Journal of Business Intelligence Management Studies
Vol. 10, No.39, Spring 2022
ims.atu.ac.ir
DOI: 10.22054/IMS.2021.63193.2038

ATU PRESS

Original Research
Received: 2021-08-30 Revised: 2021-12-18 Accepted: 2021-12-26
ISSN: 2821-0964 eISSN: 2821-0816

A Model for Customer Churn Management of an Internet Service Provider

Sahar Amiri — M.Sc of Information Technology Management, Tarbiat Modares University, Tehran, Iran.

Alireza Hassanzadeh * — Professor of Information Technology Management Dept., Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran.

Shaghayegh Sahraei — Assistant Professor of Information Technology Management Dept., Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran.

Abstract

Customer churning is one of the most important issues facing Internet Service Providers in a competitive and rapidly saturating market. Due to the high costs associated with attracting new customers, ISPs have turned to a customer retention approach that explicitly seeks to reduce churn. This study has been surveyed the churning of internet service customers in one of the largest telecommunication companies in Iran. In order to predict the churn, customer data has been collected during six months, and their churning behavior has been investigated over a period of one year after. In addition to churn prediction, the most important factors affecting churn have been identified. In the preprocessing step, the "Random Under-Sampling" method is used to balance the data set and the "minimum-Redundancy, Maximum-Relevance" method is used to feature selection. Then, the "Random Forest", "Support Vector Machine" and "K-Nearest Neighbors" algorithms were applied to classify churning and non-churning customers, and the evaluation criteria showed the superiority of the random forest algorithm. The final model, which was obtained from a combination of balancing, feature

* Corresponding Author: ar_hassanzadeh@modares.ac.ir

How to Cite: Amiri, S., Hassanzadeh, A., Sahraei, Sh. (2022). A Model for Customer Churn Management of an Internet Service Provider, *Journal of Business Intelligence Management Studies*, 10(39), 67-95.

selection and classification methods, called the RUS-mRMR-RF model, is considered as an efficient model in customer churn prediction and identifying the most important factors affecting churn. The results of this study provide valuable insights for the company to develop customer retention strategies.

Keywords: Customer Churn Prediction, Feature Importance, Random Forest, Internet Service, Telecommunication Company.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مطالعات مدیریت کسب و کار هوشمند
سال دهم، شماره ۳۹، بهار ۱۴۰۱، ص ۶۷ تا ۹۵
ims.atu.ac.ir
DOI: 10.22054/IMS.2021.63193.2038

مقاله پژوهشی تاریخ ارسال: ۱۴۰۰/۶/۸ تاریخ بازنگری: ۱۴۰۰/۹/۲۷ تاریخ پذیرش: ۱۴۰۰/۱۰/۵
ISSN: 2821-0964 eISSN: 2821-0816

مدلی برای مدیریت رویگردانی مشتریان یک شرکت ارایه‌دهنده سرویس اینترنت

سحر امیری — کارشناس ارشد رشته مدیریت فناوری اطلاعات، دانشگاه تربیت مدرس، تهران، ایران

علیرضا حسن‌زاده * — استاد گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران

شقایق صحرایی — استادیار گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران

چکیده

رویگردانی مشتری یکی از مسایل مهمی است که شرکت‌های ارایه دهنده سرویس اینترنت در بازار رقابتی و به سرعت درحال اشباع با آن روبه‌رو هستند. به دلیل هزینه‌های بالای مرتبط با جذب مشتری جدید، این شرکت‌ها به رویکرد حفظ مشتری که صریحا به دنبال کاهش رویگردانی است، روی آورده‌اند. این تحقیق، رویگردانی مشتریان سرویس‌های اینترنت یکی از بزرگ‌ترین شرکت‌های مخابراتی ایران را مورد بررسی قرار داده است. به منظور پیش‌بینی رویگردانی، داده‌های مشتریان طی شش ماه جمع‌آوری شده و رویگردانی آن‌ها در یک بازه یک ساله بررسی شده است. علاوه بر پیش‌بینی رویگردانی، مهم‌ترین ویژگی‌های موثر در طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها نیز تعیین شده است. در مرحله پیش‌پردازش از روش "کم‌نمونه‌برداری تصادفی" برای متعادل‌سازی مجموعه داده و از روش "حداقل افزونگی، حداکثر ارتباط" برای انتخاب ویژگی استفاده شده است. سپس الگوریتم‌های "جنگل تصادفی"، "ماشین بردار پشتیبان" و "کا نزدیک‌ترین همسایگان" برای طبقه‌بندی مشتریان رویگردان و غیررویگردان به کار رفتند که معیارهای ارزیابی، نشان دهنده برتری الگوریتم جنگل تصادفی است. مدل نهایی که از ترکیب

مقاله حاضر برگرفته از پایان‌نامه کارشناسی ارشد رشته مدیریت فناوری اطلاعات دانشگاه تربیت مدرس است.

* نویسنده مسئول: : ar_hassanzadeh@modares.ac.ir

روش‌های متعادل‌سازی، انتخاب ویژگی و طبقه‌بندی به دست آمد تحت عنوان مدل RUS-mRMR-RF به عنوان یک مدل کارآمد در پیش‌بینی رویگردانی مشتریان و شناسایی مهم‌ترین ویژگی‌های موثر در طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها محسوب می‌گردد. نتایج این مطالعه بینش ارزشمندی را جهت تدوین استراتژی‌های حفظ مشتری به سازمان ارایه می‌دهد.

کلیدواژه‌ها: پیش‌بینی رویگردانی مشتریان، ویژگی‌های موثر، جنگل تصادفی، سرویس اینترنت، شرکت مخابراتی.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقدمه

امروزه به دلیل رقابت شدید ناشی از بازارهای اشباع شده، شرایط پویای بازار و معرفی مداوم پیشنهادهای جدیدِ رقابتی، بسیاری از شرکت‌ها از مشکل اساسی رویگردانی مشتری رنج می‌برند. در پاسخ، بسیاری از آن‌ها به جای یک استراتژی پیشنهاد محور، که برای فروش هرچه بیشتر پیشنهادات طراحی شده است، به رویکرد حفظ مشتری که به دنبال کاهش رویگردانی است، روی آورده‌اند (Coussement et al., 2017). مدیریت رویگردانی، شناسایی و تلاش در جهت حفظ مشتریانی است که قصد انتقال روابط خود به یک ارایه‌دهنده‌ی خدمات رقیب را دارند (Hadden et al., 2007). برای شرکت‌ها و مدیران روشن شده است که اطلاعات و سوابق مشتری که در پایگاه داده مشتریان موجود می باشد، یکی از مهمترین دارایی‌ها برای مقابله با رویگردانی است، زیرا می‌تواند برای ایجاد مدل استفاده شود (Caigny et al., 2018). پیش‌بینی رویگردانی را می‌توان به عنوان روشی توصیف کرد که به شناسایی زودتر رویگردان‌های احتمالی کمک می‌کند (Dahiya & Bhatia, 2015). پس از شناسایی مشتریان رویگردانِ بالقوه، این مشتریان می‌توانند هدف کمپین‌های بازاریابی فعال برای تلاش در جهت حفظ قرار گیرند (Hadden et al., 2007).

پس از یک دوره‌ی رشد سریع خدمات اینترنت، به نظر می‌رسد بازار به آستانه‌ی مرحله بلوغ رسیده است و شرکت‌های ارایه‌دهنده خدمات، در این مرحله متقاعد شده‌اند که حفظ مشتریان موجود مهم‌تر از جذب مشتریان جدید است (Li et al., 2006). مطالعه در زمینه‌ی پیش‌بینی رویگردانی به صنعت مخابرات و بخش مدیریت ارتباط با مشتریان کمک می‌کند تا افراد متمایل به رویگردانی را، قبل از آن‌که اقدام به ترک شبکه نمایند، شناسایی کنند. مساله اصلی، بحث طبقه‌بندی است؛ طبقه‌بندی هر مشترک به‌عنوان یک رویگردان بالقوه یا غیررویگردان بالقوه (Dahiya & Bhatia, 2015). طبقه‌بندی مشتریان و همچنین کشف مهم‌ترین ویژگی[۱]های موثر در جداسازی مشتریان رویگردان و

۱. Feature

غیررویگردان می‌تواند منجر به تصمیم‌گیری استراتژیک و برنامه‌ریزی کارآمد در صنعت مخابرات شود. به عنوان مثال، براساس قوانین استخراج شده، تصمیم‌گیرندگان می‌توانند برنامه‌های جدید نگهداری را طراحی و اتخاذ کنند و عملکرد کلی سازمان را بهبود بخشند و با شناخت و تجزیه و تحلیل ویژگی‌های مهم رویگردان‌ها و غیررویگردان‌ها، سیاست حفظ مناسب را تدوین نمایند (Amin et al., 2017).

