists to quickly download 1,500+ Python/R data science packages, Manage libraries, dependencies, and environments with Conda Develop and train machine learning and deep learning models with scikit learn, TensorFlow, and Theano Analyze data with scalability and performance with Dask, NumPy, pandas, and Numba.

URL:https://www.anaconda.com/distribution/download-section

2) Eclipse

Eclipse is an integrated development environment used in computer programming, and

is the most widely used Java IDE. It contains a base workspace and also an extensible plug-in system for customizing the environment. It is basically a developer workspace and Cloud IDE.

URL:https://www.eclipse.org/downloads/

3) Xampp software

XAMPP stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing and deployment purposes.

URL:https://filehippo.com/downloadxampp/

4) Support Vector Machine

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.

Command: from sklearn import svm

Publication

[1] Mr. Shubham Bhatt, Mr. Parth Bhatt, Mr. Mann Bajaria ”Fake News Detection”, *International Journal Of Research And Analytical Reviews, Vol 7 Issue 1*, March 2020.

Certificate of Publication

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | E-ISSN 2348-1269, P- ISSN 2349-5138

An International Open Access Journal

The Board of

International Journal of Research and Analytical Reviews (IJRAR)

Is hereby awarding this certificate to

Shubham Bhatt

In recognition of the publication of the paper entitled

FAKE NEWS DETECTION

Published In IJRAR (www.ijrar.org) UGC Approved (Journal No : 43602) & 5.75 Impact Factor

Volume 7 Issue 1 , Date of Publication:March 2020 2020-03-31 02:52:32

PAPER ID : IJRAR2002034

Registration ID : 216961

R.B.Joshi

EDITOR IN CHIEF

UGC and ISSN Approved - International Peer Reviewed Journal, Refereed Journal, Indexed Journal, Impact Factor: 5.75 Google Scholar

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS | IJRAR

An International Open Access Journal / Approved by ISSN and UGC

Website: www.ijrar.org | Email id: editor@ijrar.org | ESTD: 2014

IJRAR | E-ISSN 2348-1269, P- ISSN 2349-5138

Certificate of Publication

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | E-ISSN 2348-1269, P- ISSN 2349-5138

An International Open Access Journal

The Board of

International Journal of Research and Analytical Reviews (IJRAR)

Is hereby awarding this certificate to

Mann Bajaria

In recognition of the publication of the paper entitled

FAKE NEWS DETECTION

Published In IJRAR (www.ijrar.org) UGC Approved (Journal No : 43602) & 5.75 Impact Factor

Volume 7 Issue 1 , Date of Publication:March 2020 2020-03-31 02:52:32

PAPER ID : IJRAR2002034

Registration ID : 216961

R.B.Joshi

EDITOR IN CHIEF

UGC and ISSN Approved - International Peer Reviewed Journal, Refereed Journal, Indexed Journal, Impact Factor: 5.75 Google Scholar

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS | IJRAR

An International Open Access Journal / Approved by ISSN and UGC

Website: www.ijrar.org | Email id: editor@ijrar.org | ESTD: 2014

IJRAR | E-ISSN 2348-1269, P- ISSN 2349-5138

Certificate of Publication

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | E-ISSN 2348-1269, P- ISSN 2349-5138

An International Open Access Journal

The Board of

International Journal of Research and Analytical Reviews (IJRAR)

Is hereby awarding this certificate to

Parth Bhatt

In recognition of the publication of the paper entitled

FAKE NEWS DETECTION

Published In IJRAR (www.ijrar.org) UGC Approved (Journal No : 43602) & 5.75 Impact Factor

Volume 7 Issue 1 , Date of Publication:March 2020 2020-03-31 02:52:32

PAPER ID : IJRAR2002034

Registration ID : 216961

R.B.Joshi

EDITOR IN CHIEF

UGC and ISSN Approved - International Peer Reviewed Journal, Refereed Journal, Indexed Journal, Impact Factor: 5.75 Google Scholar

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS | IJRAR

An International Open Access Journal / Approved by ISSN and UGC

Website: www.ijrar.org | Email id: editor@ijrar.org | ESTD: 2014

IJRAR | E-ISSN 2348-1269, P- ISSN 2349-5138

Fake News Detection

Shubham Bhatt[1] Parth Bhatt[2] Mann Bajaria[3] Prof.Kanchan Dabre[4]
[1,2,3]BE Student [4]Assistant Professor
[1,2,3,4]Department of Computer Science & Engineering
[1,2,3.4]Universal College of Engg, Mumbai, India

