TEBONG ROLAND

DSC680-T301 Applied Data Science (2243-1)

1.1 Project 1: Proposal and Data Selection/Milestone 1

**SuperKart Sales Prediction**

**Topic:**

"Data-Driven Sales Forecasting for SuperKart: Leveraging Machine Learning for Retail Success"

**Context:**

In the highly competitive retail industry, accurate sales forecasting is crucial for operational and strategic planning. SuperKart, a leading supermarket and food mart chain, aims to enhance its sales operations by accurately predicting future sales revenue. This project focuses on using historical sales data and store characteristics to forecast sales, enabling SuperKart to optimize inventory and strategize sales across different city tiers.

**Business Problem:**

The primary goal is to develop a predictive model that estimates the total sales for each SuperKart outlet for the upcoming quarter. This model will help the organization in strategic planning, inventory management, and optimizing sales operations in various city tiers. By understanding the key drivers of sales in different store types and locations, SuperKart can make informed decisions to boost profitability and customer satisfaction.

**Data Description:**

The dataset comprises various attributes of products and stores:

* Product\_Id - unique identifier of each product, each identifier having two letters at the beginning followed by a number.
* Product\_Weight - the weight of each product
* Product\_Sugar\_Content - sugar content of each product like low sugar, regular, and no sugar
* Product\_Allocated\_Area - the ratio of the allocated display area of each product to the total display area of all the products in a store
* Product\_Type - broad category for each product like meat, snack foods, hard drinks, dairy, canned, soft drinks, health and hygiene, baking goods, bread, breakfast, frozen foods, fruits and vegetables, household, seafood, starchy foods, others
* Product\_MRP - maximum retail price of each product
* Store\_Id - unique identifier of each store
* Store\_Establishment\_Year - the year in which the store was established
* Store\_Size - the size of the store depending on sq. feet like high, medium, and low
* Store\_Location\_City\_Type - the type of city in which the store is located like Tier 1, Tier 2, and Tier 3. Tier 1 consists of cities where the standard of living is comparatively higher than its Tier 2 and Tier 3 counterparts.
* Store\_Type - the type of store depending on the products that are being sold there like Departmental Store, Supermarket Type 1, Supermarket Type 2, and Food Mart
* Product\_Store\_Sales\_Total - total revenue generated by the sale of that particular product in that particular store

**Methodology:**

The approach will include exploratory data analysis (EDA) to understand trends and patterns, followed by feature engineering to extract meaningful information from the data. Various machine learning models like Linear Regression, Random Forest, and Gradient Boosting will be used for sales prediction. The model's performance will be evaluated using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² (Coefficient of Determination).

**Ethical Considerations:**

While working with sales data, it's essential to maintain customer privacy and data security. Any sensitive information must be anonymized, and data handling should comply with privacy laws and regulations. Additionally, the model should be free from biases that could lead to unfair advantages or disadvantages for certain stores or products.

**Challenges/Issues:**

Key challenges include dealing with missing values, high dimensionality, and the potential for overfitting. Additionally, external factors like economic trends, seasonal variations, and unforeseen events (e.g., pandemics) that impact sales should be considered.

**References:**

1. Bohanec, Marko, Mirjana Kljajić Borštnar, and Marko Robnik-Šikonja. "Explaining machine learning models in sales predictions." *Expert Systems with Applications* 71 (2017): 416-428.
2. Bajaj, Purvika, et al. "Sales prediction using machine learning algorithms." *International Research Journal of Engineering and Technology (IRJET)* 7.6 (2020): 3619-3625.
3. Odegua, Rising. "Applied Machine Learning for Supermarket Sales Prediction." *Project: Predictive Machine Learning in Industry* (2020).
4. Tsoumakas, Grigorios. "A survey of machine learning techniques for food sales prediction." Artificial Intelligence Review 52.1 (2019): 441-447.