



BANKING CUSTOMER ANALYSIS BY USING JUPYTER

¹J.Priyanka, ²R.reshetha Reddy, ³B.Madhuri, ⁴U.Naveen kumar

¹Assistant Professor, ²³⁴Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

jalagapriyanka.cse@siddhartha.co.in, 23TQ1A6749@siddhartha.co.in,

23TQ1A6733@siddhartha.co.in, 23TQ1A6741@siddhartha.co.in

ABSTRACT

Banking Customer Analysis is an important process that helps banks understand customer behavior, preferences, and financial activities. The main objective of this project is to analyze banking customer data using Jupyter Notebook and identify useful insights that can help banks improve their services and customer satisfaction. In this project, customer data such as age, account balance, transaction history, and other related details are collected and analyzed using data analysis techniques. The analysis is performed using Python along with libraries like Pandas, NumPy, and Matplotlib to clean, process, and visualize the data. The project focuses on identifying customer patterns, segmenting customers based on their financial behavior, and understanding trends in banking activities. Visualization techniques such as charts and graphs are used to represent the analyzed data clearly and effectively. The results of this analysis help in predicting customer needs, improving decision-making, and providing personalized banking services. Overall, this project demonstrates how data analysis using Jupyter Notebook can support banks in understanding their customers better and enhancing their operational strategies.

I INTRODUCTION

The banking sector generates a large amount of customer data every day through transactions, account activities, and financial services. Analysing this data helps banks understand customer behaviour, preferences, and financial patterns. Banking customer analysis plays an important role in improving customer satisfaction, enhancing banking services, and supporting better decision-making. With the growth of digital banking, financial institutions need effective tools to analyse and manage large datasets. Data analysis techniques help banks identify trends, classify customers based on their behaviour, and predict future financial needs. This allows banks to provide personalized services, improve marketing strategies, and reduce potential risks. In this project, banking customer data is analysed using Jupyter Notebook, which provides an interactive platform for data analysis and visualization. The analysis is performed using Python along with powerful libraries such as Pandas for data manipulation, NumPy for numerical operations, and Matplotlib for graphical representation of data. The main aim of this project is to analyse banking customer information to discover meaningful insights about customer demographics, account balances, and transaction behaviour. By applying data analysis methods, the project helps in understanding customer patterns and improving banking strategies. Overall, this project demonstrates how modern



data analysis tools and techniques can be used to transform raw banking data into valuable information that supports better financial services and customer relationship management.

II LITERATURE SURVEY

Data analysis in the banking sector has gained significant importance due to the rapid growth of digital financial services and the increasing availability of customer data. Researchers and financial institutions have explored various techniques for analyzing banking datasets to understand customer behavior, financial trends, and risk patterns. Traditional banking analysis methods relied mainly on manual record examination and basic statistical techniques. These approaches were useful for generating reports and summaries but lacked the capability to process large datasets efficiently. As banking data continued to grow in size and complexity, more advanced data analysis tools became necessary. Recent studies have focused on using data analytics and machine learning techniques to improve banking customer analysis. Tools such as Python, R, and data mining algorithms are widely used to analyze large financial datasets. These techniques allow researchers to identify patterns in customer demographics, spending behavior, and credit risk.

1. A Study on Traditional Banking Data Analysis Methods : Traditional banking data analysis systems mainly relied on manual examination of financial records and basic statistical techniques. Banks maintained customer records in databases and generated reports based on simple calculations and summary statistics. Analysts used these reports to understand customer demographics and account activities.

2. A Study on Data Analytics and Visualization Techniques in Banking : Recent research focuses on using data analytics and visualization techniques to analyze banking customer data more effectively. Programming languages such as Python provide powerful libraries for data processing, statistical analysis, and visualization. Tools like Pandas and NumPy are used for handling large datasets, while Matplotlib and Seaborn help create graphical representations of data

III SYSTEM ANALYSIS

The banking customer analysis system is designed to help banks understand customer behavior, preferences, and risk profiles by analyzing large datasets using Jupyter Notebook. This system collects customer information such as demographics, transaction history, account balances, and product usage. Using data analysis libraries like Pandas, NumPy, and visualization tools like Matplotlib and Seaborn, it identifies patterns, trends, and correlations among customers. The system enables predictive analytics to forecast customer churn, credit risk, and product demand. It provides insights that support strategic decision-making, marketing campaigns, and personalized customer services. Overall, this system aims to improve customer satisfaction, reduce risk, and increase profitability for banks by leveraging data-driven insights efficiently.

