

چشم‌انداز مدیریت بازرگانی

شماره ۴- پیاپی ۳۷- زمستان ۱۳۸۹

صص ۴۱-۵۵

## به کار گیری فرایند داده کاوی برای پیش بینی الگوهای رویگردانی مشتری در بیمه

احمد توکلی\*، سعید مرتضوی\*\*، محسن کاهانی\*\*\*، زهرا حسینی\*\*\*\*

### چکیده

توسعه پارادایم بازاریابی دانش محور، تقاضا برای پیوستن داده کاوی به فرایندهای تجاری کسب و کار به ویژه در حوزه‌های مختلف مدیریت روابط با مشتری را به دنبال داشته است. یکی از حوزه‌های جدید و مورد اقبال در این زمینه، "مدیریت رویگردانی مشتری" می‌باشد. با اعمال فرایند داده کاوی در مقام یکی از فناوریهای مدیریت دانش، می‌توان طی کاوش در پایگاههای داده حجیمی از تعاملات ثبت شده سازمان با مشتریان که منعکس کننده تمایلات و رفتارهای واقعی آنان است به استخراج الگوهایی برای پیش بینی رفتار رویگردانی مشتری نائل آمد و مدیران را در اخذ تصمیمات لازم برای حفظ این مشتریان و کاهش روند ریزش آنان یاری داد. پژوهش حاضر با هدف تبیین قابلیت‌های داده کاوی در مدیریت رویگردانی مشتری، و با بهره گیری از متدولوژی استاندارد داده کاوی CRISP-DM، به کاوش در پایگاههای داده یکی از شرکتهای سهامی عام بیمه ای در رشته بیمه آتش سوزی پرداخته است. نتایج نشان می‌دهد کانال جذب مشتری عامل اصلی پیش بینی کننده رویگردانی یا ماندگاری مشتری در شرکت بوده و در مراتب بعد سابقه خرید و کاربری مکان بیمه شده به عنوان عوامل پیش‌بینی کننده رویگردانی قرار می‌گیرند.

**کلید واژه‌ها:** داده کاوی، مدیریت رویگردانی مشتری، دسته بندی.

---

تاریخ دریافت مقاله: ۸۸/۱۱/۲۵ تاریخ پذیرش مقاله: ۸۹/۰۴/۱۰

\* استادیار گروه مدیریت دانشگاه فردوسی مشهد.

\*\* دانشیار گروه مدیریت دانشگاه فردوسی مشهد.

\*\*\* دانشیار گروه کامپیوتر دانشگاه فردوسی مشهد.

\*\*\*\* دانش‌آموخته کارشناسی ارشد مدیریت بازرگانی دانشگاه فردوسی مشهد (نویسنده مسئول).

Email: sz\_hoseiny@yahoo.com

## مقدمه

روند تشدید رقابت و کاهش نرخ وفاداری مشتریان در دهه‌های اخیر، به ظهور پارادایمی در حوزه مدیریت بازاریابی منجر شده که بنگاهها را به گذار از رویکرد محصول محوری و اتخاذ نگرش مشتری محوری یا به عبارت دقیقتر، تعریف راهبردهای بازاریابی بر پایه نگاه از بیرون به درون سازمان، ترغیب می‌کند. مجموعه رهیافتهایی که حول محور این پارادایم مطرح شده‌اند، بر درک نیازهای مشتری و نه تمرکز بر مشخصات محصول استوار است [۸].

بسیاری از کسب و کارهای فعال در بازارهای رقابتی که بقای خود را به تداوم جریان سود حاصل از مشتری وابسته می‌بینند به ضرورت اتخاذ این رویکردها پی برده و به اقبال فزاینده از تدوین راهبردهای کسب و کار مشتری محور روی آورده اند [۱۴]. از جمله این کسب و کارها صنعت بیمه است. تشدید رقابت و بازارهای اشباع شده در کنار رشد سریع فناوری اطلاعات، گرایش به بازاریابی مشتری محور را برای شرکتهای فعال در این صنعت به ضرورتی انکار ناپذیر مبدل کرده است.

هنگامی که یک شرکت راهبرد مشتری مداری را جایگزین راهبرد محصول محوری می‌کند، "مدیریت روابط با مشتری" از جایگاه ویژه‌ای برخوردار می‌شود [۱۶]. هدف غایی مدیریت روابط با مشتری، درک سازوکارهای کسب سود بیشتر از ناحیه مشتریان و حفظ گروههای سودآور است. این فرایند پویا رابطه سازمان و مشتری را به سمتی هدایت می‌کند که مشتری ضمن ادامه مبادلات تجاری سودمند با سازمان از شرکت در مبادلات غیرسودمند برای سازمان منصرف شود [۸].

