ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

The Development of a Corporate Financial Data Analysis System in the Context of the Digital Economy's Advancement

Xianrong Li[1,*]

[1] Smart Finance and Business School, Hefei College of Finance & Economics, Hefei, Anhui, 230000, China

SUMMARY: *The digital economy has evolved rapidly, and some companies have begun to use the digital intelligence approach to self-improvement. The company's financial data analysis system was designed to help companies make smart decisions in financial management by collecting and viewing financial data. The system combines budget management, cost control and financial risk management and selects the corresponding financial index. The study later added Benford Law to assess the quality of financial data and developed a Benford-RF prediction model, which used to do financial analysis, where random log is an integrated learning model. Therefore, it is more effective to predict the financial situation of the company. In addition, in the real case, the accuracy of the use of this model for forecasting the financial position of M is 92.5%.*

KEYWORDS: *Benford's Law; Random Forest; Predictive Modeling; Financial Analysis; Digital Economy*

1 Introduction

In recent years, the expansion of the business scale and management complexity of enterprises have made the management attach great importance to the digital transformation of the financial function, actively promote the construction of data governance, system integration and intelligent analysis capabilities, and gradually explore the digital capacity enhancement path with financial analysis as the core, trying to realize the strategic shift from traditional financial accounting to intelligent decision support [1-4]. In the process of exploration, there are still many challenges in enterprise financial analysis [5-8]: First, with the continuous construction of the enterprise in informationization and financial sharing system, it has formed a certain foundation for financial data collection and analysis, but the management still perceives insufficient support for financial analysis and insufficient exploitation of the value of the data in the actual operation. Secondly, due to the construction of various business systems at different stages and the lack of unified planning, the existing SAP ECC financial system, Oracle ERP, Kingdee CRM, self-developed MES and other platforms, there are large differences in data structure, interface protocols and coding rules. Third, the financial analysis framework composed of static financial statements and traditional ratio models, such as gearing ratio, net profit margin, current ratio and other indicators, lacks the ability to effectively portray the business logic, changes in cost structure and fluctuations in the cash cycle in the changing market environment. Fourth, the enterprise has deployed several sets of business and financial systems, but the financial analysis process

*xianrong_li2025@163.com
https://doi.org/10.65102/is2026357
Published: 30 April, 2026

is still centered on the statements after the monthly closing of accounts, and key indicators such as overdue accounts receivable, procurement payment progress, and inventory capital consumption are mostly dependent on manual collation and manual calculations, which take more than 7 days on average, and seriously lag behind the pace of business feedback.

Information technologies such as large-scale data analysis, mobile networks and machine intelligence have developed, which greatly contributed to the digital economy. Businesses can use these information platforms to quickly collect complex financial data, not to be no wards behind market changes and adjust their strategies. In the digital economy, new market demand can improve the overall industrial structure and open new areas, new models, new businesses, make the industry more professional, and the value chain extends to a high class.

The intelligent transformation of finance can make decisions with scientific and feasible for the business strategy of the enterprise through the collection and accurate analysis of data [17-18]. From the viewpoint of work responsibility, financial accounting that concentrates on after-the-fact accounting puts its sight on the accuracy of financial data [19]. Should financial accountants lack proficiency in data handling, it could exert an adverse influence on the growth of the business. For executives, the application of smart finance can remove simple and repetitive mechanical work, reduce the waste of financial personnel's manpower, facilitate financial personnel to devote their energy to innovative management, and improve personnel initiative [20, 21]. Through the organic integration of the technological merits of big-data technology and the enterprise's financial management, the enterprise can achieve high-quality and efficient handling of financial data and promote all its operations in an orderly manner. This situation is highly beneficial to the enterprise's sustainable development. In fact, this also represents the crucial significance of the digital economy's development [22-24].

Driven by the background of digital economy, this paper proposes an enterprise financial data analysis system consisting of basic data layer, data warehouse layer, application logic layer and result representation layer, and discusses its application in budget management, dissemination control and risk early warning, and selects financial indicators based on these three layers. Subsequently, Benford's law is introduced on the basis of the random forest model, Benford's law is employed to carry out a data quality assessment of financial indicators. Subsequently, the Benford's factor is taken as a variable to build a random forest prediction model along with corporate financial variable indicators. In this paper, 200 listed manufacturing enterprises are selected as research samples, and a set of financial comprehensive level index system that can be used by listed enterprises is formulated, and the company's financial score is calculated through factor analysis as a basis for classification. Then the SMOTE algorithm was utilized to deal with the problem of sample imbalance. Finally, the model prediction accuracy is derived through the classification performance assessment index operation, and the accuracy and applicability of the model is verified with cases.

2 Enterprise financial data analysis system construction

The digital economy belongs to a kind of data that can rely on the Internet to realize big data content identification, positioning, and screening, and obtain available data for application and storage. Motivated by the digital economy, an enterprise financial data analysis system is established, and an analysis of its primary application scenarios is conducted.

2.1 Overall structural design

The overall structure of the enterprise financial data analysis system exists at four levels, which makes the system more diversified and has a certain degree of scalability, so that when it is necessary to modify the functions in a certain layer, it can be adjusted independently according to the actual needs. Figure 1 depicts the general architecture of the corporate financial data analysis system.

