Securities & Exchange Organization, Research, Development & Islamic Studies (RDIS)
Journal of Securities and Exchange, Summer 2025, V. 18, No.70, pp. 165-198

Evaluation of the Financial Performance of Companies Listed on the Tehran Stock Exchange Using K-Means and K-Mediod Clustering[1]

Abdolrasoul Mostajeran[2], Peyman Pazooki[3], Mohamad Amin Rafienia[4]

Received: 2024/10/09
Accepted: 2025/03/05

Research Paper

Abstract

With the rapid growth of financial data in publicly listed companies, selecting an accurate and efficient method for analysis and clustering has become increasingly important. This study evaluates the financial performance of companies listed on the Tehran Stock Exchange (TSE) and clusters them using two methods: K-Means and K-Medoids. Key financial ratios—return on assets, return on equity, gross profit margin, debt-to-equity ratio, and price-to-book ratio—were analyzed for the period 2014–2024. The K-Means algorithm demonstrated faster clustering due to its efficiency with large datasets, whereas K-Medoids, being more robust to outliers and noise, yielded more precise groupings. ANOVA results confirmed statistically significant differences among the clusters based on the selected financial indicators. Overall, K-Medoids provided more balanced and accurate classifications, while K-Means proved more suitable for simpler, high-volume datasets. These findings offer practical insights for investors and financial analysts to make better strategic decisions aligned with their risk tolerance and expected returns.

Key Words: Stock Exchange, K-Mean Clustering, K-Medoid Clustering, Asset Returns, Equity Returns, Gross Profit Margin
JEL Classification: G32.

---

1. doi: 10.22034/JSE.2025.12511.2323

2. Assistant Professor, Department of Financial Mathematics, Faculty of Financial Sciences, Kharazmi University, Tehran, Iran. Corresponding Author. (rasoul.mostajeran@khu.ac.ir).

3. Ph.D. Student, Department of Financial Management, Faculty of Market and Business, South Tehran Branch, Islamic Azad University, Tehran, Iran. (peyman.pazouki@iau.ac.ir).

4. PhD student, Department of Financial Management, Faculty of Market and Business, South Tehran Branch, Islamic Azad University, Tehran, Iran. (amin97rf@gmail.com).

سازمان بورس و اوراق بهادار، مرکز پژوهش، توسعه و مطالعات اسلامی

فصلنامه بورس اوراق بهادار، سال هجدهم، شماره ۷۰، تابستان ۱۴۰۴، صص ۱۹۸-۱۶۵

ارزیابی عملکرد مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با خوشه‌بندی‌های K-Mean و K-Mediod[1]

عبدالرسول مستاجران[2]، پیمان پازوکی[3]، محمدامین رفیعی‌نیا[4]

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۷/۱۸

تاریخ پذیرش: ۱۴۰۳/۱۲/۱۵

چکیده

با افزایش حجم داده‌های مالی در شرکت‌های بورسی، انتخاب روشی دقیق و کارآمد برای تحلیل و خوشه‌بندی این داده‌ها از اهمیت بالایی برخوردار است. هدف این پژوهش بررسی عملکرد مالی شرکت‌های بورسی ایران و خوشه‌بندی آنها با استفاده از دو روش خوشه‌بندی K-Mean و K-Medoid است. این مطالعه بر اساس نسبت‌های مالی کلیدی شامل بازده دارایی‌ها، بازده حقوق صاحبان سهام، حاشیه سود ناخالص، نسبت بدهی به حقوق صاحبان سهام و نسبت قیمت به ارزش دفتری انجام شده است. داده‌های مالی مربوط به شرکت‌های فعال در بورس اوراق بهادار تهران در بازه زمانی ۱۳۹۳ تا ۱۴۰۳ جمع‌آوری و تحلیل شده‌اند. روش K-Mean به دلیل سرعت بالا و کارایی مناسب در داده‌های بزرگ، خوشه‌بندی سریع‌تری ارائه می‌دهد، در حالی که روش K-Medoid به دلیل مقاومت بیشتر در برابر داده‌های پرت و دارای اغتشاش، خوشه‌بندی‌های دقیق‌تری انجام داده است. تحلیل واریانس (ANOVA) نشان‌دهنده تفاوت معنادار بین خوشه‌ها بر اساس معیارهای مالی یادشده است. نتایج پژوهش نشان می‌دهد که روش K-Medoid توزیع متعادل‌تر و نتایج دقیق‌تری در تفکیک شرکت‌ها ارائه کرده است، در حالی که روش K-Mean برای داده‌های ساده‌تر و با حجم بالا مناسب‌تر است. یافته‌های این پژوهش می‌تواند به سرمایه‌گذاران و تحلیل‌گران مالی کمک کند تا بر اساس خوشه‌های مشخص‌شده و با توجه به ریسک‌پذیری و بازدهی مورد انتظار خود، تصمیمات استراتژیک بهتری اتخاذ کنند.

واژه‌های کلیدی: بورس اوراق بهادار، خوشه‌بندی K-Mean، خوشه‌بندی K-Medoid، بازده دارایی‌ها ، بازده حقوق صاحبان سهام، حاشیه سود ناخالص.

طبقه‌بندی موضوعی: G32.

---

۱. doi: 10.22034/JSE.2025.12511.2323

۲. استادیار، گروه ریاضیات مالی، دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران. (نویسنده مسئول). (rasoul.mostajeran@khu.ac.ir).

۳. دانشجوی دکترا، گروه مدیریت مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. (peyman.pazouki@iau.ac.ir).

۴. دانشجوی دکترا، گروه مدیریت مالی ، واحد تهران جنوب، دانشگاه آزاد اسلامی ، تهران، ایران.(ایمیل: amin97rf@gmail.com).

مقدمه

امروزه با پایگاه‌های داده‌ای بزرگ‌تر و پیچیده‌تری سروکار داریم که به تبع آن داده‌ها و اطلاعاتی که باید مورد پردازش و تحلیل قرار بگیرند بسیار متنوع، جامع و ناهمگن هستند. حجم بزرگ داده‌ها به تنهایی به مدیران سازمان در تصمیم‌گیری کمکی نمی‌کند، بلکه باعث سردرگمی مدیران نیز می‌شود؛ ازین رو برای تحلیل داده‌ها ضروری است پایگاه‌های داده بزرگ و همچنین اطلاعات در دسترس به نوعی سیستماتیک و مرتب شوند. در این راستا، مناسب‌ترین راه‌حل یا روش این است که داده‌ها به گروه‌های همگن‌تر و با ویژگی‌های نزدیک به هم دسته‌بندی و تقسیم شوند. ماهیت این کار مبتنی بر نزدیکی ویژگی‌های عناصر هر گروه به هم است که با عناصر گروه‌های دیگر نیز متفاوت باشند. (لوکاچ و همکاران[1]، ۲۰۲۱) در روش خوشه‌بندی، داده‌ها به چند گروه تقسیم می‌شوند به طوری که شباهت بین داده‌های درون هر خوشه حداکثر باشد. (شیری، ۱۳۸۵) با کمک گروه‌های همگن، می‌توانیم هر گروه را بهتر و دقیق‌تر ارزیابی و تحلیل کنیم و در نهایت تصویری واقعی‌تر از ارزیابی کل پایگاه داده دراختیار داشته باشیم. زمانی که می‌خواهیم به بررسی تاثیر تعداد کمی از عوامل بر هدف بپردازیم معمولا روش‌های آماری مناسب است ولی زمانی که تعداد این عوامل زیاد می‌شود دیگر این روش‌ها کارایی مناسبی ندارند. این امر سبب شد که روش‌های ابتکاری دیگری افزون بر روش‌های آماری مانند شبکه‌های عصبی[2] و درخت تصمیم‌گیری[3] برای این منظور ایجاد شود. (کمالیان و همکاران، ۱۳۹۰) از داده کاوی میتوان برای انجام کارهایی مانند طبقه‌بندی[4] پیش‌بینی[5]، تخمین[6] و خوشه‌بندی[7] داده‌ها استفاده کرد. برای انجام این کارها روش‌هایی توسعه یافته‌اند که برخی از آنها عبارت‌اند از: الگوریتم‌های خوشه‌بندی[8]، شبکه‌های عصبی، الگوریتم ژنتیک، نزدیکترین همسایگی[9] و درخت تصمیم‌گیری (ایزدپرست، ۱۳۹۰) (اعتمادی و همکاران، ۱۳۹۳) همانطور که بیان شد، یکی از روش‌های متداول برای تحلیل داده‌ها، الگوریتم‌های خوشه‌بندی است.

---

1. Lukác et al
2. Neural Network
3. Decision Tree
4. Classification
5. Perediction
6. Estimation
7. Clustering
8. Cluster Detection Algorythm
9. Nearest Neighboring

خوشه‌بندی ما را قادر می‌سازد تا به جای پردازش حجم انبوه اطلاعات، تنها اطلاعاتی را بررسی و تحلیل کنیم که بسیار به هم شبیه هستند و این خود یک گام بلند برای ساده کردن مسئله است. (ایمانی، ۱۳۹۱) در مورد تحلیل داده‌های مالی، پردازش دقیق پایگاه‌های داده بزرگ و در نتیجه ارزیابی نتایج حاصل از آن دارای اهمیت است. ارزیابی عملکرد مالی شرکت‌ها نیازمند مقایسه شرکت‌هایی است که از نظر اندازه و عملکرد تفاوت‌های زیادی دارند و این کار بر اساس پردازش و تحلیل گزارش‌های چند ساله این شرکت‌ها انجام می‌شود.

مقادیر میانگین یک پایگاه داده به علت آن که شامل مجموعه‌ای گسترده از داده‌های متغیر و به شدت متفاوت است، نمی‌تواند نمایانگر مناسبی باشد، بنابراین تحلیل خوشه‌ای بهترین راه‌حل برای تحلیل پایگاه داده است. امروزه تکنیک‌ها و روش‌های مختلفی برای انجام گروه‌بندی همگن داده‌ها وجود دارد. روش‌های خوشه‌بندی توسط مادهولاتا[1] (۲۰۱۲) ارائه شده که بینش‌های ارزشمندی در مورد الگوریتم‌های خوشه‌بندی و کاربردهای آن، ویژگی‌ها و همچنین محدودیت‌های مربوطه را ارائه می‌دهد. رایج‌ترین و متداول‌ترین روش‌ها در این زمینه، الگوریتم‌های K-mean و K-Medoid هستند (سونی و پاتل[2]، ۲۰۱۷). در روش K-mean از میانگین گروه برای گروه‌بندی داده‌ها استفاده می‌شود. از محدودیت‌های این روش می‌توان به امکان تحریف در نتایج به دلیل وجود مقادیر با تفاوت و تمایز زیاد اشاره کرد. برای تعریف نقاط مرجع، روش K-Medoid به جای مقادیر میانگین از مقادیر نماینده (مدوئیدها) استفاده می‌کند (ول موروگان و سنتانام[3]، ۲۰۱۰). روش K-Medoid در مقایسه با خوشه‌بندی K-mean نسبت به داده‌های پرت حساسیت کمتری دارد (پارک و جون[4]، ۲۰۰۹). پیش از گروه‌بندی پایگاه‌های داده، شناسایی و حذف داده‌های پرت به دلیل تأثیرات منفی آن‌ها بر فرآیند گروه‌بندی و به منظور دستیابی به نتایج دقیق‌تر ضروری است (بن‌گال[5]، ۲۰۱۰). در این راستا می‌توان از روش‌هایی مانند روش مبتنی بر فاصله، روش مبتنی بر تراکم و روش‌های گراف که برای شناسایی داده‌های پرت شناخته شده‌اند، بهره برد (اقبال و همکاران[6]، ۲۰۱۷).