مطابق قانون پارتو، هشتاد درصد سود هر شرکت توسط ۲۰ درصد از مشتریان آن تأمین می‌شود. لذا شناسایی این گروه از مشتریان برای مطالعه نیازها و ترجیحات آنان در راستای تدوین راهبردهای مناسب ارتباطی جهت جذب، وفادار سازی و رضایتمندی آنان ضروری است [۱۶]. از آنجا که جذب مشتری جدید فرایندی دشوار، پرهزینه و زمانبر می باشد، این باور به شکل فزاینده‌ای مورد قبول قرار گرفته که در آینده بهترین راهبردهای بازاریابی بر حفظ مشتریان موجود و ممانعت از رویگردانی آنان معطوف می‌شود چرا که مطالعات نشان می‌دهد تغییر مختصری در نرخ نگهداشت مشتری تأثیرات معناداری بر سودآوری شرکت خواهد داشت [۱۴].

سنگ بنای راهبردهای مدیریت ارتباط با مشتری در حالت کلی و راهبردهای حفظ مشتری به‌عنوان جزئی از آن را "دانش مشتری" تشکیل می‌دهد. دانش مشتری به الگوهایی که ترجیحات واقعی و ترتیبات رفتاری وی را می‌نمایاند، گفته می‌شود [۱۳]. این دانش در کسب و کارهای کوچک با تعداد محدودی از مشتریان حاصل هوش تجاری صاحب حرفه می‌باشد اما در کسب و کارهایی با دهها یا صدها هزار مشتری که سازمان از سیستمهای اطلاعاتی گسترده ثبت

تعاملات مشتری بهره می‌برد، "داده کاوی" نقش تولید سیستماتیک هوش تجاری را برعهده گرفته است [۱].

## مبانی نظری و پیشینه موضوع

### رویکردانی مشتری

بسیاری از کسب و کارهای رقابتی به درستی دریافته‌اند که برای بقاء در صنعت، لازم است حفظ مشتریان ارزشمند، در مرکز راهبردهای مدیریتی آنان قرار گیرد. ارزش اقتصادی حفظ مشتری در مطالعاتی مورد بررسی قرار گرفته است. ریچهد و ساسر، در پژوهشی نشان داده‌اند که افزایش ۵ درصدی در نرخ نگهداشت مشتری به افزایش ۸۵ درصدی در سود بانکها و افزایش ۵۰ درصدی در حق واسطه‌گری شرکت‌های بیمه منجر می‌شود. نتایج مشابهی از مطالعه وندن پل و لاریور، که اثرات مالی افزایش یک درصدی در نرخ جذب مشتری را بررسی کرده‌اند نیز قابل دریافت است. این قاعده حاکی از نقش "مدیریت رویگردانی مشتری" در تضمین بقای سازمان است. مفهوم رویگردانی مشتری در ادبیات این موضوع، به دو صورت رویگردانی کلی و جزئی مطرح شده است. رویگردانی کلی به "گرایش مشتری به قطع مبادلات تجاری خود با شرکت در مقطعی از زمان" [۴] و نیز "تمایل مشتری به انتقال تعاملات تجاری خود به یک بنگاه رقیب" [۵] اطلاق شده و رویگردانی جزئی "کاهش مستمر ارزش طول عمر مشتری" [۷] تعریف شده است.

در مدل "چرخه سیستم مدیریت روابط با مشتری"، که قائل به وجود چهار رکن اصلی در مدیریت روابط با مشتری شامل: شناسایی مشتری، جذب مشتری، نگهداری مشتری و توسعه مشتری است تجزیه و تحلیل رویگردانی مشتری، بخشی از برنامه‌های وفادارسازی در رکن حفظ مشتری می‌باشد [۱۰]. از منظر هوش تجاری، فعالیتهای مدیریت رویگردانی مشتری شامل دو دسته اصلی از مدل‌های تحلیلی می‌باشد: ۱. پیش بینی گروه‌هایی از مشتریان که مستعد رویگردانی هستند و ۲. تشخیص اثر بخش‌ترین واکنش در این موقعیت (حتی اگر این واکنش سکوت باشد) [۷]. در مطالعه حاضر نمونه‌ای از مدل‌های دسته اول ارائه می‌شود.

