Archive of SID.ir

Customer Value Analysis Based on WRFM Model with the Combined Data Mining Method (Case Study of Hygienic and Cosmetic Products)

Omid Bashardoust
PhD in Industrial Management; Department of Management; Faculty of Accounting and Management; Roudehen Branch; Islamic Azad University; Roudehen, Iran;
Email: dr.o.bashardoust@gmail.com

Ezattollah Asgharizadeh*
PhD in Industrial Management; Associate Professor; Department of Industrial Management; Faculty of Management; Tehran University; Tehran, Iran Email: asghari@ut.ac.ir

MohammadAli Afsharkazemi
PhD in Industrial management; Associate Professor; Department of Management; Faculty of Management; Central Tehran Branch; Islamic Azad University; Tehran, Iran;
Email: M_afsharkazemi@iauec.ac.ir

Iranian Journal of Information Processing and Management

Received: 10, Oct. 2021 **Accepted: 08, Mar. 2022**

Abstract: The accumulated volume of customer information due to the growth and development of information technology and creation of databases have led companies which want to provide better services to their customers to benefit from new tools of customer relationship. One of these tools and methods is data mining techniques that can play an important and key role in customer relationship management. The purpose of this study is to analyze customer value with a combined data mining approach based on the WRFM model. So 64858 samples from customer database in the period of 2019-2020 have been selected by available purposive sampling method. The weight of WRFM attributes has been determined by surveying 3 experts of the company using a hierarchical analysis process. Based on the initial variables of the research and the variables obtained from the attributes of the WRFM model, the purchase value of customers has been analyzed. SPSS Modeller and SPSS software were used to analyze the data.The results show that the K-Means clustering method has a better performance in customer segmentation than the TwoStep clustering and the Cohonen neural network methods. Finally, based on the criteria of purity

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 38 | No. 2 | pp. 571-606
Winter 2023
https://doi.org/10.35050/JIPM010.2022.035

* Corresponding Author

Archive of SID.ir

Archive of SID.ir

percentage, repetition, error rate and Normalized Mutual Information (NMI (index, six clusters with NMI (0.631) were selected from different K-Means clustering.
This study introduces the WRFM model for customer value analysis. The weight of the attributes of this model is based on a survey of experts and using a hierarchical analysis process based on the degree of incompatibility (0.052) obtained from the hierarchical analysis method (0.15), (0.29) and (0.56), respectively, these values indicate the greater importance of the monetary value index than the other two indices. Finally, using naming market segments methods, these six clusters that were selected from different K-Means clustering divided into 4 general categories; key and special customers, golden potential customers, missing uncertain customers and new uncertain customers. According to the research model, the company should focus more on its specific and key customers; i.e. customers who are in the first, third and fifth clusters, namely loyal customers who have higher than average values in the two attributes: monetary value and frequency. Company should consider effective marketing strategies for loyal customers in order to more profitability for the company while maintaining customer relationship management.

Keywords: Clustering, Cohonen Neural Net, Customer Relationship Management, Customer Value Analysis, Data Mining, WRFM Model

Archive of SID.ir

Archive of SID.ir

تجزیه و تحلیل ارزش مشتریان بر مبنای مدل WRFM با روش ترکیبی داده‌کاوی (مورد مطالعه: محصولات بهداشتی و آرایشی)

امید بشردوست
دانش‌آموختهٔ دکتری مدیریت صنعتی؛ دانشکدهٔ مدیریت و حسابداری؛ دانشگاه آزاد اسلامی؛ واحد رودهن؛ رودهن، ایران؛
dr.o.bashardoust@gmail.com

عزت‌اله اصغری‌زاده
دکتری مدیریت صنعتی؛ دانشیار؛ دانشکدهٔ مدیریت؛ دانشگاه تهران؛ تهران، ایران؛
پدیدآور رابط asghari@ut.ac.ir

محمدعلی افشار کاظمی
دکتری مدیریت صنعتی؛ دانشیار؛ دانشکدهٔ مدیریت؛ دانشگاه آزاد اسلامی تهران مرکزی؛ تهران، ایران؛
M_afsharkazemi@iauec.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۰/۰۷/۱۸ | پذیرش: ۱۴۰۰/۱۲/۱۷ | مقاله برای اصلاح به مدت ۴۰ روز نزد پدیدآوران بوده است.

چکیده: انباشت اطلاعاتی و ایجاد پایگاه‌های داده‌ای باعث شده است شرکت‌هایی که قصد دارند به مشتریان خود خدمات مناسب‌تری ارائه کنند، به‌سوی بهره‌مندی از ابزارهای نوین ارتباط با مشتری حرکت نمایند. یکی از این ابزارها و روش‌ها تکنیک‌های داده‌کاوی است که می‌تواند نقش مهم وکلیدی در مدیریت ارتباط با مشتری ایفا کند. هدف این پژوهش تحلیل ارزش مشتریان با رویکرد ترکیبی داده‌کاوی بر اساس مدل WRFMاست. بر این اساس، ۶۴۸۵۸ نمونه از پایگاه دادهٔ مشتریان در دورهٔ ۱۳۹۸-۱۳۹۹ با روش نمونه‌گیری هدفمند در دسترس انتخاب شده است. وزن شاخص‌های مدل WRFM با نظرسنجی از ۳ کارشناس خبرهٔ یکی از شرکت‌های پیشرو «گروه صنعتی گلرنگ» به کمک فرایند تحلیل سلسله‌مراتبی تعیین شد. بر اساس متغیرهای اولیهٔ پژوهش و متغیرهای به‌دست‌آمده از شاخص‌های مدل WRFM، ارزش خرید مشتریان تجزیه و تحلیل شده است. برای تجزیه و تحلیل داده‌ها از نرم‌افزار SPSS Modeler و SPSS استفاده شد. نتایج نشان می‌دهد که روش خوشه‌بندی K-Means نسبت به روش‌های خوشه‌بندی دو-مرحله‌ای و شبکهٔ عصبی «کوهنن» عملکرد بهتری در بخش‌بندی مشتریان داشته است. سرانجام، بر اساس معیارهای

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۸ | شمارهٔ ۲ | ص‌ص ۵۷۱-۶۰۶
زمستان ۱۴۰۱
https://doi.org/10.35050/JIPM010.2022.035

Archive of SID.ir

Archive of SID.ir

درصد خلوص، تکرار، میزان خطا و شاخص اطلاعات متقابل نرمال‌شده از خوشه‌بندی‌های مختلف K-Means، شش خوشه با امتیاز اطلاعات متقابل نرمال‌شده (۰/۶۳۱) انتخاب شد. این پژوهش مدل WRFM را برای تجزیه و تحلیل ارزش مشتری معرفی کرده است. وزن شاخص‌های این مدل با نظرسنجی از کارشناسان و با استفاده از فرایند تحلیل سلسله‌مراتبی و بر اساس میزان ناسازگاری (۰/۰۵۲) که از روش تحلیل سلسله‌مراتبی به‌دست آمده، به‌ترتیب، (۰/۱۵)، (۰/۲۹) و (۰/۵۶) تعیین شد. این مقادیر نشان‌دهندهٔ اهمیت بیشتر شاخص ارزش پولی نسبت به دو شاخص دیگر است. این شش خوشه، سرانجام، با استفاده از شیوهٔ نامگذاری بخش‌های بازار در پژوهش‌های «چانگ و تسای» و «بابایان و سرفرازی» در چهار دستهٔ کلیِ مشتریان کلیدی و ویژه، مشتریان بالقوه طلایی، مشتریان نامطمئن از دست‌رفته، و مشتریان نامطمئن جدید قرار گرفتند. طبق مدل پژوهش، باید بیشتر بر مشتریان خاص و کلیدی، یعنی مشتریانی که در خوشه‌های اول، سوم و پنجم قرار گرفته‌اند، تمرکز داشت؛ یعنی مشتریان وفاداری که ضمن تداوم خرید خود دارای مقادیر بالاتر از میانگین در دو شاخص ارزش پولی و دفعات خرید بوده و به‌تازگی نیز خریدهای با ارزش ریالی بالا داشته‌اند و همچنین باید با توجه به منابع محدود، استراتژی‌های بازاریابی کارآمدی برای این دسته از مشتریان در نظر گرفت تا ضمن حفظ مدیریت ارتباط با مشتری به سودآوری بیشتر منجر شود.

کلیدواژه‌ها: شبکه عصبی کوهنن، تجزیه و تحلیل ارزش مشتری، خوشه‌بندی، داده‌کاوی، مدل WRFM، مدیریت ارتباط با مشتری

۱. مقدمه

در محیط پرتلاطم و رقابتی امروز که دانش به‌عنوان یک منبع اساسی برای سازمان‌ها و بنگاه‌های اقتصادی مطرح است، مدیریت ارتباط با مشتری به ابزاری راهبردی و حیاتی برای بنگاه‌های اقتصادی و سازمان‌ها تبدیل شده است (عاشوری، شریف‌خانی و تارخ ۱۳۹۳، ۶۲). امروزه، تعامل شرکت‌ها با مشتریان در چارچوب مدیریت ارتباط با مشتری به‌طوری چشمگیر تغییر یافته است. شناسایی ویژگی‌ها و علاقه‌های مشتریان مختلف و تخصیص بهینهٔ منابع به آن‌ها با توجه به ارزشی که برای شرکت‌ها خلق می‌کنند، به یکی از دغدغه‌های اصلی در حوزهٔ مدیریت ارتباط با مشتری تبدیل شده است (کفاش‌پور، توکلی و علیزاده زوارم ۱۳۹۱، ۶۱). به‌منظور 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 بهینهٔ منابع محدود، به‌کارگیری راهبردهای مناسب بازاریابی و سرانجام، مدیریت سودآوری فراهم می‌کند (رنگریز و بایرامی شهریور ۱۳۹۸، ۱۷۶). شناسایی مشتریان و نیازهای آن‌ها مؤلفه‌های مهمی هستند که به تولیدکنندگان برای

Archive of SID.ir

Archive of SID.ir

کسب مزیت رقابتی در ارائهٔ محصول و نیز ارائهٔ خدمات به مشتریان کمک می‌کند. شرکت‌ها باید مشتریان خود را اولویت‌بندی کنند و تمرکز بیشتری روی مشتریان کلیدی خود داشته باشند (Babaiyan & Sarfarazi 2019, 331). سازمان‌ها می‌توانند به اهداف اصلی خود، که همانا کسب مزیت رقابتی است، از طریق شناسایی گروه‌های مختلف مشتریان و نیازهای آن‌ها و برای رضایت‌مندی مشتریان و ایجاد سودآوری بیشتر دست یابند (این عمل خود وفاداری بلندمدت مشتریان را در پی خواهد داشت). سازمان‌ها همچنین، ضمن شناسایی مشتریان اصلی و حفظ آن‌ها، نه‌تنها می‌توانند مشتریان جدیدی جذب کنند، بلکه برای پرکردن جای خالی مشتریانی که تصمیم به قطع رابطه با شرکت گرفته‌اند نیز از آن بهره ببرند؛ چرا که هزینهٔ جذب مشتری جدید ۵ الی ۱۰ برابر هزینهٔ حفظ مشتری فعلی است. شرکت‌ها همچنین، احتمال موفقیت بیشتری در فروش محصولات خود به مشتریان موجود در قیاس با مشتریان بالقوه دارند (Armstrong et al. 2014).

در کنار مباحث مدیریت ارتباط با مشتری، امکانات استفاده از شبکهٔ اینترنت باعث مواجه‌شدن با حجم زیادی از داده و اطلاعات شده است. داده‌کاوی یکی از ابزارهای استخراج دانش و الگوهای موجود در پایگاه داده‌های منتج از انباشت اطلاعات مشتریان است (رنگریز و بایرامی شهریور ۱۳۹۸). یکی از روش‌های هر تصمیم‌گیری مدیریتی، داشتن راهکارهای عملیاتی به همراه احتمال موفقیت و اثربخشی این راهکارهاست و استفاده از تکنیک‌های داده‌کاوی می‌تواند این مهم را محقق سازد (رئیسی وانانی، رئیسی وانانی و تقوی‌فرد ۱۳۹۹، ۷).

در این پژوهش، صنعت محصولات بهداشتی و آرایشی مورد بررسی قرار گرفته و با توجه به این‌که این صنعت روزانه با حجم انباشته‌شده‌ای از اطلاعات خرید و جمعیت‌شناختی مشتریان مواجه است، انجام پژوهشی که در راستای ارائهٔ خدمات بهتر به مشتریان گروه‌های مختلف بتواند با استفاده از داده‌کاوی به بخش‌بندی مشتریان بپردازد و با این عمل باعث بهبود عملکرد فروش و سودآوری شرکت شود، برای مدیران و تصمیم‌گیرندگان ارشد سازمان ضرورت می‌یابد. بنا بر مطالب عنوان‌شده و اهمیت داده‌کاویِ پایگاه داده‌های مشتریان به‌منظور مدیریت بهتر ارتباط با مشتری، این پژوهش به‌دنبال آن است که بررسی نماید کدام روش‌های 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 سه سؤال

Archive of SID.ir

Archive of SID.ir

زیر بیان کرد:

۱. بهترین تعداد خوشه برای بخش‌بندی مشتریان با استفاده از شاخص‌های درصد خلوص و امتیاز اطلاعات متقابل نرمال‌شده[1] چه تعداد است؟

۲. کدام روش خوشه‌بندی داده‌کاوی این پژوهش برای بخش‌بندی مشتریان محصولات آرایشی و بهداشتی مناسب‌تر است؟

۳. از نقطه‌نظر شاخص‌های مدل پژوهش، مشتریانی که برای شرکت سودآورترند، چه ویژگی‌هایی دارند؟

۲. پیشینهٔ پژوهش

۲-۱. پیشینهٔ نظری پژوهش

۲-۱-۱. مدیریت ارتباط با مشتری[2]

سابقهٔ مدیریت ارتباط با مشتریان و پیاده‌سازی آن به‌صورت امروزی، به دههٔ ۱۹۹۰ برمی‌گردد. مشتری‌مداری و رضایت مشتری از مباحث جدید در جهان امروز است. نگرش بازاریابی ابتدا از تولیدمداری و محوریت تولید آغاز و به فروش‌مداری، محوریت بازار و سرانجام، با استفاده از فناوری اطلاعات به مشتری‌مداری تبدیل شده است (صمدی راد ۱۳۸۰). بحث این‌که سازمان چگونه مشتریان فعلی خود را حفظ کرده و به آن‌ها خدمات بهتری ارائه کند و بحث فناوری اطلاعات و ابزارهای جدیدی که توسط این فناوری ایجاد شده، باعث شده است که مدیریت ارتباط با مشتری به مبحثی نو و تازه در مهندسی و مدیریت تبدیل شود (صالحی صدقیانی و اخوان ۱۳۸۵).

