

Analysing the Evolution of Machine Learning Techniques in Forecasting Financial Distress and Their Implications for Sustainability: A Comprehensive Review and Content Analysis

Soumya Ranjan Sethi¹, Dushyant Ashok Mahadik²

¹School of Management, Research Fellow, National Institute of Technology Rourkela, India, Mail: 521sm2009@nitrkl.ac.in

²School of Management, Faculty of Finance, National Institute of Technology Rourkela, India, Mail: mahadikd@nitrkl.ac.in

Abstract

The prediction of such conditions as financial distress or corporate failure and bankruptcy has drawn much attention recently because it gives an early signal of economic problems that corporations are likely to encounter. This field of work is essential to companies, creditors, and investors since it provides information on potential risks and, therefore, potential requirements for funding solutions. Machine learning (ML) models using large databases and better computational techniques introduced a relatively new area of financial distress prediction. To fill this gap and because the ML-based financial distress prediction literature is fragmented, this study performs a systematic literature review to analyze the existing literature. Specifically, by conducting a content analysis of the credit risk study, this paper establishes themes that show how data preprocessing enhances predictive performance. It further discusses top ML models, evaluating their applicability and discussing their advantages and disadvantages. It discusses the impact of the model on sustainability. The paper also deeply analyses the relationship between financial distress prediction and sustainability to provide worthwhile advice to researchers, practical workers, and investors.

Keywords: Machine learning, Content analysis, Data Preprocessing, financial distress prediction, Sustainability



Analysing the Evolution of Machine Learning Techniques in Forecasting Financial Distress and Their Implications for Sustainability: A Comprehensive Review and Content Analysis

PRESENTED BY

Mr. Soumya Ranjan Sethi

Research Scholar, School of Management, National Institute of Technology Rourkela, India

Email: 521sm2009@nitrkl.ac.in

Dr. Dushyant Ashok Mahadik

Assistant Professor, School of Management, National Institute of Technology Rourkela, India

Email: mahadikd@nitrkl.ac.in

Introduction

- Due to the **growing complexity and unpredictability of contemporary markets** as evidenced by the financial crisis of the past ten years, the field of financial distress prediction (FDP) research is receiving more attention. Financial distress prediction is also known as corporate failure prediction or bankruptcy prediction, which plays a crucial role in different fields of the decision-making process, including accounting, finance, business, and engineering (Sethi et al., 2024).
- **Financial distress prediction** refers to identifying early warning signs that a company, organisation, or individual is at risk of experiencing severe financial difficulties. It is the state where a business experiences a specific financial problem. **Such financial challenges include the inability to pay bills and preferred dividends and the ensuing consequences** (Beaver, 1966; Edward B. Deakin, 1972; Altman, 1968). These **difficulties could lead to outcomes such as insolvency, bankruptcy, or default on financial obligations.**



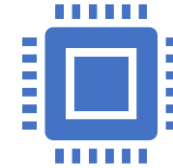
MOTIVATION



Need for Early Warnings:
Predicting financial distress enables companies, creditors, and investors to identify early warning signals, mitigating risks and safeguarding economic stability.



Integration with Sustainability
Goals: Understanding how financial distress prediction aligns with sustainability and SDG targets helps promote resilient and sustainable corporate practices.



Advancing ML Techniques:
The growing prominence of machine learning in financial distress prediction necessitates a comprehensive review of its trends, strengths, and limitations to enhance model robustness.



Bridging Literature Gaps: Addressing fragmentation in the existing literature on machine learning applications in financial distress prediction provides valuable insights for stakeholders and advances the field.

DATA & METHODOLOGY

- This study is a Review-based study.
- **Period:** 1997 to January 2024
- **Database:** Scopus
- **Final Sample:** 371 Documents
- **Software for analysis:** Excel, VOSviewer, Biblioshiny, Wordstat.
- **Search criteria- Keywords:** "Machine Learning" OR "ML" AND "Financial Distress" OR "Default" OR "Bankruptcy" AND "Prediction" OR "Forecasting" OR "Predicting".

