

# پیش‌بینی رویگردانی مشتری با استفاده از روش درخت تصمیم در صنعت بانکداری الکترونیک (مطالعه موردی بانک انصار)

دکترسید محمد میرمحمدی<sup>۱</sup>، هاجر قانعی<sup>۲</sup>، دکترعباس کرامتی<sup>۳</sup>

## چکیده

امروزه با گسترش فناوری اطلاعات، استفاده از خدمات بانکداری الکترونیک بیش از گذشته مورد توجه قرار گرفته است. با توجه به فضای رقابتی در بین بانک‌ها، حفظ مشتری در این حوزه ضروری شده و رویگردانی مشتری مساله و مشکل اساسی برای بانک‌ها به شمار می‌رود. امکان جمع‌آوری داده در پایگاه داده‌ها و استفاده از داده کاوی به‌عنوان ابزاری قدرتمند برای استخراج دانش از داده‌ها می‌تواند به بانک‌ها در شناسایی الگوهای رفتاری مشتریان کمک به‌سزایی کند. به‌دست آوردن الگوی رفتاری برای مشتریان، مدیران را در تصمیم‌گیری یاری می‌دهد. به‌این‌وسیله آنها می‌توانند مشتریان خود را که با ارزش‌ترین دارایی‌شان هستند، حفظ کنند. با استفاده از الگوهای به‌دست آمده می‌توان مشتریان رویگردان در آینده را شناسایی کرد. هدف این تحقیق ساخت مدل پیش‌بینی رویگردانی مشتری از خدمات بانکداری الکترونیک است. در این تحقیق از الگوریتم کریسپ و روش درخت تصمیم در مرحله مدل‌سازی استفاده شده است. نتایج مدل‌سازی مشخصه‌های مشتریان رویگردان از خدمات بانکداری الکترونیک را نشان می‌دهد.

## کلید واژه‌ها

رویگردانی مشتری، داده کاوی، خدمات بانکداری الکترونیک، درخت تصمیم.

۱ استادیار و عضو هیات علمی دانشگاه پیام نور تهران غرب [smhmz2005@yahoo.com](mailto:smhmz2005@yahoo.com)

۲ کارشناسی ارشد رشته مدیریت MBA دانشگاه پیام نور دماوند، نویسنده مسئول [hghaneei@yahoo.com](mailto:hghaneei@yahoo.com)

۳ دانشیار و عضو هیات علمی گروه مهندسی صنایع دانشگاه تهران [keramati@ut.ac.ir](mailto:keramati@ut.ac.ir)

## مقدمه

امروزه با توجه به جایگاه ویژه‌ی مشتری در کسب و کار و با توجه به فضای رقابتی بانک‌ها و پی بردن به اینکه نگهداری مشتریان موجود کم‌هزینه‌تر از به‌دست آوردن مشتریان جدید است (نای و همکاران، ۲۰۱۱) حفظ مشتری حائز اهمیت ویژه‌ای شده و موضوع رویگردانی مشتری جایگاه ویژه‌ای را به خود اختصاص داده است.

حفظ مشتری برای شرکت سودآور است، زیرا مشتریان در طولانی مدت حساسیت کمتری به فعالیت‌های بازار رقابتی دارند و امکان دارد از طریق تبلیغات دهان به دهان مثبت، مراجعات جدیدی را میسر سازند، در حالیکه ممکن است مشتریان ناراضی تبلیغات دهان به دهان منفی را گسترش دهند (وربک و همکاران، ۲۰۱۰). اگرچه فرایند جذب مشتری جدید مهم است، بیشتر شرکت‌های خدمات مالی اولویت برتر را به حفظ مشتری می‌دهند. به‌طور کلی مدت زمانی که مشتری با بانک باقی می‌ماند بیشتر از مشتری ارزش دارد (بنویت و ون دن پوئل، ۲۰۱۲). مطالعات متعدد اثر قدرت حفظ مشتری را به اثبات رسانده‌اند.

ارقام در یک صنعت جدی بودن مساله رویگردانی مشتری را نشان می‌دهد. برای مثال در سال ۲۰۰۱ نرخ رویگردانی مشتری حدود ۴۰ درصد برای حامل‌های تلفن بی‌سیم بوده است و هزینه‌های مرتبط با مشتری حدود ۱۰ بیلیون دلار بوده است (سارادهی و پالشیکار، ۲۰۱۱). مدیریت رویگردانی بخشی از مدیریت ارتباط با مشتری شده است و برای شرکت‌ها اهمیت ویژه‌ای دارد، زیرا آنها برای ایجاد روابط بلندمدت و حداکثرسازی ارزش بر پایه مشتریانشان تلاش می‌کنند (ریسلادا و همکاران، ۲۰۱۰). رویگردانی مشتری به علت شدت زیان ناشی از آن که سازمان‌ها متحمل می‌شوند، مورد توجه قرار گرفت (سارادهی و پالشیکار، ۲۰۱۱). با پیشرفت فناوری اطلاعات و تکنولوژی و گسترش روز افزون استفاده از خدمات بانکداری حیاتی بودن جلوگیری از رویگردانی مشتریان موجود خدمات بانکداری الکترونیک برای بانک‌ها بارزتر می‌شود. از سال‌های ۱۹۹۰ میلادی تاکنون بانکداری الکترونیک به یک کانال توزیع با بیشترین پتانسیل برای موسسات مالی تبدیل شده است. در حال حاضر اکثر شرکت‌ها به مشتریان خود دسترسی به بسیاری از خدمات را از این طریق پیشنهاد می‌دهند (لیثیانا کابانیلاز، ۲۰۱۳). با توجه به توسعه فناوری اطلاعات، بسیاری از شرکت‌ها مقدار زیادی از داده را انباشته‌اند. تجزیه و تحلیل این داده‌ها می‌تواند به مدیران در تصمیم‌گیری صحیح بازاریابی کمک کند (نای و همکاران، ۲۰۱۱). به‌دلیل توانایی داده کاوی در کشف دانش از

داده‌های مشتریان، این تحقیق به شناسایی مشخصه‌های مشتریان رویگردان و توسعه‌ی مدل پیش‌بینی رویگردانی مشتری با استفاده از روش درخت تصمیم در داده کاوی می‌پردازد.

