



Shahid Bahonar
University of Kerman



Iranian
Accounting Association

Predicting Financial Distress of Companies Using Textual Information of Board of Directors' Activity Reports

Mojtaba Alifamian¹, Mohammad Marfou², Mohamad javad Salimi³ and Iman Raeesi Vanani⁴

1. Department of Accounting, Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran.

Email: mojtaba.alifamian@yahoo.com

2. **Corresponding Author**, Department of Accounting, Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran. **Email:** marfou@atu.ac.ir

3. Department of Accounting, Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran.

Email: j_salimi@atu.ac.ir

4. Department of Information Technology Management, Allameh Tabataba'i University, Tehran, Iran.

Email: imanraeesi@atu.ac.ir

Article Info

A B S T R A C T

Article Type:

Research Article.

Article History:

Received: 22 February 2024

Received in revised form: 12 May 2024

Accepted: 13 May 2024

Available online: 22 July 2025

Keywords:

Financial Distress,
Text Mining,
Machine Learning,
Python.

JEL Classification:

G33, M41.

Objective: Financial distress is a serious concern for the economic sustainability of countries. The individual and social costs associated with financial distress have made its prediction a crucial issue for many managers, stakeholders, policymakers, and auditors. While most prior studies have relied on structured and quantitative financial data, this research aims to employ text mining techniques and machine learning algorithms to predict financial distress using unstructured data extracted from board of directors' activity reports.

Method: To achieve this objective, the board of directors' reports of 100 listed companies over the period 2011–2021 were collected. The textual content of these reports was analyzed using Python, involving preprocessing, feature extraction, and feature selection. The statistical analysis and modeling phase was carried out using various machine learning algorithms.

Results: The results of the study indicated the superiority of the two methods, the decision tree model and the support vector machine with radial kernel, over other methods (including logistic regression, random forest, nearest neighbor, and support vector machine methods with linear, sigmoid, and polynomial kernels).

Conclusion: The findings of this study suggest that, rather than solely relying on numerical data and the ratios derived from such figures, text mining techniques can also be effectively utilized for analysis and prediction. Furthermore, by integrating insights from unstructured textual data with structured quantitative information, it is possible to enhance the prediction of corporate financial distress.

Cite this article: Alifamian, M., Marfou, M., Salimi, M.J., & Raeesi Vanani, I. (2025). Predicting financial distress of companies using textual information of board of directors' activity reports. *Journal of Accounting Knowledge*, 16(2), 151-176. [In Persian].

DOI: <https://doi.org/10.22103/jak.2024.22992.4017>



Publisher: Shahid Bahonar University of Kerman.

© Marfou et al.

Introduction

In modern attitudes towards the company as a socio-economic unit, companies, in addition to their economic function, have colorful functions in non-economic fields, such as helping the labor market and employment capacity, the environment and preserving or destroying it, future generations, and transferring debt or capital. Therefore, according to the economic and non-economic consequences of the companies, the financial distress and bankruptcy of the companies or their health and efficiency have a very important effect on the socio-economic indicators both at the macro level and at the micro level. Considering these effects, predicting these situations can help in creating awareness to deal with the possible future situation (especially the economic dimension). So far, most of the research conducted on the financial distress and bankruptcy of companies has been based on structured data and quantitative disclosures. This is even though nowadays a large part of information is published through qualitative disclosures. Although the quantitative information of financial reports of companies is of great importance; this information alone cannot provide a complete picture of the company's condition for users. To use the qualitative information of financial reports, investors must first decode them and then process them. For this reason, providing and presenting written information along with quantitative information in financial reports has attracted the attention of producers, users, and regulators of accounting information to better understand quantitative information and provide information that cannot be provided in the form of numbers. In such a way, today, textual information has taken up more volume of financial reports of companies than quantitative information. Therefore, the current research seeks to first convert the unstructured data of the board of directors' report into structured data and then, use machine learning algorithms to provide a model for predicting financial distress in python.

Method

In this research, the report of the board of directors of 100 companies in the period of 2012-2022 was collected and to convert the unstructured data of the board of directors' reports into structured data, Python programming language has been used for text mining. So that after performing the text pre-processing steps, feature generation, feature selection, etc, to prepare textual data, the TF-IDF matrix has been used to select features. Then by combining these matrices and financial distress variables, machine learning algorithms are implemented and finally, the presented model has been validated using valid methods.

Results

After extracting the dependent variable (y matrix) and having explanatory variables (x matrix), we entered the modeling phase with different machine learning methods (including logistic regression, random forest, decision tree, nearest neighbor, and methods of support vector machine (SVM) with linear, RBF, sigmoid and polynomial methods). According to the results, it can be said that the two decision tree model and SVM method with radial kernels have the best performance compared to other methods based on the Precision index, which is 80 and 82 percent. According to the Recall criterion, which is the aim of identifying data that happened in financial distress, 78 and 86 percent of coverage of financial distress, respectively, based on the fscore criterion, which is a combination of two criteria of Precision and Recall, respectively 79% and 84% confirm the ability to make correct diagnoses. Other models compared to these two models have had a significant difference in diagnosing financial distress.

Conclusions

So far, many models have been presented for predicting financial distress, but what has been less discussed is the effect of the information content of the board of directors' report on predicting financial distress. The report of the board of directors contains valuable clues that, like financial data, can detect the existence of financial distress in companies. Therefore, in this

research, an attempt has been made to predict the financial distress of companies by using the information capacity of the board of directors' reports. After preprocessing and data mining operations and extraction of dependent variables, using different machine learning methods (including logistic regression, random forest, decision tree, nearest neighbor, and methods of vector machine with linear, RBF, sigmoid, and polynomial) methods, financial distress was measured. The prediction results indicate the high predictive power of two decision tree model (79%) and SVM with radial kernel (85%) compared to other models. The results of this research showed that instead of just paying attention to the numbers and the ratios derived from these numbers, the text mining technique can also be used for analysis and prediction, and by combining it with the results obtained from quantitative information, the financial distress of companies can be determined. Although the accuracy of predicting results from unstructured data compared to structured data with a predictability about 90% is less, but knowing and using the capacity of this type of information is very important so many academics and practitioners believe that quantitative disclosures alone are not effective for economic decision-making. In several ways, this research has increased knowledge of the field of accounting and data mining: First, the present research has tried to distance itself from the structured data extracted from financial statements, market data, and economic data, and for a major part of the analysis relies on the board' reports (unstructured data). Second, the results of this research showed that instead of relying on "numbers and figures", was relied on the analysis of "the text of reports" and predicted the financial distress of companies very accurately.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Acknowledgements

We would like to thank all those who contributed to the design and completion of the questionnaire by providing advice and participation.

Ethical Considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از اطلاعات متنی گزارش‌های فعالیت هیئت‌مدیره

مجتبی عالی فامیان^۱، محمد مرفوع^۲، محمدجواد سلیمی^۳ و ایمان رئیسی وانانی^۴

۱. گروه حسابداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. mojtaba.alifamian@yahoo.com **رایانامه:**

۲. نویسنده مسئول، گروه حسابداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. marfou@atu.ac.ir **رایانامه:**

۳. گروه حسابداری، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. j_salimi@atu.ac.ir **رایانامه:**

۴. گروه مدیریت عملیات و فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران. imanraeesi@atu.ac.ir **رایانامه:**

چکیده

اطلاعات مقاله

هدف: اهمیت بالای درماندگی مالی برای حیات اقتصادی کشور و هزینه‌های بالای فردی و اجتماعی آن، موضوع پیش‌بینی درماندگی مالی را به یک مسئله مهم برای استفاده‌کنندگان و ذی‌نفعان صورتهای مالی تبدیل کرده است. عمده پژوهش‌ها باتکیه بر اطلاعات مالی ساختاریافته و کمی صورتهای مالی درماندگی مالی را پیش‌بینی کردند؛ اما در این پژوهش تلاش شده است با استفاده از تکنیک متن‌کاوی و الگوریتم‌های یادگیری ماشین، از اطلاعات ساختار نیافته گزارش‌های هیئت‌مدیره جهت پیش‌بینی درماندگی مالی استفاده شود.

روش: به همین منظور گزارش هیئت‌مدیره ۱۰۰ شرکت بørsی در بازه زمانی ۱۳۹۰-۱۴۰۰ جمع‌آوری، با استفاده از زبان برنامه‌نویسی پایتون مورد متن‌کاوی (شامل مراحل پیش‌پردازش، استخراج ویژگی، انتخاب ویژگی و ...) قرار گرفتند و سپس مدل‌سازی آن با استفاده از الگوریتم‌های یادگیری ماشین در زبان برنامه‌نویسی پایتون انجام گرفت.

یافته‌ها: نتایج پژوهش حاکی از برتری دو روش مدل درخت تصمیم و ماشین بردار پشتیبان با کرنل شعاعی نسبت به سایر روش‌ها (شامل رگرسیون لجستیک، جنگل تصادفی، نزدیک‌ترین همسایگی و روش‌های ماشین بردار پشتیبان با کرنل‌های خطی، سیگموئید و چندجمله‌ای) بود.

نتیجه‌گیری: در واقع نتایج این پژوهش نشان داد که به‌جای توجه صرف بر اعداد و ارقام و نسبت‌های مشتق شده، می‌توان از تکنیک متن‌کاوی نیز جهت تجزیه و تحلیل و پیش‌بینی استفاده کرد و با تلفیق آن با نتایج حاصل از اطلاعات کمی می‌توان درماندگی مالی شرکت‌ها پیش‌بینی نمود.

نوع مقاله: مقاله پژوهشی.

تاریخ‌ها:

تاریخ دریافت: ۱۴۰۲/۱۲/۳

تاریخ بازنگری: ۱۴۰۳/۲/۲۳

تاریخ پذیرش: ۱۴۰۳/۲/۲۴

تاریخ انتشار برخط: ۱۴۰۴/۴/۳۱

واژه‌های کلیدی:

درماندگی مالی،

متن‌کاوی،

یادگیری ماشین،

پایتون.

طبقه‌بندی JEL:

G33, M41.

استناد: عالی فامیان، مجتبی؛ مرفوع، محمد؛ سلیمی، محمدجواد و رئیسی وانانی، ایمان (۱۴۰۴). پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از اطلاعات متنی

گزارش‌های فعالیت هیئت‌مدیره. *مجله دانش حسابداری*، ۱۶(۲)، ۱۷۶-۱۵۱. <https://doi.org/10.22103/jak.2024.22992.4017>



ناشر: دانشگاه شهید باهنر کرمان.

© مرفوع و همکاران.

۱- مقدمه

در نگرش‌های نوین به شرکت همچون یک واحد اجتماعی-اقتصادی، شرکت‌ها افزون بر کارکرد اقتصادی، دارای کارکردهای پررنگ در عرصه‌های غیراقتصادی از قبیل کمک به بازار کار و ظرفیت اشتغال، محیط‌زیست و حفظ یا تخریب آن، نسل‌های آینده و انتقال بدهی یا سرمایه به آنان و بسیاری از موارد حیاتی دیگر برای جامعه هستند؛ بنابراین، با توجه به پیامدهای اقتصادی و غیراقتصادی شرکت‌ها، درماندگی مالی و ورشکستگی شرکت‌ها و یا سلامت و کارایی آنها بر شاخص‌های اقتصادی-اجتماعی چه در سطح کلان و چه در سطح خرد تأثیر پر اهمیتی دارد. با توجه به این اثرات، پیش‌بینی این وضعیت‌ها می‌تواند در ایجاد آگاهی برای برخورد متناسب با وضعیت احتمالی آتی (به‌خصوص بعد اقتصادی) کمک شایانی نماید؛ چرا که استفاده از آن منجر به بهره‌گیری مناسب از فرصت‌های سرمایه‌گذاری و تخصیص بهتر منابع می‌شود. به‌این ترتیب که اولاً، ارائه هشدارهای لازم در خصوص وقوع درماندگی مالی این امکان را برای شرکت‌ها فراهم می‌کند که دست به اقدامات مقتضی بزنند و ثانیاً، سرمایه‌گذاران و اعتباردهندگان می‌توانند از طریق تشخیص فرصت‌های مطلوب سرمایه‌گذاری از فرصت‌های نامطلوب، منابع خود را در فرصت‌های مناسب سرمایه‌گذاری کنند (مهرانی و همکاران، ۱۳۸۴).

