Applied Analytic Modeling

Lab 3
Predictive Modeling Using Neural Networks-SAS Miner

Submitted to
Prof. David Parent

Submitted by

Grace Adaji
301373339

Predictive Modeling Using Neural Networks

a. In preparation for a neural network model, is imputation of missing values needed? Why or why not?

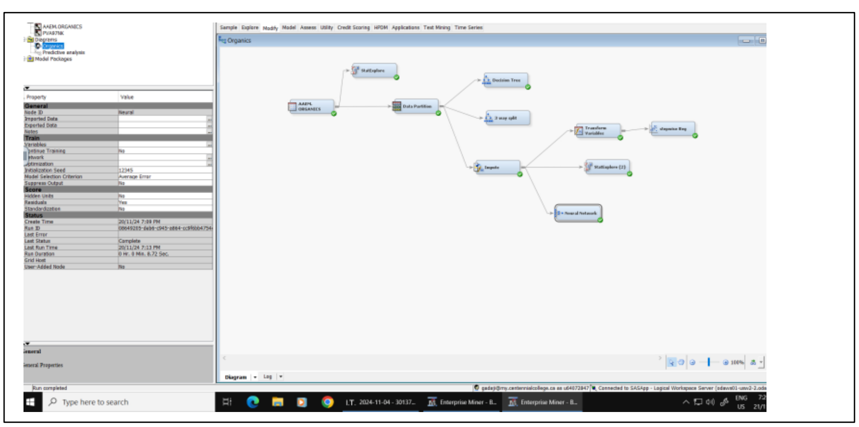
Yes, Imputation of missing values is often necessary when preparing data for a neural network model. Neural networks require a complete dataset to function effectively, as missing values can disrupt the training process and hinder the model's performance. Imputation fills in missing values with reasonable estimates (e.g., mean, median, or mode), preserving the dataset's size and avoiding the loss of valuable information that might occur if rows or columns with missing data are removed. Without imputation, the model may fail to run or produce inaccurate predictions, reducing overall performance and reliability.

However, imputation is unnecessary if the dataset has no missing values or if imputation risks introducing bias that could negatively impact the model's predictions. Additionally, if missing data is random and its absence does not significantly affect the model's performance, imputation may not be required. The decision to impute should be based on the nature and extent of the missing data and its potential impact on the modeling process.

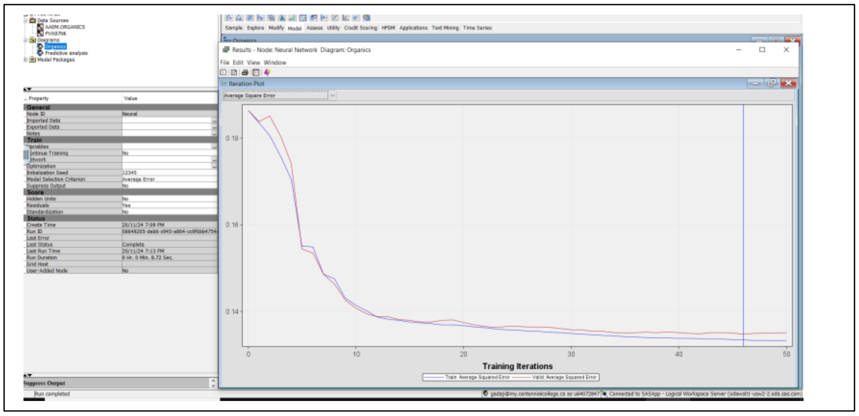
b. In preparation for a neural network model, is data transformation generally needed? Why or why not?

Click or tap here to enter text. Yes, data transformation is generally needed for neural network models. Data with extreme values or skewed distributions can make it harder for the model to converge, leading to slower training times and less accurate results. Transformations such as log transformations or other scaling techniques can help normalize and standardize the data, handle outliers and address skewed distribution for better model performance
However, neural networks are relatively tolerant to skewed data due to nonlinear nature of the activation function. In cases where the data is already well- distributed or if the model includes techniques that can handle untransformed data, transformation may not be strictly necessary. The need for transformation depends on the specific data

c. Add a Neural Network tool to the Organics diagram. Connect the Impute node to the Neural Network node. Screenshot the diagram.



d. Run the Neural Network node and examine the validation average squared error. How does it compare to other models? Screenshot your ASE window.



The Decision Tree (3 Split) model demonstrates the best performance, with the lowest validation Average Squared Error (ASE) of 0.132662, making it the most accurate predictive model in this comparison.
The Decision Tree (2 Split) model follows closely with an ASE of 0.132773, slightly higher than the 3-split variant but still outperforming the other models.
The Neural Network model has an ASE of 0.134752, which is higher than both Decision Tree models. While its performance is not the best, it remains competitive and demonstrates stronger predictive capabilities compared to the Stepwise Regression model which has the highest ASE of 0.137156, indicating the least accurate predictions.
In terms of model ranking:
Decision Tree (3 Split): ASE = 0.1326((Best)
Decision Tree (2 Split): ASE = 0.132773 (Second Best)
Neural Network: ASE = 0.134752 (Third Best)
Stepwise Regression: ASE = 0.137156 (Least Favorable)
While the Decision Tree (3 Split) is the optimal choice for accuracy, the Neural Network model’s slightly higher ASE suggests it could still be a viable option, particularly if interpretability or flexibility in capturing non-linear relationships is a priority. The Stepwise Regression model, with the highest ASE, is the least suitable for this dataset