***Abstract*- The expansion of deceiving data into regularly viewing news sources, such as internet-based life networks, news websites, and online journals, has made it possible to discern reliable news sources, thus increasing the need for computer equipment ready to give bits of information into the unwavering nature of online content. Right now, the proposed framework center around the programmed recognizable proof of false substance in online news. The research incorporates studies with various algorithms of machine learning such as Naive Bayes, SVM, Logistic Regression, Random Forest, SGD with varying sets of features. The task is isolated into two sections. First, by presenting a dataset for the undertaking of phony news discovery, the dataset is demographically blended composed of the U.S. and the Indian region. The Indian region is just considering its training space in the education sector and present many exploratory analyses to recognize linguistic discrepancies in the content of false and legitimate news. Second, the project contribution is to provide the probability of different algorithms for the given news headline. It also offers input in the form of image.**

Keywords— Support Vector Machine (SVM); Machine Learning (ML); Stochastic Gradient Descent (SGD); Natural Language Processing (NLP), Term frequency–Inverse Document Frequency (TF-IDF),Optical Character Recognition (OCR).

I. Introduction

Fake news detection has recently attracted a growing interest from the public and researchers as the circulation of miss information on-line increases, particularly in media outlets such as social media feeds, news blogs, and online newspapers. The majority of U.S. adults (62%) depend on news primarily sourced from social media. 66% of Facebook users, 59% of Twitter users, and 70% of Reddit users depend on the subsequent platforms for their news [24]. News articles that imply to be truthful, however which contain purposeful misquotes of reality with the expectation to excite interests, draw in viewership, or cheat. As of late, there have been numerous examples of unsubstantiated or false data spreading quickly finished online informal organizations. There even exist lots of websites that produce fake news almost exclusively For Example, In 2020, WhatsApp messages and Facebook posts on Indian Social Media were shared regarding Boiled thick garlic water as a treatment for Coronavirus.

This project investigates which model can offer greater accuracy and classifies the news into real or fake. It develops computational resources and models that are needed to detect fake news. The proposed model presents the fusion of dataset covering the Indian Education System and politics-specific US domain. One of the datasets is gathered physically, while the second is gathered legitimately from the web. Using this dataset, the proposed model conducts several exploratory analyses to identify linguistic properties that are predominantly present in fake content, and the proposed model is relying on linguistic features that achieve greater accuracy. To place the results in perspective, a comparison is done between the accuracy of the proposed model consisting of different ML algorithms.

II. Literature Review

There are many analysis which use the machine learning approach to calculate the creditability of news reports. Fake news is that the contents that claim folks to believe with the falsification, someday it's the sensitive messages. Once the messages were received, they're going to speedily distributed it to an alternative. The dissemination of faux news in today's digital world has accomplished on the far side a selected cluster. Combining each credible and unbelievable data on social media has created the confusion of truth, that's the

reality are hardly classified. However, the looks of faux news causes a nice threat to the protection of people's lives and property. There is information (the distributor believes they are true) or misinformation (the distributor is aware of its authenticity however he deliberately hoax) in pretend news proliferation.

Automatic deception detection: Methods for finding fake new[5] created by Conroy, Rubin, and Chen was the existing system that consisted of a good working model but few faults could be fixed with the proposed model. The following research articles are selected for review, keeping in mind the traditional and conventional approaches of fake news detection:

The lie detector[22] was used to learn the context of the fake or hoax news created by the publishers which helped to pinpoint certain phrases as well as words used to create a false news heading. Bias Aware News Recommendation System[21] was used to provide clues to how to detect the genuine and authentic news which helped us in the proposed system to detect factual news.Fake News Identification Characteristics Using Named Entity Recognition and Phrase Detection[16], Fake News Detection Model Using Media Reliability[13] and Fake News on Social Media: Brief Review on Detection Techniques[12] were all used for taking the input for preprocessing and understanding of the data set.After preprocessing and the understanding of the data set, this research uses methods to classify the believable and unbelievable message from different, there are Naïve Bayes, Random Forest, Logistic Regression, Linear Support Vector Machine(SVM) and Stochastic Gradient Descent.Naïve Bayes is a well-known classification method. We use the Naive Bayes algorithm due to the special feature for text classification and by researching the papers for reference [11][18][20] Naive Bayes was found suitable for the project.

