**1. Neural Net: Credit Risk**

In this analysis, the creditworthiness of bank customers is evaluated based on several different attributes. The ‘Neural Net’ operator is utilized to make predictions using a training and a test dataset. The original dataset had 1000 examples of which a 100 of those were removed to create a test dataset. Afterwards, the Neural Net predictive model was applied to the dataset and the results are as follows: of the test dataset examples 26% were predicted to have bad credibility while 74% were predicted to have good credibility. When compared to the training dataset, 69.6% had good credibility while 30.3% had bad credibility.

The table below shows the relationship between the credibility of a customers based on their balance of their current account. As can be seen, customers that have some balance amount in their current account have a better tendency of having good credibility than customers that have no balance or no running account at all.

|  |  |  |  |
| --- | --- | --- | --- |
| Balance of Current Account | Good | Bad | Ratio |
| <=200 | 5 | 1 | 5:1 |
| >=200 | 34 | 7 | 4.8:1 |
| No Balance | 18 | 12 | 1.5:1 |
| No Running Account | 16 | 7 | 2.3:1 |

In addition, a comparison of the confidence levels of the predictions is analyzed. There are 17 examples that were predicted to have good credibility with a confidence level of one. Out of those 17, 11 of them are customers that have a current balance of >=200DM. This indicates that any customer that has a current balance of >=200DM has a higher chance of having good credibility than other customers. On the other hand, there are 9 examples that were predicted to have bad credibility with a confidence level close to 1. Out of these 9, 5 of them are customers that have no current balance. This also indicates that customers that have no current balance have a higher chance of having bad credibility as compared to other customers. Overall, the Neural Net prediction model provides a solid basis of predicting credibility of customers with good analytical results.

**2. Model Evaluation:**

In order to determine the best learner to use for the Mesothelioma data, the “Compare ROCs” operator was used on four different learners: Decision Tree, Random Forest, Naïve Bayes and K-NN. Of these four learners, the visual comparison revealed that a decision tree would be the best one to use for our dataset. After determining this, the model below was used in to predict the class of diagnosis.

Diagram

Description automatically generated

A picture containing application

Description automatically generated

The resulting decision tree is presented below:

![Diagram

Description automatically generated]()

It can be concluded from the above result that ‘Diagnosis Method’ is the best predictor of the outcome of the treatment with an accuracy of a 100%. I reviewed the entire dataset after getting the results of the decision tree and I was able to observe that each example that had a diagnosis method value of 1 had a class of diagnosis value of 1 (healthy) and a diagnosis method value of 0 had a class of diagnosis value of 2 (has Mesothelioma). Due to this characteristic of the data, the decision tree easily classifies the class of diagnosis as 1 or 2 only by looking at the diagnosis method attribute. Furthermore, a performance of the decision tree model as well as two others (Naïve Bayes and SVM) was tested and presented in the tables below:

Decision Tree

Table

Description automatically generated

Naïve bayes

Table

Description automatically generated

SVM

Table

Description automatically generated