فصلنامه پژوهش‌های اقتصادی ایران/ سال بیستم/ شماره ۶۴/ پاییز ۱۳۹۴/ صفحات ۵۴-۲۷

برآورد یک سیستم چند متغیره تصادفی با استفاده از روش‌های داده‌کاوی
(مطالعه موردی: وجه نقد مورد نیاز شعب بانک تجارت)

فرزاد اسکندری[۱]

غزاله باغبانی[۲]

تاریخ ارسال: ۱۳۹۴/۰۳/۰۵ تاریخ پذیرش: ۱۳۹۴/۱۰/۲۰

چکیده:

در حال حاضر، بانک‌ها به صورت روزانه با چالش کفایت وجه نقد جهت پاسخگویی به مشتریان و نیز عدم تمایل به افزایش هزینه‌های ناشی از نقل و انتقال مازاد وجه نقد شعبه مواجه هستند. به همین علت، موضوع برآورد مانده وجه نقد صندوق شعب -با توجه به عملیات روزانه آن- که به عنوان یک سامانه چند متغیره محسوب می‌شود، از موارد با اهمیت در حوزه بانکداری بشمار می‌آید. در این راستا، استفاده از روش‌های داده کاوی و به خصوص روش‌های خوشه‌بندی و شبکه عصبی می‌تواند به افزایش دقت برآورد پارامتر وجه نقد مورد نیاز شعب کمک کند. شبکه‌های عصبی از لحاظ انعطاف‌پذیری، غیرخطی بودن، تحمل بیشتر نوفه‌ها و نیز وابسته نبودن به فرضیه‌های اولیه درباره داده‌های ورودی در این زمینه حائز اهمیت هستند. در این مقاله، ۲۰ شعبه بانک تجارت در بازه زمانی ۹۳/۲/۱ تا ۹۳/۶/۳۱ با توجه به تنوع بین شعب از لحاظ درجه شعبه، نوع شعبه از لحاظ سپرده‌ای یا تسهیلاتی، تعداد دستگاه خودپرداز در شعبه، شعبه کشیک/غیرکشیک در خوشه‌های متشابه دسته‌بندی شده، سپس با در نظر گرفتن نتایج حاصل از خوشه‌بندی و متغیرهای مرتبط با وجه نقد شعبه شامل متغیرهای تقویمی مانند روزهای هفته، روزهای پرداخت حقوق/واریز یارانه/واریز سود سپرده‌ها، روزهای تعطیل و مناسبت‌های رسمی و نیز متغیر میزان وجه نقد مصرفی دستگاه خودپرداز شعبه، ساختار مناسب شبکه عصبی برای برآورد وجه نقد شعب از طریق معیارهای خطا، تعیین شده و وجه نقد شعب در خوشه‌های مختلف برآورد می‌شود. نتایج تحقیق نشان می‌دهد شبکه عصبی با لحاظ نتایج خوشه‌بندی با میانگین قدر مطلق خطای ۵ درصد می‌تواند عملکرد خوبی جهت برآورد وجه نقد شعب در خوشه‌های مختلف ارائه دهد.

واژه‌های کلیدی: الگوی بانکداری، برآورد، خوشه‌بندی، داده‌کاوی، شبکه عصبی

طبقه‌بندی JEL: *C45، C13، G21*

۱- دانشیار و عضو هیات علمی دانشگاه علامه طباطبائی ، گروه آمار- نویسنده مسئول ffeskandari@yahoo.com

۲- دانشجوی دکترای آمار دانشگاه علامه طباطبائی ghazaleh_stat@yahoo.com

۱- مقدمه

تعیین رابطه بین مجموعه‌ای از متغیرهای کمی و کیفی در یک سامانه تصادفی چند متغیره و نیز ارزیابی اثرپذیری متغیرها بر متغیر پاسخ از طریق الگوهای پیشنهادی از جمله موارد با اهمیت در مباحث و مسایل آماری محسوب می‌شود. در همین زمینه و به خصوص در حوزه بانکداری، یکی از دغدغه‌های مهم در تصمیم‌گیری‌های مدیران و به خصوص روسای شعب، موضوع کفایت وجه نقد در شعب است، چراکه کمبود وجه نقد روزانه در صندوق شعبه منجر به عدم پاسخگویی به نیاز مشتری و به تبع آن نارضایتی مشتری خواهد شده و نیز مازاد وجه نقد در شعبه باعث افزایش هزینه بیمه وجوه و انتقال آن به خزانه بانک و افزایش هزینه خواب پول و از دست دادن فرصت سرمایه‌گذاری برای بانک و به تبع آن عدم امکان حصول درآمد و سود برای بانک است، بنابراین بانک‌ها همواره درصدد بهینه کردن مانده وجوه نقد صندوق شعب با توجه به عملیات روزانه آن هستند.

در حال حاضر، تامین وجه نقد مورد نیاز برای شعب در ابتدای همان روز کاری و تنها بر مبنای تجربه روسای شعب صورت می‌گیرد که گاهی با عملکرد واقعی آن روز فاصله داشته و می‌تواند منجر به افزایش نقل و انتقال پول از طریق ماشین‌های پول‌رسان مناطق شده که هزینه بانک را در این خصوص افزایش می‌دهد و یا حتی گاهی در صورت عدم تامین به موقع وجه نقد مورد نیاز منجر به نارضایتی مشتریان می‌شود. به همین منظور، لازم است تا با روشی مناسب و کارا وجه نقد مورد نیاز شعب برای دوره زمانی موردنظر در آینده، برای مثال برای روز کاری بعد، برآورد شود تا از مشکلات عنوان شده جلوگیری به عمل آید.

تاکنون از روش‌های مختلفی مانند سری‌های زمانی و شبکه‌های عصبی کلاسیک برای پیش‌بینی وجه نقد مورد نیاز شعب بانک‌ها استفاده شده است به طوری‌که در تحقیقی که توسط کاردونا و آمایا (۲۰۱۲) در بانک کلمبیا انجام شده، ابتدا با دو روش سری زمانی و شبکه عصبی کلاسیک، پیش‌بینی وجه نقد در شعب بانک انجام و سپس با مدل بهتر انتخاب شده با روش برنامه‌ریزی خطی به مینیمم کردن تابع هزینه پرداخته شده است.

در پژوهشی نیز که توسط سندی‌پان (۲۰۱۰) برای شعب بانک *ICICI* در هند صورت گرفته است، با استفاده از روش سری زمانی به پیش‌بینی وجه نقد در شعب پرداخته و سپس با روش برنامه‌ریزی خطی به بهینه‌سازی اقدام شده است.

پرم‌چند و والیا (۲۰۰۶) نیز در پیش‌بینی وجه نقد شعبه بانک دولتی هند با دو روش سری زمانی و شبکه عصبی به پیش‌بینی وجه نقد شعبه پرداخته‌اند. آنها در تحقیق خود تنها از متغیرهای تقویمی برای پیش‌بینی وجه نقد استفاده کرده‌اند.

علاوه بر این، در بررسی وضعیت تقاضای وجه نقد که توسط نوربخش و همکاران (۱۳۸۹) برای دو شعبه نمونه در بانک کارآفرین صورت پذیرفته از دو روش سری زمانی و نیز مدل انتشار استفاده شده است که با استفاده از مدل انتشار و انجام شبیه‌سازی برای توابع توزیع مقادیر واریزها و برداشت‌ها، سقف بهینه پول در صندوق شعبه نمونه با کاهش ۲۵ درصدی مواجه شد. درهمین راستا، سعیدی و شعبانی(۱۳۸۹) در تحقیقی به بررسی نحوه بکارگیری مدلی در پیش‌بینی وضعیت نقدینگی در شعب بانک‌ها با استفاده از شاخص لاندای امری و با هدف یافتن بهترین شرایط پیش‌بینی صحیح روزانه وجه نقد در یکی از شعب بانک ملی پرداخته‌اند.

بهشتی (۱۳۸۹) در تحقیق خود پیش‌بینی وجه نقد شعب بانک‌ها را با روش شبکه عصبی مورد بررسی قرار داده است. در این تحقیق، متغیرهایی برای پیش‌بینی و به عنوان ورودی مدل تعیین شده و فرآیند پیش‌بینی با استفاده از روش‌های اقتصادسنجی و تعریف تابع پیش‌بینی و بکارگیری مدلی با استفاده از شبکه عصبی پرسپترون چند لایه انجام گرفته است. او در تحقیق خود نشان داده است که روش شبکه عصبی نسبت به روش‌های دیگر برتری داشته و دارای دقت بالاتری برای پیش‌بینی است. متغیرهای ورودی در شبکه عصبی، میزان برداشت مشتریان از حساب‌های جاری، قرض‌الحسنه و مدت‌دار بوده و متغیر خروجی نیز پیش‌بینی وجه نقد شعب در نظر گرفته شده است. او در تحقیق خود متغیرهای تقویمی را که بر برداشت وجه نقد مشتریان موثر هستند، در نظر نگرفته است.

روش‌های سری‌های زمانی از دقت بالای پیش‌بینی برخوردار نبوده و نسبت به تغییرات ناگهانی در متغیرها حساس هستند. سری‌های زمانی بر دو فرض ایستایی و خطی بودن داده ها استوار هستند. فرض ایستایی در میانگین به این معنی است که سری‌های ایستا در بلندمدت در اطراف میانگین ثابت خود نوسان می کنند و همچنین واریانس این سری ثابت و محدود بوده و تابع زمان نیست و با گذشت زمان تغییر نمی‌یابد که در این‌صورت ایستا در واریانس نامیده می‌شوند. با وجود این فروض اولیه، گاهی داده‌های موجود، آن فرض‌ها را رعایت نمی کنند که بر دقت پیش‌بینی اثرگذار خواهند بود.

