THEME

Strengthening Professional Capacity: Advancing Global Valuation Competence

Financial and Non-Financial Determinants of Firm Bankruptcy in India

Aditya Deva[1] [1]and Karan Kumar[2]

Abstract:

Bankruptcy has a direct implication on valuation. The recovery rate and loss severity in the claims of various lienholders depend on the probability of going bankrupt. The study investigates a wide array of financial and non-financial variables sourced from bankruptcy records provided by the Insolvency and Bankruptcy Board of India (IBBI) spanning 2017 to 2022. We draw additional data from the Centre for Monitoring Indian Economy Prowess IQ portal (CMIE). We apply machine learning (random forest) and econometric (logistic regression) techniques to examine the determinants of bankruptcy in Indian firms. As expected, the odds of bankruptcy exhibit a negative association with return on assets (PAT/TA) and interest coverage ratio (EBIT/Interest), but a positive association with debt-to-equity and debt-to-asset ratios. We also find a significant association between a firm's, age and the type of headquarters city as significant factors affecting a firm's likelihood to go bankrupt. In particular, the association between age and the likelihood of bankruptcy is non-monotonic, suggesting an optimal seasoning life in firms. Logistic regression demonstrates an 80% accuracy rate, while random forest excels with an impressive 88% accuracy, reinforcing the potential of these models for real-world applications.

Keywords: Bankruptcy, Insolvency, Machine learning, Logistic regression, Financial Creditor (FC), Operating Creditor (OC), Corporate debtor (CD)

1 Introduction

Bankruptcy prediction involves evaluating a company's creditworthiness and financial health to determine the likelihood of it going bankrupt in the future.

[1&2]Indian Institute of Management Ahmedabad, Gujarat 380015, India

Creditors use these predictions to make informed decisions about extending credit, adjusting credit terms, and formulating risk management strategies. As per the Indian Insolvency Code 2016, Insolvency refers to a financial state in which an individual or entity is unable to meet its financial obligations or pay its debts when they come due. Bankruptcy is a legal process that provides a formal solution to address the issues of insolvency. It offers a structured framework for managing and resolving financial difficulties while protecting the rights of both debtors and creditors.

The Insolvency and Bankruptcy Code (IBC) of India, enacted in 2016, has brought about a structured process, termed as Corporate Insolvency Resolution Process (CIRP), for resolving insolvency cases. This has led to increased transparency around the duration, claims, and haircut (loss of claim amount) associated with insolvency. The IBC also enables a creditor-in-control process, where the insolvency initiation from the creditor/debtor leads to an outcome of liquidation or resolution of the defaulted company. The outcome of the insolvency process has a significant impact on the extent of credit forgone by the financial and operational creditors of the company.

The creditors of a firm include Operating Creditors (OC), and Financial Creditors (FC), they have a huge stake in the performance of the firm, and if they perceive that the firm is underperforming and unable to meet its debt obligations, they have the authority to initiate bankruptcy proceedings. The CIRP can be triggered by OC, FC, or by Corporate Debtor (CD) itself through the corporate applicant. The applicant may also suggest an interim resolution professional (IRP), whom the adjudicating authority would approve. The IRP shall collect all creditor claims and establish a Committee of Creditors (COC) composed of Financial Creditors (FC) and Operational Creditors (OC).

Once the COC is formed, a Resolution Professional will be appointed by the committee to gather data on the corporate debtor's assets, financial status, and operational activities to assess its overall financial standing. Resolution involves submitting a bankruptcy petition to the National Company Law Tribunal (NCLT) for adjudication under the Insolvency and Bankruptcy Code of 2016 (IBC 2016), with the resolution process overseen by the Insolvency and Bankruptcy Board of India (IBBI). The RP will come up with a resolution plan which is prepared based on information memorandum.

The resolution process can result in three potential outcomes:

Liquidation: This involves the complete dissolution of the firm, where the company's assets are auctioned off, and the proceeds from the liquidation are distributed among the creditors. Essentially, the firm ceases to exist as it is dismantled, and its financial resources are used to settle outstanding debts.

Restructuring: In this scenario, the creditors play a pivotal role in the firm's management and decision-making processes. They gain substantial control over the company's affairs with the aim of reviving and reorganizing its operations. This approach seeks to salvage the firm from bankruptcy and avoid liquidation.

Withdrawal: There's also the possibility that the firm's promoters or owners may decide to settle the outstanding dues with the creditors. When this occurs, the petitioner who initiated the bankruptcy proceedings may choose to withdraw their petition. In essence, this action averts the need for further legal intervention, as the financial obligations are met, and the firm avoids the bankruptcy process.

Advanced data analytics, machine learning, and neural networks have become valuable tools for assessing a firm's financial trajectory using various financial and economic variables. Researchers have worked on creating indices and models that can aid in evaluating credit risk and predicting bankruptcy. As high non-performing assets (NPAs) and haircuts in insolvency have been a major drag on the profitability of financial institutions, predictive models on company bankruptcy could help mitigate such losses.

The publicly available data of bankrupt companies through the IBC enactment provides an opportunity to build such models. However, it is important to note that these models are prone to type I and II errors, and the financial implications of both types of errors would be different. Type I error refers to the situation where a company is incorrectly classified as bankrupt, while type II error refers to the situation where a company is incorrectly classified as non-bankrupt.

The objective of this report is to develop a bankruptcy prediction model using the data of bankrupt companies in India. The model will be evaluated based on its accuracy.

2 Literature Review

Bankruptcy Prediction: A Holistic Approach

Bankruptcy prediction, a critical domain within financial analysis, has traditionally relied on quantitative financial metrics. However, recent advancements in machine learning techniques have reshaped this landscape. Notably, researchers like Sandeepa (2018) [1] and Vandana (2022) [2] in India have harnessed neural networks and random forest algorithms, demonstrating their superior predictive capabilities.

Despite these strides, a significant gap persists: the underexplored role of non-financial determinants in bankruptcy prediction. This study aims to bridge this gap by integrating qualitative factors—such as corporate governance quality and market dynamics—into traditional models. By doing so, we offer a more nuanced and holistic approach to bankruptcy prediction.

