Practical 1: Understanding Characteristics of Data Sources

GAHAN SARAIYA (18MCEC10), PRIYANKA BHATI (18MCEC02)

18mcec10@nirmauni.ac.in,18mcec02@nirmauni.ac.in

I.   AIM

Data Domain selection and Identification of Characteristics of selected Dataset of different formats.

II.   INTRODUCTION

Data mining can be performed on following types of data

- Relational databases

- Data warehouses

- Advanced DB and information repositories

- Object-oriented and object-relational databases

- Transactional and Spatial databases

- Heterogeneous and legacy databases

- Multimedia and streaming database

- Text databases

- Text mining and Web mining

This phase/practicals aims towards data understanding to check that business and data-mining goals are established as well as to make it appropriate for data-mining goals if applicable.

I.   Domain Identification

Data Mining can be used in diverse industries such as Retail market, Communications, Insurance, Education, Manufacturing, E-commerce, Banking, Bioinformatics, Crime Investigation etc. To achieve this goal this document of experiment targets on the mining process on Retail market .

Data Mining techniques help **retail** malls and grocery stores identify and arrange most sellable items in the most attentive positions. It helps store owners to comes up with the offer which encourages customers to increase their spending. Also Note that the same scenario is also applicable to local retail stores as well as e-commerce stores with slightly differ in presenting the recommendation to customers.

III.   CHARACTERIZATION OF DATASET

From the various Retail market events this experiment aims to a single event of data mining for **Black Friday Sale**.