

1. Business Understanding

Walmart is an American multinational retail corporation and the world largest retailer, has been in existence since 1962. Over the years, Walmart has built a strong and influencing brand known for its low prices, wide selection of product and their commitment to customers. Walmart business model is based on providing values to its customer through cost savings and efficiency.

♣ Problem statement

Sales is a crucial part of every business and a major determinant of business growth. Inaccurate forecast of Walmart sales can lead to over stocking or stock out, inaccurate calculation of revenue and affect investment decision. More also, Accurate sales prediction can give Walmart an idea of future sales expectation and how projected targets can be achieved as not reaching target could negatively affect stock prices and will be a problem for Walmart.

♣ Objectives

The objective of this project is to build a model which can be used to predict weekly sales for Walmart stores and department, and be able to make better future plans, limit the problem of overstocking and understocking, achieving projected targets and deciding on the best investment.

2. Data Understanding

♣ Data Capture and explanation.

Dataset used for this project was gotten from Kaggle.com, and it's a weekly data of 45 Walmart stores with 81 department containing information of the stores including the weekly sales, markdown sales (sales on discount), and store type (the store was grouped into 3 types A, B and C depending on the size). The data is from 5th February 2010 to 26th October 2012.

♣ Data Exploration (clustering, visualization, etc.)

Visualization was done for the store weekly sales, department weekly sales, yearly weekly sales.

3. Data Preparation

♣ Data Preparation

The data set for this project is a combination of 4 CSV file, one containing the store data, another containing the features, another one containing the target and the last one containing the test. All

four datasets were imported into the jupyter notebook, accessed and combined into one dataframe, after combining all, we dropped the duplicated columns and renamed some features.

♣ **Feature Engineering**

Transform the categorical data (type) to numerical data, replacing A, B and C which represent the store type to 1 replacing A, 2 replacing B and 3 recreating C. To scale the data set, we used standard scaler.

4. Modeling and pipelines creation

♣ **Use of different regression models**

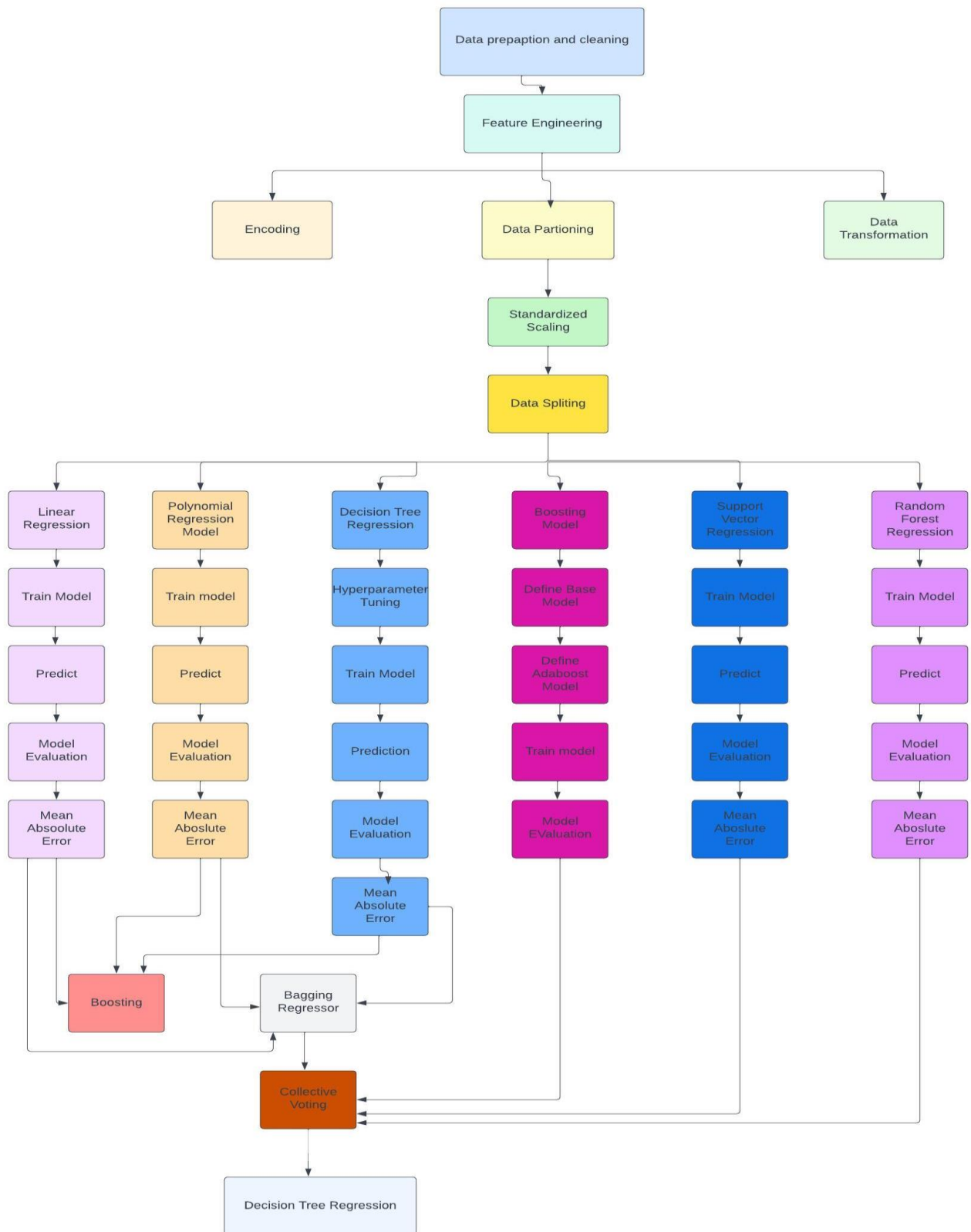
We used Regression models because it's a regression problem. The models used were the following:

- **Linear Regression Model**
- **Decision Tree Model**
- **Polynomial Regression Model**
- **Gradient Boosting Regressor**
- **Support Vector Regression**
- **Random Forest Regression**

♣ **Ensemble learning**

- **Boosting (Decision Tree Regressor – Linear Regression – Polynomial)**
- **Bagging (Decision Tree Regressor – Linear Regression – Polynomial)**
- **Stacking (Decision Tree Regressor – Linear Regression – Gradient Boost Regressor)**

♣ Pipelines creation and rationale behind it



5. Model Evaluation and Selection

♣ Evaluation metrics

Cross Validation was performed to evaluate and select the best model amongst all the models we have used.

Models	Mean Absolute Error
Linear Regression	21719.46
Decision Tree	42.55
Random Forest	1414.39
Support Vector Regression	1414.40
Gradient Boosting	8017.32

♣ Results interpretation

All the models performed with varying mean absolute errors. Decision Tree has the least mean average errors while Linear Regression, Random Forest and Support Vector had the highest mean average errors.

♣ Model Selection analysis

The best model is **Decision Tree model** because it had the least mean average error over the other models.

6. Conclusion

♣ Results analysis

Overall, the project shows that Walmart weekly sales can be predicted using machine learning algorithms, it is important to choose an appropriate model as in this case **Decision Tree model**.

♣ Recommendations

We recommend the following actions:

- ✓ From the analysis it was seen that week 51 had higher sales compared to the other weeks, and that is as a result of Christmas and new year celebration, so Walmart should stock up to avoid understocking
- ✓ **Decision Tree algorithm** should be used in predicting Walmart weekly sales.