Machine Learning for Financial Solvency Prediction

Abdullah M. (2021) provides compelling evidence for the application of machine learning in predicting financial solvency. His study achieved an impressive 88% accuracy rate for listed firms in Bangladesh. This underscores the effectiveness of advanced computational techniques within the South Asian context [3].

Altman et al. (2017) reaffirm the robustness of the Z-score model—a time-tested tool that has weathered the test of decades. With an accuracy range of 75–90%, the Z-score remains a reliable indicator of financial distress [4].

Sehgal et al. (2021) delve deeper by identifying six critical parameters essential for assessing corporate distress. Their research employs support vector machines (SVM), artificial neural networks (ANN), and logistic regression to signal early signs of distress. Importantly, their work extends beyond financial metrics, considering a broader range of variables [5].

3 Methodology

This study's approach to bankruptcy prediction modelling harnesses the predictive power of logistic regression and random forest methods, tailored to accommodate the multifaceted nature of bankruptcy determinants. Logistic regression is ideal for examining the relationship between a binary dependent variable and a set of independent variables, offering interpretable odds ratios that quantify the impact of predictor changes. The random forest model, a non-parametric algorithm, provides a robust mechanism for capturing complex interactions and non-linear relationships without the need for transformation or assumption conformity.

The dataset, drawn from bankruptcy records from the Insolvency and Bankruptcy Board of India and the CMIE ProwessIQ database, comprises both financial ratios and non-financial variables. The latter includes qualitative factors such as management quality and market perception, which have been encoded quantitatively to facilitate statistical analysis. The data preprocessing phase included normalisation, outlier detection, and treatment, ensuring a clean dataset for model training.

Data partitioning followed the 60-20-20 rule for training, validation, and testing sets. Variable selection was guided by the variance inflation factor to mitigate multicollinearity and the Akaike information criterion for model optimization. The logistic regression model was further refined using stepwise regression techniques to determine the most significant predictors, while the random forest model leveraged an iterative process of feature selection based on variable importance scores.

The final models were put through a lot of tests to see how well they could predict the future. Accuracy, sensitivity, specificity, and AUC metrics were calculated from confusion matrices and ROC curve analysis. This methodological rigour ensures the robustness and reliability of the bankruptcy prediction model, positioning it as a valuable tool for creditors and financial analysts in the Indian market.

4 Data

4.1 Data Collection

We used the IBBI data to extract the list of companies that have gone bankrupt since 2016, the list contained details of 1961 Liquidate companies & 680 restructured companies. After choosing the relevant set of independent variables to be used for modelling, we decided to source the financial and non-financial data corresponding to them from the CMIE portal. After fetching the data, we chose 247 and 87 companies from liquidation and restructured list respectively as a sample for the model. Subsequently, we also built a control sample of 338 non-bankrupt companies, by taking enterprise value, industry type and the year of bankruptcy as a reference to map non-bankrupt companies with the bankrupt ones. The enterprise value for the bankrupt companies was estimated as the total claim value raised by the creditors and for the non-bankrupt ones, it was calculated as sum of market capitalization and total debt subtracted by the available cash. Financial data of all the bankrupt and mapped non-bankrupt companies are taken with respect to t-1 year (t = Insolvency year of the company).

4.2 Data Definition

Credit institutions assess numerous variables to determine the creditworthiness of a business. To augment the current assessment process, we researched a list of important variables in different analytical sections of the financial statement to come up with an exhaustive set of financial ratios. We also chose a list of non-financial parameters with a hypothesis of it impacting the probability of bankruptcy of a company.

Among the non-financial variables, we had an “industry type” variable, which had a total of 74 distinct industry segments with varied distribution of companies among them, which made it difficult to incorporate it in the model as an independent variable. We used expert guidance about the businesses in each industry leg and clubbed them into 12 distinct industry categories based on the similarities in their core business operations. Furthermore, we used those 12 industry categories to classify companies into three asset class groups based on the tangible asset to total asset ratio.

The final list of variables decided for bankruptcy prediction model are as follows:

Table 1. Financial Variables

Variable Names	Description
C/CL	Cash to Current Liability Ratio
C/TA	Cash to Total Asset Ratio
C/TL	Cash to Total Liability Ratio
CA/CL	Current asset to Current Liability Ratio (Current Ratio)
CA/TA	Current asset to Total Asset
CL/TL	Current Liability to Total Liability Ratio
PAT/TA	PAT to Total Asset Ratio (Return on Assets (ROA))
PAT/TL	PAT to Total Liability Ratio
EBIT/IE	EBIT to Interest Expense (Interest Coverage Ratio)
D/E	Debt to Equity Ratio
D/A	Debt to Total Assets Ratio

Source: Authors' Computation

Table 2. Non-Financial Variables

Variable Names	Description
Age	Bankruptcy Year - Incorporation Year
City	Registered City (Tier I/II/III)
Govt Parity	Parity = 1 (Same political party ruling at both centre & state), otherwise 0
Industry Category	Clubbed 74 industries in 12 categories
Asset Type	3 types - Heavy, Medium, Light as per tangible asset to total asset ratio
Ownership	4 types - Private Indian, Private Foreign, Joint, Group

Source: Authors' Computation

4.3 Data Transformation

PAT to Total asset and PAT to total liability ratios are transformed to its exponential form as its absolute values were very small and it had both positive and negative values. To increase the absolute value and making it positive the variable is transformed to exponential form. The coefficient in logit model needs to be interpreted in the same.

5 Descriptive Statistics

We fetched all the relevant parameters of bankrupt companies for the year prior to their bankruptcy. The non-bankrupt companies' data is considered for the same year as of its mapped bankrupt company. We then analysed the descriptive statistics of all the employed parameters to build related hypotheses to be tested through the model. The descriptives are as follows:

Overall Summary

Table 3. Company Type Descriptives

Company Type	Non-Bankrupt	Liquidated	Restructured	Grand Total
# Companies	338	247	87	**672**

Source: Authors' Computation

Table 4. Ownership Descriptives

Ownership	Private Indian	Group	Private Foreign	Joint	Grand Total
Liquidated	185	57	4	1	**247**
Restructured	65	20	2	0	**87**
Total	250	77	6	1	**334**

Source: Authors' Computation

In our observations, two key findings emerge. First, we notice that a significant portion of bankrupt companies falls into the category of private Indian-owned firms, accounting for 75% of all bankruptcies. Additionally, 23% of the companies facing bankruptcy are part of a larger group ownership structure. Second, it's noteworthy that the sample population, consisting of 672 companies, exhibits a balanced distribution of data between bankrupt and non-bankrupt enterprises. Specifically, the dataset comprises 51% non-bankrupt companies and 49% bankrupt companies, indicating a nearly equal representation of both categories in our analysis.

