**Detection prototype of Anti-Pakistan text using ML**

Ahmad hassan S19-1125

Bisma Riaz S19-1018

Ameer Hamza Azam S19-1078

Faraz Ali F18-1025



**October 2022**

Department of Computer Science

Lahore Leads University

Lahore – PAKISTAN

## CHECKLIST:

|  |  |  |
| --- | --- | --- |
| **MEMBER NAME** | **REG. NO.** | **EMAIL ADDRESS** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

#### Submission Form for Final-Year

PROJECT REPORT

|  |  |
| --- | --- |
| **PROJECT ID** | **NUMBER OF MEMBERS** |
|  |  |
| **TITLE** |  |
|  |  |
| **SUPERVISOR NAME** |  |

|  |  |
| --- | --- |
| Number of pages in this report |  |
| I/We have enclosed the soft-copy of this document along-with the  codes and scripts created by myself/ourselves | **YES / NO** |
| My/Our supervisor has attested the attached document | **YES / NO** |
| **I/We confirm to state that this project is free from any type of plagiarism and misuse of copyrighted material** | **YES / NO** |

|  |
| --- |
| **MEMBERS’ SIGNATURES** |
|  |
|  |
| Supervisor’s Signature |

*Note 1: This paper must be signed by your supervisor*

*Note 2: The soft-copies of your project report, source codes, schematics, and executables should be delivered on a CD/DVD*

*Lahore Leads University, Lahore Department of Computer Science*

#### This work, entitled “Insert Project Title Here” has been approved for the award of

**Bachelor of Science in Computer Science**

Date

## External Examiner:

**FYP Coordinator:**

## Head of Department:

#### Department of Computer Science

LAHORE LEADS UNIVERSITY

LAHORE – PAKISTAN

# Declaration

*“No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university/institute or other institution of learning”.*

**MEMBERS’ SIGNATURES**

# Acknowledgments

We are enormously gratified by successfully completing our final year project on “Detection Prototype of Anti-Pakistan Text using ML”. We take this opportunity to thank all our faculty members and mentors who took a huge part in our progress. We would especially like to thank Engr. M. Nughman helped in completing the project with valuable suggestions and feedback ensuring our direction is correct in finalizing the project. We would like to thank our friends for helping us with their expertise in Machine learning for optimizing the Anti-text system for Pakistan. Above all, we are grateful to our parents and Allah Almighty.

# Abstract

These days, there are a ton of discussion forums available online that are created to enable people to interact, debate, and express their ideas and perspectives on various themes. For instance, individuals may generally express their thoughts by leaving comments on news websites, blogs, and social networking platforms like YouTube, Facebook, Twitter, and Instagram. It is well knowledge that user interactions in these forums may quickly degenerate and become improper, such as calling each other names or making crude jokes about respecting one's nation and its people. In the end, inappropriate words or postings are developing into a cyber-criminal, slowly harming the image of our country abroad. To enhance the quality of dialogues on social media platforms and put a stop to anti-state texts, it is thus crucial to monitor and manipulate insensitive writings. In this research, we develop an experimental machine learning method to automatically identify anti-Pakistan writings. The inappropriate messages are instantly identified by it. Since identifying improper communications is a very challenging procedure, we want to use incorrect messages to train a first prototype that will aid in accurately identifying the messages. This model can help with accurate anti-state communication identification.

# Table of contents

**CHAPTER 1** ................................................................................................................................

**1 INTRODUCTION**..................................................................................................................1

1.1 PROBLEM STATEMENT................................................................................................2

1.2 THESIS OBJECTIVE .......................................................................................................3

1.3 JUSTIFICATION ..............................................................................................................3

1.4 SCOPE................................................................................................................................3

**CHAPTER 2** .................................................................................................................................4

**2 LITERATURE REVIEW**........................................................................................................4

2.1 PROJECT SIGNIFICANCE................................................................................................4

2.2 STRUCTURE COMPONENTS..........................................................................................5

2.3 Text Classification: .............................................................................................................5

2.4 Detecting inappropriate Text..........................................................................................….6

2.5 User Conversational messages: ...........................................................................................7

2.6 Peace-building......................................................................................................................8

2.7 Religious militants and terrorism.........................................................................................9

2.8 Social Media.......................................................................................................................10

2.9 POPULAR SOCIAL MEDIA SITES:..................................................................................10

2.9.1 Facebook.......................................................................................................................10

2.9.2 Twitter...........................................................................................................................11

2.9.3 Google+..................................................................................................................…...11

2.9.4 YouTube .......................................................................................................................11

2.9.5 Pinterest........................................................................................................…..……...11

2.9.6 Instagram ......................................................................................................................12

2.9.7 WhatsApp.........................................................................................................….........12

2.9.8 Snapchat.................................................................................................... ……...........12

**CHAPTER 3** ..................................................................................................................................13

**3 METHODOLOGY AND INTERACTION OF COMPONENTS**.....…....................…......13

3.1 Methodology..................................................................................................................…....13

3.2 Tools, Mechanisms, and Software ...............................................................................….....13

3.2.1 HTML:.................................................................................................................….....13

3.2.2 CSS:...............................................................................................................….…......14

3.2.3 PyCharm:...................................................................….........…........................…......15

3.2.4 Jupyter Notebook: ....................................................................…...........….................16

3.3 Machine Learning: ...............................................................................….......................17

3.3.1 Machine Learning vs. Deep Learning vs. Neural Networks .............….......................18

3.3.2 How Machine Learning Works............................................................…….................19

3.3.3 Machine Learning Methods..................................................................……................20

3.3.4 Real-world machine learning use cases................................................…….................21

3.3.5 Challenges of Machine Learning...........................................................……................22

3.4 Artificial Intelligence ............................................................................................…..............24

3.4.1 Goals of AI................................................................................................……..............25

3.4.2 Reasoning and problem-solving.......................................................................…...........26

3.4.3 Planning........................................................................................................……...........26

3.4.4 Learning......................................................................................................…….............26

3.5 Dataset building .............................................................................................…......................27

3.6 Data Preprocessing ............................................................................…..................................28

3.7 Model Training...................................................................................................….................29

3.8 Model Evaluation....................................................................................………............31

3.8.1 Random Forest Classifier ..........................................................................….......32

3.8.2 Python Implementation of Random Forest Algorithm ........................................34

**CHAPTER 4** .............................................................................................................…….......35

**4 DESIGN AND IMPLEMENTATION** …...................................................................…35

4.1 Designing of Prototype ......................................................................................…......35

4.1.1 Flask Framework ................................................................................……........35

4.2 Libraries for Prototype..............................................................................……….....….36

4.2.1 pandas....................................................................................................................36

4.2.2 Numpy....................................................................................................................37

4.2.3 Matplotlib ...............................................................................................................37

4.2.4 Seaborn....................................................................................................................38

4.2.5 Scikit-learn .............................................................................................................39

4.2.6 NLTK .....................................................................................................................39

4.2.7 jinja..........................................................................................................................40

4.3 Twint Web Scrapping .....................................................................................................40

4.4 Model Deployment..........................................................................................................41

4.4.1 Introductory Page .........................................................................................…....42

4.4.2 About Us page........................................................................................................42

4.4.3 Project Page ...........................................................................................................43

4.4.4 Desired Output.......................................................................................................44

4.5 Coding for Implementation...............................................................................................44

**CHAPTER 5** ..........................................................................................................................49

**5 FUTURE WORK AND CONCLUSION** ........................................................................49

5.1 Future Scope ..................................................................................................................49

5.2 Conclusion or Result .....................................................................................................50

**List of Figures**

Figure 3-1: HTML language working 14

Figure 3-2: CSS language working 15

Figure 3-3: PyCharm Software 16

Figure 3-4: Jupyter Notebook Working 17

Figure 3-5: Dataset building using Twint library 28

Figure 3-6: Data preprocessing graph 28

Figure 3-7: Data preprocessing conversion graph 1 29

Figure 3-8: Data preprocessing conversion graph 2 29

Figure 3-9: Model Training Example 29

Figure 3-10: Model Training Example 30

Figure 3-11: Model Evaluation Methods 31

Figure 3-12: Random Forest Classifier working method 32

Figure 4-1: Model Deployment 41

Figure 4-2: UI of prototype 42

Figure 4-3: About our Supervisor 43

Figure 4-4: Project UI page 44

Figure 4-5: Desired Output 45

**List of Abbreviations**

**AI = Artificial Intelligence**

**ML = Machine Learning**

**UI = User Interface**

**UX = User Experience**

**URL = Uniform Resource Locator**

# Chapter 1

# Introduction:

Following those tragic incidents of 11th September 2001, a lot has been made of the possible connections between fighting terrorists and peace-building combat against terrorism and peace-building had in general, continued to carry out separately and by various sets of actors. The events of 11th September could have prompted some of the most powerful states in the world to reconsider the role of the threat of terrorism to security but it's not likely to significantly alter the nature of peace-building and peace-building tasks, ranging from building institutions and democratic structures to post-conflict reconstruction and rehabilitation. It's not unexpected that how the threat of terrorism is dealt with by those who are involved in peace-building is typically restricted to the potential impact on the security conditions to conduct their work. This is viewed as a challenge that must be addressed by the security aspect of the mission or by an ad-hoc international security force, or even by the national security structure (if there are any). There is a certain amount of reluctance towards the fight against terrorism on behalf of those in the peacebuilding community does not seem unfounded, and could be thought of as a natural response to the declaration made after the 11th September of 2001, of a worldwide "war on terrorism" that transcends traditional anti-terrorist goals and demands. In reality, the negative consequences of this global campaign result from the lack of clarity regarding its purpose and objectives.

[1] In light of this, it's not surprising that the effect of the war against terrorism on the various post-conflict and conflict areas across the globe has been mixed at most. One of the main platforms that have risen all conflicts against any state is social media. Social media is a way for any state could alter the image of another state before the world. It is also possible for terrorists to spread through it since it could serve as a method for a terrorist to convert. It is therefore essential for every state to keep an eye on social media accounts for the security and sovereignty of its citizens. Inappropriate text messages state are the most effective way to undermine the integrity of the nation.

[2] Thus, Anti-Pakistan text messages are one of the biggest issues Pakistan is facing. Text messages are also considered to be one of the most effective ways to deter Terrorist attacks across the globe, but especially in Pakistan. Pakistan is always featured in the highlights of terrorist attacks since 2000 until today. However, there has been a gradual decrease in the number of attacks, as a wide array of security measures was implemented to curb them.

[3] However, there's always an issue in securing the messages between terrorists, but it's nearly impossible to detect the messages. The primary focus of our research is the development of an application that can detect such messages. The solution is to create an Anti-Pakistan text messaging system based on machine learning. In this paper, we propose the design of a system that quickly detects the type of communication occurring to damage the security of Pakistan. For instance, if a terrorist contacts their criminal accomplices using Phrases or hidden words, the system becomes active and instantly detects the messages. In this manner, attacks are able to be controlled in Pakistan.

[4] The system is created by using machine learning. Machine learning algorithms utilize previous data to forecast new output values. Text classifiers can be utilized to organize structure and categorize almost all kinds of text in the form of messages of theft.

### Problem Statements:

In today's Digital World keeping eye on social media is vital and the government has started to organize Social Media Monitoring Units. It requires a significant amount of money and a special budget. keep track of all social media platforms it's an extremely difficult task. There has to be a method or prototype to will automatically monitor all platforms and then manipulate the accounts, or make posts immediately. The negative messages against the state directly undermine the integrity and sovereignty of the nation in front of the entire world. This can affect every aspect of the country, including lifestyle, tourism, business, and so on.

