



Comparing the Estimation Power of Machine Learning Models and Statistical Models in Predicting Profit Component Changes and Selecting the Optimal Model

Maryam Izadi

MSc., Department of Accounting Management, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: m.izadi@urmia.ac.ir

Ali Ashtab*

*Corresponding Author, Assistant Prof, Department of Accounting, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: a.ashtab@urmia.ac.ir

Akbar Zavari Rezaei

Assistant Prof, Department of Accounting, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: a.zavarirezaei@urmia.ac.ir

Abstract

Objective

The aim of predicting profit changes is to create awareness for investors, financial analysts, managers, stock market officials, creditors, and other users to judge the business unit, make decisions about buying or selling stocks, or granting or denying loans and credits. The goal of this research is to evaluate the performance and compare the accuracy of machine learning models and statistical models in predicting the direction of changes in three profit components including net profit (loss), gross profit (loss), and operating profit (loss).

Methods

In this research, using the financial information of 139 manufacturing companies listed on the Tehran Stock Exchange over a 15-year period, from 2008 to 2022, and employing

Citation: Izadi, Maryam; Ashtab, Ali & Zavari Rezaei, Akbar (2025). Comparing the Estimation Power of Machine Learning Models and Statistical Models in Predicting Profit Component Changes and Selecting the Optimal Model. *Financial Research Journal*, 27(1), 31- 57. <https://doi.org/10.22059/FRJ.2024.373472.1007580> (in Persian)



25 machine learning models and 10 statistical models, the efficiency of machine learning models and statistical models in predicting the direction of changes in profit components including net profit (loss), gross profit (loss), and operating profit (loss) has been compared. In the present study, Excel software was used for data sorting, Eviews software for extracting descriptive statistics, and data mining software SPSS Modeler and Rapidminer for predicting profit changes. The performance of machine learning models was evaluated using two criteria: accuracy (predictive accuracy of the model) and AUC (area under the curve), and the performance of statistical models was evaluated only by the accuracy criterion. Finally, in order to select the model with the best performance for predicting the direction of changes in net profit (loss), gross profit (loss), and operating profit (loss), the best model among the machine learning models was chosen using the ROC curve.

Results

After calculating the average predictive accuracy of machine learning and statistical models, it was found that the average predictive accuracy of machine learning models for dependent variables including the percentage of changes in net profit (loss), percentage of changes in gross profit (loss), and percentage of changes in operating profit (loss) ranges from 83% to 93%. It was also found that the average predictive accuracy of statistical models for all three profit components varies from 76% to 83%. After confirming the non-normality of the average accuracy of machine learning and statistical models for profit components using the Kolmogorov-Smirnov test, the non-parametric Mann-Whitney U test was used to compare the predictive accuracy of machine learning models and statistical models in predicting the direction of changes in profit components.

Conclusion

The results of the research hypotheses test indicate the high efficiency of machine learning models in predicting the direction of changes in net profit (loss), gross profit (loss), and operating profit (loss), compared to statistical models. The ROC curve results indicate that the decision tree model achieved a predictive accuracy of 100% in forecasting the direction of changes in net profit (loss) and 99.38% accuracy in predicting the direction of changes in gross profit (loss). Additionally, the rule-based inference model demonstrated a predictive accuracy of 86.76% for forecasting the direction of changes in operating profit (loss). These models exhibited the best performance and were selected as the optimal models.

Keywords: Profit change direction, Statistical models, Machine learning models.

بررسی و تطبیق قدرت تخمین مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی جهت تغییرات اجزای سود و انتخاب مدل بهینه

مریم ایزدی

کارشناس ارشد، گروه مدیریت حسابداری، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: m.izadi@urmia.ac.ir

علی آشتاب*

* نویسنده مسئول، استادیار، گروه حسابداری، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: a.ashtab@urmia.ac.ir

اکبر زواری رضایی

استادیار، گروه حسابداری، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: a.zavarirezaei@urmia.ac.ir

چکیده

هدف: هدف از پیش‌بینی تغییرات سود، آگاهی‌دادن به سرمایه‌گذاران، تحلیلگران مالی، مدیران، مسئولان بازار سهام، اعتباردهندگان و سایر استفاده‌کنندگان، به منظور قضاوت کردن واحد تجاری، تصمیم‌گیری برای خرید یا فروش سهام و اعطاء یا عدم اعطای وام و اعتبارات است. این پژوهش با هدف ارزیابی عملکرد و مقایسه دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری، برای پیش‌بینی جهت تغییرات سه مؤلفه سود، از جمله سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی اجرا شده است.

روش: در این پژوهش با استفاده از اطلاعات مالی ۱۳۹ شرکت تولیدی پذیرفته شده در بورس اوراق بهادار تهران، در دوره زمانی ۱۵ ساله، طی سال‌های ۱۳۸۷ تا ۱۴۰۱ و با به کارگیری ۲۵ مدل یادگیری ماشین و ۱۰ مدل آماری، به بررسی مقایسه کارایی مدل‌های یادگیری ماشین و مدل‌های آماری برای پیش‌بینی جهت تغییرات اجزای سود از جمله سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی پرداخته شده است. در پژوهش حاضر از نرم‌افزار اکسل برای مرتب‌سازی داده‌ها، نرم‌افزار ایویوز برای استخراج آمار توصیفی و از نرم‌افزارهای داده‌کاوی اس‌پی‌اس‌اس مدلر و ریپدماینر برای مدل‌سازی پیش‌بینی جهت تغییرات سود استفاده شده است. ارزیابی عملکرد مدل‌های یادگیری ماشین، از طریق دو معیار دقت پیش‌بینی مدل (accuracy) و ناحیه زیر منحنی (AUC) و ارزیابی عملکرد مدل‌های آماری تنها با معیار دقت پیش‌بینی مدل انجام شده است. در نهایت، به منظور انتخاب مدلی که بهترین عملکرد را برای پیش‌بینی جهت تغییرات سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی دارد، از میان مدل‌های یادگیری ماشین، به انتخاب مدل بهینه با استفاده از منحنی راک پرداخته شده است.

یافته‌ها: پس از محاسبه متوسط دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری، مشخص شد که متوسط دقت پیش‌بینی مدل‌های یادگیری ماشین برای متغیرهای وابسته (درصد تغییرات سود (زیان) خالص، درصد تغییرات سود (زیان) ناخالص و درصد تغییرات

استناد: ایزدی، مریم؛ آشتاب، علی و زواری رضایی، اکبر (۱۴۰۴). بررسی و تطبیق قدرت تخمین مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی جهت تغییرات اجزای سود و انتخاب مدل بهینه. *تحقیقات مالی*، ۲۷(۱)، ۳۱-۵۷.

سود (زیان) عملیاتی)، از ۸۳ درصد تا ۹۳ درصد و متوسط دقت پیش‌بینی مدل‌های آماری برای هر سه مؤلفه سود، از ۷۶ درصد تا ۸۳ درصد متغیر است. پس از احراز نرمال نبودن متوسط دقت مدل‌های یادگیری ماشین و مدل‌های آماری برای اجزای سود توسط آزمون کلموگروف اسمیرنوف، از آزمون ناپارامتریک یومن ویتنی برای مقایسه دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری برای پیش‌بینی جهت تغییرات اجزای سود استفاده شد.

نتیجه‌گیری: نتایج آزمون فرضیه‌های پژوهش، بیانگر کارایی بالای مدل‌های یادگیری ماشین برای پیش‌بینی جهت تغییرات سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی نسبت به مدل‌های آماری است. نتایج منحنی راک نیز نشانگر آن است که مدل درخت تصمیم با دقت پیش‌بینی معادل ۱۰۰ درصد، برای پیش‌بینی جهت تغییرات سود (زیان) خالص و دقت پیش‌بینی معادل ۳۸/۹۹ درصد برای پیش‌بینی جهت تغییرات سود (زیان) ناخالص و مدل استنتاج قوانین با دقت پیش‌بینی معادل ۷۶/۸۶ درصد در پیش‌بینی جهت تغییرات سود (زیان) عملیاتی، بهترین عملکرد را داشتند و به‌عنوان مدل بهینه انتخاب شدند.

کلیدواژه‌ها: جهت تغییرات سود، مدل‌های آماری، مدل‌های یادگیری ماشین.



مقدمه

توسعه مدل‌های پیش‌بینی سود شرکت‌ها، برای محققان حسابداری و متخصصان سرمایه‌گذاری اهمیت بسیار زیادی دارد. پیش‌بینی معیارهای مالی کلیدی، مانند سود خالص و جریان‌های نقدی عملیاتی، به کسب‌وکارها قدرت می‌دهد تا در خصوص استراتژی‌های سرمایه‌گذاری، تخصیص منابع و برنامه‌ریزی مالی کلی، تصمیم‌های آگاهانه‌ای اتخاذ کنند. توانایی پیش‌بینی عملکرد مالی آتی، به سازمان‌ها اجازه می‌دهد تا عملیات را بهینه کنند، ریسک‌ها را کاهش دهند و فرصت‌ها را به‌موقع استفاده کنند. در نتیجه، تقاضای فزاینده‌ای برای مدل‌های پیش‌بینی کارا وجود دارد که می‌تواند پیش‌بینی‌های دقیق و قابل اعتمادی را برای این شاخص‌های مالی مهم ارائه دهد (پارلینا و بودیان‌تو^۱، ۲۰۲۱). با این حال پیش‌بینی سودهای آتی دشوار است؛ زیرا با جنبه‌های متعدد یک شرکت به شیوه‌ای پیچیده و با راهنمایی اندکی از ادبیات نظری مرتبط است (لو و گو^۲، ۲۰۱۶؛ موناهان^۳، ۲۰۱۸). در این پژوهش جهت تغییرات سود، به چند دلیل بررسی شده است: نخست اینکه پیش‌بینی سطح سودهای آتی و میزان تغییرات سود (اختلاف سودهای جاری و گذشته) دشوار است؛ زیرا پژوهش‌های موجود نشان می‌دهند که پیش‌بینی سود بر اساس ویژگی‌های شرکت، دقت چشمگیری از پیش‌بینی‌های به‌دست‌آمده با استفاده از مدل‌سازی تصادفی ندارد (گراکوس و گرامسی^۴، ۲۰۱۳؛ لی و موهان‌رام^۵، ۲۰۱۴). دوم اینکه فریمن، اولسون و پنمن^۶ (۱۹۸۲) استدلال می‌کنند که تغییرپذیری در تغییرات سود، آن‌قدر بزرگ است که نمی‌توان آن را با تغییر در تغییرات سود مورد انتظار مشروط به متغیرهای توضیحی مقایسه کرد. آن‌ها پیشنهاد می‌کنند که تغییرپذیری در تغییرات سود را با تبدیل مقدار به جهت تغییرات سود کاهش دهند و پیش‌بینی کنند که کدام یک قابل دستیابی است. از سوی دیگر پیش‌بینی جهت تغییرات سود از نظر اقتصادی معنادار است و قابلیت اجرا کردن دارد؛ زیرا پژوهش‌های گسترده‌ای، پرتفوی‌هایی را براساس جهت تغییرات سود ساخته‌اند (او و پنمن^۷، ۱۹۸۹؛ والن و ویلند^۸، ۲۰۱۱). در زمینه مدل‌سازی برای پیش‌بینی، دو رویکرد برجسته جذابیت زیادی به‌دست آورده‌اند: مدل‌های آماری و مدل‌های یادگیری ماشین. مدل‌های آماری مدت‌هاست که به‌عنوان ابزاری اساسی در تحلیل و پیش‌بینی داده‌ها استفاده می‌شوند. این مدل‌ها بر اساس مفروضات از پیش تعریف شده و الگوریتم‌های ریاضی، با استفاده از فن‌هایی مانند تحلیل رگرسیون، تحلیل سری‌های زمانی و مدل‌سازی اقتصادسنجی ساخته شده‌اند (موناهان، ۲۰۱۸). هدف مدل‌های آماری بررسی الگوهای داده‌های تاریخی، گرفتن روابط و پویایی‌های اساسی برای ایجاد پیش‌بینی برای نتایج مالی آینده است (مارتینز^۹، ۲۰۲۲). از سوی دیگر، مدل‌های یادگیری ماشین، به‌عنوان جایگزینی قدرتمند ظاهر شده‌اند و از فن‌های محاسباتی پیشرفته برای یادگیری از داده‌ها و پیش‌بینی بدون برنامه‌نویسی صریح استفاده می‌کنند (پتروپولوس، سیاکولیس،

