Text Mining-based Enterprise Financial Performance Evaluation in the Context of Enterprise Digital Transformation

Changrong Guo, Jing Xing

Department of Accounting, Xinzhou Normal University, Xinzhou, 034000, China

Abstract—As enterprises gradually move towards digitalization, it is increasingly difficult to accurately evaluate changes in corporate financial performance. To improve this situation, the study uses a text mining algorithm based on the web crawler principle to extract keywords from corporate annual reports, select representative financial performance indicators through IF-FDP, and construct a corporate financial performance evaluation model using the entropy weighting method. The performance comparison experiments of the text mining algorithm proposed in the study show that the accuracy-recall rate area under the line of the text mining algorithm proposed in the study is 0.83 and the average F-value is 0.34, which are both better than other algorithms. In the empirical analysis of the financial performance evaluation model, it was found that the financial performance evaluation model had the smallest absolute error of 0.3%, which was lower than the other models. The above results indicate that both the text mining algorithm and the performance evaluation model proposed in the study outperform the comparison algorithm and model. Therefore, the performance evaluation model proposed by the study can be used to effectively evaluate the financial performance of enterprises accurately and promote the development of enterprises, which has practical application value.

Keywords—Web crawler; text mining; IF-FDP; entropy method; financial performance; evaluation model

I. Introduction

As the world enters the information age in the 21st century, the traditional economy is surging towards the digital economy. In the face of the impact of the digital economy, traditional enterprises are transforming and upgrading with the help of digital technology [1]. In order to explore the financial changes of traditional enterprises in the process of digital transformation and to understand the significance of digital transformation to the development of enterprises, the study will delve into the financial performance performance of transforming enterprises. Studies have mainly focused on the analysis of the financial mechanism of digital transformation and the analysis of the path of digital transformation, and there is less analysis of the financial performance changes in the process of digital transformation of enterprises [2]. However, in the face of the performance assessment of digitally transformed enterprises, the use of traditional performance assessment methods may not be comprehensive and accurate enough. Therefore, there is an urgent need to develop a performance evaluation model that is suitable for use by

transforming companies.[3] Text mining algorithms are intelligent algorithms that extract knowledge from large amounts of unstructured or semi-structured text, organise that knowledge into complete information and apply it [4]. Text mining algorithms have a wide range of applications in financial management because of their multi-scientific knowledge advantages due to their integration of multiple scientific fields[5]. In order to better evaluate the financial performance of enterprises accurately, the study applies text mining technology to the enterprise financial performance evaluation model, using its function of extracting textual information to complement the traditional evaluation model and thus improve the overall performance of the performance evaluation model. This research innovatively combines the Text mining algorithm based on the Pathon crawler principle with enterprise financial performance evaluation, and proposes a new enterprise financial performance evaluation model. This model not only makes up for the gap in the integration of enterprise finance and Text mining related fields, but also improves the accuracy of enterprise financial performance evaluation, It provides data support for the development of enterprise financial performance evaluation in the process of Digital transformation. This paper is mainly divided into the following five parts. The first section is the analysis of the research status in the field of performance evaluation and Text mining technology. The second section mainly constructs the financial performance evaluation model based on Text mining algorithm. The third section mainly analyzes the actual performance of the financial performance evaluation model constructed through research. The fourth section discusses the results of this experiment. The fifth section is the conclusion of this study.

II. Review of the Literature

Performance evaluation is a clear indication of performance over time, so performance evaluation metrics are widely used in many fields [6]. Xu et al. apply machine learning techniques to a performance evaluation model based on a two-layer overlay framework to address the lack of accuracy in the transaction and risk assessment of second-hand property prices. The results of the empirical analysis of the model show that the performance evaluation model has higher accuracy and outperforms traditional performance evaluation methods [7]. The Fanelli team proposed a performance evaluation model based on a specific frequency of survey results to address the issue of poor performance in the public health sector. Empirical analysis of

this performance evaluation model found that its implementation can improve the performance of the public sector, which has practical significance [8]. Karimi et al. propose a performance evaluation model based on enhanced additivity ratios to address the problem of inaccurate performance evaluations of knowledge workers, and test the model to find that it outperforms traditional methods for evaluating the performance of knowledge workers, with a significant increase in accuracy over traditional models [9]. The Galagedera team addresses the question of whether mutual funds The results of an empirical analysis of a network data envelopment analysis model of mutual fund management performance, which addresses the question of whether costs and expenses can be effectively managed, show that the performance indicator model proposed in the study can effectively improve the efficiency of payment management and has practical application [10].

Text mining has applications in the fields of intelligent systems, information retrieval, information processing, etc. Akundi et al. observed a significant increase in academic and industrial research in the field of systems engineering and proposed a comprehensive and structured integration of research in the field through the use of text mining techniques in order to understand the existing research directions in the field. "system modelling language", "physical system" and "production" are the most used terms in systems engineering research, with system modelling language being the most widely used modelling language [11]. Leem et al. propose a text mining approach to sentiment analysis of customers' online evaluations for Kakao mobile banking service, which is ambiguous and unclear. Through regular analysis, it was found that the proposed method is of practical use to improve service quality, increase customer satisfaction and assist in maintaining and upgrading the application, thus effectively increasing the completion rate of mobile banking services [12]. Zhou's team proposed to use text mining techniques to extract experimental data related to minimum streaming speed to address the problem that the selection of minimum streaming speed is easily influenced by subjective impressions, and to build a database was established. The method was validated and found to be 83% accurate in extracting the function

parameters and more objective and accurate in selecting the minimum fluidisation velocity, which can effectively solve the problem of high empirical correlation in data selection [13]. An empirical study found that anaesthesia, oestrogen receptors and fengi hydrogen receptor mediators were the most important event drivers in aquatic systems distributed across China, and the proposed approach can provide objective and valid information for water quality assessment in the era of big data [14].

To sum up, text mining method has strong text processing ability, which can meet the need of extracting and processing a large amount of text data in financial performance evaluation model. In addition, it can be found that the current financial evaluation model often only focuses on a single index, ignoring other important factors. This kind of single index evaluation is easy to ignore the overall risk and performance of the enterprise, and cannot fully and accurately evaluate the value of the enterprise. At present, there are few researches on the combination of text mining technology and financial performance evaluation model, so the combination of performance evaluation model built based on text mining algorithm is studied, hoping to improve the overall evaluation performance of enterprise financial performance evaluation model with the help of the comprehensive performance of text mining, so as to accurately evaluate the value of enterprises.

