A

Mini Project Report

on

Customer Segmentation for Financial Products using Clustering

Submitted in partial fulfillment of the requirements for the

degree

Third Year Engineering – Information Technology

by

Neha Khorne	**22104029**
Janvi Kadam	**22104158**
Divya Keni	**22104022**
Abhirami Kalathil	**22104166**

Under the guidance of

Prof. Sonal Balpande

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P. SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W)-400615

UNIVERSITY OF MUMBAI

Academic year: 2024-25

CERTIFICATE

This to certify that the Mini Project-2B report on Customer Segmentation for Financial Products using Clustering has been submitted by Neha Khorne(22104029), Janvi Kadam(22104158), Divya Keni(22104022) and Abhirami Kalathil(22104166) who are bonafide students of A. P. Shah Institute of Technology, Thane as a partial fulfillment of the requirement for the degree in Information Technology, during the second half of academic year 2024-2025 in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Sonal Balpande
Guide

Dr. Kiran Deshpande
HOD, Information Technology

Internal Examiner:

1.

External Examiner:

Place: A. P. Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide Prof. Sonal Balpande. Expressing gratitude towards our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our project coordinators **Ms. Sonal Jain & Ms. Shafaque Syed** who gave us his/her valuable suggestions and ideas when we were in need of them.

We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

Abstract

1. Introduction 6

    1.1. Purpose 7

    1.2. Problem Statement 8

    1.3. Objectives 9

    1.4. Scope 10

2. Literature Review 11

3. Project Design 13

    3.1. Features and Functionality 13

    3.2. System Architecture 15

    3.3. DFD (Data Flow Diagram) 16

    3.4. Use Case diagram 18

4. Technical Specification 19

5. Project Implementation 21

6. Project Scheduling 23

7. Results 24

8. Conclusion and Future Scope 29

References

ABSTRACT

The project "Customer Segmentation for Financial Products using Clustering" aims to analyze and segment customers in the financial sector based on their stock market behavior. By utilizing advanced machine learning techniques such as K-Means, the project focuses on clustering investors according to factors like annual income, investment risk, and preferred sectors. The goal is to provide personalized stock recommendations to users based on their segmented profile. The project uses realtime stock data from Indian markets, stored in MongoDB, and integrates data processing tools such as Pandas, NumPy, and yfinance. The application is built with a full-stack approach using Flask for the backend, and HTML/CSS with Jinja2 and Bootstrap for the frontend, providing a dynamic and interactive user experience. The dashboard allows users to view real-time stock data, manage their portfolios, and access tailored recommendations. The ultimate goal is to empower users by giving them personalized insights and recommendations to make informed financial decisions.

Chapter 1

Introduction

The Customer Segmentation for Financial Products using Clustering project is a full-stack web application designed to provide personalized stock recommendations and real-time portfolio management based on user behavior and financial data. The project aims to help investors make informed decisions by segmenting them into clusters using K-means clustering, based on their annual income, investment risk, and preferred sectors. The platform integrates with external stock market APIs (yfinance) to retrieve real-time stock prices and market data for over 100 Indian stocks, displaying them with search functionalities. The clustering model generates personalized stock recommendations, suggesting the top 4-5 Indian stocks based on the user's portfolio value, previous investments, and risk profile. The portfolio page provides dynamic updates, displaying the current value of stocks, profit/loss calculations, and overall portfolio performance. The application features interactive visualizations using Chart.js or Plotly, enabling users to track their portfolio performance, monitor profit/loss, and view market trends. The project is built using a robust tech stack, including Flask (Python) for the backend, MongoDB for data storage, and HTML, CSS, and Bootstrap for the frontend. It follows a modular MVC structure, ensuring scalability and maintainability. Inspired by financial apps like Groww and Zerodha, this project aims to offer a comprehensive investment platform with real-time data, clustering-based recommendations, and interactive portfolio management, helping users make informed financial decisions effectively.

1.1 Purpose

1. Smart Stock Recommendations:

Recommend top-performing Indian stocks based on the user's portfolio value and investment history. Use clustering insights to offer personalized suggestions, enhancing investment strategies.

2. Interactive Data Visualization:

Display portfolio performance trends and clustering results through dynamic charts and graphs. Use Chart.js and Plotly for real-time visual feedback on stock performance.

3. Seamless User Experience:

Implement a responsive and user-friendly UI using Flask templates, Jinja2, and Bootstrap. Ensure smooth navigation with dynamic content rendering and mobile compatibility

4. Efficient Technology Integration:

Utilize Flask for backend processing and MongoDB for data storage. Incorporate Flask-SocketIO for real-time updates, ensuring a seamless and interactive user experience.

1.2 Problem Statement

The lack of a robust investor segmentation method hinders financial institutions from providing personalized services, optimizing resource allocation, and maximizing profitability. Traditional segmentation methods, such as demographic or transaction-based segmentation, may not accurately capture the complexities of investor behavior and preferences.

Solution Proposed:

Customer Segmentation for financial product using clustering can help to identify distinct investor groups based on their behavior, preferences, and needs. This approach enables financial institutions to:

1. **Improve customer experience:** By understanding the unique characteristics and needs of each investor cluster, financial institutions can tailor their services, products, and communication strategies to meet the specific needs of each group.
2. **Optimize resource allocation:** Clustering investors based on their behavior and preferences enables financial institutions to allocate resources more efficiently. For example, they can focus marketing efforts on the most profitable clusters or allocate more resources to clusters with high growth potential.
3. **Increase revenue growth:** By identifying high-value investor clusters, financial institutions can develop targeted strategies to attract and retain these investors, leading to increased revenue growth.