1. Parlina & Budianto
2. Lev & Gu
3. Monahan
4. Gerakos & Gramacy
5. Li & Mohanram
6. Freeman, Ohlson & Penman
7. Ou & Penman
8. Wahlen & Wieland
9. Martins

استاورولاکیس و ولاچوگیاناکیس^۱، ۲۰۲۰). مدل‌های یادگیری ماشین، مانند شبکه‌های عصبی مصنوعی، درختان تصمیم‌گیری و ماشین‌های بردار پشتیبان، در مدیریت مجموعه داده‌های پیچیده با ابعاد زیاد قابلیت‌های چشمگیری را نشان می‌دهند (آشتاب، حقیقت و کردستانی، ۱۳۹۶). این مدل‌ها در بررسی روابط غیرخطی، شناسایی الگوهای پیچیده و تطبیق با دینامیک داده‌های در حال تغییر برتری دارند. حوزه یادگیری ماشینی در سال‌های اخیر شاهد پیشرفت‌های چشمگیری بوده که به افزایش کاربرد آن در حوزه‌های مختلف، از جمله امور مالی منجر شده است (داستیل، چلیک و پوتسان^۲، ۲۰۲۰). با توجه به مراتب فوق، سؤال اصلی پژوهش حاضر این است که آیا می‌توان به گونه‌ای منطقی و بر اساس الگوهای یادگیری ماشین، تغییرات آتی سود را پیش‌بینی کرد؟ آنچه مقاله حاضر را از پژوهش‌های مرتبط در این حوزه متمایز می‌کند، توجه به درصد تغییرات عناصر صورت‌های مالی اساسی در پیش‌بینی هر یک از عناصر سود (ناخالص، عملیاتی و خالص) است که تا کنون با بررسی صورت گرفته توسط نگارندگان این پژوهش در کشور مورد توجه قرار نگرفته است. در ادامه مقاله، به بررسی مبانی نظری و پیشینه پژوهش پرداخته شده است. در بخش بعد با معرفی روش‌شناسی پژوهش، به تحلیل یافته‌ها اشاره می‌شود. در نهایت نتیجه‌گیری و پیشنهاد‌های مرتبط با پژوهش مطرح خواهد شد.

پیشینه نظری پژوهش

یکی از کاربردهای اصلی سود، استفاده از آن به عنوان ابزاری برای پیش‌بینی است که افراد را در پیش‌بینی انواع سود و رویدادهای اقتصادی آینده یاری می‌کند که از جمله آن می‌توان به بررسی تأثیر ابعاد حساسیت سود بر پاداش هیئت مدیره شرکت اشاره کرد (واعظ، منتظر حجت و بنایی قدیم، ۱۳۹۶). به طور کلی هدف از پیش‌بینی، برنامه‌ریزی مناسب است. برنامه‌ریزی به معنای تدوین و طراحی سیاست‌ها، الگوها، طرح‌ها و ایده‌ها برای آینده، جهت تأمین اهداف سازمانی یا اهداف سیستم است. بنابراین می‌توان گفت که برنامه‌ریزی نوعی پیش‌بینی است. از این رو اطلاعات حسابداری برای اینکه بتواند در راستای تسهیل تصمیم‌گیری به سرمایه‌گذاران کمک کند، باید توان پیش‌بینی داشته باشد. برخی از پژوهشگران (از جمله شیپر و وینسنت^۳، ۲۰۰۳؛ فرانسیس، لافوند، اولسون و شیپر^۴، ۲۰۰۵) قابلیت پیش‌بینی را معیار جداگانه‌ای برای کیفیت سود در نظر گرفته‌اند. سیستم‌های پشتیبانی پیش‌بینی سود، به عنوان یکی از نیازهای اساسی مدیریت به منظور پشتیبانی از فرایند تصمیم‌گیری در نظر گرفته شده است (رندر و استیر^۵، ۲۰۱۶). امروزه هوش مصنوعی با بهره‌گیری از تفکر انسانی و قدرت یادگیری ماشین، به واکاوی داده‌ها و مقایسه آن‌ها با داده‌های قبل پرداخته و به پیش‌بینی آینده می‌پردازد که می‌توان به استفاده آن در پیش‌بینی نقدشوندگی سهام (مرادی، بحری ثالث، جبارزاده کنگرلویی و آشتاب، ۱۴۰۱) بهینه‌سازی پرتفوی (نوراحمدی و صادقی، ۱۴۰۱) و پیش‌بینی روند قیمت سهام (حیدری و

1. Petropoulos, Siakoulis, Stavroulakis & Vlachogiannakis

2. Dastile, Celik & Potsane

3. Schipper & Vincent

4. Francis, LaFond, Olsson & Schipper

5. Render & Stair

امیری، ۱۴۰۱) اشاره کرد. یادگیری ماشینی، ابزار قدرتمندی را برای پیش‌بینی فراهم می‌کند. این رویکردها به تضاد نیازی ندارند. تئوری می‌تواند راهنمایی کند که چه متغیرهایی در یک آزمایش دست‌کاری شوند؛ اما در تجزیه و تحلیل نتایج، یادگیری ماشین می‌تواند به مدیریت نتایج و تخمین‌های متعدد کمک کند. در نهایت، ابزارهای یادگیری ماشین ممکن است دامنه کار را نه فقط با ارائه داده‌های جدید یا روش‌های جدید، بلکه با تمرکز روی سؤال‌های جدید افزایش دهند (مولیناتان و اسپیس^۱، ۲۰۱۷). یادگیری ماشین که به عنوان زیرمجموعه‌ای از علوم کامپیوتر تعریف شده است، یادگیری را به طور خودکار شناسایی و پیش‌بینی الگوها را امکان‌پذیر می‌کند و به تدریج نحوه ارزیابی کسب‌وکارها و تصمیم‌گیری رهبران کسب‌وکار را تغییر می‌دهد. به همین دلیل، استفاده از یادگیری ماشین، برای ارزیابی رفتار انسان و اقتصاد در حال افزایش است. یادگیری ماشین اصولاً به این معناست که سیستم‌های خودکار می‌دانند چگونه الگوها و روابط را در داده‌ها بشناسند. مدل‌های یادگیری ماشین، رویکردی انعطاف‌پذیر و مبتنی بر داده‌ها را برای پیش‌بینی مالی ارائه می‌دهد. این مدل‌ها، از جمله شبکه‌های عصبی مصنوعی، درختان تصمیم و ماشین‌های بردار پشتیبان، برای یادگیری خودکار الگوها و روابط از داده‌ها، از مدل‌های محاسباتی پیشرفته بدون تکیه بر مفروضات صریح استفاده می‌کنند (راشکا و میرجلیلی^۲، ۲۰۱۹). مدل‌های یادگیری ماشین قابلیت‌های پیش‌بینی استثنایی را با گرفتن روابط غیرخطی پیچیده و انطباق با دینامیک داده‌های در حال تغییر نشان داده‌اند (بنزیو، کوروئل و وینسنت^۳، ۲۰۱۳). مدل‌های یادگیری ماشین، برخلاف مدل‌های آماری، دقت پیش‌بینی را بر تفسیرپذیری اولویت می‌دهند و توضیح چگونگی و چرایی پیش‌بینی‌های خاص را دشوار می‌سازد (کاروانا و همکاران^۴، ۲۰۱۵). این عدم تفسیرپذیری می‌تواند نگرانی‌های قانونی و اخلاقی ایجاد کند، به‌ویژه در حوزه‌هایی مانند امور مالی که شفافیت و پاسخ‌گویی بسیار مهم است (رودین^۵، ۲۰۱۹). علاوه بر این، مدل‌های یادگیری ماشینی از نظر محاسباتی فشرده هستند و به منابع محاسباتی شایان توجهی نیاز دارند که آن‌ها را برای برنامه‌های پیش‌بینی بی‌درنگ یا محیط‌هایی با قابلیت‌های محاسباتی محدود مناسب می‌سازد (هوانگ، وانگ و لان^۶، ۲۰۱۱). از طرفی مدل‌های آماری از دیرباز به عنوان سنگ بنای تحلیل و پیش‌بینی داده‌ها عمل کرده است. این مدل‌ها مبتنی بر نظریه‌های آماری هستند و از مدل‌های تثبیت شده، مانند تحلیل رگرسیون و تحلیل سری‌های زمانی، برای ثبت روابط بین متغیرها و ایجاد پیش‌بینی استفاده می‌کنند (چتفیلد و زینگ^۷، ۲۰۱۹). مدل‌های آماری چندین مزیت دارند؛ از جمله اینکه تفسیرپذیری، استنتاج و توانایی ترکیب دانش و فرضیه‌های قبلی در خصوص داده‌ها را در اختیار استفاده‌کنندگان قرار می‌دهند (باربوزا، کیمورا و آلتمن^۸، ۲۰۱۷). تفسیرپذیری مدل‌های آماری به متخصصان اجازه می‌دهد تا تأثیر متغیرها را درک و اهمیت ضرایب آن‌ها را ارزیابی کنند، تصمیم‌گیری را تسهیل می‌کند

1. Mullainathan & Spiess
2. Raschka & Mirjalili
3. Bengio, Courville & Vincent
4. Caruana et al
5. Rudin
6. Huang, Wang & Lan
7. Chatfield & Xing
8. Barboza, Kimura & Altman

و بینش‌های ارزشمندی را درباره پویایی‌های اساسی داده‌های مالی ارائه می‌دهند (مونتگومری، جنینگز و کولاچی^۱، ۲۰۱۹). با این حال، مدل‌های آماری محدودیت‌های خاصی دارند که باید در نظر گرفته شوند. این مدل‌ها معمولاً خطی بودن روابط بین متغیرها را فرض می‌کنند و ممکن است برای گرفتن روابط غیرخطی پیچیده تلاش کنند (بیشاپ و نصرآبادی^۲، ۲۰۰۶). آن‌ها بر مفروضاتی مانند استقلال، نرمال بودن و همسانی واریانس تکیه می‌کنند که ممکن است در مجموعه داده‌های مالی دنیای واقعی وجود نداشته باشند (میلز^۳، ۲۰۱۹). در نتیجه، مدل‌های آماری ممکن است در صورت مواجهه با توزیع داده‌های غیرخطی یا غیرنرمال، پیش‌بینی‌های کمتر از حد بهینه را نشان دهند. علاوه بر این، مدل‌های آماری ممکن است با مجموعه داده‌های با ابعاد بالا که حاوی متغیرهای متعددی است، مشکل داشته باشند؛ زیرا اغلب به انتخاب دقیق متغیر و ساده‌سازی مدل برای جلوگیری از برازش بیش از حد نیاز دارند (گلمن، هیل و وهتری^۴، ۲۰۲۰). باید دقت کرد که هر دو مدل (یادگیری ماشین و آماری) مزیت‌ها و محدودیت‌های متمایزی دارند؛ مدل‌های آماری پایه محکمی مبتنی بر نظریه آماری و امکان تفسیر و استنتاج را فراهم می‌کنند. آن‌ها اغلب به فرضیه‌های صریح در خصوص داده‌ها نیاز دارند و برای تخمین پارامتر، به مفروضات تثبیت‌شده تکیه می‌کنند (مولیناتان و اسپیس، ۲۰۱۷). در مقابل، مدل‌های یادگیری ماشین با اجازه دادن به داده‌ها برای هدایت فرایند تصمیم‌گیری مدل، پتانسیل افزایش دقت و انعطاف‌پذیری را ارائه می‌کنند. با این حال، این انعطاف‌پذیری گاهی می‌تواند به قیمت تفسیرپذیری تمام شود؛ زیرا مدل‌های یادگیری ماشین به دلیل معماری پیچیده و عملکرد داخلی پیچیده‌ای که دارند، اغلب به عنوان مدل‌های جعبه سیاه در نظر گرفته می‌شوند (آناند، برونر، ایکگوو و سوگیانیس^۵، ۲۰۱۹). روش‌های یادگیری ماشین در مقایسه با رگرسیون سه مزیت دارند: نخست اینکه این روش‌ها می‌توانند فهرست بسیار گسترده‌تری از پیش‌بینی‌کننده‌ها را برای استفاده دقیق‌تر از اطلاعات موجود در داده‌های مالی در خود جای دهند. برای مثال، می‌توان مدل‌های یادگیری ماشین را زمانی که تعداد پیش‌بینی‌ها حتی بیشتر از تعداد مشاهدات باشد استفاده کرد؛ درحالی که رگرسیون‌های سنتی برای چنین سناریویی از بین می‌روند. دوم اینکه الگوریتم‌های یادگیری ماشین شبکه گسترده‌ای را در جست‌وجوی مشخصات خود ایجاد می‌کنند تا امکان ارتباط پیچیده بین پیش‌بینی‌کننده‌های با ابعاد بالا و متغیر پیش‌بینی‌شده را فراهم کنند. سوم اینکه این الگوریتم‌ها به جای کارهای توضیحی، بیشتر برای امور پیش‌بینی به کار می‌روند. این مدل‌ها با استفاده از قاعده‌سازی^۶ برای کاهش بیش از حد برازش، پیش‌بینی نمونه با ابعاد بالایی را ارائه می‌دهند. در ادامه، تعدادی از مدل‌های پژوهش با دقت پیش‌بینی بیشتر در جهت تغییرات اجزای سود، شرح داده شده است.

1. Montgomery, Jennings & Kulahci
2. Bishop & Nasrabadi
3. Mills
4. Gelman, Hill & Vehtari
5. Anand, Brunner, Ikegwu & Sougiannis
6. Regularization

الگوریتم درخت تصمیم^۱

درخت تصمیم از روش‌های یادگیری ماشین تحت نظارت هستند که در طبقه‌بندی و رگرسیون استفاده شده‌اند. هدف از این روش، ارائه مدلی است که با یادگیری قوانین تصمیم‌گیری ساده از ویژگی‌های داده‌ها، مقدار متغیر هدف را پیش‌بینی کنند (اورسو، فیاناکا، لاروزا، راوی و ریزو^۲، ۲۰۱۸). نتیجه اجرای الگوریتم درخت تصمیم، مجموعه‌ای از شرایط منطقی به صورت «اگر..... آنگاه....» است که یک ساختار درختی، برای پیش‌بینی یک ویژگی است؛ به طوری که داده‌های قرار داده شده در برگ‌های نهایی این درخت تصمیم با یکی از مقادیر مشخصه هدف برچسب‌گذاری می‌شوند. این مدل به دلیل سهولت در تفسیرپذیری نتایج، ناپارامتری و غیرخطی بودن آن، به پیش‌فرض وجود رابطه خطی بین متغیرهای مستقل و وابسته نیازی ندارد. این الگوریتم‌ها به گونه‌ای عمل می‌کنند که سعی دارند گوناگونی یا تنوع ویژگی‌های هدف را در گره‌ها به حداقل ممکن برسانند (باقری، عزیزاده مجد، مهربخش و زیارتیان، ۱۳۹۳).

مدل ماشین بردار پشتیبان تکامل یافته^۳

ماشین بردار پشتیبان تکامل یافته، از یک راهبرد تکاملی برای بهینه‌سازی سود می‌برد. این مدل یک پیاده‌سازی ماشین بردار پشتیبان است که برای حل مسئله بهینه‌سازی دوگانه یک ماشین بردار پشتیبانی، از یک الگوریتم تکاملی استفاده می‌کند. به نظر می‌رسد این پیاده‌سازی ساده به اندازه پیاده‌سازی‌های ماشین بردار پشتیبان مرسوم در بسیاری از مجموعه داده‌ها سریع و با دقت است. علاوه بر این، می‌تواند به یادگیری با هسته‌هایی که نیمه قطعی مثبت نیستند و همچنین یادگیری چند هدفه بپردازد (میرسوا^۴، ۲۰۰۶).

مدل شبکه عصبی (پرسترون چند لایه)^۵

پرسترون یک نوع از شبکه عصبی مصنوعی است که در سال ۱۹۵۷ توسط فرانک روزنبلات ابداع شد که می‌توان آن را ساده‌ترین نوع شبکه عصبی پیشخور دانست. پرسترون یک لایه، یک ورودی مناطق تصمیم‌گیری را به شکل نیم‌صفحه تولید می‌کند. با افزودن یک لایه دیگر، هر نورون به عنوان یک پرسترون استاندارد برای خروجی‌های نورون‌های لایه قبلی عمل می‌کند، بنابراین خروجی شبکه می‌تواند مناطق تصمیم‌گیری محدب حاصل از تقاطع نیم سطوح تولید شده توسط نورون‌ها را حدس بزند. بنابراین یک پرسترون سه لایه، می‌تواند مناطق تصمیم دل‌خواه را تولید کند (پوپسکو، بالاس، پرسکو پوپسکو و ماستوراکیس^۶، ۲۰۰۹). پرسترون چند لایه، یک الگوریتم ساده برای سرعت یادگیری و تنظیم اندازه شبکه‌های عصبی در طول آموزش است. این الگوریتم ایده‌هایی از الگوریتم‌های ژنتیک و بهینه‌سازی تصادفی را ترکیب می‌کند که مجموعه کوچکی از شبکه‌ها را حفظ می‌کند که به موازات نرخ‌های مختلف و تعداد واحدهای پنهان

1. Decision tree
2. Urso, Fiannaca, La Rosa, Ravi & Rizzo
3. Support Vector Machine (Evolutionary)
4. Mierswa
5. Neural net (Multilayer perceptron).
6. Popescu, Balas, Perescu-Popescu & Mastorakis

متفاوت آموزش داده می‌شوند. پس از تعداد اندک و ثابت دوره‌ها، میزان خطا بر روی یک مجموعه اعتبارسنجی تعیین می‌شود و بدترین عملکردها با کپی‌هایی از بهترین شبکه‌ها جایگزین می‌شوند که تعداد واحدهای پنهان و نرخ یادگیری متفاوتی داشته باشند. اعداد واحد پنهان و نرخ یادگیری بر اساس توزیع احتمال از میزان موفقیت و اندازه ترسیم می‌شوند (بروئل و شفات^۱، ۲۰۱۰).

مدل استنتاج قوانین^۲

مدل استنتاج قوانین روشی برای تولید مجموعه‌ای از قوانین است که موارد را دسته‌بندی می‌کند. اگرچه درخت‌های تصمیم قادر به تولید مجموعه‌ای از قوانین هستند. روش‌های استنتاج قوانین، علاوه بر تولید یک درخت، مجموعه‌ای از قوانین مستقل را نیز ایجاد می‌کنند. استنتاج‌کننده برای تقسیم در هر سطح، قوانین اجباری ندارد و می‌تواند به آینده نگاه کند؛ از این رو می‌تواند الگوهای متفاوت و گاه بهتری برای طبقه‌بندی بیابد. برخلاف درختان تصمیم، قوانین تولید شده ممکن است همه موارد را پوشش ندهند. آن‌ها همچنین ممکن است در مفاد متناقض باشند، در این صورت باید قانونی را انتخاب کرد که از آن پیروی کند. یکی از راه‌های حل این تعارض‌ها، تعیین سطح اطمینان برای هر قانون و استفاده از قانونی است که از سطح اطمینان بالاتری برخوردار است (فایاد، پیاتتسکی و اسمیت^۳، ۱۹۹۶).

مدل رگرسیون لجستیک^۴

رگرسیون لجستیک گاهی به عنوان یک مورد خاص از الگوریتم خطی عمومی و رگرسیون خطی دیده می‌شود. مدل رگرسیون لجستیک بر مبنای فرضیه‌های بسیار متفاوت نسبت به رگرسیون خطی استوار است. تفاوت مهم این دو مدل را می‌توان در دو ویژگی رگرسیون لجستیک مشاهده کرد. اول، توزیع شرطی $y|x$ یک توزیع برنولی به جای توزیع گاوسی است؛ زیرا یک متغیر وابسته باینری است. دوم، مقادیر پیش‌بینی احتمالات هستند و بین صفر و یک محدود می‌شوند و با کمک تابع توزیع لجستیک به دست می‌آیند. رگرسیون لجستیک احتمال خروجی را پیش‌بینی می‌کند (گیتمن^۵، ۱۹۹۸).

پیشینه تجربی پژوهش

در ادامه به بررسی پیشینه تجربی پژوهش‌های خارج و داخل کشور پرداخته شده است. جونز، موزر و ویلند^۶ (۲۰۲۳) از روش‌های یادگیری ماشین برای پیش‌بینی تغییر دوره بعدی در سودآوری بر اساس مدل پیشنهادی پنمن و ژانگ استفاده کردند و متوجه شدند که روش‌های یادگیری ماشین جدید، داده‌های خارج از نمونه را به‌طور قابل توجهی بهتر از روش‌های رگرسیون سنتی پیش‌بینی می‌کنند و تفاسیر غنی‌تری درباره نقش و تأثیر متغیرهای پیش‌بینی‌کننده مختلف از

1. Breuel & Shafait
2. Rule Induction
3. Fayyad, Piatetsky & Smyth
4. Logistic Regression
5. Gitman
6. Jones, Moser & Wieland

طریق روابط غیرخطی و اثرهای متقابل آنها ارائه می‌کنند. نتایج این پژوهش در تضاد با پژوهش‌های قبلی است که نشان داده‌اند هر دو مؤلفه تجزیه دوپونت (تغییر حاشیه سود و تغییر در گردش دارایی) از تغییرات دوره بعدی در سودآوری اطلاع‌رسانی می‌کنند.

چن، چو، دو و لو^۱ (۲۰۲۲) به تجزیه و تحلیل بنیادی داده‌های مالی با رویکرد یادگیری ماشین پرداختند که یک تحلیل بنیادی از مجموعه بزرگی از اطلاعات مالی دقیق با هدف پیش‌بینی سود است. آنها در این پژوهش با استفاده از مدل‌های یادگیری ماشین (جنگل‌های تصادفی و افزایش گرادین تصادفی)، داده‌های مالی دقیق را در یک معیار خلاصه برای جهت تغییرات سود یک ساله آینده ترکیب کرده‌اند. این اندازه‌گیری، قدرت پیش‌بینی چشمگیری از نمونه را در مورد جهت تغییرات سود نشان داد.

آناند و همکاران (۲۰۱۹) سودآوری را با استفاده از یادگیری ماشینی پیش‌بینی کردند. در این پژوهش از جنگل‌های تصادفی با درخت‌های طبقه‌بندی استفاده کردند که روشی از یادگیری ماشین، برای ایجاد پیش‌بینی‌های نمونه تغییرات جهت (افزایش یا کاهش) در پنج معیار سودآوری، بازده حقوق صاحبان سهام (ROE)، بازده دارایی (ROA)، بازده خالص دارایی‌های عملیاتی (RNOA)، جریان نقدی حاصل از عملیات (CFO) و جریان نقدی آزاد (FCF) است. این پژوهش عملکرد بهتری را در معیارهای جریان نقدی نسبت به معیارهای سنتی سودآوری مبتنی بر سود نشان داد. در نهایت به این نتیجه رسیدند که روش‌های یادگیری ماشین عملکرد پیش‌بینی بهتری در مقایسه با روش‌های مبتنی بر رگرسیون سنتی ارائه می‌دهند.

توکلی و آشتاب (۱۴۰۲) کارایی مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی ریسک مالی شرکت‌ها را بررسی کردند. آنها کارایی مدل‌های یادگیری ماشین (۳۲ مدل) و مدل‌های آماری (۱۴ مدل) را برای پیش‌بینی ریسک مالی ۱۴۵ شرکت بورسی در بازه زمانی ۱۳۸۹ تا ۱۳۹۸ مقایسه کردند و به انتخاب بهترین مدل با استفاده از تکنیک‌های بهینه‌سازی پیشرفته پرداختند. یافته‌های پژوهش بیانگر آن بود که با اطمینان ۹۹ درصد، دقت پیش‌بینی مدل‌های یادگیری ماشین در پیش‌بینی ریسک مالی، بیشتر از مدل‌های آماری است. همچنین بهترین مدل یادگیری ماشین پس از بهینه‌سازی، مدل ماشین بردار پشتیبان تکاملی با دقت پیش‌بینی ۹۹/۸۶ درصد و مقدار سطح زیر منحنی برابر ۹۹/۸ درصد بوده است.

میرزایی، آشتاب و زواری رضایی (۱۴۰۲) در پژوهشی به مقایسه کارایی مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی سود خالص و جریان‌های نقدی عملیاتی پرداختند. نتایج پژوهش حاکی از برتری مدل‌های یادگیری ماشین نسبت به مدل‌های آماری در پیش‌بینی سود خالص و جریان‌های نقدی عملیاتی بود.

اسدی، میربرگ‌کار و چیرانی (۱۴۰۱) به ارائه مدل شبکه عصبی برای پیش‌بینی سود و مقایسه دقت آن با مدل‌های ARIMA و HDZ پرداختند. به‌منظور گردآوری داده‌ها از صورت‌های مالی اساسی شرکت‌ها در بازه زمانی ۱۳۹۸ تا ۱۳۹۳ استفاده شده است. نتایج حاکی از آن است که شبکه عصبی طراحی شده، توانایی پیش‌بینی روند قیمت سهام با

استفاده از شاخص‌های کل و صنعت را دارد و این امر علاوه بر تأیید دیگری بر توانایی شبکه عصبی برای پیش‌بینی حوزه‌های مالی، سودآوری استراتژی پیش‌بینی قیمت در بورس تهران را نیز تأیید می‌کند.

قادری، امینی و محمدی ملقرنی (۱۳۹۹) از مدل ترکیبی شبکه‌های عصبی مصنوعی با الگوریتم‌های فرا اکتشافی (PSO و ICA) برای پیش‌بینی مدیریت سود استفاده کردند. در مرحله اول، مدل خطی اولیه با استفاده از مدل شبکه عصبی بهینه شد؛ سپس از الگوریتم‌های ازدحام ذرات و رقابت استعماری برای بهینه‌سازی مدل استفاده شد. بنابراین، یافته‌های تجربی مربوط به بررسی ۶۲۰ مشاهدات (سال - شرکت) در بورس اوراق بهادار تهران بین سال‌های ۱۳۹۰ تا ۱۳۹۵، حاکی از سودمندی و تأثیر مستقیم روش‌های ترکیبی بر پیش‌بینی مدیریت سود بود. همچنین وجود تفاوت معنادار بین میزان سودمندی روش‌های خطی و غیرخطی را نشان داد. به عبارت دیگر، در صورت استفاده از مدل‌ها برای پیش‌بینی مدیریت سود، با حذف متغیرهای ناکارا، دقت پیش‌بینی افزایش می‌یابد. یافته‌های پژوهش نشان می‌دهد که الگوریتم رقابت استعماری نسبت به سایر مدل‌ها، از نظر کارایی بهتر بوده و مناسب‌تر عمل می‌کند.

روش‌شناسی پژوهش

جامعه آماری پژوهش کلیه شرکت‌های تولیدی پذیرفته شده در بورس اوراق بهادار است. برای انتخاب نمونه، از روش حذف سیستماتیک استفاده شده و شرکت‌هایی مد نظر قرار گرفته‌اند که قبل از سال ۱۳۸۷ وارد بورس اوراق بهادار شده باشند؛ این شرکت‌ها شامل بانک‌ها، شرکت‌های بیمه، شرکت‌های سرمایه‌گذاری یا واسطه‌گری مالی نباشند؛ اطلاعات صورت‌های مالی آن‌ها مانند صورت وضعیت مالی، صورت سود و زیان و صورت جریان‌های نقدی تا سال ۱۴۰۱ در دسترس باشند و طی دوره زمانی پژوهش تغییر سال مالی نداده باشند. بر این اساس، نمونه آماری تحقیق متشکل از ۱۳۹ شرکت پذیرفته شده در بورس اوراق بهادار تهران است. در این پژوهش برای تجزیه و تحلیل و نتیجه‌گیری در خصوص فرضیه‌ها، از نرم‌افزار اکسل برای مرتب‌سازی و دسته‌بندی داده‌ها، از نرم‌افزار ایویوز برای استخراج آمار توصیفی و از نرم‌افزارهای داده‌کاوی اس‌پی‌اس‌اس مُدلر^۱ نسخه ۱۸ و ریپیدماینر^۲ نسخه ۹ برای مدل‌سازی، استفاده شده است.

فرضیه‌های پژوهش

با توجه به قدرت پیش‌بینی عالی الگوریتم‌های یادگیری ماشینی و عاری بودن آن‌ها از محدودیت‌های موجود در مدل‌های آماری (از جمله مدل‌های غیرخطی، حساس نبودن به مقادیر پرت و نبود مشکل همخطی متغیرها) انتظار می‌رود که برای پیش‌بینی اجزای سود، مدل‌های یادگیری ماشین نسبت به مدل‌های آماری، دقت بیشتری از خود نشان دهند. بنابراین فرضیه‌های پژوهش به شرح زیر مطرح شده است:

فرضیه اول. برای پیش‌بینی جهت تغییرات سود (زیان) خالص، دقت پیش‌بینی مدل‌های یادگیری ماشین نسبت به مدل‌های آماری، به‌طور معناداری بیشتر است.

فرضیه دوم. برای پیش‌بینی جهت تغییرات سود (زیان) ناخالص، دقت پیش‌بینی مدل‌های یادگیری ماشین نسبت به مدل‌های آماری، به‌طور معناداری بیشتر است.

فرضیه سوم. برای پیش‌بینی جهت تغییرات سود (زیان) عملیاتی، دقت پیش‌بینی مدل‌های یادگیری ماشین نسبت به مدل‌های آماری، به‌طور معناداری بیشتر است.

متغیرهای پژوهش

متغیرهای پژوهش ۲۸ متغیر توضیحی و سه متغیر وابسته هستند که از صورت وضعیت مالی، صورت سود و زیان و صورت جریان‌های نقدی استخراج شده‌اند. این متغیرها به تفکیک در جدول ۱ شرح داده شده‌اند (چن و همکاران، ۲۰۲۲).

جدول ۱. متغیرهای پژوهش

منبع	شرح متغیر	نماد
متغیرهای مستقل پژوهش		
صورت وضعیت مالی	مجموع دارایی‌های سال قبل / (مجموع دارایی‌های سال قبل - مجموع دارایی‌های سال جاری) = درصد تغییرات مجموع دارایی‌ها	% Δ ASST
	دارایی‌های جاری سال قبل / (دارایی‌های جاری سال قبل - دارایی‌های جاری سال جاری) = درصد تغییرات دارایی‌های جاری	% Δ CRASST
	دارایی‌های غیرجاری سال قبل / (دارایی‌های غیرجاری سال قبل - دارایی‌های غیرجاری سال جاری) = درصد تغییرات دارایی‌های غیرجاری	% Δ NCRASST
	مجموع بدهی‌های سال قبل / (مجموع بدهی‌های سال قبل - مجموع بدهی‌های سال جاری) = درصد تغییرات مجموع بدهی‌ها	% Δ DEBT
	بدهی‌های جاری سال قبل / (بدهی‌های جاری سال قبل - بدهی‌های جاری سال جاری) = درصد تغییرات بدهی‌های جاری	% Δ CRDEBT
	بدهی‌های غیرجاری سال قبل / (بدهی‌های غیرجاری سال قبل - بدهی‌های غیرجاری سال جاری) = درصد تغییرات بدهی‌های غیرجاری	% Δ LNDEBT
	= درصد تغییرات در سفارش‌ها و پیش‌پرداخت‌ها سفارش‌ها و پیش‌پرداخت‌های سال قبل / (سفارش‌ها و پیش‌پرداخت‌های سال قبل - سفارش‌ها و پیش‌پرداخت‌های سال جاری)	% Δ PRPAY
	اندوخته‌های سال قبل / (اندوخته‌های سال قبل - اندوخته‌های سال جاری) = درصد تغییرات در اندوخته‌ها	% Δ RESRV
	موجودی نقد سال قبل / (موجودی نقد سال قبل - موجودی نقد سال جاری) = درصد تغییرات موجودی نقد	% Δ CASH
	موجودی‌های سال قبل / (موجودی‌های سال قبل - موجودی‌های سال جاری) = درصد تغییرات موجودی‌ها	% Δ INVT

