TO STUDY FAKE NEWS DETECTION IN ONLINE SOCIAL MEDIA IN CONTEXT OF MACHINE LEARNING

A PROJECT REPORT

Submitted by

Devanshi Limbasiya (150030107054)

Pooja Rada (150030107075)

Under the guidance of

Prof. Rajesh Davda

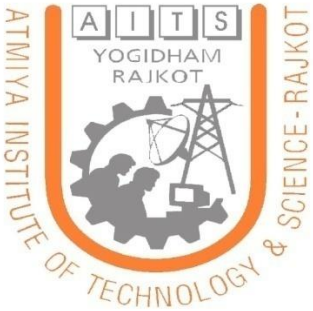
A Report submitted to Gujarat Technological University in partial fulfillment for the award of the degree

Of

BACHELOR OF ENGINEERING

in

Computer Engineering



Atmiya Institute of Technology & Science

Yogidham Gurukul, Kalawad Road,

Rajkot-360005.

Gujarat Technological University, Ahmedabad

MAY-2019

ATMIYA

INSTITUTE OF TECHNOLOGY & SCIENCE RAJKOT

This is to certify that the project entitled "**To Study Fake News Detection In Online Social Media in context of Machine Learning**" is a bonafied report of the work carried out by **Ms. Pooja Rada** under the guidance and supervision for the work of Project-II at Atmiya Institute of Technology and Science - Rajkot, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate herself, has duly been completed, fulfills the requirement of the ordinance relating to the Bachelor degree of the university and is up to the standard in respect of content, presentation and language for being referred to the examiner.

SIGNATURE
PROF.TOSAL BHALODIA
HEAD OF THE DEPARTMENT
C. E. DEPARTMENT

SIGNATURE
MR. RAJESH DAVDA
ASST PROF
C. E. DEPARTMENT

ATMIYA

INSTITUTE OF TECHNOLOGY & SCIENCE RAJKOT

This is to certify that the project entitled "**To Study Fake News Detection In Online Social Media in context of Machine Learning**" is a bonafied report of the work carried out by **Ms. Devanshi Limbasiya** under the guidance and supervision for the work of Project-II at Atmiya Institute of Technology and Science - Rajkot, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate herself, has duly been completed, fulfills the requirement of the ordinance relating to the Bachelor degree of the university and is up to the standard in respect of content, presentation and language for being referred to the examiner.

SIGNATURE
PROF.TOSAL BHALODIA
HEAD OF THE DEPARTMENT
C. E. DEPARTMENT

SIGNATURE
MR. RAJESH DAVDA
ASST PROF
C. E. DEPARTMENT

INDEX

Acknowledgement

We have taken many efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We are highly indebted to Rajesh Sir for his guidance and constant supervision as well as for providing necessary information regarding the Project Titled **"To Study Fake News Detection in Online Social Media in Context of Machine Learning".** We would like to express our gratitude towards staff members of Computer Engineering Department, Atmiya Institute of Technology & Science for their kind co-operation and encouragement which helped us in completion of this project.

We also say the big thank you to our parents for such a support and without them we can do nothing not in just project but also in life. Thankful to our family for their support.

The feeling of gratefulness to any one's help directly arises from the bottom of heart. A small but an important and timely help can prove to be a milestone in one's life.
We even thank and appreciate to our colleague in developing the project and people who have willingly helped us out with their abilities.

Very thankful to almighty of all of us "God" to give us such a best persons and all the thing he provides before we need and we always feel that without him we are nothing.

Pooja Rada
Devanshi Limbasiya

Abstract

The scourge of cyberbullying has assumed alarming proportions with an ever-increasing number of adolescents admitting to having dealt with it either as a victim or as a bystander.

Anonymity and the lack of meaningful supervision in the electronic medium are two factors that have exacerbated this social menace.

Comments or posts involving sensitive topics that are personal to an individual are more likely to be internalized by a victim, often resulting in tragic outcomes.

We decompose the overall detection problem into detection of sensitive topics, tending itself into text classification sub-problems.

We find that binary classifiers for individual tablets outperform multiclass classifiers.

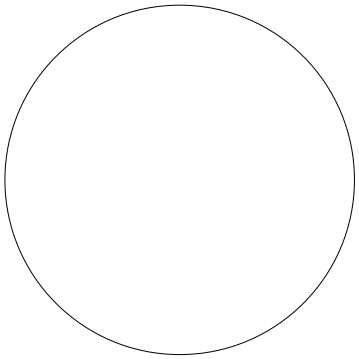
Our findings show that the detection of textual cyberbullying can be tackled by building individual topic sensitive classifiers.

List of Figures

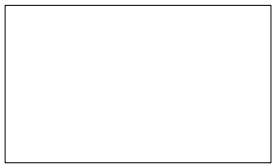
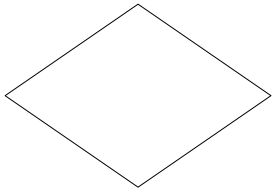
List of Tables

LIST OF SYMBOLS

DFD DIAGRAM SYMBOLS

Symbols	Meaning
	Entity
	Data Store
	Data Flow
	Data Process

E-R DIAGRAM SYMBOLS

Symbols	Meaning
	Entity
	Relationship
	Connector

Chapter 1

Introduction

1.1 Purpose

1.2 Scope

1.3 Objective

1.4 Technology and Tool

1.0 INTRODUCTION

1.1 PURPOSE:

The purpose of this project is to use machine learning algorithm to detect the fake news in online social media that travels as a real one, it is like a click bait.

It will try to enhance the user experience on the online social media platform and will also save lot of time of users that they might spent on fake news otherwise.

.

1.2 SCOPE:

The scope of this project is very diverse, it ranges from various online social media like Facebook, twitter, Instagram etc. to fake blogs, fake websites that deceive the users in one way or the other.

1.3 OBJECTIVE:

This is the standalone application that will use the dataset which is consists of various information in mixture it contains fake news and real news and also the news that appear real but are fake.

1.4 TECHNOLOGY AND TOOLS:

Front End: For designing the structure of the project following technologies are used:

1) **Jupyter notebook:**

It is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning etc.

The Notebook is a server-client application that allows editing and running notebook documents via a web browser. It can be executed on a local desktop requiring no internet access or can be installed on a remote server and accessed through the internet.

In addition to displaying/editing/running notebook documents. It has a "Dashboard" (Notebook Dashboard), a "control panel" showing local files and allowing to open notebook documents or shutting down their kernels.

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating, executing, and visualizing Jupyter notebooks.

It is similar to the notebook interface of other programs such as Maple, Mathematica, and SageMath, a computational interface style that originated with Mathematica in the 1980s.It supports execution environments (aka kernels) in dozens of languages. By default Jupyter Notebook ships with the IPython kernel but there are over 100 Jupyter kernels as of May 2018.

2) Anaconda:

Anaconda is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system *conda*.

Anaconda is a scientific Python distribution. It has no IDE of its own. Anaconda bundles a whole bunch of Python packages that are commonly used by people using Python for scientific computing and/or data science.

It provides a single download and an install program/script that install all the packages in one go. Alternate is to install Python and individually install all the required packages using pip. Additionally, it provides its own package manager (conda) and package repository. But it allows installation of packages from PyPI using pip if the package is not in Anaconda repositories. It is especially good if you are installing on Microsoft Windows as it can easily install packages that would otherwise require you to install C/C++ compilers and libraries if you were using pip. It is certainly an added advantage that conda, in addition to being a package manager, is also a virtual environment manager allowing you to install independent development environments and switch from one to the other (similar to virtualenv).

3). Python:

Python is an interpreted, object-oriented, high level programming with dynamic semantics.

Its high level built in data structures, combined with dynamic typing and binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. It supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Debugging Python program is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it causes an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on.

3). Dataset:

A **dataset** is a collection of data. Most commonly a data set corresponds to the contents of a single data basé table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. It lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum.

The dataset may comprise data for one or more members, corresponding to the number of rows. The term dataset may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event.

3) Machine learning:

Machine learning gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959).It is a subfield of computer science.

Machine learning explores the construction of algorithms which can learn and make predictions on data. Such algorithms follow programmed instructions, but can also make predictions or decisions based on data. They build a model from sample inputs.

Machine learning is done where designing and programming explicit algorithms cannot be done. Examples include spam filtering, detection of network intruders or malicious insiders working towards a data breach, fake news detection in online social media.

4) Deep Learning:

Deep learning is part of a of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, processing, social network filtering, machine translation, bioinformatics, drug design and board game programs, where they have produced results comparable to and in some cases even exceeded the human experts.

5) Naive Bayes Algorithm:

In machine learning, **naive Bayes classifiers** are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

It is a popular method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines.

Chapter 2

Project Management

2.1 Project Planning

2.2 Project Scheduling

2.3 Risk Management

2.0. PROJECT MANAGEMENT

2.1 PROJECT PLANNING

Project Planning is concerned with identifying and measuring the activities, milestones and deliverables produced by the project. Project planning is undertaken and completed sometimes even before any development activity starts. Project planning consists of following essential activities:

- Scheduling manpower and other resources needed to develop the system.
- Staff organization and staffing plans.
- Risk identification, analysis, and accurate planning.
- Estimating some of the basic attributes of the project like cost, duration and efforts the effectiveness of the subsequent planning activities is based on the accuracy of these estimations.
- Miscellaneous plans like quality assurance plan, configuration management plan, etc.

Project management involves planning, monitoring and control of the process, and the events that occurs as the software evolves from a preliminary concept to an operational implementation. Cost estimation is a relative activity that is concerned with the resources required to accomplish the project plan.

2.1.1 Project Development Approach And Justification:

A Software process model is a simplified abstract representation of a software process, which is presented from a particular perspective. A process model for software engineering is chosen based on the nature of the project and application, the methods and tools to be used, and the controls and deliverables that are required. All software development can be characterized as a problem-solving loop which in four distinct stages is encountered:

- Requirement analysis
- Coding
- Testing

- Deployment

2.1.2 Milestones and Deliverables:

As software is tangible, this information can only be provided as documents that describe the state of the software being developed without this information it is impossible to judge progress at different phases and therefore schedules cannot be determined or updated.

Milestone is an end point of the software process activity. At each milestone there should be formal output such as report that can be represented to the guide. Milestones are the completion of the outputs for each activity. Deliverables are the requirements definition and the requirements specification.

Milestone represents the end of the distinct, logical stage in the project. Milestone may be internal project results that are used by the project manager to check progress. Deliverables are usually Milestones but reverse need not be true. We have divided the software process into activities for the following milestone that should be achieved.

Software Process Activity	**Milestone**
Project Plan	Project schedule
Requirement Collection	User requirements, System Requirements
Analysis of Dataset	Choosing of appropriate dataset.
Implementation	Algorithm implementation.

Table 2.1.2.1 Milestones and Deliverables

2.1.3 Roles and Responsibilities:

This phase defines the role and responsibilities of each and every member involved in developing the system. To develop this system there was only one group with two members working on the whole application. Each member was responsible for each and every part of developing the system. Each of the group members has sufficient knowledge in several programming languages. Our team structure is of mixed control team organization as it consists of both democratic and chief programmer organization.

Task Identification :

Task	Person Name
Analysis, Analysis review	Pooja Rada, Devanshi Limbasiya, Rajesh Davda
Dataset Analysis	Pooja Rada, Devanshi Limbasiya
Algorithm Analysis	Pooja Rada , Devanshi Limbasiya
Algorithm Review and Confirmation	Rajesh Davda
Implementation	Pooja Rada, Devanshi Limbasiya
Documentation	Pooja Rada, Devanshi Limbasiya

Table 2.1.3.1 Roles And Responsibilities

2.1.4 Group Dependencies:

The structure chosen for the system is the chief programmer structure .In this system, Chief Programmer team structure is used because in the organization, a senior engineer provides the technical leadership and is designated as the chief programmer. The chief programmer partitions the task into small activities and assigns them to the team members. He also verifies and integrates the products developed by different team members and they work under the constant supervision of the chief programmer. For this system reporting entity represents myself and the role of chief programmer is played by my internal guide.

2.2 PROJECT SCHEDULING

The scheduling is the peak of a planning activity, a primary component of software project management. When combined with estimation methods and risk analysis, scheduling establishes a roadmap for project management. The characteristics of the project are used to adapt an appropriate task set for doing work.

Tasks	Started	Finished	Number of days
PPR-1	10/7/18	15/7/18	5 DAYS
PPR-2	20/7/18	24/7/18	4 DAYS
PPR-3	28/7/18	2/8/18	6 DAYS
PPR-4	3/8/18	10/8/18	8 DAYS
PSAR	11/8/18	16/8/18	6 DAYS
CANVAS	17/8/18	21/8/18	5 DAYS
NOVELTY REPORT	15/9/18	20/9/18	5 DAYS
REPORT	21/9/18	28/9/18	8 DAYS
PLAGARISM REPORT	1/10/18	2/10/18	2 DAYS

Fig. 2.2.1 shows timeline chart of this project

2.3 RISK MANAGEMENT

Risk management consists of a series of steps that help a software development team to understood and manage uncertain problems that may arise during the course of software development and can plague a software project.

Risks are the dangerous conditions or potential problems for the system which may damage the system functionalities to very high level which would not be acceptable at any cost. so in order to make our system stable and give its 100% performance we must have identify those risks, analyze their occurrences and effects on our project and must prevent them to occur.

2.3.1 Risk Identification

Risk identification is a first systematic attempt to specify risks to project plan, scheduling resources, project development. It may be carried out as a team process using brainstorming approach.

Technology risk: Technical risks concern implementation and testing problems.

- ➢ Dataset Enlargement
- ➢ Algorithm Output.

People Risks: These risks are concerns with the team and its members who are taking part in developing the system.

- ➢ Lack of knowledge
- ➢ Lack of clear vision.
- ➢ Poor communication between people.

Tools Risks:

These are more concerned with tools used to develop the project.

- ➢ Tools containing virus.

General Risks:

General Risks are the risks, which are concerned with the mentality and resources.

- ➢ Rapidly changing Datasets.
- ➢ Lack of resources can cause great harm to efficiency and timelines of project.
- ➢ Changes in dataset can cause a great harm to implementation and schedule of developing the system.
- ➢ Insufficient planning and task identification.
- ➢ Decision making conflicts.

2.3.2 Risk Analysis

"Risk analysis = risk assessment + risk management + risk communication. "

Risk analysis is employed in its broadest sense to include:

Risk assessment

Involves identifying sources of potential harm, assessing the likelihood that harm will occur and the consequences if harm does occur.

For this project It might be :-

Software(Tool) Crashing.

Risk management

Evaluates which risks identified in the risk assessment process require management and selects and implements the plans or actions that are required to ensure that those risks are controlled.

Precautions taken to make risks minimal are as under:-

Keeping the software tool up to date by updating the software periodically.

Risk communication

Involves an interactive dialogue between guide and us, which actively informs the other processes.

Steps taken for risk communication is as under: -

- All the possible risks are listed out during communication and project is developed taking care of that risks.

Chapter 3

System Requirements Study

3.1 User Characteristics

3.2 Hardware and Software Requirements

3.3 Constraints

3.4 Assumptions and Dependencies

3.0 SYSTEM REQUIREMENT STUDY

3.1 USER CHARACTERISTICS

Admin:-

- Show project and user full detail
- Manage user
- Mange project
- Manage dataset

User:-

- Upload pieces of news
- Circulation of news
- Analyze the news

3.2 HARDWARE AND SOFTWARE REQUIREMENT SPECIFICATION

This shows minimum requirements to carry on to run this system efficiently.

3.2.1 Hardware Requirements

Server side Hardware Requirement:

Devices	Description
Processor	Intel Core Duo 2.0 GHz or more
RAM	512 MB or more
Hard Disk	10 GB or more

Table 3.2.1.1 Server side Hardware Requirement

3.2.2 Software Requirements

For which	Software
Operating System	Windows XP/2003/vista/7/8/10,Linux, Mac OS x
Front End	Jupyter notebook
Back End	Numpy, Panda
Scripting Language	Python

Table 3.2.2.1 Software Requirements

3.2.3 Client side Requirements

For which	Requirement
Browser	Any Compatible browser device

Table 3.2.3.1 client-side Requirements

3.3 CONSTRAINTS

3.3.1 Hardware Limitations

The major hardware limitations faced by the system are as follows:

If the appropriate hardware is not there like processor, RAM, hard disks

-the problem in processing requests of client

-if appropriate storage is not there our whole database will crash due to less storage because our main requirement is large storage.

3.3.2 Interfacing with other systems

There should be the compatible browser to perfectly detect the fake news. The functionality of the system should be such that it can be used as sub module of some larger applications.

3.3.3 Reliability Constraints

The major reliability constraints are as follows:

- The software should be efficiently designed so as to give reliable recognition of fake news and so that it can be used for more pragmatic purpose.
- The design should be versatile and user friendly.
- The application should be fast, reliable and time saving.
- The system should have universal adaptations.
- The system be compatible with future upgradation.

3.4 DEPENDENCIES

The entire project depends on various libraries of python. The libraries are as follows:

NumPy: NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Pandas: pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the python programming language.
pandas is a NumFOCUS sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to donate to the project.

Python: This module implements a number of iterator building blocks inspired by constructs from APL, Haskell and SML. Each has been recast in a form suitable for Python.

Matplotlib: Matplotlib is a Python 2D plotting library which produces publication quality figures in variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

Scikit: Simple and efficient tools for data mining and data analysis. Accessible to everybody, and reusable in various contexts. Built on NumPy, SciPy, and matplotlib. Open source, commercially usable-BSD license.

Chapter 4

System Analysis

4.1 Study of Current System

4.2 Problem and Weaknesses of Current System

4.3 Requirements of New System

4.4 Feasibility Study

4.5 Requirements Validation

4.6 Features Of New System

4.7 Data Flow Diagram

4.8 ER Diagram

4.9 Selection Of Hardware and Software and Justification

4.1 STUDY OF CURRENT SYSTEM

Current system focus on classifying online reviews and publicly available social media posts.

4.2 PROBLEMS AND WEAKNESS OF CURRENT SYSTEM

The current system is undoubtedly well-designed for detecting the deception but it has some following limitations:

"Conroy, Rubin, and Chen outline several approaches that seem promising towards the aim of perfectly classify the misleading articles. They note that simple content-related n-grams and shallow parts-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.
Feng and Hirst implemented a semantic analysis looking at 'object: descriptor' pairs for contradictions with the text on top of Feng's initial deep syntax model for additional improvement. Rubin, Lukoianova and Tatiana analyze rhetorical structure using a vector space model with similar success. Ciampaglia et al. employ language pattern similarity networks requiring a pre-existing knowledge base. "

- Lack of an awareness of this system.
- Implementation is difficult and complex
- Some security related issues may be created.
- Cost Effectiveness

4.3 REQUIREMENTS SPECIFICATION

Requirements specification adds further information to the requirements definition.

4.3.1 Algorithm Requirements

- Dataset
- Input
- Appropriate functions
- Training
- Efficiency
- Output

4.3.2 System Requirements

- Usability:

The system should be easily able to detect the deception in blogs or news in online social media.

- Efficiency:

The system should provide easy and fast response.

4.4 FEASIBILTIY STUDY

An important outcome of the preliminary investigation is the determination that the system is feasible or not. The main aim of the feasibility study activity is to determine whether it would be financially and technically feasible to develop a project.

The feasibility study activity involves the analysis of the problem and collection of all relevant information relating to the product such as the different datasets which would be input to the system, the processing required to be carried out on these datasets, the output required to be produced by the system as well as the various constraints on the behaviors of the system.

4.4.1. Does the system contribute to the overall objectives of the organization?

The main aim of behind development of this system is to provide fake news detection that can prevent the social bullying of the persons which need it and also for the people who does want to waste their time on fake news.

4.4.2. Can the system be implemented using the current technology and within the given cost and schedule constraints?

- The system can be easily implemented using existing technology. The technology used is anaconda which is user friendly and freeware. After seeing the functionality that system provides the cost of developing the application does not matter.

- Taking the schedule constraints in consideration the time available is approximately 9 months. The time period is enough to develop the system.

4.5 REQUIREMENT VALIDATION

A requirements validation is concerned to check whether the requirements actually define the system, which the customer wants? Requirements validation is important because errors in requirements document can lead to extensive rework costs when they are subsequently discovered. We have performed the following validation checks

- **Validity checks**

Check whether the information entered is in valid format

- **Consistency checks**

A requirement in a document is not conflicting.

- **Completeness checks**

The requirements document includes requirement, which define all functions, and constraints intended by the system user.

- **Realism checks**

Using knowledge of existing technology, the requirements are checked to ensure that they could actually be implemented.

- **Verifiability**

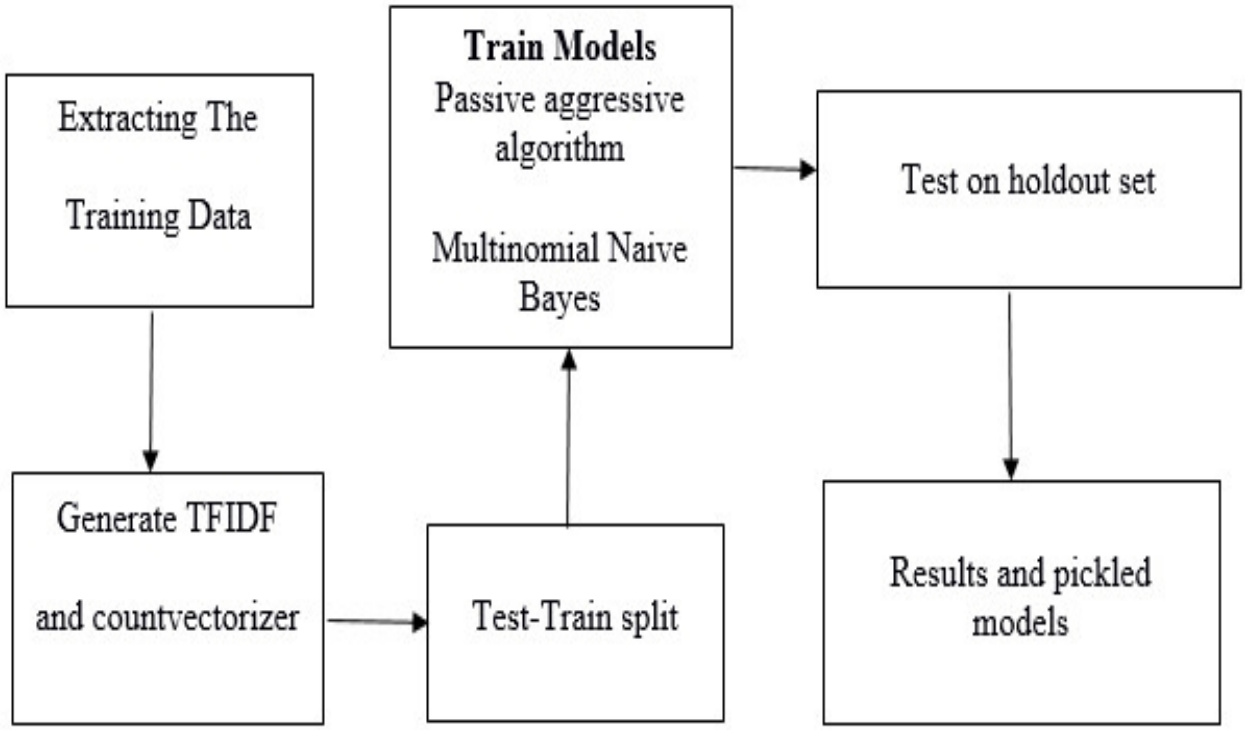
The requirements are given in verifiable manner (e.g.: Using quantifiable measures) to reduce disputes between client and developer.

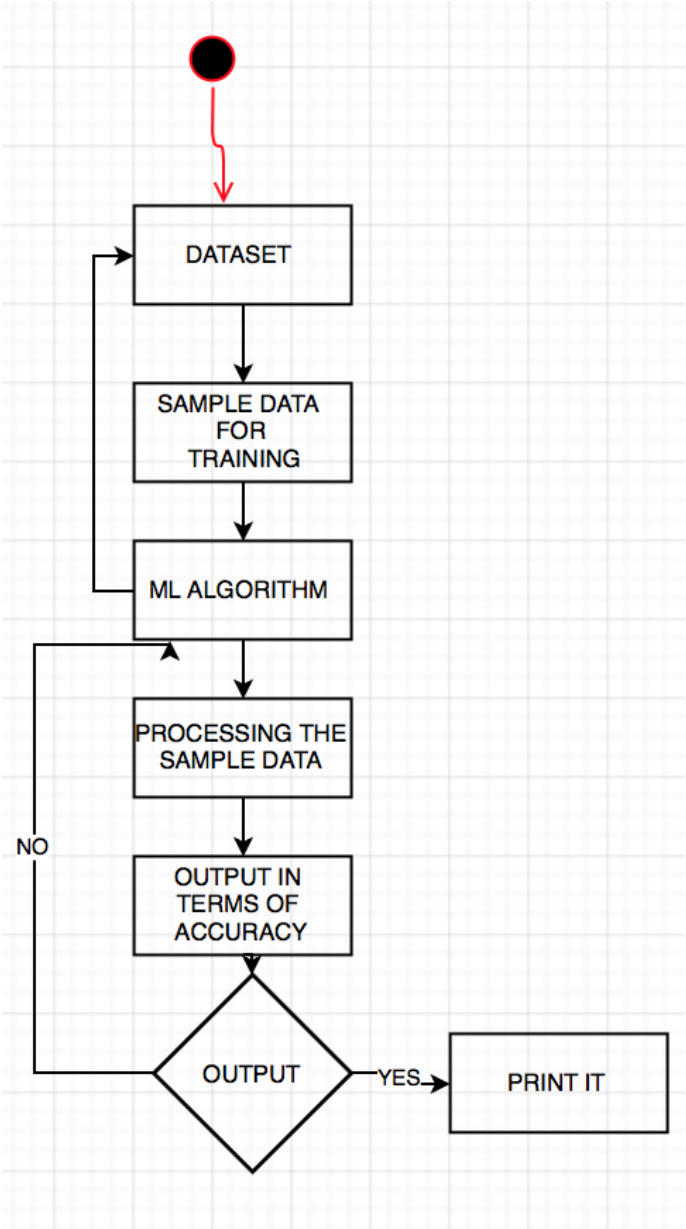
4.6 FEATURES OF NEW SYSTEM

We will try to develop application as follows:

- The system being available in regional languages.
- Provide the more awareness in our country India about this concept.
- User can upload his/her idea through description, team information, videos of his/her work, and the form of reward and main for which purpose he/she needed the money.
- One can pledge the money if one like the idea.
- Communication provided between innovators and investors.
- Safety for money transfer and surety of security of ideas.

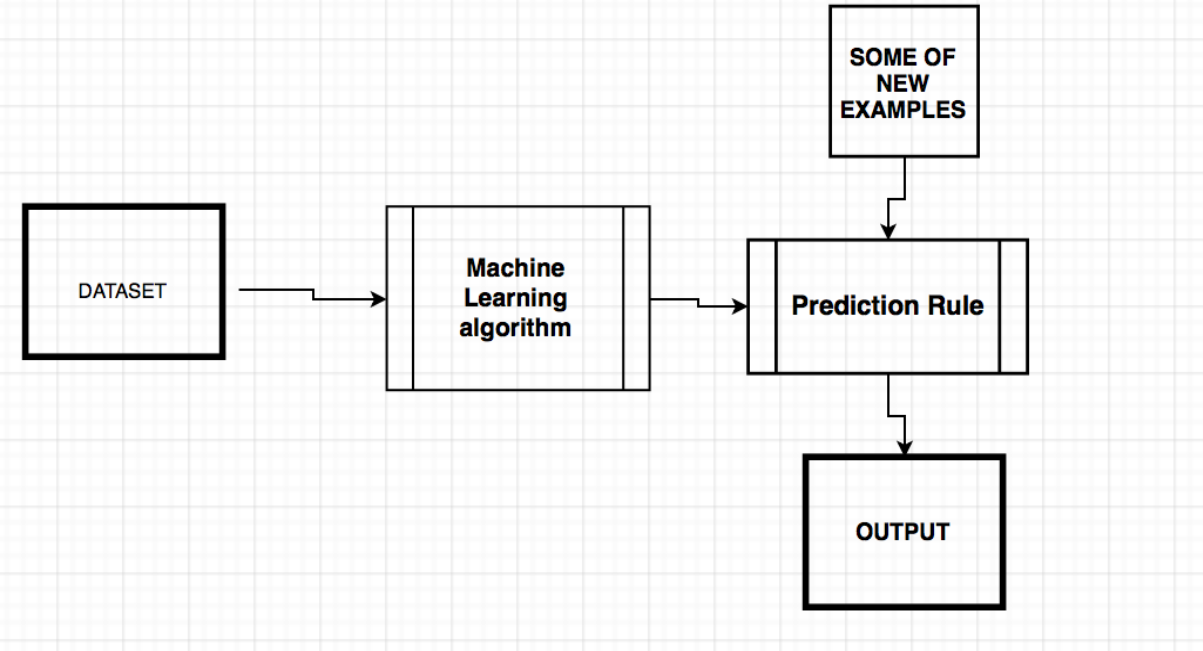
4.7 FLOWCHART OF NEW SYSTEM:





4.7 Activity Diagram

4.8 E-R Diagram



4.9 SELECTION OF HARDWARE AND SOFTWARE

The Tables 4.9.1 and 4.9.2, 4.9.3 below give idea of the hardware and software required for the system and client side requirements.

- **Hardware Selection**

Devices	Description
Processor	Intel Core Duo 2.0 GHz or more
RAM	512 MB or more
Hard Disk	10 GB or more

Table 4.9.1 Hardware Requirements

- **Software Selection**

For which	Software
Operating System	Windows XP/2003/vista/7/8/10,Linux, Mac os x
Front End	Jupyter Notebook
Back End	Numpy,Panda
Scripting Language	Python

Table 4.9.2 Software Requirements

- **Client side requirements:**

For which	Requirement
Browser	Any Compatible browser device

Table 4.9.2 Client Side Requirements

Chapter 5

System Design

5.1 Dataset Selection

5.2 Code Snippet

5.1 DATASET SELECTION

As any machine learning algorithm requires dataset, it plays an important part in algorithm. Dataset selection is a crucial process in where we try to find the appropriate dataset which in turn will act as an input to our system and will be used to train the classifier.

For our project we have selected the "Liar" dataset which was available online and had lots of recommendations for using it. It is divided into parts as train, test, valid, and will be given to the classifier.

The following is the glimpse of our wide dataset and it also shows how the data is arranged in the dataset.

```
1  12134.json	barely-true	We have less Americans working now than in the 70s.	economy,jobs	vicky-hartzler	U.S. Representative	Missouri	republican	1	0	1	0	0	an interview with ABC17 News
2  238.json	pants-fire	When Obama was sworn into office, he DID NOT use the Holy Bible, but instead the Kuran (Their equivalency to our Bible, but very different beliefs).	obama-birth-certificate,religion	chain-email			none	11	43	8	5	105
3  7891.json	false	Says Having organizations parading as being social welfare organizations and then being involved in the political combat harkens back to why the statute a hundred years ago said that they were prohibited.	campaign-finance,congress,taxes	earl-blumenauer	U.S. representative	Oregon	democrat	0	1	1	1	0	a U.S. Ways and Means hearing
4  8169.json	half-true	Says nearly half of Oregons children are poor.	poverty	jim-francesconi	Member of the State Board of Higher Education	Oregon	none	0	1	1	1	0	an opinion article
5  929.json	half-true	On attacks by Republicans that various programs in the economic stimulus plan are not stimulative, "If you add all that stuff up, it accounts for less than 1 percent of the overall package."	economy,stimulus	barack-obama		President	Illinois	democrat	70	71	160	163	9	interview with CBS News
6  9416.json	false	Says when armed civilians stop mass shootings with guns, an average of 2.5 people die; otherwise, an average of 18 people die.	guns	jim-rubens	Small business owner	New Hampshire	republican	1	1	0	1	0	in an interview at gun shop in Hudson, N.H.
7  6861.json	true	Says Tennessee is providing millions of dollars to virtual school company for results at the bottom of the bottom.	education,state-budget	andy-berke	Lawyer and state senator	Tennessee	democrat	0	0	0	0	0	a letter to state Senate education committee chairwoman Dolores Gresham.
8  1122.json	false	The health care reform plan would set limits similar to the socialized system in Britain, where people are allowed to die if their treatment would cost more than $22,000.	health-care	club-growth			none	4	5	4	2	0	a TV ad
9  13138.json	true	Says Donald Trump started his career back in 1973 being sued by the Justice Department for racial discrimination because he would not rent apartments in one of his developments to African-Americans.	candidates-biography,diversity,housing	hillary-clinton	Presidential candidate	New York	democrat	40	29	69	76	7	the first presidential debate
10 1880.json	half-true	Bill White has a long history of trying to limit or even disenfranchise military voters.	military	republican-party-texas		Texas	republican	3	1	1	3	1	an e-mail
11 12803.json	half-true	John McCains chief economic adviser during the 08 race estimated that Trumps promises would cause America to lose 3.5 million jobs.	economy	tim-kaine	U.S. Senator	Virginia	democrat	8	3	15	15	0	a speech at the Democratic National Convention in Philadelphia
12 5409.json	false	Says 21,000 Wisconsin residents got jobs in 2011, but 18,000 of them were in other states.	job-accomplishments,jobs,states	kathleen-vinehout			democrat	1	1	1	1	0	remarks
13 7313.json	half-true	State revenue projections have missed the mark month after month.	state-budget	steve-henson		State Senator	Georgia	democrat	0	0	1	0	0	a press release
```

5.1.1 Dataset

5.2 CODE SNIPPET:

The Jupyter notebook will be used for implementing our machine learning algorithm and it has many files including dataset files and python notebooks which has following extensions I.e. ".tsv" ".pynb" .

We also tried to use python libraries like torch and the famous numpy. A small level implementation of our project is shown below.

In [*]:

```
# import the required packages here
import sys
sys.path.append('../')

from data import train_data_prepare
from train import train
from test import test, test_data_prepare

def run(train_file, valid_file, test_file, output_file):
    '''The function to run your ML algorithm on given datasets, generate the output and save them into the provided

    Parameters
    ----------
    train_file: string
        the path to the training file
        valid_file: string
                the path to the validation file
        test_file: string
                the path to the testing file
    output_file: string
        the path to the output predictions to be saved
    '''


    # read data from input
    train_samples, word2num = train_data_prepare(train_file)
    valid_samples = test_data_prepare(valid_file, word2num, 'valid')

    # your training algorithm
    model = train(train_samples, valid_samples, word2num)

    # your prediction code
    test(test_file, output_file, word2num, model)

    # define other functions here


run('train.tsv', 'valid.tsv', 'test.tsv', 'predictions.txt')
```

5.2.1 Dataset files

Chapter 6

Code Implementation

6.1 Implementation Environment

6.2 Program/Module Specification

6.3 Coding Standards

6.0 IMPLEMENTATION

6.1 IMPLEMENTATION ENVIRONMENT

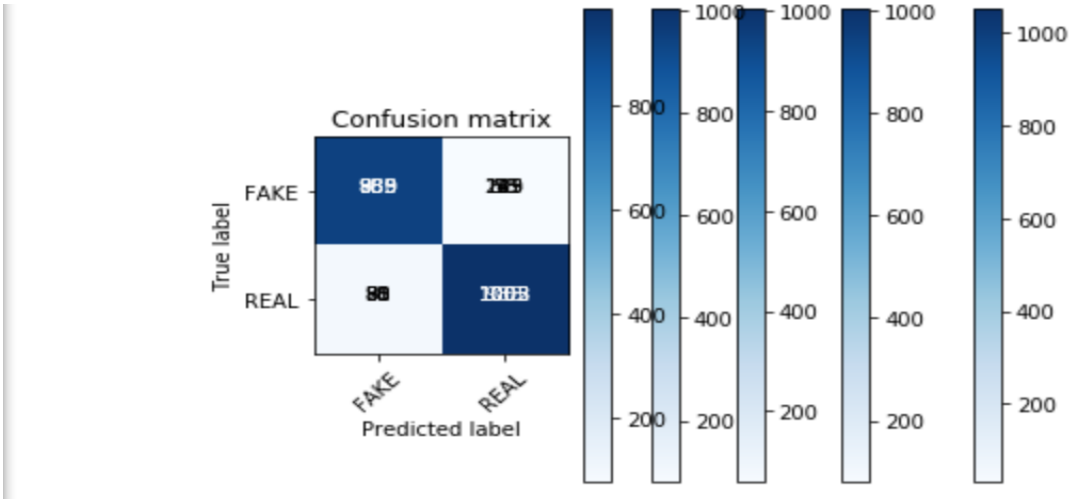
As our project is study based project and the best tool which is used at the undergraduate level is “Anaconda” . It consists of different modules in which we can code but for our project we have used Jupyter Notebook, which is used for high level python programming. Jupyter Notebook provides browser environment as it opens up in the *browser.it* can also connect to kernel and terminal.

6.2 PROGRAM/MODULE SPECIFICATION

The naive bayes classifier algorithm is the most applicable algorithm to implement fake news detection as it works on conditional probability and other major concepts of Data mining that are used in this project and we have also studied it in 7th semester which made the understanding of code quite easy.

```
#---------------------------------------------------------------
# Naive Bayes classifier for Multinomial model
#---------------------------------------------------------------

clf = MultinomialNB()

clf.fit(tfidf_train, y_train)                        # Fit Naive Bayes classifier according to X, y

pred = clf.predict(tfidf_test)                      # Perform classification on an array of test vectors X.
score = metrics.accuracy_score(y_test, pred)
print("accuracy:   %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'])
print(cm)


clf = MultinomialNB()

clf.fit(count_train, y_train)

pred = clf.predict(count_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy:   %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'])
print(cm)

```

The final output is generated with the help of matplot lib of python which helps to understand the count of various words in the articles circulated. Below it is shown with the help of color coding in the given matrix and respective range.



6.3 CODING STANDARDS

Normally, good software development organization requires their programmers to adhere to some well-defined and standard style of coding called coding standard.

6.3.1 Variable Standards:

Our project implementation uses apt variable names that makes the understanding of the domain quite easy.

6.3.2 Comment Standards:

Comments increases readability of our code and makes it easy for the third party to understand it. We have used comments everywhere needed and also used the references of the online codes.

Every code block and the different modules start with the comments, describing in brief about the code and the details.

Comments may also be used in between and along with the lines of code to explain one specific line or lines.

In python we can use. '#' to for single comment and for multiple lines we can use delimiters that is," ' ' ' " . We have used both during programming.

Chapter 7

Testing

7.1 Testing Plan

7.2 Testing Strategy

7.3 Testing Methods

7.4 Test Cases

7.0 TESTING

Various parameters like implementation environment, program modules and coding standards are explained in previous chapter while this chapter is aimed to provide brief account of testing the software.

There are two principal motives of testing the software

1. To rectify the error in execution
2. To check the viability of software

The testing ensures that the software is according to the required specification standards and performs the task meant for it. The testing is done by our in house employee that act as novice user and test the application with all possible way to find the bugs and error as well as check validation.

7.1 TESTING PLAN

Testing is carried out at the following three stages :

- Design
- Implementation
- Coding

7.1.1 Design Testing:

The design errors are to be rectified at the initial stage. Such errors are very difficult to repair after the execution of software.

7.1.2 Implementation Testing:

The errors occurred at this stage can't be overlooked because such errors do not allow the further process.

7.1.3 Coding Testing:

The coding procedure plays significant role in software designing. The improper coding of any software can generate inconsistent results. Such errors may occur due to incorrect syntax or false logic. If the errors at coding stage remain unnoticed may give rise to grave failure of the system.

7.2 TESTING STRATEGY

A strategy for software testing integrates software test case design method into a well-planned series of steps that result in the successful construction of the software.

The strategy provides the roadmap that describes the steps to be conducted as a part of testing, then these steps are planned and then undertaken, and how much effort, time and resource will be required.

- We have tested our whole system using bottom up testing strategy.
- Bottom up testing involves integrating and testing the modules to the lower levels in the hierarchy, and then working up hierarchy of modules until the final module is tested.
- Bottom up testing strategy shows how actual testing is to be done with whole system but it does not show any detail about each module testing.
- When all modules are tested successfully then I will move to one step up and continue with white box testing strategy.
- When all modules will be tested successfully then I will integrate those modules and try to test integrated system using black box testing strategy.

Why Black Box Testing in my Project?

In my project whatever I have implemented was going to be tested by guide Mr. Rajesh Davda so there was a black box testing involve directly.

7.2 TESTING METHOD

7.3.1 Unit Testing

The unit testing is meant for testing smallest unit of software. There are two approaches namely bottom-up and top-down.

In bottom up approach the last module is tested and then moving towards the first module while top down approach reverses the action. In present work we opt for the first one.

The bottom up approach for the current project is carried out as shown in.

7.3.2 Integration Testing

The integration testing is meant to test all the modules simultaneously because it is possible that all the modules may function correctly when tested individually. But they may not work altogether and may lead to unexpected outcome.

7.3.3 Validation Testing

After the integration testing software is completely assembled as a package, interfacing error have been uncovered and corrected, and then validation testing may begin. Validation can be defined in many ways but a simple definition is what a validation succeeds when software functions in a manner.

7.3.4 Storage Testing

The dataset of the system has to be stored on the hard disk. So the storage capacity of the hard disk should be enough to store all the data required for the efficient running of the software.

7.4 TEST CASES

7.4.1 Purpose

The purpose of this project is to use machine learning algorithm to detect the fake news in online social media that travels as a real one, it is like a click bait. It will try to enhance the user experience on the online social media platform and will also save lot of time of users that they might spent on fake news otherwise.

Chapter 8

Limitations and Future Enhancement

8.1 Limitations

8.2 Future Enhancement

8.1 LIMITATIONS:

Though we tried our best in developing this domain but as limitations are mere parts of any system so are of our system. Some limitations of our domains is:

- The present software uses high quality external hardware at input level. If quality of input document is poor, output may suffer due to limitation of it.
- The platform used is ANACONDA (JUPYTER NOTEBOOK) which is an open source software. This limits the cost of project.
- Limited dataset
- Limited processing speed
- When compared to real world applications our domains are not applicable as it is entirely study based.

8.2 FUTURE ENHANCEMENT:

There is always a scope for enhancements in any developed system, especially when our nature of the project is iterative which allows us to rethink on the method of development to adopt changes in the project. Below mentioned are some of the changes possible in the future to increase the adaptability, and efficiency of the system:

- Increase the dataset
- Increase the processing speed.
- Try to bring the domain as close as possible to the real world.
- Quality of dataset can be improved.

Chapter 9

Conclusion and Discussion

9.1 Self analysis and Project viabilities

9.2 Problem encountered and possible solutions

9.3 Summary of project

9.1 SELF ANALYSIS AND PROJECT VIABILITIES

This shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. We achieved classification accuracy of approximately 74% on the test set which is a decent result considering the relative simplicity of the model. These results may be improved in several ways, that are described in the article as well. Received results suggest, that fake news detection problem can be addressed with artificial intelligence methods.

9.2 PROBLEM ENCOUNTERED AND POSSIBLE SOLUTIONS:

9.2.1 Resource Availability:

An important part of checking the veracity of a specific claim is to evaluate the stance different news sources take towards the assertion. Automatic stance evaluation, i.e. stance detection, would arguably facilitate the process of fact checking.

9.2.2 Requirement Understanding:

Automatic fake news detection is a challenging problem in deception detection, and it has tremendous real-world political and social impacts. However, statistical approaches to combating fake news has been dramatically limited by the lack of labeled benchmark datasets.

9.2.3 Problem Encountered and Possible Solutions:

Problem:

Fake news

Solution:

To detect fake news and analyze it.

9.3 SUMMARY OF PROJECT

The scourge of cyberbullying has assumed alarming proportions with an ever-increasing number of adolescents admitting to having dealt with it either as a victim or as a bystander.

Anonymity and the lack of meaningful supervision in the electronic medium are two factors that have exacerbated this social menace.

Fake news is a phenomenon which is having a significant impact on our social life, in particular in the political world. Fake news detection is an emerging research area which is gaining interest but involved some challenges due to the limited amount of resources available.

We propose in this paper, a fake news detection model that use machine learning techniques. We investigate and compare two different features extraction techniques and six different machine classification techniques.

Experimental evaluation yields the best performance using Term Frequency Inverted Document Frequency (TFIDF) as feature extraction technique, and Linear Support Vector Machine (LSVM) as a classifier, with an accuracy of more than 74%.

We find that binary classifiers for individual tablets outperform multiclass classifiers. Our findings show that the detection of textual cyberbullying can be tackled by building individual topic sensitive classifiers.

REFERNCES:

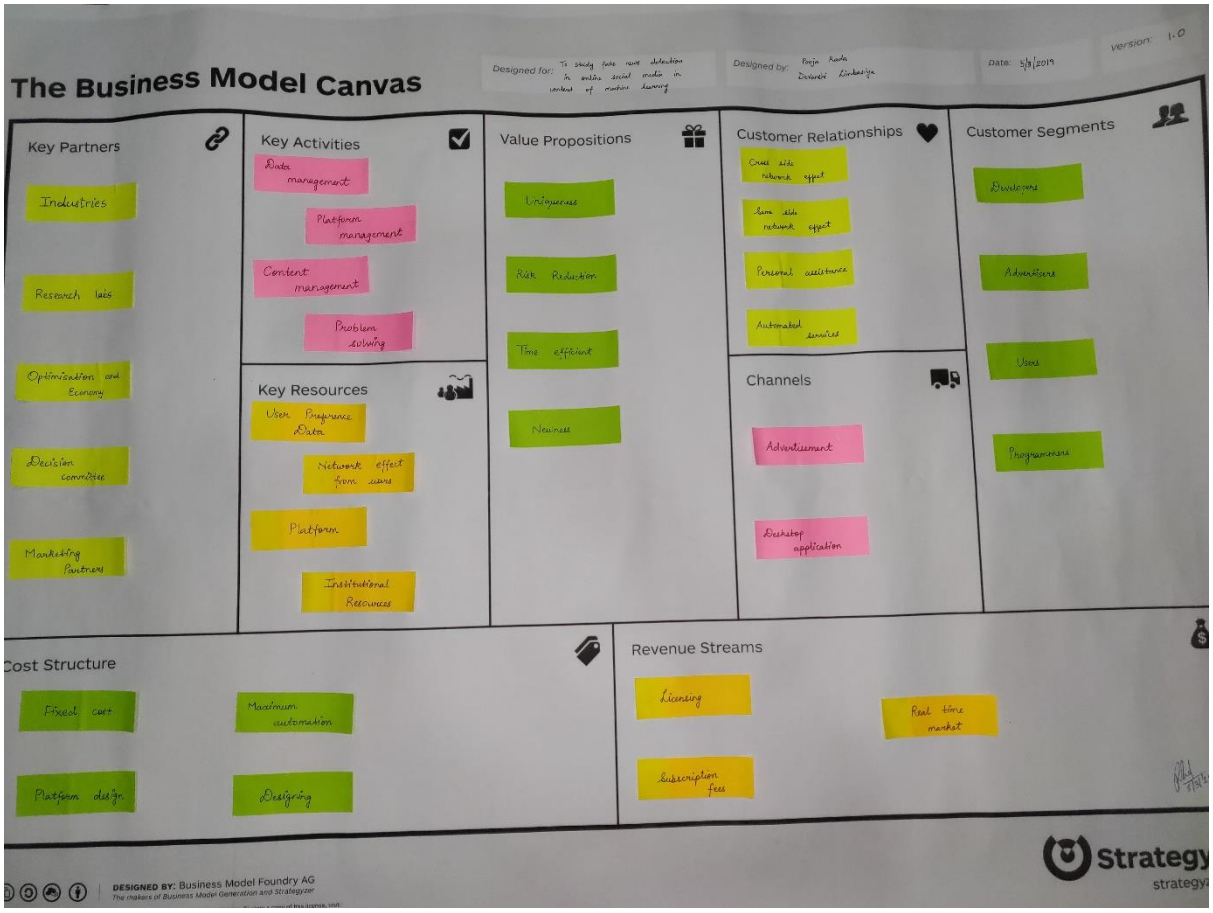
https://patents.google.com
https://patentimages.storage.googleapis.com/6c/34/81/c390e0d0b7a340/US8185448.pdf
https://patents.google.com/patent/CA2984904A1/en?oq=CA2984904A1
https://patents.google.com/patent/US20130073473A1/en?q=Social+media&q=events&q=detection&q=verification&oq=Social+media+events+detection+and+verification
https://patents.google.com/patent/CN102929918A/en?q=False&q=online&q=public+opinion&q=identification&q=method&oq=False+online+public+opinion+identification+method
https://patentimages.storage.googleapis.com/a9/be/27/26f147f349b34a/US6131085.pdf
https://patentimages.storage.googleapis.com/fa/f7/07/6ea3c92236263e/US6807558.pdf
https://patentimages.storage.googleapis.com/80/42/4a/bd509910cfedc1/US5877759.pdf
https://patentimages.storage.googleapis.com/56/2d/6e/dcbce2b8a3124d/US8194986.pdf
https://arxiv.org/abs/1705.00648
https://link.springer.com/chapter/10.1007/978-3-319-69155-8_9
https://www.aclweb.org/anthology/W16-0802
https://arxiv.org/abs/1707.03264
https://ieeexplore.ieee.org/abstract/document/8100379/

APPENDIX

A.1 Business Model Canvas
A.2 Exhibition of Canvas
A.3 Periodic Progress Reports
A.4 Patent Drafting Exercise Report
A.5 BMC Report

A.1 BUSINESS MODEL CANVAS:

This canvas teaches us to view our project as business. Which are our partners, activities, value propositions, customer relationship, cost structure and revenue.



A.2 EXHIBITION OF CANVAS :

An exhibition of these canvas made by all the teams was carried out at the Computer Department of Atmiya Institute of Technology And Science.

Event started at 10:00 am on 23st of March. All the teams gave presentation on their respective canvas and Present running project.

Our respected Principle DR. G. D. Aacharya along with our H.O.D Madam Tosal Bhalodiya and all respective faculties and Professor Mr. Rajesh Davda inspected BMC canvas and gave valuable advices on our project . A healthy participation from the students gave their opinion toward the enrichment of the project canvas and project.

Based on advices and suggestion we have reflected changes relatively.

A.3 PERIODIC PROGRESS REPORT:

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Limbasiya Devanshi Vallabhbhai
EnrollmentNo : 150030107054 Department : Computer Engineering
MobileNo : 8141135110 Discipline : BE
Email : devanshi208@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : First PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We are currently working on implementation of navies bayes algorithm that we have studied in our academic subject (DMBI).

2. What challenge you have faced ?

We are finding it difficult to implement algorithm as there are various tedious steps.

3. What support you need ?

We would like to have a better understanding of algorithm.

4. Which literature you have referred ?

We have referred to a lot of websites that explains naive bayes in detail, websites like,
https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/
https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ etc.

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Limbasiya Devanshi Vallabhbhai
EnrollmentNo : 150030107054 Department : Computer Engineering
MobileNo : 8141135110 Discipline : BE
Email : devanshi208@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Second PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We are in the implementation phase of our project, we have completed almost 30% of implementation of algorithm.

2. What challenge you have faced ?

We are having lot of errors and bugs in the implementation.

3. What support you need ?

We would like to have guidance on how to solve these errors as they are the huge block in our implementation.

4. Which literature you have referred ?

We have referred some of the personal blogs of people that might have faced the same error. one of it is , https://machinelearningmastery.com

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Limbasiya Devanshi Vallabhbhai
EnrollmentNo : 150030107054 Department : Computer Engineering
MobileNo : 8141135110 Discipline : BE
Email : devanshi208@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Third PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We have tried to solve the errors that were occurring in our implementation.

2. What challenge you have faced ?

We solved some errors but majority of the bugs are unsolved.

3. What support you need ?

We would like to have better understanding go the errors in order to solve them.

4. Which literature you have referred ?

We tried to solve it with our best understanding and concepts that we know and have taken the guidance of some people in our group.

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Limbasiya Devanshi Vallabhbhai
EnrollmentNo : 150030107054 Department : Computer Engineering
MobileNo : 8141135110 Discipline : BE
Email : devanshi208@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Forth PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We did the patent drafting exercises of PMMS(GTU), to learn and understand how the patent is draft and what kind of documents is required and at what places the patent office is available.

2. What challenge you have faced ?

We did not face as big of a challenge while doing the exercise, we just did not understand few terms.

3. What support you need ?

We dont need any support in this activity as we got everything after googling it out a few times and after watching several videos.

4. Which literature you have referred ?

We have referred to several videos and have googled out some of the terms that were hard to understand while doing the PDE of GTU PMMS module, some of the links are as follows: https://www.youtube.com/watch?v=57IdaH8Wj1g https://www.youtube.com/watch?v=TCSnj1ukOfs

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

1/1

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Pooja Rada
EnrollmentNo : 150030107075 Department : Computer Engineering
MobileNo : 9909994200 Discipline : BE
Email : bhavanarada@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : First PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We are currently working on implementation of navies bayes algorithm that we have studied in our academic subject (DMBI).

2. What challenge you have faced ?

We are finding it difficult to implement algorithm as there are various tedious steps.

3. What support you need ?

We would like to have a better understanding of algorithm.

4. Which literature you have referred ?

We have referred to a lot of websites that explains naive bayes in detail, websites like,
https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/
https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ etc.
https://www.saedsayad.com/naive_bayesian.htm

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Page 1 o

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Pooja Rada
EnrollmentNo : 150030107075 Department : Computer Engineering
MobileNo : 9909994200 Discipline : BE
Email : bhavanarada@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Second PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We are in the implementation phase of our project, we have completed almost 30% of implementation of algorithm.

2. What challenge you have faced ?

We are having lot of errors and bugs in the implementation.

3. What support you need ?

We would like to have guidance on how to solve these errors as they are the huge block in our implementation.

4. Which literature you have referred ?

We have referred some of the personal blogs of people that might have faced the same error. one of it is , https://machinelearningmastery.com

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Pooja Rada
EnrollmentNo : 150030107075 Department : Computer Engineering
MobileNo : 9909994200 Discipline : BE
Email : bhavanarada@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Third PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We have tried to solve the errors that were occurring in our implementation.

2. What challenge you have faced ?

We solved some errors but majority of the bugs are unsolved.

3. What support you need ?

We would like to have better understanding go the errors in order to solve them.

4. Which literature you have referred ?

We tried to solve it with our best understanding and concepts that we know and have taken the guidance of some people in our group.

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

Print Back

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
StudentName : Pooja Rada
EnrollmentNo : 150030107075 Department : Computer Engineering
MobileNo : 9909994200 Discipline : BE
Email : bhavanarada@gmail.com Semester : Semester 8

PPR Details

Periodic Progess Report : Forth PPR

Project : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

Status : Reviewed

1. What Progress you have made in the Project ?

We did the patent drafting exercises of PMMS(GTU), to learn and understand how the patent is draft and what kind of documents is required and at what places the patent office is available.

2. What challenge you have faced ?

We did not face as big of a challenge while doing the exercise, we just did not understand few terms.

3. What support you need ?

We dont need any support in this activity as we got everything after googling it out a few times and after watching several videos.

4. Which literature you have referred ?

We have referred to several videos and have googled out some of the terms that were hard to understand while doing the PDE of GTU PMMS module, some of the links are as follows: https://www.youtube.com/watch?v=57IdaH8Wj1g https://www.youtube.com/watch?v=TCSnj1ukOfs

Comments

Comment by Internal Guide :
Done

Comment by External Guide :
None

Comment by HOD :
None

Comment by Principal :
None

Comment by University Admin :
None

A.4 BMC REPORT:

Business Model Canvas:

This canvas teaches us to view our project as business. Which are our partners, activities, value propositions, customer relationship, cost structure and revenue

KEY PARTNERS

- **Finance Partners**

Who provide financial support for investment, purchasing row materials and maintenance cost etc.

