فصلنامه مدیریت صنعتی دانشکده علوم انسانی، دانشگاه آزاد اسلامی
واحد سنندج – سال سیزدهم/ شماره ۴۴ /تابستان ۱۳۹۷

پیشبینی رویگردانی مشتریان بانک با استفاده از روش داده کاوی

پروین نجمی
کارشناس ارشد مدیریت فناوری اطلاعات، واحد تهران شمال دانشگاه آزاداسلامی، تهران، ایران
عباس راد (نویسنده مسؤول)
استادیار دانشکده مدیریت دانشگاه شهید بهشتی
Email: a-raad@sbu.ac.ir
مریم شعار
استادیار دانشکده مدیریت واحد تهران شمال، دانشگاه آزاداسلامی، تهران، ایران
تاریخ دریافت: ۹۷/۲/۱۰ * تاریخ پذیرش: ۹۷/۸/۱۱

چکیده

شدت پیداکردن رقابت درفضای صنعتی و اقتصادی و حرکت بازار به سمت بازار رقابت کامل، باعث شده است که تمایل بنگاه‌های اقتصادی به جذب مشتریان بیشتر کاهش یافته و در عوض، تمایل به فعالیت در زمینه‌های مختلف خدماتی و تولیدی افزایش یافته است. به همین منظور، توسعه روش هایی به منظور شناسایی مشتریان رویگردان و پیش بینی رویگردانی، از مهمترین فعالیت های حوزه فروش به حساب خواهد آمد. در صورتی که بانک فرصت کافی برای پیش بینی رویگردانی مشتریان را داشته باشد؛ می تواند از این مسئله جلوگیری کرده و یا حداقل به اصلاح ساختارها و خدمات خود به منظور جلوگیری از ریزش تعداد بیشتری از مشتریان بپردازد. تحقیق حاضر، به منظور توسعه چنین مدلی برای بانک شهر صورت گرفته است. به همین منظور، از یک الگوریتم دو مرحله ای خوشه بندی، دسته بندی داده کاوی استفاده شده است. به منظور خوشه بندی مشتریان، از نقشه های خودسازمان ده شبکه عصبی که یک روش یادگیری نظارت نشده است؛ استفاده شده و برای دسته بندی از ماشین بردار پشتیبان و درخت تصمیم استفاده شده است. روش استفاده از این ابزارها به این صورت است که ابتدا از دو مشخصه میانگین موجودی و میانگین تراکنش مشتریان در دوره سه ماهه پایانی استفاده شده و به عنوان ورودی شبکه عصبی در خوشه بندی مورد استفاده قرار گرفته است. پس از آن، در مرحله کلاس بندی، از داده های مربوط به تراکنش های نقدی و اعتباری به منظور کلاس بندی و پیش بینی استفاده شده است. نتایج به دست آمده از دقت کلاس بندی صورت گرفته، حاکی از آن است که مدل پیشنهادی تا بیش از ۸۰٪ توانایی پیش بینی رویگردانی مشتری را داشته و در این مسیر، ماشین بردار پشتیبان عملکرد بهتری از درخت تصمیم نشان داده است.

کلمات کلیدی: مشتری رویگردان، شبکه عصبی، نقشه خود سازمان ده، ماشین بردار پشتیبان، درخت تصمیم.

۱- مقدمه

از زمان انقلاب صنعتی و به خصوص بعد از جنگ جهانی دوم که تولید صنعتی رونق گرفته و بنگاه های تولیدی به عرضه محصولات خود در سطح انبوه در بازار های مختلف پرداختند؛ سیاست اصلی درآمدزایی شرکت ها، کاهش هزینه ها و افزایش فروش بوده است. در واقع، افزایش سهم بازار با استفاده از قیمت های کمتر و کیفیت بالاتر، مهمترین عملکرد مدیریت بازاریابی و فروش شرکت ها بوده است. با گذشت زمان و بهبود تکنولوژی، فاصله کیفی بین کالاهای مختلف کمتر شد و عملا کالاها در بازارهای مختلف به سمت همگنی پیش رفت. همچنین، با اضافه شدن بنگاه های اقتصادی بیشتر و مصرفی شدن هرچه بیشتر فرهنگ مردم، دو سمت عرضه و تقاضا در صنایع مختلف نیز رشد بالایی را تجربه کرده و در نهایت، بازارها به سمت رقابت کامل پیش رفتند که در آن یک بنگاه اقتصادی نمی تواند با اجرای سیاست های قیمتی یا کیفی، به جلب مشتریان جدید بپردازد.

در چنین فضایی، توسعه سهم بازار و افزایش فروش، از اولویت سیستم های بازاریابی و فروش خارج شده و جای خود را به سیاست افزایش سهم کیف پول داد. این سیاست جدید، با فرض اینکه کسب مشتریان جدید در بازار یا غیرممکن بوده و یا با هزینه های بالایی امکان پذیر است؛ پیشنهاد می کند که شرکت ها به فروش کالاها و خدمات متعدد به مشتریان خود بپردازند. در واقع، به جای جلب تعداد زیادی مشتری، تعداد زیادی کالا یا خدمت به تعداد کمی مشتری فروخته شود. این سیاست مورد استقبال انواع بنگاه های اقتصادی قرار گرفته و این مسئله تا جایی پیش رفت که توسعه خطوط تولید سنگین، جای خود را به آرایش کارگاهی به منظور تولید کالاهای بیشتر و با تنوع بالاتر داد.

در فضایی که تمرکز بنگاه های اقتصادی بر فروش بیشتر به مشتریان موجود می باشد؛ قطعا حفظ مشتریان اهمیت بسیار بالایی دارد. به همین منظور، مدیران فروش و بازاریابی شرکت ها به استفاده از روش های تحلیل آماری و ریاضی در قالب علم داده کاوی به منظور تحلیل روابط مشتریان پرداخته و به وسیله آن به آنالیز مشتریان و شناسایی بهتر آنها پرداختند. یکی از مهمترین تحقیقات توسعه یافته در این زمینه، شناسایی مشتریان رویگردان می باشد. تعاریف زیادی از رویگردانی شده است و البته این تعاریف با توجه به نوع صنایع و حتی نوع بازارها می تواند متفاوت باشد؛ لیکن به صورت کلی می توان رویگردانی را به عدم تمایل مشتری به استفاده از کالا یا خدمات ایجاد شده توسط یک شرکت خاص تعبیر نمود. شناسایی مشتران رویگردان از دو جهت برای شرکت ها اهمیت دارد؛ اول اینکه می توانند جلو ریزش مشتریان را با استفاده از پیشنهادهای ویژه گرفته و آنها را حفظ کنند و دوم اینکه می توانند با یک آنالیز رویگردانی جامع، به بررسی دلایل و نقاط ضعف شرکت و بهبود آن بپردازند.

