

طبقه‌بندی مشتریان اینترنت بانک با کمک الگوریتم‌های داده‌کاوی

رضا رادفر^۱، نوید نظافتی^۲، سعید یوسفی اصلی^۳

چکیده: طبقه‌بندی مشتریان با استفاده از الگوریتم‌های داده‌کاوی، بانک‌ها را قادر به حفظ و وفاداری مشتریان قدیم و جذب مشتریان جدید خواهد کرد. یکی از روش‌های داده‌کاوی، درخت تصمیم‌گیری است و چنانچه درخت تصمیم مناسبی ساخته شود، می‌توان مشتریان را به‌طور بهینه طبقه‌بندی کرد. در این نوشتار، یک مدل مناسب برای طبقه‌بندی مشتریان بر مبنای بهره‌گیری از خدمات اینترنت‌بانک ارائه شده است. این مدل بر اساس استاندارد CRISP-DM انجام‌گرفته و داده‌های مورد نیاز از پایگاه داده مشتریان اینترنت‌بانک سینا استخراج شده است. در میان سایر درختان تصمیم‌گیری، درخت تصمیم نهایی مبتنی بر معیارهای بهینگی و دقت بوده و براساس دسته‌بندی مشتریان در سه سطح بالا، متوسط و پایین، پیش‌بینی مشتریان جدیدی که متقاضی استفاده از اینترنت‌بانک هستند، شکل می‌گیرد. پژوهش پیش رو از نظر هدف، کاربردی و از نظر گردآوری داده‌ها، پژوهشی اسنادی به‌شمار می‌رود. قوانین استخراج‌شده مربوط به مشتریان، مدیران بانک‌ها را قادر می‌کند تا بر اساس الگوهای کشف‌شده سیاست‌گذاری کنند و درک بهتری از انتظارات کنونی و آتی مشتریان داشته باشند.

واژه‌های کلیدی: استاندارد CRISP-DM، بانکداری الکترونیک، داده‌کاوی، درختان تصمیم‌گیری، طبقه‌بندی.

۱. دانشیار گروه مدیریت و اقتصاد، دانشگاه آزاد اسلامی واحد علوم تحقیقات، تهران، ایران

۲. استادیار گروه مدیریت دولتی، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران

۳. کارشناس ارشد مدیریت فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد الکترونیک، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۲/۰۳/۰۵

تاریخ پذیرش نهایی مقاله: ۱۳۹۲/۰۸/۱۱

نویسنده مسئول مقاله: سعید یوسفی اصلی

E-mail: syousefi2005@gmail.com

مقدمه

همزمان با الکترونیکی شدن تجارت کسب و کار، بانکداری نیز به شکل الکترونیکی درآمد تا بتواند به مشتریان خدمات مالی و بانکی برخط ارائه کند. توسعه و پیشرفت بانکها در زمینه بانکداری الکترونیک که یکی از پیش شرطها و ابزارهای اساسی مورد نیاز تجارت الکترونیک به شمار می رود، بسیار ضروری و حیاتی است، اما آنچه بیش از همه اینها اهمیت دارد، عمومی کردن بانکداری الکترونیک و تلاش در جهت استقبال مشتریان از خدمات الکترونیکی بانکها است؛ چراکه اگر چنین نباشد، هر تلاشی در این زمینه، حتی تجهیز بانکها به پیشرفته ترین و به روزترین سخت افزارها و نرم افزارها، نمی تواند تأثیر زیادی در الکترونیکی کردن تجارت و بانکداری در ایران داشته باشد (کازی و قاضی، ۲۰۱۲). بانکهای سراسر جهان، روزانه حجم زیادی از داده های الکترونیکی را در پایگاه داده متمرکز نگهداری می کنند. برای سازمانها غیر ممکن است که با حجم زیاد داده ها به تجزیه و تحلیل و بازیابی اطلاعات مفید بپردازند (بمبری، ۲۰۱۱). در این راستا می بایستی رفتار مشتریان در زمینه خدمات بانکداری الکترونیک به دقت بررسی شود تا بانکها بتوانند براساس آنها الگوهای مناسب را شناخته و به ترویج بانکداری الکترونیک اقدام کنند.

در گذشته عموماً استخراج اطلاعات مفید از داده های ثبت شده، به صورت دستی و برعهده تحلیل گران بوده است. با توجه به اینکه تجزیه و تحلیل دستی داده ها بسیار کند و گران بوده و هر روز بر پیچیدگی و حجم داده ها افزوده می شد، تحلیل های دستی به سمت تحلیل های غیر مستقیم خودکار و استفاده از روش های رایانه ای حرکت کرده است (تارخ و شریفیان، ۱۳۸۹) و نیاز مبرمی مبنی بر استفاده از فناوری های جدید و ابزارهای خودکار به وجود آمد تا به صورت هوشمند، حجم زیاد داده را به اطلاعات و دانش تبدیل کند. در این شرایط ضروریست از فناوری اطلاعات برای استفاده از این دانش بهره گرفت و داده کاوی^۱، پاسخی مناسب برای استخراج این دارایی است (تقوی فرد، منصوری، ناصرزاده و فراست، ۱۳۸۶).

با توجه به وجود پایگاه داده در بانکها، این پژوهش تلاش می کند با استفاده از داده کاوی که یکی از ابزارهای فناوری اطلاعات است، به طبقه بندی مشتریان اینترنت بانک پرداخته و الگوهای پنهان موجود در داده های تراکنشی موجود را به شکل مجموعه ای از قوانین نمایش دهد.

بیان مسئله

با افزایش روبه رشد استفاده مشتریان از خدمات بانکداری الکترونیک، به ویژه خدمات اینترنت بانک، تعداد تراکنشها با افزایش فزاینده ای روبه رو است و بانکها با حجم عظیمی از

داده‌ها در پایگاه‌های اطلاعاتی خود مواجه‌اند. این امر می‌تواند یک منبع ارزشمند برای کشف دانش و الگوهای رفتاری مشتریان تلقی شود.

بهره‌گیری از داده‌کاوی می‌تواند در این زمینه بسیار راه‌گشا باشد. توجه به ضرورت شناسایی دسته‌های مختلف مشتریان از جنبه‌های گوناگون و شناخت مشتریانی که از خدمات اینترنت بانک برای امور بانکی خود استفاده می‌کنند، بانک‌ها را در سیاست‌گذاری‌های آتی در خصوص مدیریت ارتباط با مشتریان یاری می‌دهد؛ چراکه بخش‌بندی مشتریان به گروه‌های همگن و ارائه خدمات متناسب با نیازهای هر گروه، در مدیریت ارتباط با مشتری اهمیت ویژه‌ای داشته و مشتریان با ادراکات متفاوت، اهمیت متفاوتی برای ویژگی‌های خدمات مشابه قائل هستند (حسینی، بحرینی‌زاده و ضیائی بیده، ۱۳۹۱). بانک‌ها می‌توانند با توجه به این مهم، تبلیغات مؤثر و سیستم‌های تشویقی خاصی برای هر دسته از مشتریان در نظر بگیرند و خدمات سفارشی‌شده‌ای را با توجه به دسته‌های مختلف مشتریان، ارائه دهند. در این پژوهش، مشتریان به‌لحاظ استفاده از خدمات اینترنت بانک با استفاده از رویکرد داده‌کاوی طبقه‌بندی می‌شوند.