Existing system

In the existing banking system, customer analysis is primarily performed manually or using basic spreadsheet tools. Bank employees collect customer data such as account details, transaction history, and personal information, then analyze it using simple calculations or static reports. The process is time-consuming, prone to human errors, and often lacks depth in



identifying hidden patterns or predicting future trends. Insights derived from this approach are limited and reactive, making it difficult for banks to offer personalized services, detect potential risks, or make data-driven decisions efficiently.

Disadvantages of existing system

- Manual analysis is time-consuming and inefficient.
- High risk of human errors in calculations and reporting.
- Limited ability to identify hidden patterns in customer behavior.
- Reactive decision-making due to lack of predictive insights.
- Difficulty in personalizing services for individual customers.

Proposed system

The proposed system for banking customer analysis uses Jupyter Notebook with advanced data analysis and visualization tools to automate and enhance the process. Customer data is collected from various sources and processed using Python libraries such as Pandas, NumPy, and Scikit-learn for analysis, while Matplotlib and Seaborn help visualize trends and patterns. This system enables predictive analytics, such as forecasting customer churn, credit risk assessment, and identifying product preferences. By automating data processing and providing interactive visualizations, the system allows banks to make faster, accurate, and data-driven decisions. It improves customer segmentation, personalizes services, reduces operational errors, and supports strategic planning, ultimately enhancing customer satisfaction and business profitability.

Advantages of proposed system

- **Faster and more efficient** data processing using automation.
- **Reduces human errors** in analysis and reporting.
- **Identifies hidden patterns** and trends in customer behavior.
- Enables **predictive analytics** for churn, credit risk, and product demand.
- Supports **personalized services** and better customer segmentation.

IV METHODOLOGY

The methodology explains the steps followed to perform banking customer data analysis using Jupyter Notebook and Python. The project follows a structured process to collect, process, analyze, and visualize banking customer data.

1. **Data Collection** The first step is collecting the banking customer dataset. The dataset may contain details such as customer ID, age, gender, account balance, transactions, and loan information. The data is usually stored in CSV or Excel format and loaded into the system using Pandas.

2. **Data Pre-processing** The collected data may contain missing values, duplicate entries, or incorrect data formats. In this stage, the data is cleaned and prepared for analysis by:

- Removing duplicate records
- Handling missing values
- Converting data into the correct



format • Filtering unnecessary data This step ensures the dataset is accurate and suitable for analysis.

3. Data Exploration In this stage, the dataset is explored to understand its structure and characteristics. Basic statistical analysis such as mean, count, and distribution of data is performed. This helps in identifying patterns and relationships within the customer data.

4. Data Analysis After exploring the data, different analytical techniques are applied to study customer behavior. The analysis focuses on: • Customer segmentation based on age or balance • Transaction behavior analysis • Identifying high-value customers • Observing trends in banking activities Numerical computations are supported using NumPy.

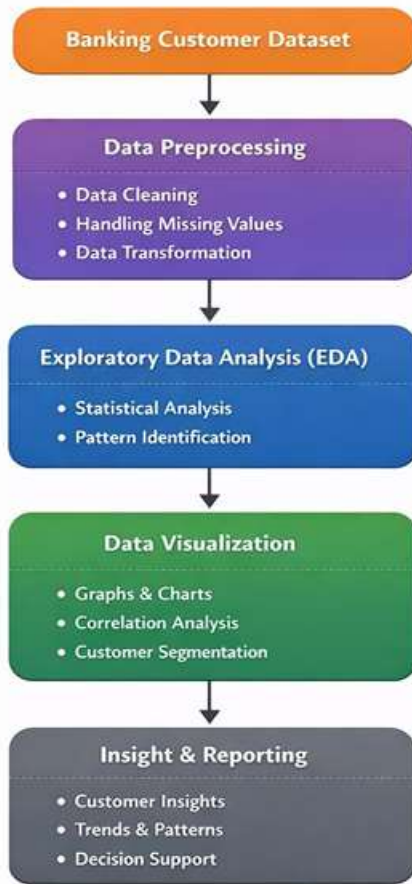
System Architecture

The system architecture of Banking Customer Analysis describes how customer data is collected, processed, analysed, and visualized to generate useful insights. The project mainly uses Jupyter Notebook with Python for performing data analysis.

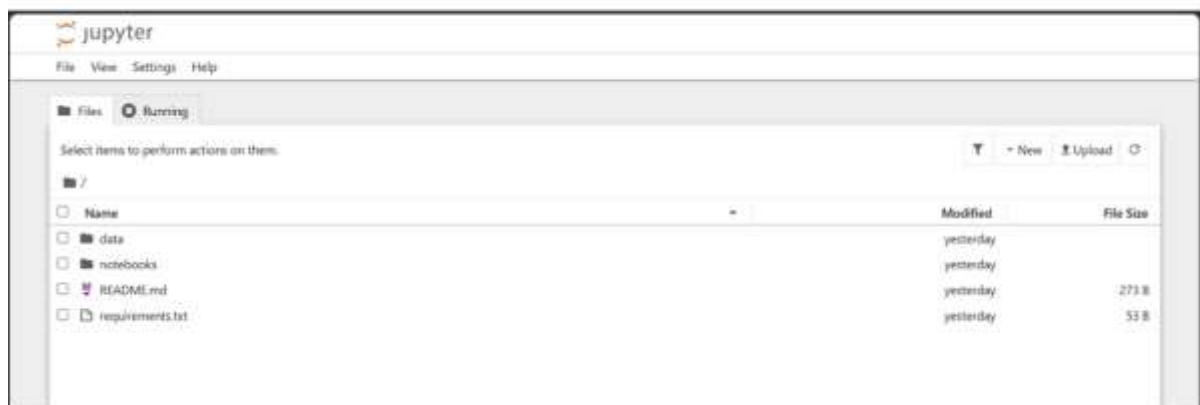
1. Data Collection Layer In this layer, the banking customer dataset is collected from sources such as CSV files, databases, or banking records. The dataset may contain information like: • Customer ID • Age • Gender • Account balance • Transaction details • Loan information This data is loaded into the system using Pandas.

