|  |  |
| --- | --- |
| Date | 11-10-2023 |
| Team ID | 8941 |
| Project Name | Product Sales Analysis |

Phase2 Project Submission

ABOUT DATASET:

REC corp LTD. is small-scaled business venture established in India. They have been selling FOUR PRODUCTS for OVER TEN YEARS. The products are P1, P2, P3 and P4. They have collected data from their retail centers and organized it into a small csv file, which has been given to you. **The excel file contains about 8 numerical parameters:**

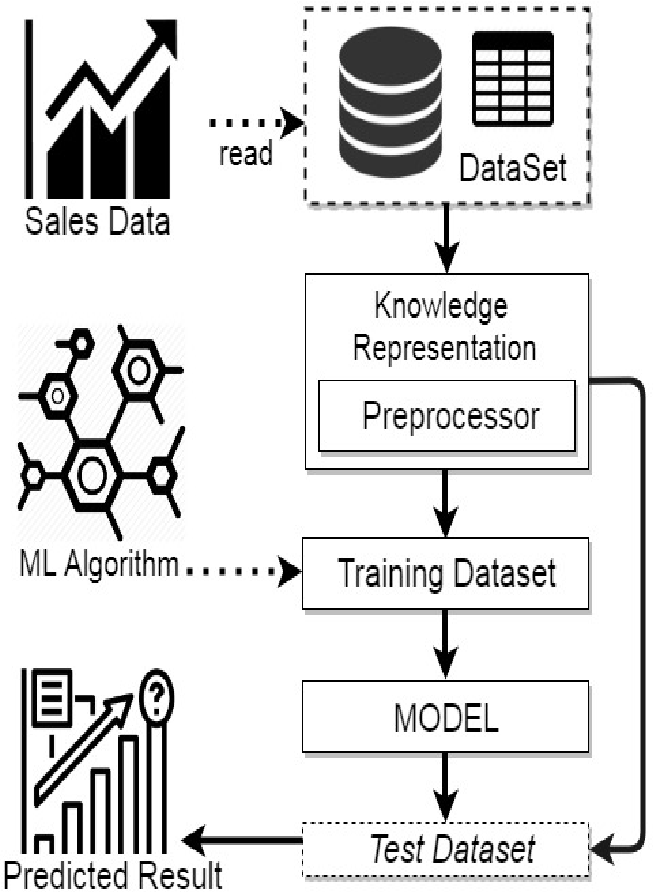
* Q1- Total unit sales of product 1
* Q2- Total unit sales of product 2
* Q3- Total unit sales of product 3
* Q4- Total unit sales of product 4
* S1- Total revenue from product 1
* S2- Total revenue from product 2
* S3- Total revenue from product 3
* S4- Total revenue from product 4

**Dataset Link:**[**https://www.kaggle.com/datasets/ksabishek/product-sales-data**](https://www.kaggle.com/datasets/ksabishek/product-sales-data)

INTRODUCTION:

ML is one of the most powerful tools available to sales and marketing professionals today. By analyzing large amounts of data, it can help you uncover patterns that would be otherwise hidden, and predict customer behavior with amazing accuracy. This has a lot of potential applications in sales and marketing. For example, you could use machine learning for [sales forecasting](https://www.akkio.com/applications/forecasting?utm_source=Akkio&utm_medium=content-marketing&utm_content=ml-sales), [predicting time-to-close](https://www.akkio.com/applications/predict-time-to-close?utm_source=Akkio&utm_medium=content-marketing&utm_content=ml-sales), predicting which customers are most likely to [respond positively](https://www.akkio.com/applications/augmented-lead-scoring?utm_source=Akkio&utm_medium=content-marketing&utm_content=ml-sales) to a new product or service, and more.



ARCHITECTURE:

DATA COLLECTION AND PREPROCESSING:

Importing the dataset: Obtain a comprehensive dataset containing relevant features such as sales of product and revenue of products etc.

Data preprocessing: Clean the data by handling missing values, outliers and categorical variables. Standardize or normalize numerical features.

EXPLORATORY DATA ANALYSIS (EDA):

* Visualize and analyze the dataset to gain insights into the relationships between variables.
* Identify correlations and patterns that can inform feature selection and engineering.
* Present various data visualizations to gain insights into the dataset
* Explore correlations between features and the target variable(sales trend)

MACHINE LEARNING METHODS FOR SALES PREDCITION:

One of the most common methods used to predict sales is **regression analysis.** This method involves using historical sales data to train a model that can predict future sales. The model can take into account factors such as **past sales, marketing campaigns, and economic indicators** to make its predictions.

Another popular method for predicting sales is **time series analysis**. This method involves using historical sales data to identify patterns and trends in sales over time. The model can then use these patterns to make predictions about future sales. This method is particularly useful for predicting sales in seasonal industries, such as retail and tourism.

Another approach is using **decision tree-based algorithms** like **Random Forest, Gradient Boosting** etc. These algorithms are particularly useful when there are many factors that can influence sales, such as product features, customer demographics, and market conditions. The algorithm can help identify the most important factors and use them to make predictions.

In addition to these methods, machine learning can also be used to predict sales through the use of **neural networks.** Neural networks are a type of machine learning algorithm that can learn to recognize patterns in data. They can be trained on large amounts of sales data and can make predictions about future sales.

Machine learning can also be used to predict sales by using **clustering algorithms,** which can help identify groups of similar customers. This information can then be used to create targeted marketing campaigns and improve sales strategies

MODEL EVALUATION AND SELECTION:

* Split the dataset into training and testing sets.
* Evaluate models using appropriate metrics (e.g., Mean Absolute Error, Mean Squared Error, R-squared) to assess their performance.
* Use cross-validation techniques to tune hyperparameters and ensure model stability.
* Compare the results with traditional linear regression models to highlight improvements.
* Select the best-performing model for further analysis.

DEPLOYMENT AND PREDICTION:

* Deploy the chosen regression model to predict sales trends.
* Develop a user-friendly interface for users to input property features and receive sales predictions

CONCLUSION AND FUTURE WORK:

In the Phase 2 conclusion, we will summarize the key findings and insights from the many machine learning algorithm. We will reiterate the impact of these techniques on improving the accuracy and robustness of sales predictions.

Future Work:

We will discuss potential avenues for future work, such as incorporating additional data sources, exploring or expanding the project with more features and interactivity.