نماد	شرح متغیر	منبع
% Δ RECBL	حساب‌های دریافتنی سال قبل / (حساب‌های دریافتنی سال قبل - حساب‌های دریافتنی سال جاری) = درصد تغییرات حساب‌های دریافتنی	
% Δ PPE	= درصد تغییرات اموال، ماشین‌آلات و تجهیزات اموال، ماشین‌آلات و تجهیزات سال قبل / (اموال، ماشین‌آلات و تجهیزات سال قبل - اموال، ماشین‌آلات و تجهیزات سال جاری)	
% Δ EQUITY	حقوق مالکانه سال قبل / (حقوق مالکانه سال قبل - حقوق مالکانه سال جاری) = درصد تغییرات حقوق مالکانه	
% Δ TAX	مالیات بر درآمد سال قبل / (مالیات بر درآمد سال قبل - مالیات بر درآمد سال جاری) = درصد تغییرات مالیات بر درآمد	صورت سود و زیان
% Δ EXPNS	= درصد تغییرات در هزینه‌های شرکت (هزینه‌های توزیع و فروش، اداری و عمومی) هزینه‌های شرکت سال قبل / (هزینه‌های شرکت سال قبل - هزینه‌های شرکت سال جاری)	
% Δ NOPROF	سود (زیان) غیرعملیاتی سال قبل / (سود (زیان) غیرعملیاتی سال قبل - سود (زیان) غیرعملیاتی جاری) = درصد تغییرات سود (زیان) غیرعملیاتی	
% Δ INTRST	هزینه بهره سال قبل / (هزینه بهره سال قبل - هزینه بهره سال جاری) = درصد تغییرات هزینه بهره	
% Δ EPS	سود (زیان) پایه هر سهم سال قبل / (سود (زیان) پایه هر سهم سال قبل - سود (زیان) پایه هر سهم جاری) = درصد تغییرات سود (زیان) پایه هر سهم	
% Δ DPS	سود نقدی هر سهم سال قبل / (سود نقدی هر سهم سال قبل - سود نقدی هر سهم سال جاری) = درصد تغییرات سود نقدی هر سهم	
% Δ SALE	فروش سال قبل / (فروش سال قبل - فروش سال جاری) = درصد تغییرات در فروش	
% Δ FINCF	= درصد تغییرات جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی (جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی سال قبل - جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی سال جاری) / جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی سال قبل	صورت جریان‌های نقدی
% Δ OPERCF	= درصد تغییرات جریان‌های نقدی ورودی حاصل از فعالیت‌های عملیاتی (جریان‌های نقدی ورودی حاصل از فعالیت‌های عملیاتی سال قبل - جریان‌های نقدی ورودی حاصل از فعالیت‌های عملیاتی سال جاری) / جریان خالص ورود نقد حاصل از فعالیت‌های عملیاتی سال قبل	
% Δ CAPEX	= درصد تغییرات پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود (پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود سال قبل - پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود سال جاری) / پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود سال قبل	

منبع	شرح متغیر	نماد
	= درصد تغییرات جریان نقدینگی فروش دارایی‌ها (فروش نقدی دارایی‌ها) جریان نقدینگی فروش دارایی‌های سال قبل / (جریان نقدینگی فروش دارایی‌های سال قبل - جریان نقدینگی فروش دارایی‌های سال جاری)	$\Delta \text{ASTCSH} \%$
	= درصد تغییرات پرداخت‌های نقد ناشی از فعالیت‌های سرمایه‌گذار (پرداخت‌های نقد ناشی از فعالیت‌های سرمایه‌گذاری سال قبل - پرداخت‌های نقد ناشی از فعالیت‌های سرمایه‌گذاری سال جاری) پرداخت‌های نقد ناشی از فعالیت‌های سرمایه‌گذاری سال قبل /	$\Delta \text{INVCF} \%$

متغیرهای وابسته پژوهش

صورت سود و زیان	سود (زیان) خالص سال قبل / (سود (زیان) خالص سال قبل - سود (زیان) خالص سال جاری) = درصد تغییرات سود (زیان) خالص	$\Delta \text{PROF} \%$
	سود (زیان) ناخالص سال قبل / (سود (زیان) ناخالص سال قبل - سود (زیان) ناخالص سال جاری) = درصد تغییرات سود (زیان) ناخالص	$\Delta \text{GPROF} \%$
	سود (زیان) عملیاتی سال قبل / (سود (زیان) عملیاتی سال قبل - سود (زیان) عملیاتی سال جاری) = درصد تغییرات سود (زیان) عملیاتی	$\Delta \text{OPROF} \%$

مدل‌های پژوهش

۲۵ مدل یادگیری ماشین و ۱۰ مدل آماری که در این پژوهش استفاده شده است، در جدول ۲ مشاهده می‌شود (چن و همکاران، ۲۰۲۲).

جدول ۲. مدل‌های پژوهش

نوع مدل	مدل‌ها
مدل‌های یادگیری ماشین	Decision Tree
	Decision Stump
	Random Tree
	Random Forest
	CHAID
	Gradient Boosted Trees
	CART
	K-NN
	Support Vector Machine
	Support Vector Machine (Lib SVM)
	Support Vector Machine (Evolutionary)
	Support Vector Machine (PSO)
	SVM (Hyper Hyper)
	Neural-net (Deep learning)
	Neural net (one-layer perceptron)

مدل‌ها	نوع مدل
Neural net (Multilayer perceptron)	
Rule induction	
Single Rule induction (single attribute)	
Naive Bayes	
Naive Bayes (Kernel)	
ID3 trees	
C5	
Quest	
Bayes structure type TAN	
Bayes structure type Markov Blanket	
Probit	مدل‌های آماری
Logit	
Extreme value	
Generalized Linear Model (GLM)	
Logistic Regression	
Logistic Regression (Evolutionary)	
Logistic Regression (Keerthi et al 2003)	
Linear Discriminant Analysis	
Regularized Discriminant Analysis	
Quadratic Discriminant Analysis	

یافته‌های پژوهش

در این قسمت یافته‌های پژوهش شامل آمار توصیفی و تحلیلی بیان شده است.

آمار توصیفی متغیرهای پژوهش

به‌منظور بررسی و تجزیه و تحلیل اولیه داده‌ها، اطلاعات مربوط به آمار توصیفی متغیرهای وابسته و مستقل این پژوهش، در جدول ۳ نشان داده شده است. شایان ذکر است که برای خنثی کردن اثرهای داده‌های پرت، کلیه متغیرها در سطح ۱ درصد و ۹۹ درصد وینسورایز شده‌اند. نتایج حاصل از آمار توصیفی متغیرهای پژوهش بیانگر این است که به‌طور مثال، متوسط درصد تغییرات پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود ۸۶/۳ درصد و درصد تغییرات موجودی نقد ۷۸ درصد است که به‌ترتیب بیانگر افزایش پرداخت‌های نقد برای خرید دارایی ثابت مشهود و افزایش موجودی‌های نقد شرکت‌های مورد بررسی در طی دوره زمانی پژوهش است. همچنین متوسط درصد تغییرات جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی ۶۰/۵- درصد و درصد تغییرات در جریان نقدینگی فروش دارایی‌ها ۲۱/۳ درصد است که به‌ترتیب بیانگر کاهش جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی و کاهش جریان‌های نقدینگی حاصل از فروش دارایی‌ها در شرکت‌های مورد بررسی در طی دوره زمانی پژوهش است.

جدول ۳. آمار توصیفی متغیرهای پژوهش

متغیرها	میانگین	میانه	پیشینه	کمینه	انحراف معیار	چولگی	کشیدگی
درصد تغییرات مجموع دارایی‌ها	۰/۲۴۵	۰/۱۶۷	۰/۹۸۱	-۰/۱۰۳	۰/۲۸۳	۱/۱۳۹	۳/۶۱۴
درصد تغییرات جریان نقدینگی فروش دارایی‌ها (فروش نقدی دارایی‌ها)	-۰/۲۱۳	۰/۰۰۰	۰/۳۹۱	-۰/۹۰۴	۰/۵۱۳	-۰/۲۷۰	۱/۴۸۰
درصد تغییرات پرداخت‌های نقد برای خرید دارایی‌های ثابت مشهود	۰/۸۶۳	۰/۱۷۶	۶/۸۳۷	-۰/۷۸۵	۱/۹۴۷	۱/۹۲۶	۵/۹۹۰
درصد تغییرات موجودی نقد	۰/۷۸۰	۰/۲۰۶	۵/۵۱۶	-۰/۷۵۱	۱/۶۲۱	۱/۶۳۴	۴/۹۵۶
درصد تغییرات دارایی‌های جاری	۰/۲۶۶	۰/۱۹۲	۱/۰۵۰	-۰/۱۷۳	۰/۳۲۰	۰/۹۲۰	۳/۲۰۱
درصد تغییرات بدهی‌های جاری	۰/۲۶۲	۰/۱۷۷	۱/۲۵۵	-۰/۲۷۲	۰/۳۸۷	۰/۹۷۲	۳/۴۵۳
درصد تغییرات مجموع بدهی‌ها	۰/۲۴۳	۰/۱۶۴	۱/۱۴۰	-۰/۲۰۷	۰/۳۴۴	۱/۰۳۴	۳/۵۷
درصد تغییرات سودنقدی هر سهم (DPS)	۰/۱۰۱	۰/۰۰۰	۱/۵۰۰	-۰/۸۱۸	۰/۶۷۸	۰/۷۱۵	۲/۷۶۹
درصد تغییرات سود (زیان) پایه هر سهم (EPS)	۰/۱۰۱	-۰/۰۰۲	۱/۷۲۸	-۰/۹۹۰	۰/۸۲۷	۰/۶۰۵	۲/۴۳۲
درصد تغییرات حقوق مالکانه	۰/۲۷۱	۰/۱۴۴	۱/۵۴۶	-۰/۳۷۳	۰/۴۶۲	۱/۲۹۵	۴/۳۳۲
درصد تغییرات در هزینه‌های شرکت (فروش، اداری و عمومی)	۰/۲۶۷	۰/۲۱۵	۱/۱۵۰	-۰/۳۰۲	۰/۳۵۳	۰/۷۶۶	۳/۳۸۹
درصد تغییرات جریان‌های نقدی حاصل از فعالیت‌های تأمین مالی	-۰/۶۰۵	-۰/۶۱۳	۲/۰۳۲	-۳/۳۳۶	۱/۵۰۲	-۰/۰۸۲	۲/۵۷۳
درصد تغییرات سود (زیان) ناخالص	۰/۲۳۱	۰/۱۴۰	۱/۳۵۲	-۰/۷۲۴	۰/۷۰۱	۰/۲۶۷	۱/۹۰۲
درصد تغییرات هزینه بهره	۰/۲۱۱	۰/۱۰۰	۱/۱۴۱	-۰/۴۱۷	۰/۴۷۲	۰/۶۵۱	۲/۴۴۵
درصد تغییرات پرداخت‌های نقد ناشی از فعالیت‌های سرمایه‌گذاری	-۰/۲۰۰	۰/۰۰۰	۰/۲۶۷	-۰/۹۴۷	۰/۴۴۸	-۰/۸۱۰	۲/۰۳۶
درصد تغییرات موجودی‌ها	۰/۲۹۱	۰/۱۹۴	۱/۲۹۷	-۰/۳۰۳	۰/۴۱۹	۰/۸۷۱	۳/۰۹۰
درصد تغییرات بدهی‌های غیرجاری	۰/۱۸۹	۰/۱۴۷	۰/۹۶۴	-۰/۳۶۰	۰/۳۸۱	۰/۵۸۱	۲/۶۶۲
درصد تغییرات دارایی‌های غیرجاری	۰/۱۸۳	۰/۱۷۳	۰/۸۲۱	-۰/۰۹۹	۰/۲۸۶	۱/۱۶۰	۳/۱۵۵
درصد تغییرات سود (زیان) غیر عملیاتی	-۰/۰۴۹	-۰/۰۱۵	۰/۲۳۶	-۰/۳۳۴	۰/۲۶۴	-۰/۰۰۱	۱/۱۳۶
درصد تغییرات سهام عادی منتشر شده و در دست سهام‌داران	۰/۱۹۱	۰/۰۰۰	۱/۰۰۰	۰/۰۰۰	۰/۳۶۲	۱/۵۶۲	۳/۶۸۹
درصد تغییرات جریان خالص ورود نقد حاصل از فعالیت‌های عملیاتی	۰/۰۷۱	-۰/۰۱۴	۲/۳۵۶	-۲/۰۲۴	۱/۲۵۳	۰/۱۹۶	۲/۴۴۴
درصد تغییرات سود (زیان) عملیاتی	۰/۲۹۸	۰/۱۷۳	۱/۶۹۶	-۰/۶۷۷	۰/۷۱۷	۰/۵۹۲	۲/۴۲۳
درصد تغییرات حساب‌های پرداختی	۰/۵۲۱	۰/۲۵۹	۲/۳۳۸	-۰/۴۵۱	۰/۸۶۰	۰/۹۶۵	۲/۸۰۴

