فصلنامه مدیریت صنعتی دانشکده علوم انسانی، دانشگاه آزاد اسلامی
واحد سنندج – سال پانزدهم/ شماره ۵۳ / پاییز ۱۳۹۹

تبیین دسته‌بندهای ماشین بردار پشتیبان و شبکه عصبی جهت درجه بندی شعب بانک

داود خسروانجم (نویسنده مسؤول)
گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
Email: d.khosroanjom@modares.ac.ir
محمد الیاسی قوپی
کارشناس ارشد مهندسی کامپیوتر گرایش هوش مصنوعی، دانشگاه شهید چمران، اهواز، ایران
بهزاد کشانچی
گروه مدیریت بانکی، دانشکده علوم اقتصادی، موسسه علوم بانکی، تهران، ایران
بهاره بوبانیان
گروه مدیریت بازرگانی، دانشکده مدیریت، واحد سنندج، دانشگاه آزاد اسلامی، سنندج، ایران
شوانه عبدالهی
گروه مدیریت دولتی، دانشکده مدیریت، واحدمهاباد، دانشگاه آزاد اسلامی، مهاباد، ایران

تاریخ دریافت: ۹۹/۰۳/۱۵ * تاریخ پذیرش ۹۹/۰۸/۱۷

چکیده

در صنعت بانکداری اطلاعات زیادی وجود دارد که شناسایی آن از اهمیت ویژه‌ای برخوردار است. استفاده از تکنیک‌های داده کاوی نه تنها کیفیت را بهبود می‌بخشد بلکه منجر به مزایای رقابتی و ارتقای جایگاه بازار نیز می‌گردد. با استفاده از داده‌کاوی و به منظور تجزیه و تحلیل الگوها و روندها, بانکها می‌توانند صحت این را پیش بینی کنند درجه بندی شعب بانک چگونه است. در این مقاله درجه‌بندی شعب یکی از بانکهای بزرگ تجاری (تعداد شعب انتخاب شده ۱۸۲۵ شعبه و تعداد ویژگی‌های مورد استفاده ۵۷ ویژگی) با استفاده از دسته‌بندهای ماشین بردار پشتیبان و شبکه عصبی چند لایه پرسپترون برروی داده‌های واقعی انجام گرفت. نتایج ارزیابی مربوط به ماشین بردار پشتیبان نشان داد که این دسته‌بند دارای کارایی پایین‌تری برای روش پیشنهادی است. اما استفاده از شبکه‌های عصبی و ترکیب آن با PCA نشان داد که دارای معیارهای کارایی بالایی است. مقادیر مربوط به کارایی و صحت با استفاده از شبکه عصبی با دقت بسیار بالایی بدست آمد.

کلمات کلیدی: داده کاوی، بانکداری، ماشین بردار پشتیبان، شبکه عصبی.

۱-مقدمه

افزایش رقابت جهانی در بازارها سازمان‌ها را مجبور می‌کند تا برای موفقیت بیشتر به راههای نوین رقابت در بلند مدت روی آورند. تنها استفاده از فناوری پیشرفته و کیفیت بالا برای حفظ مشتریان کافی نمی‌باشد زیرا رقبا نیز به سرعت در حال پیشرفت می‌باشند و به زودی به همان سطح خواهند رسید(Kim and Atuahene-Gima, 2010). به عبارت دیگر, با پیشرفت‌های فناوری, بهره‌وری افزایش یافته و تغییرات گسترده‌ای در کسب وکار اتفاق افتاده است(Karakatsanis et al., 2016). مشتریان نیز بطور مداوم از طریق کانالهای ارتباطی متنوع اطلاعات مورد نیاز خود را کسب می‌کنند و به این طریق دانش آنها افزایش می‌یابد و این افزایش دانش موجب می‌شود تا مشتری گزینه‌های جدیدی را کسب نماید (Mesforoush and Tarokh, 2013).

یکی از مهم‌ترین مواردی که در سازمان‌های امروز به خوبی به چشم می‌خورد, فراوانی و حجم بالای داده‌ها و اطلاعاتی است که در سازمان تولید می‌شود. این حجم وسیع از داده و اطلاعات، در حالی که سازمان را با مسائلی در رابطه با چگونگی نگهداری, جمع آوری و طبقه بندی داده‌ها روبه‌رو می‌کند, دانشی را در درون خود نهفته دارد که یک سازمان یادگیرنده و هوشمند, از آن به مانند فرصتی برای ارتقای توانمندی‌های خود استفاده می‌کند(Thillainayagam,2012). رقابت تنگاتنگی که بین بانکها برای ارائه‌ی خدمات بانکداری به وجود آمده, روز به روز رو به افزایش است. در چنین شرایطی, مشتریان این بانکها نیز برای دریافت خدمات با کیفیت بالاتر، انتظارات بیشتری خواهند داشت. بنابراین استفاده از داده‌کاوی می‌تواند به بانکها برای ارتقاء خدمات با استفاده از انبوه داده کمک نماید.

امروزه سازمان‌ها مزایایی را که از استفاده از داده کاوی می‌گیرند درک کرده‌اند. داده‌کاوی ابزاری با ارزش است که به سازمان‌ها کمک می‌کند مزایای رقابتی بهتری نسبت به رقباء کسب نمایند(El-Zehery et al., 2013). یک گزارش از تحقیق فارستر[1] بیان می‌کند که از ۱۰۰۰ شرکت، ۵۲ درصد از آنها از داده‌کاوی برای استراتژی‌های بازاریابی خود استفاده می‌نمایند. افزایش داده-ها تنها دلیل برای استفاده سازمان‌ها از داده‌کاوی نیست بلکه این تمایل به داده‌کاوی به دلیل کاهش هزینه داده‌های ذخیره شده و افزایش دسترسی به داده‌های جمع آوری شده, توسعه پایداری[2] و کارایی الگوریتم‌های ماشین یادگیری برای پردازش داده‌ها می‌باشد(Ltifi et al., 2016). هر چه قابلیت‌های پایگاه‌داده بزرگتر و هزینه‌ها کاهش یابند استفاده از داده‌کاوی برای خدمات تجاری بیشتر توصیه می‌گردد(Deneke et al., 2017). داده‌کاوی به سازمان‌ها کمک می‌کند که تجارت خود را بهتر درک نموده, به طور بهتری به مشتریان خود خدمات ارائه نمایند و در نهایت اثربخشی سازمان را افزایش دهند (Sing'oei and Wang,2013). داده کاوی باید در بطن استراتژی روابط با مشتری بانک قرار گیرد تا برنامه عملیاتی که بایستی انجام گیرد مشخص شود و این عمل در نتیجه آنچه از طریق داده کاوی یاد گرفته شده صورت می پذیرد(Garcia-Marti et al., 2019).

امروز داده‌کاوی در تمامی صنایع شامل مؤسسه‌های مالی, بیمه و ارتباطات مورد استفاده قرار می‌گیرد. در سال‌های گذشته سازمان‌ها علی الخصوص بانکها, اهمیت اطلاعاتی را که در مورد مشتریان جمع آوری نموده اند تشخیص داده‌اند (Crock et al. 2001).

عمده فعالیت بانک‌ها در کشور توسط شعب انجام می‌پذیرد؛ لیکن میزان مسئولیت و اختیارات هر شعبه ممکن است با دیگر شعب متفاوت باشد. تعیین سطوح مختلف اختیار و مسئولیت‌های هر شعبه توسط مکانیسمی با عنوان درجه‌بندی شعب صورت می‌گیرد. از این‌رو درجه‌بندی یکی از امور مهم و اساسی در فرآیندهای مدیریتی بانک محسوب شده و نتایج آن اثرات مستقیم و غیرمستقیم فراوانی را در شعب بر جای می‌گذارد. به عبارت دیگر, دسته‌بندی و رتبه‌بندی شعب یانکها, یکی از دغدغه‌های اصلی مدیران بانکها محسوب می‌شود. این کار باعث ایجاد یک فضای رقابتی شده و سودآوری را افزایش می‌دهد. روش‌های درجه بندی فعلی بانکها که غالباً با انتخاب شاخص‌ها و درجه اهمیت آنها به صورت میانگین وزنی توسط عده معدودی از کارشناسان انجام می‌شود, روش علمی نبوده و با تغییر کارشناسان, این شاخص‌ها و ضرایب تغییر می کند. بنابراین تجدید نظر در این روشها

[1] Forrester
[2] robust

و انتخاب روشهای علمی ضروری است. تا بحال دسته‌بندی شعب با استفاده از اطلاعات آماری و توسط فرد خبره انجام می-گرفت که قطعاً دقت لازم را ندارد. هدف این مقاله, ایجاد یک سیستم هوشمند مبتنی بر داده‌کاوی و یادگیری ماشین به منظور دسته‌بندی خودکار شعب بانک است. مطالعه‌ی موردی روی شعب یکی از بانکهای بزرگ تجاری انجام گرفت. برای این منظور اطلاعات کلیه‌ی شعب در ۵ سال اخیر بر اساس شاخص‌های تعیین شده جمع‌آوری شده و همچنین این داده‌ها توسط فرد خبره برچسب‌گذاری شده‌اند. با استفاده از یادگیری ماشین و داده‌کاوی می‌توان از دانش قبلی استفاده و ارتباط نهان بین شاخص‌ها و دسته‌ی آن‌ها را پیدا کرد. بعد از اتمام فرآیند آموزش سیستم, می‌توان از آن به عنوان یک سیستم دسته‌بندی شعب بانک استفاده نمود. به این صورت که سیستم با دریافت شاخص‌های تعیین شده می‌تواند دسته‌ی مربوط به آن شعب را تعیین کند. برای ایجاد یک سیستم بهینه از چند دسته‌بند مختلف با پارامترهای متفاوت استفاده شده تا در نهایت یک سیستم یادگیری بهینه شناسایی شود. نتایج ارزیابی توسط معیارهای دقت, کارایی و صحت بر روی شعت بانک بدست آمده است.

