Financial Research Journal

Print ISSN: 1024-8153
Online ISSN: 2423-5377

Predicting Bank Customer Churn Using Machine Learning

Mohammad Hossein Mahmoudzadeh
MSc., Department of Computer Engineering, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. E-mail: mhmahmoudzadeh@aut.ac.ir

Mohammad Hassan Shirali Shahreza *
*Corresponding Author, Assistant Prof., Department of Computer Education, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. E-mail: hshirali@aut.ac.ir

Abstract

Objective

In competitive markets, companies focus on establishing long-term relationships with their customers and strengthening their loyalty. Due to the high costs associated with acquiring new customers, businesses tend to focus on retaining existing ones. Predicting which customers are likely to churn in the future plays a crucial role in shaping effective customer retention strategies. To predict customer churn and identify its drivers, companies use customer information and historical data recorded from them. This paper investigates customer churn prediction in the banking industry using real customer data from one of the largest banks in Iran.

Methods

To analyze and predict customer behavior, two equal and consecutive periods were considered. Customer behavior in the first period was used to predict a target variable in the second period. A significant drop in the average effective balance during the second period, compared to the first, was defined as the indicator of customer churn. By processing a large volume of banking transactions in the first period and aggregating them at different levels, various behavioral features for customers. We selected. One fixed validation set and three

Citation: Mahmoudzadeh, Mohammad Hossein & Shirali Shahreza, Mohammad Hassan (2025). Predicting Bank Customer Churn Using Machine Learning. *Financial Research Journal*, 27(2), 218-245. https://doi.org/10.22059/FRJ.2024.357770.1007453 (*in Persian*)

Financial Research Journal, 2025, Vol. 27, No.2, pp. 218-245
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/FRJ.2024.357770.1007453
Article Type: Research Paper

Received: April 16, 2023
Received in revised form: October 06, 2024
Accepted: October 16, 2024
Published online: May 21, 2025

training sets of different sizes were selected. To address the issue of dataset imbalance, class weights were determined based on the ratio of class sizes, ensuring that the minority class received greater weight during the training process. To predict customer churn, widely used machine learning algorithms—including Naive Bayes, k-Nearest Neighbors, Support Vector Machine, Logistic Regression, and Decision Tree—were applied, along with ensemble learning methods such as Random Forest, Adaptive Boosting, and Gradient Boosting. Subsequently, deep learning methods were applied, and a model incorporating modern modules such as residual connections and layer normalization—similar to state-of-the-art architectures—was proposed. Exhaustive experiments were conducted to evaluate the performance of the aforementioned methods.

Results

The results showed that ensemble learning algorithms and the proposed deep learning models outperformed the baseline models. Additionally, increasing the size of the training set contributed to improved model performance. Among the traditional machine learning classification algorithms, the decision tree trained on two training sets obtained the highest AUC ROC on the validation set with 0.8531 and 0.8597. The gradient boosting model obtained the overall highest AUC ROC on the validation set with 0.8984 and 0.9010. Deep learning-based single models achieved AUC-ROC values of 0.8825, 0.8909, and 0.8958, outperforming all traditional methods and two ensemble learning approaches while performing competitively with the gradient boosting algorithm.

Conclusion

Extracting behavioral features from customers' banking transactions and applying ensemble methods, along with the proposed deep learning-based models, proves effective in predicting banking customer churn, particularly in cases of a significant decrease in the average effective balance.

Keywords: Banking, Customer churn prediction, Deep learning, Machine learning.

تحقیقات مالی
شاپا چـــاپـــی: ۱۰۲۴-۸۱۵۳
شاپا الکترونیکی: ۲۴۲۳-۵۳۷۷

پیش‌بینی ریزش مشتریان بانک با استفاده از یادگیری ماشین

محمدحسین محمودزاده
کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهـران، ایـران.
رایانامه: mhmahmoudzadeh@aut.ac.ir

محمدحسن شیرعلی شهرضا *
* نویسنده مسئول، استادیار، گروه آموزشی علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهران، ایران. رایانامه: hshirali@aut.ac.ir

چکیده

هدف: در بازارهای رقابتی، شرکت‌ها روی برقراری روابط بلندمدت با مشتریان و تقویت وفاداری تمرکز دارند. به‌علت هزینه‌هـای سـنگین جذب مشتری جدید، کسب‌وکارها روی نگهداری مشتریان موجود تمرکز می‌کنند. پیش‌بینی مشتریانی کـه احتمـال روی‌گردانـی آن‌هـا در آینده وجود دارد، بخش مهمی از راهبرد حفظ مشتری است. در این مقاله پیش‌بینی ریزش مشتری در صنعت بانکداری، روی دادگان واقعی مشتریان یکی از بانک‌های بزرگ ایران انجام شده است.

روش: در صنعت بانکداری، کاهش شدید میانگین مانده مؤثر یک مشتری در یک بازهٔ زمانی نسبت به بازهٔ زمانی قبلی، به‌عنـوان ریـزش مشتری در نظر گرفته می‌شود. در این مقاله، ابتدا با پردازش حجم زیادی از تراکنش‌های بانکی در یک بازهٔ زمانی مشـخص، ویژگـی‌هـای رفتاری متفاوت در سطوح مختلف برای مشتریان به‌دست آمد؛ سپس برای پیش‌بینی ریزش، از الگوریتم‌های پُراستفاده در یادگیری ماشـین و روش‌های یادگیری جمعی استفاده شد. در ادامه، با استفاده از روش‌های یادگیری عمیق و واحدهای نوین آن، معماری مدل مـدنظر ارائـه شد. در نهایت، با انجام آزمایش‌های جامع، عملکرد روش‌های نام برده بررسی شد.

یافته‌ها: این پژوهش در یکی از بانک‌های بزرگ ایران اجرا شد و آزمایش‌ها روی دادگان واقعی مشتریان بانک صورت پـذیرفت. در ایـن آزمایش‌ها، از اطلاعات جمعیت‌شناختی و رفتار گذشته مشتریان بهره گرفته شد؛ امـا از اطلاعـات شخصـی افـراد اسـتفاده نشـد تـا حـریم خصوصی مشتریان حفظ شود. در آزمایش‌های صورت‌گرفته، پیش‌بینی ریزش مشتری روی بازهٔ زمانی یک ماهه انجام گرفت. بدین ترتیب دو بازهٔ یک‌ماهه متوالی مدنظر قرار گرفت و ویژگی‌های رفتاری مشتریان، از بازهٔ زمـانی اول اسـتخراج شـد. متغیـر هـدف نیـز از مقایسـه میانگین مانده مؤثر در بازه‌های زمانی اول و دوم به‌دست آمد. در صورتی که میانگین ماندهٔ مؤثر یک مشتری، در بازهٔ دوم نسـبت بـه بـازهٔ اول با بیش از ۷۰درصد کاهش همراه بود، به‌عنوان ریزش در نظر گرفته شد. در نتـایج بـه‌دسـت‌آمـده الگـوریتم‌هـای یـادگیری جمعـی و

استناد: محمودزاده، محمدحسین و شیرعلی شهرضا، محمدحسن (۱۴۰۴). پیش‌بینی ریزش مشتریان بانک با استفاده از یادگیری ماشـین. *تحقیقات مالی*، ۲۷(۲)، ۲۱۸- ۲۴۵.

تحقیقات مالی، ۱۴۰۴، دوره ۲۷، شماره ۲، صص. ۲۱۸- ۲۴۵
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی

تاریخ دریافت: ۱۴۰۲/۰۱/۲۷
تاریخ ویرایش: ۱۴۰۳/۰۷/۱۵
تاریخ پذیرش: ۱۴۰۳/۰۷/۲۵
تاریخ انتشار: ۱۴۰۴/۰۲/۳۱
doi: https://doi.org/10.22059/FRJ.2024.357770.1007453

همچنین مدل‌های عمیق ارائه‌شده، عملکرد بهتری را نسبت به مدل‌های مبنا نشان دادند. افزایش اندازهٔ مجموعهٔ آموزش در عملکرد بهتر مدل‌ها مؤثر بود. مدل تقویت گرادیان با ۰/۸۹۸۴ بیشترین مساحت زیر منحنی مشخصهٔ عامل گیرنده نسبت به مجموعهٔ اعتبارسنجی را به‌دست آورد.

نتیجه‌گیری: استخراج ویژگی‌های رفتاری از تراکنش‌های بانکی مشتریان و استفاده از روش‌های یادگیری جمعی و همچنین مدل‌های ارائه‌شده مبتنی‌بر یادگیری عمیق، در پیش‌بینی ریزش مشتری مؤثرند. پس از تحلیل رفتار و شناسایی مشتریان در شرف روی‌گردانی، پیشنهادهایی برای جلوگیری از ریزش و حفظ مشتری ارائه شد. برای نمونه، تفکیک مشتریان بر اساس سن، شغل، تحصیلات و غیره به‌منظور ارائه خدمات و تولید محصولات بانکی بر این مبنا، ایجاد تنوع در خدمات موجود، ارائهٔ خدمات مورد نیاز مشتریان از طریق بسترهای مجازی و در صورت نیاز در محل فعالیت و زندگی مشتریان، تسهیل در ارائهٔ خدمات مطلوب به مشتریان، افزایش اعتماد مشتریان از طریق ارائهٔ کاربردیِ امن و همچنین، حفاظت از اطلاعات مشتریان، به حفظ مشتری کمک می‌کند.

کلیدواژه‌ها: پیش‌بینی ریزش مشتری، بانکداری، یادگیری عمیق، یادگیری ماشین.

مقدمه

برای هر کسب‌وکاری مشتریان اساس درآمد و موفقیت‌اند و یکی از دارایی‌های مهم در نظر گرفته می‌شوند. در بازارهای رقابتی که خدمات عرضه‌شده توسط ارائه‌دهندگان مختلف، مشابه یکدیگرند، مشتریان قدرت انتخاب دارند و به‌راحتی می‌توانند خدمت‌رسان خود را تغییر دهند. در چنین شرایطی، سازمان‌ها روی برقراری روابط بلندمدت با مشتریان خود تمرکز دارند و در طول زمان رفتار مشتری را رصد می‌کنند (امین و همکاران[1]، ۲۰۱۷). مدیریت ارتباط با مشتری، راهبردی جامع برای ایجاد و تقویت روابط وفادارانه و طولانی‌مدت با مشتریان است. مدیریت ارتباط با مشتری در کسب‌وکارهای مختلف مانند مخابرات، بانک، بیمه، بازار خرده‌فروشی و غیره به رسمیت شناخته شده و به‌طور گسترده استفاده می‌شود (وافیادیس، دیامانتاراس، ساریگیانیدیس و چاتزیساواس[2]، ۲۰۱۵). امروزه رویکرد مشتری‌محور برای پیش‌بینی رفتار مشتریان بر اساس داده‌های تاریخی ذخیره‌شده در بانک‌های اطلاعاتی و سامانه‌های مدیریت ارتباط با مشتری بسیار رایج است (امین و همکاران، ۲۰۱۷).

نگهداری مشتری و پیش‌بینی ریزش

یکی از اهداف اصلی در مدیریت ارتباط با مشتری، حفظ مشتریان موجود است. طبق بررسی‌های صورت‌گرفته، جذب مشتریان جدید بسیار پُرهزینه‌تر از حفظ مشتریان موجود است. در برخی موارد هزینهٔ جذب مشتری تا ۲۰ برابر بیشتر از حفظ مشتری برآورد شده است (وافیادیس و همکاران، ۲۰۱۵). این واقعیت اهمیت نگهداری مشتریان موجود را آشکار ساخته و تمرکز کسب‌وکارها نیز به‌سمت آن تغییر کرده است. همچنین تقویت وفاداری و حفظ مشتری به بهبود فروش و کاهش هزینه‌های بازاریابی منجر می‌شود (امین و همکاران، ۲۰۱۷).

