Name: *Daniel Velek, Dario Cavada, Joshua Bringhurst*
Course: *BAN615DE 23FA Predictive Business Analytics*
Professor: Dr. Shaikh
Assignment: *Final Project*
Due Date: *Saturday, December 16th*

Table of Contents

1) Abstract

The text Information Systems Education Journal (ISEDJ) Volume 19, No. 1, February 2021, features a special issue on Fraud Cases. This abstract is focused on the case titled "Can you Predict the Money Laundering Cases?" authored by Richard V. McCarthy, Wendy Ceccucci, Mary McCarthy, and Nirmalkumar Sugurmar. My understanding is this case, is now intended for business analytics courses, and revolves around predictive analytics applied to anti-money laundering efforts. There students are tasked with developing and optimizing predictive models using data from People's United Bank. The case explores the challenges of improving upon the bank's baseline model and emphasizes the importance of accurate detection in addressing the critical issue of money laundering. The significant importance of financial crimes, more specifically money laundering, poses an enlarged threat and challenge for banks in complying with regulatory requirements. In order to address this, the case "Can you Predict the Money Laundering Cases?" expands into the realm of business analytics, focusing on predictive analytics as a tool to enhance anti-money laundering efforts. This introduction provides an overview of the case's relevance, the background of money laundering, and the vital role of predictive models in identifying suspicious transactions. This case sets the stage for students to navigate through the steps of hypothesis development, data analysis, and model development, ultimately aiming to surpass the baseline model's effectiveness in detecting potential money laundering cases. Through this case, students will gain practical insights into the application of predictive analytics in a real-world scenario within the banking industry.

2)Introduction

In this Money Laundering case, the prediction of money laundering cases involved a critically thought-out approach to business analytics, specifically leveraging their predictive analytic techniques. The authors utilized a dataset that was provided by People's United Bank,

consisting of 38,515 total transactions sampled from approximately October 2014 to September 2015. The baseline model, created by the bank through six iterations of predictive models, served as the starting point in order to begin analysis. We found the best fit model by using multiple processes including these statistical measures such as the Kolmogorov-Smirnov (KS) and Receiver Operating Characteristic (ROC) statistics, with a focus on optimizing predictive accuracy.

The various predictive modeling techniques, including regression, were utilized to enhance the model's fit. Predictive models were developed and validated using a 70/30 split of the data, where 70% was used for model building, and 30% for validation. The criteria for selecting the best fit model included an emphasis on minimizing misclassification rates and evaluating Type I and Type II errors. In the final model, it aimed to pass the benchmark set by People's United Bank, providing a more defined identifier of transactions that required investigation for potential money laundering. Through this process, the authors sought to demonstrate the practical application of predictive analytics in improving the efficiency and accuracy of anti-money laundering efforts within the banking sector.

The importance of our project is underscored by its direct alignment with our SAS Enterprise Mining diagram (see appendix 1.1). Recognizing that our level of expertise may not reach that of the professionals showcased in the case study, we embarked on our predictive business statistical analysis. Our method encompassed the utilization of diverse metrics, including KS Statistics, ASE, Misclassification Rates, ROC, Decision Trees, and Neural Network nodes. These metrics played a pivotal role in the discovery of key insights, ranging from identifying missing values to evaluating skewed variables and even pinpointing potential outliers.

Through our meticulous analysis, we unearthed that the HPDMForest had emerged as the optimal Neural Network Node, a discovery validated by both the KS Statistic of 0.519 and the ROC index of 0.85. The selection of this node underscored its superior performance within the context of our predictive modeling endeavors. The application of KS Statistics and ASE of 0.123195 facilitated an in-depth assessment of the discriminatory power inherent in our models, while Misclassification Rate of 0.181965 had provided indispensable insights into the precision of our predictions. The ROC index, offering a comprehensive measure of our model's class-distinguishing capability, significantly contributed to strengthening the overall soundness of our analytical framework.