امروزه حدود دوسوم موجودی کسب وکار جهانی را بخش‌های خدماتی دربرگرفته است و صنایع متمرکز بر منابع انسانی و غیر-انسانی در حال رشد است(Jayachandran et al., 2005). انتظار می‌رود فناوری اطلاعات و مدیریت ارتباط با مشتری, سازوکاری برای ایجاد و حفظ رابطه با مشتریان از طریق ایجاد روابط بلندمدت با مشتریان فراهم آورد (Luo et al., 2010). ارائه‌ی خدمات مناسب به مشتریان, موجب افزایش اعتماد مشتریان و افزایش تمایل آنها به خرید مجدد و درمقابل خدمات رسانی نامناسب و بی‌اثر, یکی از مهم‌ترین دلایل رفتار ترک مشتری است (Torkzadeh et al., 2006). یکی از اهداف سازمان ها در استفاده از داده کاوی افزایش وفاداری مشتریان و حفظ مشتریان وفادار جهت دستیابی به مزیت رقابتی و سودآوری است(رنگریز و بایرامی, ۱۳۹۸). داده‌کاوی بطور گسترده‌ای برای ارتقاء عملکرد سازمانی و ایجاد موقعیت رقابتی مورد استفاده قرار می‌گیرد(Chen et al., 2009). و به شیوه‌های گوناگون و بر اساس نظرات متفاوت نویسندگان مختلف تعریف شده است. جدول شماره (۱) تعدادی از تعاریف داده‌کاوی را براساس ادبیات موضوع رایج در داده‌کاوی نشان می‌دهد.

جدول شماره (۱): تعاریف داده‌کاوی براساس ادبیات موضوع

ردیف	مفهوم داده‌کاوی	منبع
۱	قابلیت جستجوی پیچیده داده که از الگوریتم‌های آماری برای کشف روابط مخفی بین داده‌ها استفاده می‌کند	Rygielski et al., 2002
۲	فرآیند استخراج اطلاعات و دانش مفید از پایگاه داده ای بسیار بزرگ	Shmueli et al., 2007
۳	مجموعه‌ای از مکانیسم‌ها و تکنیک‌ها که از طریق نرم افزار, اطلاعات مخفی شده در داده‌ها را استخراج می‌کند	Coenen, 2011
۴	فرآیندی که با استفاده از تکنیک‌های آماری, ریاضی, هوش مصنوعی و یادگیری ماشین اطلاعات مفید را شناسایی و دانش را از اطلاعات گرفته شده استخراج می‌کند	Bahari and Elayidom, 2015

براساس تعاریف متفاوتی که از داده‌کاوی در جدول فوق ارائه شد این استنباط می‌شود که در همه یک ایده مشترک وجود دارد: استخراج اطلاعات مهم از داده‌های موجود به منظور ارئه‌ی تصمیم‌گیری بهتر سازمانی.

داده‌کاوی نه تنها تصمیم‌گیری را با استفاده از جستجو برای تشخیص روابط و الگو از داده‌های گسترده جمع آوری شده بهبود می‌بخشد بلکه انباشت[3] اطلاعات را نیز کاهش می‌دهد(Zhu et al.,2001). داده‌کاوی سازمان را قادر می‌سازد بر اطلاعات مهم در پایگاه داده تمرکز کند و به مدیران با پیش بینی رفتارها و روندهای آتی در اتخاذ تصمیماتی که متکی بر دانش هستند کمک می‌نماید (Chopoorian et al., 2001). به علت اینکه فرآیند جستجو در انبوه داده‌ها خیلی پرهزینه و زمانبر است مدیران با استفاده از داده‌کاوی می‌توانند قادر به اتخاذ تصمیمات بهتری باشند(Agany et al., 2020).

- داده‌کاوی و کشف دانش از پایگاه داده

[3] overload

تاریخچهٔ کشف دانش از پایگاه‌های اطلاعاتی قدمت چندانی ندارد و امروزه به داده‌کاوی مشهور است. اصطلاح کشف دانش برای نخستین بار در دههٔ ۱۹۹۰ مطرح شد و توجه پژوهشگران را به سمت الگوریتم‌های داده‌کاوی معطوف کرد. هدف داده‌کاوی, کشف دانش جدید, معتبر و قابل پیگیری با استفاده از ابزارهای هوش مصنوعی و آماری در حجم بالایی از داده‌ها است(Marban et al., 2009). داده‌کاوی، استخراج یا اقتباس دانش از مجموعهٔ داده‌ها است و به فرآیندی گفته می‌شود که دانش را از داده‌ها استخراج می‌کند و این دانش در قالب الگوها و مدلها بیان می‌شود (Yaning et al., 2008).در شکل ۱ مراحل کشف دانش از پایگاه داده‌ها نشان داده شده است(Deshpande and Thakare, 2010).این مراحل به شرح زیر هستند:

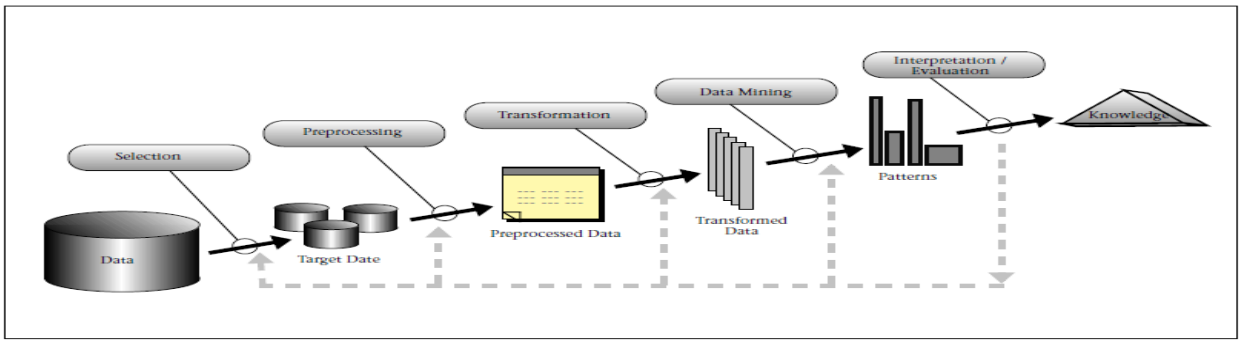
۱. انتخاب داده‌ها: داده‌های مربوط به تجزیه وتحلیل و تصمیم‌گیری از داده‌های دیگر جدا می‌شوند.

۲. پیش پردازش اطلاعات: پردازش, تمیزکردن و یکپارچه سازی داده ها انجام می‌گیرد.

۳. تبدیل داده‌ها: داده های انتخا ب شده به شکل مناسبی برای روش داده‌کاوی تبدیل می‌شوند.

۴. داده‌کاوی: در این مرحله از روشهای هوشمندانه‌ای برای استخراج الگوهای بالقوهٔ مفید استفاده می‌شود و در مورد این روشها تصمیم‌گیری می‌شود.

۵. تفسیر و ارزیابی: در این مرحله, الگوهای جالب توجه نشان دهندهٔ دانش, بر اساس اقدامات انجام شده شناخته می‌شوند و دانش کشف شده در اختیار کاربر قرار می‌گیرد. در این مرحله استفاده از روش‌های تجسم سازی برای کمک به کاربران ضروری است.



شکل شماره (۱). مراحل کشف دانش از پایگاه داده‌ها (منبع: Fayyad et al., 1997)

دو هدف اصلی داده‌کاوی، پیشگویی و توصیف است(Kiss,2003):

۱. داده‌کاوی پیشگویی کننده, مدلی را از سیستم ارائه می‌دهد که این مدل را مجموعه‌ای از داده‌های مشخص, پیش بینی می‌کنند. هدف کلی آن ایجاد الگویی برای طبقه بندی, پیش بینی و تخمین داده‌ها است.

۲. داده‌کاوی توصیفی, اطلاعات جدید و غیر بدیهی را بر اساس مجموعه‌ای از داده‌های موجود ارائه می‌دهد و هدف کلی آن درک و شناخت سیستم‌های تجزیه وتحلیل شده با استفاده از الگوها و روابط موجود است.

جدول ۲ چگونگی فعالیت‌های داده‌کاوی را براساس نویسندگان مختلف در ادبیات موضوع نشان می‌دهد.

