**TWITTER SPAM DETECTION USING MACHINE LEARNING**

A MAJOR PROJECT REPORT SUBMITTED

TO

**OSMANIA UNIVERSITY, HYDERABAD**



IN PARTIAL FULFILMENT OF THE AWARD FOR THE DEGREE OF MASTER OF COMPUTER APPLICATIONS

Submitted by

Md. Zeeshan

1304-21-862-152



DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS

AURORA’S POST GRADUATE COLLEGE (MBA)

(Affiliated to Osmania University)

Ramanthapur, Hyderabad – 500013



CERTIFICATE

This is to certify that the project work entitled

**TWITTER SPAM DETECTION USING MACHINE LEARNING**

is a bonafide work done by

Md. Zeeshan

1304-21-862-152

as part of the curriculum in the

DEPARTMENT OF MASTER OF COMPUTER

APPLICATIONS

AURORA’S POST GRADUATE COLLEGE (MCA)

Ramanthapur, Hyderabad – 500013 in partial fulfillment of requirement for the award of

Master of Computer Applications

Osmania University, Hyderabad

This work has been carried out under our guidance.

Internal Guide Head of Department Principal

Internal Examiner External Examiner



**MANAC INFOTECH Pvt Ltd**

CERTIFICATE

This is to certify that Md Zeeshan bearing the Reg. No: 1304-21-862-152 from AURORA’S POST GRADUATE COLLEGE, RAMANTHAPUR, student of MCA, has successfully completed the project titled “**TWITTER SPAM DETECTION USING MACHINE LEARNING**” as part of the course curriculum in our organization.

He/ She has done the project using python Technologies using the period 04-06-2023 to 25-09-2023, under the guidance and supervision of Mr. A. Akash, from MANAC INFOTECH Pvt Ltd, Hyderabad.

He/She has completed the assigned project well within the time frame. He/She is sincere hardworking and his conduct during the period is commendable.

We wish all the best in his/her future endeavour

For MANAC INFOTECH Pvt Ltd

M.KONDAL REDDY,

(Manager-HR)

**DECLARATION**

I, **Md. Zeeshan**hereby declare that the project entitled “**TWITTER SPAM DETECTION USING MACHINE LEARNING**” has been carried out by me in the MANAC INFOTECH Pvt Ltd. This project is submitted to Osmania University, Hyderabad in partial fulfillment of requirements for the award of the degree of “MASTERS OF COMPUTER APPLICATIONS”. The results embodied in this dissertation have not been submitted to any other University or institution for the award of Degree or Diploma.

(Signature of the student)

Md.Zeeshan

1304-21-862-152

ACKNOWLEDGEMENT

I would like to express deep gratitude and respect to all those people behind the scene who guided and inspired in the completion of this project work.

I wish to convey my sincere thanks to Principal and Head of Department of MCA, our project in charge Mr. MOHAMMED ISMAIL and project guide Ms. ANITHA for giving me the required guidance during this project work.

Last but not least I am very thankful to all the faculty members of my college and friends for their suggestions and help to me in successfully completing this project.

Md Zeeshan

1304-21-862-152

**TABLE OF CONTENTS**

### TITLE PAGE NO

1. **ORGANIZATION PROFILE**  1-3
2. **ABSTRACT 4-5** 
   1. Introduction 6
   2. Existing system 7
   3. Proposed system 7
3. **SYSTEM ANALYSIS 08** 
   1. Description of Modules 09
   2. Feasibility Study 9-10
   3. Software Requirements Specification 11
4. **SYSTEM DESIGN 12** 
   1. UML Diagrams 13-20
5. **IMPLEMENTATION 21**
   1. Software technology 22-49
   2. Sample source code 50-66
   3. Screen Shots 67-71
6. **TESTING** 72-77
7. **SCOPE**  78-79
8. **CONCLUSION 80-81**

1. **BIBLIOGRAPHY & WEBLIOGRAPHY** 82-84

**ORGANIZATION PROFILE**

**ABOUT US**

MANAC Infotech Pvt. Ltd. (MIPL) is a progressive and innovative IT company offering a spectrum of services including software development, staffing, recruitment and training. An IIM alumnus company started in 1998 as a training partner of TCS iON. Tata Consultancy Services Limited (TCS Ltd.) is a leading IT solutions company & one of the world's leading Information Technology consulting, services and business process outsourcing organizations and a part of the Tata Group, India's best-known business conglomerate. MANAC has built a suite of services around its core offering of training.

In the training arena MANAC has an unenviable track record of grooming several hundreds of students, building their skills and launching their careers. Alumni of MANAC are in senior IT positions across the globe. MANAC also offers a host of training programs tailor made for IT companies.

The development arm of MIPL focuses on providing solutions to the Education & Training sector including assessments, analytics, online course delivery and ERP for training companies. Along the way, MANAC has developed a strong expertise in JAVA, Android and open source technologies.

Relations built with HR managers over the years have helped MANAC provide specialized services in the area of staffing and fresher’s recruitment for IT & ITES companies.

Having started with a strong student focus, MANAC continues to provide other support services to B.Tech students including project guidance, internships and placement support.

**CLIENT PROFILE**

**Name : FAZAL UR RAHMAN**

**Company : MANAC INFOTECH PVT LTD.**

**Designation : SOFTWARE TRAINER**

**Experience : 10 YEARS**

**Skills : ANDROID, PYTHON, JAVA, SQL, C#.NET, ASP.NET,**

**HTML5, CSS3**

Real-Time Projects: Some of the Applications Developed recently for clients in India & abroad.

[INDO-AMERICANS RECRUITERS](http://indo-americans.com/) (http://www.indo-americans.com)

[ZAM ZAM RESTAURANT](http://zamzamrestaurant.co.nz/) (http://www.zamzamrestaurant.co.nz)

[MANAC INFOTECH](http://manacinfotech.com/) (http://www.manacinfotech.com)

ZEN RECRUITERS (http://www.zenrecruiters.com)

[EVIZA](http://eviza.in/) [(http://www.eviza.in)](http://www.eviza.in/)

[AZ SEMICON](http://azsemicon.com/) [(http://www.azsemicon.com)](http://www.azsemicon.com/)

[ZINFOTECH](http://zinfotech.com/) [(http://www.zinfotech.com)](http://www.zinfotech.com/)

[RELITE OPTICALS](http://reliteopticals.com/) (http://www.reliteopticals.com)

**ABSTRACT**

**ABSTRACT**

Now a days, social media platforms have become an important part of our existence. The social media networks like Facebook, Instagram,Twitter, SanpChat and YouTube are used for communication among people and source of promoting businesses. Twitter is an excessive communication and sharing platform, where people can share their emotions and promote their businesses by using 140 character messages. More than 42 millions Twitter accounts are created every month. Twitter’s receptiveness to spamming has prompted the prominence of activities on Twitter. Twitter spam is a very a sophisticated issue however it’s difficult to unravel. So far, previous research has suggested a number of detection and defense methods that secure the Twitter users from spammers. So, we are going to work on spam detection techniques of Twitter. This study consists of 3 sections: 1- Background about spam detection on Twitter. 2- A literature review comparative analysis of machine learning, deep learning and hybrid algorithms. 3- Discussion on limitation of previous studies and future directions.

### 2.1 INTRODUCION

In recent years, millions of Internet users have been able to communicate and collaborate on social media online networks (OSN) . Today, we have entered the age of online social networks OSN . Interest in this issue has been growing misinformation spread online on social media. Facebook, Twitter, and LinkedIn are the most prominent social media platforms that enable users to communicate with each other, use information and communicate in a meaningful way. Twitter is a great communication platform and sharing, it attracts profiles when provided services for spreading 140-character messages. Every month, the number of new accounts increasing more than 42 million on Twitter. Companies and individuals impressed the supremacy of quickly sharing information, it also performance as a smart power for the sender unsolicited and uncontroversial messages over the Twitter. This kind of data or messages are understood as spam data or messages. Though, due to the immense fame of Twitter, it also attracts the attention of cybercriminals (such as spammers).

Manual filtering of messages or data from Twitter is the starting lines. Traditional machine learning methods used for spam detection models and automated spam detection methods are also started with the utilization. For the spam filtering simple blacklisting, content-based and conversational spam detection techniques of data mining methods used. These types of methods done fairly on large data or emails messages, but identifying the spam is being a big challenge from small and noisy spam detection platforms day by day. In short, from the domain of Twitter and SMS, it is more difficult to identify the cause of the spam, noise and small length of messages and emails. In this research, use different machine learning, deep learning techinques and compare their performance on larger datasets. Also, squeeze and compare performance as well as the number of features extracted.

### 2.2 EXISTING SYSTEM

## Existing System

Traditional methods used for spam detection models and automated spam detection methods are also started with the utilization. For the spam filtering simple blacklisting, content-based and conversational spam detection techniques of data mining methods used. These type of methods done fairly on large data or emails messages, but, identifying the spam is being a big challenge from small and noisy spam detection platforms day by day. In short, from the domain of Twitter and SMS, it is more difficult to identify the cause of the spam, noise and small length of messages and emails.

## DISADVANTAGES

* + 1. Detecting the spam is being a big challenge from small and noisy spam.
    2. More difficulty identify the reason to identify

### 2.3 PROPOSED SYSTEM

Several techniques have been suggested to combat spam. To automatically detect spam, researchers have implemented data mining algorithms to make spam detection a classification issue. Data mining has many types but in this research, spamming extraction from tweets by using deep learning and machine learning. Machine learning uses two main techniques: Supervised learning allows you to collect data or produce a data output from a previous ML deployment. Unsupervised learning find the hidden pattern or data grouping without the requirement for human intercession.

## ADVANTAGES

* + 1. Using different type of machine learning algorithms
    2. Detecting spam from twitter most effectively

**SYSTEM ANALYSIS**

### 3.1 DESCRIPTION OF MODULES

**MODULES**

* + USER

**MODULES DESCRIPTION:**

* **USER:** Here user is a module, user should register with the application after successful login into the

application he can perform some operations such as register, login, update,

Load data, View data ,Preprocess and split data and logout.

Load Data: Here we load the data.

View Data: we can able to see data.

Split Data: data splitting done here.

Detect Spam: generate model and svm predictions.

### 3.2 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY

**ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

**TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**SOCIAL FEASIBILITY**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

## 3.3 SOFTWARE REQUIREMENTS SPECIFICATION

**Software Requirements**

For developing the application the following are the Software Requirements:

1. Python
2. Django

**Operating Systems supported**

1. Windows 7
2. Windows XP
3. Windows 8

**Technologies and Languages used to Develop**

1. Python

**Debugger and Emulator**

* Any Browser (Particularly Chrome)

**Hardware Requirements**

For developing the application the following are the Hardware Requirements:

* Processor: Pentium IV or higher
* RAM: 256 MB
* Space on Hard Disk: minimum 512MB

