Karafan
Quarterly Scientific Journal of National University of Skills
Special Issue 2025, Vol. 22, p.491-521
Journal Homepage: https://karafan.tvu.ac.ir/?lang=en
: 10.48301/kssa.2024.394327.2528

Predicting fraud in financial statements using supervised methods

Zahra Nemati[1] , Ali Mohammadi[2] *, Ali Bayat [3] , Abbas Mirzaei [4],
1. Department. of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran.
2. Corresponding Author, Assistant Professor of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran
3. Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran.
4. Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran.

ARTICLE INFO

Article Type:
Original Research

Received: 11.27.2023
Revised: 02.27.2024
Accepted: 06.19.2024

Keyword:
Financial Statements, Financial Ratios, Genetic Algorithm, Supervised Methods

***Corresponding Author:**
Ali Mohammadi
Email:
ali_mohammedi93@yahoo.com

ABSTRACT

The goal of this study was to offer appropriate financial measures that can be used in supervised approaches to forecast the likelihood of fraud in financial statements. The statistical sample, which was chosen using the systematic elimination sampling approach, was comprised of 180 companies that were admitted to the Tehran Stock Exchange between 2014 and 2021. 96 financial ratios were collected using the supervised methods of support vector machine, K-nearest neighbor, Bayesian network, hybrid method (Bagging), neural network, decision tree, logistic regression, and random forest to predict fraud in financial statements. The results demonstrated that the techniques perform rather poorly in the majority of the evaluation criteria. As a result, to improve performance, the genetic algorithm was employed to minimize the ratios, and after 89 iterations, 7 suitable financial ratios were recovered with a proportionality function of 0.2917 and an accuracy of 70.83%. The performance of the proposed approaches for predicting fraud in financial statements was then examined again using the extracted financial ratios. After lowering the financial ratios, the results revealed that all of the offered approaches performed better.

©2024 the authors. Published by National University of Skills, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC License) (https://creativecommons.org/licenses/by-nc/4.0/)

E-ISSN: 2538-4430
ISSN: 2382-9796

EXTENDED ABSTRACT

Introduction

In recent decades, executives and law enforcement organizations around the world have focused heavily on the subject of misleading financial reporting. With the advancement of technology and high-speed communication networks, nowadays the methods of committing fraud are very complex; committing it is easier, and detecting it has become more difficult, particularly since fraud is usually hidden and fraudsters act smartly and quickly. For this reason, its diagnosis is a complex but important task.

Thus far, many models have been put forward to help find fraud. These models can be roughly put into three groups: statistical models, innovative models, and models based on artificial intelligence. Studies have shown that the use of statistical models, due to their reliance on limiting assumptions such as normal distribution and a high classification error rate, and meta-heuristic models, due to their being stuck in local optimal points and premature convergence, have decreased and led to the acceptance of artificial intelligence models because these techniques are frequently non-parametric and require little initialization. Because of this, artificial intelligence techniques are becoming more prevalent in fraud detection methods. Hence, given the importance of fraud prediction in financial statements, metaheuristic methods were employed in this study to provide users with appropriate financial ratios for the prediction of fraud in financial statements.

Methodology

This research is descriptive-correlational because it seeks to describe the conditions or phenomena under investigation. In terms of the implementation process, it is quantitative and in terms of time position, it is retrospective (post-event). To collect the required theoretical and background information, it was conducted by studying the research carried out at the national and international levels including books and articles, using the library method and scanning. Excel software was utilized for their calculations to obtain the necessary data of the variables from the financial statements and reports of independent auditors and legal inspectors published by the Tehran Stock Exchange Organization and Rahvard Navin software package. Meta-innovative and data mining techniques were utilized to evaluate the information and test the research hypotheses, and for this, MATLAB and Datalab tools were used.

Because there is no organization or institution in Iran to identify companies that commit fraud and information about these companies is not available to the public, the condition other clauses of the audit reports of the companies that had a modified opinion (rejected opinion, no opinion, and conditional opinion) were consequently thoroughly examined in light of the examples of fraud previously mentioned. In a sample of 1440 companies (180 companies for 8 years), 532 were suspected of financial statement fraud, while 908 were found to be non-

fraudulent. Companies suspected of fraud are represented by the number one, while non-fraudulent companies are represented by the number zero.

Results and Discussion

The supervised methods used in this study included support vector machines, K-nearest neighbors, simple Bayesian models, neural networks, decision trees, logistic regression, random forests, and group classification using 96 financial ratios extracted from financial statements. The likelihood of fraud was predicted using sample companies. The genetic algorithm was used as a metaheuristic method of nature-inspired evolutionary algorithms to select a subset of features (i.e., financial ratios). The results were saved as MATLAB tables.

Learning techniques were first taught to examine and measure the ratios extracted from the proposed algorithms. To determine the training percentage of the models, 70% of the data (1008 data, of which 376 were suspected fraud companies and 632 were non-fraudulent companies) were provided as training data in the MATLAB software. Lastly, 30% of the remaining data (including 432 companies data: 156 data of companies suspected of fraud, and 276 data of non-fraudulent companies) were used as test data in the software to evaluate the algorithms and the expected amount of fraud in financial statements that should be investigated and evaluated.

Tables 1 and 2 show the average values of the evaluation criteria for the proposed methods.

Table 1. Brief results of performance evaluating the proposed method without feature reduction.

Description	Accuracy	Precision	Recall	F-measure	ROC	TP+TN
Bagging	72.45 %	61.21%	64.74%	62.93%	73.50%	313
Support vector machine	69.44%	56.90%	63.46%	60%	71%	300
k nearest neighbor	66.20%	52.72%	62.18%	57.06%	67%	286
Bayesian network	65.51%	51.98%	58.97%	55.26%	68%	283
neural network	54.39%	38.79%	45.51%	41.88%	60.70%	235
decision tree	54.16%	37.34%	39.74%	38.50%	62.60%	234
logistic regression	47.91%	30.93%	35.89%	33.22%	58%	207
Random forest	41.89%	25.38%	31.41%	31.60%	59.30%	181

Table 2. The results of performance evaluation, confusion matrix, and ROC of proposed methods with financial ratios extracted from the GA algorithm.

Description	Accuracy	Precision	Recall	F-measure	ROC	TP+TN
Bagging	80.13 %	76.41 %	65.10 %	70.30 %	78.3 %	350
Support vector machine	76.47 %	69.34 %	62.48 %	65.72 %	74.7 %	335
k nearest neighbor	74.72 %	65.51 %	63.42 %	65.40 %	67.6 %	328

Bayesian network	73.25 %	62.98 %	62.90 %	62.73 %	72.5 %	320
neural network	72.20 %	61.54 %	60.01 %	60.74 %	65.1 %	270
decision tree	72.19 %	60.04 %	61.31 %	60.65 %	65.3 %	269
logistic regression	72.51 %	60.47 %	61.89 %	61.14 %	62.6 %	252
Random forest	72.22 %	59.96 %	61.92 %	60.90 %	61.4 %	230

Conclusions

Given that many of their users, including shareholders, creditors, and investors rely on financial statements to make decisions, the presence of fraud in these cases not only results in high expenses, risk, and errors in their decisions, but also has serious non-financial effects particularly the erosion of the credibility of the accounting profession. On the other hand, as information technology has developed, fraudsters' methods have become faster and more complex, and the independent factors in the accuracy of financial statement fraud predictions are unclear. Compared to discrete and non-linear data, continuous and linear data are more predictive using statistical methods. The likelihood of accurately predicting fraud is decreased with these methods. The development of fraud detection techniques in financial statements is crucial for this reason. Genetic Algorithm methods were used in this study to extract appropriate financial ratios for fraud prediction in financial statements. A comparison of the supervised classification methods was carried out to determine which method performed best, including support vector machines, K-nearest neighbors, Bayesian networks, neural networks, decision trees, logistic regression, random forests, and group classification. Additionally, financial ratios are better able than other techniques to foresee the possibility of fraud in financial statements. According to the obtained results, the Bagging classification method performed better than other supervised classification methods for all performance evaluation criteria.

کارافن
فصلنامه علمی دانشگاه ملی مهارت
ویژه‌نامه ۱۴۰۴، دوره ۲۲، ۵۲۱-۴۹۱
آدرس نشریه: https://karafan.tvu.ac.ir/
doi: 10.48301/kssa.2024.394327.2528

پیش‌بینی تقلب در صورت‌های مالی با استفاده از روش‌های نظارت شده

زهرا نعمتی[۱]، علی محمدی*[۲]، علی بیات[۳]، عباس میرزایی[۴]

۱. گروه حسابداری، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران
۲. نویسنده مسئول، حسابداری، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران
۳. گروه حسابداری، گروه حسابداری، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران
۴. گروه کامپیوتر، واحد اردبیل، دانشگاه آزاد اسلامی، اردبیل، ایران

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

دریافت مقاله: ۱۴۰۲/۰۹/۰۶
بازنگری مقاله: ۱۴۰۲/۱۲/۰۸
پذیرش مقاله: ۱۴۰۳/۰۳/۳۰

کلید واژگان:
صورت‌های مالی، نسبت‌های مالی، الگوریتم ژنتیک، روش‌های نظارت شده

***نویسنده مسئول:** علی محمدی
پست الکترونیکی
ali_mohammedi93@yahoo.com

چکیده

هدف این پژوهش ارائه نسبت‌های مالی مناسب جهت پیش‌بینی احتمال تقلب در صورت‌های مالی با استفاده از روش‌های نظارت شده می‌باشد. اطلاعات مورد بررسی مربوط به نمونه آماری که به روش نمونه‌گیری حذف سیستماتیک انتخاب شده است، شامل ۱۸۰ شرکت پذیرفته شده در بورس اوراق بهادار تهران، طی مقطع زمانی ۱۴۰۰-۱۳۹۳ می‌باشد. با بررسی مبانی نظری و تحقیقات انجام شده۹۶ نسبت مالی گردآوری گردیده است که توسط روش‌های نظارت شده ماشین بردارپشتیبان، کای نزدیک‌ترین همسایه، شبکه بیزین، روش ترکیبی (بگینگ)، شبکه عصبی، درخت تصمیم، رگرسیون لجستیک و جنگل تصادفی جهت پیش‌بینی تقلب در صورت‌های مالی مورد استفاده قرار گرفتند. نتایج نشان دهنده عملکرد نسبتاً پایین روش‌ها در اکثر معیارهای ارزیابی می‌باشد. به همین دلیل جهت بهبود عملکرد از الگوریتم ژنتیک جهت کاهش نسبت‌ها استفاده گردیده که با مقدار تابع تناسب عدد۰/۲۹۱۷ و دقت ۷۰/۸۳ ٪ و پس از۸۹ مرحله تکرار، ۷ نسبت مالی مناسب استخراج شد. در مرحله بعد مجدداً با نسبت‌های مالی استخراج شده به ارزیابی عملکرد روش‌های پیشنهادی جهت پیش‌بینی تقلب در صورت‌های مالی پرداخته شد. نتایج نشان دهنده بهبود عملکرد در تمامی روش‌های پیشنهادی پس از کاهش نسبت‌های مالی می‌باشد.

شاپای الکترونیکی: ۲۵۳۸-۴۴۳۰
شاپای چاپی: ۹۷۹۶-۲۳۸۲

©2024 the authors. Published by National University of Skills, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC License) (https://creativecommons.org/licenses/by-nc/4.0/)

۱.مقدمه

بدلیل این‌که استفاده‌کنندگان از صورت‌های مالی مانند اعتباردهندگان، سرمایه‌گذاران و نهادهای دولتی و... براساس اطلاعات موجود در آن‌ها جهت اعطای اعتبار، سرمایه‌گذاری و برنامه‌ریزی برای آینده سهام خود در شرکت تصمیم‌گیری می‌کنند، لذا صورت‌های مالی باید براساس ارائه اطلاعات منصفانه و واقعی در مورد وضعیت مالی، جریان‌های نقدی و عملکرد عملیاتی شرکت تهیه شود (عماد الدینی و سعیدی، ۱۳۹۸). افزایش تعداد تقلب در صورت‌های مالی و موارد تجدید ارائه که اغلب با ورشکستگی شرکت‌های بزرگ در آمیخته، نگرانی‌های را درباره کیفیت گزارشگری مالی ایجاد کرده است. این گزارش‌های مالی بازارهای سرمایه کشورهای مختلف را دستخوش نوسانات شدید نموده است و تاثیرمنفی براین بازارها داشته و باعث از بین رفتن اعتماد عمومی به سرمایه‌گذاری گردیده است (زارع بهنمیری، ۱۳۹۴). در پژوهشی که توسط بازرسان رسمی تقلب در خصوص این‌که چند درصد درآمد سالانه سازمان‌ها صرف خسارت‌ها و زیان‌های ناشی از تقلب می‌شود انجام دادند، به این نتیجه رسیدندکه بطور متوسط سالانه ۵ درصد از درآمدها صرف هزینه‌ها و زیان‌های مربوط به تقلب می‌شود (صدیقی کمال، ۱۳۹۲). با وجود این‌که تقلب در صورت‌های مالی فقط ۹ درصد موارد تقلب در گزارشات می‌باشد اما متوسط زیان آن به ازاء هرتقلب۵۹۳,۰۰۰ دلار می‌باشد که از پرهزینه‌ترین موارد تقلب است (انجمن بازرسان رسمی تقلب، ۲۰۲۲)[۱]. تقلب عمده در صورت‌های مالی می‌تواند تاثیر منفی قابل توجهی بر ارزش بازار واحد تجاری، اعتبار و توانایی آن‌ها در دستیابی به اهداف استراتژیشان داشته باشد که در نهایت منجر به ورشکستگی آن‌ها و از دست رفتن دهها و یا هزاران فرصت شغلی می‌شود. در سطح جامعه نیز باعث کاهش‌کارایی بازار سرمایه، از بین رفتن اعتماد عموم به حرفه حسابداری و حسابرسی و کاهش رشد و شکوفائی اقتصادی می‌گردد (کیموناکی و همکاران، ۲۰۱۶)[۲]. علی‌رغم اهمیت تقلب در صورت‌های مالی و با توجه به کسب رتبه ۲۵ ایران در رتبه‌بندی فساد مالی توسط سازمان شفافیت بین‌الملل[۳] در سال ۲۰۲۱، که متاسفانه از لحاظ فساد مالی و گسترش آن در میان۱۸۰ کشور جهان رتبه۵۰ را کسب نموده است، هیچ نهاد قانونی بطور مستقیم برای تحقیق و کشف تقلب یا پایگاهی جهت اطلاع‌رسانی لیست شرکت‌های متقلب جهت کشف و مصادیق تقلب در صورت‌های مالی در کشور ما وجود ندارد. از سویی حجم اطلاعات صورت‌های مالی بسیار بالا و کاملا تخصصی می‌باشد که اکثر استفاده‌کنندگان از این صورت‌های مالی تخصص و وقت کافی جهت بررسی آن‌ها را جهت اتخاذ تصمیمات خود ندارند. همچنین با پیشرفت تکنولوژی و شبکه‌های ارتباطی پرسرعت و امروزه روش‌های انجام تقلب بسیار پیچیده، ارتکاب آن آسان‌تر وکشف آن مشکل-ترگردیده است و از آنجایی که تقلب معمولاً کتمان می‌شود و متقلبین، هوشمندانه و سریع عمل می‌نمایند (سادگالی و همکاران، ۲۰۱۹)[۴].