این تحقیق با هدف ارایه یک مدل کارآمد با استفاده از الگوریتم‌های یادگیری ماشین جهت پیش‌بینی رویگردانی مشتریان شرکت مورد بررسی و شناسایی ویژگی‌هایی که بیشترین تاثیر را در طبقه‌بندی و جداسازی مشتریان رویگردان و غیررویگردان دارند، انجام شده است.

پیشینه پژوهش

در ادبیاتِ رویگردانی، داده‌کاوی برای بهبود تلاش‌های بازاریابی، شناسایی تقلب و مدیریت بهتر شبکه، نقش بسیار مهمی در شرکت‌های ارتباطی و مخابراتی ایفا کرده است. (Dahiya & Bhatia, 2015). بررسی رویگردانی مشتری موضوع بسیاری از تکنیک‌ها و روش‌ها از جمله روش‌های آماری و داده‌کاوی بوده است. برتری استفاده از داده‌کاوی برای بررسی رویگردانی مشتری در مقایسه با نظرسنجی‌های تحقیقات بازار کاملا ثابت شده است. نظرسنجی‌های تحقیقات بازار براساس پرسشنامه‌ها یا مصاحبه‌های داوطلبانه صورت می‌گیرند و از هزینه‌ی بالا، دسترسی محدود به جمعیت مشتریان و خودگزارش‌دهی اطلاعات رنج می‌برند. برعکس، داده‌کاوی، دانش کل جمعیت مشتری را بر اساس تجزیه و تحلیل داده‌های فعلی و تاریخی آن‌ها فراهم می‌کند (Hassouna et al., 2016).

تحقیقات قبلی نشان داده که نتایج مدل‌های پیش‌بینی‌کننده می‌تواند در تصمیم‌گیری بهتر به مدیران کمک کند. طبق نتایج عینی، شرکت‌هایی که سیستم‌های تصمیم‌گیری مبتنی بر داده را به کار می‌گیرند، در اقدامات مالی و عملیاتی عملکرد موفق‌تری دارند (Caigny et al. 2020). از طرف دیگر، هوشمندی کسب‌وکار به عنوان یک فن‌آوری شناخته شده می‌تواند به سازمان‌ها در کسب، جذب و انتقال دانش جدید کمک کند. از

این نظر، تجزیه‌وتحلیل کسب‌وکار مولفه‌ای از هوشمندی کسب‌وکار است که بر استفاده از تکنیک‌های تحلیلی برای پاسخ به سوالات سازمانی و بهبود تصمیم‌گیری متمرکز است (Torres et al., 2018). سیستم‌های هوشمند می‌توانند استراتژی‌های تجاری را برای کمک به مدیریت سازمان در کاهش رویگردانی مشتری فراهم کنند. در این راستا، داده‌کاوی و یادگیری ماشین به عنوان ابزاری مهم در هوشمندی کسب‌وکار در نظر گرفته می‌شوند که می‌توانند از رفتارهای قبلی مشتریان به منظور پیش‌بینی الگوها، رفتارها و روندهای آینده استفاده کنند (Khodabandehlou & Rahman, 2017).

در ادبیات روش‌های موثر بسیاری برای رسیدگی به مساله مدیریت رویگردانی مشتری وجود دارد. روش‌های تحلیلی عمدتا شامل مدل‌های آماری، یادگیری ماشین و داده‌کاوی است (Bi et al., 2016). با تأمل در تحقیقات قبلی، می‌توان ادبیات بررسی رویگردانی مشتریان را بر اساس دو جنبه‌ی اصلی دسته‌بندی کرد. نخست، مطالعاتی که به طور عمده به عوامل تعیین کننده رویگردانی مربوط می باشند. این مطالعات، عوامل تعیین کننده شناخته شده رویگردانی را با استفاده از رفتارهای مشتریان تجزیه‌وتحلیل کرده و مورد بررسی قرار داده‌اند. دسته دیگری از مطالعات وجود دارند که علاقه نویسندگان آن‌ها به بهبود نتیجه مدل‌های پیش‌بینی رویگردانی با استفاده از روش‌های محاسباتی پیشرفته بوده است (Keramati et al., 2014). در این بخش به مرور برخی از پژوهش‌های حوزه رویگردانی مشتریان پرداخته می‌شود:

کرامتی و همکاران[1] (۲۰۱۴) از روش‌های طبقه‌بندی داده‌کاوی در پیش‌بینی رویگردانی مشتریان تلفن همراه یک شرکت مخابراتی استفاده نمودند؛ در این مطالعه دفعات استفاده، تعداد کل شکایات و مدت زمان استفاده به عنوان مهم‌ترین موارد موثر در رویگردانی مشتریان شناسایی گردیدند (Keramati et al., 2014). کورون و همکاران[2] (۲۰۱۸) آزمایشی را بر روی ۲۴۳ مجموعه داده‌ی واقعی برای امور مختلف پیش‌بینی انجام

۱. Keramati et al.

۲. Couronné et al.

دادند که مقایسه‌ای از عملکرد پیش‌بینی‌کنندگی دو الگوریتم جنگل تصادفی و رگرسیون لجستیک، به عنوان ابزارهای طبقه‌بندی باینری می‌باشد. این مطالعه، به طور میانگین عملکرد خوب پیش‌بینی جنگل تصادفی در مقایسه با رگرسیون لجستیک را در تقریبا ۶۹ درصد از مجموعه داده‌ها نشان می‌دهد (Couronné et al., 2018). اولاه و همکاران[۱] (۲۰۱۹) یک مدل پیش‌بینی رویگردانی را ارایه کرده‌اند که از طبقه‌بندی و همچنین تکنیک‌های خوشه‌بندی برای شناسایی مشتریان رویگردان در بخش مخابرات استفاده می‌کند. در این پژوهش تکنیک‌های رگرسیون لجستیک، درختان تصمیم، جنگل‌های تصادفی و درختان مدل لجستیک اعمال شده است و مشاهده شده که الگوریتم جنگل تصادفی با ۸۸/۶۳ درصد موارد طبقه‌بندی شده‌ی صحیح و ۸۸ درصد معیار اندازه‌گیری F، عملکرد بهتری داشته است (Ullah et al., 2019). در مطالعه جِین و همکاران[۲] (۲۰۲۰) طیف گسترده‌ای از تحقیقات در زمینه رویگردانی از سال ۲۰۰۵ تا ۲۰۲۰ بررسی شده است که شامل انواع روش‌های پیشنهادیِ تحقیقات گذشته و فن‌آوری‌های مورد استفاده در این تحقیقات است. این مطالعه نشان می‌دهد که روش قالب برای پیش‌بینی رویگردانی استفاده از تکنیک‌های یادگیری ماشین است. همچنین روش‌های انتخاب ویژگی بسیار اساسی هستند و برای افزایش عملکرد مدل، باید در پژوهش‌ها گنجانده شوند (Jain et al., 2021). وُ و همکاران[۳] (۲۰۲۱) یک مدل پیش‌بینی رویگردانی مشتری با استفاده از داده‌های غیرساختاریافته سوابق تماس مشتری، که همان مطالب گفتاری در ارتباطات تلفنی است، پیشنهاد داده‌اند. نتایج نشان می‌دهد که مدل ارایه شده می‌تواند ریسک رویگردانی مشتری را با دقت خوبی پیش‌بینی کند (Vo et al., 2021). وو و همکاران[۴] (۲۰۲۱) در پژوهشی یک چارچوب یکپارچه تجزیه‌وتحلیل مشتری برای مدیریت رویگردانی ارایه داده‌اند. در این مطالعه ابتدا وضعیت رویگردانی مشتریان پیش‌بینی شده است سپس به

