

1.INTRODUCTION

In recent years, Big Data Analytics (BDA) has emerged as a growing way to read records and extract statistics and similarities in various software environments. Due to urban sprawl and population development, crime plays a vital role in our society. However, those qualities have been further followed by the rise of violent crime and accidents. To address these issues, social scientists, analysts, and defense institutions have devoted much effort to understanding the styles and skills of the mines. When it comes to public access, there are many situations that require a lot of access to a record. As a result, new techniques and technologies need to be designed to study these different records and to be available in more places. Analysis of such large records allows us to better manage events, identify similarities, set inventory and make short decisions accordingly. This can also help by increasing our knowledge of the problems of individual history and modern conditions, over time ensuring improved security / safety and a better standard of living, in addition to cultural expansion and financial growth. The rapid growth of cloud computing and recording of grocery purchases and technology, from commercial enterprises and research institutes to many governments and groups, has created a tremendous amount of weight / sophistication from the records collected and made available to the public. It has emerged as a growing value for the importance of extracting important statistics and gaining new insights into information styles in those record assets. The BDA can effectively handle the demands of records that may be very large, unstructured, and quickly transferred to standardized methods. As a fast-growing and powerful practice, the DBA can mobilize circles to use its records and promote new freedom. In addition, the BDA can be used to help sensible organizations spread out in advance with more dynamic jobs, over-earning and happy customers.

1.1 Existing System

Many researchers have gone through this problem regarding the criminal cases being unsolved for a long period. They proposed different crime prediction algorithms. In all these models the accuracy will surely vary depending on the data set and the features or attributes we select during data preprocessing.

In Crime prediction done on the Mississippi crime data set where models like linear regression and Decision stump model are used gave a result of 83%, 88% and 67% respectively [1]. Although these accuracies of the predictions may vary accordingly because

it is discovered that many machine learning algorithms are implemented on data sets consisting of different places having distinctive features, so predictions are changing in all cases

1.2 Proposed System

The proposed system is made on the basis of the research work that is done by going through various such documentations. Nearly all of the crimes are predicting based on the location and the types of crimes that are occurring in those areas.

On surveying previous works, Linear Regression, Decision Tree and Random Forest tend to give good accuracy so these models are used in this paper to predict crimes. The dataset used in this paper is from data.world.com. The data set contains different types of crimes that being committed in India according to the state and year respectively [4]. This paper takes types of crimes as input and gives the area in which crimes are committed as output. The data preprocessing involves data cleaning, feature selection, dropping null values, data scaling by normalizing and standardizing. After data preprocessing the data is free of null values which may alter the accuracy of the model significantly and feature selection is used to select only the required features that won't affect the accuracy of mode

1.3 Literature survey

Wang, Wei Li, Gangfeng Yan and Sumian Song, "Towards a framework for cyber attack impact analysis of electric cyber physical systems", *Industrial Technology (ICIT) 2017 IEEE International Conference on*, pp. 638-643, 2017.

This paper proposes an unified framework for electrical cyber physical systems (ECPSs) and studies the mechanism of cyber attacks and cyber security. Communication networks are designed by characteristics of power grids. This model is universal to both transmission and distribution grids. The fragility of ECPSs under cyber attacks (DoS attacks and false data injection attacks) is analyzed in three scenarios based on different types and attackers' acknowledgments of the ECPSs. It has been proved that attacks happen at information uploading routers or control strategy downloading routers will influence the system differently. The effectiveness of relay protection policies and cyber security policies are verified by experimental results.

M. Khan, S. K. Pradhan and H. Fathima, "Applying data Mining Techniques in Cyber Crimes", *International Conference on Anti Cyber Crimes(ICACC)*, vol. 109, no. 2, pp. 315-426, 2017.

Globally the internet is been accessed by enormous people within their restricted domains. When the client and server exchange messages among each other, there is an activity that can be observed in log files. Log files give a detailed description of the activities that occur in a network that shows the IP address, login and logout durations, the user's behavior etc. There are several types of attacks occurring from the internet. Our focus of research in this paper is Denial of Service (DoS) attacks with the help of pattern recognition techniques in data mining. Through which the Denial of Service attack is identified. Denial of service is a very dangerous attack that jeopardizes the IT resources of an organization by overloading with imitation messages or multiple requests from unauthorized users.

Mariam Nouh, Jason RC Nurse and Michael Goldsmith, "Towards Designing a Multipurpose Cybercrime Intelligence Framework", *Intelligence and Security Informatics Conference (EISIC)*, pp. 60-67, 2016.

With the wide spread of the Internet and the increasing popularity of social networks that provide prompt and ease of communication, several criminal and radical groups have adopted it as a medium of operation. Existing literature in the area of cybercrime intelligence focuses on several research questions and adopts multiple methods using techniques such as social network analysis to address them. In this paper, we study the broad state-of-the-art research in cybercrime intelligence in order to identify existing research gaps. Our core aim is designing and developing a multipurpose framework that is able to fill these gaps using a wide range of techniques. We present an outline of a framework designed to aid law enforcement in detecting, analysing and making sense out of cybercrime data.

Akshay Kumar Singh, Neha Prasad, Nohil Narkhede and Siddharth Mehta, "Crime: Classification and Pattern Prediction", *IARJSET*, vol. 3, no. 2, pp. 41-43, February 2016.

The growing availability of information technologies has enabled law enforcement agencies to collect detailed data about various crimes. Classification techniques can be applied to these data to build decision-aid tools and facilitate investigations of law enforcement agencies. In this paper, we propose an approach for constructing a decision tree based classification model for a crime prediction. Proposed model assists law enforcement

agencies in discovering crime patterns and predicting future trends. We provide an implementation and analysis of our proposed method.

Rasoul Kiani, Silamak Mahdavi and Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", *IJARAI*, vol. 4, no. 8, pp. 1-7, 2015.

Crimes will somehow influence organizations and institutions when occurred frequently in a society. Thus, it seems necessary to study reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. The main objective of this paper is to classify clustered crimes based on occurrence frequency during different years. Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. We applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales within 1990 to 2011. We assigned weights to the features in order to improve the quality of the model and remove low value of them. The Genetic Algorithm (GA) is used for optimizing of Outlier Detection operator parameters using RapidMiner tool.

H.-K. Oh, S.-W. Kim, S. Park and M. Zhou, "Can you trust online ratings? A mutual reinforcement model for trustworthy online rating systems", *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 45, no. 12, pp. 1564-1576, Dec. 2015.

The average of customer ratings on a product, which we call a reputation, is one of the key factors in online purchasing decisions. There is, however, no guarantee of the trustworthiness of a reputation since it can be manipulated rather easily. In this paper, we define false reputation as the problem of a reputation being manipulated by unfair ratings and design a general framework that provides trustworthy reputations. For this purpose, we propose TRUE-REPUTATION, an algorithm that iteratively adjusts a reputation based on the confidence of customer ratings. We also show the effectiveness of TRUE-REPUTATION through extensive experiments in comparisons to state-of-the-art approaches.

M. Cui, J. Wang, A. R. Florita and Y. Zhang, "Generalized graph Laplacian based anomaly detection for spatiotemporal microPMU data", *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3960-3963, Sep. 2019.

This letter develops a novel anomaly detection method using the generalized graph Laplacian (GGL) matrix to visualize the spatiotemporal relationship of distribution-level phasor measurement unit (μ PMU) data. The μ PMU data in a specific time horizon are segregated into multiple segments. An optimization problem formulated as a Lagrangian function is utilized to estimate the GGL matrix. During the iterative process, an optimal update is constituted as a quadratic program problem. To perform the μ PMU-based spatiotemporal analysis, normalized diagonal elements of GGL matrix are proposed as a quantitative metric. The effectiveness of the developed method is demonstrated through real-world μ PMU measurements gathered from test feeders in Riverside, CA, USA.

H. E. Egilmez, E. Pavez and A. Ortega, "Graph learning from data under Laplacian and structural constraints", *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 6, pp. 825-841, Sep. 2017.

Graphs are fundamental mathematical structures used in various fields to represent data, signals, and processes. In this paper, we propose a novel framework for learning/estimating graphs from data. The proposed framework includes (i) formulation of various graph learning problems, (ii) their probabilistic interpretations, and (iii) associated algorithms. Specifically, graph learning problems are posed as the estimation of graph Laplacian matrices from some observed data under given structural constraints (e.g., graph connectivity and sparsity level). From a probabilistic perspective, the problems of interest correspond to maximum a posteriori parameter estimation of Gaussian-Markov random field models, whose precision (inverse covariance) is a graph Laplacian matrix. For the proposed graph learning problems, specialized algorithms are developed by incorporating the graph Laplacian and structural constraints. The experimental results demonstrate that the proposed algorithms outperform the current state-of-the-art methods in terms of accuracy and computational efficiency.

M. Cui, J. Wang and M. Yue, "Machine learning based anomaly detection for load forecasting under cyberattacks", *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5724-5734, Sep. 2019.

Accurate load forecasting can create both economic and reliability benefits for power system operators. However, the cyberattack on load forecasting may mislead operators to make unsuitable operational decisions for the electricity delivery. To effectively and accurately detect these cyberattacks, this paper develops a machine learning-based anomaly detection

(MLAD) methodology. First, load forecasts provided by neural networks are used to reconstruct the benchmark and scaling data by using the k-means clustering. Second, the cyberattack template is estimated by the naive Bayes classification based on the cumulative distribution function and statistical features of the scaling data. Finally, the dynamic programming is utilized to calculate both the occurrence and parameter of one cyberattack on load forecasting data. A widely used symbolic aggregation approximation method is compared with the developed MLAD method. Numerical simulations on the publicly load data show that the MLAD method can effectively detect cyberattacks for load forecasting data with relatively high accuracy. Also, the robustness of MLAD is verified by thousands of attack scenarios based on Monte Carlo simulation.

