

خوشه‌بندی بانک‌های کشور با استفاده از الگوریتم‌های یادگیری ماشین

محمد مهدی داودی^۱

چکیده

بازار بین‌بانکی یکی از مهم‌ترین ابزار تأمین نقدینگی مؤسسات اعتباری محسوب می‌شود. با این حال، با توجه به شرایط نامطلوب اقتصادی و بحران‌های مالی متوالی در پی بحران‌های سیاسی و اقتصادی، فعالیت در این بازار با ریسک‌های با اهمیتی همراه است. به عبارتی مهم‌ترین ریسک فعالیت در این بازار ریسک عدم ایفای تعهدات بانک‌های سپرده‌پذیر می‌باشد. از این رو در این پژوهش سعی شده است تا با بررسی عملکرد بانک‌های کشور با استفاده از داده‌های استخراج‌شده از صورت‌های مالی آن‌ها، وضعیت نظام بانکی به‌طور کلی ارزیابی گردد. در این راستا برای ارایه مدلی جامع، از الگوهای یادگیری ماشین نظارت‌نشده و به‌طور خاص از الگوهای خوشه‌بندی بهره‌برده شده است. به‌منظور پیاده‌سازی الگوی خوشه‌بندی و ارزیابی نتایج الگوهای مختلف از ۴ الگوی Hierarchical-K-، Hierarchical، K-Means و Model Based استفاده شده است. تأیید و همگرایی نتایج الگوهای مختلف به‌منظور برآورد الگوریتم خوشه‌بندی یکی از مهم‌ترین دستاوردهای گزارش حاضر به‌شمار می‌رود. نتایج الگوهای مختلف برآورد شده نشان می‌دهد که در بهترین و بهینه‌ترین حالت، بانک‌های مورد بررسی را می‌توان به ۹ خوشه طبقه‌بندی کرد. در واقع نتایج برآورد شده حاکی از اختلاف، فاصله و یا به‌عبارت‌دیگر عدم شباهت بالای بانک‌های مورد بررسی می‌باشد. این نکته نیز قابل توجه می‌باشد که ۲۳ بانک مورد بررسی به ۹ خوشه متفاوت تقسیم می‌شوند. این مهم نشان می‌دهد نسبت‌های مالی بین بانک‌های مورد مطالعه اختلاف چشمگیری با یکدیگر دارند.

واژه‌های کلیدی: بازار بین بانکی، الگوریتم k-means، الگوریتم سلسله مراتبی، مدیریت ریسک، نسبت‌های مالی
طبقه‌بندی JEL: C38، G24، D81

^۱ کارشناس ارشد علوم اقتصادی، دانشکده اقتصاد دانشگاه علامه طباطبایی و کارشناس ریسک بانک اقتصاد نوین، mdavoudi70@gmail.com

۱ مقدمه

بازار بین‌بانکی یکی از اجزای بازار پول است که در آن بانک‌ها و سایر مؤسسات اعتباری نسبت به معاملات با یکدیگر مبادرت می‌ورزند. این بازار حداقل دو نقش حیاتی در نظام‌های مالی نوین ایفا می‌کند. اولین و مهم‌ترین نقش آن، مداخله فعالانه و مؤثر بانک مرکزی در اجرای سیاست پولی از طریق راهبری نرخ‌های سود است. دوم، بازارهای بین‌بانکی کارآمد، نقدینگی را به شکل مطلوبی از مؤسسات مالی دارای مازاد و جوه به مؤسسات دارای کسری و جوه انتقال می‌دهند؛ بنابراین، سیاست‌گذاران انگیزه بالایی برای ایجاد یک بازار بین‌بانکی کارآمد و قوی دارند تا بانک مرکزی بتواند به نرخ سود مطلوب موردنظر خود دست یابد و مؤسسات مالی نیز قادر به مبادله نقدینگی به شکل کارایی در میان خود باشند.

با وجود تصویب مقررات و دستورالعمل بازار بین‌بانکی ریالی در سال ۱۳۸۳، این بازار، به دلیل آماده نبودن زمینه اجرای مقررات و اضافه برداشت قابل توجه بانک‌ها از بانک مرکزی، تا سال ۱۳۸۷ عملیاتی نگردید. این بازار در ۱۸ تیر ۱۳۸۷ به صورت رسمی آغاز به فعالیت نمود. در ابتدا بانک مرکزی با پذیرش عضویت ۱۰ بانک که در مقایسه با سایر بانک‌ها از مدیریت نقدینگی نسبتاً بهتری برخوردار بودند، بازار را فعال کرد. این در حالی است که طی سال‌های اخیر تمامی مؤسسات مالی و اعتباری فعال در کشور از ابزار این بازار برای مدیریت بهینه نقدینگی، استفاده می‌نمایند.

با وجود اینکه ابزارهای بازار بین‌بانکی نقشی مؤثر در مدیریت نقدینگی نهادهای مالی ایفا می‌کنند، در برخی از برهه‌های زمانی به دلیل افزایش نرخ سود سپرده‌گذاری، هزینه بالایی را به بانک تحمیل می‌کند. از عوامل مؤثر در نرخ سود علی‌الحساب سپرده‌پذیری/سپرده‌گذاری در این بازار می‌توان به نرخ سود علی‌الحساب تجهیز منابع بانک، نرخ سود تسهیلات بانک، ریسک بانک سپرده‌پذیر، مدت تسهیلات و محدودیت‌های اجرایی بانک مرکزی ج.ا.ا. اشاره داشت.

همان‌طور که ذکر شد یکی از مهم‌ترین عوامل تأثیرگذار بر نرخ سود بازار بین‌بانکی، ریسک بانک سپرده‌پذیر می‌باشد. به عبارتی بانک‌هایی که از نظر شاخص‌های عملیاتی از وضعیت چندان مناسبی برخوردار نیستند، با مشکلات بیشتری در سپرده‌پذیری و جوه مواجه می‌شوند. به طوری که بانک‌های سپرده‌گذار با لحاظ ریسک بازپرداخت و جوه سپرده‌گذاری شده، تمایل به سپرده‌گذاری در بانک‌های پر ریسک را ندارند. در این شرایط بانک‌های پر ریسک در راستای تأمین و جوه موردنیاز می‌بایست با نرخ‌های بالاتری اقدام به سپرده‌پذیری نمایند تا سایر بانک‌ها به سپرده‌گذاری در آن‌ها ترغیب شوند.

در راستای طبقه‌بندی و یا رتبه‌بندی بانک‌های فعال در کشور، الگوهای متفاوتی وجود دارد. با این حال می‌بایست به این نکته توجه شود که رتبه‌بندی بانک‌های فعال در کشور در قیاس با طبقه‌بندی آن‌ها، بحثی مجزا می‌باشد. در رابطه با رتبه‌بندی بانک‌های فعال در کشور باید این نکته در نظر گرفته شود که رتبه‌بندی بانک‌ها بر اساس نسبت‌های مالی می‌تواند گاهی موجب گمراهی تحلیلگران مالی شود، چراکه برخی نسبت‌های مالی قابلیت رتبه‌بندی ندارند. به‌طور مثال در خصوص نسبت تسهیلات به سپرده نمی‌توان اظهار نمود بانکی که نسبت تسهیلات به سپرده بالاتری دارد بانکی با عملکرد بهتر و یا بانکی که نسبت تسهیلات به سپرده پایین‌تری دارد، بانکی با شرایط نامناسب‌تر می‌باشد. چرا که با توجه به متون مالی، بهترین مقدار برای نسبت مذکور می‌بایست بین ۸۰ تا ۹۰ درصد باشد. سایر نسبت‌های مالی نیز چنین وضعیتی دارند. از این رو برای اجتناب از اشتباه، در این پژوهش از روش خوشه‌بندی استفاده شده است.

خوشه‌بندی^۱ فرآیند دسته‌بندی مجموعه‌ای از نمونه‌ها به خوشه‌هایی است که اعضای درونی هر خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را نسبت به اعضای سایر خوشه‌ها داشته باشند. با استفاده از این الگو می‌توان بانک‌هایی که از نظر رفتاری بیشترین شباهت را به یکدیگر دارند شناسایی و با توجه به پیشینه رفتار آن‌ها، در مواجهه با بانک‌های هم رفتار نیز تصمیم‌سازی کرد. به عبارتی زمانی که یک بانک به‌عنوان مشتری جدید مراجعه می‌نماید، می‌توان با در نظر گرفتن خوشه‌بندی انجام‌شده، بانک‌های هم‌رفتار با بانک مذکور را تعیین نمود. پس از آن می‌توان رفتار مشتری جدید را مطابق با رفتار مشتریان هم‌گروه خود در قبال اجرای تعهداتش پیش‌بینی نمود.

¹ clustering

با این وجود، دامنه مطالعات خوشه‌بندی طی سال‌های اخیر از اهمیت ویژه‌ای برخوردار شده است. امروزه خوشه‌بندی یکی از پرکاربردترین ابزارها در زمینه هوش مصنوعی به شمار می‌آید. خوشه‌بندی به علت رشد روزافزون حجم متون تحت وب، پیکره‌های متنی، مقالات و... از اهمیت خاصی برخوردار است. از این رو، روزبه‌روز الگوریتم‌های پیچیده‌تری برای خوشه‌بندی داده‌ها توسط متخصصین آماری منتشر می‌شود.

رتیب مباحث ارایه شده در این مقاله بدین شرح است: ابتدا پیشینه پژوهش مرتبط با خوشه‌بندی بانک‌ها در داخل و خارج از کشور بررسی شده است. سپس مباحث نظری مرتبط با الگوریتم‌های مورد استفاده در تحقیق ارایه شده است. پس از آن داده‌های مورد استفاده در پژوهش معرفی شده است. در آخر نیز نتایج مدل‌سازی و جمع‌بندی مباحث ارایه شده است.

۲ پیشینه پژوهش

همان‌طور که در بخش مقدمه ذکر شد طیف وسیعی از پژوهش‌ها در خصوص رتبه‌بندی و بررسی بانک‌های کشور صورت پذیرفته است. در ادامه به برخی از مهم‌ترین این پژوهش‌ها در داخل کشور اشاره شده است.