---

1. Madhulatha
2. Soni & Patel
3. Velmurugan & Santhanam
4. Park & Jun
5. Ben-Gal
6. Iqbal et al

مبانی نظری و توسعه فرضیه‌ها

بخش عمده‌ای از پژوهش‌های انجام گرفته تاکنون به پردازش آن دسته از پایگاه‌های داده می‌پردازد که شامل عناصر بزرگ و متنوع هستند و در آن‌ها، تجزیه داده‌ها به گروه‌های همگن صورت می‌پذیرد. روش‌های خوشه‌بندی در بسیاری از زمینه‌ها مانند زیست‌شناسی، طبقه‌بندی بیماری‌ها، باستان‌شناسی، تقسیم‌بندی تصاویر، طبقه‌بندی‌های اجتماعی و بخش‌بندی بازار، و همچنین در تحلیل پایگاه‌های داده در بسیاری از زمینه‌ها از جمله کارآفرینی (سابو و هرمان[1]، ۲۰۱۴)، حمل و نقل (اصفهانی و همکاران، ۲۰۱۹)، تحقیقات پزشکی (ویندگاسن و همکاران[2]، ۲۰۱۸)، مهندسی (فام و افیفی، ۲۰۰۷) و عملکرد خوشه‌های صنعتی (یادگاری و همکاران، ۲۰۱۸) مورد استفاده قرار می‌گیرند. در حوزه اقتصاد و امور مالی، همزمان با افزایش نیاز به داده‌ها و پایگاه‌های داده بزرگ و پیچیده، نیاز به پردازش داده‌های اقتصادی و مالی نیز افزایش یافته است. از میان حجم عظیمی از اطلاعاتی که برای تصمیم‌گیرندگان در دسترس است، تنها اطلاعاتی که برای آن‌ها با ارزش و مفیدتر هستند باید انتخاب شوند.

در مورد تحلیل‌های اقتصادی-مالی، منابع علمی روش‌های متعددی از گروه‌بندی را در زمینه‌های مختلف توصیف می‌کنند از جمله: تحلیل ریسک (لمیو و همکاران[3]، ۲۰۱۴)، تحلیل ریسک مالی (کو و همکاران[4]، ۲۰۱۴) (ژیائوجون و یالین[5]، ۲۰۲۱) انتخاب نسبت‌های مالی (وانگ و لی[6]، ۲۰۰۸)، فعالیت‌های تقلب اقتصادی (سابائو[7]، ۲۰۱۲)، تحلیل پرتفوی املاک و مستغلات (هپسن و واتان‌سور[8]، ۲۰۱۲)، و تحلیل عملکرد مالی (پوپا و همکاران[9]، ۲۰۲۱). کای و همکاران[10] (۲۰۱۲) بینش‌های ارزشمندی در مورد روش‌های خوشه‌بندی ارائه داده و مزایا و معایب آن‌ها را برای مجموعه داده‌های مالی احصاء کرده‌اند. آن‌ها نشان دادند که خوشه‌بندی استوار بر تراکم برای مجموعه داده‌های مالی مناسب نیست، در حالی که روش K-mean بهترین تعداد خوشه‌ها را برای کمک به درک طبقه‌بندی داده‌های مالی ارائه می‌دهد.

---

1. Szabo & Herman
2. Windgassen et al
3. Lemieux et al
4. (Kou et al; 2014)
5. Xiaojun & Yalin
6. Wang & Lee
7. Sabau
8. Hepsen & Vatansever
9. Popa et al
10. Cai et al

روش‌های تجزیه گروهی نیز برای گروه‌بندی شرکت‌ها آن گونه که توسط سربان و همکاران[1] (۲۰۱۳) نشان داده شد، بسیار مناسب و کاربردی هستند. این نویسندگان روش‌های خوشه‌بندی را برای ۱۰۶ شرکت بر اساس شاخص‌های اقتصادی شامل ارزش افزوده اقتصادی، درآمد خالص، فروش فعلی، حقوق صاحبان سهام و قیمت سهام اعمال کردند. کائور و همکاران[2] (۲۰۱۴) خوشه‌بندی را به عنوان «فرایند گروه‌بندی اشیاء داده به خوشه‌های مجزا به گونه‌ای که داده‌ها در هر خوشه مشابه باشند، اما با سایر خوشه‌ها متفاوت باشند» تعریف کردند. روش‌ها و تکنیک‌های مختلفی برای ایجاد این خوشه‌ها وجود دارد، مانند K-mean و K-Medoid. بر اساس گفته کائور و همکاران (۲۰۱۴)، روش K-mean یک روش ساده و آسان برای استفاده است. هر خوشه بر اساس مقادیر میانگین تشکیل می‌شود و مقادیر یا عناصری که به یک خوشه تعلق دارند، به این میانگین نزدیک‌تر هستند. با این حال، همان نویسندگان (۲۰۱۴) دو عیب عمده این روش را نیز مطرح کردند: حساسیت به مقادیر حدی و نبود دانش در مورد تعداد خوشه‌ها. مدللو و نوگراها[3] (۲۰۲۱) روش K-mean را به عنوان روشی توصیف کردند که مقادیر را از یک پایگاه داده به گروه‌های مشخصی گروه‌بندی می‌کند به‌طوری که سطح مشابهت داده‌ها در یک گروه تا حد امکان بالا و در مقایسه با سایر گروه‌ها تا حد امکان پایین باشد. سطح مشابهت بر اساس فاصله بین مقادیر هر عنصر در گروه و میانگین گروه تعیین می‌شود. نویسندگان بر سادگی این روش تأکید کردند.

روش خوشه‌بندی K-mean توسط ایکوتون و همکاران[4] (۲۰۲۱) از نظر مزایا، بهبودهای روش کلاسیک، نقاط قوت و ضعف پیاده‌سازی موجود و همچنین اجرای K-mean ترکیبی بر اساس الگوریتم‌های فراکاوشی الهام گرفته از طبیعت و شناسایی روندهای جدید تحلیل شده است. همان نویسندگان الگوریتم‌ها و بهینه‌سازی‌های مختلفی را ارائه دادند که می‌تواند برای جامعه پژوهشی مفید باشد. روش گروه‌بندی K-mean توسط هرنانت و همکاران[5] (۲۰۰۷) برای مقایسه سوپرمارکت‌ها بر اساس شاخص‌های مالی مختلف استفاده شد. کراماریچ و همکاران[6] (۲۰۱۸) نیز از روش خوشه‌بندی K-mean برای مقایسه شرکت‌های بیمه براساس شاخص مالی

1. Serban et al
2. Kaur et al
3. Medellu & Nugraha
4. Ikotun et al
5. Hernant et al
6. Kramaric et al

بازده حقوق صاحبان سهام استفاده کردند. تحلیل خوشه‌بندی K-mean و تحلیل مؤلفه اصلی برای طبقه‌بندی ۲۵ کشور اتحادیه اروپا و ارائه نمایی مقایسه‌ای از تعامل بین کارآفرینی دیجیتال و متغیرهای توسعه پایدار به کار گرفته شد. (هرمان[1]، ۲۰۲۲) این روش به‌عنوان محبوب‌ترین تکنیک خوشه‌بندی برای تقسیم یک مجموعه داده به یک مجموعه از k گروه (خوشه) در نظر گرفته می‌شود.

(پچ و ورچوتا، ۲۰۲۰) در مقایسه با K-Mean، روش K-Medoid با مقادیر نماینده برای هر خوشه کار می‌کند. به جای میانگین‌ها، روش K-Medoid مقادیر نماینده‌ای را انتخاب می‌کند که مقادیر مرکزی در خوشه هستند. مقادیر در پایگاه داده با آن خوشه‌هایی مرتبط هستند که مدوئیدهای آن‌ها بیشترین شباهت را دارند. کائور و همکاران[2] (۲۰۱۴) مدوئید را به‌عنوان «عنصری از یک خوشه که میانگین عدم تشابه آن با تمام اشیاء در خوشه حداقل است، یعنی نقطه‌ای که مرکزی‌ترین موقعیت را در مجموعه داده مورد نظر دارد» تعریف کردند. در این روش از بین داده‌های متعدد، تعداد مشخصی نماینده به‌طور تصادفی انتخاب می‌شوند که گروه‌ها با عناصری تشکیل می‌شوند که بیشترین شباهت را به مقادیر نماینده دارند. پس از تشکیل این خوشه‌ها، نماینده‌های جدیدی انتخاب می‌شوند که گروه تشکیل‌شده را بهتر نمایندگی کنند. این روند ادامه می‌یابد تا زمانی که هیچ نماینده‌ای موقعیت خود را تغییر ندهد (ویشواکارما و همکاران[3]، ۲۰۱۷). این روش شاید بزرگ‌ترین عیب روش K-Mean یعنی حساسیت به مقادیر حدی را برطرف می‌کند (مدلو و نوگراها[4]، ۲۰۲۱).

دو روش خوشه‌بندی محبوب K-Medoid و K-Mean توسط آرورا و همکاران (۲۰۱۶) مورد تحلیل قرار گرفتند و در نتیجه آن‌ها تأیید کردند که زمان اجرا برای روش K-Medoid بهتر از K-Mean است و همچنین روشK-Medoid نسبت به داده‌های پرت حساس نیست. برخلاف این نتایج، ولموروگان[5] (۲۰۱۲) نیز الگوریتم‌های خوشه‌بندی K-Medoid و K-Mean را مقایسه کرد و نتایج تجربی نشان داد که الگوریتم K-Mean بهترین نتایج را در مقایسه با الگوریتم K-Medoid ارائه می‌دهد. افزون بر این، دسوزا و همکاران[6] (۲۰۱۷) جنبه‌های مثبت

1. Herman
2. Kaur et al
3. Vishwakarma et al
4. Medellu & Nugraha
5. Velmurugan
6. Dsouza et al

و منفی این دو روش را مورد بررسی قرار دادند و نتیجه‌گیری کردند که K-Medoid در تمام جنبه‌ها مانند زمان اجرا، حساسیت نداشتن به داده‌های پرت و کاهش خطا بهتر است، اما با این محدودیت که پیچیدگی آن در مقایسه با K-Mean بیشتر است. خلاصه نظرات در ادبیات، مزایا و معایب روش‌های K-Mean و K-Medoid در جدول ۱ نشان داده شده است.

جدول ۱. مزایا و معایب روش‌های K-Mean و K-Medoid

	K-Mean Method	K-Medoid Method
مزایا	روش شناخته‌شده و کاربردی (دسوزا و همکاران، ۲۰۱۷؛ پچ و ورچوتا[1]، ۲۰۲۰) پیاده‌سازی ساده و سریع (کائور و همکاران، ۲۰۱۴؛ مدلو و نوگراها، ۲۰۲۱؛ دسوزا و همکاران) مقیاس‌پذیری برای داده‌های بزرگ	روش شناخته‌شده و کاربردی (دسوزا و همکاران، ۲۰۱۷) حساسیت کمتر به داده‌های پرت (آرورا و همکاران، ۲۰۱۶؛ دسوزا و همکاران، ۲۰۱۷) زمان اجرای کمتر برای هر خوشه در مقایسه با K-Mean (کائور و همکاران، ۲۰۱۴؛ دسوزا و همکاران، ۲۰۱۷؛ آرورا و همکاران، ۲۰۱۶)
معایب	زمان اجرای بیشتر برای هر خوشه در مقایسه با K-Medoid (کائور و همکاران، ۲۰۱۴؛ دسوزا و همکاران، ۲۰۱۷؛ آرورا و همکاران، ۲۰۱۶) حساسیت به داده‌های پرت (کائور و همکاران، ۲۰۱۴؛ دسوزا و همکاران، ۲۰۱۷) تعیین تعداد خوشه‌ها به صورت دستی	تعیین تعداد خوشه‌ها به صورت دستی پیچیدگی بالاتر در مقایسه با K-Mean (دسوزا و همکاران، ۲۰۱۷؛ آرورا و همکاران، ۲۰۱۶)

در هر دو روش، تعریف تعداد خوشه‌ها اهمیت زیادی پیدا می‌کند. برای این منظور، چندین روش وجود دارد، متداول‌ترین روش‌ها، روش آرنج (Elbow) و روش سیلوئت (Silhouette) هستند (پچ و ورچوتا[2]، ۲۰۲۰؛ سزوله[3]، ۲۰۱۹؛ دزوبا و کرایلوو[4]، ۲۰۲۱) تکنیک دیگری که می‌تواند مشکل تشخیص تعداد خوشه‌ها را حل کند استفاده از VAT[5] (ارزیابی بصری تمایل) برای خوشه‌بندی (پاخیرا[6]، ۲۰۱۲) و تحلیل خوشه‌ای سلسله‌مراتبی است. کودیناریا و ماکوانا[7] (۲۰۱۳) روش‌های مختلف تعیین تعداد خوشه‌ها که در الگوریتم

1. Pech & Vrchota
2. Pech & Vrchota
3. Szüle
4. Dzuba & Krylov
5. Visual Assessment of Tendency
6. Pakhira
7. Kodinariya & Makwana

خوشه‌بندی K-Mean اعمال می‌شوند را به تفصیل توضیح داده‌اند. از میان روش‌های بیان شده، روش آرنج به‌عنوان قدیمی‌ترین و متداول‌ترین روش شناخته می‌شود. این یک روش بصری است که ماهیت آن افزایش تعداد گروه‌ها به صورت تکی است که از دو شروع می‌شود. برای هر گام می‌توان واریانس را مشاهده کرد و زمانی که نمودار شکستگی آرنج را نشان داد، می‌توان فرآیند را متوقف نمود، زیرا تعداد خوشه بهینه پیدا شده است. این شکل ممکن است نقطه شکست را نشان دهد که در آن یک «جهش» در درجه ناهمگونی رخ می‌دهد (سایمون[1]، ۲۰۰۶).

