Firm Takeover Predictor

Ahmed Ahmed

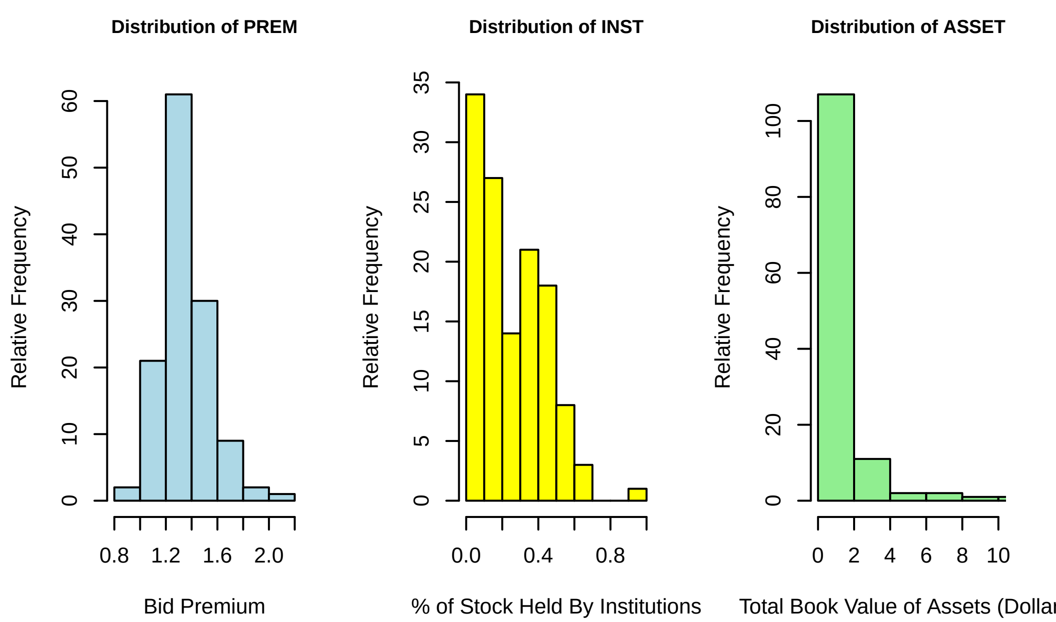
Data Description:

The data set consists of information regarding 126 firms that were targets of tender offers and taken oven within 52 weeks of initial bid.

The information can be divided into three main groups, firm specific characteristics, defensive actions by the target firm management, and intervention by federal regulators.

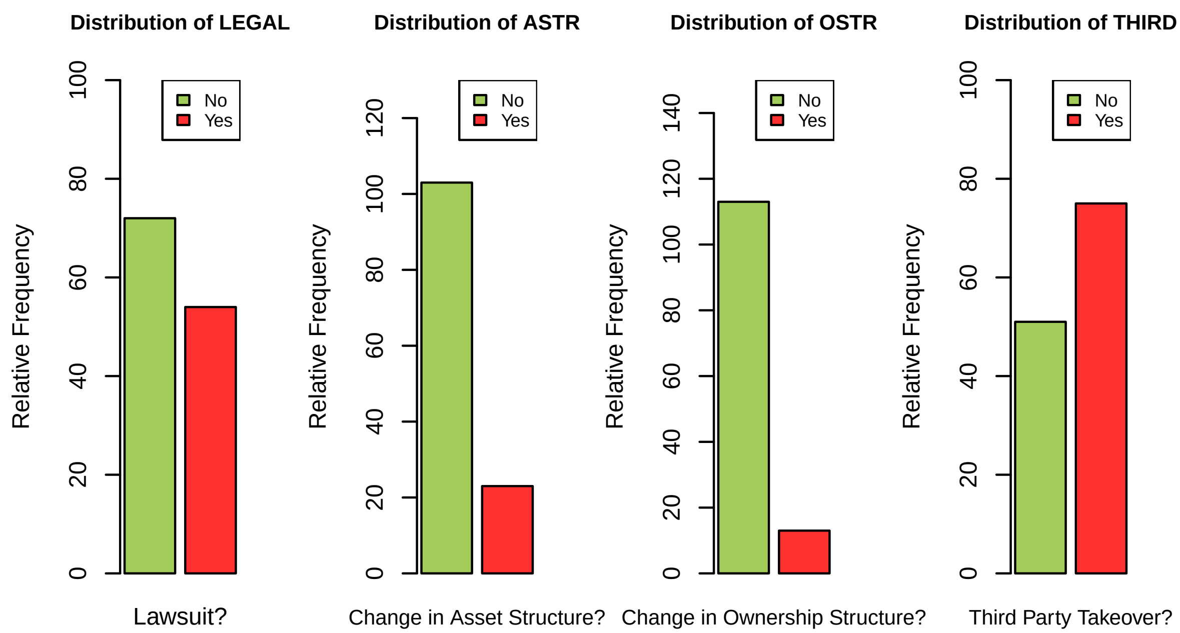
(a) Firm Specific Characteristics: These variables provide information regarding the finances of the firm.