با وجود آنکه روش شبکه‌های عصبی استاندارد در تعیین متغیرهای ورودی و نیز پیچیدگی مدل بهینه و ایجاد مدلی با پیچیدگی مناسب مشکل است، اما نسبت به سایر روش‌ها عملکرد بهتری داشته و از دقت بیشتری برخوردار هستند. غیرخطی بودن شبکه‌های‌عصبی، تحمل بیشتر نوفه‌ها، انعطاف‌پذیری و نیز وابسته نبودن آنها به فرضیه‌های اولیه درباره داده‌های ورودی از دیگر مزیت‌های استفاده از شبکه‌های عصبی است (هو و همکاران،۱۹۹۸).

علاوه بر این، شبکه‌های عصبی قادر به تخمین هر تابع پیوسته با هر دقت موردنظر هستند که از ویژگی‌های آنها بشمار می‌آید (هورنیک و همکاران،۱۹۸۹؛ آیری و می‌یاک،۱۹۸۸). همچنین امکان تعمیم در شبکه‌های عصبی از مهم‌ترین مزیت‌های آنها است به‌طوری‌که شبکه‌های عصبی پس از یادگیری از داده‌هایی که برای آموزش شبکه مورد استفاده قرار می گیرند، قادر به استنباط داده‌های مشاهده نشده آتی هستند، حتی اگر داده‌های آموزش شبکه دارای نوفه باشند. (هو و همکاران،۱۹۹۸).

این مقاله با استفاده از روش‌های داده‌کاوی به برآورد وجه نقد شعب بانک تجارت می‌پردازد. ابتدا، با توجه به آنکه شعب بانک از لحاظ ویژگی و برخی دیگر از متغیرها که در مقاله به آن اشاره خواهد شد، متفاوت هستند به خوشه‌بندی شعب پرداخته تا شعب با ویژگی‌های مشابه در خوشه‌ها دسته‌بندی شوند، سپس با روش شبکه عصبی برای خوشه‌ها، وجه نقد مورد نیاز شعب برآورد خواهد شد.

بخش دوم مقاله به بیان مساله و روش تحقیق می‌پردازد. بخش سوم به بررسی داده‌کاوی و تکنیک‌های مورد استفاده در مقاله اختصاص دارد. بخش چهارم به انجام محاسبات و پیاده‌سازی روش‌های مورد بحث بر داده‌های موجود و یافته‌های تحقیق می‌پردازد. بخش پنجم نیز به نتیجه‌گیری و پیشنهادات اختصاص یافته است.

۲- بیان مساله و روش تحقیق

وجود وجه نقد کافی از الزامات و موارد با اهمیت در هر شعبه‌ای محسوب می‌شود، زیرا در صورت عدم کفایت وجه نقد، شعبه امکان پاسخگویی نیاز مشتریان را ندارد که همین موضوع به نارضایتی مشتریان می‌انجامد. همچنین مازاد وجه نقد شعبه در پایان روز کاری، براساس قانون بیمه بانک منجر به الزام انتقال پول به خزانه بانک شده که هزینه استهلاک پول و ماشین‌های حمل پول را افزایش می‌دهد. علاوه بر این، عدم پیش‌بینی و برآورد صحیح از مقدار بهینه نگهداری وجه نقد در شعب بانک، می‌تواند فرصت‌های اعطای تسهیلات بیشتر را که باعث حصول درآمد برای بانک می‌شود از بانک سلب کند. بنابراین برآورد وجه نقد مورد نیاز هر روز کاری شعبه برای مسئول شعبه جهت تامین رضایت مشتری و همچنین کاهش هزینه‌های بانک بابت حمل‌ونقل پول از خزانه به شعبه از اهمیت بسزایی برخوردار است.

جامعه آماری در این مقاله، شعب بانک تجارت در شهر تهران هستند که با توجه به تنوع شعب در هر یک از مدیریت‌های شعب بانک، از نمونه‌گیری خوشه‌ای تصادفی جهت انتخاب شعب نمونه استفاده شده است.

با توجه به گستردگی و تنوع شعب بانک، لازم است تا شعب براساس متغیرهایی خوشه‌بندی شوند و وجه نقد مورد نیاز شعب در خوشه‌ها با توجه به شباهت‌های درون‌گروهی آنها برآورد شود. برای این منظور، ابتدا با تعریف متغیرهایی از قبیل درجه شعبه، نوع شعبه از لحاظ سپرده‌ای یا تسهیلاتی، تعداد دستگاه خودپرداز در شعبه و شعبه کشیک/ غیرکشیک به خوشه‌بندی شعب پرداخته می‌شود تا شعب براساس شباهت‌ها

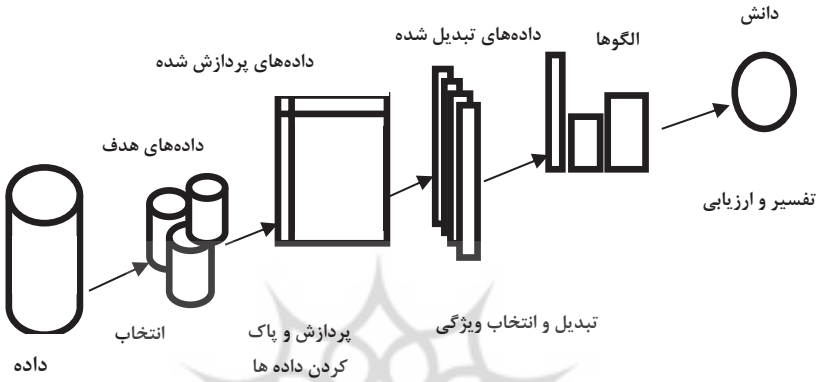
خوشه‌بندی شوند. در این خصوص، شعب بانک تجارت بر اساس برخی شاخص‌ها ازجمله سود/ زیان شعبه، میزان سپرده، ارائه انواع خدمات، میزان تسهیلات و مطالبات و... به درجات ممتاز الف، ممتاز ب، یک، دو، سه، چهار و پنج تقسیم‌بندی می‌شوند. همچنین شعب کشیک تا ساعت ۱۶:۳۰ فعالیت دارند و به تبع آن ممکن است نیاز به وجه نقد بیشتری داشته باشند. علاوه بر این، شعب غیرکشیک تا ساعت ۱۴:۳۰ فعالیت دارند و ممکن است نیاز به نقدینگی کمتری داشته باشند. نوع شعبه از لحاظ سپرده‌ای یا تسهیلاتی نیز براساس میانگین سالانه بدهکاری یا بستانکاری شعب به حساب مرکز تعیین شده است؛ اگر میانگین مانده بدهکاری شعبه به حساب مرکز از میانگین مانده بستانکاری شعبه به حساب مرکز بیشتر باشد، شعبه سپرده‌ای و درغیر این‌صورت، تسهیلاتی در نظر گرفته می‌شود که شعب سپرده‌ای معمولاً به دلیل جذب بیشتر سپرده، نیاز کمتری به وجه نقد خواهند داشت. در ضمن هرچه تعداد دستگاه‌های خودپرداز در شعبه بیشتر باشد به دلیل نیاز به پول‌گذاری دستگاه‌ها، وجه نقد مصرفی شعبه بیشتر خواهد بود.

در گام بعد، با مدنظر قرار دادن نتایج حاصل از خوشه‌بندی و شماره خوشه شعب، متغیرهای روزهای هفته (شامل ترتیب روز کاری و چندمین روز هفته)، روزهای پرداخت حقوق/ واریز یارانه/ واریز سود و نیز روزهای تعطیل و مناسبت‌های رسمی و نیز متغیر میزان وجه نقد مصرفی دستگاه خودپرداز شعبه که بر وجه نقد مورد نیاز شعبه تاثیرگذار هستند به عنوان متغیرهای ورودی در نظر گرفته شده‌اند تا وجه نقد مورد نیاز شعبه برآورد شود. داده‌های مربوط به متغیرهای نامبرده در بازه زمانی ۹۳/۲/۱ تا ۹۳/۶/۳۱ برای ۲۰ شعبه بانک، استخراج و مورد استفاده قرار گرفته است. همچنین خوشه‌بندی داده‌ها و برآورد وجه نقد به روش شبکه عصبی از طریق برنامه‌نویسی در نرم‌افزار R انجام شده است.

۳- داده‌کاوی

داده‌کاوی به مفهوم استخراج اطلاعات نهان در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ به منظور یافتن الگوها و روابط مفید بین داده‌ها است. به عبارت

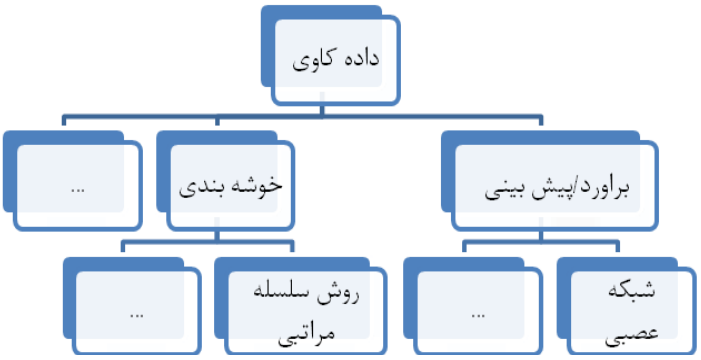
دیگر، داده‌کاوی فرآیندی است که برای استخراج اطلاعات مفید و در پی آن کسب دانش از پایگاه‌های بزرگ داده‌ها از روش‌های آماری، ریاضی، هوش مصنوعی و یادگیری ماشین استفاده می‌کند (توربان و همکاران،۲۰۰۷). با توجه به شکل (۱)، هدف اصلی داده‌کاوی کمک به فرآیند تصمیم‌گیری از طریق استخراج دانش از داده‌ها است.