جدول شماره (۲): فعالیت‌های داده‌کاوی

منبع / فعالیت‌ها	Bradley et al. (1999)	Yoon (1999)	Berry and Linoff (2002)	Chye et al. (2002)
خوشه بندی	✓	✓	✓	✓
تجسم سازی			✓	✓
مدلسازی پیشبینانه	✓		✓	✓
تحلیل همبستگی		✓		
تشخیص انحرافات		✓		

الف) کاربردهای داده‌کاوی

امروزه استفاده از داده‌کاوی در صنایع مختلفی روبه گسترش است که برخی از حوزه‌های مهم آن عبارتند از: تجزیه و تحلیل نتایج۴ دارویی، تشخیص تقلب در کارت‌های اعتباری, پیش بینی رفتار خرید مشتری, پیش بینی علایق شخصی کاربران Web و فرآیند بهینه‌سازی تولید(Carvajal et al., 2018). به طور کلی یکی از کاربردهای اصلی داده‌کاوی در بحث مدیریت ارتباط با مشتری است. ارتباط با مشتریان بدون شناخت نیازهای آنها و شناخت نیازهای مشتریان بدون به کار بردن روش‌های داده‌کاوی امکان پذیر نیست. داده‌کاوی به شرکت‌ها جهت حرکت به سمت مشتری محوری کمک می‌کند. انبار داده مناسب جز حیاتی یک مدیریت ارتباط با مشتری موفق می‌باشد. داده‌ها در حوزه‌های مختلفی جمع آوری می‌شوند. پس از جمع آوری داده‌ها لازم است داده‌ها پردازش شده و توسط روش‌های داده‌کاوی تحلیل شوند(Rygielski et al., 2002).

برخی از مزیت‌های به کار بردن داده‌کاوی در مدیریت ارتباط با مشتری به شرح زیر است(2001, LEE):

- بهبود کارائی کسب و کار با کمترین بودجه
- بهره برداری از پایگاه داده برای نگهداری روابط مشتری
- افزایش وفاداری مشتری و ارزش مشتری و کاهش نرخ از دست رفتن مشتری
- آشنایی با نیازهای مشتری برای توسعه استراتژی های پاسخگویی به نیازها
- ارزیابی اثربخشی تبلیغات و ترقی‌ها
- پاسخگویی به انتظارات مشتری و تقویت کیفیت خدمات

ب) صنعت بانکداری

در صنعت بانکداری شناسایی اطلاعات مشتریان اهمیت ویژه‌ای دارد. در این صنعت اطلاعات زیادی وجود دارد و استفاده از فناوری اطلاعات نه تنها کیفیت خدمات را بهبود می‌بخشد بلکه مزایای رقابتی را نیز بدست می‌آورد (Hwang et al., 2002). مقدار زیادی از داده‌هایی که بانکها سالیانه جمع آوری می‌نمایند می‌تواند تأثیر زیادی در استفاده بهتر از داده‌کاوی داشته باشد. با استفاده از داده‌کاوی و به منظور تجزیه و تحلیل الگوها و روندها, بانکها می‌توانند صحت این را پیش بینی کنند که چگونه مشتریان نسبت به تغییرات در نرخ سود واکنش نشان می‌دهند و چه مشتریانی نسبت به ارائه خدمات جدید رغبت بیشتری نشان می‌دهند و اینکه چه مشتریانی می‌خواهند با ریسک بیشتری وام‌های بیشتری بگیرند.

جدول ۳ حوزه‌های مختلف صنعت بانکداری در ایران را با استفاده از داده‌کاوی نشان می‌دهد.

جدول شماره (۳): حوزه های داده‌کاوی در صنعت بانکداری

ردیف	حوزه مورد استفاده	منبع
۱	رویگردانی مشتری از خدمات بانکی	سپهری و همکاران،۱۳۹۰
۲	کاربران موبایل بانک	حسن زاده و همکاران، ۱۳۹۱
۳	مشتریان اینترنت بانک	رادفر و همکاران، ۱۳۹۳
۴	طراحی سیستم مدیریت پرتفولیوی وام های بانکی	دیواندری و همکاران
۵	بخش بندی مشتریان بانک	خواجوند و همکاران،۱۳۹۱
۶	طراحی مدل انتخاب نیروی انسانی	آذر و همکاران،۱۳۸۹

منافع و دستاوردهای داده‌کاوی در صنعت بانکداری عبارتنداز (Teymourpour, 2015):

- خلق تمایز در بانک‌ها, مهمترین دستاورد بکارگیری فناوری داده‌های بزرگ و داده‌کاوی آنها است
- دانش نهان استخراج شده از داده‌ها مهمترین دارایی بانک شده و منشاء مزیت رقابتی و تمایز و اعتلای برند بانک در نزد مشتریان خواهد بود
- تحلیل حجم عظیم داده‌های بانکی, اساس تصمیم‌گیری و اقدامات استراتژیک مدیران بانکی

[4] Outcomes

- بخش بندی مشتریان برای شناسایی، ارزیابی و پیش بینی رفتار مشتریان به منظور تعریف سبدهای متنوع خدمات بانکی
- رصد کردن تغییر خواستگاه مشتریان
- سفارشی سازی خدمات بانکی بر اساس بخش‌های مختلف مشتریان

بازاریابی, مدیریت ریسک, تشخیص تقلب و حفظ مشتری کاربردهای مؤثری از داده‌کاوی در صنعت بانکداری می‌باشند.

ج) بازاریابی

بازاریابی یکی از بهترین حوزه‌هایی است که می‌تواند در آن تکنیک‌های داده‌کاوی مورد استفاده قرار گیرد. داده‌کاوی سازمان‌ها را قادر می‌سازد که براساس میزان گسترده‌ای از داده‌هایی که از مشتریان دارند مشتریان هدف را شناسایی نمایند و این یکی از اهداف اساسی بازاریابی در هر سازمانی است. با این کار به میزان قابل توجهی در پول و زمان صرفه جویی شده و سازمان می-تواند تشخیص دهد که مشتریانش چه کسانی هستند و قادر به پیش بینی الگوهای مصرف آنها را دارد. یکی از کاربردهای گسترده داده‌کاوی در صنعت بانکداری در بازاریابی است. معاونت بازاریابی بانک با استفاده از داده‌کاوی می تواند پایگاه داده مشتریانش را مورد تجزیه و تحلیل قرار دهد و از لحاظ آماری ارجحیت مشتریان خود را نسبت به خدمات بانک تشخیص دهد. بانک های آمریکایی از پایگاه داده بازاریابی به منظور بهبود خدمات به مشتریان و افزایش سود استفاده می نمایند و با یکپارچه سازی اطلاعات ۵ ساله مشتریان خود قادر بودند که خدمات ویژه‌ای را به مشتریان خود ارائه دهند(Cabena et al., 1998). با تشخیص الگوهای رفتار مشتریان قابلیت سوددهی بانک‌ها تعیین می‌گردد و آنها می‌توانند تجارت خود را توسعه دهند.

د) مدیریت ریسک

در صنعت بانکداری داده‌کاوی برای مدیریت ریسک نیز مورد استفاده قرار می‌گیرد. موسسات مالی و بانک‌ها به لحاظ ماهیت کاری نیازمند فرآیندهای مدیریت ریسک می باشند. یکی از عمده‌ترین ریسک‌های بانک‌ها به عنوان بخش اصلی نظام مالی, ریسک اعتباری است. معوقه بانک‌ها, گویای فقدان مدل‌های مناسب اندازه گیری ریسک اعتباری و سیستم های مدیریت ریسک در شبکه بانکی است(Teymourpour, 2015).

بانکها نیاز به این دارند که بدانند کدام دسته از مشتریان آنها قابلیت اعتماد بیشتری دارند. اگر بانکها اطلاعاتی از مشتریان خود نداشته باشند ارائه‌ی کارت‌های اعتباری به مشتریان جدید, توسعه خطوط[5] اعتباری به مشتریان موجود, واگذاری وام و ... برای آنها تصمیماتی مبتنی بر ریسک می‌باشند(Turki, 2011). در موضوع کارت اعتباری, داده‌کاوی از طریق اینکه کدام دسته از مشتریان نسبت به حساب هایشان کوتاهی[6] می‌کنند می‌تواند ریسک را کاهش دهد. کروک[7] و همکاران(۲۰۰۱) طی گزارشی اعلام کردند که رتبه بندی اعتباری یکی از اولین ابزارهای توسعه داده شده مدیریت ریسک مالی است. رتبه‌بندی اعتباری هنگامی که تصمیم به وام دادن گرفته شود ابزاری باارزش برای وام دهندگان در صنعت بانکداری به شمار می‌رود. وام دهندگان تعداد وام‌هایی را که ارائه می‌دهند حتما با ابزاری ارزیابی ریسک صحت و درستی آن را کنترل می‌نمایند(Crook et al., 2001). داده‌کاوی اعتبار رفتارهای شخصی وام گیرنده را با اقساط, گرونامه[8] و وام‌های کارت اعتباری تفکیک می‌نماید. امتیازی که حاصل شده به وام دهنده اجازه می‌دهد مشتری را مورد ارزیابی قرار داده و در نهایت تشخیص دهد که آیا فرد صلاحیت گرفتن وام را دارد و یا خیر و اگر دارد ریسک آن چگونه است. مشتریانی که در دوره طولانی در بانک حساب دارند و دستمزدهای بیشتری دارند احتمال گرفتن وامشان بیشتر می‌باشد نسبت به مشتریان جدید و یا مشتریانی که دستمزد پایین تری دارند. با آگاهی از اینکه نحوه رفتار و حساب مشتری چگونه است بانک می‌تواند در موقعیت بهتری برای کاهش ریسک باشد.