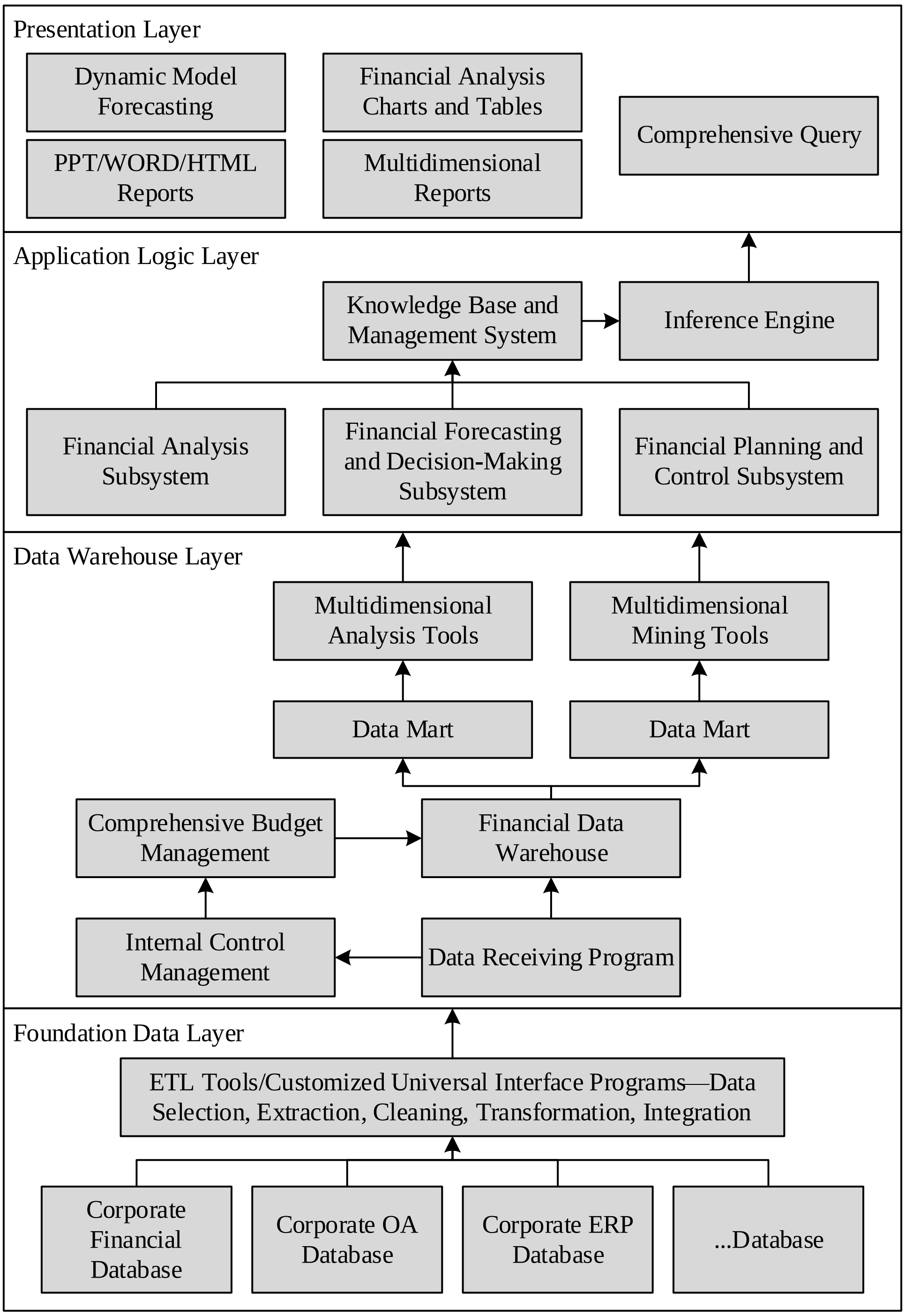

Figure 1: The comprehensive framework of the corporate financial data analysis system

(1) Basic Data Layer

This layer mainly provides basic data support, when the enterprise accumulates more financial data, effective analysis of these financial data can improve the rationality of decision-making. In this layer, a dedicated data collection structure is designed to obtain relevant data from different databases through ETL data collection tools, and integrate these data to realize effective data mining and complete data storage in a unified format.

(2) Data Warehouse Layer

This layer is located on top of the basic data layer, mainly realizes data analysis and storage, and at the same time creates the data warehouse model, adopts the star data model to complete the data storage, so it can effectively help the system to improve the decision-making ability. Once the data analysis and storage model has been developed, enterprise employees can retrieve the necessary data from the data repository based on their individual requirements. Subsequently, they can build a multi-tiered data mart in line with the data analysis models related to various topics.

(3) Application Logic Layer

Application logic layer belongs to the core part of the system, this layer uses multi-dimensional reports, multi-dimensional analysis and other tools to generate statistical reports, and respectively, different subsystems to achieve the analysis of financial data, the completion of the analysis, the use of expert knowledge base and reasoning machine to achieve more accurate analysis and decision-making.

(4) Results representation layer

This layer generates the analysis result page of the application logic layer and shares it with the users, who can browse the analysis and decision-making results through the WEB webpage, and can also download the reports in Word, Excel and other formats.

2.2 Financial analysis application scenarios

2.2.1 Budget management

(1) Accurate budgeting

The financial data analysis system aggregates past financial data and current business data to help the budget budget. The past data includes various income, expenses, expenses, etc. Thanks to the analysis of changes in time and regular modeling, cyclical changes and hidden rules in entrepreneurial activity can be found. At the same time, real-time business data, such as sales orders, inventory, production planning and market changes, makes the budget more promising and faster adapting to change.

(2) Budget execution monitoring and adjustment

The system for analyzing financial data is capable of conducting real-time monitoring of budget execution. It also undertakes a comparative analysis between the actual business figures and the budgeted data. By setting the warning threshold of key indicators, when the actual data deviates from the budget by a certain percentage, an early warning signal will be issued in time. At the same time, the financial data analysis system is able to analyze in depth the reasons for the deviation of budget execution, whether it is the business volume exceeding expectations, poor cost control or unreasonable budgeting. According to the analysis results, it provides adjustment suggestions for enterprises to ensure the realization of enterprise budget objectives.

2.2.2 Cost control

(1) Cost analysis and traceability

Financial data analysis systems can help companies scrutinize the cost structure and disassemble the total cost into several parts, such as commodity costs, labor costs, production costs, etc. D. To find out how much each cost is taken into account in the total value. Considering the historical data on the cost, you can find the law and the main factors of change in value. In addition, data tracking technology can also be used to catch up where the cost comes from.

