DDOS ATTACK DETECTION USING MACHINE LEARNING

In partial fulfilment for the award of the degree of

B.Sc. Digital and Cyber Forensic Science

A Project Report

Submitted by

RENAX.R.J

RCAS2020BDC115

Under the guidance of

MR.BHAARATHI.I ,M.SC. (S/W ENGG.)., MBA(ISM)., B.ED.(CS)
Assistant Professor
Department of Information Technology
Rathinam College of Arts and Science, Coimbatore – 21

DEPARTMENT OF INFORMATION TECHNOLOGY

RATHINAM COLLEGE OF ARTS AND SCIENCE
(AUTONOMOUS)

COIMBATORE – 641 021

MAY 2023

DECLARATION

This is to certify that the project work entitled "**DDOS ATTACK DETECTION USING MACHINE LEARNING**'' submitted to **Rathinam College of Arts and Science (Autonomous), Coimbatore**, in partial fulfilment of the requirements for the award of B.Sc. (**Digital and Cyber Forensic Science**) is the record of original work done by me during the period of study (2022-2023) in **RathinamCollege of Arts and Science (Autonomous), Coimbatore.**

Name of the Candidate: **RENAX . R. J**

Reg No : **RCAS2020BDC115**

Signature of the Candidate

RATHINAM COLLEGE OF ARTS AND SCIENCE
(AUTONOMOUS)

COIMBATORE – 641021

DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that Bonafide Project work done by the candidate under my supervision in Partial fulfillment of the requirements for the award of **B.Sc. (Digital and Cyber Forensic Science).**

Name of the Candidate **: RENAX. R. J**

Reg. No **: RCAS2020BDC115**

Signature of the Guide Signature of the HOD

Place:

Date:

Submitted for the Viva-Voce held on_____.

Internal Examiner External Examiner

ACKNOWLEDGEMENT

On successful completion for project look back to thank who made in possible. First and foremost, thank "**THE ALMIGHTY"** for this blessing on us without which we could havenot successfully our project.

I am extremely grateful to **Dr. Madan A. Sendhil., M.S., Ph.D**., Chairman, Rathinam Group of Institutions, Coimbatore and **Mrs.Shima Sendhil, MBA,** Director, Rathinam Group of Institutions, Coimbatore for giving me opportunity to study in this college.

I am extremely grateful to **Dr.R. Manickam, MCA., M.Phil., Ph.D.,** CEO & Secretary,Rathinam College of Arts & Science (Autonomous), Coimbatore.

Extend deepsense of valuation to **Dr.S. Balasubramanian., M.Sc., Ph.D(Swiss)., PDF (Swiss & USA)**Principal, Rathinam College of Arts & Science (Autonomous)who has permitted to undergo the project.

Unequally thank **Dr .T. Velumani M.Sc (CS)., B.Ed., M.Phil., MBA., M.Sc(Psy)., Ph.D.,Head, Department of Information Technology** for his constructive suggestions, and adviceduring the course of study.

I convey special thanks, to the supervisor **Mr.Bhaarathi.I, M.Sc.(S/W Engg.)., MBA(ISM)., B.Ed.(CS) Assistant Professor, Information Technology** who offered their inestimable support, guidance, valuable suggestion, motivations, helps given for the completion of the project. Also I extend my thanks to all the staff members of the department. I dedicated sincere respect to my parents for their moral motivation in completing the project.

RENAX.R.J

(**Reg.no:RCAS2020BDC115)**

RATHINAM COLLEGE OF ARTS AND SCIENCE (Autonomous)
(Affiliated to Bharathiar University NAAC, Re-Accredited With "A++" Grade,
Approved by AICTE and Recoganized by UGC Under Section 2(f) And 12(b)
Rathinam Techzone, Pollachi Road, Echanari P.O, Coimbatore-641021

Date:

Department of INFORMATION TECHNOLOGY

Project Completion Certificate

This is to certify that Mr. **RENAX. R. J**(20BDC115), **III B.Sc. (Digital and Cyber Forensic Science**), Rathinam College of Arts and Science has successfully completed the project entitled "**DDOS ATTACK DETECTION USING MACHINE LEARNING**" during the academic year 2022 -2023 under the supervision and guidance of **Mr.Bhaarathi M.Sc.(S/W Engg.), MBA(ISM)**, **B.Ed.(CS**) Assistant Professor, Information Technology.

Signature of the guide

Signature of the HOD

SEAL

NoviTech
the innovation partner

TO WHOMSOEVER IT MAY CONCERN

This is to Certify that **Mr. RENAX. R. J (Reg.No: RCAS2020BDC115)** Final year student of **B.Sc Digital and Cyber Forensic Science in** Department of Information Technology at **"RATHINAM COLLEGE OF ARTS AND SCIENCE, COIMBATORE"** Completed his Project on the title **"DETECTION OF DDOS ATTACK USING MACHINE LEARNING"** during the Period from **January 2023** to **April 2023**. We wish a great endeavor of his career.

For NoviTech R&D Pvt. Ltd.,

Mr.VinothKumar.A
Branch Manager

NoviTech R&D Pvt Ltd.,
3rd Floor, Sai Sruthi Complex,Ramar Koil Street, Ram Nagar,
Coimbatore - 641 009 Tamilnadu, INDIA.
0422 - 4645081
www.novitechrnd.com
support@novitechrnd.com

CONTENT

DDOS ATTACK DETECTION USING MACHINE LEARNING:

AlGORITHM:

A set of eight supervised machine learning algorithms are selected todetect DDoS attack and found the best model in terms of accuracy, precision, re cal and false alarm ratel. For experimental results, a standard benchmark dataset CIC-IDS2017 is used for training and testing purpose.

ABSTRACT

Distributed Denial of Service (DDOS) attack is one of the common network attacks. DDOS attack occurs when a website or server is targeted by a malicious user to deny the services by flooding with unwanted information. This causes delay of services to legitimate user. Denial of Service (DOS) attack happens when the attack is from single source, whereas Distributed Denial of Service attack (DDOS) happens when the attack is from many number ofsources say Botnet which controls the devices remotely for malicious purpose.

K-Fold cross validation is performed during the pre-processing stage. Then the eight models are trained and tested via K-Fold cross validation to find the best one to detect the DDOS attack
at the earliest stage. In the testing phase we tested the trained models with the parameters Accuracy, Precision, Recall and FAR. Among eight models we found that Random Forest is
the best model by considering all parameters into account. It has produced 99.885% accuracy, 99.88% Precision, 100% Recall and 0.05% False alarm rate to detect DDOS attack at theearliest.

INTRODUCTION

1.1INTRODUCTION:

Distributed denial of service (DDOS) attacks are the most critical threats to many areas of our life such as IoT, smart cities, healthcare, information technology and commercial parts. DDOS attacks continue to threaten the network security of all business sectors despite their size because of their continuous increases in complexity, volume and frequency. The authors of have classified DDOS attacks into two parts: (i) The first part is named reflection-based DDOS attacks. In this part, cyberspace gadgets are utilized to transmit attack traffic such as HTTP calls to the target, and the attacker's identity is hidden. These requests are sent throughthe source IP address targeting the IP addresses in the reflector servers (bots). Therefore, all of these concurrent demands are forwarded to the victim. Typically, these attacks are passed out to misuse the application protocols (i.e., TCP, UDP individually or integration of them). MSSQL or SSDP can be used in TCP based attacks, while Char Gen, NTP or TFTP can be used in UDP. A collection of these protocols is used with the confirmed attacks, which consists of the following protocols: DNS, LDAP, NetBIOS, SNMP, or PORTMAP. (ii) The second part is exploitation-based DDOS attacks, which similarly uses both TCP and UDP. The SYN flood attack is a TCP-based attack, while the UDP flood and UDP-Lag are UDP based attacks. provides a detailed DDOS attack taxonomy.

1.2 MOTIVATION:

Cybercriminals have used DDOS attacks to turn down the servers that are being targeted and penetrate venture networks that have the ability to overwhelm results. Many organizations face problems managing modern cyberattacks because of the increasing numbersof DDOS attacks' size and complexity.

With the latest technologies, because of resource restrictions such as limited memory and processing capacity, smart gadgets and IoT are particularly vulnerable to a wide range of DDOS attacks, so the cybercriminals are aware of these modern technologies and their weaknesses. Many organizations in 2016, such as Netflix, CNN and Twitter, were disconnected for nine hours because of an attack on their internet service providers.

This technical problem caused many issues, for example, financial losses, productivity losses, brand harm, insurance rating decreases, client and provider unstable relationships, and exceeding the IT financial plan.

1.3 ADVENTAGE OF DDOS ATTACK DETECTION:

- Naive Bayes algorithm the accuracy comparing to other Machine Learning Algorithms TheMain advantage of this algorithm is, Correct prediction.
- The predicted output should be more accurate compared with other Machine Learningalgorithms
- Less computation time.
- Simple and fast: Naive Bayes is a simple algorithm that is easy to implement and computationally efficient. It can quickly classify new data points once the model is trained.

1.3 PROPOSE OF DDOS ATTACK DETECTION:

A distributed denial-of-service (DDOS) attack is a malicious attempt to disrupt the normal traffic of a targeted server, service or network by overwhelming the target or itssurrounding infrastructure with a flood of Internet traffic.

DDOS attacks achieve effectiveness by utilizing multiple compromised computersystems as sources of attack traffic. Exploited machines can include computers and other networked resources such as IoT devices.

LITERATURE SURVEY

2.1 A Machine Learning-Based Classification and Prediction Technique for DDOS Attack:

Author: Mohmand MI, Hussain H, Khan AA, Ullah U, Zakarya M, Ahmed A, RazaM, Rahman IU, Haleem M. A

To access the research proposed a complete framework for DDOS attacks prediction.