III. CONSTRUCTION OF A FINANCIAL PERFORMANCE EVALUATION MODEL BASED ON TEXT MINING ALGORITHMS

A. Text Mining Algorithms based on Pathon Crawler Principles

A web crawler, also known as a web spider, is a web technology in which a computer automatically collects specific data through a program or script according to certain rules [15-16]. Python, one of the most widely used crawler programming, can not only automatically crawl text, images, audio and video information in large quantities, but Python also provides a large number of third-party libraries to assist in information crawling [17]. The principles of web crawling technology are shown in Fig. 1.
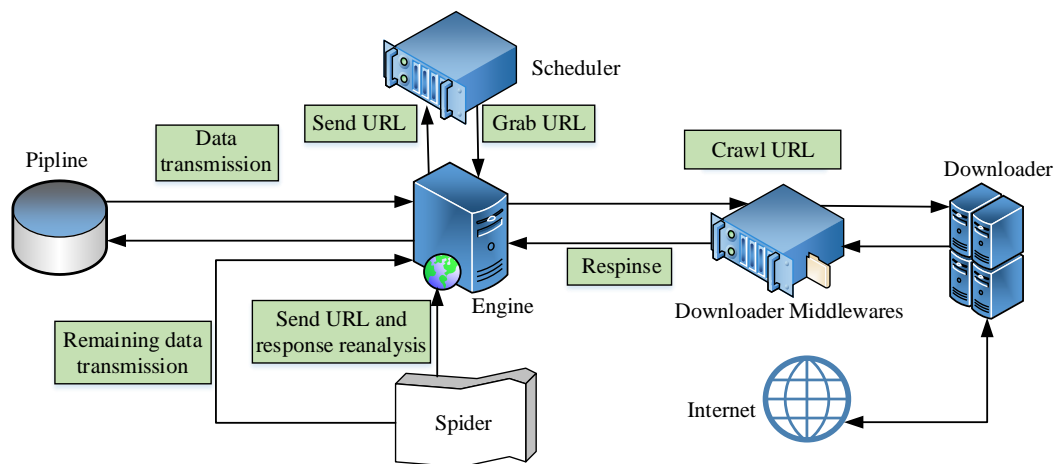

Fig. 1. Web spider principle.

As shown in Fig. 1, first the web crawler needs to send a URL to the engine, which in turn forwards the URL to the scheduler. The dispatcher then processes the URL, and the processed URL is returned to the engine. The engine takes the received URL and gets the response information through the downloader intermediate component, which is then parsed, filtered and stored by the crawler. Finally, the remaining filtered data is forwarded to the pipeline for processing. Text mining algorithms are a process of fusing word processing techniques with intelligent learning algorithms to extract valuable textual information and knowledge from text and reorganise the extracted information. The research uses text mining algorithms as semi-structured text information processing algorithms, the process of which can be divided into three parts, i.e. text information acquisition and processing, text information parsing and thesaurus building and text information quantitative analysis. The structure of the text mining algorithm proposed in the study is shown in Fig. 2.
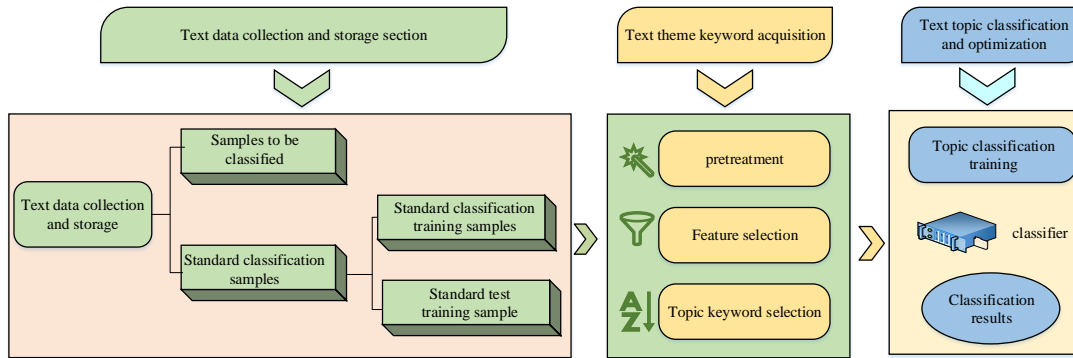

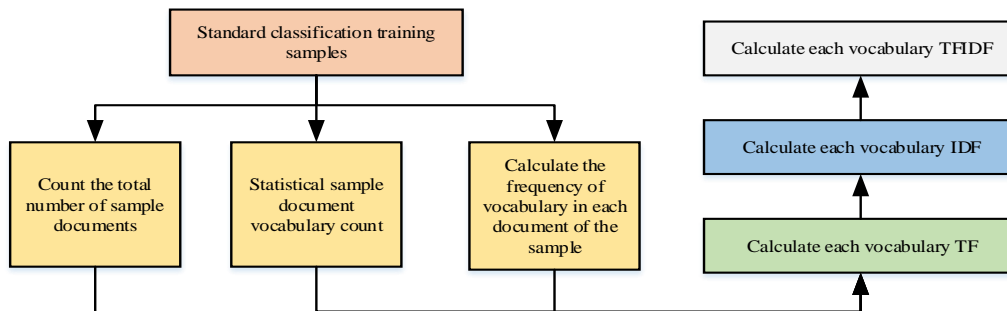Fig. 2. Text mining algorithm structure.

Fig. 3. Text word frequency feature weight calculation process.

