پژوهش‌های نوین در تصمیم‌گیری
دوره ۱۰، شماره ٤، ویژه نامه مدیریت بازرگانی ١٤٠٤، صص ۱۸۱-۲۰۳
نوع مقاله: پژوهشی

تحلیل ریزش مشتریان در صنعت بانکداری با استفاده از الگوریتم‌های یادگیری ماشین

امید ولی زاده [۱]، مهسا اخوان راد [۲]، امیر محمد فکور ثقیه [۳*]

۱ .دانشجوی دکتری تحقیق در عملیات، گروه مدیریت، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد، مشهد، ایران
۲ .دانشجوی دکتری تحقیق در عملیات، گروه مدیریت، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد، مشهد، ایران
۳. دانشیار، گروه مدیریت، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد،مشهد، ایران

تاریخ دریافت: ١٤٠٤/٠٣/٠٩ تاریخ پذیرش: ١٤٠٤/٠٦/٣٠

چکیده

در صنعت رقابتی بانکداری، حفظ مشتریان به چالشی اساسی برای مؤسسات مالی تبدیل‌شده است. هدف این پژوهش شناسایی عوامل مؤثر بر ریزش مشتریان و پیش‌بینی آن با استفاده از پنج الگوریتم یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی، XGBoost، AdaBoost و KNN است. داده‌های دموگرافیکی و بانکیِ مربوط به مشتریان شعب یک بانک خصوصی در شهر مشهد (با توجه به درخواست بانک مورد نظر، در این مطالعه از آوردن نام آن خودداری شده است.) جمع‌آوری و به‌صورت دقیق تحلیل‌شده‌اند (شناسایی مقادیر تکراری، حذف مقادیر پرت و مدیریت مقادیر گم‌شده). ارزیابی عملکرد الگوریتم‌ها بر اساس معیارهای صحت، دقت، یادآوری و F۱-Score انجام شد. الگوریتم XGBoost با صحت ۹۲ درصد، بهترین عملکرد را در پیش‌بینی ریزش مشتریان داشته است. نتایج بیانگر آن است که عواملی همچون تعداد شکایات مشتریان، نوع حساب و دسته‌بندی مشتری (طلایی، نقره‌ای و سربی) مهم‌ترین شاخص‌ها در پیش‌بینی ریزش هستند. این یافته‌ها می‌توانند به بانک‌ها در طراحی استراتژی‌های مشخص و کاربردی مانند ارسال پیشنهادهای تسهیلات با نرخ ترجیحی برای مشتریان طلایی یا افزایش سرعت پاسخ‌دهی به شکایات دارندگان حساب جاری کمک نمایند. همچنین بهره‌گیری از این الگوریتم‌ها می‌تواند منجر به کاهش هزینه‌های عملیاتی و ارتقای اثربخشی برنامه‌های وفاداری مشتریان شود.

کلیدواژه‌ها: تحلیل داده، ریزش مشتریان، صنعت بانکداری، یادگیری ماشین

---

* نویسنده مسئول مقاله Email: amf@um.ac.ir

۱- مقدمه و بیان مسئله

صنعت بانکداری به‌خصوص در بخش بانکداری کوچک و متوسط در تمام دنیا به کسب‌وکاری کاملاً رقابتی تبدیل‌شده است. با توجه به این‌که خدمات اصلی و پایه‌ای ارائه‌شده به مشتریان در اکثر بانک‌ها تفاوت چندانی با یکدیگر ندارند، ریزش مشتریان به یکی از چالش‌های اساسی و بحرانی برای بانک‌ها و مؤسسات مالی تبدیل‌شده است. شواهد نشان می‌دهد که بانک و مؤسسات مالی سالیانه حدود ۲۵ درصد از مشتریان خود را از دست می‌دهند و هر کاهش یک‌درصدی در نرخ نگهداری مشتریان می‌تواند اثر قابل‌توجهی بر سودآوری شود [۱]. بانک‌ها برای تعامل مؤثر با مشتریان خود و دستیابی به مزیت رقابتی، ناگزیرند از استراتژی‌های اختصاصی استفاده کنند. این استراتژی‌ها برای اثربخشی بیشتر معمولاً بر پایه بهره‌گیری از فرآیندها و اجزای مختلف مدیریت ارتباط با مشتری و تصمیم‌گیری داده محور شکل می‌گیرند. در این میان، مدیریت ریزش مشتریان به‌عنوان یکی از مؤلفه‌های کلیدی این استراتژی‌ها، تنها به معنای حفظ مشتریان فعلی نیست، بلکه نقشی اساسی در افزایش جذابیت برند، ارتقای جایگاه رقابتی و جذب مشتریان جدید ایفا می‌کند. ریزش مشتریان غالباً تحت تأثیر عواملی چون نارضایتی از کیفیت خدمات، اقدامات بازاریابی رقبا، عرضه نوآوری‌های فناورانه توسط سایر بانک‌ها و تحولات سریع محیط کسب‌وکار شکل می‌گیرد. تداوم این وضعیت می‌تواند علاوه بر افزایش هزینه‌های تبلیغات و جذب مجدد مشتریان، به کاهش وفاداری و اعتماد مشتریان و درنهایت تضعیف اعتبار سازمانی بانک منجر شود [۲].

با گسترش فناوری اطلاعات، بانک‌ها توانسته‌اند با بهره‌گیری از داده‌های گسترده و متنوع حاصل از تعاملات مشتریان، تحلیل دقیق‌تری از رفتار آن‌ها ارائه دهند. این تحلیل‌ها زمینه‌ساز شناسایی الگوهای رفتاری پیچیده و طراحی مداخلات مؤثر به‌منظور کاهش نرخ ریزش مشتریان است [۳]. داده‌هایی که از طریق کانال‌های مختلف مانند تراکنش‌های مالی، کانال‌های دیجیتال و تعاملات چهره به چهره جمع‌آوری می‌شوند، می‌توانند به ابزارهای کلیدی برای شناسایی علل ریزش مشتریان تبدیل شوند [٤].

یادگیری ماشین به‌عنوان یکی از ابزارهای پیشرفته تحلیل داده، به بانک‌ها این امکان را می‌دهد که با دقت بیشتری الگوها و روندهای مرتبط با ریزش مشتریان را شناسایی کنند. این تکنیک‌ها قادر به پردازش حجم بالایی از داده‌ها هستند و می‌توانند الگوهای پیچیده‌ای را شناسایی کنند

که در تحلیل‌های سنتی معمولاً نادیده گرفته می‌شوند [٥]. در این راستا، الگوریتم‌های یادگیری ماشین نه‌تنها می‌توانند در شناسایی مشتریان در معرض ریزش پرکاربرد باشند، بلکه به بانک‌ها کمک می‌کنند تا اقداماتی پیشگیرانه و هدفمند برای کاهش ریزش مشتریان انجام دهند [٦]. از مزایای یادگیری ماشین می‌توان به جمع‌آوری و تحلیل داده‌ها، مدل‌سازی و پیش‌بینی، تشخیص عوامل مؤثر بر ریزش، استفاده از تحلیل پیش‌بینی و تحلیل توصیفی و تقویت تجربه مشتری اشاره کرد [٧].

در مطالعات پیشین مربوط به پیش‌بینی ریزش مشتریان در صنعت بانکداری، اغلب از یک یا دو الگوریتم محدود (مانند درخت تصمیم یا رگرسیون لجستیک) استفاده شده است و مقایسه هم‌زمان الگوریتم‌های مختلف در بستر داده‌های واقعی بانکی، به‌ویژه در ایران، کمتر مورد توجه قرار گرفته است. علاوه بر این، بسیاری از پژوهش‌ها عمدتاً در صنایع دیگر نظیر مخابرات یا تجارت الکترونیک انجام شده‌اند و از این رو یافته‌ها به دلیل تفاوت‌های ساختاری و رفتاری مشتریان، قابلیت تعمیم مستقیم به صنعت بانکداری ایران را ندارند.

پژوهش حاضر با هدف رفع این خلأ، پنج الگوریتم متنوع شامل درخت تصمیم، جنگل تصادفی، **XGBoost**، **AdaBoost** و **KNN** را انتخاب کرده است. معیار انتخاب این الگوریتم‌ها تنوع رویکردهای آن‌ها در یادگیری ماشین بوده است: درخت تصمیم به دلیل سادگی و قابلیت تفسیر بالا، جنگل تصادفی برای کاهش بیش‌برازش، **XGBoost** و **AdaBoost** به دلیل توانایی مدیریت داده‌های نامتوازن و استخراج الگوهای پیچیده، و **KNN** به‌عنوان الگوریتم مبتنی بر شباهت‌های محلی. این انتخاب متوازن باعث می‌شود که نقاط ضعف و قوت الگوریتم‌ها به‌صورت تطبیقی آشکار گردد و تصویری جامع‌تر از عملکرد مدل‌ها در صنعت بانکداری ایران ارائه شود.