سلیمی و همکاران (۱۳۹۵) با استفاده از الگوی اعتبارسنجی به روش دلفی فازی، عوامل مدل نهایی با شش بعد و ۲۶ مؤلفه را شناسایی نمودند. در ادامه با استفاده از فرایند تحلیل شبکه‌ای و روش دیمتل، ضریب اهمیت هر یک از این عوامل را به دست آورده و بدین ترتیب مدل نهایی ارائه شده است. در ادامه بانک‌های بورسی و فرابورسی فعال در نظام بانکی کشور، بر مبنای صورت‌های مالی حسابرسی شده‌ی سال ۱۳۹۳، با استفاده از مدل ارائه‌شده و روش تاپسیس رتبه‌بندی شدند. نتایج به دست آمده نشان می‌دهد، به ترتیب بانک‌های پاسارگاد، خاورمیانه، کارآفرین، دی و سینا در سال ۹۳ عملکرد مطلوب‌تری نسبت به سایر بانک‌های مورد مطالعه داشته‌اند.

ارضاء و قاسم پور (۱۳۹۶) در پژوهش خود با استفاده از مدل کملز^۱ و با استفاده از رویکرد ترکیبی فرایند تحلیل سلسله مراتبی و آراس^۲ به رتبه‌بندی بانک‌های بورسی و فرابورسی فعال در نظام بانکی کشور پرداختند. در این پژوهش شاخص‌های کفایت سرمایه، کیفیت دارایی، مدیریت، سودآوری، نقدینگی و حساسیت نسبت به ریسک بازار برای رتبه‌بندی بانک‌ها استفاده شده است. نتایج پژوهش نشان می‌دهد که شاخص کفایت سرمایه و سودآوری بیشترین اهمیت در میان شاخص‌های مربوط به رتبه‌بندی را داشته‌اند.

علیزاده زوارم و پویا (۱۳۹۶) ۳۱ بانک و مؤسسه مالی و اعتباری کشور را در قالب ۶ مورد از مهم‌ترین شاخص‌های ترافیکی وب‌سایت شامل متوسط تعداد صفحات بازدید شده، متوسط زمان مرور سایت، درصد بازدیدکنندگان از داخل کشور، درصد بازدیدکنندگان از خارج کشور، تعداد پیوندها و سرعت بارگذاری بر اساس وب‌سایت و موتور جستجوی الکسا مورد ارزیابی قرار دادند. سپس با استفاده از روش خوشه‌بندی سلسله مراتبی، بانک‌ها و مؤسسات شبیه به هم در قالب خوشه‌هایی همگن دسته‌بندی شدند. کیفیت و اعتبار هر یک از خوشه‌های ایجاد شده با استفاده از روش تحلیل تشخیصی مورد تأیید قرار گرفت و در نهایت، به تحلیل ویژگی‌های هر یک از این خوشه‌ها پرداخته شده است. یافته‌ها حاکی از آن است که بین خوشه‌های مختلف از نظر شاخص سرعت بارگذاری تفاوت معناداری وجود ندارد اما در مورد میزان سایر شاخص‌ها، تفاوت بین خوشه‌های مختلف، معنادار بوده است.

خاتمی فیروزآبادی و همکاران (۱۳۹۶) به کمک یک مدل عمومی که بیانگر تأخر زمانی مراجعات مشتری به بانک، تعداد تراکنش‌های مالی و مقدار سپرده‌گذاری آن‌ها می‌باشد، از دو منظر تسهیلات و معوقات، مدل را جهت خوشه‌بندی و رتبه‌بندی مشتریان در شبکه بانکی توسعه داده‌اند. در این تحقیق طیف وسیعی از مشتریان با ویژگی‌های مختلف در گروه‌های مشابه خوشه‌بندی شده و سپس با استفاده از روش مجموع وزن دهی ساده به رتبه‌بندی آن‌ها پرداخته شده است. از مهم‌ترین

¹ CAMELS

² ARAS

نتایج تحقیق می‌توان به سه خوشه از مشتریان اشاره نمود که در مهم‌ترین آن‌ها تنها ۵۶۴ مشتری، ۰/۰۳ درصد از کل مشتریان حقیقی و خوش حساب بانک بوده و نیم درصد کل سپرده‌ها و ۴ درصد کل تسهیلات را دارند. در خوشه دوم که شامل دو گروه از مشتریان بوده و یک درصد از کل مشتریان حقیقی بانک را شامل شده، ۳۶ درصد کل سپرده‌های بانک را در اختیار دارند و جزء مشتریانی می‌باشند که ممکن است بانک را با ریسک نقدینگی مواجه نمایند. خوشه سوم که ۶۷ درصد کل مشتریان خرد بانک را در برمی‌گیرد دارای رفتار نرمال می‌باشند. آن‌ها از لحاظ شاخص‌ها تأخر، متوسط مانده سپرده و مقدار تسهیلات، کمتر از متوسط هر شاخص در کل مشتریان بانک می‌باشند. از مهم‌ترین کاربردهای این مقاله، تبیین سیاست‌های بازاریابی ویژه برای هر خوشه می‌باشد.

یوسفی (۱۳۹۷) به رتبه‌بندی بانک‌ها و ارائه یک مدل تصمیم‌گیری چند معیاره فازی برای رتبه‌بندی بانک‌های کشور پرداخته است. برای به دست آوردن اطلاعات مالی جامعه آماری پژوهش که شامل ۳۱ بانک و ۲ موسسه مالی و اعتباری می‌باشد از ترازنامه و صورت سود و زیان حسابرسی شده مربوط به پایان سال ۱۳۹۳ استفاده شده است. برای تجزیه و تحلیل اطلاعات و رتبه‌بندی بانک‌ها از ترکیب روش فرآیند تحلیل سلسله مراتبی فازی (FAHP) و تکنیک اولویت‌بندی ترجیحات براساس نزدیکی به راه‌حل ایدئال (TOPSIS) استفاده شده است. در مرحله اول وزن اهمیت هر یک از شاخص‌های مالی با استفاده از AHP فازی به دست آمد و در مرحله بعد، با استفاده از روش TOPSIS رتبه‌بندی بانک‌ها صورت پذیرفته است. نتایج حاصل از پژوهش بیانگر آن است که شاخص سرمایه از مهم‌ترین عوامل تأثیرگذار در رتبه‌بندی بانک‌ها می‌باشد و همچنین بانک‌هایی که از لحاظ شاخص‌های مالی در وضعیت خوبی قرار ندارند باید به دنبال تقویت جایگاه خود از طریق بهبود در وضعیت شاخص‌های مالی و یا ادغام با شرکای استراتژیک و یا بانک‌های بزرگ‌تر باشند.

پس از بررسی پژوهش‌های صورت گرفته در داخل کشور به برخی از مهم‌ترین پژوهش‌های صورت گرفته در خارج از کشور پرداخته شده است.

درداک و بویتان (۲۰۰۹)^۱ با گردآوری مؤسسات اعتباری دارای ویژگی‌های مشابه از نظر فعالیت واسطه‌گری مالی، ثابت کردند که خوشه‌بندی می‌تواند کمک شایانی در تجزیه و تحلیل گروه هم‌تارانه نماید. هدف از این مطالعه این بود که نمونه‌ای نماینده از مؤسسات اعتباری رومانیایی را در خوشه‌های کوچک‌تر و همگن گنجانده تا نشان دهد کدام مؤسسات اعتباری الگوهای مشابهی با توجه به مشخصات ریسک و سودآوری خود دارند. آن‌ها دریافتند که طی دوره ۲۰۰۴-۲۰۰۶، خوشه‌ها از نظر شباهت ریسک‌ها و سودآوری نسبتاً پایدار باقی مانده‌اند.

ارکان و سایاسنگ (۲۰۱۶)^۲ با استفاده از الگوهای خوشه‌بندی در بخش بانکداری اروپا و بهره‌برداری از شاخص‌های بانکی تلفیقی از بانک مرکزی اروپا توانستند بانک‌های مذکور را خوشه‌بندی نمایند. این مطالعه بانک‌های مشابه را با توجه به نسبت‌های بخش بانکی و تغییرات در گروه‌های خوشه‌ای متأثر از بحران مالی تعیین کرد. هدف از این پژوهش پاسخ به این سؤال است که آیا مالکیت خارجی بانک‌ها به ویژگی یا خوشه‌بندی این بانک‌ها کمک می‌کند یا یک ترکیب خاص یک کشور است. یافته‌های پژوهش تأیید می‌کند که گروه‌بندی بخش‌های بانکی بر اساس نسبت‌های بانکی نشان می‌دهد که کشورهای اتحادیه اروپا در منطقه جغرافیایی مشابه و با مشارکت اقتصادی بالاتر تمایل دارند در خوشه مشابه گروه‌بندی شوند.

مولین (۲۰۱۷)^۳ از تحلیل خوشه‌بندی برای داده‌های بانکی که با معیار کالینسکی-هاراباسز^۴ و روش وارد^۵ پردازش شده است استفاده کرده است. مدل‌های تجاری که این خوشه‌ها نشان می‌دهند، بانکداری جهانی با تمرکز خرده‌فروشی و بانکداری

¹ Dardac, N., & Boitan, I. A. (2009)

² Ercan, H., & Sayaseng, S. (2016)

³ Molin, F. (2017)

⁴ Calinski-Harabasz Criterion

⁵ Ward's method

جهانی با تمرکز عمده‌فروشی هستند. این مدل‌های کسب‌وکار در طول زمان (۲۰۰۷-۲۰۱۶) تحلیل شده‌اند که نشان می‌دهد مؤسسات اعتباری در جهتی سالم توسعه یافته‌اند. بنابراین، مؤسسات اعتباری در سال ۲۰۱۶ نسبت به سال ۲۰۰۷ از نظر مالی قابل اعتمادتر بودند. با توجه به روند داده‌ها، این توسعه احتمالاً ادامه خواهد داشت.

لاگاسیو و قارانتا (۲۰۲۱)^۱ مدل‌های تجاری ۱۲۳۷ بانک در طول سال‌های ۲۰۱۱-۲۰۱۷ را با استفاده از یک تحلیل خوشه‌بندی شامل نسبت‌های هر دو طرف ترازنامه بانک و صورت سود و زیان تحلیل کردند. سپس، از طریق تجزیه و تحلیل ارتباطات در پی این سال بودند که آیا بین پذیرش یک مدل تجاری خاص و اندازه، سودآوری، کارایی و مشخصات ریسک ارتباط وجود دارد یا خیر. آن‌ها در نهایت دریافته‌اند که یک مدل تجاری بانک بر اساس عوامل درونزای مختلف (اقلام صورت‌های مالی بانک و استراتژی عملیاتی) است و می‌تواند با عوامل برون‌زا (بحران مالی، سیاست‌های بانک‌های مرکزی) مرتبط باشد.