As shown in Fig. 2, the proposed text mining algorithm firstly crawls the annual report text of enterprises through Python web crawler technology technique, and pre-processes, feature selection and keyword selection for the crawled text. Finally, the keywords are classified and calculated statistically through classifier and word frequency statistics techniques. In the text data acquisition and storage part of the text mining algorithm, the study collected the text of annual reports of A-share high-end manufacturing enterprises in Shanghai and Shenzhen from 2012 to 2022, converted the text to TXT format through PDFminer, and then carried out word separation processing, de-duplication operations, word extraction and information storage on the converted text. In the subject keyword acquisition part, the algorithm uses the jieba word splitter to cut words and clean the stored information, eliminating companies with abnormal financial or other conditions, as well as samples with serious data deficiencies, and completing the operation of removing obsolete words, thus obtaining the text word set. The text mining algorithm filters out the discontinued words according to the text requirements, selects the appropriate keywords and calculates their corresponding feature weights. Finally, the selected keywords are fed into the classifier to build a subject classification keyword database. As an important technique for text mining, the word frequency statistical calculation method is an objective statistical method that can identify hot words and quantify the change trend by the frequency of occurrence of keyword words of a specific topic. The process of calculating the word frequency feature weights is shown in Fig. 3.

As can be seen from Fig. 3, after the study has counted the pre-processed text information, the feature weights will be calculated by the feature algorithm, i.e. Term Frequency-Inverse Document Frequency (TF-IDF) algorithm for the feature words. tP is the number and frequency of occurrences of the feature words in the text, and its calculation formula The formula is shown in equation (1).

$$TP_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

In equation (1), $n_{i,j}$ indicates the number of times the keyword t_i appears in the document d_j . IDF is the frequency of the subject keyword in other texts, and is a quantitative indicator of the prevalence of the keyword. If the IDF value of a keyword is smaller, it means that this keyword has a good classification function; if its IDF value is larger, it means that this keyword is common. The formula for calculating IDF is shown in equation (2).

$$IDF_i = \log \frac{|D|}{1 + |j : t_i \in d_j|}$$

(2)

In equation (2), $|D|$ is the number of all documents; $|j : t_i \in d_j|$ indicates the number of documents containing the keyword t_i . TF-IDF is a quantitative indicator of the importance of keywords in the text and is calculated as shown in equation (3).

$$TP - IDF_{i,i} = \frac{n_{i,j}}{\sum_k n_{k,j}} * \log \frac{|D|}{1 + |j : t_i \in d_j|}$$

(3)

In equation (3), $n_{i,j}$ indicates the number of occurrences of the keyword t_i in the document d_j ; $|D|$ is the number of all documents; $|j : t_i \in d_j|$ indicates the number of documents containing the keyword t_i . After calculating the TF-IDF values of the keywords, the study used the cardinality check combined with the weights of the feature terms to evaluate and select the feature terms. The formula for calculating the keyword cardinality value is shown in equation (4).

$$x^2 = \frac{N(AD - BC)^2}{(A+C)(A+B)(B+D)(B+C)}$$

(4)

In equation (4), indicates the frequency of the keyword A t_i in positive documents; indicates the frequency of the keyword B t_i in positive documents; C indicates the frequency of the keyword t_i not in positive documents; D indicates the frequency of the keyword t_i not in negative documents. The study derived the text feature word selection score based on normalisation, which is shown in equation (5).

$$Score(t_i) = \alpha \cdot \left[\frac{n_i}{N} * 100 \right] + \beta \left[\frac{c_i}{N} * 100 \right]$$

(5)

In equation (5), n_i denotes the keyword t_i in the TF-IDF positive order number; c_i denotes the keyword t_i in the chi-square test positive order number; α and β set the parameter value to 0.5. The feature words were sorted into sequences according to the selected scores, and the weight scores of the keyword propagation nodes were calculated as shown in equation (6).

$$S(V_i) = (1-d)W'(V_i) + d * W'(V_i) \sum_{j \in Ln(V_i)} \frac{w_{ji}}{\sum_{k=V_j} w_{ji}} S(V_j)$$

(6)

In equation (6), is the weight of the node $S(V_i)$ V_i , w_{ji} is the weight of the nodes V_i and V_j , is the set of nodes , and $j \in Ln(V_i)$ V_i $k = V_j$ is the set of nodes V_j refers to. After convergence of the weights of the nodes, the nodes are normalised for sorting, so that keywords can be selected. The normalisation formula is shown in equation (7).

$$Score(t_i) = \left[\frac{n_i}{N} * 100 \right]$$

(7)

In equation (7), n_i denotes the positive ordinal number of the keyword t_i in the keyword weighting. After counting the frequency of the main keywords, the study selects the companies that are in digital transformation from 2012-2022 by portraying the degree of digital transformation through the weight of digital keywords. The specific calculation formula equation (8) is shown.

$$q_i = \frac{a_{it}}{\sum_t^i a_{it}}$$

(8)

In equation (8), a_{it} indicates the total number of digitised keywords for the company i in the year t .