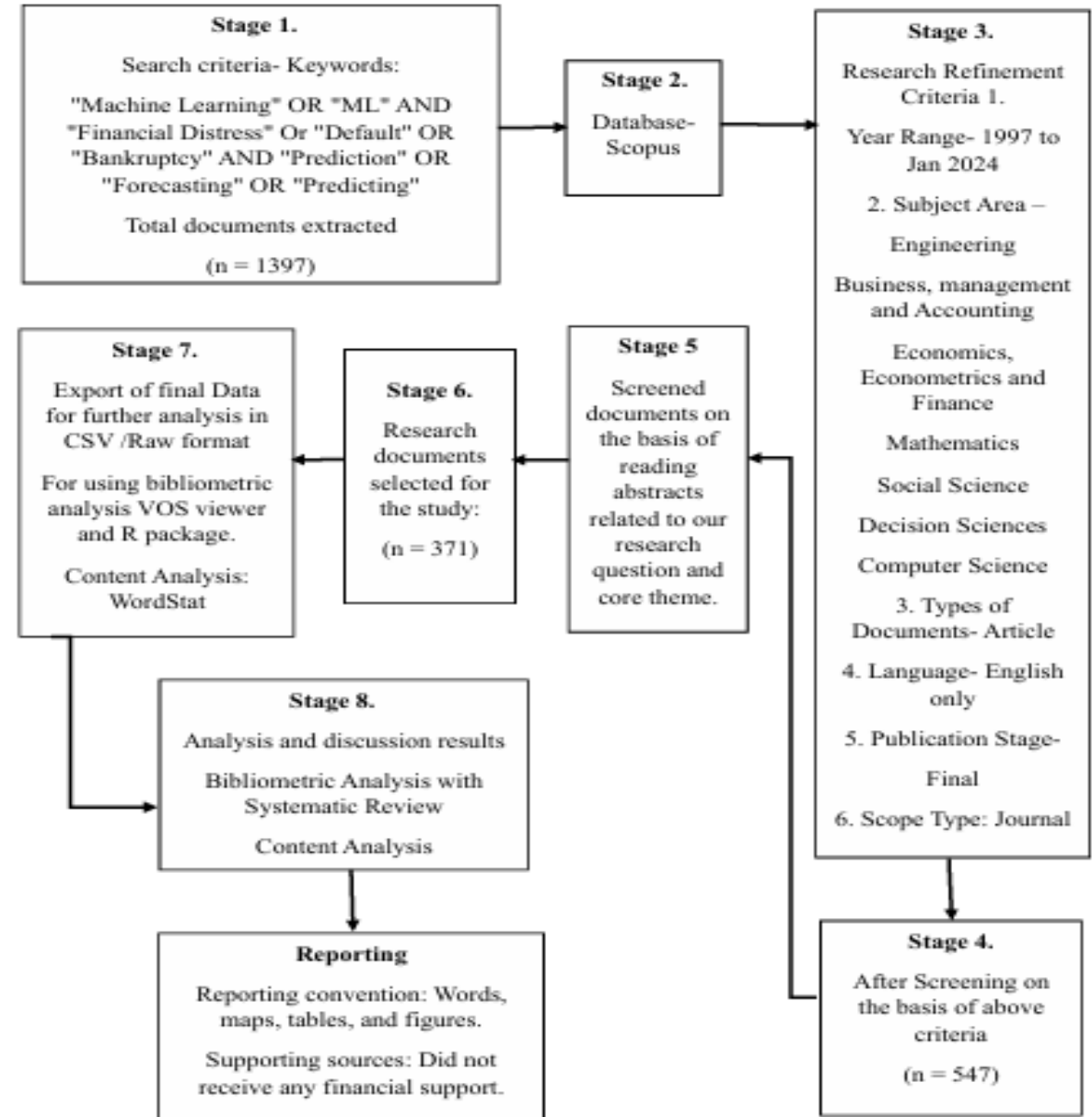


Figure 1. Methodological flowchart (PRISMA) framework

Research questions



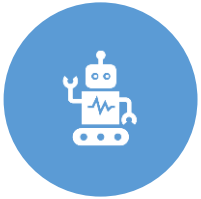
The following six research questions in this review paper:



RQ1: Identify the major contributions to the FDP's examination of machine learning technologies, i.e., the most influential authors, organisations, countries, and papers in this field of study.



RQ2: Which research themes are the most significant in this domain?



RQ3: What are the data pre-processing techniques used in machine learning technology?



RQ4: What are the different FDP models and their advantages and disadvantages?

