

کشف الگوهای رفتار خرید مشتریان با استفاده ترکیبی از خوشه‌بندی و قوانین انجمنی

(مورد مطالعه: محصولات شوینده و بهداشتی)

امید بشردوست^۱، عزت اله اصغری زاده^{۲*}^۱ دکترای تخصصی، گروه مدیریت صنعتی، دانشکده علوم تربیتی مشاوره (مدیریت و حسابداری)، واحد رودهن،

دانشگاه آزاد اسلامی، رودهن، ایران.

^۲ دانشیار، مدیریت صنعتی، دانشکده مدیریت دانشگاه تهران، تهران، ایران. (نویسنده مسئول)

تاریخ دریافت: 1402/03/04 تاریخ پذیرش: 1402/05/16

Extract the patterns of customers buying behavior using a combination of clustering and association rules

(Case study: hygienic and detergent products)

Omid Bashardoust¹, Ezattollah Asgharizadeh^{2*}¹ Department of Industrial Management, Faculty of Education Science and Counseling (Management and Accounting) Roudehen Branch, Islamic Azad University, Roudehen. Iran.² Department of Industrial Management, Faculty of Management, Tehran University. Tehran, Iran. (corresponding author)

Received: (25/05/2023)

Accepted: (07/06/2023)

شناسه یکتا: asm.pgu.ac.ir/article_706598

Abstract

The present study is of applied-developmental type and has a descriptive exploratory nature that seeks to discover customer behavior patterns using a combined approach of clustering and association rules to identify and determine different segments of customers can meet their needs and expectations and ultimately lead to profitability. The statistical population of this research is retail, consumer cooperative and local customers of a company active in the supply of hygienic and detergent products in the period 2021-2022, of which 65,534 samples have been collected by purposive sampling method. Analysis and data mining was performed using SPSS Modeler software. To determine customer segments, the K-Means method with 5 clusters and a silhouette value of 0.2 and appropriate quality has been selected. Finally, with this number of clusters and using the Apriori algorithm with the assumption of at least 60% confidence, 35 general rules for the sample customers of this research have been obtained So that managers can easily identify the buying behaviors and patterns of their key customers and other customers by using data mining tools in a very short period of time, and based on this, the marketing and sales strategies of their company, prioritize and plan can design and plan their marketing and customer relationship management strategies based on this.

KeyWords: Association Rules, Clustering, Customer relationship Management (CRM), Data Mining, Silhouette index

چکیده

پژوهش حاضر از نوع کاربردی- توسعه‌ای بوده و دارای ماهیت توصیفی اکتشافی است که به دنبال کشف الگوهای رفتاری خرید مشتریان به کمک رویکرد ترکیبی خوشه‌بندی و قوانین انجمنی است تا ضمن شناسایی و تعیین بخش‌های مختلف مشتریان بتواند نیازها و انتظارات آن‌ها را برآورده نموده و در نهایت به سودآوری منجر شود؛ جامعه آماری این پژوهش مشتریان خرده‌فروشی، تعاونی مصرف و محلی شرکتی فعال در عرضه محصولات بهداشتی و شوینده در بازه زمانی ۱۴۰۰-۱۴۰۱ است که تعداد ۶۵۵۳۴ نمونه به روش نمونه‌گیری هدفمند در دسترس گردآوری شده است. تجزیه و تحلیل و داده‌کاوی به کمک نرم‌افزار SPSS Modeler انجام شده است. برای تعیین بخش‌بندی‌های مشتریان روش K-Means با تعداد ۵ خوشه و با مقدار سیلوئت ۰/۲ و کیفیت مناسب انتخاب شده است. در نهایت با این تعداد خوشه و با استفاده از الگوریتم Apriori با فرض حداقل اطمینان ۶۰ درصد ۳۵ قانون کلی برای مشتریان نمونه این پژوهش حاصل شده است که مدیران به راحتی بتوانند با بهره‌گیری از پایگاه داده‌های مشتریان خود و با استفاده از ابزار داده‌کاوی در بازه زمانی بسیار کوتاهی رفتارها و الگوهای رفتاری خرید مشتریان کلیدی و سایر مشتریان خود را شناسایی کرده و بر این اساس راهبردهای بازاریابی و فروش شرکت خود را اولویت‌بندی و برنامه‌ریزی نمایند.

واژگان کلیدی: خوشه‌بندی، داده‌کاوی، شاخص سیلوئت، قوانین وابستگی،

مدیریت ارتباط با مشتری (CRM)

مقدمه

چالش روند فزاینده تغییرات محیطی، میزان رقابت و گذر از دوران انحصاری به فضای رقابت، شرکت‌ها را به سمت بازاریابی پویا (بازاریابی هدفمند) کشانده است. با توجه به این که ارتقای سطح رضایتمندی مشتریان و همچنین افزایش سودآوری و رشد پایدار از راهبردهای اصلی شرکت‌ها به شمار می‌روند استفاده از ابزارهایی که در این راستا باعث سهولت فرآیند تصمیم‌گیری شود بیش از پیش اهمیت یافته است (اخوان و همکاران، ۱۳۹۹: ۱). شناخت مشتریان و کشف الگوهای رفتاری خرید آن‌ها و ارائه راهبردهای بازاریابی متناسب با نیاز، خواسته و سلیقه‌های در حال تغییر مشتریان که به واسطه رشد و تغییرات سریع فناوری‌های ارتباطی و اطلاعاتی و رقابتی تر شدت عرصه کسب و کارهای مختلف ایجاد شده است برای هر سازمان به منظور کسب نتایج مؤثر و ایجاد مزیت رقابتی بیش از پیش اهمیت یافته است. می‌توان عنوان کرد افزایش سرعت پردازش اطلاعات توسط نرم‌افزارهای مختلف آماری و تحلیل داده‌ها و همچنین انباشت اطلاعات جمعیت شناختی مشتریان در کنار اطلاعات خرید هر مشتری در پایگاه داده‌ای شرکت‌ها را فراهم نموده است تا شرکت بتواند بدون نیاز به انجام فرآیندهای زمانبر و هزینه‌بر تحقیقات بازاریابی سنتی به تحلیل الگوهای رفتاری هر مشتری که جنبه پویا نیز داشته پردازد. در عرصه کنونی کسب و کار بهره‌مندی از ابزارهای مؤثر و هوشمند مدیریت ارتباط با مشتری می‌تواند جوابگوی نیازهای روز سازمان‌ها و شرکت‌ها در عرضه خدمات و محصولات متناسب با نیاز و سلیقه‌های هر مشتری باشد.

برای اینکه بتوان مدیریت ارتباط با مشتری مؤثری داشت می‌بایست درباره ارزش مشتری و بخش‌بندی آن‌ها اطلاعاتی را جمع‌آوری نمود تا راحت‌تر و سریعتر پاسخگوی نیازهای منحصر به فرد هر بخش از مشتریان بود. بازار رقابتی، امروزه شاهد تغییر و تحولات پیچیده‌ای است و حاوی ویژگی‌های خاصی مانند: تکرار خرید مشتریان در بازه‌های زمانی مختلف، حجم بالای خرید مشتریان، اطلاعات با ارزش از رفتار خرید مشتریان است. در این گونه از بازارها درک و پیش‌بینی الگوی خرید و شناسایی نیازهای مشتریان و عرضه متناسب با خواسته و انتظارات مشتری از اولویت‌های هر کسب و کاری است (ها و بئه^۱، ۲۰۰۶).

بسیاری از پژوهشگران و مدیران بازاریابی بخش‌بندی بازار و بازاریابی را از جمله مهم‌ترین مقوله‌های اصلی در حوزه مدیریت ارتباط با مشتری دانسته‌اند که بر این موضوع نیز تأکید بسیار شده است. هدف اصلی در بخش‌بندی مشتریان، تخصیص منابع محدود در اختیار سازمان و شرکت برای شناسایی سودآورترین مشتریان و تحلیل رفتار، نیازها و اولویت‌های آنهاست (تسیپتسیس و چریانوپولوس^۲، ۲۰۰۹ و دیب^۳، ۱۹۹۸).

در کنار مباحث مدیریت ارتباط با مشتری تسهیلات استفاده از شبکه اینترنت باعث انباشت حجم زیادی از داده و اطلاعات شده است (رنگریز و بایرامی، ۱۳۹۸). با توجه به حجم گسترده انباشت داده در پایگاه داده‌ها، استفاده از راهکارها و ابزارهای مناسب برای دسته‌بندی و تولید اطلاعات از میان انبوهی از داده‌ها برای سازمان امری حیاتی است. داده‌کاوی پاسخی به این نیاز به منظور استخراج دانش پنهان درون داده‌ها و تأمین اطلاعات مورد نیاز است (ریسی و انانی، ریسی و تقوی فرد، ۱۳۹۹: ۷).

^۱ Ha, S.H. & Bae, S.M.

^۲ Tsipitsis, K. & Chorianopoulos, A.

^۳ Dibb, S.

داده کاوی از طریق جستجو روابط و الگوها از داده‌های جمع آوری شده توسط سازمان موجب کاهش داده‌های اضافی می‌شود. یکی از گسترده‌ترین زمینه‌های کاربرد داده کاوی مدیریت ارتباط با مشتری است. با مدیریت مؤثر ارتباط با مشتری به کمک تجزیه و تحلیل پایگاه داده مشتریان و با تأکید بر ارائه خدماتی که مشتریان واقعاً خواهان آن هستند، می‌توان در هزینه‌ها صرفه جویی نمود و بیشتر بر مشتریان خود متمرکز شد و بر پایه تحلیل رفتار مشتریان می‌توان ریسک حاصل از سرمایه گذاری را پیش‌بینی و کاهش داد و بخش‌بندی مناسبی از نیازمندی‌های مشتریان به دست آورد. با این شیوه ضریب پذیرش خدمات نیز افزایش خواهد یافت (تقوا و حسینی بامکان، ۱۳۹۰: ۱۹۲-۱۹۳).

بنابر مطالب عنوان شده و اهمیت کشف و شناخت الگوهای رفتاری خرید مشتریان و کارآمدی ابزارهای داده کاوی در مدیریت ارتباط با مشتری این پژوهش در صدد آن است با استفاده از ابزارهای ترکیبی خوشه‌بندی و قوانین انجمنی داده کاوی به شناخت دانش مشتری و استخراج الگوهای رفتاری پویای خرید مشتریان در عرصه محصولات بهداشتی و شوینده که از تنوع بازاری برخوردار است بپردازد تا مشخص نماید چه الگوهای رفتاری خریدی در هر خوشه مشتریان حاکم است تا با شناسایی این الگوها بتوان خدمات تخصصی تر و متناسب با سلیقه‌ها و علایق مشتریان به آن‌ها ارائه کرد؛ در این راستا اهداف فرعی این پژوهش را می‌توان در قالب سؤال‌های ذیل بیان کرد:

۱. الگوهای غالب رفتاری خرید درون هر خوشه مشتریان چگونه است؟
۲. کدام روش خوشه‌بندی این پژوهش برای بخش‌بندی مشتریان محصولات بهداشتی و شوینده با توجه به شاخص‌های نسبت اندازه و معیار سیلوئت مناسبتر است؟

مبانی نظری پژوهش

در ادامه به معرفی داده کاوی و روش‌های مختلف خوشه‌بندی آن پرداخته شده و بعد از آشنایی با فضای داده کاوی به پژوهش‌های داخلی و خارجی که در زمینه به کارگیری داده کاوی برای بخش‌بندی مشتریان صورت گرفته اشاره می‌شود.

داده کاوی

ایده‌های اصلی داده کاوی برای مدیریت ارتباط با مشتری این است که داده‌های تاریخی دربرگیرنده اطلاعات مفید درباره آینده رفتار مشتریان هستند چرا که رفتار مشتریان در قالب داده‌های شرکت نشان می‌دهد این رفتارها، اتفاقی نیستند بلکه نیازها، تمایلات، سلیقه‌ها و عملکردهای مشتریان را نشان می‌دهد. شناسایی الگوهای اساسی و مفید در درون متغیرهای به ظاهر اتفاقی، یکی از نقش‌های مهم داده کاوی است (بهوجانی و بهات^۱، ۲۰۱۶: ۲۵۲). از ویژگی‌های خاص و مهم فرآیند داده کاوی، ماهیت پویا و دینامیکی آن است (شهرابی، ۱۳۹۴: ۷۸-۷۹). تکنیک‌های داده کاوی را می‌توان به منظور ساخت سه نوع مدل برای سه فعالیت به کار برد: ۱- نمایه‌سازی توصیفی، ۲- نمایه‌سازی هدایت شده، ۳- پیش‌بینی (آزمون-های فرضیه معمولاً منجر به تولید مدل‌های توصیفی می‌شود) (شهرابی، ۱۳۹۴).

