Karafan

Quarterly Scientific Journal of Technical and Vocational University

Special Issue 2023, Vol. 20, p. 777-800

Journal homepage: https://karafan.tvu.ac.ir/?lang=en

doi:10.48301/KSSA.2022.316104.1852

TVU

Original Research

E-ISSN: 2538-4430
ISSN: 2382-9796

Classification and Prediction of Customer Categories Using Combination of LRFM Method, Quartiles and Multi-Class Data Mining Methods

Hossein Ghaffarian[1*], Ali Reza Bamohabbat[2]

[1]Assistant Professor, Department of Computer Engineering, Faculty of Engineering, Arak University, Arak, Iran.

[2]MSc. Graduated from department of Information Technology, E-commerce, Atrak Quchan Non-profit University. Quchan, Iran.

ARTICLE INFO

Received: 12.30.2021
Revised: 04.08.2022
Accepted: 05.16.2022

Keyword:
Customer relationship management
Data mining
Clustering
LRFM
Classification
Multi-class performance evaluation metrics

***Corresponding Author:**
Hossein Ghaffarian
Email: *h-ghaffarian@araku.ac.ir*

ABSTRACT

Today, Customer Relationship Management (CRM) has become an inevitable necessity in organizations. However, one of the primary problems in this area is the lack of a clear criterion for classifying customers. Creating predictive models from customer categories is also one of the challenges of customer relationship management. For this purpose, in this paper, the combination of LRFM method with quartile concepts and multi-class data mining methods was used to categorize customers. In this solution, customer information records were first reviewed and refined to remove invalid data. Customers were then categorized by combining the concepts of the LRFM method with quartiles. In order to identify and classify new customers, the obtained outputs are subjected to feature selection and their additional features removed. Then, the remaining features were transferred to different classifiers to create predictive customer category models. Contrary to previous research achievements, in this paper, multi-class microscopic and macroscopic evaluation criteria were used to evaluate the performance of predictors. In order to evaluate the results, in a case study, part of the customer information of Sarchin Khorasan Company was used. Evaluations showed that the accuracy of the predicted models of the proposed method in customer classification was higher than the conventional method of customer classification using K-Means. In addition, the proposed method performed better in group detection of records.

©2023 Technical and Vocational University, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0 license) (https://creativecommons.org/licenses/by-nc/4.0/).

EXTENDED ABSTRACT

Introduction

Customer relationship management (CRM) refers to all processes and technologies that are used in companies and organizations to identify, encourage, expand, maintain and provide service to customers. In this article, using data mining methods, a new model is presented to classify and predict the assignment of customers to different categories. For this purpose, a proposed combined LRFM method with the concepts of quantiles and data mining methods was proposed to categorize customers. In this method, first, customers are placed in different categories based on the LRFM model and statistical data. For this purpose, concepts of quartiles and outlier data were used. Then, with the help of data mining classifiers, multi-class predictive models of customer categories were created. This step included removing redundant features and selecting superior features in the data set before creating predictive models.

Methodology

The proposed method of this article consists of two parts. First, the method of customer categorization is discussed, and in the second part, the class of customers is determined.

Customer categorization

The proposed method of this article to categorize customers is a method based on the LRFM model. The proposed method for categorizing customers included 3 steps. At first, customer data was extracted. This data can have a different structure in each set. Therefore, extracting and refining these data is considered as the first step of the proposed method. The purpose of refining customer data was to correct or, if necessary, remove unwanted or null values provided by customers. The presence of these values has adverse effects on the quality of the final results of data analysis. In the second step of this method, the parameters related to the LRFM model were divided into two 3- and 5-part forms, and the result of the LRFM value was obtained from the product of these values. In the last stage of the proposed method, first, second and third quartiles were used to place customers in different categories. These three values showed well the state of data dispersion in a data set with skewness.

Determining the class of customers

In the method used in this article to classify and build decision-making models, first, irrelevant features were removed from the data set. Then, with the help of feature selection methods, redundant features were also removed. In the final step, a decision-making model was built with the help of the remaining features, and finally this model was tested and its accuracy used to determine the class of other records. Here, irrelevant features are those features whose presence or absence has no effect on the decision-making process. In addition, in this article, Pearson's criterion was used to select useful

features. Uncorrelated variables were kept and one of each pair of positively or negatively correlated variables was removed.

Results and discussion

In the present work, a data set containing 1134 information records of the sale of products of Sarchin Khorasan Company was used. The data set had 12 features.

Categorizing customers

Because of the popularity of the K-Means method in categorizing customers, in order to compare the results, a classification structure based on K-Means was also implemented. In addition, in order to determine the best number of clusters in K-Means, with the help of this met Silhouette parameter, customers were categorized into four different clusters.

Also, the selected features for the LRFM model were as follows:

- The features of purchasing a product or similar products were used for *Recency*.
- The number of visits the store was used to characterize *Frequency*.
- The number of products purchased by each customer was used for *Monetary*.
- To characterize the *Length* of the customer companionship period, the visit feature of the company products was used.

The LRFM formula used in this article was the product of the first 4 parameters and the sum of their results with the fifth parameter. Furthermore, based on the values of the quantiles, the customers were divided into four categories: gold, silver, bronze and ordinary customers. The choice of number 4 was to match and compare the results of this method with the results of the combination of K-Means and Silhouette. Customers with LRFM value less than or equal to 5 (first quartile) were categorized as normal customers, customers with LRFM value greater than 5 and smaller equal to 6 (second quartile) as bronze customers, customers with LRFM value greater than 6 and smaller equal to 10 (third quartile) as silver customers and other customers (LRFM value greater than 10) as gold customers.

Evaluation of classifications

To evaluate the results, 13 different classifiers were used in IBM SPSS Modeler 18 software. In addition, the dataset was divided into training set (including 66% or 753 records) and test set (including 34% or 381 records).

Considering that four different classes of customers were defined in this data set, the problem was a multi-class problem. Thus, to evaluate the results of multi-class models, we used the relationships of the article (Skolova and Lapalmi, 2009). In multi-class models, two microscopic and macroscopic procedures were defined for this purpose. The difference between these two procedures was that in the microscopic model, the average of all sample values in different classes was used. However, in the macroscopic model, the average values obtained for the classes were used. Since the SPSS Modeler software was not able to analyze the performance of multi-class models, the results of determining the class of each class were collected and entered into MATLAB.

Reviewing the results shows that, in general, the used classifiers with the outputs of the proposed method performed much better. While in the results of classification with K-Means, the best classification accuracy was related to the logistic regression class, which was close to 100%; in the results of the proposed method, six classifiers were able to obtain 100% accuracy. This obvious superiority indicates that the performance of the proposed method for customer clustering was quite effective.

In addition, the average results for microscopic parameters in classification with K-Means were higher than the similar values in the proposed method. However, in the macroscopic parameters of the observed procedure, it was completely the reverse. This shows that the combination method of K-Means and Silhouette was focused on improving the result of individual class recognition, and in contrast, the proposed method was based on improving the result of group recognition. These observations can be used to determine the desired method. The selected evaluation approach can be a tool for determining organizational goals for CRM and how to interact with customers.

Conclusion

In this article, a novel method for categorizing and predicting the category of each customer was presented. The results show that the outputs of the proposed method performed better not only in the individual evaluation of the class members, but also in the performance of group recognition of customers.

کارافن
فصلنامه علمی دانشگاه فنی و حرفه‌ای
ویژه‌نامه ۱۴۰۲، دوره ۲۰، ۸۰۰-۷۷۷
آدرس نشریه: https://karafan.tvu.ac.ir/
doi:10.48301/KSSA.2022.316104.1852

مقاله پژوهشی
شاپای الکترونیکی: ۲۵۳۸-۴۴۳۰
شاپای چاپی: ۲۳۸۲-۹۷۹۶

دسته‌بندی و پیش‌بینی دسته‌های مشتریان به کمک تلفیق روش LRFM، چندک‌ها و روش‌های داده‌کاوی چندکلاسه

حسین غفاریان[۱*]، علیرضا بامحبت[۲]

۱- استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه اراک، اراک، ایران.
۲- فارغ التحصیل کارشناسی ارشد، گروه فناوری اطلاعات، گرایش تجارت الکترونیک، دانشگاه غیرانتفاعی اترک قوچان، قوچان، ایران.

اطلاعات مقاله

دریافت مقاله: ۱۴۰۰/۱۰/۰۹
بازنگری مقاله: ۱۴۰۱/۰۱/۱۹
پذیرش مقاله: ۱۴۰۱/۰۲/۲۶

کلید واژگان:
مدیریت ارتباط با مشتریان
داده‌کاوی
خوشه‌بندی
LRFM
کلاس‌بندی

***نویسنده مسئول:** حسین غفاریان
پست الکترونیکی:
h-ghaffarian@araku.ac.ir

چکیده

امروزه مدیریت ارتباط با مشتری به ضرورتی اجتناب‌ناپذیر در سازمان‌ها تبدیل شده است. با این حال، یکی از مشکلات اولیه در این زمینه، فقدان معیار مشخص برای طبقه‌بندی مشتریان است. ایجاد مدل‌های پیش‌بینی از دسته‌های مشتری نیز یکی از معضلات مدیریت ارتباط با مشتری است. برای این منظور در این مقاله، از تلفیق روش LRFM به همراه مفاهیم چندک‌ها و روش‌های داده‌کاوی چندکلاسه برای دسته‌بندی مشتریان استفاده شده است. در این راهکار، ابتدا رکوردهای اطلاعاتی مشتریان بررسی و پالایش شدند تا داده‌های نامعتبر حذف گردند. سپس با ترکیب مفاهیم روش LRFM و چندک‌ها، مشتریان دسته‌بندی شده‌اند. در ادامه، به‌منظور تشخیص و طبقه‌بندی مشتریان جدید، خروجی‌های به‌دست‌آمده، در معرض انتخاب ویژگی قرار گرفتند و ویژگی‌های اضافی آنها حذف شدند. سپس، ویژگی‌های باقی‌مانده برای ایجاد مدل‌های پیش‌بینی‌کننده دسته مشتریان به طبقه‌بندهای مختلف منتقل شدند. برخلاف تحقیقات پیشین، در این مقاله، از معیارهای ارزیابی چندکلاسه میکروسکوپی و ماکروسکوپی برای ارزیابی عملکرد پیش‌بینی‌کننده‌ها استفاده شده است. به‌منظور ارزیابی نتایج، در یک مطالعه موردی، بخشی از اطلاعات مشتریان شرکت سرچین خراسان مورد استفاده قرار گرفته است. ارزیابی‌ها نشان می‌دهد که دقت مدل‌های پیش‌بینی روش پیشنهادی در طبقه‌بندی مشتریان، بالاتر از دقت روش مرسوم دسته‌بندی مشتریان به کمک K-Means است. همچنین روش پیشنهادی در تشخیص گروهی رکوردها عملکرد بهتری دارد.

