Data Mining Assignment 1

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

Weather Prediction-

1. What the data is.

- The data includes the current state of weather like rain, heat, wind and fog.

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

- We can know if the weather is going to be sunny or rainy or cloudy etc
- It is useful for transport purposes(travelling by airways) and also in agriculture for farmers
- We can observe changes in the patterns of weather conditions which change over time by comparing with previous patterns and get to know about natural disasters ahead

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

- Decision trees with neural networks would be relevant combination of data mining techniques for weather prediction as they can easily recognize patterns within the large unstructured data and give accurate predictions
- For example the decision tree would give us the uncertain information and reduce its uncertainty and decide which attributes go into decision nodes of the tree. After forming a tree structure then, we can attain a result weather a day is either sunny or rainy based on the path of tree

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

- As mentioned above, a combination of data mining techniques will be useful for weather prediction but if only one of these is used then it wont show any result that we desire. For example, just performing clustering will not help in weather prediction.

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.

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.

Weather prediction is an everlasting area where data mining can be performed. As long as this earth exists, we need weather prediction. Climate can be defined as the effect the sun's radiation has in the long term on the varied earth atmosphere and surface when it is in rotation. As agriculture is a huge part of a nation's economy, weather prediction has become an important factor in day to day life.

The different data mining techniques used for weather prediction are-

- Group Method Data Handling
- Decision Tree
- Neural Networks
- K-means clustering analysis

As there is a huge amount of data from past records and also digging presently, just using one kind of technique will not give us a desired output. The unstructured data needs to be sorted out by removing redundant data, unnecessary data etc and then they have to be evaluated using different techniques. As the values always change and there are many other factors which impact climate change, it is always not possible to get 100% accurate results. After a lot of research, it was established that decision trees along with neural networks give 87% accuracy in prediction and 98% accuracy in precision.