



پژوهش‌های نوین در تصمیم‌گیری

دوره ۷، شماره ۱، بهار ۱۴۰۱، صص ۱۹۸-۲۲۳

نوع مقاله: پژوهشی

تحلیل خوشه‌ای مشتریان بر مبنای مدل WRFM با رویکرد داده‌کاوی غیرنظارتی (مورد مطالعه محصولات بهداشتی)

امید بشر دوست^۱، عزت اله اصغری زاده^۲، محمد علی افشار کاظمی^۳

۱. دانشجوی دکترای مدیریت، گروه مدیریت، دانشکده حسابداری و مدیریت، واحد رودهن، دانشگاه آزاد اسلامی، رودهن، ایران

۲. دانشیار، گروه مدیریت، دانشکده مدیریت، دانشگاه تهران، تهران، ایران

۳. دانشیار، گروه مدیریت، دانشکده مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ پذیرش: ۱۴۰۰/۰۸/۲۹

تاریخ دریافت: ۱۴۰۰/۰۶/۲۱

چکیده

در دنیای رقابتی امروز که شرکت‌ها با حجم انبوهی از اطلاعات مشتریان به علت رشد و پیشرفت فناوری‌های اطلاعاتی و ایجاد پایگاه‌های داده‌ای مختلف مواجه‌اند استفاده از ابزار مدیریت ارتباط با مشتری که بتواند به درستی و به موقع نیازها و انتظارات مشتریان را شناسایی و رصد کند بیش از پیش ضرورت می‌یابد؛ یکی از تکنیک‌هایی که می‌تواند در این برهه نقشی کلیدی و اساسی ایفا کند داده‌کاوی پایگاه داده‌هاست. هدف این پژوهش تحلیل داده‌های مشتریان بر پایه مدل WRFM به کمک روش‌های داده‌کاوی غیرنظارتی است؛ پژوهشگران درصددند که با کشف الگوهای موجود به آرایه استراتژی‌های مؤثرتری برای هر گروه از مشتریان و خصوصاً مشتریان کلیدی بپردازند تا برای سازمان سودآوری و عملکرد بهتری در برداشته باشد. جامعه آماری این پژوهش مشتریان محصولات بهداشتی در بازه زمانی سال‌های ۱۳۹۷-۱۳۹۸ است که به روش نمونه‌گیری هدفمند در دسترس تعداد ۶۴۸۵۸ نمونه از پایگاه داده مشتریان انتخاب شده است. به کمک ۳ تن از خبرگان (مدیران ارشد) فروش شرکت وزن شاخص‌های مدل WRFM تعیین شده است. برای تجزیه و تحلیل داده‌ها از نرم‌افزار کلمنتاین و SPSS استفاده شده است. با توجه به مدل پژوهش، ۴ دسته مشتری: خاص و کلیدی، طلایی بالقوه، نامطمئن از دست رفته، نامطمئن جدید شناسایی و نامگذاری شدند که برای هر یک از این دسته‌ها استراتژی‌های متفاوتی ارائه شده است. ضمناً نتایج نشان می‌دهد که خوشه‌بندی K- میانگین شش خوشه‌ای با خلوص ۰/۷۴۴ درصد نسبت به دیگر روش‌ها عملکرد بهتری داشته است.

کلیدواژه‌ها: تجزیه و تحلیل خوشه‌ای، تشخیص مغایرت، خوشه‌بندی، داده‌کاوی، مدل WRFM



۱- مقدمه

در محیط رقابتی امروز که دانش به عنوان یک منبع اساسی برای سازمان‌ها و بنگاه‌های اقتصادی مطرح است مدیریت ارتباط با مشتری به ابزاری استراتژیک و کلیدی برای بنگاه‌های اقتصادی و سازمان‌ها تبدیل شده است [۱، ص ۶۲]. امروزه تعامل شرکت‌ها با مشتریان در چارچوب مدیریت ارتباط با مشتری به‌طور چشمگیری تغییر یافته است. شناسایی ویژگی‌ها و سلیقه‌های مشتریان مختلف و تخصیص بهینه منابع به آنها با توجه به ارزشی که برای شرکت‌ها خلق می‌کنند به یکی از دغدغه‌های اصلی در حوزه مدیریت ارتباط با مشتری تبدیل شده است [۲، ص ۶۱]. به منظور دستیابی به مزیت رقابتی و افزایش سودآوری، شناسایی و تحلیل ویژگی‌ها و بخش‌بندی مشتریان بر اساس ارزشی که برای سازمان دارند، زمینه را برای تخصیص بهینه منابع محدود، به‌کارگیری استراتژی‌های مناسب بازاریابی و در نهایت مدیریت سودآوری فراهم می‌کند [۳، ص ۱۷۶]. شناسایی مشتریان و نیازهای آن‌ها فاکتورهای مهمی‌اند که به تولیدکنندگان برای کسب مزیت رقابتی هم در ارایه محصول و هم ارایه خدمات به مشتریان کمک می‌کند. شرکت‌ها باید مشتریان خود را اولویت‌بندی کنند و تمرکز بیشتری روی مشتریان کلیدی خود داشته باشند [۴، ص ۳۳۱]. امروزه تعامل شرکت‌ها با مشتریان در قالب مدیریت ارتباط با مشتری به طور قابل توجهی تغییر یافته است. شناسایی ویژگی‌های مشتریان مختلف و تخصیص بهینه منابع به آنها با توجه به ارزشی که برای شرکت‌ها دارند، به یکی از دغدغه‌های اصلی در حوزه مدیریت ارتباط با مشتری تبدیل شده است [۵، ص ۶۵]. در کنار مباحث مدیریت ارتباط با مشتری امکانات استفاده از شبکه اینترنت باعث مواجه شدن با حجم زیادی از داده و اطلاعات شده است که داده‌کاوی یکی از ابزارهای استخراج دانش و الگوهای موجود در این پایگاه داده‌های گسترده است [۳]. یکی از روش‌های هر تصمیم‌گیری مدیریتی در دست داشتن راهکارهای عملیاتی به همراه احتمال موفقیت و اثربخشی این راهکارهاست که استفاده از تکنیک‌های داده‌کاوی می‌تواند این مهم را محقق سازد [۶، ص ۷].

صنعت مورد مطالعه در این پژوهش یعنی محصولات بهداشتی، روزانه با حجم انباشته شده‌ای از اطلاعات خرید و جمعیت‌شناختی مشتریان مواجه است؛ انجام پژوهشی که بتواند با استفاده از داده‌کاوی به ارایه استراتژی‌های بهتری به مشتریان در گروه‌های مختلف بپردازد و باعث بهبود عملکرد فروش و سودآوری شرکت شود برای مدیران و کارشناسان ضرورت می‌یابد. بنابر مطالب عنوان شده و اهمیت داده‌کاوی پایگاه داده‌های مشتریان به منظور مدیریت بهتر ارتباط با مشتری، این پژوهش به دنبال آن است که بررسی نماید کدام روش داده‌کاوی



برای بخش‌بندی مشتریان با هدف بهبود مدیریت ارتباط با مشتری می‌تواند عملکرد، دقت و کارایی بالاتری داشته باشد. در راستای اهداف این پژوهش می‌توان سؤال‌های این پژوهش را در قالب ذیل بیان کرد:

۱. بهترین تعداد خوشه برای بخش‌بندی مشتریان با استفاده از شاخص‌های بدست آمده چیست؟
۲. از نقطه نظر شاخص‌های مدل پژوهش مشتریانی که برای شرکت سودآورترند چه ویژگی‌هایی دارند؟

۲- پیشینه و مبانی نظری پژوهش

۲-۱- مدیریت ارتباط با مشتری^۱

سابقه مدیریت ارتباط با مشتری و پیاده‌سازی آن به صورت امروزی به دهه ۱۹۹۰ برمی‌گردد. مشتری‌مداری و رضایت مشتری از مباحث جدید در جهان امروز است [۷]. بحث اینکه سازمان چگونه مشتریان را جذب کند؟ و یا اینکه مشتریان فعلی خود را چگونه حفظ کرده و به آنها خدمات بهتری ارائه کند؟ بحث فناوری اطلاعات و ابزارهای جدیدی که توسط این فناوری ایجاد شده باعث شده است تا مدیریت ارتباط با مشتری به مبحثی نو و تازه در مهندسی و مدیریت تبدیل شود [۸].

۲-۲- بخش‌بندی^۲

بخش‌بندی به شرکت‌ها اجازه می‌دهد تا گروه‌های مشتری را شناسایی کرده و نیازهای آنها را مشخص نموده و انتظارات مشتریان را به‌طور تخصصی‌تر برآورده کرده و پاسخگویی به این خواسته‌ها را افزایش دهد و در حالی که کاربران جدید و بالقوه را با استراتژی‌های بازاریابی مرتبط جذب می‌کند باعث افزایش درآمد کلی شرکت می‌شود. امروزه تقسیم‌بندی‌های بازارها با توجه به موارد متعددی صورت می‌گیرد که می‌تواند شامل موارد ذیل باشد: ۱- تقسیم‌بندی براساس خصوصیات جمعیت‌شناسی مشتریان (ویژگی‌هایی مانند: سن، جنسیت، درآمد، تحصیلات، شغل)؛ ۲- تقسیم‌بندی بر اساس خصوصیات روانشناسی^۳ (در این شیوه تقسیم بازار، مردم بر اساس گرایش‌ها، ارزش‌ها و سبک‌های زندگی گروه‌بندی می‌شوند و معمولاً داده‌ها با استفاده از پرسشنامه جمع‌آوری می‌شوند)؛ ۳- تقسیم-



بندی بازار بر اساس خصوصیات رفتاری مشتریان (در این شیوه تقسیم‌بندی بازار، خرید و مصرف یا عدم خرید و مصرف محصول، تعداد دفعات خرید و مقدار مصرف مورد توجه قرار می‌گیرد)؛ ۴- تقسیم‌بندی مشتریان با توجه به منافع محصول برای مصرف‌کننده؛ ۵- تقسیم‌بندی بازارها بر اساس ویژگی‌های جغرافیایی (بازارهای که در یک منطقه جغرافیایی قرار دارند به یکدیگر نزدیک‌اند و می‌توان به طور هم‌زمان با آنها ارتباط برقرار کرد و یا از آنها بازدید کرد) [۹، ص ۱۹۳-۱۹۹].

