A Project report on

MACHINE LEARNING BASED ANALYSIS OF CRYPTO CURRENCY MARKET FINANCIAL RISK MANAGEMENT

Submitted in partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

By

S. ESWARI	**(204G1A0530)**
B. AMRUTHA	**(204G1A0509)**
J. AISWARYA	**(204G1A0505)**
S. DIYA FARNAAZ	**(204G1A0529)**
V. MANOHAR	**(204G1A0553)**

Under the Guidance of

Ms. G. Lokeswari M.Tech., (Ph.D)

Assistant Professor

Department of Computer Science & Engineering

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

Rotarypuram Village, B K Samudram Mandal, Ananthapuramu - 515701

2023-2024

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

(Affiliated to JNTUA, Accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE)

Rotarypuram Village, BK Samudram Mandal, Ananthapuramu-515701

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Certificate

This is to certify that the project report entitled **MACHINE LEARNING-BASED ANALYSIS OF CRYPTO CURRENCY MARKET FINANCIAL RISK MANAGEMENT** is the bonafide work carried out by **S. Eswari, B. Amrutha, J. Aiswarya, S. Diya Farnaaz, V. Manohar** bearing Roll Number **204G1A0530, 204G1A0509, 204G1A0505, 204G1A0529, 204G1A0553** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering** during the academic year 2023-2024.

Project Guide

Ms. G. Lokeswari M.Tech., (Ph.D)

Assistant Professor

Head of the Department

Mr. P. Veera Prakash M.Tech., (Ph.D)

Assistant Professor

Date:

Place: Rotarypuram

EXTERNAL EXAMINER

DECLARATION

We, Ms. S. Eswari with reg no: 204G1A0530, Ms. B. Amrutha with reg no: 204G1A0509, Ms. J. Aishwarya with reg no: 204G1A0505, Ms. S. Diya Farnaaz with reg no: 204G1A0529, Mr. V. Manohar with reg no: 204G1A0553 students of SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY, Rotarypuram, hereby declare that the dissertation entitled **"MACHINE LEARNING BASED ANALYSIS OF CRYPTO CURRENCY MARKET FINANCIAL RISK MANAGEMENT"** embodies the report of our project work carried out by us during IV year Bachelor of Technology under the guidance of **Ms. G. Lokeswari,** Assistant Professor, Department of CSE, and this work has been submitted for the partial fulfillment of the requirements for the award of the Bachelor of Technology degree.

The results embodied in this project have not been submitted to any other University of Institute for the award of any Degree or Diploma.

S. ESWARI	Reg no: 204G1A0530
B. AMRUTHA	Reg no: 204G1A0509
J. AISWARYA	Reg no: 204G1A0505
S. DIYA FARNAAZ	Reg no: 204G1A0529
V. MANOHAR	Reg no: 204G1A0553

VISION & MISSION OF THE INSTITUTION

Vision:

To become a premier Educational Institution in India offering the best teaching and learning environment for our students that will enable them to become complete individuals with professional competency, human touch, ethical values, service motto, and a strong sense of responsibility towards environment and society at large.

Mission:

- Continually enhance the quality of physical infrastructure and human resources to evolvein to a center of excellence in engineering education.

- Provide comprehensive learning experiences that are conducive for the students to acquire professional competences, ethical values, life-long learning abilities and understanding of the technology, environment and society.

- Strengthen industry institute interactions to enable the students work on realistic problems and acquire the ability to face the ever-changing requirements of the industry.

- Continually enhance the quality of the relationship between students and faculty which is a key to the development of an exciting and rewarding learning environment in the college.

VISION & MISSION OF THE DEPARTMENT OF CSE

Vision:

To evolve as a leading department by offering best comprehensive teaching and learning practices for students to be self-competent technocrats with professional ethics and social responsibilities.

Mission:

DM 1: Continuous enhancement of the teaching-learning practices to gain profound knowledge in theoretical & practical aspects of computer science applications.

DM 2: Administer training on emerging technologies and motivate the students to inculcate self-learning abilities, ethical values and social consciousness to become competent professionals.

DM 3: Perpetual elevation of Industry-Institute interactions to facilitate the students towork on real-time problems to serve the needs of the society.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we have now the opportunity to express our gratitude for all of them.

It is with immense pleasure that we would like to express our indebted gratitude to our Guide **Ms. G. Lokeswari, Assistant Professor, Computer Science & Engineering**, who has guided us a lot and encouraged us in every step of the project work. We thank him for the stimulating guidance, constant encouragement and constructive criticism which have made possible to bring out this project work.

We express our deep-felt gratitude to **Mr. C. Lakshminatha Reddy, Assistant Professor** and **Mr. M. Narasimhulu, Assistant Professor,** Project Coordinators for their valuable guidance and unstinting encouragement enabled us to accomplish our project successfully in time.

We are very much thankful to **Mr. P. Veera Prakash, Assistant Professor & Head of the Department, Computer Science & Engineering** for his kind support and for providing necessary facilities to carry out the work.

We wish to convey our special thanks to **Dr. G. Balakrishna, Principal** of **Srinivasa Ramanujan Institute of Technology** for giving the required information in doing our project work. Not to forget, We thank all other faculty and non-teaching staff, and our friends who had directly or indirectly helped and supported us in completing our project in time.

We also express our sincere thanks to the Management for providing excellent facilities.

Finally, we wish to convey our gratitude to our families who fostered all the requirements and facilities that we need.

Project Associates

204G1A0530

204G5A0509

204G1A0505

204G1A0529

204G1A0553

ABSTRACT

Crypto currency is a well-known financial state in the globe, posing a variety of dangers that have an impact on the intrinsic risk assessment of risk auditors. Since its inception, the rise of crypto currenci.es has presented financial institutions with a wide range of risks in terms of money laundering. In the institution of financial supports such as anti-money laundering, banks, and bank secrecy, continue as a risk specialist, bank manager, and compliance officer who has a provocation for the connected transaction through crypto currencies and the users who conceal the illicit funds. In this study, the crypto currency framework was subjected to Hierarchical Risk Parity and unsupervised machine learning. The professional accounting procedure in terms of the inherent risk associated with bit-coin. The professional crypto currency experience in transaction cause the lower risk comparing the less experienced one. The Hierarchical Risk Parity gives the better output in term of returning the adjusted risk tail to get the better risk management result. The result section shows the proposed model is robust to various intervals which are re-balanced and the co-variance window estimation.

Keywords: *Decision tree, Random Forest, MLP Classifier, Adaboost, Extra tree Classifier, ML techniques.*

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

EPU	Economic Policy Uncertainty
RAM	Random Access Memory
MLP	Multilayer Perceptron
ANN	Artificial Neural Network
SRS	Software Requirements Specification
DLL	Dynamic Load Library
DRAC	Dynamic Random Access Memory
SDD	Solid State Drive
IDE	Integrated Development Environment
WSGI	Web Server Gateway Interface
UML	Unified Modelling Language
PCA	Principal Component Analysis
HRP	Hierarchical Risk Parity
RL	Reinforcement Learning

CHAPTER 1

INTRODUCTION

Financial market is one of the complex systems that the definition of complexity didn't get accepted from universities and this cause the agreement in term of interacting the elements of complex systems together. Complex system modeling is similar to daunting task which the structure of this system organized based on hierarchical manner that collected their own subsystems [1]–[3]. This resources extracted by the name of hierarchical models. Unfortunately, in the process of portfolio construction there is a hug challenge regarding the lack of correlation matrix in hierarchical structure. This issue worsen the matrices for large covariance. In recent decades, around 2500 type of cryptocurrencies which contains the 252.5 trillion dollar of trading in this market [4]–[6]. The cryptocurrency reverberation transpire in, out of order environment [7]–[10]. Even news publishers had more interest and closer attention to the price changes and the large remote of actions to the soar unmitigated. Rules set up is for investors protecting and try to stop the money laundry. Similarly, stop the crowd for the fiat currency. Regarding all the mentioned good wills, implementation and theories shows the dedicated movement of price of cryptocurrency market. Lahre et al. Markowitz optimization with the high ratio.

Walid et al. [15] proposed the relationship between cryptocurrencies based on the highest frequency. The presented system gives the output of useful marketing insights and gives the allowance to the agent to improve the system stability. Platanakis et al. [16], demonstrates the estimation error in term of return estimation rather than naively diversified (1/N) strategy. Similarly, they used [17] the model of Black Litterman based on the variance constraints to support the sophisticated portfolio technique for estimation control of the simple methods to manage the cryptocurrency. Saba et al. [18] applied the wavelet-based analysis for cryptocurrency multi-scale dynamic interdependence between the liquid cryptocurrencies to count the traders and investors heterogeneous behavior. Corbet et al. [19] compare the different rules of trading in term of average-oscillator to breakout the range of trading strategies. Based on the reports of cryptocurrency related audit considerations abuilding the general awareness for the intrinsic risks of the ecosystem of digital assets recommended.

1.1 Problem Statement

The problem statement of machine learning-based analysis of cryptocurrency market financial risk management is to address the challenge of accurately assessing and predicting the volatile nature of cryptocurrency markets, including price fluctuations, market manipulation, and regulatory uncertainties. The aim is to develop robust models that can provide timely and reliable risk assessments, aiding investors and financial institutions in making informed decisions and managing their exposure to financial risks in the cryptocurrency market.

1.2 Objectives

The main objective of this research is to classify machine learning-based analysis of crypto currency market financial risk management.

1. Gather a relevant dataset, cleaning and filtering it meticulously.
2. Train your chosen algorithm rigorously and evaluate its accuracy meticulously.
3. Lastly, we aim to determine whether there is a risk or not based on the specified criteria.

1.3 Scope of the Project

The following are the boundaries that have established in the proposed system which defines scope.

1. Filter the data set by eliminating null values and missing values.
2. Algorithms that we used has high accuracy
3. Prediction of risk found or not.

CHAPTER 2

LITERATURE SURVEY

[1]. C. Y. Kim and K. Lee, ''Risk management to cryptocurrency exchange and investors guidelines to prevent potential threats,'' in Proc. Int. Conf. Platform Technol. Service (PlatCon), Jan. 2018, pp. 1–6.

Investment and interest in cryptocurrency is rapidly growing. The price of each bitcoin, in particular, has exceeded 10,000 dollars as of November 2017, so we do not know how long the uptrend will continue. Although blockchain technology is more open and security oriented than conventional currency issuing methods, it is relatively ineffective in terms of distribution and management of cryptocurrency.

The most common way to get cryptocurrency is trading through exchange and mining, which novices sometimes invest in without sufficient knowledge.Management vulnerability of investors and management plan. Server management plan and personal action tips will be provided.

[2]. I. U. Haq, A. Maneengam, S. Chupradit, W. Suksatan, and C. Huo, ''Economic policy uncertainty and cryptocurrency market as a risk management avenue: A systematic review,'' Risks, vol. 9, no. 9, p. 163, Sep. 2021.

The role of the cryptocurrency market as a risk management avenue has got the attention of researchers. However, it is an immature asset class and requires gaps in current literature for future research directions. This research provides a systematic review of the vast range empirical literature based on the cryptocurrency market as a risk management avenue against economic policy uncertainty (EPU). Additionally, heterogeneous EPU requires heterogeneous solutions to deal with stock market volatility and economic policy uncertainty in different economies.

Likewise, the divergent protocol and administration of currencies in the crypto market consequently vicissitudes the hedging and diversification performance against each economy. Many research lines can benefit investors, policymakers, fund managers, or portfolio managers.

[3]. J. Gold and S. D. Palley, ‘‘Protecting cryptocurrency assets,’’ Risk Manage., vol. 68, no. 3, pp. 12–13, 2021

According to the advisory, companies the exploit impacts should immediately contact law enforcement, and take a number of technical risk mitigation steps. These include generating new keys for cryptocurrency wallets, and/or moving funds to new wallets; using hardware wallets to keep the private keys in a separate, secured storage area; introducing two-factor authentication as an extra layer of verification; removing impacted hosts from the network and changing all passwords to any accounts associated with impacted hosts; and updating and installing patches for all software and hardware, including any antivirus software, host-based intrusion detection software and firewall firmware.

[4]. I. Barkai, T. Shushi, and R. Yosef, ‘‘A cryptocurrency risk–return analysis for bull and bear regimes,’’ J. Alternative Investments, vol. 24, no. 1, pp. 95–118, Jun. 2021.

In this article, the authors develop a new analytical lens through which to examine the risk–return profiles of bitcoin, Litecoin, ripple, and Ethereum. Their focus is to understand better the price behavior of individual cryptocurrencies and their influence on one another. To achieve this, they segment each cryptocurrency’s time series of returns into disparate bull and bear regimes. They then examine the nature and extent of overlap between these regimes and whether they change over time. They also collect and plot several indicative distributed-denial-of-service attacks against the time series to investigate their possible impact on regime change episodes. Their findings shed light on previously unexplored systemic risk indicators within the crypto market as a whole and on the relationship between specific cryptocurrency pairs.

These findings enhance the risk management toolkit for investors by revealing potential price behavior contagion patterns between cryptocurrencies pertinent to blended portfolio management.

CHAPTER 3

METHODOLOGY

3.1 Machine Learning:

Machine learning, a subset of artificial intelligence, enables computers to learn from data and improve without explicit programming. It includes supervised, unsupervised, and reinforcement learning. Supervised learning uses labeled data for predictions, unsupervised learning discovers patterns in unlabeled data, and reinforcement learning guides agents' behavior through environmental interactions and rewards. Deep learning employs multi-layer neural networks for complex pattern recognition. Feature engineering optimizes model performance by selecting and transforming relevant data features. Addressing challenges like overfitting and underfitting ensures robust model development. Machine learning transforms industries like healthcare, finance, transportation, and entertainment with predictive insights and data-driven solutions.

Basic Terminology:

- **Dataset:** A collection of data examples used for solving a problem, containing relevant features.
- **Features:** Key data elements that aid in understanding a problem and are input to Machine Learning algorithms for learning.
- **Model:** The learned representation of a phenomenon by a Machine Learning algorithm, derived from training data. It serves as the algorithm's output after training.

Types of Machine Learning:

There are multiple forms of Machine Learning; supervised, unsupervised, semi supervised and reinforcement learning.