به همین دلیل تشخیص آن یک کار پیچیده اما مهم است. تاکنون مدل‌های مختلفی جهت شناسایی احتمال وقوع تقلب ارائه شده است. مطالعات نشان می‌دهد که استفاده از مدل‌های آماری به دلیل اتکا بر مفروضات

---

۱ Association Chief Police Officers

۲ chimonaki et al

۳ Transparency International Organization

4 Sadgali et al

محدود کننده، مانند توزیع نرمال و نرخ خطای بالای طبقه بندی (یائو و همکاران،۲۰۱۹)[۱] کاهش یافته و موجب استقبال از مدل‌های هوش مصنوعی شده است، زیرا که این تکنیک‌ها اغلب ناپارامتریک بوده و در بکارگیری آنها، نیاز چندانی به مفروضات اولیه و اطلاعات مربوط به چگونگیِ توزیعِ ویژگی‌های مالی نیست. لذا این پژوهش با توجه به موارد ذکر شده، بدنبال یافتن بهترین نسبت‌های مالی و استفاده از روش‌های هوش مصنوعی جهت پیش‌بینی احتمال تقلب در صورت‌های مالی می‌باشد.

۲.مبانی نظری و پیشینه پژوهش

۱,۲. تعریف تقلب

تقلب دارای اثرات زیان باری درابعاد مختلف اقتصادی، فرهنگی و اجتماعی می‌باشد. در واقع فساد و تقلب، سوء استفاده از قدرت و یا منفعت شخصی و یا گروهی است که در نتیجه عدم کنترل برقدرت و ایجاد فرصت برای تحریف از قدرت موجود و عدم قطعیت ضمانت‌های اجرایی است. فساد و تقلب به عنوان جرم غیرعادی از طریق خدشه وارد کردن به اخلاق و فرهنگ جامعه و سیاست‌های دولت در تأمین منافع مردم باعث کاهش رقابت سطح ملی و مختل شدن رشد اقتصادی می‌شود و به کیفیت رابطه‌های تجاری لطمه می‌زند و در نهایت منجر به خنثی شدن تلاش‌ها در جهت کاهش فقر و تبعیض اجتماعی است (امار و پاربا، ۲۰۲۰)[۲].

ابادین و دایکمور (۲۰۲۰)[۳] هم اعتقاد دارندکه تقلب دارای مفهوم گسترده‌ای است که جهت دست یافتن به منافع غیرقانونی با فریب عمدی اشاره دارد و دارای انواع متفاوت فساد مالی، گزارشگری متقلبانه و سواستفاده از دارایی‌ها می‌باشد.

جاویر و همکاران (۲۰۱۸)[۴] بیان می‌کنند تقلب یک فعالیت بزهکارانه و فریبکارانه با قصد منفعت مالی و سایر منافع می‌باشد، که در برگیرنده حقه بازی، حیله‌گری و رفتار غیرمنصفانه بوسیله یک شخص فریبکار و متقلب است.

در بیانیه شماره ۹۹ استانداردهای حسابرسی آمریکا، آمده است :

" تقلب فعلی عمدی است که به ارائه نادرست با اهمیت در صورت‌های مالی مورد حسابرسی بی‌انجامد. تقلب معمولاً در صورت بروز سه وضعیت زیر رخ می‌دهد:

۱- مدیریت یا سایر کارکنان در پی‌انگیزه‌ای یا تحت فشار خاصی مرتب تقلب می‌شوند.

۲- وضعیت موجود بطور مثال نبود کنترل، نامناسب بودن کنترل‌های موجود، یا توانایی مدیریت در زیر پا گذاشتن کنترل‌ها، فرصتی را برای ارتکاب تقلب فراهم می‌کند.

۳- افرادی که مرتکب اعمال متقلبانه می‌شوند، قادر به توجیه عمل خود هستند و بعضی از این افراد دارای نگرش، ویژگی یا مجموعه‌ای از اصول اخلاقی هستند که به آن‌ها اجازه می‌دهد تا آگاهانه و عمداً تخلفاتی را مرتکب شوند" (راموس،۲۰۰۳) [۵].

---

۱ Yao et al

۲ Umar& Purba

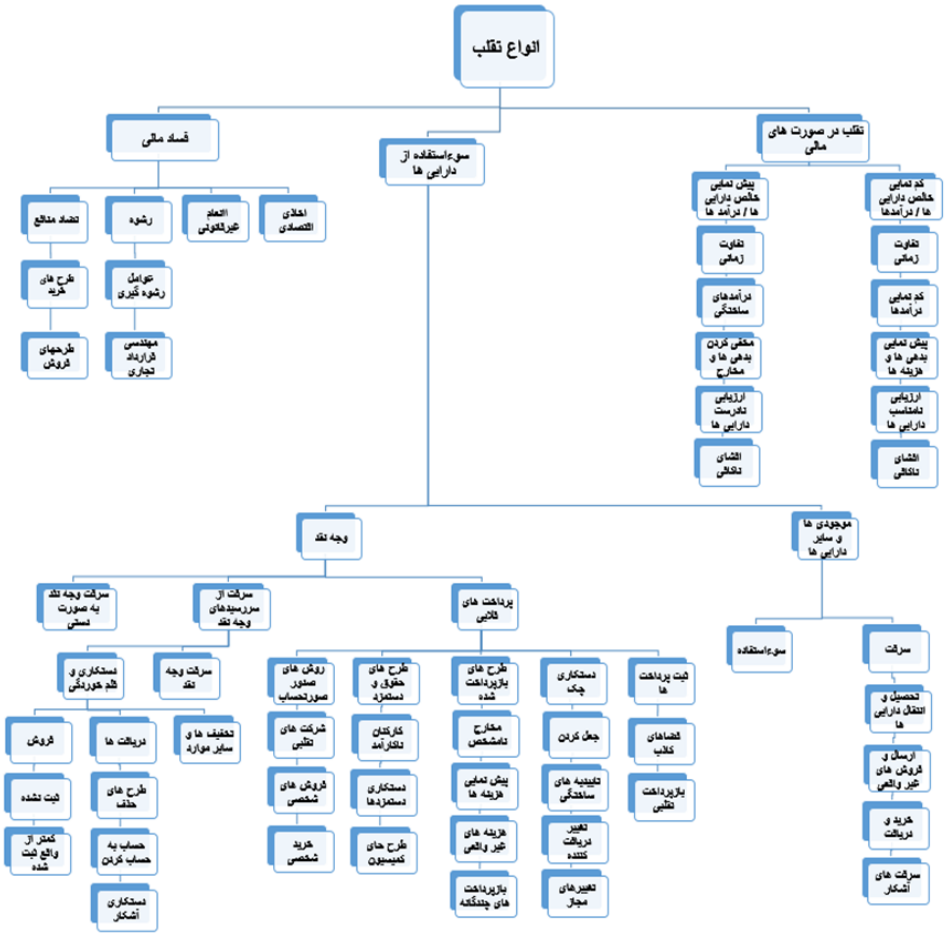
۳ Ibadin & Dikemor

۴ Javier et al

۵ Ramos

۲،۲. انواع تقلب

انجمن بازرسان رسمی تقلب (گزارش انجمن بازرسان تقلب، ۲۰۲۲)[۱]، انواع تقلب حرفه‌ای را در سه مقوله فساد مالی، سوءاستفاده از دارایی‌ها و تقلب در صورت‌های مالی طبقه‌بندی می‌کند.



شکل۱- طبقه‌بندی انواع تقلب واجزاء آن‌ها (گزارش انجمن بازرسان تقلب، ۲۰۲۲)

فساد مالی: تقلبی است که در آن کارکنان یک واحد از موقعیت و نفوذ خود به منظور دستیابی به منافع مستقیم یا غیرمستقیم سوء استفاده می‌کنند. مانند: دریافت پورسانت، درگیرشدن در تضاد منافع (طرح‌های خرید و فروش)، رشوه خواری و... (صدیقی، ۲۰۱۳).

۱ Association of Certified Fraud Examiners

براساس آمار انجمن بازرسان رسمی تقلب، فساد مالی در سال ۲۰۲۲، ۵۰ درصد موارد تقلب را شامل می‌شود و خسارتی معادل ۱۵۰,۰۰۰ دلار داشته است.

سوء استفاده از دارایی‌ها: شامل دزدی یا استفاده نادرست از دارایی‌های یک سازمان است. مانند ترفندهای نقد دزدی، دزدی کالا و سایر دارایی‌ها و استفاده غیرقانونی از دارای‌ها به عنوان دارایی شخصی (وکیلی فرد، ۲۰۱۳). استاندارد حسابرسی شماره ۲۴۰ ایران نیز انواع عوامل خطر تقلب مربوط به سوء استفاده از دارایی‌ها را در دو دسته شامل عوامل مرتبط با آسیب پذیری دارایی‌ها در برابر سوء استفاده و عوامل مرتبط با کنترل‌ها طبقه‌بندی نموده است.

طبق گزارش انجمن بازرسان رسمی تقلب در سال ۲۰۲۲، سوءاستفاده از دارایی‌ها شایع‌ترین نوع تقلب است. این تقلب در بیش از ۸۶ درصد از موارد تقلب، کشف و گزارش شده و بطور متوسط زیانی معادل ۱۰۰,۰۰۰ دلار در هر بنگاه اقتصادی داشته است.

تقلب در صورت‌های مالی: طبق بخش۲۴ استاندارد حسابرسی این تقلب شامل فریبکاری، از قبیل سندسازی و دستکاری یا تغییر سوابق حسابداری یا مدارک پشتوانه تهیه صورت‌های مالی، ارائه نادرست رویدادها در صورت‌های مالی یا حذف عمدی آن‌ها و به کارگیری نادرست استانداردهای حسابداری مرتبط با اندازه‌گیری، شناسایی، طبقه-بندی و ارائه یا افشاء به عمد است (کمیته تدوین استانداردهای حسابرسی، ۲۰۲۰).

با وجود این‌که تقلب در صورت‌های مالی فقط ۹ درصد موارد تقلب است اما متوسط زیان آن به ازاء هر تقلب۵۹۳,۰۰۰ دلار برآورد شده است که از پرهزینه‌ترین موارد تقلب است (انجمن بازرسان تقلب، ۲۰۲۲).

عمده ترین روش تقلب در صورت‌های به صورت بیش نمایی دارایی‌ها و درآمدها و یا کم نمایی بدهی‌ها و هزینه‌ها است. گاهی اوقات هم ممکن است برعکس موارد یادشده، یعنی با کم‌نمایی دارایی‌ها یا درآمد و بیش نمایی بدهی‌ها یا هزینه‌ها انجام شود.

انگیزه‌های انجام تقلب در صورت‌های مالی می‌تواند بدلیل پاداش مبتنی بر سود گزارش شده (واتز و زیمرمن،۱۹۸۶) [۱]، رسیدن به پیش‌بینی‌های داخلی و خارجی (فاماو جنسن، ۱۹۸۳) [۲]، حفظ یا افزایش قیمت بازار سهام (دیچو و همکاران، ۱۹۹۵)[۳]، جلوگیری از تخلف در قراردادهای بدهی (اسپاتیس، ۲۰۰۲)[۴]، حداقل کردن بدهی مالیاتی(دیچو و همکاران، ۱۹۹۶) و تأمین مالی ارزان قیمت (کوتسیانتیس،۲۰۰۶) باشد[۵].