۱. Ullah et al.

۲. Jain et al.

۳. Vo et al.

۴. Wu et al.

کشف برخی ویژگی‌های مهم برای تقسیم‌بندی مشتریان رویگردان پرداخته شده است (Wu et al., 2021). کاظمی و حجازی‌نیا (۱۳۹۲) مروری بر پژوهش‌های انجام شده در خصوص متغیرهای تاثیرگذار بر رویگردانی مشتریان شرکت‌های مخابراتی، با هدف ارایه یک مدل جامع برای پیش‌بینی رویگردانی مشتریان انجام داده‌اند. نتیجه‌ی این تحقیق نشان می‌دهد که عواملی از جمله کیفیت، هزینه، تبلیغات، رضایت، اطلاعات جمعیت شناختی، ورود رقبا با فن‌آوری برتر و نگرانی‌های امنیتی در رویگردانی مشتریان شرکت‌های ارایه‌دهنده‌ی خدمات مخابراتی موثر می‌باشد (کاظمی و حجازی‌نیا، ۱۳۹۲). عاشوری و البدوی (۱۳۹۴) در مطالعه‌ای عوامل تاثیرگذار بر رویگردانی مشتریان یک شرکت ارایه‌دهنده‌ی خدمات اینترنت[۱] را ارزیابی کرده‌اند و از نتایج به‌دست آمده برای تحلیل تاثیر پارامترهای مختلف بر رویگردانی بهره گرفته‌اند. بر اساس نتایج این مطالعه، متغیر مدت زمان نصب و راه‌اندازی بیشترین تاثیر را بر رفتار رویگردانی مشتریان شرکت‌های ارایه دهنده‌ی خدمات اینترنتی داشته است (عاشوری و البدوی، ۱۳۹۴). مطالعه کاظمی و حجازی‌نیا (۱۳۹۵) با جمع‌آوری داده‌های مشتریان شرکت همراه اول از طریق پرسشنامه‌ی محقق ساخته، و استفاده از آزمون‌های تی تک نمونه‌ای، تحلیل رگرسیون و تحلیل واریانس یک طرفه نشان داده است که به جز دو متغیر قیمت خدمات و نگرانی‌های امنیتی، سایر متغیرها یعنی نارضایتی مشتری، کیفیت خدمات، هزینه تغییر، تبلیغات و ورود رقبا با فن‌آوری برتر، و همچنین برخی از ویژگی‌های جمعیت شناختی و کاربری مشتریان (سن، شغل، زمان خریداری سیم کارت و تعداد سیم کارت) با رویگردانی مشتریان رابطه معناداری دارند (کاظمی و حجازی‌نیا، ۱۳۹۵). خانلری و همکاران (۱۳۹۵) روشی برای پیش‌بینی رویگردانی مشتریان یک شرکت ایرانی ارایه دهنده خدمات اینترنت ارایه کرده‌اند. به عنوان عوامل موثر بر رویگردانی مشتریان، داده‌هایی شامل اطلاعات دموگرافیک، رفتار خرید، رفتار مصرف و رضایت مشتریان مورد بررسی قرار گرفته است. این مطالعه با استفاده از یک روش ترکیبی و با تمرکز بر بهینه‌سازی انتخاب ویژگی، انجام

۱. Internet Service Provider (ISP)

شده است و برای پیش‌بینی رویگردانی از ترکیبی از الگوریتم‌های شبکه عصبی مصنوعی و درخت تصمیم استفاده شده است (خانلری و همکاران، ۱۳۹۵). سپهری و کوشا (۱۳۹۶) با تحقیق بر روی داده‌های یک شرکت مخابراتی در ایران با استفاده از تکنیک ماشین بردار پشتیبان، مدلی برای پیش‌بینی رویگردانی مشتریان ارایه کرده‌اند (سپهری و کوشا، ۱۳۹۶). عسگری و همکاران (۱۳۹۸) در مطالعه‌ای به معرفی مفهوم زنجیره وضعیت که تغییرات وضعیت رویگردانی جزئی مشتریان طی زمان را معرفی می‌کند، پرداخته‌اند و با به‌کارگیری آن و استفاده از روش خوشه‌بندی سلسله‌مراتبی و ماشین‌های بردار پشتیبان، یک مدل پیش‌بینی رویگردانی جزئی مشتریان بانکی ارایه کرده‌اند. نتایج این مطالعه نشان می‌دهد با افزایش طول زنجیره‌های وضعیت، مدل به دقت بالایی دست پیدا می‌کند (عسگری و همکاران، ۱۳۹۸).

روش

داده‌کاوی به روشی استاندارد نیاز دارد تا کمک کند مشکلات کسب‌وکار در قالب مسایل داده‌کاوی دربیاید، تغییرات مناسب در داده‌ها و تکنیک‌های داده‌کاوی را پیشنهاد دهد و ابزاری برای ارزیابیِ اثربخشیِ نتایج و مستندسازی تجربیات فراهم کند. روش فرآیند استاندارد صنعتی برای داده‌کاوی[۱]، که به طور مختصر کریسپ دی اِم نامیده می‌شود، به راهکاری برای حل این مشکلات پرداخته است. هدف از ارایه این روش، پروژه‌های داده‌کاوی بزرگ‌تر، کم‌هزینه‌تر، قابل اعتمادتر، تکرارپذیرتر، قابل کنترل‌تر و سریع‌تر است (Wirth & Hipp, 2000; Mariscal et al., 2010).

کریسپ دی اِم فرآیند داده‌کاوی را به شش مرحله اصلی تقسیم می‌کند (Harper & Pickett, 2006):

۱. شناخت کسب و کار[۲]

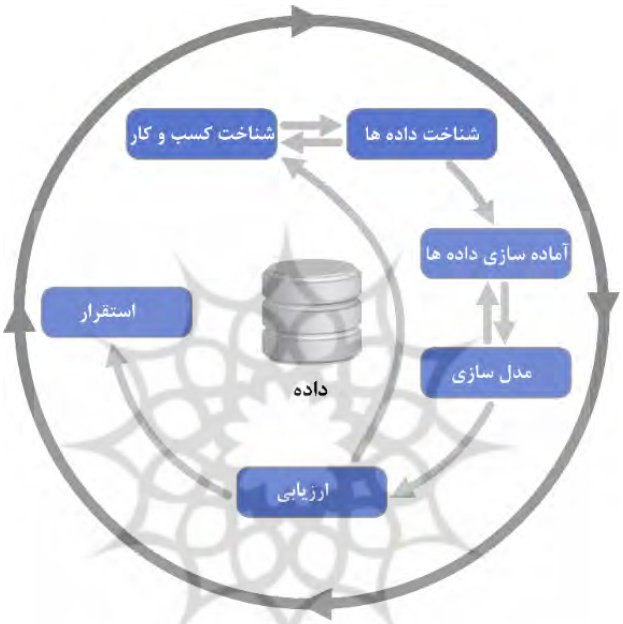
۲. شناخت داده‌ها[۱]

---

۱. Cross-Industry Standard Process for Data Mining (CRISP-DM)

2. Business Understanding

۳. آماده‌سازی داده‌ها[۲]
۴. مدل سازی[۳]
۵. ارزیابی[۴]
۶. استقرار[۵]



شکل ۱. نمودار فرآیند کریسپ دی ام (Wirth & Hipp, 2000)

شناخت کسب‌وکار

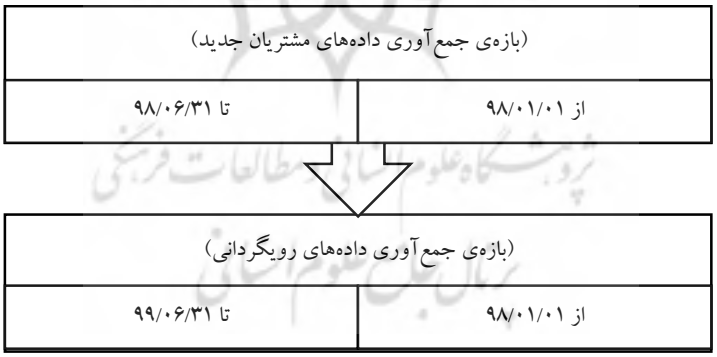
در این پژوهش به مساله روی گردانی مشتریان سرویس‌های اینترنت پس‌پرداخت[۶] یکی از شرکت‌های مخابراتی ایران پرداخته شده است. این مشتریان تا زمان برقراری سرویس