بخشی از راهبرد حفظ مشتری که از اهمیت زیادی برخوردار است، پیش‌بینی ریزش مشتری است. فرایند جست‌وجو و شناسایی مشتریانی که تمایل زیادی به ترک شرکت نشان می‌دهند، به‌عنوان پیش‌بینی ریزش مشتری شناخته می‌شود. برای مواجهه با مسئلهٔ ریزش، می‌توان از داده‌های نگهداری‌شده از مشتریان استفاده و آن را به دانش معنادار تبدیل کرد. در پیش‌بینی ریزش مشتری، یک مدل امتیازدهنده بر اساس رفتار گذشته مشتری، برآورد احتمال روی‌گردانی داوطلبانه در آینده را فراهم می‌آورد (دی کاینی، کوسمان و دی باک[3]، ۲۰۱۸). هدف مدل‌های پیش‌بینی ریزش، تشخیص نشانه‌های اولیه روی‌گردانی و شناسایی مشتریانی است که احتمال ترک آن‌ها افزایش یافته است. بدین‌ترتیب می‌توان قبل از روی‌گردانی، مشتریان هدف را شناسایی کرد و به‌منظور ارائه پیشنهادها و برگزاری پویش‌های بازاریابی، از آن‌ها بهره برد. ریزش مشتری می‌تواند نتیجهٔ سطح پایین رضایت، راهبردهای رقابتی تهاجمی، محصولات جدید، مقررات یا موارد دیگر باشد (وافیادیس و همکاران، ۲۰۱۵). علاوه‌بر پیش‌بینی مشتریان محتمل، محققان در پی این هستند تا به‌دلایل ریزش مشتری و محرک‌های مهم آن، مانند رضایت مشتری پی ببرند (دی کاینی و همکاران، ۲۰۱۸). از این رو، توانایی عملکردی مدل‌های پیش‌بینی ریزش مشتری حائز اهمیت است. استفاده از این مدل‌های پیش‌بینی‌کننده، باید به

1. Amin et al.
2. Vafeiadis, Diamantaras, Sarigiannidis & Chatzisavvas
3. De Caigny, Coussement & De Bock

بینش‌های قابل اجرا و درک بهتر محرک‌های ریزش مشتری، به‌منظور اتخاذ تصمیم‌های آگاهانهٔ مدیریتی منجر شود (دی کاینی و همکاران، ۲۰۱۸). این حقایق در نهایت باعث می‌شود که فعالیت پیش‌بینی ریزش مشتری، به‌عنوان بخشی ضروری از تصمیم‌گیری راهبردی و فرایند برنامه‌ریزی شرکت‌ها و سازمان‌ها باشد (امین و همکاران، ۲۰۱۷). بدین ترتیب اهمیت اطلاعات تاریخی مشتریان که می‌تواند برای ساختن مدل‌های پیش‌بینی‌کننده استفاده شود، روشن است و به‌عنوان دارای مهم برای مقابله با ریزش مشتری شمرده می‌شود. از این رو شرکت‌های زیادی برای سرمایه‌گذاری در مدیریت ارتباط با مشتری و حفظ اطلاعات مشتریان اقدام کرده‌اند.

شرح مسئله و چالش‌های آن

پیش‌بینی ریزش مشتری، مسئله‌ای برای دسته‌بندی مشتریان به گروه‌های ریزش و عدم ریزش است. بدین منظور با در اختیار داشتن اطلاعات مشتریان و داده‌های تاریخی ثبت‌شده برای آن‌ها، باید به مهندسی ویژگی پرداخت. ویژگی‌های استخراج‌شده برای مشتری، به‌عنوان ورودی به مدل پیش‌بینی‌کننده داده می‌شود. مدل تخمین خود را از دسته‌ای که مشتری به آن مربوط است، به‌عنوان خروجی ارائه می‌دهد. خروجی مدل می‌تواند به‌صورت برآوردی از احتمال تعلق مشتری به دستهٔ ریزش باشد.

از پیش‌بینی این مدل‌ها، برای انتخاب مشتریان هدف به‌منظور انجام بررسی‌های بیشتر و تلاش برای جلوگیری از ترک مشتری استفاده می‌شود. علاوه‌بر شناسایی مشتریان در شرف روی‌گردانی، یافتن علل و محرک‌های مؤثر بر ریزش مشتری به‌منظور اتخاذ تصمیم‌های مدیریتی نیز ارزش‌آفرین است. از این رو توانایی عملکردی این مدل‌ها در پیش‌بینی احتمال ریزش حائز اهمیت است. بدین منظور، استفاده از الگوریتم‌های دسته‌بندی کارا و وجود ویژگی‌های مرتبط و مفید برای مشتریان لازم است. این ویژگی‌ها باید از پردازش و تجمیع حجم زیادی از دادگان ذخیره‌شده از فعالیت مشتریان در گذر زمان مهندسی و استخراج شوند.

از دیگر چالش‌هایی که در مسئلهٔ پیش‌بینی ریزش با آن مواجه می‌شویم، مشکل عدم تعادل مجموعه داده است. مشتریان مربوط به دستهٔ ریزش که مورد توجه قرار دارند، درصد کمتری از مجموعه داده را تشکیل می‌دهند. بدین ترتیب برای مقابله با عدم تعادل مجموعه داده نیز، باید راه‌کاری را اتخاذ کرد.

پیش‌بینی ریزش مشتری در زمینه‌های مختلف مانند مخابرات، خدمات مالی، بیمه و غیره مطالعه و بررسی شده و مقاله‌های زیادی نیز با محوریت آن انتشار یافته است. در بیشتر مقاله‌ها با استفاده از الگوریتم‌های دسته‌بندی یادگیری ماشین سنتی، فرایند پیش‌بینی ریزش مشتری صورت پذیرفته است (احمد، جعفر و الجمعه[1]، ۲۰۱۹؛ کاروانا، یزید، سیالیم و مورسانتو[2]، ۲۰۱۹؛ لالوانی، میشرا، چادا و ستی[3]، ۲۰۲۲). همچنین روش‌هایی از یادگیری عمیق نیز برای پیش‌بینی ریزش مشتری، به‌طور محدود استفاده شده است (خان و همکاران[4]، ۲۰۱۹؛ اسپانودس و نگوین[5]، ۲۰۱۷).

1. Ahmad, Jafar & Aljoumaa
2. Karvana, Yazid, Syalim & Mursanto
3. Lalwani, Mishra, Chadha & Sethi
4. Khan et al.
5. Spanoudes & Nguyen

در این مقاله، مسئلهٔ پیش‌بینی ریزش مشتری در صنعت بانکداری مطالعه و بررسی می‌شود. این پژوهش در یکی از بزرگ‌ترین بانک‌های ایران انجام شده و از داده‌های واقعی مشتریان بانک بهره گرفته شده است. برای حفظ حریم خصوصی مشتریان، از ذکر نام بانک در این مقاله خودداری شده است. با فرض اینکه حجم زیاد داده، از تعامل مشتریان بانک در طول زمان در اختیار است، نحوهٔ استفاده از دادگان به‌طور مؤثر و استخراج ویژگی‌های مفید برای پیش‌بینی ریزش مشتری، مورد توجه قرار گرفته است. با استفاده از معیار مهم میانگین ماندهٔ حساب مشتری و مقایسهٔ آن در بازه‌های زمانی مختلف، متغیر هدف را مشخص می‌کنیم. با بررسی تراکنش‌های انجام‌شده مشتری در یک بازه زمانی و فراداده‌های مربوط به آن، ویژگی‌هایی از نحوهٔ رفتار مالی مشتری به‌دست می‌آوریم.

همان‌طور که اشاره شد، در اغلب کارهای پیشین در زمینهٔ پیش‌بینی ریزش مشتری، از روش‌های یادگیری ماشین سنتی استفاده شده است. با توجه به اهمیت عملکرد مدل پیش‌بینی کنندهٔ مورد استفاده و ظرفیت یادگیری آن، این سؤال مطرح می‌شود که آیا استفاده از یادگیری عمیق نوین، می‌تواند به بهبود عملکرد کمک کند؟ در این راستا آزمایش‌های جامعی برای بررسی استفاده از روش‌های مختلف یادگیری ماشین و تأثیر اندازه مجموعه دادگان آموزش انجام می‌دهیم. بدین منظور از الگوریتم‌های پُرکاربرد و مرسوم دسته‌بندی در یادگیری ماشین بهره می‌بریم. از روش‌های یادگیری عمیق استفاده می‌کنیم و با به‌کارگیری برخی از واحدهای نوین آن، شبکه‌هایی را تشکیل می‌دهیم که معماری آن‌ها مشابه با آخرین دستاوردهای این حوزه است.

ساختار این مقاله در ادامه بدین شرح است: در ادامه، به پیشینهٔ تحقیق در حوزهٔ پیش‌بینی ریزش مشتری و روش‌های پُراستفاده پرداخته و چند نمونه از کارهای صورت‌گرفته در سال‌های اخیر مرور می‌شود. سپس مسئلهٔ پژوهش و روش تحقیق تشریح می‌شود. در بخش بعد، نتایج حاصل از آزمایش‌های انجام‌شده ارائه می‌شود. در پایان مقاله نیز نتیجه‌گیری و جمع‌بندی نهایی ارائه خواهد شد.

پیشینهٔ تجربی پژوهش

تحلیل رفتار مشتریان بانک و دسته‌بندی آن‌ها از مباحثی است که در سال‌های اخیر، در کانون توجه پژوهشگران حوزه تحقیقات مالی بوده است. به‌طور نمونه، در مقالهٔ رحیمی، روستا و آسایش (۱۴۰۳) به ارائهٔ الگوی ارتقای توان رقابت‌پذیری خدمات ارزی مشتریان در صنعت بانکداری پرداخته شده است. برای تعیین ابعاد، مؤلفه‌ها و شاخص‌ها، از آزمون دلفی و برای تعیین وضع موجود و همچنین، تعیین عوامل اثرگذار، از آزمون تحلیل معادلات ساختاری و تحلیل عاملی اکتشافی استفاده شده است. بر اساس نتایج آزمون معادلات ساختاری و آماره‌های به‌دست‌آمده، دریافته‌اند که ضرایب تمامی عوامل در نظرگرفته‌شده، بر مدل نهایی تأثیر معناداری می‌گذارند و اثرگذاری تمامی عوامل، بر الگوی ارتقای توان رقابت پذیری خدمات ارزی مشتریان در صنعت بانکداری تأیید شده است. در این مقاله اشاره شده است که بانک ها بایستی با ارزیابی وضعیت در حوزهٔ بانکی، به ایجاد مزیت رقابتی برای جذب، حفظ و نگهداری مشتریان اقدام کنند. در مقالهٔ احمدی کوشا، احمدی، رنجبر و کردلوئی (۱۴۰۳) شناسایی شاخص‌های اعتبارسنجی و رتبه‌بندی مشتریان

در تسهیلات خُرد در بانک خاورمیانه صورت گرفته است. در این پژوهش روش های آماری در دو بخش آمـار توصـیفی و استنباطی انجام گرفته است. در بخش آمـار توصـیفی، برخـی عوامـل شخصـیتی همچـون سـن، جنسـیت، تحصـیلات، کسب‌وکار و... تجزیه‌وتحلیل و از طریق جدول و نمودارها بررسی شده است. الگوریتم‌هایی بیز ساده اجرا و برای رده‌بندی معیارها و ایجاد الگو استفاده شده است. به‌منظور ارزیابی نهایی مدل اعتبارسنجی و رتبه‌بنـدی مشـتریان تسـهیلات خُـرد بدون پشتوانه، از آزمون تی استفاده شده است. یافته‌های این مقاله نشان می‌دهد که بانک باید بر اساس شاخص‌هایی در خصوص پرداخت یاعدم پرداخت وام تصمیم گیری کند و دربارهٔ اشخاص متقاضی تسهیلات، شناخت کامل داشـته باشـد. بدین منظور باید داده‌های لازم را از آن‌ها جمع‌آوری کند و از حسن اعتبار و شهرت آن‌ها اطمینان یابد.

در مقالهٔ باجلان، فلاح‌پور و رئیسی (۱۴۰۳) مدلی برای اندازه‌گیری ریسک و بهینه‌سازی پرتفوی اعتباری بانـک‌هـا ارائه شده است. ابتدا به بررسی ریسک اعتباری پرتفوی تسهیلات بانکی با استفاده از رویکرد اکچوئری پرداخته و سـپس با استفاده از شبکهٔ عصبی پرسپترون و با توجه به محدودیت های بانک در ارائه تسهیلات، ترکیب بهینه پرتفوی اعتبـاری تعیین شده است. نتایج به‌دست‌آمده در این مقاله نشان می‌دهد که بازده پرتفوی بهینه از پرتفوی فعلی بانک بالاتر اسـت. براساس یافته ها و تأیید فرضیه های پژوهشی نتیجه گرفته شده است که استفاده از مدل اکچوئری بـرای تعیـین ریسـک اعتباری و سپس بهینه‌سازی با استفاده از شبکهٔ عصبی مصنوعی، به بهبود فرایند بهینه‌سازی پرتفوی اعتبـاری بانـک‌هـا منجر می‌شود.

در مقالهٔ احمدی سرتختی، هژبر کیـانی، حسـینی و معمـارنژاد (۱۴۰۲) انتخـاب بهتـرین روش بـرای اعتبارسـنجی مشتریان ضمانت‌نامه های اعتباری و تفکیک اشخاص خوش‌حساب از بدحساب، برای کاهش نکول اعتبارات اعطـا شـده صندوق ضمانت صادرات ایران صورت گرفته است. به‌کمک مدل شبکهٔ عصبی مصنوعی، مـدلی بـرای ارزیـابی ریسـک متقاضیان تسهیلات و ضمانت نامه ها از این صندوق طراحی شده که بیشترین قدرت پیش‌بینی احتمال نکـول تسـهیلات اعطایی را داشته باشد. بر اساس نتایج حاصل از این پژوهش، نتیجه‌گیری شده است که می‌توان ضمن تفکیک مشـتریان خوش‌حساب از بدحساب و بر اساس رتبهٔ اعتباری مشتریان، میزان وثـایق اخذشـده از مشـتریان را متناسـب بـا وضـعیت اعتباری گروه‌های اعتباری تنظیم کرد.