There was a total of 6 variables with a right skewness, starting largest to smallest:

- Wires_Size – 17.0522
- Num_Related – 15.3631
- Num_Trigger – 15.2936
- Wires_Mult – 15.0994
- Num_Acct_Alert – 8.9476
- Num_Tran_Alert – 6.6717

Furthermore, our exploration of Decision Trees (see appendix 2.1 & 2.3) significantly enriched our comprehension of variable importance and relationships within the dataset. This combination of many and various techniques in a predictive analysis highlights our steadfast dedication to withdraw meaningful insights despite the potential limitations in our practical sophistication. This all-inclusive approach instilled confidence in our required transformation and in addressing the potential outliers, thereby reinforcing the reliability of our analytical outcomes.

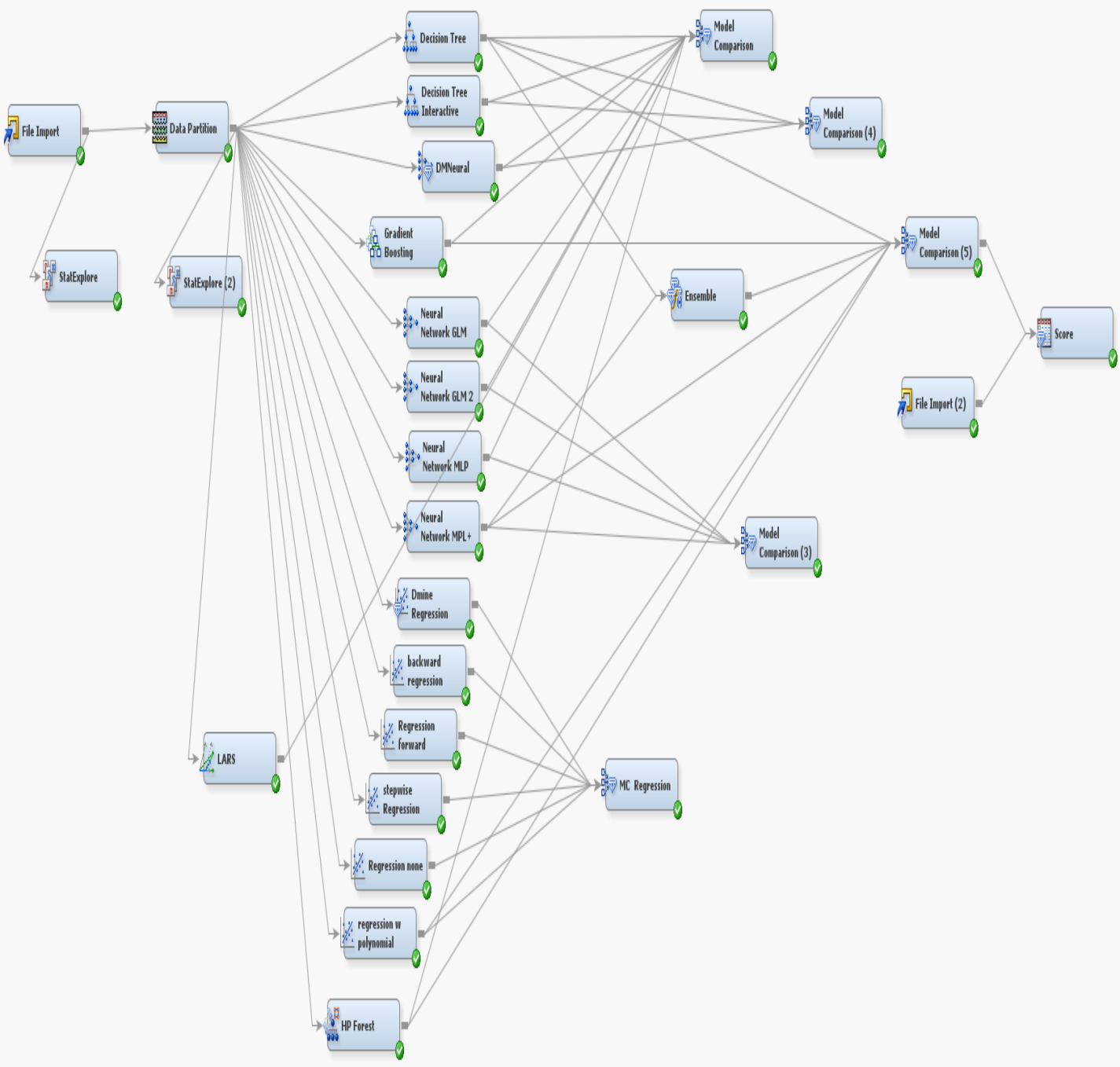
Lastly, our journey through this predictive business statistical analysis has been both enlightening and fruitful. The alignment of our project with the SAS Enterprise Mining diagram provided a solid foundation for our exploration. Acknowledging that our proficiency does not rival that of the professionals who are featured in the case study, we navigated through diverse metrics, including KS Statistics, ASE, Misclassification Rates, ROC, Decision Trees, and Neural Network nodes. What was deemed surprising is the fact that the HPDMForest node emerged as the optimal node, supported by both the KS Statistic and the ROC index, signifies a critical milestone in our predictive modeling efforts. This finding underscores the sturdiness and effectiveness of our chosen approach. The application of KS Statistics and ASE allowed us to rigorously assess the power of our models, while insights from Misclassification Rates provided a tangible measure of accuracy. The ROC index, with its comprehensive evaluation of its capabilities, further solidified the reliability of our analytical work.