## مرور ادبیات پژوهش

### رویگردانی مشتری

رویگردانی مشتری تبدیل به یک مساله حیاتی شده است و یکی از چالش‌های اصلی است که شرکت‌های بسیاری در سطح جهانی با آن مواجه هستند (یو و همکاران، ۲۰۱۱). مشتریانی که به ارتباط خود با یک شرکت خاتمه می‌دهند و ارتباط جدید با شرکت دیگر ایجاد می‌کنند، رویگردان نامیده می‌شوند (چن و همکاران، ۲۰۱۴). مشتریان رویگردان می‌توانند به دو دسته اصلی رویگردانان ارادی و رویگردانان اجباری تقسیم شوند. رویگردانی ارادی زمانی اتفاق می‌افتد که مشتری به‌طور داوطلبانه به استفاده از خدمات یک شرکت خاتمه می‌دهد و معمولاً به سمت استفاده از خدمات شرکت رقیب تغییر مسیر می‌دهد. رویگردانی اجباری زمانی رخ می‌دهد که شرکت خودش به ادامه خدمت به مشتری پایان می‌دهد که عموماً به خاطر عدم پرداخت حق عضویت توسط مشتری است (ساراده‌ی و پالشیکار، ۲۰۱۱). رویگردانان اجباری به سادگی قابل شناسایی هستند، آنها مشتریانی هستند که شرکت از ارائه خدمت به آنها صرف‌نظر کرده است (هادن و همکاران، ۲۰۰۵). در این تحقیق به بررسی رویگردانی ارادی از نوع تعمدی پرداخته می‌شود.

مطالعات زیادی تحلیل رویگردانی مشتری را در زمینه‌های مختلفی به کار گرفته‌اند، مانند صنعت کارت اعتباری، صنعت تلفن بی‌سیم و صنعت خدمات مالی (لین و همکاران، ۲۰۱۱). مشتری در دوره مشاهده، تراکنشی با بانک نداشته باشد، رویگردان نامیده می‌شود (نای و همکاران، ۲۰۱۱). در این تحقیق مشتری رویگردان است که در دوره زمانی اول فروردین ۹۲ تا آخر اسفند ۹۳ از طریق درگاه‌های بانکداری الکترونیک بررسی شده تراکنشی نداشته باشد.

### داده کاوی

با کشف دانش در پایگاه داده‌ها، دانش سودمند، با قاعده‌ها و اطلاعات سطح بالا از مجموعه داده‌های مربوط در پایگاه داده استخراج می‌شود و از زوایای مختلف مورد بررسی قرار می‌گیرد (سنگ و چن، ۲۰۱۰). داده کاوی یک رشته مربوط به رشته‌های مختلف علمی است که با هدف کلی پیش‌بینی نتایج و به‌کارگیری الگوریتم‌های پیچیده، برای کشف الگوهای پنهان، وابستگی‌ها، ناهنجاری‌ها یا ساختار از داده‌های بزرگ موجود در پایگاه داده‌ها

یا دیگر مخازن اطلاعات، اطلاعات ضروری را از پایگاه داده‌های بزرگ استخراج می‌کند (تسای و لو، ۲۰۰۹). داده کاوی یک حوزه میان رشته‌ای و با رشد سریع است که حوزه‌های گوناگونی مانند پایگاه داده، آمار، یادگیری ماشین و سایر زمینه‌های مختلف را با هم تلفیق کرده است تا اطلاعات و دانش ارزشمند پنهان را در حجم بزرگی از داده‌ها استخراج کند (غضنفری و همکاران، ۱۳۸۷). داده کاوی امکان جستجو برای اطلاعات ارزشمند در حجم زیادی از داده‌ها را فراهم می‌کند (لیائو و همکاران، ۲۰۱۲). مزیت آن کمک به پیش‌بینی‌هایی در مورد روندهای آتی با استفاده از داده‌های گذشته و حال است. استفاده از آن به مدیران کمک می‌کند تا با حذف عدم کارایی‌ها، مشکلات تجاری سازمان را بهتر حل کنند و تصمیمات مهم را سریع‌تر و با درجه اطمینان بالاتری بگیرند. داده کاوی، ابزار قدرتمندی برای کشف دانش است (حاجی صفری و الهی، ۱۳۹۰).

### داده کاوی در رویگردانی مشتری

به‌منظور باقی ماندن در بازار به شدت رقابتی شرکت‌های بسیاری در زمینه‌های مختلف از تکنیک‌های داده کاوی برای تجزیه و تحلیل رویگردانی استفاده می‌کنند، برای پیش‌بینی رویگردانی مشتری، بسیاری از محققان به‌طور عمده بر دو روش زیر تمرکز کرده‌اند:

**روش اول:** روش‌های طبقه‌بندی سنتی شامل درخت تصمیم، رگرسیون لاجستیک و غیره. این روش‌ها می‌توانند برای تجزیه و تحلیل داده‌های کیفی و داده‌های پیوسته به کار روند.

**روش دوم:** روش‌های هوش مصنوعی شامل شبکه عصبی مصنوعی و غیره (یو و همکاران، ۲۰۱۱). در ادامه به پژوهش‌های انجام شده در این زمینه اشاره می‌شود:

نای و همکاران (۲۰۱۱) در مقاله خود مدل رویگردانی مشتری را با استفاده از دو الگوریتم داده کاوی رگرسیون لاجستیک و درخت تصمیم ساخته‌اند. نتایج نشان می‌دهد که اطلاعات دموگرافیک سهم کمتری در رویگردانی مشتری دارد و عملکرد<sup>۱</sup> در روش رگرسیون کمی بهتر از درخت تصمیم است. شارما و پانیگرایی (۲۰۱۱) در مقاله خود در اشتراک خدمات بی‌سیم تلفنی با استفاده از تکنیک شبکه عصبی، رویگردانی مشتری را پیش‌بینی می‌کنند. آنها تکنیک شبکه عصبی مصنوعی را برای پیدا کردن بهترین مدل از داده‌های ذخیره شده مشتری برای پیش‌بینی رویگردانی پیشنهاد می‌دهند. کرامتی و اردبیلی (۲۰۱۱) در مقاله خود در یکی از اپراتورهای تلفن همراه در ایران با