1.3 Objectives

1. To segment customers based on financial behavior using clustering algorithms.
2. To provide real-time portfolio management.
3. To offer personalized stock recommendations.
4. To create interactive data visualizations for better financial insight and analysis.
5. To ensure real-time updates.

1.4 Scope

1. Can be used for clustering algorithms to group customers based on financial behaviors like income, spending, and investments.
2. Can be used for portfolio performance tracking monitors returns and evaluates portfolio composition over time, enabling investors to identify areas for improvement and optimize portfolio performance. It also compares portfolio returns to benchmarks and industry averages to assess performance.
3. Can be used for risk assessment modeling evaluates market risk, credit risk, and liquidity risk using statistical models, enabling investors to identify potential risks and opportunities.
4. Can be used for daily returns calculation evaluates daily portfolio performance and identifies trends, enabling investors to monitor returns in real-time and adjust strategies. It also compares daily returns to benchmarks and industry averages to assess performance.
5. Can be used for stock correlation analysis evaluates relationships between stocks to identify potential correlations, enabling investors to optimize portfolio composition and reduce risk. It also identifies opportunities to diversify portfolios and mitigate potential risks.

Chapter 2

Literature Review

1. **Risk Assessment and Portfolio Optimization using Machine Learning Techniques by A. K. Singh and R. K. Singh (2020) :**

   This paper proposes a machine learning-based approach for risk assessment and portfolio optimization, using techniques such as Random Forest and Genetic Algorithm. The approach aims to help investors and financial institutions make informed decisions about portfolio composition and risk management.

2. **Stock Correlation Analysis using Machine Learning Algorithms by S. S. Rao and S. K. Singh (2019) :**

   This paper applies machine learning algorithms to analyze stock correlations and identify potential investment opportunities. The study uses techniques such as Principal Component Analysis and Artificial Neural Networks to uncover hidden patterns in stock price movements and correlations.

3. **A Comparative Study of Clustering Algorithms for Investor Segmentation by S. K. Goyal and S. Singh (2018) :**

   This paper compares the performance of various clustering algorithms for segmenting investors based on their demographic and transactional characteristics. The study aims to identify the most effective clustering algorithm for investor segmentation, enabling financial institutions to tailor their services to specific investor groups.

Chapter 3

Project Design

3.1 Features and Functionality

1. **Customer Segmentation Using Clustering Algorithms.**
    - **Feature:** Segmentation of customers into distinct groups based on financial behaviors (e.g., income, spending patterns, and investment activities).
    - **Functionality:** Automatically clusters users into groups using machine learning algorithms (K-Means), enabling businesses to identify target audiences for personalized services.
2. **Interactive Data Visualizations**
    - **Feature:** Interactive charts, graphs, and tables for exploring customer segments and key financial metrics.
    - **Functionality:** Users can click, hover, and filter data visualizations such as pie charts, bar charts, and scatter plots to analyze customer groups and financial data dynamically.
3. **Real-time Data Updates.**
    - **Feature**: Real-time updating of financial data and customer segments.
    - **Functionality:** As new data is added or customer behavior changes, the platform will automatically refresh to reflect the latest customer groupings and metrics.
4. **User-Friendly Dashboard.**
    - **Feature:** A responsive and intuitive dashboard for users to explore customer data and insights.
    - **Functionality**: The dashboard is designed for easy navigation, allowing users to quickly access and understand the customer segments, metrics, and visualizations.
5. **Stock Recommendations**
    - **Feature:** Personalized Stock Recommendations
    - **Functionality:** Suggests top 4-5 Indian stocks to users based on their portfolio value, existing stock ownership, and clustering insights, ensuring relevant and tailored investment options**.**