After getting some success in implementing the Naive Bayes algorithm we started to implement the different machine learning algorithms[2][3] by using different articles to study the algorithms to improve efficiency from the existing system by using Linear Support Vector Machine, Stochastic Gradient Descent, Logistic Regression, and Random Forest Classifier.Hoax Web Detection For News in Bahasa Using Support Vector Machine[6] was referred to learn how to input the Support Vector Machine in the proposed system and increase the speed of processing since it requires less computation then Naive Bayes.Fake Data Analysis and Detection Using Ensembled Hybrid Algorithm[9] which helped the project in determining the most useful algorithm Random Forest Classifier included in the Ensembled Algorithm used for improving the accuracy.Fake News Detection Enhancement with Data Imputation[19] was used to compare the different algorithms and use different preprocessing techniques to improve the efficiency of the project.Liar a benchmark dataset for Fake News Detection was created by William Yang Wang under the name "Liar, Liar Pants on Fire" which has been used for training the system[23].

III. Proposed System

For the detection of fake news, a dataset is passed through multiple machine learning algorithms. Dataset used consist of the U.S. dataset of political facts called the"Liar, Liar Pants on Fire"[23]. The other dataset is manually entered and mainly consists of the information regarding the education system. The literature survey provided the opportunity to create a proposed system that performs better than the existing system. The findings helped us to use the algorithms in the proposed system which are Naive Bayes, SGD, SVM, Random Forest, and Logistic Regression. The algorithm which performs best is going to be selected for the Fake News detection after training it with a well-defined dataset. The dataset consisted of two field statements and labels where statements comprise of the news headlines while the label contained the value whether the news was True or False.

Certain hardware components or other software resources are to be present on the proposed system so that it can be used efficiently. These prerequisites are known as system requirement and are often used as a guideline as opposed to an absolute rule. The Software that is required for the project is Windows 8,8.1 and 10. The software used to run the program can be any depending on the user with python as a core language, for the program to run efficiently the proposed system requires minimum of 200gb of HDD/SSD with at least of Intel

core i3 or AMD Ryzen 3 processors coupled with 4gb of ram. The platform used for executing Python code is Visual Studio Code and Jupyter Notebook.