همان‌طور که ملاحظه گردید بخش عمده‌ای از مقالات و پژوهش‌های موردبررسی به‌خصوص در داخل کشور به رتبه‌بندی بانک‌ها پرداخته بودند. این در حالی است که در برخی موارد با توجه به اهداف سیاست‌گذار، خوشه‌بندی بانک‌های کشور و استخراج و گردآوری بانک‌های هم رفتار می‌تواند کمک شایانی در برنامه‌ریزی مؤسسات مالی داشته باشد. از این‌رو وجه تمایز این پژوهش با مقالات مبحث‌عنه‌الگوی مورد استفاده در پژوهش مبتنی بر الگوهای یادگیری ماشین می‌باشد.

۳ مباحث نظری

در این پژوهش از ۴ الگوی خوشه‌بندی متفاوت به شرح ذیل استفاده شده است:

- الگوی ۱: K-Means یا کای میانگین
 - الگوی ۲: Hierarchical یا سلسله مراتبی
 - الگوی ۳: Hierarchical-K-Means ترکیب الگوهای سلسله مراتبی و کای میانگین
 - الگوی ۴: Model-Based یا بر مبنای مدل
- در ادامه هر چهار الگوی فوق به اختصار توضیح داده خواهد شد.

۱.۳ الگوی ۱: روش خوشه‌بندی K-Means

در دهه‌های ۱۹۵۰ تا ۱۹۷۰ الگوهای متعددی توسط پژوهشگران برای خوشه‌بندی به روش K-means یا کای میانگین ارایه شده است (پرز-اورتگا و همکاران، ۲۰۱۹)^۲. اولین بار این الگو توسط لوید (۱۹۵۷)^۳ معرفی گردید. باین‌حال به دلیل عدم پیشرفت لازم در حوزه کامپیوتر تا سال ۱۹۸۲ در هیچ مجله‌ای مورد استفاده قرار نگرفته بود^۴. این الگو بر مبنای منطق یادگیری ماشین^۵ طراحی شده است. در این الگو داده‌های موردبررسی به K گروه مختلف طبقه‌بندی می‌شوند، به طوری که فاصله بین عناصر درون گروه‌ها کمترین مقدار ممکن می‌باشد.

این الگوریتم پارامتر K را به‌عنوان ورودی گرفته و مجموعه n شیء را به k خوشه، افراز می‌کند. به طوری که سطح شباهت داخلی خوشه‌ها زیاد و سطح شباهت اشیا بیرون خوشه‌ها کم باشد. شباهت هر خوشه نسبت به متوسط اشیا آن خوشه سنجیده شده که این متوسط، مرکز خوشه نامیده می‌شود. کارکرد این الگوریتم بدین شرح است:
ورودی: K ، تعداد خوشه‌ها. پایگاه داده X شامل n شیء:

¹ Lagasio, V., & Quaranta, A. G. (2021)

² Pérez-Ortega, J., et al (2019)

³ Lloyd, S. P. (1957)

⁴ Lloyd, S. (1982).

⁵ Unsupervised Machine Learning

$$X = \{x_1, x_2, \dots, x_n\}, x_i \in R^m \quad (1)$$

خروجی: یک مجموعه از K خوشه که معیار مربع خطا (sum of square) را حداقل می‌کند.
مراحل اجرای الگوریتم خوشه‌بندی K-Means بدین صورت است مک کی (۲۰۰۳):^۱

۱. به صورت تصادفی، K داده دلخواه را به عنوان مراکز خوشه‌های ابتدایی انتخاب می‌کنیم.
۲. هر داده را با توجه به نزدیکی آن به مراکز خوشه‌ها، به خوشه‌ها تخصیص می‌دهیم. معیار نزدیکی را فاصله اقلیدسی در نظر می‌گیریم:

$$d(x_i, m_j) = \|x_i - m_j\| = \sqrt{\sum_{k=1}^m (x_{ik} - m_{jk})^2} \quad (2)$$

که در آن x_i ، i امین داده ورودی و m_j مرکز (میانگین) خوشه j ام است. داده مذکور در خوشه‌ای قرار می‌گیرد که کم‌ترین فاصله را با مرکز آن خوشه داشته باشد. مرکز خوشه نیز میانگین حسابی داده‌های آن خوشه است. مراکز خوشه‌ها را به روز می‌کنیم یعنی برای هر خوشه میانگین اعضای خوشه را به دست می‌آوریم. با توجه مراکز جدید خوشه‌ها به مرحله ۲ برمی‌گردیم و فرآیند فوق را تا جایی ادامه می‌دهیم که هیچ تغییری در خوشه‌ها رخ ندهد (در این حالت الگوریتم پایان یافته است).
برای مراکز خوشه‌ها $\{\mu_j = 1, \dots, k\}$ داریم:

$$k_j = \min \sum_{j=1}^k S_j \quad (3)$$

که k_j خوشه j ام و مجموع مربعات برای خوشه j به صورت رابطه (۴) است:

$$S_j = \sum_{i=1}^n z_{ji} |x_i - \mu_j|^2 \quad (4)$$

که در آن $z_{ji} = 1$ است، اگر در گروه j باشد و در غیر این صورت برابر صفر است. پارامتر μ_j برابر میانگین گروه j است:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^n z_{ji} x_i \quad (5)$$

از مزایای این الگوریتم سادگی و همگرایی سریع در رسیدن به جواب نهایی است. همچنین، از معایب این الگوریتم این موضوع می‌باشد که چون در ابتدا تعداد خوشه‌ها به الگوریتم داده می‌شود بهینه بودن جواب به دست آمده را نمی‌توان تضمین کرد، زیرا تعداد خوشه‌ها به صورت دستی انتخاب می‌گردد و به تبع آن اشتباه در تعیین این عدد در جواب نهایی تأثیر زیادی دارد. کارایی الگوریتم نیز بسیار وابسته به نحوه تعیین مراکز اولیه برای خوشه‌هاست.

برای تخمین الگوی k -means از ۴ الگوی لوید (۱۹۵۷)^۲، فورگی (۱۹۶۵)^۳، مک کوپین (۱۹۶۷)^۴ و هارتینگان و وانگ (۱۹۷۹)^۵ استفاده شده است. برای مطالعه هر کدام از روش‌های فوق و مقایسه تفاوت‌های آنها می‌توان به (پرز-اورتگا و همکاران، ۲۰۱۹) مراجعه کرد.

¹ MacKay, D. J. (2003)

² Lloyd, S. P. (1957)

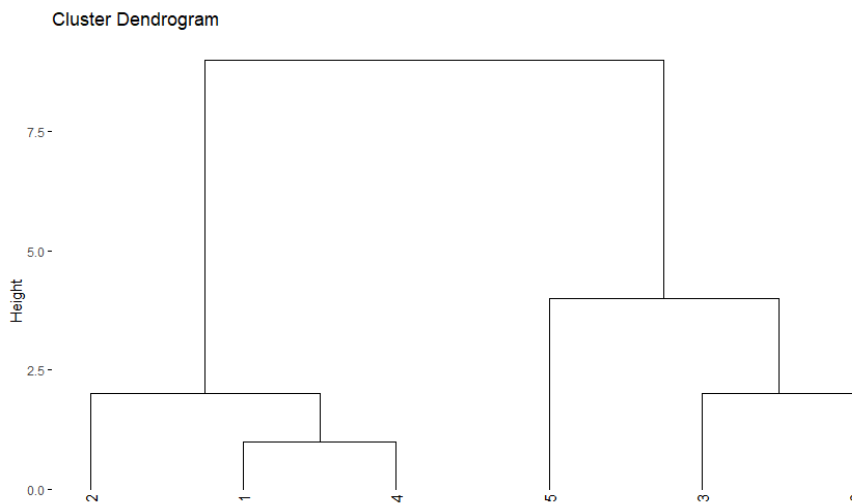
³ Forgy, E. W. (1965)

⁴ MacQueen, J. (1967)

⁵ Hartigan, J. A. and Wong, M. A. (1979).

۲.۳ الگوی ۲: روش خوشه‌بندی Hierarchical

یکی دیگر از پرکاربردترین الگوهای خوشه‌بندی، الگوی خوشه‌بندی سلسله مراتبی است. در این روش، برخلاف الگوریتم خوشه‌بندی K-Means، هر مشاهده ممکن است در بیش از یک خوشه قرار گیرد، زیرا خوشه‌ها بر اساس سطوح مختلف فاصله تشکیل می‌شود (جیوردانی و همکاران، ۲۰۲۰)^۱. بنابراین هر خوشه ممکن است زیرمجموعه خوشه دیگر در سطحی از فاصله قرار گیرد. شکل زیر نتیجه یک خوشه‌بندی سلسله مراتبی نمونه را به صورت درختواره نگار^۲ برای متغیرهای ۱، ۲، ۳، ۴، ۵ و ۶ نشان می‌دهد.



شکل ۱. خوشه‌بندی داده‌های فرضی بر اساس الگوی سلسله مراتبی

همان‌طور که در شکل شماره ۱ مشاهده می‌شود، الگوریتم خوشه‌بندی سلسله مراتبی متغیرهای فرضی مذکور را در دو گروه اصلی (۱، ۲ و ۳) و (۴، ۵ و ۶) قرار داده است. همچنین با در نظر گرفتن منطق بالا به پایین^۳ می‌توان متغیرهای فرضی را به چهار گروه (۱ و ۲)، (۳ و ۴)، (۵ و ۶) تقسیم کرد. به‌منظور اجرای محاسبات مربوط به این روش خوشه‌بندی، دو معیار فاصله (شباهت) موردنیاز می‌باشد. ۱- میزان فاصله بین زوج مشاهدات. ۲- میزان فاصله بین خوشه‌ها.

