

کاربرد داده کاوی در بهبود مدیریت ارتباط با مشتری

دکتر محمد جعفر تارخ*
کبری شریفیان**

چکیده

امروزه مدیریت ارتباط با مشتری فقط یک مزیت رقابتی نیست بلکه شناسایی و جذب مشتریان وفادار و نگهداشتن آنها برای ادامه حیات در بازار رقابت، امری ضروری است. بانک‌ها و موسسات اعتباری در گذشته به علت محدودیت تنوع خدمات و عدم دسترسی به اطلاعات مشتریان، از روش‌های ساده کیفی جهت سنجش اعتبار مشتریان خود استفاده می‌کردند ولی امروزه به دلیل وجود بانکداری الکترونیکی ثبت اطلاعات تراکنشی مشتریان راحت تر صورت می‌گیرد و همین امر موجب شده است روش‌های کمی جایگزین روش‌های کیفی شود. در بررسی‌های انجام شده یکی از مشکلات بانک ملت عدم شناسایی مشتریان و عدم اتخاذ تصمیمات مناسب مدیریت و سلیقه‌ای بودن تصمیمات در برخورد با

* عضو هیات علمی دانشگاه صنعتی خواجه نصیر الدین طوسی Mjtarokh@kntu.ac.ir

** عضو هیات علمی دانشگاه آزاد اسلامی K_sharif121@yahoo.com

مشتریان می‌باشد. بنابراین در این تحقیق به منظور شناسایی مشتریان بانک ملت و تدوین استراتژی‌های مناسب برای برخورد با آنها از داده کاوی و ابزار آن مانند الگوریتم ژنتیک و الگوریتم میانگین k استفاده شد. بدین منظور ابتدا بانک ملت و بانک اطلاعاتی آن در بخش های مختلف بررسی شد، پس از استخراج داده از بانک اطلاعاتی و پاک سازی آن، مشتریان در گروه های مختلف، با استفاده از الگوریتم میانگین k به دو روش فازی و غیر فازی گره بندی شدند. برای تعیین تعداد بهینه خوشه‌ها ابتدا مجدداً با استفاده از الگوریتم ژنتیک مشتریان در گروه‌های مختلف تقسیم بندی شده با استفاده از روابط خاص تعریف شده تعداد بهینه خوشه‌ها تعیین شد. برای ارزیابی کیفیت خوشه‌های بدست آمده از معیار سنجش تراکم خوشه‌ها استفاده شد. سپس با استفاده از معیارهای تازگی مراجعه هر مشتری، تکرار مراجعات آنها و ارزش پولی هر مشتری ارزش خوشه‌ها تعیین شد و در انتها برای تحلیل خوشه‌ها و تبیین استراتژی مناسب هر خوشه از هرم ارزش مشتری استفاده می‌شود. در انتها نیز با استفاده از درخت تصمیم‌گیری قوانین شناسایی مشتریان جدید استخراج شد.

کلمات کلیدی: داده کاوی، خوشه‌بندی، هرم ارزش مشتری، الگوریتم ژنتیک، استخراج قانون، مشتری.

مقدمه

شناخت مشتریان و همچنین شناخت نیازهای مشتریان، عمل موثری در کسب برتری در ارائه خدمات به مشتری است. مدیران باید مشتریان خود را اولویت بندی کرده و کانون توجه خود را بر مشتریان کلیدی متمرکز کنند و روز به روز هزینه از دست دادن مشتری را بیشتر درک می‌کنند.

لذا وقتی مشتریان از تجارت با ما دست می‌شویند و شروع به تجارت با رقبای ما می‌نمایند، چند موقعیت ناخوشایند اتفاق می‌افتد:

اول این که ما درآمد جاری خود را به خاطر رابطه تجاری موجود از دست می‌دهیم. دوم این که اگر رابطه تجاری با مشتری از دست برود نیازی به استخدام

افراد جدید نداریم. سوم این که شهرت و اعتبار ما از دست می‌رود. مشتریان ما احتمالاً تجربیات خودشان را با مشتریان دیگر در میان خواهند گذاشت این زیان ممکن است منجر به از دست رفتن اعتماد مشتریان کنونی ما و هر مشتری بالقوه‌ای شود. امروزه تدوین سیاست‌های و استراتژی‌های رقابتی از سوی صاحب نظران تاکید می‌شود. لذا موسسات اعتباری (مانند بانک‌ها و...) نمی‌توانند اهداف اساسی کار خود نظیر دستیابی به مزیت رقابتی یا ایجاد سود را نادیده بگیرند. شناسایی گروه‌های مختلف مشتریان و تعیین خواسته‌ها و نیازهای آنان می‌تواند باعث به وجود آمدن رضایت مشتری برای موسسات مالی شود و همین امر موجب وفاداری بیشتر مشتریان می‌شود. شناسایی مشتریان کلیدی و حفظ این گونه مشتریان در بلند مدت نسبت به جذب مشتریان جدید، برای جایگزینی مشتریانی که قطع رابطه با بانک کرده‌اند، سودمندتر است. زیرا هزینه جذب یک مشتری کلیدی جدید ۵ برابر هزینه حفظ یک مشتری است. مشتریانی که بانک بتواند رضایت آنها را جلب کند، تجربیات مثبت خود را برای دیگران بیان می‌کنند و به این ترتیب وسیله تبلیغ برای بانک می‌شوند و در نتیجه هزینه جذب مشتریان جدید کاهش می‌یابد.

در گذشته عموماً تحلیل گران وظیفه استخراج اطلاعات مفید از داده‌های ثبت شده رابه صورت دستی انجام می‌دادند. اما به علت این که این تجزیه و تحلیل های دستی داده‌ها، بسیار کند، گران و موضوعی بوده و حجم داده‌ها و پیچیدگی داده‌ها روز به روز در حال افزایش است، حرکت از سمت تحلیل های مستقیم دستی به سمت تحلیل های غیر مستقیم اتوماتیک با استفاده از روش های کامپیوتری شروع شده است. داده کاوی به وسیله کاوش داده‌های موجود در بانک‌های اطلاعاتی، اطلاعات و دانش مفید از داده‌ها را استخراج می‌کند، بنابراین یکی از ابزار مدیریت ارتباط با مشتری می‌باشد. بخش بندی یکی از مباحث مطرح در حوزه مدیریت ارتباط با مشتری است. بخش بندی شکستن جمعیت زیادی از مشتریان به بخش های مختلف است به طوری که مشتریان موجود در هر بخش به یکدیگر شبیه و مشتریان بخش‌های مختلف با یکدیگر متفاوت هستند. بخش‌بندی دیدگاهی کلی از تمام

بانک اطلاعاتی مشتریان ارائه می‌دهد و به صاحبان کسب و کار امکان اعمال رفتار و سیاست‌های متفاوت را به مشتریان هر بخش می‌دهد. در حالت ایده آل هر سازمان باید هر یک از مشتریانش را بطور کامل بشناسد. ولی این کار در عمل امکان پذیر نیست و در واقع بخش بندی این امکان را فراهم می‌آورد تا مشتریانی که شبیه به هم هستند در یک بخش قرار گیرند. در این صورت مدیریت و شناخت این بخش‌ها بسیار ساده‌تر از شناخت تک تک مشتریان است. داده کاوی برای بخش‌بندی مشتریان، دارای ابزارهای متفاوتی می‌باشد مانند الگوریتم ژنتیک، الگوریتم میانگین k و غیره. بنابراین در این مقاله با استفاده از این ابزار یک سیستم خبره تولید که مدیریت را در تصمیم‌گیری و اخذ استراتژی‌های مناسب برای برخورد با مشتریان پشتیبانی می‌کند.