(2) Cost optimization strategy development

After completing the book with the financial data analysis system, the company can develop a method of reducing costs. If it is found that certain raw materials are expensive, you can talk to the supplier about the price, find a new supplier, or change the process of purchasing to save money on the purchase. A company can also save labor costs by training employees by optimizing production processes or buying automation equipment.

2.2.3 Early warning of risks

Businesses can use a financial data analysis system to calculate data and update the information in real time so they can look at changes in financial risks. The system will quickly place business data in real-time in the risk index model to see how the company is now financial. If the corresponding indicators jumped abnormally and the preset value is also achieved, the system immediately gave a risk alarm, pushed to the management personnel, allowed them to check the source of the risk and quickly find a way to cope with it.

3 Benford-RF financial data analysis forecasting model

Financial data analysis is used in budget management, cost management and risk warning. But when conducting a financial assessment, companies need to consider many factors that are easily overlooked and that may not allow the final analysis to be unthinking of Woods. To solve this problem, we chose a high-quality model of a random forest (RF).

3.1 Benford's Law

The Benford method is a method of theoretical evaluation of distribution, commonly used to assess the quality of data.

Benford distribution law refers to the first digit of data in a large number of natural data sets obeys a specific distribution.Suppose that the initial digit of a natural data set is denoted as $N\left(N=1,2\cdots 9\right)$. Then, the likelihood that the initial digit of the data set $N=n$ is as follows:

$$P(N=n)=\log_{10}\left(1+\frac{1}{n}\right) \tag{1}$$

The specific probability of the first digit (1.2...9) in the natural dataset can be calculated from equation (1).

The Benford Act states that the first number of data is more likely to be consistent with this law when the sample is large enough and the amount of data increases. With this function, if the data is problematic or inconsistency, it will significantly deflect from the Benford Act. Since the higher the quality of the data and the more complies with the Benford law, the large-scale dataset and this law are very different, it will make people doubt the quality of data.

(1) A good rate test in chi-square is a statistical method to see if the actual number of accidents and the number of standards at the beginning of the dataset corresponds. With this method, two assumptions were made: a zero hypothesis and an alternative hypothesis. The zero hypothesis believes that the actual number of accidents at the beginning coincides with

the number of times the standard. The constructed χ^2 statistic is:

$$\chi^2 = N\sum_{i=1}^{9}\left[\frac{(e_i - b_i)^2}{b_i}\right]\left(i = 1, 2\cdots 9\right) \tag{2}$$

where e_i is the actual observed frequency of the first digit (second or third...ninth) i of the tested dataset and b_i is the standardized frequency of the first digit (second or third...ninth) of the tested data under Benford's law. The levels of statistical significance are 10%, 5%, and 1% correspondingly. For the χ^2 test with 8 degrees of freedom, the respective critical values are 13.362, 15.507, and 20.090.

(2) The modified Kolmogorov-Smirnov-of-fit-of-fit has its method. When testing data using K-S, you must first learn the cumulative function of distributing sample data and standard data. Then calculate the cumulative value of the sample data distribution and the cumulative value of the distribution of standard control data. The constructed K-S goodness-of-fit statistic is as follows:

$$M = \max\left|F_e(x) - F_b(x)\right| \tag{3}$$

After Kuiper and Stephens improved the K-S statistic, the modified Kolmogorov-Smirnov statistic was obtained:

$$V_N^* = V_N\left(N^{1/2} + 0.155 + 0.24N^{-\frac{1}{2}}\right) \tag{4}$$

$$V_N = \max(F_e(x) - F_b(x)) + \max(F_b(x) - F_e(x)) \tag{5}$$

The table shows that the critical value at 10% significance level is 1.183. The test to assess the data quality by constructing the modified Kolmogorov-Smirnov statistic is called modified Kolmogorov-Smirnov goodness-of-fit test.

(3) A distance test is employed to ascertain whether the leading digit of the data under examination adheres to the Benford distribution. This is accomplished by computing the disparity between the actual observed frequency of the leading digit in the tested data and the standard frequency of the leading digit as stipulated by Benford's law. This type of test is referred to as a distance test. The distance statistic of the statistic is:

$$M = \max_{i=1,2,\ldots,9}\left\{\left|b_i - e_i\right|\right\} \tag{6}$$

$$D = \sqrt{\left[\sum_{i=1}^{9}(b_i - e_i)^2\right]} \tag{7}$$

(4) The judgment standard of Pearson's correlation coefficient is as follows: when the correlation coefficient falls within the range of 0.99<r≤1, it indicates that the leading digit of the dataset under examination adheres to Benford's law. When the correlation coefficient is 0.97<r≤0.99, it means that there is a possibility of forgery in the detected data, which needs to be given special attention. When the correlation coefficient r≤0.97 indicates that the detected data have a high possibility of forgery, and need to pay more attention.

The four statistical testing approaches mentioned above are among the most frequently employed ones. Among them, the chi-square goodness-of-fit test is particularly well-known. In this research paper, the chi-square goodness-of-fit test will be utilized to evaluate the quality of financial data.

3.2 Constructing the Benford factor

Let us assign the symbol $X_j \left(j=1,2,3,\cdots,k \right)$ to represent the financial indicator variable. Let $r_d^{(j)}$ signify the disparity between the observed occurrence rate of the initial digit d of the indicator X_j and the theoretical occurrence rate stipulated by Benford's law. Subsequently, the formula for $r_d^{(j)}$ is as follows:

$$r_d^{(j)} = f_d^{(j)} - f_{B,d}^{(j)} \quad j=1,2,\cdots,k \tag{8}$$

Then there are:

$$\sum_{d=1}^{9} r_d^{(j)} = 0 \ j=1,2,\cdots,k \tag{9}$$

Let the initial digit that has the greatest absolute value of the frequency difference of the first-digit be denoted as $a^{(j)}$. That is:

$$a^{(j)} = \arg\max_d \left| r_d^{(j)} \right| \ j=1,2,\cdots,k \tag{10}$$

$$B_{i,j} = \begin{cases} 1 & X_{i,j} \text{ first digit is } a^{(j)} \\ 0 & \text{Others} \end{cases} \tag{11}$$

3.3 Benford-RF model

The model used in this study makes full use of the advantages of the random forest method and combines multiple decision tree models. By doing so, we can avoid potential defects and improve the prediction performance of financial data analysis and prediction model.