M. Cui, J. Zhang, A. R. Florita, B.-M. Hodge, D. Ke and Y. Sun, "An optimized swinging door algorithm for identifying wind ramping events", *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 150-162, Jan. 2016.

With the increasing penetration of renewable energy in recent years, wind power ramp events (WPRES) have started affecting the economic and reliable operation of power grids. In this paper, we develop an optimized swinging door algorithm (OpSDA) to improve the state of the art in WPRES detection. The swinging door algorithm (SDA) is utilized to segregate wind power data through a piecewise linear approximation. A dynamic programming algorithm is performed to optimize the segments by: 1)merging adjacent segments with the same ramp changing direction; 2)handling wind power bumps; and 3)postprocessing insignificant-ramps intervals. Measured wind power data from two case studies are utilized to evaluate the performance of the proposed OpSDA. Results show that the OpSDA provides 1)significantly better performance than the SDA and 2)equal-to-better performance compared to the L1-Ramp Detect with Sliding Window (L1-SW) method with significantly less computational time.

2. METHODOLOGY

A support vector machine is a very important and versatile machine learning algorithm, it is capable of doing linear and nonlinear classification, regression and outlier detection. Support vector machines also known as SVM is another algorithm widely used by machine learning people for both classification as well as regression problems but is widely used for classification tasks. It is preferred over other classification algorithms because it uses less computation and gives notable accuracy. It is good because it gives reliable results even if there is less data.

We will explain in this blog What is SVM, how SVM works, pros and cons of SVM, and hands on problem using SVM in python.

What Is Support Vector Machine (Svm)?

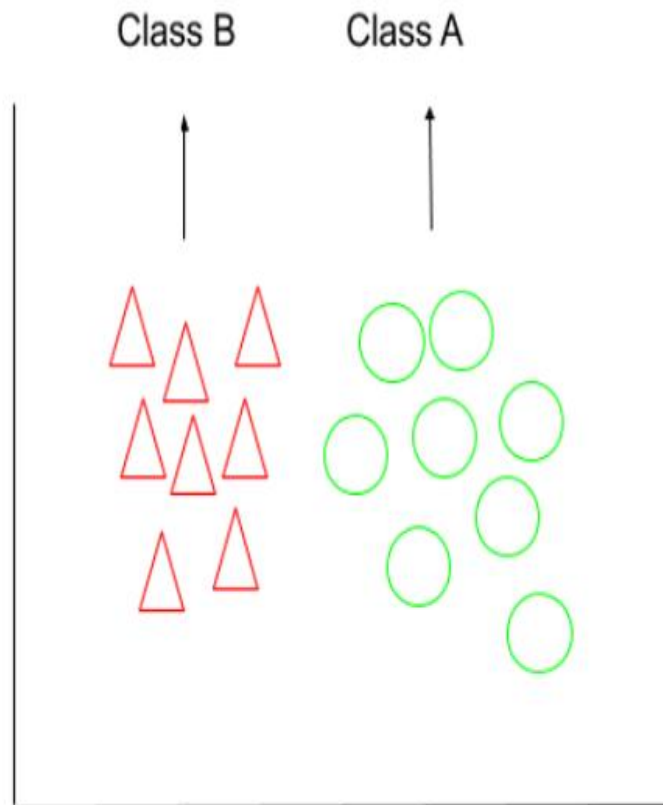
A support vector machine is a machine learning model that is able to generalise between two different classes if the set of labelled data is provided in the training set to the algorithm. The main function of the SVM is to check for that hyperplane that is able to distinguish between the two classes.

There can be many hyperplanes that can do this task but the objective is to find that hyperplane that has the highest margin that means maximum distances between the two classes, so that in future if a new data point comes that is two be classified then it can be classified easily.

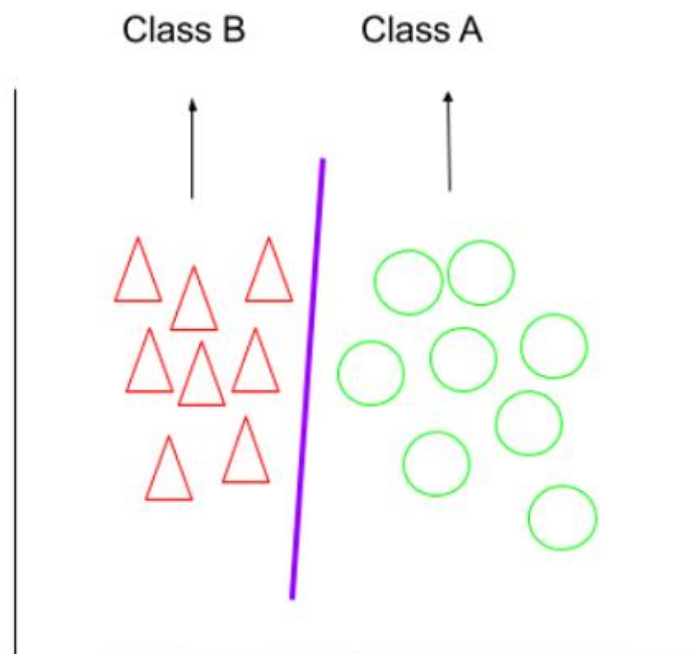
How Does Svm Works?

1. Linearly Separable Data

Let us understand the working of SVM by taking an example where we have two classes that are shown in the below image which are a class A: Circle & class B: Triangle. Now, we want to apply the SVM algorithm and find out the best hyperplane that divides the both classes.

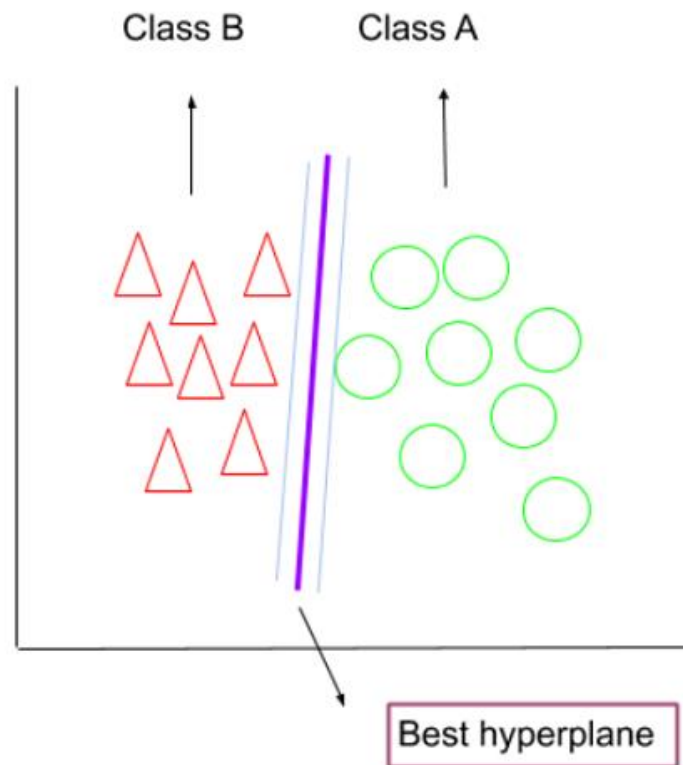


Class A and B



Labelled Data

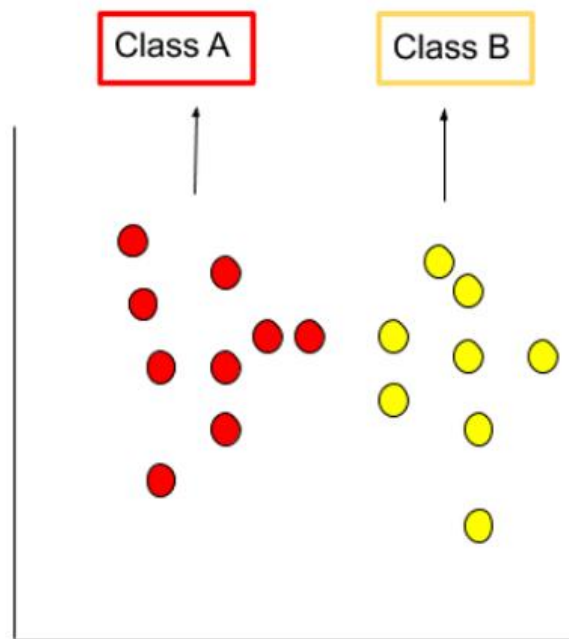
SVM takes all the data points in consideration and gives out a line that is called '**Hyperplane**' which divides both the classes. This line is termed as '**Decision boundary**'. Anything that falls in circle class will belong to the class A and vice-versa.



All hyperplanes are not good at classification

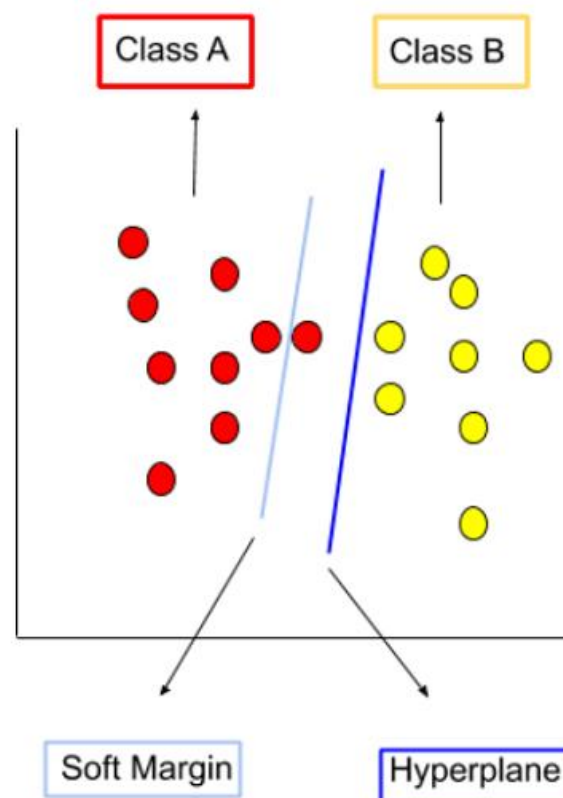
There can be many hyperplanes that you can see but the best hyper plane that divides the two classes would be the hyperplane having a large distance from the hyperplane from both the classes. That is the main motive of SVM to find such best hyperplanes.

There can be different dimensions which solely depends upon the features we have. It is tough to visualize when the features are more than 3.



Class A- Red & Class- B Yellow

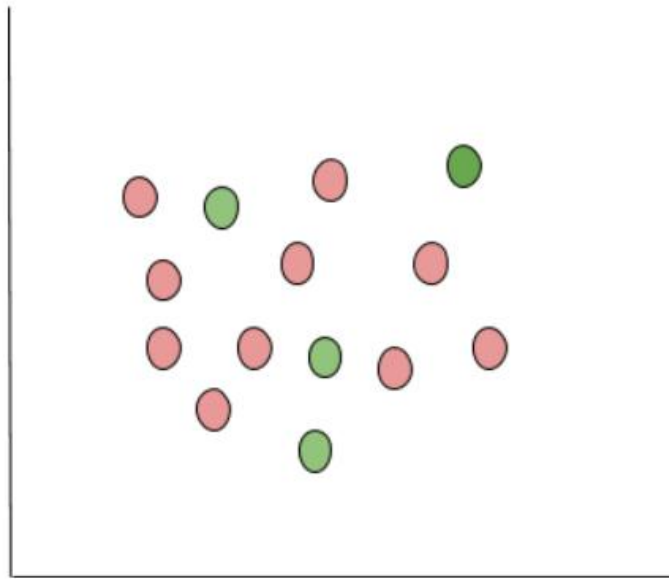
Consider we have two classes that are red and yellow class A and B respectively. We need to find the best hyperplane between them that divides the two classes.



Soft margin and hyperplane

Soft margin permits few of the above data points to get misclassified. Also, it tries to make the balance back and forth between finding a hyperplane that attempts to make less misclassifications and maximize the margin.

2. Linearly Non-separable Data



Linearly non-separable dataset

If the data is non linearly separable as shown in the above figure then SVM makes use of kernel tricks to make it linearly separable. The concept of transformation of non-linearly separable data into linearly separable is called **Cover's theorem** - *“given a set of training data that is not linearly separable, with high probability it can be transformed into a linearly separable training set by projecting it into a higher-dimensional space via some non-linear transformation”*. Kernel tricks help in projecting data points to the higher dimensional space by which they became relatively more easily separable in higher-dimensional space.

Kernel Tricks:

Kernel tricks also known as Generalized dot product. Kernel tricks are the way of calculating dot product of two vectors to check how much they make an effect on each other. According to Cover's theorem the chances of linearly non-separable data sets becoming linearly

separable increase in higher dimensions. Kernel functions are used to get the dot products to solve SVM constrained optimization.

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.SOFTWARE REQUIREMENT SPECIFICATION

3.1 HARDWARE AND SOFTWARE REQUIREMENTS

SYSTEM REQUIREMENTS

4.1 Hardware Requirements

- ☐ System : Core i3.
- ☐ Hard Disk : 40 GB.
- ☐ Mouse : Optical Mouse.
- ☐ Ram : 4GB.

4.2 Software Requirements

- ☐ Operating system : Windows 7 Ultimate.
- ☐ Coding Language : Python.
- ☐ Front-End : Python.
- ☐ Designing : Html,css,javascript.
- ☐ Data Base : MySQL.

3.2 DESCRIPTION ABOUT TOOLS USED

4. TECHNOLOGY

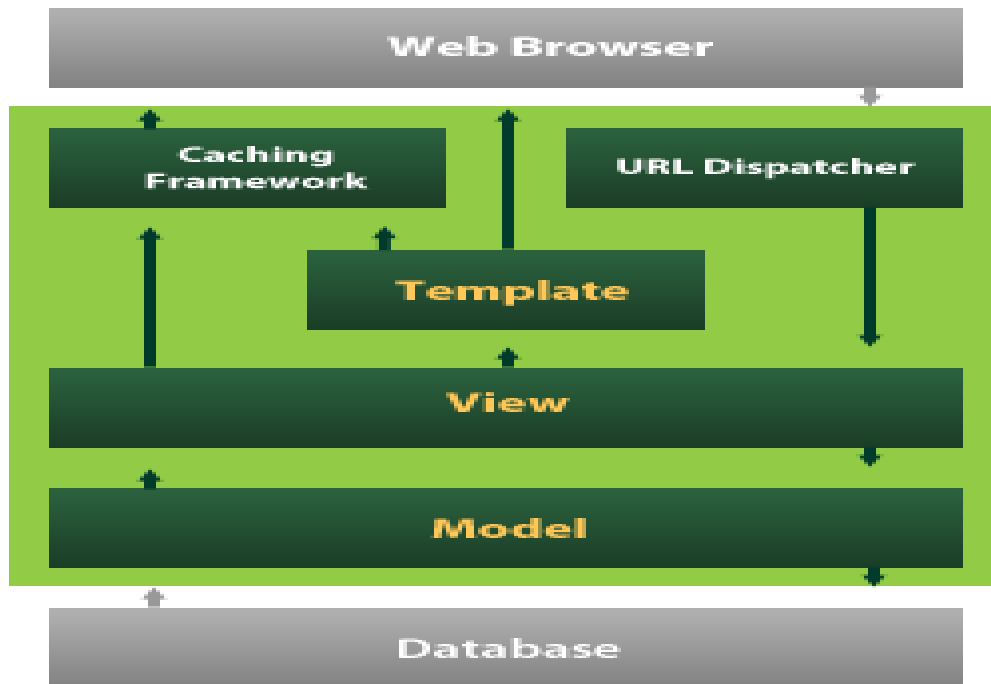
PYTHON

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. An interpreted language, Python has a design philosophy that emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++ or Java. It provides constructs that enable clear programming on both small and large scales. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit Python Software Foundation. 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

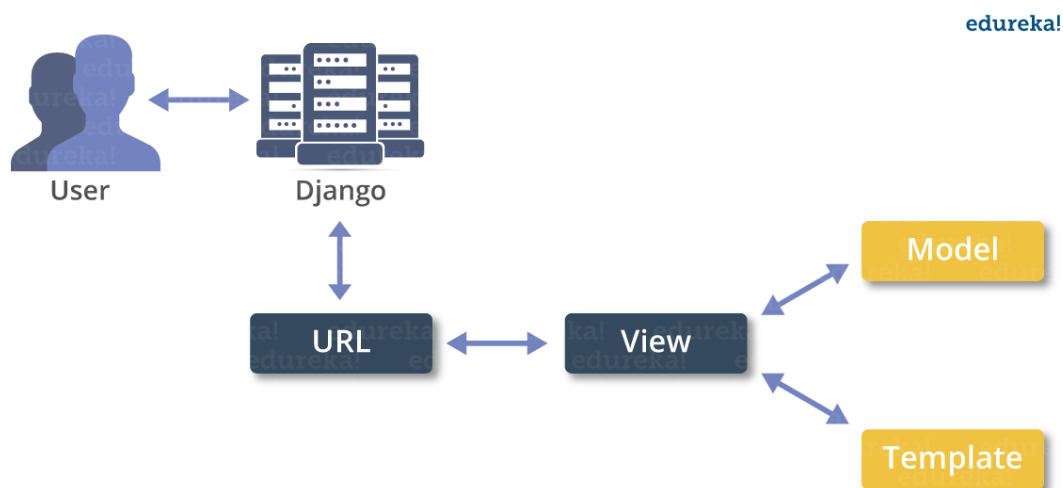
DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.



Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models



Installation work

PyCharm is a cross-platform editor developed by JetBrains. Pycharm provides all the tools you need for productive Python development.

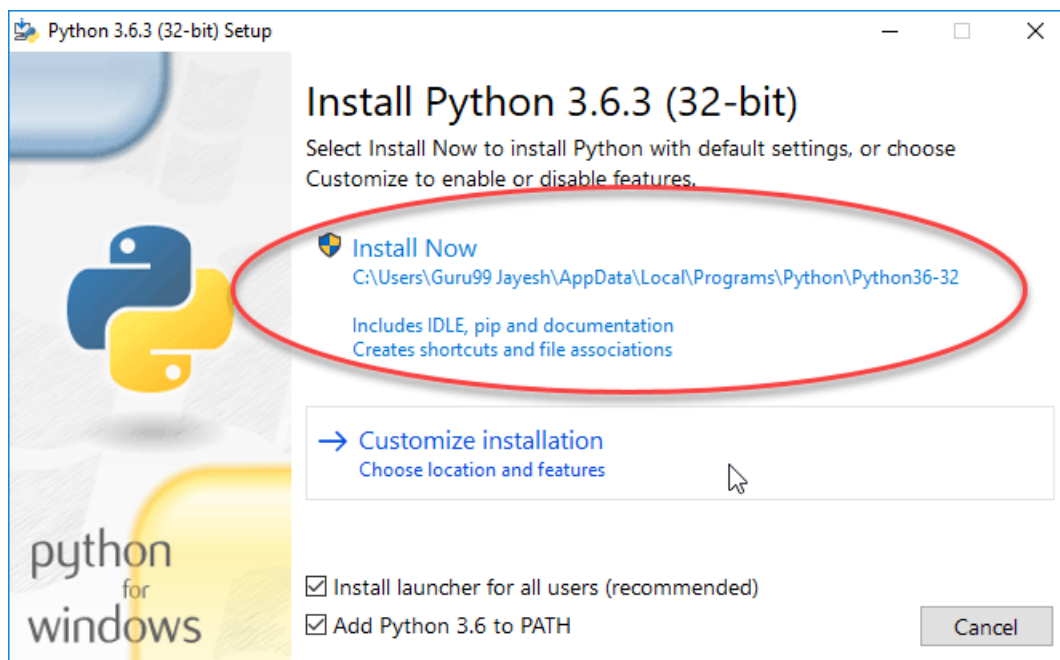
Below are the detailed steps for installing Python and PyCharm

Installing Python

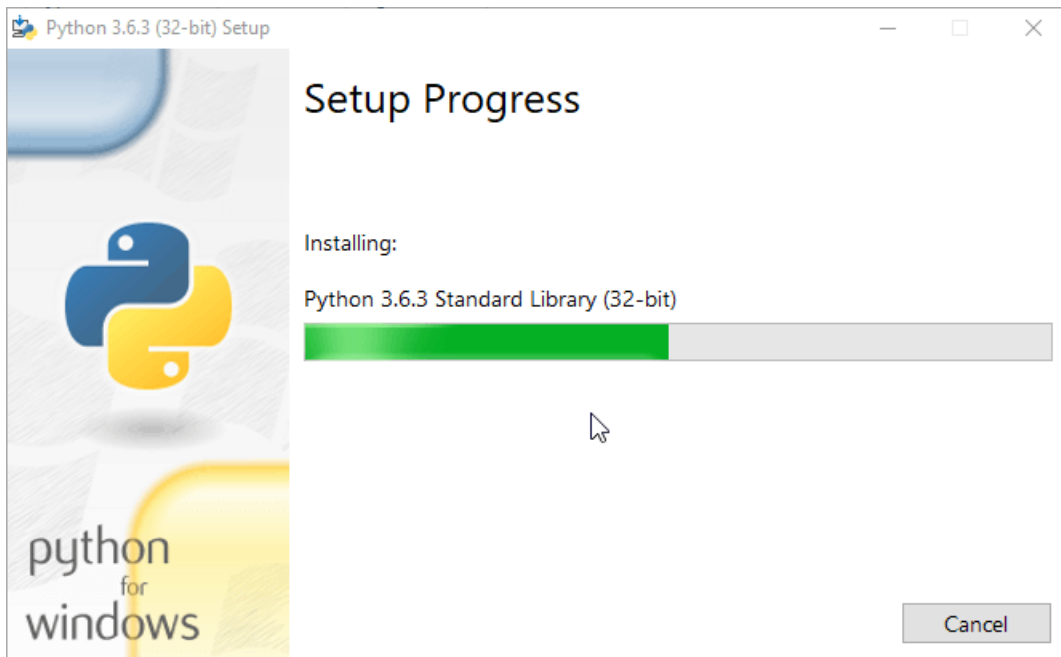
Step 1) To download and install Python visit the official website of Python <http://www.python.org/downloads/> and choose your version. We have chosen Python version 3.6.3



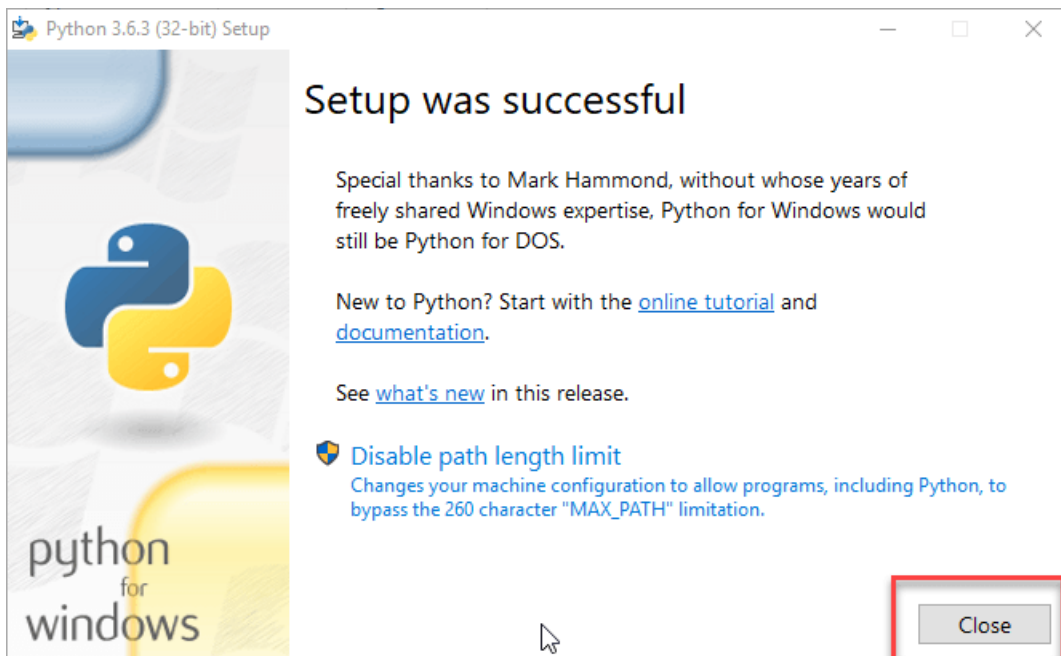
Step 2) Once the download is complete, run the exe for install Python. Now click on Install Now.



Step 3) You can see Python installing at this point.



Step 4) When it finishes, you can see a screen that says the Setup was successful. Now click on "Close".



Installing Pycharm

Step 1) To download PyCharm visit the website <https://www.jetbrains.com/pycharm/download/> and Click the "DOWNLOAD" link under the Community Section.

Download PyCharm

Windows

macOS

Linux

Professional

Full-featured IDE
for Python & Web
development

DOWNLOAD

Free trial

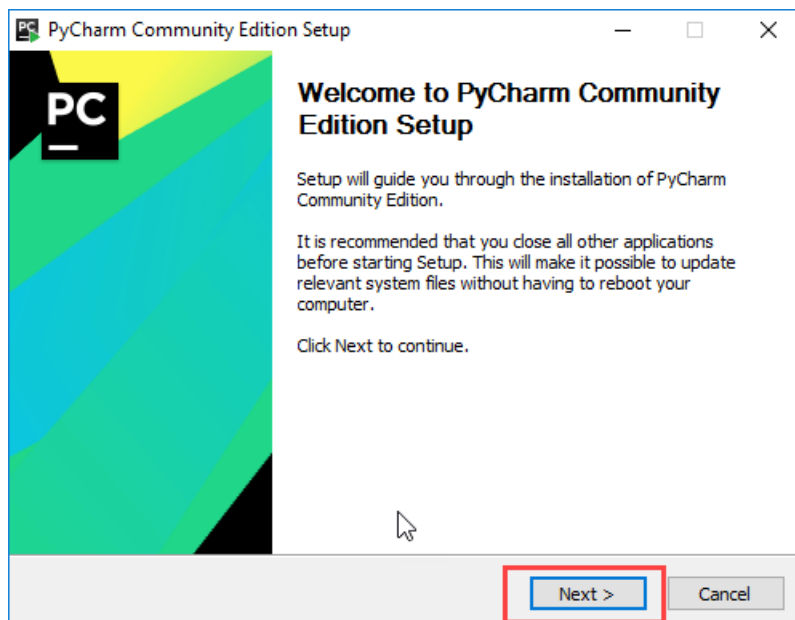
Community

Lightweight IDE
for Python & Scientific
development

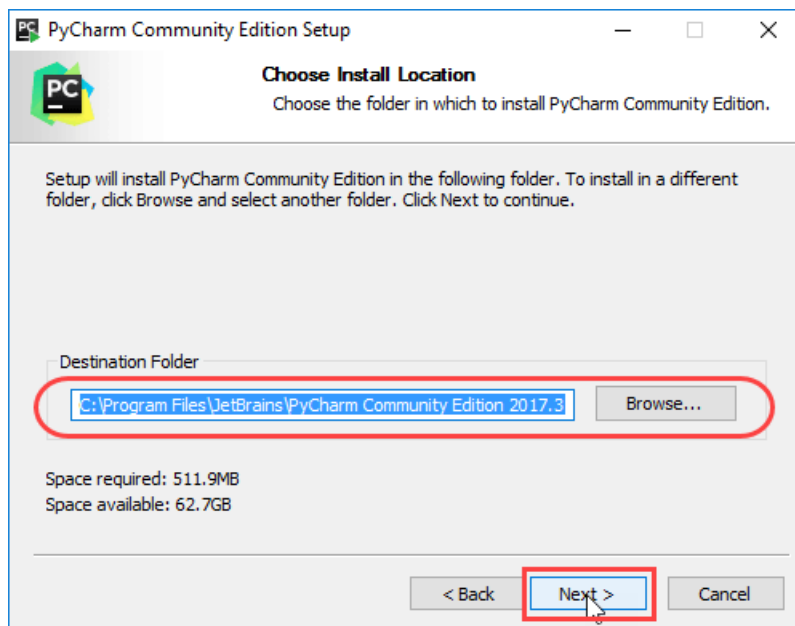
DOWNLOAD

Free, open-source

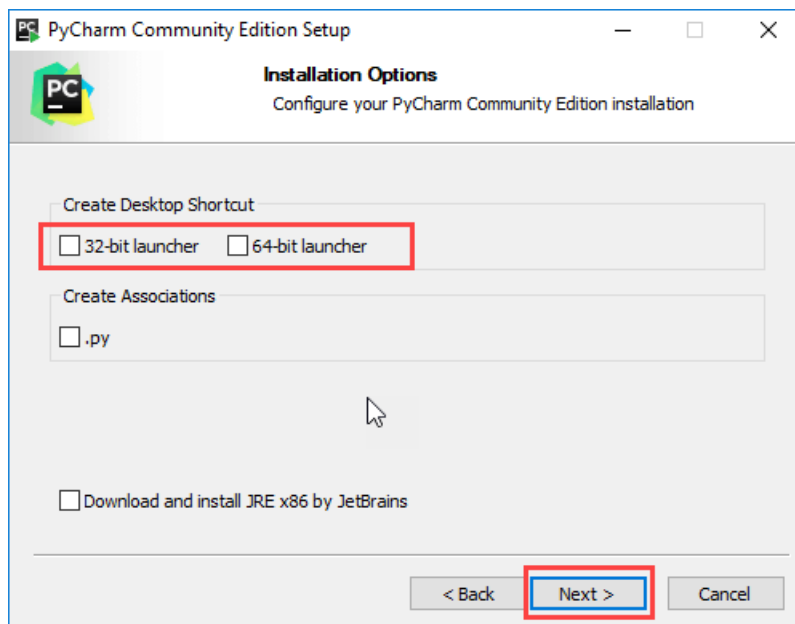
Step 2) Once the download is complete, run the exe for install PyCharm. The setup wizard should have started. Click “Next”.



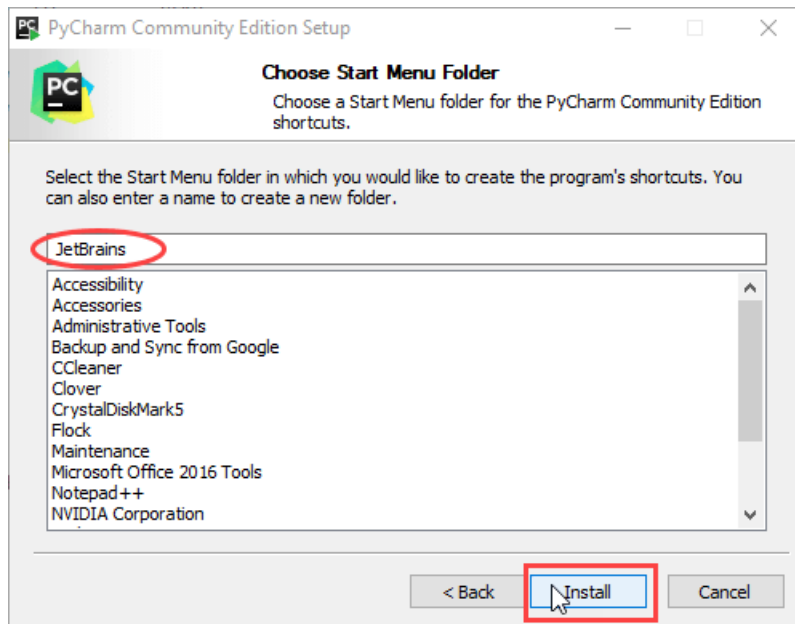
Step 3) On the next screen, Change the installation path if required. Click “Next”.



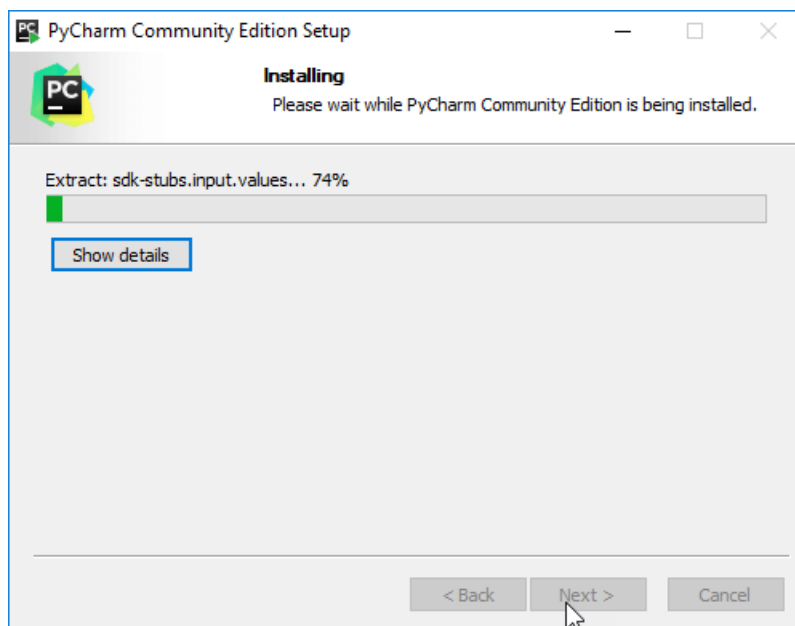
Step 4) On the next screen, you can create a desktop shortcut if you want and click on “Next”.



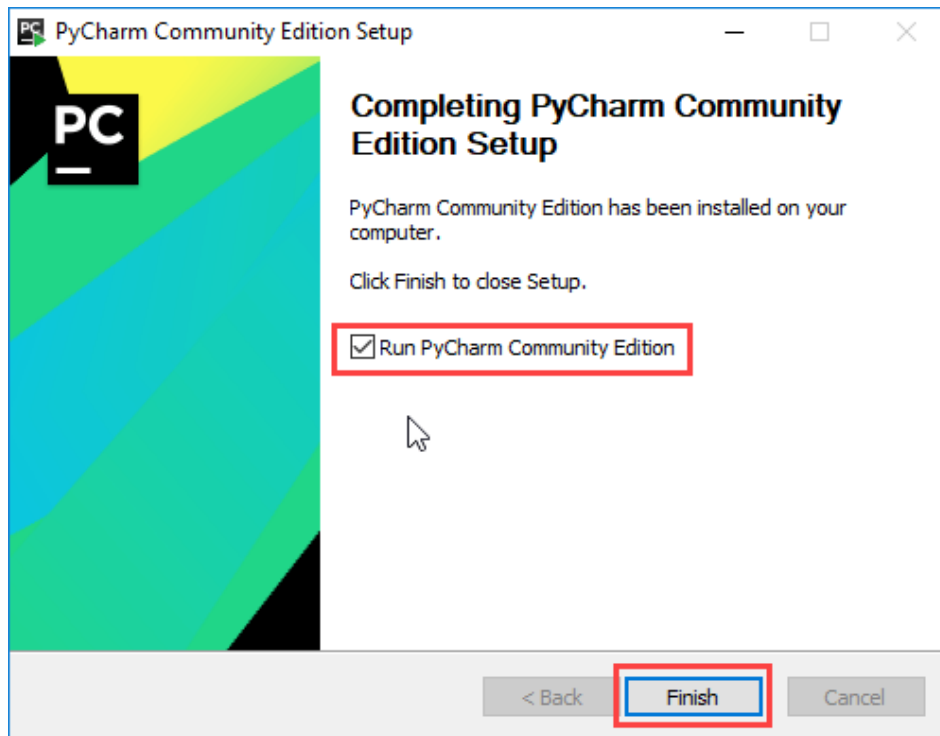
Step 5) Choose the start menu folder. Keep selected JetBrains and click on “Install”.



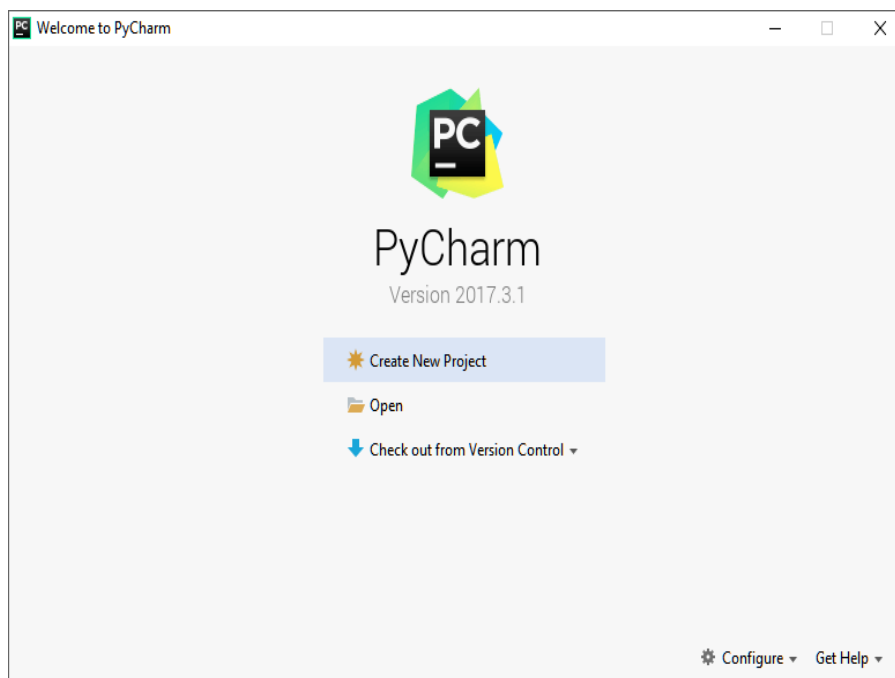
Step 6) Wait for the installation to finish.



Step 7) Once installation finished, you should receive a message screen that PyCharm is installed. If you want to go ahead and run it, click the “Run PyCharm Community Edition” box first and click “Finish”.



Step 8) After you click on "Finish," the Following screen will appear.



Creating First Program

Step 1) Open PyCharm Editor. You can see the introductory screen for PyCharm. To create a new project, click on “Create New Project”.

5. MODULE DESCRIPTION

Data Collection:

The data set used is the crimes that are committed in India during the year 2001-2018 which is available in the dataset world. It consists of features like the states of India and the districts of every state where

the crimes are committed. It also gives the type of crimes that are being committed such as kidnapping, raping, robbery, theft, criminal breach of trust,

Transaction in websites has high risk of users performing fraudulent activities such as doing money laundry, using stolen identity and credit card, etc. due to unscreened and diversified background of users. The rise in artificial intelligent (machine learning in specific) enables to detect these fraudulent activities with accuracy and real-time. However, the activities and background of a new user visiting a website are usually not directly learnable by traditional machine learning algorithms. Feature extraction from activity records and feature transformation from user's background are a necessity as well as a prerequisite

Data Processing:

The first and major step in data PreProcessing is done in order to remove the null values and the features or attributes that are unnecessary. Nine thousand entries are present in the dataset that is being used in this [8]. All the null values are removed. To use the data consisting of string values there is a need to convert that string values to float to use the machine learning algorithms efficiently. This conversion of data can be done in mainly two ways one is one hot encoding and the other one is label encoding. The one which is used here is label encoding

The dataset contain the information from the amazon web site where frwads might have happened.. it contain many unwanted contents, so those contents needs to be removed for that purpose, data processing is applied.

Feature Selection

Feature Selection is the method done in order to avoid the alteration of accuracy or to increase the accuracy by only selecting the required features or attributes in given data. This increases the accuracy of the model by removing unnecessary attributes.

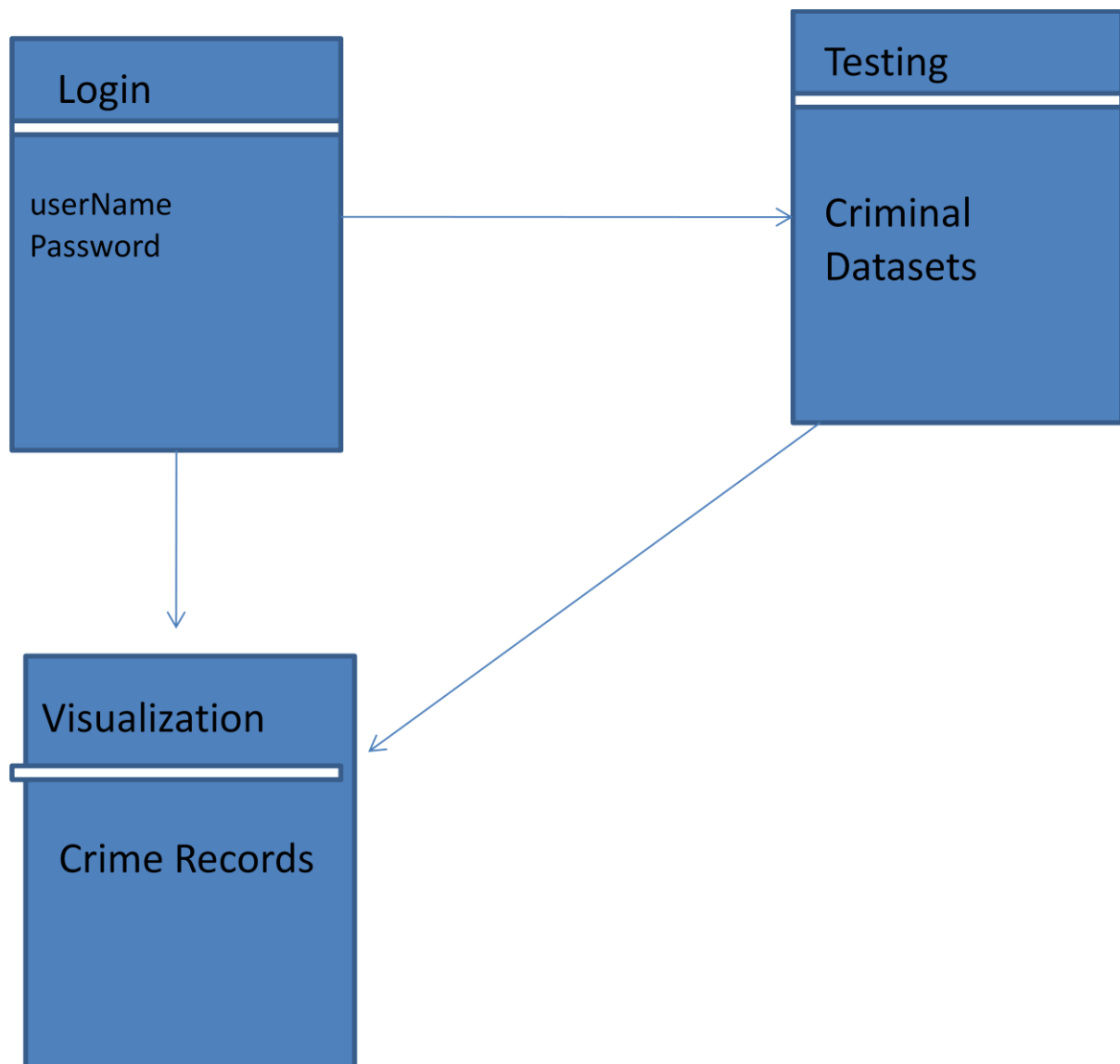
SYSTEM MODULING USING UML

INTRODUCTION

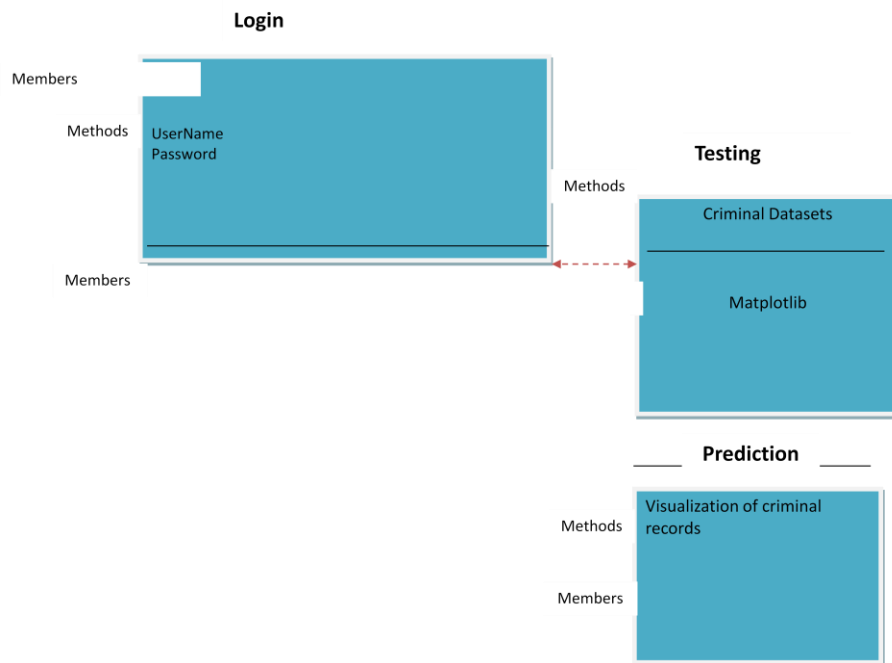
A support vector machine is a machine learning model that is able to generalise between two different classes if the set of labelled data is provided in the training set to the algorithm. The main function of the SVM is to check for that hyperplane that is able to distinguish between the two classes.

There can be many hyperplanes that can do this task but the objective is to find that hyperplane that has the highest margin that means maximum distances between the two classes, so that in future if a new data point comes that is to be classified then it can be classified easily.

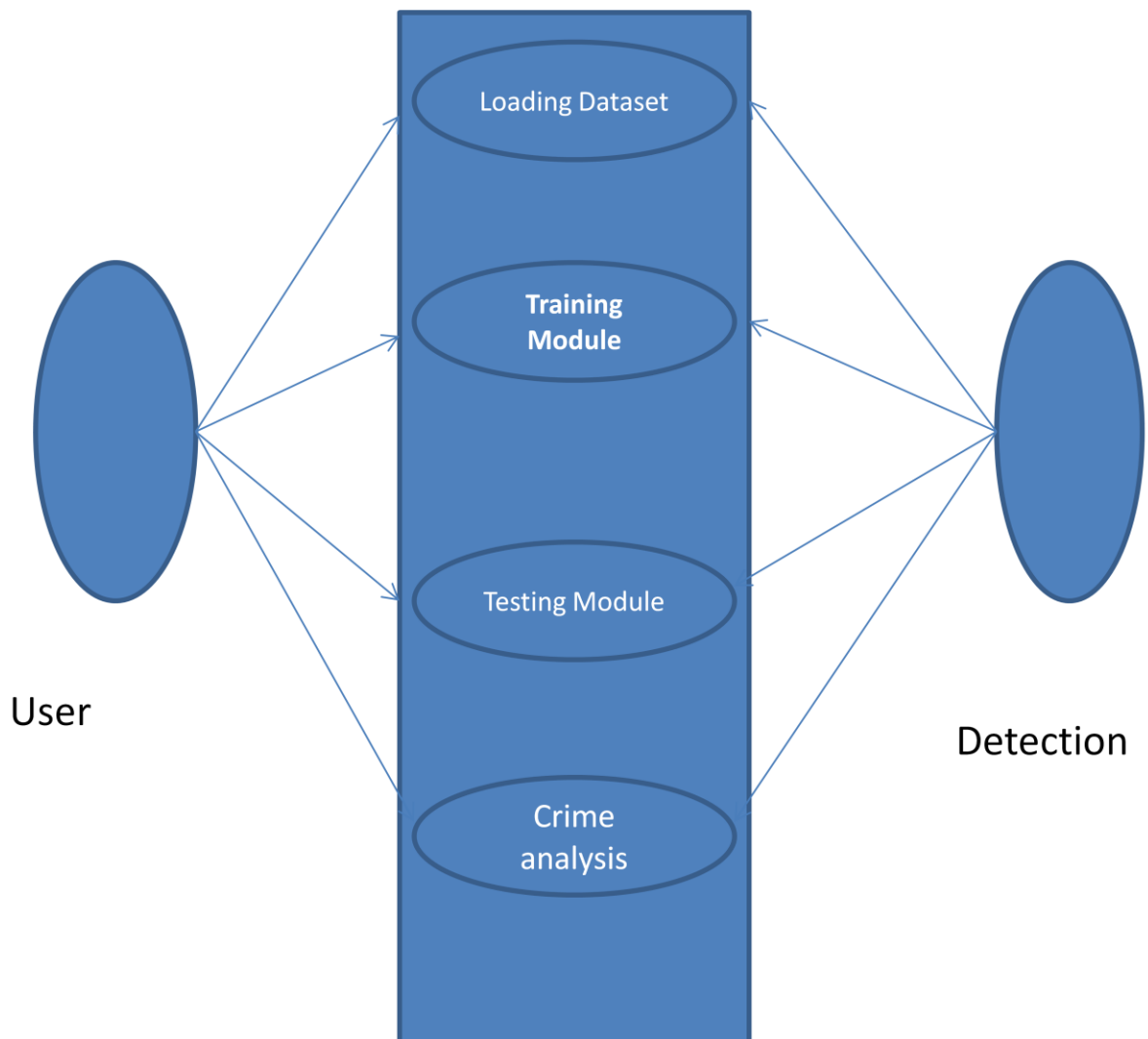
CLASS DIAGRAM



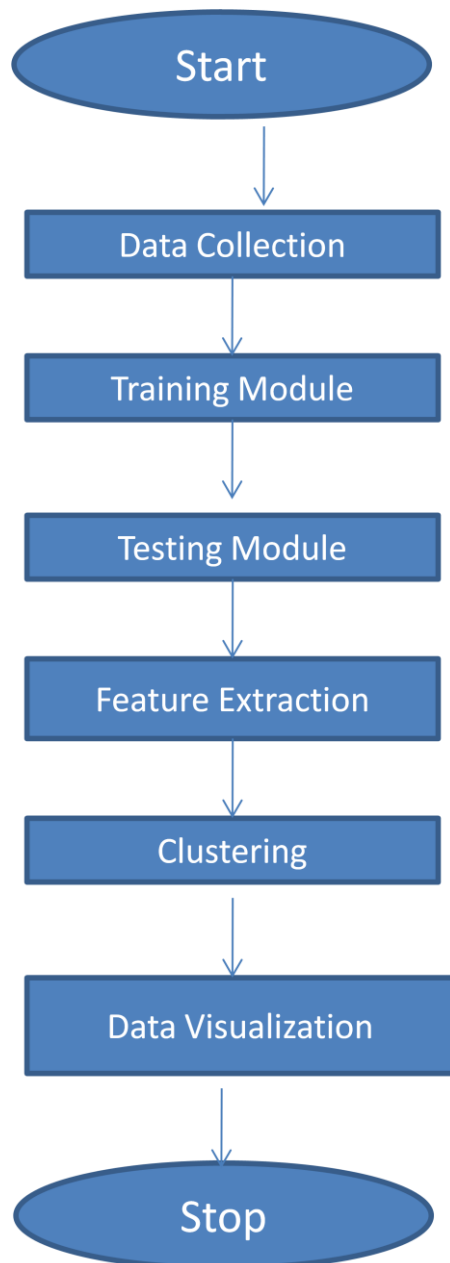
OBJECT DIAGRAM



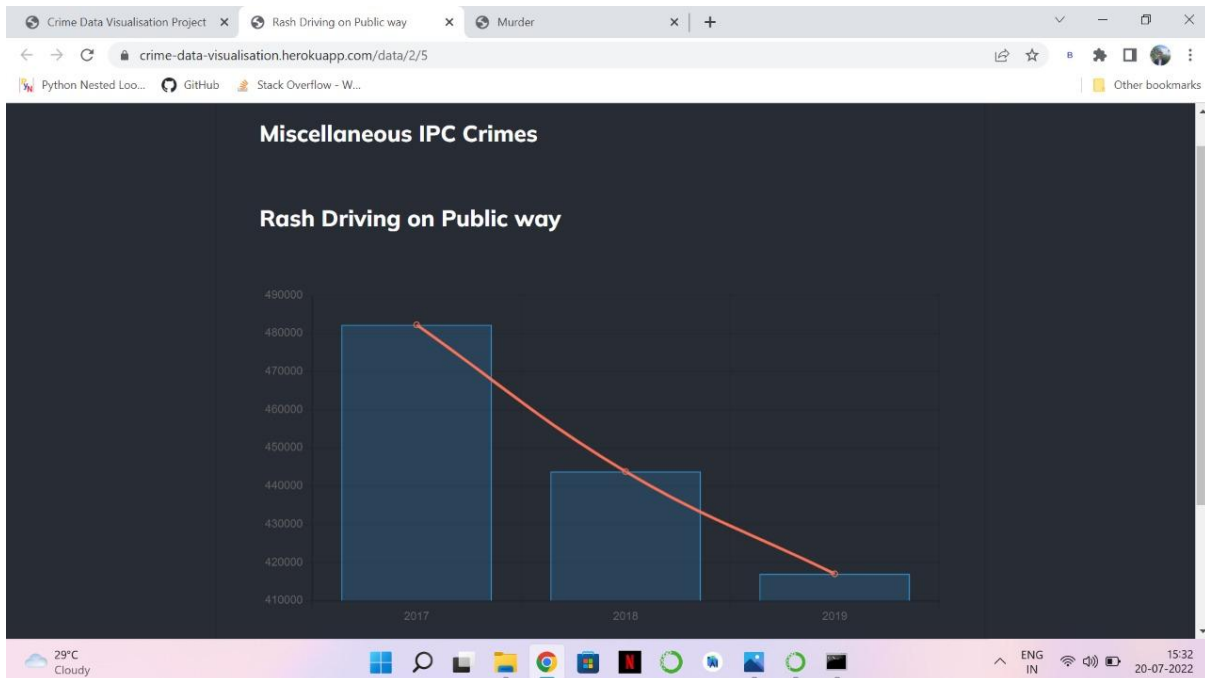
USECASE DIAGRAM

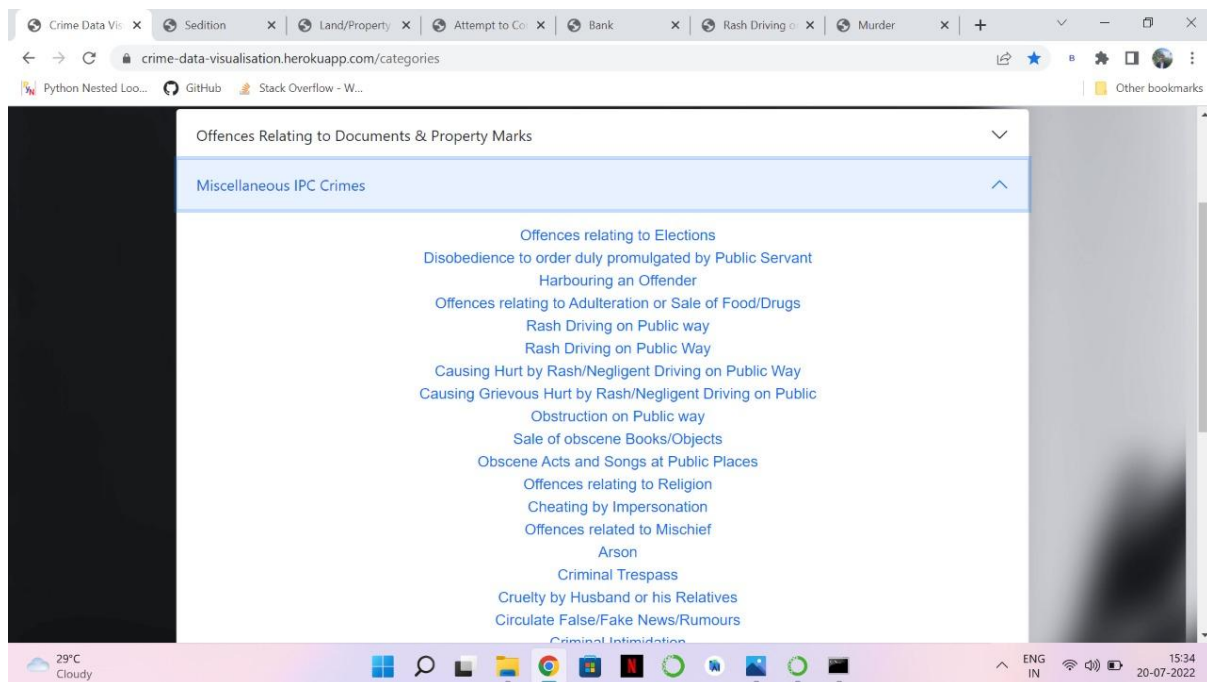
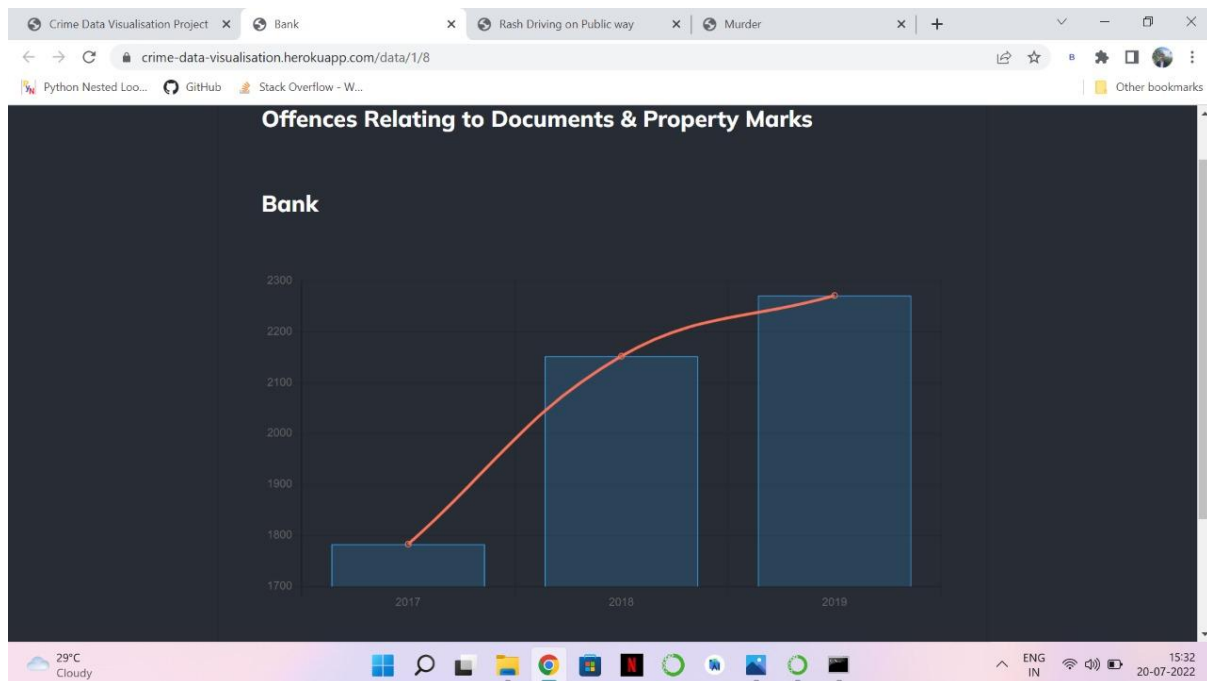


ACTIVITY DIAGRAM



6. SCREEN SHOTS







7. SYSTEM TESTING:

Integration:

Integration in software engineering means combining software parts (so-called subsystems) into one system. Integrated systems demonstrate better performance comparing to the sum of all independent ones. What is more important, a holistic application delivers greater functional significance. It's more convenient to distribute and use. Integrating is trending as never before these days. More and more mergeable systems appear on the Internet. Therefore, connecting with them is vital.

Integration testing is the process of testing the interface between two software units or module. It's focus on determining the correctness of the interface. The purpose of the integration testing is to expose faults in the interaction between integrated units. Once all the modules have been unit tested, integration testing is performed.

Integration test approaches –

There are four types of integration testing approaches. Those approaches are the following:

1. Big-Bang Integration Testing –

It is the simplest integration testing approach, where all the modules are combining and verifying the functionality after the completion of individual module testing. In simple words, all the modules of the system are simply put together and tested. This approach is practicable only for very small systems. If once an error is found during the integration testing, it is very difficult to localize the error as the error may potentially belong to any of the modules being integrated. So, debugging errors reported during big bang integration testing are very expensive to fix.

2. Bottom-Up Integration Testing –

In bottom-up testing, each module at lower levels is tested with higher modules until all modules are tested. The primary purpose of this integration testing is, each subsystem is to test the interfaces among various modules making up the subsystem. This integration testing uses test drivers to drive and pass appropriate data to the lower level modules.

Advantages:

- In bottom-up testing, no stubs are required.
- A principle advantage of this integration testing is that several disjoint subsystems can be tested simultaneously.

Disadvantages:

- Driver modules must be produced.
- In this testing, the complexity that occurs when the system is made up of a large number of small subsystem.

3. Top-Down Integration Testing –

Top-down integration testing technique used in order to simulate the behaviour of the lower-level modules that are not yet integrated. In this integration testing, testing takes place from top to bottom. First high-level modules are tested and then low-level modules and finally integrating the low-level modules to a high level to ensure the system is working as intended.

Advantages:

- Separately debugged module.
- Few or no drivers needed.
- It is more stable and accurate at the aggregate level.

Disadvantages:

- Needs many Stubs.
- Modules at lower level are tested inadequately.

4. Mixed Integration Testing –

A mixed integration testing is also called sandwiched integration testing. A mixed integration testing follows a combination of top down and bottom-up testing approaches. In top-down approach, testing can start only after the top-level module have been coded and unit tested. In bottom-up approach, testing can start only after the bottom level modules are ready. This sandwich or mixed approach overcomes this shortcoming of the top-down and bottom-up approaches. A mixed integration testing is also called sandwiched integration testing.

Advantages:

- Mixed approach is useful for very large projects having several sub projects.
- This Sandwich approach overcomes this shortcoming of the top-down and bottom-up approaches.

Disadvantages:

- For mixed integration testing, require very high cost because one part has Top-down approach while another part has bottom-up approach.
- This integration testing cannot be used for smaller system with huge interdependence between different modules.

Testing:

SYSTEM TEST

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.

8. CONCLUSION

In this report, a progression of best-in-class enormous information investigation and representation strategies were used to dissect wrongdoing large information from three US urban areas, which permitted us to distinguish designs and acquiring patterns. By investigating the Prophet model, a neural organization model, and the profound learning calculation LSTM, we found that both the Prophet model and the LSTM calculation perform better compared to traditional neural organization models. We likewise figured out the ideal time for the preparation test to be 3 years, to accomplish the best forecast of patterns regarding the RMSE and spearman relationship. Ideal boundaries for the Prophet and the LSTM models are additionally decided. Extra outcomes clarified before will give new bits of knowledge into wrongdoing patterns and will help both police offices and law implementation organizations in their dynamic. In the future, we intend to finish our on-going stage for conventional huge information investigation which will be fit for preparing different sorts of information for a wide scope of utilizations. We additionally plan to join multivariate perception chart mining procedures and fine-grained spatial examination to uncover more expected examples and patterns inside these datasets. Also, we mean to direct more sensible contextual investigations to additionally assess the viability and adaptability of the various models in our framework.

9.REFERENCES

- [1] Niu K, Zhang H, Zhou T, Cheng C and Wang C 2019 A novel spatio-temporal model for cityscale traffic speed prediction. *IEEE Access*, vol 7, 30050-57.
- [2] Peral J, Ferrandez A, Mora H, Gil D and Kauffmann E 2019 A review of the analytics techniques for an efficient management of online forums: An architecture proposal. *IEEE Access*, vol 7, 12220-40.
- [3] A. Haldorai and A. Ramu, Security and channel noise management in cognitive radio networks, *Computers & Electrical Engineering*, vol. 87, p. 106784, Oct. 2020. doi:10.1016/j.compeleceng.2020.106784
- [4] A. Haldorai and A. Ramu, Canonical Correlation Analysis Based Hyper Basis Feedforward Neural Network Classification for Urban Sustainability, *Neural Processing Letters*, Aug. 2020. doi:10.1007/s11063-020-10327-3
- [5] Wang Z, Ren J, Zhang D, Sun M and Jiang J 2018 A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. *Neurocomputing*, 287, 68-83.
- [6] Yan Y, Ren J, Zhao H, Sun G, Wang Z, Zheng J and Soraghan J 2018 Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. *Cognitive Computation*, 10 (1), 94-104.
- [7] Dai J, Song H, Sheng G and Jiang X 2018 May LSTM networks for the trend prediction of gases dissolved in power transformer insulation oil. In 2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM) pp. 666-669 IEEE.