به منظور شناسایی و تشخیص رویگردانی، شرکت ها باید اطلاعات مرتبط با مشتری های خود را تحلیل کرده و یک الگو بین آنها ایجاد کنند؛ برای مثال با تحلیل روابط مشتری بتوانند الگویی برای رویگردانی تعریف کرده و بر این اساس مشتریان رویگردان را شناسایی کنند. بنابراین، گام اول تحلیل رویگردانی مشتریان، مربوط به جمع آوری داده های مرتبط با مشتریان می باشد. با توجه به اینکه روابط مشتریان با شرکت ها در بسیاری از موارد محدود بوده و یا شرکت ها توانایی دسترسی به اطلاعات تمامی مشتریان ندارند؛ تحلیل رویگردانی به حوزه هایی مربوط می شود که توانایی شناسایی مشتریان و ردگیری استفاده او از خدمات یا کالاها را دارند. به همین دلیل تحلیل رویگردانی بیشتر در صنعت اطلاعات و ارتباطات انجام می پذیرد؛ چرا که شرکت های ارائه دهنده این خدمات به راحتی توانایی دسترسی به اطلاعات مشتریان نظیر طول مدت مکالمه، تعداد مکالمه، پیامک ها، بسته های اینترنتی و سایر موارد را دارند و لذا می توانند با تحلیل این اطلاعات، مشتریانی که نسبت به شرکت رویگردان شده اند را شناسایی کنند.

امیرشاهی و روان بد (۱۳۸۵) نشان داد ۱۲ عامل مؤثر توسط تکنیک آنالیز عاملی به منظور تقویت اعتماد مشتریان شناخته شده است که از جمله می توان به مشاوره، کارکنان، خدمات مکمل، شهرت و پاسخگویی بانک در قبال مشتریان اشاره کرد. محققان توانایی این عوامل در توجیه حس اعتماد مشتریان در مجموع را تا ۶۵٪ تخمین زده اند. بهرام زاده و مقرب (۱۳۸۹) اعتقاد دارند که میزان سوددهی بانک، رابطه مستقیم با میزان اعتماد مشتریان داشته و به منظور آزمون این موضوع، از مشتریان بانک های پاسارگاد، اقتصاد نوین و سینا در استان خوزستان استفاده شده است. بر اساس نتایج به دست آمده، وفاداری، رضایتمندی، اعتماد و تعهد از مهمترین عوامل مؤثر بر وفاداری مشتریان بوده اند. محمدی (۱۳۹۰) عوامل کارکنان، خدمات و محصولات، فرآیندهای

کاری، قیمت خدمات و محصولات، ویژگی شعب، فرآیندهای کاری و فعالیت های بازارافزایی با ترتیب های مختلف، به عنوان مهمترین عوامل مؤثر بر وفاداری شناسایی کرده‌اند.

نخعی و همکاران (۱۳۹۳) نتیجه گرفته اند که نرخ کارمزد پایین، اولین عامل جذب مشتریان می باشد. دومین عامل، تقویت امکانات و فناوری بوده و عوامل مدیریت سرمایه، عوامل فیزیکی و ارتباطات مستمر با مشتری به عنوان عوامل مرثر بعدی شناخته شده اند.

هایگورت (۲۰۱۵) بیان می کند که هزینه ایجاد مشتریان جدید به دلیل رقابت شدید، بسیار زیاد است. محقق، در این تحقیق به استفاده از ابزارهای داده کاوی به منظور پیش بینی رویگردانی پرداخته است. به منظور انتخاب مشخصه های مؤثر در رویگردانی، از نظر خبرگان بازار بیمه و تحقیقات مشابه انجام شده، استفاده شده است و در نهایت سن و میزان استفاده از امکانات بیمه، مهمترین مشخه ها انتخاب شده اند. ابزارهای استفاده شده توسط محقق، شامل رگرسیون لاجیت، درخت تصمیم، شبکه عصبی و ماشین بردار پشتیبان بوده است پس از تعیین پارامترهای بهینه برای هریک از روش ها و تعیین نمونه های آموزش و کنترل در هر مورد، کارایی شبکه عصبی و رگرسیون لاجیت از دو روش دیگر به صورت معناداری بالاتر تخمین زده شده است.

هانگ و همکاران (۲۰۱۶) بر این موضوع که شناسایی مشتریان رویگردان به منظور جلوگیری از ریزش مشتریان، اهمیت بالایی دارد به توسعه یک مدل کلاس بندی فازی پرداخته اند. بر اساس نظر مؤلفان، روش های کلاس بندی مبتنی بر قواعد، کارایی بهتری در کلاس بندی مشتریان رویگردان دارند؛ چرا که قواعد استخراج شده از این روش ها می تواند مورد استفاده شرکت جهت بهبود عملکرد و شناسایی نقاط ضعف قرار بگیرد؛ لیکن، این روش ها با فرض پایدار بودن نمونه تحقیق و پایگاه داده مورد استفاده ایجاد شده اند و این در حالی است که عدم تعادل در پایگاه داده می تواند نتایج را تحت تأثیر قرار دهد. به این ترتیب، محققان یک مدل کلاس بندی بر مبنای قواعد فازی طراحی کرده اند که این اشکال را برطرف می کند. بر اساس نتایج به دست آمده، کارایی و دقت روش پیشنهادی در مقایسه با روش های دیگر کلاس بندی بر مبنای قاعده، مناسب بوده است.

گردینی و وجلیو (۲۰۱۶) اظهار می دارند که تجارت الکترونیک چارچوب خوبی برای آسودگی مشتریان و آسانی معاملات فراهم کرده است؛ لیکن ریسک رویگردانی مشتری را نیز افزایش داده است؛ چرا که هزینه رویگردانی و گرایش به رقبا برای مشتریان کم است. به همین دلیل، در این تحقیق از ماشین بردار پشتیبان برای شناسایی و پیش بینی رویگردانی به منظور پیشگیری و اصلاحات مرتبط استفاده شده است. به منظور بهینه سازی پارامترهای SVM نیز از تکنیک AUC استفاده شده است. محققان، کارایی روش پیشنهادی را با روش های رگرسیون لجستیک، شبکه عصبی و SVM کلاسیک (بدون بهینه سازی پارامترها) مقایسه کرده اند و نتیجه گیری کرده اند که AUC نقش مهمی در کارایی SVM بازی می کند و باعث کارایی بهتر آن در مقایسه با سه روش رقیب شده است.

