



<https://nmrj.ui.ac.ir/?lang=en>
New Marketing Reserch Journal
E-ISSN: 2228- 7744
Vol. 13, Issue 2, No.49, Summer 2023
Document Type: Research Paper
Received: 16/11/2022 Accepted: 11/07/2023

Customer Churn Analysis Based on the Data-mining Approach: Hybrid Algorithm Incorporates Decision Tree and Bayesian Network

Atefeh Aghakhani Bezdi Langari

MA Student, Department of Industrial Engineering and Management, Shahrood University of Technology,
Shahrood, Iran
aqakhani396@gmail.com

Aliakbar Hasani  *

Associate Professor, Department of Industrial Engineering and Management, Shahrood University of
TechnologyShahrood, Iran
aa.hasani@shahroodut.ac.ir

Abstract

Today, companies and organizations are aware of the fact that customer retention leads to greater profitability. Increasing competition causes the rate of customer churn to grow. Therefore, studying the features influencing the tendency of customer churn is important. In the present study, a hybrid model based on the data mining approach is presented to analyze the features of churn customers. In the first step, the feature selection node has been used to identify the features with higher importance and remove redundant items. Then, the C5.0 Decision Tree and Bayesian network were used to classify the customers into two groups, turning customers, and non-turning customers. These are data mining techniques and terms that can help in forecasting. Finally, the proposed model has been implemented in the chain store industry as a case study. Key findings indicate that both the decision tree model and the Bayesian network can predict churn customers with different accuracies, the area under the

*Corresponding author



receiver operating curve in the decision tree model is greater than the Bayesian network model and it has better performance. The results indicate the optimal efficiency of the proposed method. In addition, the results show that three features of gender, marital status, and age from the set of demographic characteristics and five factors of average monthly income level, number of purchases per month, the share of online shopping, how to get to know the store, type of market-related to customer transaction records are among the most effective factors.

Introduction

Customers are among the most important assets for businesses. Customer relationship management has been introduced as a comprehensive key strategy to stay focused on customers' needs and integrate methods of dealing with customers in the organization. In the key area of customer relationship management, the importance of customer orientation is very evident. This term means leaving the organization on behalf of customers and turning to competing organizations to receive services. Customers may leave the organization for various obvious or hidden reasons. The goal of organizations is to maintain existing customers by using customer retention methods. Customer loss creates a situation for competing organizations to attract customers of an organization. This situation has made the importance of predicting the churn of customers to double. Researchers have concluded that a small change in customer retention rate can have a huge impact on overall business improvement. Predicting customer churn is a suitable tool for describing the customer retention process of an organization, and the purpose of using it is to identify a group of customers who are prone to churn. Knowing this group of customers and taking preventive measures can play a very important role in preventing customers from turning away. In such a situation, predicting customer turnover has attracted a lot of attention in management and marketing studies. To efficiently manage customer churn forecasting within an organization, it is of great importance to provide an effective and highly accurate customer churn prediction model.

Methodology

The research is practical according to the purpose. Because the results of this research can be practically used, the nature of this research is post-event. Considering the existence of three types of research methods: quantitative, qualitative, and hybrid, the method of this research is quantitative. Since a specific methodology must be used to perform data mining operations, standard Crisp methodology has been used in this regard. Also, various data mining techniques including feature selection and classification have been used in this research. Finally, a case study is the method used in this research. In this research, based on the real data of the customers of the chain store industry, customer churn is predicted and the

effective factors of churn are identified. In this research, a hybrid model based on a data mining approach is presented to analyze the factors of customer churn. In the first step, the Bayesian network algorithm was used to identify factors with higher importance. Bayesian network is a non-circular directed graph where each node represents a variable and arcs represent direct causal relationships between connected nodes and conditional probability tables are assigned to nodes that have conditional dependence. In the second step, the C5.0 decision tree technique was used to classify customers based on their churning status. The C5.0 decision tree is an improvement over the C4.5 and ID3 decision trees. The division of each node is calculated based on information gain. This index is used to select the fragile variable in the process of tree growth.

Findings

The results of the research indicate that the demographic characteristics and purchase records of customers are effective in the behavior of customers. Based on the results of the classification in this research and the rules provided by the decision tree, the eight key factors identified in turning away customers with a significant effect on the studied case have been analyzed in the following. The classification results show that people with a monthly income of 50 million Rials and above are among those customers who are likely to buy again. People with a monthly income of less than 50 million Rials are not customers. People whose number of purchases was more than 2 times a month are among those loyal customers who are likely to buy again, and those customers who are likely to buy less than 2 times would not repurchase and they are part of the churning customers. The obtained results show that people with an internet shopping share of more than 20% are loyal customers and people with a lower shopping share are turned away (churn) customers. The percentage of women's purchases is higher than that of men, and women are more loyal than men. Married people have a higher percentage of the store's customers and are more loyal than single people. The overall amount of people's purchases from a chain store is much higher than from a traditional supermarket.

Conclusions


Analyzing and predicting customer behavior is very important because the cost of losing a customer is very high for an organization. In this regard, in the current research, the method of combining the Bayesian network and a C5.0 decision tree has been developed for predicting the analysis of customer churn. For this purpose, the data of customers of a chain store in Mashhad City has been examined as a case study. The variable of customer repurchase probability is considered as a dependent variable and then the most important independent variables for the implementation of the C5.0 decision tree and Bayesian network are identified by the feature selection node. The results of the present research show that the application of

the feature selection algorithm can help the decision makers to accurately classify the model and choose the best model and focus on the variables with the highest importance in predicting the turning of customers. The results also indicate that the eight factors of age, marital status, average monthly income, number of purchases per month, familiarity with the store, type of market, share of online shopping, and special sales are among the most important factors affecting diversion. According to the comparison of two Bayesian network algorithms and the C5.0 decision tree based on ROC diagram results, it is emphasized that the C5.0 decision tree with the highest accuracy has a better performance in identifying returning customers. Finally, a set of managerial insights for formulating marketing plans and facing all kinds of customers has been presented.

Keywords: Customer Churn, C5.0 Decision Tree, Bayesian Network, Data Mining, Machine Learning.

مقاله پژوهشی

تحلیل روی گردانی مشتریان مبتنی بر رویکرد داده کاوی: الگوریتم ترکیبی درخت تصمیم و شبکه بیزین (مورد مطالعه: فروشگاه‌های زنجیره‌ای)

عاطفه آقاخانی بزدی لنگری^۱، علی اکبر حسنی^{۲*} 

۱- دانشجوی کارشناسی ارشد مدیریت صنعتی، دانشکده مهندسی صنایع و مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران

۲- دانشیار گروه مهندسی صنایع، گروه مهندسی صنایع، دانشکده مهندسی صنایع و مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران

aa.hasani@shahroodut.ac.ir

چکیده

امروزه سازمان‌ها به این آگاهی رسیده‌اند که حفظ مشتریان باعث سودآوری بیشتر می‌شود. همچنین، افزایش رقابت نیز باعث می‌شود تا میزان روی گردانی مشتریان افزایش یابد؛ از این رو مطالعه عوامل مؤثر بر تمایل به روی گردانی یا عدم روی گردانی مشتریان برای پژوهشگران و فعالان کسب و کارها اهمیت دارد. در پژوهش حاضر یک مدل ترکیبی مبتنی بر رویکرد داده کاوی برای تحلیل عوامل روی گردانی مشتریان ارائه شده است. در گام نخست برای شناسایی عوامل با درجه اهمیت زیادتر و حذف موارد زائد از گره انتخاب ویژگی استفاده و در گام دوم نیز برای طبقه‌بندی و پیش‌بینی مشتریان به دو دسته مشتریان روی گردان و مشتریان غیر روی گردان از درخت تصمیم C5.0 و شبکه بیزین استفاده شده است. در نهایت، مدل پیشنهادی در صنعت فروشگاه‌های زنجیره‌ای به‌عنوان مطالعه موردی پیاده‌سازی شده است. یافته‌های پژوهش حاکی از آن است که هر دو مدل درخت تصمیم و شبکه بیزین توانایی مناسب را برای پیش‌بینی روی گردانی مشتریان دارد و سطح زیر منحنی عملیاتی گیرنده در مدل درخت تصمیم بیشتر از مدل شبکه بیزین بوده است؛ در نتیجه مدل درخت تصمیم عملکرد بهتری دارد. همچنین، نتایج نشان می‌دهد که سه عامل جنسیت، وضعیت تأهل و سن از مجموعه مشخصه‌های دموگرافیک و پنج عامل متوسط سطح درآمد ماهیانه، تعداد خرید در ماه، سهم خرید اینترنتی، نحوه آشنایی با فروشگاه و نوع بازار مربوط به سوابق تراکنش‌های مشتریان از مهم‌ترین عوامل مؤثر بر روی گردانی مشتریان است.

کلیدواژه‌ها: روی گردانی مشتری، داده کاوی، الگوریتم درخت تصمیم C5.0، الگوریتم شبکه بیزین، یادگیری ماشین.

۱. مقدمه

مشتریان از جمله مهم‌ترین دارایی‌ها برای کسب و کارها هستند. در دهه‌های اخیر، به دلیل اهمیت مشتریان در کسب و کار، تمرکز بسیاری از سازمان‌ها از محصول‌گرایی به مشتری‌مداری بوده است (Ye et al., 2013). شناسایی گروه‌های مختلفی از مشتریان و ایجاد طرح‌های بازاریابی، فروش و خدمات مناسب با نیازها و ویژگی‌های هر گروه، از جمله هدف‌های مهم مدیریت ارتباط با مشتری است (Kotler, 2000). مدیریت ارتباط با مشتری به عنوان یک استراتژی کلیدی جامع برای تمرکز بر نیازهای مشتریان و یکپارچه‌سازی روش‌های برخورد با مشتریان در سازمان معرفی شده است (Brown & Cooper, 1999). در صورتی که سازمان‌ها قادر باشند میزان نگهداری مشتریان خود را به ۵ درصد افزایش دهند، سود کسب‌شده برای سازمان‌ها بین ۲۵ تا ۸۵ درصد افزایش پیدا می‌کند. در این حالت، مشتریان فعلی خرید بیشتری را در زمان کمتری انجام می‌دهند. همچنین، اگر مشتریان فعلی به نوسان‌های قیمت حساسیت کمتری را نشان دهند در این صورت می‌توانند تبلیغی برای جذب مشتریان جدید باشند (Ganesh et al., 2000). در حوزه کلیدی مدیریت ارتباط با مشتری، اهمیت موضوع روی گردانی مشتریان بسیار نمایان است. در واقع، این اصطلاح به معنای ترک سازمان از جانب مشتریان و مراجعه به سازمان‌های رقیب برای دریافت خدمات است. مشتریان ممکن است به دلایل گوناگون آشکار یا پنهان سازمان ارائه‌دهنده خدمت یا محصول را ترک کنند. در واقع، روی گردانی مشتری نقش بسیار مهمی در بازارهای اشباع‌شده امروزی دارد. هر سازمان برای پیشتازی خود در بازار، کسب سود و مدیریت هزینه‌های خود تلاش می‌کند. این در حالی است که جذب مشتریان جدید پنج تا شش برابر گران‌تر از حفظ مشتریان موجود است