در این پژوهش برای تقسیم‌بندی بازار مشتریان از ویژگی‌های خصوصیات رفتاری، جمعیت‌شناختی و داده‌های مالی و خرید مشتریان استفاده شده است همچنین بر پایه نتایج داده‌کاوی، بخش‌بندی مختلف مشتریان صورت می‌گیرد که در ادامه بعد از پیشینه پژوهش به آن اشاره می‌شود.

۳-۲- مدل تازگی، تکرار و ارزش پولی^۵

در اوایل دهه ۶۰ بود که کالینان برای اینکه بتواند مشتریانی که احتمال خرید دارند را از آنهایی که قصد خرید ندارند جدا کند روش کمی، ساده و مؤثری را ابداع کرد. وی روش خود را تجزیه و تحلیل تازگی، فراوانی و ارزش پولی یا به طور خلاصه RFM نامید. از این زمان به بعد پژوهشگران مختلفی از این مدل و مدل‌های ارتقایافته آن برای بخش‌بندی و تعیین ارزش مشتریان خود استفاده کردند. تکنیک تازگی، تکرار و ارزش پولی رفتار مشتری را شناسایی می‌کند و ویژگی‌های رفتار مشتری را بر اساس سه متغیر ذیل نشان می‌دهد: ۱- تکرار: آخرین خرید که به فاصله زمانی بین زمانی که آخرین رفتار خرید مصرف‌کننده اتفاق افتاده و حال حاضر اشاره دارد؛ ۲- فراوانی خریده‌ها: که به تعداد معاملات در یک دوره خاص اشاره دارد؛ ۳- ارزش پولی: که به مبلغ پول مصرف شده در یک دوره خاص اشاره دارد. پیاده‌سازی مدل تازگی، تکرار و ارزش پولی سنتی، هر مشتری را از نظر ارزشمندی آنها بر اساس پارامترها در مقابل سایر مشتریان رتبه‌بندی می‌کند و یک امتیاز تازگی، تکرار و ارزش پولی را برای هر مشتری یا محصول ایجاد می‌کند [۱۰، ص ۵۲۶۰].

بیشتر پژوهش‌های اولیه در مورد مدل تازگی، تکرار و ارزش پولی به برابری وزن هر یک از شاخص‌ها تأکید داشته‌اند مانند پژوهش‌های بلتبرگ و دیگران [۱۱] و خواجه‌وند و دیگران [۱۲]. پژوهش‌هایی درباره بهبود دقت بازاریابی هدف با استفاده از مدل تازگی، تکرار و



ارزش پولی انجام شده است که یکی از این بهبودها ارایه مدل تازگی، تکرار و ارزش پولی موزون است که در این مدل برای هر شاخص، وزنی متناسب با نوع کسب و کار و صنعت باتوجه به نظر خبرگان صنعت در نظر گرفته می‌شود [۱۳]: بنابراین هر وزن متناسب با نظر خبرگان هر صنعت با استفاده از تکنیک فرآیند تحلیل سلسله مراتبی تعیین می‌شود که این موضوع باعث می‌شود محدودیت‌های مدل تازگی، تکرار و ارزش پولی موجود به‌طور مؤثری برطرف شود [۱۴].

بلتبرگ و دیگران (۲۰۰۸)، در مورد اهمیت بازاریابی پایگاه‌های داده‌های مشتریان پژوهشی انجام داده‌اند همچنین بیان کرده‌اند که مدل تجزیه و تحلیل تازگی، تکرار و ارزش پولی اولین مدل پیش‌بینی کننده‌ای بوده است که برای بازاریابی پایگاه داده‌های مشتریان استفاده شده است [۱۱]. اولین گام در این مدل این است که پرونده مشتری را بر اساس اینکه چگونه مشتریان به تازگی از شرکت خرید کرده‌اند مرتب کرده و سپس پایگاه داده‌ها به پنج قسمت مساوی (پنجک) تقسیم کرده و این پنجک‌ها از شماره ۱ تا ۵ شماره‌گذاری شود بنابراین ۲۰ درصد از مشتریانی که به تازگی خریدهایی از شرکت داشته‌اند به شماره ۵ تعلق خواهند گرفت و ۲۰ درصد بعدی به شماره ۴ و این کار تا شماره یک ادامه دارد. مراحل بعد شامل مرتب کردن مشتریان بر اساس تعداد خرید و بعد از آن براساس ارزش پولی خرید است. در نهایت پایگاه داده به ۱۲۵ گروه تقریباً مساوی (سلول) با توجه به تازگی، تعداد خرید و ارزش پولی تقسیم می‌شود. مشتریان یا محصول‌های با نمرات بالاتر معمولاً سودآورترین اند [۱۳].

۴-۲- داده‌کاوی^۷

داده‌کاوی شامل فرایند استخراج الگوهای جالب، کوچک اما بااهمیت، ضمنی، نا شناخته و بالقوه مفید و یا دانش از مقدار بسیار عظیمی از داده‌هاست. این فرایند در بردارنده مجموعه‌ای از فعالیت‌ها برای پیدا کردن الگوهای جدید، ضمنی و یا غیرمنتظره در داده‌ها یا الگوهای غیر معمول در آنهاست [۱۵، ص ۵۶].

شش فعالیت اصلی داده‌کاوی شامل: ۱- دسته‌بندی، ۲- تخمین، ۳- پیش‌بینی، ۴- گروه-بندی شباهت، ۵- خوشه‌بندی، ۶- توصیف و نمایه‌سازیست؛ که سه دسته اول آن داده‌کاوی هدایت‌شده بوده و دسته ۴ و ۵ داده‌کاوی غیرهدایت شده است و دسته آخر به داده‌کاوی هدایت‌شده و هدایت‌نشده تقسیم می‌شود که گروه‌بندی شباهت گاهی با عنوان الگوریتم کشف



مغایرت^{۱۶} مطرح می‌شود [۱۶].

مدل‌های کشف مغایرت به منظور تشخیص نقاط پرت یا موارد غیرمعمول در داده‌ها مورد استفاده قرار می‌گیرند. این مدل‌ها تنها اطلاعات مربوط به رفتارهای نرمال را ذخیره‌سازی می‌کنند. این مدل کشف مغایرت نیازی به مجموعه داده‌ای برای آموزش و یادگیری ندارد [۱۷، ص ۱۲۵-۱۲۶].

خوشه‌بندی، تقسیم داده‌ها به گروه‌هایی از اشیای مشابه است. فرآیند قراردادن داده‌ها در گروه‌هایی که بیشترین شباهت را در برخی ویژگی‌ها دارند خوشه‌بندی نامیده می‌شود. هر خوشه شامل مجموعه‌ای از داده‌های مشابه درون خوشه‌ای بوده و متفاوت از داده‌های سایر خوشه‌هاست [۱۸].

بسیاری از الگوریتم‌های خوشه‌بندی توسعه یافته‌اند که از چندین جنبه طبقه‌بندی شده‌اند مانند: الگوریتم‌های تقسیم بندی^۱ بر اساس تقسیم کل داده‌ها به گروه‌های متفاوت، روش‌های تجمع سلسله‌مراتبی^۱، روش‌های (تفرقه‌افکنانه-تقسیمی^۱)، روش‌های مبتنی بر تراکم^۲، روش-های مبتنی بر شبکه^۲ و مبتنی بر مدل^۴ [۱۹، ص ۲۵۳] که شناخته شده‌ترین و پرکاربردترین روش تقسیم‌بندی روش خوشه‌بندی k - میانگین است که در این پژوهش از این روش در کنار روش‌های خوشه‌بندی دومرحله‌ای و روش تشخیص مغایرت استفاده شده است.

۳- پیشینه تجربی پژوهش

در اینجا تنها به بعضی از پژوهش‌هایی که در بعضی موارد شباهت‌هایی با هدف و روش این پژوهش داشته است اشاره می‌شود:

هوانگ و لی^۵ (۲۰۲۱)، پژوهشی را با هدف شناسایی و اولویت‌بندی مشتری برای محصولات جدید با استفاده از مدل تازگی، تکرار و ارزش پولی و رتبه‌بندی متنی^۶ انجام داده‌اند. این پژوهشگران مدلی با عنوان RWF ارائه کرده‌اند که برای جامعه آماری و اعتبار سنجی مدل پیشنهادی خود از داده‌های وبسایت کاربران موبایل کره جنوبی در بازه زمانی ۱ می تا ۳۱ جولای ۲۰۲۰ بهره برده‌اند؛ برای تجزیه و تحلیل اوزان از الگوریتم رتبه‌بندی متنی استفاده کرده‌اند. با مدل پیشنهادی نرخ پاسخگویی مشتریانی عالی سازمان ۳۰ درصد رشد داشته است، همچنین نتایج پژوهش این امکان را برای متخصصان فراهم ساخته تا بتوانند از مشتریان هدف خود



بدون داشتن سابقه خرید قبلی اطلاعات با ارزشی تهیه کنند [۲۰].

بشردوست و دیگران (۲۰۲۱)، پژوهشی را با عنوان ارایه الگوی دسته‌بندی مشتریان با رویکرد داده‌کاوی ترکیبی انجام داده‌اند که در این پژوهش بر اساس الگوریتم مؤلفه‌های مهم که بر پایه مقدار پی^۷ اهمیت هر متغیر جمعیت شناختی و اطلاعات خرید مشتری تعیین شده و بر اساس شاخص مغایرت ۲/۱۴۵ از مجموع اطلاعات مشتریان ۶۵۵ رکورد مغایر تشخیص داده شده است که بعد از حذف این تعداد از مشتریان پژوهشگران از روش‌های ترکیبی داده‌کاوی برای دسته‌بندی مشتریان استفاده کرده‌اند که در این پژوهش در روش دومرحله‌ای شاخص درصد خلوص خوشه‌ای ۰/۵۳۳ و در روش k- میانگین ۰/۷۶۱ بدست آمده است. در نهایت در این پژوهش مشخص شده است که روش k- میانگین ۴ خوشه‌ای بهترین عملکرد را برای دسته‌بندی مشتریان داشته است و متناسب با این نتایج به ارایه پیشنهادهایی پرداخته‌اند [۲۱].