در حوزه درماندگی و ورشکستگی شرکت‌ها عمده مطالعات صورت گرفته شده با استفاده از داده‌های کمی و ساختاریافته گزارش‌ها انجام گرفته است. این در حالی است که امروزه بخش زیادی از اطلاعات از طریق افشاهای کیفی مانند گزارش‌ها و سندهای متنی همچون فایل‌ها و افشاهای اجباری مانند فرم‌های 10-k، 10-Q و 8-k اعلامیه‌های سود و دیگر بیانیه‌های منتشر شده، کنفرانس‌های خبری و رسانه‌های مالی، اخبار، یادداشت‌های تحلیلی، گزارش‌های مالی تحلیل‌گران، بیانیه‌ها و تفاسیر قانون‌گذاران مانند مقررات کمیسیون بورس اوراق بهادار، اخبار کلان اقتصادی و احساسات، پیام‌های اینترنتی و مطالب شبکه‌های اجتماعی که همگی نمونه‌ای از داده‌های مالی ساختار نیافته می‌باشند منتشر می‌شوند (گو^۱ و همکاران، ۲۰۱۶). علی‌رغم اهمیت اطلاعات کمی و ساختاریافته گزارش‌های مالی برای استفاده کنندگان، ارائه تصویر کامل از وضعیت شرکت صرفاً با استفاده از این اطلاعات امکان‌پذیر نیست. به همین دلیل از دیرباز استفاده کنندگان، تهیه کنندگان و مقررات‌گذاران اطلاعات حسابداری جهت درک بهتر اطلاعات کمی و ارائه اطلاعاتی که در قالب اعداد نمی‌گنجد به تهیه و ارائه اطلاعات نوشتاری در کنار اطلاعات کمی گزارش‌های مالی علاقه‌مند بوده‌اند، به‌طوری که امروزه بخش زیادی از گزارش‌های مالی به اطلاعات متنی اختصاص یافته است (هاجک^۲ و همکاران، ۲۰۱۳).

پس همان‌طور که اشاره شد پیش‌بینی دقیق وضعیت شرکت نیازمند گزارشی است که بتواند اطلاعاتی جامع و کامل در خصوص شرکت افشا و ارائه کند و یکی از این گزارش‌ها، گزارش هیئت‌مدیره به مجمع است. در این گزارش سالانه، مدیریت علاوه بر مرور و ارائه گزارش در خصوص فعالیت‌های صورت گرفته سال گذشته و نحوه انجام امور شرکت، اقدام به ترسیم چشم‌انداز و اهداف آینده شرکت می‌کند. از این رو این گزارش به‌عنوان یک سند مهم برای تحلیل‌گران و سرمایه‌گذاران که مایل به بررسی اصول مالی و عملکرد مدیریت این شرکت به‌منظور سرمایه‌گذاری و پیش‌بینی وضعیت آتی شرکت هستند است. در واقع اهمیت نوشتارهای کیفی گزارش‌های فعالیت هیئت‌مدیره از این جهت است که برخلاف ارقام کمی، به فعالان بازار سرمایه انتظاراتی را در خصوص عملکرد آینده شرکت منتقل می‌کند (بلک ریشنان^۳ و همکاران، ۲۰۱۰) و از طرفی با توجه به اینکه تجویز روشی در خصوص نحوه بیان اطلاعات نوشتاری وجود ندارد مدیران می‌توانند

¹ Guo² Hajek³ Balakrishnan

انتظارات عملکردی خود را با به کارگیری سبک بیان و انتخاب واژگان خاص به مشارکت کنندگان بازار سرمایه منتقل کنند. در نتیجه براین اساس و مطالعات تتلاک^۱ و همکاران (۲۰۰۸)، هنری^۲ (۲۰۰۸)، لوران و مک دونالد^۳ (۲۰۱۱) و هنری و لیون^۴ (۲۰۱۶) می توان گفت پیام های نوشتاری گزارش های مدیریتی دارای محتوای اطلاعاتی می باشند و استفاده از این اطلاعات جهت پیش بینی وضعیت شرکت نیازمند کد گشایی و پردازش اطلاعاتی آن است.

عمده پژوهش های صورت گرفته در این حوزه در ایران را می توان بر دو گروه کلی تقسیم کرد: گروه اول مطالعاتی هستند که از فرهنگ واژگان برای ارزیابی درماندگی مالی استفاده کردند و گروه دوم مطالعاتی هستند که به بررسی و سنجش ارتباط بین خوانایی (قابلیت فهم)، میزان شباهت افشا و طول متن افشا شده با درماندگی مالی پرداختند. این در حالی است که به دلیل مناسب نبودن این روش های برای تحلیل متون تجاری و مالی (لوران و مک دونالد، ۲۰۱۰)، این پژوهش تلاش کرده است با استفاده از روش متن کاوی محتوای اطلاعاتی گزارش های فعالیت هیئت مدیره را کد گشایی و در جهت پیش بینی درماندگی مالی استفاده کند. متن کاوی^۵ که تحت عناوین تجزیه و تحلیل هوشمند متن، کاوش داده های متنی و کشف دانش از متون نیز شناخته می شود، به فرایند استخراج اطلاعات و دانش جالب توجه و غیربدیهی از متن بدون ساختار اشاره دارد (گوپتا و لهال^۶، ۲۰۰۹). اما این روش پژوهش (متن کاوی) در ایران به دلیل پیچیدگی و دشواری فرایند و عدم دسترسی مناسب به منابع سخت افزاری کمتر مورد استفاده قرار گرفته است به گونه ای که در کشورمان و در رشته حسابداری، پژوهشی که درماندگی مالی را از طریق تجزیه و تحلیل متن گزارش ها ارزیابی کرده باشد، یافت نشد.

از این رو پژوهش حاضر به دنبال آن است که نخست داده های ساختار نیافته گزارش فعالیت هیئت مدیره را به داده های ساختار یافته تبدیل کند و سپس با استفاده از الگوریتم های یادگیری ماشین، مدلی برای پیش بینی درماندگی مالی ارائه کند. دوم اینکه در اکثر تحقیقات گذشته، دو مفهوم کاملاً متمایز درماندگی مالی و ورشکستگی به صورت مترادف و با مفهوم یکسان در نظر گرفته شده است. به طوری که از ماده ۱۴۱ قانون تجارت جهت شناسایی شرکت های درمانده و ورشکسته استفاده شده است؛ لذا، در این پژوهش نخست با مطالعه ادبیات پژوهش معیارهای درماندگی مالی مشخص و سپس شرکت ها دارای واجد شرایط وارد فرایند مدل سازی شدند لذا، سؤال اصلی پژوهش حاضر این است که آیا با تجزیه و تحلیل متن گزارش های فعالیت هیئت مدیره می توان به پیش بینی درماندگی مالی شرکت ها پرداخت؟

در قسمت های بعدی مقاله، به ترتیب مبانی نظری و پیشینه پژوهش، روش اجرای پژوهش و یافته های استنباطی ارائه خواهد شد و بخش پایانی مقاله نیز به بیان نتیجه گیری، پیشنهادها و محدودیت های پژوهش اختصاص یافته است.

۲- مبانی نظری و مروری بر پیشینه پژوهش

درماندگی مالی وضعیتی است که شرکت دارای مشکلات مالی باشد، در ادبیات سنتی، مصادیق مشکلات مالی شامل ناتوانی در پرداخت بدهی یا سود سهام ممتاز و پیامدهای بعدی آن شامل اضافه برداشت از سپرده های بانکی، از بین رفتن منافع اعتبار دهندگان و حتی ورود به فرایند ورشکستگی قانونی است (آلتمن^۷، ۱۹۶۸ و دیکین^۸، ۱۹۷۲)، چنین تعریفی از درماندگی مالی بر مبنای چارچوب نظری مربوط به جریان های نقدی یا نقد شوندگی دارایی ها است. به اعتقاد بیور^۹ (۱۹۶۶)

¹ Tetlock

² Henry

³ Loughran & McDonald

⁴ Henry & Leone

⁵ Text Mining

⁶ Gupta & Lehal

⁷ Altman

⁸ Deakin

⁹ Beaver

واحد تجاری شبیه به مخزنی است که شامل جریان‌های نقدی ورودی و خروجی است و شرکتی که دچار درماندگی مالی است، شبیه به مخزنی است که آب آن در حال تخلیه است. دو مشکل اصلی که شرکت‌های درمانده با آن مواجه هستند عبارت‌اند از کمبود نقدینگی و وجود تعهدات زیاد. به عبارت دیگر، کمبود نقدینگی در هنگام درماندگی مالی، به طور موقت شرکت را در ایفای تعهدات خود با مشکل روبرو می‌سازد و سعی در جبران این کمبود از طریق فروش دارایی‌ها و اخذ تسهیلات چیزی جز کاهش ظرفیت و عملکرد شرکت و افزایش اهرم مالی برای شرکت به همراه ندارد.

فرایند درماندگی مالی یک فرایند پویا و به‌طور کلی طولانی است که بر ساختار سرمایه شرکت، سیاست‌های سرمایه‌گذاری و عملکرد تأثیر می‌گذارد (کال^۱، ۲۰۰۲). هنگامی که یک شرکت دچار درماندگی مالی می‌شود، به این معنی است که شرکت در حال ورود به یک منطقه خطرناک است. با تشخیص درماندگی مالی، شرکت باید اقداماتی را در عملیات خود اتخاذ کند تا از اثرات مخرب درماندگی مالی جلوگیری کند. تشخیص درماندگی مالی در مرحله مقدماتی و اقدامات اصلاحی فوری باعث می‌شود شرکت در اسرع وقت از منطقه خطر خارج شود و با زیان‌های جزئی بر این فرایند غلبه کند. در مقابل، تشخیص دیر هنگام درماندگی مالی می‌تواند آسیب را عمیق‌تر کند و ممکن است برای نجات شرکت خیلی دیر باشد و در نتیجه ورشکستگی اجتناب‌ناپذیر خواهد بود. در شرایط رقابتی امروز دلایلی همچون کمبود منابع، ضعف مدیریت در به کارگیری منابع موجود و همچنین وضعیت نابسامان اقتصادی کشورها سبب شده است که شرکت‌ها در ایفای به موقع تعهدات خود به مشکل بخورند و در آستانه حذف از این رقابت قرار بگیرند و هزینه‌های زیادی را بر شرکت، مؤسسات مالی، سهام‌داران، مدیران، کارکنان و در سطحی کلان بر کل اقتصاد و دولت تحمیل کنند؛ از این رو بسیاری از پژوهشگران به دنبال آن هستند که با استفاده از نسبت‌های مالی و سایر متغیرهای صورت‌های مالی، درماندگی مالی شرکت‌ها را پیش‌بینی کنند و در مراحل مختلف درماندگی مالی با تشخیص وضعیت شرکت و جلوگیری از اتلاف منابع و تجدید ساختار مدیریت، شرایط شرکت جهت ادامه رقابت را فراهم کنند.

استفاده از شاخص‌های مالی جهت پیش‌بینی درماندگی مالی اولین بار توسط **فیتزپاتریک^۲** (۱۹۳۲) استفاده شد. چند سال بعد **مروین^۳** (۱۹۴۲) در پژوهش خود نشان داد میانگین شاخص‌های مالی بین شرکت‌های درمانده و غیر درمانده متفاوت است. اما نقطه عطف تحقیقات مربوط به درماندگی مالی به نیمه دوم سال ۱۹۶۰ بر می‌گردد. زمانی که **بیور^۴** (۱۹۶۶) از نسبت‌های مالی برای پیش‌بینی درماندگی مالی استفاده کرد. با این وجود مدل ارائه شده **آلتن^۵** (۱۹۶۸) که مبنای توسعه بسیاری از مدل‌های دیگر ورشکستگی است را می‌توان یک کار پیش‌گام در حوزه جهانی در نظر گرفت. با گذشت زمان و پیشرفت علم آمار تکنیک‌های جدیدی همچون **لاجیت** و **پرویت** جهت پیش‌بینی درماندگی مالی توسعه یافته که از معروف‌ترین پژوهشگران این حوزه (**لاجیت**) می‌توان به **اولسون^۵** (۱۹۸۰) و حوزه **پرویت** می‌توان به **زمیجوسکی^۶** (۱۹۸۴) (**پرویت**) اشاره کرد. در سال‌های بعد مدل **مرتون^۷** (۱۹۷۳) به عنوان یک مدل مبتنی بر مبنای نظری، جایگزینی برای رویکرد مبتنی بر تجزیه و تحلیل داده‌های تاریخی ارائه شد که علی‌رغم کاربرد عملی گسترده، ساختار پیشرفته ریاضی این مدل‌ها به عنوان یک محدودیت اساسی برای آنان در نظر گرفته می‌شد. توسعه حوزه علمی تحقیقات هوش مصنوعی و توسعه سریع فن آوری کامپیوتر تأثیر قابل بر توسعه مدل‌ها پیش‌بینی شکست شرکت‌ها داشته است (**آلتن و همکاران**، ۲۰۱۷).

