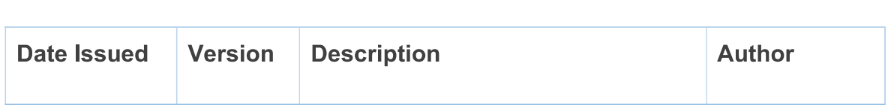


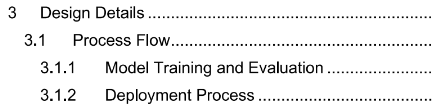
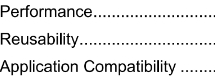
stores Sales Prediction

-Derrick T





|  |  |  |  |
| --- | --- | --- | --- |
| 19-07-2023 | 1.0 | Created | Derrick T |
|  |  |  |  |
|  |  |  |  |



2

4

5

5

5

5

5

5-6

6

6

6

6

6-7

7

7

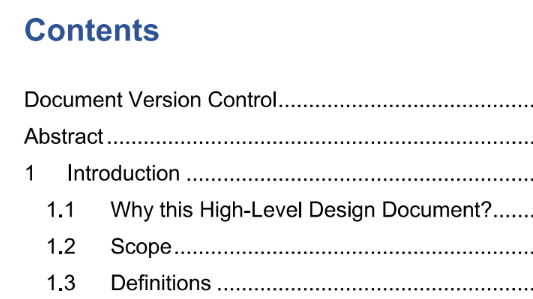
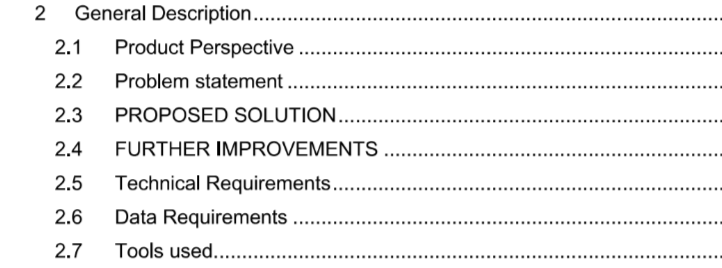
8

8

8

9

10-11



Abstract:

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.

This project focuses on predicting the sales based on multiple factors.



 The stores sales prediction solution is a basically a prediction model (Multiple ML models tried and tested). This solution will help us predict the sales for the upcoming years based on the data (past data).

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.



The propose solution is building a machine model that is trained using past data and evaluated, and the final evaluated model is fed in the unseen data to make predictions, in this case the sales amount is predicted.



Stores sales prediction can be made more accurate by reducing irrelevant data and adding relevant data which has higher probability of affecting the sales amount.



There were no technical requirements for developing this solution, the default requirements were a Personal Computer with good computing power.



The data required was based on the problem statement and the variables or the dependencies provided.

The client provides the dataset with relevant columns as mentioned in the problem statement.

After the solution is developed, user data of the user’s choice will be required.

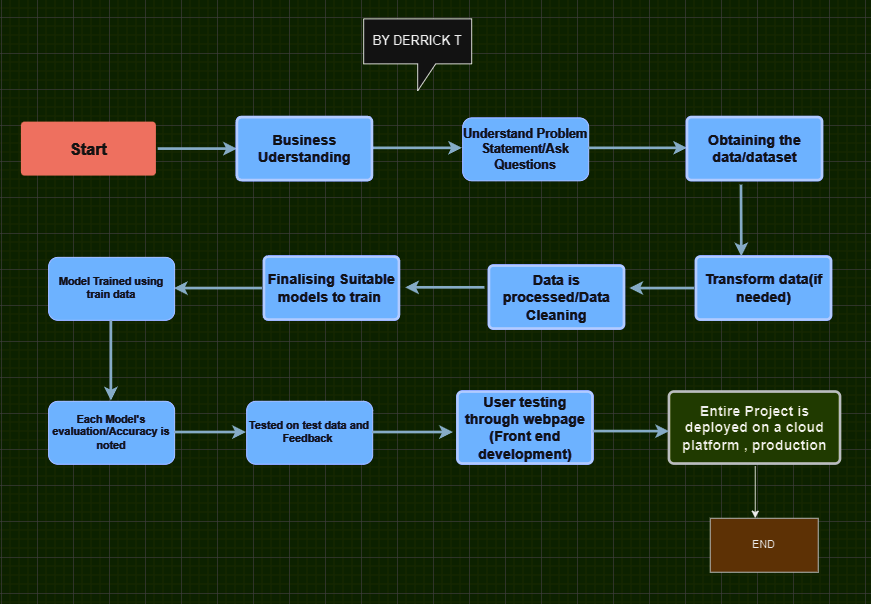






* Python was used as the programming language – to load, pre-process, scale, build ML model and for saving it.
* For visualization, data wrangling etc, frameworks such as NumPy, pandas, matplotlib, scikit learn were used
* PyCharm was used as the IDE for developing a flask application.
* Jupyter notebook as the IDE for developing the solution.
* Visual Studio Code to develop a webpage using HTML and CSS.