نوآوری اصلی پژوهش حاضر در دو بعد است: نخست، مقایسه هم‌زمان و نظام‌مند این پنج الگوریتم بر داده‌های واقعی مشتریان یک بانک خصوصی در ایران، که تاکنون در ادبیات داخلی کمتر مورد توجه بوده است؛ دوم، شناسایی و رتبه‌بندی عوامل کلیدی مؤثر بر ریزش مشتریان (مانند تعداد شکایات و نوع حساب) که می‌تواند به‌طور مستقیم مبنای طراحی استراتژی‌های عملی بانک‌ها در مدیریت وفاداری و کاهش ریزش قرار گیرد. این رویکرد ترکیبی، علاوه بر پر

کردن شکاف موجود در مطالعات پیشین، به افزایش قابلیت کاربردی و تصمیم‌گیری مبتنی بر داده در صنعت بانکداری ایران کمک خواهد کرد.

بنابراین تحقیق حاضر سعی دارد با استفاده از الگوریتم‌های مختلف یادگیری ماشین به تجزیه‌وتحلیل عوامل مؤثر بر ریزش مشتریان بپردازد.

۲- پیشینهٔ پژوهش

۲-۱- ریزش مشتریان در صنعت بانکداری

ریزش مشتریان به معنای ترک آن‌ها به دلایل مختلفی همچون نامناسب بودن سرویس‌دهی، ارتباط ضعیف با مشتری و عدم توجه به نیازهای مشتری و تغییر مقررات یک کسب‌وکار است و این موضوع در صنعت بانکداری از اهمیت زیادی برخوردار است [۵]. از دست دادن مشتریان مستقیماً منجر به کاهش درآمد بانک‌ها و درعین‌حال افزایش هزینه‌های مربوط به جذب مشتریان جدید می‌گردد. مطالعات نشان داده‌اند که هزینه جذب یک مشتری جدید چندین برابر بیشتر از حفظ مشتری موجود است. همچنین، فرآیند اعتمادسازی و شکل‌گیری ارتباط مؤثر با مشتریان جدید مستلزم صرف منابع زمانی، مالی و انسانی قابل‌توجهی است. در کنار این موارد، ریزش مشتریان می‌تواند موجب آسیب به اعتبار برند بانک شده و شهرت سازمان را در معرض خطر قرار دهد. این آسیب‌ها با توجه به گسترش شبکه‌های اجتماعی و افزایش قدرت انتقال تجربیات منفی مشتریان، می‌تواند تشدید شود و بر تصمیم‌گیری سایر مشتریان بالقوه نیز تأثیر منفی بگذارد [۸]. مدیریت ریزش مشتریان از طریق تکنیک‌های پیش‌بینی و تحلیل داده‌ها، راهبردی مؤثر در بهبود وفاداری مشتریان و افزایش سهم بازار به شمار می‌رود [۹]. اهمیت بررسی ریزش مشتریان باعث شده مطالعات بسیاری به توسعه مدل‌های پیش‌بینی ریزش مشتریان بپردازند و با استفاده از داده‌های مربوطه الگوهای ریزش مشتریان را شناسایی کنند.

۲-۲- روش‌های سنتی پیش‌بینی ریزش مشتریان

پیش از گسترش روش‌های نوین یادگیری ماشین، ابزارهای آماری کلاسیک مانند رگرسیون از رایج‌ترین رویکردها برای مدل‌سازی رفتار مشتری و پیش‌بینی احتمال ریزش او بودند [۱۰]. این دسته از روش‌ها غالباً بر مبنای فرض وجود روابط خطی میان متغیرهای مستقل و وابسته شکل‌گرفته‌اند. در این میان، رگرسیون به دلیل سادگی ساختاری و امکان تفسیر نتایج، یکی از

پرکاربردترین رویکردها به شمار می‌رود که در مطالعات اولیه مرتبط با ریزش مشتری به‌طور گسترده مورداستفاده قرارگرفته است. هدف اصلی این روش، برآورد احتمال وقوع رویدادهایی مانند ترک یا وفاداری مشتریان است [۱۱]. اگرچه این رویکردها برای پیش‌بینی ریزش در گذشته موفقیت‌هایی داشتند، اما با ظهور داده‌های حجیم و پیچیدگی‌های فزاینده رفتار مشتریان، محدودیت‌ها و نقاط ضعف این روش‌ها بیشتر نمایان شد. روش‌های سنتی نمی‌توانستند به خوبی با داده‌های چندبعدی و غیرخطی که امروزه در بانکداری رایج است، سازگار شوند [۱۲]. ازجمله این چالش‌ها می‌توان به عدم توانایی در شناسایی تعاملات غیرخطی میان متغیرها و محدودیت در پردازش داده‌های حجیم اشاره کرد. به همین دلیل، با پیشرفت تکنولوژی و ظهور روش‌های نوین داده‌کاوی و یادگیری ماشین، استفاده از الگوریتم‌های پیچیده‌تر و کارآمدتر موردتوجه قرار گرفت. این روش‌ها می‌توانند با بهره‌گیری از تکنیک‌های پیشرفته، الگوهای پنهان در رفتار مشتریان را شناسایی کرده و به‌صورت دقیق‌تر احتمال ریزش آن‌ها را پیش‌بینی کنند [۱۳].

۲-۳- استفاده از الگوریتم‌های یادگیری ماشین در پیش‌بینی ریزش مشتریان

با پیشرفت فناوری و افزایش توان در پردازش داده‌های کلان، الگوریتم‌های یادگیری ماشین به‌عنوان رویکردی مؤثر برای تحلیل داده‌ها و پیش‌بینی ریزش مشتریان موردتوجه قرارگرفته‌اند [١٤]. یادگیری ماشین این امکان را فراهم می‌سازد تا سازمان‌ها با بهره‌گیری از حجم بالای داده‌ها، الگوهای پنهان و پیچیده را شناسایی کرده و اقدامات پیشگیرانه‌ای برای کاهش نرخ ریزش مشتریان اتخاذ کنند. با توجه به رقابت شدید در صنعت بانکداری و اهمیت حفظ مشتریان، بانک‌ها به‌طور گسترده‌ای از تکنیک‌های یادگیری ماشین برای تحلیل رفتار مشتریان و پیش‌بینی ریزش آن‌ها استفاده می‌کنند [١٥].

مدل‌های یادگیری ماشین با قابلیت تحلیل هم‌زمان داده‌های متنوع و چندبعدی، قادرند تغییرات تدریجی در الگوهای رفتاری مشتری را ردیابی کرده و نشانه‌های کاهش تعامل را پیش از وقوع ریزش شناسایی کنند. این فرآیند، نه‌تنها به شناسایی مشتریان در معرض ریسک کمک می‌کند، بلکه اطلاعات ارزشمندی را برای طراحی اقدامات در زمان مناسب در اختیار واحدهای بازاریابی و مدیریت ارتباط با مشتری قرار می‌دهد؛ اقداماتی که می‌توانند بر اساس نوع و شدت تغییر رفتار، به‌صورت هدفمند و متناسب با نیاز هر مشتری تنظیم شوند [١٦].

۲-۴- پیشینه تجربی

در ادامه به پیشینه تحقیق در رابطه با ریزش مشتریان می‌پردازیم. بررسی پیشینه تحقیق نشان می‌دهد در سال‌های اخیر مطالعات در مورد ریزش مشتری در صنایع مختلف افزایش چشمگیری داشته است. با توجه به پیشرفت‌های روزافزون تکنیک‌های هوش مصنوعی، استفاده از روش‌های سنتی (مانند رگرسیون) جای خود را به‌روش‌های جدیدتر داده است. درعین‌حال، معرفی روش‌های کارآمدتر در حوزه هوش مصنوعی، نیاز به مطالعات تازه در مورد ریزش مشتریان را ضروری می‌سازد. در جدول ۱ برخی مطالعات در حوزه ریزش مشتریان آمده است.