<sup>۱</sup> performance

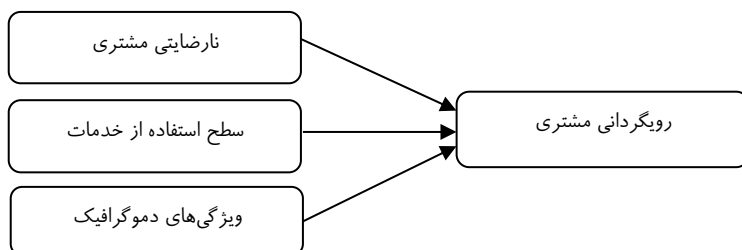
استفاده از روش رگرسیون لاجستیک دو سطحی به بررسی عوامل تاثیرگذار بر رویگردانی مشتری پرداخته‌اند. نتایج این تحقیق نشان می‌دهد که نارضایتی مشتری، سطح استفاده آنها از خدمات مورد استفاده و ویژگی‌های دموگرافیک مشتری، بیشترین تاثیر را بر رویگردانی مشتریان دارند. یو و همکاران (۲۰۱۱) در مقاله خود از روش‌های شبکه عصبی مصنوعی، درخت تصمیم، ماشین‌های بردار پشتیبان و ماشین‌های بردار پشتیبان توسعه یافته برای بررسی اینکه کدام مدل در پیش‌بینی رویگردانی مشتری بهتر است، استفاده کرده‌اند. آنها روش ماشین‌های بردار پشتیبان توسعه یافته را در مقایسه با سه روش ذکر شده پیشنهاد می‌دهند. این الگوریتم دارای این مزیت است که نمونه‌های نامتوازن و غیرخطی را بررسی می‌کند. نوروزی (۱۳۸۸) در پایان‌نامه خود با هدف شناسایی عمده‌ترین دلایل شکل‌گیری رویگردانی در میان مشتریان حساب‌های جاری بانک کشاورزی و با اجرای آن از روش استاندارد کریسپ در خصوص کشف دانش از داده‌ها تبعیت می‌کند. قواعد کلاس‌بندی از درخت تصمیم استخراج شدند. تمدنی جهرمی (۲۰۰۹) در پایان نامه خود بیان داشته است که مسئله‌ای که شرکت تلفن همراه تالیا با آن سرو کار داشت، تشخیص مشتریانی با احتمال زیاد رویگردانی در آینده نزدیک و هدف قرار دادن آنها با عواملشان به منظور متعاقد کردن آنها برای ماندن بود، اما به علت فقدان یک مدل صحیح برای زیر نظر گرفتن رفتار مشتریان‌شان، شرکت قادر نبود رویگردانان را از غیر رویگردانان تشخیص دهد. تکنیک درخت تصمیم برای ساخت مدل انتخاب شده است. توکلی و همکاران (۱۳۸۹) در مقاله خود با هدف تبیین قابلیت‌های داده کاوی در مدیریت رویگردانی مشتری و با بهره‌گیری از متدولوژی استاندارد داده کاوی کریسپ به کاوش در پایگاه‌های داده یکی از شرکت‌های سهامی عام بیمه‌ای در رشته‌ی بیمه آتش سوزی پرداخته است. در تحقیق ایشان اشاره شده است که رایج‌ترین ابزار دسته‌بندی مشتریان بر اساس رویگردانی، درخت تصمیم و شبکه عصبی است که در تحقیقشان از ابزار درخت تصمیم استفاده شده است. نتایج تحقیق آنها نشان می‌دهد کانال جذب مشتری عامل اصلی پیش‌بینی‌کننده رویگردانی یا ماندگاری مشتری در شرکت بوده و در مراتب بعد سابقه خرید و کاربری مکان بیمه شده به‌عنوان عوامل پیش‌بینی کننده رویگردانی به کار می‌رود. احمدی‌نژاد (۱۳۹۱) در تحقیق خود با استفاده از روش رگرسیون سلسله مراتبی تعدیل شده در نرم افزار SPSS به بررسی متغیرهای تاثیرگذار بر رویگردانی مشتری می‌پردازد و نشان می‌دهد که مولفه‌های سن، تحصیلات، سطح درآمد، مدت زمان همکاری با بانک، موقعیت فیزیکی

شعبه، رضایت مشتری، کیفیت خدمات و اعتماد رابطه‌ی معناداری با متغیر رویگردانی مشتری دارند. او با استفاده از تکنیک شبکه عصبی مشتریان را خوشه‌بندی کرده و آنها را به خوشه‌های وفاداران، تحصیلکردگان، جوانان و بانوان تقسیم‌بندی کرده است. سپس با استفاده از روش درخت تصمیم احتمال رویگردانی هر یک از این خوشه‌ها پیش‌بینی شده است. خلاصه تحقیقات پیشین در جدول شماره ۱ آورده شده است.

جدول شماره ۱: تحقیقات پیشین مرتبط با داده کاوی در رویگردانی مشتری

نویسنده	سال	تکنیک‌های استفاده شده	صنعت
نوروزی	۱۳۸۸	خوشه‌بندی با روش k-means. درخت تصمیم	بانک
تمدنی چهرمی	۲۰۰۹	خوشه‌بندی بر اساس مشخصات آر. اف. ام، درخت تصمیم، شبکه عصبی	تلفن همراه
توکلی و همکاران	۱۳۸۹	درخت تصمیم	بیمه
نای و همکاران	۲۰۱۱	رگرسیون لاجستیک، درخت تصمیم	بانک
شارما و کومار پانیگراهی	۲۰۱۱	شبکه عصبی	ارائه‌دهندگان خدمات تلفنی
کرامتی و اردبیلی	۲۰۱۱	رگرسیون لاجستیک دو سطحی	اپراتور تلفن همراه
یو و همکاران	۲۰۱۱	شبکه عصبی مصنوعی، درخت تصمیم، ماشین‌های بردار پشتیبان، ماشین‌های بردار پشتیبان توسعه یافته	تجارت الکترونیک
احمدی نژاد	۱۳۹۱	رگرسیون سلسله مراتبی، شبکه عصبی، درخت تصمیم	بانک

## مدل مفهومی تحقیق



شکل شماره ۱: الگوی مفهومی تحقیق

این الگو برگرفته از الگوی مفهومی است که کرامتی و اردبیلی (۲۰۱۱) در مقاله خود به آن اشاره کرده‌اند. با استفاده از این مدل، داده‌ها از پایگاه داده بانک انصار استخراج شد.

### نارضایتی مشتری

در تعریف ویژگی نارضایتی مشتری، کرامتی و همکاران (۱۳۸۸) در مقاله خود آورده‌اند که به اعتقاد کاتلر اگر عملکرد شرکت انتظارات مشتری را برآورده کند، مشتری احساس رضایت خواهد کرد و در غیر این صورت احساس نارضایتی می‌کند. به این دلیل که تحقیق به روش داده کاوی است و از داده‌های موجود در پایگاه داده بانک استفاده شده است، امکان استفاده از نظر شخصی مشتریان و ارزیابی با استفاده از نظرات آنها وجود ندارد و تنها می‌توان از روی اطلاعات موجود در پایگاه داده بانک به تفسیر رفتار مشتریان پرداخت. کرامتی و اردبیلی (۲۰۱۱) در مقاله خود برای سنجش ویژگی نارضایتی مشتری شاخص‌های شکایت داشتن مشتری و طول دوره ارتباط با او را در نظر گرفته‌اند. در این تحقیق نیز داده‌های مربوط به این دو ویژگی از پایگاه داده بانک انصار استخراج شد.

### سطح استفاده از خدمات

جهان بکام (۱۳۸۶) در پایان نامه خود دستگاه خودپرداز، تلفن بانک، سیستم پایانه‌ی فروش، خدمات اینترنتی را به‌عنوان شاخص‌های بانکداری الکترونیک معرفی کرده است. ایمانی‌خواه (۱۳۹۱) در پایان نامه خود شاخص‌های بانکداری الکترونیک را با توجه به نیازهای بازار بانکداری به بانکداری اینترنتی، بانکداری مبتنی بر تلفن همراه، بانکداری تلفنی، بانکداری مبتنی بر نامبر، بانکداری مبتنی بر دستگاه خودپرداز، بانکداری مبتنی بر پایانه فروش و بانکداری مبتنی بر شعب الکترونیک تقسیم کرده است. در تحقیق حاضر با توجه به داده‌های موجود در پایگاه داده بانک، داده‌های مربوط به تعداد و حجم تراکنش‌ها از طریق درگاه‌های بانکداری الکترونیک اینترنت بانک، همراه بانک، تلفن بانک، آنی بانک و دستگاه خودپرداز از پایگاه داده بانک برای سنجش سطح استفاده از خدمات استخراج شد.

### ویژگی‌های دموگرافیک مشتری

جهان بکام (۱۳۸۶) در پایان‌نامه خود شاخص‌های ویژگی‌های دموگرافیک را به صورت سن، جنسیت، میزان تحصیلات و سطح درآمد مشتری در نظر گرفته است. کلیمس و همکاران (۲۰۱۰) در مقاله خود آورده‌اند که ویژگی‌های دموگرافیک مشتری می‌تواند به سن، تحصیلات، درآمد، فرهنگ یا ملیت دسته‌بندی شود. همچنین او اشاره می‌کند که شغل بر احتمال استفاده از کانال‌های بانکداری الکترونیک اثر می‌گذارد. در تحقیق حاضر با توجه به محدودیت پایگاه داده ویژگی‌های سن، جنسیت، شغل و سطح تحصیلات مورد بررسی قرار گرفت.

### روش‌شناسی پژوهش

در این تحقیق از الگوریتم کریسپ که شامل ۶ گام زیر است، پیروی شده است.

۱. درک نیازهای کسب و کار

۲. درک داده

۳. آماده‌سازی داده

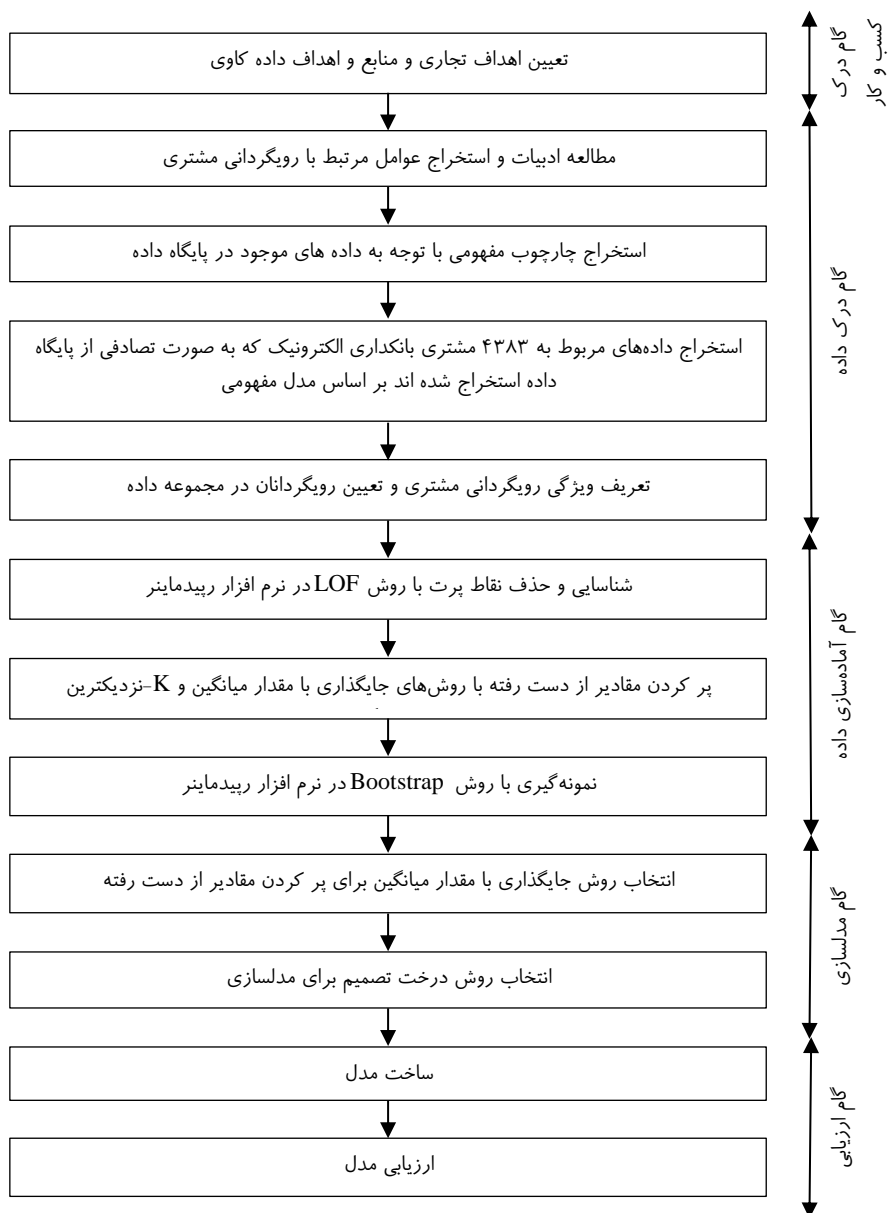
۴. مدل‌سازی

۵. اعتبارسنجی

۶. به کارگیری نتایج

در تحقیق حاضر از پنج گام اول این الگوریتم استفاده شده است و از گام ششم به دلیل اینکه در حیطه این تحقیق نیست صرف‌نظر شده است. مراحل روش تحقیق در شکل ۲ آورده شده است.





شکل شماره ۲: مراحل روش تحقیق

در گام مدل‌سازی از روش درخت تصمیم استفاده شده است که در ادامه توضیح داده شده است:

### درخت تصمیم

در درخت تصمیم هر گره نشان دهنده برخی ویژگی‌های نمونه است و هر شاخه با یکی از مقدارهای ممکن برای این ویژگی در ارتباط است (تسای و چن، ۲۰۱۰). درخت تصمیم یک تکنیک شناخته شده است و کاربردهای موفق بسیاری در مشکلات جهان واقعی دارد. به دلیل اینکه خروجی درخت تصمیم می‌تواند به شکل درخت یا قوانین سازمان یابد، درک نتایج درخت تصمیم، آسان است (نای و همکاران، ۲۰۱۱). در تحقیق حاضر با استفاده از این روش، به منظور به دست آوردن مشخصه‌های مشتریان رویگردان به مدل‌سازی رویگردانی مشتری از خدمات بانکداری الکترونیک پرداخته شده است.

### تحلیل داده‌ها

ذکر این نکته لازم است که هر تحقیق بنا به درون مایه خود بر روی تعدادی از گام‌ها و زیرمرحله‌های الگوریتم کریسپ تمرکز دارد. بنابراین در این تحقیق تنها زیرمراحل از الگوریتم کریسپ که در چارچوب کلی مساله موضوعیت دارد، استفاده شده است. در ادامه مراحل اجرای این الگوریتم آورده شده است.

### گام اول: درک کسب و کار

در این مرحله برای رسیدن به هدف و تعریف مساله داده کاوی و ایجاد یک طرح و روش مناسب، باید درک روشنی از اهداف و نیازمندی‌های پروژه داشت (چن و هوانگ، ۲۰۱۱). درک کسب و کار شامل درک اهداف و نیازمندی‌های داده کاوی است (زن و همکاران، ۲۰۰۷). اهداف تجاری این تحقیق شامل یافتن مشخصه‌های مشتریان رویگردان از خدمات بانکداری الکترونیک است. در این تحقیق از منابع علمی، انسانی و مالی استفاده شده است. به منظور دستیابی به هدف تجاری ذکر شده، با استفاده از روش درخت تصمیم در مدل‌سازی، مشخص می‌شود که مشتری با چه ویژگی‌هایی رویگردان است.

### گام دوم: درک داده

در این گام به مسائلی همچون جمع آوری داده، توصیف داده و اکتشاف در داده پرداخته شده است (لاروس، ۲۰۰۵). در این تحقیق داده‌های ۴۳۸۳ مشتری بانکداری

الکترونیک بانک انصار که به صورت تصادفی از پایگاه داده بانک انصار انتخاب شدند، مورد بررسی قرار گرفت. به این صورت که اطلاعات شکایت مشتریان از خدمات بانکداری الکترونیک، اطلاعات دموگرافیک و تراکنشی در دوره‌ی زمانی ابتدای فروردین سال ۹۲ تا آخر اسفند سال ۹۳ از پایگاه داده بانک استخراج شد.

داده‌های تحقیق به صورت زیر هستند:

داده‌های دموگرافیک مشتریان بانکداری الکترونیک شامل:

**سن:** (کمترین: ۱۹ سال، بیشترین: ۷۶ سال و میانگین سن: ۳۹ سال)

**جنسیت:** (۴,۶۱ درصد زن و ۹۵,۳۹ درصد مرد)

**تحصیلات:** (۰,۵۹ درصد بی‌سواد، ۰,۸ درصد خواندن و نوشتن، ۲,۵۱ درصد ابتدایی، ۱۲,۳۹

درصد سیکل، ۴۲,۱۲ درصد دیپلم، ۱۸,۹۴ درصد فوق دیپلم، ۱۹,۴۴ درصد کارشناسی، ۱,۷۶

درصد کارشناسی ارشد و ۱,۴۶ درصد دکترا)

**شغل مشتریان:** (۰,۹ درصد تولیدی، ۹۴,۶ درصد خدمات، ۰,۵ درصد مسکن و ساختمان،

۳,۷ درصد بازرگانی و ۰,۳ درصد کشاورزی) است.

داده‌های تراکنشی مشتریان بانکداری الکترونیک: تعداد و حجم تراکنش مشتریان

بانکداری الکترونیک از درگاه‌های مختلف بانکداری الکترونیک شامل دستگاه خودپرداز، تلفن

بانک، اینترنت بانک، همراه بانک و آئی بانک در دوره‌ی زمانی تحقیق.

سایر اطلاعات شامل ویژگی شکایت داشتن مشتری (به ازای تمامی مشتریان مقدار صفر

دارد)، طول دوره ارتباط مشتری با بانکداری الکترونیک (کمترین ۰ سال، بیشترین ۱۲ سال

و میانگین ۷ سال)، همچنین رویگردانی مشتری (۱,۴۴ درصد رویگردان و ۹۸,۵۶ درصد

غیررویگردان) هستند. متغیر شغل از نوع اسمی است.

متغیرهای جنسیت، شکایت داشتن مشتری، رویگردانی از نوع باینری هستند. متغیر

جنسیت دارای دو حالت زن و مرد می‌باشد که به ترتیب با اعداد ۰ برای زن و ۱ برای مرد

کدگذاری شده‌اند، متغیرهای شکایت داشتن و رویگردانی مشتری دارای دو حالت ۰ و ۱

هستند که عدد ۰ نشان دهنده شکایت نداشتن و رویگردان نبودن مشتری و عدد ۱

نشان‌دهنده شکایت داشتن و رویگردان بودن مشتری است. متغیر سطح تحصیلات از نوع

ترتیبی است و سایر آنها از نوع عددی گسسته هستند.

### گام سوم: آماده‌سازی داده

در این گام انتخاب داده و پاکسازی داده انجام می‌شود (لاروس، ۲۰۰۵). هان و همکاران (۲۰۱۲) نمونه‌گیری را جزئی از گام آماده‌سازی داده قلمداد کرده‌اند.

در مرحله انتخاب داده متغیرهای مرتبط با رویگردانی مشتری که از پایگاه داده بانک استخراج شده‌اند به جز متغیر شکایت داشتن مشتری به دلیل داشتن ستون تک مقداره صفر، انتخاب می‌شوند.

پاکسازی داده به‌منظور پر کردن داده‌های از دست رفته، شناسایی و از بین بردن اختلال و داده‌های پرت و همچنین درست کردن تناقض‌ها در مجموعه داده استفاده می‌شود (هان و همکاران، ۲۰۱۲). در تحقیق حاضر، مشکل نمونه‌های پرت در مجموعه داده به وسیله روش LOF در نرم افزار رپیدماینر حل می‌شود. با استفاده از این روش داده‌های پرت، شناسایی و حذف می‌شوند. همچنین برای پر کردن مقادیر از دست‌رفته از روش جایگذاری با مقدار میانگین و تخمین داده‌های از دست رفته با روش k- نزدیکترین همسایگی استفاده شده است، در اینجا مقدار k برابر با ۵ در نظر گرفته شده است. در روش جایگذاری با مقدار میانگین، مقادیر تهی با مقدار میانگین ویژگی جایگزین می‌شوند، در روش تخمین نیز برای پر کردن مقدار تهی یک روش رده‌بندی مورد نیاز است که در اینجا k- نزدیکترین همسایگی است که این مقدار با ۵ نزدیکترین همسایگی اش تخمین زده می‌شود.

به دلیل اینکه تعداد مشتریان رویگردان ۶۳ نفر در مقابل تعداد مشتریان غیررویگردان ۴۳۲۰ نفر هستند، با مشکل نامتوازن بودن داده‌ها از نظر تعداد افراد رویگردان مواجه شده و برای رفع این مشکل، لین و همکاران (۲۰۱۲) برای متوازن کردن مجموعه داده، استفاده از زیرنمونه‌گیری تصادفی از کلاس اکثریت را به‌عنوان استراتژی موثر پیشنهاد می‌دهند. در اینجا از روش نمونه‌گیری بوتسترپینگ در نرم‌افزار رپیدماینر استفاده شده است. در این روش به صورت تصادفی و با جایگذاری از رکوردهای مشتریان نمونه‌گیری می‌شود. پس از انجام مراحل این بخش، با استفاده از درخت تصمیمی که دارای پارامترهای شاخص جینی و عمق ۲۰ است، ارزیابی انجام شده و از شاخص‌های ارزیابی استفاده می‌شود که بتوان آنها را با یکدیگر مقایسه کرد. در این درخت از روش اعتبارسنجی تقسیمی استفاده شده است و ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای تست در نظر گرفته شده است. این شاخص‌ها و روش‌های اعتبارسنجی در قسمت ارزیابی مدل در گام پنجم به‌طور کامل توضیح داده خواهند شد.

در جدول شماره ۲ بهترین دقت به دست آمده از انتخاب داده، پاکسازی داده و نمونه‌گیری آورده شده است.

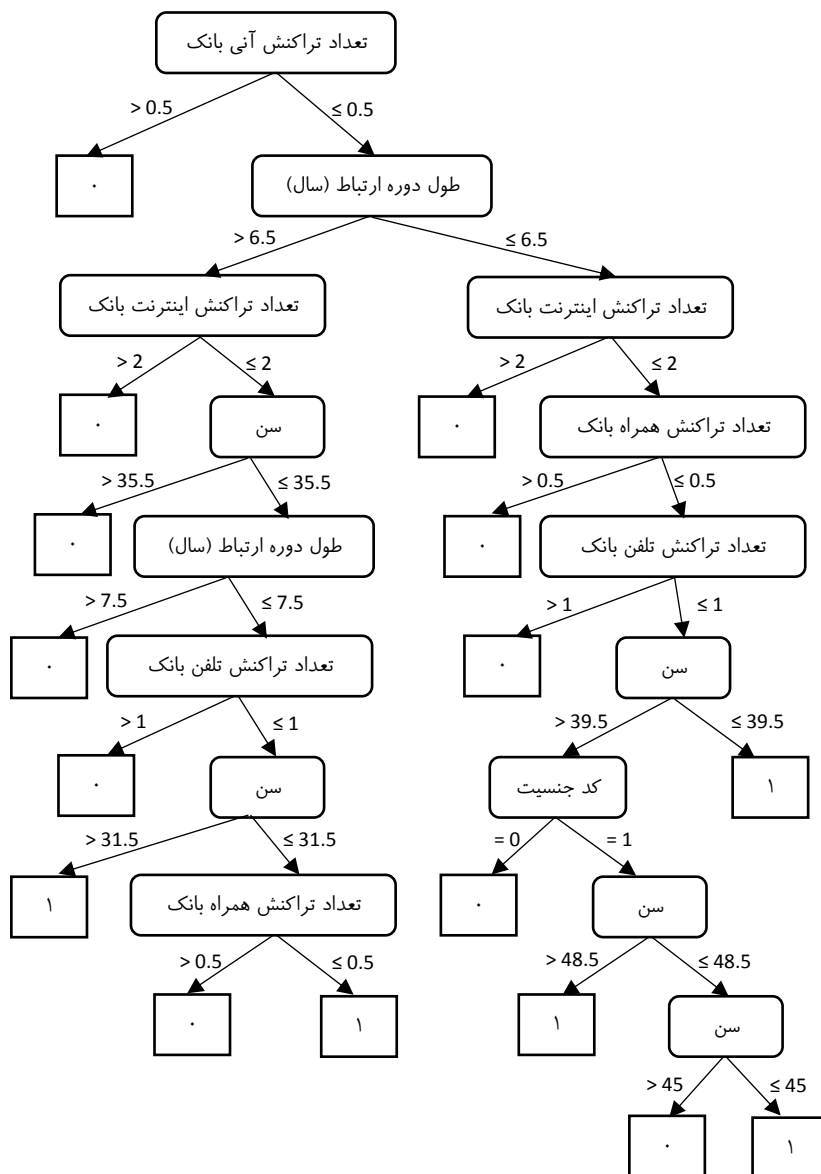
جدول شماره ۲: مقدار معیارهای ارزیابی بعد از پاکسازی و نمونه‌گیری

Noise & Outliers	Missing values	Sampling	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
LOF	Replace (AVG)	Bootstrap	۹۹/۳۱	۷۲/۲۲	۸۶/۶۷	۷۸/۷۹
	Impute (k=5)		۹۹/۳۱	۷۲/۲۲	۸۶/۶۷	۷۸/۷۹

با توجه به یکسان بودن نتایج ارزیابی، روش جایگذاری با مقدار میانگین برای پر کردن داده‌های از دست رفته انتخاب می‌شود.

### گام چهارم: مدل‌سازی

مدل‌سازی شامل انتخاب و به کارگیری روش مناسب مدل‌سازی، سنجیدن پارامترهای مدل برای رسیدن به نتایج بهینه می‌باشد (لاروس، ۲۰۰۵). در تحقیق حاضر با استفاده از روش درخت تصمیم بر روی مجموعه داده به دست آمده از مرحله آماده‌سازی داده، مدل‌سازی انجام می‌شود. برای انتخاب پارامترها بهترین دقت به دست آمده به عنوان معیار در نظر گرفته شده است. بهترین مدل درخت تصمیم به دست آمده در شکل شماره ۳ آورده شده است.



شکل شماره ۳: مدل درخت تصمیم

این مدل درخت تصمیم مشخصه‌های مشتریان رویگردان را به گروه‌های زیر دسته‌بندی می‌کند:

۱. مشتری که از طریق آنی بانک تراکنشی نداشته باشد و طول دوره ارتباط او کمتر و مساوی ۶/۵ سال و تعداد تراکنش اینترنت بانک او کمتر و مساوی ۲ و از طریق همراه بانک تراکنشی نداشته باشد و تعداد تراکنش تلفن بانک او کمتر و مساوی ۱ و سن او کمتر و مساوی ۳۹/۵ سال باشد، رویگردان است.
۲. مشتری که از طریق آنی بانک تراکنشی نداشته باشد و طول دوره ارتباط او کمتر و مساوی ۶/۵ سال و تعداد تراکنش اینترنت بانک او کمتر و مساوی ۲ و از طریق همراه بانک تراکنشی نداشته باشد و تعداد تراکنش تلفن بانک او کمتر و مساوی ۱ و سن او بیشتر از ۴۸/۵ سال و جنسیت او مرد باشد، رویگردان است.
۳. مشتری که از طریق آنی بانک تراکنشی نداشته باشد و طول دوره ارتباط او کمتر و مساوی ۶/۵ سال و تعداد تراکنش اینترنت بانک او کمتر و مساوی ۲ و از طریق همراه بانک تراکنشی نداشته باشد و تعداد تراکنش تلفن بانک او کمتر و مساوی ۱ و سن او بیشتر از ۳۹/۵ سال و کمتر و مساوی ۴۵ سال و جنسیت او مرد باشد، رویگردان است.
۴. مشتری که از طریق آنی بانک تراکنشی نداشته باشد و طول دوره ارتباط او بیشتر از ۶/۵ سال و کمتر و مساوی ۷/۵ سال و تعداد تراکنش اینترنت بانک او کمتر و مساوی ۲ و سن او کمتر و مساوی ۳۵/۵ سال و بیشتر از ۳۱/۵ سال و تعداد تراکنش تلفن بانک او کمتر و مساوی ۱ باشد، رویگردان است.
۵. مشتری که از طریق آنی بانک تراکنشی نداشته باشد و طول دوره ارتباط او بیشتر از ۶/۵ سال و کمتر و مساوی ۷/۵ سال و تعداد تراکنش اینترنت بانک او کمتر و مساوی ۲ و سن او کمتر و مساوی ۳۱/۵ سال و تعداد تراکنش تلفن بانک او کمتر و مساوی ۱ و از طریق همراه بانک تراکنشی نداشته باشد، رویگردان است.

ارزیابی و اعتبارسنجی این مدل در گام بعد بیان شده است.

### گام پنجم: ارزیابی

ارزیابی مدل‌های به‌دست آمده از مرحله مدل‌سازی به منظور بررسی کیفیت و موثر بودن قبل از اینکه به کار گرفته شوند (لاروس، ۲۰۰۵). برای تعریف شاخص‌های ارزیابی accuracy، precision، recall، f\_measure ابتدا مفاهیم FN، FP، TN و TP توضیح داده می‌شود.

TN: بیانگر تعداد رکوردهایی است که دسته‌ی واقعی آنها منفی بوده و الگوریتم دسته‌بندی نیز دسته آنها را به درستی برچسب‌گذاری کرده است.

FP: این مقدار نشان‌دهنده تعداد رکوردهایی است که دسته‌ی واقعی آنها منفی بوده و الگوریتم دسته‌بندی دسته‌ی آنها را به اشتباه مثبت تشخیص داده است.

FN: این مقدار نشان‌دهنده تعداد رکوردهایی است که دسته‌ی واقعی آنها مثبت بوده و الگوریتم دسته‌بندی دسته آنها را به اشتباه منفی تشخیص داده است.

TP: این مقدار نشان‌دهنده تعداد رکوردهایی است که دسته‌ی واقعی آنها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آنها را به درستی برچسب‌گذاری کرده است (کرامتی و همکاران، ۲۰۱۴).

شاخص‌های ارزیابی accuracy، precision، recall، f\_measure به صورت زیر تعریف می‌شوند (هان و همکاران، ۲۰۱۲).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})$$

$$\text{Precision (class = yes)} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall (class = yes)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F\_measure} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

برای بالا بردن اعتبار مدل در این تحقیق از روش اعتبارسنجی دسته‌ای<sup>۱</sup> استفاده شده است. در این روش تمام مجموعه داده هم برای آموزش و هم برای تست در نظر گرفته می‌شود. در این روش داده‌های اولیه به صورت تصادفی به k زیرمجموعه با اندازه تقریباً مساوی تقسیم می‌شوند. در تکرار i ام، قسمت Di به عنوان مجموعه تست و قسمت‌های باقی‌مانده برای آموزش مدل به کار می‌روند و مدل بر اساس آنها ساخته می‌شود. از هر کدام از این قسمت‌ها به تعداد مساوی برای آموزش و یک بار برای تست استفاده می‌شود (هان و همکاران، ۲۰۱۲). به طور معمول مقدار k برابر با ۱۰ در نظر گرفته می‌شود. در این تحقیق نیز مقدار k برابر با ۱۰ در نظر گرفته شده است. در این تحقیق با استفاده از روش اعتبارسنجی دسته‌ای مدل‌سازی انجام شد. بهترین نتایج پارامترهای ارزیابی برای مدل‌سازی درخت تصمیم در ادامه آورده شده است:

<sup>۱</sup> Cross-Validation



جدول شماره ۳: نتایج اعتبارسنجی مدل‌سازی درخت تصمیم بر روی مجموعه داده

(%) Accuracy	Precision (%)	(%) Recall	F-measure (%)
۹۹/۵۵	۸۹/۳۹	۸۵/۶۷	۸۶/۴۶

### گام ششم: به‌کارگیری نتایج

این گام شامل کاربرد عملی مدل‌های ساخته شده می‌باشد (لاروس، ۲۰۰۵). این گام در حیطه این تحقیق نیست و از آن صرف‌نظر شده است.

### نتیجه‌گیری و پیشنهادات

در این تحقیق ابتدا با مرور ادبیات پیشین، عوامل مرتبط با رویگردانی مشتری شناسایی شدند. سپس مدل مفهومی این تحقیق به‌دست آمد. سنجه‌های متناسب با هر یک از عوامل مدل تحقیق از ادبیات پیشین و با نظر خبرگان تعیین شد. داده‌های مرتبط با این سنجه‌ها از پایگاه داده بانک انصار استخراج و با استفاده از روش کریسپ به ساخت مدل پیش‌بینی رویگردانی مشتری از خدمات بانکداری الکترونیک با استفاده از روش درخت تصمیم پرداخته شد. مشخصه‌های مشتریان رویگردان از خدمات بانکداری الکترونیک با استفاده از این درخت تصمیم شناسایی شد.

