

## Introduction

The gaming industry has expanded rapidly and is one of the fastest growing industries in USA. Over the last 12 years consumers' attitudes towards and acceptance of the gaming industry has changed resulting into widely diverse demographics of players and large sales volumes. As a result of its vastness and demographic expansion, the cost of development in the gaming industry has also considerably increased. The average age of people playing games has also considerably increased. Although many people believe that there are less number of female gamers, the actual proportion is now close to 40%. Overall, the game industry has done a remarkable job over the last few years in coming up with new games targeting different groups of demographics. The gaming industry has also partnered with Hollywood to make video games for movies such as Harry Potter, Spider Man, Alice in Wonderland and so on. Video game which was originally meant only for entertainment and hobby is now a profession by itself. However, little research exists with regards to which variables and models may predict sales success of a game. In this research, I compare the performance of a logistic regression, decision tree and a neural net in predicting the success for video games using published data from commonly available sources.

## Tool

JMP is used for data exploration and building predictive models using Neural Network, Partition and Logistic Regression. 600 data points are collected from one of the popular website: [www.metacritic.com](http://www.metacritic.com)

## Overview

“Sales” is the target variable which is modeled to forecast the block buster games. The variable “Sales” (measured in ‘000\$) are converted to a binary variable Hits(1) & Otherwise (0) where Hits are those games that reaches a sales of \$5m or more. The independent variables are “release date, names, genres (action, adventure, sports, golf, flight, party, simulation, platformer, racing, puzzle, war games, strategy) metacritic, user rating, developers, players and game ratings (T, M & E10+). Although, “genre” is a nominal variable with many categories, it is transformed into multiple binary variables (0, 1) and those binary variables are used to predict the target variable. Three predictive models are built and compared: Partition, Regression and Neural network. In order to judge the best performer among the predictive models, Model Comparison was applied. Almost all of them performed good in giving the results, however Regression was the best among all of them.

Regression performed the best among the three predictive models followed by Partition and Neural Network

| Creator            | 2.4.6.8 | RSquare | Generalized RSquare | Mean-Log p | RMSE   | Abs Dev | Mean Misclassification Rate |
|--------------------|---------|---------|---------------------|------------|--------|---------|-----------------------------|
| Neural             | 0.2495  | 0.3375  | 0.5244              | 0.4165     | 0.3517 |         | 0.3575                      |
| Partition          | 0.2737  | 0.4235  | 0.5116              | 0.4117     | 0.3441 |         | 0.2333                      |
| FitNominalLogistic | 0.4514  | 0.8293  | 0.3718              | 0.3520     | 0.2475 |         | 0.1778                      |

