**ISGB7967 – Data Mining for business Dr. Kamesam**

Assignment 3 – **Artificial Neural Networks**

**Question 1**

Both Decision Tree model and Neural Networks model set 50% data for training and the rest for testing.

First, the **Decision Tree** model can predict the churn status of an unknown customer with overall **85%** accuracy.The **Neural Networks** model can only predict the churn status of an unknown customer with overall **76%** accuracy.

Second, since the company wants to seize voluntary leavers, whether the model classified Vol well or not is very important.The decision tree model has overall **95%** accuracy in predicting voluntary leavers while The Neural Network model has overall **72.8%** accuracy in predicting voluntary leavers.Unfortunately, Neural Network model is not an ideal model to tell who is voluntary leaver.

The partition of Neural Network model decides there’s **fewer training data**(30% used for validation).In addition, the sample space is not large enough for Neural Network model. So Neural Network model not perform as well as Decision Tree model.

Therefore,I recommend the company to use the Decision Tree model, since the decision tree model can predict an unknown customer much more accurately than Neural Network model and is more effective in predicting voluntary leavers.

|  |  |
| --- | --- |
| 571979792476882535  **Figure1.1 Neural Network Analysis Results** | 509034465139836447  **Figure1.2 Decision Tree Analysis Results** |
| 35833009241664433  **Figure1.3 Neural Net Model Stream** | |

**Question 2**

**Problem Statement**

Given data on some features of individuals and individual’s respond to charity status, is it possible to build a model which can predict the response of an unknown individual with reasonable accuracy? The given data has individuals who respond to charity and individuals who don’t respond to charity.

**Analysis**

Using SPSS, a “neural network” model was constructed. 70% of the data was used for building the model and 30% for testing. For an unknown individual, We expected the model to tell us whether he or she would respond to the campaign or not.

The predicted accuracy of the neural network is **75.1%** .The **most important fields** within the network are **Pre-campaign** expenditure and **Pre-campaign** visits.

The model can predict the response status of an unknown individual with overall **74%** accuracy.That’s not so good but okay.However, the model shows high accuracy in predicting individuals who are **Non-responder**(with 95% accuracy but only 30% accuracy in predicting responders). The reason why is that **a large proportion of data** is Non-responder(68.7%) while only 31.3% is responder. The performance of Neural Network model is highly depend on the volume of data.

**How make use of model**

The model can predict whether an unknown individual will response to campaign(with overall 73.7% accuracy). The model is very effective in predicting individuals who tend not to response campaign.

|  |  |
| --- | --- |
| 356893043408473028  **Figure2.1Model Summary** | 595380359706188190  **Figure2.2 Classification Accuracy** |
| 294623576054241274  **Figure2.3 analysis results** | 628989009264339728  **Figure2.4 Proportion of response** |
| 785882451620058836  **Figure 2.5 Model Stream** | |

**Question 3**

**MLR Model**

The MLR model with all inputs shows that **61.25%** of Concrete Compressive Strength can be explained by age and ingredients. The F Signification indicates that we should reject null hypothesis that there is NO linear relation between CCS, age and ingredients. From P-value of the t-stat, **FAG and CAG** should be eliminated from the regression. After eliminating FAG then eliminating CAG, the MLR model suggests **61.17%** of Concrete Compressive Strength can be explained by age and ingredients.

**Regression Equation:**

CCS = 29.03 + 0.1054\*CEM + 0.0865\*BFS + 0.0687\*FA - 0.218\*H2O + 0.239\*SPR + 0.1135\*AGE

**Conclusion**

The mean absolute error of MLR model is 8.26 and mean absolute percentage error is 31.47%.The model is **not good**.

**ANN Regression Model**

Set 70% of data for building the model and the 30% for testing. The ANN model has overall accuracy of **87.8%** and **Mean Absolute Error** for ANN model is **4.739**, which is much lower than that of MLR model(**8.26**). For the comparison of predicted ccs and actual ccs, ANN model performs better than regression model.In ANN, Linear Correlation(**0.932**) and bubble plot(Figure3.6) show that the model is very effective.But in regression model, the correlation between the two ccs is **0.783**(Figure3.4), much less than that in ANN.For the standard deviation, ANN(**6.194**) is also better than regression model(**10.379**).

The ANN model is stronger than MLR model, because the ANN model is a non-linear regression model.The regression model is a linear model, which limited in predicting some values that are non linear.

|  |  |
| --- | --- |
| excel1  Figure3.1 MLR regression with all inputs | excel2  Figure3.2 MLR regression without FAG&CAG |
| excel3  Figure3.3 MLR Inter correlation matrix | mean  Figure3.4 MLR 3 measures |

|  |  |
| --- | --- |
| q31  Figure 3.5 ANN Model Summary | q32  Figure 3.6 ANN actual and predict ccs |
| q34  Figure 3.7 ANN analysis results | |
| q33  Figure 3.8 Model Stream | |