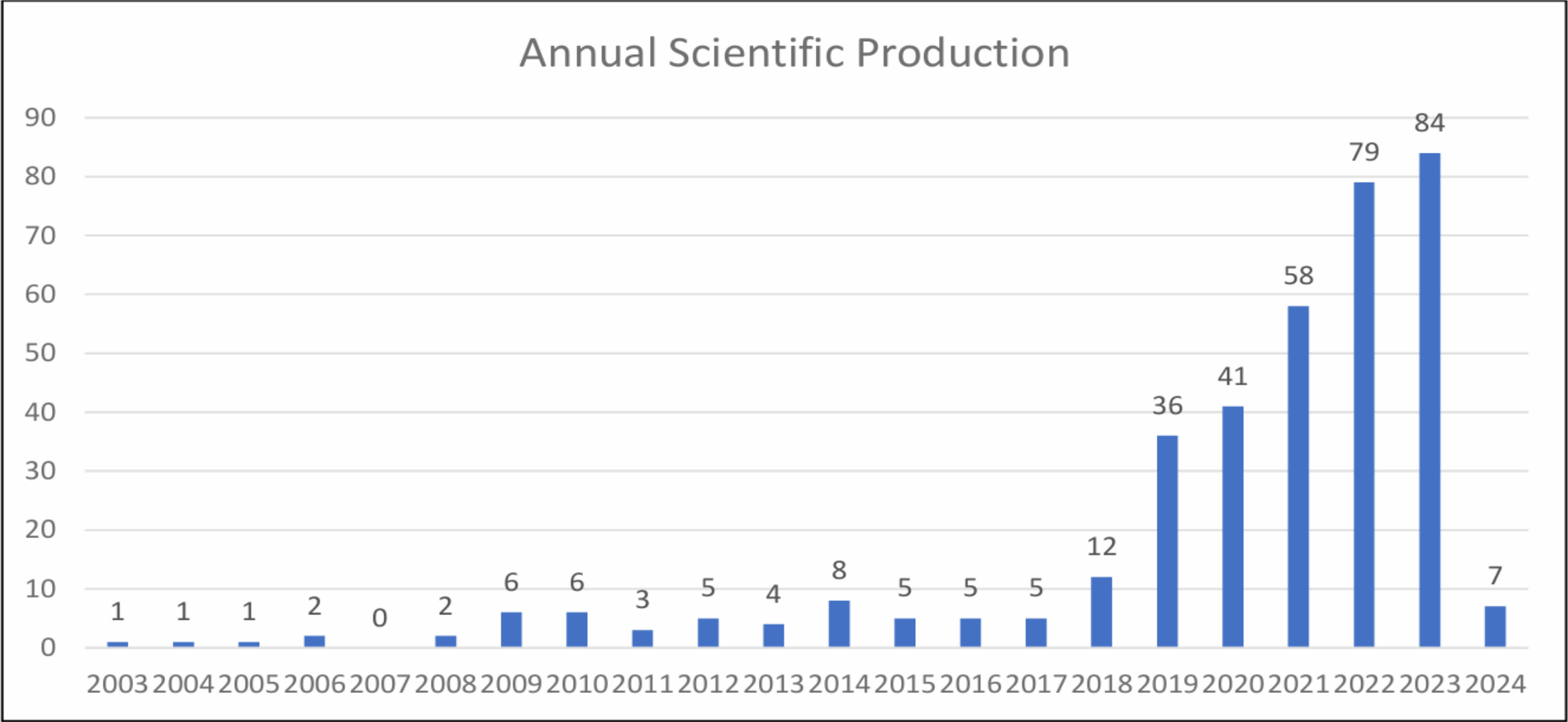


RQ5: How does this study impact sustainability and its contribution towards SDGs?