بارز و وندن پل، مدیریت رویگردانی مشتری را متأثر از دو نگرش مجزا معرفی می‌کنند: رویکرد پیشرو و رویکرد انفعالی. در رویکرد انفعالی شرکت فقط هنگامی که یک مشتری در خواست قطع مراودات خود را دارد به ترغیب وی برای ماندگاری اقدام می‌کند. در مقابل شرکتی که رویکرد فعال را برمی‌گزیند، تلاش می‌کند تا پیش از اقدام مشتریان برای خروج، گروه‌های مستعد ترک سازمان را شناسایی کرده و برنامه‌های ترفیعی یا تشویقی خاصی را با هدف ممانعت از رویگردانی آنان به اجرا درآورد. تجارب مدیران نشان می‌دهد، احتمال اثر بخشی برنامه‌های ترفیعی منتج از

رویکرد دوم برای حفظ مشتریان بیش از رویکرد اول است، ضمن اینکه در این حالت هزینه‌های نهایی پایین تری به شرکت تحمیل می‌شود (بخشی از هزینه از دست دادن مشتری، به شکل سود احتمالی است که نصیب شرکتهای رقیب می‌شود به‌علاوه بازگویی تجربه تعامل ناموفق از سوی مشتری ناراضی می‌تواند اثر بخشی راهبردهای جذب مشتری سازمان را نیز تحت تأثیر قرار دهد). درعین حال مدل‌های پیش‌بینی کننده رویگردانی مشتری باید از ضریب صحت قابل قبولی برخوردار باشند تا از هدررفت منابع برای حفظ مشتریانی که به شکل بالقوه در صدد ترک سازمان نبوده‌اند پیشگیری شود.

همانگونه که قبلاً اشاره شد پیش‌بینی رفتار مشتری برای به اجرا درآوردن هر یک از راهبردهای مدیریت ارتباط با مشتری از جمله مدیریت رویگردانی، منوط به جمع‌آوری اطلاعات مشتری و تجزیه و تحلیل الگوهای رفتاری وی است. در کسب و کارهای بزرگ اجرای فرایند داده کاوی می‌تواند به استقرار سیستمی که بتواند قبل از اقدام مشتری برای ترک سازمان، واکنش بهنگام و مناسب را برگزیند، کمک کرده و چارچوبی برای اعمال کنترل پیش‌نگر، که شکل مطلوب کنترل‌های مدیریتی است را به وجود آورد.

### داده کاوی

ریژلسکی وهمکاران، فناوری داده کاوی را موجد انقلابی در عینیت بخشیدن به مفاهیم مدیریت روابط با مشتری در کسب و کارهای بزرگ معرفی می‌کنند. قابلیت داده کاوی در اعطای جنبه عملیاتی به مفاهیم مدیریت روابط با مشتری، این فناوری را در ردیف اجزای اصلی همه ارکان مدیریت روابط با مشتری تحلیلی و راهبردهای مطرح در آن از جمله مدیریت رویگردانی مشتری قرار داده است [۱۰].

داده کاوی را فرایند اکتشاف دانش از پایگاه‌های حجیم داده دیجیتال نامیده‌اند [۶]. دو هدف کارکردی داده کاوی توصیف و پیش‌بینی معرفی شده است. توصیف به یافتن الگوهای قابل تفسیر از داده‌ها متمرکز شده و پیش‌بینی به برآورد مقادیر یک متغیر هدف بر اساس ارزشهای یک یا چند متغیر دیگر می‌پردازد. داده کاوی برای پیشبرد هر دو این اهداف از مدل‌های متنوعی نظیر طبقه‌بندی، خوشه‌بندی، رگرسیون و نمایه سازی بهره می‌برد و در اجرای این مدل‌ها متکی به استفاده از فنون آماری، یادگیری ماشینی و فناوری پایگاه داده است [۱].

فنون داده کاوی در ارائه مدل‌های پیش‌بینی رویگردانی مشتری به شکل وسیعی مورد استفاده قرار گرفته است. در اکثر این تحقیقات، تکنیک داده کاوی اصلی مورد استفاده "طبقه‌بندی" و دو ابزار بسیار رایج برای اجرای این تکنیک "شبکه‌های عصبی" و "درختان تصمیم" می‌باشند [۱].