عباسی‌مهر و شعبانی (۲۰۲۱)، پژوهشی را با هدف ارایه روشی جدید برای تجزیه و تحلیل و پیش‌بینی رفتار پویای مشتری با استفاده از خوشه‌بندی سری زمانی در مورد مشتریان بانک انجام داده‌اند. داده‌های این پژوهش شامل ویژگی‌های تازگی، دفعات و ارزش پولی خرید بوده است که از داده‌های نقطه فروش^{۱۸} بخش تجزیه و تحلیل داده‌های بانک بدست آمده است. پس از انجام یک مطالعه تجربی بر روی داده‌های ۲۵۳۱ مشتری، روند غالب رفتار مشتریان با استفاده از روش پیشنهادی کشف شده است. بر اساس تجزیه و تحلیل ویژگی پولی، مشتریان به چهار بخش اصلی تقسیم شدند. در این پژوهش یک روش جدید برای جذب رفتار پویای مشتری با استفاده از خوشه‌بندی سری‌های زمانی بر روی داده‌های مرتب شده بر اساس زمان ارایه شده است. از نظر پژوهشگران، این پژوهش اولین مطالعه‌ای بوده است که ترکیبی از زمان، فراوانی، ارزش پولی و خوشه‌بندی سری‌های زمانی را برای نشان دادن روند رفتار پویای مشتریان استفاده کرده است [۲۲].

کاباساکال^{۱۹} (۲۰۲۰)، مطالعه‌ای موردی را با هدف بخش‌بندی مشتریان بر اساس مدل تازگی، تکرار و ارزش پولی در یک خرده‌فروشی الکترونیکی انجام داده‌است؛ در این پژوهش، یک نمونه اولیه نرم‌افزار معرفی شده است که با استفاده از مدل تازگی، تکرار و ارزش پولی به روش امتیازدهی کمک می‌کند. داده‌های فروش به‌دست آمده برای خوشه‌بندی به کمک نمونه اولیه نرم‌افزار تجزیه و تحلیل شده و خوشه‌های کشف شده با انواع مقادیر تازگی، تکرار و ارزش پولی با استفاده از معیارهای ارزیابی خوشه مقایسه شده‌اند. سرانجام برای هر بخش -



پیشنهادهایی مربوط به استراتژی‌های بازاریابی ارائه شده است [۲۳].

بابایان و سرفرازی (۲۰۱۹)، پژوهشی را با عنوان تجزیه و تحلیل مشتریان مخابرات خراسان جنوبی با گسترش مدل تازگی، تکرار و ارزش پولی به مدل طول ارتباط مشتری، تازگی، تکرار و ارزش پولی انجام داده‌اند. هدف این پژوهشگران طبقه‌بندی مشتریان در خوشه‌های مختلف بوده است تا مشتریانی که مصرف بالایی دارند و یا مشتریانی که ارزش بالایی برای شرکت داشته و از دست رفته‌اند شناسایی شوند. در این پژوهش از نرم‌افزار کلمنتاین برای تجزیه و تحلیل داده‌ها استفاده شده است و برای خوشه‌بندی از روش خوشه‌بندی k - میانگین و خوشه‌بندی دومرحله‌ای استفاده شده است؛ در نهایت ۹ خوشه توسط شاخص دیویس- بولدین برای روش دومرحله‌ای انتخاب شده است که بر اساس مقدار شاخص LRFM، شرکت مشتریان را در ۵ گروه وفادار، بالقوه، جدید، از دست‌رفته و پر مصرف قرار داده است که با این تقسیم‌بندی‌ها شرکت قادر خواهد بود تصمیم‌های بهتری را در آینده برای هر گروه از مشتریان خود داشته باشد [۴].

صالحی و سالاری (۲۰۱۷)، پژوهشی را به منظور شناسایی رفتار مشتریان کاشی و سرامیک با مقایسه فنون داده‌کاوی و منطق فازی انجام داده‌اند که نتایج نشان داده است که شبکه‌عصبی فازی با درصد پیش‌بینی صحیح ۰/۹۴۱ نسبت به سایر روش‌های مورد استفاده، نتیجه ویژگی‌های مدنظر مشتریان را بهتر می‌تواند پیش‌بینی کند [۲۴، ص ۱۷۳].

بشیری موسوی و دیگران (۲۰۱۵)، با استفاده از روش داده‌کاوی و تحلیل سلسله مراتبی فازی^{۱۲} به تجزیه و تحلیل ارزش مشتری در شعب بانک تجارت زنجان پرداخته‌اند. این پژوهشگران به کمک اطلاعات جمعیت‌شناختی، فراوانی، مقدار خرید و ویژگی اعتماد به کمک خوشه‌بندی k - میانگین به تجزیه و تحلیل مشتریان در قالب هرم ارزشی پرداخته‌اند که در نهایت ۴ طیف کلی برای مشتریان خود تعریف نموده‌اند: مشتریان پلاتینیومی، طلایی، نقره‌ای و سربی [۲۵].

از بررسی پیشینه پژوهش می‌توان بیان کرد که پژوهش‌هایی که با استفاده از روش‌های مختلف داده‌کاوی به خصوص خوشه‌بندی و دسته‌بندی چه به صورت مجزا و چه به صورت ترکیبی بر اساس معیارهای متفاوت به منظور کشف و استخراج دانش از پایگاه داده‌های مشتریان اجرا شده، بنا به هدف، نوع و ماهیت پژوهش‌ها دستاوردها و نتایج متفاوتی در پی داشته‌است. بهبود شیوه‌های خوشه‌بندی و دسته‌بندی مشتریان به خصوص در عرصه خدمات بانکداری و صنعت خرده‌فروشی باعث صرفه‌جویی در زمان، هزینه و رضایت‌مندی کاربران و



استفاده‌کنندگان (مصرف‌کنندگان) شده است؛ همچنین در صنعت خرده فروشی باعث افزایش فروش و تخصصی‌تر شدن مباحث بازاریابی مدیریت ارتباط با مشتری شده است. نتایج پژوهش‌ها نشان می‌دهد که روش‌های داده‌کاوی در بخش‌بندی علمی و دقیق‌تر بازار هدف مشتریان چقدر می‌تواند مؤثر و کارا باشد. در این پژوهش ضمن بررسی روش‌های مختلف خوشه‌بندی، به دنبال بررسی بهترین روش خوشه‌بندی با توجه به شرایط پژوهش و بازار مشتریان مورد بررسی در حیطه مدیریت ارتباط با مشتریان هستیم.

۴- روش پژوهش

در مورد ماهیت پژوهش قمست‌هایی که به نمایش اطلاعات جمعیت شناختی مشتریان جامعه هدف می‌پردازد ماهیت توصیفی داشته و قسمت‌هایی که مربوط به یافته‌های پژوهش بوده و از داده‌کاوی استفاده شده جنبه تحلیلی و اکتشافی دارد (اکتشافی بودن از این منظر است که دانش درون پایگاه داده‌های مشتریان کشف و استخراج می‌شود)؛ به طور کلی می‌توان این پژوهش را از نظر ماهیت انجام پژوهش توصیفی- تحلیلی اکتشافی دانست.

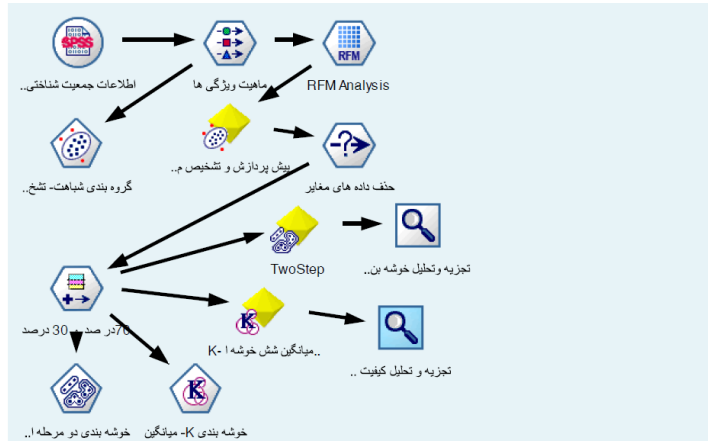
جامعه آماری این پژوهش شامل مشتریانی از استان تهران شامل مناطق ۸، ۱۲، ۱۳، ۱۴، ۱۵ شهرداری تهران و شهرهای گرمسار، ورامین، قرچک، ایوانکی، دماوند، پیشوا و پاکدشت بوده که در بازه سال‌های ۱۳۹۷-۱۳۹۸ اطلاعات این مشتریان در قالب صفحه‌های گسترده^{۲۲} از پایگاه داده‌های شرکتی از گروه صنعتی گلرنگ که در زمینه عرضه محصولات بهداشتی و آرایشی فعالیت می‌کند گردآوری شده است و از این بین تعداد ۶۴۸۵۸ نمونه با روش نمونه‌گیری هدفمند در دسترس انتخاب شده است (پایگاه اولیه داده‌های شرکت شامل ۶۴۸۵۸ رکورد (سطر) و ۱۹ ویژگی (ستون) بوده که هر سطر اطلاعات مربوط به یک مشتری است) (این روش نمونه‌گیری به این علت انتخاب شده است که در داده‌کاوی مهمترین مسأله دسترسی به اطلاعات مرتبط و داده‌های با حجم بالاست که بتوان از آن اطلاعات ارزشمند تری استخراج کرد).