B. Construction of a Financial Performance Evaluation Model based on the Entropy Method

Financial performance evaluation models are models that quantify financial data through performance evaluation indicators, and then make a comprehensive evaluation of performance based on set evaluation criteria [18]. The most widely used performance evaluation models are the financial management theory performance evaluation model, and the mathematical and statistical model financial performance evaluation model [19]. The financial evaluation model proposed in the study belongs to the mathematical statistical model, which is a performance evaluation model based on the text mining algorithm and the entropy weight method.
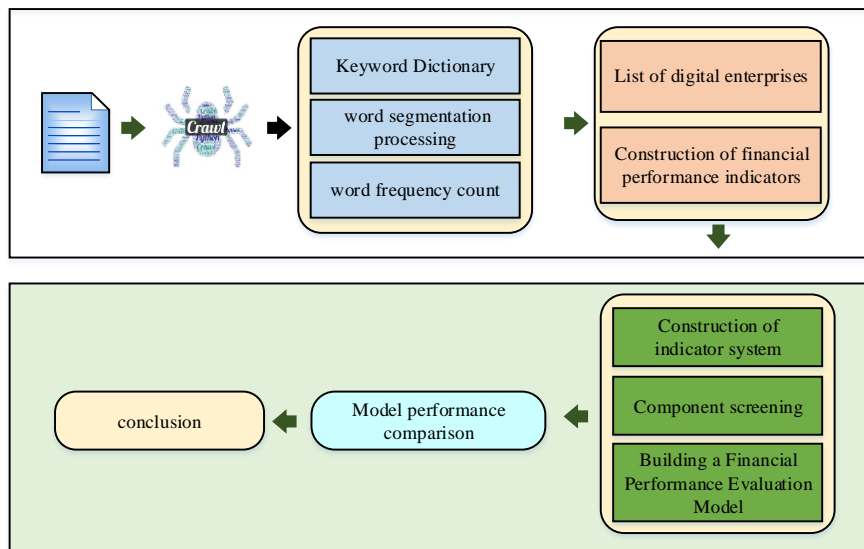
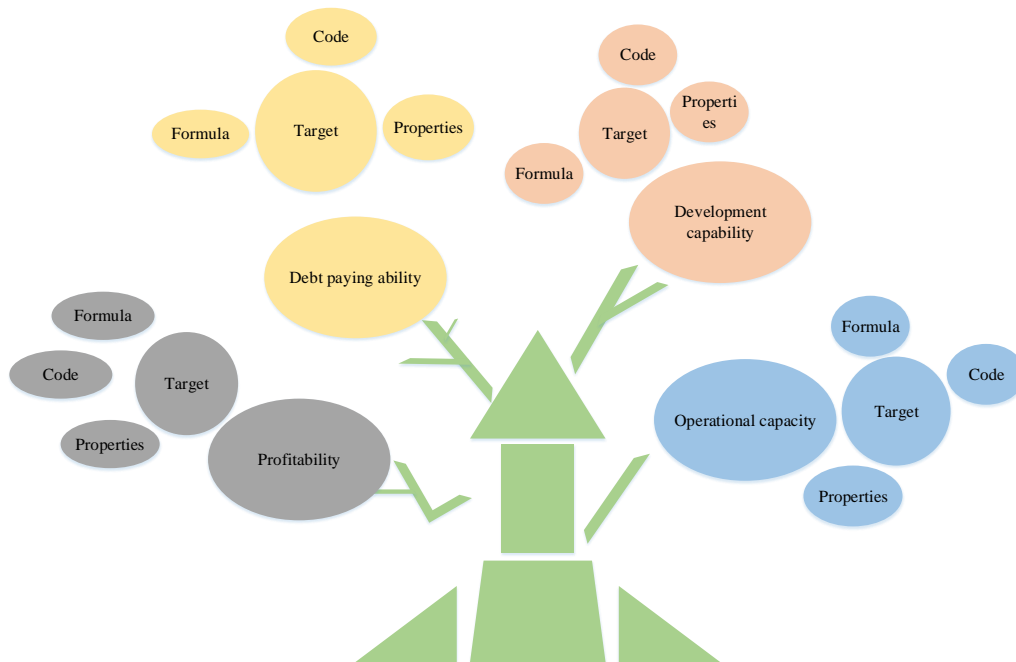Fig. 4.   Structure of performance evaluation model based on text mining.

Fig. 5.   Structure of enterprise financial performance evaluation index system.

As can be seen from Fig. 4, the first step of the model is the processing of text information. The model constructs a digital keyword dictionary through a text mining algorithm based on the principle of crawlers, and carries out word separation processing to obtain the required digital keywords, and finally carries out quantitative analysis on the obtained keywords, calculates the keyword word frequency and inverse document frequency, and selects a list of enterprises for digital transformation. The second step is to build an enterprise financial performance evaluation index model. In order to comprehensively reflect the business quality and financial performance of the enterprise into the whole cycle, and to solve the problem that a single financial indicator can only evaluate the unilateral financial situation of the enterprise, the study integrates the entropy weight method with each indicator of the financial dimension to build a more scientific and reasonable enterprise financial performance evaluation index model. An empirical analysis of the current financial situation is carried out, and statistical descriptions of the full range of financial conditions in each dimension are separately conducted to explore the changes in the financial performance of manufacturing enterprises. Finally, the entropy method financial performance evaluation model is compared with the performance of other algorithmic financial performance evaluation models to verify the reasonableness of the improved performance evaluation model proposed by the study. The structure of the enterprise financial performance evaluation index system is shown in Fig. 5.

As shown in Fig. 5, the study establishes an enterprise financial performance evaluation index system in terms of the relevance, systematicity, importance and feasibility of the enterprise's financial performance in accordance with four dimensions: profitability, debt servicing, development and operational capability. The study constructed a total of 17 indicators for the preliminary index system of enterprise financial performance evaluation, as shown in Table I.

TABLE I. PRELIMINARY SELECTION INDEX SYSTEM FOR ENTERPRISE FINANCIAL PERFORMANCE EVALUATION

Dimension	Indicator code	Financial index	Indicator Definition
Profitability	A1	Return on assets	Pre tax profit/average total assets
	A2	Net profit margin of total assets	Net profit/average balance of total assets
	A3	Net return on assets	Net profit/average balance of shareholders' equity
	A4	Operating profit margin	Operating profit/revenue
	A5	Return on invested capital	(Net profit+financial expenses)/(Total assets -Current liabilities+Notes payable+Short term borrowings+Non current liabilities due within one year) year)
	A6	Cost profit margin	Total profit/total cost expenses
	A7	Earnings per share	Current net profit attributable to ordinary shareholders/weighted average number of outstanding ordinary shares in the current period
Debt repayment ability	A8	Quick ratio	Quick assets/current liabilities
	A9	Current ratio	Current assets/current liabilities
	A10	Asset liability ratio	Total liabilities/total assets
	A11	Cash to current liability ratio	Net cash flow from operating activities/total liabilities
Development capability	A12	Total asset growth rate	Increase in total assets this year/total assets at the beginning of the year
	A13	Net asset growth rate	Increase in net assets this year/total net assets last year
	A14	Operating revenue growth rate	(Current year's operating revenue - previous year's operating revenue)/previous year's operating revenue
	A15	Current Asset turnover	Operating income/average occupancy of current assets
Operational capacity	A16	Net profit growth rate	(Current net profit - Previous net profit)/Previous net profit
	A17	Total Asset turnover	Main business income/total assets

