

شناسایی و تحلیل الگوهای رفتاری مشتریان خدمات بانکی با استفاده از تکنیکهای داده کاوی

محمدپریدی ۱

تاریخ دریافت: ۱۳۹۹/۷/۲۰

حسن صابری ۲

تاریخ پذیرش: ۱۳۹۹/۹/۲

زین العابدین امینی ۳

احسان ساده ۴

چکیده

شناسایی و تحلیل الگوهای رفتاری مشتریان یکی از چالشهای اساسی بانکها در تعیین مشتریان هدف، مدیریت ارتباط با مشتریان و تخصیص بهینه منابع محدود بانکها به مشتریان می باشد. لذا با توجه به حجم عظیم اطلاعات مشتریان در پایگاه اطلاعات بانکها، تعدد و تنوع خدمات بانکی و همچنین ویژگیهای فتاری خاص مشتریان در این پژوهش از تکنیکهای داده کاوی جهت ارائه یک الگوی مناسب برای شناسایی، بخش بندی، رتبه بندی و تحلیل الگوهای رفتاری مشتریان استفاده گردید، لذا داده های تراکنشی مشتریان بانک رفاه کارگران به عنوان آماری در نظر گرفته شد و تراکنش هایی که از اول مهر ۱۳۹۸ تا ۲۹ اسفند ۱۳۹۸ انجام گرفته بود به عنوان نمونه انتخاب گردید. سپس عملیات پیش پردازش بر روی آنها انجام و مدلسازی داده ها با هدف بخش بندی و رتبه بندی مشتریان به وسیله الگوریتم های K-means و SOM انجام گردید و در نهایت به وسیله الگوریتم Apriori قوانین حاکم بر روابط بین خوشه های مشتریان و نوع حساب ها کشف گردید. از آنجا که تعیین ارزش هریک بخش های مشتریان اهمیت بسزایی دارد لذا از معیار WRFM بهینه شده مدل (RFM) جهت محاسبه و تعیین ارزش طول عمر مشتریان (CLV) استفاده و براساس ارزش تعیین شده، استراتژیهای مناسب برای ارائه خدمات بانکی به هریک از بخشهای موصوف تعیین گردید.

کلمات کلیدی: بخش بندی، رتبه بندی، هوش مصنوعی، یادگیری ماشین، K-means، SOM، WRFM، Apriori

۱- دانشجوی دکتری گروه مدیریت دانشکده علوم انسانی واحد ساوه، دانشگاه آزاد اسلامی ساوه. ایران. mo.paridari@yahoo.com

۲- گروه مدیریت دانشکده علوم انسانی واحد ساوه، دانشگاه آزاد اسلامی ساوه. ایران.

۳- گروه مدیریت دانشکده علوم انسانی واحد ساوه، دانشگاه آزاد اسلامی ساوه. ایران. drsajadami@yahoo.com

۴- گروه مدیریت دانشکده علوم انسانی واحد ساوه، دانشگاه آزاد اسلامی ساوه. ایران. ehsansadeh@gmail.com

امروزه شناسایی و رتبه بندی مشتریان یکی از چالشهای اساسی بانکها در امر بخش بندی بازار، تعیین مشتریان هدف و مدیریت ارتباط با مشتریان می باشد (شهرابی، ۱۳۹۰). تعداد بسیار زیاد مشتریان و ویژگی های رفتاری خاص هر یک از ایشان، تعدد و تنوع محصولات و خدمات بانکی، عدم جذابیت و کارایی بازاریابی انبوه و در مقابل هزینه های بسیار زیاد اجرای بازاریابی تک به تک و همچنین عوامل محیطی بیرونی تاثیر گذار بر عملکرد بانکها از قبیل رقابت شدید در افزایش سهم بازار بانکهای خصوصی و موسسات مالی و اعتباری، رشد سریع تکنولوژی و بانکداری الکترونیکی و ... از مهمترین عوامل تشدید این چالش بوده و عدم توجه به این مهم مخاطرات جدی همچون خروج مشتریان ارزشمند، افزایش هزینه برای جذب مشتریان جدید و همچنین الزام به حفظ مشتریان قدیمی و ... برای بانکها در بر خواهد داشت که پیشگیری از این امر مستلزم داشتن پایگاه اطلاعاتی قوی و منسجم و استفاده از سیستم های پشتیبان مدیریت جهت بهینه نمودن تصمیم گیریهای مرتبط با مشتریان یا به عبارت بهتر مدیریت ارتباط با مشتری می باشد (Edelstein, ۲۰۰۳; Smit, ۲۰۰۶; He, ۲۰۱۴, Shi, Wan, & Zhao, ۲۰۱۴) و (شهرابی، ۱۳۹۰; سلیمانی بشلی و اکبری اصل (۱۳۹۰)، تکنیک هایی از قبیل (خوشه بندی) خوشه بندی سلسله مراتبی مبتنی بر مدل و مبتنی بر بخش بندی (، دسته بندی و ...، با استفاده از داده کاوی، تحلیل داده ها و بخش بندی مشتریان می توان به اطلاعات با ارزشی از قبیل شناسایی نیازهای مشتری دست یافت). سلیمانی بشلی و اکبری اصل (۱۳۹۰)، لذا استفاده از تکنیک های داده کاوی جهت بخش بندی و شناسایی الگوهای رفتاری مشتریان برای بانکها امری ضروری و حیاتی است). کاتلر و ۱۳۹۶؛ سلیمانی بشلی و اکبری اصل (۱۳۹۰)،

هر بخش از مشتریان سود متفاوتی برای بانک ایجاد می نمایند لذا تعیین ارزش حال و آینده هر بخش از مشتریان ارزش بسزایی دارد بدیهی است هر بخش از مشتریان که دارای ارزش بالاتری باشند بایستی از بهترین سطح خدمات و مزایای ویژه استفاده نمایند تا باعث افزایش وفاداری و کاهش تمایل ایشان به روی آوردن به بانکهای رقیب گردد (سلیمانی بشلی و اکبری اصل (۱۳۹۰)، مهمترین معیاری که برای محاسبه

ارزش طول عمر مشتری می توان استفاده نمود معیار RFM می باشد. مدل WRFM مدل بهینه شده ی مدل RFM می باشد که W معرف وزن برای هر یک شاخص های مذکور می باشد (Lemon & Mark, ۲۰۰۶) پس از تعیین ارزش طول عمر هر بخش از مشتریان می توان نسبت به تعیین اولویت و رتبه هر یک اقدام نمود تا بر اساس میزان اهمیت و رتبه هر بخش از مشتریان خدمات متناسب با آن بخش ارائه گردد. (مصلحی, کفاش پور, و ناجی عظیمی (۱۳۹۳),

با توجه به فقدان الگوی مناسب جهت شناسایی و رتبه بندی مشتریان در بانک و همچنین حجم انبوه اطلاعات مشتریان در پایگاه اطلاعات بانک و به منظور طراحی و ارائه یک الگوی مناسب جهت شناسایی و رتبه بندی مشتریان بانک، این پژوهش مبادرت به استفاده از تکنیک های داده کاوی جهت بخش بندی، رتبه بندی و شناسایی الگوهای رفتاری مشتریان برای بانک ها نموده است.

• مروری بر چارچوب نظری پژوهش

۱-۲- بخش بندی

بخش بندی بندی بازار که فرایند تعریف و تقسیم یک بازار همگن بزرگ به بخش های واضح و مشخص که دارای نیازها و خواسته های مشابه است، بخش مهمی از استراتژی کسب و کار می باشد که اگر به درستی و به طور مناسب اجرا شود، از ارزش کلی شرکت پشتیبانی می کند و تاثیر مثبتی بر فروش، سود و سهم بازار دارد. (Kannisto, ۲۰۱۶) تقسیم بندی بازار یک استراتژی است که شامل تقسیم یک بازار بزرگ به بخش هایی می باشد که در آن مشتریان دارای نیازها و برنامه های کاربردی مشترک برای محصولات و خدمات ارائه شده در بازار هستند. این بخش ها بسته به ترکیب هر گروه می توانند با تعدادی از ویژگی های مختلف شناسایی شوند. یکی از دلایل اصلی ایجاد بخش های مختلف بازار، کسب اطلاعات بیشتر در مورد مشتریان و به کار بردن راهبردهای مختلف برای دستیابی به رضایت بیشتر مشتری و افزایش سود است. مشتریان هر بخش باید تا حد ممکن به یکدیگر شبیه باشند. این ویژگی به شرکت اجازه می دهد که استراتژی های مختلف بازاریابی

محصول را برای برآورده ساختن نیازهای مشتری و کسب سود بیشتر به کار بندند (Kieu, Ou, & Cai, ۲۰۱۸)؛ Huerta-Muñoz, Ríos-Mercado, & Ruiz, ۲۰۱۷) می توان به عنوان یک تکنیک تعریف کرد که در آن، گروه ها ویژگی های مشابهی دارند (Y. Liu, Kiang, & Brusco, ۲۰۱۲).

۲-۲ خوشه بندی

خوشه بندی به معنای تقسیم داده ها به گروه های مشابه است. داده ها بر اساس اصل حداکثر کردن شباهت داخل گروه ها و حداقل کردن شباهت بین گروه ها، خوشه بندی می شوند. خوشه بندی یک روش متداول توصیفی است که در جستجوی تشخیص تعداد محدودی خوشه برای توصیف داده ها است (KAUFMAN, ۱۹۹۰) & ROUSSEUW, تجزیه و تحلیل خوشه یا خوشه بندی، زمینه تحقیقاتی است که با گروهی از اشیا یا اهداف به نام خوشه ها یا بخش ها تعریف می شود. به گونه ای که اعضای یک خوشه شبیه هم و متفاوت از اعضای سایر خوشه ها باشند. تکنیک های خوشه بندی در بسیاری از زمینه های تحقیقی، مانند یادگیری ماشین، آمار، بیوانفورماتیک و بازاریابی) به عنوان مثال، تقسیم بندی بازار (افزایش یافته است (Trindade, Dias, & Ambrósio, ۲۰۱۷). تکنولوژی خوشه بندی به طور گسترده ای در بسیاری از زمینه ها مورد استفاده قرار می گیرد. در زمینه تجاری، می توان از آن برای تجزیه و تحلیل رفتار مشتریان و ارائه یک مبنای مهم برای توسعه استراتژی بازاریابی تجاری استفاده کرد. علاوه بر این، در زمینه تجارت الکترونیک، می توان آن را برای تجزیه و تحلیل ویژگی های مشتری های مشابه، برای ارائه خدمات بهتر به مشتریان به کار برد (Zhang, Zhang, & Zhang, ۲۰۱۸).

الگوریتم K-means معمول ترین روش خوشه بندی ساده است و برای تعداد زیادی از داده های عددی با ابعاد بزرگ به کار می رود و یک روش کارآمد برای طبقه بندی داده های مشابه به یک خوشه مشابه ارائه می دهد (Yu, Chu, Wang, Chan, & Chang, ۲۰۱۸). همچنین نقشه های خودسازمانده یا به اختصار SOM اولین بار توسط کوهنن معرفی گردید. این شبکه ها بر اساس یادگیری رقابتی عمل می کنند و دارای یادگیری بدون ناظر هستند به این مفهوم که نیازی به داشتن خروجی مورد انتظار در طی فرآیند آموزش نیست.

(Chuang, Wei, Zhifu, & Zhi, ۲۰۱۷) ورودی این شبکه ها مجموعه ای از داده های آموزش است که معمولاً با هدف خوشه بندی یا دسته بندی به شبکه اعمال می شوند (López, Valero, Senabre, Aparicio, & Gabaldon, ۲۰۱۲).

Aparicio, & Gabaldon, ۲۰۱۲)

۳-۲ قوانین انجمنی

هدف اصلی داده کاوی در پیدا کردن وابستگی، یافتن قانونهای محکم و قابل توجه است. قوانین وابستگی یا قوانین انجمنی اولین بار در سال ۱۹۹۳ برای مقادیر صفر و یک توسط آگراوال و همکاران مطرح شد و در سال ۲۰۰۲ منتشر گردید (Aggarwal, Procopiuc, & Yu, ۲۰۰۲). الگوریتم Apriori در سال ۱۹۹۶ توسط Cheung ابداع شد. این الگوریتم معروف ترین الگوریتم قوانین انجمنی می باشد و از یک روش تکراری برای یافتن مجموعه اشیاء مکرر استفاده می کند. نکته بسیار جالبی که می بایست در اینجا به آن اشاره نماییم آن است که بسیاری از الگوریتمهای شناسایی الگو که برای داده کاوی استفاده می شوند، در علم یادگیری ماشین توسعه یافته اند. اما الگوریتم Apriori که برای استخراج مجموعه اشیاء مکرر و البته قوانین انجمنی به کار می رود ماهیتاً یک الگوریتم داده کاوی است. (صنعی آباده و همکاران، ۱۳۹۳)