برای تهیه و تدوین پیشینه پژوهش از روش مطالعات کتابخانه‌ای و با کمک موتورهای جستجوگر در بین نشریه‌های علمی داخلی و خارجی معتبر موضوعات مربوط با هدف پژوهش گردآوری شد. بخشی که مربوط به اطلاعات مشتریان شرکت بوده از پایگاه داده‌های شرکت در قالب اکسل گردآوری شده است سپس این اطلاعات که شامل اطلاعات جمعیت شناختی مشتریان و اطلاعات کمی خرید مشتریان است در محیط نرم‌افزار (اس.پی.اس.اس.^{۲۳}) برده شده



تا شناخت کلی نسبت به ماهیت داده‌های اولیه پژوهش ایجاد شود سپس با استفاده از مدل تازگی، تکرار و ارزش پولی شاخص‌های مربوط به هر مشتری استخراج شد و به متغیرهای اولیه پژوهش اضافه شده است؛ در ادامه از ۳ نفر از مدیران ارشد شرکت (خبرگان) که دارای سابقه و دانش لازم در زمینه محصولات آرایشی و بهداشتی بوده‌اند خواسته شده که در قالب ماتریس مقایسه‌های زوجی نظر خود را درباره اوزان هر یک از شاخص‌های مدل تازگی، تکرار و ارزش پولی بیان کنند (بر پایه میانگین حسابی سطری، وزن هر شاخص ایجاد شده است) سپس با استفاده از روش فرآیند تحلیل سلسله مراتبی و به روش میانگین هندسی نظرهای خبرگان با هم ادغام شده و برای تعیین ثبات و اعتبار وزن‌های ایجاد شده شاخص ناسازگاری نیز محاسبه و بررسی شده است. بعد از اطمینان از وزن‌های بدست آمده، این وزن‌ها در مقادیر مربوط به هر شاخص مدل تازگی، تکرار و ارزش پولی ضرب شده تا مقدار نهایی تازگی، تکرار و ارزش پولی موزون برای هر مشتری بدست آید. سپس این چند متغیر جدید به متغیرهای اولیه پژوهش اضافه و داده‌کاوی بر روی این داده‌ها انجام شده است.

برای تجزیه و تحلیل داده‌های پژوهش از نرم‌افزارهای (اس. پی. اس. اس.) نسخه ۲۱ و کلمنتاین^{۲۴} نسخه ۱۲ استفاده شده است بدین صورت که کدگذاری و بررسی اولیه داده‌ها در محیط (اس. پی. اس. اس.) انجام شده و شاخص تازگی، تکرار و ارزش پولی نیز توسط این نرم افزار برای هر مشتری محاسبه شده سپس از روش‌های k - میانگین و دو مرحله‌ای^{۲۵} نیز به کمک نرم‌افزار کلمنتاین، برای خوشه‌بندی اطلاعات مشتریان استفاده شده است. همچنین برای تعیین بهترین روش خوشه‌بندی و کیفیت خوشه‌های ایجاد شده از شاخص‌های درصد خلوص^{۲۶} و اطلاعات متقابل نرمال‌شده^{۲۷} نیز استفاده شده است. از ۷۰ درصد داده‌ها برای یادگیری داده‌کاوی و از مابقی آن برای تعیین میزان دقت و صحت آزمون نتایج داده‌کاوی استفاده شده است. در شکل ۱ کلیه مراحل انجام مدلسازی (مدل مفهومی پژوهش) تا رسیدن به خوشه‌بندی که توسط نرم‌افزار داده‌کاوی کلمنتاین ایجاد شده به تصویر کشیده شده است.



شکل ۱. فرآیند مدلسازی داده‌کاوی پژوهش (مدل مفهومی)

۵- یافته‌های پژوهش

در این قسمت ابتدا به متغیرهای اولیه و ثانویه که در جریان این پژوهش ایجاد و به کار رفته اند در قالب جدول ۲ اشاره می‌شود.

جدول ۲. متغیرهای اولیه پژوهش به تفکیک نوع، مقیاس و نقش

ردیف	نام متغیر	نوع متغیر	مقیاس اندازه‌گیری	وضعیت	نقش
۱	گروه مشتریان سطح ۱ (سه حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۲	گروه مشتریان سطح ۲ (چهار حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۳	گروه مشتریان سطح ۳ (دو حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۴	شهر (هشت حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۵	وضعیت مشتری (چهار حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۶	منطقه (هفت حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۷	زمان سپری شده از آخرین خرید سال ۹۷	numeric	Scale	متغیر پایه اولیه	مستقل
۸	ماه خرید (دوازده حالت)	String	Nominal	متغیر پایه اولیه	مستقل
۹	خرید در روز چندم ماه	numeric	Scale	متغیر پایه اولیه	مستقل
۱۰	گروه بازار محصولات (۲۶ قلم کالا)	String	Nominal	متغیر پایه اولیه	مستقل
۱۱	فاکتور خرید (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۱۲	خالص خرید (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۱۳	تعداد خط فاکتور خرید	numeric	Scale	متغیر پایه اولیه	مستقل



ردیف	نام متغیر	نوع متغیر	مقیاس اندازه‌گیری	وضعیت	نقش
۱۴	متوسط خرید هر فاکتور (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۱۵	کل تخفیفات (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۱۶	زمان سپری شده از آخرین خرید	numeric	Scale	متغیر پایه اولیه	مستقل
۱۷	تعداد فروش	numeric	Scale	متغیر پایه اولیه	مستقل
۱۸	مقدار برگشتی (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۱۹	تخفیفات کالایی (ریال)	numeric	Scale	متغیر پایه اولیه	مستقل
۲۰	مقدار تازگی خرید موزون	numeric	Scale	متغیر ثانویه	مستقل
۲۱	مقدار دفعات خرید موزون	numeric	Scale	متغیر ثانویه	مستقل
۲۲	مقدار ارزش پولی خرید موزون	numeric	Scale	متغیر ثانویه	مستقل
۲۳	مقدار WRFM	numeric	Scale	متغیر ثانویه	(هدف)

در این پژوهش ۱۹ متغیر اولیه وجود داشته که بر اساس این ۱۹ متغیر و متغیرهایی ثانویه که بعداً بر اساس مدل تازگی، تکرار و ارزش پولی موزون به پژوهش اضافه شده فرآیند داده‌کاوی انجام شده است. که از این بین، از زمان سپری شده از آخرین خرید مشتری برای محاسبه مقدار تازگی خرید (R)، از تعداد خط فاکتور فروش برای مقدار دفعات خرید (F) و همچنین از متوسط خرید هر فاکتور برای محاسبه مقدار ارزش پولی (M) استفاده شده است. مقدار نهایی WRFM برای هر داده که همان متغیر ردیف ۲۳ جدول ۳ است بر اساس رابطه ۱ بدست می‌آید.

$$\text{رابطه ۱)} \quad \text{WRFM}_v = \text{WR} \times \text{Recency}_v + \text{WF} \times \text{Frequency}_v + \text{WM} \times \text{Monetary}_v$$

$\text{W}_{R, F, M}$ وزن‌های هر یک از شاخص مدل تازگی، تکرار و ارزش پولی خرید و $V = \text{Value}$

مقدار هر یک از شاخص‌هاست. مقدار وزن شاخص‌های مدل تازگی، تکرار و ارزش پولی خرید مشتری به ترتیب ۰/۱۵، ۰/۲۹ و ۰/۵۶ بدست آمده است. مقدار شاخص ناسازگاری پژوهش ۰/۰۵۲ بدست آمده است که چون کمتر از ۰/۱ است پس می‌توان با اطمینان از این وزن‌ها برای این پژوهش استفاده کرد.

۱-۵- پیش پردازش داده‌ها

برای داده‌کاوی باید داده‌های پژوهش پیش پردازش شوند که این کار ابتدا با روش تشخیص مغایرت انجام گردیده است تا داده‌های پرت و دورافتاده را مشخص کرده و از کل داده‌ها حذف کنیم (چراکه روش داده‌کاوی k-میانگین به این نوع داده‌ها حساس است)؛ سپس از روش‌های



خوشه‌بندی در ادامه استفاده می‌شود. در جدول ۳ نتایج روش تشخیص مغایرت (ناهنجاری) آمده است.

جدول ۳. نتایج گروه‌بندی تشخیص مغایرت

نام روش	تعداد و گروه‌های غالب ^{۲۸}		تعداد داده‌های هرگروه به تفکیک	تعداد داده‌های مغایر	مقدار شاخص‌های مدل WRFM		انحراف استاندارد
	گروه غالب یک	گروه غالب دو			WR	WF	
تشخیص مغایرت	گروه غالب یک	۳	۱۳۵۰۷	۲۴۷	WR	۰/۱۷۹	۰/۰۸۴
					WF	۰/۹۸۸	۰/۰۴۴
					WM	۱/۸۶۴	۰/۸۳۸
	گروه غالب دو	۳	۲۴۴۵۵	۲۲۹	WR	۰/۵۶	۰/۱۶۶
					WF	۱/۱۳	۰/۲۷۴
					WM	۲/۲۸۹	۰/۴۳۱
	گروه غالب سه	۳	۲۶۸۹۶	۱۷۲	WR	۰/۴۸۷	۰/۱۷۶
					WF	۰/۶۳۴	۰/۳۶۸
					WM	۱/۰۱۷	۰/۴۳۱
جمع کل			۶۴۸	۶۴۸۵۸			

(منبع: یافته‌های پژوهش)

باتوجه به نتایج حاصل شده از روش گروه‌بندی تشخیص مغایرت جدول ۳، در این روش از مجموع ۲۳ متغیر اولیه و ثانویه پژوهش، ۲۲ متغیر به عنوان ورودی و از متغیر مقدار کل WRFM به عنوان متغیر خروجی برای عملیات تشخیص مغایرت که یک روش داده‌کاوی غیرنظارتی بوده استفاده شده است که در نهایت با توجه به مقدار برش (حد آستانه) شاخص مغایرت ۲/۴۶۴۵۸ که توسط خود سیستم محاسبه شده است سه گروه غالب تشخیص داده شد (در این پژوهش، معیار تشخیص مغایرت بر اساس درصد بیشترین رکوردهای مغایر در داده‌های یادگیری^{۲۹} است که در اینجا یک درصد به عنوان پیش فرض انتخاب شده است). در جدول ۳، گروه غالب سه دارای بیشترین فراوانی با ۲۶۸۹۶ رکورد بوده است همچنین کمترین فراوانی نیز مربوط به گروه غالب یک با ۱۳۰۵۷ رکورد است. با توجه به یافته‌های این بخش در کل ۶۴۸ رکورد یا داده مغایر از کل ۶۴۸۵۸ رکورد اولیه تشخیص داده شده است که این داده‌ها از کل داده‌ها حذف شده تا فرآیند خوشه‌بندی با دقت بالاتری انجام شود.