For the proposed work, the UNWS-np-15 dataset was extracted from the GitHub repository and Python was used as a simulator. After applying the machine learning models, heregenerated a confusion matrix for identification of the model performance. In the firstclassification,

The results showed that both Precision (PR) and Router call (RE) are ∼89 % for the Random Forest algorithm. The average Accuracy (AC) of proposed model is ~89% which is superb and enough good. In the second classification, the results showed that both Precision (PR) and Recall (RE) are approximately 90% for the XG Boost algorithm. The average Accuracy (AC) of suggested model is ~90%. By comparing work to the existing research works, the accuracy of the defect determination was significantly improved which isapproximately 85% and 79%, respectively.

2.2 Information and Random Forest Feature Importance Method:

Author: Alduailij M, Khan QW, Tahir M, Sardaraz M, Alduailij M, Malik

This project presents a method for DDOS attack detection in cloud computing. The primary objective of this article is to reduce misclassification error in DDOS detection. In the proposed work, here select the most relevant features, by applying two featureselection techniques.

i.e., the Mutual Information (MI) and Random Forest Feature Importance (RFFI) methods. Random Forest (RF), Gradient Boosting (GB), Weighted Voting Ensemble (WVE),K Nearest Neighbour (KNN), and Logistic Regression (LR) are applied to selected features. The experimental results show that the accuracy of RF, GB, WVE, and KNN with 19 featuresis 0.99.

To further study these methods, misclassifications of the methods are analyzed, which lead to more accurate measurements. Extensive experiments conclude that the RF performed well in DDOS attack detection and misclassified only one attack as normal. Comparative results are presented to validate the proposed method.

2.3 An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks:

Author: Rezvy S, Luo Y, Petridis M, Lasebae A, Zebin T

For the purpose of protecting 5G and IoT networks from intrusion, pre have implemented an auto-encoded dense neural network algorithm. Here, use the standard-setting Aegean Wi-Fi Intrusion dataset to test the efficacy of the algorithm.

findings demonstrated superior performance, with a combined detection accuracy of 99.9 percent against Flooding, Impersonation, and Injection attacks. Additionally, method demonstrated how the proposed algorithm outperforms state-of-the-art methods in the literature, both in terms of detection accuracy and speed.

2.4 Deep learning approaches for detecting DDOS attacks:

Authors: Mittal M, Kumar K, Behal S

I looked through four popular digital libraries (IEEE, ACM, Science Direct, and Springer) and one academic search engine (Google Scholar) to find recent articles. Having reviewed the applicable literature, model have broken down the SLR's findings into five broadareas of inquiry:

(i) the various forms of deep learning techniques for detecting DDOS attacks;

(ii) the methodologies, strengths, and weaknesses of currently available deep learning techniques for detecting DDOS attacks.

(iii) datasets and attack types from the existing literature that have been used as benchmarks; and

(iv) pre-processing methods, hyper-parameter values, experimental setups, and performance metrics from the current literature.

(v) areas where more research is neededand potential new avenues of study.

2.5 DDoS attack detection method based on machine learning:

Authors: Pei J, Chen Y, Ji W

proposes a machine learning-based approach to detecting DDOS attacks, with the twostages of detection being feature extraction and model detection. In the feature extraction phase, thedata packages are compared based on their classification, revealing the characteristics of DDOS attack traffic that make up the majority of it.

After features have been extracted, they are fed into a machine learning model as inputs, and the random forest algorithm is used to hone the attack detection model. The results of the experiments demonstrate that the proposed DDOS attack detection method based on machinelearning has a high detection rate for the most common DDOS attacks

REQUIREMENT ANALYSIS:

3.1 PLATEFORM REQUIREMENT:

PYTHON:

Python is an open source programming language. Python was made to be easy-to-read and powerful. A Dutch programmer named Guido van Rossum made Python in 1991. He named it after the television show Monty Python's Flying Circus. Many Python examples and tutorials include jokes from the show.

Python is an interpreted language. Interpreted languages do not need to be compiled to run. A program called an interpreter runs Python code on almost any kind of computer. This means that a programmer can change the code and quickly see the results. This also means Python is slower than a compiled language like C, because it is not running machine code directly.

Python is a good programming language for beginners. It is a high-level language, which means a programmer can focus on what to do instead of how to do it. Writing programs
in Python takes less time than in some other languages.

Python drew inspiration from other programming languages like C, C++, Java, Perl,and Lisp.

Python has a very easy-to-read syntax. Some of Python's syntax comes from C, because that is the language that Python was written in. But Python uses whitespace to delimit code: spaces or tabs are used to organize code into groups. This is different from C. In C, there is a semicolon at the end of each line and curly braces ({}) are used to group code. Using whitespaceto delimit code makes Python a very easy-to-read language.

Python is used by hundreds of thousands of programmers and is used in many places. Sometimes only Python code is used for a program, but most of the time it is used to do simple jobs while another programming language is used to do more complicated tasks.

Its standard library is made up of many functions that come with Python when it isinstalled.

On the Internet there are many other libraries available that make it possible for the Python

language to do more things. These libraries make it a powerful language; it can do many different things.

Some things that Python is often used for are:

Web development Game programming
Desktop GUIs Scientific programmingNetwork programming.

Version 1

Python reached version 1.0 in January 1994. The major new features included in thisrelease were the functional programming tools lambda, map, filter and reduce. Van Rossum stated that "Python acquired Lambda, reduce (), filter () and map (), courtesy of a Lisp hacker who missed them and submitted working patches"

The last version released while Van Rossum was at CWI was Python 1.2. In 1995, Van Rossum continued his work on Python at the Corporation for National Research Initiatives(CNRI) in Reston, Virginia whence he released several versions.

By version 1.4, Python had acquired several new features. Notable among these are the Modula-3 inspired keyword arguments (which are also similar to Common Lisp's keyword arguments) and built-in support for complex numbers. Also included is a basic form of data hiding by name mangling, though this is easily bypassed.

During Van Rossum's stay at CNRI, he launched the Computer Programming for Everybody (CP4E) initiative, intending to make programming more accessible to more people, with a basic "literacy" in programming languages, similar to the basic English literacy and mathematics skills required by most employers. Python served a central role in this: because of its focus on clean syntax, it was already suitable, and CP4E's goals bore.

Similarities to its predecessor, ABC. The project was funded by DARPA. As of 2007, theCP4E project is inactive, and while Python attempts to be easily learnable and not too arcane in its syntax and semantics, reaching out to non-programmers is not an active concern.

In 2000, the Python core development team moved to BeOpen.com to form the Be Open Python Labs team. CNRI requested that a version 1.6 be released, summarizing Python's development up to the point at which the development team left CNRI. Consequently, therelease schedules for 1.6 and 2.0 had a significant amount of overlap. Python 2.0 was the only release from BeOpen.com. After Python 2.0 was released by BeOpen.com, Guido van Rossum and the other Python Labs developers joined Digital Creations.

The Python 1.6 release included a new CNRI license that was substantially longer than the CWI license that had been used for earlier releases. The new license included a clause stating that the license was governed by the laws of the State of Virginia. The Free Software Foundation argued that the choice-of-law clause was incompatible with the GNU General Public License. Be Open, CNRI and the FSF negotiated a change to Python's free software license that would make it GPL-compatible. Python 1.6.1 is essentially the same as Python 1.6, with a few minor bug fixes, and with the new GPL-compatible license.

Version 2

Python 2.0 introduced list comprehensions, a feature borrowed from the functional programming languages SETL and Haskell. Python's syntax for this construct is very similar

to Haskell's, apart from Haskell's preference for punctuation characters and Python's preference for alphabetic keywords. Python 2.0 also introduced a garbage collection system capable of collecting reference cycles.

Python 2.1 was close to Python 1.6.1, as well as Python 2.0. Its license was renamed Python Software Foundation License. All code, documentation and specifications added, from the time of Python 2.1's alpha release on, is owned by the Python Software Foundation (PSF), a non- profit organization formed in 2001.model after the Apache Software Foundation.

The release included a change to the language specification to support nested scopes, like other statically scoped languages. (The feature was turned off by default, and not required, untilPython 2.2.)

A major innovation in Python 2.2 was the unification of Python's types (types written in C)and classes (types written in Python) into one hierarchy. This single unification made Python's

object model purely and consistently object oriented. Also added were generators which wereinspired by Icon.

Python 2.5 was released on September 2006 and introduced the with statement, which encloses a code block within a context manager (for example, acquiring a lock before the blockof code is run and releasing the lock afterwards, or opening a file and then closing it), allowing Resource Acquisition Is Initialization (RAII)-like behaviour and replacing a common try/finally idiom.

Python 2.6 was released to coincide with Python 3.0, and included some features from that release, as well as a "warnings" mode that highlighted the use of features that were removedin Python 3.0. Similarly, Python 2.7 coincided with and included features from Python 3.1, which was released on June 26, 2009.

parallel 2.x and 3.x releases then ceased, and Python 2.7 was the last release in the 2.xseries. In November 2014, it was announced that Python 2.7 would be supported until 2020, but users were encouraged to move to Python 3 as soon as possible.

Version 3

python 3.0 (also called "Python 3000" or "Py3K") was released on December 3, 2008 It was designed to rectify fundamental design flaws in the language—the changes required could not be implemented while retaining full backwards compatibility with the 2.x series,which necessitated a new major version number. The guiding principle Python 3 was:"reduce feature duplication by removing old ways of doing things".

Python 3.0 was developed with the same philosophy as in prior versions. However, as Python had accumulated new and redundant ways to program the same task, Python 3.0 hadan emphasis on removing duplicative constructs and modules, in keeping with "There should be one— and preferably only one —obvious way to do it".

Nonetheless, Python 3.0 remained a multi-paradigm language. Coders still had options among object-orientation, structured programming, functional programming and otherparadigms, but within such broad choices, the details were intended to be more obvious in Python 3.0 than they were in Python 2.x.