جدول ۱: خلاصه‌ای از پیشینه تحقیق

ردیف	رفرنس	مطالعه موردی	سال انتشار	الگوریتم‌های به‌کاررفته				
				XGBoost	جنگل تصادفی	درخت تصمیم	KNN	AdaBoost
۱	[۱۷]	شرکت مخابراتی	۲۰۲۲		✓	✓		
۲	[۱۸]	مشاغل اشتراکی	۲۰۲٤	✓		✓	✓	
۳	[٦]	صنعت بانکداری	۲۰۲۳	✓		✓		
٤	[۱۲]	مؤسسه مالی	۲۰۲٤		✓	✓	✓	
٥	[۱۹]	بازارهای مالی	۲۰۲٤			✓		
٦	[۲۰]	تجارت الکترونیک	۲۰۲۳		✓	✓		
۷	[۲۱]	بازارهای مالی	۲۰۲٤		✓		✓	
۸	[۲۲]	صنعت هتل داری	۲۰۲٤	✓	✓		✓	
۹	[۲۳]	صنعت نرم افزار	۲۰۲۲		✓	✓		✓
۱۰	[۲٤]	صنعت لوازم خانگی	۲۰۲۳	✓	✓	✓		
۱۱	مطالعه حاضر	بانک خصوصی		✓	✓	✓	✓	✓

طبق جدول ۱ تعداد قابل توجهی از مطالعات به تحلیل و پیش‌بینی ریزش مشتریان با استفاده از الگوریتم‌های یادگیری ماشین پرداخته‌اند. مطالعاتی همچون [۱۷]، [۱۲]، [۲۱]، [۲۵] و [۲۰] از الگوریتم‌هایی مثل جنگل تصادفی و درخت تصمیم و *KNN* به بررسی و تحلیل ریزش مشتریان در شرکت‌های مخابراتی و مؤسسات مالی پرداختند. در این مطالعات متغیرهایی همچون مدت عضویت مشتری، تعداد شکایات ثبت‌شده و نوع خدمات دریافت‌شده موردبررسی قرارگرفته‌اند. نتایج این پژوهش‌ها نشان داده‌اند که الگوریتم‌های مبتنی بر تجمیع مانند جنگل تصادفی در محیط‌هایی که داده‌ها پیچیده هستند، دقت بالاتری ارائه می‌دهند. همچنین، متغیرهای رفتاری مانند تعداد شکایات و سطح تعاملات به‌عنوان مهم‌ترین عوامل تأثیرگذار در پیش‌بینی ریزش مشتریان شناسایی‌شده‌اند.

همچنین تحقیقاتی مثل [۱۸]، [۲۲]، [۲۳] و [٢٤] با الگوریتم‌های مشابه و علاوه برآن الگوریتم‌های جدیدی مثل *XGBoost* و *Adaboost* به بررسی و تحلیل ریزش مشتریان در صنایعی نظیر خدمات اشتراکی، هتلداری، نرم‌افزار به‌عنوان سرویس (*SaaS*) و اجاره لوازم‌خانگی پرداخته‌اند. این مطالعات از متغیرهایی مانند تعداد تمدید قراردادها، رضایت مشتریان از خدمات، تعاملات دیجیتال، تعداد اقامت‌های مکرر در هتل‌ها استفاده کرده‌اند. نتایج این پژوهش‌ها نشان داده است که الگوریتم‌های تقویتی نظیر *XGBoost* و *AdaBoost* به دلیل توانایی در مدیریت داده‌های نامتوازن و استخراج الگوهای پیچیده، دقت بالاتری در پیش‌بینی ارائه می‌دهند. این الگوریتم‌ها همچنین توانایی بالایی در تشخیص مشتریان با ریسک بالای ریزش داشته و امکان اقدامات پیشگیرانه برای حفظ مشتریان را فراهم کرده‌اند.

تمرکز بیشتر مطالعات انجام‌شده بر پیش‌بینی و شناسایی عوامل مؤثر ریزش مشتریان در صنایع مختلف بوده است که عمدتاً از الگوریتم‌های مختلف یادگیری ماشین جهت تحلیل این پدیده استفاده کرده‌اند. بااین‌حال، مطالعات کمتری به تحلیل ویژگی‌های خاص، الگوهای رفتاری و عوامل مرتبط با تعاملات مشتریان در صنعت بانکداری پرداخته‌اند. به‌ویژه، توجه به داده‌های چندبعدی، تحلیل متغیرهای رفتاری و تعاملات مالی مشتریان و ارائه راهکارهای هدفمند جهت کاهش ریزش در محیط‌های بانکی کمتر موردبررسی قرارگرفته است. مطالعه حاضر این شکاف را با رویکردی جامع و دقیق پر می‌کند. ابتدا با تحلیل داده‌های مرتبط با مشتریان یک بانک خصوصی در مشهد، متغیرهایی مانند مدت عضویت، تعداد شکایات، نوع حساب بانکی و...

بررسی قرارگرفته‌اند. سپس با استفاده از الگوریتم‌های یادگیری ماشین نظیر *XGBoost*، جنگل تصادفی، درخت تصمیم، *AdaBoost* و *KNN* مدل‌هایی برای پیش‌بینی ریزش مشتریان به کار گرفته‌شده‌اند.

۳- روش‌شناسی پژوهش

هدف اصلی این پژوهش تحلیل ریزش مشتریان با استفاده از الگوریتم‌های یادگیری ماشین است. برای این منظور، داده‌های مربوط به ریزش مشتریان در یکی از بانک‌های خصوصی فعال در شهر مشهد مقدس مورد تحلیل و بررسی قرارگرفته است. جامعه آماری پژوهش حاضر، شامل ٦٤ شعبه فعال بانک در شهر مشهد می باشد. با توجه به درجه‌بندی کیفی شعب بانک (درجه ۱ تا ٤)، از روش نمونه‌گیری هدفمند استفاده شد؛ به‌گونه‌ای که تنها شعب دارای درجه اول و دوم به دلیل اهمیت راهبردی و حجم بالای مشتریان، وارد نمونه شدند. بر این اساس، ۳۰ شعبه به‌عنوان نمونه پژوهش انتخاب شد و تعداد ۳٦٥۱۳ مشتری مربوط به این شعب می‌باشد.

برای پاک‌سازی داده‌ها، مجموعه‌ای از اقدامات شامل حذف داده‌های تکراری، شناسایی و حذف مقادیر پرت و مدیریت مقادیر گمشده انجام شد. داده‌های تکراری از طریق بررسی تطابق کامل مقادیر در تمامی ستون‌ها شناسایی و حذف گردیدند. برای شناسایی مقادیر پرت، از روش آماری مبتنی بر بازه‌ی بین چارکی[۱] استفاده شد و داده‌های خارج از محدوده قابل‌قبول حذف شدند. درزمینه‌ی مدیریت مقادیر گمشده، از روش *KNN Imputation*[۲] بهره گرفته شد. در این روش، مقدار گمشده‌ی هر ویژگی با میانگین مقادیر همان ویژگی در k نمونه‌ی مشابه جایگزین می‌شود؛ میزان شباهت بین نمونه‌ها بر اساس فاصله‌ی اقلیدسی در فضای ویژگی‌ها محاسبه می‌گردد. مقدار $k = 5$ در نظر گرفته شد تا تعادلی میان تعداد همسایه‌ها و دقت جایگذاری برقرار شود.

پس از آماده‌سازی داده‌ها، الگوریتم‌های مختلف روی آن‌ها اعمال و مدل‌های مربوطه با استفاده از معیارهای صحت[۳]، دقت[٤]، بازیابی[٥] و *F۱-Score* ارزیابی شدند. در ادامه هر یک از معیارها

---

۱. Interquartile Range
۲. K-Nearest Neighbors Imputation
۳. Accuracy
٤. precision
٥. Recall

و نقش آن‌ها در ارزیابی مدل پیش‌بینی ریزش مشتریان بانک بررسی‌شده است. در این پژوهش، داده‌ها به صورت تصادفی به دو مجموعه آموزش (۷۰ درصد) و تست (۳۰ درصد) تقسیم شدند تا مدل‌های یادگیری ماشین آموزش‌دیده و سپس ارزیابی شوند.

معیار صحت [۲٦] یکی از شاخص‌های مهم برای ارزیابی عملکرد مدل‌های طبقه‌بندی در پیش‌بینی ریزش مشتریان است و نشان می‌دهد چه نسبتی از پیش‌بینی‌ها درست انجام شده است. صحت بالا بیانگر توانایی مدل در شناسایی مشتریان در حال ریزش است که به بانک کمک می‌کند با تصمیمات به‌موقع، نرخ ریزش را کاهش دهد. بااین‌حال، به دلیل نامتعادل بودن داده‌ها (تعداد بیشتر مشتریان ماندگار نسبت به در معرض ریزش)، اتکا به صحت به‌تنهایی کافی نیست و استفاده از معیارهایی مانند دقت، یادآوری و *F۱-Score* تصویری جامع‌تر از عملکرد مدل ارائه می‌دهد.

دقت [۲٦]، یکی از معیارهای کلیدی در ارزیابی مدل‌های یادگیری ماشین محسوب می‌شود که نسبت پیش‌بینی‌های مثبت صحیح به کل پیش‌بینی‌های مثبت را نشان می‌دهد. در مسئله پیش‌بینی ریزش مشتریان بانک، دقت بالا بیانگر توانایی مدل در کاهش هشدارهای کاذب و تمرکز صحیح بر مشتریان در معرض ریزش است. بااین‌حال، به‌منظور ارزیابی جامع عملکرد مدل، توجه هم‌زمان به معیارهایی نظیر یادآوری و *F۱-Score* ضروری است تا از شناسایی ناقص مشتریان در آستانه ریزش جلوگیری شود و اقدامات نگهداشت به‌درستی مورد استفاده قرار گیرد.

