Data Mining Assignment 1

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

Problem:Covid-19

1. What the data is.

Dats is measured,collected,reported and analysed and can be visualized using graphs,images etc..,. Data is defined as the facts and statistics collected together for references.

Data for the problem is: We have to get the informtion about number of persons got affected,number of persons having symptoms,number of persons recovered,number of persons dead etc..,

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

We will collect lots of data and warehouse it.Analyse large amount of data to discover meaningful patterns.There are 2 methods of DM.They are:

1. Prediction methods--Used to predict unknown/future values using data we have.
2. Description methods--We go through facts rather predicting from the data.

From the above example,after collecting the data we have to predict about different people with same symptons based on their health issues.

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

Classification is the process where we can classify the data with the help of class labels. Example:Gender

In clustering,there are no predefines class labels and it is unsupervised learning. Example:Discovery of patterns

Regression:Used to identify and analyze the relationship between variables.

And there are other types of datamining.According to me,Regression is relevant type of datamining.

Regression is the type of datamining technique that is used to predict a range of numerical values in a dataset.So we can know the count of persons who may effect using death rates or based on the cases in that area etc..,

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.

Classification helps deal with different kinds of symptoms like cough,fever,etc.., can have different numerical data like death rate, recovery rate and different symptoms.So, these types can be classified for easy analysis. So, which is why classification can’t be done.

Write one to two pages of 11 point single-spaced typeset text - you aren't writing a paper, but it isn't short answer either.