**DATA MINING FINAL PROJECT**

**Risk Management for Insurance Companies against Firms**

Team Paramount (Group 5 - Tuesday Morning Batch)

* Akash Mishra
* Gopichand Yenkipalli
* Hiranmai Vallabhaneni
* Rahul Bankey
* Shanmukh Peela
* Surya Teja Pidakala

**Business Question**:

Failure of a corporation or bankruptcy can be detrimental to both the company and the global economy. In order to lessen the financial loss brought on by bankruptcy, business professionals, investors, governments, and academic scholars have long researched strategies to recognize the possible risk of business failure. For many associated financial institutions, anticipating insolvency is a critical duty. In general, the goal is to forecast the probability that a company would fail. For financial institutions to make the right lending decisions, effective prediction models are essential.

Making the right financing decisions requires effective bankruptcy prediction on the part of financial institutions. The most significant aspects influencing the forecast are typically the input variables (or features), such as financial ratios, and prediction approaches, such as machine learning techniques.

Corporate governance indicators (CGIs), in addition to financial ratios (FRs), have been identified as another crucial category of input variable. The prediction performance attained by fusing CGIs with FRs hasn't, however, been thoroughly looked at. Only a few carefully chosen CGIs and FRs have been utilized in related investigations, and the features used may vary from one study to the next. The experimental findings, which are supported by a real-world dataset from Taiwan, demonstrate that the CGI categories of board structure and ownership structure and the FR categories of solvency and profitability are the most crucial factors in bankruptcy prediction.

In particular, a combination of prediction accuracy, ROC curve, and misclassification cost yields the greatest prediction model performance. However, in some markets, such as the Chinese market, where the definition of distressed enterprises is ambiguous and the characteristics of corporate governance indicators are not immediately apparent, these findings might not be applicable. The insurance companies have dealt with numerous problems. Following are few of them.

**Lack of trust:**

This is the rationale behind why many businesses forego insurance. Many insurance companies deny providing certain benefits and frequently underpay claims. As a result, the majority of individuals simply consider insurance to be an unnecessary expense. Financial difficulties do lead to the closure of several insurance companies.

**Competition:**

There is a significant difficulty for insurers today because there are so many insurance companies on the market. Each business searches for the optimum strategy for marketing and selling its insurance offerings. Most insurance firms, particularly new ones, are the ones that people have the most doubts about. Therefore, for the new insurance companies to succeed and hold their own against the competition, they should develop clever business plans.

**Economic instability:**

All insurance firms will be impacted by a downturn in the national economy. Similar to how interest rates on credit facilities offered by financial institutions are affected in such circumstances, insurance rates may be affected to the point where insurance companies are compelled to raise their prices.

Even if it is explicitly specified in the contract that the insurance premiums may occasionally vary, of course, no client will welcome this. Because customers can quickly distribute information about a service or product, they were unhappy with, such circumstances could harm a company's reputation.

**Driving factors:**

* To Assist the upcoming insurance companies.
* To help the existing insurance firms to make better business decisions.
* To detect the fraudulent bankruptcy issue.
* To predict the economic instability of the market.
* To reevaluate the current customer and plan the strategy to overcome the impact.

**Data source Description:**

For this study, data from the Taiwan Economic Journal1 for the years 1999–2009 were gathered. The definition of corporate bankruptcy was developed using the Taiwan Stock Exchange's business rules. Additionally, two criteria were applied for gathering the data samples. First, the sample companies had to have three years' worth of full public disclosure prior to the financial crisis. In order to compare the bankruptcy and non-bankruptcy cases, there should be a sufficient number of comparable businesses in the same sector that are of a similar size. The resulting sample includes businesses from the manufacturing sector, which consists of industrial and electronics businesses (346 businesses), the service sector, which consists of businesses in the travel, retail, and shipping sectors (39 businesses), and other sectors (93 businesses), but not financial businesses.

It should be noted that if there is a significant difference between the number of bankrupt and non-bankrupt cases, this results in a class imbalance problem.

We will analyze the dataset containing 6280 observations with 96 columns. The following features of the data will allow us to predict whether an insurance company any must give insurance to firms or not based on their chances of bankruptcy.

Below are some of the major impacting variables:

|  |  |
| --- | --- |
| Sr No | **Target Variable:** |
| 1 | Bankrupt: 1 if the company is going to be bankrupted, 0 if not |

|  |  |
| --- | --- |
| Sr No | **Predictor variables:** |
| 1 | Net Income to Total Assets |
| 2 | Interest Coverage Ratio (Interest expense to EBIT) |
| 3 | Cash Flow to Liability |
| 4 | Retained Earnings to Total Assets |
| 5 | Total Asset Growth Rate |
| 6 | Operating Profit Rate |
| 7 | Cash flow rate |
| 8 | After-tax net Interest Rate |
| 9 | Operating Profit Per Share |
| 10 | Operating Expense Rate |
| 11 | Cash Flow to Sales |
| 12 | Cash Reinvestment % |
| 13 | Cash Flow Per Share |
| 14 | CFO to Assets |
| 15 | Cash/Current Liability |

**Past Analysis on the Dataset:**

To anticipate the success of predicting whether a company will go bankrupt or not, the author of this dataset has already taken on this task and offered a few data mining (DM) approaches.

* Constructed a correlation matrix to observe the strength of relationships of each variable with bankruptcy.
* Built a few models with an accuracy of around 70 per cent.

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
|  | **Actual Yes** | **Actual No** |
| **Predicted Yes** | * Actual – Company will go bankrupt. * Predicted – Company will go bankrupt.   Our model is predicting the company will go bankrupt and in actuality, it will go bankrupt.  **(True positive)** | * Actual – Company will not go bankrupt. * Predicted – Company will go bankrupt.   Our model is predicting the company will go bankrupt and in actuality, it will not go bankrupt.  **(False Positive)** |
| **Predicted No** | * Actual – Company will go bankrupt. * Predicted – Company will not go bankrupt.   Our model is predicting the company will not go bankrupt and in actuality, it will go bankrupt.  **(False Negative)** | * Actual – Company will not go bankrupt. * Predicted – Company will not go bankrupt.   Our model is predicting the company will not go bankrupt and in actuality, it will not go bankrupt.  **(True Negative)** |

**Analysis of the Dataset:**

* We are dropping columns that are unwanted for the analysis.
* We will be checking which sampling technique can be used to balance our dataset.
* Perform data cleaning, transforming, and converting column datatypes if necessary.
* We then perform binary classifiers (classification models) like the KNN model, Decision Tree, AdaBoost, Gradient Boost, Random Forest, XGBoost, and Logistic Regression and we are predicting which model best fits the dataset with a greatrecall score.
* Also, we will be checking the model recall value with different parameter tuning techniques.

Selection of “**Recall’’** as a deciding parameter here, is because we think that the FN case has a greater impact on the scenario as compared with the FP case. i.e. Our model is predicting the company will not go bankrupt and in actuality, it will go bankrupt in this case the insurance company will have to pay for the losses.

**Dataset Attachment:**

****