۲-۳- نسبت‌های مالی

در جوامع بشری ضرورت اطلاعات مالی همواره باعث ایجاد چالش‌های فکری شده تا اطلاعاتی مفید و سودمند در اختیار تصمیم‌گیرندگان قرار گیرد. یکی از این موارد، تلاش در جهت یافتن ارتباط و تناسب بین اطلاعات مالی مختلف بوده است. در طول دهه۱۸۹۰ با افزایش وسیع حجم و گردش اطلاعات مالی در ابتدا گزارش‌های مالی

۱ Watts & Zimerman

۲ Fama & Jensen

۳ Dechow et al

۴ Spathis

۵ Kotsiantis

برحسب درخواست و اقلام مورد نیاز ارائه می‌شد، سپس در شکل ستونی و مقایسه‌ای تهیه گردید و در نهایت از قرن نوزدهم به صورت نسبت‌های مالی با هدف به دست آوردن اطلاعات معنادار از صورت‌های‌مالی استخراج می‌شوند. از این نسبت‌ها در حیطه‌های کاربردی جهت بررسی عملکرد شرکت و مقایسه آن در سنوات مختلف، پیش‌بینی موفقیت و عملکرد آتی شرکت و مقایسه شرکت‌ها با یکدیگر و در پژوهش‌ها در طراحی و ارائه مدل استفاده می‌شود (زارع بهنمیری و ملکیان، ۱۳۹۵).

نسبت‌های مالی به دسته‌های زیر تقسیم می‌شوند:

نسبت‌های سنجش توان مالی کوتاه‌مدت یا نسبت‌های نقدینگی: این نسبت‌ها همانطور که از نامشان پیداست، اطلاعاتی در مورد وضعیت نقدینگی شرکت فراهم می‌آورند و نشان دهنده توانایی شرکت در انجام تعهدات مالی کوتاه مدت خود می‌باشند. برخی ازپژوهش‌ها از قبیل سنت پیرو اندرسون (۱۹۸۴) نشان داده شده است که دارایی‌های جاری در شرکت‌های متقلب به مقدار زیادی ناشی از موجودی کالا و حساب‌های دریافتنی می‌باشند (پرسونز، ۲۰۱۱)[۱].

نسبت‌های سنجش توان مالی بلندمدت یا نسبت‌های اهرمی: این نسبت‌ها توانایی شرکت برای پرداخت تعهدات بلندمدت را اندازه‌گیری می‌کنند و نشان می‌دهد در تامین مالی و ساختار سرمایه چه میزان از بدهی استفاده شده است. پرسونز (۲۰۱۱) معتقد است که اهرم بالاتر بعضاً باپتانسیل بالاتر برای خلاف‌های توافق وام و توانایی کمتر برای دست یافتن به سرمایه اضافی از طریق استقراض در ارتباط باشد.

نسبت‌های مدیریت دارایی‌ها یا نسبت‌های گردش دارایی‌ها (کارایی): با استفاده از این نسبت‌ها می‌توان میزان کارایی شرکت در به کارگیری و استفاده از دارایی‌ها را اندازه‌گیری نمود (دالنیال، ۲۰۱۴)[۲]. کرکوس، اسپاتیس و مانولوپولوس(۲۰۰۷)[۳] در پژوهش خود نشان دادند که میانگین نسبت گردش دارایی‌ها در شرکت‌های غیرمتقلب و متقلب بطور معناداری متفاوت است.

نسبت‌های سودآوری: این نسبت‌ها ابزاری جهت سنجش میزان توانایی شرکت برای تولید درآمد و تحصیل سود هستند (جهانخانی و شوری،۱۳۹۲). پژوهش‌های کروتزفلد و والاس(۱۹۸۶) [۴] نشان داد که بنگاه‌ها با سوددهی کمتر با اعمال تقلب در صورت‌های مالی بطور معنی‌داری نسبت به بنگاه‌ها با سوددهی بیشتر در ارتباط هستند.

۴،۲.الگوریتم‌های[۵] پیشنهادی:

الگوریتم ژنتیک[۶]: الگوریتم ژنتیک روش جستجو و بهینه‌سازی است که با توجه به اصول انتخاب و ژنتیک طبیعی ایجاد می‌شوند (خو و وونش، ۲۰۰۵) [۷]. در این الگوریتم، از تکنیک‌های زیست‌شناسی فراگشتی مانند وراثت، جهش زیست‌شناسی و اصول انتخابی داروین برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو استفاده می‌شود. طبق نظریه تکامل تدریجی داروین، تنها افراد با ژن‌های مرغوب می‌توانند زنده بمانند و فرزندان جدیدی تولید

---

۱ persons

۲ Dalnial

۳ Kirkos, Spathis & Manolopoulos

۴ Kreutzfeldt & Wallace

۵ Han et al

۶ Genetic Algorithms

۷ Xu & Wunsch

نمایند و افراد با ژن‌های نامرغوب در فرایندی با عنوان بقای اصلح حذف خواهند شد و این فرایند تا رسیدن به جواب بهینه یا پایان یافتن زمان اجرا ادامه می‌یابد (تقوی و نوبری، ۱۳۸۵).

الگوریتم ماشین بردار پشتیبان[۱]: ماشین بردار پشتیبان یکی از روش‌های طبقه‌بندی با یادگیری با نظارت است که می‌توان برای حل مسائل طبقه‌بندی یا رگرسیون استفاده کرد. این الگوریتم توسط واپنیک (۱۹۹۵)[۲] براساس تئوری یادگیری آماری و به حداقل رساندن خطر ساختاری معرفی شده است. این الگوریتم ابر صفحه‌هایی را در فضا ترسیم می‌کند، که عمل تمایز نمونه‌های مختلف داده را به طور بهینه انجام می‌دهند، یعنی دو گروه را به گونه‌ای از هم متمایز می‌کند که با نزدیک‌ترین نقاط از هر گروه بیشترین فاصله را داشته باشند، بهترین ابر صفحه، به معنای صفحه‌ای است که دارای بیشترین فاصله از دو گروه است. این روش، داده‌ها را با یافتن بهترین ابر صفحه‌هایی که تمامی داده‌های یک گروه را از داده‌های گروه دیگر مجزا می‌کنند، طبقه‌بندی می‌کند(پرادان،۲۰۱۲)[۳].

الگوریتم کی نزدیکترین همسایه[۴] : این الگوریتم یکی از ساده‌ترین و در عین حال مهم‌ترین روش‌ها در طبقه‌بندی است. مبنای این الگوریتم پیدا کردن تعداد معینی از نزدیکترین عناصرموجود در جامعه آماری به عنصر جدید وارد شده در آن جامعه است که براساس آن بتوان نزدیکترین داده موجود به عنصرجدید را از لحاظ ویژگی‌های مختلف پیدا نموده و درهمان طبقه‌ای قرارداده که عناصر نزدیک به آن قرار دارند. به گفته ییینگ کوان و همکاران (۲۰۰۲)[۵] این الگوریتم نوعی طبقه‌بندی غیر پارامتریک جهت بدست آوردن تابع توزیع از روی داده‌های توزیع شده می‌باشد. درآن یک سند و یا داده‌ی آموزشی برای دسته‌بندی وجود دارد و این الگوریتم در میان سندهای آموزشی پیش دسته‌بندی شده، براساس یک معیار، شباهت پیدا نموده و دسته‌های این الگوریتم برای پیش‌بینی دسته سند آزمایشی به وسیله امتیازدهی سندهای هر دسته منتخب استفاده خواهد شد (گوآ و همکاران، ۲۰۰۳)[۶]. به طورکلی، روش کای‌نزدیک‌ترین همسایه یک مورد خاص از یادگیری مبتنی بر نمونه است که با داده‌های نمادین سروکار دارد. این روش همچنین نمونه‌ای از روش یادگیری تنبل است، تکنیکی است که منتظر می‌ماند تا پرس و جو فراتر از داده‌های آموزشی تعمیم یابد (کونچوا،۲۰۱۴)[۷].

الگوریتم شبکه بیزین[۸]: تاریخچه شبکه بیزین به کشف فرمول بیز در سال ۱۷۶۳ توسط کشیش انگلیسی توماس بیز برمی‌گردد. این الگوریتم مبتنی بر قضیه احتمال بیز است و احتمال تعلق به یک گروه خاص را تخمین می‌زند (لئونگ، ۲۰۰۷)[۹].

قضیه بیز به این صورت بیان می‌شود:

---

۱ Support vector machine Algorithms

۲ Vapnik

۳ Pradhan

۴ K-near neighbors Algorithms

۵ Yingquan et al

۶ Guo et al

۷ Kuncheva

۸ Bayesian Network Algorithms

۹ Leung

$$P(Y|X) = \frac{P(Y)*P(X|Y)}{P(X)} \quad (۱)$$

X مشاهده (یا مجموعه‌ای از صفات) وY نتیجه (یا برچسب گروه) است که می‌تواند از مجموعه داده‌ها به دست آید.

P(Y|X) نشان دهنده احتمالات پسین متغیرX بر طبقات ممکن است.

P(Y) بیانگر احتمالات پیشین هر طبقه بدون اطلاع درباره متغیرX است.

P(X|Y) نشان دهنده احتمالات مشروط متغیرX با احتمالY است.

P(X) اساساً احتمال شواهد است.

برای طبقه‌بندی یک نمونه جدید، می‌توان P(Y|X) را برای هر گروه خاص Y به دست آورده و بررسی کرد که کدام یک از این گروه‌ها، مقدار بیشتری دارند. برچسب گروه خاصY با بیشترین مقدار P(Y | X) برای یک صفت خاصX، به عنوان گروه تخمینی برای نمونه جدید در نظر گرفته می‌شود. از آنجا که P(X) برای هر مقدار گروه خاص نتیجه یکسانی دارد، لازم نیست که برای هر نمونه جدید محاسبه شده و بطور ثابت فرض می‌شود (شینده و همکاران، ۲۰۱۴)[۱].

الگوریتم روش ترکیبی (بگینگ)[۲]: این الگوریتم یک تکنیک یادگیری جمعی است که توسط بریمان[۳] در سال ۱۹۹۶ مطرح شد و هدف آن کاهش خطای با استفاده از مجموعه‌ای از مدل‌های یادگیری ماشین هم نوع می‌باشد. در روش بگینگ هر روش طبقه‌بند یک مدل را بر روی داده‌های آموزشی می‌سازد تا به وسیله‌ی آن بتواند تفاوت‌ها را در طبقه‌های مختلف درک کند. این الگوریتم، به جای این‌که خود یک مدل بسازد از مدل‌های ساخته شده توسطِ بقیه‌ی روش‌های طبقه‌بندی استفاده کرده و با یک رأی‌گیری، مشخص می‌کند که کدام طبقه را برای نمونه‌ی جاری باید برگزینند. هر کدام از طبقه‌بندها، به مجموعه‌ی داده دسترسی دارند. در این روش یک زیر مجموعه از مجموعه داده‌ی اصلی به هر کدام از طبقه‌بندها داده می‌شود. یعنی هر طبقه‌بند یک قسمت از مجموعه‌ی داده (ویژگی‌ها) را مشاهده کرده و باید مدل خود را براساس همان قسمت از داده‌ها که در اختیارش قرار گرفته است، بسازد (همه ویژگی‌ها به همه روش‌های طبقه‌بندی داده نمی‌شود) (شینده و همکاران، ۲۰۱۴).

شکل۲- الگوریتم روش ترکیبی (بگینگ) (وانگ و همکاران، ۲۰۱۴) [۴]

۱ Shinde et al

۲ Bagging Algorithms

۳ Breiman

۴wang et al

الگوریتم شبکه‌های عصبی[۱]: طرح این الگوریتم بر گرفته از ساختار مغز انسان است. همانطور که مغز انسان به شناسایی الگوهای مختلف داده‌ها و دسته بندی انواع اطلاعات می‌پردازد، می‌توان شبکه‌های عصبی را به شیوه‌ای مشابه با رفتار مغز انسان آموزش داد تا به تشخیص الگو بپردازد و دسته‌بندی داده‌ها را انجام دهند. یک الگوریتم شبکه عصبی متشکل از تعداد سلول عصبی و امثال آن به صورت واحدهای پردازش به هم پیوسته است (هان و همکاران، ۲۰۱۲) [۲]. این روش برای پیش‌بینی یا طبقه‌بندی متغیرهایی که اندازه‌گیری آن‌ها دشوار است و حل مشکلات طبقه‌بندی غیرخطی به درستی عمل می‌کند و نسبت به داده‌های پرت حساس نیست. ورودی‌ها را از طریق پردازش اطلاعات اطلاعات غیر خطی به خروجی‌هایی که با اهداف مطابقت داده می‌شوند در یک گروه متصل از نرون‌های مصنوعی که لایه‌های واحدهای پنهان را تشکیل می‌دهند، متصل می‌کند (دواو دو، ۲۰۱۱)[۳]. از ترکیب ورودی X و مجموعه وزن‌های نرون W، فعالیت هر واحد پنهان و خروجی Y مشخص می‌شود :

$$Y=f(X,W) \qquad (۲)$$

الگوریتم درخت تصمیم[۴]: این الگوریتم یک ساختار درختی به عنوان یک روش پیش‌بینی نتیجه یا مقدار ویژگی،بر اساس متغیرهای ورودی است. این روش، ابزاری رایج در یادگیری ماشین است که به یافتن استراتژی مناسب برای رسیدن به نتیجه رضایت بخش کمک می‌کند زیرا می‌توان با تطابق شروط از گره ریشه تا گره‌های برگ به مجموعه‌ای از قوانین مهم دست یافت درختان طبقه‌بندی، همان‌گونه که از نام آن پیداست، برای جدا کردن یک مجموعه داده به طبقات متعلق به متغیر پاسخ، استفاده می‌شود. معمولاً متغیر پاسخ دو طبقه بله یا خیر (۱ یا ۰) دارد. اگر متغیر پاسخ بیش از دو دسته باشد، ازانواع مختلفی از الگوریتم درخت تصمیم که ساخته شده‌اند استفاده می‌شوند (بورجیلی و همکاران، ۲۰۱۹)[۵].