^۱ Bhojani, S. & Bhatt, N

خوشه بندی

خوشه‌بندی یک تکنیک یادگیری بدون نظارت است که اشیا بر اساس اصل بیشترین شباهت بین اعضای هر خوشه و کمترین شباهت بین خوشه‌های مختلف گروه‌بندی می‌شوند (هان و کمبر^۱، ۲۰۱۰) که با این روش به تصمیم‌گیرندگان کمک می‌کند تا بخش‌های بازار را با وضوح بیشتری شناسایی کنند و در نتیجه استراتژی‌های بازاریابی و فروش مؤثرتری برای حفظ مشتری ایجاد کنند (جوتمات و موری^۲، ۲۰۱۹: ۱۸). روش‌های خوشه‌بندی به روش‌های سلسله‌مراتبی^۳، پارتیشن‌بندی، مبتنی بر چگالی^۴، مبتنی بر مدل، مبتنی بر شبکه و محاسبات نرم^۵ تقسیم می‌شوند (روکاچ و مایمن^۶، ۲۰۰۵: ۳۲۱). اخیراً یک مدل خوشه‌بندی بدون نظارت برای کشف قوانین کمیاب جالب پیشنهاد شده است اما عملکرد این مدل از نظر زمان و دقت کاهش یافته است (آخوندزاده و همکاران، ۱۳۹۴: ۹).

بخش بندی مشتریان

از اوایل دهه ۱۹۸۰ مفهوم مدیریت ارتباط با مشتری در حوزه بازاریابی اهمیت پیدا کرده است. شاید به توان گفت قویترین ابزارهای مورد استفاده برای بازاریابی، پیش‌بینی رفتار خرید و بخش‌بندی مشتریان است که از این طریق می‌توان خریداران را از غیر خریداران مجزا نمود و گروه‌های مشتریان را شناسایی و آن‌ها را از یکدیگر متمایز کرد. هنگام برخورد با مجموعه داده‌های بزرگ، بخش‌بندی بازار اغلب در پیش‌بینی کسب و کار استفاده می‌شود. بسیاری از مشتریان بر اساس معیارهایی از شباهت گروه‌بندی می‌شوند. سپس، از پیش‌بینی در سطح بخش برای نشان دادن جمعیت در هر بخش استفاده می‌شود (موری و همکاران^۷، ۲۰۱۸: ۳۳). هر مشتری ارزش متفاوتی ایجاد می‌کند. تقسیم‌بندی مشتری به کسب و کار کمک می‌کند تا ارزش مشتریان مختلف و نیازهای آنها را به منظور شناسایی استراتژی‌های بازاریابی مختلف یا مدیریت ارتباط با مشتری درک کند. گروه‌بندی بر اساس ارزش مشتری تنها بر اساس درآمد مشتری نیست، بلکه می‌تواند عوامل مختلف را در بر گیرد (جوتمات و موری، ۲۰۱۹: ۱۸). بخش‌بندی مشتریان به شناسایی مشتریان با ویژگی‌های مشابه گفته شده که توسط بازاریابان برای هدفگذاری مؤثر و تخصیص بهینه منابع مورد استفاده قرار می‌گیرد (گریفین^۸، ۲۰۰۳). بخش‌بندی بازار از دو مرحله مجزا تشکیل شده است: اندازه‌گیری تفاوت بین مشتریان (معروف به اندازه‌گیری فاصله) و سپس بخش‌های مشابه ایجاد کردن (فرآیندی که به عنوان خوشه‌بندی شناخته می‌شود) (موری و همکاران، ۲۰۱۸: ۳۴). روش‌های زیادی برای تقسیم بازار به بخش‌ها وجود دارد (چاکرابورتی^۹، ۲۰۱۳)؛ اما به طور کلی، روش‌ها یا بر اساس ویژگی‌های توصیفی یا ویژگی‌های رفتاری هستند (موری و آگارد و باراجاس^{۱۰}، ۲۰۱۷).

^۱ Han, J. & Kamber, M.

^۲ Jutamat, J. and Mori, T.

^۳ Hierarchical clustering

^۴ Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

^۵ Soft computing

^۶ Rokach L., Maimon O.

^۷ Murray et al.

^۸ Griffin, J.

^۹ Chakraborty

^{۱۰} Murray, Agard & Barajas

راه‌های زیادی برای تقسیم‌بندی بازارها برای یافتن مخاطب هدف مناسب وجود دارد. پنج راه برای تقسیم‌بندی بازارها شامل تقسیم‌بندی جمعیت‌شناختی^۱، روانشناختی^۲، رفتاری^۳، جغرافیایی^۴ و شرکت‌شناختی^۵ است.

بر اساس معیارهای تقسیم‌بندی و اهداف هر حیطه کاری بخش‌بندی مشتریان می‌تواند شامل موارد ذیل باشد: ۱- بخش‌بندی بر اساس ارزش^۶ (با استفاده از این روش مشتریان با ارزش و مهم شناسایی شده و تغییرات ارزش آن‌ها را در طول زمان می‌توان تحلیل نمود)؛ ۲- بخش‌بندی بر اساس رفتار^۷؛ ۳- بخش‌بندی بر اساس احتمال رفتار آتی مشتریان^۸ (مشتریان بر اساس میزان رویگردانی و میزان ورشکستگی و... بخش‌بندی شده که به کمک مدل‌های رده‌بندی تخمین زده می‌شوند)؛ ۴- بخش‌بندی بر اساس وفاداری^۹ (مشتریان وفادار و رویگردان را مشخص می‌کند)؛ ۵- بخش‌بندی بر اساس نیاز و خواسته مشتریان (بر اساس اطلاعات تحقیقات بازار، نیاز، خواسته و علائق مشتری شناسایی شده و بر این اساس گروه‌بندی می‌شوند) (تقوی فرد و همکاران، ۱۴۰۰).

تقسیم‌بندی بازار از دو مرحله مجزا تشکیل شده است (موری و همکاران، ۲۰۱۸: ۳۴): اندازه‌گیری تفاوت بین مشتریان (معروف به اندازه‌گیری فاصله) و سپس ایجاد بخش‌های مشابه (فرآیندی که به عنوان خوشه‌بندی شناخته می‌شود). روش‌های زیادی برای تقسیم بازار به بخش‌ها وجود دارد (چاکرابورتی، ۲۰۱۳؛ هان و کامبر، ۲۰۰۶؛ لی و همکاران، ۲۰۰۹)، اما به طور کلی، روش‌ها بر اساس ویژگی‌های توصیفی یا ویژگی‌های رفتاری است (موری و همکاران، ۲۰۱۷). از ویژگی‌های توصیفی (مانند اندازه، مکان، جنس، سن یا فراوانی تراکنش) معمولاً استفاده می‌شود زیرا این نوع متغیرها به راحتی قابل کمی‌سازی است. موفقیت در تقسیم‌بندی بر اساس ویژگی‌های توصیفی به شدت بر دو فرض متکی است: اینکه داده‌ها برای پشتیبانی از متغیرها در دسترس هستند و اینکه ویژگی‌های توصیفی واقعاً برای ایجاد بخش‌ها مرتبط هستند (گورعلی و پینار^{۱۰}، ۲۰۱۶). عوامل خارجی زیادی مانند اقتصاد، اقدامات رقبا و ادراک اجتماعی وجود دارند که رفتار مشتری تأثیر می‌گذارد (سولیان و راجاکوپال^{۱۱}، ۲۰۱۸)، با این حال، انتخاب عوامل مناسب می‌تواند چالش برانگیز باشد. بخش‌بندی بر اساس ویژگی‌های رفتاری نیازمند داده‌های تاریخی است که رفتار و روشی برای استخراج و شناسایی رفتار درون داده‌ها را نشان می‌دهد (باراگان، فونتس و امبیروچو^{۱۲}، ۲۰۱۶؛ کاشوان و ولو^{۱۳}، ۲۰۱۳). البته امروزه کار بندی بازار و مشتریان و یا تعیین بخش‌های مشتریان با استفاده از نرم‌افزارهای داده‌کاوی و با بهره‌مندی از الگوریتم‌های خوشه‌بندی K-میانگین، دومرحله‌ای، دومرحله‌ای باگزینش اتوماتیک^{۱۴}، شبکه عصبی کوهنن^{۱۵} (نگاشت‌های

^۱ demographic

^۲ psychographic

^۳ behavioral

^۴ geographic

^۵ Firmographic

^۶ Value Based Segmentation

^۷ Behavioral Segmentation

^۸ Needs/ attitudinal Segmentation

^۹ loyalty based Segmentation

^{۱۰} Gur Ali, O., & Pinar, E.

^{۱۱} Sevlian, R., & Rajagopal, R.

^{۱۲} Barragan, J. F., Fontes, C. H., & Embiruçu, M.

^{۱۳} Kashwan, K. R., & Velu, C. M.

^{۱۴} TwoStep-AS

^{۱۵} Kohonen Neural Net(KNN)

(مورد مطالعه: محصولات شوینده و بهداشتی)

خودسازمانده^{۱)}، خوشه‌بندی عصبی فازی، خوشه‌بندی بر پایه چگالی، سلسله مراتبی، تشخیص مغایرت^{۲)} و... به راحتی و با کمترین زمان ممکن امکان‌پذیر شده است. در سال‌های اخیر مدل‌های مختلفی برای بخش‌بندی و تعیین گروه مشتریان معرفی، بازبینی و به روز رسانی شده‌اند که هر یک دارای مزایایی بوده‌اند که به تعدادی از آنها اشاره می‌شود: مثلاً مدل تازگی، تکرار و ارزش پولی خرید (RFM)^{۳)} (بلت برگ و همکاران، ۲۰۰۸)، مدل تازگی، تکرار و ارزش پولی موزون (WRFM)^{۴)} (بشردوست و همکاران، ۱۴۰۱)، مدل وفاداری، تازگی، تکرار و ارزش پولی (LRFM)^{۵)} (باباییان و سرفرازی، ۲۰۱۹) و... که این مدل‌ها بر اساس اینکه مشتری به تازگی از شرکت خرید کرده، چند بار خرید داشته و حجم خرید آنها چگونه بوده و فاصله بین اولین و آخرین خرید چقدر بوده است (معیاری برای وفاداری مشتریان) پایه ریزی شده است.

پژوهشگران و مدیران بازاریابی در بسیاری از صنایع مختلف به بخش‌بندی مشتری به عنوان فعالیتی استراتژیک اهمیت داده‌اند؛ برای مثال می‌توان به بخش‌بندی مشتریان فروشگاه‌های زنجیره‌ای (غفاری و داودی، ۱۳۹۵)، مشتریان بانک‌ها و صنعت بانکداری (خواجوند و همکاران، ۱۳۹۱؛ غریب و همکاران، ۱۳۹۸؛ تقوی فرد و همکاران، ۱۴۰۰؛ تیموریان و همکاران، ۱۴۰۲)، مشتریان خدمات الکترونیکی (ایزدی و همکاران، ۱۳۹۵) مشتریان رستوران‌های زنجیره‌ای (حسین‌زاده شهری و همکاران، ۱۳۹۴) و یا در تشخیص و گروه‌بندی بیماران (محمودی و همکاران، ۱۳۹۶؛ ویشناوردهان و لکشیمپاد^{۶)}، ۲۰۱۸) اشاره کرد.

قوانین انجمنی^{۷)}

یکی از روش‌های توصیفی و غیرنظارتی داده‌کاوی قوانین انجمنی است. این روش با هدف کمی‌سازی ارتباط بین ویژگی‌ها به مطالعه ویژگی‌هایی که همراه یکدیگرند می‌پردازد که به شکل قوانین اگر و آنگاه بیان می‌شود (هان و کمبر، ۲۰۱۰).

پیشینه تجربی پژوهش

در این قسمت به تعدادی از پژوهش‌هایی که از روش‌های ترکیبی داده‌کاوی به منظور خوشه‌بندی و تعیین بخش‌های بازار مشتریان و تحلیل الگوهای رفتار خرید مشتریان استفاده کرده‌اند در قالب جدول (۱) اشاره مختصری می‌شود:

جدول ۱. پیشینه پژوهش‌های انجام شده

ردیف	سال	نویسندگان	عنوان / هدف	روش / ابزار	نتایج
۱	۱۳۹۵	ایزدی و همکاران	بخش‌بندی و طبقه‌بندی مشتریان بازارهای خدمات الکترونیکی	داده‌کاوی تجمیعی، دلفی فازی، خوشه-بندی، شبکه عصبی، یادگیری ماشینی، ماشین بردار پشتیبان،	کارآمدی روش داده‌کاوی برنامه‌ریزی خطی در دسته‌بندی مشتریان

^۱ Self-Organizing Maps(SOM)

^۲ Anomaly Detection

^۳ Recency, Frequency and Monetary (RFM)

^۴ weighted recency frequency monetary value (WRFM)

^۵ Length, Recency, Frequency, Monetary (LRFM)

^۶ LakshmiPadmaja, D. and Vishnuvardhan, B.