با بررسی این مقاله‌ها می‌توان به اهمیت دسته‌بندی مشتریان در نظام بانکی پی برد. همچنین در تعدادی از کارهـا، به اعتبارسنجی مشتریان به‌منظور ارائه تسهیلات پرداخته شده است. اما متفاوت با پژوهش‌های ذکرشده، ایـن مقالـه بـه پیش‌بینی ریزش مشتریان بانک با تحلیل رفتار مشتری بر اساس تراکنش‌های بانکی انجام‌شده در بـازه زمـانی مشـخص می‌پردازد.

در گذشته مسئله پیش‌بینی ریزش مشتری در حوزه‌های کسب‌وکاری مختلف مانند مخـابرات، بانکـداری و خـدمات مالی، بیمه و غیره مورد مطالعه قرار گرفته است. در بسیاری از مطالعات ابتدا روی دادگان ذخیـره شـده از سـابقه فعالیـت مشتریان مهندسی ویژگی صورت گرفته است. با تجمیع دادگان و استخراج ویژگی‌های رفتاری و همچنین اضـافه کـردن اطلاعات جمعیت‌شناختی، مجموعه داده‌های ساختاریافته به‌دست آمده است. در این موارد مطالعات روی داده‌های واقعـی

شرکت‌ها یا مؤسسه‌هایی که پژوهش در آن‌ها صورت گرفته، انجام شده است (احمد و همکاران، ۲۰۱۹؛ کاروانا و همکاران، ۲۰۱۹). در تعدادی از کارها نیز استفاده از معدود مجموعه داده های عمومی مرتبط با پیش‌بینی ریزش مشتری در دستور کار قرار گرفته است (هالیباس و همکاران[1]، ۲۰۱۹؛ پوستوخینا، پوستوخین، نگوین، الهوسنی و شانکار[2]، ۲۰۲۱). در اغلب پژوهش‌ها عدم تعادل مجموعه داده، از چالش‌های موجود بوده که برای مواجهه با آن، از روش‌های کم‌نمونه‌برداری و بیش‌نمونه‌برداری استفاده شده است (رحمان و کومار[3]، ۲۰۲۰). برای پیش‌بینی ریزش مشتری، از مدل‌هایی مانند درخت تصمیم[4]، بیز ساده[5]، k نزدیک‌ترین همسایه[6]، رگرسیون لجستیک[7]، ماشین بردار پشتیبان[8] استفادهٔ بیشتری شده است. همچنین مدل‌های یادگیری جمعی مانند جنگل تصادفی[9]، تقویت تطبیقی[10]، تقویت گرادیان[11] و غیره نیز استفاده شده‌اند. بدین ترتیب، در اغلب مقاله‌های منتشرشده، استفاده از الگوریتم‌های دسته‌بندی یادگیری ماشین سنتی روی دادگان ساختاریافته بررسی شده و الگوریتم‌های یادگیری جمعی عملکرد خوبی داشته است.

استفاده از روش‌های یادگیری ماشین

در ادامه به چند نمونه از مطالعات انجام‌شده در سال‌های اخیر می‌پردازیم. در مقالهٔ احمد و همکاران (۲۰۱۹) پیش‌بینی ریزش مشتری در صنعت مخابرات بررسی شده است. بدین منظور از روش های یادگیری ماشین استفاده شده و برای سنجش عملکرد مدل نیز معیار مساحت زیر منحنی انتخاب شده است. مجموعه دادگان شامل، اطلاعات مشتریان در طول ۹ ماه فعالیت بوده است. علاوه‌بر ویژگی های آماری، نوع دیگری از ویژگی ها بر اساس فعالیت اجتماعی مشتریان از طریق پیامک و تماس محاسبه شده است. برای تأیید اعتبار و بهینه سازی ابرپارامترها، اعتبارسنجی متقابل ده برابری اجرا شده است. برای رفع نامتعادل بودن داده‌ها از زیرنمونه‌برداری یا اجرای الگوریتم‌های مبتنی‌بر درخت که نامتعادل بودن دادگان تأثیری روی آن ها ندارد، استفاده شده است. چهار الگوریتم مبتنی بر درخت شامل درخت تصمیم، جنگل تصادفی، ماشین تقویت گرادیان و تقویت گرادیان شدید آزمایش شده‌اند. استفاده از ویژگی‌های شبکهٔ اجتماعی، بیشترین تأثیر را روی بهبود عملکرد مدل داشته و معیار مساحت زیر منحنی را از ۸۴ به ۹۳/۳ افزایش داده است. بهترین نتایج توسط الگوریتم تقویت گرادیان شدید به‌دست آمد و پس از آن نیز به‌ترتیب الگوریتم‌های تقویت گرادیان، جنگل تصادفی و درخت تصمیم قرار گرفتند. در مقالهٔ لالوانی و همکاران (۲۰۲۲) نیز پیش‌بینی ریزش مشتری در صنعت مخابرات انجام شده است. سپس با استفاده از الگوریتم جست‌وجوی گرانشی، انتخاب ویژگی اجرا شده است. مدل هایی مانند رگرسیون

1. Halibas et al.
2. Pustokhina, Pustokhin, Nguyen, Elhoseny & Shankar
3. Rahman & Kumar
4. Decision Tree
5. Naive Bayes
6. k-Nearest Neighbors
7. Logistic Regression
8. Support Vector Machine
9. Random Forest
10. Adaptive Boosting (AdaBoost)
11. Gradient Boosting

لجستیک، بیز ساده، ماشین بردار پشتیبان، درخت تصمیم و غیره، روی دادگان آموزش اعمال شده است. همچنین تأثیر شیوه‌های تقویتی و جمعی مانند تقویت تطبیقی، تقویت گرادیان شدید، جنگل تصادفی و غیره نیز بر دقت مدل بررسی شده است. برای تنظیم ابرپارامترها و جلوگیری از بیش‌برازش مدل‌ها از اعتبارسنجی k برابری روی مجموعه آموزش استفاده شده است. عملکرد مدل های مختلف روی مجموعه آزمایش با استفاده از ماتریس درهم‌ریختگی و مساحت زیر منحنی ارزیابی شده‌اند. نتایج به‌دست‌آمده نشان می‌دهد که دسته‌بندهای تقویت تطبیقی و تقویت گرادیان شدید، به‌ترتیب با ۸۱/۷۱ درصد و ۸۰/۸ درصد بیشترین دقت ها را به‌دست آوردند. همچنین بیشترین مساحت زیر منحنی نیز ۸۴ درصد توسط دسته‌بندهای تقویت تطبیقی و تقویت گرادیان شدید به‌دست آمد.

در مقالهٔ پوستوخینا و همکاران (۲۰۲۱) مدلی برای پیش‌بینی ریزش مشتری در بخش مخابرات ایجاد شده است. بدین منظور شیوه‌ای برای بیش‌نمونه‌برداری ساختگی از دستهٔ اقلیت همراه با مدل یادگیری ماشین شدید با وزن‌های بهینه ارائه شده است. برای تعیین نرخ بهینهٔ نمونه‌برداری و تنظیم پارامترهای مدل از الگوریتم بهینه‌سازی باران با چند هدف استفاده شده است. برای مدیریت عدم تعادل مجموعه داده، بیش‌نمونه‌برداری به‌کار گرفته شده و در نهایت، مدل دسته‌بندی روی داده‌های آماده‌سازی‌شده اعمال شده است. آزمایش‌های گسترده‌ای برای مدل ارائه‌شده روی چند مجموعه دادهٔ مخابراتی برای پیش‌بینی ریزش انجام شده است. نتایج نشان می‌دهد که مدل ارائه‌شده در سه مجموعه داده نسبت به مدل‌های دیگر برتری دارد.

در مقالهٔ ویجایا و سیواسانکار[1] (۲۰۱۹) چند الگوریتم یادگیری ماشین برای پیش‌بینی ریزش مشتری در صنعت مخابرات بررسی شده است. علاوه‌براین استفاده از الگوریتم‌های فراابتکاری برای پیش‌بینی ریزش نیز به بحث گذاشته شده است. بدین منظور، تحلیلی روی استفاده از الگوریتم بهینه‌سازی ازدحام ذرات، به‌عنوان یک پیش‌بینی‌کننده ریزش انجام شده است. دسته‌بندهای پیشنهادی برای بررسی سطوح پیش‌بینی و جنبه‌های عملکردی با مدل‌هایی مانند درخت تصمیم، بیز ساده، k نزدیک ترین همسایه، ماشین بردار پشتیبان، جنگل تصادفی و سه مدل ترکیبی مقایسه شده‌اند. برای بررسی عملکرد مدل‌ها از معیارهایی مانند دقت، درستی، امتیاز F1، نمودار مشخصه عامل گیرنده و غیره استفاده شده است. آزمایش‌های صورت‌گرفته نشان می‌دهد که بهینه‌سازی ازدحام ذرات و انواع دیگر آن روی مجموعه داده‌های بزرگ و نامتعادل عملکرد خوبی داشته‌اند. همچنین مشاهده شد که استفاده از بهینه‌سازی ازدحام ذرات برای انتخاب ویژگی و بازپخت شبیه‌سازی‌شده به‌عنوان الگوریتم جست‌وجوی محلی در شرایط نامتعادل به خوبی عمل می‌کنند.

در مقالهٔ کاروانا و همکاران (۲۰۱۹) پیش‌بینی ریزش مشتری در صنعت بانکداری مورد بررسی قرار گرفته است. روش‌های دسته‌بندی درخت تصمیم، شبکهٔ عصبی، ماشین بردار پشتیبان، بیز ساده و رگرسیون لجستیک روی مجموعه داده‌ای در یک بانک خصوصی در اندونزی آزمایش شده است. داده‌های به‌دست‌آمده شامل ویژگی‌های جمعیت‌شناختی مشتریان از قبیل سن، حرفه، محصولات مورد استفاده، ویژگی‌هایی مربوط به تراکنش‌های ورودی یا خروجی، تراکنش‌های داخلی یا خارجی و همچنین ویژگی‌هایی مانند میانگین موجودی در بازه‌های زمانی مختلف، تعداد

1. Vijaya & Sivasankar

حساب‌های تحت مالکیت و غیره بوده است. آزمایش‌ها با نسبت‌های نمونه‌برداری مختلف از دسته‌های ریزش و عدم ریزش انجام شده است. نتایج به‌دست‌آمده نشان می‌دهند که تعداد نمونه‌های مورد استفاده برای آموزش تأثیر زیادی بر نتایج مدل دارد. همچنین نسبت اندازه دسته‌ها تأثیر زیادی بر نتایج معیار یادآوری دارد و با نسبت برابر دسته‌ها به معیار یادآوری بیشتری دست یافتند. ماشین بردار پشتیبان با نسبت نمونه‌گیری برابر از دسته‌ها بیشترین سود حاصل از پیش‌بینی ریزش مشتری رابه‌دست آورد.

در مقالهٔ چلیک و عثمان اوغلو[1] (۲۰۱۹) الگوریتم‌های یادگیری ماشین مانند شبکهٔ عصبی، درخت تصمیم، ماشین بردار پشتیبان، بیز ساده، k نزدیک‌ترین همسایه، تقویت گرادیان شدید و غیره برای تحلیل و پیش‌بینی ریزش مشتریان مقایسه شده‌اند.

در مقالهٔ هالیباس و همکاران (۲۰۱۹) تحلیل اکتشافی دادگان و مهندسی ویژگی برای پیش‌بینی ریزش مشتری در صنعت مخابرات مورد مطالعه قرار گرفته است. همچنین چندین روش دسته‌بندی شامل بیز ساده، رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، تقویت گرادیان و... روی مجموعه داده اعمال شده است. برای بررسی عملکرد مدل‌ها از معیارهای مختلفی مانند دقت، خطای دسته بندی، درستی، یادآوری، امتیاز F1 و مساحت زیر منحنی استفاده شده است. با توجه به نتایج به‌دست آمده در آزمایش، الگوریتم تقویت گرادیان بهترین عملکرد را داشته و تقریباً در تمام معیارها از سایر دسته‌بندها بهتر عمل کرده است. علاوه‌براین، عملکرد همهٔ دسته‌بندها پس از انجام بیش‌نمونه‌برداری از دسته اقلیت، با بهبود همراه بوده است.

