

Austin Craft

10/14/25

CSCI460

Naive Bayes Analysis

The purpose of this project was to evaluate how the performance of a Naive Bayes classifier is affected by the changes in train/test data ratios. The python script loads the given data set into a data frame, then processes all the categories into numerical values so the model can process them. Null or missing values are disregarded, and the dataset is split between features and the target, whether the client subscribed to a deposit or not. The program then uses the sklearn library to train the dataset on different testing splits, and the results of each split is printing individually. These results show the accuracy and f1 score for that particular split. Overall the results for this model show that the accuracy of the model remained consistent across several test/train Splits. The model was able to generate it's best f1 score with only 20 percent training data, suggesting that this model was able to generalize the data fairly well. Unfortunately the model was not able to rapidly improve when given more and more training data, showing that the model may have oversaturated it's learning fairly quickly. The overall simplicity and naive assumptions limit the model's ability to predict accurately based on more training data. So while the model was consistent it is still limited in its accuracy due to the limitations of the naive bayes model itself. So while the model is able to make fairly accurate predictions, increasing the training data does not have as large an impact as one would expect due to the limitations and assumptions that are required when working with a naive bayes model.