Financial Descriptives

Table 5. Financial Ratios Descriptives

Parameter		C/CL	C/TA	C/TL	CA/CL	CA/TA	CL/TL	PAT/TA	PAT/TL	EBIT/IE	D/E	D/A
N	Bankrupt	332	333	333	332	333	333	333	333	275	333	333
	Non-Bankrupt	339	339	339	339	339	339	339	339	316	339	339
Missing	Bankrupt	1	0	0	1	0	0	0	0	58	0	0
	Non-Bankrupt	0	0	0	0	0	0	0	0	23	0	0
Mean	Bankrupt	0.02	0.01	0.01	1.55	0.48	0.43	-0.68	-0.23	-11	2	2
	Non-Bankrupt	0.16	0.02	0.07	2.75	0.53	1.39	0.03	0.08	20	2	0.71
Median	Bankrupt	0	0	0	0.47	0.49	0.3	-0.15	-0.11	-0.38	-2	1.25

	Non-Bankrupt	0.03	0.01	0.01	1.43	0.57	1.04	0.02	0.04	4	1	0.56
Std. Dev.	Bankrupt	0.15	0.06	0.03	11.5	0.32	0.42	5.71	0.79	48	22	3.41
	Non-Bankrupt	0.65	0.05	0.24	6.06	0.25	1.85	0.69	0.33	70	11	2.19
Min.	Bankrupt	-0.78	-0.6	-0.4	0	0	0	-1.04	-12	-341	-97	0.05
	Non-Bankrupt	-0.57	-0.2	-0.4	0	0	0	-3.7	-3	-76	-26	0.03
Max.	Bankrupt	2.2	0.6	0.2	195	1	2.4	0.4	0.5	226	202	39.8
	Non-Bankrupt	7.7	0.42	3	66	1	19	12	1.52	818	184	39.5

Source: Authors' Computation

Our observations in the above data descriptives highlighted two significant findings. First, we encountered a total of 83 missing data points in the dataset, with a majority of these missing values (81 in total) originating from the EBIT/Interest expense variable. These gaps in the data resulted from the absence of interest expense information for the mentioned 81 companies in the CMIE prowess dataset. To address this data gap, we employed the K-Nearest Neighbour imputation method with a k-value of 10 to effectively fill in the missing values.

Secondly, our analysis revealed the presence of several outliers within the dataset, which became evident through the extreme minimum and maximum values observed for various variables. Notably, the outlier issue was more pronounced among smaller companies that had encountered financial difficulties in the year leading up to their bankruptcy. To address this concern and ensure the robustness of our model, we implemented methods designed to identify and remove both influential and standard residual outliers. These observations underscore the steps taken to enhance the dataset's quality and suitability for our analysis.

5.1 Time Series Analysis of Bankruptcy Filing

In our time series analysis, we focused on the period spanning from 2017 to 2022, encompassing the entire list of companies sourced from the IBCC data. To gain deeper insights into the patterns of insolvency in India, we employed a decomposition method that effectively disentangles the underlying trend and seasonality from the time series plot. This approach offers valuable clarity regarding the temporal dynamics of insolvency trends over the years.

Furthermore, we compared the insolvency pattern within our dataset, specifically the data related to bankrupt companies, to the overall trend observed in the time series analysis. By aligning our dataset's insights with the overarching insolvency trend, we aimed to gain a more comprehensive understanding of insolvency dynamics in the Indian business landscape.

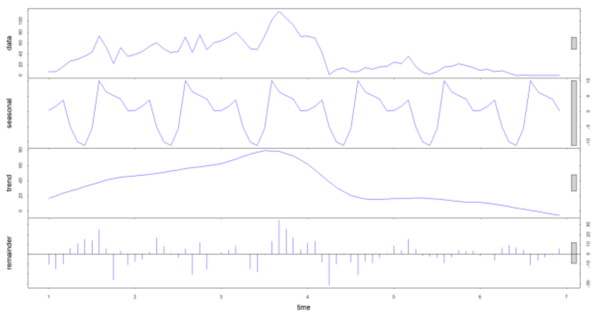


Fig. 1. Overall Bankruptcy Trend and Seasonality across six years (2017-2022)

Source: Authors' Computation

Table 6. Bankruptcy Trend in Our Dataset Population

Bankruptcy Year	2017	2018	2019	2020	2021	2022
Liquidated	71	75	79	11	8	3

Restructured	23	25	27	8	4	0
Total Bankrupt	94	100	106	19	12	3

Source: Authors' Computation

Our observations during the analysis of the dataset have unveiled significant insights into the bankruptcy trends over time. First and foremost, the bankruptcy trend within our dataset aligns with the broader bankruptcy trends, as depicted in the accompanying figure. This alignment underscores the representativeness of our dataset in capturing the larger bankruptcy landscape.

Additionally, our analysis revealed a noteworthy seasonality in bankruptcy filings, particularly highlighting peaks during the Pre-Diwali months, specifically August, September, and October. This seasonal pattern suggests an increased likelihood of insolvency proceedings during these months, followed by a subsequent decline.

Furthermore, a broader trend emerged from our dataset, showing a consistent increase in bankruptcy filings up until 2019, after which a decline commenced, coinciding with the onset of the COVID-19 pandemic. These observations provide valuable insights into the cyclical and temporal aspects of bankruptcy trends, reflecting the interplay of economic, cultural, and external factors on insolvency occurrences.