The Benford-RF model is built in the following sequence:

(1) We derive Benford features from the predictors in $D=\left\{(X_1,Y_1),(X_2,Y_2),\cdots,(X_N,Y_N)\right\}$, where $X_i=\left(X_{i,1},X_{i,2},\cdots,X_{i,k}\right)\left(i=1,2,\cdots,N\right)$. For each variable, we compare its empirical first-digit frequencies with the Benford distribution and compute the maximum deviation (Eq. 11), which defines the Benford factor(s). We then augment each Xi with the Benford terms to form $X_i^B=\left(X_{i,1},X_{i,2},\cdots,X_{i,k},B_{i,1},B_{i,2},\cdots,B_{i,k}\right)\left(i=1,2,\cdots,N\right)$ and obtain the expanded dataset $D^B=\left\{(X_1^B,Y_1),(X_2^B,Y_2),\cdots,(X_N^B,Y_N)\right\}$.

(2) We generate n bootstrap samples D^B from $D^{B(s)}\left(s=1,2,\cdots,n\right)$ (sampling with replacement) and fit one decision tree h_s to each sample. The number of trees is chosen from candidate values (e.g., 50/100/200) using a learning-curve criterion (validation or OOB), increasing n until performance gains plateau.

(3) We aggregate n decision trees to form the random forest; the predicted class is determined by majority vote:

$$H\left(X^{B}\right) = \arg\max \sum\nolimits_{s=1}^{n} I\left(h_{s}\left(X^{B(s)}\right) = Y\right) \tag{12}$$

The Benford's Law serves the purpose of assessing the data quality of financial metrics. It helps in pinpointing the sample-point data that might carry the risk of fraudulent activity. Subsequently, the Benford factor is designated as a key variable representing data quality and incorporated into the Random Forest model. The Random Forest model employs an ensemble learning algorithm along with Bootstrap resampling.

4 Analysis of forecasting models for financial data analysis

4.1 Selection of financial indicators

This paper analyzes corporate financial data based on budget management, cost control and risk early warning, and selects financial indicators to assess the overall financial status of the company, with 24 indicators initially selected:

A1 is earnings per share (EPS). A2 and A3 capture cash returns, measured as cash return on net assets and total assets, respectively. A4 reports the cost-of-goods-sold (COGS) ratio (COGS/sales), and A5 is operating cash flow per share. A6 describes asset structure via the share of current assets in total assets (current assets/total assets). A7 and A8 summarize capital structure and leverage, using the effective leverage (gearing) ratio and the long-term debt share in long-term capital. A9-A11 focus on solvency and liquidity: debt-to-asset (asset-liability) ratio, quick (acid-test) ratio, and cash ratio. A12 is interest coverage (times interest earned), while A13-A15 provide cash- and maturity-related debt indicators, including cash-to-maturing-debt, operating-cash-flow-to-net-debt, and the share of current liabilities in total debt. A16-A18 scale cash receipts from merchandise sales by net assets, total assets, and inventory, respectively. A19 and A20 are efficiency metrics (inventory turnover and working-capital turnover). Finally, A21-A24 are growth indicators, covering year-on-year growth in net income, operating cash flow per share, operating income, and net assets.

4.2 Classification of the financial situation

4.2.1 Sample Selection

This paper focuses on A-share listed manufacturing companies, a total of 200 companies were screened, and the annual report data of the sample companies in 2023 were selected as the research basis of this paper.

4.2.2 Factor analysis process

The samples were linearly transformed [0,1] by the departure standardization method, in order to determine whether the data were normally distributed, Bartlett sphericity test was used, the result of KMO coefficient was 0.671, the correlation was good, and the factor analysis was feasible, and the Bartlet results showed that the significance was close to 0, smaller than 0.05, the correlation matrix was non-unit matrix, the indexes had a relationship with each other, the factor analysis was feasible, and the samples passed the test.

(1) Factor extraction

Extraction of principal components through factor analysis, aimed at reducing the refinement of the data, the total explained variance of the indicator data is shown in Table 1. The factor analysis of financial indicators has 24 components, among which 7 components have eigenvalues more than 1 and the rest are lower than 1. And their cumulative variance is

79.221%, which can be asserted that these 7 factors have good interpretability.

Table 1: The total interpretation variance of the index data

Constituent	Initial eigenvalue			Rotational load squared		
	Total	Variance%	Cumulation%	Total	Variance%	Cumulation%
1	5.570	23.208	23.208	5.421	22.588	22.588
2	4.517	18.821	42.029	4.847	20.196	42.784
3	2.621	10.921	52.950	2.725	11.354	54.138
4	2.032	8.467	61.416	2.089	8.704	62.842
5	1.718	7.158	68.575	1.594	6.642	69.484
6	1.394	5.808	74.383	1.269	5.288	74.771
7	1.161	4.838	79.221	1.068	4.450	79.221
8	0.957	3.988	83.208			

(2) Analysis of Principal Components

Employing the maximum variance approach for the orthogonal rotation of the component matrix, when arranging the first factor loadings in a descending sequence, the rotated component matrix of the index data is presented in Figure 2. 24 indicators on at least one principal component loadings greater than 0.5, the index division is appropriate, the text selected 24 financial indicators are left.