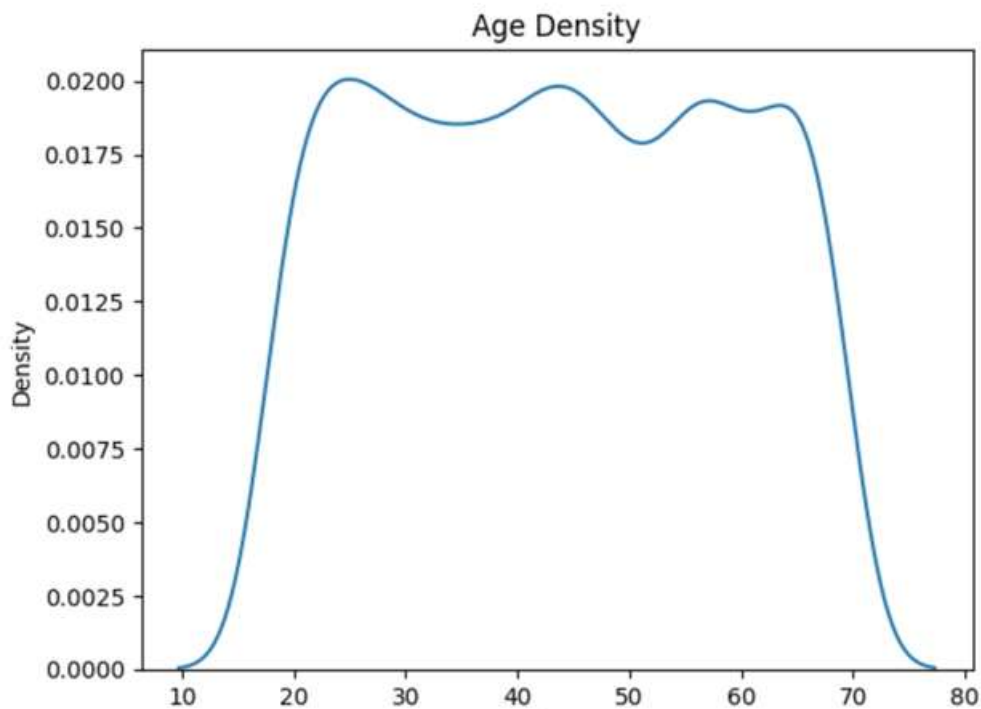
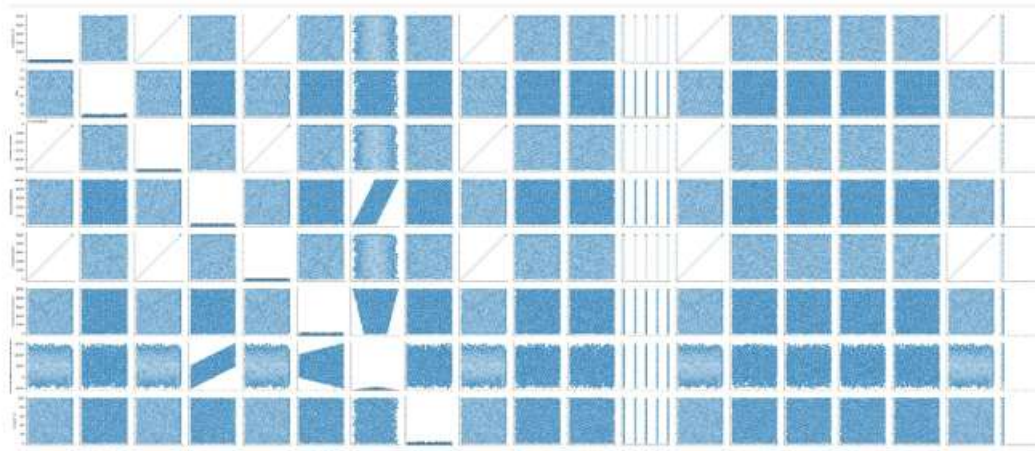
2. Data Pre-processing Layer Raw data often contains missing values, duplicate records, or incorrect data. In this stage: • Missing values are handled • Unnecessary columns are removed • Data is cleaned and formatted Libraries like NumPy and Pandas are used to prepare the data for analysis

3. Data Analysis Layer After cleaning the data, different analysis techniques are applied to understand customer behaviour.



V RESULTS&OUTPUT





VI CONCLUSION

The Banking Customer Analysis System successfully demonstrates how data analysis and visualization techniques can be used to extract meaningful insights from banking customer datasets. By using Python programming and powerful data analysis libraries, the system processes large amounts of customer data and presents useful information in a clear and understandable format. The project focuses on analyzing important customer attributes such as age, gender, credit score, account balance, and estimated salary. Through data preprocessing and exploratory data analysis, the system identifies patterns and trends present in the dataset. The use of graphical visualizations such as histograms,



bar charts, and correlation heatmaps helps represent complex data in an easy-to-understand manner. The implementation of the project using Python, Pandas, NumPy, Matplotlib, and Seaborn provides an efficient environment for performing data analysis tasks. These tools allow the system to handle large datasets, perform statistical analysis, and generate meaningful visual insights. Overall, the Banking Customer Analysis System provides a simple and effective approach for analyzing banking customer data. The project highlights the importance of data-driven analysis in the financial sector and demonstrates how modern data analytics tools can be used to improve understanding of customer behavior and financial trends

REFERENCE

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
- [7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.



- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.