1. Data Understanding
2. Data Preparation
3. Modeling
4. Evaluation
5. Deployment
6. Post Paid

اینترنت بر روی خط تلفن، ملزم به پرداخت هزینه هستند؛ لذا درخواست جمع‌آوری سرویس از طرف مشتری و قطع آن، به عنوان رویگردانی تعریف می‌گردد. بنا به نظر صاحب‌نظران سازمان، بازه‌ی یک ساله برای بررسی رویگردانی مشتریان در نظر گرفته شده است، به این ترتیب مشتریانی که در طول یک سال پس از برقراری سرویس اقدام به قطع آن کرده‌اند به عنوان مشتریان رویگردان و آنان که در این بازه زمانی به حفظ رابطه با شرکت ادامه داده‌اند به عنوان مشتریان غیررویگردان شناخته شده‌اند.

شناخت داده‌ها

مجموعه داده تحقیق مربوط به مشتریانی است که در شش ماهه اول سال ۱۳۹۸ اقدام به خرید خدمات اینترنت پس‌پرداخت از شرکت مورد بررسی نموده‌اند. داده‌های ۷۳۸۳۵ مشتری از پایگاه داده شرکت استخراج شد، سپس رویگردانی یا عدم رویگردانی این مشتریان تا یک سال پس از برقراری سرویس بررسی شد و مشخص شد از این تعداد ۵۸۱۵ مشتری رویگردان شده‌اند یعنی تقریبا ۷/۸ درصد مشتریان مورد نظر در زمانی کم‌تر از یک سال اقدام به قطع خدمات اینترنت خود نموده‌اند.



شکل ۲. جدول زمانی مدل پیش‌بینی رویگردانی

آماده‌سازی داده‌ها

کارآیی مدل پیش‌بینی رویگردانی، بر اساس سیستم طبقه‌بندی[1]، به یادگیری حاصل از مجموعه داده‌های موجود متکی است. مجموعه داده پیش‌پردازش شده مناسب به طبقه‌بند[2] کمک می‌کند تا به سطح آموزش مورد نیاز برسد، که در نهایت به عملکرد مطلوبی تبدیل می‌شود. شرکت‌های مخابراتی اطلاعات زیادی را در مورد مشتریان به‌دست آورده و ذخیره می‌کنند، مانند اطلاعات جمعیتی، نوع قرارداد، صورتحساب‌ها و پرداخت‌ها، گزارش خدمات و غیره. چنین داده‌هایی دارای ابعاد بالا و توزیع نامتوازن هستند. این موضوع ممکن است باعث یادگیری ضعیف طبقه‌بند شود. بنابراین، مرحله پیش‌پردازش، اساسا به یک نمونه‌برداری مناسب و استراتژیِ کاهش ویژگی برای انجام یادگیری خوب توسط طبقه‌بند نیاز دارد (Idris et al., 2012). محصول پیش‌پردازش داده‌ها، مجموعه داده نهایی است که به عنوان مجموعه آموزش جهت یادگیری الگوریتم طبقه‌بند به کار می‌رود (Kotsiantis et al., 2006).

متعادل‌سازی داده‌ها

عدم تعادل کلاس، یک ویژگی ذاتی مجموعه داده‌های پیش‌بینی رویگردانی است. شناسایی گروه رویگردانِ اقلیت از تعداد زیادی غیررویگردان، هسته اصلی مدل‌سازی رویگردانی موفق محسوب می‌شود زیرا اکثر الگوریتم‌های طبقه‌بندی استاندارد، غالبا نسبت به کلاس اکثریت، سوگیری نشان می‌دهند (Zhu et al., 2017). عدم تعادل کلاس شناختن گروه اقلیت را برای برخی از تکنیک‌های یادگیری ماشین با مشکل مواجه می‌کند، اگرچه ممکن است به دقت کلی بالایی دست پیدا کنند (Amin et al., 2017).

یکی از متداول‌ترین تکنیک‌های مقابله با کمیاب بودن داده‌ها، نمونه‌برداری مجدد است (لو و همکاران، ۲۰۱۲). سه روش اساسی نمونه‌برداری مجدد شامل بیش‌نمونه‌برداری تصادفی[3]، کم‌نمونه‌برداری تصادفی[4] و بیش‌نمونه‌برداری اقلیت مصنوعی[1] است (Gui,

1. cassification
2. classifier
3. Random Oversampling (ROS)
4. Random Undersampling (RUS)

2017; Yap et al., 2014) زمانی که به اندازه کافی نمونه‌های کلاس اقلیت در دسترس باشند، کم‌نمونه‌برداری از کلاس اکثریت به عنوان وسیله خوبی برای افزایش حساسیت یک طبقه‌بند نسبت به کلاس اقلیت محسوب می‌گردد. برای نمونه‌برداری از مجموعه داده‌های بزرگ مخابراتی نیز معمولا روش کم‌نمونه‌برداری تصادفی به کار می‌رود (Chawla et al., 2002). در این پژوهش نیز از این روش استفاده شده است.

انتخاب ویژگی

انتخاب ویژگی، فرآیند شناسایی ویژگی‌های موثر و حذف موارد اضافی، نویزی و دارای اطلاعات کم، برای بهبود یک مدل پیش‌بینی است (Hadden et al., 2007). تکنیک‌های کاهش ویژگی با استخراج مرتبط‌ترین متغیرها، که بیشترین قدرت برای ایجاد تمایز بین رویگردان‌ها و غیررویگردان‌ها را دارند (انتخاب متغیر)، یا انتخاب نماینده‌ترین مشتریان (نمونه‌گیری)، کاهش ابعادی مجموعه داده‌ها را در پی دارند (Coussement et al., 2017).

حداقل افزونگی و حداکثر ارتباط[2] یک روش انتخاب ویژگی با پیچیدگی محاسباتی کم و هزینه اندک است که زیرمجموعه مناسبی از ویژگی‌های منتخب را پیدا می‌کند و یک مطالعه جامع نشان داده که این روش دقت طبقه‌بندی را به میزان قابل توجهی بهبود می‌بخشد (Peng et al., 2005). در این پژوهش نیز از این روش برای انتخاب ویژگی استفاده شده است.

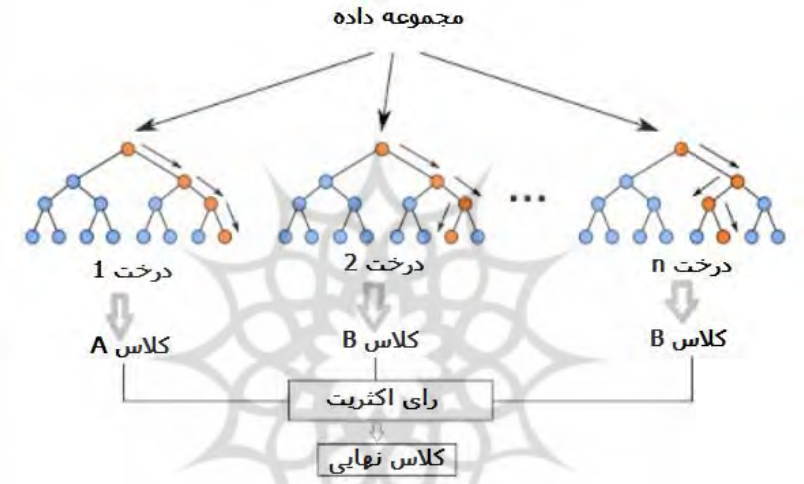
مدل‌سازی

مساله اصلی در مرحله مدل‌سازی انتخاب یک الگوریتم طبقه‌بند مناسب است که بین قابل درک بودن و عملکرد پیش‌بینی‌کنندگیِ مدل تعادل خوبی برقرار کند (Caigny et al., 2018). از منظر شناخت مدل، پیش‌بینی رویگردانی را می‌توان به عنوان یک مساله

1. Synthetic Minority Over-Sampling (SMOTE)
2. Minimum Redundancy and Maximum Relevance (mRMR)

یادگیری نظارت شده[1] تعریف کرد یعنی با توجه به افق پیش‌بینیِ از پیش تعیین شده، رویگردان‌های آینده را پیش‌بینی می‌کند (Xiao et al., 2016).