بیشتر پژوهشهای صورت گرفته در مقوله پیش‌بینی رویگردانی مشتری در حوزه‌های خدمات مالی، خرده‌فروشی و به ویژه ارتباطات سیار و خدمات آنلاین قرار می‌گیرند [۱۵]. از جمله این پژوهشها می‌توان به موارد زیر اشاره کرد:

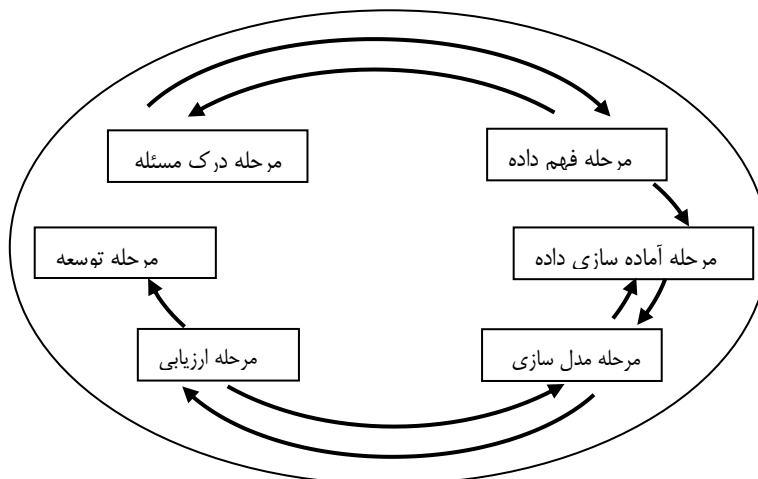
وانگ و همکاران، در پژوهشی روی داده‌های یک شرکت ارتباطات بی‌سیم با استفاده از ابزار درختان تصمیم به کاوش در ۶۰۰۰۰ رکورد مربوط به مراودات ۴۰۰۰ مشتری در یک دوره سه ماهه پرداخته‌اند. در قواعد حاصله از مدل، دو متغیر فاصله زمانی از آخرین اتصال مشتری به شبکه و تعداد دفعات اتصال به شبکه، دارای قابلیت پیش‌بینی رفتار رویگردانی یا ماندگاری مشتری تشخیص داده شده‌اند.

در پژوهش تی-سای وچن، در یک شرکت خدمات چند رسانه‌ای آنلاین، پس از کاوش در بیش از ۳۷۰۰۰ رکورد ثبت شده‌ی تعاملات مشتریان با ابزار درختان تصمیم دو متغیر زمان اتصال مشتری به شبکه و میزان تخفیف نرخ خدمات مهمترین عوامل پیش‌بینی کننده رویگردانی مشتری از شرکت تشخیص داده شده است.

لی ئو، با کاوش در داده‌های یک شرکت خطوط هوایی، با استفاده از ابزار درختان تصمیم قواعدی برای پیش‌بینی رفتار مشتری در سه کلاس: وفادار، دارای پتانسیل رویگردانی و رویگردان به دست آورد. متوسط درآمد، اشتغال در بخش دولتی، نرخ محصولات و استفاده از خدمات طول پرواز از جمله متغیرهایی بودند که در این پژوهش پیش‌بینی کننده رفتار رویگردانی مشتری معرفی شدند.

### روش تحقیق

اجرای هر فرایند داده کاوی بر دو گام اصلی استوار است: پیش پردازش داده‌ها و مدل‌سازی. زمان و نوع فعالیتهای لازم برای اجرای گام اول اقتضایی بوده و به ساختار مجموعه داده و همخوانی آن با اهداف تحقیق متکی است. اما گام دوم در همه پژوهشها تقریباً به شکل یکنواختی اجرا می‌شود. از اینرو با وجود متدولوژیهای استاندارد داده کاوی، اکثر پژوهشهای انجام گرفته در این حوزه با تغییراتی از این روشها استفاده کرده‌اند. در مطالعه حاضر نیز شکل تعدیل شده‌ای از متدولوژی استاندارد داده کاوی CRISP-DM (جدول شماره ۱)، به شرح زیر اجرا شده است.