با توجه به مقادیر مختلف شاخص‌های مدل تازگی، تکرار و ارزش پولی موزون و مقادیر



انحراف استاندارد مربوط به هر شاخص در گروه‌های ایجاد شده، مقدار انحراف استاندارد شاخص ارزش پولی موزون (WM) نسبت به سایر مقادیر انحراف استاندارد بیشتر بوده که این می‌تواند به معنی تغییرپذیری زیاد مقادیر این شاخص باشد. در ادامه بعد از حذف داده‌های مغایر (۶۴۸ داده ناهنجار) از کل داده‌ها، تعداد ۶۴۲۱۰ داده باقیمانده است که از ۷۰ درصد این داده‌ها برای یادگیری و از مابقی ۳۰ درصد آن برای آزمون دقت مدل‌های خوشه‌بندی پژوهش، استفاده شده است.

در ادامه دو روش کارآمدتر خوشه‌بندی این پژوهش یعنی، k - میانگین و دومرحله‌ای را با توجه به تعداد تکرار خوشه‌بندی، میزان خطا، درصد خلوص و همچنین شاخص اطلاعات متقابل نرمال‌شده بررسی و مقایسه می‌شود تا از نتیجه این مقایسه روش بهینه و باکیفیت‌تر خوشه‌بندی تعیین و همچنین تعداد خوشه‌های بهینه پژوهش مشخص شود تا در ادامه از این نتایج برای تعیین ارزش مشتریان با توجه به هر خوشه اقدام شود.

۲-۵- یافته‌های خوشه‌بندی K - میانگین

در بیشتر پژوهش‌های انجام شده مانند: پژوهش‌های سیدحسینی و دیگران [۱۰]، بین و دیگران [۱۴]، رنگیز و بایرامی [۳] و کاباساکال [۲۳]، برای خوشه‌بندی از این روش به علت کارآمدتر بودن آن استفاده شده است بنابراین به علت فراوانی استفاده و کارایی بالای آن در بیشتر پژوهش‌های صورت گرفته، در این پژوهش از این روش نیز استفاده می‌شود. در این قسمت به خوشه‌بندی‌هایی که با روش k - میانگین با فرض ۳، ۴، ۵، ۶ و ۷ خوشه ایجاد شده است اشاره می‌شود. برای این روش تعداد ۲۰ تکرار^{۲۰} به عنوان حداکثر تکرار عمل خوشه‌بندی تارسیدن به نتیجه مربوطه انتخاب شده است. چنانچه تعداد خوشه‌های متفاوتی انتخاب شود (حداکثر ۲۰ تکرار) درصد خطاهای متفاوتی ایجاد می‌شود که شرح آن در جدول ۴ آمده است.

جدول ۴. خوشه‌بندی‌های متفاوت k -means

تعداد تکرار	درصد خطا	فراوانی هر خوشه	تعداد و شماره خوشه	تعداد تکرار*	درصد خطا	فراوانی هر خوشه	تعداد و شماره خوشه
۲۰	۰/۰۰۱	۸۰۰۵	۱	۲۰	۰/۰۰۱	۱۲۴۵۳	۱
		۱۰۸۷۴	۲			۲۳۳۲۷	۲
		۸۵۵۲	۳			۹۲۶۰	۳
		۱۷۳۰۹	۴			۴۵۰۴۰	جمع کل
		۴۵۰۴۰	جمع کل				



تعداد تکرار	درصد خطا	فراوانی هر خوشه	تعداد و شماره خوشه		تعداد تکرار*	درصد خطا	فراوانی هر خوشه	تعداد و شماره خوشه	
۱۳	۰/۰۰	۶۳۳۳	۱	۶	۱۹	۰/۰۰	۶۴۰۲	۱	۵
		۱۰۸۱۹	۲				۷۴۹۹	۲	
		۵۰۲۶	۳				۸۵۲۹	۳	
		۱۲۶۱۸	۴				۱۷۰۸۵	۴	
		۴۲۵	۵				۵۵۲۵	۵	
		۹۸۲۹	۶				۴۵۰۴۰	جمع کل	
		۴۵۰۴۰	جمع کل		تعداد تکرار	درصد خطا	فراوانی هر خوشه	تعداد و شماره خوشه	
*منظور از تعداد تکرار عملیات خوشه‌بندی است تا به خطای صفر برسد یا به حد تعیین شده برای تعداد تکرار عملیات خوشه‌بندی (هرکدام زودتر اتفاق بیفتد) عملیات خوشه‌بندی اتمام می‌یابد					۲۰	۰/۰۰	۶۳۰۹	۱	۷
							۵۴۱۱	۲	
							۳۴۹۷	۳	
							۱۲۰۱۵	۴	
							۳۷۹	۵	
							۹۵۲۴	۶	
							۷۹۰۵	۷	
							۴۵۰۴۰	جمع کل	

(منبع: یافته‌های پژوهش)

با توجه به جدول ۴ در خوشه‌بندی‌های مختلف K- میانگین با تعداد خوشه‌های متفاوت به جز سه خوشه‌ای خوشه چهارم دارای بیشترین فراوانی در هر گروه خوشه بوده. وقتی تعداد خوشه‌ها ۵،۶ یا ۷ انتخاب می‌شود خطای خوشه‌بندی صفر خواهد شد ولی از این بین تنها خوشه‌بندی شش‌تایی در تعداد تکرار کمتری به این میزان خطا دست پیدا کرده است که می‌تواند به عنوان گزینه ایده‌آل از نظر کیفیت خوشه‌بندی مورد بررسی قرارگیرد.



جدول ۵. مقایسه کیفیت خوشه‌بندی‌های K-means و TwoStep

روش خوشه‌بندی	تعداد خوشه یا گروه‌ها	تعداد تکرار تنظیم اولیه	میزان خطا ^{۳۱}	درصد خلوص	شاخص اطلاعات متقابل نرمال شده
TwoStep	۶	بدون داده‌های پرت ^{۳۲}		۰/۷۴۵۸	۰/۵۲۱
TwoStep	۲	با تنظیم‌های اولیه		۰/۵۸۷	۰/۱۸۳
K-Means	۳	۲۰	۰/۰۰۱	۰/۵۹۰	۰/۵۵
	۴	۲۰	۰/۰۰۱	۰/۵۶۴	۰/۴۲۸
	۵	۱۹	۰/۰	۰/۷۲۲	۰/۶۴۲
	۶	۱۳	۰/۰	۰/۷۴۴	۰/۶۳۱
	۷	۲۰	۰/۰	۰/۷۴۸	۰/۵۹۱

(منبع: یافته‌های پژوهش)

با توجه به جدول ۵ از بین خوشه‌بندی‌های مختلف دومرحله‌ای، خوشه‌بندی با شش خوشه انتخاب می‌شود چراکه درصد خلوص بالاتری داشته و شاخص اطلاعات متقابل نرمال شده آن ۰/۵۲۱ است (مقدار خلوص بین صفر و یک است؛ خوشه خالص دارای مقدار یک است و شاخص اطلاعات متقابل نرمال شده نیز عددی بین صفر و یک است و هر چقدر این شاخص به یک نزدیکتر باشد نشان از کیفیت بهتر خوشه‌بندی و ایجاد اطلاعات با ارزش‌تری است [۲۶]). از بین خوشه‌بندی‌های k - میانگین، خوشه‌بندی‌های پنج‌تایی، شش‌تایی و هفت‌تایی وضعیت بهتری داشته‌اند (از نظر شاخص تعداد تکرار خوشه‌بندی تا رسیدن به خطای صفر، میزان خطای خوشه‌بندی، درصد خلوص و شاخص اطلاعات متقابل نرمال شده). با مقایسه این چهار شاخص در این سه دسته خوشه‌بندی، خوشه‌بندی شش‌تایی در میزان تکرار کمتر و درصد خلوص نسبت به خوشه‌بندی پنج‌تایی وضعیت بهتری داشته است ولی خوشه‌بندی پنج‌تایی تنها در شاخص اطلاعات متقابل نرمال شده وضعیت بهتری داشته است؛ در مقایسه بین خوشه‌بندی شش‌تایی با هفت‌تایی، خوشه‌بندی شش‌تایی در دو شاخص تعداد تکرار کمتر تا رسیدن به درصد خطای صفر و شاخص اطلاعات متقابل نرمال شده تفاوت چشمگیری دارد ولی خوشه‌بندی هفت‌تایی تنها در درصد خلوص مقدار اندکی وضعیت بهتری داشته است. بین خوشه‌بندی هفت‌تایی و پنج‌تایی، خوشه‌بندی هفت‌تایی تنها در شاخص درصد خلوص وضعیت بهتری دارد در حالی که خوشه‌بندی پنج‌تایی در دو شاخص تعداد تکرار کمتر برای خوشه‌بندی تا رسیدن به خطای صفر و شاخص اطلاعات متقابل نرمال شده وضعیت بهتری داشته است که



در نهایت بین این سه تعداد خوشه‌بندی، خوشه‌بندی شش‌تایی نسبت به بقیه خوشه‌بندی بهتر و با کیفیت‌تری ایجاد کرده است.

از بین دو روش مختلف خوشه‌بندی k - میانگین و خوشه‌بندی دمرحله‌ای روش k - میانگین با تعداد ۶ خوشه به عنوان روش بهینه انتخاب می‌شود چراکه در شاخص اطلاعات متقابل نرمال شده تفاوت چشمگیری با هم دارند ولی از نظر شاخص درصد خلوص تفاوت چندانی با هم ندارد (تقریباً برابر بوده‌اند).

