Summary

Food is a basic necessity in life, India being a country where farming is a primary occupation and most of India's economy is dependent on the agricultural produce. The agricultural output of the country is largely dependent on climatic conditions and India being such a large country has a wide array of geographical conditions and temperatures across the land mass. The process of determining which crops to grow in a specific area is very important and is mostly dependent on the soil, climate and abundance of water in the region. Our intention through this project is to predict the ideal weather that is beneficial to crop growth in a specific region.

The primary reason for us to choose this project is that we are really interested in weather forecasting patterns and we would like for it to be of some help to people. Numerical weather predicate is the use of current and past observations of meteorological data to predicate weather in the near future.

In this paper the data mining technique used is the K-means cluster algorithm on the data set which was modified into suitable from the raw format after pre-processing stage. After that J48 algorithm was applied on to it. J48 are the improved versions of the C4.5 algorithms or can be called as optimized implementation of the C4.5. Over that regression techniques were applied.

The J48 algorithm was applied to generate a decision tree and each node in the tree consists of a decision and that decision leads to the result. The K-means clustering algorithm is a data mining/ machine learning algorithm used to cluster observations into groups of related observations without any prior knowledge of those relationships.

The method used was to consider different geographical regions as clusters and the data used to predict the weather were:

* Annual-seasonal Minimum temperatures
* Annual-seasonal Maximum temperatures
* Annual-seasonal Mean temperatures.

Through this project we intend to learn more about data mining algorithms and their use in weather forecast. We are a very long way from controlling the weather but we are doing considerably better at predicting it. We cannot be a 100% certain at any point about the results we get from the predictive techniques but they will definitely aid in decision making which is the main aim of data mining.

Submitted by:

Abhay Arnikar

Siddharth Rawat