مدیریت ارتباط با مشتری از فرایندهای سنتی بازاریابی، فروش و ارائهٔ خدمات فراتر رفته و تا مرزهای سازمانی گسترش می‌یابد. سازمان با بهره‌گیری از اهرم فناوری پایگاه داده‌ها برای هماهنگی ارتباط مشتری با شرکت می‌تواند ارزش ارتباط با مشتری را بیشینه کند و میزان حفظ مشتری را بهبود بخشد (صنایعی ۱۳۸۱، ۳۰).

۲-۱-۲. بخش‌بندی مشتریان[3]

بخش‌بندی به شرکت‌ها اجازه می‌دهد تا گروه‌های مشتری را شناسایی کرده و نیازهای آن‌ها را مشخص نموده و انتظارهای مشتریان را به‌طور تخصصی‌تر برآورده ساخته

1. normalized mutual information (NMI) 2. customer relationship management (CRM) 3. segmentation

Archive of SID.ir

Archive of SID.ir

و پاسخگویی به این خواسته‌ها را افزایش دهند و در حالی که کاربران جدید و بالقوه را با راهبردهای بازاریابی مرتبط جذب می‌کنند، باعث افزایش درآمد کلی شرکت شوند. امروزه، تقسیم‌بندی‌های بازارها با توجه به موارد متعددی مانند ۱) تقسیم‌بندی بر اساس خصوصیات جمعیت‌شناختی مشتریان (ویژگی‌هایی مانند سن، جنسیت، درآمد، تحصیلات، شغل و ...)، ۲) تقسیم‌بندی بر اساس خصوصیات روان‌شناسی[1] (مردم بر اساس گرایش‌ها، ارزش‌ها و سبک‌های زندگی گروه‌بندی می‌شوند و به‌طور معمول، داده‌ها با استفاده از پرسشنامه جمع‌آوری می‌شوند)، ۳) تقسیم‌بندی بازار بر اساس خصوصیات رفتاری مشتریان (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[2]، و ۵) تقسیم‌بندی بازارها بر اساس ویژگی‌های جغرافیایی صورت می‌گیرد (کیگان ۲۰۰۴). در این پژوهش برای تقسیم‌بندی بازار مشتریان از ویژگی‌های رفتاری، جمعیت‌شناختی و داده‌های مالی و خرید مشتریان استفاده شده است. همچنین، با استفاده از تعدادی از روش‌های داده‌کاوی غیرنظارتی (خوشه‌بندی)، بخش‌بندی مشتریان صورت می‌گیرد که در ادامه، در بخش یافته‌های پژوهش به آن اشاره می‌شود.

۲-۱-۳. مدل تازگی، تکرار و ارزش پولی[3]

شاید به‌توان گفت که در اوایل دهۀ ۶۰ میلادی بود که برای این‌که بتوان مشتریانی را که احتمال خرید داشتند، از آن‌هایی که قصد خرید نداشتند، جدا کنند، روش کمّی، ساده و مؤثری ابداع شد. این روش، تجزیه و تحلیل تازگی، فراوانی و ارزش پولی یا به‌طور خلاصه RFM نامیده شد. از این زمان به بعد، پژوهشگران مختلفی از این مدل و مدل‌های ارتقایافتۀ آن برای بخش‌بندی و تعیین ارزش مشتریان خود استفاده کرده‌اند. تکنیک تازگی، تکرار و ارزش پولی رفتار مشتری را شناسایی می‌کند و ویژگی‌های رفتار مشتری را بر اساس سه متغیرِ ۱) تکرار (آخرین خرید که به فاصلۀ زمانی بین زمانی که آخرین رفتار خرید مصرف‌کننده اتفاق افتاده و حال حاضر اشاره دارد)، ۲) فراوانی خریدها (که به تعداد معاملات در یک دورۀ خاص اشاره دارد)، و ۳) ارزش پولی خرید (که به مبلغ پول مصرف‌شده در یک دورۀ خاص اشاره دارد) نشان می‌دهد. پیاده‌سازی

1. psychographic segmentation 2. benefit segmentation
3. recency, frequency and monetary value model (RFM)

Archive of SID.ir

Archive of SID.ir

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 (Hosseini, Maleki & Gholamian 2010, 5260).

بیشتر پژوهش‌های اولیه در مورد مدل RFM به برابری وزن هر یک از شاخص‌ها تأکید دارند؛ مانند پژوهش‌های Hughes (1994); Blattberg, Kim & Neslin (2008) ، و «خواجوند، تقوی‌فرد و نجفی» (۱۳۹۱). همچنین، Hosseini, Maleki & Gholamian (2010) و Zaheri et al. (2012) و Stone (1995) پژوهش‌هایی را دربارهٔ بهبود دقت بازاریابی هدف با استفاده از مدل RFM انجام داده‌اند. یکی از این بهبودها ارائهٔ مدل تازگی، تکرار و ارزش پولی موزون[1] است که در این مدل برای هر شاخص، وزنی متناسب با نوع کسب‌وکار و صنعت با توجه به نظر خبرگان صنعت در نظر گرفته می‌شود (Stone 1995). بنابراین، هر وزن متناسب با نظر خبرگان هر صنعت با استفاده از تکنیک‌های فرایند تحلیل سلسله‌مراتبی تعیین می‌شود و این موضوع باعث می‌شود که محدودیت‌های مدل تازگی، تکرار و ارزش پولی موجود به‌طوری مؤثر برطرف شود (Hosseini, Maleki & Gholamian 2010; Khajvand & Tarokh 2011; Zaheri et al. 2012; (Bin, Peiji and Dan 2008;.

«بلت‌برگ، کیم و نسلین» از سه منظرِ زیر به اهمیت بازاریابی پایگاه‌های داده‌های مشتریان پرداخته‌اند: ۱) افزایش بهره‌وری بازاریابی، ۲) ایجاد و تقویت ارتباط با مشتری، و ۳) ایجاد مزیت رقابتی پایدار. آن‌ها همچنین اذعان داشته‌اند که مدل تجزیه و تحلیل RFM اولین مدل پیش‌بینی‌کننده‌ای بوده است که برای بازاریابی پایگاه داده‌های مشتریان استفاده شده است Blattberg, Kim & Neslin (2008).

اولین گام در مدل تازگی، تکرار و ارزش پولی این است که پروندهٔ مشتریان بر اساس این‌که آن‌ها به‌تازگی از شرکت چگونه خرید کرده‌اند، مرتب شده و سپس، پایگاه داده‌ها به پنج قسمت مساوی (پنجک) تقسیم و این پنجک‌ها از شماره ۱ تا ۵ شماره‌گذاری شود. بنابراین، ۲۰ درصد از مشتریانی که به‌تازگی خریدهایی از شرکت داشته‌اند، به شمارهٔ ۵ تعلق خواهند گرفت و ۲۰ درصد بعدی به شمارهٔ ۴ و این کار تا شمارهٔ ۱ ادامه خواهد یافت. مراحل بعد شامل مرتب کردن مشتریان بر اساس تعداد خرید

1. weighted recency, frequency and monetary value model (WRFM)

Archive of SID.ir

Archive of SID.ir

و بعد از آن بر اساس ارزش پولی خرید است. سرانجام، پایگاه داده با توجه به تازگی، تعداد خرید و ارزش پولی به ۱۲۵ گروه تقریباً مساوی (سلول) تقسیم می‌شود. مشتریان و یا محصول‌های با نمرات بالاتر به‌طور معمول، سودآورترین هستند (Hughes 1994; Stone 1995).

۲-۱-٤. داده‌کاوی[1]

داده‌کاوی یک گام اساسی در کشف دانش در پایگاه داده‌هاست. داده‌کاوی به کشف الگوهای جدید از انبوه داده‌ها با تمرکز بر الگوریتم‌های کشف دانش مفید در پایگاه داده اشاره دارد. به‌طور کلی، سه مرحلهٔ اصلی در داده‌کاوی وجود دارد: آماده‌سازی داده‌ها، کاهش داده‌ها، و کشف اطلاعات با ارزش (Bhojani and Bhatt 2016, 252).

شش فعالیت اصلی داده‌کاوی شاملِ ۱) دسته‌بندی، ۲) تخمین، ۳) پیش‌بینی، ۴) گروه‌بندی شباهت، ۵) خوشه‌بندی، و ۶) توصیف و نمایه‌سازی است. سه دستهٔ اولِ آن داده‌کاوی هدایت‌شده بوده و دستهٔ ۴ و ۵ داده‌کاویِ غیرهدایت شده است و دستهٔ آخر به داده‌کاوی هدایت‌شده و هدایت‌نشده تقسیم می‌شود. این گروه‌بندیِ شباهت گاهی با عنوان الگوریتم کشف مغایرت[2] مطرح می‌شود (شهرابی ۱۳۹۴).

۲-۱-۵. گروه‌بندی شباهت (الگوریتم کشف مغایرت)

مدل‌های کشف مغایرت به‌منظور تشخیص نقاط دورافتاده[3] یا موارد غیرمعمول در داده‌ها مورد استفاده قرار می‌گیرند. این مدل‌ها تنها اطلاعات مربوط به رفتارهای نرمال را ذخیره‌سازی می‌کنند. این مدل‌های کشف مغایرت نیازی به مجموعه داده‌ای برای آموزش و یادگیری ندارند (علیزاده و ملک‌محمدی ۱۳۹۳، ۱۲۵- ۱۲۶).

۲-۱-٦. خوشه‌بندی[4]

خوشه‌بندی، تقسیم داده‌ها به گروه‌هایی از اشیای مشابه است. فرایند قرار دادن داده‌ها در گروه‌هایی که بیشترین شباهت را در برخی ویژگی‌ها دارند، خوشه‌بندی نامیده می‌شود. هر خوشه شامل مجموعه‌ای از داده‌های مشابه درون‌خوشه‌ای بوده و متفاوت از داده‌های سایر خوشه‌هاست (Han, Kamber and Pei 2011). فاصله، معیار شباهت است و فرمول اندازه‌گیری فاصله در خوشه‌بندی بسیار مهم است. راه‌های اندازه‌گیری فاصله

1. data mining
2. anomaly detection
3. outlier
4. clustering

Archive of SID.ir

Archive of SID.ir

بسیار متفاوت‌اند (Jintana and Mori 2019). بسیاری از الگوریتم‌های خوشه‌بندی توسعه یافته و از چندین جنبه طبقه‌بندی شده‌اند؛ مانند الگوریتم‌های تقسیم‌بندی[1] بر اساس تقسیم کل داده‌ها به گروه‌های متفاوت، روش‌های تجمع سلسله‌مراتبی[2]، روش‌های تفرقه‌افکنانه یا تقسیمی[3]، روش‌های مبتنی بر تراکم[4]، روش‌های مبتنی بر شبکه[5] و روش‌های مبتنی بر مدل[6] (Bhojani and Bhatt 2016, 253). البته، «روکاچ» درکتابی با عنوان «کتاب راهنمای داده‌کاوی و کشف دانش» روش‌های خوشه‌بندی را چنین نام می‌برد: روش‌های سلسله‌مراتبی[7]، پارتیشن‌بندی، مبتنی بر چگالی، مبتنی بر مدل، مبتنی بر شبکه و محاسبات نرم[8] (Rokach 2010).

شناخته‌شده‌ترین و پرکاربردترین روش تقسیم‌بندی روش خوشه‌بندی k- میانگین است که در پژوهش حاضر از این روش در کنار روش‌های خوشه‌بندی دو-مرحله‌ای[9] و کوهنن[10] (شبکه‌های خودسازمانده) و الگوریتم کشف مغایرت[11] استفاده شده است.

۲-۲. پیشینهٔ تجربی پژوهش

در این قسمت به تعدادی از پژوهش‌هایی که در مواردی شباهت‌هایی با هدف و روش پژوهش حاضر دارند، اشاره می‌شود:

«هووانگ و لی» پژوهشی را با هدف شناسایی و اولویت‌بندی مشتریان محصولات جدید با استفاده از مدل RFM و الگوریتم رتبه‌بندی متنی[12] انجام داده‌اند. این پژوهشگران مدلی با عنوان RWF ارائه کرده و برای اعتبارسنجی مدل پیشنهادی خود از داده‌های وب‌سایت کاربران موبایل کرهٔ جنوبی در بازهٔ زمانی اول ماه می تا ۳۱ ماه جولای ۲۰۲۰ بهره برده‌اند. آن‌ها برای تجزیه و تحلیل اوزان از الگوریتم مذکور استفاده کرده‌اند. نتایج پژوهش آن‌ها نشان می‌دهد که با مدل پیشنهادی نرخ پاسخگویی مشتریان عالی سازمان ۳۰ درصد رشد داشته است. همچنین، نتایج پژوهش این امکان را برای متخصصان فراهم ساخته که بتوانند از مشتریان هدف خود بدون داشتن سابقهٔ خرید قبلی اطلاعات باارزشی تهیه کنند (Hwang and Lee 2021).

«عباسی‌مهر و شعبانی» پژوهشی را با هدف ارائهٔ روشی جدید برای تجزیه و تحلیل

1. partitioning methods
2. hierarchical agglomerative
3. divisive methods
4. density-based methods
5. grid-based methods
6. model-based methods
7. hierarchical,
8. soft-computing
9. two-step clustering
10. Kohonen (self-organizing map)
11. anomaly
12. text rank

Archive of SID.ir

Archive of SID.ir

و پیش‌بینی رفتار پویای مشتریان با استفاده از خوشه‌بندی سری زمانی در مورد مشتریان بانک انجام داده‌اند. داده‌های پژوهش آن‌ها که شامل ویژگی‌های تازگی، تکرار و ارزش پولی خرید بود، از داده‌های دستگاه کارتخوان[1] بخش تجزیه و تحلیل داده‌های بانک به دست آمده است. پس از انجام یک مطالعهٔ تجربی بر روی داده‌های ۲۵۳۱ مشتری که از دستگاه‌های کارت‌خوان استفاده کرده‌اند، روند غالب رفتار مشتریان با استفاده از روش پیشنهادی کشف شده است. در پژوهش آن‌ها بر اساس تجزیه و تحلیل ویژگی پولی، مشتریان به چهار بخش اصلی تقسیم شدند. این پژوهش یک روش جدید برای جذب رفتار پویای مشتریان با استفاده از خوشه‌بندی سری‌های زمانی بر روی داده‌های مرتب‌شده بر اساس زمان ارائه کرده 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 (Abbasimehr & Shabani 2020).