در جدول ۶ نتایج مقادیر مدل تازگی، تکرار و ارزش پولی موزون و پراکندگی آن در دو روش خوشه‌بندی k - میانگین و دمرحله‌ای آورده شده است (داده‌های این جدول بر اساس پارتیشن‌بندی^{۳۳} ۷۰ درصد داده‌های آموزشی و ۳۰ درصد داده‌های آزمون تهیه شده است).

جدول ۶. مقایسه مقادیر WRFM و پراکندگی آن در دو روش متفاوت خوشه‌بندی (منبع: یافته‌های پژوهش)

روش	پوشه	میانگین WRFM	Min	Max	پراکندگی	انحراف استاندارد ^{۳۳}	خطای استاندارد از میانگین	ضریب واریانس ^{۳۵} (درصد)
k- میانگین	۱	۶۳۲۳	۴/۱۲	۵	۰/۰۹۲	۰/۳۰۳	۰/۰۰۴	۶/۵۷
	۲	۱۰۸۱۹	۲/۰۲۳	۲/۱۲	۰/۴۴۸	۰/۶۶۹	۰/۰۰۶	۲۳/۰۷
	۳	۵۰۲۶	۴/۱۴۸	۴/۵۵	۰/۱۸۳	۰/۴۲۷	۰/۰۰۶	۱۰/۲۹
	۴	۱۲۶۱۸	۲/۱۶۸	۱/۳	۰/۳۱۲	۰/۵۵۹	۰/۰۰۵	۲۵/۷۸
	۵	۴۲۵	۳/۷۷۴	۱	۱/۲۹۵	۱/۱۳۸	۰/۰۵۵	۳۰/۱۵
	۶	۹۸۲۹	۳/۵۲۶	۳/۰۱	۰/۱	۰/۳۱۷	۰/۰۰۳	۸/۹۹
جمع کل		۴۵۰۴۰						
Two Step	۱	۸۰۱۰	۴/۵۳۸	۳/۵۵	۰/۰۷۲	۰/۲۶۸	۰/۰۰۳	۵/۹۱
	۲	۶۸۸۸	۳/۳۶۴	۱	۰/۹۸۱	۰/۹۹	۰/۰۱۲	۲۹/۴۳
	۳	۵۸۷۳	۲/۳۵۶	۳	۰/۲۲۲	۰/۴۷۱	۰/۰۰۶	۱۹/۹۹
	۴	۶۸۴۹	۲/۳۰۵	۱	۰/۲۴۴	۰/۴۹۴	۰/۰۰۶	۲۱/۴۳
	۵	۴۴۱۲	۲/۵۰۴	۲/۹۹	۰/۶۲۷	۰/۷۹۲	۰/۰۱۲	۳۱/۶۳
	۶	۸۲۷۳	۳/۵۲۷	۳/۰۱	۰/۰۹۸	۰/۳۱۴	۰/۰۰۳	۸/۹
جمع		۴۰۳۰۵						
جمع کل داده‌ها		۶۴۲۱۰	۳/۰۰۸	۱	۱/۲۶	۱/۱۲۳	۰/۰۰۴	۳۷/۳۳



با توجه به جدول ۶ که براساس بهترین خوشه‌های ایجاد شده از دو روش متفاوت خوشه-بندی تهیه شده است آنچه مشهود است در هر دو روش، خوشه اول بهترین خوشه بوده چراکه از نظر مقدار تازگی، تکرار و ارزش پولی موزون این شاخص عددی بین ۴/۵ تا ۴/۷ داشته است و این مقدار بالاتر از میانگین کل آن یعنی (۳/۰۰۸) است؛ همچنین با توجه به اینکه مقدار عددی در فاصله ۴ تا ۵ قرار گرفته است بنابراین می‌توان گفت اکثر مشتریان این خوشه، مشتریان طلایی هستند؛ همچنین از آنجاییکه ضریب واریانس این خوشه مقدار ۶/۵۷ درصد است از پراکندگی کمتری نسبت به سایر خوشه‌بندی‌ها برخوردار بوده است که می‌تواند نشان از یکدست بودن نوع مشتریان این خوشه داشته باشد (ضریب واریانس کوچکتر پراکندگی کمتر). در خوشه‌بندی ۶ تایی به روش K- میانگین خوشه اول و سوم که جمعاً دارای ۱۱۳۴۹ داده از ۴۵۰۴۰ داده بخش آموزشی این پژوهش را دربرمی‌گیرند با فراوانی تجمعی ۲۵/۲ درصد و با توجه به اینکه مقدار تازگی- تکرار- ارزش پولی موزون این دو خوشه بیشتر از میانگین کل بوده و در محدود ۴-۵ قرار گرفته‌اند این دو دسته با هم تقریباً با قانون ۲۰-۸۰ (بیست درصد مشتریان ۸۰ درصد درآمد شرکت را ایجاد می‌کنند) همخوانی داشته و این دودسته خوشه، خوشه‌های باارزشی برای شرکت به حساب می‌آیند. همانطور که از جدول ۱۰ مشهود است تنها خوشه اول، سوم، پنجم و ششم روش خوشه‌بندی k- میانگین مقدار تازگی، تکرار و ارزش پولی موزون آن‌ها از مقدار میانگین کل (۳/۰۰۸) بیشتر است که این می‌تواند نشان از متفاوت و با اهمیت بودن این خوشه‌ها برای شرکت باشد. در روش خوشه‌بندی دو مرحله‌ای از آنجایی که به هنگام محاسبه خوشه‌بندی‌ها تعدادی مقادیر خالی برای بعضی از داده‌های ثبت شده وجود داشته است که برای خوشه‌بندی‌ها استفاده نشده است که به‌عنوان خوشه خالی^{۳۶} در نظر گرفته شده است برای همین است که داده-های آموزشی دو روش خوشه‌بندی k- میانگین و دو مرحله‌ای با هم متفاوت است. در خوشه‌بندی دو مرحله‌ای خوشه اول، دوم و ششم از متوسط میانگین کل مقدار تازگی، تکرار و ارزش پولی موزون بیشتر بوده‌اند که این به نوعی اهمیت نسبی این سه خوشه را برای شرکت تداعی می‌کند. از بین این سه خوشه، خوشه اول با فراوانی ۸۰۱۰ و با مقدار تازگی، تکرار و ارزش پولی ۴/۵۳۸ و درصد فراوانی ۱۹/۸۷ درصد بهترین خوشه این روش بوده که از نظر ارزشی نیز بسیار مهم است چراکه تقریباً طبق قانون ۲۰-۸۰، این مشتریان ۸۰ درصد درآمد شرکت را ایجاد کرده‌اند که برای شرکت بسیار مهم است.

در این قسمت با استفاده از روشی که در پژوهش‌های باباییان و سرفرازی [۴] و چانگ



تسای^{۳۷} [۲۷] برای نام گذاری دسته خوشه‌ها از آن کمک گرفته‌اند در این پژوهش برای تجزیه و تحلیل مشتریان و رتبه‌بندی آن‌ها بر اساس شاخص‌های مدل تازگی، تکرار و ارزش پولی استفاده شده است (با این تفاوت که بعد وفاداری که با مدل LRFM همخوانی دارد در این پژوهش بررسی نشده است). در هر خوشه با توجه به مقایسه مقدار هر شاخص با مقدار میانگین کل آن شاخص اگر شاخص بزرگتر از میانگین کل باشد از علامت (↑) استفاده می‌شود و اگر این شاخص کمتر از میانگین باشد از علامت (↓)، به همین ترتیب برای شاخص‌های دیگر این عملیات تکرار می‌شود. برای تعیین رتبه هر شاخص در هر خوشه با توجه به خوشه‌های دیگر از اعداد ۱ تا ۶ برای این مقایسه استفاده می‌شود (عدد شش یعنی این خوشه در این شاخص دارای کمترین مقدار و بدترین رتبه و یک یعنی بالاترین مقدار و بهترین رتبه در بین شش خوشه داشته است و مابقی خوشه‌ها در بین این مقادیر قرار خواهند گرفت). در مورد شاخص تازگی خرید این قانون برعکس است یعنی هر چقدر این مقدار کوچکتر باشد بهتر است (عدد کوچکتر در شاخص تازگی موزون (WR) نشان از تازگی خرید مشتری دارد و عدد بزرگتر نشان از آن دارد که مشتری مدت طولانی است که از شرکت خرید نداشته است). رتبه (۱،۶،۶) یعنی این خوشه در دو شاخص تکرار و ارزش پولی خرید مدل RFM دارای بالاترین رتبه بوده است (وضعیت خوبی ندارد) ولی در شاخص تازگی خرید رتبه اول را دارد. همچنین بهترین دسته خوشه و مشتری، مشتری است که دارای کمترین مقدار از نظر شاخص تازگی خرید بوده و در دو شاخص دیگر دارای بالاترین مقدار خود است. در جدول ۷ به نتایج رتبه‌بندی بین خوشه‌ای و الگوهای ایجاد شده پرداخته می‌شود.