Fig. 3.1: Types of Machine Learning

Supervised Learning: Supervised learning is a popular machine learning paradigm where an algorithm learns from labeled data to predict labels for new, unseen examples based on the relationships it has learned from the training data. When fully-trained, the supervised learning algorithm will be able to observe a new, never before-seen example and predict a good label for it.

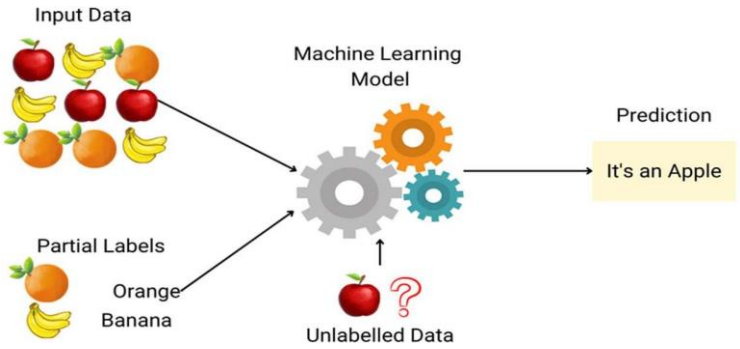

Fig. 3.2: Process of Supervised Learning

Unsupervised learning: Unsupervised learning doesn't use labeled data. Instead, algorithms analyze data's inherent properties to group, cluster, or organize it. This data-driven approach allows the algorithm to uncover patterns and structures without predefined labels, producing results based solely on the data's format.

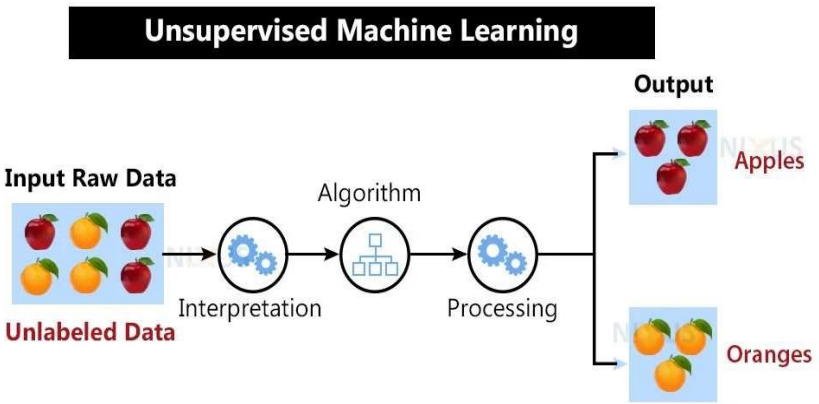

Fig. 3.3: Process of Unsupervised Learning

Reinforcement learning: Reinforcement learning differs significantly from supervised and unsupervised learning, focusing on behavior-driven learning inspired by neuroscience and psychology. In this paradigm, an agent interacts with an environment by taking actions that influence it. The environment provides feedback to the agent through updated states and rewards, creating a feedback loop that guides the agent's learning and decision-making process.

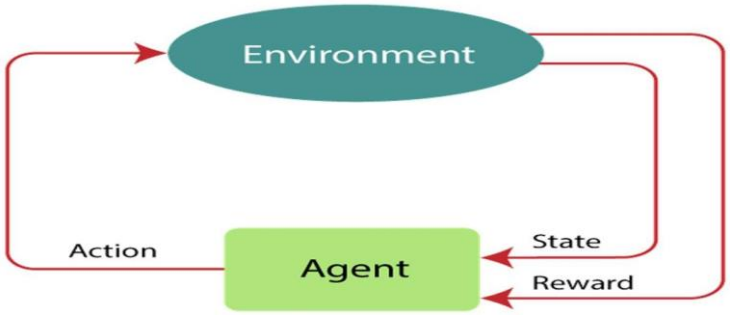

Fig. 3.4: Reinforcement Learning

3.2 Algorithm Used

Decision Tree:

Decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking.

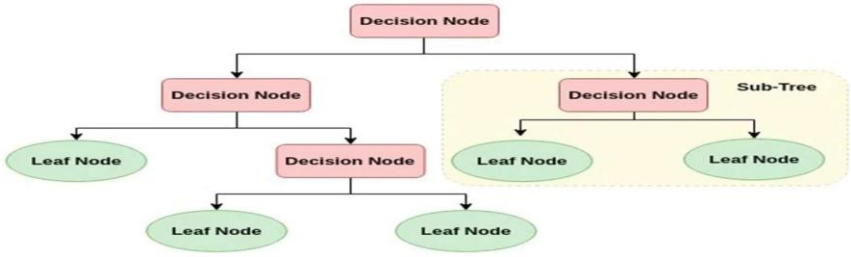

Fig. 3.5: Decision Tree Algorithm

The basic idea behind any decision tree algorithm is as follows:

- Select best attribute using Attribute Selection Measures (ASM) for splitting.
- Designate attribute as decision node and divide dataset into subsets.
- Recursively build tree by repeating above steps for each subset.
- Terminate when subsets have uniform attributes, no more attributes, or no instances.

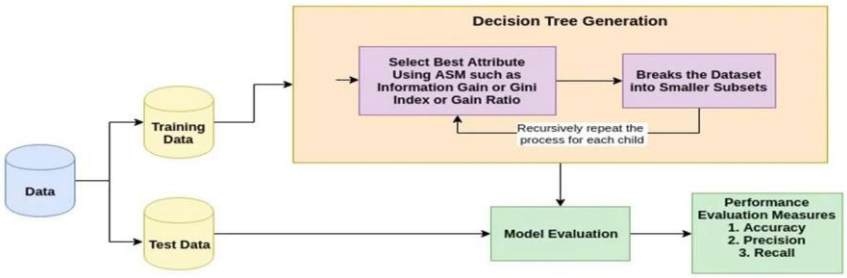

Fig. 3.6: Decision Tree Algorithm Implementation

Random Forest Classifier:

Random forest is an ensemble learning method combining multiple decision trees for regression and classification. Trained with bagging, it improves accuracy by averaging predictions across the tree ensemble.

The random forest algorithm aggregates predictions from multiple decision trees by averaging their outputs, enhancing precision with more trees. It overcomes decision tree limitations by reducing overfitting and improving accuracy without needing extensive configurations in packages like Scikit-learn.

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

Decision trees serve as the foundational components of the random forest algorithm, presenting a tree-like structure in decision support. Understanding decision

trees is key to grasping the mechanics of random forests. Nodes in the tree symbolize attributes used for predictions, with decision nodes linking to leaves. The tree encompasses three node types, as illustrated in the following diagram.

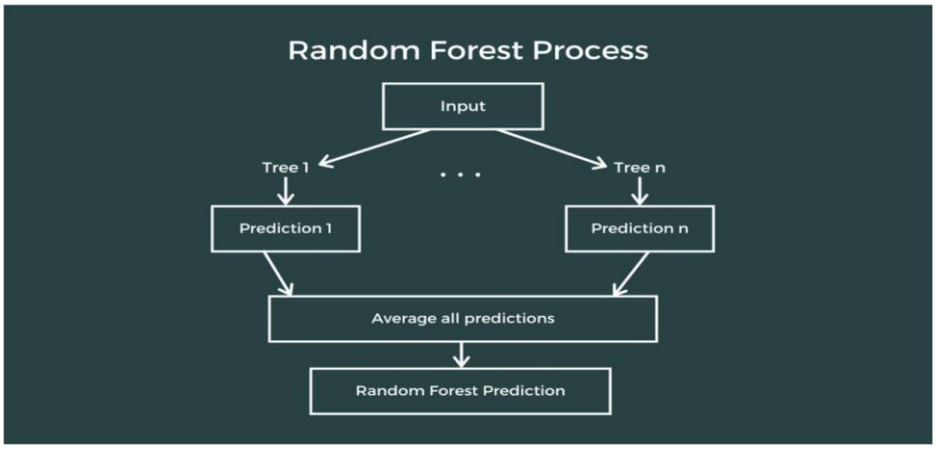

Fig. 3.7: Random Forest Algorithm

Entropy quantifies uncertainty, while information gain measures the reduction in uncertainty of a target variable using independent variables or features. Information gain assesses how features inform the prediction of a target class. It's calculated by subtracting the conditional entropy of the target variable given a feature (Y|X) from the entropy of the target variable (Y).

Information gain guides decision tree training by reducing uncertainty; higher gain indicates more uncertainty removed. It's crucial for branch splitting in tree construction. For instance, predicting a customer's phone purchase decision can be visualized with a Decision tree. Features like serve, internal space, as a decision nodes.

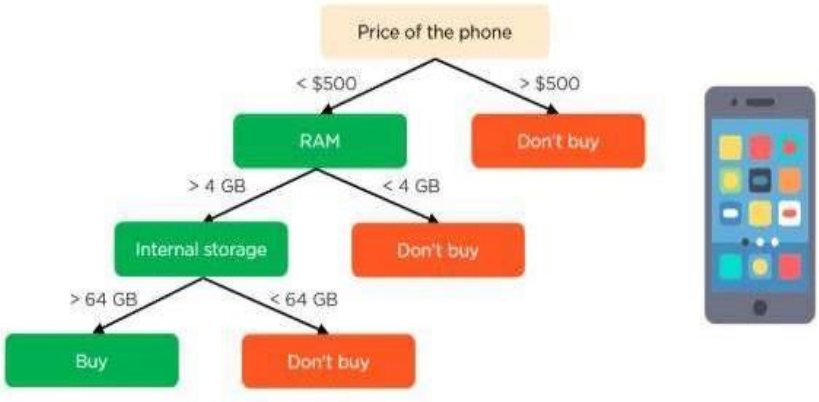

Fig. 3.8: Random Forest Algorithm Implementation

In random forests, unlike decision trees, root and segregating nodes are randomized. Using bagging, multiple data samples train individual decision trees. These trees produce varied outputs based on the data. The final prediction is determined by majority voting: the most frequently predicted outcome across trees is chosen. For instance, if three trees predict 'buying' and one 'not buying', the final prediction is 'buying', suggesting the customer will purchase the phone.

AdaBoost:

AdaBoost, or Adaptive Boosting, is an Ensemble Method that iteratively adjusts instance weights, emphasizing misclassified samples. This boosting technique aims to minimize bias and variance in supervised learning by sequentially growing learners. Unlike traditional boosting, AdaBoost prioritizes misclassified records from the previous model, refining weak learners into stronger ones. It generates 'n' decision trees in training, where each subsequent tree focuses on the misclassifications of its predecessor. Notably, all boosting methods allow record repetition during training.

Since we now know the boosting principle, it will be easy to understand the AdaBoost algorithm. Let's dive into AdaBoost's working. When the random forest is used, the algorithm makes an 'n' number of trees. It makes proper trees that consist of a start node with several leaf nodes. Some trees might be bigger than others, but there is no fixed depth in a random forest. With AdaBoost, however, the algorithm only makes a node with two leaves, known as Stump.

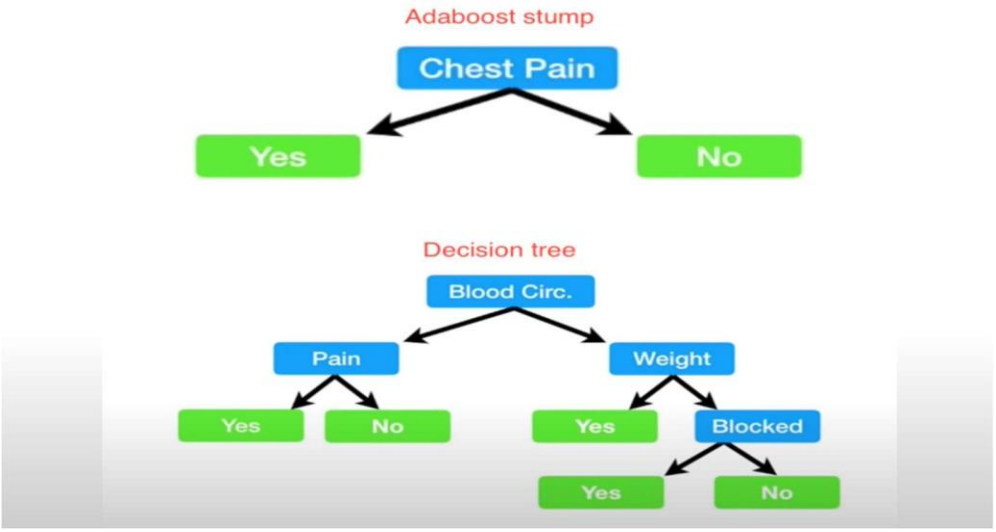

Fig. 3.9: AdaBoost Algorithm

The figure here represents the stump. It can be seen clearly that it has only one node with two leaves. These stumps are weak learners and boosting techniques prefer this. The order of stumps is very important in AdaBoost. The error of the first stump influences how other stumps are made. Let's understand this with an example.

The output is in categorical form, here in the form of Yes or No. All these records will be assigned a sample weight. The formula used for this is **'W=1/N'** where N is the number of records. In this dataset, there are only 5 records, so the sample weight becomes 1/5 initially. Every record gets the same weight. In this case, it's 1/5.

Multilayer Perceptron:

A multilayer perceptron (MLP) is a type of feedforward artificial neural network (ANN). While the term "MLP" can be used broadly to describe any feedforward ANN, it specifically refers to networks with multiple layers of perceptrons, often termed "vanilla" neural networks when having a single hidden layer. An MLP comprises at least three layers: input, hidden, and output. Each non-input node, or neuron, employs a nonlinear activation function. MLPs use the backpropagation algorithm for supervised learning, allowing them to handle non-linearly separable data due to their multiple layers and activation functions, distinguishing them from linear perceptrons.

The term "multilayer perceptron" does not refer to a single perceptron that has multiple layers. Rather, it contains many perceptron's that are organized into layers. An alternative is "multilayer perceptron network".