# SYSTEM DESIGN

### 4.1 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

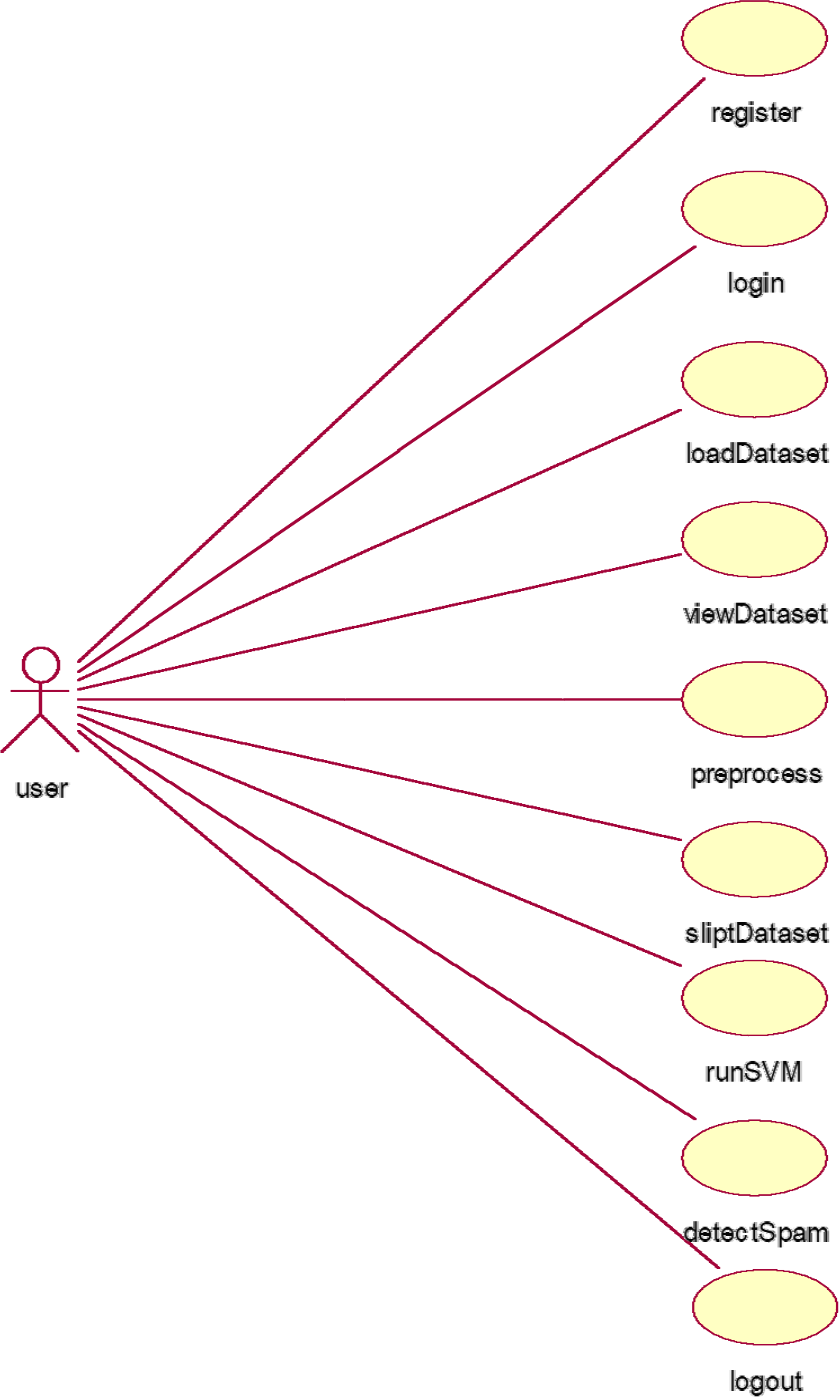
**GOALS:**

The Primary goals in the design of the UML are as follows:

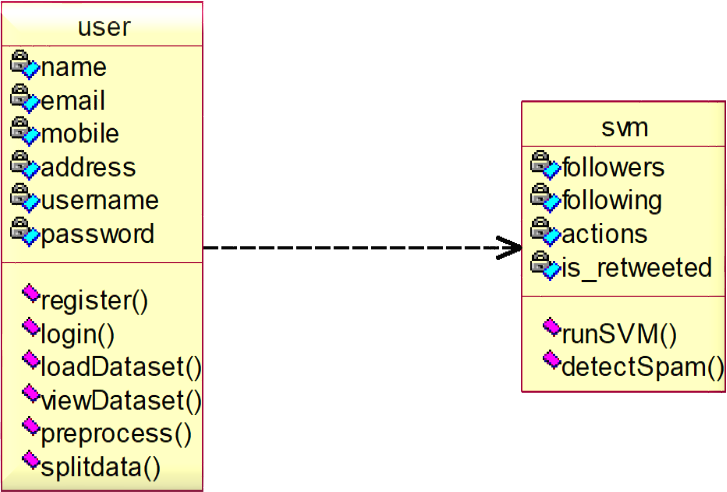
1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components. 7. Integrate best practices.

**USE CASE DIAGRAM:**

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

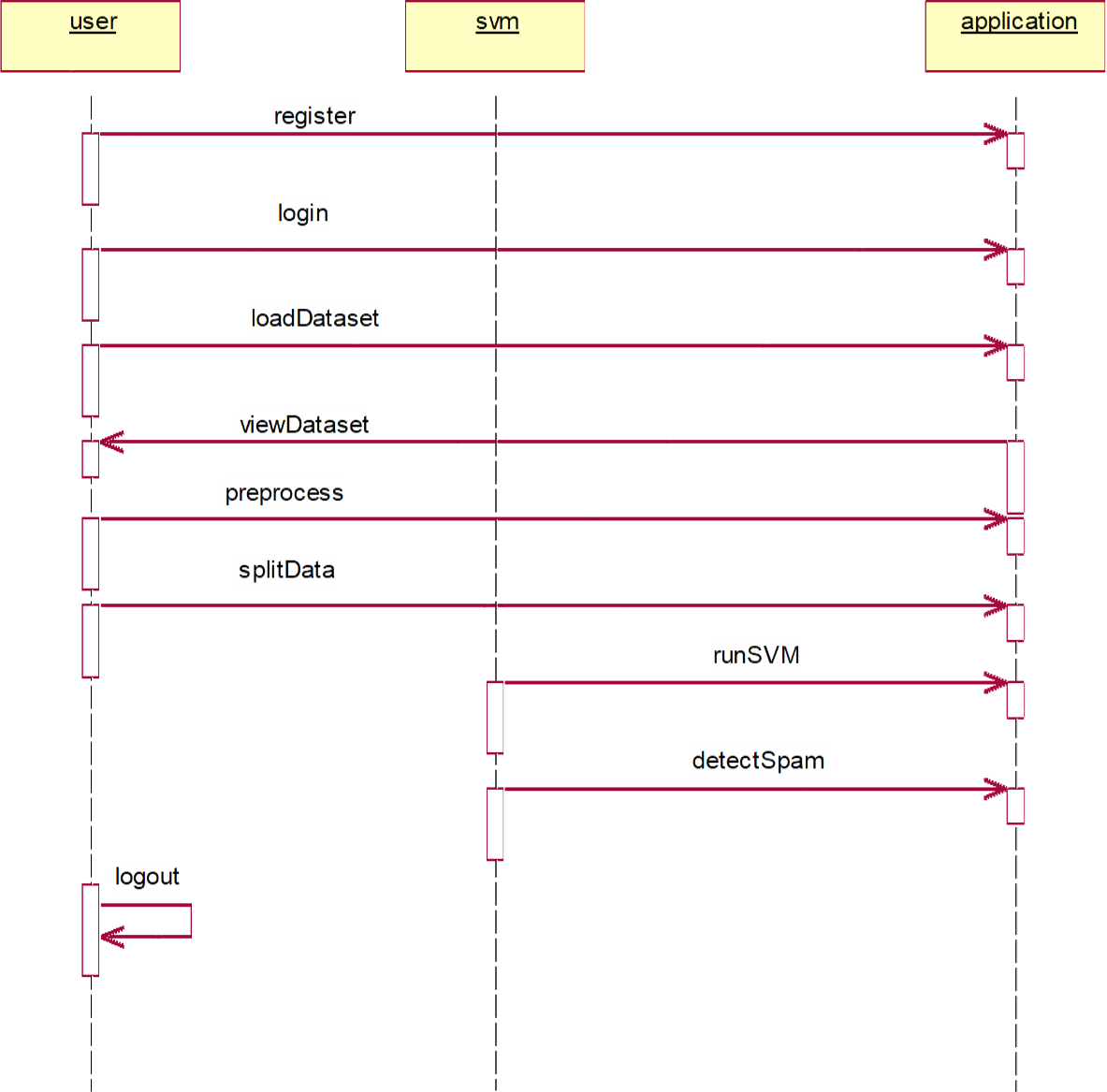


### CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

### SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.





### Collaboration Diagram:

Component diagram:

The main elements of component diagram are components.it has dependency, generalization, association and realization relationships. It includes interfaces to specify the external behaviour of a component. Component diagrams are basically used to model static view of a system. This can be achieved by modelling various physical components.



Deployment Diagram:

A deployment diagram shows the configuration of run time processing nodes and the components that live on them. The deployment view contains the node instances connected by communication links. Deployment diagrams are used to visualize the static aspect of the physical nodes and their relationship and specifies their details for constructions. The main contents of a deployment diagram are nodes, dependency and association reltionships. These diagrams may contain nodes and constraints, apart from components. These diagrams may also contain packages or sub systems. The deployment view usually addresses the distribution, delivery and installation of the parts that makeup the physical system.



# IMPLEMENTATION

### 5.1 SOFTWARE TECHNOLOGY

### What is Python :-

Below are some facts about Python.

Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.

Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.

The biggest strength of Python is huge collection of standard library which can be used for the following –

* [Machine Learning](https://www.geeksforgeeks.org/machine-learning/)
* GUI Applications (like Kivy, Tkinter, PyQt etc. )
* Web frameworks like Django (used by YouTube, Instagram, Dropbox)
* Image processing (like Opencv, Pillow)
* Web scraping (like Scrapy, BeautifulSoup, Selenium)
* Test frameworks
* Multimedia

Advantages of Python :-

Let’s see how Python dominates over other languages.

1. Extensive Libraries

Python downloads with an extensive library and it contain code for various purposeslike regular expressions, documentation-generation, unit-testing, web browsers,

threading, databases, CGI, email, image manipulation, and more*.* So, we don’t have to write the complete code for that manually.

1. Extensible

As we have seen earlier, Python can be **extended to other languages**. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.

1. Embeddable

Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add **scripting capabilities** to our code in the other language.

1. Improved Productivity

The language’s simplicity and extensive libraries render programmers **more productive** than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.

1. IOT Opportunities

Since Python forms the basis of new platforms like Raspberry Pi, it finds the future bright for the Internet Of Things. This is a way to connect the language with the real world.

1. Simple and Easy

When working with Java, you may have to create a class to print **‘Hello World’**. But in Python, just a print statement will do. It is also quite **easy to learn, understand,** and **code.** This is why when people pick up Python, they have a hard time adjusting to other more verbose languages like Java.

1. Readable

Because it is not such a verbose language, reading Python is much like reading English. This is the reason why it is so easy to learn, understand, and code. It also does not need curly braces to define blocks, and **indentation is mandatory.** This further aids the readability of the code.

1. Object-Oriented

This language supports both the **procedural and object-oriented** programming paradigms. While functions help us with code reusability, classes and objects let us model the real world. A class allows the **encapsulation of data** and functions into one.

1. Free and Open-Source

Like we said earlier, Python is **freely available.** But not only can you [**download**](https://data-flair.training/blogs/install-python-windows/)[**Python**](https://data-flair.training/blogs/install-python-windows/)for free, but you can also download its source code, make changes to it, and even distribute it. It downloads with an extensive collection of libraries to help you with your tasks.

1. Portable

When you code your project in a language like C++, you may need to make some changes to it if you want to run it on another platform. But it isn’t the same with Python. Here, you need to **code only once**, and you can run it anywhere. This is called **Write Once Run Anywhere (WORA)**. However, you need to be careful enough not to include any system-dependent features.

1. Interpreted

Lastly, we will say that it is an interpreted language. Since statements are executed one by one, **debugging is easier** than in compiled languages.

Any doubts till now in the advantages of Python? Mention in the comment section.

### Advantages of Python Over Other Languages

1. Less Coding

Almost all of the tasks done in Python requires less coding when the same task is done in other languages. Python also has an awesome standard library support, so you don’t have to search for any third-party libraries to get your job done. This is the reason that many people suggest learning Python to beginners.

1. Affordable

Python is free therefore individuals, small companies or big organizations can leverage the free available resources to build applications. Python is popular and widely used so it gives you better community support.

### The 2019 Github annual survey showed us that Python has overtaken Java in the most popular programming language category.

1. Python is for Everyone

Python code can run on any machine whether it is Linux, Mac or Windows. Programmers need to learn different languages for different jobs but with Python, you can professionally build web apps, perform data analysis and [**machine learning**](https://data-flair.training/blogs/machine-learning-tutorials-home/), automate things, do web scraping and also build games and powerful visualizations. It is an all-rounder programming language.

