**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Rishi kumar 2. Email ID:- politerishi@gmail.com 3. **Contribution : -  All  Self by me**   **Keywords:**  Data Collection and Preprocessing  - Importing important libraries and modules - Data Cleaning - Missing Data Handling - Merging the Datasets  Exploratory Data Analysis  - Categorical Features - Numerical Features - EDA Conclusion  Feature Selection and Outlier Detection  - Feature Engineering - Outlier Detection and Treatment  Modeling  - Train Test Split , Linear , Polynomial , Lasso , Ridge , Decision Tree,  Feature Importance |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/politerishi/Retail-Sales-Prediction-main-.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  **Sales forecasting refers to the process of estimating demand for or sales of a particular product over a specific period of time. Businesses use sales forecasts to determine what revenue they will be generating in a particular timespan to empower themselves with powerful and strategic business plans. Important decisions such as budgets, hiring, incentives, goals, acquisitions and various other growth plans are affected by the revenue the company is going to make in the coming months and for these plans to be as effective as they are planned to be it is important for these forecasts to also be as good. The work here predicts the sales for a drug store chain in the European market for a time period of six weeks and compares the results of different machine learning algorithms**  **Understanding the Data:**  **First step involved is understanding the data and getting answers to some basic questions like; What is the data about? How many rows or observations are there in it? How many features are there in it? What are the data types? Are there any missing values? And anything that could be relevant and useful to our investigation. Let’s just understand the dataset first and the terms involved before proceeding further. Our dataset consists of two csv files, the first consists of historical data with 1017209 rows or observations and 9 columns with no null values. The second dataset was supplementary information about the stores with 1115 rows and 10** **columns and a lot of missing values in a few columns. The data types were of integer, float and object in nature.** |
|  |