Project Proposal

Data Analysis: Trends In Weather

Marcus Lorenzana, Robert Vasquez

**Project Description:**

For our project we will be doing a data analysis of weather data over a period of perhaps 10 years although it depends on the file size cap. In terms of the actual data being processed, we are most likely going to use temperature and precipitation as our main measure. The goal of the project is to detect trends in certain regions of the United States and discover whether there is evidence for longer winters/summers and if the trend will get better or worse in the future. We will most likely use Weka as our data analysis tool to find these trends.

**Proposed Timeline and Milestones:**

We hope that by the progress proposal we will have reached close to the half point mark in our progress. By this time we should have gathered all our appropriate data and have done all the necessary data cleaning and preprocessing. During this time we hope to have some simple to medium complexity measures of the data and be in the data mining phase. By the final project proposal we expect to have completed most of the project and have interesting trends to support or dispel our hypothesis that changes in climate is taking place.

**Additional Information:**

For the actual data sets the most likely option will be National Climatic Data Center, specifically:

<https://www.ncdc.noaa.gov/cdo-web/datasets>

Our primary source right now for gathering ideas on the project is from this article:

<http://www.academia.edu/5173396/Effectiveness_of_Using_Data_Mining_for_Predicting_Climate_Change_in_Sri_Lanka>

It goes into a lot of detail into the process of using data mining for detecting patterns in the climate although we do not plan on implementing everything completely due to time. We also are not going to look specifically at Sri Lanka and will probably just gather the important steps in data mining climate in a more universal region. As such we will mainly be looking into “Temperature rise” and “High Intensity of Rainfall” which will probably coincide with “Sea Level Rise” and “Drought”.

Another great source we are taking a look at, which provides a better overview of the require steps is this series of presentation slides:

<http://web.ornl.gov/sci/knowledgediscovery/ClimateDataMining/docs/Ganguly_ICDM-SSTDM-slides_2008.pdf>

If finding patterns in climate change proves too difficult, we will most likely look for patterns in a database of temperatures and precipitation and form a conclusion off of the results we obtain.