معمولاً قبل از شروع مراحل خوشه‌بندی سلسله مراتبی تجمیعی، از یک ماتریس فاصله^۴ یا ماتریس شباهت^۵ برای تسریع در محاسبات کمک گرفته می‌شود. این ماتریس نشان می‌دهد که فاصله بین هر زوج از مشاهدات به چه میزان است. البته نوع تابعی که فاصله باید به‌وسیله آن اندازه‌گیری شود، در مقادیر موجود در این ماتریس تأثیرگذار است. در خوشه‌بندی سلسله مراتبی تجمیعی یا HAC، با توجه به مقدارهای این ماتریس، مشاهدات یا خوشه‌هایی که دارای کمترین فاصله (بیشترین شباهت) هستند باهم ادغام می‌شوند و خوشه جدیدی می‌سازند. در مرحله بعد مجدداً فاصله بین مشاهدات و یا خوشه‌های جدید، توسط ماتریس فاصله که به‌روزرسانی شده، محاسبه و کار ادغام ادامه پیدا می‌کند تا تنها

¹ Giordani, P., et al (2020)

² Dendrogram

³ Top-Down

⁴ Distance Matrix

⁵ Similarity Matrix

یک خوشه باقی بماند. به منظور استخراج بیشترین شباهت بین داده‌ها می‌بایست از الگوهای مختلف تخمین فاصله و یا شباهت استفاده نمود که در جدول شماره ۱، ۵ الگویی که در محاسبات این پژوهش تخمین زده شده، ارائه شده است.

جدول ۱

معرفی شیوه محاسبه فاصله بین مشاهدات برای داده‌های کمی

ردیف	عنوان الگو	فرمول
۱	Euclidean distance	$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
۲	Manhattan distance	$d_{man}(x, y) = \sum_{i=1}^n x_i - y_i $
۳	Pearson correlation distance	$d_{cor}(x, y) = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$
۴	Spearman correlation distance	$d_{spear}(x, y) = 1 - \frac{\sum_{i=1}^n (\hat{x}_i - \hat{\bar{x}})(\hat{y}_i - \hat{\bar{y}})}{\sqrt{\sum_{i=1}^n (\hat{x}_i - \hat{\bar{x}})^2 \sum_{i=1}^n (\hat{y}_i - \hat{\bar{y}})^2}}$
۵	Kendall correlation distance	$d_{kend}(x, y) = 1 - \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$

برای مطالعه بیشتر در خصوص سایر الگوهای مورد استفاده به (جیوردانی و همکاران، ۲۰۲۰) رجوع شود.

۳.۳ الگوی ۳: روش خوشه‌بندی Hierarchical-K-Means

یکی از محدودیت‌های الگوی K-Means، انتخاب مقادیر اولیه برای مرکز خوشه‌ها می‌باشد. در الگوی ترکیبی مورد استفاده، مزیت‌های دو الگوی K-Means و Hierarchical با یکدیگر ترکیب شده است. پیاده‌سازی این الگو بدین شرح است (کاسامارا، ۲۰۱۷):^۱

- در گام اول خوشه‌بندی بر اساس الگوی Hierarchical صورت گرفته و پس از آن درختواره به K خوشه تقسیم‌بندی می‌شود.
- در گام دوم مراکز هر خوشه (میانگین) مشخص می‌گردد.
- در گام سوم با استفاده از مراکز خوشه‌ها (که در گام دوم توضیح داده شد) الگوی K-Means برآورد می‌گردد.

۴.۳ الگوی ۴: خوشه‌بندی بر مبنای مدل (Model-Based Clustering)

در روش‌های پیشین، فرضیاتی مبنی بر وجود یک توزیع آماری برای داده‌ها وجود نداشت. در روش خوشه‌بندی بر مبنای مدل^۲ یک توزیع آماری برای داده‌ها فرض می‌شود. هدف در اجرای خوشه‌بندی بر مبنای مدل، برآورد پارامترهای توزیع آماری به همراه متغیر پنهانی است که به عنوان برچسب خوشه‌ها در مدل معرفی شده است. با توجه به تعداد خوشه‌هایی که در نظر گرفته شده است، تابع درست‌نمایی برای پیدا کردن خوشه‌ها به صورت زیر تعریف می‌شود (کاسامارا، ۲۰۱۷):

¹ Kassambara, A. (2017).

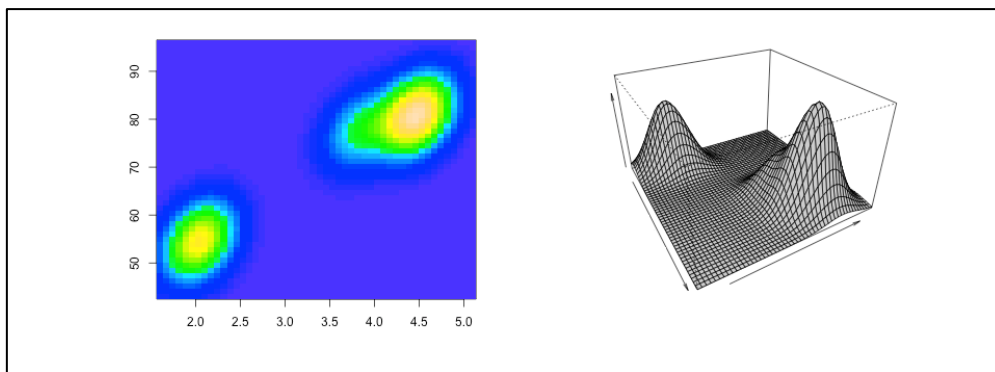
² Model-Based Clustering

$$L_M(\theta_1, \theta_2, \dots, \theta_k; t_1, t_2, \dots, t_k) = \prod_{i=1}^n \sum_{j=1}^k t_j f_j(x_i, \theta_j) \quad (6)$$

در رابطه شماره ۶، t_j بیانگر احتمال تعلق نقطه به خوشه j ام و f_j نیز توزیع آمیخته با پارامترهای θ_j است. در اکثر مواقع به منظور تخمین، از توزیع نرمال آمیخته استفاده می‌شود. در چنین حالتی توزیع آمیخته را به فرمی می‌نویسند که درصد آمیختگی (درصدی از داده‌ها که متعلق به هر توزیع هستند) نیز به‌عنوان پارامتر در مدل حضور دارد.

$$f(x) = \sum_{j=1}^k p_j \Phi(X; \mu_j, \Sigma_j) \quad (7)$$

در رابطه ۷، Φ توزیع نرمال با پارامترهای μ_j و Σ_j برای توزیع j ام هستند و p_j نیز درصد آمیختگی برای توزیع j ام محسوب می‌شود. در شکل شماره ۲ نمونه‌ای از توزیع آمیخته نرمال دومتغیره نمایش داده شده است.



شکل ۲. توزیع آمیخته نرمال دومتغیره

انتخاب‌های مختلفی برای ماتریس واریانس-کوواریانس یا Σ وجود دارد. یکی از این حالت‌ها در نظر گرفتن استقلال و یکسان بودن واریانس برای همه توزیع‌ها است. همچنین می‌توان واریانس هر توزیع را متفاوت از توزیع‌های دیگر در نظر گرفت و بدون شرط استقلال عمل کرد. در حالت اول پیچیدگی مدل و تعداد پارامترهای برآورد شده کم و در حالت دوم پیچیدگی مدل و تعداد پارامترهای برآورد شده زیاد خواهد بود.

برای برآورد پارامترهای چنین مدلی باید از تابع درستنمایی استفاده کرد و با استفاده از عملیات مشتق‌گیری مقدار حداکثر را برای این تابع به دست آورد. بدین‌صورت، پارامترهای مدل نیز برآورد می‌شوند. ولی با توجه به پیچیدگی تابع درستنمایی می‌توان از روش‌هایی ساده و سریع‌تری نیز کمک گرفت. یکی از این روش‌ها استفاده از الگوریتم EM^۱ است. در الگوریتم EM که بر اساس متغیر پنهان عمل می‌کند، می‌توان به حداکثر مقدار تابع درستنمایی همراه با برآورد پارامترها دست یافت. در چنین حالتی می‌توان برچسب تعلق هر مقدار به خوشه را به‌عنوان متغیر پنهان در نظر گرفت. اگر متغیر پنهان را z_{ij} در نظر بگیریم مقداری برابر با ۱ دارد اگر مشاهده j ام در خوشه j ام باشد و در غیر این صورت مقدار آن ۰ خواهد بود.

۴ داده‌ها

در راستای بررسی وضعیت عملکرد بانک‌ها در بازار بین‌بانکی، از نسبت‌های مالی آن‌ها طی در پایان سال مالی ۱۳۹۹ استفاده شده است. نسبت‌های مالی مورد استفاده در این گزارش شامل ۱۵ نسبت متفاوت در چارچوب نسبت‌های نقدینگی، سودآوری، کفایت سرمایه و کیفیت دارایی است که پس از بررسی بیش از ۱۵ نسبت فوق‌الذکر طبقه‌بندی شده است. در