الگوریتم رگرسیون لجستیک[۶]: برای انجام رگرسیون روی مجموعه‌ای از متغیرها از این الگوریتم که یک روش آماری طبقه‌بندی داده‌های باینری از مدل خطی معروف به مدل لجستیکی یا لوجیت استفاده میکند. این روش شامل تکنیک‌های زیادی برای مدل‌سازی و تحلیل متغیرهای خاص و منحصر بفرد، با تمرکز بر رابطه بین متغیر وابسته و یک یا چند متغیر مستقل، می‌باشد. در روش لوجیت، متغیر تبدیل شده با استفاده از یک تابع خطی درست مانند رگرسیون خطی، تقریب زده می‌شود. همان طور که در رگرسیون خطی نشان داده شده است، وزن‌ها باید طوری انتخاب شوند که داده‌های آموزشی درست را نتیجه دهند. رگرسیون خطی مناسب بودن را با استفاده از روش مربع خطا اندازه گیری می‌کند. از معادله ذیل در روش رگرسیون لجستیک استفاده می‌شود:

$$\sum_{n=1}^{\infty}(1-x^i)\log(1-p_r[1|a_1+a_2+\cdots+a_4])+x^i\log(pr[a^i_1+a^i_2+\cdots+a^i_k]) \qquad (۳)$$

۱ Neural Networks algorithms

۲ Han et al

۳Dua &Du

۴ Decision Tree Algorithm

۵ Bou Rjeily et al

۶ Logistic Regression Algorithm

که x^i برابر با صفر یا یک را ترجیح و بیشتر از آن استفاده می‌کنند (پارک و همکاران، ۲۰۱۱) [۱].

الگوریتم جنکل تصادفی[۲]: الگوریتم جنگل تصادفی در برگیرنده تعداد بسیاری درخت تصمیم می‌باشد که خروجی این الگوریتم بر مبنای آرای مشخص هر کدام از درخت‌ها مشخص و طبقه بندی نهایی انجام می‌شود. این الگوریتم مجموعه داده‌های تصادفی را با استفاده از نمونه گیری با جایگزینی برای آموزش هر یک از درختان جهت تصمیم گیری ایجاد می‌کند. در هر گره یک درخت، RF با استفاده از آنتروپی و معیار Gain Information، مهم‌ترین ویژگی را بین ویژگی‌ها انتخاب می‌کند. آنتروپی اندازه گیری ناخالصی یا عدم قطعیت صفات و Gain Information وسیله‌ای برای یافتن آموزنده ترین صفت است. بنابراین، هدف به حداقل رساندن آنتروپی و به حداکثر رساندن Gain Information با انتخاب ویژگی‌ها است. علاوه بر این، RF در هر گره از درخت، جایی که زیر مجموعه‌ای از m ویژگی برای تصمیم گیری در آن گره در نظر گرفته می‌شود، انتخاب فضای فرعی را به صورت تصادفی انجام می‌دهد (بلجی و همکاران، ۲۰۱۶) [۳].

۵،۲.پیشینه پژوهش:

(چن، ۲۰۲۳)[۴] یک روش انتخاب ویژگی یکپارچه را برای ساخت سیستم ویژگی تشخیص تقلب در صورت‌های مالی شرکت‌های فهرست شده پیشنهاد کرده و برای به دست آوردن یک مدل تشخیص تقلب در صورت‌های مالی، مدل‌های یادگیری ماشین Random Forest، GBDT، XGBoost و LightGBM را مقایسه کرده و با اضافه کردن الگوریتم SMOTE[۵] به طور موثر مشکل توزیع نامتعادل را حل کرده و به طور قابل توجهی توانایی تشخیص تقلب را بهبود داده است. در بین چهار الگوریتم یادگیری ماشین انتخاب شده، GBDT از نظر عملکرد AUC[۶] و حساسیت[۷] بهترین عملکرد را داشت. (شوگو و شن یانگ،۲۰۲۲)[۸] یک سیستم تشخیص تقلب مالی با استفاده از مدل یادگیری عمیق مبتنی برترکیبی از ویژگی‌های عددی و داده‌های متنی ارائه کردند، که از صورت‌های مالی جهت استخراج شاخص‌های مالی و همچنین گزارش مدیریتی جهت استخراج ویژگی‌های متنی در بخش MD&A[۹] با استفاده از بردار کلمه از گزارش‌های سالانه ۵۱۳۰ شرکت چینی استفاده نمودند، سپس از مدل‌های یادگیری عمیق استفاده کرده و عملکرد آن‌ها را به ترتیب با داده‌های عددی، داده‌های متنی و ترکیبی از آن‌ها مقایسه کردند. نتایج نشان داد که روش پیشنهادی با استفاده از طبقه‌بند LSTM با دقت ۹۴/۹۸ درصد و طبقه‌بند GRU با دقت ۹۴/۴۸ درصد نسبت به روش‌های طبقه‌بند سنتی کارایی بهتری دارند. (چنگ و همکاران ۲۰۲۱) [۱۰] در

---

۱ Park et al

۲ Random Forest Algorithm

۳ Belgiu et al

۴ Chen

۵ Synthetic Minority Oversampling Technique

۶ Area under the ROC Curve

۷ Recall

۸ Xiuguo & Shengyong

۹ Managment discussion & analysis

۱۰ Cheng et al

پژوهش خود جهت پیش‌بینی تقلب در صورت‌های مالی ابتدا برای انتخاب ویژگی از روش‌های پیش پردازش داده‌ها وسه روش انتخاب ویژگی مقادیرگمشده و مدیریت کلاس‌های نامتعادل، ویژگی ادغام شده استفاده نموده و۷۲ نسبت مالی را به ۱۸ نسبت مالی کاهش دادند و سپس چهار طبقه‌بند شبکه عصبی، درخت تصمیم، درختان اضافی و جنگل تصادفی را جهت طبقه‌بندی شرکت‌ها به متقلب و غیر متقلب بکار بردند. نتایج پژوهش نشان داد که ۱۸ نسبت انتخاب شده توسط روش ادغام شده با طبقه‌بند جنگل تصادفی دارای دقت ۹۸/۹۲ درصد است که نسبت به سایر روش‌ها دقت بیشتری دارد. (هدایت الله و همکاران،۲۰۲۰)[۱] به منظور تشخیص تقلب صورت‌های مالی در شرکت‌های اندونزی با استفاده از یادگیری ماشین براساس بهینه‌سازی فراابتکاری در ابتدا ۱۸ نسبت مالی که اطلاعات آن‌ها وجود داشت انتخاب کرده سپس با استفاده از روش تحلیل مؤلفه اصلی۱۰ نسبت مالی استخراج گردیده که در طبقه‌بندی مورد استفاده قرارگرفته است. سپس از چندین رویکرد یادگیری ماشین مبتنی بر بهینه‌سازی فراابتکاری برای توسعه مدل‌های پیش‌بینی تقلب در صورت‌های مالی استفاده نمودند. در این پژوهش از الگوریتم ژنتیک برای کاهش متغیرهای مالی و روش‌های ماشین‌بردار پشتیبان و شبکه عصبی پس انتشار[۲] بهینه‌سازی شده برای طبقه‌بندی استفاده شده است که نسبت‌های مالی استخراج شده توسط الگوریتم ژنتیک با طبقه‌بند ماشین بردار با دقت ۹۶/۱۵ درصد نسبت به سایر روش‌ها دارای دقت بیشتری بوده است. (امیدی و همکاران، ۲۰۱۹)[۳] در پژوهشی با استفاده از پنج روش با نظارت، یعنی شبکه عصبی چندلایه پیش‌خور، شبکه عصبی احتمالی، ماشین‌بردار پشتیبان، مدل خطی- لگاریتمی چند جمله‌ای و تحلیل افتراقی و ۱۸ تا داده مالی به پیش‌بینی تقلب در صورت‌های مالی پرداختند. نتایج پژوهش نشان داد که شبکه عصبی چند لایه پیش خور، با دقت بالای۹۰ درصد عملکرد بهتری در تشخیص تقلب درگزارش‌های مالی نسبت به سایر روش‌ها دارد. (جان ۲۰۱۸)[۴] در پژوهشی با بکارگیری روش‌های شبکه عصبی و ماشین بردارپشتیبان به منظور استخراج متغیرهای مناسب جهت پیش‌بینی تقلب از۲۲ متغیرمالی و غیرمالی در یک دوره یازده ساله استفاده نمودند. در روش شبکه عصبی۱۰ متغیر و ماشین بردار پشتیبان۳ متغیر بدست آمد. سپس با چهار تکنیک درخت تصمیم (QUEST، C5.0، CART، CHAID) دقت تشخیص تقلب در صورت‌های مالی را بررسی نمودند. نتایج پژوهش نشان داد که۱۰ متغیر استخراج شده توسط روش شبکه عصبی مصنوعی و طبقه‌بندی با درخت تصمیمCART با عدد۹۰/۲۱ درصد بالاترین دقت تشخیص تقلب را در صورت‌های مالی بین سایر مدل‌ها دارد. (عمرو همکاران، ۲۰۱۷)[۵] به بررسی پیش‌بینی تقلب در گزارش‌های مالی با استفاده از روش‌های رگرسیون لجستیک، ماشین بردار پشتیبان، الگوی تصمیم‌گیری چند معیاره و شبکه عصبی مصنوعی پرداختند در این پژوهش از۱۰ نسبت مالی بهترین نسبت‌ها جهت پیش‌بینی تقلب مورد استفاده قرار گرفته است. نتایج نشان داد که روش شبکه عصبی مصنوعی با سطح پیش‌بینی ۹۴/۸۷درصد توانایی بالاتری در پیش‌بینی تقلب نسبت به سایر روش‌ها دارد. (آفتابی و

۱ Hidayattullah et al

۲ Back Propagation Neural Network (BPNN)

۳ Omidi, Min, Moradinaftchali & Piri

۴ Jan

۵ Omar et al

همکاران،۱۴۰۲) یک رویکرد جدید مبتنی برGAN[۱] پیشنهاد کرده که با جمع‌آوری صورت‌های مالی سالانه ده بانک ایرانی و سپس استخراج سه نوع ویژگی پیشنهادی، مجموعه داده جدیدی ساخته است. نتایج تجربی روی این مجموعه داده نشان می‌دهد که روش پیشنهادی در تولید نمونه‌های مصنوعی مشکوک به تقلب با روش طبقه‌بندی XGBoost [۲] با دقت ۹۹ درصد و روش ماشین بردار پشتیبان با دقت ۱۰۰ درصد عملکرد خوبی نسبت به سایر روش‌های طبقه‌بندی دارد. علاوه بر این، در تشخیص دقیق نمونه‌های مشکوک به تقلب، عملکرد مقایسه‌ای با مدل‌های نظارت شده، عملکرد بهتری نسبت به مدل‌های بدون نظارت دارد. (کامرانی و عابدینی، ۱۴۰۱) با استفاده از بررسی مقالات و روش گلوله برفی و مصاحبه با متخصصان و خبرگان، ۲ نسبت غیرمالی و ۱۹ نسبت مالی را استخراج نمودند و با روش‌های شبکه عصبی و ماشین بردارپشتیبان به پیش‌بینی و کشف تقلب پرداختند. نتایج نشان داد که قدرت پیش‌بینی ماشین بردار پشتیبان ۸۶ درصد است و نسبت به شبکه عصبی عملکرد بهتری دارد. (کاظمی و پیری ۱۴۰۱) جهت پیش‌بینی تقلب در گزارش‌گری مالی از رویکرد یادگیری ماشین در فضای چند کلاسه استفاده کردند. در این پژوهش روش‌های رگرسیون لجستیک، درخت تصمیم، الگوریتم گرادیان تقویت شده و ماشین بردار پشستیبان در محیط پایتون بکاربرده شد که نتایج نشان دهنده تفاوت معنادار در عملکرد الگوهای یادگیری ماشین در فضای چند کلاسه وجود دارد و روش ماشین بردار پشتیبان نسبت به سایر روش‌ها عملکرد بهتری دارد. (یوخنه القیانی و همکاران، ۱۴۰۰) با استفاده از رویکر د ترکیبی داده‌کاوی کلاسیک، ANFIS و الگوریتم فراابتکاری به پیش‌بینی احتمال تقلب در گزارشگری مالی مالیاتی پرداختند. نتایج نشان دهنده افزایش قدرت پیش‌بینی مدل شناسایی گزارشگری مالی– مالیاتی با بکارگیری الگوریتم‌های بهینه‌سازی مختلف در رویکرد داده‌کاوی می‌باشد و بهبنه‌سازی با الگوریتم ازدحام ذرات بهینه‌ترین مدل را حاصل نموده است. (رضائی و همکاران،۱۳۹۹) جهت کشف تقلب از ۴۱ متغیر مالی و غیر مالی با ۵ تکنیک شبکه بیزین، درخت تصمیم، شبکه عصبی، ماشین بردار پشتیبان و روش ترکیبی استفاده نمودند. نتایج پژوهش نشان داد روش ترکیبی با نرخ پیش‌بینی۹۶/۲ درصد دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک‌ها می‌باشد. (تشدیدی و همکاران، ۱۳۹۸) به منظور ارائه رویکردی نوین در پیش‌بینی کشف تقلب در صورت‌های مالی با جستجو در شواهد تجربی در ابتدا ۲۳ نسبت مالی که اطلاعات آن‌ها در ایران وجود داشت انتخاب کرده سپس با استفاده از روش آنتروپی متقابل، ۱۶ نسبت بعنوان بهترین و موثرترین نسبت‌ها استخراج گردید. برای طبقه‌بندی شرکت‌ها به متقلب و غیرمتقلب هم از سه روش رگرسیون لجستیک، الگوریتم‌های ژنتیک و زنبورعسل استفاده نمودند که نتایج نشان داد که الگوریتم زنبورعسل نسبت به سایر روش‌ها با دقت پیش‌بینی ۸۲/۵ درصد بیشتری دقت جهت پیش‌بینی تقلب را دارد. (ابراهیمی و خواجوی،۱۳۹۶) در پژوهشی به منظور انتخاب متغیرهای که بیشترین تاثیر را برای کشف تقلب در صورت‌های مالی دارد روش انتخاب ویژگی مبتنی بر همبستگی بکار برده‌اند. بدین منظور از۴۰ متغیرمالی و غیرمالی استفاده نمودند که نتایج پژوهش حاکی از سودمندی نسبت‌های مالی نقدی، پوشش بهره، حساب‌های دریافتنی به کل دارایی‌ها، موجودی کالا به فروش خالص، لگاریتم طبیعی فروش، سودخالص به فروش و دارایی جاری به کل دارایی‌ها می‌باشد. همچنین برای پیش‌بینی تقلب از روش‌های داده‌کاوی شبکه عصبی مصنوعی، شبکه بیزین و جنگل تصادفی استفاده کردندکه الگوریتم جنگل تصادفی با دقت ۹۶/۷۷ درصد برای

