

بژوهش‌های نوین در تصمیم‌گیری  
دوره ۳، شماره ۴، زمستان ۱۳۹۷

## دسته‌بندی سرویس‌های مشتریان به لحاظ سطح استفاده از خدمات شبکه شتاب با استفاده از مدل‌های ترکیبی

شهرزاد بهناز<sup>۱</sup>، راحیل حسینی<sup>۲\*</sup>

- ۱- استادیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد الکترونیک، دانشگاه آزاد اسلامی، تهران، ایران  
۲- استادیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد شهر قدس، دانشگاه آزاد اسلامی، تهران، ایران

پذیرش: ۱۳۹۷/۴/۲۰

دریافت: ۱۳۹۶/۱۱/۲

### چکیده

با تجهیز بانک‌ها به سیستم‌های پرداخت و دریافت الکترونیکی و استفاده از کارت‌های اعتباری، مشتریان اکثر تعاملات بانکی خود را از طریق کارت‌های اعتباری و کانال‌های اعتباری نظیر دستگاه‌های خودپرداز ATM، پایانه‌های فروش POS، تلفن‌بانک، موبایل بانک و اینترنت بانک انجام می‌دهند. مشتریان با افزایش تعداد بانک‌ها و محصولات و خدمات آنان، زمانی که خدمات و محصولات بهتری پیدا کنند، به راحتی بانک خود را تغییر می‌دهند؛ بنابراین، مدیریت ارتباط با مشتری یک انتخاب اجتناب‌ناپذیر برای بانک‌هاست. یکی از ابزارهایی که در زمینه مدیریت ارتباط با مشتری می‌تواند به سازمان‌ها کمک‌کننده باشد، داده‌کاوی است. در این تحقیق برای شناسایی خدمات سودآوری که توسط مشتریان خوب مورد استقبال قرار گرفته‌اند، چهار روش داده‌کاوی درخت تصمیم، بیز ساده، K همسایه نزدیک و مدل ترکیبی استفاده شدند. هر یک از این روش‌ها بر روی داده‌های واقعی، آزمایش و کارایی هر روش سنجیده شد. نتایج ارزیابی مدل‌ها نشان داد که دقت مدل K همسایه نزدیک در شناسایی خدمات سودآور ۹۳/۲۶ درصد، دقت مدل بیز ۷۴/۸۳ درصد

E-mail: Rahil.hosseini@godsiau.ac.ir

\* نویسنده مسئول مقاله:

و دقت مدل درخت تصمیم ۹۷/۱۸ درصد است و پس از ترکیب مدل‌ها، دقت روش ترکیبی ۹۴/۸۰ درصد شده است. همچنین مدل ترکیبی ۹۶/۰۱ درصد از خدمات عادی را درست شناسایی کرده و در مجموع ۹۴/۴۴ درصد از خدمات را به درستی تشخیص داده است و عملکرد بهتری نسبت به روش بیز ساده و K همسایه نزدیک داشته است. از مقایسه نتایج ارزیابی مدل‌ها مشخص شد که استفاده از روش‌های ترکیبی در مقایسه با روش‌های دسته‌بندی مبتنی بر یک مدل، از دقت بیشتری برخوردار است.

**واژگان کلیدی:** دسته‌بندی تجمیعی؛ بیز ساده؛ درخت تصمیم؛ K همسایه نزدیک؛ رأی‌گیری.

#### ۱- مقدمه

استفاده از خدمات بانکداری الکترونیک توسط مشتریان به‌ویژه خدمات کارت، تراکنش‌های بسیار زیادی را برای بانک‌ها ایجاد نموده است و بانک‌ها با حجم عظیمی از داده‌ها در پایگاه اطلاعاتی خود مواجه هستند که می‌تواند فارغ از استفاده در عملیات بانکی به‌عنوان یک منبع ارزشمند جهت کشف دانش و الگوهای رفتاری مشتریان تلقی شود. با توجه به ضرورت شناسایی رفتار مشتریان از جهات مختلف، شناخت درگاه‌ها و خدماتی که مشتریان برای امور بانکی خود استفاده می‌کنند، می‌تواند بانک‌ها را در سیاست‌گذاری‌های آتی در خصوص ارتباط با مشتریان و ارائه خدمات متنوع و کارا یاری دهد. استفاده از راهکارهای مناسب برای دسته‌بندی و تولید اطلاعات از میان انبوهی از داده‌ها برای بانک امری ضروری و حیاتی است و روش‌های رتبه‌بندی، پاسخی به این نیاز در جهت تأمین اطلاعات موردنیاز مدیران بانک‌ها است [۱، ص ۷۱-۸]. سیستم‌های مدیریت بانکداری، حجم انبوهی از داده مشتریان را در اختیار دارند و بنابراین، تجزیه و تحلیل این داده‌ها و تصمیم‌گیری روی محصولات موردنیاز مشتریان به‌صورت دستی امکان‌پذیر نیست. برای حل این مشکل فنون داده‌کاوی استفاده می‌شوند. با کمک روش‌های خوشه‌بندی و دسته‌بندی، مشتریان بر اساس شباهت ویژگی‌هایشان دسته‌بندی شده و این ویژگی‌ها مدل می‌شوند [۲]. هدف این پژوهش، پیدا کردن مدل‌هایی با توجه به مشخصات تراکنش‌ها و پراکندگی خدمات استفاده‌شده توسط مشتریان در شبکه شتاب است تا به این وسیله خدماتی که توسط سایر بانک‌ها ارائه شده و با استقبال مشتریان روبرو شده است، شناسایی شود. این شناسایی به مدیران بانکی کمک

می‌کند تا با ارائه این خدمات در بانک خود بتوانند در بازار رقابتی امروز از سایر بانک‌ها عقب نمانند. همچنین خروجی این پژوهش به بانک مذکور کمک می‌کند تا درک مناسب و بهتری درباره خدماتی که ارائه می‌دهد و میزان رضایت‌مندی مشتریان خود و مشتریان سایر بانک‌ها از کیفیت این خدمات پیدا کند؛ به این ترتیب می‌تواند کیفیت و تنوع خدمات‌ها را افزایش دهد و به تبع، از این راه به جذب مشتری سایر بانک‌ها و مهم‌تر از آن افزایش رضایت‌مندی مشتریان خود و در نتیجه جذب منابع بیشتر و کسب سود بیشتر کمک کند.

با تشدید رقابت، حفظ مشتری هزینه‌ای به مراتب کمتر از جذب مشتری جدید دارد. بر اساس نتایج مطالعات، بانک‌ها و مؤسسات مالی به طور میانگین سالیانه حدود ۲۵ درصد از مشتریان خود را از دست می‌دهند. در نظام بانکی کشور ما نیز یکی از دغدغه‌های اصلی مدیران ریزش مشتری است. کاهش ۱ درصد نرخ ریزش مشتری منجر به ۶ درصد افزایش در میزان سود می‌شود [۳، ص ۷۹-۵۸].

