Predicting Firm Bankruptcy

Team: Saurav S Borah, Durga M Dash, Jacob K Shaw

In this project, we were asked to develop a predictive model that combines various econometric measures to predict if a firm will file for bankruptcy. The training data provided consisted of 64 predictors and 10000 datapoints.

We trained the data using various classification models such as Logistic Regression, Adaptive LASSO, SVM, Decision Tree, Random Forest, Bagging, Gradient Boosting, Neural Network, HP Neural, etc. After comparing the ROC index values (from the Model Comparison node) on the validation set, we have selected the two best fitting models mentioned below:

Project Model 1:

Our first project model focused on optimizing classification predictions using two key prediction models: the HP Neural and Gradient Boosting. We leveraged the Ensemble node, employing a voting ensemble method to combine their outputs. This approach synergized diverse learning strategies, enhancing predictive accuracy and robustness.

The Ensemble node facilitated collective decision-making, aggregating predictions from both models. By mitigating individual biases and harnessing multiple model strengths, our final model achieved superior performance. This amalgamation of techniques not only optimized predictive power but also provided a reliable framework for informed decision-making based on classification predictions.

In summary, our project's success stemmed from the amalgamation of the HP Neural and Gradient Boosting Models through Ensemble modeling, resulting in a high-performing classification prediction model in SAS Enterprise Miner.

Project Model 2:

For the second project model ,we leveraged gradient boosting and neural network techniques. The Gradient Boosting Model and Neural Network nodes were utilized to develop individual predictive models due to their abilities to handle complex data relationships and identify intricate patterns, respectively.

To improve accuracy and reliability, an Ensemble node was employed using the average ensemble method. This approach combined predictions from both models, mitigating biases and enhancing overall performance by leveraging the strengths of multiple algorithms. Rigorous validation through cross-validation techniques ensured the model's efficacy.

The resulting ensemble model represents a culmination of diverse methodologies, offering a balanced and dependable classification prediction mechanism. Its deployment potential spans industries where precise classification is crucial, underscoring its value in finance, healthcare, marketing, and beyond.