امروزه، با پیشرفت تکنولوژی و توسعه استفاده از خدمات الکترونیک، زمینه تعامل بیشتر شرکت ها با مشتریان در بسیاری از صنایع مهیا شده است. از جمله این صنایع، می توان به بانکداری اشاره نمود. بانکداری الکترونیک باعث افزایش تعامل مشتریان با بانک ها و در نتیجه افزایش استفاده آنها از خدمات مالی شده است. علاوه بر این، خدمات مالی ارائه شده توسط بانک ها معمولا بسیار جامع بوده و از ارائه انواع خدمات بانکی حضوری، اعتباری و اینترنتی تا خریدهای غیرحضوری و همچنین خدمات متفاوت دیگر همچون کارگزاری و مشاوره را شامل می گردد. بنابراین، با توجه به این خدمات گسترده و امکان استفاده غیرحضوری از این خدمات، بانک ها به سادگی به اطلاعات مشتریان دسترسی پیدا کرده و توانایی تحلیل آنها به منظور شناسایی رویگردانی را پیدا می کنند.

پژوهش حاضر، با در نظر داشتن اهمیت و ضرورت تحلیل رویگردانی، به شناسایی مشتریان رویگردان و پیش بینی رویگردانی پرداخته و به این منظور، در ابتدا الگوی رویگردانی را با استفاده از روش خوشه بندی شبکه عصبی کشف کرده و سپس با استفاده از ماشین بردار پشتیبان، به پیش بینی رویگردانی می پردازد. در همین راستا، در ابتدای تحقیق، پژوهش های مرتبط انجام شده در داخل و خارج از کشور به صورت مختصر مرور می گردد. در قسمت بعد، مفاهیم مرتبط با روش های خوشه بندی و دسته بندی در داده کاوی به روش های شبکه عصبی و ماشین بردار پشتیبان شرح داده شده و در نهایت نتایج به دست آمده از روش تحقیق، نشان داده شده و جمع بندی نهایی صورت می پذیرد.

۲- روش شناسی

الف) تدوین الگوی مفهومی

با توجه به تحقیقات انجام شده تاکنون و متغیرهای در دسترس محقق با استفاده از پایگاه های داده بانکی، متغیرهای تحقیق شامل موارد زیر می گردند:

جدول شماره (۱): الگوی مفهومی

مشخصات دموگرافیک	تعداد تراکنش	ارزش تراکنش	تراکنش های اعتباری
سن مشتری (Prasad and Mahdavi, 2012)	تعداد تراکنش	میانگین ارزش تراکنش نقدی (Mutanen, 2006)	تعداد وام های دریافتی (Prasad and Mahdavi, 2012)
مدت زمان استفاده از خدمات بانک (Adwan et al., 2014)	تعداد تراکنش نقدی (Mutanen, 2006)	میانگین ارزش تراکنش های کارت اعتباری (Prasad and Mahdavi, 2012)	تعداد وام های جاری (Prasad and Mahdavi, 2012)
درآمد (Adwan et al., 2014)	تعداد تراکنش های کارت اعتباری (Prasad and Mahdavi, 2012)	میانگین ارزش تراکنش های اینترنت	ارزش وام های دریافتی (Prasad and Mahdavi, 2012)
سطح تحصیلات (Prasad and Mahdavi, 2012)	تعداد تراکنش های اینترنتی		ارزش وام های جاری (Prasad and Mahdavi, 2012)

یک نکته در استفاده از متغیرهای معرفی شده، در نظر گرفتن مشخصات مشتریان به صورت نسبی و عدم مقایسه به صورت مطلق می باشد. برای مثال، ممکن است یک مشتری وفادار با درآمد پایین، به نسبت درآمد خود، تراکنش های بالایی را انجام داده باشد؛ ولی این میزان در مقابل تراکنش های یک مشتری رویگردان با درآمد بالا، پایین به نظر بیاید. بنابراین، تعداد و ارزش تراکنش ها باید به نسبت درآمد مشتریان تعدیل گردد.

علاوه بر این، در نظر گرفتن عملکرد مشتریان به نسبت عملکرد گذشته خود نیز ضروری می باشد. به این معنا که عملکرد تعدیل شده با درآمد یک مشتری، می تواند نسبت به مشتری دیگر بالاتر باشد؛ لیکن می تواند به نسبت عملکرد گذشته همان مشتری، کاهش چشمگیری را داشته باشد و در واقع مشتری به سمت رویگردانی حرکت کرده باشد. بنابراین، علاوه بر درآمد، متغیرهای تحقیق به نسبت میانگین متحرک ۱۲ ماهه خود نیز تعدیل می گردند :

$$x_{ij} = \frac{X_{ij}}{MA(X_{ij})} \qquad (۱)$$

در رابطه مذکور، X_{ij} مقادیر هریک از متغیرها (i) برای هریک از مشتریان (j) بوده و $MA(X_{ij})$ نیز نشان دهنده میانگین متحرک ۱۲ ماهه همان متغیر می باشد.

از آنجا که در تحقیق حاضر از روش های شبکه عصبی خودسازمان ده و ماشین بردار پشتیبان به منظور خوشه بندی و دسته بندی مشتریان استفاده می شود؛ در این قسمت به صورت مختصر به مرور روش های مذکور پرداخته شده و سپس متغیرهای تحقیق و مراحل اجرای روش تحقیق شرح داده می شود.

ب) نقشه خودسازمان ده شبکه عصبی

در یک شبکه ی خود سازمان ده، واحد های پردازش گر در گره های یک شبکه ی یک بعدی، دو بعدی یا بیشتر قرار داده می شوند. واحد ها در یک فرآیند یادگیری رقابتی نسبت به الگوهای ورودی منظم می شوند. محل واحدهای تنظیم شده در شبکه به گونه ای نظم می یابد که برای ویژگیهای ورودی، یک دستگاه مختصات معنی دار روی شبکه ایجاد شود. لذا یک نقشه ی خود سازمان ده، یک نقشه ی توپوگرافیک از الگوهای ورودی را تشکیل می دهد که در آن، محل قرار گرفتن واحدها، متناظر ویژگیهای ذاتی الگوهای ورودی است.

