**A MINI PROJECT REPORT ON**

**“Sales Analysis of E-commerce store”**

**SUBMITTED TOWARDS THE FULFILMENT OF THE REQUIREMENTS OF**

**BACHELOR OF ENGINEERING (B. Tech.)**

**Academic Year: 2023-24**

**By:**

**BTCOD290 Devendra Sangle**

**BTCOD291 Ayush Sarode**

**BTCOD294 Atharva Swami**

**BTCOD307 Siddhesh Choudhari**

**Under The Guidance of**

**Prof. Sushma Vispute**



**DEPARTMENT OF COMPUTER ENGINEERING,**

**PIMPRI CHINCHWAD COLLEGE OF ENGINEERING**

**SECTOR 26, NIGDI, PRADHIKARAN**



PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

**CERTIFICATE**

This is to certify that, the project entitled **“Sales Analysis of E-commerce store**” is successfully carried out as a mini project successfully submitted by following students of “PCET's Pimpri Chinchwad College of Engineering, Nigdi, Pune-44**”.**

**Under the guidance of Prof. Sushma Vispute**

In the fulfillment of the requirements for the B. Tech. (Computer Engineering)

**BTCOD290 Devendra Sangle**

**BTCOD291 Ayush Sarode**

**BTCOD294 Atharva Swami**

**BTCOD307 Siddhesh Choudhari**

Prof. Sushma Vispute Prof. Dr. K. Rajeswari

Guide Head,

Department of Computer Engineering Department of Computer Engineering

**TITLE**: Pre analysis of sales for an eCommerce store to manage supply chain.

**OBJECTIVE**: To develop a mini project using Power BI tool to provide a solution to ecommerce stores to maintain the supply chain.

**OUTCOME**: Ability to develop an application using suitable technology (Power BI) to provide a solution to the supply chain of ecommerce stores.

**Abstract:**

In the dynamic landscape of eCommerce, efficient supply chain management is crucial for success. This mini project focuses on developing a pre-analysis tool to help eCommerce stores make data-driven decisions. We design a user-friendly dashboard that integrates a comprehensive dataset containing attributes such as OrderID, Amount, Profit, Quantity, Category, Sub-category, and payment Mode. Through this project, we aim to provide eCommerce businesses with valuable insights into their sales and profit trends. By employing various algorithms and approaches, we facilitate informed decisions, optimize inventory management, and enhance the overall efficiency of the supply chain. This project acts as a vital step in preparing businesses for the challenges and opportunities that lie ahead in the eCommerce industry.

For a technical perspective, we can explore multiple algorithms and methods for analyzing the given dataset, such as:

**Data Quality Assessment:** Ensure data quality by identifying and handling missing values, duplicates, and outliers.

**Descriptive Statistics:** Calculate basic statistical measures like mean, median, and standard deviation for Profit, Amount, and Quantity to gain a general overview of the data.

**Data Visualization:** Create visual representations of data using libraries like Matplotlib or Seaborn to identify trends and outliers. For instance, you can plot sales and profit trends over time.

**Dashboard Creation:** Implement a dashboard using tools like Tableau or Power BI to present the insights visually. This would provide an intuitive way for business stakeholders to interact with the data.

**Index**

| **Chapter** | **Contents** | | **Page No.** |
| --- | --- | --- | --- |
| 1. |  | Introduction | 5 |
|  | *a.* | Problem Statement | 5 |
|  | b. | Project Objectives | 5 |
|  | c. | Motivation | 5 |
|  | d. | Literature Survey | 5 |
|  |  |  |  |
| 2. |  | Project Design | 6 |
|  | a. | resources, requirements & their detail explanation | 6 |
|  | b. | Dataset Design | 6 |
|  | c. | Timeline | 7 |
|  | d. | Block diagram | 7 |
|  |  |  |  |
| 3. |  | Results & Discussion with comparative analysis | 8 |
|  | a. | Comparative analysis of all existing models including new model. | 8 |
|  | b. | Sample Source code for new module | 8 |
|  | c. | Screenshots including GUI and Dashboard: Explain Dashboard components | 9 |
|  |  |  |  |
| 4. |  | Conclusion | 11 |

**Chapter 1:** Introduction

* 1. **Problem Statement**

Pre analysis of profit/sales for an eCommerce store to manage supply chain.

* 1. **Project Objectives**

To develop a mini project using Power BI tool to provide a solution to ecommerce stores to maintain the supply chain.