Moreover, MLP "perceptron's" are not perceptron's in the strictest possible sense. True perceptron's are formally a special case of artificial neurons that use a threshold activation function such as the Heaviside step function. MLP perceptron's can employ arbitrary activation functions. A true perceptron performs binary classification, an MLP neuron is free to either perform classification or regression, depending upon its activation

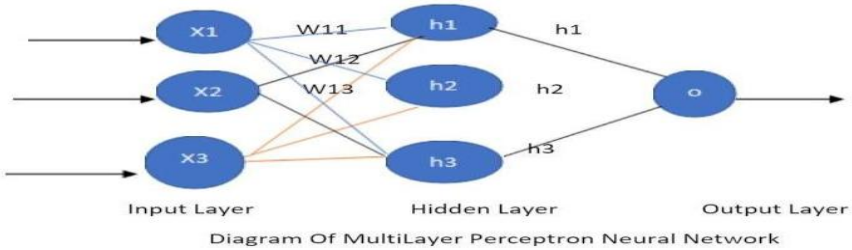

Fig. 3.10: Multi-Layer Perceptron Algorithm

The term "multilayer perceptron" later was applied without respect to nature of the nodes/layers, which can be composed of arbitrarily defined artificial neurons, and not perceptron's specifically. This interpretation avoids the loosening of the definition of "perceptron" to mean an artificial neuron in general.

MLPs are universal function approximators as shown by Benko's theorem,[4] so they can be used to create mathematical models by regression analysis. As classification is a particular case of regression when the response variable is categorical, MLPs make good classifier algorithms.

Extra Tree Classifier:

Extremely Randomized Trees Classifier (Extra Trees Classifier) is a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output its classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest.

In the Extra Trees Classifier, each decision tree is built using the original training sample. At test nodes, trees receive a random subset of k features, choosing the best split based on criteria like the Gini Index. This randomness fosters the creation of diverse, de-correlated trees. For feature selection, the Gini Importance of each feature is computed during forest construction, representing the reduction in decision criteria due to each feature. Features are then ranked by Gini Importance, allowing users to select the top features based on their preference.

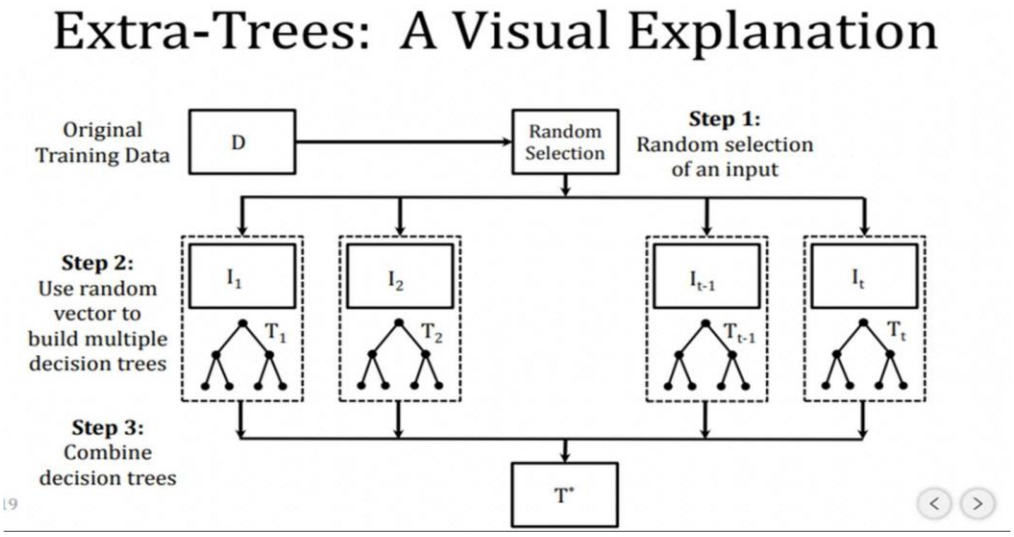

Fig. 3.11: Extra -Tree Classifier Algorithm

CHAPTER 4

SYSTEM REQUIREMENTS SPECIFICATIONS

A Software Requirements Specification (SRS) defines functional and nonfunctional requirements for a software project, including use cases. It clarifies project goals, aids planning, and helps identify limitations and risks early in development.

A Software Requirements Specification (SRS) provides a detailed description of a software system's functional and non-functional requirements, including user interactions through use cases. It ensures consistency in project development by capturing all necessary requirements. Clear understanding of the software system is vital for development, necessitating continuous communication with customers to gather requirements. A comprehensive SRS outlines interactions with modules, hardware, other programs, and user scenarios. It's crucial for testers to thoroughly understand the SRS to avoid errors in test cases and expected results.

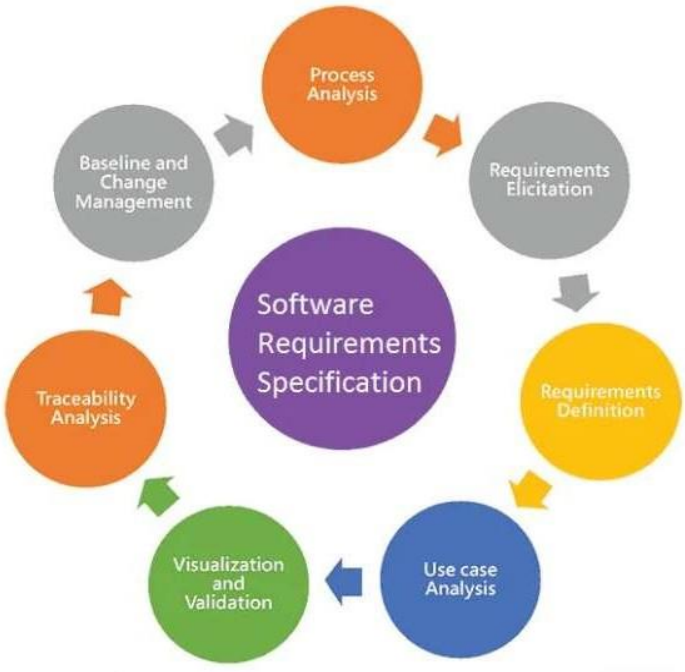

Fig. 4.1: Types of Requirements in SRS

4.1 Functional Requirements:

Functional requirements are specific demands from end-users outlining essential system features and capabilities. These requirements, integral to the development contract, define the inputs, operations, and expected outputs of the system, directly observable in the final product. Unlike non-functional requirements, which address performance and user experience, functional requirements focus on tangible functionalities users directly interact with.

4.2 Basic Requirements

1. Data collection: Data set collection starts with verifying and loading available data into CSV files, establishing a systematic foundation for analysis. Efficient systems for verification, extraction, and storage ensure data availability for informed decision-making.

2. Data Preprocessing: Before model training, data pre-processing ensures quality and compatibility by handling missing values, encoding categories, scaling features, and removing outliers. This enhances model accuracy and deepens data understanding.

3. Model training: After pre-processing, data is split into training and testing subsets training for algorithm training and testing for model performance evaluation. This ensures assessing the model's generalization capabilities on unseen data.

4. Model evaluation and testing: Diverse machine learning algorithms, like decision trees, SVMs, random forests, or neural networks, are used to predict student personality or adaptability levels. The aim is to construct an accurate model predicting the target variable from available features.

Application Requirements

1. Users can register and login to the application.
2. User details can be updated within the system.
3. Risk prediction algorithms provide accurate risk assessments.
4. Users input currency details to predict associated risks.
5. All application modules function correctly.
6. Users can access algorithm accuracy and log out from the system.

4.2 Non-Functional Requirements

Non-Functional Requirements (NFRs) define the quality attributes like responsiveness, usability, security, and portability that a software system must meet. They assess system performance and adherence to critical standards, essential for user satisfaction and system success. NFRs impose constraints on system design across agile backlogs, such as requiring a site to load within 3 seconds with over 10,000 simultaneous users. Unlike functional requirements that specify system behavior, NFRs focus on system properties like reliability, response time, and resource utilization.

Non-functional requirements stem from user needs, budget limitations, organizational policies, and interoperability demands with other systems. They encompass product, organizational, user, and operational requirements, among other external factors.

Benefits of Non-Functional Requirements:

- They ensure good user experience and ease of operating the software.
- They help in formulating security policy of the software system.

Requirements

1. Data Security and Accessibility
2. Application Adaptability and Responsiveness
3. User Experience and Intuitiveness
4. Performance, Speed, and Stability
5. Maintenance, Updates, and User Engagement
6. User-Friendliness and Satisfaction
7. Performance Speed and System Reliability

4.4 Python Libraries:

In programming, a library consists of precompiled codes for specific tasks, reducing redundancy. Python libraries, with related modules, streamline development by eliminating repetitive coding. They are vital for efficiency in Machine Learning, Data Science, and other fields. Libraries enhance functionality by providing reusable code and tools, making programming more efficient and convenient.

A Python library is a set of related modules providing reusable code for diverse programs, simplifying Python programming by reducing redundancy.

Working of Python Library

A Python library is a collection of modules designed for specific programming tasks, eliminating the need for repetitive code. In the MS Windows environment, library files often use the DLL extension, representing Dynamic Load Libraries. When linked to a program, the linker automatically integrates the library's functionalities, enhancing program efficiency and reducing code redundancy. Integrating libraries into Python programs streamlines development, making coding more efficient and versatile.

Python standard library

The Python Standard Library encompasses Python's syntax, semantics, and core modules, offering access to fundamental system functionalities like I/O. With over 200 core modules, it elevates Python to a high-level programming language. While most Python libraries are written in C, enhancing efficiency, other specialized libraries further simplify a programmer's tasks, augmenting Python's versatility and ease of use. Let's have a look at some of the commonly used libraries:

1. Pandas: Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.

2. NumPy: The name "NumPy" stands for "Numerical Python". It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data. It consists of in-built mathematical functions for easy computations. Even libraries like TensorFlow use NumPy internally to perform several operations on tensors. Array Interface is one of the key features of this library.

3. Flask: Flask is a lightweight web framework in Python, termed "micro" due to its minimalistic design without mandatory tools or libraries. Unlike frameworks with built-in functionalities, Flask relies on third-party extensions for features like database abstraction, form validation, and authentication. These extensions integrate seamlessly, enhancing Flask's capabilities as if they were native, making it flexible and adaptable to diverse web development needs.

4. Scikit-Learn: Scikit-Learn is a robust Python library for machine learning, offering a user-friendly interface and a wide array of algorithms. Integrated with libraries like NumPy, SciPy, and matplotlib, it streamlines data processing, model training, and

evaluation. Popular among data scientists, it supports various tasks like classification, regression, and clustering. Additionally, Scikit-Learn simplifies model selection, hyper parameter tuning, and performance assessment, making it essential for developing accurate and scalable machine learning solutions.

Use of Libraries in Python Program

Maintaining code modularity is crucial for efficient upkeep in extensive Python programs. Modules in Python encapsulate frequently used functions, eliminating code duplication and complexity across programs. By segmenting code into reusable modules, we streamline the development process and facilitate easy importation of functionalities when needed. Python's simple syntax further simplifies this by enabling straightforward module imports, ensuring quick access to interrelated modules within a library as required.

4.5 Hardware Requirements

The hardware requirements include the requirements specification of the physical computer resources for a system to work efficiently. The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. The Hardware Requirements are listed below:

System Processor	:	Intel I3
Hard Disk	:	500 GB
Ram	:	4 GB

1. Processor: A processor is an integrated electronic circuit that performs the calculations that run a computer. A processor performs arithmetical, logical, input/output (I/O) and other basic instructions that are passed from an operating

system (OS). Most other processes are dependent on the operations of a processor. A minimum 1 GHz processor should be used, although we would recommend S2GHz or more. A processor includes an arithmetical logic and control unit (CU), which measures capability in terms of the following:

- Ability to process instructions at a given time
- Maximum number of bits/instructions

The proposed system requires a 2.4 GHz processor or higher.

Fig. 4.2: Processor

2. Ethernet connection (LAN) OR a wireless adapter (Wi-Fi):

Wi-Fi is a radio technology used for wireless local area networking (WLAN) based on the IEEE 802.11 standards. It enables devices like computers, smartphones, TVs, and drones to connect wirelessly over a range of approximately 20 meters indoors and further outdoors. Devices can connect via a wireless access point or hotspot to access the Internet and communicate with Ethernet-connected devices. Hotspot coverage can vary from a single room to extensive areas, achieved by using overlapping access points.

Fig. 4.3: Ethernet Connection

3. Hard Drive: A hard drive is an electro-mechanical device that uses magnetic storage on rapidly rotating disks, called platters, to store and retrieve digital data. Paired with magnetic heads on a moving actuator arm, it reads and writes data randomly, allowing for non-sequential access. Unlike volatile storage, hard drives retain data even when powered off. For the proposed system, a minimum of 32 GB storage capacity is recommended.

Fig. 4.4: Hard Disk

4. Memory (RAM): Random-access memory (RAM) is a form of computer data storage that stores data and machine code currently being used. A random- access memory device allows data items to be read or written in almost the same amount of time irrespective of the physical location of data inside the memory. In today's technology, random-access memory takes the form of integrated chips. RAM is normally associated with volatile types of memory (such as DRAM modules), where stored information is lost if power is removed, although non- volatile RAM has also been developed. A minimum of 4 GB RAM is recommended for the proposed system.

Fig. 4.5: RAM

4.6 Software Requirements

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The requirements can be obvious or hidden, known or unknown, expected or unexpected from client‘s point of view.

H/W Configuration:

Operating system: Windows 7 or 7+

RAM: 8 GB

Hard disc or SSD: More than 500 GB

Processor: Intel 3rd generation or high or Ryzen with 8 GB Ram

S/W Configuration:

Software’s: Python 3.6 or high version

IDE: PyCharm.

Framework: Flask, pandas, NumPy and Scikit-Learn

1. PyCharm: PyCharm is the most popular IDE for Python, and includes great features such as excellent code completion and inspection with advanced debugger and support for web programming and various frameworks. The intelligent code editor provided by PyCharm enables programmers to write high quality Python code. The editor enables programmers to read code easily through color schemes, insert indents on new lines automatically.