۱ Generative Adversarial Networks

۲ Extreme Gradient Boosting

پیش بینی تقلب عملکرد بهتری نسبت به سایر تکنیک‌ها داشته است. (زارع بهنمیری و ملکیان، ۱۳۹۵) به ارائه مدلی جهت پیش‌بینی احتمال تقلب مالی پرداختند. آن‌ها با استفاده از روش نرم افزار متلب و طی دو مرحله با استفاده از آزمون‌های Stepwise Regression و Elastic net، ۷ نسبت سرمایه درگردش به دارایی، حساب دریافتنی به فروش، وجه نقد به بدهی جاری، موجودی به دارایی جاری، بدهی به حقوق صاحبان سهام، سود ناخالص به دارایی و قدر مطلق تغییرات در نسبت جاری انتخاب شده‌اند. در نتیجه پژوهش، با استفاده از آزمون لاجیت ۶۴/۰۴ درصد مدل برآورد شده قابلیت پیش‌بینی دارد.

۳.فرضیه پژوهش

کاهش نسبت‌های مالی باعث بهبود عملکرد روش‌های طبقه‌بندی جهت پیش‌بینی احتمال تقلب می‌باشد.

٤.روش پژوهش

این پژوهش از نظر نتایج و پیامد کاربردی، از نظر هدف بدلیل‌اینکه بدنبال توصیف شرایط یا پدیده‌های مورد بررسی است توصیفی– همبستگی، از نظر فرآیند اجرا کمی و از نظر جایگاه زمان گذشته‌نگر(پس رویدادی) می‌باشد. برای جمع‌آوری اطلاعات مورد نیاز مبانی نظری و پیشینه، با مطالعه پژوهش‌های انجام شده در داخل و خارج از کشور شامل کتب و مقالات به روش کتابخانه‌ای و فیش‌برداری، انجام گردیده است. جهت جمع‌آوری اطلاعات مورد نیاز متغیرها از صورت‌های مالی و گزارش حسابرسان مستقل و بازرسان قانونی منتشره سازمان بورس اوراق بهادار تهران و بسته نرم‌افزار ره‌آوردنوین و برای محاسبات مربوطه به آن‌ها از نرم‌افزار اکسل استفاده شده است. جهت تجزیه و تحلیل اطلاعات و آزمون فرضیه پژوهش از روش‌های فراابتکاری و داده‌کاوی استفاده می‌شود و برای این منظور نرم‌افزارهای متلب و دیتالب را بکار می‌بریم.

جامعه آماری پژوهش حاضر، شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران بین سال‌های ۱۳۹۳ تا ۱۴۰۰ می‌باشد. نمونه با روش حذف نظام‌مند و نمونه‌گیری شرطی انتخاب شده است، بدین صورت که پایان سال مالی شرکت‌ها منتهی به ۲۹ اسفند باشد، جزء شرکت‌های واسطه‌گری‌مالی، مانند سرمایه‌گذاری، هلدینگ، بانک‌ها و بیمه‌ها نباشند. اطلاعات مورد نیاز برای متغیرهای پژوهش موجود باشند. با توجه به شروط در نظر گرفته شده تعداد۱۸۰ شرکت انتخاب گردید.

۵. متغیرهای پژوهش

متغیر وابسته : جهت تعریف و مشخص نمودن تقلب در صورت‌های مالی بعنوان متغیر وابسته، با مراجعه و مطالعه استاندارد حسابرسی۲۴۰ با عنوان مسئولیت حسابرس و مبانی نظری تحقیقات داخلی و خارجی در ارتباط با تقلب، مصادیق تقلب شناسایی و استخراج شد.

در پیوست سه- استاندارد حسابرسی۲۴۰ نمونه‌هایی از شرایط نشان‌دهنده احتمال وجود تقلب به شرح ذیل ذکر شده است:

- ✓ معاملاتی که به طور کامل یا به موقع ثبت نشده‌اند یا مبلغ، تاریخ ثبت طبقه‌بندی یا رویه حسابداری آن‌ها نامناسب است.
- ✓ نبود مستندات: نداشتن مستندات کافی جهت مشخص کردن تاثیر تعدیلات احتمالی

- ✓ وجود اقلام عمده بدون توضیح در صورت مغایرت‌ها، تغییرات غیرعادی در ترازنامه، تغییرات در روندها یا نسبت‌ها یا روابط مهم صورت‌های مالی- بطور مثال: رشد سریع مطالبات در مقایسه با رشد درآمدها
- ✓ آخرین تعدیلات انجام شده که نتایج مالی را به شدت تحت تاثیر قرار می‌دهند.

همچنین با بررسی تحقیقات خارجی و داخلی انجام شده در مورد تقلب مهم‌ترین مصادیق تقلب به شرح ذیل بیان گردیده است :

- ✓ بیش‌نمایی و کم‌نمایی در درآمدها و دارایی‌ها
- ✓ بیش‌نمایی و کم‌نمایی در هزینه‌ها و بدهی‌ها
- ✓ صورت‌های مالی تجدید ارائه شده و وجود تعدیلات سنواتی با اهمیت
- ✓ وجود اختلاف مالیاتی با حوزه‌های مالیاتی و عدم کفایت ذخیره بابت مالیات عملکرد
- ✓ وجود دارایی و اقلام راکد مانند موجودی
- ✓ فرض تداوم فعالیت شرکت برای چندین دوره متوالی زیر سوال رفته باشد و اظهارنظر حسابرس مشروط باشد اما شرکت همچنان صورت‌های مالی خود را برمبنای تداوم فعالیت تهیه نماید بطور مثال شرکتی که تولید آن از دو سال قبل متوقف شده و فروشی نداشته است.
- ✓ بکارگیری اشتباه استانداردهای حسابداری مرتبط با شناسایی، اندازه‌گیری، طبقه‌بندی و ارائه و افشا

بدلیل این‌که در ایران سازمان و یا نهادی برای شناسایی شرکت‌هایی که مرتکب تقلب می‌شوند وجود ندارد و اطلاعات مربوط به این شرکت‌ها در دسترس عموم قرار نمی‌گیرد و از طرفی با توجه به این‌که در برخی از تحقیقات انجام شده رابطه بین تقلب و اظهار نظر حسابرس تایید شده است، به همین دلیل با توجه به مصادیق تقلب که در بالا عنوان گردیده است، بندهای شرط و سایر بندهای گزارش‌حسابرسی شرکت‌هایی که دارای اظهارنظر تعدیل‌شده (اظهارنظر مردود، عدم اظهارنظرو مشروط) می‌باشند، مورد بررسی و تحلیل جامع قرارگرفت و از بین ۱۴۴۰سال – شرکت (۱۸۰شرکت برای ۸ سال) بعنوان نمونه جامعه آماری، ۵۳۲ سال – شرکت، بعنوان شرکت مشکوک به تقلب در صورت‌های مالی و ۹۰۸ سال – شرکت، بعنوان غیر متقلب شناسایی گردید. شرکت‌های مشکوک به تقلب با عدد یک و شرکت‌های غیرمتقلب با عدد صفر نشان داده شده‌است.

متغیرمستقل: از نسبت‌های مالی در این پژوهش بعنوان متغیر مستقل یا متغیرهای پیش‌بینی‌کننده تقلب در صورت‌های مالی استفاده شده است. با بررسی مبانی نظری و پژوهش‌های انجام شده نسبت مالی در چهاردسته نقدینگی، اهرمی، کارایی وسودآوری استخراج گردیده که در بررسی ابتدایی برخی از نسبت‌ها که مشابه و معکوس یکدیگر بودن حذف گردید و در نهایت۹۶ نسبت‌مالی به شرح جدول ذیل باقی مانده است.

جدول۱- عناوین ۹۶ نسبت مالی

عنوان نسبت‌های مالی	شرح نسبت

نقدینگی	موجودی نقد به بدهی جاری، موجودی نقد به دارایی جاری، موجودی نقد عملیاتی به بدهی جاری، موجودی کالا به بدهی جاری، دارایی جاری به بدهی جاری، قدر مطلق تغییرات در نسبت جاری، مجموع موجودی کالا و حساب‌های دریافتی به کل دارایی، موجودی نقد به کل بدهی، بدهی جاری به کل دارایی، حساب‌های دریافتی به کل دارایی، سرمایه در گردش به کل دارایی، معادل نقد به بدهی جاری، دارایی جاری به کل دارایی، دارایی ثابت به بدهی-های غیرجاری، درصد تغییرات موجودی نقد عملیاتی، جریان نقدی عملیاتی به کل دارایی، دارایی آنی به بدهی جاری، موجودی نقد به کل دارایی، موجودی کالا به کل دارایی، موجودی کالا به دارایی جاری، دارایی جاری به بدهی جاری
اهرمی	کل بدهی به کل دارایی، هزینه بهره به کل بدهی، کل بدهی به کل دارایی منهای دارایی نامشهود، هزینه مالی به کل دارایی، حقوق صاحبان سهام به کل دارایی، بدهی بلند مدت به کل دارایی، بدهی بلند مدت به حقوق صاحبان سهام، کل بدهی به حقوق صاحبان سهام، فروش به کل بدهی، موجودی کالا به کل بدهی، سود قبل از بهره ومالیات به هزینه مالی(پوشش بهره)
کارایی	حساب‌های دریافتی به فروش، هزینه‌های عملیاتی به کل دارایی، تغییرات نسبت حساب‌های دریافتی به فروش، سود انباشته به کل دارایی، درصد تغییرات دارایی ثابت، تغییرات نسبت حساب‌های دریافتی به فروش نسبت به سال قبل، درصد تغییرات در بهای تمام شده فروش رفته، تغییرات موجودی کالا به فروش، درصد تغییرات کل دارایی، دارایی‌های ثابت به کل دارایی، فروش به کل دارایی، تغییرات موجودی کالا به تغییرات فروش، حساب‌های دریافتی به موجودی کالا، بهای تمام شده کالای فروش رفته به موجودی کالا، حساب‌های دریافتی به سود و زیان عملیاتی، فروش به دارایی ثابت، درصد تغییرات در موجودی کالا، دوره گردش موجودی کالا، سود قبل از مالیات به کل دارایی، حساب‌های دریافتنی به فروش * ۳۶۵، موجودی کالا به فروش، جریان نقدی عملیاتی به فروش، حساب‌های پرداختنی به خرید *۳۶۵، بهای تمام شده کالای فروش رفته به فروش، فروش به سرمایه در گردش، تغییرات حساب دریافتی به کل دارایی
سودآوری	سود قبل از مالیات به حقوق صاحبان سهام، سود قبل از مالیات به بدهی‌های جاری، سود خالص به دارایی ثابت، سود ناخالص به کل دارایی، سود قبل از بهره و مالیات به بدهی جاری، هزینه‌های عملیاتی به فروش، سود خالص به حقوق صاحبان سهام،درصد تغییرات در سود عملیاتی، درصد تغییرات فروش، سود خالص به کل دارایی، سود قبل از بهره و مالیات به کل دارایی، سود خالص به فروش، سود و زیان انباشته به سود خالص، موجودی نقد عملیاتی به سود

عملیاتی، درصد تغییرات در سود خالص، سود قبل از مالیات به سود قبل از بهره و مالیات، سود ناخالص به سود و زیان عملیاتی، سود عملیاتی به کل دارایی، سود عملیاتی منهای جریان نقدی عملیاتی به کل دارایی، سود قبل از بهره و مالیات به فروش، سود و زیان انباشته به حقوق صاحبان سهام، سود قبل از مالیات به فروش، سود ناخالص منهای جریان نقدی عملیاتی به فروش، حساب‌های دریافتی به سود و زیان عملیاتی، سود ناخالص به فروش، درصد تغییرات هزینه‌های عملیاتی، درصد تغییرات حاشیه سود ناخالص، فروش به حقوق صاحبان سهام، اقلام تعهدی به کل دارایی، سود خالص به سود ناخالص، سود عملیاتی به حقوق صاحبان سهام، درصد تغییرات سود ناخالص به فروش، سود قبل از بهره و مالیات به دارایی ثابت، درصد تغییرات سودخالص به فروش، سود عملیاتی به فروش، سود سهام پرداخت شده به سود خالص، سود عملیاتی به دارایی ثابت، سود قبل از مالیات به دارایی‌های ثابت، سود خالص به سرمایه در گردش

٦.یافته‌های پژوهش

در این پژوهش از روش‌های نظارت شده شامل ماشین بردار پشتیبان، K-نزدیک‌ترین همسایه، شبکه بیزین، روش ترکیبی (بگینگ)، شبکه‌های عصبی، درخت تصمیم، رگرسیون لجستیک و جنگل تصادفی در دو مرحله جهت پیش‌بینی احتمال تقلب در صورت‌های مالی بکار گرفته شده است. مرحله اول با استفاده از ۹۶ نسبت مالی استخراج شده از اطلاعات صورت‌های مالی شرکت‌های نمونه و مرحله دوم پس از کاهش این نسبت‌های مالی توسط الگوریتم ژنتیک استفاده شده است. سپس عملکرد روش‌های پیشنهادی جهت پیش‌بینی احتمال تقلب، قبل و بعد از کاهش نسبت‌های مالی مقایسه گردیده است.