مرور ادبیات

در برخی تحقیقات صورت گرفته در این حوزه، محققان در مقاله ای در سال ۲۰۰۲ مشتریان یک بانک و یک کلوپ کتاب با استفاده از داده کاوی برای شناسایی مشتریان طبقه‌بندی شدند. [۶]

در مقاله‌ای دیگر در سال ۲۰۰۴ با در نظر گرفتن سهم سود ایجاد شده، سود بالقوه و تعریف سود آوری مشتری یک مدل پیشنهاد کردند و مشتریان را بخش بندی نمودند. [۵]

بحث اعتبارسنجی مشتریان بانک توسط مدل‌های اعتبارسنجی در تحقیقات اخیر بسیار به کار می‌رود در سال ۲۰۰۸ محققان در مقاله‌ای از مدل رتبه‌بندی تحلیل لینک با استفاده از ماشین بردار پشتیبان استفاده کردند. [۱۳]

گروهی از محققان در سال ۲۰۰۸ در مقاله‌ای با استفاده از مدل شبکه عصبی برای اعتبارسنجی مشتریان استفاده کردند. [۱]

گروهی دیگر از محققان در سال ۲۰۰۹ از مدل طبقه‌کننده‌های ترکیبی استفاده کردند. [۱۰]

روند توسعه داده کاوی را می‌توان در جدول ۱ بررسی نمود. [۲]

جدول ۱. روند توسعه داده کاوی

دوران	سوال تجاری	تکنولوژی	تولید کنندگان	ویژگی‌ها
گرد آوری داده (سالهای ۱۹۶۰)	متوسط عایدی شرکت در ۵ سال گذشته چقدر بوده است؟	کامپیوتر، نوارها، دیسکها	CDC-IBM	تحویل داده‌های استاتیک و تاریخ گذشته
دسترسی داده‌ها (سالهای ۱۹۸۰)	فروش در واحد در NewEngland در مارس گذشته چقدر بوده است؟	بانکهای اطلاعاتی رابطه ای مانند SQL	Oracle-Sybase-IBM-Microsoft	تحویل داده‌های پویای تاریخ گذشته
راهبری داده‌ها (سالهای ۱۹۹۰)	فروش در واحد در NewEngland در مارس گذشته چقدر بوده و جزئیات شهر Boston چیست؟	OLAP- انبار داده- بانک اطلاعاتی چند بعدی	Arbor-IRI-Pilot-Redbrick-Evolutionarytech	تحویل اطلاعات پویای تاریخ گذشته در چندین سطح
داده کاوی (سال ۲۰۰۰)	در واحد شهر Boston در ماه آینده میزان فروش چقدر خواهد بود؟ چرا؟	الگوریتم‌های پیشرفته- بانک‌های اطلاعاتی عظیم	Lockheed-IBM-SGI-Nascent Industry	تحویل اطلاعات دورنما و آینده نگر

داده کاوی^۱

امروزه به دلیل وجود حجم انبوهی از داده‌ها در پایگاه داده‌ها، امکانات کامپیوترها و الگوریتم‌های قوی‌تر، جانشین تحلیل‌های دستی شده است تا اطلاعات و دانش را از داده‌های موجود استخراج کند. داده کاوی فرآیند کشف رابطه‌ها، الگوها و روند معنی‌داری است که به بررسی حجم وسیعی از اطلاعات ذخیره شده در پایگاه داده‌ها با فناوری تشخیص الگو می‌پردازد.

ابتدا تعاریفی از مفاهیم پایه را ارائه می‌کنیم. [۷]

داده: کوچک‌ترین جزء ذخیره شده در پایگاه داده را داده می‌نامیم. مانند: سن کارمندان

اطلاعات: با اضافه شدن مفهومی به داده‌ها اطلاعات بدست می‌آید. مانند: سن ۲۰٪ کارمندان بالای ۵۰ سال است.

دانش: از اضافه شدن قواعد به اطلاعات، دانش بدست می آید. مانند: کارمندان بالای ۵۰ سال، از بهره وری بیشتری برخوردارند.

فرایند داده کاوی شامل چندین مرحله است:

۱- تعیین هدف از داده کاوی، در راستای جواب به مسئله مطرح در سیستم

۲- شناخت و ایجاد داده هدف (فیلدهای مورد بحث در هدف)

۳- آماده سازی داده، در این قسمت بیش از ۸۰ درصد نیروی کار و زمان صرف می شود، در این مرحله به تعیین استراتژی در مورد داده های گم شده و نقاط پرت، کم کردن ابعاد داده و... را در بر می گیرد.

۴- تطبیق هدف با یکی از روش های داده کاوی از قبیل کلاس بندی، خوشه بندی، رگرسیون، جمع بندی و ...

۵- انتخاب الگوریتم مناسب هر یک از روشهای اتخاذ شده

۶- تعبیر و تفسیر الگو با توجه به هدف تعیین شده در بند اول، این قسمت شاید تفاوت اصلی داده کاوی و کشف دانش باشد اما در بسیاری منابع داده کاوی شامل کل این ۶ مرحله است و معادل کشف دانش می باشد.

می توان فعالیت های داده کاوی را به یکی از دو طبقه بندی زیر تقسیم کرد:

الف - داده کاوی پیش بینی کننده، که مدل سیستم را بوسیله مجموعه داده از پیش تعیین شده، تولید می کند.

ب - داده کاوی توصیف کننده، که اطلاعات جدید و بدیع را بر مبنای مجموعه داده موجود ایجاد می کند.

چند روش از این دو نوع را مرور می کنیم:

کلاس بندی: تابع پیش بینی که نمونه را به یکی از چند کلاس معین تخصیص می دهد.

رگرسیون: تابع پیش بینی، که یک نمونه داده را به یک مقدار حقیقی متغیر پیش بینی می نگارد.

خوشه بندی یا کلاس ترینگ: تابع توصیفی که تعداد محدودی خوشه را جهت توصیف داده، جستجو می کند.

جمع بندی: تابع توصیفی که توصیف فشرده برای یک مجموعه از داده پیدا می‌کند.

مدل وابستگی: تابع توصیفی برای پیدا کردن وابستگیهای مهم بین متغیرها شناخت انحراف و تغییر: کشف تغییرات با بیشترین اهمیت در مجموعه داده بسته به خواص داده و مسئله موجود، می‌توان با تصمیمی راجع به کاربرد یکی یا بیشتر از یکی از تکنیک‌های کشف دانش، آنها را مورد استفاده قرار داد.

کاربردهای داده کاوی

در واقع اکثر مسائل اقتصادی و تجاری می‌توانند در قالب شش ناحیه زیر وظیفه بندی شوند:

۱: طبقه بندی

برای فهم و برقراری ارتباط با محیط اطراف همواره اشیاء را دسته بندی و درجه بندی می‌کنیم.

۲: تخمین

عمل طبقه بندی مربوط به نتایج گسسته است در حالی که عمل تخمین مربوط به نتایج پیوسته می‌باشد. با دادن مقادیری به عنوان ورودی، از تخمین استفاده می‌کنیم تا یک متغیر ناشناس مانند درآمد و... را تخمین بزنیم.

۳: پیش‌بینی

پیش‌بینی همانند طبقه بندی و تخمین می‌باشد با این تفاوت که رکوردهایی که طبقه‌بندی می‌شوند با توجه به پیش‌بینی‌هایی است که از رفتار آینده و یا تخمین مقادیری که در آینده به خود می‌گیرد می‌باشد.

۴: وابستگی گروهی یا تحلیل سبد خرید

عمل وابستگی گروهی یا تحلیل سبد خرید در واقع تشخیص می‌دهد که چه چیزهایی به صورت مرتبط با هم کنار یکدیگر قرار گیرند.

۵: خوشه بندی

خوشه‌بندی در واقع تقسیم‌بندی یک جمعیت ناهمگون به تعدادی زیر مجموعه همگون است.

۶: توصیفی

گاهی اوقات هدف از داده کاوی توصیف وضعیتی است که در پایگاه اطلاعاتی وجود دارد به گونه ای که سبب افزایش دریافتی‌ها شود.

پس به طور کلی داده کاوی را به یکی از دو طبقه زیر تقسیم می‌کنند:

۱- داده کاوی پیش بینی کننده که مدل سیستم را به وسیله مجموعه ای از متغیرها یا فیلدهای معلوم پیش بینی می‌کند. هدف در این نوع داده کاوی ایجاد مدلی است که در قالب یک کد قابل اجرا بتواند وظایف کلاس بندی، پیشگویی، تخمین و... را انجام دهد.

۲- داده کاوی توصیف که طرح هایی که داده‌ها را توصیف می‌کند ایجاد می‌کند که این طرح‌ها عموماً توسط انسان تعبیر می‌شود. هدف در این نوع داده کاوی تحلیل سیستم به وسیله الگوها و ارتباطات کشف شده است.

در جدول ۲ این دو طبقه با وظایف و روشهای به کار رفته در هر کدام بیان شده است.

جدول ۲. انواع مختلف داده کاوی و روشهای مورد استفاده

روشهای به کار گرفته شده	وظایف داده کاوی	طبقه بندی داده کاوی
درخت تصمیم گیری، تحلیل سبد خرید، استنتاج براساس حافظه	طبقه بندی	مدلهای پیش بینی کننده
شبکه‌های عصبی، درخت تصمیم گیری، الگوریتم ژنتیک	برازش	
شبکه‌های عصبی، الگوریتم ژنتیک، سری‌های زمانی	تخمین زدن	مدلهای توصیف کننده
بصری سازی	خلاصه سازی	
منطق فازی	خوشه بندی	
قواعد انجمنی	وابستگی گروهی	

در گذشته روش‌های کلاسیک مختلفی نظیر تجزیه و تحلیل خطی، رگرسیون خطی برای اعتبارسنجی مشتریان استفاده شده اند که در بین این دو روش، تجزیه و تحلیل خطی از مقبولیت بیشتری برخوردار می باشد. اما نقطه ضعف روش تجزیه و تحلیل خطی، فرض خطی بودن رابطه بین متغیرها و نرمال بودن آنهاست، در حالی که این رابطه معمولاً غیرخطی و غیرنرمال می باشد. لذا دقت این دو روش بسیار پایین است. [۹]

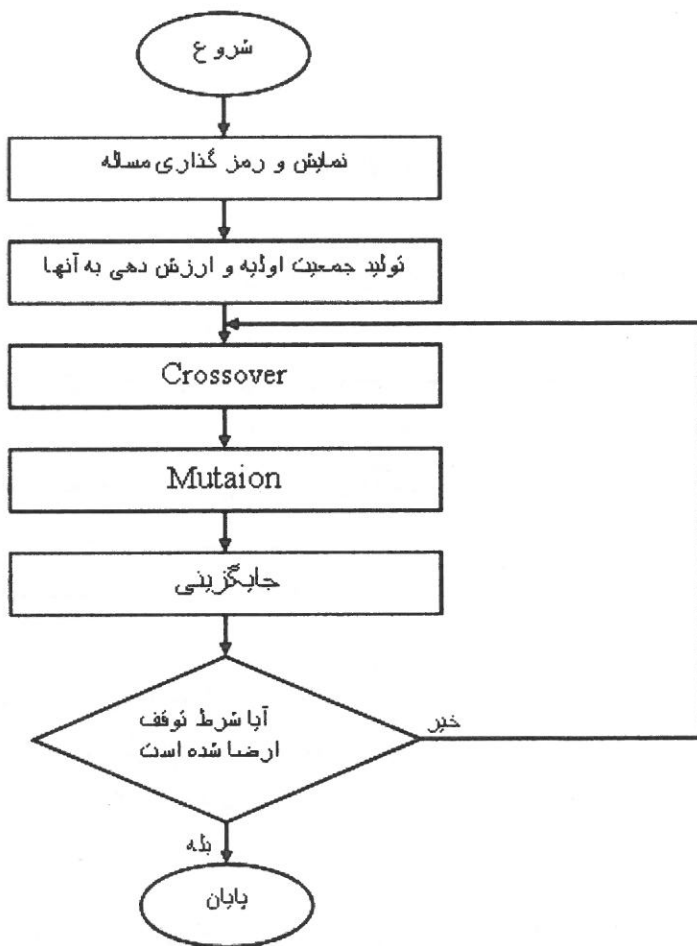
علاوه بر این، چند روش برجسته در حوزه هوش مصنوعی همچون سیستم‌های خبره، سیستم‌های فازی، شبکه‌های عصبی و الگوریتم ژنتیک نیز برای اعتبارسنجی مورد استفاده قرار گرفته‌اند. در بین این روش‌ها شبکه‌های عصبی به علت توانایی تخمین روابط پیچیده غیرخطی بیشتر مورد توجه بوده‌اند. [۸]

الگوریتم ژنتیک^۱

الگوریتم ژنتیک از اصول ژنتیک طبیعی و بقای بهترین‌ها نشأت گرفته است که به توصیف مجموعه‌ای از رویه‌های طبیعی جستجو می پردازد. این تکنیک بهینه‌سازی در مسائلی که روش‌های ریاضی دقیق و بهینه برای حل آن وجود ندارد استفاده

می‌شود. روش تولید فضای جواب بصورت تصادفی یا ابتکاری^۱ و به صورت یک تکنیک جستجو جهت یافتن جواب بهینه برای مسئله است. مراحل اجرای الگوریتم ژنتیک شامل مراحل زیر است: [۱۴]

۱. شناسایی ژنها: در الگوریتم ژنتیک برای نمایش یک ژن از یک بیت استفاده می‌شود. یک ژن نشان دهنده رفتار یک بخشی از راه حل مسئله است.
۲. سازماندهی ژنها در کروموزمها: به مجموعه ژنها کروموزم می‌گویند.
۳. ایجاد یک جمعیتی از راه حل‌های ممکن: به مجموعه‌ای از راه حل‌ها یا کروموزمها یک جمعیت می‌گویند.
۴. ارزیابی کروموزمها: در این مرحله به ارزیابی اثر بخشی جمعیت اولیه برای مقایسه با جمعیت‌های دیگر می‌پردازد.
۵. ایجاد جمعیت جدید توسط عملیات تقاطع و جهش: برخی راه حل‌ها ممکن است در یک جمعیت از راه حل‌های دیگر برتر باشد. به همین دلیل با انتخاب این مجموعه راه حل‌ها و توسط عملیات جهش، جمعیت دیگری شکل می‌گیرد. مجدداً مراحل ۴ تا ۵ شرط توقف الگوریتم محقق می‌شود. شکل ۱ مراحل الگوریتم ژنتیک را به صورت فلوجارت نمایش می‌دهد.



شکل ۱. فلوچارت الگوریتم ژنتیک

خوشه بندی و الگوریتم میانگین k

آنالیز خوشه بندی مجموعه ای از روش شناسی هایی برای تقسیم اتوماتیک نمونه ها به تعدادی گروه می باشد بطوری که نمونه ها در یک گروه شبیه به هم و نمونه هایی در گروه های متفاوت شبیه نمی باشند یک نتیجه کلی از آنالیز خوشه بندی، توصیف کلی از هر خوشه می باشد که این خصوصاً برای تحلیل عمیق تری از خواص مجموعه داده ها لازم می باشد.

اغلب الگوریتم های خوشه بندی براساس دو روش زیر پایه گذاری شده اند:

۱- خوشه بندی سلسله مراتبی

۲- خوشه بندی تقسیمات تکرار خطا

تکنیک های سلسله مراتبی، داده ها را در دنباله ای تو در تو از گروه ها سازمان می دهد، که می تواند در قالب ساختار درختی نشان داده شود. درحالی که الگوریتم های تقسیمات مجذور خطا تلاش می کند تا تقسیماتی بدست آورد که پراکندگی داخل خوشه را کمینه یا پراکندگی بین خوشه ها را افزایش دهد. پرکاربردترین استراتژی خوشه بندی قسمتی بر مبنای معیار مجذور خطا می باشد.