3.2. System Architecture

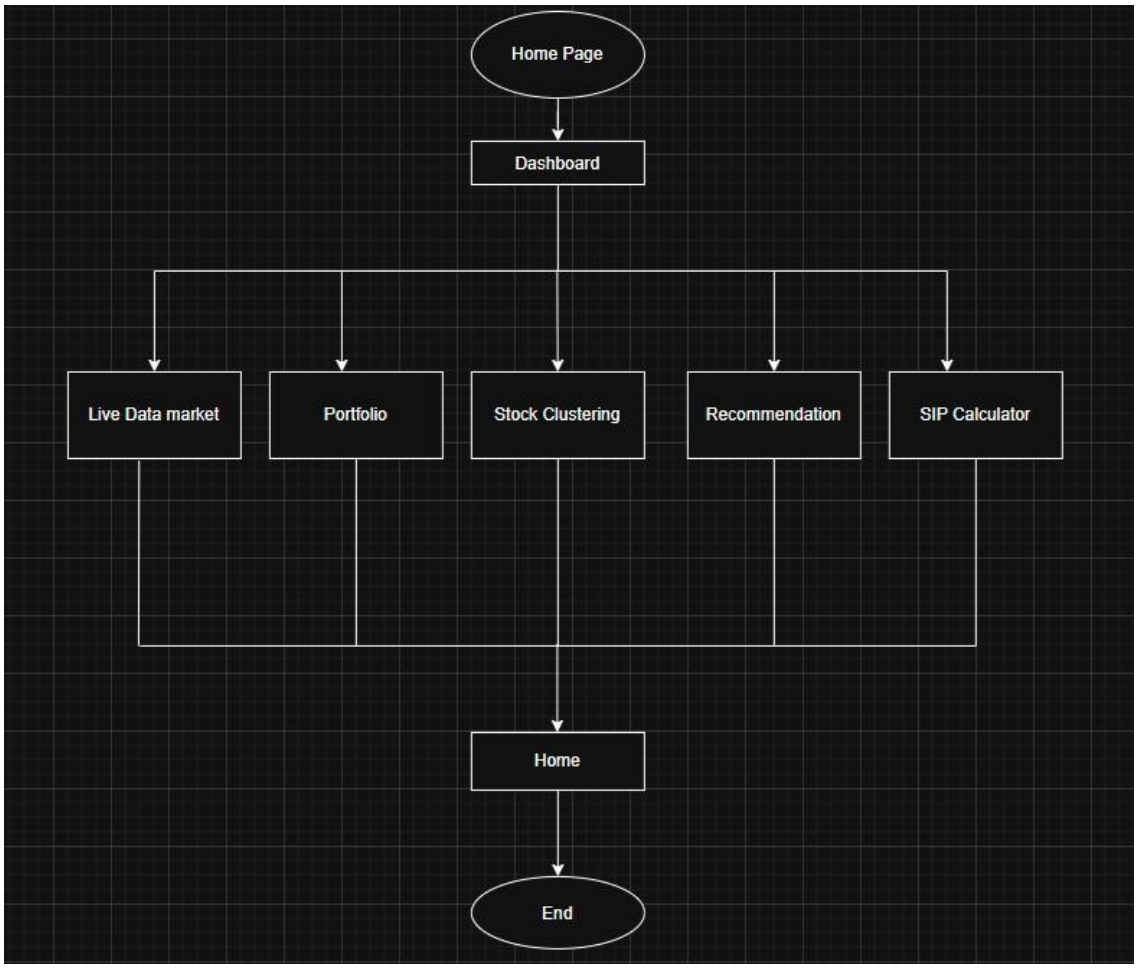


Fig 3.2: Block Diagram

Fig 3.2 The images show a flowchart representing the navigation structure of a financial project dashboard. The dashboard includes paths to "Live Data Market" (showing real-time stock prices), "Portfolio" (displaying top and worst-performing customer portfolios), "Stock Clustering" (segmenting customers based on three factors), "Recommendation" (recommending stocks for individual customers), and a "SIP Calculator"(providing SIP investment plans).

3.3 DFD (Data Flow Diagram)

DFD Level 0:

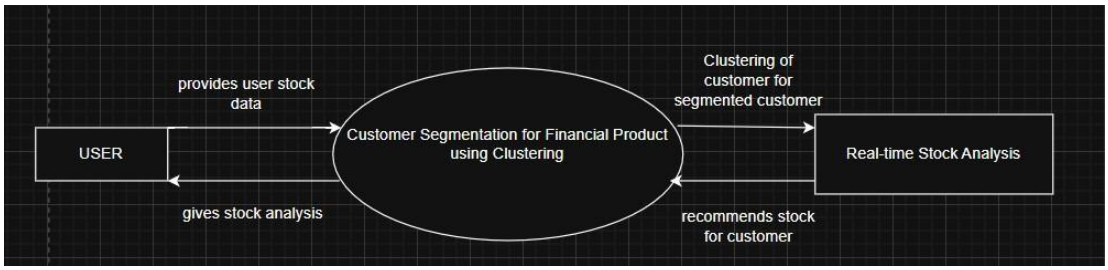


Fig 3.3.(a) : DFD level 0

Fig 3.3(a) The image outlines a system that combines customer segmentation (via clustering) with real-time stock analysis to offer personalized financial recommendations or services. It bridges user needs with data-driven insights, emphasizing automation and customization in financial product offerings

DFD level 1 :