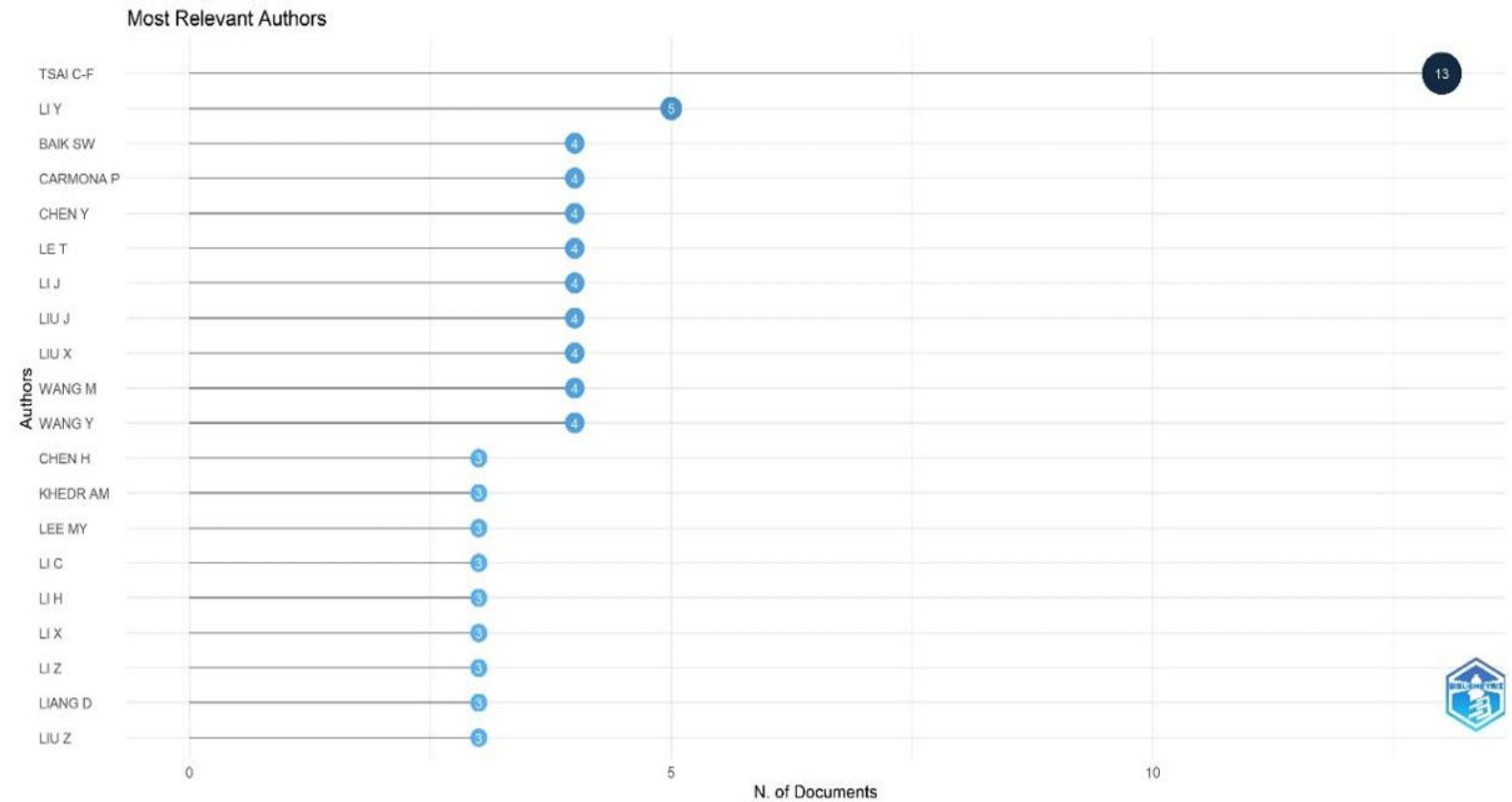


RQ6: What are the gaps and areas for future research direction?

Analysis & Interpretation



Most Relevant Authors



Top 10 most frequent words on ML in FDP

Term	Frequency
Machine learning	180
Forecasting	152
Learning systems	75
Finance	57
Bankruptcy prediction	53
Decision trees	51
Risk assessment	48
Learning algorithms	46
Classification (of information)	37
Machine learning models	33

Source: Compiled by Authors

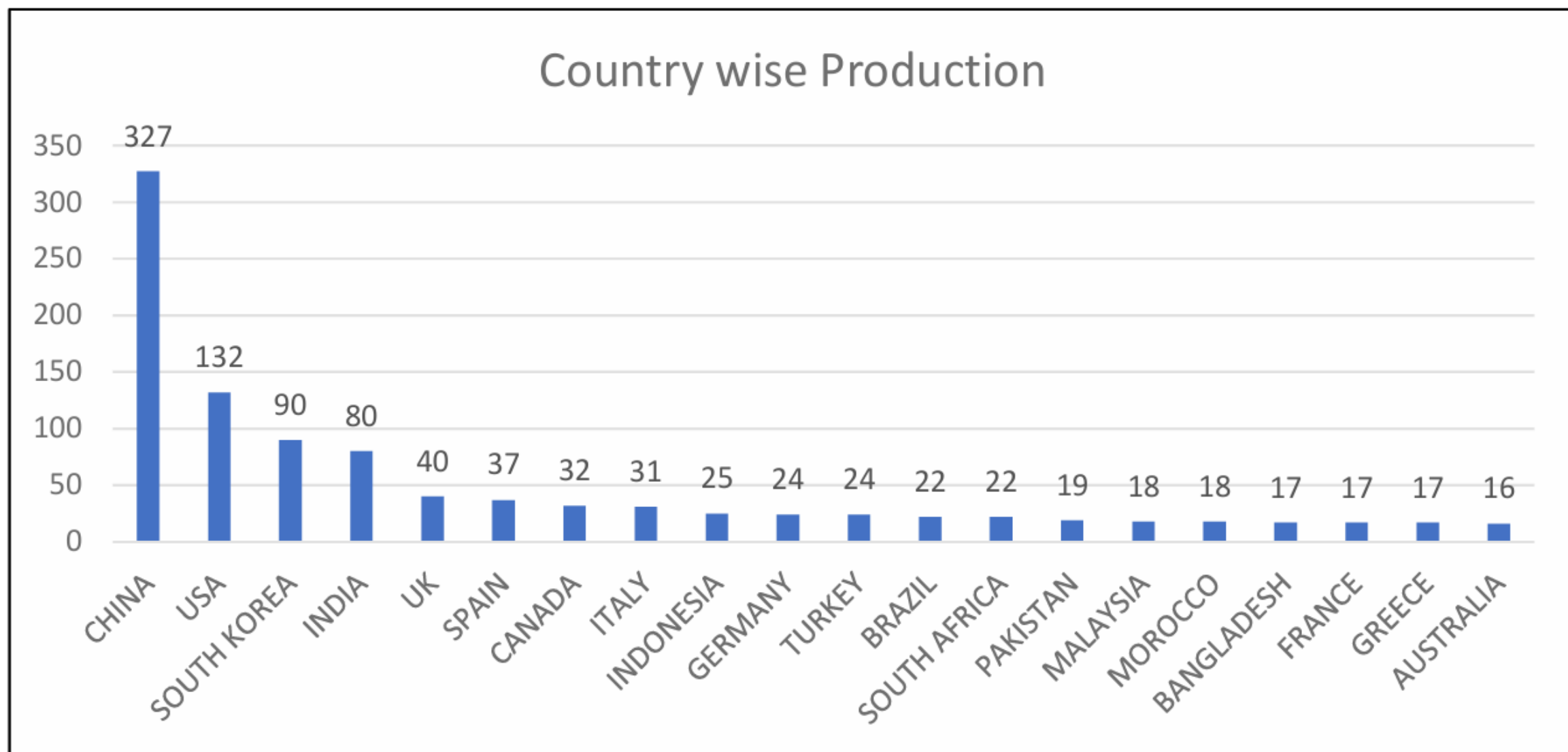
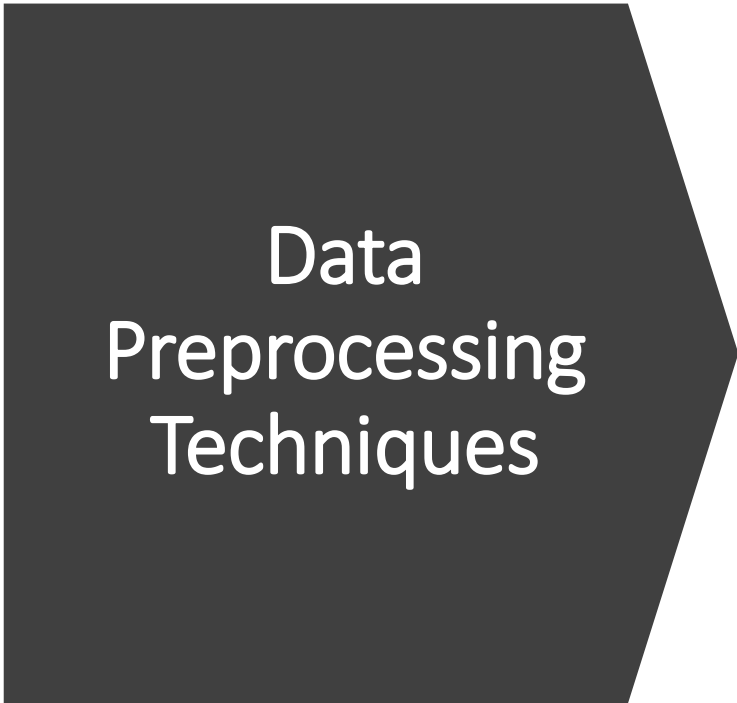


