DOI: http://dx.doi.org/10.22059/jitm.2016.57230

مدیریت فناوری اطلاعات
Journal of Information Technology Management

دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۸، شمارهٔ ۲
تابستان ۱۳۹۵
ص. ۴۱۴- ۳۹۳

طراحی مدلی برای بهبود سیستم‌های پیشنهاددهندهٔ بانکی بر اساس پیش‌بینی علایق مشتریان: کاربرد روش‌های داده‌کاوی

مریم سادات مطهری نژاد[۱]، محمدمهدی ذوالفقارزاده[۲]، احسان خدنگی[۳]، علی اصغر سعدآبادی[۴]

چکیده: امروزه بانک‌ها برای جذب و حفظ مؤثر مشتریان به ابزارهای جدیدی مانند سیستم‌های پیشنهاددهنده نیاز دارند. برخلاف عموم سیستم‌های پیشنهاددهنده که پیشنهاد بر اساس شباهت میان ترجیحات سایر کاربران به وی داده می‌شود؛ در این پژوهش از روش دسته‌بندی استفاده شده است که در آن پیشینهٔ علایق خود مشتری، مهم‌ترین ویژگی برای تصمیم‌گیری دربارهٔ خدمات بانکی مناسب به اوست. در این پژوهش از چهار دسته‌بندی‌کنندهٔ پرسپترون چندلایه، ماشین بردار پشتیبانی، K- نزدیک‌ترین همسایه و بیز ساده استفاده شد. ابتدا پس از پیش‌پردازش مجموعهٔ دادهٔ مربوط به سرویس‌های مورد استفادهٔ مشتریان مختلف بانک با چهار روش مختلف دسته‌بندی آموزش داده شدند؛ سپس اعتبار آنها با روش اعتبارسنجی ضربدری ده‌تایی به تأیید رسید و بهترین روش انتخاب شد. در انتها پیشنهاددهندهٔ نهایی که ترکیبی از چهار روش دسته‌بندی بیز ساده با عملکرد ۸۵/۴ درصد، K- نزدیک‌ترین همسایه با عملکرد ۸۳/۳ درصد، پرسپترون چندلایهٔ ۱ با عملکرد ۸۱/۴ درصد و پرسپترون چندلایهٔ ۲ با عملکرد ۹۲/۶ درصد، به‌ترتیب برای پیشنهاد چهار سرویس بانکی اینترنت، موبایل، انتقال وجه با اینترنت و پرداخت صورت‌حساب با تلفن است، ارائه شد.

واژه‌های کلیدی: بانکداری الکترونیک، داده‌کاوی، دسته‌بندی، سیستم پیشنهاددهنده.

---

۱. کارشناس ارشد مدیریت فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد الکترونیک، تهران، ایران
۲. استادیار دانشکدهٔ مدیریت دانشگاه تهران، تهران، ایران
۳. دانشجوی دکتری مهندسی نرم‌افزار، دانشکدهٔ کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران
۴. دانشجوی دکتری سیاستگذاری علم و فناوری، دانشکدهٔ مدیریت دانشگاه تهران، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹٤/۰٤/۰۳
تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۰۲/۰۳
نویسندهٔ مسئول مقاله: محمدمهدی ذوالفقارزاده
E-mail: zolfaghar@ut.ac.ir

مقدمه

طی دو دههٔ گذشته رقابت در صنعت بانکداری به‌طور شایان توجهی افزایش یافته است. با توجه به افزایش سطح رقابت، به‌دست‌آوردن و حفظ مشتریان اهمیت بسیاری دارد. امروزه گزینه‌های زیادی برای انتخاب نوع خدمات و ارائه‌دهندهٔ آن پیش رو مشتریان قرار دارد. مدیران بانک‌ها باید آگاه باشند که اگر به هر مشتری به‌طور خاص توجه نکنند، مشتری ممکن است وفاداری خود را به بانک از دست بدهد و بانک دیگری را برای دریافت خدمات انتخاب کند. از سوی دیگر، حفظ مشتری به‌مراتب کم‌هزینه‌تر از جذب مشتری جدید است. با توجه به اینکه وفاداری مشتریان نزد بانک‌ها اهمیت بسیاری دارد، بانک‌ها باید برای جذب و نگاه‌داشتن مشتریان از راهبرد مناسبی استفاده کنند و با هر یک از مشتریان بر اساس الگوی رفتاری خاص آن مشتری عمل نمایند. راهبرد کنونی بسیاری از بانک‌ها، ارائهٔ محصول مناسب به مشتری در موقعیت و زمان مناسب است؛ اما در عمل اجرای عملی و اثربخش این اهداف چندان ساده نیست. از سوی دیگر در عصر انفجار اطلاعات، سازمان‌ها هر روزه با حجم زیادی از داده روبه‌رو هستند. در این میان، مسئلهٔ اساسی سازمان‌ها استخراج اطلاعات مفید از داده‌ها و تبدیل آنها به نتایج عملی است؛ به‌ویژه بانک‌ها به‌دلیل حجم گستردهٔ داده‌های موجود به پردازش و استخراج دانش نهفته در داده‌ها نیاز دارند که توجه ویژه به داده‌کاوی[1] و روش‌های کشف دانش را می‌طلبد. بنابراین می‌توان سامانه‌های مشتری‌محور و امن را با استخدام تکنیک‌های داده‌کاوی در بانکداری الکترونیک پیاده‌سازی کرد؛ این کار به خلق مزیت رقابتی پایدارتر و غیرقابل الگوبرداری منجر خواهد شد (زین‌العابدین، مهدوی و خان‌بابایی، ۱۳۹۱).

روش مناسب برای بهبود ارتباط با مشتریان، ارائهٔ پیشنهادهای خدماتی به آنان در زمان و موقعیت مناسب است. برای ارائهٔ پیشنهاد مناسب به کاربر، باید دانش سطح بالایی از داده‌ها استخراج شود تا بتوان بر اساس این دانش دربارهٔ علایق مشتریان و پیشنهاد سرویس به ایشان تصمیم‌گیری کرد. این امر از مهم‌ترین چالش‌ها و مسائل سیستم‌های پیشنهاددهنده یا توصیه‌گر[2] در بانکداری الکترونیک به‌شمار می‌رود که برای حل آن می‌توان از تکنیک‌های داده‌کاوی بهره برد. داده‌کاوی شامل دو رویکرد رایج خوشه‌بندی[3] و دسته‌بندی[4] است.

عموم مطالعات حوزهٔ سیستم‌های پیشنهاددهندهٔ بانکی، حاکی از به‌کارگیری رویکرد خوشه‌بندی است (گالگو، هیوکاس و رودریگز، ۲۰۱۲ و تقوا، برمکان و توفانی، ۲۰۱۱).

---

1. Data mining
2. Recommender system
3. Clustering
4. Classification

خوشه‌بندی با تمام اعضای یک خوشه رفتار یکسانی دارد و به هر کاربر بر اساس شباهت میان ترجیحاتی که بین اعضای خوشه برقرار است، پیشنهادهایی می‌دهد. هرچند روش‌های خوشه‌بندی کاربرد و مزیت‌های زیادی دارند؛ مزیت و برتری ویژهٔ تکنیک‌های دسته‌بندی در این است که می‌توانند با هر مشتری به‌صورت منحصربه‌فرد برخورد کنند. در این روش پیشینهٔ رفتاری و علایق مشتری مبنایی مهم برای تصمیم‌گیری دربارهٔ چگونگی رفتار با مشتری قرار می‌گیرد و از این جهت می‌تواند برای ارائهٔ پیشنهادهای مطلوب به آنان بسیار مناسب باشد. پژوهش حاضر از این مزیتِ روش‌های دسته‌بندی در فنون داده‌کاوی بهره برده است تا به‌وسیلهٔ آن الگویی برای بهبود سیستم‌های پیشنهاددهندهٔ بانکی به مشتریان ارائه کند و کیفیت پیشنهادها را ارتقا دهد. از این رو، پژوهش حاضر بر آن است که الگویی را برای ارائهٔ پیشنهاد خدمات بانکی مناسب با دقت و عملکرد زیاد بر اساس شناخت و دسته‌بندی پیشینهٔ رفتاری مشتریان به آنها ارائه نماید. بدین منظور، سؤال اصلی پژوهش این‌گونه مطرح می‌شود: پیشنهاد خدمات بانکی مناسب به مشتریان بانک با توجه به تکنیک‌های دسته‌بندی بر اساس پیشینهٔ رفتاری و مشخصات جامعه‌شناختی آنها چیست؟ درواقع این پژوهش سعی در کشف چگونگی دسته‌بندی مشتریان بانک بر اساس تکنیک‌های داده‌کاوی و چیستی پیشنهاد خدمات بانکی مناسب و بهترین روش دسته‌بندی برای این پیشنهاد دارد. در این نوشتار ابتدا مختصری به پیشینهٔ پژوهش و تعریف مفاهیم اصلی مرتبط با داده‌کاوی و تکنیک‌های آن پرداخته می‌شود، سپس در مرحلهٔ اول فرایند پژوهش، داده‌ها پیش‌پردازش می‌شوند تا به‌صورت مطلوب درآیند. در گام بعد، چهار مدل دسته‌بندی بیز ساده[1]، k- نزدیک‌ترین همسایه[2]، پرسپترون چندلایه[3] و ماشین بردار پشتیبان[4] با استفاده از داده‌های پیش‌پردازش‌شده آموزش داده می‌شوند و در نهایت با ارزیابی روش‌های دسته‌بندی برای هر یک از سرویس‌های بانکی، مدل نهایی سیستم پیشنهاددهنده ارائه می‌شود.