As Pakistan is among the countries that are most well-known from all angles and even a single unsuitable message could create an impression that is not right. Therefore, it is essential to have a method to identify state-sponsored propaganda automatically, which will protect our reputation from all angles and monitors social media in a separate manner.

### Aims and Objectives:

The main objective of the study is to develop an Anti-Pakistan text detection system to proactively detect and notify the relevant messages to the higher authority.

The following specific objectives have been formulated:

• To identify the Anti-Pakistan messages

• To review every communication or tweet on social media

• To develop a system that detects the messages and notify properly

### Scope of Project:

The research aims to develop a model for anti-state text messages found in social media. This is due to the fact that solutions could be a significant step towards improving accountability, proactive adaptability, and better use of social media when monitored, which will result in the safety and cleanliness of the social media environment for all. The goal of this study is to the vast size of Pakistan that is operating all day and night to protect Pakistan as well as its security. The concept developed in this study will allow users to keep track of social media over a certain period of time and increase its usage.

# Chapter 2

# Literature Review:

This section provides a summary of the literature regarding the Anti-Pakistan detection of text messages to help identify the elements to take into consideration in the Pakistan defense system. It will also and to identify areas of need in the literature, develop a conceptual and theoretical framework and provide a base for the analysis of empirical data. The results of the data will be utilized to add significance or references to the research that will be used in the writing of the thesis and the development of the anti-Pakistan text message system.

**2.1 PROJECT SIGNIFICANCE:**

The aim of this endeavor is to design an experimental prototype that can work for us to identify inappropriate content on social media. Also, it must be an easy-to-use prototype that any layman can handle. [6] People's habits with regard to technology have drastically changed over the span of a few months. The number of homes that have Internet broadband has dramatically grown, and the number of regular Internet users (i.e. people who log on every week at least) has also grown. Nowadays, everyone has a regular Internet connection, whether it be for social media use, academic purposes, or simply for leisure. The majority of young people use the Internet to send and receive instant messages, and emails, participate in social networks, listen to the radio, view video content from sharing services (like YouTube), and research products and services, according to a recent Eurostat survey on Internet access and use statistics for households and individuals. As in the real world and the real world, the Internet contains a variety of risky sites in which a single comment can be damaging to the country. In a recent study, the primary online dangers for children include interactions in a non - threatening manner with people who are not their friends (i.e. unwelcome friend requests or messages that are offensive) as well as inappropriate messages with the intent of causing violence, hate, and bullying. 5 To prevent anyone from posting inappropriate content this includes the ability to monitor the apps the minor uses to monitor his or her connection's time and frequency as well as the messages the minor receives and sends. Thus, there is a necessity for a model that can safeguard everyone as well as the integrity of the nation.

**2.2 STRUCTURE COMPONENTS:**

Structure components of the prototype to develop are very important and their fast and efficient working is an advantage because they must be fast in order of their fast working.

A detail of the components and software requirements are as follows:

• System: 64-bit system with 8 GB RAMS.

• Platform: Windows 10 Pro.

• Programming Language: Python.

• Python Version: Python Version – 3.7.6

• Python Distribution Platform: Pycharm

• Programming Environment: Jupyter Notebook.

• Python Library: Scikit-learn.

• Computing Library: Numpy.

• Visualization Library: Matplotlib.

• Package: Pandas.

• Machine Learning Framework: TensorFlow.

• Micro Web Framework: Flask.

• Web Template Engine: Jinja2.

• Front End Development: HTML, CSS, JavaScript.

• Cloud Application Platform: Heroku. (For deployment of GUI system).

• Repository Hosting Service: GitHub.

• Version Control Service: Git.

These are the requirements and components that must be installed on the computer and with the help of these packages the prototype I developed.

**2.3 Text Classification:**

The volume of data (i.e. images, text, and video) accessible via the Internet has increased exponentially over the past few years, causing the issue of overloading in multimedia environments present. However, the time and resources of users aren't sufficient to handle all the data. This has led to a desire to develop automatic text classification systems. They 6 were initially used initially to retrieve information and later for information filtering, text categorization, recommendation systems, sentiment analysis, and document 12 summarization and document summarization, among others. In addition, these methods have been employed in numerous fields like engineering, medicine, psychology, and business, among others. [8]

Most processes for determining text can be classified into five phases:

(a) Data acquisition and labeling,

(b) Preprocessing,

(c) Feature extraction or text encoding

(d) Reduction in dimensionality,

(e) Classifier training and

(f) Evaluation.

Different studies address various issues which arise within this area.

**2.4 Detecting inappropriate Text:**

Recently, the science of machine learning has seen a rise in interest in the subject of identifying objectionable information (such as video messages, photos, or videos). Various efforts have been made, in particular, to remove undesirable text content using ML approaches.

Text mining techniques based on the combination of fundamental (binominal) classifiers are available to weed out content that is improper for the web. A few techniques for website analysis that provide data in other languages are also suggested.

The issue of identifying foul language in user-to-user interactions and erroneous search phrases has been addressed in. In this study, the authors propose deep learning models that, when applied to analyze real-world search queries and conversations, dramatically outperform the manually produced pattern-based and other features-based baselines.

Alternatives evaluate comments that are unsuitable for news. In particular, the impact of text distortion is assessed as a consequence of performance improvements.

Threatening or abusive language, as well as comments that are offensive due to a person's gender, religion, or race, are instances of improper remarks. To address the issue of the inappropriateness of user-generated content posted on the Internet, certain strategies are being created. As a consequence, the study of hate speech is a particular emphasis of many programmers. For instance, Hammer used an LR model to find violent material in 24840 manually annotated lines gathered from comments on YouTube against gays, minorities, and immigration. Only 5% of nonviolent texts could be identified as violent, yet the classifier made mistakes in that category at a rate of around 10%. [9].

Others tackle the problem of cyberbullying detection on Twitter. Thus, Chen et al. compared two pre-trained word embedding techniques (i.e., Glove and Word2Vec) for text encoding as well as well-known techniques like TF-IDF. It turned out that the simple TFIDF outperformed the more complex word embedding approaches. Muralist al. evaluated the Naïve Bayes and RF classifiers using features derived from Twitter such as activity, user, and tweet contents.

In this project, we'll employ machine learning approaches to recognize sexually explicit remarks posted by individuals on social media. We will examine if it is feasible to develop an automated categorization model that effectively blocks this kind of information. If so, filters that are appropriate for children or any other user who wants to block this sort of material might be made. [10]

**2.5 User Conversational messages:**

We have gathered real-world conversation data from Chabot and gaming apps for our present study. Search inquiries and conversations are quite different from one another. The following characteristics help to distinguish between questions and discussion 1. Generally speaking, we found that talks are longer than inquiries. Around 3.5 words are used on average in inquiries, whereas 8 words are used on average in talks. Furthermore, interactions might range from 14 simple "hello" to lengthy replies. After removing any discussions that were longer than 0.1 percent, the average chat lasted 250 words, according to our research. We discovered a maximum length of 18 words for inquiries by using the same methodology. 2. Conversations are natural language phrases that are often intended to ask another person or entity (a human or a machine) for something, with the inquiries normally being confined to the keywords. 3. Conversations often include misspelled words. 70% of talks in our survey had several misspelled words. We discovered several misspellings. Sometimes the user's purpose cannot be established from their writing alone and must be ascertained in conjunction with the context and prior conversational interactions. 5. Smileys and symbols may also be used in conversations. For example,

– You are great \_/\\_

– Thank you :): D

– Awesome job (y)

Methods based on deep learning are used to address these issues. We mix two strengths for query completion recommendations. To aid in automated query recognition, Convolutional Neural Networks (CNN) and Bi-Directional LSTM (BLSTM) deep learning models suggest a special structure called "Convolutional, Bi-Directional LSTM (C-BiLSTM)". CBiLSTM uses a 15 convolutional layer to create a feature representation of the word s in the query when it receives one. The BLSTM then receives the feature representation and processes it to provide a more precise representation of the patterns by encapsulating the sequential patterns. A highly connected network processes the learned representation of the query in order to forecast the target class. The variety of discussions we discovered that LSTM and BLSTM with character gram embedding performed better. The three suggested models are trained in a single, unified model and successfully include both local and global meanings of their features. They are not reliant on any hand-crafted feature engineering. They perform better than pattern-based models and other manually created feature-based base models when tested against actual conversations and search queries. [11]

**2.6 Peace-building:**

While the bulk of disputes about language and meanings are strangely related to those about terrorism, certain clarifications are also necessary on the other side. In this research, the word "peace-building" is not restricted to the post-conflict period in particular, especially when it is used to refer to the period after hostilities have ended. The degree of violence and its geographical and temporal distribution vary greatly in numerous conflict zones. In a region that is often referred to as a "conflict zone," fighting is likely to be taking place in some locations while remaining relatively stable in others for extended periods. It seems sensible in this situation to focus on a period when host ileitis has subsided to the point that even modest measures to promote peace are possible. The actions taken at the national level as part of national stabilization efforts in post-conflict regions, such as rehabilitation or reconstruction, reintegration, and so forth, are also considered peace-building actions or components of peace-building for the purposes of conducting this research. In the conclusion, this research strongly urges the adoption of a more particular and limited definition of terrorist activities while also pressing for the expansion of the term peace-building.

**2.7 Religious militants and terrorism:**

There is a danger of entering the realm of subjectivity when the term "terrorism" is used to characterize certain organizations or the operations they carry out. September 11, 2001, terrorist attacks and subsequent acts of terrorism associated with al-Qaeda have made it more difficult to describe terrorism and have led to the usage of new terminology and phrases like "international terrorism," "global terrorism," "catastrophic terrorism," and others. The militant religious groups in Pakistan are involved in terrorism, but they are also all fundamental parts of the extreme Islamic ideology that supports and links terrorism. Experts and academics have made an effort to define terrorism in a way that is acceptable to everybody, but they have not been successful. There are several recurring themes if one examines the problem from the perspective of history and intellectual discussion since the French Revolution and its regimen de la terrier, when the idea of terror first gained popularity. Contrary to its French origins, the bulk of the time, the idea of terrorism was associated with anti-government revolutionary movements. 17 Only in the years leading up to and during World War II was the phrase often used to describe widespread persecution of people by governments. 48 Although there was a time when radical organizations using terror tactics would advocate and defend their actions as "propaganda by deeds," after World War II no revolutionary group wanted to be labeled "terrorist," even though it was targeting civilians and non-combatants. As a result, there is increasing confusion about the term's definition and it has come to mean different things to different people. Different methods of comprehending the word "terrorist" have resulted from the uncertainty in what constituted a terrorist. According to some terrorists, their brutality against people supports their claims that they are freedom-loving warriors defending their rights. Due to the existing uncertainty, several regimes have defined independence fighters and guerrillas as "terrorist groups" and exploited their threats to justify the use of lethal force against them.

It is important to develop a plan to defend Pakistan in all of these circumstances and warn it before any assault happens to stop these threats and attacks on the nation.