Python 2.7.0

Note: A bug fix release, 2.7.13, is currently available. Its use is recommended.Python 2.7.0was released on July 3rd, 2010.
Python 2.7 is scheduled to be the last major version in the 2.x series before it moves into an extended maintenance period. This release contains many of the features that were first releasedin Python 3.1. Improvements in this release include:

- ✓ An ordered dictionary type
- ✓ New unit test features including test skipping, new assert methods, and test discovery
- ✓ A much faster io module
- ✓ Automatic numbering of fields in the str.format() method
- ✓ Float re improvements backported from 3.x
- ✓ Tile support for T kinter
- ✓ A backport of the memory view object from 3.x
- ✓ Set literals
- ✓ Set and dictionary comprehensions
- ✓ Dictionary views
- ✓ New syntax for nested with statements

3.1.1SUPPORTING OPERATING SYSTEM:

WINDOWS 10:

Windows 10 introduced the **Universal Windows Platform** (UWP), which provides a common app platform on every device that runs Windows. The UWP core APIs are the same on all Windows devices. If your app only uses the core APIs, it will run on any Windows device no matter whether you are targeting a desktop PC, Xbox, Mixed-reality headset, and soon.

3.2 SOFTWARE REQUIREMENT:

PYTHON:

The Platform module is used to retrieve as much possible information about the platform on which the program is being currently executed. Now by platform info, it means informationabout the device, it's OS, node, OS version, Python version, etc.

This module plays a crucial role when you want to check whether your program is compatible with the python version installed on a particular system or whether the hardware specificationsmeet the requirements of your program.

This module already exists in the python library and does not require any installation using pip.

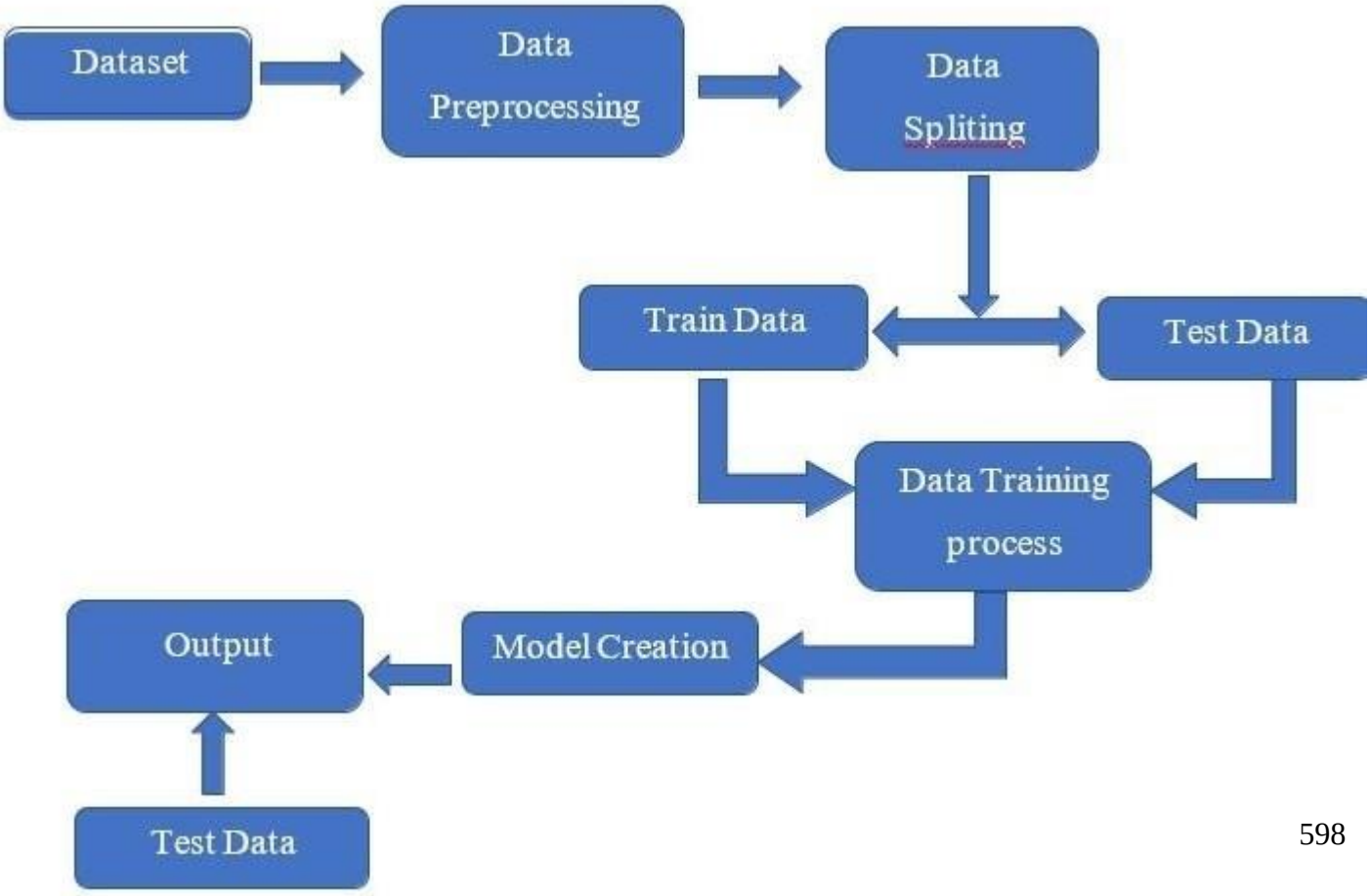
3.3 HARDWARE REQUIEMENT:

- Windows -10
- Ram -4GB or 8GB
- Processor - i3 or i5
- Python- 3.9

APPENDIX

4.1 PROJECT DESING:

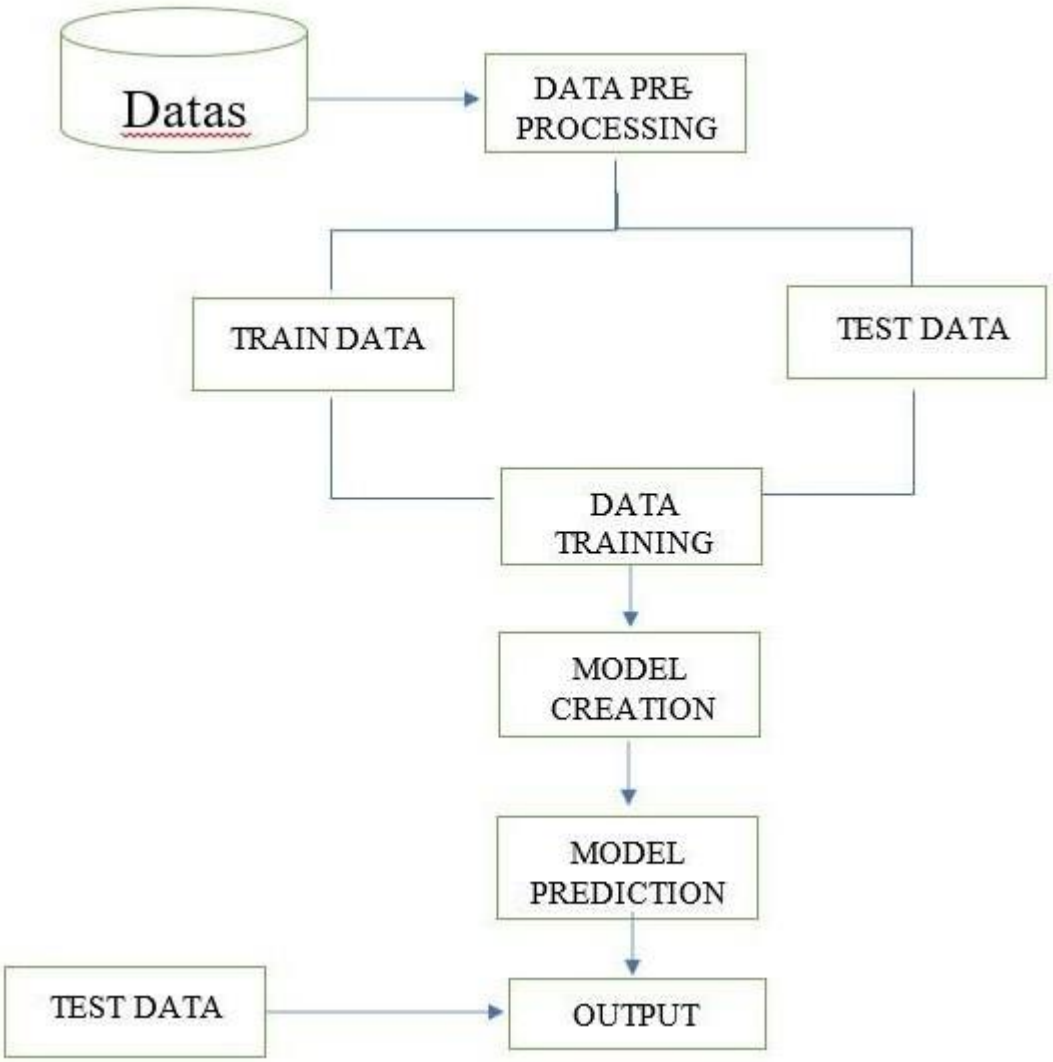
Block Diagram:



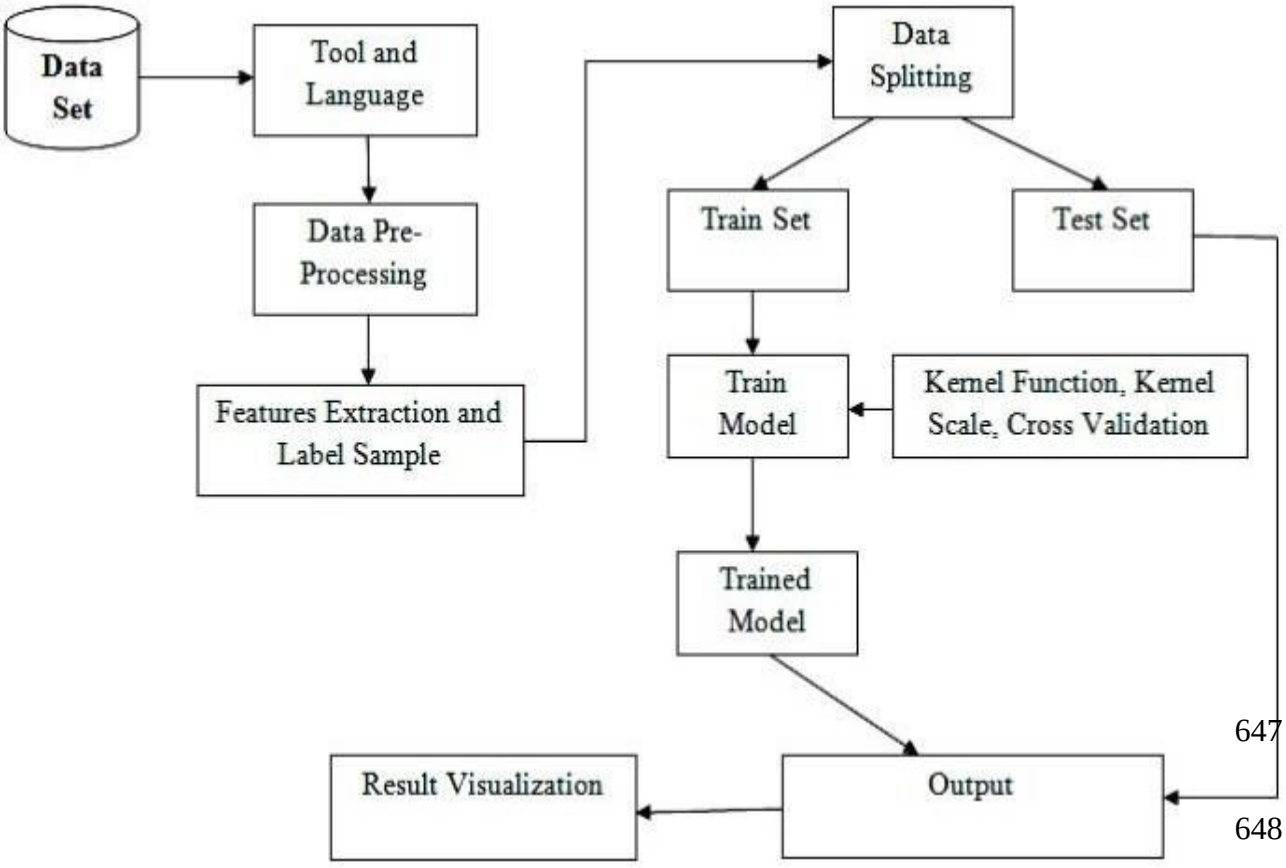
598

4.2 DATA SET DIAGRAM:

Block Diagram:



4.3 ER DIAGRAM:



647

648

SOFTWARE DESCRIPTION

4.4 SOFTWARE DESCRIPTION:

MODULES:

1. NUMPY
2. PANDAS
3. SKLEARN
4. FLASK

NUMPY:

NumPy is a Python package. It stands for ' Numerical Python'It is a library consisting of multidimensional array objects and a collection of routines for processing ofarray.

eric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray wasalso developed, having some additional functionalities. In 2005, Travis Oliphant createdNumPy package by incorporating the features of Numarray into Numeric package. There are many contributors to this open source project.

Operations using NumPy

- Using NumPy, a developer can perform the following operations
- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.

Operations related to linear algebra. NumPy has in-built functions for linear algebra and randomnumber generation.

NumPy – A Replacement for MatLab

NumPy is often used along with packages like SciPy (Scientific Python) andMat−plotlib (plotting library).

This combination is widely used as a replacement for MatLab, a popular platform for technical computing. However, Python alternative to MatLab is now seen as a more modern andcomplete programming language.

It is open source, which is an added advantage of NumPy.INSTALLATION:

pip install "numpy”

PANDAS:

as is an open-source, BSD-licensed Python library providing high- performance, easy-to-use

data structures and data analysis tools for the Python programming language. Python withPandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features

of Python Pandas and how to use them in practice.

Pandas deals with the following three data structures − Series

DataFrame:

- Panel

These data structures are built on top of Numpy array, which means they are fast.

INSTALLATION

Pip install pandas

SKLEARN:

Scikit-learn is a machine learning library for Python. It features several regression, classification and clustering algorithms including SVMs, gradient boosting, k- means, randomforests and DBSCAN.

It is designed to work withPython Numpy and SciPy .

The scikit-learn project kicked off as a Google Summer of Code (also known as GSoC) project by David Cournapeau as scikits.learn. It gets its name from “Scikit”, a separate third-party extension to SciPy.

ython Scikit-learn

Scikit is written in Python (most of it) and some of its core algorithms are written in Cythonfor even better performance.
Scikit-learn is used to build models and it is not recommended to use it for reading, manipulating and summarizing data as there are better frameworks available for the purpose.It is open source and released under BSD license.

Install Scikit Learn:

Scikit assumes you have a running Python 2.7 or above platform with NumPY

(1.8.2 and above) and SciPY (0.13.3 and above) packages on your device. Once we have thesepackages installed we can proceed with the installation.

pip install scikit-learn

Keras:
keras is a high-level neural networks API, capable of running on top of Tensorflow, Theano, and CNTK.

It enables fast experimentation through a high level, user-friendly, modular andextensible API.

Keras can also be run on both CPU and GPU.Keras was developed and is maintained by Francois Chollet and is part of the Tensorflow core, which makes it Tensorflows preferred high-level API.

Keras including the two most used Keras models (Sequential and Functional), the corelayers as well as some preprocessing functionalities.

Installation: pipinstall keras **Flask:**

What is Flask?

Flask is an API of Python that allows us to build up web-applications.

It was developed by Armin Ronacher. Flask's framework is more explicit than Django'sframework and is also easier to learn because it has less base code to implement a simple web- Application. A Web-Application Framework or Web Framework is the collection of modules and libraries that helps the developer to write applications without writing the low-level codes such as protocols, thread management, etc. Flask is based on WSGI (Web Server
Gateway Interface) toolkit and Jinja2 template engine.

Routing:

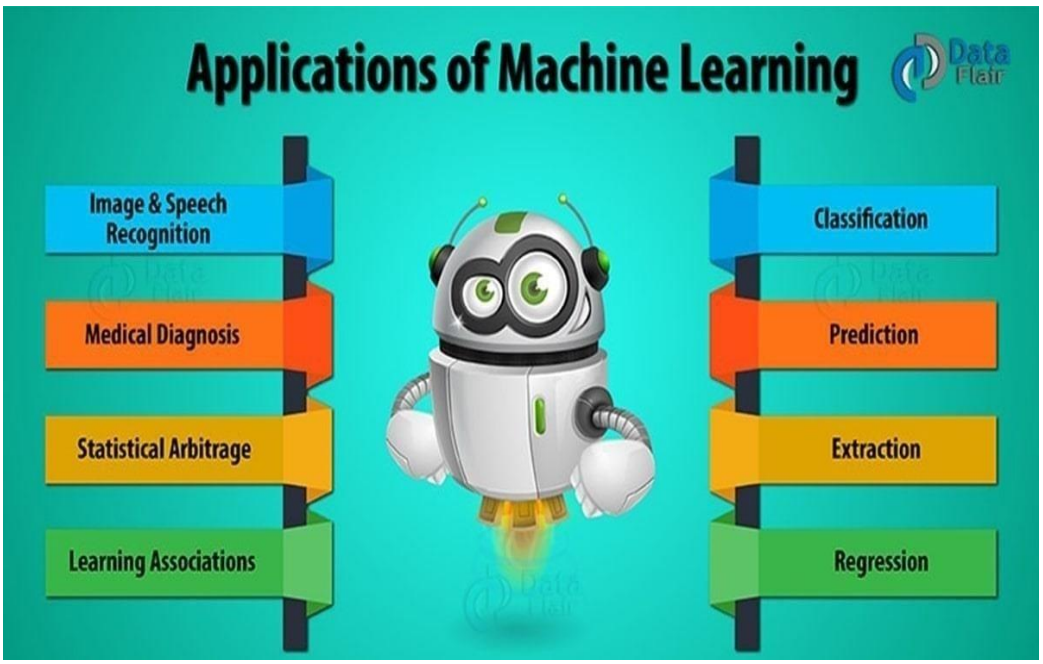
Nowadays, the web frameworks provide routing technique so that user can rememberthe URLs.

It is useful to access the web page directly without navigating from the Home page. It isdone through the following route () decorator, to bind the URL to a functionBuilding URL in FLask:

Dynamic Building of the URL for a specific function is done using url_for () function.The function accepts the name of the function as first argument, and one or more keyword arguments. See this example.

IMPLEMENTATION DETAILS

5.1 IMPLEMENTATION OF DDOS ATTACT DETECTION



- Machine Learning Applications in Healthcare
- Machine Learning Applications in Finance
- Machine Learning Applications in Retail
- Machine Learning Applications in Travel
- Machine Learning Applications in Media

MACHINE LEARNING IN HEALTHCARE:

Doctors and medical practitioners will soon be able to predict with accuracy on how long patients with fatal diseases will live. Medical systems will learn from data and help patients save money by skipping unnecessary tests. Radiologists will be replaced by machine learning algorithms.