^۷ Association Rules

	برنامه ریزی خطی - معیار سیلوئت				
۲	۱۳۹۵	غفاری و داودی	بررسی و تحلیل سبد بازار و چیدمان محصولات یکی از شعب فروشگاه‌های زنجیره‌ای اراک با رویکرد داده کاوی	الگوریتم اجتماع، قواعد وابستگی، Apriori، روش استاندارد Crisp-DM	استفاده از داده کاوی قواعد انجمنی، افزایش رضایتمندی مشتریان و فروش سازمان و آرایه راهکارهای بهبود
۳	۱۳۹۶	محمودی و همکاران	بررسی و تعیین عوامل مؤثر بر بروز سرطان معده با رویکرد داده کاوی	Apriori، الگوریتم ID3	قدرت خوب مدل پیشنهادی برای پیش بینی سرطان معده
۴	۱۳۹۷	حدادها و همکاران	خوشه‌بندی، ارزیابی و انتخاب پروژه‌ها با روش کوله پشتی و روش تصمیم‌گیری چند معیاره	-K خوشه‌بندی، میانگین، ANN، نرم افزار متلب، Supper، Decision، و مدل‌سازی کوله پشتی با GAMS	استخراج ۴ معیار اصلی عوامل سازمانی، اجرایی، اقتصادی و بازار با ۱۳ زیر معیار - تأیید کارآمدی ابزار پیشنهادی
۵	۱۳۹۸	غریب و همکاران	ارایه الگوی ترکیبی داده کاوی با استفاده از قوانین انجمنی و خوشه‌بندی جهت شناسایی الگوهای غالب رفتار خرید مشتریان (مطالعه موردی: بانک انصار)	-K میانگین، Apriori	رفتار پویای مشتری در جا به جایی الگوها در بازه‌های زمانی بررسی شده و چهار گروه الگوی رفتاری غالب شناسایی شده است
۶	۱۳۹۹	اخوان و همکاران	تبیین بخش‌بندی مشتریان بازارهای صنعتی با نقشه‌های خودسازمانده (شرکت فولاد مبارکه)	بخش‌بندی مشتریان، نقشه‌های خود سازمانده SOM، مدل RFM	مشتریان بازار داخلی شرکت فولاد مبارکه اصفهان، بر اساس ۹۵ معیار مستخرج از ۴۸ شاخص (جمعیت‌شناختی، جغرافیایی عملیاتی، رفتاری و وضعیتی)، در پنج خوشه که بر مبنای متغیرهای شیوه‌های خرید (RFM)، به صورت مشتریان طلایی، مشتریان خاص، مشتریان وفادار، مشتریان رویگردان و مشتریان مشکوک نام‌گذاری شده‌اند، قرار گرفته اند.
۷	۱۴۰۰	تقوی فرد و همکاران	خوشه‌بندی مشتریان در حوزه بانکداری الکترونیک با بهره‌گیری از تراکنش‌های الکترونیکی و اطلاعات دموگرافیک (مورد مطالعه: بانک رفاه)	-K میانگین، RFMW، خوشه‌بندی، شاخص سیلوئت و دیویس بولدین	مدل جامع بخش‌بندی تراکنش‌های شتابی مشتریان بر مبنای میزان فعالیت و کارمزد در شبکه شتاب بر اساس مدل توسعه یافته تازگی، تکرار و میزان خرید موزون
۸	۱۴۰۰	بشردوست و همکاران	ارایه الگوی دسته‌بندی مشتریان با رویکرد داده کاوی ترکیبی - (محصولات بهداشتی و آرایشی)	خوشه‌بندی، WRFM، AHP، SPSS و کلمنتاین، درصد خلوص، امتیاز جینی	روش خوشه‌بندی K- میانگین روشی کارآمدتر با درصد خلوص (۰/۶۷۴) بوده و ۴ خوشه تعداد خوشه بهینه بوده است
۹	۱۴۰۱	مهاجر باجگیران	طراحی مدل تلفیقی برای خوشه‌بندی اقلام انبار و تخصیص آنها به محل انبارش در شرکت خانه‌سازی در مشهد با هدف کاهش هزینه حمل و نقل	-K میانگین، برنامه ریزی تخصیص تعمیم یافته، GAMS	خوشه‌بندی در تخصیص محل‌های انبارش - آرایه مدلی کلی برای مسایل انبارداری

(مورد مطالعه: محصولات شوینده و بهداشتی)

۱۰	۱۴۰۱	بشردوست و همکاران	تحلیل خوشه‌ای مشتریان بر مبنای مدل تازگی، تکرار و ارزش پولی موزون با رویکرد داده کاوی	تشخیص مغایرت، WRFM، خوشه‌بندی، میانگین، درصد خلوص، شاخص اطلاعات متقابل نرمال شده (NMI)	k- میانگین شش خوشه‌ای با خلوص (۰/۷۴۴) بهترین روش خوشه‌بندی بوده و ۴ دسته مشتری: خاص و کلیدی، طلایی بالقوه، نامطمئن از دست رفته، نامطمئن جدید بر پایه خوشه‌بندی برای مشتریان تعیین و برچسب گذاری شده است
۱۱	۱۴۰۲	تیموریان و همکاران	ارائه مدل مدیریت ارتباط با مشتری با تاکید بر مدیریت دانش مشتری در صنعت بانکداری	رویکرد داده بنیاد و مصاحبه نیمه ساختارمند	در الگوی پارادایمیک ارائه شده، ۳ دسته از کدهای محوری عوامل علی ۳ دسته عوامل زمینه‌ای ۳ دسته عوامل مداخله‌گر ۵ دسته و استراتژی‌ها ۳ دسته و "تحقق بازاریابی داخلی" به عنوان پیامدها، در مدل مدیریت ارتباط با مشتری با تاکید بر مدیریت دانش مشتری در صنعت بانکداری مشخص شدند
۱۲	۲۰۲۲	داراب و همکاران	یک مدل مبتنی بر خوشه‌بندی پیشرفته برای کشف قوانین غیر منتظره	طبقه‌بندی کننده‌های مبتنی بر یادگیری ماشینی	مدل پیشنهادی باعث کاهش عملیات پس پردازش، تنها مهمترین و فشرده‌ترین قوانین ایجاد می‌شود- اعتبار مدل با نمونه داده‌های واقعی پزشکی تأیید شد.
۱۳	۲۰۲۱	وو و همکاران	مدل ماشین یادگیری افراطی با الگوریتم خوشه‌بندی و کرم شب تاب به منظور تخمین تبخیر و تعرق داده‌های دمایی دریاچه پویانگ جنوب چین	ارایه مدل‌سازی جدید و سریع پردازش داده- های بزرگ برای بخش کشاورزی و منابع آن در مقیاس منطقه‌ای	
۱۴	۲۰۲۰	هوآنگ و یو ^۱	بررسی مدل RFM بهبود یافته بخش‌بندی مشتریان بر اساس الگوریتم K- میانگین (صنعت خرده فروشی)	RFM، خوشه‌بندی، K- میانگین	مدل RFMC پیشنهادی دارای عملکرد بهتری بوده است که توسط داده‌های شرکت تجارت الکترونیکی آموزشی تست شده است.
۱۴	۲۰۲۰	گونی، پکر و تورهان	رویکرد ترکیبی برای پروفایل مشتری در خدمات ویدیویی بر حسب تقاضا (یک اپراتور تلویزیون)	خوشه‌بندی، قوانین انجمنی، K- میانگین	تعیین چهار گروه مشتری: مشتریان با ارزش، مصرف بالا، مشترکین کم مصرف مشترکین کم مصرف وفادار، مشترکین غیر وفادار، داده کاوی ابزاری کارآمد برای تشکیل بخش‌های مشتریان، اجرای مؤثر مدیریت ارتباط با مشتری
۱۵	۲۰۲۰	ناتالیا ^۲	تقسیم‌بندی کاربران برنامه‌های موبایل: طبقه‌بندی رفتار وفاداری مشتری در بستر خرید آنلاین (سفر آنلاین)	خوشه‌بندی، دسته بندی	دسته‌بندی مشتریان در سه دسته کلی جویندگان درمان‌های ویژه، جویندگان ارزش پول، طرفداران برندها (وفادارترین) ، رایه توصیه‌های مدیریتی برای اجرای کمپین‌های وفاداری متناسب مشخصات جمعیتی و رفتاری هر بخش خوشه.

^۱ Huang , Zhang, Yue He

^۲ Natalia, Fortuna, Leonita

۱۶	۲۰۱۹	جوامات و موری	خوشه‌بندی مشتریان برای روشی جدید از پشتیبانی استراتژی بازاریابی در کسب و کار پیک(تایلند)	خوشه‌بندی، RFM، روش پیشنهادی مفهوم RFM و الگوریتم‌های خوشه‌بندی k-means را برای طبقه‌بندی و شناسایی مشتریان سودآور معرفی می‌کند.
۱۷	۲۰۱۸	ویشاواردهان و لکشیمپاد	ارایه الگوریتم اصلاح شده برای خوشه‌بندی- با استفاده از داده‌های واقعی تشخیص سرطان	بهبود داده کاوی و پایداری مدل نزدیکترین K همسایگی
۱۸	۲۰۱۸	موری و همکاران	پیش‌بینی تقاضای تک تک مشتریان از یک مجموعه داده بزرگ و پرسر و صدا	مدل پیشنهادی دارای عملکرد بهتری نسبت به مدل سنتی خوشه‌بندی بوده است.
۱۹	۲۰۱۸	توکلی و همکاران	تقسیم بندی مشتری و توسعه استراتژی بر اساس تحلیل رفتار کاربر، مدل RFM و تکنیک‌های داده کاوی: مطالعه موردی دیجی کالا	روش جدید دسته بندی باعث افزایش تعداد خریده‌ها و میانگین مبالغ سبدهای خرید شده است.

با نگاهی به پژوهش‌های پیشین انجام شده در زمینه کشف و یا شناخت الگوهای رفتاری خرید مشتریان آنچه مشخص است زمینه به کارگیری رویکرد داده کاوی در مسایل مدیریت ارتباط با مشتریان بسیار متنوع و گسترده است به خصوص در مسایل بهبود فرآیندهای تصمیم‌گیری، که بیشتر پژوهش‌های انجام شده به منظور پیش‌بینی رفتارهای انسانی و الگوهای رفتاری خرید مشتریان در زمینه بانکداری، صنعت مخابرات، تشخیص بیماری‌ها و بهبود شیوه‌های خدمت‌رسانی بوده است و کمتر در زمینه صنعت خرده فروشی محصولات خاصی چون محصولات بهداشتی و شوینده که جنبه زود مصرف و کوتاه مدت داشته، پرداخته‌اند. ماهیت این صنعت با رشد چشمگیر رقبا و مصرف و انباشت اطلاعات مواجه است و از آنجایی که در این صنعت همچنان از شیوه‌های سنتی تحقیقات بازار و بازاریابی که بسیار پر هزینه و زمانبر است استفاده می‌شود بنابراین انجام پژوهشی که بتواند به ارایه شیوه‌های کارآمدی برای مدیریت ارتباط با مشتری و شناسایی الگوهای رفتاری خرید مشتریان محصولات بهداشتی و شوینده که در چند سال اخیر مورد بی‌توجهی مدیران فروش و بازاریابی شرکت‌های فعال در این صنعت قرار گرفته است بپردازد بیش از پیش جلوه می‌نماید. می‌توان گفت نوآوری این پژوهش در استفاده از روش ترکیبی خوشه‌بندی و قوانین انجمنی با تحلیل رفتاری پویای مشتریان در هر خوشه الگوی خرید مشتریان است که در پژوهش‌های پیشین کمتر به جنبه پویای هر الگو در بازه‌های زمانی خرید مشتریان تأکید شده است و بیشتر با الگوی رفتاری متوسط یا الگوی غالب تحلیل رفتار مشتریان صورت گرفته است در حالی که در این پژوهش رفتار مشتریان که با تغییر شرایط محیطی و همچنین مقوله کرونا و پیامدهای آن با تغییرات بسیاری مواجه بوده است مورد بررسی و تجزیه و تحلیل واقع شده است.

روش پژوهش

از آنجایی که هدف این پژوهش توسعه کاربرد داده کاوی بر بستر بازاریابی به منظور دستیابی به الگوهای رفتاری مشتریان است تا بتواند از ماحصل آن الگوهای رفتاری (قوانین رفتاری) خرید مشتریان را شناسایی نماید بنابراین این پژوهش جزو پژوهش‌های کاربردی-توسعه‌ای است؛ این پژوهش از حیث شیوه انجام اکتشافی است چرا که در ابتدا

دانشی نسبت به رفتار مشتریان وجود نداشته و این دانش با انجام عملیات داده‌کاوی در حین پژوهش ایجاد می‌شود و در بستر آن فرضیه‌های پژوهش نیز در ذهن پژوهشگر ایجاد می‌شود.

جامعه آماری و روش نمونه‌گیری

جامعه آماری این پژوهش مشتریانی است که در یکی از قالب‌های گروه مشتری شامل: خرده‌فروشی، تعاونی مصرف و محلی از محصولات بهداشتی و شوینده شرکت استفاده کرده‌اند و محدود به استان تهران هستند؛ که از این بین ۶۵۵۳۴ مشتری به عنوان نمونه آماری با روش نمونه‌گیری هدفمند در دسترس از اطلاعات مشتریان شعبه جنوب شرق استان تهران که شامل مناطق ۸، ۱۲، ۱۳، ۱۴، ۱۵ (شهرداری تهران) و شهرهای دماوند، گرمسار، قرچک، ورامین، پاکدشت، ایوانکی می‌شوند انتخاب و گردآوری شده‌اند.

