Article

Financial Risk Assessment of Photovoltaic Industry Listed Companies Based on Text Mining

Yuanying Chi, Mingjian Yan, Yuexia Pang * and Hongbo Lei

Economics and Management School, Beijing University of Technology, Beijing 100124, China
* Correspondence: pangyuexia@bjut.edu.cn

Abstract: At present, the research on photovoltaic companies’ financial risk early warning model mainly focuses on financial indicators and non-financial indicators from corporate governance structure and external audit opinions. There are few literature studies on the companies’ internal information from their annual report. To solve the above problem, firstly, this paper aims to establish a comprehensive assessment indicators system including financial and non-financial indicators considering the companies’ internal information. Secondly, this paper uses text mining and a binary logistic regression model to evaluate the financial risk for 37 listed companies in the photovoltaic industry. The results showed that profitability was the most significant factor. Probability, as well as negative sentiment ratios, are both negatively correlated with the occurrence of financial risk, while development capability is positively associated with financial risk. These findings can be used as an effective supplement for financial risk evaluation in the photovoltaic industry and provide reference strategies for developing listed companies in the photovoltaic industry.

Keywords: renewable energy; photovoltaic company; financial risk; logistic regression; text mining

1. Introduction

On 22 September 2020, carbon peaking and carbon neutrality goals were proposed at the general debate of the 75th United Nations General Assembly. Solar cells that convert sunlight into electricity have obtained increasing interest from academia and the market in recent years, since they represent the most important way to use renewable energy to meet the development of society and relieve the pressure of environmental pollution [1–6]. Recently, the attention to the photovoltaic (PV) industry has begun to increase rapidly, and photovoltaic-related funds have been issued one after another. Policy banks and commercial banks have also given preference to renewable energy industries. More and more listed companies in the photovoltaic industry have emerged and expanded production and operations by selling stocks and bonds. However, photovoltaic listed companies still have many hidden financial dangers. From an industry perspective, compared with traditional industries, photovoltaic listed companies have higher capital and technology-intensive levels, slow asset turnover, and a long period of construction and financing cycles. On the other hand, photovoltaic listed companies are highly sensitive to policies, such as subsidies. As an emerging industry, policy changes and adjustments have a greater impact on the operating conditions of photovoltaic companies, which will lead to financial risks. Therefore, it is of great significance to construct a financial risk evaluation system for photovoltaic companies [7].

The companies’ financial status is particularly important for listed companies and is related to the fluctuations of each company’s stock prices and bond prices. Establishing an accurate and effective financial early warning model helps investors or financial institutions to discover the financial risks, and prevent them from causing greater losses. Additionally, it will help the company carry out risk prevention and control internally,
review the management’s decision-making level internally, and avoid financial crises caused by decision-making mistakes. It can be traced back to Fitzpatrick’s financial early warning model [8]. By summarizing the financial characteristics of bankrupt and normal operating companies, Beaver [9] proposed a univariate early warning model and found that the use of property rights ratio and return on net assets can more accurately judge the occurrence of financial risks. Based on the in-depth study of previous models, it is found that cash flow/total liabilities and net profit/total assets have better financial forecasting accuracy. Subsequently, Altman [10] used a multivariate evaluation model for the first time. The Z-score model is established for the five indicators: capital/total assets, retained earnings/total assets, EBIT/total assets, preferred stock plus common stock market value/total liabilities, and sales/total assets. Moreover, the bankruptcy probability of the company is judged according to the final calculation result of the Z score. To improve the models’ accuracy, Altman [11] and other scholars expanded the five variables to seven variables based on the Z-score model, and the ZETA model is established on the asset rate of return, income stability, solvency, profit accumulation ability, liquidity, capitalization, and scale. Due to the limitations of univariate and multivariate model assumptions, Ohlson [12] used the multiple logistic regression method to conduct empirical analysis and compared the results with other models. He believed that the logistic model prediction was more accurate. In recent years, Chinese scholars have comprehensively used the logistic model to study and continuously expanded and improved the logistic model to improve the accuracy of financial risk prediction. Gu et al. [13] and others introduced accrual earnings management variables and real earnings management variables into the logistic financial early warning model. Additionally, it can improve the accuracy of early warning for firms with financial difficulties and avoid erroneous judgments brought by financial information distortion through the comparative analysis of special treatment companies and normal operating companies. Pan [14] selected firms that issued bonds and defaulted in China’s bond market as samples, then used the logistic model to predict the default risk of corporate debt using financial indicators. Given the low significance of the financial variables of small and micro enterprises on their loan defaults, He et al. used non-financial indicators as variables to establish a credit default early warning model for small and micro enterprises [15]. Yang et al. [16] constructed the Benford factor to reflect the financial data quality indicators. The introduction of the Benford factor can improve the model prediction accuracy and extend to other situations such as risk assessment and default forecasting. Based on the four leading indicators of profitability, solvency, operating ability and development ability, Chen and Guo [17] introduced profit quality and market valuation indicators, and jointly used factor analysis and logistic regression to predict the financial risks of listed agricultural companies. Wang and Zhang [18] reconstructed financial indicators from the level of capital stock and capital flow and established a logistic early warning model for financial distress. However, Tarighi et al. [19] used the logistic model to investigate the relationship between corporate social responsibility disclosure and distressed financial risk. The results showed that a high level of exposure may downgrade the firm’s credit rating, resulting in financial distress easily. At present, scholars have used neural network models to predict financial risks. Yang et al. [20] used Benford’s law and Myer index to represent data quality and construct the BM-BP neural network financial risk early warning model. By introducing a machine-learning forest into the framework of evidence, Zhu et al. [21] proposed an evidence theory-random forest (DS-RF) model with four dimensions of profitability, asset quality, debt risk, and operating growth. Wu et al. [22] presented a crisis early warning model that combines a multi-layer perceptron artificial neural network (MLP-ANN) with the traditional Altman Z-Score model, and the hybrid model shows a better performance than either pure neural network method or Altman Z-Score model.

