The research question for this problem was “What is the highest prediction accuracy that we can have regarding RFID adoption in a supply management firm?” A survey was filled by 755 managers and due to incomplete data, 604 observations were used in the analysis. There are multiple classification methods that can be used for this problem. In this paper, we have decided to use Logistic Regression since Discriminant Analysis requires the independent variables to be MV normal and the dataset contains categorical variables. However Logistic regression does not require this assumption and only requires that the error follows a Bernoulli distribution.

Since the objective of this study is to achieve the highest prediction accuracy for RFID adoption, the recommended evaluation metric for model performance would be its sensitivity and false positives rates. We started this analysis with a stepwise logistic regression having α=0.1 for a variable to stay in the stepwise. Four variables eventually stayed in the stepwise model. These variables are (ordered by their Chi-square score): The total number of other information technology adoptions at firm (NUMIT), the firm’s obligation to do as the channel/supply chain leader suggests (CHLEADER), the level of firm revenue (REVENUE), and finally the level of “pick-to-light” system adoption (P2LS).

The next step was to build the logistic regression model by investigating four-way, three-way, and two-way interactions. None of the interaction effects was significant. Therefore, the main effects model using the four variables above will have the highest performance. In terms of selecting the best threshold for predicting adoption. We have noticed from examining the confusion matrix that the sensitivity and false positive rates are not yielding high prediction accuracies. For example, if we take a low probability rate for adoption such as 0.3, we will find that the model had correctly classified only 71 out of possible 153 observations that adopted RFID. This means that the sensitivity score for the best model was around 46% which is a low score. But even at that rate, the false positive rate was 60% which is high and means that in addition to the 71 observations, there were 110 more observations that were classified as RFID adopters but in reality, they are not adopters. The area under the ROC curve was 0.7 which makes the model statistically useful and better than a random classifier.

After analyzing the odds ratio estimates for the CHLEADER variable, the table shows that for every 1 unit increase in the survey item, the odds of adopting RFID increase by 21%. As for the NUMIT variable, the table shows that for every increase in other IT adoptions at the firm, the odds of adopting RFID increase by 32%. As for the variable P2LS, the table shows that switching from no-use to hi-use reduce the odds by 53% and switching from no-use to low-use increases the odds of RFID adoption by 49%. Finally for the variable REVENUE, the table shows that switching from low revenue to high revenue increases the odds of RFID adoption by 120%.

**Appendix**

The following are the outputs of the SAS code