در مقالهٔ جین، یاداو و مانوف[2] (۲۰۲۱) پیش‌بینی ریزش مشتری در سه حوزهٔ مختلف بانکداری، مخابرات و فناوری اطلاعات با استفاده از روش‌های یادگیری ماشین مورد مطالعه قرار گرفته است. برای پیش‌بینی ریزش مشتری در حوزه بانکی، از مجموعه داده‌ای عمومی، شامل ۱۰۰۰۰ نمونه و ۱۴ ویژگی استفاده شده است. ویژگی‌های مشتریان شامل امتیاز اعتباری، موقعیت جغرافیایی، جنسیت، سن، موجودی حساب، تعداد محصولات مورد استفاده، داشتن کارت اعتباری و غیره بوده است. برای پیش‌بینی ریزش از چهار مدل الگوریتمی رگرسیون لجستیک، جنگل تصادفی، ماشین بردار پشتیبان و تقویت گرادیان شدید استفاده شده و عملکرد الگوریتم‌های مختلف در حوزه‌های متفاوت مقایسه شده است. همچنین با تحلیل اکتشافی داده‌ها، راهبردهایی برای حفظ مشتری به‌دست آمده است. طبق نتایج به‌دست‌آمده، جنگل تصادفی در بخش بانکی با دقت ۸۶/۳۱۲ درصد، رگرسیون لجستیک در بخش فناوری اطلاعات با دقت ۹۰/۱۳۶ درصد و تقویت گرادیان شدید در بخش مخابرات با دقت ۸۲/۹۴۲ درصد بهترین عملکردها را به‌دست آوردند.

در مقالهٔ رحمان و کومار (۲۰۲۰) پیش‌بینی ریزش مشتری با استفاده از یادگیری ماشین در صنعت بانکداری انجام شده است. در این مطالعه دسته‌بندهای k نزدیک‌ترین همسایه، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی تحت شرایط مختلف آزمایش شده‌اند. همچنین برخی از روش‌های انتخاب ویژگی برای یافتن ویژگی‌های مرتبط‌تر و بررسی عملکرد سیستم مورد استفاده قرار گرفته‌اند. این آزمایش‌ها روی مجموعه دادهٔ ذکرشده در مورد قبلی انجام شده

1. Çelik & Osmanoglu
2. Jain, Yadav & Manoov

است. در نتایج به‌دست‌آمده، مدل جنگل تصادفی پس از بیش‌نمونه‌برداری با دقت پیش‌بینی ۹۵/۷۴ درصد عملکرد بهتری را نسبت به سایر مدل‌های آزمایش‌شده به‌دست آورده است.

در مقالهٔ سید، عبدالفتاح و خلیف[1] (۲۰۱۸) نیز پیش‌بینی ریزش مشتری در صنعت بانکداری مطالعه شده است. به‌منظور تولید مدلی برای پیش‌بینی احتمال روی‌گردانی مشتریان بانک از الگوریتم درخت تصمیم استفاده شده است.

در مقالهٔ لی و همکاران[2] (۲۰۲۱) پیش‌بینی ریزش مشتری در صنعت پخش رسانه‌های جمعی رادیو و تلویزیون انجام شده است. بدین منظور عوامل تأثیرگذار بر ریزش مشتری در صنعت شبکه‌های کابلی جمع‌آوری و بررسی شده است. برای کاهش ابعاد متغیرهای موجود از روش تحلیل مؤلفه اصلی و برای پیش‌بینی ریزش مشتری از الگوریتم دسته‌بندی رگرسیون لجستیک استفاده شده است.

جدول ۱ حوزهٔ کسب‌وکاری و مدل‌های استفاده‌شده در چند مقالهٔ بیان شده آورده شده است.

جدول ۱. حوزهٔ کسب‌وکار و مدل‌های استفاده‌شده برای پیش‌بینی ریزش مشتری در تعدادی از مقاله‌های سال‌های اخیر

مقاله	حوزه	مدل‌ها
(احمد و همکاران، ۲۰۱۹)	مخابرات	تقویت گرادیان، تقویت گرادیان شدید، جنگل تصادفی، درخت تصمیم
(لالوانی و همکاران، ۲۰۲۲)	مخابرات	بیز ساده، تقویت تطبیقی، تقویت گرادیان شدید، جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک، ماشین بردار پشتیبان
(ویجایا و سیواسانکار، ۲۰۱۹)	مخابرات	k نزدیک‌ترین همسایه، بیز ساده، جنگل تصادفی، درخت تصمیم، ماشین بردار پشتیبان
(کاروانا و همکاران، ۲۰۱۹)	بانکداری	بیز ساده، درخت تصمیم، رگرسیون لجستیک، شبکهٔ عصبی، ماشین بردار پشتیبان
(هالیباس و همکاران، ۲۰۱۹)	مخابرات	بیز ساده، تقویت گرادیان، جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک
(جین و همکاران، ۲۰۲۱)	بانکداری، مخابرات، فناوری اطلاعات	تقویت گرادیان شدید، جنگل تصادفی، رگرسیون لجستیک، ماشین بردار پشتیبان
(رحمان و کومار، ۲۰۲۰)	بانکداری	k–نزدیک‌ترین همسایه، جنگل تصادفی، درخت تصمیم، ماشین بردار پشتیبان
(سید و همکاران، ۲۰۱۸)	بانکداری	درخت تصمیم
(لی و همکاران، ۲۰۲۱)	پخش رسانه‌ای	رگرسیون لجستیک

استفاده از یادگیری عمیق

در میان روش های یادگیری ماشین، از یادگیری عمیق نیز در برخی مقالات پیش‌بینی ریزش مشتری استفاده شده است. در اکثر این کارها از چندین لایه پیش‌خور با تابع فعال‌سازی یک‌سوساز[3] به‌عنوان لایه‌های پنهان در ساختار شبکهٔ عصبی

1. Sayed, Abdel-Fattah & Kholief
2. Li et al.
3. Rectifier (ReLU)

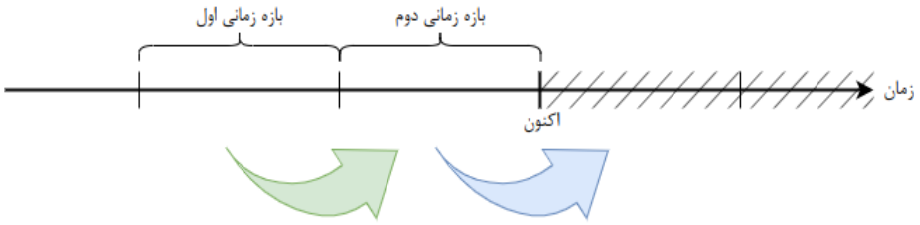
استفاده شده است. در ادامه به چند نمونه از این موارد اشاره می‌کنیم. در مقالهٔ اسپانودس و نگوین (۲۰۱۷) یک شبکهٔ عصبی پیش‌خور عمیق برای پیش‌بینی ریزش مشتری پیاده‌سازی شده است. لایه‌های پنهان با تابع فعال‌سازی یک‌سوساز و لایهٔ خروجی با دو نورون و تابع بیشینه هموار و به‌کارگیری روش‌های منظم‌سازی L1 و L2 پیاده‌سازی شده است. همچنین از عملیات حذف تصادفی[1] برای قدرت تعمیم بهتر استفاده شده است. برای آموزش مدل از الگوریتم بهینه‌سازی نزول گرادیان تصادفی همراه با تکانه و روش توقف زودهنگام استفاده شده است. در مقالهٔ خان و همکاران (۲۰۱۹) از شبکهٔ عصبی به‌صورت پرسپترون چندلایه برای پیش‌بینی ریزش مشتری در صنعت مخابرات استفاده شده است. در مقالهٔ اومایاپاراواتی و ایاکوتی[2] (۲۰۱۷) از شبکهٔ عصبی پیش‌خور و شبکهٔ عصبی پیچشی یک‌بُعدی برای یادگیری خودکار ویژگی از دادگان ورودی برای پیش‌بینی ریزش مشتری در صنعت مخابرات استفاده شده است. در مقالهٔ دومینگوس، اوجمه و دارامولا[3] (۲۰۲۱) تحلیل تجربی تأثیر ابرپارامترهای مختلف در استفاده از شبکه‌های عصبی عمیق برای پیش‌بینی ریزش مشتری در صنعت بانکی انجام شده است. از آزمایش‌های انجام‌شده نتیجه‌گیری شده است که شبکه‌های عصبی عمیق با استفاده از تابع فعال‌سازی یک‌سوساز در لایه‌های پنهان، نسبت به مدل‌های پرسپترون چندلایه عملکرد بهتری دارند. همچنین استفاده از الگوریتم بهینه سازی انتشار جذر میانگین مربعات در مقایسه با سایر الگوریتم‌ها، به یادگیری مدل با عملکرد بهتری منجر شده است.

روش‌شناسی پژوهش

پیش‌بینی ریزش مشتری یک مسئله دسته‌بندی دودویی است. برای مشتری iام بردار ویژگی $x_i \in R^n$ تشکیل می‌شود و $y_i \in \{۰,۱\}$ نشان‌دهندهٔ متغیر هدف به‌معنای ریزش یا عدم ریزش است. بدین ترتیب با در اختیار داشتن مجموعه داده $\{(x_1, y_1), \cdots, (x_m, y_m)\}$، هدف یادگیری تابع f است؛ به‌طوری که $f(x_i)$ تخمین مناسبی از مقدار y_i باشد. برای این منظور باید با استفاده از داده‌هایی که از مشتریان در اختیار است، ویژگی‌هایی را برای اشخاص جمع‌آوری کرد. با استفاده از این ویژگی‌ها، مشتریانی که احتمال روی‌گردانی آن‌ها در آینده وجود دارد، شناسایی می‌شوند.

برای پیش‌بینی ریزش مشتریان، دو بازه زمانی هم‌اندازه و متوالی را درنظر می‌گیریم. انتهای بازه زمانی دوم باید قبل‌تر از زمان حال باشد تا دادگان ثبت‌شده در اختیار باشند. رفتار مشتریان در بازهٔ زمانی اول را بررسی و ویژگی‌های رفتاری را برای مشتریان استخراج می‌کنیم. برای مشتریان کمیت معیار را در بازه‌های زمانی اول و دوم مقایسه کرده و متغیر هدف را مشخص می‌کنیم. بدین ترتیب با توجه به رفتار مشتری در بازه زمانی اول، سعی در پیش‌بینی متغیر هدف نسبت به بازه زمانی دوم داریم. پس از آموزش مدل دسته بند روی بازه زمانی اول، در نهایت از رفتار مشتری در بازه زمانی دوم استفاده می شود تا پیش‌بینی روی بازه زمانی متوالی در آینده صورت پذیرد. در شکل ۱ شمایی از نحوهٔ انتخاب بازه‌های زمانی برای پیش‌بینی رفتار مشتری به تصویر کشیده شده است.

1. Dropout
2. Umayaparvathi & Iyakutti
3. Domingos, Ojeme & Daramola



شکل ۱. نحوهٔ انتخاب بازه‌های زمانی برای پیش‌بینی رفتار مشتری

مهندسی ویژگی

در اجرای این طرح، دادگانی شامل اطلاعات مشتریان، اطلاعات حساب‌ها و کارت‌ها، مانده روزانه حساب‌ها و اطلاعات تراکنش‌های بانکی برای استفاده در دسترس بودند. با استفاده از اطلاعات ذکرشده به مهندسی ویژگی برای مشتریان بانک می‌پردازیم. مشتریان متشکل از اشخاص حقیقی و حقوقی هستند. برای اشخاص اطلاعات جمعیت‌شناختی مانند نوع مشتری، جنسیت، تاریخ تولد، وضعیت اقامت و غیره را در اختیار داریم. علاوه‌بر اطلاعات جمعیت شناختی، نیاز داریم تا ویژگی هایی از نحوهٔ رفتار مشتریان نیز به‌دست آوریم. مشتریان بانک می توانند دارای حساب های مختلفی باشند و با آن ها به انجام تراکنش های مالی بپردازند. در این پژوهش به‌منظور تحلیل رفتار مشتریان، اطلاعاتی را که برای حساب‌های مختلف به‌دست می‌آیند، تجمیع و ویژگی‌ها را به ازای اشخاص محاسبه می‌کنیم. در این پژوهش حساب‌های از نوع مشترک را در نظر نمی‌گیریم.

برای اشخاص مختلف، میانگین مانده مؤثر در طول بازه‌های زمانی اول و دوم را به‌طور جداگانه محاسبه می کنیم. طبق گفت‌وگو با متخصصان بانکی، ماندهٔ مؤثر معیار مهمی بوده و به‌معنای بخشی از سرمایهٔ بانک است که متعلق به هر مشتری است و بر مبنای آن اهمیت مشتریان سنجیده می شود. میانگین مانده مؤثر در بازه زمانی مورد مطالعه را به‌عنوان ویژگی برای مشتری در نظر می گیریم. همچنین از مقایسه میانگین مانده مؤثر در بازه های زمانی اول و دوم، متغیر هدف را به‌دست می‌آوریم. اگر میانگین مانده مؤثر شخصی در بازه زمانی دوم نسبت به خود شخص در بازه زمانی اول کاهش شدیدی داشته باشد، می تواند به‌معنای خروج سرمایه از بانک باشد. بدین ترتیب درصد زیاد کاهش در میانگین مانده مؤثر مشتری را به‌عنوان ریزش مشتری در نظر می‌گیریم.

