فصلنامه علمی- پژوهشی تحقیقات بازاریابی نوین
سال ششم، شماره اول، شماره پیاپی (۲۰) بهار ۱۳۹۵
تاریخ دریافت: ۹۴/۱۱/۸ تاریخ پذیرش: ۹۵/۳/۳
صص: ۲۲-۱

مطالعهٔ مقایسه‌ای الگوریتم‌های خوشه‌بندی در راستای سنجش ارزش مشتری در سیستم مدیریت ارتباط با مشتری در صنعت بیمه

محمد صالح ترکستانی[۱]، طاها منصوری[۲]، یاسمین تقی‌زاده[۳]*
۱- استادیار گروه مدیریت بازرگانی، دانشگاه علامه طباطبایی
۲- دانشجوی دکتری مدیریت فناوری اطلاعات دانشگاه علامه طباطبایی
۳- دانشجوی کارشناسی ارشد، گروه مهندسی فناوری اطلاعات، موسسهٔ آموزش عالی مهر البرز

چکیده

دنیای رقابتی امروز، تعامل شرکت‌ها و سازمان‌ها را با مشتریان خود به طور قابل توجهی تغییر و بهبود داده است. از این رو، یکی از چالش‌های اساسی سازمان‌ها، شناسایی مشتریان فعلی خود، تخصیص بهینهٔ منابع به آنها و ایجاد تمایز بین گروه‌های مختلف مشتری و در نهایت سنجش و رتبه‌بندی هر گروه بر مبنای ارزش آن گروه است. هدف اصلی این پژوهش، مطالعهٔ مقایسه‌ای بین الگوریتم‌های خوشه‌بندی به عنوان یکی از مهمترین تکنیک‌های داده کاوی و به‌کارگیری آنها در راستای سنجش ارزش مشتریان(ارزش طول عمر مشتری) است. صنعت مورد مطالعه در این پژوهش، صنعت بیمه و به صورت جزئی‌تر، حوزهٔ بیمهٔ بدنهٔ اتومبیل است. روش پیشنهادی این پژوهش، برگرفته از روش CRISP-DM است و نیز در راستای بخش‌بندی و سنجش ارزش طول عمر مشتریان، از یکی از مدل‌های توسعه یافته RFM استفاده شده است. در این پژوهش اطلاعات نزدیک به ۸۱۲ مشتری از شرکت منتخب در حوزهٔ بیمه(بیمهٔ بدنهٔ خودرو) استخراج شده است و سپس با استفاده از نرم افزارهای توانمند در حوزهٔ داده‌کاوی و محاسبات آماری به نام‌های RapidMiner و SPSS و Matlab و با به‌کارگیری الگوریتم‌های مختلف، مشتریان خوشه‌بندی شده‌اند. پس از تحلیل نتایج هر خوشه و تعیین بهترین روش خوشه‌بندی بر مبنای شاخص‌های Silhouette و SSE، به تحلیل و مقایسهٔ ارزش مشتریان در خوشه‌های به‌دست آمده از بهترین شیوهٔ خوشه‌بندی می‌پردازیم. در نهایت با استفاده از فرایند پیشنهادی این پژوهش می توان ارزش هر گروه از مشتریان را در صنعت بیمه به دست آورد و استراتژی‌های بازاریابی متناسب با هر گروه از مشتریان را ارائه نمود.

واژه‌های کلیدی: بیمه، خوشه‌بندی، داده‌کاوی، مدیریت ارتباط با مشتری، ارزش طول عمر مشتری، مدل RFM

* نویسنده مسؤول y_taghizadeh2003@yahoo.com

Archive of SID
www.SID.ir

۱-۱- مقدمه

از سال ۱۹۸۰ مدیریت ارتباط با مشتری از اهمیت خاصی در حیطهٔ کسب وکار نوین مشتری محور، برخوردار شده است. امروزه سازمان‌ها در راستای شناسایی و سنجش ارزش مشتریان خود، به بررسی ویژگی‌های رفتاری آنان پرداخته و سعی نموده‌اند با هر مشتری به شیوهٔ خاص او برخورد کنند و این امر ضمن ایجاد ارتباط و تعامل بیشتر با مشتریان، به ایجاد مزیت رقابتی و سودآوری برای سازمان منجر می‌شود(قره خانی و ابولقاسمی، ۱۳۹۱).

شرکت‌های بیمه، از نهادهای پویا و مهم در بسیاری از کشورهای جهان، از جمله ایران، به حساب می‌آیند و نیز به عنوان یک بخش خدماتی، در اقتصاد کشورها نقش جبران کننده و محافظتی دارند، به گونه‌ای که عملیات موفقیت‌آمیز آنها، انگیزه و محرکی برای دیگر صنایع و توسعهٔ اقتصاد کشور می‌شود(حسن زاده، ۱۳۸۷). موجودیت شرکت‌های بیمه و رشد آنها برای ایفای هر چه بهتر نقش خود در صحنهٔ اقتصادی هر کشور، تا حد زیادی به رضایت مشتریان بستگی دارد. در گذشته، شرکت‌های بیمه روابط نزدیکی با مشتریان خود داشتند و با توجه به نیازهای مشتریان، سعی می‌کردند با ارائهٔ خدمات مناسب به آنها، نیازها و خواسته‌های مشتریان خود را برطرف کنند؛ ولی در طول سال‌های بعد، با افزایش تعداد مشتریان و رقابت در بازارکاری، سازمان بیمهٔ مذکور، به منظور شناسایی، جذب و حفظ مشتریان با ارزش خود، به دنبال استراتژی‌های جدید بر مبنای فناوری‌های نوین بوده است(چسی و کیس[1]، ۲۰۰۰، ۲۷۵-۲۹۶). با همین هدف، ابزارها و متدهای متعدد داده‌کاوی به وجود آمده است. یکی از این متدها، خوشه‌بندی است که در کشف تصادفی دانش کاربرد دارد.

امروزه، یکی از پرکاربردترین شاخص‌های مهمی که در بخش‌بندی(خوشه‌بندی) مشتریان مطرح می‌شود، ارزش طول عمر مشتری است که سازمان‌ها را در شناسایی مشتریان با ارزش خود توانمند می‌کند(هو و جینگ[2]، ۲۰۰۸). ارزش طول عمر مشتری یکی از پارامترهای قابل اندازه‌گیری است که نقش مهمی در بخش‌بندی مشتریان ایفا نموده و به عنوان یکی از ورودی‌های اصلی برای الگوریتم‌های خوشه‌بندی مشتری مورد استفاده قرار می‌گیرد.

پروژه جاری سعی دارد با توجه به شرایط صنعت بیمهٔ ایران(حوزهٔ بیمهٔ بدنهٔ اتومبیل)، ابتدا مدل مناسب را برای سنجش ارزش مشتری در این حوزه معرفی می‌کند و سپس روش‌ها و متدهای مختلف خوشه بندی را در راستای سنجش ارزش مشتریان مورد تحلیل و بررسی، مقایسه و اعتبار سنجی قرار می‌دهد و در نهایت با استفاده از پارامتر قابل اندازه‌گیری ارزش طول عمر مشتری، به تحلیل و بررسی ارزش مشتریان در هر خوشه می‌پردازد.

لازم به ذکر است که در راستای سنجش ارزش طول عمر مشتری، مدل‌های مختلف وجود دارد، ولی در این پژوهش ما از یک مدل توسعه یافته بر مبنای مدل RFM (LRFMC) مناسب برای صنعت بیمه استفاده می‌کنیم.

اولین هدف در این پژوهش، معرفی الگوریتم مناسبی برای خوشه‌بندی مشتریان بر مبنای ارزش طول عمر آنها است. بدین منظور ابتدا باید مدلی مناسب در صنعت بیمه در راستای سنجش ارزش مشتریان بیابیم و سپس با تبدیل این مدل به پارامترهای قابل اندازه‌گیری،

1 Chassey, Case

2 Hu, W., Jing, Z.

Archive of SID

www.SID.ir

به عنوان ورودی الگوریتم‌های خوشه‌بندی، به بخش-بندی مشتریان صنعت بیمه بپردازیم. در این راستا اولین سؤالی که مطرح می‌شود این است که به منظور سنجش ارزش مشتریان در صنعت بیمه، چه متدها و مدل‌هایی وجود دارد و کدام بهینه‌تر است؟ دیگر اینکه برای خوشه‌بندی در راستای سنجش ارزش مشتریان، با استفاده از داده‌های سیستم مدیریت ارتباط با مشتری، از چه متدها و الگوریتم هایی می‌توان استفاده کرد؟ نیز اینکه نقاط قوت و ضعف هر کدام از این متدها و الگوریتم‌ها چیست؟ و در نهایت سوال اصلی که این پژوهش سعی دارد به آن پاسخ دهد این است که بنا به شرایط صنعت بیمهٔ ایران، چه فرآیندی برای بخش‌بندی مشتریان، به منظور سنجش ارزش مشتریان، پیشنهاد می-شود؟

لازم به ذکر است که با توجه به وجود تنوع بسیار در صنعت بیمهٔ ایران، بیمهٔ بدنهٔ اتومبیل به دلیل پر اهمیت بودن، اختیاری بودن و نیز ارتباط مستقیم این نوع بیمه با مشتریان حقیقی خود انتخاب شده است و سازمان مورد بررسی، شرکت منتخب در حوزهٔ بیمهٔ بدنهٔ اتومبیل است. به منظور سنجش ارزش مشتری در سازمان مورد مطالعه، با استفاده از پرسشنامه و مطالعهٔ میدانی، پارامترهای مدل توسعه یافتهٔ LRFMC و وزن هر یک از آنها استخراج می‌شود.