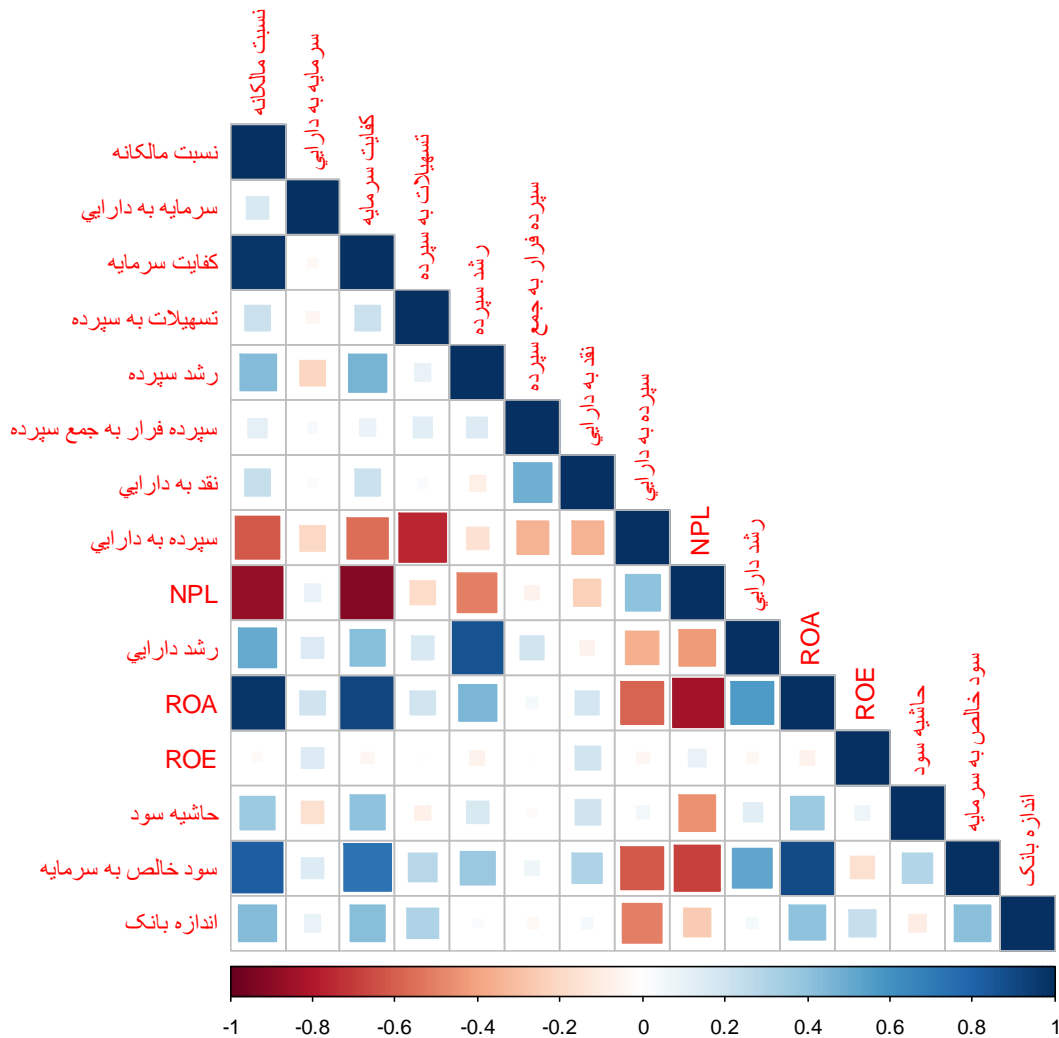
¹ Expectation Maximization

بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

کنترل تورم و ثبات مالی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱

شکل شماره ۳ میزان همبستگی نسبت‌های مورد استفاده و میزان همبستگی آن‌ها با یکدیگر نشان داده شده است. شایان ذکر است بخش عمده‌ای از نسبت‌های مورد مطالعه به دلیل همبستگی بالا با یکدیگر از مدل حذف شده‌اند. نسبت‌های مالی مذکور در جدول ۲ ارایه شده است.



شکل ۳. توزیع آمیخته نرمال دومتغیره

بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

کنترل تورم و ثبات مالی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱

جدول ۲

نسبت‌های مالی مورد استفاده در گزارش

نسبت مالکانه = کل دارایی‌ها/ حقوق صاحبان سهام	کفایت سرمایه
کل دارایی‌ها/ کل سرمایه	
نسبت کفایت سرمایه	
کل سپرده‌ها/ کل تسهیلات	نقدینگی
نرخ رشد سپرده = سپرده اول دوره / (سپرده پایان دوره - سپرده اول دوره)	
کل سپرده‌ها/ سپرده‌های فرار	
مجموع دارایی‌ها/ مجموع دارایی‌های نقدی	
کل دارایی‌ها/ کل سپرده‌ها	کفایت دارایی
نسبت NPL = تسهیلات/ نسبت مطالبات غیر جاری	
نرخ رشد دارایی = دارایی اول دوره / (دارایی پایان دوره - دارایی اول دوره)	
اندازه بانک	سودآوری
بازده دارایی‌ها (ROA) = کل دارایی/ سود خالص	
بازده حقوق صاحبان سهام (ROE) = حقوق صاحبان سهام/ سود خالص	
حاشیه سود = درآمدهای مشاع و غیرمشاع/ سود خالص	
سرمایه/ سود خالص	

علاوه بر این با توجه به عدم ارایه صورت‌های مالی منتهی به ۱۳۹۹/۱۲/۲۹ از سوی برخی از بانک‌های مورد بررسی، تنها امکان بررسی عملکرد ۲۳ بانک فراهم بوده است که در جدول ۳ لیست بانک‌های مورد بررسی ارایه شده است.

جدول ۳

بانک‌های مورد مطالعه در پژوهش و نوع حسابرسی صورت‌های مالی

ردیف	عنوان بانک	نوع حسابرسی
۱	اقتصاد نوین	حسابرسی شده
۲	قرض الحسنه رسالت	حسابرسی شده
۳	ایران زمین	حسابرسی نشده
۴	کارآفرین	حسابرسی شده
۵	خاورمیانه	حسابرسی شده
۶	مسکن	حسابرسی شده
۷	ملت	حسابرسی شده
۸	پارسیان	حسابرسی شده
۹	رفاه کارگران	حسابرسی شده
۱۰	کشاوری	حسابرسی شده
۱۱	شهر	حسابرسی شده
۱۲	پاسارگاد	حسابرسی شده
۱۳	پست بانک	حسابرسی شده
۱۴	صادرات	حسابرسی شده
۱۵	سامان	حسابرسی شده
۱۶	صنعت معدن	حسابرسی شده
۱۷	سرمایه	حسابرسی شده
۱۸	تجارت	حسابرسی شده
۱۹	توسعه صادرات	حسابرسی شده
۲۰	دی	حسابرسی نشده
۲۱	سینا	حسابرسی شده
۲۲	گردشگری	حسابرسی شده
۲۳	ملی	حسابرسی شده

۵ نتایج

به منظور پیاده‌سازی الگوهای مختلف خوشه‌بندی می‌بایست فاصله آن‌ها از یکدیگر یا به عبارت دیگر میزان شباهت یا عدم شباهت پارامترهای متغیرها مورد ارزیابی قرار گیرد. از این رو در گام اول میزان فاصله بانک‌های مورد بررسی بر اساس نسبت‌های مالی آن‌ها ارزیابی شده است.

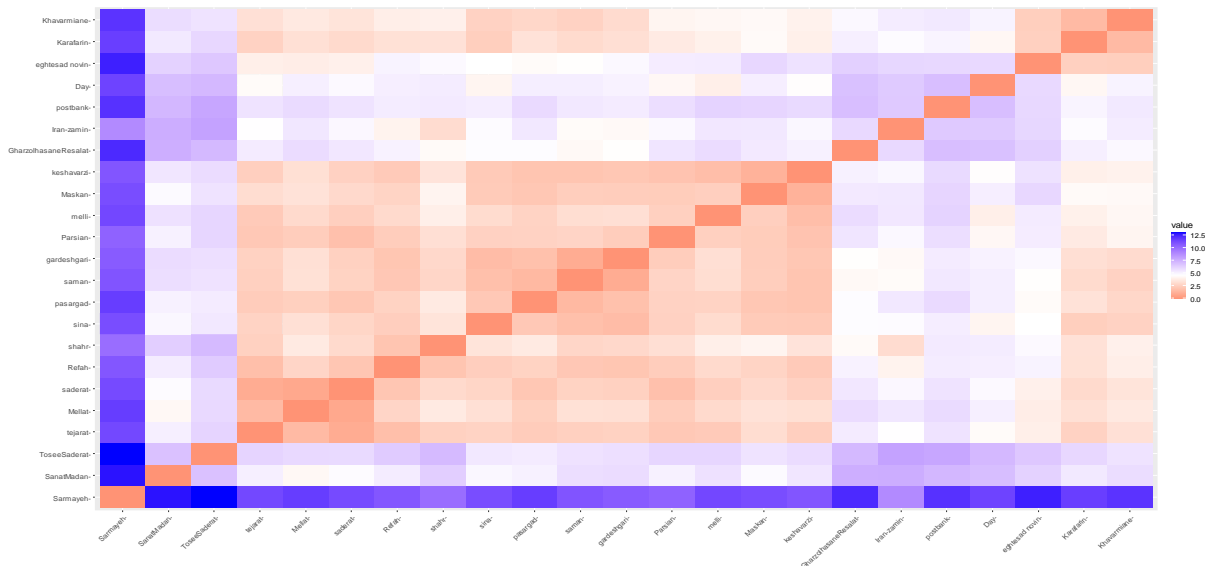
در شکل‌های ۴ میزان فاصله اقلیدسی پارامترهای بانک‌های مورد مطالعه برای داده‌های مذکور، نشان داده شده است. همان‌طور که در شکل‌های مذکور قابل مشاهده است، فاصله پارامترهای برخی بانک‌ها از سایر بانک‌ها بسیار زیاد می‌باشد. لازم به ذکر است فاصله بانک‌های مورد بررسی بر اساس نسبت‌های مالی، با الگوهای مختلفی همچون همبستگی پیرسون و اسپیرمن و معیارهایی چون منهتن و کندال نیز برآورد شده است که با توجه به نتایج مطلوب معیار اقلیدسی، تنها این معیار در ادامه گزارش استفاده شده است.

در گام بعد با استفاده از فاصله‌های برآورد شده، خوشه‌بندی با الگوهای مختلف صورت گرفته است.

بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

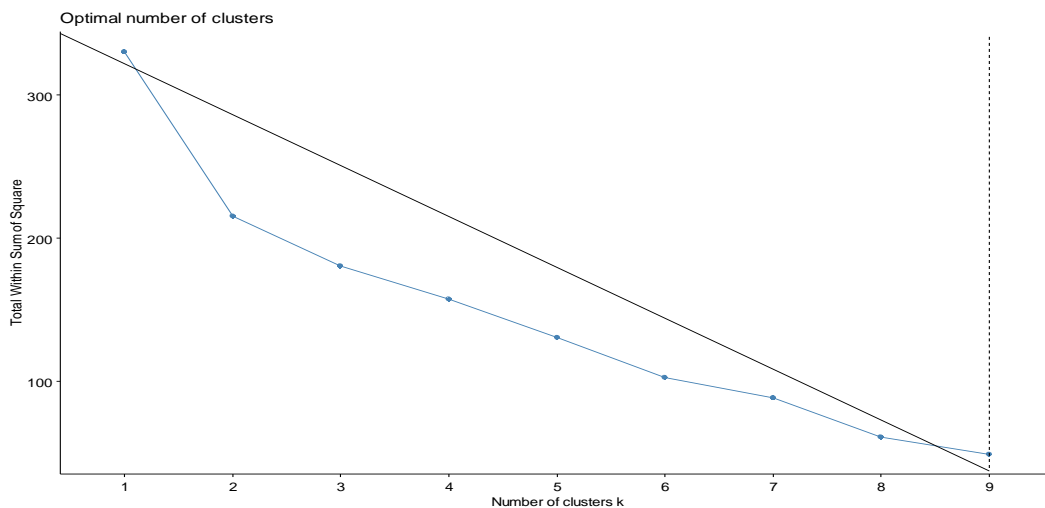
کنترل تورم و ثبات مالی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱



شکل ۴. فاصله اقلیدسی پارامترهای بانک‌های مورد بررسی با توجه به داده‌های منتهی به ۱۳۹۹/۱۲/۳۰

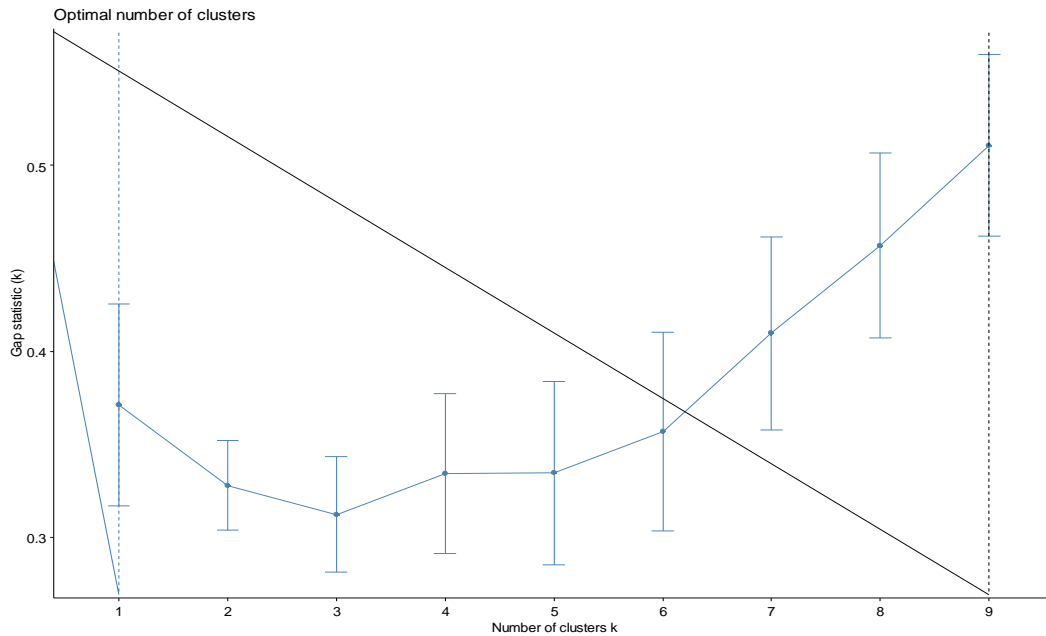
به منظور تقسیم بانک‌ها به خوشه‌های مناسب از الگوی انتخاب تعداد خوشه‌های مختلفی همچون Gap, Elbow, statistic و BIC استفاده شده است. در نمودار شماره ۱ تا ۳ مشخص شده است که بهترین تعداد خوشه‌ها معادل ۹ خوشه است.



نمودار ۱. انتخاب بهترین مقدار خوشه‌ها با معیار Elbow

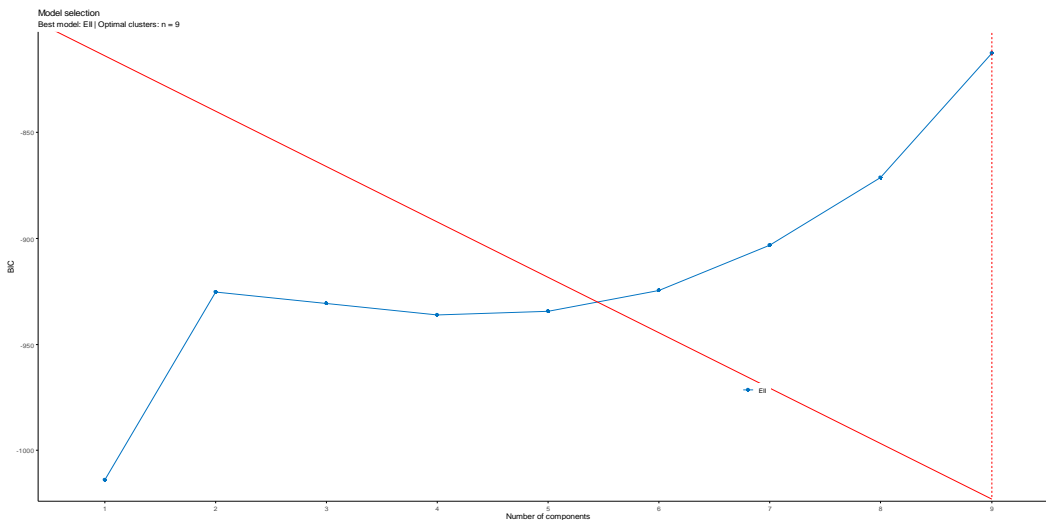
به منظور مطالعه بیشتر در خصوص الگوی تعیین تعداد خوشه بهینه با روش Elbow می‌توان به مقاله همیرا و راسیده (۲۰۲۰)^۱ رجوع کرد.

^۱ Humaira, H., & Rasyidah, R. (2020)



نمودار ۲. انتخاب بهترین مقدار خوشه‌ها با معیار Gap statistic

به‌منظور مطالعه بیشتر در خصوص الگوی تعیین تعداد خوشه بهینه با روش Gap statistic می‌توان به مقاله پاتیل و بایداری (۲۰۱۹)^۱ رجوع کرد.



نمودار ۳. انتخاب بهترین مقدار خوشه‌ها با معیار BIC

به‌منظور مطالعه بیشتر در خصوص الگوی تعیین تعداد خوشه بهینه با روش BIC می‌توان به مقاله ژائو و همکاران (۲۰۰۸)^۲ رجوع کرد.

¹ Patil, C., & Baidari, I. (2019)

² Zhao, Q., et al (2008)

- در این بخش نتایج خوشه‌بندی بانک‌ها بر اساس الگوهای مختلف نشان داده شده است. همان‌طور که نتایج جدول ۴ نشان می‌دهد. تفسیر نتایج برآورد شده به شرح ذیل می‌باشد.
- همان‌طور که در بخش مباحث نظری مطرح شد، از جمله مشکلات الگوهای مرکزی^۱ همچون الگوی k-means این است که بهیچ‌گونه آن وابسته به انتخاب اولیه مراکز بوده و بنابراین بهینه نیست. مشکلات دیگر آن تعیین تعداد خوشه‌ها و صفر شدن خوشه‌ها می‌باشد. به‌طور مثال در نمودار ۴ مشاهده می‌شود که دو بانک توسعه صادرات و صنعت معدن با وجود نزدیکی در شکل در یک خوشه قرار نگرفته‌اند. دلیل آن هم انتخاب اولیه مراکز بوده است که با وجود تکرار الگوسازی در هیچ‌یک از نتایج این دو بانک در یک خوشه قرار نگرفته‌اند. لذا با استفاده از نمودار می‌توان به برخی از ابهامات و اشکالات این الگو فائق آمد. علاوه بر این با توجه به الگوهای مورد اشاره در بخش قبل تعداد خوشه‌ها برابر ۹ در نظر گرفته شده است. باین‌حال هیچ خوشه‌ای بدون داده و صفر نبوده است.
 - در خصوص الگوریتم سلسله مراتبی این نکته حائز اهمیت است که این الگوریتم‌ها یک پارتیشن‌بندی مجموعه داده را ارائه نمی‌دهند، بلکه یک سلسله‌مراتب گسترده‌ای از خوشه‌هایی که در فاصله‌های معینی با یکدیگر ادغام می‌شوند، ارائه می‌دهد. از این‌رو بحث تعداد خوشه‌ها در این الگو مطرح نیست و در این پژوهش به دلیل همخوانی این الگو با سایر الگوها تعداد خوشه‌ها ۹ در نظر گرفته شده است.
 - به‌منظور بهبود نتایج الگوی k-means و غلبه بر مشکل انتخاب مراکز داده‌ها از الگوی ترکیبی^۲ Hierarchical-K-Means استفاده شده است. نتایج این الگو نسبت به k-means بهبود داده شده است و قابلیت قیاس با یکدیگر را دارند.
 - در الگوی Model-Based پس از تخمین ۸ الگوی مختلف از الگوی ترکیبی چندمتغیره^۳ استفاده شده است. به‌منظور مطالعه سایر الگوهای ممکن برای تخمین این الگو می‌توان به (کاسامارا، ۲۰۱۷) رجوع کرد. خوشه‌بندی مبتنی بر توزیع، مدل‌های پیچیده‌ای را برای خوشه‌ها ایجاد می‌کند که می‌تواند همبستگی و وابستگی ویژگی را نشان دهد.
 - بر اساس نتایج الگوهای تخمین زده‌شده، بانک‌های سرمایه‌دی، توسعه صادرات و پست بانک در هر چهار الگوی برآورد شده در هیچ خوشه مشترکی قرار نگرفته‌اند. به عبارتی هرکدام از این بانک‌ها در خوشه‌ای مجزا طبقه‌بندی شده‌اند. دلیل آن نسبت‌های مالی ناپایدار آن‌ها در برخی موارد است که در برخی موارد عملکرد ضعیف و در سایر موارد تفاوت ماهوی نوع فعالیت دلیل آن است.
 - یکی دیگر از خوشه‌های مورد بحث، خوشه پنجم، شامل بانک‌های خاورمیانه، اقتصاد نوین و کارآفرین می‌باشد. به عبارتی در هر ۴ الگوی Hierarchical-K-Means، Hierarchical و K-Means و مدل پایه این سه بانک در یک خوشه قرار گرفته‌اند.
 - بر اساس نتایج در هر ۴ الگوی Hierarchical-K-Means، Hierarchical و K-Means و مدل پایه ۳ بانک صادرات، تجارت و ملت که در پرتفوی صندوق دارا یکم هستند در یک خوشه قرار گرفته‌اند.
 - بانک سرمایه با توجه به نسبت‌های مالی بسیار وخیم فاصله زیادی با سایر بانک‌ها دارد و در نمودارهای ۵ و ۶ کاملاً مشخص است که از ابتدا در خوشه دیگری قرار گرفته است.
 - به نظر می‌رسد معقولانه‌ترین نتایج را الگوی Hierarchical-K-Means ارائه داده است. در این الگو بانک‌های دولتی مسکن و کشاورزی در یک خوشه قرار دارند. همچنین بانک ملی نیز با این دو بانک در ارتباط است.

¹ centroid

² Hybridy

³ multivariate mixture

بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

کنترل تورم و ثبات مالی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱

جدول ۴

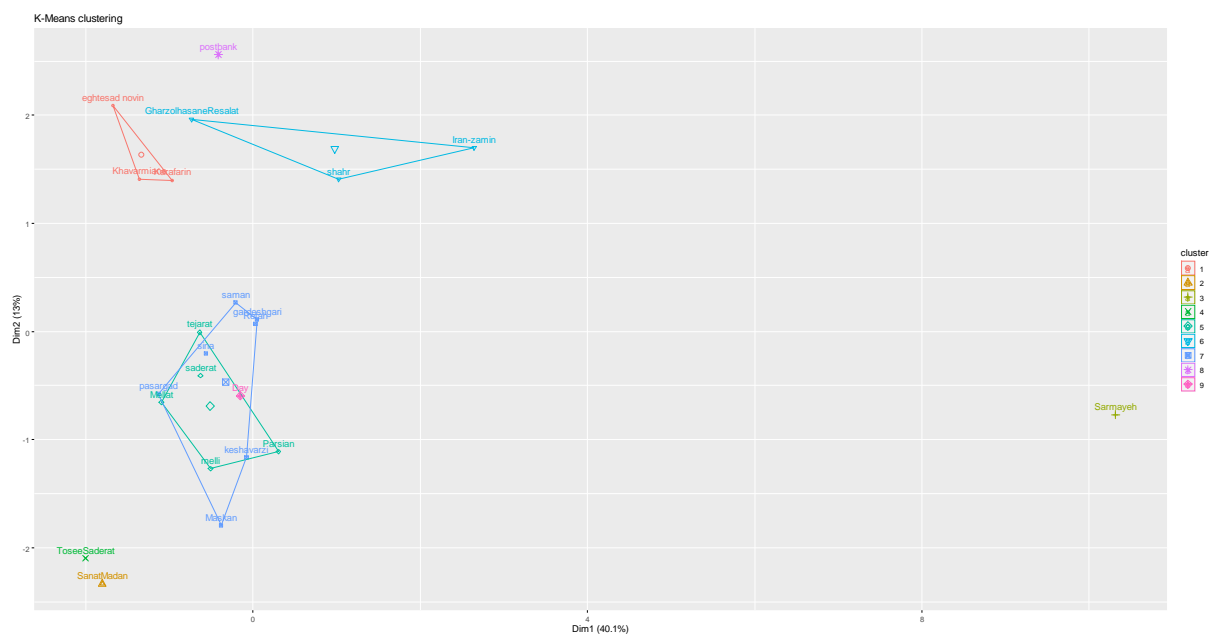
نتایج خوشه‌بندی بانک‌های کشور براساس ۴ الگو

گروه خوشه‌بندی	الگوی K-Means	گروه خوشه‌بندی	الگوی Hierarchical	گروه خوشه‌بندی	الگوی Hierarchical-K-Means	گروه خوشه‌بندی	الگوی مدل پایه
1	توسعه صادرات	۱	توسعه صادرات	۱	توسعه صادرات	۱	قرض الحسنه رسالت
۲	صنعت معدن	۲	صنعت معدن	۲	سرمایه	۲	ایران زمین
۳	سرمایه	۳	سرمایه	۳	دی	۳	پست بانک
۴	پست بانک	۴	پست بانک	۴	پست بانک	۴	صنعت معدن
۵	دی	۵	دی	۵	قرض الحسنه رسالت	۵	سرمایه
۶	اقتصاد نوین	۶	ایران زمین	۶	ایران زمین	۶	توسعه صادرات
	کارآفرین		قرض الحسنه رسالت		دی		
۷	خاورمیانه	۷	اقتصاد نوین	۷	اقتصاد نوین	۷	اقتصاد نوین
	ملت		کارآفرین		کارآفرین		
	پارسیان		خاورمیانه		خاورمیانه		
۸	صادرات	۸	تجارت	۸	صنعت معدن	۸	مسکن
	تجارت		تجارت		ملت		
	ملی		ملت		صادرات		
	قرض الحسنه رسالت		صادرات		رفاه کارگران		
۹	ایران زمین	۹	شهر	۹	سامان	۹	رفاه کارگران
	شهر		گردشگری		کشاورزی		
۹	مسکن	۹	پاسارگاد	۹	پاسارگاد	۹	کشاورزی
	رفاه کارگران		سامان		سامان		
	کشاورزی		گردشگری		گردشگری		
	پاسارگاد		پارسیان		پارسیان		
	سینا		ملی		ملی		
	سامان		مسکن		مسکن		
	گردشگری		کشاورزی		کشاورزی		
گردشگری	پارسیان	پارسیان					
	رفاه	رفاه	ملی				

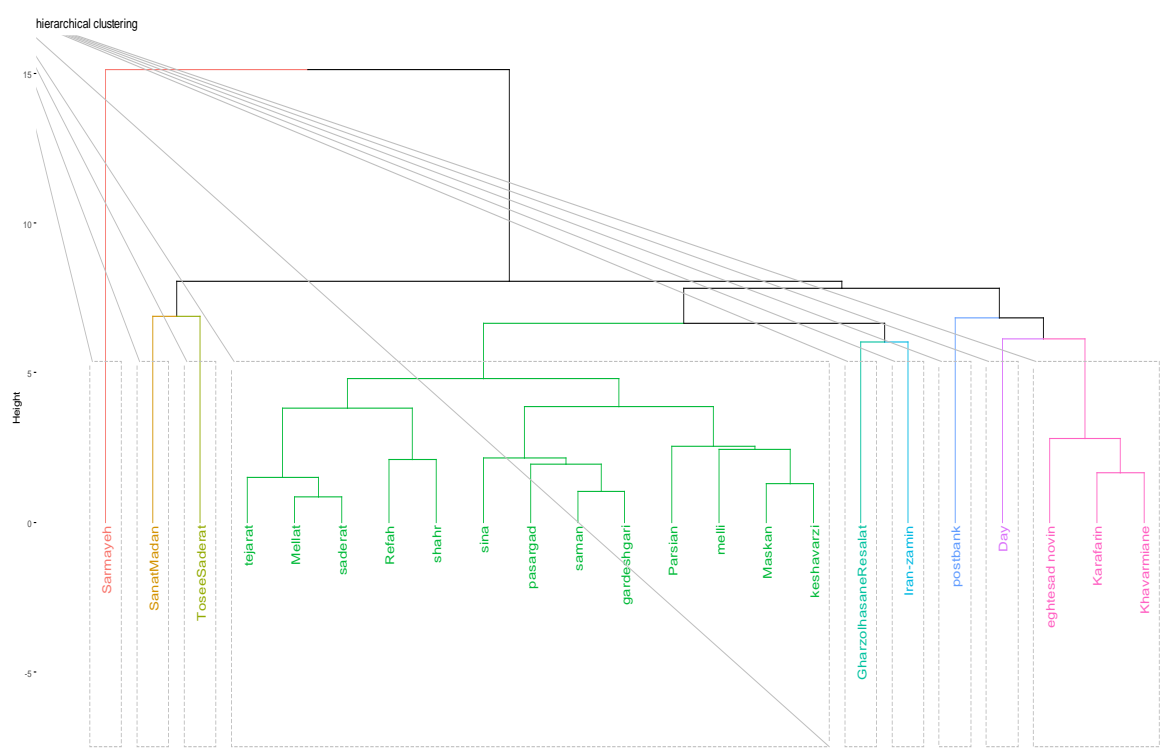
بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

کنترل تورم و ثبات مالی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱



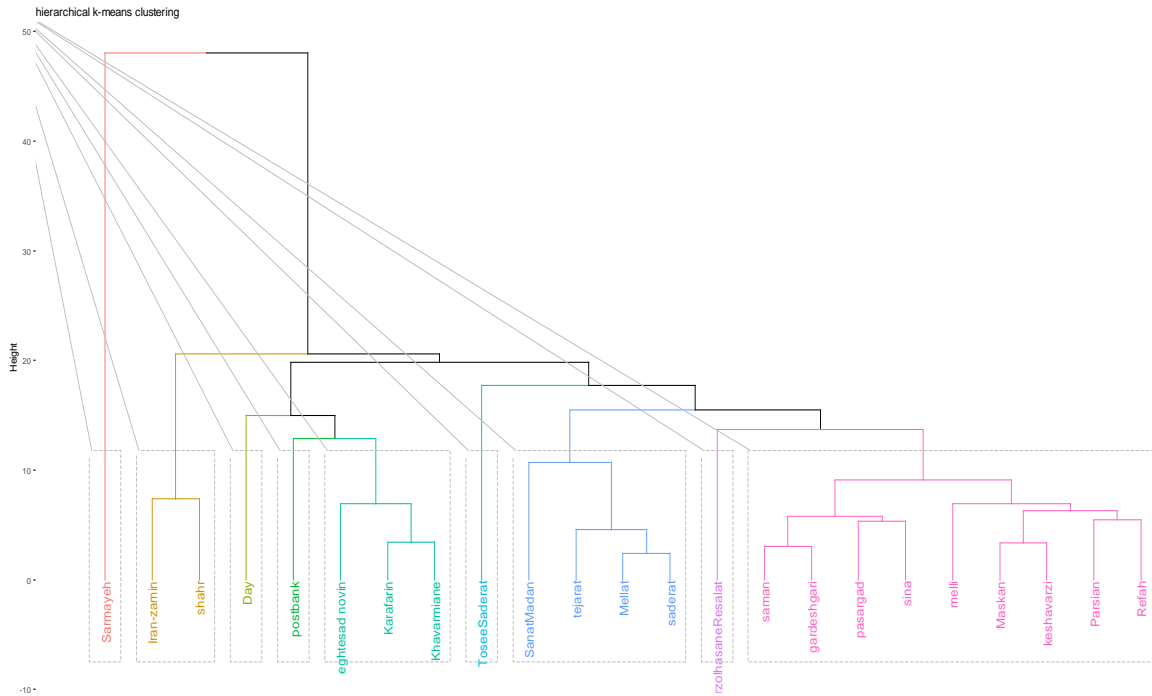
نمودار ۴. نتایج خوشه‌بندی براساس الگوی k-means