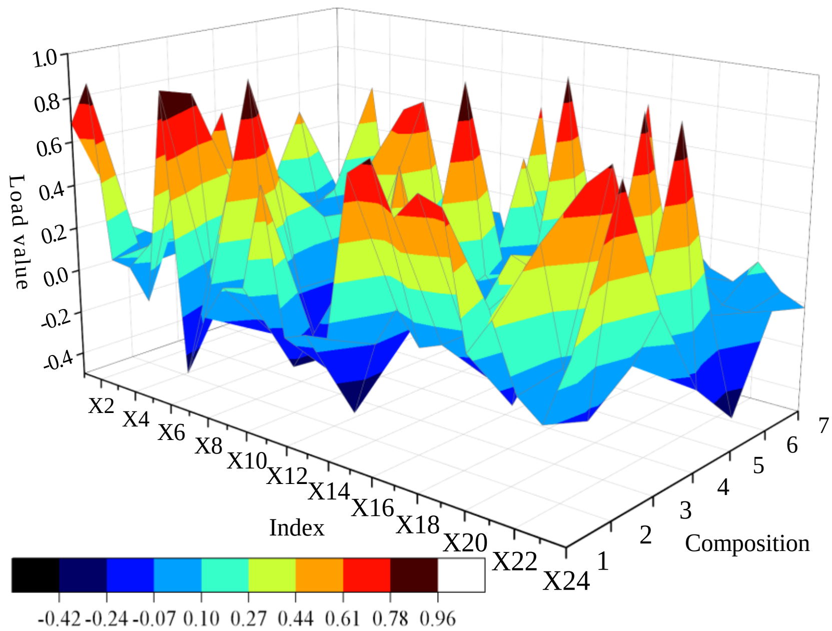

Figure 2: Index data rotation composition matrix

4.2.3 Factor Score Results

Based on the factor score coefficient matrix, we computed each firm's factor scores using the corresponding linear score functions. We then derived factor weights from the rotated variance contribution rates and aggregated the weighted factor scores to obtain a composite financial performance score for each enterprise. The resulting factor scores for the sample firms are reported in Figure 3. Most of the companies are located between 1.00 and 1.50, taking into account the distribution traits of the data, this paper classifies scores lower than 1 into the category of subpar financial performance, higher than 1 has a better financial ability. This classification will be used as the basis for modeling.

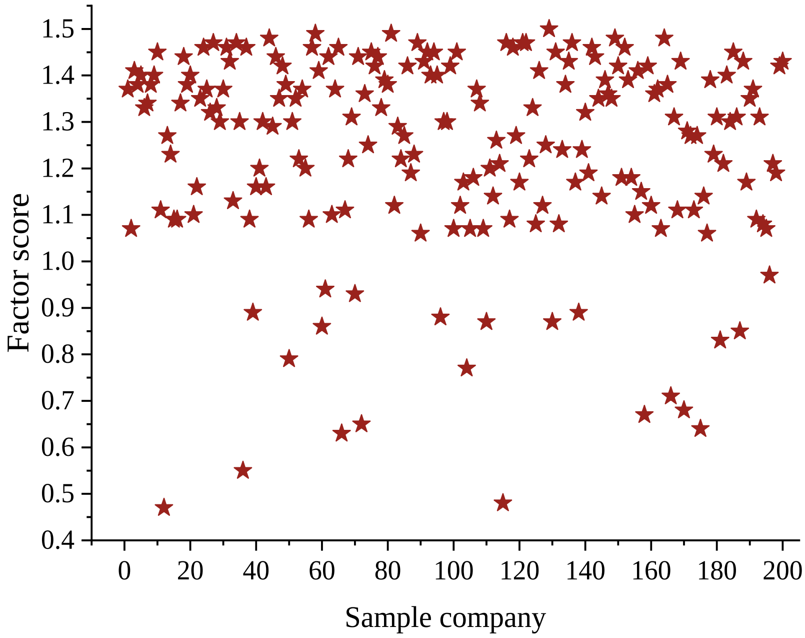

Figure 3: The factor score of the financial index of the sample company

4.3 Sample processing

4.3.1 SMOTE Principles

In order to resolve its interference with the Benford-RF model, this paper uses the SMOTE algorithm to balance the unbalanced data, and then applies the Benford-RF model for classification.

The SMOTE algorithm mainly generates new samples for several classes. N particular practice is given a linear interpolation between an existing sample of several classes and its near-bone sample. The algorithm is performed in several stages: first for each minority sample, calculate the Euclideanity distance between it and another minority sample, and then determine the samples of the N; then randomly select a minority sample from these areas N and use this selected sample to create a new minority sample, and the synthetic formula is given in equation (1).

$$x_{new} = x_i + rand(0,1)*(x_j - x_i) \tag{13}$$

where x_{new} represents the new synthetic sample, x_i is the ith minority class sample, and x_j is the random nearest neighbor sample of x_i.

4.3.2 Sample imbalance handling process

In this paper, sample oversampling is implemented with the help of SMOTE in over_sampling module of imblearn library in Python.

In this paper, the initial dataset classification is plotted with EPS and NAV as axes. The categorization under the unbalanced sample grouping is shown in Figure 4, where the red points with a financial no status of 1 are the majority and the blue points with a financial status of 0 are distributed in a small number. Combined with the actual situation of the sample, set sampling_strategy=0.4, that is, the proportion of positive and negative class samples after oversampling is 0.4. In this paper, the classification of the balanced dataset is plotted again using earnings per share and cash return on net assets as axes as shown in Figure 5. The red points in the figure are samples with a financial no status of 1 and the blue points with a financial status of 0. The distribution of samples with two financial statuses is better balanced and the imbalance of the data is improved.