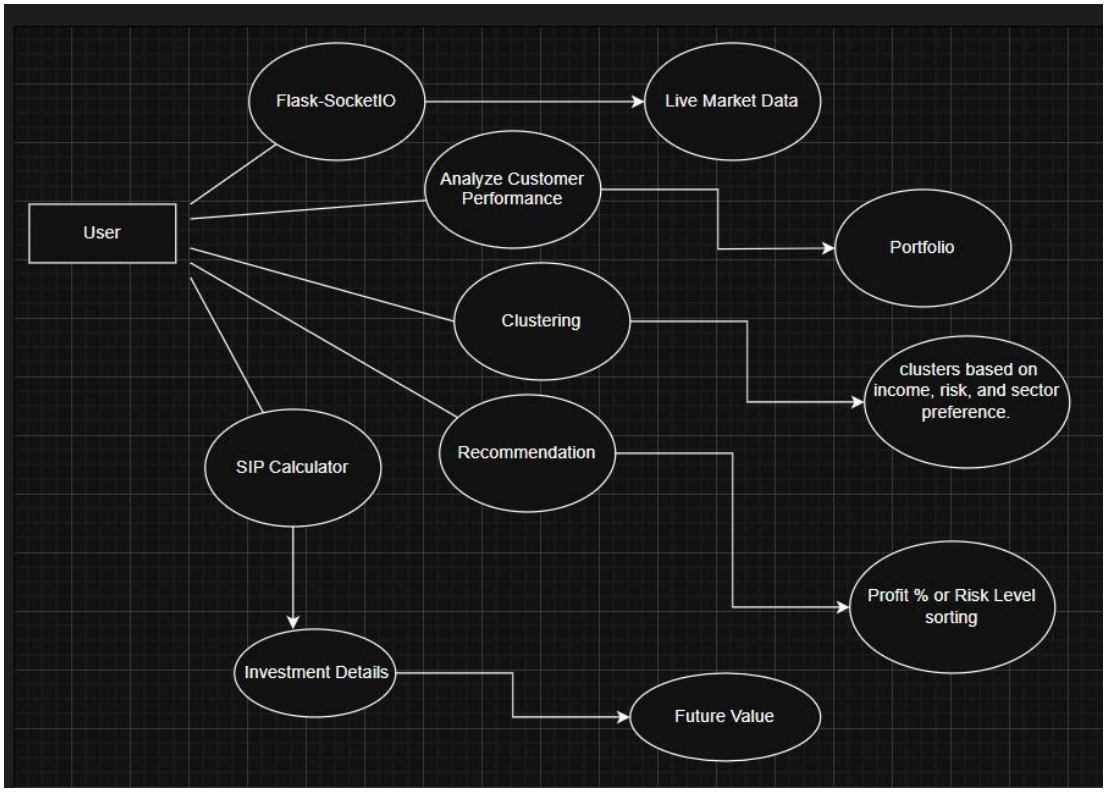


Fig 3.3(b) : DFD level 1

Fig 3.3.(b) : This image outlines real-time investment recommendation system using Flask-SocketIO. It clusters users by income, risk, and sector preference to provide personalized stock suggestions. The system analyzes live market data, tracks portfolios, and sorts options by profit or risk. Users get tailored investment details and future value predictions. The goal is data-driven, automated financial advice.

3.4 Use Case Diagram

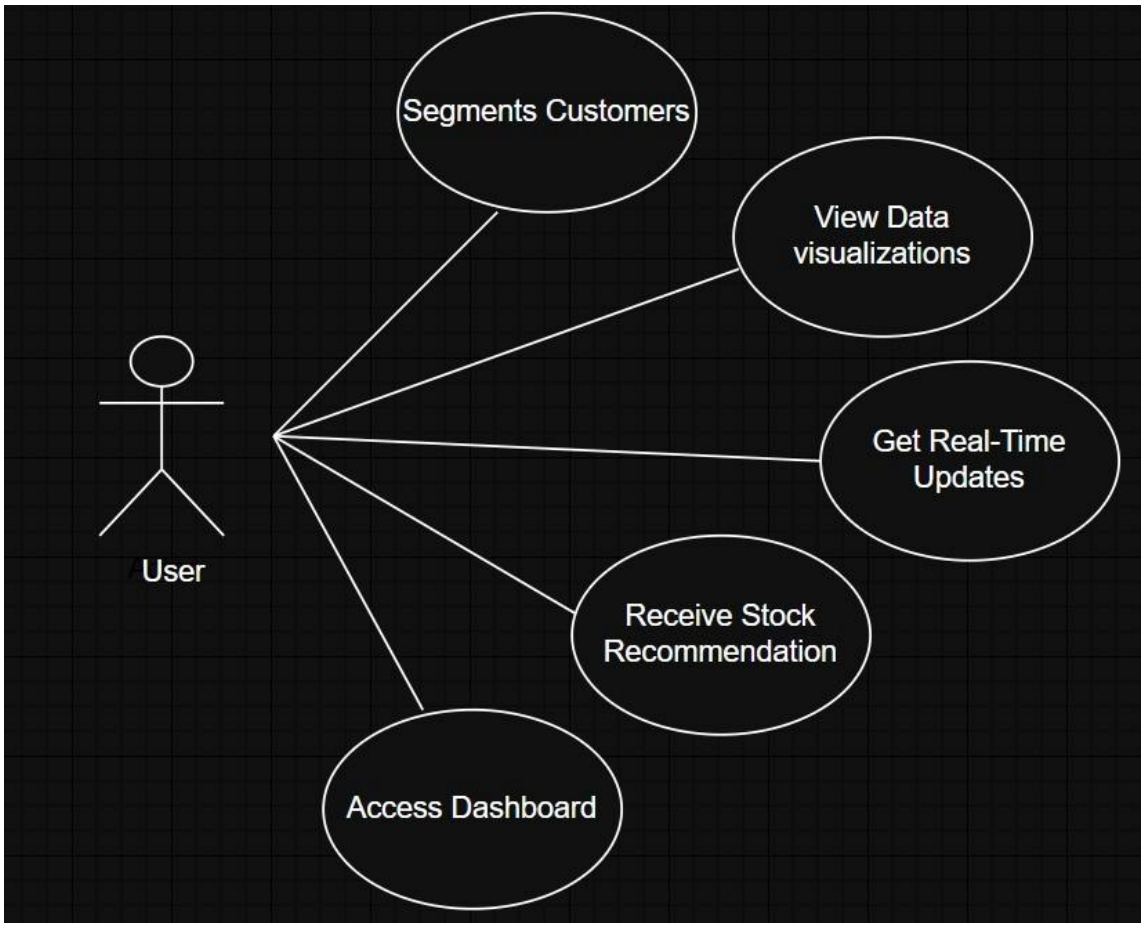


Fig:3.4: Use Case Diagram

Fig 3.4The use case diagram illustrates the admin's workflow in a stock market system, managing Live Market Data, Portfolio, Stock Clustering, AI Stock Picks, and SIP Calculator for informed investment decisions. It enables efficient financial analysis and personalized stock recommendations for optimized portfolio management.