هدف از انجام این پژوهش، افزایش توانایی مدیران بانک‌ها برای ارائه بهتر خدمات بانکداری اینترنتی به مشتریان، بر اساس سطوح شناسایی‌شده است تا بر اساس الگوهای کشف‌شده، سیاست‌های آتی خود را در جهت منافع مشتریان و بانک تعدیل کنند. پژوهش پیش رو تلاش می‌کند تا به پرسش‌های زیر پاسخ دهد:

چگونه می‌توان مشتریان بانک را به‌لحاظ سطح استفاده از خدمات اینترنتی در تراکنش‌های بانکداری الکترونیک با کمک الگوریتم‌های داده‌کاوی طبقه‌بندی کرد؟

مهم‌ترین شاخص‌ها در ارزیابی مشتریان اینترنت بانک کدام خصیصه است؟

بهترین درخت تصمیم در طبقه‌بندی مشتریان، تولیدشده از کدام الگوریتم داده‌کاوی است؟

پیشینه پژوهش

پیشینه نظری

بانکداری الکترونیک: با به‌وجودآمدن اینترنت، مفهوم اصلی بانکداری الکترونیک در سال ۱۹۹۱ شکل گرفت، به این معنا که مشتریان بدون حضور در شعبه، امور بانکی خود را با مراجعه به فضای الکترونیکی انجام دهند (هاشمیان، عیسانی، میکائیلی و طباطبائی، ۱۳۹۱).

بانکداری الکترونیک مزایای بسیاری چون، نداشتن محدودیت زمانی و مکانی، دسترسی آسان به اطلاعات، کاهش دادن هزینه خدمات و صرفه‌جویی در زمان مشتریان را به‌ارمغان آورده است که این امر، موجب رشد سریع استفاده از خدمات بانکداری الکترونیک شده است (دیواندری،

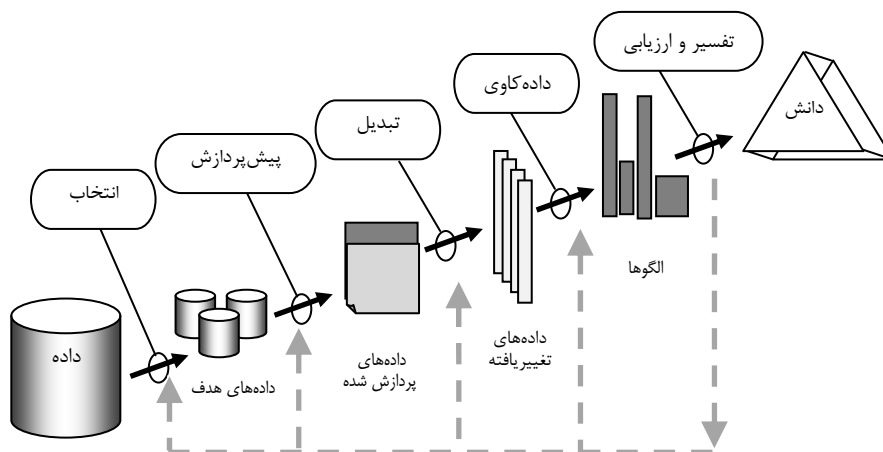
عابدی و ناصرزاده، ۱۳۹۲). استفاده از خدمات بانکداری الکترونیک، یکی از راه‌های کسب مزیت رقابتی برای بانک‌ها به‌شمار می‌رود و موجب به‌وجودآمدن رقابتی تنگاتنگ در این زمینه شده است، در چنین شرایطی سطح انتظار مشتریان برای دریافت این‌گونه خدمات نیز افزایش یافته است (رسولی و مانیان، ۱۳۹۱).

بانکداری الکترونیک مزایایی برای بانک‌ها داشته که مهم‌ترین آنها به‌شرح زیر است:

- متمرکز شدن بر کانال‌های توزیع جدید؛
- ارائه خدمات اصلاح‌شده به مشتریان؛
- استفاده از راهبردهای تجارت الکترونیک (عزیزی سرخنی، اله‌قلی‌زاده آذری و کردلوئی، ۱۳۸۷).

داده‌کاوی و کشف دانش از پایگاه داده: تاریخچه کشف دانش از پایگاه‌های اطلاعاتی قدمت چندانی ندارد و امروزه به داده‌کاوی مشهور است. اصطلاح کشف دانش برای نخستین بار در دهه ۱۹۹۰ مطرح شد و توجه پژوهشگران را به سمت الگوریتم‌های داده‌کاوی معطوف کرد. هدف داده‌کاوی، کشف دانش جدید، معتبر و قابل پیگیری با استفاده از ابزارهای هوش مصنوعی^۱ و آماری در حجم بالایی از داده‌ها است (مریان، سگویا، مناسالواس و فرناندز بیزن، ۲۰۰۹). داده‌کاوی، استخراج یا اقتباس دانش از مجموعه داده‌ها است و به فرایندی گفته می‌شود که دانش را از داده‌ها استخراج می‌کند و این دانش در قالب الگوها و مدل‌ها بیان می‌شود (تقوی‌فرد و نادعلی، ۱۳۹۱). در شکل ۱ مراحل کشف دانش از پایگاه داده‌ها نشان داده شده است (دسپاند و تاکر، ۲۰۱۰)، این مراحل به شرح زیر هستند:

۱. انتخاب داده‌ها: داده‌های مربوط به تجزیه و تحلیل و تصمیم‌گیری از داده‌های دیگر جدا می‌شوند.
۲. پیش‌پردازش اطلاعات: پردازش، تمیز کردن و یکپارچه‌سازی داده‌ها انجام می‌گیرد.
۳. تبدیل داده‌ها: داده‌های انتخاب‌شده به شکل مناسبی برای روش داده‌کاوی تبدیل می‌شوند.
۴. داده‌کاوی: در این مرحله از روش‌های هوشمندانه‌ای برای استخراج الگوهای بالقوه مفید استفاده می‌شود و در مورد این روش‌ها تصمیم‌گیری می‌شود.
۵. تفسیر و ارزیابی: در این مرحله، الگوهای جالب توجه نشان‌دهنده دانش، بر اساس اقدامات انجام‌شده شناخته می‌شوند و دانش کشف‌شده در اختیار کاربر قرار می‌گیرد. در این مرحله استفاده از روش‌های تجسم‌سازی برای کمک به کاربران ضروری است.



شکل ۱. مراحل کشف دانش از پایگاه داده‌ها

منبع: فایاد، پیاتسکی و اسمیت، ۱۹۹۷

دو هدف اصلی داده‌کاوی، پیشگویی و توصیف است (کیس، ۲۰۰۳):

۱. داده‌کاوی پیشگویی‌کننده، مدلی را از سیستم ارائه می‌دهد که این مدل را مجموعه‌ای از داده‌های مشخص، پیش‌بینی می‌کنند. هدف کلی آن ایجاد الگویی برای طبقه‌بندی، پیش‌بینی و تخمین داده‌ها است.
 ۲. داده‌کاوی توصیفی، اطلاعات جدید و غیر بدیهی را بر اساس مجموعه‌ای از داده‌های موجود ارائه می‌دهد و هدف کلی آن درک و شناخت سیستم‌های تجزیه و تحلیل شده با استفاده از الگوها و روابط موجود است.
- شایان ذکر است که داده‌کاوی از ترکیب چندین رشته نشئت می‌گیرد. آمار، یادگیری ماشین، روش‌های بهینه‌سازی، روش‌های تشخیص و شناخت الگو، بانک اطلاعاتی، تجسم‌سازی، شبکه‌های عصبی، مدل‌های ریاضی، بازیابی اطلاعات، الگوریتم ژنتیک و هوش مصنوعی فزونی هستند که داده‌کاوی از آنها بهره می‌برد (محمودی، رستمی، سایبانی و مرادی، ۱۳۹۲).