(Chung et al., 2016)؛ بنابراین هدف سازمان‌ها بر نگهداری مشتریان موجود با استفاده از روش‌های علمی و کارآمد حفظ مشتری متمرکز است. در واقع، از دست دادن مشتری فعلی منجر به ایجاد جذب مشتریان برای رقبا می‌شود (Lin et al., 2011). این وضعیت باعث شده است تا اهمیت پیش‌بینی روی گردانی مشتریان سازمان دوچندان شود. محققان به این نتیجه دست یافته‌اند که اندکی تغییر در میزان نگهداری مشتری می‌تواند بر بهبود کلی کسب و کار تأثیر بسیار زیادی بگذارد. پیش‌بینی روی گردانی مشتریان ابزاری مناسب برای توصیف فرآیند نگهداری مشتریان یک سازمان است که هدف از به کارگیری آن شناسایی گروهی از مشتریان است که مستعد روی گردانی هستند. با شناسایی این گروه از مشتریان و انجام دادن اقدام‌های پیشگیرانه می‌توان نقش بسیار مهمی در جلوگیری از روی گردانی مشتریان ایفا کرد. در چنین شرایطی، پیش‌بینی روی گردانی مشتریان توجه زیادی را در مطالعات مدیریت و بازاریابی به خود جلب کرده است. برای مدیریت کارآمد پیش‌بینی روی گردانی مشتری در درون یک سازمان، ارائه مدل پیش‌بینی اثربخش با صحت زیاد روی گردانی مشتری اهمیت فراوانی دارد. مدل‌های پیش‌بینی کننده متعددی برای این مسئله ارائه شده است. در این میان، تکنیک‌های داده‌کاوی به‌طور مؤثری، قادر به شناسایی مشتریان متمایل به روی گردانی است. این تکنیک‌ها قابلیت کشف الگوها و ارتباط‌های درونی داده‌ها را دارند. همچنین، می‌توانند به دسته‌بندی و پیش‌بینی رفتار مدل براساس داده‌های در دسترس پردازند. به عبارت دیگر، داده‌کاوی یک مبحث میان‌رشته‌ای است که با به کارگیری الگوریتم‌های مختلف به کشف الگوهای پنهانی مجموعه داده‌های گسترده می‌پردازد (Tsai & Lu, 2009). در صنعت فروشگاه‌های زنجیره‌ای نیز به دلیل افزایش رقابت و توجه

ویژه مشتریان به کیفیت و قیمت کالا، مسئله تحلیل روی گردانی مشتریان اهمیت فراوانی پیدا کرده است. بدیهی است که در این فضای رقابتی و هزینه اندک تغییر فروشگاه، مشتریان به راحتی فروشگاه مدنظر خود را به یک فروشگاه دیگر با خدمات و محصولات بهتر تغییر می دهند. روی گردانی مشتریان از یک فروشگاه به معنای کاهش سود عملیاتی است. به همین دلیل، فروشگاه‌ها باید تلاش کنند تا مشتریان با ارزش فعلی خود را حفظ کنند؛ بنابراین هدف از پژوهش حاضر ارائه مدلی کارآمد برای پیش بینی احتمال روی گردانی مشتریان فروشگاه زنجیره‌ای و همچنین، ارائه پیشنهادهایی اثربخش برای کاهش روی گردانی مشتریان است.

۲. پیشینه پژوهش

در این بخش مطالعات انجام شده در حوزه تحلیل روی گردانی مشتریان از منظر معیارها و روش تحلیل مرور می شود. در پیشینه موضوع، روی گردانی مشتری به فرآیندی اطلاق می شود که مشتریان از یک ارائه دهنده خدمات به ارائه دهنده دیگری تغییر جهت می دهند؛ از این رو تکرار نشدن خرید به منزله روی گردانی از خدمت دهنده خواهد بود (Shobana et al., 2023). روی گردانی مشتری منجر به از دست دادن مشتریان بالقوه و کاهش درآمد سازمان خواهد شد. در این راستا، شناسایی و تحلیل علت‌های روی گردانی مشتریان اهمیت دارد و سازمان را قادر می کند تا خدمات خود را برای پاسخگویی بهتر به نیازهای مشتریان بهبود دهد. به طور کلی، روی گردانی مشتری به دو دسته اصلی اختیاری و اجباری تقسیم بندی می شود. شناسایی وضعیت روی گردانی اجباری راحت تر از وضعیت اختیاری است؛ زیرا مشتریانی را در بر می گیرد که به دلیل پاسخگو نبودن به موقع و مناسب سازمان، تصمیم به قطع ارتباط گرفته‌اند. این در حالی است که

شناسایی روی گردانی اختیاری در دو نوع عمدی و تصادفی و زمان رخداد آن برای سازمان بسیار دشوار است. روی گردانی اختیاری تصادفی می تواند به دلایل پیش بینی نشده همچون تغییرات شرایط مالی یا مکانی مشتری رخ دهد. روی گردانی اختیاری عمدی نیز تحت تأثیر عواملی همچون کیفیت سطح خدمات دریافت شده در مقابل رقبای، تحولات فناورانه و مسائل مشابه خواهد بود. شناسایی و تحلیل عوامل مؤثر بر روی گردانی یک مسئله پیچیده و تأثیر گذار بر مدیریت عملکرد سازمان‌هاست. به طور کلی، عوامل مؤثر بر روی گردانی به دو دسته درونی و بیرونی تقسیم بندی شده است (امامی و همکاران، ۱۳۹۴). برخی از مهم ترین عوامل درونی عبارت است از: قوانین، عوامل مدیریتی، تبلیغات، خدمات، ساختار هزینه، کارکنان، تعامل و عوامل محیطی. برخی از مهم ترین عوامل بیرونی نیز عبارت است از: شهرت و اعتبار، روی گردانی اجباری، شلوغی و ازدحام، عوامل اجتماعی، تجربه‌های گذشته، ملاحظه‌های قومی و هزینه روی گردانی. برای نمونه، در یک پژوهش برای تحلیل عوامل مؤثر بر روی گردانی مشتریان در صنعت بانکداری از داده‌های تراکنش مشتریان بانک، اطلاعات دموگرافیک، طول مدت ارتباط مشتریان با بانک و شکایت مشتریان استفاده شده است (Keramati et al., 2016). نتایج به دست آمده با درخت تصمیم نشان دهنده ویژگی‌های مشتریان روی گردان است. در پژوهشی دیگر از مدل یادگیری ماشین افراطی برای پیش بینی روی گردانی مشتریان در صنعت بانکداری استفاده شده است. نتایج کسب شده حاکی از دقت بیشتر مدل ارائه شده در مقایسه با تکنیک‌های ماشین‌های بردار پشتیبان و شبکه‌های عصبی پس انتشار است (Mohanty & Sree, 2018). محققان در یک پژوهش به این نتیجه دست یافته‌اند که تأثیر وفاداری و رضایت مشتریان بر

قصد جابه‌جایی معنادار است و موانع جابه‌جایی این ارتباط را تقویت می‌کند (Mosavi et al., 2018). محققان در یک مطالعه قیاسی برای پیش‌بینی روی‌گردانی مشتریان، Apache Sark ML و MLlib را بررسی کرده‌اند (Sayed et al., 2018). در یک پژوهش دیگر از تکنیک شبکه عصبی برای تحلیل روی‌گردانی مشتریان استفاده شده است (Rosa, 2019). نویسنده در این پژوهش از داده‌های یک دوره پنج‌ماهه برای تعریف مشتریان وفادار با توجه به متغیرهای RFM استفاده کرده است. سپس از روش یادشده برای پنج ماه بعد به جهت تشخیص اینکه کدام یک از مشتریان وفادار مشخص شده در دوره قبلی در این دوره روی‌گردان شده‌اند، استفاده می‌شود. در پژوهشی دیگر با استفاده از شبکه‌های عصبی به پیش‌بینی مشتریان روی‌گردان در صنعت مخابراتی پرداخته شده است (Khan et al., 2019). در ابتدا مجموعه داده از شرکت‌های مخابراتی پاکستان جمع‌آوری و سپس برای حذف اطلاعات نامناسب، مراحل پیش‌پردازش بر روی داده‌ها پیاده‌سازی و اجرا و در ادامه، ۲۶ ویژگی مهم که برای تشخیص روی‌گردانی مشتریان تأثیرگذار بوده انتخاب شده است. در نهایت، مدلی با به کارگیری الگوریتم شبکه‌های عصبی ایجاد شد که دقتی برابر با ۷۹ درصد را برای پیش‌بینی مشتریان روی‌گردان نشان می‌دهد. در یک پژوهش دیگر با استفاده از تکنیک‌های یادگیری ماشین، الگوریتم‌های درخت تصمیم، جنگل تصادفی گرادیان بوسستینگ و XGBoost به پیش‌بینی مشتریان روی‌گردان در حوزه مخابرات با استفاده از محیط Hadoop پرداخته شده است (Ahmad et al., 2019). در ابتدا مجموعه داده به کاررفته با استفاده از مراحل پیش‌پردازش و فرآیند تحلیل ویژگی‌ها بررسی شد. در ادامه، برای مدل‌سازی پیش‌بینی و طبقه‌بندی مشتریان روی‌گردان، ویژگی‌های