## JMP OUTPUTS

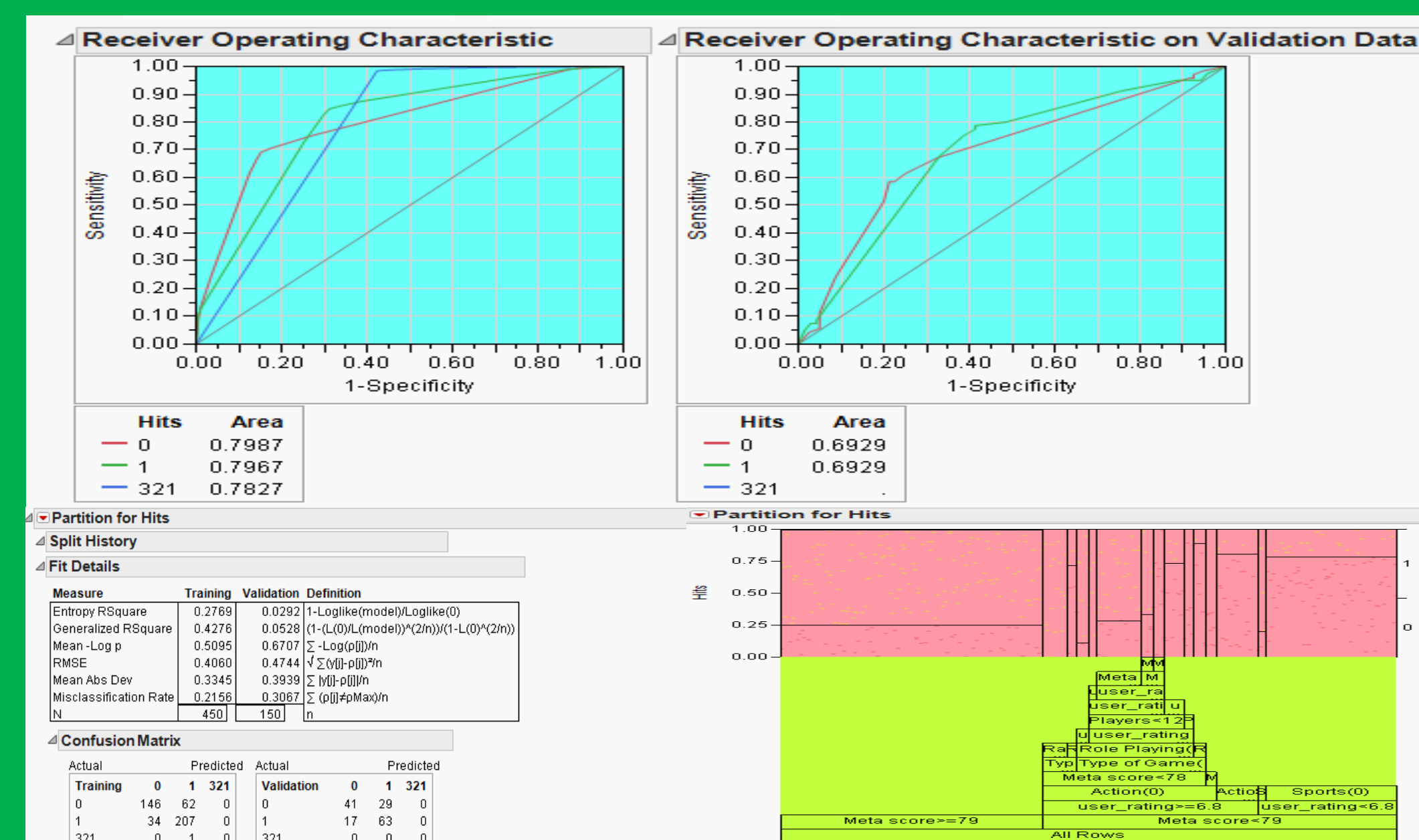


Figure 1. Screen shots of Partition(Decision Tree)

