CSE523 Machine Learning
Weekly Project Report

Group 10

Project title: Big Mart Sales Prediction

1. **Task performed in this week**

   - We found a dataset for our project from Kaggle.
   - We researched the four methods that we are going to use for the sales prediction:
     1. Linear Regression
     2. Random Forest Regression
     3. XGBoost
     4. SARIMAX
   - We read different articles and papers on these methods.
   - We explored various tools for the project such as Jupyter notebook, Google CoLab, PyCharm.

2. **Outcomes of task performed**

   - The dataset we finalised is the following:
     https://www.kaggle.com/datasets/shivan118/big-mart-sales-prediction-datasets
   - The following are the references for the articles of the methods:
     - https://www.jetir.org/papers/JETIREJ06034.pdf
     - https://www.researchgate.net/publication/356824621_Comparative_Analysis_of_Supervised_Machine_Learning_Techniques_for_Sales_Forecasting
     - https://dl.acm.org/doi/10.1145/3357254.3357290
   - We finalised the approach for implementation of our project as the following steps:
     1. Data Structure and Content
     2. Exploratory Data Analysis
     3. Missing Value Treatment
     4. Feature Engineering
     5. Encoding Categorical Variables
     6. Label Encoding
     7. Pre-Processing Data

8. Modeling

9. Linear Regression

10. Random Forest Regressor

11. XGBoost

12. SARIMAX

13. Deployment

3. Tasks to be performed in the upcoming week

- We will implement the data analysis on the dataset and perform necessary cleaning.
- Then we will apply feature engineering on the data and finalise the number of features.
- Then we will try to model the linear regression method and analyse its results.