PyCharm offers some of the best features to its users and developers in the following aspects:

- Code completion and inspection
- Support for web programming and frameworks such as Django and Flask

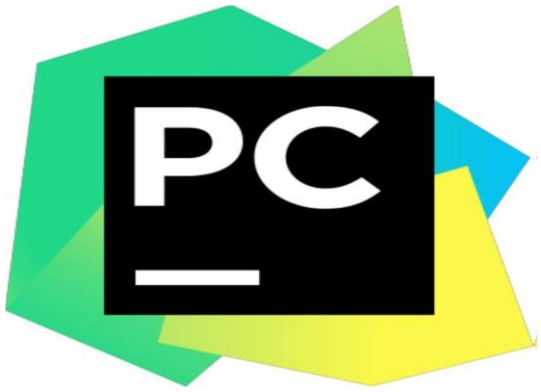

Fig. 4.6: PyCharm image

2. Python: Python is a high-level, object-oriented programming language favored for web and app development due to its dynamic semantics. Its simplicity and unique, readable syntax make it easy to learn and maintain, reducing development costs. Python's support for modules and packages promotes modular design and code reuse across different projects, facilitating collaborative work and overcoming language barriers within teams.

Fig. 4.7: Python Icon

3. Flask Framework: A Web Framework is a collection of libraries and modules that simplifies web application development by handling low-level details like protocols and thread management. Flask, a Python-based web framework developed by Armin Ronacher and the Pocco group, relies on the Werkzeug WSGI toolkit and Jinja2 template engine. WSGI serves as a standard interface between web servers and Python web applications, streamlining the development process.

Werkzeug is a WSGI toolkit, which implements requests, response objects, and other utility functions. This enables building a web framework on top of it. The Flask framework uses Werkzeug as one of its bases.

Fig. 4.8: Flask Python Logo

4. Visual Studio Code

Visual Studio Code, commonly known as VS Code, is a versatile and lightweight code editor developed by Microsoft. Designed for developers, it offers a plethora of features to enhance productivity and streamline coding processes. Its intuitive user interface, coupled with extensive customization options, makes it a favorite among programmers of all levels. VS Code supports a wide range of programming languages and frameworks, making it highly adaptable for various development tasks. Additionally, its built-in Git integration, debugging tools, and extensions marketplace further enrich its capabilities. With regular updates and a vibrant community, VS Code continues to evolve, setting the standard for modern code editors in the industry.

Fig. 4.9: Visual Studio Code

CHAPTER 5

SYSTEM ANALYSIS AND DESIGN

Systems development is a structured approach involving stages like planning, analysis, design, deployment, and maintenance. System analysis involves gathering and interpreting data to identify system problems and goals, enhancing efficiency and functionality. Meanwhile, system design plans a new or updated system by outlining its components or modules to meet specific requirements, optimizing how the system achieves its objectives based on a thorough understanding of the existing system.

5.1 UML Diagrams:

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

UML aims to be a universal language for modeling object-oriented software, consisting of a Meta-model and notation components. It serves as a standard for specifying, visualizing, constructing, and documenting software and business systems, incorporating best practices for modeling large and complex systems. Utilizing graphical notations, UML plays a crucial role in object-oriented software development and the broader software development process.

5.1.1 Use Case Diagram:

- A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis.

- Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.

- The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

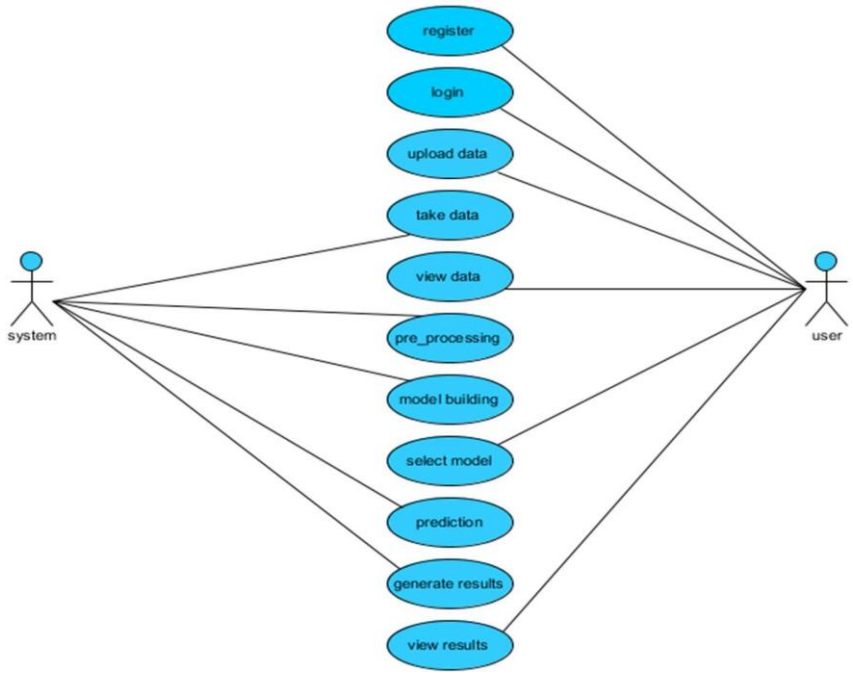

Fig. 5.1: Use Case Diagram

5.1.2 Class Diagram

A class diagram in UML depicts the system's classes, their attributes, methods, and relationships among classes, illustrating the information contained within each class.

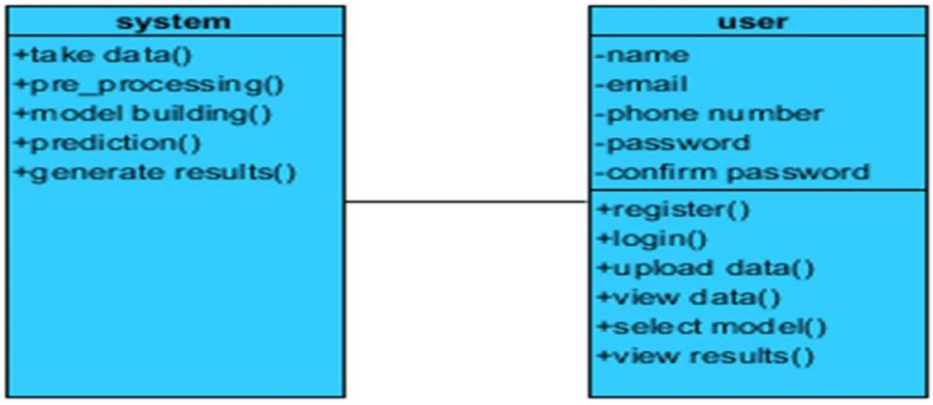

Fig. 5.2: Class Diagram

5.1.3 Sequence Diagram:

A sequence diagram in UML displays interactions between processes and their order, derived from Message Sequence Charts, also known as event diagrams or timing diagrams.

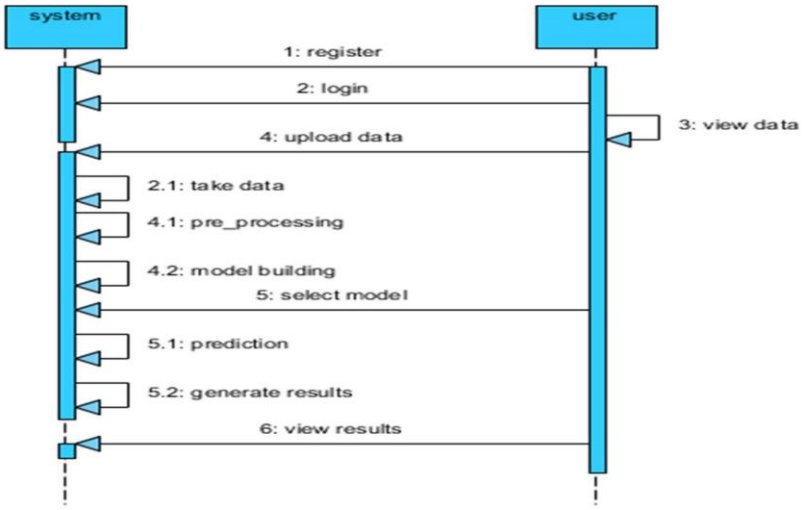

Fig. 5.3: Sequence Diagram

5.1.4 Collaboration Diagram:

Collaboration diagrams show method call sequences with a numbering technique, emphasizing object organization alongside the order of operations, unlike sequence diagrams, which focus solely on method sequences.

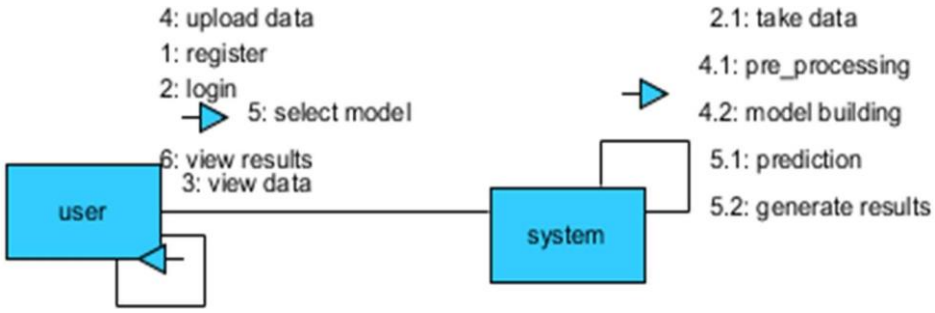

Fig. 5.4: Collaboration Diagram

5.1.5 Activity Diagram:

Activity diagrams visually represent workflows, detailing stepwise activities with options for choices, iterations, and concurrent actions. They provide a clear view of business and operational processes, illustrating the overall control flow within a system.

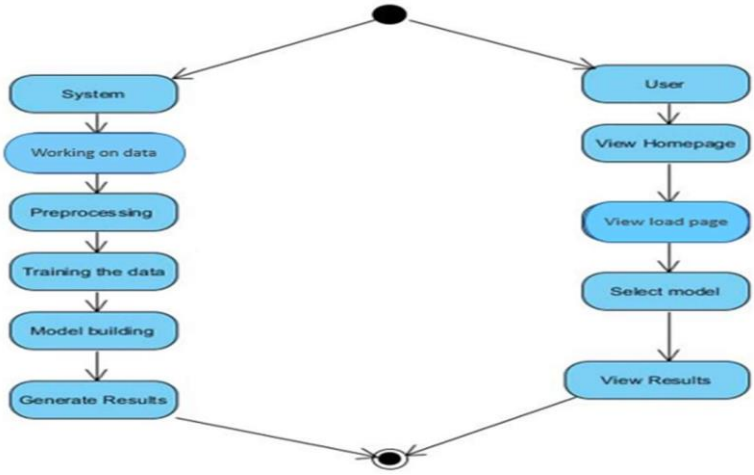

Fig. 5.5: Activity Diagram

5.2 System Architecture

In machine learning, system architecture is the blueprint for computational frameworks supporting model training, evaluation, and deployment. It covers hardware, software, and data flow design for efficient and scalable processing. This architecture dictates data flow, computation methods, and model integration, playing a vital role in achieving high performance and scalability in machine learning applications.

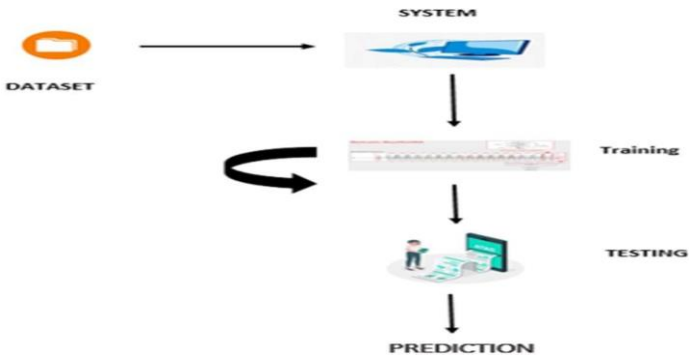

Fig. 5.6: System Architecture

5.3 Flowchart

Flowcharts in machine learning algorithms visually depict the sequential steps involved in the training and deployment process. They outline the stages of data preprocessing, feature engineering, model selection, training, evaluation, and deployment. Flowcharts serve as roadmaps for developers, aiding in understanding the algorithmic workflow and identifying potential bottlenecks or areas for optimization. These visual representations facilitate clear communication and documentation of complex machine learning pipelines, enhancing collaboration and efficiency in algorithm development.

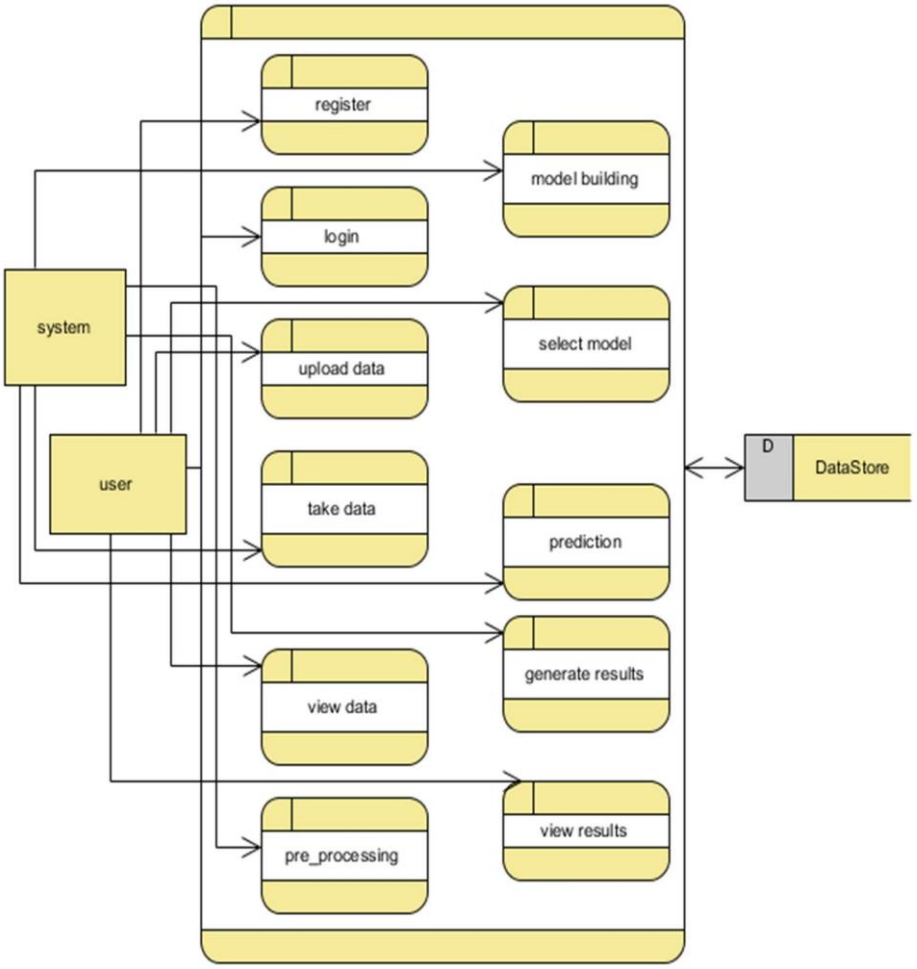

Fig. 5.7: Flowchart of the system

CHAPTER 6
IMPLEMENTATION

In this project we have mainly two objectives

Objective 1: - To improve the dataset's quality, we filtered out noise data, including null values and missing data, sourced from Kaggle. After preprocessing, we trained the dataset using machine learning techniques. Subsequently, we evaluated the dataset's accuracy using five models: Decision Tree, Random Forest, AdaBoost, MLP Classifier, and Extra Tree Classifier.