- **Decision committee**

It is Main Staff of Company Who Takes Major decisions about company.

- **Row material partners**

It is an industry or organization who supplies row materials required for our product.
Here,

- Internet Company
- Mobile Company
- Software Company
- Digital Equipment Company

- **Marketing Partners**

This Partner plays main role to advertise the product to the consumers via different channels.
Here,

- Gmail
- Facebook
- Twitter

❖ KEY ACTIVITIES

- **Analytical Unit**

It does the analysis of complete problem.

- Which functionalities should be added?
- Which are the problems are solved?

➢ Designing Unit

It performs the designing of the software.

- Using Jupyter Notebook designing is done.

➢ Coding Unit

It performs coding of the portal.

- Database handling
- Upload the project
- See the project
- Backed on project

➢ Testing Unit

It is unit in which testing of product is done. It is helpful to maintain quality of working of portal.

- Testing is done by giving different inputs.

➢ Customer Service

This department provides consumer satisfaction by giving them service like guidelines about the portal.

- Guidebook
- Social Sharing
- Testing is performed time to time.
- User-friendly

➢ R&D Department

This dept. is helpful to develop product as per future technologies. It takes review of product to customers and upgrade portal with new features.

- By taking feedback

❖ KEY RESOURCES

➢ Software and Technologies:

Our main resources to develop the portal are included here.

- Text editor
- Anaconda
- Browser
- Internet connectivity
- Jupyter Notebook
- Python

❖ VALUE PROPOSITIONS

In this section we include the features which make our portal more attractive.

- Newness
- Customization
- Risk Reduction
- Performance
- Accessibility
- Usability

❖ CUSTOMER RELATIONSHIPS

It is all about how company maintains relationship with their customers by giving them facilities and services.

➢ Feedback& Service

It takes complains of customers and solve their problems.

- Feedback form

➢ Maintenance

It is a facility provided by company to recover the error in portal in minimum time period.

- Testing

➢ Advertisements

It is a process to show feature and information about product to the customers.

- Liking
- Sharing
- Facebook
- Twitter
- Gmail
- Through notification

❖ CHANNELS

In this section it includes the devices through which our portal can be accessible.

- Mobile Phone
- Laptop
- PC
- Tablet
- Advertisement

❖ CUSTOMER SEGMENTS

In this section it includes the customers of our product.

- Innovator
- Investor
- Student
- Professor
- Scientist
- Researcher
- Developer
- Users

❖ COST STRUCTURE

It includes cost that company has to make portal.

- Net usage
- Electric Usage

- Software purchasing cost

❖ REVENUE STREAMS

It includes the revenue that generate from this portal to company.
Here,

- Licensing
- Subscription fees
- Real time Market

A.5 PATENT DRAFTING EXERCISE:

College : ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT
Department : Computer Engineering
Discipline : BE
Semester : Semester 8
Project Name : To Study Fake News Detection In Online Social Media In Context Of Machine Learning.
Team ID : 29915

Form 1 – APPLICATION FOR GRANT OF PATENT

Applicants :

Sr. No	Name	Nationality	Address	Mobile No.	Email Id
1	Pooja Rada	Indian	Computer Engineering , ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT , Gujarat Technologycal University.	9909994200	bhavanarada@gmail.com
2	Limbasiya Devanshi Vallabhbhai	Indian	Computer Engineering , ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT , Gujarat Technologycal University.	8141135110	devanshi208@gmail.com

Inventors :

Sr. No	Name	Nationality	Address	Mobile No.	Email Id
1	Pooja Rada	Indian	Computer Engineering , ATMIYA INSTITUTE OF TECHNOLOGY & SCIENCE, RAJKOT , Gujarat Technologycal University.	9909994200	bhavanarada@gmail.com
2	Limbasiya Devanshi Vallabhbhai	Indian	Computer Engineering , ATMIYA INSTITUTE OF	8141135110	devanshi208@gmail.com

			TECHNOLOGY & SCIENCE, RAJKOT , Gujarat Technologycal University.		

I/We, the applicant(s) hereby declare(s) that:

Following are the attachments with the applications :

Form 2 - PROVISIONAL/COMPLETE SPECIFICATION

1 . Title of the project/invention :

To Study Fake News Detection In Online Social Media In Context Of Machine Learning.

2. Preamble to the description :

Provisional

3. Description

a) Field of Project / Invention / Application :

The applications of NLP (Natural Language Processing) techniques for detecting the fake news, that is, misleading news stories that comes from the non-reputable sources. Only by building a model based on a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) tfidf matrix, (word tallies relative to how often they're used in other articles in your dataset) can only get you so far. But these models do not consider the important qualities like word ordering and context. It is very possible that two articles that are similar in their word count will be completely different in their meaning.

b) Prior Art / Background of the Project / Invention :

We studied a lot of websites online and also tried to study different algorithms that are used in our project and would make it better.

c) Summary of the Project / Invention :

Owing to the growing popularity of social media, spammers have made these their prime target to mislead others with an intent to make financial or political gains. In this work, we have developed a fake news detection system which allows users to alter potentially deceptive information. The prediction that a particular news article is deceptive is based on the analysis of previously seen news as well as information available online from reliable sources. A scarcity of deceptive news for predictive modeling, is a major stumbling block in this field.information is shared/forwarded/re-tweeted without much thought.This work therefore aims to prevent fake news from trending on social media and enables us to make informed decisions especially during high impact events

d) Objects of Project / Invention :

The main objective of our project is to study fake news in social media and analyze it to see if we can have better control over it or not.

e) Drawings :

f) Description of Project / Invention : (full detail of project) :

Owing to the growing popularity of social media, spammers have made these their prime target to mislead others with an intent to make financial or political gains. In this work, we have developed a fake news detection system which allows users to alter potentially deceptive information. The prediction that a particular news article is deceptive is based on the analysis of previously seen news as

well as information available online from reliable sources. A scarcity of deceptive news for predictive modeling, is a major stumbling block in this field.information is shared/forwarded/re-tweeted without much thought.The applications of NLP (Natural Language Processing) techniques for detecting the fake news, that is, misleading news stories that comes from the non-reputable sources. Only by building a model based on a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) tfidf matrix, (word tallies relative to how often they're used in other articles in your dataset) can only get you so far. But these models do not consider the important qualities like word ordering and context. It is very possible that two articles that are similar in their word count will be completely different in their meaning.This work therefore aims to prevent fake news from trending on social media and enables us to make informed decisions especially during high impact events

g) Examples :

h) Claims (Not required for Provisional Application) / Unique Features of Project

The unique feature of our project is that of analysis, as we are trying to implement it after studying others and would make it more effective.

4. Claims

5. Date and signature

6. Abstract of the project / invention :

Owing to the growing popularity of social media, spammers have made these their prime target to mislead others with an intent to make financial or political gains. In this work, we have developed a fake news detection system which allows users to alter potentially deceptive information. The prediction that a particular news article is deceptive is based on the analysis of previously seen news as well as information available online from reliable sources. A scarcity of deceptive news for predictive modeling, is a major stumbling block in this field.information is shared/forwarded/re-tweeted without much thought.

This work therefore aims to prevent fake news from trending on social media and enables us to make informed decisions especially during high impact events

Form 3 – STATEMENT AND UNDERTAKING UNDER SECTION 8

Name of the applicant(s) : I/We, Pooja Rada ,Limbasiya Devanshi Vallabhbhai

Hereby declare :

Name,Address and Nationality of the joint applicant : (i) that I/We have not made any application for the same/substantially the same victim invention outside India.

(ii) that the rights in the application(s) has/have been assigned to

Name of the Country	Date of Application	Application Number	Status of the Application	Date of Publication	Date of Grant
N/A	N/A	N/A	N/A	N/A	N/A

(iii)That I/We undertake that upto the date of grant of the patent by the Controller, I/We would keep him informed in writing the details regarding corresponding applications for patents filed outside India within three months from the date of filing of such application.

Dated this 5 day of April 2019

To be signed by the applicant Signature.....

or his authorised registered patent agent :	
Name of the Natural Person who has signed :	Pooja Rada ,Limbasiya Devanshi Vallabhbhai

To,
The Controller of Patents,
The Patent Office,
At Mumbai