**2.8 Social Media:**

People from all around the globe may express their opinions and discuss their concerns on social media. One has to be informed of what social media genuinely means before comprehending its practical applications. It is a term used to describe interactions between individuals or groups who utilize the 18 internet and virtual communities to produce, distribute, and even trade ideas, photos, videos, images, and more. Smartphones and interactive social networking sites like Twitter, Myspace, and Facebook Orkut are influencing how children are reared, making social media an essential part of their daily existence. The way that young people connect with their parents, peers, and technology is changing as a result of social media. Social networking has two advantages. Social networking sites may undoubtedly be a helpful resource for professionals. They do this by helping recent graduates, market their skills and locate business possibilities. Effective network building may also be achieved through social networks. On the other side, there are many threats on the internet that are connected to online groups. One of those risks is cyberbullying sometimes defined as the act of tormenting someone via internet means. In this article, we cover all facets of social media and both its beneficial and detrimental consequences. The emphasis is on certain facets of youth, business, and society as well as health education. In this essay, we'll talk about how various media have a broad impact on society. [13]

**2.9 POPULAR SOCIAL MEDIA SITES:**

**2.9.1 Facebook**

It is the most popular social networking platform on the Internet in terms of both user numbers and brand awareness. Since its founding on February 4, 2004, Facebook has 11 amassed more than 1.59 billion monthly active users, making it one of the most effective methods to connect customers worldwide 19 with your business. More than a million small and medium-sized enterprises are anticipated to use the platform to sell their brands.

**2.9.2 Twitter:**

Likely, keeping our material to 140 characters won't be the best way to market our business. We will be surprised to see that there are more users on the social media platform than the 320 million monthly active users who can utilize the 140-character restriction to transmit information. Businesses may use Twitter to interact with prospective consumers, answer inquiries, and offer the latest information. They can also utilize targeted marketing to reach certain groups. Twitter's headquarters are in San Francisco, California, and the service began on March 21 of that year. [14]

**2.9.3 Google+:**

One of the most well-known social networking platforms nowadays is Google+. It is a crucial tool for small enterprises due to its SEO-related worth alone. Google+ was launched on December 15, 2011, and as of December 15, 2015, it was a part of key alliances with 418 million active members.

**2.9.4 YouTube:**

Three former PayPal workers launched YouTube, the biggest and most well-known video-based internet network, on February 14, 2005. On November 14, 2006, the business was purchased from Google for $1.65 billion. More than 1 billion people use YouTube's website each month, making it the second most popular internet search engine after Google.

**2.9.5 Pinterest:**

In terms of internet networking, Pinterest is often a newbie. It is a stage with computer-generated announcement sheets on which businesses may adhere their material. In September 2015, Pinterest said that it had exceeded 100 million users. Since Pinterest's users are overwhelmingly female, private businesses with a large female target market should invest in it.

**2.9.6 Instagram:**

Instagram is an image-based social media platform. Facebook manages the platform, which has more than 400 million active users. The majority of users use it to publish content on topics like work, travel, and food. In addition to photo-editing and video features, Instagram is also known for its amazing channels. Facebook and Instagram are both used by around 90% of users.

**2.9.7 WhatsApp:**

A cross-platform instant messaging program for computers, smartphones, and tablets is called WhatsApp Messenger. To send users who have the program running on their smartphone photographs, texts, documents, videos, and audio messages, Internet access is required. It was made available in January 2010. Facebook acquired WhatsApp Inc. on February 19, 2004, for about $19.3 billion. More than one billion individuals use the service today to communicate with their friends, partners, family, and even clients. [15]

**2.9.8 Snapchat:**

When they were Stanford University students, Reggie Brown, Evan Spiegel, and Bobby Murphy developed Snapchat, an image-based application training tool. The software was launched in September 2011 and within a relatively short period of concentration; it amassed a sizable user base, reaching 100 million active users in May 2015. Snapchat is used by the vast majority of people using social media.

**CHAPTER 3**

**METHODOLOGY AND INTERACTION OF**

**COMPONENTS**

We will talk about the project's chosen languages and techniques in this part.

**3.1 Methodology:**

The project was built using a methodology based on the RANDOMFORESTCLASIFIER

Algorithm. The popular machine learning algorithm Random Forest belongs to the supervised

Learning methodology. It is a good option for ML issues involving both regression and

Classification. This is based on the concept of ensemble learning, which is the act of mixing

many classifiers to solve a challenging issue and increase the effectiveness of the

Algorithm.

Random Forest, as its name indicates, is a classifier that uses several decision trees on

different subsets of the provided dataset and averages them to increase the dataset's

prediction accuracy. The random forest uses the best forecasts from each decision tree

rather than just one, and it predicts the result depending on which guesses received the

most support. [16]

Increasing tree density throughout the forest may improve accuracy and solve the

overfitting problem.

**3.2 Tools, Mechanisms, and Software:**

A brief introduction of all the software and tools used in the development of the prototype

are as follows:

**3.2.1 HTML:**

The acronym HTML stands for Hyper Text Markup Language. The markup language used

to create Web pages is called HTML. The structure of a Web page is specified by HTML.

Various elements make up HTML. HTML components provide the browser

13instructions on what to show on the screen. It is supported by CSS (Cascading Style Sheets)

technology and JavaScript programming languages.

Web browsers receive HTML files from a server or local storage and transform them into

multi-media web pages. HTML was originally accompanied by hints to the design in the

documents, and it conceptually defines the structure of a website page.

HTML components These are the fundamental components of HTML pages. HTML is the

primary building block of HTML pages. With HTML coding, images and other

components like interactive forms It might be included into the finished website. By

referencing the structural semantics for text, such as headers for lists and paragraphs,

HTML enables you to create structured documents. Links Other things, sayings, and other

stuff. Tags, which are expressed using angle brackets, create HTML elements. Content is

introduced into the page directly through tags like and. other tags, including.

The surroundings provide details about the document's content. It may also include sub

elements with additional tags. HTML tags are used by browsers to interpret the material

on a web page, not to display them. [17]



Figure 3-1: HTML language working

**3.2.2 CSS:**

Cascading Style Sheets (CSS) are a form of stylesheet that govern how documents written

in HTML or XML should be laid up (including XML 25 dialects such as SVG, MathML

or XHTML). CSS specifies how elements must appear on displays, paper, in writing,

during speaking, and in other mediums. [18]

The separation of content and presentation, which includes the layout, colours, and fonts,

may be made easier using CSS. This division could enhance the accessibility of the content

and give designers more freedom and control over presentation features that allow multiple

websites to use the same formatting and CSS in separate.css documents, which can reduce

repetition and complexity in the text's structure. It could also enable the.css do cument to

be stored to speed up the loading of pages that share the file as well as its for.[19]

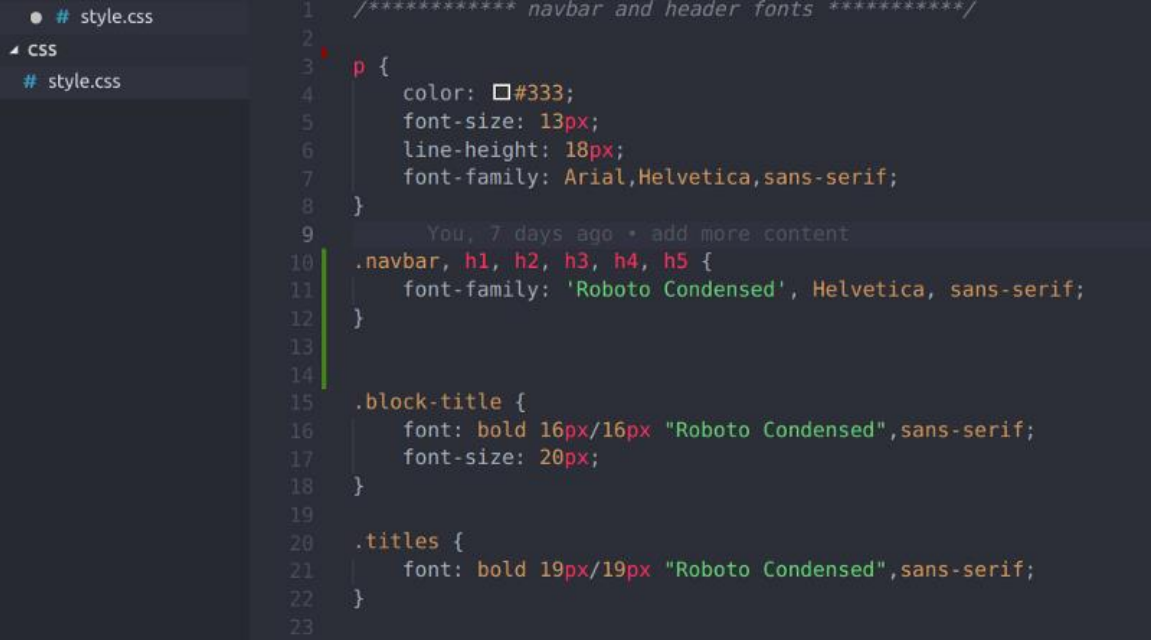


Figure 3-2: CSS language working

The ability to separate content from formatting enables you to render the same page markup

in a variety of ways, including written, on -screen, spoken (using screen readers and

speech-based browsers), braille-based touchscreen devices, and printed. When the material

is accessible through a portable device, CSS additionally offers instructions for alternative

layout. When there are many style rules that are compatible with one element, the

priority mechanism used to choose which style rule will apply is where the name

"cascading" comes from. The cascading priority system is straightforward to comprehend.

[20]

**3.2.3 PyCharm:**

One of the most well-known Python IDEs is PyCharm. This is due to a number of factors,

including the fact that it was made by JetBrains, the company behind the famed IntelliJ

15IDEA IDE, one of the three main Java IDEs, and WebStorm, the "smartest JavaScript IDE."

Another strong argument is the fact that it supports Django-based web development.

Python is an integrated development environment (IDE) used to program computers,

specifically for the Python programming language. It was made by Czech company

JetBrains, which created it (formerly called IntelliJ). It may be used to create web apps

with Django and data science applications with Anaconda since it has code analysis, a

visual debugger, a unit tester, and interaction with version control systems (VCSes).

PyCharm is cross-platform; it is available in versions for Windows, macOS, and Linux.

Additionally to an educational version and a professional edition with added functionality,

it also has a Community Edition Community Edition that is distributed under the Apache

License (released under the terms of a 27 subscription-based exclusive licensing). [21]

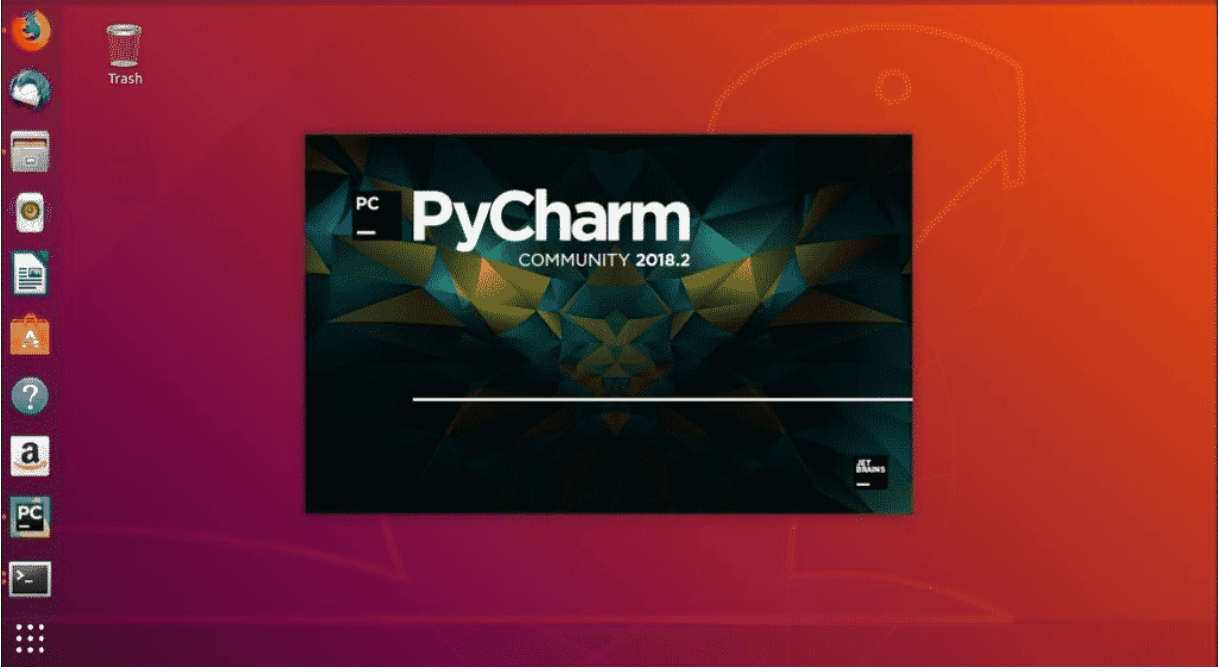


Figure 3-3: PyCharm Software

**3.2.4 Jupyter Notebook:**

Project Jupyter is a project and community with the goal of "developing open - source

software, open-standards, and services for interactive computing across dozens of

programming languages." Like other planets, it is a word that some people pronounce as

"py." Fernando Perez created Project Jupyter in 2014 after splitting it off from IPython.