یادآوری [۲۷]، معیاری است که توانایی مدل در شناسایی تمام موارد مثبت واقعی را می‌سنجد. این معیار در پیش‌بینی ریزش مشتریان بانک اهمیت دارد زیرا نشان می‌دهد مدل تا چه حد توانایی شناسایی مشتریان در معرض ریزش را دارد. یادآوری بالا بیانگر توانایی مدل در شناسایی اکثر مشتریان در معرض ریزش است، حتی اگر برخی پیش‌بینی‌ها نادرست باشند. این معیار زمانی اهمیت پیدا می‌کند که هزینه از دست دادن مشتریان واقعی بالا باشد؛ بنابراین افزایش یادآوری به بانک‌ها در حفظ مشتریان و کاهش ریزش کمک می‌کند.

F۱-Score [۲۷] به‌عنوان میانگین هارمونیک دقت و یادآوری، معیاری متعادل برای ارزیابی عملکرد مدل‌های طبقه‌بندی محسوب می‌شود. در زمینه پیش‌بینی ریزش مشتریان، این شاخص توانایی مدل را در شناسایی صحیح مشتریان در حال ریزش در کنار کاهش هشدارهای

نادرست نشان می‌دهد. مقدار بالای F۱-Score بیانگر تعادل مطلوب بین دقت و یادآوری بوده و به بانک‌ها در اتخاذ راهکارهای مؤثر برای حفظ مشتریان کمک می‌کند.

برای تجزیه‌وتحلیل داده‌ها از روش‌های یادگیری ماشین استفاده‌شده است. در همین راستا، تحقیق حاضر از پنج الگوریتم قدرتمند یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی، *XGBoost*، *AdaBoost* و *KNN* استفاده کرده است. در ادامه نتایج حاصل از پنج الگوریتم ازنظر کیفیت و دقت با یکدیگر مقایسه شده‌اند. در ادامه به معرفی روش‌های یادگیری ماشین پرداخته‌شده است.

۳-۱- الگوریتم *XGBoost*

الگوریتم *XGBoost* برای اولین بار توسط چن و گوسترین[۱] (۲۰۱۶) مطرح شد. سازوکار اصلی این الگوریتم مبتنی بر پایه روش تقویت است؛ به‌گونه‌ای که پیش‌بینی‌های حاصل از چند یادگیرنده‌ی ضعیف به‌صورت مرحله‌ای ترکیب می‌شوند تا در نهایت یک مدل قوی شکل گیرد. از مزایای این روش می‌توان به پشتیبانی از مدل‌های خطی و غیرخطی، انعطاف‌پذیری بالا و تنظیمات پارامترهای مختلف و سرعت‌بالا در پردازش داده‌ها اشاره کرد. در این الگوریتم، مدل اولیه با داده‌های ورودی تطبیق داده می‌شود و در ادامه مدل دوم برای جبران خطاهای مدل اول اعمال می‌شود. این فرایند چندین بار تکرار می‌شود تا زمانی که معیار توقف محقق شود. درنهایت، پیش‌بینی نهایی از ترکیب پیش‌بینی‌های هر مدل به دست می‌آید.

۳-۲- الگوریتم درخت تصمیم

درخت تصمیم توسط راس کوینلان[۲] (۱۹۸۶) معرفی شد. این الگوریتم به‌عنوان یکی از ابزارهای یادگیری نظارت‌شده مطرح‌شده و برای طبقه‌بندی و تحلیل رگرسیون استفاده می‌شود. درخت تصمیم به‌وسیله تقسیم داده‌ها به گروه‌های کوچک‌تر در هر مرحله، یک ساختار درختی ایجاد می‌کند که در آن هر گره یک ویژگی را نمایان می‌سازد و هر شاخه نشان‌دهنده یک مقدار ممکن برای آن ویژگی است. در این روش، هدف این است که در هر مرحله، تقسیم‌بندی بهینه‌ای انجام شود تا بهترین جداسازی میان دسته‌ها اتفاق بیفتد. معیارهایی مانند آنتروپی[۳] یا شاخص جینی[۴]

---

۱. Chen & Guestrin
۲. Quinlan
۳. Entropy
۴. Gini Index

برای انتخاب ویژگی بهینه استفاده می‌شوند. درنهایت، هر برگ در درخت نمایانگر یک تصمیم یا برچسب نهایی است.

۳-۳- الگوریتم جنگل تصادفی

مدل جنگل تصادفی توسط برایمن[۱] (۲۰۰۱) جهت توسعه روش درخت تصمیم عرضه شد. کارکرد جنگل تصادفی بر طبق روش تجمیعی[۲] می‌باشد که از چندین درخت تصمیم به‌عنوان مدل‌های پایه جهت پیش‌بینی استفاده می‌کند. در این الگوریتم هر درخت در جنگل تصادفی بر طبق یک زیرمجموعه کاملاً تصادفی از ورودی‌ها و خروجی‌ها تشکیل می‌شود به‌گونه‌ای که هر درخت به‌عنوان یک پیش‌بینی مستقل عمل می‌کند. جهت پیش‌بینی مدل نتایج تمامی درخت‌ها باهم تجمیع می‌شوند. برای مسائل طبقه‌بندی، پیش‌بینی نهایی بر اساس رأی‌گیری اکثریت درخت‌ها صورت می‌گیرد و برای مسائل رگرسیون میانگین پیش‌بینی‌های همه درخت‌ها به‌عنوان خروجی ارائه می‌شود. این الگوریتم به دلیل انعطاف در برابر نویز داده‌ها کارایی مطلوبی در مدل‌های پیچیده و غیرخطی دارد.

۳-۴- الگوریتم *KNN*

الگوریتم *KNN* توسط توماس کاور و پیتر هرت[۳] مطرح شد. این روش از نوع الگوریتم‌های نظارت‌شده است که برای طبقه‌بندی و همچنین تحلیل رگرسیون استفاده می‌شود. در الگوریتم *KNN*، فرآیند یادگیری مبتنی بر داده‌های جدید بوده و دسته‌بندی هر نمونه با توجه به ویژگی‌های مشترک آن با نزدیک‌ترین همسایگان صورت می‌گیرد. درواقع می‌توان گفت دسته‌بندی داده‌های جدید طبق شباهت به k نمونه نزدیک در مجموعه‌ی آموزشی(ورودی) انجام می‌شود. لازم به ذکر است مقدار k توسط کاربر مشخص می‌شود.

۳-۵- الگوریتم *AdaBoost*

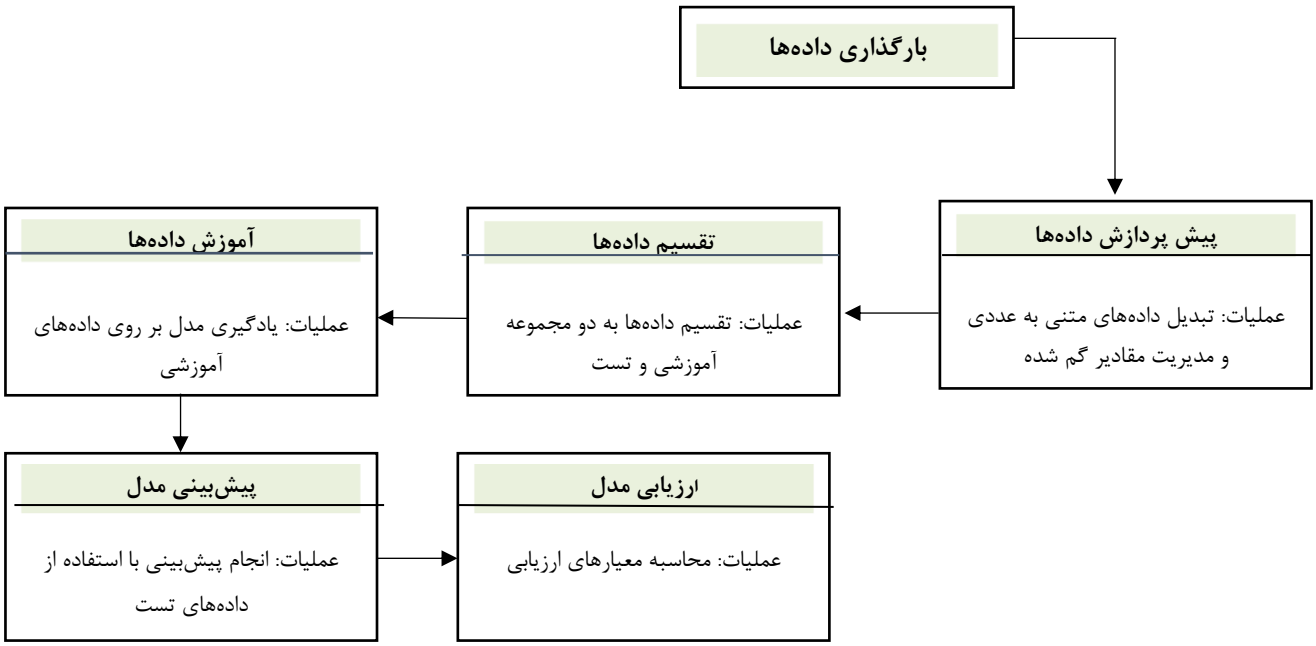
الگوریتم AdaBoost توسط فروند و رابرت شاپایر[٤] (۱۹۹۷) معرفی شد. این الگوریتم با هدف ترکیب چندین یادگیرنده‌ی ضعیف و تبدیل آن‌ها به یک یادگیرنده‌ی قوی طراحی شده است. سازوکار اصلی AdaBoost مبتنی بر تخصیص وزن به نمونه‌های آموزشی است. به‌طوری‌که