As shown in Table I, the study divided the financial performance indicators into four dimensions, in which profitability is the ability of an enterprise to earn profits within a certain period of time; solvency is the ability of an enterprise to use its assets to repay short-term and long-term debts, and cash payments; development capacity refers to the trend of future cash flows from operating activities; and operational capacity is the ability of an enterprise to manage its operations within a certain operating period [20]. The study uses principal component analysis to screen the 17 primary indicators under these four dimensions, determines the principal components through the cumulative variance contribution rate, and selects representative indicators according to the criterion of cumulative variance contribution rate of 75%. The formula for calculating the contribution value of financial indicators' contribution ratio is shown in equation (9).

$$a_i = \frac{\lambda_i}{\sum_i^m \lambda_i}$$

(9)

After determining the representative indicators for performance evaluation, the weighting of each financial indicator needs to be assigned, and the weighting method used in the study is the entropy method. The entropy weighting method is a method of objectively assigning weights to the indicator system using known data, which can avoid errors arising from subjective impressions and has the advantages of high calculation precision and accuracy. The higher the entropy value, the less information the indicator contains, and the more information it contains. The entropy weighting method first requires pre-processing of the original data, including the standardisation of positive indicators, negative indicators and moderate indicators. The formula for the standardisation of positive indicators is shown in equation (10).

$$Y_{ij} = \frac{X_{ij} - \min(X_{ij})}{\max(Xi_j) - \min(X_{ij})}$$

(10)

The formula for the standardisation of negative indicators is shown in equation (11).

$$Y_{ij} = \frac{\max(X_{ij}) - X_{ij}}{\max(Xi_j) - \min(X_{ij})}$$

(11)

The formula for the standardisation of moderate indicators is shown in equation (12).

$$Y_{ij} = 1 - \frac{|X_{ij} - Xj|}{\max|X_{ij} - Xj|}$$

(12)

In equation (12), X_j is the fixed value of the moderate indicator. After pre-processing the indicators, the characteristic contribution of the indicators is calculated as shown in equation (13).

$$P_{ij} = \frac{Y_{ij}}{\sum_{ij}^{n} Y_{ij}}$$

(13)

In equation (13), P_{ij} is the characteristic contribution of the j indicator for the i company. The higher the value, the more information is contained in the indicator. The calculation formula is shown in equation (14).

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} P_{ij} \ln(P_{ij}), 0 \le e_j \ge 1$$

(14)

For the calculation of the degree of variation of the indicators, the formula for calculating the coefficient of variation is shown in equation (15).

$$g_j = 1 - e_j$$

(15)

Finally, in order to calculate the comprehensive evaluation score of the indicators and determine the weight of the

evaluation indicators, the calculation formula is shown in equation (16).

$$W_j = \frac{g_j}{\sum_{i-1}^{m} g_j}, j = 1, 2, \dots m$$

(16)

IV. TEXT MINING ALGORITHM PERFORMANCE COMPARISON AND PERFORMANCE EVALUATION MODEL EMPIRICAL ANALYSIS

The study extracts keywords from corporate annual report texts by text mining algorithms and classifies the results, using the TF-IDF metric to calculate the frequency of keywords in corporate annual report texts. In order to better analyse the keywords with higher hotspots, the study results only show the top ten digitisation-related keywords that appear frequently in the text of corporate annual reports from 2012 to 2022. Also to verify the practicality of the text mining algorithm, the study verifies the accuracy of the classification results of the text mining algorithm. The keyword frequency calculation results and text mining ROC curves are shown in Fig. 6.
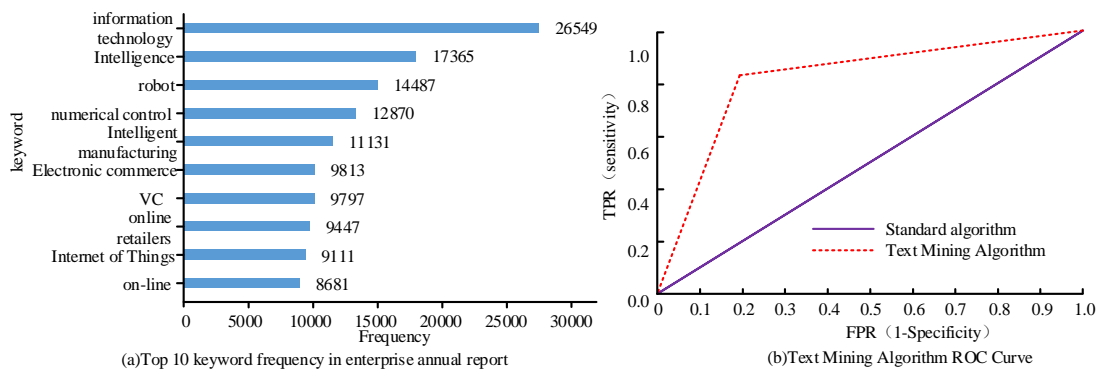

(a)Top 10 keyword frequency in enterprise annual report

(b)Text Mining Algorithm ROC Curve

Fig. 6. Keyword frequency results and text mining algorithm ROC curve.

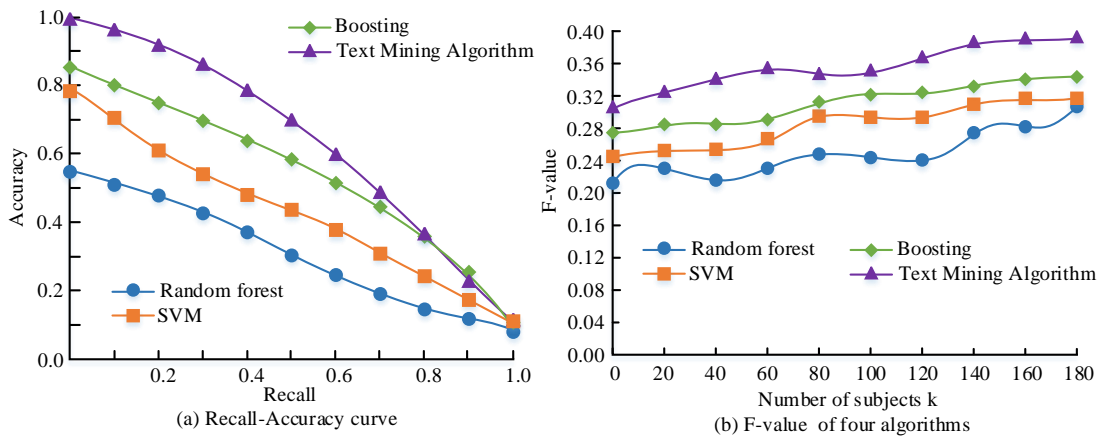(a) Recall-Accuracy curve

(b) F-value of four algorithms

Fig. 7. Accuracy-recall rate and f-value results of four algorithms.