¹ Kahl

² Fitzpatrick

³ Merwin

⁴ Beaver

⁵ Ohlson

⁶ Zmijewski

⁷ Merton

به طوری که امروزه حجم بسیاری از مقالات در حوزه مالی و حسابداری با استفاده از این ابزارها به بررسی عوامل، اثرات و پیش‌بینی درماندگی پرداختند.

تمرکز عمده پژوهش‌های که تاکنون در خصوص درماندگی مالی صورت گرفته است بر داده‌های کمی و نسبت‌های مالی بوده است. این در حالی است استفاده از نسبت‌های مالی در پیش‌بینی درماندگی مالی به دلیل تاریخی بودن اطلاعات و تحت تأثیر بودن رویه‌های حسابداری همواره مورد انتقاد قرار گرفته است و بیان می‌شود اعداد مستخرج از صورت‌های مالی توانایی کافی جهت پیش‌بینی درماندگی مالی شرکت‌ها را ندارد. در واقع علی‌رغم اهمیت بسیار زیاد اطلاعات کمی گزارش‌های مالی، این اطلاعات به‌تنهایی تصویری کاملی از وضعیت شرکت برای استفاده‌کنندگان فراهم نمی‌کند. مطالعات گذشته بسیاری بر اهمیت اطلاعات نوشتاری موجود در گزارش‌های مدیریتی تأکید کردند و نشان دادند که این اطلاعات و رویکردهای قضاوتی زبان، تأثیر بسزایی در تصمیم‌گیری‌های سرمایه‌گذاران در خصوص خرید و فروش سهام، اعطای اعتبار، کاهش فاصله انتظارات حسابرسی و نظایر آن دارد. (کلت ورتی و جونز^۱، ۲۰۰۳ و مهرانی و نونهال نهر، ۱۳۹۲). مهم‌ترین بارزه اصلی این اطلاعات، امکان انتقال انتظارات در خصوص عملکرد آینده شرکت به مشارکت‌کنندگان بازار است (بلک ریشنان و همکاران، ۲۰۱۰). در واقع به دلیل اینکه در خصوص نحوه ارائه اطلاعات نوشتاری تجویزی از سوی سازمان‌های ذی‌صلاح صورت نگرفته است مدیران با به کارگیری سبک بیان و گزینش واژگان خاص، انتظارات عملکردی خود را به بازار سرمایه و اعتباردهندگان انتقال دهند. دو رویکرد کلی در خصوص استفاده مدیریت از این اطلاعات جهت شکل‌دهی انتظارات مشارکت‌کنندگان بازار سرمایه از عملکرد خود وجود دارد: ۱- رویکرد اطلاعاتی ۲- رویکرد فرصت‌طلبانه.

در رویکرد اطلاعاتی مدیران با به کارگیری از شرح نوشته‌های کیفی، به دنبال سازگار کردن انتظارات سرمایه‌گذاران با انتظارات خود و علامت‌دهی مناسبی در خصوص عملکرد آینده می‌باشند. در واقع مدیران با انتخاب و تکرار واژه‌هایی خاص در متن اطلاعات بر جنبه متقاعدکنندگی تمرکز دارند و به دنبال آن هستند که با انتخاب لحن مساعد و تأکید بر جوانب خاص، در خصوص عملکرد خود به بازار سرمایه اطلاع‌رسانی کنند. ازجمله پژوهش‌هایی که در نتایج خود بر این جنبه تأکید کردند می‌توان به پژوهش‌های هنری (۲۰۰۸)، لوران و مک‌دونالد^۲ (۲۰۱۱) و هنری و لیون^۳ (۲۰۱۶) اشاره کرد. در طرف مقابل بر اساس رویکرد فرصت‌طلبانه مدیران به دلیل آزادی عمل و اختیار در انتخاب نوع واژگان می‌توانند به دنبال مقاصد و منافع شخصی خود باشند و امکان متهم کردن مدیریت به رفتار غیراخلاقی، به دلیل استفاده بیشتر از واژگان مثبت یا منفی، بسیار سخت است. ازاین‌رو، به کارگیری رویکرد فرصت‌طلبانه، راه‌حلی کم‌هزینه جهت نیل به مقاصد شخصی و تحت قرارداد استفاده‌کنندگان است. ازجمله پژوهش‌هایی که به شواهدی در خصوص وجود رویکرد فرصت‌طلبانه در نحوه نوشتاری گزارش‌های مدیریتی دست پیدا کردند می‌توان به پژوهش یانگ و لیو^۴ (۲۰۱۷) و لوران و همکاران (۲۰۰۹) اشاره کرد.

ازاین‌رو اطلاعات کیفی و متنی در کنار اطلاعات کمی گزارش‌های مالی مورد توجه تهیه‌کنندگان، استفاده‌کنندگان و قانون‌گذاران اطلاعات حسابداری است و اعتقاد بر این است که انتقال این منابع اطلاعاتی به چارچوب‌های قابل استفاده،

¹ Clatworthy & Jones

² Loughran & Mcdonald

³ Henry & Leone

⁴ Yang & Liu

ارزش بالقوه تجاری بالایی دارد. اما مشکلی که در ارتباط با بهره‌برداری از این اطلاعات کیفی وجود دارد حجم بالای آن نسبت به اطلاعات کمی است؛ لذا نیاز است که نخست سرمایه‌گذاران اقدام به کدگذاری از این اطلاعات نموده و سپس آنها رو مورد پردازش قرار دهند. یکی از ابزارهایی که برای این منظور مورد استفاده قرار می‌گیرد متن کاوی است. لی^۱ (۲۰۱۰) بیان کرد دو رویکرد در ارتباط با تجزیه و تحلیل زبان به کاررفته در متن گزارش‌های شرکت‌ها وجود دارد. رویکرد اول همانند مطالعات قبلی صورت گرفته در حوزه‌های روان‌شناسی و زبان‌شناسی بر جعبه واژه‌ها^۲ تکیه می‌کند که درواقع لیستی از واژه‌های از پیش مشخص شده است. در این رویکرد فرض بر ارتباط این لیست از پیش تعیین شده با فضای احساسی موجود در متن، همچون خوش‌بینی، فریبه‌کاری، بدبینی یا نااطمینانی است. باین حال لوران و مک دونالد (۲۰۱۰) رویکرد جعبه واژه‌ها را مناسب تحلیل متون تجاری و مالی نمی‌دانند (هامفریس^۳ و همکاران، ۲۰۱۱). رویکرد دوم به جای استفاده از لیستی از واژه‌های از پیش مشخص شده، به دنبال آن است که با استفاده از روش‌های آماری، داده‌ها اهمیت هر یک از واژه‌ها را تعیین کنند. رویکرد دوم، برخلاف رویکرد اول که نیازمند لیستی از واژه‌های تخصصی حوزه مورد بررسی است نیازی به لیست تخصصی ندارد؛ لذا به دلیل مزایایی که رویکرد دوم دارد پژوهش حاضر از این رویکرد برای تجزیه و تحلیل متن (متن کاوی) استفاده کرده است. در میان گزارش‌های مختلف شرکت، گزارش فعالیت هیئت‌مدیره از اهمیت بسزایی برخوردار است (کول و جونز، ۲۰۰۵).

گزارش فعالیت هیئت‌مدیره به عنوان گزارشی جامع در خصوص فعالیت سالانه شرکت، افزون بر ارائه اطلاعات تکمیلی در خصوص اطلاعات مندرج در صورت‌های مالی، فرصتی برای مدیران جهت اطلاع‌رسانی جنبه‌های مهم عملکردی خود به سهام‌داران ایجاد می‌کند. از آنجایی که این گزارش یک گزارش گذشته، حال و آینده‌نگر است به عنوان ابزاری برای پیش‌بینی وضعیت شرکت مورد استفاده قرار می‌گیرد (برزگری خانقاه و همکاران، ۱۳۹۸).

۳- پیشینه پژوهش

جو و شین^۴ (۲۰۱۶) در پژوهشی تحت عنوان پیش‌بینی درماندگی مالی بر اساس آنالیز داده‌های بزرگ نشان دادند که متغیرهای احساسی استخراج شده از روزنامه‌های اقتصادی بر پیش‌بینی ورشکستگی شرکت‌ها اثرگذار است. و در واقع لغات احساسی منفی نشان‌دهنده وضعیت اقتصادی ضعیف آن شرکت است و احتمال ورشکستگی آن را بالا می‌برد. هاجک و هنریگس^۵ (۲۰۱۷) در پژوهشی با ترکیب ۳۲ متغیر مالی با متغیرهای زبانی (متنی) به روش یادگیری ماشین به این نتیجه رسیدند که ترکیب متغیرهای کمی با زبانی نسبت متغیرهای کمی توان کمتری برای پیش‌بینی گزارشات متقلبانه دارند. وانگ^۶ و همکاران (۲۰۱۷) به این نتیجه رسیدند اطلاعات متنی مستخرج شده از گزارش‌های مدیریتی به همراه اطلاعات مالی امکان شناسایی شرکت‌های متقلب را افزایش می‌دهد. همچنین نتایج آنان حاکی از برتری روش‌های مبتنی بر یادگیری ماشین در تشخیص شرکت‌های متقلب و روش‌های مبتنی بر تحلیل نیز در تشخیص شرکت‌های غیر متقلب بود.

مای^۷ و همکاران (۲۰۱۸) در پژوهشی به بررسی استفاده از یادگیری عمیق بر پیش‌بینی ورشکستگی شرکت‌ها پرداختند. نتایج پژوهش حاکی از آن بود که استفاده از یادگیری عمیق قدرت پیش‌بینی ورشکستگی شرکت‌ها را بالا می‌برد. تانگ^۸ و همکاران (۲۰۲۰) نشان دادند فاکتورهای متنی گزارشات مدیریتی (تعداد کلمات مثبت، منفی، طول متن و ...) می‌تواند سبب

¹ Li

² Bag Of The Words

³ Humpherys

⁴ Jo&Shin

⁵ Hajek & Michalak

⁶ Wang

⁷ Mai

⁸ Tang

افزایش دقت پیش‌بینی درماندگی مالی شود. **سریدهاران^۱ و همکاران (۲۰۲۰)** دقت پیش‌بینی الگوریتم‌های یادگیری ماشین با یکدیگر مقایسه کردند. نتایج پژوهش آنان حاکی از آن بود که علی‌رغم قابل قبول بودن دقت تمام الگوریتم‌ها، شبکه عصبی بیش‌ترین دقت را در پیش‌بینی درماندگی مالی دارد. **لی^۲ و همکاران (۲۰۲۱)** با استفاده از روش یادگیری عمیق، اقدام به ایجاد یک فرهنگ لغت جهت پیش‌بینی درماندگی مالی نمودند.