جامعهٔ آماری در این پژوهش شامل اطلاعات مشتریان در بازهٔ زمانی ابتدای سال ۱۳۹۱ تا ابتدای سال ۱۳۹۳ است که در حدود ۸۱۲ رکورد را شامل می‌شود. از آنجا که داده‌کاوی، فرآیند استخراج دانش قابل فهم و قابل اعتماد از پایگاه‌های بزرگ داده است، می‌تواند کاربران را برای برقراری ارتباط مستقیم با حجم عظیم داده‌ها توانمند کند(صنیعی، محمودی، طاهرپور، ۱۳۹۳). به همین دلیل در پژوهش جاری، نمونهٔ آماری همان جامعهٔ آماری است. در ابتدا کلیهٔ اطلاعات جامعهٔ آماری به عنوان ورودی پژوهش در نظر گرفته می‌شود، ولی ممکن است در طول تحقیق یکسری از داده‌ها برای آماده‌سازی و پیش پردازش داده‌ها به منظور نرمالسازی و پاکسازی، حذف گردد.

۲-مبانی نظری پژوهش

یکی از متدهای پرکاربرد در داده‌کاوی، متد CRISP-DM[1] است(شیرر[2]، ۲۰۰۰). در پژوهش جاری با توجه به اهداف بیان شده، از متدولوژی CRISP، به منظور تجزیه و تحلیل آماری داده‌ها و اطلاعات استفاده می‌شود. این متدولوژی از مراحل زیر تشکیل شده است:

۱.فهم تجاری/ ۲. درک داده/ ۳.آماده سازی داده/ ۴.مدل سازی/ ۵.ارزیابی/ ۶. توسعه.

در شکل ۱، بر اساس این مدل، طرح کلی پژوهش را مشاهده می‌کنید.

۲-۱-فاز۱: درک محیط کسب و کار

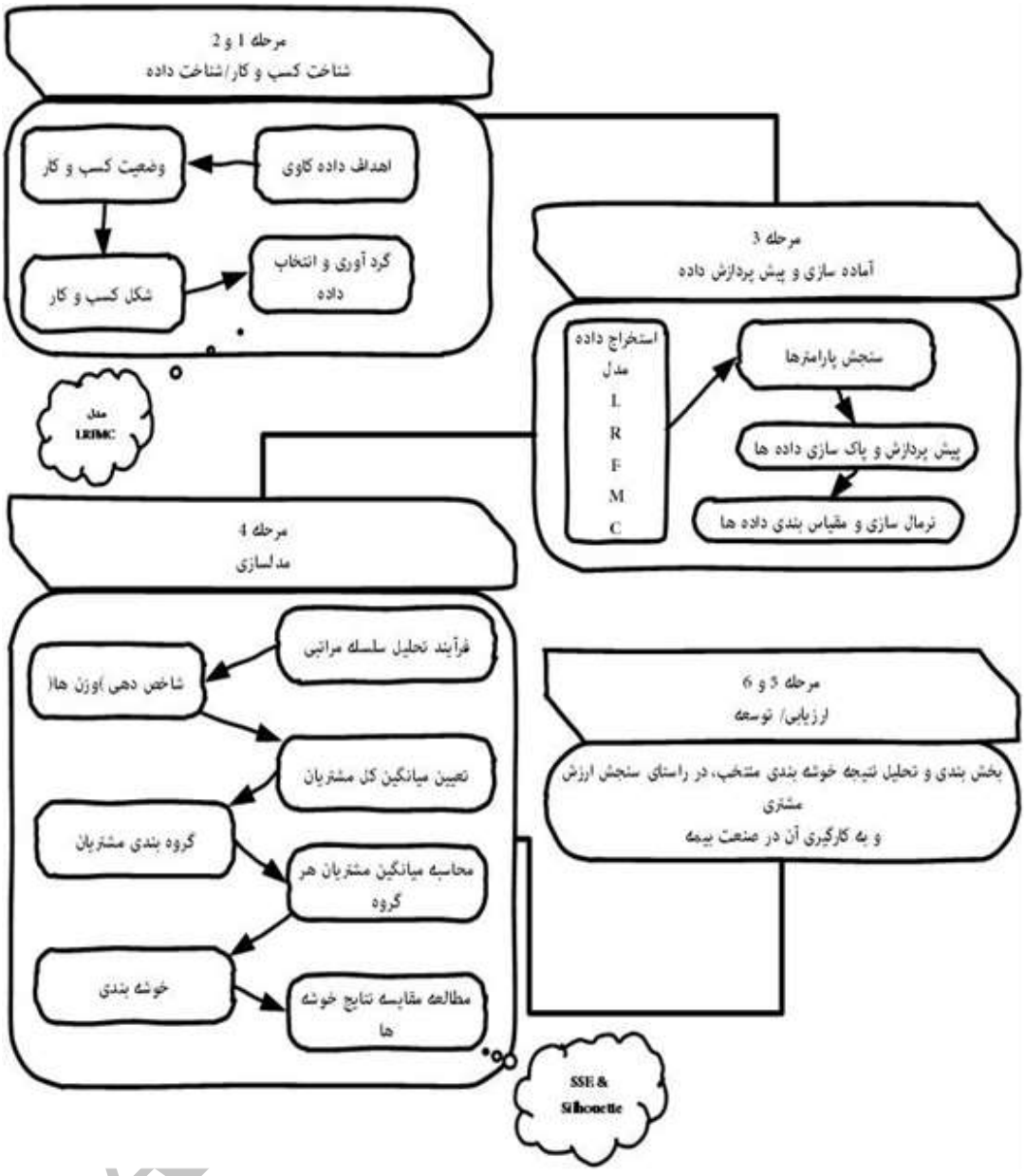
۲-۱-۱-بیمه

همان گونه که قبلا نیز گفتیم، صنعت بیمه به عنوان یکی از عمده‌ترین شاخص‌های توسعه‌یافتگی نهادهای اقتصادی مطرح می‌شود(ترکستانی و همکاران، ۱۳۹۳).

1 Cross Industry Standard Process For Data-Mining
2 Shearer C.

Archive of SID

www.SID.ir



شکل ۱. چارچوب پژوهش

۲-۱-۲-بیمهٔ خودرو[1]

استفاده از اتومبیل از همان آغاز، با خطرات و حوادث جانی و مالی همراه بوده است. خوشبختانه می‌توان با ایجاد و تعمیم انواع مختلف بیمه(بیمهٔ اتومبیل) و به ویژه اجباری کردن بیمهٔ شخص ثالث، تا حدودی عواقب زیان‌بار آن را قابل تحمل و جبران نمود(ماهنامهٔ بیمه-پژوهشکدهٔ بیمه، ۱۳۹۱، ش ۳۵):

یکی از انواع بیمه‌های خودرو در ایران، بیمهٔ بدنهٔ اتومبیل است. این نوع بیمه، جبران کنندهٔ خسارات ناشی از حوادثی است که برای اتومبیل بیمه شده اتفاق می‌افتد و بطور معمول و در بیمه نامه‌های عادی، سه خطر: حادثه، آتش‌سوزی و دزدی را شامل می‌شودو

1 Vehicle Insurance

www.SID.ir

همچنین انواع پوشش‌های تکمیلی، مانند غرامت، نوسانات قیمت، خطرات ناشی از مواد شیمیایی و... را دربرمی‌گیرد. ما در این پژوهش به دلیل غیر اجباری بودن بیمهٔ بدنهٔ خودرو و نیز ارتباط مستقیم شرکت بیمه با مشتری، این حوزه از بیمه را برای مطالعه انتخاب نموده‌ایم(ترکستانی و همکاران، ۱۳۹۳).

۲-۲- مدیریت ارتباط با مشتری[1]

مدیریت ارتباط با مشتری، به همهٔ فرآیندها و فناوری‌هایی گفته می‌شود، که در شرکت ها و سازمان‌ها برای شناسایی، ترغیب، گسترش، حفظ و ارائهٔ خدمات و محصولات بهتر به مشتریان به کار می‌رود(شاء، رابرت[2]، ۱۹۹۱). CRM از سه بخش اصلی روابط ، مدیریت و مشتری تشکیل شده است(صداقتی، ۱۳۹۱).

۲-۳- فاز۲: درک داده

۲-۳-۱-ارزش دورهٔ عمر مشتری

ایدهٔ اصلی ارزش چرخهٔ عمر مشتری این است که مشتریان را باید بر اساس سودآوری آنها برای سازمان، مورد قضاوت قرار داد. از طرفی پژوهشگران و مدیران، همه بر این عقیده هستند که شرکت‌ها نباید به دنبال جذب هرگونه مشتری باشند، بلکه آنها بایستی مشتری مناسب را جذب کنند. براساس مطالعات ارائه شده توسط ای. تی. کیرنی، می‌بینیم که چگونه برخی از مشتریان برای ما بسیار ارزشمند هستند، در حالی‌که برخی دیگر می‌توانند بر سود سازمان اثر منفی داشته باشند(ای تی کرنر[3]، ۲۰۰۳).

دیدگاه‌های مختلفی نسبت به موضوع ارزش دورهٔ عمر مشتری وجود دارد. در ادامه به بررسی متداول‌ترین این روش‌ها می‌پردازیم:

۱. روش ارزش فعالی خالص(برگر و نادا، ۱۹۹۸)،

۲. روش سهم کیف پول(کومارورفیندتز، ۲۰۰۶)،

۳. روش زنجیرهٔ مارکوف(نیزوکاراوی، ۲۰۰۰)،

۴. روش ارزش گذشته مشتری(رزمی و قنبری، ۱۳۸۸)،

۵.روش بازگشت سرمایه(نرو،۲۰۰۱)،

۶.مدل RFM: مبنای تعیین ارزش دورهٔ عمر مشتری، سه شاخص RFM، یعنی تازگی، تعداد دفعات مبادله، ارزش پولی مبادله است که توسط هوگس در سال ۱۹۹۴ ارائه شده است. مزیت این مدل نسبت به سایر مدل‌ها، نگرش‌های غیر مالی آن به جای نگرش‌های صرفاً مالی به ویژگی مشتریان است و دیگر دیدگاه چند بعدی آن نسبت به روش‌های دیگر که دارای دیدگاه تک بعدی هستند.

حال با توجه به صنعت بیمه و مشتریان، برای سنجش ارزش طول عمر مشتریان از مدل توسعه یافتهٔ LRFMC استفاده می‌کنیم.

شاخص‌های مدل LRFMC (سوئینی[4]، ۲۰۱۲):

1 CRM-customer relationship management
2 Shaw, Robert
3 A. T. Kearney
4 Soeini

Archive of SID

www.SID.ir

جدول ۱.شاخص‌های مدل LRFMC

شاخص	تعریف	واحد
(Length) L	بازهٔ زمانی ارتباط با مشتری(از اولین قرارداد تا آخرین قرارداد مشتری با سازمان)	تعداد سال ها
(Recency) R	مدت زمان بین آخرین قرارداد مشتری با سازمان (که بر مبنای اصول شرکت های بیمه سالیانه در نظر گرفته می شود)	تعداد سال ها
(Frequency) F	تعداد قراردادهای مشتری در طول ارتباط با سازمان	تعداد دفعات
(Monetary) M	حق بیمه مشتری(که باید در بازهٔ زمانی مذکور توسط مشتری پرداخت شود)	واحد پولی- ریال
(Costs) C	میانگین خسارت وارده توسط مشتری در طول ارتباط با سازمان	واحد پولی- ریال

به منظور نرمالسازی پارامترهای مدل مذکور، با استفاده از فرمول‌های زیر، متغیرها را، ما بین اعداد صفر تا یک نرمال سازی می‌کنیم(به ازای کوچکترین داده، مقدار صفر و به ازای بزرگترین داده، مقدار یک به دست خواهد آمد). نکتهٔ مهم در اینجا تأثیر منفی شاخص R و C در ارزش دورهٔ عمر مشتری است (۱).

تأثیر مثبت $Z_{ij} = \frac{X_{ij} - R_{min}}{R_i}$ تأثیر منفی $Z_{ij} = \frac{R_{max} - X_{ij}}{R_i}$

R_{min}: کوچکترین مقدار / R_i: دامنهٔ متغیر(اختلاف کوچکترین و بزرگترین مقدار شاخص)/ R_{max} : بزرگترین مقدار

۲-۴-فاز۳: آماده‌سازی و پیش‌پردازش داده‌ها

۲-۴-۱-پاک‌سازی داده‌ها

یکی از مشکلات شایع داده‌ها،که به ایجاد نتایج نامطلوب در داده‌کاوی منجر می‌شود، پایین بودن کیفیت آنها است(حسینی، فرشته، ۱۳۹۲). در این پژوهش با استفاده از برنامهٔ RapidMiner، به شناسایی و حذف داده‌های پرت می‌پردازیم و نیز با استفاده از تکنیک‌های فرآیند تحلیل سلسله مراتبی[1]، شاخص مدل مذکور را وزن‌دهی خواهیم کرد. به گونه‌ای که مجموعهٔ وزن‌های به دست آمده مساوی یک شود.

مجموعهٔ وزن‌ها: $W_L + W_R + W_F + W_M + W_C = 1$

بعد از یافتن وزن شاخص‌ها به تعیین ارزش متغیرها برای هر مشتری می‌پردازیم. در این مرحله، برای تعیین ارزش متغیرها برای هر مشتری، لازم است مقدار نرمال شدهٔ متغیرها را در وزن به دست آمده در مرحلهٔ قبل ضرب کرد(۲):

$$X_w^n = w_L \times X_n$$

این مقادیر، ورودی الگوریتم‌های خوشه‌بندی در فاز بعد، خواهد بود.

۲-۵-فاز۴: مدل‌سازی

۲-۵-۱-داده‌کاوی

داده‌کاوی، در فرآیند مدیریت ارتباط با مشتری، نقش مهمی را بازی می‌کند که از سویی با مراکز داده و یا انبار داده‌ای تعامل دارد و از سوی دیگر، با نرم افزار مدیریت رقابت در تعامل است(وانگ، ین[2]، ۲۰۰۲: ۱۱). داده‌کاوی با تمرکز بر روی مهمترین اطلاعات، به تحلیلگران شرکت‌های بیمه کمک شایانی می‌کند. در حقیقت، داده‌کاوی با پیش‌بینی خسارت‌های تقلبی و پوشش‌های درمانی و همچنین با پیش‌بینی نیازهای مشتریان و تقسیم بندی مشتریان، به صنعت بیمه کمک فراوانی می‌کند(اوهام هسواری و همکاران، ۲۰۱۴).

1 AHP Analytical Hierarchy process

2 Wang J., Yen D.

Archive of SID

www.SID.ir

۲-۵-۲-خوشه‌بندی

خوشه‌بندی برای تشخیص گروه‌های مشابه از داده‌ها و گروه بندی براساس رفتار یا ارزش مشتریان به کار می‌رود. با این شیوه می‌توان مشتریان و بازارهای هدف را دسته‌بندی کرد. خوشه ندی به این شکل انجام می‌شود که رکوردهایی که بیشترین شباهت به یکدیگر دارند(با توجه به معیار شباهت تعریف شده) در یک خوشه قرار می‌گیرند. در نتیجه، داده‌های موجود در خوشه‌های متفاوت کمترین شباهت را به یکدیگر خواهند داشت(مدهولاتا ، ۲۰۱۲).

در ادامه انواع الگوریتم‌هایی که در این پژوهش مورد استفاده قرار گرفته است به اختصار توضیح داده می‌شود:

۱- الگوریتم سلسله مراتبی Ward CA :

درتحلیل خوشه‌ای(سلسله مراتبی)، هر داده به عنوان یک خوشهٔ مجزا شروع می‌شود؛ برای نمونه، به اندازهٔ داده‌ها خوشه وجود دارد، پس خوشه‌ها را به ترتیب ترکیب می‌کند و تعداد خوشه‌ها را در هر قدم کاهش می‌دهد تا اینکه فقط یک خوشه باقی بماند. به نمودار درختی سلسله مراتبی، دندروگرام گفته می‌شود که نشان دهندهٔ نقاط اتصال است. معمولاً اگر تعداد خوشه‌ها قبل از گروه‌بندی مشخص نباشد، از خوشه‌بندی سلسله مراتبی استفاده می‌شود و نیز اندازه‌های فاصله را می‌توان از طرق مختلف اندازه‌گیری کردکه رایج‌ترین آن فاصلهٔ اقلیدسی است.

۲- الگوریتم چند میانگینی K-means CA:

روش K-means یکی از روش‌های خوشه‌بندی مبنا در داده‌کاوی است. در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شود. سپس داده‌ها با توجه به میزان نزدیکی(شباهت)، به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال، می‌توان در هر تکرار با میانگین‌گیری از داده‌ها، مراکز جدیدی برای آنها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود. الگوریتم K-means برای اجرا نیاز به یک مقدار اولیه(k) یعنی تعداد خوشه‌ها دارد. این امر را می‌توان به عنوان یکی از بزرگترین عیوب این الگوریتم دانست. همچنین از آنجا که این الگوریتم همیشه نزدیکترین شئ به مرکز خوشه‌ها را می‌جوید، ایجاد خوشه‌هایی با اندازهٔ یکسان را ترجیح می‌دهد و این موضوع سبب می‌شود که غالباً مرز بین خوشه‌ها به درستی مشخص نگردد. علی‌رغم اینکه خاتمه‌پذیری در این الگوریتم تضمین شده است، ولی جواب نهایی واحد نیست و همواره در حال تغییر و وابسته به مقدار خوشه‌های اولیه است.

تعیین تعداد اولیهٔ خوشه‌ها:

- تحلیلگران معمولاً در مدل RFM تعداد خوشه‌ها را هشت تا در نظر می‌گیرند. از مقایسهٔ دو به دوی شاخص‌های مدل مذکور، هشت حالت(۲×۲×۲) حاصل می‌شود.

- روش دیگر برای تعیین خوشه‌های بهینه، استفاده از الگوریتم‌های سلسله مراتبی است. یکی از این الگوریتم‌ها Ward یا الگوریتم نگهبان SOM است که در این روش، ابتدا تعداد مناسب خوشه تخمین زده می‌شود(حبیب پور و صفری، ۱۳۹۰).

- روش دیگر برای تخمین تعداد خوشه‌ها، استفاده از شاخص دان است.

در این پژوهش برای رسیدن به تعداد خوشه‌های بهینه با کیفیت بالا، از روش یا شاخص وارد استفاده می‌شود.

Archive of SID

www.SID.ir

۳- الگوریتم Two Step & X-means CA:

الگوریتم‌های X-means و Two-Steps، به هدف بهبود و رفع نقایص الگوریتم K-means ایجاد و تولید شده‌اند. در این الگوریتم‌ها دیگر نیازی به تعیین تعداد خوشه‌ها از سوی کاربر نیست و با گرفتن کمینه و پیشینهٔ تعداد خوشه‌ها، بهترین تعداد خوشه را تخمین می‌زنند. تفاوت این دو الگوریتم، در شکل خوشه‌هایی است که ارائه می‌دهند. الگوریتم x-means مانند الگوریتم K-means، همواره سعی در تشکیل خوشه‌های کروی دارد، اما الگوریتم Two-Steps، خوشه‌هایی با شکل دلخواه ایجاد می‌کند.

۴- الگوریتم DBSCAN[1] CA:

روشDBSCAN، معروفترین روش خوشه‌بندی مبتنی بر چگالی است. در خوشه‌بندی مبتنی بر چگالی، خوشه‌ها به عنوان نواحی چگال از مجموعه‌داده، تعریف می‌شوند. ولی با توجه به مزایای بالای این الگوریتم، مانند عدم نیاز به مشخص بودن تعداد خوشه‌ها و یا ایجاد خوشه‌هایی با اشکال مختلف و شناسایی داده‌های پرت، معایبی نیز داشته است(مولوی، چمپلو و سندر[2]، ۲۰۱۳). این الگوریتم نمی‌تواند مجموعهٔ داده‌هایی با اختلاف چگالی زیاد را خوشه‌بندی کند. سربار بالای محاسباتی دارد. کیفیت DBSCAN به نوع اندازه‌گیری فاصلهٔ نقاط بستگی دارد. فاصلهٔ اقلیدسی، رایج‌ترین نوعی است که استفاده می‌شود. در مورد داده‌هایی در ابعاد بالا، این نوع اندازه‌گیری فاصله بی‌فایده است.

۵- الگوریتم Kohonen CA:

این الگوریتم بر مبنای الگوریتم خود سازمانده شبکه‌های عصبی طراحی شده است. یکی از معایب این الگوریتم، عدم تفسیر خروجی آن توسط کاربران و برنامه‌های دیگر است(تی. کوهونن[3]، ۱۹۸۹).

۶- الگوریتم NKA CA:

این الگوریتم ترکیبی از الگوریتم K-means و عملگر جهش از الگوریتم‌های تکاملی است. الگوریتم NKA یک الگوریتم مبنایی معمولی نیست. بخش اصلی الگوریتم NKA از الگوریتم معمولی K-means به دست آمده است و به منظور بهبود عملکرد بهتر، پارامتر جهش با ایجاد یک نویز از در دام افتادن، در حالت‌های بهینهٔ محلی جلوگیری می‌کند. زمانی که الگوریتم NKA به سمت حالت بهینهٔ محلی می‌رود، عملگر جهش با احتمال P و به صورت رندومی، یکی از نودهای والد را انتخاب می‌کند و این امر سبب می‌شود که مراکز خوشه‌هایی که توزیع یکنواخت دارند، تغییر کند. در ابتدا مقدار جهش کم است و زمانی که هیچ بهبودی در اجرای الگوریتم مشاهده نشود، یعنی احتمالاً الگوریتم NKA در حالت بهینهٔ محلی قرار گرفته است. زمانی که عملکرد تابع هدف افت پیدا کند، مقدار جهش افزایش می‌یابد(منصوری و غلامیان، ۲۰۱۴).

۳- روش تحقیق

فاز ۵-ارزیابی

در این پژوهش، پس از خوشه‌بندی مشتریان با استفاده از روش‌های مختلف، با استفاده از دو پارامتر SSE و Silhouette به مطالعهٔ مقایسه‌ای بین روش‌های خوشه‌بندی می‌پردازیم و در نهایت با انتخاب برترین روش در این مدل، خوشه‌های حاصل از آنها را بخش‌بندی و رتبه‌بندی می‌کنیم. ارزش طول عمر مشتری در هر خوشه را محاسبه می‌شود و خوشه‌ها در

1 Density Based Clustering
2 Champello & Moulavi & Sander
3 T. Kohonen

Archive of SID

www.SID.ir

هـرم ارزش دورهٔ عمـر مشـتری قـرار مـی‌گیـرد و استراتژی‌های هر بخش با توجه بـه هـرم مـذکور تعیـین می‌گردد.

فرمـول محاسـبهٔ ارزش دورهٔ عمـر مشـتریان در هـر خوشه:

$$CLV=\overline{L_{cl}}+\overline{R_{cl}}+\overline{F_{cl}}+\overline{M_{cl}}+\overline{C_{cl}}$$

۴-تجزیه و تحلیل داده‌ها

۴-۱- فاز ۶-تدوین(کاربری):

در این پژوهش، باید از نتایج حاصل از اجرای مـدل پیشنهادی در کسب و کار مد نظر، یعنی صـنعت بیمـه و حوزهٔ بیمهٔ بدنهٔ اتومبیـل، بـرای سـنجش ارزش مشـتریان استفاده کرد. همان طور که بیان شد مـدل پیشـنهادی در این پژوهش، برای شناسـایی و سـنجش ارزش مشـتریان در صنعت بیمه مدلی ساخته است.

۴-۲- خلاصهٔ داده‌ها

پس از شناسـایی واسـتخراج داده‌هـا بـر مبنـای مـدل توسعه یافته LRFMC، در جدول ۲، خلاصه‌ای از داده-هـای اسـتخراج شـده از سـازمان بیمـهٔ مـورد مطالعـه را مشاهده می‌کنید.

جدول ۲. وضعیت شاخص داده‌های استخراج شده

شاخص	واحد	max	min	Avg
(Length) L	تعداد سال ها	۳	۱	۱٫۸
(Recency) R	تعداد سال ها	۲	۰	۱
(Frequency) F	تعداد دفعات	۳	۱	۱٫۸
(Monetary) M	واحد پولی- ریال	۴۸۴,۷۶۵,۹۰۰	۳۱۳,۹۵۰	۴,۰۹۱,۰۸۳
(Costs) C	واحد پولی- ریال	۲۵۰,۰۰۰,۰۰۰	۰	۱,۴۶۶,۵۷۲,۹

در گام بعدی با استفاده از فرمول ۱ بـه نرمـال‌سـازی داده‌ها می‌پردازیم.

۴-۳-آماده‌سازی و پیش‌پردازش داده‌ها

وزن‌دهی شاخص‌های مدل LRFMC (W):

براسـاس پاسـخی کـه پـنج نفـر از خبرگـان و صاحبنظران صنعت بیمه به پرسشنامهٔ ارائه شـده داده‌انـد، مقایسات زوجی را بـه منظـور تعیـین وزن شـاخص‌هـای مدل LRFMC در جدول زیر مشاهده می‌کنید:

جدول ۳. ماتریس مقایسه‌های زوجی

	L	R	F	M	C
L	۱	۳	۳	۴	۴
R		۱	۳	۳	۳
F			۱	۳	۴
M				۱	۶
C					۱

بر مبنای این ماتریس، در نرم افزار اکسپرت چویس وزن W هر شاخص تعیین می‌گردد.

Archive of SID

www.SID.ir

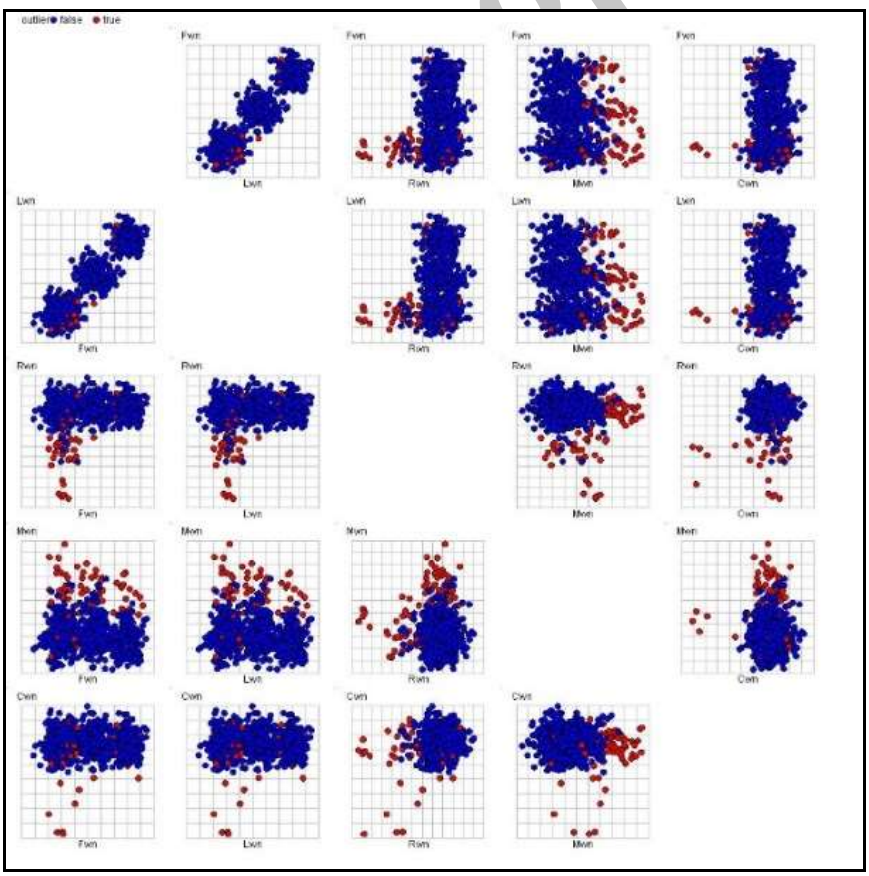
جدول ۴. وزن شاخص‌های مدل LRFMC

نماد وزن شاخص‌ها	مقادیر وزن نسبی شاخص‌ها
W_L	۰,۱۰۱
W_R	۰,۱۵۶
W_F	۰,۲۵۴
W_M	۰,۴۴۰
W_C	۰,۰۴۹
Sum	۱

با بهره‌گیری از وزن هر شاخص، ارزش هر شاخص با توجه به فرمول ۲ محاسبه می‌شود.

پاک‌سازی داده‌ها:

همان گونه که گفتیم با استفاده از برنامهٔ RapidMiner به شناسایی و حذف داده‌های پرت می‌پردازیم. در این فرآیند، حدود ۲۱۴ دادهٔ پرت شناسایی و حذف شد. در شکل زیر، ماتریس پراکندگی دادهٔ پرت را مشاهده می‌کنید. در این ماتریس داده‌های پرت با رنگ قرمز مشخص شده‌اند.



شکل۲. ماتریس پراکندگی داده‌ها پرت

www.SID.ir

۴-۴-مدل‌سازی

لازم به ذکر است که بر مبنای اصل اساسی حد مرکزی برای دست‌یابی به توزیع نرمال نتایج و نیز کاهش خطا در اجرای الگوریتم‌های خوشه‌بندی، هر الگوریتم ۳۰ بار اجرا شده است و ماتریس پراکندگی هر الگوریتم در شکل ۳ نشان داده شده است.

الگوریتم سلسله مراتبی Ward CA :

در جدول زیر بخشی از مراحل روش وارد را مشاهده می‌کنید:

جدول ۵. مراحل اجرای الگوریتم Ward

مجموعه سازی/Agglomeration Schedule						
مرحله	تشکیل خوشه برای اولین بار		ضرایب	ترکیب خوشه ها		مرحله
	Cluster 2	Cluster 1		Cluster 2	Cluster 1	
....	....	....	...	...	...	...
594	573	586	.031	200	7	590
593	583	587	.042	9	6	591
596	585	584	.055	3	1	592
595	558	591	.091	332	6	593
....	....	....	...	...	...	...

بر مبنای جدول بالا، مشخص می‌شود که بهترین مرحله برای توقف خوشه‌بندی، مرحلهٔ ۵۹۲ است. با توقف خوشه‌بندی در این مرحله، تعداد خوشهٔ بهینه جهت خوشه‌بندی برابر با چهار خوشه خواهد بود(۵۹۸-۵۹۲=۶)(حبیب پور و صفری، ۱۳۹۰). با توجه به این امر که الگوریتم‌های سلسله مراتبی، به دلیل کند بودن، برای داده‌هایی با ابعاد بالا مناسب نیست، نتایج اجرای این الگوریتم در بخش ارزیابی و مقایسهٔ خوشه در نظر گرفته نمی‌شود و تنها به عنوان مبنایی برای تعداد خوشه‌های بهینه در دیگر الگوریتم‌ها مورد استفاده قرار می‌گیرد.

بنابر نتایج مشاهده شده در الگوریتم DBSCAN – خوشه‌های حاصل: دو خوشه با ابعاد صفر و ۵۹۸ (کل داده‌ها)-که عدم توانایی آن در خوشه‌بندی درست داده‌های با چگالی و ابعاد بالا را نشان می‌دهد، این الگوریتم را در بخش مطالعهٔ مقایسه‌ای الگوریتم‌های خوشه‌بندی حذف می‌کنیم.

ما در این پژوهش برای بررسی و مقایسه بین الگوریتم‌های خوشه‌بندی، از دو پارامتر Silhouette و SSE استفاده می‌کنیم. در این جا به بیان تعریف مختصری از این پارامترها پرداخته، سپس روش برتر خوشه‌بندی را بر مبنای این پارامترها می‌یابیم.

پارامتر Silhouette:

شاخص Silhouette یکی از معیارهای متداول اعتبارسنجی در خوشه‌بندی است. این شاخص، معیار فواصل درون خوشه‌ای و برون خوشه‌ای را هم زمان در نظر می‌گیرد. Silhouette به عنوان یک روش گرافیکی تفسیری، به خوبی می‌تواند جایگاه هر چیز را در خوشه‌ها مشخص کند(پی راووس[1]، ۱۹۹۶). به بیان

1 P. Rouseeuw

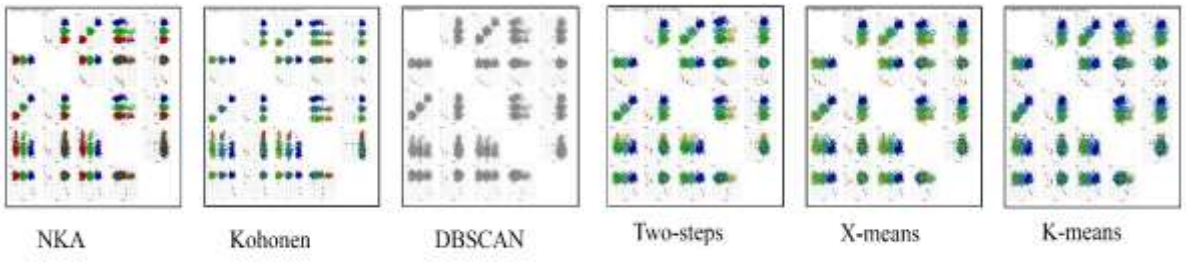
Archive of SID

www.SID.ir

ساده، مقدار میانگین شاخص .Sil، نشان دهندهٔ میزان تراکم در خوشه‌ها است.

پارامتر SSE (Sum of Squared Error)

SSE به عنوان یک پارامتر سنجش کیفیت الگوریتم خوشه بندی، در این تحقیق مورد استفاده قرار گرفته است. شاخص SSE مجموعهٔ مربعات فواصل بین داده‌ها در هر خوشه با میانگین آن خوشه است. به بیان ساده‌تر، این پارامتر مربعات فواصل درون خوشه‌ای از مرکز خوشه را حساب می‌کند و با توجه به تعریف این پارامتر، هر چقدر مقدار این شاخص کمتر باشد، تراکم بیشتری در خوشه‌ها حاصل شده است. اگر همهٔ نقاط درون خوشه، در مرکز خوشه قرار گیرد، این شاخص صفر می‌شود.



شکل ۳. ماتریس پراکندگی الگوریتم های خوشه بندی

با توجه به نتایج به دست آمده(نمودار ۴)، بهترین شیوهٔ خوشه‌بندی بر مبنای پارامتر Silhouette، روش خوشه ندی NKA است. میانگین پارامتر Silhouette در این الگوریتم از مابقی روش‌ها بهتر بوده است. ما در الگوریتم K-means مقدار Silhouette بهتری نیز مشاهده می‌کنیم، ولی این روش خوشه‌بندی، به دلیل واریانس بالا، ثبات مناسبی در خوشه‌بندی ندارد. بنا بر این نتیجه که الگوریتم Kohonen کمترین میزان SSE را دارد، می‌توان نتیجه گرفت که الگوریتم Kohonen متراکم‌ترین خوشه‌ها را ایجاد کرده است.

فاز ۵-ارزیابی و فاز ۶-تدوین

به منظور استخراج دانش از خوشه‌های حاصل از الگوریتم خوشه‌بندی، در راستای شناسایی رفتار مشتری و تشخیص عوامل مؤثر در سنجش ارزش مشتریان، از الگوریتم C5.0 در برنامهٔ SPSS modeler بهره می‌گیریم. این الگوریتم را به دلیل دقت بالاتر(۹۹/۹۸) نتایج ارزیابی آن نسبت به سایر الگوریتم‌های هم-کاربرد، انتخاب کرده‌ایم(تقوی فرد و خواجه وند، ۱۳۹۲).

نتایج الگوریتم NKA و Kohonen:

به منظور تحلیل نتایج خوشه‌بندی این الگوریتم، یک مرحله از نتایج این الگوریتم را به عنوان معیار نتایج خوشه‌بندی در نظر می‌گیریم[۱]. نتایج این دو الگوریتم را در ادامه مشاهده می‌نمایید.

به منظور تحلیل هر خوشه و تأثیر مقدار هر شاخص مدل LRFMC در تشکیل خوشه‌ها، ابتدا مقادیر میانگین هر شاخص مدل مذکور در هر خوشه با مقدار میانگین هر شاخص در کل داده‌ها که در جدول ۷و۸ ارائه شده است، مقایسه می‌شود و اگر متوسط ارزش آن شاخص در یک خوشه از متوسط ارزش آن شاخص در کل داده‌ها بیشتر باشد، با نماد ↑(نشان دهندهٔ وضعیت

۱. مرحله‌ای که بهترین مقدار Silhouette,SSE را دارد.

Archive of SID

www.SID.ir

مطلوب) و در صورتی کـه متوسـط ارزش شاخصـی در یک خوشه کمتـر از متوسـط ارزش آن در کـل داده‌هـا باشد، این وضعیت بـا علامـت ↓(نشـان دهنـده وضـعیت نامطلوب) نشان داده می‌شود .

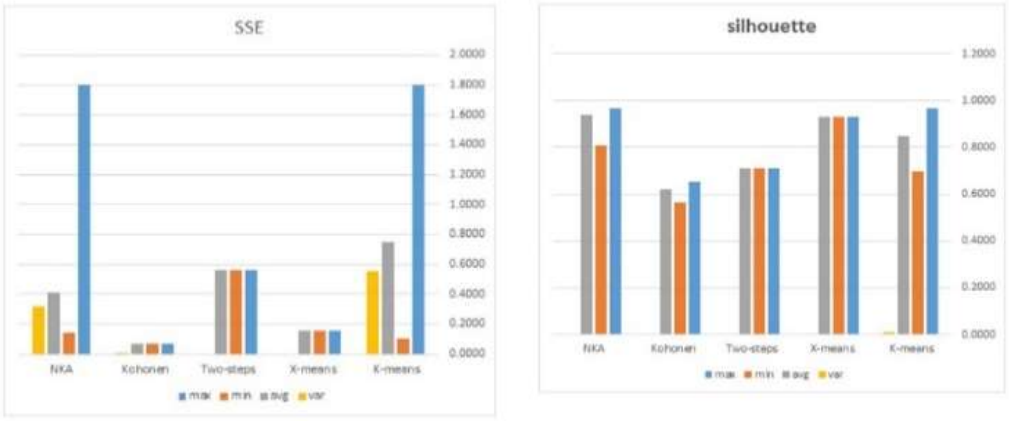
جدول ۷. NKA

عنوان خوشه	وضعیت متوسط ارزش هر شاخص در هر خوشه				
	L	R	F	M	C
خوشه یک	↑↑	↑	↑↑	↓	-
خوشه دو	↑	↑	↑	↑	-
خوشه سه	↓	↓	↓	↑	-

جدول۶. شاخص های SSE و Silhouette

SSE				clustering	No#
var	avg	min	max		
0.5510	0.7471	0.1002	1.7977	K-means	1
0.0000	0.1535	0.1535	0.1535	X-means	2
0.0000	0.5632	0.5632	0.5632	Two-steps	3
0.0000	0.0667	0.0644	0.0705	Kohonen	4
0.3187	0.4120	0.1434	1.7977	NKA	5

silhouette				clustering	No#
var	avg	min	max		
0.0096	0.8474	0.7001	0.9681	K-means	1
0.0000	0.9320	0.9320	0.9320	X-means	2
0.0000	0.7114	0.7114	0.7114	Two-steps	3
0.0013	0.6198	0.5632	0.6545	Kohonen	4
0.0031	0.9378	0.8061	0.9681	NKA	5



نمودار ۴. شاخص های SSE وSilhouette

www.SID.ir

جدول ۸. Kohonen

عنوان خوشه	میانگین شاخص های مدل های LRFMC				
	L	R	F	M	C
Y=0،X=0	↑	↑	↑	↓	-
Y=2،X=0	↑↑	↑	↑↑	↓	-
Y=0،X=1	↑	↑	↑	↑	-
Y=2،X=1	↑↑	↑	↑↑	↑	-
Y=2،X=2	↑	↑	↑	↑↑	-
Y=0،X=3	↓	↓	↓	↓	-
Y=1،X=3	↓	↑	↓	↑	-
Y=2،X=3	↓	↑	↓	↑↑	-

حال ارزش طول دورهٔ عمر مشتریان در هر شاخص را محاسبه می‌کنیم.

جدول ۹. ارزش دورهٔ عمر مشتری

عنوان خوشه	ارزش طول دورهٔ عمر مشتری
NKA	
خوشهٔ دو	۰,۴۰۶۱
خوشهٔ سه	۰,۲۲۸۱
Kohonen	
Y=0،X=0	0.3936
Y=2،X=0	0.5726
Y=0،X=1	0.4064
Y=2،X=1	0.5979
Y=2،X=2	0.4352
Y=0،X=3	0.2176
Y=1،X=3	0.2405
Y=2،X=3	0.2673

در این مرحله با استفاده از الگوریتم تصمیم‌گیری C5.0، قواعد حاکم در خوشه‌های به دست آمده را استخراج می‌کنیم.

قواعد تصمیم گیری در الگوریتم NKA:

Archive of SID

www.SID.ir

L <= 1 [Mode: cluster-3] => cluster-3
L > 1 [Mode: cluster-2]
L <= 2 [Mode: cluster-2] => cluster-2
L > 2 [Mode: cluster-1] => cluster-1

F <= 1 [Mode: cluster-3] => cluster-3
F > 1 [Mode: cluster-2]
F <= 2 [Mode: cluster-2] => cluster-2
F > 2 [Mode: cluster-1] => cluster-1

M <= 779،100 [Mode: cluster-1] => cluster-1

M > 779،100 [Mode: cluster-3] => cluster-3

قواعد تصمیم گیری در الگوریتم Kohonen:

L <= 1 [Mode: X=3، Y=0]

M <= 3،441،950 [Mode: X=3، Y=0] => X=3، Y=0

M > 3،441،950 [Mode: X=3، Y=1]

M <= 5،648،750 [Mode: X=3، Y=1] => X=3، Y=1

M > 5،648،750 [Mode: X=3، Y=2] => X=3، Y=2

L > 1 [Mode: X=0، Y=2]

L <= 2 [Mode: X=0، Y=0]

M <= 2،188،200 [Mode: X=0، Y=0] => X=0، Y=0

M > 2،188،200 [Mode: X=1، Y=0]

M <= 4،122،350 [Mode: X=1، Y=0] => X=1، Y=0

M > 4،122،350 [Mode: X=2، Y=2] => X=2، Y=2

L > 2 [Mode: X=0، Y=2]

M <= 3،853،100 [Mode: X=0، Y=2] => X=0، Y=2

M > 3،853،100 [Mode: X=1، Y=2] => X=1، Y=2

جدول ۱۰. NKA

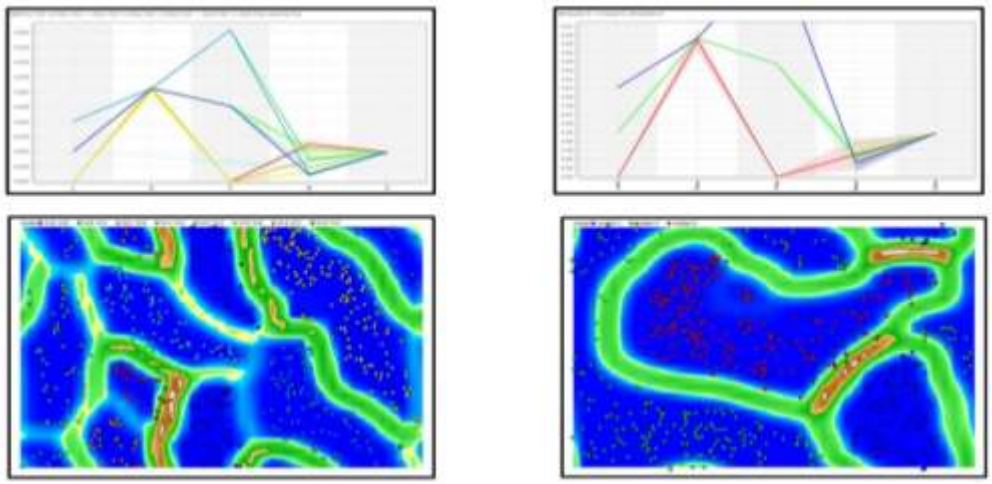
عنوان خوشه	ابعاد		میانگین شاخص های مدل های LRFMC				
			L	R	F	M	C
خوشه یک	۱۳۹	٪۲۳	۰،۱۰۱۰	۰،۱۵۶۰	۰،۲۵۴۰	۰،۰۱۵۳	۰،۰۴۹۰
خوشه دو	۲۲۴	٪۳۷	۰،۰۵۰۵	۰،۱۵۶۰	۰،۱۲۷۰	۰،۰۲۳۶	۰،۰۴۹۰
خوشه سه	۲۳۵	٪۳۹	۰	۰،۱۵۴۰	۰	۰،۰۲۵۱	۰،۰۴۹۰
Silhouette			Average	۰،۹۶۸۱	Var.	۰،۰۰۱۱	
SSE			۰،۱۸۹۷				

جدول ۱۱. Kohonen

Archive of SID

www.SID.ir

میانگین شاخص های مدل های LRFMC					ابعاد		عنوان خوشه
C	M	F	R	L			
0.0490	0.0111	0.1270	0.1560	0.0505	18%	۹۰۵	X=0,Y=0
0.0490	0.0126	0.2540	0.1560	0.1010	21%	۱۲۲	X=0,Y=2
0.0490	0.0239	0.1270	0.1560	0.0505	13%	۷۵	X=1,Y=0
0.0487	0.0382	0.2540	0.1560	0.1010	3%	۱۵	X=1,Y=2
0.0490	0.0527	0.1270	0.1560	0.0505	7%	۴۲	X=2,Y=2
0.0490	0.0155	0.0000	0.1531	0.0000	27%	۱۶۲	X=3,Y=0
0.0490	0.0355	0.0000	0.1560	0.0000	7%	۴۰	X=3,Y=1
0.0490	0.0623	0.0000	0.1560	0.0000	5%	۳۱	X=3,Y=2
	۰.۲۱۶	Var.	۰.۵۴۲۲	Average	Silhouette		
				۰.۰۶۲۲	SSE		



نمودار ۵. شاخص‌های الگوریتم NKA وKohonen - شکل ۶. ماتریس‌های SOM

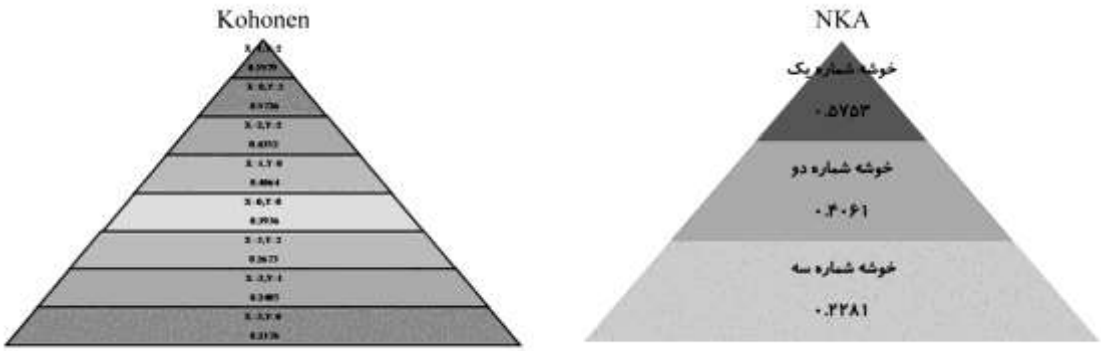
F <= 1 [Mode: X=3، Y=0] => X=3، Y=0

F > 1 [Mode: X=0، Y=2]

F <= 2 [Mode: X=0، Y=0] => X=0، Y=0

F > 2 [Mode: X=0، Y=2] => X=0، Y=2

واریانس شاخص‌های C و R پس از حذف داده‌های پرت، صفر است و در نتیجه در خوشه‌بندی تأثیر ندارد.



شکل ۷. هرم ارزش مشتری برای الگوریتم های NKA و Kohonen

Archive

www.SID.ir

فصلنامه علمی- پژوهشی تحقیقات بازاریابی نوین
سال ششم، شماره اول، شماره پیاپی (۲۰) بهار ۱۳۹۵
تاریخ دریافت: ۹۴/۱۱/۸ تاریخ پذیرش: ۹۵/۳/۳
صص: ۲۲-۱

۵-بخش‌بندی مشتریان براساس ارزش دورهٔ عمر آنها

پس از اینکه ارزش دورهٔ عمر مشتریان در هر خوشه تعیین می‌شود، با توجه به ارزش‌های تعیین شدهٔ مشتریان (قالب هرم ارزش دورهٔ عمر که در فصل دوم به توضیح آن پرداختیم)، به بخش‌بندی و رتبه‌بندی خوشه‌ها در هرم ارزش طول عمر مشتری می‌پردازیم و با توجه به ارزش طول عمر مشتریان هر خوشه که در مرحلهٔ قبل محاسبه می‌شود، رتبهٔ خوشه و وضعیت آنها را در هرم ارزش دورهٔ عمر مشتری تعیین می‌کنیم.

همان گونه که در شکل ۷ مشاهده می‌کنید، خوشه-های هرم ارزش دورهٔ عمر مشتری، با توجه به میزان ارزش دورهٔ عمر مشتریان، از پایین به بالا(کم به زیاد) مرتب شده‌اند. در رأس هرم، خوشه‌ای با بیشترین میزان ارزش دورهٔ عمر مشتریان قرار دارد.

۵-نتایج پژوهش

در راستای پاسخ به سؤالات مطرح شده در این پژوهش، ابتدا به بررسی و سپس مقایسهٔ مدل‌های سنجش ارزش مشتری پرداخته‌ایم. در این میان مدل RFM، به دلیل برتری‌هایی که نسبت به روش‌های دیگر مدل‌های سنجش ارزش طول عمر مشتری دارد، انتخاب شده است. برای افزایش بازده این مدل در صنعت بیمه، مدل را با اضافه کردن المان‌های دیگر بهبود و توسعه داده‌ایم(LRFMC) و سپس با استفاده از مقایسات زوجی سلسله مراتبی گروهی AHP وزن‌های هر شاخص مدل مذکور را با توجه به صنعت بیمهٔ مورد مطالعه، استخراج کرده‌ایم. با بررسی اجمالی بین الگوریتم‌های مختلف خوشه‌بندی و بررسی مقالات و تحقیقات انجام شده در این زمینه، هفت الگوریتم مختلف خوشه‌بندی را انتخاب کردیم و به بررسی، مطالعه و مقایسهٔ نتایج حاصل از این هفت الگوریتم خوشه‌بندی با استفاده از داده‌های استخراجی از حوزه بیمهٔ بدنه در صنعت بیمه پرداختیم.

نتیجهٔ اصلی حاصل از این پژوهش ارائهٔ فرآیندی برای سنجش ارزش مشتری در صنعت بیمه است. به طور کلی نتایج حاصل از فعالیت‌های انجام شده را، برای تکمیل این پژوهش، به صورت زیر می وان بیان نمود:

- معرفی و جمع‌آوری مفاهیم مرتبط با خوشه‌بندی.
- تحلیل و بررسی روش‌ها و الگوریتم‌های خوشه-بندی مطرح در دنیا.
- بررسی مدل‌های مختلف سنجش ارزش مشتری، توسعه و گسترش مدل مذکور در راستای سنجش ارزش مشتری در صنعت بیمه.
- مطالعه و بررسی تحقیقات انجام گرفته در زمینهٔ خوشه‌بندی با توجه به پارامترهای مؤثر بر ارزش طول عمر مشتری.

* نویسنده مسؤول
y_taghizadeh2003@yahoo.com

Archive of SID
www.SID.ir

- ارائهٔ چارچوبی پیشنهادی در راستای سنجش ارزش مشتری در صنعت بیمه.

- خوشه‌بندی مشتریان شرکت بیمه(حوزهٔ بیمهٔ بدنهٔ اتومبیل) براساس فاکتورهای مدل LRFMC به منظور سنجش ارزش طول عمر مشتریان.

- مثمرثمر بودن این تحقیق در تبیین استراتژی‌های بازاریابی و تدوین برنامه‌های سیستم مدیریت ارتباط با مشتری براساس تفکیک مشتریان به گروه‌های مختلف بر مبنای ارزشی که مشتری برای سازمان دارد.

علاوه بر این، با مقایسه بین هفت الگوریتم خوشه-بندی، می‌توان گفت که عملکرد الگوریتم‌های خوشه-بندی با توجه به داده‌ها متفاوت است و با استفاده از دو پارامتر Silhouette و SSE الگوریتم‌های NKA و Kohonenنتایجی بهینه داشته است. ثبات در بدست آوردن نتایج بهینه، یکی از نکات دیگری است که در این مطالعهٔ مقایسه‌ای مدنظر قرار گرفته است، به همین دلیل الگوریتم K-means با وجود جواب‌های بهینه، با اجرای مکرر در حالت‌های بهینهٔ محلی به دام افتاده و جواب‌های متفاوتی را در هر مرحله به دست داده است.

۵-۲-پیشنهادها

این پژوهش به بررسی خوشه‌بندی مشتریان صنعت بیمه(حوزهٔ بیمهٔ بدنهٔ اتومبیل) با تمرکز بر ارزش طول عمر مشتری پرداخته است. با توجه به شرایط و محدودیت‌های پژوهش مبتنی بر دریافت داده‌ها از سوی یک سازمان بیمه، سعی بر این بود که بهترین روش برای تأمین اهداف اولیهٔ پژوهش ارائه شود. این موضوع را می‌توان در زمینه‌های دیگر گسترش داد و به نتایج مفیدی برای این دانش رسید.

با توجه به چارچوب و فرآیند پژوهش حاضر، پیشنهادهای زیر ارائه می‌شود:

- می‌توان موضوع این پژوهش را در مورد دیگر حوزه‌های بیمه(مانند بیمه‌های شخص ثالث، بیمه‌های درمانی، بیمهٔ حوادث طبیعی و..) مورد تحلیل و بررسی قرار داد.

- خوشه‌بندی یکی از تکنیک‌های پایهٔ داده‌کاوی است و نتایج آن مبنای پژوهش‌های دیگر قرار خواهد گرفت. به طور کلی هدف از خوشه بندی، شناسایی نیازها و رفتارهای مشتریان است. بنابراین تشخیص و تفسیر دقیق خوشه‌ها اهمیت بالایی دارد. در پژوهشهای آینده می‌توان با استفاده از تکنیک‌های دیگر داده-کاوی، ارزش آتی خوشه‌های مشتریان را محاسبه کرد و با استفاده از تکنیک‌های دیگر درخت‌های تصمیم-گیری، به استخراج دانش و الگوهای پنهان در رابطه با مشتریان پرداخت.

- هدف از این پژوهش، بخش‌بندی مشتریان در راستای سنجش ارزش آنها است. اعتبار سنجی و حتی ریسک‌سنجی در صنعت بیمه موضوع دیگری است که می‌توان مورد بررسی قرار داد.

- با استفاده از الگوریتم‌های دیگر خوشه‌بندی و نیز مدل‌های دیگر سنجش ارزش مشتری که در فصل دو به اختصار آنها را بیان کردیم، می‌شود چارچوبی نوین برای خوشه‌بندی مشتریان برمبنای ارزش مشتری ارائه و پیشنهاد کرد.

- مطالعهٔ شرکت‌های خصوصی و نوپا در حوزهٔ بیمه، زمینهٔ دیگری برای پژوهش است که علاوه بر بررسی عملکرد شرکت مزبور در بازار کسب و کار، می‌تواند به کسب دیدگاهی کلی نسبت به عمکرد شرکت‌های خصوصی بیمه بیانجامد.

- بررسی مشتریان در بازه‌های زمانی طولانی‌تر نتایج کارآمدتر و گسترده‌تری در شناسایی ارزش مشتریان به

Archive of SID

www.SID.ir

دست می‌دهد و تأثیر اقتصاد و عوامل بیرونی را بر این نتایج کمرنگ‌تر می‌کند.

فهرست منابع

۱- برادران، وحید و بیگلری، محمد. (۱۳۹۴). بخش بندی مشتریان صنایع تولید و بخش کالای پرگردش براساس مدل بهبود یافته RFM (مطالعه موردی : شرکت گلستان). در مدیریت بازرگانی، ۷: ۲۳-۴۲.

۲- پورمدرس سرچشمه، موسی و افشارکاظمی، محمد. (۱۳۹۴). بکارگیری الگوریتم Kohonen در خوشه‌بندی تأمین کنندگان خودرو ساز. در همایش ملی مدیریت و آموزش. ملایر: دانشگاه ملایر.

۳- تارخ، محمد جعفر و شریفیان، کبری. (۱۳۸۹). کاربرد داده‌کاوی در بهبود مدیریت ارتباط با مشتری. در فصلنامهٔ علمی -پژوهشی مطالعات مدیریت، ۶: ۱۵۳-۱۸۱.

۴- ترکستانی، محمد صالح و ده پناه، آرمان. (۱۳۹۱). تعیین خسارت بالقوهٔ بیمه‌گذاران در رشتهٔ بدنهٔ اتومبیل با استفاده از مدل شبکه‌های عصبی.

۵- ترکستانی، محمد صالح و دیگران. (۱۳۹۳). مفاهیم و اصول بیمه. تهران: سادس.

۶- تقی تقوی فرد، محمد و خواجوند، سمانه. (۱۳۹۲). بخش بندی خوشه‌ای مشتریان بانک(مورد مطالعه بانک صادرات ایران). در فصلنامهٔ علمی- پژوهشی کاوش‌های مدیریت بازرگانی، ۳۹-۶۴.

۷- حری، هومن. (۱۳۸۹). داده‌کاوی در مدیریت ارتباط با مشتری. پایان نامهٔ کارشناسی ارشد: دانشگاه آزاد اسلامی.

۸- حسینی، فرشته. (۱۳۹۳). روش تشخیص داده‌های پرت Outlier detection . در همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات. تهران: شرکت علم و صنعت طلوع فرزین. http://www.civilica.com/Paper-CSITM01-CSITM01_237.html.

۹- خدپور، آمنه؛ رزمی، زهرا و حامدی، پروشات. (پاییز ۱۳۹۲). خوشه‌بندی مشتریان به منظور اصلاح استراتژی قیمت‌گذاری و تدوین استراتژی تخفیف‌دهی(مطالعهٔ موردی: شرکت کدبانو). در فصلنامهٔ علمی و پژوهشی تحقیقات بازاریابی نوین، ۳ : ۱۳۵-۱۵۰ .

۱۰- دانش، ملیحه؛ یغمایی مقدم، محمدحسین و اکبرزاده توتونچی، محمدرضا. (۱۳۸۹). خوشه-بندی داده با استفاده از ترکیب K-harmonic means و PS. درکنفرانس مهندسی برق ایران. اصفهان: دانشگاه صنعتی اصفهان: ۲۱۳۳-۲۱۳۸ . www.ICEE2010.ir.

۱۱- زینب ارزانی، معصومه. (۱۳۹۱). بررسی ارزش طول عمر مشتریان با استفاده از روش داده-کاوی. پایان نامهٔ کارشناسی ارشد: دانشگاه صنعتی خواجه نصیرالدین طوسی.

۱۲- صداقتی، آزاده. (۱۳۹۱). مدیریت ارتباط با مشتری و داده کاوی. تهران: مگستان.

۱۳- صنیعی آباده، محمد؛ محمودی، سینا و طاهرپرور، محدثه. (۱۳۹۳). داده‌کاوی کاربردی. ج ۲. تهران: نیاز دانش.

۱۴- غضنفری، مهدی؛ ملک محمدی، سمیرا؛ علیزاده، سمیه و فتح اله زاده، مهدی. (۱۳۸۹). بخش‌بندی مشتریان صادراتی میوه‌های خوراکی. در فصلنامهٔ پژوهشنامهٔ بازرگانی، ۵۵: ۱۵۱-۱۸۱.

Archive of SID

www.SID.ir

۱۵- قره خانی، محسن و مریم ابوالقاسمی. (۱۳۹۱). کاربرد داده‌کاوی در صنعت بیمه. در تازه‌های جهان بیمه، ۱۵۸: ۵-۲۱.

۱۶- قره نژاده، سحر. لزوم حفظ مشتریان بیمه با استفاده از ابزار داده‌کاوی. در تازه‌های جهان بیمه، ۱۵۰-۱۵۱: ۱۵-۲۳.

۱۷- قیاسی، راضیه؛ باقری دهنوی، ملیحه و مینایی، بهروز. (۱۳۹۰). ارائهٔ مدلی جهت اندازه‌گیری میزان وفاداری و ارزش مشتریان با استفاده از تکنیک RFM و الگوریتم‌های خوشه‌بندی. در دومین کنفرانس ملی محاسبات نرم و فناوری اطلاعات (NCSCIT02_107). http://www.civilica.com/Paper-NCSCIT02-NCSCIT02_107.html.

۱۸- الله یاری و فتحعلی زاده. (۱۳۹۰). مدلی برای سنجش ارزش طول عمر مشتری. فصلنامهٔ پژوهشکدهٔ بیمهٔ ایران .

۱۹- هان، ژیاوی؛ کمبر، میشیلین و پی، ژان. (۱۳۹۳). داده‌کاوی- مفاهیم و تکنیک‌ها. ترجمهٔ مهدی اسماعیلی. تهران: نیاز دانش.

۲۰- هرابی، بابک و امیر خانلری. (۱۳۹۱). اندازه‌گیری ارزش طول عمر مشتری بر مبنای مدل RFM. در پژوهش‌های مدیریت عمومی، ۵ (۱۵): ۶۳-۸۴.

۲۱- بین، شرکت بیمه. (۱۳۹۴). «نقش داده‌کاوی در صنعت بیمه و مطالعهٔ موردی با استفاده از تکنیک طبقه‌بندی». در صنعت بیمه در رسانه‌ها: ۵-۱۱.< http://sarmadins.ir>

22- Ahmed, M.; El-Zehery, Hazem M. Elbakry, Mohamed S. El-Ksasy & Nikos Mastorakis. (2013). Applying Data Mining Techniques for Customer Relationship Management: A Survey. Recent Advances in Computer Engineering, Communications and Information Technology, p. 398-406.

23- Almana، Amal M, Mohmet Saloih Aksoy & Rasheed Alzahroni. (2014). A Survey on Data Mining Techniques in Customer Churn Analysis for Telecom Industry. Journal of Engineering Research and Applications, 4 (2248-9622): 165-171. www.Ijera.Com

24- Ansari, Azarnoosh & Shermineh Ghalamkari. (2014). Segmenting Online Customers Based On Their Lifetime Value And Rfm Model By Data Mining Techniques. International Journal of Information Science & Managemen, p. 69-82.

25- Brawner, Austin. (2014). Determining Customer Lifetime Value: How Fewer Customers Can Mean More Profit. The Ecommerce Influence. May 9. http://www.Ecommerceinfluence.Com/Determining-Customer-Lifetime-Value-How-Fewer-Customers-Can-Mean-More-Profit/

26- Dhandayudam, Prabha & Ilango Krishnamurthi. (2012). Customer Behavior Analysis Using Rough Set Approach. Journal of Theoretical & Applied Electronic Commerce Research 8 (2): 21-33. Doi: 10.4067/S0718-1.

27- Du, Xiaoshan. (2006). Data Mining Analysis and Modeling For Marketing Based On Attributes Of Customer Relationship. School of Mathematics and Systems Engineering، Växjö University.

28- El Seddawy, Ahmed Bahgat, Dr. Ramadan Moawad, & Dr. Maha Attia Hana. (2005). "Applying Data Mining Techniques in Crm". UN Publish 11.

29- Ellaban, Mahmoud Ayesh Abu. (2013). The Role of Data Mining Technology in Building Marketing and Customer Relationship Management (Crm) for Telecommunication Industry (Case Study: Jawwal Mobile Operator – Gaza Strip). A Thesis for Master Degree: Islamic University-Gaza.

30- Fader, P. S. Hardie, B. G. & Lee, K. L. (2005). Rfm and Clv: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research, 42(4): 415-430.

Archive of SID

www.SID.ir

Real Time Clustering. International Journal of Data Warehousing and Mining, p. 1-14.

40- Manu, Chopra. (2012). Analysis of Clustering Technique For Crm. International Journal of Engineering And Managment Sciences 3 (2229-600x): 402-408.

41- Nejad, Mohammad Behrouzi; Ebrahim Behrouzian Nejad & Ali Karami. (2012). Using Data Mining Techniques to Increase Efficiency of Customer Relationship Management Process. Research Journal of Applied Sciences, Engineering and Technology, 4: 5010-5015.

42- Ngai, E.W.T.; Li Xiu, & D.C.K. Chau. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. Expert Systems with Applications, p. 2592-2602. Www.Elsevier.Com/Locate/Eswa.

43- Ogbi, Masoud & Hussein Farisat. (2014). Survey The Necessities Of Data Mining In The Banking System And Crm. Ndian Journal of Fundamental & Applied Life Sciences, 4: 127-131. Http://Www.Cibtech.Org/Sp.Ed/Jls/2014/01/Jls.Htm.

44- Oliveira, Vera L´Ucia Migu´Eis. (2012). Analytical Customer Relationship Management in Retailing Supported By Data Mining Techniques. Thesis For Doctoral Degree: Faculdade De Engenharia Da Universidade Do Porto.

45- Rajagopal, Sankar. (2011). Customer Data Clustering Using Data Mining Technique. International Journal of Database Management Systems, 4: 1-11. Doi:10.5121/Ijdms.2011.3401.

46- Rashidi, Aydin. (2010). Customer Relationship Management and Its Use in Insurance Industry. UN Publish (Mba from University Of Tehran), 18.

47- Rousseeuw, Peter J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics, p. 53-65. Doi: 10.1016/0377-0427(87)90125-7.

48- Rygielski, Chris; Jyun-Cheng Wang & David C. Yen. (2002). Data Mining Techniques for Customer Relationship

31- Fader, P. S. Hardie, B. G. & Lee K. L. (2005). Rfm and Clv: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research, p. 415-430.

32- Fluegemann, Joseph K. Misty D. Davies & Nathan D. Aguirre. (2011). Determining the Optimal Number of Clusters with the Clustergram. New Mexico: NASA Usrp.

33- Hosseini, Seyed Mohammad Seyed, Anahita Maleki & Mohammad Reza Gholamian. (2010). Cluster Analysis Using Data Mining Approach to Develop Crm Methodology to Assess The Customer Loyalty. Expert Systems with Applications (0957-4174/$): 5259–5264. www.Elsevier.Com/Locate/Eswa.

34- Hsu, Spring C. (2012). The Rfm-Based Institutional Customers Clustering: Case Study of a Digital Content Provider. Information Technology Journal 11 (9): 1193-1201. Doi:10.3923/Itj.2012.1193.1201.

35- Hu, Wang & Zhang Jing. (2008). Study of Segmentation for Auto Services Companies Based On Rfm Model. Pontifícia Universidade Católica De São Paulo. http://www.Pucsp.Br/Icim/Ingles/Downloads/Pdf_Procceeding_2008/66. Pdf

36- Hu, Y-H. & T-W. Yeh. (2014). Discovering Valuable Frequent Patterns Based On Rfm Analysis Without Customer Identification Information. Knowledge-Based Systems 76-88. Http://Www.Sciencedirect.Com/Science/Journal/09507051.

37- Janakiraman, S. & K. Umamaheswari. (2014). A Survey on Data Mining Techniques for Customer Relationship Management. International Journal of Engineering, Business and Enterprise Applications (2279-0039): 55-61.

38- Maheswari, R. Uma, S. Saravana Mahesan, Dr. Tamilarasan & A. K. Subramani. (2014). Role of Data Mining In Crm. International Journal of Engineering Research 3 (2): 75-78.

39- Mansouri, Taha; Ahad Zare Ravasan; & Mohammad Reza Gholamian. (2014). A Novel Hybrid Algorithm Based On K-Means and Evolutionary Computations for

Archive of SID

www.SID.ir

Management. Technology in Society 483–502. Www.Elsevier.Com/Locate/Techsoc.

49- Seyed Hosseini, Mohammad; Anahita Maleki & Mohammad Reza Gholamian. (2010). Cluster Analysis Using Data Mining Approach To. Expert Systems with Applications, 37: 5259-5264.

50- Shearer C. The Crisp-Dm Model: The New Blueprint For Data Mining, J Data Warehousing (2000); 5:13—22

51- Soeini, Reza Allahyari & Ebrahim Fathalizade. (2012). Customer Segmentation Based On Modified Rfm Model in the Insurance Industry. Proceedings of 2012 4th International Conference On Machine Learning And Computing, p. 101-104.

52- Sohrabi, Babak; & Amir Khanlari. (2007). Customer Lifetime Value (Clv) Measurement Based On Rfm Model. Iranian Accounting & Auditing Review, 14: 7-20.

53- Tan, Pang-Ning; Michael Steinbach & Vipin Kumar. (2006). Cluster Analysis: Basic Concepts and Algorithms. In Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach and Vipin Kumar, p. 487-568. Minnesota: Addison-Wesley Companion Book Site.

54- Tsiptsis, Konstantinos; & Antonios Chorianopoulos. (2009). Data Mining Techniques in Crm: Inside Customer Segmentation. Wiley.

55- Turk, Mohammad Amir. (2007). Data Mining and Mobile Crm: An Update. A Thesis for Master Degree: Blekinge Institute of Technology.

56- Woo, J. Y, S. M. Bae, And S. C. Park. (2009). Data Ming and Improve Value of Customers. Expert Systems with Application, p. 763–772.

57- Wu, H. E. Chang & Ch Lo. (2009). Applying Rfm Model and K-Means Method in Customer Value Analysis of an Outfitter. International Conference on Concurrent Engineering. New York.

Archive of SID

www.SID.ir