Chapter 4

Technical Specification

Backend

- Python (Flask)
- MongoDB (NoSQL Database)
- Flask-SocketIO (WebSockets)
- Gunicorn + Eventlet (Production server)
- Threading for concurrent data fetching

Frontend:

- HTML5
- CSS3
- JavaScript
- Bootstrap (for styling)
- Chart.js / Plotly (for visualizations)

Data Science Libraries

- Pandas
- NumPy (Data manipulation)
- Scikit-learn (Clustering models)
- yfinance (Real-time stock data fetching) Matplotlib (Data visualization)

Chapter 5

Project Implementation

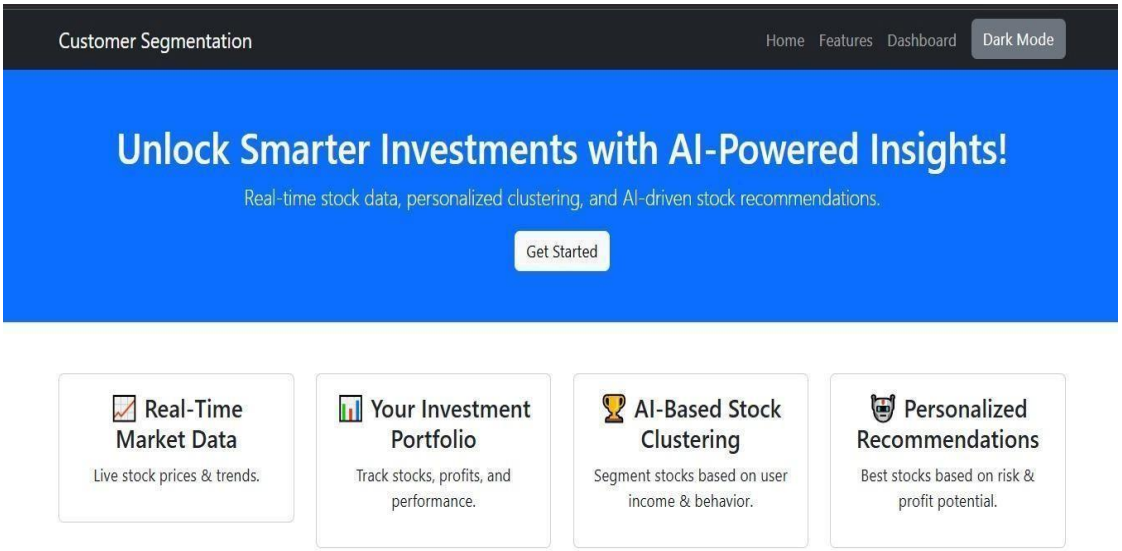


Fig 5.1 Home page

Fig 5.1 This image promotes an AI-driven investment platform offering real-time stock data, portfolio tracking, and personalized recommendations based on income, risk, and behavior. It helps users make smarter investment decisions through automated clustering and AI-powered insights.

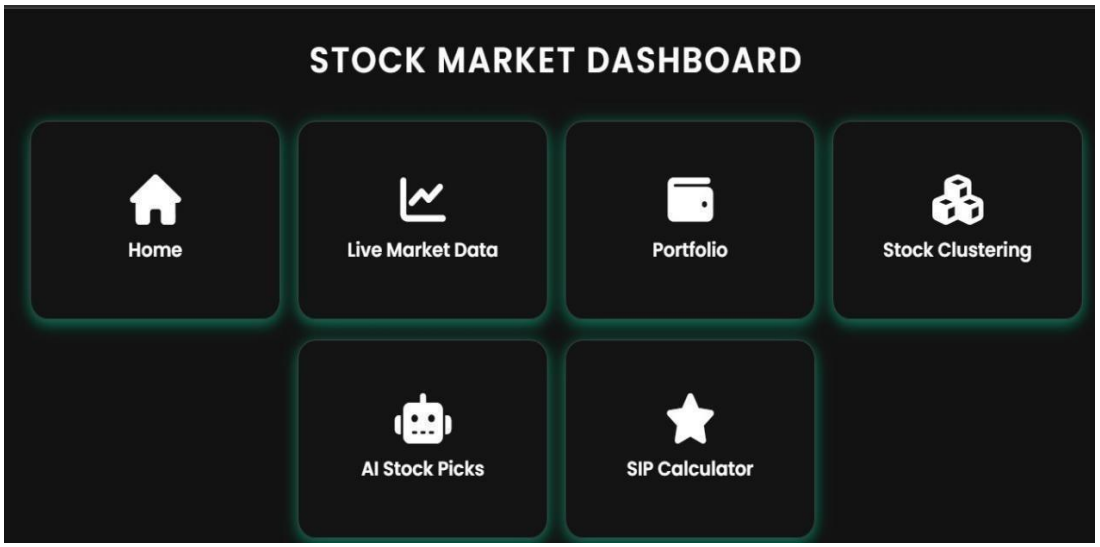


Fig 5.2 Dashboard

Fig 5.2 This image shows a stock market dashboard featuring live market data, portfolio tracking, AI-driven stock clustering, and investment tools like SIP calculators. It provides investors with real-time insights and AI-powered stock recommendations for smarter trading decisions.