هاونگ^۳ و همکاران (۲۰۲۳) با استفاده از تحلیل احساسات گزارشات سالانه مدیران به پیش‌بینی درماندگی مالی پرداختند. نتایج تحلیل آنان نشان داد امتیاز حاصل از تحلیل احساسات گزارشات تجزیه و تحلیل مدیران در مقایسه با گزارشات حسابرسی بسیار خوش‌بینانه است و کاربرد تحلیل احساسات هر دو گزارش می‌تواند توانایی پیش‌بینی درماندگی مالی را افزایش دهد. **لی و وانگ^۴ (۲۰۲۳)** با ترکیب داده‌های مالی و داده‌های متنی گزارش‌های سالانه اقدام به پیش‌بینی درماندگی مالی با استفاده از روش‌های مختلف پرداختند. نتایج آنان حاکی از برتری روش یادگیری عمیق بر روش سنتی یادگیری ماشین بود. **سانگ^۵ و همکاران (۲۰۲۴)** با ترکیب شاخص‌های مالی، شاخص‌های مسئولیت اجتماعی، مدیریتی و متنی الگوی جدیدی برای پیش‌بینی درماندگی مالی ارائه دادند و نشان دادند شاخص‌های متنی مهم‌ترین نقش را در تکمیل شاخص‌های مالی سنتی پیش‌بینی درماندگی مالی ایفا می‌کنند و سیگنال‌های درماندگی مالی آنها در مقایسه با شاخص‌های مدیریت و مسئولیت اجتماعی زودتر ظاهر می‌شوند.

بت‌شکن و همکاران (۱۳۹۷) در پژوهشی نخست با استفاده از الگوریتم‌های تصمیم‌گیری دیمتل اقدام به انتخاب بهترین نسبت‌های مالی پیش‌بینی‌کننده درماندگی مالی نموده و سپس با استفاده از نسبت‌های مالی منتخب، درماندگی مالی شرکت را پیش‌بینی کردند. نتایج پژوهش آنان نشان داد دقت مدل ارائه شده در سطح ۹۵ درصد اختلاف معناداری با دقت مدل‌های آلتمن و رگرسیون لجستیک دارد. **ابراهیمی سرو علیا و همکاران (۱۳۹۷)** در پژوهشی نخست با استفاده از روش دلفی اقدام به شناسایی متغیرهای مالی پیش‌بینی‌کننده درماندگی مالی نمودند و سپس با استفاده از مدل کاکس اقدام به پیش‌بینی پویای درماندگی مالی شرکت‌ها کردند که بر اساس نتایج صحت و دقت تخمین با استفاده از منحنی ROC و Brier مورد تأیید قرار گرفت. **بولو و احمدوند (۱۳۹۸)** در پژوهشی به ارائه الگویی برای پیش‌بینی نکول شرکتی پرداختند. نتایج پژوهش آنان نشان داد که در بورس اوراق بهادار تهران صرفاً نسبت‌های حسابداری به‌عنوان متغیرهای نکول مورد توجه قرار گرفته است و سایر متغیرهای بالقوه در پیش‌بینی نکول نقشی ندارند.

هروی دستجردی و همکاران (۱۳۹۷) خطر تقلب مدیران با استفاده از روش داده کاوی را مورد ارزیابی قرار دادند. نتایج پژوهش آنان نشان داد که رگرسیون Lasso با استفاده از واژه‌های هیئت‌مدیره می‌تواند با دقت ۸۹ تا ۹۱ درصد خطر بالای تقلب در شرکت‌ها را پیش‌بینی کند. در پژوهشی، **پله و همکاران (۱۳۹۸)** نشان دادند لحن گزارش فعالیت هیئت‌مدیره بر سودآوری و جریان نقد عملیاتی سال آتی تأثیر منفی معناداری دارد. **زارع بهمیری و همکاران (۱۳۹۸)** نشان دادند رابطه مثبت و معناداری بین قدرت مدیران با لحن اعلان سود مدیران وجود دارد و درصد اعضای غیرموظف هیئت‌مدیره سبب تضعیف این رابطه می‌شود. **حاجیها و صابری روجی (۱۴۰۰)** در پژوهشی نشان دادند که کلمات پر تکرار گزارش حسابرسی می‌تواند در پیش‌بینی درماندگی مالی مؤثر واقع شود. **نمازی و ابراهیمی (۱۴۰۰)** نشان دادند که مدل ماشین

¹ Sreedharan² Li³ Huang⁴ Li & Wang⁵ Song

بردار پشتیبان به نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۰/۵۵، ۰/۵۶، ۰/۷۸ و ۰/۴۳ بر روی مجموعه آموزشی دست یافت و در نهایت نتایج حاصل از مجموعه آزمایشی کنار گذاشته شده با مدل پیشنهادی به نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۰/۵۰، ۰/۵۰، ۰/۶۸ و ۰/۴۰ منجر شد.

حسینی و حاجیان نژاد (۱۴۰۱) با بررسی شکل توزیع نسبت های مالی استفاده شده در مدل آلتمن نشان دادند که آگاهی از شکل توزیع آماری متغیرهای x_1 ، x_2 ، x_3 و x_4 تأثیری در آگاهی از توزیع مقادیر متغیر Z-score ندارد. **هاشمی گل سفیدی و همکاران (۱۴۰۱)** کاربرد یادگیری ماشین در ارائه الگوی پیش‌بینی ورشکستگی را بررسی کردند و نشان دادند نسبت درآمد خالص به کل دارایی و نسبت ارزش بازار حقوق صاحبان سهام به کل ارزش بازار بیشترین تأثیر در پیش‌بینی درماندگی مالی را دارند. **عباسیان و همکاران (۱۴۰۲)** با به کارگیری اطلاعات مبتنی بر شبکه های مالی و روش ترکیبی درخت تصمیم تقویت گرادیان رویکردی نوین در پیش‌بینی درماندگی مالی ارائه کردند. نتایج آنان نشان داد عملکرد مدل تصمیم تقویت گرادیان از نظر دقت و هم از نظر خطای نوع یک عملکرد بهتری در مقایسه با دو مدل K نزدیک ترین همسایه و رگرسیون لجستیک داشته است. **رحیمی و همکاران (۱۴۰۳)** در پژوهشی اقدام به ارائه رویکردی برای پویاسازی پیش‌بینی درماندگی مالی با استفاده از انواع معیارهای مالی نمودند. نتایج پژوهش آنان نشان داد مدل تحلیلی پوششی داده های پویا مبتنی بر بدترین عملکرد (WPF-DEA) توان ارزیابی ناکارایی واحدهای مورد بررسی را دارا است.

۴- جامعه آماری و روش نمونه‌گیری

- جامعه آماری پژوهش حاضر کلیه شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران در بازه زمانی ۱۳۹۰ تا ۱۴۰۰ است. در این پژوهش حجم نمونه براساس معیارهای زیر و با استفاده از حذف سیستماتیک تعیین شده است:
۱. جزء شرکت‌های مالی، (مانند بانک‌ها و مؤسسات مالی) و شرکت‌های سرمایه‌گذاری یا شرکت‌های واسطه‌گری مالی نباشد. دلیل این موضوع، ماهیت متفاوت عملیات شرکت‌های مذکور است (افلاطونی، ۱۳۹۲).
 ۲. طی سال‌های مورد بررسی تغییر فعالیت یا تغییر سال مالی نداده باشد؛ زیرا استفاده از داده‌ها با سال‌های مالی متفاوت تفسیر نتایج تحقیق را مشکل خواهد کرد.
 ۳. دوره مالی آن‌ها منتهی به پایان اسفندماه باشد. دلیل این شرط، خشی کردن تأثیر چرخه‌های تجاری مؤثر بر عملکرد و وضعیت مالی شرکت‌ها است (افلاطونی، ۱۳۹۲).
 ۴. اطلاعات مالی آن قابل دسترس باشد؛ زیرا بدون وجود داده‌ها عملاً هیچ کاری نمی‌توان انجام داد (افلاطونی، ۱۳۹۲).
 ۵. در طی دوره پژوهش توقف عملیاتی بیش از شش ماه نداشته باشد؛ زیرا وقفه معاملاتی می‌تواند نتایج آزمون را دچار انحراف سازد.
 ۶. گزارش پی‌دی‌اف هیئت‌مدیره شرکت با نرم‌افزارهای موجود قابل تبدیل به ورد باشند. با توجه به این محدودیت‌ها تعداد ۱۰۰ شرکت برای انجام پژوهش انتخاب شد.

برای جمع‌آوری داده‌های موردنیاز از روش اسناد کاوی استفاده شد و داده‌های مالی شرکت‌ها از نرم‌افزار رهاورد نوین استخراج شده و گزارش‌های هیئت‌مدیره به مجمع عمومی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران نیز از سامانه جامع اطلاع‌رسانی ناشران دانلود و بررسی شده است. در این پژوهش از ابزارهای مختلفی برای تجزیه و تحلیل استفاده شده است. برای تبدیل فایل پی‌دی‌اف گزارش‌های هیئت‌مدیره به مجمع عمومی صاحبان سهام

به ورد از ربات‌های موجود در این حوزه استفاده شده است. همچنین به منظور متن‌کاوی گزارش‌های هیئت‌مدیره و استفاده از تکنیک یادگیری ماشین و نرم‌افزار پایتون^۱ نسخه ۳.۱۲.۳ استفاده شده است.

۵- چارچوب کلی پژوهش

این مطالعه با هدف شناسایی متغیرها و ویژگی‌های کیفی اثرگذار در پیش‌بینی درماندگی مالی و منظور نمودن آن‌ها در یک الگو جهت پیش‌بینی، صورت می‌گیرد در چندین مرحله به شرح ذیل انجام پذیرفت:

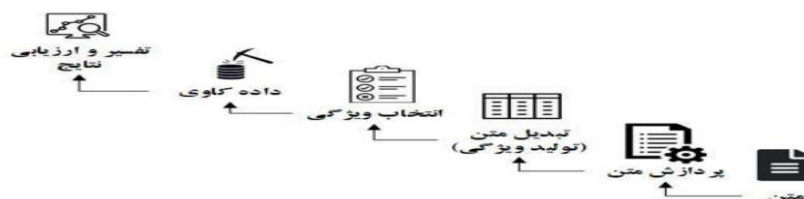
۵-۱- مرحله یک: جمع‌آوری اطلاعات مربوط به شرکت‌های درمانده

داده‌های پژوهش از طریق گزارش‌ها و اطلاعیه‌های منتشر شده در شبکه کدال، بانک‌های اطلاعاتی سازمان بورس و اوراق بهادار و بانک اطلاعاتی شرکت‌های پردازش اطلاعات مالی جمع‌آوری می‌شود. در این مرحله، نخست برخلاف اکثر پژوهش‌های پیشین صورت گرفته که دو مفهوم کاملاً متمایز درماندگی مالی و ورشکستگی را به صورت مترادف و با مفهوم یکسان در نظر گرفته‌اند و مشمول ماده ۱۴۱ بودن را شاخص درماندگی در نظر گرفته‌اند این پژوهش شرکت‌هایی که حداقل یکی از معیارهای زیر را داشته باشد به عنوان شرکت درمانده در نظر گرفته خواهد گرفته است:

۱. سود خالص یا سود عملیاتی قبل از مالیات آن، حداقل سه سال متوالی منفی باشد. (گیلبرت^۲ و همکاران، ۱۹۹۰)
۲. بیش از دو سال متوالی، زیان انباشته و عملکرد ضعیف داشته باشد (گیلبرت و همکاران، ۱۹۹۰)
۳. ارزش دفتری حقوق صاحبان سهام برای سه سال متوالی کمتر از سرمایه اسمی آن باشد (گیلبرت و همکاران، ۱۹۹۰).
۴. سود قبل از بهره، مالیات و استهلاک برای دو سال متوالی کمتر از هزینه‌های مالی آن بوده یا نسبت سود قبل از بهره، مالیات و استهلاک به هزینه‌های مالی آن کمتر از ۸۰ درصد باشد (پیندادو^۳ و همکاران، ۲۰۰۸).
۵. سود نقدی آن سه سال متوالی حداقل کاهش ۴۰ درصدی نسبت به سال قبل داشته باشد (دنيس^۴ و دنيس، ۱۹۹۵)
۶. خالص جریان نقدی عملیاتی آن، دو یا سه سال مالی متوالی منفی باشد (پلات^۵، ۲۰۰۶).
۷. بازده سهام و رشد فروش آن دو سال مالی متوالی منفی باشد (اپلر و تیتمن^۶، ۱۹۹۴).