شیوه گردآوری داده‌ها

از روش مطالعه کتابخانه‌ای و مستندات علمی شامل مقاله‌ها و پایان‌نامه‌های تحصیلات تکمیلی مرتبط با موضوع پژوهش برای جمع‌آوری اطلاعات استفاده شده است، به این صورت که از طریق موتورهای جستجوگر مقاله‌ها، کتاب‌ها و مستندات، پایان‌نامه‌ها و نظر کارشناسان مرتبط با موضوع پژوهش بدست آمده و بعد از مطالعه و خلاصه‌برداری از مطالب مرتبط در پیشینه این پژوهش استفاده شده است و همچنین برای بخش داده‌کاوی این پژوهش با هدف کشف الگوهای رفتاری خرید مشتریان، از داده‌های جمع‌آوری و انباشته شده در پایگاه داده شرکت که به صورت فایل صفحه گسترده است که بخشی از آن مربوط به اطلاعات جمعیت‌شناختی مشتریان بوده و بخش دیگر مربوط به اطلاعات خرید و مالی مشتریان است نیز استفاده شده است.

شیوه تجزیه و تحلیل داده‌ها

این پژوهش به دنبال پیدا کردن الگوهای غالب رفتاری خرید است بنابراین از روش‌های مختلف و ترکیبی داده‌کاوی مانند خوشه‌بندی، قوانین انجمنی استفاده شده است که برای تجزیه و تحلیل و تعیین خوشه‌بندی‌ها از روش‌های مرسوم داده‌کاوی مانند k- میانگین، خوشه‌بندی دو مرحله‌ای استفاده شده و برای کشف الگوهای غالب رفتار خرید از الگوریتم قوانین انجمنی (اپریوری) استفاده شده است چراکه این الگوریتم دارای حجم محاسباتی کمتری نسبت به سایر روش‌هاست زیرا اصل اپریوری به این صورت است که اگر مجموعه‌ای از عناصر مکرر باشد، تمام زیرمجموعه‌های غیر تهی آن نیز مکرر خواهند بود (تان و همکاران^۱، ۲۰۰۶).

برای تعیین مؤلفه‌های مهم در تعیین خوشه‌بندی و پیش‌بینی از روش گزینش متغیرهای مهم به کمک آزمون آماری و بر اساس مقدار پی^۳ استفاده شده است. به طور کلی می‌توان گفت برای تجزیه و تحلیل اطلاعات مشتری که جنبه کیفی دارد از روش داده‌کاوی با استفاده از نرم‌افزار SPSS Modeler نسخه ۱۸ استفاده شده است. در ادامه به نحوه انجام فرآیند خوشه‌بندی و کشف قوانین انجمنی مورد استفاده این پژوهش اشاره می‌شود:

^۱ Excel

^۲ Tan, P.N., Steinbach, M. & Kumar, V.

^۳ P-Value

روش تحلیل پژوهش

نحوه عملکرد الگوریتم خوشه‌بندی K میانگین:

الگوریتم K- میانگین از روش‌های معمول و کارا در خوشه‌بندی است که K (تعداد خوشه‌ها) را به منزله ورودی می‌گیرد و مجموعه n شی را به K خوشه تخصیص می‌دهد. مراحل این الگوریتم به شرح ذیل است:

- به صورت تصادفی، K شی را به منزله مراکز خوشه‌های ابتدایی انتخاب می‌کند.
- هر شی را با توجه به بیشترین شباهت آن به مراکز خوشه‌ها، به خوشه‌ها تخصیص می‌دهد.
- مراکز خوشه‌ها را به روز می‌کند. به طوری که برای هر خوشه مقدار متوسط اشیای آن خوشه محاسبه می‌شود. به نقل از تان و همکاران تا هنگامی که هیچ تغییری در خوشه‌ها رخ ندهد به مرحله دوم بازمی‌گردد (آخوندزاده و همکاران، ۱۳۹۴).

برای اعتبارسنجی و کیفیت خوشه‌بندی روش‌های گوناگونی وجود دارد مانند: شاخص دان، شاخص دیویس بولدین، شاخص امتیاز جینی، بهره اطلاعاتی، شاخص اطلاعات متقابل نرمال شده و معیار سیلوئت (ضریب نیم‌رخ) که در این پژوهش ما از معیار سیلوئت که بر پایه محاسبات فاصله‌های درون خوشه‌ای و برون خوشه‌ای بنا نهاده شده استفاده کرده‌ایم. شاخص سیلوئت برای سنجش اعتبار خوشه‌ها، نسبت بین مجموع مربعات درون خوشه‌ای و مجموع مربعات بین خوشه‌ای را می‌سنجد. این شاخص همواره مقداری بین ۱ تا ۱- را کسب می‌کند. اگر شاخص سیلوئت نزدیک به یک باشد، نشان‌دهنده ساختار قوی خوشه‌بندی است، اما اگر این شاخص صفر باشد، خوشه‌بندی، ساختاری ضعیف دارد. اگر به شاخص مذکور، عدد منفی تعلق گیرد، به این معنی است که خوشه‌بندی هیچ ساختار منطقی ندارد (ال سراگ، ۲۰۱۲).

نحوه عملکرد قوانین انجمنی به منظور کشف قوانین وابستگی:

قوانین انجمنی به شکل گزاره‌های اگر آنگاه مطرح می‌شوند که به همراه معیارهای پشتیبان و اطمینان در چارچوب رابطه‌های ۱ تا ۴ بیان می‌شوند:

$$X \Rightarrow Y \text{ (support, confidence)} \quad (1)$$

$$\text{Support } (X \rightarrow Y) = P(X \cap Y) \quad (2)$$

$$\text{Confidence } (X \rightarrow Y) = P(X|Y) \quad (3)$$

$$\text{Lift } (X \rightarrow Y) = P(X|Y) / P(Y) \quad (4)$$

که معمولاً در تحلیل قواعد وابستگی از معیار لیفت^۱ یا اهمیت^۲ استفاده می‌شود که این معیار به ارزیابی میزان وابستگی مقدم و موخر بر اساس فرآیند ذیل می‌پردازد:

اگر مقدار اهمیت در این روش برابر یک شود، یعنی خرید مؤلفه C و مؤلفه D دو رویداد کاملاً از هم مستقل‌اند؛ اگر مقدار اهمیت کوچکتر از یک است یعنی اگر یک مشتری C را بخرد احتمالاً D را نمی‌خرد؛ اگر مقدار اهمیت بزرگتر از یک شود، یعنی اگر یک مشتری C را بخرد به احتمال زیاد D را نیز خواهد خرید؛ اگر مقدار اهمیت برابر صفر شود، اینگونه می‌توان تفسیر کرد که بین این دو محصول هیچ اجتماعی (ارتباطی) وجود ندارد (مک لنان و همکاران، ۱۳۹۲).

روایی و پایایی ابزار

در این پژوهش از نرم‌افزار داده‌کاوی SPSS Modeler استفاده شده است که به دفعات توسط مجامع علمی سراسر دنیا استفاده شده و از نظر اعتبار مورد پذیرش و استفاده بسیار قرار گرفته‌اند. برای تعیین دقت و اعتبارسنجی کلی داده‌ها،

^۱ Lift

^۲ Importance

(مورد مطالعه: محصولات شوینده و بهداشتی)

داده‌های پژوهش به دو بخش یادگیری و آزمون^۱ تقسیم شده است. همچنین روایی و پایایی این پژوهش مورد قبول خبرگان صنعت مورد بررسی و نخبگان علمی کشور قرار گرفته است. برای بیان میزان دقت، صحت و همچنین اعتبارسنجی خوشه‌ها برای عملیات داده‌کاوی از شاخص‌های میزان خطای خوشه‌بندی، معیار سیلوئت (ضریب نیم رخ) و نسبت اندازه بزرگترین خوشه به کوچکترین خوشه در هر روش و همچنین برای سنجش قوانین انجمنی و قواعد ایجاد شده از شاخص اطمینان، پشتیبان و اهمیت استفاده شده است.

قلمرو تحقیق

از نظر مکانی این پژوهش در یکی از شرکت‌های فعال در عرضه محصولات بهداشتی و شوینده از گروه صنعتی گلرنگ در بازه زمانی سال‌های ۱۴۰۰-۱۴۰۱ انجام شده است.

یافته‌های پژوهش

در این پژوهش از روش‌ها و الگوریتم‌های داده‌کاوی ذیل استفاده شده است: ۱- الگوریتم تشخیص مؤلفه‌های مهم^۲، ۲- الگوریتم تشخیص مغایرت، ۳- روش‌های خوشه‌بندی K-میانگین، خوشه‌بندی دو مرحله‌ای، ۴- الگوریتم قوانین انجمنی اperiوری. در ادامه به یافته‌ها و نتایجی که از هر یک از این روش‌ها حاصل شده اشاره می‌شود.

از بین ۲۴ (فیلد) اولیه داده‌های پژوهش (پایگاه داده‌های این پژوهش دارای ۶۵۵۳۴ رکورد مشتری بوده که هر رکورد مشتری دارای ۲۴ ستون (ویژگی) شامل جمعیت شناختی و اطلاعات خرید است)، با فرض نوع قلم کالایی محصولات خریداری شده به عنوان متغیر خروجی (هدف) به منظور کاهش ابعاد پژوهش برای استفاده در خوشه‌بندی و بقیه مراحل داده‌کاوی از روش گزینش ویژگی‌های مهم استفاده شده است که در این مرحله بر اساس محاسبه ضریب پیرسون و مقدار پی^۳ با اطمینان ۹۵ درصد با توجه به مقدار معناداری (Sig) فیلدهایی که در پیش‌بینی نوع محصول خریداری شده تأثیری نداشته و یا کم اهمیت تشخیص داده شوند یعنی ($0/05 \leq \text{Sig}$) از فیلدهای اولیه پژوهش حذف می‌شوند. که بر این اساس ۱۴ فیلد مهم ارزیابی شده‌اند و ۲ فیلد ID مشتریان و طول جغرافیایی مشتریان هم کم اهمیت بوده‌اند و فیلدهای گروه مشتریان، گروه مشتریان سطح یک، گروه مشتریان سطح دو، گروه مشتریان سطح سه، سال، موقعیت جغرافیایی، آخرین خرید از سال ۱۴۰۰، اولین خرید از سال ۱۴۰۰ بنا به این دلیل که تعداد فراوانی یک دسته بسیار زیاد بوده و یا اینکه ضریب تغییر واریانس متغیر مربوطه از حد آستانه تعیین شده کمتر بوده است حذف شده‌اند (پیش فرض انجام تحلیل حداقل ضریب تغییرات ۰/۱ و حداقل انحراف استاندارد ۰/۰ تعیین شده است).

روش تشخیص مغایرت

وقتی پس از تعیین ویژگی‌های مهم برای پیش‌بینی متغیر هدف (گروه بازار محصولات) از الگوریتم تشخیص مغایرت استفاده می‌شود؛ ۶ گروه غالب^۴ که هر گروه ویژگی‌های جمعیت شناختی مخصوص به خود دارد ایجاد شده که این شرایط

^۱ Training, Testing

^۲ Feature selection

^۳ P-value

^۴ Peer-Group

با توجه به مقدار شاخص مغایرت^۱ ۲/۸۱۷۱۴ (داده‌هایی که دارای انحراف از میانگین بیش از حدود ۲/۸ باشند مغایر تشخیص داده شده) بدست آمده است. به منظور انجام ادامه فرآیند داده کاوی پژوهش و بهبود کیفیت داده‌ها، جمعاً ۶۵۵ رکورد مغایری که در این روش تشخیص داده شده است را از کل نمونه اولیه حذف کرده تا در نهایت ۶۴۸۷۹ نمونه برای ادامه کار پژوهش که خوشه‌بندی است آماده شود. نتایج این روش به شرح جدول (۲) ذیل است.

جدول ۲. نتایج روش تشخیص مغایرت

ماه	منطقه	شهر	وضعیت مشتری *	روز خرید*	زمان سپری شده از آخرین سال ۱۳۹۸	خالص خرید (ریال)	تعداد داده‌های مغایر	فراوانی	گروه غالب ^۲	روش
تیر*	۱۵*	تهران*	غیرفعال *	۲۱	۳۰۱	۲۶۹۴۸۶۵	۸۳	۹۵۰۰	۱	تشخیص مغایرت
اسفند *	۸*	تهران	فعال	۲۰	۳۲	۷۴۵۶۱۵	۱۰۱	۱۲۰۸۲	۲	
بهمن *	۱۳*	تهران	فعال	۲۰	۲۶	۷۹۸۰۸۷	۱۲۲	۱۴۷۳۹	۳	
آذر*	۸	تهران	فعال	۲۰	۲۷	۸۴۵۲۵۲	۲۱۷	۸۵۹۰	۴	
تیر*	۸*	تهران*	فعال*	۱۸	۷۷	۰*	۴۴	۶۸۷۸	۵	
بهمن *	...	پاکدشت*	فعال		۳۱	۱۲۲۵۱۶۹	۸۸	۱۳۷۴۵	۶	
							۶۵۵	۶۵۵۳۴	جمع کل	

منبع: یافته‌های پژوهشگر

اعداد داخل جدول (۲) فوق گرد شده‌اند. در مورد وضعیت مشتری در این پژوهش چهار وضعیت کلی برای مشتریان وجود دارد: مشتریان فعال، مشتریان موقت، مشتریان غیرفعال و مشتریان لیست سیاه. روز خرید مشتریان همگی به‌طور متوسط بیان شده‌اند و در مورد شهر، منطقه، ماه و روز خرید وضعیت غالب مشتریان ذکر شده است. قسمت‌هایی که با علامت ستاره مشخص شده‌اند یعنی مقدار و یا ویژگی پیش‌بینی شده دارای مقداری خطا است ولی ستون‌هایی که دارای این علامت نیستند دارای مقدار پیش‌بینی قطعی بوده‌اند در مورد خالص خرید گروه غالب پنجم علت صفر شدن خالص خرید این دسته از مشتریان به این خاطر بوده است که این مشتریان محصولات را سفارش داده‌اند و حتی خرید هم داشته‌اند و ثبت سفارش برایشان صورت گرفته ولی بنا به دلایلی محصولات را عودت داده‌اند برای همین مقدار صفر در جدول (۳) ایجاد شده است.

