Data Mining Assignment 1

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

1. **What the data is?**

- The data includes **weather** reports also contain information about precipitation, wind speed and direction, relative humidity, atmospheric pressure, and other things as well. A typical **weather report** tells you the high and low temperatures for the past day. It also tells you the present temperature.

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

- Know about the present climate like if the climate is going to be sunny or rainy or cloudy etc
- climate report useful for farmers like loosening the soil, seeding, special watering, moving plants when they grow bigger, and **harvesting**, among others.
- Can predict climatic changes and occurrences of natural disaster.

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

● Decision trees with clustering 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 a short answer either.

Weather prediction or knowing about the climate is an everlasting area where data mining can be performed. As long as this earth exists, we need weather prediction. Generally climate is defined as the weather conditions prevailing in an area in general or over a long period.**Agriculture** is one of the major sectors of the Indian economy. It is present in the country for thousands of years. Over the years it has developed and the use of new technologies and equipment replaced almost all the traditional methods of farming.So weather prediction became important.

From ancient times to till now we have a huge amount of data regarding the climate changes, 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.

The different data mining techniques used for weather prediction are-

- Decision Tree
- Neural Networks
- K-means clustering analysis