McKinsey Global Institute estimates that applying machine learning techniques to better inform decision making could generate up to $100 billion in value based on optimized innovation, enhanced efficiency of clinical trials and the creation of various novel tools for physicians, insurers and consumers. Computers and Robots cannot replace doctors or nurses, however the use of life-saving technology (machine learning) can definitely transform healthcare domain. When we talk about efficiency of machine learning, more data produces effective results – and the healthcare industry is residing on a data goldmine.

i) Drug Discovery/Manufacturing:

Manufacturing or discovering a new drug is expensive and lengthy process as thousands of compounds need to be subjected to a series of tests, and only a single one might result in a usable drug. Machine learning can speed up one or more of these steps in this lengthymultistep process.

Machine Learning Examples in Healthcare for Drug Discovery:

Pfizer is using IBM Watson on its immuno-oncology (a technique that uses body's immune system to help fight cancer) research. This is one of the most significant uses of IBMWatson for drug discovery. Pfizer has been using machine learning for years to sieve throughthe data to facilitate research in the areas of drug discovery (particularly the combination of multiple drugs) and determine the best participant for a clinical trial.

ii)Personalized Treatment/Medication

Imagine when you walk in to visit your doctor with some kind of an ache in your stomach. After snooping into your symptoms, the doctor inputs them into the computer that extracts the latest research that the doctor might need to know about how to treat your ache.

MRI and a computer helps the radiologist detect problems that possibly could be too small for the human eye to see. In the end, a computer scans all your health records and family medical history and compares it to the latest research to advice a treatment protocol thatis particularly tailored to your problem. Machine learning is all set to make a mark inpersonalized care.

Personalized treatment has great potential for growth in future, and machine learning could play a vital role in finding what kind of genetic makers and genes respond to a particular treatment or medication. Personalized medication or treatment based on individual health records paired with analytics is a hot research area as it provides better disease assessment. In future, increased usage of sensor integrated devices and mobile apps withsophisticated remote monitoring and health-measurement capabilities, there would be another data deluge that could be used for treatment efficacy. Personalized treatment facilitates health optimization and also reduces overall healthcare costs.

Machine Learning Examples in Healthcare for Personalized Treatment:

A major problem that drug manufacturers often have is that a potential drug sometimes work only on a small group in clinical trial or it could be considered unsafe becausea small percentage of people developed serious side effects. Genentech, a member of the Roche Group collaborated with GNS Healthcare to innovate solutions and treatments using biomedical data. Genentech will make use of GNS Reverse Engineering and Forward Simulation to look for patient response markers based on genes which could lead to providing targeted therapies for patients.

MACHINE LEARNING APPLICATIONS IN FINANCE:

More than 90% of the top 50 financial institutions around the world are using machinelearning and advanced analytics. The application of machine learning in Finance domain helps banks offer personalized services to customers at lower cost, better compliance and generate greater revenue.

Machine Learning Examples in Finance for Fraud Detection one of the core machine learning use cases in banking/finance domain is to combat fraud. Machine learning is best suited for this use case as it can scan through huge amounts of transactional data and identify if there is any unusual behaviour. Every transaction a customer makes is analyzed in real-time and given a fraud-score that represents the likelihood of the transaction being fraudulent. If the fraud score is above a particular threshold, a rejection will be triggeredautomatically which would otherwise be difficult without the application of machine learning techniques as humans cannot reviews 1000's of data points in seconds and make a decision.

- Citibank has collaborated with Portugal based fraud detection company Feed zai that works in real-time to identify and eliminate fraud in online and in-person banking by alerting thecustomer.

- PayPal is using machine learning to fight money laundering. PayPal has several machine learning tools that compare billions of transactions and can accurately differentiate between what is a legitimate and fraudulent transaction amongst the buyers and sellers.

MACHINE LEARNING APPLICATIONS IN RETAIL:

- Machine learning in retail is more than just a latest trend, retailers are implementing big datatechnologies like Hadoop and Spark to build big data solutions and quickly realizing the fact that it's only the start.

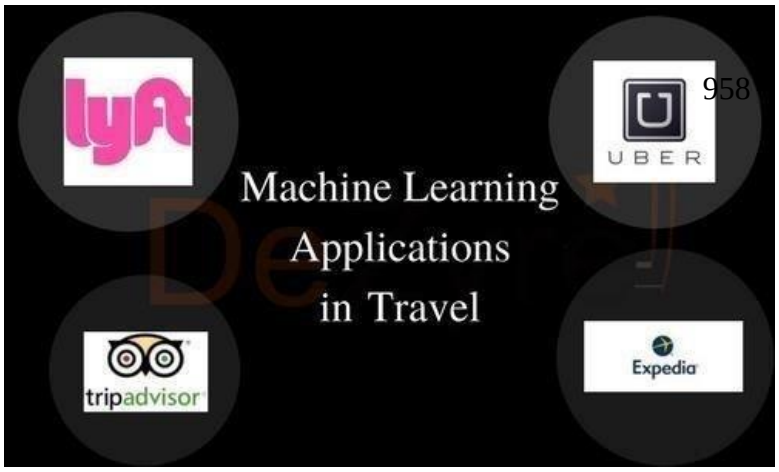
They need a solution which can analyse the data in real-time and provide valuable insights that can translate into tangible outcomes like repeat purchasing. Machine learning algorithms process this data intelligently and automate the analysis to make this supercilious goal possible for retail giants like Amazon, Target, Alibaba and Walmart.

The moment you start browsing for items on Amazon, you see recommendations for products you are interested in as “Customers Who Bought this Product Also Bought” and“Customers who viewed this product also viewed”, as well specific tailored product recommendation on the home page, and through email. Amazon uses Artificial Neural Networks machine learning algorithm to generate these recommendations for you.

To make smart personalized recommendations, Alibaba has developed “E-commerce Brain” that makes use of real-time online data to build machine learning models for predicting what customers want and recommending the relevant products based on their recent order history, bookmarking, commenting, browsing history, and other actions.

MACHINE LEARNING APPLICATIONS IN TRAVEL:

One of Uber’s biggest uses of machine learning comes in the form of surge pricing,
a machine learning model nicknamed as “Geosurge” at Uber. If you are getting late for ameeting and you need to book an Uber in crowded area, get ready to pay twice the normal fare.
In 2011, during New Year’s Eve in New York, Uber charged $37 to $135 for onemile journey.



Uber leverages predictive modelling in real-time based on traffic patterns, supply and demand. Uber has acquired a patent on surge pricing. However, customer backlash on surgepricing is strong, so Uber is using machine learning to predict where demand will be highso that drivers can prepare in advance to meet the demand, and surge pricing can be reduced toa greater extent.

MACHINE LEARNING APPLICATIONS IN SOCIAL MEDIA:

Machine learning offers the most efficient means of engaging billions of social media users. From personalizing news feed to rendering targeted ads, machine learning is the heart ofall social media platforms for their own and user benefits.

Social media and chat applications have advanced to a great extent that users do notpick up the phone or use email to communicate with brands – they leave a comment on Facebook or Instagram expecting a speedy reply than the traditional channels.

Earlier Facebook used to prompt users to tag your friends but nowadays the social networks artificial neural networks machine learning algorithm identifies familiar faces fromcontact list. The ANN algorithm mimics the structure of human brain to power facialrecognition. The professional network LinkedIn knows where you should apply for your nextjob, whom you should connect with and how your skills stack up against your peers as you search for new job.

PRINCIPAL COMPONENT ANALYSIS:

The main idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent. The same is done by transforming the variables to a new set of variables.

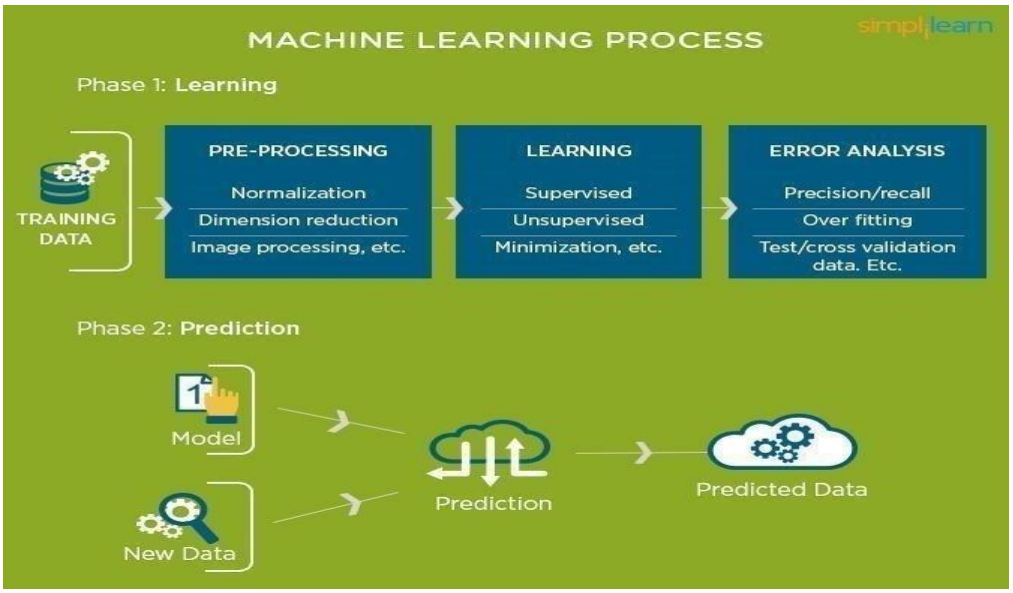
which are known as the principal components (or simply, the PCs) and are orthogonal, ordered such that the retention of variation present in the original variables decreases as we move down in the order. So, in this way, the 1st principal component retains maximum variation that was present in the original components. The principal components arethe eigenvectors of a covariance matrix, and hence they are orthogonal.

5.2 DETECTION OF DDOS ATTACK ON WEB APPLICATION:

To better understand the uses of machine learning, consider some of the instances where machine learning is applied: the self-driving Google car, cyber fraud detection, online recommendation engines—like friend suggestions on Facebook, Netflix showcasing the movies and shows you might like, and "more items to consider" and "get yourself a little something" on Amazon—are all examples of applied machine learning.

