15CSE401 - Machine Learning and Data Mining

Case Study - Project Proposal

Team Number: 3

Serial No.	Student Name	Roll Number
1	Dev Mithran J	CB.EN.U4CSE18015
2	Sowmiyanarayan S	CB.EN.U4CSE18053

Project Title: Store Item Demand Forecasting

Abstract:

The prosperity of a store lies not only in offering quality products to the customers, but also in organising product supply based on demand. A prudential supply management system would be one designed on the basis of well substantiated facts. To maximize the benefits in the longer run, one must collect data over a period and mine it to predict the demand trend. The data in the dataset depicts the purchase trend of objects from different stores for a period of 5 years. The objective here is to try and get the sales figures for the next couple of months using a predictive model.

Description:

The case study aims at maximising the profit as well as minimising the wastage in resources. Many-a-times the shop administrator tends to overbuy items that are purchased rarely and some of them expire which is a blow to sustainability as well as the store profits.

In order to avoid this, our model studies the recorded product purchase trends, thus understanding the needs of the people and then goes on to predict the future trends. This is a typical case of a regression problem. The veracity of the model's predictions can be evaluated by taking the deviation between the existing record and the values predicted. The lesser the deviation, the better our results will be.

For this purpose, we use the demand forecasting dataset from Kaggle which is available at Store Item Demand Forecasting Challenge. The dataset contains 9,13,001 examples each defined by four attributes. We proceed by reading the csv file, filling the missing values (if any) and condensing the data by aggregating them into months. We then proceed to scale the data, split the dataset into training and testing datasets, following which we train the model.

To achieve this, we plan to use Scikit Learn predominantly, followed by other libraries that enable us to use regression models. To better predict the final results, we plan to test the results of multiple prediction models against the ones available and choose the best model before moving on to the forecasts.