Abstract:

Retail establishments like Big Mart use data mining to anticipate customer demand and optimize inventory management. This paper introduces a predictive model using the Random Forest Regressor technique to forecast sales for businesses like Big Mart. The model offers increased precision and reliability in sales forecasting, surpassing existing approaches in this area.

Introduction:

Big Mart, a leading supermarket chain, has challenged Data Scientists to create a model that accurately predicts product sales at each store. They have used Kaggle sales data from multiple cities to identify key products and stores driving sales, ensuring business success.

Data Collection Module:

This module encompasses the preprocessing of raw data obtained from Big Mart, which involves handling missing data, anomalies, and outliers. Subsequently, an algorithm will be trained to build a model using this processed data. This module integrates three functions, serving to extract and transform data from one database into a suitable format.

Data Pre-Processing Module:

The dataset utilized comprises sales data from Big Mart, encompassing nine attributes including Item\_Type, Item\_MRP, Item\_Fat\_Content, Outlet\_Size, Outlet\_Type, Item\_Weight, Item\_Visibility, Outlet\_Establishment\_Year, and Outlet\_Location\_Type. Among these attributes, "Item\_Outlet\_Sales" serves as the target variable, while the remaining attributes function as independent variables. Data pre-processing involves preparing and adapting raw data for learning models, constituting a crucial initial step in constructing a machine learning model. Real-world data often contain noise, missing values, and may require formatting to render them suitable for machine learning applications.

Evaluation Module:

Machine learning model evaluation is crucial for its development, requiring the construction and use of various metrics to derive insights. The evaluation process is iterative, guided by metric enhancements, and the Root Mean Squared Error (RMSE) metric is used to differentiate between outcomes.

Prediction Module:

Our proposed approach involves implementing a predictive model utilizing the Random Forest technique to forecast sales for companies resembling Big Mart. Through our analysis, we have determined that this model exhibits superior performance compared to existing models.