الگوریتم جنگل تصادفی، که توسط لئو بریمن[2] در سال ۲۰۰۱ پیشنهاد شد، به عنوان یک روش طبقه‌بندی و رگرسیونی همه منظوره، بسیار موفق بوده است. در این روش، چندین درخت تصمیم تصادفی با هم ترکیب می‌شوند و پیش‌بینی‌های آن‌ها توسط میانگین‌گیری تجمیع می‌گردد (Speiser et al., 2019).



شکل ۳. دیاگرام جنگل تصادفی[3]

جنگل تصادفی به اندازه کافی تطبیق‌پذیر است و می‌تواند برای مسایلی در ابعاد بزرگ استفاده شود؛ به راحتی با مسایل مختلف یادگیری سازگار می‌شود (Biau & Scornet, 2016) و میزان اهمیت ویژگی‌های[4] موثر در پیش‌بینی را ارزیابی می‌کند (Liaw & Wiener, 2002; Couronné et al., 2018).

1. Supervised Learning
2. Leo Breiman 1928-2005
3. https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page
4. Feature Importance

جنگل تصادفی ابزاری موثر در پیش‌بینی است زیرا دچار بیش‌برازش نمی‌شود و تزریق نوع صحیحی از تصادفی بودن باعث می‌شود طبقه‌بندی دقیق و تنظیم کننده‌ای داشته باشد (Breiman, 1999). همچنین انواع متغیرهای ترکیبی را به خوبی کنترل می‌کند، یعنی می‌تواند متغیرهای طبقه‌ای و ترتیبی[1] را به روشی ساده و طبیعی کنترل کند. این ویژگی را از پیش‌بینی کننده‌های درختی به ارث می‌برد (Breiman et al., 1984).

با توجه به این توضیحات انتظار می‌رود الگوریتم جنگل تصادفی انتخاب مناسبی برای رسیدن به اهداف این پژوهش باشد. با این حال برای اطمینان از کارایی آن، دو الگوریتم ماشین بردار پشتیبان[2] و کا نزدیکترین همسایگان[3] نیز که از پرکاربردترین الگوریتم‌ها در زمینه طبقه‌بندی و پیش‌بینی هستند، بر روی مجموعه داده اعمال و نتایج مقایسه شده‌اند. لازم به ذکر است در این پژوهش برای پیاده‌سازی الگوریتم‌ها از کدهای نوشته شده در زبان برنامه‌نویسی پایتون[4] استفاده شده است.

ارزیابی

عملکرد الگوریتم‌های یادگیری ماشین به‌طور معمول توسط ماتریس درهم‌ریختگی[5] ارزیابی می‌شود، همان‌طور که در جدول ۱ نشان داده شده است، برای مساله‌ای که دارای دو طبقه (کلاس) است، ستون‌ها کلاس "پیش‌بینی" و ردیف‌ها کلاس "واقعی" هستند. به این ترتیب جایگاه موارد مثبت واقعی[6] و مثبت کاذب[7] و همچنین منفی واقعی[8] و منفی کاذب[9] در ماتریس درهم‌ریختگی مشخص می‌گردند (Keramati et al., 2014; Bahari & Elayidom, 2015; Vafeiadis et al., 2015). در این پژوهش، مثبت واقعی،

1. Ordered
2. Support Vector Machine
3. K-Nearest Neighbors
4. Python
5. Confusion Matrix
6. True Positive
7. False Positive
8. True Negative
9. False Negative

رویگردان‌ها هستند. مثبت کاذب، غیر رویگردان‌هایی هستند که به اشتباه توسط مدل به عنوان رویگردان طبقه‌بندی شده‌اند. منفی واقعی، غیررویگردان‌ها هستند و منفی کاذب نیز رویگردان‌هایی هستند که به اشتباه توسط مدل به عنوان غیررویگردان طبقه‌بندی شده‌اند.

جدول ۱. ماتریس درهم‌ریختگی برای ارزیابی طبقه‌بندی رویگردانی (Vafeiadis et al., 2015)

		کلاس پیش‌بینی شده	
		رویگردان‌های پیش‌بینی شده	**غیررویگردان‌های پیش‌بینی شده**
کلاس واقعی	**رویگردان‌های واقعی**	TP	FN
	غیررویگردان‌های واقعی	FP	TN

بر اساس آرایه‌های ماتریس درهم‌ریختگی، مهم‌ترین معیارهای ارزیابی به صورت زیر تعریف می‌شوند:

صحت[1]: به نسبت تعداد موارد طبقه بندی شده‌ی صحیح اشاره دارد و برابر است با تعداد کل پیش‌بینی‌های درست یعنی مجموع مثبت واقعی و منفی واقعی، تقسیم بر مجموع پیش‌بینی‌های انجام شده (چه غلط و چه درست) (Bahari & Elayidom, 2015; Vafeiadis et al., 2015):

$$\text{صحت} = \frac{TP + TN}{TP + FP + TN + FN}$$

حساسیت[2] **(یا فراخوانی**[3]**)**: برابر است با نسبت موارد مثبتی که به درستی شناسایی شده‌اند به تعداد کل موارد مثبت و به میزان مثبتی که به درستی طبقه‌بندی شده (نرخ مثبت واقعی) اشاره دارد (Bahari & Elayidom, 2015; Vafeiadis et al., 2015):

$$\text{فراخوانی} = \frac{TP}{TP + FN}$$

1. Accuracy
2. Sensitivity
3. Recall

تشخیص‌پذیری[1]: برابر است با تعداد موارد منفی واقعی تقسیم بر مجموع منفی واقعی و مثبت کاذب و به میزان منفی طبقه‌بندی شده صحیح (نرخ منفی واقعی) اشاره دارد (Bahari & Elayidom, 2015):

$$\text{تشخیص‌پذیری} = \frac{TN}{TN + FP}$$

معیار F[2] **یا F1**[3]: دقت و فراخوانی به تنهایی نمی‌توانند کارایی یک طبقه‌بند را توصیف کنند زیرا عملکرد خوب در یکی از این شاخص‌ها لزوما به معنی عملکرد خوب در دیگری نیست. به همین دلیل، معیار F1، ترکیب محبوبی است که معمولا به عنوان یک معیار واحد برای ارزیابی عملکرد طبقه‌بند استفاده می‌شود. معیار F1 به عنوان میانگین هارمونیک دقت و فراخوانی تعریف می‌شود (Keramati et al., 2014; Vafeiadis et al., 2015):

$$\text{معیار F1} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

مقداری نزدیک به یک بیانگر این است که دقت و فراخوانیِ ترکیبیِ بهتری توسط طبقه‌بند حاصل شده است (Vafeiadis et al., 2015).