جدول ۱. فرایند CRISP-DM

### اخذ و فهم مجموعه داده تحقیق

مجموعه داده در اختیار این تحقیق که جامعه آماری آن را تشکیل می‌دهد، شامل ۳۱۶۱۶ مشاهده است که هر مشاهده معادل یک قرارداد عقد بیمه آتش سوزی، در بازه زمانی ۸۴/۱۲/۱ تا ۸۶/۰۶/۳۱ می‌باشد. داده‌های هر مشاهده در یک رکورد بانک اطلاعاتی بیمه آتش سوزی در سیستم‌های اطلاعاتی شرکت مورد مطالعه ثبت بوده و از آن طریق اخذ شده است. اجرای این مرحله از فرایند داده کاوی که در گام پیش پردازش قرار می‌گیرد، شامل تبدیل فرمت داده‌ها و ساماندهی آن در قالب جداول اطلاعاتی قابل باز خوانی برای نرم افزار داده کاوی می‌شود. با طی این مرحله ۳۶ متغیر متمایز حاوی ویژگی‌هایی از خریدار، محصول، محل مورد بیمه و نمایندگی بیمه‌گر در هر رکورد این مجموعه داده شناسایی شدند. آماده سازی داده شامل انتخاب، تمیز کردن و تبدیل داده‌ها

به منظور تعقیب هدف تحقیق ایجاد تغییراتی در ساختار هر رکورد مجموعه داده اخذ شده ضروری بود. با بهره‌گیری از سه قلم اطلاعاتی شماره بیمه نامه قبلی، کد نمایندگی بیمه‌گر قبلی و کد شعبه سرپرستی در قرارداد قبلی، که سابقه خرید مشتری را منعکس می‌کردند و استفاده از پرس و جوهای نسبتاً پیشرفته SQL، کد "وضعیت خروج مشتری" استخراج و به هر رکورد مجموعه داده الحاق شد. واضح است که این کد وضعیت رفتار رویگردانی یا ماندگاری مشتری را در دو دوره متوالی خدمات نشان می‌دهد. غربالگری مجموعه داده اولیه به منظور حذف اطلاعات نامرتبط با هدف دیگر اقدام این مرحله است. به منظور ایجاد شرایط استفاده از فنون داده کاوی مورد نیاز این تحقیق، پاره‌ای دیگر از تبدیلات و تغییرات نظیر: محدود کردن مدت زمان قرارداد

بیمه از حالت روزشمار به سه مقدار کوتاه مدت، بلند مدت و مدت معمول (یکساله)، صورت گرفت. در نهایت خطاهای کدگذاری موجود در داده‌ها با توجه به دامنه مقادیر ممکن متغیرها و با اعمال فیلترهای چندگانه روی داده‌ها، تا حد امکان تصحیح شد.

### مدل سازی

با اتمام مراحل پیش پردازش داده، مرحله مدل سازی با دو زیرمرحله: تعیین کارکردی از داده کاوی که برای حل مسئله تحقیق مورد نیاز است، و انتخاب ابزار مناسب برای پیشبرد آن کارکرد، آغاز می‌شود.

از آنجا که هدف نهایی این پژوهش نگاشت مشتریان به دو گروه: "روگردان و غیر روگردان (ماندگار)" است، کارکرد داده کاوی متناسب با این هدف "دسته‌بندی" خواهد بود. کارکرد دسته‌بندی که از جمله کارکردهای دارای ماهیت پیش‌بینی‌گری داده کاوی است، به ساختن مدل‌هایی برای پیشگویی کلاس(رده) یک شیء بر اساس برخی ویژگیها یا صفات آن معطوف می‌شود. واژه‌های عمومی شیء و کلاس به تناسب موضوع پژوهش تفسیر می‌شوند. در این تحقیق منظور از شیء "مشتری" و منظور از کلاس یا رده "رفتار دو وجهی ماندگاری یا روگردانی مشتری" است.

در ایجاد یک مدل دسته‌بندی، مراحل: انتخاب تکنیک دسته‌بندی، انتخاب مجموعه داده مدل، تعیین متغیرهدف و معرفی متغیرهای پیشگویی پیش فرض(صفات)، طی می‌شود. پیش از این اشاره شد که در مرحله آماده‌سازی داده، متغیر دو ارزشی "وضعیت خروج مشتری" به آن بخش از مجموعه داده در اختیار که بررسی روگردانی یا ماندگاری مشتری در بازه زمانی داده‌های تحقیق برای آنها ممکن بود الحاق شد. به این طریق یک زیر مجموعه ۱۰۵۱۹ موردی از مشاهدات تحقیق به عنوان: "مجموعه داده مدل روگردانی" انتخاب شدند.