«کاباساکال» مطالعه‌ای موردی با هدف بخش‌بندی مشتریان بر اساس مدل RF در یک خرده‌فروشی الکترونیکی انجام داده است. در این پژوهش یک نمونهٔ اولیه نرم‌افزار معرفی شده است که با استفاده از مدل RFM به روش امتیازدهی کمک می‌کند. داده‌های فروش به‌دست‌آمده برای خوشه‌بندی به کمک نمونهٔ اولیهٔ نرم‌افزار تجزیه و تحلیل شده و خوشه‌های کشف‌شده با انواع RFM با استفاده از معیارهای ارزیابی خوشه مقایسه شده‌اند. سرانجام، برای هر بخش پیشنهادهایی مربوط به استراتژی‌های بازاریابی ارائه شده است (Kabasakal 2020).

«بابایان و سرفرازی» پژوهشی را با عنوان «تجزیه و تحلیل مشتریان مخابرات خراسان جنوبی با گسترش مدل RFM به LRFM[2]» انجام داده‌اند. این پژوهشگران از سوابق خرید ۹۹۹ مشتری که در یک سال توسط مخابرات جمع‌آوری شده، استفاده کرده‌اند. هدف این پژوهشگران طبقه‌بندی مشتریان در خوشه‌های مختلف بوده است تا مشتریانی که مصرف بالایی داشته و یا مشتریانی که ارزش بالایی برای شرکت داشته و از دست رفته‌اند، شناسایی شوند. در این پژوهش از نرم‌افزار «کلمنتاین»[3] برای تجزیه وتحلیل داده‌ها، و برای خوشه‌بندی از روش خوشه‌بندی K- میانگین و خوشه‌بندی دو-مرحله‌ای استفاده شده است.

1. point of sale (POS)
2. Length, Recency, Frequency, Monetary (LRFM)
3. Clementine

Archive of SID.ir

Archive of SID.ir

سرانجام، ۹ خوشه توسط شاخص «دیویس-بولدین»[1] برای روش دو-مرحله‌ای انتخاب شده است و شرکت بر اساس مقدار شاخص‌های مدل LRFM، مشتریان را در ۵ گروه وفادار، بالقوه، جدید، ازدست‌رفته، و پرمصرف قرار داده است. با این تقسیم‌بندی‌ها شرکت قادر خواهد بود تصمیم‌های بهتری را درآینده برای هرگروه از مشتریان خود داشته باشد (Babaiyan & Sarfarazi 2019).

«رنگریز و بایرامی شهریور» پژوهشی را با استفاده از تکنیک‌های داده‌کاوی با هدف بررسی تأثیر مدیریت ارتباط با مشتریان الکترونیکی[2] بر وفاداری مشتریان بر روی داده‌های تراکنشی ۴۰۶۶۶ مشتری بانک ملت در بازهٔ زمانی شش ماه خاص از سال ۱۳۹۴ انجام داده‌اند. در این پژوهش از روش‌های خوشه‌بندی k- میانگین و شبکه‌های عصبی (با الگوریتم پس انتشار خطا)[3] و مدل LRFM به همراه برنامه‌نویسی در نرم‌افزارهای «متلب» و «اکسل» استفاده شده است. نتایج پژوهش حکایت از آن داشته که با افزایش میزان استفادهٔ مشتریان از خدمات E-CRM وفاداری آن‌ها بیشتر می‌شود. رابطه بین مدیریت ارتباط با مشتری الکترونیکی و مؤلفه‌های مدل LRFM و وفاداری، یک رابطه غیرخطی بوده و میزان تغییر در وفاداری به ازای تغییر E-CRM، مقداری ثابت نیست. میزان بالا رفتن وفاداری مشتریان تابعی از مؤلفه‌های LRFM، مقدار E-CRM و وزن‌های به‌دست‌آمده در شبکهٔ عصبی است و سرانجام، دو خوشهٔ کلی وفادار و غیروفادار برای مشتریان به‌دست آمده است و با توجه به خوشه‌ای که مشتریان در آن قرار گرفته‌اند، راهبردهای خاص بازاریابی و فروش برای مشتریان در نظر گرفته شده است (۱۳۹۸).

«مارجیانتی» و همکاران، پژوهشی را با هدف شناسایی مشتریان کلیدی و سودآور شرکت با عنوان «بخش‌بندی مشتریان بر اساس الگوریتم انتشار وابستگی و مدل RFM، بر روی ۱۰۰۰ مشتری استفاده‌کننده از خدمات موزیک آنلاین انجام دادند. این پژوهشگران به این نتیجه رسیدند که اگر از روش فاصلهٔ اقلیدسی با ترجیح میانه و فاصلهٔ «منهتن»[4] با ترجیح حداقلی برای محاسبهٔ الگوریتم انتشار وابستگی استفاده شود، نتایج بهتری ایجاد می‌شود. این پژوهشگران از نرم‌افزار «متلب» و روش استاندارد شش-مرحله‌ای داده‌کاوی[5] بهره بردند. سرانجام، از بین ۵۱ خوشهٔ ایجادشده، سه دستهٔ کلی به شرح باارزش‌ترین

1. Davies Bouldin index
2. Electronic- Customer Relationship Management (E-CRM)
3. back propagation algorithm
4. Manhattan
5. cross industry standard process for data mining (CRISP-DM)

Archive of SID.ir

Archive of SID.ir

مشتریان با (۳/۷) درصد معادل ۳۷ نفر، رشدپذیرترین مشتریان با (۲۶/۳) درصد معادل ۲۶۳ نفر و سرانجام، مشتریان زیر صفر با ۷۰ درصد معادل ۷۰۰ نفر به‌دست آمد. نتایج پژوهش حکایت از آن داشته که به‌کارگیری روش انتشار وابستگی درکنار مدل تازگی، تکرار و ارزش پولی جزئیات بیشتری را می‌تواند در مورد مشتریان در اختیار شرکت قرار دهد تا مدیران مربوط بتوانند راهبردهای مؤثرتری را به‌کار گیرند (Margianti et al. 2016).

از بررسی پیشینهٔ پژوهش می‌توان گفت که زمینهٔ به‌کارگیری داده‌کاوی به‌منظور خوشه‌بندی و دسته‌بندی مشتریان، چه به‌صورت مجزا و چه به‌صورت ترکیبی به‌منظورکشف و استخراج دانش از پایگاه داده‌های مشتریان، بسیار متنوع و متفاوت بوده و بنا به هدف، نوع و ماهیت پژوهش‌ها، روش‌های مختلفی از خوشه‌بندی از جمله K-میانگین، دو-مرحله‌ای و شبکهٔ عصبی «کوهنن»، عملکرد، کیفیت و دستاوردها و نتایج متفاوتی در پی داشته است. همچنین، بهبود شیوه‌های خوشه‌بندی و دسته‌بندی مشتریان به‌ویژه در عرصهٔ خدمات بانکداری باعث صرفه‌جویی در زمان، هزینه و رضایتمندی کاربران و استفاده‌کنندگان (مصرف‌کنندگان) شده و در صنعت خرده‌فروشی نیز باعث افزایش فروش و تخصصی‌تر شدن مباحث بازاریابی مدیریت ارتباط با مشتری شده است. در پژوهش حاضر، ضمن بررسی روش‌های متداول خوشه‌بندی (K- میانگین، دو-مرحله‌ای، شبکهٔ عصبی کوهنن)، با توجه به شرایط پژوهش و بازار مشتریان مورد بررسی، به‌دنبال بررسی بهترین روش خوشه‌بندی در حیطهٔ مدیریت ارتباط با مشتریان هستیم.

۳. روش پژوهش

این پژوهش از نظر هدف، کاربردی-توسعه‌ای است، ولی از نظر ماهیت انجام پژوهش جنبهٔ توصیفی اکتشافی دارد. در این پژوهش، قسمت‌هایی که به نمایش اطلاعات جمعیت‌شناختی مشتریان جامعهٔ هدف می‌پردازد، ماهیت توصیفی داشته و قسمت‌هایی که مربوط به یافته‌های پژوهش بوده و از داده‌کاوی استفاده کرده، جنبه تحلیلی و اکتشافی دارد.

جامعهٔ آماری این پژوهش را مشتریان خرده‌فروشی، تعاونی مصرف و محلی که محصولات بهداشتی و آرایشی یکی از شرکت‌های پیشرو در «گروه صنعتی گلرنگ» را در بازهٔ زمانی ۱۳۹۸-۱۳۹۹ خریداری کرده‌اند و محدود به استان تهران هستند، تشکیل می‌دهد. از این بین، ۶۴۸۵۸ نمونهٔ آماری با روش نمونه‌گیری هدفمندِ در دسترس از

Archive of SID.ir

Archive of SID.ir

اطلاعات مشتریان شعبهٔ 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 شده است (این روش نمونه‌گیری به این علت انتخاب شده است که در داده‌کاوی مهم‌ترین مسئله، دسترسی به اطلاعات مرتبط، صحیح و با حجم بالاست؛ به‌طوری که بتوان از آن، اطلاعات باارزشی را برای جامعهٔ هدف به‌دست آورد). این اطلاعات در قالب خروجی‌های صفحه‌های گسترده (اکسل) شامل اطلاعات جمعیت‌شناختی و کمّی خرید مشتریان بوده و هر سطر آن مربوط به یک مشتری و هر مشتری دارای ۱۹ ویژگی[1] (ستون) بوده است. تجزیه و تحلیل اولیه و تعیین مقدار شاخص‌های مدل تازگی، تکرار و ارزش پولی موزون به کمک نرم‌افزار «اس‌پی‌اس‌اس» نسخهٔ ۲۱ انجام شد تا شناخت کلی نسبت به ماهیت داده‌های اولیهٔ پژوهش ایجاد شود. سپس، با استفاده از مدل تازگی، تکرار و ارزش پولی شاخص‌های مربوط به هر مشتری استخراج و به متغیرهای اولیهٔ پژوهش اضافه شد تا در ادامه، از ۳ نفر از مدیران ارشد (خبرگانِ) شرکت که دارای سابقه و دانش لازم در زمینهٔ محصولات آرایشی و بهداشتی بوده‌اند، خواسته شد که نظر خود را دربارهٔ اوزان هر یک از شاخص‌های مدل تازگی، تکرار و ارزش پولی در قالب ماتریس مقایسه‌های زوجی بیان کنند (بر پایهٔ میانگین حسابی سطری، وزن هر شاخص ایجاد شده است). سپس، با استفاده از فرایند تحلیل سلسله‌مراتبی و به روش میانگین هندسی نظرهای خبرگان با هم ادغام شده و برای تعیین ثبات و اعتبار وزن‌های ایجادشده شاخص ناسازگاری نیز محاسبه و بررسی شده است.

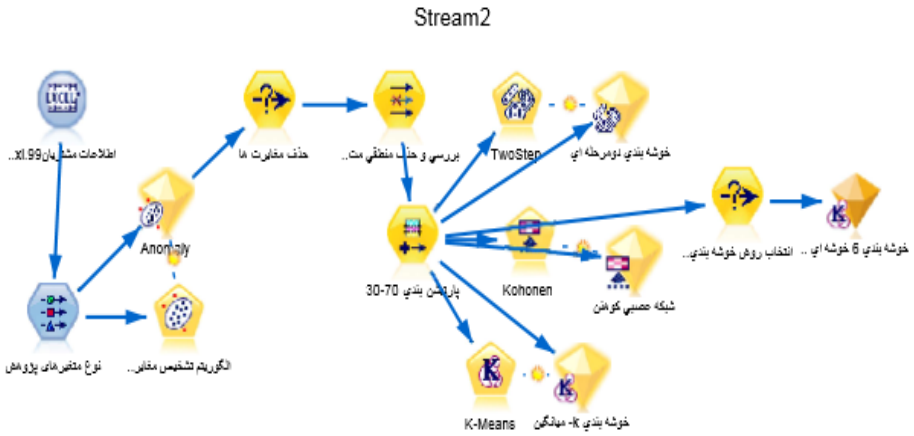
بعد از اطمینان از وزن‌های به‌دست‌آمده، این وزن‌ها در مقادیر مربوط به هر شاخص مدل تازگی، تکرار و ارزش پولی ضرب شد تا مقدار نهایی شاخص تازگی، تکرار و ارزش پولی موزون (WRFM)[2] برای هر مشتری به‌دست آید. آنگاه، این چند متغیر جدید به متغیرهای اولیهٔ پژوهش اضافه و داده‌کاوی با هدف تحلیل خوشه‌ای ارزش مشتریان به کمک «اس‌پی‌اس‌اس مودلر»[3] نسخهٔ ۱۸ بر روی کلیهٔ داده‌های ایجادشده انجام شد. برای تعیین بهترین روش خوشه‌بندی و کیفیت خوشه‌های ایجادشده از شاخص‌های درصد خلوص[4] و اطلاعات متقابل نرمال‌شده نیز استفاده شده است. از ۷۰ درصد داده‌ها برای

1. feature
2. weighted recency frequency monetary value (WRFM)
3. SPSS Modeler
4. purity

Archive of SID.ir

Archive of SID.ir

یادگیری داده‌کاوی و از بقیهٔ آن برای تعیین میزان دقت و صحت آزمون نتایج داده‌کاوی استفاده شد. در شکل ۱، مدل مفهومی مراحل اجرای داده‌کاوی پژوهش به تصویر کشیده شده است.



شکل ۱. فرایند مدل‌سازی داده‌کاوی پژوهش حاصل از خروجی نرم‌افزار (مدل مفهومی)

٤. تجزیه و تحلیل یافته‌ها

آمارهٔ توصیفی شاخص‌های مدل WRFM، پیش از عملیات داده‌کاوی و پیش‌پردازش داده‌های پژوهش، بر اساس ۶۴۸۵۸ دادهٔ رکورد مشتری به‌شرح جدول ۲ است.

جدول ۲. یافته‌های آمارهٔ توصیفی شاخص‌های مدل WRFM

شاخص	حداقل	حداکثر	میانگین	انحراف استاندارد	ضریب واریانس (درصد)
تازگی موزون (WR)	۰/۱۵	۰/۷۵	۰/۴۵	۰/۲۱۲	۴۷/۱۱
تکرار موزون (WF)	۰/۲۹	۱/۴۵	۰/۸۹۵	۰/۳۷۸	۴۲/۲۳
ارزش پولی موزون (WM)	۰/۵۶	۲/۸	۱/۶۷۳	۰/۷۸۹	۴۷/۱۶
تازگی، تکرار و ارزش پولی موزون (WRFM)	۱	۵	۳/۰۱۸	۱/۱۳۱	۳۷/۴۷

با توجه به یافته‌های آمارهٔ توصیفی مربوط به شاخص‌های مدل تازگی، تکرار و ارزش پولی موزون مطابق جدول ۲، میانگین کل امتیاز مدل WRFM (۳/۰۱۸) به‌دست آمده است که در ادامهٔ پژوهش، میانگین شاخص‌های این مدل برای تجزیه و تحلیل وضعیت

Archive of SID.ir

Archive of SID.ir

مشتریان و ارزش آن‌ها مورد استفاده قرار می‌گیرد. در ادامه، به متغیرهای اولیه و ثانویه که در جریان این پژوهش استفاده و ایجاد شده‌اند، در قالب جدول ۳، اشاره می‌شود.