روش خوشه بندی قسمتی میانگین k ، ساده ترین و پر استفاده ترین الگوریتم مورد استفاده در معیار مجذور خطا می باشد و از الگوریتم های مناسب برای داده های متریک و مطلق می باشد و قدم های مبنای آن بصورت زیر می باشد. [۱۱]

۱- انتخاب یک تقسیم بندی اولیه با k خوشه شامل نمونه های انتخابی تصادفی و محاسبه مراکز خوشه ها

۲- تولید یک تقسیم بندی جدید با تخصیص هر نمونه به نزدیک ترین مرکز خوشه

۳- محاسبه مراکز جدید خوشه ها به عنوان مراکز خوشه ها

۴- تکرار قدم های ۲ و ۳ تا اینکه یک مقدار بهینه برای تابع معیار پیدا شود (یا اینکه اعضای خوشه ثابت شود)

خوشه بندی و الگوریتم ژنتیک و تعیین تعداد بهینه خوشه ها

در خوشه بندی با استفاده از الگوریتم ژنتیک یک معیار خوشه بندی به نام μ به صورت زیر استفاده می شود:

$$\mu(c_1, \dots, c_k) = \sum \sum \|x_i - z_j\| \quad (1)$$

وظیفه الگوریتم ژنتیک یافتن مراکز خوشه به گونه ای است که معیار خوشه بندی μ حداقل شود.

هر رشته دنباله ای از اعداد واقعی است که نشان دهنده k خوشه می باشند. در یک فضای n بعدی طول هر کروموزوم $n * k$ خواهد بود. برای ساخت جمعیت اولیه مراکز

k خوشه به صورت تصادفی از داده‌های موجود انتخاب شده و در یک کروموزم جای می‌گیرند. این فرآیند برای تمام جمعیت تکرار می‌شود.

در ابتدا خوشه‌ها بر اساس مراکز که کروموزم‌ها را تشکیل داده‌اند شکل می‌گیرد. به این صورت که x_i به یکی از خوشه‌ها تخصیص داده می‌شود. پس از این که خوشه‌بندی انجام شد، مراکز خوشه‌های موجود در کروموزم‌ها با میانگین نقاط هر خوشه جایگزین می‌شود. برای خوشه جدید مرکز به صورت زیر محاسبه می‌شود.

$$Z_i^* = 1/n \sum x_i (i=1 \dots k) \quad (2)$$

با یک مثال این رابطه را توضیح می‌دهیم:

فرض کنید مرکز خوشه اول (۷۲/۳-۵۱/۶) باشد. پس از خوشه‌بندی دو نقطه (۷۰-۵۰) و (۷۴-۵۲) در این خوشه قرار گرفتند. بدین ترتیب مرکز جدید این خوشه

برابر $(52+70)/3$ و $(51/6+74)/3$ یا $(72/3+70+74)/3$ خواهد بود

معیار خوشه بندی یا همان μ از رابطه زیر محاسبه می‌شود:

$$\mu = \sum_{i=1}^k \mu_i \quad (3)$$

$$\mu_i = \sum \|x_i - z_j\| \quad (4)$$

$$f = 1/\mu \quad (5)$$

حداکثر تابع فوق زمانی است که μ حداقل باشد که این همان هدف ما است. در این الگوریتم شرط توقف به اندازه حداکثر تعداد صورت گرفته است. مثلاً تعداد دفعات اجرای این الگوریتم را ۴ در نظر می‌گیریم.

بدین ترتیب می‌توان تعداد خوشه‌هایی را انتخاب نمود که به ازای آن فاصله بین مراکز خوشه‌ها و شباهت مراکز خوشه با اعضای درون هر خوشه حداکثر گردد. [۳]

ارزیابی کیفیت خوشه‌بندی

معیارهای زیادی برای محاسبه کیفیت خوشه‌های حاصل از خوشه‌بندی وجود دارد. خوشه‌بندی‌ای مطلوب‌تر است که شباهت درون خوشه‌ها یا همگنی درون

خوشه‌ها و فاصله بین خوشه‌ها یا ناهمگنی برون خوشه‌ای را حداکثر نمایند. [۸]

یکی از روشهای ارزیابی عملکرد خوشه‌بندی، روش سنجش تراکم درون خوشه‌ای است. این معیار میزان تراکم خوشه‌ها را زمانی که تعداد خوشه‌ها ثابت است نشان می‌دهد. در این معیار باید متغیرها در یک محدوده باشند. با توجه به این که در این تحقیق داده‌ها گسسته هستند برای محاسبه تراکم یک خوشه، فاصله مشتریان درون یک خوشه باید با مرکز خوشه سنجیده شود. در این روش ابتدا هر مشتری با توجه به مقادیر شاخص‌های آن (مانده حساب روزانه و مبلغ چک برگشتی و...) در یک فضای چند بعدی به صورت یک نقطه در نظر گرفته می‌شود.

پس از خوشه‌بندی نیز مرکز هر خوشه به صورت یک نقطه با مقادیر خاص می‌باشد. حال ابتدا فاصله اقلیدسی هر مشتری از مرکز خوشه را بدست می‌آوریم. سپس میانگین مربعات فاصله اقلیدسی بین هر مشاهده و میانه خوشه‌ای که به آن تعلق دارد (پراکندگی یا واریانس هر خوشه را) محاسبه می‌کنیم.

$$\text{Dist}(x_i, z_j) = \sqrt{\sum_{i=1}^l \sum_{j=1}^j (\text{average}_i - \text{average}_j)^2 + (\text{val} - \text{chek}_i - \text{val} - \text{chek}_j)^2 + \dots} \quad (6)$$

سپس تراکم خوشه برابر است با:

$$F(k) = \frac{1}{k} \sum_{i=1}^k \text{Dist}(x_i, z_j) \quad (7)$$

در این رابطه C^n مراکز خوشه‌ها می‌باشد. در واقع $F(k)$ میانگین مربعات فاصله اقلیدسی بین هر مشاهده و میانه خوشه‌ای که به آن تعلق دارد، است. هر چه $F(k)$ کمتر باشد نشان می‌دهد که خوشه‌ها متراکم‌ترند و خوشه‌بندی بهتر صورت گرفته است.

مدلی براساس تازگی، تکرار و ارزش پولی برای ارزیابی خوشه‌ها

ارزش مشتریان را براساس معیارهای تازگی (R) تکرار (F) ارزش پولی (M) اندازه‌گیری می‌کنیم.

معیار تازگی، فاصله بین تازه‌ترین زمان تراکنش و زمان تحلیل را اندازه می‌گیرد و با توجه به این که مورد مطالعاتی بانک می‌باشد، ما اطلاعات ۱ سال مالی را در نظر می‌گیریم، بنابراین معیار تازگی آخرین باری است که مشتری در طی این ۱ سال به

بانک مراجعه داشته است و بین ۰ تا ۳۶۵ می باشد. معیار تکرار، تکرار مراجعات در بازه زمانی ۱ سال را اندازه می گیرد و معیار ارزش پولی، مانده پایان دوره مشتری در پایان ۱ سال است.

از این مدل برای سنجش ارزش هر خوشه استفاده شده است و به سازمان‌ها کمک می کند تا بتوانند ارزش خوشه‌های خود را در مشخص نمایند و خوشه‌های هدف را برای برنامه‌ریزی و انجام، اقدامات لازم را برای افزایش سودآوری در هر خوشه تعیین نمایند. بر مبنای این معیار، ارزش هر مشتری را می توان به صورت زیر مشخص نمود:

$$v(c_i) = w^R * R(c_i) + w^F * F(c_i) + w^M * M(c_i) \quad (8)$$

در این رابطه $R(c_i)$ و $F(c_i)$ و $M(c_i)$ به ترتیب امتیازات هر مشتری با توجه به معیارهای تازگی، تکرار و ارزش پولی است و w^R ، w^F و w^M وزن‌ها معیارها است. امتیازاتی که از بانک اطلاعاتی استخراج می‌نمایم قبل از محاسبه ارزش هر مشتری ابتدا نرمال می‌شود. بنابراین $R(c_i)$ و $F(c_i)$ و $M(c_i)$ را به صورت زیر تعریف می‌کنیم.

$$R(c_i) = \frac{Q^R - Q_{Min}^R}{Q_{Max}^R - Q_{Min}^R} \quad (9)$$

$$F(c_i) = \frac{Q^F - Q_{Min}^F}{Q_{Max}^F - Q_{Min}^F} \quad (10)$$

$$M(c_i) = \frac{Q^M - Q_{Min}^M}{Q_{Max}^M - Q_{Min}^M} \quad (11)$$

که Q^R ، Q^F و Q^M مقدار اصلی امتیازات هر مشتری مطابق تعریف تازگی، تکرار و ارزش پولی است و Q_{Min} حداقل مقدار، این معیارها Q_{Max} حداکثر مقدار این معیارها برای همه مشتریان را نشان می‌دهد.

سودآوری خوشه O^n با محاسبه میانگین ارزش همه مشتریان خوشه π ام بدست می‌آید و به صورت معادله زیر است:

$$v(O^n) = w^R * R(O^n) + w^F * F(O^n) + w^M * M(O^n) \quad (12)$$

$$R(O^n) = \frac{\sum_{c_i \in O^n} R(c_i)}{\|O^n\|} \quad (13)$$

$$F(O^n) = \frac{\sum_{c_i \in O^n} F(c_i)}{\|O^n\|} \quad (14)$$

$$M(O^n) = \frac{\sum_{c_i \in O^n} M(c_i)}{\|O^n\|} \quad (15)$$

که $R(O^n)$ ، $F(O^n)$ و $M(O^n)$ امتیارات n امین خوشه با توجه به معیارهای R ، F و M است. پس از مشخص شدن سودآوری هر خوشه، خوشه‌ها مرتب می‌شود و با ارزش‌ترین آنها مشخص شده و برای تبیین استراتژی برای هر خوشه و خدمت رسانی به مشتریان استفاده می‌شود.

تحلیل خوشه‌ها با استفاده از هرم ارزش مشتری

مفهوم بازاریابی مشتری و هرم مشتری به عنوان ابزاری برای شناخت و تجزیه و تحلیل رفتار مشتری در سال ۲۰۰۱ معرفی شد. [۱۲]

هرم زیر را به صورت زیر تفسیر می‌کنیم:

- مشتریان فعال: سازمان یا شرکت هایی که در یک محدوده زمانی مشخص، کالا یا خدمات خریداری کرده‌اند. مشتریان فعال به ۴ بخش بالا، بزرگ، متوسط و کوچک تقسیم بندی می‌شوند.
- مشتریان غیرفعال: شرکت هایی که قبلا کالا یا خدمات خریداری می‌کردند ولی این کار را در محدوده زمانی مشخصی انجام نداده‌اند. مشتریان غیر فعال منبع مهمی برای درآمد بالقوه هستند. آنها هم چنین منبع اطلاعات ارزشمندی هستند و از طریق این اطلاعات می‌توان از تبدیل مشتریان فعال به غیرفعال جلوگیری کرد.
- مشتریان احتمالی یا بالقوه: شرکت هایی که به نوعی با سازمان ارتباط دارند، ولی هنوز کالا یا خدماتی خریداری نکرده‌اند. مشتریان احتمالی شرکت هایی هستند که انتظار داریم در آینده نزدیک به مشتریان فعال ارتقا یابند.

- مشتریان حدسی: شرکت هایی هستند که می توان محصولات و یا خدماتی را به آنها سرویس داد ولی هنوز با آنها ارتباط برقرار نشده است. معمولاً تلاش می شود تا با مشتری حدسی ارتباط برقرار شده و با این کار به هدف تبدیل آنان به مشتریان فعال در دراز مدت نزدیک شد.
- سایرین: شرکت هایی که خیلی ساده، نیاز یا علاقه ای به خرید یا استفاده از کالاها و خدماتها ندارند. در حالی که از این گروه پولی بدست نخواهد آمد، ولی مهم است که شناخته شوند و میزان هزینه و زمان بازاریابی که برای برقراری ارتباط با این گروه افراد و شرکتها که هرگز با سازمان معامله نخواهند کرد، مشخص شود.

مفهوم مدیریت ارتباط با مشتری را می توان از طریق هرم مشتری به سه بخش زیر مطرح نمود:

- الف) جذب مشتریان جدید به هرم مشتری
- ب) ارتقای مشتریان به سمت بالای هرم
- ج) حفظ مشتریان در برابر ترک هرم

درخت تصمیم گیری و استخراج قوانین

بعد از دسته بندی مشتریان به گروه های مختلف با استفاده از روش های مختلف خوشه بندی، دانش موجود در سیستم را باید در قالب قواعداگر... آن گاه... استخراج کنیم. تا این مرحله چون هیچ دانشی نسبت به سیستم نداشتیم برای تقسیم بندی مشتریان از روش های مختلف خوشه بندی استفاده کردیم اما در مرحله قبل با بدست آوردن گروه های مختلف، در حقیقت دانش اولیه را نسبت به سیستم بدست آورده ایم و خروجی گروه بندی را برای کلاس بندی بقیه مشتریان و استخراج قواعد کلی سیستم استفاده می کنیم. این روش دارای ساختار درختی می باشد به طوری که هر برگ مربوط به یک کلاس و هر گره مربوط به یک فیلد یا ویژگی می باشد. درخت تصمیم گیری با گرهی به نام ریشه آغاز می شود و هر قاعده به یک مسیر از ریشه اختصاص دارد. این الگوریتم در هر تقسیم، فیلدی را انتخاب می کند که با

کمترین تقسیمات به برگ برسد. ویژگیهای مطرح برای قوانین خروجی در این نرم افزار به شرح زیر می باشد: پشتیبانی: تعداد نمونه های وارد شده به هر گره را پشتیبانی گویند. بدیهی بودن: در مورد بعضی از نمونه ها، قوانین با ترکیب گره ها بدست می آید. این ویژگی درصد صحت هر قانون را بیان می کند. شماره خطا: این ویژگی تعداد نمونه هایی را بیان می کند به گروه مذکور تعلق ندارند ولی به گره وارد شده اند. بنابراین قانونی بیشترین اعتبار را دارد که دارای بیشترین پشتیبانی و کمترین شماره خطا باشد.

خوشه بندی مشتریان بانک ملت با استفاده از داده کاوی

۱- یافتن شاخص های موثر در تعیین مشتریان کلیدی
جهت یافتن مشتریان کلیدی در ابتدا لازم بود شاخص های موثر در تعیین این گونه مشتریان در خصوص حساب های جاری قرض الحسنه شناسایی شوند. جهت تعیین این شاخص ها از دو روش استفاده گردید که عبارتند از:

۱: استفاده از مطالعات قبلی انجام گرفته در خصوص مشتریان کلیدی
در این روش با بررسی و مطالعه برخی از تحقیقات انجام گرفته در این زمینه چه در مرکز تحقیقات و برنامه ریزی بانک ملت و چه در مراکز دانشگاهی، تعدادی از شاخص ها شناسایی و مشخص گردید.

۲: استفاده از نظر کارشناسان خبره
جهت یافتن نظر کارشناسان خبره در مورد یافتن شاخص های موثر در تعیین مشتریان کلیدی طی جلساتی نظرات آنان جمع آوری گردید و مهم ترین شاخص ها از نظر آنان تعیین گردید. پس از انجام دو مرحله فوق مهم ترین شاخص هایی که در خصوص حساب های مشتریان کلیدی در حساب جاری قرض الحسنه شناسایی شدند به شرح زیر می باشند:

- ۱- مانده حساب روزانه ۲- مبلغ چک برگشتی روزانه ۳- گردش بدهکار روزانه
- ۴- گردش بستانکار روزانه ۵- مانده پایان دوره روزانه ۶- مانده بستانکار روزانه

استخراج داده‌های مربوط به شاخص‌ها

پس از تعیین شاخص‌ها، لازم بود تا اطلاعات مربوط به مشتریان در رابطه هر یک از شاخص‌ها جمع گردد. در این مرحله با استفاده از داده‌های موجود در بانک اطلاعاتی ابتدا کلیه بانک‌های ملت در سطح تهران به تفکیک شعبه متمایز گردید و اطلاعات حدود ۵۵۰۰ مشتری به صورت تصادفی انتخاب شدند. ولی برای مشاهده رفتار مشتریان در ۱ سال نیاز به بدست آوردن فیلدهایی مانند معدل مانده حساب روزانه در ۱ سال، جمع گردش بدهکار در ۱ سال، جمع گردش بستانکار در ۱ سال، جمع مانده بستانکار مانده پایان دوره در ۱ سال و مبلغ چک‌های برگشتی در ۱ سال بود و این فیلدهای تجمعی وجود نداشت، بنابراین با استفاده از نرم افزار (SQL) برنامه‌هایی برای بدست آوردن داده‌های مربوط به این شاخص‌ها نوشته شد.

بررسی وضعیت داده و آماده سازی آن

پس از انتخاب فیلدها و جمع آوری داده‌ها نوبت به پاکسازی این داده‌ها از داده‌های پرت می‌باشد. برای حذف نقاط پرت در داده‌های موجود ابتدا با استفاده از نرم افزار (Excel) نمودار مقادیر هریک از شاخص‌ها رسم شده و انحراف معیار این داده‌ها از میانگین بدست آورده شد سپس داده‌هایی که بیشترین انحراف معیار را داشته، حذف شدند و تعداد داده‌های موجود به ۲۹۰۰ داده رسید.

برای بازسازی رکوردهایی که برخی از فیلدهای آن خالی بود نیز با استفاده از روش مولدهای هم نهشتی خطی که یکی از روش‌های تولید اعداد تصادفی است مقادیری برای داده‌های گم شده تولید شد. سپس با توجه به اینکه در داده‌های باقیمانده اختلاف بین بیشترین و کمترین مقدار داده زیاد بود و این مسئله دقت خروجی را کاهش می‌داد، داده‌های موجود با استفاده از روش‌های آماری نرمال سازی شد.

بررسی عدم همبستگی فیلدها با استفاده از آنالیز واریانس

برای اطمینان از عدم وابستگی بین شاخص‌های انتخابی و مستقل بودن آنها با

استفاده از نرم افزار (Minitab) تست آنالیز واریانس انجام شد. در یک فاصله اطمینان ۹۵٪ فرض برابری میانگین‌ها رد شد، بنابراین فیلدهای انتخابی کاملاً از یکدیگر مستقل می باشند.

انتخاب روش داده کاوی و تکنیک آن

در این تحقیق به علت اینکه در ابتدا هیچ گونه برداشتی از گروه های مشتریان وجود نداشت، از روش خوشه بندی قسمتی با معیار مجذور خطا استفاده شد. همچنین همان طور که قبلاً بیان شد الگوریتم های تقسیمات مجذور خطا تلاش دارد تا تقسیماتی بدست آورد که پراکندگی بین خوشه‌ها را ماکزیمم و پراکندگی داخل خوشه‌ها کمینه کند که این همان چیزی است که در مسئله ما صدق می کند. با توجه به این که داده‌های ما از نوع مقادیر عددی، مطلق و پیوسته هستند الگوریتم میانگین تعداد خوشه‌ها انتخاب می شود.

خوشه بندی

پس از انجام مراحل آماده سازی، یکپارچگی و استخراج داده‌ها با فرمت مناسب، با استفاده از الگوریتم میانگین k خوشه بندی شدند. خوشه بندی یک بار باروش غیر فازی و یک بار با روش فازی انجام شد. تعداد خوشه‌ها ۵ در نظر گرفته شده است که در ادامه علت این انتخاب توضیح داده خواهد شد.

تقسیم بندی مشتریان در ۵ گروه به صورت غیر فازی

در این مرحله برای تقسیم بندی مشتریان از نرم افزار (Spss) استفاده شد. جدول ۳ مراکز خوشه‌ها را ارائه می کند.

جدول ۳. مراکز ۵ خوشه به روش غیر فازی

نام خوشه	مانده حساب روزانه	مبلغ چک برگشتی	مانده بستانکار روزانه	گردش بستانکار روزانه	گردش بدهکار روزانه	مانده پایان دوره روزانه
A	۱۷۳۶۷۶	۲۶۶۶۲۷۱	۵۶۵۷۶۵۶	۱۱۹۸۸۳	۱۲۰۸۳۵۵	۱۶۵۶۹۷
B	۱۲۹۹۶۴	۰	۷۸۱۳۸۳۸	۱۷۷۷۳۶	۱۵۱۸۶۹۱	۳۸۶۶۸۶
C	۶۱۲۶۴۳	۰	۲۱۲۲۳۲۲	۱۰۲۳۸۶	۱۰۱۷۹۹۰	۶۱۷۵۵۱
D	۱۴۸۲۳۹	۴۱۲۰۳۸۱	۳۱۷۷۲۶۶	۱۰۲۳۸۵	۱۱۰۲۴۸۰	۷۰۲۰۵۵
E	۴۸۴۱۰۵	۰	۷۸۲۱۰۷۱	۱۳۰۰۲۷	۱۳۰۰۲۷۵	۱۰۱۸۶۶

خروجی دیگر نرم افزار شامل گروه های اختصاص داده شده به هر یک از نمونه ها و فاصله هر نقطه تا مرکز خوشه می باشد. جدول ۴ نمونه ای از این خروجی را ارائه می کند.

جدول ۴. نمونه ای از خروجی نرم افزار Spss

مانده حساب روزانه	مبلغ چک برگشتی	مانده بستانکار روزانه	گردش بستانکار روزانه	گردش بدهکار روزانه	مانده پایان دوره روزانه	گروه	فاصله
۱۰۵۳	۰۳۴۶	۰۰۳۸	۰۰۷۸	۰۰۷۸۴	۱۶۳	۱	۲۵۵۶۰۵
۰۰۰۱	۱۰۷۱۷	۰۰۰۳	۰۰۰۳	۰۰۰۳۶	۰۰۰	۱	۲۷۳۳۱۱

تقسیم بندی مشتریان به ۵ گروه به صورت فازی

برای خوشه بندی مشتریان در حالت فازی از نرم افزار (DataEngin) استفاده می شود. جدول ۵ نمونه ای از خروجی نرم افزار (DataEngin) و تخصیص مشتریان مختلف به صورت فازی به گروه های مختلف است.

جدول ۵. نمونه ای از خروجی نرم افزار DataEngin

E	D	C	B	A	مانده پایان دوره روزانه	گردش بدهکار روزانه	گردش بستانکار روزانه	مانده بستانکار روزانه	مبلغ چک برگشتی	مانده حساب روزانه
۰.۰۰۵	۰.۹۵	۰.۰	۰.۰	۰.۰۰	۱.۶	۰.۰۷	۰.۰	۰.۰	۰.۳	۱
۰.۰۰۶	۰.۹۴	۰.۰	۰.۰	۰.۰۱	۰.۰	۰.۰۰	۰.۰	۰.۰	۱۱.۷	۰

برای محاسبه تراکم خوشه ها، با استفاده از فرمول شماره ۷ زیر برنامه ای توسط نرم افزار ویژوال بیسیک نوشته شد و تراکم خوشه های حاصل شده به دو روش فازی و غیر فازی محاسبه گردید. همان طور که در جدول ۶ مشاهده می شود تراکم خوشه‌هایی که با روش فازی ایجاد شده است کمتر است و این روش برای خوشه‌بندی مناسب تر است.

جدول ۶. مقایسه تراکم خوشه‌ها با استفاده از روش غیر فازی و فازی

روش خوشه بندی	تراکم خوشه ها
روش غیر فازی	۰.۵
روش فازی	۰.۱

برای تعیین تعداد بهینه خوشه‌ها از الگوریتم ژنتیک استفاده شد. بدین منظور با استفاده از نرم افزار متلب کد الگوریتم ژنتیک برای خوشه بندی مشتریان نوشته شد. شکل ۱ بخشی از کد نوشته شده می باشد.

```

npop=5500
end
npop=5500;
nvar=200;
pc=0.8;
nc=2*round(pc*npop/2);
mu=0.2;
nmu=round(mu*npop);
pop=randint(npop,nvar)
maxit=400;
bestpop=zeros(maxit,nvar)
bestcosts=zeros(maxit,1)
for it=1:maxit
    %calculation of costs
    costs=Cost(pop)
    %sorting the population according to the costs values
    [costs ind]=sort(costs)
    pop=pop(ind(1:npop),:);
    costs=costs(1:npop)
    %update bestpop and bestcosts
    bestpop(it,:)=pop(1)
    bestcosts(it)=costs(1)
    %display results
    disp(['Iteration ' num2str(it) ':
    Best Cost = ' num2str(bestcosts(it))
    if it==maxit
        break;
    end

```

شکل ۱. بخشی از کد الگوریتم ژنتیک توسط متلب

براساس فرمول شماره ۳ نتایج برای پارامتر μ بر طبق جدول زیر می باشد:

با توجه به جدول ۷ مشاهده می شود که مقادیر μ برای $k=5$ بیشتر است. بنابراین براساس فرمول شماره ۵ مقدار f کمتر و بهترین تعداد خوشه ۵ است.

جدول ۷. مقادیر بدست آمده برای μ با تعداد خوشه های مختلف

تعداد خوشه ها=۵	تعداد خوشه ها=۳	دفعات اجرا
۸۷۶.۸۷	۲۱۲.۱۲	۱
۸۵۰.۵۷	۱۹۸.۴۲	۲
۸۴۳.۹۸	۲۱۹.۰۹	۳
۸۶۹.۸۶	۲۰۵.۶۸	۴

تحلیل خوشه‌ها

حال ارزش مشتریان را براساس معیارهای تازگی (R) تکرار (F) ارزش پولی (M) اندازه گیری می‌کنیم.

معیار تازگی مراجعه مشتری نشان دهنده آخرین باری است که مشتری به بانک مراجعه کرده است. مقدار این متغیر بین ۰ و ۳۶۵ است که ارزش امتیاز ۳۶۵ از همه بیشتر است و بدین معناست که به تازگی مشتری به بانک مراجعه کرده است. با توجه به فرمول (۹) مقدار R برای هر مشتری محاسبه شده است. تعداد تکرار مراجعات مشتریان در طی ۱ سال گذشته نشان می‌دهد که این متغیر بین ۰ تا ۳۰ است و با استفاده از معادله (۱۰) مقدار F برای هر مشتری محاسبه شده است. برای محاسبه ارزش پولی هر مشتری مانده پایان دوره در نظر گرفته شود. با توجه به معادله (۱۱) مقدار M برای هر مشتری محاسبه می‌شود. سپس ارزش خوشه‌ها با توجه به معادله (۱۶) محاسبه می‌شود. در این معادله وزن متغیرها براساس نظر خبرگان تعیین شده است.

با توجه به اهمیت تکرار مراجعات برای آن وزن ۰.۵ در نظر گرفته شده است و برای متغیر M وزن ۰.۳ و برای متغیر R وزن ۰.۲ در نظر گرفته شده است. با توجه به معادلات (۱۳)، (۱۴) و (۱۵) تازگی مراجعه مشتری، تکرار و ارزش پولی برای هر خوشه تعیین شده است. سپس با استفاده از معادله (۱۲) ارزش هر خوشه تعیین و نتایج در جدول ۸ است:

جدول ۸. مقادیر محاسبه شده برای ارزیابی خوشه‌ها با استفاده از معادلات ذکر شده

شماره خوشه	میانگین تازگی هر خوشه	میانگین تکرار هر خوشه	میانگین ارزش پولی هر خوشه	میانگین ارزش هر خوشه	تعداد مشتریان عضو خوشه
۴	۱.۳۲	۰.۹۵	۰.۸	۱.۲	۱۵۱۹
۵	۱.۵	۰.۷	۰.۶	۱	۲۵۰
۱	۱.۹۸	۰.۶۶	۰.۴	۰.۹۹	۶۳۴
۳	۱.۷۵	۰.۴۷	۰.۳	۰.۹۲	۱۳۲
۲	۱.۹۸	۰.۳۸	۰.۱	۰.۸۷	۳۴۵

با توجه به هرم ارزش مشتری می‌توان خوشه‌ها را به صورت زیر تحلیل کرد و استراتژی‌های مدیریت ارتباط با مشتری برای هر خوشه را تبیین نمود:

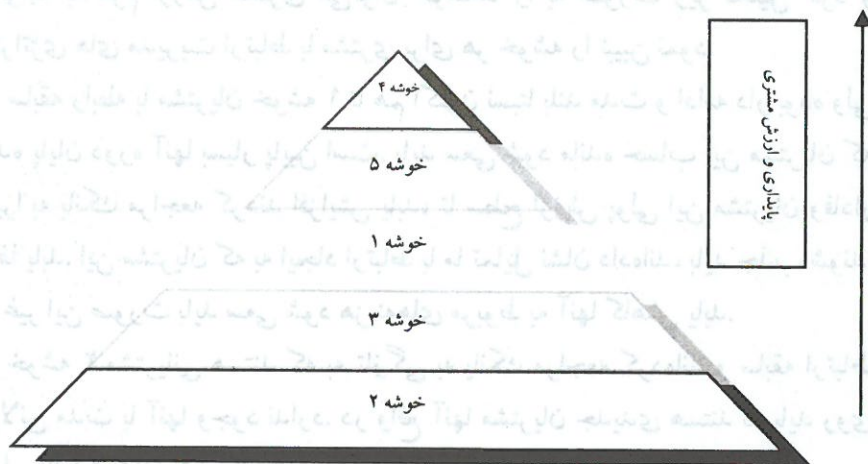
سابقه رابطه با مشتریان خوشه ۱ تا هم اکنون نسبتاً بلند مدت و ادامه دار بوده ولی مانده پایان دوره آنها بسیار پایین است. باید سعی شود مانده حساب این مشتریان که اخیراً به بانک مراجعه کردند افزایش یابد، تا سطح ارزش پولی این مشتریان وفادار ارتقا یابد. این مشتریان که به ایجاد ارتباط با ما تمایل نشان داده‌اند، باید جذب شوند. در غیر این صورت باید سعی شود هزینه‌های مربوط به آنها کاهش یابد.

خوشه ۲ مشتریانی هستند که به تازگی به بانک مراجعه کرده‌اند و سابقه ارتباط طولانی مدت با آنها وجود ندارد. در واقع آنها مشتریان جدیدی هستند که باید روی آنها مطالعات شناخت صورت گیرد و سعی به جذب آنها گردد.

خوشه ۳ مشتریانی هستند که دارای سابقه ارتباطات قبلی هستند ولی به تازگی ارتباط با آنها کمرنگ‌تر شده است. این مشتریان در آستانه رویگردانی و ریزش هستند. باید علت رویگردانی آنها مشخص شده و استراتژی لازم جهت جلوگیری از ریزش آنها اتخاذ گردد. این مشتریان در واقع مشتریان غیر فعال می‌باشند.

خوشه ۴ با ارزش ترین خوشه در بین مشتریان است. مانده پایان دوره این مشتریان بیشتر است. سابقه ارتباط با این مشتریان طولانی مدت است و این تداوم ارتباط تا کنون نیز ادامه داشته است. این مشتریان کلیدی هستند و باید در حفظ و نگهداری آنها تلاش شود تا ریزش نمایند. این مشتریان جزء مشتریان وفادار و فعال در بالاترین سطح سودآوری هستند.

خوشه ۵ مشتریانی هستند که نسبتاً به تازگی و به صورت مداوم حجم قابل قبولی از آنها به بانک مراجعه کرده‌اند. این مشتریان مشتریان معمولی هستند و نسبتاً وفادارند اما از ارزش پایین تری نسبت به خوشه ۴ برخوردار هستند. باید سعی شود مانده حساب این مشتریان افزایش یابد. در شکل ۲ با توجه به نتایج بدست آمده مشتریان در قالب هرم ارزش مشتری پیاده سازی شدند.



شکل ۲. هرم ارزش مشتری بر اساس ۵ خوشه بدست آمده

استخراج قواعد

تا این مرحله چون هیچ دانشی نسبت به سیستم نداشتیم برای تقسیم بندی مشتریان از روش‌های مختلف خوشه بندی استفاده کردیم اما در این مرحله با بدست آوردن گروه‌های مختلف، در حقیقت دانش اولیه را نسبت به سیستم بدست آورده‌ایم و خروجی گروه بندی را برای کلاس بندی بقیه مشتریان و استخراج قواعد کلی سیستم استفاده می‌کنیم. بدین صورت چنان چه مشتری جدیدی وارد شود با دریافت اطلاعات اولیه شاخ‌ها و استفاده از قوانین زیر می‌توان تشخیص داد این مشتری جزء کدام گروه از مشتریان می‌باشد و تصمیمات لازم را در مورد او اتخاذ نماییم. برای استخراج قوانین با استفاده از درخت تصمیم گیری ابتدا خروجی خوشه بندی را که همان فیلدهای انتخابی به اضافه گروه‌های تخصیص یافته به هر کدام می‌باشد را به ۲ قسمت داده آموزش و داده تست تقسیم می‌شود که حدود ۷۵٪ از رکوردها به عنوان داده آموزش و ۲۵٪ باقیمانده به عنوان داده تست معرفی می‌شوند.

البته در منابع مختلف این نسبت متفاوت می‌باشد، ولی در این نسبت باید نهایت دقت در انتخاب مجموعه‌های آموزش و تست داشته باشیم تا بیشترین تنوع داده‌ها برای ورودی و خروجی در هر دو مجموعه حفظ شود و بهترین نتیجه حاصل شود.

سپس این خروجی‌ها را در دو فایل جداگانه Excel ذخیره می‌کنیم. تفسیر این قوانین به این صورت است که به عنوان مثال اگر مانده پایان دوره مشتری کوچکتر از ۴۲۷.۲۳۱ باشد و معدل حساب روزانه مشتری ۲۹۷.۰۰۱ باشد مشتری جزء کلاس B می‌باشد و به همین صورت بقیه قوانین تفسیر می‌شود.

پیشنهادات

۱. افزایش اطلاعات مدیران از تکنولوژی‌های داده کاوی و مزایای آن
۲. وجود زیر ساخت‌های اطلاعاتی یکپارچه و پایگاه‌های اطلاعاتی نرمال
۳. آگاهی برنامه نویسان بانک‌های اطلاعاتی از نیازمندی‌های بانک به برخی اطلاعات ضروری جهت اجرا تکنولوژی‌های جدید
۴. ثبت کامل تمامی اطلاعات در فیله‌های بانک اطلاعاتی
۵. افزایش آگاهی مدیران نسبت به سود آوری حاصل از بازاریابی
۶. توجه به مسائل کیفیت عملکرد و نحوه‌گزینش مدیریت برای پست‌های سازمانی
۷. افزایش دوره‌های نظر سنجی از مشتریان و احترام به خواست و نیاز مشتری

نتیجه گیری و پیشنهادات برای تحقیقات آتی

با افزایش اهمیت رضایت مشتری در محیط تجاری امروز، بسیاری از سازمان‌ها روی مباحث مرتبط با شناخت مشتری، وفاداری و سودآوری مشتری برای افزایش سهم بازار خود و کسب رضایت بیشتر مشتری تمرکز نموده‌اند. مدیریت ارتباط با مشتری به عنوان یک مزیت رقابتی برای سازمان‌ها محسوب می‌گردد. یکی از روش‌های شناخت مشتری، بخش بندی مشتریان به گروه‌های همگن و اتخاذ سیاست‌های بازاریابی متناسب با هر بخش است. در این مقاله پس از خوشه بندی مشتریان به دو روش فازی و غیر فازی با توجه به اینکه معیار تراکم خوشه‌ها در روش فازی کمتر است نتیجه گرفتیم همواره روش فازی برای خوشه بندی مشتریان از دقت بیشتری برخوردار است. همچنین برای اینکه تعداد بهینه خوشه‌ها را بدست آوریم بهترین روش، از الگوریتم ژنتیک استفاده کردیم. در این مقاله سپس ارزش

خوشه‌ها تعیین شد و در انتها برای تحلیل خوشه‌ها و تبیین استراتژی مناسب برای هر خوشه از هرم ارزش مشتری بهره گرفته شد و با توجه به نتایج بدست آمده دیگر نیازی به شناخت تک تک مشتریان نمی‌باشد، بلکه هر خوشه معرف ویژگی‌های مشتریان موجود در آن گروه است. در انتها نیز با استفاده از درخت تصمیم‌گیری قوانین شناسایی مشتریان استخراج شد.

در این مقاله با یکپارچه سازی چند مازول از جمله بانک اطلاعاتی، پیش پردازش‌های اولیه و درخت تصمیم‌گیری پایگاه دانشی ایجاد شد که می‌تواند برای شناسایی و اعتبارسنجی مشتریان کارا باشد. گرچه با بکارگیری الگوریتم ژنتیک در الگوریتم‌های درخت تصمیم و شبکه‌های عصبی ممکن است بتوان دانش بهتری کسب کرد. از سوی دیگر، با بکارگیری مجموعه داده با استفاده از سیستم‌های استنتاج فازی و روش‌های کلاسیک نظیر رگرسیون می‌توان نتایجی را کسب کرد و نتایج حاصله را مورد مقایسه قرار داد و با کاربرد تکنیک‌های آماری اعتبار مدل را بررسی نمود.

تحقیقاتی که در آینده می‌تواند مورد توجه قرار گیرند عبارتند از:

- ۱- خوشه‌بندی داده‌ها براساس سایر روش‌های خوشه‌بندی مانند شبکه عصبی، الگوریتم مورچگان و.... و مقایسه الگوریتم‌ها
- ۲- پیش‌بینی رفتار مشتریان و برنامه ریزی تقاضا
- ۳- ارائه متد جدید برای خوشه‌بندی و بهبود الگوریتم‌های موجود

1. Abdou , H . , & Pointon , J ,(2008) “**Neutral nets versus conventional techniques in credit scoring in Egyptian banking**”, Expert System with Application , Vol, xxx,pp.1
2. Alex Berson& Stephen Smith& kurt Thearling , (2001) , “**Bulding Data Mining Application for CRM**” , McGraw- Hill.
3. C. Y. Tsai, C. Chiu, (2004) , " A purchase- based market segmentation methodology" , Expert Systems with Applications 27 (2004) 265– 276
4. Chen, M- Ch., Huang, Sh- H. (2003) “**Credit scoring and rejected instances reassigning through evolutionary computation techniques**”, Expert Systems with Applications, 24(4), pp. 433- 441.
5. Hyunseok Hwang, Taesoo Jung (2004)," **An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry**", Expert Systems with Applications 26 (2004) 181–188
6. Jussi Ahola, Esa Rinta-Runsala, (2002) "**Data Mining Case Studies in Customer Profiling**", VTT information Technology
7. Kantardzic, Mehmed, (2003), "**Data Mining: Concepts, models, methods, and algorithms**", wiley- inter science
8. Kim, Y. S., & Sohn, S. Y. (2004). **Managing loan customers using misclassification patterns of credit scoring model**. Expert S ystems with Applications, 26, pp. 567– 573.
9. Lee, T. S., Chen, I. F. (2005). **A two- stage hybrid credit scoring model using Systems with Applications**, 28, 743–725
10. Nanni, L. & Lumini , A,(2009) “**An experimental comparison of ensemble of classifiers for bankruptcy predication and credit scoring**”, Expert System with Application ,Vol. 36 , pp. 1-4
11. Olson, D. & , Shi. Y. , “**Introduouction to Bussiness Data Maning. Singapore**” :MCGrow Hill Education , 2007
12. Valarie A. Zeithaml, Roland T. Rust, and Katherine N. Lemon,(2001), “ **The Customer Pyramid: Creating and Serving Profitable Customers**”, California Management Review.
13. Xu , X. , Zhou , C. , & Wang , Z," “**Creadit scoring algorithm based on link analysis ranking with support vector machne**”, Expert System with Applicatoin , Vol.xxx.pp.6, 2008
14. Zhang, Y., & Bhattacharyya,S., “**Genetic Programing in classifying large-scale data:an ensemble method**”.Information Sciences,vol 163,pp.2,6,2009