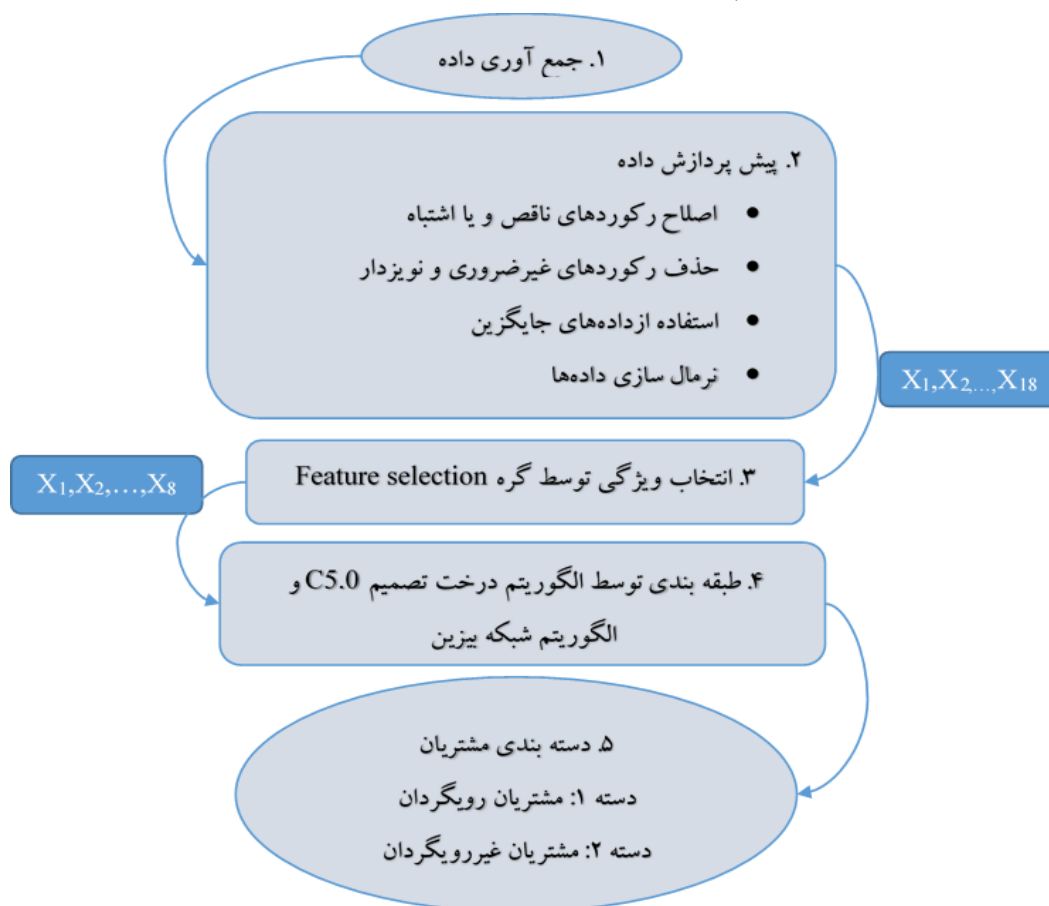
مناسب انتخاب و در اختیار الگوریتم‌ها قرار داده شد. نتایج نشان می‌دهد که مدل ایجادشده با به کارگیری XGBoost نسبت به مدل‌های ایجادشده دیگر دقت بیشتر و عملکرد بهتری را برای طبقه‌بندی مشتریان روی‌گردان دارد. در یک پژوهش دیگر با به کارگیری رویکرد متن‌کاوی، ابعاد رضایت مشتری در خطوط هواپیمایی با استفاده از نظرهای برخط مشتریان بررسی شده است (Lucini et al., 2020). محققان نظرهای مشتریان را از خطوط هواپیمایی و مسافران جمع‌آوری و ابعاد رضایت مشتریان را در ۲۷ بُعد شناسایی و همچنین، نظرهای مشتریان را با به کارگیری روش طبقه‌بندی بیزین ساده به مثبت و منفی تقسیم کرده‌اند و راهکارهایی را برای بهبود رضایت مشتریان ارائه داده‌اند؛ بنابراین حفظ مشتریان فعلی و جذب مشتریان جدید اهمیت زیادی دارد و انجام‌دادن یک پژوهش جامع را می‌طلبد. تحلیل مناسب رفتار روی‌گردانی و پیش‌بینی دقیق آن می‌تواند منجر به کسب سود بیشتر شود. در اغلب مطالعات رفتار روی‌گردانی مشتریان به صورت مستقل از رفتار دیگر مشتریان در نظر گرفته می‌شود. در یک مطالعه محققان عوامل مؤثر بر روی‌گردانی مشتریان و پیش‌بینی آن را با در نظر گرفتن وابستگی رفتار آنها با استفاده از ابزار تحلیل شبکه و الگوریتم یادگیری ماشین تحلیل کرده‌اند (Ljubičić et al., 2023). همچنین، پژوهشگران در مطالعه‌ای دیگر یک مدل پیش‌بینی را برای تحلیل روی‌گردانی مشتریان در فضای تجارت بنگاه‌به‌بنگاه با استفاده از ابزار داده‌کاوی مبتنی بر الگوریتم جنگل تصادفی ارائه کرده‌اند (Gattermann-Itschert & Thonemann, 2022). در یک مطالعه دیگر، پژوهشگران یک مدل پیش‌بینی برای تحلیل عوامل مؤثر بر روی‌گردانی مبتنی بر الگوریتم تحولی و شبکه بیزین در صنعت مخابرات ارائه کرده‌اند (Amin et al., 2023). در یک مطالعه

(Shrestha & Shakya, 2022).

۳. روش پژوهش

پژوهش حاضر از نظر هدف کاربردی و از نظر رویکرد تجزیه و تحلیل داده‌ها کمی است. شناسایی و تعریف شاخص‌های روی گردانی و ارزیابی مشتریان مبتنی بر پیشینه موضوع و نظر خبرگان انجام شده است. برای این منظور در ابتدا مطالعه معیارها براساس مطالعات کتابخانه‌ای و یافته‌های حاصل از پژوهش‌های پیشین جمع آوری شده است. سپس مجموعه ویژگی‌ها با روش دلفی فازی دو مرحله‌ای به بررسی و تأیید خبرگان پژوهش رسیده است. گام‌های کلی پژوهش حاضر در شکل ۱ نشان داده شده است.

دیگر محققان روش‌های مختلف مبتنی بر یادگیری ماشین را برای پیش‌بینی عوامل مؤثر بر روی گردانی مشتریان بررسی کرده‌اند (Geiler et al., 2022). همچنین، محققان در یک مطالعه دیگر روش‌های مختلف طبقه‌بندی را برای پیش‌بینی روی گردانی مشتریان در فضای تجارت الکترونیک بررسی کرده‌اند (Baghla & Gupta, 2022). همچنین، محققان در یک مطالعه‌ای دیگر روش‌های مختلف طبقه‌بندی را برای پیش‌بینی روی گردانی مشتریان در فضای تجارت الکترونیک بررسی کرده‌اند (Kim & Lee, 2022). محققان با استفاده از رویکرد یادگیری ماشین مبتنی بر پیاده‌سازی و با تقویت گرادیان درخت تصمیم‌گیری اقدام به پیش‌بینی عوامل مؤثر بر روی گردانی مشتریان در صنعت مخابرات کشور نپال کرده‌اند



شکل ۱: گام‌های الگوریتم ترکیبی مبتنی بر درخت تصمیم C5.0 و شبکه بیزین (منبع: یافته‌های پژوهش)

Figure 1: Steps of Hybrid Algorithm Based on C5.0 Decision and Bayesian Network

۳-۱. جمع آوری داده‌ها

داده‌های بسیار زیادی در بانک اطلاعاتی فروشگاه‌های زنجیره‌ای ثبت شده است. ویژگی‌های لازم پژوهش حاضر با نظر خبرگان و پژوهش‌های پیشین انتخاب شده است. با استفاده از پژوهش‌های گذشته و مبانی نظری درباره موضوع مدنظر، عوامل اصلی روی گردانی شناسایی و استخراج شده است. سپس با کمک تکنیک دلفی فازی در دو مرحله که مبتنی بر نظرهای خبرگان بود، ویژگی‌های اصلی در دو مرحله شناسایی شد تا در طراحی مدل از آنها استفاده شود.

۳-۲. پیش پردازش داده‌ها

پیش پردازش داده‌ها به عنوان اولین مرحله در ساخت مدل اهمیت فراوانی دارد. به کارگیری روش‌های مناسب پیش پردازش بر روی داده‌ها قبل از ورود آنها به مدل‌های هوش مصنوعی از جمله مواردی است که می‌تواند نتایج حاصل از شبیه‌سازی‌ها را به واقعیت نزدیک کند (حبیبی‌پور و همکاران، ۱۳۹۶). در پژوهش حاضر داده‌ها در ابتدا با استفاده از گره انتخاب ویژگی حاضر در نرم‌افزار مدلر، بررسی و محدودسازی شده است؛ به گونه‌ای که ویژگی‌هایی که رکودهای ناقص و یا اشتباه داشتند، اصلاح شدند. درآمد، ویژگی‌هایی که اطلاعات نوین‌دار داشتند و یا از رکودهای غیرضروری تشکیل شده بودند با به کارگیری داده‌های جایگزین اصلاح شدند. در نهایت، داده‌های موجود برای ارتقا کیفیت نتایج حاصل، طبق معادله ۱ نرمال‌سازی شد. به این ترتیب، داده‌ها در محدوده ۰ و ۱ مقدار می‌گیرند.

ارزش واقعی - مقدار پیشینه داده‌ها

(۱) مقدار پیشینه داده‌ها - مقدار کمینه داده‌ها

کردن نرمال =

۳-۳. انتخاب ویژگی

در روش‌های انتخاب ویژگی با انتخاب زیرمجموعه‌ای بهینه و با حداقل اندازه ممکن از ویژگی‌های اولیه، ابعاد داده‌ها کاهش می‌یابد. روش‌های گوناگون انتخاب ویژگی سعی بر این دارند که از میان 2^N زیرمجموعه کاندید برای یک مجموعه N عضو، بهترین زیرمجموعه را بر اساس یک تابع ارزیابی پیدا کنند. در تمام این روش‌ها بر اساس کاربرد، زیرمجموعه‌ای به عنوان جواب انتخاب می‌شود که بتواند مقدار تابع ارزیابی را بهینه کند. اگرچه هر روش تلاش می‌کند تا بهترین ویژگی‌ها انتخاب شود، با توجه به وسعت جواب‌های ممکن و افزایش تعداد آنها به صورت توانی از N ، پیدا کردن جواب بهینه مشکل است و در N های متوسط و بزرگ هزینه زیادی دارد. به طور کلی، روش‌های انتخاب ویژگی به سه دسته اصلی بسته‌بندی، فیلترینگ و ترکیبی تقسیم‌بندی می‌شود (Kumar & Elavarasan, 2014). در پژوهش حاضر از روش فیلترینگ برای انتخاب ویژگی‌ها استفاده شده است. در روش فیلترینگ با توجه به ویژگی‌های کلی داده‌ها، ویژگی‌های بهینه از میان کل ویژگی‌های موجود انتخاب می‌شود. این روش هزینه محاسباتی کم و سرعت بالایی دارد و به راحتی می‌توان آن را پیاده‌سازی کرد (Tomar & Agarwal, 2014). در پژوهش حاضر از گره انتخاب ویژگی برای حذف داده‌های زائد استفاده شده است. این گره با استفاده از ضریب همبستگی به انتخاب فیلهایی که می‌پردازد که ضریب همبستگی آنها به عدد یک نزدیک است. در طبیعت حوادث متعددی اتفاق می‌افتد که بین آنها همبستگی وجود دارد؛ به طوری که افزایش در مقدار یک متغیر منجر به افزایش یا کاهش در مقدار یک متغیر دیگر می‌شود که نشان‌دهنده همبستگی است.