متغیر هدف در مدل حاضر کد دو ارزشی "وضعیت خروج مشتری"، و صفات شامل هفت فیلد معرف ویژگیهای جمعیت‌شناختی یا رفتاری متقاضیان خرید بیمه آتش سوزی یعنی: کاربری محل بیمه شده، شهرستان محل مورد بیمه، خرید یا عدم خرید پوششهای تکمیلی، ماهیت بیمه گذار، سابقه خرید، نوع حریق و ارزش بیمه شده، که از هر رکورد از مشاهدات ثبت شده قابل استخراج است، می‌باشد.

قبلاً اشاره شد که رایج‌ترین ابزارهای مورد استفاده در دسته‌بندی مشتریان بر اساس رفتار روگردانی "شبکه‌های عصبی" و "درختهای تصمیم" بوده‌اند. با در نظرگرفتن این نکته که الگوهای استخراج شده از مدل درخت تصمیم به شکل دنباله‌هایی از قوانین اگر-آنگاهی می‌باشند، به صورت کارآمدتری زمینه تدوین راهبردهای بازاریابی برای هر دسته از مشتریان را بر

طبق ویژگی‌های جمعیت‌شناختی و رفتاری آنان فراهم می‌کنند. لذا ابزار "درخت تصمیم" به عنوان گزینه بهینه برای پیشبرد هدف مطالعه حاضر انتخاب شد.

#### درخت تصمیم

درخت تصمیم، مشاهدات وارد شده به مدل را با مرتب کردن آنها در درخت از گره ریشه به سمت گره‌های برگ (گره‌های انتهایی درخت که فقط از یک سو به سایرگره‌ها متصل هستند) دسته‌بندی می‌کند. هر گره داخلی (غیر برگ) از درخت، متناظر با یک صفت از مشاهدات بوده و هر شاخه‌ای که از آن گره خارج می‌شود متناظر با یک مقدار ممکن برای آن صفت می‌باشد. در نهایت به هر گره برگ از درخت، یک دسته‌بندی منتسب می‌شود. هر یک از مشاهدات تحقیق، از گره ریشه وارد درخت شده و پس از آزمون صفت مشخص شده توسط این گره و حرکت در شاخه متناظر با مقدار صفت در آن مشاهده، در یک کلاس جای می‌گیرد. این فرایند برای هر زیر درختی که گره جدید ریشه آن می‌باشد، تکرار می‌شود.

ساخت درخت تصمیم با الگوریتم‌های مختلفی میسر است. الگوریتم مورد استفاده برای تشکیل درخت تصمیم مدل این پژوهش، الگوریتم CHAID می‌باشد. این الگوریتم گره‌های داخلی درخت را بر اساس میزان همبستگی هر صفت با متغیر هدف سازماندهی می‌کند و در ایجاد گره‌های برگ، با توجه به اینکه متغیر هدف در مدل این تحقیق گسسته، کیفی و دو ارزشی است، برای انتساب هر یک از مشاهدات مدل به یکی از دو کلاس متغیر هدف (رویگردان و ماندگار) بر پایه صفات مشاهده، از آزمون استقلال بین متغیر هدف و هر یک از صفات مشاهده، بهره می‌برد. به منظور انجام این آزمون برای هر صفت یک جدول توافقی تشکیل می‌شود که تعداد سطرهای آن متناظر با رده‌های این صفت و دو ستون آن متناظر با رده‌های متغیر هدف می‌باشد. آماره آزمون از فرمول زیر محاسبه می‌شود.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

که در آن:

$O_{ij}$ : فراوانی مورد انتظار برای سلول واقع در سطر  $i$  و ستون  $j$ ؛

$e_{ij}$ : فراوانی مشاهده شده برای سلول واقع در سطر  $i$  و ستون  $j$ ؛

$r$ : تعداد سطرها و  $c$ : تعداد ستونهای آن است.



این آماره تقریباً دارای توزیع کی دو با  $(c-1)*(r-1)$  درجه آزادی است. مقدار آماره با مقدار بحرانی کی دو در جدول در سطح معناداری مشخص شده در مدل مقایسه شده و فرض صفر (فرض استقلال) رد یا قبول می شود.

### ارزیابی مدل

هنگام استفاده از کارکردهایی همچون دسته بندی که ماهیت پیش بینی دارند، ارزیابی مدل های استخراج شده ضروری است. در این تحقیق ارزیابی مدلها با استفاده از روش "ماتریس انطباقی" صورت گرفته است (جدول شماره ۲). در این روش، کل مجموعه داده مدل به دو قسمت داده آموزش و داده آزمون تقسیم می شود. نسبت این تقسیم برای این مطالعه با توجه به تعداد مشاهدات و قاعده معمول در تحقیقات مشابه ۷۰ به ۳۰ در نظر گرفته شده است. در روش ارزیابی ماتریس انطباقی، مجموعه داده آزمون به عنوان داده های ورودی، به خروجی مدل حاصل از داده آموزش، وارد شده و ضریب صحت مدل از فرمولهای زیر محاسبه می شود:

$$AccuracyRate = \frac{N_1 + N_2}{N_1 + N_2 + M_1 + M_2}$$

$$ErrorRate = \frac{M_1 + M_2}{N_1 + N_2 + M_1 + M_2}$$

که در آن :

$N_1$ : تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۱ قرار دارند و مدل هم کلاس آنان را به درستی پیش بینی کرده است.

$N_2$ : تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۲ قرار دارند و مدل هم کلاس آنان را به درستی پیش بینی کرده است.

$M_1$ : تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۱ قرار دارند و مدل کلاس آنان را به غلط ۲ پیش بینی کرده است.

$M_2$ : تعداد مشاهداتی از مجموعه داده آزمون که در کلاس ۲ قرار دارند و مدل کلاس آنان را به غلط ۱ پیش بینی کرده است.

جدول ۲. ماتریس انطباقی

پیش بینی مدل			
کلاس ۱		کلاس ۲	
$M_1$	$N_1$	کلاس ۱	موقعیت
$N_2$	$M_2$	کلاس ۲	واقعی مشاهده

### یافته‌های کاربردی

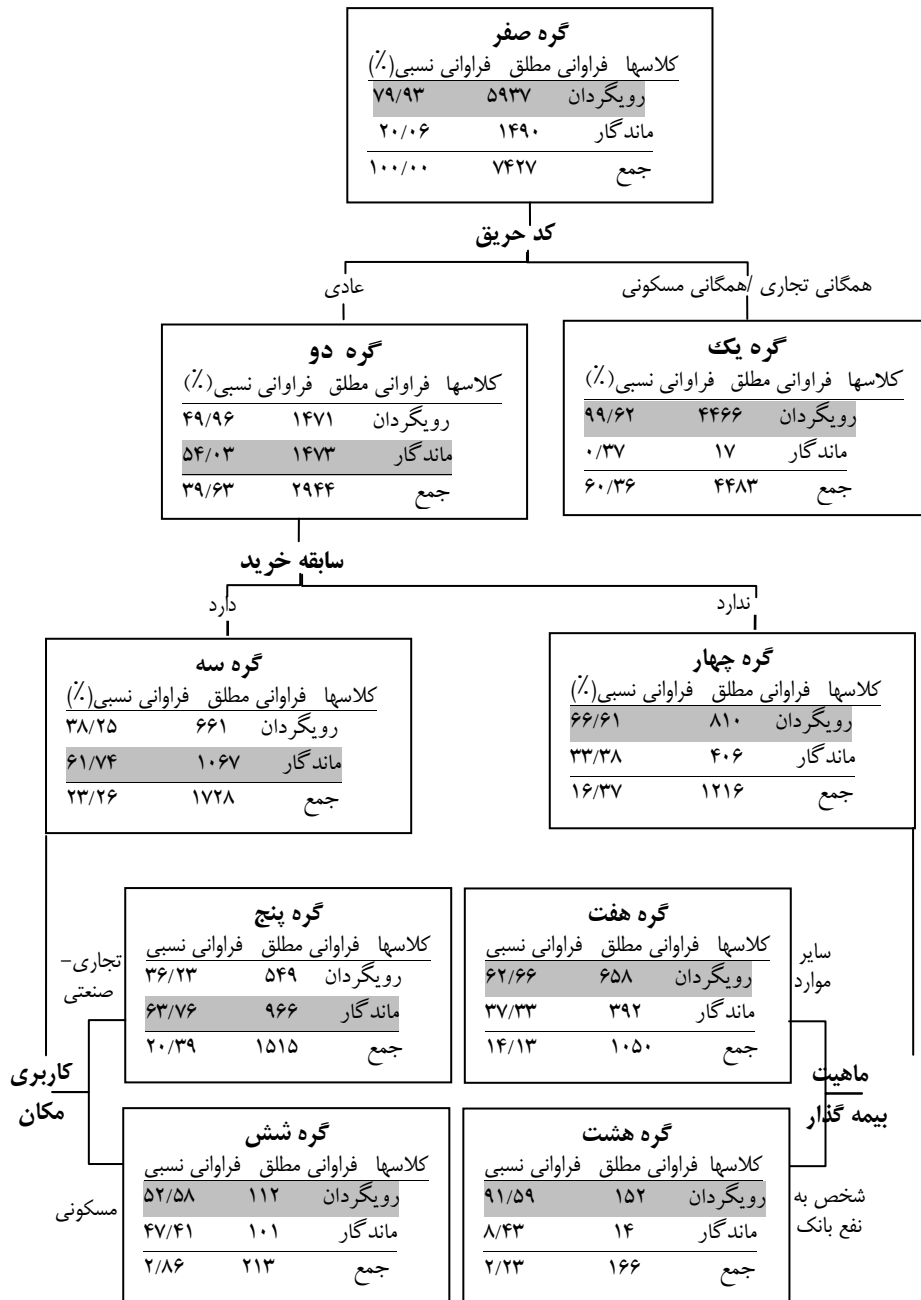
خروجی فرایند داده کاوی در پیش بینی رویگردانی مشتری در این تحقیق، شامل: استخراج صفات مؤثر در پیش بینی ارزش متغیر هدف و تعیین ضریب اثر نسبی هریک در این پیش بینی، تولید درخت معرف الگوهای دسته‌بندی مشتریان بر اساس صفات پیشگو در رده‌های متغیر هدف و ماتریس انطباقی ارزیابی ضریب صحت پیش بینی مدل می‌شود. جدول‌های ۳ تا ۶ این خروجی‌ها را نشان می‌دهد.