---

۱. Breiman
۲. Ensemble Learning
۳. Cover & Hart
٤. Freund & Schapire

در هر مرحله، وزن بیشتری به نمونه‌هایی که به‌اشتباه طبقه‌بندی شده‌اند داده می‌شود. سپس یک یادگیرنده‌ی ضعیف جدید بر اساس داده‌های وزن‌دهی‌شده آموزش می‌بیند. این فرآیند به‌صورت تکراری ادامه می‌یابد تا مجموعه‌ای از مدل‌ها ساخته شود که ترکیب آن‌ها منجر به افزایش دقت پیش‌بینی می‌گردد. از مزایای اصلی AdaBoost می‌توان به عملکرد قابل‌قبول آن حتی در مواجهه با داده‌های دارای نویز و ضعف ساختاری اشاره کرد.

در شکل ۱ مراحل تحلیل ریزش مشتریان با استفاده از مدل‌های یادگیری ماشین آمده است:



شکل ۱: مراحل تحلیل ریزش مشتریان با استفاده از مدل‌های یادگیری ماشین

۴- یافته‌های پژوهش

همان‌طور که اشاره شد، در این تحقیق اطلاعات مربوط به ریزش مشتریان در یک بانک خصوصی فعال در شهر مشهد مقدس مورد تحلیل و بررسی قرارگرفته است. جدول ۲ ویژگی‌های مشتریان انتخابی را نشان می‌دهد.

جدول ۲: ویژگی‌های مشتریان مورد مطالعه

شماره ویژگی	نام ویژگی	توضیحات	منبع	ورودی یادگیری ماشین	خروجی یادگیری ماشین
۱	سن مشتری	(بین ۱۸ تا ۳۵ سال)، (بین ٣٦ تا ٦٠ سال) و (بین ٦١ به بالا)	[٥]	✓	----
۲	جنسیت	زن یا مرد	[۱۲]	✓	----
۳	دسته‌بندی مشتری	طلایی، نقره‌ای و سربی	[۱۲]	✓	
٤	مدت‌زمان سابقه اعتباری (ماه)	----	[٦]	✓	
٥	سطح تحصیلات	دیپلم، کارشناسی، کارشناسی ارشد به بالا	[۱۲]	✓	
٦	مدت عضویت مشتری (ماه)	----	[٦] و [٩]	✓	
۷	تعداد شکایات در ماه	----	[٦] و [۱۳]	✓	

شماره ویژگی	نام ویژگی	توضیحات	منبع	ورودی یادگیری ماشین	خروجی یادگیری ماشین
۸	نوع حساب	حساب‌جاری، حساب پس‌انداز، حساب قرض‌الحسنه و حساب‌های ارزی	[۹]	✓	
۹	تعداد خدمات بانکی استفاده‌شده در ماه	----	[۵]	✓	
۱۰	ریزش یا عدم ریزش	----	----		✓

جدول ۳ و ٤ نیز تحلیل‌های آماری ویژگی‌های منتخب را نشان می‌دهد.

جدول ۳: ویژگی‌های کیفی مشتریان

تعداد فراوانی	درصد فراوانی	ویژگی
١٤٦٠٥	٤٠	سن (بین ۱۸ تا ۳۵ سال)
١٨٢٥٧	٥٠	سن (بین ٣٦ تا ٦٠ سال)
٣٦٥١	١٠	سن (بین ٦١ به بالا)
٢١١٧٧	٥٨	جنسیت (مرد)
١٥٣٣٦	٤٢	جنسیت (زن)
٧٣٠٣	٢٠	دسته مشتری (طلایی)
١٧٨٩١	٤٩	دسته مشتری (نقره‌ای)
١١٣١٩	٣١	دسته مشتری (سربی)
١٦٤٣١	٤٥	سطح تحصیلات (دیپلم)
١٢٧٧٩	٣٥	سطح تحصیلات (لیسانس)
٧٣٠٣	٢٠	سطح تحصیلات (ارشد و بالاتر)
١٤٢٤٠	٣٩	نوع حساب (پس‌انداز)
١٠٩٥٤	٣٠	نوع حساب (جاری)
٦٩٣٧	١٩	نوع حساب (قرض‌الحسنه)
٣٢٨٦	٩	نوع حساب (ارزی)
٢٢٦٣٨	٦٢	عدم ریزش

ویژگی	درصد فراوانی	تعداد فراوانی
ریزش	۳۸	۱۳۸۷۵

جدول ۴: ویژگی‌های کمی مشتریان

ویژگی	میانگین	واریانس	حداقل	حداکثر
مدت عضویت مشتری (ماه)	٦۰/۰٦	۳٤/۳۹	۱	۱۱۹
مدت‌زمان سابقه اعتباری (ماه)	٥۳/٦٥	۲۷/۱٤	٦	۱۰۰
تعداد خدمات بانکی استفاده‌شده در ماه	۲/۲۹۹	۲/۸۸٥	۱	۹
تعداد شکایات در ماه	٦/۹۹۹	٤/۳۳۰	۰	۱٤

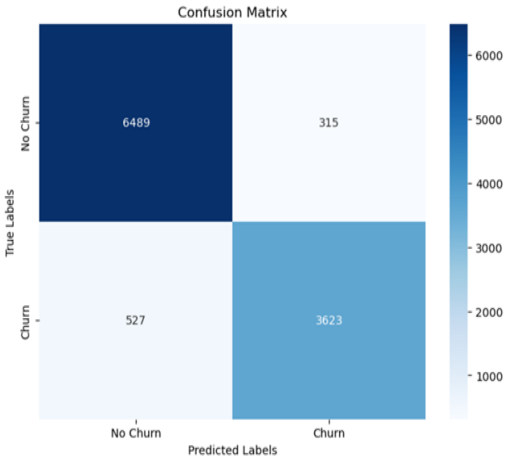
۴-۱- ارزیابی مدل‌های یادگیری ماشین

در این مرحله ماتریس‌های درهم‌ریختگی[۱] برای هر مدل یادگیری ماشین موردبررسی قرار می‌گیرد. ماتریس درهم‌ریختگی شامل جدولی است که جهت ارزیابی عملکرد یک مدل طبقه‌بندی با نمایش تعداد پیش‌بینی‌های درست و غلط مورداستفاده قرار می‌گیرد. این ماتریس متشکل از سطرها و ستون‌هایی است که نشان‌دهنده کلاس‌های واقعی و پیش‌بینی‌شده می‌باشند. ماتریس درهم‌ریختگی دارای ٤ رکن اصلی است: تعداد مثبت‌های واقعی[۲](*TP*): تعداد مواردی که مدل یادگیری ماشین به‌درستی کلاس مثبت را پیش‌بینی کرده است. تعداد منفی‌های واقعی[۳](*TN*): تعداد مواردی که مدل به‌طور صحیح کلاس منفی را شناسایی کرده است. مثبت‌های کاذب[٤](*FP*): مواردی که مدل به‌اشتباه کلاس مثبت را پیش‌بینی کرده است (خطای نوع اول). منفی‌های کاذب[٥](*FN*): مواردی که مدل به‌اشتباه کلاس منفی را پیش‌بینی کرده است (خطای نوع دوم). از این مقادیر جهت محاسبه معیارهایی همچون صحت مدل، دقت مدل، یادآوری و *F-۱ Score* استفاده می‌شود.