برای جمع‌آوری سایر ویژگی‌های رفتاری، تاریخچه تراکنش‌های مشتریان در بازه زمانی مورد مطالعه را بررسی می‌کنیم. تراکنش‌های صورت‌حساب مشتریان می‌توانند به صورت واریز یا برداشت باشند. همچنین تراکنش‌ها می‌توانند به‌صورت جزئی‌تر از انواع مختلفی مانند برداشت از پایانه فروش، پرداخت قبض، خرید شارژ، خرید اینترنتی و غیره باشند. از انواع دیگر تراکنش‌ها می‌توان به انتقال از/به کارت شتاب یا داخلی، انتقالی داخلی، پایا و ساتنا نیز اشاره کرد.

برای مشتریان از تراکنش‌های آن‌ها در سطوح مختلف آماره‌هایی از قبیل تعداد، مجموع و میانگین مبالغ، مبالغ کمینه و بیشینه را محاسبه می کنیم. ابتدا آماره‌های ذکرشده را برای تراکنش‌ها به تفکیک واریز و برداشت محاسبه می کنیم. یعنی برای یک شخص در بازه زمانی مشخص به صورت کلی تعداد تراکش‌های واریز و برداشت، مجموع و

میانگین مبالغ واردشده به حساب‌ها و مبالغ خارج‌شده از حساب‌ها، کمترین و بیشترین مبالغ واردشده و خارج‌شده مشخص می‌شوند. سپس همین آماره‌ها را برای تراکنش‌های واریز و برداشت به تفکیک نوع تراکنش نیز به‌دست می‌آوریم. برای مثال برای یک شخص در بازه زمانی مشخص تعداد برداشت از پایانه فروش و مجموع، میانگین، کمینه و بیشینه مبالغ برداشت‌شده مشخص می‌شوند. همچنین این اعداد برای نوع تراکنش‌های دیگر مانند انتقالی کارت داخلی و خارجی و انتقالی حساب داخلی و خارجی به‌دست می‌آیند. بدین ترتیب با استفاده از تراکنش‌های بانکی به ویژگی‌هایی برای مشتریان که نمایانگر رفتار مالی آن‌ها باشد می‌رسیم.

آماده‌سازی و پیش‌پردازش دادگان

دادگان را به مجموعه‌های آموزش، اعتبارسنجی و آزمایش تقسیم می‌کنیم. مدل‌های مختلف را روی دادگان آموزش یادگیری کرده و برای مقایسه، معیار مساحت زیر منحنی مشخصه عامل گیرنده نسبت به دادگان اعتبارسنجی را مدنظر قرار می‌دهیم. هر ویژگی اسمی را به تعداد مقادیر متمایزش به متغیرهای دودویی تبدیل می‌کنیم. هر ویژگی عددی را نیز به‌منظور داشتن توزیع نرمال با میانگین $\mu = 0$ و انحراف معیار $\sigma = 1$ مقیاس می‌کنیم. بدین ترتیب با الحاق کردن اعداد به‌دست‌آمده، برای هر مشتری به یک بردار $x \in \mathbb{R}^n$ می‌رسیم.

از چالش‌هایی که برای مسئله پیش‌بینی ریزش مشتری وجود دارد، نامتعادل بودن مجموعه دادگان است. بدین صورت که مشتریان مربوط به دستهٔ ریزش، درصد کمتری از مجموعه دادگان را شامل می‌شوند. برای مقابله با این چالش، عکس نسبت اندازه دسته‌ها را به‌عنوان وزن دسته‌ها درنظر می‌گیریم. در این صورت طی مراحل آموزش مدل، وزن بیشتری به نمونه‌های دسته ریزش اختصاص داده می‌شود تا تعادل برقرار شود.

الگوریتم‌های دسته‌بندی یادگیری ماشین

ابتدا به‌عنوان دسته‌بندهای اولیه، مدل‌های k نزدیک‌ترین همسایه، درخت تصمیم، بیز ساده، رگرسیون لجستیک و ماشین بردار پشتیبان را آزمایش می‌کنیم. همچنین از روش‌های یادگیری جمعی جنگل تصادفی، تقویت تطبیقی و تقویت گرادیان نیز استفاده می‌کنیم.

استفاده از یادگیری عمیق

بعد پنهان مدل را برابر d در نظر گرفته و $h^{(0)} = x \in \mathbb{R}^n$ را فرض می‌کنیم. با انجام یک تبدیل خطی روی $h^{(0)}$، بعد آن را به d تغییر داده و سپس تابع غیرخطی یک‌سوساز را اعمال می‌کنیم تا به $h^{(1)} \in \mathbb{R}^d$ برسیم. به این تبدیل صورت‌گرفته، لایه پیش‌پردازش می‌گوییم. در ادامه بازنمایی پنهان مدل در لایه l–ام ($l > 1$) برابر خواهد بود با:

$$h^{(l)} = \max\left(0, W_l h^{(l-1)} + b_l\right) \quad \text{رابطهٔ ۱)}$$

که $W_l \in \mathbb{R}^{d \times d}$، $b_l \in \mathbb{R}^d$ و $h^{(l)} \in \mathbb{R}^d$ خواهند بود. با تکرار لایه ذکرشده می‌توان شبکه‌هایی عمیق را تشکیل داد.

در ادامه به‌منظور سهولت آموزش شبکه‌های عمیق‌تر، اتصال باقی‌مانده‌ای (هی، ژنگ، رِن و سان[1]، ۲۰۱۶) به‌صورت $\mathrm{F}(h) + h$ را به‌کار می‌گیریم. در حالت اول، به‌ازای هر لایه و در حالت دوم، برای هر دو لایه اتصال باقی‌مانده‌ای را برقرار می‌کنیم. در حالت تک لایه، تابع F مشابه رابطه ۱ خواهد بود. در حالت اتصال باقی‌مانده‌ای برای دو لایه، تابع F را به صورت زیر در نظر می‌گیریم:

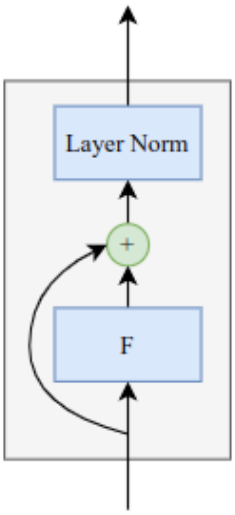
$$F(h) = W_2 max(0, W_1 h + b_1) + b_2$$ رابطهٔ ۲)

که عبارت است از یک تبدیل خطی، اعمال تابع غیرخطی یک‌سوساز و سپس یک تبدیل خطی دیگر. برای انجام عمل جمع، ابعاد ورودی و خروجی تابع F را برابر در نظر می‌گیریم. در صورت قرارگیری دو لایه پشت سر هم، بعد بردار میانی حاصل از تبدیل خطی اول می‌تواند با بعد ورودی و خروجی برابر نباشد. بنابراین $W_1 \in \mathbb{R}^{d' \times d}$ و $W_2 \in \mathbb{R}^{d \times d'}$ که d و d' ممکن است نابرابر باشند.

همچنین برای کاهش زمان آموزش، نرمال‌سازی لایه (با، کیروس و هینتون[2]، ۲۰۱۶) را پس از اتصال باقی‌مانده‌ای قرار می‌دهیم. در این صورت بازنمایی پنهان مدل در لایه l–ام $(l > 1)$ برابر خواهد بود با:

$$h^{(1)} = Layer\ Norm\left(F\left(h^{(1-1)}\right) + h^{(1-1)}\right)$$ رابطهٔ ۳)

با تکرار بلوک‌های فوق می‌توان شبکه‌هایی عمیق را تشکیل داد. بلوک حاصل مشابه زیرلایه‌های استفاده شده در معماری تبدیل‌کننده (وسوانی و همکاران[3]، ۲۰۱۷) است. نمایشی از این بلوک در شکل ۲ نشان داده شده است. همچنین نرمال‌سازی لایه را پس از لایهٔ پیش‌پردازش قرار می‌دهیم.



شکل ۲. بلوک شامل اتصال باقی‌مانده‌ای و نرمال‌سازی لایه

1. He, Zhang, Ren & Sun
2. Ba, Kiros & Hinton
3. Vaswani et al.

سرانجام برای اجرای دسته بندی نهایی، روی بازنمایی پنهان حاصل از آخرین بلوک $h^{(L)}$، یک رگرسیون لجستیک اعمال می‌کنیم. در این رابطه، که $w \in \mathbb{R}^d$، $b \in \mathbb{R}$، σ تابع سیگموید و $\hat{y} = \mathrm{P}(y = 1 \mid x)$ است.

رابطهٔ ۴) $$\hat{y} = \sigma\left(w^{\top} h^{(L)} + b\right)$$

یافته‌های پژوهش

این پژوهش در یکی از بزرگ‌ترین بانک‌های ایران اجرا شده و آزمایش‌های انجام‌شده روی دادگان واقعی مشتریان بانک صورت پذیرفته است. در این آزمایش ها از اطلاعات جمعیت شناختی و رفتار گذشتهٔ مشتریان بهره گرفته شده و اطلاعات قابل‌شناسایی شخصی افراد استفاده‌ای نداشته است. در آزمایش‌های صورت‌گرفته، پیش‌بینی ریزش مشتری روی بازهٔ زمانی یک ماهه انجام شده است. بدین ترتیب، دو بازهٔ یک ماهه متوالی در نظر گرفته شده و ویژگی‌های رفتاری مشتریان از بازهٔ زمانی اول استخراج شده است. متغیر هدف نیز از مقایسهٔ میانگین ماندهٔ مؤثر در بازه های زمانی اول و دوم به‌دست می‌آید. در صورتی که میانگین ماندهٔ مؤثر یک مشتری در بازهٔ دوم نسبت به بازهٔ اول با بیش از ۷۰ درصد کاهش همراه باشد، به‌عنوان ریزش در نظر گرفته می‌شود.

چنانچه گفته شد، تراکنش‌های بانکی شامل انواع مختلفی هستند. تعداد نوع‌های متمایز تراکنش‌ها زیاد بوده و برخی از آن‌ها برای مصارف خاص در نظر گرفته شده‌اند و به‌صورت مکرر استفاده نمی‌شوند. از این جهت برای استخراج ویژگی‌های رفتاری، نوع تراکنش‌های پُرتکرار مورد استفاده قرار گرفتند. بدین ترتیب تنها انواعی که در مجموع بیش از ۹۰ درصد تراکنش‌های انجام‌شده در بازهٔ زمانی اول را شامل می‌شدند، برای محاسبهٔ آماره‌ها انتخاب شدند.

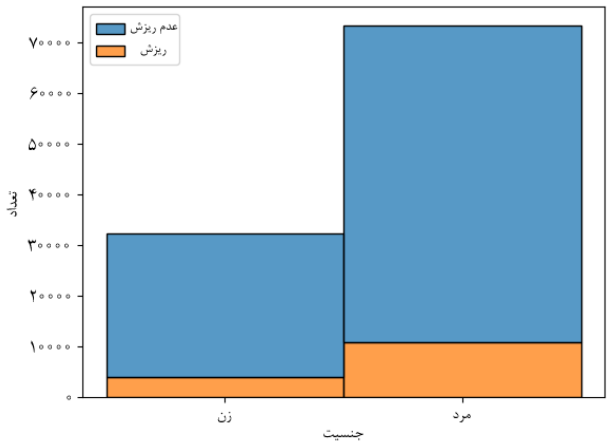
در مراحل پالایش دادگان، مشتریانی که در اطلاعات ثبت‌شده برای آن‌ها نقص یا ناسازگاری وجود داشت، کنار گذاشته شدند. به‌علت تعداد زیاد مشتریان بانک، برای انجام پیش‌بینی ریزش زیرمجموعه‌ای از اشخاص انتخاب و جداسازی شدند. در اطلاعات موجود برای تراکنش‌های انتقالی کارت داخلی، مبدأ و مقصد جابه‌جایی پول مشخص است. بدین ترتیب برای امکان بررسی تأثیر ارتباطات مالی میان اشخاص، جامعهٔ هدف به مشتریانی که در بازهٔ زمانی مورد مطالعه، تراکنش انتقالی کارت داخلی داشتند، محدود شد. مشتریان به‌دست‌آمده بر اساس میانگین ماندهٔ مؤثر در بازهٔ زمانی مورد مطالعه، به‌صورت نزولی مرتب‌سازی شدند. برای آزمایش‌های مختلف با اندازهٔ مجموعه داده آموزش متفاوت، تعدادی از مشتریان با ماندهٔ مؤثر بیشتر به‌صورت تصادفی انتخاب شدند.

دادگانی به اندازهٔ تقریبی ۱۱۰۰۰ نمونه به‌عنوان مجموعه اعتبارسنجی جداسازی شدند. سه مجموعه با اندازه‌های مختلف به‌عنوان دادگان آموزش در نظر گرفته شدند. اندازهٔ تقریبی هر مجموعه و درصد دسته های ریزش و عدم ریزش در جدول ۲ نشان داده شده است. عدم تعادل مجموعهٔ دادگان نیز در جدول ۲ مشاهده می‌شود. به ازای هر مجموعه آموزش، مدل‌ها را روی آن مجموعه یادگیری بررسی و عملکرد مدل روی مجموعه اعتبارسنجی را گزارش می‌کنیم. در این جدول K به‌معنای هزار است.

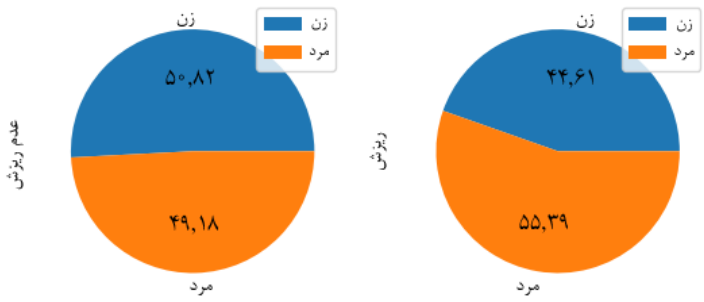
جدول ۲. اندازهٔ مجموعه‌های آموزش و درصد دسته‌های ریزش و عدم ریزش در هر مجموعه

ریزش عدم	ریزش	
۸۶/۲۹ درصد	۱۳/۷۱ درصد	۱۰۶K
۸۶/۲۵ درصد	۱۳/۷۵ درصد	۲۳۶K
۸۶/۲۳ درصد	۱۳/۷۷ درصد	۱۰۲۷K

در ادامه، به برخی از ویژگی‌های موجود در مجموعه آموزش ۱۰۶K نگاه می‌کنیم. در شکل ۳، نمودار ستونی فراوانی مقادیر متمایز ویژگی جنسیت در میان مشتریان حقیقی موجود در مجموعه داده ۱۰۶K نمایش داده شده است. همچنین در هر ستون، تعداد مشتریان بر حسب متغیر هدف با رنگ متفاوت مشخص شده است. مشاهده می‌شود که تعداد بیشتری از اشخاص حقیقی موجود در مجموعه داده را مردان تشکیل می‌دهند. در شکل ۴، درصد مردان و زنان به تفکیک دسته‌های ریزش و عدم ریزش در میان مشتریان حقیقی مجموعه داده ۱۰۶K به‌صورت نمودار دایره‌ای نمایش داده شده است. مشاهده می‌شود که ریزش در میان مردان بیشتر از زنان اتفاق افتاده است.

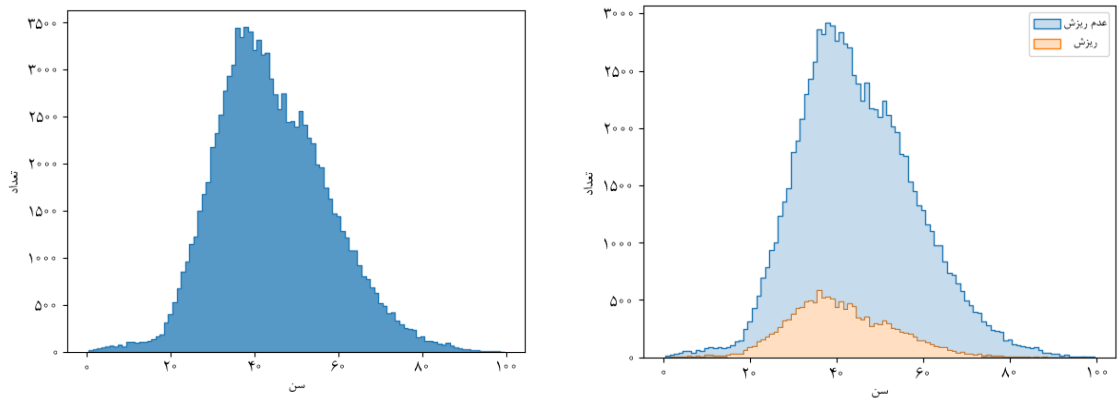


شکل ۳. نمودار فراوانی مقادیر متمایز ویژگی جنسیت در میان مشتریان حقیقی مجموعه داده ۱۰٦K

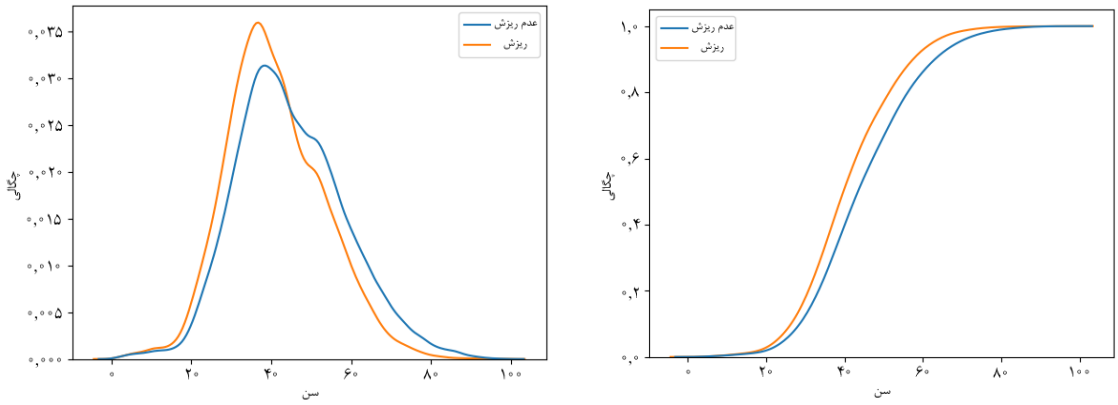


شکل ٤. نمودار دایره‌ای درصد شمول مردان و زنان، به‌تفکیک تعلق به دستهٔ ریزش در مشتریان حقیقی مجموعه داده ۱۰٦K

در شکل ۵ توزیع ویژگی سن در میان مشتریان حقیقی مجموعه داده ۱۰۶K به‌صورت کلی و به تفکیک تعلق به دستهٔ ریزش نمایش داده شده است. در شکل ۶ نیز نمودار تخمین چگالی کرنلی ویژگی سن مشتریان حقیقی، به تفکیک تعلق به دستهٔ ریزش، به‌صورت غیرتجمعی و تجمعی نمایش داده شده است. مشاهده می‌شود که ریزش در سنین پایین‌تر، بیشتر اتفاق افتاده است.



شکل ۵. نمودارهای توزیع ویژگی سن به‌صورت کلی و به‌تفکیک تعلق به دستهٔ ریزش برای اشخاص حقیقی مجموعه داده ۱۰۶K



شکل ۶. نمودارهای تخمین چگالی کرنلی ویژگی سن به تفکیک تعلق به دسته ریزش به صورت غیر تجمعی و تجمعی برای اشخاص حقیقی مجموعه داده ۱۰۶K

در جدول ۳ آماره‌هایی از ویژگی میانگین ماندهٔ مؤثر و ویژگی‌های مربوط به تراکنش‌های واریز و برداشت نمایش داده شده است. اعداد ارائه‌شده برای مبالغ، به ریال هستند. همان طور که مشاهده می‌شود، ویژگی میانگین ماندهٔ مؤثر دارای انحراف زیادی از میانگین است. با توجه به محل تجمع دادگان، توزیع این ویژگی چولگی مثبت دارد. این اتفاق به‌علت وجود اشخاصی با میانگین ماندهٔ مؤثر بیشتر نسبت به اکثریت مشتریان موجود در مجموعه داده رخ داده است.

اشخاص حقیقی موجود در مجموعه داده ۱۰۶K در بازه زمانی یک ماهه مورد مطالعه، به‌طور میانگین تعداد ۷۸ تراکنش برداشت داشته‌اند و حدود نیمی از اشخاص، بیش از ۶۱ برداشت از حساب انجام داده‌اند. حدود نیمی از مشتریان در تراکنش‌های برداشت خود، در مجموع بیش از ۴۰۶/۸ میلیون ریال و به‌طور میانگین بیش از ۶/۹ میلیون ریال، مبالغی را پرداخت کرده‌اند. همچنین حدود نیمی از مشتریان در تراکنش‌های واریز به حساب، در مجموع کمتر از ۴۶۴ میلیون ریال و به‌طور میانگین کمتر از ۳۰/۳ میلیون ریال، مبالغی را دریافت کرده‌اند. بیشترین مبلغ برداشت‌شده برای حدود ۲۵ درصد از اشخاص بیش از ۴۰۰ میلیون ریال بوده است. همچنین، بیشترین مبلغ واریزشده برای حدود ۲۵ درصد از مشتریان کمتر از ۵۰ میلیون ریال بوده است. سایر اطلاعات آماری مشابه نیز در جدول ۳ مشاهده می‌شود.

جدول ۳. آماره‌هایی از ویژگی میانگین مانده مؤثر و ویژگی‌های مربوط به تراکنش‌های واریز و برداشت در مجموعه داده ۱۰٦K

ویژگی	میانگین	معیار انحراف	25 صدک	۵۰ صدک	۷۵ صدک
مؤثر مانده میانگین	۸۱۴/۷ میلیون	۳/۰ میلیارد	۱۶۸/۹ میلیون	۲۹۰/۳ میلیون	۶۳۸/۱ میلیون
برداشت تعداد	۷۸	۷۰	۲۷	۶۱	۱۱۰
مبالغ مجموع	۱/۶ میلیارد	۶/۵ میلیارد	۱۲۵/۷ میلیون	۴۰۶/۸ میلیون	۱/۳ میلیارد
مبالغ میانگین	۲۶/۱ میلیون	۱۱۳/۶ میلیون	۲/۶ میلیون	۶/۹ میلیون	۲۰/۶ میلیون
مبالغ کمینه	۴۶۱/۰ هزار	۱۳/۹ میلیون	۱/۲ هزار	۳/۰ هزار	۲۰/۰ هزار
مبالغ بیشینه	۴۵۵/۴ میلیون	۱/۶ میلیارد	۳۰/۰ میلیون	۱۰۰/۰ میلیون	۴۰۰/۰ میلیون
واریز تعداد	۲۴	۷۵	۷	۱۴	۲۷
مبالغ مجموع	۱/۸ میلیارد	۸/۶ میلیارد	۱۴۰/۴ میلیون	۴۶۴/۰ میلیون	۱/۴ میلیارد
مبالغ میانگین	۸۴/۳ میلیون	۲۲۱/۹ میلیون	۱۲/۵ میلیون	۳۰/۳ میلیون	۷۹/۱ میلیون
مبالغ کمینه	۳/۷ میلیون	۵۶/۱ میلیون	۴/۳ هزار	۱۰۰/۰ هزار	۷۳۷/۵ هزار
مبالغ بیشینه	۵۴۸/۴ میلیون	۱/۷ میلیارد	۵۰/۰ میلیون	۱۰۴/۷ میلیون	۵۰۰/۰ میلیون

در جدول ۴ آماره‌هایی از ویژگی‌های مربوط به تراکنش‌های برداشت از پایانه فروش، به‌عنوان نمونه‌ای از نوع تراکنش‌های مختلف، نشان داده شده است. مشاهده می‌شود که اشخاص حقیقی موجود در مجموعه داده ۱۰۶K در بازه زمانی یک ماهه مورد مطالعه، به‌طور میانگین ۵۰ بار تراکنش برداشت از پایانه فروش انجام داده‌اند. همچنین در حدود نیمی از مشتریان حقیقی بیش از ۳۷ بار برداشت از پایانه فروش داشته‌اند. حدود نیمی از اشخاص در تراکنش‌های برداشت‌شده از پایانه فروش، در مجموع بیش از ۷۴/۹ میلیون ریال و به‌طور میانگین بیش از ۱/۸ میلیون ریال، مبالغی را پرداخت کرده‌اند. کمترین مبلغ برداشت‌شده در یک تراکنش توسط یک شخص، برای حدود ۲۵ درصد از مشتریان بیش از ۸۵ هزار ریال بوده است. به همین ترتیب، بیشترین مبلغ برداشت‌شده در یک تراکنش برای حدود ۲۵ درصد از مشتریان بیشتر از ۶۴ میلیون ریال بوده است. سایر اطلاعات مشابه، در جدول ۴ درج شده است.

جدول ٤. آماره‌هایی از ویژگی‌های مربوط به تراکنش‌های برداشت از پایانه فروش در مجموعه داده ۱۰٦K

ویژگی	میانگین	انحراف معیار	صدک ۲۵	صدک ۵۰	صدک ۷۵
تعداد تراکنش‌ها	۵۰	۴۹	۱۴	۳۷	۷۴
مجموع مبالغ	۲۰۷/۱ میلیون	۴۶۴/۸ میلیون	۲۵/۶ میلیون	۷۴/۹ میلیون	۲۰۰/۷ میلیون
میانگین مبالغ	۷/۷ میلیون	۳۴/۴ میلیون	۹۳۹/۸ هزار	۱/۸ میلیون	۴/۶ میلیون
کمترین مبلغ	۱/۵ میلیون	۲۳/۰ میلیون	۱۸/۰ هزار	۳۰/۰ هزار	۸۵/۰ هزار
بیشترین مبلغ	۶۹/۹ میلیون	۱۷۱/۶ میلیون	۶/۰ میلیون	۱۶/۲ میلیون	۶۴/۰ میلیون

جزئیات آموزش

مدل‌های مبتنی بر شبکهٔ عصبی را با استفاده از تابع ضرر آنتروپی متقاطع دودویی و الگوریتم بهینه‌سازی برآورد گشتاور تطبیقی (کینگما و با[1]، ۲۰۱۴) با نرخ یادگیری اولیه ۰/۰۱ آموزش می‌دهیم. در محاسبه کردن تابع ضرر، وزن بیشتری به نمونه‌های دسته ریزش اختصاص داده می‌شود تا تعادل برقرار شود. برای بررسی و مقایسهٔ عملکرد مدل‌ها در دسته‌بندی دودویی، از معیار مساحت زیرمنحنی مشخصه عامل گیرنده استفاده می‌کنیم. مدل‌هایی با پیکربندی‌های مختلف را روی مجموعه‌های آموزش یادگیری کرده و نتایج عملکردی آن‌ها روی دادگان اعتبارسنجی را به‌دست می‌آوریم. یادگیری مدل در صورت سپری شدن سه دوره متوالی و عدم افزایش مساحت زیرمنحنی مشخصه عامل گیرنده نسبت به مجموعه اعتبارسنجی، متوقف می‌شود. در تمامی مدل‌های شرح‌داده‌شده پس از اعمال تابع F در هر بلوک و همچنین پس از لایه پیش‌پردازش، عملیات حذف تصادفی با نرخ ۰/۱ اعمال شده است.

آزمایش‌ها

به‌عنوان مبنایی برای مقایسه‌ها، از چندین مدل یادگیری ماشین سنتی پُراستفاده، شامل درخت تصمیم، بیز ساده، kنزدیک‌ترین همسایه، رگرسیون لجستیک و ماشین بردار پشتیبان استفاده می‌کنیم. همچنین الگوریتم‌هایی از یادگیری جمعی شامل جنگل تصادفی، تقویت تطبیقی و تقویت گرادیان مورد استفاده قرار گرفته‌اند. به‌منظور یادگیری مدل‌های ذکرشده، در موارد ممکن برای برقرار تعادل میان دسته‌های ریزش و عدم ریزش با توجه به نسبت اندازه دسته‌ها وزن بیشتری به دستهٔ اقلیت داده شده است. نتایج به‌دست‌آمده از این آزمایش‌ها در جدول ۵ آمده است. هر کدام از مدل‌ها به‌ازای پیکربندی‌های متفاوت روی مجموعه‌های آموزش ۱۰۶K و ۲۳۶K یادگیری شده و تنها بهترین نتیجه به‌دست آمده روی مجموعه اعتبارسنجی، برحسب معیار مساحت زیرمنحنی مشخصه عامل گیرنده گزارش شده است. مشاهده می‌شود که در میان روش‌های سنتی یادگیری ماشین، درخت تصمیم بهترین عملکرد را با توجه به معیار مساحت زیرمنحنی نتیجه داده است. همچنین روش‌های جمعی نسبت به مدل‌های تکی نتایج بهتری را کسب کرده‌اند که از میان آن‌ها روش تقویت گرادیان منجر به به‌دست‌آمدن بیشترین مساحت زیر منحنی مشخصه عامل گیرنده شده است.

1. Kingma & Ba

جدول ۵. عملکرد الگوریتم‌های یادگیری ماشین سنتی به همراه روش‌های جمعی روی مجموعه‌های آموزش مختلف (برحسب معیار مساحت زیر منحنی مشخصه عامل گیرنده)

مدل	۱۰٦K	۲۳٦K
بیز ساده	۰/۷۴۵۴	۰/۷۴۵۴
K نزدیک‌ترین همسایه	۰/۷۵۰۳	۰/۷۴۸۹
ماشین بردار پشتیبان	۰/۷۷۷۰	۰/۷۸۵۱
رگرسیون لجستیک	۰/۸۳۸۲	۰/۸۳۹۶
درخت تصمیم	۰/۸۵۳۱	۰/۸۵۹۷
جنگل تصادفی	۰/۸۷۴۱	۰/۸۷۹۱
تقویت تطبیقی	۰/۸۷۱۶	۰/۸۷۲۵
تقویت گرادیان	۰/۸۹۸۴	۰/۹۰۱۰

در جدول ۶ نتایج به‌دست‌آمده از مدل‌های حاصل از تکرار لایه‌های کاملا متصل به‌صورت متوالی آمده است. حدودی از تعداد پارامترهای قابل یادگیری مدل‌ها به ازای مقادیر مختلف d و L ذکر شده است. مقدار L = ۰ به‌معنای قرارگیری لایهٔ دسته‌بندی‌کننده نهایی بلافاصله بعد از لایه پیش‌پردازش است. اعداد ذکرشده، نشان‌دهندهٔ معیار مساحت زیرمنحنی مشخصه عامل گیرنده نسبت به مجموعه اعتبارسنجی به ازای مجموعه‌های آموزش با اندازه متفاوت است. مشاهده می‌شود که بهترین نتیجه به‌دست‌آمده به‌ازای هر مجموعه آموزش با افزایش اندازه بهبود یافته است؛ اما افزایش تعداد لایه‌های پشت‌سرهم قرارگرفته، فرایند یادگیری مدل‌ها را دشوار کرده است و اغلب به کاهش معیار مساحت زیر منحنی منجر شده است.

جدول ٦. نتایج مدل‌های حاصل از انباشته‌کردن چندین لایه کاملا متصل پس از لایه پیش‌پردازش

تعداد پارامتر (×۱۰۰۰)	۱۰۲۷K	۲۳٦K	۱۰٦K	L	d
۱۸	۰/۸۷۵۰	۰/۸۷۴۳	۰/۸۶۲۱	۰	۱۲۸
۳۴	۰/۸۶۶۸	۰/۸۶۵۱	۰/۸۶۳۴	۱	
۵۱	۰/۸۲۰۲	۰/۸۴۰۹	۰/۸۳۶۵	۲	
۶۷	۰/۵۰۰۰	۰/۵۰۰۰	۰/۷۴۰۷	۳	
۸۴	۰/۵۰۰۰	۰/۷۷۶۸	۰/۵۰۰۰	۴	
۳۶	۰/۸۷۱۹	۰/۸۷۸۹	۰/۸۶۴۸	۰	۲۵۶
۱۰۱	۰/۸۵۹۲	۰/۸۶۵۶	۰/۸۵۳۷	۱	
۱۶۷	۰/۵۰۰۰	۰/۸۱۰۹	۰/۸۰۴۰	۲	
۲۳۳	۰/۵۰۰۰	۰/۵۰۰۰	۰/۵۰۰۰	۳	
۲۹۹	۰/۵۰۰۰	۰/۵۰۰۰	۰/۶۶۴۰	۴	

در جدول ۷ نتایج مدل‌هایی که برای هر لایهٔ کاملاً متصل، اتصال باقی‌مانده‌ای و نرمال‌سازی لایه قرار داده شده، آمده است. مقدار L = ۰ به‌معنای اعمال نرمال‌سازی لایهٔ پس از لایهٔ پیش‌پردازش و سپس قرارگیری لایهٔ دسته‌بندی‌کنندهٔ نهایی است. در مقایسه با آزمایش قبل، مشاهده می‌شود که استفاده از اتصال‌های پرشی به‌صورت باقی‌مانده‌ای و واحدهای نرمال‌سازی لایه، فرایند یادگیری مدل‌های عمیق‌تر را راحت‌تر و پایدارتر می‌سازند. افزایش اندازهٔ مجموعه آموزش نیز، به بهبود عملکرد مدل روی مجموعه اعتبارسنجی منجر شده است.

جدول ۷. نتایج حاصل از مدل‌های شامل اتصال باقی‌مانده‌ای برای هر یک لایه کاملاً متصل و نرمال‌سازی لایه

تعداد پارامتر (×۱۰۰۰)	۱۰۲۷K	۲۳۶K	۱۰۶K	*L*	*d*
۱۸	۰/۸۸۹۷	۰/۸۸۷۷	۰/۸۸۰۶	۰	۱۲۸
۳۵	۰/۸۹۵۵	۰/۸۹۰۹	۰/۸۸۲۵	۱	
۵۱	۰/۸۹۴۸	۰/۸۸۵۹	۰/۸۸۰۵	۲	
۶۸	۰/۸۹۳۶	۰/۸۹۰۶	۰/۸۸۱۷	۳	
۸۵	۰/۸۹۵۸	۰/۸۸۹۷	۰/۸۸۱۰	۴	
۳۶	۰/۸۹۱۵	۰/۸۸۶۰	۰/۸۸۱۲	۰	۲۵۶
۱۰۲	۰/۸۹۴۳	۰/۸۸۵۸	۰/۸۸۰۷	۱	
۱۶۹	۰/۸۹۴۸	۰/۸۸۸۳	۰/۸۸۰۷	۲	
۲۳۵	۰/۸۹۳۴	۰/۸۸۷۹	۰/۸۸۰۱	۳	
۳۰۱	۰/۸۹۳۰	۰/۸۸۶۷	۰/۸۷۹۳	۴	

در جدول ۸، نتایج به‌دست‌آمده از مدل‌های متشکل از اتصال باقی‌مانده‌ای برای هر دو لایه کاملاً متصل و نرمال‌سازی لایه آمده است. مشابه با نتایج قبل، در این آزمایش افزایش اندازهٔ مجموعه دادگان آموزش، به بهبود عملکرد مدل‌ها منجر شده است. با مقایسهٔ نتایج جداول ۷ و ۸، مشاهده می‌کنیم که بهترین عملکردهای به‌دست‌آمده به ازای هر مجموعه آموزش، بسیار نزدیک به هم هستند؛ اما تعداد پارامترهای قابل یادگیری بهترین مدل‌ها در جدول ۷، بسیار کمتر از بهترین مدل‌ها در جدول ۸ است؛ بنابراین استفاده از مدل‌های مشتمل بر اتصال باقی‌مانده‌ای برای هر یک لایهٔ کاملاً متصل، به لحاظ هزینهٔ محاسباتی مناسب‌تر به نظر می‌آید.

بدین‌ترتیب مدل‌های پُراستفاده در پیشینه پژوهش پیاده‌سازی شده و به‌منظور مقایسه نتایج با روش ارائه‌شده، بر روی مجموعه داده به‌دست‌آمده در این تحقیق اعمال شدند. به‌طور کلی مشاهده می‌کنیم که با استفاده از الگوریتم‌های یادگیری جمعی و همچنین روش‌های یادگیری عمیق، عملکرد مدل‌ها نسبت به مدل‌های مبنا بهبود می‌یابند. همچنین استفاده از مجموعه آموزش بزرگ‌تر، به بهبود عملکرد مدل روی مجموعه اعتبارسنجی منجر می‌شود؛ بنابراین می‌توان از مجموعه داده‌های بزرگ‌تر و روش‌های یادگیری عمیق بهره گرفت و پیش‌بینی ریزش مشتری را بهبود بخشید.

جدول ۸. نتایج حاصل از مدل‌های شامل اتصال باقی‌مانده‌ای برای هر دو لایه متوالی و نرمال‌سازی لایه

L	d	d'	۱۰۶K	۲۳۶K	۱۰۲۷K	تعداد پارامتر (×۱۰۰۰)
۱	۱۲۸	۱۲۸	۰/۸۸۱۱	۰/۸۸۹۰	۰/۸۹۴۹	۵۱
		۲۵۶	۰/۸۸۰۴	۰/۸۸۷۹	۰/۸۹۴۲	۸۴
		۵۱۲	۰/۸۸۲۳	۰/۸۹۰۲	۰/۸۹۴۹	۱۵۰
	۲۵۶	۲۵۶	۰/۸۸۲۱	۰/۸۹۰۸	۰/۸۹۵۰	۱۶۸
		۵۱۲	۰/۸۸۰۴	۰/۸۸۷۰	۰/۸۹۵۷	۳۰۰
		۱۰۲۴	۰/۸۷۹۵	۰/۸۸۸۱	۰/۸۹۵۸	۵۶۲
۲	۱۲۸	۱۲۸	۰/۸۷۹۱	۰/۸۸۸۳	۰/۸۹۳۹	۸۴
		۲۵۶	۰/۸۷۹۶	۰/۸۸۸۷	۰/۸۹۳۴	۱۵۰
		۵۱۲	۰/۸۸۲۲	۰/۸۸۷۶	۰/۸۹۳۵	۲۸۲
	۲۵۶	۲۵۶	۰/۸۸۰۴	۰/۸۸۸۹	۰/۸۹۵۳	۳۰۰
		۵۱۲	۰/۸۷۹۴	۰/۸۸۹۶	۰/۸۹۵۲	۵۶۳
		۱۰۲۴	۰/۸۷۸۸	۰/۸۸۷۴	۰/۸۹۴۵	۱۰۸۸

مدل برتر

در جدول ۹ دقت دسته‌بندی در پیش‌بینی ریزش مشتری برای بهترین مدل‌های به‌دست‌آمده از بخش قبل روی مجموعه اعتبارسنجی ذکر شده است. دقت‌های گزارش‌شده به‌ازای حد آستانه‌ای هستند که به به‌دست آمدن بیشترین امتیاز F1 برای هر مدل روی مجموعه اعتبارسنجی منجر شده است. باید توجه شود که نتایج به‌دست‌آمده برای مدل عمیق به‌صورت تکی و مدل تقویت گرادیان به‌صورت جمعی بوده است. عملکرد مدل عمیق پیشنهادشده از تمام روش‌های سنتی به‌صورت تکی و همچنین دو روش یادگیری جمعی استفاده‌شده دیگر، یعنی جنگل تصادفی و تقویت تطبیقی، برتر بوده است و با نتایج به‌دست آمده از الگوریتم تقویت گرادیان رقابت می کند. ممکن است با استفاده از مدل عمیق در قالب یادگیری جمعی، بهبود عملکردی بیشتری نیز حاصل شود.

جدول ۹. دقت دسته‌بندی الگوریتم تقویت گرادیان و مدل عمیق بر روی مجموعه اعتبارسنجی

مدل	۱۰۶K	۲۳۶K
گرادیان تقویت	۰/۸۷۴۴	۰/۸۷۲۴
عمیق	۰/۸۵۶۲	۰/۸۵۴۱

همچنین برای آموزش تقویت گرادیان و سایر مدل‌های جمعی استفاده‌شده روی یک ماشین، باید کل مجموعه داده در حافظه قرار گیرد؛ اما در خصوص مدل‌های مبتنی‌بر شبکهٔ عصبی، می‌توان مجموعه داده را در دسته‌های کوچک داخل حافظه بارگذاری کرد و آموزش مدل را با استفاده از دادگان بیشتر انجام داد. در مجموع بهترین عملکرد در پیش‌بینی

ریزش مشتری پس از آموزش روی مجموعه داده‌های ۱۰۶K و ۲۳۶K از منظر معیارهای مساحت زیر منحنی مشخصه عامل گیرنده و دقت دسته‌بندی نسبت به مجموعه اعتبارسنجی، توسط مدل تقویت گرادیان به‌دست آمد.

نتیجه‌گیری و پیشنهادها

در این مقاله مسئله پیش‌بینی ریزش مشتریان بانک را مورد بررسی قرار دادیم. هدف از مسئله پیش‌بینی ریزش، شناسایی مشتریانی است که احتمال روی‌گردانی آن‌ها در آینده وجود دارد. این پژوهش روی داده‌های واقعی مشتریان یکی از بزرگ‌ترین بانک‌های ایران انجام شد. با استفاده از حجم زیادی از تراکنش‌های بانکی و تجمیع آن‌ها در سطوح متفاوت، ویژگی های مختلفی برای مشتریان استخراج کردیم. به‌منظور پیش‌بینی ریزش، روش های یادگیری ماشین را روی ویژگی‌های مشتریان اعمال کردیم.

برای پیش‌بینی رفتار مشتریان، دو بازهٔ زمانی متوالی را در نظر گرفته و با استفاده از رفتار مشتری در بازهٔ اول، تلاش کردیم تا متغیر هدف در بازهٔ دوم را پیش‌بینی کنیم. بر مبنای میزان تغییر میانگین ماندهٔ مؤثر مشتری در بازهٔ زمانی دوم نسبت به بازهٔ اول، متغیر هدف را مشخص کردیم. با تجمیع تراکنش‌های مشتریان در سطوح مختلف، آماره‌هایی از رفتار مالی مشتریان را محاسبه کردیم و به‌عنوان ویژگی‌های رفتاری در نظر گرفتیم. یک مجموعه اعتبارسنجی ثابت برای مقایسهٔ مدل‌ها و تنظیم ابرپارامترها و سه مجموعه آموزش با اندازه‌های مختلف برای یادگیری مدل‌ها جداسازی کردیم. برای مشکل عدم تعادل مجموعه داده، عکس نسبت اندازهٔ دسته‌ها را به‌عنوان وزن دسته‌ها در فرایند آموزش تأثیر دادیم تا تعادل برقرار شود.

به‌عنوان مبنایی برای مقایسه، تعدادی از الگوریتم های پُراستفاده یادگیری ماشین سنتی را روی دادگان اعمال کردیم. علاوه‌بر آن‌ها، تعدادی از روش‌های یادگیری جمعی مرسوم را نیز استفاده کردیم. در ادامه، روش‌های یادگیری عمیق را استفاده کردیم و تعدادی از واحدهای نوین آن را به‌کار گرفتیم. آزمایش‌های گسترده‌ای را روی روش‌های ذکر شده به‌طور جداگانه انجام دادیم و نتایج را گزارش کردیم.

در نتایج به‌دست‌آمده مشاهده کردیم که روش‌های یادگیری جمعی و مدل‌های یادگیری عمیق، در پیش‌بینی ریزش مشتری، نسبت به مدل‌های مبنا عملکرد بهتری را ارائه می‌دهند. افزایش اندازهٔ مجموعه آموزش، در بهبود عملکرد مدل‌ها تأثیر دارد. در میان الگوریتم‌های دسته‌بندی یادگیری ماشین سنتی آموزش‌دیده روی مجموعه دادگان ۱۰۶K، درخت تصمیم بیشترین مساحت زیرمنحنی مشخصه عامل گیرنده به مقدار ۰/۸۵۳۱ را کسب کرد. در مجموع روی دادگان ۱۰۶K مدل تقویت گرادیان با ۰/۸۹۸۴ و پس از آن، مدل مبتنی‌بر یادگیری عمیق با ۰/۸۸۲۵، بیشترین مساحت زیرمنحنی مشخصه عامل گیرنده نسبت به مجموعه اعتبارسنجی را به‌دست آوردند.

پس از تحلیل رفتار و شناسایی مشتریان در شرف روی‌گردانی، می‌توان از پیشنهادهای کاربردی ذیل برای جلوگیری از ریزش و حفظ مشتری استفاده کرد. تفکیک مشتریان بر اساس سن، شغل، تحصیلات و غیره، به‌منظور ارائهٔ خدمات و تولید محصولات بانکی بر این مبنا، ایجاد تنوع در خدمات موجود، ارائهٔ خدمات مورد نیاز مشتریان از طریق

بسترهای مجازی و در صورت نیاز در محل فعالیت و زندگی مشتریان، تسهیل در ارائهٔ خدمات مطلوب به مشتریان، افزایش اعتماد مشتریان از طریق ارائه کاربردی، امن و همچنین حفاظت از اطلاعات مشتریان و غیره می‌تواند به حفظ مشتری کمک کند.

به‌علت حجم زیاد دادگان، محدودیت منابع سخت‌افزاری و حافظهٔ ذخیره‌سازی در اختیار، محدودیت زمانی و زمان‌بر بودن فرایند استخراج، پالایش و بارگذاری دادگان از پایگاه داده بانک، این تحقیق برای مطالعهٔ رفتار مشتریان روی بازهٔ زمانی یک ماهه انجام شده است. در ادامهٔ راه، می‌توان با تخصیص منابع سخت‌افزاری بیشتر، پیش‌بینی ریزش مشتری را روی بازه‌های زمانی بلندتری انجام داد. می‌توان علاوه‌بر تراکنش‌های مالی، از دادگان سایر زمینه‌های بانکی مانند تسهیلات و غیره نیز برای استخراج ویژگی‌های مشتریان استفاده کرد. همچنین، انتظار داریم که چارچوب ایجاد شده و راه‌کار ارائه‌شده در این طرح در تحلیل و پیش‌بینی رفتار مشتریان برای سایر مسائل موجود در زمینه مدیریت ارتباط با مشتری نیز کاربرد داشته باشد.

منابع

احمدی کوشا، آزاده؛ احمدی، فائق؛ رنجبر، محمدحسین و کردلوئی، حمیدرضا (۱۴۰۳). شناسایی شاخص‌های اعتبارسنجی و رتبه‌بندی مشتریان در تسهیلات خُرد در بانک خاورمیانه. *تحقیقات مالی*، ۲۶(۲)، ۴۱۵- ۴۳۸.

احمدی سرتختی، فرشید؛ هژبر کیانی، کامبیز؛ حسینی، سید شمس الدین و معمارنژاد، عباس (۱۴۰۲). طراحی مدلی برای ارزیابی ریسک اعتباری مشتریان ضمانت‌نامه‌های صادر شده توسط صندوق ضمانت صادرات ایران با کمک مدل شبکهٔ عصبی مصنوعی. *تحقیقات مالی*، ۲۵(۴)، ۶۹۵- ۷۱۶.

باجلان، سعید؛ فلاح‌پور، سعید و رئیسی، سارا (۱۴۰۳). بهینه‌سازی پرتفوی اعتباری بانک‌ها با استفاده از رویکرد اکچوئری و شبکهٔ عصبی مصنوعی. *تحقیقات مالی*، ۲۶(۳)، ۷۱۰- ۷۳۳.

رحیمی، سعیده؛ روستا، علی و آسایش، فرزاد (۱۴۰۳). ارزیابی ارتباط میان عوامل ارتقای توان رقابت‌پذیری خدمات ارزی مشتریان در صنعت بانکداری. *تحقیقات مالی*، ۲۶(۲)، ۴۳۹- ۴۶۲.

References

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.

Ahmadi Kousha, A., Ahmadi, F., Ranjbar, M. & Kordlouie, M. (2024). Validation Indicator Identification and Customer Ranking in Microloans: A Study at Middle East Bank in Iran. *Financial Research Journal*, 26(2), 399-423. (*in Persian*)

Ahmadi Sartakhti, F., Hojabr Kiani, K., Hoseini, S. & Memarnejad, A. (2023). Designing a Model for Credit Risk Assessment of Customers for Guarantees Issued by the Export Guarantee Fund of Iran via Artificial Neural Network Model. *Financial Research Journal*, 25(4), 641- 660. (*in Persian*)

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.

Ba, L. J., Kiros, J. R., & Hinton, G. E. (2016). *Layer Normalization*. arXiv preprint:1607.06450. https://doi.org/10.48550/arXiv.1607.06450

Çelik, O. & Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis. *Journal of Multidisciplinary Developments*, 4(1), 30-38.

De Caigny, A., 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.

Domingos, E., Ojeme, B. & Daramola, O. (2021). Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. *Computation*, 9(3), 34.

Halibas, A. S., Matthew, A. C., Pillai, I. G., Reazol, J. H., Delvo, E. G. & Reazol, L. B. (2019). Determining the Intervening Effects of Exploratory Data Analysis and Feature Engineering in Telecoms Customer Churn Modelling. *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, 1–7.

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Jain, H., Yadav, G. & Manoov, R. (2020). Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019* (pp. 137-156). Singapore: Springer Singapore.

Karvana, K. G. M., Yazid, S., Syalim, A. & Mursanto, P. (2019, October). Customer churn analysis and prediction using data mining models in banking industry. In *2019 international workshop on big data and information security (IWBIS)* (pp. 33-38). IEEE.

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).

Kingma, D. P. & Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015*, Conference Track Proceedings.

Lalwani, P., Mishra, M. K., Chadha, J. S. & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 104, 271-294.

Li, Y., Hou, B., Wu, Y., Zhao, D., Xie, A. & Zou, P. (2021). Giant fight: Customer churn prediction in traditional broadcast industry. *Journal of Business Research*, 131, 630-639.

Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M. & Shankar, K. (2023). Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. *Complex & Intelligent Systems*, 9, 3473–3485.

Raeesi, S., Bajalan, S. & Fallahpour, S. (2024). Bank's Credit Portfolio Optimization Using Actuarial Approach and Artificial Neural Networks. *Financial Research Journal*, 26(3), 710-733. (*in Persian*)

Rahimi, S., Rousta, A. & Asayesh, F. (2024). Evaluating the Relationship between Factors Enhancing the Competitiveness of Customer Foreign Currency Services in the Banking Industry. *Financial Research Journal*, 26(2), 424-446. (*in Persian*)

Rahman, M. & Kumar, V. (2020). Machine Learning Based Customer Churn Prediction in Banking. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1196–1201.

Sayed, H., Abdel-Fattah, M. A. & Kholief, S. (2018). Predicting potential banking customer churn using apache spark ML and MLlib packages: a comparative study. *International Journal of Advanced Computer Science and Applications*, 9(11).

Spanoudes, P. & Nguyen, T. (2017). *Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company Independent Feature Vectors*. https://arxiv.org/abs/1703.03869

Umayaparvathi, V., & Iyakutti, K. (2017). Automated feature selection and churn prediction using deep learning models. *International Research Journal of Engineering and Technology (IRJET)*, 4(3), 1846-1854.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems,* pp. 5998–6008.

Vijaya, J., & Sivasankar, E. (2019). An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, 22, 10757-10768.