Data Mining Assignment 1

**Identify a problem from your own experience that you think would be amenable to data mining. For that problem describe:**

* Data mining is a process of discovering patterns in a large set of data and data warehouses.

Also,Data mining is basically an automatic or semi-automatic technical process that analyses large amounts of scattered information to make sense of it and turn it into knowledge.Be it predicting,or pattern search or anomaly detection etc.

The problem which i thought was the swiggy app which requires data mining. For this, data of the previous orders by customers need to be discovered and stored to guess the choices of customers accurately.

**1. What the data is.?**

* In computing, data is information that will be translated into a form that is efficient for processing any sort of problem.. This information can be in the form of audio clips, images and programs. Based on the requirement of the problem, the data can be modified into the required format.
* In this problem, the data is the previous purchases of various customers using the application, also keeping track of the delivery times and the issues faced by the customers if any.

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

* From data mining we can understand the structure of the data and that helps in identifying the type of data that needs to be used to provide the solution to the problem we have. Before collecting, the processing of data is a greater step and one which is difficult. We have to clean the data and make it ready to make predictions on the data.It really consumes lots of time.
* By retrieving the data for this application, the orders can be improvised as after collecting the data from the stored data base, as we can clean the data and then the similarity patterns can be identified.And, can know what sort of orders are being places with which age group and also, what are the issues that are being faced by customers with delivery agents and vice versa, be it time issues,or the way people respond to each other.
* So that we can recruit the persons who value punctuality,and can give good suggestions for the customers.

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

* Based on the problem we have to see which technique we want to use. If a value is needed to be predicted, classification but when we need to identify a pattern from the data we have to use a clustering method on the data.
* So, for the problem I considered, I would use clustering because I need to identify the pattern by using which I can predict a value. Instead of just saying Yes or No for every single order placed through app.
* For example, if a customer ‘xy’ orders breakfast everyday , then I can predict that he would order so and so food on the particular days..
* Hence, I think clustering is relevant for this problem.

**4. Name one type of data mining that you think would not be relevant, and describe briefly why not.  
For each, illustrate with an example, e.g., if you think clustering is relevant, describe what you think a likely cluster might contain and what the real-world meaning would be.**

* I think classification is not relevant for the problem I considered. Classification is used where I use the trained data set to classify the values I have.  
  In this problem, if I need to follow classification, based on the previous products a customer bought, I should be able to say whether he’ll order a product I have or not.
* So, I think classification is not relevant for this problem.