Project Proposal

Group Details:

- Fayaz Shaik
- Mahesh Maddineni
- Nathaniel Yee
- Vivekananda Reddy Pyda

Problem Statement:

We propose a data-driven project to analyze US sales data, focusing on identifying customer purchase trends across various modes such as online, in-store, wholesale, and through distributors. By examining key parameters including purchase quantity, product price, delivery duration, discounts, and warehouse information, we aim to discern customer preferences and regional variations. Subsequently, feature engineering will enable the training of machine learning models for predictive and recommendation capabilities, guiding customers toward their preferred purchase mode based on relevant factors.

Motivation:

The above problem statement emphasizes the possible benefits for both companies and their customers. By analyzing and understanding customer purchase trends and preferences, companies can optimize their operations, inventory management, and marketing strategies. This can lead to increased customer satisfaction, improved efficiency, and potentially higher revenue. Additionally, providing customers with tailored recommendations for their preferred purchase mode can enhance their overall shopping experience, resulting in greater loyalty and engagement. The project aims to harness data-driven insights to create a win-win situation for both businesses and customers.

Data Mining Task

Type of Data: The data is presented in a comma-separated values format, featuring a combination of text, dates, and numerical information.

Task: We intend to explore various multi-class classification algorithms to address the problem, given the limited unique values in the target variable. The prominent algorithms in consideration include Naive Bayes, K-Nearest Neighbors, Stochastic Gradient Descent, Decision Tree, Random Forest, and Support Vector Machines.

Data:

The dataset can be accessed on Kaggle through the following link: https://www.kaggle.com/datasets/talhabu/us-regional-sales-data . We will be obtaining the data by downloading the CSV file provided at this location.

Schedule:

- Preprocessing (25th Sep – 1st Oct)
- Data Analysis (2nd Oct – 8th Oct)
- Feature Engineering (9th Oct – 15th Oct)
- Model Development (16th Oct – 22nd Oct)
- Mid Term Report (23rd Oct – 29th Oct)
- Recommendation System Development (30th Oct – 5th Nov)
- Final Testing and Validation (6th Nov – 12th Nov)
- Presentation Preparation (13th Nov – 19th Nov)
- Project Presentation (27th Nov)