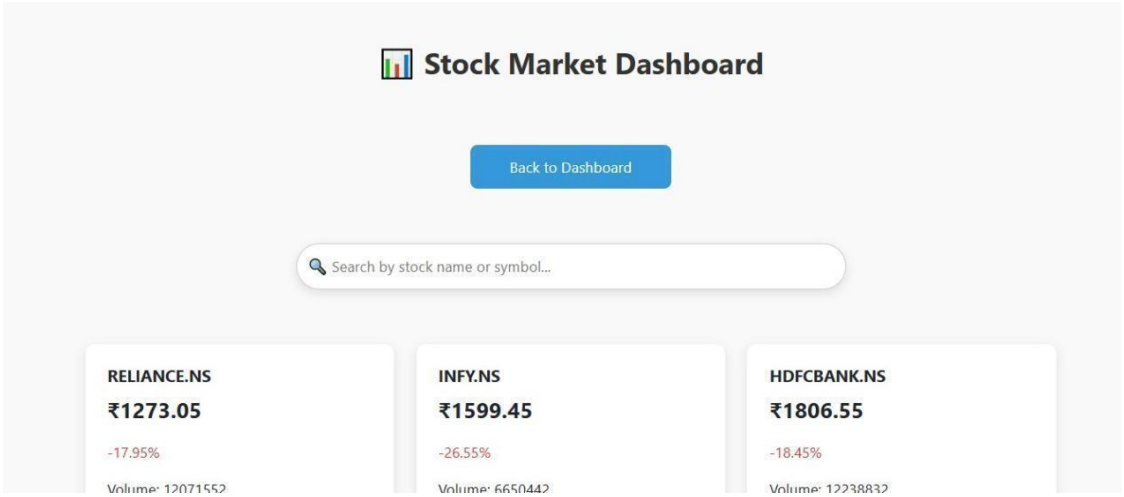


Fig 5.3 Live Market Data

Fig 5.3 This image shows a stock market dashboard displaying real-time data for stocks like RELIANCE.NS, INFXNS, and HDFCBANK.NS, including their prices, percentage changes, and trading volumes. It allows users to search for stocks and track market performance at a glance.

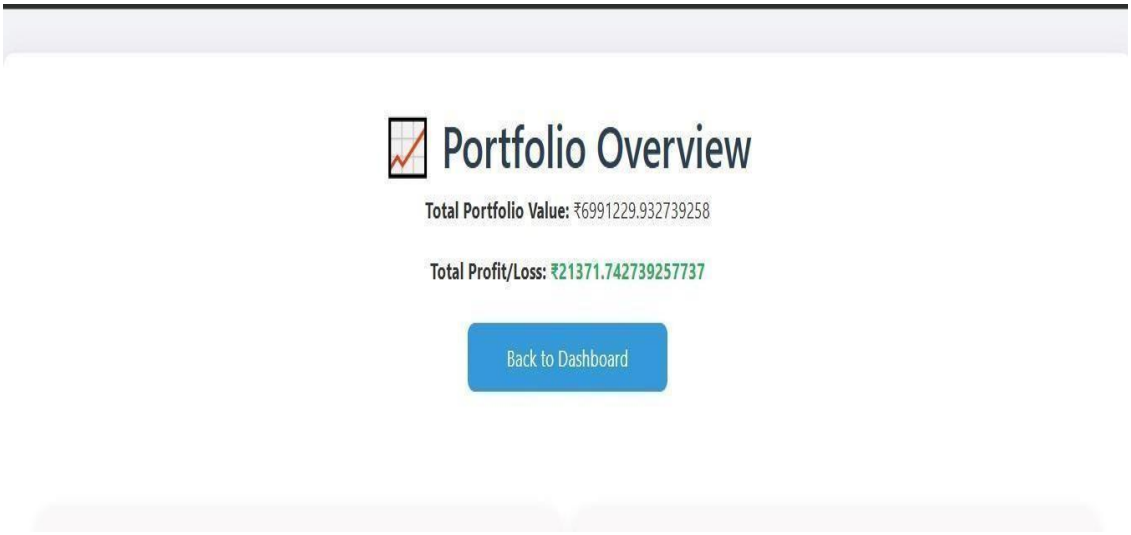


Fig 5.4 Portfolio

Fig 5.4 This image displays a **portfolio summary** showing an extremely large total value (₹86.99 quadrillion) and profit (₹121,371.74), likely placeholder/test data. It includes a "Back to Dashboard" button for navigation.

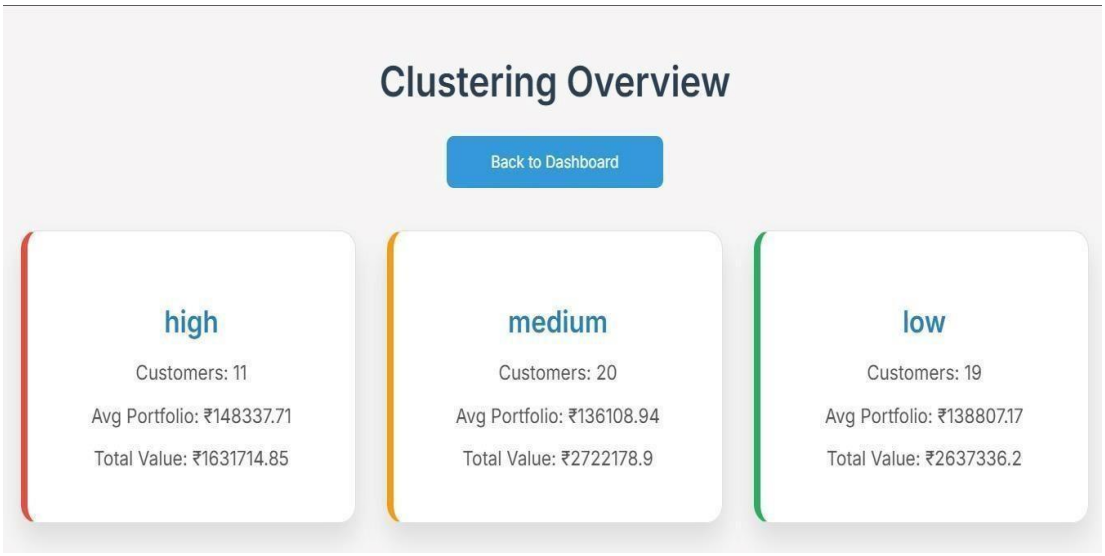


Fig 5.5 Clustering

Fig 5.5 This image displays a customer clustering dashboard segmenting users into high, medium, and low tiers based on portfolio metrics, showing customer counts, average portfolio values, and total cluster values. It includes a "Back to Dashboard" navigation option for user convenience**.**