در این تحقیق، برخلاف تحقیقات گذشته که بیشتر بر مباحث سنجش ریسک مشتریان، اعتبارسنجی مشتریان، خوش‌حساب و بدحساب بودن آن‌ها، شناسایی مشتریان ارزنده و میزان رضایت آن‌ها متمرکز بوده‌اند، به ارائه مدلی جهت دسته‌بندی تراکنش‌های مشتریان بر اساس نوع درگاه و نوع خدمات مورد استفاده و ارائه مدل‌های داده‌کاوی برای بررسی تراکنش‌های شبکه شتاب در یک بانک نمونه می‌پردازد که کمتر مورد توجه قرار گرفته یا می‌توان گفت تاکنون در ایران، پژوهشی در این خصوص صورت نگرفته است. با استفاده از این مدل، بانک می‌تواند درگاه‌های مختلف را از جنبه میزان فعالیت و خدمات مختلف استفاده شده توسط مشتریان در شبکه شتاب را شناسایی و با خدمات ارائه شده روی درگاه‌های محلی خود مقایسه کند و برای هر خدمت روی هر درگاه، تصمیمات جداگانه‌ای اتخاذ کند.

## ۲- مبانی نظری و پیشینه تحقیق

از زمانی که ابزارهایی به منظور کاوش داده‌ها به وجود آمد، استفاده از آن‌ها در حوزه‌های مالی مخصوصاً بانکداری مرسوم شد و همچنان نیز در حال گسترش است. تاکنون تحقیق‌ها و مقالات بسیار زیادی در مورد ارزیابی مشتریان بانک‌ها

صورت پذیرفته است که به برخی از آن‌ها در جدول ۱ به‌طور خلاصه اشاره شده است.

مصطفی اسماعیلی و بهروز مینایی در مقاله‌ای با عنوان تحلیل مشتریان بانک بر اساس میزان فعالیت در شبکه شتاب با استفاده از فنون داده‌کاوی، یک مدل برای دسته‌بندی مشتریان بر مبنای فعالیت (میزان رویگردانی، میزان ورشکستگی یا احتمال عدم پرداخت وام و...) و میزان کارمزد آن‌ها در شبکه شتاب ارائه کرده و میزان سودمندی مشتریان در شبکه شتاب را بررسی کردند. با استفاده از نتایج این تحقیق، مدیران بانکی توانستند مشتریان فعال خود که کمترین میزان کارمزد را داشته‌اند، شناسایی کنند. همچنین با ترکیب یک مدل  $R^*FMK$  و الگوریتم خوشه‌بندی K-Means، یک مدل پیشنهاد شده که مشتریان را به ۱۱ خوشه تقسیم می‌کند [۴]. اما هدف این تحقیق، تشخیص خدمات سودآور، خدمات عادی و خدمات سودآوری که توسط مشتریان وفادار مورد استقبال قرار گرفته‌اند، به‌منظور سیاست‌گذاری برنامه‌های متناسب با هر درگاه و هر خدمت است. در این تحقیق با استفاده از روش‌های دسته‌بندی تجمیعی و با ترکیب متغیرهای داده‌های خدمات، از جمله میانگین مبلغ کارمزد هر خدمت، میانگین تعداد هر خدمت و میانگین مبلغ گردش هر خدمت با مدل RFM، مدلی پیشنهاد داده شده است که خدمات سودآور و خدمات عادی را بتوان مبتنی بر آن شناسایی کرد. مدل ارائه‌شده می‌تواند در بانک‌ها به منظور پیش‌بینی سودآور بودن راه‌اندازی یک خدمت یا حتی یک درگاه، مورد استفاده قرار گیرد.

#### جدول ۱ مروری بر مبانی نظری و سوابق مربوطه

عنوان مقاله	روش
کاربرد فناوری داده‌کاوی در سیستم CRM بانک‌های تجاری [۵]	<ul style="list-style-type: none"> <li>معرفی فناوری داده‌کاوی و کاربردهای آن در بانک‌های تجاری؛</li> <li>به این نتیجه رسیده است که اغلب بانک‌های تجاری در کشور چین هنوز روی مفهوم فناوری داده‌کاوی مانده‌اند؛ بنابراین، بانک‌های تجاری و دانشگاه‌ها به تقویت تحقیق و عمل در این زمینه نیاز دارند.</li> </ul>
کاربرد داده‌کاوی در بخش بانکی با روش خوشه‌بندی و طبقه‌بندی [۶]	<ul style="list-style-type: none"> <li>پیشنهاد یک دسته‌بندی با روش‌های مبتنی بر تجمیع برای پاسخ دادن به عدم تعادل کلاس؛</li> </ul>

عنوان مقاله	روش
	<ul style="list-style-type: none"> <li>• نشان می‌دهد که روش‌های کم شمردن<sup>۱</sup> تصادفی، بسته‌بندی<sup>۲</sup> و تقویت<sup>۳</sup> گروه‌ها رفتار خوبی دارند؛</li> <li>• مشخص شد که الگوریتم‌های گروهی ارزشمند هستند و با استفاده از فنون پیش‌پردازش اطلاعات و آموزش دسته‌بندی تکی، بهبود نتایج به دست خواهد آمد؛</li> <li>• استفاده از دسته‌بندی‌های بیشتر باعث پیچیده‌تر شدن کار می‌شود، اما این افزایش پیچیدگی می‌تواند با نتایج بهتری که به دست می‌آیند، توجیه شود.</li> </ul>
<p>چارچوب کارآمد داده‌کاوی- CRM برای پیش‌بینی رفتار مشتری (۷، ص ۷۳۱-۷۲۵)</p>	<ul style="list-style-type: none"> <li>• عامل کلیدی در توسعه یک استراتژی کسب‌وکار رقابتی، درک و تجزیه و تحلیل رفتار مشتری است؛</li> <li>• چارچوب داده‌کاوی- CRM به سازمان‌ها کمک می‌کند تا مشتریان ارزشمند را شناسایی و رفتار آینده آن‌ها را پیش‌بینی کنند؛</li> <li>• پیشنهاد یک چارچوب داده‌کاوی- CRM کارآمد؛</li> <li>• مقایسه مدل بیز ساده با شبکه‌های عصبی؛</li> <li>• دقت شبکه عصبی به میزان ۸۸/۶۳ درصد نسبت به بیز ساده بهتر است.</li> </ul>
<p>ارائه مدل برای پیش‌بینی فنآوری‌های موردنیاز در بانک‌ها با استفاده از SOM و ARM [۸]</p>	<ul style="list-style-type: none"> <li>• با ترکیب خوشه‌بندی و قوانین وابستگی، نسخه‌ای از سیستم پیشنهاددهنده برای پیشنهاد خدمات بانکی ارائه کردند.</li> </ul>
<p>پیش‌بینی رفتار مشتریان بانکداری الکترونیک با استفاده از روش‌های داده‌کاوی در بانک شهر [۹]</p>	<ul style="list-style-type: none"> <li>• خوشه‌بندی مشتریان بر اساس سن، تحصیلات و جنسیت؛</li> <li>• الگوریتم PNN؛</li> <li>• الگوریتم SOTA؛</li> <li>• ارائه الگویی جهت ایجاد وفاداری در مشتریان.</li> </ul>
<p>طراحی مدلی برای بهبود سیستم‌های پیشنهاددهنده بانکی براساس پیش‌بینی علایق مشتریان: کاربرد روش‌های داده‌کاوی (۱۰، ص ۴۱۴-۳۹۳)</p>	<ul style="list-style-type: none"> <li>• پیشینه رفتاری و علایق مشتری، مبنای مهم برای تصمیم‌گیری درباره چگونگی رفتار با مشتری قرار می‌گیرد؛</li> <li>• ارائه الگویی برای بهبود سیستم‌های پیشنهاددهنده بانکی به مشتریان و ارتقا کیفیت پیشنهادها؛</li> <li>• ارائه الگویی برای پیشنهاد خدمات بانکی مناسب با دقت و عملکرد زیاد، براساس شناخت و دسته‌بندی پیشینه رفتاری مشتریان؛</li> <li>• مقایسه عملکرد روش‌های بیز ساده، K- نزدیک‌ترین همسایه، پرسپترون چندلایه و ماشین بردار پشتیبان؛</li> <li>• با استفاده هم‌زمان از چند روش، می‌توان از مزیت‌های تک‌تک روش‌ها بهره‌جداگانه‌ای برد.</li> </ul>
<p>تحلیل مشتریان بانک براساس میزان فعالیت در شبکه شتاب با استفاده از تکنیک‌های داده‌کاوی [۴]</p>	<ul style="list-style-type: none"> <li>• شناسایی مشتریان فعالی که کمترین میزان کارمزد را داشته‌اند، توسط مدیران بانکی؛</li> <li>• خوشه‌بندی مشتریان با الگوریتم K-Means؛</li> <li>• با بازتعریف متغیر R به R* و تعریف پارامتر K و تعیین اوزان با روش ماتریس مقایسه زوجی به نتایج بهتری جهت تحلیل میزان سودمندی مشتریان در شبکه تبادل بین‌بانکی رسیده است.</li> </ul>

