Prediction of Rise in Sea Level by Memory-Based Reasoning Model using SAS® Enterprise Miner™ 12.1

Prasanna K S Sailaja Bhamidi, Oklahoma State university, OK, USA
Faculty Advisor: Dr. Goutam Chakraborty, Oklahoma State university, OK, USA

Abstract

Increase in the sea levels is a potential problem that is affecting the human race and marine ecosystem. Many models are being developed to find the factors that are responsible for it. In this research, Memory Based Reasoning model appears more effective than most other models. The data was collected from NASA. The data contains 1,072 observations and 10 variables, such as Emissions of Carbon-dioxide, Temperature and other contributing factors such as electric power consumption, total number of industries established, etc. Results of Memory Based Reasoning models such as RD Tree and Scan Tree, Neural Networks, Decision Tree, Logistic Regression are compared. Fit Statistics, such as Misclassification Rate and Average Squared Error are used to evaluate each model's performance. The results can be used to predict the rise in sea-levels in the near future and to take the necessary actions to protect the environment from global warming and natural disasters.

Data Preparation

The data was collected from National Oceanic and Atmospheric Administration website, Census.gov and NASA website. This data had 1072 observations and 10 variables. The demographic data included the Population of every State in USA, environmental data included the Maximum and the Minimum Temperatures, Carbon Dioxide emissions and Average Snowfall.

Model Building

The dataset is divided into 70% training and 30% validation data. Missing values were replaced using Replacement Node and imputation techniques like Count and Tree. Variables which were not normally distributed were transformed using the Max. Normal and Square root methods. Data preparation and cleaning were done by SAS® Enterprise Guide® 5.1. Predictive Models for the data were built and compared using SAS® Enterprise Miner™ 12.1.
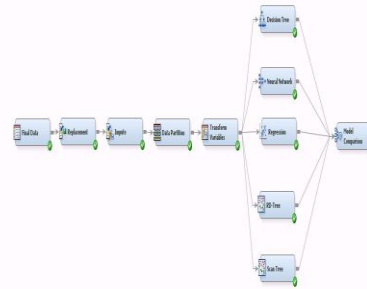

Figure 1. Model

Evaluation and Results

The models built were compared and evaluated using the fit statistics Misclassification Rate, Average Square Error, Cumulative Lift and ROC curve. Scan Tree – MBR model was the best one with a Misclassification Rate of 0.1461 and Average Squared Error of 0.101 and cumulative lift of 2.01 at 20th decile.

Name Of Model	Validation Misclassification Rate	Validation Average Squared Error
Memory Based Reasoning – Scan Tree	0.1461	0.101
Memory Based Reasoning – RD Tree	0.1495	0.109
Decision Tree	0.1485	0.112
Neural Network	0.1528	0.113
Logistic Regression	0.2691	0.181

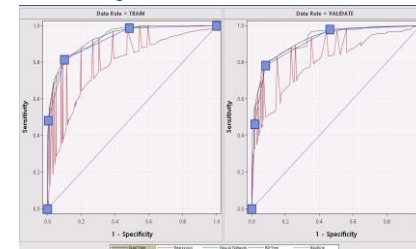Figure 2. Model Results

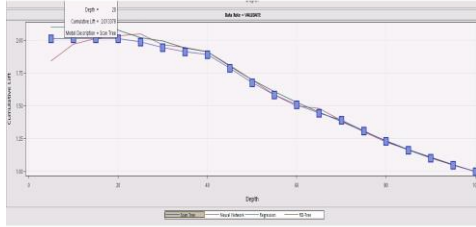Figure 3. ROC Curve Comparison

Figure 4. Cumulative Lift graph

Inferences based on the model

- According to the analysis, Carbon-Dioxide Emissions are the most contributing factors for the rise in sea level.

- Due to the higher power consumption, there is a significant increase in the temperature of a particular place.

- Due to the rapid increase in the temperature, the mean sea level of the places also increased significantly.

- The average maximum temperature of about 35% of the places increased drastically over the last decade.

Conclusions and Recommendations

- The Memory Based Reasoning model produced good results as this node carefully predicted the values taking the k-similar values of the previous years into consideration.

- The Memory Based Reasoning model works better than the traditional models like Neural Network, Decision Tree and Regression because Memory Based Reasoning validates the model.

- I have used all the long-term influencing factors, such as Carbon Dioxide Emissions and Power Consumption. Other factors, such as sedimentary deposits and tide gauge data, may also be taken into consideration for an even clearer picture of the solution for future research.

References

- National Oceanic and Atmospheric Administration. (2012) . *U.S. Daily Snowfall and snow depth amounts. Retrieved from:* http://www.ncdc.noaa.gov/snow-and-ice/dly-data.php
- National Oceanic and Atmospheric Administration. (2012) . *U.S. Temperature Data. Retrieved* from http://www.ncdc.noaa.gov/cag/timeseries/us

The author can be contacted at
E.Mail: sailaja.bhamidi@okstate.edu.
Work Phone: (405)612-5884