Fig. 6(a) shows the top ten digital keywords appearing in the text of annual reports from 2012 to 2022, of which the top three are "information technology", "intelligence" and "robotics". The top three keywords are "informatization", "intelligence" and "robotics". Fig. 6(b) shows the ROC curve of the text mining algorithm. From Fig. 6(b), it can be seen

that the area under the ROC curve of the text mining algorithm is 0.87, which is significantly higher than that of the traditional standard algorithm and has practical application value. The study compares the performance of the text mining algorithm with other algorithms. The experiment was conducted in MATLAB 2017b and simulated using Simulink.

The basic experimental environment settings are shown in Table II.

Parameter variables	Parameter selection
Overall implementation platform	Simulink
Operating system	Windows10
Operating environment	MATLAB
PC side memory	12G
CPU main frequency	2.62Hz
Global procurement unit	RTX-2070
Central Processing Unit	i7-8700
Data storage	MySQL data bank
Data regression analysis system	SPSS26.0

The results of the accuracy-recall curve and F-value comparison of the four algorithms are shown in Fig. 7.

Fig. 7(a) shows the accuracy-recall curves of the text mining algorithm and the comparison algorithm, the comparison algorithm is the random forest algorithm, the support vector machine (SVM) algorithm, and the boosting method algorithm. From Fig. 8(a), it can be seen that the accuracy-recall curves of both the text mining algorithm and the comparison algorithm algorithms show a decreasing trend, with the area under the line of the text mining algorithm being 0.83, the area under the line of the SVM algorithm being 0.56, the area under the line of the boosting method algorithm being 0.62, and the area under the line of the random forest algorithm being 0.46. From the above results, it can be concluded that the area under the line of the accuracy-recall rate of the text mining algorithm is significantly larger than that of the rest of the algorithms, with the best performance. Fig. 7(b) shows the F-values of the text mining algorithm and the comparison algorithm. From Fig. 7(b), it can be seen that the F-values of the four algorithms show an overall increasing trend with the increase of k. The average value of the F-value of the text mining algorithm is 0.34, the average value of the SVM algorithm is 0.31, the average value of the boosting method is 0.29, and the average value of the random forest algorithm is 0.23, and all the F-values of the text mining algorithm are higher than the F-values of the remaining algorithms values, and the keyword selection accuracy performance was significantly better than the other compared algorithms. The study conducted principal component analysis on the nine indicators under the profitability dimension in Table I, and selected representative indicators by calculating the correlation coefficient selection and score coefficients through SPSS, and the correlation coefficient matrix is shown in Fig. 8.

As shown in Fig. 8, the horizontal and vertical coordinates are the coefficients of the corresponding indicators A1-A9. From the results of Fig. 8, it can be seen that the index coefficients of five indicators, A1, A2, A3, A4 and A6, are relatively high, so these five indicators are selected as the main component representative factors. The study then evaluates the 2012-2022 annual reports of enterprises through the entropy weighting method financial performance evaluation model, and Fig. 9 shows the results of the evaluation of enterprise profitability and development capability.
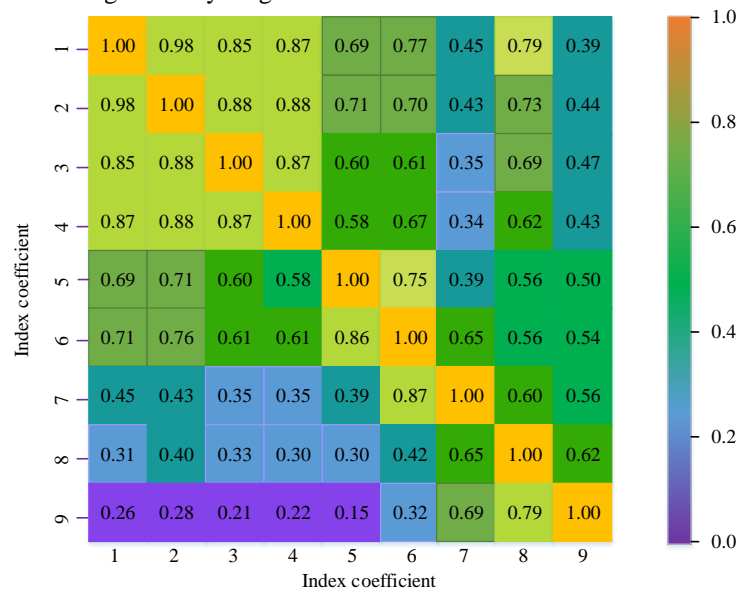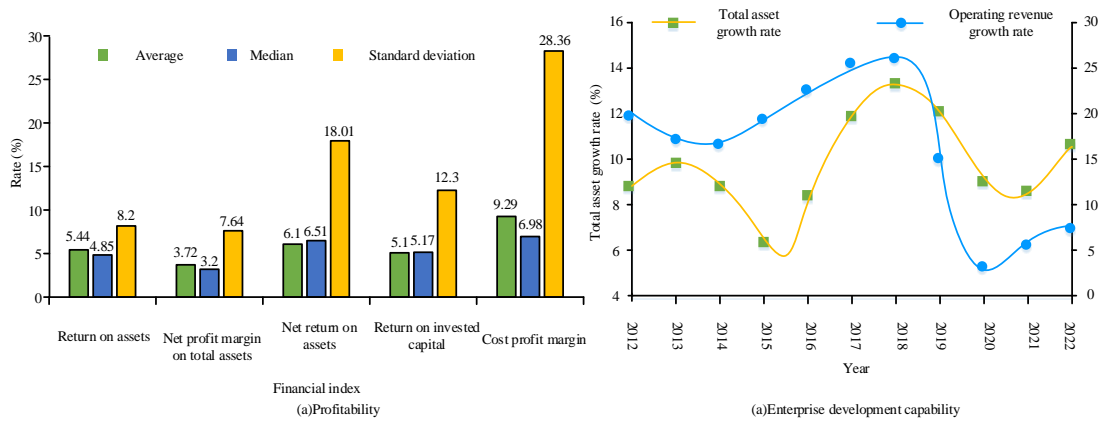

Fig. 8. Correlation matrix.