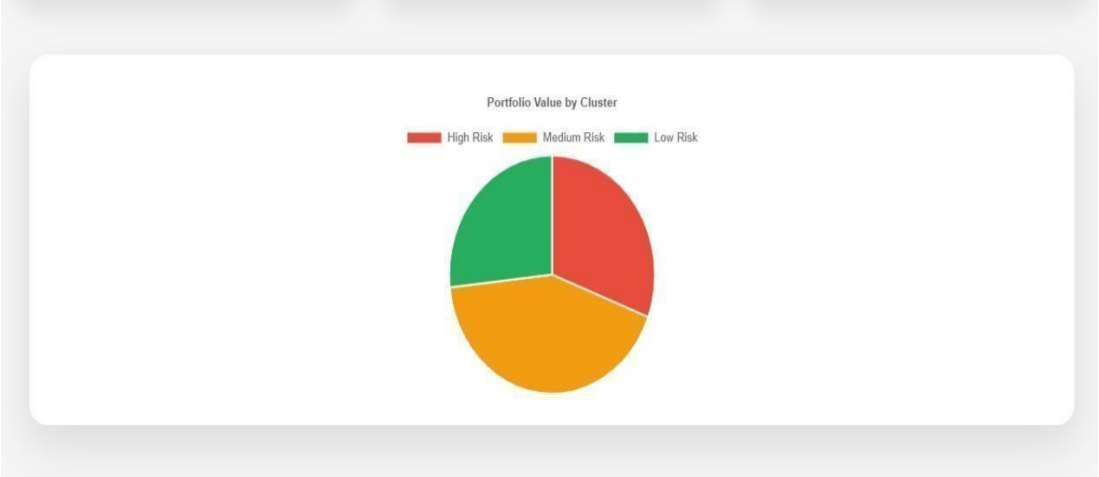


Fig 5.6 Clustering pie chart

Fig 5.6 This image shows a risk-based portfolio breakdown, categorizing investments into High, Medium, and Low Risk clusters. It appears to be a visualization (likely a chart/graph) comparing portfolio values across these risk categories.

Fig 5.7 Recommendation System

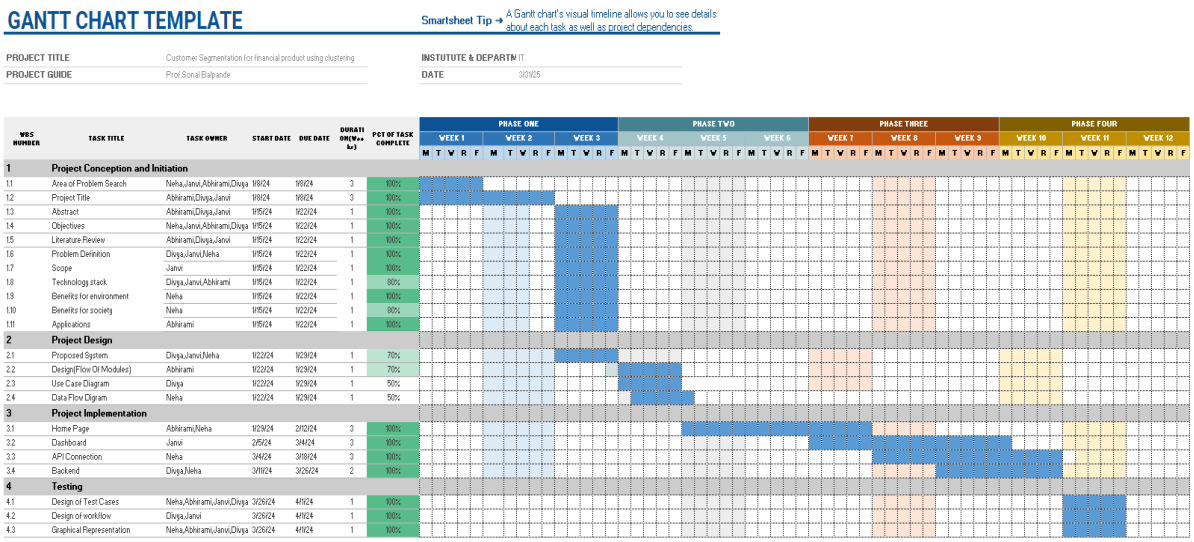
Fig 5.7 This image displays a personalized stock recommendations screen generated using the K-means clustering algorithm to group and suggest stocks based on user profiles. It includes a "Back to Dashboard" button for easy navigation to other features.

Fig 5.8 SIP Calculator

Fig 5.8 This image shows a SIP (Systematic Investment Plan) calculator that allows users to estimate returns across market caps (Large/Mid/Small) with preset interest rates. Users can input investment amount (€5000), rate (9%), duration (10 years), and start date to calculate projected gains.

Chapter 6

Project Scheduling



GANTT CHART TEMPLATE

Smartsheet Tip → A Gantt chart's visual timeline allows you to see details about each task as well as project dependencies.