Figure 3. Country-wise production using biblioshiny

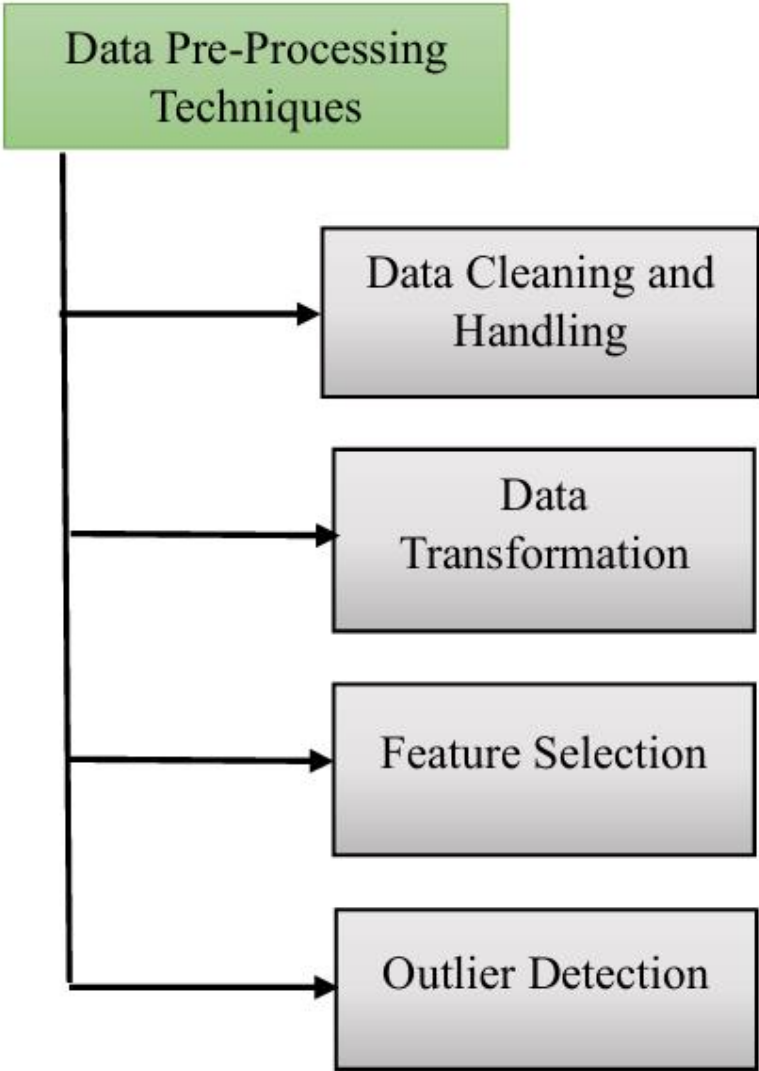
Content Analysis (Source: Wordstat)

SL. No	Topic	Keywords	Coherence (NPMI)	Freq	Cases	% Cases
1	Machine Learning	Machine Learning; Advance Machine Learning; Supervised Machine Learning; Classification and Regression; Artificial Intelligence; Support Vector Machine;	0.529	399	57	96.61%
2	Feature Selection Genetic Algorithm	Feature Selection; Feature Importance; Variable Selection; Algorithm; Prediction; Classification; Bankruptcy	0.480	97	43	72.88%
3	Credit Risk	Risk; Credit; Lending Rate; Default; Credit Risk Evaluation; Risk Management; Credit Losses; Interest Rate	0.472	61	18	30.51%

Data Preprocessing Techniques



Data Pre-Processing Techniques



```
graph TD; A[Data Pre-Processing Techniques] --> B[Data Cleaning and Handling]; A --> C[Data Transformation]; A --> D[Feature Selection]; A --> E[Outlier Detection];
```

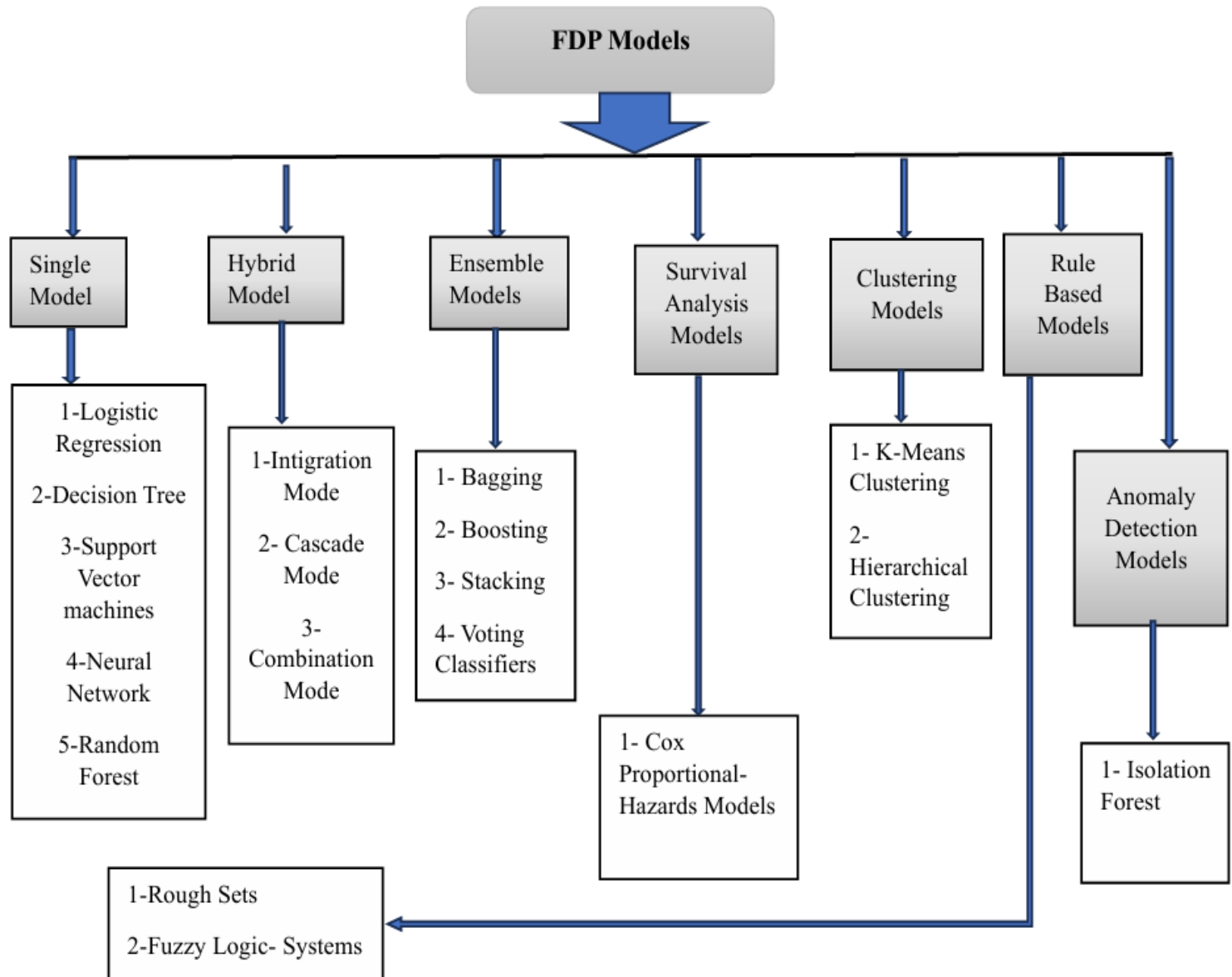
Data Cleaning and
Handling

Data
Transformation

Feature Selection

Outlier Detection

Structure of FDP Models



Benefits and drawbacks of different methods.

Methods	Benefits	Drawbacks
Logistic regression	▪ Simple to use ▪ Working well with less dataset ▪ No presumptions on the distribution of the data	▪ It creates limits that are linear. ▪ Not appropriate for high dimensional data ▪ Unable to learn features
Support vector machine	▪ Operational in High-Resolution Environments ▪ Robust to overfitting ▪ worldwide optimization ▪ Working with small datasets	▪ Reliability of Kernel Selection ▪ Not appropriate for large dataset ▪ Struggle with imbalance dataset
Neural network	▪ The system comprehends intricate non-linear relationships between input and output. ▪ Ability to parallel processing ▪ It can handle noisy data ▪ It can adapt different types data	▪ It's a complex model with numerous parameters ▪ Required large data for training ▪ Internal works are not easily interpretable by individuals
Decision trees	▪ Simple and easily understandable ▪ beneficial for data exploration ▪ able to identify non-linear connections	▪ Require more time to run the model ▪ Not appropriate for missing data in a data set ▪ There are too many rules produced by models.

