Adrian Stahl
Deep Learning Challenge
Module 20

Credit Risk Analysis Report

- Provide an overview that explains the purpose of this analysis.

  - The objective of the analysis is to develop a model capable of classifying and assessing the creditworthiness of borrowers.
  - The dataset consists of historical lending data obtained from a peer-to-peer lending services company, and the goal is to identify individuals who are likely to default on a loan.
  - The variable being predicted is referred to as "y_prediction." Initially, I extracted the loan_status values and analyzed them using the value_counts method to determine the frequency of good and bad loans within the dataset. This information is utilized to construct the test data for training and prediction.
  - Two models are constructed and applied to the task: one utilizes the LogisticRegression method, while the other incorporates the RandomOverSampler technique.

- Using a bulleted list, describe the accuracy, precision, and recall scores of the machine learning model.

  - Machine Learning Model 1 LogisticRegression:
    - balanced_accuracy_score = 0.9520479254722232
    - precision 0=1.00, 1=0.85
    - recall 0=0.99, 1=0.91
  - Machine Learning Model 2 RandomOverSampler:
    - balanced_accuracy_score = 0.9936781215845847
    - precision 0=1.00, 1=0.84
    - recall 0=0.99, 1=0.99

Adrian Stahl
Deep Learning Challenge
Module 20

- Summarize the results from the machine learning model. Include your justification for recommending the model for use by the company. If you don't recommend the model, justify your reasoning.

  - The RandomOverSampler exhibits superior performance by effectively identifying a greater number of individuals who are at risk of loan default based on their credit risk assessment.

  - The model's performance is contingent upon the specific problem being addressed. In this scenario, there is no significant concern about overclassifying good loans as it is advantageous. However, the primary focus lies in improving the classification of individuals who fall under the 1's category.

  - Given these considerations, I strongly advise utilizing the RandomOverSampler model exclusively for this particular problem.