پیشینهٔ نظری پژوهش

داده‌کاوی یکی از مراحل فرایند کشف دانش است که به کشف دانش و الگوهای معنادار از مجموعهٔ داده با حجم زیاد می‌پردازد. به بیان دیگر، هدف داده‌کاوی استخراج دانش از مجموعهٔ داده است (هان، کامبر و پی، ۲۰۱۲). در عصر حاضر با وجود پایگاه داده‌های بسیار بزرگ از اطلاعات رفتاری و فردی مشتریان، فرایند داده‌کاوی می‌تواند مدیران را در پیشبرد هر چه بهتر

1. Naïve Bayes
2. K-Nearest Neighbor (KNN)
3. Multilayer Perceptron
4. Support vector machine

ارتباط با مشتریان یاری نماید. سیستم‌های پیشنهاددهنده یکی از سیستم‌هایی هستند که می‌توانند با استخدام تکنیک‌های داده‌کاوی و روش‌های هوشمند دیگر، سازمان‌ها را در این زمینه یاری رسانند. سیستم پیشنهاددهنده سیستمی است که با استفاده از تحلیل رفتار کاربران، اطلاعات مفیدی را به کاربران توصیه می‌کند یا راهبردهایی را که کاربران برای رسیدن به اهدافشان به‌کار می‌گیرند، پیشنهاد می‌نماید (کنگاس، ۲۰۰۲). هدف اصلی سیستم‌های پیشنهاددهنده، فراهم‌آوردن ابزاری است که بتواند به کاربران در یافتن سریع و مناسب اطلاعات و نیازهایش یاری دهد. بسیاری از فروشگاه‌های اینترنتی از این سیستم‌ها استفاده می‌کنند؛ ازجمله پیشنهاد کتاب، CD و سایر کالاها در Amazon و Ebay، دوست در Facebook، فیلم در Netflix. دو تکنیک اصلی برای طراحی سیستم‌های پیشنهاددهنده وجود دارد که فیلتر همبستگی[1] و فیلتر مبتنی بر محتوا[2] نام دارند (کیم، یام، سونگ و کیم، ۲۰۰۵؛ مارتین گررو، لیسبوا، سوریا ـ الیواس، پالوماریس و بالاگر، ۲۰۰۷).

در مدل فیلتر همبستگی، بر اساس شباهت میان ترجیحاتی که بین گروه‌های مختلف مشتریان وجود دارد، محصول پیشنهاد می‌شود. در این روش عموماً شبیه‌ترین مشتریان به مشتری جاری بر اساس یک معیار شباهت محاسبه می‌شود؛ سپس بر اساس ترجیحات مشتریان شبیه به مشتری مدنظر، پیشنهاد داده می‌شود (کیم و همکاران، ۲۰۰۵ و مارتین ـ گررو و همکاران، ۲۰۰۷). دو تکنیک در فیلتر همبستگی وجود دارد.

۱. فیلتر همبستگی مبتنی بر کاربر[3]: در این روش شباهت میان کاربران بر اساس الگوی ترجیحاتشان محاسبه می‌شود؛ سپس پیشنهاد به کاربر بر اساس کاربران مشابه به او ارائه می‌شود.

۲. فیلتر همبستگی مبتنی بر آیتم[4]: در این روش به‌جای محاسبهٔ شباهت میان کاربران، شباهت میان آیتم‌های مورد توجه کاربران محاسبه می‌شود. برای مثال اگر گروهی آیتم A و B را خرید کرده باشند، آیتم‌های A و B به یکدیگر شباهت دارند و پیشنهاد نهایی بر اساس شباهت بین آیتم‌های خریداری‌شده ارائه می‌شود (هوروزو، ناراسیمهان، واسودوان، ۲۰۰۶).

فیلتر مبتنی بر محتوا با استفاده از بررسی روند خریدهای گذشتهٔ مشتریان و ترجیحات گذشتهٔ آنها، محصولات مورد علاقهٔ مشتریان را به آنها پیشنهاد می‌دهد (کیم و همکاران، ۲۰۰۵ و مارتین ـ گررو و همکاران، ۲۰۰۷). پژوهش حاضر از جنبه‌ای زیرمجموعه این دسته است.

---

1. Collaborative Filtering(CF)
2. Content based Filtering (CBF)
3. User-based collaborative filtering
4. Item-based collaborative filtering

روش‌های ترکیبی

برخی از سیستم‌های پیشنهاددهنده برای کاهش محدودیت‌های روش‌های قبلی از ترکیب دو یا چند روش دیگر استفاده می‌کنند. بر این اساس پژوهش‌های متعددی صورت گرفته است. شبیب و همکارانش سیستم پیشنهاددهنده‌ای برای پیشنهاد فیلم ارائه کردند؛ به‌نحوی که اولویت‌های مشابه کاربران را با استفاده از تکنیک‌های خوشه‌بندی شناسایی می‌کرد و از رفتار مشتری و موقعیت کاربر برای شخصی‌سازی بهره می‌برد (شبیب، نعمت‌بخش و قهرمانی، ۱۳۸۷). گالگو و هیوکاس با خوشه‌بندی مشتریان بر اساس اطلاعات پروفایل و تراکنش‌هایی که مشتری انجام داده است، به پیشنهاد محصول به کاربران خوشه با استفاده از سیستم موبایل بانک پرداختند (گالگو و هیوکاس، ۲۰۱۲). کنو و همکاران، نیز با اعمال فیلترینگ ترکیبی روی داده‌ها بر اساس فرکانس بازدید کالا، نوعی سیستم پیشنهاددهندهٔ ساده ارائه داد (کنو، لویولا، پریرا، بالوین، ۲۰۱۱). چن و فانگ نیز به بررسی مفهوم اعتماد بین افراد برای استفاده از آن در فیلترینگ کاربران در شبکهٔ اجتماعی فیس‌بوک پرداختند (چن و فانگ، ۲۰۱۰). هانگ و همکاران، گراف دوبخشی کاربر- محصولی را برای ارائهٔ نوعی سیستم پیشنهاددهنده تشریح کردند و درنهایت، جواب‌دهی فیلتر همبستگی را برای مسئله خود مناسب دانستند (هانگ، زینگ و چن، ۲۰۰۷). سو و همکارانش برای ارائهٔ نوعی سیستم پیشنهاددهنده، از ترکیب دو روش مبتنی بر ساختار شبکه و فیلتر همبستگی استفاده کردند (هسو، کینگ، پارادسی، پیدیماری و ونینگر، ۲۰۰۶).

همان‌طور که بیان شد بسیاری از محققان، به‌خصوص در کاربرد بانکداری الکترونیک، از روش‌های خوشه‌بندی بهره برده‌اند. هرچند روش‌های خوشه‌بندی مزایای فراوانی دارند، در این روش‌ها به هر مشتری بر اساس شباهت میان ترجیحات خوشه‌ای که در آن قرار دارد، پیشنهاد می‌شود؛ اما در مواردی، بهتر و دقیق‌تر است که به مشتری بر اساس تحلیل پیشینهٔ رفتاری و تراکنش‌های خود او پیشنهاد داده شود که لزوماً ارتباطی به رفتار سایر مشتریان ندارد. با استفاده از تکنیک‌های دسته‌بندی می‌توان به این مزیت دست یافت که در پژوهش حاضر برای پیشنهادهای بانکی به مشتریان، از این تکنیک‌ها استفاده شده است.