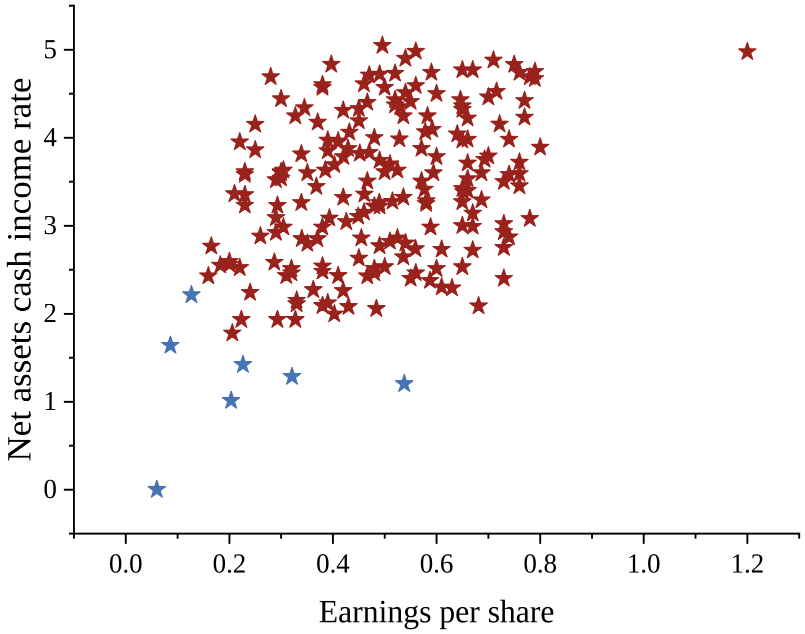

Figure 4: Classification of non-balanced sample groups

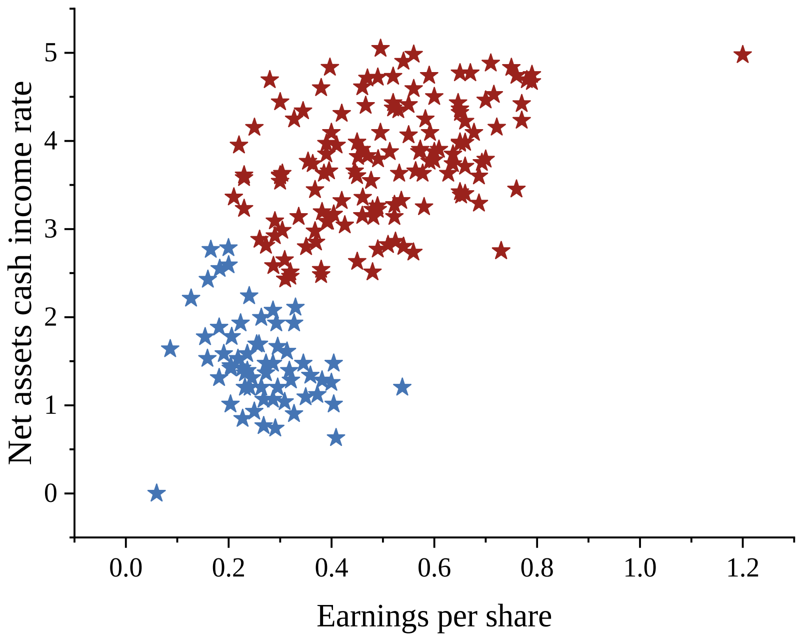

Figure 5: Classification of balanced sample groups

4.4 Model test analysis

We divided the dataset into two parts: training and testing, 7 to 3. In particular, 200 listed companies in the production industry used a test. Then we compared and looked at the following test results.

4.4.1 Test results

After balancing the training data with SMOTE, we trained the candidate classifiers and evaluated their performance using four standard classification metrics. Figure 6 summarizes the test-set results across models.

In terms of accuracy, performance ranked (highest to lowest) as Benford-RF, CatBoost, XGBoost, Random Forest, LightGBM, Logistic Regression, Neural Network, k-Nearest Neighbors, and Decision Tree. The Benford-RF model achieved the strongest overall performance (P=0.925, R=0.952, F1=0.938) and exceeded the other classifiers in accuracy by 1.2-14.4 percentage points. Relative to the standard Random Forest, Benford-RF improved accuracy by 6.0 points and increased precision, recall, and F1 by 6.9, 3.9, and 5.5 points, respectively. These gains suggest that incorporating Benford-based features materially enhances predictive performance for financial-status classification.

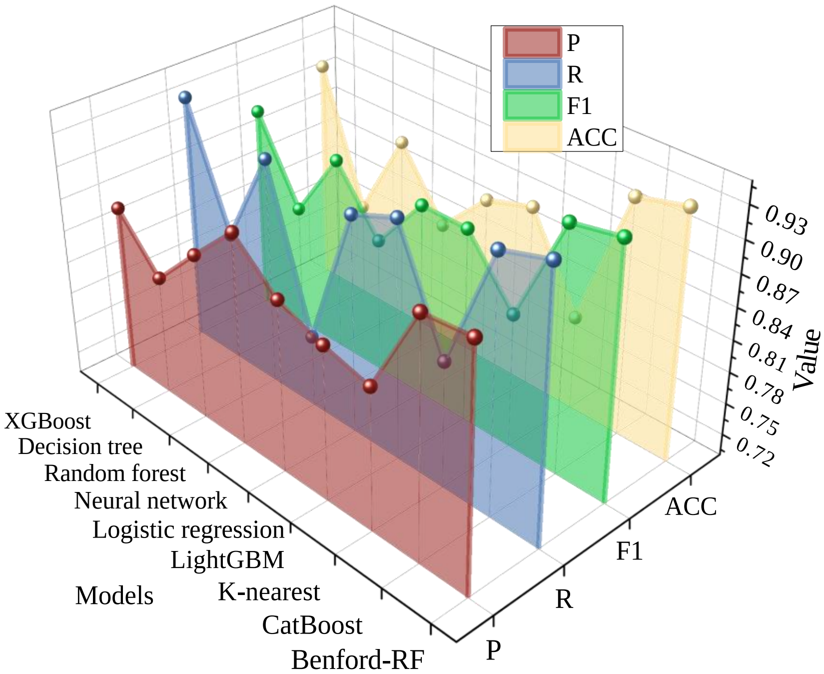

Figure 6: Comparison of test results of different models

4.4.2 AUC visualization and analysis

In contrast to accuracy, review estimation and F1, AUC values are more adapted to the distribution of samples. Simply put, even if the proportion of positive and negative samples is very bad, the AUC values can accurately determine whether the model is good or not. Accuracy and withdrawal easily affect the sample imbalance because they only look at the classification results of a certain class.

The AUC curves of different models are shown in Fig. 7. The AUC value of Benford-RF model is as high as 0.922, which is better than the other models. In addition, relative to the other eight machine learning models, the Benford-RF model in this paper has significantly better predictive performance as an integrated learning classification algorithm.

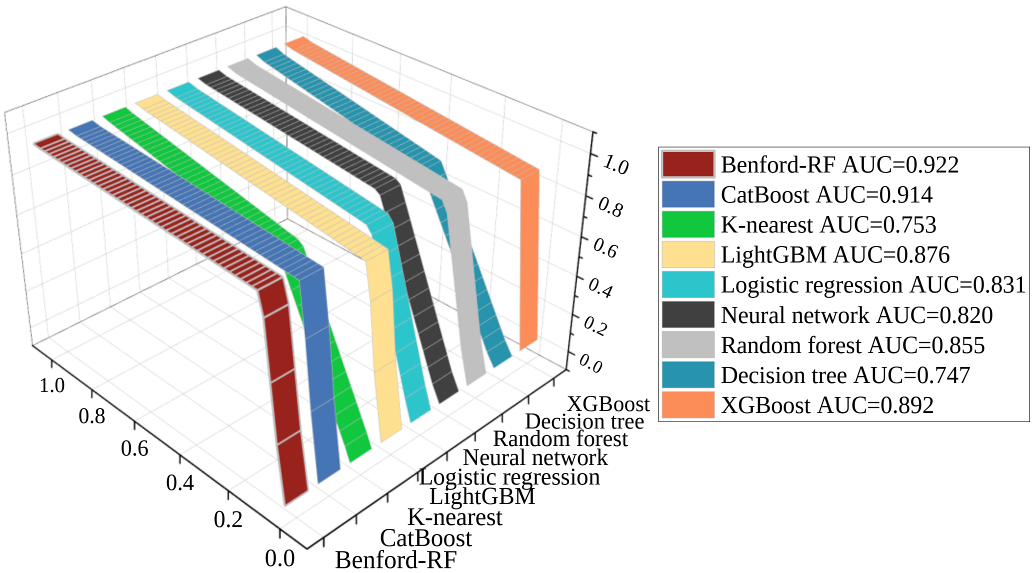

Figure 7: The AUC curve of different models

4.5 Analysis of model applications

This paper takes M listed company as the case study object, selects its financial data information as the characteristics, and uses the trained model to verify its early warning effect. Firstly, the relevant index data of M listed companies are collected and standardized, and the eigenvalues of the warning indexes are converted to the same scale of [0,1]. Finally, the Benford-RF model is used for analysis and seaborn is used to visualize the experimental results.

The results of the analysis and forecasting of the company's financial data are shown in Figure 8. Its average probability value is 1.161, and its financial position is normal, and the forecast results are consistent with the actual situation of the listing of M in 2023. In particular, 185 out of 200 times in the re-examination were correctly predicted, 15 predictions were incorrect, and the prediction accuracy ratio was 92.5%.

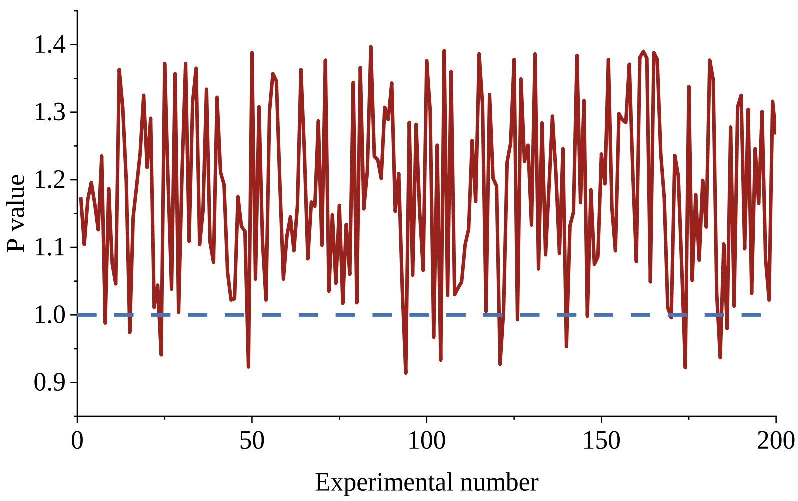

Figure 8: The company's financial data analysis predictive results

5 Conclusion

In this research document, we explore a system for analyzing the financial data of the enterprise, based on the digital economy. The system uses a layered modular frame with a data stream in the nucleus. It will complete the entire process from basic data collection to high-level decision-making. We have built Benford-RF financial data analysis and forecasting models in terms of financial management, and also empirically analyzed the impact of models on forecasting the financial situation of enterprises. The main research results are as follows: first, the financial data index system is established, and the comprehensive factor scores are calculated through the factor analysis method to effectively classify the comprehensive financial capability of the enterprise, and the enterprises with scores less than 1 and scores greater than 1 are categorized into the financial bad category and the financial normal category, respectively. Secondly, the SMOTE method is utilized to deal with the data imbalance problem, which finally makes the ratio of positive and negative category samples reach 0.4. Thirdly, the value of each evaluation index of Benford-RF model in this paper performs the best in the comparative test, and its precision rate, recall rate and F1 score are all greater than 0.92, with the accuracy rate higher than other classification algorithms ranging from 1.2% to 14.4%, and the AUC value is the highest, too. It reflects the superior financial status prediction performance of Benford-RF model. Fourthly, as a case-study entity, the model is implemented on M listed companies. The prediction precision attains 92.5%. This validates the accuracy and practicality of the prediction model developed in this research.

