

Article

A Machine Learning Approach for Investigating the Determinants of Stock Price Crash Risk: Exploiting Firm and CEO Characteristics

Yan Li , Huiyuan Xue, Shiyu Wei, Rongping Wang  and Feng Liu * 

Business School, Shandong University, Weihai 264209, China

* Correspondence: liufeng@sdu.edu.cn

Abstract: This study uses machine learning to investigate the effects of firm and CEO characteristics on stock price crash risk by collecting massive data on publicly listed firms in China. The results show that eXtreme Gradient Boosting (XGBoost) is the most effective model for predicting stock price crash risk, with relatively satisfactory performance. Meanwhile, the SHapley Additive exPlanations (SHAP) method is used to interpret the importance of features. The results show that the average weekly return of a firm over a year (RET) contributes the most and is negatively associated with crash risk, followed by Sigma, IPO age, and firm size. We also found that, among CEO characteristics, CEO pay contributes substantially to crash risk at the firm level. Our findings have important implications for research into the impact of firm and CEO characteristics on stock price crash risk and provide a novel way for investors to plan their investment decisions and risk-taking behavior rationally.

Keywords: stock price crash risk; firm characteristics; CEO characteristics; machine learning



Citation: Li, Y.; Xue, H.; Wei, S.; Wang, R.; Liu, F. A Machine Learning Approach for Investigating the Determinants of Stock Price Crash Risk: Exploiting Firm and CEO Characteristics. *Systems* **2024**, *12*, 143. <https://doi.org/10.3390/systems12050143>

Academic Editor: William T. Scherer

Received: 5 March 2024

Revised: 7 April 2024

Accepted: 17 April 2024

Published: 23 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Stock price crash is undoubtedly a disaster for society and investors, particularly retail investors who concentrate their money on a handful of firms; a stock price crash in a portfolio reduces their wealth [1]. The severe economic losses caused by stock price crash have prompted extensive research into the internal formation mechanisms of stock price crash risk. Because of information opacity and asymmetry, managers can conceal bad news to keep their high salaries [2,3]. When bad news accumulates above a certain threshold, it is immediately disclosed to the market, resulting in a significant decrease in stock price and company reputation [4].

Firm characteristics influence stock price movement and corporate risk-taking, as is common in risk-related research [5]. For instance, Deng et al. [6] proposed that stock price crash risk is related to several firm characteristics, including cash flow, operating capacity, debt-paying ability, growth potential, and profitability. They also suggested that we could use machine learning techniques to find each factor's effects and feature importance. Large companies are likelier to experience a stock price crash because they imply discretionary disclosure [4,7]. However, other studies have stated that a stock price crash occurs when the “bad news hoarding” phenomenon accumulates and reaches a critical value, at which point the bad news floods the stock market without warning [3,8,9]. Therefore, we require a systematic investigation of the influence of firm characteristics to accurately reflect each determinant's impact.

Previous studies have also identified CEO characteristics as key factors influencing firm-specific stock price crash risk [10–13]. For example, a younger CEO early in his career has the incentive to defer bad news for a string of consecutive earnings, increasing the likelihood of a crash risk [10]. Furthermore, a CEO with a greater position of power can withhold bad news for financial gain [14], resulting in stock price crashes. In addition, an overconfident CEO with poor management skills is likelier to overstate returns and ignore

bad news, increasing crash risk [15]. According to Habib, Hasan, and Jiang [16], female CEOs positively impact stock price crash risk, but the relationship varies depending on whether the female CEO is in her first appointment. Various aspects of CEO characteristics can influence crash risk, so a comprehensive investigation into how CEO characteristics contribute to crashes is urgently needed.

Understanding the causes of stock price crashes is critical for instructing investors on how to protect shareholder value and reduce wealth losses. The large number of firms in the Chinese market across a wide range of industries and sectors allows for the collection of a large amount of data, which helps to analyze and predict the risk of stock crashes more accurately. Thus, this study uses data from Chinese listed companies from 2010 to 2020 to provide comprehensive information on the factors that influence crash risk in the Chinese stock market and explore how firms and CEO characteristics affect crash risk using machine learning algorithms. Unlike traditional analytical tools, machine learning methods can precisely analyze large and complex datasets and produce convincing results [17].

Our study contributes to two areas of risk management research. First, we developed a novel stock price crash determinants model by combining firm and CEO characteristics to provide a new perspective on crash risk research. Previous studies in this field have focused solely on the impact of firm or CEO characteristics [6,14,15], but this research combined the two. Second, our study contributes to revealing important rankings in firm and CEO characteristics and finding specific relationships between factors and crash risk. These relationships have empirical implications for increasing the detectability of firm-specific stock price crashes and improving stock market regulation.

The remainder of this study is structured as follows. Section 2 shows the data source and measurements, and Section 3 presents the analysis results. Section 4 discusses these findings. Finally, Section 5 concludes the paper by discussing the study's contributions and limitations.

2. Research Methodology

Using machine learning algorithms, this study analyzed crash risk based on firm and the CEO characteristics. Machine learning can process a wide range of complex data, and its superior accuracy and explanatory power have made it a popular method for prediction and analysis [18–21]. In addition to its ability to identify long-term and delicate temporal patterns that are difficult for human analysts to detect, machine learning is particularly effective at modeling nonlinear behavior in financial data and accurately predicting the interaction effects of leading indicators of financial volatility [22]. Many scholars have shown that combining advanced deep learning and machine learning techniques is the best approach for financial forecast performance [23,24]. Recognizing the value of machine learning, we used 11 machine learning methods to investigate the relationships between firm characteristics and crash risk, including ridge regression, least absolute shrinkage and selection operator (Lasso), elastic net, multilayer perceptron regressor (MLPRegressor), decision tree, bagging, random forest, Extra-Trees, adaptive boosting (AdaBoost), Gradient Boosting Decision Tree (GBDT), and eXtreme Gradient Boosting (XGBoost).

2.1. Data and Sample

Our data cover listed companies in China from 2010 to 2020. The Chinese market, which has long been the world's second largest and most dynamic market, has significantly affected the global market. The Chinese stock market has a short history, and its laws and regulatory systems are insufficient. Thus, relying on Chinese stock market data is reasonable. We used the Choice dataset to collect firm characteristics and stock returns, and the China Stock Market and Accounting Research (CSMAR) database to obtain CEO characteristics. Then, we matched the firm characteristics data, CEO characteristics data, and stock return data using the same stock code and fiscal year. Furthermore, we winsorized all variables at 1% and 99% to account for potential bias, as extreme values affect the accuracy of the analyzed results. To accurately measure the crash risk, we excluded

observations with missing values and those with fewer than 30 weeks of stock returns. The final sample source included 1,999 firms (11,915 firm-year observations) from 2010 to 2020. Table 1 shows companies' frequency and percentage distributions across the industries examined in this study. The dataset covers a wide range of industries from computer communication to production and supply of electric power and heat power, highlighting the diversity and intricacy of these industries.

Table 1. Distribution of the sample across industries.

| Panel A. Distribution of Sample Firms Across Industries | | |
|--|------------------|-------------------|
| Industry | Frequency | Percentage |
| Computers, Communication Equipment, and Other Electronic Equipment Manufacturing | 201 | 10.06% |
| Pharmaceutical Manufacturing | 168 | 8.40% |
| Electrical Machinery and Equipment Manufacturing | 138 | 6.90% |
| Software and Information Technology Services | 136 | 6.80% |
| Raw Chemical Materials and Chemical Products Manufacturing | 134 | 6.70% |
| Special-purpose Machinery Manufacturing | 134 | 6.70% |
| Real Estate | 77 | 3.85% |
| General-purpose Machinery Manufacturing | 68 | 3.40% |
| Automobile Manufacturing | 57 | 2.85% |
| Retail | 53 | 2.65% |
| Nonmetal Mineral Products | 48 | 2.40% |
| Rubber and Plastic Products | 45 | 2.25% |
| Wholesale | 41 | 2.05% |
| Internet and Other Related Services | 41 | 2.05% |
| Smelting and Pressing of Nonferrous Metals | 36 | 1.80% |
| Metal Products | 34 | 1.70% |
| Instrument and Apparatus Manufacturing | 34 | 1.70% |
| Wine, Beverage, and Refined Tea Manufacturing | 30 | 1.50% |
| Railway, Shipping, Aerospace, and Other Transport Equipment Manufacturing | 29 | 1.45% |
| Agricultural and Sideline Product Processing | 28 | 1.40% |
| Business Services | 27 | 1.35% |
| Food Manufacturing | 26 | 1.30% |
| Ecological Protection and Environmental Governance | 24 | 1.20% |
| Civil Engineering Construction | 23 | 1.15% |
| Textile Garments and Clothing | 22 | 1.10% |
| Production and Supply of Electric Power and Heat Power | 20 | 1.00% |
| Others | 325 | 16.26% |
| Total | 1999 | 100.00% |
| Panel B. Distribution of Firm-year Observations Across Industries | | |
| Industry | Frequency | Percentage |
| Computers, Communication Equipment, and Other Electronic Equipment Manufacturing | 1117 | 9.37% |
| Pharmaceutical Manufacturing | 1040 | 8.73% |
| Electrical Machinery and Equipment Manufacturing | 807 | 6.77% |
| Software and Information Technology Services | 796 | 6.68% |
| Raw Chemical Materials and Chemical Products Manufacturing | 761 | 6.39% |
| Special-purpose Machinery Manufacturing | 757 | 6.35% |
| Real Estate | 560 | 4.70% |
| General-purpose Machinery Manufacturing | 400 | 3.36% |
| Retail | 367 | 3.08% |
| Automobile Manufacturing | 333 | 2.79% |
| Nonmetal Mineral Products | 311 | 2.61% |
| Wholesale | 281 | 2.36% |
| Rubber and Plastic Products | 256 | 2.15% |
| Metal Products | 243 | 2.04% |
| Wine, Beverage, and Refined Tea Manufacturing | 230 | 1.93% |
| Internet and Other Related Services | 222 | 1.86% |
| Smelting and Pressing of Nonferrous Metals | 220 | 1.85% |
| Agricultural and Sideline Product Processing | 186 | 1.56% |
| Instrument and Apparatus Manufacturing | 176 | 1.48% |
| Business Services | 170 | 1.43% |
| Textile Garments and Clothing | 147 | 1.23% |
| Railway, Shipping, Aerospace and Other Transport Equipment Manufacturing | 147 | 1.23% |
| Food Manufacturing | 138 | 1.16% |
| Civil Engineering Construction | 138 | 1.16% |
| Textiles | 120 | 1.01% |
| Ecological Protection and Environmental Governance | 119 | 1.00% |
| Production and Supply of Electric Power and Heat Power | 105 | 0.88% |
| Others | 1768 | 14.84% |
| Total | 11,915 | 100.00% |

2.2. Measuring Stock Crash Risk

Following Chen et al. [7], we used negative conditional return skewness (NCSKEW) and down-to-up volatility (DUVOL) in our study to estimate stock price crash risk. We used weekly return data to estimate the weekly return of each firm using the following regression model [4]:

$$R_{i,w} = \beta_0 + \beta_1 R_{r,w-2} + \beta_2 R_{r,w-1} + \beta_3 R_{r,w} + \beta_4 R_{r,w+1} + \beta_5 R_{r,w+2} + \varepsilon_{i,w} \quad (1)$$

where $R_{i,w}$ is the stock return of firm i in week w , $R_{r,w}$ is weighted average return, and $\varepsilon_{i,w}$ is the bias term, representing the parts of stock returns that do not relate to market returns. We also used two lead and two lag terms to alleviate potential problems that result in asynchronous stock trading [25]. We measured $W_{j,t}$ as the natural logarithm of the residual ($\varepsilon_{i,w}$) plus 1.

NCSKEW, the negative conditional return skewness of a specific firm's weekly returns, is the first indicator of a stock price crash risk. We computed NCSKEW using the following model:

$$NCSKEW_{j,t} = -1 \left[n(n-1)^{3/2} \right] \sum W_{j,r}^3 / \left[(n-1)(n-2) \left(\sum W_{j,r}^2 \right)^{3/2} \right] \quad (2)$$

where n is the number of observations of daily returns for firm j in year t . $W_{j,t}$ is the weekly returns of a specific firm, measured as the natural logarithm of their residual plus 1. An increasing risk of stock price crash exists for a firm with a higher NCSKEW.

DUVOL, the weekly return down-to-up volatility, is the second indicator of a stock price crash risk. We calculated DUVOL with the following model:

$$DUVOL_{j,t} = \ln \left\{ \left[(n_u - 1) \sum_{DOWN} W_{j,t}^2 \right] / \left[(n_d - 1) \sum_{UP} W_{j,t}^2 \right] \right\} \quad (3)$$

where n_u is the number of up weeks and n_d is the number of down weeks. Specifically, we divided the total weeks into up weeks and down weeks. Furthermore, we calculated the standard deviations of the subsamples and found that firms with a higher DUVOL have an increased risk of stock price crash.

2.3. Measuring Determinants

We can divide the factors that influence crash risk into two categories: firm characteristics and CEO characteristics. The former includes firm age, IPO age, LogSize, leverage, goodwill, brand capital, cash, return on assets (ROA), return on equity (ROE), sigma, RET, and DTURN. The latter includes the CEO's gender, age, education, MBA, duality, tenure, pay, shareholdings, board experience, academic experience, and overseas experience. We used previous studies to assess firm and CEO characteristics [10,19,21,23,24,26–28]. Table 2 shows a detailed description of these variables.

Table 2. Variable definitions.

| Variable | Definition | Mean | SD |
|-----------------------------|---|--------|-------|
| Firm characteristics | | | |
| Firm age | Years since the firm's founding | 18.535 | 5.531 |
| IPO age | Years since the firm's IPO | 9.795 | 6.872 |
| LogSize | ln(Total assets) | 22.055 | 1.145 |
| Leverage | Total debt/Total assets | 0.413 | 0.207 |
| Goodwill | Goodwill/Total assets; Zero replaces the missing values in goodwill | 0.038 | 0.082 |
| Brand capital | Advertising expenses/Total assets | 0.038 | 0.082 |

Table 2. Cont.

| Variable | Definition | Mean | SD |
|----------------------------|--|--------|--------|
| Cash | (Cash+Short-term Investments)/Total assets | 0.167 | 0.129 |
| ROA | Net income/Total assets | 0.055 | 0.070 |
| ROE | Net income/Total equity | 0.059 | 0.141 |
| Sigma | Standard deviation of weekly stock revenue over a year | 0.052 | 0.020 |
| RET | Average weekly returns of a specific firm over a year | 0.000 | 0.007 |
| DTURN | Current-year mean monthly share turnover–Last-year mean monthly share turnover | 0.140 | 24.296 |
| CEO characteristics | | | |
| CEO gender | “1” if the CEO is male and “0” if the CEO is female | 0.930 | 0.256 |
| CEO age | Age of the CEO | 49.724 | 6.508 |
| CEO education | “1” if the CEO holds a postgraduate degree and “0” otherwise | 0.465 | 0.499 |
| CEO MBA | “1” if the CEO holds an MBA degree and “0” otherwise. | 0.077 | 0.267 |
| CEO duality | “1” if the CEO also serves as chairman and “0” otherwise | 0.314 | 0.464 |
| CEO tenure | Number of years in a CEO position with a particular company | 1.315 | 0.798 |
| CEO pay | ln(Total Annual Salary + 1) | 13.127 | 1.813 |
| CEO shareholdings | ln(Outstanding Shares held by CEO + 1) | 9.055 | 7.903 |
| CEO board experience | “1” if the CEO is also a director and “0” otherwise | 0.918 | 0.274 |
| CEO academic experience | “1” if the CEO has the experience of (a) teaching at a college, (b) working at a research laboratory, and (c) researching at an institute, and “0” otherwise | 0.234 | 0.423 |
| CEO overseas experience | “1” if the CEO has overseas experience and “0” otherwise | 0.094 | 0.292 |
| CEO production | “1” if the CEO has career experience in the production area and “0” otherwise | 0.127 | 0.333 |
| CEO RD | “1” if the CEO has career experience in the R&D area and “0” otherwise | 0.266 | 0.442 |
| CEO design | “1” if the CEO has career experience in the design area and “0” otherwise | 0.031 | 0.172 |
| CEO HRM | “1” if the CEO has career experience in the human resource management area and “0” otherwise | 0.022 | 0.146 |
| CEO administration | “1” if the CEO has career experience in the administration area and “0” otherwise | 1.000 | 0.000 |
| CEO marketing | “1” if the CEO has career experience in the marketing area and “0” otherwise | 0.291 | 0.454 |
| CEO finance | “1” if the CEO has career experience in the finance area and “0” otherwise | 0.136 | 0.342 |
| CEO accounting | “1” if the CEO has career experience in the accounting area and “0” otherwise | 0.105 | 0.306 |

3. Results

3.1. Evaluation Criterion

We used MSE to evaluate the model performance, computed as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

where Y_i is the true value of NCSKEW or DUVOL, and \hat{Y}_i is the predicting value. MSE is a good indicator for evaluating machine learning models [29,30]. Thus, we employ MSE to compare our 11 machine learning models.

3.2. Model Evaluation and Comparison

This study trained machine learning models such as ridge regression, Lasso, elastic net, MLPLRegressor, decision tree, bagging, random forest, Extra-Trees, AdaBoost, GBDT, and XGBoost on 80% of the data. We used the remaining 20% to assess the model. We evaluated the model using five-fold cross-validation, which divides the dataset into five subsamples. The model extracts four subsamples to train models, and the remaining subsample serves as a test set to evaluate the model. Finally, we constructed and evaluated five models using a different test dataset. In addition, we chose MSE to evaluate the model. Models with higher MSE values perform poorly. Table 3 displays the MSE results of the five-fold cross-validation process (MSE kf1–MSE kf5). When measured with NCSKEW, XGBoost’s MSE values ranged from 0.4654 to 0.5518, whereas when measured with DUVOL, they

ranged from 0.2293 to 0.2657, which is lower than other methods. XGBoost performed best with a minimum mean MSE value, regardless of whether we used NCSKEW or DUVOL as the measure.

Table 3. Descriptive results of machine learning methods.

| Model | MSE | MSE kf1 | MSE kf2 | MSE kf3 | MSE kf4 | MSE kf5 | Mean |
|--------------------------------------|--------|---------|---------|---------|---------|---------|--------|
| Panel A. Using NCSKEW as the measure | | | | | | | |
| XGBoost | 0.4557 | 0.4883 | 0.4654 | 0.5337 | 0.5222 | 0.5518 | 0.5123 |
| GBDT | 0.4578 | 0.4831 | 0.4853 | 0.5522 | 0.5036 | 0.5372 | 0.5123 |
| AdaBoost | 0.4522 | 0.5032 | 0.4898 | 0.5417 | 0.5109 | 0.5502 | 0.5191 |
| Bagging | 0.4582 | 0.5055 | 0.4939 | 0.5503 | 0.5168 | 0.5553 | 0.5243 |
| Random forest | 0.5196 | 0.4971 | 0.5305 | 0.5722 | 0.5430 | 0.6794 | 0.5644 |
| Extra-Trees | 0.5385 | 0.5402 | 0.5351 | 0.5810 | 0.5647 | 0.6586 | 0.5759 |
| Lasso | 0.5241 | 0.6591 | 0.6528 | 0.5668 | 0.5531 | 0.6562 | 0.6176 |
| Ridge | 0.5248 | 0.6666 | 0.6528 | 0.5675 | 0.5550 | 0.6551 | 0.6194 |
| MLPRegressor | 0.9078 | 0.7692 | 0.6892 | 0.7641 | 0.7241 | 0.8027 | 0.7499 |
| Elastic net | 0.8124 | 0.7649 | 0.8174 | 0.9246 | 0.8383 | 0.8271 | 0.8345 |
| Decision tree | 1.0060 | 1.0545 | 1.0502 | 0.9974 | 1.1352 | 1.1662 | 1.0807 |
| Panel B. Using DUVOL as the measure | | | | | | | |
| XGBoost | 0.2186 | 0.2459 | 0.2657 | 0.2589 | 0.2472 | 0.2293 | 0.2494 |
| GBDT | 0.2215 | 0.2366 | 0.2819 | 0.2715 | 0.2295 | 0.2302 | 0.2499 |
| AdaBoost | 0.2235 | 0.2559 | 0.2851 | 0.2717 | 0.2377 | 0.2491 | 0.2599 |
| Bagging | 0.2243 | 0.2555 | 0.2808 | 0.2717 | 0.2333 | 0.2585 | 0.2600 |
| Random forest | 0.2683 | 0.2501 | 0.3257 | 0.3008 | 0.2610 | 0.3519 | 0.2979 |
| Extra-Trees | 0.2938 | 0.2915 | 0.3245 | 0.3298 | 0.2878 | 0.3279 | 0.3123 |
| Lasso | 0.2847 | 0.4207 | 0.4454 | 0.3260 | 0.2801 | 0.3447 | 0.3634 |
| Ridge | 0.2853 | 0.4245 | 0.4458 | 0.3259 | 0.2808 | 0.3437 | 0.3641 |
| MLPRegressor | 0.6372 | 0.4650 | 0.5544 | 0.6304 | 0.4168 | 0.4075 | 0.4948 |
| Elastic net | 0.4766 | 0.5588 | 0.5457 | 0.5194 | 0.4912 | 0.4720 | 0.5174 |
| Decision tree | 0.5534 | 0.5155 | 0.5811 | 0.6509 | 0.5470 | 0.4683 | 0.5526 |

Note: The table compares the performance of 11 models using MSE aided by five-fold cross-validation.

Next, we compared the models' performance using MSE as the measure. We found that the MSE value to measure crash risk in the XGBoost was 0.4557 using NCSKEW and 0.2186 using DUVOL, which were slightly higher than those in ridge regression, Lasso, bagging, Extra-Trees, and GBDT. However, the XGBoost results outperformed the elastic net, MLPRegressor, and decision tree models. Although the MSE value of AdaBoost was slightly lower than that of XGBoost when measured using NCSKEW, the mean value of the five-fold cross-validation in XGBoost was lower than that of AdaBoost, indicating that XGBoost has greater stability than AdaBoost. The results are summarized in Table 3, and show that XGBoost is the best model to explore the determinants of crash risk.

3.3. Model Interpretation

Our study interpreted the model results using Lundberg and Lee's [31] proposed SHapley Additive exPlanation (SHAP) model. SHAP indicates the contributions of each variable using game theory. We calculate SHAP values as follows:

$$SHAP\ value_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (5)$$

where i is a feature that we need to interpret. N is the set of all features we input. M is the number of features that we need to interpret, and S is the subset of N . f_x is the predicted result of x in the models. $f_x(S)$ indicates the predicted result of the set of S , and $f_x(S \cup \{i\})$ is the predicted result with the set of S adding i .

Feature importance is critical for determining which features contribute the most to the model's performance. This study used SHAP to estimate the model by determining

each factor's importance and the feature's specific impact. Table 4 shows the results of the SHAP summary analysis. Using the summary table, we calculated the mean absolute SHAP value, which reflected the contributions of each variable's characteristics. Furthermore, the table efficiently conveys the effects of the variables on the model. Finally, we calculated the average feature importance rank because NCSKEW produces results that differ from those of DUVOL.

Table 4. The results of the SHAP summary analysis.

| Variable | SHAP1 | Effects1 | Rank1 | SHAP2 | Effects2 | Rank2 | Average Rank |
|-------------------------|--------|----------|-------|--------|----------|-------|--------------|
| RET | 0.4565 | −0.9212 | 1 | 0.4441 | −0.9330 | 1 | 1 |
| Sigma | 0.1206 | −0.9230 | 2 | 0.0738 | −0.8016 | 2 | 2 |
| IPO age | 0.0779 | −0.8919 | 3 | 0.0641 | −0.8554 | 3 | 3 |
| LogSize | 0.0624 | −0.8781 | 4 | 0.0518 | −0.9182 | 4 | 4 |
| DTURN | 0.0296 | −0.1388 | 5 | 0.0215 | −0.5543 | 5 | 5 |
| Brand capital | 0.0257 | −0.1839 | 6 | 0.0178 | −0.3067 | 7 | 6.5 |
| CEO pay | 0.0193 | −0.4289 | 9 | 0.0203 | −0.5393 | 6 | 7.5 |
| Leverage | 0.0169 | 0.3330 | 10 | 0.0155 | 0.5731 | 8 | 9 |
| Cash | 0.0217 | −0.1021 | 7 | 0.0081 | −0.1376 | 12 | 9.5 |
| Goodwill | 0.0115 | 0.2983 | 11 | 0.0119 | 0.7344 | 9 | 10 |
| ROE | 0.0197 | −0.0529 | 8 | 0.0080 | −0.2135 | 13 | 10.5 |
| Firm age | 0.0104 | −0.3905 | 13 | 0.0096 | −0.7212 | 10 | 11.5 |
| ROA | 0.0110 | 0.5675 | 12 | 0.0085 | 0.6158 | 11 | 11.5 |
| CEO accounting | 0.0086 | 0.8266 | 15 | 0.0046 | 0.7113 | 15 | 15 |
| CEO shareholdings | 0.0079 | 0.3512 | 17 | 0.0058 | −0.2465 | 14 | 15.5 |
| CEO age | 0.0081 | 0.1521 | 16 | 0.0045 | 0.4151 | 16 | 16 |
| CEO marketing | 0.0090 | −0.7914 | 14 | 0.0041 | −0.8544 | 18 | 16 |
| CEO tenure | 0.0067 | 0.2150 | 18 | 0.0039 | 0.3601 | 19 | 18.5 |
| CEO education | 0.0016 | −0.8944 | 21 | 0.0042 | −0.7742 | 17 | 19 |
| CEO academic experience | 0.0018 | 0.6689 | 19 | 0.0015 | 0.8396 | 20 | 19.5 |
| CEO RD | 0.0006 | −0.0171 | 24 | 0.0006 | −0.4033 | 21 | 22.5 |
| CEO duality | 0.0007 | −0.3051 | 23 | 0.0004 | 0.0793 | 26 | 24.5 |
| CEO board experience | 0.0004 | 0.4694 | 26 | 0.0005 | 0.0919 | 24 | 25 |
| CEO design | 0.0017 | −0.5892 | 20 | 0.0000 | −0.0667 | 30 | 25 |
| CEO finance | 0.0004 | 0.4053 | 27 | 0.0005 | 0.5884 | 23 | 25 |
| CEO HRM | 0.0010 | 0.7714 | 22 | 0.0001 | 0.4582 | 28 | 25 |
| CEO production | 0.0006 | −0.5677 | 25 | 0.0004 | −0.3166 | 25 | 25 |
| CEO gender | 0.0002 | 0.3752 | 30 | 0.0005 | −0.7119 | 22 | 26 |
| CEO overseas experience | 0.0002 | −0.3707 | 29 | 0.0002 | −0.2451 | 27 | 28 |
| CEO MBA | 0.0003 | 0.5626 | 28 | 0.0001 | 0.5475 | 29 | 28.5 |
| CEO administration | 0.0000 | 0.0000 | 31 | 0.0000 | 0.0000 | 31 | 31 |

Note: SHAP1, Effects1, and Rank1 are the results using NCSKEW as a measure and SHAP2, Effects2, and Rank2 are the results using DUVOL. Rank is the rank of feature importance on average.

The SHAP summary table shows that the eight most important firm characteristics are RET, sigma, IPO age, LogSize, DTURN, brand capital, leverage, and cash. In addition, the top eight CEO characteristics include CEO pay, CEO accounting, CEO shareholdings, CEO age, CEO marketing, CEO tenure, CEO education, and CEO academic experience. The SHAP summary table displays the specific impact of features on the risk of stock price crash. Crash risk correlates negatively with RET, sigma, IPO age, LogSize, DTURN, brand capital, CEO pay, cash, ROE, firm age, CEO marketing, CEO education, CEO RD, CEO design, CEO production, and CEO overseas experience. Meanwhile, a positive relationship exists between crash risk and leverage, goodwill, ROA, CEO accounting, CEO age, CEO tenure, CEO academic experience, CEO board experience, CEO finance, CEO HRM, and CEO MBA. However, using NCSKEW to estimate the impact of CEO duality and CEO gender yields different results than using DUVOL.

4. Discussion

Intense fluctuations in stock prices have created uncertainty in investors' reactions and behaviors, as well as in the daily operations of companies, increasing the need to manage stock price crash risk for companies' long-term development. Given that China is the world's largest developing country, with a huge consumer base and diverse market demand, there is significant uncertainty about stock market changes. Based on the background of the Chinese market, we used 11 machine learning methods to explore the influential factors of crash risk using firm and CEO characteristics from a large-scale sample of Chinese listed companies. This section discusses the study's significant theoretical and practical implications.

This study contributes to the literature on stock price crash risk by developing a stock price crash determinants model using machine learning techniques considering firm and CEO characteristics. Although much previous literature has studied the relationship between firm and CEO characteristics and crash risk, most research has not explored the joint influence and influential degree of large firm and CEO characteristics [4,6]. As a result, the complexity of analyzing the determinants of crash risk has increased due to the presence of multiple features at once. However, our study fills the gap by applying machine learning methods to explore the factors that significantly impact stock price crash risk.

This study provides evidence of how firm and CEO characteristics impact crash risk, including firm characteristics (i.e., RET, sigma, and firm age) and CEO characteristics (i.e., CEO pay, CEO accounting, and CEO shareholdings). Extending Zhang et al.'s [32] study with the SHAP method, we found that RET, or the weekly return of a firm over a year, is the most significant factor among these, negatively impacting crash risk. However, unlike Xu and Zou [33], who argued that the relationship between CEO pay and stock price crash risk is unclear, our study found that CEO pay has the greatest significance in terms of CEO characteristics and has a negative influence on crash risk. Higher pay encourages them to focus more on firm performance, resulting in lower crash risk.

Furthermore, we found that a firm with a CEO with accounting experience is more likely to crash. We think that a CEO with accounting experience can easily manipulate company performance, resulting in a higher crash risk. Additionally, CEO finance and crash risk are positively associated, consistent with Jiang et al.'s [11] findings. However, our study's results on CEO age's impact on crash risk contradict previous research. According to other researchers, a firm with a younger CEO is more likely to experience a crash risk [10]. However, our results show a positive relationship between CEO age and crash risk.

This study has some practical implications for investors and supervisors. First, there is a strong association between firm characteristics and crash risk. Our study found that RET, IPO age, firm size, and brand capital affect crash risk. To avoid loss, investors should focus on these factors. For example, they can invest in stocks with a higher and more stable average weekly return, a higher ROE, and a lower ROA. RET, as the most important determinant in current research, should be noted that it negatively influences crash risk. A low or declining RET for a company may indicate poor profitability or business challenges, leading investors to become pessimistic about its future performance and stock price, resulting in share sales. Consequently, a mass sell-off can put downward pressure on the stock price, potentially leading to a market disaster if left unchecked. During a crash, stock prices may plummet, market turnover may sharply decrease, investor confidence may suffer, and the stock market's operational mechanism may sustain significant damage. Therefore, a decreasing or downward-trending RET is an early indicator of an impending stock crash. Investors and regulators should closely monitor this indicator and respond quickly to potential risks. Simultaneously, companies should strive to improve performance and profitability to maintain share prices and bolster investor trust.

Our findings suggest that regulators should be vigilant in monitoring firms' characteristics, particularly those associated with higher crash risk. They can use the insights from this study to develop more targeted surveillance mechanisms and policies to prevent potential crashes. For example, they could impose stricter disclosure requirements on firms

with specific characteristics, such as a high IPO age or a large size, to ensure that investors can access all relevant information when making investment decisions.

5. Conclusions

This study comprehensively identified the impact of various firm and CEO characteristics on crash risk and their respective contributions to stock price crash prediction. We applied 11 machine learning models to analyze data from Chinese listed firms between 2010 and 2020. The practical results show that XGBoost has the best machine learning performance and effectively examines the relationships between 31 input variables and the risk of a stock price crash. Furthermore, the SHAP method used in this study shows the importance of firm and CEO characteristics in interpreting the XGBoost model. We found among the characteristics of the firm and the CEO, the most ten important factors that impact the stock price crash risk are RET, sigma, IPO age, LogSize, DTURN, Brand capital, CEO pay, Leverage, Cash, and Goodwill.

Although this article contributes to a comprehensive presentation of firm and CEO characteristics that may be associated with the risk of a stock price crash, there are still research limitations. First, our study only included 31 features in our models, which may explain the models' limited performance. The study may have overlooked important variables impacting model performance. Second, we do not differentiate between the equity sectors to which the stock belongs, which may affect the accuracy of the results. Future research can use advanced machine learning methods to analyze samples from various equity sectors. Moreover, we only considered stocks from companies listed in the Chinese stock market. Future research can use a longer period of observations and a broader range of stocks from different markets to explore crash risk more comprehensively.

Author Contributions: Conceptualization, Y.L., S.W. and F.L.; methodology, Y.L., S.W., H.X., R.W. and F.L.; software, S.W.; formal analysis, Y.L., S.W. and F.L.; resources, F.L.; data curation, S.W.; writing—original draft preparation, Y.L., H.X., S.W. and F.L.; writing—review and editing, Y.L., H.X., R.W. and F.L.; supervision, F.L.; project administration, F.L.; funding acquisition, F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Humanities and Social Sciences Foundation of the Ministry of Education of China (Grant No. 21YJC630076).

Data Availability Statement: Data will be available from the corresponding author upon reasonable request.

Acknowledgments: We gratefully acknowledge the insightful suggestions from the editors and anonymous reviewers that substantively improved this article. We would also like to thank the members of Star-lights Research Team for their comments on earlier versions of the manuscript.

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

References

1. Barber, B.M.; Odean, T. The behavior of individual investors. *Handbook Econ. Financ.* **2013**, *2*, 1533–1570. [\[CrossRef\]](#)
2. Bleck, A.; Liu, X. Market transparency and the accounting regime. *J. Account.* **2007**, *45*, 229–256. [\[CrossRef\]](#)
3. Jin, L.; Myers, S.C. R2 around the world: New theory and new tests. *J. Financ. Econ.* **2006**, *79*, 257–292. [\[CrossRef\]](#)
4. Hutton, A.P.; Marcus, A.J.; Tehranian, H. Opaque financial reports, R2, and crash risk. *J. Financ. Econ.* **2009**, *94*, 67–86. [\[CrossRef\]](#)
5. Liu, C.; Chen, Y.; Huang, S.; Chen, X.; Liu, F. Assessing the Determinants of Corporate Risk-Taking Using Machine Learning Algorithms. *Systems* **2023**, *11*, 263. [\[CrossRef\]](#)
6. Deng, S.; Zhu, Y.; Duan, S.; Fu, Z.; Liu, Z. Stock Price Crash Warning in the Chinese Security Market Using a Machine Learning-Based Method and Financial Indicators. *Systems* **2022**, *10*, 108. [\[CrossRef\]](#)
7. Chen, J.; Hong, H.; Stein, J.C. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *J. Financ. Econ.* **2001**, *61*, 345–381. [\[CrossRef\]](#)
8. Baik BO, K.; Farber, D.B.; Lee SA, M. CEO ability and management earnings forecasts. *Contemp. Account. Res.* **2011**, *28*, 1645–1668. [\[CrossRef\]](#)
9. Xie, W.; Ye, C.; Wang, T.; Shen, Q. M&A goodwill, information asymmetry and stock price crash risk. *Econ. Res.-Ekon. Istraz.* **2020**, *33*, 3385–3405. [\[CrossRef\]](#)
10. Andreou, P.C.; Louca, C.; Petrou, A.P. CEO age and stock price crash risk. *Int. Rev. Financ.* **2017**, *21*, 1287–1325. [\[CrossRef\]](#)

11. Jiang, X.; Zhu, J.; Akbar, A.; Hou, Z.; Bao, X. The dark side of executives' professional background: Evidence from Chinese firm's stock price crash risk. *Manag. Decis. Econ.* **2022**, *43*, 3771–3784. [\[CrossRef\]](#)
12. Kim, J.B.; Liao, S.; Liu, Y. Married CEOs and stock price crash risk. *Eur. Financ. Manag.* **2022**, *28*, 1376–1412. [\[CrossRef\]](#)
13. Li, Y.; Zeng, Y. The impact of top executive gender on asset prices: Evidence from stock price crash risk. *J. Corp. Financ.* **2019**, *58*, 528–550. [\[CrossRef\]](#)
14. Al Mamun, M.; Balachandran, B.; Duong, H.N. Powerful CEOs and stock price crash risk. *J. Corp. Financ.* **2020**, *62*, 101582. [\[CrossRef\]](#)
15. Kim, J.B.; Wang, Z.; Zhang, L. CEO overconfidence and stock price crash risk. *Contemp. Account. Res.* **2016**, *33*, 1720–1749. [\[CrossRef\]](#)
16. Habib, A.; Hasan, M.M.; Jiang, H. Stock price crash risk: Review of the empirical literature. *Account. Financ.* **2018**, *58*, 211–251. [\[CrossRef\]](#)
17. Zhu, Y.; Xie, C.; Wang, G.J.; Yan, X.G. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. *Neural Comput. Appl.* **2017**, *28*, 41–50. [\[CrossRef\]](#)
18. Choudhury, P.; Allen, R.T.; Endres, M.G. Machine learning for pattern discovery in management research. *Strateg. Manag. J.* **2021**, *42*, 30–57. [\[CrossRef\]](#)
19. Liu, F.; Wang, R.; Fang, M. Mapping green innovation with machine learning: Evidence from China. *Technol. Forecast. Soc. Chang.* **2024**, *200*, 123107. [\[CrossRef\]](#)
20. Liu, F.; Long, X.; Dong, L.; Fang, M. What makes you entrepreneurial? Using machine learning to investigate the determinants of entrepreneurship in China. *China Econ. Rev.* **2023**, *81*, 102029. [\[CrossRef\]](#)
21. Liu, F.; Huang, W.; Zhang, J.; Fang, M. Corporate social responsibility in family business: Using machine learning to uncover who is doing good. *Technol. Soc.* **2024**, *76*, 102453. [\[CrossRef\]](#)
22. Chatzis, S.P.; Siakoulis, V.; Petropoulos, A.; Stavroulakis, E.; Vlachogiannakis, N. Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Syst. Appl.* **2018**, *112*, 353–371. [\[CrossRef\]](#)
23. Wang, M.; Yu, Y.; Liu, F. Does digital transformation curb the formation of zombie firms? A machine learning approach. *Technol. Anal. Strat. Manag.* **2023**, 1–17. [\[CrossRef\]](#)
24. Zhang, J.; Zhu, M.; Liu, F. Find who is doing social good: Using machine learning to predict corporate social responsibility performance. *Oper. Manag. Res.* **2023**, *17*, 253–266. [\[CrossRef\]](#)
25. Dimson, E. Risk measurement when shares are subject to infrequent trading. *J. Financ. Econ.* **1979**, *7*, 197–226. [\[CrossRef\]](#)
26. Ben-Nasr, H.; Bouslimi, L.; Zhong, R. Do patented innovations reduce stock price crash risk? *Inter. Rev. Financ.* **2021**, *21*, 3–36. [\[CrossRef\]](#)
27. Chen, Y.; Fan, Q.; Yang, X.; Zolotoy, L. CEO early-life disaster experience and stock price crash risk. *J. Corp. Financ.* **2021**, *68*, 101928. [\[CrossRef\]](#)
28. Hasan, M.M.; Taylor, G.; Richardson, G. Brand capital and stock price crash risk. *Manag. Sci.* **2022**, *68*, 7221–7247. [\[CrossRef\]](#)
29. Baek, Y.; Kim, H.Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Syst. Appl.* **2018**, *113*, 457–480. [\[CrossRef\]](#)
30. Lv, P.; Shu, Y.; Xu, J.; Wu, Q. Modal decomposition-based hybrid model for stock index prediction. *Expert Syst. Appl.* **2022**, *202*, 117252. [\[CrossRef\]](#)
31. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
32. Zhang, M.; Xie, L.; Xu, H. Corporate philanthropy and stock price crash risk: Evidence from China. *J. Bus. Ethics.* **2016**, *139*, 595–617. [\[CrossRef\]](#)
33. Xu, J.; Zou, L. The impact of CEO pay and its disclosure on stock price crash risk: Evidence from China. *China Financ. Rev. Int.* **2019**, *9*, 479–497. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.