جدول ۷. رتبه‌بندی و تعیین الگوی خرید مشتریان

رتبه خوشه	میانگین WR	میانگین WF	میانگین WM	میانگین WRFM	رتبه شاخص (R,F,M) بین خوشه‌ها	الگوی خوشه-ها [۲۷] (R,F,M)	فراوانی
۱	۰/۶۲۷	۱/۳۳۵	۲/۶۵	۴/۶۱۲	(۶،۱،۱)	(۱،۱،۱)	۶۳۳۳
۲	۰/۲۲۲	۰/۶۷۱	۱/۱۳	۲/۰۲۳	(۲،۵،۵)	(۱،۱،۱)	۱۰۸۱۹
۳	۰/۲۰۱	۱/۳۲۶	۲/۶۲۶	۴/۱۵۳	(۱،۲،۲)	(۱،۱،۱)	۵۰۲۶
۴	۰/۵۸۳	۰/۶۱۵	۰/۹۷۳	۲/۱۷۱	(۵،۶،۶)	(۱،۱،۱)	۱۲۶۱۸
۵	۰/۳۸۲	۱/۱۴۷	۲/۱۹۹	۳/۷۲۸	(۳،۳،۳)	(۱،۱،۱)	۴۲۵



فرآوانی	الگوی خوشه- ها [۲۷] (R,F,M)	رتبه شاخص (R,F,M) بین خوشه‌ها	میانگین WRFM	میانگین WM	میانگین WF	میانگین WR	خوشه	ردیف
۹۸۲۹	(۱،۱،۱)	(۴،۴،۴)	۳/۵۲۵	۲/۰۰۷	۰/۹۷۴	۰/۵۴۴	۶	
۴۵۰۴۰	جمع(داده‌های یادگیری)							
۶۴۲۱۰			۳/۰۰۸	۱/۶۶۶	۰/۸۹۲	۰/۴۵	جمع کل	

(منبع: یافته‌های پژوهش)

از بین تمامی ۸ حالت الگوی ترکیبی که به واسطه وضعیت دوگانه (یا بزرگتر از میانگین و یا کوچکتر از میانگین) هر شاخص مدل تازگی، تکرار و ارزش پولی قابل تصور است؛ در این پژوهش طبق نتایج جدول ۷، تنها چهار الگو رویت شده است که عبارتند از: ۱- الگوی (۱،۱،۱)، که خوشه اول و ششم دارای این نوع الگویند؛ ۲- الگوی (۱،۱،۱)، که خوشه سوم و پنجم دارای این الگویند؛ ۳- الگوی (۱،۱،۱)، که خوشه چهار دارای این الگو است و ۴- الگوی (۱،۱،۱)، که خوشه دوم از این الگو تبعیت می‌کند. برای نامگذاری این چهار الگو که تعیین کننده تعداد بخش‌های بازار خواهد بود از ترکیب شیوه‌های مختلف نامگذاری در پژوهش‌های باباییان و سرفرازی [۴]، چانگ و تسای [۲۷]، و مهدیراجی و دیگران [۲۸] استفاده شده است. الگوی اول مشتریان بالقوه طلایی نامیده می‌شوند چراکه از نظر شاخص‌های دفعات خرید و ارزش پولی خرید دارای مقداری بیشتری از میانگین‌اند و همچنین از نظر شاخص تازگی خرید دارای مقدار بیشتر از میانگین بوده و این بدان معناست که در قدیم این مشتریان خریدهای خوبی داشته‌اند و این بخش از بازار تقریباً ۳۶ درصد معادل ۱۶۱۵۲ مشتری را شامل می‌شود. الگوی دوم که بهترین وضعیت را در بین سایر الگوها دارد یعنی مشتریان با تعداد دفعات خرید و حجم پولی بالا که تازگی نیز خریدهایی داشته‌اند این مشتریان را مشتریان خاص و کلیدی می‌نامیم که این بخش از بازار تقریباً ۱۲ درصد معادل ۵۴۵۱ نفر را شامل می‌شود؛ الگوی سوم، مشتریان نامطمئن و از دست‌رفته نامیده می‌شود چراکه این مشتریان خریدهای کمی داشته و متوسط ارزش ریالی خریدشان نیز کمتر از میانگین کل بوده است و چون خیلی وقت هست که خریدی نداشته‌اند مشتری از دست‌رفته شناخته می‌شوند که شامل ۱۲۶۱۸ مشتری معادل ۲۸ درصد از سهم بازار شرکت می‌شوند. در نهایت الگوی چهارم مشتریان نامطمئن جدید نامگذاری می‌شود چراکه به علت تازگی خرید و کم بودن دفعات خرید و ارزش ریالی خریدشان شرکت نسبت به آنها



شناخت کافی نداشته و این مشتریان برای شرکت جنبه تازه و ناآشنا دارد که این بخش از بازار ۱۰۸۱۹ مشتری معادل ۲۴ درصد از سهم بازار شرکت را در بر می‌گیرد.

۶- نتیجه‌گیری و پیشنهادها

در این پژوهش در راستای نظر جی کیگان [۹]، برای بخش‌بندی مشتریان، از خصوصیات رفتاری به همراه خصوصیات جمعیت شناختی مشتریان که در قالب ویژگی‌های اسمی نمود پیدا کرده استفاده شده است. این پژوهش مدل تازگی، تکرار و ارزش پولی موزون را برای تجزیه و تحلیل ارزش و رفتار خرید مشتریان معرفی کرده است که وزن شاخص‌های این مدل با توجه به نظر خبرگان و با استفاده از فرآیند تحلیل سلسله مراتبی و با کمک جدول مقایسات زوجی و با توجه به نرخ ناسازگاری ۰/۰۵۲ به ترتیب ۰/۱۵، ۰/۲۹ و ۰/۵۶ بدست آمد که این مقادیر نشان از اهمیت نسبی بالاتر شاخص ارزش پولی از دید خبرگان این پژوهش دارد؛ نتایج این قسمت (وزن‌دهی به شاخص‌ها برای ایجاد مدل تازگی، تکرار و ارزش پولی موزون) با پژوهش‌های [۱۰]، [۱۳]، [۱۴] که اذعان داشته‌اند برای هر صنعت باید با توجه به نظر خبرگان همان صنعت و با استفاده از فرآیند تحلیل سلسله مراتبی اوزان مدل RFM تعیین شود همخوانی داشته ولی با پژوهش‌های [۱۱]، [۱۲] که به برابری اوزان برای هر سه شاخص تازگی، تکرار و ارزش پولی تأکید داشته‌اند تفاوت دارد. از کل ۶۴۸۵۸ داده پایگاه داده‌های اولیه مشتریان شرکت، ۶۴۸ داده مغایر به روش الگوریتم تشخیص مغایرت با مقدار شاخص مغایرت ۲/۴۶۴۵۸ تشخیص داده شده که بعد از حذف این تعداد داده، عملیات خوشه‌بندی بر اساس مدل ارائه شده پژوهش انجام شد.

به طور کلی می‌توان گفت این پژوهش بیشترین همخوانی را از نظر متدولوژی پژوهش و نمونه آماری با پژوهش [۲۱] داشته است؛ ولی در پژوهش مذکور از شاخص ناسازگاری برای بررسی اعتبار اوزان استفاده نشده؛ همچنین ترتیب اولویت اهمیت شاخص‌های مدل RFM به ترتیب تازگی، ارزش پولی و تکرار خرید بوده است و تحلیل درون خوشه‌ای نیز برای مشتریان صورت نگرفته است و برای تعیین اهمیت متغیرهای پیش‌بینی برای خوشه‌بندی از روش تعیین ویژگی‌های مهم^{۳۸} استفاده کرده‌اند. از منظر بازار مورد بررسی با پژوهش‌های [۳]، [۲۲]، [۲۸] که از داده‌کاوی برای خوشه‌بندی مشتریان خدمات بانکداری استفاده کرده‌اند تفاوت آشکار دارد. همچنین در این پژوهش برای تعیین تعداد خوشه بهینه از شاخص‌های تعداد تکرار، درصد خطا،



درصد خلوص و شاخص اطلاعات متقابل نرمال شده استفاده شده است درحالی‌که در پژوهش [۲۳] از شاخص سیلوئت؛ [۲] از شاخص وارد؛ [۴] و [۱۰] از شاخص دیویس- بولدین برای شناسایی تعداد خوشه‌های اولیه بهینه استفاده کرده‌اند.

در اینجا بحثی که وجود دارد این است که از آنجایی‌که نتایج این پژوهش به اوزان بدست آمده (بر پایه نظر خبرگان و ماتریس مقایسات زوجی) درباره هر یک از شاخص‌های مدل تازگی، تکرار و ارزش پولی تا حدود بسیاری بستگی دارد پس تغییر در نظر خبرگان می‌تواند باعث جابه‌جاشدن نوع دسته‌بندی مشتریان در این پژوهش شود و همچنین این احتمال وجود دارد که اگر درزمینه‌های دیگری این پژوهش انجام شود ترتیب اولویت وزنی شاخص‌ها تغییر کند.

در نهایت با توجه به هدف پژوهش و بهره‌مندی از مدل تازگی، تکرار و ارزش پولی موزون و استفاده از روش‌های خوشه‌بندی دومرحله‌ای و خوشه‌بندی k - میانگین، ۶ خوشه با ۵ نوع مشتری شناسایی شدند که با استفاده از شیوه نامگذاری که در پژوهش‌های [۴]، [۲۷]، [۲۸] استفاده شده است ۴ عنوان کلی به شرح: مشتریان خاص و کلیدی به تعداد ۵۴۵۱ مشتری، مشتریان طلایی بالقوه با ۱۶۱۵۲ مشتری، مشتریان نامطمئن از دست‌رفته با ۱۲۶۱۸ مشتری و مشتریان نامطمئن جدید با ۱۰۸۱۹ مشتری برای بخش بندی مشتریان این بخش‌ها تعیین شد.

نتایج این پژوهش به مدیران شرکت خصوصاً مدیران فروش و بازاریابی کمک خواهد کرد تا بهتر و دقیقتر مشتریان خود را شناخته و الگوهای رفتاری خرید آن‌ها را درک نمایند همچنین بتوانند ارزش هر مشتری را بر اساس مدل پیشنهادی به درستی تشخیص داده و نسبت به مشتریانی که در وضعیت بحرانی قرار داشته و یا احتمال دارد در این دسته قرار بگیرند اقدامات پیشگیرانه و تشویقی را در پیش بگیرند.

به مدیران شرکت پیشنهاد می‌شود تمرکز بیشتری روی مشتریان خاص و کلیدی خود داشته باشند یعنی مشتریان خوشه سه و پنج و سعی کند این مشتریان را از طریق توسعه امکانات ارتباطی مناسب و ابزارهای تبلیغاتی و ارتقای فروش حفظ کنند چراکه این خوشه و مشتریانی که در این خوشه قرار گرفته‌اند خیلی ارزشمند بوده و جزو مشتریان خاص شرکت هستند که سهم عمده‌ای از درآمد شرکت را ایجاد می‌کنند.