Brian Granger, too. The name Project Jupyter pays respect to Galileo's notebooks, which

include the moons of Jupiter that Galileo found, as well as to the three programming

16languages that Jupyter uses: Julia, Python, and R. The interactive computing tools Jupyter

Notebook, JupyterHub, and JupyterLab have been created by Project Jupyter. NumFOCUS

[22]

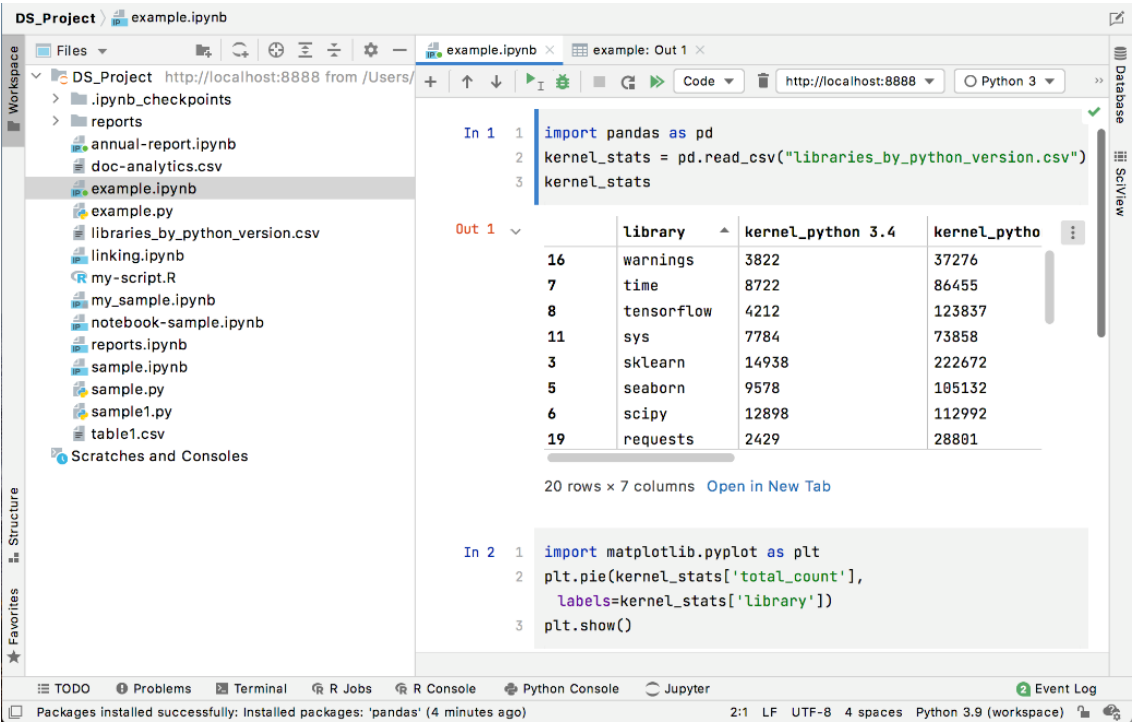


Figure 3-4: Jupyter Notebook Working

**3.3 Machine Learning:**

Machine Learning (ML) is a field of study that focuses on comprehending and creating

ways for learning, that is, procedures that make use of data to improve performance in

specific tasks. It is seen as having some aspects of artificial intelligence. Machine learning

algorithms build an algorithm using training data, sometimes referred to as data samples,

and are capable of making predictions or taking choices without needing to be explicitly

programmed to do so. These algorithms may be used in a range of industries, including

computer vision, email filtering, voice recognition, and medicine, when it's difficult or

impossible to develop conventional algorithms to do the necessary tasks. [23]

Although computation statistics, which is focused on using computers to make predictions,

is not the only sort of statistical, it is strongly related to a particular subset of machine

learning. The field of machine learning benefits from the methodologies, ideas, and

application fields that may be learned fr om studying mathematical optimization. Data

mining is a related field of study that specialises in the exploratory exploration of data

17using unsupervised techniques. Some machine learning applications employ data and

neural networks in a way that closely resembles how the brain operates. Machine learning

is frequently referred to as prescriptive analytics since it is used to solve business problems.

In 1962, Robert Nealey, a self-described checkers expert, used an IBM 7094 computer to

play the game. He was defeated by the computers. This achievement seems little in light

of what is now possible, yet it is nonetheless recognised as a significant advance in the area

of artificial intelligence. The development of processing and storage technology in the

future years will provide some ground-breaking innovations that we now use and love, like

Netflix's recommendation engine or driverless cars. [24]

The data science research community is expanding quickly, and machine learning is a key

component of that discipline. The most crucial insights in the data mining project are

revealed by teaching computers to construct classifications or predictions using statistical

approaches. These insights then influence business and application choices that may have

an effect on ke y growth metrics. The demand for data scientists will grow as the amount

of big data keeps growing, and they will be needed to help identify the most important

business problems and the data that will give the solutions.

**3.3.1 Machine Learning vs. Deep Learning vs. Neural Networks:**

The differences between deep learning and machine learning should be understood since

they may and are often used interchangeably. Neural networks, deep learning, and machine

learning are all branches of artificial intelligence. However, neural networks a branch of

deep learning, which is itself a subject of machine learning.

The way that machines and deep learning learn is the main distinction between the two.

Deep learning can automate several feature extraction process step s, minimizing the need

for some human involvement and enabling the use of larger datasets. As noted by Lex

Freidman in this MIT lecture, one may see deep learning as "scalable machine learning"

(01:08:05). Classical machine learning, sometimes referred to as "non-deep" machine

learning, relies more on human input for learning. The sorts of characteristics needed to

18distinguish different data inputs are chosen by human specialists, and these features often

need more extensive data to grasp. [25]

The supervised learning approach, also known as labelled data sets, may be used by "deep"

machine learning to direct its algorithm, although it is not a requirement. It can rapidly

identify the characteristics that set apart various sorts of data from one another after

ingesting unstructured data in its most basic form (such as text or photos). Machine

learning, however, does not need the involvement of people in the data processing process,

allowing us to develop machine learning in more creative ways. Deep learning an d neural

networks are often attributed with accelerating technical progress in fields like voice

recognition, computer vision, and natural language processing.

Artificial neural networks (ANNs), commonly referred to as neural networks, are made up

of layers of nodes that include an input layer, one or more hidden layers, and an output

layer. Each artificial neural network node is connected to the next and given a threshold

and a weighting. A node will be activated and transfer data to the subsequent layer if its

output exceeds the threshold that has been specified. If not, no information is sent to the

layer below. The layers of depth in a neural system are simply described by the word "deep"

in deep learning. A deep learning method or even a deep neural system may be thought of

as a neural network with more than three layers, inclusive of both inputs and outputs. One

or two-layer neural networks are considered to be basic neural networks. [26]

**3.3.2 How Machine Learning Works:**

A machine-learning algorithm learns in three main stages, according to UC Berkeley.

1. The Choice Algorithms for machine learning in general may be used to estimate or

categories data. The method may predict the presence of a pattern in the data based

on the input data, which can be labelled or unlabeled. [27]

2. The Function Error In order to evaluate the models' correctness, an error function

is utilized. If there exist instances, it may be possible to compare them using the

error function to judge how accurate the model is. [28]

3. A method for model optimization The weights are changed to close the discrepancy

between the real situation and the predicted model if the model performs better with

the data points of the training set. When a minimal level of accuracy is attained, the

algorithm will repeat the assessment and optimization procedure while

independently altering the weights. [29]

**3.3.3 Machine Learning Methods:**

* **Supervised machine learning:**

A supervised machine learning system supervised education the use of labelled data to

construct algorithms that can accurately categories data or predict outcomes is often

referred to as supervised learning. The weights of the input data are changed when more

data is added to the model, until it is adequately fitted. This takes place during g the cross

validation process to ensure that the model does not exhibit underfitting or improper fitting.

Large-scale real-world difficulties may be resolved by companies with the assistance of

supervised learning. One such example is the separation of spam into a distinct folder from

your inbox. Neural networks, naive Bayes logistic regression, linear regression, random

forests, support vector machines (SVM), and many more techniques are used for

supervised learning. [30]

* **Unsupervised machine learning:**

Unsupervised education Unsupervised machine learning, often known as ML, uses

machine learning algorithms to organize and evaluate unlabeled data. These algorithms

automatically identify hidden patterns or data clusters. It is the perfect tool for cross-selling

techniques and more exploratory data analysis since it can show how similar and different

data are. Additionally, it supports client segmentation, picture recognition, and pattern

recognition. The reduction of dimensionality may also be used to reduce the number of

features that are included in a model. Singular value decomposition (SVD) and principal

component analysis (PCA) are two of the most widely used techniques for doing this.

Neural networks, k-means-based clustering, probabilistic clustering methods, and many

other algorithms are used to learn unsupervised. [31]

* **Semi-supervised learning:**

Unsupervised and supervised learning are well-balanced by semi-supervised learning. It

uses a smaller, unlabeled sample of data to help with classification during training and also

extracts features from a larger, unlabeled collection of data. The problem of not having

enough labelled data or the resources to label enough data might be resolved by semi

supervised learning instead of developing a supervised algorithm.

**3.3.4 Real-world machine learning use cases:**

Here are just a few examples of machine learning you might encounter every day:

**Speech recognition:** The system is known as automated speech recognition (ASR), often referred to as voice-to-text and computer speech recognition. It is a skill that turns spoken

words into written text using natural language processing (NLP). Many mobile devices

include speech recognition capabilities into their operating systems, enabling voice search,

like Siri, or enhancing messaging accessibility. [32]

**Customer service:** Online chatbots are replacing human customer service representatives

at every stage of the customer experience. They provide personal suggestions, cross-sell

items, or provide size recommendations for users, which is altering how we see consumer

involvement on social media and online platforms. They also react to frequently asked

questions (FAQs) on subjects like shipping. Examples include the jobs often carried out by

voice assistants and virtual agents, chatbots on e-commerce websites with virtual assistants,

and messaging services like Slack and Facebook Messenger.