Objective 2: - Prediction of Risk.

Here we collect basic information from user like Currency ID, Name, percent change in 1 hour, 24 hours, 7 days, Available supply, Price_btc etc. Based on these values we predict whether risk found or not.

The two modules that make up the proposed system are the user registration and login, load data set, find accuracy and prediction.

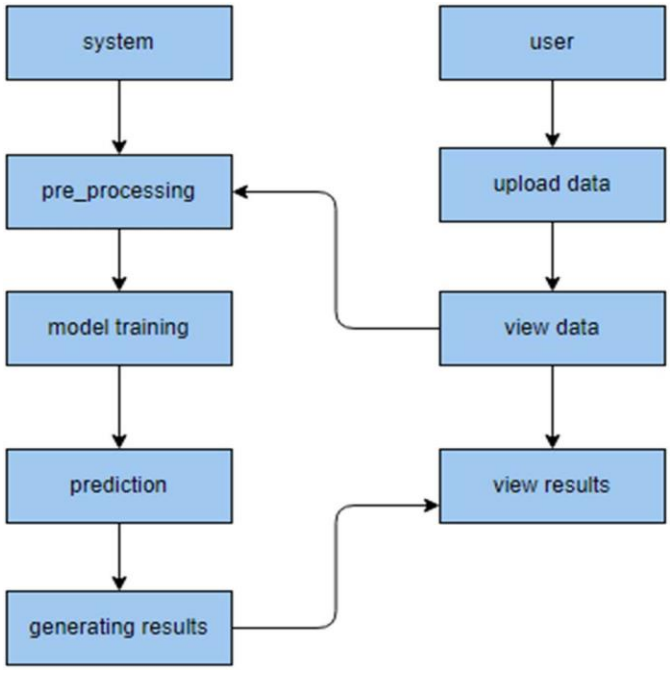

Fig. 6.1: Block diagram of Admin

Working Flow of the System

1. Users register by providing personal details including name, email, password, age, and contact number.
2. Upon successful registration, users log in using their email and password.
3. After login, users are directed to the homepage.
4. Users upload a CSV data file with noise-filtered data.
5. Users can view the uploaded data.
6. Users select one of the five available machine learning models: Decision Tree, Random Forest, MLP, AdaBoost, or Extra Tree Classifier.
7. The chosen model displays its accuracy based on the uploaded data.
8. Users input specific values to predict the presence or absence of risk, followed by viewing the prediction and model accuracy graph before logging out.

6.1 Datasets:

Datasets are vital for machine learning, serving as the basis for training and insights extraction. They vary in type and size, aiming to mirror real-world scenarios. The quality and diversity of datasets significantly influence model performance, underscoring the need for meticulous selection and preprocessing to enhance accuracy.

For risk prediction, we sourced data from Kaggle on various cryptocurrencies. Using machine learning methods, we initially cleaned the dataset by addressing null values, filling missing data, and removing unnecessary columns.

S/N	Unnamed: 0	24h_volume_usd	available_supply	id	last_updated	market_cap_usd	max_supply	name	percent_change_1h	percent_change
1	0	1262	453	128	69	1031	29	145	268	715
2	1	1260	719	455	68	1030	124	441	241	378
3	2	1259	455	130	93	1029	29	147	375	328
4	3	1261	939	664	86	1028	107	609	128	1076
5	4	1252	983	1005	56	1027	122	976	309	389
6	5	1251	311	331	57	1026	25	333	349	402
7	6	1256	640	723	57	1025	51	699	242	541
8	7	1248	451	132	97	1024	29	150	194	246
9	8	1257	432	789	59	1023	124	763	147	934
10	9	1239	976	236	94	1022	119	239	232	322
11	10	1255	723	457	71	1021	124	443	240	395
12	11	1230	956	816	74	1020	124	784	204	498
13	12	1253	868	442	85	1019	124	396	401	960
14	13	1243	662	817	73	1018	124	785	350	278
15	14	1254	967	1108	64	1017	124	1078	121	1007
16	15	1245	643	788	59	1016	124	760	557	1089

Fig. 6.2: Dataset Collection

6.2 Data Pre-Processing

Data pre-processing is vital in machine learning, refining raw data for algorithm use through stages like cleaning, integration, transformation, and reduction. Cleaning handles missing values and outliers, integration merges data sources, transformation adjusts data for algorithms, and reduction optimizes computational efficiency. Proper pre-processing ensures accurate and insightful machine learning analysis.

6.2.1 Data Cleaning:

Data cleaning is crucial in data pre-processing to handle real-world data imperfections that can hinder machine learning model performance. Methods like imputation fill missing values, noise reduction smooths irregularities, and outlier detection identifies anomalies, enhancing data quality. Clean data improves analysis accuracy, reliability, and minimizes erroneous conclusions, making it a vital step in effective machine learning.

6.2.2 Data Integration:

Data integration is essential for understanding market dynamics and investor behavior. Combining data from various sources like exchanges requires accurate matching of identifiers such as customer IDs to maintain data integrity. Techniques like record linkage and entity resolution are used to reconcile data across databases by comparing attributes and identifying related records. Leveraging advanced algorithms and domain expertise ensures effective integration, enabling more informed risk management strategies in crypto markets.

6.2.3 Data Reduction:

Data reduction techniques streamline large datasets to improve storage efficiency and lower computational costs. By reducing dimensionality through methods like feature selection and principal component analysis (PCA), essential information is retained while simplifying analysis, facilitating faster and more cost-effective ML algorithms**.**

6.3 Forget Password

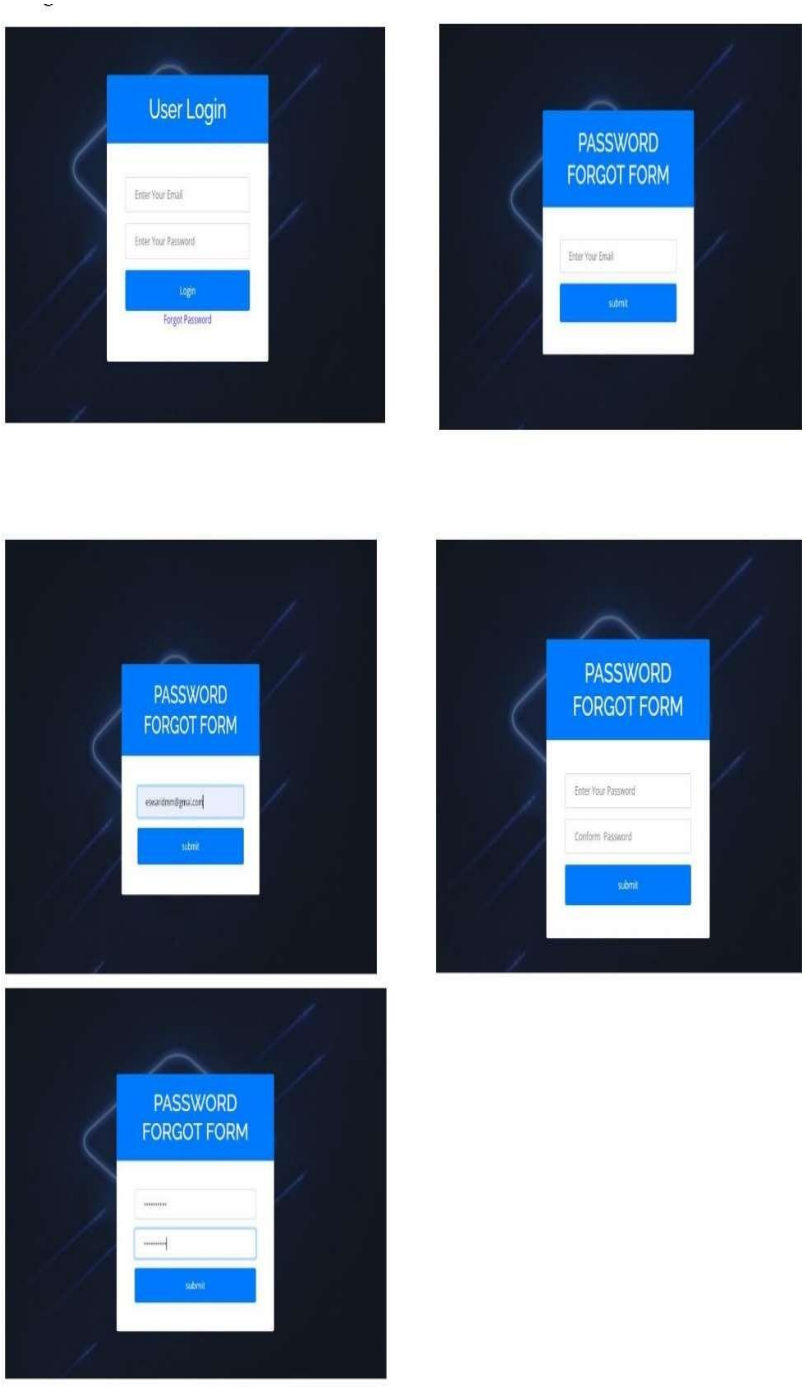

Fig. 6.3: Forget Password

CHAPTER 7

TESTING

Software testing is an essential process to evaluate the quality and performance of a software product or service. It offers an objective view to stakeholders, helping them understand and mitigate potential risks. By executing the software and identifying bugs or defects, testing ensures the product meets the required standards and is fit for its intended use.

Software testing assesses various aspects of a software component or system to ensure it meets specific criteria:

- Adherence to design and development requirements
- Correct response to diverse inputs
- Efficient performance within acceptable timeframes
- Usability for end-users
- Compatibility with intended environments for installation and operation
- Achievement of desired outcomes by stakeholders

7.1 Functionality Testing

- Database connection is successfully established.
- The flow of the application from one page to another is correct, accurate and quick.
- All the forms included in the application are working as expected.
- Proper alert messages are displayed in case of wrong inputs.
- After every action on the application the appropriate data is fetched from the backend.

7.2 Usability Testing

- The application enables smooth navigation, hence gives a user-friendly experience.
- The inputs taken from the user are via dropdown hence correct inputs are provided to the system.
- Wrong inputs given by the system are handled effectively.
- The content provided by the application is verified and is taken by the trusted sources.

7.3 Interface Testing

- The application connects correctly with the server. In case of failure an appropriate message is displayed.
- Interruptions by the server or by the user are handled efficiently.
- If the user enters wrong credentials or invalid email id, the application handles it efficiently by displaying appropriate messages.
- The interaction with the user is smooth and easy.

7.4 Performance Testing

- It works fine with moderate internet speed.
- The connection is secured and user details are stored in a secured manner.
- The switch from one screen to another is quick and smooth.
- The inputs from users are taken correctly and response is recorded quickly.

7.5 Unit testing

Unit testing focuses on verifying the correctness of individual software units by designing test cases that validate internal logic and input-output relationships. Conducted after completing a software unit and before integration, it ensures each component functions correctly and follows the specified business process or system configuration. This structural testing method examines all decision branches and code flows to confirm accurate performance according to documented specifications, using clear inputs and expected outcomes.

7.6 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.

7.7 System Test

System testing validates the complete, integrated software system against specified requirements to ensure consistent and predictable outcomes. This process-oriented testing method focuses on testing configurations, including pre-defined process links and integration points, to confirm that the system functions as expected and meets user requirements

7.8 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.

CHAPTER 8

RESULTS

The web application is designed to provide a user-friendly platform for managing and analyzing datasets. Upon initial use, new users are required to register using their email addresses and create a secure password to ensure data privacy and security. Subsequently, they can log in using their registered email and password to gain access to the application's features. Once logged in, users can navigate to the "load data" section, where they have the ability to upload disordered datasets in CSV format.

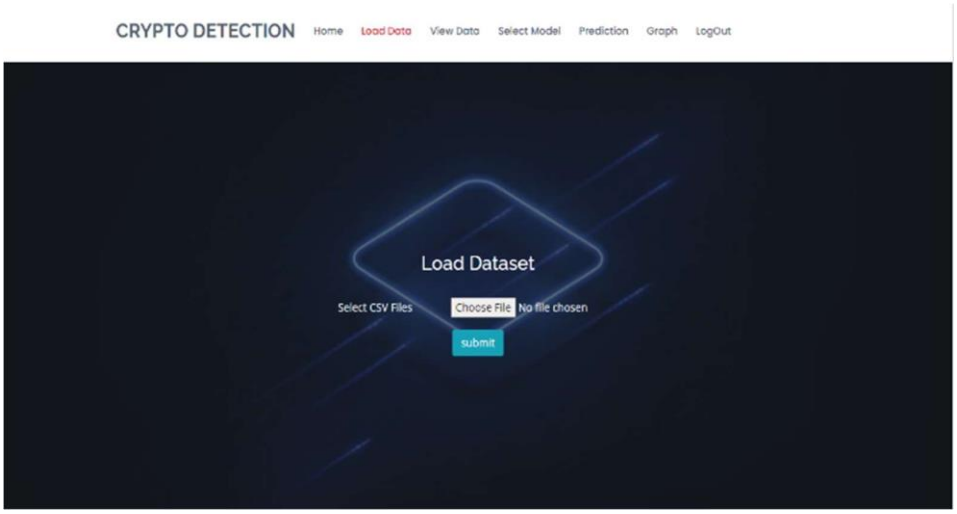

Fig. 8.1: Load Data Page

The application then processes the uploaded data to create an ordered dataset, which users can subsequently view and manipulate. Furthermore, users can select from a range of machine learning algorithms to train their data and evaluate accuracy.