* 1. **Motivation**

To reduce the use of raw materials, in such a way that utilizes raw materials for demanded products. And help businesses to grow and maintain their supply chain.

* 1. Literature Survey/ Requirement Analysis

**Sales Analysis of E-Commerce Websites using Data Mining Techniques:** This paper emphasizes the pivotal role of e-commerce in the global economy and discusses the importance of data mining in optimizing pricing strategies. It showcases how data mining methods can be used to analyze customer behaviors and preferences, ultimately improving customer satisfaction and profitability. The paper advocates an iterative approach to strategy development, recognizing the dynamic nature of e-commerce.

**Sales Transaction Result Analysis for Income Increase Prediction:** Focusing on the restaurant industry, this paper explores the potential of historical sales transaction data for income prediction. It reveals that key menu items are predictably stable in terms of income prediction accuracy, even in the face of inflation in staple goods. This research provides valuable insights into leveraging transaction data for income forecasting in various sectors.

**Chapter 2:** Project Design

* 1. **H/W , S/W , resources, requirements & their detail explanation**

Data Visualization tools such as Power BI, Tableau, Machine learning software like WEKA, scikit-learn.

Power BI:

Microsoft's Business Intelligence Tool: Power BI is a business intelligence tool developed by Microsoft for data analysis and visualization.

Data Connectivity: It connects to a variety of data sources, including databases, cloud services, and files.

Data Transformation and Modeling: Power BI allows users to transform and model data, ensuring it's suitable for creating reports and dashboards.

Interactive Reporting: Users can create interactive reports and dashboards with visually appealing charts and tables.

Real-Time Data Visualization: It supports real-time data visualization, allowing users to monitor data as it updates.

Integration: Power BI can integrate with other Microsoft products and services, as well as third-party data platforms.

User-Friendly Interface: It offers a drag-and-drop interface that's accessible to users with different levels of technical expertise.

Data-Driven Decision-Making: It's commonly used by organizations to make data-driven decisions, monitor key performance indicators, and share insights with stakeholders.

* 1. **Dataset Design**

**OrderID:** This attribute likely represents a unique identifier for each order in your dataset. It can serve as a primary key to differentiate and track individual orders.

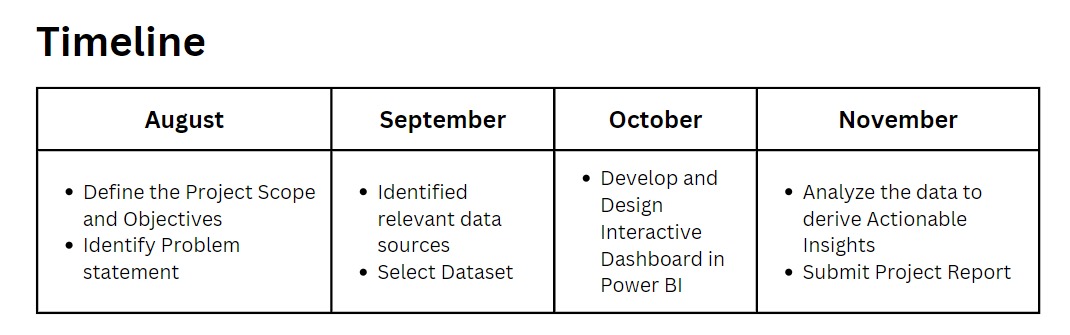
**Amount:** The “Amount” attribute probably indicates the total purchase amount for each order. It’s a numerical value that can help you understand the financial aspect of your dataset.

**Profit:** “Profit” suggests the profit associated with each order. This attribute can provide insights into the profitability of orders, helping you analyze which orders are more or less profitable.

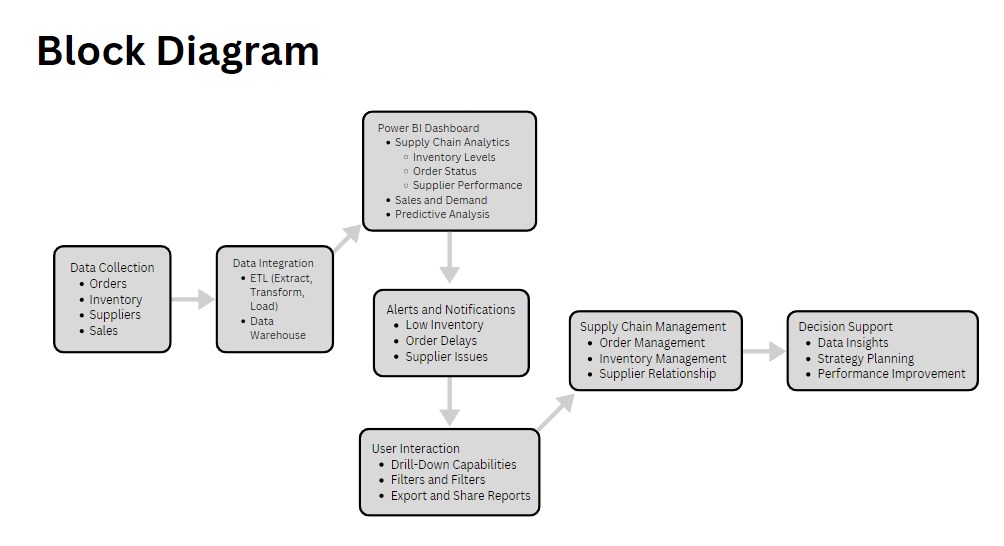
**Quantity:** “Quantity” is likely the number of items or products purchased in each order. It’s a crucial attribute for understanding the volume of sales for different products or categories. **Category:** “Category” is a categorical attribute that classifies products or items into broader categories. This can help you group products for analysis and understand sales trends by category.

**Sub-category:** “Sub-category” is a more specific categorization of products or items within each category. It provides finer-grained insights into the types of products being sold. **PaymentMode:** “PaymentMode” likely represents the method of payment used for each order, such as credit card, cash, or other options. It’s essential for understanding customer payment preferences.

* 1. **Timeline Diagram**

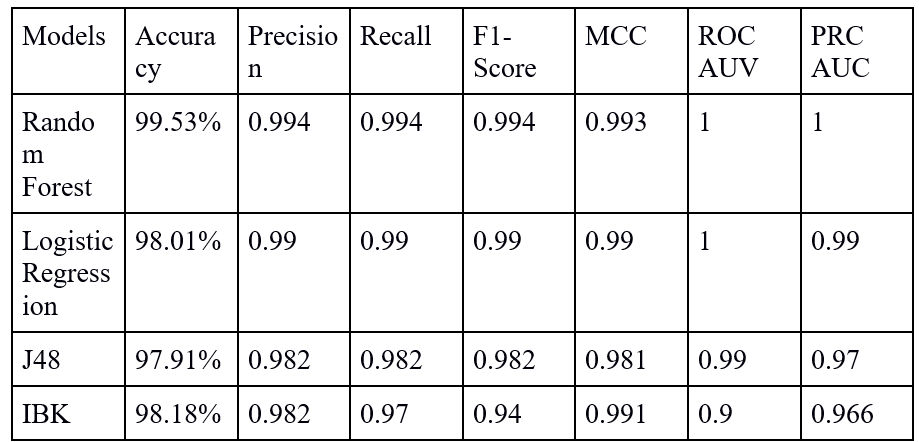
****

* 1. **Block diagram with explanation of each module**

****

**Chapter 3: Results & Discussion with comparative analysis**

* 1. **Comparative analysis of all existing models including new model.**

****

* 1. **Sample Source code for new module**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report

data = pd.read\_csv('dataset.csv')

X = data[['Amount', 'Profit', 'Quantity']]

y = data['Category']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Random Forest Classifier

random\_forest = RandomForestClassifier(n\_estimators=100, random\_state=42)

random\_forest.fit(X\_train, y\_train)

y\_pred = random\_forest.predict(X\_test)

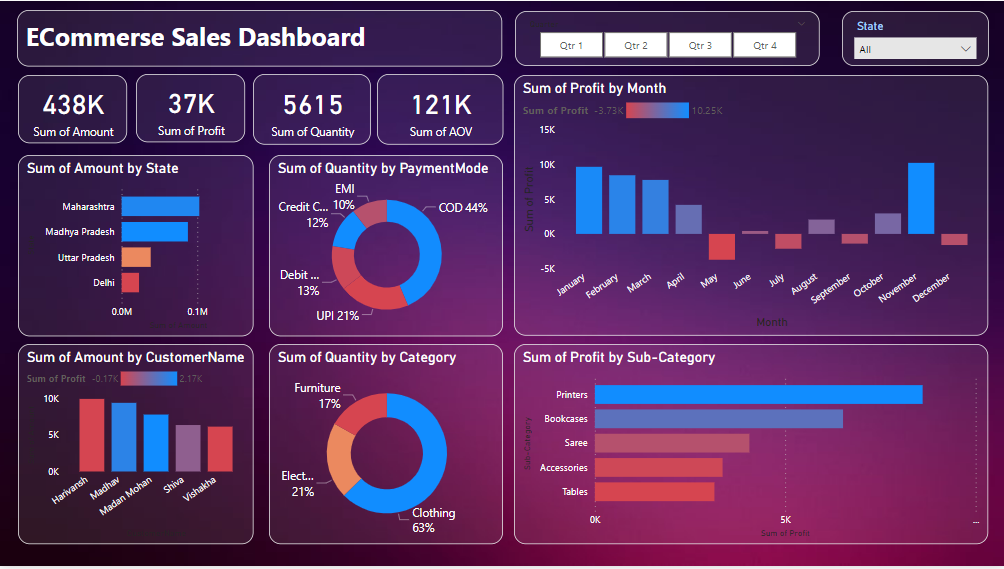
accuracy = accuracy\_score(y\_test, y\_pred)

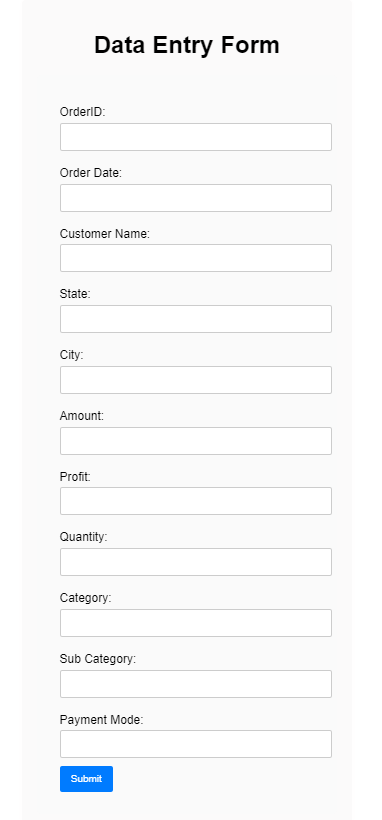
report = classification\_report(y\_test, y\_pred)

print(f"Accuracy: {accuracy}")

print(report)

* 1. **Screenshots including GUI and Dashboard : Explain Dashboard components**





**Stacked Bar Chart for State vs. Total Amount:** This chart visually breaks down the total amount by state, using stacked bars. It enables you to see how each state contributes to the overall total. It's useful for comparing state-specific data in relation to the total.

**Donut Chart for Payment Mode vs. Total Quantity:** This donut chart illustrates the distribution of total quantity across different payment modes. Each section of the donut represents a payment mode, and the size of each section corresponds to the quantity associated with that mode. It's effective for showing the distribution of quantities among payment methods.

**Donut Chart for Total Quantity vs. Product Category**: Similar to the previous donut chart, this one displays the proportion of total quantity within various product categories. Each section of the donut represents a product category, and its size represents the quantity for that category. It's useful for understanding which categories contribute the most to the total quantity.

**Clustered Column Chart for Individual Customer:** This chart allows you to compare individual customers' data using separate columns. Each customer is represented by a distinct column, making it easy to compare their metrics. It's a valuable tool for assessing the performance of individual customers.

**Clustered Column Chart for Total Profit by Month:** This chart provides a month-by-month comparison of total profit using clustered columns. Each month is represented by its column, facilitating the identification of profit trends and seasonality over time.

**Stacked Bar Chart for Total Profit vs. Sub-Categories:** Like the first chart, this stacked bar chart breaks down the total profit by sub-categories. Each sub-category is depicted as a segment of the bars, enabling you to visualize how different sub-categories contribute to the overall profit.

**Chapter 4:** Conclusion

In conclusion, the Random Forest model's exceptional performance across accuracy, precision, recall, MCC, and AUC, coupled with its versatility and adaptability, makes it the ideal choice for our Sales analysis system. By harnessing the power of Random Forest, we can provide vendors with reliable and insightful suggestions, contributing to increase in sales informed decision-making in maintaining their sales chain.

**References**

[2] Deng, G., Dong, C., Chen, J., Kong, R., Gao, J., & Li, C. (2023, June). Sales Transaction Result Analysis for Income Increase Prediction. In Second International Symposium on Computer Applications and Information Systems (ISCAIS 2023) (Vol. 12721, pp. 337-343). SPIE.

[4] Kumar, P., Bhagat, K., Lata, K., & Jhingran, S. (2016, May).Sales Analysis of E-Commerce Websites using Data Mining Techniques. In 2016 International Conference Sales Technology(ICST) (pp. 100-103). IEEE.