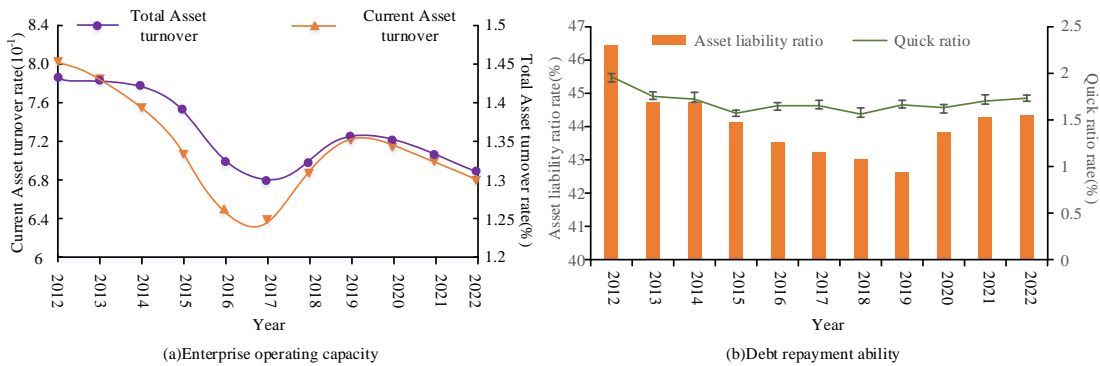Fig. 9.   Profitability and development ability of the enterprise.

Fig. 10. Enterprise operational capability and debt repayment capability.

Fig. 9(a) shows the results of the enterprise profitability assessment. From Fig. 9(a), it can be found that the mean of return on assets, total profitability of assets, return on assets, return on assets and cost margin in enterprise finance are 5.44%, 3.70%, 6.10%, 5.15% and 9.29% respectively, and the standard deviation in enterprise profitability indicators are 8.2%, 7.64%, 18.01%, 12.3% and 28.36%, indicating a high degree of dispersion in the profitability indicators. Fig. 9(b) shows the results of the assessment of the development capability of the enterprise. The development curve of the enterprise fluctuates greatly during the period 2012-2022, with the growth rate of total assets and operating income reaching a peak of 15.4% and 23.2% respectively in 2018; the growth rate of total assets reaches a minimum of 5.3% in 2020; and the growth rate of operating income reaches a minimum of 12.4% in 2021. Fig. 10 shows the results of the evaluation of the company's operating capacity and debt servicing capacity for the period 2012-2022.

Fig. 10(a) shows the analysis of the operating capacity of the enterprise during the period 2012-2022. In 2012, the company's operating capacity reached a maximum with a total asset turnover ratio of 1.46% and a current asset turnover ratio of 0.78%; in 2017, the company's operating capacity reached a minimum with a total asset turnover ratio of 1.29% and a current asset turnover ratio of 0.68%. Fig. 10(b) shows the analysis of corporate debt service capacity results for the period 2012-2022, with the corporate quick ratio reaching a maximum value of 1.95% in 2012 and the quick ratio reaching a minimum value of 1.52% in 2016. In 2019, it reaches a minimum value of 42.9%, with a difference of 3.5%. In order to test the performance of the entropy method performance evaluation model proposed by the study, the study compared the performance of the entropy method financial performance evaluation model with other algorithmic models for performance experiments. The study used the scores of the four dimensions of performance and the absolute difference with the actual value as the comparison index for performance analysis. Fig. 11 shows the performance comparison results between the entropy method performance evaluation model and the comparison model.

(a)Comparison of scores for various methods

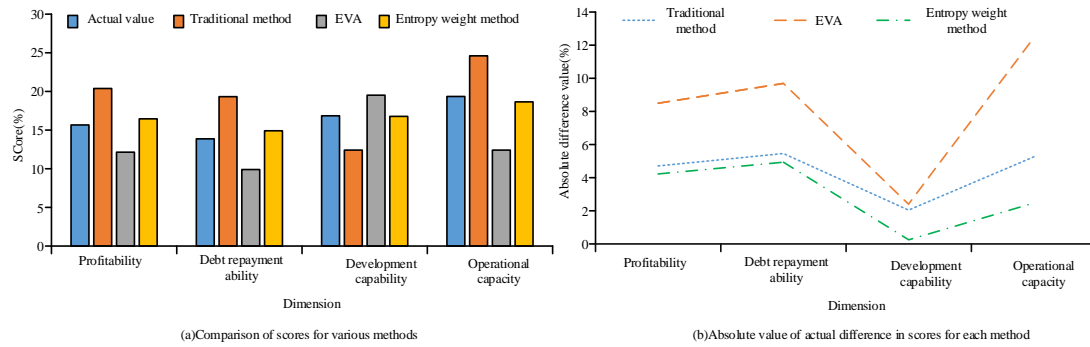(b)Absolute value of actual difference in scores for each method

Fig. 11. Performance comparison results of performance evaluation models.

Fig. 11(a) shows the results of the entropy method of evaluating the financial performance of enterprises, which are the actual value, the traditional model and the Economic Value Added (EVA) algorithm model. As can be seen from Fig. 11(a), the actual scores for profitability, solvency, growth and operational capability are 15.7%, 14.8%, 16.2% and 19.8%. The traditional, EVA and entropy models have profitability scores of 20.5%, 12.6% and 16.4%, respectively; debt servicing scores of 18.6%, 10.6% and 15.9%, respectively; development scores of 13.3%, 18.8% and 15.9%, respectively; and operational capacity scores of 26.7%, 12.7% and 18.3%, respectively. Fig. 11(b) shows the absolute error curves of the performance evaluation models and the actual values. As can be seen from Fig. 11(b), the absolute error curve of the entropy method performance evaluation model is lower than that of the other algorithm models, and the absolute error of the three algorithm models is the lowest in the evaluation of development capability. The absolute error of the entropy method performance evaluation model is 0.3%, which is lower than that of the traditional model (2.4%) and the EVA algorithm (2.6%); the absolute error of the entropy method performance evaluation model is 1.1%, which is lower than that of the traditional model (3.8%) and EVA algorithm (4.2%). The entropy method performance evaluation model is the closest to the actual value and has the best performance.

