10/12/2015

PROJECT PROPOSAL

1. TITLE OF PROJECT AND GROUP INFORMATION
   - A. Project Title – Sales Forecast.
   - B. Group Information:
     i. Leader – Jeet Takwani
     ii. Members – Jeet Takwani, Hriya Maharaja.
     iii. Name – Miners
     iv. Id – 8.

2. PROBLEM AND GOAL:
   - A. What do you want to solve?
     i. Rossmann is a pharmaceutical store in Europe which has over 3000 stores. Their managers are tasked to predict daily sales. Given daily sales for 1115 stores located across Germany we hope to predict 6 weeks of daily sales.
   - B. Why do you think this is important?
     i. This is important because store sales are influenced by many factors such as location, promotions, marketing etc. Reliable sales forecast will help the stores to be more effective in terms of their staff schedules, promotional activities and managing inventory.
   - C. What Result do you expect?
     i. The result that we expect is better accuracy in predicting sales forecast for each day in next six weeks giving better result in terms of management staff schedules and promotional

activities. This will benefit the store in long run and the store managers.

3. FORMALIZE THE PROBLEM INTO DATA MINING TASK.

A. Which data type to be address?

i. The data type to be addressed here is Matrix data.
ii. There are about 1 million rows and 9 columns.
iii. These rows have data for all the 1115 stores. The data about daily sales is between 1st Jan 2013 to 31st July 2015 .

B. What data mining tasks are needed?

i. The data mining task that is needed to predict the sales forecast is to use various prediction algorithms on Matrix data.

4. DATA PLAN

A. What kind of data will be used?

i. The data type to be addressed here is Matrix data.
ii. There are about 1 million rows and 9 columns.
iii. These rows have data for all the 1115 stores. The data about daily sales is between 1st Jan 2013 to 31st July 2015 .
iv. The data is in csv file format.

B. Where and how do you get the data?

i. The problem is an open competition on Kaggle.com and data is obtained from that site.

C. Get data on time.

i. The data is already available on Kaggle as it is an active open competition.

5. DETAILED PLAN.

A. The plan:

i. First step is to analyze the problem and data given for the problem.
ii. There will be data pre processing if needed on the data.
iii. The second step will be to do research and learn various prediction algorithms.
iv. After learning and researching about prediction algorithms, the next step will be to code and develop different models for each algorithm
v. After using training and testing data and finding the prediction given by different algorithms, next step will be to evaluate the prediction and research why a particular algorithm predicted in a manner that it did.
vi. Based on the prediction we will choose the best algorithm with maximum accuracy in prediction.