جدول ۳. متغیرهای اولیهٔ پژوهش به‌تفکیک نوع، مقیاس و نقش

ردیف	نام متغیر	نوع متغیر	مقیاس اندازه‌گیری	وضعیت	نقش
۱	گروه مشتریان سطح ۱ (سه حالت)	رشته‌ای[1]	اسمی[2]	متغیر پایهٔ اولیه	مستقل
۲	گروه مشتریان سطح ۲ (چهار حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۳	گروه مشتریان سطح ۳ (دو حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۴	شهر (هشت حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۵	وضعیت مشتری (چهار حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۶	منطقه (هفت حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۷	زمان سپری‌شده از آخرین خرید ۱۳۹۸	عددی[3]	نسبی[4]	متغیر پایهٔ اولیه	مستقل
۸	ماه خرید (دوازده حالت)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۹	خرید در روز چندم ماه	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۰	گروه بازار محصولات (۲۶ قلم کالا)	رشته‌ای	اسمی	متغیر پایهٔ اولیه	مستقل
۱۱	فاکتور خرید (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۲	خالص خریدر (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۳	تعداد خط فاکتور خرید	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۴	متوسط خرید هر فاکتور (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۵	کل تخفیفات (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۶	زمان سپری‌شده از آخرین خرید	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۷	تعداد فروش	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۸	مقدار برگشتی (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۱۹	تخفیفات کالایی (ریال)	عددی	نسبی	متغیر پایهٔ اولیه	مستقل
۲۰	مقدار تازگی خرید موزون	عددی	نسبی	متغیر ثانویه	مستقل
۲۱	مقدار دفعات خرید موزون	عددی	نسبی	متغیر ثانویه	مستقل

1. string
2. nominal
3. numeric
4. scale

Archive of SID.ir

Archive of SID.ir

ردیف	نام متغیر	نوع متغیر	مقیاس اندازه‌گیری	وضعیت	نقش
۲۲	مقدار ارزش پولی خرید موزون	عددی	نسبی	متغیر ثانویه	مستقل
۲۳	مقدار تازگی، تکرار و ارزش پولی موزون WRFM	عددی	نسبی	متغیر ثانویه	وابسته (هدف)

در این پژوهش ۱۹ متغیر اولیه وجود داشت که بر اساس این ۱۹ متغیر و متغیرهای ثانویه که بعداً بر اساس مؤلفه‌های مدل WRFM به پژوهش اضافه شده، فرایند داده‌کاوی انجام شد و از این بین، از زمان سپری شده از آخرین خرید مشتری برای محاسبهٔ مقدار تازگی خرید[1] (R)، از تعداد خط فاکتور فروش برای مقدار دفعات خرید[2] (F) و همچنین، از متوسط خرید هر فاکتور برای محاسبهٔ مقدار ارزش پولی[3] (M) استفاده شده است. مقدار نهایی تازگی، تکرار و ارزش پولی موزون برای هر داده که همان متغیر ردیف ۲۳ جدول ۳، است، بر اساس فرمول ۱، به‌دست می‌آید.

$$WRFMv = WR \times Recencyv + WF \times Frequencyv + WM \times Monetaryv$$ فرمول ۱)

$W_{R, F, M}$ وزن‌های هر یک از شاخص مدل تازگی، تکرار و ارزش پولی خرید و V=Value مقدار هر یک از شاخص‌هاست. مقدار وزن شاخص‌های مدل تازگی، تکرار و ارزش پولی خرید مشتری به‌ترتیب، ۰/۱۵، ۰/۲۹ و ۰/۵۶ به‌دست آمده است. پژوهشگران برای این‌که مطمئن شوند می‌توانند از این اوزان استفاده کنند، مقدار شاخص ناسازگاری را به کمک روش فرایند تحلیل سلسله‌مراتبی[4] محاسبه نموده و این مقدار در حدود ۰/۰۵۲ به‌دست آمده است که چون کمتر از ۰/۱ است، پس می‌توان با اطمینان از این وزن‌ها برای این پژوهش استفاده کرد.

٤-١. نامگذاری مشتریان

نامگذاری (کدینگ) دسته‌بندی‌های مختلف مشتریان بدین‌صورت انجام می‌گیرد که اگر مقدار مدل تازگی، تکرار و ارزش پولی موزون که از فرمول ۱ به‌دست می‌آید کمتر از ۱ باشد، مشتریان بحرانی اگر بین ۱ تا ۲ باشد مشتری پُرریسک، اگر بین ۲ تا ۳ باشد مشتری معمولی، اگر بین ۳ تا ۴ باشد مشتری نقره‌ای، و بالاتر از ۴ مشتری طلایی خواهد بود که از ماحصل این قسمت ردیف (ویژگی) جدیدی تحت عنوان نوع (گروه) مشتریان

1. recency 2. frequency 3. monetary
4. analytical hierarchy process (AHP)

Archive of SID.ir

Archive of SID.ir

به جدول ۳، اضافه خواهد شد و بر این اساس خروجی وضعیت مشتریان شرکت قبل از حذف داده‌های مغایر به‌شرح جدول ۴، به‌دست آمده است.

جدول ٤. یافته‌های نوع مشتریان به‌تفکیک تعداد و درصد فراوانی در ابتدای عملیات داده‌کاوی پژوهش

نوع مشتریان	طلایی	نقره‌ای	معمولی	پریسک	بحرانی	جمع
تعداد	۱۴۲۴۳	۱۶۳۲۸	۱۹۶۵۴	۱۲۱۹۶	۱۷۸۹	۶۴۸۵۸
درصد فراوانی	۲۲/۱۸	۲۵/۴۳	۳۰/۶۱	۱۸/۹۹	۲/۷۹	۱۰۰

به‌منظور انجام خوشه‌بندی می‌بایست پیش‌پردازشی بر روی داده‌ها صورت گیرد تا کیفیت و دقت خوشه‌بندی بهتر شود. بنابراین، این عملیات در این پژوهش توسط گروه‌بندی غالب یا روش تشخیص مغایرت انجام شده است تا داده‌های دورافتاده را مشخص کرده و از کل داده‌ها حذف کرد (چرا که روش داده‌کاوی k- میانگین به این نوع داده‌ها حساس است)، سپس از روش‌های خوشه‌بندی در ادامه استفاده می‌شود. در جدول ۵، نتایج این روش آمده است.

جدول ۵. یافته‌های نتایج گروه‌بندی تشخیص مغایرت

نام روش	تعداد و گروه‌های غالب[1]		تعداد داده‌های هر گروه به‌تفکیک	تعداد داده‌های مغایر	نوع مشتریان در هر گروه	مقدار شاخص‌های مدل تازگی، تکرار و ارزش پولی موزون		انحراف استاندارد
تشخیص مغایرت	۳	گروه غالب یک	۱۳۵۰۷	۲۴۷	طلایی، معمولی، نقره‌ای بحرانی، پرریسک	تازگی موزون	۰/۱۷۹	۰/۰۸۴
						تکرار موزون	۰/۹۸۸	۰/۴۴
						ارزش پولی موزون	۱/۸۶۴	۰/۸۳۸
		گروه غالب دو	۲۴۴۵۵	۲۲۹	نقره‌ای و طلایی	تازگی موزون	۰/۵۶	۰/۱۶۶
						فراوانی موزون	۱/۱۳	۰/۲۷۴
						ارزش پولی موزون	۲/۲۸۹	۰/۴۳۱
		گروه غالب سه	۲۶۸۹۶	۱۷۲	معمولی، پرریسک و بحرانی	تازگی موزون	۰/۴۸۷	۰/۱۷۶
						تکرار موزون	۰/۶۳۴	۰/۲۶۸
						ارزش پولی موزون	۱/۰۱۷	۰/۴۳۱
جمع کل			۶۴۸۵۸	۶۴۸				

1. peer-groups

Archive of SID.ir

Archive of SID.ir

با توجه به نتایج به‌دست‌آمده از روش گروه‌بندی تشخیص مغایرت جدول ۵، در این روش از ۲۲ متغیر از مجموع ۲۳ متغیر اولیه و ثانویهٔ پژوهش به‌عنوان ورودی، و از متغیر مقدار کل تازگی، تکرار و ارزش پولی موزون به‌عنوان متغیر خروجی برای عملیات تشخیص مغایرت که یک روش داده‌کاوی غیرنظارتی است، استفاده شده و سرانجام، با توجه به مقدار برش (حد آستانه) شاخص مغایرت (۲/۴۶۴۵۸) که توسط خود سیستم بر اساس درصد بیشترین رکوردهای مغایر در داده‌های یادگیری[1] (که در اینجا یک درصد است)، به‌عنوان پیش‌فرض انتخاب‌شده محاسبه شده و سرانجام، سه گروه غالب بر اساس این شاخص و حد آستانه تعیین شده است.

در جدول ۵، گروه غالب سه دارای بیشترین فراوانی با ۲۶۸۹۶ رکورد است. در این گروه مشتریان معمولی، پرریسک و بحرانی به‌ترتیب، دارای بیشترین فراوانی بودند. همچنین، کمترین فراوانی مربوط به گروه غالب یک، با ۱۳۰۵۷رکورد است و بنابراین، در این گروه هر پنج نوع مشتری به‌ترتیب فراوانی مشتری طلایی، معمولی، نقره‌ای، بحرانی و پرریسک قرار گرفته‌اند. با توجه به یافته‌های این بخش در کل، ۶۴۸ رکورد یا داده مغایر از کل ۶۴۸۵۸ رکورد اولیه تشخیص داده شد و این داده‌ها از کل داده‌های مورد پژوهش حذف شد تا فرایند خوشه‌بندی با دقت بالاتری انجام شود. با توجه به مقادیر مختلف شاخص‌های مدل WRFM و مقادیر انحراف استاندارد مربوط به هر شاخص درگروه‌های غالب، مقدار انحراف استاندارد شاخص ارزش پولی موزون (WM) نسبت به سایر مقادیر، انحراف استاندارد بیشتری داشته است و این می‌تواند به معنای تغییرپذیری زیاد مقادیر این شاخص باشد. در ادامه، از ۷۰ درصد (۶۴۲۱۰) دادهٔ باقی‌مانده، برای یادگیری و از بقیهٔ آن برای آزمون دقت مدل‌های خوشه‌بندی پژوهش استفاده شده است.

٤-۲. یافته‌های خوشه‌بندی شبکهٔ عصبی کوهنن[2]

از آنجا که عملکرد این روش در بعضی از پژوهش‌ها مانند پژوهش‌های Hanafizadeh & Mirzazadeh (2011); Chang & Tsay (2004); Blattberg et al. (2008); Wei et al. (2012) نسبت به سایر روش‌های متداول مانند k- میانگین بهتر بوده است، در این پژوهش در کنار روش‌های دیگر خوشه‌بندی مورد بررسی قرار می‌گیرد تا اگر نتایج این

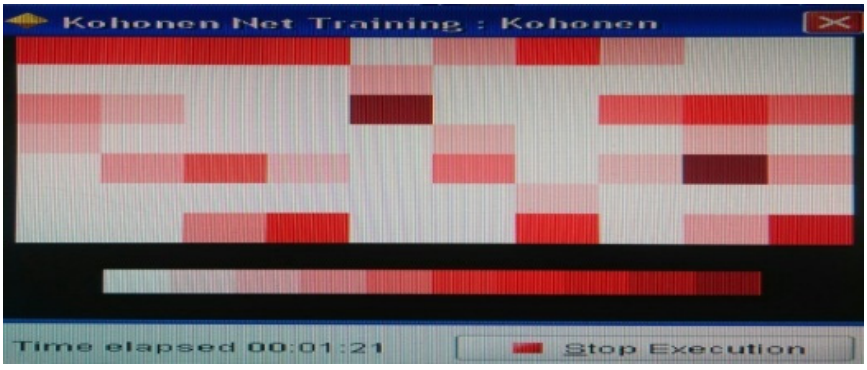
1. training data
2. Kohonen neural net clustering

Archive of SID.ir

Archive of SID.ir

روش خوشه‌بندی بهتر باشد، بتوان از آن به‌عنوان روش اصلی برای خوشه‌بندی مشتریان استفاده کرد (خوشه‌بندی «کوهنن» مراحلی دارد که خوشه‌بندی‌ها را به اشکالی مصور در ابعاد ماتریس دو-بعدی تبدیل می‌کند و برای پژوهشگران بسیار مفید و قابل درک است).

ماتریس شبکهٔ عصبی «کوهنن» (نگاشت‌های خودسازمانده)[1] این پژوهش از ۷۰ سلول عصبی به‌عنوان لایهٔ ورودی[2] و ۶۳ سلول عصبی[3] به‌عنوان لایهٔ خروجی[4] برای ایجاد خوشه‌بندی‌ها استفاده کرده است. خروجی این روش خوشه‌بندی یک ماتریس سلولی ۷×۱۰ مطابق با شکل ۶، است. این شکل بر اساس زمان‌بندی یک دقیقه و بیست‌ویک ثانیه توسط نرم‌افزار «اس‌پی‌اس‌اس مودلر» ایجاد شده است، ولی نتایج بیان‌شده بر اساس خروجی نهایی این روش است (امکان گرفتن خروجی و ثبت لحظه‌ای تصاویر ماتریس «کوهنن» توسط نرم‌افزار مقدور نبوده و به‌طور پیوسته فضای رنگی و خالی آن در حال تغییر بوده است. بنابراین، پژوهشگران با عکس‌برداری و ثبت لحظه‌ای این ماتریس را ثبت کرده‌اند). ماتریس «کوهنن» ایجادشده شامل ۷۰ سلول است که ۱۴ سلول آن خالی از داده بوده و تنها ۶ سلول آن حاوی رکوردهایی بیشتر از ۲۰۰۰ داده است و بقیهٔ آن‌ها حتی فراوانی کمتر از ۱۰۰ نیز داشته‌اند. در ادامه، تنها به این ۶ سلول که حاوی بیشترین رکورد فراوانی است، اشاره می‌شود (رنگ‌های تیره‌تر نشان از مراکز خوشه‌ای قوی‌تر و با فراوانی دادهٔ بیشتر است و رنگ‌های روشن‌تر نشان از فراوانی کمتر و مراکز خوشه‌ای ضعیف‌تر دارد).