یادگیری رقابتی که در این قبیل شبکه ها بکار گرفته می شود بدین صورت است که در هر قدم یادگیری، واحدها برای فعال شدن با یکدیگر به رقابت پرداخته و در پایان یک مرحله رقابت تنها یک واحد برنده می شود، که وزنهای آن نسبت به وزنهای سایر واحدها به شکل متفاوتی تغییر داده می شود. این نوع از یادگیری را یادگیری نظارت نشده می نامند. شبکه های خودسازمان ده به لحاظ ساختاری به چند دسته تقسیم می شوند که از جمله می توان به شبکه همینگ، کلاه مکزیکی و کوهونن اشاره نمود. شبکه کوهونن که در این تحقیق از آن استفاده شده است؛ به صورت چند لایه ای می باشد. یک لایه کوهونن آرایه ای از نورون ها به صورت یک بعدی، دو بعدی یا بیشتر است که نمونه ای از آن را در شکل فوق مشاهده می شود. در فاز یادگیری هر یک از واحدها فاصله ی بردار ورودی تا وزن های خود را به صورت زیر محاسبه می کنند.

$$I_i = D(X, w_i) \qquad (۲)$$

که D تابع سنجش فاصله است و می توان هر یک از توابع مرسوم برای سنجش فاصله را استفاده نمود؛ مثلا فاصله ی کمان کروی یا فاصله ی اقلیدسی را می توان استفاده نمود. واحدها با این محاسبه می خواهند بدانند نزدیکترین بردار وزن به بردار ورودی را دارند یا خیر؛ که این همان بخش رقابتی در اینگونه از شبکه ها می باشد.

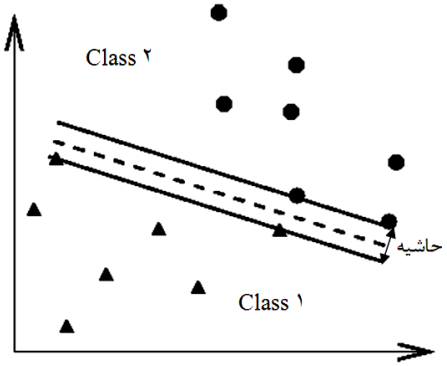
واحد دارای نزدیکترین وزن به بردار ورودی، برنده این مرحله از رقابت خواهد بود که برای آن Z_i مربوطه برابر ۱ قرار داده می شود و سایر Z_i ها برابر صفر خواهند بود. آنگاه قانون کوهونن که به صورت ذیل است برای به روز رسانی وزن ها استفاده می شود:

$$w_i^{new} = w_i^{old} + \alpha(X - w_i^{old})Z_i \qquad (۳)$$

به این ترتیب، توسط این قانون، می توان به صورت جداگانه به بررسی اوزان و بهینه سازی آنها در هر مرحله از آموزش شبکه پرداخت.

ج) ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از تکنیک های دسته بندی است که بر اساس تئوری یادگیری آماری پایه گذاری و معرفی شده است(Vapnik, 1998). ماشین بردار پشتیبان به دنبال یافتن ابرصفحه‌ای است که با استفاده از آن داده‌ها را دسته بندی کند. این ابرصفحه به گونه‌ای یافته می‌شود که حاشیه دسته بندی را حداکثر نماید، شکل شماره (۱).



شکل شماره (۱): ابرصفحه ماشین بردار پشتیبان

مدل کلی برای یافتن ابرصفحه ای مطابق با شکل فوق را می توان به صورت رابطه زیر نوشت:

$$\max \min_{i=1,\dots,n} d(X_i,H) \quad (۳)$$

و محدودیت های مسئله باید به گونه ای باشند که داده های دسته یک در نیم فضای مثبت و داده های دسته دوم در نیم فضای منفی قرار بگیرند. به عبارت دیگر می توان محدودیت های مسئله را همانند زیر بدست آورد:

$$\begin{cases} \omega^T X_i + b \geq 0 \quad \forall X_i \in class\ 1 \\ \omega^T X_i + b \leq 0 \quad \forall X_i \in class\ 2 \end{cases} \quad (۴)$$

از آنجایی که ابرصفحه دسته بند در ماشین بردار پشتیبان متعارفی است. فاصله هر کدام از بردارهای پشتیبان تا ابر صفحه متعارفی طبق رابطه روابط زیر است:

$$\frac{\omega^T X_i + b}{\|\omega\|} = \frac{1}{\|\omega\|} \quad \forall\ supportvector \in class\ 1 \quad (۵)$$

$$\frac{\omega^T X_i + b}{\|\omega\|} = \frac{-1}{\|\omega\|} \quad \forall\ supportvector \in class\ 2$$

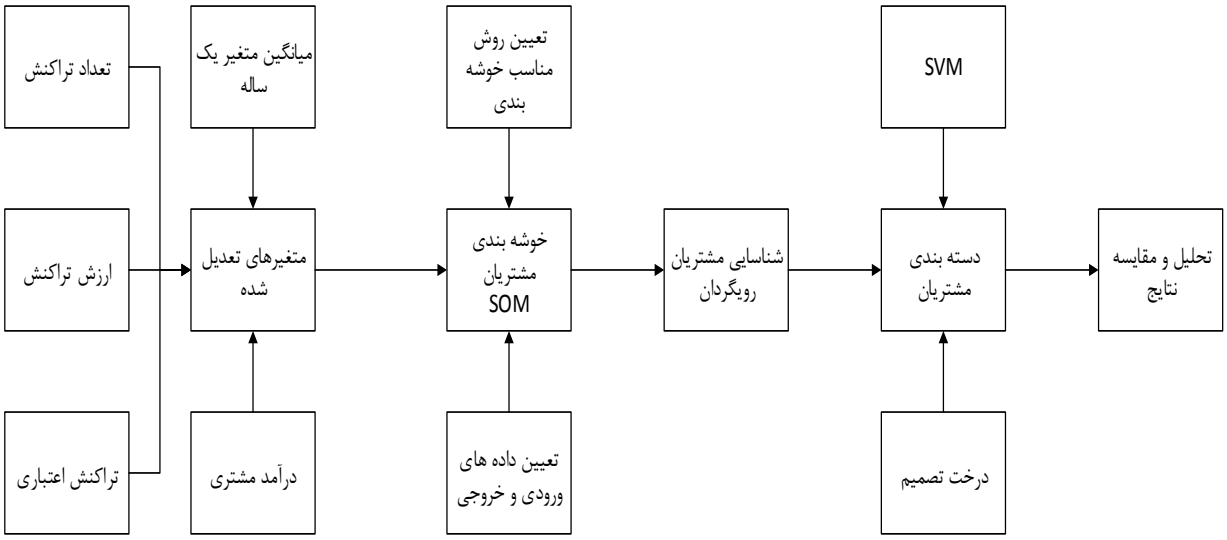
بنابراین حاشیه دسته بندی از رابطه زیر محاسبه می شود:

$$\frac{1}{\|\omega\|} - \frac{1}{\|\omega\|} = \frac{2}{\|\omega\|} \quad (۶)$$

تابع هدف ماشین بردار پشتیبان ببه صورت ماکزیمم سازی حاشیه تشکیل می گردد.

۳- نتایج و بحث

مراحل اجرای تحقیق به صورت کلی مطابق شکل (۲) می باشد؛ محقق در ابتدا داده های مربوط به مشتریان را از پایگاه های داده مربوطه استخراج کرده و سپس متغیرهای تحقیق را ایجاد نموده است. در ادامه، متغیرها به منظور قابل مقایسه بودن، با درآمد و میانگین تعدیل شده و آماده استفاده در فرآیند خوشه بندی و دسته بندی می گردند. خوشه بندی تحقیق با استفاده از روش نقشه خود سازمان ده شبکه عصبی و دسته بندی با استفاده از ماشین بردار پشتیبان صورت می گیرد.



شکل شماره (۲): توانایی ماشین بردار پشتیبان

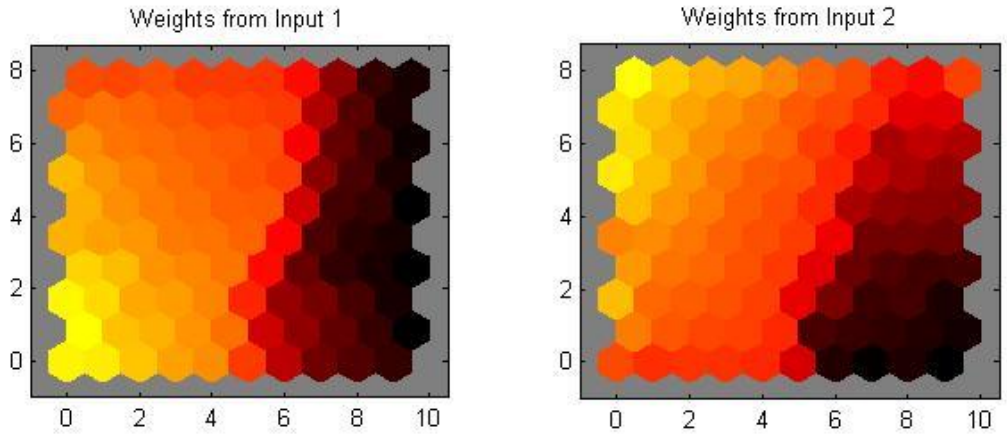
همانگونه که از شکل مشخص می گردد؛ به منظور بررسی توانایی ماشین بردار پشتیبان در دسته بندی و پیش بینی رویگردانی، از درخت تصمیم نیز به عنوان معیار مقایسه استفاده شده است. مورد مطالعاتی تحقیق، مشتریان بانک شهر در بازه سال های ۱۳۹۳ تا ۱۳۹۵ می باشد که تعداد مورد نیاز نمونه با استفاده از روش کوکران و نمونه گیری توسط روش خوشه بندی تصادفی صورت گرفته است. مشخصه های آماری مربوط به نمونه ایجاد شده همانند جدول (۱) می باشد.

جدول شماره (۲): مشخصه های آماری مربوط به نمونه ایجاد شده

آماره	تغییرات ارزش تراکنش نقدی	تغییرات تعداد تراکنش نقدی	تغییرات تراکنش اعتباری	تغیرات میانگین موجودی	تغییرات ارزش کل تراکنش
کمینه	-۰/۴۸۰	-۰/۴۷۸	-۰/۴۹۱	-۰/۸۸۳	-۰/۶۷۴
بیشینه	۰/۵۲۳	۰/۵۵۷	۰/۴۸۷	۰/۵۳۴	۰/۳۷۴
میانه	۰/۰۶۴	۰/۰۵۷	۰/۰۵۳	۰/۰۳۲	۰/۰۰۴
میانگین	۰/۰۴۲	۰/۰۴۱	۰/۰۳۶	-۰/۱۲۲	-۰/۰۹۴
انحراف معیار	۰/۲۰۴	۰/۱۹۸	۰/۱۹۸	۰/۳۹۱	۰/۲۶۳
چولگی	-۰/۳۷۲	-۰/۳۹۶	-۰/۵۳۰	-۰/۶۳۷	-۰/۷۴۲
کشیدگی	۲/۸۸۷	۲/۹۰۶	۲/۹۷۴	۱/۹۸۶	۲/۲۶۰

لازم به ذکر است که مشخصه های آماری مربوطه در جدول (۱)، مربوط به متغیرهای تعدیل شده با درآمد می باشد. مشاهده می شود که مقدار چولگی و کشیدگی در مورد تمامی متغیرها در حد نرمال یا با مقدار کم می باشد و لذا می توان فرم تمام آنها را به صورت نرمال فرض کرد.