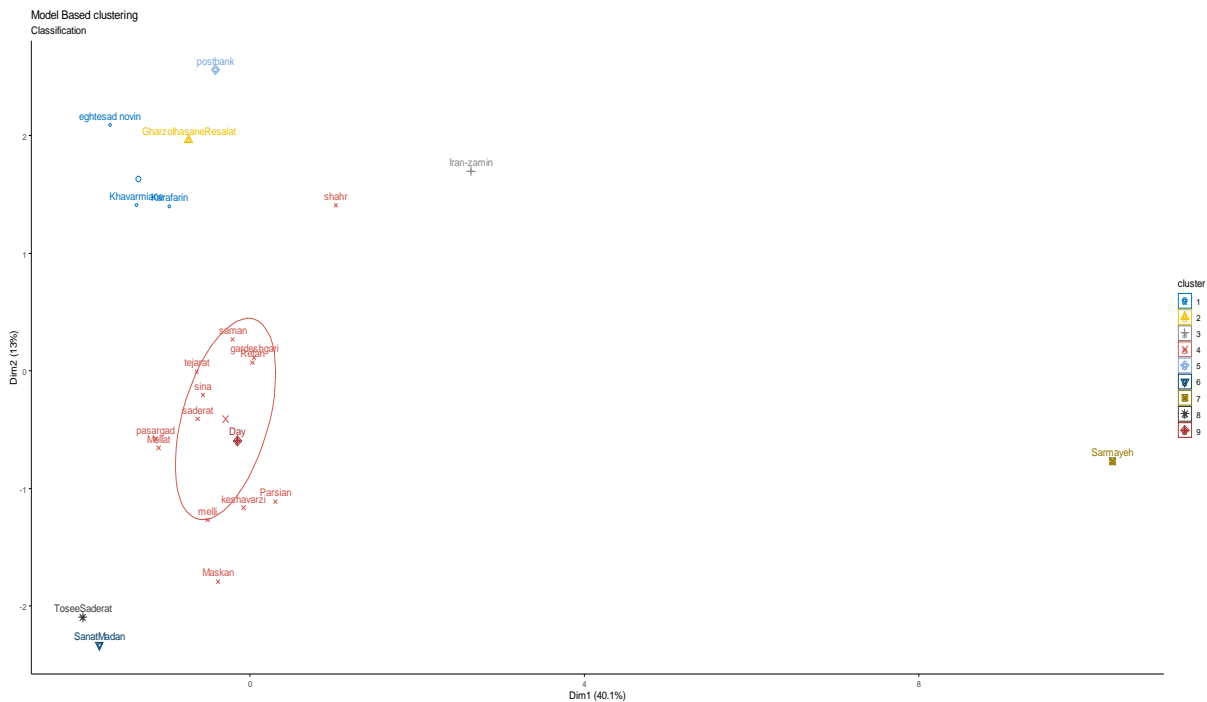
نمودار ۵. نتایج خوشه‌بندی بر اساس الگوی سلسله مراتبی

بیست و نهمین همایش سالانه سیاست‌های پولی و ارزی

تهران، بانک مرکزی جمهوری اسلامی ایران - ۹ و ۱۰ خرداد ۱۴۰۱



نمودار ۶. نتایج خوشه‌بندی براساس الگوی سلسله مراتبی و k-means



نمودار ۷. نتایج خوشه‌بندی براساس الگوی مدل پایه

۶ جمع‌بندی

همان‌طور که ذکر شد بازار بین‌بانکی یکی از مهم‌ترین ابزار تأمین نقدینگی مؤسسات اعتباری محسوب می‌شود. با این حال، با توجه به شرایط نامطلوب اقتصادی و بحران‌های مالی متوالی در پی بحران‌های سیاسی و اقتصادی، فعالیت در این بازار با ریسک‌های بااهمیتی همراه است. به عبارتی مهم‌ترین ریسک فعالیت در این بازار ریسک عدم ایفای تعهدات بانک‌های سپرده‌پذیر می‌باشد. از این‌رو در این پژوهش سعی شد تا با بررسی عملکرد بانک‌های کشور با استفاده از داده‌های استخراج‌شده از صورت‌های مالی آن‌ها، وضعیت نظام بانکی به‌طور کلی ارزیابی گردد.

در این راستا برای آرایه مدلی جامع، از الگوهای یادگیری ماشین نظارت‌نشده (Unsupervised Machine Learning) و به‌طور خاص از الگوهای خوشه‌بندی (Clustering) بهره‌برده شده است. به‌منظور پیاده‌سازی الگوی خوشه‌بندی و ارزیابی نتایج الگوهای مختلف از ۴ الگوی Hierarchical، K-Means، Hierarchical-K-Means و Model Based استفاده شد. تأیید و همگرایی نتایج الگوهای مختلف به‌منظور برآورد الگوریتم خوشه‌بندی یکی از مهم‌ترین دستاوردهای گزارش حاضر به شمار می‌رود.

نتایج الگوهای مختلف برآورد شده نشان داد که در بهترین و بهینه‌ترین حالت، بانک‌های موردبررسی را می‌توان به ۹ خوشه طبقه‌بندی کرد. در واقع نتایج برآورد شده حاکی از اختلاف، فاصله و یا به‌عبارت‌دیگر عدم شباهت بالای بانک‌های موردبررسی می‌باشد. این نکته نیز قابل توجه می‌باشد که ۲۳ بانک موردبررسی به ۹ خوشه متفاوت تقسیم می‌شوند. این مهم نشان می‌دهد نسبت‌های مالی بین بانک‌های مورد مطالعه اختلاف چشمگیری با یکدیگر دارند. نمودارهای ۵ و ۶ به‌طور آشکار فاصله بسیار زیاد بانک‌های سرمایه‌دی، توسعه صادرات و پست‌بانک را با سایر بانک‌های مورد مطالعه نشان می‌دهد. با این حال این نکته می‌بایست در نظر گرفته شود که الگوریتم‌های خوشه‌بندی و رتبه‌بندی تفاوت ماهوی بسیار زیادی با یکدیگر دارند. الگوریتم‌های رتبه‌بندی امکان قضاوت بین متغیرهای مختلف را فراهم می‌آورد. این در حالی است که الگوریتم خوشه‌بندی تنها به دسته‌بندی متغیرها در گروه‌های هم رفتار یا مشابه می‌پردازد. از این‌رو در الگوریتم خوشه‌بندی نمی‌توان انتظار داشت که خوشه‌ها نسبت به یکدیگر برتری داشته باشند. با این وجود به نظر می‌رسد بانک‌های سرمایه‌دی، ایران‌زمین، شهر و دی که بیشترین فاصله را از میانگین نسبت‌های مالی سایر بانک‌ها دارند وضعیت بحرانی‌تری را داشته باشند. با این وجود در مورد برتری سایر خوشه‌ها می‌توان اظهار داشت خوشه بانک اقتصاد نوین، کارآفرین و خاورمیانه بهترین عملکرد را در نظام بانکی دارند.

فهرست منابع

- ارضاء، ا و قاسم پور، ش (۱۳۹۶)، رتبه‌بندی بانک‌های خصوصی ایران براساس مدل کملز با استفاده از رویکرد ترکیبی فرایند تحلیل سلسله مراتبی و آراس، فصلنامه راهبرد مدیریت مالی، دوره ۵، پاییز، ص ۹۹-۱۱۸.
- خاتمی فیروزآبادی، م؛ سجادی، خ؛ تقوی فرد، م و بامداد صوفی، ج (۱۳۹۶)، خوشه‌بندی و رتبه‌بندی مشتریان بانک به کمک مدل RFM-LD، کنفرانس بین‌المللی مدیریت صنعتی، دوره ۲.
- سلیمی، م؛ ابراهیمی سروعلیا، م؛ قاسم پور، ش (۱۳۹۵)، طراحی مدل بومی رتبه‌بندی بانک‌های ایرانی بر مبنای سلامت بانکی، نشریه تحقیقات مالی، دوره ۱۸، ص ۶۵۳-۶۷۴.
- علیزاده زوارم، ع و پویا، علیرضا (۱۳۹۶)، ارزیابی و خوشه‌بندی بانک‌ها و مؤسسات مالی و اعتباری ایران بر اساس شاخص‌های ترافیکی وبسایت، پژوهش‌های مدیریت منابع سازمانی، دوره ۷، بهار، ص ۱۸۹-۲۰۶.
- یوسفی، م (۱۳۹۷)، رتبه‌بندی بانک‌های کشور براساس شاخص‌های مالی و با به‌کارگیری رویکرد ترکیبی AHP فازی و TOPSIS، نشریه پژوهش‌های پولی بانکی، دوره ۱۱، ص ۲۵-۵۴.

Dardac N. - Boitan A. I. (2009): A Cluster Analysis Approach for Bank's Risk Profile: The Romanian Evidence. *European Research Studies Journal*, 1, pp. 109-118.

Ercan, H., & Sayaseng, S. (2016). The cluster analysis of the banking sector in Europe. *Economics and Management of Global Value Chains*, 111-127.

- Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics*, 21, 768–769.
- Giordani, P., Ferraro, M. B., & Martella, F. (2020). *An introduction to clustering with R*. Springer Singapore.
- Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28, 100–108. Doi: 10.2307/2346830.
- Humaira, H., & Rasyidah, R. (2020). Determining the Appropriate Cluster Number Using Elbow Method for K-Means Algorithm. *WMA-2*.
- Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). Sthda.
- Lagasio, V., & Quaranta, A. G. (2021). Cluster analysis of bank business models: the connection with performance, efficiency and risk. *Finance Research Letters*, 102640.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129–137.
- Lloyd, S. P. (1957). Least squares quantization in PCM. Technical Note, Bell Laboratories. Published in 1982 in *IEEE Transactions on Information Theory*, 28, 128–137.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.
- Molin, F. (2017). Cluster analysis of European banking data.
- Patil, C., & Baidari, I. (2019). Estimating the optimal number of clusters k in a dataset using data depth. *Data Science and Engineering*, 4(2), 132-140.
- Pérez-Ortega, J., Almanza-Ortega, N. N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., & Martínez-Rebollar, A. (2019). The k-means algorithm evolution. In *Introduction to Data Science and Machine Learning*. IntechOpen.
- Zhao, Q., Hautamaki, V., & Fränti, P. (2008, October). Knee point detection in BIC for detecting the number of clusters. In *International conference on advanced concepts for intelligent vision systems* (pp. 664-673). Springer, Berlin, Heidelberg.