۱. Under sampling

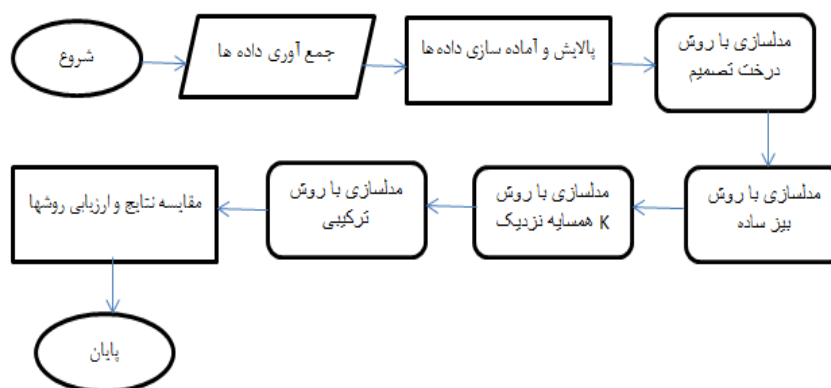
۲. Bagging

۳. Boosting

عنوان مقاله	روش
تکنیک بازار یابی محصولات مبتنی بر داده‌کاوی برای محصولات بانکی [۱۱]	<ul style="list-style-type: none"> <li>پیشنهاد یک سیستم دولایه که اول مشتریان را خوشه‌بندی می‌کند و سپس یک مدل دسته‌بندی برای محصولات و کانال‌های ارتباطی می‌سازد؛</li> <li>پیشنهاد دو راه‌کار برای تعیین محصولات بانکی و نوع کانال: روش افرازی و روش مبتنی بر مدل با افرازی؛</li> <li>ثابت شد که روش افرازی، به‌تنهایی، صحت را افزایش می‌دهد اما مدل ترکیبی دقت بالاتری دارد.</li> </ul>

## ۲. روش اجرای تحقیق

نمودار زیر مراحل انجام کار در این تحقیق را ارائه می‌نماید:



## نمودار ۱ مراحل انجام کار

ابتدا مختصری در مورد مفاهیم استفاده شده و تعاریف الگوریتم‌های داده‌کاوی به کار گرفته شده در این تحقیق، شرح و سپس مراحل کار توضیح داده شده است.

### ۱-۲- مدیریت ارتباط با مشتری (CRM)

مدیریت ارتباط با مشتری، ارتباط مشتریان با سازمان و نیازمندی‌های آنان را موردبررسی و تجزیه و تحلیل اصولی قرار می‌دهد. در واقع، CRM فرآیندی جهت گردآوری و یکپارچه‌سازی اطلاعات به‌منظور بهره‌برداری مؤثر و هدف‌دار از آن‌هاست. این اطلاعات می‌تواند در برقراری ارتباط با مشتریان، فروش، بازاریابی مؤثر یا نیازهای بازار باشد. مدیریت ارتباط با مشتری یک روش، یک نظام و از همه مهم‌تر یک راهبرد در کسب‌وکار است که هدف آن دسته‌بندی مشتریان و مدیریت

دسته‌بندی سرویس‌های مشتریان به ... شهرزاد بهناز و همکار

آن‌ها به منظور بهینه‌سازی ارزش مشتری در درازمدت و بهره‌گیری سازمان از آن است. مدیریت ارتباط با مشتری، در واقع در فرآیندهای پیدا کردن مشتری، نزدیک شدن به آن، مدیریت و ایجاد رضایت در مشتریان و نگهداری آن‌ها است. برقراری ارتباط با مشتریان به منظور شناسایی، حفظ مشتریان کلان و بهینه‌سازی ارائه خدمات در راستای افزودن به دوره ماندگاری مشتریان به سبب افزایش میزان رضایت آن‌ها و همچنین بالا بردن سودآوری مشتریان ناشی از افزایش قدرت رقابتی بیشتر و در نتیجه، کسب درآمد حاصل از فروش یا ارائه خدمت، در موفقیت بانک‌ها نقش بسزایی دارد [۱۲].

#### ۲-۲- دسته‌بندی مشتریان در سیستم بانکی

هرروزه، مشتریان مختلفی برای دریافت انواع خدمات به بانک مراجعه می‌کنند. همیشه مشتریان دارای موقعیت یکسانی نیستند. از این رو، بانک می‌بایست مشتریان وفادار خود را شناسایی کند. معیارهای وفاداری مبتنی بر رفتار (نظیر تازگی، تکرار و ارزش پولی) که در بازاریابی مستقیم کاربرد زیادی دارند، به طور روزافزونی مورد استفاده قرار می‌گیرند. این معیارها تصویر واقعی‌تری از اینکه یک شرکت تا چه اندازه نسبت به رقبایش خوب عمل می‌کند، فراهم می‌آورند. توانایی پیش‌بینی احتمال خرید را بهبود می‌دهند و اندازه‌گیری ارزش دوره عمر مشتری را ممکن می‌سازند.

#### ۲-۳- داده‌کاوی

داده‌کاوی عبارت است از فرآیند استخراج اطلاعات نهان، الگوهای جالب و روابط مشخص در حجم زیادی از داده‌ها [۱۳].

یکی از روش‌های داده‌کاوی برای تحلیل داده‌ها، دسته‌بندی است که در آن، تأثیر چند متغیر یا چند نوع داده بر روی یک نوع داده یا یک متغیر بررسی می‌شود. به متغیرها و انواع داده‌هایی که تأثیر آن‌ها بررسی می‌شود، متغیر مستقل و متغیری که تأثیر بقیه داده‌ها بر روی آن بررسی خواهد شد، متغیر وابسته یا برچسب گویند؛ بنابراین، اولین کار برای استفاده از الگوریتم‌های دسته‌بندی، مشخص نمودن متغیرهای وابسته و متغیر مستقل است. خروجی الگوریتم‌های دسته‌بندی، مدلی برای توصیف دسته‌های موجود در

داده‌هاست که پس از اینکه ثابت شد که چقدر قابل اعتماد است و چنانچه صحت آن قابل قبول بود، می‌توان از آن برای پیشگویی داده‌های جدید استفاده نمود.

#### ۱-۳-۲- درخت تصمیم

یکی از محبوب‌ترین روش‌های دسته‌بندی، درخت‌های تصمیم هستند؛ چراکه استفاده از آن‌ها به دانش گسترده‌ای نیاز ندارد. بنابراین، برای کاربردهای متفاوت در علوم مختلف بسیار مناسب است، می‌تواند داده‌های چندبُعدی و داده‌های عددی و غیرعددی را تحلیل کند، خروجی و مدل ایجادشده به راحتی قابل درک و فهم است و مراحل آموزش و دسته‌بندی درخت‌های تصمیم آسان است و به سرعت صورت می‌پذیرد. دقت دسته‌بندی درخت‌های تصمیم مناسب و زیاد است و حساس به داده‌های پرت نیست؛ چراکه می‌توان از طریق هرس کردن درخت اثر داده‌های پرت و نویزها را کاهش داد. روش درخت تصمیم در تقسیم‌بندی داده‌ها به گونه‌ای است که هیچ داده‌ای حذف نمی‌شود؛ به عبارت دیگر، تعداد داده‌ها در گره اولیه با مجموع داده‌ها در شاخه‌های درخت ایجادشده برابر است.

#### ۲-۳-۲- روش‌های دسته‌بندی بیزین

اساس این روش، قوانین احتمال است و مبتنی بر نظریه بیز است. در این روش، احتمال این‌که یک نمونه داده عضو یک دسته باشد، محاسبه می‌شود. شرط استفاده از دسته‌بندی بیز ساده، استقلال شرطی ویژگی‌ها است؛ به این معنی که تأثیر یک ویژگی بر روی برچسب داده‌ها مستقل از تأثیر سایر ویژگی‌ها بر روی برچسب داده‌ها باشد. در غیر این صورت، می‌بایست از روش‌های تعمیم‌یافته قانون بیز استفاده نمود که به شبکه‌های بیزین معروف‌اند. از ویژگی‌های بیز ساده علاوه بر سادگی و اجرای راحت آن، این است که با یک بار پیمایش داده‌ها می‌توان کلیه احتمالات را محاسبه کرد. بنابراین، سرعت و دقت خوبی را برای پایگاه‌های داده بزرگ به همراه دارد.

#### ۳-۳-۲- دسته‌بندی‌های K همسایه نزدیک

یکی از بهترین و پرکاربردترین الگوریتم‌های دسته‌بندی که از آن استفاده گسترده‌ای در کاربردهای مختلف می‌شود KNN است. اساس این روش، تشابه بین داده‌ها است؛

به عبارتی روش کار این الگوریتم به این صورت است که ابتدا نمونه‌های آموزشی که با کمک  $n$  ویژگی خود توصیف می‌شوند، در یک فضای  $n$  بُعدی ذخیره شده و سپس برای پیش‌بینی یک داده جدید، این الگوریتم دنبال  $k$  نمونه آموزشی که به داده نمونه شبیه‌تر است، می‌گردد که اصطلاحاً به آن  $k$  همسایه نزدیک گفته می‌شود. برای محاسبه شباهت و نزدیکی داده‌ها، روش‌های متفاوتی وجود دارد؛ از جمله فاصله منهایتن، اما شناخته‌شده‌ترین آن‌ها که اغلب استفاده می‌شود، فاصله اقلیدسی است.

#### ۴-۳-۲- روش‌های تجمیعی<sup>۱</sup>

یکی از روش‌های افزایش صحت دسته‌بندی، استفاده از روش‌های تجمیعی است. در این روش‌ها، مدلی برای دسته‌بندی انتخاب می‌شود که ترکیبی از چندین دسته‌بندی است. هر دسته‌بندی رأی خود را صادر می‌کند و نتیجه نهایی در مورد برچسب کلاس براساس این رأی‌ها صادر می‌شود. معمولاً صحت روش تلفیقی بهتر از دسته‌بندی‌های پایه است؛ زیرا روش تلفیقی در صورتی برچسب کلاس را به نادرستی تعیین می‌کند که بیش از نیمی از دسته‌بندی‌های تشکیل‌دهنده آن، مرتکب خطا یا اشتباه شده باشند. هر چه تنوع مدل‌های تشکیل‌دهنده روش‌های تلفیقی بیشتر باشد، نتایج بهتری از خود نشان خواهند داد.

### ۳- تجزیه و تحلیل داده‌ها

#### ۳-۱- مراحل آماده‌سازی داده‌ها

##### ۳-۱-۱- جمع‌آوری داده‌ها

بانک انتخابی قریب به ۲۰ میلیون مشتری و روزانه حدود ۵ یا ۶ میلیون تراکنش دارد که در روزهای پایان ماه و روزهای واریز یارانه، این تعداد تا ۱۰ میلیون تراکنش نیز افزایش خواهد یافت. لذا حجم داده‌های موجود بسیار زیاد است و با سخت‌افزارها و منابعی که در اختیار عموم می‌باشد، امکان پردازش این داده‌ها وجود ندارد؛ بنابراین جامعه آماری کوچک‌تر شد. به این ترتیب که در ابتدا با استفاده از روش نمونه‌گیری، حدود ۱۰۰۰۰ مشتری به صورت تصادفی انتخاب شدند و اطلاعات مربوط به این مشتریان از جمله جنسیت، تاریخ تولد، تاریخ افتتاح حساب و شماره کارت آن‌ها از

۱. Ensemble

بانک اطلاعاتی استخراج شد. سپس اطلاعات تراکنش‌های مربوط به دو سال گذشته این مشتریان یعنی از خردادماه ۱۳۹۴ تا پایان مردادماه ۱۳۹۶ جمع‌آوری شدند. در قدم بعد با نظر کارشناسان خبره بانکی، فیلدهای مؤثر مربوط به تراکنش‌های این مشتریان شناسایی و سایر فیلدها از مجموعه داده‌ها حذف شد.

### ۲-۱-۳- یکپارچه‌سازی داده‌ها

از آنجایی‌که معمولاً داده‌ها در سازمان‌ها و بانک‌ها متمرکز نیستند و سیستم‌ها پراکنده و حتی نوع داده‌ها در زیر سیستم‌های مختلف متفاوت است، لذا برای به دست آوردن آمار درست و دقیق می‌بایست داده‌ها را از زیر سیستم‌های مختلف جمع‌آوری، متمرکز و یکپارچه نمود.

### ۳-۱-۳- بررسی وضعیت داده‌ها و آماده‌سازی آن‌ها

هدف این تحقیق، بررسی تراکنش‌های غیرحضور از درگاه‌های مبتنی بر کارت‌های عضو شتاب است؛ بنابراین پس از انتخاب فیلدها و جمع‌آوری داده‌های مربوط به تراکنش‌های مشتریان، تراکنش‌های غیرحضور از درگاه‌های مبتنی بر حساب و تراکنش‌های حضور مشتریان شناسایی و حذف شدند.

پس از آن، نوبت به پاک‌سازی این داده‌ها رسید. برای شناسایی و حذف داده‌های پرت، براساس نظر کارشناسان خبره، ابتدا داده‌های مربوط به کارت‌های هدیه در مجموعه داده‌ها شناسایی شده و از آنجاکه دارندگان کارت هدیه به‌عنوان مشتری بانک محسوب نمی‌شوند، لذا تراکنش‌های کارت هدیه از مجموعه داده‌ها حذف شد. در مرحله بعد، دستگاه‌های متعلق به سایر بانک‌ها که تراکنش‌های بسیار زیادی داشتند، به‌عنوان نقاطی که احتمالاً دستگاه بانک نمونه در آن منطقه وجود ندارد و مشتریان بانک نمونه به خاطر عدم وجود دستگاه بانک نمونه و محدودیت جغرافیایی، ناچار به استفاده از دستگاه‌های سایر بانک‌ها شده‌اند، شناسایی و کلیه تراکنش‌های مشتریان از طریق این دستگاه‌ها، از مجموعه داده‌ها حذف شدند.

برای بازسازی رکوردهایی که برخی از فیلدهای آن خالی بود نیز با استفاده از تولید اعداد تصادفی مقادیری برای داده‌های گم‌شده تولید شد و به‌این‌ترتیب، تعداد مشتریان انتخابی به ۵۱۰۳ مشتری و تعداد تراکنش‌های جمع‌آوری‌شده حدود

دسته‌بندی سرویس‌های مشتریان به ... ----- شهرزاد بهناز و همکار

۱۵۱۲۴۱۱ تراکنش رسید. همچنین به منظور سهولت کار و بالا بردن میزان دقت، داده‌های غیر عددی تبدیل به داده‌های عددی شدند؛ به عنوان مثال فیلد نوع تراکنش که دارای مقادیری از جمله FT، BLPY، NFT و غیره بود با مقادیر عددی جایگزاری شد.

#### ۳-۱-۴- یافتن شاخص‌های مؤثر در تعیین مشتریان وفادار

جهت یافتن مشتریان وفادار، در ابتدا لازم بود شاخص‌های مؤثر در تعیین این‌گونه مشتریان از طریق تراکنش‌های بانکی مشخص شوند. جهت تعیین این شاخص‌ها، نظرات کارشناسان خبره از جمله کارشناسان بخش بازاریابی، روابط عمومی و فناوری اطلاعات چند بانک نمونه از طریق مصاحبه جمع‌آوری شد؛ از جمله مهم‌ترین شاخص‌هایی که در خصوص مشتریان وفادار در تراکنش‌های بانکی شناسایی شدند، می‌توان طول عمر مشتری، میانگین گردش حساب هفتگی و میانگین تعداد تراکنش هفتگی را نام برد.

#### ۳-۱-۵- یافتن شاخص‌های مؤثر در تعیین خدمات سودآور

جهت یافتن خدمات سودآور، در ابتدا لازم بود شاخص‌های مؤثر در تعیین این‌گونه خدمات از طریق تراکنش‌های بانکی مشخص شوند. جهت تعیین این شاخص‌ها نظرات کارشناسان خبره جمع‌آوری شد؛ از جمله مهم‌ترین شاخص‌هایی که در خصوص خدمات سودآور در تراکنش‌های بانکی شناسایی شدند، می‌توان به تعداد تراکنش‌ها، میانگین مبلغ گردش هفتگی و میانگین مبلغ کارمزد هفتگی اشاره نمود.

#### ۳-۱-۶- تجمیع داده‌ها

پس از تعیین شاخص‌های مشتریان وفادار و خدمات سودآور، لازم بود تا اطلاعات مربوط به مشتریان و خدمات در رابطه با هر یک از شاخص‌ها جمع‌آوری شود. در این مرحله با استفاده از نرم‌افزار (SQL) و زبان برنامه‌نویسی JAVA برنامه‌هایی نوشته شد و اطلاعات به دست آمده برای هر مشتری در یک رکورد خلاصه‌سازی و ذخیره شد. همچنین اطلاعات به دست آمده برای هر خدمت نیز در یک رکورد خلاصه‌سازی و ذخیره شد. انواع خدمات تشخیص داده شده بانک نمونه ۸۶ نوع خدمت است.

### ۳-۱-۷- ترکیب اطلاعات مشتریان و خدمات

در این مرحله، اطلاعات خدمات و مشتریان ترکیب شدند؛ به این ترتیب که پس از تشخیص انواع خدمات و دسته‌بندی خدمات بر اساس نوع درگاه و تراکنش‌های مشتریان به تفکیک نوع خدمت دسته‌بندی شدند. پس از انجام این مرحله، تعداد ۱۱۲۹۸۷ رکورد اطلاعاتی به دست آمد که این داده‌ها وارد مدل‌سازی شدند.

### ۳-۷-۸- بارگذاری داده‌ها

برای ساختن یک مدل، به داده‌هایی از هر دو دسته مشتریان وفادار و بی‌وفا و هر دو دسته خدمات سودآور و عادی نیاز است. در هر یک از مدل‌های مورد استفاده در این بخش جهت شناختن سبب خدمات سودآور مورد استقبال مشتریان وفادار، خدمت ارزنده یا عادی به عنوان متغیر وابسته در نظر گرفته می‌شود. مقدار یک برای متغیر وابسته به معنای ارزنده و سودآور بودن سرویس و مقدار صفر به معنای عادی بودن آن خدمت است. در این مطالعه از ۳۷ متغیر مستقل ثبت شده، ۲۷ مورد از آن‌ها برای فرآیند شناسایی خدمت ارزنده انتخاب شده است.

### ۳-۷-۹- مشخص کردن برچسب کلاس

در این تحقیق، روش یادگیری در مدل‌های دسته‌بندی، روش‌های با ناظر<sup>۱</sup> است. در این روش‌ها با استفاده از داده‌هایی که برچسب کلاس آن‌ها از قبل دانسته شده است، مدل‌ها ساخته یا آموزش داده می‌شوند. این نحوه آموزش، یادگیری تحت نظارت<sup>۲</sup> نامیده می‌شود؛ زیرا که مقادیر محاسبه‌شده یا تخمین زده‌شده با نتایج معلومی مقایسه می‌شوند. در اینجا برچسب داده‌های آموزشی توسط افراد خبره و کارشناسان بانکی و تعاریف نظام بانکی کشور [۱۵] تعیین شده است.

## ۴- ساخت مدل

پس از آماده‌سازی داده‌ها، در قدم اول داده‌های ورودی به‌طور تصادفی به دو دسته تقسیم می‌شوند. دسته اول که شامل ۷۰ درصد از داده‌های ورودی است، به عنوان

۱. Service Usage

۲. Supervised Methods

۳. learning supervised

دسته‌بندی سرویس‌های مشتریان به ... ----- شهرزاد بهناز و همکار

مجموعه آموزشی وارد مرحله آموزش می‌شوند. دسته دوم که شامل ۳۰ درصد دیگر از داده‌ها است، به منظور اعتبارسنجی به کار می‌روند تا مدل به دست آمده با آن‌ها ارزیابی شود. در این تحقیق به منظور افزایش دقت پاسخ‌ها، ابتدا از ۳ مدل یا دسته‌بندی پایه یعنی درخت تصمیم، قانون بیز و k همسایه نزدیک برای ایجاد مدل دسته‌بندی استفاده شد و سپس با استفاده از روش رأی‌گیری<sup>۱</sup> یک مدل ترکیبی ساخته شد؛ ورودی دسته‌بندی‌های فوق یکسان است. برای دسته‌بندی یک نمونه جدید در روش رأی‌گیری، هر یک از دسته‌بندی‌های پایه پیشگویی کلاس خود را به عنوان یک رأی برمی‌گردانند و روش، بر اساس آرای جمع‌آوری‌شده، رأی نهایی یعنی برچسب کلاس را تعیین می‌کند.

جدول پارامترهای تعیین‌شده در ساخت مدل‌های فوق به شرح زیر است:

جدول ۲ پارامترهای مدل‌های پایه

مقدار	نام پارامتر	مدل پایه
۳	حداقل تعداد برگ	درخت تصمیم
۶	حداقل تعداد ریشه	
۳۳۸۹۶	تعداد نمونه‌ها در هر تکرار	
۵	تعداد همسایه	K همسایه نزدیک

در مرحله اول، هر یک از مدل‌های فوق تک‌تک ساخته و ارزیابی شدند و در مرحله بعد، به منظور افزایش دقت و عملکرد مدل‌ها در پیش‌بینی خدمات سودآور، مدل‌های فوق با روش رأی‌گیری با یکدیگر ترکیب شدند.

## ۵- نتیجه‌گیری

### ۵-۱- مقایسه نتایج و ارزیابی روش‌ها

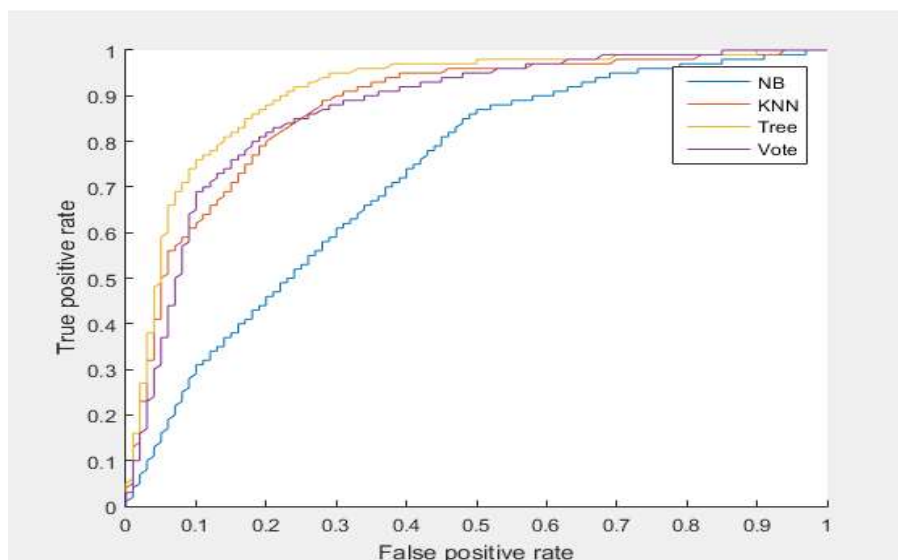
نتایج حاصل از ارزیابی هر یک از مدل‌های پایه (به‌تنهایی) و روش ترکیبی، با معیارهای منحنی مشخصه عملکرد سیستم روی کل داده‌ها، در جدول ۳ آورده شده است:

جدول ۳ نتایج ارزیابی روش ترکیبی

۱. Voting

نام روش	سطح زیر منحنی <sup>۱</sup>	بازخوانی <sup>۲</sup>	صحت <sup>۳</sup>	حساسیت <sup>۴</sup>	وضوح <sup>۵</sup>
درخت تصمیم	۹۶/۹۶	۹۷/۴۱	۹۷/۱۸	۹۷/۴۱	۹۶/۵۰
K همسایه نزدیک	۹۳/۰۶	۹۳/۴۲	۹۳/۲۶	۹۳/۴۲	۹۲/۷۱
بیز ساده	۶۸/۷۰	۷۵/۵۱	۷۴/۸۳	۷۵/۵۰	۶۱/۸۹
ترکیبی	۹۵/۲۲	۹۴/۴۴	۹۴/۸۰	۹۴/۴۴	۹۶/۰۱

همان‌طور که در جدول ۳ مشاهده می‌شود، دقت مدل K همسایه نزدیک در شناسایی خدمات سودآور ۹۳/۲۶ در صد، دقت مدل بیز ۷۴/۸۳ در صد و دقت مدل درخت تصمیم ۹۷/۱۸ درصد است، اما پس از ترکیب مدل‌ها، دقت روش ترکیبی ۹۴/۸۰ درصد شده است. مدل ترکیبی ۹۴/۴۴ درصد از خدمات سودآور را به‌درستی شناسایی کرده است؛ درحالی‌که مدل بیز ساده ۷۵/۵۰ درصد و k همسایه نزدیک ۹۳/۴۲ درصد از خدمات سودآور را درست تشخیص داده‌اند. همچنین مدل ترکیبی، ۹۶/۰۱ درصد از خدمات عادی را درست شناسایی کرده و در مجموع، ۹۴/۴۴ درصد از خدمات را به‌درستی تشخیص داده و عملکرد بهتری نسبت به روش بیز ساده و K همسایه نزدیک داشته است.



۱. AUC
۲. Recall
۳. Accuracy
۴. Sensitivity
۵. Specificity

### نمودار ۲ نمودار منحنی مشخصه عملکرد روش‌های ترکیبی

نمودار ۲ سطح زیر منحنی روش ترکیبی را در مقایسه با هر یک از مدل‌های پایه نشان می‌دهد. مدل بیز ساده با سطح زیر منحنی ۶۸/۷۰ در صد، عملکرد خوبی روی داده‌ها نداشته است، اما با ترکیب آن با مدل درخت تصمیم که سطح زیر منحنی آن ۹۵/۲۲ شده است و دقت خوبی را در تشخیص خدمات سودآور داشته است. همچنین مدل ترکیبی عملکرد بهتری را در مقایسه با مدل K همسایه نزدیک با سطح زیر منحنی ۹۳/۰۶ دارد. محور افقی در این نمودار نمایش‌دهنده خدمات سودآوری است که به‌درستی تشخیص داده نشده‌اند و محور عمودی نمایش‌دهنده خدمات سودآوری است که درست تشخیص داده شده‌اند.

جدول ۴ ماتریس درهم‌ریختگی هر یک از مدل‌های پایه و روش ترکیبی را نشان می‌دهد:

جدول ۴ نتایج الگوریتم‌های روش ترکیبی

نام روش	مثبت درست <sup>۱</sup>	مثبت نادرست <sup>۲</sup>	منفی نادرست <sup>۳</sup>	منفی درست <sup>۴</sup>
درخت تصمیم	۲۴۶۷۰	۳۰۰	۶۵۵	۸۲۷۱
K همسایه نزدیک	۲۴۴۰۳	۵۶۷	۱۷۱۹	۷۲۰۷
بیز ساده	۲۴۳۲۱	۶۳۸	۷۸۹۰	۱۰۳۶
ترکیبی	۲۴۶۵۹	۳۱۱	۱۴۵۱	۷۴۷۵

نتایج ذکرشده در جدول ۴ نشان می‌دهد که مدل ترکیبی در مقایسه با روش‌های دسته‌بندی با یک مدل، چه تعدادی را درست تشخیص داده است. مدل ترکیبی از کل تراکنش‌های مشتریانی که خدمات سودآور را مورد استقبال قرار داده‌اند، ۲۴۶۵۹ تراکنش را تشخیص داده است؛ درحالی‌که مدل K همسایه نزدیک ۲۴۴۰۳ تراکنش و بیز ساده

۱. True Positive
۲. False Positive
۳. False Negative
۴. True Negative

۲۴۳۲۱ تراکنش را شناسایی کرده‌اند. همچنین مدل ترکیبی از کل تراکنش‌های خدمات عادی ۷۴۷۵ تراکنش را تشخیص داده است و در مقابل، مدل K هم‌سایه نزدیک ۷۲۰۷ و بیز ساده تنها ۱۰۳۶ تراکنش را شناسایی کرده‌اند. بنابراین همان‌طور که از مقایسه نتایج ارزیابی فوق مشخص است، استفاده از روش‌های ترکیبی در مقایسه با روش‌های دسته‌بندی با یک مدل معمولاً از دقت بیشتری برخوردار است. همچنین پس از ترکیب چند مدل پایه و استفاده از روش رأی‌گیری، سطح زیر منحنی نیز بهبود یافته و به محور ۱ نزدیک‌تر شده است.

اما همان‌طور که از نتایج ارزیابی مشخص است، عملکرد روش درخت تصمیم با این جامعه آماری، در مقایسه با روش ترکیبی بهتر بوده است که به علت نتایج ضعیف روش بیز ساده و تأثیر آن است. بنابراین، هرچقدر تعداد مدل‌های پایه بیشتر شود، زمان مدل‌سازی و پیچیدگی کار افزایش خواهد یافت. انتخاب مدل پایه نیز در دقت نهایی مؤثر خواهد بود؛ به‌عنوان مثال، در اینجا، مدل بیز روی داده‌ها عملکرد خوبی نشان نداده است و بهتر این است که با مدل دیگری جایگزین شود. به‌عنوان ادامه کار این تحقیق، می‌توان مدل‌های پایه دیگر از جمله شبکه‌های عصبی و ماشین بردار پشتیبان را نیز ارزیابی نمود و چنانچه عملکرد بهتری داشتند، در مدل ترکیبی استفاده کرد تا دقت مدل افزایش یابد یا با توجه به نتایج خوب درخت تصمیم، می‌توان از مدل‌های ترکیبی گردآوری خود راه‌انداز<sup>۱</sup> و توان‌افزایی<sup>۲</sup> با مدل پایه درخت تصمیم استفاده نمود.

## ۶- بحث و جمع‌بندی

در این تحقیق به‌منظور مقایسه بین نتایج دسته‌بندی، در حالتی که از یک مدل پایه استفاده شود با زمانی که از مدل‌های متعدد و متفاوت استفاده شود، روش ترکیبی و روش رأی‌گیری به کار گرفته شد. در روش ترکیبی، سه مدل پایه متفاوت یعنی k هم‌سایه نزدیک، بیز ساده و درخت تصمیم مورد استفاده قرار گرفت و نتایج تک‌تک این روش‌ها با یکدیگر مقایسه شد. در نهایت با ترکیب این سه مدل و استخراج یک مدل ترکیبی که بر اساس رأی اکثریت، برچسب کلاس را تعیین می‌نمود، به مدلی رسیدیم که دقت آن از تک‌تک مدل‌ها به‌تنهایی، بالاتر بود. اگرچه داده‌های ورودی در هر سه روش بالا یکسان

۱. Bagging

۲. Boosting

دسته‌بندی سرویس‌های مشتریان به ... شهرزاد بهناز و همکار

بودند، اما نتایج ارزیابی مدل‌ها تفاوت داشتند. از مقایسه نتایج ارزیابی مدل‌های پایه و روش‌های تجمیعی می‌توان نتیجه گرفت که همواره روش‌های تجمیعی از دقت بیشتری برخوردار هستند.

در این تحقیق با یکپارچه‌سازی چند ماژول از جمله بانک اطلاعاتی، پیش‌پردازش‌های اولیه و مدل دسته‌بندی تجمیعی، پایگاه دانشی ایجاد شد که می‌تواند برای شناسایی و اعتبارسنجی مشتریان و بررسی وضعیت خدمات ارائه‌شده روی درگاه‌های بانک نمونه کارا باشد.

همان‌طور که در جدول ۵ مشاهده می‌کنید مصطفی اسماعیلی و بهروز مینایی در مقاله تحلیل مشتریان بانک بر اساس میزان فعالیت در شبکه شتاب با استفاده از فنون داده‌کاوی، با خوشه‌بندی مشتریان، میزان سودمندی آن‌ها در شبکه شتاب را بررسی کردند [۴]. بشیری و همکاران نیز در مقاله تحلیل ارزش مشتری در بانک با استفاده از روش‌های داده‌کاوی و تحلیل سلسله‌مراتبی فازی، مشتریان را بر اساس متغیرهای داده‌های تراکنشی آن‌ها به ۴ گروه پلاتینیومی، طلایی، نقره‌ای و سربی دسته‌بندی کردند [۱۴، ص ۴۳-۲۳]؛ اما در این تحقیق خدمات بانک به‌منظور سیاست‌گذاری برنامه‌های متناسب با هر درگاه و هر خدمت، دسته‌بندی شدند. مدل ارائه‌شده می‌تواند در بانک‌ها برای پیش‌بینی سودآور بودن راه‌اندازی یک خدمت یا حتی یک درگاه مورد استفاده قرار گیرد.

جدول ۵ مقایسه سایر مقالات با مدل پیشنهادی

عنوان مقاله	مدل استفاده‌شده	دستاوردها
تحلیل مشتریان بانک بر اساس میزان فعالیت در شبکه شتاب با استفاده از تکنیک‌های داده‌کاوی [۴]	الگوریتم K-Means مدل بسط‌یافته RFM	• ارائه مدلی برای دسته‌بندی مشتریان بر مبنای فعالیت و میزان کارمزد آن‌ها در شبکه شتاب؛ • بررسی میزان سودمندی مشتریان در شبکه شتاب؛ • شناسایی مشتریان فعال با کمترین میزان کارمزد.
تحلیل ارزش مشتری در بانک با استفاده از تکنیک داده‌کاوی و تحلیل سلسله‌مراتبی فازی [۱۴، ص ۴۳-۲۳]	فرآیند تحلیل سلسله‌مراتبی فازی الگوریتم K-Means مدل بسط‌یافته DFMT	• خوشه‌بندی مشتریان بر مبنای متغیرهای داده‌های تراکنشی آن‌ها؛ • میزان تأثیر هر یک از معیارهای مدل dfmt <sup>۱</sup> در امتیازدهی به مشتریان مشخص شده است.
مقاله حاضر	مدل دسته‌بندی تجمیعی	• ارائه مدلی برای دسته‌بندی خدمات بانک؛ • شناسایی درگاه‌ها و خدمات سودآور در شبکه شتاب؛

۱. Demographic-Frequency-Money-Trust

• ارائه مدل پیشنهادی با ترکیب متغیرهای داده‌های خدمات با مدل RFM		
---	--	--

از آنجایی که حجم داده‌های موجود در بانک‌ها بسیار زیاد است و پردازش این حجم داده نیاز به منابع و سخت‌افزارهای مناسب دارد، لذا در این تحقیق با استفاده از روش نمونه‌گیری جامعه آماری کوچک‌تر شده است؛ بنابراین قطعاً پاسخ به دست آمده جواب نهایی بانک نخواهد بود و چنانچه بانک بخواهد در این زمینه فعالیت کند و نتیجه آن را برای سیاست‌گذاری و تصمیم‌گیری ملاک عمل قرار دهد، می‌بایست این روش را تعمیم داده و روی کل داده‌ها و کل مشتریان لحاظ کند و با در اختیار گرفتن سرورهای قوی و سخت‌افزار و منابع مناسب مدل را از کل داده‌ها استخراج کند تا جواب نهایی بوده و بتوان بر اساس آن تصمیمات مدیریتی گرفت. همچنین می‌بایست از مباحث BigData کمک گرفت که موضوع این تحقیق نبوده و خود یک حوزه تحقیق مفصل را می‌طلبد.

از سوی دیگر، چنانچه امکان دسترسی به داده‌های کلیه زیرسیستم‌ها فراهم باشد، متغیرهای دیگری نیز وجود دارند که به شرط در اختیار داشتن آن‌ها، می‌توانند با ورود به مدل به تشخیص بهتر رفتار مشتریان و کشف الگوهای دیگر کمک کنند. با توجه به هزینه‌های هنگفتی که بانک‌ها برای راه‌اندازی شعبه جدید و حفظ شعب موجود متحمل می‌شوند، گسترش خدمات الکترونیکی و ترغیب مشتریان به استفاده از خدمات غیرحضوری می‌تواند کارگشا باشد. از سوی دیگر، درآمد حاصل از دریافت کارمزد خدمات الکترونیکی، سود سرشاری عاید بانک‌ها خواهد نمود. با توجه به دقت خوبی که روش‌های دسته‌بندی تجمیعی برای بررسی رفتار مشتریان دارند، پیشنهاد می‌شود از این روش‌ها برای بررسی کلاهبرداری و تراکنش‌های تقلب و پول‌شویی نیز استفاده شود.

## ۷- منابع

- [1] Jamal Shahrabi, Esmail Hadavandi, Data mining in the banking industry, Amirkabir University of Technology Branch, 2011, pp. 8-71.
- [2] Ravichandran, S. S., et al., Rule-base data mining systems for customer queries, Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on, IEEE, 2012.

- [3] Mohammad Taghi Taghavifard, Reza Habibi, Hamed Abdolahi, Determination of bank survivability and profitability using decision tree and regression forest model, *Journal of Modern Researches in Decision Making* vol. 2, no. 4, 2017, pp. 58-79.
- [4] Mostafa Esmaeili, Behrouz Minaei, Bank customer analysis based on the amount of activity in the accelerated network using data mining techniques, *Information Technology Management Conference Shahid Beheshti University*, 2015.
- [5] XI, Y.-p. and C. Min, Application of Data Mining Technology in CRM System of Commercial Banks, *DEStech Transactions on Engineering and Technology Research (eeta)*, 2017.
- [6] Çaliş, A., Data mining application in banking sector with clustering and classification methods, *Industrial Engineering and Operations Management (IEOM)*, 2015 International Conference on, IEEE, 2015.
- [7] Bahari, T. F. and M. S. Elayidom, An efficient CRM-data mining framework for the prediction of customer behaviour. *PROcedia Computer Science* 46: pp. 725-731, 2015.
- [8] Shakouri, H., Presenting a model for predicting needed technologies in banks using SOM and ARM, *Knowledge & Technology*, in Press, 2016.
- [9] Sajad Shokouhyar, Ali Otarkhani, Prediction of the behavior of e-banking customers by using data mining methods (Case study: Bank of the city), *International Conference on Management and Economic Development*, 2016.
- [10] Maryam Sadat Motahari, designing a Model to Improve Banking Systems Based on Predicting Customer Interests: Using Data Mining Methods, *Journal of Information Technology Management* Vol. 8, no. 2, 2017, pp. 393-414.
- [11] Mitik, M., Data Mining Based Product Marketing Technique for Banking Products, *Data Mining Workshops (ICDMW)*, 2016 IEEE 16th International Conference on, IEEE, 2016.
- [12] Azim Zarei, Designing a Structural Model for Customer Disruption in Public Banks (Case study: Selected Banks of Semnan), *Journal of Management Research in Iran* vol. 21, no. 1, 2017.

- [13] Han, J., et al., Data mining: concepts and techniques, Elsevier, 2011.
- [14] Seyed Alireza Mousavi Bashiri, Amir Afsar and Arash Mahjubi Fard, Analysis of customer value in the bank using data mining technique and fuzzy hierarchical analysis, Journal of Management Research in Iran vol.19,no1,2015,pp.23-43
- [15] "Concepts, Definitions and Methods", the website of the Central Bank of the Islamic Republic of Iran, available on 15/7/1397, available at the link: <https://www.cbi.ir/simplelist/4505.aspx>