**PROJECT TITLE:**

Future sales prediction

**PROBLEM DEFINITION:**

* This paper proposes an ensembling of decision tree model based future sales pre -diction. In particular, we present a detailed description of feature engineeringnecessary to generate trainable parameters and then perform detailed model performance study with various state of the art decision tree based models. Our bestperforming model resulted in a RMSE of 0.87605 achieving a class rank of 6and Global Kaggle leaderboard rank of 62.

**DESIGN THINKING** :

**Define Objectives and Scope :**

Clearly outline the project's goals, what you want to predict and the time frame for predictions

**Data Sources Identification :**

Determine where your data will come from. This could include internal databases, external sources, customer surveys, or market research.

**Data Gathering :**

A Internal Data Collect historical sales data, including sales figures, product attributes, pricing, promotions, and customer information.

**Data Preprocessing :**

a Data Cleaning Clean and format the data, removing duplicates and irrelevant information.

**Data Integration :**

Combine data from various sources into a single dataset for analysis.

**Data Exploration :**

a Perform exploratory data analysis (EDA) to understand data distributions, correlations, and patterns.

**LITERATURE SURVEY:**

|  |  |  |
| --- | --- | --- |
| Title | Methodology | Outcome |
| Kagglefuture sales | 1. description 2. Data-Preprocessing 3. Feature Engineering | In this work we presented future sales prediction models based on decision treestructures. Ourevaluation showed the best performing model can be achieved through ensembling of LGBM andrandom-forestgiving equal weight to each. |
| Intelligent Sales Prediction Using Machine Learning Techniques. | 1. Logistic regression model 2. ARIMA, SVM 3. Random forest algorithm, neural network. 4. Decision Tree, XGBoost. | Sales forecasting is the process of predicting future sales. It is the vital part of the financial planning of the business. Most of the companies heavily depend on the future prediction of the sales. |
| Machine Learning Project on Sales Prediction | 1. Apache Spark 2. Spark SQL 3. Apache Spark MLLib 4. DataFrame-based API | In this project, looking at the various Stores Sales around the world are tasked with predicting their daily sales in advance. |
| A MACHINE LEARNING ANALYSIS OF ROSSMAN STORE SALES | 1. Data Exploration 2. Data Wrangling 3. Data Visualization | Sales predictive analysis utilizing machine learning algorithms is a useful tool that can help firms get insights into their sales trends and anticipate future sales outcomes. Businesses may create reliable predictive models using machine learning techniques such as linear regression and gradient boosting. |

**REFERENCES:**

[1]. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boostingsystem. In Proceedingsof the 22nd ACM SIG-KDD international conference on knowledge discovery and data mining,pages 785–794, 2016 .[2]. Avinash kumar, Neha Gopal & Jatin Rajput(2020). An Intelligent Model For Predicting the Sales of a Product.[3]. Zixuan Huo(2021) ‘Sales Prediction based on Machine Learning’, IEEE 2nd International Conference on ECommerce and Internet Technology (ECIT)