Lab Report on Analysis of Purchase Data and IRCTC Stock Data

P Praneeth Reddy

Department of Computer Science and Engineering
Amrita Vishwa Vidhyapeetam, Bengaluru
bl.en.u4cse21142@bl.students.amrita.edu

Abstract **- In this report, we understand consumer purchasing patterns and focus on markets with high demand, a quantitative study is essential. This paper addresses stock market analysis, which will help clients choose equities with high potential for profit. We analyze data from the "Purchase Data" and "IRCTC Stock Price" We will perform various tasks including data segregation, matrix operations, statistical analysis, and classification.**

I. INTRODUCTION

This lab assessment focuses on harnessing data analysis techniques to gain valuable insights from real-world data sourced from the "Purchase Data" and "IRCTC Stock Price" worksheets.Through this analysis, we aim to showcase our ability to manipulate data, perform statistical calculations, and apply classification models effectively.In this report, we will detail our approach and findings, demonstrating the practical application of data analysis in solving real-world problems.

II. Data Description

1) Purchase Data: This dataset consists of information related to customer transactions. It serves as the foundation for our analysis of customer behavior and product cost prediction.

2) IRCTC Stock Price Data: This dataset contains information related to the stock prices of a company. This dataset forms the basis for our stock price analysis.

III. Methodology

Data Segregation and Matrix Operations
We segregated the purchase data into two matrices, A and C, as per the instructions. Matrix A represents a set of data with specific characteristics, and Matrix C represents another set, with potentially different attributes.

Dimensionality and Number of Vectors

The dimensionality of this vector space is determined by the number of columns in each matrix, representing the attributes of the data.

Rank Calculation for Matrix A

The rank of Matrix A provides insights into the fundamental structure of the data, indicating whether there are any redundant or dependent variables among the attributes.

Using Pseudo-Inverse for Cost Calculation

By applying the Pseudo-Inverse, we obtained these coefficients, which enable us to estimate product costs based on the given data.

Customer Classification

Customers were categorized as RICH or POOR based on their payment behaviour. Those with payments above Rs. 200 were labelled as RICH, while others were classified as POOR.

Calculating Mean and Variance of Price Data

The mean provides a measure of central tendency, while the variance quantifies the spread or dispersion of the data,

Calculating Probability of Making a Loss

we determined the probability of making a loss in the stock. This involved identifying negative Chg% values, which indicate a decline in stock value.

Probability of Making a Profit on Wednesdays

We calculated the probability of making a profit on Wednesdays by considering positive Chg% values for Wednesdays in the dataset. This probability provides insights into the likelihood of profitable trades on Wednesdays.

Scatter Plot of Chg% Data

A scatter plot of the Chg% data against the day of the week. This visualization helps us understand how stock price changes vary throughout the week.
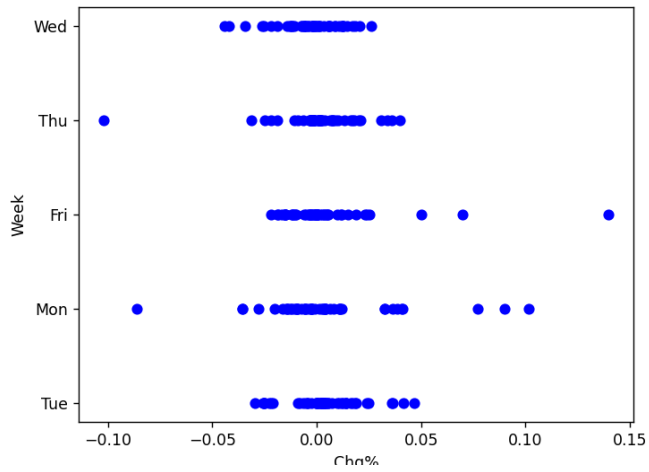
Fig. 1 Scatter Plot of Chg% Data

IV. Conclusion

Building a stock price prediction system is a complex task that requires careful data handling, model selection, and ongoing monitoring. It's essential to keep in mind that predicting stock prices is inherently uncertain, and no model can guarantee accurate predictions at all times.