3) Conclusion

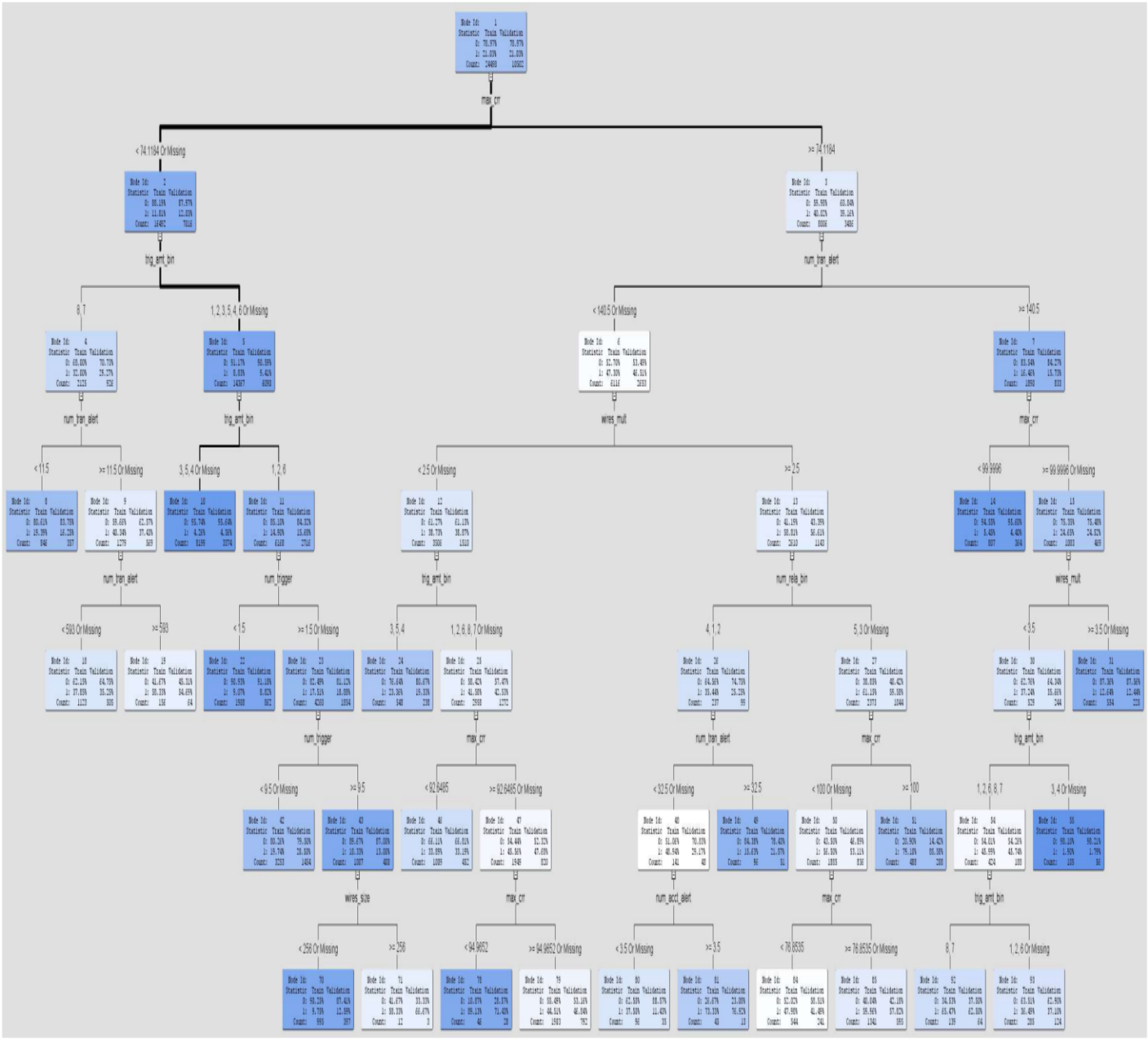
In conclusion, our journey through predictive analytics has significantly enhanced our understanding of all variable importance and its relationships within the dataset, adding depth to the overall analysis. With these techniques, our group has successfully conducted a thorough predictive analysis, showcasing our commitment to extracting meaningful insights, even when faced with limitations in hi-tech sophistication. This extensive approach instilled confidence in the integrity of us and our data, affirming that no transformational adjustments were required, and all outliers were effectively addressed.