PROJECT TITLE: Customer Segmentation for financial product using clustering
INSTITUTE & DEPARTM: IT
PROJECT GUIDE: Prof.Sonal Ballpande
DATE: 3/31/25

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION(Days)	PCT OF TASK COMPLETE
1	**Project Conception and Initiation**					
1.1	Area of Problem Search	Neha,Janvi,Abhirami,Divya	1/8/24	1/8/24	3	100%
1.2	Project Title	Abhirami,Divya,Janvi	1/8/24	1/8/24	3	100%
1.3	Abstract	Abhirami,Divya,Janvi	1/15/24	1/22/24	1	100%
1.4	Objectives	Neha,Janvi,Abhirami,Divya	1/15/24	1/22/24	1	100%
1.5	Literature Review	Abhirami,Divya,Janvi	1/15/24	1/22/24	1	100%
1.6	Problem Definition	Divya,Janvi,Neha	1/15/24	1/22/24	1	100%
1.7	Scope	Janvi	1/15/24	1/22/24	1	100%
1.8	Technology stack	Divya,Janvi,Abhirami	1/15/24	1/22/24	1	90%
1.9	Benefits for environment	Neha	1/15/24	1/22/24	1	100%
1.10	Benefits for society	Neha	1/15/24	1/22/24	1	90%
1.11	Applications	Abhirami	1/15/24	1/22/24	1	100%
2	**Project Design**					
2.1	Proposed System	Divya,Janvi,Neha	1/22/24	1/29/24	1	70%
2.2	Design(Flow Of Modules)	Abhirami	1/22/24	1/29/24	1	70%
2.3	Use Case Diagram	Divya	1/22/24	1/29/24	1	50%
2.4	Data Flow Digram	Neha	1/22/24	1/29/24	1	50%
3	**Project Implementation**					
3.1	Home Page	Abhirami,Neha	1/29/24	2/12/24	3	100%
3.2	Dashboard	Janvi	2/5/24	3/4/24	3	100%
3.3	API Connection	Neha	3/4/24	3/18/24	3	100%
3.4	Backend	Divya,Neha	3/11/24	3/26/24	2	100%
4	**Testing**					
4.1	Design of Test Cases	Neha,Abhirami,Janvi,Divya	3/26/24	4/1/24	1	100%
4.2	Design of workflow	Divya,Janvi	3/26/24	4/1/24	1	100%
4.3	Graphical Representation	Neha,Abhirami,Janvi,Divya	3/26/24	4/1/24	1	100%

PHASE ONE: WEEK 1, WEEK 2, WEEK 3
PHASE TWO: WEEK 4, WEEK 5, WEEK 6
PHASE THREE: WEEK 7, WEEK 8, WEEK 9
PHASE FOUR: WEEK 10, WEEK 11, WEEK 12

Figure 6.1: Project Scheduling Gantt Chart

Chapter 7

Result

The project successfully achieved customer segmentation by applying clustering algorithms to financial data, grouping users based on attributes such as income, spending behavior, and savings patterns. These clusters represent distinct customer profiles, which can be leveraged for more personalized financial strategies and services. The segmentation reveals valuable insights into customer diversity, financial habits, and potential investment behaviors.

The web-based dashboard presents these results through interactive visualizations, making it easy to interpret and explore customer groups. Users can view charts and graphs that dynamically display the distribution of customers across clusters, along with key statistics like average income or spending levels. This enhances the overall understanding of each segment and supports effective decision-making by financial analysts or service providers.

Overall, the project demonstrates the practical application of machine learning in financial analytics. It transforms raw data into actionable insights through real-time visual updates and an intuitive user interface. These results lay the foundation for further features such as stock recommendations, risk profiling, and tailored financial planning, making the solution scalable and adaptable to various financial institutions.

Chapter 8

Conclusion and Future Scope

The "Customer Segmentation for Financial Products using Clustering" project demonstrates the effective use of machine learning, specifically K-Means clustering, to segment investors based on their financial behavior, such as annual income, investment risk appetite, and preferred sectors. By integrating real-time stock data from Indian markets, dynamic portfolio management, and personalized stock recommendations, the platform empowers users to make smarter financial decisions. The project employs a full-stack architecture with Flask for the backend, MongoDB for storage, and Bootstrap with Chart.js for the frontend, ensuring a scalable, responsive, and interactive user experience. This work highlights how behavioral segmentation can significantly enhance customer experience compared to traditional demographic-based methods.

In the future, the platform can be further enhanced by incorporating advanced features such as sentiment analysis using Natural Language Processing (NLP) to adjust stock recommendations based on real-time market news and trends. Expanding the system to support global stock markets would offer users opportunities for international diversification. Additional features like more detailed risk assessment modules, automated alerts for stock price movements, and real-time performance comparisons with market benchmarks could make the platform even more powerful and useful for investors seeking dynamic portfolio management.

Moreover, developing a dedicated mobile application would ensure easier accessibility and real-time monitoring for users on the go. Introducing AI-driven virtual financial advisors could offer automated, personalized financial guidance, enhancing the platform's advisory capabilities. As the platform grows, implementing user feedback mechanisms and continuous model improvement will ensure that stock recommendations remain relevant, personalized, and aligned with evolving market conditions. Thus, with future upgrades, the project can evolve into a comprehensive financial advisory solution that bridges machine learning and practical investment management.

REFERENCE

1. Risk Assessment and Portfolio Optimization using Machine Learning Techniques by A. K. Singh and R. K. Singh (2020).
2. Stock Correlation Analysis using Machine Learning Algorithms by S. S. Rao and S. K. Singh (2019).
3. Financial Risk Prediction Using Machine Learning Algorithms by M. Patel and K. Shah (2021)
4. A Comparative Study of Clustering Algorithms for Investor Segmentation by S. K. Goyal and S. Singh (2018).