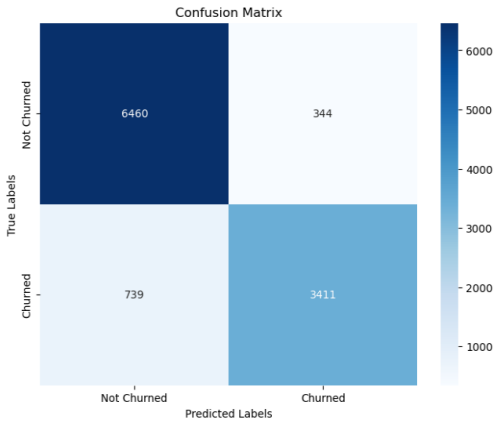
شکل ۲ و ۳ به‌عنوان نمونه ماتریس درهم‌ریختگی الگوریتم *XGBoost* و *Adaboost* رسم شده است. در این ماتریس محور افقی نشان‌دهنده برچسب‌های پیش‌بینی‌شده توسط مدل و محور

۱. confusion matrix
۲. True Positive
۳. True negative
٤. False Positive
٥. False negative

عمودی نشان‌دهنده برچسب‌های واقعی است. همچنین جدول ۵ خلاصه‌ای از اندازه‌گیری‌های عملکرد مدل‌های یادگیری ماشین را نشان می‌دهد. لازم به ذکر است کلیه تحلیل‌های لازم با نرم‌افزار پایتون ورژن (۳.۱۱) و کتابخانه‌های مربوطه انجام‌گرفته است.



شکل ۲: ماتریس درهم‌ریختگی الگوریتم *XGBoost*



شکل ۳: ماتریس درهم‌ریختگی الگوریتم *Adaboost*

جدول۵: معیارهای اندازه‌گیری عملکرد مدل‌های یادگیری ماشین

نام مدل	صحت عملکرد	دقت مدل	یادآوری	*F*۱-score
XGBoost	۰/۹۲	۰/۹۲	۰/۸۸	۰/۹۰
AdaBoost	۰/۹۰	۰/۹۱	۰/۸۵	۰/۸۸
جنگل تصادفی	۰/۸۷	۰/۸۷	۰/۸۵	۰/۸٦
درخت تصمیم	۰/۸۵	۰/۸۵	۰/۸٤	۰/۸٤
KNN	۰/۸۱	۰/۷۹	۰/۷۸	۰/۷۸

جهت بررسی اختلاف عملکرد مدل‌ها از نظر آماری، آزمون ویلکاکسون بر روی مقادیر *F*۱-*score* حاصل از ۵ بار اعتبارسنجی متقابل اجرا شد. بر اساس نتایج این آزمون و با توجه به (p <۰/۰۰۵) عملکرد مدل *XGBoost* به‌طور معناداری بهتر از سایر الگوریتم‌ها بوده است.

طبق نتایج جدول ۵، *XGBoost* بالاترین عملکرد و پس‌ازآن *AdaBoost* در رتبه دوم قرار دارد. از دلایلی که *XGBoost* بالاترین کارایی را نسبت به سایر الگوریتم‌های یادگیری ماشین و

همچنین عملکرد بهتری نسبت به سایر مدل‌های ذکرشده دارد می‌توان به توانایی این الگوریتم در مدیریت کارآمد و بهینه داده‌های بزرگ اشاره کرد که این الگوریتم را به گزینه‌ای مناسب جهت بررسی سناریوهای مختلفی تبدیل می‌کند. *XGBoost* از یک رویکرد مبتنی بر تقویت گرادیان استفاده می‌کند که سرعت و دقت بالا را در مسائل مختلف یادگیری ماشین (طبقه‌بندی و تحلیل رگرسیون) فراهم می‌آورد. این الگوریتم از چندین تکنیک بهینه‌سازی استفاده می‌کند تا عملکرد مدل‌ها را بهبود بخشد و زمان آموزش را تا حد امکان کاهش دهد. یکی از مهم‌ترین تکنیک‌ها در *XGBoost*، *Tree Pruning* است که به کمک روش قطع شاخه‌های نامربوط رشد بیش‌ازحد درخت‌ها را کنترل می‌کند. این امر باعث می‌شود که مدل نهایی دچار بیش برازش نشود. همچنین *XGBoost* از تکنیک نمونه‌برداری *Column Subsampling* بهره می‌برد که مشابه الگوریتم جنگل تصادفی برخی از ویژگی‌ها را به‌صورت تصادفی در هر مرحله انتخاب می‌کند. این تکنیک به کاهش همبستگی بین درخت‌ها و بهبود عملکرد عمومی مدل کمک می‌کند. در رتبه‌های سوم تا پنجم هم الگوریتم‌های جنگل تصادفی، درخت تصمیم و *KNN* قرار دارند که عملکرد صحت این سه الگوریتم بین ۸۱ تا ۸۷ درصد است. در جدول ٦ درصد اهمیت هر یک از ویژگی‌ها با استفاده از الگوریتم *XGBoost* در ریزش مشتریان مشخص‌شده است. همچنین به‌منظور بررسی پایداری اهمیت ویژگی‌ها، از روش بازنمونه‌گیری *Bootstrap* با ۱۰۰ تکرار استفاده شد. در این روش، با نمونه‌گیری تصادفی با جایگزینی از داده‌ها، مدل چندین بار آموزش داده شد و انحراف معیار اهمیت هر ویژگی به‌عنوان شاخصی از ناپایداری نسبی آن گزارش گردید.

جدول ۶: اهمیت ویژگی‌ها بر ریزش مشتری

ویژگی	درصد اهمیت ویژگی‌ها بر ریزش مشتری	انحراف معیار(%)
تعداد شکایات در ماه	۳۰	۲/۷۱۱
دسته‌بندی مشتری (طلایی)	۱۰	۰/۹٤۹
دسته‌بندی مشتری (نقره‌ای)	۷	۰/۷٥٥
دسته‌بندی مشتری (سربی)	۳	۰/۲٤٦
مدت عضویت مشتری (ماه)	۱۲	۱/۲۷
تعداد خدمات بانکی استفاده‌شده در ماه	۱۳	۱/۱۹٥
نوع حساب (جاری)	۲/۸	۰/۸۷۱
نوع حساب (پس‌انداز)	۲/۱	۰/۱۱٦

ویژگی	درصد اهمیت ویژگی‌ها بر ریزش مشتری	انحراف معیار(%)
نوع حساب (قرض‌الحسنه)	۱/٤	۰/۳۸۲
نوع حساب (ارزی)	۰/۷	۰/۰٦۷
سطح تحصیلات (دیپلم)	۰/۲	۰/۰۱۹
سطح تحصیلات (کارشناسی)	۰/۳۵	۰/۰۳۹
سطح تحصیلات (اشد و بالاتر)	۰/٤۵	۰/۰٤۵
جنسیت (مرد)	۲/٤	۰/۲۳۳
جنسیت (زن)	۱/٦	۰/۱٦
سن (۱۸-۳۵)	۱	۰/۱۰۳
سن (۳٦-٦۰)	۰/۷	۰/۰٦٦
سن (٦۰ به بالا)	۰/۳	۰/۰۲۸
مدت‌زمان سابقه اعتباری (ماه)	۱۱	۱/۰۰٦

۵- نتیجه‌گیری و پیشنهادها

در این مطالعه، پیش‌بینی ریزش مشتریان بانک با استفاده از الگوریتم‌های یادگیری ماشین مورد تجزیه‌وتحلیل قرار گرفت. هدف اصلی این پژوهش، شناسایی ویژگی‌های مؤثر بر ریزش مشتریان و ارزیابی عملکرد پنج الگوریتم یادگیری ماشین شامل *XGBoost*، جنگل تصادفی، درخت تصمیم، *AdaBoost* و *KNN* بود. نتایج نشان داد که الگوریتم *XGBoost* با صحت ۹۲ درصد بهترین عملکرد را در پیش‌بینی ریزش مشتریان دارد. این یافته‌ها تأکید می‌کنند که استفاده از الگوریتم‌های پیشرفته یادگیری ماشین می‌تواند به بانک‌ها در شناسایی و حفظ مشتریانی که احتمال ریزش بالایی دارند کمک کند و به بهبود تصمیم‌گیری استراتژیک و افزایش سودآوری آن‌ها منجر شود. الگوریتم *XGBoost* به دلیل استفاده از تقویت گرادیان و تکنیک‌های بهینه‌سازی مانند *Tree Pruning* و نمونه‌برداری تصادفی، توانست داده‌های پیچیده و چندبعدی را با دقت بالایی پردازش کند، بنابراین، استفاده از این الگوریتم برای مدیریت داده‌های بزرگ و متنوع در محیط‌های بانکی توصیه می‌شود. در مقایسه، الگوریتم‌های دیگر مانند *AdaBoost* با صحت ۹۰ درصد عملکرد نسبتاً خوبی داشتند، اما به دلیل پیچیدگی محاسباتی و زمان بیشتری که برای آموزش نیاز دارند، ممکن است برای پیاده‌سازی‌های عملی در مقیاس وسیع مناسب نباشند. الگوریتم‌های جنگل تصادفی، درخت تصمیم و *KNN* نیز

دقت‌های متفاوتی از ۸۱ تا ۸۷ درصد داشتند که نشان می‌دهد هر الگوریتم بر اساس ویژگی‌های داده‌ها و ساختار مدل می‌تواند نتایج متفاوتی ارائه دهد؛ بنابراین، انتخاب الگوریتم مناسب باید با توجه به ویژگی‌های خاص داده‌ها و نیازهای دقیق سازمان انجام شود.