V. CONCLUSION

Due to the low accuracy of traditional performance evaluation models in assessing enterprise performance, in order to more comprehensively and accurately explore the changes in financial performance of digital enterprises during the period 2012-2022, the study proposes to use a text mining algorithm based on crawler technology to filter the keywords of enterprise annual reports and construct a list of digital enterprises. The financial performance data of the selected enterprises were input into the financial performance evaluation model proposed in the study, and it was found that the digital enterprises had the best growth capacity in 2018, with the annual growth rate of total assets and operating revenue reaching the peak of 15.4% and 23.2%, and the maximum operating capacity in 2012, with the total assets turnover ratio of 1.46% and current assets turnover ratio of 0.78%. A performance comparison experiment of the text mining algorithm revealed that the area under the accuracy-recall line of the algorithm was 0.83 and the mean value of F-value was 0.34, which was better than other

algorithms. The entropy method financial performance evaluation model has a minimum absolute error of 0.3% and a maximum of 1.1% from the actual value, and its absolute error is lower than the traditional model absolute error of 3.8% and the EVA algorithm absolute error of 4.2% at its maximum, and its performance is optimal. The above results indicate that the model proposed in the study has good evaluation effect and has good potential for application in enterprise performance evaluation. The shortcoming of this study is that the model proposed in this study has a narrow scope of application, which can only be applied to the period of enterprise digital transformation, and cannot accurately evaluate the financial status of enterprises before and after the transformation. The subsequent research direction is to improve the scope of application of enterprise financial evaluation model.

VI. FINDINGS

REFERENCES

[1] Mosiagin I, Pallitsch K, Klose I, Preinfalk A, Maulide N. As Similar As Possible, As Different As Necessary-On-Site Laboratory Teaching during the Journal of Chemical Education, 2021, 10(98):3143-3152.

[2] Lin C H, Wu J X, Hsu J C, Chen P Y, Pai N S, Lai H Y. Tremor Class Scaling for Parkinson Disease Patients Using an Array X-Band Microwave Doppler-Based Upper Limb Movement Quantizer. IEEE sensors journal, 2021, 21(19):21473-21485.

[3] Upadhyay S, Weech-Maldonado R, Opoku-Agyeman W. Hospital Cultural Competency Leadership and Training is Associated with Better Financial Journal of healthcare management / American College of Healthcare Executives, 2022, 67(3):149-161.

[4] Sharma H, Hefele J, Xu L, Conkling B, Wang X. First Year of Skilled Nursing Facility Value-based Purchasing Program Penalizes Facilities With Poorer Financial Performance. Medical care, 2021, 59(12):1099-1106.

[5] Chinyamurindi W, Kyogabiirwe J B, Kabagabe J B, Mafabi S, Dywili M. Antecedents of small business financial performance: the role of human resource management practices and strategy. Employee Relations, 2021, 5(43):1214-1231.

[6] Burch K, Burch B D. Activities of Daily Living Performance in Persons With Dementia. Alzheimer disease and associated disorders, 2021, 35(2):153-. 159.

[7] Xu L, Li Z. Amman H. A New Appraisal Model of Second-Hand

Housing Prices in China First-Tier Cities Based on Machine Learning Algorithms. Computational Economics, 2021, 57(2):617-637.

[8] Fanelli S, Salvatore F P. Indicators and criteria for efficiency and quality in public hospitals: a performance evaluation model [J]. Global Business and Economics Review, 2021, 25(3/4):212-230.

[9] Karimi H, Nikkhah-Farkhani Z. Performance Appraisal of Knowledge Workers Using Augmented Additive Ratio Assessment (A-ARAS) Method: A Case Study. IEEE Transactions on Engineering Management, 2020, 69(5):2285-2295

[10] Galagedera D, Fukuyama H, Watson J, Tan E. Do mutual fund managers earn their fees? New measures for performance appraisal. European Journal of European Journal of Operational Research, 2020, 287(2):653-667.

[11] Akundi A, Mondragon O. Model based systems engineering-A text mining based structured comprehensive overview. systems Engineering , 2022, 25(1):51-67.

[12] Leem B, Eum S. Using text mining to measure mobile banking service quality. Industrial Management & Data Systems, 2021, 121(5):993-1007.

[13] Zhou J, Liu D, Ye M, Liu Z. Data-Driven Prediction of Minimum Fluidization Velocity in Gas-Fluidized Beds Using Data Extracted by Text Mining. Industrial & Engineering Chemistry Research, 2021, 60(37):13727-13739.

[14] Cheng F, Li H, Brooks B, You J. Signposts for Aquatic Toxicity Evaluation in China: Text Mining using Event-Driven Taxonomy within and among Regions. Environmental Science and Technology, 2021, 55(13):8977-8986.

[15] Lim M, Li Y, Song X. Exploring customer satisfaction in cold chain logistics using a text mining approach. industrial management & data systems. 2021, 121(12):2426-2449.

[16] Tran M, Panchal S, Chauhan V, Mevawalla A, Fraser R, Fowler M. Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion batteries. Research, 2022, 46(2):786-794.

[17] Bauer M, Lee W, Papadakis M, Zalewski M, Dubey A. Supercomputing in Python With Legate. Computing in Science and Engineering, 2021, 23(4):73-79.

[18] X Lü, Deng R, Li X, Wu Y. Comprehensive performance evaluation and optimization of hybrid power robot based on proton exchange membrane fuel cell. International Journal of Energy Research, 2022, 46(2):1934-1950.

[19] Vakamalla T R, Mangadoddy N. Comprehensive Dense Slurry CFD Model for Performance Evaluation of Industrial Hydrocyclones. Industrial & Engineering Chemistry Research, 2021, 33(60):12403-12418.

[20] Xue W, Chen B, Hong D, Yu J, Liu G. Research on the Comprehensive Evaluation Method for the Automatic Recognition of Raman Spectrum under Analytical chemistry, 2022, 21(94):7628-7636.