۵-۲- مرحله دوم: متن‌کاوی

در این مرحله گزارش‌های شرکت‌های عضو نمونه با استفاده از ربات‌های موجود در این حوزه از پی‌دی‌اف^۷ به ورد^۸ تبدیل شد. سپس شرکت‌هایی که گزارش‌های آن‌ها پس از تبدیل شدن به ورد به دلایل مختلف (همچون به هم چسبیده و قابل خواندن نبودن واژه‌ها و یا تبدیل شدن به حروف و علائم ناخوانا و غیرقابل تشخیص) از کیفیت لازم برخوردار نبودند از نمونه حذف شدند. پس از این مرحله گزارش‌های فعالیت هیئت‌مدیره مورد متن‌کاوی قرار گرفت. متن‌کاوی در برگیرنده مراحل مختلفی است که می‌توان پیش‌پردازش، استخراج ویژگی و انتخاب ویژگی را به عنوان مراحل اصلی آن عنوان کرد:



شکل ۱. مراحل متن‌کاوی

¹ Python

² Gilbert

³ Pindado

⁴ Denis

⁵ Platt

⁶ Opler & Titman

⁷ Pdf

⁸ Word

۵-۲-۱- پیش‌پردازش متن

باتوجه به اهمیت پیش‌پردازش متن در فرایند متن‌کاوی از چندین مرحله جهت پیش‌پردازش و حذف داده‌های اضافی و یکسان‌سازی استفاده خواهد شد:

۱. نرمال‌سازی متن (Normalize)
۲. تبدیل متن به جملات و سپس به کلمات (Tokenization)
۳. حذف کلمات توقف (stop words)
۴. ریشه‌یابی لغات (stemming word)
۵. ذخیره تمامی کلمات و حذف کلمات و کاراکترهای زائد.

۵-۲-۲- فرایند نرمال‌سازی

در فرایند نرمال‌سازی همه متون از طریق ویرایشات و اصلاحات رسم‌الخطی استاندارد و یکپارچه می‌شوند.

۵-۲-۳- حذف کلمات توقف

منظور از کلمات توقف، کلماتی است که از لحاظ معنایی دارای اهمیت کمی بوده؛ ولی در جملات به کرات استفاده می‌شود و معمولاً در جملات، کاربرد ربطی دارند مثل «اگر»، «و»، «با». این نوع کلمات به دلیل تعدد بسیار و بار معنایی کم، در اغلب فعالیت‌های حوزه پردازش زبان طبیعی در فاز پیش‌پردازش حذف می‌شوند.

۵-۲-۴- ریشه‌یابی لغات

ریشه‌یابی لغات به گونه‌ای است که کلماتی که دارای ریشه مشابهی هستند به عنوان یک لغت در نظر گرفته خواهد شد. به عنوان مثال، دو عبارت «فروش» و «فروش‌ها» به عنوان یک عبارت مشابه تلقی خواهد شد.

۵-۲-۵- حذف کلمات و کاراکترهای زائد

در این مرحله تمامی کاراکترهای اضافی همچون (*، {، }، @، ؟، + و...) حذف می‌گردند.

۵-۳- مرحله سوم: انتخاب ویژگی

پس از پیش‌پردازش گزارش‌های هیئت‌مدیره در مرحله دوم از ماتریس TF-IDF و یا سایر ماتریس‌ها جهت انتخاب ویژگی استفاده شد که به صورت اختصار نحوه عمل ماتریس TF-IDF توضیح داده شده است: ماتریس TF-IDF (تکرار کلمه × تکرار سند معکوس) بر اساس فرمول زیر امتیاز هر کلمه را برای مرحله انتخاب ویژگی مشخص می‌کند.

$$TF - IDF_{w_i,j} = tf_{i,j} \times \log\left(\frac{N}{df}\right)$$

که در آن $tf_{i,j}$ تعداد تکرار کلمه w_i در متن d_j ، df تعداد متونی که در برگیرنده کلمه w_i بوده، N تعداد متون مورد بررسی و $w_{i,j}$ نیز وزن واژه w_i در متن j است.

در این روش تکرار کلمات در سند نشان‌دهنده اهمیت اصطلاح در اسناد است و تکرار سند در این اصطلاح (df) درصد اسناد حاوی این اصطلاح (نشان‌دهنده میزان اهمیت در کل متن است. مقدار df پایین نشان‌دهنده منحصر به فرد بودن این اصطلاح در اسناد است چرا که در بسیاری از اسناد ظاهر نمی‌شود. از این رو در این پژوهش به منظور وزن‌دهی از idf

به جای df استفاده خواهد شد؛ لذا وزن بالا در روش tf×idf تکرار زیاد یک کلمه در یک سند و اسناد کم حاوی این متن را نشان خواهد داد. در این مرحله کلماتی که بیشترین امتیاز را دارند ویژگی‌ها را مشخص خواهند کرد.

۴-۵- مرحله چهارم: مرحله آموزش مدل و ارزیابی و مقایسه نتایج

پس از این مرحله، از ترکیب ماتریس کلمات و داده‌های مالی منتخب و درماندگی یا عدم درماندگی شرکت‌ها ماتریس ورودی مدل تشکیل خواهد شد. در قدم بعدی، الگوریتم‌های یادگیری ماشین برای ارائه مدل پیش‌بینی اجرا خواهد شد. در پژوهش حاضر برای تشخیص درماندگی مالی از رابطه زیر استفاده خواهد شد:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{32} & \dots & X_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \times \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ \vdots \\ B_n \end{bmatrix}$$

متغیر وابسته (Y) به صورت یک ماتریس $m \times 1$ است که اعضای آن، برای شرکت‌هایی که شاخص درماندگی مالی در «تقسیم‌بندی اولیه» برای آنها بالا تشخیص داده شده است معادل یک و برای سایر شرکت‌ها معادل صفر است. همچنان ماتریس متغیرهای توضیحی (X) یک ماتریس $m \times n$ است که در آن m معادل تعداد شرکت سال‌ها و n معادل تعداد واژه‌ها و شاخص‌های مالی است که به عنوان متغیرهای توضیح‌دهنده شاخص درماندگی مالی، استفاده شده و باید از بین آنها واژه‌ها و شاخص‌هایی که بیشترین قدرت را در تبیین درماندگی مالی دارند، مشخص شوند. هر عضو این ماتریس که محل تلاقی یک سطر با یک ستون است، فراوانی نسبی (TF-IDF) یک واژه را نشان می‌دهد. این پژوهش پس از تشکیل ماتریس مذکور، وارد فاز آماری و مدل‌سازی با الگوریتم‌های یادگیری ماشین شده است.

۵-۵- یادگیری ماشین^۱

یادگیری ماشین از زیرمجموعه‌های هوش مصنوعی است که این امکان را برای سیستم‌ها فراهم می‌کند تا به صورت خودکار یادگیری و پیشرفت داشته باشند. تمرکز اصلی یادگیری ماشین بر توسعه برنامه‌های رایانه‌ای است تا بتوانند به داده‌ها دسترسی پیدا کرده و از آنها برای یادگیری خود استفاده کنند. در واقع هدف این روش کشف الگوهای پنهان و دستیابی به تصمیمات و نتایج قابل اطمینان و تکرارپذیر از طریق یادگیری روابط و روندهای گذشته است (هکلینگ، ۲۰۱۷).

پس از مدل‌سازی، به منظور ارزیابی قدرت پیش‌بینی مدل ارائه شده داده‌ها به دودسته آموزش و ارزیابی تقسیم خواهند شد. برای این منظور نخست یادگیری مدل پژوهش از طریق داده‌های آموزش انجام و سپس داده‌های ارزیابی به منظور محاسبه دقت و صحت الگوریتم‌ها روی داده‌هایی که تاکنون مشاهده نکرده است، استفاده خواهد شد. قابل ذکر است به منظور ارزیابی مناسب نیاز است که اجرای الگوریتم‌ها چندین بار تکرار شود؛ لذا، زمانی که یک مجموعه داده در اختیار است به منظور ارزیابی نهایی نیاز است که بخشی از آن کنار گذاشته شود و از مابقی داده‌ها جهت یادگیری استفاده شود و مجدداً دو مجموعه را تغییر داد و دوباره مدل مورد ارزیابی قرار گیرد. اعتبارسنجی ۱۰ قسمتی یکی از روش‌هایی است که برای این منظور مورد استفاده قرار می‌گیرد در این روش مجموعه داده‌ها به صورت تصادفی به K قسمت مساوی، تقسیم می‌گردد. در اجرای اول قسمت اول از K قسمت به منظور ارزیابی، K-1 قسمت باقیمانده برای یادگیری استفاده می‌شود. در اجرای دوم قسمت دوم از K قسمت به منظور ارزیابی، K-1 قسمت باقیمانده برای یادگیری استفاده می‌شود الگوریتم

¹ Machine Learning

مورد اشاره K مرتبه به همین روال اجرا میگردد و در نهایت میانگین نتایج پس از اجرای روش اعتبارسنجی ۱۰ قسمتی به‌عنوان شاخص ارزیابی مدل ارائه شده در سال t در نظر گرفته خواهد شد.

۵-۶- مرحله ارزیابی دقت پیش‌بینی الگو

متدهای مختلفی جهت ارزیابی عملکرد مدل یادگیری و نتایج دسته‌بندی وجود دارد. ماتریس درهم‌ریختگی یک روش اندازه‌گیری عملکرد برای مسئله طبقه‌بندی یادگیری ماشین است. ماتریس درهم‌ریختگی نتایج حاصل از طبقه‌بندی را بر اساس اطلاعات واقعی موجود، نمایش می‌دهد و تعداد پیش‌بینی‌های صحیح و نادرست هر کلاس را مشخص می‌کند. جدول زیر ماتریس درهم‌ریختگی برای یک مدل دوتایی را نشان می‌دهد:

جدول ۱. ماتریس درهم‌ریختگی الگوریتم‌های طبقه‌بندی

کلاس/تشخیص	تشخیص مثبت	تشخیص منفی
مثبت	TP	FN
منفی	FP	TN

TP: کلاس مثبت که به‌درستی مثبت تشخیص داده شده است.

FN: کلاس مثبتی که به‌اشتباه منفی شده است.

FP: کلاس منفی که به‌اشتباه مثبت تشخیص داده شده است.

TN: کلاس منفی که به‌درستی منفی تشخیص داده شده است.

بر اساس این ماتریس انواع شاخص‌های دقت، صحت و امتیاز F-1 و منحنی AUC جهت ارزیابی قدرت پیش‌بینی کنندگی الگوریتم‌های یادگیری ماشین استفاده شده است:

دقت^۱: عبارت است از میزان پیش‌بینی‌های صحیح در مجموع دو کلاس. این پارامتر در واقع نشانگر میزان الگوهای است که درست تشخیص داده شده‌اند:

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Fn + Fp + Tn}$$

صحت^۲ درستی پیش‌بینی‌ها را در یک کلاس نسبت به کل موارد درست را نشان می‌دهد (نسبت پیش‌بینی صحیح به کل موارد واقعی در آن گروه) فرمول محاسبه به‌صورت زیر است:

$$\text{precision} = \frac{Tp}{Tp + Fp}$$

tp تعداد پیش‌بینی‌های درست و fp تعداد پیش‌بینی‌های غلط در آن کلاس است.

پوشش^۳: در شاخص پوشش نسبت مقدار موارد صحیح طبقه‌بندی‌شده توسط الگوریتم از یک کلاس به کل اعضای پیش‌بینی شده در آن گروه را محاسبه می‌شود. (نسبت پیش‌بینی به کل موارد پیش‌بینی شده در آن گروه) فرمول محاسبه به‌صورت زیر است:

$$\text{Recall} = \frac{Tp}{Tp + Fn}$$

^۱ Accuracy

^۲ Precision

^۳ Recall

امتیاز F1-^۱: این معیار دقت (Precision) و پوشش (Recall) را با هم در نظر می‌گیرد. معیار F1 در بهترین حالت، یک و در بدترین حالت صفر است. این معیار توصیف‌کننده میانگین وزن‌دار مابین دو کمیت Precision و Recall است. این معیار میزان دقت و پوشش یک مدل را به صورت هم‌زمان بررسی می‌کند و میزان کیفیت کلاس‌بندی را تعیین می‌کند. برای محاسبه امتیاز F1 از فرمول زیر استفاده می‌شود.

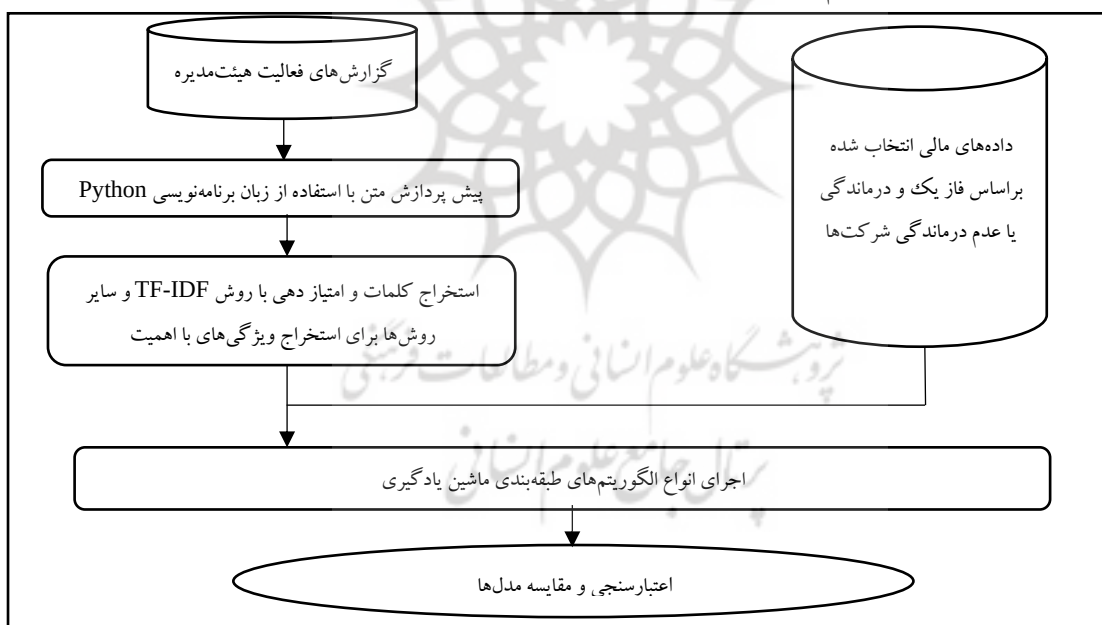
$$F1\ Score = \frac{precision * Recall}{precision + Recall}$$

معیار ناحیه تحت منحنی ROC، یکی دیگر از معیارهای مهمی است که برای تعیین کارایی مدل‌های طبقه‌بندی مورد استفاده قرار می‌گیرد. ناحیه تحت منحنی، نشان‌دهنده سطح زیر نمودار ROC است که هر چه مقدار این عدد مدل بزرگ‌تر باشد، کارایی نهایی مدل مطلوب‌تر ارزیابی می‌شود. در واقع، منحنی‌های ROC منحنی‌های دوبعدی هستند که در آن درصد تشخیص صحیح دسته مثبت روی محور Y و به طور مشابه، درصد تشخیص غلط دسته منفی روی محور X رسم می‌شود. به بیان دیگر، منحنی ROC مصالحه نسبی میان سودها و هزینه‌ها را نشان می‌دهد.

$$TPR = \frac{Tp}{Tp + Fn} * 100$$

$$AUC = \frac{1 + TPR - FPR}{2} * 100$$

با توجه مراحل اشاره شده الگوریتم اجرای تحقیق به صورت زیر است:



شکل ۲. الگوریتم اجرای پژوهش

۶- یافته‌های پژوهش

همان‌طور که اشاره شد هدف از انجام مراحل پیش‌پردازش آماده‌سازی متون گزارش‌های فعالیت هیئت‌مدیره برای عملیات متن‌کاوی است. از این رو در این مرحله با استفاده از زبان برنامه‌نویسی پایتون عملیات نرمال‌سازی، حذف کلمات توقف، ریشه‌یابی و حذف کلمات و کاراکترهای اضافی صورت گرفت که چند نمونه از کدهای استفاده شده در شکل ۳ و ۴ ارائه شده است.

^۱ F1 Score

جدول ۲. خلاصه نتایج حاصل از استفاده از الگوریتم‌های مختلف یادگیری ماشین

نام مدل	Accuracy	Precision	Recall	F1 Score	AUC
مدل جنگل تصادفی					
رگرسیون لجستیک					
درخت تصمیم					
نزدیک‌ترین k همسایگی					
SVM با کرنل خطی					
SVM با کرنل شعاعی					
SVM با کرنل چندجمله‌ای					
SVM با کرنل سیگموئید					

منبع: یافته‌های پژوهشگر

در برازش هر یک از مدل‌ها از روش اعتبار متقاطع (cross validation) به روش kfold استفاده شده است. از این رو با در نظر گرفتن ۱۰ فولد برای هر نمونه‌گیری، آموزش و اعتبارسنجی مدل‌ها در هر فولد انجام شده است. همان‌طور که اشاره شد در این مرحله جهت ارزیابی قدرت پیش‌بینی‌کنندگی هریک از مدل‌های یادگیری ماشین از شاخص‌های معرفی شده در جدول به ترتیب دقت مدل (Accuracy)، صحت مدل یا همان درصد تشخیص درست موارد درماندگی مالی (Precision)، پوشش یا همان درصد موارد درست در کل داده‌ها (Recall)، معیار امتیاز F که به عنوان ترکیب متعادل بین معیارهای دقت و صحت است (F1 score) و همچنین AUC به عنوان سطح زیر منحنی ROC استفاده شده است و نتایج حاصل برای هر یک از شاخص‌ها از میانگین نتایج هر فولد (جدول ۳) حاصل شده است.

باتوجه به جدول شماره ۲ می‌توان گفت دو روش درخت تصمیم و SVM با کرنل شعاعی بهترین عملکرد نسبت به سایر روش‌ها داشته‌اند به طوری که بر اساس شاخص صحت که تمرکز آن درستی تشخیص‌های مربوط به وقوع درماندگی مالی دارای ۸۰ و ۸۲ درصد، بر اساس معیار پوشش (Recall) که هدف تشخیص داده‌هایی است که واقعاً وقوع درماندگی مالی روی آن‌ها داده است به ترتیب ۷۸ و ۸۶ درصد پوشش دهی وقوع درماندگی مالی، بر اساس معیار fscore نیز که نوعی ترکیبی از دو معیار صحت و پوشش است به ترتیب ۷۹ و ۸۴ درصد توانایی تشخیص‌های درست را تأیید می‌کنند. سایر مدل‌ها نسبت به این دو مدل تفاوت قابل توجهی در تشخیص درماندگی مالی داشته‌اند. در ادامه نتایج مربوط به هر یک از ۱۰ فولد برای هر یک از روش‌های به کار گرفته شده نمایش داده شده است.

جدول ۳. جزئیات نتایج حاصل ۱۰ فولد برای هر یک از روش‌های به‌کاررفته

نام مدل	شماره فولد	Accuracy	Precision	Recall	F1 Score	AUC
مدل جنگل تصادفی	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
	۸					
	۹					
	۱۰					
مدل رگرسیون لجستیک	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
	۸					
	۹					
	۱۰					
مدل درخت تصمیم	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					

نام مدل	شماره فولد	Accuracy	Precision	Recall	F1 Score	AUC
مدل نزدیکترین k همسایگی	۸					
	۹					
	۱۰					
	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
SVM با کرنل خطی	۸					
	۹					
	۱۰					
	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
SVM با کرنل شعاعی	۸					
	۹					
	۱۰					
	۱					
	۲					

نام مدل	شماره فولد	Accuracy	Precision	Recall	F1 Score	AUC
SVM با کرنل چندجمله‌ای	۶					
	۷					
	۸					
	۹					
	۱۰					
	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
	۸					
	۹					
	۱۰					
	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
	۸					
	۹					
	۱۰					
SVM با کرنل سیگموئید	۱					
	۲					
	۳					
	۴					
	۵					
	۶					
	۷					
	۸					
	۹					
	۱۰					

منبع: یافته‌های پژوهشگر

۶-۱- تجزیه و تحلیل حساسیت

مطالعات اخیر انجام شده در حوزه یادگیری ماشین، روش‌های قابل توضیح جدیدی را معرفی کرده‌اند که می‌توانند پیش‌بینی‌های طبقه‌بندی‌کننده‌های یادگیری ماشین را به شیوه‌ای قابل تفسیر و دقیق توضیح دهند. یکی از این روش‌ها،

مدل لایم^۱ (LIME) است. مدل LIME می‌تواند استنتاج‌های تولید شده توسط مدل‌های یادگیری ماشین را با انجام یک تقریب محلی نقطه استنتاج، توضیح دهد. در این مدل ویژگی‌های با وزن مثبت بالا، در تقریب رگرسیون خطی از تصمیم پیش‌بینی پشتیبانی می‌کنند، درحالی‌که ویژگی‌هایی با وزن منفی مخالف تصمیم هستند. در واقع خروجی این مدل، مجموعه‌ای از توضیحات است که نشان می‌دهد سهم هر ویژگی در پیش‌بینی نمونه چقدر است. در جدول ۴، نتایج حاصل از اجرای این مدل بر روی بهترین الگوریتم پیش‌بینی‌کننده درماندگی مالی ارائه شده است. قابل ذکر است به دلیل بالای حجم کلمات صرفاً کلمات با ضریب بااهمیت بالا در جدول ۴ ارائه شده است:

جدول ۴. نتایج حاصل از به کارگیری مدل لایم جهت توضیح پذیری نتایج حاصل از الگوریتم یادگیری ماشین

کلمات اثر گذار در جهت پیش بینی سالم بودن شرکت		کلمات اثر گذار در جهت پیش بینی درماندگی مالی شرکت	
اهمیت	کلمه	اهمیت	کلمه
قسمت	کار	محصول	مدیر
گزارش	بنیاد	انجام	پایش
سفارش	گمراه	باعث	مواد
نسبت	متوسط	تمام	سند
کافی	مرجع	توجه	قیمت
همراه	میانگین	روبرو	دنیا
سرمایه	رای	تأمین	وقت
مشابه	مجوز		

منبع: یافته‌های پژوهشگر

۷- نتیجه گیری

تاکنون مدل‌های زیادی درباره پیش‌بینی درماندگی مالی ارائه شده است، اما چیزی که کمتر به آن پرداخته شده اثر محتوای اطلاعاتی گزارش فعالیت‌های هیئت‌مدیره در پیش‌بینی درماندگی مالی است. در واقع متن گزارش فعالیت‌های

¹ Local Interpretable Model-Agnostic Explanations

هیئت‌مدیره به مجمع، حاوی سرنخ‌های ارزشمندی است که همانند داده‌های مالی، می‌توانند وجود درماندگی مالی در شرکت‌ها را تشخیص دهند. از این رو در این پژوهش تلاش شده است تا بتوان با استفاده از ظرفیت اطلاعات گزارش‌های فعالیت هیئت‌مدیره، درماندگی مالی شرکت‌ها را پیش‌بینی نمود. پس از انجام فرایند پیش‌پردازش و عملیات متن‌کاوی و استخراج متغیرهای وابسته، با استفاده از روش‌های مختلف یادگیری ماشین (شامل رگرسیون لجستیک، جنگل تصادفی، درخت تصمیم، نزدیک‌ترین همسایگی و روش‌های ماشین بردار با کرنل‌های خطی، شعاعی، سیگموئید و چندجمله‌ای) قدرت پیش‌بینی‌کنندگی درماندگی مالی اندازه‌گیری شد. نتایج پیش‌بینی حاکی از قدرت پیش‌بینی‌کنندگی بالای دو مدل درخت تصمیم (۷۹ درصد) و SVM با کرنل شعاعی (۸۵ درصد) نسبت به سایر مدل‌ها است. در واقع نتایج این پژوهش نشان داد که به‌جای توجه صرف بر اعداد و ارقام، نسبت‌های مشتق شده از این اعداد و ارقام، می‌توان از تکنیک متن‌کاوی نیز جهت تجزیه و تحلیل و پیش‌بینی استفاده کرد و با تلفیق آن با نتایج حاصل از اطلاعات کمی می‌توان درماندگی مالی شرکت‌ها پیش‌بینی نمود که نتایج پژوهش در راستای پژوهش **مای و همکاران (۲۰۱۹)**، **تانگ و همکاران (۲۰۲۰)** و **سانگ و همکاران (۲۰۲۴)** است که بیان می‌کنند استفاده از اطلاعات متنی سبب افزایش قدرت پیش‌بینی‌کنندگی درماندگی مالی می‌شود. در واقع علی‌رغم اینکه دقت پیش‌بینی‌کنندگی نتایج حاصل از داده‌های ساختاریافته نسبت به ساختاریافته (همچون پژوهش **وانگ و همکاران، ۲۰۱۷**) با دقت پیش‌بینی‌کنندگی حدود ۹۰ درصد، کمتر است؛ اما شناخت و استفاده از ظرفیت این نوع اطلاعات بسیار بااهمیت است؛ به طوریکه بسیاری از دانشگاهیان و دست‌اندرکاران معتقدند افشاهای کمی به تنهایی برای تصمیم‌گیری اقتصادی مؤثر نیستند (**گو و همکاران، ۲۰۱۶**).