پیشینهٔ تجربی

در سال‌های اخیر نسخه‌های متعددی از سیستم‌های پیشنهاددهنده برای پیشنهاد مفاهیم مختلف در سیستم‌های تجارت الکترونیک و شبکه‌های اجتماعی ارائه شده است. از جمله مفاهیمی که محققان بیشتر به آنها توجه کرده‌اند، پیشنهادهای خبر (کامپن و بیلیکاوا، ۲۰۱۰)، دوست (یین، هانگ و دیویسون، ۲۰۱۱) و محتوا (وانگ، زو، یانگ، لی و وو، ۲۰۱۳) است. با وجود این،

به‌دلایل مختلفی چون دسترسی‌نداشتن به مجموعهٔ دادهٔ مناسب، کار زیادی در حوزهٔ پیشنهاد خدمات بانکی انجام نشده است.

مطالعهٔ آسوشه، باقرپور و یحیی‌پور (۲۰۰۸) از اولین مطالعاتی است که به سیستم‌های پیشنهاددهندهٔ بانکی توجه کرده است. ایشان مدل‌های پذیرش مختلف را با یکدیگر مقایسه کردند و پارامترهای مهم مدل پذیرش مناسبی که بتواند نتایج استفاده از سیستم‌های پیشنهاددهنده را گارانتی کند، استخراج نمودند. سپس بر اساس مدل پذیرش فناوری (TAM) مدلی برای تطبیق با سیستم پیشنهاددهندهٔ بانکی ارائه دادند.

شکوری، امیری، موسی‌خانی و خدنگی (۱۳۹۵) با ترکیب خوشه‌بندی و قوانین وابستگی، نسخه‌ای از سیستم پیشنهاددهنده برای پیشنهاد خدمات بانکی ارائه کردند. در این کار ابتدا با استفاده از شبکه‌های خودسازمان‌دهنده، مشتریان به خوشه‌های متشکل از افراد با متغیرهای رفتاری مشابه تقسیم شدند. سپس با استفاده از قوانین وابستگی، رابطهٔ میان سرویس‌های بانکی برای خوشه‌های مختلف مشتریان شناسایی شد. در انتها با تحلیل شکاف سرویس‌های مورد علاقهٔ هر خوشه و سرویس‌های استفاده‌شدهٔ هر مشتری، سرویس مناسب شناسایی و پیشنهاد شد.

رادفر، نظافتی و یوسفی اصل (۱۳۹۳) برای طبقه‌بندی مشتریان اینترنت‌بانک به‌کمک الگوریتم‌های داده‌کاوی، از درخت تصمیم برای طبقه‌بندی مشتریان بانک سینا بر اساس میزان استفادهٔ ایشان از سرویس اینترنت‌بانک استفاده کردند. در این پژوهش، ویژگی‌های ورودی درخت تصمیم، سن، سطح تحصیلات، جنسیت، تأهل، شغل، محل سکونت و برچسب خروجی (بالا، متوسط و پایین) مبتنی بر تعداد و نوع عملیات بانکی در نظر گرفته شد. در نهایت از روی طبقه‌بندی به‌دست‌آمده قوانینی استخراج شد که توانست به مدیران بانک‌ها برای مدیریت ارتباط با مشتری کمک کند.

کریمی علویجه، عسکری و پاراسایت (۱۳۹۳) نیز در مقاله‌ای با عنوان «فروشگاه اینترنتی هوشمند: سیستم پیشنهاددهندهٔ مبتنی بر تحلیل رفتار کاربران»، برای ارائهٔ پیشنهاد، به ترکیب فیلترهای همبستگی و خوشه‌بندی کاربران بر اساس ویژگی‌های جمعیت‌شناختی اقدام کردند. آنها از تاریخچهٔ امتیازهای کاربران به کالاهای مشاهده‌شده برای پیش‌بینی امتیاز کالاهای مشاهده‌نشده استفاده کردند و برای ارائهٔ پیشنهاد به کاربران تازه‌وارد، از خوشه‌بندی و میانگین وزن‌دار امتیازهای اعضای خوشه بهره بردند. نتایج پیاده‌سازی‌ها در این مقاله نشان می‌دهد استفاده از اطلاعات جمعیت‌شناختی در کنار تاریخچهٔ امتیاز کاربران، نتایج فیلتر همبستگی در ارائهٔ پیشنهاد کالا را بهبود می‌دهد.

در پژوهشی دیگر، رسولی و مانیان (۱۳۹۱) با هدف طراحی سیستم استنتاج فازی برای انتخاب خدمات بانکداری الکترونیک در بانک سپه، پس از بررسی عوامل مؤثر بر انتخاب خدمات

الکترونیک، بر مبنای تعریف توابع عضویت (مثلثی و ذوزنقه‌ای)، نوعی سیستم استنتاج فازی برای سنجش میزان رضایت مشتریان بانک‌ها ارائه دادند. این سیستم توسط تعدادی از مشتریان بانک سپه مورد استفاده قرار گرفت. بر اساس نتایج، میزان رضایت مشتریان از «کیفیت خدمات ارائه‌شده و اعتماد به آنها» و «سهولت استفاده از خدمات» خوب و میزان رضایت از «میزان پاسخگویی ارائه‌دهندهٔ خدمات» و «کیفیت دسترسی به اطلاعات» متوسط ارزیابی شد. بر این اساس قوت‌ها و ضعف‌های بانک سپه به‌دست آمد که این بانک می‌تواند با توجه به آنها، رضایت مشتریان فعلی و قابلیت ایجاد فرصت برای جذب مشتریان جدید را افزایش دهد.

روش‌شناسی پژوهش

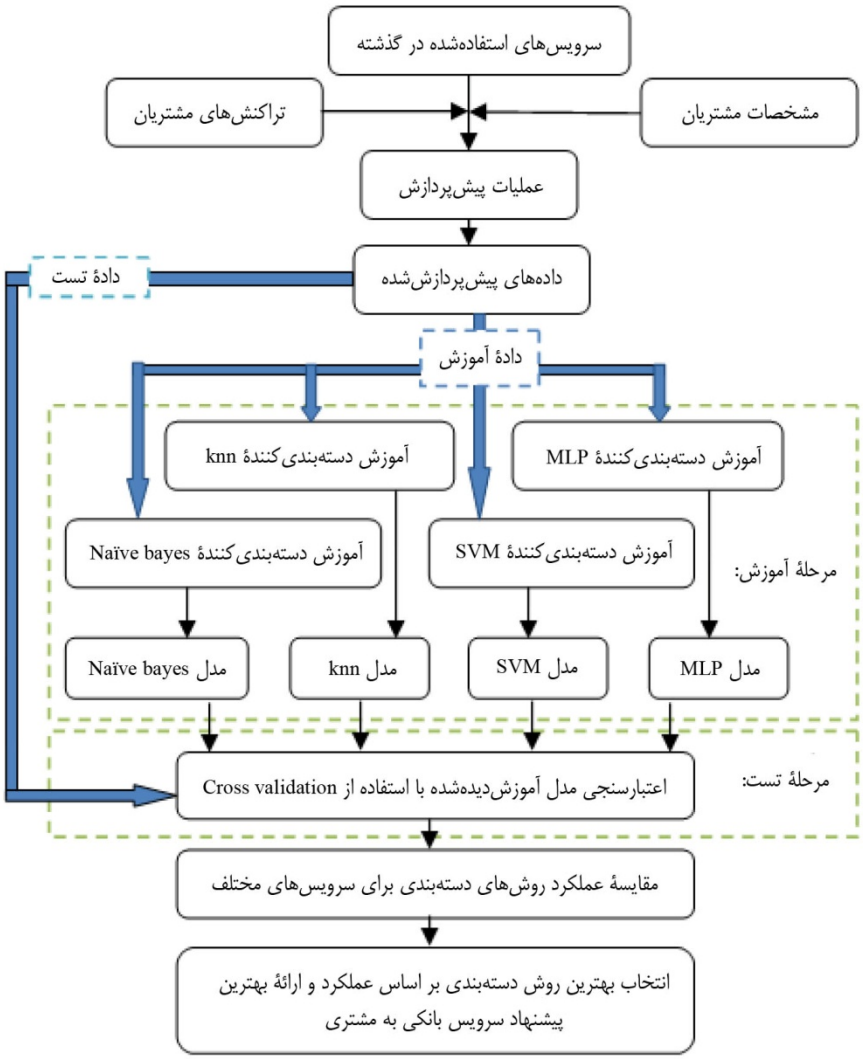
این پژوهش به‌دنبال ارائهٔ مدلی به‌منظور بهبود سیستم‌های پیشنهاددهندهٔ بانکی برای پیش‌بینی علایق مشتریان بر اساس داده‌های بانکی است. این داده‌ها شامل برخی ویژگی‌های جمعیت‌شناختی، تراکنش‌های مالی مشتریان و سرویس‌های استفاده‌شدهٔ آنها درگذشته است. ابتدا داده‌های خام با استفاده از فنون آماری، تجزیه‌وتحلیل شده و در فرایند پیش‌پردازش پاک‌سازی می‌شوند. در ادامهٔ فرایند پژوهش، چهار روش دسته‌بندی‌کنندهٔ MLP، SVM، KNN و بیز ساده با استفاده از داده‌های پیش‌پردازش‌شده به‌طور مجزا آموزش داده می‌شوند و سپس تحت آزمون قرار می‌گیرند. برای استقلال نتیجهٔ اعتبارسنجی از نحوهٔ انتخاب مجموعهٔ آموزش و آزمون، از اعتبارسنجی ضربدری ده‌تایی استفاده می‌شود. با مقایسهٔ نتایج به‌دست‌آمده بر اساس معیارهای دقت مختلف، عملکرد مدل نهایی برای پیشنهاد سرویس‌های مختلف به مشتریان بر اساس معیارهای میزان دقت به‌دست‌آمده و سرعت آموزش محاسبه می‌گردد. مدل نهایی ترکیبی از مدل‌ها با بهترین عملکرد برای توصیهٔ سرویس‌های بانکی به مشتری است. شایان ذکر اینکه برای آموزش دسته‌بندی‌کننده و اعتبارسنجی آنها از نرم‌افزار رپیدماینر[1] استفاده شده است. رپیدماینر نرم‌افزار تخصصی در حوزه‌های مختلف داده‌کاوی از جمله دسته‌بندی، خوشه‌بندی، رگرسیون و قوانین وابستگی است. این پژوهش به لحاظ هدف کاربردی است و از حیث روش جمع‌آوری اطلاعات پیمایشی محسوب می‌شود.

