Low Level Design

Google play store

Written By	Dhiraj Jesani & Dharvi Sanghavi
Last Revised Date	

……2 Low Level Design

1. Introduction

1.1 What is Data mining?

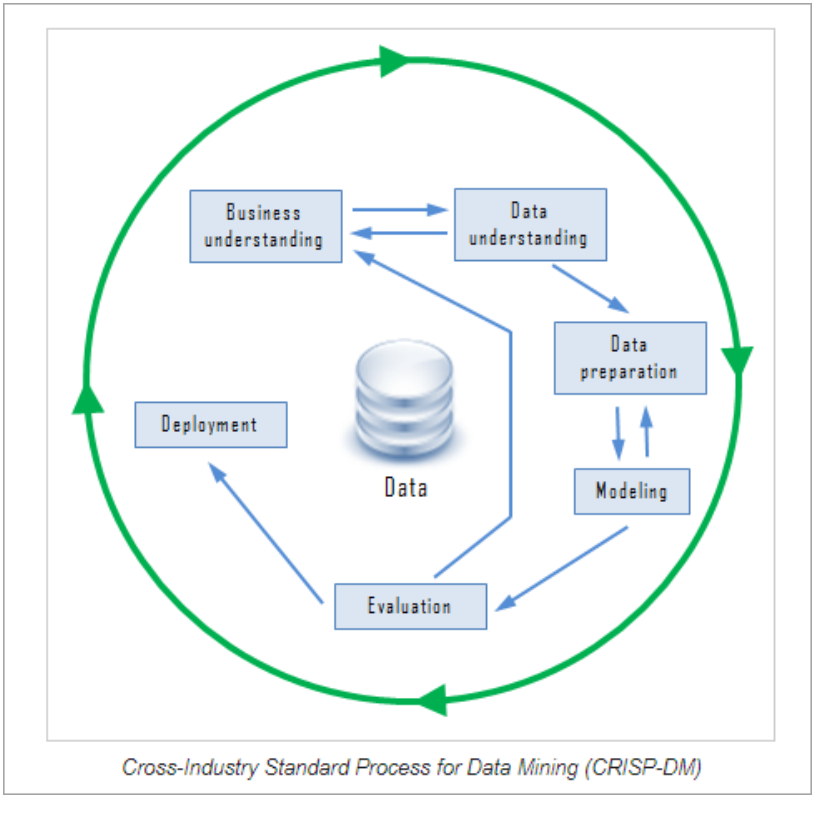
Data mining is a process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

1.2 Scope

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities.

.....4 Low Level Design

2. Architecture



Cross-Industry Standard Process for Data Mining (CRISP-DM)

3. Architecture Description

3.1 Business Understanding

This phase consists of a very precise specification of the problem together with methods of evaluating the achievement of the goal. A correct solution to the problem involves a correct understanding of the problem. Real business problems are rarely obvious.

3.2 Data Understanding

Data understanding is the knowledge that you have about the data, the needs that the data will satisfy, its content and location. Data understanding is expressed in organizations as business glossaries, data dictionaries, models and other forms of metadata or other places where information about the data is stored.

We have to understand data and gives some solution based on the questions. We can remove outlier values, remove missing data, and check relationship between different attributes/columns.

3.3 Data preparation

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

We have to find the outlier values with boxplot, and if there is any missing value in dataset then We have to replace. Then change the data type if particular data type not match with current Data type.

3.4 Modelling

A data mining model gets data from a mining structure and then analyzes that data by using a data mining algorithm. The mining structure and mining model are separate objects. A mining model stores information derived from statistical processing of the data, such as the patterns found as a result of analysis.

3.5 Evaluation

Model evaluation aims to estimate the generalization accuracy of a model on future (unseen/out-of-sample) data. Methods for evaluating a model's performance are divided into 2 categories: Train, and holdout (Test). Both methods use a test set (i.e. data not seen by the model) to evaluate model performance.

3.6 Deployment

The last step in the data mining process is to deploy the models to a production environment. Deployment is important because it makes the models available to users so that you can perform any of the following tasks like: Create reports that let users request predictions, view trends, or compare models.