این پژوهش از چند جهت، برای حوزه حسابداری و حوزه داده‌کاوی، حائز دانش‌افزایی است: اول، پژوهش حاضر سعی کرده است از داده‌های ساختاریافته مستخرج شده از صورت‌های مالی، داده‌های بازار و داده‌های اقتصادی فاصله گرفته و برای بخش عمده‌ای از تحلیل‌ها بر گزارش‌های هیئت‌مدیره (داده‌های ساختار نیافته) به مجمع تکیه کند که این موضوع در راستای نتایج **تتلاک و همکاران (۲۰۰۸)**، **هنری (۲۰۰۸)**، **لوران و مک دونالد (۲۰۱۱)** و **هنری و لیون (۲۰۱۶)** مبنی بر دارا بودن محتوای اطلاعاتی گزارش‌های نوشتاری برای پیش‌بینی وضعیت عملکرد آتی شرکت است. دوم اینکه نتایج این پژوهش نشان داد که به‌جای تکیه بر «اعداد و ارقام»، بر تجزیه و تحلیل «متن گزارش‌ها» تکیه کرد و درماندگی مالی شرکت‌ها را بادقت بالایی پیش‌بینی کرد.

با توجه به نتایج حاصل از پژوهش که نشان داد با متن‌کاوی اطلاعات کیفی و ساختار نیافته هیئت‌مدیره می‌توان درماندگی مالی شرکت‌ها را بادقت بالایی پیش‌بینی نمود به سرمایه‌گذاران، تحلیل‌گران، مدیران، حساب‌رسان، پژوهشگران و سایر استفاده‌کنندگان از اطلاعات منتشر شده توسط شرکت پیشنهاد می‌شود که در تحلیل‌ها و پیش‌بینی‌های خود از تکنیک‌های متن‌کاوی برای تحلیل بیشتر گزارش‌های کیفی و ساختار نیافته بهره‌گیرند. همچنین به پژوهشگران آتی پیشنهاد می‌شود که باتکیه بر سایر گزارش‌های کیفی مشابه، اطلاعاتی مانند گزارش تفسیری مدیریت، متن صورت‌جلسات هیئت‌مدیره، متن سخنرانی‌های هیئت‌مدیره و مدیران شرکت، متن کنفرانس‌های تلفنی، متن گزارش حساب‌رس، متن نامه مدیران، متن اطلاعیه‌های سود، و ... توانایی این متون را در سنجش و ارزیابی درماندگی مالی بررسی نمایند. همچنین پیشنهاد می‌شود پیش‌بینی‌پذیری سایر موضوعات حوزه حسابداری و مالی همچون پیش‌بینی قیمت سهام، ریسک و ... را با تکیه بر متن‌کاوی گزارش‌های کیفی و ناساختاریافته سنجش و ارزیابی کنند.

۱-۷- محدودیت‌ها

اولین محدودیت مربوط به دسترسی به فایل WORD گزارش‌های هیئت‌مدیره بود که پژوهشگر مجبور شد از ربات‌های موجود در این حوزه استفاده نماید که در برخی موارد به دلیل مواردی چون اسکن بودن گزارش، فرمت گزارش پس از تبدیل به فایل متنی (text) به صورت به هم ریخته یا عربی نمایش داده می‌شد که امکان خواندن آن اصلاً وجود نداشت و این موضوع سبب کاهش حجم نمونه شد و همچنین محدودیت مربوط به عدم وجود الگوریتم‌ها و نرم‌افزارهای نشانه‌گذاری (Qtagger) برای متون فارسی بود که سبب شد تجزیه و تحلیل گزارش‌های فعالیت هیئت‌مدیره و پایش واژه‌ها به صورت دستی انجام پذیرد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت و این موضوع مورد تأیید همه آنهاست.

مشارکت نویسندگان

این مقاله برگرفته از رساله آقای مجتبی عالی‌فامیان در رشته حسابداری به راهنمایی دکتر محمد مرفوع در دانشگاه علامه طباطبائی است.

نویسنده اول: تهیه و آماده‌سازی نمونه‌ها، انجام آزمایش و گردآوری داده‌ها، انجام محاسبات، تجزیه و تحلیل آماری داده‌ها، تحلیل و تفسیر اطلاعات و نتایج، تهیه پیشنویس مقاله.

نویسنده دوم: استاد راهنمای رساله، طراحی پژوهش، نظارت بر مراحل انجام پژوهش، بررسی و کنترل نتایج، اصلاح، بازبینی و نهایی‌سازی مقاله.

نویسنده سوم: استاد مشاور رساله، مشارکت در طراحی پژوهش، نظارت بر پژوهش، مطالعه و بازبینی مقاله.

نویسنده چهارم: استاد مشاور رساله، مشارکت در طراحی پژوهش، نظارت بر پژوهش، مطالعه و بازبینی مقاله.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

حامی مالی

نویسنده/ نویسندگان هیچگونه حمایت مالی برای تحقیق، تألیف و انتشار این مقاله دریافت نکرده‌اند.

تقدیر و تشکر

نویسندگان مقاله بر خود واجب می‌دانند از همکاری و مساعدت تمامی عزیزانی که در این مطالعه ما را یاری نموده‌اند، سپاسگزاری نمایند.

منابع

- ابراهیمی سرو علیا، محمدحسن؛ باباجانی، جعفر؛ آخوند، محمدرضا و فاخر، اسلام (۱۳۹۷). ارائه الگویی برای پیش‌بینی پویای درماندگی مالی با استفاده از تحلیل بقاء. *فصلنامه اقتصاد مقداری*، ۱۵(۴)، ۱۶۷-۱۹۸. https://jqe.scu.ac.ir/article_14147.html
- افلاطونی، عباس (۱۳۹۲). تحلیل آماری با Eviews در پژوهش‌های حسابداری و مدیریت مالی. تهران، انتشارات ترمه. <https://www.gisoom.com/11397041>
- بت‌شکن، محمد هاشم؛ سلیمی، محمدجواد و فلاح‌نگر متحدجو، سعید (۱۳۹۷). ارائه یک روش ترکیبی به‌منظور پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۲۰(۲)، ۱۷۳-۱۹۲. https://jfr.ut.ac.ir/article_67697.html
- برزگری خانقاه، جمال؛ انصاری سامانی، حبیب و رزاززاده، لیدا (۱۳۹۸). پیش‌بینی وضعیت مالی شرکت‌ها با استفاده از تحلیل محتوای گزارش‌های هیئت‌مدیره. *مطالعات تجربی حسابداری مالی*، ۱۶(۶۴)، ۱۳۵-۱۶۰. https://qjma.atu.ac.ir/article_10959.html
- بولو، قاسم و احمدوند، میثم (۱۳۹۸). الگویی برای پیش‌بینی نکول شرکتی در بورس اوراق بهادار تهران. *مجله دانش حسابداری*، ۱۰(۱)، ۳۸-۱. https://jak.uk.ac.ir/article_2262.html
- پله، مولود؛ ایزدی‌نیا، ناصر و امیری، هادی (۱۳۹۸). بررسی تأثیر لحن گزارش‌های فعالیت هیئت‌مدیره بر عملکرد آتی شرکت‌ها مبتنی بر دو دیدگاه علامت‌دهی و رفتار فرصت‌طلبانه مدیران. *دو فصلنامه حسابداری ارزشی و رفتاری*، ۴(۸)، ۳۱-۱. <http://aapc.khu.ac.ir/article-1-663-fa.html>
- حاجیها، زهره و صابری روچی، محمدرضا (۱۴۰۰). متن‌کاوی و پیش‌بینی درماندگی مالی. *مطالعات حسابداری و حسابرسی*، ۳۹(۳۸)، ۵۳. https://www.iaaaas.com/article_134533.html
- حسینی، سید رسول و حاجیان‌نژاد، امین (۱۴۰۱). تأثیر توزیع آماری نسبت‌های مالی بر مقادیر مدل آلتمن با استفاده از شبیه‌سازی مونت کارلو. *مجله دانش حسابداری*، ۱۳(۳)، ۱۶۱-۱۹۳. https://jak.uk.ac.ir/article_3168.html
- رهروی دستجردی، علیرضا، فروغی، داریوش، کیانی، غلامحسین. (۱۳۹۷). ارزیابی خطر تقلب مدیران با استفاده از روش داده‌کاوی. *مجله دانش حسابداری*، ۹(۳۳)، ۹۱-۱۱۴. https://jak.uk.ac.ir/article_1932.html
- رحیمی، حمید؛ مینویی، مهرزاد و فتحی، محمدرضا (۱۴۰۳). پیش‌بینی پویا درماندگی مالی: مطالعه موردی. *فصلنامه اقتصادی مالی*، ۱۸(۶۶)، ۳۸۵-۴۰۸. <https://sanad.iau.ir/fa/Article/1063299?FullText=FullText>
- زارع بهمنیری، محمدجواد؛ حیدری سورشجانی، زهرا و جودکی چگکی، زهرا (۱۳۹۸). قدرت مدیران و لحن اعلان سود با تأکید بر نقش نظارت هیئت‌مدیره. *فصلنامه حسابداری مالی*، ۴۴(۱۱۹)، ۱۴۳. <https://qfaj.mobarakheh.iau.ir/article-1-2005-fa.html>
- عباسیان، عزت‌اله؛ شهرکی، کاوه؛ فلاح‌پور، سعید و نمکی، علی (۱۴۰۲). رویکردی نوین در پیش‌بینی درماندگی مالی با به‌کارگیری اطلاعات مبتنی بر شبکه مالی و روش ترکیبی درخت تصمیم تقویت‌گرا دیان. *مدیریت دارایی و تأمین مالی*، ۳(۴۲)، ۱۱۳-۱۴۰. <https://ensani.ir/fa/article/583773>
- مهرانی، ساسان و نونهال‌نهر، علی اکبر (۱۳۹۲). آزمون تأثیر رویکردهای قضاوتی زبان در کاهش فاصله انتظارات حسابرسی. *بررسی‌های حسابداری و حسابرسی*، ۲۰(۳)، ۱۲۳-۱۴۴. https://acctgrev.ut.ac.ir/article_35679.html
- مهرانی، ساسان؛ مهرانی، کاوه؛ منصفی، یاشار و کرمی، غلامرضا (۱۳۸۴). بررسی کاربردی الگوهای پیش‌بینی ورشکستگی زیمسکی و شیرانا در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *بررسی‌های حسابداری و حسابرسی*، ۴۱(۱۰۵-۱۳۱). https://acctgrev.ut.ac.ir/article_18462.html
- نمازی، محمد و ابراهیمی، شهاب (۱۴۰۰). پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران با استفاده از ماشین بردار پشتیبان. *راهبرد مدیریت مالی*، ۹(۳۲)، ۱۱۵-۱۳۲. https://jfam.alzahra.ac.ir/article_5588.html
- هاشمی گل سفیدی، افشین؛ لشگری، زهرا و حاجیها، زهره (۱۴۰۱). کاربرد یادگیری ماشین در ارائه الگویی برای پیش‌بینی ورشکستگی. *تحقیقات حسابداری و حسابرسی*، ۱۴(۵۶)، ۱۷۱-۱۹۰. DOI: 10.22034/iaar.2022.168271