بر اساس یافته‌های این پژوهش و با توجه به الگوریتم *XGBoost* که عملکرد بهتری نسبت به سایر روش‌ها دارد، تعداد شکایات مشتریان در ماه با اثرگذاری ۳۰ درصد، مهم‌ترین عامل در ریزش مشتریان است؛ بنابراین، بهبود فرایندهای مدیریت شکایات و افزایش سرعت و کیفیت پاسخ‌دهی به مشکلات مشتریان اهمیت ویژه‌ای در صنعت بانکداری دارد. دسته‌بندی مشتریان (طلایی، نقره‌ای و سربی) بااهمیت مجموع ۲۰ درصد، حاکی از آن است که مشتریان طلایی با سهم ۱۰ درصد، بیشترین حساسیت را نسبت به نحوه و کیفیت ارائه خدمات دارند. این پدیده از چند جنبه قابل تبیین است. نخست، مشتریان طلایی به دلیل دارایی، گردش مالی بالا و سابقه اعتباری قوی، همواره در کانون توجه بانک‌های رقیب قرار دارند و اغلب پیشنهادهای جذاب‌تری همچون نرخ سود ترجیحی، خدمات اختصاصی یا تسهیلات آسان‌تر دریافت می‌کنند. دوم، این گروه از مشتریان معمولاً به سطحی از خدمات ممتاز و شخصی‌سازی‌شده عادت دارند؛ ازاین‌رو، هرگونه افت در کیفیت خدمات، تأخیر در پاسخ‌گویی یا نادیده‌گرفتن انتظارات آن‌ها می‌تواند منجر به نارضایتی سریع و ترک بانک شود.

از طرف دیگر، حساب‌های جاری با تأثیر ۲/۸ درصدی، نقش پررنگ‌تری در بین انواع حساب در ریزش مشتریان دارند که نشان می‌دهد ارائه مشوق‌ها و تسهیلات خاص برای دارندگان این حساب‌ها می‌تواند در نگهداشت آن‌ها مؤثر باشد. همچنین مردان نسبت به زنان بیشتر در معرض ریزش قرار دارند که این تفاوت بیانگر ضرورت طراحی استراتژی‌های بازاریابی و تعامل متناسب با جنسیت است. در میان گروه‌های سنی، مشتریان ۱۸ تا ۳۵ سال با سهم ۱ درصد، به دلیل انتظارات بالا، استفاده از خدمات الکترونیک و وفاداری پایین‌تر، نیاز به توجه ویژه دارند. علاوه بر این، مدت عضویت و سابقه اعتباری نشان می‌دهد مشتریان باسابقه، دارایی ارزشمندی برای بانک هستند و لازم است با مشوق‌های مستمر و برنامه‌های وفاداری از آن‌ها نگهداری شود. درنهایت، سطح تحصیلات با سهم ۱ درصد، کمترین تأثیر را در ریزش دارد، اما مشتریان دارای تحصیلات عالی، معمولاً آگاه‌تر و حساس‌تر هستند و در برابر خدمات غیرحرفه‌ای، واکنش سریع‌تری از خود نشان می‌دهند.

نتایج این تحقیق نشان می‌دهد که به‌کارگیری الگوریتم‌های پیشرفته یادگیری ماشین، به‌ویژه مدل‌هایی مانند *XGBoost*، می‌تواند به بانک‌ها در شناسایی دقیق مشتریان در معرض ریزش کمک کرده و زمینه‌ساز تصمیم‌گیری‌های استراتژیک در سطح مدیریت شود. این مدل‌ها با تحلیل الگوهای رفتاری مشتریان، قابلیت به‌کارگیری در سیستم‌های مدیریت روابط مشتری[۱] را دارند و می‌توانند بانک‌ها را در طراحی کمپین‌های بازاریابی هدفمند، بهبود خدمات و ارائه پیشنهادهای ویژه یاری رسانند. افزون بر این، نتایج مدل می‌تواند مبنایی برای تدوین سیاست‌های عملیاتی نظیر ارائه تسهیلات با نرخ ترجیحی به مشتریان طلایی، افزایش سرعت پاسخ‌گویی به شکایات دارندگان حساب جاری و ارسال پیام‌های شخصی‌سازی‌شده به مشتریان خاکستری فراهم آورد.

افزون بر این، یافته‌های این پژوهش با نتایج مطالعات اخیر در حوزه مالی همچون [۱۹]، [۲۱] هم‌راستا است. در مطالعات مذکور نیز الگوریتم‌های تقویتی مانند *XGBoost* به‌عنوان مدل‌هایی با دقت بالا در پیش‌بینی ریزش مشتریان معرفی‌شده‌اند و بر نقش کلیدی متغیرهایی همچون تعداد شکایات و نوع حساب تأکید شده است. این هم‌سویی، اعتبار نتایج به‌دست‌آمده در پژوهش حاضر را تقویت کرده و شواهد تجربی معتبری برای اثربخشی مدل‌های پیشنهادی فراهم می‌سازد. مزیت نسبی پژوهش حاضر در مقایسه با مطالعات پیشین، در به‌کارگیری مجموعه‌ای جامع‌تر از الگوریتم‌های یادگیری ماشین و تحلیل مقایسه‌ای ساخت‌یافته آن‌ها نهفته است؛ این رویکرد، زمینه‌ساز ارزیابی دقیق‌تر عملکرد مدل‌ها در بستر داده‌های واقعی و متغیرهای متنوع بانکی شده است.

از چالش‌های اصلی این تحقیق می‌توان به کیفیت و کمیت داده‌های موجود اشاره کرد. درحالی‌که داده‌ها شامل اطلاعات دموگرافیکی و اقتصادی مشتریان بود، برخی از ویژگی‌های مهم رفتاری و تعاملی مشتریان، مانند اطلاعات مربوط به تعاملات آنلاین یا داده‌های شبکه‌های اجتماعی در دسترس نبود. تحقیقات آتی می‌تواند با در نظر گرفتن این ویژگی‌ها به نتایج دقیق‌تری برسد. این نوع داده‌ها می‌توانند به بهبود دقت مدل‌ها کمک کنند زیرا اطلاعات بیشتری درباره رفتار و ترجیحات مشتریان فراهم می‌کنند. همچنین پیشنهاد می‌گردد در تحقیقات آتی شاخص رضایت مشتریان به عنوان یکی از معیارهای تاثیرگذار در مدل‌های پیش‌بینی ریزش

---

[۱] . Customer Relationship Management

مشتری در نظر گرفته شود تا نتایج دقیق‌تری حاصل شود. همچنین استفاده از الگوریتم‌های *LightGBM* و *CatBoost* به‌ویژه در مواجهه با داده‌های پیچیده یا دارای ویژگی‌های غیرعددی توصیه می‌شود.

۶- منابع

[۱] Zorić, B., *Predicting customer churn in banking industry using neural networks.* Interdisciplinary Description of Complex Systems: INDECS, ۲۰۱٦. ۱٤(۲): p. ۱۱٦-۱۲٤.https://doi.org/۱۰,۷۹۰٦/indecs.۱٤,۲,۱.

[۲] Najafi, A. and Akhondzadeh Noughabi, E., *Pattern Mining of customer dynamics through different customer value states by using sequence pattern mining and big data analytics.* Modern Research in Decision Making, ۲۰۲٤. ۹(٤): p. ٦۸-۹۳.(in persian)

[۳] Liu, X., et al., *Customer churn prediction model based on hybrid neural networks.* Scientific Reports, ۲۰۲٤. ۱٤(۱): p. ۳۰۷۰۷.https://doi.org/۱۰,۱۰۳۸/s٤۱٥۹۸-۰۲٤-۷۹٦۰۳-۹

[٤] Keramati, A., Ghaneei, H., and Mirmohammadi, S.M., *Developing a prediction model for customer churn from electronic banking services using data mining.* Financial Innovation, ۲۰۱٦. ۲: p. ۱-۱۳.https://doi.org/۱۰,۱۱۸٦/s٤۰۸٥٤-۰۱٦-۰۰۲۹-٦.

[٥] Guliyev, H. and Yerdelen Tatoğlu, F., *Customer churn analysis in banking sector: Evidence from explainable machine learning model.* Journal of Applied Microeconometrics, ۲۰۲۱. ۱(۲).

[٦] Tran, H., Le, N., and Nguyen, V.-H., *CUSTOMER CHURN PREDICTION IN THE BANKING SECTOR USING MACHINE LEARNING-BASED CLASSIFICATION MODELS.* Interdisciplinary Journal of Information, Knowledge & Management, ۲۰۲۳. ۱۸.https://doi.org/۱۰,۲۸۹٤٥/٥۰۸٦.