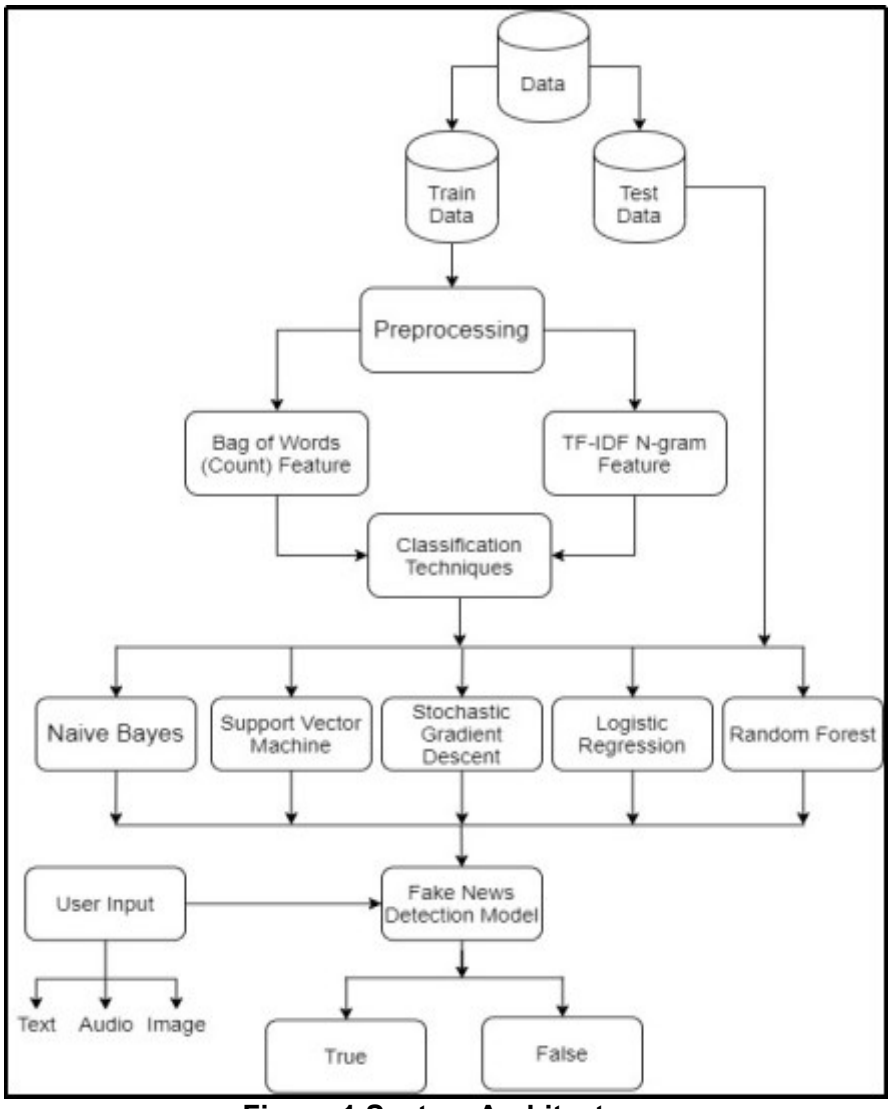

Figure 1 System Architecture

Figure 1 elaborates on the system architecture of the proposed system. The first step is the dataset which is a fusion of Indian and U.S. data. The dataset is read from the CSV file in the Data preparation file which is then evenly distributed for training and testing purposes. The training data is sent to the preprocessing to check for missing labels and quality of data, any multiple copies of data are deleted and then the data is for feature selection.

In the Feature selection, feature extraction and selection methods are obtained from sci-kit learn python libraries. For feature selection, methods like simple bag-of-words and n-grams and then term frequency like TF-IDF weighting. The system also uses word2vec and POS tagging to extract the features which in then turn is used for creating a featured embedded matrix.

After Feature selection classification models are created. The bag of words feature is selected which uses NLP for counting the number of occurrences of each word in a sentence.

This feature is then used with all the algorithms and the F1 Scores are calculated.

Now, term frequency weights with n-grams are used. This feature uses TF-IDF with 1 gram as well as 2 gram. The classifiers use this feature to calculate the F1 Score.

In terms of precision and recall it is found that Logistic Regression and Random Forest are performing the best so they are nominated as the candidate algorithms.

By running both random forest and logistic regression with GridSearch's best parameter estimation, it was concluded that the Logistic Regression algorithm with n-gram TF-IDF as well as Random Forest algorithm with n-gram TF-IDF has better accuracy than other algorithms in the proposed model.

In the proposed system the user enters the input which can be in the form of text, audio, image via GUI, the Fake News Detection System will apply the machine learning algorithm to detect whether the news is hoax or genuine and provide probabilities of each algorithm used.

Methodology of the proposed system in the form of representation of working of the model shown in Figure 2. It divides the flow of the proposed model into three parts. The first phase being Input data, the second phase being classifier and the third phase being the result.

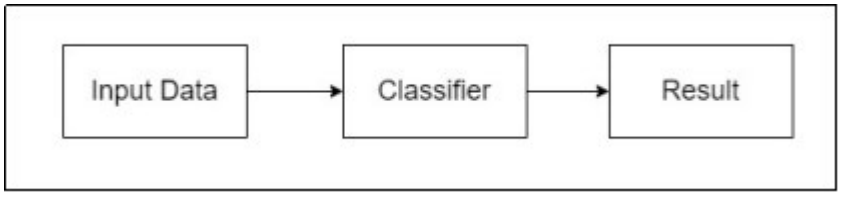

Figure 2 Methodology

a) In First Phase, the proposed model accepts the input from the user from the GUI which was created using Flask. The input can be provided in the form of text, audio as well as visual.
b) In the Second Phase, the proposed model will apply the machine learning algorithm to classify into fake or true data using the most optimum algorithm and feature from the trained dataset.
c) In the Third Phase, the proposed model display the output based on the classification and provides the probability for each algorithm for that particular news.

IV. Results and Discussion

This section depicts the entire setup of Fake news detection. All the components are connected and screenshots depict the entire procedure.
Following are the screenshot and procedure in an orderly manner:

a) Accept the input:

User needs to create an account and login to the website.After logging in user can give their input in three forms:-

Figure 3 Text an Voice Input

1) Text:- Accept the input from the user in the form of text, paragraphs, links, etc. in the search box.
2) Audio:- The input can also be in the form of voice shown in Figure 3 which is converted to text using text to speech one of the fundamentals of NLP.The Fig 3 shows the propose model working using the website where the users input a headline of the article.

Figure 4 Image Input

3) Visual:- The Figure 4 depicts the upload step to take input in the form of visual representation. The input is in the form of an image that needs to be uploaded from the local storage of device and submitted for next step as shown in Figure 4.