شکل (۱)- فرآیند کلی کشف دانش و جایگاه داده‌کاوی در آن

با توجه به آنکه داده‌کاوی از روش‌های آمار ناپارامتری استفاده می‌کند، امکان بکارگیری آن برای انواع داده‌ها وجود دارد که از مزیت‌های آن بشمار می‌رود، چراکه داده‌های جمع‌آوری شده اغلب خیلی از فرض‌های قدیمی آماری (از قبیل داشتن توزیع مشخص و...) را درنظر نمی‌گیرند و عدم رعایت آنها می‌تواند صحت و درستی نتایج تحقیق را دچار مشکل کند.

داده‌کاوی شامل عملیات و تکنیک‌های متعددی جهت رده‌بندی، خوشه‌بندی، پیش‌بینی رفتار داده و تخمین/ برآورد است که با توجه به داده‌ها و هدف موردنظر انتخاب می‌شوند (السن و دلن،۲۰۰۸). نمودار (۱)، تکنیک‌ها و الگوریتم‌های مرتبط با داده‌کاوی را که در این تحقیق مورد استفاده قرار گرفته است، نشان می‌دهد که در بخش‌های۳-۱ و ۳-۲ به تشریح آنها پرداخته خواهد شد.



نمودار (۱)- تکنیک‌های داده‌کاوی مورد استفاده در تحقیق

۳-۱- خوشه‌بندی

خوشه‌بندی یکی از تکنیک‌های پرکاربرد در داده‌کاوی است که به یافتن یک ساختار درون مجموعه‌ای از داده‌های بدون برچسب پرداخته و از نوع یادگیری بدون نظارت محسوب می‌شود. به‌عبارت دیگر، خوشه‌بندی فرآیند گروه‌بندی مجموعه‌ای از مشاهدات به کلاس‌هایی مشابه است. اعضای درون یک خوشه شباهت‌های زیادی به یکدیگر داشته، اما اعضای هر خوشه با اعضای موجود در خوشه‌های دیگر شباهت خیلی کمتری دارند (هان و کامبر،۲۰۰۶؛ برامر،۲۰۰۷؛ ونگ،۲۰۰۸).

بکارگیری روش خوشه‌بندی، زمانی که هیچ نوع اطلاعات قبلی از ساختار داخلی پایگاه داده‌ها وجود ندارد و نیز امکان استفاده از آن برای انواع گوناگون داده‌ها ، از جمله مزایای استفاده از روش خوشه‌بندی است. همچنین در خوشه‌بندی، نیازی به داده‌های آموزش نیست (السن و دلن،۲۰۰۸). با این حال، حساس بودن این روش به پارامترهای اولیه از قبیل تعداد خوشه‌ها و اندازه‌های دقیق فواصل نزدیکی از محدودیت‌های خوشه‌بندی محسوب می‌شود.

مراحل خوشه‌بندی شامل انتخاب معیار شباهت یا نزدیکی مشاهدات، انتخاب روش تجزیه و تحلیل خوشه‌ای، تصمیم‌گیری در خصوص تعداد خوشه‌ها و تفسیر خوشه‌های ایجاد شده است.

۳-۱-۱- روش سلسله مراتبی

از جمله روش‌های پرکاربرد در خوشه‌بندی، روش سلسله مراتبی[1] است. در روش خوشه‌بندی سلسله مراتبی به خوشه‌های نهایی بر اساس میزان عمومیت آنها، ساختاری سلسله مراتبی معمولا به صورت درختی نسبت داده می‌شود که به آن، درخت سلسله مراتبی دندوگرام[2] می‌گویند.

روش‌های خوشه‌بندی بر اساس ساختار سلسله مراتبی تولیدی توسط آنها معمولا به دو دسته بالا به پایین یا تقسیم‌کننده[3] و پایین به بالا یا متراکم‌شونده[4] تقسیم می‌شوند. در روش بالا به پایین، ابتدا تمام داده‌ها به عنوان یک خوشه در نظر گرفته می‌شوند و سپس در طی یک فرآیند تکراری در هر مرحله داده‌هایی که شباهت کمتری به هم دارند به خوشه‌های مجزایی شکسته می‌شوند و این فرآیند تا رسیدن به خوشه‌هایی که دارای یک عضو هستند، ادامه پیدا می‌کند. در روش پایین به بالا یا متراکم‌شونده نیز ابتدا هر یک از داده‌ها به عنوان خوشه‌ای مجزا در نظر گرفته می‌شود و طی فرآیندی تکراری در هر مرحله خوشه‌هایی که شباهت بیشتری با یکدیگر ترکیب می‌شوند تا در نهایت یک خوشه و یا تعداد مشخصی خوشه حاصل شود (برامر،۲۰۰۷).

تفاوت اصلی در بین این روش‌ها به انتخاب معیار شباهت و نحوه محاسبه شباهت بین خوشه‌ها مربوط می‌شود. به عنوان مثال، در الگوریتم نزدیک‌ترین همسایه برای محاسبه شباهت بین دو خوشه A و B از معیار $d_{AB} = \min_{i \in A, j \in B} d_{ij}$ استفاده می‌شود که i یک مشاهده متعلق به خوشه A و j یک مشاهده متعلق به خوشه B می‌باشد. در الگوریتم دورترین همسایه نیز برای محاسبه شباهت بین دو خوشه A و B از معیار $d_{AB} = \max_{i \in A, j \in B} d_{ij}$ استفاده می‌شود که در واقع شباهت بین دو خوشه بیشترین فاصله بین یک عضو از یکی با یک عضو از دیگری است. معیارهای شباهت وقتی که همه

1- Hierarchical Clustering
2- Dendogram
3- Divisive
4- Agglomerative

متغیرها از نوع عددی هستند، معمولاً با فاصله اقلیدسی بیان می‌شوند و در غیر این صورت، یک معیار مناسب برای آن در نظر گرفته می‌شود (هان و کامبر، ۲۰۰۶)

روش‌های خوشه‌بندی سلسله مراتبی استاندارد با استفاده از ضریب تشابه گوور[1](گوور، ۱۹۷۱) و یا دیگر معیارهای تشابه که در تحقیق گودا و بیدی (۱۹۹۱) آمده است، می‌توانند برای داده‌های آمیخته (داده‌های عددی و اسمی) مورد استفاده قرار گیرند (آندربرگ، ۱۹۷۳؛ جین و دوبس، ۱۹۸۸).

با توجه به آنکه داده‌های مربوط به متغیرهای مرتبط با خوشه‌بندی در این تحقیق از نوع داده‌های آمیخته (ترکیب داده‌های اسمی و عددی) هستند، بنابراین از ضریب شباهت گوور به عنوان معیار شباهت استفاده شده است. قبل از توضیح ضریب شباهت گوور، تعریف کوتاهی از ماتریس ارمیتی و ماتریس معین نیمه مثبت به این صورت ارائه می‌شود:

ماتریس ارمیتی[2]، ماتریس مربعی است که با ترانهاده مزدوج خود برابر باشد. به عبارت دیگر، اگر A^H ماتریس ترانهاده مزدوج باشد، داریم: $A = A^H$

ماتریس معین نیمه مثبت[3]، به صورت $q(x) = x^T Ax, q(x) \geq 0, \forall\, x \in R^n$ تعریف می‌شود. به بیان دیگر، ماتریس معین نیمه مثبت، ماتریس ارمیتی است که همه مقادیر ویژه آن نامنفی هستند.

۳-۱-۲- ضریب شباهت گوور

یکی از معروف‌ترین ضرایب جهت اندازه‌گیری شباهت در داده‌های آمیخته، ضریب شباهت گوور است. در حالت کلی، شباهت بین دو مشاهده i و j به صورت رابطه (۱) تعریف می‌شود (گوور، ۱۹۷۱):

$$S_{ij} = \frac{\sum_k w_{ijk} s_{ijk}}{\sum_k w_{ijk}} \qquad (۱)$$

1- Gower Coefficient of Similarity
2- Hermitian Matrix
3- Positive Semidefinite Matrix

w_{ijk} دارای مقدار یک است، اگر هر دو مشاهده i و j برای متغیر kام موجود و امکان مقایسه وجود داشته باشد و در غیر این صورت، برابر صفر خواهد بود. اگر w_{ijk} برای تمام مشاهدات برابر صفر باشد، S_{ij} قابل‌تعریف نیست. همچنین اگر تمام مقایسه‌ها امکانپذیر باشند، $\sum w_{ijk} = k$ خواهد بود که k تعداد کل متغیرها است. مقدار s_{ij} نیز به نوع داده‌ها بستگی دارد به طوری که برای متغیرهای اسمی، می‌توان رابطه (۲) نوشت:

$$s_{ijk} = \begin{cases} 1, & if \ \ xik = x\,jk \\ 0, & if \ \ xik \neq x\,jk \end{cases} \qquad (۲)$$

به عبارت دیگر، اگر دو مشاهده i و j در متغیر kام با یکدیگر مساوی باشند، S_{ij} برابر یک و اگر با یکدیگر متفاوت باشند، صفر درنظر گرفته می‌شود. همچنین در متغیرهای عددی با مقادیر $x_1,...,x_n$ برای متغیر k، می‌توان رابطه (۳) را نوشت:

$$s_{ijk} = 1 - \frac{\left|x_{ik} - x_{jk}\right|}{R_k} \qquad (۳)$$

که R_k تفاوت بین ماکزیمم و مینیمم مقدار متغیر kام است. مشابه متغیرهای اسمی، اگر $x_{ik} = x_{jk}$ باشد، آنگاه $s_{ijk} = 1$ و اگر x_{ik} و x_{jk} نشان‌دهنده ماکزیمم و مینیمم مقدار متغیر باشند، s_{ijk} برابر صفر خواهد بود. داده‌های دوحالتی نیز مانند داده‌های اسمی در نظر گرفته می‌شوند.