Furthermore, users can select from a range of machine learning algorithms to train their data and evaluate accuracy.

Decision Tree

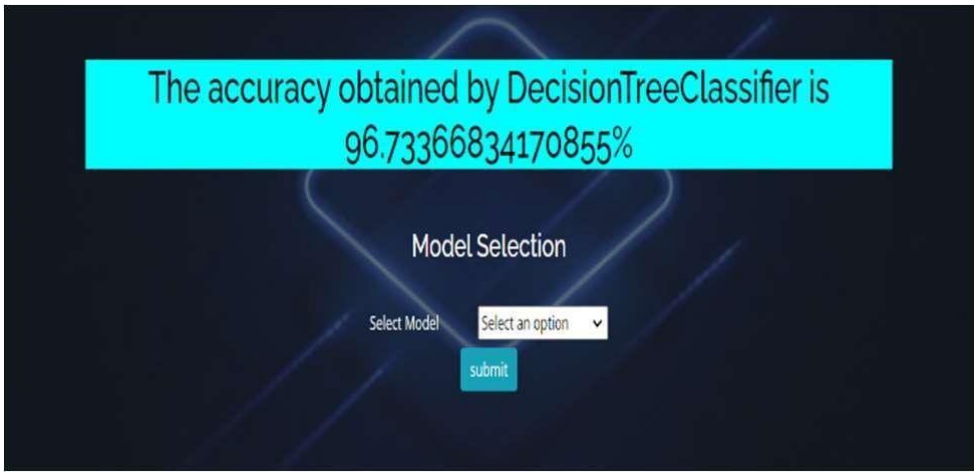

Fig. 8.2: Decision Tree Accuracy

Random Forest

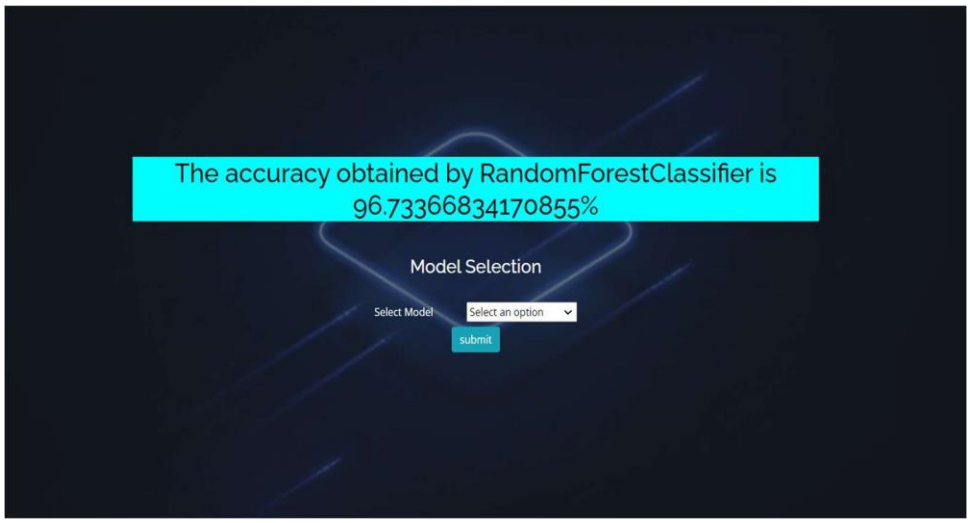

Fig. 8.3: Random Forest Accuracy

Extra Tree Classifier

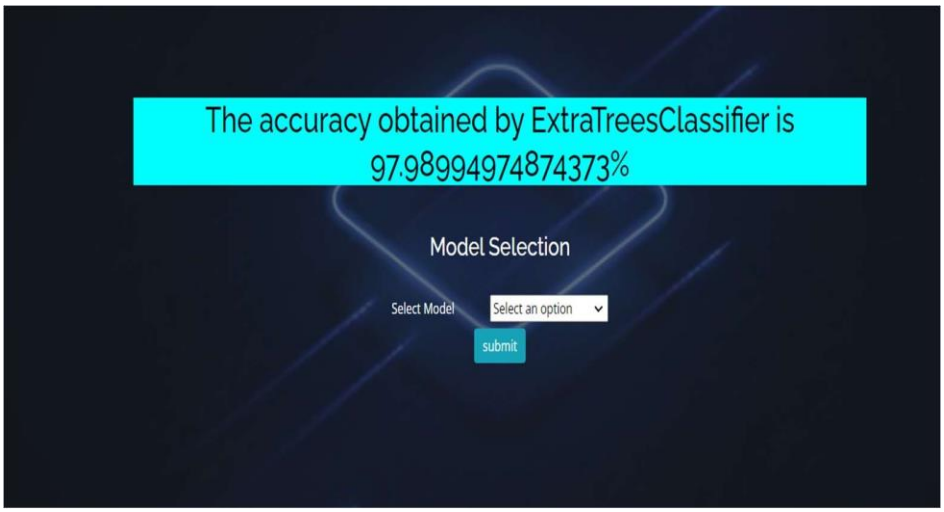

Fig. 8.4: Extra Tree Classifier Accuracy

AdaBoost Classifier

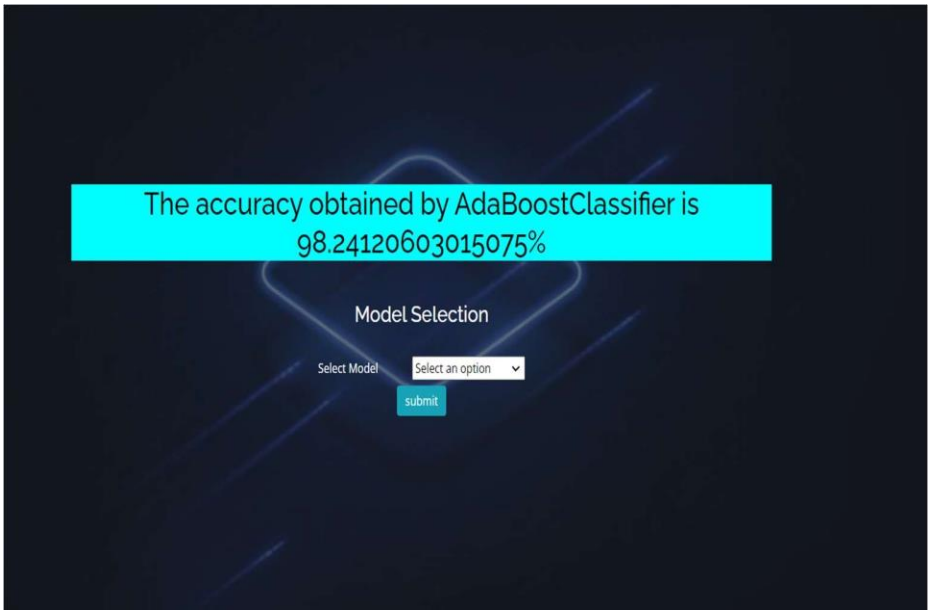

Fig. 8.5: AdaBoost Accuracy

Multi Layer Perceptron Classifier

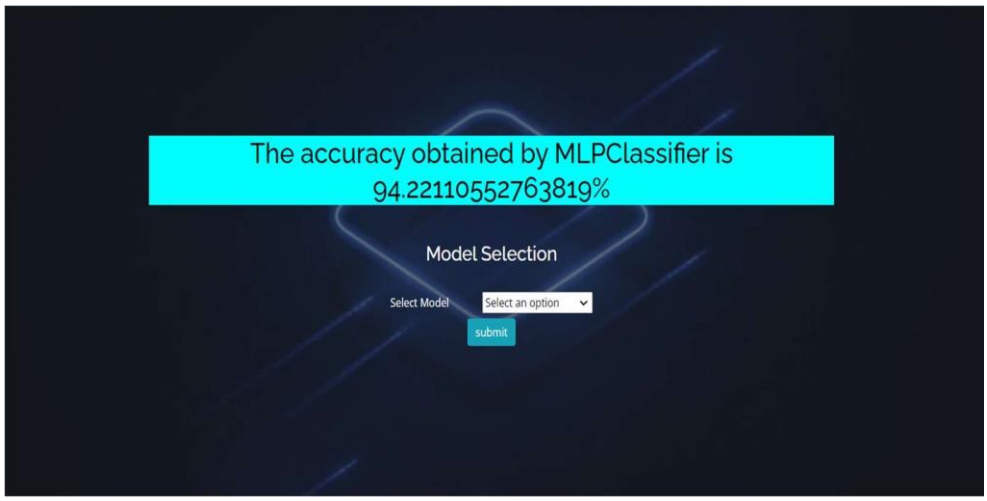

Fig. 8.6: Multi-Layer Perceptron Accuracy

The "prediction" section allows users to input various parameters and receive risk assessment results.

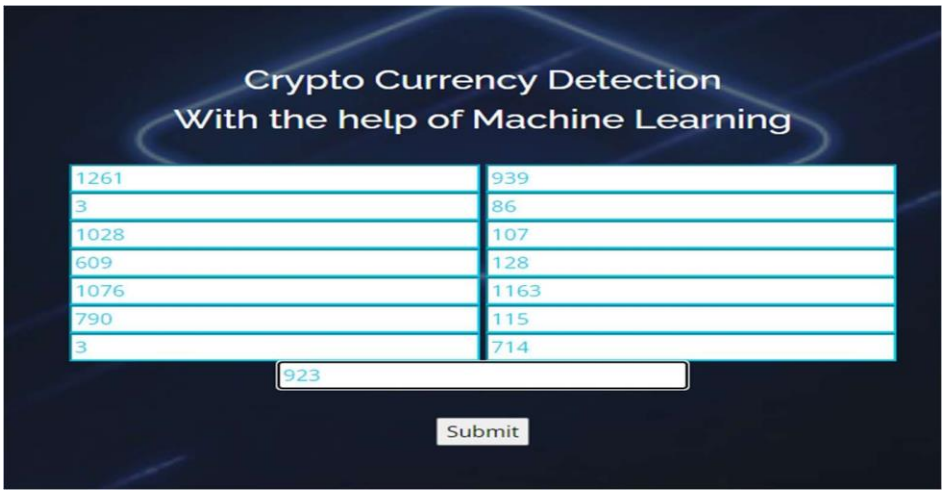

Fig. 8.7: Prediction of Risk 1

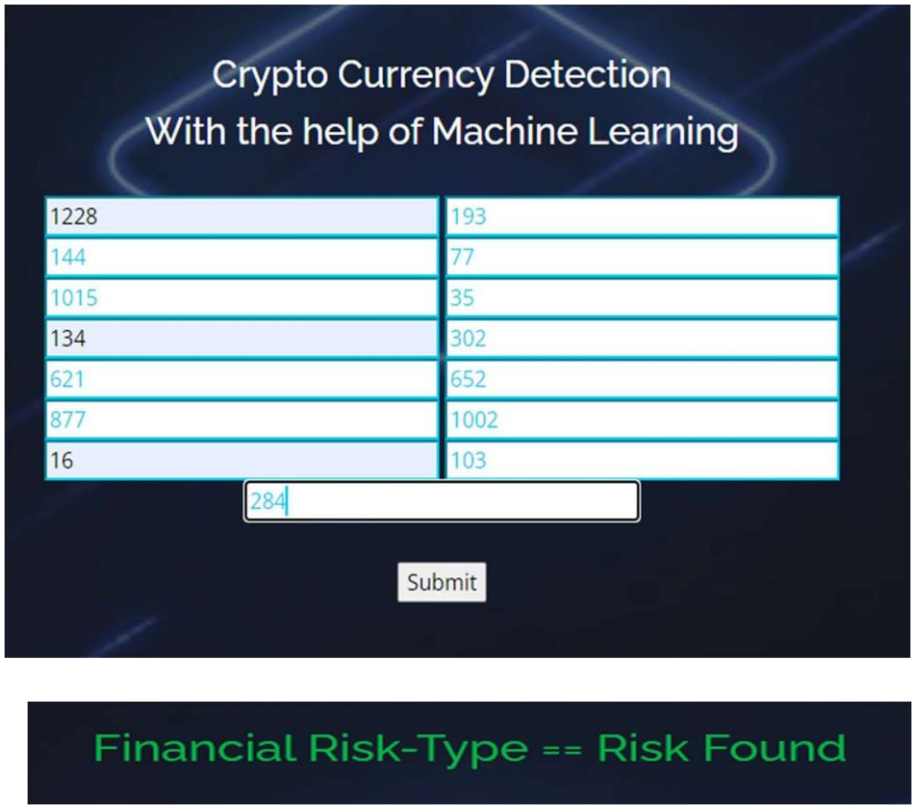

Fig. 8.8: Prediction of Risk 2

Additionally, the application provides graphical representations of data model accuracy in the "graph" section, enabling users to visually assess their data analysis results.

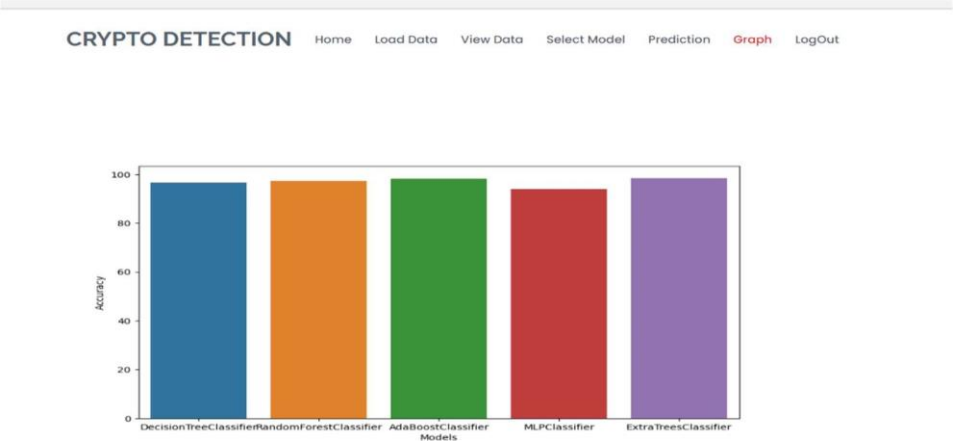

Fig. 8.9: Graph of Accuracy

CONCLUSION AND FUTURE WORK

In the current system, risk management practices rely on Reinforcement Learning, which is recognized for its limitations in accuracy within algorithms. To address this challenge, the proposed system adopts a diverse range of Machine Learning Techniques renowned for their capacity to deliver higher accuracy and more precise predictions. This strategic shift aims to enhance the effectiveness of risk assessment processes. As part of the data preprocessing stage, meticulous attention is given to eliminating null and missing values, ensuring that the datasets used for analysis are robust and reliable. Subsequently, these refined datasets are subjected to comprehensive training procedures, enabling the system to generate highly accurate predictions regarding various risk factors associated with cryptocurrency networks. Through this approach, the proposed system endeavors to significantly improve the accuracy and reliability of risk management practices within the realm of cryptocurrency.