With the continuous advancement of risk forecasting, scholars have gradually found that the whitewashing of corporate financial data leads to a particular bias in using financial data to predict future risks. It is impossible to predict risks more accurately if limited
to the estimation of financial indicators. Some scholars have found that text mining can effectively explore non-financial information and is essential in supporting financial forecasting, corporate financing, and technological innovation [23]. Text mining is applied to the analysis of non-financial information by scholars. Song and Huang [24] used text mining tools to study the information disclosure content of the notes to the financial statements. It is conducive to rapidly extracting crucial information in the annual report. It plays an important role in supplementing and explaining the financial statement analysis’s reliability, accuracy, and comprehensiveness. Decomposing corporate social responsibility (CSR) into numerous dimensions by the linear discriminant analysis method and text mining technique, Lee and Huang revealed that a specific CSR indicator shows the degree of corporate operating performance, regarded as the crisis warning signal before financial distress occurrence [25]. Based on the theoretical framework of the TEI@I methodology, Xiao et al. [26] used integrated text mining and deep learning to construct an early warning model for corporate financial risks and introduced non-financial indicators such as corporate governance structure. Due to the disadvantages of hysteresis and indirection in the estimation of quantitative indicators, Ouyang et al. [27] combined the analysis of online public opinion texts and proposed the Attention-LSTM neural network model to build an early warning model, which has a higher accuracy rate for systemic risk indicators compared to SVR and ARIMA. Li et al. [28] constructed a sentiment dictionary including capital market, stock market, and the company’s internal environment and politics, and predicted the company’s future financial distress through sentiment analysis of the company’s annual report. Scholars such as Jiang et al. [29] proposed a semantic feature extraction method based on word embedding technology to predict the financial distress of non-listed listed companies, and provided analysis help for stakeholders. Li et al. [30] used text mining tools from financial reports to identify and mine the risks facing the insurance industry in the future. Yang et al. [31] researched the market sentiment information of listed companies, measured the risk expectation index of listed companies, and conduct empirical research on the impact mechanism of corporate investment strategy choice and future risk expectations. Their findings aid decision-making between physical investments and financial assets when companies monitor financial risks.

With the promotion of the Chinese carbon peaking and carbon neutrality goals, some scholars have begun to use financial models for photovoltaic companies to make corresponding risk predictions. Luo [32] took 45 A-share solar photo-voltaic companies as the research object and used the logistic model for risk early warning prediction. Dong [33] used EVA theory to optimize some indicators based on solvency, profitability, and other indicators and calculated the probability of a financial crisis for 14 photovoltaic companies. Lu (2019) [34] used the Pearson correlation analysis method to screen the company’s financial indicators and build a financial risk early warning indicator system. Some scholars have begun to introduce machine learning methods for in-depth research. Sun [35] used the OLAP method to process financial data and analyze company financial risks from four dimensions: debt repayment, operation, profitability, and payment.

To sum up, photovoltaic companies are characterized by more considerable initial investments with a longer payback duration, are highly dependent on subsidies, and have poor capital liquidity. These features may lead to substantial financial danger to firms. According to the current subsidy situation, the gradual reduction of financial support will have an enormous impact on the PV industry, creating great uncertainty about the future development of the PV industry. Predicting the financial risk for photovoltaic firms seems necessary not only in the degree of firms and investors, but also for regulators.

Identifying and predicting corporate financial risks is conducive to maintaining investors’ income levels and the company’s sustainable operations. The company financial risk model began with the initial univariate early warning model, then evolved from the multivariate early warning model to the multivariate logistic model. It has grown with deep learning, logistic methods, and neural network models. Evaluation indicators have also evolved from a single financial indicator to multiple financial indicators and
gradually evolved into a combination of economic indicators and non-financial indicators. Currently, the logistic financial early warning model is relatively mature, and text mining is steadily recognized by everyone in the financial early warning model. It can assist in obtaining information other than financial data and assist in predicting future financial risks of companies. However, the selection of non-financial indicators currently focuses on corporate governance structure and external audit opinions, and less attention was paid to the disclosure of the company’s text information from the annual report text. For the photovoltaic industry, it has characteristics of large initial investment and a long payback duration, which influences business and management status. It is closely related to financial risks. The company’s annual report is a better way to acquire first-hand information about the firm’s yearly operation situation. In addition, some scholars find that text information in the annual report helps predict the occurrence of financial risk. Most of the research is merely based on text information, but few combine that with financial indicators.

This paper aims to construct a reasonable and reliable risk prediction model and hopes it will give some insights to the firms, investors, and regulators in order to prevent risks. The research object of this paper is listed companies in the Chinese PV industry. The structure of this paper is as follows: Section 1 introduces the background and related literature review. Section 2 shows preparation works, including discussing the possible influencing factors and the influence mechanism, the data sources, and the methodologies used in the article. Section 3 presents the two sets of regression results (with non-financial indicators or not). Section 4 discusses the results, and Section 5 is the conclusion.

2. Influencing Factors, Materials, and Methods
2.1. Influencing Factors

A listed company’s continued operating ability is necessary to prevent financial storms. When a company exposes too many financial risks, the company will face delisting problems. According to previous studies [36–38], there are two ways to predict the financial risks of companies. One is financial indicators such as profitability, operating ability, solvency, and development ability. They can be obtained from the financial statements published in the company’s annual report by calculation. The other is the indicators extracted from textual information by text mining. Applications such as text similarity and text sentiment analysis are commonly used in text mining.

Profitability capability reflects company profits ability from its production and operation. Strong profitability guarantees the company’s continued operation, and protects the company against financial risk losses. With the rapid expansion of the solar industry, the price of photovoltaic products has gradually declined, because of intensified industrial competition. Strong profitability ensures that the company avoids being eliminated in the competitive market. Profitability indicators mainly include operating profit rate, cost-profit rate, surplus cash guarantee multiple, return on total assets, return on equity and return on capital, etc.

Operational capability reflects the operational efficiency and capital recovery capability of a company. For instance, with the rapid growth of the company’s scale, the company’s raw material procurement and warehouse-scale, semi-finished products, and inventories are increased. Currently, the speed of component transportation between companies has slowed, and the inventory balance level has accumulated, affecting the company’s capital operation efficiency and cash flow generated by operating activities. Operating ability can be reflected by driving ability indicators such as total assets and inventory turnover rates.

Solvency capability reflects a company’s ability to repay its debts. The solvency is closely related to the company’s debt scale and profitability. Since photovoltaic companies are capital-intensive, they generally have those features, which include large assets and liabilities, average short-term solvency, poor profitability, corporate profitability, and
cash flow. Therefore, the solvency indicator is good for analyzing the financial risk capability of the photovoltaic company. Solvency capability can be reflected by indicators such as the current ratio, cash ratio, and asset-liability ratio.

Development capability mainly measures the ability of a company to expand its business scale and grow stronger. Affected by the Chinese country’s carbon peaking and carbon neutrality goals, new energy industries such as photovoltaics have received more national policy subsidies. The company’s sustainable operation scale is sensitive primarily to those policies. The development indicator mainly measures the potential of a company to expand its production scale through its production and operation activities.

Text similarity reflects the level of inertial disclosure of the corporate financial report. If a listed company has textual inertial disclosure, the similarity of its textual information is high. High text similarity usually implies that the company repeatedly discloses the same information as in previous years, which is suspected of hiding unfavorable information about the company.

Text sentiment analysis analyzes, processes, induces, and reasons on subjective texts with emotional color, and calculates the sentiment information of text sentences. When the analysis results show more negative emotions, it reflects that the management has lower confidence in the current company development and future market, and the company is in danger of facing financial difficulties.

2.2. Data Source

The “Special Treatment During Abnormal Condition of Listed Companies”, promulgated in 1998, stipulates that the listed company shall be given special treatment (ST) when the listed company has an “abnormal condition”. Briefly, the company operating losses for two consecutive years are marked as special treatment. The China Securities Regulatory Commission will add ST, *ST before the stock code of these listed companies with anomalies. Therefore, listed companies with stock codes ST and *ST can be classified as listed companies in risk, and then the information of ST companies and non-ST companies in listed companies can be used to construct a risk prediction model to predict the risk.

According to the scope of existing listed companies, the listed companies classified as the photovoltaic concept of flush are selected as the research objects. Flush data show that, as of August 10, 2022, there are 362 listed companies belonging to the photovoltaic concept, among which are 10 listed companies with risks marked ST. Excluding one firm with incomplete data, the ST sample is finally 9 listed companies. According to Li (2017) [39], the prediction accuracy of binary logistic regression prediction is better when the sample configuration is 1:2 to 1:1. However, the number of samples of risky companies in the photovoltaic industry is limited, so the ratio of samples of risk-free companies is increased. The ratio of 1:3 is used for sample selection. Therefore, 28 risk-free companies were selected as matching samples.

The public time of the t year annual report of listed companies is t + 1 year, so the t – 1 year annual report of listed companies and in t year whether the special treatment of these two events occurs in the same year. Therefore, this paper adopts the financial data and annual report information of listed companies in years t – 2 to establish a model to predict the possibility of whether risk in year t will occur.

According to the 9 existing companies with ST mark, the annual report data and text of the annual report two years ago when they were marked with ST were selected, and the data were selected from 2016 to 2018. Non-ST companies select the data and text information of the 2019 annual report.

The data for this paper comes from the CSMAR database. The text of the annual report is downloaded from the Flush website. The 37 sample companies are finally selected, as shown in Table 1.
Table 1. Sample of listed companies in the photovoltaic industry.

| Code   | Name                                           | Code   | Name                                           | Code   | Name                                           |
|--------|------------------------------------------------|--------|------------------------------------------------|--------|------------------------------------------------|
| 000040 | Tunghsu Azure Renewable Energy Co., Ltd.        | 002684 (ST) | Dynavolt Renewable Energy Technology (Henan) Co., Ltd. | 600509 | Xinjiang Tianfu Energy Co., Ltd.                |
| 000155 | Sichuan New Energy Power Company Limited       | 300029 (ST) | Jiangsu Huasheng Tianlong Photovoltaic Co., Ltd. | 600537 | Eging Photovoltaic Technology Co., Ltd.         |
| 000591 | Cecep Solar Energy Co., Ltd.                   | 300187 | Yonker Environmental Protection Co., Ltd.       | 600703 | Sanan Optoelectronics Co., Ltd.                |
| 000791 | Gepic Energy Development Co., Ltd.             | 600011 | Huaneng Power International, Inc.               | 600736 | Suzhou New District Hi-Tech Industrial Co., Ltd. |
| 000862 | Ning Xia Yin Xing Energy Co., Ltd.             | 600089 | Tbea Co., Ltd.                                  | 600821 (ST) | Nyocor Company Limited                         |
| 000875 | Jilin Electric Power Co., Ltd.                 | 600151 | Shanghai Aerospaceautomobile Electromechanical Co., Ltd. | 601222 | Jiangsu Linyang Energy Co., Ltd.                |
| 002121 (ST) | Shenzhen Clou Electronics Co., Ltd. | 600163 | Zhongmin Energy Co., Ltd.                       | 601619 | Ningxia Jiaze Renewables Corporation Limited   |
| 002255 (ST) | Suzhou Hailu Heavy Industry Co., Ltd. | 600256 | Guanghui Energy Co., Ltd.                       | 601991 | Datang International Power Generation Co., Ltd. |
| 002256 (ST) | Shenzhen Sunrise New Energy Co., Ltd. | 600277 | Elion Energy Company Limited                   | 603105 | Zhejiang Sunoren Solar Technology Co., Ltd.     |
| 002323 (ST) | Shandong Yabo Technology Co., Ltd. | 600310 | Guangxi Guidong Electric Power Co., Ltd.        | 603126 | Sinoma Energy Conservation Ltd.                 |
| 002499 (ST) | Kelin Environmental Protection Equipment, Inc | 600406 | Nari Technology Co., Ltd.                     | 603628 | Clenergy Technology Co., Ltd.                   |
| 002529 (ST) | Jiangxi Haiyuan Composites Technology Co., Ltd. | 600481 | Shuangliang Eco-Energy Systems Co., Ltd.           | 603693 | Jiangsu New Energy Development Co., Ltd.        |
| 002529 (ST) | Beijing Jingyuntong Technology Co., Ltd.       | 601908 | Beijing Jingyuntong Technology Co., Ltd.        | 601908 | Beijing Jingyuntong Technology Co., Ltd.        |

Note: (ST) means a company is marked as special treatment due to the operating losses for two consecutive years. All the companies mentioned in Table 1 are registered in China.

2.3. Method

Based on the availability of financial data and company annual reports, a suitable number of listed companies are selected as research subjects. Combining financial and text disclosure indicators, an assessment indicator framework was established. Due to the excessive number of financial indicators, the original indicators were transformed into five leading indicators by using the principal component analysis (PCA) method to reduce the dimensionality. Text similarity and sentiment values were calculated using text mining techniques. To verify that adding text disclosure indicators will improve the prediction effect, financial indicators are used first in the logistic regression. Then, the regression will be combined with financial and text disclosure indicators. The results of the two regressions will be compared and analyzed to verify the validity of text disclosure indicators. The whole empirical research procedure is shown in Figure 1.
Figure 1. The empirical research procedures.

First, financial indicators are selected to construct a comprehensive assessment model. Based on the research of Zhou et al. [40], the four first-level indicators of operation management capability, profitability, solvency, and development capability are summarized. Then, 18 secondary financial indicators are selected as the financial indicators of the financial risk assessment model. The corporate financial indicators are shown in Table 2.

Table 2. Financial risk assessment indicators.

| First-Level Indicator               | Secondary Indicators                      | Variable |
|------------------------------------|--------------------------------------------|----------|
| Solvency indicator                 | Current ratio                             | X1       |
|                                    | Quick ratio                               | X2       |
|                                    | Cash ratio                                | X3       |
|                                    | Assets and liabilities                     | X4       |
| Operation management capability indicator | Accounts receivable turnover               | X5       |
|                                    | Inventory turnover                         | X6       |
|                                    | Total asset turnover                       | X7       |
|                                    | Fixed asset growth rate                    | X8       |
|                                    | Total asset growth rate                    | X9       |
| Development capability indicator    | Operating income growth rate               | X10      |
|                                    | Owner’s equity growth rate                 | X11      |
|                                    | Return on assets rate                      | X12      |
| Profitability indicator            | Net profit margin on total assets          | X13      |
Net profit margin of current assets \( X_{14} \)
Return on net assets rate \( X_{15} \)
Return on invested capital rate \( X_{16} \)
Operating gross profit margin rate \( X_{17} \)
Operating net profit margin rate \( X_{18} \)

Source: CSMAR database (http://cndata1.csmar.com/, accessed 12 October 2021).

Then, we screen for key influencing factors using the PCA method. The PCA is a statistical analysis method that divides the original multiple variables into a few comprehensive indicators. Before using PCA to process the original data, it is necessary to test and analyze the sample data for whether they are suitable for PCA extraction. Suppose there are \( n \) samples, and each sample has a total of \( p \) variables, forming a data matrix of \( n \times p \) order. When \( p \) is large, it is more troublesome to investigate the problem in \( p \)-dimensional space. Therefore, it is necessary to carry out dimensionality reduction processing, replacing the original more variable indicators with fewer comprehensive ones. Additionally, these fewer comprehensive indicators reflect the information of initial more variable indicators as much as possible. Moreover, they are independent of each other. This paper will extract the principal components and conduct model construction and analysis based on the indicators in Table 2.

Based on the KMO test, the KMO value is 0.629. According to the standard given by Kaiser, if the KMO value is greater than 0.5, it is suitable for PCA. Moreover, according to the Bartlett sphericity test, the associated probability is 0.000, which is less than the significance level of 0.05. Therefore, the Bartlett sphericity test null hypothesis is rejected, and the sample is considered suitable for PCA.

The extraction results of the principal components' common factors of the financial indicators are shown in Table 3. It is clear to see that the cumulative variance of the five principal components reaches 84.010\%, which means that these five main component variables can explain 84.010\% of the information of all indicators.

Table 3. Total variance explanation.

| Component | Initial Eigenvalue | Extract the Sum of Squared Loads | Sum of Squares of Rotational Loads |
|-----------|-------------------|----------------------------------|-----------------------------------|
|           | Total              | Percentage of Variance           | Accumulation\%                    | Total              | Percentage of Variance | Accumulation\% |
| Y_1       | 7.121              | 39.560                           | 39.560                            | 7.121              | 39.560                  | 39.560         | 5.236               | 29.087          | 29.087          |
| Y_2       | 3.622              | 20.124                           | 59.684                            | 3.622              | 20.124                  | 59.684         | 3.117               | 17.319          | 46.406          |
| Y_3       | 1.748              | 9.708                            | 69.392                            | 1.748              | 9.708                   | 69.392         | 2.741               | 15.229          | 61.635          |
| Y_4       | 1.366              | 7.587                            | 76.979                            | 1.366              | 7.587                   | 76.979         | 2.277               | 12.648          | 74.283          |
| Y_5       | 1.266              | 7.031                            | 84.010                            | 1.266              | 7.031                   | 84.010         | 1.751               | 9.727           | 84.010          |

The maximum variance method is adopted to obtain the rotated factor load matrix and transfer the original financial indicators into 5 principal component indicators: \( Y_1, Y_2, Y_3, Y_4, \) and \( Y_5 \). The following conclusions can be drawn:

- The principal component, \( Y_3 \), is mainly explained by the variables \( X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, \) and \( X_{18} \). These variables reflect the growth rate of the owner’s equity, return on assets, net profit margin on total assets, net profit margin on current assets, return on equity, return on invested capital, gross operating profit margin, and net operating profit rate separately. It mainly expresses the profitability of the company.
- The principal component \( Y_2 \) is mainly explained by the variables \( X_{1}, X_{2}, X_{3}, \) and \( X_{4} \), reflecting the current ratio, quick ratio, cash ratio, and asset-liability ratio, primarily reflecting the debt repayment ability of the company.
The principal component, $Y_3$, is mainly explained by variables $X_8$, $X_9$, and $X_{10}$, reflecting the growth rate of total assets, the growth rate of owners’ equity, and the growth rate of fixed assets, and mainly representing the development ability of the company.

The principal component $Y_4$ is mainly explained by the variable $X_5$, which reflects the turnover of accounts receivable and primarily represents the operating ability of the company.

The principal component, $Y_5$, is mainly explained by variables $X_6$ and $X_7$, reflecting inventory turnover and total asset turnover, and mostly representing the company’s operating ability.

Next, we extract the indicators from text disclosure based on text mining. This paper selects the section “Discussion and Analysis of Operation” in the annual report for further analysis. The text content is derived from the annual reports of chosen listed companies from 2017 to 2019 in the photovoltaic concept classification. Before identifying the valuable information from text, the text should be preprocessed through cleaning, segmentation, and deleting stop words to eliminate the structural data. In this paper, we adopt the commonly used method, cosine similarity calculation, to calculate the text similarity calculation (SIM). The larger the value, the more similar the text is. SIM indicator calculating steps are as follows. This paper uses natural language processing for text mining and applies text similarity and sentiment analysis, using the TF-IDF method and the lexicon-based method to calculate the corresponding indicator value, respectively. Firstly, use the TF-IDF (term frequency-inverse document frequency) algorithm to find the keywords in the excerpt part of the reports. Taking out 100 keywords for each annual report and combining them into a set, then calculating the word frequency of each word in this set. Generate the respective word frequency vector for each text,

$$S_i = (w_{i1}, w_{i2}, \ldots w_{in})$$

where $S_i$ denotes a sentence, and $w_{in}$ denotes the frequency of a word in the sentence. The whole text can be transformed into the following representation:

$$T_i = (S_{i1}, S_{i2}, \ldots S_{in})$$

$$T_t = (S_{t1}, S_{t2}, \ldots S_{tn})$$

where $T_i$ denotes the sample text, $T_t$ denotes the target text.

Then, calculate the text sentiment value by the lexicon-based method. The annual report text’s negative sentiment ratio (OP) is used as a sentiment analysis indicator. In this paper, the method of calculating the sentiment value of annual reports is chosen based on the sentiment dictionary (BosonNLP Sentiment Dictionary). If the calculation result score is less than 0, it means the sentiment tendency of the text is negative, while vice versa, greater than 0 means positive. The sentiment words, negative words, and degree adverbs in each sentence were first identified. Then, determine whether each emotion word is preceded by a negation word and an extent adverb, and divide the negation words and extent adverbs before it into a group. If there is a negation, the sentiment weight of the sentiment word is multiplied by $-1$. Or, if there is a degree adverb, it should multiply by the degree value of the last degree adverb to the score. Greater than 0 is positive, and less than 0 is negative (the absolute magnitude of the score reflects the degree of positivity or negativity). The sentiment positive and negative scores of all clauses were accumulated, and the sentiment degree percentage of the annual report text was obtained.

$$WS_i = W_i \times \text{Deny}_i \times \text{Degree}_i$$

$$S_i = \sum (W_{i1}, W_{i2}, \ldots W_{in})$$

$$Po = \sum S_{iv}, \; (WS_i > 0); \; Ne = \sum S_{iv}, \; (WS_i < 0)$$

$$OP = \frac{Ne}{Po + Ne}$$
In Equations (4)–(7), \( WS_i \) denotes the value of the sentiment word weighting score, \( Wi \) denotes the score of the \( i \)th sentiment word in the sentence, \( Deny_i \) denotes the negation before the sentiment word, default 1 for no negative words; otherwise, \(-1\). \( Degree_i \) denotes the degree adverb weighting value before the sentiment word, default 1 for no degree word. \( St \) denotes the score of the sentence. \( Po \) and \( Ne \) indicate the sentence’s positive and negative scores, respectively. \( St \) indicates the score of the sentence, and \( OP \) identifies the negative percentage of the total report textual content. After selecting financial indicators from financial statements and extracting text disclosure indicators from annual reports, a financial risk assessment indicator system for listed companies in the photovoltaic industry is constructed.

Based on the sentiment analysis of the text of the annual reports of the sample companies, the positive sentiment value (Po) and the negative sentiment value (Ne) can be obtained. In this article, the magnitude of Po represents the positive degree of the company’s management on the operation and future development, and the volume of Ne represents the negative degree. Since the length of the text presented by each annual report varies, the OP indicator can be used to reduce errors since it calculates the percentage instead of the absolute value. The integrated assessment indicators are shown in Table 4.

Table 4. Financial risk assessment indicators combining mining text and financial indicators.

| First-Level Indicator | Secondary Indicators | Variable | Explanation |
|-----------------------|----------------------|----------|-------------|
| Financial indicator   | Profitability        | \( Y_1 \) | \( X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, \) and \( X_{18} \) |
| Financial indicator   | Solvency             | \( Y_2 \) | \( X_1, X_2, X_3, X_4 \) |
| Financial indicator   | Development          | \( Y_3 \) | \( X_5, X_6, X_7, X_8, X_9 \) |
| Financial indicator   | Management capacity   | \( Y_4 \) | \( X_9 \) |
| Financial indicator   | Management capacity   | \( Y_5 \) | \( X_6 \) and \( X_7 \) |
| Text Disclosure indicator | Text similarity   | SIM      | |
| Text Disclosure indicator | Text negativity   | OP       | Po and Ne take negative percentages |

SPSS (software version is SPSS 26, Chicago, IL, USA) is used as an analysis tool to construct a logistic regression financial risk prediction model. The logistic regression model built in this paper does not include the constant term, which identifies the ratio of the probability that \( ST \) is equal to 1 and \( ST \) is equal to 0 under the baseline state when all independent variables have 0 values. In different studies, the specific meaning of the constant term may be different. In this paper, the constant term is not considered to have practical significance, so the model is constructed without the constant term.

3. Results

3.1. Prediction Based on Financial Indicators

To verify that combining textual disclosure indicators is more effective than using only financial indicators for forecasting, logistic regression is run first using only financial indicators.

According to the results of the PCA, this paper selected five principal components \( Y_1–Y_5 \) extracted from 18 financial indicators as the variables of the logistic regression model based on financial indicators.

The goodness-of-fit test of the regression equation passed the Hosmer–Lemeshow Test, and the significance of the Hosmer–Lemeshow Test was 1, greater than 0.05, indicating that the regression equation had a good degree of fit. That is, the binary logistic regression model could well reflect the relationship between variables.
The financial risk early warning model of listed companies in the photovoltaic industry was constructed by the logistic regression method based on financial indicators. The prediction accuracy is 100%, and the results are shown in Table 5.

Table 5. Financial risk prediction based on financial indicators.

| Items       | Correct | Error | Accuracy of Prediction |
|-------------|---------|-------|------------------------|
| ST marked   |          |       |                        |
| Non-risk    | 28      | 0     | 100                    |
| Risk        | 0       | 28    | 100                    |
| Total percentage |       |       | 100                    |

According to the score coefficient of the regression factor, the variables of the risk model equation are shown in Table 6.

Table 6. Variables in model based on financial indicators.

| B       | Standard Error | Wald | DOF | Significance | Exp(B) |
|---------|----------------|------|-----|--------------|--------|
| Y₁      | -115.087       | 0.001| 1   | 0.981        | 0.000  |
| Y₂      | -193.120       | 0.001| 1   | 0.982        | 0.000  |
| Y₃      | 113.822        | 0.001| 1   | 0.981        | 2.707 × 10^11 |
| Y₄      | -500.615       | 0.001| 1   | 0.981        | 0.000  |
| Y₅      | 38.423         | 0.001| 1   | 0.982        | 48,650,255,124,666,608.000 |

B is the regression coefficient of each variable, and the expression of the logistic regression model of financial indicators can be obtained as Equation (8).

$$P = \frac{1}{1 + e^{-(-115.087Y_1-193.120Y_2+113.822Y_3-500.615Y_4+38.423Y_5)}}$$

The results show that although the prediction accuracy is high, the p-value (significance column in Table 6) of the principal components indicators Y₁, Y₂, Y₃, Y₄, and Y₅ are all far greater than 0.05. Under the significance text with 95% confidence level, the coefficient of these variables cannot pass the test. It means using the prediction based on five principal components cannot be used directly in the model. More variables need to be considered and the model should be revised. The regression with both financial indicators and text disclosure indicators would be necessary.

3.2. Prediction Based on Financial Indicators and Text Disclosure Indicators

Since the five principle component indicators are not significant, text disclosure indicators such as OP and SIM are taken into account in the model. Based on the significant text with a 95% confidence level, the explanatory variable with the Y₁ + Y₃ + OP is finally selected to build the model.

The goodness-of-fit test of the regression equation passed the Hosmer–Lemeshow Test, and the significance of the Hosmer–Lemeshow Test was 0.882, greater than 0.05, indicating that the regression equation had a good fitting degree. The binary logistic regression model could very well reflect the relationship between variables. As shown in Table 7.

Table 7. Hosmer–Lemeshow Test.

| Chi-Square | DOF | Significance |
|------------|-----|--------------|
| 3.037      | 7   | 0.882        |

The prediction accuracy of the financial risk early warning model of listed companies in the photovoltaic industry constructed by integrating text mining and the logistic regression method of financial indicators is 91.9%, as shown in Table 8.
Table 8. Risk prediction based on financial indicators and text disclosure indicators.

| Items      | Correct | Error | Accuracy of Prediction |
|------------|---------|-------|------------------------|
| ST marked  | Non-risk| 27    | 1                      | 96.4                   |
| Risk       |         | 2     | 7                      | 77.8                   |
| Total percentage |       |        |                        | 91.9                   |

According to the score coefficient of the regression factor, the variables of the risk model equation are shown in Table 9.

Table 9. Variables in model based on financial indicators and text disclosure indicators.

| Indicator | B       | Standard Error | Wald  | DOF | Significance | Exp(B) |
|-----------|---------|----------------|-------|-----|--------------|--------|
| $Y_1$     | -6.538  | 2.727          | 5.746 | 1   | 0.017        | 0.001  |
| $Y_3$     | 1.726   | 0.743          | 5.398 | 1   | 0.020        | 5.618  |
| OP        | -13.070 | 6.058          | 4.654 | 1   | 0.031        | 0.000  |

$B$ is the regression coefficient of each variable, and the expression of the logistic regression model of financial indicators can be obtained as Equation (9).

$$P = \frac{1}{1 + e^{-(6.538Y_1 + 1.726Y_3 - 13.070OP)}}$$  \hspace{1cm} (9)

The results show that the significance of variable indicators is all less than 0.05, indicating that the represented $Y_1$ profitability indicator, $Y_3$ development indicator, and OP text negative sentiment ratio have strong significance on the occurrence of marked ST in listed companies in the photovoltaic industry. $Y_1$ profitability has the highest significance, with the most substantial explanatory power. It is considered that these three indicators are the most important influencing factors of ST in listed companies in the photovoltaic industry.

It can be seen from the results that the coefficient of company profitability $Y_1$ is estimated at $-6.538$. When $Y_1$ increases by one unit, the OR value becomes the original $e^{-6.538}$ time. That is, the value of Exp(B) is 0.001. Therefore, the higher the value of the company profitability indicator, the smaller the possibility of financial risk, which is a negative impact indicator and a protective factor.

According to the results, the estimated value of the coefficient of company development ability $Y_3$ is $1.726$. When $Y_3$ increases by one unit, it becomes $e^{1.726}$ times the original value. That is, the value of Exp(B) is 5.618. Therefore, the greater the value of the company’s development ability indicator, the greater the possibility of financial risk, which is a positive impact indicator and a risk factor.

In addition, the estimated coefficient of the negative sentiment ratio OP in the company’s annual report is $-13.070$. Whenever OP increases by one unit, the OR value becomes the original value’s $e^{-13.070}$ time. The higher the negative emotion ratio indicator, the lower the possibility of financial risk for the company, which is a negative influence indicator and belongs to the protective factor.

4. Discussion

4.1. Model Analysis

The empirical analysis model shows that although the prediction accuracy of the model constructed by selecting the financial principal component indicator is high, the coefficient is not significant. However, that problem was solved when we used the financial indicators combined with the text disclosure indicator OP. The significant test could be passed using $Y_1 + Y_3 + OP$ as the explanatory variable, meaning the variables have practical meanings. In other words, combining financial indicators with text disclosure indicators has a better fitting result. The results show that profitability and negative text
sentiment ratio belong to negative correlation protective factors. The stronger the indicator ability, the photovoltaic listed firms have a lower possibility of financial risk. Development ability is a positively related risk factor. The stronger its indicator ability is, the higher the possibility of financial risk for listed companies in the photovoltaic industry. Among them, profitability has a strong significance to the occurrence of ST of listed companies in the photovoltaic industry, and it is the main factor in predicting the possibility of financial risk under this model. Development ability and company annual report text emotional negative ratio of significance is relatively weak, but still has a strong explanatory power.

The coefficients of the business ability indicator, profitability indicator, and text similarity indicator selected are not significant and will be eliminated in the subsequent model construction of the comprehensive text indicator. After removing the insignificant financial principal component indicator and adding the text similarity indicator and text sentiment indicator, the \( Y_1 + Y_3 + \text{OP} \) combination indicator with a good fitting degree and good model significance was finally compared to construct the model. The prediction accuracy of the model reaches 91.9%. Additionally, the significance of the Hosmer–Lemeshow test of the model is 0.8882, greater than 0.05, indicating a good degree of fit.

The results show that the significance of all indicators is less than 0.05, which will impact the future occurrence of ST in listed companies in the photovoltaic industry. The profitability indicator is a protective factor of negative correlation. It implies that the greater the indicator value, the lower the possibility of financial risk of listed companies in the photovoltaic industry, which plays a decisive factor. Negative text sentiment ratio and development ability are positively correlated risk factors. The higher the indicator value, the higher the possibility of financial risk for listed companies in the photovoltaic industry. The significance of the indicator is 0.020 and 0.031, respectively, which also plays a role in predicting the possibility of financial risk as a secondary influencing factor.

4.2. Results Analysis

We integrated text disclosure indicators and financial indicators in the empirical study to build a logistic regression prediction model. The results show that the financial indicators of company profitability indicators are the most important protective factor, while corporate debt repayment ability indicator is a secondary risk factor. The indicator of text disclosure is the most important risk factor, and the indicator of text similarity is the second risk factor.

As the most important protective factor indicator, profitability indicators in this paper include return on assets, net profit margin on total assets, net profit margin on current assets, return on equity, return on invested capital, operating gross profit margin, operating net profit rate and owner’s equity growth rate. The empirical results show that the improvement of profitability of listed companies in the photovoltaic industry can significantly reduce the probability of ST. It shows that the profitability of listed companies in the photovoltaic industry is a core ability in the operation process and is also the most important indicator of preventing the occurrence of financial risks for listed companies in the photovoltaic industry. Therefore, the improvement of profitability indicators is a powerful guarantee for the rapid development of listed companies in the photovoltaic industry.

As the most important protective factor indicator, the negative sentiment ratio of the annual report text is the protective factor. The annual report text adopts the sentiment analysis method based on the sentiment dictionary, carries on the sentiment polarity analysis to the sample, and obtains the negative emotion tendency ratio of the annual report. The results show that in the process of financial risk prediction, the emotional information in the text of the annual report has more realistic feedback on the company’s current situation. The management and the board of directors do not embellish too many words in the statements to clarify their operating conditions, avoiding the transmission of false information to investment institutions and investors to prevent the occurrence of information asymmetry. The
improvement of the company’s operating conditions is more likely to achieve a turnaround in subsequent years. The company is likely to turn a profit in the following years and realize the removal of the ST cap.

Development ability is a risk factor. In this paper, the indicators of development ability include the growth rate of fixed assets, the growth rate of total assets, and the growth rate of operating income. The empirical results show that the increase in the development ability indicator of listed companies in the photovoltaic industry will also increase the probability of ST. This indicates that the photovoltaic industry listed companies should first have a reasonable use of corporate assets. A high growth rate means that the company’s fixed assets investment grows too fast, leading to less liquidity and defective equipment, and turnover. It certainly will increase the incidence of ST, reasonable control of asset growth in a proper scope, and risk aversion.

5. Conclusions

According to the regression results, we find, through using 37 photovoltaic firms and the appropriate method, that taking text disclosure indicators into account is more effective than using financial indicators to predict risk solely. Additionally, profitability and the negative sentiment ratio have a negative inhibitory effect on the occurrence probability of future ST, and a higher value will reduce the future ST probability. Development ability has a positive promotion effect on the occurrence probability of ST in the future, and a higher value will lead to a higher occurrence probability of ST. This model fully combines financial indicators and text information content, which has reference significance for the future financial risk assessment of the company. However, there are some shortcomings in this paper. Firstly, the number of experimental subjects selected is somewhat inadequate due to the small number of companies meeting the requirements. Secondly, the experiment aims at whether the company’s annual report information can improve the company’s financial risk prediction without extracting textual information from the content of news, announcements, etc. Additionally, text indicators were only selected for text similarity and text sentiment values, and further exploration of usable indicators is needed.

According to the results of the model construction in this paper, investors or investment institutions should pay attention to the company’s financial indicators and focus on the company’s regular disclosure of annual reports when judging the company’s development status and the probability of financial risks. When reading the company’s annual report documents, it is necessary to have the ability to identify the information behind the text, to avoid being influenced by the beautification of the company’s annual report information on the company’s future risk, valuation, operation, and other basic information judgment. Therefore, it is increasingly important to pay attention to the trend of textual information in annual reports. Investors should strengthen the analysis of the annual report of listed companies, which covers the management and operation current situation of the development of evaluation on the company and the company’s future development outlook, focusing on financial condition, based on deep analysis combining the annual report text information for making scientific investment decisions.

At the same time, the empirical results show that the photovoltaic industry listed company’s profitability is negatively related to the photovoltaic industry listed company’s financial risk factors. The development ability of photovoltaic industry listed companies’ financial risk is related to factors, suggesting that investors should be strengthened for the photovoltaic industry according to the analysis of the financial statements of listed companies, the profitability of listed companies’ focus on photovoltaic industry, and a good profit level ensures the sustainable development of the company, has a strong profit guarantee in risk aversion, and has a solid ability to resist risks.

For listed companies in the PV industry, profitability ability, development ability, and text negativity should be focused on, which are the key factors to avoid financial warnings. It can be examined from the following perspectives:
From the aspect of scientific research, photovoltaic listed companies can adjust their product structure according to the changes in market demand, according to the national development situation of new energy, increase research and development efforts, and develop innovative products with independent and independent property rights to meet market demand. At the same time, photovoltaic technology updates quickly and requires high levels of research personnel, which is the key to improving the innovation capacity of photovoltaic enterprises. This requires PV-listed companies to attract excellent researchers with a good research environment, advanced equipment, and high salaries. This will enable the enterprises to occupy the market share, promote industry progress and reduce business risks.

In the process of production and operation, uncontrolled and blind expansion to capture the market share is likely to accelerate the bankruptcy process before the company enters the maturity cycle. It is important to expand business operations gradually and seek professional advice from finance professionals when making investment decisions to ensure the liquidity and sustainability of the company’s cash. When selecting investment options, focus on opportunity costs and the time value of money, and weigh investment options holistically to find the most profitable use for limited funds. At the same time, managers should keep track of the real-time trend of corporate funds and adjust investment plans on time according to the actual situation. When it is found that the cash flow of the enterprise is about to break, timely and effective find means to raise funds to relieve the pressure of capital.

For investors, ignoring the textual information of annual reports and focusing only on company fundamentals and share prices can cause certain investment risks. Investors should pay attention to the management’s attitude towards the company’s current operating status and future planning in the company’s disclosure of annual reports and quarterly reports to judge whether the company is currently in operational difficulties and make scientific investment decisions. Secondly, for financial data, focus on the sub-indicators contained under profitability and development capacity to improve investment efficiency and reduce the interference of other indicators.

For the regulator, there may be limitations in judging whether the formula is to be ST only through financial data. At present, the exchange judges whether a listed company should be given special treatment based on the following two conditions: first, the net profit of the listed company for both audited fiscal years is negative; second, the audited net assets per share of the listed company for the latest fiscal year is lower than the par value of the stock. Neither of them take into account the factors that affect non-financial information. The exchange should strengthen the audit of listed companies’ annual and quarterly reports to detect financial fraud and financial malpractice at an early stage and ensure that companies make their company operation information open and transparent. Supervisory authorities can quantify the annual and quarterly reports of companies, extract information on key indicators, and discover clues from the text review of annual reports as early as possible to avoid the occurrence of systemic financial risks caused by corporate chains. Therefore, the relevant departments should strengthen the regulation of the text format of annual reports and supervise the disclosure content of the texts of annual and quarterly reports of listed companies, which is conducive to improving the overall quality of financial reports and perfecting the exchange supervision system.

Author Contributions: Conceptualization, Y.P.; Formal analysis, Y.P.; Methodology, Y.P. and H.L.; Supervision, Y.C.; Writing—original draft, M.Y.; Writing—review & editing, M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Postdoctoral Science Foundation (2021M700300) and the Major Program of the National Natural Science Foundation of China (7214000).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Financial data are available at http://cndata1.csmar.com/, and financial reports are available at https://www.10jqka.com.cn/.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.; Li, B.; Ma, C.Q.; Huang, F.; Feng, G.; Chen, H.; Hou, J.; Yan, L.; Wei, Q.; Luo, Q. Recent progress in organic solar cells (Part I material science). *Sci. China Chem.* 2021, 65, 224–268.

2. Liu, X.; Liang, Z.; Du, S.; Niu, X.; Tong, J.; Yang, C.; Lu, X.; Bao, X.; Yan, L.; Li, J.; et al. Two compatible acceptors as an alloy model with a halogen-free solvent for efficient ternary polymer solar cells. *ACS Appl. Mater. Interfaces* 2022, 14, 9386–9397.

3. Ren, Y.; Liu, X.; Li, H.; Qin, J.; Du, S.; Lu, X.; Tong, J.; Yang, C.; Li, J. Utilizing non-conjugated small-molecular tetrasodium iminodisuccinateas electron transport layer enabled improving efficiency of organic solar cells. *Opt. Mater.* 2022, 129, 112520.

4. Xue, L.; Liu, X.; Wang, Q.; Yang, M.; Du, S.; Yang, C.; Tong, J.; Xia, Y.; Li, J. Improved Performance of Organic Solar Cells by Utilizing Green Non-halogen Additive to Modulate Active Layer Morphology. *Energy Technol.* 2022. 202200504. https://doi.org/10.1002/ente.202200504.

5. Yang, C.; Zhang, C.; Chen, C.; Ren, Y.; Shen, H.; Tong, J.; Du, S.; Xia, Y.; Li, J. A New Alcohol-Soluble Polymer PFN-ID as Cathode Interlayer to Optimize Performance of Conventional Polymer Solar Cells by Increasing Electron Mobility. *Energy Technol.* 2022, 10, 2200199. https://doi.org/10.1002/ente.202200199.

6. Lin, R.; Xu, J.; Wei, M.; Wang, Y.; Qin, Z.; Liu, Z.; Wu, J.; Xiao, K.; Chen, B.; Park, S.M.; et al. All-perovskite tandem solar cells with improved grain surface passivation. *Nature* 2022, 603, 73–78.