فرایند کلی

این پژوهش با هدف شناسایی سرویس‌های مورد علاقهٔ مشتریان و ارائهٔ پیشنهاد مناسب به آنها طی فرایند پنج‌مرحله‌ای اجرا شده است که مراحل آن در شکل ۱ مشاهده می‌شود.

---

1. Rapidminer



شکل ۱. فرایند کلی پژوهش

همان‌طور که در شکل ۱ مشاهده می‌شود، مراحل پژوهش عبارت است از: ۱. پیش‌پردازش داده‌ها؛ ۲. آموزش مدل؛ ۳. آزمون و اعتبارسنجی مدل آموزش‌دیده؛ ۴. مقایسهٔ عملکرد روش‌های دسته‌بندی‌شده و ۵. انتخاب بهترین روش برای پیشنهاد سرویس بانکی مناسب به مشتریان که در ادامه به توضیح بیشتر این مراحل پرداخته می‌شود.

پیش‌پردازش داده‌ها

۱. نرمال‌کردن داده‌ها

داده با مقدار بزرگ می‌تواند روش دسته‌بندی را به سمت خود متمایل کند یا تأثیر داده‌های با مقادیر کوچک را کاهش دهد؛ بنابراین بهتر است قبل از اعمال روش‌های داده‌کاوی، خصوصیات مختلف نرمال شوند. برای نرمال‌سازی داده‌ها[1] از رابطهٔ ۱ استفاده می‌شود؛ این رابطه داده‌ها را بین ۰ و ۱ نرمال می‌کند.

$$X_{normalized} = (\frac{X_i}{X_{max} - X_{min}}) \quad \text{رابطهٔ ۱)}$$

۲. تحلیل همبستگی

برای محاسبهٔ همبستگی متغیرها از ضریب همبستگی پیرسون استفاده شد. جدول ۱ همبستگی بین برخی از متغیرهای پژوهش را نمایش می‌دهد.

جدول ۱. همبستگی بین متغیرهای پژوهش

متغیرها	موبایل	پرداخت تلفنی	انتقال تلفنی	تلفن	خودپرداز	باجه	تعداد خرید	فاصلهٔ زمانی	میانگین تراکنش	تحصیلات	جنسیت
جنسیت	۰	-۰/۰۱	۰/۰۱	۰/۰۲	۰	۰	۰	۰	-۰/۰۱	۰/۰۱	۱
تحصیلات	۰/۶۷	۰/۲۳	۰/۱۶	۰/۳۶	۰/۳۱	۰/۵۵	۰/۵۵	۰/۱۷	۰/۳۱	۱	
میانگین تراکنش	۰/۲۸	۰/۱۸	۰/۱۱	۰/۳۷	۰/۳۶	۰/۴۳	۰/۵۱	۰/۰۱	۱		
فاصلهٔ زمانی	۰/۱۱	۰/۰۵	۰/۰۱	۰/۱۶	۰/۱۴	۰/۴۵	۰	۱			
تعداد خرید	۰/۵۷	۰/۳۴	۰/۱۷	۰/۷۲	۰/۶۸	۰/۸۳	۱				
باجه	۰/۵۴	۰/۳۱	۰/۱۵	۰/۶۸	۰/۶۳	۱					
خودپرداز	۰/۳۱	۰/۳۸	۰/۱۴	۰/۴	۱						
تلفن	۰/۳۳	۰/۲۱	۰/۰۴	۱							
انتقال تلفنی	۰/۰۶	۰/۰۶	۱								
پرداخت تلفنی	۰/۱۸	۱									
موبایل	۱										

1. Normalization

با مطالعهٔ همبستگی بین متغیرهای مختلف، می‌توان به نتایج جالبی دست یافت. برای مثال همان‌طور که در جدول ۲ مشاهده می‌شود، متغیر جنسیت[1] همبستگی ضعیفی با تمام متغیرهای دیگر دارد. این نشان می‌دهد جنسیت مشتری تأثیر محسوسی بر میزان تراکنش‌های مشتریان یا نوع سرویس بانکی استفاده‌شده ندارد؛ از این رو با حذف متغیر جنسیت، می‌توان پیچیدگی مدل را کاهش داد. این کار سرعت آموزش و آزمایش مدل دسته‌بندی را افزایش می‌دهد. همچنین دیده می‌شود که ویژگی میانگین تراکنش بیشترین همبستگی را با باجه و دستگاه خودپرداز دارد؛ بر این اساس می‌توان نتیجه گرفت مشتریانی که تراکنش‌های زیادی دارند، به‌طور عمده از باجه یا سیستم خودپرداز استفاده می‌کنند و تراکنش‌های زیاد را با استفاده از اینترنت و تلفن انجام نمی‌دهند. همچنین می‌توان نتیجه گرفت که مشتریان به سرویس اینترنت‌بانک، تلفن‌بانک یا موبایل‌بانک ارائه‌شده اطمینان کمی دارند؛ از این رو بانک باید تلاش کند با فرهنگ‌سازی اطمینان مشتریان به این سرویس‌ها را افزایش دهد. میزان تحصیلات ویژگی‌ای است که همبستگی زیادی با ویژگی‌های دیگر دارد. به بیان دیگر، هر چه تحصیلات بالاتر باشد، میزان تراکنش زیادتر و استفاده از سرویس‌های مختلف بانکی بیشتر خواهد بود. از سوی دیگر، تحصیلات بیشترین همبستگی را با سرویس موبایل دارد (۰/۶۷). این موضوع را می‌توان به این صورت تحلیل کرد که افراد با تحصیلات بالاتر، نسبت به سایر افراد بیشتر از سرویس موبایل استفاده می‌کنند و افراد کم‌سوادتر از سرویس موبایل کمتر استفاده می‌کنند. با توجه به میزان همبستگی بین خصوصیات مختلف، می‌توان تصمیماتی را دربارهٔ حذف برخی ویژگی‌ها اتخاذ کرد که در بهبود نتایج تأثیرگذار است. از این نتایج می‌توان در ارتقای سیستم پیشنهاددهندهٔ نهایی استفاده کرد؛ به‌طوری که سیستم پیشنهاددهنده علاوه‌بر اطلاع‌رسانی و آموزش مناسب‌تر کاربران در مورد فواید اینترنت‌بانک یا تلفن‌بانک، این سرویس‌ها را با ملاحظات مشخصی به مشتریان پیشنهاد دهد؛ به‌ویژه اینکه استفاده از این دو سرویس به تحصیلات مشتری وابستگی زیادی دارد و می‌توان دسته‌بندی‌کننده را به‌گونه‌ای آموزش داد که ویژگی تحصیلات در پیشنهاد سرویس موبایل و اینترنت تأثیر بیشتری داشته باشد.

۳. پاک‌سازی داده

داده‌ها ممکن است در زمان جمع‌آوری خطاهایی داشته باشند که این موضوع می‌تواند دقت دسته‌بندی‌کننده را تحت تأثیر قرار دهد. از این رو قبل از تجزیه‌وتحلیل داده، آنها را پاک‌سازی می‌کنیم. اصلاح دادهٔ گم‌شده و حذف دادهٔ پرت دو بخش اصلی از فرایند پاک‌سازی داده است که به‌صورت خلاصه در زیر تعریف می‌شود:

---

1. Gender

اصلاح داده گم‌شده: دادهٔ گم‌شده زمانی اتفاق می‌افتد که یک یا چند خصوصیت از یک نمونهٔ داده به‌دلایل مختلف در مرحلهٔ جمع‌آوری داده جاافتاده باشد و مقدار نداشته باشد. برای حل این مشکل راهکارهای مختلفی وجود دارد. در این پژوهش دادهٔ گم‌شده را با مقدار میانگین ویژگی مدنظر جایگزین می‌کنیم.

حذف دادهٔ پرت[1]**:** به نمونهٔ دادهٔ آموزشی که با نمونه‌های آموزشی از کلاس‌های دیگر احاطه شده است، دادهٔ پرت گفته می‌شود. همچنین به داده‌هایی که از نمونهٔ داده‌های دیگر دور بوده و به‌دلیل دور بودن، دسته‌بندی‌کننده را به سمت خود متمایل می‌کند و باعث خطا می‌شود نیز، دادهٔ پرت می‌گویند. در این پژوهش از فاکتور محلی outlier[2] ارائه‌شده توسط برونینگ، کریگل و ساندر (۲۰۰۰)، برای شناسایی دادهٔ پرت استفاده کردیم. برای محاسبهٔ این مقدار از رابطهٔ ۲ استفاده می‌شود.

$$LOF_k(A) = \frac{\sum_{B \in N_k(A)} \frac{LRD(B)}{LRD(A)}}{k}$$ رابطهٔ ۲)

که در آن LRD[3] چگالی دسترسی محلی نمونهٔ داده و $N_k(A)$ مجموعهٔ k نزدیک‌ترین همسایه به نمونهٔ دادهٔ A است. همان‌طور که مشاهده می‌کنید به میزانی که چگالی A بیشتر باشد، مقدار LOF نهایی کمتر می‌شود. زمانی که مقدار LOF نزدیک ۱ باشد، A دادهٔ پرت نیست و هر چه این مقدار بیشتر شود، احتمال پرت‌بودن نمونهٔ دادهٔ A بیشتر خواهد بود. مقدار LRD نیز از رابطهٔ ۳ محاسبه می‌شود.

$$LRD(A) = 1/(\frac{\sum_{B \in N_k(A)} distance(A,B)}{k})$$ رابطهٔ ۳)

$\text{distance}(A, B)$ فاصلهٔ دسترسی‌پذیری A و B بر اساس فاصلهٔ اقلیدسی است. بر این اساس هر چه فاصلهٔ نمونهٔ دادهٔ A از همسایه‌هایش بیشتر باشد، چگالی‌اش کمتر و در نهایت احتمال پرت‌بودن آن بیشتر خواهد بود (برونینگ، کریگل و ساندر، ۲۰۰۰).

٤. کم‌کردن ابعاد داده[4]

ابعاد زیاد داده علاوه‌بر پیچیده‌کردن دسته‌بندی‌کننده، سرعت آموزش و ارزیابی آن را کاهش می‌دهد. همچنین می‌تواند مشکلاتی از جمله بیش‌برازش[5] ایجاد کند. بنابراین برای افزایش

1. Outlier Removal
2. Local Outlier Factor (LOF)
3. Local Reachability Density
4. Dimension reduction
5. Over fitting

قابلیت تعمیم دسته‌بندی‌کننده، بهتر است تعداد ویژگی‌های دادهٔ ورودی را با استفاده از روش‌های کاهش بُعد، کاهش دهیم.

یکی از روش‌های کاهش بُعد، بهرهٔ اطلاعات است. به زبان ساده، به میزان کاهش آنتروپی با داشتن دانش دربارهٔ یک ویژگی را بهرهٔ اطلاعات می‌گویند که به‌صورت زیر محاسبه می‌شود:

$$IG(P,a) = E(P) - E(P|a)$$ رابطهٔ ۴)

که $E(P|a)$ آنتروپی مجموعهٔ داده با تقسیم داده‌ها بر اساس ویژگی a است. همچنین آنتروپی از رابطهٔ ۵ به‌دست می‌آید.

$$E(P) = -\sum_{i=1}^{n} P_i \log(P_i)$$ رابطهٔ ۵)

که P_i نسبت داده از دسته i به کل داده‌هاست. با استفاده از تکنیک بهرهٔ اطلاعات اهمیت ویژگی‌های مختلف برای پیش‌بینی سرویس‌های بانکی مختلف محاسبه شد. در هر چهار سرویس وزن ویژگی جنسیت صفر بود؛ به این معنا که در نظر گرفتن این ویژگی تأثیر زیادی بر پیش‌بینی و پیشنهاد سرویس‌های مختلف بانکی ندارد. بنابراین متغیر جنسیت از مجموعه حذف شد. دربارهٔ سایر ویژگی‌ها بر اساس نوع سرویس پیشنهادی تصمیمات مشابهی اتخاذ شد و از مجموعه حذف شدند. جدول ۲ اهمیت ویژگی‌های مختلف در پیش‌بینی سرویس اینترنت را نشان می‌دهد. به‌دلیل طولانی‌شدن، از آوردن جدول مربوط به سرویس‌های دیگر پرهیز شده است.

جدول ۲. اهمیت ویژگی‌های مختلف در پیش‌بینی و پیشنهاد سرویس اینترنت با تکنیک بهرهٔ اطلاعات

ویژگی	وزن	ویژگی	وزن
جنسیت	۰/۰۰۰	تلفن همراه	۰/۰۴۸
انتقال تلفنی	۰/۰۰۹	اینترنت	۰/۰۵۴
پرداخت خودپرداز	۰/۰۰۹	تلفن	۰/۰۶۸
انتقال خودپرداز	۰/۰۱۲	میانگین تراکنش	۰/۱۰۴
پرداخت تلفنی	۰/۰۱۳	تعداد خرید	۰/۱۵۱
فاصلهٔ زمانی	۰/۰۱۵	باجه	۰/۱۵۷
پایانهٔ فروش	۰/۰۳۸	تحصیلات	۱/۰۰۰
خودپرداز	۰/۰۴۷		

مجموعهٔ داده و جامعهٔ آماری

مجموعهٔ داده استفاده‌شده مختص به ۴۰۰۰ مشتری است که از پایگاه اطلاعاتی بانک اقتصاد نوین در بازهٔ زمانی ۱۳۸۷ تا ۱۳۸۸ جمع‌آوری شده است. ویژگی‌های مختلف مجموعهٔ دادهٔ تحقیق را می‌توان به سه بخش کلی تقسیم کرد: مشخصات مشتریان، تراکنش‌های مالی مشتریان و تعداد استفاده کاربر از سرویس‌های بانکی درگذشته که در جدول ۳ نمایش داده شده‌اند.

جدول ۳. ویژگی‌های مختلف مجموعهٔ داده

نام متغیر	توضیح
جنسیت	جنسیت مشتری
تحصیلات	تحصیلات مشتریان
میانگین تراکنش	میانگین مبالغ تراکنش‌های مشتری در بازهٔ زمانی خاص
فاصلهٔ زمانی	فاصلهٔ میان زمان آخرین تراکنش مشتری تا زمان ارزیابی
تعداد خرید	تعداد خریدهای مشتریان در بازهٔ زمانی خاص
باجه	مراجعهٔ مشتریان به باجه‌های بانک
خودپرداز	استفاده از سیستم‌های خودپرداز برای برداشت پول
اینترنت	استفاده از سیستم حساب اینترنتی برای مشاهدهٔ حساب و درخواست تسهیلات از بانک
تلفن	استفاده از تلفن برای دریافت صورت‌حساب و گردش موجودی
تلفن همراه	استفاده از موبایل برای دریافت صورت‌حساب و گردش موجودی
پایانهٔ فروش	استفاده از سیستم پایانهٔ فروش (POS) برای پرداخت مبالغ خرید
انتقال اینترتی	استفاده از سیستم حساب اینترنتی برای انتقال وجه
خرید اینترنتی	استفاده از سیستم حساب اینترنتی برای خرید
انتقال خودپرداز	استفاده از سیستم‌های خودپرداز برای انتقال وجه
پرداخت خودپرداز	استفاده از سیستم‌های خودپرداز برای پرداخت قبوض
انتقال تلفنی	استفاده از تلفن برای انتقال وجه
پرداخت تلفنی	استفاده از تلفن برای پرداخت قبوض
پرداخت موبایلی	استفاده از موبایل برای پرداخت قبوض