۴-۲ مدل RFM برای محاسبه ارزش عمر مشتریان

روش RFM، که توسط Hughes پیشنهاد شده (Hughes, ۲۰۱۲) یک روش موثر برای تقسیم بندی بازار در بازاریابی است که مخاطبان را از طریق تجزیه و تحلیل طول مدت یک دوره از زمان آخرین خرید (R)، تعداد خرید درون یک دوره زمان مشخص (F) و مقدار پول صرف شده در این مدت زمان مشخص شده (M) یا بر اساس تازگی، تکرار و ارزش پولی دسته بندی می کند (Qin, Yuan, & Wang, ۲۰۱۷). این روش یکی از مهم ترین روش های اندازه گیری ارزش عمر مشتری می باشد زیرا با بخش بندی مشتریان به وسیله ی داده های تراکنشی و بر مبنای این مدل، بخش های مشتریان با ارزش تعیین می شوند (D. Liu & Shih, ۲۰۰۵) باب استون معتقد است که بسته به صنعت مورد بررسی باید به هر یک از معیارهای تازگی، تکرار و ارزش پولی، وزن داد. خود او در پژوهشی که مرتبط با کارت اعتباری بود، اولویت اول را به تکرار،

سپس تازگی و در آخر ارزش پولی نسبت داد اما اولویت ها را به صورت شهودی تعیین نمود (Stone, ۲۰۰۷). مدت زمانی است که از آخرین خرید مشتری گذشته است، پژوهش گران دریافتند هر چه این مدت کمتر باشد احتمال پاسخ مشتری غالباً بیشتر است. تکرار، دفعات خرید مشتری از ابتدا تاکنون یا در یک دوره زمانی مشخص است که هر چه بیشتر باشد احتمال پاسخ را اغلب افزایش می دهد و ارزش پولی، میزان پول پرداخت شده توسط مشتریان در گذشته یا در یک دوره زمانی مشخص و یا میانگین خرید هر سفارش است که هر چه بیشتر باشد احتمال پاسخ افزایش می یابد. البته روابط منفی و مثبت ذکر شده همواره صحیح نیست. با این حال وقتی رابطه بین تازگی و احتمال پاسخ که اغلب منفی است مشخص شد، می توان آهنگ پاسخ را با استفاده از داده های قبلی مشتری محاسبه کرد. رابطه تکرار و احتمال پاسخ نیز باید به صورت تجربی تعیین شود، اما غالباً مثبت است. تکرار به صورت تعداد دفعات خرید مشتری از ابتدا تاکنون یا تعداد دفعات خرید از ابتدا تاکنون تقسیم بر مدت زمان بیان می شود (Bult & Wansbeek, ۱۹۹۵).

• روش شناسی پژوهش

از آنجا که این پژوهش همزمان از داده های تراکنشی و دموگرافیک مشتریان استفاده می کند، توصیفی اکتشافی می باشد. از نظر زمانی نوع تحقیق گذشته نگر یا پس رویدادی می باشد که در یک زمان مشخص اقدام به جمع آوری داده ها شده است و در پایان آن زمان به تجزیه و تحلیل داده ها پرداخته شده است. در این پژوهش به منظور آزمون الگو و روش های ارائه شده برای بخش بندی و طبقه بندی از داده های تراکنشی مشتریان بانک رفاه کارگران استفاده شد. بنابراین داده های تراکنشی مشتریان این بانک به عنوان جامعه آماری در نظر گرفته شد. تراکنش هایی که از اول مهر ۱۳۹۷ تا ۲۹ اسفند ۱۳۹۷ انجام گرفته بود به عنوان نمونه انتخاب شد. این داده ها، شامل دو قسمت اصلی است. قسمت اول مربوط به اطلاعات دموگرافیک مشتریان شامل جنسیت، سن و... بوده و قسمت دوم شامل داده های تراکنشی مشتریان این بانک می باشد که از آن به عنوان ابزاری برای جمع آوری داده ها و اطلاعات استفاده گردیده است. در نتیجه داده های مربوط به ۱۱۲۳۷۳۵ تراکنش گرد آوری و ثبت شد. از این داده ها برای بررسی مدل های طبقه بندی و رتبه بندی کننده مشتریان استفاده گردید.

به دلیل استفاده از مدل های داده کاوی و همچنین تغییرات شرایط اقتصادی و اولویت های مشتریان، نیاز به استفاده از جدیدترین داده ها احساس می شد، بنابراین از تمامی داده های مربوط به بازه زمانی ذکر شده که نسبتاً کامل بودند استفاده شده است. با توجه به مطالبی که قبلاً بیان شد، سه متغیر تازگی، تکرار و ارزش پولی برای بخش بندی بازار مورد استفاده قرار می گیرد. برای بدست آوردن این سه متغیر از تاریخ آخرین تراکنش به عنوان تازگی، تعداد تکرار تراکنش هر مشتری، به عنوان تکرار و از مانده حساب هر مشتری به عنوان ارزش پولی یا حجم مبادله استفاده گردیده و نتایج با استفاده از داده های سن، جنسیت و شغل مورد تجزیه و تحلیل قرار گرفته است. همچنین برای بخش قوانین انجمنی نیز از نوع حساب های مشتریان و خوشه های بدست آمده برای کشف روابط استفاده گردیده است. برای تعیین اعتبار و روایی نتایج روش های بخش بندی از معیار سیلوئت استفاده شده است و از آنجا که برای گردآوری داده ها در این مرحله از پرسشنامه استفاده نشده است، نیازی به تعیین پایایی وجود ندارد. داده ها مستقیماً از پایگاه داده مشتریان که به صورت دقیق ذخیره شده است استخراج گردیده و سپس برای کشف خطاهای احتمالی در داده ها عملیات پیش پردازش بر روی آنها انجام گردیده است. در این پژوهش برای انجام محاسبات مربوط به بخش داده کاوی از نرم افزار IBM SPSS Modeler ویرایش ۱۸ استفاده شده است.