روش‌شناسی پژوهش

پایگاه داده مورد استفاده در این مطالعه شامل شامل داده‌ها و شاخص‌های مالی شرکت‌های فعال در بورس اوراق بهادار تهران و فرابورس ایران در بازه زمانی ۱۰ سال، از سامانه بورس اوراق بهادار تهران، نرم افزار TSEclient، رهاورد ۳۶۰ و سایر سامانه‌های مرتبط جمع‌آوری شده است. با استفاده از نرم افزار SPSS و برنامه نویسی پایتون[2] مراحل پیش‌پردازش داده‌ها و نرمال‌سازی، شناسایی و حذف داده‌های پرت با استفاده از شناسایی داده‌های پرت با استفاده از الگوریتم جنگل ایزوله [3]و کنترل با نمودار جعبه‌ای[4]، انتخاب تعداد خوشه بهینه با استفاده از روش آرنج و دندوگرام در SPSS، خوشه‌بندی با الگوریتم K-Mean به‌وسیله SPSS، خوشه‌بندی با الگوریتم K-Medoid با استفاده از پایتون و در نهایت مقایسه دو روش خوشه‌بندی انجام شد.

این جدول آمار توصیفی، اطلاعاتی از پنج شاخص مالی مهم را نشان می‌دهد. داده‌های بررسی شده شامل ۶۶۴ مشاهده برای هر شاخص است. بیشترین چولگی و کشیدگی مربوط به شاخص نسبت بدهی به حقوق صاحبان سهام است که نشان‌دهنده توزیع غیرمتقارن داده‌ها است و احتمال وجود نقاط پرت را افزایش می‌دهد. سایر شاخص‌ها مانند بازده دارایی‌ها دارای چولگی و کشیدگی کمتری هستند که نشان‌دهنده توزیع متقارن‌تر و مسطح‌تر داده‌هاست.

---

1. Simon
2. Python
3. Isolation Forest
4. Box plot

جدول ۲. آمار توصیفی داده های استخراج شده

شرح	تعداد N	حداقل مقدار	حداکثر مقدار	میانگین	انحراف معیار	چولگی	خطای چولگی	کشیدگی	خطای کشیدگی
نسبت قیمت به ارزش دفتری	۶۶۴	-۰,۵۱۹۷۰۸۴۱۳	۲,۷۹۶۶۷۰۵۴۷	-۰,۱۰۳۰۰۴۷۸۸	۰,۴۱۱۳۵۹۱۶۹	۲,۷۸۸	۰,۰۹۵	۱۱,۱۱۶	۰,۱۸۹
بازده حقوق صاحبان سهام	۶۶۴	-۱,۸۱۱۳۶۴۷۳۸	۲,۵۴۹۰۳۰۳۹۵	-۰,۰۸۰۸۹۳۰۹۷	۰,۷۳۵۹۷۴۵۱۶	۰,۸۷۷	۰,۰۹۵	۰,۵۴۹	۰,۱۸۹
حاشیه سود ناخالص	۶۶۴	-۳,۲۶۸۶۵۲۱۱	۴,۹۹۶۴۸۱۳۳۲۵	-۰,۰۱۱۲۲۶۵۳۴	۰,۹۶۸۱۴۳۸۷۵	۱,۱۱۸	۰,۰۹۵	۲,۱۶۴	۰,۱۸۹
نسبت بدهی به حقوق صاحبان سهام	۶۶۴	-۰,۴۲۶۵۶۳۷۴۱	۴,۰۵۹۷۱۳۶۱	-۰,۰۹۷۱۱۰۷۲۷	۰,۴۷۹۸۵۰۳۰۳	۳,۹۸	۰,۰۹۵	۲۲,۴۳۷	۰,۱۸۹
بازده دارایی	۶۶۴	-۲,۴۱۷۸۶۲۲۵	۲,۸۱۷۴۲۹۳۳۶	-۰,۰۵۲۲۶۴۹۷	۰,۸۳۴۴۸۷۸۵	۰,۶۸	۰,۰۹۵	۰,۱۰۵	۰,۱۸۹
معتبر به صورت لیست‌وار	۶۶۴								

میانگین بیشتر شاخص‌ها منفی بوده که ممکن است عملکرد منفی در این دوره را نشان دهد. پراکندگی بیشتر در «حاشیه سود ناخالص» با انحراف معیار بالا مشاهده می‌شود، در حالی که بازده حقوق صاحبان سهام پایداری بیشتری دارد. این یافته‌ها می‌تواند برای ارزیابی ریسک‌ها و تصمیم‌گیری‌های مالی مفید باشد.

متغیرهای و شاخص‌های عملکرد مالی

شاخص‌های مورد بررسی در این مقاله شامل قیمت به ارزش دفتری هر سهم[1]، بازده حقوق صاحبان سهام[2]، بازده دارایی‌ها[3]، حاشیه سود ناخالص[4]، نسبت بدهی به حقوق صاحبان سهام[5] هستند.

1. Price-to-Book Ratio (P/B)
2. Return on Equity (ROE)
3. Return on Assets (ROA)
4. Gross Profit Margin
5. debt to equity ratio

هرچیو و همکاران (۲۰۱۰) معتقد بودند که شاخص بازده دارایی‌ها که نشان‌دهنده سودآوری استفاده از دارایی‌ها است، باید در همه موارد حداقل ۵ درصد باشد. در مورد بازده حقوق صاحبان سهام که مهم‌ترین شاخص برای سرمایه‌گذاران بوده و نشان می‌دهد مدیریت یک شرکت چگونه از پول سرمایه‌گذاران به‌طور مؤثر استفاده می‌کند، بیشتر سرمایه‌گذاران حرفه‌ای به دنبال سرمایه‌گذاری‌هایی هستند که بازدهی بیش از ۱۵ درصد داشته باشند.

هاتم مقایسه بین‌المللی شرکت‌های سه کشور را با استفاده از شاخص‌های ROS[1]، ROA، ROE و سایر شاخص‌ها انجام داد (هاتم[2]، ۲۰۱۴؛ انگوین و همکاران[3]، ۲۰۱۹). سودآوری ۵۸ شرکت بورسی در ویتنام را نیز بر اساس شاخص‌های مالی ROS، بازده دارایی‌ها و بازده حقوق صاحبان سهام تحلیل کرده و مقادیر میانگین برای این شرکت‌ها را ۹۰ درصد برای ROS، ۱,۵۰ درصد برای بازده دارایی‌ها و ۳,۹۴ درصد برای بازده حقوق صاحبان سهام گزارش کردند.

برای شرکت‌های بورسی رومانی، پوپا و همکاران[4] (۲۰۲۱) شاخص‌های بازده دارایی‌ها و بازده حقوق صاحبان سهام را مطالعه کردند و به مقادیر زیر دست یافتند: میانگین ۲ درصد برای بازده دارایی‌ها با حداقل ۱۴۷٪ و حداکثر ۲۹٪ و میانگین ۸٪ برای بازده حقوق صاحبان سهام با حداقل ۱,۲۰۱٪ – و حداکثر ۱,۰۱۳٪. همان‌طور که در بالا نشان داده شد، مقادیر ROS، بازده دارایی‌ها و بازده حقوق صاحبان سهام بسیار متفاوت هستند، اما در همه موارد اولین الزام برای این شاخص‌ها مثبت بودن آن‌ها است و پیشنهاد می‌شود که این مقادیر برای ROS و بازده دارایی‌ها بالای ۵٪ و برای بازده حقوق صاحبان سهام حداقل ۱۰٪ و در حال افزایش باشد.

افزون بر این، مطالعات و تحلیل‌های اقتصادی-مالی بسیاری وجود دارند که نشان می‌دهند شرکت‌های بسیاری برای ارزیابی عملکرد مالی خود در فعالیت‌های اقتصادی مختلف از این سه شاخص و همچنین حاشیه سود ناخالص و قیمت به ارزش دفتری هر سهم استفاده می‌کنند که از این جمله

---

۱. Return on Sales یا بازده فروش یک نسبت مالی است که نشان می‌دهد یک شرکت چقدر از درآمد حاصل از فروش خود را به سود عملیاتی تبدیل می‌کند. این نسبت از تقسیم سود عملیاتی بر فروش خالص به دست می‌آید.

2. Hatem

3. Nguyena et al

4. Sabău Popa et al

شرکت‌ها می‌توان شرکت‌های بورسی (ژوسوپوا و همکاران[۱]، ۲۰۱۸؛ بایارا[۲]، ۲۰۱۷؛ ین و همکاران[۳]، ۲۰۲۲)، شرکت‌های بورسی غیرمالی (سزوله[۴]، ۲۰۱۹؛ اختر و همکاران[۵]، ۲۰۲۲) [۳۴،۴۹]، شرکت‌های حمل‌ونقل و انبارداری (دو[۶]، ۲۰۲۱)، کشاورزی (گورال و سولیوودا[۷]، ۲۰۲۱؛ هلوشکووا و لکه‌شووا[۸]، ۲۰۲۰) ، صنعت آرایشی (د بلازیو و همکاران[۹]، ۲۰۲۰) مواد غذایی و نوشیدنی (فنیوش[۱۰] و همکاران، ۲۰۲۰) ، صنعت خودروسازی (زین‌الدین[۱۱] و همکاران، ۲۰۲۱) و سایر صنایع (خور و همکاران[۱۲]، ۲۰۱۹؛ پاولکووا و همکاران[۱۳]، ۲۰۲۱) را نام برد.

پوپا و همکاران[۱۴] (۲۰۲۱) همچنین شاخص‌های بازده دارایی‌ها و بازده حقوق صاحبان سهام (به‌علاوه شش شاخص دیگر) را برای ساخت یک شاخص مالی مرکب به‌منظور تعیین عملکرد مالی شرکت‌های بورسی انتخاب کردند. پلونئووا (۲۰۲۱) از شاخص‌های بازده حقوق صاحبان سهام در تعیین ارزش افزوده اقتصادی[۱۵]، بازده دارایی‌ها و ROS برای مقایسه عملکرد مالی شرکت‌های منتخب که در خوشه‌های مختلف در جمهوری چک و اسلواکی قرار داشتند، استفاده کرد. به‌دلیل تفاوت‌های بین شرکت‌ها، آفرمایانی و دویانتو[۱۶] (۲۰۲۱) در زمینه قیمت‌های سهام و (گولاغیز و ساهین[۱۷]، ۲۰۱۷؛ ساجتوس و میتو[۱۸]، ۲۰۰۷) [۶۷] در زمینه عملکرد مالی، از روش خوشه‌بندی و شاخص‌های مالی بازده دارایی‌ها و بازده حقوق صاحبان سهام برای مقایسه عملکرد مالی شرکت‌های بورسی استفاده کردند.

از شاخص‌هایی همچون نسبت جاری و نسبت آنی به‌عنوان معیارهای نقدینگی به دلیل آنکه تمرکز مطالعه بر عملکرد بلندمدت شرکت‌ها بود، نه وضعیت نقدینگی کوتاه‌مدت و نسبت سود

---

1. Zhussupova et al
2. Bayaraa
3. Yen et al
4. Szüle
5. Akhtar et al
6. Do
7. Goral & Soliwoda
8. Hloušková & Lekešová
9. De Blasio et al
10. Fenyves et al
11. Zainudin et al
12. Khour et al
13. Pavelkova et al
14. Popa et al
15. Economic Value Added (EVA)
16. Afrimayani & Devianto
17. Güla˘gız, F.K; ¸Sahin, S
18. Sajtos, L; Mitev, A

هر سهم (EPS) به دلیل نوسانات کوتاه مدت با تغییرات فصلی و گردش دارایی‌ها به دلیل همبستگی بالای آن با بازده دارایی‌ها(ROA) استفاده نشده است.