برای بررسی و سنجش نسبت‌های استخراج شده از الگوریتم‌های پیشنهادی، ابتدا تکنیک‌های یادگیری آموزش داده می‌شود. برای این منظور از ۷۰ ٪ داده‌ها (۱۰۰۸داده که۳۷۶ داده شرکت‌های مشکوک به تقلب و۶۳۲ داده شرکت‌های غیرمتقلب) به عنوان داده آموزشی[۱] در نرم‌افزار متلب به الگوریتم‌های مورد استفاده در این پژوهش داده شده است تا میزان درصد آموزش مدل‌ها مشخص شود و در نهایت۳۰٪ بقیه داده‌ها (شامل۴۳۲ داده، ۱۵۶داده شرکت‌های مشکوک به تقلب و ۲۷۶ داده شرکت‌های غیرمتقلب) بعنوان داده‌های آزمایشی[۲] در نرم‌افزار اعمال گردیده تا ارزیابی اصلی الگوریتم‌ها صورت گرفته و میزان پیش‌بینی تقلب در صورت‌های مالی بررسی و ارزیابی گردد.

۱،۶. بررسی توانایی روش‌های پیشنهادی جهت پیش بینی تقلب بدون کاهش ویژگی:

۱،۶. ۱.معیارهای‌سنجش عملکرد روش‌های پیشنهادی جهت پیش بینی تقلب بدون کاهش ویژگی:

۱ Train

۲ Test

به منظور ارزیابی عملکرد توانایی روش‌های پیشنهادی طبقه برای پیش‌بینی تقلب از معیارهای ارزیابی زیر استفاده می‌شود (شارما و جین، ۲۰۱۹[۱]):

(۴) دقت: $Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP+TN+FP+FN}$

(۵) صحت: $\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$

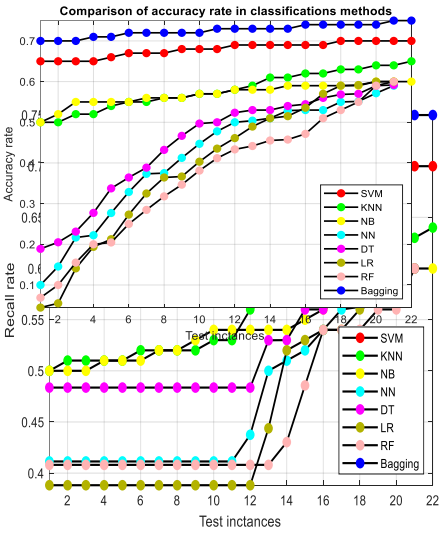
(۶) حساسیت: $Recall = \frac{\mathrm{TP}}{TP+FN}$

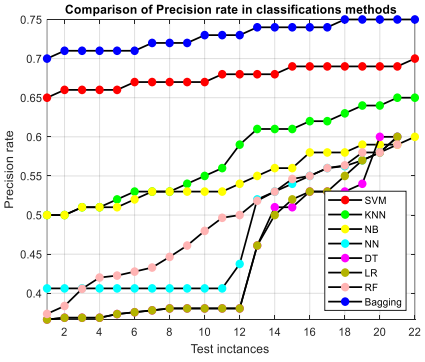
(۷) معیارF: $F - measure = \frac{2*Precision*Recall}{Precision+Recall}$

نتایج حاصل از این معیارهای ارزیابی عملکرد برای هر یک از روش‌های پیشنهادی با توجه به تمامی نسبت‌های مالی (۹۶ نسبت مالی) به شرح جدول 2 می‌باشد:

جدول2- خلاصه نتایج ارزیابی عملکرد روش‌های پیشنهادی بدون کاهش ویژگی

شرح	دقت	صحت	حساسیت	معیار F
روش ترکیبی (بگینگ)	٪ ۷۲/۴۵	٪ ۶۱/۲۱	٪ ۶۴/۷۴	٪ ۶۲/۹۳
ماشین بردار پشتیبان	٪ ۶۹/۴۴	٪ ۵۶/۹۰	٪ ۶۳/۴۶	٪ ۶۰
کای نزدیکترین همسایه	٪ ۶۶/۲۰	٪ ۵۲/۷۲	٪ ۶۲/۱۸	٪ ۵۷/۰۶
شبکه بیزین	٪ ۶۵/۵۱	٪ ۵۱/۹۸	٪ ۵۸/۹۷	٪ ۵۵/۲۶
شبکه عصبی	٪ ۵۴/۳۹	٪ ۳۸/۷۹	٪ ۴۵/۵۱	٪ ۴۱/۸۸
درخت تصمیم	٪ ۵۶/۱۶	٪ ۳۷/۳۴	٪ ۳۹/۷۴	٪ ۳۸/۵۰
رگرسیون لجستیک	٪ ۴۷/۹۱	٪ ۳۰/۹۳	٪ ۳۵/۸۹	٪ ۳۳/۲۲
جنگل تصادفی	٪ ۴۱/۸۹	٪ ۲۵/۳۸	٪ ۳۱/۴۱	٪ ۳۱/۶۰





۱ Sharma & Jain

شکل۳- خلاصه نتایج ارزیابی عملکرد الگوریتم‌های پیشنهادی بدون کاهش ویژگی

باتوجه به شکل بالا مشاهده می‌شود، در میان روش‌های طبقه‌بندی، روش ترکیبی(بگینگ) با عدد ۷۲/۴۵، ۶۱/۲۱، ۶۴/۷۴، ۶۲/۹۳ درصد نسبت به سایر روش‌های طبقه‌بندی به ترتیب دقت، صحت، حساسیت و معیار F بالاتری را نشان می‌دهد. روش ماشین بردارپشتیبان با عدد ۶۹/۴۴، ۵۶/۹۰، ۶۳/۴۶ و ۶۰ درصد رتبه دوم را بدست آورده‌اند.

۲،۱،۶. ماتریس آشفتگی روش‌های پیشنهادی برای پیش بینی تقلب بدون کاهش ویژگی:

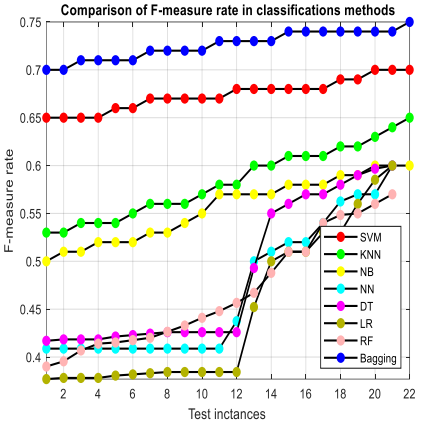
تعداد سطرها و ستون‌های ماتریس آشفتگی بستگی به تعداد طبقات[۱] دارد که دراین پژوهش دارای دو طبقه شرکت‌های غیرمتقلب و مشکوک به تقلب می‌باشد، بنابراین ماتریس آشفتگی در این پژوهش عبارت است از:

مثبت واقعی(TP): صورت‌های مالی که مشکوک به تقلب است و به درستی تشخیص داده شده‌اند.

مثبت کاذب (FP): صورت‌های مالی که مشکوک به تقلب است و به اشتباه غیرمتقلب تشخیص داده شده‌اند.

منفی واقعی(TN): صورت‌های مالی که غیرمتقلب است و به درستی تشخیص داده شده‌اند.

منفی کاذب(FN): صورت‌های مالی که غیرمتقلب است و به اشتباه مشکوک به تقلب تشخیص داده شده‌اند.



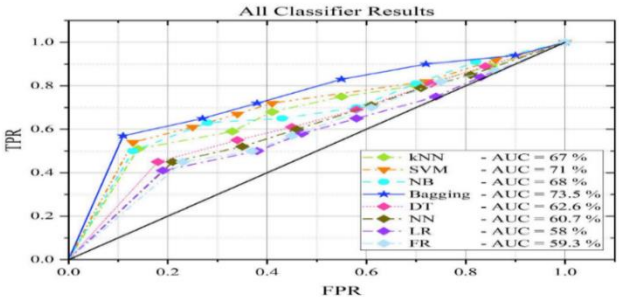
جدول۳- نتایج ماتریس آشفتگی روش‌های پیشنهادی بدون کاهش ویژگی

شرح	منفی کاذب (FN)	منفی واقعی (TN)	مثبت کاذب (FP)	مثبت واقعی (TP)
روش ترکیبی (بگینگ)	۵۵	۲۱۲	۶۴	۱۰۱
ماشین بردار پشتیبان	۵۷	۲۰۱	۷۵	۹۹
کای نزدیکترین همسایه	۵۹	۱۸۹	۸۷	۹۷
شبکه بیزین	۶۴	۱۹۱	۸۵	۹۲
شبکه عصبی	۸۵	۱۶۴	۱۱۲	۷۱
درخت تصمیم	۹۴	۱۷۲	۱۰۴	۶۲
رگرسیون لجستیک	۱۰۰	۱۵۱	۱۲۵	۵۶
جنگل تصادفی	۱۰۷	۱۳۲	۱۴۴	۴۹

۳،۱،۶.بررسی کارایی(ROC) روش‌های پیشنهادی برای پیش بینی تقلب بدون کاهش ویژگی:

۱ Class

منحنیROC ترسیم دو بعدی عملکرد روش‌های پیشنهادی است که درآن مثبت واقعی برروی نمودار X و مثبت کاذب برروی نمودارY ترسیم شده است. یک معیار متداول در این روش محاسبه مساحت زیر نمودار ROC است. میزان کارایی روش‌های طبقه‌بندی پیشنهادی، جهت پیش‌بینی احتمال تقلب در صورت‌های مالی، با توجه به مقادیر دقت، حساسیت، صحت، نرخ مثبت واقعی و نرخ مثبت کاذب در نمودار ROC ذیل نشان داده شده است.



شکل۴- سنجش و ارزیابی میزان کارایی روش‌های طبقه‌بندی پیشنهادی بدون کاهش ویژگی

مقادیر استخراج شده در خصوص ارزیابی معیارهای سنجش عملکرد، ماتریس آشفتگی و کارایی روش‌های پیشنهادی برای طبقه‌بندی با تمامی نسبت‌های مالی نسبتاً پایین و مناسب نمی‌باشد و نشان‌دهنده این موضوع می‌باشدکه باید با انتخاب ویژگی‌های مناسب و استفاده از آن‌ها به منظور طبقه‌بندی شرکت‌ها، باعث بهبود نتایج شویم.

۲,۶. انتخاب نسبت‌های مالی با استفاده از الگوریتم ژنتیک

الگوریتم‌های فرا ابتکاری برای سنجش کیفیت پاسخ‌های به دست آمده نیاز به یک تابع برازش دارند که عمدتا با عنوان تابع هدف نیز شناخته می‌شوند. از این رو، در راهکار پیشنهادی نیز، برای ارزیابی زیر مجموعه ویژگی‌های بدست آمده از الگوریتم ژنتیک، نیاز به در نظر گرفتن یک تابع هدف می‌باشد. مسئله انتخاب ویژگی، ذاتاً یک مسئله چند هدفه است. بدین معنا که از یکسو نیاز به کاهش تعداد ویژگی‌ها با انتخاب زیرمجموعه شبه بهینه از ویژگی‌ها است و از طرفی باید خطای پیش‌بینی مدل طبقه‌بندی را نیز کمینه نماید. از این رو، تابع هدف باید شامل هر دو مورد باشد. بر این اساس، راهکار گوناگونی تاکنون توسط محققین ارائه شده است که از جمله آنها، ترکیب هر دو هدف فوق بصورت مجموع و کمینه کردن هم زمان آن‌ها می‌باشد. بنابراین، رابطه زیر بعنوان تابع هدف در الگوریتم ژنتیک استفاده شده، برای ارزیابی پاسخ‌های یافت شده استفاده خواهد شد. تابع هدف مورد استفاده بصورت:

(۸)

$$\min f(x)=\text{Error Rate}+\frac{\#\text{selected feature}}{\text{Total Feature}}$$

که در رابطه فوق، هدف کمینه سازی همزمان نسبت تعداد نسبت‌های مالی انتخاب شده به تعداد کل نسبت-های مالی و خطای پیش‌بینی مدل دسته‌بندی می‌باشد. با کاهش خطای پیش‌بینی مدل دسته‌بندی کیفیت و

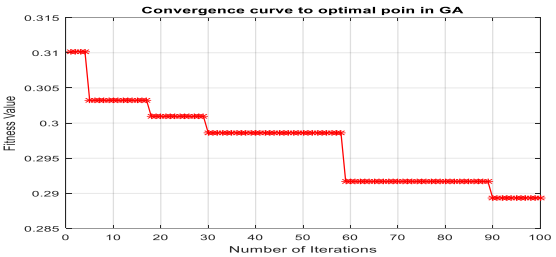
کارایی دسته‌بندی با انتخاب زیرمجموعه نسبت‌های مالی شبه بهینه افزایش پیدا می‌کند و از طرفی کاهش تعداد نسبت‌های مالی انتخاب شده نیز، پیچیدگی مدل دسته‌بندی کاهش و در نتیجه قابلیت تعمیم آن افزایش پیدا می‌کند

جدول زیر بهترین نسبت‌های مالی انتخاب شده توسط الگوریتم ژنتیک را مشخص کرده است.

جدول ۴- زیرمجموعه ویژگی‌های انتخاب شده توسط الگوریتم ژنتیک

عنوان نسبت مالی	تعداد تکرار ویژگی در ۳۰ مرتبه اجرا	عنوان نسبت مالی	تعداد تکرار ویژگی در ۳۰ مرتبه اجرا
کل بدهی به کل دارایی	۲۲	کل بدهی به حقوق صاحبان سهام	۱۷
سرمایه در گردش به کل دارایی	۲۲	سود و زیان انباشته به حقوق صاحبان سهام	۲۹
حسابهای دریافتنی به فروش	۳۰	دوره گردش موجودی کالا	۱۳
حسابهای دریافتنی به کل دارایی	۳۰		

این نسبت‌های مالی به عنوان ویژگی‌های بهینه از میان ویژگی‌های با همبستگی بالا برگزیده شده است. با توجه به این که معیار اصلی ارزیابی راه‌حل در الگوریتم ژنتیک، خطای تشخیص صورت‌های مالی است، از این رو هر چه مقدار تابع تناسب برای یک ذره کمتر باشد، ویژگی‌های انتخاب شده توسط آن ذره بهینه‌تر خواهد بود. با توجه به این‌که الگوریتم ژنتیک به صورت تصادفی مقدار دهی اولیه و عملکردهای تصادفی دارند، نتایج یک بار اجرای آن‌ها برای استناد کافی نیست، از این رو برای داشتن نتایج قابل استفاده و به دست آوردن دیدی از میانگین عملکرد الگوریتم‌های فرا ابتکاری، میانگین نتایج ۳۰ بار اجرای مستقل مورد استناد و بررسی قرار گرفته است.



شکل ۵- همگرایی تابع تناسب به نقطه بهینه در الگوریتم ژنتیک

همانطورکه درشکل ۵ نشان داده شده است، مقادیر تابع تناسب الگوریتم ژنتیک در مسئله انتخاب زیرمجموعه ویژگی‌ها با افزایش تکرار به نقطه بهینه با خطای صفر افزایش می‌یابد. مقدار تابع تناسب این الگوریتم پس از ۱۰۰ مرحله تکرار عدد ۰/۲۹۱۷ را بدست آورده است. و دقت نسبت‌های مالی انتخاب شده به ازای داده‌های آموزشی به منظور شناسایی صورت‌های مالی مشکوک به تقلب و غیر متقلب ۷۰/۸۳ ٪ می‌باشد. که پس از۸۹ مرحله تکرار بهترین نسبت‌های مالی را به دست آورده است.

۲،۶. ۱.روایی الگوریتم ژنتیک:

جهت بررسی روایی نسبت‌های مالی استخراج شده توسط الگوریتم زنتیک از داده‌های آزمایشی استفاده گردید که، نتایج زیر به دست آمده است.

جدول۵- روایی الگوریتم ژنتیک

عنوان	غیرمتقلب	مشکوک به تقلب	جمع	صحت‌پیش‌بینی
غیرمتقلب	۲۴۷	۲۹	۲۷۶	
مشکوک به تقلب	۵۳	۱۰۳	۱۵۶	٪ ۷۷/۷۶
جمع	۳۰۰	۱۳۲	۴۳۲	

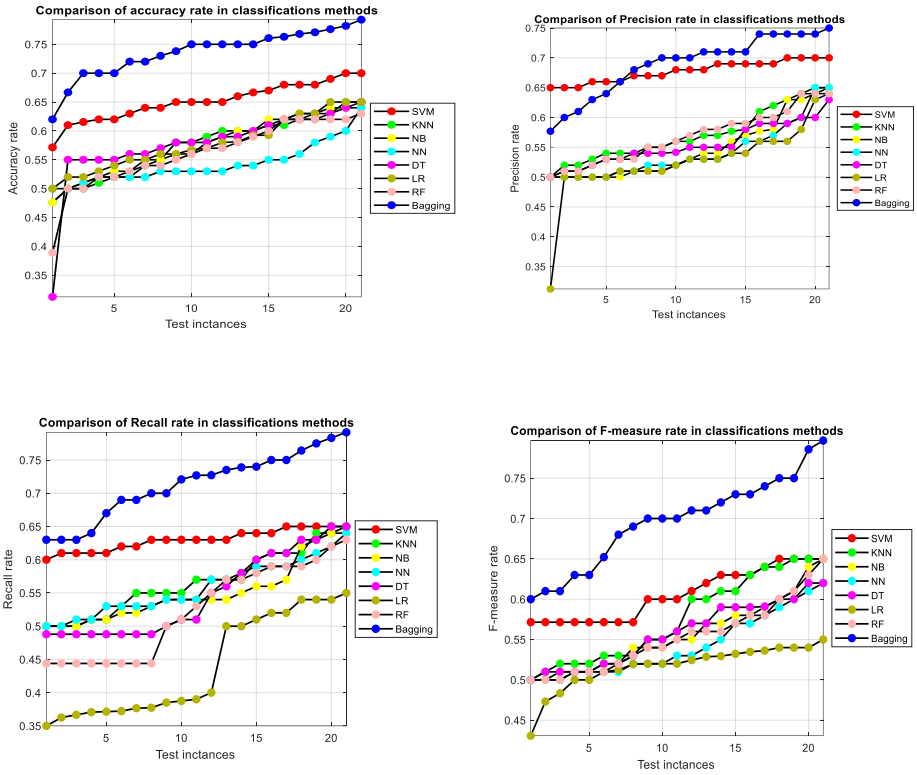
۳,۶. بررسی توانایی روش‌های پیشنهادی جهت پیش‌بینی تقلب با استفاده از نسبت‌های مالی انتخاب شده:

۳,۶. ۱.معیارهای‌سنجش عملکرد روش‌های پیشنهادی جهت پیش بینی تقلب با استفاده از نسبت‌های مالی انتخاب شده :

خلاصه نتایج توانایی روش‌های طبقه‌بندی پیشنهادی با ۷ نسبت مالی استخراج شده توسط الگوریتم ژنتیک با استفاده از چهار معیار ارزیابی عملکرد حاصل از ۳۰ مرتبه اجرا توسط داده‌های آزمایشی، به شرح جداول ذیل ارائه گردیده است.

جدول۶- خلاصه نتایج ارزیابی عملکرد روش‌های پیشنهادی با استفاده از نسبت‌های مالی انتخاب شده

شرح	دقت	صحت	حساسیت	معیار F
روش ترکیبی (بگینگ)	٪ ۸۰/۱۳	٪ ۷۶/۴۱	٪ ۶۵/۱۰	٪ ۷۰/۳۰
ماشین بردار پشتیبان	٪ ۷۶/۴۷	٪ ۶۹/۳۴	٪ ۶۲/۴۸	٪ ۶۵/۷۲
کای نزدیکترین همسایه	٪ ۷۴/۷۲	٪ ۶۵/۵۱	٪ ۶۳/۴۲	٪ ۶۵/۴۰
شبکه بیزین	٪ ۷۳/۲۵	٪ ۶۲/۹۸	٪ ۶۲/۹۰	٪ ۶۲/۷۳
شبکه عصبی	٪ ۷۲/۲۰	٪ ۶۱/۵۴	٪ ۶۰/۰۱	٪ ۶۰/۷۴
درخت تصمیم	٪ ۷۲/۱۹	٪ ۶۰/۰۴	٪ ۶۱/۳۱	٪ ۶۰/۶۵
رگرسیون لجستیک	٪ ۷۲/۵۱	٪ ۶۰/۴۷	٪ ۶۱/۸۹	٪ ۶۱/۱۴
جنگل تصادفی	٪ ۷۲/۲۲	٪ ۵۹/۹۶	٪ ۶۱/۹۲	٪ ۶۰/۹۰



شکل۶- خلاصه نتایج ارزیابی عملکرد الگوریتم‌های پیشنهادی با استفاده از نسبت‌های مالی انتخاب شده

باتوجه به شکل بالا مشاهده می‌شود، در میان روش‌های طبقه‌بندی، روش ترکیبی(بگینگ) با عدد ۸۰/۱۳، ۷۶/۴۱، ۶۵/۱۰، ۷۰/۳۰ درصد نسبت به سایر روش‌های طبقه‌بندی به ترتیب دقت، صحت، حساسیت و معیار F بالاتری را نشان می‌دهد. روش ماشین بردارپشتیبان با عدد ۷۶/۴۷، ۶۹/۳۴، ۶۲/۴۸، ۶۵/۷۲ درصد رتبه دوم را بدست آورده‌اند.

۲،۳،۶. ماتریس آشفتگی روش‌های پیشنهادی برای پیش‌بینی تقلب با استفاده از نسبت‌های مالی انتخاب شده :

نتایج بدست آمده حاصل از ماتریس آشفتگی توسط داده‌های آزمایشی، مربوط به روش‌های پیشنهادی و نسبت‌های مالی استخراج شده، در جدول ۷ ارائه شده است.

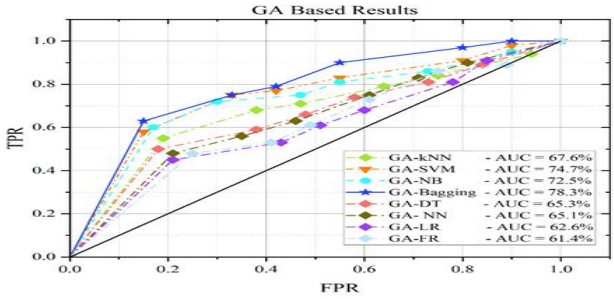
جدول۷- نتایج ماتریس آشفتگی روش‌های پیشنهادی با استفاده از نسبت‌های مالی انتخاب شده

شرح	منفی کاذب (FN)	منفی واقعی (TN)	مثبت کاذب (FP)	مثبت واقعی (TP)
روش ترکیبی (بگینگ)	۵۳	۲۴۷	۲۹	۱۰۳
ماشین بردار پشتیبان	۵۶	۲۳۵	۴۱	۱۰۰
کای نزدیکترین همسایه	۵۵	۲۲۷	۴۹	۱۰۱

شبکه بیزین	۵۶	۲۲۰	۵۶	۱۰۰
شبکه عصبی	۷۷	۱۹۱	۸۵	۷۹
درخت تصمیم	۷۷	۱۹۰	۸۶	۷۹
رگرسیون لجستیک	۸۶	۱۸۲	۹۴	۷۰
جنگل تصادفی	۹۷	۱۷۱	۱۰۵	۵۹

۳،۲،۶.بررسی کارایی(ROC) روش‌های پیشنهادی برای پیش‌بینی تقلب با استفاده از نسبت‌های مالی انتخاب شده:

میزان کارایی نسبت‌های مالی استخراج شده توسط الگوریتم ژنتیک با روش‌های طبقه‌بندی پیشنهادی، با توجه به مقادیر دقت، حساسیت، صحت، نرخ مثبت واقعی و نرخ مثبت کاذب در نمودارROC شکل ۷ نشان داده شده است.



شکل۷- سنجش و ارزیابی میزان کارایی روش‌های طبقه‌بندی پیشنهادی با استفاده از نسبت‌های مالی انتخاب شده

۷.بحث و نتیجه‌گیری

صورت‌های مالی منتشر شده شرکت‌های پذیرفته شده در بورس اوراق بهادار، مهمترین منبع اطلاعاتی در بازار سرمایه است که بایستی درباره وضعیت مالی و عملکرد یک شرکت برای استفاده‌کنندگان از این اطلاعات (سرمایه‌گذاران، اعتباردهندگان، مدیران و...) جهت تصمیم‌گیری‌های اقتصادی و تخصیص سرمایه‌شان، شفاف و قابل اتکا باشند. تقلب در صورت‌های مالی علاوه براینکه باعث به هدر رفتن منابع و عدم دستیابی به اهداف سازمانی می‌گردد، می‌تواند باعث به وجود آمدن بحران‌های اقتصادی در سطح کلان و باعث ایجاد نگرانی و سلب اعتماد عمومی نسبت به حرفه حسابداری و حسابرسی وگزارش‌های مالی شود (رضایی و رایلی،۲۰۱۰)[۱]. باتوجه به روند خصوصی‌سازی و افزایش تعداد شرکت‌های پذیرفته‌شده در بورس اوراق بهادار کشور و عضویت این سازمان در جمع اعضای سازمان بین‌المللی کمیسیون‌های اوراق بهادار، بدون شک نیاز به ارتقاء کیفیت اطلاعات مالی و پیش‌بینی احتمال تقلب در صورت‌های مالی ضروری و از اهمیت زیادی برخوردار است.