تبدیل مسئلهٔ پیشنهاد سرویس به مسئلهٔ دسته‌بندی

در نگاه نخست، مسئلهٔ پیشنهاد محصول به مشتری ماهیت مسئلهٔ دسته‌بندی را ندارد و همان‌طور که در بخش پیشینهٔ پژوهش نیز مشاهده شد، روش‌های ارائهٔ پیشنهاد عموماً از رویکرد دسته‌بندی استفاده نکرده‌اند. در این پژوهش تلاش شد به مسئلهٔ پیشنهاد سرویس بانکی به‌صورت مسئلهٔ دسته‌بندی نگاه شود؛ به این صورت که سرویسی که می‌خواهد به کاربر پیشنهاد شود، برچسب خروجی (متغیر وابسته) دسته‌بندی‌کننده می‌گیرد. همچنین مشخصات مشتریان، تراکنش‌های مالی آنها و استفاده از سرویس‌های دیگر، ویژگی‌های ورودی (متغیرهای مستقل) در نظر گرفته می‌شوند. از سوی دیگر، برچسب خروجی بر اساس مقدار آستانهٔ مناسب به مقدار دودویی تبدیل شد و درنتیجه پیشنهاد سرویس بانکی به‌صورت مسئلهٔ دسته‌بندی دودویی کاهش یافت. در این مسئله خروجی صفر نشان‌دهندهٔ پیشنهاد ندادن سرویس به مشتری و خروجی یک نشان‌دهندهٔ پیشنهاد سرویس به مشتری است.

یافته‌های پژوهش

در زیر به بررسی مراحل تجزیه‌وتحلیل داده‌ها و در نهایت به جدول‌های یافته‌های پژوهش پرداخته می‌شود.

آموزش و ارزیابی دسته‌بندی‌کننده

همان‌طور که پیش از این مطرح شد، در این پژوهش از چهار دسته‌بندی‌کنندهٔ پرسپترون چندلایه، k- نزدیک‌ترین همسایه، بیز ساده و ماشین بردار پشتیبان برای دسته‌بندی داده‌ها و ارائهٔ پیشنهاد استفاده شده است. ابتدا این چهار روش با توجه به داده‌های پیش‌پردازش‌شده پس از آموزش، تحت آزمون قرار می‌گیرند. برای اعتبارسنجی این روش‌ها، از روش اعتبارسنجی ضربدری ده‌تایی[1] استفاده شد. داده‌ها به ۱۰ قسمت مساوی تقسیم شدند؛ به این صورت که نُه قسمت از داده‌ها برای آموزش مدل و یک قسمت باقی‌مانده برای آزمون مدل استفاده شد. همین کار برای ۱۰ حالت مختلف تقسیم داده به دادهٔ آموزش و آزمون انجام می‌گیرد و بین دقت مدل‌های مختلف میانگین گرفته می‌شود. این کار سبب استقلال نتایج اعتبارسنجی از الگوی توزیع داده ورودی می‌شود. باید توجه داشت که هر چهار روش دسته‌بندی برای سرویس‌های مختلف آموزش داده می‌شود و بهترین مدل برای هریک از سرویس‌ها انتخاب می‌گردد.

---

1. 10-fold cross validation

پیشنهاد سرویس‌های بانکی مختلف

برای پیشنهاد هر سرویس بانکی دسته‌بندی‌کننده‌های مختلف آموزش داده می‌شوند و پس از آزمون و اعتبار سنجی مدل‌های یادگیری‌شده، دقت به‌دست‌آمده از هر روش دسته‌بندی بررسی می‌شود. میزان دقت و عملکرد دسته‌بندی‌کننده‌های دودویی را می‌توان از روش‌های مختلف محاسبه کرد. مهم‌ترین این روش‌ها عبارت‌اند از:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$ رابطهٔ ۶)

$$precision = \frac{TP}{TP + FP}$$ رابطهٔ ۷)

$$recall = \frac{TP}{TP + FN}$$ رابطهٔ ۸)

$$F - measure = 2 \times \frac{precision.recall}{precision + recall}$$ رابطهٔ ۹)

که در آن TP دسته‌های مثبتی است که درست دسته‌بندی‌شده‌اند؛ FP دسته‌های مثبتی است که غلط دسته‌بندی‌شده‌اند؛ FN دسته‌های منفی غلط دسته‌بندی‌شده و TN دسته‌های منفی درست دسته‌بندی‌شده است. مشخص است که تمام این مقادیر در بازهٔ (۰,۱) هستند که هر چه مقدار به‌دست‌آمده بزرگ‌تر باشد، نشان‌دهنده عملکرد بهتر دسته‌بندی‌کننده است.

نمودار ROC نیز نموداری بر اساس ترکیب دو پارامتر recall در محور عمودی و $precision - ۱$ در محور افقی است. AUC مساحت زیر منحنی ROC است که معیار دیگری برای مقایسهٔ عملکرد روش‌های دسته‌بندی است.

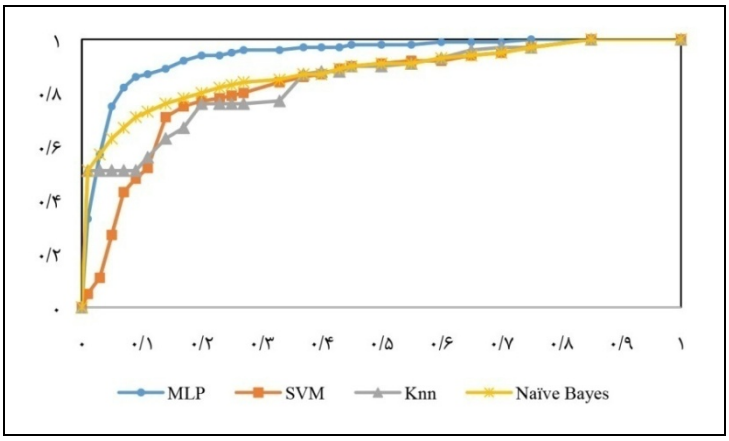
پیشنهاد سرویس اینترنت با استفاده از روش‌های دسته‌بندی

برای پیشنهاد سرویس اینترنت، همان‌طور که بیان شد پس از آموزش دسته‌بندی‌کننده‌های مختلف و اعتبارسنجی مدل‌ها با روش اعتبارسنجی ضربدری ده‌تایی، به نتایج جدول ۴ رسیدیم.

همان‌گونه که در جدول ۴ مشاهده می‌شود، روش MLP دارای بهترین دقت نسبت به سایر روش‌هاست. نمودار ROC روش‌های مختلف، در شکل ۲ نمایش داده شده است.

جدول ٤. دقت دسته‌بندی‌کننده‌های مختلف برای پیشنهاد سرویس اینترنت

	ماشین بردار پشتیبان	پرسپترون چندلایه	۵-نزدیک‌ترین همسایه	بیز ساده
صحت	٪۸۱	٪۸۹	٪۸۰	٪۸۲
دقت	٪۸۶	٪۹۳	٪۸۴	٪۸۸
یادآوری	٪۹۳	٪۹۳	٪۹۱	٪۸۸
سطح زیر نمودار ROC	۰/۸۵	۰/۹۴	۰/۸۵	۰/۸۹
معیار F	۰/۸۹	۰/۹۳	۰/۸۷	۰/۸۸



شکل ۲. نمودار ROC دسته‌بندی‌کننده‌ها مختلف در پیشنهاد سرویس اینترنت

نمودار ROC نیز نشان می‌دهد روش MLP بیشترین سطح زیرنمودار را دارد و این نشان‌دهندهٔ دقت بالای این روش نسبت به سایر روش‌هاست؛ اما نکته‌ای که در این میان وجود دارد زمان و هزینهٔ آموزش هر روش است. از آنجا که زمان نیز برای ما اهمیت دارد، لازم است تأثیر آن سنجیده شود. از این رو به‌منظور انتخاب بهترین مدل برای پیشنهاد سرویس، باید هر دو معیار زمان و دقت با هم سنجیده شود.

از آنجا که تأثیر زمان در این تحقیق کم‌اهمیت‌تر از دقت است، از مدل ۲۰–۸۰ برای ترکیب این دو استفاده شد. از رابطه‌های ۱۰ و ۱۱ برای به‌دست‌آوردن عملکرد نهایی استفاده می‌شود.

$$speed = \left(100 - (time \times 10)\right) \qquad \text{رابطهٔ ۱۰)}$$

$$Performance = \left(\frac{1}{5} \times speed\right) + \left(\frac{4}{5} * accuracy \times 100\right) \qquad \text{رابطهٔ ۱۱)}$$

در نهایت، عملکرد دسته‌بندی‌کننده‌های مختلف با دخالت متغیر زمان (بر حسب دقیقه) به‌دست می‌آید (جدول ۵). همان‌طور که مشاهده می‌شود روش بیز ساده با مقدار عددی ۸۵/۴ درصد، بیشترین عملکرد را دارد؛ از این رو، روش بیز ساده برای پیشنهاد سرویس اینترنت به کاربران انتخاب می‌شود. همین فرایند برای آموزش دسته‌بندی‌های مختلف به‌منظور پیشنهاد سرویس‌های بانکی دیگر نیز تکرار شد. این نتایج در جدول ۶ خلاصه شده است.