5.2 Bankruptcy Analysis across NCLT Benches and Tier Cities

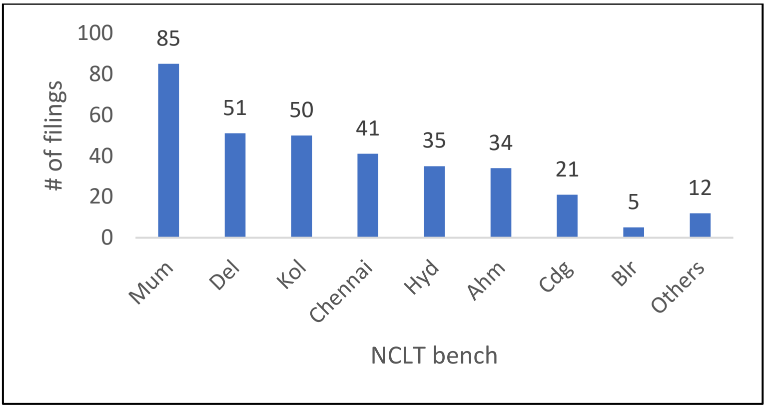


Fig. 2. Bankruptcy filing distribution across the NCLT bench

Source: Authors' Computation

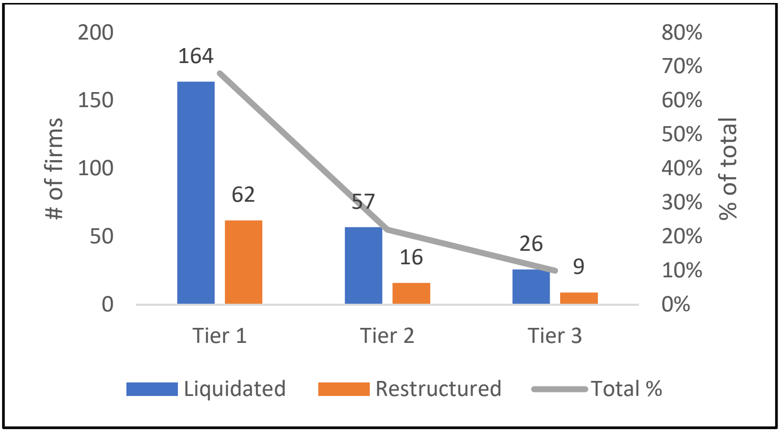


Fig. 3. Bankruptcy filing distribution across different tier cities

Source: Authors' Computation

Our observations regarding the geographic distribution of bankrupt firms within our dataset have unveiled intriguing insights into the relationship between city tiers and bankruptcy occurrences. Specifically, we found that a substantial 68% of the total bankrupted firms were located in tier-I cities. This observation aligns with the broader trend, indicating that a higher number of firms tend to establish themselves in tier-I cities. The prevalence of bankruptcies in these cities reflects their greater economic and industrial activity.

Moreover, a similar analogy extends to our observation of a high number of firms filing for insolvency in the Mumbai NCLT bench. This observation is consistent with Mumbai's status as a financial and commercial hub in India, drawing numerous companies and economic activities. As such, it is not surprising to find a heightened incidence of insolvency filings in this particular NCLT bench, given its prominence in the business landscape.

These observations highlight the correlation between the geographical concentration of firms, particularly in tier I cities and prominent commercial centres like Mumbai, and the likelihood of bankruptcy occurrences. They underscore the impact of location and urbanization on insolvency trends within the dataset.

5.3 Analysis of claim distribution across different bankruptcy claimants

In this data descriptive analysis, we delve into a comprehensive exploration of factors influencing the dynamics of creditor decisions. Specifically, we focus on understanding the patterns and trends in creditor claims within the context of corporate bankruptcy. By closely examining the data related to creditor claims, we aim to shed light on why certain factors, such as the type of bankruptcy (liquidation or restructuring), claim values, and claim ratios, become instrumental in creditors' lending determinations.

Table 7. Claim distribution among claimants.

Claimant	CD	FC	OC	Total
No. of Companies	36	202	95	333
Total Claim (INR Cr)	31,304	5,43,192	1,29,981	7,04,477
Total OC Claim (INR Cr)	26,013	4,97,125	1,20,239	6,43,377
Total FC Claim (INR Cr)	5,291	46,067	9,666	61,024
Average Claim (INR Cr)	870	2,689	1,354	2,109

Note: Corporate Debtors (CD), Financial Creditors (FC), Operational Creditors (OC)

Source: Authors' Computation

Table 8. OC Claim to FC Claim Ratio

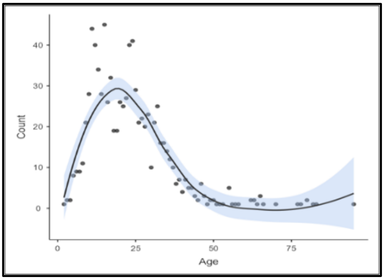
Bankruptcy Type	2017	2018	2019	2020	2021	2022	Overall
Liquidation	7.8	21	10.8	26.1	86	1.7	11.4
Restructure	10.4	9	6.2	8.8	33.3	-	9.4

Source: Authors' Computation

In our examination of the dataset, we've made several notable observations that shed light on the dynamics of bankruptcy cases and their implications for creditors. Firstly, we found that financial creditors-initiated bankruptcy tends to have significantly higher average claim values (2-3X) compared to others. This finding indicates a strong correlation between FC claims and companies with higher enterprise values, suggesting that financial creditors tend to be more heavily involved with larger enterprises.

Secondly, we observed a consistent pattern regarding the operational creditor (OC) to FC claim ratio in bankrupt companies. Specifically, this ratio tends to be higher for companies that undergo liquidation as opposed to those that are restructured. This insight implies that the distribution and proportion of claims between operational and financial creditors may play a crucial role in the bankruptcy outcomes, with liquidated companies having a higher OC to FC claim ratio. These observations underscore the intricate relationship between claim values and creditor decisions in bankruptcy cases, offering valuable insights for further analysis and decision-making in the context of lending and insolvency.