All these examples echo the vital role machine learning has begun to take in today's data-rich world. Machines can aid in filtering useful pieces of information that help in major advancements, and we are already seeing how this technology is being implemented in a widevariety of industries.

The process flow depicted here represents how machine learning works.



With the constant evolution of the field, there has been a subsequent rise in the uses, demands, and importance of machine learning. Big data has become quite a buzzword in thelast few years;

that's in part due to increased sophistication of machine learning, which helps analyze those big chunks of big data. Machine learning has also changed the way data extraction, and interpretation is done by involving automatic sets of generic methods that have replacedtraditional statistical techniques.

Uses of Machine Learning:

Earlier in this article, we mentioned some applications of machine learning. Tounderstand the concept of machine learning better, let's consider some more examples: web search results, real-time ads on web pages and mobile devices, email spam filtering, networkintrusion detection, and pattern and image recognition. All these are by-products of applying machine learning to analyze huge volumes of data.

Additionally,data analysis was always being characterized by trial and error, an

approach that becomes impossible when data sets are large and heterogeneous. Machinelearning comes as the solution to all this chaos by proposing clever alternatives to analyzing huge volumes of data. By developing fast and efficient algorithms and data-driven models for real-time processing of data, machine learning is able to produce accurate results and analysis.

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model of a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images

with and without that object (the input), and each image would have a label (the output)designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback. Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample inputs are missing the desired output.

- Classification algorithms and regression algorithms are types of supervised learning.
- Classification algorithms are used when the outputs are restricted to a limited set ofvalues.

For a classification algorithm that filters emails, the input would be an incoming email, and theoutput would be the name of the folder in which to file the email.

For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

Supervised Machine Learning:

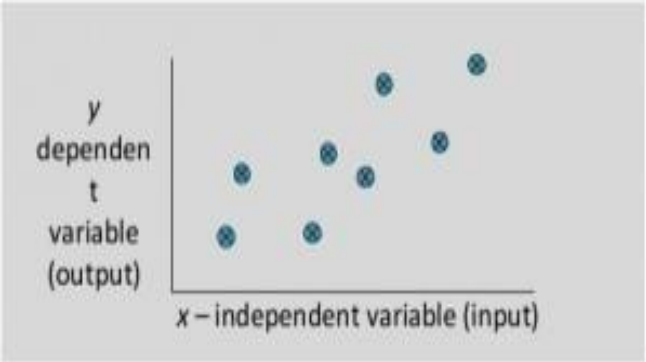
The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use analgorithm to learn the mapping function from the input to the output $Y = f(X)$. The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

Techniques of Supervised:

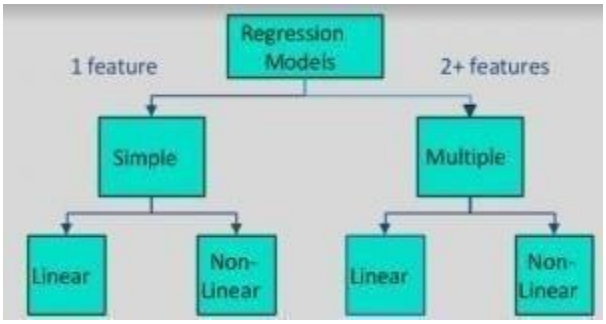
Machine Learning algorithms include linear and logistic regression, multi-class classification, Decision Trees and support vector machines. Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. For example, a classification algorithm will learn to identify animals after being trained on a dataset of images that are properly labeled with the species of the animal and some identifying characteristics.

supervised learning problems can be further grouped into Regression and Classification problems. Both problems have as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification.

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used, the simplest is the linearregression. It tries to fit data with the best hyper-plane which goes through the points.



Types of Regression Models:

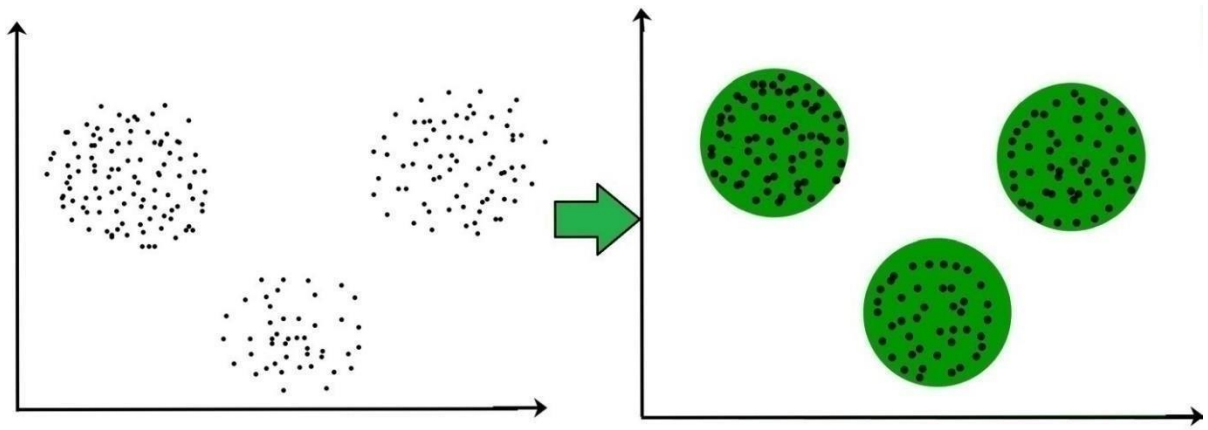


unsupervised learning, the algorithm builds a mathematical model of a set of data which contains only inputs and no desired outputs. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data.

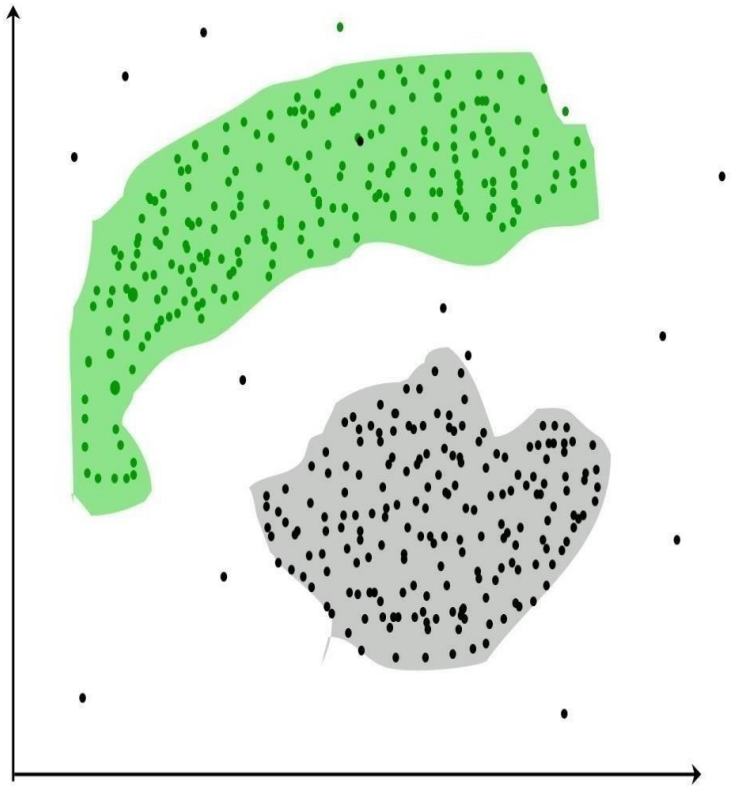
An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

For ex– The data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the belowpicture.



It is not necessary for clusters to be a spherical. Such as:

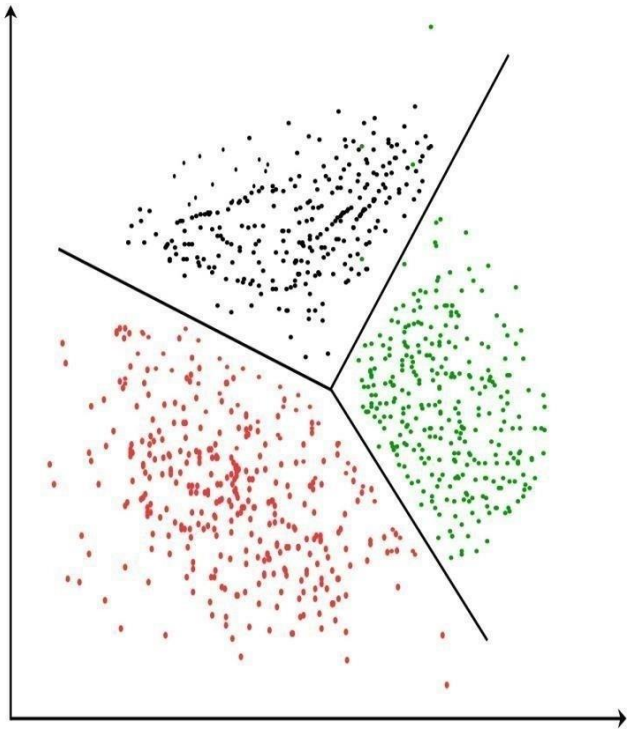


These data points are clustered by using the basic concept that the data point lies within the given constraint from the cluster center. Various distance methods and techniques are usedfor calculation of the outliers.

Clustering is very much important as it determines the intrinsic grouping among the un labeled data present. There are no criteria for a good clustering. It depends on the user, what is the criteria they may use which satisfy their need. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), in finding "natural clusters" and describe their unknown properties ("natural" data types), in finding useful and suitable groupings ("useful" data classes) or in finding unusual data objects (outlier detection). This algorithm must make some assumptions which constitute the similarity of points and eachassumption make different and equally valid clusters.