7. Wang, H.; Zhu, X.; Li, P. Research on the Impact of Government Subsidies on Solar Photovoltaic Industry Innovation. *Bus. Manag.* 2022, 44, 57–72.

8. Fitzpatrick, P.J. A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies; The Ohio State University: Columbus, OH, USA, 1932.

9. Beaver, W.H. Financial ratios as predictors of failure. *J. Acc. Res.* 1966, 4, 71–111.

10. Altman, E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Financ.* 1968, 23, 589–609.

11. Altman, E.I.; Haldeman, R.G.; Narayanan, P. A new model to identify bankruptcy risk of corporations. *J. Bank. Finance* 1977, 1, 29–54.

12. Ohlson, J.A. Financial ratios and the probabilistic prediction of bankruptcy. *J. Acc. Res.* 1980, 18, 109–131.

13. Gu, X.; Wang, B.; Li, W. A Research on Improving the Early Warning Accuracy of Logistic Financial Early Warning Model: An Analysis Introducing Earnings Management Variables. *J. Nanjing Audit. Univ.* 2018, 15, 45–52.

14. Pan, Z. Logistic Regression Early Warning Model for Enterprise Debt Default Risk. *Shanghai J. Econ.* 2018, 8, 73–83.

15. He, Y.; Li, X.; Wang, X. Early warning of non-financial factors for credit default of small and micro enterprises. *Stat. Decis.* 2019, 35, 180–183.

16. Yang, G.; Zhou, Y.; Sun, L. Enterprise Financial Early Warning Method Based on Benford-Logistic Model. *J. Quant. Tech.* 2019, 36, 149–165.

17. Chen, X.; Guo, H. Study on Financial Crisis Warning of Agricultural Listed Companies Based on Factor Analysis and Logistic Regression Model. *J. Appl. Stat. Manag.* 2022, 41, 11–24.

18. Wang, Z.; Zhang, X. Research on Early Warning of Financial Distress from the Perspective of the Relationship between Capital Supply and Demand. *Account. Econ. Res.* 2021, 35, 21–36.

19. Tanighi, H.; Appolloni, A.; Shirzad, A.; Azad, A.J.S. Corporate social responsibility disclosure (CSR) and financial distressed risk (FDR): Does institutional ownership matter? *Sustainability* 2022, 14, 742.

20. Yang, G.; Du, F.; Jia, X. Financial Early Warning Model of BP Neural Network Based on the First and the Last Quality Factors. *Stat. Decis.* 2022, 38, 166–171.

21. Zhu, W.; Zhang, T.; Wu, Y.; Li, S.; Li, Z. Research on optimization of an enterprise financial risk early warning method based on the DS-RF model. *Int. Rev. Financ. Anal.* 2022, 81, 102140.

22. Wu, D.; Ma, X.; Olson, D.L. Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decis. Support Syst.* 2022, 159, 113814.

23. Gupta, A.; Dengre, V.; Kheruwala, H.A.; Shah, M. Comprehensive review of text-mining applications in finance. *Financ. Innov.* 2020, 6, 39.

24. Song, Y.; Huang, L. Research on Disclosure of Notes to Financial Statements Based on Text Mining Technology. *Friends Account.* 2019, 601, 142–147.

25. Lee, M.; Huang, Y. Corporate social responsibility and corporate performance: A hybrid text mining algorithm. *Sustainability* 2020, 12, 3075.

26. Xiao, Y.; Xiong, K.; Zhang, X. Enterprise Financial Risk Early Warning Model Based on TEI@I Methodology. *Manag. Rev.* 2020, 32, 226–235.

27. Ouyang, Z.; Lai, Y. Systemic financial risk early warning of financial market in China using Attention-LSTM model. *N. Am. J. Econ. Finance.* 2021, 56, 101383.

28. Li, S.; Shi, W.; Wang, J.; Zhou, H. A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction. *Inf. Process. Manag.* 2021, 58, 102673.
29. Jiang, C.; Lyu, X.; Yuan, Y.; Wang, Z.; Ding, Y. Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China. *Int. J. Forecast.* 2022, 38, 1086–1099.
30. Li, B.; Wang, Y.; Zhu, X.; Li, J. Identification and evolution analysis of important risk in insurance industry based on the textual risk disclosures in financial reports. *Syst. Eng.-Theory Pract.* 2022, 42, 333–344.
31. Yang, B.; Yang, Y.; Du, J. Enterprise Risk Expectation and Investment Strategy Selection: An Empirical Study based on Text Mining of Annual Reports of Listed Enterprises. *Bus. Manag. J.* 2022, 44, 122–140.
32. Luo, C. Research on Financial Risk Early Warning of Solar PV Enterprises in China. Master’s Thesis, East China University of Technology, Fuzhou, China, 2017.
33. Li, B.; Wang, Y.; Zhu, X.; Li, J. Identification and evolution analysis of important risk in insurance industry based on the textual risk disclosures in financial reports. *Syst. Eng.-Theory Pract.* 2022, 42, 333–344.
34. Yang, B.; Yang, Y.; Du, J. Enterprise Risk Expectation and Investment Strategy Selection: An Empirical Study based on Text Mining of Annual Reports of Listed Enterprises. *Bus. Manag. J.* 2022, 44, 122–140.
35. Luo, C. Research on Financial Risk Early Warning of Solar PV Enterprises in China. Master’s Thesis, East China University of Technology, Fuzhou, China, 2017.
36. Dong, X. Research on Financial Crisis Early Warning of Listed Companies in Photovoltaic Industry Based on EVA Theory. Master’s Thesis, Northwest University, Xi’an, China, 2017.
37. Lu, C. Research on the Construction and Application of GM Company’s Financial Risk Early Warning Indicator System. Master’s Thesis, Xiangtan University, Xiangtan, China, 2019.
38. Sun, J. Research on the Financial Risk OLAP Analysis and Early Warning in PV Enterprises. Master’s Thesis, North China Electric Power University, Beijing, China, 2017.
39. Bai, J. Research on Financial Risk Early Warning of Listed Companies in my country—Financial Risk Early Warning Model Based on Principal Component Analysis and Logistic Regression. *Manag. Adm.* 2022, 8, 8–15.
40. Li, C.; Jia, H.; Zhao, G.; Fu, H. Credit Risk Warning of Listed Companies Based on Information Disclosure Text Empirical evidence from management discussion and analysis of the Chinese annual report. *Chin. J. Manag. Sci.* 2022, 1, 1–14.
41. Liang, L.; Liu, B. Research on Financial Risk Early Warning of Listed Companies Based on Text Mining. *Comput. Eng. Appl.* 2022, 58, 255–266.
42. Zhou, L.; He, Z.; Meng, T. Credit Risk Contagion in an Enterprise Group with Dynamic Copula Models. *Chin. J. Manag. Sci.* 2019, 27, 71–82.