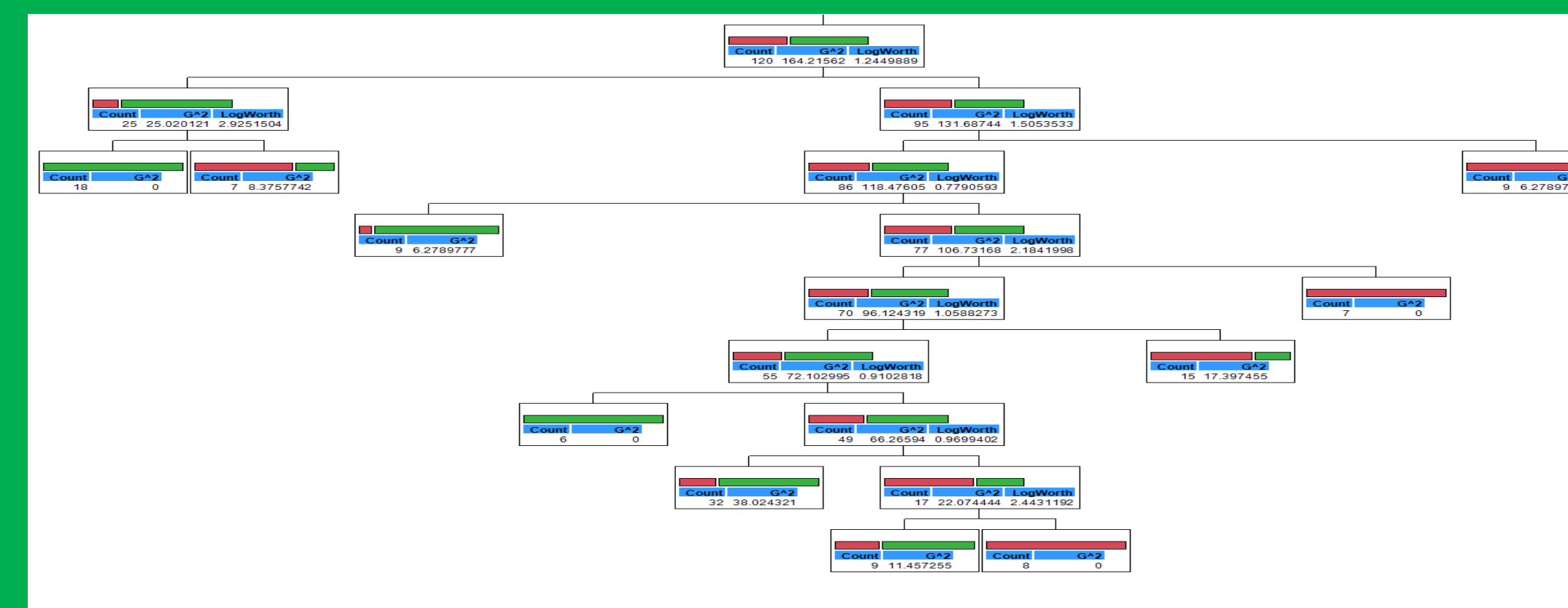


Figure 2. Screen shots showing the results for the "Hits" Games by Partition

## Predictive Modeling

Three models were applied to do the prediction for the success rate of video games. Before the models were applied a partition method was applied (training 75%) to build the models and (validation 25%) to test the predictive model's results. A model comparison node was used to test and compare the performance of each model against other models. The parameters for judging were based on the misclassification rate and Rsquare values. In the final stage the best model was chosen by the Model comparison node as the **Logistic Regression** based on those selection method.