Finally, the Benford-Random Forest (Benford-RF) model developed as part of this study can help companies predict their financial position and provide useful reference information for businesses. This is an optional part of the corporate financial data analysis system. Using this model, he can direct companies to manage the budget, cost supervision, and risk warning.

About the Author

Xianrong Li was born in Weinan, Shaanxi, P.R. China, in 1983. She obtained a bachelor's degree from Yichun University in China. She is currently working at Smart Finance and Business School Hefei College of Finance & Economics. Her main research direction is Intelligent Finance and Data Analysis.

References

[1] Snihur, Y., & Tarzijan, J. (2018). Managing complexity in a multi-business-model organization. Long Range Planning, 51(1), 50-63.

[2] Ji, L., Jiang, W., & Deng, B. (2024). Corporate digital transformation and cost of equity capital. China Journal of Accounting Studies, 12(4), 693-721.

[3] Zaitsev, A., Kichigin, O., & Korotkova, A. (2019, October). Standard dynamic financial analysis and control tools of an enterprise in the time of digital economy. In Proceedings of the 2019 International SPBPU Scientific Conference on Innovations in Digital Economy (pp. 1-7).

[4] Jia, T., Wang, C., Tian, Z., Wang, B., & Tian, F. (2022). Design of digital and intelligent financial decision support system based on artificial intelligence. Computational

intelligence and neuroscience, 2022(1), 1962937.

[5] Liu, L. X., Sun, Z., Xu, K., & Chen, C. (2024). AI-driven financial analysis: Exploring ChatGPT’s capabilities and challenges. International Journal of Financial Studies, 12(3), 60.

[6] Pena, Y., Correal, D., & Miranda, E. (2017). A review of the current situation, challenges and opportunities in the financial analysis of enterprise architectures. World Review of Science, Technology and Sustainable Development, 13(2), 145-173.

[7] Zheng, H. (2024). The Transformation and Challenges of Enterprise Financial Management Models in the Era of Big Data. Journal of Modern Business and Economics, 1(3).

[8] Gao, J. (2022). Analysis of enterprise financial accounting information management from the perspective of big data. International Journal of Science and Research (IJSR), 11(5), 1272-1276.

[9] Lu, L., Yang, S., & Li, Q. (2024). The interaction of digital economy, artificial intelligence and sports industry development--based on China PVAR analysis of provincial panel data. Heliyon, 10(4).

[10] Li, H. (2023). Analysis of Effective Paths for Informationization Construction of Enterprise Financial and Accounting Management. Accounting and Corporate Management, 5(11), 27-32.

[11] Song, Y., & Jiang, Y. (2024). How does the digital economy drive the optimization and upgrading of industrial structure? the mediating effect of innovation and the role of economic resilience. Sustainability, 16(4), 1352.

[12] Bogachov, S., Pluhatar, T., Plakhotnik, O., Alieksieieva, O., & Bondar, V. (2021). Innovative business development in the digital economy. Entrepreneurship and Sustainability Issues, 8(4), 682.

[13] Kan, D., Lyu, L., Huang, W., & Yao, W. (2022). Digital economy and the upgrading of the global value chain of China’s service industry. Journal of Theoretical and Applied Electronic Commerce Research, 17(4), 1279-1296.

[14] Fatticia, R., Harjoni, H., Christiaan, P., Julyarman, N., & Ariyanti, R. (2024). The Influence of Fintech on Traditional Financial Management. Journal Markcount Finance, 2(2), 194-204.

[15] Wu, W. (2022). Credit risk measurement, decision analysis, transformation and upgrading for financial big data. Complexity, 2022(1), 8942773.

[16] Agyapong, D. (2021). Implications of digital economy for financial institutions in Ghana: an exploratory inquiry. Transnational Corporations Review, 13(1), 51-61.

[17] Siddiqui, N. A. (2025). Optimizing Business Decision-Making Through AI-Enhanced Business Intelligence Systems: A Systematic Review of Data-Driven Insights in Financial And Strategic Planning. Strategic Data Management and Innovation, 2(1),

202-223.

[18] Al-Okaily, M., & Al-Okaily, A. (2024). Financial data modeling: an analysis of factors influencing big data analytics-driven financial decision quality. Journal of Modelling in Management, 20(2), 301-321.

[19] Al-Omush, A., Almasarwah, A., & Al-Wreikat, A. (2025). Artificial intelligence in financial auditing: redefining accuracy and transparency in assurance services. EDPACS, 70(6), 1-20.

[20] Alawneh, O. M., Tawalbeh, J., & AlOqaily, A. N. (2025). Reveling The Interrelationship Of Digital Transformation And Business Intelligence Techniques To Boost Human Capital Development. Economics-Innovative And Economics Research Journal, 13(1), 509-524.

[21] Megawaty, M., Muhammad, N., Nawir, F., & Hamdat, A. (2025). Cognitive Learning in Banking Human Resource Development: Strategies for Enhancing Competence and Adaptability in the Digital Era. Jurnal Manajemen Perbankan Keuangan Nitro, 1(1), 22-28.

[22] Wang, M. (2025, March). Research on efficiency evaluation of financial management in universities based on GWO-BPNN algorithm. In Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024) (Vol. 13550, pp. 1204-1212). SPIE.

[23] Li, P., & Zhang, L. (2025). Application of big data technology in enterprise information security management. Scientific Reports, 15(1), 1022.

[24] Khan, S. A. R., Godil, D. I., Jabbour, C. J. C., Shujaat, S., Razzaq, A., & Yu, Z. (2025). Green data analytics, blockchain technology for sustainable development, and sustainable supply chain practices: evidence from small and medium enterprises. Annals of Operations Research, 350(2), 603-627.