- PREM: Bid Premium = Bid price (U.S Dollars)/ Price of Bid 2 weeks prior (U.S Dollars)
- INST: % of stock held by institutions
- ASSET: Total book value of assets (***In Billions of dollars***)



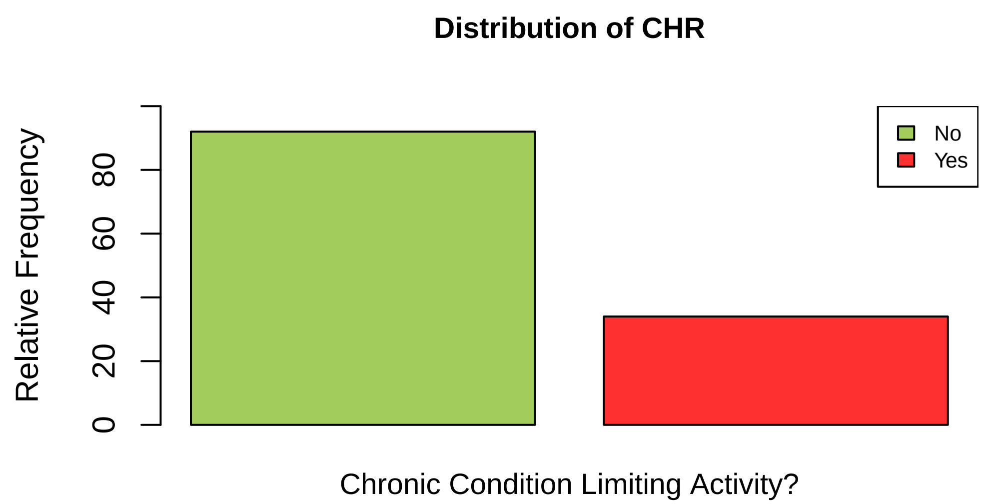
(b) Defensive Actions by target firm management: These variables provide information on how the firms reacted to the takeover.

- LEGAL: Binary variable that indicates whether there was a lawsuit
- ASTR: Binary variable that indicates whether there was a proposed change in asset structure
- OSTR: Binary variable that indicates whether there was a proposed change in ownership structure.
- THIRD: Binary variable that indicate whether there was management invitation for a third party takeover.



(c) Intervention by federal regulators:

- CHR: Binary variable that indicates whether the firm had a chronic condition that limited activity.



Finally, there's miscellaneous information such as:

- FIRM ID: The ID of the firms that were taken over within 52 weeks of the initial bid.
- Weeks: ***The number of weeks*** between the time of the initial bid and takeover.
- TOVER: Indicates whether firm was taken over. **For this data set, all firms were taken over within 52 weeks of initial bid.**

Response Variables:

- **BIDNUM: Number of bids received**
- **BIRY: Binary response which indicates whether the number of bids was less than 2, or greater than or equal to 2.**

Goal of Project:

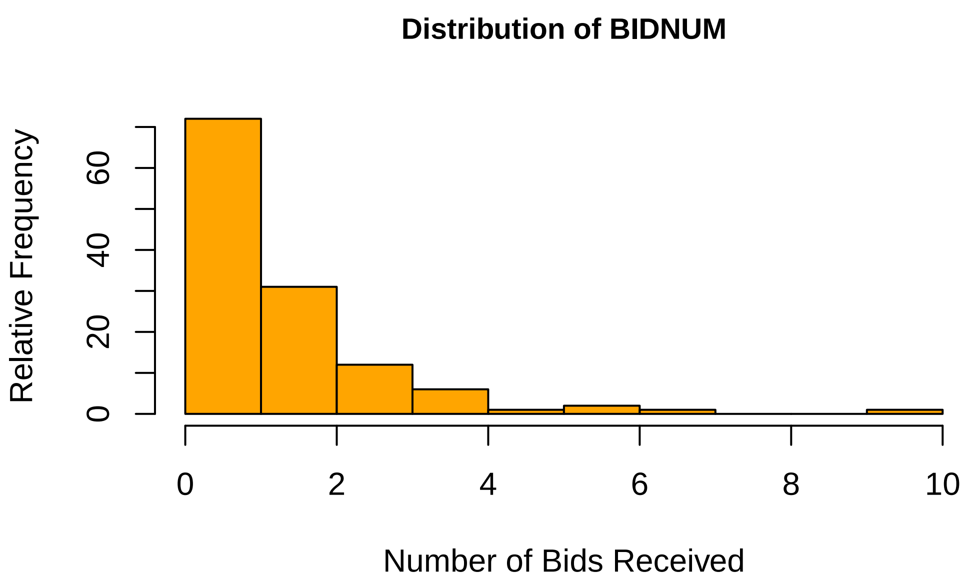
The goal of this project is to build the most accurate and simplest model, where the firm characteristics will be used as the predictor variables and the results will provide a better understanding of which groups of variables are most significant in predicting:

- Number of bids that will lead to a firm getting taken over within 52 weeks of initial bid (BIDNUM)
- Probability that number of bids that lead to a firm getting taken over will be less than 2 (BIRY = 0, otherwise = 1).

In other words, the predictive model will allow one to anticipate whether a firm will get taken over within 52 weeks based on information regarding the firm.

Statistical Methods:

BIDNUM



The data type is a count response , therefore the Generalized Linear Model (GLM) approach was taken to build the simplest model with the best predictive capabilities. The link function applied to the count response is the log link function.

Poisson Log-Linear Model :

The Poisson Log-Linear model is essentially the starting point to applying the GLIM approach for predicting our count response. This model assumes that the count response follows a Poisson Distribution. Recall that in a Poisson Distribution:

The probability distribution function (p.d.f):

$$p(y; \lambda) = \exp(-\lambda)\lambda^y/y!, \quad y = 0, 1, 2, \dots, \lambda > 0$$

Then, $E(Y) = \lambda$ and $\mathrm{Var}(Y) = \lambda$, and $Y (Count Response) $ is said to be *nominally dispersed* (mean = variance).

The model was then tested for adequecy by comparing the fitted full model (model that includes all the predictitor variables) to that of the null model (model that doesn't include any predictor variables). In addition, the fitted full model was compared to the saturated model (model with no error in the predicted response).

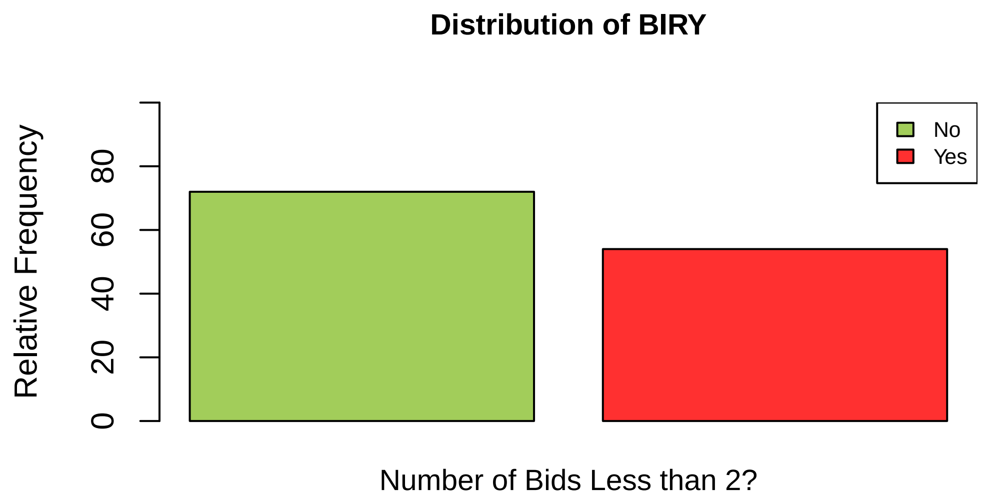
To compare the models, the Chi-Squared Test was applied to determine whether there is a significant difference between the residual deviances from each of the models. Residual deviance is a measure of how much variability in our response was accounted for or explained by the predictor variables. Furthermore, the null model is simply the intrinsic variability of our count response.