[۷] Alizadeh, M., et al., *Development of a customer churn model for banking industry based on hard and soft data fusion.* IEEE Access, ۲۰۲۳. ۱۱: p. ۲۹۷٥۹-۲۹۷٦۸.https://doi.org/۱۰,۱۱۰۹/ACCESS.۲۰۲۳,۳۲٥۷۳٥۲.

[۸] Chang, V., et al., *Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study.* Annals of Operations Research, ۲۰۲٤: p. ۱-۲٥.https://doi.org/۱۰,۱۰۰۷/s۱۰٤۷۹-۰۲٤-۰٦۱۳٤-x

[۹] Singh, P.P., et al., *Investigating customer churn in banking: A machine learning approach and visualization app for data science and management.* Data Science and Management, ۲۰۲٤. ۷(۱): p. ۷-۱٦.https://doi.org/۱۰,۱۰۱٦/j.dsm.۲۰۲۳,۰۹,۰۰۲

[۱۰] Jafari, M.J., Tarokh, M.J., and Soleimani, P., *A data-driven Agent-based model and framework for Churn prediction in Telecommunication Industry.* Modern Research in Decision Making, ۲۰۲٤. ۹(۲): p. ١٦٤-١٩٠.(in persian)

[۱۱] Yousefi Ghaleh Roudkhani, M.A., Tehrani, R., and Mirlouhi, S.M., *Investigating the Impact of Financial performance metrics on Financial Stability of Banks in the Financial Crisis.* Management Research in Iran, ۲۰۲۱. ۲۵(۲): p. ۱-۲۱.(in persian)

[۱۲] de Lima Lemos, R.A., Silva, T.C., and Tabak, B.M., *Propension to customer churn in a financial institution: A machine learning approach.* Neural Computing and Applications, ۲۰۲۲. ۳٤(١٤): p. ١١٧٥١-١١٧٦٨.https://doi.org/١٠,١٠٠٧/s٠٠٥٢١-٠٢٢-٠٧٠٦٧-x.

[۱۳] Matuszelański, K. and Kopczewska, K., *Customer churn in retail e-commerce business: Spatial and machine learning approach.* Journal of Theoretical and Applied Electronic Commerce Research, ۲۰۲۲. ۱۷(۱): p. ١٦٥-١٩٨.https://doi.org/١٠,٣٣٩٠/jtaer١٧٠١٠٠٠٩.

[۱٤] zarei, G., Mohammad khani, R., and fathi, h., *Investigating and identifying the consequences of using artificial intelligence in marketing.* Management Research in Iran, ۲۰۲٤. ۲۸(۲): p. ۱-۳۱.(in persian)

[۱٥] Keramati, A., Ghaneei, H., and Mirmohammadi, S.M., *Investigating factors affecting customer churn in electronic banking and developing solutions for retention.* International Journal of Electronic Banking, ۲۰۲۰. ۲(۳): p. ١٨٥-٢٠٤.https://doi.org/١٠,١٥٠٤/IJEBANK.٢٠٢٠,١١٤٢٧.

[١٦] Haddadi, S.J., et al., *Customer churn prediction in imbalanced datasets with resampling methods: A comparative study.* Expert Systems with Applications, ۲۰۲٤. ۲٤٦: p. ١٢٣٠٨٦.https://doi.org/١٠,١٠١٦/j.eswa.٢٠٢٣,١٢٣٠٨٦.

[۱۷] Lalwani, P., et al., *Customer churn prediction system: a machine learning approach.* Computing, ۲۰۲۲. ۱۰٤(۲): p. ٢٧١-٢٩٤.https://doi.org/١٠,١٠٠٧/S٠٠٦٠٧-٠٢١-٠٠٩٠٨-Y.

[۱۸] Mahesh, B.S., et al., Predicting Customer Churn in Subscription-Based Enterprises Using Machine Learning, in Book·Predicting Customer Churn in Subscription-Based Enterprises Using Machine Learning.٢٠١٧,Springer Place·Published.p.٣٦٥-٣٧٧.https://doi.org/١٠,١٠٠٧/٩٧٨-٩٨١-٩٩-٨٤٣٨-١_٢٦

[۱۹] El Khair Ghoujdam, M., et al., *Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques.* Cogent Economics & Finance, ۲۰۲٤. ۱۲(۱): p. ٢٤١٤٩٢٦.https://doi.org/١٠,١٠٨٠/٢٣٣٢٢٠٣٩,٢٠٢٤,٢٤١٤٩٢٦.

[۲۰] Shobana, J., et al., *E-commerce customer churn prevention using machine learning-based business intelligence strategy.* Measurement: Sensors, ۲۰۲۳. ۲۷: p. ١٠٠٧٢٨.https://doi.org/١٠,١٠١٦/j.measen.٢٠٢٣,١٠٠٧٢٨.

[۲۱] Gurung, N., et al., *AI-Based Customer Churn Prediction Model for Business Markets in the USA: Exploring the Use of AI and Machine Learning*

Technologies in Preventing Customer Churn. Journal of Computer Science and Technology Studies, ۲۰۲٤. ٦(۲): p. ۱۹-۲۹.https://doi.org/۱۰,۳۲۹۹٦/jcsts.

[۲۲] Dursun-Cengizci, A. and Caber, M., *Using machine learning methods to predict future churners: an analysis of repeat hotel customers.* International Journal of Contemporary Hospitality Management, ۲۰۲٤.https://doi.org/۱۰,۱۱۰۸/IJCHM-۰٦-۲۰۲۳-۰۸٤٤.

[۲۳] Çallı, L. and Kasım, S., *Using Machine Learning Algorithms to Analyze Customer Churn in the Software as a Service (SaaS) Industry.* Academic Platform Journal of Engineering and Smart Systems, ۲۰۲۲. ۱۰(۳): p. ۱۱۵-۱۲۳.https://doi.org/۱۰,۲۱٥٤۱/apjess.۱۱۳۹۸٦۲.

[۲٤] Suh, Y., *Machine learning based customer churn prediction in home appliance rental business.* Journal of big Data, ۲۰۲۳. ۱۰(۱): p. ٤۱.https://doi.org/۱۰,۱۱۸٦/s٤۰٥۳۷-۰۲۳-۰۰۷۲۱-۸.

[۲٥] He, Y., Xiong, Y., and Tsai, Y. *Machine learning based approaches to predict customer churn for an insurance company*. in *2020 Systems and Information Engineering Design Symposium (SIEDS)*. ۲۰۲۰. IEEE.https://doi.org/۱۰,۱۱۰۹/SIEDS٤۹۳۳۹,۲۰۲۰,۹۱۰٦٦۹۱

[۲٦] Allen, K., et al., *Machine literature searching VIII. Operational criteria for designing information retrieval systems.* American Documentation (pre-۱۹۸٦), ۱۹٥٥. ٦(۲): p. ۹۳.https://doi.org/۱۰,۱۰۰۲/asi.٥۰۹۰۰٦۰۲۰۹.

[۲۷] Goutte, C. and Gaussier, E. *A probabilistic interpretation of precision, recall and F-score, with implication for evaluation*. in *European conference on information retrieval*. ۲۰۰٥. Springer.https://doi.org/۱۰,۱۰۰۷/۹۷۸-۳-٥٤۰-۳۱۸٦٥-۱_۲٥

[۲۸] Chen, T. and Guestrin, C. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ۲۰۱٦.https://doi.org/۱۰,۱۱٤٥/۲۹۳۹٦۷۲,۲۹۳۹۷۸٥

[۲۹] Quinlan, J.R., *Induction of decision trees.* Machine learning, ۱۹۸٦. ۱: p. ۸۱-۱۰٦.https://doi.org/۱۰,۱۰۰۷/BF۰۰۱۱٦۲٥۱.

[۳۰] Breiman, L., *Random forests Mach Learn 45 (1): 5–32*. ۲۰۰۱, ed.https://doi.org/۱۰,۱۰۲۳/A:۱۰۱۰۹۳۳٤۰٤۳۲٤

[۳۱] Cover, T. and Hart, P., *Nearest neighbor pattern classification.* IEEE transactions on information theory, ۱۹٦۷. ۱۳(۱): p. ۲۱-۲۷.https://doi.org/۱۰,۱۱۰۹/TIT.۱۹٦۷,۱۰٥۳۹٦٤.

[۳۲] Freund, Y. and Schapire, R.E., *A decision-theoretic generalization of on-line learning and an application to boosting.* Journal of computer and system sciences, ۱۹۹۷. ٥٥(۱): p. ۱۱۹-۱۳۹.https://doi.org/۱۰,۱۰۰٦/jmsp.۱۹۹۷,۱٥۰٤.