Furthermore, the proposed system utilizes anomaly detection algorithms to identify unusual patterns or outliers in transaction data, which could indicate potential security threats or fraudulent activities within cryptocurrency networks. By flagging these anomalies, the system can alert risk managers to take timely actions, thus mitigating potential risks effectively.

REFERENCES

[1] C. Y. Kim and K. Lee, ‘‘Risk management to cryptocurrency exchange and investors guidelines to prevent potential threats,’’ in Proc. Int. Conf. Platform Technol. Service (PlatCon), Jan. 2018, pp. 1–6.

[2] I. U. Haq, A. Maneengam, S. Chupradit, W. Suksatan, and C. Huo, ‘‘Economic policy uncertainty and cryptocurrency market as a risk management avenue: A systematic review,’’ Risks, vol. 9, no. 9, p. 163, Sep. 2021.

[3] J. Gold and S. D. Palley, ‘‘Protecting cryptocurrency assets,’’ Risk Manage., vol. 68, no. 3, pp. 12–13, 2021.

[4] I. Barkai, T. Shushi, and R. Yosef, ‘‘A cryptocurrency risk–return analysis for bull and bear regimes,’’ J. Alternative Investments, vol. 24, no. 1, pp. 95–118, Jun. 2021.

[5] V. Boiko, Y. Tymoshenko, R. Y. Kononenko, and D. Goncharov, ‘‘The optimization of the cryptocurrency portfolio in view of the risks,’’ J. Manage. Inf. Decis. Sci., vol. 24, pp. 1–9, Sep. 2021.

[6] G. Köchling, ‘‘Essays in finance: Corporate hedging, mutual fund managers’ behavior, and cryptocurrency markets,’’ M.S. thesis, Universitätsbibliothek Dortmund, Dortmund, Germany, 2021.

[7] Z. Umar, N. Trabelsi, and F. Alqahtani, ‘‘Connectedness between cryptocurrency and technology sectors: International evidence,’’ Int. Rev. Econ. Finance, vol. 71, pp. 910–922, Jan. 2021.

[8] T. Kurosaki and Y. S. Kim, ‘‘Cryptocurrency portfolio optimization with multivariate normal tempered stable processes and foster-hart risk,’’ Finance Res. Lett., vol. 45, Mar. 2022, Art. no. 102143.

[9] A. Masharsky and I. Skvortsov, ‘‘Cryptocurrency market development in Latvia and the Baltic states,’’ Eur. Cooperation, vol. 1, no. 49, pp. 7–22, 2021.

[10] S. Bhattacharya and K. Rana, ‘‘A case study on cryptocurrency driven euphoria in 2020-21,’’ Int. J. Res. Eng., Sci. Manage., vol. 4, no. 3, pp. 9–11, 2021.

Smart Risk Management in Crypto Markets

Lokeswari Gara,[1,a)] Eswari Singanamala, Diya Farnaaz Syed, Amrutha Bandi, Aishwarya Jambugumpala , Manohar Vadde[2,3,4,5,6b)]

Assistant Professor[1], Student[2,3,4,5,6], Department of Computer Science and Engineering
Srinivasa Ramanujan Institute of Technology, Anantapur, Andhra Pradesh, India

[a)]*Lokeswarig.cse@srit.ac.in*,
[b)]*204g1a0530@srit.ac.in*,
[c)]*204g1a0529@srit.ac.in*,
[d)]*204g1a0509@srit.ac.in*,
[e)]*204g1a0505@srit.ac.in*,
[f)]*204g1a0553@srit.ac.in*

Abstract. Crypto money may be a well-known monetary state within the globe, posturing a assortment of perils that have an affect on the inborn hazard evaluation of hazard evaluators. Since its beginning, the rise of crypto monetary standards has displayed money related educate with a wide extend of dangers in terms of cash washing. Within the framework of the organization monetary bolsters such as opposed to currency laundering, bank enigma continue accordingly hazard master, bank supervisor, and compliance officer who includes a incitement for the associated exchange through crypto monetary forms and the clients who conceal the unlawful reserves. In this think about, the crypto money system was subjected to Progressive Chance Equality and unsupervised machine learning. The proficient bookkeeping method in terms of the inborn hazard related with bit-coin. The proficient crypto money involvement in the exchange mitigates risk compared to less experienced counter parts. In order to provide a better chance administration outcome, the Various levels Chance Equality provides a much higher yield in terms of returns or returns the balanced hazard tail. The resultant region shows up. Strong to various interims is the recommended show. the ones that undergo recalibration and the estimate of the covariance window.

Keywords : Machine Learning, Decision tree, Random Forest, MLP Classifier, Adaboost, Extra tree Classifier, ML techniques.

INTRODUCTION

Financial markets are intricate systems that involve the interaction of various elements. The structure of these systems is hierarchical, with subsystems that play their own role. However, when it comes to Portfolio assembly, there is a substantial challenge due to the absence of correlation matrix within the hierarchical framework, especially for large covariance. In recent years, the cryptocurrency market has seen a surge in popularity, with around 2500 types of cryptocurrencies trading a total value of 252.5 trillion dollars. The cryptocurrency market operates in a volatile environment, attracting attention from news publishers and investors alike.

Efforts have been made to regulate the cryptocurrency market, with rules in place to Combatting money laundering and protect investors. In spite of these endeavors, the market remains dynamic and prone to price fluctuations. Various models and techniques have been proposed to analyze and manage cryptocurrency portfolios. For example, Markowitz optimization and high-frequency relationships have been explored to provide insights and improve system stability. Estimation errors in return estimation and sophisticated portfolio techniques, such as the Black Litterman model, have been used to manage cryptocurrency investments more effectively.

Wavelet-based analysis has been applied to study the dynamic interdependence between liquid cryptocurrencies, revealing heterogeneous behavior among traders and investors. Different trading rules have been compared to develop effective trading strategies. It is recommended that investors and professionals in the industry be aware of the inherent risks associated with digital assets. Chartered Professional Accountants Canada (CPAC) has highlighted the importance of considering audit considerations specific to the cryptocurrency ecosystem.

Overall, the cryptocurrency market presents both opportunities and challenges, requiring a nuanced approach to portfolio management and risk assessment.

LITERATURE SURVEY

[1]. C. Y. Kim and K. Lee, ‘‘Risk management to cryptocurrency exchange and investors guidelines to prevent potential threats,’’ in Proc. Int. Conf. Platform Technol. Service (PlatCon), Jan. 2018, pp. 1–6.

The allure and capital infusion into cryptocurrency are burgeoning, The value of Bitcoin surpassing $10000 as of November 2017. The future of this trend is uncertain. While blockchain technology offers enhanced security compared to traditional currency systems, falls short in terms of cryptocurrency distribution and management. Many people acquire cryptocurrency through exchanges and mining, often without adequate knowledge. This paper examines the vulnerabilities of cryptocurrency exchanges and user wallets and proposes risk management strategies following international standards like NIST and ISO. The focus is on analyzing blockchain weaknesses, implementing effective countermeasures, addressing investor vulnerabilities, and outlining server management and personal security tips.

[2]. I. U. Haq, A. Maneengam, S. Chupradit, W. Suksatan, and C. Huo, ‘‘Economic policy uncertainty and cryptocurrency market as a risk management avenue: A systematic review,’’ Risks, vol. 9, no. 9, p. 163, Sep. 2021.

Cryptocurrency literature is growing quickly these days, with a focus on how the cryptocurrency market can help manage risks. This type of asset is still new and has gaps in existing research that need to be filled for future studies. A recent review looked at a lot of different studies on how cryptocurrencies can be used to manage risks related to economic policy ambiguity (EPU).The examination discovered that cryptocurrencies have different levels of connection to EPU in different countries, which affects how well they can help manage risks. These different connections are because EPU depends on the policies and decisions made by each country's regulatory authorities. Dealing with these varied connections requires different approaches to handle stock market ups and downs and economic policy ambiguity in various economies. The unique rules and regulations of the cryptocurrency market also affect how well it can protect against risks in each economy. Many different groups like investors, policymakers, fund custodians, or asset managers can benefit from this research. The authors suggest that future studies should look at different topics, data frequencies, and research methods to keep improving our understanding of how cryptocurrencies can help manage risks.

[3]. J. Gold and S. D. Palley, ‘‘Protecting cryptocurrency assets,’’ Risk Manage., vol. 68, no. 3, pp. 12–13, 2021

Literature, as the advisory explains, suggests that companies facing exploit impacts should act swiftly by contacting law enforcement and implementing various technical measures to reduce risks. These measures involve creating new keys for cryptocurrency wallets or transferring assets to new wallets, employing physical wallets to secure private keys in a segregated location, setting up two-factor authentication for added security, removing affected hosts from the network, changing all associated passwords, and updating patches for software and hardware components. It is advised to scan systems thoroughly as a precaution for potential lateral movement by threat actors and the installation of additional malware. By taking these recommended actions promptly, companies can enhance their cybersecurity defenses and prevent unauthorized access to their systems.

[4]. I. Barkai, T. Shushi, and R. Yosef, ‘‘A cryptocurrency risk–return analysis for bull and bear regimes,’’ J. Alternative Investments, vol. 24, no. 1, pp. 95–118, Jun. 2021.

The authors of this article introduce a fresh perspective for analyzing the risk-reward characteristics of prominent cryptocurrencies such as bitcoin, litecoin, ripple, and ethereum. They emphasize on examining how the price behavior of each individual cryptocurrency can impact the others. By dividing the returns of each cryptocurrency into different bull and bear market phases, they investigate how these phases overlap and change over time. Additionally, they explore the effect of distributed-denial-of-service attacks on these market phases to understand their influence on regime changes. Their research uncovers systemic risk indicators in the cryptocurrency market and reveals potential contagion patterns between different cryptocurrencies. These findings offer valuable insights for investors looking to manage risk in their portfolios and pave the way for future research in this rapidly evolving market. Overall, the study highlights the increasing systemic risk in the cryptocurrency market and the diminishing diversification benefits of holding a mixed portfolio of cryptocurrencies.

EXISTING SYSTEM

In the current system, deploying machine learning algorithms is somewhat challenging to construct due to insufficient data regarding the data visualization. The RL was compared to existing research endeavors and the prevalent benchmarks in this domain, the duo management evolve algorithms for portfolios and the fundamental Deep Q-Network(DQN) of the trading system. In terms of bit-coin risk management, the proposed RL algorithms are compared to other existing studies To surmount all these challenges, we leverage machine learning packages accessible in the scikit-learn library.

Disadvantages:

- High intricacy
- time-intensive

PROPOSED SYSTEM

We have put forward various machine learning models for categorizing the machine learning-based assessment of cryptocurrency market financial risk management. Correspondingly, analogous research endeavors have suggested models for appraising such performance. Hence, we advocate the utilization of Decision Trees, Random Forests, MLP Classifiers, AdaBoost, and Extra Tree Classifiers to forecast the risks.

Advantages:

- Maximum precision
- Simplifies time complexity
- Is simple to operate

BLOCK DIAGRAM

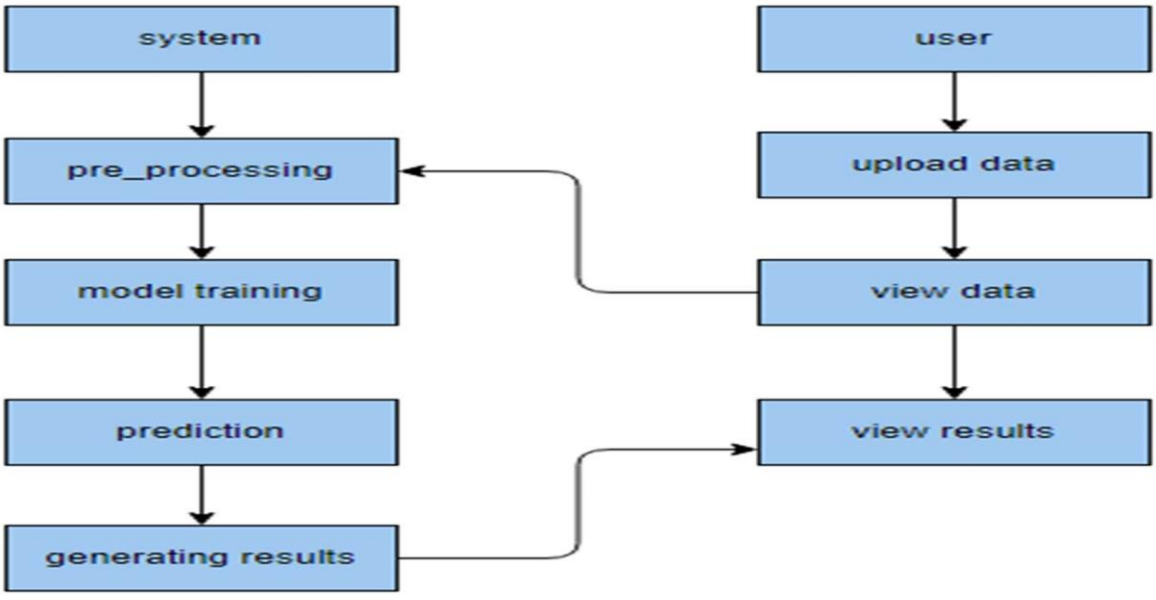

Figure 1 : Block Diagram

METHODOLOGIES

1. Decision Tree:

A decision tree is like a flowchart that makes a difference us make choices. It contains a tree structure where each inner hub speaks to a highlight, like a color or measure, and the branches appear the choice rules. At the conclusion of the branches, we have leaf hubs that appear the ultimate result. The beat hub is called the root hub, and the tree learns how to part based on distinctive highlights. This prepare of part is called recursive apportioning. Choice trees are incredible for decision-making since they see like flowcharts that people can effortlessly get it. To construct a choice tree, we to begin with choose the finest include to part the information. At that point, we keep part the information into littler bunches based on distinctive highlights until we reach a point where all the information focuses have a place to the same bunch, we run out of highlights, or there are no more information focuses cleared out. This way, choice trees offer assistance us make sense of complicated information in a straightforward way.