گوور (۱۹۶۶) نشان داد که فاصله بین مشاهدات iام و jام که متناسب با $(1 - S_{ij})^{\frac{1}{2}}$ هستند را می‌توان جایگزینی برای فضای اقلیدسی در نظر گرفت. برای این کافی است که S، ماتریس معین نیمه مثبت باشد. در این‌خصوص در صورتی که مقادیر گمشده در S نباشد، آنگاه S ماتریس معین نیمه مثبت خواهد بود.

۳-۱-۳- ارزیابی اعتبار در خوشه‌بندی

از آنجا که نتایج حاصل از پیاده‌سازی الگوریتم‌های خوشه‌بندی بر داده‌ها با توجه به انتخاب پارامترها (شامل تعداد خوشه‌ها) می‌تواند بسیار متفاوت از یکدیگر باشد، از این رو

باید با در نظرگرفتن شاخص‌هایی، اعتبار و عملکرد خوشه‌ها مقایسه شوند تا خوشه‌هایی که بهترین تناسب را با داده‌ها دارند، انتخاب شوند. شاخص مورد استفاده در این مقاله، شاخص سیلهوئت‌روسو[1] است. این شاخص از دسته شاخص‌های درونی[2] در ارزیابی اعتبار خوشه‌ها است که عمل ارزیابی خوشه‌ها را با استفاده از مقادیری که از خوشه‌ها محاسبه می‌شود، انجام می‌دهد (کافمن و روسو، ۱۹۹۰). در ادامه، خلاصه‌ای از شاخص سیلهوئت روسو که در مقاله استفاده شده، ارائه می‌شود.

شاخص سیلهوئت‌روسو به صورت روابط (۴) تا (۸) تعریف می‌شود:

$$silhouette = \sum_{i=1}^{n} S(i)/n \ , \ silhouette \in [-1, 1] \qquad (۴)$$

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i); b(i)\}} \qquad (۵)$$

$$a(i) = \frac{\sum_{j \in C_r} d_{ij}}{n_r - 1} \qquad (۶)$$

$$b(i) = \min_{s \neq r} d_{iC_s} \qquad (۷)$$

$$d_{iC_s} = \frac{\sum_{j \in C_s} d_{ij}}{n_s} \qquad (۸)$$

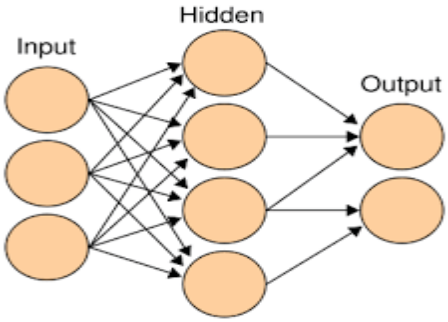
که $a(i)$ میانگین عدم تشابه iامین مشاهده نسبت به بقیه مشاهدات در خوشه C_r است. C_r و C_s، خوشه‌های r و s و n_r و n_s نیز تعداد مشاهدات در خوشه‌های C_r و C_s هستند. شاخص سیلهوئت روسو بین دو مقدار ۱- و ۱ قرار دارد و مقدار بزرگ‌تر برای این معیار نشان‌دهنده تعداد خوشه‌های مطلوب‌تر است.

۳-۲- شبکه عصبی

شبکه عصبی از جمله رایج‌ترین تکنیک‌های داده‌کاوی جهت برآورد و پیش‌بینی محسوب می‌شود. از ساختارهای رایج و معروف که در برآورد و پیش‌بینی کاربرد بسیاری داشته و

1- Rousseeuw's Silhouette
2- Internal Criteria

در این تحقیق نیز مورد استفاده قرار گرفته، شبکه‌های چند لایه پرسپترون است. شکل (۲) نمونه‌ای از شبکه عصبی چندلایه پرسپترون را نشان می‌دهد.



شکل (۲)- شبکه عصبی پرسپترون

این نوع از شبکه‌ها دارای مجموعه‌ای از ورودی‌ها (x_i ها) هستند که از طریق یک یا چند لایه پنهان به لایه خروجی متصل شده و مقدار خروجی $f(x)$ محاسبه می‌شود. به عنوان مثال، خروجی‌های یک شبکه با یک لایه ورودی به صورت زیر محاسبه می‌شوند:

$$f_k(x) = b_k + \sum_j v_{jk} h_j(x) \quad (۹)$$

$$h_j(x) = tanh(a_j + \sum u_{ij}\ x_i) \quad (۱۰)$$

که در آن، u_{ij} وزن یال مربوط به ورودی i به واحد پنهان j بوده و همچنین v_{jk} وزن مربوط به واحد پنهان j به خروجی k است. a_j و b_k نیز اریبی‌های واحدهای پنهان و خروجی هستند. هر مقدار خروجی مجموع وزنی مقادیر ورودی به علاوه یک اریبی است. هر واحد لایه پنهان به طور مشابه یک مجموع وزنی مقادیر ورودی را محاسبه کرده و حاصل را از طریق یک تابع غیرخطی فعال‌کننده مانند تانژانت هیپربولیک و یا لجستیک منتقل می‌کند. یک تابع غیرخطی فعال‌کننده به واحدهای لایه پنهان این امکان را می‌دهد که «مشخصه‌های پنهان» ورودی را که برای محاسبه خروجی‌ها مفید هستند، بیان کند.

برای ساخت یک مدل شبکه عصبی و استفاده از آن باید تعداد لایه‌ها و نورون‌های هر لایه در شبکه، نوع شبکه و توابع فعال‌کننده مشخص شوند تا پس از آن مقادیر وزن‌های شبکه برای نمونه‌های متعدد طی فرآیند آموزش، اصلاح شوند و شبکه از این طریق، فرآیند یادگیری را طی کرده تا برای داده‌های ورودی جدید دارای عملکرد خوبی باشد.

یکی از الگوریتم‌های پرکاربرد جهت آموزش شبکه، الگوریتم پس‌انتشارخطا[1] است که هدف آن حداقل کردن تابع خطا در شبکه است. در این الگوریتم با در نظر گرفتن نرخ یادگیری مناسب، داده خروجی با داده هدف مقایسه و میزان خطا محاسبه می‌شود، سپس خطای محاسبه شده به لایه‌های قبلی انتقال می‌یابد و ضرایب وزنی اصلاح می‌شوند تا درنهایت تابع خطا حداقل شود.

برای تعیین تعداد لایه پنهان از روش آزمون و خطا استفاده می‌شود. یک روش، رویکرد پیشرو[2] است که از تعداد کمی از نورون‌ها در لایه پنهان آغاز و در هر گام پس از محاسبه معیار خطا که شامل ریشه میانگین توان دوم خطا[3] (*RMSE*)، میانگین قدر مطلق انحراف[4] (*MAD*) و میانگین توان دوم خطای نرمال شده[5] (*NMSE*) است به تعداد نورون‌ها افزوده می‌شود تا در هر مرحله که مقدار معیار خطا حداقل شد، تعداد نورون‌ها برای لایه پنهان تعیین شوند (پرم‌چند و والیا، ۲۰۰۶).

ضریب تعیین R^2 نیز که در واقع از میانگین توان دوم خطای نرمال شده (*NMSE*) به دست می‌آید، جهت تغییرات متغیرهای مستقل و وابسته را نشان داده و مقداری بین صفر و یک دارد. مقدار یک نشان‌دهنده تطابق کامل داده‌ها است. روش محاسبه هر یک از معیارهای نامبرده در جدول (۱) خلاصه شده است.

1- Back Propagation
2- Forward
3- Root Mean Square Error
4- Mean Absolute Deviation
5- Normalized Mean Square Error

جدول (۱)- معیارهای خطا در ارزیابی عملکرد شبکه عصبی

معیار خطا	روش محاسبه *
میانگین توان‌دوم خطا (MSE)	$MSE = \frac{\sum (A-P)^2}{N}$
ریشه میانگین توان‌دوم خطا ($RMSE$)	$RMSE = \sqrt{MSE}$
میانگین قدر مطلق درصد خطا ($MAPE$)	$MAPE = \frac{100}{N} \times \sum \left\| \frac{A-P}{A} \right\|$
میانگین قدر مطلق انحراف (MAD)	$MAD = \frac{\sum \|A-P\|}{N}$
میانگین توان دوم خطای نرمال شده ($NMSE$)	$NMSE = \frac{\sum (A-P)^2}{\sum (A-\bar{A})^2}$
ضریب تعیین R^2	$R^2 = 1 - NMSE$

* A: مقدار واقعی داده، P: مقدار پیش‌بینی شده داده، $\bar{A}$: میانگین A و N تعداد داده‌ها است.

٤- نتایج محاسبات و یافته‌های تحقیق

٤-١- نتایج خوشه‌بندی شعب

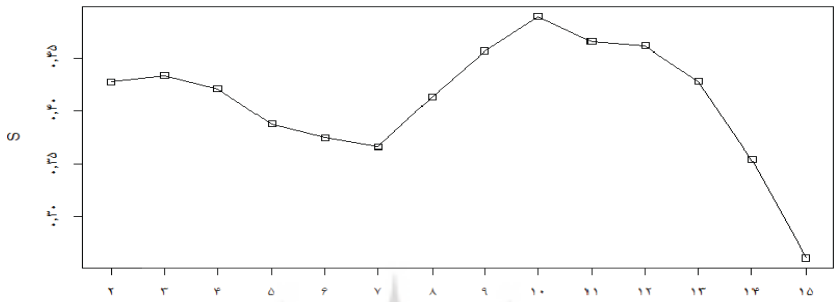
با توجه به آنکه در این مقاله، داده‌های مورد استفاده در خوشه‌بندی شعب از نوع داده‌های آمیخته (اسمی و عددی) هستند، از این رو با استفاده از ضریب تشابه گوور و سپس انجام خوشه‌بندی سلسله مراتبی، شعب به خوشه‌های متشابه دسته‌بندی شده‌اند. اعتبار خوشه‌ها با استفاده از شاخص سیلهوئت‌روسو اندازه‌گیری شده است. افزایش مقدار شاخص سیلهوئت‌روسو، بیانگر تعداد خوشه‌های بهینه‌تر است. جدول (۲) ، مقدار این شاخص را برای تعداد خوشه‌های متفاوت نشان می‌دهد.

جدول (۲)- مقادیر شاخص‌های ارزیابی اعتبار خوشه‌ها با تعداد متفاوت

تعداد خوشه‌ها (K)	۲	۳	٤	٥	٦	۷	۸
مقدار شاخص سیلهوئت روسو	۰/٤۲۷٤	۰/٤۳۳٥	۰/٤۲۰۹	۰/۳۸۷٦	۰/۳۷٤٦	۰/۳٦٥٦	۰/٤۱۳۳
تعداد خوشه‌ها (K)	۹	۱۰	۱۱	۱۲	۱۳	۱٤	۱٥
مقدار شاخص سیلهوئت روسو	۰/٤٥٦۲	۰/٤۸۸٥	۰/٤٦٤۹	۰/٤٦۰۹	۰/٤۲۷۹	۰/۳٥۳۷	۰/۲٦۱٤

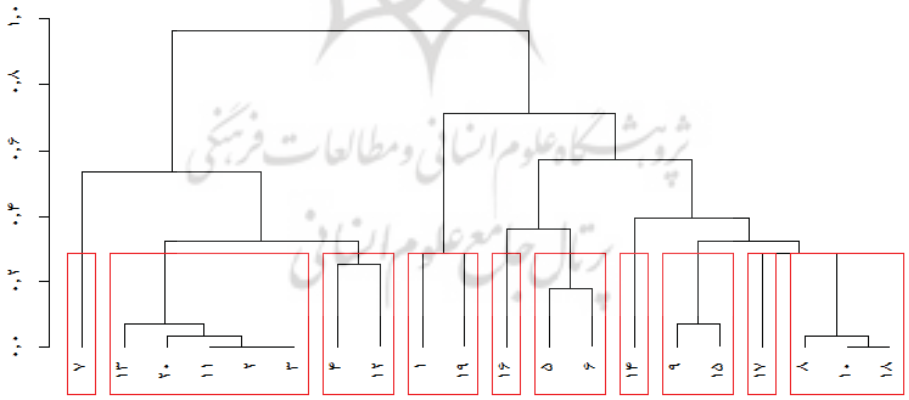
مأخذ: نتایج تحقیق

همانطور که نمودار (۲) نشان می‌دهد، بیشترین مقدار شاخص سیلهوئت برای تعداد ۱۰ خوشه است، از این رو عملکرد خوشه‌بندی برای تعداد ۱۰ خوشه، نسبت به سایر تعداد خوشه‌ها بهینه است.



نمودار (۲)- مقدار شاخص سیلهوئت نسبت به تعداد خوشه‌ها

با استفاده از الگوریتم دورترین همسایه در روش خوشه‌بندی سلسله مراتبی با تعداد ۱۰ خوشه، شماره شعب در داده‌ها استخراج شده‌اند که در نمودار (۳) ارائه شده است. همچنین جدول (۳) نشان‌دهنده کد شعب تشکیل‌دهنده هر خوشه است.



نمودار (۳)- خوشه‌بندی سلسله مراتبی با تعداد ۱۰ خوشه

جدول (۳)- کد شعب تشکیل‌دهنده هر خوشه

شماره خوشه	کد شعبه
۱	۱٤٦-٤٥٠
۲	۱۰۷-٦٦-۳٦۲-۷۳٦-۷۰٦
۳	۳۹۲-۱۳٥
٤	٥٤-۱٦٥
٥	۳۱۷
٦	۲۷۰-۳۱۲-٤۱۱
۷	۰-۲۱
۸	۱۰۱۹
۹	٥٦
۱۰	۲۳۱

مأخذ: نتایج تحقیق

٤-۲- نتایج حاصل از تحلیل شبکه عصبی براساس نتایج خوشه‌بندی

با مدنظر قرار دادن نتایج حاصل از خوشه‌بندی شعب در ۱۰ دسته متشابه، در گام بعد با استفاده از شبکه عصبی به برآورد وجه نقد مورد نیاز شعب پرداخته می‌شود. به‌منظور افزایش دقت و سرعت در آموزش شبکه عصبی، ابتدا تمامی داده‌ها با استفاده از فرمول $z_i = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})}$ استاندارد شده‌اند به‌طوری‌که تمام داده‌ها بین صفر و یک قرار گرفته‌اند. در این رابطه، z_i داده استاندارد شده، x_i داده مورد استفاده، x_{max} و x_{min} به ترتیب بیشترین و کمترین داده در هریک از متغیرها است. همچنین ۷۰ درصد داده‌ها به منظور آموزش شبکه و ۳۰ درصد داده‌ها نیز جهت آزمون شبکه به طور تصادفی استخراج و مورد استفاده قرار گرفته‌اند.

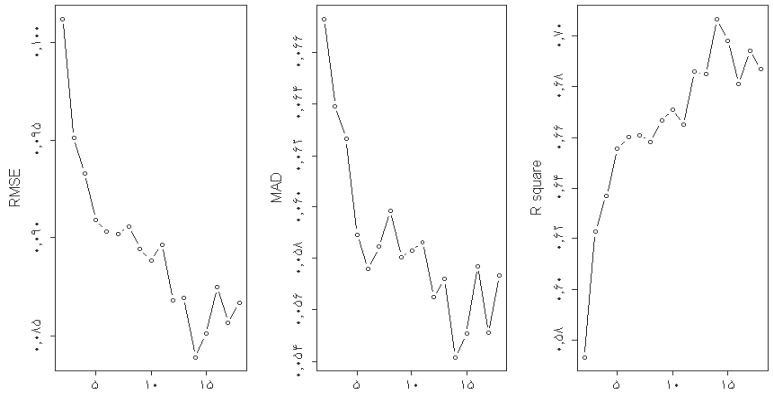
لایه ورودی شبکه عصبی دارای ۶ نورون شامل متغیرهای «روزکاری (*work.day*)»، «روزهفته (*week.day*)»، «روز پرداخت حقوق/ واریز یارانه/ واریزسود (*salary.day.effect*)»، «روز تعطیل و مناسبت‌های رسمی (*off.day.effect*)»، «میزان وجه نقد مصرفی دستگاه خودپرداز شعبه (*ATM*)» و نیز «شماره خوشه (*nc*)» بوده و لایه خروجی نیز دارای یک نورون شامل متغیر «وجه نقد مورد نیاز شعبه (*cash*)» است.

متغیرهای روزکاری، روز هفته، روز پرداخت حقوق/واریز یارانه/واریز سود، روزهای تعطیل و مناسبت‌های رسمی، متغیرهای تقویمی هستند به‌طوری‌که متغیر «روزکاری»، براساس ترتیب روزهای‌کاری از ابتدای ماه تا آخرین روزکاری هر ماه و متغیر «روز هفته»، براساس ترتیب روزهای هفته کدگذاری شده‌اند. همچنین متغیر «روز پرداخت حقوق/ واریز یارانه/ واریز سود» در روزهای اول تا سوم و نیز بیستم تا بیست و پنجم هرماه وزن‌دهی شده‌اند. روزهای کاری قبل و پس از تعطیلات رسمی نیز به دلیل تقاضای بیشتر وجه نقد در متغیر «روز تعطیل و مناسبت‌های رسمی» مدنظر قرار گرفته‌اند. متغیر «میزان وجه نقد مصرفی دستگاه خودپرداز شعبه» نیز میزان مصرف دستگاه خودپرداز شعبه بوده و متغیر «شماره خوشه» نیز براساس نتایج بخش ۴-۱ مورد استفاده قرار گرفته است.

از آنجا که برای تعیین تعداد نورون‌ها در لایه پنهان از روش آزمون و خطا استفاده می‌شود، به همین جهت در این تحقیق با رویکرد پیشرو، ابتدا تعداد ۲ نورون در لایه پنهان را در نظر گرفته و با افزایش تعداد نورون‌ها و محاسبه معیارهای خطا شامل ریشه میانگین توان دوم خطا ($RMSE$)، میانگین قدر مطلق انحراف (MAD) و ضریب تعیین R^2، در نهایت تعداد ۱۴ نورون در لایه پنهان که منجر به ایجاد شبکه‌ای با کمترین مقدار خطا در معیارهای خطا و بیشترین مقدار ضریب تعیین R^2 شده، انتخاب شده است.

نمودار (۴) مقدار معیارهای خطا نسبت به تعداد نورون‌ها در لایه پنهان را جهت تعیین ساختار شبکه با کمترین مقدار خطا نشان می‌دهد.

مقادیر معیارهای خطا و نیز ضریب تعیین برای تعداد نورون‌های مختلف در جدول (۴) ارائه شده است. براساس جدول (۴)، ریشه میانگین توان دوم خطا ($RMSE$) با مقدار۰/۰۸۳ و میانگین قدر مطلق انحراف (MAD) با مقدار ۰/۰۵۴ دارای کمترین مقدار در معیارهای خطا هستند. همچنین ضریب تعیین R^2 با مقدار ۰/۷ دارای بیشترین مقدار بوده که متعلق به تعداد ۱۴ نورون در لایه پنهان است.



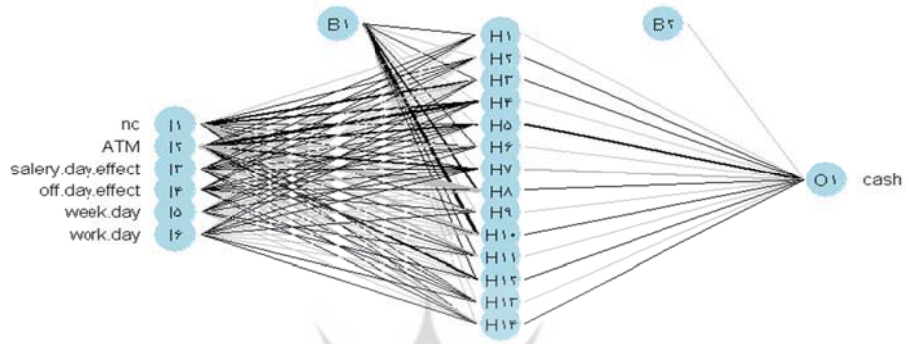
نمودار (٤)- معیارهای خطا و ضریب تعیین نسبت به تعداد نورون در لایه پنهان

جدول (٤)- مقادیر معیارهای خطا و ضریب تعیین نسبت به تعداد نورون در لایه پنهان

تعداد نورون	R^2	*MAD*	*RMSE*
۲	۰/۵۷۳	۰/۰۶۷۲	۰/۱۰۱۱
۳	۰/۶۲۲	۰/۰۶۳۹	۰/۰۹۵۱
٤	۰/۶۳۷	۰/۰۶۲۶	۰/۰۹۳۳
۵	۰/۶۵۵	۰/۰۵۸۹	۰/۰۹۰۹
۶	۰/۶۶۰	۰/۰۵۷۵	۰/۰۹۰۳
۷	۰/۶۶۱	۰/۰۵۸٤	۰/۰۹۰۱
۸	۰/۶۵۸	۰/۰۵۹۸	۰/۰۹۰۵
۹	۰/۶۶۶	۰/۰۵۸۰	۰/۰۸۹٤
۱۰	۰/۶۷۰	۰/۰۵۸۲	۰/۰۸۸۸
۱۱	۰/۶۶۵	۰/۰۵۸۶	۰/۰۸۹۶
۱۲	۰/۶۸۵	۰/۰۵۶٤	۰/۰۸۶۷
۱۳	۰/۶۸٤	۰/۰۵۷۱	۰/۰۸۶۹
۱٤	۰/۷۰۶	۰/۰۵٤۱	۰/۰۸۳۹
۱۵	۰/۶۹۸	۰/۰۵۵۰	۰/۰۸۵۱
۱۶	۰/۶۸۱	۰/۰۵۷۶	۰/۰۸۷٤
۱۷	۰/۶۹٤	۰/۰۵۵۱	۰/۰۸۵۶
۱۸	۰/۶۸۶	۰/۰۵۷۳	۰/۰۸۶۶

مأخذ: نتایج تحقیق

با توجه به کمترین مقدار خطا و بیشترین مقدار ضریب تعیین برای تعداد ۱۴ نورون در لایه پنهان، مدل شبکه عصبی تحقیق با سه لایه (دارای ۶ نورون در لایه ورودی، ۱۴ نورون در لایه پنهان و ۱ نورون در لایه خروجی) به صورت شکل (۳) تعیین شده است.



شکل (۳)- مدل شبکه عصبی تحقیق با سه لایه

پس از تعیین تعداد نورون‌ها در لایه پنهان با در نظر گرفتن توابع فعال‌کننده تانژانت هیپربولیک و لجستیک برای لایه‌های پنهان و خروجی، مقدار مجموع توان دوم خطا[1] (*SSE*) در شبکه عصبی، محاسبه شده که نتایج آن در جدول (۵) ارائه شده است.

براساس جدول (۵)، مقدار مجموع‌توان دوم خطا در شبکه با تابع فعال‌کننده تانژانت هیپربولیک در لایه پنهان و تابع خطی برای لایه خروجی از مقدار خطا در بقیه توابع کمتر است، از این رو توابع فعال‌کننده در شبکه عصبی این تحقیق برای لایه‌های پنهان و خروجی به ترتیب، توابع تانژانت هیپربولیک و خطی در نظر گرفته شده است.

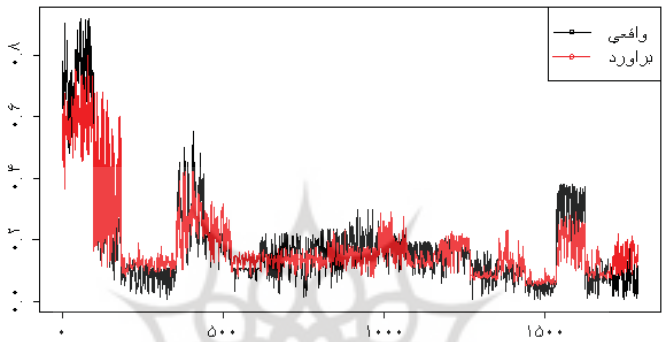
جدول (۵)- توابع فعال‌کننده برای لایه‌های پنهان و خروجی در شبکه عصبی تحقیق

تابع فعال‌کننده	مجموع توان دوم خطا (*SSE*)
لجستیک- خطی	٦/٤٠
لجستیک- لجستیک	٦/٥٣
تانژانت هیپربولیک- خطی	٥/٧٤
تانژانت هیپربولیک- تانژانت هیپربولیک	٦/٣٥

مأخذ: نتایج تحقیق

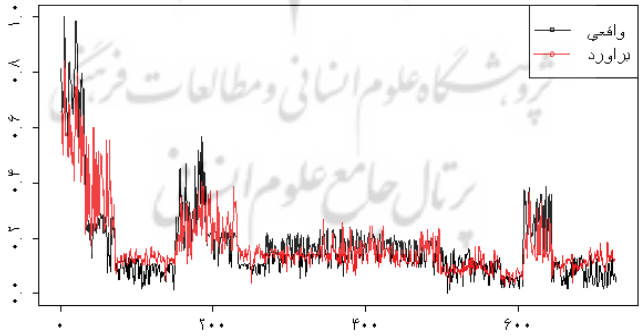
1- Sum of Squared Errors

شبکه عصبی تعیین شده با بهترین عملکرد ازطریق ۷۰ درصد داده‌ها (تعداد ۱۷۹۱ داده)، حداکثر ۵۰۰ تکرار و با الگوریتم پس انتشار خطا با نرخ یادگیری ۰/۲، آموزش داده می‌شود تا وزن‌های مناسب برای پیش‌بینی و برآورد وجه نقد شعب حاصل شوند. نمودار (۵)، مقادیر واقعی وجه نقد شعب و مقادیر برآورد شده آن با شبکه عصبی را برای داده‌های آموزش نشان می‌دهد.



نمودار(۵)- مقادیر واقعی وجه نقد شعب و مقادیر پیش‌بینی شده آن با شبکه عصبی برای داده‌های آموزش

علاوه بر این، داده‌های واقعی وجه نقد شعب و مقادیر برآوردشده توسط شبکه عصبی مرتبط با داده‌های آزمون، شامل ۳۰ درصد داده‌ها (تعداد ۷۲۹ داده) در نمودار۶ نشان داده‌شده است.



نمودار (٦)- مقادیر واقعی وجه نقد شعب و مقادیر پیش‌بینی شده آن با شبکه عصبی برای داده‌های آزمون

براساس جدول (۶)، میانگین قدر مطلق خطا و نیز ریشه میانگین توان دوم خطای برآورد وجه نقد برای تمام ۲۰ شعبه نمونه آماری (در ۱۰ خوشه) در بازه زمانی ۵ ماه (از اردیبهشت تا شهریورماه ۱۳۹۳) به ترتیب برابر ۰/۰۵ و ۰/۰۸ به دست آمده است. این مقادیر نشان می‌دهند، برآورد وجه نقد شعب بانک با استفاده از شبکه عصبی و با لحاظ نتایج حاصل از خوشه‌بندی شعب دارای دقت مطلوبی است.

جدول (٦)- مقادیر خطا برای برآورد وجه نقد در ۲۰ شعبه بانک در بازه زمانی ۵ ماه

میانگین قدر مطلق خطا (*MAD*)	ریشه میانگین توان دوم خطا (*RMSE*)
۰/۰۵	۰/۰۸

مأخذ: نتایج تحقیق

به دلیل حجم زیاد داده‌ها برای ۲۰ شعبه بانک در ۵ ماه از ارائه تمام مقادیر خودداری شده است. به همین دلیل، مقادیر واقعی وجه نقد شعب و مقدار برآورد آن از طریق شبکه عصبی، تنها برای خوشه اول (شامل دو شعبه ۴۵۰ و ۱۴۶) و پنجم (شامل شعبه ۳۱۷) برای اردیبهشت‌ماه ۱۳۹۳ در جداول (۷) و (۸) ارائه گشده است. به‌منظور اندازه‌گیری مقدار خطای پیش‌بینی هرروز کاری از معیارهای قدر مطلق خطا و توان دوم خطا و نیز جهت اندازه‌گیری خطای یک ماه هر خوشه از میانگین قدر مطلق خطا و ریشه میانگین توان‌دوم خطا استفاده شده است.

براساس نتایج جداول (۷) و (۸)، میانگین قدر مطلق خطا در اردیبهشت‌ماه ۱۳۹۳ برای دو شعبه ۴۵۰ و ۱۴۶ که در خوشه اول هستند، برابر ۰/۱۶۱ و این معیار برای شعبه ۳۱۷ که در خوشه پنجم است، برابر ۰/۰۲۴ به دست آمده است. همچنین ریشه میانگین توان‌دوم خطا برای خوشه اول، برابر ۰/۱۹۳ و برای خوشه پنجم ۰/۰۳۴ است که نشان‌دهنده مقدار خطای کم در برآورد وجه نقد است.[1]

۱- به منظور رعایت امانت در داده‌های بانک، مقادیر به صورت استاندارد شده ارائه شده است.

جدول (۷)- مقدار واقعی وجه نقد و مقدار برآورد آن از طریق شبکه عصبی، برای خوشه شماره یک در اردیبهشت ماه ۱۳۹۳

شعبه	روز	مقدار واقعی	مقدار برآورد شده	قدرمطلق خطا	توان دوم خطا
٤٥٠	۱۳۹۳۰۲۰۱	۰/٦۲٤۲	۰/٦۹۱٦	۰/۰٦۷۳	۰/۰۰٤
	۱۳۹۳۰۲۰۲	۰/۷۸۳۱	۰/٥٤۰۳	۰/۲٤۲۸	۰/۰٥۸
	۱۳۹۳۰۲۰۳	۰/٦٤٤۸	۰/٥۹۷۰	۰/۰٤۷۸	۰/۰۰۲
	۱۳۹۳۰۲۰٤	۰/٥٤۰۷	۰/٥۳۱۱	۰/۰۰۹۷	۰/۰۰۰
	۱۳۹۳۰۲۰٦	۰/۸۱۳۸	۰/۷٤۳٦	۰/۰۷۰۲	۰/۰۰٤
	۱۳۹۳۰۲۰۷	۰/۷۳۳٦	۰/٥۳٥۹	۰/۱۹۷۷	۰/۰۳۹
	۱۳۹۳۰۲۰۸	۰/۹۰٤۱	۰/٥٦۷۷	۰/۳۳٦٤	۰/۰۱۱
	۱۳۹۳۰۲۰۹	۰/۷٤۰۰	۰/٦۰۷۸	۰/۱۳۲۲	۰/۰۱۷
	۱۳۹۳۰۲۱۰	۰/۷۸۳٥	۰/٦۱٦۰	۰/۱٦۷٥	۰/۰۲۸
	۱۳۹۳۰۲۱۱	۰/۷۱۸٥	۰/٥۹۷۷	۰/۱۲۰۸	۰/۰۱٤
	۱۳۹۳۰۲۱۳	۰/٦۸٦۰	۰/٥٦۳٥	۰/۱۲۲٥	۰/۰۱٥
	۱۳۹۳۰۲۱٤	۰/٦٥۹۰	۰/٥۷۷٥	۰/۰۸۱٥	۰/۰۰٦
	۱۳۹۳۰۲۱٥	۰/۸۳۱۳	۰/٤۲۹۰	۰/٤۰۲۳	۰/۱٦۱
	۱۳۹۳۰۲۱٦	۰/٦٤۲۲	۰/٦٤۷۰	۰/۰۰٤۹	۰/۰۰۰
	۱۳۹۳۰۲۱۷	۰/۷٦۰۷	۰/٦۰۰۷	۰/۱٦۰۱	۰/۰۲٥
	۱۳۹۳۰۲۱۸	۰/٦۸۰۸	۰/٥٥۰۲	۰/۱۳۰٦	۰/۰۱۷
	۱۳۹۳۰۲۲۰	۰/۷۷۸۷	۰/٦۱٦۷	۰/۱٦۲۱	۰/۰۲٦
	۱۳۹۳۰۲۲۱	۰/۸٥۰٦	۰/٦۳٤٤	۰/۲۱٦۲	۰/۰٤٦
	۱۳۹۳۰۲۲۲	۰/۷۱۲۹	۰/٥۰۰٥	۰/۲۱۲٥	۰/۰٤٥
	۱۳۹۳۰۲۲٤	۰/٦۷٥٦	۰/٤٥٤٦	۰/۲۲۱۰	۰/۰٤۸
	۱۳۹۳۰۲۲٥	۰/٤۹۷۳	۰/٥٥۱۷	۰/۰٥٤٤	۰/۰۰۲
	۱۳۹۳۰۲۲۷	۰/٦٦۱۰	۰/٥۸۰۹	۰/۰۸۰۰	۰/۰۰٦
	۱۳۹۳۰۲۲۸	۰/٤۸۳۰	۰/٥۷۸۲	۰/۰۹٥۲	۰/۰۰۹
	۱۳۹۳۰۲۲۹	۰/۷۱۷۳	۰/٥۹۹۰	۰/۱۱۸۳	۰/۰۱٤
	۱۳۹۳۰۲۳۰	۰/٦۷٥۱	۰/٦۰۲٦	۰/۰۷۲٥	۰/۰۰٥
	۱۳۹۳۰۲۳۱	۰/٦۸۲۳	۰/٥٦۲٦	۰/۱۱۹۷	۰/۰۱٤

ادامه جدول ۷-

توان دوم خطا	قدرمطلق خطا	مقدار برآورد شده	مقدار واقعی	روز	شعبه
۰/۰۳۷	۰/۱۹۴۹	۰/۴۵۷۹	۰/۲۶۳۱	۱۳۹۳۰۲۰۱	۱۴۶
۰/۰۹۰	۰/۳۰۱۶	۰/۵۲۵۳	۰/۲۲۳۷	۱۳۹۳۰۲۰۲	
۰/۰۱۵	۰/۱۲۲۸	۰/۱۴۵۸	۰/۲۶۸۷	۱۳۹۳۰۲۰۳	
۰/۰۰۰	۰/۰۰۴۱	۰/۲۳۹۰	۰/۲۴۳۲	۱۳۹۳۰۲۰۴	
۰/۰۱۳	۰/۱۱۶۱	۰/۱۴۷۶	۰/۲۶۳۷	۱۳۹۳۰۲۰۶	
۰/۱۳۸	۰/۳۷۱۲	۰/۵۹۲۱	۰/۲۲۰۹	۱۳۹۳۰۲۰۷	
۰/۱۱۸	۰/۳۴۳۷	۰/۵۹۱۹	۰/۲۴۸۲	۱۳۹۳۰۲۰۸	
۰/۰۱۶	۰/۱۲۸۳	۰/۴۰۳۸	۰/۲۷۵۵	۱۳۹۳۰۲۰۹	
۰/۰۱۲	۰/۱۱۰۵	۰/۳۳۷۸	۰/۲۲۷۳	۱۳۹۳۰۲۱۰	
۰/۰۰۰	۰/۰۲۶۵	۰/۲۴۷۲	۰/۲۷۳۷	۱۳۹۳۰۲۱۱	
۰/۰۴۰	۰/۲۰۰۱	۰/۴۴۷۰	۰/۲۴۶۹	۱۳۹۳۰۲۱۳	
۰/۰۲۹	۰/۱۷۳۲	۰/۴۲۹۶	۰/۲۵۶۴	۱۳۹۳۰۲۱۴	
۰/۰۱۴	۰/۱۱۸۷	۰/۱۵۲۷	۰/۲۷۱۴	۱۳۹۳۰۲۱۵	
۰/۰۰۲	۰/۰۴۵۹	۰/۲۰۱۹	۰/۲۴۷۸	۱۳۹۳۰۲۱۶	
۰/۰۱۰	۰/۱۰۳۷	۰/۱۵۸۲	۰/۲۶۱۹	۱۳۹۳۰۲۱۷	
۰/۱۷۷	۰/۴۲۰۸	۰/۶۴۲۱	۰/۲۲۱۴	۱۳۹۳۰۲۱۸	
۰/۰۳۶	۰/۱۹۱۲	۰/۴۵۸۹	۰/۲۶۷۷	۱۳۹۳۰۲۲۰	
۰/۱۵۸	۰/۳۹۸۰	۰/۶۳۲۹	۰/۲۳۴۸	۱۳۹۳۰۲۲۱	
۰/۰۱۰	۰/۱۰۱۱	۰/۳۸۵۴	۰/۲۸۴۳	۱۳۹۳۰۲۲۲	
۰/۰۱۳	۰/۱۱۶۷	۰/۱۵۰۱	۰/۲۶۶۸	۱۳۹۳۰۲۲۴	
۰/۰۵۲	۰/۲۲۸۱	۰/۴۹۴۸	۰/۲۶۶۷	۱۳۹۳۰۲۲۵	
۰/۰۰۵	۰/۰۷۶۳	۰/۲۴۶۳	۰/۱۷۰۰	۱۳۹۳۰۲۲۷	
۰/۱۳۸	۰/۳۷۱۶	۰/۵۹۱۷	۰/۲۲۰۱	۱۳۹۳۰۲۲۸	
۰/۰۰۷	۰/۰۸۹۱	۰/۱۸۲۲	۰/۲۷۱۳	۱۳۹۳۰۲۲۹	
۰/۰۷۹	۰/۲۸۲۱	۰/۵۲۳۵	۰/۲۴۱۵	۱۳۹۳۰۲۳۰	
۰/۰۰۷	۰/۰۸۵۹	۰/۱۴۵۶	۰/۲۳۱۴	۱۳۹۳۰۲۳۱	

میانگین قدر مطلق خطا= ۰/۱۶۱، میانگین توان دوم خطا= ۰/۰۳۷، ریشه میانگین توان دوم خطا= ۰/۱۹۳

مأخذ: نتایج تحقیق

جدول (۸)- مقدار واقعی وجه نقد و مقدار برآورد آن از طریق شبکه عصبی، برای خوشه شماره پنج در اردیبهشت ماه ۱۳۹۳

شعبه	روز	مقدار واقعی	مقدار برآورد شده	قدرمطلق خطا	توان دوم خطا
۳۱۷	۱۳۹۳۰۲۰۱	۰/۰۵۶۱	۰/۰۸۵۱	۰/۰۲۹۰	۰/۰۰۰۸
	۱۳۹۳۰۲۰۲	۰/۰۸۳۷	۰/۰۸۷۲	۰/۰۰۳٤	۰/۰۰۰۰۱
	۱۳۹۳۰۲۰۳	۰/۰٦۸۸	۰/۰۹۷۳	۰/۰۲۸۵	۰/۰۰۰۸
	۱۳۹۳۰۲۰٤	۰/۱۰۲٦	۰/۱۱۲٦	۰/۰۱۰۰	۰/۰۰۰۱
	۱۳۹۳۰۲۰٦	۰/۰۷۷۳	۰/۱٦۱۳	۰/۰۸٤۰	۰/۰۰۷
	۱۳۹۳۰۲۰۷	۰/۰۷٦٤	۰/۱۳۰۷	۰/۰۵٤٤	۰/۰۰۲
	۱۳۹۳۰۲۰۸	۰/۱۱۹٦	۰/۱۲٤۱	۰/۰۰٤۵	۰/۰۰۰۰۲
	۱۳۹۳۰۲۰۹	۰/۱۲۷۰	۰/۱۱۷۹	۰/۰۰۹۱	۰/۰۰۰۰۸
	۱۳۹۳۰۲۱۰	۰/۰۸۲٤	۰/۱۱۲۳	۰/۰۲۹۸	۰/۰۰۰۸
	۱۳۹۳۰۲۱۱	۰/۱۰۵٦	۰/۱۱۱۸	۰/۰۰٦۲	۰/۰۰۰۸
	۱۳۹۳۰۲۱۳	۰/۰۵۰۰	۰/۱۵۹۵	۰/۱۰۹۵	۰/۰۰۰۰٤
	۱۳۹۳۰۲۱٤	۰/۱۳۷۰	۰/۱۳۲۵	۰/۰۰٤۵	۰/۰۱۱۹۹
	۱۳۹۳۰۲۱۵	۰/۱۰٤۰	۰/۱۲٦۵	۰/۰۲۲۵	۰/۰۰۰۰۲
	۱۳۹۳۰۲۱٦	۰/۰۹۷۷	۰/۱۲۱۰	۰/۰۲۳۲	۰/۰۰۰۵
	۱۳۹۳۰۲۱۷	۰/۱۱۰۲	۰/۱۱۵۷	۰/۰۰۵٦	۰/۰۰۰۵
	۱۳۹۳۰۲۱۸	۰/۱۱۱۰	۰/۱۰۵۲	۰/۰۰۵۸	۰/۰۰۰۰۳
	۱۳۹۳۰۲۲۰	۰/۰٦٤۸	۰/۰۹۲۲	۰/۰۲۷٤	۰/۰۰۰۰۳
	۱۳۹۳۰۲۲۱	۰/۰۷۹۰	۰/۰۸۸٦	۰/۰۰۹٦	۰/۰۰۰۷
	۱۳۹۳۰۲۲۲	۰/۱۰۷۵	۰/۰۹۷٤	۰/۰۱۰۱	۰/۰۰۰۰۹
	۱۳۹۳۰۲۲٤	۰/۰۸۵٤	۰/۰۹۰٤	۰/۰۰۵۰	۰/۰۰۰۱
	۱۳۹۳۰۲۲۵	۰/۰۷۹۷	۰/۰۸۹٤	۰/۰۰۹۷	۰/۰۰۰۰۳
	۱۳۹۳۰۲۲۷	۰/۰۹۰۳	۰/۱۳۱۰	۰/۰٤۰۷	۰/۰۰۰۰۹
	۱۳۹۳۰۲۲۸	۰/۱۲۲۲	۰/۱۱۵٦	۰/۰۰٦۷	۰/۰۰۱٦
	۱۳۹۳۰۲۲۹	۰/۰۹۵٤	۰/۱۱۳٤	۰/۰۱۸۰	۰/۰۰۰۰٤
	۱۳۹۳۰۲۳۰	۰/۰٤۹۵	۰/۱۱٤۰	۰/۰٦٤٤	۰/۰۰۰۳
	۱۳۹۳۰۲۳۱	۰/۰۹۵۸	۰/۱۰۲۲	۰/۰۰٦٤	۰/۰۰٤

میانگین قدر مطلق خطا = ۰/۰۲٤۲، میانگین توان دوم خطا= ۰/۰۰۱، ریشه میانگین توان دوم خطا= ۰/۰۳٤

مأخذ: نتایج تحقیق

۵- نتیجه‌گیری و پیشنهاد

۵-۱- نتیجه‌گیری

نتایج مقاله حاضر را می‌توان به شکل زیر بیان کرد:

۱- با توجه به آنکه داده‌کاوی از روش‌های آمار ناپارامتری محسوب می‌شود، امکان بکارگیری آن برای انواع داده‌ها وجود دارد که از مزیت‌های آن بشمار می‌رود، چراکه داده‌های جمع‌آوری شده اغلب خیلی از فرض‌های قدیمی آماری(از قبیل داشتن توزیع مشخص و...) را در نظر نمی‌گیرند و عدم رعایت آنها می‌تواند صحت و درستی نتایج تحقیق را دچار مشکل کند. در این خصوص استفاده از نتایج خوشه‌بندی در شبکه عصبی می‌تواند دقت پیش‌بینی و برآورد وجه نقد شعب را افزایش دهد.

۲- در این مقاله، ۲۰ شعبه بانک تجارت در بازه زمانی ۵ ماه با توجه به وجود تنوع بین آنها، ابتدا در خوشه‌های متشابه دسته‌بندی شده، سپس با مدنظر قرار دادن نتایج حاصل از خوشه‌بندی و استفاده از معیارهای مختلف خطا، ساختار شبکه عصبی که منجر به بهبود عملکرد شبکه و برآورد وجه نقد با دقت بالاتر می‌شود، تعیین شده است.

۳- نتایج نشان می‌دهد شبکه عصبی با سه لایه (یک لایه ورودی با ۶ نورون، یک لایه پنهان با ۱۴ نورون و یک لایه خروجی با یک نورون)، الگوریتم یادگیری پس‌انتشارخطا و با نرخ یادگیری ۰/۲، می‌تواند با میانگین قدر مطلق خطا و نیز ریشه میانگین توان دوم خطا به ترتیب برابر ۰/۰۵ و ۰/۰۸ تخمین مناسب‌تری برای وجه نقد مورد نیاز شعبه ارائه کند که تاییدی قابل قبول برای مدل است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی

۵-۲- پیشنهاد برای مطالعات آتی

پرتال جامع علوم انسانی

استفاده از روش شبکه‌های عصبی می‌تواند به مشکل بیش برازش منجر شود به طوری که آموزش شبکه با حداقل خطا انجام شود، اما شبکه در داده‌های آزمون و پیش‌بینی داده‌های جدید عملکرد خوبی نداشته باشد. در این‌خصوص، بکارگیری رویکرد بیزی در شبکه‌های عصبی می‌تواند مشکل بیش برازش را حل کند که این موضوع می‌تواند به عنوان تحقیقات آتی مدنظر قرار بگیرد.

فهرست منابع

بهشتی، شروین (۱۳۸۹)، مدلی برای پیش‌بینی نقدینگی مورد نیاز شعب بانک با استفاده از شبکه‌های عصبی و پیاده‌سازی آزمایشگاهی آن، دانشکده حسابداری و مدیریت، دانشگاه علامه طباطبایی، پایان‌نامه کارشناسی‌ارشد.

سعیدی، علی و شعبانی‌مطلق، مریم (۱۳۸۹)، «ریسک نقدینگی در صنعت بانکداری با استفاده از شاخص لاندای امری»، *فصلنامه بورس اوراق بهادار*، سال‌سوم، شماره۱۲، صفحات ۱۴۹-۱۲۹.

نوربخش، ایمان، حیدری، هادی و زواریان، زهرا (۱۳۸۹)، «مدیریت نقدینگی وجوه نقد صندوق شعب با استفاده از مدل انتشار»، گزارش هفتم، پژوهشکده پولی و بانکی، بانک مرکزی جمهوری اسلامی ایران.

Anderberg, M. (1973), *Cluster Analysis for Applications*, Academic Press.

Bramer, M. (2007), Principles of Data Mining, Springer Verlag Berlin Heidelberg.

Cardona, L. and M. L. Amaya (2012),"Cash Management Cost Reduction Using Data Mining to Forecast Cash Demand and LP to Optimize Resources", *Memetic Comp*,vol.4, pp.127-134.

Gowda, K. and E. Diday (1991),"Symbolic Clustering Using a New Dissimilarity Measure", *Pattern Recognition* , vol.24, no.6 , pp. 567-578.

Gower, J. (1971),"A General Coefficient of Similarity and Some of its Properties", *Bio Metrics* ,vol. 27, pp. 857-874.

Gower, J. (1966),"Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis", *Bio Metrics*, vol.53 , pp. 315-328.

Han, J. and M. Kamber, (2006), *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufman.

Hornik,K.,M. Stinchcombe and H.White(1989),"Multilayer Feedforward Networks are Universal Approximators", *Neural Networks*, vol.2, pp. 359-366.

Hu, M., G. Zhang, and B. Patuwo (1998),"Forecasting with Artificial Neural Networks:The State of the Art", *International Journal of Forecasting* , vol.14, pp. 35-62.

Irie, B. and S. Miyake (1988),"Capabilities of Three-layered Perceptrons", *In Proceedings of the IEEE International Conference on Neural Networks*, vol. 1, pp. 641-648.

Jain, A. and R. Dubes (1988), *Algorithms for Clustering Data*, Prentice Hall.

Kauffman, L. and P. Rousseeuw (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley.

Olson, D. L. and D. Delen (2008), *Advanced Data Mining Techniques*, Springer Verlag Heidelberg.

Premchand, K. and E. Walia (2006), "Cash Forecasting: An Application of Artificial Neural Networks in Finance", *International Journal of Computer Science & Applications*, vol.3 , no. 1, pp. 61-77.

Sandipan, R. (2010),"Determining Optimal Cash Allocation at ICICI Bank Branches", *Mumbai: ICICI B*ank, SAS Global Forum, pp. 1-13.

Turban, E., J. Aronson, T. Liang, and R. Sharda (2007), *Decision Support and Business Intelligence Systems*, Pearson Education.

Wang, P. (2008), "Clustering and Classification Techniques for Nominal Data Application", Dissertation, Department of Electronic and Engineering, City University of Hong Kong.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی