نشریه علمی مطالعات مدیریت کسب و کار هوشمند- سال هفتم ـ شماره ۲۸ـ تابستان ۱۳۹۸
صفحات ۱۱۰- ۶۷

پیش‌بینی رویگردانی جزئی مشتریان بانک‌ها با استفاده از مدل زنجیره وضعیت

محسن عسگری *
محمدرضا تقوا **
محمدتقی تقوی‌فرد ***

چکیده

بانک‌ها در فضای رقابتی شدید تلاش می‌کنند تا به منابع مالی بیشتری دست پیدا کنند. با توجه به بالاتر بودن هزینه‌های جذب مشتری جدید نسبت به نگهداری مشتریان موجود، عمده تلاش بانک‌ها روی حفظ سپرده‌های موجود مشتریان در بانک متمرکز است. لذا پیش‌بینی رویگردانی مشتریان پیش از وقوع برای بانک‌ها از اهمیت ویژه‌ای برخوردار است. تقریباً در تمامی تحقیقات مرتبط در بانک‌ها مشتریان به دو دسته رویگردان و غیر رویگردان با یک تعریف ثابت از رویگردانی تقسیم شده‌اند؛ اما در شرایط بانکداری ایران نمی‌توان از یک تعریف ثابت برای رویگردانی استفاده نمود؛ بنابراین لازم است که رویگردانی را به‌صورت دینامیک و در قالب وضعیت‌های مختلف تعریف کنیم. برای این منظور در این تحقیق مفهوم زنجیره وضعیت معرفی می‌شود که تغییرات وضعیت رویگردانی جزئی مشتریان طی زمان را مشخص می‌کند. با به‌کارگیری این زنجیره‌ها و استفاده ترکیبی از تکنیک‌های خوشه‌بندی سلسله‌مراتبی و همچنین ماشین‌های بردار پشتیبان، مدلی برای پیش‌بینی رویگردانی جزئی مشتریان بانک‌ها ساخته شد. برای ساختن نمونه عملی و ارزیابی دقت پیش‌بینی، پنج سال داده‌های واقعی مشتریان یک بانک اروپایی و همچنین سه سال داده‌های مشتریان سه بانک ایرانی مورد استفاده قرار گرفتند. نتایج حاکی از دقت بالای پیش‌بینی در مدل‌های ساخته‌شده روی هر چهار بانک به‌خصوص با افزایش طول زنجیره‌های وضعیت در داده‌های آزمون است.

کلیدواژگان: رویگردانی جزئی، مدل زنجیره وضعیت، مشتریان بانک، خوشه‌بندی سلسله‌مراتبی، ماشین بردار پشتیبان

---

* دانشجوی دکتری، مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران.

** عضو هیأت علمی، گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران. (نویسنده مسئول)؛ Taghva@atu.ac.ir

*** عضو هیأت علمی، گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران.

تاریخ دریافت:۱۳۹۷/۰۲/۲۷ تاریخ پذیرش:۱۳۹۷/۰۷/۱۱

مقدمه

در فضای رقابتی امروز یکی از مهم‌ترین چالش‌های سازمان‌ها در مدیریت ارتباط با مشتریان[1] حفظ و نگهداری مشتریان موجود است. بین محققین در این خصوص اتفاق‌نظر نسبی وجود دارد که میانگین هزینه‌هایی که برای جذب مشتری جدید به نسبت نگهداری مشتری موجود صرف می‌شود به‌مراتب بیشتر است (ریبه و همکاران[2]، ۲۰۱۴). لذا مفهوم رویگردانی مشتری[3] از اهمیت ویژه‌ای برخوردار است.

صنعت بانکداری در ایران پس از ظهور بانک‌های خصوصی به‌شدت متحول شد. تعداد زیاد بانک‌های خصوصی با چابکی زیاد و ارائه سریع محصولات و خدمات جدید، فشار رقابتی بسیار زیادی را ایجاد کرده است. رقابت شدید در این بازار موجب شده است که بانک‌ها به دنبال جذب مشتریان و به‌تبع آن جذب منابع بیشتر باشند. با ورود بانک‌های خصوصی شاخص رقابت در صنعت بانکداری از ۰/۴۰ در فاصله سال‌های ۱۳۷۶ تا ۱۳۸۲ به ۰/۸۲ در فاصله سال‌های ۱۳۸۳ تا ۱۳۸۹ رسیده است (نجارزاده و همکاران[4]، ۲۰۱۳).

بدیهی است که در این فضای رقابتی و با توجه به هزینه تغییر پایین[5]، مشتریان بانک‌ها به‌راحتی بانک موردنظر خود را به یک بانک دیگر با خدمات و محصولات بهتر تغییر می‌دهند. ریزش مشتریان از یک بانک به معنی کاهش منابع مالی بوده و ریسک بالایی برای آن بانک به شمار می‌آید. لذا بانک‌ها بایستی به‌شدت تلاش کنند که مشتریان ارزشمند فعلی خود را حفظ کنند. محاسبه احتمال رویگردانی مشتریان پیش از وقوع آن، این فرصت را فراهم می‌کند تا سازمان بتواند با ایجاد مشوق‌هایی مشتری را برای ادامه فعالیت قانع نماید (وربکه و همکاران[6]، ۲۰۱۲). محاسبه احتمال رویگردانی مشتریان بانک می‌تواند مبنای محاسبه ریسک منابع بانک قرار گیرد که ریسک موردنظر یکی از مهم‌ترین پارامترهای تأثیرگذار در ارزیابی وضعیت مالی بانک و همچنین تصمیمات مالی مختلف است. درواقع اینکه یک بانک بداند روی چه مقدار

1. Customer Relationship Management
2. Riebe et al.
3. Customer Churn
4. Najarzadeh et al.
5. Switching Cost
6. Verbeke et al.

از منابع حال حاضر خود و تا چه زمانی می‌تواند با اطمینان حساب باز کند از اهمیت بسیاری برخوردار است. لذا محاسبه شاخص ذکرشده غیرازاینکه در مدیریت ارتباط با مشتریان مؤثر است در محاسبه ریسک منابع بانک نیز کاربرد دارد.

با توجه به توسعه سیستم‌های بانکداری متمرکز[1] بانک‌ها بیشترین اطلاعات را در مورد رفتار مشتریان می‌توانند از طریق تحلیل داده‌های تراکنش‌های مشتریان به دست آورند. با استفاده از این داده‌ها و با کمک الگوریتم‌ها و تکنیک‌های داده‌کاوی می‌توان مدل‌هایی برای پیش‌بینی رفتار رویگردانی مشتریان بانک‌ها ساخت.

البته رفتار رویگردانی مشتری می‌تواند بسیار پیچیده‌تر از قطع ارتباط کامل او با سازمان باشد. مثلاً یک مشتری بانک که رویگردانی نموده ممکن است حساب بانکی خود را نبسته و حتی فعالیت مالی هم داشته باشد و از برخی از خدمات بانک نیز استفاده کند اما در واقع بیشتر مقدار سپرده خود را از بانک خارج کرده باشد. برای این منظور نوعی از رویگردانی برحسب میزان فعالیت مشتری تحت عنوان رویگردانی جزئی[2] تعریف می‌شود. لذا اینکه بتوانیم رفتار مشتری را در بازه‌های زمانی مشخص (مثلاً ماهانه) از منظر رویگردانی شناسایی کنیم بسیار اهمیت دارد. همان‌طور که در بخش مرور ادبیات نیز اشاره کرده‌ایم، در بیشتر مقالات حوزه رویگردانی در بانکداری، مشتریان به دو دسته رویگردان و غیر رویگردان[3] تقسیم شده‌اند و از تکنیک‌های طبقه‌بندی[4] برای ساختن مدل‌های پیش‌بینی استفاده می‌شود؛ اما ارائه تعریفی ثابت برای رویگردانی به‌خصوص در صنعت بانکداری ایران بسیار ساده‌انگارانه است. در واقع خود اینکه مشتری در کدام وضعیت از رویگردانی جزئی قرار دارد موضوعی است که بایستی با محاسبه پارامترهای مؤثر در رویگردانی و تکنیک‌های داده‌کاوی توصیفی[5] مانند خوشه‌بندی شناسایی شود. همچنین در ساختن مدل پیش‌بینی بایستی به تغییرات وضعیت رویگردانی جزئی مشتری طی زمان نیز توجه نمود. در واقع تنها یک وضعیت از مشتری نمی‌تواند مبنای پیش‌بینی

1. CORE Banking
2. Partial Churn
3. Churner and non-churner
4. Classification
5. Descriptive

قرار بگیرد بلکه بایستی تعدادی از وضعیت‌های مختلف رویگردانی پشت سر هم که مشتری در گذشته داشته است مبنای ساختن مدل پیش‌بینی قرار بگیرد.

جنبه‌های نوآوری در این مقاله به‌اختصار شامل: استخراج پارامترهای رویگردانی جزئی ویژه مشتریان بانک‌ها، شناسایی وضعیت رویگردانی جزئی مشتریان بانک‌ها، معرفی مفهوم مدل زنجیره وضعیت رویگردانی جزئی[1] مشتریان و نهایتاً ساختن مدل پیش‌بینی بر اساس این مفهوم و بالا بردن دقت آن است.

در این مقاله ابتدا مروری بر مفهوم رویگردانی مشتریان و تحقیقات مرتبط با آن در بانکداری انجام می‌شود. سپس روش تحقیق، داده مورد استفاده و روش ساختن مدل پیش‌بینی بحث می‌شوند. در بخش بعدی نتایج حاصله ارائه شده و در خصوص آن‌ها بحث می‌شود. درنهایت جمع‌بندی از نکات مهم ارائه شده در مقاله خواهد آمد.

مرور ادبیات

در این بخش با تکیه‌بر مرور تحقیقات پیشین ابتدا به ارائه تعریفی مختصر از رویگردانی مشتریان و اهمیت و انواع آن می‌پردازیم. سپس روش‌های تحلیل رویگردانی مشتریان بحث می‌شود. در ادامه مقالات مرتبط با پیش‌بینی رویگردانی در صنایع مختلف مرور می‌شوند. پس‌ازآن بر اساس اهداف تحقیق مقالات مرتبط با حوزه تحقیق در صنعت بانکداری مرور می‌شوند که در این مرور ابتدا انواع تکنیک‌های داده‌کاوی مورد استفاده در پیش‌بینی رویگردانی صنعت بانکداری، سپس موضوع رویگردانی جزئی بحث می‌شوند و در ادامه خلاصه برخی تحقیقات مهم در این موضوع آورده می‌شود. درنهایت هم جمع‌بندی از مرور ادبیات انجام می‌شود که در آن مختصراً به نکات برداشت‌شده از ادبیات در راستای اهداف مقاله اشاره می‌شود.

رویگردانی، اهمیت و انواع آن

رویگردانی را می‌توان به‌صورت عدم تمایل مشتری به تداوم استفاده از محصولات یا خدمات

1. Partial churn state model

سازمان تعریف نمود (ینگ و همکاران[1]، ۲۰۱۸).

پیش‌بینی رویگردانی مشتری یکی از مهم‌ترین موضوعات در مدیریت ارتباط با مشتریان است. امروزه به دلیل رقابت شدید و بهبود دسترسی مشتریان به اطلاعات امکان جایگزین شدن رقبا با یکدیگر بسیار بیشتر شده است؛ بنابراین شرکت‌ها بایستی به پیش‌بینی رویگردانی مشتریان بپردازند تا از این طریق از کاهش مشتریان خود جلوگیری کنند (ژو و همکاران[2]، ۲۰۱۸).

در بازارهای رقابتی رویگردانی مشتری وقتی رخ می‌دهد که او از خدمات و محصولات سازمان در مقایسه با رقبا ناراضی شده و تصمیم به ترک استفاده از خدمات و محصولات سازمان می‌گیرد. هدف مدل‌های رویگردانی شناسایی به‌موقع مشتریانی است که احتمال خروج آن‌ها زیاد شده است (وفیادیس و همکاران[3]، ۲۰۱۵).

رویگردانی را می‌توان ازنظر نوع ارتباط مشتری با سازمان به دو نوع رویگردانی کامل و رویگردانی جزئی[4] تقسیم نمود. رفتار رویگردانی جزئی به آن دسته از مشتریانی نسبت داده می‌شود که لزوماً به‌طور کامل ارتباط خود را با سازمان قطع نکرده و هنوز از بخشی از خدمات و محصولاتش استفاده می‌کنند (میگوئیس و همکاران[5]، ۲۰۱۲).

روش‌های تحلیل رویگردانی

تحقیقات در زمینه رویگردانی از منظر روش می‌توان به دو دسته تقسیم کرد. گروه اول با استفاده از ابزارهایی مانند پرسشنامه به تحلیل رفتار و ذهنیت مشتریان می‌پردازند. از آن جمله می‌توان به تحقیق کیم و یون[6] در خصوص سنجش اثر رضایت مشتریان بر رویگردانی در صنعت مخابرات کشور کره اشاره نمود (کیم و یون، ۲۰۰۴). تحقیق ماهاجان و همکارانش[7] در خصوص عوامل مؤثر بر رویگردانی مشتری در صنعت مخابرات نیز از این نوع هستند

1. Yang et al.
2. Zhu et al.
3. Vafeiadis et al.
4. Partial Churn
5. Migueis et al.
6. Kim & Yoon
7. Mahajan et al.

(ماهاجان و همکاران، ۲۰۱۷). این دسته از تحقیقات هم نیازمند صرف زمان و هزینه زیاد برای دستیابی به نتیجه هستند و هم قابلیت تعمیم‌پذیری پایینی دارند (ینگ و همکاران، ۲۰۱۸).

گروه دیگری از تحقیقات به ساختن مدل‌هایی بر مبنای داده‌های موجود از تراکنش‌های مشتریان و تحلیل آن‌ها با روش‌های داده‌کاوی می‌پردازند. این گروه از تحقیقات داده محور بوده و بر مبنای تحلیل داده‌های موجود از تراکنش‌های مشتریان با سازمان استوار هستند. با توسعه سیستم‌های اطلاعاتی پیشرفته الکترونیک معمولاً سازمان‌ها حجم بالایی از داده‌های مربوط به رفتار مشتریان را به‌صورت ساده و ارزان در اختیار دارند که می‌توانند با روش‌های مختلف داده‌کاوی به تحلیل آن‌ها بپردازند. عمده تحقیقات اخیر پیش‌بینی رویگردانی مشتریان از این گروه هستند (ینگ و همکاران، ۲۰۱۸).

پیش‌بینی رویگردانی در صنایع مختلف

تحقیقات در زمینه پیش‌بینی رویگردانی در صنایع مختلفی انجام شده که از آن جمله می‌توان به صنعت مخابرات (بابو و آنانتانارایانان[1]، ۲۰۱۸؛ کوسیمنت و همکاران[2]، ۲۰۱۷؛ لو و همکاران[3]، ۲۰۱۴؛ کیم و همکاران، ۲۰۱۴؛ پراساد و مدهوی[4]، ۲۰۱۲ و داهیا و بهاتیا[5]، ۲۰۱۵)، بانکداری (علی و اونترک[6]، ۲۰۱۴؛ ژو و همکاران، ۲۰۱۴؛ کائور و همکاران[7]، ۲۰۱۳ و فرقاد و همکاران[8]، ۲۰۱۴)، بیمه (گونتر[9]، ۲۰۱۴) و سرگرمی (کوسیمنت و همکاران، ۲۰۱۳) اشاره نمود.

نقطه مشترک تمامی این صنایع این است که حجم بالایی از داده‌های تراکنش‌های مشتریان را به‌صورت الکترونیک نگهداری می‌کنند. در واقع به جهت اینکه فرآیندهای ارائه خدمات به

1. Babu & Ananthanarayanan
2. Coussement et al.
3. Lu et al.
4. Prasad & Madhavi
5. Dahiya & Bhatia
6. Ali & Anturk
7. Kaur et al.
8. Farquad et al.
9. Günther

مشتریان در صنایع مذکور از طریق سیستم‌های الکترونیک نسبتاً پیشرفته ارائه می‌شود، طبعاً داده‌های زیادی نیز از رفتار و عملکرد کاربران موجود هستند (چو و همکاران[1]، ۲۰۱۶). به جهت مقایسه باید گفت که توسعه مدل‌های پیش‌بینی رویگردانی مشتری در صنعت مخابرات به دلیل رشد سریع و رقابت شدید بین اپراتورها به نسبت سایر صنایع بسیار بیشتر بوده است (امین و همکاران[2]، ۲۰۱۷).

پیش‌بینی رویگردانی در صنعت بانکداری

در تحقیقات مرتبط با پیش‌بینی رویگردانی در بانکداری تکنیک‌های مختلف داده‌کاوی شامل مورد استفاده قرار گرفته است. از آن جمله می‌توان به شبکه‌های عصبی[3] (زوریک[4]، ۲۰۱۶)، درخت تصمیم[5] (علی و اونترک، ۲۰۱۴)، دسته‌بندی و درخت رگرسیون[6] (پراساد و مدهوی، ۲۰۱۲)، ماشین‌های بردار پشتیبان[7] (ژو و همکاران، ۲۰۱۴)، بیز ساده[8] (کائور و همکاران، ۲۰۱۳) و الگو کاوی ترتیبی[9] (چیانگ و همکاران[10]، ۲۰۰۳) اشاره نمود. در سال‌های اخیر استفاده از مدل‌های ترکیبی[11] برای بهینه‌سازی مدل‌های پیش‌بینی گسترش یافته است. یک مدل ترکیبی معمولاً از دو یا چند تکنیک بهره می‌گیرد. برای مثال طبقه‌بندی[12] و خوشه‌بندی[13] می‌توانند با یکدیگر ترکیب شوند (هوآنگ و کچادی[14]، ۲۰۱۳). همچنین ترکیب درخت تصمیم و رگرسیون در ساختن مدل پیش‌بینی رویگردانی (کایگنی و همکاران[15]، ۲۰۱۸)، یا

1. Chu et al.
2. Amin et al.
3. Artificial Neural Network
4. Zoric
5. Decision Tree
6. Classification and Regression Tree (CART)
7. Support Vector Machines
8. Naıve Bayes
9. Sequential Pattern Mining
10. Chiang et al.
11. Hybrid
12. Classification
13. Clustering
14. Huang & Kechadi
15. Caigny et al.

ترکیب شبکه‌های عصبی[1] با رگرسیون چندگانه[2] به‌منظور افزایش دقت پیش‌بینی ترکیب شدند (خاشعی و همکاران[3]، ۲۰۱۲). محققین بر این باور هستند که استفاده از مدل‌های ترکیبی می‌تواند کارایی بهتری نسبت به مدل‌های ساده ایجاد نماید (سای و لو[4]، ۲۰۰۹).

در بسیاری از تحقیقات مرتبط به پیش‌بینی رویگردانی در بانکداری، مفهوم رویگردانی به‌صورت کامل و مشتریان به دو دسته رویگردان و غیر رویگردان تقسیم شده‌اند (علی و اونترک، ۲۰۱۴؛ ژو و همکاران، ۲۰۱۴؛ کائور و همکاران، ۲۰۱۳؛ فرقاد و همکاران، ۲۰۱۴ و پراساد و مدهوی، ۲۰۱۲). هرچند در سایر صنایع مانند خرده‌فروشی از مدل‌های رویگردانی جزئی استفاده شده‌اند، اما تا جایی که نویسندگان این مقاله اطلاع دارند، موضوع رویگردانی جزئی در تحقیقات مرتبط با بانکداری مغفول مانده است.

در مورد رویگردانی جزئی[5] در صنعت خرده‌فروشی می‌توان به کار میگوئیس و همکارانش (۲۰۱۲) اشاره نمود که بر اساس رفتار خرید مشتریان و توالی آن در خواربارفروشی، مراحل مختلفی از رویگردانی را تعریف نموده و بر اساس رفتار نمونه‌ای ۷۵ هزارتایی از مشتریان یک مدل پیش‌بینی رویگردانی جزئی ساختند. نتایج آزمون این مدل، حاکی از عملکرد مناسب این مدل نسبت به مدل‌های سنتی بود (میگوئیس و همکاران، ۲۰۱۲).

با توجه به مفهوم رویگردانی جزئی می‌توان مراحلی را در مورد رویگردانی مشتری از آغاز فعالیت او تا هنگام رویگردانی کامل متصور بود؛ بنابراین پیش‌بینی اینکه کدام مشتری و با چه احتمالی در کدام مرحله از رویگردانی قرار دارد موجب توفیق بیشتر در نگهداری مشتری خواهد شد. به‌خصوص در مواردی که مشتریان موردنظر ارزشمند نیز باشند، اهمیت این پیش‌بینی به دلیل تأثیر مستقیم در سودآوری، دوچندان می‌شود (هوآنگ و کچادی، ۲۰۱۳).

1. Artificial Neural Network
2. Multiple Regression
3. Khashei et al.
4. Tsai & Lu
5. Partial Churn

در ادامه جهت تأکید بیشتر به‌طور خلاصه به برخی تحقیقات مهم در حوزه پیش‌بینی رویگردانی در صنعت بانکداری پرداخته می‌شود.

در تحقیق علی و اونترک (۲۰۱۴) چارچوبی برای پیش‌بینی رویگردانی ارائه شده است که در آن به ساختن مدل‌های پیش‌بینی از روی داده‌های تراکنش‌های مشتریان پرداخته‌اند. مهم‌ترین نوآوری آن‌ها ساختن مدل‌های پویایی است که امکان تعریف بازه‌های زمانی مختلف در مدل وجود داشته باشد و مدل به یک بازه زمانی مشخص محدود نباشد (علی و اونترک، ۲۰۱۴).

در تحقیقی چیانگ و همکاران (۲۰۰۳) با استفاده از روش قواعد انجمنی[1] به تحلیل ترتیبی[2] تراکنش‌های مشتریان در بانکداری الکترونیک پرداخته‌اند که حاصل آن قواعدی است که مشخص می‌کند مشتری‌هایی که از این نوع خدمات رویگردانی کرده‌اند چه زنجیره‌ای از انواع تراکنش‌ها را در یک پنجره زمانی قبل از رویگردانی معمولاً انجام می‌دهند تا طبق آن بتوان نوعی پیش‌بینی قبل از وقوع رویگردانی داشت و نتیجه آن نیز این بود که ۸۰ درصد از مشتریان پس از اینکه چند بار ورود ناموفق به خاطر رمز عبور اشتباه داشته‌اند دیگر از سرویس استفاده نکرده‌اند (چیانگ و همکاران، ۲۰۰۳).

آدیمو و اوینیی (۲۰۱۵) در تحقیقشان مدلی برای پیش‌بینی رفتار مشتریان مبتنی بر برخی تکنیک‌های داده‌کاوی شامل k-means و JRip پیشنهاد کرده‌اند که بر اساس داده‌های رفتار مشتریان در یک بانک خصوصی در نیجریه انجام شده است (آدیمو و اوینیی[3]، ۲۰۱۵)

کیم و همکارانش (۲۰۱۳) در تحقیق خود با استفاده از تکنیک PLS[4] یک مدل پیش‌بینی رویگردانی توسعه دادند و سپس با در نظر گرفتن متغیرهای قابل‌کنترل بازاریابی و همچنین هزینه‌های مدیریتی مترتب بر آن‌ها، به ساخت و حل مدل بهینه‌سازی هزینه‌های بازاریابی در مدیریت نگهداشت مشتری پرداختند (کیم و همکاران، ۲۰۱۳).

1. Association Rules
2. Sequential
3. Adeyemo & Oyeniyi
4. Partial Least Square

جمع‌بندی مرور ادبیات

پس از مرور مقالات مرتبط با موضوع پیش‌بینی رویگردانی، به‌عنوان جمع‌بندی مطالب ارائه شده در راستای اهداف تحقیق حاضر می‌توان به نکاتی که در ادامه می‌آید اشاره نمود.

در تقریباً تمامی مقالات مطالعه شده بیان شده که با توجه به هزینه بالای جذب مشتریان جدید به نسبت نگهداری مشتریان موجود، موضوع پیش‌بینی رویگردانی مشتریان از آن جهت اهمیت دارد که می‌تواند ابزاری کارا و نسبتاً ارزان برای سازمان‌ها فراهم کند که با استفاده از آن مشتریان در خطر خروج که هنوز سازمان را ترک نکرده‌اند شناسایی کرده و با ارائه مشوق‌هایی از رویگردانی آن‌ها جلوگیری کنند.

در اکثر مقالات مطالعه شده تعریف استفاده‌شده از رویگردانی به‌صورت کامل در نظر گرفته شده است و مشتریان به دو گروه رویگردان و غیر رویگردان تقسیم شده‌اند. تعداد کمی از مقالات روی تعریف رویگردانی به‌صورت جزئی (وقتی مشتری هنوز از بخشی از خدمات و محصولات سازمان استفاده می‌کند) تمرکز کرده‌اند که بیشتر در صنعت خرده‌فروشی بوده‌اند. تا جایی که نویسندگان مقاله بررسی کرده‌اند، تحقیقی با موضوع پیش‌بینی رویگردانی در صنعت بانکداری که بر اساس تعریف رویگردانی جزئی انجام شده باشد مشاهده نگردید. عمده تحقیقات در صنعت بانکداری نیز بر اساس تقسیم مشتریان به دو گروه رویگردان و غیر رویگردان انجام شده‌اند. لذا موضوع مقاله حاضر می‌تواند جایگاه ویژه‌ای در ادبیات موضوع پیش‌بینی رویگردانی در بانک‌ها داشته باشد.

برخی تحقیقات با استفاده از ابزارهایی مانند پرسشنامه و مصاحبه به جمع‌آوری داده و تحلیل رفتار و ذهنیت مشتریان می‌پردازند که مستلزم صرف زمان و هزینه زیاد بوده و تعمیم‌پذیری پایینی دارند. با توجه به توسعه سیستم‌های اطلاعاتی پیشرفته استفاده از روش‌های داده‌کاوی برای پیش‌بینی و تحلیل رویگردانی مشتریان بسیار متداول‌تر شده است. در مقاله حاضر نیز

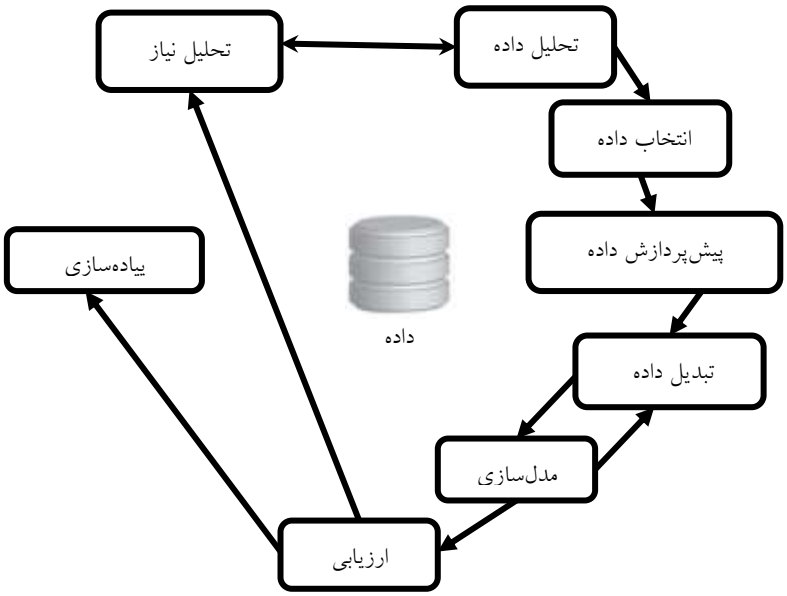
منحصراً از داده‌های تراکنش‌های بانکی و داده‌کاوی برای ساختن مدل پیش‌بینی رویگردانی استفاده شده است.

بیشتر تحقیقات مرتبط با پیش‌بینی رویگردانی در صنایعی مانند مخابرات، بانکداری، بیمه و سرگرمی انجام شده است که حجم بالایی از داده‌های تراکنش‌های مشتریان را به‌صورت الکترونیک نگهداری می‌کنند. در این مقاله نیز مدل پیش‌بینی رویگردانی بر اساس داده‌های تراکنش‌های مشتریان در صنعت بانکداری ایران ساخته شده است.

بیشتر تحقیقات بر اساس نمونه‌برداری از داده‌ها و مدل‌سازی روی نمونه کوچکی از داده‌ها انجام شده است و بر اساس آن نمونه یک جمع‌بندی کلی در خصوص رویگردانی مشتریان ارائه شده است؛ اما ساختن مدل روی داده‌های تمامی مشتریان می‌تواند منجر به شناسایی تک‌تک مشتریانی شود که در خطر خروج قرار دارند و خود این موضوع می‌تواند بخشی از سیستم مدیریت ارتباط با مشتریان باشد. در این مقاله در یک بازه زمانی سه‌ساله تمامی داده‌های تراکنش‌های مشتریان برای مدل‌سازی مورد استفاده قرار گرفته است و که مدل ساخته شده می‌تواند به ازای هر مشتری در خصوص وضعیت رویگردانی پیش‌بینی ارائه نماید.

روش تحقیق

در این بخش روش تحقیق استفاده شده در این مقاله شرح داده می‌شود. در ابتدا نموداری خلاصه از روش تحقیق اتخاذ شده آورده می‌شود و سپس با ارائه برخی تعاریف اولیه بخش‌های مختلف روش تحقیق تشریح می‌شوند. روش کلی تحقیق استفاده شده در این مقاله در شکل ۱ قابل مشاهده است.



شکل ۱: شمای کلی روش تحقیق

هدف این تحقیق شناسایی و پیش‌بینی رفتار رویگردانی جزئی با توجه به رفتار مالی مشتریان بانک‌ها است. بر اساس این هدف و مطالعات انجام شده برخی تعاریف جهت روشن کردن نتایج به شرح زیر مشخص شدند:

- **رویگردانی**: رفتاری که مشخص می‌کند مشتری تمایلی به ادامه ارتباط با بانک و استفاده از خدمات و محصولاتش را ندارد.
- **رویگردانی جزئی**: معمولاً مشتری ارتباط خود را با بانک به‌طور کامل قطع نمی‌کند و صرفاً سطح فعالیت خود را در بانک نسبت به مشتریان دیگر پایین می‌آورد.
- **مراحل رویگردانی**: بر اساس مدل رویگردانی جزئی مراحل مختلف رویگردانی از منظر سطح ارتباط مشتری با بانک را می‌توان به چهار مرحله تقسیم نمود:

- **روی‌آوری**: مشتری به‌تازگی فعالیت مالی خویش را آغاز کرده است و در مقطع مورد نظر در حال افزایش ارتباط خود با بانک است. سطح رفتار مشتری در این مرحله بالا است.
- **ثبات**: مشتری در مقطع زمانی مورد نظر رفتار نسبتاً با ثباتی دارد و تغییر شدیدی در میزان ارتباط خود با بانک ندارد. سطح رفتار مشتری در این مرحله متوسط است.
- **رویگردانی**: مشتری در مقطع زمانی مورد نظر در حال کاهش میزان فعالیت خود به نحو قابل‌توجهی بوده (که مقدار آن با کمک پروفایل رویگردانی مشخص می‌شود) و در حال کاهش ارتباط با بانک است. سطح رفتار مشتری در این مرحله پایین است.
- **ریزش**: مشتری در مقطع زمانی مورد نظر تقریباً به‌طور کامل ارتباط خود را با بانک قطع کرده و میزان فعالیت خود را به نزدیک صفر می‌رساند.

- **پارامترهای رویگردانی**: تعریف دقیق اینکه مشتری در کدام‌یک از مراحل رویگردانی جزئی قرار دارد نیاز به مشخص کردن پارامترهایی دارد که بتوان رفتار مشتری را با استفاده از آن به یکی از مراحل مذکور رویگردانی جزئی نسبت داد.
- **بازه زمانی**: برای محاسبه هر یک از پارامترهای رویگردانی بایستی یک بازه زمانی مشخص شود که پارامتر مورد نظر در آن بازه زمانی محاسبه شود. این بازه زمانی برای این تحقیق ماه شمسی در نظر گرفته شده است.
- **مدل وضعیت**: این مدل مشخص می‌کند که وضعیت سطح فعالیت مشتری از منظر رویگردانی (اینکه مشتری در کدام مرحله رویگردانی شامل روی‌آوری، ثبات، رویگردانی یا ریزش قرار دارد) برای مشتریان مختلف چگونه است.
- **زنجیره وضعیت**: این زنجیره مشخص می‌کند که در تعدادی ماه متوالی در هر ماه وضعیت رویگردانی مشتری در کدام حالت قرار دارد. طول این زنجیره به تعداد ماه‌هایی است که مشتری از ابتدای فعالیت تاکنون در بانک تراکنش دارد. هر عضو

این زنجیره حالت رویگردانی مشتری در هر ماه را مشخص می‌کند. این حالت رویگردانی می‌تواند یکی از مراحل رویگردانی جزئی تعریف شده در این تحقیق شامل روی‌آوری، ثبات، رویگردانی و ریزش باشد؛ بنابراین اولین عضو زنجیره مشخص می‌کند که در اولین ماهی که مشتری در بانک تراکنش داشته است، در کدام مرحله از چهار مرحله مذکور قرار گرفته است. دومین عضو مشخص می‌کند که در ماه دوم وضعیت مشتری چه بوده است و به همین ترتیب تا آخرین عضو این زنجیره وضعیت مشتری را در آخرین ماه فعالیت در بانک (آخرین ماهی که داده موجود است) مشخص می‌کند. تعریف مراحل رویگردانی، پارامترها و نحوه شناسایی اینکه مشتری در هر ماه در کدام مرحله از رویگردانی قرار دارد در ادامه تحقیق بیان خواهد شد.

روش کلی استفاده شده در این تحقیق را می‌توان از نظر نوع توصیفی – پیمایشی دانست. همچنین از نظر پارادایم حاکم بر تحقیق چنان‌که پیش‌ازاین نیز آمد، فرآیند گرا است. در این تحقیق در واقع کشف یک واقعیت نهفته در انبوهی از داده دنبال می‌شود و عملاً توصیفی از وضعیت موجود در مورد مشتریان بانک با تحلیل رفتار آن‌ها روی داده‌های موجود انجام می‌شود.

متدولوژی‌های مختلفی برای پارادایم فرآیند گرا به‌عنوان پرکاربردترین رویکرد در تحقیقات داده‌کاوی وجود دارند. پرکاربردترین این متدولوژی‌ها شامل متدولوژی کشف دانش در داده (KDD[1])، فرآیند استاندارد داده‌کاوی در صنایع مختلف (CRISP-DM[2]) و متدلوژی SEMMA[3] هستند. با توجه به اینکه این متدولوژی‌ها نقاط مشترک زیادی دارند، جهت بهره‌گیری از مزایای تمامی آن‌ها، ترکیبی از آن‌ها به شرح جدول ۱ زیر مورد استفاده قرار می‌گیرند.

1. Knowledge Discovery in Databases
2. CRoss Industry Process for Data Mining
3. Sample, Explore, Modify, Model, Assess

جدول ۱: متدولوژی تحقیق

عنوان مرحله	شرح مرحله	مرحله متناظر در CRISM-DM	مرحله متناظر در KDD	مرحله متناظر در SEMMA
تحلیل نیاز	در این مرحله با توجه به نیازهای کسب‌وکار صورت مسئله داده‌کاوی روشن می‌شود.	تحلیل کسب‌وکار	درک نیاز	
تحلیل داده	در این مرحله داده‌ها بررسی شده و نکات مهم در مورد داده‌ها مشخص می‌شود.	تحلیل داده	درک داده	
انتخاب داده	در این مرحله بخشی از داده که با اهداف مشخص‌شده سازگار است انتخاب می‌شود.		انتخاب داده	نمونه داده
پیش‌پردازش	در این مرحله مشکل داده‌های ناقص و معیوب با حذف یا اصلاح برطرف می‌شود.	آماده‌سازی داده	پیش‌پردازش	کاوش در داده
تبدیل داده	در این مرحله داده‌ها به شکل مورد نیاز برای مدل‌سازی داده‌کاوی تبدیل می‌شوند.		تبدیل داده	تغییر شکل داده
مدل‌سازی	در این مرحله روی داده‌های تبدیل شده مدل داده‌کاوی ساخته می‌شود.	مدل‌سازی	داده‌کاوی	مدل‌سازی
ارزیابی	در این مرحله دقت و کیفیت مدل ساخته شده مورد ارزیابی قرار می‌گیرد.	ارزیابی	تفسیر و ارزیابی	ارزیابی
پیاده‌سازی	این مرحله پایانی بوده که مدل ساخته شده مورد استفاده قرار می‌گیرد.	پیاده‌سازی مدل	پیاده‌سازی	

البته لازم به ذکر است، طبق ماهیت تکاملی پارادایم فرآیند گرا، کلیه مراحل ذکرشده با توجه به نتایج حاصل از هر مرحله ممکن است نیاز به انجام اصلاحات و تکرار مرحله یا مراحل قبلی داشته باشد. این رفت‌وبرگشت و انجام اصلاحات و تکرار مراحل آن‌قدر ادامه می‌یابد تا اینکه بر اساس شاخص‌های کمی و یا کیفی نتایج تا حد خوبی قابل‌قبول باشند. در ادامه به شرح

جزئیات روش تحقیق در قالب مراحل مذکور در جدول بالا می‌پردازیم:

تحلیل نیاز

در این مرحله از طریق نظرسنجی از خبرگان با استفاده از شیوه مصاحبه نیمه ساختاریافته مراحل مختلف و پارامترهای مؤثر بر قرار گرفتن مشتری در هر یک از مراحل رویگردانی جزئی استخراج شدند. معیارهای خبرگی در این تحقیق با توجه به اهداف و سؤالات تحقیق عبارت‌اند از افرادی که حداقل تجربه ۴ سال در بخش‌های بازاریابی یا ارتباط با مشتریان بانک‌ها یا سطوح مدیریت ارشد بانک‌ها (مدیریت‌های ستاد مرکز و بالاتر) همچنین طراحان نرم‌افزار در شرکت‌های معتبر فناوری اطلاعات که محصول تحلیلی مبتنی بر انبار داده و هوش تجاری در حداقل یک بانک خصوصی داشته باشند. با توجه به این معیارها و دسترسی محقق نهایتاً از نظرات ۱۸ خبره در این تحقیق بهره‌برداری شد. لیست پارامترها و تعاریف آن‌ها در بخش نتایج آمده است.

در این تحقیق در دو مرحله از نظرسنجی از خبرگان فوق‌الذکر استفاده شد. در مرحله اول از طریق به‌کارگیری روش مصاحبه نیمه ساختاریافته با تشریح اهداف تحقیق از آن‌ها خواسته شد تا پارامترها و شاخص‌هایی که از نظر آنان می‌تواند نشان‌دهنده رفتار رویگردانی مشتریان باشد را مشخص کنند. پس از اجرای این مرحله و جمع‌بندی نتایج مصاحبه‌ها، تعداد ۱۰ پارامتر که لیست آن‌ها در بخش نتایج آورده شده‌اند استخراج شد.

پس‌ازاین مرحله با توجه به تعداد زیاد پارامترها و حجم بالای داده‌های بانکی و بالطبع زمان و هزینه بالای انجام محاسبات این پارامترها به ازای تک‌تک مشتریان، از خبرگان خواسته شد تا به اولویت‌بندی ده پارامتر استخراج‌شده بپردازند. این اولویت‌بندی با استفاده از روش مقایسات زوجی انجام شد که در بخش نتایج تشریح شده است.

تحلیل داده

این تحقیق روی داده‌های مربوط به رفتار مالی مشتریان بانک انجام می‌شود. با توجه به تلاش

انجام شده داده‌های تراکنش‌های مالی مشتریان چهار بانک خصوصی در این تحقیق مورد استفاده قرار گرفته است. سه تا از آن‌ها بانک‌های خصوصی ایرانی هستند و یکی از این پایگاه‌های داده مربوط به یک بانک اروپایی است که به‌عنوان یک نمونه داده استاندارد برای ارزیابی مدل‌ها و الگوریتم‌های استفاده شده در این تحقیق به کار گرفته شده‌اند. این نمونه داده واقعی یک بانک اروپایی، برای اولین بار در اختیار شرکت‌کنندگان سومین تجارب کشف دانش در پایگاه‌های داده (۱۹۹۹) قرار گرفت و بعدازآن نیز در تحقیقات دیگری نیز مورد استفاده قرار گرفت (باربارا و وو[1]، ۲۰۰۱؛ پرلیچ و هوآنگ[2]، ۲۰۰۵؛ گو و ویکتور[3]، ۲۰۰۶ و ایکونوموسکا و زروسکی[4]، ۲۰۱۱). با توجه به توافق محرمانگی داده اسم بانک‌های ایرانی و همچنین مشخصات مشتریان در این تحقیق پنهان شده‌اند.

انتخاب و پیش‌پردازش داده

نکته مهم در مورد این تحقیق این است که تمامی داده‌های موجود تراکنش‌های مشتریان در قلمرو زمانی تحقیق مورد استفاده قرار می‌گیرد و در این تحقیق از نمونه‌برداری خاصی استفاده نمی‌شود. صرفاً داده‌های ناقص و یا مشکل‌دار که ممکن است کیفیت نتایج تحقیق را دچار خدشه نمایند در مراحل ابتدایی اجرای تحقیق حذف شده‌اند.

داده‌های مورد استفاده در این تحقیق از نظر قلمرو زمانی، داده‌های تراکنش‌های روی سپرده‌های مشتریان سه بانک خصوصی ایرانی در ۳ سال (از سال ۱۳۹۳ تا ۱۳۹۶) است. در مورد بانک اروپایی داده‌های ۵ سال (از سال (۱۹۹۳ تا ۱۹۹۸) مورد استفاده قرار گرفته‌اند. در بازه‌های زمانی ذکر شده هر کدام از مشتریان روی هر نوع از سپرده‌ها هر نوع تراکنش مالی که داشته باشند در این چهار بانک ذکر شده استخراج شده و در این تحقیق از آن‌ها استفاده شده است.

1. Barbará & Wu
2. Perlich & Huang
3. Guo & Viktor
4. Ikonomovska & Džeroski

در این تحقیق داده‌های بانک اروپایی صرفاً به این دلیل مورد استفاده قرار گرفتند تا امکان تعمیم‌پذیری مدل ارائه شده نیز مورد بررسی قرار گیرد. همچنین لازم به ذکر است که برای هر یک از بانک‌ها مدل‌سازی روی داده‌های خود بانک‌ها انجام می‌شود و تنها چیزی که بین بانک‌ها مقایسه می‌شود، دقت پیش‌بینی مدل‌سازی انجام شده روی داده‌های آن‌هاست.

تبدیل داده

در این تحقیق برای مدل‌سازی مبتنی بر داده‌کاوی از داده‌های واقعی تراکنش‌های مالی حساب‌های مشتریان سه بانک خصوصی ایرانی و یک بانک اروپایی استفاده می‌شود. این داده‌ها در قالب پایگاه‌های داده اوراکل دریافت شده و جهت استفاده‌های بعدی به پایگاه داده Microsoft SQL Server منتقل شدند.

دسترسی به داده در این تحقیق از طریق اتصال به پایگاه‌های داده سامانه‌های یکپارچه بانکی برقرار می‌شود. بخشی از داده که بر اساس نیاز تعریف شده در مورد رفتار مشتریان مورد نیاز است انتخاب شده و پس از پیش‌پردازش با استفاده از ابزارهای متداول در محیط‌های مدیریت پایگاه‌های داده، پارامترهای مورد نظر به ازای هر مشتری در هر ماه محاسبه شده و در پایگاه داده دیگری به جهت تحلیل‌های بعدی ذخیره شدند.

به دلیل مواجهه به داده‌های بزرگ عمدتاً تلاش می‌شود که در مراحل گردآوری و آماده‌سازی داده‌ها از ابزارهای نرم‌افزاری متناسب برای مدیریت داده‌ها استفاده شود. هرچند که ابزار نهایی مورد استفاده در حین تحقیق مشخص می‌شود اما هم به جهت تجربه به‌کارگیری و هم به جهت امکانات در این تحقیق استفاده از ابزارهای شرکت Oracle در مدیریت داده‌ها همچنین ابزارهای Microsoft SQL Server در مدیریت انبار داده و ابزارهای پاک‌سازی و انتقال و تبدیل داده و ابزارهای Microsoft Data Mining برای خوشه‌بندی مبنا قرار خواهد گرفت. دلیل استفاده از این ابزار برای خوشه‌بندی امکان دسترسی سریع‌تر این ابزار به داده‌ها در محیط پایگاه داده به داده‌های تراکنشی مشتریان بود. در واقع به جهت وجود نسبت قابل‌توجه عملیات خواندن و نوشتن (I/O) مورد نیاز برای خوشه‌بندی،

محدودیت‌های سخت‌افزاری و مهم‌تر از همه نحوه عدم داشتن مجوز استفاده از پایتون در سرور و محیط پایگاه داده فراهم شده توسط بانک برای محققین، این تصمیم گرفته شد که خوشه‌بندی با خود ابزار داده‌کاوی موجود در سیستم مدیریت پایگاه داده انجام شود. پس از انجام این مرحله و انتقال داده‌های پردازش‌شده به سرور دیگر به جهت سرعت بیشتر برای ساختن زنجیره وضعیت و پیاده‌سازی مدل SVM از زبان برنامه‌نویسی و کتابخانه‌های Python کمک گرفته شده است.

تکنیک‌های مورد استفاده در تحلیل داده‌ها را می‌توان به دو دسته تکنیک‌های بدون نظارت[۱] که بیشتر روی استخراج مفاهیم مشخص از درون داده تمرکز دارند و همراه با نظارت[۲] که بیشتر مبتنی بر یادگیری رفتارهای نمونه از طریق یک مجموعه داده و یافتن رفتارهای مشابه در مجموعه مشابه دیگری از داده‌ها هستند، تقسیم نمود.

مدل‌سازی

در تحقیق حاضر از یک مدل ترکیبی استفاده شده است که هم تکنیک‌های بدون نظارت و هم تکنیک‌های همراه با نظارت را در بر می‌گیرد. از تکنیک‌های بدون نظارت در شناسایی حالت‌های مختلف رفتار مشتریان از منظر رویگردانی جزئی (از ورود مشتری تا خروج کامل مشتری) استفاده می‌شود و از تکنیک‌های همراه با نظارت در پیش‌بینی رفتار مشتری در تغییر حالت از یک حالت رویگردانی به حالتی دیگر، استفاده خواهد شد. در این مرحله بخشی از داده که برچسب مشخص دارد و رفتار شناخته‌شده‌ای را تبیین می‌کند به‌عنوان مجموعه داده یادگیری[۳] به مدل داده می‌شود و در اصطلاح مدل آموزش[۴] داده می‌شود. مدل آموزش‌دیده روی هر مجموعه مشابه دیگری از داده‌ها رفتارهای مورد نظر را جستجو نموده و در صورت کشف مشابهت موارد یافته شده را مشخص می‌کند.

1. Unsupervised
2. Supervised
3. Training Set
4. Train

پیاده‌سازی و ارزیابی

دقت مدل‌های ساخته شده معمولاً از طریق خود مجموعه داده مورد استفاده برای آموزش، ارزیابی می‌شود. روش کار به این طریق است که بخشی از داده‌هایی که برچسب رفتار مورد انتظار را خورده‌اند (مثلاً حدود ۲۰ درصد) از ابتدا جدا شده و در مرحله آموزش مورد استفاده قرار نمی‌گیرند. در واقع آموزش با ۸۰ درصد داده انجام می‌شود. از ۲۰ درصد جدا شده در آزمون مدل بهره گرفته می‌شود که بر حسب اینکه مدل چقدر خوب بتواند نتایج مورد انتظار که در ۲۰ درصد جدا شده مشخص است را پیش‌بینی نماید، دقت مدل مشخص می‌شود. البته مقدار درصد جدا شده برای آزمون در هر مرحله از تحقیق بر اساس نیاز و در نظر گرفتن ملاحظات داده‌های بزرگ متفاوت در نظر گرفته شده است.

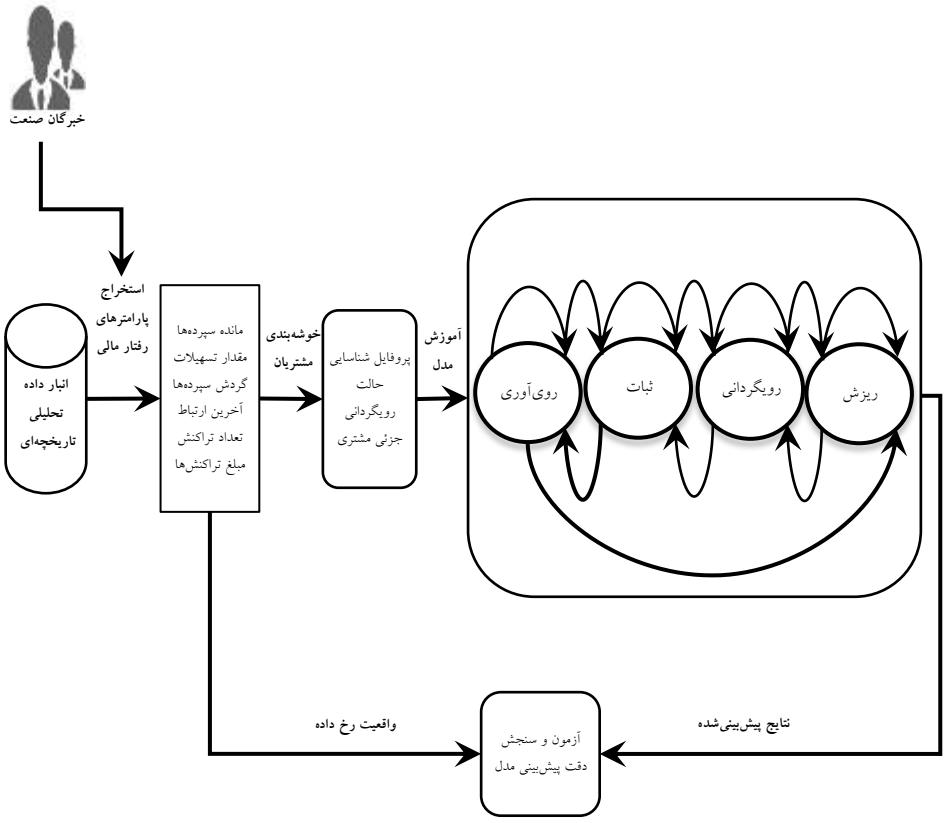
به این طریق می‌توان به‌جز پیش‌بینی رفتار رویگردانی که در تحقیق حاضر مدنظر است، احتمال دقت این پیش‌بینی را نیز بیان کرد؛ یعنی در واقع می‌توان گفت که مشتری به چه احتمالی در چه مرحله‌ای از رویگردانی قرار دارد و این پیش‌بینی هم احتمالاً چقدر دقت دارد.

چارچوب کلان تحقیق

چارچوب کلان نظری تحقیق حاضر در شکل ۲ قابل مشاهده است. منبع داده‌ای اصلی مورد استفاده برای این تحقیق انبار داده تحلیلی است که داده‌های مشتریان و رفتار مالی آن‌ها را طی سه سال به‌صورت تاریخچه‌ای نگهداری می‌کند. ابتدا مشتریان بر مبنای میزان فعالیتشان با استفاده از تکنیک‌های داده‌کاوی خوشه‌بندی می‌شوند. از سوی دیگر با استفاده از تجارب خبرگان قواعدی استخراج می‌شود که بر مبنای این قواعد و همچنین نتایج خوشه‌بندی هر یک از مشتریان مورد نظر در هر مقطع زمانی ماهیانه وضعیت مشتری از منظر رویگردانی مشخص می‌شود. در هر مقطع زمانی مشتری می‌تواند یکی از چهار حالت روی‌آوری، ثبات، رویگردانی و ریزش را داشته باشد.

پس از مشخص شدن پروفایل رویگردانی مشتری در مقاطع زمانی مختلف، وضعیت رویگردانی مشتری برای حدود ۳۰ ماه به‌عنوان ورودی به مدل پیش‌بینی داده می‌شود و از طریق تکنیک طبقه‌بندی داده‌کاوی مدل پیش‌بینی ساخته می‌شود. مدل ساخته شده این قابلیت را دارد که با داشتن آخرین وضعیت مشتری بتواند وضعیت او در مقطع یا مقاطع بعدی را پیش‌بینی کند. مثلاً درصورتی‌که آخرین وضعیت مشتری ثبات باشد، مدل مورد نظر پیش‌بینی می‌کند که با چه احتمالی در همان وضعیت باقی‌مانده و با چه احتمالی از وضعیت ثبات به وضعیت‌های دیگر تغییر می‌کند.

به‌منظور آزمون مدل و سنجش دقت پیش‌بینی وضعیت مشتریان، حدود ۶ ماه از داده تاریخچه‌ای (به‌عنوان داده آزمون) نگه داشته شده و وارد مدل نمی‌شود. پس از ساخته‌شدن مدل پیش‌بینی، برای تعدادی از مشتریان وضعیت رویگردانی در این ۶ ماه از طریق مدل مورد پیش‌بینی قرار می‌گیرد. با توجه به اینکه واقعیت رخ داده در مورد مشتریان نیز در ۶ ماه مورد بررسی از طریق پروفایل رویگردانی در اختیار قرار دارد، میزان صحت پیش‌بینی مدل قابل‌اندازه‌گیری است. لذا می‌توان با این روش برای مدل میزانی از دقت در پیش‌بینی را محاسبه نمود. در واقع با این چارچوب نه‌تنها می‌توان وضعیت رویگردانی یک مشتری را پیش‌بینی نمود بلکه بر مبنای نتایج آزمون می‌توان گفت که این پیش‌بینی تا حدی دقت دارد.



شکل ۲: چارچوب کلان نظری تحقیق

نتایج

خلاصه نتایج حاصله از هر مرحله روش تحقیق ارائه شده به شرح زیر ارائه می‌شود:

تحلیل نیاز

پارامترهای تحلیل رفتار رویگردانی مشتریان با توجه به مصاحبه با خبرگان به شرح جدول ۲ زیر خلاصه می‌شوند:

جدول ۲: پارامترهای استخراج شده مؤثر در رویگردانی

نشانه	نام پارامتر	تعریف	نشانه رویگردانی
AcSt	وضعیت حساب	حسابی که بسته شده یا مدت زیادی به‌صورت راکد باقی مانده باشد نشان‌دهنده رویگردانی مشتری است.	حساب‌های راکد و غیرفعال
AcNo	تعداد حساب	هر چه تعداد حساب‌های فعال مشتری بیشتر باشد نشان‌دهنده سطح فعالیت بالاتر مشتری با بانک است.	تعداد کم حساب فعال
M	گردش حساب	هرچه مبالغ تراکنش‌های مشتری با بانک کمتر باشد نشان‌دهنده سطح پایین فعالیت و رویگردانی مشتری است.	گردش کمتر حساب
L	طول ارتباط	هرچه مدت‌زمانی که مشتری در بانک افتتاح حساب کرده بیشتر باشد نشان‌دهنده عدم تمایل مشتری به رویگردانی است.	طول ارتباط کوتاه
R	تازگی ارتباط	هرچه از مدت‌زمانی که مشتری آخرین تراکنش را با بانک انجام داده بیشتر گذشته باشد نشان‌دهنده سطح کمتر فعالیت مشتری است.	تازگی کمتر ارتباط
F	تعداد تراکنش‌ها	هر چه تعداد تراکنش‌های مالی مشتری با بانک در بازه زمانی مورد بررسی بیشتر باشد نشان‌دهنده سطح بالاتر فعالیت مشتری است.	تعداد کمتر تراکنش‌ها
LtBl	مانده‌حساب (مانده آخر روز)	میانگین ماهانه مانده آخر روز هر سپرده مشتری در بانک هر چه بیشتر باشد سطح فعالیت مشتری بالاتر است.	مانده‌حساب پایین‌تر
MnBl	مانده‌حساب (کمترین مانده روز)	میانگین ماهانه کمترین مانده‌حساب‌های مشتری در هرروز در بانک هر چه بیشتر باشد سطح فعالیت مشتری بالاتر است.	مانده‌حساب پایین‌تر
AvBl	مانده‌حساب (میانگین مانده روز)	میانگین ماهانه میانگین مانده‌حساب‌های مشتری در هرروز در بانک هر چه بیشتر باشد، سطح فعالیت مشتری بالاتر است.	مانده‌حساب پایین‌تر
MCr	گردش بستانکار حساب	هرچه جمع مبالغ ورودی به‌حساب مشتریان بالاتر باشد سطح فعالیت مشتری بالاتر است.	گردش بستانکار پایین‌تر

به جهت تشخیص دقیق اینکه کدام پارامترها از مجموع ۱۰ پارامتر فهرست شده اهمیت بیشتری دارند از روش مقایسه‌های زوجی در تحلیل سلسله‌مراتبی[1] (AHP) استفاده می‌شود. در این روش با استفاده از پرسشنامه‌های مقایسه‌ای بین معیارها با طیف پنج نقطه‌ای به شرح جدول ۳ از نظرات خبرگان اخذ شد:

جدول ۳: طیف نقطه‌ای مقایسه‌های زوجی

ترجیح یکسان	کمی بهتر	بهتر	خیلی بهتر	کاملاً بهتر
۱	۳	۵	۷	۹

روی داده‌های مقایسه‌های زوجی برای خبرگانی که پرسشنامه‌ها را تکمیل کردند، در مورد هر مقایسه میانگین حسابی گرفته شده و برای تشکیل بردار ویژه از میانگین هندسی استفاده شده است. ضمناً بردار ویژه بر اساس روش MiniMax نرمال‌سازی شده است.

در جدول ۴ ماتریس نتیجه بردار ویژه مقایسه‌های زوجی انجام شده توسط خبرگان آمده است.

جدول ٤: نتیجه مقایسه‌های زوجی پارامترهای رویگردانی

پارامتر	AcSt	AcNo	M	L	R	F	LtBl	MnBl	AvBl	MCr	بردار ویژه	بردار ویژه نرمال
AcSt	۱	۰/۲۳٦	۰/۱۲۹	۰/۱۹٤	۰/۱٤	۰/۱۳٤	۰/۱٤۹	۰/۲۲	۰/۱۷۳	۰/۱۱۳	۰/۱۹۳	۰/۰۱۲
AcNo	٤/۲۳۱	۱	۰/۱٤٦	۰/۲٤۱	۰/۱٦۳	۰/۱٥٥	۰/۱۷٦	۰/۲۸۳	۰/۲۱	۰/۱۲	۰/۲۹۳	۰/۰۱۸
M	۷/۷٦۹	٦/۸٤٦	۱	٥/۷٦۹	۰/۳۰۲	۰/۲۷۷	٤/۰۷۷	٦/۲۳۱	٥/۱٥٤	۰/۱۲۷	۱/۸۳۳	۰/۱۱۱
L	٥/۱٥٤	٤/۱٥٤	۰/۱۷۳	۱	۰/۱۹٤	۰/۱۹٤	۰/۲٤٥	۳/۱٥٤	۰/۳۱۷	۰/۱۳٦	۰/٥۸٥	۰/۰۳٥
R	۷/۱٥٤	٦/۱٥٤	۳/۳۰۸	٥/۱٥٤	۱	۰/۳۱۷	٤/٥۳۸	٦/۰۷۷	٤/۸٤٦	۰/۱۳۸	۲/۳۱۳	۰/۱٤
F	۷/٤٦۲	٦/٤٦۲	۳/٦۱٥	٥/۱٥٤	۳/۱٥٤	۱	٤/۲۳۱	٥/٤٦۲	٤/٥۳۸	۰/۱٤۲	۲/۹۰۱	۰/۱۷٥
LtBl	٦/٦۹۲	٥/٦۹۲	۰/۲٤٥	٤/۰۷۷	۰/۲۲	۰/۲۳٦	۱	٤/٥۳۸	۳/۷٦۹	۰/۱٤٤	۱/۱۷۲	۰/۰۷۱
MnBl	٤/٥۳۸	۳/٥۳۸	۰/۱٦	۰/۳۱۷	۰/۱٦٥	۰/۱۸۳	۰/۲۲	۱	۰/۲٦٥	۰/۱٤۸	۰/٤۲۹	۰/۰۲٦
AvBl	٥/۷٦۹	٤/۷٦۹	۰/۱۹٤	۳/۱٥٤	۰/۲۰٦	۰/۲۲	۰/۲٦٥	۳/۷٦۹	۱	۰/۱٦	۰/۸۱۱	۰/۰٤۹

1. Analytical Hirearchy Processing (AHP)

نتیجه مقایسه‌های زوجی پارامترها نشان می‌دهد که گردش بستانکار حساب (تراکنش‌های واریز به سپرده) مهم‌ترین پارامتر مؤثر در سطح رفتار مشتری از منظر رویگردانی است. در مرحله بعدی پارامترهای تعداد تراکنش سپرده، شاخص تازگی، جمع گردش (جمع مبلغ واریز و برداشت) و میانگین مانده آخر روز هر سپرده، به ترتیب روی رفتار رویگردانی مؤثر هستند. همچنین این موضوع نشان می‌دهد که پارامترهایی مثل وضعیت حساب، تعداد حساب‌ها، میانگین کمترین مانده مشتری بر مبنای کمترین و میانگین مانده روز و طول مدتی که مشتری در بانک افتتاح سپرده نموده است، تأثیر معنی‌داری در رفتار رویگردانی مشتری ندارد. لذا این پارامترها در این تحقیق مورد استفاده قرار نگرفته و حذف می‌شوند.

تحلیل و انتخاب داده

در جدول ۵ زیر بررسی آماری داده‌های مورد استفاده در این تحقیق آورده شده است:

جدول ۵. بررسی آماری داده‌های منابع داده‌ای تحقیق

ردیف	ویژگی	بانک اروپایی	بانک ۱	بانک ۲	بانک ۳
۱	تعداد سپرده‌ها	۴۵۰۰	۱۹۹۹۹۸۱	۳۴۴۵۰۹۵	۶۴۳۵۳۹۶
۲	تعداد تراکنش‌ها	۱۰۵۶۳۲۰	۳۵۴۹۰۰۳۰۶	۱۱۸۸۷۱۱۳۹۸	۱۳۱۹۹۱۶۵۷۸
۳	کوچک‌ترین تاریخ	۱۹۹۳-۰۱-۰۱	۱۳۹۳/۰۷/۳۰	۱۳۹۳/۰۷/۳۰	۱۳۹۳/۰۷/۳۰
۴	بزرگ‌ترین تاریخ	۱۹۹۸-۱۲-۳۱	۱۳۹۶/۰۵/۳۱	۱۳۹۶/۰۵/۳۱	۱۳۹۶/۰۵/۳۱
۵	میانگین مبلغ تراکنش‌ها	۵۹۲۴ دلار	۴۲۲۰۰۲۴۳ ریال	۳۹۵۱۶۶۲۸ ریال	۴۰۵۸۵۸۹۲ ریال
۶	انحراف معیار مبلغ تراکنش‌ها	۹۵۲۳ دلار	۳۳۹۶۷۴۸۷۶۵ ریال	۶۳۰۷۲۶۸۶۱۱ ریال	۱۶۴۵۳۵۸۵۶۴ ریال

همچنین با در نظر گرفتن اهداف تحقیق، با توجه به نحوه رفتار خاص مشتریان حقوقی و سازمانی در رویگردانی، در این تحقیق صرفاً داده‌های مشتریان حقیقی انتخاب شدند. ضمناً در

خصوص نوع سپرده‌ها تنها سپرده‌های کوتاه‌مدت روزشمار انتخاب می‌شوند. در خصوص تراکنش‌ها نیز تنها تراکنش‌های غیر خودکار که مشتری با بانک انجام می‌دهد، انتخاب می‌شوند.

پیش‌پردازش

انحراف معیار بالای مبلغ تراکنش‌های بانک‌های ایرانی نشانه وجود داده‌های پرت است. به جهت حذف داده‌های پرت از این داده‌ها از روش نرمال‌سازی Z-Score برای شناسایی داده‌های پرت استفاده شد. بدین ترتیب که داده‌هایی که در فاصله ۶ برابر انحراف معیار با میانگین قرار دارند به‌عنوان مبالغ خیلی بزرگ و مبالغ خیلی کوچک شناسایی و حذف شدند که تعداد آن‌ها در ادامه آمده است. ازآنجاکه کیفیت خوشه‌بندی تا حد زیادی به داده‌های پرت وابسته است پس از حذف این داده‌ها به دلیل سادگی کار از خود مبلغ تراکنش به‌جای Z-Score آن استفاده شد که کیفیت بالای خوشه‌بندی پس از حذف داده‌های پرت مؤید این مطلب است.

در این مرحله برخی داده‌های مشکل‌دار که با اهداف تحقیق هم‌راستا نبودند، شناسایی و حذف شدند. تراکنش‌های خودکار (۱۴۴۳۱۵ مورد در بانک اروپایی)، تراکنش‌های با مبلغ صفر (۲۸۹ مورد در بانک اروپایی، ۴۶۵۵۶۹، ۶۵۳۳۴۲ و ۲۳۴۷۹ مورد به ترتیب در بانک‌های ایرانی شماره ۱ و ۲ و ۳)، تراکنش‌های با مبلغ خیلی پایین (۳۱۶۹۶۹، ۲۹۷۰۵۰ و ۸۱۹۸۵۹ مورد به ترتیب در بانک‌های ایرانی شماره ۱ و ۲ و ۳) و خیلی بالا (۴۳۹۶۵۷، ۲۱۱۹۸۶ و ۲۱۳۰۲۰ مورد به ترتیب در بانک‌های ایرانی شماره ۱ و ۲ و ۳) حذف شدند.

همچنین مانده پس از تراکنش در مورد تعداد ۲۲۴۵ از تراکنش‌های بانک اروپایی، ۵۳۶۷ از تراکنش‌های بانک شماره ۱، ۱۲۵۶ از تراکنش‌های بانک شماره ۲ و ۱۰۷۵۳ از تراکنش‌های بانک شماره ۳ مغایرت داشت که در این مرحله اصلاح گردید.

ارتباط شماره مشتری و شماره سپرده نیز در ۲۱۵ سپرده از بانک شماره ۱ و ۴۳۲ سپرده از بانک شماره ۳ مشکل داشت که به‌ناچار این سپرده‌ها نیز حذف شدند.

تبدیل داده

به جهت آماده‌سازی برای تحلیل، کلیه داده‌های هر چهار بانک به شرح جدول ۶ به تفکیک تجمیع و تبدیل شدند.

جدول ٦: تبدیل داده‌های منابع داده‌ای

فیلد	معنی	توضیحات
C_No	شماره مشتری	
YEAR	سال	شمسی/میلادی
MONTH	ماه	شمسی/میلادی
MCr	گردش بستانکار حساب (مجموع مبالغ واریزشده)	ریال/ دلار در ماه
FCr	تعداد تراکنش واریز انجام شده در ماه	
RCr	آخرین روزی که مشتری تراکنش واریز داشته است	در ماه ۱ تا ۳۱
LtBt	میانگین مانده آخر روز سپرده مشتری	ریال/ دلار در ماه

لازم به ذکر است که در این مرحله، داده‌های تمام سپرده‌های مشتریان (در صورت وجود بیش از یک سپرده) تجمیع شدند. همچنین در ماه‌هایی که احیاناً مشتریان تراکنشی نداشته‌اند تمامی پارامترها صفر در نظر گرفته شده و در جدول داده تبدیل شده وارد شدند.

برای تبدیل داده‌ها به ساختار فوق‌الذکر ابتدا داده‌ها به‌صورت خام از منابع داده‌ای اوراکل استخراج شد. سپس داده‌های با ابزار Microsoft SQL Server Data Import در پایگاه داده ماکروسافت ذخیره شد. سپس با ابزار Data Quality Service و یک پروژه کیفیت داده تعریف شده داده‌ها مورد بررسی قرار گرفتند. در این مرحله مشکلات ورود داده به پایگاه داده SQL Server شناسایی و مرتفع شد. سپس برای پالایش محتوایی داده‌ها که در بخش قبلی به آن اشاره شد از اسکریپت‌های T-SQL استفاده شد.

مدل‌سازی

همان‌طور که در بخش مقدمه نیز اشاره شد، به دلیل ماهیت رویگردانی جزئی به‌خصوص در صنعت بانکداری ایران، در این مرحله ابتدا بایستی کلاس رفتاری مشتری در هر ماه از منظر رویگردانی تعیین شود؛ یعنی رویگردانی به‌جای یک مرحله کامل در چهار مرحله جزئی (شامل روی‌آوری، ثبات، رویگردانی یا ریزش) بایستی تعریف شود. همچنین تعریف رویگردانی نباید با تعریف ثابت انجام شود و بر اساس رفتار مالی مشتریان در هر مجموعه داده بایستی بر اساس داده‌ها انجام شده و در طی زمان هم تغییر کند.

برای انجام این مهم از مدل خوشه‌بندی با الگوریتم Hirearchial K-Means استفاده شد. تعداد خوشه‌ها سه عدد در نظر گرفته شد. پنجره زمانی برای انجام این خوشه‌بندی شامل سه سال برای بانک‌های ایرانی و پنج سال برای بانک اروپایی در نظر گرفته شد که خوشه مشتریان در هر ماه مشخص شود. به ترتیب خوشه‌های اول تا سوم مراحل روی‌آوری، ثبات و ریزش را مشخص می‌کنند. در موارد صفر بودن تراکنش‌های مشتری در ماه مرحله ریزش در نظر گرفته می‌شود.

پس از این مرحله در خصوص ساختن مدل پیش‌بینی ابتدا بایستی زنجیره‌ای از تغییرات وضعیت مشتری در ماه‌های مختلف از شروع تا پایان بازه زمانی مورد بررسی، در خصوص هر مشتری ساخته می‌شود. این زنجیره که در فرمول تعریف شده است، از ابتدای دوره زمانی منبع داده‌ای شروع می‌شود و تا پایان آن ادامه دارد.

$$CSEQ_i = \{S_1, S_2, \cdots, S_n\} \qquad S_i \begin{cases} \text{ریزش} : 0 \\ \text{رویگردانی} : 1 \\ \text{ثبات} : 2 \\ \text{روی‌آوری} : 3 \end{cases}$$

فرمول (۱)

به ازای هر مشتری i یک زنجیره به طول n از وضعیت‌های رویگردانی ماهانه مشتری تهیه می‌شود. لازم به ذکر است که طول این زنجیره‌ها برای مشتریان بانک اروپایی ۶۰ و برای بانک‌های ایرانی ۳۶ است. برای استفاده بهتر در ذخیره‌سازی صفرهای ابتدای زنجیره حذف شدند؛ یعنی اینکه زنجیره وضعیت مشتری از اولین ماهی شروع می‌شود که در آن تراکنش دارد.

برای ساختن مدل پیش‌بینی از ماشین بردار پشتیبان[1] (SVM) استفاده می‌کنیم که در آن بردار ورودی زنجیره وضعیت مشتری در t ماه گذشته و خروجی آن احتمال قرار گرفتن مشتری در وضعیت‌های مختلف در ماه t+1 است.

برای ساختن داده آموزش تمامی زیرمجموعه‌های مرتب t+1 عضوی از زنجیره‌های وضعیت مشتریان را استخراج می‌کنیم. سپس t عضو اول زنجیره را به‌صورت بردار ورودی و عضو t+1 را به‌عنوان کلاس خروجی به مدل SVM می‌دهیم. پس از تکمیل شدن آموزش مدل پیش‌بینی SVM(t) تهیه می‌شود. این مدل می‌تواند با گرفتن بردار t عضوی (t ماه پشت سر هم) به‌عنوان ورودی، احتمال قرار گرفتن در وضعیت‌های ریزش، رویگردانی، ثبات و روی‌آوری را در ماه t+1 را به‌عنوان خروجی محاسبه نماید.

$$CSEQ_i = \{S_1, S_2, \cdots, S_n\} \rightarrow SVM(t) \rightarrow PS_{t+1} \begin{cases} P_0 : \text{احتمال ریزش} \\ P_1 : \text{احتمال رویگردانی} \\ P_2 : \text{احتمال ثبات} \\ P_3 : \text{احتمال روی‌آوری} \end{cases} \quad \sum_{j=0}^{3} P_j = 1$$

فرمول (۲)

برای مثال در جدول ۷ نمونه‌ای از SVM(2) برای ۱۰۰۰ مشتری از داده‌های بانک اروپایی آورده شده است.

1. Support Vector Machine

جدول ۷: نمونه‌ای از SVM(2) برای ۱۰۰۰ مشتری بانک اروپایی

	مدل پیش‌بینی SVM(2)		احتمال وضعیت برای ماه بعد (t+1)			
			ریزش	رویگردانی	ثبات	روی‌آوری
وضعیت مشاهده‌شده دو ماه اخیر t=2	ریزش	ریزش	٪ ۱۰۰	۰	۰	۰
	رویگردانی	ریزش	٪ ۱۰۰	۰	۰	۰
	ثبات	ریزش	٪ ۱۰۰	۰	۰	۰
	روی‌آوری	ریزش	٪ ۱۰۰	۰	۰	۰
	ریزش	رویگردانی	عدم وجود داده			
	رویگردانی	رویگردانی	۰	٪ ۹۲/۱۱	٪ ۴/۷۵	٪ ۳/۱۴
	ثبات	رویگردانی	۰	٪ ۶۵/۸۷	٪ ۲۱/۹۳	٪ ۱۲/۲۰
	روی‌آوری	رویگردانی	۰	٪ ۴۱/۴۲	٪ ۱۵/۸۲	٪ ۴۲/۷۶
	ریزش	ثبات	عدم وجود داده			
	رویگردانی	ثبات	۰	٪ ۴۹/۳۶	٪ ۳۷/۷۰	٪ ۱۲/۹۴
	ثبات	ثبات	۰	٪ ۲/۶۰	٪ ۸۹/۴۵	٪ ۷/۹۵
	روی‌آوری	ثبات	۰	٪ ۹/۷۳	٪ ۶۷/۸۳	٪ ۲۲/۴۴
	ریزش	روی‌آوری	عدم وجود داده			
	رویگردانی	روی‌آوری	۰	٪ ۳۸/۲۱	٪ ۱۶/۰۷	٪ ۴۵/۷۲
	ثبات	روی‌آوری	۰	٪ ۱۰/۹۸	٪ ۶۴/۱۶	٪ ۲۴/۸۶
	روی‌آوری	روی‌آوری	۰	٪ ۱۲/۹۳	٪ ۹/۸۰	٪ ۷۷/۲۷

در این تحقیق مدل پیش‌بینی با استفاده از SVM روی داده‌های آموزش هر چهار بانک با طول زنجیره ۱ تا ۱۹ ساخته شد.

ارزیابی

برای ارزیابی دقت مدل‌های پیش‌بینی زنجیره وضعیت مشتریان انتخاب شده برای داده آزمون را به طول مورد نظر مدل پیش‌بینی به مدل داده و نتیجه پیش‌بینی با واقعیت مقایسه می‌کنیم. میانگین احتمال پیش‌بینی‌شده وقوع وضعیتی توسط مدل که در واقع اتفاق افتاده است را به‌عنوان دقت پیش‌بینی مدل در نظر می‌گیریم. برای مثال برای ارزیابی دقت مدل نمونه SVM(2) که در بالا مطرح شد، ۱۰ مشتری از داده‌های آزمون بانک اروپایی به‌طور تصادفی انتخاب می‌کنیم که زنجیره فعالیت آن‌ها به شرح جدول ۸ است.

جدول ۸: زنجیره فعالیت ۱۰ مشتری نمونه

شماره مشتری	زنجیره وضعیت
۱۰۶۴۴	۳۳۳۳۳۳۳۳۲۲۳۳۳۳۳۳
۳۰۳۹	۱۲۲۲۲۲۲۲۲۲۲۲۲۲۲۲۳۲۲۲۳۲۲۲۲۲۳۲۲۲۲۱۱۱۲۲۲۲۲۳۱۲۲۲۲
۱۲۷۰	۲۲۲۲۲۲۲۲۲۲۲۲۲۲۲۱۲۲۲۲۲۲۲۱۲۲۳۲۲۲
۲۲۶	۲۲۲۲۲۲۲۲۳۳۲۲۲۲۲۲۲۲۲۱۲۲
۲۹۰۹	۲۲۳۲۲۲۲۲۲۲۱۲۲۲۲۲۲۲۲۲۲۲۲۲۲۳۲۲۲۲
۶۹۲	۲۲۲۲۲۳۲۲۲۲۲۱۲۲۲۲۳۲۲۲۲۲۲۲۲۲
۳۲۷۱	۱۱۱۱۲۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۳۱۱۱۱
۳۸۲۰	۱۲۳۲۲۲۲۲۲۲۲۲۲۲۲۲۲۲۱۲۲۲۳۲۲۲۳۲۲۲۲۲۲۲۲۲۲۲۲۳۲۳۲۲۲۳۲۲۲۲۲۲۲۲۲۲
۶۷۹	۲۲۳۱۲۲۳۲۲۲۲۳۲۲۲۳
۹۱۸۸	۳۳۳۳۳۳۳۳۲۲۳۲۳۳۲۲۱۱۳۳۳۳۳۲۳۳۳۳۳۲۲۲۲۳۲۳۳۲۲۳۳۲۳۳۳۳۲۲۱۲۳۳

سپس با مقایسه وضعیت سه ماه ابتدایی این زنجیره‌ها با خروجی نمونه SVM(2) که در جدول ۷ نشان داده شدند می‌توانیم به ارزیابی دقت SVM(2) به شرح جدول ۹ بپردازیم.

جدول ۹: ارزیابی دقت مدل نمونه ساخته شده

شماره مشتری	وضعیت در ماه اول و دوم		وضعیت در ماه سوم	احتمال وقوع وضعیت ماه دوم از نظر مدل SVM(2)
۱۰۶۴۴	۳ (روی‌آوری)	۳ (روی‌آوری)	۳ (روی‌آوری)	٪ ۷۷/۲۷
۳۰۳۹	۲ (ثبات)	۲ (ثبات)	۲ (ثبات)	٪ ۸۹/۴۵
۱۲۷۰	۲ (ثبات)	۲ (ثبات)	۲ (ثبات)	٪ ۸۹/۴۵
۲۲۶	۲ (ثبات)	۲ (ثبات)	۲ (ثبات)	٪ ۸۹/۴۵
۲۹۰۹	۲ (ثبات)	۲ (ثبات)	۳ (روی‌آوری)	٪ ۷/۹۵
۶۹۲	۲ (ثبات)	۲ (ثبات)	۲ (ثبات)	٪ ۸۹/۴۵
۳۲۷۱	۱ (رویگردانی)	۱ (رویگردانی)	۱ (رویگردانی)	٪ ۹۲/۱۱
۳۸۲۰	۱ (رویگردانی)	۲ (ثبات)	۳ (روی‌آوری)	٪ ۱۲/۹۴
۶۷۹	۲ (ثبات)	۲ (ثبات)	۳ (روی‌آوری)	٪ ۷/۹۵
۹۱۸۸	۳ (روی‌آوری)	۳ (روی‌آوری)	۳ (روی‌آوری)	٪ ۷۷/۲۷
میانگین				**٪ ۶۳/۳۳**

نتایج نهایی

برای تعیین وضعیت رویگردانی مشتریان در هر ماه ابتدا خوشه‌بندی سلسله‌مراتبی با روش Hirearchail K-Means بر اساس پارامترهای RCr، FCr، MCr و LtBl انجام شد. لازم به تأکید مجدد است که خوشه‌بندی روی رکوردهای داده‌ای پارامترهای محاسبه شده برای هر مشتری در هر ماه (مطابق جدول ۶) انجام شد. نتایج این خوشه‌بندی در قالب جداول زیر قابل مشاهده است. میانگین ضرائب عضویت در خوشه‌ها نشان می‌دهد که خوشه‌بندی انجام شده به خوبی مشتریان را افراز می‌کند.

جدول ۱۰: نتیجه خوشه‌بندی بانک اروپایی

توصیف خوشه	مرکز خوشه (بانک اروپایی)				میانگین ضریب عضویت
	RCr	FCr	MCr	LtBl	
رویگردانی	۲۸/۵	۱/۹	۴۳۷۲	۲۴۶۳	۱/۹۲٪
ثبات	۳۰/۴	۲/۲	۱۸۰۱۴	۴۰۰۹۲	۲/۸۷٪
روی‌آوری	۳۰/۴	۲/۹	۵۱۱۳۳	۵۰۶۸۱	۵/۹۳٪

جدول ۱۱: نتیجه خوشه‌بندی بانک شماره ۱

توصیف خوشه	مرکز خوشه (بانک شماره ۱)				میانگین ضریب عضویت
	RCr	FCr	MCr	LtBl	
رویگردانی	۸/۳	۱/۳	۱۲۸۲۱۹۶۳	۵۵۸۰۶۳	۲/۹۴٪
ثبات	۱۵/۸	۵/۲	۲۱۸۱۱۹۷۴۵	۴۷۲۸۸۰۹	۵/۹۳٪
روی‌آوری	۲۷/۲	۱۶/۴	۱۳۲۹۶۶۷۸۱۱	۲۹۶۲۳۰۶۰	۸/۸۹٪

جدول ۱۲: نتیجه خوشه‌بندی بانک شماره ۲

توصیف خوشه	مرکز خوشه (بانک شماره ۲)				میانگین ضریب عضویت
	RCr	FCr	MCr	LtBl	
رویگردانی	۹/۳	۱/۲	۱۲۵۶۵۵۲۴	۴۱۲۹۶۷	۳/۹۱٪
ثبات	۱۴/۲	۴/۵	۱۸۳۲۲۰۵۸۶	۶۰۵۲۸۷۶	۵/۹۰٪
روی‌آوری	۲۰/۴	۱۹/۸	۱۷۱۵۲۷۱۴۷۶	۳۶۱۴۰۱۳۳	۸/۹۳٪

جدول ۱۳: نتیجه خوشه‌بندی بانک شماره ۳

توصیف خوشه	مرکز خوشه (بانک شماره ۳)				میانگین ضریب عضویت
	RCr	FCr	MCr	LtBl	
رویگردانی	۶/۳	۱/۱	۱۳۵۳۹۹۹۳	۵۸۹۳۱۵	۲/۹۶٪
ثبات	۱۲/۸	۴/۹	۳۰۷۱۱۲۶۰۱	۵۴۷۷۹۱۰	۶/۸۳٪
روی‌آوری	۲۳/۴	۱۲/۸	۱۶۰۸۸۹۸۰۵۱	۳۳۵۶۲۹۲۷	۱/۸۸٪

خروجی این خوشه‌بندی منجر به تعیین کلاس وضعیت رویگردانی (ریزش - بدون تراکنش، رویگردانی، ثبات و روی‌آوری) در هر ماه برای همه مشتریان چهار بانک مذکور شد.

سپس برای هر بانک ۱۹ مدل پیش‌بینی با طول زنجیره وضعیت‌های ورودی بین ۱ تا ۱۹ ساخته شد. سپس خروجی هر ۱۹ مدل برای داده‌های آزمون برای هر بانک مورد ارزیابی قرار گرفت که نتایج ارزیابی به شرح جدول ۱۴ زیر بود.

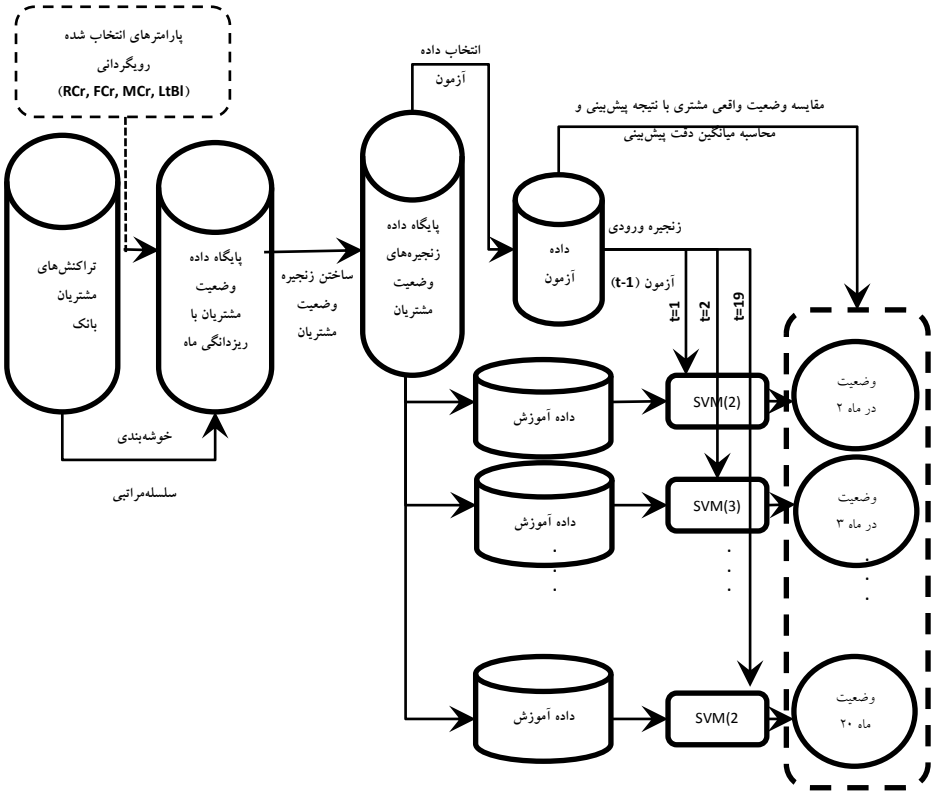
جدول ١٤: دقت پیش‌بینی مدل‌ها

تعداد ماه ورودی به مدل (t)	دقت پیش‌بینی مدل بانک اروپایی SVMBKEU(t)	دقت پیش‌بینی مدل بانک شماره ۱ SVMBK01(t)	دقت پیش‌بینی مدل بانک شماره ۲ SVMBK02(t)	دقت پیش‌بینی مدل بانک شماره ۳ SVMBK03(t)
۱	٪ ۶۸/۸	٪ ۵۸/۱	٪ ۶۳/۷	٪ ۶۰/۷
۲	٪ ۷۱/۱	٪ ۵۷/۷	٪ ۶۵/۰	٪ ۶۱/۱
۳	٪ ۶۹/۷	٪ ۶۰/۹	٪ ۶۴/۸	٪ ۶۲/۰
۴	٪ ۷۰/۰	٪ ۶۴/۷	٪ ۶۵/۱	٪ ۶۵/۹
۵	٪ ۷۴/۸	٪ ۶۳/۸	٪ ۶۶/۷	٪ ۶۸/۳
۶	٪ ۷۶/۹	٪ ۶۸/۹	٪ ۶۷/۸	٪ ۶۹/۸
۷	٪ ۷۶/۹	٪ ۷۰/۳	٪ ۶۸/۹	٪ ۷۰/۹
۸	٪ ۷۵/۳	٪ ۷۶/۸	٪ ۷۱/۷	٪ ۷۱/۹
۹	٪ ۷۸/۹	٪ ۷۸/۲	٪ ۷۳/۲	٪ ۷۵/۸
۱۰	٪ ۸۹/۸	٪ ۷۷/۳	٪ ۷۲/۹	٪ ۷۸/۴
۱۱	٪ ۹۰/۹	٪ ۸۰/۴	٪ ۷۵/۳	٪ ۸۰/۷
۱۲	٪ ۸۹/۳	٪ ۸۱/۱	٪ ۷۸/۱	٪ ۸۲/۹
۱۳	٪ ۹۳/۱	٪ ۸۵/۱	٪ ۸۰/۴	٪ ۸۲/۳
۱۴	٪ ۹۲/۷	٪ ۸۴/۶	٪ ۸۰/۶	٪ ۸۴/۱
۱۵	٪ ۹۲/۴	٪ ۸۵/۹	٪ ۸۱/۷	٪ ۸۵/۷
۱۶	٪ ۹۶/۱	٪ ۸۵/۷	٪ ۸۲/۱	٪ ۸۶/۶
۱۷	٪ ۹۴/۶	٪ ۸۷/۱	٪ ۸۳/۶	٪ ۸۸/۲
۱۸	٪ ۹۵/۶	٪ ۸۹/۰	٪ ۸۵/۹	٪ ۹۰/۴
۱۹	٪ ۹۶/۱	٪ ۹۰/۶	٪ ۸۷/۱	٪ ۹۱/۱

روند بهبود نتایج ارزیابی با افزایش طول زنجیره وضعیت در مدل‌سازی در شکل ۳ قابل مشاهده است.

شکل ۳: تغییرات دقت پیش‌بینی با افزایش طول زنجیره

به جهت حفظ یکپارچگی، خلاصه نتایج استخراج شده در تحقیق در قالب شکل ۴ به‌صورت زیر ارائه می‌شود.



شکل ٤: چارچوب یکپارچه یافته‌ها

در شکل بالا مشخص است که در ابتدا پارامترهای مؤثر در رویگردانی مشتریان که با استفاده از نظرات خبرگان انتخاب شده‌اند برای هر مشتری در هر ماه محاسبه شده، سپس روی داده‌ها خوشه‌بندی انجام شده و پس از مشخص شدن وضعیت (خوشه) هر مشتری در هر ماه، نتایج در یک انبارک داده ذخیره شدند. سپس برای هر مشتری یک زنجیره از وضعیت‌های مختلف او در ماه‌های مختلف تهیه و در قالب یک پایگاه داده ذخیره شد. پس از آن بخشی از داده‌های زنجیره وضعیت به‌منظور سنجش دقت پیش‌بینی کنار گذاشته شد و سپس

زیرمجموعه‌هایی دو عضوی، سه عضوی و ... بیست عضوی جهت آموزش استخراج شده و مدل‌های پیش‌بینی با ورودی زنجیره وضعیت به طول یک، دو، سه و ... و نوزده عضو ساخته شد. پس از ساخته شدن ۱۹ مدل برای هر بانک، وضعیت مشتریان در داده‌های آزمون یک‌به‌یک به تک‌تک مدل‌ها به‌عنوان ورودی ارائه شد و نتیجه پیش‌بینی حاصل از مدل با نتیجه واقعی مشاهده‌شده در داده آزمون مقایسه شد. تعداد پیش‌بینی درست تقسیم‌بر تعداد کل پیش‌بینی‌های انجام شده توسط هر مدل به‌عنوان درصد دقت پیش‌بینی مدل در نظر گرفته شد که در جداول و نمودارهایی که پیش‌ازاین ارائه شد، قابل مشاهده است.

نتایج تحقیق

در این تحقیق تلاش شد تا مدلی برای شناسایی و پیش‌بینی وضعیت رویگردانی مشتریان بانک‌ها بر اساس داده‌های تراکنش‌های بانکی در ایران ارائه شود. به‌منظور دستیابی به این هدف ابتدا پارامترهایی که می‌تواند به‌عنوان سنجه‌هایی برای شناسایی وضعیت مشتریان از منظر رویگردانی مورد استفاده قرار گیرد استخراج شد. همچنین این پارامترها اولویت‌بندی شده و چهار پارامتر با بیشترین تأثیر شامل آخرین تراکنش بستانکار (RCr)، تعداد تراکنش بستانکار (FCr)، مجموع مبالغ واریزی مشتری در ماه (MCr) و میانگین ماهانه آخرین مانده روز (LtBL) انتخاب شدند. در انتخاب و نحوه تعریف پارامترها به‌طور ویژه‌ای به شرایط خاص بانکداری ایران توجه شد.

با استفاده از پارامترهای مذکور و تکنیک داده‌کاوی خوشه‌بندی سلسله‌مراتبی، مدلی برای شناسایی وضعیت مشتریان در هر ماه از منظر رویگردانی جزئی ارائه شد که با شرایط خاص بانکداری در ایران تناسب بیشتری دارد. با توجه به کم بودن مقالاتی که به موضوع رویگردانی جزئی در بانک‌ها پرداخته‌اند، این مدل یکی از جنبه‌های نوآوری این تحقیق به شمار می‌آید. میانگین ضرایب عضویت بالای مشتریان در خوشه‌ها و همچنین کیفیت بالای خوشه‌بندی انجام شده حاکی از نتایج مناسب در این مرحله است. در مورد بانک‌های ایرانی کیفیت

خوشه‌بندی با پارامترهای انتخاب شده کمی بهتر است که تناسب بیشتر تعریف این پارامترها را با صنعت بانکداری ایران نشان می‌دهد.

در ادامه مفهوم زنجیره تغییرات وضعیت مشتریان معرفی شد که این مفهوم می‌تواند وضعیت رویگردانی جزئی مشتریان در ماه‌های متوالی را نشان دهد. این زنجیره وضعیت می‌تواند نمایشگر پویایی تغییرات وضعیت رویگردانی برای هر مشتری در بانک باشد. با استفاده از این ابزار می‌توان سابقه رفتار مشتری از منظر رویگردانی را به‌عنوان داده ورودی برای آموزش در اختیار مدل‌های پیش‌بینی قرار داد.

برای ساختن مدل پیش‌بینی از تکنیک داده‌کاوی ماشین‌های بردار پشتیبان که جزء متداول‌ترین ابزارهای ساختن مدل‌های پیش‌بینی رویگردانی در اکثر صنایع است، استفاده شد. با توجه به اینکه مقدار داده‌های تاریخچه‌ای در دسترس از مشتریان (طول زنجیره قابل تهیه برای هر مشتری) برای پیش‌بینی رویگردانی ممکن است متفاوت باشد، با استفاده از داده‌های آموزش برای هر بانک ۱۹ مدل مختلف هر یک با داده‌های زنجیره‌ها با طول از ۲ تا ۲۰ وضعیت، ساخته شد. این مدل‌ها بر اساس اینکه چقدر داده تاریخچه‌ای از مشتری در دسترس است می‌توانند برای پیش‌بینی وضعیت مشتری مورد استفاده قرار گیرند. به این شکل که زنجیره فعالیت مشتری در گذشته در به‌عنوان ورودی به مدل داده شده و مدل می‌تواند وضعیت مشتری در ماه بعدی را پیش‌بینی نماید.

برای سنجش دقت پیش‌بینی مدل‌ها از تعدادی زنجیره فعالیت مشتریان که به‌منظور استفاده در آزمون مدل‌ها کنار گذاشته شده بودند استفاده شد. برای هر مدل ساخته شده زنجیره فعالیت مشتریان با طول مورد نظر به‌عنوان ورودی به مدل داده شده و سپس نتیجه پیش‌بینی وضعیت مشتری توسط مدل در ماه بعد استخراج می‌شد. این نتیجه با وضعیت واقعی مشتری در ماه بعدی که در زنجیره آزمون مشخص بود مقایسه شده و میزان درست بودن پیش‌بینی هر مدل نسبت به کل پیش‌بینی‌های انجام شده، به‌عنوان دقت پیش‌بینی مدل محاسبه شد.

نتایج به‌دست‌آمده از محاسبه دقت پیش‌بینی مدل‌ها این موضوع را نشان دادند که با افزایش تعداد طول زنجیره ورودی به مدل‌ها، دقت پیش‌بینی به‌طور قابل‌توجهی افزایش می‌یافت. این

یافته نشان می‌دهد که هر چه میزان داده‌ای که از فعالیت گذشته مشتری در دسترس بانک قرار داشته باشد، بیشتر باشد می‌توان امید داشت که مدل‌های ساخته شده با دقت پیش‌بینی بالاتری عمل نمایند.