روش آماری

شاخص‌های آماری مورد استفاده در این مطالعه شامل شاخص‌های توصیفی مانند میانگین، انحراف معیار، حداقل، حداکثر، چولگی و کشیدگی بوده است. این شاخص‌ها برای بررسی همگن بودن مقادیر موجود در پایگاه داده انتخاب شدند. نتایج نشان داد که داده‌ها ناهمگن بوده و تفاوت‌های بزرگی با یکدیگر دارند، که ضرورت دسته‌بندی گروه‌ها را توجیه می‌کرد. پس از تشکیل خوشه‌ها، گروه‌های همگن به دست آمدند. برای گروه‌بندی همگن، از تحلیل خوشه‌بندی با دو روش غیر سلسله‌مراتبی K-Mean و K-Medoid استفاده شد. روش‌های خوشه‌بندی می‌توانند سلسله‌مراتبی (ساختار درختی) یا غیر سلسله‌مراتبی باشند. روش سلسله‌مراتبی خوشه‌ها را به تدریج تشکیل می‌دهد و هزینه پردازشی بسیار بالایی دارد زیرا همه اشیاء قبل از هر مرحله خوشه‌بندی با هم مقایسه می‌شوند. در مورد پایگاه‌های داده بزرگ، این امر می‌تواند منجر به صرف زمان اجرای چشمگیری شود (هرمان و همکاران، ۲۰۲۲).

الگوریتم‌های K-Mean و K-Medoid مراکز خوشه‌ها را تا زمانی که تمام نقاط با مراکز مرتبط شوند، تغییر می‌دهند. خوشه‌بندی غیر سلسله‌مراتبی در مقایسه با خوشه‌بندی سلسله‌مراتبی نیاز به صرف زمان کمتری دارد. علاوه بر زمان اجرا، تفسیر و استفاده از نتایج تحلیل خوشه‌بندی سلسله‌مراتبی در نمونه‌های بزرگ‌تر به‌طور چشمگیری پیچیده‌تر است. بنابراین استفاده از روش K-Mean یا K-Medoid پیشنهاد می‌شود (هرمان و همکاران، ۲۰۲۲). گروه‌بندی شرکت‌ها بر اساس نسبت‌های قیمت به ارزش دفتری هر سهم (P/B)، بازده دارایی‌ها (ROA)، بازده حقوق صاحبان سهام (ROE) ، حاشیه سود ناخالص (GPM) و نسبت بدهی به حقوق صاحبان سهام انجام شد. تعداد خوشه‌ها با استفاده از روش آرنج (Elbow)و مشاهده بصری نمودار دندوگرام تعیین شد و همچنین از تحلیل خوشه‌بندی سلسله‌مراتبی با روش Ward و مجذور فاصله اقلیدسی استفاده شد. در نهایت، برای پایگاه داده شرکت‌های بورسی ۵ خوشه انتخاب شد. تحلیل آماری پایگاه‌های داده و ویرایش نمودارها با استفاده از نرم‌افزارهای مایکروسافت اکسل و SPSS و استفاده از قابلیت برنامه‌نویسی نرم‌افزار پایتون انجام شد.

بصورت خلاصه مراحل کار بصورت زیر است:

- ✓ جمع‌آوری داده های شرکت‌های بورسی
- ✓ پیش‌پردازش داده‌ها و استانداردسازی داده‌ها
- ✓ شناسایی داده‌های پرت با استفاده از الگوریتم Isolation Forest و کنترل دوباره با Boxplot
- ✓ انتخاب تعداد خوشه بهینه با استفاده از روش‌های دندوگرام SPSS و الگوریتم Elbow در پایتون
- ✓ خوشه‌بندی با الگوریتم K-mean به وسیله SPSS
- ✓ خوشه‌بندی با الگوریتم K-Medoid با استفاده از پایتون
- ✓ مقایسه دو روش خوشه‌بندی

استانداردسازی و شناسایی داده‌های پرت با استفاده از الگوریتم جنگل ایزوله[1]

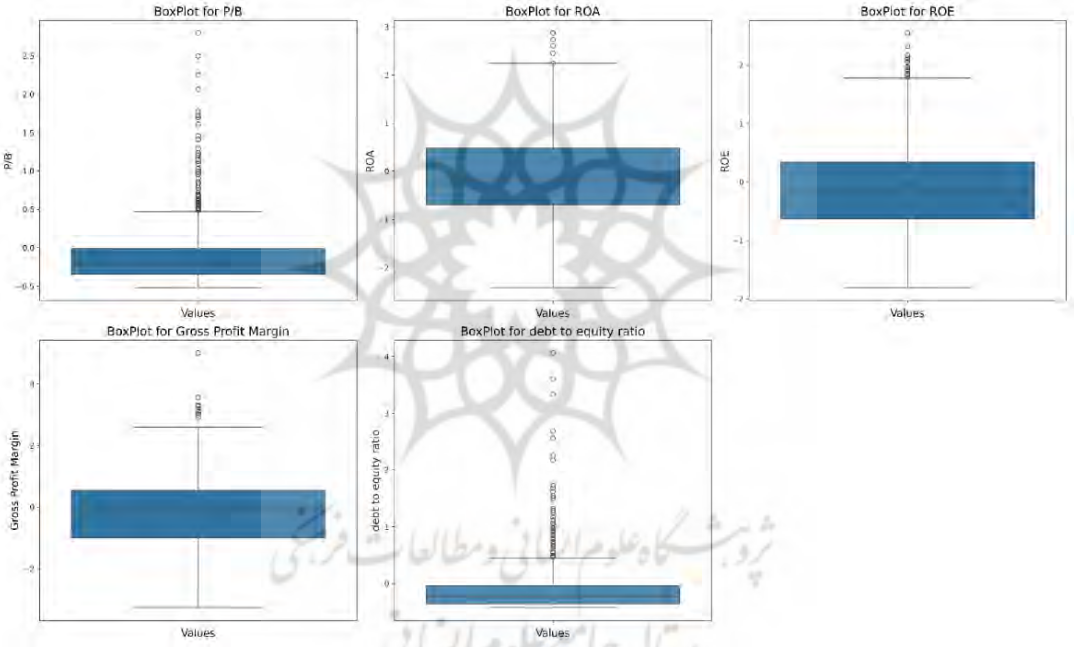
بعد از جمع‌آوری داده‌های خام، به دلیل تفاوت مقیاس ویژگی‌های مختلف، لازم بود داده‌ها استانداردسازی شوند تا الگوریتم‌های یادگیری ماشین بتوانند با داده‌ها در یک مقیاس مشابه کار کنند. برای این کار از استانداردسازی در نرم افزار پایتون (StandardScaler) استفاده شد که مقادیر هر ویژگی را به گونه‌ای تغییر داد که میانگین آن‌ها برابر با صفر و انحراف معیار برابر با یک باشد. این روش به دلیل ویژگی‌های موثر آن در همگن‌سازی داده‌ها برای مدل‌هایی که مبتنی بر فاصله هستند، انتخاب شد. برای بهبود فرآیند شناسایی داده‌های پرت از الگوریتم جنگل ایزوله استفاده شد. این الگوریتم بر مبنای یادگیری ماشین و ساختار درختی عمل می‌کند و به طور خاص برای شناسایی داده‌های پرت طراحی شده است. الگوریتم جنگل ایزوله داده‌هایی که به آسانی و با انجام تعداد کمی تقسیم‌بندی، قابل تفکیک هستند را شناسایی و آن‌ها را به عنوان داده‌های پرت مشخص می‌کند. الگوریتم جنگل ایزوله با ایجاد درخت‌های تصمیم‌گیری به صورت تصادفی، داده‌ها را به قسمت‌های کوچکتر تقسیم می‌کند. داده‌هایی که به سرعت از سایر داده‌ها جدا می‌شوند یعنی در عمق کمتری از درخت قرار می‌گیرند و به همین دلیل نیز داده‌های پرت

1. Isolation Forest

در نظر گرفته می‌شوند. این روش برای داده‌های پیچیده و چندبعدی بسیار مناسب است و می‌تواند داده‌های پرت را با دقت بیشتری شناسایی کند (هستی و همکاران[1]، ۲۰۰۹).

ابزار بصری کنترل پراکندگی داده‌های پرت

نمودار جعبه‌ای یک ابزار گرافیکی در تحلیل داده‌هاست که برای نمایش توزیع داده‌ها و شناسایی داده‌های پرت استفاده می‌شود. این نمودار اطلاعاتی درباره میانه، چارک اول(Q1) ، چارک سوم(Q3) ، دامنه بین چارکی (IQR) و نقاط پرت را نشان می‌دهد. استفاده از Boxplotدر تحلیل داده‌ها به پژوهشگران کمک می‌کند تا پراکندگی و میزان نرمال بودن داده‌ها را به سرعت ارزیابی کنند. همچنین، این ابزار برای مقایسه توزیع داده‌ها در گروه‌های مختلف مفید است (مک‌کیننی[2]، ۲۰۱۷).



این نمودارها وضعیت پراکندگی داده‌های مالی استانداردشده را برای متغیرهای کلیدی نشان می‌دهند. متمرکز بودن داده‌ها در محدوده مرکزی هر نمودار نشان از تأثیر مثبت استانداردسازی

1. Hastie et al
2. McKinney

دارد. در نمودارهای جعبه‌ای شاخص‌های آماری با وضوح مشخص هستند و نقاط میانه و چارک‌ها به‌خوبی نمایش داده شده‌اند. هرچند نقاط پرت دیده می‌شود، تعداد آن‌ها نسبت به کل داده‌ها محدود است و داده‌های مرکزی همچنان بدون نوسان شدید قابل بررسی‌اند. نمودارها امکان شناسایی الگوها و بررسی تأثیر متغیرها را فراهم می‌کنند و به تحلیل‌های آماری پیشرفته و پیش‌بینی‌ها کمک می‌کنند. این وضعیت مبنای مطمئنی برای تحلیل‌های مالی به شمار می‌آید. همان‌طور که در تحلیل نمودار BoxPlot مشاهده شد، روش جنگل ایزوله توانسته است داده‌های پرت را شناسایی و حذف کند. این نشان‌دهنده دقت بالاتر این روش در شناسایی داده‌های غیرمعمول و پرت است. این الگوریتم به دلیل ساختار درختی و الگوریتم منحصربه‌فرد خود، می‌تواند داده‌های پرت پیچیده و چندبعدی را بهتر شناسایی کند که این ویژگی برای داده‌های مالی و اقتصادی با ویژگی‌های متنوع و نوسانات بالا بسیار مفید است.

روش آرنج[1] و نمودار دندوگرام برای تعیین تعداد بهینه خوشه‌ها

یک تکنیک گرافیکی برای تعیین تعداد بهینه خوشه‌ها در الگوریتم‌های خوشه‌بندی، الگوریتم‌های K-Means و K-Medoid است. این روش بر اساس تغییرات مجموع مربعات خطا درون خوشه‌ها (WCSS[2]) برای تعداد متفاوتی از خوشه‌ها عمل می‌کند. هدف اصلی این روش یافتن نقطه‌ای است که پس از آن افزایش تعداد خوشه‌ها تأثیر چشمگیری در کاهش مجموع خطا ندارد و به عنوان نقطه بهینه انتخاب می‌شود (بریمن[3]، ۲۰۰۱). پس از محاسبه WCSS برای تعداد مختلفی از خوشه‌ها، نمودار Elbow رسم می‌شود که محور افقی تعداد خوشه‌ها و محور عمودی مقدار WCSS است. در این نمودار، نقطه‌ای که نمودار از کاهش شدید به کاهش ملایم تغییر می‌کند، به عنوان نقطه آرنج[4] شناخته می‌شود (هان و همکاران[5]، ۲۰۱۱). نقطه آرنج نشان‌دهنده تعداد بهینه خوشه‌ها است. پس از این نقطه، افزایش تعداد خوشه‌ها تأثیر زیادی در کاهش WCSS ندارد و نشان می‌دهد که خوشه‌بندی بهینه حاصل شده است (هستی و همکاران[6]، ۲۰۰۹).

---

1. Elbow Method
2. Within-Cluster Sum of Squares
3. Breiman
4. Elbow Point
5. Han et al
6. Hastie et al

در الگوریتم K-Means، مرکز هر خوشه به عنوان میانگین نقاط خوشه محاسبه می‌شود و WCSSمجموع مربعات فاصله‌ها از این مرکز است. در حالی که در الگوریتم K-Medoid، مرکز خوشه (مدوئید) یکی از نقاط واقعی داده‌هاست و مجموع فاصله‌ها به جای WCSS به عنوان ابزار اندازه گیری استفاده می‌شود. بنابراین، در K-Medoid از مجموع فاصله‌های معمولی برای ترسیم نمودار آرنج استفاده می‌شود، که این الگوریتم نسبت به داده‌های پرت مقاوم‌تر عمل می‌کند (کافمن و روسو[1]، ۱۹۹۰).

یافته‌های پژوهش

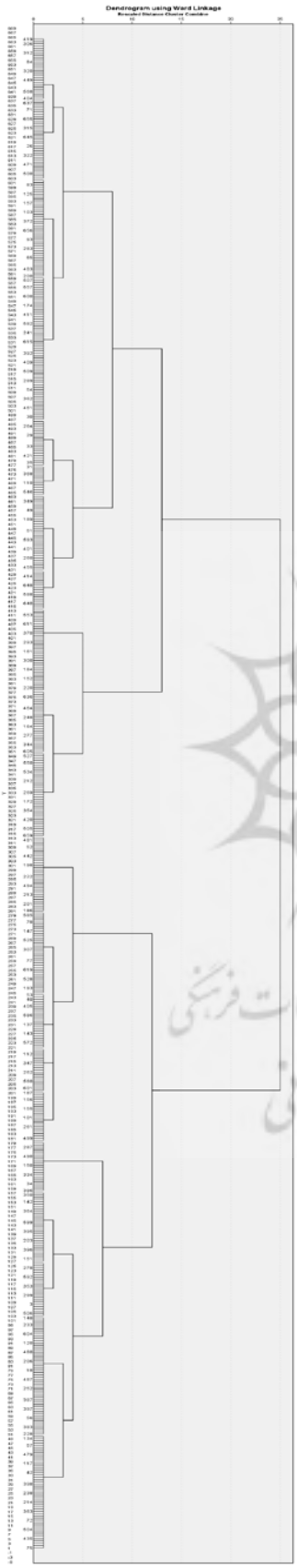
در الگوریتم K-Means، مرکز هر خوشه به عنوان میانگین نقاط خوشه محاسبه می‌شود و WCSSمجموع مربعات فاصله‌ها از این مرکز است. در حالی که در الگوریتم K-Medoid، مرکز خوشه (مدوئید) یکی از نقاط واقعی داده‌هاست و مجموع فاصله‌ها به جای WCSS به عنوان ابزار اندازه گیری استفاده می‌شود. بنابراین، در K-Medoid از مجموع فاصله‌های معمولی برای ترسیم نمودار آرنج استفاده می‌شود، که این الگوریتم نسبت به داده‌های پرت مقاوم‌تر عمل می‌کند (کافمن و روسو[2]، ۱۹۹۰).

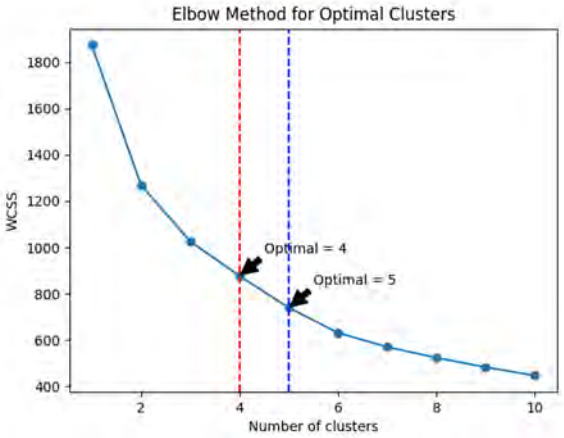
خوشه‌بندی به روش K-Mean

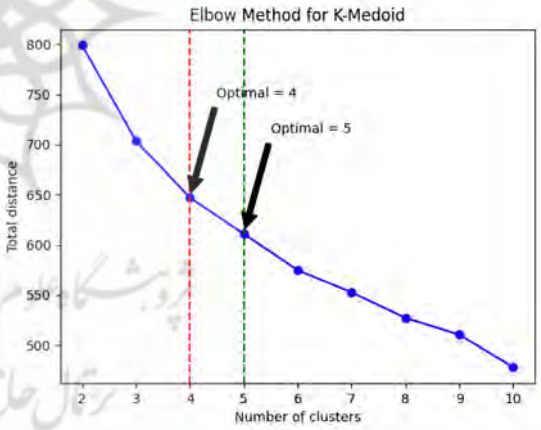
روش K-mean یکی از الگوریتم‌های متداول در حوزه یادگیری ماشین برای خوشه‌بندی داده‌ها است که هدف آن تقسیم داده‌ها به K خوشه مجزا با کمترین میزان خطای داخلی است. در این روش، هر نقطه داده به نزدیک‌ترین مرکز خوشه اختصاص می‌یابد و مراکز خوشه‌ها تا رسیدن به حالت پایدار به‌روزرسانی می‌شوند. این الگوریتم به دلیل سادگی، سرعت در پیاده‌سازی و کاربرد وسیع در تحلیل داده‌های بزرگ همچنان یکی از روش‌های پرکاربرد در مسائل دسته‌بندی و خوشه‌بندی به‌شمار می‌آید. پژوهش‌های اخیر نشان می‌دهند که با بهبود الگوریتم‌های K-mean، دقت و کارایی آن در مسائل پیچیده افزایش یافته و کاربردهای آن در تحلیل داده‌های پیچیده و چندبعدی همچنان رشد یافته است (علی و همکاران[3]، ۲۰۲۲).

---

1. Kaufman & Rousseeuw
2. Kaufman & Rousseeuw
3. Ali et al







- تعداد خوشه های بهینه با توجه به نمودار دندوگرام و الگوریتم Elbow در هر دو روش K-mean و K-medoid ، در حدود ۵ خوشه می باشد. Elbow در شکل معلوم و مشخص نیست

تابع هدف در فرمول k-mean که به دنبال کمینه‌سازی مجموع مربعات فاصله‌ها است، به شرح زیر تعریف می‌شود:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

J مجموع مربعات خطا، K تعداد خوشه‌ها، C_i مجموعه داده‌های متعلق به خوشه i، مرکز خوشه i μ_i و $\|x - \mu_i\|^2$ فاصله اقلیدسی بین داده X و مرکز خوشه μ_i است.

جدول ۳. خوشه‌بندی با الگوریتم K-Mean

شرح	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴	خوشه ۵
قیمت به ارزش دفتری هر سهم	۰,۱۴۸۱۱۸۹۳	-۰,۲۴۳۶۲۶۸	-۰,۰۰۲۲۹۵۳	-۰,۱۴۷۱۹۳۵۹۴	۰,۰۸۶۲۳۷۹۱۶
بازده دارایی‌ها	-۰,۸۹۵۵۸۳۷	-۰,۵۶۶۷۳۵۸	-۰,۰۱۹۷۶۴۴	-۱,۳۳۰۵۳۸۳۲	۱,۱۲۵۶۸۷۵۳۷
بازده حقوق صاحبان سهام	-۰,۱۵۰۲۵۶۷	-۰,۶۵۱۴۱۶۱	-۰,۳۲۲۶۷۲۸	-۱,۰۷۴۶۷۰۱	۰,۹۴۵۴۲۵۸۹۷
حاشیه سود ناخالص	-۰,۴۲۴۱۵۷۸	-۰,۷۸۳۱۷۶۱	۲,۵۴۹۳۳۶۶۵	۰,۳۳۴۹۶۶۲۸۹	۰,۲۲۱۵۷۴۸۶۸
بدهی به حقوق صاحبان سهام	۱,۳۱۳۱۷۵۱۸	-۰,۲۴۸۲۶۱۹	-۰,۲۹۱۸۴۳۶	-۰,۱۴۸۴۶۶۶۴۷	-۰,۱۵۳۱۵۲۰۰۵

این جدول نشان‌دهنده مراکز نهایی خوشه‌بندی داده‌ها با الگوریتم K-Mean برای پنج خوشه مختلف است. تحلیل کلی این جدول به شرح زیر است: به طور خلاصه این جدول نشان‌دهنده مراکز نهایی خوشه‌بندی با الگوریتم K-Mean برای پنج خوشه است. در ویژگی قیمت به ارزش دفتری هر سهم(P/B) ، خوشه‌های ۱ و ۵ مقادیر مثبت دارند که نشان‌دهنده عملکرد بالاتر از میانگین است، در حالی که سایر خوشه‌ها مقادیر منفی دارند. برای بازده دارایی‌ها(ROA)، خوشه ۵ بهترین عملکرد مثبت را دارد و دیگر خوشه‌ها مقادیر منفی دارند که نشان‌دهنده بازدهی پایین‌تر از میانگین است. بازده حقوق صاحبان سهام (ROE) نیز در خوشه ۵ بیشترین مقدار مثبت را دارد، در حالی که سایر خوشه‌ها مقادیر منفی دارند. حاشیه سود

ناخالص در خوشه ۳ بیشترین مقدار مثبت را نشان می‌دهد و سایر خوشه‌ها مقادیر منفی دارند. نسبت بدهی به حقوق صاحبان سهام در خوشه ۱ بیشترین مقدار مثبت را دارد که نشان‌دهنده نسبت بدهی بالاتر است، در حالی که سایر خوشه‌ها مقادیر منفی دارند. این تحلیل به ارزیابی ریسک و بازده شرکت‌ها کمک می‌کند و تفاوت‌های عملکردی بین خوشه‌ها را مشخص می‌سازد.

خوشه ۱ شامل شرکت‌هایی با قیمت مناسب به ارزش دفتری و بازده دارایی‌ها و حقوق صاحبان سهام در سطح متوسط است. حاشیه سود پایین و نسبت بدهی کم، ریسک مالی این خوشه را پایین نگه می‌دارد. خوشه ۲ شامل شرکت‌هایی با بازده دارایی بالا و بازده حقوق صاحبان سهام چشمگیر است. با این حال، حاشیه سود ناخالص منفی نشان‌دهنده وجود مشکلات در سودآوری عملیاتی است. بدهی کم و منفی، ساختار مالی این خوشه را مطمئن‌تر می‌کند.خوشه ۳ شرکت‌هایی با حاشیه سود ناخالص بسیار بالا اما بازده منفی دارایی و حقوق صاحبان سهام را نشان می‌دهد. به‌رغم سود ناخالص بالا، بازده عملیاتی ضعیف و عملکرد مالی منفی می‌تواند نشان‌دهنده ناکارآمدی در مدیریت باشد. خوشه ۴ شامل شرکت‌هایی با بازده منفی شدید در دارایی‌ها و حقوق صاحبان سهام و قیمت بسیار پایین نسبت به ارزش دفتری است. حاشیه سود ناخالص متوسط مثبت است، اما بدهی بسیار بالای شرکت‌ها، ساختار مالی پرریسک این خوشه را نشان می‌دهد.خوشه ۵ شامل شرکت‌هایی با بازده دارایی و حقوق صاحبان سهام بسیار بالا و سودآوری مناسب است. با این حال، نسبت بدهی بسیار بالاست که نشان‌دهنده ساختار مالی پرریسک این شرکت‌ها است. به طور کلی می‌توان بیان کرد خوشه ۵ به عنوان خوشه با بهترین عملکرد مالی و سودآوری شناخته می‌شود، در حالی که خوشه‌های ۱ و ۲ به دلیل مقادیر منفی در بازده دارایی‌ها و بازده حقوق صاحبان سهام نشان‌دهنده عملکرد ضعیف‌تر هستند. خوشه ۳ نشان‌دهنده شرکت‌هایی با حاشیه سود بالا و بازده متوسط است و خوشه ۴ شامل شرکت‌هایی با نسبت‌های منفی و چالش‌های مالی است. به صورت خلاصه: خوشه ۱ و ۲ : کم‌ریسک‌تر و عملکرد مطلوب با بدهی پایین؛ خوشه ۳ : سود ناخالص بالا اما مشکلات جدی در بازده عملکرد؛ خوشه ۴ : ضعیف‌ترین خوشه با بازده منفی شدید و بدهی بالا؛ خوشه ۵ : بهترین عملکرد در بازده، اما ریسک مالی بالا به دلیل بدهی زیاد.

جدول ۴. تحلیل واریانس ANOVA روش K-Mean

شاخص	میانگین مربع خوشه	درجه آزادی (خوشه)	میانگین مربع خطا	درجه آزادی (خطا)	آماره F	سطح معناداری
نسبت قیمت به درآمد (P/B)	۳,۳۰۹	۴	۰,۱۵	۶۵۹	۲۲,۰۴	<.001
بازده دارایی‌ها (ROA)	۷۳,۶۵۹	۴	۰,۲۵۴	۶۵۹	۲۹۰,۶	<.001
حاشیه سود ناخالص	۱۰۱,۸۷۳	۴	۰,۳۲۵	۶۵۹	۳۱۳,۸	<.001
بازده حقوق صاحبان سهام (ROE)	۵۷,۵۸۱	۴	۰,۱۹۵	۶۵۹	۲۹۴,۶	<.001
نسبت بدهی به حقوق صاحبان سهام	۲۳,۲۸۴	۴	۰,۰۹	۶۵۹	۲۵۷,۸	<.001

این جدول نتایج آزمون ANOVA یک‌طرفه را نشان می‌دهد که تفاوت معناداری بین خوشه‌ها در متغیرهای مالی بررسی‌شده شامل قیمت به ارزش دفتری هر سهم (P/B)، بازده دارایی‌ها (ROA)، حاشیه سود ناخالص (GP)، بازده حقوق صاحبان سهام(ROE) و نسبت بدهی به حقوق صاحبان سهام را تأیید می‌کند. مقادیر F بسیار بالا و Sig کمتر از ۰,۰۰۱ در تمام متغیرها، تفاوت میانگین‌ها را از نظر آماری معنادار می‌کند. متغیرهایی مانند حاشیه سود ناخالص GP و بازده دارایی‌ها (ROA) بیشترین نقش را در تفکیک خوشه‌ها داشته‌اند. با توجه به اینکه خوشه‌ها برای بیشینه‌سازی تفاوت‌ها انتخاب شده‌اند، نتیجه نشان می‌دهد خوشه‌بندی به‌خوبی و براساس تفاوت متغیرهای مالی انجام شده است.

خوشه‌بندی‌ها دقیق انجام شده و تفاوت معنادار میان خوشه‌ها نشان می‌دهد که ویژگی‌های هر خوشه به خوبی تعریف شده‌اند. می‌توان از این نتایج برای بهینه‌سازی تصمیم‌گیری‌های مدیریتی استفاده کرد. شاخص‌هایی که بالاترین مقدار F را دارند (مانند حاشیه سود ناخالص) می‌توانند اولویت بیشتری در تحلیل‌های بعدی داشته باشند.

جدول ۵. تعداد شرکت‌های هر خوشه در روش K-Mean

تعداد شرکت‌های هر خوشه	
خوشه ۱	۲۵
خوشه ۲	۲۴۴
خوشه ۳	۳۲
خوشه ۴	۲۲۶
خوشه ۵	۱۳۷
معتبر	۶۶۴
ناموجود	۰

جدول بالا، توزیع تعداد شرکت‌ها در هر خوشه را نشان می‌دهد. خوشه ۲ با ۲۴۴ شرکت بزرگ‌ترین خوشه است در حالی که خوشه ۱ با تنها ۲۵ شرکت کوچک‌ترین خوشه محسوب می‌شود. این توزیع نشان‌دهنده تراکم شرکت‌ها و چگونگی تقسیم‌بندی آن‌ها در خوشه‌های مختلف است.

جدول ۶. فاصله بین مراکز نهایی خوشه ها در روش K-Mean

خوشه	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴	خوشه ۵
خوشه ۱	۰	۱,۷۵۵	۳,۴۹۸	۱,۸۳۹	۲,۸۰۴
خوشه ۲	۱,۷۵۵	۰	۳,۴۰۲	۱,۳۲۴	۲,۵۵۸
خوشه ۳	۳,۴۹۸	۳,۴۰۲	۰	۲,۲۳۷	۲,۸۹۳
خوشه ۴	۱,۸۳۹	۱,۳۲۴	۲,۲۳۷	۰	۱,۶۶۲
خوشه ۵	۲,۸۰۴	۲,۵۵۸	۲,۸۹۳	۱,۶۶۲	۰

جدول فاصله‌های بین مراکز نهایی خوشه‌ها اطلاعات مهمی درباره تمایز یا شباهت خوشه‌ها ارائه می‌دهد. فاصله‌های بیشتر نشان‌دهنده تفاوت‌های روشن‌تر میان خوشه‌ها است، در حالی که فاصله‌های کمتر نشان‌دهنده شباهت بیشتر بین خوشه‌ها است. به عنوان مثال، خوشه‌های ۲ و ۴ نزدیک‌ترین فاصله را دارند که بیانگر شباهت بالای ویژگی‌های این دو

خوشه است، در حالی که خوشه‌های ۱ و ۳ با بیشترین فاصله، تمایز بیشتری با هم دارند. این تحلیل به ارزیابی کیفیت خوشه‌بندی کمک می‌کند و می‌تواند به عنوان مبنایی برای تصمیم‌گیری‌های بعدی، مانند ادغام خوشه‌های نزدیک یا تحلیل دقیق‌تر خوشه‌های متفاوت، مورد استفاده قرار گیرد. این داده‌ها برای بهبود مدل‌های خوشه‌بندی و ارزیابی اثربخشی الگوریتم‌های مورد استفاده ضروری هستند.

خوشه‌بندی به روش K-Medoid

روش K-Medoid یکی از الگوریتم‌های خوشه‌بندی است که مشابه روش K-Means عمل می‌کند، با این تفاوت که به جای استفاده از میانگین نقاط برای تعیین مرکز هر خوشه، از یک نمونه واقعی به عنوان مرکز استفاده می‌کند. این ویژگی باعث می‌شود K-Medoid در برابر داده‌های پرت و دارای اغتشاش مقاوم‌تر باشد.

فرض کنید مجموعه‌ای از نقاط داده به صورت $\{{}_{n}x_1, x_2, \ldots, x\} = X$ در اختیار داریم، K-Medoid به صورت زیر تعریف می‌شود:

$$\arg\min_{y \in X} \sum_{i=1}^{n} d(y, x_i) = \text{medoid}x$$

که در آن $d(y, x_i)$ فاصله بین نقاط y و $\mathbf{X}$ است.

جدول ۷. خوشه‌بندی با الگوریتم K-Medoid

روش K-medoid	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴	خوشه ۵
قیمت به ارزش دفتری هر سهم	-۰,۱۰۴۲۶۰۳۸۷	-۰,۲۴۹۱۵۰۲۰۱	۰,۱۵۵۴۲۸۵۲۱	-۰,۲۵۶۱۷۵۸۹۶	-۰,۰۰۷۰۸۹۱۴۷
بازده دارایی‌ها	-۰,۲۶۳۰۲۹۷۳۶	-۰,۸۱۵۱۲۷۳۴۷	۱,۱۳۳۳۸۹۴۹۶	۰,۷۵۱۰۹۳۳۲۷	-۰,۱۳۴۶۱۰۸۵۷
بازده حقوق صاحبان سهام	-۰,۳۶۷۸۲۲۲۳۴	-۰,۷۳۷۲۶۴۱۰۱	۱,۰۷۹۰۳۴۰۴۷	۰,۰۸۲۱۲۸۲۸۹	۰,۰۸۳۶۳۴۷۹۲
حاشیه سود ناخالص	۱,۴۴۵۸۹۱۵۱۷	-۰,۷۰۵۷۲۴۹۹۳	۰,۷۲۴۲۵۶۱۴۱	-۰,۹۵۱۱۵۳۹۹۷	۰,۰۳۹۶۳۶۷۳۱
بدهی به حقوق صاحبان سهام	-۰,۲۲۸۳۷۰۶۴۲	-۰,۰۳۰۴۰۶۸۹۴	-۰,۱۱۹۹۲۸۰۳۵	-۰,۳۵۰۲۷۶۲۳۴	۰,۰۳۴۵۲۷۷۵۱

در خوشه‌بندی با روش K-Medoid، تفاوت‌های معناداری بین ویژگی‌های مالی خوشه‌ها دیده می‌شود. خوشه ۱ شامل شرکت‌هایی با بازده حقوق صاحبان سهام (ROE) و بازده دارایی‌ها (ROA) پایین و نسبت بدهی بالا است که نمایانگر ریسک مالی بالاتر و کارایی کمتر است. خوشه ۲ بهترین عملکرد مالی را با بالاترین ROE و ROA و حاشیه سود قوی‌تر دارد و شرکت‌های این خوشه از نظر مالی دارای جایگاه بهتری هستند. خوشه ۳ عملکرد مالی میانه‌ای دارد و تعادل بین ریسک و بازده دیده می‌شود. خوشه ۴ دارای بازده پایین اما نسبت بدهی کمتر است که نشان‌دهنده مدیریت مالی محتاطانه است. خوشه ۵ مشابه خوشه ۱ است اما عملکرد مالی ضعیف‌تری دارد. خوشه ۲ از لحاظ پارامترهای مالی، بهترین عملکرد را دارد و شامل شرکت‌هایی با بازده بالا و حاشیه سود مناسب است.

خوشه ۱ شامل شرکت‌هایی با سودآوری متوسط، اما بدهی بالا و نسبت قیمت به ارزش دفتری پایین است. این شرکت‌ها ممکن است در مراحل اولیه رشد یا دارای ساختار مالی پرریسک باشند. خوشه ۲ شامل شرکت‌هایی با بازدهی بهتر، سودآوری مناسب، و ساختار مالی متعادل‌تر است. شرکت‌ها ممکن است در وضعیت تثبیت و رشد باشند. خوشه ۳ شامل شرکت‌هایی با سودآوری متوسط، اما بدهی بالا و قیمت به ارزش دفتری پایین است. ریسک مالی این شرکت‌ها به طور کلی بالا است. خوشه ۴ شامل شرکت‌هایی با بازدهی بالا، سودآوری قوی و ساختار مالی کم‌ریسک‌تر است. این شرکت‌ها ممکن است رهبران بازار یا شرکت‌های پایدار باشند. خوشه ۵ شامل شرکت‌هایی با ساختار مالی بسیار پرریسک، بازدهی منفی، و سودآوری ضعیف است. این شرکت‌ها ممکن است در وضعیت بحران یا زیان‌دهی باشند.

جدول ۸. مشخصات خوشه‌ها - خوشه‌بندی به روش Kmedoid

خوشه‌بندی ۱	شرکت‌های با سودآوری متوسط اما بدهی بالا
خوشه‌بندی ۲	شرکت‌های با بازدهی مناسب، سودآوری بالا، و بدهی متعادل
خوشه‌بندی ۳	شرکت‌های با سودآوری متوسط و بدهی بالا
خوشه‌بندی ۴	شرکت‌های با بازدهی بالا، سودآوری قوی، و بدهی کم (بهترین وضعیت)
خوشه‌بندی ۵	شرکت‌های پرریسک با بازدهی منفی و بدهی زیاد (بدترین وضعیت)

جدول ۹. تعداد شرکت‌های هر خوشه-روش K-Medoid

تعداد شرکت‌های هر خوشه	
خوشه ۱	۹۵
خوشه ۲	۲۰۸
خوشه ۳	۱۰۵
خوشه ۴	۸۵
خوشه ۵	۱۷۱
معتبر	۶۶۴
ناموجود	۰

این جدول نشان می‌دهد که خوشه ۲ با ۲۰۸ شرکت بزرگ‌ترین خوشه است و شرکت‌هایی با ویژگی‌های مالی عمومی‌تر را شامل می‌شود. خوشه ۵ با ۱۷۱ شرکت در رتبه دوم قرار دارد و خوشه‌های کوچک‌تر (۱، ۳ و ۴) به ترتیب با ۹۵، ۱۰۵، و ۸۵ شرکت نشان‌دهنده تمایز بیشتر در ویژگی‌های مالی هستند. تعداد کل شرکت‌های معتبر ۶۶۴ شرکت است. توزیع شرکت‌ها تأیید می‌کند که الگوریتم خوشه‌بندی به‌خوبی داده‌ها را بر اساس شباهت‌ها تفکیک کرده است.

