Write Up - Credit Card Default Machine Learning Analysis

For my project, I utilized the Default of Credit Card Clients dataset as found on UC Irvine’s Machine Learning repository (<https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients>). The data was collected in Taiwan, and the objective of this dataset is to predict whether a customer will or will not default on their credit card payment. A value of def\_payment = “1” denotes a defaulter, and a value of “0” denotes a non-defaulter. Therefore, the dataset indicates a binary classification problem.

Before creating any models, exploration, engineering, and scaling were performed on the dataset. Then, I split the data into training, test, and validation dataframes using a 70%, 20%, and 10% split, respectively. From there, I used a MinMaxScaler to rescale each feature to a common range, also ensuring the data was in a state to make the best predictions possible.

To solve the underlying binary classification problem, two machine learning models were created. The first was a Logistic Regression model. Here, I was somewhat surprised at my results. The ROC AUC score was average, coming in at about 0.75 for train, 0.76 for test, and 0.74 for validation. However, there was a bit of a drop off for the F1 and accuracy scores, with values approximately equal to 0.69 for both metrics across the train, test, and validation datasets. Later in the dataset, I attempted to re-run the Logistic Regression model with only the top 3 features present after determining the most important features. The model performed worse. In a Production environment, this “refined” Logistic Regression candidate model would not have taken the role of champion due to its lesser performance against the same data.

Finally, I created a Random Forest model. Initially, without tuning, my Random Forest model provided 0.71 for both accuracy and F1 score against train and test. For validation, both values were 0.69. Upon tuning my model, the results were worse for both test and validation datasets. I did not include my training results as credible, as 0.98 for both metrics indicates strongly to me that there was most likely overfitting on this dataset since it was used to derive the “best model” from k-fold cross validation.

In conclusion, while my model has decent performance, I do not think it would suffice to present to executive level stakeholders at a credit card company. My ROC AUC score of 0.74 in my Logistic Regression model within the validation dataset indicates a decent TPR against FPR as the threshold varies. Additionally, my accuracy in the validation dataset of 0.69 indicates that approximately 69% of the predictions are correct. Finally, an F1 score of 0.69 across the Logistic Regression and Random Forest models for my validation dataset displays a reasonable tradeoff between precision (ability to correctly predict positive instances) and recall (ability to capture all positive instances). However, given the context and industry, I would want all these values to be higher to make the determination of whether a customer should be issued a credit card based on their risk of default. Were I a data scientist at this credit card company, I would recommend capturing additional features of interest such as credit score and debt-to-income ratio. From there, I would refine and rebuild my models to yield better results.