نکته بسیار مهم که در نتایج این تحقیق مشاهده می‌شود این است که درمجموع با افزایش طول زنجیره ورودی به مدل پیش‌بینی، دقت پیش‌بینی مدل افزایش می‌یابد. البته این افزایش لزوماً خطی نبوده و در برخی از اعداد طول زنجیره به‌صورت مقطعی کاهش نیز داشته است که این می‌تواند به دلیل ماهیت انتخاب تصادفی داده‌های آزمون باشد. هرچند روند کلی تغییرات دقت پیش‌بینی با افزایش طول زنجیره وضعیت، افزایشی است. این مطلب به این معنی است که هر چه سابقه بیشتری از مشتری از منظر رویگردانی جزئی را بتوانیم به مدل پیش‌بینی ارائه کنیم می‌توانیم انتظار داشته باشیم که مدل با دقت بهتری عمل نماید.

همچنین در خصوص دقت پیش‌بینی ملاحظه می‌شود که مدل پیش‌بینی ساخته شده در بانک اروپایی دقتی کمی بالاتر از بانک‌های ایرانی دارد که این می‌تواند به دلیل ثبات بیشتر رفتار در مشتریان بانک اروپایی نسبت به مشتریان ایرانی باشد.

همچنین عدم وجود اختلاف معنادار بین دقت پیش‌بینی مدل‌های ساخته شده روی داده‌های بانک اروپایی که در دسترس عمومی قرار دارد و داده‌های بانک‌های ایرانی تا حد زیادی می‌تواند نشانه‌ای از امکان تعمیم‌پذیری نحوه مدل‌سازی به بانک‌های مختلف باشد. همچنین با توجه به متفاوت بودن بانک‌های ایرانی با یکدیگر از منظر اندازه، تعداد مشتری و نوع فعالیت بانک، عدم وجود اختلاف معنادار در دقت پیش‌بینی، قدرت تعمیم‌پذیری نحوه مدل‌سازی ارائه شده را نشان می‌دهد.

درنهایت به نظر می‌رسد که مدل‌های ساخته شده با توجه به حجم نسبتاً بالای داده‌های به کار گرفته شده برای ساختن مدل، به‌خوبی جوابگوی اهداف بانک‌ها در پیش‌بینی رویگردانی مشتریان خواهد بود.

جمع‌بندی

با نگاهی بر تحقیقات انجام شده در زمینه رویگردانی مشتریان بانک‌ها، درمی‌یابیم که تقریباً همه آن‌ها از یک تعریف رویگردانی به‌صورت کامل بهره جسته و با تقسیم مشتریان به رویگردان و غیر رویگردان از تکنیک‌های دسته‌بندی برای ساختن مدل‌های پیش‌بینی استفاده نموده‌اند. در این تحقیق به دلیل ماهیت پیچیده‌تر رفتار مشتریان به‌خصوص در صنعت بانکداری ایران از مفهوم رویگردانی جزئی استفاده شده و رفتار مشتریان با استفاده از داده‌های چهار بانک مختلف در قالب چهار وضعیت مختلف از منظر رویگردانی (ریزش، رویگردانی، ثبات و روی‌آوری) با روش خوشه‌بندی مشخص شد. همچنین با توجه به تغییرات احتمالی مشتریان در وضعیت رویگردانی در طی زمان، زنجیره‌ای از وضعیت مشتریان در ماه‌های متوالی برای هر مشتری تهیه شد که مبنای پیش‌بینی رفتار بعدی مشتری قرار گیرد.

در ادامه با استفاده از تکنیک SVM، مدل‌های پیش‌بینی با زنجیره‌های با طول متفاوت از وضعیت مشتریان در هر یک از بانک‌ها ساخته شد. دقت پیش‌بینی مدل‌های ساخته شده با استفاده از داده‌های آزمون سنجیده شده و نشان داده شد که هر چه طول زنجیره‌هایی که با استفاده از آن مدل پیش‌بینی ساخته و آموزش داده می‌شود، بیشتر باشد دقت پیش‌بینی بیشتر است.

به‌جز نتایج حاصله جنبه‌های مهمی در استفاده از نتایج این تحقیق وجود دارند که در ادامه به آن می‌پردازیم. همان‌طور که ذکر شد در این تحقیق پارامترهای دقیقی از رفتار مشتریان بانک‌ها از منظر رویگردانی استخراج و تعریف شد که با صنعت بانکداری به‌ویژه در ایران تناسب دارد. با توجه به اینکه اکثر مقالات این حوزه از تعاریف ساده برای رویگردانی مشتریان بانک بهره جسته‌اند، این پارامترها می‌تواند توسط محققین این حوزه در صنعت بانکداری در آینده مورد استفاده قرار بگیرد.

مفهوم زنجیره وضعیت رویگردانی جزئی مشتریان می‌تواند از دیگر جنبه‌های کاربرد این تحقیق باشد. همان‌طور که پیش‌ازاین نیز ذکر شد، در تمامی مدل‌های مشاهده‌شده توسط محققین در ادبیات پیش‌بینی رویگردانی در صنعت بانکداری، مشتریان به دو دسته رویگردان

و غیر رویگردان تقسیم می‌شوند که این موضوع با پیچیدگی‌هایی که صنعت بانکداری به‌خصوص در ایران دارد سازگار نیست. در این تحقیق تلاش شده پیچیدگی رفتار مشتریان بانک‌ها از منظر رویگردانی با استفاده از مفهوم زنجیره وضعیت بهتر توضیح داده شود.

همچنین استفاده ترکیبی از روش‌های داده‌کاوی خوشه‌بندی سلسله‌مراتبی و ماشین‌های بردار پشتیبان در ساختن مدل پیش‌بینی رویگردانی مشتریان در صنعت بانکداری از جنبه‌های قابل‌توجه این مقاله است. لازم به ذکر است که استفاده از ترکیبی از تکنیک‌های داده‌کاوی برای ساختن مدل‌های پیش‌بینی رویگردانی در صنایعی مانند مخابرات و خرده‌فروشی سابقه داشته است اما محققین به مقاله‌ای که این استفاده ترکیبی را در صنعت بانکداری به کار گرفته باشد، برنخوردند.

آخرین و شاید مهم‌ترین جنبه قابل‌استفاده از نتایج این تحقیق خود مدل پیش‌بینی ساخته شده است که به می‌تواند توسط مدیریت‌های ارشد بانک‌ها به‌خصوص بازاریابی و ارتباط با مشتریان در بانک‌ها مورد استفاده جدی قرار بگیرد. با توجه به‌تناسب مدل‌سازی ارائه شده در این تحقیق با کسب‌وکار بانکداری در ایران و اجرای موفقیت‌آمیز آن روی داده‌های واقعی سه بانک مختلف ایرانی، مدیران مرتبط در بانک‌ها می‌توانند این مدل‌سازی را روی داده‌های خود انجام دهند. نتایج حاصل از مدل‌سازی روی تک‌تک مشتریان بانک‌ها می‌تواند منجر به تهیه فهرستی ماهانه از مشتریانی که در خطر خروج یا کم کردن سطح فعالیت قرار دارند، شود. این لیست می‌تواند بر حسب اندازه و اهمیت مشتریان اولویت‌بندی شده و مدیران بانک می‌توانند با ابزارهایی که در اختیار دارند و ارائه مشوق‌های مختلف از رویگردانی این مشتریان جلوگیری نمایند.

لذا بنا بر توضیحات ارائه شده، نتایج این تحقیق می‌تواند برای گروه‌های زیر کاربرد داشته باشد:

- مدیران ارشد بانک‌ها به‌خصوص معاونت‌های بازاریابی و ارتباط با مشتریان
- افراد و شرکت‌های توسعه‌دهنده نرم‌افزارهای تحلیلی بانکی
- محققین حوزه‌های مرتبط با بانکداری، بازاریابی، فناوری اطلاعات و علوم کامپیوتر

استفاده از مدل‌های پیش‌بینی رویگردانی توسط بانک‌ها برای مشتریان نیز به‌صورت غیرمستقیم مفید خواهد بود؛ زیرا عملاً به‌محض اینکه مشتری به هر دلیلی از کیفیت خدمات بانک در مقایسه با بانک‌های دیگر رضایت نداشته باشد، این خود بانک است که این موضوع را پیش از اینکه مشتری تصمیم به رویگردانی بگیرد متوجه شده و واکنش نشان می‌دهد. خود این نوع واکنش پیش‌دستانه بانک می‌تواند به افزایش رضایت مشتری منجر شود.

موانع تحقیق

اصلی‌ترین مانعی که در مسیر این تحقیق وجود داشت دسترسی به داده‌های واقعی از رفتار مشتریان بانک خصوصی است. در واقع بانک‌ها به دلایل محرمانگی و هم به دلیل رقابت علاقه‌ای به انتشار داده‌های مربوط به مشتریان خود ندارند. در این تحقیق تلاش این محدودیت با پنهان کردن مشخصات مشتریان و همچنین نام بانک مورد استفاده تا حدی مرتفع شد.

از دیگر موانع این تحقیق نحوه دسترسی به داده‌ها و محدودیت‌های نصب نرم‌افزارها روی سرورهای بانک و همچنین محدودیت‌های دسترسی به محیط مدیریت پایگاه داده بود که با به‌کارگیری ابزارهای خود محیط مدیریت پایگاه داده در بخشی از تحقیق این محدودیت رفع شد.

منابع

Ali, Ö. G., & Arıtürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, *41*(17), 7889-7903.

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.

Babu, S., & Ananthanarayanan, N. R. (2018). Enhanced Prediction Model for Customer Churn in Telecommunication Using EMOTE. In *International Conference on Intelligent Computing and Applications* (pp. 465-475). Springer, Singapore.

Barbará, D., & Wu, X. (2001, July). Finding dense clusters in hyperspace: an approach based on row shuffling. In *International Conference on Web-Age Information Management* (pp. 305-316). Springer, Berlin, Heidelberg.

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.

Chiang, D. A., Wang, Y. F., Lee, S. L., & Lin, C. J. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, *25*(3), 293-302.

Chu, C., Xu, G., Brownlow, J., & Fu, B. (2016, November). Deployment of churn prediction model in financial services industry. In *Behavioral, Economic and Socio-cultural Computing (BESC), 2016 International Conference on* (pp. 1-2). IEEE.

Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, *66*(9), 1629-1636.

Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27-36.

Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on* (pp. 1-6). IEEE.

Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*, *19*, 31-40.

Guo, H., & Viktor, H. L. (2006, August). Mining relational data through correlation-based multiple view validation. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 567-573). ACM.

Günther, C. C., Tvete, I. F., Aas, K., Sandnes, G. I., & Borgan, Ø. (2014). Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, *2014*(1), 58-71.

Huang, Y., & Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, *40*(14), 5635-5647.

Ikonomovska, E., & Džeroski, S. (2011, March). Regression on evolving multi-relational data streams. In *Proceedings of the 2011 Joint EDBT/ICDT Ph. D. Workshop* (pp. 1-7). ACM.

Kaur, M., Singh, K., & Sharma, N. (2013). Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers. *International Journal on Recent and Innovation Trends in Computing and Communication*, *1*(9), 720-725.

Khashei, M., & Bijari, M. (2012). A new class of hybrid models for time series forecasting. *Expert Systems with Applications*, *39*(4), 4344-4357.

Kim, K., Jun, C. H., & Lee, J. (2014). Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Systems with Applications*, *41*(15), 6575-6584.

Kim, H. S., & Yoon, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications policy*, *28*(9-10), 751-765.

Lu, N., Lin, H., Lu, J., & Zhang, G. (2014). A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, *10*(2), 1659-1665.

Mahajan, V., Misra, R., & Mahajan, R. (2017). Review on factors affecting customer churn in telecom sector. *International Journal of Data Analysis Techniques and Strategies*, *9*(2), 122-144.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & e Cunha, J. F. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert systems with applications*, *39*(12), 11250-11256.

Najarzadeh, R., Reed, M., & Mirzanejad, H. (2013). A Study of the Competitiveness of Iran's Banking System. *Journal of Economic Cooperation & Development*, *34*(1), 93.

Oyeniyi, A. O., Adeyemo, A. B., Oyeniyi, A. O., & Adeyemo, A. B. (2015). Customer churn analysis in banking sector using data mining techniques. *Afr J Comput ICT*, *8*(3), 165-174.

Perlich, C., & Huang, Z. (2005). Relational learning for customer relationship management. In *Proceedings of international workshop on customer relationship management: data mining meets marketing*.

PKDD'99. *3rd European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD99) Discovery Challenge*: http://lisp.vse.cz/pkdd99/chall.htm

Prasad, U. D., & Madhavi, S. (2012). Prediction of churn behavior of bank customers using data mining tools. *Business Intelligence Journal*, *5*(1), 96-101.

Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013, September). Telecommunication subscribers' churn prediction model using machine learning. In *Digital Information Management (ICDIM), 2013 Eighth International Conference on* (pp. 131-136). IEEE.

Riebe, E., Wright, M., Stern, P., & Sharp, B. (2014). How to grow a brand: Retain or acquire customers?. *Journal of Business Research*, *67*(5), 990-997.

Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*(10), 12547-12553.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229.

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.

Yang, C., Shi, X., Luo, J., & Han, J. (2018). I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application.

Zhu, B., Xiao, J., & He, C. (2014). A Balanced Transfer Learning Model for Customer Churn Prediction. In *Proceedings of the Eighth International Conference on Management Science and Engineering Management* (pp. 97-104). Springer, Berlin, Heidelberg.

Zhu, B., Baesens, B., Backiel, A. E., & vanden Broucke, S. K. (2018). Benchmarking sampling techniques for imbalance learning in churn prediction. *Journal of the Operational Research Society*, *69*(1), 49-65.

Zorić, B. A. (2016). Predicting customer churn in banking industry using neural networks. *Interdisciplinary Description of Complex Systems: INDECS*, *14*(2), 116-124.