نتایج آزمون ANOVA نشان می‌دهد که تفاوت‌های آماری معناداری بین خوشه‌ها وجود دارد، یعنی متغیرهای مالی که برای خوشه‌بندی استفاده شده‌اند، به خوبی توانسته‌اند گروه‌ها را از هم جدا کنند. متغیرهایی مثل بازده دارایی‌ها(ROA)، بازده حقوق صاحبان سهام(ROE) و حاشیه سود ناخالص (Gross Profit Margin) بیشترین تأثیر را در این جداسازی داشته‌اند. متغیر نسبت قیمت به ارزش دفتری (P/B) هم تأثیر قابل توجهی دارد، اما کمتر از سه متغیر اول است.

جدول ۱۰. تحلیل واریانس ANOVA- روش K-Medoid

شاخص	آماره F	سطح معناداری
قیمت به ارزش دفتری هر سهم	۲۵,۴۷۱۵	<.001
بازده دارایی‌ها	۴۰۸,۰۰۹۴	<.001
بازده حقوق صاحبان سهام	۳۵۶,۳۹۳۳	<.001
حاشیه سود ناخالص	۳۸۲,۵۳۹۱	<.001
بدهی به حقوق صاحبان سهام	۱۲,۸۲۸۱	<.001

نسبت بدهی به حقوق صاحبان سهام[1] (D/E) نیز تفاوت‌هایی بین خوشه‌ها نشان داده، اما تأثیر آن نسبت به بقیه متغیرها کمتر است. به طور خلاصه، این آزمون نشان می‌دهد که این متغیرها برای خوشه‌بندی داده‌ها مناسب بوده‌اند.

مقایسه خوشه‌بندی به دو روش K-Mean و K-Medoid

نتایج کلی مقایسه روش‌های خوشه‌بندی K-Mean و K-Medoid بر اساس جدول‌های ارائه شده، نشان می‌دهد که هر دو روش قادر به تفکیک شرکت‌ها به خوشه‌های مجزا بر اساس معیارهای مالی هستند. روش K-Mean به دلیل سرعت و کارایی بالا در پردازش داده‌های بزرگ، نتایج مشابهی در تعیین خوشه‌ها ارائه داده است. با این حال، روش K-Medoid در برخورد با داده‌های پرت، دقیق‌تر عمل کرده و در این تحلیل خوشه‌بندی‌ها را با انحراف کمتری انجام داده است. از منظر ریسک مالی، خوشه‌های با مقادیر بالاتر در متغیرهایی همچون بازده دارایی‌ها (ROA) و بازده حقوق صاحبان سهام (ROE)، به‌عنوان خوشه‌هایی با عملکرد مالی بهتر شناخته می‌شوند. همچنین، نسبت بدهی به حقوق صاحبان سهام در برخی خوشه‌ها نشان‌دهنده سطح ریسک مالی متفاوت است، به نحوی که خوشه‌های با مقدار کمتر این نسبت، ریسک کمتری را نشان می‌دهند. از نظر تعداد شرکت‌های موجود در خوشه‌ها، روش K-Medoid توزیع متوازن‌تری را بین خوشه‌ها ایجاد کرده است. این موضوع نشان‌دهنده عملکرد مناسب‌تر این روش در روبرویی با داده‌های متنوع و پیچیده است. همچنین، فاصله بین مراکز خوشه‌ها در روش K-Medoid نسبت به K-Mean کمتر بوده که نشان‌دهنده همگنی بیشتر در خوشه‌ها است. در نهایت، اگرچه روش K-Mean به دلیل سرعت بالاتر برای داده‌های بزرگ مناسب است اما روش K-Medoid به دلیل دقت بالاتر و مقاومت در برابر داده‌های پرت، برای تحلیل‌های حساس‌تر و دقیق‌تر پیشنهاد می‌شود. از نظر پارامترهای بازدهی و سودآوری، روش K-Mean در خوشه‌بندی دقیق‌تر بازدهی شرکت‌ها موفق‌تر عمل کرده درحالی که K-Medoid برای تفکیک شرکت‌ها با داده‌های پیچیده و متنوع عملکرد بهتری داشته است. بنابراین، انتخاب روش بسته به نوع داده‌ها و هدف از خوشه‌بندی متفاوت خواهد بود.

1. Debt to Equity Ratio

جدول ۱۱. مقایسه خوشه‌بندی به دو روش K-mean و K-medoid

خوشه	خوشه‌بندی K-mean	خوشه‌بندی K-medoids
خوشه ۱	کم ریسک و باثبات، بازده متوسط، بدهی کم و وضعیت مطلوب مالی	سودآوری متوسط و کم‌ریسک، بازده دارایی‌ها قابل قبول و بدهی کنترل‌شده
خوشه ۲	ریسک بالا با بازده متغیر، بازده مثبت و بدهی ناپایدار	ریسک مالی بالا، بازده منفی و سودآوری پایین
خوشه ۳	نوسان بالا در سودآوری، حاشیه سود بسیار بالا، بازده مالی ضعیف	سودآوری بالا و پایدار، بازده مطلوب و تعادل مالی
خوشه ۴	ضعیف‌ترین وضعیت مالی، بازده منفی و بدهی بالا	سودآوری منفی با ریسک کمتر، بهبود اندک بازده و کاهش بدهی
خوشه ۵	سودآوری بالا با بدهی زیاد، بازده بالا و ریسک مالی زیاد	ریسک مالی متوسط، سودآوری پایین و وضعیت بدهی کنترل‌شده

بحث و نتیجه‌گیری

افزودن شاخص‌های جدید به مدل خوشه‌بندی برخلاف پژوهش هرمان و همکاران (۲۰۲۲) که تنها بر ۳ شاخص مالی (ROA، ROE، ROS) تمرکز داشت، این پژوهش از ۵ شاخص مالی دیگر شامل نسبت بدهی به حقوق صاحبان سهام (D/E) و نسبت قیمت به ارزش دفتری (P/B) استفاده کرده است، همچنین بازار بورس اوراق بهادار ایران به‌عنوان نمونه موردی انتخاب شده است .این امر باعث شده که خوشه‌بندی‌ها با دقت بیشتری صورت بگیرد و تحلیل یافته‌ها و مقایسه با مطالعات پیشین نتایج این پژوهش نشان داد که روش K-Medoid خوشه‌هایی متعادل‌تر ایجاد می‌کند، درحالی‌که K-Means اگرچه سرعت بیشتری دارد اما حساسیت بالاتری به داده‌های پرت نشان می‌دهد. این یافته‌ها با مطالعه هرمان و همکاران (۲۰۲۲)، همخوانی دارد که نشان داد K-Medoid در داده‌های مالی عملکرد پایدارتری دارد و خوشه‌بندی متوازن‌تری ارائه می‌دهد. در مقایسه با مطالعه (ول موروگان و سنتانام، ۲۰۱۰)، نتایج این پژوهش نیز تأیید می‌کند که روش K-Medoid به دلیل انتخاب نقاط واقعی داده به‌عنوان مرکز خوشه، مقاومت بیشتری در برابر داده‌های پرت دارد، درحالی‌که K-Means مستعد ایجاد خوشه‌های با اندازه نامتعادل است.

همچنین، پژوهش آرورا و همکاران (۲۰۱۶) نشان داد که K-Means در تحلیل داده‌های مالی، سرعت بالاتری دارد اما نتایج آن در حضور داده‌های پرت قابل اطمینان نیست. نتایج این مقاله نیز نشان می‌دهد که K-Medoid توانسته است در تحلیل شرکت‌های بورسی ایران،

خوشه‌های با پایداری مالی مشخص ایجاد کند. مطالعات قبلی همچون کای و همکاران[1] (۲۰۱۲) نیز تأکید کرده‌اند که افزودن شاخص‌های مالی نظیر نسبت بدهی به حقوق صاحبان سهام (D/E) و نسبت قیمت به ارزش دفتری (P/B) می‌تواند باعث بهبود خوشه‌بندی شرکت‌های بورسی شود. این یافته با مقاله حاضر همخوانی دارد که نشان داد افزودن این شاخص‌ها موجب بهبود تفکیک خوشه‌های مالی شد. در مقایسه با پژوهش ایکوتون و همکاران[2]، (۲۰۲۱) این مطالعه نشان می‌دهد که K-Medoid برای مجموعه‌های مالی ایران که دارای داده‌های با نوسانات بالا هستند، بهتر عمل می‌کند. درحالی‌که آن پژوهش اشاره داشت که در بازارهای پایدار، K-Means به دلیل سادگی پردازش، انتخاب مناسب‌تری است، اما در این مطالعه مشخص شد که به دلیل وجود داده‌های پرت و ناپایداری اقتصادی، K-Medoid نتایج دقیق‌تری ارائه می‌دهد.

از کاربردهای خوشه‌بندی‌های مورد استفاده در این پژوهش و نتایج آن در حوزه مالی و سرمایه‌گذاری می توان موارد زیر را بیان کرد. کمک به سرمایه‌گذاران در شناسایی شرکت‌های با ریسک کمتر همچنین سرمایه‌گذاران می‌توانند از خوشه‌بندی برای شناسایی شرکت‌هایی با سودآوری پایدار و سطح بدهی مطلوب استفاده کنند. این مدل می‌تواند به مدیران شرکت‌های بورسی در بهبود استراتژی‌های مالی و مدیریت بدهی‌ها کمک کند.

مطالعه و پژوهش‌های آینده

لازم به بیان است که این مطالعه، مانند هر مطالعه دیگر، محدودیت‌هایی دارد و می‌تواند در جهت‌های خاصی ادامه و بهبود یابد، درحالی‌که عناصر اصیل آن که توسط سایر متخصصان بی‌چون و چرا باقی مانده‌اند، ارائه می‌شود. به‌عنوان محدودیت‌های این مطالعه می‌توان به موارد زیر اشاره کرد:

۱. تعداد شاخص‌های مالی استفاده‌شده در پژوهش (تنها شاخص‌های مالی رایج انتخاب شده‌اند (نسبت‌های قیمت به ارزش دفتری هر سهم (P/B)، بازده دارایی‌ها (ROA)، بازده حقوق صاحبان سهام (ROE) ، حاشیه سود ناخالص و نسبت بدهی به حقوق صاحبان سهام) هستند.

1. Cai et al
2. Ikotun et al

۲. دوره زمانی تحلیل صورت‌های مالی شرکت‌هایی که تحلیل بر اساس آن‌ها انجام شده است، تنها شامل ۱۰ سال متوالی است.

۳. انتخاب تنها بخش تولیدی و بازرگانی به‌عنوان زمینه فعالیت شرکت‌های تحلیل شده بدین معنی که شرکت‌های موجود در پایگاه داده تنها شرکت‌هایی هستند که فعالیت اصلی آن‌ها تولیدی و بازرگانی است.

۴. انتخاب چندین زمینه فعالیت افزون بر تولید، خدمات. استفاده از روش‌ها و تکنیک‌های بیشتر در تحلیل داده‌های توصیفی، تحلیل عاملی، تشکیل گروه‌های همگن و تحلیل عملکرد مالی (معیارهای ارزیابی عملکرد).

۵. ساخت یک شاخص ترکیبی برای توصیف عملکرد مالی (شامل ROS، ROA و ROE در کنار شاخص‌های دیگر مانند نقدینگی، نسبت جاری، دوره وصول مطالبات، گردش حسابهای پرداختنی و غیره).

۶. استفاده از روش‌های دیگر خوشه‌بندی آماری و مقایسه با نتایج حاصل از خوشه‌بندی‌های انجام شده در این پژوهش.

References

Afrimayani; Devianto, D. (2021). The time series clustering of stock price in LQ45 index and its financial performance analysis. *J. Phys .Conf. Ser.* 1943, 1–8.

Akhtar, M; Yusheng, K; Haris, M; Ul Ain, Q; Javaid, H.M. (2022). Impact of financial leverage on sustainable growth, market performance, and profitability. *Econ. Chang. Restrict.* 55, 737–774.

Ali, S; Khan, M; Ahmed, R. (2022). Advances in K-Mean Clustering Algorithm: Enhancing Performance and Accuracy. *Journal of Machine Learning Research*.23, 1125-1145.

Arora, P; Deepali, D; Varshney, S. (2016). Analysis of K-Means and K-Medoids algorithm for big data. *Procedia Comput. Sci.* 78, 507–512.

Bayaraa, B. (2017).Financial performance determinants of organizations: The case of Mongolian companies.*J.Compet.*9, 22–33.

Ben-Gal, I. Outlier Detection. In Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, 2nd ed; Maimon, O; Rockach, L; Eds; Springer: New York, NY, USA, 2010; pp. 117–130.

Brealey, R. A; Myers, S. C; Allen, F. Principles of Corporate Finance; McGraw-Hill Education: 2020.

Cai, F; Le-Khac, N.A; Kechadi, M.T. (2012). Clustering Approaches for Financial Data Analysis. In Proceedings of the 8th International Conference on Data Mining (DMIN 2012), Las Vegas, NV, USA, 15–18 December.

De Blasio, V; Pavone, P; Migliaccio, G. Cosmetics companies: Income developments in time of crisis. J. Small Bus. Enterp. Dev.2022. ahead-of-print.

Deloitte. Global Powers of Retailing. 2021.

Do, D.T. A study on financial performance of transport & warehouses firms listed on the Hanoi stock exchange. Econ. Financ. Lett. 2021, 8, 44–52.

Dsouza, S; Dsouza, J.D; Vanitha, T. Analysis of data using k-means and k-medoids algorithms. Int. J. Latest Trends Eng. Technol. Spec. Issue SACAIM 2017, 370–373.

Dzuba, S; Krylov, D. (2021). Cluster analysis of financial strategies of companies. *Mathematics*. 9, 3192.

Esfahani, R.K; Shahbazi, F; Akbarzadeh, M. (2019). Three-phase classification of an uninterrupted traffic flow: A k-means clustering study. *Transp. B-Transp. Dyn.* 7, 546–558.

Etemadi, H; Hisarzadeh, R; Rahmani, A; & Azar, A. (2014). Modeling information uncertainty management using fuzzy decision tree methodology:

Evidence from feasibility study of information-uncertainty-based theory. Quarterly Journal of the Tehran Stock Exchange, (25). (In Persian)

Fenyves, V; Zsido, K.E; Bircea, I; Tarnoczi, T. (2020). Financial performance of Hungarian and Romanian retail food small businesses.Br. Food J. 122, 3451–3471.

Goral, J; Soliwoda, M. (2021). On the profitability of Polish large agricultural holdings. Acta. Oecon. 71, 137–159.

Güla˘gız, F.K; Sahin, S. (2017). Comparison of hierarchical and non-hierarchical clustering algorithms. *Int. J. Comput. Eng. Inf. Technol*. 2017, 9, 6–14.59.

Hatem, B.S. (2014) Determinants of firm performance: A comparison of European countries. Int. J. Econ. Financ. 6, 243–249.

Hennig, C. (2015). What are the true clusters? *Pattern Recognit. Lett*. 64, 53–62.

Hepsen, A; Vatansever, M. (2012). Using hierarchical clustering algorithms for Turkish residential market. *Int. J. Econ. Financ.* 4, 138–150.

Herman, E. The interplay between digital entrepreneurship and sustainable development in the context of the EU digital economy: A multivariate analysis. Mathematics 2022, 10, 1682.

Herman, E; Zsido, K.-E; & Fenyves, V. (2022). Cluster analysis with K-Mean versus K-Medoid in financial performance evaluation. Applied Sciences, 12(12), 7985. https://doi.org/10.3390/

Herman, E; Zsido, K.-E; Fenyves, V. (2022). Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation. *Appl. Sci.* 12, 7985.

Hernant, M; Andersson, T.B; Hilmola, O.P. (2007) Managing retail chain profitability based on local competitive conditions: Preliminary analysis. *Int. J. Retail Distrib. Manag*. 35, 912–935.

Hloušková, Z; Lekešová, M. (2020). Farm outcomes based on cluster analysis of compound farm evaluation. *Agric. Econ.-Czech.* 66, 435–443.

Ikotun, A.M; Almutari, M.S; Ezugwu, A.E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Appl. Sci.* 11, 11246.

Imani, M. (2012). Clustering Persian texts (Master's thesis). Department of Computer Engineering, Sharif University of Technology. (In Persian)

Iqbal, M.Z; Riaz, M; Nasir, W. (2017). Multivariate Outlier Detection: A comparison among two clustering techniques. Pak. J. Agric. Sci. 54, 227–231.

Izadparast, M. (2011). Customer classification in insurance using data mining. Monthly Journal of World Insurance News, (161), Payame Noor University, Tehran. (In Persian)

J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.

Kamalian, A.-R; Niknafs, A.-A; Afsharizad, O; & Gholamalipour, R. (2011). Identifying the most important factors affecting the information disclosure ranking of companies listed on the Tehran Stock Exchange using a data mining approach. Quarterly Journal of the Tehran Stock Exchange, (11). (In Persian)

Kaur, N.M; Kaur, U; Singh, D. (2014). K-medoid clustering algorithm—A review. *Int. J. Comput. Appl. Technol*. 1, 42–45.

Khour, S; Elexa, L; Istok, M; Rosova, A. (2019). A Study from Slovakia on the transfer of Slovak companies to tax havens and theirimpact on the sustainability of the status of a business entity. *Sustainability*. 11, 2803.

Kodinariya, T.M; Makwana, P.R. (2013). Review on determining number of cluster in K-Means clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud*. 1, 90–95.

Kou, G; Peng, Y; Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Inf. Sci*. 275, 1–12.

Kramaric, T.P; Bach, M.P; Dumicic, K; Zmuk, B; Zaja, M.M. (2022). Exploratory study of insurance companies in selected post-transition countries: Non-hierarchical cluster analysis. Cent. Eur. J. Oper. Res. 2018, 26, 783–807.

L. Breiman, Random Forests, Machine Learning, 2001.

L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, 1990

Lemieux, V; Rahmdel, P.S; Walker, R; Wong, B.L.W; Flood, M. (2014). Clustering techniques and their effect on portfolio formation and risk analysis.

Lukác, J; Teplická, K; Culková, K; Hrehová, D. (2021). Evaluation of the financial performance of the municipalities in Slovakia in the context of multidimensional statistics. *J. Risk Financ. Manag.* 14, 570.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise (KDD), 1996.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, (2010). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), 1996.

McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.

Medellu, J.V.C; Nugraha, E.S. (2021). K-means and k-medoid algorithm application in clustering stock data in Indonesia. In Proceedings of the Symposium on Data Science Petra, Jordan, 5–7 April.

Nguyena, V.C; Nguyena, T.N.L; Tranb, T.T; Nghiemc, T.T. (2019). The impact of financial leverage on the profitability of real estate companies: A study from Vietnam stock exchange. *Manag. Sci. Lett*. 9, 2315–2326

Pakhira, M.K. (2012). Finding Number of Clusters before Finding Clusters. Proc. Technol. 4, 27–37.

Park, H.S; Jun, C.H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Syst. Appl*, 36, 3336–3341.

Pavelkova, D; Zizka, M; Homolka, L; Knapkova, A; Pelloneova, N. (2021). Do clustered firms outperform the non-clustered? Evidence of financial performance in traditional industries. Ekon. Istraz. 34, 3270–3292.

Pech, M; Vrchota, J. (2020). Classification of Small- and Medium-Sized Enterprises Based on the Level of Industry 4.0 Implementation. *Appl. Sci*; 10, 5150.

Pelloneová, N. (2021). Are there differences in the financial performance of Czech and Slovak cluster organizations? Ekon. Cas; 69, 907–927.

Pham, D.T; Afify, A.A. (2007). Clustering techniques and their applications in engineering. *Proc. Inst. Mech. Eng. Part C-J. Eng. Mech.Eng. Sci*. 221, 1445–1459.

Popa, D.N; Bogdan, V; Sabau Popa, C.D; Belenesi, M. (2021). Performance mapping in two-step cluster analysis through ESEG disclosures and EPS. Kybernetes, 51, 98–118.

Ren, J; Hua, K; & Cao, Y. (2022). Global Optimal K-Medoids Clustering of One Million Samples. *In Advances in Neural Information Processing Systems 35* (NeurIPS 2022).

Sabău Popa, C.D; Popa, D.N; Bogdan, V; Simut, R. (2021). Composite financial performance index prediction—A neural networks approach. *J. Bus. Econ. Manag*. 22, 277–296.

Sabau, A.S. (2012). Survey of clustering based financial fraud detection *research. Inform. Econ*. 16, 110–122.

Serban, E.C; Bogeanu, A; Tudor, E. (2013). Clustering techniques in financial data analysis applications on the U.S. financial market. *Ann. Constantin Brâncu¸si Univ. Târgu Jiu Econ. Ser*. 4, 176–194.

Shiry. (2006). Introduction to clustering. Lecture notes for Machine Learning course, Amirkabir University of Technology. Retrieved from http://www.aut.ac.ir/shiry/clustering.html. (In Persian).

Simon, J. (2006). Applications of cluster analysis in marketing research. *Stat. Szle*. 84, 627–650.

Soni, K.G; Patel, A. (2017). Comparative Analysis of K-means and K-medoids algorithm on IRIS Data. *Int. J. Comput. Intell. Syst.*13, 899–906.

Szabo, Z.K; Herman, E. (2014). Productive entrepreneurship in the EU and its barriers in transition economies: A cluster analysis. *ActaPolytech. Hung.* 11, 73–94.

Szüle, B. (2019). Comparison of cluster number determination methods. *Stat. Szle.* 97, 421–438.

T. Hastie, R. Tibshirani, and J. Friedman, the Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Springer, 2009.

T. Hastie, R. Tibshirani, and J. Friedman, the Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009.

Velmurugan, T. (2012). Efficiency of K-Means and K-Medoids algorithms for clustering arbitrary data points. *Int. J. Comput. Appl. Technol*. 3, 1758–1764

Velmurugan, T; Santhanam, T. (2010). Computational complexity between K-Means and K-Medoids clustering algorithms for normal and uniform distributions of data points. *J. Comput. Sci*. 6, 363–368.

Vishwakarma, S; Nair, P.S; Rao, D.S. (2017). A comparative study of K-means and K-medoid clustering for social media text mining. *Int.J. Adv. Sci. Res. Eng. Trends*. 2, 297–302.

Wang, Y.J; Lee, H.S. (2008). A clustering method to identify representative financial ratios. *Inf. Sci.* 178, 1087–1097.

Windgassen, S; Moss-Morris, R; Goldsmith, K; Chalder, T. (2018). The importance of cluster analysis for enhancing clinical practice: An example from irritable bowel syndrome. *J. Ment. Health*. 27, 94–96.

Xiaojun, S; Yalin, L. (2021). Research on financial early warning of mining listed companies based on BP neural network model. *Resour. Policy* 73, 102223.

Yadegari, R; Rahmani, K; Modarres Khiyabani, F. (2018). Providing a comprehensive model to measure the performance dimensions of industrial clusters using the hybrid approach of Q factor analysis and cluster analysis. *Int. J. Qual. Res*. 13, 235–248.

Yen, M.-F; Huang, Y.-P; Yu, L.-C; Chen, Y.-L. (2022). A two-dimensional sentiment analysis of online public opinion and future financial performance of publicly listed companies. *Comput. Econ*. 59, 1677–1698.

Zainudin, R; Mahdzan, N.S; Mohamad, N.N. (2021) Internationalisation and financial performance: In the case of global automotivefirms. *Rev. Int. Bus. Strategy*. 31, 80–102.

Zhussupova, Z; Onyusheva, I; El-Hodiri, M. (2018). Corporate governance and firm value of Kazakhstani companies in the conditions of economic instability. Pol. *J. Manag. Stud*; 17, 235–245.