Model comparison allows us to determine whether the assumption that our count response fits a Poisson Distribution is a reasonable assumption. This is determined by how well the fitted full model performs in comparison to the null and saturated models. If the fitted full model is inferior to both models, this can be an indicator of over dispersion. The presence of over dispersion in our data is determined by how close the dispersion parameter of the fitted full model is to the value of 1. To address over dispersion, the data can be fit to the Quasi-Poisson LogLinear and negative binomial models, where the assumption that the mean and variance are equal are relatively less strict in comparison to the Poisson LogLinear Model.

The adequacy of Quasi-Poisson LogLinear and negative binomial models is determined similarly to the Poisson LogLinear Model. After dealing with over dispersion, the model is refined by selecting only the variables that were found to be significant using the obtained p-values. Finally, the models are assessed using the Mean Absolute Error (MAE) which measures how well the predicted values of the model compares to the observed responses. MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{\lambda}_i|$$

BIRY



The data type is a binary response . Therefore, both the GLMs and tree-based model approaches were applied to determine which of the approaches provide the simplest model with the best predictive capabilities.

For the GLMs approach, the Logit and Probit models were employed. In the Logit model, the logit link function is applied to the odds ratio of our binary response. This is denoted by the equation:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

Let $\eta = \text{logit}(\pi)$. After some algebra, we see that we can uniquely write π as a function of η, i.e., the inverse transformation is

$$\pi = \frac{\exp(\eta)}{1+\exp(\eta)}$$

However, in the Probit model, the probit link function is applied to the odds ratio of our binary response. Simillarily to the Logit model, the odds ratio can be extrapolated. This is denoted by the equations:

$$\Phi^{-1}(\pi_i) = \eta_i$$

$$\pi_i = \Phi(\eta_i)$$

Moreover, both the Logit and Probit models apply a function to the odds ratio of the binary response. Then, the function of the odds ratio is treated as the response in the model. Therefore, by understanding how the change in the predictor variables effects the change in the function applied to the odds ratio of the response variable, we can extrapolate the resultant effect on the odds ratio itself.

Model adequacy is determined by fitting and comparing the full binary models (all predictor variables), null models (no predictor variables), and reduced models. The reduced models are obtained by applying the step() function, which uses the null and full models as the starting and end point for selecting the predictor variables that explain the binary response. For comparison of the models, the Chi-Squared Test is applied to determine whether there is a significant difference between the residual deviance of the models. Also, its important to point out that when assessing the adequacy of a model, the interactions between the predictor variables should also be taken into consideration and compared to the models with no interactions, to determine whether including these effects improves the predictive capability of the model. The model with the lowest AIC score will be selected as the simplest model, and the accuracy of the model is assessed by evaluating the ROC curves on the test data set. The ROC curve evaluates the prediction accuracy of a model by assessing the sensitivity and specificity of the model. Sensitivity is defined as the probability that the positive class (in this case the positive class is = 0) is predicted correctly in the test data set. Furthermore, specificity is defined as the probability that the "1" response is predicted correctly (keep in mind that the positive class is defined by the confusion matrix code used). The ROC plots allows us to compare the models built using the GLIM and tree-based approaches, to select the simplest and most accurate model.

For the tree-based approach, the CART Algorithm, Random Forest and XGBoost models were applied and evaluated in determining the simplest and most accurate model in predicting the binary response in the test data set. These methods are similar in the sense that a binary tree is constructed by repeatedly splitting a node into two child nodes to determine which predictor variables provide the best predictive capabilities of the binary response variable. However, the models are different in the manner in which the nodes are split, the manner in which the accuracy of the model is assessed after a split, and how the algorithm of the model decides when to stop splitting.

In the CART Algorithm, the binary tree is initially constructed by taking into consideration all predictor variables in the train data. The algorithm applies a “greedy” approach, where the very best split point is chosen each time at node. After building the tree, the Cost of the Tree is evaluated. The Cost is defined as:

$$\text{Cost}_{CP}(\text{Tree}) = \text{Error}(\text{Tree}) + CP \times \mathcal{N}(\text{Tree}),$$

Where CP denotes a complexity parameter, which is the minimum improvement in the model required at each node. As CP increases, a penalty is incurred proportional to the number of leaf nodes $\mathcal{N}(\text{Tree})$. Therefore, we select the value of CP which gives the subtree resulting in the smallest cross-validated prediction error (xerror) from the output. Xerror is a measure of the predictive performance of the model. To further optimize the initial model, the binary tree is pruned to determine the number of splits required and CP value that corresponds to the smallest xerror. Its important to note that the CART algorithm splits the tree based on the best possible choice at each stage, without taking into consideration whether those choices remain optimal. Furthermore, pruning the tree is a crucial step in obtaining the simplest and most accurate model using this method.

The random forest (RF) model is an ensemble learning method. The term “random” is used to define this model because predictors are chosen randomly from the full set of predictors in the training data when building a decision tree. In addition, the term “forest” denotes the fact that the model uses outputs from multiple trees to make a decision. In other words, unlike the CART Algorithm, the RF approach generates multiple decision trees to determine the optimal criterias for splitting at a node, and by majority vote the RF method applies the learnings to determine which predictor variables have a significant impact on the binary outcomes. Finally, the trained model can then be applied to the test data for comparing its predictive capabilities.

The XGBoost method is similar to the RF in the sense that it uses multiple binary decision trees to generate a model with optimal predictive capabilities. However, the distinguishing factor between RF and XGBoost is that in RF, the multiple decision trees are created independently of each other and they are relatively deep (many splits per tree). However, XGBoost is a gradient boosting algorithm that successfully builds an ensemble of shallow trees (relatively less number of splits), where each shallow tree is generated using findings from the previous one. As a result, the combined trees provide a highly accurate predictive algorithm.

In conclusion, the CART, RF and XGBoost tree based methods explore the improvement in the predictive capability of a model as the splitting criteria at each node becomes more complex when comparing the different methods. The predictive capability of each model is determined by quantifying the probability that the “1” and “0” responses were predicted correctly in the test data set. As a result, this allows us to determine whether an enhanced splitting criteria is necessary when selecting **the simplest and most accurate model.**

Results:

Response Variable- BIDNUM:

Poisson Loglinear Model Results

Model	Residual Deviance	Deviance Parameter
Fitted Full Model	76.1	0.85
Null Model	105.7	-
Reduced Model	80.2	-

Table 2: Chi-Squared Test Comparison

Models Compared	p-value obtained from Chi-Squared Test
Fiitted Full and Null models	0.00099
Fitted Full and Saturated models	0.85
Fitted Full and Reduced models	0.77

Assuming an alpha of 0.05 (95% Confidence), **the Chi-Squared test favors the fitted full model when compared to the null and saturated models. In addition, when comparing the fitted full to the reduced model, the test shows that dropping the insignificant variables provides similar predictive capabilities as the fitted full model.**

Table 3: Reduced Model Coefficients

Significant Predictors	Model Coefficient
ASSET	0.035
LEGAL	0.37
THIRD	0.53

In the Reduced Poisson Loglinear model, LEGAL, ASSET and THIRD were determined to be significant predictor variables. In addition, THIRD was determined to be the most significant predictor based on the value of its coefficient.

Effect of Significant Predictors on BIDNUM

Table 4: **In conclusion, management invitation for a third party takeover was found to have the highest effect on number of bids received for a firm takeover, where the change in number of bids increased by 3 bids. This is followed by the presence of a lawsuit, where the count response increased by 2. Finally, the firms total book value resulted in the number of bids to change by 1.**

Predictor Variable	Reduced Model Coefficient	Effect on BIDNUM = Log Inverse of η
LEGAL (Presence of Lawsuit)	0.37	2.1
ASSET (Total Book Value in Billions)	0.035	0.97
THIRD (Invitation for takeover)	0.53	3.03

Accuracy of Model:

Simplest Model	Mean Absolute Error
Reduced Poisson LogLinear Model	0.88

The reduced model was approximately 88% accurate in predicting the count response in the test data set. In addition, there were no indication of over dispersion and this can be confirmed by comparing the MAE score to that obtained using the Negative Binomial Regression Model .

Negative Binomial Model Results:

Table 6: ***Residual Diagnostics Results***

Test Empoloyed	Coefficient Value	p-value
Shapiro-Wilk Test	0.976	0.0681

Table 7: ***Comparing Fitted Full to Reduced Model***

Models to be compared	p-value obtained from Chi-Squared Test
Fitted Full and Null models	0.002
Fitted Full and Reduced models	0.77

The Chi-Squared test favors the fitted full model when compared to the null model. In addition, when comparing the fitted full and reduced models, the test showed that the dropped variables don't have a significant effect on the predicted number of bids received.

Effect of Significant Predictors on BIDNUM

Predictor Variable	Reduced Model Coefficient	Effect on BIDNUM = Log Inverse of η
LEGAL (Presence of Lawsuit)	0.37	2.1
ASSET (Total Book Value in Billions)	0.035	0.97
THIRD (Invitation for takeover)	0.53	3.03

Accuracy of Reduced Negative Binomial Model:

Simplest Model	Mean Absolute Error
Reduced Poisson LogLinear Model	0.88

The accuracy of the reduced Poisson and Negative Binomial Regression models were almost identical when predicting the test data set. This confirms that there isn't evidence of over dispersion in our data. Both models show that LEGAL, ASSET and THIRD are significant predictor variables of the count response, "number of bids received" (BIDNUM) with similar effects.

Response Variable-BIRY

Logit Model Results:

Table 10: ***Ch-Squared Test Comparison***

Models Compared	p-value obtained from Chi-Squared Test
Fitted Full and Null models	0.006
Fitted Full and Reduced models	0.96

The Chi-squared test was used to compare the fitted full to the null and reduced models. The fitted full was found to be significantly better than the null model. (Assuming alpha = 0.1). In other words, the reduced model is the most parsimonious model obtained using the Logit approach.

Table 11: ***Coefficients of Significant Predictors Obtained Using Reduced Model and its effect on the Logit Function of the Binary Response (BIRY)***

Significant Predictor	Reduced Model Coefficient	Effect on Logit Function
THIRD	1.33	2.99
PREM	-2.92	-1.25
WEEKS	0.10	1.76

The significant predictors were determined to be THIRD (Invitation from management for third party takeover),PREM (Bid Premium), and Weeks (Number of weeks between initial bid and takeover). THIRD and WEEKS were found to significantly increase the odds that the number of bids received for a firm takeover will be greater than 2. However, PREM had the opposite effect, it was determined that the higher the bid premium, the odds that the number of bids received for a firm takeover will be greater than 2 significantly decreases.

Assessing Model Accuracy:

The accuracy of the reduced logit model was compared to that of the reduced logit models (with interactions).

Model	Accuracy on Test Data	Sensitivity	Specificity
Interactions	52.1	100%	0%
No interactions	71.5	100%	0%

Including the interaction between the predictor variables doesn't improve the accuracy of the model. Moreover, the Reduced Logit Model (No interactions) was found to be the simplest and most accurate model obtained using the Logit approach.

Probit Model Results:

Table 13: ***Chi-Squared Test Model Comparison***

Models Compared	p-value obtained from Chi-Squared Test
Fitted Full and Null models	0.005
Fitted Full and Reduced models	0.9506

The Chi-squared test was used to compare the fitted full to the null and reduced models. The fitted full was found to be significantly better than the null model. (Assuming alpha = 0.01). In other words, the reduced model is the most parsimonious model obtained using the Logit approach.

Table 14: ***Coefficients of Significant Predictors Obtained Using Reduced Model and its effect on the Logit Function of the Binary Response (BIRY)***

Significant Predictor	Reduced Model Coefficient	Effect on Probit Function
THIRD	0.83	1.86
WEEKS	0.06	1.09
PREM	-1.80	-0.77

The significant predictors were determined to be THIRD (Invitation from management for third party takeover),PREM (Bid Premium), and Weeks (Number of weeks between initial bid and takeover). THIRD and WEEKS were found to significantly increase the odds that the number of bids received for a firm takeover will be greater than 2. However, PREM had the opposite effect, it was determined that the higher the bid premium, the odds that the number of bids received for a firm takeover will be greater than 2 significantly decreases.

Assessing Model Accuracy:

The accuracy of the reduced probit model was compared to that with interactions.

Reduced Model	Accuracy	Sensitivitity	Specificity
Interactions	53.9%	100%	0%
No Interactions	70.9%	100%	0%

CART Algorithm Approach:

Table 16: ***Optimized Fitted Tree***

CP	nsplit	xerror
0.18	0	1

The lowest xerror value is obtained at the “0” split, pruning the decision tree is unnecessary.

Table 17: ***Assessing Model Accuracy***

Sensitivity	Specificity	Mis-classification Rate
66.7%	63.6%	34.6%

Random Forest Approach:

Table 18: ***Top 3 Significant Predictor Variables***

Predictor Variable
PREM
WEEKS
ASSET

Table 19: ***Assessing Accuracy of Model***

Sensitivity	Specificity
86.7%	54.5%

Compared to the logit/probit models, the only variable that wasn't considered as one of the most significant was THIRD. In additon, compared to the CART method, the model performed better in predicting the "0" response, but dropped slightly when predicting the "1" response.

XGBoost Approach:

Table 20: ***Optimized Fitted Tree***

Optimized Test-AUC Score	Number of Rounds
0.76	100

Table 21: ***Assessing Accuracy of Model***

Accuracy	Sensitivity	Specificity	Mis-classification Rate
73.1%	86.7%	54.5%	26.9%

The XGBoost model performed similarly to the Random Forest Model in regards to predictive capability. In comparison to the CART method. these models were more accurate in predicting the binary response variable in the testing data set. However, its important to note that models built using Random Forest and XGBoost performed slightly worse in predicting the the "1" response in the testing data set.

Summary and Conclusion:

BIDNUM:

The Poisson LogLinear Model was used to predict the response variable *BIDNUM*. The data fit the model with no signs of over-dispersion. This was confirmed by comparing the accuracy of the model to that of the Negative Binomial Regression model. Both models had similar accuracy in correctly predicting the number of bids received in the test data set. Therefore, the Reduced Poisson Loglinear Model was considered to be the simplest and most accurate model for predicting BIDNUM.

BIRY:

Both the GLM and tree-based approaches were applied to build the simplest and most accurate model for predicting the binary response in the test data set. When comparing the models built using the GLM and tree-based approaches, the Random Forest and XGBoost were slightly more accurate in predicting the binary response variable in the test data set. It was found that in the GLM models suggested that THIRD is one of the most significant predictor variables. However, the Random Forest model suggests otherwise, where THIRD wasn't even considered one of the top 3 predictor variables. Furthermore, this could explain why the models built using the GLM approach were slightly less accurate.

In conclusion, the Reduced Poisson LogLinear Model was determined to be the simplest and most accurate model for predicting BIDNUM. Findings from this model shows that the total book value (ASSET), presence of a lawsuit (LEGAL) and invitation from management for a third party takeover (THIRD) play a significant role in the number of bids that will lead to a firm takeover.

The Random Forest and XGBoost (Tree Based Models) provide the best predictive capabilities for predicting BIRY. Findings from these models show that the Premium Price (PREM), number of weeks between initial bid and firm takeover (WEEKS), and total book value of firm (ASSET) play a significant role in the probability that the number of bids that will lead to a firm takeover is greater than 2. However, for the purpose of selecting the simplest and most accurate model, the Random Forest approach would be considered the best. The reason for this is because the XGBoost approach applies a more enhanced splitting criteria for building the optimized decision tree. **Since both models were similar in accuracy, this shows that the enhanced splitting criteria is unnecessary given our data, and that the Random Forest approach is sufficient to build the simplest and best model for predicting BIRY.**