نتایج روش خوشه‌بندی (دو مرحله‌ای)

از آنجایی که در بعضی از پژوهش‌ها مانند (خواجوند، تقوی فرد و نجفی ۱۳۹۱؛ باباییان و سرفرازی ۲۰۱۹)، این روش نسبت به دیگر روش‌های خوشه‌بندی عملکرد بهتری داشته است بنابراین در این پژوهش نیز سعی شده است این روش در کنار سایر روش‌ها مورد بررسی و تحلیل قرار گیرد که اگر نتایج و دقت بالاتری نسبت به دیگر روش‌های خوشه‌بندی داشته باشد بتوان از آن استفاده کرد. این روش نیز بر پایه تعدادی تنظیمات اولیه بنا شده که عبارتند از: ۱- حداقل تعداد

^۱ Anomaly index Cutoff^۲ Peer- Group

(مورد مطالعه: محصولات شوینده و بهداشتی)

خوشه ۲ و حداکثر ۱۵؛ ۲- برای تعیین فاصله مراکز خوشه‌ای از احتمال لگاریتمی^۱ استفاده شده؛ ۳- برای تعیین کیفیت و عملکرد خوشه‌بندی‌ها از معیار بیزین شوارتس^۲ (BIC) استفاده شده است؛ ۴- در این روش ۱۴ متغیر به عنوان ورودی انتخاب شده است. نتایج این روش خوشه‌بندی در جدول (۳) آمده است. در این پژوهش از بین ۱۴ متغیر اولیه برای ورودی تحلیل، ویژگی‌های پوشش مشتری با ضریب اهمیت بسیار ناچیز در حدود صفر، مقدار برگشت محصول برحسب ریال با ضریب اهمیت ۰/۰۲، روز خرید با ضریب اهمیت ۰/۰۵، تعداد خط فاکتور با ضریب اهمیت ۰/۲۸ و ماه خرید با ضریب اهمیت ۰/۵ کمترین تأثیر و اهمیت را در پیش‌بینی خوشه‌بندی با این شیوه داشته‌اند و مابقی متغیرها دارای ضریب اهمیت یک بوده‌اند.

جدول ۳. نتایج روش خوشه‌بندی دو مرحله‌ای

نام روش	تعداد خوشه‌ها	کیفیت خوشه‌بندی	مقدار میانگین سیلوئت	نسبت فراوانی (بزرگترین خوشه به کوچکترین خوشه)	اندازه بزرگترین خوشه	بزرگترین خوشه (درصد)	اندازه کوچکترین خوشه	کوچکترین خوشه (درصد)
خوشه‌بندی دو مرحله‌ای	۳	مناسب ^۳	۰/۲	۴/۳۹	۳۰۰۱۴	۶۵/۸	۶۸۴۲	۱۵

منبع: یافته‌های پژوهشگر

نتایج خوشه‌بندی K- میانگین

در بیشتر پژوهش‌های انجام شده مانند پژوهش‌های: (بشردوست و همکاران ۱۴۰۱؛ بشیری و افشار ۲۰۱۸؛ شیخ، غلامیان و قنبرپور ۲۰۱۹، آریونی و میراندا، ۲۰۱۸؛ جوماتام و موری، ۲۰۱۹) برای خوشه‌بندی از این روش به علت کارایی بالای آن استفاده شده است بنابراین به علت فراوانی استفاده و کارایی بالای آن در این پژوهش از این روش نیز استفاده می‌شود تا چنانچه نتایج آن از دقت بالاتری نسبت به دیگر روش‌های مورد استفاده در این پژوهش برخوردار باشد از آن به عنوان روش بهتر و مناسبتر برای ادامه تحلیل خوشه‌ای و داده کاوی پژوهش استفاده کرد. در این قسمت به خوشه بندی‌هایی که با روش K- میانگین با فرض ۴، ۳، ۵، ۶، ۷ خوشه ایجاد شده است اشاره می‌شود (در روش خوشه‌بندی K- میانگین لازم است در ابتدا پژوهشگر یا کاربر مقدار اولیه‌ای را برای تعداد دسته‌های خوشه‌بندی تعیین کند که چنانچه هر کدام از این تعداد خوشه‌ها نتایج بهتری داشت انتخاب شود). وقتی که ویژگی گروه بازار محصولات (۲۶ قلم کالایی) به عنوان متغیر خروجی انتخاب می‌شود نتایج این روش خوشه‌بندی به شرح جدول (۴) خواهد بود.

جدول ۴. نتایج خوشه‌بندی K-Means

نام روش	تعداد خوشه‌ها	کیفیت خوشه‌بندی	مقدار میانگین سیلوئت	تعداد تکرار مراحل خوشه‌بندی	میزان خطا ^۵	نسبت فراوانی (بزرگترین خوشه به کوچکترین خوشه)	اندازه بزرگترین خوشه	بزرگترین خوشه (درصد)	اندازه کوچکترین خوشه	کوچکترین خوشه (درصد)
	۳	ضعیف	۰/۱	۷	۰/۰	۲/۵۶	۲۲۹۸۴	۵۰/۴	۸۹۸۳	۱۹/۷

^۱ Log- Likelihood

^۲ Schwarz's Bayesian Criterion(BIC)

^۳ Fair

^۴ Iteration

^۵ Error

۱۴/۵	۶۶۲۶	۴۱/۵	۱۸۹۳۴	۲/۸۶	۰/۰	۷	۰/۲	ضعیف	۴	میانگین -k
۱۱/۵	۵۲۳۰	۳۰/۳	۱۳۸۳۳	۲/۶۴	۰/۰	۱۵	۰/۲	مناسب	۵	
۴/۴	۲۰۰۸	۳۱/۴	۱۴۳۲۸	۷/۱۴	۰/۰	۱۱	۰/۲	ضعیف	۶	
۴/۴	۲۰۰۸	۳۱/۱	۱۴۲۰۵	۷/۰۷	۰/۰	۱۰	۰/۱	ضعیف	۷	

منبع: یافته‌های پژوهشگر

باتوجه به جدول (۴) از بین تعداد خوشه‌های مختلفی که به روش خوشه‌بندی K- میانگین بررسی شده است تعداد ۵ خوشه با توجه به مقدار معیار سیلوئت و کیفیت خوشه‌بندی انتخاب می‌شود. در ادامه به بررسی نحوه تقسیم مشتریان به تفکیک هر خوشه در دو روش خوشه‌بندی K- میانگین و دو مرحله‌ای به تفکیک داده‌های یادگیری و آزمون پرداخته می‌شود که این اطلاعات در جدول (۵) به نمایش درآمده است.

جدول ۵. مقایسه زوجی شیوه تقسیم مشتریان در دو روش مختلف خوشه‌بندی

جمع کل	خوشه‌بندی دو مرحله‌ای						نام روش
	خوشه سوم		خوشه دوم		خوشه اول		
	داده‌های آزمون	داده‌های یادگیری	داده‌های آزمون	داده‌های یادگیری	داده‌های آزمون	داده‌های یادگیری	
۱۱۱۷۴	۰	۰	۳۲۷۸	۷۷۸۶	۲۱	۸۹	خوشه ۱
	۰		۱۱۰۶۴		۱۱۰		جمع
۱۹۵۳۹	۵۵۲۵	۱۳۴۳۶	۰	۰	۱۸۱	۳۹۷	خوشه ۲
	۱۸۹۶۱		۰		۵۷۸		جمع
۱۷۶۰۸	۵۲۵۳	۱۲۰۴۱	۰	۰	۹۹	۲۱۵	خوشه ۳
	۱۷۲۹۴		۰		۳۱۴		جمع
۹۰۹۱	۰	۰	۲۱۸	۵۴۹	۲۴۴۸	۵۸۷۶	خوشه ۴
	۰		۷۶۷		۸۳۲۴		جمع
۷۴۶۷	۱۹۱۷	۴۵۳۷	۱۹۳	۴۲۸	۱۲۷	۲۶۵	خوشه ۵
	۶۴۵۴		۶۲۱		۳۹۲		جمع
۶۴۸۷۹	۱۲۶۹۵	۳۰۰۱۴	۳۶۸۹	۸۷۶۳	۲۸۷۶	۶۸۴۲	جمع کل
	۴۲۷۰۹		۱۲۴۵۲		۹۷۱۸		

منبع: یافته‌های پژوهشگر

در ادامه به مقایسه دو روش خوشه‌بندی مورد بررسی در این پژوهش پرداخته می‌شود تا در نهایت از این مقایسه روش خوشه‌بندی بهتر انتخاب شود و بر این اساس مشتریان به بخش‌های مختلف تقسیم شده و برای هر بخش مدیران فروش و بازاریابی برنامه‌های ویژه‌ای تدارک ببینند.

جدول ۶. مقایسه دو روش مختلف خوشه‌بندی

روش خوشه‌بندی	تعداد خوشه‌بندی	کیفیت خوشه‌بندی	مقدار سیلوئت	نسبت فراوانی (بزرگترین خوشه به کوچکترین خوشه)
دو مرحله‌ای	۳	مناسب	۰/۲	۴/۳۹
K- میانگین	۵	مناسب	۰/۲	۲/۶۴

منبع: یافته‌های پژوهشگر

(مورد مطالعه: محصولات شوینده و بهداشتی)

در جدول (۶) مقایسه دو روش مختلف خوشه‌بندی که در این پژوهش مورد بررسی قرار گرفته آمده است. با توجه به جدول بالا و بر اساس کیفیت خوشه‌بندی از روش‌های مختلف، خوشه‌بندی K- میانگین با تعداد ۵ خوشه دارای کیفیت بهتری بوده است چرا که دارای وضعیت پراکندگی خوشه‌ای مناسبتری با توجه به شاخص نسبت فراوانی است و در دو معیار کیفیت خوشه‌بندی و مقدار اعتبارسنجی شاخص سیلوئت دارای وضعیتی مشابه خوشه‌بندی دو مرحله‌ای است پس این روش به عنوان روش نهایی برای تحلیل خوشه‌بندی و کشف الگوهای رفتاری خرید مشتریان مورد استفاده قرار می‌گیرد.

برای تعیین الگو و رتبه‌بندی خوشه‌های روش خوشه‌بندی K- میانگین از شیوه مورد استفاده در پژوهش‌های (آریونی و میراندا، ۲۰۱۸؛ چانگ و تسای، ۲۰۰۴؛ بابایان و سرفرازی، ۲۰۱۹) استفاده می‌شود بدین صورت که ابتدا با توجه به مقایسه دو به دو هر یک از مؤلفه‌های میانگین زمان سپری شده از آخرین خرید مشتری با مقدار میانگین کل و همچنین با مقایسه میانگین مقدار خرید قطعی با مقدار میانگین کل خرید مشتریان و میانگین تعداد خط فاکتور فروش با میانگین کل این مقدار با توجه به اینکه ۵ خوشه انتخاب شده است پس رتبه‌بندی خوشه‌ها بین ۱ تا ۵ خواهد بود؛ به بهترین وضعیت رتبه یک داده می‌شود و به بدترین حالت رتبه ۵ مثلاً برای ویژگی میانگین زمان سپری شده از آخرین خرید که جنبه معکوس دارد و هرچه اعداد این قسمت کوچکتر باشد یعنی مشتریان به تازگی از شرکت خرید داشته‌اند و رتبه بهتری به آنها تعلق خواهد گرفت و برای دو ویژگی دیگر هر چقدر مقادیر ایجاد شده بیشتر باشد رتبه بهتری تعلق می‌گیرد. رتبه (۳، ۱، ۱) که مربوط به خوشه اول است یعنی این خوشه در شاخص تازگی خرید در بین ۵ خوشه دارای رتبه ۳ بوده ولی در دو شاخص تعداد دفعات خرید و ارزش پولی خرید دارای بهترین رتبه بوده است و به همین ترتیب برای دیگر خوشه‌ها، رتبه‌ها به شرح جدول (۷) تعیین می‌شود. برای تعیین نوع الگوهای حاکم درون خوشه‌ها از علامت (↑) برای نشان دادن ویژگی‌هایی که بالاتر از میانگین هستند استفاده شده و از علامت (↓) برای نشان دادن ویژگی‌های که دارای مقادیر پایینتر از میانگین هستند استفاده شده است و با توجه به دو وضعیتی بودن هر ویژگی (بالا تر یا پایین تر از میانگین) و داشتن سه ویژگی برای مقایسه ۸ حالت کلی برای این نوع نحوه نام‌گذاری الگوها قابل تصور است که با توجه به شرایط این پژوهش و داشتن ۵ خوشه برای مقایسه ۴ الگوی کلی: الگوی اول (↓، ↑، ↓)، با فراوانی ۱۸۶۴۱، الگوی دوم (↑، ↓، ↓)، با فراوانی ۱۹۵۳۹، الگوی سوم (↓، ↓، ↓)، با فراوانی ۱۷۶۰۸ و الگوی چهارم (↓، ↓، ↑)، با فراوانی ۹۰۹۱ بدست آمده است که این نتایج در جدول (۷) به نمایش درآمده است.

جدول ۷. تعیین رتبه و الگوی درون خوشه‌ای

روش	خوشه	فراوانی	درصد فراوانی	میانگین زمان سپری شده از آخرین خرید	میانگین مقدار خرید قطعی (ریال)	میانگین تعداد خط فاکتور فروش (ریال)	میانگین روز خرید	وضعیت رتبه‌بندی	نوع الگو
میانگین	۱	۱۱۱۷۴	۱۷/۲۲	۳۱/۴۹۷	۱۲۱۸۱۲۷/۲۴	۲/۴۵۶	۲۰/۲۵۸	(۳، ۱، ۱)	(↓، ↓، ↓)
	۲	۱۹۵۳۹	۳۰/۱۲	۲۹/۲۶	۱۱۳۶۱۸۶/۲۵۸	۲/۳۲۲	۱۹/۷۳۲	(۳، ۴، ۲)	(↓، ↓، ↓)
	۳	۱۷۶۰۸	۲۷/۱۴	۲۶/۹۷	۷۶۸۷۲۸/۹۳۹	۱/۹۷۵	۱۹/۷۱	(۵، ۵، ۱)	(↓، ↓، ↓)
	۴	۹۰۹۱	۱۴/۰۱	۳۲۲/۵۹	۱۱۷۸۴۵۱/۹۳	۲/۲۵	۲۰/۴۳۹	(۴، ۲، ۵)	(↑، ↓، ↓)
	۵	۷۴۶۷	۱۱/۵۱	۵۷/۹۱۷	۱۱۵۴۳۶۶/۴۸۳	۲/۳۹۳	۱۹/۸۸	(۲، ۳، ۴)	(↓، ↓، ↓)
جمع کل / میانگین		۶۴۸۷۹	۱۰۰	۷۳/۶۴۳	۱۱۵۰۵۷۹/۰۳۲	۲/۲۹۵	۱۹/۹۳۴		
کمترین مقدار				۴	۰	۰	۱		
بیشترین مقدار				۵۷۵	۴۱۱۰۱۵۱۶۸	۲۹	۳۱		

انحراف استاندارد ^۱	۱۱۷/۱۶۳	۳۴۸۳۴۵/۸۱۱	۲/۱۲۰	۸/۲۴۲
-------------------------------	---------	------------	-------	-------

منبع: یافته‌های پژوهشگر



نمودار ۸. نمودار فراوانی روند رفتاری خرید مشتریان به تفکیک ماه و خوشه

منبع: یافته‌های پژوهشگر

آنچه از بررسی نمودار (۸) روند رفتاری خرید مشتریان در طی ماه‌های مختلف درون هر خوشه قابل مشاهده است این است که مشتریان درون خوشه دوم به طور میانگین از یک روند صعودی خرید با شیب تندی تا پایان سال برخوردار بوده‌اند. مشتریان درون خوشه چهارم در ابتدای سال با روند صعودی خرید با شیب تندی مواجه بوده‌اند ولی از تیر ماه به یکباره با روند نزولی خرید مواجه شده‌اند. مشتریان خوشه سوم با روند صعودی خرید مواجه بوده‌اند که البته شیب این روند نسبتاً تند بوده است و دو خوشه دیگر تقریباً از روند نسبتاً پایدارتری برخوردار بوده‌اند (بیشترین خرید مشتریان خوشه اول در تیر ماه با فراوانی ۱۱۲۹ مورد، بیشترین خرید مشتریان خوشه دوم در بهمن ماه با فراوانی ۲۱۴۱ مورد، بیشترین خرید مشتریان خوشه سوم در آذر ماه با فراوانی ۱۸۴۱ مورد و در خوشه چهارم بیشترین خرید مشتریان در تیرماه با فراوانی ۱۴۰۹ بوده و در نهایت بیشترین خرید مشتریان خوشه پنجم در ماه بهمن با فراوانی ۸۰۴ مورد صورت گرفته است). در جدول (۹) فراوانی مشاهده الگوهای رفتاری خرید مشتریان به تفکیک هر الگو در هر خوشه آمده است.

جدول ۹. الگوهای رفتاری مشتریان به تفکیک هر خوشه

الگوی رفتاری غالب میانگین	جمع کل	الگوهای غالب هر خوشه								شماره خوشه
		(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	(۱،۱،۱)	
(۱،۱،۱)	۱۱۱۷۴	۵۱۹۲	۱۲۹	۱۶۶۸	۳۷۶	۱۷۶	۱۱۵۶	۲۳۷۶	۱۰۱	۱
(۱،۱،۱)	۱۹۵۳۹	۱۰۳۱۴	۳۶۴	۲۷۵۳	۸۹۶	۲۸۸	۱۴۱۴	۳۳۱۸	۱۹۲	۲
(۱،۱،۱)	۱۷۶۰۸	۱۱۹۰۵	۱۲۲	۱۸۴۶	۹۱۶	۱۲۰	۸۴۳	۱۷۸۲	۷۴	۳
(۱،۱،۱)	۹۰۹۱	۵۹	۱۲۳۹	۷۶	۵۲۳۷	۱۶۳۱	۳۰	۱۳	۸۰۶	۴
(۱،۱،۱)	۷۴۶۷	۲۹۸۴	۲۲۳	۷۸۲	۱۰۱۶	۴۵۸	۳۹۴	۱۴۱۵	۱۹۵	۵
جمع کل	۶۴۸۷۹	۳۰۴۵۴	۲۰۷۷	۷۱۲۵	۸۴۴۱	۲۶۷۳	۳۸۳۷	۸۹۰۴	۱۳۶۸	

منبع: یافته‌های پژوهشگر

^۱ Standard Deviation (Std. Dev.)

با توجه به نتایج الگوهای رفتاری مشاهده شده در هر خوشه مطابق جدول (۹) هر ۸ گروه الگو در هر خوشه مشاهده شده است (چون سه ویژگی زمان سپری شده از آخرین خرید، مقدار خالص خرید و تعداد خط فاکتور خرید مؤلفه‌های مورد بررسی هستند که هر یک می‌توانند بر اساس اینکه دارای مقادیر کمتر یا بیشتر از میانگین کل بوده باشد دو حالت را اختیار کنند بنابراین ۸ حالت قابل تصور است).

یافته‌های قوانین انجمنی (تحلیل وابستگی)^۱

در این قسمت به الگوهایی که در رفتار خرید مشتریان جامعه آماری این پژوهش حاکم است اشاره می‌شود که برای کشف این الگوها از بین الگوریتم‌های کارما^۲، تالی^۳ و آپریوری^۴ در این پژوهش از روش آپریوری استفاده شده است. برای این منظور تنظیمات اولیه‌ای می‌بایست بر روی نرم‌افزار SPSS Modeler صورت گیرد. این الگوریتم شامل قوانین اگر و آنگاه (مقدم^۵ و تالی^۶) می‌باشد و به طور معمول از دو شاخص پشتیبان^۷ و اطمینان^۸ برای سنجش قوانین ایجاد شده استفاده می‌کند (شاخص پشتیبان بیان می‌کند چه تعداد قوانین شامل هم بخش مقدم است و هم بخش تالی (فراوانی صحت وقوع یک گزاره)، شاخص اطمینان بیان می‌کند چه تعداد قوانین فقط با بخش مقدم ایجاد شده است).

در این پژوهش از تنظیمات اولیه خود نرم‌افزار مبنی حداقل ۱۰ درصد پشتیبان و حداقل ۸۰ درصد اطمینان و همچنین وجود حداکثر ۵ مؤلفه (مقدم) در بخش اگر قوانین، با فرض بهینه‌سازی سرعت خروجی نرم‌افزار استفاده شده است که بعد از ثبت این اطلاعات از نرم‌افزار خواسته می‌شود بر این اساس قوانینی را کشف و استخراج^۹ نماید. بعد از اجرا نرم‌افزار با گزارش خطایی روبرو شده است مبنی بر اینکه حداستانه^{۱۰} تعیین شده بسیار بالا است پس پژوهشگران مجدد حداقل درصد پشتیبان را به صفر درصد و حداقل درصد اطمینان را نیز به ۴۵ کاهش داده‌اند تا قوانین کلی‌تری ایجاد شوند بر این اساس ۷۹ قانون تشخیص داده شده است که از آنجایی که بسیاری از این قوانین دارای درصد اطمینان و حمایت بسیار ناچیز بوده‌اند و همچنین چون به دنبال قوانین قویتری هستیم بنابراین با بررسی مجدد قوانین بدست آمده از بین ۷۹ قانون ایجاد شده، قوانین را در نظر گرفته‌ایم که ضمن اینکه دارای فراوانی (نمونه^{۱۱}) بیشتری باشند دارای مقادیر اطمینان بالاتری نیز هستند. از آنجایی که در این پژوهش مقدار اهمیت (لیفت) تمامی قوانین بیشتر از یک است پس بخش مقدم و تالی به هم وابسته هستند به نوعی می‌توان گفت اگر اطلاعات جمعیت‌شناختی مقدم اتفاق بیفتد آنگاه محصول مشخص شده خریداری می‌شود. در نهایت بر اساس شاخص اطمینان ۶۰ درصد و حداقل پشتیبانی صفر درصد تعداد ۴ قانون برای شامپو بدن، ۶ قانون برای مایع نرم‌کننده حوله و لباس، ۱۳ قانون برای مایع دستشویی، ۲ قانون برای دستمال کاغذی، ۴ قانون برای

^۱ Association Rules

^۲ Carma

^۳ Sequenc

^۴ Apriori

^۵ Antecedent

^۶ Consequent

^۷ Support

^۸ Confidence

^۹ Extract Rules

^{۱۰} Threshold

^{۱۱} Instances

شامپو بچه، ۴ قانون پودر لباسشویی ماشینی، ۱ قانون برای پودر لباسشویی دستی، ۱ قانون برای مایع شیشه پاک کن جمعاً ۳۵ قانون کلی از بین ۷۹ قانون اولیه بدست آمد که برای تحلیل و بررسی الگوهای خرید مشتریان مورد تحلیل قرار گرفته است که در اینجا تنها به بخشی از این قوانین اشاره می‌شود:

۳ قانون از ۴ قانون کلی برای مشتریانی که شامپو بدن سفارش می‌دهند:

۱. اگر مشتری در وضعیت موقت باشد و تیر ماه خرید کرده باشد آنگاه احتمال دارد شامپو بدن سفارش دهد؛
۲. اگر مشتری در وضعیت موقت باشد و تیر ماه خرید کرده باشد و از منطقه ۱۵ باشد آنگاه احتمال دارد شامپو بدن سفارش دهد.
۳. اگر مشتری در وضعیت موقت باشد و تیر ماه خرید کرده باشد و از تهران باشد آنگاه احتمال دارد شامپو بدن سفارش دهد.

۴ قانون از ۶ قانون برای مشتریانی که مایع نرم کننده حوله و لباس سفارش می‌دهند:

۱. اگر مشتری در وضعیت موقت باشد و مهر ماه خرید کرده باشد آنگاه احتمال دارد مایع نرم کننده حوله و لباس سفارش دهد؛
۲. اگر مشتری در وضعیت موقت باشد و مهر ماه خرید کرده باشد و از منطقه ۱۵ باشد آنگاه احتمال دارد مایع نرم کننده حوله و لباس سفارش دهد؛
۳. اگر مشتری در وضعیت موقت باشد و مهر ماه خرید کرده باشد و از شهر تهران باشد آنگاه احتمال دارد مایع نرم کننده حوله و لباس سفارش دهد؛
۴. اگر مشتری در وضعیت لیست سیاه باشد و فروردین ماه خرید کرده باشد و از منطقه ۱۵ باشد آنگاه احتمال دارد مایع نرم کننده حوله و لباس سفارش دهد؛

۶ قانون از ۱۳ قانون برای مشتریانی که مایع دستشویی سفارش می‌دهند:

۱. اگر مشتری از دماوند باشد و در تیر ماه خرید کند آنگاه احتمال دارد مایع دستشویی سفارش دهد؛
۲. اگر مشتری از دماوند باشد و در تیر ماه خرید کرده باشد و همچنین در وضعیت فعال باشد آنگاه احتمال دارد مایع دستشویی سفارش دهد؛
۳. اگر مشتری از پیشوا باشد و در دی ماه خرید کرده باشد و در وضعیت غیرفعال باشد آنگاه احتمال دارد مایع دستشویی سفارش دهد؛
۴. اگر مشتری در لیست سیاه باشد و از ورامین بوده و در آذرماه خرید کرده باشد آنگاه احتمال دارد مایع دستشویی سفارش دهد؛
۵. اگر مشتری در لیست سیاه باشد در دی ماه خرید کرده باشد و از منطقه ۱۲ باشد آنگاه احتمال دارد مایع دستشویی سفارش دهد؛
۶. اگر مشتری در لیست سیاه باشد و مرداد ماه خرید کند و از منطقه ۱۳ آنگاه احتمال دارد مایع دستشویی سفارش دهد؛

۱ قانون از ۲ قانون برای مشتریانی که دستمال کاغذی سفارش می‌دهند:

۱. اگر مشتری از ورامین بوده و آذر ماه خرید کرده باشد و در وضعیت غیرفعال قرار گرفته باشد آنگاه احتمال دارد دستمال کاغذی سفارش دهد؛

۲ قانون از ۴ قانون برای مشتریانی که شامپو بچه سفارش می‌دهند:

۱. اگر مشتری از پیشوا باشد و بهمن ماه خرید کرده و در وضعیت غیرفعال باشد آنگاه احتمال دارد شامپو بچه سفارش دهد.
 ۲. اگر مشتری در لیست سیاه باشد و از شهر پاکدشت و در اسفند ماه خرید کرده باشد آنگاه احتمال دارد شامپو بچه سفارش دهد.
- ۲ قانون از ۴ قانون کلی برای مشتریانی که پودر لباسشویی ماشینی سفارش می دهند:
۱. اگر مشتری در لیست سیاه باشد و اردیبهشت ماه خرید کرده باشد و از شهر پاکدشت آنگاه احتمال دارد پودر لباسشویی ماشینی سفارش دهد؛
 ۲. اگر مشتری در لیست سیاه باشد و اسفند ماه خرید کرده باشد و از منطقه ۱۵ باشد آنگاه احتمال دارد پودر لباسشویی ماشینی سفارش دهد؛
- تنها قانون کلی برای مشتریانی که پودر لباسشویی دستی سفارش می دهند:
۱. اگر مشتری در مهر ماه باشد و از پاکدشت و در وضعیت غیرفعال آنگاه احتمال دارد پودر لباسشویی دستی سفارش دهد؛
- تنها قانون کلی برای مشتریانی که مایع شیشه پاک کن سفارش می دهند:
۱. اگر مشتری در بهمن ماه خرید کرده باشد و در وضعیت غیرفعال باشد و از منطقه (۲۳) (شهرهای حومه تهران و همچنین از شهر تهران باشد آنگاه احتمال دارد مایع شیشه پاک کن سفارش دهد؛
- در ادامه به منظور درک بهتر وجود ارتباط بین مؤلفه‌های جمعیت‌شناختی این پژوهش (منطقه، شهر، ماه، وضعیت مشتریان) و شماره خوشه هر مشتری از نمودار وب^۱ که بر اساس فراوانی ارتباطات بین ویژگی‌ها ایجاد می‌شود برای درک بهتر ارتباطات بین مؤلفه‌های جمعیت‌شناختی و متغیر هدف (گروه بازار محصولات) در دو روش خوشه‌بندی k- میانگین و دو مرحله‌ای استفاده شده است. نتایج این ارتباطات در قالب شکل (۱۰) به نمایش درآمده است. در این شکل هر چقدر تعداد ارتباط بین مؤلفه‌ها بیشتر باشد با رنگ آبی تیره و پررنگتر نمایش داده می‌شوند و اگر تعداد فراوانی ارتباطها کمتر باشد با رنگ روشنتر نمایش داده می‌شوند مثلاً در این شکل مشخص است بیشترین مشتریان در خوشه دوم k- میانگین از شهر تهران بوده و از منطقه ۸ هستند یا مثلاً بیشترین مشتریان در داده‌های آزمون این پژوهش فعال بوده و مربوط به خوشه دوم k- میانگین بوده و یا در خوشه سوم خوشه‌بندی دو مرحله‌ای قرار گرفته‌اند.

رفتار میانگین الگوها مقایسه شوند تا برنامه‌های بازاریابی و برنامه‌ریزی‌های استراتژیک مدیریت ارتباط با مشتری بر این اساس طراحی و تدوین شوند.

نتایج این پژوهش با بررسی تفاوت‌های الگوهای رفتاری خرید مشتریان به تفکیک هر خوشه، ماهیت پویای داده کاوی را نیز تأیید کرده است که این موضوع در راستای نتایج پژوهش (غریب و همکاران، ۱۳۹۸) هست که به دنبال ارائه الگوی تحلیلی پویای رفتار مشتریان بوده‌اند.

از بین ۲۶ قلم کالایی محصولات با استفاده از الگوریتم اپریوری با مقدار شاخص اطمینان ۶۰ درصد تنها برای ۸ قلم کالایی محصول شامل: شامپو بدن با ۴ قانون، مایع نرم‌کننده حوله لباس با ۶ قانون، دستمال کاغذی با ۲ قانون، شامپو بچه با ۴ قانون، مایع دستشویی ۱۳ قانون، پودر لباسشویی ماشینی با ۴ قانون، پودر لباسشویی دستی با ۱ قانون، مایع شیشه پاک‌کن با ۱ قانون، کلاً ۳۵ قانون و الگو حاصل شده است که نشان می‌دهد مشتریان این ۸ قلم کالایی را بسیار سفارش داده‌اند پس مدیران فروش و مدیران انبار شرکت مورد بررسی می‌بایست این قلم کالاها را با توجه به شرایط منطقه‌ای و شهری و ماه خرید مورد توجه داشته باشند تا نسبت به درخواست تولید به موقع این دسته اقلام در بازه‌های زمانی تعیین شده اقدام نمایند. همچنین با توجه به نتایج گروه‌بندی غالب جدول ۲، میانگین خریدها در بازه زمانی ۱۸ تا ۲۱ هر ماه صورت گرفته است پس شرکت باید برای این زمان‌ها خود را آماده‌تر کند به نوعی برای این روزها تلاش بیشتر و تمرکز خاص و ویژه‌ای داشته باشد (عملکرد نیروی فروش جهت ثبت سفارش دقیق مورد پایش قرارگیرد و ارسال به موقع به دست مشتریان با هماهنگی با واحد انبار صورت گیرد).

با مقایسه الگو رفتاری میانگین هر خوشه با الگوهای بدست آمده برای هر خوشه می‌توان گفت تنها در خوشه سوم الگوی غالب (۱،۱،۱) با الگوی میانگین تقریباً هم‌خوانی دارد که می‌تواند نشان از ثبات رفتاری خریداران این خوشه داشته باشد که در طول سال تقریباً این روند الگو را حفظ کرده‌اند و مشتریان این خوشه دارای رفتار اینچنین بوده‌اند: تازگی خرید داشته ولی دفعات خرید و میانگین ریالی خرید آن‌ها کمتر از میانگین کل بوده است درحالی‌که در دیگر خوشه‌ها الگوهای رفتاری غالب و الگوی رفتاری میانگین کاملاً متفاوت بوده که این خود نشان از عدم ثبات رفتاری خرید مشتریان درون سایر خوشه‌ها دارد که این موضوع باید مورد توجه تصمیم‌گیرندگان و مدیران ارشد سازمان قرارگیرد چراکه الگوهای رفتاری متفاوت و متغیر بحث راهبردهای بازاریابی و تحقیقاتی و تبلیغاتی متفاوتی می‌طلبد و البته هزینه‌های شرکت نیز در این زمینه افزایش خواهد یافت ولی به راحتی می‌توان برای آینده خرید مشتریان خوشه سوم برنامه‌ریزی مشخص‌تری داشت و از تبلیغات و بازاریابی یک دست و مشابه‌ای استفاده کرد که با این تفاسیر این خوشه مشتریان از بعد تبلیغاتی هزینه کمتری برای شرکت خواهند داشت ولی از بعد فروش و سود برای سازمان هزینه بر خواهند بود چراکه که الگوی غالب آنها خرید کمتر از میانگین است.

این پژوهش مسیری برای تصمیم‌گیری مدیران فعال در صنعت خرده‌فروشی خصوصاً محصولات شوینده و بهداشتی در زمینه مدیریت ارتباط با مشتری بر پایه پویایی‌های رفتاری خرید مشتریان در مسیر کسب مزیت رقابتی ترسیم می‌نماید که مدیران و شرکت‌های فعال در زمینه محصولات خرده‌فروشی به راحتی بتوانند با بهره‌گیری از پایگاه داده‌های مشتریان خود و با استفاده از ابزار داده کاوی در بازه زمانی بسیار کوتاهی رفتارها و الگوهای رفتاری خرید مشتریان کلیدی و سایر مشتریان خود را شناسایی کرده و بر این اساس راهبردهای بازاریابی و فروش شرکت خود را اولویت‌بندی و برنامه‌ریزی نمایند.

محدودیت‌ها و پیشنهادات

از آنجایی که این پژوهش مربوطه به یکی از شعب فعال شرکت است و نه کل شعب بنابراین با توجه به ماهیت داده کاوی که جنبه پویا داشته و همچنین با توجه به تغییر سلیقه مشتریان که ارتباط نزدیکی با شرایط درآمدی، اقتصادی، فرهنگی و موقعیت جغرافیایی و آب و هوایی دارد می تواند دستخوش تغییراتی شود بنابراین لازم است مدیران شرکت این پژوهش را در مورد سایر شعب در بازه های زمانی متفاوتی انجام دهند تا نتایج دقیقتری با قابلیت تعمیم پذیری بیشتری حاصل شود.

با توجه به شیوع و گسترش پاندمی کرونا که استفاده از مواد شوینده را بیشتر ساخته و همچنین به علت محدودیت های کرونایی و استفاده از ماسک و... این احتمال وجود دارد که با عادی شدن شرایط جهانی و کشور در مواجهه با پاندمی کرونا بسیاری از الگوهای رفتاری خرید مشتریان با تغییراتی مواجه شوند که شاید بتوان گفت یک پژوهش جالب مقایسه الگوهای رفتاری مشتریان برای سه دوره زمانی قبل از شیوع و بروز بیماری کرونا (حالت عادی) سال های قبل از ۱۳۹۹، در شرایط شیوع و پاندمی کرونا (اوج زمان همه گیری کرونا) سال های ۱۳۹۹-۱۴۰۰ و سال های بعد از اتمام این همه گیری (که می توان شش ماهه سال های ۱۴۰۱ به بعد را در نظر گرفت). که مقایسه این سه وضعیت احتمالاً کشفیات بسیاری را عاید شرکت مورد بررسی در مورد الگوهای رفتاری مشتریان خود خواهد نمود.

پیشنهاد می شود این پژوهش را بار دیگر با استفاده از شاخص های دان، دیویس بولدین و دیگر روش های ارزیابی کیفیت و بهینگی خوشه بندی انجام داده و با نتایج پژوهش حاضر مقایسه کرده تا چنانچه هر کدام از روش ها یا تکنیک ها نتایج بهتری داشته باشد مدیران شرکت از همان روش مناسب تر بهره ببرند و خوشه بندی و دسته بندی مشتریان را بر اساس نتایج جدید حاصله تعیین نمایند.

پیشنهاد می شود مدیران محترم شرکت و یا پژوهشگران علاقمند به مقوله داده کاوی در شرکت های خرده فروشی از دیگر روش های دسته بندی و کشف الگوهای رفتاری خرید مشتریان مانند درخت های تصمیم C5.0، (درخت دسته بندی و رگرسیون) C&RT و... به عنوان ابزاری کمکی در کشف الگوهای رفتاری بهره برده تا با مقایسه آن با نتایج این پژوهش بتوان با قدرت تحلیل بالاتری قضاوت نمود البته شاید بتوان از ابزارهای داده کاوی ترکیبی با رویکرد عصبی فازی نیز استفاده کرد تا شاهد نتایج جالبی بود که در صورت بهبود نتایج از روش مربوطه برای تصمیم گیری و ارایه الگوهای مصرف و بالتبع آن راهبردهای بازاریابی و فروش و تبلیغات شرکت بهره برد.

این پژوهش در یکی از شرکت های پیشروی گروه صنعتی گلرنگ در عرضه محصولات بهداشتی و شوینده انجام شده است که سهم بزرگی از بازار مربوطه را در ایران دارد لکن چون در استان تهران انجام شده است بهتر است در تعدادی از شعبه های استانی نیز پژوهش مجدد انجام شده و در کنار آن برای افزایش قدرت تعمیم پذیری الگوهای رفتاری خرید مشتریان را در دو سه شرکت دیگر گروه صنعتی گلرنگ که در این زمینه فعالیت دارند را نیز مورد تجزیه و تحلیل قرارداد تا بتوان الگوهای حاکم بر شیوه خرید و نوع محصولات خریداری شده را به تفکیک هر استان و شهر استخراج نمود تا در نهایت بتوان سند چشم انداز الگوی مصرف کشور را استخراج کرد.

منابع

آخوندزاده نوقایی، الهام، البدوی، امیر، سپهری، محمد مهدی، (۱۳۹۴) "تحلیل تغییرات ساختاری بخش های مشتریان با روش ترکیبی خوشه بندی و قوانین انجمنی"، **مدیریت بازرگانی**، شماره ۷، دوره ۳، صص ۵۱۵-۵۴۲.