Continue..

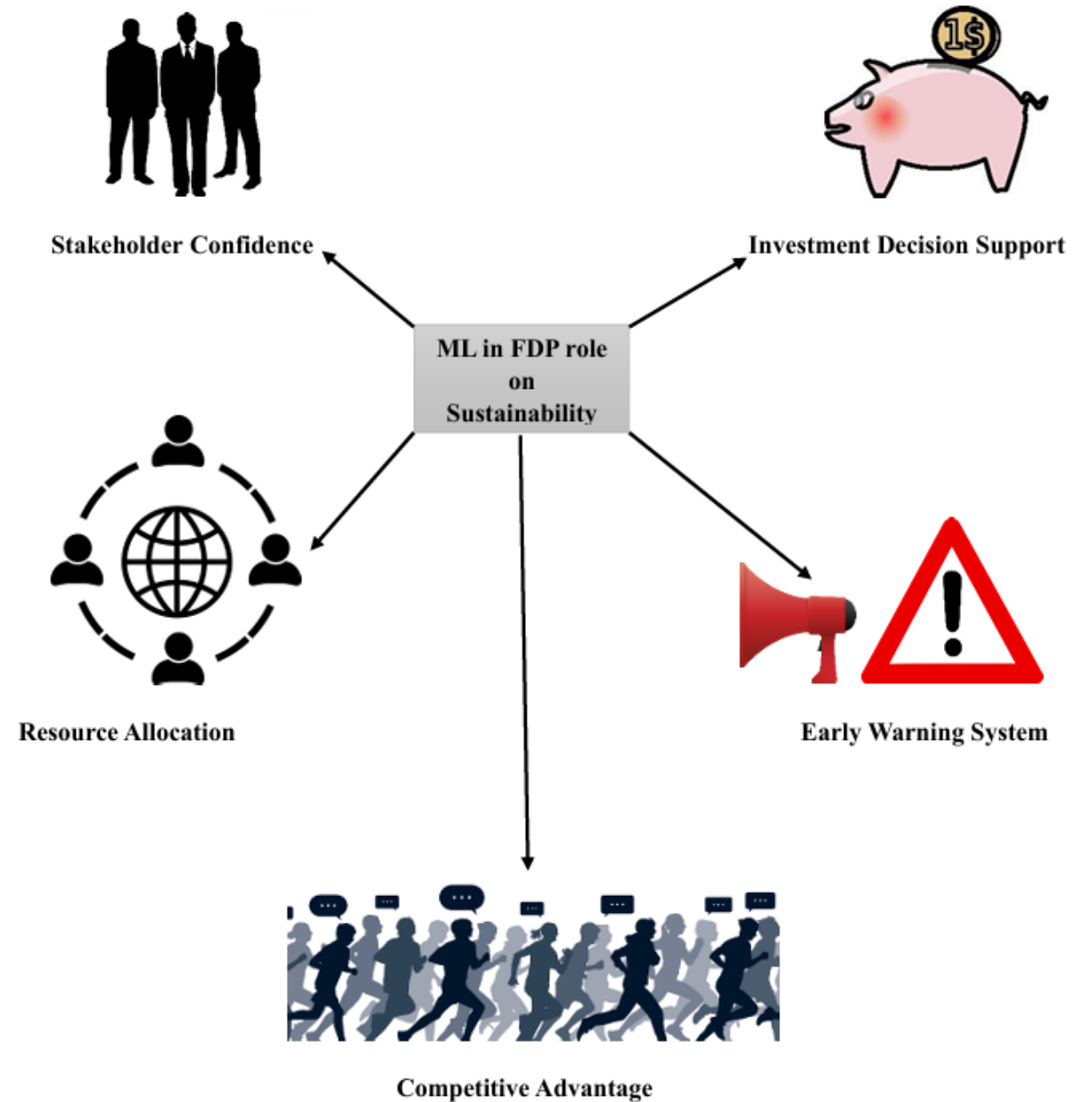
Random forest	▪ Among the most reliable ML algorithms ▪ High accuracy model ▪ Handles non- linearity ▪ It can handle numerical and categorical features ▪ It prone to overfitting	▪ Not easily interpret ▪ Black box model ▪ Not appropriate for imbalanced data ▪ Require feature engineering
Bagging	▪ Reduces the variance of the model ▪ It improves model accuracy ▪ Strong Against Outliers ▪ Effective management of a complicated dataset	▪ Increased Complexity of Computation ▪ Requirements of more memory ▪ Restricted gains in performance for proficient learner
Boosting	▪ Provide higher accuracy ▪ Good robust ML algorithm ▪ Versatility ▪ Reduce overfitting	▪ Black box nature ▪ Disregard overfitting ▪ It can introduce bias ▪ Models are More complex in nature
Cox Proportional-Hazards Models	▪ Its flexible ▪ It believes that throughout time, the hazard functions of various groups will be proportionate ▪ uses partial likelihood estimation ▪ Handle censoring effectively	▪ It may not be appropriate for forecasting outside of the range of available data or for calculating absolute risk ▪ Sensitive to outliers ▪ Limited Results

Continue..

Rough Sets	▪ Easy to interpretable ▪ Handle uncertainty data ▪ It does not make assumptions ▪ It can use both numerical and categorical data	▪ Scalability issues ▪ Depends on quality of input data ▪ Its predicted accuracy is not always as high as that of other machine learning models
Fuzzy Logic- Systems	▪ It can handle uncertainty data ▪ Flexible in nature ▪ Noise tolerance ▪ Easy to implement	▪ Complex in computational point of view ▪ susceptible to overfitting ▪ Harder to analyse
Isolation forest	▪ Efficient in nature ▪ Handle larger datasets ▪ Robust to outliers and anomalies ▪ Easy to implement	▪ not function well on heavily skewed datasets ▪ The method may not be optimal for datasets with complex relationships or multiple anomalies groups.

Role of this study on Sustainability

- Early Warning System
- Resource Allocation
- Investment Decision Support
- Stakeholder Confidence
- Competitive Advantage



Towards SDGs

8 DECENT WORK AND
ECONOMIC GROWTH



9 INDUSTRY, INNOVATION
AND INFRASTRUCTURE



11 SUSTAINABLE CITIES
AND COMMUNITIES



12 RESPONSIBLE
CONSUMPTION
AND PRODUCTION



17 PARTNERSHIPS
FOR THE GOALS



CONCLUSION

1. This study finds the most productive authors, organisations, and nations in the digital inclusion literature by utilising a variety of trends, bibliometrics, evolution and systematic reviews.
2. This study suggested three main area research themes in the following areas:
 - (1) Credit risk
 - (2) Feature selection and genetic algorithms
 - (3) machine learning.
3. We observed that the research trend is growing, indicating this field's importance in this current era.
4. Additionally, we found that the study has a significant contribution towards sustainability.
5. Better data preprocessing is better for model accuracy.

Policy and Implications

1.Proactive Risk Management: Policymakers should mandate the adoption of ML-based financial distress prediction tools to enhance early warning systems and prevent bankruptcy, ensuring economic stability and sustainability

2.Support for Sustainable Practices: Governments and institutions must incentivize businesses to use predictive insights for resource optimization, reducing inefficiencies, and aligning financial decisions with environmental and social sustainability goals.

3.Global Collaboration: International cooperation is essential for developing robust, adaptable ML frameworks that address financial distress and foster economic resilience in diverse economic contexts.

4.Capacity Building: Invest in training programs and infrastructure to enable organizations and policymakers to effectively implement advanced ML techniques for sustainable financial management.

Limitations and future research direction

We acknowledge the limitations that still exist in our study.

Here are the study's limitations and future research direction summarised in Some points:

1. We used only the Scopus database. Different databases could yield completely different results is one of the research's drawbacks.
2. while our study aligns with the objectives and values of the bibliometric analysis, we encourage future researchers to employ scientometric analysis, meta-analysis, and topographic view to provide a comprehensive overview of the digital inclusion literature.
3. Future researchers may come up with more innovative models to add to the existing body of knowledge.
4. Future researchers need to consider the possibility of using sustainability metrics as early warning signs of financial trouble.
5. To enhance the accuracy of predictions, they might try out several sophisticated ensemble approaches, including stacking and blending.
6. Researchers may include other data sources, such as social media, satellite images, and text analysis, to improve the model's prediction ability. However, they may also do cross-country studies to identify global patterns and differences in financial distress and sustainability.

References

- Sethi, S. R., & Mahadik, D. A. (2024). Spotting Trouble Before It Starts: Has Financial Distress Prediction Evolved During 1985–2022. *Applied Econometrics and International Development*, 24(1), 181-206.
- W.H. Beaver, Financial ratios and predictors of failure, *Journal of Accounting Research: Empirical Research in Accounting—Selected Studies* 4 (1966) 71–111
- Altman, E.I., “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *Journal of Finance* 23 (1968), pp. 589–609
- E.B. Deakin, A discriminate analysis of predictors of business failure, *Journal of Accounting Research Spring* (1972) 167–179.
- Yap, B. C. F., Yong, D. G. F., & Poon, W. C. (2010). How well do financial ratios and multiple discriminant analysis predict company failures in Malaysia? *International Research Journal of Finance and Economics*, 54(13), 166–175.
- Altman, E. I., Iwanicz-Drozowska, M., Laitinen, E. K., & 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 & accounting*, 28(2), 131-171.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business*, 74(1), 101-124.
- Ceylan, I. E. (2021). The impact of firm-specific and macroeconomic factors on financial distress risk: A case study from Turkey. *Universal Journal of Accounting and Finance*, 9(3), 506-517.
- Sairin, N. H., Salisi, M. S., & Bujang, I. (2019). Determining Macroeconomic Factor of Financial Distress in Malaysia. *Malaysian Journal of Business and Economics (MJBE)*, (2).