وجود همبستگی بین متغیرهای تحت ارزیابی، به این معنا نیست که یک متغیر، علت متغیر دیگر است. ضریب همبستگی، شاخص آماری است که میزان و حدود ارتباط بین متغیرها را نشان می دهد. در این گره پنج آیتم وجود دارد که در آیتم اول می توان براساس مقدارهای مفقوده انتخاب کرد؛ یعنی در این آیتم ۷۰٪ رکوردها و فیلدهایی که مقدارهای مفقود شده را دارند، به ترتیب حذف و غربال می شوند. در آیتم دوم فیلدهایی غربال می شود که رکوردهای زیادی را براساس فیلدهایی که در دسته های مشابهی قرار گرفته، داشته باشد؛ برای مثال، ۹۰٪ مشتریان فروشگاه اگر یک نوع کالا را خریداری کرده باشند این ویژگی نمی تواند تمایز دهنده یک مشتری از مشتری دیگر باشد؛ در نتیجه این متغیر به عنوان یک متغیر بی اهمیت می تواند از مدل تحلیل حذف شود. در آیتم سوم فیلدهایی غربال می شود که تعداد زیادی دسته های جزئی و متفاوت با سایر دسته ها وجود داشته باشد؛ یعنی اگر هر مشتری کالای متفاوتی را خریداری کند این اطلاعات در مدل سازی مشتری بعدی به طور طبیعی، نقشی ندارد و مفید نخواهد بود و به همین صورت دو آیتم بعدی براساس ضریب اطمینان، واریانس و انحراف معیار انتخاب و ویژگی را انجام می دهند. برای انجام دادن انتخاب و ویژگی، احتمال خرید مجدد مشتری به عنوان متغیر وابسته پژوهش در نظر گرفته می شود که مقدارهای ۰ و ۱ به ترتیب شامل عدم خرید مجدد و خرید مجدد می شود. متغیرهای پیش بینی کننده های پژوهش حاضر موارد ۱ تا ۱۷ جدول ۱ است. در این پژوهش از گره انتخاب و ویژگی واقع در نرم افزار مدلر برای متغیر وابسته و پیش بینی کننده استفاده و عملکرد رویکرد انتخاب و ویژگی نیز با دقت حاصل از فرآیند طبقه بندی سنجیده شده است. به این ترتیب که یکبار از

الگوریتم انتخاب و ویژگی برای انجام دادن طبقه بندی استفاده و بار دیگر بدون استفاده از الگوریتم انتخاب و ویژگی سعی در طبقه بندی داده ها می شود. در نهایت، با مقایسه میزان دقت هر کدام، می توان به اهمیت به کارگیری الگوریتم انتخاب و ویژگی پی برد. در پایان نیز با استفاده از شبکه بیزین به انتخاب و ویژگی ها (اولویت بندی ویژگی ها) پرداخته شده و سپس درخت تصمیم مبتنی بر خروجی بیزین ترسیم شده است.

۴-۳. درخت تصمیم

تکنیک درخت تصمیم یکی از قدیمی ترین تکنیک ها برای ایجاد مدل دسته بندی است. الگوریتم های مبتنی بر درخت تصمیم، نتیجه کار را به صورت درختی از کلیه حالت های مختلف مقدارهای ویژگی ها ارائه می دهند (صنعی آباده و همکاران، ۱۳۹۱). از این رو درخت تصمیم قادر به ایجاد توصیف های درک پذیر برای انسان از مجموعه روابط موجود در یک مجموعه داده است و همچنین، می تواند برای وظایف دسته بندی و پیش بینی از درخت تصمیم استفاده شود (Wu et al., 2022). این ساختار تصمیم گیری به شکل های گوناگون مانند تکنیک های ریاضی و محاسباتی که به توصیف، دسته بندی و عام سازی یک مجموعه از داده ها کمک می کند نیز معرفی می شود (Tsai & Chiou, 2009). از جمله نقاط قوت درخت تصمیم عبارت است از: ۱- قابل فهم و درک بودن مدل ایجاد شده؛ ۲- ارائه پیش بینی ها در قالب یک رشته قوانین؛ ۳- بی نیاز بودن به محاسبات طولانی و پیچیده برای دسته بندی داده ها؛ ۴- قابلیت به کارگیری برای انواع مختلف داده ها؛ ۵- شناسایی متغیرهای تأثیرگذار در پیش بینی و دسته بندی.

در پژوهش حاضر از درخت تصمیم C5.0 استفاده

۳-۵. شبکه بیزین

شبکه بیزین یک روش گرافیکی است که محققان در مطالعات تحلیل ریسک و ایمنی بر پایه دانش احتمال و عدم قطعیت به آن توجه کرده‌اند. شبکه بیزین یک گراف جهت‌دار غیرمدور است که هر گره نشان‌دهنده یک متغیر و کمان‌ها نیز نشان‌دهنده روابط علی مستقیم بین گره‌های بهم متصل شده، است. جدول‌های احتمال شرطی به گره‌های با وابستگی شرطی اختصاص داده می‌شود. براساس استقلال شرطی به دست آمده از مفهوم d-separation و قاعده زنجیره ای، شبکه بیزین توزیع احتمال مشترک $p(x)$ از متغیرهای $X = \{U_1, \dots, U_n\}$ را طبق رابطه ۴ نشان می‌دهد (Nielsen & Jensen, 2009; Kjaerulff & Madsen, 2008).

$$P(X) = \prod_{i=1}^n p(U_i | P_a(U_i)) \quad (۴)$$

$P_a(U_i)$ نشان دهنده احتمال والد متغیر U_i است؛ از این رو احتمال متغیر U_i به صورت رابطه ۵ محاسبه می‌شود.

$$P(U_i) = \sum_{x \setminus U_i} P(X)P(U_i) = \sum_{x \setminus U_i} P(X) \quad (۵)$$

یکی دیگر از ویژگی‌های بی نظیر شبکه بیزین در مقایسه با سایر روش‌ها، قابلیت بهره‌گیری از تئوری بیز برای به‌روزرسانی احتمال وقوع رویدادهای اولیه به محض دریافت شواهد جدید (مانند آمار وقوع حوادث، داده‌های فرآیندی لحظه‌ای برای محاسبه مقادارهای احتمالات به‌روزشده) است. (رابطه ۶) (Nielsen & Jensen, 2009; Kjaerulff & Madsen, 2008).

$$P(X|E) = \frac{P(X \cdot E)}{P(E)} = \frac{P(X \cdot E)}{\sum_X P(X \cdot E)} \quad (۶)$$

در رابطه (۶) X نشان‌دهنده متغیر تحت ارزیابی و E شواهد دریافتی است.

شده است که از جمله مزایای الگوریتم C5.0 درخت تصمیم، سرعت زیاد، کارایی حافظه، ایجاد درخت‌های تصمیم کوچک‌تر، وزن دهی و غربال‌سازی است (بحرینی‌نژاد و سروش، ۱۳۸۸).

الگوریتم درخت تصمیم C5.0 نسخه بهبودیافته بر پایه درخت‌های تصمیم C4.5 و ID3 است که در آن شکسته شدن هر گره بر پایه بهره اطلاعاتی خواهد بود. از شاخص بهره اطلاعاتی برای گزینش متغیر شکننده در فرآیند رشد درخت به کار گرفته می‌شود. همچنین، میزان همگنی نمونه‌ها در یک گره با شاخص آنتروپی ارزیابی و از آن در تعیین بهره اطلاعاتی نیز استفاده می‌شود. در صورتی که متغیر هدف، C مقدار متفاوت داشته باشد، شاخص آنتروپی S وابسته به C کلاس از رابطه ۲ محاسبه خواهد شد.

$$Entropy(s) = - \sum_{i=1}^c P_i \cdot \log_2 P_i \quad (۲)$$

شاخص P_i نسبتی از S است که به کلاس i تعلق می‌گیرد. در واقع، شاخص بهره اطلاعاتی میزان کاهش انتظار را در شاخص آنتروپی ارائه می‌کند. به عبارت دیگر، بهره اطلاعاتی تعیین‌کننده اثرگذاری یک متغیر در فرآیند کلاس‌بندی است. بهره اطلاعاتی برای متغیر A وابسته به داده‌های S با استفاده از رابطه ۳ محاسبه خواهد شد.

$$Entropy(S, A) = Entropy(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (۳)$$

در بخش نخست رابطه (۳) میزان شاخص آنتروپی S در حالت اولیه و در بخش دوم آن میزان شاخص آنتروپی انتظار پس از تقسیم براساس متغیر A نمایش داده شده است.

۴. یافته‌ها و بحث

در این مطالعه، اطلاعات مرتبط با خرید محصولات مصرفی از نوع تندمصرف از شعبه‌های منتخب یک فروشگاه زنجیره‌ای در شهرستان مشهد با استفاده از ابزار داده کاوی تحلیل شده است. مجموعه داده استفاده شده در این پژوهش، شامل اطلاعات ۱۰۰۰ مشتری با ۱۸ ویژگی است که ۷ ویژگی آن مربوط به اطلاعات دموگرافیک مشتریان مانند سن، جنسیت، وضعیت تأهل، مدرک تحصیلی، تعداد خانوار، وضعیت اشتغال و متوسط سطح درآمد ماهیانه و سایر ویژگی‌ها درباره سوابق خرید هر مشتری است. محققان اطلاعات را با

استفاده از پرسشنامه طراحی شده جمع‌آوری کردند که دربرگیرنده سؤال‌های مرتبط با ویژگی‌های یادشده در فوق بوده است. ویژگی‌های استفاده‌شده برای انجام دادن این پژوهش (جدول ۱) با توجه به پیشینه موضوع، نظرهای خبرگان با روش دلفی دو مرحله‌ای جمع‌آوری شده است. جامعه خبرگان پژوهش دربرگیرنده ۲۰ نفر از مدیران ارشد فروشگاه زنجیره‌ای شهرستان مشهد و استادان دانشگاه در حوزه بازاریابی بوده است.

جدول ۱: ویژگی‌های انتخاب‌شده

Table 1: Selected Features

شماره ویژگی	نام ویژگی	منبع	شماره ویژگی	نام ویژگی	منبع
۱	جنسیت	Keramati et al. (2016)	۱۰	تعداد خرید در ماه	Xue & Lu (2011) Baghla & Gupta (2022) Kim & Lee (2022)
۲	سن		۱۱	ارزش خرید در ماه	
۳	وضعیت تأهل	Gattermann & Thonemann (2022) Amin et al. (2023)	۱۲	تعداد خرید در سال	
۴	مدرک تحصیلی		۱۳	تأثیر فروش ویژه در خرید	
۵	تعداد خانواده		۱۴	ارزش آخرین خرید در زمان فروش ویژه	
۶	وضعیت اشتغال		۱۵	نحوه آشنایی با فروشگاه	
۷	متوسط سطح درآمد ماهیانه		۱۶	سهم خرید اینترنتی در طی سال	
۸	روش آخرین خرید		۱۷	نوع بازار	
۹	ارزش آخرین خرید	Xue & Lu (2011)	۱۸	احتمال خرید مجدد	Shobana et al. (2023)

منبع: یافته‌های پژوهش

برای انجام دادن و تحلیل مدل ارائه‌شده در این پژوهش از نرم‌افزار IBM SPSS MODELER 18.0 استفاده شده است. بعد از انجام دادن پیش‌پردازش از گره انتخاب ویژگی برای شناسایی ویژگی‌های نامرتب

و تکراری که مسئله را با کمترین کاهش درجه کارایی تشریح می‌کند، استفاده و سپس از الگوریتم درخت تصمیم C5.0 و الگوریتم شبکه بیزین برای طبقه‌بندی داده‌ها استفاده می‌شود.

۴-۱. بررسی اثربخشی رویکرد انتخاب ویژگی

در پژوهش حاضر داده‌ها برای ارزیابی مدل ایجاد شده به دو بخش آموزش و آزمایشی تقسیم‌بندی شده است. درباره تعداد رکوردهای آموزش و آزمون می‌توان گفت که اگر تعداد داده آموزشی زیاد باشد، مدل ساخته شده به واقعیت نزدیک‌تر خواهد شد. در واقع، از ۷۵ درصد داده‌ها برای آموزش و از ۲۵ درصد نیز برای آزمون و اعتبار مدل استفاده شده است. داده‌های بخش آموزش باعث می‌شود که مدل به‌طور هوشمند، خود را به‌روز و مدل بهینه را ایجاد کند.

داده‌های بخش آزمون نیز مدل ایجاد شده را ارزیابی می‌کند. برای بررسی اثربخشی رویکرد انتخاب ویژگی از معیار دقت طبقه‌بندی استفاده شده است. بدین جهت، برای بررسی اثربخشی این رویکرد در میزان دقت طبقه‌بندی یکبار از گره انتخاب ویژگی و بار دیگر بدون استفاده از گره انتخاب ویژگی سعی در طبقه‌بندی داده‌ها شده است که نتایج الگوریتم درخت تصمیم و شبکه بیزین به صورت جداگانه در جدول‌های ۲ و ۳ آمده است.

جدول ۲: بررسی اثربخشی رویکرد انتخاب ویژگی درخت تصمیم C5.0

Table 2: Evaluation of the Effectiveness of the C5.0 Decision Tree Feature Selection Approach

مجموعه داده آزمون		مجموعه داده یادگیری		الگوریتم درخت تصمیم C5.0
درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	
۶/۰۲	۹۳/۹۸	۱/۳۸	۹۸/۶۲	دقت بدون انتخاب ویژگی
۲/۴۱	۹۷/۵۹	۰	۱۰۰	دقت با انتخاب ویژگی

منبع: یافته‌های پژوهش

جدول ۳: بررسی اثربخشی رویکرد انتخاب ویژگی شبکه بیزین

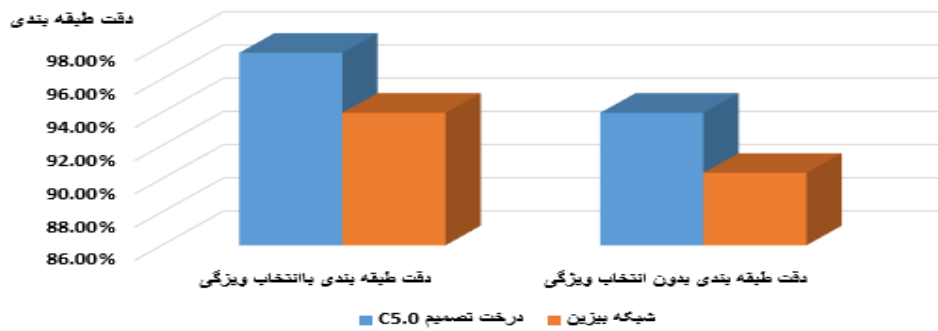
Table 3: Evaluation of the Effectiveness of the Bayesian Network Feature Selection Approach

مجموعه داده آزمون		مجموعه داده یادگیری		الگوریتم شبکه بیزین
درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	
۹/۶۴	۹۰/۳۶	۰/۹۲	۹۹/۰۸	دقت بدون انتخاب ویژگی
۶/۰۲	۹۳/۹۸	۳/۲۳	۹۶/۷۷	دقت با انتخاب ویژگی

منبع: یافته‌های پژوهش

در مجموعه داده آزمون به ترتیب برابر با ۹۷/۵۹ و ۹۳/۹۸ محاسبه شده است. با مقایسه نتایج به دست آمده می‌توان نتیجه گرفت که استفاده از گره انتخاب ویژگی (به عنوان تکنیک استفاده شده برای رویکرد انتخاب ویژگی) تأثیر بسزایی در دقت طبقه‌بندی دارد.

همان‌گونه که در جدول‌های ۲ و ۳ و شکل ۲ مشاهده می‌شود، دقت طبقه‌بندی دو الگوریتم درخت تصمیم C5.0 و شبکه بیزین بدون استفاده از انتخاب ویژگی در مجموعه داده آزمون به ترتیب برابر با ۹۳/۹۸ و ۹۰/۳۶ است و دقت طبقه‌بندی دو الگوریتم درخت تصمیم C5.0 و شبکه بیزین با استفاده از انتخاب ویژگی



شکل ۲: نمودار میله‌ای بررسی اثربخشی رویکرد انتخاب ویژگی (منبع: یافته‌های پژوهش)

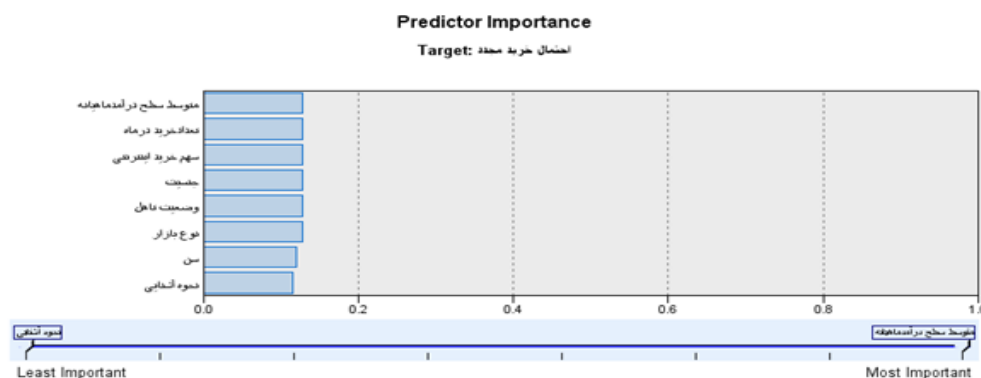
Figure 2: Bar Graph of the Effectiveness of the Feature Selection Approach

انجام دهد. این فرآیند به صورت بازگشتی تکرار می‌شود تا بازمی‌زیر گروه‌ها به زیرگروه‌های دیگری شکسته شود. در این فرآیند زیرنمونه‌ها مجدداً شاخه زده می‌شود و تا زمانی ادامه می‌یابد که نتوان زیرنمونه‌ها را شاخه زد. در پایان، پایین‌ترین سطح شاخه‌ها آزموده می‌شود و با حذف یا هرس شاخه‌هایی که نقش مهمی در مدل ندارد، عملیات پایان می‌یابد (علیزاده و ملک محمدی، ۱۳۹۳). با توجه به نتایج به دست آمده در شکل ۳ با اهمیت‌ترین متغیرها با گره انتخاب ویژگی انتخاب و به ترتیب میزان اهمیت آنها شناسایی شده است که با اهمیت‌ترین آنها متوسط سطح درآمد مشتریان را نشان می‌دهد.

۲-۴. بررسی اثربخشی رویکرد طبقه بندی با

استفاده از الگوریتم درخت تصمیم C5.0

در پژوهش حاضر برای ایجاد درخت تصمیم، داده‌ها به دو بخش آموزش و آزمون با نسبت ۳ به ۱ تقسیم و از الگوریتم درخت تصمیم انتخابی C5.0 استفاده شده است. همچنین، از متغیر احتمال خرید مجدد به عنوان متغیر هدف (متغیر وابسته) و از متغیرهای شناسایی شده با گره انتخاب ویژگی در قالب متغیرهای با اهمیت به عنوان متغیرهای ورودی (متغیرهای مستقل) برای ساخت درخت تصمیم C5.0 استفاده شده است. در واقع، درخت تصمیم C5.0 با بررسی تمامی فیلدهای پایگاه داده قصد دارد تا به فیلدی برسد که بهترین دسته‌بندی و پیش‌بینی را با تقسیم داده‌ها به زیرگروه‌ها



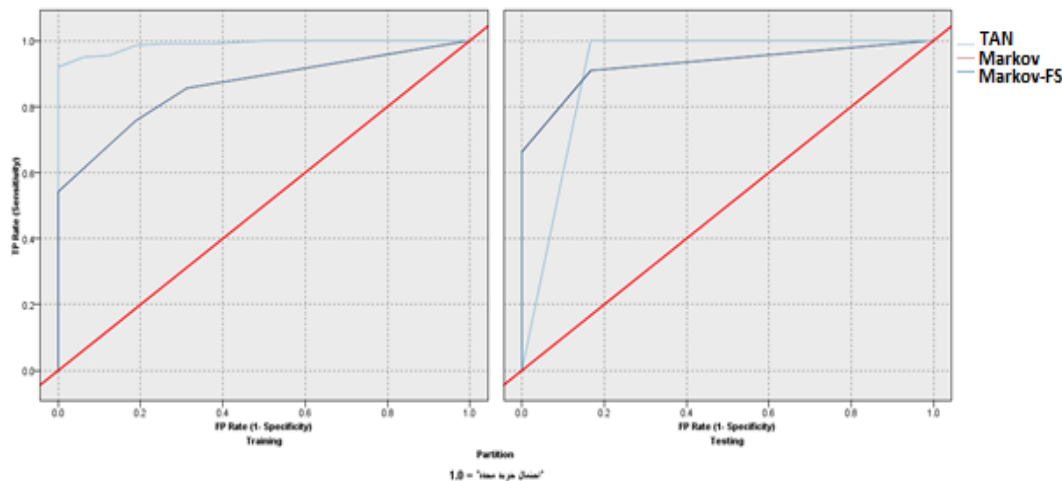
شکل ۳: نتایج طبقه‌بندی با الگوریتم درخت تصمیم C5.0 (منبع: یافته‌های پژوهش)

Figure 3: Classification Results with C5.0 Decision Tree Algorithm

۳-۴. بررسی اثربخشی رویکرد طبقه بندی با استفاده از شبکه بیزین

در پژوهش حاضر با استفاده از الگوریتم شبکه بیزین و درخت تصمیم، مدل و قوانین رویکرد طبقه بندی در راستای احتمال خرید مجدد مشتریان ارائه و استخراج شده است. پس از جمع آوری داده های مربوط به متغیرها به ساخت مجموعه آموزش و آزمون پرداخته شده است که ۷۵ درصد داده ها به عنوان مجموعه آموزش و ۲۵ درصد به عنوان مجموعه آزمون به صورت تصادفی انتخاب شده است. داده های بخش آموزش مدل را می سازند و داده های بخش آزمون نیز به ارزیابی مدل ایجاد شده می پردازند. در واقع، یکی از نیازمندی های شبکه بیزین و درخت تصمیم، فیلد هدف طبقه ای است. فیلد هدف (متغیر وابسته) این پژوهش احتمال خرید مجدد و فیلدهای ورودی ۱۷ متغیر مستقل است. به طور معمول، در پژوهش های مرتبط با مدل سازی برای بررسی میزان دقت مدل و میزان

پیش بینی از نمودار receiver operating characteristics (ROC) استفاده می شود. سطح زیر منحنی، میزان توانایی مدل را در تفاوت قائل شدن بین چند نتیجه نشان می دهد که میزان تمایز نامیده می شود. هرچه قدر عدد سطح زیرین منحنی ROC به ۱ نزدیک تر باشد، می توان نتیجه گرفت که دقت مدل در معیار خوب است و برعکس. همچنین، هرچه این عدد به ۰/۵ نزدیک تر باشد، می توان نتیجه گرفت که دقت مدل پایین است و پیش بینی نامناسب دارد. در ادامه، طبقه بندی داده ها با سه الگوریتم شبکه بیزین به نام های TAN MARKOV و MARKOV-FS بررسی شده است و با توجه به مقایسه دقت به دست آمده در نمودار ROC (شکل ۴) می توان نتیجه گرفت که تابع TAN از دو تابع MARKOV و MARKOV-FS عملکرد بهتری دارد.



شکل ۴: مقایسه دقت سه الگوریتم شبکه بیزین (منبع: یافته های پژوهش)

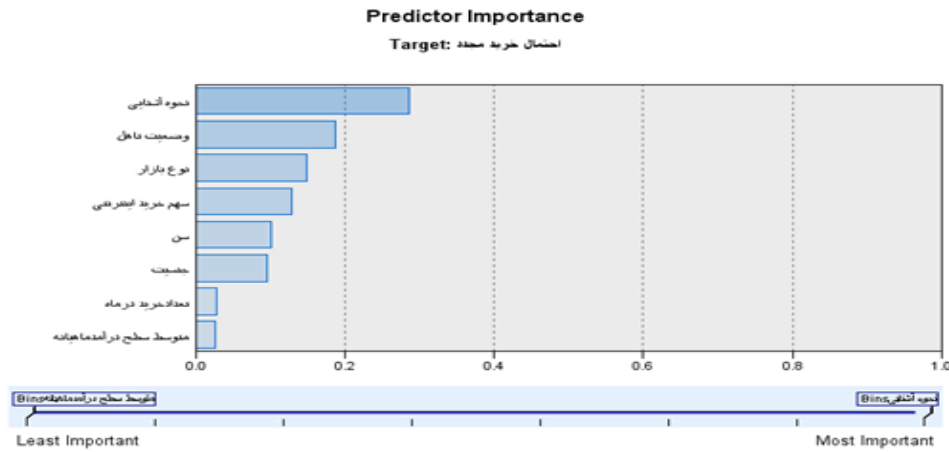
Figure 4: Comparison of the Accuracy of Three Bayesian Network Algorithms

عملکرد الگوریتم جست و جوی TAN به گونه ای است که با ایجاد یک شبکه بیزین ساده این امکان را فراهم می کند که هر پیش بینی کننده علاوه بر وابستگی به متغیر

باتوجه به نتایج به دست آمده در نمودار ROC برای اینکه بتوانیم شبکه ای با کیفیت بالا داشته باشیم باید از الگوریتم جست و جوی شبکه بیزین TAN استفاده شود.

شکل ۵، نمودار اهمیت طبقه بندی متغیرهای (۱۳۹۳). به دست آمده را با گره انتخاب ویژگی نشان می دهد.

هدف بتواند به دیگر پیش بینی کننده ها نیز وابستگی داشته باشد؛ بنابراین ساختار TAN باعث می شود تا دقت طبقه بندی افزایش یابد (علیزاده و ملک محمدی،

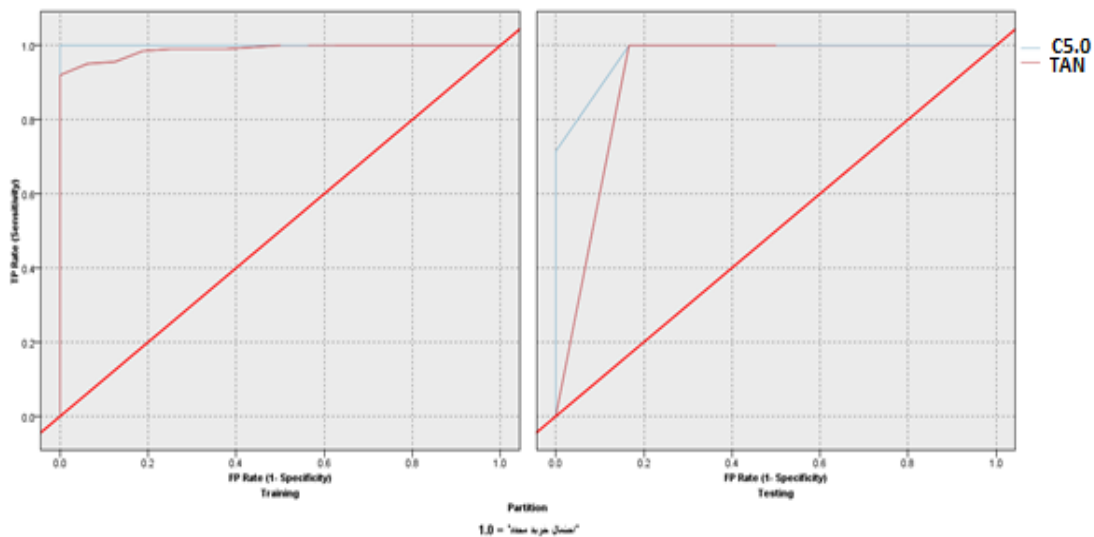


نمودار ۵: طبقه بندی متغیرها با توجه به میزان اهمیت آنها با شبکه بیزین (منبع: یافته های پژوهش)

Diagram 5: Classification of Variables According to Their Importance with Bayesian Network

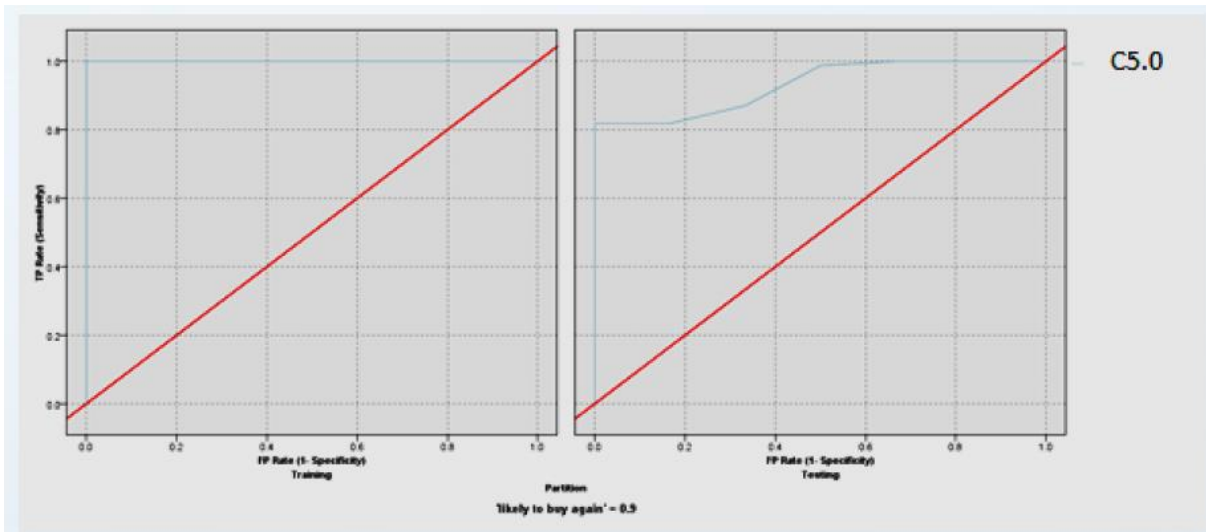
تصمیم C5.0 عملکرد بهتری در پیش بینی روی گردانی مشتریان دارد.

در شکل های ۶ و ۷ مشخصه عملکرد سیستم ROC برای الگوریتم ها آورده شده است. این نتایج بیانگر آن است که از نظر سطح زیر منحنی ROC مدل درخت



شکل ۶: نمودار ROC مقایسه دو الگوریتم درخت تصمیم C5.0 و شبکه بیزین (منبع: یافته های پژوهش)

Figure 6: ROC Diagram Comparing Two C5.0 Decision Tree Algorithms and Bayesian Network



شکل ۷: نمودار ROC درخت تصمیم C5.0 (منبع: یافته‌های پژوهش)

Figure 7: ROC Diagram of C5.0 Decision Tree

دقت بیشتری را دارد. به‌طور کلی، با توجه به دو معیار دقت طبقه‌بندی و منحنی ROC می‌توان نتیجه گرفت که مدل درخت تصمیم C5.0 دقت بیشتری دارد.

همچنین، مطابق جدول ۴ مقدار خروجی و یا مقدار پیش‌بینی شده که با مدل ارائه‌شده به‌دست آمده، نشان می‌دهد درخت تصمیم C5.0 با میزان دقت ۹۷/۵۹ درصد نسبت به شبکه بیزین با میزان دقت ۹۳/۹۸ درصد

جدول ۴: مقایسه دقت دو مدل ارائه‌شده

Table 4: Comparison of the Accuracy of the Two Presented Models

مجموعه داده آزمون		مجموعه داده یادگیری		الگوریتم
درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	درصد پیش‌بینی نادرست	درصد پیش‌بینی درست	
۲/۴۱	۹۷/۵۹	۰	۱۰۰	درخت تصمیم C5.0
۶/۰۲	۹۳/۹۸	۳/۲۳	۹۶/۷۷	شبکه بیزین

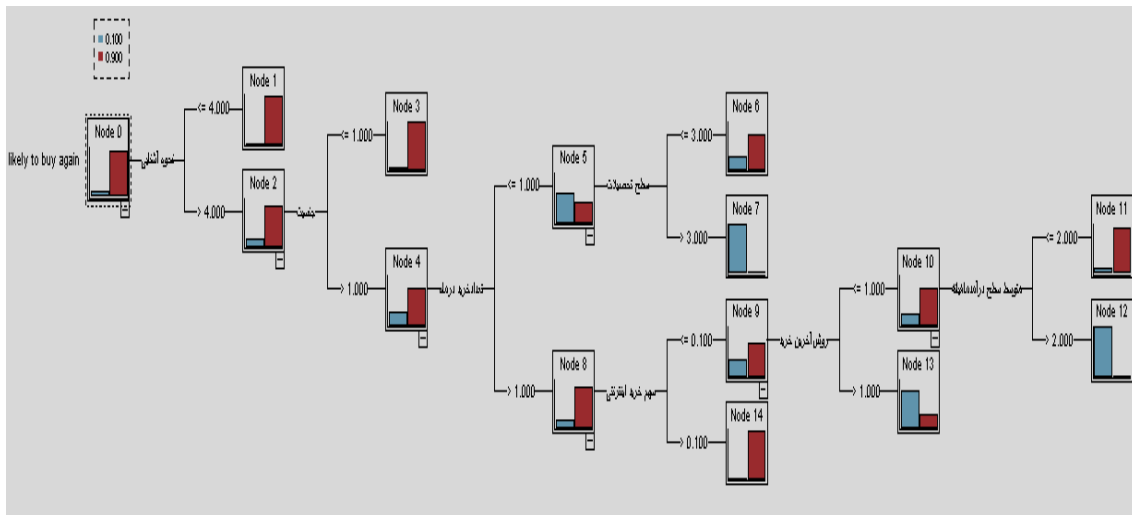
منبع: یافته‌های پژوهش

که متوسط سطح درآمد پایین‌تری دارند و به تعبیری میزان خرید کمی از سازمان دارند، احتمال بیشتری دارد که از سازمان دریافت‌کننده خدمت روی گردان شوند. از سویی برخلاف آنچه تصور می‌شد، متغیر نحوه آشنایی جزء کم‌تأثیرترین متغیرها بر پیش‌بینی روی گردانی مشتریان است.

در نهایت، پس از اولویت‌بندی معیارها با الگوریتم شبکه بیزین (شکل ۸)، نتایج حاصل به‌عنوان ورودی درخت تصمیم C5.0 به کار گرفته می‌شود. نتایج حاصل از پیاده‌سازی درخت تصمیم در شکل ۹ نمایش داده شده است. براساس قوانین استخراج‌شده و متغیرهای تأثیرگذار بر پیش‌بینی، آن دسته از مشتریان



شکل ۸: ویژگی‌های انتخاب‌شده با الگوریتم شبکه بیزین (منبع: یافته‌های پژوهش)
Figure 8: Selected Features with Bayesian Network Algorithm



شکل ۹: نتایج طبقه‌بندی با الگوریتم درخت تصمیم C5.0 (منبع: یافته‌های پژوهش)
Figure 9: Classification Results with C5.0 Decision Tree Algorithm

۴-۴. پیشنهاد‌های مدیریتی

نتایج پژوهش حاکی از آن است که مشخصه‌های دموگرافیک و سوابق خرید مشتریان در رفتار روی گردانی مشتریان اثرگذار است. مبتنی بر نتایج حاصل از طبقه‌بندی در پژوهش حاضر و قوانین ارائه‌شده با درخت تصمیم (شکل ۸)، ۸ عامل کلیدی شناسایی شده در روی گردانی مشتریان با تأثیر فراوانی

برای شهرستان مشهد بررسی و در ادامه، تحلیل شده است. پیشنهاد‌های کاربردی به تأیید خبرگان پژوهش رسیده است.

۴-۱-۱. متوسط سطح درآمد ماهیانه

نتایج طبقه‌بندی انجام‌شده نشان می‌دهد که افراد با درآمد ماهیانه پنج میلیون تومان و بیشتر جزء آن دسته از

- خدمات باارزش به هر دسته از مشتریان با عنوان باشگاه مشتریان و دسته‌بندی آنها در قالب‌های مختلف مانند طلایی، نقره‌ای و برنزی ارائه شود.

۳-۱-۴. سهم خرید اینترنتی در طی سال

نتایج کسب‌شده نشان می‌دهد که افراد با سهم خرید اینترنتی بیشتر از ۲۰ درصد جزء مشتریان وفادار و افراد با سهم خرید پایین‌تر جزء مشتریان روی‌گردان هستند. از آنجایی که امروزه افزایش فروش و جذب مشتریان جدید وابستگی زیادی به استراتژی‌های بازاریابی دیجیتال سازمان دارد، مجموعه پیشنهادهای کاربردی ذیل ارائه می‌شود.

- فعالیت و تولید محتوا در شبکه‌های اجتماعی با تعداد مخاطب زیاد انجام شود؛

- سایت اینترنتی و نرم‌افزارهای موبایلی برای خرید از فروشگاه و ارائه اطلاعات از فعالیت‌های فروشگاه، طرح‌های تخفیف و غیره توسعه یابند؛

- طرح‌های فروش مانند تخفیف ویژه خرید اینترنتی و یا تسهیلات ویژه برای تحویل خرید اینترنتی ارائه شود تا انگیزه مشتریان به انتخاب خرید اینترنتی بیشتر شود.

۴-۱-۴. جنسیت

نتایج پژوهش نشان می‌دهد که میزان خرید بانوان بیشتر از آقایان بوده است و بانوان وفاداری بیشتری نسبت به آقایان داشته‌اند؛ از این رو پیشنهادهایی در زمینه توجه ویژه به جامعه مشتریان بانوان در تدوین طرح‌های فروش مانند ارائه تخفیفات و تسهیلات برای تکرار خرید و افزایش دفعات و ارزش کلی خرید می‌شود. همچنین، پیشنهاد می‌شود که در ارسال محتوای تبلیغاتی و اطلاع‌رسانی به ویژگی جنسیت نیز توجه شود.

مشتریانی هستند که امکان خرید مجدد آنها وجود دارد. از سوی مقابل، افراد با درآمد ماهیانه کمتر از پنج میلیون تومان جزء مشتریان روی‌گردان هستند؛ از این رو پیشنهادهای کاربردی ذیل ارائه می‌شود.

- پیشنهاد می‌شود به تأمین کالاهای مشابه با گروه قیمتی متفاوت دقت شود تا مشتریان با درآمد پایین نیز توانایی خرید را داشته باشند و در نهایت، از روی‌گردانی مشتریان فروشگاه جلوگیری شود.

- پیشنهادهای قیمتی ویژه برای خرید انواع روش‌های تخفیف مبتنی بر اندازه خرید، زمان و حجم خرید برای مشتریان با سطح درآمد پایین‌تر ارائه شود تا این دسته از مشتریان به خرید از فروشگاه تمایل بیشتری پیدا کنند و بدین نحو از روی‌گردانی آنها اجتناب شود.

۲-۱-۴. تعداد خرید در ماه

نتایج پژوهش حاکی از آن است که افرادی که تعداد خرید آنها بیشتر از ۲ بار در ماه بوده است، جزء آن دسته از مشتریان وفادار قرار می‌گیرند؛ زیرا احتمال خرید مجدد آنها وجود دارد و آن دسته از مشتریانی که احتمال خرید آنها کمتر از ۲ بار بوده است، امکان خرید مجدد آنها وجود ندارد و جزء مشتریان روی‌گردان خواهند بود؛ از این رو پیشنهادهای کاربردی ذیل ارائه می‌شود.

- برگزاری جشنواره‌های فروش ویژه نظیر جشنواره پایان فصل و ارائه تخفیف‌های جذاب و اطلاع‌رسانی از طریق باشگاه مشتریان با هدف ترقیب مشتریان به تکرار خرید برگزار شود؛

- طرح‌های تخفیف سفارشی‌سازی شده برای هر گروه از مشتریان با توجه به اطلاعات کسب‌شده از سوابق خرید گذشته با هدف نزدیکی بیشتر به مشتریان و ایجاد وفاداری ارائه شود؛

۵-۱-۴. وضعیت تأهل

باتوجه به نتایج به دست آمده، افراد متأهل درصد بیشتری از مشتریان فروشگاه را دارند و نسبت به افراد مجرد وفاداری بیشتری نیز دارند؛ از این رو پیشنهاد می شود که در ارسال محتوای تبلیغاتی و اطلاع رسانی به ویژگی جنسیت نیز توجه شود.

۶-۱-۴. نوع بازار

نتایج حاصل حاکی از آن است که میزان کلی خرید افراد از فروشگاه زنجیره ای بسیار بیشتر از سوپرمارکت سنتی است. دلیل این پدیده را می توان در ارائه طیف متنوعی از محصولات هم خانواده به صورت متمرکز در یک محل فیزیکی جست و جو کرد. پاسخگویی به طیف گسترده از نیازهای مشتریان، ارائه خدمات همزمان مانند محصولات مصرفی و غیرمصرفی در یک مکان، در نظر گرفتن خدمات تفریحی و سرگرمی برای مشتریان همچون رستوران، محل بازی و نگهداری کودکان و غیره همگی در ارتقا جذابیت

خرید از یک فروشگاه زنجیره ای مؤثر است. از دیگر دلایل افزایش خرید افراد در این فروشگاه ها این است که افراد می توانند مدت زمان بیشتری و ارزش خرید بیشتری را در فروشگاه داشته باشند.

۷-۱-۴. سن

نتایج نشان می دهد افرادی که بیش از ۲۰ سال دارند، امکان خریدشان بیشتر است و افرادی که کمتر از بیست سال دارند، امکان خرید مجددشان اندک است؛ از این رو پیشنهاد می شود برای جذب مشتریان در بازه سنی کمتر، ضمن شناسایی نیازهای مصرفی خریداران (نظیر گروه ها و سبک های کالایی تقاضاشده) به این جامعه از مشتریان به عنوان مشتریان بالقوه آینده نیز توجه شود.

۸-۱-۴. نحوه آشنایی با فروشگاه

به طور کلی، چهار روش آشنایی افراد با فروشگاه در این پژوهش بررسی شده است (جدول ۵).

جدول ۵: روش های آشنایی مشتریان با فروشگاه

Table 5: Ways to Familiarize Customers with the Store

شماره روش	نام روش
۱	دوستان و آشنایان
۲	کارکنان فروشگاه
۳	صفحات مجازی فروشگاه
۴	تبلیغات تلویزیونی

منبع: یافته های پژوهش

نتایج حاکی از آن است که آشنایی افراد با فروشگاه از طریق معرفی دوستان و آشنایان، کارکنان فروشگاه، صفحات مجازی و تبلیغات تلویزیونی به ترتیب بیشترین اهمیت را در آشنایی افراد با فروشگاه و امکان خرید آنها دارد؛ از این رو، پیشنهاد استفاده از ابزار تبلیغات دهان به دهان ارائه می شود. همچنین،

پیشنهاد می شود سرمایه گذاری در تبلیغات با توجه به درجه اهمیت هر یک از شیوه های آشنایی باشد. در این میان، اثر تبلیغات دهان به دهان مهم است. این در حالی است که این اثر سنتی در فضای مجازی نیز نقش بازی کرده است و برنامه ریزان باید به اثر دهان به دهان الکترونیکی توجه کنند.

۵. نتیجه گیری

تحلیل و پیش‌بینی رفتار مشتریان اهمیت فراوانی دارد؛ زیرا هزینه‌ی از دست دادن مشتری برای یک سازمان بسیار زیاد است. در همین راستا، در پژوهش حاضر کوشش شد تا برای پیش‌بینی تحلیل روی گردانی مشتریان، روش ترکیبی شبکه‌ی بیزین و درخت تصمیم C5.0 توسعه داده شود. برای این منظور، داده‌های مشتریان یک فروشگاه زنجیره‌ای در شهرستان مشهد به‌عنوان مطالعه موردی بررسی شد. متغیر احتمال خرید مجدد مشتریان به‌عنوان متغیر وابسته در نظر گرفته و سپس با اهمیت‌ترین متغیرهای مستقل برای پیاده‌سازی درخت تصمیم C5.0 و شبکه‌ی بیزین با گره انتخاب ویژگی واقع در نرم‌افزار IBM SPSS MODELER 18.0 شناسایی شد. نتایج پژوهش حاضر نشان می‌دهد که به‌کارگیری الگوریتم انتخاب ویژگی می‌تواند دقت مدل طبقه‌بندی را بهبود دهد. همچنین، انتخاب بهترین مدل طبقه‌بندی و تمرکز بر روی متغیرها با بالاترین اهمیت بر کیفیت پیش‌بینی روی گردانی مشتریان اثرگذار است. نتایج پژوهش حاکی از آن است که ۸ عامل سن، وضعیت تأهل، متوسط سطح درآمد ماهیانه، تعداد خرید در ماه، نحوه‌ی آشنایی با فروشگاه، نوع بازار، سهم خرید اینترنتی و فروش ویژه از مهم‌ترین عوامل مؤثر بر روی گردانی است. با توجه به مقایسه‌ی دو الگوریتم شبکه‌ی بیزین و درخت تصمیم‌گیری C5.0 که مبتنی بر نتایج نمودار ROC است، نتایج حاصل بر این مسئله تأکید دارد که درخت تصمیم C5.0 با بیشترین دقت، عملکرد بهتری در شناسایی مشتریان روی گردان دارد. در نهایت، در پژوهش حاضر مجموعه‌ای از پیشنهادهای کاربردی مبتنی بر نتایج پژوهش برای تدوین طرح‌های بازاریابی، فروش و مواجهه با انواع مشتریان ارائه شد.

برای انجام دادن پژوهش‌های آتی پیشنهاد می‌شود که طیف گسترده‌تری از متغیرهای اثرگذار مرتبط با مشتریان (هزینه و زمان دسترسی به فروشگاه)، فروشگاه ارائه‌دهنده خدمت (تنوع ارائه محصولات مصرفی و غیرمصرفی، خدمات رفاهی و سرگرمی) و رقبا (کیفیت خدمات ارائه‌شده رقبا و طرح‌های فروش) در توسعه مدل پیش‌بینی در نظر گرفته شود. علاوه بر آن، استفاده از دیگر روش‌های پیش‌بینی و ارزیابی عملکرد آن نیز به محققان پیشنهاد شود.

منابع

- اسماعیلی، مهدی (۱۳۹۲). مفاهیم و تکنیک‌های داده‌کاوی. کاشان: دانشگاه آزاد اسلامی واحد کاشان.
- امامی، لطیف، پوراشرف، یاسان‌اله، و طولابی، زینب (۱۳۹۴). ارائه مدلی برای رویکردانی مشتریان از بانک ملی با استفاده از معادلات ساختاریافته (مطالعه موردی: شعب بانک ملی استان ایلام). مدیریت بازاریابی، ۱۰ (۲۶)، ۲۵-۴۶.
- بحرینی‌نژاد، اردشیر، و سروش، علیرضا (۱۳۸۸). هوشمندی کسب و کار و داده‌کاوی: یک استراتژی برای به‌کارگیری داده‌ها و برگشت سرمایه. تهران: انتشارات ناقوس.
- حبیبی‌پور، اعظم، طالبی، علی، کریمیان، علی‌اکبر، دهقانی، فرهاد، و مختاری، محمدحسین (۱۳۹۶). تعیین روش بهینه پیش‌پردازش داده‌ها به‌منظور افزایش دقت شبیه‌سازی‌های شوری خاک سطحی (مطالعه موردی: منطقه مروت). آب و خاک، ۳۱ (۳)، ۹۱۵-۹۲۸. Doi: 10.22067/jsw.v31i3.55462
- صنعی‌آباده، محمد، محمودی، سینا، و طاهر پرور، محدثه (۱۳۹۱). داده‌کاوی کاربردی. تهران: انتشارات نیاز دانش.

- from national bank by applying structured equations (Case study: National bank branches in Ilam province). *Journal of Marketing Management*, 10(26), 25-46 [In Persian].
- Fenton, N. E., & Neil, M. D. (2007). *Managing risk in the modern world: Applications of Bayesian networks*. London Mathematical Society.
- Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *Journal of Marketing*, 64(3), 65-87. Doi: 10.1509/jmkg.64.3.65.18028
- Gattermann-Itschert, T., & Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Journal of Industrial Marketing Management*, 107, 134-147. Doi: 10.1016/j.indmarman.2022.09.023
- Geiler, L., Affeldt, S., & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. *Journal of Data & Knowledge Engineering*, 142, 102100. Doi: 10.1016/j.datak.2022.102100
- Habibipoor, A., Talebi, A., Karimian, A. A., Dehghani, F., & Mokhtari, M. H. (2017). Investigation of the optimal method of data processing to increase the accuracy of simulation of surface soil salinity (Case study: Marvast). *Water and Soil*, 31(3), 915-928. Doi: 10.22067/jsw.v31i3.55462 [In Persian].
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques* (3rd ed). University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.
- Ismaili, M. (2012). *Data mining concepts and techniques*. Kashan: Islamic Azad University Press [In Persian].
- Keramati, A., Ghaneei, H., Mirmohammadi, S., علیزاده، سمیه، و ملک محمدی، سمیرا (۱۳۹۳). داده کاوی و کشف دانش گام به گام با نرم افزار Clementine. تهران: انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی.
- Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 1-24. Doi: 10.1186/s40537-019-0191-6
- Alizadeh, S., & Malek Mohammadi, S. (2014). *Data mining and knowledge discovery step by step with Clementine software*. Tehran: Khajeh Nasiruddin Tousi University of Technology [In Persian].
- Amin, A., Adnan, A., & Anwar, S. (2023). An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. *Applied Soft Computing*, 137, 110103. Doi: 10.1016/j.asoc.2023.110103
- Baghla, S., & Gupta, G. (2022). Performance evaluation of various classification techniques for customer churn prediction in e-commerce. *Microprocessors and Microsystems*, 94, 104680. Doi: 10.1016/j.micpro.2022.104680
- Bahraini Nejad, A., & Soroush, A. (1388). *Business intelligence and data mining: A strategy for data utilization and return on investment*. Tehran: Naghoos Publication [In Persian].
- Brown, S. A., & Coopers, P. W. (1999). *Customer relationship management: A strategic imperative in the world of e-business*. John Wiley & Sons.
- Chung, B. D., Park, J. H., Koh, Y. J., & Lee, S. (2016). User satisfaction and retention of mobile telecommunications services in Korea. *International Journal of Human-Computer Interaction*, 32(7), 532-543. Doi: 10.1080/10447318.2016.1179083
- Emami, L., Pourashraf, Y., & Toulabi, Z. (2016). A model for customers switching

- customer reviews. *Journal of Air Transport Management*, 83, 101760. Doi: 10.1016/j.jairtraman.2019.101760
- Mohanty, R., & Naga Ratna Sree, C. (2018). Churn and non-churn of customers in banking sector using extreme learning machine. In *Proceedings of the Second International Conference on Computational Intelligence and Informatics: ICCII 2017* (pp. 51-58). Springer Singapore. Doi: 10.1007/978-981-10-8228-3_6
- Mousavi, S. M., Sangari, M. S., & Keramati, A. (2018). An integrative framework for customer switching behavior. *The Service Industries Journal*, 38(15-16), 1067-1094. Doi: 10.1080/02642069.2018.1428955
- Nadkarni, S., & Shenoy, P. P. (2004). A causal mapping approach to constructing Bayesian networks. *Decision Support Systems*, 38(2), 259-281. Doi: 10.1016/S0167-9236(03)00095-2
- Nielsen, T. D., & Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Second Edition. New York, USA: Springer Science & Business Media.
- Pearl, J. (1987). Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence*, 32(2), 245-257. Doi: 10.1016/0004-3702(87)90012-9
- Rosa, N. B. D. C. (2019). *Gauging and foreseeing customer churn in the banking industry: A neural network approach*. Master Thesis. The New University of Lisbon
- Sanii Abadeh, M., Mahmoudi, S., & Taher Parvar, M. (2011). *Applied data mining*. Tehran: Niaz Danesh Publication [In Persian].
- Sayed, H., Abdel-Fattah, M. A., & Kholief, S. (2018). Predicting potential banking customer churn using Apache Spark ML and MLib packages: A comparative study. *International Journal of Advanced*
- M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2, 10. Doi: 10.1186/s40854-016-0029-6
- Khan, Y., Shafiq, S., Naeem, A., Ahmed, S., Safwan, N., & Hussain, S. (2019). Customers churn prediction using artificial neural networks (ANN) in telecom industry. *International Journal of Advanced Computer Science and Applications*, 10(9), 132-142. Doi: 10.14569/IJACSA.2019.0100918
- Kim, S., & Lee, H. (2022). Customer churn prediction in influencer commerce: An application of decision trees. *Procedia Computer Science*, 199, 1332-1339. Doi: 10.1016/j.procs.2022.01.169
- Kjaerulff, U. B., & Madsen, A. L. (2008). Bayesian networks and influence diagrams. *Springer Science & Business Media*, 200, 114. Doi: 10.1007/978-0-387-74101-7
- Kotler, P., & Keller, K. L. (2006). *Marketing management* (12th ed). New Jersey.
- Kumar, V. A., & Elavarasan, N. (2014). A survey on dimensionality reduction technique. *International Journal of Emerging Trends and Technology in Computer Science*, 3(6), 36-41.
- Lin, C. S., Tzeng, G. H., & Chin, Y. C. (2011). Combined rough set theory and flow network graph to predict customer churn in credit card accounts. *Journal of Expert Systems with Applications*, 38(1), 8-15. Doi: 10.1016/j.eswa.2010.05.039
- Ljubičić, K., Merćep, A., & Kostanjčar, Z. (2023). Churn prediction methods based on mutual customer interdependence. *Journal of Computational Science*, 67, 101940. Doi: 10.1016/j.jocs.2022.101940
- Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J. (2020). Text mining approach to explore dimensions of airline customer satisfaction using online

Guangping, Z. (2013). Customer segmentation for telecom with the k-means clustering method. *Information Technology Journal*, 12(3), 409.

Computer Science and Applications, 9(11), 674-677.

- Shobana, J., Gangadhar, C., Arora, R. K., Renjith, P. N., Bamini, J., & devidas Chincholkar, Y. (2023). E-commerce customer churn prevention using machine learning-based business intelligence strategy. *Measurement: Sensors*, 27, 100728. Doi: 10.1016/j.measen.2023.100728
- Shrestha, S. M., & Shakya, A. (2022). A customer churn prediction model using XGBoost for the telecommunication industry in Nepal. *Procedia Computer Science*, 215, 652-661. Doi: 10.1016/j.procs.2022.12.067
- Tomar, D., & Agarwal, S. (2014). A survey on pre-processing and post-processing techniques in data mining. *International Journal of Database Theory and Application*, 7(4), 99-128.
- Tsai, C. F., & Chiou, Y. J. (2009). Earnings management prediction: A pilot study of combining neural networks and decision trees. *Expert Systems with Applications*, 36(3), 7183-7191. Doi: 10.1016/j.eswa.2008.09.025
- Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), 12547-12553. Doi: 10.1016/j.eswa.2009.05.032
- Wu, X., Li, P., Zhao, M., Liu, Y., Crespo, R. G., & Herrera-Viedma, E. (2022). Customer churn prediction for web browsers. *Expert Systems with Applications*, 209, 118177. Doi: 10.1016/j.eswa.2022.118177
- Xue, H., & Lu, W. C. (2011). Research of customer churn prediction model in a supermarket. In *2011 International Conference on E-Business and E-Government (ICEE)* (pp. 1-5). IEEE. Doi: 10.1109/ICEBEG.2011.5886886
- Ye, L., Qiuru, C., Haixu, X., Yijun, L., &