منحنی مشخصه عملیاتی گیرنده[4]: منحنی مشخصه عملیاتی گیرنده را می‌توان نماینده خانواده بهترین مرزهای تصمیم‌گیری برای هزینه‌های نسبی مثبت واقعی و مثبت کاذب دانست. این منحنی یک روش استاندارد برای خلاصه کردن عملکرد طبقه‌بند در طیف وسیعی از داد و ستدها بین نرخ خطای مثبت واقعی و مثبت کاذب است. محور افقی منحنی نشان دهنده درصد مثبت کاذب:

1. Specificity
2. F-Measure
3. F1-Score
4. Receiver Operating Characteristic (ROC)

$$\%\text{FP} = \frac{FP}{TN + FP}$$

و محور عمودی نمایانگر درصد مثبت واقعی است:

$$\%\text{TP} = \frac{TP}{TP + FN}$$

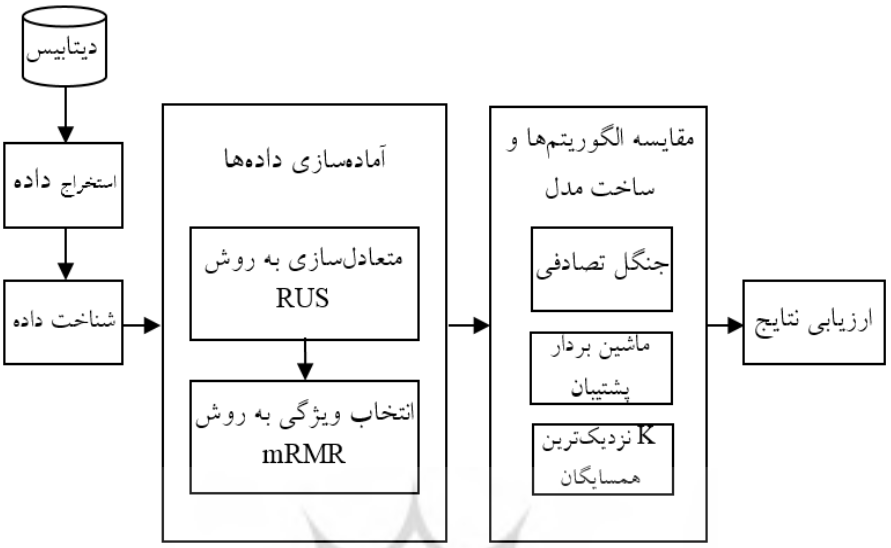
نقطه ایده‌آل روی منحنی نقطه (۱و۰) خواهد بود (یا نقطه‌ی (۱۰۰و۰) درصد) که همه مثال‌های مثبت به درستی طبقه‌بندی شده‌اند و هیچ نمونه منفی به عنوان مثبت، طبقه‌بندی نشده است. سطح زیر منحنی[1] نیز یک معیار عددی واحد از نمودار گرافیکی حساسیت، یا نرخ مثبت واقعی، در مقابل نرخ مثبت کاذب برای یک سیستم طبقه‌بندی باینری است و معیاری مفید برای تجسم و ارزیابی عملکرد طبقه‌بند به حساب می‌آید (Idris et al., 2012).

گام‌های تحقیق

با توجه به توضیحاتی که داده شد، گام‌های تحقیق را می‌توان به صورت شکل ۴ نشان داد:

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Area Under the Curve (AUC)



شکل ٤. گام‌های تحقیق

یافته‌ها

با اعمال روش‌های متعادل‌سازی و انتخاب ویژگی، یک مجموعه داده متعادل شامل ۱۲ ویژگی نهایی به دست آمد. الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان و کا نزدیک‌ترین همسایگان بر روی این مجموعه داده پیاده‌سازی شدند و نتایج عملکرد هرکدام در قالب معیارهای ارزیابی به دست آمد که در جدول ۲ نشان داده شده است.

جدول۲. نتایج معیارهای ارزیابی الگوریتم‌ها

الگوریتم	صحت	حساسیت	تشخیص پذیری	معیار F1	AUC
جنگل تصادفی	0.750	0.726	0.773	0.749	0.843
ماشین بردار پشتیبان	0.528	0.485	0.569	0.500	0.629
کا نزدیک‌ترین همسایگان	0.505	0.490	0.521	0.505	0.659

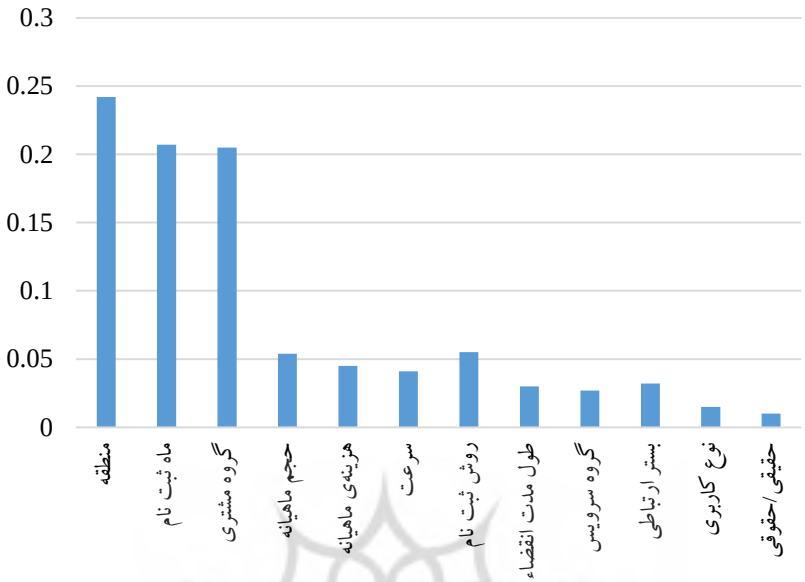
نتایج به دست آمده حاکی از برتری عملکرد الگوریتم جنگل تصادفی بر روی مجموعه داده تحقیق، در مقایسه با دو الگوریتم ماشین بردار پشتیبان و کا نزدیک‌ترین همسایگان

است. معیارهای ارزیابی نشان می‌دهد مدل نهایی برای پیش‌بینی و تعیین مهم‌ترین ویژگی‌های موثر بر طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها، که به صورت ترکیبی از روش‌های پیش‌پردازش و طبقه‌بندی، RUS-mRMR-RF نامگذاری می‌شود، توانسته با استفاده از داده‌های تحقیق، جداسازی خوبی از مشتریان رویگردان و غیررویگردان انجام دهد.

با استفاده از مدل ساخته شده، میزان اهمیت هریک از ویژگی‌های موثر بر طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها در شرکت مورد بررسی مطابق جدول۳ و نمودار ۱ به دست آمد:

جدول ۳. میزان اهمیت ویژگی‌های موثر بر طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها

ترتیب اهمیت	ویژگی	میزان اهمیت
۱	منطقه	۰/۲۴۲
۲	ماه ثبت نام	۰/۲۰۷
۳	گروه مشتری	۰/۲۰۵
۴	روش ثبت نام	۰/۰۵۵
۵	حجم ماهیانه	۰/۰۵۴
۶	هزینه‌ی ماهیانه	۰/۰۴۵
۷	سرعت	۰/۰۴۱
۸	بستر ارتباطی	۰/۰۳۲
۹	طول مدت انقضاء	۰/۰۳۰
۱۰	گروه سرویس	۰/۰۲۷
۱۱	نوع کاربری	۰/۰۱۵
۱۲	حقیقی/ حقوقی	۰/۰۱۰



نمودار ۱. میزان اهمیت ویژگی‌های موثر بر طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها

بحث و نتیجه‌گیری

در این تحقیق سعی شد مدلی برای مدیریت رویگردانی مشتریان یک شرکت مخابراتی ایرانی ارایه گردد که شامل پیش‌بینی رویگردانی و شناسایی مهم‌ترین ویژگی‌های موثر در طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها می‌باشد. برای دستیابی به این هدف داده‌های ۷۳۸۳۵ مشتری از پایگاه داده سازمان استخراج شد. سپس، در مرحله آماده‌سازی با استفاده از تکنیک "کم‌نمونه‌برداری تصادفی" داده‌ها متعادل‌سازی شدند و با استفاده از روش انتخاب ویژگی "حداقل افزونگی و حداکثر ارتباط" ویژگی‌های زاید در مجموعه داده حذف شدند و موثرترین ویژگی‌ها برای انجام پیش‌بینی استخراج گردیدند. سپس سه الگوریتم جنگل تصادفی، ماشین بردار پشتیبان و کا نزدیکترین همسایگان که از مطرح‌ترین الگوریتم‌های یادگیری ماشین در زمینه پیش‌بینی و طبقه‌بندی محسوب می‌شوند، جهت پیش‌بینی و شناسایی مشتریان رویگردان و غیررویگردان بر روی مجموعه داده تحقیق اعمال شدند و نتایج حاصل در قالب معیارهای استانداردِ ارزیابی مدل‌های پیش‌بینی، به

دست آمد. مقایسه معیارهای ارزیابی، برتری الگوریتم جنگل تصادفی را در مقایسه با دو الگوریتم دیگر به خوبی نشان می‌دهد. مدل نهایی که از ترکیب روش‌های متعادل‌سازی، انتخاب ویژگی و طبقه‌بندی به دست آمد تحت عنوان مدل RUS-mRMR-RF به عنوان یک مدل جدید پیش‌بینی رویگردانی در این تحقیق ارایه شد که عملکرد قابل قبولی در پیش‌بینی رویگردانی مشتریان شرکت و شناسایی مهم‌ترین عوامل موثر بر طبقه‌بندی رویگردان‌ها و غیررویگردان‌ها داشت.