References

- Abbasian, E., Shahraki, K., Fallahpour, S., & Namaki, A. (2023). A novel approach in predicting financial distress by financial network-based information and the integrated method of gradient boosting decision tree. *Journal of Asset Management and Financing*, <https://ensani.ir/fa/article/583773> [In Persian].
- Aflatoni, A. (2013). *Statistical Analysis with Eviews in accounting and financial management research*. Tehran, Termeh Publications. <https://www.gisoom.com/book/11397041> [In Persian].
- Altman E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.2307/2978933>.
- Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131-171. <https://doi.org/10.1111/jifm.12053>.
- Balakrishnan, R., Qiu, X.Y., & Srinivasan, P. (2010). On the predictive ability of narrative disclosures in annual reports. *European Journal of Operational Research*, 202(3), 789-801. <https://doi.org/10.1016/j.ejor.2009.06.023>.
- Barzegari Khanagha, J., Ansari Samani, H., & Razzazzadeh, L. (2019). Predicting the financial status of companies using content analysis the reports of the board of directors. *Empirical Studies in Financial Accounting*, 16(64), 135-160. https://qjma.atu.ac.ir/article_10959.html?lang=en [In Persian].
- Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 5, 71-111. <https://doi.org/10.2307/2490171>.
- Beaver, W.H. (1966). Financial ratios as predictors of failure, empirical research in accounting: Selected studies. *Journal of Accounting Research*, 5, 71-111. <https://doi.org/10.2307/2490171>.
- Blue, Gh., & Ahmadvand, M. (2019). A model for predicting corporate default in Tehran Stock Exchange. *Journal of Accounting Knowledge*, 10(1), 1-38. https://jak.uk.ac.ir/article_2262.html?lang=en [In Persian].
- Botshekan, M., Salimi, M., & Falahatgar Mottahedjoo, S. (2018). Developing a hybrid approach for financial distress prediction of listed companies in Tehran stock exchange. *Financial Research Journal*, 20(2), 173-192. https://jfr.ut.ac.ir/article_67697.html?lang=en [In Persian].
- Cole, C.J., & Jones, C.L. (2005). Management discussion and analysis: A review and implications for future research. *Journal of Accounting Literature*, 24, 135. [ProQuest].
- Clatworthy, M., & Jones, M.J. (2003). Financial reporting of good news and bad news: evidence from accounting narratives. *Accounting and Business Research*, 33(3), 171-185. <https://doi.org/10.1080/00014788.2003.9729645>.
- Deakin, E.B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 10(1), 167-179. <https://doi.org/10.2307/2490225>.
- Denis, D.J., & Denis, D.K. (1995). Causes of financial distress following leveraged recapitalizations. *Journal of Financial Economics*, 37(2), 129-157. [https://doi.org/10.1016/0304-405X\(94\)00792-Y](https://doi.org/10.1016/0304-405X(94)00792-Y).
- Ebrahimi sarv olia, M.H., Babajani, J., Akhond, M., & Fakher, E. (2019). A pattern for dynamic prediction of financial distress by using survival analysis. *Quarterly Journal of Quantitative Economics*, 15(4), 167-198. https://jqe.scu.ac.ir/article_14147.html [In Persian].
- Fitzpatrick, P. (1932). A comparison of ratios of successful industrial enterprises with those of failed companies. The certified public accountant (October, November, December), 598-605, 656-662, and 727-731, respectively. <https://www.sci epub.com/reference/297635>.
- Gilbert, L.R., Menon, K., & Schwartz, K.B. (1990). Predicting bankruptcy for firms in financial distress. *Journal of Business Finance and Accounting*, 17 (1), 161-171. <https://doi.org/10.1111/j.1468-5957.1990.tb00555.x>.
- Guo, L., Shi, F., & Tu, J. (2016). Textual analysis and machine leaning: Crack unstructured data in finance and accounting. *The Journal of Finance and Data Science*, 2(3), 153-170. <https://doi.org/10.1016/j.jfds.2017.02.001>.

- Gupta, V., & Lehal, G.S. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60-76. <https://doi.org/10.4304/jetwi.1.1.60-76>.
- Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods. *Knowledge-Based Systems*, 128, 139-152. <https://doi.org/10.1016/j.knosys.2017.05.001>.
- Hajek, P., & Michalak, K. (2013). Feature selection in corporate credit rating prediction. *Knowledge-Based Systems*, 51, 72-84. <https://doi.org/10.1016/j.knosys.2013.07.008>.
- Hajiha, Z., & Saberi Rochi, M.R. (2021). Text-mining technique and prediction of financial distress. *Accounting and Auditing Studies*, 10(38), 39-52. https://www.iaaaas.com/article_134533.html?lang=en [In Persian].
- Hashemi Gelsefidi, A., Lashgari, Z., & Hajiha, Z. (2022). Using machine learning to provide a model for predicting bankruptcy. *Accounting and Auditing Research*, 14(56), 171-190. DOI: [10.22034/iaar.2022.168271](https://doi.org/10.22034/iaar.2022.168271) [In Persian].
- Hackeling, G. (2017). Mastering machine learning with scikit-learn. Packt Publishing Ltd. [Amazon].
- Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, 45(4), 363-407. <https://doi.org/10.1177/0021943608319>.
- Henry, E., & Leone, A.J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, 91(1), 153-178. <https://doi.org/10.2308/accr-51161>.
- Hosseini, S.R., & Hajiannejad, A. (2022). The effect of statistical distribution of financial ratios on Altman model values using Monte Carlo simulation. *Journal of Accounting Knowledge*, 13(3), 161-193. https://jak.uk.ac.ir/article_3168.html?lang=en [In Persian].
- Huang, B., Yao, X., Luo, Y., & Li, J. (2023). Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research*, 330(1), 457-484. <https://doi.org/10.1007/s10479-022-04633-3>.
- Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K., & Felix, W.F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585-594. <https://doi.org/10.1016/j.dss.2010.08.009>.
- Jo, N.O., & Shin, K.S. (2016). Bankruptcy prediction modeling using qualitative information based on big data analytics. *Journal of Intelligence and Information Systems*, 22(2), 33-56. <https://doi.org/10.13088/jiis>.
- Kahl, M. (2002). Economic distress, financial distress and dynamic liquidation. *The Journal of Finance*, 57(1), 135-168. <https://doi.org/10.1111/1540-6261.00418>.
- Kang, T., Park, D.H., & Han, I. (2018). Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics. *Telematics and Informatics*, 35(2), 370-381. <https://doi.org/10.1016/j.tele.2017.12.014>.
- Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, 29, 143-165. <https://ssrn.com/abstract=1756926>.
- Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A deep learning-based approach to constructing a domain sentiment lexicon: a case study in financial distress prediction. *Information Processing & Management*, 58(5), 102673. <https://doi.org/10.1016/j.ipm.2021.102673>.
- Li, J., & Wang, C. (2023). A deep learning approach of financial distress recognition combining text. *Electronic Research Archive*, 31(8), 4683-4707. DOI: [10.3934/era.2023240](https://doi.org/10.3934/era.2023240).
- Loughran, T., & McDonald, B. (2010). Measuring readability in financial text. *SSRN eLibrary*, 64(4), 1643-1671. <http://dx.doi.org/10.2139/ssrn.1920411>.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.

- Loughran, T., McDonald, B., & Yun, H. (2009). A wolf in sheep's clothing: The use of ethics-related terms in 10-K reports. *Journal of Business Ethics*, 89, 39-49. <https://doi.org/10.1007/s10551-008-9910-1>.
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743-758. <https://doi.org/10.1016/j.ejor.2018.10.024>.
- Mehrani, S., & Nonahal Nahr, A.A. (2013). Evaluating the effect of the judgmental approaches of language in reducing audit expectation gap. *Accounting and Auditing Review*, 20(3), 123-144. DOI: [10.22059/acctgrev.2013.35679](https://doi.org/10.22059/acctgrev.2013.35679) [In Persian].
- Mehrani, S., Mehrani, K., Monsefi, Y., & Karami, Gh. (2005). An application study of Zimsky and Shirana bankruptcy prediction models in companies listed on the Tehran Stock Exchange. *Accounting and Auditing Reviews*, 41, 105-131. https://acctgrev.ut.ac.ir/article_18462.html?lang=en [In Persian].
- Merton, R.C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 141-183. <https://doi.org/10.2307/3003143>.
- Merwin, C.L. (1942). Financing small corporations in five manufacturing industries, 1926-36. [NBER Books].
- Namazi, M., & Ebrahimi, S. (2021). Financial distress prediction of the listed companies on Tehran Stock Exchange (TSE) and Iran Fara Burse (IFB) using support vector machine. *Financial Management Strategy*, 9(1), 115-132. https://jfm.alzahra.ac.ir/article_5588.html?lang=en [In Persian].
- Opler, T.C., & Titman, S. (1994). Financial distress and corporate performance. *The Journal of Finance*, 49(3), 1015-1040. <https://doi.org/10.1111/j.1540-6261.1994.tb00086.x>.
- Peleh, M., Izadinia, N., & Amiri, H. (2020). The effect of tone in board's activity reports on future performance based on signaling and opportunistic approach. *Iranian Journal of Behavioral and Valued Accounting*, 4(8), 1-31. <http://aapc.khu.ac.ir/article-1-663-fa.html> [In Persian].
- Pindado, J., Rodrigues, L., & De la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, 61(9), 995-1003. <https://doi.org/10.1016/j.jbusres.2007.10.006>.
- Platt, H.D., & Platt, M.B. (2006). Understanding differences between financial distress and bankruptcy. *Review of Applied Economics*, 2(2), 141-157. <https://ideas.repec.org/a/ags/reapec/50146.html>.
- Rahimi, H., Minoui, M., & Fathi, M. (2024). Dynamic prediction of financial distress: A case study. *Financial Economic Quarterly*, 18(66), 385-408. <https://sanad.iau.ir/fa/Article/1063299?FullText=FullText> [In Persian].
- Rahrovi Dastjerdi, A., Foroghi, D., & Kiani, G. (2018). Assessing managers fraud through analysis of board of directors report by data mining. *Journal of Accounting Knowledge*, 9(1), 91-114. https://jak.uk.ac.ir/article_1932.html [In Persian].
- Song, Y., Jiang, M., Li, S., & Zhao, S. (2024). Class-imbalanced financial distress prediction with machine learning: Incorporating financial, management, textual, and social responsibility features into index system. *Journal of Forecasting*, 43(3), 593-614. <https://doi.org/10.1002/for.3050>.
- Sreedharan, M., Khedr, A.M., & El Bannany, M. (2020). A comparative analysis of machine learning classifiers and ensemble techniques in financial distress prediction. In *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, pp. 653-657. <https://doi.org/10.1109/SSD49366.2020.9364178>.
- Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting*, 39(5), 769-787. <https://doi.org/10.1002/for.2661>.
- Tetlock, P.C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437-1467. <https://doi.org/10.1111/j.1540-6261.2008.01362.x>.
- Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 109-131. <https://doi.org/10.2307/2490395>.

- Wang, Z., Chen, M.H., Chin, C.L., & Zheng, Q. (2017). Managerial ability, political connections, and fraudulent financial reporting in China. *Journal of Accounting and Public Policy*, 36(2), 141-162. <https://www.researchgate.net/publication/315499649>.
- Yang, J.H., & Liu, S. (2017). Accounting narratives and impression management on social media. *Accounting and Business Research*, 47(6), 673-694. <https://doi.org/10.1080/00014788.2017.1322936>.
- Zare Bahnamiri, M.J., Heidary Surshjani, Z., & Joudaki Chegeni, Z. (2020). CEO power and earnings announcement tone with emphasis on the role of board oversight. *Iranian Journal of Behavioral and Valued Accounting*, 11(44), 119-143. <https://qfaj.mobarakeh.iau.ir/article-1-2005-fa.html> [In Persian].
- Zmijewski, M.E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82. <https://doi.org/10.2307/2490859>.

