Data Project 7

Sree Teja Kongara,

1. Which of these Classification/ Prediction algorithms produces the best results?  What is it about the way the algorithm works that explains the quality of your results?

**Ans:** In my opinion, for the dataset in question, the Random Forest algorithm produces the best results. These results are relevant because the Random Forest algorithm produces the most accurate values among all current algorithms, and it can be seen in the output for our particular dataset as well. The classifier’s accuracy can be verified using methods such as the Confusion Matrix.

1. Which of these two algorithms works better than last week's kNN and Naive Bayes?  Compare how the algorithms work, then use that comparison to explain your results.

**Ans:** The Random Forest algorithm works better in comparison with the other algorithms. It also has an effective method for estimating the missing data, thus maintaining accuracy. The Random Forest algorithm has methods for balancing error in class population unbalanced data sets. This sets is apart from the Decision Tree algorithm, which makes sequential, hierarchical decisions about the outcome variable based on the predictor data.

1. What do the results from questions 2 and 3 tell you about the data in your dataset?  Explain your results so that a non-technical manager will see the benefit of your work for the future of the organization.

**Ans:** The results show that the dependent variable ‘Poverty’ is closely related to the chosen independent variable of ‘Income’, more so than the other variables Drive and Unemployment. This means that to understand the poverty levels in a particular county, the income levels of its residents must be studied, which would lead to deeper insight.

1. What business, political, or medical decision could this manager make on the basis of your results?

**Ans:** A political or government decision that can be made by utilizing the insights gathered from this analysis is that of analyzing income levels in the counties, so that a close correlation to its poverty levels can be reached. This could lead to a better understanding of poverty and might help in reducing poverty levels effectively.