جدول ۵. عملکرد دسته‌بندی‌کننده‌ها در پیشنهاد سرویس اینترنت

	بیز ساده	۵-نزدیک‌ترین همسایه	پرسپترون چندلایه	ماشین بردار پشتیبان
دقت	٪۸۱	٪۸۹	٪۸۰	٪۸۲
زمان (دقیقه)	۱	۶	۰/۴	۰/۲
عملکرد	۸۳/۸	۸۵/۲	۸۳/۶	۸۵/۴

جدول ٦. عملکرد دسته‌بندی‌کننده‌ها برای پیشنهاد سرویس‌های بانکی مختلف

	بیز ساده	۵-نزدیک‌ترین همسایه	پرسپترون چندلایه	ماشین بردار پشتیبان
عملکرد دسته‌بندی‌کننده‌ها در سرویس موبایل				
دقت	٪۸۱	٪۸۹	٪۸۰	٪۷۹
زمان (دقیقه)	۱	۶	۰/۴	۰/۲
عملکرد	۸۲/۶	۷۹/۲	۸۳/۳	۸۳/۱
عملکرد دسته‌بندی‌کننده‌ها در سرویس اینترنت برای انتقال وجه و پرداخت صورت‌حساب				
دقت	٪۶۷/۶۲	٪۸۴/۲۸	٪۶۶/۱۰	٪۶۶/۷۵
زمان (دقیقه)	۱	۶	۰/۴	۰/۲
عملکرد	۷۳/۱	۸۱/۴	۷۲/۵	۷۳/۲
عملکرد دسته‌بندی‌کننده‌ها در سرویس تلفن برای پرداخت صورت‌حساب				
دقت	٪۸۷/۷۷	٪۹۸/۲۵	٪۸۶/۸۲	٪۸۱/۹۲
زمان (دقیقه)	۱	۶	۰/۴	۰/۲
عملکرد	۸۹/۲	۹۲/۶	۸۹/۱	۸۵/۳

بر اساس نتایج جدول ۵، برای پیشنهاد سرویس موبایل روش 5-nn با عملکرد ۸۳/۳ درصد، انتقال وجه و پرداخت صورت‌حساب روش MLP با عملکرد ۸۱/۴ درصد و برای انتقال وجه و پرداخت صورت‌حساب با تلفن، روش MLP با عملکرد ۹۲/۶ درصد انتخاب شد.

مدل پیشنهاددهندهٔ نهایی

از آنجا که این پژوهش به‌دنبال مدل پیشنهاددهنده‌ای با دقت زیاد برای ارائهٔ پیشنهاد سرویس‌های بانکی به مشتریان با رویکرد دسته‌بندی است، می‌توان مدل نهایی را بر اساس دقت روش‌های دسته‌بندی مختلف به‌دست آورد. مدل‌های انتخابی و پارامترهای آنها قبل از دخالت‌دادن پارامتر زمان مطابق جدول ۷ است و بر اساس محاسبهٔ عملکرد و دخالت پارامتر زمان مدل‌های بهینهٔ نهایی به‌صورت جدول ۸ تغییر می‌کنند.

جدول ۷. نتایج ارزیابی بر اساس دقت برای پیشنهاد سرویس‌های بانکی

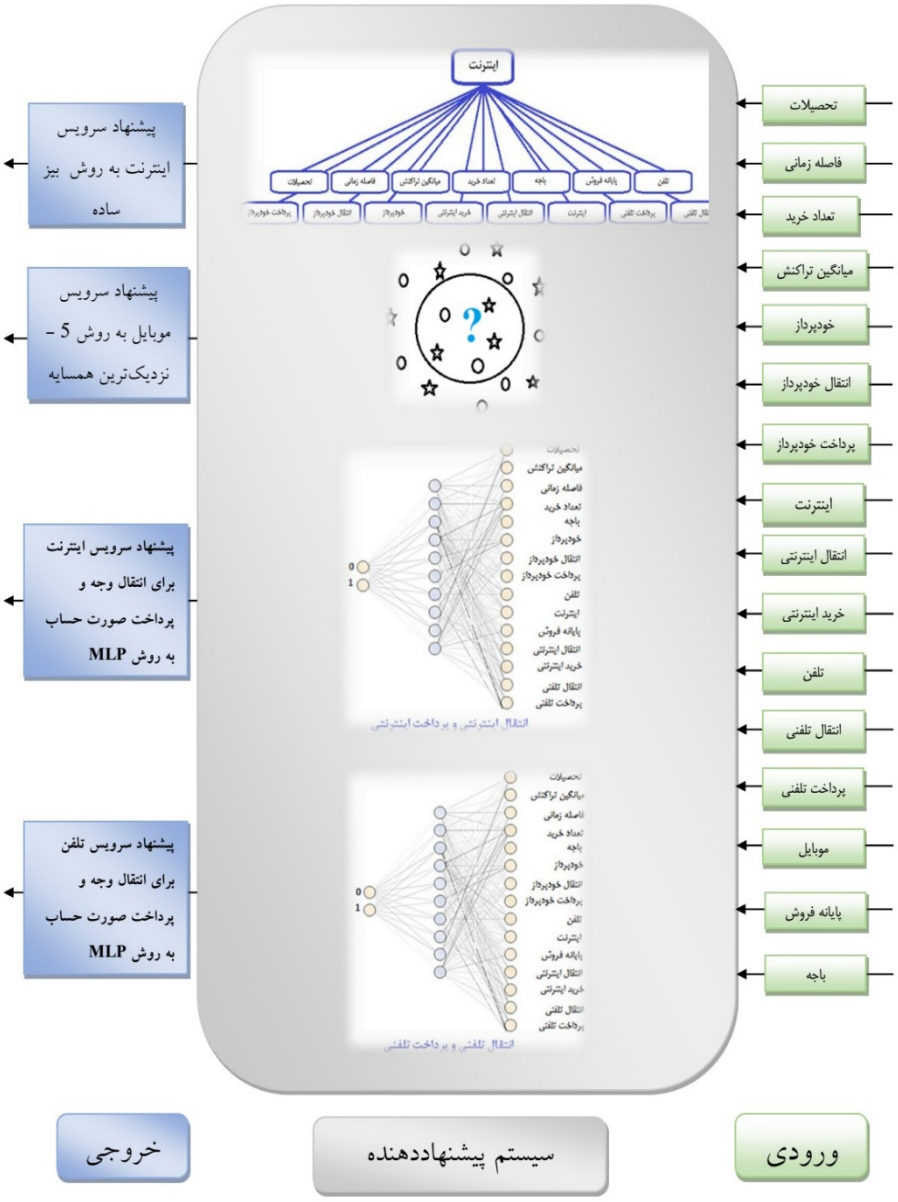
	روش	صحت	دقت	یادآوری	سطح زیر نمودار ROC	معیار F
سرویس اینترنت	MLP	٪۸۹	٪۹۳	٪۹۳	۰/۹۴	۰/۹۳
سرویس موبایل	MLP	٪۸۹	٪۷۹/۴	٪۷۵/۲	۰/۹۴	۰/۷۷
سرویس اینترنت برای انتقال وجه و پرداخت	MLP	٪۸۴/۲۸	٪۸۷/۲۹	٪۸۶/۱۴	۰/۹۲۳	۰/۸۷
سرویس تلفن برای انتقال وجه و پرداخت	MLP	٪۹۸/۲۵	٪۹۹	٪۹۹/۰۱	۰/۹۹۵	۰/۹۹

جدول ۸. نتایج ارزیابی بر اساس عملکرد برای پیشنهاد سرویس‌های بانکی

	روش	صحت	زمان(دقیقه)	عملکرد
سرویس اینترنت	بیز ساده	٪۸۲	۰/۲	۸۵/۴
سرویس موبایل	۵-نزدیک‌ترین همسایه	٪۸۰	۰/۴	۸۳/۳
سرویس اینترنت برای انتقال وجه و پرداخت	پرسپترون چندلایه	٪۸۴/۲۸	۶	۸۱/۴
سرویس تلفن برای انتقال وجه و پرداخت	پرسپترون چندلایه	٪۹۸/۲۵	۶	۹۲/۶

بدین ترتیب ساختار مدل پیشنهاددهندهٔ نهایی در قالب شکل ۳ به‌دست آمد. همان‌طورکه در شکل ۳ مشخص است مدل ارائه‌شده در این پژوهش از بیز ساده برای پیشنهاد سرویس اینترنت و از 5-nn برای پیشنهاد سرویس موبایل استفاده می‌کند. همچنین از MLP نیز برای پیشنهاد سرویس اینترنت و تلفن برای انتقال وجه و پرداخت صورت‌حساب استفاده می‌کند. شایان ذکر است که این مدل می‌تواند روی مجموعهٔ دادهٔ استفاده‌شده به‌خوبی عمل کند، اما برای استفاده

روی مجموعهٔ دادهٔ دیگر باید دوباره مراحل آموزش و اعتبارسنجی اجرا شود تا مدل‌های مناسب برای ارائهٔ سرویس‌های بانکی مجموعهٔ دادهٔ مدنظر شناسایی شوند.