4) Appendix

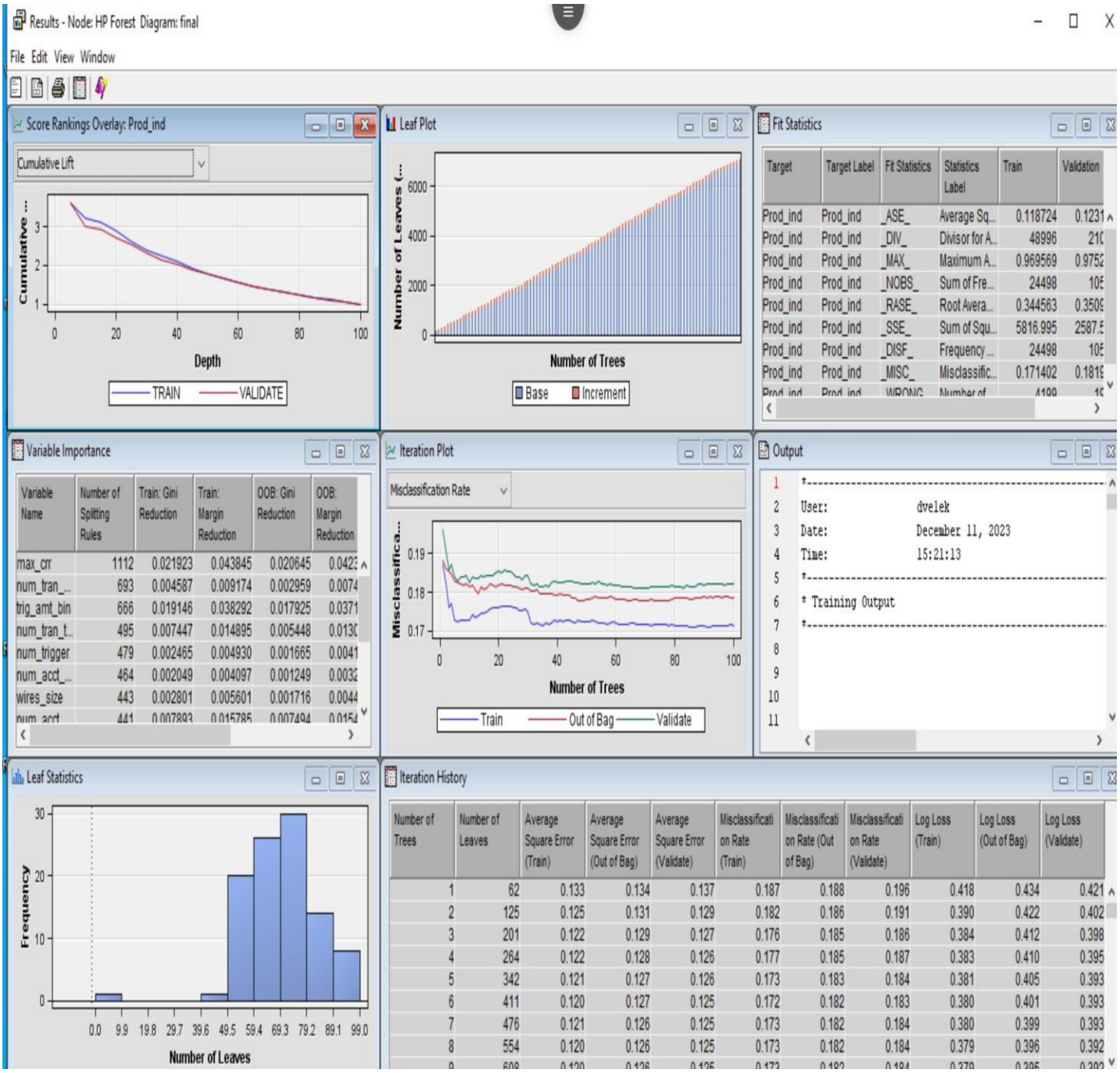
4.1 SAS Diagram



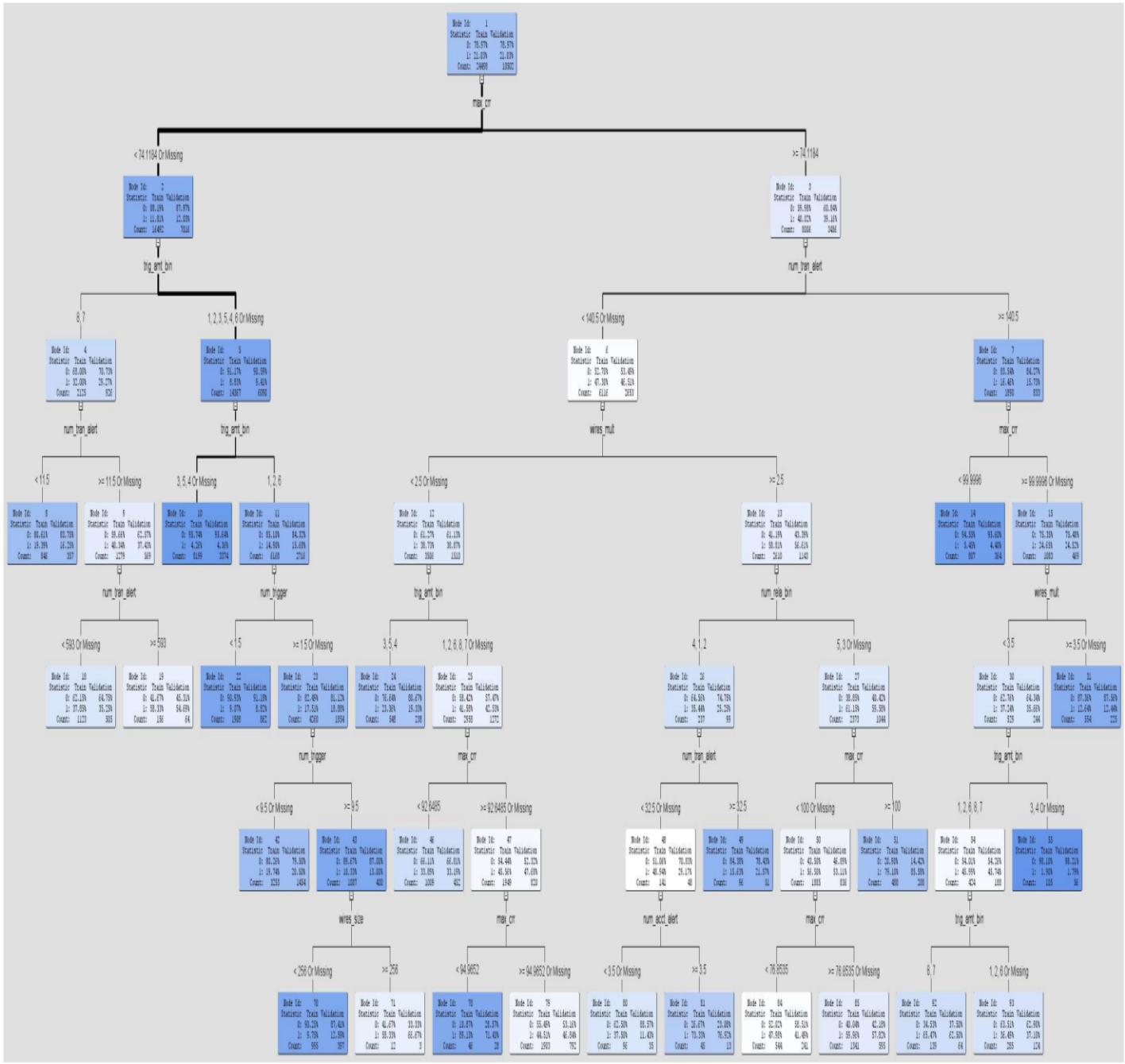
4.2 Decision Tree



4.3 Decision Tree HP Forest



4.4 Decision Tree Interactive


max_cr
< 74.1184 Or Missing
>= 74.1184
trig_amt_bin
num_tran_alert
8,7
1,2,3,5,4,6 Or Missing
< 140.5 Or Missing
>= 140.5
num_tran_alert
trig_amt_bin
wires_mul
max_cr
< 11.5
>= 11.5 Or Missing
3,5,4 Or Missing
1,2,6
< 2.5 Or Missing
>= 2.5
< 99.9996
>= 99.9996 Or Missing
num_tran_alert
num_trigger
trig_amt_bin
num_rela_bin
wires_mul
< 593 Or Missing
>= 593
< 1.5
>= 1.5 Or Missing
3,5,4
1,2,6,8,7 Or Missing
4,1,2
5,3 Or Missing
< 3.5
>= 3.5 Or Missing
num_trigger
max_cr
num_tran_alert
max_cr
trig_amt_bin
< 9.5 Or Missing
>= 9.5
< 92.6485
>= 92.6485 Or Missing
< 32.5 Or Missing
>= 32.5
< 100 Or Missing
>= 100
1,2,6,8,7
3,4 Or Missing
wires_size
max_cr
num_acct_alert
max_cr
trig_amt_bin
< 256 Or Missing
>= 256
< 94.9652
>= 94.9652 Or Missing
< 3.5 Or Missing
>= 3.5
< 78.8535
>= 78.8535 Or Missing
8,7
1,2,6 Or Missing