متغیرها	میانگین	میانه	پیشینه	کمینه	انحراف معیار	چولگی	کشیدگی
درصد تغییرات اموال، ماشین‌آلات و تجهیزات	۰/۱۷۲	۰/۰۵۲	۰/۸۵۴	-۰/۱۰۳	۰/۲۹۶	۱/۲۹۸	۳/۴۷۱
درصد تغییرات سود (زیان) خالص	۰/۲۱۳	۰/۲۸	۱/۰۴۶	-۰/۵۰۷	۰/۵۷۱	۰/۲۲۷	۱/۶۵۹
درصد تغییرات در سفارش‌ها و پیش‌پرداخت‌ها	۰/۵۴۶	۰/۱۹۴	۲/۸۳۷	-۰/۵۵۹	۱/۰۵۰	۱/۰۵۰	۲/۹۷۸
درصد تغییرات حساب‌های دریافتی	۰/۳۳۴	۰/۲۰۰	۱/۴۵۵	-۰/۳۲۹	۰/۵۴۰	۰/۸۱۳	۲/۶۵۵
درصد تغییرات در اندوخته‌ها	۰/۱۷۹	۰/۰۰۰	۰/۸۰۹	۰/۰۰۰	۰/۲۷۸	۱/۳۷۵	۳/۳۷۵
درصد تغییرات سود (زیان) انباشته	۰/۳۱۹	۰/۱۸۴	۱/۷۳۵	-۰/۶۹۶	۰/۷۰۹	۰/۵۹۰	۲/۵۶۷
درصد تغییرات در فروش	۰/۲۷۹	۰/۲۱۳	۱/۰۷۹	-۰/۲۵۸	۰/۳۵۲	۰/۶۹۴	۲/۷۹۲
درصد تغییرات مالیات بر درآمد	۰/۱۴۷	۰/۰۰۰	۰/۷۲۸	-۰/۲۸۶	۰/۳۶۷	۰/۵۰۸	۱/۸۵۶

آمار تحلیلی

جدول ۴ نتایج حاصل از دقت پیش‌بینی مدل‌های آماری و مدل‌های یادگیری ماشین برای هر یک از اجزای سود از جمله سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی را به تفکیک نشان می‌دهد.

جدول ۴. دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری

مدل‌ها		درصد تغییرات سود (زیان) خالص		درصد تغییرات سود (زیان) ناخالص		درصد تغییرات سود (زیان) عملیات	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Decision Tree		۱	۱	۰/۹۹۳۸	۰/۹۹۷	۰/۸۵۵۶	۰/۹۲۵
Decision Stump		۱	۱	۰/۹۷۶۵	۰/۹۷۷	۰/۸۱۲۵	۰/۸۱۲
Random Tree		۰/۹۸۳	۰/۵	۰/۹۸۱	۰/۵	۰/۹۰۶	۰/۵
Random Forest		۱	۱	۰/۹۹۶۲	۱	۰/۹۰۱۲	۰/۹۷۳
CHAID		۰/۹۴۳۸	۰/۹۸۷	۰/۹۳۲۳	۰/۹۸۴	۰/۸۱۴۳	۰/۹۰۳
Gradient Boosted Trees		۰/۸۲۲۷	۰/۸۳۷	۰/۹۳۶۱	۰/۹۸	۰/۸۲۲۷	۰/۸۳۷
CART		۰/۹۵۳۵	۰/۹۷۴	۰/۹۳۴۳	۰/۹۵۲	۰/۸۳۳۶	۰/۸۶۲
K-NN		۰/۹۵۵۱	۰/۹۸۷	۰/۹۲۸۱	۰/۹۷۸	۰/۸۲۱۱	۰/۹۵
Support Vector Machine (SVM)		۰/۹۹۷۶	۱	۰/۹۸۸۵	۰/۹۹۸	۰/۸۵۱۸	۰/۹۳۹
Support Vector Machine (Lib SVM)		۰/۹۶۰۷	۰/۹۹۱	۰/۹۶۰۷	۰/۹۹۲	۰/۸۶۱۴	۰/۹۳۱
Support Vector Machine (Evolutionary)		۰/۸۹۹۸	۰/۹۸۱	۰/۸۹۷۸	۰/۹۷۹	۰/۷۸۴۲	۰/۸۵۸
Support Vector Machine (PSO)		۱	۱	۱	۱	۰/۹۹۹۵	۱
SVM (Hyper Hyper)		۰/۹۷۹۵	۰/۹۹۹	۰/۹۷۹۹	۰/۹۹۹	۰/۹۷۶۵	۰/۹۹۷

درصد تغییرات سود (زیان) عملیات		درصد تغییرات سود (زیان) ناخالص		درصد تغییرات سود (زیان) خالص		مدل‌ها
AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	
۰/۳۶	۰/۴۴۴۶	۰/۲۴۸	۰/۴۲۸۳	۰/۷۵۳	۰/۵۹۴۷	Neural-net (Deep learning)
۰/۸۳	۰/۷۴۵۸	۰/۹۹۵	۰/۹۶۷۴	۰/۹۹۸	۰/۹۷۸۹	Neural net (one-layer perceptron)
۰/۹۶۳	۰/۹۴۹۲	۰/۹۹۷	۰/۹۹۴۷	۱	۱	Neural-net (Multilayer perceptron)
۰/۹۲۵	۰/۸۶۷۶	۰/۹۷۶	۰/۹۷۶	۱	۱	Rule induction
۰/۹۹۷	۰/۹۷۲۲	۰/۹۹۸	۰/۹۷۷۹	۱	۱	Single Rule induction (single attribute)
۰/۸۴۶	۰/۷۷۱۲	۰/۹۷۳	۰/۸۹۷۴	۰/۹۷۵	۰/۸۹۹۳	Naive Bayes
۰/۸۸۵	۰/۸۰۳۸	۰/۹۸۹	۰/۹۴۲۹	۰/۹۹۳	۰/۹۵۳	Naive Bayes (Kernel)
۰/۷۰۳	۰/۶۳۳۱	۰/۶۹۱	۰/۶۲۰۶	۰/۷۰۶	۰/۶۳۶	ID3 trees
۰/۸۹۹	۰/۹۴۸	۰/۹۵۹	۰/۹۸۶	۰/۹۶۰	۰/۹۸۶	C5
۰/۷۳۱	۰/۷۳۱	۰/۹۴۵	۰/۹۴۵	۰/۹۴۶	۰/۹۴۶	Quest
۰/۸۲۷	۰/۹۱۲	۰/۹۳۵	۰/۹۸۳	۰/۹۳۶	۰/۹۸۶	Bayes structure type TAN
۰/۹۳۹	۰/۹۹	۰/۹۹۹	۱	۰/۹۹۸	۱	Bayes structure type Markov Blanket
-	۰/۷۹۹	-	۰/۹۲۳۳	-	۰/۹۲۶۱	Probit
-	۰/۸۰۰۵	-	۰/۹۲۶۱	-	۰/۹۳	Logit
-	۰/۷۹۴۷	-	۰/۹۲۰۴	-	۰/۹۱۹۹	Extreme value
-	۰/۸۲۱۱	-	۰/۹۷۸۴	-	۰/۹۸۹۹	Generalized Linear Model (GLM)
-	۰/۸۲۲۱	-	۰/۹۸۱۳	-	۱	Logistic Regression
-	۰/۵۰۰۲	-	۰/۴۹۹۸	-	۰/۴۹۹۸	Logistic Regression (Evolutionary)
-	۰/۸۱۲	-	۰/۹۷۵۵	-	۰/۹۸۶۶	Logistic Regression (Keerthi et al., 2003)
-	۰/۸۲۰۱	-	۰/۹۵۶۸	-	۰/۹۵۹۲	Linear Discriminant Analysis
-	۰/۷۹۴۷	-	۰/۹۶۰۷	-	۰/۹۶۴۵	Regularized Discriminant Analysis
-	۰/۷۵۴	-	۰/۹۵۵۹	-	۰/۹۵۹۳	Quadratic Discriminant Analysis

AUC و Accuracy در جدول فوق به ترتیب، نشان دهنده سطح زیر منحنی و دقت پیش‌بینی مدل‌ها هستند که از آن برای ارزیابی عملکرد مدل استفاده شده است. متمایل شدن AUC (سطح زیر منحنی) هر مدل به سمت ۱، گویای عملکرد بهتر آن مدل است. همان‌طور که از جدول ۴ مشهود است، دقت پیش‌بینی مدل‌های یادگیری ماشین از ۸۲/۹ درصد تا ۹۳/۳ درصد و دقت پیش‌بینی مدل‌های آماری نیز از ۷۶/۴ درصد تا ۸۹/۸ درصد متغیر است.

تحلیل فرضیه‌های پژوهش

حال سؤال کلیدی این است که کدام رویکرد، مدل‌های آماری یا مدل‌های یادگیری ماشینی، دقت بیشتری را برای پیش‌بینی سود (زیان) خالص و سود (زیان) ناخالص و سود (زیان) عملیاتی ارائه می‌دهد؟ در جدول ۵ به مقایسه متوسط دقت مدل‌های آماری و مدل‌های یادگیری ماشین پرداخته شده است.

جدول ۵. مقایسه متوسط دقت مدل‌های یادگیری ماشین و مدل‌های آماری

دقت	درصد تغییرات سود (زیان) خالص	درصد تغییرات سود (زیان) ناخالص	درصد تغییرات سود (زیان) عملیاتی
متوسط دقت کلیه مدل‌ها	۰/۹۲۳۶	۰/۹۱۱۳	۰/۸۱۱۰
متوسط دقت مدل‌های یادگیری ماشین	۰/۹۳۳۵	۰/۹۱۸۳	۰/۸۲۹۶
متوسط دقت مدل‌های آماری	۰/۸۹۸۵	۰/۸۹۳۳	۰/۷۶۴۸

مقایسه نتایج متوسط دقت پیش‌بینی مدل‌ها، نشان‌دهنده دقت بیشتر مدل‌های یادگیری ماشین برای پیش‌بینی جهت تغییرات سود (زیان) خالص، تغییرات سود (زیان) ناخالص و تغییرات سود (زیان) عملیاتی، نسبت به مدل‌های آماری است. برای مقایسه آماری دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری، نخست نرمال بودن یا غیر نرمال بودن متوسط دقت مدل‌های یادگیری ماشین و مدل‌های آماری برای هر یک از متغیرهای وابسته، با استفاده از آزمون کلموگروف اسمیرنوف مورد آزمون قرار گرفت. نتایج این آزمون در جدول ۶ ارائه شده است.

جدول ۶. آزمون نرمال

Smirnov–Kolmogorov	درصد تغییرات سود (زیان) خالص	درصد تغییرات سود (زیان) ناخالص	درصد تغییرات سود (زیان) عملیاتی
آماره	۰/۳۶۴	۰/۴۰۱	۰/۲۸۰
سطح معناداری	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱

با توجه به آماره کلموگروف - اسمیرنوف که برای درصد تغییرات سود (زیان) خالص، درصد تغییرات سود (زیان) ناخالص و درصد تغییرات سود (زیان) عملیاتی به ترتیب ۰/۳۶۴، ۰/۴۰۱ و ۰/۲۸۰ است و سطح معناداری مربوطه که برای هر یک ۰/۰۰۱ است، هیچ کدام از متغیرهای تغییرات سود (زیان) خالص، تغییرات سود (زیان) ناخالص و تغییرات سود (زیان) عملیاتی نرمال نیستند؛ بنابراین برای مقایسه متوسط دقت مدل‌های یادگیری ماشین و مدل‌های آماری، از آزمون ناپارامتریک یومن ویتنی استفاده شده است که نتایج آن به شرح جدول ۷ است.

جدول ۷. آزمون یومن ویتنی

آزمون یومن ویتنی	درصد تغییرات سود (زیان) خالص	درصد تغییرات سود (زیان) ناخالص	درصد تغییرات سود (زیان) عملیاتی
آماره	-۲/۲۷۸	-۲/۷۰۲	-۲/۶۶۷
سطح معناداری	۰/۰۲۳	۰/۰۰۷	۰/۰۰۸

آماره یومن ویتنی برای تغییرات سود (زیان) خالص، تغییرات سود (زیان) ناخالص، تغییرات سود (زیان) عملیاتی به ترتیب ۲/۲۷۸-، ۲/۷۰۲- و ۲/۶۶۷- است و سطح معناداری برای سه متغیر وابسته، از جمله درصد تغییرات سود (زیان) خالص، درصد تغییرات سود (زیان) ناخالص و درصد تغییرات سود (زیان) عملیاتی، به ترتیب ۰/۰۰۸، ۰/۰۰۷، ۰/۰۲۳ و کمتر از ۵ درصد است؛ از این رو دقت مدل های یادگیری ماشین برای پیش بینی جهت تغییرات اجزای سود نسبت به مدل های آماری به صورت معناداری بیشتر است و هر سه فرضیه پژوهش تأیید می شود.

انتخاب مدل بهینه

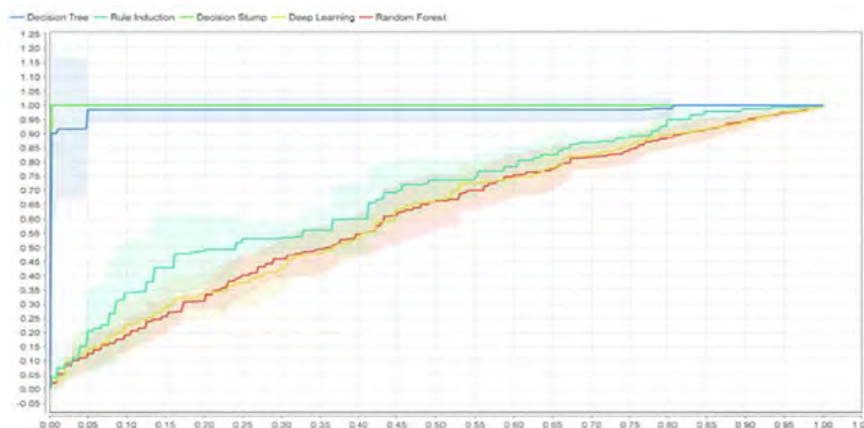
در این مرحله، به منظور انتخاب مدلی که برای پیش بینی تغییرات سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی بهترین عملکرد را دارد، از میان مدل های یادگیری ماشین، به انتخاب مدل بهینه با استفاده از منحنی راک^۱ پرداخته شده است. از میان مدل هایی یادگیری ماشین، مدل هایی با دقت پیش بینی و سطح زیرمنحنی ۱۰۰ درصد یا نزدیک به ۱۰۰ درصد، برای هر یک از اجزای سود انتخاب شد و منحنی راک آن ها با هم مقایسه و مدل بهینه انتخاب شد. مدل های انتخابی برای هر یک از اجزای سود به شرح جدول ۸ است.

جدول ۸. انتخاب مدل هایی با دقت پیش بینی بیشتر برای اجزای سود

متغیرها	مدل	Accuracy	AUC
درصد تغییرات سود (زیان) خالص	Decision Tree	۱	۱
	Decision Stump	۱	۱
	Random Forest	۱	۱
	Rule induction	۱	۱
	Neural-net (Deep learning)	۰/۹۹۷۶	۱
درصد تغییرات سود (زیان) ناخالص	Decision Tree	۰/۹۹۳۸	۰/۹۹۷
	Random Tree	۰/۹۸۱	۰/۵
	Random Forest	۰/۹۹۶۲	۱
	Rule induction	۰/۹۷۶	۰/۹۷۶
	Neural-net (Deep learning)	۰/۹۸۸۵	۰/۹۹۸
درصد تغییرات سود (زیان) عملیاتی	Random Tree	۰/۹۰۶	۰/۵
	Random Forest	۰/۹۰۱۲	۰/۹۷۳
	Neural-net (Deep learning)	۰/۸۵۱۸	۰/۹۳۹
	Rule induction	۰/۸۶۷۶	۰/۹۲۵
	single rule induction (single attribute)	۰/۹۷۲۲	۰/۹۹۷

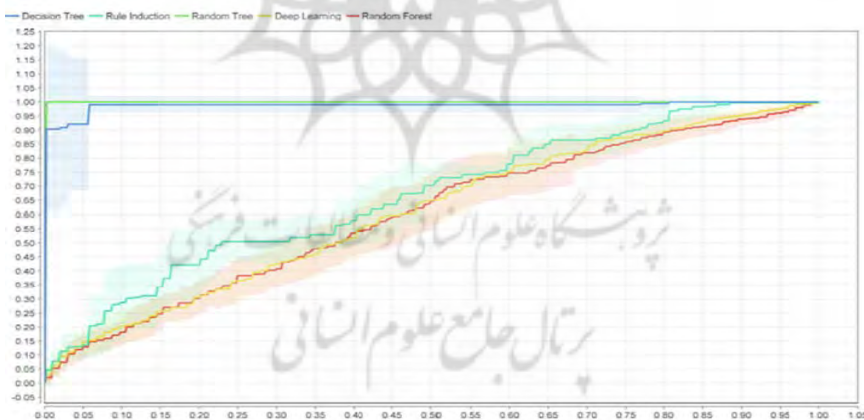
۱. منحنی مشخصه عملکرد گیرنده یا به اختصار منحنی ROC با رسم نرخ مثبت واقعی (TPR) در برابر نرخ مثبت کاذب (FPR) در تنظیمات آستانه های مختلف ایجاد می شود. نرخ مثبت واقعی به عنوان حساسیت، بازیابی یا احتمال تشخیص نیز شناخته می شود. نرخ مثبت کاذب نیز به عنوان احتمال هشدار کاذب شناخته می شود.

نتایج منحنی نمودار راک برای هر یک از اجزای سود به شرح زیر است. شکل ۱ نتایج منحنی راک را برای متغیر درصد تغییرات سود (زیان) خالص نشان می‌دهد.



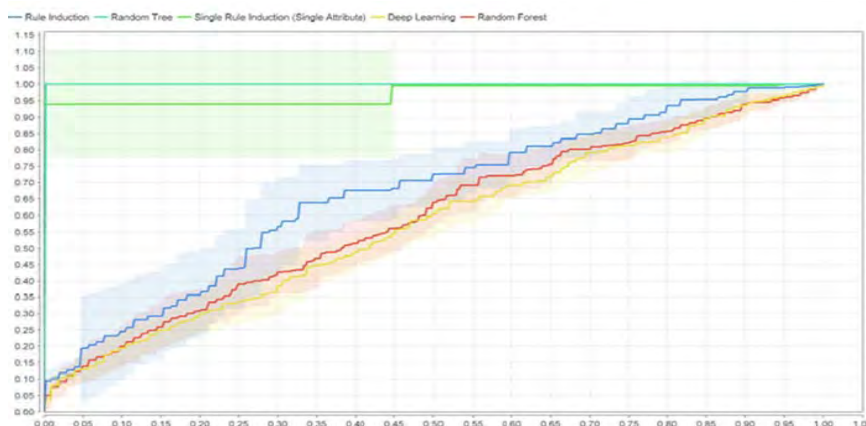
شکل ۱. منحنی راک مربوط به متغیر درصد تغییرات سود (زیان) خالص

شکل ۱ کارایی مدل درخت تصمیم را در مقایسه با سایر مدل‌ها در پیش‌بینی درصد تغییرات سود (زیان) خالص نشان می‌دهد. شکل ۲ نیز نتایج منحنی راک را برای متغیر درصد تغییرات سود (زیان) ناخالص نشان می‌دهد.



شکل ۲. منحنی راک برای درصد متغیر تغییرات سود (زیان) ناخالص

شکل ۲ نیز کارایی بیشتر مدل درخت تصمیم را در مقایسه با سایر مدل‌ها در پیش‌بینی جهت تغییرات سود (زیان) ناخالص را نشان می‌دهد. در ادامه شکل ۳ نتایج منحنی راک را برای متغیر تغییرات سود (زیان) عملیاتی نشان می‌دهد. همان گونه که مشخص است، مدل استنتاج قوانین در مقایسه با سایر مدل‌ها در پیش‌بینی تغییرات سود (زیان) عملیاتی، کارایی بیشتری را از خود نشان می‌دهد.



شکل ۳. منحنی راک مربوط به متغیر درصد تغییرات در سود (زیان) عملیاتی

نتیجه‌گیری و پیشنهادها

هدف از این پژوهش مقایسه دقت پیش‌بینی مدل‌های یادگیری ماشین و مدل‌های آماری برای پیش‌بینی جهت تغییرات اجزای سود، از جمله سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی بود. برای آزمون فرضیه‌ها، از ۲۵ مدل یادگیری ماشین و ۱۰ مدل آماری با به‌کارگیری ۲۸ متغیر مستقل و ۳ متغیر وابسته استخراج شده از صورت‌های مالی مربوط به ۱۳۹ شرکت تولیدی پذیرفته شده در بورس اوراق بهادار، طی سال‌های ۱۳۸۷ تا ۱۴۰۱ استفاده شد. دقت مدل‌های یادگیری ماشین برای پیش‌بینی جهت درصد تغییرات سود (زیان) خالص، درصد تغییرات سود (زیان) ناخالص و درصد تغییرات سود (زیان) عملیاتی، از ۸۲/۹ درصد تا ۹۳/۳ درصد و دقت مدل‌های آماری نیز از ۷۶/۴ درصد تا ۸۹/۸ درصد متغیرند. پس از مقایسه متوسط دقت مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی جهت تغییرات اجزای سود، مشخص شد که دقت پیش‌بینی مدل‌های یادگیری ماشین در مقایسه با مدل‌های آماری، به صورت معناداری بیشتر است. سپس برای انتخاب مدل بهینه (با استفاده از منحنی راک) مشخص شد که مدل درخت تصمیم با دقت پیش‌بینی معادل ۱۰۰ درصد برای پیش‌بینی سود (زیان) خالص و دقت پیش‌بینی معادل ۹۹/۳۸ درصد برای پیش‌بینی سود (زیان) ناخالص و مدل استنتاج قوانین با دقت پیش‌بینی معادل ۸۶/۷۶ درصد برای پیش‌بینی سود (زیان) عملیاتی، کارایی بیشتری دارند. می‌توان انتظار داشت که با توجه به دقت پیش‌بینی مدل‌های یادگیری ماشین در برآورد تغییرات اجزای سود، سرمایه‌گذاران بالقوه و بالفعل در کنار اطلاعات بنیادی و فنی، از این مدل‌ها برای سرمایه‌گذاری و پیش‌بینی عملکرد مالی شرکت استفاده کنند. نتایج این پژوهش با یافته‌های میرزایی و همکاران (۱۴۰۲)، آناند و همکاران (۲۰۱۹) و چن و همکاران (۲۰۲۲) مطابقت دارد. پیشنهاد می‌شود که در پژوهش‌های آتی با استفاده از مدل‌های یادگیری ماشین، به پیش‌بینی حساسیت سود (ضریب واکنش سود، نوسان‌های غیرعادی بازده و خطای پیش‌بینی سود)، کیفیت سود و پایداری سود در صنایع مختلف و برای شرکت‌های بورسی و فرابورسی، پژوهش‌های جداگانه‌ای انجام شود.

منابع

- اسدی، مسعود؛ میربرگ کار، سیدمظفر و چیرانی، ابراهیم (۱۴۰۱). ارائه یک مدل شبکه عصبی برای پیش‌بینی سود شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران و مقایسه دقت آن با مدل‌های HDZ و ARIMA. *حسابداری مدیریت*. ۱۵(۵۴)، ۱۶۳-۱۸۰.
- آشتاب، علی؛ حقیقت، حمید و کردستانی، غلامرضا (۱۳۹۶). مقایسه دقت مدل‌های پیش‌بینی بحران مالی و تأثیر آن بر ابزارهای مدیریت سود. *بررسی‌های حسابداری و حسابرسی*، ۲۴(۲)، ۱۴۷-۱۷۲.
- باقری، فاطمه؛ علیزاده مجد، حکیمه؛ مهربخش، زهرا و زیارتیان، مجید (۱۳۹۳). استفاده از الگوریتم‌های داده‌کاوی در بررسی عوامل مؤثر بر پیش‌بینی وضعیت بدو تولد نوزادان. *فصلنامه علمی پژوهشی زیست پزشکی جرجانی*، ۲(۲)، ۵۹-۶۸.
- توکلی، سامان؛ آشتاب، علی (۱۴۰۲). مقایسه کارایی مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی ریسک مالی. *راهبرد مدیریت مالی*، ۱۱(۱)، ۵۳-۷۶.
- حیدری، مهدی و امیری، امیررضا (۱۴۰۱). بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران، *تحقیقات مالی*، ۲۴(۴)، ۶۰۲-۶۲۳.
- قادری، اقبال؛ امینی، پیمان و محمدی ملقرنی، عطاءالله (۱۳۹۹). به‌کارگیری الگو ترکیبی شبکه‌های عصبی مصنوعی با الگوریتم‌های فراکاوشی (ICA, PSO) در پیش‌بینی مدیریت سود. *پژوهش‌های تجربی حسابداری*، ۱۰(۲)، ۲۱۳-۲۴۸.
- مرادی، بابک؛ بحری ثالث، جمال؛ جبارزاده کنگرلویی، سعید و آشتاب، علی (۱۴۰۱). تبیین و ارائه مدلی برای پیش‌بینی نقدشوندگی سهام در بورس اوراق بهادار تهران، *تحقیقات مالی*، ۲۴(۱)، ۱۳۴-۱۵۶.
- میرزایی، سجاد؛ آشتاب، علی و زواری رضایی، اکبر (۱۴۰۲). مقایسه کارایی مدل‌های آماری و یادگیری ماشین و انتخاب مدل بهینه در پیش‌بینی سود خالص و جریان‌های نقدی عملیاتی، *مدیریت دارایی و تأمین مالی*، ۱۱(۲)، ۵۳-۷۴.
- نوراحمدی، مرضیه و صادقی، حجت‌الله (۱۴۰۱). یادگیری ماشین مبتنی بر رویکرد سلسله‌مراتبی برابری ریسک (مطالعه موردی: پرتفولیو سهام متشکل از ۳۰ شرکت برتر بورس اوراق بهادار تهران)، *تحقیقات مالی*، ۲۴(۲)، ۲۳۶-۲۵۶.
- واعظ، علی؛ منتظر حجت، امیرحسین و بنایی قدیم، رحیم (۱۳۹۶). تأثیر ابعاد حساسیت سود (ضریب واکنش سود، نوسان‌های غیرعادی بازده و خطای پیش‌بینی سود) بر پاداش هیئت مدیره، *تحقیقات مالی*، ۱۹(۴)، ۶۱۵-۶۴۲.

References

- Anand, V., Brunner, R., Ikegwu, K. & Sougiannis, T. (2019). *Predicting profitability using machine learning*. Available at SSRN 3466478.
- Asadi, M., Mirbargkar, S. & Chirani, E. (2022). Providing a neural network model to predict the profits of companies listed on the Tehran Stock Exchange and comparing its accuracy with HDZ and ARIMA models. *Management Accounting*, 15(54), 163-180. (in Persian)

- Ashtab, A., Haghighat, H. & Kordestani, G. (2017). Comparison of Financial Distress Prediction Models Accuracy and its Effect on Earnings Management Tools. *Accounting and Auditing Review*, 24(2): 147-172. (in Persian)
- Bagheri F, Alizadeh Majd H, Mehrbakhsh Z, Ziaratban M(2014). Use of data mining algorithms in assessing the affecting factors on predicting the health status of newborns. *Jorjani Biomed Journal*, 2 (2), 59- 68. (in Persian)
- Barboza, F., Kimura, H. & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417.
- Bengio, Y., Courville, A. & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Bishop, C. M. & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, 4(4), 738. New York: springer
- Breuel, T. & Shafait, F. (2010). Automl: Simple, effective, fully automated learning rate and size adjustment. In *The Learning Workshop* (Vol. 4, p. 51). Cliff Lodge.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M. & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*: 1721-1730.
- Chatfield, C. & Xing, H. (2019). *The analysis of time series: an introduction with R*. CRC press.
- Chen, X., Cho, Y. H., Dou, Y. & Lev, B. (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research*, 60(2), 467-515.
- Dastile, X., Celik, T. & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- Francis, J., LaFond, R., Olsson, P. & Schipper, K. (2005). The market pricing of accruals quality. *Journal of accounting and economics*, 39(2), 295-327.
- Freeman, R. N., Ohlson, J. A. & Penman, S. H. (1982). Book rate-of-return and prediction of earnings changes: An empirical investigation. *Journal of accounting research*, 639-653.
- Ghaderi, E., Amini, P. & Mohammadi Molqarny, A. (2020). Application of Artificial Neural Network Hybrid Models with Metaheuristic Algorithms (PSO, ICA) in Earnings Management Forecast. *Empirical Research in Accounting*, 10(2), 213-248. doi: 10.22051/jera.2018.19246.1952 (in Persian)
- Gelman, A., Hill, J. & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.
- Gerakos, J. & Gramacy, R. (2013). Regression-based earnings forecasts. *Chicago Booth Research Paper*, 12-26.

- Gitman, L. J. (1998). *Principles of managerial finance*. Addison Wesley Longman Higher Education.
- Heidari, M. & Amiri, A.R. (2023). Inspecting the Predictive Power of Artificial Intelligence Models in Predicting the Stock Price Trend in Tehran Stock Exchange, *Financial Research Journal*, 24(4), 602-623. (in Persian)
- Huang, G. B., Wang, D. H. & Lan, Y. (2011). Extreme learning machines: a survey. *International journal of machine learning and cybernetics*, 2, 107-122.
- Jones, S., Moser, W. J. & Wieland, M. M. (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research*, 40(4), 2643-2672.
- Lev, B. & Gu, F. (2016). *The end of accounting and the path forward for investors and managers*. John Wiley & Sons.
- Li, K. K. & Mohanram, P. (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies*, 19, 1152-1185.
- Martins, A. I. (2022). *Earnings prediction using machine learning methods and analyst comparison* (Doctoral dissertation).
- Mierswa, I. (2006, July). Evolutionary learning with kernels: A generic solution for large margin problems. In *Proceedings of the 8th annual conference on genetic and evolutionary computation*, 1553-1560.
- Mills, T. C. (2019). *The econometric modelling of financial time series*. Cambridge University Press
- Mirzaei, S., Ashtab, A., & Zavari Rezaei, A. (2023). Comparing the Efficiency of Statistical Models and Machine-Learning Models and Choosing the Optimal Model for Predicting Net Profit and Operating Cash Flows. *Journal of Asset Management and Financing*, 11(2), 53-74. (in Persian)
- Monahan, S. J. (2018). Financial statement analysis and earnings forecasting. *Foundations and Trends® in Accounting*, 12(2), 105-215.
- Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2019). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Moradi, B., Bahri Sales, J., Jabarzadeh Kangharlui, Said & Ashtab, A. (2022). Explaining and Proposing a Market Liquidity Prediction Model in Tehran Stock Exchange, *Financial Research Journal*, 24(1), 134-156. (in Persian)
- Mullainathan, S. & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- Nourahmadi, M. & Sadeghi, H. (2022). A Machine Learning-Based Hierarchical Risk Parity Approach: A Case Study of Portfolio Consisting of Stocks of the Top 30 Companies on the Tehran Stock Exchange, *Financial Research Journal*, 24(2), 236-256. (in Persian)
- Ou, J. A. & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of accounting and economics*, 11(4), 295-329.

- Parlina, N. D. & Budianto, E. (2021). Implementation of cash flow as a measuring tool in predicting future net income: (case study at kedai nyobian 8 daily period september–october 2020). *Journal of management, accounting, general finance and international economic issues*, 1(1), 16-24.
- Petropoulos, A., Siakoulis, V., Stavroulakis, E. & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3), 1092-1113.
- Popescu, M. C., Balas, V. E., Perescu-Popescu, L. & Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579-588.
- Raschka, S. & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.
- Render, B. & Stair Jr, R. M. (2016). *Quantitative Analysis for Management*, 12e. Pearson Education India.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5), 206-215.
- Schipper, K. & Vincent, L. (2003). Earnings quality. *Accounting horizons*, 17, 97-110.
- Tavakoli, S. & Ashtab, A. (1402). Comparing the effectiveness of machine learning models and statistical models in predicting financial risk. *Financial management strategy*, 11(1), 53-76. (in Persian)
- Urso, A., Fiannaca, A., La Rosa, M., Ravì, V. & Rizzo, R. (2018). Data mining: Prediction methods. *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma*, 1, 3.
- Vaez, S.A., Montazer Hojat, A.H. & Ghadim, R.B. (2017). The Effect of Profit Sensitivity Dimensions (Earnings Response Coefficient, Returns Abnormal Fluctuations and Earning Prediction Error) on Board of Director's Compensation. *Financial Research Journal*, 19(4), 615-642. (in Persian)
- Wahlen, J. M. & Wieland, M. M. (2011). Can financial statement analysis beat consensus analysts' recommendations? *Review of Accounting Studies*, 16, 89-115.