**Computer vision:** Is an artificial intelligence (AI) technology that enables computers and

other systems to gather pertinent data from digital video, digital pictures, and other visual

inputs and make decisions based on those inputs. It differs from conventional picture

recognition jobs in that it may provide recommendations. Convolutional neural networks

based Social-media picture tagging, radiological imaging for healthcare, and autonomous

cars in the automotive sector all employ computer vision. [33]

**Recommendation engines:** AI algorithms can assist in identifying trends in data that can be used to create more effective cross-selling strategies by using historical consumption

behavior data. When a customer is checking out with an online retailer, this can be used to

offer pertinent recommendations for them to add on.

**Automated stock trading:** designed to increase stock portfolios Without human

interaction, high-frequency trading systems driven by AI execute hundreds, or maybe

millions, of deals per day.

**3.3.5 Challenges of Machine Learning:**

Our lives have become more easier as a result of improvements in machine learning

technology. However, the use of machine learning in businesses has also given rise to

certain ethical questions about AI. These consist of Singularity of technology.

* **Technological singularity:**

Although the general public pays a lot of attention to this issue, most academics aren't

concerned about the potential of AI soon surpassing human capabilities. It is sometimes

referred to as superintelligence, which is defined by Nick Bostrum as "any intellect that

significantly excels the finest human brains in nearly every discipline, including scientific

innovation, general knowledge, and social abilities." Although Superintelligence and

powerful AI are not likely to affect our civilization in the near future, the idea of them

creates intriguing issues when we consider the prospect of autonomous systems, such as

driverless automobiles. It is unrealistic to think that a driverless car won't get in an accident,

but who would be responsible and liable in these circumstances? The verdict is still out on

this one, but these are the types of moral disputes that are being waged as innovative, new

AI technology develops. [34]

* **AI impact on jobs:**

Despite the fact that the bulk of people's views on artificial intelligence center on job loss

This issue has to be one that is likely to change. We are seeing a change in the market needs

22for certain employment positions with each disruptive, creative technology. For instance,

when we look at the automobile industry, we can see that many businesses, like GM, are

putting their attention on electric cars in order to support the green effort. The energy sector

won't disappear, but the main fuel that powers it is shifting from a petroleum economy to

an electronic one. Artificial intelligence should be seen similarly, since it has the potential

to move the demand for work to other sectors. As the quantity of data grows and changes

every day, there will be people who can assist in managing these systems. Within the

sectors most vulnerable to changing demand, such as customer service, resources must still

be available to handle increasingly challenging problems. Helping workers transition into

these new fields in line with market expectations is the most crucial aspect of AI and its

implications on the labor market. [35]

* **Privacy:**

Data privacy, data security, and data security are topics that are usually brought up while

talking about privacy. These problems have made it possible for policymakers to progress

recently. For example, the GDPR regulation was passed in 2016 to protect the personal

data of people who reside in the European Union and European Economic Area and to

offer people gr eater control over their personal data. For the United States, various states

are developing regulations, such as the California Consumer Privacy Act (CCPA), which

mandates that businesses notify their clients when their personal information is processed.

Companies have been obliged to reevaluate how they retain and use personally identifiable

information as a result of the new regulations (PII). As a result, businesses are making

security-related expenditures a high priority in their attempts to close any gaps or openings

for hacking, spying, and cyberattacks.

* **Bias and discrimination:**

A number of ethical questions surrounding artificial intelligence and AI have been

highlighted by prejudice and discrimination in a variety of intelligent technologies. When

the training data itself may be skewed, how can we defend against discrimination and bias?

Although the majority of businesses have good intentions when it comes to their

23automation endeavours, Reuters points out some of the unforeseen consequences of

incorporating AI into recruiting procedures. As they attempted to simplify and automate a

process In technical positions that are available to everyone, Amazon purposefully slanted

prospective applications based on gender, but eventually had to abandon the scheme.

Discrimination and bias may be seen in a wide range of applications, from face recognition

software to social media algorithms, and are not only confined to the human resources

department.

Businesses have been increasingly interested in the discussion of AI ethics and values as

they become more aware of the risks posed by AI. For instance, Arvind K. Krishna, CEO

of IBM, stated in an announcement from the previous year that the company had

discontinued its general-purpose facial analysis and recognition software "IBM

vehemently opposes and will not support the use of any other technology, including third

party facial recognition technology, for mass surveillance or racial profiling, for violations

of fundamental liberties and rights of individuals, or for any other use that is inconsistent

with our core principles or the Principles of Trust and Transparency. [36]

* **Accountability:**

There is no substantial regulation that regulates AI activities, and there is no enforcement

system to guarantee that ethical AI is being employed. The detrimental effects of an

unreliable AI system on your bottom line are the main drivers for businesses to adhere to

these standards. To close this gap, researchers and ethicists have worked together to

establish ethical frameworks that control the development and use of AI models in society.

Research has demonstrated that the combination of shared responsibility and a lack of

understanding of the possible repercussions isn't helpful to preserving society at this

moment since they merely act as guidelines. [37]

**3.4 Artificial Intelligence:**

In contrast to the natural intelligence shown by all creatures, including humans, artificial

intelligence (AI) is the capacity to identify intelligence expressed by computers. The study

of intelligent agents is known as the field of artificial intelligence. It refers to any system

24that has the ability to detect its environment and make choices that will increase the

likelihood that it will accomplish its goals.

The phrase "artificial intelligence" used to be used to describe robots that mimic and

demonstrate "human" cognitive abilities that are a feature of the human brain, such as

"learning" and "problem-solving." Major AI researchers eventually rejected the notion,

describing AI as a function that results from the rational process operating rationally and

in accordance with what is not restricting how intelligence may be characterized.

Applications of AI include cutting-edge online search engines like Google,

recommendation systems like YouTube, Amazon, and Netflix, voice recognition software

like Siri and Alexa, self-driving automobiles like Tesla, automated decision-making, and

dominating games that need strategic thinking (such as Chess and Go).

Tasks that are considered to need "intelligence" are often omitted from the definition of

AI, which is referred to by "the impact of AI. For instance, although being a commonplace

technique today, optical character recognition is often not considered to be AI.

In 1956, the study of artificial intelligence was first established as a discipline. Since then,

there have been several waves of excitement that were followed by disappointment and

budget cuts (known as "AI winter"), new approaches, successes, and increased investment.

Since its inception, AI research has explored and disproved a wide range of approaches,

including researching human brains, recreating human brains, problem-solving, formal

logic, massive databases of knowledge, and mimicking the behavior of animals. Machine

learning that is heavily based on mathematics and statistics has dominated the subject in

the first two decades of the twenty-first century. This approach has been very effective in

solving many difficult issues in both business and academia. [38]

**3.4.1 Goals of AI:**

These days, the replication (or creation) of intelligence difficulty is divided into more

manageable problems. Researchers hope that intelligent robots will exhibit these particular

qualities and skills. The following characteristics have garnered the greatest attention.

**3.4.2 Reasoning and problem solving:**

In the beginning, scientists created algorithms that replicated the sequential reasoning that

people employ to solve problems or re ach conclusions. Using ideas from economics and

probability, AI research in the late 1990s and the 1980s had created methods for coping

with ambiguous or inadequate information.

Many of these algorithms were discovered to suffer from a "combinatorial explosion,"

meaning that when problems increased in complexity, they became exponentially slower.

This made them unsuitable for handling large reasoning problems. The sequential

deductions that early AI research might imitate are not even used by humans. The bulk of

their issues are solved through rapid intuitions.

**3.4.3 Planning:**

A clever agent with the ability to plan builds a picture of the world as it is right now,

predicts the actions they will take to change it, and makes decisions that maximize the

value (or "value") of the possibilities at hand. It is conceivable for an agent to assume it is

the only system working in the universe when confronted with conventional planning

problems, giving it confidence in the outcomes of its actions. If it isn't the only actor,

though, the agent must be able to think under ambiguity, constantly assess its environment,

and make adjustments. A multi-agent strategy employs both inter-agent cooperation and

inter-agent rivalry to accomplish a goal. Swarm intelligence and algorithmic evolution may

both benefit from the behavior of an evolving type like this. [39]

**3.4.4 Learning:**

The most basic concept in AI research since its inception is machine learning (ML), which

examines computer-based algorithms that are enhanced automatically via use.

Unsupervised learning recognises patterns in the input stream. Unsupervised learning

comes in two main varieties: classification and numerical regression, both of which need a

person to manually classify the input data first. You may use classification to discover what

category something belongs to. As it learns to classify new inputs, the programme will

study a number of instances of things in various categories. Regression is the process of

attempting to develop a function that explains the relationship between outputs and inputs

26and foresees how changes in inputs are likely to effect outputs. Both regression and

classifiers are seen as "function approximators" that are attempting to find an unknown

(potentially implicit) function. For example, the spam classifier may be seen as learning a

function that categorizes email content as "spam" or "not spam." In reinforcement learning,

agents get rewards for good replies and punishment for poor ones. To create a strategy for

navigating its problem area, the agent classifies its replies. Transfer learning is the process

through which the information gained from one specific issue is transferred to another. [40]

The idea of computational learning may be used to assess learners in terms of the

complexity of the sample (how many data are required), the computational difficulty, and

other optimization-related notions.

**3.5 Dataset building:**

It could seem like gathering data for your AI project is a straightforward operation that can

be completed in the background while you focus the majority of your time and energy on

building the machine-learning model. However, managing data might take a long time

because of the magnitude that the work could grow to, as experience has repeatedly shown.

Therefore, understanding what machine learning is in its most basic form is essential. It's

crucial to understand the ideal method for data collection as well as the qualities that make

up a quality data set. [41]

Simply described, a machine learning dataset is a collection of data points that computers

may use for analysis and predicting. Since a computer cannot interpret data in the same

manner that people do, the data must be homogeneous and simple to understand. Once the

data has been gathered, it is essential to clean it up, finalize the data, and add relevant tags

that computers can understand in order to do this.

A well-constructed dataset must also adhere to specific requirements for quantity and

quality. You must make sure that your data is well-balanced and relevant if you want 38

training to go smoothly and quickly. Utilize real-time data wherever possible, and consult

knowledgeable professionals on its quantity and source.



Figure 3-5: Dataset building using Twint library

**3.6 Data Preprocessing:**

Data preprocessing is a process of preparing the raw data and making it suitable for a

machine learning model. It is the first and crucial step while creating a machine learning

model. [42]

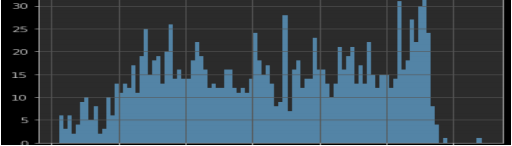


Figure 3-6: Data preprocessing graph

When creating a machine learning project, it is not always a case that we come across the

clean and formatted data. And while doing any operation with data, it is mandatory to clean

it and put in a formatted way. So for this, we use data preprocessing task.

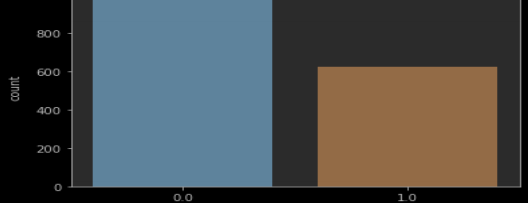


Figure 3-7: Data preprocessing conversion graph 1



Figure 3-8: Data preprocessing conversion graph 2

**3.7 Model Training:**

A training model is a dataset that is used to train an ML algorithm. It consists of the sample

output data and the corresponding sets of input data that have an influence on the output.

The training model is used to run the input data through the algorithm to correlate the

processed output against the sample output. The result from this correlation is used to

modify the model. [43]

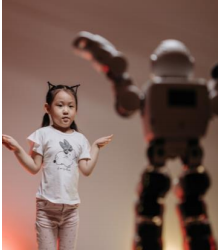


Figure 3-9: Model Training Example

Model fitting is the term used to describe this procedure. The correctness of the suggested

model depends critically on the precision of the training data or validation dataset.

The act of supplying the ML algorithm with data to help in the discovery and acquisition

of the optimal values for each characteristic associated with it is referred to as "model

training" for machine languages. There are several machine learning models available,

however the supervised and unsupervised learning models are the most popular ones.

When the data used for training contains both input and output values, supervised learning

may be accomplished. An supervisory signal is any piece of data that contains both the

anticipated outcome and the input. When inputs are supplied into the model, the training

process is carried out in line with the variance of the result that is processed from the

recorded result. [44]

Unsupervised learning involves determining patterns in the data. Additional data is then

used to fit patterns or clusters. This is also an iterative process that improves the accuracy

based on the correlation to the expected patterns or clusters. There is no reference output

dataset in this method.

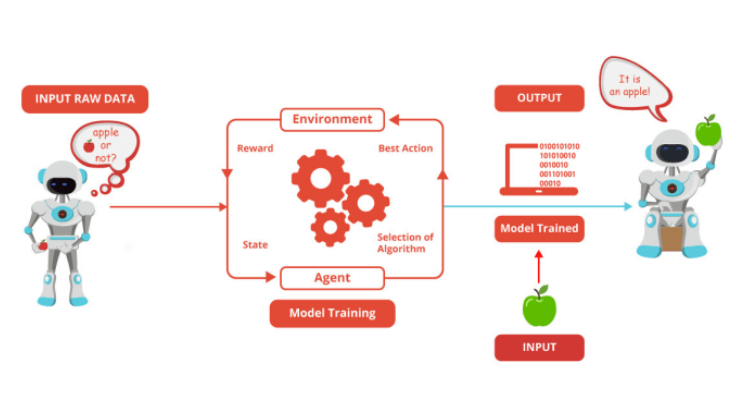


Figure 3-10: Model Training Example

**3.8 Model Evaluation:**

Model evaluation is the process of using different evaluation metrics to understand a

machine learning model's performance, as well as its strengths and weaknesses. Model

evaluation is important to assess the efficacy of a model during initial research phases, and

it also plays a role in model monitoring. Following are the ways in which the model was

evaluated:

i. Using SVM

ii. Logistic Registration

iii. Random Forest Classifier

iv. Kneighbors Classifier

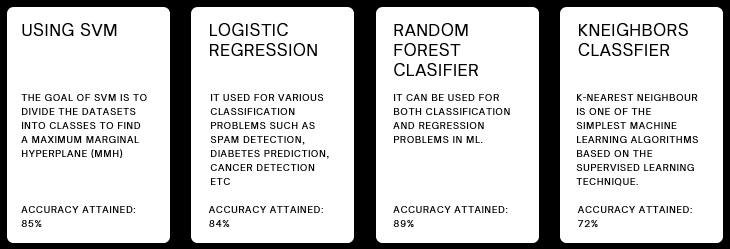


Figure 3-11: Model Evaluation Methods

The degree of precision obtained using each of the four techniques varied. For instance,

the accuracy of the SVM approach is 85%. With Logistic Regression, accuracy was 84%;

with Random Forest Classifier, accuracy was 89 percent; and with Neighbors Classifier,

accuracy was 72%. Random Forest Classifier is chosen because it is believed to be the best

algorithm to utilize as a model for this kind of prototype based on these results of accuracy.

**3.8.1 Random Forest Classifier:**

Random Forest is a popular machine learning algorithm that belongs to the supervised

learning technique. It can be used for both Classification and Regression problems in

ML. It is based on the concept of ensemble learning, which is a process of *combining*

*multiple classifiers to solve a complex problem and to improve the performance of the*

*model.*

As the name suggests, "Random Forest is a classifier that contains a number of decision

trees on various subsets of the given dataset and takes the average to improve the

predictive accuracy of that dataset." Instead of relying on one decision tree, the random

forest takes the prediction from each tree and based on the majority votes of predictions,

and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the

problem of overfitting.

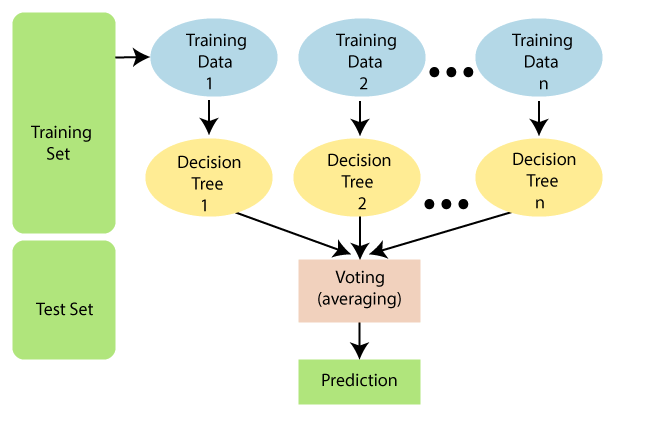


Figure 3-12: Random Forest Classifier working method

**Assumptions for Random Forest:**

Since the random forest combines several decision trees in order to anticipate the data's

nature, it is likely that certain decision trees will be able to predict the desired outcome

32while others won't. But when all trees are included, they can anticipate the right result.

Here are two presumptions that might be used to create a better Random Forest classifier:

o To guarantee that the classification algorithm can provide accurate results rather

than an ignorant output, there must be real results in the value variable in the

dataset.

o The predictions of the tree need to be relatively uncorrelated.

**Why use Random Forest?**

Below are some points that explain why we should use the Random Forest algorithm:

<="" li="">

o It takes less training time as compared to other algorithms.

o It predicts output with high accuracy, even for the large dataset it runs efficiently.

o It can also maintain accuracy when a large proportion of data is missing.

**How does Random Forest algorithm work?**

Construction of the random forest using a mixture of N choices trees is the first phase of

random forest's two-phase operation. Predicting each tree built in the first step is the second

phase. Working process can be explained in the below steps and diagram:

**Step-1:** Pick K data points at random from the training set.

**Step-2:** Construct the decision trees for the chosen data points (Subsets).

**Step-3:** Decide how many N-level decision trees you want to build. Repeat steps 1 and 2

in step 4.

**Step-4:** Repeat Step 1 & 2.

**Step-5:** Assign the newly found numbers to the categories that earn the most votes for new

data points by identifying the expected results from each tree. [45]

**3.8.2 Python Implementation of Random Forest Algorithm:**

The Random Forest Algorithm tree will then be produced using Python. We'll utilize the

same dataset, "user data.csv," to do this. The Random Forest classifier may be compared

against other classification models such as Decision Tree Classifiers, KNN, SVM, Logistic

Regression, etc. using the same dataset. [46]

Implementation Steps are given below:

✓ Data Pre-processing step

✓ Fitting the Random forest algorithm to the Training set

✓ Predicting the test result

✓ Test accuracy of the result (Creation of Confusion matrix)

✓ Visualizing the test set result.

# Chapter 4

**DESIGN AND IMPLEMENTATION**

**4.1 Designing of Prototype:**

Our prototype was created using the Python Flask framework, which allows data from the application to be delivered to the user interface and user experience. A web-based framework is Flask. Since CSS and HTML are the languages used to create web-based apps, they are used to design the prototype.

**4.1.1 Flask Framework:**

A web framework is Flask. It's a Python module that enables easy development of web apps. It is a small-scale micro framework with an easy-to-expand fundamental framework but no object relationship manager (ORM) or additional features. It contains a variety of intriguing features, such as URL routing and a template engine. It is a framework for WSGI web applications. Python was used to create the Flask framework for web-based applications. It was developed by Armin Ronacher, the leader of Poocco, a global organization of Python aficionados. The Workzeg WSGI toolkit and Jinja2 templates engine serve as the foundation for Flask. These are both Pocco projects. [47]

* **WSGI:**

The Web Server Gateway Interface, sometimes known as WSGI, has been used as a standard in the development of Python web applications. The definition of a standard interface for web servers and web -based applications is WSGI.

* **Werkzeug:**

An WSGI toolkit called Werkzeug may implement request objects, response objects, utility functions, and other things. This makes it possible to build a web frame on top of it. With Werkzeug as its foundation, the Flask framework was created.

* **Jinja2:**

A well-known template engine created for Python is called Jinja2. Systems that use web templates integrate the template with a specific data source to produce dynamic online pages.

* **Micro framework:**

Many people refer to Flask as a micro framework. It was intended to keep the application's fundamentals simple and adaptable. Instead of providing a data support abstraction layer, Flask may be enhanced to include these features directly into the application. [48]

**Why is Flask a good web framework choice?**

Unlike the Django framework, Flask is very Pythonic. It’s easy to get started with Flask, because it doesn’t have a huge learning curve. On top of that it’s very explicit, which increases readability. To create the “Hello World” app, you only need a few lines of code.

**4.2 Libraries for Prototype:**

The phrase "library" in the context of computer technology refers to a collection of nonvolatile materials used by computer programs, generally for the creation of software. Documentation, configuration data, help data, and message templates all fall under this category. Classes, variables, and subroutines that have already been created, as well as type definitions.

The following libraries are used in this project:

**4.2.1 Pandas:**

Data analysis is made possible via the Python module Pandas. As a result of the project's founding by Wes McKinney in 2008 and the need for reliable and adaptable tools for quantitative research, pandas has developed into one of the most in-demand Python libraries. Its user base is quite large. Pandas is built on top of two essential Python libraries: NumPy for mathematical operations and Matplotlib for data presentation. Pandas serves as an addition to these libraries that makes it possible to utilize various NumPy and matplotlib methods with just a little amount of code. The pandas'.plot() function, for instance, combines numerous Matplotlib 48 functions into a single method that enables users to generate a graph in only a few lines. [49] Before Pandas, a lot of analysts used Python for data proces sing and preparation before switching to a particular language, like R, for the rest of their job. Pandas offered two different types of data storage objects, which simplify analysis and eliminate the need to switch between tools: Series having a list-like structure and DataFrames with a tabular structure are also examples.

**4.2.2 Numpy:**

The abbreviation for numerical is NumPy. Multidimensional arrays and algorithms to handle them may be found in the Python library. NumPy enables the execution of mathematical and logical operations on arrays. The foundational elements of NumPy, such as its environment and structure, will be covered in this lesson. Additionally, it describes various indexing techniques, array operations, and more. There is also a quick explanation of Matplotlib. Illustrations are used to describe every step of the procedure to make sure you understand. [50]

**4.2.3 Matplotlib:**

Python's Matplotlib package makes it possible to create 2D graphs and plots using scripts. It includes a module called Pyplot that simplifies plotting by providing options to modify line styles as well as font properties, format axes, etc. A wide variety of plots and graphs, including histograms, bar charts, spectra, error charts, and more, may be supported by it. It is used in combination with NumPy to provide a powerful open source environment that can replace MatLab. Additionally, it works with graphics programes like PyQt and WXPython. [51]

**4.2.4 Seaborne:**

Python statistical graph charting uses the amazing visualization tool Seaborne. In order to improve the aesthetics of statistical charts, Seaborne comes with lovely colors and styles as standard. The matplotlib library serves as the program's foundation, and the pandas data structures work closely with it. A project called Seaborne aims to make data interpretation and analysis primarily based on visualization. Seaborne offers dataset-focused APIs 49 that let us switch between several visual representations for the same variables, facilitating a deeper comprehension of the information.

**Different categories of plot in Seaborne:**

Plots are basically used for visualizing the relationship between variables. Those variables can be either be completely numerical or a category like a group, class or division.

**Seaborne divides plot into the below categories:**

**• Relational plots:** This plot is used to understand the relation between two variables.

**• Categorical plots:** This plot deals with categorical variables and how they can be visualized.

**• Distribution plots:** This plot is used for examining univariate and bivariate distributions

**• Regression plots:** The regression plots in seaborne are primarily intended to add a visual guide that helps to emphasize patterns in a dataset during exploratory data analyses.

**• Matrix plots:** A matrix plot is an array of scatterplots.

**• Multi-plot grids**: It is an useful approach is to draw multiple instances of the same plot on different subsets of the dataset.

**4.2.5 Sickest-learn:**

A free machine learning library for Python is called Scikit-learn. It has support for a number of methods, including vector machines, random forests, and k-neighbors. Additionally, it is compatible with NumPy and SciPy, two Python libraries for math and science.

With the aid of the sickest-learn package, which was created to simplify the process of machine learning in Python simple and more robust, we'll learn how to program in Python in this video lesson. [52]

We'll utilize the Sales Win Loss database set from IBM's Watson repository for this. We'll use pandas to import the data set, examine it using pandas functions like heads(), tail(), and dtypes(), and then test utilizing Seaborne charting tools to present our data.

After that, we'll apply preprocessing and go into Sickest-Learn. Sickest-Label Encoder() Learns and train test split() functions were used to analyze the data and divide it into train and test samples, respectively. The cheat sheet will also be used to decide the algorithm to utilize to process our data sets. Additionally, we'll forecast the future using three distinct algorithm types (Naive-Bayes, Linear SVC, and K-Neighbors Classifier) and assess each one's effectiveness using tools like the sickest-learn library's accuracy score() function. Using sickest-learn and Yellow brick visualization, we will also display the performance score of the various models.

**4.2.6 NLTK:**

The Natural Language Toolkit (NLTK) is a programming environment for Python applications that deal with human language data and are used to analyze language statistically. It contains text processing packages that provide semantic reasoning, tokenization, parsing, and classification stemming. Along with cookbooks and a booklet that explains the principles behind the language processing that NLTK is able to handle, it also includes graphical demos and data sets with samples. [53]

**4.2.7 Jinja:**

Jinja is a Python software engine used to create HTML, XML, and other markup forms. It is sometimes referred to as "Jinja2" to denote the most current version. The HTTP response contains these formats for the end user. Because Jinja2 employs a standard template tag syntax and is properly extracted to be an open source application, it may be used as a dependency for many code libraries.

**4.3 Twint Web Scrapping:**

We can extract information from numerous websites on the internet and send it to your local computer via web scraping. Using Hypertext Transfer Protocols, it extracts data from numerous web-based portals and uses it for our purposes. Numerous businesses use this method to gather data and build web-based search bots.

Python has a large selection of modules and packages that help with website scraping, including selenium and beautiful soup. Web scraping can be automated using a number of libraries, including Auto scraper. All of these tools make use of different APIs, which allow us to scrape data and then save it in a data frame on our computer.

Twint is an open-source Python tool that may be used to scrape tweets, or to retrieve data from Twitter without using the Twitter API. Twint differs from competing Twitter scraping APIs in a number of ways that make it more user-friendly and unique, including: [54]

• Twitter API has a limit of fetching only 3200(last) tweets while twint has no limit of downloading tweets, it can download almost all the tweets.

• Easy to use and very fast.

• No initial Sign-in or Sign-up required for fetching data. Twint can be used to scrape tweets using different parameters like hash tags, usernames, topics, etc. It can even extract information like phone number and email id’s from the tweets.

**4.4 Model Deployment:**

Pushing updates or changes from one deployment environment to another are how web and software development is deployed. Your live site, also known as the live site or the production, is constantly there when you establish a website.

The simplest way to implement a machine learning algorithm is to create a web-based service that can make predictions. In this case, we'll deploy the machine learning models using the Flask web platform. This is then changed into a web page that is easy to use. [55]

The deployment process often consists of a number of interrelated operations that may be affected by modifications made to one another. The actions might take place on either the consumer or producer side, or perhaps on both. Because every software system is unique in some manner, it is impossible to pinpoint the precise procedures or approaches for each activity. Because of this, "deployment" should be seen as a basic process that has to be customized to fit particular requirements or features.

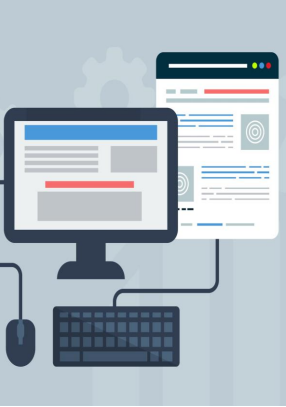


Figure 4-1: Model Deployment

**4.4.1 Introductory Page:**

Introductory page of prototype is given below. Here it is the UI design of the prototype

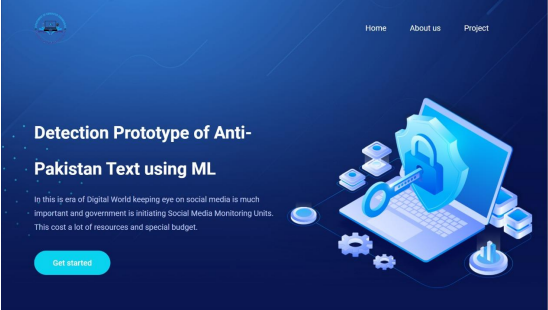


Figure 4-2: UI of prototype

The whole operation of the prototype could be done from this page. UI is designed to provide accessibility. If someone who is not a professional wants to determine the appropriate text, then this interface is ideal for that. With only a single click, one can determine the appropriate text. Thus, instead of running the program every time, we've designed this user interface for easy access and speedy working. The user can visit the Home page, About Us webpage, Project pages from this page.

**4.4.2 About Us page:**

In the About Us section the team members as well as our supervisor's brief introduction are provided. It is also stated which team member was responsible for what portion of the work.

**4.4.3 Project Page:**

Project page is basically that portion where any user can detect the inappropriate text from the cv’s file. User just need to choose the CSV file by clicking on the Choose File button then user needs to write the limits of the tweets which he/she wants to manipulate. After that submit button is needed to press. Then it will show that text is inappropriate or not.

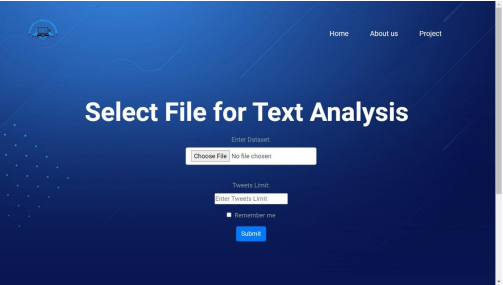


Figure 4-3: Project UI page

**4.4.4 Desired Output:**

Here in this section, the text is completely analyzed with ML and results will be given as “Not Anti Pakistan” or “Anti Pakistan”.

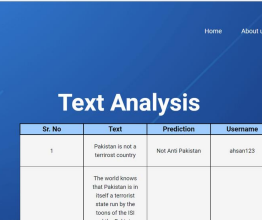


Figure 4-4: Desired Output

Furthermore, the output is as expected and completing all our requirements of the project.

**4.5 Coding for Implementation:**

import pickle

import re

import string

import pandas as pd

from flask import Flask, render\_template, request, url\_for

from werkzeug.utils import redirect

sentiment\_analyzer\_model = pickle.load(open('my\_model.pkl', 'rb'))

def cleanText(x):

x = x.encode('ascii', 'ignore').decode() # remove emojis

x = re.sub(r'https\*\S+', '', x) # remove urls

x = re.sub(r'@\S+', '', x) # remove mentions

x = re.sub(r'#\S+', '', x) # remove hashtags

x = re.sub(r'\'w+', '', x)

return x

def preporocess(tweet):

tweet = [t for t in tweet if t not in string.digits] # removing digits

tweet = ''.join(tweet)

tweet = [t for t in tweet if t not in string.punctuation] # removing punctuations

return ''.join(tweet)

app = Flask(\_\_name\_\_)

@app.route('/', methods=['GET', 'POST'])

def index():

if request.method == "POST":

run\_path = "/home"

return redirect(url\_for('run', run\_path=run\_path))

@app.route('/home')

def svm():

return render\_template("svm.html")

@app.route('/s\_t', methods=["GET", "POST"])

def s\_t():

if request.method == "POST":

f = request.files['file'] limit = int(request.form['limit'])

f.save(f.filename)

# f.load(f.filename)

data = pd.read\_csv(f.filename)

# data['tweet']=data['tweet'].apply(cleanText) # cleaning the text using cleanText function , removing hashes , mentions emojis and urls etc

data['tweet'] = data['tweet'].apply( preporocess) # preprocessing the tweets removing stopwords , punctuations , digits

data['tweet'] = data['tweet'].apply(lambda x: x.strip()) # removing spaces in the begining of the twee

t data['tweet'] = data['tweet'].dropna()

# return f.filename + "file" + str(len(data)) + "<p>" + str( data['tweet'][:10]) + "</p>"

# preds=[]

# pred.append(sentiment\_analyzer\_model.predict([data['tweet'][i]]))

# for i in range(limit):

preds = sentiment\_analyzer\_model.predict(data['tweet'][:limit])

# result={'tweet':data['tweet'][:limit] ,'prediction':pred}

result = {}

for i, j in zip(data['tweet'][:limit], preds, data['username']):

if j == 1.0:

result[i] = 'Anti State'

else:

result[i] = 'Not Anti State'

return render\_template("result.html", result=result)

"""sumary\_line {% for key,value in r.items()%}

<tr>

<td>{{key}} </td>

<td>{{value}} </td>

</tr>

{%endfor%}

data['tweet']=data['tweet'].apply(cleanText)

data['tweet']=data['tweet'].apply(preporocess)

data['tweet']=data['tweet'].apply(lambda x : x.strip())

{% for key,value in r.items()%}

<tr>

<td> {{key}}</td>

<td> {{value}}</td>

</tr>

{%endfor%} return render\_template("svm.html" ,tweet=data['tweet'][:limit], pred\_text=pred)

<td>{% for i in tweet %} {{i}} {%endfor%}</td>

<td> {% for j in pred\_text %} {{j}} {%endfor%} </td>

<table board=”1px solid black”>

|  |
| --- |
| {% for i in tweet%} |
| <tr>  <td>{{i}}</td>  </tr>  {%endfor%}  </table>  Keyword arguments:  argument -- description Return:  return\_description |

if request.method=="POST":

tweet=request.form["tweet"]

pred=int(sentiment\_analyzer\_model.predict([tweet]))

if pred==4:

pred\_text='Positive

elif pred==0:

pred\_text='Negative'

else:

pred\_text='Neutral

return render\_template("svm.html" ,tweet=tweet, pred\_text=pred\_text)"""

if \_\_name\_\_ == "\_\_main\_\_":

app.run(debug=True)'

# Chapter 5

# Conclusions & Future Work:

**5.1 Future Scope:**

• This is a model developed to serve our country. It can easily be expanded to other countries. If this model is extended to a global level, everyone will be connected and won't allow anyone to write any unsuitable content about the sovereignty of the nation.

• Around 707 million cyber security hacks in 2015, with 554 million during the first quarter of 2016 all by itself. Businesses are struggling to keep just one step in front of cybercriminals. USC experts suggest that the auto-learning and automation capabilities provided by AI will help protect data more efficiently and effectively protecting people from terrorist attacks or even small-scale identity theft. AI powered tools search for patterns that are associated with malware and viruses before they are able to steal huge quantities of data or cause chaos. This is why AI is utilized in our project. In the future, it will be able to detect the cyber-security threat and stop it.

• If we can improve the design to alert the defense system of our country, it'll be an extremely effective system. It will instantly alert the defense system of the post that is not appropriate and it will handle it efficiently.

• This project is available on the web, thus it is an online prototype should we improve it in the field of mobile phones that all mobile phones must include this detection system, it could prevent the user from posting the wrong content, which can be very beneficial for security systems.

**5.2 Conclusion:**

A simple and useful project called the Detection Prototype of Anti-Pakistan text that utilizes ML is created and developed here. Through this project, untrue text that is threatening our country is identified and altered. Simply upload the CSV file to the project, and then it'll detect the issue and display the results.

This project is implemented using ML that will detect any irrelevant content on social media and alert the system that it's an Anti-Pakistan text.

The idea of creating an intelligent country that is full of respect and dignity, to show the image of a tranquil nation before the world, which can improve our quality of living. We have shown an idea that uses the most recent programming language, which is Python. AI as well as ML is used in this prototype, which operate independently in a way that is automated and precise. With this prototype our nation is safe from any inappropriate messages from social media.

# References:

1. Bar-Yossef, Z., Kraus, N.: Context-sensitive query autocompletion. In: WWW’11, pp. 107– 116. ACM, New York, NY, USA (2011)

2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J Mach Learn Res3:993–1022. http://dl.acm.org/citation.cfm? id=944919.944937 (2003)

3. Cai, F., de Rijke, M.: A survey of query auto completion in information retrieval. Found. Trends® Inf. Retr. 10(4), 273–363 (2016). <https://doi.org/10.1561/1500000055>

4. Chiu, J.P.C., Nichols, E.: Named Entity Recognition with Bidirectional LTM-CNNs. CoRR abs/1511.08308. http://arxiv.org/abs/ 1511.08308 (2015)

5. Werbos, P.J.: Backpropagation through time: what it does and how to do it. Proc. IEEE78(10), 1550–1560 (1990). [https://doi.org/10. 1109/5.58337](https://doi.org/10.%201109/5.58337)

6. K. Pte, Digital 2019: Global Digital Overview, 2019. <https://datareportal.com/reports/digital-2019-global-digital-overview>

7. Eurostat, Internet Access and Use Statistics – Households and Individuals, 2016. https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Archive:Internet\_access\_and\_use\_statistics\_- \_households\_and\_individuals

8. NSPCC, Net Aware Report 2017: “Freedom to Express Myself Safely”: Exploring How Young People Navigate Opportunities and Risks in Their Online Lives, NSPCC, London, England, 2017. <https://learning.nspcc.org.uk/research-resources/2017/net-awarereport-2017-freedom-toexpress-myself-safely>

9. C.-S. Li, G. Xiong, and E.M. Tapia, New frontiers in cognitive content curation and moderation, APSIPA Trans. Signal Inf. Process., Vol. 7, 2018.

10. M. Yar, A failure to regulate? The demands and dilemmas of tackling illegal content and behaviour on social media, Int. J. Cybersecur. Intell. Cyber., Vol. 1, 2018, pp. 5-20.

11. N. Duarte, E. Llanso, and A. Loup, Mixed messages? The limits of automated social media content analysis, in FAT (New York City, NY, USA), 2018.

12. T. Davidson, D. Warmsley, M. Macy, and I. Weber, Automated hate speech detection and the problem of offensive language, in Eleventh International AAAI Conference on Web and Social Media (Montreal, Quebec, Canada), 2017

13. Abhimanyu Shankhdhar, JIMS / Social media and businss / 52

14. Mahmoudi Sidi Ahmed et al., “Detection of Abnormal Motions in Multimedia”, Chania ICMI-MIAUCE’08 workshop, Crete, Greece, 2008.

15. S.Shabnoor,S.Tajinder,Social Media its Impact with Positive and NegativeAspects IJCATR, Volume 5– Issue 2, 71 - 75, 2016

16. J. Chen, S. Yan, and K.-C. Wong, Verbal aggression detection on Twitter comments: Convolutional neural network for short-text sentiment analysis, Neural Comput. Appl., 2018, pp. 1-10.

17. B.K. Narayanan, S. Moses, M. Nirmala, et al., Adult content filtering: restricting minor audience from accessing inappropriate internet content, Educ. Inf. Technol., Vol. 23, 2018, pp. 2719-2735.

18. R. Baeza-Yates, B. Ribeiro-Neto, et al., Modern Information Retrieval, ACM Press, New York, NY, USA, Vol. 463, 1999.

19. Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807- 7. OCLC 36417892

20. Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F., "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning" IEEE Transactions on Vehicular Technology, 2020.

21. Scott Patterson (13 July 2010). "Letting the Machines Decide". The Wall Street Journal. Retrieved 24 June 2018.

22. "Introduction to AI Part 1". Edzion. 2020-12-08. Retrieved 2020-12-09.

23. Alpaydin, Ethem (2010). Introduction to Machine Learning. MIT Press. p. 9. ISBN 978-0- 262-01243-0.

24. "9 Reasons why your machine learning project will fail". www.kdnuggets.com. Retrieved 2018-08-20.

25 . Vincent, James (2019-04-10). "The first AI-generated textbook shows what robot writers are actually good at". The Verge. Retrieved 2019-05-05.

26. Poole, David; Mack worth, Alan; Goebel, Randy (1998). Computational Intelligence: A Logical Approach. New York: Oxford University Press. ISBN 978-0-19-510270- 3. Archived from the original on 26 July 2020. Retrieved 22 August 2020.

27. Russell, Stuart J.; Norvig, Peter (2003), Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2

28. Nilsson, Nils (1998). Artificial Intelligence: A New Synthesis. Morgan Kaufmann. ISBN 978-1-55860-467-4. Archived from the original on 26 July 2020. Retrieved 18 November 2019.

29. Luger, George; Stubblefield, William (2004). Artificial Intelligence: Structures and Strategies for Complex Problem Solving (5th ed.). Benjamin/Cummings. ISBN 978-0- 8053-4780-7. Archived from the original on 26 July 2020. Retrieved 17 December 2019.

30. Zhang, T.: Solving large scale linear prediction problems using stochastic gradient descent algorithms. In: Proceedings of the Twenty-first International Conference onMachine Learning, ACM, New York, NY, USA, ICM

31. Nagar, Himanshu, Chetna Dabas, and J. P. Gupta. "Navie Bayes and K-Means Hybrid Analysis for Extracting Extremist Tweets", ACM Conference, pp 27-32.

32. A.T.M Shahjahan, K.Chisty, “Social Media research and its effect on our society” International journal of Information 7 communication Engineering , Vol:8, No:6,2014

33. W.Tariq, M.Mehboob, M.A.Khan, F.Ullah “The Impact of social Media and Social Networking on education and Students of Pakistan” international Journal of Computer sciences issues, Vol:9,No:3,July 2012

34. . Xu, Z., Zhu, S.: Filtering offensive language in online communities using grammatical relations. In: Proceedings of the Seventh Annual CEAS 2010 (2010)

35. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by backpropagating errors. Cogn. Model. 5(3), 1 (1988)

36. Razavi, A.H., Inkpen, D., Uritsky, S., Matwin, S.: Offensive language detection using multilevel classification. In: AI’10, Springer-Verlag, pp. 16–27 (2010

37. 2. Huang, P.S., He, X., Gao, J., Deng, L., Acero, A., Heck, L.: Learning deep structured semantic models for web search using Clickthrough Data. In: CIKM’13 (2013) 38. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735– 1780 (1997). [https://doi.org/10.1162/neco. 1997.9.8.1735](https://doi.org/10.1162/neco.%201997.9.8.1735)

39. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. Technical Report UCB/EECS-2010-24. http://www2.eecs.berkeley.edu/ Pubs/TechRpts/2010/EECS-2010-24.html (2010)

40. Chiu, J.P.C., Nichols, E.: Named Entity Recognition with Bidirectional LTM-CNNs. CoRR abs/1511.08308. http://arxiv.org/abs/ 1511.08308 (2015) 54

41. Amir Wasim. "Terrorism dogs Pakistan in ’06: Over 900 killed in 657 attacks" Dawn, 7 January 2007

42. Attacks, clashes claim 3,448 lives in 2007 Daily Times, 5 January 2008

43. Muhammad Ali Siddiqi. "8,000 dead: is the world aware?" Dawn, 27 January 2009

44. Ali Usman. "HRCP launches report on human rights situation: 2008: a year of new challenges" Daily Times, 7 April 2009

45. Javed Aziz Khan. "889 killed, 2,072 hurt in 61 suicide attacks this year" The News, 22 December 2008

46. Amir Wasim. "Over 12,800 militants caught in 2009" Dawn, 11 January 2010

47. "Home - PakistanBodyCount.org". pakistanbodycount.org. Retrieved 6 August 2015.

48. "Suicides & Drones Analytics - PakistanBodyCount.org". pakistanbodycount.org. Retrieved 6 August 2015.

49. National Consortium for the Study of Terrorism and Responses to Terrorism. (2016). Global Terrorism Database (globalterrorismdb\_0616dist.xlsx Archived 10 July 2016 at the Wayback Machine). Retrieved from https://www.start.umd.edu/gtd University of Maryland

50. Islamabad, Kabul look inwards as Tehran blames US, Express Tribune 51. V. Vapnik, Statistical Learning Theory, Wiley-Interscience, 1998.

52. X. Shu, R. Cohen, et al., Natural Language Toolkit (NLTK), 2010.

53. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, in Advances in Neural Information Processing Systems (Nevada, USA), 2013, pp. 3111-3119.

54. T. Mikolov, Q.V. Le, and I. Sutskever, Exploiting Similarities among Languages for Machine Translation, 2013. arXiv preprint arXiv:1309.4168

55. X. Zhang, J. Zhao, and Y. LeCun, Character-level convolutional networks for text classification, in Advances in Neural Information Processing Systems (Montreal, Quebec, Canada), 2015, pp. 649-657.

# Appendix B: Project Timeline

**DATE**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PROJECT ID** | | |  | |  | **TOTAL NUMBER OF WEEKS IN**  **PLAN** | |  |
|  | | | | | | | | |
| **TITLE** | |  | | | | | | |
|  | | | | | | | | |
| **No.** | **STARTING WEEK** | | | **DESCRIPTION OF MILESTONE** | | | **DURATION** | |
| 1 |  | | |  | | |  | |
| 2 |  | | |  | | |  | |
| 3 |  | | |  | | |  | |
| 4 |  | | |  | | |  | |
| 5 |  | | |  | | |  | |
| 6 |  | | |  | | |  | |
| 7 |  | | |  | | |  | |
| 8 |  | | |  | | |  | |
| 9 |  | | |  | | |  | |
| 10 |  | | |  | | |  | |
|  |  | | |  | | |  | |
|  |  | | |  | | |  | |
|  |  | | |  | | |  | |
|  |  | | |  | | |  | |
|  |  | | |  | | |  | |
|  |  | | |  | | |  | |

*\* You can provide Gantt chart instead of filling this form, if you like*

# Appendix C: User Manual

How a person will use your project.