• یافته های پژوهش

۴-۱- پیش پردازش داده ها

از آنجایی که داده های ارائه شده توسط بانک رفاه کارگران شامل داده های مربوط به شغل، سن، جنسیت، آخرین گردش حساب، متوسط مانده حساب، گردش بدهکار، گردش بستانکار، شرح حساب و نوع حساب مشتریانی بود که در طی شش ماه تراکنش انجام داده بودند، بنابراین برای دست یافتن به داده های کامل و بی نقص باید تمام این ویژگی ها به طور جداگانه بررسی و داده های پرت آنها شناسایی می شد که مراحل انجام این مرحله به شرح ذیل می باشد:

- محدوده سنی مورد بررسی برای این پژوهش از ۱۸ سال تا ۷۰ سال در نظر گرفته شد و بقیه داده ها از تجزیه و تحلیل ها حذف گردید.

- کد ۱ برای نشان دادن جنسیت مرد و کد ۲ برای نشان دادن جنسیت زن در نظر گرفته شد .

- از آنجایی که تاریخ تراکنش های داده ها از تاریخ اول مهر ۱۳۹۷ تا ۲۹ اسفند ۱۳۹۷ می باشد بنابراین داده هایی که تاریخ آخرین تراکنش آنها در این محدوده تاریخی نبود حذف گردید .

- حساب و نوع حسابی که داده های مربوط به آن با دیگر داده ها هماهنگی و همسویی نداشت و یا تعداد آن بسیار کم بود داده پرت در نظر گرفته شد و حذف گردید.

- از آنجایی که تعداد تراکنش ها برای تجزیه و تحلیل داده ها بسیار زیاد است بنابراین باید گروه بندی هایی انجام می گردید . در این پژوهش گروه بندی ها بر اساس جنسیت، سن و شغل انجام گردید . به عنوان مثال مردان ۳۵ ساله که دارای کد شغلی ۷۰۰ بودند در یک گروه قرار گرفتند . سپس ویژگی های تازگی، تکرار و حجم مبادله میانگین به عنوان تازگی، تکرار و حجم مبادله آن گروه در نظر گرفته شد.

۲-۴ مقایسه زوجی فرایند تحلیل سلسله مراتبی (AHP)، طراحی پرسشنامه خبره و گروه بندی مشتریان برای تعیین وزن معیارها و رتبه بندی گزینه ها از پرسشنامه مقایسه های زوجی) نه درجه ساعتی (فرایند تحلیل سلسله مراتبی (AHP) استفاده شد که توسط خبرگان صنعت بانکداری تکمیل گردید.

ارزش هر شاخص مدل RFM از ضرب مقدار نرمال شده شاخص در وزن آن تعیین گردیده است . ارزش این شاخص ها به صورت زیر است:

$$M = WM$$

$$F = WF$$

$$R = WR$$

جدول ۱ نشان می دهد در رتبه بندی مشتریان بانک حجم مبادله در مقایسه با تازگی ۶ بار مهمتر و در مقایسه با تکرار ۳ بار مهمتر است.

جدول ۱- مقایسه زوجی و تعیین وزن نرمال شده معیارها

تازگی	تعداد دفعات	حجم مبادله	ت اول
۰.۱۱۱	۰.۱۰۵	۰.۲۸۹	۰.۱۶۸۳۳۳
۰.۳۳۳	۰.۱۱۲	۰.۱۹۶	۰.۲۱۳۶۶۷
۰.۵۵۶	۰.۷۸۳	۰.۵۱۵	۰.۶۱۸

0.157545 0.149900601

0.203973 0.194076422

0.638482 0.607502301

0.951479

ارزش هر مشتری در هر خوشه از مجموع متوسط ارزش شاخص های RFM در آن خوشه محاسبه می گردد:

$$CLV = WR + WF + WM$$

۳-۴ گروه بندی مشتریان به روش RFM

نحوه امتیازدهی رویکرد RFM به این صورت است که ابتدا ماکزیمم و مینیمم مقدار هر شاخص از هم کم شده و حاصل آن به ده گروه تقسیم می شود. بطور مثال برای شاخص تکرار خرید، گروه اول شامل ۱۰٪ از مشتریانی هستند که بیشترین تکرار خرید در بازه زمانی مورد نظر را دارند و امتیاز ۹ به آن ها تخصیص داده می شود، ۱۰٪ دوم امتیاز هشت می گیرند و با همین روند ادامه می دهد تا امتیاز صفر به ۱۰٪ پایانی تعلق گیرد که کمترین تکرار خرید را دارند، برای شاخص مبلغ خرید نیز چنین روندی دنبال می شود در خصوص شاخص تأخر یا تازگی، با توجه به ماهیت آن، هر چه مقدار آن بالاتر باشد تأثیر آن منفی تر است زیرا نشان دهنده روزهایی است که مشتری تراکنشی نداشته است، لذا امتیازدهی مربوط به شاخص تأخر بصورت معکوس انجام می شود، امتیاز صفر تعلق می گیرد به ۱۰٪ مشتریانی که بالاترین مقدار تأخر را دارند یعنی تعداد روزهای بیشتری است که تراکنشی نداشته اند و به همین ترتیب ادامه می دهد تا به ۱۰٪ مشتریانی که کمترین مقدار تأخر را دارا هستند امتیاز نه اختصاص داده شود. جدول ۲ امتیاز دهی نهایی را نشان می دهد.

جدول ۲- مقیاس گذاری ویژگی های RFM

از؟ امت	تازگی-R (%)	تکرار-F (%)	حجم-M (%)
از؟ امت ۹	۰-۱۰	۹۰-۱۰۰	۹۰-۱۰۰
از؟ امت ۸	۱۰-۲۰	۸۰-۹۰	۸۰-۹۰
از؟ امت ۷	۲۰-۳۰	۷۰-۸۰	۷۰-۸۰
از؟ امت ۶	۳۰-۴۰	۶۰-۷۰	۶۰-۷۰
از؟ امت ۵	۴۰-۵۰	۵۰-۶۰	۵۰-۶۰
از؟ امت ۴	۵۰-۶۰	۴۰-۵۰	۴۰-۵۰
از؟ امت ۳	۶۰-۷۰	۳۰-۴۰	۳۰-۴۰
از؟ امت ۲	۷۰-۸۰	۲۰-۳۰	۲۰-۳۰
از؟ امت ۱	۸۰-۹۰	۱۰-۲۰	۱۰-۲۰
از؟ امت ۰	۹۰-۱۰۰	۰-۱۰	۰-۱۰

پس از محاسبه و تخصیص امتیازات صفر تا نه، وزن های بدست آمده در هر کدام ضرب شده و با جمع کردن امتیازات نهایی سه شاخص برای کلیه مشتریان، امتیاز WRFM برای هر مشتری بدست می آید، هر چه امتیاز WRFM برای یک مشتری بیشتر باشد، بهتر است.