شکل ۳. ساختار نهایی مدل پیشنهاددهنده

نتیجه‌گیری و پیشنهادها

در بانک‌ها پاسخ به این سؤال که هریک از مشتریان به چه سرویس‌هایی نیاز و علاقه دارند، بسیار حائز اهمیت است. در این پژوهش سعی شد مدلی برای پیشنهاد سرویس‌های مختلف بانکی با استفاده از چندین روش دسته‌بندی ارائه شود. با استفادهٔ همزمان از این چند روش، می‌توان از مزیت‌های تک‌تک روش‌ها بهرهٔ جداگانه‌ای برد.

هرچند در نگاه نخست، مسئلهٔ پیشنهاد سرویس‌های بانکی از دسته مسائل دسته‌بندی نیست، به‌دلیل مزیت‌های این روش در برخورد منحصربه‌فرد با هر مشتری، در این پژوهش تلاش شد این مسئله از دید مسئلهٔ دسته‌بندی مطالعه شود؛ به این صورت که سرویس آماده برای پیشنهاد به کاربر، به‌عنوان برچسب خروجی انتخاب شد و بقیهٔ متغیرها مانند داده‌های جمعیت‌شناختی، سرویس‌ها و تراکنش‌های انجام‌شده نیز ویژگی‌های ورودی محسوب شدند. پس از پیش‌پردازش داده‌های مشتریان، روش‌های دسته‌بندی تحت آموزش و آزمایش قرار گرفتند. پس از اعتبارسنجی نتایج، پیشنهاددهندهٔ نهایی از ترکیب سه دسته‌بندی‌کنندهٔ بیز ساده، ۵-نزدیک‌ترین همسایه و MLP به‌دست آمد که در مرحلهٔ ارزیابی دقت بالایی داشت.

ابتدا داده‌های مشتریان پس از عبور از مرحلهٔ پیش‌پردازش، به‌صورت مجزا وارد هر یک از روش‌های دسته‌بندی شد؛ سپس هر چهار روش بیان‌شده آموزش داده شدند و بعد از آن مدل آموزش‌داده شده آزمایش شد. در مرحلهٔ بعد، عملکرد هر روش برای پیشنهاد سرویس مدنظر به مشتریان، بر اساس میزان دقت و زمان آموزش مدل محاسبه گردید. مدل نهایی ترکیبی از روش‌های ارائه‌شده با بهترین عملکرد برای ارائهٔ سرویس بانکی به مشتری است که شامل ارائهٔ پیشنهاد سرویس اینترنت با روش بیز ساده و عملکرد ۸۵/۴ درصد، پیشنهاد سرویس موبایل با روش 5-nn و عملکرد ۸۳/۳ درصد، پیشنهاد سرویس اینترنت برای انتقال وجه و پرداخت صورت‌حساب با روش MLP و عملکرد ۸۱/۴ درصد و پیشنهاد سرویس تلفن برای انتقال وجه و پرداخت صورت‌حساب با روش MLP و عملکرد ۹۲/۶ درصد است.

References

Asosheha, A., Bagherpour, S. & Yahyapour, N. (2008). Extended acceptance models for recommender system adaption, case of retail and banking service in Iran. *WSEAS transactions on business and economics*, 5(5): 189-200.

Breunig, M. M., Kriegel, H. P., Ng, R. T. & Sander, J. (2000). LOF: identifying density-based local outliers. *in ACM Sigmod Record*, 29(2): 93-104.

Chen, W. & Fong, S. (2010). Social network collaborative filtering framework and online trust factors: a case study on Facebook. *The Fifth International Conference on Digital Information Management (ICDIM)*. Thunder Bay, 5-8 July 2010.

Gallego, D., Huecas, G. & Salvachúa Rodríguez, J. (2012). Generating context-aware recommendations using banking data in a mobile recommender system. *The Sixth International Conference on Digital Society,* Valencia, 30 January - 4 February 2012.

Gallego, D. & Huecas, G. (2012). An empirical case of a context-aware mobile recommender system in a banking environment. *Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing*. Vancouver, 26-28 June 2012.

Han, J. & Kamber, M. & Pei, J. (2012). *Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition.* Amsterdam: Elsevier.

Horozov, T., Narasimhan, N. & Vasudevan, V. (2006). Using location for personalized POI recommendations in mobile environments. *International Symposium on Applications and the Internet,* Phoenix, 23-27 January 2006.

Hsu, W. H., King, A. L, Paradesi, M. S., Pydimarri, T. & Weninger, T. (2006). Collaborative and Structural Recommendation of Friends using Weblog-based Social Network Analysis. *in AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. Palo Alto, 27–29 March 2006.

Huang, Z., Zeng, D. D. & Chen, H. (2007) . Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommender Systems. *Management Science*, 53(7): 1146-1164.

Kangas, S. (2002). Collaborative filtering and recommendation systems. *in: VTT information technology.* Espoo: VTT.

Karimi Alavije, M., Askari, S. & Parasite, S. (2015). Intelligent Online Store: User Behavior Analysis based Recommender System. *Journal of Information Technology Management*. 7(2): 385-406.

Kim, Y. S., Yum, B. J., Song, J. & Kim, S. M. (2005). Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications.* 28(1): 381-393.

Kompan, M. & Bieliková, M. (2010). Content-based news recommendation. *International Conference on Electronic commerce and web technologies*

(EC-Web 2010), University of Deusto, Bilbao, 30 August - 3 September 2010.

Konow, R., Tan, W., Loyola, L., Pereira, J. & Baloian, N. (2011). A Visited Item Frequency Based Recommender System: Experimental Evaluation and Scenario Description. *Journal of Universal Computer Science,* 17(14): 2009-2028.

Martín-Guerrero, J. D. & Lisboa, P. J. & Soria-Olivas, E. & Palomares, A. & Balaguer, E. (2007). An approach based on the Adaptive Resonance Theory for analyzing the viability of recommender systems in a citizen Web portal. *Expert Systems with Applications*, 33(3): 743-753.

Radfar, R. & Nezafati, N. & Yousefi Asli, S. (2014). Classification of Internet banking customers using data mining algorithms. *Journal of Information Technology Management*, 6(1): 71-90. (*in Persian*)

Rasoli, H. & Manian, A. (2012). Designing a Fuzzy Inference System for Selecting e-Banking Services (Case Study: Sepah Bank). *Journal of Information Technology Management*, 4(12): 41-64.

Shabib, N. & Nematbakhsh, M. & Ghahramani, F. (2009). Using data mining methods for recognizing customer behavior in mobile commerce. *The Second Iran data mining conference*. (*in Persian*)

Shakouri, H., Amiri, B., Mousakhani, M. & Khadangi, E. (2016). Presenting a model for predicting needed technologies in banks using SOM and ARM. *Knowledge & Technology*, in Press. (*in Persian*)

Taghva, M. R., Bamakan, S. M. H. & Toufani, S. (2011). A data mining method for service marketing: A case study of banking industry. *Management Science Letters,* 1(3): 253-262.

Wang, Z., Sun, L., Zhu, W., Yang, S., Li, H. & Wu, D. (2013). Joint social and content recommendation for user-generated videos in online social network. *IEEE Transactions on Multimedia*, 15(3): 698-709.

Yin, D. & Hong, L. & Davison, B. D. (2011). Structural link analysis and prediction in microblogs. *Proceedings of the 20th ACM international conference on Information and knowledge management*. Glasgow, 24-28 October 2011.

Zeinolabedin, F. & Mahdavi, M. & Khanbabaee, M. (2012). Customers clustering using data mining techniques and RFM model to marketing banking services. *The 4th international conference on Banking Services marketing*. Tehran, 14-15 October 2012. (*in Persian*)