عملکرد مدل پیشنهادی بدین صورت است که با استفاده از مجموعه داده‌های دردسترس از مشتریان رویگردان و غیررویگردان پیشین، به کشف الگوهای پنهان در این داده‌ها می‌پردازد و از یادگیریِ به دست آمده، برای شناسایی مشتریان رویگردان بالقوه در آینده استفاده می‌کند. همچنین میزان اهمیت ویژگی‌های موجود در مجموعه داده را برای طبقه‌بندی و جداسازی مشتریان رویگردان از غیررویگردان مشخص می‌کند. بر این اساس پیشنهاد می‌شود شرکت با اعمال مدل پیشنهادی بر روی داده‌های مشتریان جدید، رویگردان‌های بالقوه را شناسایی کند و با این شناسایی زودهنگام، پیش از آن‌که اولین نشانه‌های رویگردانی نمایان شود و تاثیر فعالیت‌ها برای حفظ مشتری کاهش یابد، در جهت افزایش رضایتمندی این مشتریان و متعاقبا نگهداری آن‌ها برای مدت زمان طولانی‌تر تلاش نماید.

همچنین با تعیین میزان اهمیت ویژگی‌ها توسط مدل پیشنهادی، مشخص شد سه ویژگی منطقه، ماه ثبت‌نام و گروه مشتری به طور قابل توجهی بیش از سایر ویژگی‌ها در طبقه‌بندی و جداسازی مشتریان رویگردان از غیررویگردان موثر هستند. از این رو بر اساس هر یک از این سه ویژگی موثر، پیشنهادات کاربردی زیر به شرکت ارایه می‌گردد:

منطقه: طی بررسی‌هایی مشخص شد نرخ رویگردانی در برخی مناطق مشخصا بیشتر از مناطق دیگر است. لذا پیشنهاد می‌شود در مناطق دارای نرخ رویگردانی بالاتر، بر اساس عوامل اجتماعی، اقتصادی، بافت جمعیتی و حضور رقبا در این مناطق، فعالیت‌های بازاریابی بازبینی و یا افزایش داده شود. همچنین با توجه به وجود مرکز عملیات و نگهداری مجزا در هر منطقه، که وظیفه‌ی رسیدگی به درخواست‌های پشتیبانی مشتریان و رفع مشکلات شبکه را بر عهده دارد، بررسی و در صورت نیاز اصلاح روند رسیدگی به

درخواست‌های مشتریان در مراکز عملیات و نگهداری مناطق دارای نرخ رویگردانی بالاتر، پیشنهاد می‌گردد.

ماه ثبت نام: به بخش‌های تجاری و فروش شرکت پیشنهاد می‌شود جشنواره‌های ویژه‌ی فروش را در ماه‌هایی برگزار کنند که نرخ رویگردانی ثبت‌نامی‌های آن کمتر بوده است. و از طرفی در ماه‌هایی که نرخ رویگردانی ثبت‌نامی‌های آن بیشتر بوده، مشتریان را به بستن قراردادهای بلندمدت‌تر ترغیب نمایند.

گروه مشتری: گروه مشتری بر اساس تعریفی که توسط شرکت ارایه شد به مزایایی که در هریک از کمپین‌های بازاریابی برای گروه‌های مختلف مشتریان در نظر گرفته شده است اشاره دارد. مشخص شد که گروه‌هایی از مشتریان با مزایای خاص مانند بهره‌مندی از مودم رایگان، زمان‌های استفاده رایگان، تخفیف‌های ویژه و غیره در مقایسه با دیگر گروه‌های مشتریان رویگردانی کمتری دارند. از این رو پیشنهاد می‌گردد شرکت کمپین‌های آینده را با توجه بیشتر به این موارد طراحی نماید.

یافته‌های این تحقیق به سازمان مورد بررسی کمک می‌کند تا به ایجاد استراتژی‌هایی برای حفظ و تقویت روابط با مشتریان خود بپردازد. با این قابلیت‌ها، مدل پیشنهادی اطلاعات مهمی را هم برای بخش‌های فنی و هم بخش‌های فروش و ارتباط با مشتریان فراهم می‌کند تا منابع، استراتژی‌ها و نیروی کار، در جهت ارتقاء کسب‌وکار و اتخاذ تصمیمات مبتنی بر دانش، هماهنگ شوند.

محدودیت‌های تحقیق

- عدم دسترسی به اطلاعات مصرف مشتریان و تعداد تماس‌های مشتریان با واحد پشتیبانی شرکت، باعث شد که این فاکتورها در تجزیه‌وتحلیل‌های رویگردانی و ساخت مدل به کار نروند و تاثیر آن‌ها قابل بررسی نباشد.
- مدل‌های داده‌کاوی عمر منقضی نسبتا کوتاهی دارند. بازار ارایه‌دهندگان خدمات اینترنت روزانه با فن‌آوری‌های جدیدی روبه‌رو می‌شود، در نتیجه، داده‌های تاریخی برای پیش‌بینی‌ها به تدریج از ارزش کمتری برخوردار می‌شوند. مدل ساخته شده در این پژوهش نیز شامل این محدودیت می‌باشد.

• تجزیه و تحلیل رویگردانی مشتریان به تنهایی برای بهبود روابط با آن‌ها کافی نیست. پیش‌بینی و بررسی سایر رفتارهای مشتریان مانند کاهش میزان خرید، ثبت شکایات و غیره، می‌تواند به عنوان مکمل پژوهش حاضر در نظر گرفته شود.

• واحد پشتیبانی شرکت مورد پژوهش روزانه پیام‌ها و درخواست‌های متنی زیادی را از مشتریان خود دریافت می‌کند که شامل اعلام خرابی سرویس، درخواست راهنمایی، ابراز نارضایتی و غیره است. استفاده از این پیام‌ها که می‌تواند به عنوان یک منبع باارزش داده جهت تجزیه و تحلیل رویگردانی به کار رود، در این تحقیق در نظر گرفته نشده است.

تعارض منافع

تعارض منافع وجود ندارد.

ORCID

Sahar Amiri **https://orcid.org/0000-0001-7795-5559**
Alireza Hassanzadeh **http://orcid.org/0000-0003-3015-3019**
Shaghayegh Sahraei **http://orcid.org/0000-0002-7265-8739**

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

خانلری، الف.، رییسی وانانی، الف.، و مقدسی، ز. (۱۳۹۵). پیش بینی رویگردانی مشتریان شرکت‌های ارایه‌دهنده خدمات اینترنت با ترکیبی از الگوریتم‌های داده کاوی؛ مطالعه موردی یک شرکت ایرانی. *دومین کنفرانس بین‌المللی در مدیریت، حسابداری و اقتصاد*، تهران. https://civilica.com/doc/610791/

سپهری، س.، و کوشا، ح. (۱۳۹۶). پیش‌بینی رویگردانی مشتریان با استفاده از کرنل‌های ترکیب شده در تکنیک ماشین بردار پشتیبان. *سومین کنفرانس بین‌المللی مهندسی صنایع و سیستم‌ها (ICISE 2017)*، مشهد. https://civilica.com/doc/669097

عاشوری، ع.، و البدوی، الف. (۱۳۹۴). مدل ترکیبی برای پیشبینی دلایل رویگردانی مشتریان شرکت‌های ارایه‌دهنده خدمات اینترنتی ISP. *دوازدهمین کنفرانس بین‌المللی مهندسی صنایع*، تهران. https://civilica.com/doc/515893

