

## Data Mining Assignment 1

1. What the data is.

The data is about shopping daily essentials for my house.

2. What type of benefit you might hope to get from data mining.

It helps me to understand how can i buy essentials needed for my house in my budget preventing me spending on unnecessary things. It helps me to analyse the amount of money i am spending on shopping each time and In what ways can I reduce it?

3. What type of data mining (classification, clustering, etc.) you think would be relevant.

Association rule mining is more relevant because Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets found in various kinds of databases such as relational databases, transactional databases, and other forms of repositories.

Association Rule Mining is sometimes referred to as “Market Basket Analysis”, as it was the first application area of association mining. The aim is to discover associations of items occurring together more often than you’d expect from randomly sampling all the possibilities.

4. Name one type of data mining that you think would not be relevant, and describe briefly why not.

Regression wouldn't be relevant because Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.

First, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Second, in some situation's regression analysis can be used to infer causal relationships between the independent and dependent variables.

We use regression analysis to describe the relationships between a set of independent variables and the dependent variable. Regression analysis produces

a regression equation where the coefficients represent the relationship between each independent variable and the dependent variable.