۱ Rezaee & Riley

از طرفی با توجه به پیشرفت فناوری اطلاعات، روش‌های انجام تقلب پیچیده‌تر و سریع‌تر شده است و روش‌های آماری بدلیل این‌که برای داده‌های پیوسته و خطی قابلیت پیشبینی بیشتری نسبت به داده‌های گسسته و غیرخطی دارد (رانگاناتان و همکاران،۲۰۱۷)[۱]. لذا امکان پیش‌بینی درست تقلب با این روش‌ها کاهش می‌یابد. به همین دلیل توسعه روش‌های کشف تقلب درصورت‌های مالی از روش‌های آماری و سنتی به سمت تکنیک‌های هوش مصنوعی الزامی است.

همچنین با عنایت به حجم بالای اطلاعات مربوط به صورت‌های مالی که ممکن است برخی از این اطلاعات زائد، نامربوط و منسوخ بوده و باعث کاهش کارایی روش‌های مورد استفاده جهت پیش‌بینی احتمال تقلب گردد، لذا در پژوهش حاضر بدنبال انتخاب نسبت‌های مالی مناسب توسط الگوریتم ژنتیک، به منظور افزایش توانایی و کارایی عملکرد روش‌های پیشنهادی (ماشین بردارپشتیبان، کای نزدیک‌ترین همسایه، شبکه بیزین، روش ترکیبی(بگینگ)، شبکه عصبی، درخت تصمیم، رگرسیون لجستیک و جنگل تصادفی) که از روش‌های نظارت شده، تکنیک‌های هوش مصنوعی هستند، جهت پیش‌بینی احتمال تقلب در صورت‌های مالی شرکت‌ها می‌باشیم.

نتایج بدست آمده در این پژوهش نشان داد، که عملکرد تمامی روش‌های پیشنهادی، پس از کاهش نسبت‌های مالی توسط الگوریتم ژنتیک، نسبت به قبل از کاهش این نسبت‌ها بهبود یافته و معیارهای دقت، صحت و کارایی بالاتری داشته اند، که بیان‌گر تایید فرضیه پژوهش"کاهش نسبت‌های مالی باعث بهبود عملکرد روش‌های طبقه-بندی جهت پیش‌بینی احتمال تقلب می‌باشد"، است.

۸. پیشنهادهای پژوهش

با توجه به اهمیت تقلب درصورت‌های مالی شرکت‌ها موارد ذیل بعنوان پیشنهاد جهت پیش‌بینی و پیشگیری ارائه می‌گردد:

۱- با توجه به این‌که شناسایی شرکت‌های که مرتکب تقلب در صورت‌های مالی می‌شوند بسیار سخت، تخصصی و زمانبر هست و اکثر استفاده کنندگان از اطلاعات مالی، تخصص لازم و کافی جهت در این خصوص را ندارند، پیشنهاد می‌شود، سازمان یا نهادی رسیدگی به مقوله گزارش‌گری مالی متقلبانه را بیش از پیش مورد توجه قرار داده و با تشکیل هیأت تخصصی شرکت‌هایی که مرتکب تقلب می شوند را شناسایی و اطلاعات مربوط به این شرکت‌ها را در دسترس عموم قرار دهند.

۲- از آنجائیکه الگوریتم‌های هوش مصنوعی و هوشمند با توجه به حجم بالای داده‌ها از دقت و سرعت پیش‌بینی بالایی برخوردار هستند، به پژوهش‌گران پیشنهاد می‌شود در پژوهش‌های خود از این روش‌ها جهت کشف سریع‌تر موارد تقلب و تحمیل زیان کمتر به ذینفعان استفاده کنند.

۳- در این پژوهش از شرکت‌های تجاری بعنوان جامعه آماری استفاده شده است پیشنهاد می شود از مدل این پژوهش برای سایر صنایع مانند شرکت‌های سرمایه‌گذاری، هلدینگ‌ها، بیمه‌ها و موسسات مالی و اعتباری و بیمه‌ها استفاده شود.

۱ Ranganathan et al

۴- برای پیش‌بینی تقلب در صورت‌های مالی در این پژوهش از نسبت‌های مالی استفاده شده است. توصیه می‌شود در پژوهش‌های آینده از عوامل تاثیرگذار دیگری که غیرمالی (ویژگی‌های شخصیتی کارکنان حسابرسان و مدیران،کمیته‌های حسابرسی و کنترل‌های داخلی و...) هستند، استفاده شود.

۵- در این پژوهش از سه روش فراابتکاری استفاده شده است، توصیه می‌شود، از سایر الگوریتم‌های بهینه‌سازی مانند طوفان مغزی، جهش قورباغه، فاخته، جستجوی سنجاب و... استفاده نموده و نتایج مقایسه گردد.

۶- با توجه به این‌که ممکن است پژوهش‌های مشابهی در سایر کشورها انجام گرفته باشد، پیشنهاد می شود، مقایسه تطبیقی بین کشورهای مختلف در رابطه با این عوامل در هر یک از کشورها می‌تواند در قالب پژوهش‌های آتی مورد توجه قرار گیرد.

۷- پیشنهاد می شود از مدل پیشنهادی این پژوهش جهت پیش‌بینی احتمال ورشکستگی در شرکت‌ها استفاده شود.

References

Aftabi, S. Z., Ahmadi, A., & Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. Expert Systems with Applications, 227, 120144.

Arboleda, F. J. M., Guzman-Luna, J. A., & Torres, I. D. (2018). Fraud detection-oriented operators in a data warehouse based on forensic accounting techniques. Computer Fraud & Security, 2018(10), 13-19.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

Bou Rjeily, C., Badr, G., Hajjarm El Hassani, A., & Andres, E. (2019). Medical data mining for heart diseases and the future of sequential mining in medical field. Machine learning paradigms: Advances in data analytics, 71-99.

Chen, Y. (2023). Financial Statement Fraud Detection based on Integrated Feature Selection and Imbalance Learning. Frontiers in Business, Economics and Management, 8(3), 46-48.

Cheng, C. H., Kao,Y.F., &lin, H. P. (2021). A financial statement fraud model based on synthesized attribute selection and a dataset with missing values and imbalanced classes. Applied Soft Computing 108: 107487.

Chimonaki, C., Papadakis, S., Vergos, K., & Shahgholian, A. (2019). Identification of financial statement fraud in Greece by using computational intelligence techniques. In Enterprise Applications, Markets and Services in the Finance Industry: 9th International Workshop, FinanceCom 2018, Manchester, UK, June 22, 2018, Revised. Springer International Publishing.

Committee of Audit Standards (2020). Standards and Principles of Accounting and Auditing: Audit Standards. Audit Organisation Publications: Tehran, Iran.(In Persian)

Corruption Perceptions Index. (2021).https://www.transparency.org/en/cpi

Dua, S., and X. Du. (2011). Data mining and machine learning in cybersecurity. 1st edition, Taylor and Francis Group.

Dalnial, H., A. Kamaluddin., Z. Mohd Sanusi & K. Syafiza Khairuddin. (2014). Accountability in financial reporting: detecting fraudulent firms, Procedia - Social and Behavioral Sciences145: 61– 69.

Dechow, M. P. , Sloan, R. G. , & Sweeney,A. P. (1996). “ Couses And ConsequencesOf Earnings Manipulation: An Analysis Of Firms Subject To Enforcement Actions By The Sec″ Contemporary Accounting Research

Dechow, M. P. , Sloan, R. G. , & Sweeney,A. P. (1995). " Detecting earnings management " The Accounting Review, 70(2) , 193.

Ebrahimi, Mehrdad; Khajavi, Shokrollah (2017). Modeling Variables Affecting Fraud Detection in Financial Statements through Data Mining Techniques. Financial Accounting (In Persian)

Emadoddini, A., & Saeedi, P. (2019). Investigating the relationship between unusual audit fees and accruals management and actual earnings management. Journal of Accounting and Management Vision, 2(9), 78-97.

Fama, E. F. and Jensen, M. C. (1983).“Scparation of ownership and control”,Journal of Law and Economics

Guo,G., Wang, H., Bell, D., Bi, Y., & Greer,K. (2003).KNN Model-Based Approach inClassification, Lecture Notes in Computer Science

Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.

Hidayattullah, S., Surjandari, I., & Laoh, E. (2020). Financial Statement Fraud Detection in Indonesia Listed Companies using Machine Learning based on Meta-Heuristic Optimization. International Workshop on Big Data and Information Security (IWBIS). IEEE. 79-84.

Ibadin, P. O., & Kemebradikemor, E. (2020). Tax Fraud in Nigeria: A Review of Causal Factors. Journal of Taxation and Economic Development, 19(1), 64-80.

Jan, C. L. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. Sustainability 10(2): 513.

Kazemi, T., & Piri, P. (2023). Financial Reporting Fraud Scheme Prediction via Machine Learning Approach–Multiclass Classification. Empirical Research in Accounting, 12(4), 255-280.(In Persian)

Kirkos, S., C. Spathis & Y. Manolopoulos. (2007). Data Mining Techniques for the Detection of Fraudulent Financial Statements, Expert Systems with Applications 32: 995-1003.

Kamrani, H., & Abedini, B. (2022). Formulation of Financial Statement Fraud Detection Model Using Artificial Neural Network and Support Vector Machine Approaches in Companies Listed in Tehran Bahador Stock Exchange. Journal of Management Accounting and Auditing Knowledge, 11(41), 285-314 (In Persian)

Kotsiantis, S., (2006), ″Method Of Financing″, Australian Accounting Review (Springer-Verlag Berlin, Heidelberg@ 2006).

Kuncheva, L. I. (2014). Combining pattern classifiers: methods and algorithms. John Wiley & Sons

Leung, K. M. (2007). Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2007, 123-156.

Occupational Fraud. (2022).A Report To Nations ,https://acfepublic.s3.us-west-mazonaws.com

Omar, N., Z. A. Johari & M. Smith. (2017). Predicting fraudulent financial reporting using artificial neural network. Journal of Financial Crime. 24(2):362-387.

Omidi, M., Q. Min., V. Moradinaftchali & M. Piri. (2019). The efficacy of predictive methods in financial statement fraud. Discrete Dynamics in Nature and Society 4989140: 1-12.

Persons, Obeua S., (2011). Using Financial Statement Data to Identify Factors Associated with Fraudulent Financial Reporting, Journal of Applied Business Research 3: 38-46.

Pradhan, A. (2012). Support vector machine-a survey. International Journal of Emerging Technology and Advanced Engineering, 2(8), 82-85.

Park, Y., & Reeves, D. (2011, March). Deriving common malware behavior through graph clustering. In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security.

Ramos,M.(2003). Auditors responsibility for fraud detection. the Journal of accountancy.

Ranganathan, P., Pramesh,C.S., & Aggarwal,R. (2017). Common pitfalls in statistical analysis: Logistic regression. Perspectives in clinical research 8(3): 148

Rezaee, Z., & Riley, R. (2011). Financial Statement Fraud Defined. John Wiley & Sons.

Rezaei, Mahdi; Nazemi Ardakani, Mahdi; Naser Sadr Abadi, Alireza (2020). Fraud Detection in Financial Statements through Audit Reports of Financial Statements, Management Accounting Journal, (45): 141–153. (In Persian).

Ross, S., Randolph, W., & Jordan, B. (2013). Modern Financial anagement, translated by Ali Jahankhani & Mojtaba Shouri, SAMT Publications, Tehran, Iran.

Sadgali, I., N. Sael & F. Benabbou. (2019). Performance of machine learning techniques in the detection of financial frauds. Procedia computer science 148: 45-54.

Sedighi Kamal, L. (2013). Fraud in Financial Statements Based on Reports by Association of Certified Fraud Examiners. Auditor (In Persian)

Sharma, S., & Jain, A. (2019). Dynamic social network analysis and performance evaluation. International Journal of Intelligent Engineering Informatics, 7(2-3), 180-202.

Shinde, A., Sahu, A., Apley, D., & Runger, G. (2014). Preimages for variation patterns from kernel PCA and bagging. Iie Transactions, 46(5), 429-456.

Spathis, C. , (2002). Detecting False Financial Statements Using Published Data:Some Evidence from Greece. Managerial Auditing Journal.

Taghavi, M., & Nazak, N. (2006). Use of Evolutionary Algorithms in Data Mining. First International Conference on Research Methods in Engineering. (In Persian)

Tashdidi, Elaheh; Sepasi, Sahar; Etemadi, Hossein; Azar, Adel (2019). A Novel Approach to Prediction and Detection of Fraud in Financial Statements through Bee Colony Optimisation Algorithm. Accounting Knowledge (In Persian)

Umar, H., Purba, R. (2020), "HU Model: Incorporation of Fraud Star in Detection of Corruption", International Journal of Economics and Management Studies, 13(6).

Vakilifard, H., Jabarzadeh Kangarlouei, S., Pourreza, A., & Ahmadi, S. (2009). Analysis of Fraud Characteristics in Financial Statements. Accountant(In Persian)

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. Decision support systems, 57, 77-93.

Watts, Rl. , &, Zimerman, J. L., (1986),Pozetive Accounting Theory ″, 2(124)

Xiuguo, W., & Shengyong, D. (2022). An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning. IEEE Access, 10, 22516-22532.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on neural networks, 16(3), 645-678.

Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: a multi-analytic approach. Sustainability, 11(6), 1579.

Youkhneh Alghiani, Mariam; Jamal Bahri Sales, Jamal; Jabarzadeh Kangarlouei, Saedi; Zavari Rezaei, Akbar (2021). Explaining Financial Tax Cross Reporting of Companies: Hybrid Method of Classic Data Mining, ANFIS, and Metaheuristic Algorithms, Empirical Studies of Financial Accounting, 18 (71), 89–111.(In Persian)

Zareh Bahmanmiri, M, & Malekian Kaleh Basi, E. (2015). Fraud Prediction in Financial Statements through Financial Ratios. Financial Management Perspective (In Persian)