عسگری، م. تقوا، م. و تقوی فرد، م. (۱۳۹۸). پیش‌بینی رویگردانی جزئی مشتریان بانک‌ها با استفاده از مدل زنجیره وضعیت. *فصلنامه مطالعات مدیریت کسب و کار هوشمند، ۷(۲۸)، ۶۷-۱۱۰*. https://doi.org/10.22054/IMS.2019.10230

کاظمی، م.، و حجازی نیا، ر. (۱۳۹۲). ارایه مدلی به منظور پیش‌بینی رویگردانی مشتریان شرکت‌های مخابراتی. *دومین همایش ملی علوم مدیریت نوین*، گرگان. https://civilica.com/doc/231685

کاظمی، م.، و حجازی نیا، ر. (۱۳۹۵). بررسی متغیرهای موثر در رویگردانی مشتریان تلفن همراه. *فصلنامه مدیریت توسعه و تحول*، دوره ۱۳۹۵، ویژه نامه: ۱۱۵-۱۲۱. http://jdem.ir/article_534563.html?lang=en

References

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237: 242-254. https://doi.org/10.1016/j.neucom.2016.12.009

Bahari, T.F., & Elayidom, M.S. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. *Procedia Computer Science*, 46: 725-731. https://doi.org/10.1016/j.procs.2015.02.136

Bi, W., Cai, M., Liu, M., & Li G. (2016). A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn. *IEEE Transactions on*

Industrial Informatic, 12(3): 1270-1281. https://doi.org/10.1109/tii.2016.2547584

Biau, G., & Scornet E. (2016). A random forest guided tour. *TEST*, 25: 197–227. https://doi.org/10.1007/s11749-016-0481-7

Breiman, L. (1999). Radnom Forests – Random Features. *Technical Report 567, Department of Statistics, University of California: Berkeley, CA, USA*. https://www.stat.berkeley.edu/~breiman/random-forests.pdf

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). Classification And Regression Trees (1st ed.). *Routledge*. https://doi.org/10.1201/9781315139470

Caigny, A.D., Coussement, K., & De Bock, K.W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2): 760-772. https://doi.org/10.1016/j.ejor.2018.02.009

Caigny, A.D., Coussement, K., & De Bock, K.W. (2020). Leveraging fine-grained transaction data for customer life event predictions. *Decision Support Systems*, 130: 113232. https://doi.org/10.1016/j.dss.2019.113232

Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16: 321-357. https://doi.org/10.1613/jair.953

Couronné, R., Probst, P., & Boulesteix, A.L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*, 19: 270. https://doi.org/10.1186/s12859-018-2264-5

Coussement, K., Lessmann, S., & Verstraetenc, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, 95: 27-36. https://doi.org/10.1016/j.dss.2016.11.007

Dahiya, K., & Bhatia, S. (2015). Customer churn analysis in telecom industry. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*: 1-6. https://doi.org/10.1109/icrito.2015.7359318

Gui, C. (2017). Analysis of imbalanced data set problem: The case of churn prediction for telecommunication. *Artificial Intelligence Research*, 6(2): 93-99. https://doi.org/10.5430/air.v6n2p93

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10): 2902-2917. https://doi.org/10.1016/j.cor.2005.11.007

Harper, G., & Pickett, S.D. (2006). Methods for mining HTS data. *Drug Discovery Today*, 11(15–16): 694-699. https://doi.org/10.1016/j.drudis.2006.06.006

Hassouna, M., Tarhini, A., Elyas, T., & AbouTrab, M.S. (2016). Customer churn in mobile markets a comparison of techniques. *International Business Research*, 8(6): 224-237. https://doi.org/10.48550/arXiv.1607.07792

Idris, A., Rizwan, M., & Khan, A. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, 38(6): 1808-1819. https://doi.org/10.1016/j.compeleceng.2012.09.001

Jain, H., Khunteta, A., & Srivastava, S. (2021). Telecom churn prediction and used techniques, datasets and performance measures: a review. *Telecommun Systems*, 76: 613-630. https://doi.org/10.1007/s11235-020-00727-0

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24: 994-1012. https://doi.org/10.1016/j.asoc.2014.08.041

Khodabandehlou, S., & Rahman, M.Z. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, 19(1/2): 65-93. https://doi.org/10.1108/JSIT-10-2016-0061

Kotsiantis, S.B., Kanellopoulos, D., & Pintelas, P.E. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science*, 1(1): 111-117. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.8413&rep=rep1&type=pdf

Li, S.T., Shue, L.Y., & Lee, S.F. (2006). Enabling customer relationship management in ISP services through mining usage patterns. *Expert Systems with Applications*, 30(4): 621-632. https://doi.org/10.1016/j.eswa.2005.07.016

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3): 18-22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2): 137-166. https://doi.org/10.1017/S0269888910000032

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8): 1226-1238. https://doi.org/10.1109/TPAMI.2005.159

Speiser, J.L., Miller, M.E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134: 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

Torres, R., Sidorova, A., & Jones, M.C. (2018). Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective. *Information & Management*, 55(7): 822-839. https://doi.org/10.1016/j.im.2018.03.010

Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S.U., & Kim, S.W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7: 60134-60149. https://doi.org/10.1109/ACCESS.2019.2914999

Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G., & Chatzisavvas, K.Ch. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55: 1-9. https://doi.org/10.1016/j.simpat.2015.03.003

Vo, N.N.Y., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, 212: 106586. https://doi.org/10.1016/j.knosys.2020.106586

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*: 29-39. http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

Wu, S., Yau, W.C., Ong, T.S., & Chong, S.C. (2021). Integrated Churn Prediction and Customer Segmentation Framework for Telco Business. *IEEE Access*, 9: 62118-62136. https://doi.org/10.1109/ACCESS.2021.3073776

Xiao, J., Jiang, X., He, C., & Teng, G. (2016). Churn prediction in customer relationship management via GMDH-based multiple classifiers ensemble. *IEEE Intelligent Systems*, 31(2): 37-44. https://doi.org/10.1109/MIS.2016.16

Yap, B.W., Abd Rani, K., Abd Rahman, H.A., Fong, S., Khairudin, Z., & Abdullah, N.N. (2014). An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*: 13-22. https://doi.org/10.1007/978-981-4585-18-7_2

Zhu, B., Baesens, B., & vanden Broucke, S.K.L.M. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, 408: 84-99. https://doi.org/10.1016/j.ins.2017.04.015

References [In Persian]

Asgari, M., Taghva, M., & Taghavifard, M.T. (2019). Prediction of Bank Customers' Partial Churn Using State Chain Model. *BI Management Studies*, 7(28): 67 – 110. [In Persian] https://doi.org/10.22054/IMS.2019.10230

Ashoori, A., & Albadavi, A. (2016). Combined model for predicting customer churn reasons of Internet service providers (ISP). *12th International Conference in Industrial Engineering, Tehran*. [In Persian] https://civilica.com/doc/515893

Kazami, M., & Hejazinia, R. (2013). Provide a Model for Customer Churn Prediction of telecommunication companies. The Second Conference on Modern Management Sciences, Gorgan. [In Persian] https://civilica.com/doc/231685

Kazami, M., & Hejazinia, R. (2017). Study affective variables in mobile customers churn. Journal of Development Evolution Management, 1395(special issue): 115-121. [In Persian] http://jdem.ir/article_534563.html?lang=en

Khanlari, A., Vanani, I.R., & Moghadasi, Z. (2017), Customer Churn Prediction of Internet service providers with a combination of data mining algorithms; Case study of an Iranian company. *2nd International Conference of Management, Accounting and Economics, Tehran*. [In Persian] https://civilica.com/doc/610791/

Sepehri, S., & Koosha, H. (2017). Customer Churn Prediction using kernels combined in Support Vector Machine technique. *3rd International Conference on /industrial and Systems Engineering (ICISE 2017), Mashhad*. [In Persian] https://civilica.com/doc/669097

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

استناد به این مقاله: امیری، سحر.، حسن‌زاده، علیرضا.، صحرایی، شقایق. (۱۴۰۱). مدلی برای مدیریت رویگردانی مشتریان یک شرکت ارایه‌دهنده سرویس اینترنت، *مطالعات مدیریت کسب وکار هوشمند*، ۱۰(۳۹)، ۶۷-۹۵.

DOI: 10.22054/IMS.2021.63193.2038

Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی