

Lab 12

Real World Data Mining

Objective

Use learned data mining techniques to identify relevant patterns and useful information in a given data set.

Understanding the dataset

Go through the data set and try to gather information about its domain. Understand the meaning of all fields individually. Identify the dataset in terms of the attributes present in the data, the number of instances, missing values, and other relevant characteristics.

Objectives of Data Mining Experiments

Figure out 3-4 specific, interesting questions about the domain that you want to answer with the data mining techniques. Try to choose questions that are about the domain not about the techniques or the experimental parameters.

Preprocessing of the data

Identify the set of preprocessing techniques depending upon your choice of model and the given dataset. Explain the necessity of each such step.

Model Interpretation

Describe the resulting model (e.g., size of the model, readability, accuracy, etc.). Summarize the model in your own words focusing on the most interesting/relevant patterns. Elaborate on if and how the model answers the objectives identified.

Results

- What is/are the model(s) with the highest performance for each objective of each data set?
- Discuss how well a data mining technique worked on these datasets. What combination of parameters yield particularly good results?