ه) تشخیص تقلب[9]

---

5 Lines
6 Default
7 Crook
8 Mortgage
9 Fraud

حوزه‌ی دیگری که داده‌کاوی می‌تواند در بانکداری صنعتی بکار رود تشخیص تقلب است. امروزه داده کاوی قابلیت گسترده‌ای در تشخیص تقلب برای مؤسسه های مالی دارد (Lin et al., 2015). در واقع گسترده‌ترین کاربرد داده‌کاوی در بانکداری, تشخیص تقلب است (Groth, 1998). توانایی تشخیص فعالیت‌های کلاهبردارنه دغدغه روبه فزاینده‌ای برای برخی کسب و کار است و با استفاده از داده‌کاوی فعالیت‌های کلاهبردارنه را می‌توان کشف و آن را متوقف نمود. مؤسسه‌های مالی دو رویکرد متفاوتی برای تشخیص اقدامات کلاهبردارانه مورد استفاده قرار می‌دهند(Decker, 1998). در رویکرد اول، بانک با استفاده از داده‌های انبارداده(اطلاعات تراکنش‌های برخی از شرکت‌ها) از برنامه‌های داده‌کاوی برای شناسایی الگوهای تقلب استفاده می-کند. بانک می‌تواند سپس با مراجعه به الگوهای آنان و مقایسه آن با پایگاه داده خود علائم مشکلات را شناسایی نماید. در رویکرد دوم, شناسایی الگوی تقلب بر اساس اطلاعات داخلی خود بانک می‌باشد. یکی از سیستم‌هایی که در شناسایی تقلب موفق بوده است سیستم ارزیابی تقلب فالکون[۱۰] می‌باشد که بوسیله HNC توسعه داده شده است. مبنای کار فالکون براساس شبکه‌های عصبی می‌باشد و بوسیله تعدادی از بانکها برای تشخیص تراکنش‌های مشکوک مورد استفاده قرار می-گیرد(Brachman et al., 1996).

و) حفظ و اکتساب مشتری[۱۱]

داده‌کاوی نه تنها کمک به معرفی مشتریان جدید در صنعت بانکداری می‌نماید بلکه به حفظ و نگهداری مشتریان موجود نیز کمک می‌نماید. نگهداری و اکتساب مشتری یکی از موضوعات مهم برای هر صنعتی علی الخصوص صنعت بانکداری است. امروزه مشتریان حق انتخاب‌های گوناگونی با توجه به کسب و کار خود دارند. فعالان صنعت بانکداری باید به این آگاه باشند که اگر به مشتریان خود توجه کامل نداشته باشند مشتری می‌تواند به سادگی شرکت دیگری را که خود می‌خواهد انتخاب نماید. داده‌کاوی می‌تواند با ارائه محصولات و خدمات به مشتریان جدید و کشف الگوهای خرید مشتریان قبلی در رسیدن به هدف کمک نماید و بانک را قادر می‌سازد با ارائه پیشنهادهای جذاب مشتریان موجود را حفظ نماید. هنگامی که بانک Chase Manhattan در مقایسه با رقباء مشتریان خود را از دست داد با استفاده از داده‌کاوی به تجزیه و تحلیل حساب های مشتریان خود اقدام نمود و سازوکارهای الزامات حساب مشتریان خود را تغییر داد و لذا با این اقدام مجدداً توانست نسبت به جذب مشتریان جدید اقدام نماید(Fabris,1998).

۲- روش شناسی تحقیق

دراین مقاله درجه‌بندی شعب بانک تجارت توسط روش‌های ماشین بردار پشتیبان و شبکه عصبی به کمک داده‌های واقعی انجام می‌گیرد. در ابتدا روش کار با ماشین بردار پشتیبان معرفی می‌شود و در ادامه شبکه عصبی بیان می‌شود. روش کار به این صورت است که در ابتدا یک پیش‌پردازش اولیه شامل نرمال‌سازی داده‌ها و کاهش ابعاد انجام می‌گیرد و سپس درجه‌بندی شعب به صورت جداگانه توسط دو دسته‌بند ماشین بردار پشتیبان و شبکه عصبی چند لایه پرسپترون انجام می‌گیرد.

الف) ماشین بردار پشتیبان

فرض کنید تعدادی از بردارهای ویژگی یا الگوهای آموزشی بصورت {x1, x2, ..., xi, ..., xN} داریم که هر یک بردار ویژگی d بعدی بوده و دارای برچسب yi که $y_i \in \{-1, +1\}$. هدف حل یک مسأله دسته بندی دو کلاسه بصورت بهینه است. فرض کنید این دو کلاس را با تابع تمایز f(x) و با یک ابر صفحه H با رابطه (۱) بخواهیم از هم جدا کنیم.

$$H: W.X + b = 0 \quad (۱)$$

$$f(x) = sign(w.x + b) \quad (۲)$$

بردار وزن W, بردار عمود بر ابر صفحه جدا کننده و b مقدار بایاس است. منظور از w.x حاصلضرب داخلی است. واپنیک [۱۲] ثابت کرد که بعد VC برای طبقه‌بندی کننده‌هایی از نوع ابر صفحات کانونی[۱۳], دارای یک کران بالاست که این کران بالا با توان

[10] Falcon
11 Customer Acquisition and Retention
[12] Vapnik

دوم نُرم بردار وزنی یعنی $\|w\|^2$ نسبت مستقیم دارد. در واقع اگر ما $\|w\|^2$ را محدود کرده و مینیمم کنیم, بعد VC طبقه‌بندی کننده را می‌نیمم کرده‌ایم و تخمین ما از مقدار ریسک واقعی بصورت احتمالی دقیق‌تر بوده و خاصیت تعمیم دسته‌بندی کننده بیشتر خواهد شد.

$$\|w\| = \left(\sum_{i=1}^{d} w_i^2\right)^{1/2} \quad (۳)$$

رابطه بین خاصیت تعمیم طبقه‌بندی کننده با نُرم بردار وزن $\|w\|$ را می‌توان به طریق دیگر نیز توجیه کرد: فرض کنید داده‌های دو کلاس جدایی پذیر باشند و بردارهای ویژگی مرزی کلاس اول روی ابرصفحه H^+ و بردارهای ویژگی مرزی کلاس دوم روی ابر صفحه H^- قرار گیرند. ابر صفحات H^+ و H^- به صورت رابطه (۴) تعریف می‌شوند.

$$H^+: w.x + b = +1$$
$$H^-: w.x + b = -1 \quad (۴)$$

الگوهایی که بر روی ابر صفحات H^+ و H^- قرار می‌گیرند, بردار پشتیبان نامیده می‌شوند. ناحیه بین دو ابر صفحه H^+ و H^- را حاشیه یا ناحیه مرزی[۱۴] گویند. فاصله بین دو ابر صفحه H^+ و H^- برابر $\frac{2}{\|w\|}$ خواهد بود. طراحی ابر صفحه با بیشترین عرض ناحیه مرزی یا ناحیه مرزی بهینه بر این استوار است که با شرط درست طبقه‌بندی شدن الگوها, عرض ناحیه مرزی حداکثر شود, یعنی $\frac{2}{\|w\|}$ ماکزیمم شود و $\|w\|$ مینیمم گردد. هدف این است که اولاً الگوها درست طبقه‌بندی گردند و ثانیاً بر روی و یا خارج از ناحیه مرزی واقع شوند, یعنی برای تمامی نمونه‌ها رابطه (۵) برقرار باشد. بنابراین ما با یک مسئله بهینه‌سازی موجود در رابطه (۵) روبرو هستیم و هدف پیدا کردن w به نحوی است که رابطه (۵) حداقل شود.

$$Min \frac{1}{2}\|w\|^2 + C\sum_i \varepsilon_i$$
$$y_i(w.x_i + b) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0 \quad (۵)$$

که در آن C پارامتر تنظیم کننده حاشیه می‌باشد که وظیفه‌ی آن برقراری تعادل بین حداکثر کردن حاشیه و حداقل کردن خطای دسته‌بندی را بوده و همواره بزرگ تر از صفر است. اگر C عددی بزرگ انتخاب شود, توجه بیشتری به خطا معطوف می‌گردد. ماشین بردار پشتیبانی که به این صورت تعریف شده باشد را ماشین بردار پشتیبان حاشیه–نرم[۱۵] می‌نامند.

ب) داده‌های جداناپذیر خطی

در حالتی که داده‌ها جداناپذیر بوده و همچنین کلاس‌ها دارای همپوشانی هستند, جدا کردن کلاس‌ها توسط مرز خطی همواره با بروز خطا همراه می‌باشد. به منظور حل مشکل مزبور می‌توان ابتدا داده‌ها را با استفاده از یک تبدیل غیرخطی φ از فضای اولیه به فضایی با بعد بالاتر منتقل کرد با این هدف که در فضای جدید, کلاس ها تداخل کمتری با یکدیگر داشته باشند. پس از نگاشت داده‌ها به فضای بالاتر, با استفاده از معادلات قبل و جایگزین کردن $\boldsymbol{x_i}$ با $\boldsymbol{\varphi(x_i)}$ ابر صفحه بهینه بدست خواهد آمد. انتقال از بعد پایین به بعدی بالاتر را می‌توان توسط توابع کرنل انجام داد. تعدادی از توابع کرنل موجود به شرح زیر می‌باشند:

(۶) کرنل خطی $K(x_i, x_j) = x_i^T x_j + c$

(۷) کرنل چند جمله‌ای $K(x_i, x_j) = (\alpha x_i^T x_j + c)^d$

(۸) کرنل تابع پایه شعاعی $K(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$

در معادلات فوق c عبارت ثابت, d درجه چند جمله‌ای و σ پارامتر قابل تنظیم است.

---

[13] Canonical Hyperplane
[14] Margin
[15] Soft-Margin

کرنل تابع پایه شعاعی در دو مقوله نسبت به کرنل‌ها دارای نقطه قوت بوده و در نتیجه عملکرد نمونه‌ها را به ابعاد فضایی بالاتری نگاشت می‌نماید, بنابراین بر خلاف کرنل خطی, در حالت‌هایی که روابط بین کلاس‌ها و ویژگی‌ها به صورت غیر خطی است, کاربرد خواهد داشت. دومین نقطه قوت, تعداد پارامترهایی است که بر پیچیدگی انتخاب مدل تأثیر گذار می‌باشند. کرنل چند جمله‌ای تعداد پارامترهای بیشتری نسبت به تابع پایه شعاعی دارد. از این رو در این مقاله از کرنل تابع پایه شعاعی استفاده شده است. بنابراین مسئله بهینه‌سازی به صورت رابطه (۹) تعریف می‌شود.

$$Max \ \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$
$$\sum_i \alpha_i y_i = 0\,, \qquad \forall\, i, 0 \le \alpha_i \le C \tag{۹}$$

برای دسته‌بندی داده جدیدی همانند x می‌توان تصمیم‌گیری را بر اساس تابع (۱۰) انجام داد.

$$f(x) = sign \sum \alpha_i y_i k(x_i, x) + b \tag{۱۰}$$

ج) ماشین بردار پشتیبان در حالت چندکلاسه

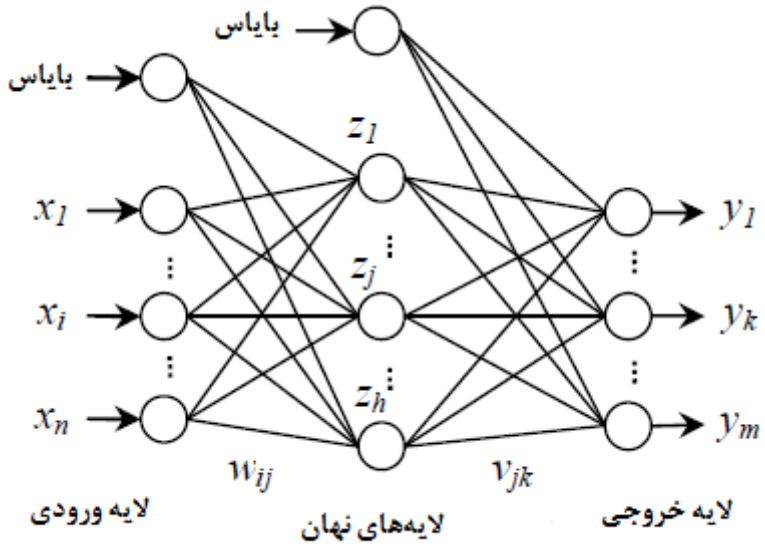
رابطه های بیان شده تا به اینجا مربوط به یک مسئله کلاس‌بندی برای نمونه‌هایی است که تنها متعلق به دو کلاس هستند و ابر صفحه‌های توانایی تنفکیک داده‌های به دو کلاس را دارند. اما مسئله ما بیش از دو کلاس دارد. در مسئله مورد نظر در مورد درجه‌بندی شعبات بانک ما با ۷ درجه یا در واقع ۷ کلاس مختلف روبرو هستیم. هر چند که تئوری طراحی ابر صفحه با ناحیه مرزی بهینه, برای یک مسئله دو کلاسه ارائه شده است ولی به راحتی می‌توان یک ماشین بردار پشتیبان ساخت که در حالت چندکلاسه نیز عمل دسته‌بندی را به خوبی انجام دهد. برای این کار چند استراتژی مختلف وجود دارد: ۱- یکی در مقابل بقیه[۱۶] ۲- دسته‌بندی کردن زوج- زوج[۱۷] ۳- تعریف توابع هدف چند کلاسه[۱۸] ۴- استفاده از کدهای تصحیح خطا[۱۹]. در این مقاله از روش دوم برای یک مسئله چند کلاسه استفاده شده است.

د) شبکه عصبی پرسپترون چند لایه

شبکه‌های عصبی مصنوعی با الهام گرفتن از ماهیت کارکرد سیستم مغز, توانایی بالایی در حل مسائل پیچیده دارند. شبکه‌های عصبی پرسپترون چند لایه که در این مقاله مورد استفاده قرار می‌گیرند یکی از پرکاربردترین نوع از این شبکه‌های عصبی مصنوعی است. در این شبکه‌ها مدل یادگیری باناظر است. ویژگی‌های مربوط به نمونه‌های ما به عنوان ورودی سلول‌های عصبی در نظر گرفته می‌شوند و برچسب هر نمونه در لایه خروجی مشخص می‌شود. هدف لایه‌های میانی، ایجاد یک رابطه محاسباتی بین لایه ورودی و لایه خروجی است. همان‌طور که در شکل (۲) نشان داده شده است, تعداد لایه‌های پنهان می‌تواند بیش از یک لایه باشد. لایه خروجی خطی است اما تابع عملیاتی استفاده شده برای لایه‌های پنهان در این مقاله یک تابع غیر خطی شبه هذلولی می‌باشد که در رابطه (۱۱) نشان داده شده است.

$$\sigma(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{۱۱}$$

[16] One versus the Rest
[17] Pairwise Classification
[18] Multiclass Objective Functions
[19] Error-Correcting Codes



شکل شماره (۲): ساختار شبکه عصبی پرسپترون چند لایه

فرض کنید مقادیر ورودی و خروجی سلول‌های عصبی لایه ورودی و لایه خروجی به ترتیب (x1, x2, ..., xn) و (y1, y2, ..., ym) باشند و مقادیر ورودی سلول‌های عصبی لایه مخفی (s1, s2, ..., sh) ومقادیر خروجی آن‌ها (z1, z2, ..., zh) باشند. در این صورت مقادیر آن‌ها از رابطه‌های (۱۲) و (۱۳) محاسبه می‌شوند.

$$z_i = \sigma\left(\sum_{j=1}^{n} w_{ij} + w_{0i}\right), \quad 1 \le i \le h \tag{۱۱}$$

$$y_k = \sigma\left(\sum_{i=1}^{h} v_{ki} z_i + v_{0k}\right), 1 \le k \le m \tag{۱۲}$$

که در آن w_{ij} وزن‌های اتصالات بین لایه ورودی و لایه مخفی است, $\boldsymbol{w_{0i}}$ مقادیر حد آستانه سلول‌های عصبی مخفی هستند و $\boldsymbol{v_{ki}}$ وزن‌های اتصالات بین لایه مخفی و خروجی است. حال می‌توان مقادیر خروجی را با استفاده از رابطه (۱۳) بدست آورد.

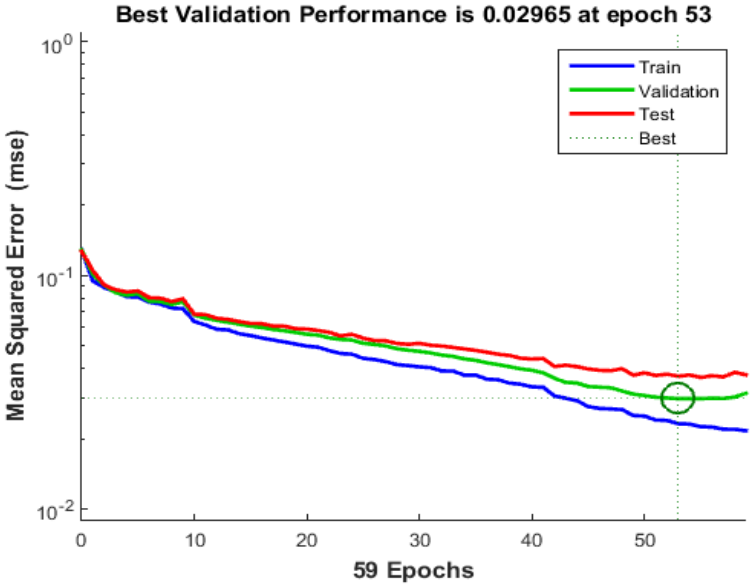
$$y_k = \sigma\left(\sum_{i=1}^{h} v_{ki}\sigma\left(\sum_{j=1}^{n} w_{ij}x_i + w_{0i}\right) + v_{0k}\right) \tag{۱۳}$$

مقادیر مروبوط به $\boldsymbol{y_k}$, مقادیر خروجی هر نمونه است. اما انتظار داریم که این خروجی, نزدیک به برچسب انتخاب شده برای آن نمونه باشد. در این مقاله از روش یادگیری پس انتشار خطا استفاده می‌شود. روند کار به این صورت است در هر بار اختلاف بین خروجی شبکه با برچسب انتخاب شده محاسبه می‌شود و سعی می‌شود با تکرارهای متوالی میزان این خطا به حداقل برسد. اما نکته‌ای که وجود دارد این است که این فرآیند آموزش چه وقت باید متوقف شود. اگر فرایند آموزش خیلی طولانی باشد, بیش برازش[۲۰] پیش می‌آید یعنی شبکه خیلی به داده‌های آموزش حساس می‌شود و اگر داده‌های جدید کمی متفاوت باشند, نتیجه‌ی دقیقی حاصل نمی‌شود. به همین دلیل در این مقاله داده‌ها یه سه دسته‌ی آموزش[۲۱]، اعتبارسنجی[۲۲] و آزمایش[۲۳] تقسیم می‌شوند.

اهمیت داده‌های اعتبارسنجی این است که از وقوع بیش برازش جلوگیری می‌کند. در شکل (۳) یک نمونه فرآیند آموزش و میزان خطاهای آموزش, اعتبارسنجی و آزمایش در هر اپوک نشان داده شده است. همانطور که مشاهده می‌شود هر چه تعداد اپوک‌ها بیشتر شود, میزان خطای آموزش کاهش می‌یابد اما به نقطه‌ای می‌رسیم که کم کم خطای اعتبارسنجی افزایش می‌یابد؛ این

[20] Overfitting
[21] Train
[22] Validation
[23] Test

نقطه همان جایی است که ممکن است از آن به بعد بیش برازش اتفاق بیافتد به همین دلیل فرایند آموزش در آن متوقف می-شود.



شکل شماره (۳): خطای مربوط به فرآیند آموزش در شبکه عصبی

ه) مجموعه آموزشی

داده‌های مورد استفاده در این مقاله مربوط به رتبه‌بندی شعب بانک در کشور است. تعداد شعب انتخاب شده ۱۸۲۵ شعبه است و تعداد ویژگی‌های مورد استفاده ۵۷ ویژگی است. رتبه‌بندی این شعبه‌ها بر اساس محاسبات و نظر فرد خبره انجام شده است و در نهایت شعب به هفته درجه مختلف دسته‌بندی شده‌اند. تعداد شعب مربوط به هر درجه در جدول ۴ نشان داده شده است.

جدول شماره (۴): تعداد شعب مربوط به هر دسته در مجموعه آموزشی

درجه	درجه ۱	درجه ۲	درجه ۳	درجه ۴	درجه ۵	درجه ۶	درجه ۷
تعداد شعبات	۱۷	۲۴	۹۰	۲۲۵	۷۳۴	۷۱۶	۱۹

و) پیش‌پردازش داده‌ها

از بین ۵۷ ویژگی موجود, تعدادی از ویژگی‌ها تأثیر چندانی در رتبه‌بندی ندارند و برای اکثر شعب بدون مقدار هستند به همین دلیل در ابتدا این ویژگی‌ها حذف شده‌اند. از بین ویژگی‌ها ۷ مورد چنین خصوصیتی داشتند بنابراین تعداد ویژگی‌های باقیمانده برای فرآیند آموزش توسط دسته‌بندها[24] ۵۰ عدد است. با توجه به اینکه ویژگی‌های مورد استفاده در مقیاس‌های عددی مختلف تعریف شده‌اند بنابراین کار با آن‌ها تا حدودی می‌توان دقت را کاهش دهد به همین دلیل از یک روش نرمال‌سازی استفاده شده است. در نرمال‌سازی، تمامی ویژگی‌ها به یک محدوده عددی یکسان نگاشت می‌شوند. در این مقاله از روش نرمال‌سازی مینیمم- ماکزیمم استفاده شده است. این روش, تبدیل خطی را روی داده اصلی انجام می‌دهد، ارتباط بین مقادیر داده اولیه را حفظ می‌کند. فرض کنید که $\max_A$ و $\min_A$ مقادیر ماکزیمم و مینیمم ویژگی A باشند, نرمال‌سازی مینیم- ماکزیمم مقدار v مربوط به ویژگی A را به مقدار v' در محدوده [new-$\min_A$, new-$\max_A$] توسط رابطه (۱۴) نگاشت می‌کند.

$$v' = \frac{v - min_A}{max_A - min_A}(new_max_A - new_min_A) + new_min_A \quad (۱۴)$$

ز) کاهش ابعاد داده‌ها

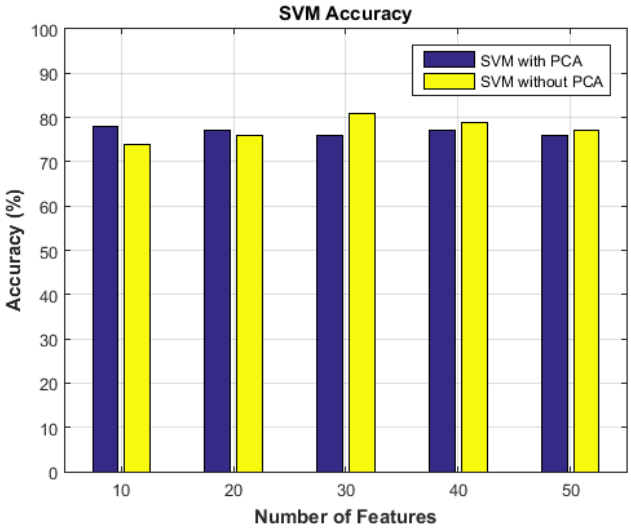
[24] Classifier

زمانی که در مسائل یادگیری با تعداد ویژگی‌های زیاد روبرو هستیم, کاهش ابعاد می‌توان حجم محاسبات را بدون دخالت تأثیرگذار در دقت دسته‌بند, کاهش دهد. در این مقاله برای کاهش ابعاد از روش تحلیل اجزای اصلی[۲۵] استفاده شده است. در این روش با استفاده از مقادیر ویژه, بردارهای ویژه و ماتریس کوواریانس ارتباطات بین ویژگی‌ها شناسایی شده و تحت عنوان ویژگی‌های جدید مشخص می‌شوند. در روش پیشنهادی ۵۰ ویژگی به عنوان ورودی الگوریتم تحلیل اجزای اصلی داده می‌شوند و ۵۰ ویژگی خروجی جدید به ما داده می‌شود. اما ویژگی‌های خروجی به ترتیب اولویت ارزش برای دسته‌بندی, مرتب شده‌اند.

ح) نتایج ارزیابی

درجه بندی شعب بانک به‌وسیله دو روش ماشین بردار پشتیبان و شبکه عصبی انجام گرفت و نتایج آن‌ها به صورت جداگانه در ادامه آمده است. برای ارزیابی از معیارهای استاندارد دقت[۲۶]، کارایی[۲۷] و صحت[۲۸] استفاده شده است.

نتایج ارزیابی استفاده از ماشین بردار پشتیبان به عنوان دسته‌بند و توسط معیار صحت در شکل (۴) نشان داده شده است. نتایج نشان می‌دهد که کارایی این دسته‌بند در روش پیشنهادی خیلی زیاد نیست و میزان صحت آن اغلب کمتر از ۸۰ درصد است. آزمایشات یک بار بدون استفاده از PCA و بار دیگر با استفاده از PCA انجام شده است. زمانی که دسته‌بند از تعداد کم ویژگی-ها استفاده می‌کند مشاهده می‌شود که PCA باعث افزایش میزان صحت شده است اما در مجموع تأثیر آن در روش ماشین بردار پشتیبان کم است.



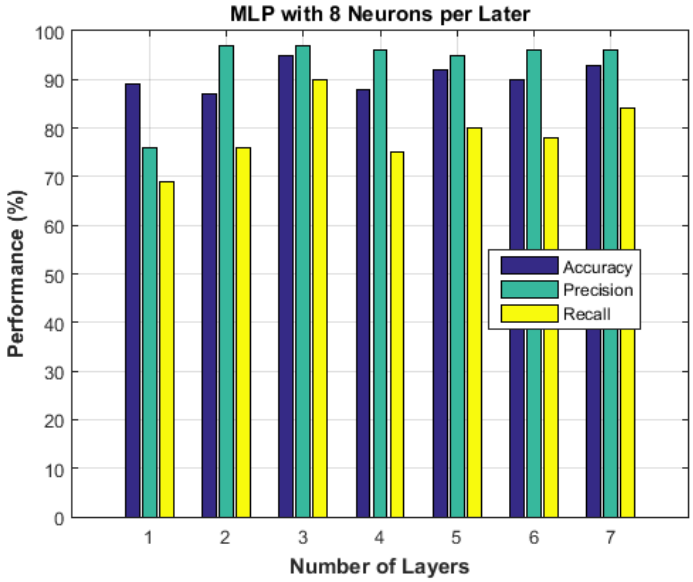
شکل ۴. نتایج ارزیابی ماشین بردار پشتیبان

ط) ارزیابی شبکه عصبی

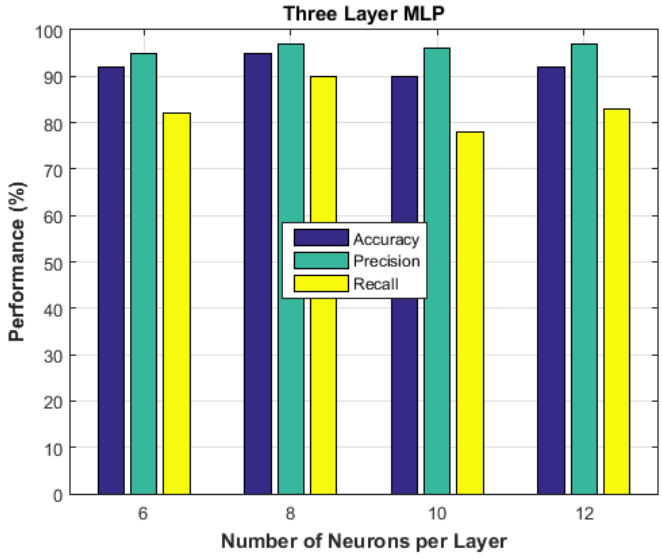
برای ارزیابی بهتر شبکه‌های عصبی تعداد لایه‌های مورد استفاده و همچنین تعداد نرون‌های هر لایه بصورت متغیر درنظر گرفته شده است. در آزمایش اول مربوط به استفاده از شبکه عصبی تعداد لایه‌های از یک تا ۷ انتخاب شدند و تعداد نرون‌های هر لایه ۸ عدد است. نتایج مربوط به معیارهای ارزیابی در شکل (۵) نشان داده شده است. زمانی که تعداد لایه‌ها کم باشد شبکه خیلی ساده است و توانایی یافتن ارتباط دقیق بین ورودی و خروجی را ندارد به همین دلیل دسته‌بند نسبتاً ضعیف عمل می‌کند و زمانی که تعداد لایه‌های زیاد باشند, شبکه خیلی پیچیده شده و به داده‌های آموزش خیلی حساس می‌شود در این حالت هم نتایج دسته‌بند خیلی عالی نیست. در این آزمایش بهترین مقادیر مربوط به دقت, کارایی و صحت مربوط به زمانی است که از سه لایه پنهان استفاده شده است. در این حالت مقادیر به ترتیب ۹۸ درصد, ۹۰ درصد و ۹۶ درصد است که به نسبت ماشین بردار پشتیبان نتایج بسیار بهتری است.

[25] Principal Component Analysis (PCA)
[26] Precision
[27] Recall
[28] Accuracy

در یک آزمایش دیگر به منظور اینکه بهترین حالت تعداد نرون‌های لایه مخفی را مشخص کنیم یک شبکه با سه لایه مخفی درنظر گرفته شد. تعداد نرون‌های لایه مخفی ۶, ۸, ۱۰ و ۱۲ در نظر گرفته شدند. نتایج مربوط به معیارهای ارزیابی آن در شکل (۶) نشان داده شده است. نتایج نشان می‌دهد بهترین نتایج مربوط به زمانی است که تعداد نرون‌های لایه مخفی ۸ نرون است.



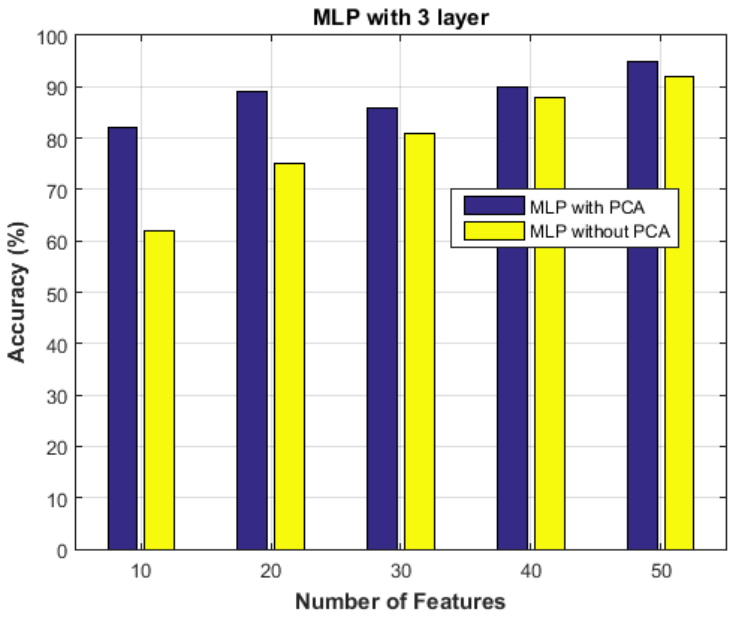
شکل شماره (۵). بررسی تعداد لایه‌های پنهان در کارایی شبکه عصبی



شکل شماره (۶). بررسی تعداد نرون‌های لایه مخفی در کارایی شبکه عصبی

برای اینکه تأثیر استفاده از PCA در شبکه عصبی مشخص شود در یک آزمایش دیگر با استفاده از تعداد ویژگی‌های مختلف, یک بار بدون استفاده از PCA با دیگر با استفاده از آن, ویژگی‌های به شبکه عصبی با سه لایه پنهان و تعداد ۸ نرون برای هر لایه داده شد. نتایج در شکل (۷) نشان داده شده است. نتایج نشان می‌دهد که انجام PCA برخلاف ماشین بردار پشتیبان, تأثیر بسیار زیادی در نتایج شبکه عصبی دارد. به عنوان مثال با استفاده از PCA و فقط توسط ۱۰ ویژگی اول آن, میزان صحت دسته‌بند ۸۲ درصد است در صورتی که با استفاده از ماشین بردار پشتیبان و توسط تمامی ۵۰ ویژگی میزان صحت ۷۷ درصد بود. این آزمایش همچنین به ما نشان می‌دهد که زمانی که تعداد ویژگی‌های انتخابی برای دسته‌بند کم باشد PCA می‌تواند نقش بسیار مفیدی برای شبکه عصبی داشته باشد. نکته جالب این است که زمانی که تمامی ویژگی‌ها استفاده شده است میزان صحت

با استفاده از PCA, ۹۶ درصد و بدون استفاده از آن ۹۲ درصد می‌باشد، یعنی PCA صرفنظر از کاهش ابعاد می‌توان داده‌های بهتری را برای ورودی شبکه عصبی تولید کند.



شکل شماره (۷). بررسی تاثیر PCA روی کارایی شبکه عصبی

۳- نتایج و بحث

بانکها در سه بخش تجهیز منابع، تخصیص منابع و ارائه خدمات فعالیت می‌کنند، این فعالیت‌ها توسط شعب در سرتاسر کشور بعمل می‌آید. در مکانسیم درجه‌بندی، شعب بانک با توجه به شاخص‌های مختلفی از جمله حجم فعالیت بانکی، مکان جغرافیایی، دسترسی به امکانات نرم افزاری و سخت افزاری به شعب ممتاز و درجات یک، دو، سه، چهار و پنج طبقه‌بندی می‌گردند. سطح فعالیت و مسئولیت بانکها به درجه آنها بستگی دارد، تعیین صحیح درجه یک شعبه از آنجا که مشخص کننده سطح اختیارات و مسئولیت‌های آن شعبه است می‌تواند منجر به رشد شعبه و ایجاد انگیزه برای پرسنل شده و به دنبال آن سودآوری برای بانک به ارمغان آورد و برعکس درجه‌بندی اشتباه ممکن است علاوه بر تضییع حقوق پرسنل زمینه‌ساز بروز فسادهای مالی و اداری باشد. لذا درجه بندی شعب مورد توجه مدیریت بانکها می‌باشد. روش‌های درجه بندی فعلی بانکها که غالباً با انتخاب شاخص‌ها و درجه اهمیت آنها به صورت میانگین وزنی توسط عده معدودی از کارشناسان انجام می‌شود، روش علمی نبوده و با تغییر کارشناسان، این شاخص‌ها و ضرایب تغییر می‌کند. بنابراین تجدید نظر در این روش‌ها و انتخاب روش‌های علمی ضروری است.

در این مقاله یک روش خودکار جهت دسته‌بندی شعب بانک معرفی گردید. برای شناسایی دسته‌بندی بهینه در فرآیند یادگیری، از دسته‌بندهای ماشین بردار پشتیبان و شبکه عصبی چند لایه پرسپترون استفاده شده است. آزمایش‌ها بر روی داده‌های واقعی جمع‌آوری شده مربوط به شعب مختلف بانک تجارت انجام گرفت. نتایج ارزیابی مربوط به ماشین بردار پشتیبان نشان داد که این دسته‌بند دارای کارایی پایین‌تری برای روش پیشنهادی است. اما استفاده از شبکه‌های عصبی و ترکیب آن با PCA نشان داد که دارای معیارهای کارایی بالایی است. مقادیر مربوط به دقت کارایی و صحت با استفاده از شبکه عصبی به ترتیب ۹۸ دصد، ۹۰ درصد و ۹۶ درصد بدست آمد.

یکی از چالش‌های موجود در این تحقیق عدم وجود مراجع مشابه جهت مقایسه‌ی عملکرد روش پیشنهادی است. که بدین منظور جهت رفع این نقیصه سعی شده است که پارامترهای مختلف جهت شناسایی بهترین دسته‌بند شناسایی شود. نتایج بدست آمده نشان می‌دهد که سیستم پیشنهادی توانسته است که با دقت بسیار مناسبی درجه‌بندی شعب مختلف بانک را انجام دهد. لذا یافته‌های این تحقیق می‌تواند به بانکها در تعیین دقیق پارامترها به منظور انجام درجه بندی و در نتیجه تعیین سطح مطلوب اختیارات و مسئولیت های شعبه کمک نماید و این امر می تواند منجر به رشد شعبه (ارتقای درجه) و ایجاد انگیزه برای پرسنل و به دنبال آن سودآوری برای بانک گردد.

محققان دیگر می‌توانند با بکارگیری فرآیند تحلیل شبکه‌ای (ANP) و یا دیگر روش‌های تصمیم گیری چندمعیاره (MCDM) برای انتخاب دقیق‌تر ویژگی‌ها و محاسبه ضرایب اهمیت ابعاد و شاخص‌ها اقدام کنند.

٤-منابع

1. Agany, D., Pietri, J., and Gnimpieba, E. (2020). Assessment of vector-host-pathogen relationships using data mining and machine learning. *Computational and Structural Biotechnology Journal*, 18 (1), 1704-1721.
2. Azar, A., Ahmadi, P., & Sabt, M.V. (2010). Model Design for Personnel Selection with Data Mining Approach (Case Study: A Commerce Bank of Iran).*Journal Of Information Technology Management*, 2(4), 3-22 (In Farsi).
3. Bahari, F., and Elayidom, S.(2015). An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. International Conference on Information and Communication Technologies. Procedia Computer Science. 46(1),725 – 731
4. Barchman,R.J., Khabaza,T., Kloesgen,W., Piatetsky-Shapiro,G., and Simoudis,E. (1996). Mining Business Databases. *Communications of the ACM*.39 (11).
5. Berry,M.J.A. and Linoff,G.S.(2002).Mastering Data Mining. John Wiley & Sons, New York.
6. Bradley,P.S.,Fayyad.U.M., and Mangasarian,O.L.(1999). Mathematical Programming for Data Mining: Formulation and Challenges. *Journal on Computing*, 11(3),217-238.
7. Cabena,P., Hadjinian,P., Stadler,R., Verhees,J., and Zanasi,A. (1998). Discovering Data Mining: From Concept to Implementation. Prentice Hall. New Jersey.
8. Carvajal, M., Viacrusis, K., Hernandez, L., Amalin, D., and Watanabe K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC*. 18(1), 183-196.
9. Chen, Y., Kuo, M., Wu, S., Tang, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data, Electronic Commerce Research and Applications, 8 (5), 241–251.
10. Chye,K.H. and Gerry, C.K.L. July-Dec.(2002). Data Mining and Customer Relationship Marketing in the Banking Industry:Singapore. *Management Review*, 24(2),1-27.
11. Coenen, F. (2011). Data mining: past, present and future. The Knowledge Engineering Review. 26(1), 25-29.
12. Crook, J.N., Edelman,D.B., and Thomas,L.C. (2001). Editorial Review. Journal of the Operation Research Society. 52(9).972-973.
13. Deneke C, Rentzsch R, Renard Y. and Paprbag. (2017). a machine learning approach for the detection of novel pathogens from NGS data. 7(1), 25-38.
14. Decker,P. (1998). Data Mining's Hidden Dangers. Banking Strategy.6-14.
15. Deshpande, S.P. & Thakare, V.M. (2010). Data Mining System and Applications: A Review. International Journal of Distributed and Parallel systems (*IJDPS*), 1(1), 32-44.
16. Divandari, A., Shabahang, R., Mohammdpoor, M., Moosavi, R. (2007). Designing Portfolio Management System for Banking Loans By Using Data Mining Technology. 6(4), 207-236 (In Persian).
17. El-zehry A.M., El-bakry H.M, and El-ksasy M.S.,(2013). Applying Data Mining Techniques for Customer Relationship Management: A Survey. *(IJCSIS) International Journal of Computer Science and Information Security*, 11(11), 1-8.
18. Fabris, P. 1998. Advanced Navigation.CIO. May 15.
19. Fayyad, U. & Piatetsky-Shapiro, G. & Smyth, P. (1997). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence (AAAI), 17 (3), 37-54.
20. Groth. R. (1998). Data Mining: A Hands-On Approach for Business Professionals. Prentice Hall, New Jersey.
21. Hasanzadeh, A., Ghanbari, M.H., & Elahi, S. (2012). Classification Of Mobile Banking Users By Data Mining Approach: Comparison Between Artificial Neural Networks And Naive Bayes Techniques. Management Research In Iran (*Modares Human Sciences*), 16(2), 57-71 (In Persian).

22. Hwang,H., Ku,C., Yen,D.C., and Cheng,C. (2002). Critical Factors Influencing the Adoption of Data Warehouse Technology: A Study of the Banking Industry in Taiwan. *Decision Support Systems*, 10(39), 1-21.
23. Jayachandran, S., Sharma, S., Kaufman, P., Raman, P. (2005). The role of relational information process and technology use in customer relationship management. *Journal of Marketing*, 69(4), 177–192.
24. Karakatsanis, L., AlKhader, W., MacCrory, F., Alibasic, A., Atif, M.O., Aung, Z., and Lee,W.W.(2016). Data Mining Approach to Monitoring the Requirements of the Job Market: A Case Study. Information Systems. 4-21.
25. Khajevand, S., Taghavi, M., & Najafi, E. (2012). Customer classification of Bank Saderat Iran using data mining. *Journal of Management Studies*, 22(179), 67-82 (In Persian).
26. .Khanbabei, M., Movahedi, F., Alborzi, M., & Radfar, R. (2019). Framework of Using Data Mining for Business Process Improvement. *Production and Operations Management,* 10(1), 25-45 (In Persian).
27. Kim, N. & Atuahene-Gima, K. (2010). Using Exploratory & Exploitative Market Learning for New Product Development. *Journal of Product Innovation Management*, 27(4), 519-536.
28. Kiss, F. (2003). Credit scoring process from a knowledge management prospective. Periodica Polytechnica Ser. *SOC. MAN. SCI*, 11 (1), 95-110.
29. LEE, S. J. & SIAU, K. (2001). A review of data mining techniques. *Industrial Management & Data System*, 101(1), Pages 41-46.
30. Lin, C.C., Chiu, A.A., Huang, Y.S., and Yen, D.C.(2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89(1), 459–470.
31. Luo, Y., Zheng, Q., Jayaraman, V. (2010). Managing Business Process Outsourcing. *Journal of science Direct*, 3(1), 205-217.
32. Ltifi, H., Benmohamed,E., Kolski,C., and Ben Ayed,m.(2016). Enhanced visual data mining process for dynamic decision-making. *Knowledge-Based Systems*. 112(1), 166–181.
33. Marbán, O. & Segovia, J. & Menasalvas, E. & Fernández-Baizán, C. (2009). Toward data mining engineering: a software engineering approach. *Information Systems*, 34 (1), 87-107.
34. Mesforoush, A., Tarokh, M.J. (2013). Customer Profitability Segmentation for SMEs Case Study: Network Equipment Company. *Novel science, Research in Industrial Engineering*, 2(1), 30-44.
35. Pourzarandi, M.E., Shabahang, R., and Lucas, C., (2004).Designing and Developing an Appropriate Inventory Management Model using Data Mining and Knowledge Discovery Technology, The Application of Neural Networks, Economic & Management. *Quarterly Journal of the Islamic Azad University,* (61), 92-105.
36. Radfar, R., Nezafati, N., & Yoosefi, S. (2014).Classification Of Internet Banking Customers Using Data Mining Algorithms. *Journal of Information Technology Management*, 6(1), 71-90 (In Persian).
37. Rangriez, H., & Bayrami, Z. (2019).The Impact of E-CRM on Customer Loyalty Using Data Mining Techniques. *Journal of Business Intelligence Management Studies*, 7(27), 175-205 (In Persian).
38. Rygielski C., Wang J. C.,Yen D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(1), 483-502.
39. Sepehri M M, norouzi A, Chubdar S, & Teymourpour B. (2015). Developing a model for discovering the causes of customer churn from banking services via hybrid approach of data mining and survey. *IQBQ*. 15 (4), 97-125 (In Persian).
40. Shmueli G., Patel Nitin R., Bruce Peter C.(2007). Data mining for business intelligence concepts, techniques, and applications";in Microsoft Office Excel with Xlminer, Wiley-Interscience
41. Sing'oei, L. and Wang,J. (2013). Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing. *International Journal of Computer Science Issues*, 10(2), 1-7
42. Thillainayagam V., (2012). Data Mining techniques and applications-A review. *I-manager's Journal on Software Engineering*, 6(1), 1-6.
43. Torkzadeh, G., Cha-Jan Chang, J., Hansen, G.W. (2006). Identifying issues in customer relationship management at Merck-Medco. *Decision Support Systems, Journal of science Direct*, 42 (2), 1116-1130.

44. Teymourpour, B. (2015). Applications and challenges of data mining in the banking industry, Fifth Annual Conference on Electronic Banking and Payment Systems (In Persian).
45. Turki M. Alkheliwi. (2011). Enhancing risk prediction in financial application using data mining and game theory principles. THESIS Submitted in partial satisfaction of the requirements for the degree of MASTER OF SCIENCE, Michigan State University.
46. Yaning, L. & Kia, L. & Fei, G. (2008). Data mining approach and application in CRM project for internet-focused banking. IEEE, Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on, 12-14 Oct. 1-4.
47. Yoon, Y. (1999). Discovering Knowledge in Corporate Database. *Information Systems Management*, 16(2), 64-71.

Explaining the Categories of Support Vector Machine and Neural Network for Ranking of Bank Branches

Davod Khosroanjom(Corresponding Author)
Department of Management, College of Management, & Economic, PhD Student at Tarbiat Modres. Tehran. Iran
Mohammad Alyasi
Master of soft ware engenering
Behzad Keshanchi
Department of Banking Management, College of Economic Sciences, School of Banking Sciences, Tehran, Iran
Bahare Boobanin
Department of Business Management, College of Management, Sannadaj Branch, Islamic azad university Sannadaj , Iran.
Shovane Abdollai
Department of Public Management, College of Management, Mahabad Branch, Islamic Azad University

Abstract
There is a lot of information in the banking industry that is of particular importance in identifying it. The use of data mining techniques not only improves quality but also leads to competitive advantages and market positioning. By using data mining and in order to analyze patterns and trends, banks can predict the accuracy of how bank branches are ranked. In this paper, the branches of one of the large commercial banks (number of selected branches 1825 branches and the number of features used 57 features) were performed on real data using support vector machine categories and multi layer perceptron neural network. The evaluation results related to the support vector machine showed that this classifier has lower efficiency for the proposed method. However, the use of neural networks and its combination with PCA showed that it has high performance criteria. Values related to efficiency and accuracy was obtained using neural network with very high accuracy.
Keywords: Data mining, Banking, Support vector machine, neural network.