5.4 Age distribution of bankrupt companies



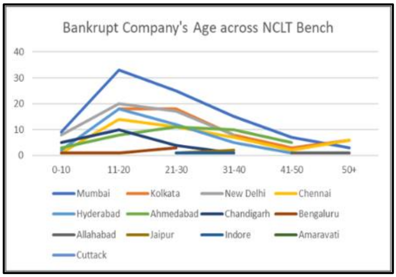


Fig 4. Age Distribution of Bankrupt Companies

Fig. 5. Age Distribution across NCLT Benches

Source: Authors' Computation

Source: Authors' Computation

A smooth and symmetrical age distribution curve was observed, which interestingly aligns with a quadratic pattern. Within this age distribution, a notable concentration of bankrupt companies emerges, primarily falling within the 15-25 years age range. Furthermore, our examination of different National Company Law Tribunal (NCLT) benches reveals a general conformity to the overarching trend. However, it is noteworthy that Ahmedabad presents a distinctive peak in the higher age range, setting it apart from the standard pattern observed in the other regions. These age-related findings underscore the significance of age as a variable in our bankruptcy prediction model and encourage further exploration into the unique dynamics within the Ahmedabad NCLT bench.

5.5. State Ruling Party Congruency during the time of Insolvency

The inclusion of government parity as a bankruptcy-predicting variable in our analysis is rooted in a fundamental hypothesis. We propose that the alignment and harmonization between state and central governments can exert a significant influence on the policy framework and support systems available within a given region. It is our belief that this, in turn, may have a substantial bearing on a company's susceptibility to insolvency. In this context, we conduct a data descriptives analysis to explore the role and impact of government parity as a vital factor in predicting bankruptcy cases. By examining the descriptive statistics of this variable, we aim to gain a deeper understanding of its significance and potential implications within the broader context of insolvency prediction.

Table 9. Government Congruency against bankruptcy

Govt Congruence	Non-Bankrupt	Bankrupt	Total
Congruence	184	146	330
Non-Congruence	155	187	342

Source: Authors' Computation

Table 10. Government Congruency against type of bankruptcy

Govt. Congruency	Liquidation	Restructure	Total	Total%	Avg. Claim (In Cr)
Congruence	109	37	146	43%	2376
Non-Congruence	138	50	188	55%	1902

Note: Govt. Congruence implies the same ruling party in federal and state government

Source: Authors' Computation

Table 11. State-wise bankruptcy and govt. congruency details

	Govt. Congruency			Govt. Non-Congruency		
States	Non-Bankrupt	Bankrupt	Total	Non-Bankrupt	Bankrupt	Total
Maharashtra	79	78	157	8	4	12
NCT of Delhi				32	52	84
Gujarat	58	25	83			
West Bengal				24	47	71
Tamil Nadu				17	39	56
Telangana				23	33	56
Haryana	12	9	21			

Karnataka	6	3	9	9	1	10

Source: Authors' Computation

It is observed that there is almost equal representation of companies from both government-congruent and non-congruent states. To ensure fair industry representation, we matched non-bankrupt companies with bankrupt companies based on similar enterprise value within the same industry, without imposing specific regional constraints, thus randomizing government congruency within industries. Despite the current political landscape in India, where the prevailing federal government party governs the majority or forms alliances in 16 out of the 29 states, it's intriguing to note that only 43% of recorded bankruptcies have occurred within government-congruent regions. Using this observation, we build a hypothesis that companies in non-congruent states (e.g., West Bengal) might be more prone to bankruptcy than their counterparts in similar domains operating in congruent states (e.g., Gujarat), whose significance will be tested in the model.

5.6 Asset classification

Table 12. Asset Classification on Tangible Asset%

Tangible Asset % Bracket	#Company	% Total	Asset Type
0-10%	149	22.20%	Asset Light
10-20%	50	7.40%	Asset Light
20-30%	65	9.70%	Asset Medium
30-40%	69	10.30%	Asset Medium
40-50%	86	12.80%	Asset Medium
50-60%	62	9.20%	Asset Heavy
60-70%	68	10.10%	Asset Heavy
70-80%	64	9.50%	Asset Heavy
80-90%	37	5.50%	Asset Heavy
90-100%	22	3.30%	Asset Heavy

Source: Authors' Computation

The dataset has nearly equal distribution of companies across different asset classes. (Asset light- 30%, Asset Medium -33%, Asset Heavy – 37%)

6 Bankruptcy Prediction Model

To construct our bankruptcy prediction model and gain valuable insights into influential variables, we opted for two robust methodologies: logistic regression and random forest models. Our evaluation of model performance hinges on crucial metrics, including R-square and prediction accuracy.

6.1 Logistic Regression Model

In the logistic regression model, we meticulously divided the dataset into training, validation, and test sets in a proportion of 60:25:15, ensuring a robust model development process. The training set was utilized to train the model, while the validation set played a pivotal role in determining the threshold value "c" for managing the different costs associated with type I and type II errors. Finally, the model's predictive performance was assessed using the test data, providing a measure of accuracy and other vital metrics.

As a first step to the model building, the correlations between financial variables were checked to make a judgment about the presence of collinearity in the model.

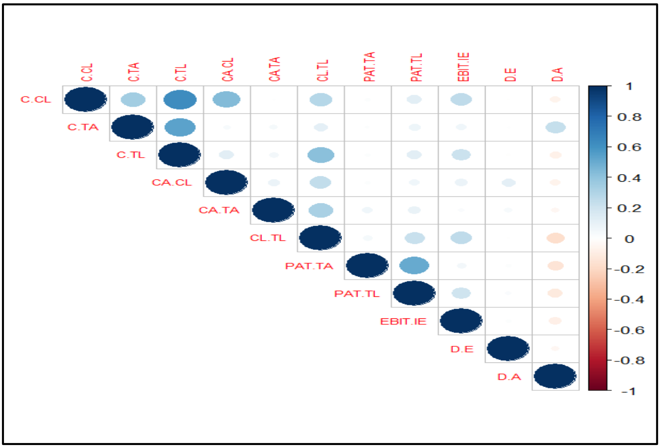


Fig. 6. Correlation Matrix of financial variables

Source: Authors' Computation

The correlation matrix unveiled noteworthy associations between variables, such as cash-to-current liability and cash-to-total liability ratios, along with PAT-to-total assets and PAT-to-total liability ratios. To address this multicollinearity, we employed the Generalized Variance Inflation Factor (GVIF), with a threshold set at GVIF>2 for variable removal. Consequently, the variable "Cash to current liability" was excluded from the model due to concerns regarding multicollinearity.

In addition to addressing multicollinearity, we also conducted an evaluation of outliers present in the data. Our approach involved the elimination of three data points exhibiting standard residuals exceeding 2.5, as suggested by descriptive statistics of financial variables. Further analysis was carried out to identify influential observations using Cook's distance, which led to the removal of two data points with Cook's distance values exceeding 1.

After addressing the outliers, we proceeded with the model, which revealed that certain variables turned out to be statistically insignificant. Through a process of iterative rounds that involves dropping insignificant variables and verifying multicollinearity, we arrived at the final model presented below:

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	4.4327	1.752E+00	2.53	0.01142
PAT.TA	-5.5467	1.527E+00	-3.633	0.0003
EBIT.IE	-0.0174	7.400E-03	-2.334	0.0196
D.E	0.0207	1.011E+00	1.96	0.04997
D.A	2.9781	5.014E-01	5.939	2.86E-09
factor(Govt. Parity) 1	-0.2504	3.208E-01	-0.781	0.43505
Age	-0.0739	2.497E-02	-2.963	0.00304
factor(City_Type)Tier 2	-0.7416	3.430E-01	-2.162	0.0306
factor(City_Type)Tier 3	-0.6421	8.517E-01	-0.754	0.4509

Null Deviance: 551.49 on 397 degrees of freedom

Residual Deviance: 274.35 on 388 degrees of freedom

AIC: 294.35

Table 13. Logistic Regression Results (Bankruptcy Model)

Source: Authors' Computation

Logistics Regression Equation

(For Govt Parity =0 & Tier 1 City)

P(Bankruptcy=1) = 4.4327 -5.5467*(PAT.TA) -0.0174*(EBIT.IE) +0.0207*(D.E) +2.9781*(D.A) -0.0739*(Age)

(For Govt Parity =0 & Tier 2 City)

P(Bankruptcy=1) = 4.4327 -0.7416 -5.5467*(PAT.TA) -0.0174*(EBIT.IE) +0.0207*(D.E) +2.9781*(D.A) -0.0739*(Age)

```
Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                              1.017e+00  2.601e-01   3.911 9.19e-05 ***
Age_year:factor(Asset_class)Asset Heavy -3.568e-02  1.791e-02  -1.993 0.046296 *
Age_year:factor(Asset_class)Asset Light -4.810e-02  1.768e-02  -2.720 0.006521 **
Age_year:factor(Asset_class)Asset Medium -5.189e-02 1.355e-02  -3.829 0.000129 ***
factor(Asset_class)Asset Heavy:I(Age_year^2) -2.662e-05 3.031e-04 -0.088 0.930012
factor(Asset_class)Asset Light:I(Age_year^2)  3.271e-04 2.592e-04  1.262 0.206998
factor(Asset_class)Asset Medium:I(Age_year^2) 3.011e-04 1.387e-04  2.171 0.029938 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 931.54  on 671  degrees of freedom
Residual deviance: 896.68  on 665  degrees of freedom
AIC: 910.68

Number of Fisher Scoring iterations: 5
```

Fig. 7. Age relation with bankruptcy prediction across different asset classes

Source: Authors' Computation

Model Interpretations

The interpretations drawn from the analysis of the results offer valuable insights into the factors influencing bankruptcy predictions. Firstly, an increase in a company's return on assets (PAT.TA) and interest coverage ratio (EBIT/ Interest) is associated with a reduced likelihood of bankruptcy. Conversely, higher debt-to-equity and debt-to-asset ratios increase the probability of bankruptcy. Age is a significant variable, with its coefficient indicating that as a company's age increases, the odds of bankruptcy tend to decrease. To pinpoint the age at which this decrease becomes significant, an age^2 variable was introduced, and the peak age was calculated as 73 years (for asset-light companies). This suggests that when a company reaches this age, the odds of bankruptcy significantly decrease. Moreover, the type of city also matters, as Tier II cities exhibit a lower likelihood of bankruptcy compared to Tier I cities. Lastly, while government parity may seem insignificant, its inclusion in the model highlights the potential impact of government policies on a company's risk of bankruptcy. These insights provide a comprehensive understanding of the variables influencing bankruptcy predictions, aiding stakeholders in making informed decisions.

Model Evaluation

We constructed the confusion matrix to gauge the model accuracy performance on the test dataset. To determine the classification threshold, we assigned a significantly higher cost (10X) to the misclassification of a bankrupt company as non-bankrupt as a credit default case leads to significantly higher loss for a creditor as compared to taking the opportunity cost in losing out a credible customer. It resulted in a threshold value c of 0.53 (Note: the threshold value decides the bankruptcy prediction; here, the test data probability of bankruptcy greater than 0.53 would be categorized as P(Bankruptcy=1)). The model's prediction accuracy turns out to be 80.2%. Moreover, the plotted ROC curve with a high area under the curve of 0.92 suggests a good model performance of efficiently distinguishing the non-bankrupt and bankrupt classes.

Table 14. Confusion Matrix of Logistics Regression (Bankruptcy Model)

Confusion Matrix			
	Prediction\Actual	Non Bankrupt	Bankrupt
	Non Bankrupt	42	15
	Bankrupt	5	39
Accuracy		0.802	
95% CI		(0.7109, 0.8746)	
No information rate		0.5347	
P-Value [Acc >NIR]		1.97E-08	

Kappa	0.6075
Mcnemar's Test P-value	0.04417
Sensitivity	0.8936
Specificity	0.7222
Pos Pred value	0.7368
Neg Pred value	0.8864
Prevalence	0.4653
Detection rate	0.4158
Detection Prevalence	0.5644
Balanced Accuracy	0.8079

Source: Authors' Computation

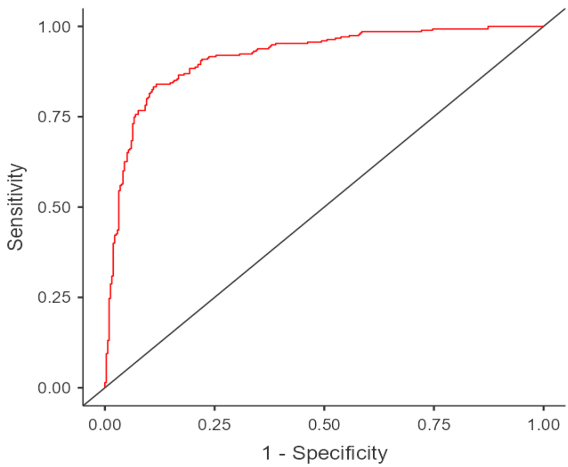


Fig. 8. ROC Curve of Bankruptcy Prediction Logit Model

(Area Under Curve: 0.916)

Source: Authors' Computation

6.2 Random Forest

The application of a random forest model has enriched our understanding of variable importance and their order. This model follows a methodology analogous to the logistic regression, with data partitioned into training and testing sets at an 80:20 ratio. The central aim of incorporating the random forest model was to conduct a comparative assessment of its predictive performance against the logistic regression model.

Table 15. Confusion matrix of Random Forest model

Confusion Matrix			
	Prediction\Actual	Non-Bankrupt	Bankrupt
	Non-Bankrupt	54	3
	Bankrupt	9	35
Accuracy		0.8812	
95% CI		(0.8017, 0.9371)	
No information rate		0.6238	
P-Value [Acc >NIR]		6.98E-09	
Kappa		0.7546	
Mcnemar's Test P-value		0.1489	
Sensitivity		0.8571	
Specificity		0.9211	
Pos Pred value		0.9474	
Neg Pred value		0.7955	
Prevalence		0.6238	
Detection rate		0.5347	
Detection prevalence		0.5644	
Balanced Accuracy		0.8891	

Source: Authors' Computation

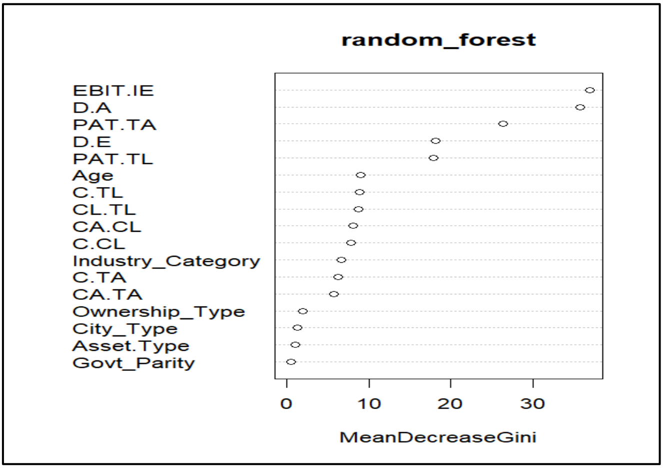


Fig. 9. Mean decrease gini map for variable importance

Source: Authors' Computation

Model Interpretations

In our analysis, we draw several important conclusions from the model's performance. Firstly, the model exhibits an impressive accuracy rate of 88%, indicating its proficiency in effectively categorizing bankruptcy cases. This high level of accuracy underscores its reliability for predicting financial insolvency. Secondly, the model demonstrates a strong level of sensitivity, correctly identifying 85% of actual positive bankruptcy cases. This highlights its ability to recognize firms on the verge of financial distress accurately, signifying its practical utility in real-world scenarios.

Moreover, key financial variables, including the interest coverage ratio, debt to asset ratio, and the ratio of profit after tax to total assets (PAT to TA), which were previously identified as significant contributors in the bankruptcy prediction process, maintain their importance in the random forest model. This reaffirms their role as critical indicators of financial distress and emphasizes their significance in enhancing bankruptcy predictions. Additionally, within the non-financial category, the variable "age" emerges as both an important and statistically significant factor in predicting a firm's likelihood of bankruptcy. These insights underscore the multifaceted nature of the model and the pivotal role played by these variables in improving bankruptcy predictions. The findings highlight the value of these variables in assessing a company's financial health and its potential for bankruptcy.

7 Bankruptcy Type Prediction Model

While our bankruptcy prediction model effectively assesses a company's vulnerability to insolvency based on significant variables, it does not address the specific type of bankruptcy that might occur when a company is predicted as bankrupt. Distinguishing between liquidation and restructuring is of utmost importance since it can greatly impact creditors' credit risk assessment and influence their lending decisions. Considering this, we extended our analysis beyond the bankruptcy prediction model and developed a separate model to predict the type of bankruptcy that a company may be prone to.

To achieve this, we employed the same training and test datasets used for the bankruptcy prediction model. These datasets were meticulously filtered to isolate companies that had undergone bankruptcy. Subsequently, through rigorous assessments for multicollinearity and variable significance, we successfully constructed a logit model that provides valuable insights into the specific type of bankruptcy that companies are likely to face. This comprehensive approach not only enhances our understanding of bankruptcy dynamics but also equips creditors with essential information to make well-informed lending decisions in a dynamic financial landscape.

Table 16. Logistic Regression Results (Bankruptcy Type Prediction Model)

Coefficients:				
	Estimate	St. Error	z value	PR(>|z|)
(Intercept)	2.987E-01	3.453E-01	0.865	0.3869
D.A	2.042E-01	1.271E-01	1.607	0.1080
OC_FC Ratio	8.403E-05	4.496E-05	1.869	0.0616
(Asset.Type) Asset Light	8.190E-01	4.352E-01	1.882	0.0599
(Asset.Type) Asset Medium	4.007E-01	4.025E-01	0.995	0.3195

Null Deviance: 222.15 on 198 degrees of freedom

Residual Deviance: 208.85 on 194 degrees of freedom

AIC: 218.85

Source: Authors' Computation

Table 17. Logit Confusion matrix (Bankruptcy Type model)

Confusion Matrix		
Prediction\Actual	Non-Bankrupt	Bankrupt
Non-Bankrupt	6	5
Bankrupt	11	32
Accuracy	0.7037	
95% CI	(0.5639, 0.8202)	
No information rate	0.6852	
P-Value [Acc >NIR]	4.49E-01	
Kappa	0.2408	
Mcnemar's Test P-value	0.2113	
Sensitivity	0.3529	
Specificity	0.8649	
Pos Pred value	0.5455	
Neg Pred value	0.7442	
Prevalence	0.3148	
Detection rate	0.1111	
Detection prevalence	0.2037	
Balanced Accuracy	0.6089	

Source: Authors' Computation

Model Interpretations

Notably, at a 90% significance level, the ratio of operational creditor (OC) claims to financial creditor (FC) claims emerged as a crucial predictor. An increase in this ratio corresponds to a higher likelihood of liquidation, signifying that a company with a more substantial burden of operational claims relative to financial claims is more prone to liquidation in the event of bankruptcy. Furthermore, our findings underscore the importance of a company's asset structure in bankruptcy outcomes. Companies characterized by lighter assets, as indicated by a low tangible-to-total asset ratio, are more likely to undergo liquidation when compared to their asset-heavy counterparts. This highlights the critical role of asset composition in determining the fate of a company during bankruptcy proceedings.

8 Conclusion

The analysis of our predictive models has yielded valuable insights into bankruptcy prediction. It is evident that a high return on assets and interest coverage ratio are strong indicators of a company's resilience against bankruptcy. These metrics underscore a firm's ability to generate profits and cover interest expenses, making it a more secure prospect for both investors and creditors. Additionally, low debt-to-asset and debt-to-equity ratios offer assurance to creditors considering lending resources to a company. Such lower leverage ratios signify reduced financial risk. On the flip side, a high ratio of operational creditors (OC) to financial creditors (FC) claims often corresponds to an increased likelihood of liquidation, indicating challenges in negotiating with operational creditors. Finally, our analysis reveals that as a firm age, the probability of bankruptcy decreases, emphasizing the significance of experience and longevity in navigating financial challenges. These findings collectively offer practical implications for risk assessment and financial decision-making in the corporate landscape, providing a comprehensive understanding of bankruptcy dynamics and enhancing the ability to make informed decisions.

9 Limitation of the study

The data collected from CMIE is generally reliable; however, the utilization of K-Nearest Neighbours (K-NN) imputation to fill in missing data for specific variables, such as the interest coverage ratio, introduces a level of uncertainty. Imputed values may not perfectly represent the true nature of the missing data, potentially affecting prediction accuracy. The variables chosen were comprehensive and screened for multicollinearity factors using VIF. Furthermore, the study faced limitations in terms of available financial data. While approximately 2,400 firms went through bankruptcy (liquidation and restructuring) between 2017 and 2022, financial data was only available for 400 of these firms, which could restrict the sample size and generalizability of the bankruptcy prediction model. However, it's important to underscore that the dataset used in this study was meticulously curated and suitable for developing a bankruptcy prediction model. The study's findings offer valuable insights into bankruptcy prediction within the Indian business context.

10 Future Scope of Work

Sector-Specific Bankruptcy Prediction Model: A promising avenue for future research involves the expansion of this study to create bankruptcy prediction models tailored to specific industry sectors. Different industries exhibit distinct risk factors and financial structures that can exert a profound influence on the probability of bankruptcy. Developing sector-specific models would allow for a more precise and nuanced assessment of bankruptcy risk, enhancing the overall predictive accuracy and offering valuable sector-specific insights.

Real-Time Financial Health Monitoring: The integration of real-time financial health monitoring mechanisms is a prospective area of exploration. Such monitoring could prove instrumental in aiding the decision-making processes of both company boards of directors and creditors. By providing continuous updates on a company's financial status, stakeholders can make more informed and timely decisions, potentially averting financial crises or bankruptcy.

Incorporating Non-Financial Data: To further enrich the understanding of bankruptcy dynamics, the incorporation of non-financial data sources holds great promise. Beyond financial metrics, sources such as market sentiment towards the company, customer reviews, and assessments of management quality can offer valuable insights. Integrating these non-financial indicators into bankruptcy prediction models can contribute to a more comprehensive and accurate assessment of a company's bankruptcy risk.

11 References

1. *Gupta, V. (2022). Bankruptcy prediction using machine learning techniques: Evidence on Indian companies under insolvency and bankruptcy code: BANKRPUTCY PREDICTION OF INDIAN COMPANIES UNDER IBC. The Journal of Prediction Markets, 16(2), 77-100. doi:https://doi.org/10.5750/jpm.v16i2.1947*
2. *Kaur, S. Comparative analysis of bankruptcy prediction models: An Indian perspective. CABELL'S DIRECTORY, USA, 19.*
3. *Abdullah, M. (2021), "The implication of machine learning for financial solvency prediction: an empirical analysis on public listed companies of Bangladesh", Journal of Asian Business and Economic Studies, Vol. 28 No. 4, pp. 303-320. https://doi.org/10.1108/JABES-11-2020-0128*
4. *Altman, E., Iwanicz-Drozdowska, M., Laitinen, E. and Suvas, A. (2017), "Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-score model", Journal of International Financial Management and Accounting, Vol. 28 No. 2, pp. 131-171.*
5. *Sandeepa (2018). "Comparative analysis of bankruptcy prediction models: An Indian perspective." Cabbell's Direction, USA, 19.*
6. *Sehgal, S., Mishra, R.K., Deisting, F. and Vashisht, R. (2021), "On the determinants and prediction of corporate financial distress in India", Managerial Finance, Vol. 47 No. 10, pp. 1428-1447. https://doi.org/10.1108/MF-06-2020-0332*
7. *Kukreja, G., Gupta, S.M., Sarea, A.M. and Kumaraswamy, S. (2020), "Beneish M-score and Altman Z-score as a catalyst for corporate fraud detection", Journal of Investment Compliance, Vol. 21 No. 4, pp. 231-241. https://doi.org/10.1108/JOIC-09-2020-0022*
8. *Goh, E., Mat Roni, S. and Bannigidadmath, D. (2022), "Thomas Cook(ed): using Altman's z-score analysis to examine predictors of financial bankruptcy in tourism and hospitality businesses", Asia Pacific Journal of Marketing and Logistics, Vol. 34 No. 3, pp. 475-487. https://doi.org/10.1108/APJML-02-2021-0126*