**PROJECT PROPOSAL**

**Members in the group:**

|  |  |
| --- | --- |
| **Names** | **CWID** |
| **Ajinkya Bijwe** | **802332916** |
| **Purva Chakravarti** | **802344432** |
| **Gurpreet Kaur** | **891517658** |

**Project Name: Store sales estimation**

**Introduction**

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information

This project is about using data mining process for store sales estimation. Huge store salesdata will be analyzed along with the holiday markdown events of the store for its estimation.

**Problem Statement:**

* The sales of a company is very unpredictable. It depends mainly on the quality, quantity, prices of different products and of course high discounts in any classic event.
* The company needs some basic estimated plan for the sales improvement and get a proper growth in it from the markdown holidays’ and the everyday sales
* In order to get the store sales estimation perfectly, we need to go over through the historical sales data for many departments of a company.
* The holiday markdown events are also included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact
* Lack of awareness and knowledge about various prediction techniques, makes it difficult for companies to survive the competition.
* These prediction techniques should yield consistent and highly accurate results.

**Description:**

* The main objective of this paper is to analyze the historical sales data available for many departments of a company by exploring different data mining techniques.
* Data mining is a method of extracting unknown projecting information from large databases which is a widespread technology that helps organizations to focus on the most important information in data repositories with great potential.
* In this project, a historical dataset of sales of multiple departments of a company will be undertaken for our research.
* Different data mining tasks, techniques and tools will be used for the sales estimation from the extracted data of the dataset.
* The dataset will be stored on MS DOS database and will be extracted though a MS DOS object
* For prediction, the extracted data will undergo through various data mining tasks:
  + - Analysis of survey data
    - Explanatory simulation
    - Analytical modeling
    - Identify patterns and rules
    - Acquisition summary

**For backend operations**: MS DOS database

**Language to be used**: R programming language