به‌علت استفاده از پایگاه داده بانک امکان بررسی همه عوامل تاثیرگذار بر رویگردانی مشتری میسر نبود. محدودیت تحقیق این بود که تنها به بررسی عواملی پرداخته شد که داده‌های مرتبط با آنها در پایگاه داده وجود داشت. همچنین بدلیل حجم بسیار زیاد پایگاه داده و محرمانه بودن برخی از آنها، استخراج داده‌های مرتبط با تحقیق زمان‌بر بوده است. در تحقیقات آتی می‌توان به شناسایی عوامل دیگر مرتبط با رویگردانی مشتری با استفاده از پایگاه داده‌های صنایع دیگر پرداخت. همچنین مدل پیش‌بینی رویگردانی مشتری را در آن صنایع به‌دست آورد. پیشنهاد می‌شود در تحقیقات آتی با استفاده از تکنیک‌های دیگر داده‌کاوی نظیر شبکه‌های عصبی، نیو بیز، ماشین‌های بردار پشتیبان مدل‌سازی انجام شود.

## منابع و مآخذ

- احمدی نژاد، مصطفی (۱۳۹۱). طراحی مدل کاهش رفتار رویگردانی مشتریان (مطالعه موردی بانک تجارت)، تهران: دانشگاه پیام نور، رساله دکتری مدیریت بازرگانی.
- ایمانی خواه، سید مهدی (۱۳۹۱). بررسی عوامل موثر بر بهبود و ارتقا تصویر ذهنی مشتریان از خدمات بانکداری الکترونیک (مورد مطالعه: بانک رفاه)، دانشگاه پیام نور، پایان نامه کارشناسی ارشد مدیریت بازاریابی.
- توکلی، احمد، مرتضوی، سعید، کاهانی، محسن و حسینی، زهرا (۱۳۸۹). به کارگیری فرآیند داده کاوی برای پیش بینی الگوهای رویگردانی مشتری در بیمه، چشم انداز مدیریت بازرگانی، پیاپی ۳۷، شماره ۴، ص ۴۱-۵۵.
- جهان بکام، سعیده (۱۳۸۶). بررسی میزان پذیرش بانکداری الکترونیک در بین مشتریان بانک های شهر شیراز با استفاده از مدل پذیرش تکنولوژی، دانشگاه پیام نور، پایان نامه کارشناسی ارشد مدیریت دولتی.
- حاجی صفری، هانیه و الهی، شعیان (۱۳۹۰). چارچوبی جهت بررسی تاثیر تصمیم گیرندگان سازمانی بر موفقیت فرایند داده کاوی، فصلنامه علمی پژوهشی پژوهشگاه علوم و فناوری اطلاعات ایران، دوره ۲۷، شماره ۱، ص ۲۷۹-۳۰۰.
- غضنفری، مهدی، علیزاده، سمیه و تیمورپور، بابک (۱۳۸۷). داده کاوی و کشف دانش (چاپ اول)، تهران: انتشارات دانشگاه علم و صنعت ایران.
- کرامتی، عباس، سیدین اردبیلی، سید محسن و سهرابی، بابک (۱۳۸۸). تحلیل رویگردانی مشتریان بررسی وضعیت یکی از اپراتورهای تلفن همراه ایران با کمک روش های داده کاوی، فصلنامه علوم مدیریت ایران، سال چهارم، شماره ۱۴، ص ۶۳-۹۱.
- نوروزی، اشرف (۱۳۸۸). توسعه مدل پیش بینی کننده رفتار رویگردانی مشتری از خدمات بانکداری با کمک روش های داده کاوی، تهران: دانشگاه تربیت مدرس، پایان نامه کارشناسی ارشد مهندسی فناوری اطلاعات.
- Benoit, Dries F., & Van den Poel, D. (2012). Improving customer retention in financial services using kinship network information. *Expert Systems With Applications*, 39, pp. 11435-11442.
- Chen, S. C., & Huang, M. Y. (2011). Constructing credit auditing and control & management model with data mining technique. *Expert Systems With Applications*, 38, pp. 5359-5365.
- Chen, K., Hu, Y.-H. & Hsieh, Y.-Ch. (2014). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Inf Syst E-Bus Manage*, DOI 10.1007/s10257-014-0264-1
- Clemes, M. D., Gan, Ch., & Zhang, D. (2010). Customer switching behaviour in the Chinese retail banking industry. *International Journal of Bank Marketing*, 28(7), pp. 519-546.
- Hadden, J., Tiwaria, A., Roy, R., & Ruta, D. (2005). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34, pp. 2902 – 2917.
- Han, j., Kamber, M., & Pei, J. (2012). *DATA MINING Concepts and Techniques*, USA: Morgan Kaufmann.

- 
- Keramati, A., & Ardabili, S. M. S. (2011). Churn analysis for an Iranian mobile operator. *Telecommunications Policy*, 35, pp. 344-356.
  - Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozzafari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24, pp. 994-1012.
  - Laros, D. T. (2005). *An introduction to data mining*, New Jersey: John Wiley & Sons.
  - Liébana-Cabanillas, F., Noguera, R., Herrera, L.J., & Guillén, A. (2013). Analysing user trust in electronic banking using data mining methods. *Expert Systems With Applications*, 40, pp. 5439-5447.
  - Lane, P. C. R., Clarke, D. & Hender, P. (2012). On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data. *Decision Support Systems*, 53, pp. 712-718.
  - Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems With Applications*, 38, pp. 15273-15285.
  - Risselada, H., Verhoef, P. C., & Bijmolt, T. H. A. (2010). Staying Power of Churn Prediction Models. *Journal of Interactive Marketing*, 24, pp. 198-208.
  - Seng, J.-L., & Chen, T.C. (2010). An analytic approach to select data mining for business decision. *Expert Systems With Applications*, 37, pp. 8042-8057.
  - Tsai, Ch.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks, *Expert Systems with Applications*, 36, pp. 12547-12553.
  - Tsai, Ch.-F., & Chen, M.-Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems With Applications*, 37, pp. 2006-2015.
  - Verbeke, W., Martens, D., Mues, Ch., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems With Applications*, 38, pp. 2354-2364.
  - Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems With Applications*, 38, pp. 1999-2006.
  - Sharma, A., & Panigrahi, P. K. (2011). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications*, 27(11), pp. 26-31.
  - Yu, X., Guo, Sh., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems With Applications*, 38, pp. 1425-1430.
  - Zan, M., Shan, Z., Li, L., & Ai-jun, L. (2007). A Predictive Model of Churn in Telecommunications Based on Data Mining. *IEEE International Conference on Control and Automation ThAI-2 Guangzhou*, pp. 809- 813.