پیشینه تجربی

البرزی و همکاران در مقاله‌ای با عنوان «به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها» به ارائه مدلی برای اعتبارسنجی مشتریان

بانک‌ها، به منظور اعطای تسهیلات اعتباری پرداخته‌اند. در ساخت این مدل، فرایند توسعه در شناخت الگو و فرایند CRISP به کار رفته است. مدل طبقه‌بندی پیشنهادی در بانک ملت مورد بررسی قرار گرفت و نتایج پژوهش نشان داد، امکان استفاده از مدل طبقه‌بندی پیشنهادی برای ساخت و آزمون درختان تصمیم‌گیری به منظور اعتبارسنجی مشتریان بانک وجود دارد (البرزی، محمدپورزند و خان بابایی، ۱۳۸۹).

«به‌کارگیری داده‌کاوی برای کشف مدل امتیازبندی و تحلیل رفتاری مشتریان بانک» عنوان مطالعه‌ای است که مینایی و اصغری (۱۳۸۷) در خصوص به‌کارگیری داده‌کاوی برای کشف مدل امتیازبندی و تحلیل ریسک مشتریان بانک انجام دادند. در این پژوهش از روش شبکه‌های عصبی برای دسته‌بندی مشتریان متقاضی وام بانکی استفاده شده است؛ به‌گونه‌ای که اطلاعات ۲۴۲۰۸۴ مشتری که از ابتدای سال ۸۳ تا انتهای سال ۸۶ وام دریافت کرده‌اند را، در چارچوب ابزار SQL Server 2005 گردآوری کردند و پس از پردازش اولیه، مدل‌سازی را انجام دادند. همچنین در انتها از روش‌های قوانین وابستگی برای کشف الگوهای رفتاری مشتریان در بازپرداخت وام‌ها استفاده شده است که خصیصه‌های مؤثر در آن، اطلاعاتی چون منطقه و زمان پرداخت وام و میزان بهره آن است.

یانینگ، کیا و فی (۲۰۰۸)، در پژوهشی به‌کارگیری رویکرد داده‌کاوی در پیاده‌سازی سیستم مدیریت با مشتریان را مورد توجه قرار داده‌اند و مزایای بهره‌گیری از داده‌کاوی را از دیدگاه کسب‌وکار در پذیرش مشتریان بانکداری اینترنتی، در قالب یک مدل مفهومی بیان کردند. این مطالعه بیشتر به تحلیل داده‌های کارت‌های اعتباری می‌پردازد.

فنگ و ما (۲۰۰۹) درخصوص کاربرد تکنولوژی داده‌کاوی در مدیریت ارتباط با مشتری^۱، بیان کردند که داده‌کاوی نقش مهمی در بهبود سطح و کارایی مدیریت ارتباط با مشتری در بانک‌های تجاری بازی می‌کند. این مقاله بر کاربرد تکنولوژی داده‌کاوی در CRM بانک‌های اقتصادی تأکید داشته و در مورد بعضی از مسائل کلیدی مرتبط با این کاربرد، داده‌کاوی در CRM شامل مخازن داده، نگهداری مشتری و به‌طور خاص روی دسته‌بندی مشتریان و نقش مهم آن در تجزیه و تحلیل ریسک بحث می‌کند و در نهایت به این نتیجه می‌رسند که تنها با یکپارچه‌سازی سیستم‌های مبتنی بر داده‌کاوی، سیستم‌های بانکداری و سیستم‌های پشتیبانی تصمیم‌گیری، می‌توان سطح خدمات و سوددهی بانک‌ها را بیشتر از گذشته ارتقا داد.

حسینی و همکاران در مقاله‌ای به تحلیل اهمیت عملکرد ویژگی‌های خدمت برپایه بخش‌بندی مشترکان تلفن همراه در استان یزد پرداخته‌اند و برای انجام آن، از نقشه‌های

خودسازمانده و برای محاسبه اهمیت ویژگی‌ها در هر بخش، از شبکه عصبی پیشخور چندلایه استفاده شده است و مشترکان تلفن همراه در استان یزد را در سه بخش دسته‌بندی کرده‌اند (حسینی، بحرینی‌زاده و ضیائی بیده، ۱۳۹۱).

آذر، احمدی و سبط (۱۳۸۹)، در مقاله‌ای با عنوان «طراحی مدل انتخاب نیروی انسانی با رویکرد داده‌کاوی»، به شناسایی شاخص‌های نیروی انسانی که بر عملکرد یا ارتقا مؤثر هستند، پرداخته و در نهایت، مدلی برای انتخاب متغیرهای تأثیرگذار، متغیر هدف و الگوریتم‌های مناسب، از بین قواعد به‌دست‌آمده، ارائه کرده‌اند.

جدول ۱. مقایسه پژوهش‌های ارائه‌شده با پژوهش حاضر

پژوهشگر	سال	روش داده‌کاوی	زمینه پژوهش
یانینگ، کیا و فی	۲۰۰۸	الگوریتم ژنتیک و شبکه عصبی	کارت اعتباری در بانکداری اینترنتی
فنگ و ما	۲۰۰۹	شبکه عصبی	بررسی ریسک در مدیریت ارتباط با مشتری
مینایی و اصغری	۱۳۸۷	شبکه عصبی	بازپرداخت تسهیلات
البرزی، محمدپورزندی و خان‌بابایی	۱۳۸۹	الگوریتم ژنتیک	اعتبارسنجی مشتریان
آذر، احمدی و سبط	۱۳۸۹	C5	انتخاب نیروی انسانی
حسینی، بحرینی‌زاده و ضیائی بیده	۱۳۹۱	شبکه عصبی	صنعت خدمات تلفن همراه
پژوهش حاضر	۱۳۹۲	C5	بررسی مشتریان اینترنت بانک

با توجه به پیشینه مطرح‌شده، این پژوهش به ارائه یک چارچوب پیشنهادی نوین برای طبقه‌بندی مشتریان بانک در سطوح مختلف استفاده از خدمات اینترنت بانک به کمک مدل فرایندی CRISP-DM پرداخته و درخت‌های تصمیم‌گیری منتج به انتخاب الگوریتم C5 مقایسه شده‌اند. این امر به کشف الگوهای پنهان موجود در داده‌های تراکنشی موجود در پایگاه داده‌های بانک، به صورت مجموعه‌ای از قوانین کمک می‌کند.

روش‌شناسی پژوهش

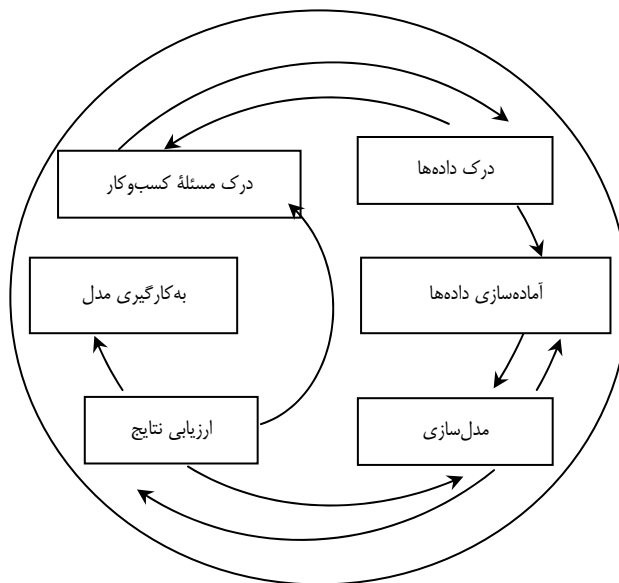
پژوهش پیش رو از نظر هدف، پژوهشی کاربردی و از نظر گردآوری داده‌ها، پژوهشی اسنادی شمرده می‌شود. برای دستیابی به اطلاعات مشتریان بانک از پایگاه داده^۱، فهرستی از اطلاعات

1. Data Center

مورد نیاز استخراج شد. همچنین برای شناسایی متغیرهای تأثیرگذار در رفتار مشتریان در سیستم اینترنت بانک (فیلتر کردن خصیصه‌ها)، از مصاحبه آزاد با خبرگان بانک استفاده شده است. نوع داده‌های این پژوهش به صورت عددی (پیوسته) و اسمی (گسسته) است.

در این پژوهش از مدل استاندارد CRISP-DM استفاده شده است که یک روش استاندارد داده‌کاوی بوده و در اواخر سال ۱۹۹۶، سه شرکت بزرگ دایملر کرایسلر (بنز)، اس.پی.اس.اس.^۱ و ان.سی.آر.^۲ آن را ایجاد کرده‌اند (چپمن و همکاران، ۲۰۰۰: ۱).

این روش، یک مدل فرایندی برای داده‌کاوی ارائه می‌دهد که مروری بر چرخه عمر هر پروژه داده‌کاوی است، چرخه عمر یک پروژه داده‌کاوی، شامل شش مرحله است: درک مسئله کسب‌وکار، درک داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی نتایج و به‌کارگیری مدل که در شکل ۲ نشان داده شده است (آذر و همکاران، ۱۳۸۹).



شکل ۲. چرخه عمر یک پروژه داده‌کاوی

منبع: چپمن و همکاران، ۲۰۰۰: ۱۰

برای داشتن پژوهشی موفق در زمینه داده‌کاوی، باید متخصص داده‌کاوی از توان و تجربه متخصص کسب‌وکار در تمام فرایند داده‌کاوی بهره‌مند شود (فایاد و یوتورسامی، ۱۹۹۶).

1. SPSS
2. NCR

با توجه به اینکه برای انجام فرایند پژوهش، استاندارد جهانی CRISP-DM مورد استفاده قرار گرفته است؛ در زیر ساختار اجرایی پژوهش بر اساس مراحل این استاندارد تشریح می‌شود:

- **درک مسئله کسب‌وکار:** این مرحله ابتدا بر درک اهداف و ملزومات پروژه از دیدگاه کسب‌وکار متمرکز می‌شود که اصلی‌ترین هدف کسب‌وکار بانک، پیش‌بینی وضعیت رفتار مشتریان حقیقی بانک در استفاده از خدمات اینترنتی است و سپس تبدیل این دانش به یک تعریف مسئله داده‌کاوی و یک برنامه طراحی شده مقدماتی برای دستیابی به اهداف را بررسی می‌کند. هدف داده‌کاوی در این پژوهش «طبقه‌بندی» و نوع آن «پیش‌بینی‌کننده» است؛ بدین‌گونه که با توجه به بررسی رفتار مشتریان اینترنت‌بانک، مشتریان در سه طبقه خوب، متوسط و ضعیف تقسیم‌بندی شده‌اند، سپس دانش مربوط به پیش‌بینی مشتریان جدید بر اساس الگوریتم‌های داده‌کاوی استخراج شد.

- **درک داده‌ها:** این مرحله با جمع‌آوری اولیه داده‌ها شروع می‌شود و به توصیف داده‌ها و تعیین کیفیت آنها می‌پردازد. در این پژوهش برای ارزیابی اولیه از داده‌ها، ابتدا پایگاه‌های داده موجود در بانک که اطلاعات مشتریان حقیقی اینترنت‌بانک را در خود ذخیره کرده است، مورد بررسی قرار گرفت و سپس با نمونه‌هایی از داده‌های موجود در پایگاه‌های داده، نحوه اخذ داده‌ها و اطلاعات مورد نیاز از سیستم‌های عملیاتی رایانه‌ای، سیستم رتبه‌بندی، نحوه ذخیره‌سازی داده‌ها و کسب اطلاعات از آنها، آشنایی کامل حاصل شد. پیش از جمع‌آوری و انتخاب داده‌ها برای شروع عملیات پیش‌پردازش، ابتدا کیفیت داده‌های موجود و دسترسی به آنها مورد سنجش قرار گرفت. این ارزیابی کیفیت به این دلیل انجام گرفت که در داده‌ها، خصیصه‌های اصلی تأثیرگذار در رفتار مشتریان وجود داشته باشد و داده‌های گم‌شده یا ناقص در میان داده‌ها، میزان قابل قبولی داشته باشند.

- **آماده‌سازی داده‌ها:** مرحله آماده‌سازی داده‌ها، شامل کلیه فعالیت‌هایی است که برای ساختن مجموعه داده‌های نهایی (داده‌هایی که برای مدل‌سازی آماده شده‌اند) از داده‌های خام اولیه به کار می‌رود. هرچه کیفیت این آماده‌سازی بهتر باشد، مدل‌سازی نیز بهتر خواهد بود (آذر، احمدی و سبط، ۱۳۸۹). وظایف آماده‌سازی داده‌ها در چند دوره انجام می‌گیرد و هیچ ترتیب از پیش تعریف‌شده‌ای ندارد. این وظایف شامل انتخاب جداول، رکوردها و خصیصه‌ها و همچنین

انتقال و پاک‌سازی داده برای مدل‌سازی است. در این مرحله، داده‌های مرتبط با مشتریان اینترنت‌بانک از پایگاه داده‌هایی که در مرحله قبل شناسایی شدند، استخراج و ثبت شدند. سپس داده‌ها در یک پایگاه داده جامع و یکپارچه (البته پالایش نشده) به‌منزله یک پایگاه داده رابطه‌ای قرار گرفتند. سپس داده‌ها پالایش شده و ساختار مورد نظر برای مدل‌سازی روی آنها اعمال شد. برای پاک‌سازی و پیش‌پردازش داده‌ها، دو عملیات کاهش داده و اعمال تغییرات در شکل داده‌ها، روی پایگاه داده رابطه‌ای صورت گرفت.

• **مدل‌سازی^۱:** در این مرحله، انواع روش‌های مدل‌سازی انتخاب شده و به کار گرفته می‌شوند. در کل، برای یک نوع مسئله داده‌کاوی چندین روش وجود دارد. برخی از روش‌ها نیازمند فرمت ویژه‌ای از داده‌ها هستند. بنابراین اغلب لازم است که به مرحله آماده‌سازی داده بازگشت. برای اعمال روش‌ها در پژوهش حاضر، از نرم‌افزار Clementine12 استفاده شده است. برای مدل‌سازی داده‌ها، به‌طور جداگانه روی هر الگوریتم طبقه‌بندی انجام گرفت و الگوریتمی که بالاترین صحت را داشت، مبنای مدل‌سازی و استخراج دانش از آن قرار گرفت. هدف از این کار، استخراج دانش با توجه به داده‌های آموزشی از الگوریتم انتخابی است.

• **ارزیابی نتایج^۲:** در این مرحله از پروژه، مدلی که از دیدگاه تحلیل داده، کیفیت بالایی دارد، ساخته شده است. پیش از اقدام برای به‌کارگیری مدل، باید مدل به‌طور کلی ارزیابی شده و گام‌های اجرایی برای تطابق با اهداف کسب‌وکار مرور شوند. در اینجا با توجه به مقایسه‌های صورت گرفته، روشی که از بقیه پیش‌بینی دقیق‌تری را انجام می‌دهد، برای استفاده بهره‌وران (بانک مورد مطالعه) پیشنهاد می‌شود.

• **به‌کارگیری مدل^۳:** ایجاد مدل، به‌معنای پایان پروژه نیست؛ حتی اگر هدف مدل ارتقای دانش از داده‌ها باشد، باز هم دانش حاصل شده به سازماندهی نیاز دارد و باید به‌شکلی ارائه شود که بهره‌وران بتوانند از آن استفاده کنند. بسته به ملزومات کار، فاز به‌کارگیری می‌تواند به‌سادگی ایجاد یک گزارش، یا به پیچیدگی اجرای یک فرایند قابل تکرار داده‌کاوی باشد. با توجه به اینکه این پژوهش یک پژوهش کاربردی است، نتایج می‌تواند به‌صورت یک سیستم سایه‌ای، به‌موازات کار کارشناسان بانک کاربرد داشته باشد. به این معنا که نتایج طبقه‌بندی و ارزیابی سیستم داده‌کاوی، بر ارزیابی کارشناسان بانک تأثیرگذار باشد. برای مثال

1. Modeling
2. Evaluation
3. Deployment

اگر سیستم داده‌کاوی مشتری را خوب ارزیابی کند، بانک می‌تواند سرویس‌های بیشتری را در اینترنت‌بانک در اختیار مشتری قرار دهد و تبلیغات بیشتری در مورد سایر محصولات بانکداری مدرن داشته باشد.

یافته‌های پژوهش

مراحل اجرایی پژوهش به سه قسمت توصیف داده‌ها، تحلیل داده‌ها و ارزیابی مدل، تفکیک شده‌اند که در ادامه جزئیات آنها تشریح شده است.

توصیف داده‌ها

جامعه آماری پژوهش متشکل از داده‌های مربوط به حدود ۴۷۰۰۰ مورد از مشتریان بانک سینا هستند که در بازه سه‌ساله (۱۳۸۸ تا ۱۳۹۱) سرویس اینترنت‌بانک برای آنها فعال شده است. با توجه به اینکه داده‌های موجود در جامعه آماری از لحاظ کیفیت مورد بررسی قرار گرفته و تا حد امکان پالایش شده‌اند، بنابراین مدل‌های پژوهش برای نمونه نهایی، فقط روی ۷۰۴۷ مورد از داده‌های مشتریان حقیقی اجرا شد که این داده‌ها در یک پایگاه داده در نرم‌افزار اکسل ذخیره شده‌اند. رکوردهای حذف‌شده مربوط به مشتریان حقوقی، مشتریانی با اطلاعات ناقص و مشتریانی که هرگز از خدمات اینترنت‌بانک استفاده نکردند، بوده است. شش خصیصه مشتریان متغیرهای مستقل پژوهش و طبقه مشتری، متغیر وابسته آن هستند که در جدول ۲ نشان داده شده است.

جدول ۲. عناوین فیلدها

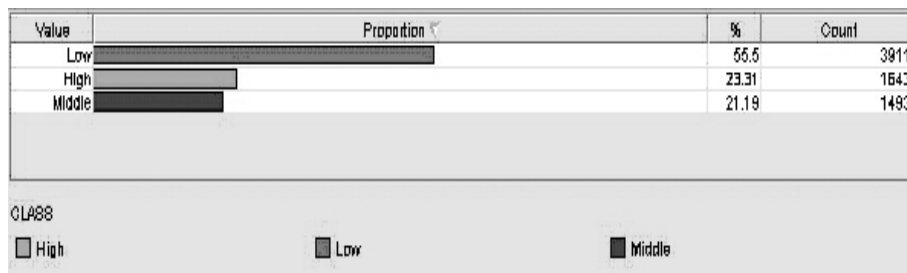
نام فیلد خصیصه	سن	سطح تحصیلات	جنسیت	تأهل	شغل	محل سکونت
نام اختصاری	AGE	EDUCATION	GENDER	MARIED	JOB	CITY
نام فیلد طبقه	کلاس					
نام اختصاری	Class					

در جدول ۳ وضعیت توزیع داده‌ها برای هر یک از فیلدها نمایش داده شده است.

جدول ۳. توزیع داده‌ها

فیلد خصیصه	نمودار	نوع داده	تعداد گروه‌ها	تعداد رکوردهای معتبر
سن		پیوسته	۴	۷۰۴۱
سطح تحصیلات		پیوسته	۳	۶۳۶۵
جنسیت		گسسته	۲	۷۰۴۷
تأهل		گسسته	۲	۶۵۲۹
محل سکونت		پیوسته	۴	۷۰۴۷
شغل		پیوسته	۵	۷۰۴۷
کلاس		پیوسته	۳	۷۰۴۷

در این جدول توزیع داده‌ها در هر یک از فیلدها مشخص شده است. رنگ مشکی طبقه مشتریان متوسط، رنگ خاکستری مشتریان طبقه بد و رنگ سفید مشتریان طبقه خوب را نمایان می‌کند. شکل ۳ توزیع سه طبقه مشتریان خوب، بد و متوسط در فیلد کلاس را نشان می‌دهد. طبقه مشتریان بد ۵۵/۵ درصد (۳۹۱۱ رکورد) از کل مشتریان، طبقه مشتریان متوسط ۲۱/۱۹ درصد (۱۴۹۳ رکورد) و طبقه مشتریان خوب ۲۳/۳۱ درصد (۱۶۴۳ رکورد) است.



شکل ۳. توزیع طبقه‌بندی مشتریان

نوع داده‌های این پژوهش از هر دو نوع پیوسته و گسسته است. گفتنی است، برخی داده‌های مربوط به متغیرهای عددی، به داده‌های اسمی تبدیل شده‌اند و هر کدام از متغیرهای اسمی در این داده‌ها، انواع مختلفی دارند که در جدول ۴ مشاهده می‌شود.

جدول ۴. انواع متغیرهای اسمی

متغیرهای اسمی	نوع مربوطه
سطح تحصیلات	پایین (دیپلم و پایین‌تر)، متوسط (فوق دیپلم، لیسانس)، بالا (فوق لیسانس و بالاتر)
جنسیت مشتری	مرد، زن
محل سکونت	پایتخت (تهران)، مراکز استان، شهرستان‌های استان و روستاها، کشورهای خارجی
شغل مشتری	کارمند، آزاد (تولیدی، کاسب، روزمزد)، فرهنگی (مدرس، استاد، دانشجو، محصل)، پزشکی (جراح، پزشک، داروساز، بیکار (خانه‌دار، بازنشسته))
وضعیت تأهل	مجرد، متأهل
سن	نوجوان (کمتر از ۲۰)، جوان (بین ۲۰ و ۳۵)، میان‌سال (بین ۳۵ و ۴۵)، مسن (بیش از ۴۵)

فیلد کلاس مربوط به طبقه مشتریان است که بر اساس نحوه استفاده از سرویس‌های اینترنت بانک (به‌منزله معیار تعیین شده) در سه وضعیت خوب، متوسط و بد تفسیر می‌شود. انواع عملیات بانکی در اینترنت بانک (انواع انتقال وجه، عملیات چک، عملیات تسهیلات، عملیات کارت نقدی، صورت حساب سپرده و کارت اعتباری) معیارهای تعیین کننده هستند. برای وزن دهی به این معیارها از پنج نفر خبره در امور بانکی کمک گرفته شد و هر پنج نفر به معیارهای مطرح شده وزنی دادند که بعد از نرمال سازی نهایی، هر یک از انواع عملیات بانکی، وزن مربوطه را گرفتند. مشتریان براساس تعداد و نوع عملیات بانکی، در فیلد کلاس امتیازی بین صفر تا صد دریافت کردند. این امتیاز، معیار قرار گرفتن مشتریان فعلی در هر یک از طبقات است (جدول ۵). مشتریانی که امتیاز آنها کمتر از ۲۰ باشد، در طبقه مشتریان بد اینترنت بانک، مقادیر بین ۲۰ تا ۶۰ مشتریان متوسط و مقادیر بالاتر از ۶۰ جزء مشتریان خوب اینترنت بانک محسوب می‌شوند.

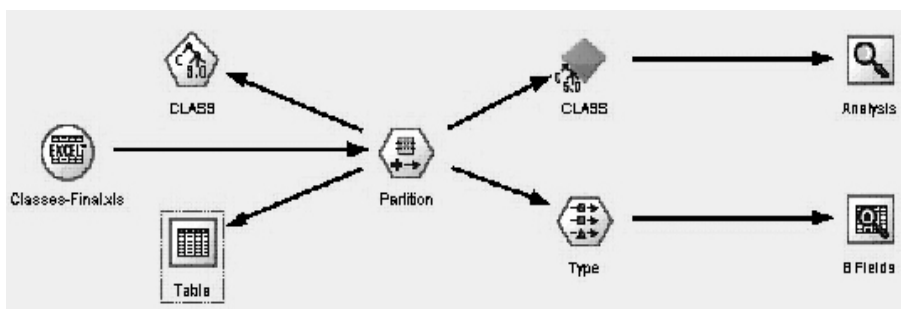
جدول ۵. طیف طبقه‌ها

طبقه مشتری	بد	متوسط	خوب
امتیاز دریافت شده از سرویس‌های اینترنت بانک	کمتر از ۲۰	بین ۲۰ و ۶۰	بالاتر از ۶۰

تحلیل داده‌ها

در این بخش داده‌ها را مدل سازی کرده و تحلیل‌های لازم در خصوص اعتبارسنجی مدل و کارایی آن ارائه می‌شود.

پس از حذف فیلدهای غیر مفید، پایگاه داده نهایی برای طبقه‌بندی مشتریان به‌منظور ورود به نرم‌افزار، دارای ۷۰۴۷ رکورد و ۶ فیلد از خصیصه‌های مشتریان و فیلد کلاس مشتری بود. برای فراخوانی طبقه‌بندی، باید فیلد کلاس را خروجی و بقیه فیلدها را به‌طور ورودی معرفی کرد. الگوریتم اصلی اعمال شده در این پژوهش، درخت تصمیم C5 بوده و استریم حاصل از مدل‌سازی در نرم‌افزار Clementine 12 به‌صورت شکل ۴ است.



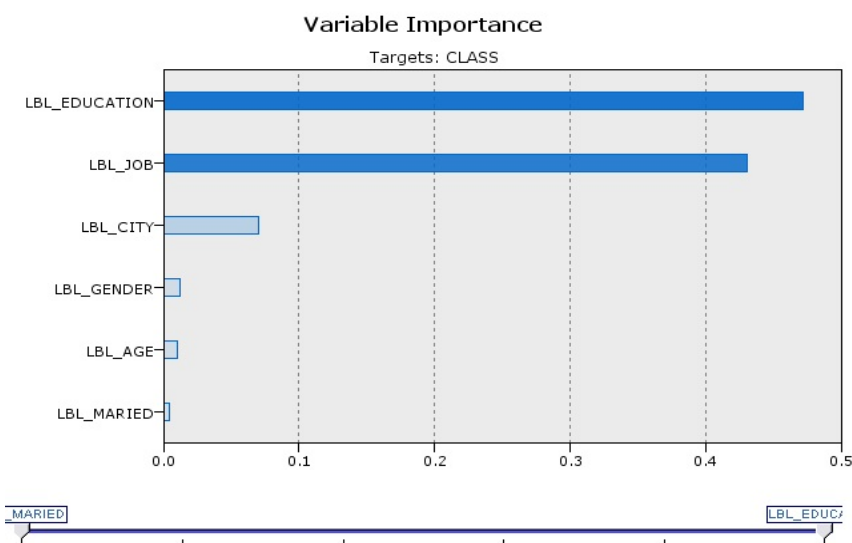
شکل ۴. طرح مدل‌سازی الگوریتم طبقه‌بندی C5

میزان دقت الگوریتم به‌لحاظ طبقه‌بندی صحیح مشتریان با توجه به میزان تکرارهای آزمایش (دفعات اعمال الگوریتم روی داده‌ها)، به‌صورت جدول ۶ است. همان‌گونه که در این جدول مشخص است، برای داده‌های آموزش از کل ۴۹۱۶ رکورد، تعداد ۳۶۵۲ رکورد (۷۴/۲۹ درصد) از داده‌ها به‌طور صحیح طبقه‌بندی شدند و تعداد ۱۲۶۴ رکورد (۲۵/۷۱ درصد) طبقه‌بندی نادرستی داشتند. برای داده‌های آزمایشی، از کل ۲۱۳۱ رکورد، تعداد ۱۵۹۷ رکورد (۷۴/۹۴ درصد) طبقه‌بندی درست و تعداد ۵۳۴ رکورد (۲۵/۰۶ درصد) نادرست طبقه‌بندی شدند.

جدول ۶. دقت الگوریتم درخت تصمیم C5

آزمایش		آموزش		پار تیشن
۱۵۹۷	۷۴/۹۴٪	۳۶۵۲	۷۴/۲۹٪	نمونه‌های طبقه‌بندی‌شده صحیح
۵۳۴	۲۵/۰۶٪	۱۲۶۴	۲۵/۷۱٪	نمونه‌های طبقه‌بندی‌شده نادرست
۲۱۳۱		۴۹۱۶		کل نمونه‌ها

الگوریتم C5، اهمیت خصیصه‌های مشتریان را به صورت شکل ۵ اولویت‌بندی می‌کند که این موضوع، آگاهی سازمانی ارزشمندی را در اختیار بانک قرار می‌دهد تا شاخص‌هایی با اهمیت بالا را مبنای تصمیم‌گیری قرار دهد.



شکل ۵. اهمیت شاخص‌ها

ارزیابی مدل

از آنجا که باید اعتبار روش ارائه‌شده در هر پژوهشی مورد سنجش قرار گیرد، برای سنجش اعتبار و صحت مدل، از تقسیم نمونه به دو مجموعه داده آموزشی و آزمایشی استفاده شده است. میزان اعتبار با نتایج داده‌های جدید آزمون می‌شود و داده‌های آزمایشی به منزله ناظر به الگوریتم وارد شده و نتایج میزان صحت آن را ارزیابی می‌کند. در این پژوهش، ۷۰ درصد از داده‌ها مجموعه داده‌های آموزشی و ۳۰ درصد نیز، مجموعه داده‌های آزمایشی انتخاب شده‌اند. معیار اعتبار و صحت مدل، بسته به صحت طبقه‌بندی یا تفکیک داده‌های آزمایشی است. در این پژوهش از اعتبارسنجی متقابل با ده تکرار استفاده شده است؛ به این معنا که مجموعه داده‌ها به ده قسمت تقسیم می‌شود و هر بار ۷۰ درصد از مجموعه داده‌های آموزشی و ۳۰ درصد از مجموعه داده‌های آزمایشی انتخاب شده و دقت طبقه‌بندی سنجیده می‌شود، پس از اینکه فرایند ده بار تکرار شد، دقت نهایی مدل ارائه می‌شود.

برای سنجش کارایی مدل ارائه شده درخت تصمیم C5، صحت نتایج این مدل با الگوریتم‌های طبقه‌بندی دیگری چون شبکه عصبی، درخت تصمیم CHAID و رگرسیون لجستیک^۱ مورد مقایسه قرار گرفت که نتایج آن در جدول ۷ نمایش داده شده است.

جدول ۷. مقایسه تطبیقی صحت الگوریتم‌ها

الگوریتم	C5	Neural Net	CHAID	Logistic
صحت	٪۷۴/۹۴	٪۷۴/۳۳	٪۷۳/۹۱	٪۶۱/۳۸

همان‌طور که در جدول ۷ مشاهده می‌شود، درخت تصمیم‌گیری C5 به میزان ۰/۶۱ درصد دارای دقت بالاتری نسبت به بقیه درختان است، اما با توجه به اینکه تعداد برگ‌ها و اندازه این درخت خیلی بیشتر از تعداد برگ‌ها و اندازه درختان دیگر است، بنابراین بهترین الگوریتم انتخاب شده است. همچنین این درخت از پیچیدگی کمتری نسبت به سایر درختان برخوردار است که به دلیل گستردگی آن، از آوردن نمودار فرم درختی و ذکر کامل فرم اگر آنگاه^۲ خودداری شده است.

الگوریتم استخراج شده قابلیت طبقه‌بندی رکوردهای جدید مشتریان را داشته و می‌توان به میزان صحت آن در تشخیص وضعیت رفتار مشتریان جدید اطمینان کرد. هنگامی که اطلاعات مشتری جدید وارد سیستم می‌شود، آن را از ابتدای درخت (گره ریشه)، بر اساس ویژگی‌های بیوگرافیک هدایت می‌کنیم تا زمانی که به یک گره برگ برسیم، در آنجا رده مربوطه را به مشتری نسبت داده و می‌توانیم پیش‌بینی کنیم مشتری در کدام طبقه اینترنت بانک قرار می‌گیرد.

نتیجه‌گیری و پیشنهادها

داده‌کاوی ابزار مهمی برای استفاده سودمند از داده‌ها به شمار می‌رود و یکی از مهم‌ترین فناوری‌ها برای بهره‌برداری مؤثر و دقیق از داده‌های حجیم است. بانک‌ها برای ارائه خدمات به مشتریان، نیازمند شناسایی و دسته‌بندی آنها هستند و در این راستا پژوهش پیش رو، از الگوریتم‌های طبقه‌بندی و ساخت درختان تصمیم‌گیری استفاده کرده است.

با توجه به مطالعات انجام گرفته در این زمینه، می‌توان گفت بهره‌گیری از داده‌کاوی در مدیریت ارتباط با مشتریان در سیستم‌های بانکداری، بیشتر در حوزه طبقه‌بندی مشتریان کارت

1. Logistic -R
2. If-Then

اعتباری و تسهیلات بانکی و پیش‌بینی ریسک آنها بوده است و کمتر به طبقه‌بندی مشتریان اینترنت بانک پرداخته شده است.

از آنجاکه درخت تصمیم C5، بهینگی و دقت بالاتری نسبت به کلیه الگوریتم‌های طبقه‌بندی داده‌کاوی مورد استفاده در پژوهش دارد، بنابراین دانش استخراج‌شده از این درخت، مورد اعتمادترین دانش حاصل از داده‌های مورد بررسی به‌شمار رفته و می‌تواند مبنای استخراج قوانین داده‌کاوی باشد. مهم‌ترین شاخص‌ها در این الگوریتم، تحصیلات و شغل است و شاخص‌های شهر، جنسیت، سن و تأهل، به‌ترتیب از اهمیت‌های کمتری برخوردارند. برخی قوانین به‌دست‌آمده از این الگوریتم به‌شرح زیر هستند:

- در خصوص افراد فرهنگی:

- نوجوانان در طبقه بد؛
- مردان جوان با تحصیلات متوسط در طبقه خوب؛
- افراد جوان (مردان و زنان) با تحصیلات بالا در طبقه خوب؛
- مردان میان‌سال با سطح تحصیلات بالا در طبقه خوب؛
- زنان میان‌سال با سطح تحصیلات بالا در طبقه متوسط.

- در خصوص افراد با مشاغل آزاد:

- افراد با سطح تحصیلات بالا در طبقه بد؛
- افراد با سطح تحصیلات متوسط در طبقه خوب؛
- افراد با سطح تحصیلات پایین در طبقه بد.

- در خصوص کارمندان:

- افراد مسن ساکن پایتخت در طبقه متوسط؛
- مردان میان‌سال ساکن پایتخت در طبقه خوب؛
- زنان میان‌سال ساکن پایتخت در طبقه متوسط؛
- افراد ساکن مراکز استان‌ها با توجه به سطح تحصیلات (بالا، متوسط و پایین)، طبقه کلاسشان (خوب، متوسط و بد) تعیین می‌شود.

- پزشکان در طبقه متوسط و خوب.

- افراد بیکار در طبقه بد.

با در نظر گرفتن محدوده پژوهش، موارد زیر برای مطالعات بعدی پیشنهاد می‌شود:

- انجام پژوهشی مشابه روی شبکه‌های دیگر بانکداری مدرن، از جمله همبانک که به‌تازگی وارد حوزه بانکی شده است؛

- ارائه انواع مدل الگوریتم‌های طبقه‌بندی ترکیبی مبتنی بر داده‌کاوی، در شناسایی رفتار مشتریان در حوزه بانکداری مدرن؛
- از آنجایی که داده‌های این پژوهش مربوط به یک بانک است، پیشنهاد می‌شود پژوهشی با روش‌های به کار رفته در این پژوهش، در بانک‌ها یا مؤسسه‌های مشابه دیگر انجام گیرد و نتایج آن با نتایج این پژوهش و روش‌های دیگر، مقایسه شوند.
- در پایان پیشنهادهایی برای بهره‌وران از این پژوهش و به‌ویژه برای بانک مورد مطالعه به شرح زیر ارائه می‌شود:
- آموزش کارکنان شعب، به‌منظور راهنمایی مشتریان در استفاده از این خدمات بر اساس طبقه مشتریان، سطح استفاده از سرویس‌های بانکداری الکترونیک را افزایش می‌دهد.
- تهیه و تنظیم سیستمی کارآمد، به‌منظور پیش‌بینی رفتار مشتریان جدید در استفاده از خدمات اینترنت‌بانک، می‌تواند نوع خدمات و سرویس‌های ارائه‌شده به مشتریان را شخصی‌سازی کرده و زمینه رضایت مشتریان را فراهم کند.

منابع

- آذر، ع؛ احمدی، پ. و سبط، م. (۱۳۸۹). طراحی مدل انتخاب نیروی انسانی با رویکرد داده‌کاوی (استخدام داوطلبان آزمون‌های ورودی یک بانک تجاری در ایران). *مدیریت فناوری اطلاعات*، ۲ (۴): ۲۲-۳.
- البرزی، م؛ محمد پورزندی، م؛ خان بابایی، م. (۱۳۸۹). به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها. *مدیریت فناوری اطلاعات*: ۲ (۴): ۲۳-۳۸.
- تارخ، م؛ شریفیان، ک. (۱۳۸۹). کاربرد داده‌کاوی در بهبود مدیریت ارتباط با مشتری. *مطالعات مدیریت صنعتی*، ۶ (۱۷): ۱۸۱-۱۵۳.
- تقوی‌فرد، م. ت؛ منصوری، ط؛ ناصرزاده، م. ر؛ فراست، ع. ر. (۱۳۸۶). داده‌کاوی و کاربرد آن در تصمیم‌گیری. *دانش مدیریت*، ۲۰ (۷۹): ۱۴-۳.
- تقوی‌فرد، م. ت؛ نادعلی، ا. (۱۳۹۱). طبقه‌بندی متقاضیان تسهیلات اعتباری بانکی با استفاده از داده‌کاوی و منطق فازی. *مطالعات مدیریت صنعتی*، ۹ (۵۲): ۹۱-۵۲.
- حسینی، ی؛ بحرینی‌زاده، م. و ضیائی‌بیده، ع. (۱۳۹۱). تحلیل اهمیت - عملکرد ویژگی‌های خدمت برپایه بخش‌بندی مشتریان با رویکرد داده‌کاوی (پژوهشی در بازار خدمات تلفن همراه در استان یزد). *مدیریت فناوری اطلاعات*، ۴ (۱۳): ۷۰-۴۵.

دیواندری، ع.؛ عابدی، ا.؛ ناصرزاده، م. ر. (۱۳۹۲). ارائه مدل مفهومی برای تبیین عوامل کلیدی مؤثر بر کیفیت سیستم‌های ارائه‌دهنده خدمات بانکداری اینترنتی (پیمایشی پیرامون بانک ملت). مدیریت فناوری اطلاعات، ۵ (۱): ۳۶-۱۹.

رسولی، ه.؛ مانیان، ا. (۱۳۹۱). طراحی سیستم استنتاج فازی برای انتخاب خدمات بانکداری الکترونیک (مطالعه موردی بانک سپه). مدیریت فناوری اطلاعات، ۴ (۱۲): ۶۴-۴۱.

عزیزی سرخنی، م.؛ اله‌قلی‌زاده آذری، م.؛ کردلوئی، ح. (۱۳۸۷). بررسی زیرساخت‌های موجود بانک تجارت برای استقرار بانکداری الکترونیکی. فصلنامه علمی ترویجی مدیریت (پژوهشگر)، ۵ (۱۰): ۱-۱۱.

محمودی، ک.؛ رستمی، ح.؛ سایبانی، م.؛ مرادی، ع. (۱۳۹۲). مروری بر علم داده‌کاوی و کاربردهای آن در صنایع فراساحل. پنجمین همایش ملی صنایع فراساحل OIC. دانشگاه صنعتی شریف. ۳۱ اردیبهشت و ۱ خرداد ۱۳۹۲.

مینایی، ب.؛ اصغری، ف. (۱۳۸۷). به‌کارگیری داده‌کاوی برای کشف مدل امتیازبندی و تحلیل رفتاری مشتریان بانک. دومین کنفرانس داده‌کاوی ایران، دانشگاه صنعتی امیرکبیر. ۲۱ و ۲۲ آبان ۱۳۸۷.

هاشمیان، م.؛ عیسانی، م. ت.؛ میکائیلی، ف.؛ طباطبائی، م. (۱۳۹۱). عوامل مؤثر بر پذیرش ابزارهای بانکداری الکترونیک از سوی مشتریان (پیمایشی درباره بانک سامان). مدیریت فناوری اطلاعات، ۴ (۱۱): ۱۷۴-۱۵۵.

Bhambri, V. (2011). Application of Data Mining in Banking Sector. *IJCST*, 2 (2): 199-202.

Chapman, P. & Clinton, J. & Kerber, R. & Khabaza, T. & Reinartz, T. & Shearer, C. & Wirth, R. (2000). *CRISP-DM Step-Data Mining Guide*. SPSS Inc. CRISPMWP.

Deshpande, S.P. & Thakare, V.M. (2010). Data Mining System and Applications: A Review. *International Journal of Distributed and Parallel systems (IJDPS)*, 1 (1): 32-44.

Fang, B. & Ma, S. (2009). Data Mining Technology and Its Application in CRM of Commercial Banks. *First international workshop on database technology and applications, IEEE*. 243-246.

Fayyad, U. & Uthurusamy, R. (1996). Data Mining and Knowledge Discovery in Databases. *communications of the acm*, 39: 24-26.

- Fayyad, U. & Piatetsky-Shapiro, G. & Smyth, P. (1997). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence (AAAI)*, 17 (3): 37-54.
- Kazi Imran, M. & Qazi Baseer, A. (2012). Use of Data Mining in Banking. *International Journal of Engineering Research and Applications (IJERA)*, 2: 738-742.
- Kiss, F. (2003). credit scoring process from a knowledge management prospective. *Periodica Polytechnica Ser. SOC. MAN. SCI*, 11 (1): 95-110.
- Marbán, O. & Segovia, J. & Menasalvas, E. & Fernández-Baizán, C. (2009). Toward data mining engineering: a software engineering approach. *Information Systems*, 34 (1): 87-107.
- Yaning, L. & Kia, L. & Fei, G. (2008). Data mining approach and application in CRM project for internet-focused banking. *IEEE, Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, 12-14 Oct. 1-4.