Clustering Algorithms:

K-means clustering algorithm – It is the simplest unsupervised learning algorithmthat solves clustering problem.K-means algorithm partition n observations into k clusters whereeach observation belongs to the cluster with the nearest mean serving as a prototype of the cluster



Applications of Clustering in different fields:

- Marketing: It can be used to characterize & discover customer segments for marketing purposes.
- Biology: It can be used for classification among different species of plants and animals.
- Libraries: It is used in clustering different books on the basis of topics and information. Insurance: It is used to acknowledge the customers, their policies and identifying the frauds.
- City Planning: It is used to make groups of houses and to study their values based ontheir geographical locations and other factors present.
- Earthquake studies: By learning the earthquake affected areas we can determine the dangerous zones.

Active learning algorithms access the desired outputs (training labels) for a limited set of inputs based on a budget, and optimize the choice of inputs for which it will acquire training labels.

When used interactively, these can be presented to a human user for labeling. Reinforcement learning algorithms are given feedback in the form of positive or negative reinforcement in a dynamic environment, and are used in autonomous vehicles or in learning to play a game against a human opponent. Other specialized algorithms in machine learning include topic modeling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Machine learning seems to be the most straightforward case of all. It is for the most part associated with terms referring to different scientific methods for knowledge discovery or prediction (labelled as machine or statistical learning methods). Towards Data Science provides a platform for thousands of people to exchange ideas and to expand our understanding of data science. Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured similar to data mining.

5.3 MODULE DESCRIPTION:

- Data collection
- Data Pre-Processing
- Training data and Test data
- Model Creation
- Model Prediction

Data collection

The data set obtained from Kaggle Website, is used in this paper for the experimental verification. This data set comprises some attributes and so many tuples. The categorical valuesare converted to numeric values in order to make the classification algorithm more efficient.

For example, categorical attribute ‘salary’ contains three values such as low, medium and high. Hence it is converted to 0, 1 and 2 respectively. The misspelled attributes are also corrected.

Data Pre-Processing

Pre-processing refers to the transformations applied to our data before providing the data to the algorithm. Data Pre-processing technique is used to convert the raw data into anunderstandable data set. In other words, whenever the information is gathered from various

Training data and Test data

- For choosing a model we split our dataset into train and test
- edata’s are split into 3:1 ratio that means
- Training data having 70 percent and testing data having 30 percent
- In this split process preforming based on train_test_split model • After splitting we get xtrainxtest and ytrain ytest

Model Creation

- Contextualise machine learning in your organisation.
- Explore the data and choose the type of algorithm.
- Prepare and clean the dataset.
- Split the prepared dataset and perform cross validation.
- Perform machine learning optimisation.
- Deploy the model.

Model Prediction

By using Machine learning learning algorithm decision tree will predict whether the employee will quit the organization or not. The predicted value is compared with the actual value in the database.

Decision Tree

A decision tree is a popular machine learning algorithm that is commonly used for classification and regression problems. It is a type of supervised learning algorithm that is easy to understand and interpret, making it a popular choice for both novice and experienced data scientists.

A decision tree algorithm builds a model by recursively splitting the training data intosmaller and smaller subsets, using a set of decision rules based on the features of the data. The algorithm starts at the root node of the tree and uses a set of conditions to split the data into two or more child nodes. This process is repeated for each child node until a stopping criterion is met, such as when all the instances in a node belong to the same class or when a certain depth is reached.

The decision tree algorithm creates a tree-like model that can be used to make predictions on new data. When a new instance is presented to the model, the algorithm traverses the tree from the root to a leaf node, using the decision rules at each node to determine which child node to move to. Once the algorithm reaches a leaf node, it outputs the predicted class or value for the instance.

There are many variations of the decision tree algorithm, including ID3, C4.5, CART, and Random Forests. These variations differ in the way they handle missing data, how they select the best feature to split on, and how they handle over fitting.

🔔 Random Forest

Random Forest is a popular ensemble learning algorithm in machine learning, which isa combination of multiple decision trees to improve the predictive accuracy and reduce overfitting.

A random forest algorithm randomly selects a subset of features and a subset of data from the original dataset, then builds a decision tree on each subset. It combines the results ofeach decision tree to make a final prediction. The final prediction is determined by taking themajority vote of the predictions from each decision tree.

The random selection of features and data subsets helps to reduce overfitting andimprove the generalization of the model. It also makes the algorithm more robust to outliers and missing data.

Random Forests can be used for both classification and regression problems. For classification, the algorithm predicts the class label with the highest frequency of votes from the decision trees. For regression, the algorithm predicts the average value of the target variablefrom the predictions of all decision trees.

One of the main advantages of Random Forests is that they are easy to use and do notrequire much data preprocessing. They can handle large datasets with many features and are relatively insensitive to the choice of hyperparameters. They also provide information about the relative importance of each feature in the prediction.

However, Random Forests can be slow and computationally expensive, especially for large datasets with many features. They may also not perform well for imbalanced datasets where one class is much more frequent than the other classes. Finally, since Random Forests use a combination of decision trees, their predictions can be difficult to interpret and visualize.

Naïve Bayes:

Naive Bayes is a popular machine learning algorithm for classification tasks. It is based on Bayes' theorem, which states that the probability of a hypothesis (such as a class label) given some observed evidence (such as features of an instance) is proportional to the probability ofthe evidence given the hypothesis, multiplied by the prior probability of the hypothesis.

In simpler terms, Naive Bayes algorithm assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature, given the class variable. This assumptionis called the "naive" assumption and hence the name Naive Bayes.

he algorithm builds a model by calculating the conditional probability of each featuregiven each class in the training data. Then, when presented with a new instance, the algorithm uses Bayes' theorem to calculate the probability of each class given the observed features. Theclass with the highest probability is then selected as the predicted class for the instance.

Naive Bayes algorithm is simple and computationally efficient, making it suitable for large datasets with high-dimensional feature spaces. It is often used in natural language processing tasks such as text classification, spam filtering, and sentiment analysis.

There are three main types of Naive Bayes algorithms:

Gaussian Naive Bayes: It assumes that the features follow a Gaussian (normal) distribution.

Multinomial Naive Bayes: It is used for discrete data such as text data, where each feature represents the frequency of a word or token.

Bernoulli Naive Bayes: It is similar to Multinomial Naive Bayes, but it is used for binarydata where each feature is either present or absent.

5.4 Proposed System:

The main contribution of this research was to use modern machine learning techniques to construct an intuitive medical prediction system for prediction of cyber attck. Different typesof machine learning classifier algorithms were trained in this study, including Random Forest
Classifier, Gaussian naive bayes and Decision Tree to select the best predictive model foraccurate cyberattack prediction.

5.5 Algorithm Implementation:

Naive Bayes:

Overall, Naive Bayes is a powerful and flexible algorithm that can be applied in manydifferent scenarios, from text classification to image recognition to fraud detection.

- simple and fast: Naive Bayes is a simple algorithm that is easy to implement and computationally efficient. It can quickly classify new data points once the model is trained.
- Works well with small datasets: Naive Bayes can perform well even with small datasets, whichmakes it useful in scenarios where there is limited data available.
- Robust to irrelevant features: Naive Bayes is robust to irrelevant features, meaning that it can still produce accurate results even if there are some features in the dataset that do not contributeto the classification task.
- Can handle high-dimensional data: Naive Bayes can handle datasets with a high number offeatures, which is often challenging for other machine learning algorithms.
- Interpretable results: The probabilities calculated by Naive Bayes can be interpreted as thelikelihood of a data point belonging to a particular class, which can provide insight into how
  the model is making its predictions.
- Can handle categorical and continuous data: Naive Bayes can handle both categorical andcontinuous data, making it a versatile algorithm that can be used in a wide range of classification tasks.

working process of naïve Bayes algorithm:

The Naive Bayes algorithm is a classification algorithm based on Bayes' theorem.The algorithm works by first training a model on a labeled dataset, and then using this modelto classify new, unlabeled data.

Data preprocessing: The first step is to prepare the data by cleaning it and converting it into asuitable format for the algorithm to work with.

Training: Once the data is preprocessed, the algorithm uses it to train a model. During this step, the algorithm calculates the probabilities of each feature occurring given each class. Forexample, if the feature is "age" and the class is "spam/not spam," the algorithm will calculatethe probability of each age group (e.g. 18-24, 25-30, etc.) occurring in spam emails vs. non spam emails.

Testing: After the model is trained, it is tested on a separate dataset to evaluate its accuracy. During this step, the algorithm takes in new, unlabeled data and calculates the probabilities ofit belonging to each class based on the probabilities calculated during the training step.

Prediction: Finally, the algorithm predicts the class with the highest probability for the new,unlabeled data.

One assumption made by the Naive Bayes algorithm is that all features are independent of eachother given the class. This assumption is often not true in real-world scenarios, but thealgorithm can still perform well in practice.

1. DDoS-PSH-ACK:

PSH-ACK is a DDoS attack designed to disrupt network activity by saturating bandwidth and resources on stateful devices in its path. By continuously sending ACK-PSH packets towards a target, stateful defenses can go down (In some cases into a fail open mode).

2. DDoS-ACK:

An ACK flood attack is when an attacker attempts to overload a server with TCP ACKpackets. Like other DDoS attacks, the goal of an ACK flood is to deny service to other users by slowing down or crashing the target using junk data.

3. Benign:

Visual explanation of a benign append attack. “M” refers to malicious and “B” refers to benign. This attack type is often seen in the real world in the form of benign library injections. In that case, malicious code is injected into a large benign file

SYSTEM TESTING

Software testing for a DDoS attack detection project can be quite challenging, as it requires simulating an attack and verifying that the detection and mitigation mechanisms work asexpected. Here are some suggestions for testing such a project:

6.1 UNIT TESTING:

Start by testing individual components of the software, such as the algorithms used for detecting and mitigating DDoS attacks. Use mock data to ensure that the individual components arefunctioning correctly.

6.2 INTEGRATION TESTING:

Test the integration of the different components to ensure they work together as expected. This includes verifying that the detection and mitigation mechanisms are triggered when they shouldbe and that they work together as a cohesive system.

6.3 LOAD TESTING:

Simulate a large number of requests to the system to see how it handles the load. This can include both legitimate requests and requests designed to mimic a DDoS attack. Verify that thedetection and mitigation mechanisms are triggered appropriately and that they can handle the load without impacting the system's performance.

6.4 PENETRATION TESTING:

Conduct penetration testing to verify that the system is secure and that there are no vulnerabilities that attackers can exploit to bypass the detection and mitigation mechanisms.

6.5 RED TEAM TESTING:

Hire a red team to simulate an actual DDoS attack and see how the system responds. This canhelp identify any weaknesses in the system and allow for improvements to be made.

6.6 REGRESSION TESTING:

As changes are made to the software, ensure that previous functionality has not been negativelyimpacted.

6.7 PERFORMANCE TESTING:

Monitor the system's performance during a simulated attack to ensure that it can detect and mitigate the attack within an acceptable time frame.

Overall, testing a DDoS attack detection project requires a comprehensive approach that takes into account the various aspects of the software and its potential use cases. By implementing a robust testing strategy, developers can ensure that their software is effective in protectingagainst DDoS attacks.

CODING

7.1 FRONTEND:HTML

```
<html>
    <head>
    <style>
    body{
    background-colo;pred;          background-
image:url("static/image5.gif");    background-size:1700px 750px;


    }
    </style>
            <meta charset="utf-8">
      <meta http-equiv="X-UA-Compatible" content="IE=edge">
      <meta name="viewport" content="width=device-width, initial-scale=1">
      <title>Login</title>
     <meta name="description" content="">
     <meta name="author" content="templatemo">
            <!--favicon-->
     <link rel="shortcut icon" href="favicon.ico" type="image/icon">
     <link rel="icon" href="favicon.ico" type="image/icon">
      <meta charset="utf-8">
```

```
 <meta name="viewport" content="width=device-width, initial-scale=1">
 <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
 <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.4.1/jquery.min.js"></script>
 <script       src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
<!-- Footer -->
      <link type="text/css" rel="stylesheet" href="../../Homepage/css/style.css">
<style>
            .type
            {
                  padding:5px;
    border-radius:10px;
    background-color :#ff4000;
            }
            </style>
            </body>
    </head>

    <body>
    <br><br><br><br><br><br>
    <p       style="font-size:45px;text-align:center;font-
family:Timesnewroman;color:black;fontweight:bolder;margin-top:100px;margin-left:-
600px;">GAIT CLASSIFICATION</p> <br>
<div style=margin-left:300px;margin-right:900px;>
```

```
    <form                                    style="text-align:center;color:black;font-size:17px;"
action="http://127.0.0.1:5000/login" method = "post">
 <br>
            <label  for="inputFirstName"                  style="margin-left:10px;font-
 size:25px;fontweight:bolder;color:black;">Upload The File</label>  
           <input type="file"  class="form-control" name="files" accept=".csv"> <br>
            <input type="submit" class="btn btn-primary"  name="submit" value="Submit">


     <br><br><br>
         </form>
       </div>


    </body>
</html>
```

7.2 BACKEND: PYTHON

```python
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
import pickle
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv("APA-DDoS-Dataset.csv")
print(data)


le=LabelEncoder()
data['frame.time'] = le.fit_transform(data['frame.time'])
data['ip.dst'] = le.fit_transform(data['ip.dst'])
data['ip.src'] = le.fit_transform(data['ip.src'])
data['Label'] = le.fit_transform(data['Label'])

X=data.drop(['Label','frame.time'],axis=1)

print(X)

Y=data['Label']
print(Y)

##To check the number of samples and features in your data
print("X shape:", X.shape)
print("y shape:", Y.shape)
```

```
32
33  ##To check the mean, standard deviation, minimum, maximum, and quartile values of each feature.
34  print("Describe",X.describe())
35
36  x_train,x_test,y_train,y_test = train_test_split(X,Y,shuffle=True,test_size=0.25, random_state=0)
37
38  from sklearn.naive_bayes import GaussianNB
39  NB = GaussianNB()
40  NB.fit(x_train, y_train)  #train the data
41  y_pred=NB.predict(x_test)
42  ##print(y_pred)
43  ##print(y_test)
44  print('Naive Bayes ACCURACY is', accuracy_score(y_test,y_pred))
45
46
47
48  from flask import *
49  import pickle
50  import pandas as pd
51  from sklearn.metrics import accuracy_score
52
53  app = Flask(__name__)
54
55  @app.route("/")
56  def home():
57      return render_template("browser1.html")
58  @app.route('/login',methods = ['POST'])
59  def login():
60          uname=request.form['files']
61          rr=pd.read_csv(uname)
```

```
        rr['ip.dst'] = le.fit_transform(rr['ip.dst'])
        rr['ip.src'] = le.fit_transform(rr['ip.src'])
        type(rr)
        y_pre=NB.predict(rr)
        if y_pre[0]==0:
            return render_template('index1.html')
        elif y_pre[0]==1:
            return render_template('index2.html')
        elif y_pre[0]==2:
            return render_template('index3.html')

if __name__ == '__main__':
    app.run()

```

8. CONCLUSION

Distributed Denial of Service (DDoS) attacks are a type of cyberattack that target a network or a website by overwhelming it with traffic from multiple sources. DDoS attacks are a serious threat to organizations and can cause significant damage, including downtime, data theft, and financial loss. Machine learning-based classification and prediction can be used to detect and mitigate DDoS attacks.In this project, we were able to successfully implement the proposed system, which uses supervised machine learning based naïve bayes algorithm. In project. Projects, three types of attacks are expected. For example, DDoSPSHACK, DDoS-ACK, and Benign. The final results will be shown in a Web Application made with HTML and CSS.

9. REFERENCE

[1] Mohmand MI, Hussain H, Khan AA, Ullah U, Zakarya M, Ahmed A, Raza M, Rahman IU, Haleem M. A machine learning-based classification and prediction technique for DDoS attacks. IEEE Access. 2022 Feb 17;10:21443-54.

[2] Alduailij M, Khan QW, Tahir M, Sardaraz M, Alduailij M, Malik F. Machine-LearningBased DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method. Symmetry. 2022 May 27;14(6):1095.

[3] Rezvy S, Luo Y, Petridis M, Lasebae A, Zebin T. An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks. In2019 53rd Annual Conference on information sciences and systems (CISS) 2019 Mar 20 (pp. 1-6). IEEE.

[4] Mittal M, Kumar K, Behal S. Deep learning approaches for detecting DDoS attacks: A systematic review. Soft Computing. 2022 Jan 27:1-37.

[5] Pei J, Chen Y, Ji W. A DDoS attack detection method based on machine learning. InJournal of Physics: Conference Series 2019 Jun 1 (Vol. 1237, No. 3, p. 032040). IOP Publishing.

10. SCREENSHORTS

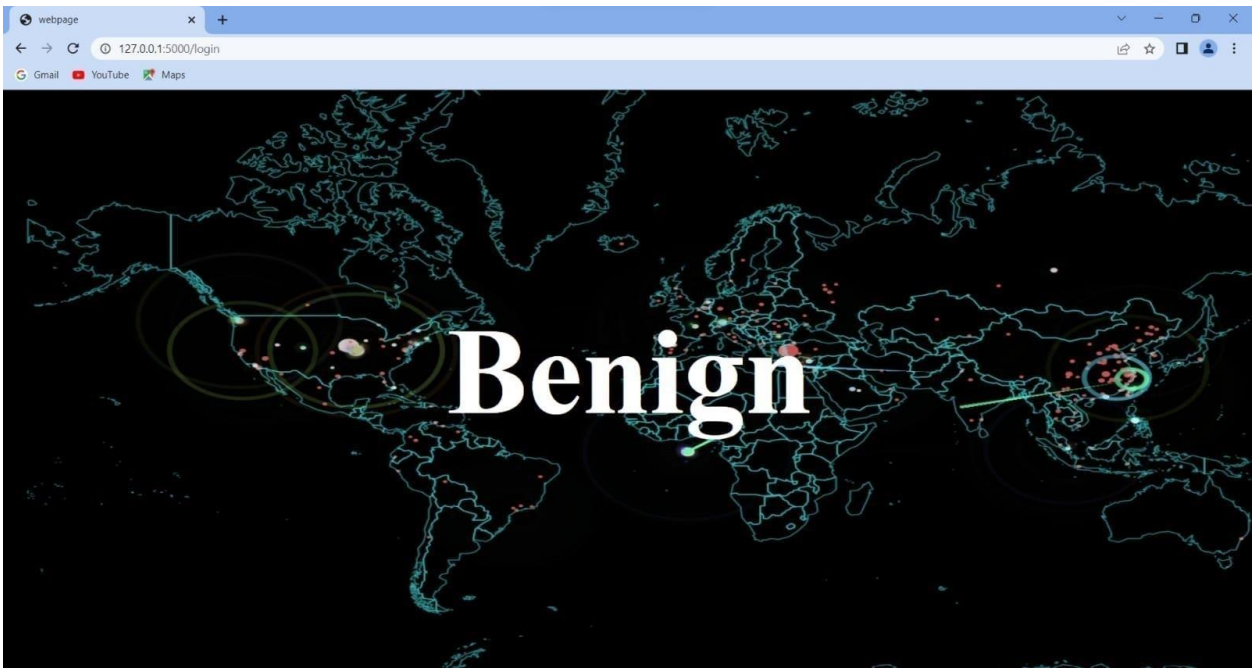
WEBPAGE:

TEST 1:

TEST 2:


webpage
127.0.0.1:5000/login
Gmail
YouTube
Maps
DDoS-PSH-ACK

TEST 3:


webpage
127.0.0.1:5000/login
Gmail
YouTube
Maps
Benign