Formula : $Entropy(s) = -P(yes)log2\ P(yes) - P(no)\ log2\ P(no)$

2. Random Forest Classifier

A random forest is a cool way to use a bunch of decision trees to make smart choices. Instead of just relying on one tree, it combines the predictions of many trees to make better guesses about things. This helps avoid mistakes like overfitting and actually improves how accurately we can predict stuff. The random forest algorithm trains all these decision trees in a "forest" by using a special method called bagging. Bagging is like boosting the accuracy of machine learning by working together as a team. Some cool things about random forests are that they can handle missing data really well, they can make pretty good predictions without needing a lot of tuning, and they're great at stopping overfitting. Plus, they randomly pick different groups of features when making decisions. They also use clever ideas like entropy and information gain to make sure they're making the best choices possible. These ideas help reduce uncertainty and make the predictions more accurate. So basically, random forests are like a smart team of decision-makers that work together to get things right.

3. Multi Layer Perceptron:

Multilayer Perceptron (MLP) in Scikit-learn simplifies deep learning in Python but lacks an activation function in the output layer. It uses varied loss functions for regression and classification tasks and supports single and multiple target values regression without GPU support for accelerated computation. Unlike Keras, Scikit-learn's MLP restricts fine-tuning parameters per layer. MLPs, a subset of feedforward neural networks, comprise input, hidden, and output layers, trained using backpropagation. They consist of at least three layers: input, hidden, and output, with each node, except inputs, acting as a neuron with a nonlinear activation function. MLPs distinguish from linear perceptrons by handling non-linearly separable data. The term "multilayer perceptron" refers to multiple perceptrons organized into layers. MLP "perceptrons" diverge from true perceptrons by allowing arbitrary activation functions and handling both classification and regression tasks. MLPs offer stochastic problem-solving capabilities for complex problems like fitness approximation. Cybenko's theorem demonstrates MLPs' ability to approximate universal functions, facilitating mathematical model development. MLPs are effective classifiers, treating classification as a form of regression for categorical response variables.

Final Data Set

	last_updated	market_cap_usd	max_supply	name	percent_change_1h	percent_change_24h	percent_change_7d	price_btc	price_usd	rank	symbol	total_supply	Pred
128	69	1031	29	145	268	715	684	889	1214	0	181	406	1
455	68	1030	124	441	241	378	313	880	1205	1	423	656	1
130	93	1029	29	147	375	328	340	885	1210	2	107	408	1
664	86	1028	107	609	128	1076	1163	790	1115	3	714	923	0
1005	56	1027	122	976	309	389	227	466	789	4	1257	1005	0
331	57	1026	25	333	349	402	582	882	1207	5	300	270	1
723	57	1025	51	699	242	541	462	864	1189	6	671	576	0
132	97	1024	29	150	194	246	266	876	1201	7	191	407	1
789	59	1023	124	763	147	934	919	875	1200	8	1239	385	0
236	94	1022	119	239	232	322	304	361	678	9	24	988	1
457	71	1021	124	443	240	395	307	846	1171	10	421	658	1
816	74	1020	124	784	204	498	642	485	808	11	1225	952	0
442	85	1019	124	396	401	960	1027	783	1109	12	411	877	1
817	73	1018	124	785	350	278	399	854	1179	13	758	668	0
1108	64	1017	124	1078	121	1007	961	369	687	14	1235	1009	0
788	59	1016	124	760	557	1089	1161	837	1162	15	726	581	0
144	77	1015	35	134	302	621	652	877	1202	16	103	284	1
720	68	1014	124	693	338	427	656	813	1138	17	669	700	0
1306	74	1013	124	1275	188	757	370	878	1203	18	1282	151	0
854	87	1012	124	826	203	266	466	816	1141	19	799	719	0

Figure 2: Data Set

4. AdaBoost

AdaBoost, also known as Adaptive Boosting, is a method used in machine learning to make weak learners stronger. Unlike traditional boosting methods, AdaBoost pays more attention to wrongly classified data points, aiming to improve overall performance. The process involves creating decision trees one after the other, with each new model focusing more on fixing mistakes made by the previous one. This cycle continues until the desired number of base learners is reached. In AdaBoost, each tree is quite simple, with just a single node and two leaves, often called a "stump." This is different from random forests, which have a fixed number of trees. Imagine a dataset with three features and a Yes-No output; AdaBoost assigns a starting weight of 1/5 to each record to keep things fair. Over time, as the algorithm learns, these weights are adjusted to contribute to accurate predictions.

Formula : W=1/N

5. Extra Tree Classifier

The Extra Trees Classifier is an ensemble learning method that aggregates the outcomes of multiple uncorrelated decision trees to make classification predictions. It operates similarly to the Random Forest Classifier but varies in how it constructs decision trees within the forest. Within the Extra Trees Forest framework, each decision tree is created from the primary training sample. At every testing node, a randomized subset of features (termed 'k') is presented to each tree from the feature set. The decision tree then identifies the prime feature for data segmentation based on a mathematical criterion, commonly the Gini Index. This stochastic feature sampling leads to the formation of numerous uncorrelated decision trees. Regarding feature selection, while constructing the forest, the Gini Importance of each feature is computed, representing its contribution to the decision-making process. Features are ranked based on their Gini Importance, allowing users to select the top k features for further analysis.

$$Gain(S,A)=Entropy(S)-\Sigma_{vEValues(A)}\frac{|Sv|}{|S|}Entropy(S_v)$$

RESULT SECTION

The web application is intended to offer a user friendly platform for managing and analysing datasets. Upon initial use, new users are required to register using their email addresses and create a secure password to ensure data privacy and security. Subsequently, they can log in using their registered email and password to gain access to the application's features. Once logged in, users can navigate to the "load data" section, where they have the ability to upload disordered datasets in CSV format. The application then processes the uploaded data to create an ordered dataset, which users can subsequently view and manipulate. Furthermore, users can select from a range of machine learning algorithms to train their data and evaluate accuracy.

Home page

Figure 3: Home Page

Accuracy of different algorithms are given below.

	Models	Accuracy
0	DecisionTreeClassifier	96.733668
1	RandomForestClassifier	97.236181
2	AdaBoostClassifier	98.241206
3	MLPClassifier	93.969849
4	ExtraTreesClassifier	98.492462

Figure 4: Accuracy table

Graph

Among those algorithms Extra Tree Classifier has high accuracy.
Below graph describe the comparison among the algorithms

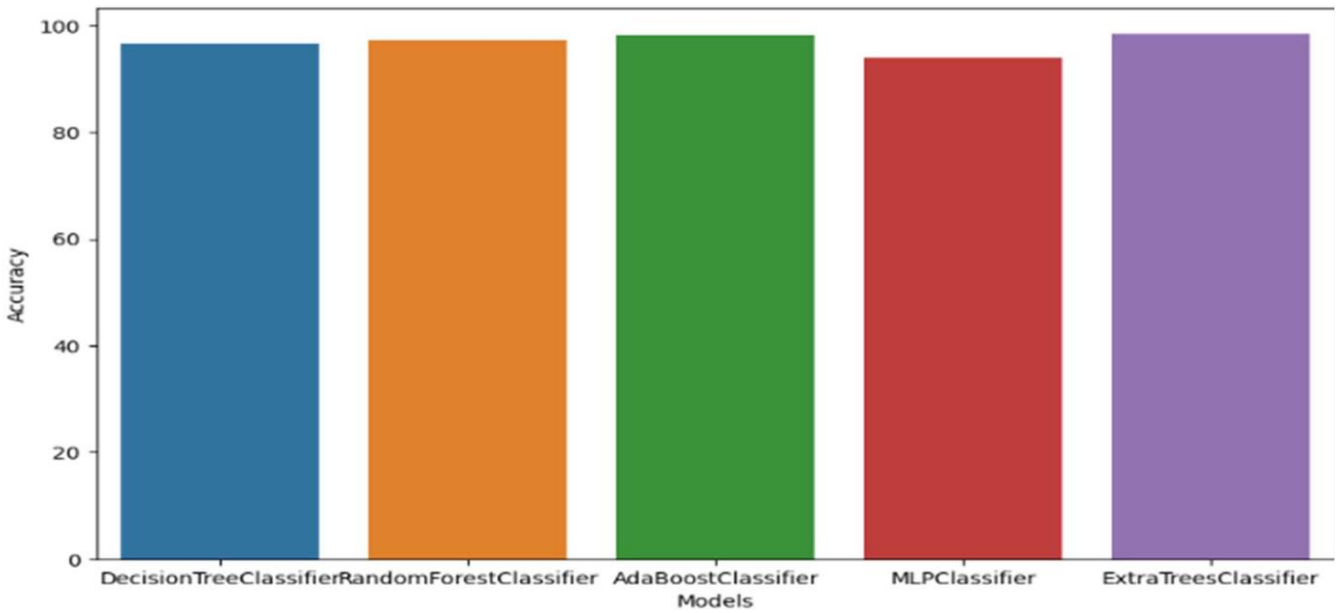

Figure 4: Graph

CONCLUSION

The risk management of a crypto currency network was examined in this study utilizing the Reinforcement Learning (RL) approach and an asset allocation strategy referred to as Hierarchical Risk Parity (HRP) were utilized for a cryptocurrency portfolio. When compared to other machine learning algorithms employed in this field, reinforcement learning produces superior performance evaluation discoveries. The primary justification for employing RL in this situation process is the learning-based characteristic of this technique enables system structures to attain high intricacy regarding providing the correct knowledge to the system's operations. Furthermore, the HRP framework possesses the best features and the most favourable diversification traits. The data were studied utilizing multiple estimating windows and approaches, as well as re-adjusting within the selected timeframe. The implemented HRP provides a viable alternative for transitional asset allocations.

REFERENCES

1. C .Y .Kin and K. Lee, ‘‘*Risk management to cryptocurrency exchange and investors guidelines to prevent potential threats,*’’ in Proc. Int. Conf. Platform Technol. Service (PlatCon), Jan. 2018, pp. 1–6.
2. I. U. Haq, A. Maneengam, S. Chupradit, W. Suksatan, and C. Huo, *‘‘Economic policy uncertainty and cryptocurrency market as a risk management avenue: A systematic review,’’* Risks, vol. 9, no. 9, p. 163, Sep. 2021.
3. J. Gold and S. D. Palley, *‘‘Protecting cryptocurrency assets,’’* Risk Manage., vol. 68, no. 3, pp. 12–13, 2021.
4. I. Barkai, T. Shushi, and R. Yosef, *‘‘A cryptocurrency risk–return analysis for bull and bear regimes,’’* J. Alternative Investments, vol. 24, no. 1, pp. 95–118, Jun. 2021.
5. V. Boiko, Y. Tymoshenko, R. Y. Kononenko, and D. Goncharov, *‘‘The optimization of the cryptocurrency portfolio in view of the risks,’’* J. Manage. Inf. Decis. Sci., vol. 24, pp. 1–9, Sep. 2021.
6. G. Köchling, *‘‘Essays in finance: Corporate hedging, mutual fund managers’ behavior, and cryptocu rrency markets,’ ’*M.S. thesis, Universitätsbibliothek Dortmund, Dortmund, Germany, 2021.
7. Z. Umar, N. Trabelsi, and F. Alqahtani, *‘‘Connectedness between cryptocurrency and technology sec tors: International evidence,’’* Int. Rev. Econ. Finance, vol. 71, pp. 910–922, Jan. 2021.
8. T. Kurosaki and Y. S. Kim, *‘‘Cryptocurrency portfolio optimization with multivariate normal tempered stable processes and foster-hartrisk,’’* Finance Res. Lett., vol. 45, Mar. 2022, Art. no. 102143.
9. [9] A. Masharsky and I. Skvortsov, ‘‘Cryptocurrency market development in Latvia and the Baltic states,’’ Eur. Cooperation, vol. 1, no. 49, pp. 7–22, 2021.
10. [10] S. Bhattacharya and K. Rana, ‘‘A case study on cryptocurrency driven euphoria in 2020-21,’’ Int. J. Res. Eng., Sci. Manage., vol. 4, no. 3, pp. 9–11, 2021.

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)
5th & 6th April 2024
Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss Lakeswari Gara, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in Crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor

PARTICIPATION CERTIFICATES

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)
5th & 6th April 2024
Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss. Eswari Singamala, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)
5th & 6th April 2024
Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss. Amrutha Bandi, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)

5^{th} & 6^{th} *April* 2024

Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss. Aishwarya Jambugumpala, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)
5th & 6th April 2024
Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss. Divya Farnaaz syed, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in Crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor

AIP Conference Proceedings

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(UGC- AUTONOMOUS)
PRODDATUR
Vidya Nagar, Proddatur, YSR Kadapa (Dist.), Andhra Pradesh 516360

ICIAET-24 Proceedings

International Conference on Innovative Approaches in Engineering & Technology (ICIAET-24)
5th & 6th April 2024
Organized by Department of Electrical & Electronics Engineering

Certificate of Appreciation

This certificate is awarded to Dr./Mr./Mrs./Miss. Manohar Vadde, of SRIT, Anantapur has participated and presented a paper entitled Smart Risk Management in Crypto Markets with Paper ID: ICIAET-P218 in ICIAET-2024.

Dr V Mahesh Kumar Reddy
Convenor & Organising Chair

Dr G Sreenivasula Reddy
Principal, CBIT

turnitin
Official sponsor