Data Project 6

Sree Teja Kongara,

1. Which attribute(s) did you pick as the class attribute?  Explain your decision.

Ans: The attribute ‘Poverty’ was chosen as the class attribute since after drawing a correlation matrix, it was observed to have the highest correlation coefficient with other variables. The other variables having the next greatest values are Income, Unemployment and Drive respectively.

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

Ans: In my opinion, the kNN algorithm works best since we are comparing with k=1 and k=5 values. This shows us the closest data sets to our variable and so a relation can be established easily.

In the Naïve Bayes algorithm we are trying to establish the probability that the variable fits into a certain group. Thus the results obtained can be slightly ambigious when compared to the kNN result.

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

Ans: The kNN algorithm works better than the logistic regression model, since it easily establishes the nature of the class variable by testing the closeness to its neighbors. In my opinion this is an effective method to determine the class variable.

The Naïve Bayes algorithm tests for the probability that something fits into a certain group. It shows how likely the variable is to fit into a certain group. Whereas the kNN algorithm classifies an observation based on the closeness to it’s nearest neighbor(s) and establishes the relationship based on this. In our results, the kNN algorithm showed a more comprehensive relationship than the Naïve Bayes model.

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 attribute Poverty is closely related to the independent attributes Income and Unemployment. This can be seen in the plotted graph and the closeness of the data points shows the magnitude of effect the independent variables have on the class variable.

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

Ans: One of the possible decisions that can be made is a political one where the established relationship between Income and Unemployment in the counties can be observed to establish the effect they have on the Poverty levels in the respective county.