©2022 Technical and Vocational University, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0 license) (https://creativecommons.org/licenses/by-nc/4.0/).

مقدمه

مدیریت ارتباط با مشتری (CRM[1]) به همه فرایندها و فناوری‌هایی گفته می‌شود که در شرکت‌ها و سازمان‌ها برای شناسایی، ترغیب، گسترش، حفظ و ارائه خدمت به مشتریان به‌کار می‌رود (حری، ۲۰۱۴). فزونی و دسترسی انبوه به اطلاعات، مشتریان را آگاه‌تر و ماهرتر ساخته است؛ بنابراین سازمان‌ها باید از خط‌مشی‌هایی چون مدیریت ارتباط با مشتری، برای جذب و نگه‌داشت مشتری بهره ببرند (ویسی و غروی، ۲۰۱۵). ارزیابی‌های انجام‌شده نشانگر آن است که رضایت و اعتماد، هردو تأثیر مثبت و معنی‌داری بر چسبندگی مشتریان به فروشندگان دارند (سعیدا اردکانی و همکاران، ۲۰۲۱). (دل‌انگیزان و همکاران، ۲۰۲۲) نشان داده‌اند که در تدوین راهبردهای تجاری‌سازی محصولات، قوانین و مقررات پشتیبان از عوامل تأثیرگذار می‌باشد.

مدیریت ارتباط با مشتریان، فرایندی است تعاملی که به دنبال رسیدن به یک تعادل بهینه بین سرمایه‌گذاری شرکت و برآورده‌کردن نیازهای مشتریان به‌منظور ایجاد حداکثر سود قابل‌تحصیل می‌باشد. مدیریت ارتباط با مشتری توسط چهار عنصر دانش، هدف، فروش و خدمت تعریف شده است (ادلستین، ۲۰۰۱)[2]. ابزار CRM مستلزم حرکت به سمت مشتری‌مداری و تعریف استراتژی بازار از دیدگاه برون‌سازمانی است. این روند تحت عنوان چرخه زندگی مشتری مطرح می‌شود که شامل جذب مشتریان جدید، افزایش ارزش مشتری و حفظ مشتریان خوب است. مدیریت ارتباط با مشتری، زیرساختاری است که ارزش مشتری را آشکار می‌کند و افزایش می‌دهد و همچنین ابزار صحیحی را برای وفادار نگه‌داشتن مشتریان ارزشمند در اختیار قرار می‌دهد (مایمون و روکاچ، ۲۰۱۰)[3].

ارزش دوره عمر مشتری از جمله مهم‌ترین مباحث حوزه CRM می‌باشد. این پارامتر ابزاری‌ست که جایگاه و حوزه فعالیت مشتریان در رابطه با سازمان را مشخص می‌کند؛ بنابراین استفاده از ابزارهایی که با دقت و سهولت بیشتری بتواند به این مهم دست‌یابند و پاسخگوی نیازهای مشتریان باشد، در درجه بالایی از اهمیت قرار دارد. راه‌کارهای داده‌کاوی، ابزارهای سودمندی را برای این موضوع در اختیار سازمان‌ها قرار می‌دهند. داده‌کاوی یکی از اجزای مهم فرایند کشف اطلاعات از داده‌هاست که برای این منظور از روش‌های هوشمند برای استخراج الگوها استفاده می‌کند (هان و همکاران، ۲۰۱۱)[4]. به کمک داده‌کاوی می‌توان مشتریان را به گروه‌های متفاوت تقسیم کرد و براساس آن محصولات و خدمات مشتری‌محور ارائه داد.

داده‌کاوی، علمی است که از آغاز آن تنها سه دهه می‌گذرد و در این فاصله کوتاه توانسته است به شکل تأثیرگذاری در سایر علوم خودنمایی کند. امروز داشتن یک درک بهتر از امور کسب و کار در سازمان‌ها بسیار مهم است. این درک شامل درک مشتریان، بازار، منابع، روش تأمین منابع و رقبای تجاری شرکت‌هاست و امروز از آن به‌عنوان هوش تجاری یاد می‌شود (هان و همکاران، ۲۰۱۱). بدون بهره‌گیری از روش‌های داده‌کاوی، بسیاری از تلاش-ها در زمینه هوش تجاری محکوم به شکست می‌باشد. مواردی همچون تحلیل مؤثر بازار، ارزیابی بازخوردهای مشتریان، ارزیابی نقاط ضعف و قوت رقبا، حفظ مشتریان ارزشمند و اتخاذ تصمیمات هوشمندانه بدون بهره بردن از داده‌کاوی اگر غیرممکن نباشند، به‌سختی میسر خواهند شد.

---

[1] Customer Relationship Management
[2] Edelstein
[3] Maimon & Rokach
[4] Han

در این مقاله با استفاده از روش‌های داده‌کاوی، مدل‌هایی برای دسته‌بندی و پیش‌بینی تعلق مشتریان به دسته‌های مختلف در رکوردهای کسب شده از شرکت سرچین خراسان ارائه شده است. برای این منظور برای دسته‌بندی مشتریان، روشی ترکیبی LRFM با مفاهیم چندک‌ها و روش‌های داده‌کاوی پیشنهاد شده است. در این روش، ابتدا مشتریان براساس مدل LRFM و داده‌های آماری در دسته‌های مختلف قرار گرفته‌اند. برای این منظور از مفاهیم چارک‌های اول، دوم و سوم و مفهوم داده‌های پرت استفاده شده است. مدل LRFM یک نمونه برتر از مدل RFM برای دسته‌بندی مشتریان است که براساس طول دوره وفاداری مشتری، تازگی خرید، تعداد دفعات خرید و ارزش پول خرید کار خود را انجام می‌دهد. در ادامه، مدل‌های پیش‌بینی‌کننده چندکلاسه دسته مشتریان به کمک کلاس‌بندهای داده‌کاوی ایجاد شده است. این مرحله شامل حذف ویژگی‌های زائد و انتخاب ویژگی‌های برتر در مجموعه داده‌ها، قبل از ایجاد مدل‌های پیشگو می‌باشد. همچنین برای ارزیابی نتایج روش‌های پیشنهادی این مقاله از نرم‌افزارهای MATLAB و IBM SPSS Modeler 18 استفاده شده است. نتایج، نشانگر عملکرد مطلوب روش پیشنهادی است.

ادامه ساختار مقاله حاضر به شرح زیر می‌باشد: در بخش دوم، پیشینه تحقیق موضوع حاضر بررسی شد. در بخش سوم جزئیات روش پیشنهادی برای ایجاد و دسته‌بندی مشتریان ارائه گردید. چگونگی ساخت مدل‌های پیشگو برای تعیین دسته مشتریان در بخش چهارم مورد بحث قرار گرفت. در بخش پنجم مقاله به ارزیابی روش پیشنهادی پرداخته شد و در نهایت جمع‌بندی مقاله در بخش ششم ارائه شد.

مروری بر پیشینه تحقیق

به‌منظور تعیین اولویت‌بندی قرار گرفتن مشتریان در دسته‌ها از سبک‌ها و روش‌های گوناگونی استفاده می‌شود. برخی از محققان براساس ارزش فعلی، ارزش بالقوه و وفاداری مشتری، مشتریان را بخش‌بندی کرده‌اند. برخی دیگر براساس طول عمر ارزش مشتری و مؤلفه‌های مدل RFM مشتریان را بخش‌بندی و استراتژی بازاریابی برای هر بخش را تبیین کرده‌اند.

(هوانگ و همکاران، ۲۰۰۴)[1] با در نظر گرفتن سهم سود ایجاد شده، سود بالقوه و تعریف سودآوری مشتری یک مدل ارزش دوره عمر مشتری پیشنهاد کردند و براساس ارزش فعلی، ارزش بالقوه و وفاداری مشتریان را خوشه‌بندی کردند. (کیم و همکاران، ۲۰۰۶)[2] با در نظر گرفتن اهمیت شناخت مشتری در زمینه ایجاد ارتباط بلندمدت، کسب وفاداری و سودآوری بیشتر مشتری، به ارائه چارچوبی برای تحلیل ارزش مشتری، دسته‌بندی مشتریان برمبنای این ارزش و سپس تبیین استراتژی‌های مناسب برای هر بخش پرداختند. آنها مدل جدید طول عمر مشتری و بخش‌بندی مشتریان برمبنای ارزش مشتری و فرصت‌های فروش کناری را پیشنهاد کردند. (جونکر و همکاران، ۲۰۰۴)[3] ابتدا مشتریان را به گروه‌های همگن بخش‌بندی کردند و سپس سیاست‌های بهینه بازاریابی را برای هر بخش تعیین کردند. این تحقیق بر روی بانک اطلاعاتی یک شرکت که شامل اطلاعات ارسال نامه‌های تبلیغاتی بین سال‌های ۱۹۹۴ و ۱۹۹۹ است، صورت گرفته و متغیرهای تعریف شده برای RFM به ترتیب تعداد نامه‌های ارسال شده بعد از آخرین پاسخ مشتری را R، درصد نامه‌های پاسخ داده شده به کل نامه‌های ارسالی را F و میانگین حجم درخواستی مشتری را

[1] Hwang
[2] Kim
[3] Jonker

M تعیین کردند. (بوکینکس و ون دن پوئل، ۲۰۰۵)[1] نیز مدلی مبنی بر پیش‌بینی نرخ ریزش مشتریان وفادار طراحی کردند که تمرکز آن بر انتخاب وفادارترین مشتریان است. براساس پارامترهای تعریف شده در این تحقیق، مشتریانی که تکرار خریدهای آنها بیشتر از میانگین باشد و انحراف از استاندارد زمان صرف شده برای خرید آنها از متوسط زمان میانگین کمتر باشد به‌عنوان مشتریان وفادار انتخاب شده‌اند.

دسته‌بندی مشتریان را می‌توان معادل خوشه‌بندی داده‌ها در دنیای داده‌کاوی دانست.؛ زیرا در هر دو روش، از ابتدا ملاک مشخصی برای تفکیک اطلاعات وجود ندارد و باید این ملاک‌ها به شکل جداگانه تعریف گردند. یک روش خوشه‌بندی خوب سعی می‌کند تا خوشه‌هایی با کیفیت بالا را براساس دو معیار شباهت بالای نقاط داخلی و شباهت کم نقاط خارجی ایجاد کند (کومار و رینارتز، ۲۰۰۵)[2].

(تسای و چیو، ۲۰۰۴)[3] به بسط فراروش جدید بخش‌بندی بازار بر مبنای متغیرهای خاص محصولات، مثل موارد خریداری‌شده و میزان درآمد مربوط به آنها از روی تراکنش‌های قبلی مشتریان پرداخته‌اند. در این تحقیق آنها یک مقیاس اندازه‌گیری شباهت بر مبنای خرید، الگوریتم خوشه‌بندی و تابع کیفیت خوشه‌بندی معرفی کردند و برای تضمین شباهت الگوهای خرید مشتریان یک خوشه، از الگوریتم ژنتیک استفاده کردند.

(لیو و شیه، ۲۰۰۵)[4] با ارائه یک مدل ترکیبی AHP و ماتریس مقایسات زوجی، وزن‌های متغیرهای RFM در ارزیابی وفاداری و ارزش دوره عمر مشتری را تعیین کردند و سپس برای خوشه‌بندی مشتریان براساس مدل RFM از روش‌های داده کاوی استفاده کردند.

(سیدحسینی و همکاران، ۲۰۱۱) به کمک ترکیب مدل RFM برای ارزشیابی مشتریان و الگوریتم خوشه‌بندی K-Means و معیار (دیویس و بولدین، ۱۹۷۹)[5] مشتریان را خوشه‌بندی و براساس آن درجه و تابع وفاداری مشتریان را محاسبه کردند.

(برهمانا و همکاران، ۲۰۲۰)[6] در یک کار مشابه اقدام به خوشه‌بندی مشتریان مبتنی بر مدل RFM و براساس روش‌های خوشه‌بندی K-Means، K-Medoids و DBSCAN کرده‌اند. در این مقاله نیز از معیارهای دیویس-بولدین و Silhouette برای تعیین تعداد خوشه مناسب استفاده شده است.

(ارناواتی و همکاران، ۲۰۲۱)[7] اقدام به بررسی کاربردهای داده کاوی در تقسیم‌بندی مشتریان مبتنی بر RFM کرده‌اند و چارچوبی کلی برای این مسئله در محیط سیستم‌های اطلاعات جغرافیایی ارائه داده‌اند.

(وو و همکاران، ۲۰۲۰)[8] در مقاله خود به بررسی تأثیر خرید مشتریان یک سازمان به کمک مدل RFM و K-Means پرداخته‌اند. در این تحقیق، مشتریان به چهار گروه تقسیم شده‌اند و سعی شده است تا به کمک مدل‌سازی و ارزیابی، رفتارهای سازمان در قبال مشتریان به شکل صحیح مدیریت شود.

[1] Buckinx & Van den poel
[2] Kumar & Reinartz
[3] Tsai & Chiu
[4] Liu & Shih
[5] Davies & Bouldin
[6] Brahmana
[7] Ernawati
[8] Wu

(کریستی و همکاران، ۲۰۲۱)[1] از روش‌های K-Means، Fuzzy C-Means و یک روش K-Means اصلاح شده، در کنار RFM برای دسته‌بندی مشتریان استفاده کرده‌اند. در این مقاله از معیار Silhouette برای تأیید تعداد خوشه‌ها استفاده شده است.

مقالاتی نظیر (چنگ و چن، ۲۰۰۹)[2]، (دوگن و همکاران، ۲۰۱۸)[3]، (ماریانی و همکاران، ۲۰۱۸)[4]، (توکلی و همکاران، ۲۰۱۸)، (هوانگ و همکاران، ۲۰۲۰)[5] و (وهرا و همکاران، ۲۰۲۰)[6] نیز از دیگر منابعی هستند که نویسندگان آنها اقدام به ترکیب و استفاده توأم از RFM و K-Means برای دسته‌بندی مشتریان سازمان‌های مختلف کرده‌اند.

دسته‌بندی مشتریان

مسئله مدیریت ارتباط با مشتریان، همواره یکی از مهم‌ترین دغدغه‌های مدیران واحدهای تولیدی و تجاری بوده است و تا کنون راه‌کارهای متنوعی برای دسته‌بندی مشتریان پیشنهاد شده است. در این بخش به بیان روش پیشنهادی مورداستفاده در این مقاله برای دسته‌بندی مشتریان خواهیم پرداخت. از آنجاکه بخشی از روش پیشنهادی بر پایه مدل LRFM است، ابتدا توضیحاتی در خصوص این مدل ارائه شده است.

مدل RFM و توسعه‌های آن

مشتریان، منبع نهایی رشد و تمام کسب و کارها محسوب می‌شوند و از آنجاکه ارزیابی رفتار آنها دارای اهمیت به‌سزایی است، روش‌های متعددی برای محاسبه ارزش دوره عمر مشتری وجود دارد (سیدحسینی و همکاران، ۲۰۱۱). یکی از معروف‌ترین این مدل‌ها، مدل RFM است که ارائه کرد (هاگز، ۱۹۹۴)[7] ارائه کرد. امروز RFM به‌عنوان یک روش عمده و مشهور مورد استفاده قرار می‌گیرد (زلقی و عباس‌نژاد ورزی، ۲۰۱۴). نام این مدل از ابتدای سه کلمه و به عبارت دیگر سه زیربنای اصلی این مدل گرفته شده است:

- **تازگی خرید**[8]: فاصله زمانی بین آخرین خرید صورت‌گرفته تا امروز
- **تعداد دفعات خرید**[9]: تعداد دفعاتی که مشتری در دوره زمانی خاص اقدام به خرید کرده است.
- **ارزش پولی خرید**[10]: مقدار پولی که مشتری در دوره زمانی خاص برای خرید اختصاص داده است.

با وجود آنکه این مدل، بسیار عمومی است اما مدل محاسبه شاخص‌های آن بسیار متنوع است. ساده‌ترین مدل به‌کاررفته برای RFM این است که هرکدام از شاخص‌های فوق به پنج دسته مجزا تقسیم می‌شود و سپس به جای مقادیر واقعی هر شاخص، یک عدد بین ۱ تا ۵ جانشین می‌گردد. عدد بزرگ‌تر به معنی تازگی، دفعات خرید و ارزش خرید بیشتر مشتری است. در مرحله نهایی، از ترکیب مقادیر عددی سه شاخص برای استنتاج استفاده می‌شود. برای

---

[1] Christy
[2] Cheng & Chen
[3] Dogan
[4] Maryani
[5] Huang
[6] Vohra
[7] Hughes
[8] Recency
[9] Frequency
[10] Monetary Value

این منظور می‌توان از حاصل‌ضرب، حاصل جمع یا حتی چینش مقادیر به فرم XYZ (هر کرکتر نماد مقدار عددی یکی از شاخص‌هاست و مقدار ۱ تا ۵ می‌گیرد) استفاده کرد. نویسندگان کتاب آموزش کاربردی SPSS (طالع و همکاران، ۲۰۱۷) در خصوص رویکردهای متفاوت برای رتبه‌بندی یا امتیازدهی مدل RFM به سه دسته رتبه‌بندی مقدماتی، رتبه‌بندی میان سلولی (خوشه‌ای) و رتبه‌بندی با استفاده از وزن‌دهی هر رویکرد اشاره دارند.

مدل LRFM یکی از روش‌های زیرمجموعه مدل RFM است که در سال ۲۰۰۴ ارائه گردید. در این مدل یکی از ارکان وفاداری مشتریان به یک سیستم، طول دوره همراهی و ارتباط مشتری محسوب شده است. به همین دلیل حرف L (ابتدای عبارت Length) به ابتدای مدل RFM اضافه شده است (چانگ و تسای، ۲۰۰۴)[1]. مدل LRFM امروزه به یکی از مدل‌های مناسب در خوشه‌بندی مشتریان در مدیریت ارتباط با مشتری تبدیل شده است که برای نمونه می‌توان به مقالات (ماریسا و همکاران، ۲۰۱۹)[2]، (علیزاده زوارم و کریمی مزیدی، ۲۰۱۸)، (کائو و همکاران، ۲۰۱۱)[3] و (بابائیان و سرافرازی، ۲۰۱۹) اشاره کرد.

ضعف‌های RFM/LRFM

در مدل RFM و نمونه‌های تکامل‌یافته آن، میزان اهمیت یک مشتری براساس ارزش فعلی، ارزش بالقوه و وفاداری وی به یک سازمان یا کسب و کار سنجیده می‌شود. اما این مدل معمولاً کامل نمی‌باشد و نیازمند استفاده از روش‌های متفاوتی برای ارزیابی بهتر دسته‌ها و اعداد اختصاص‌یافته به مشتریان است. برای درک بهتر مشکلات موجود، ابتدا لازم است تا نقاط ضعف مدل‌سازی فعلی بررسی گردد.

اولین نقطه ضعف مدل فعلی این است که ممکن است هیچ ملاک مشخصی برای دسته‌بندی مشتریان ارائه نشده باشد. برای حل این مشکل می‌توان از روش‌های مختلفی استفاده کرد؛ برای مثال با استفاده از روش‌های خوشه‌بندی، مشتریان به تعدادی خوشه تقسیم می‌شوند و سپس با توجه به ویژگی‌های کلی هر خوشه، یک نام برای آن خوشه در نظر می‌گیریم که در واقع این نام همان نام دسته خواهد بود. اما نکته مهم در خوشه‌بندی این است که چگونه می‌توان تعداد خوشه‌ها را تعیین کرد. بنابراین لازم است تا معیار مناسبی برای تعیین تعداد خوشه‌ها در نظر گرفت.

دومین معضل موجود این است که بعد از محاسبه مقدار RFM/LRFM (به روش حاصل‌جمع، حاصل‌ضرب یا هر روش دیگر)، نیازمند اختصاص کلاس به هر دسته هستیم. برای این کار لزوماً تقسیم مستقیم اعداد به ۳ یا ۵ بخش مساوی در اعدادی که نمودار پراکندگی آنها به لحاظ آماری به یک سمت خاص متمایل است (چولگی به راست یا چپ) نمی‌تواند پاسخ مناسبی را ارائه دهد. پس نیازمند یافتن راه‌کار بهتری برای ایجاد این دسته‌ها هستیم.

دسته‌بندی به کمک تحلیل آماری و مدل LRFM

روش پیشنهادی این مقاله در راستای دسته‌بندی مشتریان، روشی مبتنی بر مدل LRFM است. اما در این مدل و برای فرار از معضلات و ابهامات LRFM سعی شده است از تحلیل‌های آماری برای به‌دست آوردن مقدار LRFM و دسته هر مشتری استفاده شود. شکل ۱ نمودار روش پیشنهادی را نشان می‌دهد.

---

[1] Chang & Tsay
[2] Marisa
[3] Kao

براساس شکل ۱، در ابتدا داده‌های مشتریان استخراج می‌گردد. این داده‌ها در هر مجموعه (و حتی در زیرمجموعه‌های یک مجموعه) می‌تواند ساختار متفاوتی داشته باشد. بنابراین استخراج و پالایش این داده‌ها به عنوان اولین گام روش پیشنهادی در نظر گرفته شده است. منظور از پالایش داده‌های مشتریان، اصلاح و یا در صورت ضرورت، حذف مقادیر ناخواسته یا تهی ارائه‌شده توسط مشتریان می‌باشد. وجود این مقادیر تأثیرات نامطلوبی بر کیفیت نتایج نهایی تجزیه و تحلیل داده‌ها می‌گذارد. مجموعه کاملی از روش‌های ممکن برای این پالایش در (هان و همکاران، ۲۰۱۱) ارائه شده است. انتخاب ترکیب روش‌های مختلف وابستگی کاملی به مجموعه داده‌های در دسترس دارد و نمی‌توان یک الگوی یکسان را برای تمامی موارد اعمال کرد.

شکل ۱. نمودار کلی روش پیشنهادی برای دسته‌بندی مشتریان

در دومین گام از این روش، پارامترهای وابسته به مدل LRFM به دو شکل ۳ و ۵ قسمتی تقسیم شده و نتیجه مقدار LRFM از حاصل‌ضرب این مقادیر در یکدیگر به‌دست می‌آید. همان‌طور که بیان گردید برای این منظور یک روش مشخص وجود ندارد و ممکن است براساس شرایط مسئله و داده‌های دریافتی از مشتریان، این فرمول حتی با اعمال ضرایب یا ترکیبی از جمع و ضرب همراه باشد. با توجه به باز بودن این مسئله، فرمول پیشنهادی این مقاله که براساس داده‌های دریافتی از شرکت سرچین خراسان پیشنهاد شده است، در بخش ارزیابی نتایج بررسی می‌شود.

در آخرین مرحله از روش پیشنهادی، مشتریان باید در دسته‌های مختلف قرار بگیرند. با توجه به اینکه ممکن است پراکندگی داده‌ها به شکلی باشد که نمودار پراکندگی داده‌ها به سمت چپ یا راست متمایل باشد (چولگی به راست و چپ)، در این صورت استفاده از پارامتری همانند میانگین نمی‌تواند معیار خوبی برای قرار دادن یک مشتری در یک دسته خاص باشد. نکته مهم دیگری که در اینجا باید به آن اشاره کرد این است که ممکن است برخی از مشتریان برخلاف بسیاری دیگر از مشتریان، حجم بالای خرید داشته باشند. این موضوع نیز به‌شدت بر پارامتری همانند میانگین، تأثیر منفی دارد؛ بنابراین پارامتری همانند میانه (عنصر وسط در یک لیست مرتب شده صعودی)، در این شرایط نتیجه بهتری را نسبت به میانگین ایجاد خواهد کرد.

نکته مهم دیگری که در این میان باید به آن اشاره کرد این است که حتی برای تعیین دامنه‌های تقسیم پارامترهای مدل LRFM به ۳ یا ۵ (یا هر تعداد دلخواه دیگر) نیز لازم است تا بیش از یک نقطه برای تقسیم داده‌ها به

بازه‌های مشخص مورد استفاده قرار بگیرد. با توجه به اینکه در مدل LRFM عموماً بازه‌های $n^{1/n}$ (n تعداد بازه‌های درخواستی است) مورد پذیرش هستند، در آخرین بخش پایانی روش پیشنهادی، استفاده از مقادیر چندک‌ها مورد توجه قرار گرفته است. اگر یک مجموعه را به شکل صعودی مرتب کنیم، در k امین چندک یک مجموعه، k درصد مقادیر کوچک آن مجموعه قرار دارند.

معروف‌ترین چندک‌های مورداستفاده در علم آمار، چارک‌های اول، دوم و سوم هستند که به ترتیب معادل صدک بیست‌وپنجم، میانه و صدک هفتادوپنجم می‌باشند (هان و همکاران، ۲۰۱۱). این سه مقدار، وضعیت پراکندگی داده‌ها را در یک مجموعه داده با چولگی به‌خوبی نشان می‌دهند. براساس یک قاعده ساده، داده‌های یک مجموعه که در فاصله‌ای به اندازه ۱/۵ برابر اندازه فاصله چارک اول و سوم کمتر از چارک اول یا ۱/۵ برابر اندازه فاصله چارک اول و سوم بیشتر از چارک سوم باشند، داده‌های پرت آن مجموعه در نظر گرفته می‌شوند.

در روش پیشنهادی، از مقادیر چندک‌ها برای تعیین دسته مشتریان استفاده شده است. با توجه به آنچه گفته شد مزیت این روش این است که علاوه بر دسته‌بندی مشتریان، می‌توان مشتریان با مقادیر LRFM بسیار بزرگ را که در این روش به عنوان مقادیر پرت شناسایی شده‌اند، شناسایی کرد و در آینده برای ترغیب آنها به ادامه خرید از مجموعه، امتیازهای خاصی را در نظر گرفت. همچنین به این ترتیب می‌توان براساس بودجه، طرح‌های تبلیغاتی متفاوتی را برای دسته‌های برتر مشتریان ایجاد نمود و به آنها اطلاع‌رسانی کرد.

تعیین کلاس مشتریان

یکی از مباحث جذاب در دنیای مدیریت روابط مشتریان این است که چگونه می‌توان به کمک داده‌های فعلی، مدلی را ساخت که به کمک آن بتوان مشتریان جدید را دسته‌بندی کرد. در این بخش به این موضوع خواهیم پرداخت. با انطباق مفهوم دسته مشتریان با مفهوم کلاس، می‌توان از مدل‌های کلاس‌بندی داده‌کاوی برای حل این مشکل بهره برد.

امروزه روش‌های بسیار متفاوتی برای فرایند کلاس‌بندی ارائه شده است که هرکدام از آنها از جنبه‌های مختلف سرعت، دقت، سهولت اجرا، قابلیت مقاومت و کار با داده‌های مختلف، داده‌های پرت و حتی داده‌های تهی، ویژگی‌های متفاوتی دارند. معروف‌ترین و پرکاربردترین کلاس‌بندی امروزی، کلاس‌بندهای درختی، کلاس‌بندهای رگرسیون و کلاس‌بندهای تلفیقی می‌باشند. کلاس‌بندهای درختی دارای ساختار درک ساده هستند. مدل‌های مبتنی بر رگرسیون عمدتاً به دلیل دقت بالا مورد توجه هستند و نهایتاً کلاس‌بندهای تلفیقی به دلیل تلفیق آرای چند کلاس‌بند و نتیجه-گیری براساس برایند نظرات آنها عمدتاً بهترین عملکرد را در بین روش‌های مختلف دارند (هان و همکاران، ۲۰۱۱). با این وجود کلاس‌بندی مبتنی بر شبکه عصبی یا قانون بیز نیز کاربردهای خود را دارند.

یکی از مواردی که در مباحث کلاس‌بندی مورد توجه است، انتخاب یک زیرمجموعه از ویژگی‌های ممکن برای کلاس‌بندی و ساخت مدل‌های تصمیم‌گیری می‌باشد. در این حالت تنها برخی از ویژگی‌های ممکن انتخاب می‌شوند و به یک لیست نهایی منتقل می‌گردند و سایر ویژگی‌ها از مجموعه داده حذف می‌شوند. انتخاب ویژگی از دو منظر مهم می‌باشد: نخست آنکه با کاستن از حجم اطلاعات، سرعت و نیازمندی‌های پردازشی کم می‌شود. همچنین در موارد متعددی مشاهده شده است که پس از انتخاب ویژگی‌های مطلوب و حذف سایر ویژگی‌ها، دقت مدل‌های تصمیم گیری نیز افزایش می‌یابد.

اگرچه روش‌های متنوعی برای انتخاب ویژگی وجود دارد اما عمده این روش‌ها بر این حقیقت استوار است که اگر دو ویژگی با هم همبستگی مثبت/منفی داشته باشند (افزایش و کاهش مقدار هر دو ویژگی با/برعکس همدیگر اتفاق بیافتد) می‌توان یکی از آنها را حذف کرد و برعکس اگر هیچ وابستگی خاصی بین دو ویژگی وجود نداشته باشد، این دو ویژگی می‌توانند عوامل متمایزکننده خوبی برای افزایش دقت کلاس‌بند باشند.

در روش به‌کاررفته این مقاله برای کلاس‌بندی و ساخت مدل‌های تصمیم‌گیری، ابتدا ویژگی‌های غیرمرتبط از مجموعه داده حذف می‌گردند. در ادامه به کمک روش‌های انتخاب ویژگی، تنها ویژگی‌های مؤثر در ساخت مدل برای مرحله نهایی انتخاب می‌شوند و بقیه ویژگی‌ها حذف می‌شوند. در آخرین مرحله به کمک ویژگی‌های باقی‌مانده، مدل تصمیم‌گیری ساخته می‌شود و در نهایت این مدل امتحان می‌شود و از دقت آن برای تعیین کلاس سایر رکوردها استفاده می‌شود. شکل ۲ جریان کاری روش به‌کاررفته در این مقاله برای ساخت مدل تصمیم‌گیری را نشان می‌دهد.

شکل ۲. نمودار کلی روش کلاس‌بندی و ایجاد مدل تصمیم‌گیرنده در خصوص تعیین کلاس یک مشتری

در روش مورداستفاده این بخش، منظور از ویژگی‌های غیرمرتبط، آن دسته از ویژگی‌ها هستند که وجود یا نبود آنها تأثیری در روند تصمیم‌گیری ندارد. برای مثال در بسیاری از مجموعه‌های داده، ویژگی‌های ردیف یا شناسه کاربر وجود دارد که فاقد هرگونه اطلاعات کاربردی برای یک سیستم تصمیم‌گیری می‌باشند. کشف و حذف این دسته از ویژگی‌ها متناسب با مجموعه داده به‌کار رفته و دانسته‌های موجود در خصوص هریک از ویژگی‌های آنهاست.

همچنین در این مقاله به‌منظور انتخاب ویژگی‌های مفید از معیار پیرسون[1] (هان و همکاران، ۲۰۱۱) استفاده شده است. معیار پیرسون، یک معیار تعیین شباهت براساس میزان همبستگی خطی بین دو متغیر تصادفی است. مقدار این معیار از تقسیم مقدار کواریانس دو متغیر بر مقدار انحراف معیار آنها به دست می‌آید. مقدار ضریب همبستگی پیرسون یک عدد بین ۱+ تا ۱- است که مقادیر مثبت به معنای همبستگی مثبت و مقادیر منفی به معنی همبستگی منفی بین مقادیر دو متغیر است. همچنین مقدار صفر به معنای نبودهمبستگی مقادیر دو متغیر می‌باشد. متغیرهای ناهمبسته حفظ می‌شوند و از بین هر جفت متغیر همبسته مثبت یا منفی، یکی حذف می‌شود.

[1] Pearson

نتایج و بحث

در تحقیق حاضر، از یک مجموعه داده شامل ۱۱۳۴ رکورد اطلاعاتی فروش محصولات شرکت سرچین خراسان استفاده شده است. شرکت سرچین خراسان در زمینه تولید رب و انواع کمپوت فعال می‌باشد. در ادامه این بخش ابتدا ویژگی‌های این مجموعه داده معرفی و سپس مراحل مختلف تحلیل داده‌های این مجموعه بررسی می‌گردد.

جدول ۱. فیلدهای مجموعه داده مورداستفاده

ردیف	نام ویژگی	شرح ویژگی	نوع ویژگی	دامنه مقادیر ویژگی
۱	ردیف	شماره ردیف رکورد	عددی	مقدار عددی بین ۱ تا ۱۱۳۴
۲	شناسه مشتری	شماره عددی اختصاصی هر مشتری	عددی	یک عدد صحیح ۴ یا ۵ رقمی
۳	جنسیت	جنسیت مشتری	دودویی	مرد یا زن
۴	سن	سن مشتری	عددی	عدد صحیح
۵	وضعیت ازدواج	وضعیت ازدواج مشتری	دودویی	متأهل یا مجرد
۶	دفعات مراجعه به فروشگاه	وضعیت مراجعه مشتری به فروشگاه	اسمی	مکرر، معمولی، به‌ندرت
۷	بازدید از محصولات تولیدی در ۱۲ ماه گذشته	آیا مشتری در طول یک سال گذشته از محصولات تولیدی شرکت بازدید کرده است؟	دودویی	بله یا خیر
۸	خرید از محصولات در ۱۲ ماه گذشته	آیا مشتری در طول یک سال گذشته خریدی انجام داده است؟	دودویی	بله یا خیر
۹	تعداد خرید	تعداد خریدهای یک مشتری	عددی	یک عدد صحیح صفر یا مثبت
۱۰	خرید محصولات مشابه در ۱۸ ماه گذشته	آیا مشتری محصولات مشابه شرکت را در ۱/۵ سال گذشته خریده است؟	دودویی	بله یا خیر
۱۱	خرید سایر محصولات	آیا مشتری سایر محصولات شرکت را خریده است؟	دودویی	بله یا خیر
۱۲	نحوه پرداخت	نحوه پرداخت پول توسط مشتری	اسمی	کارت بانکی، انتقال وجه از طریق بانک، پرداخت ماهیانه

مجموعه داده مورد استفاده

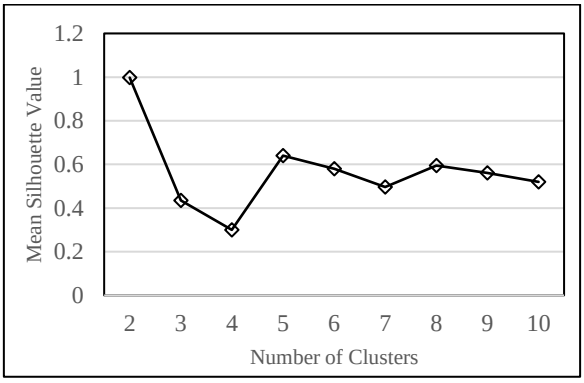
مجموعه داده‌های به‌کاررفته در این پژوهش دارای ۱۲ ویژگی است که هریک اطلاعات متفاوتی را در خصوص مشتریان ارائه می‌کنند. اطلاعات کامل این ویژگی‌ها در جدول ۱ آورده شده است.

دسته‌بندی مشتریان به کمک ترکیب K-Means و Silhouette

همان‌طور که در بخش ۲ این مقاله بیان شد روش K-Means روشی بسیار پرکاربرد و متداول در بین محققان برای دسته‌بندی مشتریان می‌باشد. برای مقایسه نتایج روش پیشنهادی، یک ساختار دسته‌بندی مبتنی بر K-Means نیز اجرا شد که جزئیات آن در ادامه این بخش ارائه شده است. همچنین با توجه به حساسیت روش K-Means به موضوع تعداد خوشه‌ها، مشابه پژوهش‌های پیشین نظیر (برهمانا و همکاران، ۲۰۲۰) و (کریستی و همکاران، ۲۰۲۱)، در این مقاله از پارامتر Silhouette برای این منظور استفاده شده است. مقدار Silhouette از طریق مقایسه میانگین فاصله تا عناصر در خوشه‌های مشابه با میانگین فاصله تا عناصر در خوشه‌های دیگر، محاسبه می‌شود. به منظور یافتن تعداد خوشه بهینه، کافی است تا به کمک یک حلقه، مقادیر مختلف تعداد خوشه به

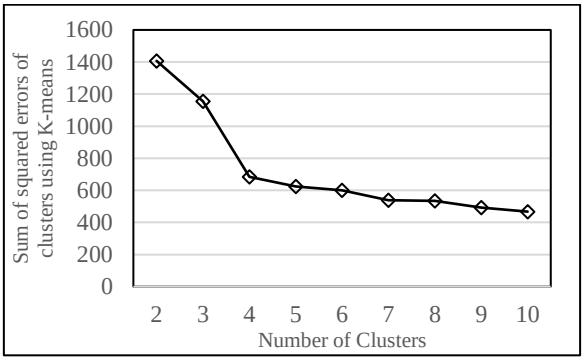
الگوریتم K-Means داده شود و نتیجه خوشه‌بندی این الگوریتم به کمک پارامتر Silhouette ارزیابی گردد. بهترین تعداد خوشه زمانی به‌دست می‌آید که مقدار Silhouette کمینه گردد. پس از یافتن تعداد خوشه‌های بهینه، کافی‌ست تا یک بار دیگر و این بار با تعیین دقیق تعداد خوشه‌ها، داده‌ها را خوشه‌بندی کرد و شماره هر خوشه (با شروع از یک) را به‌عنوان دسته مشتری در نظر بگیریم.

شکل ۳ نشان‌دهنده مقدار Silhouette برای تعداد مختلف خوشه، برروی مجموعه داده مورد بحث می‌باشد. همانطور که در این شکل نشان داده شده است، بهترین حالت خوشه‌بندی در زمانی به دست می‌آید که چهار خوشه داشته باشیم. در این حالت پارامتر Silhouette کمترین مقدار ممکن خود را دارد.



شکل ۳. ارزیابی تأثیر تعداد خوشه‌ها به کمک پارامتر Silhouette

به‌منظور ارزیابی صحت ادعایی طرح‌شده در خصوص دقت پارامتر Silhouette برای تعیین تعداد خوشه‌ها، در شکل ۴ میزان مجموع مربع خطای خوشه‌بندی به کمک روش K-Means به‌ازای تعداد خوشه‌های مختلف نشان داده شده است. همان‌طور که در این شکل دیده می‌شود، کاهش مقدار خطا تا عدد ۴ خوشه با شیب زیادی همراه است اما پس از آن کاهش مقدار این شیب بسیار ملایم می‌شود. در علم داده‌کاوی، این نقطه به نقطه آرنج معروف است و معرف تعداد خوشه‌های مناسب کار می‌باشد. شایان ذکر است که با افزایش تعداد خوشه‌ها، مقدار مربع خطاها در هر صورت کاهش می‌یابد که دلیل این امر کاهش فاصله عناصر یک خوشه از مرکز این خوشه، به دلیل کاسته شدن از ابعاد خوشه‌ها می‌باشد.



شکل ۴. ارزیابی مجموع مربع خطاهای دسته‌بندی به روش K-Means به ازای تعداد خوشه‌های مختلف

در نهایت به کمک این روش، مشتریان در قالب چهار خوشه مختلف دسته‌بندی می‌شوند. سپس شماره این خوشه‌ها (یک عدد صحیح بین ۱ تا ۴) به صورت یک ستون جدید با نام class به مجموعه داده اضافه می‌گردد تا در مرحله ساخت مدل تصمیم‌گیری مورد استفاده قرار بگیرد.

دسته‌بندی مشتریان به کمک روش پیشنهادی

همان‌طور که در بخش ۳ بیان گردید در روش پیشنهادی از مدل LRFM استفاده شده است. برای این منظور ابتدا باید پارامترهای مؤثر در محاسبه LRFM مشخص گردند. بررسی‌های مجموعه داده در دسترس برای مدل LRFM به شرح زیر است:

- برای ویژگی تازگی خرید ترکیبی، از ویژگی‌های خرید یک محصول یا محصولات مشابه استفاده شده است.
- برای ویژگی تعداد دفعات خرید، از ویژگی تعداد دفعات مراجعه به فروشگاه استفاده شده است.
- برای ویژگی ارزش پولی خرید، از تعداد کالاهای خریداری شده توسط هر مشتری استفاده شده است.
- برای ویژگی طول دوره همراهی مشتری، ویژگی بازدید از محصولات شرکت مورد استفاده قرار گرفته است.

فرمول LRFM به کار رفته در این مقاله به شرح زیر است:

$$RFM = F1*F2*F3*F4+F5 \quad (۱)$$

از آنجاکه برخی ویژگی‌های به‌کاررفته در رابطه فوق از نوع غیرعددی هستند، لازم است تا ابتدا معادل عددی آنها به‌کار گرفته شود. جزئیات تبدیل و مقادیر جدید پارامترهای F1 تا F5 در جدول ۲ ارائه شده است. در این رابطه، نظر به اینکه همراهی مشتری نیازمند بازدید وی از محصولات مختلف کارخانه است اما این بازدیدها الزاماً منجر به خرید یک محصول خاص نمی‌گردد، پارامتر F5 برخلاف سایر پارامترها به شکل حاصل‌جمع نشان داده شده است. پس از محاسبه مقدار LRFM هر رکورد اطلاعاتی به کمک رابطه ۱، نتایج حاصل شده به شرح زیر است:

- مقدار حداقل: ۲
- مقدار حداکثر: ۳۸
- مقدار چارک اول (صدک ۲۵): ۵
- مقدار چارک دوم (میانه): ۶
- مقدار چارک سوم (صدک ۷۵): ۱۰

جدول ۲. جزئیات تبدیل مقادیر ویژگی‌های به‌کاررفته در محاسبه LFRM

عبارت اختصاری	نام ویژگی	مقدار
F1	تعداد خرید	تا ۴ خرید عدد یک ۵ یا ۶ خرید عدد دو بیش از ۶ خرید عدد سه
F2	دفعات مراجعه به فروشگاه	«به‌ندرت» عدد یک «معمولی» عدد دو «مکرر» عدد سه
F3	خرید از محصولات در ۱۲ ماه گذشته	«بله» عدد دو «خیر» عدد یک
F4	خرید محصولات مشابه در ۱۸ ماه گذشته	«بله» عدد دو «خیر» عدد یک
F5	بازدید از محصولات تولیدی در ۱۲ ماه گذشته	«بله» عدد دو «خیر» عدد یک

همچنین براساس مقادیر چندک‌ها، مشتریان در چهار دسته مشتریان طلایی، نقره‌ای، برنزی و معمولی قرار داده شده‌اند. انتخاب عدد ۴ برای تعداد دسته‌های مشتریان به دلیل هم‌خوانی و قابلیت مقایسه نتایج این روش با نتایج ترکیب K-Means و Silhouette می‌باشد. مشتریان با مقدار LRFM کمتر یا مساوی ۵ (چارک اول) به‌عنوان مشتری معمولی، مشتریان با مقدار LRFM بزرگ‌تر از ۵ و کوچک‌تر مساوی ۶ (چارک دوم) به‌عنوان مشتریان برنزی، مشتریان با مقدار LRFM بزرگ‌تر از ۶ و کوچک‌تر مساوی ۱۰ (چارک سوم) به‌عنوان مشتریان نقره‌ای و سایر مشتریان (مقدار LRFM بزرگ‌تر از ۱۰) به‌عنوان مشتریان طلایی علامت خورده‌اند. در خروجی نهایی این مرحله، مقادیر LRFM و کلاس‌های هر رکورد در قالب دو ستون جدید به نام‌های LRFM و class به مجموعه داده اضافه گردید.

ارزیابی مدل‌های تصمیم‌گیری مبتنی بر کلاس‌بندی

پس از تعیین کلاس رکوردهای مختلف، نیازمند تعیین ویژگی‌های برتر هر کلاس هستیم که برای این منظور از نرم‌افزار IBM SPSS Modeler 18 استفاده شده است. نتایج این ارزیابی به کمک معیار پیرسون به‌منظور تعیین ویژگی‌های برتر در هر روش، به شرح زیر است:

- در خروجی حاصل از دسته‌بندی مشتریان به کمک روش ترکیب Silhouette و K-Means، ویژگی‌های سن، دفعات مراجعه به فروشگاه، خرید از محصولات در ۱۲ ماه گذشته، تعداد خرید، خرید محصولات

مشابه در ۱۸ ماه گذشته، خرید سایر محصولات و نحوه پرداخت را به همراه ویژگی اضافه شده کلاس به عنوان ویژگی‌های برتر انتخاب شده‌اند.

- در خروجی حاصل از روش پیشنهادی برای دسته‌بندی مشتریان، ویژگی‌های وضعیت ازدواج، دفعات مراجعه به فروشگاه، خرید از محصولات در ۱۲ ماه گذشته، تعداد خرید، خرید محصولات مشابه در ۱۸ ماه گذشته و نحوه پرداخت را به همراه ویژگی‌های اضافه شده LRFM و کلاس به عنوان ویژگی‌های برتر انتخاب شده‌اند.

در ادامه کار خروجی‌های فیلتر شده روش‌های مورد بحث به ۱۳ مدل کلاس‌بند مختلف موجود در SPSS Modeler وارد گردید. همچنین برای انجام مراحل اجرا و آزمون عملکرد تمامی کلاس‌بندها، داده‌ها به دو بخش آموزش (شامل ۶۶ درصد یا ۷۵۳ رکورد) و تست (شامل ۳۴ درصد یا ۳۸۱ رکورد) تقسیم شده‌اند.

تحلیل نتایج مدل‌های کلاس‌بندی نیازمند چهار تعریف پایه زیر می‌باشند (هان و همکاران، ۲۰۱۱):

- True Positive (TP): تعداد رکوردهای مثبتی که به‌درستی توسط کلاس‌بند برچسب کلاس مثبت را پذیرفته‌اند.
- True Negative (TN): تعداد رکوردهایی منفی که به‌درستی توسط کلاس‌بند برچسب کلاس منفی را پذیرفته‌اند.
- False Positive (FP): تعداد رکوردهای منفی که به‌اشتباه توسط کلاس‌بند برچسب کلاس مثبت را پذیرفته‌اند.
- False Negative (FN): تعداد رکوردهای مثبتی که به‌اشتباه توسط کلاس‌بند برچسب کلاس منفی را پذیرفته‌اند.

بر پایه این تعاریف، پارامترهای زیر به عنوان پارامترهای متداول برای ارزیابی نتایج روش‌های مختلف کلاس‌بندی مورد استفاده قرار می‌گیرند (هان و همکاران، ۲۰۱۱):

- **دقت[1] کلاس‌بند:** درصدی از رکوردهایی که به‌درستی در کلاس خودشان طبقه‌بندی شده‌اند.
- **نرخ خطا:** درصدی از رکوردها که به‌اشتباه در کلاسی غیر از کلاس خودشان طبقه‌بندی شده‌اند.
- **وضوح[2]:** درصدی از رکوردهای اعلام شده برای یک کلاس که به‌درستی متعلق به آن کلاس هستند.
- **فراخواندن[3]:** درصدی از کل رکوردهای یک کلاس که به‌درستی در آن کلاس طبقه‌بندی شده‌اند.
- **سنجه F:** میانگین هارمونیک وزن‌دار دو معیار وضوح و فراخواندن می‌باشد.

با توجه به اینکه در این مجموعه داده، چهار کلاس مختلف مشتریان تعریف شده است، مسئله ما یک مسئله چندکلاسه می‌باشد. با توجه به اینکه روابط پایه پارامترهای فوق، مختص مدل‌های دو کلاسه و نه چندکلاسه است، برای ارزیابی نتایج مدل‌های چند کلاسه، از روابط مقاله (سوکولووا و لاپالمه، ۲۰۰۹)[4] استفاده شده است. در مدل-های چند کلاسه دو رویه میکروسکوپیک و ماکروسکوپیک برای این منظور تعریف شده است. تفاوت این دو رویه در

[1] Accuracy
[2] Precision
[3] Recall
[4] Sokolova & Lapalme

این است که در مدل میکروسکوپیک میانگین تمام مقادیر نمونه‌ها در کلاس‌های مختلف مورد استفاده قرار می‌گیرد. اما در مدل ماکروسکوپیک میانگین مقادیر به‌دست‌آمده برای کلاس‌ها مورد استفاده قرار می‌گیرند. سکولوا و لاپالمی (سوکولووا و لاپالمه، ۲۰۰۹) پارامترهای ارزیابی فوق را به شکل زیر برای یک مجموعه داده با L کلاس مختلف فرموله کرده‌اند:

$$Average \quad Accuracy = \frac{\sum_{i=1}^{L} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{L} \tag{۲}$$

$$Error \quad Rate = \frac{\sum_{i=1}^{L} \frac{FP_i + FN_i}{TP_i + TN_i + FP_i + FN_i}}{L} \tag{۳}$$

$$\Pr ecision_{\mu} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} (TP_i + FP_i)} \tag{۴}$$

$$\mathrm{Re}\, call_{\mu} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} (TP_i + FN_i)} \tag{۵}$$

$$FScore_{\mu} = \frac{(\beta^2 + 1) \times \Pr esicion_{\mu} \times \mathrm{Re}\, call_{\mu}}{\beta^2 \times \Pr esicion_{\mu} + \mathrm{Re}\, call_{\mu}} \tag{۶}$$

$$\Pr ecision_{M} = \frac{\sum_{i=1}^{L} \frac{TP_i}{(TP_i + FP_i)}}{L} \tag{۷}$$

$$\mathrm{Re}\, call_{M} = \frac{\sum_{i=1}^{L} \frac{TP_i}{(TP_i + FN_i)}}{L} \tag{۸}$$

$$FScore_{M} = \frac{(\beta^2 + 1) \times \Pr esicion_{M} \times \mathrm{Re}\, call_{M}}{\beta^2 \times \Pr esicion_{M} + \mathrm{Re}\, call_{M}} \tag{۹}$$

در روابط فوق، روابط ۴ تا ۶ مربوط به مدل مکروسکوپیک و روابط ۷ تا ۹ مربوط به مدل ماکروسکوپیک در مدل‌های چند کلاسه می‌باشد. همچنین برای محاسبه سنجه F معمولاً برای β مقدار یک درنظر گرفته می‌شود که به

آن F1-Score نیز گفته می‌شود. شایان ذکر است که در روابط فوق به‌ازای هر کلاس باید درستی یا نادرستی تعلق یک رکورد به آن کلاس را به شکل مجزا از سایر کلاس‌ها محاسبه کرد.

از آنجاکه نرم‌افزار SPSS Modeler قادر به تحلیل کارایی مدل‌های چند کلاسه نیست؛ نتایج تعیین کلاس هر کلاس‌بند در قالب جداولی مجزا جمع‌آوری و به MATLAB وارد شده‌اند.

جداول ۳ و ۴ نتایج حاصل از دقت و عملکرد هریک از کلاس‌بندها برروی خروجی‌های دو روش بیان‌شده برای دسته‌بندی را نشان می‌دهند. همان‌طور که در این جدول دیده می‌شود، در مجموع کلاس‌بندهای مورداستفاده با خروجی‌های روش پیشنهادی دارای عملکرد بسیار بهتری بوده‌اند.

در حالی که در جدول ۳ بهترین دقت کلاس‌بندی مربوط به کلاس‌بند رگرسیون لجستیک می‌باشد که نزدیک به عدد ۱۰۰ می‌باشد، در نتایج ارائه‌شده در جدول ۴، شش کلاس‌بند توانسته‌اند برروی مقادیر حاصل از روش پیشنهادی به دقت ۱۰۰ درصد برسند. این برتری مشهود نشانگر آن است که عملکرد روش پیشنهادی برای خوشه‌بندی مشتریان کاملاً مؤثر بوده است. با این وجود، کمترین مقدار متوسط دقت در جدول ۳ برابر با ۹۲/۵۲ درصد (کلاس‌بند Tree-AS) است و این عدد برای جدول ۴ برابر با ۸۴/۹۱ درصد (کلاس‌بند بیز) می‌باشد.

جدول ۳. نتایج دقت کلاس‌بندهای مختلف برروی خروجی‌های روش K-Means با هدف دسته‌بندی مشتریان

ردیف	کلاس‌بند	متوسط دقت	نرخ خطا	وضوح (میکرو)	فراخواندن (میکرو)	سنجه F1 (میکرو)	وضوح (ماکرو)	فراخواندن (ماکرو)	سنجه F1 (ماکرو)
۱	رگرسیون لجستیک	۰.۹۹۷۴	۰.۰۰۲۶	۰.۹۹۴۸	۰.۹۹۴۸	۰.۹۹۴۸	۰.۹۹۶۱	۰.۹۹۶۱	۰.۹۹۶۱
۲	CHAID	۰.۹۸۶۹	۰.۰۱۳۱	۰.۹۷۳۸	۰.۹۷۳۸	۰.۹۷۳۸	۰.۰۰۰۰	۰.۷۳۴۹	۰.۰۰۰۰
۳	C۵.۰	۰.۹۹۴۸	۰.۰۰۵۲	۰.۹۸۹۵	۰.۹۸۹۵	۰.۹۸۹۵	۰.۹۹۱۶	۰.۸۷۰۰	۰.۹۲۶۸
۴	جنگل تصادفی	۰.۹۹۰۸	۰.۰۰۹۲	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۶۴	۰.۸۶۴۴	۰.۹۲۱۴
۵	Tree-AS	۰.۹۲۵۲	۰.۰۷۴۸	۰.۸۵۰۴	۰.۸۵۰۴	۰.۸۵۰۴	۰.۰۰۰۰	۰.۶۴۱۲	۰.۰۰۰۰
۶	درخت C&R	۰.۹۹۳۴	۰.۰۰۶۶	۰.۹۸۶۹	۰.۹۸۶۹	۰.۹۸۶۹	۰.۰۰۰۰	۰.۷۴۵۰	۰.۰۰۰۰
۷	شبکه عصبی	۰.۹۹۰۸	۰.۰۰۹۲	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۶۴	۰.۸۶۴۴	۰.۹۲۱۴
۸	Discriminant	۰.۹۹۶۱	۰.۰۰۳۹	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۳۶	۰.۹۹۵۰	۰.۹۹۴۳
۹	SVM	۰.۹۶۷۲	۰.۰۳۲۸	۰.۹۳۴۴	۰.۹۳۴۴	۰.۹۳۴۴	۰.۹۵۰۶	۰.۹۵۳۱	۰.۹۵۱۹
۱۰	LSVM	۰.۹۸۹۵	۰.۰۱۰۵	۰.۹۷۹۰	۰.۹۷۹۰	۰.۹۷۹۰	۰.۹۸۴۸	۰.۹۸۴۵	۰.۹۸۴۶
۱۱	شبکه بیز	۰.۹۳۴۴	۰.۰۶۵۶	۰.۸۶۸۸	۰.۸۶۸۸	۰.۸۶۸۸	۰.۹۰۶۱	۰.۹۱۳۹	۰.۹۱۰۰
۱۲	KNN	۰.۹۷۷۷	۰.۰۲۲۳	۰.۹۵۵۴	۰.۹۵۵۴	۰.۹۵۵۴	۰.۹۶۷۰	۰.۹۶۷۷	۰.۹۶۷۳
۱۳	تلفیقی	۰.۹۹۶۱	۰.۰۰۳۹	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۳۶	۰.۹۹۵۰	۰.۹۹۴۳
۱۴	میانگین نتایج	۰.۹۸۰۰	۰.۰۲۰۰	۰.۹۶۰۰	۰.۹۶۰۰	۰.۹۶۰۰	۰.۷۵۰۵	۰.۸۸۶۵	۰.۷۳۶۰

همچنین میانگین نتایج به‌ازای پارامترهای میکروسکوپیک در جدول ۳ بالاتر از مقادیر مشابه در جدول ۴ هستند. این در حالی است که در پارامترهای ماکروسکوپیک رویه مشاهده‌شده، کاملاً برعکس است و نتایج حاصل از روش پیشنهادی در جدول ۴ برتری محسوسی را بر نتایج جدول ۳ دارند. این موضوع نشانگر این است که روش ترکیب K-Means و Silhouette بر بهبود نتیجه تشخیص انفرادی کلاس هر رکورد متمرکز است و در مقابل روش پیشنهادی زمینه‌ساز بهبود نتیجه تشخیص گروهی رکوردها می‌باشد. از این مشاهدات می‌توان برای تعیین روش موردنظر استفاده کرد. رویکرد ارزیابی انتخاب شده می‌تواند ابزاری برای تعیین اهداف سازمانی برای CRM و نحوه تعامل با مشتریان باشد.

جدول ۴. نتایج دقت کلاس‌بندهای مختلف برروی خروجی‌های روش پیشنهادی به‌منظور دسته‌بندی مشتریان

ردیف	کلاس‌بند	متوسط دقت	نرخ خطا	وضوح (میکرو)	فراخواندن (میکرو)	سنجه F1 (میکرو)	وضوح (ماکرو)	فراخواندن (ماکرو)	سنجه F1 (ماکرو)
۱	رگرسیون لجستیک	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۲	CHAID	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۳	C۵.۰	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۴	جنگل تصادفی	۰.۹۹۶۱	۰.۰۰۳۹	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۲۱	۰.۹۹۲۸	۰.۹۹۲۲	۰.۹۹۲۵
۵	Tree-AS	۰.۹۶۰۶	۰.۰۳۹۴	۰.۹۲۱۳	۰.۹۲۱۳	۰.۹۲۱۳	۰.۹۲۸۸	۰.۹۲۳۰	۰.۹۲۵۹
۶	درخت C&R	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۷	شبکه عصبی	۰.۹۹۷۴	۰.۰۰۲۶	۰.۹۹۴۸	۰.۹۹۴۸	۰.۹۹۴۸	۰.۹۹۵۱	۰.۹۹۴۸	۰.۹۹۵۰
۸	Discriminant	۰.۹۸۵۶	۰.۰۱۴۴	۰.۹۷۱۱	۰.۹۷۱۱	۰.۹۷۱۱	۰.۹۷۳۵	۰.۹۷۰۳	۰.۹۷۱۹
۹	SVM	۰.۹۴۴۹	۰.۰۵۵۱	۰.۸۸۹۸	۰.۸۸۹۸	۰.۸۸۹۸	۰.۸۹۴۴	۰.۸۹۴۰	۰.۸۹۴۲
۱۰	LSVM	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۱۱	شبکه بیز	۰.۸۴۹۱	۰.۱۵۰۹	۰.۶۹۸۲	۰.۶۹۸۲	۰.۶۹۸۲	۰.۷۴۵۲	۰.۶۹۴۴	۰.۷۱۸۹
۱۲	KNN	۰.۹۹۰۸	۰.۰۰۹۲	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۱۶	۰.۹۸۲۲	۰.۹۸۱۴	۰.۹۸۱۸
۱۳	تلفیقی	۱.۰۰۰۰	۰.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰	۱.۰۰۰۰
۱۴	میانگین نتایج	۰.۹۷۸۸	۰.۰۲۱۲	۰.۹۵۷۶	۰.۹۵۷۶	۰.۹۵۷۶	۰.۹۶۲۵	۰.۹۵۷۷	۰.۹۶۰۰

نتیجه‌گیری

نظر به اهمیت دسته‌بندی و تعیین دسته مشتریان متفاوت در صنایع مختلف، ارائه مدل‌های کارآمد در این زمینه مورد توجه محققان قرار دارد. در این مقاله یک روش جدید برای دسته‌بندی مشتریان ارائه گردید و سپس براساس خروجی‌های این روش، یک روش برای کلاس‌بندی و پیش‌بینی دسته هر مشتری ارائه شد. نتایج نشانگر آنست که خروجی‌های روش پیشنهادی نه‌تنها در ارزیابی انفرادی کلاس‌بندها عملکرد بهتری داشته‌اند بلکه در عملکرد تشخیص

گروهی مشتریان نیز موفق‌تر هستند. استفاده از مدل‌های دقیق‌تر خوشه‌بندی می‌تواند به یک عنوان یک کار آتی برای دسته‌بندی بهتر محتواهای اطلاعاتی مربوط به مشتریان مورد توجه قرار بگیرد.

References

Alizadeh Zoeram, A., & Karimi Mazidi, A. R. (2018). New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System. *Iranian Journal of Management Studies*, *11*(2), 351-378. https://doi.org/10.22059/ijms.2018.242528.672839

Babaiyan, V., & Sarfarazi, S. A. (2019). Analyzing Customers of South Khorasan Telecommunication Company with Expansion of RFM to LRFM Model. *Journal of Artificial Intelligence and Data Mining*, *7*(2), 331-340. https://doi.org/10.22044/jadm.2018.6035.1715

Brahmana, R. W. S., Mohammed, F. A., & Chairuang, K. (2020). Customer segmentation based on RFM model using K-means, K-medoids, and DBSCAN methods. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, *11*(1), 32-43. https://doi.org/10.24843/LKJITI.2020.v11.i01.p04

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252-268. https://doi.org/10.1016/j.ejor.2003.12.010

Chang, H. H., & Tsay, S. F. (2004). Integrating of SOM and K-mean in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal Of Information Management*, *11*(4), 161-203. https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=387779

Cheng, C-H., & Chen, Y-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*(3), 4176-4184. https://doi.org/10.1016/j.eswa.2008.04.003

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, *33*(10), 1251-1257. https://doi.org/10.1016/j.jksuci.2018.09.004

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224-227. https://doi.org/10.1109/TPAMI.1979.4766909

Delangizan, S., Papzan, A., & Armand, S. (2022). Design and Development of a Model for Commercialization of Organic Products Based on Fundamental Theory (Case Study: Kermanshah Province). *Karafan Quarterly Scientific Journal*, *18*(4), 33-48. https://doi.org/10.48301/kssa.2021.271211.1375

Dogan, O., Ayçin, E., & Bulut, Z. A. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, *8*(1), 1-19. https://avesis.kocaeli.edu.tr/yayin/894de1af-d068-4e33-ad18-d0d727c24fbe/customer-segmentation-by-using-rfm-model-and-clustering-methods-a-case-study-in-retail-industry

Edelstein, H. (2001). Building profitable customer relationships with data mining. In *Customer Relationship Management: The Ultimate Guide to the Efficient Use of CRM*. Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-84961-8_26

Ernawati, E., Baharin, S., & Kasmin, F. (2021, November 28). *A review of data mining methods in RFM-based customer segmentation*. 2nd Annual Conference of Science and Technology, Malang, Indonesia. https://doi.org/10.1088/1742-6596/1869/1/012085

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques* (3 ed.). Morgan Kaufmann. https://www.amazon.com/Data-Mining-Concepts-Techniques-Management/dp/0123814790

Horrei, H. (2014). *Apply Data Mining in Customer Relationship Management*. Tebyan. https://library.tebyan.net/en/165757/

Huang, Y., Zhang, M., & He, Y. (2020, June 19-21). *Research on improved RFM customer segmentation model based on K-Means algorithm.* 2020 5th International Conference on Computational Intelligence and Applications, Beijing, China. https://doi.org/10.1109/ICCIA49625.2020.00012

Hughes, A. M. (1994). *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program*. Probus Publishing Company. https://www.amazon.com/Strategic-Database-Marketing-Masterplan-Customer-Based/dp/1557385513

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, *26*(2), 181-188. https://doi.org/10.1016/S0957-4174(03)00133-7

Jonker, J-J., Piersma, N., & Van den Poel, D. (2004). Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, *27*(2), 159-168. https://doi.org/10.1016/j.eswa.2004.01.010

Kao, Y-T., Wu, H-H., Chen, H-K., & Chang, E-C. (2011). A case study of applying LRFM model and clustering techniques to evaluate customer values. *Journal of Statistics and Management Systems*, *14*(2), 267-276. https://doi.org/10.1080/09720510.2011.10701555

Kim, S-Y., Jung, T-S., Suh, E-H., & Hwang, H-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*(1), 101-107. https://doi.org/10.1016/j.eswa.2005.09.004

Kumar, V., & Reinartz, W. J. (2005). *Customer relationship management: A databased approach*. Wiley. https://www.amazon.com/Customer-Relationship-Management-Databased-Approach/dp/0471271330

Liu, D-R., & Shih, Y-Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, *42*(3), 387-400. https://doi.org/10.1016/j.im.2004.01.008

Maimon, O., & Rokach, L. (2010). *Data mining and knowledge discovery handbook* (2, Ed.). Springer New York. https://doi.org/10.1007/978-0-387-09823-4

Marisa, F., Ahmad, S. S. S., Yusof, Z. I. M., Hunaini, F., & Aziz, T .M. A. (2019). Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. *International Journal of Integrated Engineering*, *11*(3), 169-180. https://doi.org/10.30880/ijie.2019.11.03.018

Maryani, I., Riana, D., Astuti, R. D., Ishaq, A., & Pratama, E. A. (2018, October 17-18). *Customer Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm*. 2018 Third International Conference on Informatics and Computing, Palembang, Indonesia https://doi.org/10.1109/IAC.2018.8780570

Saeida Ardakani, S., Konjkav Monfared, A., & Hosseini Tavabe, Z. (2021). A Model of Customer's Stickiness in Online Retail with Emphasis of Their Perceived Value (Case study: DigiKala). *Karafan Quarterly Scientific Journal*, *19*(5), 573-595. https://doi.org/10.48301/kssa.2021.287817.1546

Seyedhosseini, S., Gholamian, M., & Maleki, A. (2011). A methodology based on RFM using data mining approach to assess the customer loyalty. *International Journal of Industrial Engineering*, *22*(2), 171-179. http://ijiepm.iust.ac.ir/article-1-661-en.html

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Tale, A., Khani, H., Zendehdel, T., Hajtalebi, H., Behzadfar, F., & Razaqiazar Azar, M. (2017). *SPSS Software Applied Training*. Vandad. https://www.gisoom.com/book/11321411 /

Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018, October 12-14). *Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study.* 2018 IEEE 15th International Conference on e-Business Engineering Xi'an, China https://doi.org/10.1109/ICEBE.2018.00027

Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, *27*(2), 265-276. https://doi.org/10.1016/j.eswa.2004.02.005

Veisi, H., & Gharavi, E. (2015) .Data mining based approach in customer relationship management and marketing. *Consumer Behavior Studies Journal*, *2*(2), 25-44. https://cbs.uok.ac.ir/article_7657.html?lang=en

Vohra, R., Pahareeya, J., Hussain, A., Ghali, F., & Lui, A. (2020). Using Self Organizing Maps and K Means Clustering Based on RFM Model for Customer Segmentation in the Online Retail Business. In D. Huang & P. Premaratne (Eds.), *Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part III*. Springer Cham. https://doi.org/10.1007/978-3-030-60796-8_42

Wu, J., Shi, L., Lin, W-P., Tsai, S-B., Li, Y., Yang, L., & Xu, G. (2020). An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Means Algorithm. *Mathematical Problems in Engineering*, *2020*, 1-7. https://doi.org/10.1155/2020/8884227

Zalaghi, Z., & Abbasnejad Varzi, Y. (2014). Measuring customer loyalty using an extended RFM and clustering technique. *Management Science Letters*, *4*(5), 905-912. https://doi.org/10.5267/j.msl.2014.3.026