- اخوان، آفرین، جهادی نائینی، فرزانه، ابویی، محمدحسین (۱۳۹۹). "تبیین بخش‌بندی مشتریان بازارهای صنعتی با نقشه‌های خودسازمانده" **کاوش‌های مدیریت بازرگانی**، شماره ۱۲، دوره ۲۴، صص ۱-۱۸. Doi: 10.22034/bar.2021.4273.1534
- ایزدی، بهرام، بهرام رنجبریان، سعیده کتابی، و فریا نصیری مفخم، (۱۳۹۵) "یک رویکرد جامع برای بخش‌بندی بازار و طبقه‌بندی مشتریان با استفاده از روش‌های داده‌کاوی و برنامه‌ریزی خطی"، **مدیریت تولید و عملیات**، شماره ۷، دوره ۱، صص ۱-۲۲.
- بشردوست، امید، اصغری‌زاده، عزت اله، افشارکاظمی، محمد علی، (۱۴۰۰)، "ارایه الگوی دسته‌بندی مشتریان با رویکرد داده‌کاوی ترکیبی"، **مدیریت کسب و کار**، شماره ۵۰، صص ۸۵-۱۱۱.
- بشردوست، امید، اصغری‌زاده، عزت اله، افشارکاظمی، محمد علی، (۱۴۰۱)، "تحلیل خوشه‌ای مشتریان بر مبنای مدل WRFM با رویکرد داده‌کاوی غیرنظارتی (مورد مطالعه محصولات بهداشتی و آرایشی)"، **پژوهش‌های نوین در تصمیم‌گیری**، شماره ۷، دوره ۱، صص ۱۹۸-۲۲۳.
- تقوا، محمد رضا، حسینی بامکان، سید مجتبی، (۱۳۹۰)، "ارایه خدمات مناسب به مشتریان بالقوه با استفاده از تکنیک‌های داده‌کاوی در حوزه بانکداری الکترونیک"، **مطالعات مدیریت صنعتی**، شماره ۹، دوره ۲۳، صص ۱۸۷-۲۰۷.
- تقوی فرد، محمد تقی، خانی، امیر محمد، لطفی، فاطمه، (۱۴۰۱). "خوشه‌بندی مشتریان در حوزه بانکداری الکترونیک با بهره‌گیری از تراکشن‌های الکترونیکی و اطلاعات دموگرافیک (مورد مطالعه: بانک رفاه)". **مدیریت تبلیغات و فروش**، شماره ۱، دوره ۳، صص ۴۲-۵۶. Doi: 10.52547/JABM.3.1.42.
- تیموریان، محمدشاهین، خدایاری، بهناز، سعیدی، حمید. (۱۴۰۲). "ارائه مدل مدیریت ارتباط با مشتری با تاکید بر مدیریت دانش مشتری در صنعت بانکداری". **مدیریت تبلیغات و فروش**، شماره دو، دوره ۴. Doi :10.52547/JABM.3.2.14335۴
- حسین زاده شهری، معصومه، کرمی، مسعود، مهربانی، مهناز، (۱۳۹۴)، "بخش‌بندی مشتریان در رستوران‌های زنجیره‌ای بر اساس سبک تغذیه (مطالعه موردی: رستوران‌های زنجیره‌ای فست فود بوف در شهر تهران)"، **مدیریت بازرگانی**، شماره ۷، دوره ۱، صص، ۸۳-۹۹. Doi: 10.22059/jibm.2015.50723
- خواجهوند، سمانه، تقوی فرد، محمد تقی، نجفی، اسماعیل، (۱۳۹۱)، "بخش‌بندی مشتریان بانک صادرات ایران با استفاده از داده‌کاوی"، **مطالعات مدیریت بهبود و تحول**، شماره ۲۱، دوره ۶۷، صص ۱۷۹-۲۰۰.
- رنگریز، حسن، بایرامی شهریور، زهرا، (۱۳۹۸)، "اثر E-CRM بر روی وفاداری مشتری با استفاده از روش‌های داده‌کاوی"، **مطالعات مدیریت کسب و کار هوشمند**، شماره ۷، دوره ۲۷، صص ۱۷۵-۲۰۵.
- ریسی وانانی، سینا، ریسی وانانی، ایمان، تقوی فرد، محمد تقی، (۱۳۹۹)، "مدلی برای بخش‌بندی یادگیرندگان و بهبود عملکرد آموزشی با استفاده از الگوریتم‌های داده‌کاوی"، **مطالعات مدیریت کسب و کار هوشمند**، شماره ۹، دوره ۳۳، صص ۵-۳۸.
- شهرابی، جمال، (۱۳۹۴) **داده‌کاوی**، تهران، سروش گیتا.
- غریب، ایمان، طلوعی، عباس، حیدرزاده، کامبیز، (۱۳۹۸)، "ارایه الگوی ترکیبی داده‌کاوی با استفاده از قواعد انجمنی و خوشه‌بندی جهت شناسایی الگوهای غالب رفتار مشتریان (مطالعه موردی: بانک انصار)"، **آینده پژوهی مدیریت**، شماره ۳۰، دوره ۳، صص ۱۸۹-۲۰۱.

غفاری آشتیانی، پیمان، داودی، معصومه، (۱۳۹۵)، "بررسی و تحلیل سبد بازار و چیدمان محصولات فروشگاه‌های زنجیره‌ای"، **کاوش‌های مدیریت بازرگانی**، شماره ۸، دوره ۱۶، صص ۱۶-۱۸۴.

کشاوری حدادها، اعظم، جلیلی بال، زهرا، حاجی یخچالی، سیامک، (۱۳۹۷)، "خوشه‌بندی، ارزیابی و انتخاب پروژه‌ها با رویکرد کوله پشتی و روش‌های تصمیم‌گیری چندمعیاره"، **مطالعات مدیریت صنعتی**، شماره ۱۶، دوره ۵۰، صص
Doi: 10.22054/jims2018.9112. ۲۵۵-۲۲۹

کوهزادی، فواد، قره بیگلر، حسین، بودافی خواجه نوبر، حسین، علوی متین، یعقوب. (۱۴۰۱). "طراحی مدل مدیریت ارتباط با مشتری مبتنی بر کلان داده". **مدیریت تبلیغات و فروش**، شماره ۱، دوره ۳، صص ۱۱۲-۱۳۰. Doi: 10.52547/JABM.3.1.112.

محمودی، سید عباس، میرزایی، کمال، محمودی مصطفی، (۱۳۹۶)، "تعیین عوامل مؤثر بر سرطان معده با استفاده از رویکرد داده کاوی"، **مجله دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران (پیاورد سلامت)**، شماره ۱۱، دوره ۳، صص ۳۳۲-۳۴۱.

مک لنان، جیمی؛ تانگ، جاثوهوی و کریوات، بوگدان (۱۳۹۲)، **داده کاوی با اس.کیو.ال سرور ۲۰۰۸**. ترجمه خازناد، امیرعلی، کوشش، پاتریس، مشهد، انتشارات پرتونگار.

مهاجر باجگیران، مرتضی، پویا، علیرضا، ناجی عظیمی، زهرا، فدائی، سمیه، (۱۴۰۱)، "طراحی مدلی تلفیقی برای خوشه‌بندی اقلام انبار و تخصیص آنها به محل انبارش"، **مطالعات مدیریت صنعتی**، شماره ۲۰، دوره ۶۶، صص ۱۵۳-۱۹۵
doi: 10.22054/jims.2019.41124.2299.

Babaiyan, V. and S.A. Sarfarazi, S.A. (2019) "Analyzing Customers of South Khorasan Telecommunication Company with Expansion of RFM to LRFM Model". **Journal of AI and Data Mining**, vol.7, no. 2, pp. 331-340.

Barragan, J. F., Fontes, C. H., & Embiruçu, M. (2016). A wavelet-based clustering of multivariate time series using a multiscale SPCA approach. *Computers & Industrial Engineering*, 95, 144-155.

Bashiri Mosavi, A. and Afsar, A. (2018) "Customer value analysis in banks using data mining and fuzzy analytic hierarchy processes" **International Journal of Information Technology & Decision Making**, vol. 17, no. 3, pp. 819-840.

Bhojani, S. and Bhatt, N. (2016) "Data Mining Techniques and Trends A Review" **Global Journal for Research Analysis (GJRA)**, Vol. 5, no. 5, pp. 252-254.

Chakraborty, G. (2013). "Customer segmentation using SAS Enterprise Miner". Cary, NC: SAS.Institute Inc.

Chang, H.H. and Tsay, S.F. (2004) "Integrating of SOM and K-means in data mining clustering: An empirical study of CRM and profitability evaluation" **Journal of Information Management**, vol.11, no. 4, pp. 161-203.

Darrab, S., Bhardwaj, P., Broneske, D. and Saake, G. (2022) "OPECUR: An Enhanced Clustering-Based Model for Discovering Unexpected Rules. In: Li B. et al. (Eds) *Advanced Data Mining and Applications*" ADMA 2022. **Lecture Notes in Computer Science**, vol 13088. Springer, Cham. 2022 <https://doi.org/10.1007/978-3-03095408-6-3>.

Dibb, S. (1998) "Market segmentation: strategies for success" **Marketing Intelligence and Planning**, Vol. 16, no.7, pp. 394- 40.

El-Serag, H.B. (2012) "Epidemiology of viral hepatitis and hepatocellular carcinoma" **Gastroenterology**, Vol. 142, no.6, pp. 1264-1273.

Griffin, J., Bahasa, A. and Kartini Yahya, DWI. (2003) **Customer loyalty, Menumbuhkan dan Mempertahankan Kesetiaan Pelanggan**, Jakarta, Indonesia, Penerbit Erlangga.

Guney, S., Peker, S. and Turhan, C.(2020) "A Combined Approach for Customer Profiling in Video on Demand Services Using Clustering and Association Rule Mining," in **IEEE Access**, vol. 8, pp. 84326-84335. Doi: 10.1109/ACCESS.2020.2992064.

Han, J. and Kamber, M. (2010) "**Data mining concepts and techniques**", 3rd ed. Elsevier, Morgan Kaufmann Publishers, Waltham, MA, USA.

Ha, S.H. and Bae, S.M. (2006) "Keeping Track of Customer Life Cycle to Build Customer Relationship, Lecture Notes in Computer Science", **Advanced Data Mining and Applications: Second International Conference, ADMA**, Xi'an, China, August 14-16, 2006, Proceedings. Germany: Springer.

Huang, Z. & Yue H., (2020) "Research on improved RFM customer segmentation model based on K-Means algorithm", **5th International Conference on Computational Intelligence and Applications (ICCIA)**, pp. 24-27.

Jutamat, J. and Mori, T. (2019) **Customer clustering for a new method of marketing strategy support within the courier business**, **Academia Book Chapter**, Vol. 31, no. 2, pp. 1-19.

Kashwan, K. R., & Velu, C. M. (2013). Customer segmentation using clustering and data mining techniques. **International Journal of Computer Theory and Engineering**, 5(6), 856–861.

LakshmiPadmaja, D. and Vishnuvardhan, B. (2018) "Classification and Performance Improvement Using Random Subset Feature Selection Algorithm for Data Mining", **big data research**, Elsevier. Vol. 12, pp. 1-12. <https://doi.org/10.1016/j.bdr.2018.02.007>.

Le, T., Agard, B., & Deveault, S. (2009). Application du data mining à la segmentation du marché des meubles aux États-Unis. In **Eighth Congress International de Genie Industriel**.

Mediana, A., Madyatmadja, E.D. and E. Miranda.(2018)"Customer segmentation in XYZ bank using K-means and K-medoids clustering".**International Conference on Information Management and Technology (ICIMTech)**, pp. 412-416.

Murray, P., Agard, B., & Barajas, M. (2017). "Market segmentation through data mining: a method to extract behaviors from a noisy data set". **Computers & Industrial Engineering**, 109, 233–252.

Murray, P., Agard, B., & Barajas, M. (2018). "Forecast of individual customer's demand from a large and noisy dataset". **Computers & Industrial Engineering**, 118, 33–43.

Natalia, F. L., 2020" Segmentation of Mobile Applications Users: Classification of Customer Loyalty Behavior in Online Shopping Platform", **International Conference on Information Management and Technology (ICIMTech)**, pp.126- 129

Rokach L., Maimon O. (2005) Clustering Methods. In: Maimon O., Rokach L. (Eds) **Data Mining and Knowledge Discovery Handbook**. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_15.

Sevlian, R., & Rajagopal, R. (2018). A scaling law for short term load forecasting on varying levels of aggregation. **International Journal of Electrical Power & Energy Systems**, 98, 350–361.

Sheikh, Alireza, Tohid Ghanbarpour, and Gholamiangonabadi Davoud, (2019). "A Preliminary Study of Fintech Industry: A Two-Stage Clustering Analysis for Customer Segmentation in the B2B Setting", **Journal of Business-to-Business Marketing**, 26(2): 197-207.

Tan, P.N., Steinbach, M. & Kumar, V. (2006). **Introduction to Data Mining**. Pearson Education Inc, USA

Tavakoli, M., Masoumi, M., & Rahmani, E.,(2018) "Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and Data Mining Techniques: A Case Study, **15th International Conference on e-Business Engineering (ICEBE)**, pp.119-126.

Tsiptsis, K. & Chorianopoulos, A. (2009). **Data mining techniques in CRM: Inside customer segmentation**, John Wiley & Sons, Chichester

Wu, Lifeng & Peng, Youwen & Fan, Junliang & Wang, Yicheng & Huang, Guomin,(2021)
"A novel kernel extreme learning machine model coupled with K-means clustering and firefly
algorithm for estimating monthly reference evapotranspiration in parallel computation,"
Agricultural Water Management, Elsevier, vol. 245(C). DOI:
10.1016/j.agwat.2020.106624.