شکل ٦. نتایج خوشه‌بندی به روش «کوهنن»

1 .self-organizing maps (SOM) 2. input layer 3. neurons
4. output layer

Archive of SID.ir

Archive of SID.ir

فراوانی سلول‌های ماتریس نهایی به‌دست‌آمده با بیشترین تعداد داده فراوانی به‌ترتیب، عبارت‌اند از: ۱. سلول (X_9,Y_6) با ۲۶۵۶ داده، ۲. سلول (X_0,Y_2) با ۲۴۵۸ داده، ۳. سلول (X_9,Y_3) با ۲۲۰۳ داده، ۴. سلول (X_0,Y_4) با ۲۰۸۱ داده، ۵. سلول (X_7,Y_6) با ۲۰۷۷ داده، و ۶. سلول (X_8,Y_0) با ۲۰۶۵ داده.

٤-۳. یافته‌های خوشه‌بندی K- میانگین

از این روش در بیشتر پژوهش‌های انجام‌شده مانند Alvandi, Fazli & Abdoli (2012)؛ «کفاش‌پور، توکلی و علیزاده» (۱۳۹۱)؛ «ایزدی» و همکاران (۱۳۹۵)؛ Bashiri Mosavi & Afsar (2018); Sheikh, Ghanbarpour & Gholamiangonabadi (2019); Bin, Peiji & Dan (2008); Wu, Chang & Lo (2009); Chiu et al. (2009); Aryuni & Miranda (2018); Jintana & Mori (2019) به‌علت کارایی بالای آن برای خوشه‌بندی استفاده شده است. بنابراین، به‌علت فراوانی استفاده و کارایی بالای آن در پژوهش حاضر نیز از این روش استفاده می‌شود تا چنانچه نتایج آن از دقت بالاتری نسبت به دیگر روش‌های مورد استفاده در این پژوهش برخوردار باشد، از آن به‌عنوان روش بهتر و مناسب‌تر برای ادامهٔ تحلیل خوشه‌ای و داده‌کاوی پژوهش استفاده شود. در این قسمت به خوشه‌بندی‌هایی که با روش K- میانگین با فرض ۵، ۶، و ۷ خوشه ایجاد شده، اشاره می‌شود (در روش خوشه‌بندی K- میانگین لازم است پژوهشگر یا کاربر ابتدا مقدار اولیه‌ای را برای تعداد دسته‌های خوشه‌بندی تعیین کند تا چنانچه هر کدام از این تعداد خوشه‌ها نتایج بهتری داشت، انتخاب شود). برای این روش تعداد ۲۰ تکرار[1] به‌عنوان حداکثر تکرار عمل خوشه‌بندی تا رسیدن به نتیجهٔ مربوط انتخاب شده است. چنانچه تعداد خوشه‌های متفاوتی انتخاب شود (با فرض حداکثر ۲۰ تکرار)، درصد خطاهای متفاوتی متناسب با هر تکرار فرایند خوشه‌بندی K- میانگین ایجاد می‌شود که شرح آن در جدول ۷، آمده است.

1. iterations

Archive of SID.ir

Archive of SID.ir

جدول ۷. یافته‌های خوشه‌بندی‌های متفاوت به روش K- میانگین

روش	تعداد و شمارهٔ خوشه‌ها		فراوانی	نوع مشتریان به‌تفکیک فراوانی				تعداد تکرار خوشه‌بندی	درصد خطا
				طلایی	نقره‌ای	معمولی	پریسک	بحرانی		
K- میانگین میانگین	۵	۱	۶۴۰۲	۶۴۰۲	۰	۰	۰	۰	۱۹	۰/۰۰
		۲	۷۴۹۹	۰	۰	۰	۷۴۹۹	۰		
		۳	۸۵۲۹	۳۶۰۶	۴۹۰۴	۱۹	۰	۰		
		۴	۱۷۰۸۵	۰	۶۵۸۶	۱۰۴۹۹	۰	۰		
		۵	۵۵۲۵	۰	۱۰	۳۲۲۷	۱۰۲۵	۱۲۶۳		
	جمع کل		۴۵۰۴۰	۱۰۰۰۸	۱۱۵۰۰	۱۳۷۴۵	۸۵۲۴	۱۲۶۳		
	۶	۱	۶۳۲۳	۶۳۲۳	۰	۰	۰	۰	۱۳	۰/۰۰
		۲	۱۰۸۱۹	۰	۳	۶۰۳۱	۳۵۳۷	۱۲۴۸		
		۳	۵۰۲۶	۳۴۲۹	۱۵۹۷	۰	۰	۰		
		۴	۱۲۶۱۸	۰	۰	۷۶۶۳	۴۹۵۵	۰		
		۵	۴۲۵	۲۵۶	۷۱	۵۱	۳۲	۱۵		
		۶	۹۸۲۹	۰	۹۸۲۹	۰	۰	۰		
	جمع کل		۴۵۰۴۰	۱۰۰۰۸	۱۱۵۰۰	۱۳۷۴۵	۸۵۲۴	۱۲۶۳		
	۷	۱	۶۳۰۹	۶۳۰۹	۰	۰	۰	۰	۲۰	۰/۰۰
		۲	۵۴۱۱	۰	۱۰۷۷	۲۴۵۶	۸۹۳	۹۸۵		
		۳	۳۴۹۷	۳۴۴۳	۵۴	۰	۰	۰		
		۴	۱۲۰۱۵	۰	۴۳۴۹	۷۶۶۶	۰	۰		
		۵	۳۷۹	۲۵۶	۷۱	۴۸	۴	۰		
		۶	۹۵۲۴	۰	۵۹۴۹	۳۵۷۵	۰	۰		
		۷	۷۹۰۵	۰	۰	۰	۷۶۲۷	۲۷۸		
	جمع کل		۴۵۰۴۰	۱۰۰۰۸	۱۱۵۰۰	۱۳۷۴۵	۸۵۲۴	۱۲۶۳		

با توجه به جدول ۷، در خوشه‌بندی‌های مختلف K- میانگین اکثر مشتریان طلایی در خوشهٔ اول قرار گرفته است و همچنین، در تمامی خوشه‌بندی‌ها، به‌جز خوشه‌بندی ۵تایی، در خوشهٔ دوم، سه گروه مشتری معمولی، پرریسک و بحرانی حتماً حضور دارند. از بین خوشه‌بندی‌های ۵، ۶، و ۷تایی K- میانگین، خوشه‌بندی شش‌تایی با تکرار کمتری

Archive of SID.ir

Archive of SID.ir

به خطای صفر می‌رسد که می‌تواند به‌عنوان گزینهٔ مناسب از نظر کیفیت خوشه‌بندی مورد بررسی قرار گیرد.

٤-٤. یافته‌های خوشه‌بندی دو-مرحله‌ای

از آنجا که این روش در بعضی از پژوهش‌ها مانند «خواجوند، تقوی‌فرد و نجفی» (۱۳۹۱) و Babaiyan & Sarfarazi (2019)، نسبت به دیگر روش‌های خوشه‌بندی عملکرد بهتری داشته، بنابراین، در پژوهش حاضر نیز سعی شده است این روش درکنار سایر روش‌ها مورد بررسی و تحلیل قرار گیرد تا اگر نسبت به دیگر روش‌های خوشه‌بندی نتایج و دقت بالاتری داشته باشد، بتوان از آن استفاده کرد. این خوشه‌بندی روشی غیرنظارتی است؛ بدین‌صورت که کاربر و یا پژوهشگر فقط لازم است محدوده‌ای را برای تعداد اولیهٔ خوشه‌ها تعیین کند. بقیهٔ عملیات و تعیین تعداد خوشهٔ بهینه توسط محاسبات درونی خود الگوریتم مشخص می‌شود که در این پژوهش حداقل ۲ و حداکثر ۱۵ خوشه به‌عنوان تنظیم‌های اولیه تعیین شده است.

جدول ۸. یافته‌های خوشه‌بندی دو-مرحله‌ای با احتساب داده‌های دورافتاده و بدون آن

روش	تعداد و شمارهٔ خوشه		فراوانی خوشه	نوع مشتریان به‌تفکیک					مقدار شاخص‌های مدل تازگی، تکرار و ارزش پولی موزون		
				طلایی	نقره‌ای	معمولی	پریسک	بحرانی	تازگی موزون	تکرار موزون	ارزش پولی موزون
دو-مرحله‌ای	۶	۱	۸۰۱۰	۸۰۰۲	۸	۰	۰	۰	۰/۴۶۵	۱/۳۶۰	۲/۷۱۵
		۲	۶۸۸۸	۱۹۸۴	۲۱۵۴	۲۰۸۹	۶۵۴	۷	۰/۵۰۶	۱/۰۲۳	۱/۸۳۴
		۳	۵۸۷۳	۰	۰	۴۵۳۱	۱۳۱۳	۲۹	۰/۴۵۷	۰/۷۱۴	۱/۱۸۴
		۴	۶۸۴۹	۰	۰	۴۸۸۴	۱۹۵۹	۶	۰/۴۶۶	۰/۷۲	۱/۱۱۹
		۵	۴۴۱۲	۰	۱۰۸۷	۲۲۴۱	۸۵۲	۲۳۲	۰/۱۵۲	۰/۸۴۱	۱/۵۱۲
		۶	۸۲۷۳	۲۲	۸۲۵۱	۰	۰	۰	۰/۵۲۷	۰/۹۶۶	۲/۰۳۳
	جمع کل		۴۰۳۰۵	۱۰۰۰۸	۱۱۵۰۰	۱۳۷۴۵	۴۷۷۸	۲۷۴			
	۲[1]	۱	۳۰۰۵۹	۹	۱۱۳۱۲	۱۳۶۹۲	۴۷۷۳	۲۷۳	۰/۴۳۹	۰/۸۲۹	۱/۴۹۶
		۲	۱۰۲۴۶	۹۹۹۹	۱۸۸	۵۳	۵	۱	۰/۴۸	۱/۳۵۲	۲/۶۸۵
	جمع کل		۴۰۳۰۵	۱۰۰۰۸	۱۱۵۰۰	۱۳۷۴۵	۴۷۷۸	۲۷۴			

1. exclude outliers

Archive of SID.ir

Archive of SID.ir

همان‌طور که از نتایج یافته‌های خوشه‌بندی دو-مرحله‌ای جدول ۸، مشخص است، وقتی از خوشه‌بندی دو-مرحله‌ای بدون در نظر گرفتن داده‌های دورافتاده استفاده شود، تنها دو خوشه ایجاد می‌شود، ولی در حالت عادی، ۶ خوشه ایجاد شده است. در حالت ۶خوشه‌ای، در خوشهٔ اول بیشتر مشتریان از نوع مشتریان طلایی بودند، ولی در خوشهٔ ششم بیشتر مشتریان از نوع نقره‌ای بودند. همچنین، در خوشه‌های دیگر ترکیب‌های مختلفی از مشتریان حضور داشتند. در حالت دوخوشه‌ای، در خوشهٔ دوم بیشتر مشتریان از نوع مشتریان طلایی بودند و مقدار شاخص‌های تازگی، تکرار و ارزش پولی موزون حکایت از این موضوع دارد. با مقایسهٔ نتایج دو روش مختلف خوشه‌بندی دو-مرحله‌ای (با داده‌های دورافتاده و بدون آن)، به نظر می‌رسد که خوشه‌بندی شش‌تایی کیفیت بالاتری داشته است (در خوشه‌بندی، خوشه‌هایی که دارای فراوانی از یک یا نهایت دو رکورد (متغیر) متفاوت با درصد فراوانی بالا باشند، کیفیت بالاتری خواهند داشت). از بین ۶۴۲۱۰ داده و بر اساس ۷۰ درصد داده‌های آموزشی انتظار می‌رفت که تعداد ۴۵۰۴۰ داده در بخش جمع کل داده‌های روش خوشه‌بندی دو-مرحله‌ای مشاهده می‌شد، ولی در حال حاضر، این مقدار ۴۰۳۰۵ است. علت این اختلاف، وجود داده‌های خالی است که در نتیجهٔ محاسبات درونی این روش ایجاد شده است. تعداد کل داده‌های با مقادیر خالی در این روش (مجموع داده‌های خالی بخش یادگیری و آزمون) برابر ۶۶۶۸ بود و این مورد را می‌توان در جدول ۹، مشاهده نمود.

جدول ۹. یافته‌های مقایسهٔ فراوانی خوشه‌ها در دو روش خوشه‌بندی بر اساس داده‌های یادگیری و آزمون

روش خوشه‌بندی	داده‌های پژوهش	خوشهٔ اول	خوشهٔ دوم	خوشهٔ سوم	خوشهٔ چهارم	خوشهٔ پنجم	خوشهٔ ششم	خوشهٔ خالی[1]	جمع کل داده‌ها	درصد داده‌ها
k- میانگین	یادگیری	۶۳۲۳	۱۰۸۱۹	۵۰۲۶	۱۲۶۱۸	۴۲۵	۹۸۲۹	۰	۴۵۰۴۰	۷۰/۱۴
	آزمون	۲۶۶۱	۴۵۵۱	۲۰۸۸	۵۵۰۰	۱۹۶	۴۱۷۴	۰	۱۹۱۷۰	۲۰/۸۶
جمع کل		۸۹۸۴	۱۵۳۷۰	۷۱۱۴	۱۸۱۱۸	۶۲۱	۱۴۰۰۳	۰	۶۴۲۱۰	۱۰۰
دو-مرحله‌ای	یادگیری	۸۰۱۰	۶۸۸۸	۵۸۷۳	۶۸۴۹	۴۴۱۲	۸۲۷۳	۴۷۳۵	۴۵۰۴۰	۷۰/۱۴
	آزمون	۳۳۸۱	۲۹۷۵	۲۵۹۹	۲۹۱۲	۱۸۳۴	۳۵۴۶	۱۹۳۳	۱۹۱۷۰	۲۰/۸۶
جمع کل		۱۱۳۹۱	۹۸۶۳	۸۴۷۲	۹۷۶۱	۶۲۳۷	۱۱۸۱۹	۶۶۶۸	۶۴۲۱۰	۱۰۰

1. null

Archive of SID.ir

Archive of SID.ir

در این پژوهش از بین شاخص‌های مختلف مانند درصد خلوص[1] و پراکندگی، امتیاز جینی[2] (پراکندگی جمعیت)، آنتروپی[3] (بهرهٔ اطلاعاتی)، نسبت بهرهٔ اطلاعاتی[4]، آزمون مربع کای (χ^2) (به‌طور معمول، برای تعیین کیفیت و میزان خلوص خوشه‌بندی و دسته‌بندی داده‌ها استفاده می‌شود)، تنها از دو روش درصد خلوص و شاخص اطلاعات متقابل نرمال‌شده استفاده شده است که با توجه به فرمول ۲ و فرمول‌های ۳ و ۴ این شاخص‌ها به‌دست می‌آیند. مقدار خالص بودن بین یک (خوشه‌بندی بهینه) و صفر (خوشه‌بندی بد) قرار دارد (ویسی ۱۳۹۶). درصد خلوص را می‌توان از فرمول ۲ به‌دست آورد. در این فرمول، K نشان‌دهندهٔ تعداد خوشه‌ها و j نشان‌دهندهٔ تعداد دسته‌هاست و N نشان‌دهندهٔ تعداد کل داده‌هاست.

فرمول ۲)
$$Purity(Cluster, Class) = \frac{1}{N\sum_{j=1}^{j} Max_j[\text{Cluster}k \cap \text{Class j}]}$$

معیار خارجی اطلاعات متقابل نرمال نشان می‌دهد که آیا اطلاعات مفید و غیرتصادفی از خوشه‌بندی به‌دست آمده است یا خیر (عدد این شاخص بین صفر و یک است). مقدار یک، نشان‌دهندهٔ آن است که هر خوشه دقیقاً بیان‌کنندهٔ یک دسته است؛ مقدار صفر (خوشه‌بندی تصادفی) یعنی دانستن خوشه به افزایش اطلاعات ما از دسته کمکی نمی‌کند؛ شاخص اطلاعات متقابل نرمال‌شده به تعداد خوشه‌ها حساس است که از فرمول ۳ و ۴ به‌دست می‌آید (ویسی ۱۳۹۶). در فرمول ۳، NMI، شاخص اطلاعات متقابل نرمال شده است و Pkj احتمال (فراوانی نسبی) دستهٔ j در خوشهٔ K است. K نشان‌دهندهٔ تعداد خوشه‌هاست و I مقدار اطلاعات متقابل قبل از نرمال کردن است.

فرمول ۳)
$$\text{NMI}(\text{Cluster}, \text{Class}) = \frac{\text{I}(\text{Cluster},\text{Class})}{\frac{[\text{H}(\text{Cluster})+\text{H}(\text{Class})]}{2}}$$

فرمول ۴) (
$$\text{I}(\text{Cluster}, \text{Class}) = \sum_{k=1}^{k}\sum_{j=1}^{j} \text{P}\left(\text{Cluster}_k \cap \text{class}_j\right) * \log\frac{\text{P}\left(\text{Cluster}_k \cap \text{class}_j\right)}{\text{P}(\text{cluster}_k) * \text{P}\left(\text{class}_j\right)}$$

1. purity
2. gini score
3. entropy
4. information gain ratio

Archive of SID.ir

Archive of SID.ir

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

جدول ۱۰. یافته‌های مقایسهٔ کیفیت خوشه‌بندی‌های k- میانگین و دو-مرحله‌ای

روش خوشه‌بندی	تعداد خوشه یا گروه‌ها	تعداد تکرار تنظیم اولیه	میزان خطا[1]	درصد خلوص	شاخص اطلاعات متقابل نرمال‌شده
دو-مرحله‌ای	۶	بدون داده‌های دورافتاده		۰/۷۴۵۸	۰/۵۲۱
دو-مرحله‌ای	۲	با تنظیم‌های اولیه		۰/۵۸۷	۰/۱۸۳
K- میانگین	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

با توجه به یافته‌های مقایسهٔ دو روش مختلف خوشه‌بندی K- میانگین و دو-مرحله‌ای مطابق با جدول ۱۰، از روش خوشه‌بندی دو-مرحله‌ای، خوشه‌بندی با شش خوشه انتخاب می‌شود؛ چرا که درصد خلوص بالاتری داشته و شاخص اطلاعات متقابل نرمال‌شدهٔ آن ۰/۵۲۱ است (مقدار خلوص بین صفر و یک است. خوشهٔ خالص دارای مقدار یک است و شاخص اطلاعات متقابل نرمال‌شده نیز عددی بین صفر و یک است که هرچه این شاخص به یک نزدیک‌تر باشد، نشان از کیفیت بهتر خوشه‌بندی و ایجاد اطلاعات باارزش‌تری است).

از بین خوشه‌بندی‌های K- میانگین، خوشه‌بندی‌های پنج-تایی، شش-تایی و هفت-تایی (از نظر شاخص تعداد تکرار خوشه‌بندی تا رسیدن به خطای صفر، میزان خطای خوشه‌بندی، درصد خلوص و شاخص اطلاعات متقابل نرمال‌شده) وضعیت بهتری داشتند. با مقایسهٔ این چهار شاخص در این سه دسته خوشه‌بندی، خوشه‌بندی شش-تایی در میزان تکرار کمتر و درصد خلوص نسبت

1. error

Archive of SID.ir

Archive of SID.ir

به خوشه‌بندی پنج‌تایی وضعیت بهتری داشته است، ولی خوشه‌بندی پنج-تایی تنها در شاخص اطلاعات متقابل نرمال‌شده وضعیت بهتری داشته است. در مقایسه بین خوشه‌بندی شش-تایی با هفت-تایی، خوشه‌بندی شش-تایی در دو شاخص تعداد تکرار کمتر تا رسیدن به درصد خطای صفر و شاخص اطلاعات متقابل نرمال‌شده تفاوت چشمگیری دارد، ولی خوشه‌بندی هفت-تایی تنها در درصد خلوص مقدار اندکی وضعیت بهتری داشته است. بین خوشه‌بندی هفت-تایی و پنج-تایی، خوشه‌بندی هفت-تایی تنها در شاخص درصد خلوص وضعیت بهتری دارد؛ در حالی که خوشه‌بندی پنج-تایی در دو شاخص تعداد تکرارکمتر برای خوشه‌بندی تا رسیدن به خطای صفر و شاخص اطلاعات متقابل نرمال‌شده وضعیت بهتری داشته است که سرانجام، از بین این سه خوشه‌بندی، خوشه‌بندی شش-تایی نسبت به بقیه خوشه‌بندی بهتر و باکیفیت‌تری ایجاد کرده است. حال، از بین دو روش مختلف خوشه‌بندی K- میانگین و خوشه‌بندی دو-مرحله‌ای روش K-میانگین با تعداد ۶ خوشه به‌عنوان روش مناسب انتخاب می‌شود؛ چرا که در شاخص اطلاعات متقابل نرمال‌شده تفاوت چشمگیری با هم داشته، ولی از نظر شاخص درصد خلوص تفاوت چندانی با هم ندارد (تقریباً برابر بوده‌اند). در جدول ۱۱، نتایج مقادیر مدل تازگی، تکرار و ارزش پولی موزون و پراکندگی آن در دو روش خوشه‌بندی k- میانگین و دو-مرحله‌ای آورده شده است (داده‌های این جدول بر اساس پارتیشن‌بندی ۷۰ درصد داده‌های آموزشی و ۳۰ درصد داده‌های آزمون تهیه شده است).

جدول ۱۱. مقایسهٔ یافته‌های مقادیر WRFM و پراکندگی آن در دو روش متفاوت خوشه‌بندی

روش	شمارهٔ خوشه	فراوانی خوشه	میانگین WRFM	حداقل	حداکثر	واریانس	انحراف استاندارد[1]	خطای استاندارد از میانگین	ضریب واریانس[2] (درصد)
k- میانگین	۱	۶۳۲۳	۴/۶۱	۴/۱۲	۵	۰/۰۹۲	۰/۳۰۳	۰/۰۰۴	۶/۵۷
	۲	۱۰۸۱۹	۲/۰۲۳	۱	۳/۱۲	۰/۴۴۸	۰/۶۶۹	۰/۰۰۶	۳۳/۰۷
	۳	۵۰۲۶	۴/۱۴۸	۳/۱۴	۴/۵۵	۰/۱۸۳	۰/۴۲۷	۰/۰۰۶	۱۰/۲۹
	۴	۱۲۶۱۸	۲/۱۶۸	۱/۳	۳	۰/۳۱۲	۰/۵۵۹	۰/۰۰۵	۲۵/۷۸
	۵	۴۲۵	۳/۷۷۴	۱	۵	۱/۲۹۵	۱/۱۳۸	۰/۰۵۵	۳۰/۱۵
	۶	۹۸۲۹	۳/۵۲۶	۳/۰۱	۴	۰/۱	۰/۳۱۷	۰/۰۰۳	۸/۹۹
جمع کل		۴۵۰۴۰							

1. standard deviation (SD)
2. coefficient of variance (CV)

Archive of SID.ir

Archive of SID.ir

روش	شمارهٔ خوشه	فراوانی خوشه	میانگین WRFM	حداقل	حداکثر	واریانس	انحراف استاندارد	خطای استاندارد از میانگین	ضریب واریانس (درصد)
دو-مرحله‌ای	۱	۸۰۱۰	۴/۵۳۸	۳/۵۵	۵	۰/۰۷۲	۰/۲۶۸	۰/۰۰۳	۵/۹۱
	۲	۶۸۸۸	۳/۳۶۴	۱	۵	۰/۹۸۱	۰/۹۹	۰/۰۱۲	۲۹/۴۳
	۳	۵۸۷۳	۲/۳۵۶	۱	۳	۰/۲۲۲	۰/۴۷۱	۰/۰۰۶	۱۹/۹۹
	۴	۶۸۴۹	۲/۳۰۵	۱	۳	۰/۲۴۴	۰/۴۹۴	۰/۰۰۶	۲۱/۴۳
	۵	۴۴۱۲	۲/۵۰۴	۱	۳/۹۹	۰/۶۲۷	۰/۷۹۲	۰/۰۱۲	۳۱/۶۳
	۶	۸۲۷۳	۳/۵۲۷	۳/۰۱	۴/۷۱	۰/۰۹۸	۰/۳۱۴	۰/۰۰۳	۸/۹
جمع		۴۰۳۰۵							
جمع کل داده‌ها		۶۴۲۱۰	۳/۰۰۸	۱	۵	۱/۲۶	۱/۱۲۳	۰/۰۰۴	۳۷/۳۳

با توجه به یافته‌های شاخص‌های پراکندگی در دو روش خوشه‌بندی دو-مرحله‌ای و K- میانگین مطابق جدول ۱۱، که بر اساس بهترین خوشه‌های ایجادشده از دو روش متفاوت خوشه‌بندی تهیه شده، آنچه مشهود است در هر دو روش، خوشهٔ اول بهترین خوشه بود؛ چرا که 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 این مقدار در فاصلهٔ ۴ تا ۵ قرار گرفته، بنابراین، می‌توان گفت که اکثر مشتریان این خوشه، مشتریان طلایی هستند. همچنین، از آنجا که ضریب واریانس این خوشه مقدار (۶/۵۷) درصد است، از پراکندگی کمتری نسبت به سایر خوشه‌بندی‌ها برخوردار بوده که می‌تواند نشان از یک دست بودن نوع مشتریان این خوشه باشد (ضریب واریانس کوچک‌تر پراکندگی کمتر).

در خوشه‌بندی ۶تایی به روش K- میانگین خوشهٔ اول و سوم، که در جمع ۱۱۳۴۹ داده بخش آموزشی این پژوهش را دربرمی‌گیرد، با فراوانی تجمعی ۲۵/۲ درصد و با توجه به این‌که مقدار تازگی، تکرار، و ارزش پولی موزون این دو خوشه بیشتر از میانگین کل بوده و در محدودهٔ ۴ تا 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 دو دسته خوشه، خوشه‌های باارزشی برای شرکت به حساب می‌آیند. همان‌طور که از جدول ۱۱، 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 و

Archive of SID.ir

Archive of SID.ir

این می‌تواند نشان از متفاوت و بااهمیت بودن این خوشه‌ها برای شرکت باشد.

در روش خوشه‌بندی دو-مرحله‌ای از آنجا که هنگام محاسبهٔ خوشه‌بندی‌ها تعدادی مقادیر خالی برای بعضی از داده‌های ثبت‌شده وجود داشته که برای خوشه‌بندی‌ها استفاده نشده و به‌عنوان خوشهٔ خالی[1] در نظر گرفته شده، برای همین است که داده‌های آموزشی دو روش خوشه‌بندی K- میانگین و دو-مرحله‌ای با هم متفاوت است. در خوشه‌بندی دو-مرحله‌ای خوشهٔ اول، دوم و ششم از متوسط میانگین کل مقدار تازگی، تکرار و ارزش پولی موزون بیشتر بوده‌اند که این به‌نوعی اهمیت نسبی این سه خوشه را برای شرکت تداعی می‌کند. از بین این سه خوشه، خوشهٔ اول با فراوانی ۸۰۱۰ و با مقدار تازگی، تکرار و ارزش پولی (۴/۵۳۸) و درصد فراوانی (۱۹/۸۷) بهترین خوشهٔ این روش بوده که از نظر ارزشی نیز برای شرکت بسیار مهم هستند؛ چرا که تقریباً طبق قانون ۲۰-۸۰، این مشتریان ۸۰ درصد درآمد شرکت را ایجاد کرده‌اند.

در این قسمت از شیوه‌ای که در پژوهش‌های Chang & Tsay (2004); Alvandi, Fazli & Abdoli (2012); و Babaiyan & Sarfarazi (2019) برای تجزیه و تحلیل ارزش مشتریان و رتبه‌بندی آن‌ها بر اساس شاخص‌های مدل RFM به‌کار رفته، کمک گرفته‌ایم؛ با این تفاوت که بُعد وفاداری که با مدل LRFM همخوانی دارد، در این پژوهش بررسی نشده و به‌جای استفاده از علامت الگوی (‹) برای نشان دادن مقادیر بالاتر از میانگین و علامت الگوی (Œ) برای نشان دادن مقادیر کمتر از میانگین، از حرف H برای نشان دادن الگوهایی که دارای مقادیر بالاتر از میانگین بوده و از حرف L برای نشان دادن الگوهایی که دارای مقادیر کمتر از میانگین بوده استفاده شده است. در هر خوشه با توجه به مقایسهٔ مقدار هر شاخص با مقدار میانگین کل آن شاخص، اگر شاخص بزرگ‌تر از میانگین کل باشد، از حرف H استفاده می‌شود و اگر این شاخص کمتر از میانگین باشد، از حرف L. به همین ترتیب، برای شاخص‌های دیگر این عملیات تکرار می‌شود. برای تعیین رتبهٔ هر شاخص در هر خوشه با توجه به خوشه‌های دیگر از اعداد ۱ تا ۶ برای این مقایسه استفاده می‌شود (۶ یعنی این خوشه در این شاخص دارای کمترین مقدار و رتبه و ۱ یعنی بالاترین مقدار و رتبه را در بین شش خوشه داشته است و بقیهٔ خوشه‌ها در بین این مقادیر قرار خواهند گرفت). این قانون در مورد شاخص تازگی خرید برعکس است؛ یعنی هر قدر این مقدار

1. null cluster

Archive of SID.ir

Archive of SID.ir

کوچک‌تر باشد، بهتر است (عدد کوچک‌تر در شاخص تازگی خرید موزون نشان از تازگی خرید مشتری دارد و عدد بزرگ‌تر نشان از آن دارد که مشتری مدت طولانی است که از شرکت خرید نداشته است). رتبهٔ (۶،۶،۱) یعنی این خوشه در دو شاخص تکرار و ارزش پولی خرید مدل RFM دارای بالاترین رتبه بوده (وضعیت خوبی ندارد)، ولی در شاخص تازگی خرید رتبهٔ اول را دارد. همچنین، بهترین دستهٔ خوشه و مشتری، آن مشتری است که دارای کمترین مقدار از نظر شاخص تازگی خرید بوده و در دو شاخص دیگر دارای مقادیر بالاتر از میانگین است. به‌منظور به‌دست آوردن رتبهٔ کلی هر خوشه با توجه به ضرایب اهمیت نسبی هر مؤلفهٔ مدل تازگی، تکرار و ارزش پولی و همچنین با مقایسهٔ زوجی رتبهٔ هر مؤلفه در بین خوشه‌ها سرانجام، رتبهٔ کلی هر خوشه به‌دست آمده است و با این تفاسیر خوشهٔ اول دارای رتبهٔ یک، خوشهٔ سوم دارای رتبهٔ دو، خوشهٔ پنجم دارای رتبهٔ سه، خوشهٔ ششم دارای رتبهٔ چهار، خوشهٔ دوم دارای رتبهٔ پنج و سرانجام، خوشهٔ چهارم دارای رتبهٔ شش است که نتایج این رتبه‌بندی وتعیین الگوها در جدول ۱۲، بیان شده است.

جدول ۱۲. نتایج رتبه‌بندی و تعیین الگوی خرید مشتریان به تفکیک هر خوشه

روش	شمارهٔ خوشه	میانگین تازگی خرید موزون	میانگین تکرار خرید موزون	میانگین ارزش پولی موزون	میانگین WRFM	رتبهٔ شاخص (R,F,M) بین خوشه‌ها	رتبهٔ کلی خوشه	الگوی خوشه‌ها (R,F,M)	فراوانی
k- میانگین	۱	۰/۶۲۷	۱/۳۳۵	۲/۶۵	۴/۶۱۲	(۱،۱،۶)	۱	(H,H,H)	۶۳۲۳
	۲	۰/۲۲۲	۰/۶۷۱	۱/۱۳	۲/۰۲۳	(۵،۵،۲)	۵	(L,L,L)	۱۰۸۱۹
	۳	۰/۲۰۱	۱/۳۲۶	۲/۶۲۶	۴/۱۵۳	(۲،۲،۱)	۲	(L,H,H)	۵۰۲۶
	۴	۰/۵۸۳	۰/۶۱۵	۰/۹۷۳	۲/۱۷۱	(۶،۶،۵)	۶	(H,L,L)	۱۲۶۱۸
	۵	۰/۳۸۲	۱/۱۴۷	۲/۱۹۹	۳/۷۲۸	(۳،۳،۳)	۳	(L,H,H)	۴۲۵
	۶	۰/۵۴۴	۰/۹۷۴	۲/۰۰۷	۳/۵۲۵	(۴،۴،۴)	۴	(H,H,H)	۹۸۲۹
جمع(داده‌های یادگیری)									۴۵۰۴۰
جمع کل داده‌ها		۰/۴۵	۰/۸۹۲	۱/۶۶۶	۳/۰۰۸				۶۴۲۱۰

به‌واسطهٔ وضعیت دوگانهٔ بزرگ‌تر و یا کوچک‌تر از میانگین بودن هر شاخص مدل *RFM* هشت حالت الگوی ترکیبی قابل تصور است. در این پژوهش طبق نتایج جدول ۱۲، تنها چهار الگو رؤیت شده که عبارت‌اند از: ۱) الگوی اول (H,H,H)، که خوشهٔ اول

Archive of SID.ir

Archive of SID.ir

و ششم دارای این نوع الگو هستند، ۲) الگوی دوم (L,H,H)، که خوشهٔ سوم و پنجم دارای این الگو هستند، ۳) الگوی سوم (H,L,L)، که تنها خوشهٔ چهار دارای این الگوست، و ۴) الگوی چهارم (L,L,L)، که تنها خوشهٔ دوم از این الگو تبعیت می‌کند. برای نامگذاری این چهار الگو که تعیین‌کنندهٔ تعداد بخش‌بندی‌های بازار شرکت مورد بررسی خواهد بود، از ترکیب شیوه‌های مختلف نامگذاری در پژوهش‌های «خواجوند، تقوی‌فرد و نجفی» (۱۳۹۱)، Indika et al. (2015); Alvandi, Fazli & Abdoli (2012); Chang & Tsay (2004); Margianti et al. (2016); Aryuni & Madyatmadja & Miranda (2018); Mahdiraji et al. (2019); استفاده شده که شرح آن در ادامه آمده است:

۱. الگوی اول (H,H,H): مشتریانی که از نظر شاخص‌های دفعات خرید و ارزش پولی دارای مقدار بیشتری از میانگین بوده و همچنین، از نظر شاخص تازگی خرید دارای مقدار بیشتر از میانگین هستند و این بدان معناست که این مشتریان پیش‌تر، خریدهای خوبی داشته‌اند. این بخش از بازار شامل ۱۶۱۵۲ مشتری است که تقریباً ۳۶ درصد سهم بازار شرکت را شامل است. از این پس این دسته از مشتریان را مشتریان طلایی بالقوه می‌نامیم.

۲. الگوی دوم (L,H,H): مشتریانی با تعداد دفعات خرید و ارزش پولی بالاتر از میانگین که به‌تازگی نیز خریدهای خوبی داشته‌اند، برای شرکت سودآوری بسیاری دارند و این بخش از بازار شامل ۵۴۵۱ مشتری است که تقریباً ۱۲ درصد سهم بازار شرکت را شامل است. از این پس، این دسته از مشتریان را مشتریان خاص و کلیدی می‌نامیم.

۳. الگوی سوم (H,L,L): مشتریانی که خریدهای کمی داشته و متوسط ارزش ریالی خریدشان نیز کمتر از میانگین کل بوده و مدت زیادی است که خریدی نداشته‌اند. این بخش از بازار شامل ۱۲۶۱۸ مشتری است که معادل ۲۸ درصد سهم بازار شرکت را شامل است و از این پس این دسته از مشتریان را مشتریان نامطمئنِ ازدست‌رفته می‌نامیم.

۴. الگوی چهارم (L,L,L): مشتریانی که شرکت به‌علت تازگی خرید و کم بودن دفعات خرید و ارزش ریالی خریدشان نسبت به آن‌ها شناخت کافی نداشته و این مشتریان برای شرکت جنبهٔ مشتریان تازه و ناآشنا را دارد. این بخش از بازار شامل ۱۰۸۱۹ مشتری، معادل ۲۴ درصد از سهم بازار شرکت است و از این پس مشتریان نامطمئنِ جدید نامیده می‌شوند.

Archive of SID.ir

Archive of SID.ir

۵. نتیجه‌گیری

در این پژوهش به‌منظور تجزیه و تحلیل ارزش مشتریان و همچنین، با هدف رفع بعضی از معایب مدل تازگی، تکرار و ارزش پولی از ویژگی‌های رفتاری و جمعیت‌شناختی مشتریان در کنار مدل تازگی، تکرار و ارزش پولی استفاده شده است. پژوهشگران به‌منظور لحاظ نمودن اهمیت نسبی هر شاخص که در صنایع مختلف بسته به ماهیت بازار می‌تواند متفاوت باشد، مدل WRFM را پیشنهاد کرده‌اند. در این پژوهش وزن‌های شاخص‌های این مدل با توجه به نظر خبرگان و به کمک جدول مقایسات زوجی و با توجه به نرخ ناسازگاری (۰/۰۵۲) حاصل از روش فرایند تحلیل سلسله‌مراتبی (AHP)، به‌ترتیب، (۰/۱۵)، (۰/۲۹)، و (۰/۵۶) برای شاخص‌های تازگی، تکرار و ارزش پولی تعیین شده است و این از دید خبرگان پژوهش حاضر نشان از اهمیت نسبی بیشتر شاخص ارزش پولی داشته است. نتایج اوزان مدل پیشنهادشدهٔ این پژوهش در راستای پژوهش‌های Stone (1995); Bin, Peiji & Dan (2008); Hosseini, Maleki & Gholamian (2010); Khajvand & Tarokh (2011); Zaheri et al. (2012); Hughes (1994)، Blattberg, Kim & Nesli (2008)، «خواجوند، تقوی‌فرد و نجفی» (۱۳۹۱) که به برابری اوزان برای هر سه شاخص تازگی، تکرار و ارزش پولی تأکید داشته‌اند، تفاوت دارد.

به‌طور کلی، می‌توان گفت این پژوهش از نظر روش‌شناسی پژوهش بیشترین همخوانی را با پژوهش Khajvand & Tarokh (2011) دارد، ولی از منظر بازار مورد بررسی با پژوهش‌های Zaheri et al. (2012); Alvandi, Fazli & Abdoli (2012)، «خواجوند، تقوی فرد و نجفی» (۱۳۹۱)، Mahdiraji et al. (2019); Aryuni & Madyatmadja & Miranda (2018); «رنگریز و بایرامی» (۱۳۹۸) و Abbasimehr & Shabani (2020) تفاوت آشکار دارد. همچنین، در این پژوهش برای تعیین تعداد خوشهٔ بهینه از شاخص‌های تعداد تکرار، درصد خطا، درصد خلوص و شاخص اطلاعات متقابل نرمال‌شده استفاده شده، در حالی که در پژوهش‌های «ایزدی» و همکاران (۱۳۹۵) و Kabasakal (2020) از شاخص «سیلوئت»[1]، Khajvand & Tarokh (2011) از شاخص «دان»[2]، «کفاش‌پور، توکلی و علیزاده» (۱۳۹۱) و Wen & Tseng (2011) از شاخص «وارد»[3]، Hosseini, Maleki & Gholamian (2010)، Babaiyan & Sarfarazi (2019) از شاخص «دیویس-بولدین» برای شناسایی تعداد خوشه‌های بهینه استفاده شده است.

1. silhouette
2. Dunn
3. Ward

Archive of SID.ir

Archive of SID.ir

از بین ۶۴۲۱۰ دادهٔ باقی‌مانده بعد از عملیات تشخیص مغایرت بر اساس بازه‌ای که امتیاز کلی شاخص‌های مدل WRFM هر مشتری قرار داشته است (بازهٔ یک تا پنج) برای مشتریان برچسب‌هایی تحت عناوینِ طلایی، نقره‌ای معمولی، پُرریسک و بحرانی تعیین شد. از کل داده‌ها تعداد ۱۹۶۵۴ مشتری معمولی، ۱۶۳۲۸ مشتری نقره‌ای، ۱۴۲۴۳ مشتری طلایی، ۱۲۱۹۶ مشتری پُرریسک و ۱۷۸۹مشتری بحرانی تعیین و مشخص شد. با نگاهی به درصد فراوانی مشتریان طلایی (۲۲/۱۸ درصد) همچنین، با توجه به این‌که اساس مدل RFM نیز تا حدود بسیاری بر پایهٔ قانون ۲۰-۸۰ استوار است، پس می‌توان گفت که این دسته از مشتریان ۸۰ درصد از درآمد شرکت را ایجاد خواهند کرد. و این با اساس کار مدل کلی RFM همخوانی داشته و در این پژوهش مورد تأیید قرار گرفته است.

با توجه به هدف پژوهش و بهره‌مندی از مدل تازگی، تکرار و ارزش پولی موزون و استفاده از روش‌های خوشه‌بندی شبکهٔ عصبی «کوهنن»، خوشه‌بندی دو-مرحله‌ای و خوشه‌بندی K- میانگین، ۶ خوشه با ۵ نوع مشتری شناسایی شدند و با استفاده از شیوهٔ نامگذاری در پژوهش‌های Chang & Tsay (2004)، Aryuni, Madyatmadja & Miranda (2018)، Mahdiraji et al. (2019)، Babaiyan & Sarfarazi (2019) چهار عنوان و دستهٔ کلی به شرح زیر برای مشتریان تعیین شد: ۱) مشتریان خاص و کلیدی با ۵۴۵۱ مشتری، ۲) مشتریان طلایی بالقوه با ۱۶۱۵۲مشتری، ۳) مشتریان نامطمئنِ ازدست‌رفته با ۱۲۶۱۸ مشتری، و ۴) مشتریان نامطمئنِ جدید با ۱۰۸۱۹ مشتری.

با بررسی وضعیت خاص مشتریان خوشهٔ اول مطابق جدول ۱۲، که دارای بالاترین مقدار میانگین از نظر مدل تازگی، تکرار و ارزش پولی موزون با مقدار ۴/۶۱۲ بود و همچنین با توجه به دارا بودن بیشترین میانگین ارزش پولی خرید که نشان از ارزشمندی ویژهٔ این خوشهٔ مشتریان برای شرکت دارد و همچنین، با توجه به واریانس و ضریب تغییرات مؤلفه‌های مدل RFM این دسته از مشتریان، پژوهشگران با مشورت با مدیران ارشد فروش شرکت تصمیم گرفتند که این خوشه از مشتریان را به‌رغم داشتن الگوی اول (H,H,H)، در دستهٔ مشتریان خاص و کلیدی قرار دهند. بنابراین، تعداد مشتریان خاص و کلیدی به ۱۱۷۷۴ مشتری و سهم بازار این دسته به ۲۶/۱۴ درصد افزایش یافت. این مقدار تقریباً با قانون ۲۰-۸۰ تطابق دارد (۲۰ درصد مشتریان ۸۰ درصد درآمد شرکت را ایجاد می‌کنند) که شرکت باید طبق این دسته‌بندی با توجه به منابع محدود خود اولویت‌بندی‌هایی را لحاظ نماید تا بر مشتریانی که بیشترین سودآوری را برای شرکت در

Archive of SID.ir

Archive of SID.ir

بر دارند، متمرکز شود و برنامه‌های ویژه‌ای (مانند ارائهٔ تخفیفات کالایی نقدی و غیرنقدی خاص، امکان پرداخت چکی محصولات خریداری‌شده با بازپرداخت طولانی‌مدت‌تر، کاهش زمان ارسال محصول و تسهیلات ویژهٔ دیگر) را به این دسته از مشتریان اختصاص دهد تا ضمن حفظ مشتریان کلیدی، به افزایش خرید در آن‌ها نیز کمک کرده باشد. البته، نباید گروه مشتریان طلایی بالقوه را نیز نادیده بگیرد، زیرا این مشتریان پتانسیل لازم برای تبدیل شدن به مشتریان خاص و کلیدی شرکت را دارند؛ فقط لازم است مدیران فروش و بازاریابی شرکت راهبردهای تشویقی و ترغیبی برای جذب آن‌ها در نظر بگیرند. همچنین، برای دستهٔ مشتریان نامطمئن ولی جدید، به‌منظور حفظ و تداوم خرید آن‌ها از ابزارهای مختلف تبلیغاتی بهره برده و به ارائهٔ تخفیفات مختلف اقدام نمایند. لکن همیشه باید این موضوع را مد نظر داشته باشند که این مشتریان ریسک بالایی برای شرکت داشته و دارند و شرکت نباید منابع زیادی از بودجهٔ محدود خود را به این دسته تخصیص دهد.

در این پژوهش برای سنجش کیفیت و انتخاب روش مناسب خوشه‌بندی از معیارهای تعداد تکرار، میزان خطا، درصد خلوص و شاخص اطلاعات متقابل نرمال شده استفاده شده است که پیشنهاد می‌شود به‌منظور بررسی میزان تفاوت‌هایی که ممکن است در نتایج این پژوهش به‌واسطهٔ نوع روش انتخابی ایجاد شود، این پژوهش را با شاخص‌های «دیویس بولدین»، شاخص «دان»، معیار «سیلوئت»، امتیاز جینی وآنتروپی نیز مورد بررسی قرار دهند.

این پژوهش در بازهٔ زمانی خاصی از سال‌های ۱۳۹۸- ۱۳۹۹ انجام شده و مربوط به مشتریان خرده‌فروشی، تعاونی مصرف و محلی مناطق شرقی، جنوب و جنوب شرقی استان تهران بود که با آغاز شیوع پاندمی کرونا مصادف بود؛ دوره‌ای که مردم ایران شاید به‌طور کامل به شرایط موجود عادت نکرده بودند. بنابراین، انجام پژوهشی که اثر تغییر شیوهٔ رفتاری خرید محصولات بهداشتی و آرایشی مردم در دوران اوج (پیک) پاندمی کرونا را بر میزان خرید مشتریان خرده‌فروشی، تعاونی مصرف و محلی بررسی نماید، می‌تواند خود موضوع جالب و مهمی برای پژوهش باشد؛ چرا که تغییر نوع شیوهٔ زندگی اثرات خود را به‌طور محسوس بر شیوهٔ خرید و نوع خرید محصولات خواهد گذاشت و نتایج آن می‌تواند دستاوردهای مفیدی برای سرمایه‌گذاران در بخش تولیدات محصولات بهداشتی وآرایشی و همچنین برای مدیران تولید و بازاریابی شرکتِ مورد بررسی دربر داشته باشد.

Archive of SID.ir

Archive of SID.ir

فهرست منابع

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* ۷ (۱) :۱-۲۲.

خواجوند، سمانه، محمدتقی تقوی‌فرد، و اسماعیل نجفی. ۱۳۹۱. بخش‌بندی مشتریان بانک صادرات ایران با استفاده از داده‌کاوی. *مطالعات مدیریت بهبود و تحول* ۲۲ (۶۷): ۱۷۹-۲۰۰.

رنگریز، حسن، و زهرا بایرامی شهریور. ۱۳۹۸. تأثیر مدیریت ارتباط با مشتری الکترونیکی بر وفاداری مشتریان با استفاده از تکنیک‌های داده‌کاوی. *مطالعات مدیریت کسب‌وکار هوشمند* ۷ (۲۷): ۱۷۵-۲۰۵.

رئیسی وانانی، سینا، ایمان رئیسی وانانی، و محمدتقی تقوی‌فرد. ۱۳۹۹. مدلی برای بخش‌بندی یادگیرندگان و بهبود عملکرد آموزشی با استفاده از الگوریتم‌های داده‌کاوی. *مطالعات مدیریت کسب‌وکار هوشمند* ۹ (۳۳): ۵-۳۸.

شهرابی، جمال. ۱۳۹۴. *داده‌کاوی*. تهران: سروش گیتا.

صالحی صدقیانی، جمشید، و مریم اخوان. ۱۳۸۵. مدیریت ارتباط با مشتری. *حسابدار* ۲۱ (۱۷۶): ۲۳-۳۶.

صمدی راد، برات‌اله. ۱۳۸۰. اصول مشتری‌گرایی و بازاریابی (با نگاهی به نقش ارتباطات انسانی در مشتری‌گرایی). *مجله هنر هشتم*، ۲۱ و ۲۲: ۹۴-۹۸.

صنایعی، علی. ۱۳۸۱. *بازاریابی و تجارت الکترونیک*. اصفهان: جهاد دانشگاهی.

عاشوری، مریم، مونا شریف‌خانی، و محمدجعفر تارخ. ۱۳۹۳. توسعهٔ مدل فرایندی مدیریت دانش مشتری با استفاده از سیستم‌های مدیریت ارتباط با مشتری. *رشد فناوری* ۴ (۴۰): ۶۲-۶۸.

علیزاده، سمیه، و سمیرا ملک‌محمدی. ۱۳۹۳. *داده‌کاوی و کشف دانش گام‌به‌گام با نرم‌افزار Clementine*. تهران: دانشگاه صنعتی خواجه نصیرالدین طوسی.

کفاش‌پور، آذر، احمد توکلی، و علی علیزاده زوارم. ۱۳۹۱. بخش‌بندی مشتریان بر اساس ارزش دوره عمر آن‌ها با استفاده از داده‌کاوی بر مبنای مدل (RFM). *پژوهش‌های مدیریت عمومی* ۵ (۱۵): ۶۳-۸۴.

کیگان، وارن جی. ۲۰۰۴. *مدیریت بازاریابی جهانی*. ترجمه و تلخیص عبدالحمید ابراهیمی. ۱۳۸۳. تهران: دفتر پژوهش‌های فرهنگی.

ویسی، هادی. ۱۳۹۶. *جزوه درسی روش‌های آماری در پردازش زبان طبیعی (خوشه‌بندی)*. دانشکده فناوری و علوم مدرن، دانشگاه تهران.

References

Abbasimehr, H., and M. Shabani. 2020. A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers. *Kybernetes* 50 (2): 221-242.

Archive of SID.ir

Archive of SID.ir

Alvandi, M., S. Fazli, and F. S. Abdoli. 2012. K-Means clustering method for analysis customer lifetime value with LRFM relationship model in banking services. *International Research Journal of Applied and Basic Sciences* 3 (11): 2294-2302.

Armstrong, G., Ph. Kotler, L. Harris, and Nigel Piercy. 2014. *Principles of marketing*. 6th edition.?: Pearson.

Aryuni, M., E. D. Madyatmadja, and E. Miranda. 2018. Customer segmentation in XYZ bank using K-means and K-medoids clustering. In 2018 *International Conference on Information Management and Technology (ICIMTech)* (pp. 412-416). IEEE Jakarta, Indonesia.

Azadnia, A. H., M. Z. Mat Saman, K. Yew Wong, and A. R. Hemdi. 2011. Integration model of Fuzzy C-means clustering algorithm and TOPSIS Method for Customer Lifetime Value Assessment. *In Industrial Engineering and Engineering Management (IEEM)*, 2011 IEEE International Conference on (pp. 16-20), IEEE Singapore.

Babaiyan, V. and S. A. Sarfarazi. 2019. Analyzing Customers of South Khorasan Telecommunication Company with Expansion of RFM to LRFM Model. *Journal of AI and Data Mining* 7 (2): 331-340.

Bashiri Mosavi, A., and A. Afsar. 2018. Customer value analysis in banks using data mining and fuzzy analytic hierarchy processes. *International Journal of Information Technology & Decision Making* 17 (3): 819–840.

Bhojani, Shital H., and Nirav Bhatt. 2016. Data Mining Techniques and Trends – a Review. *Global Journal for Research Analysis (GJRA*), 5 (5): 252-254.

Bin, Deng, Shao Peiji, and Zhao Dan. 2008. Data mining for needy students identify based on improved RFM model: a case study of university. *Proceedings of the 2008 International Conference on Information Management, Innovation Management, and Industrial Engineering*, 1: 244-247. Taipei, Taiwan.

Blattberg, R. C., B. D. Kim, and Scott A. Neslin. 2008. RFM Analysis in Database Marketing. NewYork, NY: Springer.

Chang, H. H., and S. F. Tsay. 2004. Integrating of SOM and K-means in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal of Information Management*:(4) 11 161-203.

Chiu, Ch.Y., Ch. Yi- Feng, I. T. Kuo, and H. Ch. Ku. 2009. An intelligent market segmentation system using k-means and particle swarm optimization.*Expert Systems with Applications* 36: 4558-4565.

Han, J., M. Kamber, and J. Pei. 2011. *Data Mining: Concepts and techniques*. (3rd Ed.). Morgan Kaufmann publications Printed in the United States of America: Elsevier.

Hanafizadeh, P., and M. Mirzazadeh. 2011. Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method – ADSL market of a telecommunication company. *Expert Systems with Applications* 38: 198-205.

Hosseini, S. M., A. Maleki, and M. R. Gholamian. 2010. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications* 37: 5259–5264.

Hu, W., and J. Zhang. 2008 (December 10-11). Study of segmentation for auto services companies based on RFM model.*Proceedings from the 5th International Conference on Innovation and Management*. Maastricht, the Netherlands.

Hughes, A. M. 1994. *Strategic database marketing: the master plan for starting and managing a profitable, customer-based marketing program*. Chicago, IL: Probus Publishing.

Hwang, S., and Y. Lee. 2021. Identifying customer priority for new products in target marketing: Using RFM model and TextRank. *Innovative Marketing* 17 (2): 125-136.

Archive of SID.ir

Archive of SID.ir

Indika, H. A, R. M. Kapila Tharanga Rathnayaka, & S. K. Illangarathne. 2015. Mining Profitability of Telecommunication Customers Using K-Means Clustering. *Journal of Data Analysis and Information Processing* 3 (3): 63-71.

Jintana, J., and T. Mori. 2019. Customer clustering for a new method of marketing strategy support within the courier business. *Academia Book Chapter* 31 (2): 1-19.

Kabasakal, İ. 2020. Customer Segmentation Based On Recency Frequency Monetary Model: A Case Study *E-Retailing. Bilişim Teknolojileri Dergisi* 13 (1): 47-56.

Khajvand, M. and M. J. Tarokh. 2011. Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science* 3: 1327–1332.

Khajvand, M., K. Zolfaghar, S. Ashoori, and S. Alizadeh. 2011. Estimating Customer Life Time Value based on RFM analysis of customer purchase behavior: case study. *Procedia Computer Science* 3: 57-63.

Li, D. Ch., L. D. Wen, and W. T. Tseng. 2011. A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. *Expert Systems with Applications* 38 (6): 7186-7191.

Mahdiraji, H. A., E. Kazimieras Zavadskas, A. A. Kazeminia, and A. A. Abbasi Kamardi. 2019. Marketing strategies evaluation based on big data analysis: aClusteing-MCDM approach. *Economic Research-Ekonomska Istraživanja* 32 (1): 2882-2898.

Margianti, Eko. Sri, Rina Refianti, A. Benny Mutiara, and Nuzulina. 2016. Affinity Propagation and RFM-Model for CRM'S DATA Analysis. *Journal of Theoretical and Applied Information Technology*. 84 (2): 272-282.

Rokach, L. 2010. Data mining ad Knowledge discovery handbook: A survey of clustering algorithms. US: Springer.

Sheikh, A., T. Ghanbarpour, and D. Gholamiangonabadi. 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.

Stone, B. 1995. *Successful direct marketing methods*. Lincolnwood, IL: NTC Business Books.

Wu, H. H., E. C. Chang, & C. F. Lo. 2009. Applying RFM model and K-means method in customer value analysis of an outfitter. In *Global Perspective for Competitive Enterprise, Economy and Ecology - Proceedings of the 16th ISPE International Conference on Concurrent Engineering|Global Perspect. Compet. Enterp., Econ. Ecol. - Proc. ISPE Int. Conf. Concurrent Eng.* (pp. 665-672). Springer-Verlag London Limited 2009.

Wei, J. T., L. Shih-Yen. W. Chih-Chien, and W. Hsin-Hung. 2012. A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Systems with Application* 39 (5): 5529-553

Zaheri, F., H. Farughi, H. Soltanpanah, S. Alaniazar, and F. Naseri. 2012. Using multiple criteria decision making models for ranking customers of bank network based on loyalty properties in weighted RFM model. *Management Science Letters* 2 (2): 697-704.

Archive of SID.ir

Archive of SID.ir



امید بشردوست

متولد سال ۱۳۶۲، دارای مدرک تحصیلی دکتری در رشتهٔ مدیریت صنعتی، گرایش سیستم‌ها از دانشگاه جامع آزاد اسلامی واحد رودهن است.

مدل‌سازی، شبیه‌سازی سیستم‌های پویا و گسسته پیش‌آمد، تحلیل داده‌ها و داده‌کاوی، ارزیابی عملکرد و تحلیل راهبردی شرکت‌ها و مدیریت ارتباط با مشتری، گردشگری و توسعهٔ پایدار از جمله علایق پژوهشی وی است.



عزت‌اله اصغری‌زاده

متولد۱۳۳۶، دارای مدرک دکتری تخصصی در رشتهٔ مدیریت مهندسی و عملیات از دانشگاه کوئینزلند استرالیاست. ایشان هم‌اکنون دانشیار دانشکدهٔ مدیریت دانشگاه تهران است.

تصمیم‌گیری چندمعیاره، ارزیابی عملکرد، زنجیره تأمین، تحقیق در عملیات، تحلیل آماری، و داده‌کاوی از جمله علایق پژوهشی وی است.



محمدعلی افشار کاظمی

متولد ۱۳۴۵، دارای مدرک دکتری تخصصی در رشتهٔ مدیریت صنعتی گرایش تحقیق در عملیات از دانشگاه آزاد اسلامی واحد علوم و تحقیقات است. ایشان هم‌اکنون دانشیار دانشکدهٔ مدیریت دانشگاه آزاد اسلامی واحد تهران مرکزی است.

منطق فازی، تصمیم‌گیری فازی، شبکه عصبی، الگوریتم ژنتیک، شبیه‌سازی، سیستم‌های پویا، ارزیابی عملکرد، داده‌کاوی، تحلیل پوششی داده‌ها، و تحقیق در عملیات از جمله علایق پژوهشی وی است.

Archive of SID.ir