۴-۴-۴ مدل سازی رتبه بندی مشتریان و ارزش طول عمر مشتری

۴-۴-۱- خوشه بندی داده ها با الگوریتم K-MEANS

یکی از مسائل مهم در خوشه بندی، تعیین تعداد بهینه خوشه هاست که در اکثر الگوریتم ها مانند K - means باید توسط خود کاربر معین شود و راه خاصی برای تعیین آن مشخص نشده است. یک راه حل ممکن امتحان kهای مختلف و تعیین مقدار بهینه بر اساس یکسری شاخص های از پیش تعریف شده است. بنابراین ابتدا با الگوریتم K - Means بر اساس روش RFM، خوشه بندی برای ۲ تا ۱۰ خوشه انجام گرفت و سپس برای تعیین تعداد خوشه بهینه از معیار سیلوئیت استفاده شد. بر اساس معیار Silhouette تعداد ۳ خوشه به عنوان تعداد خوشه بهینه انتخاب گردید. جدول زیر مقدار این متغیرها را به ازای تعداد خوشه ۳ نشان می

دهد .

جدول ۳- مشخصه هر خوشه با استفاده از الگوریتم k-means

تعداد رکورد ها	ن؟ انگ؟ م؟ CLV	ن؟ انگ؟ م؟ WM	ن؟ انگ؟ م؟ WF	ن؟ انگ؟ م؟ WR
۸۱۳	۶.۳۹	۴.۲۹	۱.۲۷	۰.۸۳
۱۰۶۲	۳.۸۳	۱.۸۱	۱.۵۲	۰.۵
۱۲۰۳	۰.۷۷	۰.۳۱	۰.۱۵	۰.۳۱

در این جدول مقادیر میانگین متغیرهای تازگی، تکرار و ارزش پولی هر بخش به صورت مجزا و همچنین میانگین CLV هر خوشه و تعداد مشتریان موجود در هر خوشه داده شده است. مطابق با اطلاعات ارائه شده، بخش اول با ۸۱۳ مشتری دارای بالاترین میانگین CLV بوده و ارزشمندترین بخش است.

۴-۲-۴ خوشه بندی داده ها با الگوریتم SOM

روشی دیگر برای بررسی تعداد بهینه خوشه ها، بررسی نقشه خروجی نقشه های خودسازمان ده کوهنن یا الگوریتم SOM است. داده های رتبه بندی مشتریان (CLV) به عنوان ورودی به شبکه وارد و ابعاد نقشه خروجی مدل کوهنن به صورت یک نقشه دو بعدی 10×7 که حالت پیشنهادی اولیه است در نظر گرفته شد. تفاوت رنگ موجود در نقشه نشانگر تعداد مشتری در هر یک از خانه ها است، هر چه تعداد مشتری بیشتری در یک خانه قرار گرفته باشد، آن خانه پررنگ تر است. با بررسی نتایج بدست آمده از این بخش، مشاهده می شود که می توان مشتریان را در سه بخش قرار داد.

جدول ۴- مشخصه خوشه های بدست آمده از الگوریتم SOM

تعداد رکورد ها	ن؟ انگ؟ م؟ CLV	ن؟ انگ؟ م؟ WM	ن؟ انگ؟ م؟ WF	ن؟ انگ؟ م؟ WR
۷۲۴	۶.۷۸	۴.۵۶	۱.۳۳	۰.۸۹
۱۱۰۴	۴.۰۱	۱.۸۳	۱.۵۴	۰.۶۴
۱۲۵۰	۰.۵۳	۰.۲۴	۰.۱۲	۰.۱۷

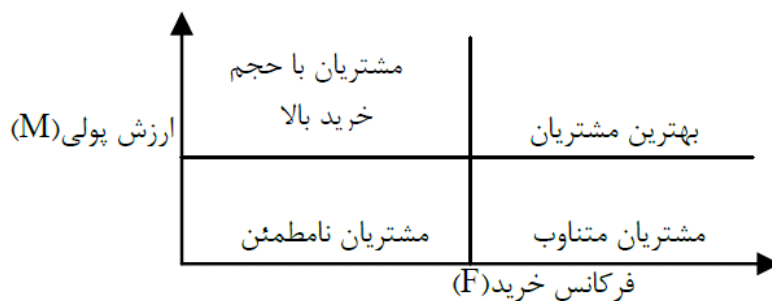
در این جدول دیده می شود که خوشه اول با ۷۲۴ مشتری و نمره میانگین CLV برابر با ۷۸/۶ دارای بیشترین مقادیر تازگی، تکرار و ارزش پولی می باشد.

مقادیر بدست آمده از خوشه بندی با الگوریتم SOM شبیه به الگوریتم k-means است ولی از آنجایی که معیار سیلوئت مربوط به الگوریتم SOM کمتر از معیار سیلوئت مربوط به الگوریتم k-means است بنابراین کیفیت بخش بندی مدل کوهنن یا SOM نسبت به مدل k-means کمتر است.

مقادیر تازگی، تکرار و ارزش پولی مشتریان سه بخش، نشان گر تفاوت بخش ها است. مشتریان بخش اول، دارای بالاترین مقادیر تازگی، تکرار و ارزش پولی نسبت به دو بخش دیگر هستند. نام این دسته از مشتریان را می توان مشتریان با ارزش زیاد نام گذاری کرد. در مقابل بخش سوم دارای پایین ترین مقادیر تازگی، تکرار و ارزش پولی نسبت به دو بخش دیگر است و از این رو نام آن گروه مشتریان با ارزش کم تعیین می شود. بخش دوم با مقادیر تازگی، تکرار و ارزش پولی متوسط به نام بخش با ارزش متوسط نامیده می شود.

-۳-۴- آنالیز خوشه ها بر مبنای RFM

مارکوس (Marcus, ۱۹۹۸)، ماتریس ارزش مشتری ارائه داد که تکرار خرید مشتری (F) و ارزش خرید مشتری (M) دو بردار این ماتریس را تشکیل داده اند. همچنین شاخص طول مدت همکاری و تأخر و تازگی خرید، به وفاداری مشتری مرتبط است؛ مارکوس معتقد است هر چه طول مدت همکاری بیشتر باشد و هر چه مدت کمتری از آخرین تراکنش مشتری گذشته باشد، مشتری وفادار تر است. در این مطالعه به دلیل نبود داده های مدت زمان همکاری مشتریان، این عامل حذف گردیده و بخش بندی بدون در نظر گرفتن طول مدت همکاری انجام شد.



شکل ۲- ماتریس ارزش مشتری (Marcus, ۱۹۹۸)

در این مطالعه با توجه به توصیف بخش بندی مشتریان (Ha & Park, ۱۹۹۸)، از علامت (↑) برای نشان دادن بالا تر بودن میانگین گروه از میانگین کل و از علامت (↓)، برای نشان دادن پایین تر بودن مقدار میانگین گروه از میانگین کل استفاده شده است.

جدول ۵- مقادیر میانگین کل متغیرهای R، F و M

WR-average	WF-average	WM-average
0.51	0.92	1.88

جدول ۶ به مقایسه مقادیر میانگین WR، WF و WM خوشه ها با میانگین کل این متغیرها می پردازد.

جدول ۶- نتایج مقادیر میانگین WR، WF و WM هر خوشه با میانگین کل

درصد خوشه ها	تعداد رکورد هلی	انگ؟م WM	انگ؟م WF	انگ؟م WR
خوشه ۱	813	?4.29	?1.27	?0.83
خوشه ۲	1062	?1.81	?1.52	?0.52
خوشه ۳	1203	?0.31	?0.15	?0.3

به دلیل ناقص بودن داده ها، طول ارتباط مشتریان بانک در دسترس نمی باشد ولی اگر در نظر بگیریم که سن مشتری رابطه مستقیم با طول ارتباط مشتری با بانک دارد) به طوری که از نتایج بخش های قبل نیز مشخص شد که هر چه سن افراد بالا تر باشد در خوشه بهتری جای می گیرند (می توانیم سن را به جای طول ارتباط با مشتری قرار داده و بدون وزن دادن به آن، مشتریان را از نظر ماتریس وفاداری بررسی نماییم).

بر طبق مطالعات چانگ و تسای (Chang & Tsay, ۲۰۰۴) با اضافه شدن شاخص L طول ارتباط مشتری (زمینه جهت تحلیل دقیق تر مشتریان فراهم می گردد. همچنین ادعا می کند که ارتباط طولانی تر مشتری، وفاداری بالاتر و کوتاه تر بودن زمان تعامل اخیر) تازگی یا تاخر، بالاتر بودن وفاداری مشتری را نشان می دهد. دو شاخص طول ارتباط مشتری (L) و زمان معامله اخیر (R) به عنوان ماتریس وفاداری مشتری تعریف می شود (Chang & Tsay, ۲۰۰۴).



شکل ۳- ماتریس وفاداری مشتری (Chang & Tsay, ۲۰۰۴)

اگر سن را L در نظر بگیریم و بازه سنی بین ۱۸ تا ۷۰ را به ده قسمت تقسیم کرده و به هر کدام از بازه ها امتیاز ۰ تا ۹ اختصاص دهیم بنابراین میانگین طبق جدول ۷ خواهد بود.

جدول ۷- مقادیر میانگین کل متغیر های L ، WR ، WF و WM

L-average	WR-average	WF-average	WM-average
3.5	0.51	0.92	1.88

جدول ۸ به مقایسه مقادیر میانگین L ، WR ، WF و WM خوشه ها با میانگین کل این متغیرها می پردازد.

درصد خوشه ها	تعداد رکورد های	؟انگ؟م WM	؟انگ؟م WF	؟انگ؟م WR	؟انگ؟م L
خوشه	813	?4.29	?1.27	?0.83	?5.5
خوشه	1062	?1.81	?1.52	?0.52	?3
خوشه	1203	?0.31	?0.15	?0.3	?3.2

بنابر جدول ۸ در ماتریس وفاداری مشتری، مشتریان خوشه سوم جزو مشتریان نامطمئن مشتریان خوشه دوم جزو مشتریان جدید و مشتریان خوشه اول جزو مشتریان وفادار می باشند.

۴-۵- قوانین انجمنی (الگوریتم Apriori)

از آنجایی که داده های اخذ شده از بانک رفاه کارگران داده های حساب های مشتریان را نیز در بر می گیرد بنابراین باید قوانین حاکم بر آنها نیز کشف گردد تا مشخص شود کدام حساب ها تراکنش بیشتر داشته اند و قوانینی که حاکم بر آنهاست چه می باشد. در داده های به کار برده شده در پژوهش از ۱۰۸ نوع حساب برای بیان حساب مشتریان استفاده شده است.

در این بخش از پژوهش، مشتریان هر خوشه از نظر نوع حساب هایی که استفاده می نمایند مورد بررسی قرار گرفته و بدین منظور حساب هایی که در مجموع بیش از یک درصد تراکنش ها را در طول شش ماه بازه زمانی مورد مطالعه داشتند با استفاده از قوانین انجمنی و الگوریتم Apriori تجزیه و تحلیل شدند. قواعد بدست آمده با حداقل پشتیبان ۴/۰ درصد و حداقل درجه اطمینان ۷۵ درصد به شرح زیر می باشد.

Consequent	Antecedent	Support %	Confidence %
خوشه 2	حساب 97 حساب 25 حساب 79	0.422	92.308
خوشه 1	حساب 98 حساب 79	0.487	86.667
خوشه 1	حساب 108 حساب 51 حساب 79	0.552	82.353
خوشه 1	حساب 108 حساب 41	0.747	78.261
خوشه 3	حساب 64	28.428	75.543

شکل ۴- قواعد حاصل از قوانین انجمنی

• نتایج پژوهش

تجزیه و تحلیل های بدست آمده از پرسشنامه خبرگان که توسط مدیران و خبرگان صنعت بانکداری تکمیل شده است نشان می دهد که حجم مبادله نسبت به تازگی و تکرار بسیار پر اهمیت بوده و ارزش بیشتری را به خود اختصاص می دهد. همچنین نتایج نشان می دهد با توجه به معیار سیلوئت تعداد بهینه سه خوشه برای خوشه بندی مشتریان مناسب تر بوده و نتایج برای خوشه ها به شرح زیر می باشد:

خوشه اول:

این خوشه دارای ۸۱۳ رکورد بوده و ۴/۲۶ درصد از کل مشتریان را به خود اختصاص داده و از نظر هر سه شاخص بالا تر از میانگین کل بوده و مشتریان این خوشه به عنوان **بهترین مشتریان** معرفی می گردند. همچنین مشتریان این خوشه جزو مشتریان وفادار می باشند که ۵۳ درصد این دسته از مشتریان را مردان و ۴۷ درصد آن را زنان تشکیل می دهند. بازه سنی این خوشه ۱۸ تا ۷۰ سال بوده و میانگین سنی این خوشه ۵۲ سال می باشد ۵/۷۴. درصد مشتریان این دسته را افراد بالای ۴۵ سال تشکیل می دهند و تنها ۵/۲۵ درصد آن مربوط به رده سنی ۱۸-۴۴ سال است و شغل های با کد ۷۰۶، ۳ و ۷۰ به ترتیب بیشترین درصد را در این خوشه

دارند که نشانگر شغل های دکترای غیرپزشکی و تاجر بوده و داده مربوط به کد ۷۰ ناقص بوده و شغل مربوط به کد ۷۰ در داده هایی که از بانک اخذ شده بیان نگردیده است.

خوشه دوم:

این خوشه که تعداد ۱۰۶۲ رکورد را به خود اختصاص داده است از نظر تازگی تراکنش تقریباً در حد میانگین بوده و از نظر تکرار تراکنش ها از میانگین کل بالاتر می باشد. حجم مبادله که در وزن دهی خبرگان مهمترین عامل بود در این خوشه کمتر از حد میانگین کل بوده و از نظر ماتریس ارزش مشتری مارکوس جزو مشتریان متناوب می باشند. همچنین مشتریان این خوشه جزو مشتریان جدید بوده که ۴۸ درصد این دسته از مشتریان را مردان و ۵۲ درصد آن را زنان تشکیل می دهند. بازه سنی این خوشه ۱۸ تا ۷۰ سال بوده و میانگین سنی افراد این خوشه ۳۹ سال می باشد. ۳۸ درصد این خوشه افراد زیر ۳۰ سال و ۳۰ درصد بین ۳۰ تا ۴۵ سال داشته که بیشترین درصد را دارند و افراد ۷۰-۵۷ ساله کمترین درصد را در این خوشه دارا می باشند (۱۰ درصد). در این خوشه بیشترین تعداد مشتریان دارای شغل هایی با کد ۱۳، ۱۲ و ۲ می باشند که کد ۲ معرف شغل آزاد بوده و کدهای ۱۲ و ۱۳ از نظر عنوان شغلی در داده های اخذ شده نامشخص می باشند.

خوشه سوم:

این خوشه که شامل ۱۲۰۳ رکورد می باشد از نظر هر سه شاخص R ، F و M وضعیت نامطلوبی دارد و در هر سه شاخص پایین تر از حد میانگین کل می باشد و از نظر ماتریس ارزش مشتری مارکوس جزو مشتریان نامطمئن محسوب می گردند. همچنین مشتریان این خوشه جزو مشتریان نامطمئن بوده که ۴۸ درصد این خوشه را مردان و ۵۲ درصد آن را زنان تشکیل می دهند. بازه سنی این خوشه ۱۸ تا ۷۰ سال بوده و میانگین سنی افراد این خوشه ۴۲ سال می باشد. ۳۶ درصد این خوشه افراد زیر ۳۰ سال و ۲۳ درصد بین ۳۰ تا ۴۵ سال داشته که بیشترین درصد را دارند و افراد ۷۰-۵۷ ساله کمترین درصد را در این خوشه دارا می

باشند (۵/۱۱) درصد. (بیشترین تعداد مشتریان این خوشه دارای کد ۱۳، ۱۲ و ۲ می باشند که کد ۲ معرف شغل آزاد بوده و کدهای ۱۲ و ۱۳ از نظر عنوان شغلی در داده های اخذ شده نامشخص می باشند.

خوشه اول بیشتر از حساب های سپرده سرمایه گذاری بلندمدت و نوع حساب طرح سودمند رفاه حقیقی (۹۸)، سپرده قرض الحسنه پس انداز و نوع حساب، مشتریان (۷۹)، گواهی سپرده مرحله شانزدهم (۱۰۸)، سپرده قرض الحسنه جاری اشخاص و نوع حساب، جاری توثیق تسهیلات-جاری بدون دسته چک (۵۱)، سپرده قرض الحسنه جاری اشخاص و نوع حساب جاری با کارت الکترونیکی (۴۱) استفاده می کنند. همچنین اکثر افراد این خوشه هر کدام دو یا سه حساب با هم دارند. خوشه دوم از حساب سپرده سرمایه گذاری بلندمدت و نوع حساب، سپرده یکساله جدید (۹۷)، سپرده قرض الحسنه پس انداز و نوع حساب، مشتریان (۷۹) و سپرده سرمایه گذاری کوتاه مدت اشخاص و نوع حساب، سپرده کوتاه مدت (۲۵) استفاده کرده و اکثر مشتریان این سه حساب را با هم دارند. خوشه سوم تنها از سپرده قرض الحسنه جاری اشخاص و نوع حساب، مددجویان سازمان بهزیستی (۶۴) استفاده می نمایند.

منابع:

- سلیمانی بشلی رع & ، اکبری اصل ر. (۱۳۹۰). *بازاریابی خدمات بانکی*. اتحاد.
- شهرابی رج. (۱۳۹۰). *داده کاوی در صنعت بانکداری*. تهران: انتشارات جهاد دانشگاهی.
- شهرابی رج & ، ذوالقدر شجاعی رع. (۱۳۹۱). *داده کاوی پیشرفته، مفاهیم و الگوها* چاپ دوم. (انتشارات جهاد دانشگاهی).
- صنعی آباده رم ، محمودی رس & ، طاهر پرور رم. (۱۳۹۳). *داده کاوی کاربردی* دوم. (تهران: نیاز دانش).
- کاتلر ف. (۱۳۹۶). *مدیریت بازاریابی* چاپ بیست و زب. فروزنده. (ed.) تهران: نشر آموخته.
- مصلحی رن ، کفاش پور آ & ، ناجی عظیمی ز. (۱۳۹۳). استفاده از مدل LRFM برای بخش بندی مشتریان بر اساس ارزش چرخه عمر آنها. *پژوهشهای مدیریت عمومی*. ۱۴۰-۱۱۹، 25،

- Aggarwal, C. C., Procopiuc, C., & Yu, P. S. (۲۰۰۲). Finding localized associations in market basket data. *IEEE Transactions on Knowledge and Data Engineering*, 14(۱), ۵۱-۶۲. [./https://doi.org/10.1109/TKDE.2002.1016212](https://doi.org/10.1109/TKDE.2002.1016212)
- Bult, J. R., & Wansbeek, T. (۱۹۹۵). Optimal Selection for Direct Mail. *Marketing Science*, 14(۴), ۳۷۸-۳۹۴. [...https://doi.org/10.1287/mksc.14.4.378](https://doi.org/10.1287/mksc.14.4.378)
- Chang, H. ., & Tsay, S. . (۲۰۰۴). Integrating of SOM and K mean in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal of Information Management*, 11(۴), ۱۶۱-۲۰۳.
- Chuang, C., Wei, Z., Zhifu, W., & Zhi, L. (۲۰۱۷). The diagnosis method of stator winding faults in PMSMs based on SOM neural networks. *Energy Procedia*, 105, ۲۲۹۵-۲۳۰۱. [...https://doi.org/10.1016/j.egypro.2017.03.113](https://doi.org/10.1016/j.egypro.2017.03.113)
- Edelstein, H. (۲۰۰۳). *Building profitable customer relationships with data mining*.
- Ha, S. H., & Park, S. C. (۱۹۹۸). Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications*, 15(۱), ۱-۳۱. [-https://doi.org/10.1016/S0957-4174\(98\)00081-7](https://doi.org/10.1016/S0957-4174(98)00081-7)
- Han, J., & Kamber, M. (۲۰۰۶). Data Mining: Concepts and Techniques. *Annals of Physics*, 54, ۷۷۰. [-https://doi.org/10.5860/CHOICE.49-33.0](https://doi.org/10.5860/CHOICE.49-33.0)
- Huerta-Muñoz, D. L., Ríos-Mercado, R. Z., & Ruiz, R. (۲۰۱۷). An iterated greedy heuristic for a market segmentation problem with multiple attributes. *European Journal of Operational Research*, 261(۱), ۷۵-۸۷. [...https://doi.org/10.1016/j.ejor.2017.02.013](https://doi.org/10.1016/j.ejor.2017.02.013)
- Hughes, A. M. (۲۰۱۲). *Strategic Database Marketing 4e : The Masterplan for Starting and Managing a Profitable , Customer-Based Marketing Program*.
- Kannisto, P. (۲۰۱۶). "I'M NOT A TARGET MARKET": Power asymmetries in market segmentation. *Tourism Management Perspectives*, 20, ۱۷۴-۱۸۰. [...https://doi.org/10.1016/j.tmp.2016.09.001](https://doi.org/10.1016/j.tmp.2016.09.001)
- KAUFMAN, L., & ROUSSEEUW, P. J. (۱۹۹۰). *Finding Groups in Data An Introduction to Cluster Analysis*. A JOHN WILEY & SONS.
- Kieu, L. M., Ou, Y., & Cai, C. (۲۰۱۸). Large-scale transit market segmentation with spatial-behavioural features. *Transportation Research Part C: Emerging Technologies*, 90(August ۲۰۱۷), ۹۷-۱۱۳. [...https://doi.org/10.1016/j.trc.2018.03.003](https://doi.org/10.1016/j.trc.2018.03.003)
- Lemon, K. N., & Mark, T. (۲۰۰۶). Customer Lifetime Value as the Basis of Customer Segmentation Customer Lifetime Value as the Basis of Customer Segmentation : Issues and Challenges. *Journal of Relationship Marketing*, (December ۲۰۱۴), ۳۷-۴۱. [vnhttps://doi.org/10.1300/J377v.0n.2](https://doi.org/10.1300/J377v.0n.2)

- Liu, D., & Shih, Y. (۲۰۰۵). *Integrating AHP and data mining for product recommendation based on customer lifetime value*. 42, ۳۸۷-۴۰۰. ...<https://doi.org/10.1016/j.im.2005.10.008>
- Liu, Y., Kiang, M., & Brusco, M. (۲۰۱۲). A unified framework for market segmentation and its applications. *Expert Systems with Applications*, 39(۱۱), ۱۰۲۹۲-۱۰۳۰۲. ...<https://doi.org/10.1016/j.eswa.2012.02.111>
- López, M., Valero, S., Senabre, C., Aparicio, J., & Gabaldon, A. (۲۰۱۲). Application of SOM neural networks to short-term load forecasting: The Spanish electricity market case study. *Electric Power Systems Research*, 91, ۱۸-۲۷. ...<https://doi.org/10.1016/j.epsr.2012.04.009>
- Marcus, C. (۱۹۹۸). A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, 15(۵), ۴۹۴-۵۰۴.
- Nadali, A., Kakhky, E. N., & Nosratabadi, H. E. (۲۰۱۱). Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system. *ICECT 2011 - 2011 3rd International Conference on Electronics Computer Technology*, 6(January ۲۰۱۱), ۱۶۱-۱۶۵. ...<https://doi.org/10.1109/ICECTECH.2011.5942073>
- Qin, R., Yuan, Y., & Wang, F. (۲۰۱۷). Exploring the Optimal Granularity for Market Segmentation in RTB Advertising via Computational Experiment Approach. *Electronic Commerce Research and Applications*. ...<https://doi.org/10.1016/j.elerap.2017.07.001>
- Stone, B. (۲۰۰۷). *Successful Direct Marketing Methods*. In *Journal of Advertising* (Eighth Edi). Lincolnwood: NTC Business Books, IL.
- Trindade, G., Dias, J. G., & Ambrósio, J. (۲۰۱۷). Extracting clusters from aggregate panel data: A market segmentation study. *Applied Mathematics and Computation*, 296, ۲۷۷-۲۸۸. ...<https://doi.org/10.1016/j.amc.2016.10.012>
- Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (۲۰۱۸). Two improved k-means algorithms. *Applied Soft Computing Journal*, 68, ۷۴۷-۷۵۵. ...<https://doi.org/10.1016/j.asoc.2017.08.032>
- Zhang, G., Zhang, C., & Zhang, H. (۲۰۱۸). Improved K-means Algorithm Based on Density Canopy. *Knowledge-Based Systems*. ...<https://doi.org/10.1016/j.knosys.2018.01.031>