0.Meta data

Course Name: Programming and Problem Solving for Data Analytics

Course Code: 04802-E

Instructor: George Okeyo.

TAs: Faustin, Janvier, Pacifique

Assignment Title: **Building a Credit Approval Model using Neural Networks**.

Name: Manzi Patrick(**pmanzi**)

Submission Date: 9/12/2021

1.Background and problem Description
Nowadays, financial institutions such as banks are receiving massive credit applications. This tremendous rise in number of credit applicants has challenged the manual ways of processing these credit requests from customers where bank staff must manually look through applicant's history like income levels and day to day account transactions among others in order to approve or reject the application. However, this process is time consuming. As result, my research will focus on **automating credit approval tasks** using machine learning technique called artificial neural networks.

2.Methods

In this research, python programming language was used as it is believed to have powerful tools for automating tasks and conduct data analysis that goes beyond simple data analytics. A variety of techniques for data preprocessing was applied for tracing and handling missing data. Different functions like info and value count were used to detect missing values and then the missing values was replaced by mode for categorical attributes and mean for continuous attributes.

To make sure that dataset is clean, i applied several techniques to investigate if no outliers were in our data. I employed varies boxplots to perform un-variate and scatter plots multi-variate outlier analysis. Finally, after finding outliers, I have handled them using standard score (z-score) which basically gives an idea of how far the data point is from the mean and then by setting a threshold, I could detect and remove those data points that were far from the mean. Outlier analysis was only

conducted on continuous attributes. After, I employed one hot encoding and label encoder of treating categorical features and then used standard scaler for scaling for continuous variables to avoid model bias.

Finally, neural network model called MLPclassifier from skit-learn library was built, trained and evaluated using cross validation technique. Metrics like accuracy and F1-score was employed to verify model predictability power and greedy search hyperparameter tuning was used to give the best estimator.

3.Results and discussion

This research had identified that 67 missing values and 52 outliers. Missing categorical values was replaced by mode whereas continuous got filled with mean and then outliers were completely dropped. Appropriate label encoding was undertaken, and standard scaling was performed to make data ready for modeling.

Accuracy score performance metric which is the fraction of predictions our model got right had shown that our model is accurate at 84.94%. whereas the F1 score which is the weighted average of Precision and Recall was found to be 84.16%. this results generally prove that our model to perform well.

Hyperparameter optimization had revealed set of optimal parameters for our best model including identity activation function, 11 hidden layer sizes,90 number of epochs or maximum iterations and lbfgs solver or optimizer.

Surprisingly, the performance of our classifier had not significantly boosted after hyperparameter tuning. for example, the accuracy was increased from 84.94% to 85.10% whereas the f1-score h ad a little increase from 84.16% to 84.57%. by comparing these performance results with bench marks this credit approval model works better.

4.Conclusion

In conclusion, this research has made the automation of credit approval tasks possible using machine learning artificial neural networks which can be paramount in financial institutions such as banks. Even in a situation where banks receive tons of credit applications, this model can still handle the tasks with approximately 85% accuracy. I believe that in a near future, several financial institutions will embrace this technology and it will be a game changer.

5.Reference

[1]"Predicting Credit Card Approvals using ML Techniques", *Medium*, 2021. Available: https://medium.datadriveninvestor.com/predicting-credit-card-approvals-using-ml-techniques-9cd8eaeb5b8c.

[2]"Ways to Detect and Remove the Outliers", *Medium*, 2021. Available: https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba.

[3]"1.17. Neural network models (supervised)", *scikit-learn*, 2021. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html.

[4]V. Zach, "K-Fold Cross Validation in Python (Step-by-Step)", *Statology*, 2021. Available: https://www.statology.org/k-fold-cross-validation-in-python/.

[5]"Neural Net with GridSearch", *Kaggle.com*, 2021. Available: https://www.kaggle.com/hhllcks/neural-net-with-gridsearch.

[6]"UCI Machine Learning Repository: Credit Approval Data Set", *Archive.ics.uci.edu*, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Credit+Approval.