Figure 5 Text Conversion

The input image is converted to text using Pytesseract OCR as shown in Figure 5. Here, the image consisted of the famous quote from Nikola Tesla stating “Intelligent people tend to have less friends than the average person. The smarter you are, the more selective you become.” After the text conversion system proceeds to creation of the embedded matrix.

b) Create an Embedding matrix

Embedding Matrix is used to convert words into mathematical format, this is done using NLP.

$$X = \begin{array}{c} \\ I \\ like \\ enjoy \\ deep \\ learning \\ NLP \\ flying \\ . \end{array} \begin{array}{c} \begin{array}{cccccccc} I & like & enjoy & deep & learning & NLP & flying & . \end{array} \\ \begin{bmatrix} 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Figure 7 Embedding Matrix

For example, The occurrence of “prime minister” will be greater than the occurrence of prime India. In figure 6 Embedding Matrix the words of a sentence are converted to numeric values which help in the further processing of the model.

c) Check the existing data sets.

The input data is first checked in the dataset if not present then it is sent to the system for classification purposes where it tries to find the closest relation to any of the following statements for classification..

	Statement	Label
2	Building a wall on the U.S.-Mexico border will take literally years.	TRUE
3	Wisconsin is on pace to double the number of layoffs this year.	FALSE
4	Says John McCain has done nothing to help the vets.	FALSE
5	Suzanne Bonamici supports a plan that will cut choice for Medicare Advantage seniors.	TRUE

Figure 7 Dataset

In Figure 7, dataset consist of Statement and Label associated with it. Here, Statement consist of news headlines and Label consist of authenticity of news in the form of value.

d) Evaluate the matrix using various algorithm and features:

The proposed model will use Naive Bayes, SVM, SGD, Logistic Regression, Random Forest classifier algorithm to classify the news into fake or true and provide the probabilities.

All the below graphs show the learning curve of average training and test scores for various Classifiers where the color Bright Turquoise indicates training and Violet indicates the Cross-Validation score. Here, only 2,555 datasets were included for plotting purposes but as more datasets are included the accuracy of the model increases significantly.

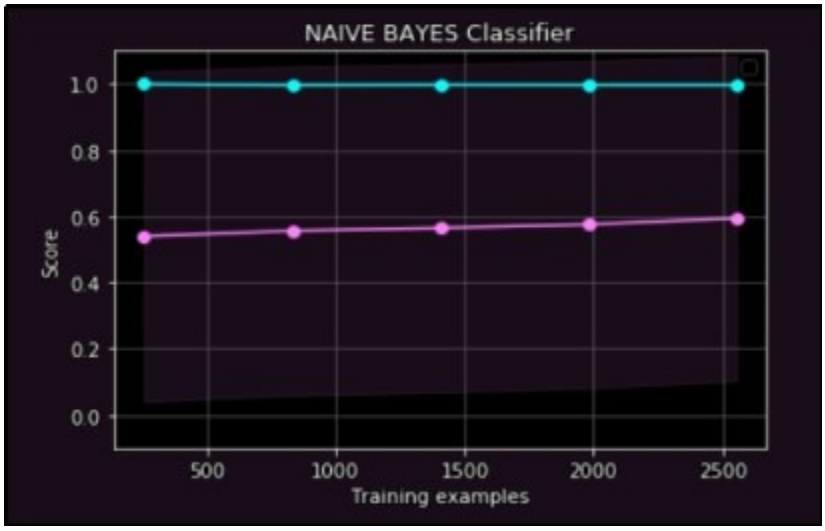

Figure 8 Learning Curve of Naive Bayes

The proposed model will use the Naive Bayes classifier algorithm to classify the news into fake or true since the application of this algorithm is for text classification and spam filtration itself. Figure 8 shows the learning curve for Naive Bayes classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for Naive Bayes is 0.669.

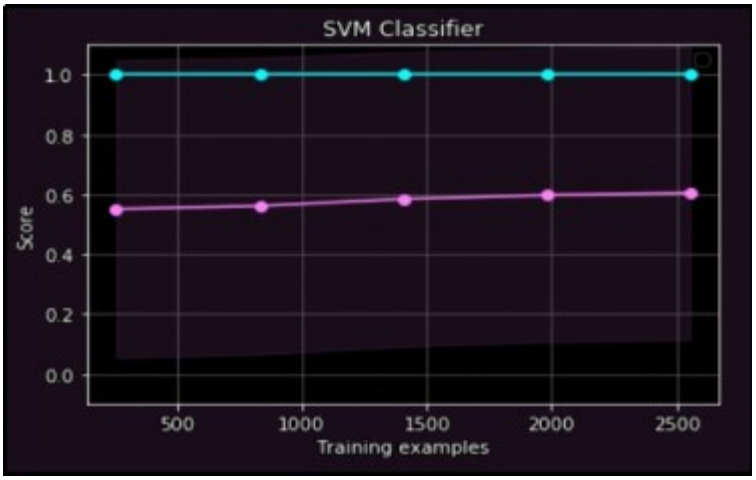

Figure 9 Learning Curve of SVM

The proposed system uses inputs provided by the user to process a hyperplane by using the test and training data to compute Support Vector which is then used to distinguish between fake and authentic news. Figure 9 shows the learning curve for the SVM classifier where the average score increases from a range of 0.5 to 0.6 as datasets are increased and reaches the 0.6 value at 2,000. The F1 Score for SVM is 0.679.

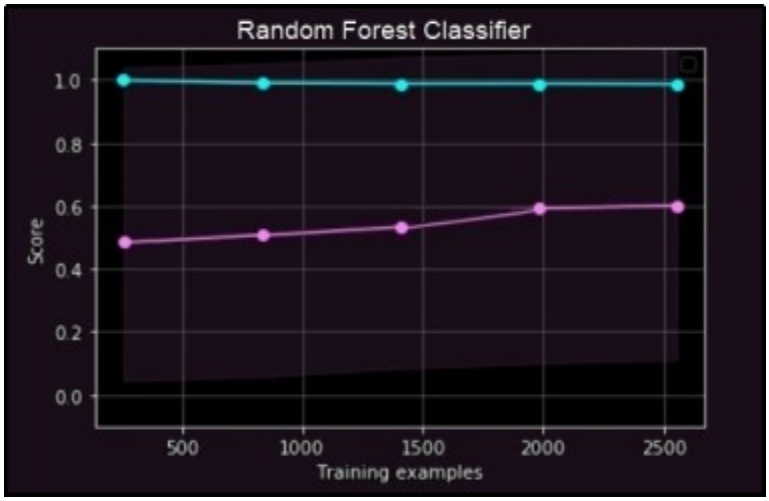

Figure 10 Learning Curve of Random Forest

Random Forest is specifically used for classification and it is useful for text classification where it consists of various decision trees correlated to each other using the training data. The feature selection is done randomly for making subsamples of data set in a form of multiple decision trees. Figure 10 shows the learning curve for the Random Forest classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value before 2,000. The F1 Score for Random Forest is 0.703.

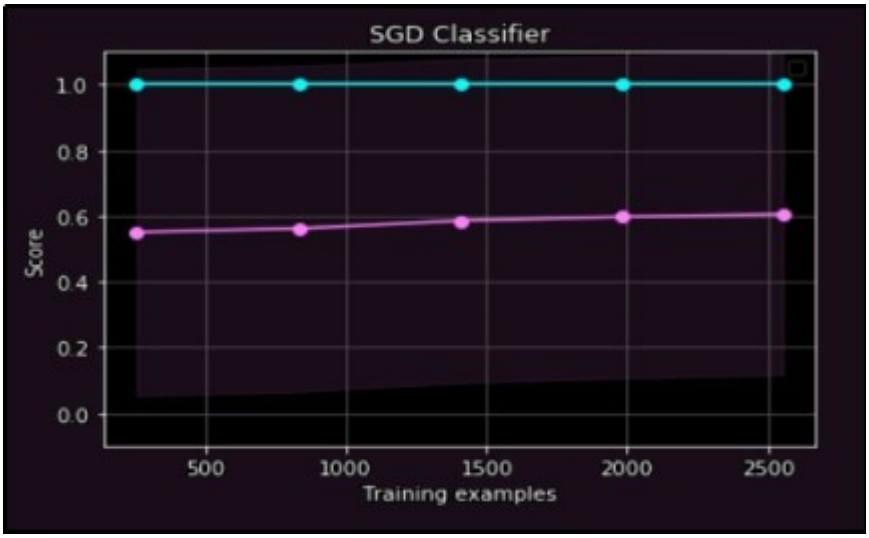

Figure 11 Learning Curve of SGD

SGD classifier was chosen for text classification due to the fast computing speed provided by the algorithm which cause the decrease in the accuracy shown in the F1 score. Figure 11 shows the learning curve for the SGD classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for SGD is 0.641.

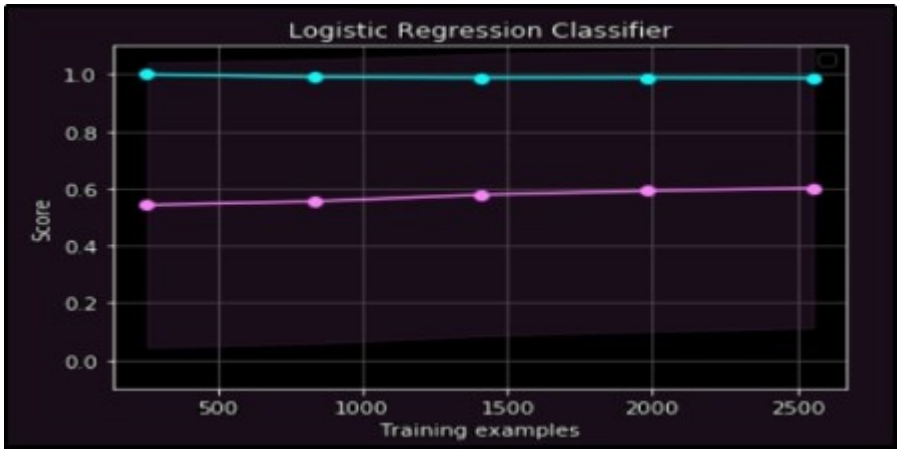

Figure 12 Learning Curve of Logistic Regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there would be only two possible classes that are genuine and hoax. Figure 12 shows the learning curve for the Logistic Regression classifier where the score increases from a range of 0.4 to 0.6 as datasets are increased and reaches the 0.6 value at 2,555. The F1 Score for Logistic Regression is 0.701.

e) Check the result:

The proposed model will give the output on the next page along with the calculated probabilities for each algorithm as well as classify the news as true or false and display in the format shown in Figure 7 Result.

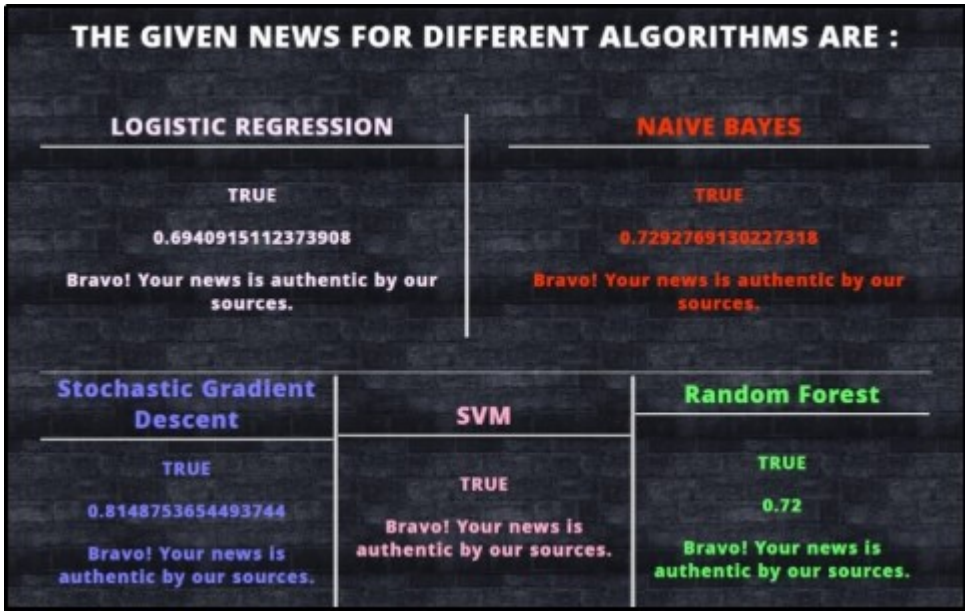

Figure 13 Result

For example, The probability for Narendra Modi is prime minister of India for Naive Bayes is 0.729, SGD is 0.815, Random Forest is 0.720 while that of Logistic Regression is 0.694 and SVM calculates the distance and not probability so it is not specified in the result. The result also depicts the true and false value for the news for each algorithm.

v. Conclusion

Artificial news is the troublesome problem since it is the bits of gossip which is too difficult to even think about identifying the reality in contents. However Fake news can be precisely placed utilizing AI strategies. Logistic Regression, Random Forest, are truly adept at identifying Fake news with high certainty.

Clearly, in the real world, it may not refer to the whole spectrum of News. There's convincing evidence that fake news isn't too hard to even think about finding in some domain selected at any level. It is additionally difficult to decide with sureness how far the result of the entire examination could be adjusted to genuine world news. The scope for information assortment was expanded and applied the strategy in a progressively broad manner.