در این پژوهش از مدل تازگی- تکرار- ارزش پولی موزون استفاده شده است که با توجه به نتایج بدست آمده می‌توان این موضوع را مطرح کرد که در این تحلیل تداوم (طول) مدت وفاداری مشتریان مورد بررسی قرارنگرفته است که اگر شرکت بخواهد مشتریانی را که وفادار



و با ارزش هستند را شناسایی نماید باید مطالعه جدیدی را با مدل WLRFM انجام دهد تا با بررسی نتایج آن متوجه شوند آیا تفاوت چشمگیری در مشتریان وفادار با ارزش و دیگر مشتریان وجود دارد یا خیر.

نتایج بدست آمده این پژوهش در مورد مشتریان یکی از شعب استان تهران بوده است که نتایج این پژوهش حتی با افزایش داده‌ها در پایگاه داده‌های مشتریان دستخوش تغییراتی خواهد شد و همچنین در شعب مختلف شرکت بنا به دلایل مختلفی چون داشتن سلیقه‌ها، الگوهای رفتاری خرید متفاوت و همچنین وجود تفاوت‌های آب و هوایی و اقلیمی لازم است شرکت به تناوب محاسبات خود را تکرار کرده تا نتایج دقیقتر و جامع‌تری را در اختیار داشته باشد. باتوجه به اینکه نتایج این پژوهش بر اساس اطلاعات سال‌های ۱۳۹۷-۱۳۹۸ بدست آمده است قطعاً با شیوع، گسترش و اپیدمی و پاندمی کرونا و موج‌های مختلف این بیماری در کنار تغییر شیوه‌های زندگی و الگوهای مصرف مشتریان، نتایج این پژوهش با تغییرهایی مواجه خواهد بود که این شرایط می‌تواند خود موضوعی برای پژوهش‌های آتی پژوهشگران باشد که نتایج آنرا با این پژوهش مقایسه نمایند.

۷- پی‌نوشت‌ها

۱. Customer Relationship Management(CRM)
۲. Segmentation
۳. Psychographic Segmentation
۴. Benefit Segmentation
۵. Recency, Frequency and Monetary Value (RFM)
۶. Blattberg R.C., et al.
۷. Data mining
۸. Anomaly detection
۹. Partitioning Methods
۱۰. Hierarchical Agglomerative
۱۱. Divisive Methods
۱۲. Density-based Methods
۱۳. Grid-based Methods
۱۴. Model-based Methods
۱۵. Hwang & Lee
۱۶. Text Rank
۱۷. P-Value
۱۸. Point of sale (POS)
۱۹. Kabasakal
۲۰. Length, Recency, Frequency, Monetary Value Model(LRFM)
۲۱. Fuzzy Analytic Hierarchy Process (FAHP)
۲۲. Excel
۲۳. SPSS
۲۴. Clementine
۲۵. Two Step
۲۶. Purity
۲۷. Normalized Mutual Information (NMI)
۲۸. Peer- Groups
۲۹. Training Data
۳۰. Iterations
۳۱. Error
۳۲. Exclude outliers
۳۳. Partition



۳۴. Standard Deviation(SD)
 ۳۵. Coefficient of Variance (CV)
 ۳۶. Null Cluster
۳۷. Chang, H. H., & Tsay, S. F
 ۳۸. Feature Selection

۸- منابع

- [۱] Ashouri, M., Sharifkhani, M., & Tarokh, M. J., Developing Customer Knowledge Management Process Model Using CRM Systems. Journal of Roshd-E-Fanavari, vol. ۱۰, ۲۰۱۴, pp. ۶۲-۶۸.
- [۲] Kafashpoor, A., Tavakoli, A. & Alizade Zavarem, A., Customer Segmentation According to Customer Life Time Value Using Data Mining Based on RFM Model, journal of Public Management Researches, vol. ۵, ۲۰۱۲, pp. ۶۳-۸۴.
- [۳] Rangriz, H., and Bayrami Shahrivar, Z., The Impact of E-CRM on Customer Loyalty Using Data Mining Techniques. IT Management Studies, vol. ۷, ۲۰۱۹, pp. ۱۷۵-۲۰۵.
- [۴] Babaiyan, V. & Sarfarazi, S., Analyzing Customers of South Khorasan Telecommunication Company with Expansion of RFM to LRFM Model. Journal of AI and Data Mining, Vol. ۷, ۲۰۱۹, pp. ۳۳۱-۳۴۰.
- [۵] Shokohyar, S., Rezaeian, A. & Boroufar, A., Identifying the customer behavior model in life insurance Sector using data mining. Management Research in Iran, Vol. ۲۰, ۲۰۱۶, pp. ۶۵-۹۴.
- [۶] Raesi Vanani, S., Raesi Vanani, I. & Taghavifard, M., A Model for Learners Segmentation and Educational Performance Improvement Using Data Mining Algorithms. IT Management Studies, Vol. ۹, ۲۰۲۰, pp. ۵-۳۸.
- [۷] Samadi Rad, B., Principles of customer orientation and marketing (looking at the role of human communication in customer orientation), Eighth Art Mag., Vol. ۲۱ & ۲۲, ۲۰۰۱, pp. ۹۴-۹۸.
- [۸] Salehi Sadaghiyani, J. & Akhavan, M., Customer Relationships Management, Accountants, Vol. ۲۱, ۲۰۰۶, pp. ۲۳-۳۶.
- [۹] Keegan, Warren J., Global Marketing Management. Translated and averaged by Abdolhamid Ebrahimi, (۲ed.) Iran Cultural Studies Center, Tehran, ۲۰۰۱, pp. ۱۹۳-۱۹۹.
- [۱۰] Hosseini, S.M.S., Maleki, A. & Gholamian, M.R., Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, Vol. ۳۷, ۲۰۱۰, pp. ۵۲۵۹-۵۲۶۴.
- [۱۱] Blattberg, R.C., Kim, B.D. & Neslin, S.A., RFM Analysis in Database Marketing. International Series in Quantitative Marketing, ۱۸, Springer, New York, ۲۰۰۸.
- [۱۲] Khajvand, M., Zolfaghari, K., Ashoori, S., & Alizadeh, S. Estimating Customer Life Time Value based on RFM analysis of customer purchase behavior: case



- study. *Procedia Computer Science*, Vol. ۳, ۲۰۱۱, pp. ۵۷-۶۳.
- [۱۳] Stone, B., *Successful direct marketing methods*, Lincolnwood, IL: NTC Business Books, ۱۹۹۵, pp. ۳۷-۵۷.
- [۱۴] Bin, D., Peiji, S., & Dan, Z., Data mining for needy students identify based on improved RFM model: a case study of university. *Proceedings of the ۲۰۰۸ International Conference on Information Management, Innovation Management, and Industrial Engineering*, vol. ۱, ۲۰۰۸, pp. ۲۴۴-۲۴۷.
- [۱۵] Sohrabi, B., RaeesiVanani, I. & ZarehMirkabad, F., "Designing a Recommender System for Optimizing and Managing Bank Facilities through the Utilization of Clustering and Classification Algorithms". *Modern Research in Decision Making*, Vol. ۱, ۲۰۱۶, pp. ۵۳-۷۶.
- [۱۶] Shahrabi, J., *Data Mining*, Soroush Gita, Tehran, ۲۰۱۵.
- [۱۷] Alizadeh, S. & Malekmohammadi, S., *Data Mining and knowledge discovery step by step by Clementine Software*, Khajeh NasiruddinTusi University of Technology Press. Tehran, ۲۰۱۴.
- [۱۸] Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and techniques* (۳rd Ed.), Elsevier, ۲۰۱۱.
- [۱۹] Bhojani, S. & Bhatt, N., *Data Mining Techniques and Trends – A Review*. *Global Journal for Research Analysis (GJRA)*, Vol. ۵, ۲۰۱۶, pp. ۲۵۲-۲۵۴.
- [۲۰] Hwang, S. & Lee, Y., Identifying customer priority for new products in target marketing: Using RFM model and Text Rank, *Innovative Marketing*, Vol. ۱۷, ۲۰۲۱, pp. ۱۲۵-۱۳۶.
- [۲۱] Bashardoust, O., Asgharizadeh, E. & AfsharKazemi, M., Presenting a customer classification Pattern with a combined data mining approach (case study: Hygienic and Cosmetic products Industry), vol. ۱۳, ۲۰۲۱, pp. ۸۵-۱۱۱.
- [۲۲] Abbasimehr, H. and Shabani, M. "A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers", *Kybernetes*, Vol. ۵۰, ۲۰۲۱, pp. ۲۲۱-۲۴۲.
- [۲۳] Kabasakal, İnanç. *Customer Segmentation Based On Recency Frequency Monetary Model: A Case Study in E-Retailing*. Vol. ۱۳, ۲۰۲۰, pp. ۴۷-۵۶.
- [۲۴] Salehi, M. & Salari, M. "Comparing data mining and fuzzy logic techniques to identify behavior of customers", *Modern Research in Decision Making*, Vol. ۲, ۲۰۱۷, pp. ۱۷۳-۱۹۲.
- [۲۵] Bashirimosavi, S.A., Afsar, A. and MahjoubiFard, A., Bank customers value analysis using data mining techniques and fuzzy hierarchical analysis, *Management Researches in Iran*, Vol. ۱۹, ۲۰۱۵, pp. ۲۳-۴۳.
- [۲۶] Veysi, H., *Booklet of Statistical Methods in Natural Language Processing (Clustering)*. Faculty of Modern Sciences and Technologies, University of Tehran. ۲۰۱۷.
- [۲۷] Chang, H. H., & Tsay, S. F., Integrating of SOM and K-means in data mining clustering: An empirical study of CRM and profitability evaluation *Journal of Information Management*, Vol. ۱۱, ۲۰۰۴, pp. ۱۶۱-۲۰۳.



- [۲۸] Mahdiraji, H.A., Zavadskas, E.K., Kazemina, A., Abbasi Kamardi, A. Marketing strategies evaluation based on big data analysis: a CLUSTERING-MCDM approach, *Economic Research-Ekonomska Istraživanja*, Vol. ۳۲, ۲۰۱۹, pp. ۲۸۸۲-۲۸۹۸.