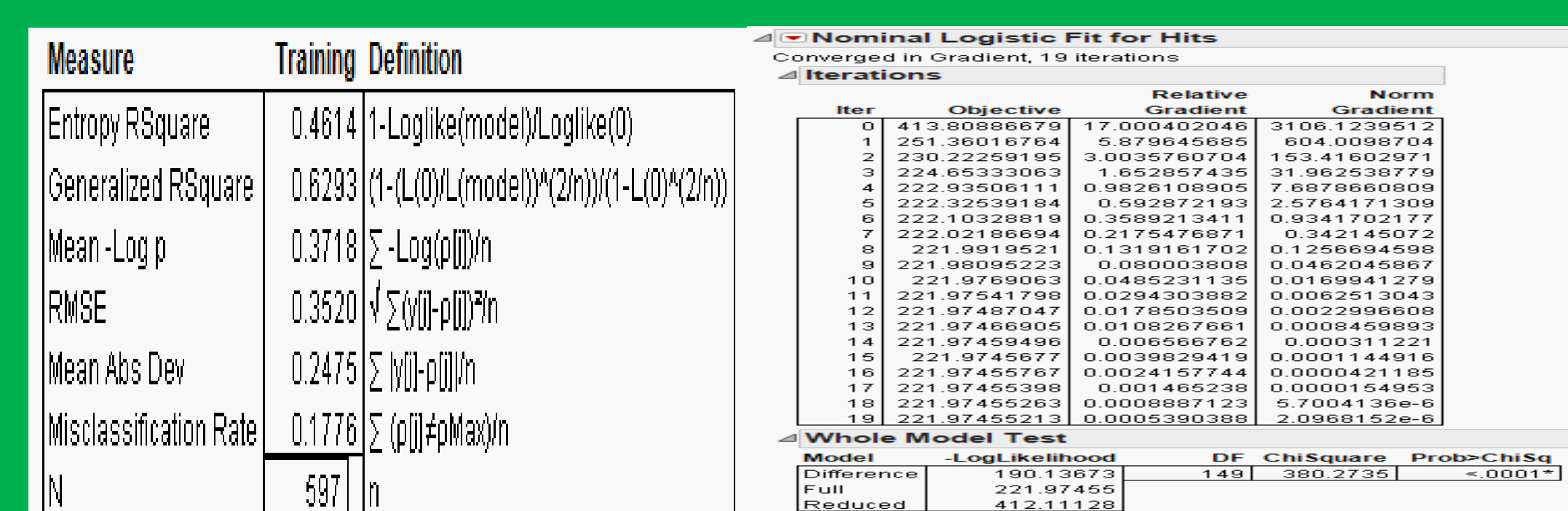


Figure 3: Screen shots of Regression showing the iterations and Results

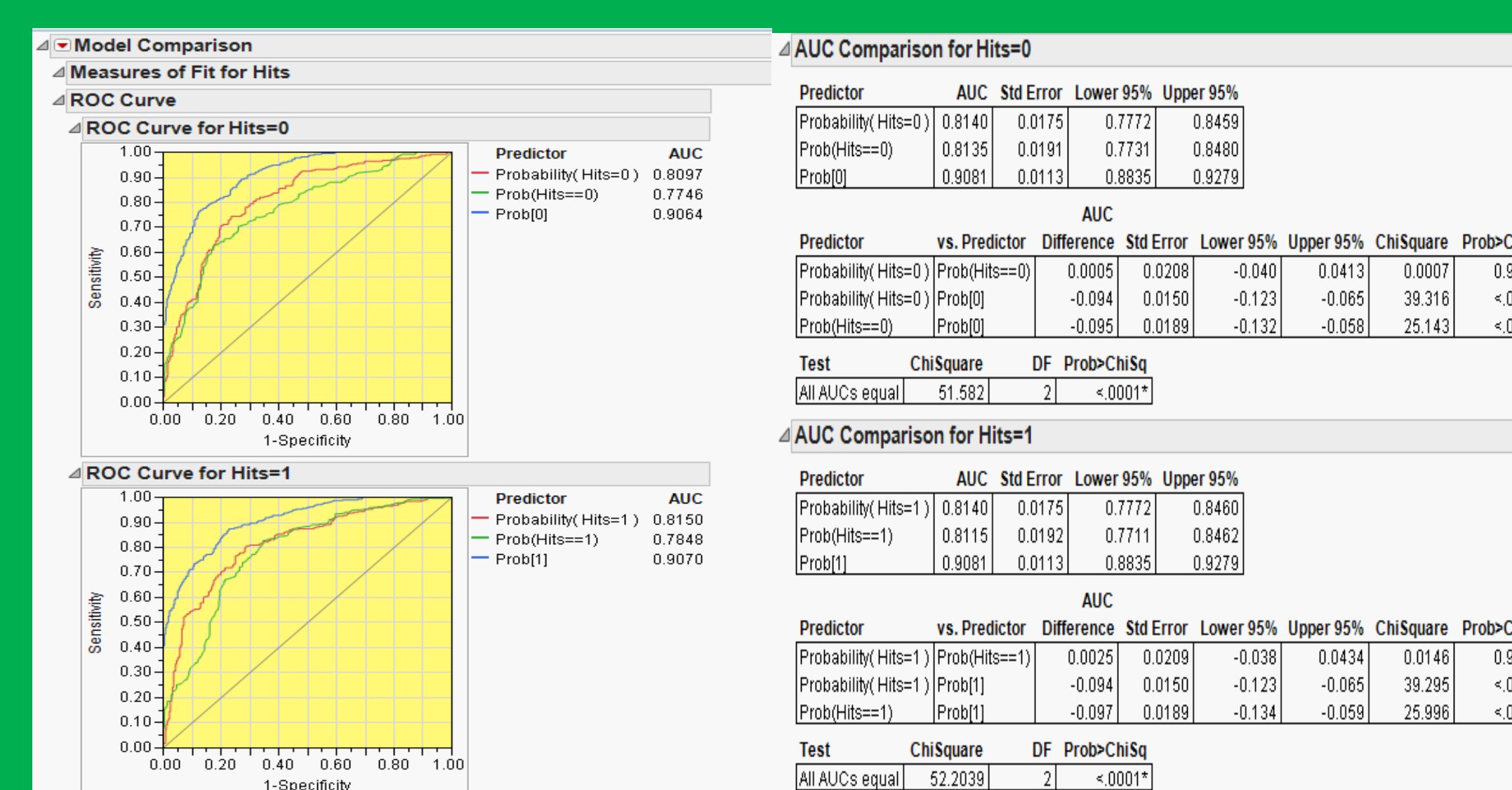


Figure 4: showing the Model Comparison of the models NN, Regression & Partition

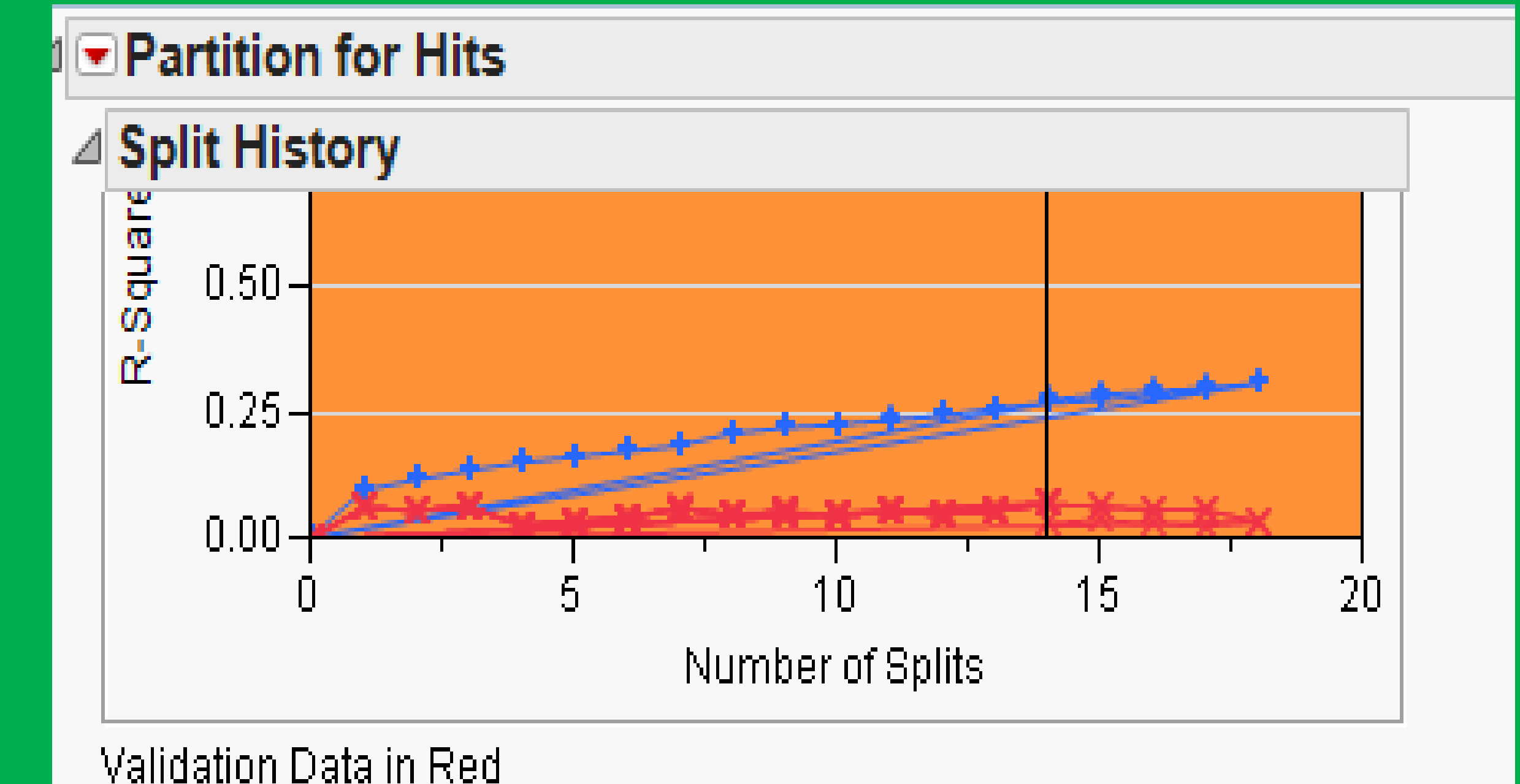


Figure 5. Partition split history

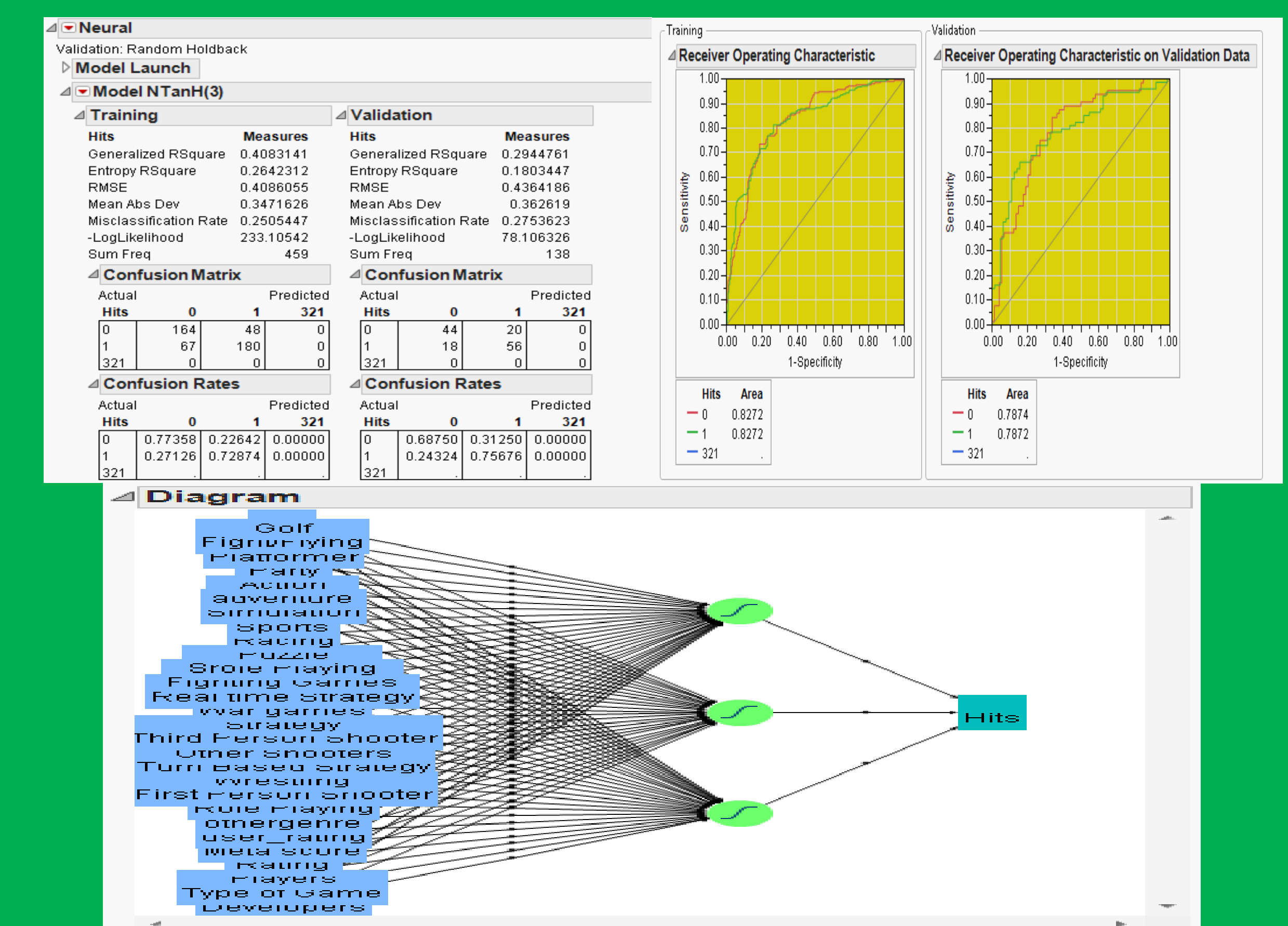


Figure 6 : Screen shots of Neural network showing the Hidden Layers and ROC

## Conclusion

Based on the results of Regression and Partition, some common variables emerged as important predictor variables. “Metacritic” and “users ratings” are the most important variables followed by “Types of games” for both the models. Regression also pointed out “Developers” as another significant variable to be taken into consideration.

The prediction of the success rate of video games would help the video game companies to come up with new and innovative demanding types of Games. The reviews and the users scores provide a lot of information regarding customer's satisfaction levels and expectations. It is gratifying to see that the marketplace success of a video game is driven by users and critics' ratings. This suggests that game developers should involve users and experts (critics) early in the game development process to make sure their views are heard and incorporated in game design.

## Reference

- John Sterman, Khan Jekarl, Cate Reavis : Sony's Battle for Video Game Supremacy (Sept 2008)
- Efraim Turban, Ramesh Sharda, Dursun Delen: Decision Support and Business Intelligence Systems(9<sup>th</sup> edition)

## Acknowledgements

Dr. Goutam Chakraborty and Dr. Ramesh Sharda