References

1 "Karishnu Poddar ,Geraldine Bessie Amali D. & K.S. Umadevi,"Comparison of Various Machine Learning Models for Accurate Detection of Fake News",Jan 2020.

2 Supaya Aphiwongsophon and Prabhat Chongstitvatana,"Detecting fake news with machine learning method",IEEE,July 2018.

3 Akshay Jain ,Amey Kasbe,"Fake news detection",IEEE,Feb 2018.

4 S. Maheshwari,"How fake news goes viral: A case study", Nov 2016

5 Niall J. Conroy ,Victoria L. Rubin & Yimin Chen,"Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52,Feb 2016.

6 Rahmat M.A, Indrabayu & Intan Sari Areni,"Hoax Web Detection For News in Bahasa Using Support Vector Machine",Dec 2019.

7 Abhishek Verma ,Vanshika Mittal & Suma Dawn,"FIND: Fake Information and News Detections using Deep Learning",Sep 2019.

8 Bhavika Bhutani ,Neha Rastogi ,Priyanshu Sehgal & Archana Purwar,"Fake News Detection Using Sentiment Analysis",Sep 2019.

9 Palagati Bhanu Prakash Reddy ,Mandi Pavan Kumar Reddy ,Ganjikunta Venkata Manaswini Reddy ,K. M. Mehata,"Fake Data Analysis and Detection Using Ensembled Hybrid Algorithm",Aug 2019.

10 Rohit Kumar Kaliyar,"Fake News Detection Using A Deep Neural Network",July 2018.

11 Mykhailo Granik & Volodymyr Mesyura,"Fake news detection using naive Bayes classifier",Nov 2017.

12 Zaitul Iradah Mahid ,Selvakumar Manickam & Shankar Karuppayah,"Fake News on Social Media: Brief Review on Detection Techniques",July 2019.

13 Youngkyung Seo ,Deokjin Seo & Chang-Sung Jeong,"Fake News Detection Model Using Media Reliability",Feb 2019.

14 Ye-Chan Ahn & Chang-Sung Jeong,"Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning",Oct 2019.

15 Ammu Kuriakose ,Dinnu Sebastian ,Esther Mahima Mathew ,Hannu Mathew & G Er.Gokulnath,"ALIKAH- A Clickbait and Fake News Detection System using Natural Language Processing",Oct 2019.

16 Herley Shaori Al-Ash & Wahyu Catur Wibowo,"Fake News Identification Characteristics Using Named Entity Recognition and Phrase Detection",Nov 18.

17 Oluwaseun Ajao ,Deepayan Bhowmik & Shahrzad Zargari,"Sentiment Aware Fake News Detection on Online Social Networks",April 2019.

18 Inggrid Yanuar Risca Pratiwi ,Rosa Andrie Asmara & Faisal Rahutomo,"Study of hoax news detection using naïve bayes classifier in Indonesian language",Jan 2018.

19 Chandra Mouli Madhav Kotteti ,XIshuang Dong ,Na Li & Lijun Qian,"Fake News Detection Enhancement with Data Imputation",Oct 2018.

20 Fatma Najar ,Nuha Zamzami & Nizar Bouguila,"Fake News Detection Using Bayesian Inference",Sep 2019.

21 Anish Patankar ,Joy Bose & Harshit Khanna,"A Bias Aware News Recommendation System",Mar 2019.

22 Rada Mihalcea & Carlo Strapparava,"The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language",Aug 2009.

23 William Yang Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection", Aug 2017.

24 Gottfried J. & Shearer E,"News use across social media platforms”,May 2016.

Acknowledgement

We take this opportunity to express our deep sense of gratitude to our project guide **Prof. Kanchan Dabre** and Head of Computer Engineering Department **Dr. Jitendra Satturwar** for their continuous guidance and encouragement throughout the duration of our project work. It is because of their experience and wonderful knowledge that we can fulfill the requirement of completing the project within the stipulated time.We would also like to thank **Prof. Vishakha Shelke**,our project coordinator for her encouragement, whole-hearted cooperation and support. We would also like to thank our Principal **Dr.J. B. Patil**, the management of Universal College of Engineering, Vasai, Mumbai for providing us all the facilities and the work friendly environment. We acknowledge with thanks, the assistance provided by departmental staff, library and lab attendants.

Mann Arish Bajaria (CS 08)

Parth Bhadresh Bhatt (CS 15)

Shubham Divyeshkumar Bhatt (CS 16)