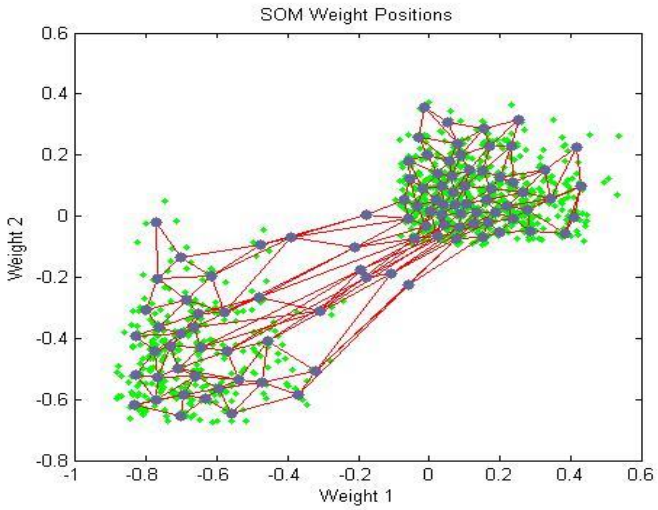
خوشه بندی مشتریان، همانگونه که پیش از این نیز ذکر شد؛ توسط روش SOM صورت می پذیرد که متغیرهای ورودی آن شامل میانگین موجودی و ارزش کل تراکنش ها می باشد. نقشه اولیه تشکیل شده، شامل یک شبکه ۱۰۰ نرونی ۱۰*۱۰ می باشد که در آن از ۷۰٪ داده های نمونه تحقیق به منظور آموزش نرون ها و بهینه سازی اوزان آنها استفاده شده است. شکل (۳) نشان دهنده اوزان اخذ شده توسط نرون ها به تفکیک دو متغیر ورودی می باشد.



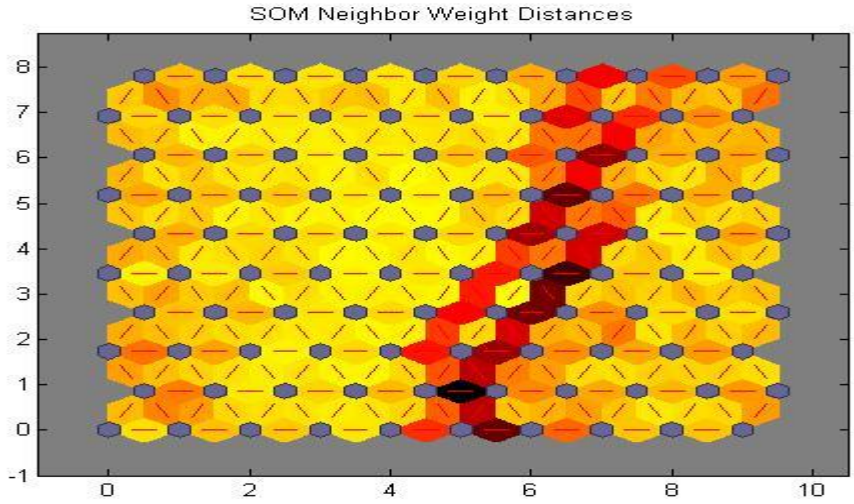
شکل شماره (۳): وزنهای درون داده ۱

بر اساس شکل (۳) مشخص است که از چپ به راست، اوزان نسبت داده شده به نرون ها در مورد هر دو متغیر ورودی، افزایش یافته است و به صورت مشخص، می توان انتظار داشت که نرون های با وزن کمتر در یک خوشه و نرون های با وزن بیشتر در خوشه دیگر قرار گیرند. به منظور درک بهتر فاصله اوزان تخصیص داده شده به نرون ها، می توان آنها را در فضای دو بعدی متغیرهای ورودی یعنی میانگین موجودی و ارزش کل تراکنش مطابق شکل (۴) ترسیم نمود.

شکل شماره (۴): وضعیتهای وزن SOM



همانگونه که مورد انتظار بود؛ اوزان تخصیص داده شده به نرون ها، آنها را با فاصله زیاد از یکدیگر و در دو خوشه مشخص قرار می دهند. بنابراین، می توان نرون های با فاصله کمتر از یکدیگر را در یک خوشه در نظر گرفت. فاصله میان نرون ها در شبکه آموزش دیده، به خوبی در شکل (۵) قابل مشاهده است.



شکل شماره (۵): فواصل وزن های همجوار SOM

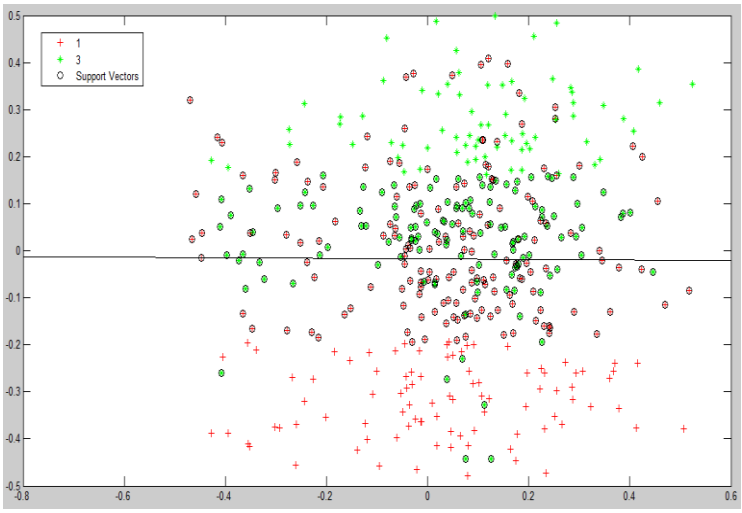
در شکل (۵)، نرون های سمت راست در فاصله اندک با یکدیگر و با فاصله زیاد از نرون های سمت چپ قرار گرفته اند. به این ترتیب، می توان مشتریان تخصیص داده شده به نرون های سمت راست را در یک خوشه و مشتریان تخصیص داده شده به نرون های سمت چپ را در خوشه دیگر طبقه بندی نمود و به این ترتیب، تعداد مشتریان حاضر در هر خوشه مطابق جدول (۲) خواهد بود.

جدول شماره (۳): خوشه‌بندی

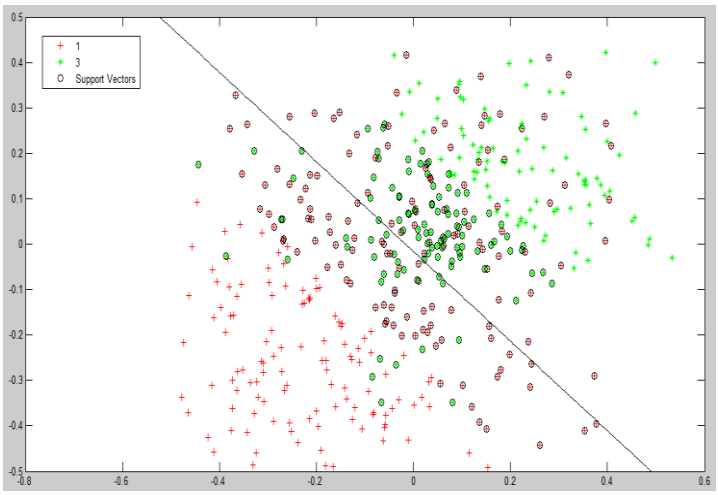
خوشه	مرکز خوشه	تعداد اعضای خوشه
مشتریان وفادار	(۰/۶۷- و ۰/۴۵-)	۴۳۸
مشتریان معمولی	(۰/۰۴و ۰/۰۷)	۲۴۸
مشتریان رویگردان	(۰/۲۷و ۰/۰۵)	۳۴۱

برای محاسبه مرکز خوشه، از میانگین اوزان نرون های حاضر در هر خوشه به نسبت هریک از ورودی ها، استفاده شده است. برای مثال، میانگین اوزان تمامی نرون های خوشه مشتریان رویگردان در متغیر اول، نشان دهنده پارامتر اول و میانگین اوزان نرون ها در متغیر دوم، نشان دهنده پارامتر دوم می باشد. در پایان این مرحله، داده های حاضر در نمونه تحقیق برچسب گذاری شده و لذا می توان با استفاده از ماشین بردار پشتیبان به پیش بینی رویگردانی پرداخت.

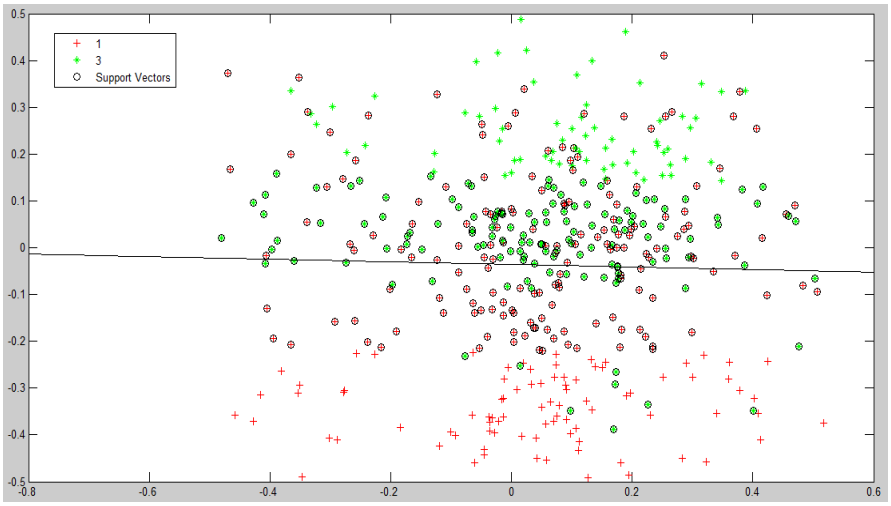
همانگونه که گفته شد؛ ماشین بردار پشتیبان با استفاده از یک ابر صفحه جداکننده به دسته بندی مشتریان در دو دسته رویگردان و غیررویگردان شامل معمولی و وفادار می پردازد. بر اساس روش تحقیق، سه متغیر به منظور دسته بندی یا همان پیش بینی رویگردانی به عنوان ورودی ماشین بردار پشتیبان استفاده شده اند که نتایج به دست آمده از دسته بندی در فضای دو بعدی هر جفت از متغیرهی مربوطه در مطابق شکل (۶) می باشد.



شکل شماره (۶): تعداد تراکنش نقدی – ارزش تراکنش نقدی



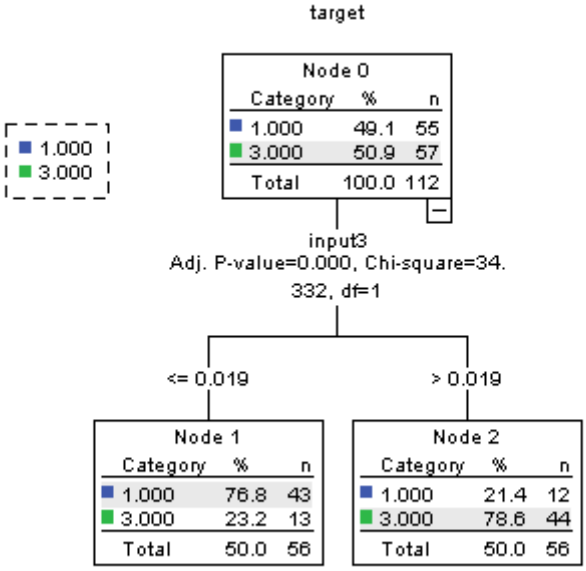
شکل شماره (۷): تعداد تراکنش اعتباری – ارز تراکنش اعتباری



شکل شماره (۸): ارزش تراکنش نقدی – اعتباری

در شکل (۶)، رنگ سبز نشان دهنده مشتریان رویگردان و رنگ قرمز مشتریان معمولی و وفادار را نشان می دهد. نقاط دایره شکل، نقاطی هستند که به عنوان بردار پشتیبان انتخاب شده اند و ابر صفحه جداکننده به گونه ای تشکیل شده است که فاصله آن با تمامی این نقاط بیشینه گردد.

علاوه بر روش ماشین بردار پشتیبان، از درخت تصمیم نیز به عنوان یک معیار به منظور رویگردانی استفاده شده و به منظور دسته بندی با استفاده از آن از الگوریتم CHAID استفاده شده است. شکل (۷) نشان دهنده درخت تشکیل شده بر اساس داده های آموزش می باشد.



مشاهده می شود که درخت تصمیم تنها از یک متغیر و آن هم تراکنش های اعتباری به منظور دسته بندی بین مشتریان رویگردان و وفادار استفاده کرده است و دو متغیر دیگر، بهبودی در دسته بندی توسعه داده شده، ایجاد نکرده اند. بر اساس نتایج به دست آمده از درخت تصمیم و ماشین بردار پشتیبان، می توان نتایج به دست آمده از این دو روش را با جدول (۳) مقایسه نمود.

روش	تعداد پیش بینی صحیح	تعداد پیش بینی غلط	صحت پیش بینی
SVM	۹۱	۲۱	٪۸۱ /۳
درخت تصمیم	۸۶	۲۶	٪ ۷۷/۷

بر اساس نتایج به دست آمده، ماشین بردار پشتیبان در بیش از ٪۸۱ موارد توانسته است مشتریان رویگردان را به درستی تشخیص دهد و این آمار برای درخت تصمیم ٪۷۷ بوده است که حاکی از عملکرد بهتر SVM در پیش بینی رویگردانی می باشد.

٤- نتایج پژوهش

تحلیل روابط مشتریان و پیش بینی رفتار آنها در قبال بانک، اهمیت بسیار بالایی دارد. دلیل این امر، این است که یک مشتری، تنها از یک خدمت بانکی استفاده نمی کند؛ بلکه از مجموعه ای از خدمات بانکی و مالی که گستردگی بالایی دارد؛ بهره می برد و در واقع، هزینه از دست دادن مشتری برای بانک ها، زیاد است. در همین راستا، در این تحقیق جهت شناسایی مشتریان رویگردان، از دو متغیر میانگین موجودی و میانگین تراکنش به صورت تعدیل شده با درآمد استفاده شده است. بر اساس این دو متغیر و با استفاده از نقشه خود سازمان ده شبکه عصبی، مشتریان رویگردان شناسایی شده و برچسب وفادار، معمولی و رویگردان به مشتریان تخصیص داده شد. در ادامه، در یک فضای سه بعدی تشکیل شده توسط متغیرهای تعداد و ارزش تراکنش های نقدی و تراکنش های اعتباری و با استفاده از ماشین بردار پشتیبان، مشتریان دسته بندی شده و رویگردانی مشتریان پیش بینی گردید. بر اساس نتایج به دست آمده، روش پیشنهادی قابلیت پیش بینی رویگردانی تا ٪۸۱ موارد را داشته است. علاوه بر این، ماشین بردار پشتیبان نیز عملکرد بهتری در دسته بندی مشتریان نسبت به درخت تصمیم با ٪۷۷ کارایی، داشته است. بر اساس نتایج به دست آمده، بانک می تواند با اعمال روش پیشنهادی، ۶ ماه قبل از رویگردانی مشتریان، این مسئله را پیش بینی کرده و نسبت به اقدامات اصلاحی و پیشگیراته، اقدام نمایند.

٤- منابع

1. Adwan, Omar et al. (2014). Predicting customer churn in telecom industry using multilayer perception neural networks: Modeling and analysis. *Life Science Journal*, 11(3): 75-81.

2. Amir Shahi, Mirahmad, Siahtiri, Vida, Ravanbod, Fariba. (2010). Identifying the factors affecting the creation of "trust" in key bank customers: key customers of the Bank of Entrepreneur in Tehran. *Management Researches in Iran*, 4(64): 61-67.
3. Haghighi Kafash, Mehdi, & Akbari, Masoud. (2011).Prioritizing Factors Affecting Customer Loyalty Using the ECSI Model. *Marketing Management*, 6 (10): 95-118.
4. Bahramzadeh, Mohammad Mehdi, Shokati Maghareb, Somayeh. (2010). Identification and ranking of factors affecting customer loyalty in private banks of Khuzestan province. Second International Financial Services Marketing Conference.
5. Castanedo, Federico, et al. (2014). Using deep learning to predict customer churn in a mobile telecommunication network.
6. Datta, P., Masand, B., Mani, D., & Li, B. (2000). Automated cellular modeling and prediction on a large scale. *Artificial Intelligence Review*, 14: 485–502.
7. Ghanbarterani, Nasim. (2014). Customer clustering based on RFM model and data mining approach to increase customer loyalty. MSc Thesis, Tarbiat Moalem University, Faculty of Engineering.
8. Gordini, Niccolò, and Valerio, Veglio. (2017). Customers churn prediction and marketing retention strategies: An application of support vector machines based on the auc parameter-selection technique in b2b e-commerce industry. *Industrial Marketing Management*, 62: 100-107.
9. Huang, Bingquan et al. (2016). A Fuzzy Rule-Based Learning Algorithm for Customer Churn Prediction. Industrial Conference on Data Mining. Springer International Publishing, 2016.
10. Huigevoort, Chantine, and Remco, Dijkman. (2015). Customer churn prediction for an insurance company. Diss. M. Sc. Thesis, Eindhoven University of Technology, Eindhoven, Netherland, 2015.
11. Lee, W. S., Tzeng, G. H., Guan, J. L., Chein, K. T., and Huang, J. M. (2009). Combined MCDM Techniques for Exploring Stock Selection Based on Gordon Model. *Expert Systems with Applications*, 36(7): 6421-6430.
12. Mutanen, Teemu. (2006). Customer churn analysis–a case study. *Journal of Product and Brand Management*, 14(1): 4-13.
13. Óskarsdóttir, María, et al. (2016). A comparative study of social network classifiers for predicting churn in the telecommunication industry. *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE.
14. Prasad, U. Devi, and S. Mahdavi. (2012). Prediction of churn behavior of bank customers using data mining tools. *Business Intelligence Journal* 5(1): 96-101.
15. Sharma, Anuj, Panigrahi, and Prabin Kumar. (2013). A neural network based approach for predicting customer churn in cellular network services. *Ar Xiv preprint Ar Xiv:*1(309):39-45.
16. Vafeiadis, Thanasis et al. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modeling Practice and Theory*, 55: 1-9.
17. Vapnik, Viladmir, N. (1998). An overview of statistical learning theory. John Wiley & Sons Inc., New York.
18. Zhang, L., and Zhou, W. D. (2011). Density-Induced Margin Support Vector Machines. *Pattern Recognition*, 44(8): 1448-1460.

Using Data Mining to Predict Bank Customers Churn

Abstract

The intensity of finding competition in the industrial and economic space and the market move towards a complete competition market has made the inclination of firms to attract more customers and, instead, have increased the tendency to operate in various service and manufacturing areas. This policy, which is known for increasing the share of wallet, makes it more important to maintain customer relationships and analyze their relationships and it is necessary to conduct customer behavioral analysis, customer relationship analysis, and customer behavior forecasting. The present research seeks to identify customers who are turning away and anticipates the decline of customers in order to prevent customers from falling. In this regard, the variables associated with the reversal analysis are first identified and then the bank customers are clustered using a neural network and classified into three categories of loyal, regular, and negative clients. With the receipt of the above labels, a backup vector machine has been used to classify and reverse prediction. Based on the results, the proposed method has the ability to predict rotational deviation of up to 80% and, moreover, has a better performance than the classical decision tree.

Keywords: Neutral analysis, Neural network, Support vector machine, Decision tree.