### Disadvantages of Python

So far, we’ve seen why Python is a great choice for your project. But if you choose it, you should be aware of its consequences as well. Let’s now see the downsides of choosing Python over another language.

1. Speed Limitations

We have seen that Python code is executed line by line. But since [Python](https://www.python.org/) is interpreted, it often results in **slow execution**. This, however, isn’t a problem unless speed is a focal point for the project. In other words, unless high speed is a requirement, the benefits offered by Python are enough to distract us from its speed limitations.

1. Weak in Mobile Computing and Browsers

While it serves as an excellent server-side language, Python is much rarely seen on the **client-side**. Besides that, it is rarely ever used to implement smartphone-based applications. One such application is called **Carbonnelle**.

The reason it is not so famous despite the existence of Brython is that it isn’t that secure.

**25**

1. Design Restrictions

As you know, Python is **dynamically-typed**. This means that you don’t need to declare the type of variable while writing the code. It uses **duck-typing**. But wait, what’s that? Well, it just means that if it looks like a duck, it must be a duck. While this is easy on the programmers during coding, it can **raise run-time errors**.

1. Underdeveloped Database Access Layers

Compared to more widely used technologies like **JDBC (Java DataBase Connectivity)** and **ODBC (Open DataBase Connectivity)**, Python’s database access layers are a bit underdeveloped. Consequently, it is less often applied in huge enterprises.

1. Simple

No, we’re not kidding. Python’s simplicity can indeed be a problem. Take my example. I don’t do Java, I’m more of a Python person. To me, its syntax is so simple that the verbosity of Java code seems unnecessary.

This was all about the Advantages and Disadvantages of Python Programming Language.

### History of Python : -

What do the alphabet and the programming language Python have in common? Right, both start with ABC. If we are talking about ABC in the Python context, it's clear that the programming language ABC is meant. ABC is a general-purpose programming language and programming environment, which had been developed in the Netherlands, Amsterdam, at the CWI (Centrum Wiskunde &Informatica). The greatest achievement of ABC was to influence the design of Python.Python was conceptualized in the late 1980s. Guido van Rossum worked that time in a project at the CWI, called Amoeba, a distributed operating system. In an interview with Bill Venners1, Guido van Rossum said: "In the early 1980s, I worked as an implementer on a team building a language called ABC at Centrum voor Wiskunde en Informatica (CWI). I don't know how well people know ABC's influence on Python. I try to mention ABC's influence because I'm indebted to everything I

learned during that project and to the people who worked on it."Later on in the same Interview, Guido van Rossum continued: "I remembered all my experience and some of my frustration with ABC. I decided to try to design a simple scripting language that possessed some of ABC's better properties, but without its problems. So I started typing. I created a simple virtual machine, a simple parser, and a simple runtime. I made my own version of the various ABC parts that I liked. I created a basic syntax, used indentation for statement grouping instead of curly braces or begin-end blocks, and developed a small number of powerful data types: a hash table (or dictionary, as we call it), a list, strings, and numbers."

### What is Machine Learning : -

Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parametersthat can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain.Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

### Categories Of Machine Leaning :-

At the most fundamental level, machine learning can be categorized into two main types: supervised learning and unsupervised learning.

Supervised learninginvolves somehow modeling the relationship between measured features of data and some label associated with the data; once this model is determined, it can be used to apply labels to new, unknown data. This is further subdivided into classificationtasks and regressiontasks: in classification, the labels are discrete categories, while in regression, the labels are continuous quantities. We will see examples of both types of supervised learning in the following section.

Unsupervised learninginvolves modeling the features of a dataset without reference to any label, and is often described as "letting the dataset speak for itself." These models include tasks such as clusteringand dimensionality reduction*.* Clustering algorithms identify distinct groups of data, while dimensionality reduction algorithms search for more succinct representations of the data. We will see examples of both types of unsupervised learning in the following section.

Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven’t surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, “to make decisions, based on data, with efficiency and scale”.

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data- driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programing logic, in the problems that cannot be programmed inherently. The fact is that we can’t do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

Challenges in Machines Learning :-

While Machine Learning is rapidly evolving, making significant strides with cybersecurity and autonomous cars, this segment of AI as whole still has a long

way to go. The reason behind is that ML has not been able to overcome number of challenges. The challenges that ML is facing currently are −

**Quality of data** − Having good-quality data for ML algorithms is one of the biggest challenges. Use of low-quality data leads to the problems related to data preprocessing and feature extraction.

**Time-Consuming task** − Another challenge faced by ML models is the consumption of time especially for data acquisition, feature extraction and retrieval.

**Lack of specialist persons** − As ML technology is still in its infancy stage, availability of expert resources is a tough job.

**No clear objective for formulating business problems** − Having no clear objective and well-defined goal for business problems is another key challenge for ML because this technology is not that mature yet.

**Issue of overfitting & underfitting** − If the model is overfitting or underfitting, it cannot be represented well for the problem.

**Curse of dimensionality** − Another challenge ML model faces is too many features of data points. This can be a real hindrance.

**Difficulty in deployment** − Complexity of the ML model makes it quite difficult to be deployed in real life.

Applications of Machines Learning :-

Machine Learning is the most rapidly growing technology and according to researchers we are in the golden year of AI and ML. It is used to solve many real-world complex problems which cannot be solved with traditional approach. Following are some real-world applications of ML −

* Emotion analysis
* Sentiment analysis
* Error detection and prevention
* Weather forecasting and prediction
* Stock market analysis and forecasting
* Speech synthesis
* Speech recognition
* Customer segmentation
* Object recognition
* Fraud detection
* Fraud prevention
* Recommendation of products to customer in online shopping

How to Start Learning Machine Learning?

Arthur Samuel coined the term **“Machine Learning”** in 1959 and defined it as a **“Field of study that gives computers the capability to learn without being explicitly programmed”.**

And that was the beginning of Machine Learning! In modern times, Machine Learning is one of the most popular (if not the most!) career choices. According to [Indeed](http://blog.indeed.com/2019/03/14/best-jobs-2019/), Machine Learning Engineer Is The Best Job of 2019 with a *344%* growth and an average base salary of **$146,085** per year.

But there is still a lot of doubt about what exactly is Machine Learning and how to start learning it? So this article deals with the Basics of Machine Learning and also the path you can follow to eventually become a full-fledged Machine Learning Engineer. Now let’s get started!!!

### How to start learning ML?

This is a rough roadmap you can follow on your way to becoming an insanely talented Machine Learning Engineer. Of course, you can always modify the steps according to your needs to reach your desired end-goal!

Step 1 – Understand the Prerequisites

In case you are a genius, you could start ML directly but normally, there are some prerequisites that you need to know which include Linear Algebra, Multivariate Calculus, Statistics, and Python. And if you don’t know these, never fear! You don’t

need a Ph.D. degree in these topics to get started but you do need a basic understanding.

1. Learn Linear Algebra and Multivariate Calculus

Both Linear Algebra and Multivariate Calculus are important in Machine Learning. However, the extent to which you need them depends on your role as a data scientist. If you are more focused on application heavy machine learning, then you will not be that heavily focused on maths as there are many common libraries available. But if you want to focus on R&D in Machine Learning, then mastery of Linear Algebra and Multivariate Calculus is very important as you will have to implement many ML algorithms from scratch.

1. Learn Statistics

Data plays a huge role in Machine Learning. In fact, around 80% of your time as an ML expert will be spent collecting and cleaning data. And statistics is a field that handles the collection, analysis, and presentation of data. So it is no surprise that you need to learn it!!! Some of the key concepts in statistics that are important are Statistical Significance, Probability Distributions, Hypothesis Testing, Regression, etc. Also, Bayesian Thinking is also a very important part of ML which deals with various concepts like Conditional Probability, Priors, and Posteriors, Maximum Likelihood, etc.

1. Learn Python

Some people prefer to skip Linear Algebra, Multivariate Calculus and Statistics and learn them as they go along with trial and error. But the one thing that you absolutely cannot skip is [Python](https://www.geeksforgeeks.org/python-programming-language/)! While there are other languages you can use for Machine Learning like R, Scala, etc. Python is currently the most popular language for ML. In fact, there are many Python libraries that are specifically useful for Artificial Intelligence and Machine Learning such as [Keras](https://keras.io/), [TensorFlow](https://www.tensorflow.org/), [Scikit-learn](https://scikit-learn.org/stable/), etc.

So if you want to learn ML, it’s best if you learn Python! You can do that using various online resources and courses such as [**Fork Python**](https://practice.geeksforgeeks.org/courses/fork-python)available Free on GeeksforGeeks.

### Step 2 – Learn Various ML Concepts

Now that you are done with the prerequisites, you can move on to actually learning ML (Which is the fun part!!!) It’s best to start with the basics and then move on to the more complicated stuff. Some of the basic concepts in ML are:

1. Terminologies of Machine Learning

* **Model –** A model is a specific representation learned from data by applying some machine learning algorithm. A model is also called a hypothesis.
* **Feature –** A feature is an individual measurable property of the data. A set of numeric features can be conveniently described by a feature vector. Feature vectors are fed as input to the model. For example, in order to predict a fruit, there may be features like color, smell, taste, etc.
* **Target (Label) –** A target variable or label is the value to be predicted by our model. For the fruit example discussed in the feature section, the label with each set of input would be the name of the fruit like apple, orange, banana, etc.
* **Training –** The idea is to give a set of inputs(features) and it’s expected outputs(labels), so after training, we will have a model (hypothesis) that will then map new data to one of the categories trained on.
* **Prediction –** Once our model is ready, it can be fed a set of inputs to which it will provide a predicted output(label).

1. Types of Machine Learning

* **Supervised Learning –** This involves learning from a training dataset with labeled data using classification and regression models. This learning process continues until the required level of performance is achieved.
* **Unsupervised Learning –** This involves using unlabelled data and then finding the underlying structure in the data in order to learn more and more about the data itself using factor and cluster analysis models.
* **Semi-supervised Learning –** This involves using unlabelled data like Unsupervised Learning with a small amount of labeled data. Using labeled data vastly increases the learning accuracy and is also more cost-effective than Supervised Learning.
* **Reinforcement Learning –** This involves learning optimal actions through trial and error. So the next action is decided by learning behaviors that are based on the current state and that will maximize the reward in the future.

### Advantages of Machine learning :-

1. Easily identifies trends and patterns -

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

1. No human intervention needed (automation)

With ML, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

1. Continuous Improvement

As [**ML algorithms**](https://data-flair.training/blogs/machine-learning-algorithms/)gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

1. Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

1. Wide Applications

You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

### Disadvantages of Machine Learning :-

1. Data Acquisition

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

1. Time and Resources

ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

1. Interpretation of Results

Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose.

1. High error-susceptibility

[Machine Learning](https://en.wikipedia.org/wiki/Machine_learning) is autonomous but highly susceptible to errors. Suppose you train an algorithm with data sets small enough to not be inclusive. You end up with biased predictions coming from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of ML, such blunders can set off a chain of errors that can go undetected for long periods of time. And when they do get noticed, it takes quite some time to recognize the source of the issue, and even longer to correct it.

### Python Development Steps : -

Guido Van Rossum published the first version of Python code (version 0.9.0) at alt.sources in February 1991. This release included already exception handling, functions, and the core data types of list, dict, str and others. It was also object oriented and had a module system. Python version 1.0 was released in January 1994. The major new features included in this release were the functional programming tools lambda, map, filter and reduce, which Guido Van Rossum never liked.Six and a half years later in October 2000, Python 2.0 was introduced. This release included list comprehensions, a full garbage collector and it was supporting unicode.Python flourished for another 8 years in the versions 2.x before the next major release as Python 3.0 (also known as "Python 3000" and "Py3K") was released. Python 3 is not backwards compatible with Python

2.x. The emphasis in Python 3 had been on the removal of duplicate programming constructs and modules, thus fulfilling or coming close to fulfilling the 13th law of the Zen of Python: "There should be one -- and preferably only one -- obvious way to do it."Some changes in Python 7.3:

* + Print is now a function
  + Views and iterators instead of lists
  + The rules for ordering comparisons have been simplified. E.g. a heterogeneous list cannot be sorted, because all the elements of a list must be comparable to each other.
  + There is only one integer type left, i.e. int. long is int as well.
  + The division of two integers returns a float instead of an integer. "//" can be used to have the "old" behaviour.
  + Text Vs. Data Instead Of Unicode Vs. 8-bit

### Purpose :-

We demonstrated that our approach enables successful segmentation of intra-retinal layers—even with low-quality images containing speckle noise, low contrast, and different intensity ranges throughout—with the assistance of the ANIS feature.

### Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

* + Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
  + Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

### Modules Used in Project :-

**Tensorflow**

TensorFlow is a [free](https://en.wikipedia.org/wiki/Free_software) and [open-source](https://en.wikipedia.org/wiki/Open-source_software) [software library for dataflow and differentiable](https://en.wikipedia.org/wiki/Library_(computing)) [programming](https://en.wikipedia.org/wiki/Library_(computing)) across a range of tasks. It is a symbolic math library, and is also used for [machine learning](https://en.wikipedia.org/wiki/Machine_learning) applications such as [neural networks](https://en.wikipedia.org/wiki/Neural_networks). It is used for both research and production at [Google](https://en.wikipedia.org/wiki/Google).

TensorFlow was developed by the [Google Brain](https://en.wikipedia.org/wiki/Google_Brain) team for internal Google use. It was released under the [Apache 2.0](https://en.wikipedia.org/wiki/Apache_License) [open-source license](https://en.wikipedia.org/wiki/Open-source_license) on November 9, 2015.

**NUMPY**

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

* A powerful N-dimensional array object
* Sophisticated (broadcasting) functions
* Tools for integrating C/C++ and Fortran code
* Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi- dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of databases.

### Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

### Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and [IPython](http://ipython.org/) shells, the [Jupyter](http://jupyter.org/) Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the [sample plots](https://matplotlib.org/tutorials/introductory/sample_plots.html) and [thumbnail gallery.](https://matplotlib.org/gallery/index.html)

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control

of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

### Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. **Python**

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

* + Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
  + Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

### Install Python Step-by-Step in Windows and Mac :

Python a versatile programming language doesn’t come pre-installed on your computer devices. Python was first released in the year 1991 and until today it is a very popular high-level programming language. Its style philosophy emphasizes code readability with its notable use of great whitespace.

The object-oriented approach and language construct provided by Python enables programmers to write both clear and logical code for projects. This software does not come pre-packaged with Windows.

How to Install Python on Windows and Mac :

There have been several updates in the Python version over the years. The question is how to install Python? It might be confusing for the beginner who is willing to start learning Python but this tutorial will solve your query. The latest or the newest version of Python is version 3.7.4 or in other words, it is Python 3.

**Note:** The python version 3.7.4 cannot be used on Windows XP or earlier devices.

Before you start with the installation process of Python. First, you need to know about your **System Requirements**. Based on your system type i.e. operating system and based processor, you must download the python version. My system type is a **Windows 64-bit operating system**. So the steps below are to install python version 3.7.4 on Windows 7 device or to install Python 3. [Download the Python Cheatsheet here.](https://myelearninghub.com/python-cheat-sheet/)The steps on how to install Python on Windows 10, 8 and 7 are **divided into 4 parts** to help understand better.

Download the Correct version into the system

**Step 1:** Go to the official site to download and install python using Google Chrome or any other web browser. OR Click on the following link: [https://www.python.org](https://www.python.org/)

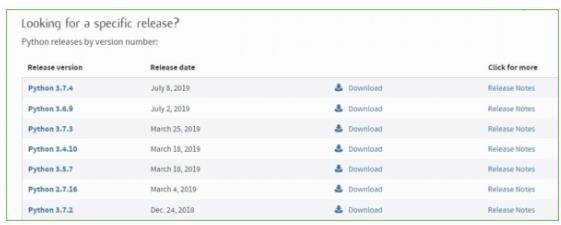


Now, check for the latest and the correct version for your operating system.

**Step 2:** Click on the Download Tab.



**Step 3:** You can either select the Download Python for windows 3.7.4 button in Yellow Color or you can scroll further down and click on download with respective to their version. Here, we are downloading the most recent python version for windows 3.7.4



**Step 4:** Scroll down the page until you find the Files option.

**Step 5:** Here you see a different version of python along with the operating system.



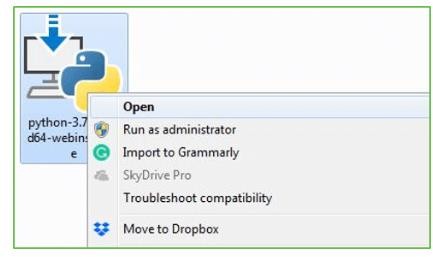
* + - To download Windows 32-bit python, you can select any one from the three options: Windows x86 embeddable zip file, Windows x86 executable installer or Windows x86 web-based installer.
    - To download Windows 64-bit python, you can select any one from the three options: Windows x86-64 embeddable zip file, Windows x86-64 executable installer or Windows x86-64 web-based installer.

Here we will install Windows x86-64 web-based installer. Here your first part regarding which version of python is to be downloaded is completed. Now we move ahead with the second part in installing python i.e. Installation

**Note:** To know the changes or updates that are made in the version you can click on the Release Note Option.

Installation of Python

**Step 1:** Go to Download and Open the downloaded python version to carry out the installation process.



**Step 2:** Before you click on Install Now, Make sure to put a tick on Add Python

3.7 to PATH.



**Step 3:** Click on Install NOW After the installation is successful. Click on Close.



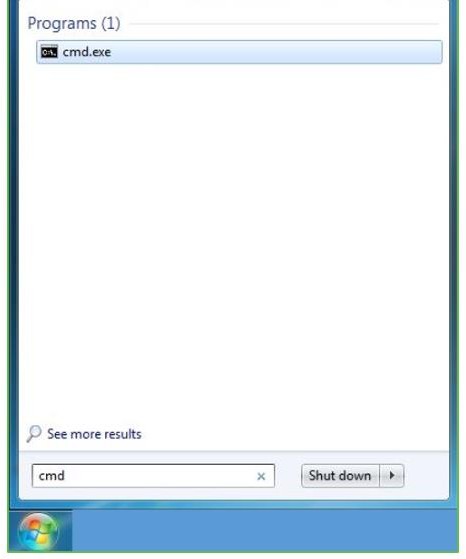
With these above three steps on python installation, you have successfully and correctly installed Python. Now is the time to verify the installation.

**Note:** The installation process might take a couple of minutes.

Verify the Python Installation

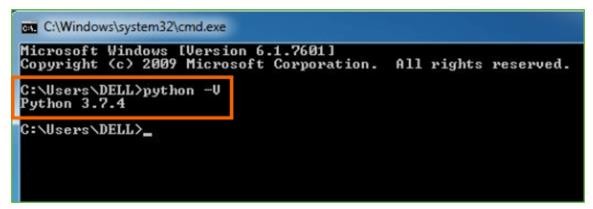
**Step 1:** Click on Start

**Step 2:** In the Windows Run Command, type “cmd”.



**Step 3:** Open the Command prompt option.

**Step 4:** Let us test whether the python is correctly installed. Type **python –V** and press Enter.



**Step 5:** You will get the answer as 3.7.4

**Note:** If you have any of the earlier versions of Python already installed. You must first uninstall the earlier version and then install the new one.

Check how the Python IDLE works

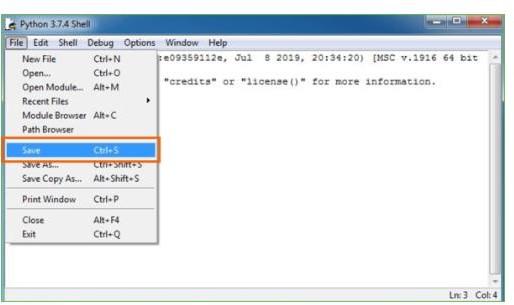
**Step 1:** Click on Start

**Step 2:** In the Windows Run command, type “python idle”.



**Step 3:** Click on IDLE (Python 3.7 64-bit) and launch the program

**Step 4:** To go ahead with working in IDLE you must first save the file. **Click on File > Click on Save**



**Step 5:** Name the file and save as type should be Python files. Click on SAVE. Here I have named the files as Hey World.

**Step 6:** Now for e.g. **enter print**

**Step 3:** Click on Install NOW After the installation is successful. Click on Close.



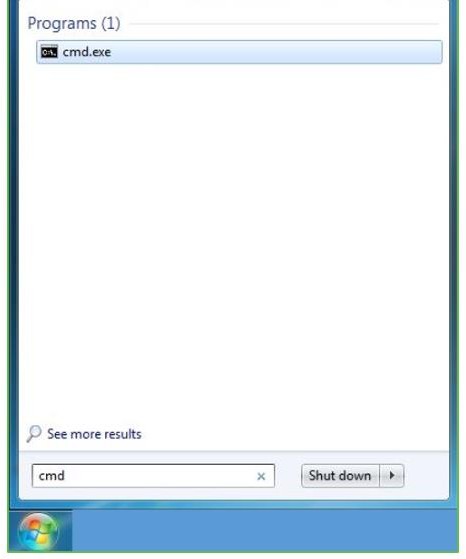
With these above three steps on python installation, you have successfully and correctly installed Python. Now is the time to verify the installation.

**Note:** The installation process might take a couple of minutes.

Verify the Python Installation

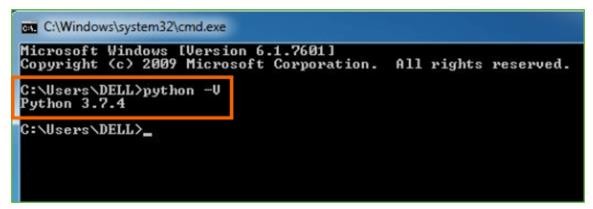
**Step 1:** Click on Start

**Step 2:** In the Windows Run Command, type “cmd”.



**Step 3:** Open the Command prompt option.

**Step 4:** Let us test whether the python is correctly installed. Type **python –V** and press Enter.



**Step 5:** You will get the answer as 3.7.4

**Note:** If you have any of the earlier versions of Python already installed. You must first uninstall the earlier version and then install the new one.

Check how the Python IDLE works

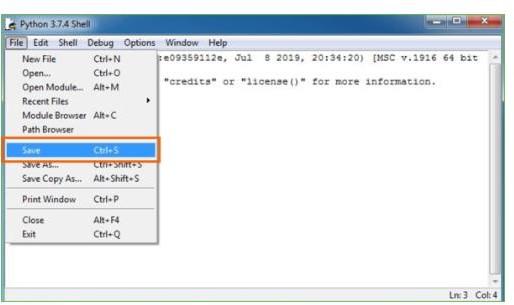
**Step 1:** Click on Start

**Step 2:** In the Windows Run command, type “python idle”.



**Step 3:** Click on IDLE (Python 3.7 64-bit) and launch the program

**Step 4:** To go ahead with working in IDLE you must first save the file. **Click on File > Click on Save**



**Step 5:** Name the file and save as type should be Python files. Click on SAVE. Here I have named the files as Hey World.

**Step 6:** Now for e.g. **enter print**

### 5.2 SOURCE CODE

## HOME.HTML

<!DOCTYPE html>

{% load static %}

<html lang="en">

<head>

<meta charset="UTF-8">

<meta http-equiv="X-UA-Compatible" content="IE=edge">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Detect Spam</title>

<link href="{%static 'https://fonts.googleapis.com/css2?family=Source+Sans+Pro:wght@300;400&display=swap'

%}" rel="stylesheet">

<link rel="stylesheet" href="{%static 'css/bootstrap.min.css'%}">

<link rel="stylesheet" href="{%static 'css/slick.css'%}" type="text/css" />

<link rel="stylesheet" href="{%static 'css/templatemo-style.css'%}">

</head>

<body>

<video autoplay muted loop id="bg-video">

<source src="{%static 'video/gfp-astro-timelapse.mp4'%}" type="video/mp4">

</video>

<div class="page-container">

<div class="container-fluid">

<div class="row">

<div class="col-xs-12">

<div class="cd-slider-nav">

<nav class="navbar navbar-expand-lg" id="tm-nav">

<a class="navbar-brand" href="#">Detect Spam From Twitter</a>

<button class="navbar-toggler" type="button" data-bs- toggle="collapse" data-bs-target="#navbar-supported-content" aria- controls="navbarSupportedContent" aria-expanded="false" aria-label="Toggle navigation">

content">

<span class="navbar-toggler-icon"></span>

</button>

<div class="collapse navbar-collapse" id="navbar-supported-

no="1">Home</a>

<ul class="navbar-nav mb-2 mb-lg-0">

<li class="nav-item selected">

<a class="nav-link" aria-current="page" href="#0" data-

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="#0" data-no="2">User</a>

<div class="circle"></div>

</li>

</ul>

</div>

</nav>

</div>

</div>

</div>

</div>

<div class="container-fluid tm-content-container">

<ul class="cd-hero-slider mb-0 py-5">

<li class="px-3" data-page-no="1">

<div class="page-width-1 page-left">

<div class="d-flex position-relative tm-border-top tm-border-bottom

intro-container">

<div class="intro-left tm-bg-dark">

<font color="red">{{data}}</font>

<h2 class="mb-4">Spam Detection</h2>

<p class="mb-4">

Several techniques have been suggested to combat spam. To

automatically detect spam, researchers have implemented data mining algorithms to

make spam detection a classification issue. Data mining has many types but in this research, </p>

<p class="mb-0">

Supervised learning allows you to collect data.

</a> for more information. </p>

</div>

<div class="intro-right">

<img src="{%static 'img/home-img-1.jpg'%}" alt="Image" class="img-fluid intro-img-1">

<img src="{%static 'img/home-img-2.jpg'%}" alt="Image" class="img-fluid intro-img-2">

</div>

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

</div>

</div>

</li>

<li data-page-no="2" class="px-3">

<div class="position-relative page-width-1 page-right tm-border-top tm- border-bottom">

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

<div class="tm-bg-dark content-pad">

<h2 class="mb-4">User Login</h2>

<form action="LogAction" method="POST" class="contact-

form">

{% csrf\_token %}

<div class="input-group tm-mb-30">

<input name="username" type="text" class="form-control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="UserName">

</div>

<div class="input-group tm-mb-30">

<input name="password" type="password" class="form- control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Password">

</div>

value="Login">

<div class="input-group justify-content-end">

<input type="submit" class="btn btn-primary tm-btn-pad-2"

&nbsp;

<a href="register"><input type="button"

class="btn btn-primary tm-btn-pad-2" value="Register"></a>

</div>

</form>

</div>

</div>

</li>

<li data-page-no="3" class="px-3">

<div class="position-relative page-width-1 page-right tm-border-top tm- border-bottom">

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

<div class="tm-bg-dark content-pad">

<h2 class="mb-4">Admin Login</h2>

class="contact-form">

<form action="ALogAction" method="POST"

{% csrf\_token %}

<div class="input-group tm-mb-30">

<input name="username" type="text" class="form-control

rounded-0 border-top-0 border-end-0 border-start-0" placeholder="UserName">

</div>

<div class="input-group tm-mb-30">

<input name="password" type="password" class="form- control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Password">

</div>

value="Login">

<div class="input-group justify-content-end">

<input type="submit" class="btn btn-primary tm-btn-pad-2"

</div>

</form>

</div>

</div>

</li>

</ul>

</div>

<div class="container-fluid">

<footer class="row mx-auto tm-footer">

<div class="col-md-6 px-0">

Copyright 2021 . All rights reserved.

</div>

<div class="col-md-6 px-0 tm-footer-right">

Designed by <a rel="sponsored" href="https://templatemo.com" target="\_blank" class="tm-link-white">GKV</a>

</div>

</footer>

</div>

</div>

<!-- Preloader, https://ihatetomatoes.net/create-custom-preloading-screen/ -->

<div id="loader-wrapper">

<div id="loader"></div>

<div class="loader-section section-left"></div>

<div class="loader-section section-right"></div>

</div>

<script src="{%static 'js/jquery-3.5.1.min.js'%}"></script>

<script src="{%static 'js/bootstrap.min.js'%}"></script>

<script src="{%static 'js/slick.js'%}"></script>

<script src="{%static 'js/templatemo-script.js'%}"></script>

</body>

</html>

## REGISTER.HTML

<!DOCTYPE html>

{% load static %}

<html lang="en">

<head>

<meta charset="UTF-8">

<meta http-equiv="X-UA-Compatible" content="IE=edge">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Detect Spam</title>

<link href="{%static 'https://fonts.googleapis.com/css2?family=Source+Sans+Pro:wght@300;400&display=swap'

%}" rel="stylesheet">

<link rel="stylesheet" href="{%static 'css/bootstrap.min.css'%}">

<link rel="stylesheet" href="{%static 'css/slick.css'%}" type="text/css" />

<link rel="stylesheet" href="{%static 'css/templatemo-style.css'%}">

</head>

<body>

<video autoplay muted loop id="bg-video">

<source src="{%static 'video/gfp-astro-timelapse.mp4'%}" type="video/mp4">

</video>

<div class="page-container">

<div class="container-fluid">

<div class="row">

<div class="col-xs-12">

<div class="cd-slider-nav">

<nav class="navbar navbar-expand-lg" id="tm-nav">

<a class="navbar-brand" href="#">Detect Spam From Twitter</a>

<button class="navbar-toggler" type="button" data-bs- toggle="collapse" data-bs-target="#navbar-supported-content" aria- controls="navbarSupportedContent" aria-expanded="false" aria-label="Toggle navigation">

content">

no="1">Home</a>

<span class="navbar-toggler-icon"></span>

</button>

<div class="collapse navbar-collapse" id="navbar-supported-

<ul class="navbar-nav mb-2 mb-lg-0">

<li class="nav-item selected">

<a class="nav-link" aria-current="page" href="#0" data-

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="#0" data-no="2">User</a>

<div class="circle"></div>

</li>

</ul>

</div>

</nav>

</div>

</div>

</div>

</div>

<div class="container-fluid tm-content-container">

<ul class="cd-hero-slider mb-0 py-5">

<li class="px-3" data-page-no="1">

<div class="page-width-1 page-left">

<div class="d-flex position-relative tm-border-top tm-border-bottom

intro-container">

<div class="intro-left tm-bg-dark">

<h2 class="mb-4">Spam Detection</h2>

<p class="mb-4">

Several techniques have been suggested to combat spam. To

automatically detect spam, researchers have implemented data mining algorithms to make spam detection a classification issue. Data mining has many types but in this research, </p>

<p class="mb-0">

Supervised learning allows you to collect data.

</a> for more information. </p>

</div>

<div class="intro-right">

<img src="{%static 'img/home-img-1.jpg'%}" alt="Image" class="img-fluid intro-img-1">

<img src="{%static 'img/home-img-2.jpg'%}" alt="Image" class="img-fluid intro-img-2">

</div>

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

</div>

</div>

</li>

<li data-page-no="2" class="px-3">

<div class="position-relative page-width-1 page-right tm-border-top tm- border-bottom">

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

<div class="tm-bg-dark content-pad">

<h2 class="mb-4">Register Here</h2>

form">

<form action="RegAction" method="POST" class="contact-

{% csrf\_token %}

<div class="input-group tm-mb-30">

<input name="name" type="text" class="form-control

rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Name">

</div>

<div class="input-group tm-mb-30">

<input name="email" type="email" class="form-control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Email">

</div>

<div class="input-group tm-mb-30">

<input name="mobile" type="number" class="form-control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Mobile">

</div>

<div class="input-group tm-mb-30">

<input name="address" type="text" class="form-control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Address">

</div>

<div class="input-group tm-mb-30">

<input name="username" type="text" class="form-control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="UserName">

</div>

<div class="input-group tm-mb-30">

<input name="password" type="password" class="form- control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Password">

</div>

value="Register">

<div class="input-group justify-content-end">

<input type="submit" class="btn btn-primary tm-btn-pad-2"

&nbsp;

<a href="index"><input type="button"

class="btn btn-primary tm-btn-pad-2" value="Login"></a>

</div>

</form>

</div>

</div>

</li>

<li data-page-no="3" class="px-3">

<div class="position-relative page-width-1 page-right tm-border-top tm- border-bottom">

<div class="circle intro-circle-1"></div>

<div class="circle intro-circle-2"></div>

<div class="circle intro-circle-3"></div>

<div class="circle intro-circle-4"></div>

<div class="tm-bg-dark content-pad">

<h2 class="mb-4">Admin Login</h2>

class="contact-form">

<form action="ALogAction" method="POST"

{% csrf\_token %}

<div class="input-group tm-mb-30">

<input name="username" type="text" class="form-control

rounded-0 border-top-0 border-end-0 border-start-0" placeholder="UserName">

</div>

<div class="input-group tm-mb-30">

<input name="password" type="password" class="form- control rounded-0 border-top-0 border-end-0 border-start-0" placeholder="Password">

</div>

value="Login">

<div class="input-group justify-content-end">

<input type="submit" class="btn btn-primary tm-btn-pad-2"

</div>

</form>

</div>

</div>

</li>

</ul>

</div>

<div class="container-fluid">

<footer class="row mx-auto tm-footer">

<div class="col-md-6 px-0">

Copyright 2021 . All rights reserved.

</div>

<div class="col-md-6 px-0 tm-footer-right">

Designed by <a rel="sponsored" href="https://templatemo.com" target="\_blank" class="tm-link-white">GKV</a>

</div>

</footer>

</div>

</div>

<!-- Preloader, https://ihatetomatoes.net/create-custom-preloading-screen/ -->

<div id="loader-wrapper">

<div id="loader"></div>

<div class="loader-section section-left"></div>

<div class="loader-section section-right"></div>

</div>

<script src="{%static 'js/jquery-3.5.1.min.js'%}"></script>

<script src="{%static 'js/bootstrap.min.js'%}"></script>

<script src="{%static 'js/slick.js'%}"></script>

<script src="{%static 'js/templatemo-script.js'%}"></script>

</body>

</html>

## USER HOME.HTML

<!DOCTYPE html>

{% load static %}

<html lang="en">

<head>

<meta charset="UTF-8">

<meta http-equiv="X-UA-Compatible" content="IE=edge">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Detect Spam</title>

<link rel="stylesheet" href="{%static 'css/bootstrap.min.css'%}">

<link rel="stylesheet" href="{%static 'css/slick.css'%}" type="text/css" />

<link rel="stylesheet" href="{%static 'css/templatemo-style.css'%}">

</head>

<body>

<video autoplay muted loop id="bg-video">

<source src="{%static 'video/gfp-astro-timelapse.mp4'%}" type="video/mp4">

</video>

<div class="page-container">

<div class="container-fluid">

<div class="row">

<div class="col-xs-12">

<div class="cd-slider-nav">

<nav class="navbar navbar-expand-lg" id="tm-nav">

<a class="navbar-brand" href="#">Detect Spam</a>

<button class="navbar-toggler" type="button" data-bs- toggle="collapse" data-bs-target="#navbar-supported-content" aria- controls="navbarSupportedContent" aria-expanded="false" aria-label="Toggle navigation">

content">

<span class="navbar-toggler-icon"></span>

</button>

<div class="collapse navbar-collapse" id="navbar-supported-

<ul class="navbar-nav mb-2 mb-lg-0">

no="1">Home</a>

Data</a>

Data</a>

<li class="nav-item selected">

<a class="nav-link" aria-current="page" href="#0" data-

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="loaddata" data-no="2">Load

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="viewdata" data-no="3">View

<div class="circle"></div>

</li>

no="4">Preprocess</a>

<li class="nav-item">

<a class="nav-link" href="preprocess" data-

<div class="circle"></div>

</li>

</ul>

</div>

</nav>

<nav class="navbar navbar-expand-lg" id="tm-nav">

<button class="navbar-toggler" type="button" data-bs- toggle="collapse" data-bs-target="#navbar-supported-content" aria- controls="navbarSupportedContent" aria-expanded="false" aria-label="Toggle navigation">

<span class="navbar-toggler-icon"></span>

</button>

content">

<div class="collapse navbar-collapse" id="navbar-supported-

<ul class="navbar-nav mb-2 mb-lg-0">

<li class="nav-item selected">

<a class="nav-link" href="splitdata" data-

no="5">SplitData</a>

SVM</a>

Spam</a>

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="generatemodel" data-no="6">Run

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="detectspam" data-no="7">Detect

<div class="circle"></div>

</li>

<li class="nav-item">

<a class="nav-link" href="index" data-no="8">Logout</a>

<div class="circle"></div>

</li>

</ul>

</div>

</nav>

</div>

</div>

</div>

</div>

<div class="container-fluid tm-content-container">

</hr>

{{data}}

</div>

<div class="container-fluid">

<footer class="row mx-auto tm-footer">

<div class="col-md-6 px-0">

Copyright 2021 . All rights reserved.

</div>

<div class="col-md-6 px-0 tm-footer-right">

Designed by <a rel="sponsored" href="https://templatemo.com" target="\_blank" class="tm-link-white">GKV</a>

</div>

</footer>

</div>

</div>

<!-- Preloader, https://ihatetomatoes.net/create-custom-preloading-screen/ -->

<div id="loader-wrapper">

<div id="loader"></div>

<div class="loader-section section-left"></div>

<div class="loader-section section-right"></div>

</div>

<script src="{%static 'js/jquery-3.5.1.min.js'%}"></script>

<script src="{%static 'js/bootstrap.min.js'%}"></script>

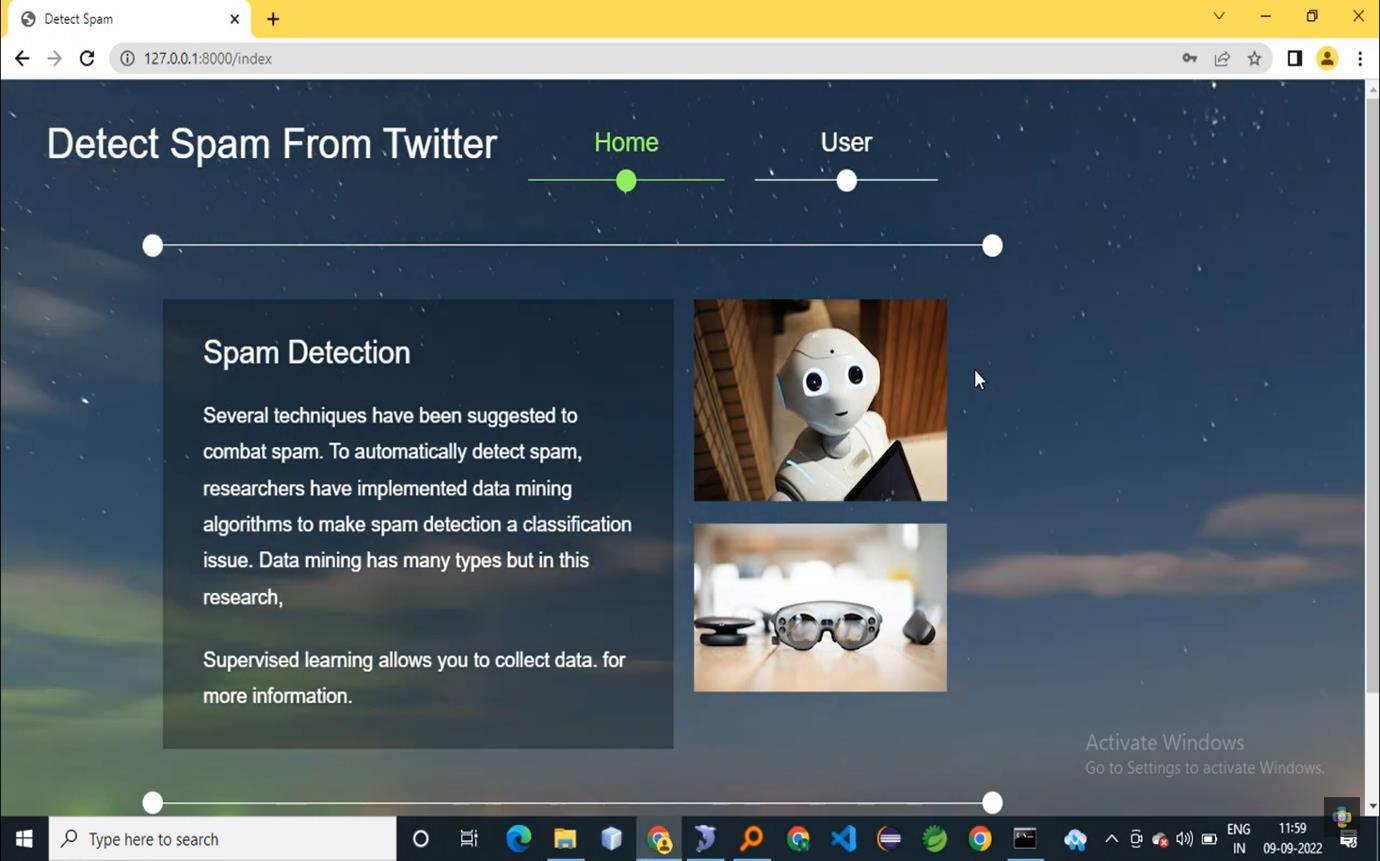
<script src="{%static 'js/slick.js'%}"></script>

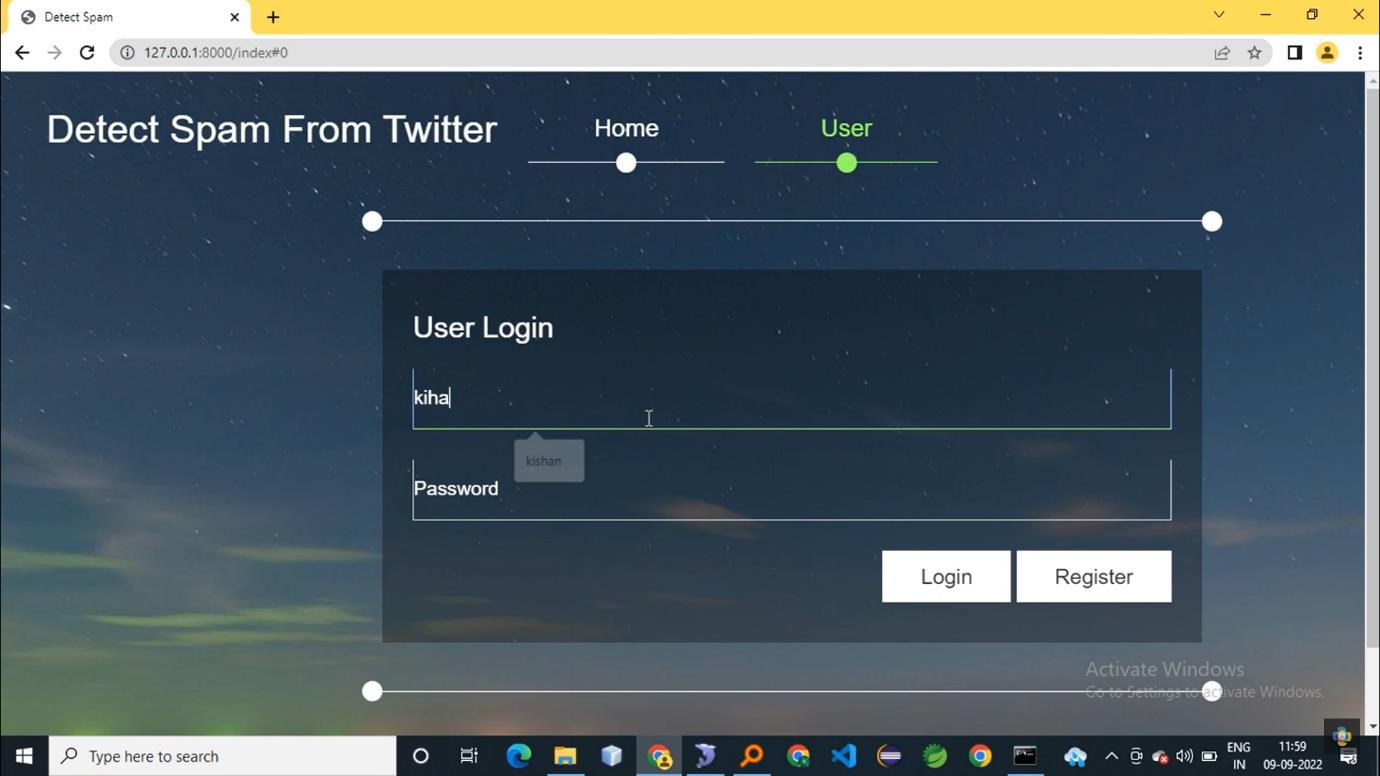
<script src="{%static 'js/templatemo-script.js'%}"></script>

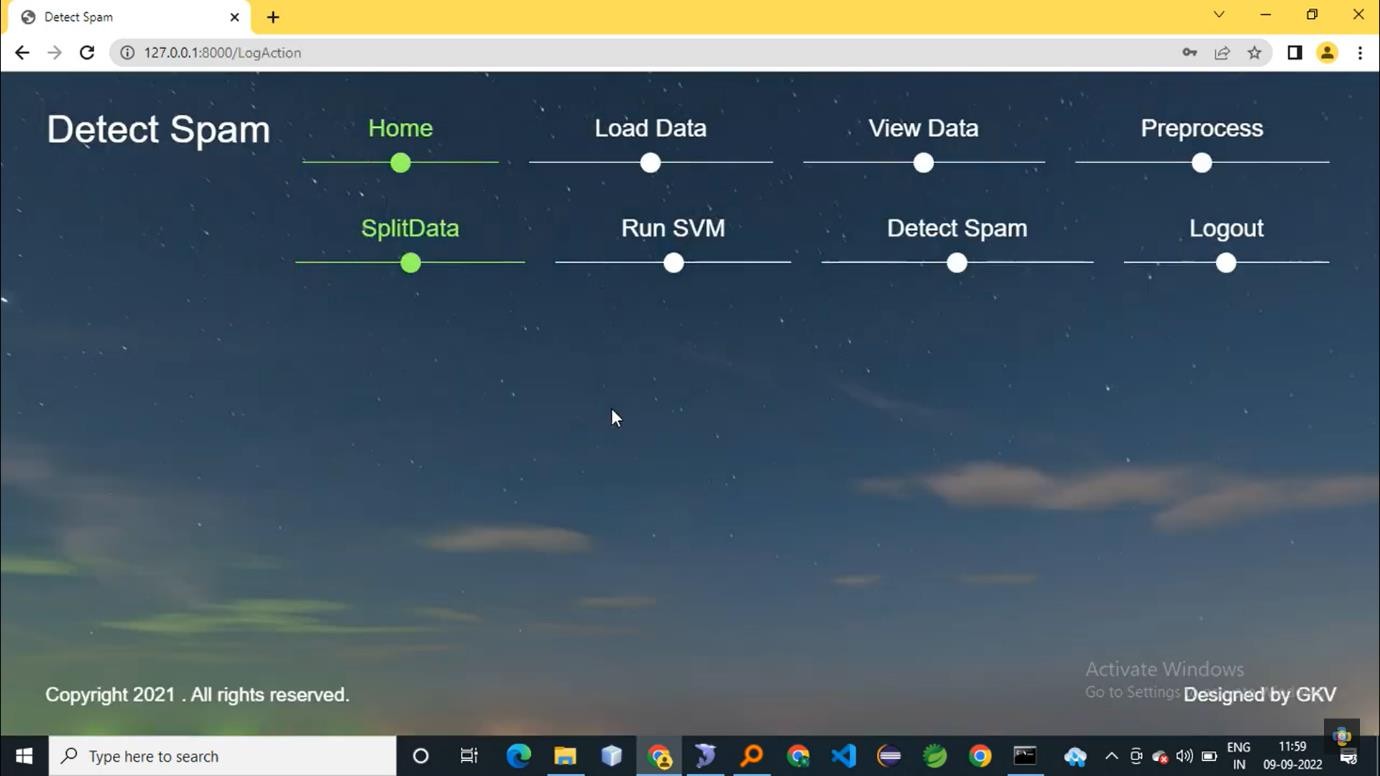
</body>

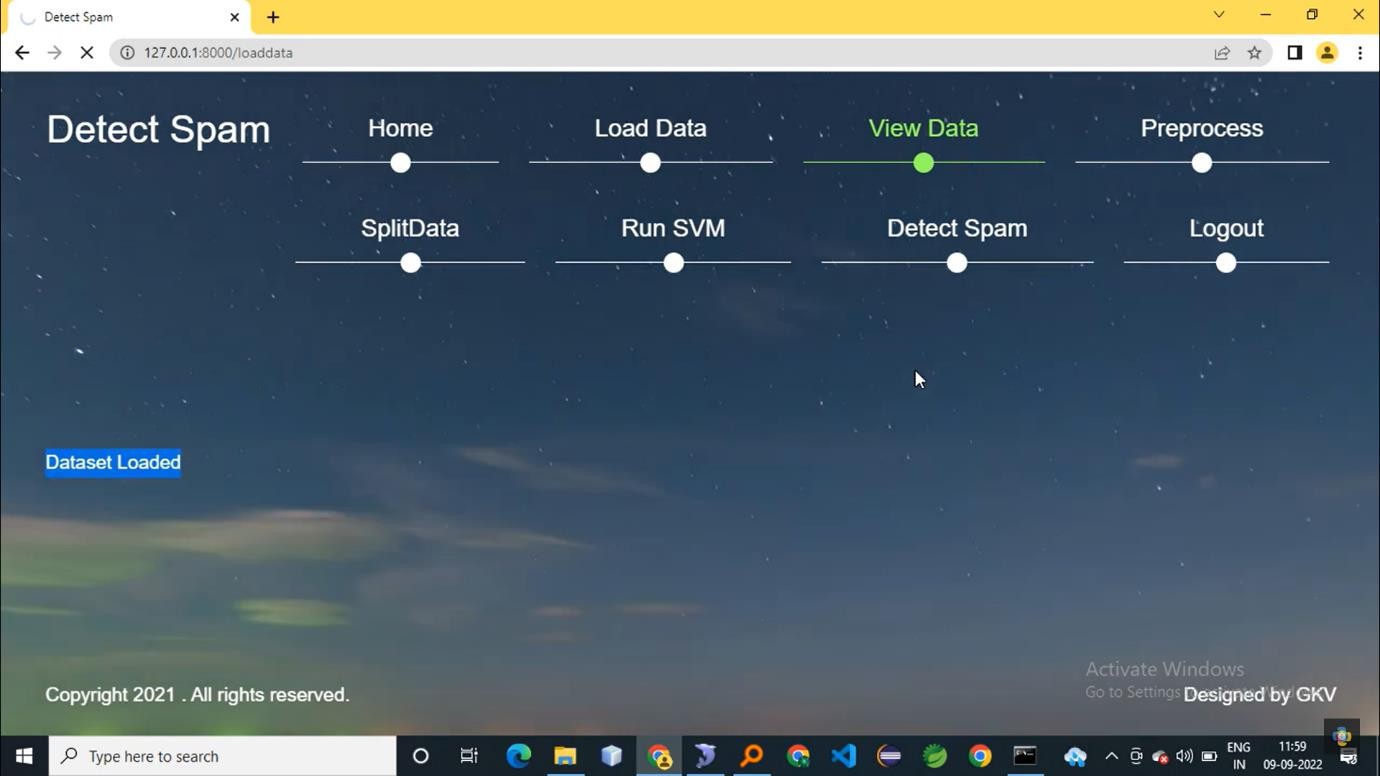
</html>

**5.3 SCREENSHOTS**

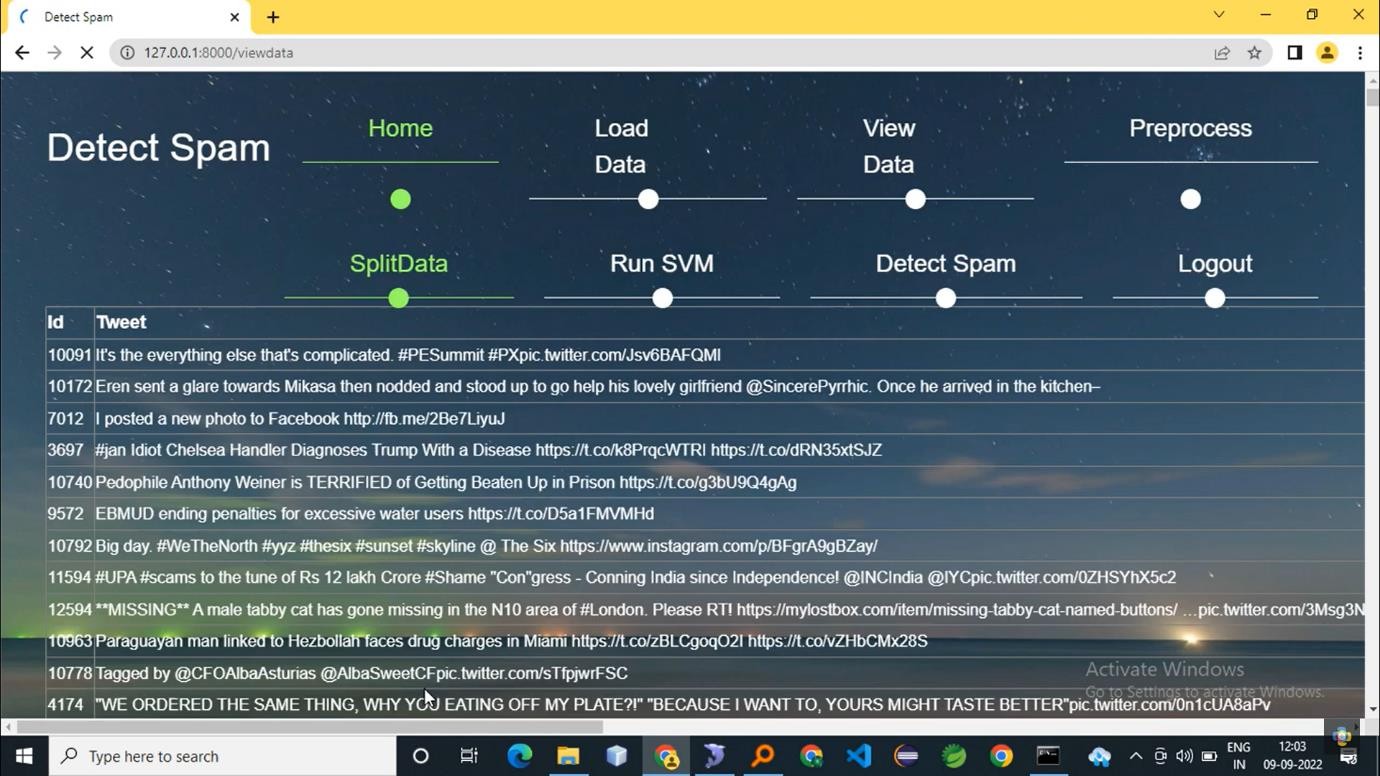
 HOME:

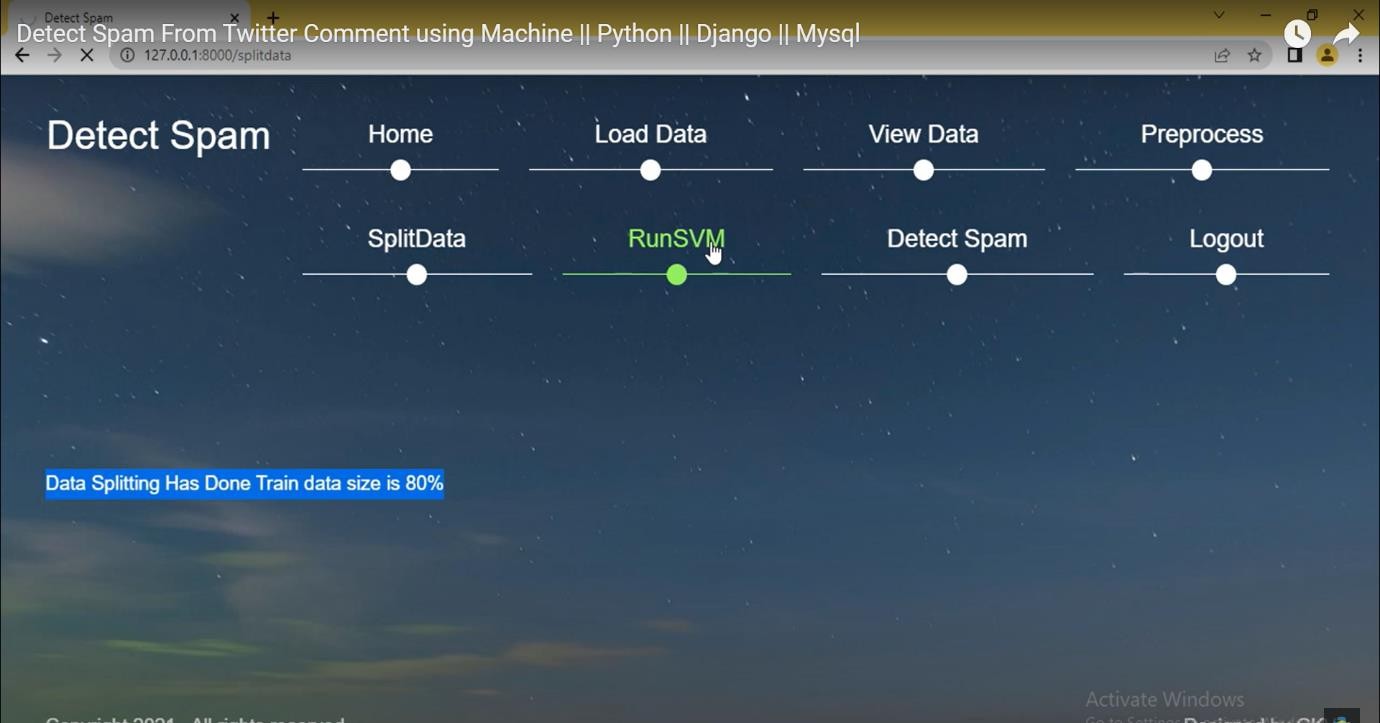
USER:

 LOGACTION:

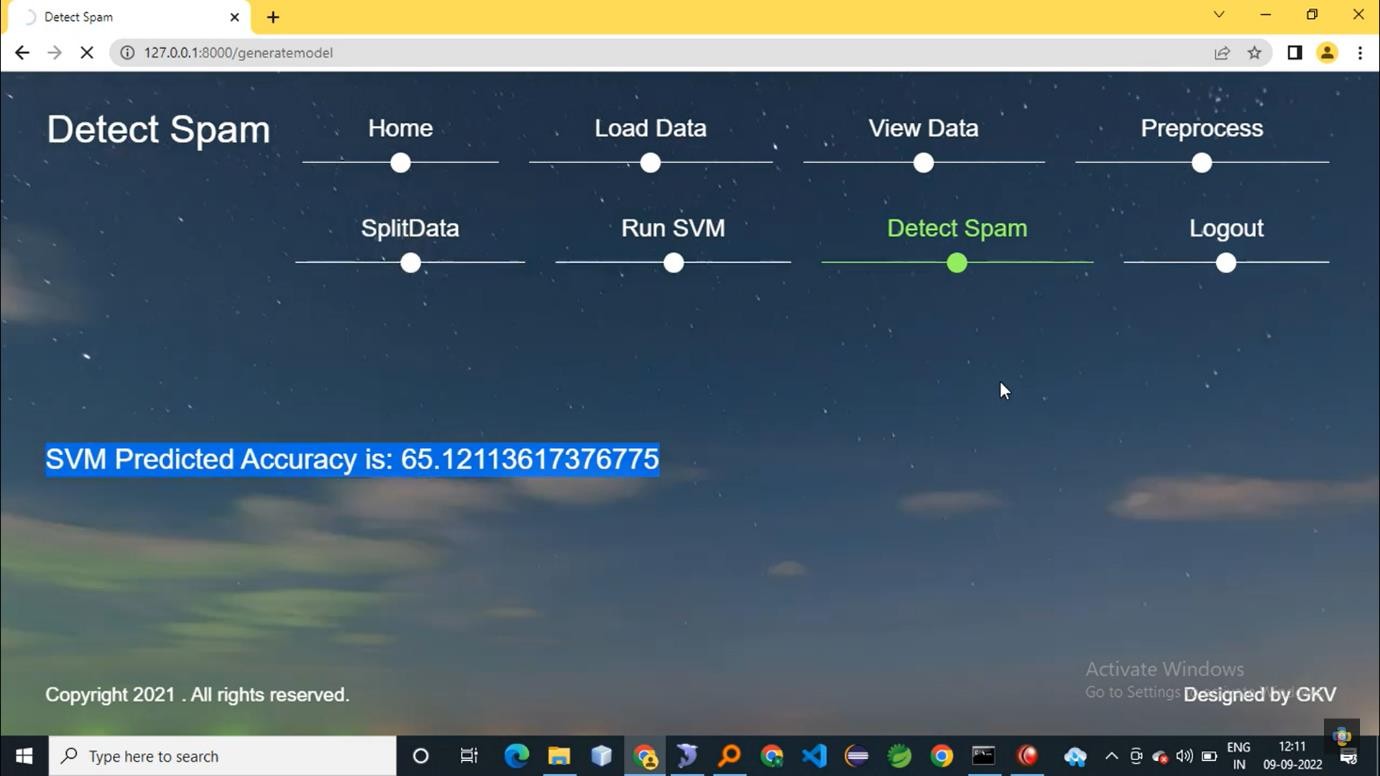
LOAD DATA:

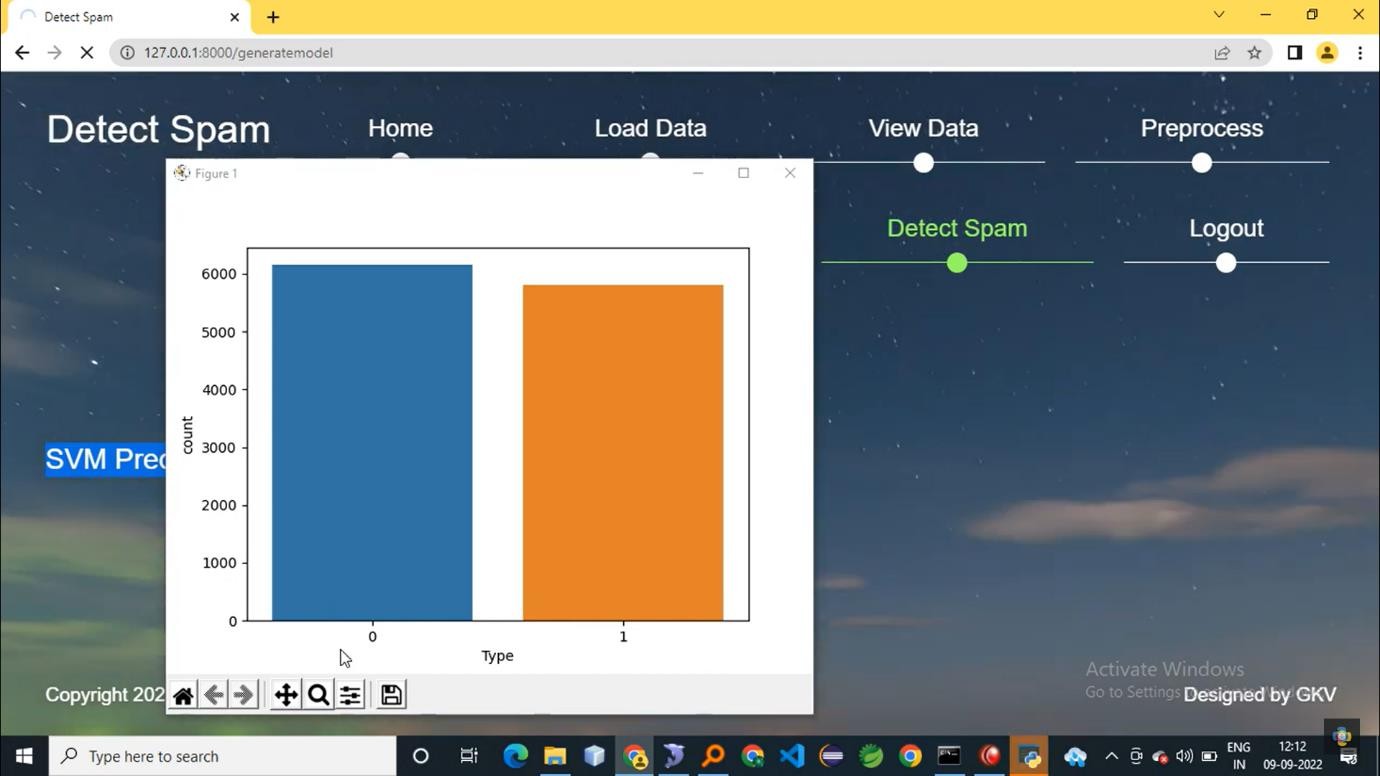
VIEW DATA:



SPLIT DATA:

GENERATE MODEL:



GRAPHICAL REPRESENTATION

**TESTING**

**6.1 SYSTEM TESTING**

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

### TYPES OF TESTS

### Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

### Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

### Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

### System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

### White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose.

It is used to test areas that cannot be reached from a black box level.

### Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

**Unit Testing**

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

### Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

**Test objectives**

* All field entries must work properly.
* Pages must be activated from the identified link.
* The entry screen, messages and responses must not be delayed.

**Features to be tested**

* Verify that the entries are of the correct format
* No duplicate entries should be allowed
* All links should take the user to the correct page.

### Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

### Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered

### 6.2 TEST CASES

| **Test Case ID** | **Test Case Description** | **Expected Outcome** |
| --- | --- | --- |
| TC-001 | Text-only spam tweet with repetitive hashtags | Detected as spam and flagged for review. |
| TC-002 | Image-only spam tweet with misleading content | Detected as spam based on image analysis. |
| TC-003 | Combination of text and image spam | Detected as spam using both text and image analysis. |
| TC-004 | Retweet of a known spam tweet | Detected as spam based on the content and source. |
| TC-005 | Legitimate tweet with a few hashtags | Not detected as spam. |
| TC-006 | Legitimate tweet with trending hashtags | Not detected as spam. |
| TC-007 | Shortened URLs leading to malicious websites | Detected as spam based on URL analysis. |
| TC-008 | Mention spam with misleading information | Detected as spam based on content analysis. |
| TC-009 | Suspicious account with repetitive posting | Detected as spam based on user behavior analysis. |
| TC-010 | Tweet with excessive use of emojis | Detected as spam due to excessive use of emojis. |
| TC-011 | Tweet with misspelled keywords | Detected as spam based on language analysis |

**SCOPE**

**Scope of the project**:

The scope of a project focused on Twitter spam detection using machine learning algorithms can be quite extensive and can vary depending on the specific objectives and resources available. Here is a broad outline of what such a project might entail.

Clearly define what constitutes "spam" on Twitter (e.g., fake accounts, promotional content, phishing links, etc.).

Determine the specific goals of the spam detection project (e.g., reducing the visibility of spam, reporting spam accounts, etc.

Gather a large dataset of tweets from Twitter. This dataset should include a mix of spam and non-spam (legitimate) tweets.

Ensure that the data is labeled or categorized as spam or non-spam.

Clean and preprocess the Twitter data. This may include text normalization, removing stop words, handling URLs, and other data cleaning tasks.

Refine the models based on evaluation results.

Experiment with different feature engineering techniques and model architectures to improve performance.

Deploy the trained model for real-time or batch processing of incoming tweets.

Develop a user-friendly interface for interacting with the spam detection system.

**CONCLUSION**

**Conclusion:**

Today is the time of social media and Twitter is the most well known social media network where anyone can post their thoughts, send their messages and promote their business. Followers have been increased on Twitter to capture attention of the spammers. In previous research, there are many algorithms of data mining that are used for spam detection on Twitter's collected datasets. In literature review, we have compared the different data mining algorithms in the category of machine learning, deep learning and hybrid algorithms. All of these algorithms researchers use for different types of spam detection. But the previous algorithms are not enough to extract and detect the spam on Twitter accurately. So, we need to expand the research for the high classification rate of spam detection. In future, we will apply previous methods on further social media stages like Instagram, SnapChat, Facebook and YouTube.

**BIBLIOGRAPHY**

* + 1. Saleh Beyt Sheikh Ahmad & Mahnaz Rafie & Seyed Mojtaba Ghorabie, "Spam detection on Twitter using a support vector machine and users’ features by identifying their interactions" 06 January 2021.
    2. Rumi Juwairiyyah, Nanditha Sriram, Jyotshna Bhushan Sharma, Ba- beetha, “Social media spam detection using deep learning” June 2020.
    3. Deepali Prakash Sonawane Dr. Baisa L. Gunjal, “New Approach for Detecting Spammers on Twitter using Machine Learning Framework” May 31st 2020.
    4. M Mostafa, A Abdelwahab, H M Sayed, “Detecting spam campaign in twitter with semantic similarity” 2020.
    5. Francesco Pierri, Carlo Piccardi, Stefano Ceri, “A multi-layer approach to disinformation detection on Twitter“2020.
    6. Ratul Chowdhury, Kumar Gourav Das ,Banani Saha, Samir Kumar Bandyopadhyay, "A Method Based on NLP for Twitter Spam detec- tion”2020.
    7. Ricardo Resende de Mendonça, Daniel Felix de Brito, and Ferrucio de Franco Rosa, Júlio César dos Reis, Rodrigo Bonacin, " A Framework for Detecting Intentions of Criminal Acts in Social Media: A Case Study on Twitter "2020.
    8. Hadi Tajalizadeh and Reza Boostani," A Novel Stream Clustering Framework for Spam Detection in Twitter” 2019.
    9. Isa Inuwa-Dutse, Mark Liptrott, Ioannis Korkontzelos, “Detection of spam-posting accounts on Twitter" 2018.
    10. Zulfikar Alom, Barbara Carminati, Elena Ferrari, “Detecting spam accounts on Twitter “2018.
    11. Rutuja Katpatal, Aparna Junnarkar, “An Efficient Approach of Spam Detection in Twitter” 2018.
    12. XIANCHAO ZHANG, ZHAOXING LI, SHAOPING ZHU, WENXIN LIANG,”Detecting Spam and Promoting Campaigns in Twitter “2016.
    13. Olubodunde Stephen Agboola, “Spam Detection Using Machine Learn- ing” 2020.
    14. Surendra Sedhai and Aixin Sun, “Semi-Supervised Spam Detection in Twitter Stream” 2017.
    15. Claudia Meda, Edoardo Ragusa, Christian Gianoglio, Rodolfo Zunino, Augusto Ottaviano, “Spam Detection of Twitter Traffic: A Framework based on Random Forests and non-uniform feature sampling” = 2016.
    16. Xianchao Zhang, Shaoping Zhu, Wenxin Liang, “Detecting Spam and Promoting Campaigns in the Twitter Social Network” 2012.
    17. Aryo Pinandito, Rizal Setya Perdana, Mochamad Chandra Saputra, Hanifah Muslimah Az-zahra, “Spam Detection Framework for Android Twitter Application Using Naïve Bayes and K-Nearest Neighbor Classifiers” 2017.
    18. Buket Ersahin, Özlem Aktas, Deniz Kılınç, Ceyhun Akyol “Twitter Fake Account Detection” 2017.
    19. Rutuja Katpatal, Aparna Junnarkar, “Spam Detection Techniques for Twitter” 2018.
    20. Abdullah Talha Kabakus, Resul Kara ““TwitterSpamDetector” A Spam Detection Framework for Twitter” 2019.