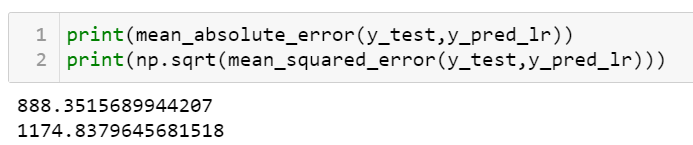




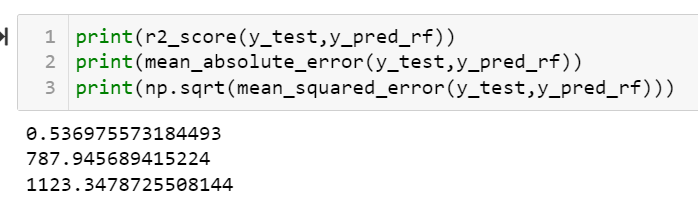


The model was trained using the train data and accuracy was observed on:

For linear regression model:



For random forest model:





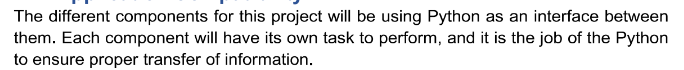
The entire solution was deployed on a local cloud platform ‘Render’ linked to GitHub repository which automatically lured the resources needed.



The solution developed should be very accurate as the sales might differ each time, all the factors when entered correctly by the user will be a step to provide accurate predictions.

Performance can be improved by training the model on new data and evaluating it often.















In this machine learning project, we set out to predict store sales using a dataset containing historical sales data, store information, and other relevant features. Throughout the project, we explored and implemented various machine learning techniques to build and evaluate predictive models. The main goal was to create a robust and accurate model that could assist store owners and managers in forecasting sales and making informed business decisions.

Key Findings:

1. Data Preprocessing: We started by carefully cleaning and preprocessing the dataset, handling missing values, and encoding categorical variables. Additionally, we performed feature engineering to extract relevant information from the available data, which helped improve the performance of the models.

2. Exploratory Data Analysis (EDA): Through EDA, we gained valuable insights into the relationships between different features and the target variable (store sales). This analysis enabled us to identify correlations, trends, and patterns, which guided our feature selection process and provided a better understanding of the data.

3. Model Selection: We experimented with various machine learning algorithms, including linear regression, decision trees, random forests. After thorough evaluation and comparison, we identified the best-performing models based on metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared.

4. Performance Evaluation: To ensure the reliability of our models, we utilized cross-validation techniques and performed hyperparameter tuning. This process allowed us to prevent overfitting and obtain models that generalize well to unseen data.

5. Model Interpretability: Understanding the factors influencing sales predictions is crucial for making informed business decisions.

6. Business Insights: With the deployment of our best-performing model, store owners and managers can now gain valuable insights into their sales forecasting. By using our predictive model, they can anticipate future sales trends, optimize inventory management, plan staffing requirements, and devise effective marketing strategies to drive revenue growth.

**Challenges Faced:**

1. Limited Data: One of the challenges we encountered was the availability of limited historical sales data. A larger dataset could potentially have improved model performance and generalization.

2. Seasonality and External Factors: Sales data in the retail industry can be influenced by various external factors such as seasonality, economic conditions, and marketing campaigns. Incorporating external data could further enhance the accuracy of our predictions.

3. Feature Engineering: Extracting the most informative features requires domain knowledge and expertise, and in some cases, obtaining additional external data sources may be necessary for more effective feature engineering.

Future Directions:

1. Time Series Models: To better capture the temporal aspects of the data, exploring time series models such as ARIMA, SARIMA, or Prophet could lead to more accurate predictions.

2. Integration of External Data: Incorporating external data sources, such as weather data, social media trends, or economic indicators, may help capture the influence of external factors on store sales.

3. Real-Time Predictions: Implementing the model in a real-time setting would enable store owners to receive up-to-date sales forecasts and respond promptly to changing market conditions.4. A/B Testing: Conducting A/B testing for different marketing strategies and promotions can provide valuable insights into their effectiveness on store sales and further optimize business operations.

In conclusion, our machine learning project on store sales prediction has demonstrated the potential of predictive modelling in the retail industry. By leveraging the power of data and machine learning, store owners and managers can make data-driven decisions, leading to improved efficiency, profitability, and customer satisfaction. Nevertheless, there are opportunities for further improvements and future research to enhance the accuracy and applicability of the models in real-world retail scenarios.