جدول ۳. صفات مؤثر در پیشگویی رویگردانی مشتری و ضرایب اهمیت نسبی هر یک

صفت	ضریب اهمیت	در ورود به الگو
پیش بینی تغییرات متغیر هدف		
نوع حریق	۰.۶۷	√
سابقه خرید	۰.۲۱	√
کاربری مکان بیمه شده	۰.۱۱	√
ماهیت بیمه گذار	۰.۰۱	√
شهرستان محل بیمه	۰.۰۰	-
خرید پوشش تکمیلی	۰.۰۰	-
ارزش مورد بیمه	۰.۰۰	-

در جدول ۳ دیده می‌شود که مدل، از میان هفت صفت پیشگوی پیش فرض، سه صفت شهرستان محل بیمه، خرید پوشش‌های تکمیلی و ارزش مورد بیمه را در پیش بینی متغیر هدف بی‌اهمیت تشخیص داده است. سایر صفات، درخت الگوهای رویگردانی یا ماندگاری مشتریان را به وجود آورده اند (جدول ۴). جدول ۵ صحت پیش بینی مدل برای هر یک از دو وضعیت متغیر هدف را در ماتریس انطباقی می‌سنجد و جدول ۶ پیش بینی‌های مدل را به صورت کلی ارزیابی می‌کند.

جدول ۴. درخت تشکیل شده برای مدل پیش بینی رویگردانی مشتری



جدول ۵. ماتریس انطباقی ارزیابی مدل رویگردانی

	ماندگاری	رویگردانی
ماندگاری	۶۰۸	۲۱
رویگردانی	۵۳۲	۱۹۳۱

جدول شماره ۶. نتیجه کلی ارزیابی مدل رویگردانی

نتیجه پیش بینی	تعداد مشاهدات (داده آزمون)	درصد
درست	۲۵۳۹	۸۲.۱۲
نادرست	۵۵۳	۱۷.۸۸
جمع	۳۰۹۲	۱۰۰

### تحلیل یافته‌ها

بررسی اولین خروجی مدل (جدول ۳)، اولویت نسبی هفت عامل پیش فرض را در تشخیص گروه‌های مشتریان رویگردان نشان می‌دهد. مشاهده می‌شود کدحریق خریداری شده که معرف کانال جذب مشتری به شرکت است عامل اصلی پیش بینی رفتار رویگردانی می‌باشد. متغیرهای سابقه خرید مشتری از شرکت و کاربری مکان بیمه شده در مراتب بعد و متغیر ماهیت بیمه‌گذار نیز با تأثیر نه چندان قابل توجه پیش بینی کننده رفتار رویگردانی هستند.

وارد نشدن چهار متغیر دیگر یعنی شهرستان محل بیمه شده، وضعیت خرید یا عدم خرید پوشش‌های مکمل و ارزش ریالی محل بیمه شده، به مدل نشان می‌دهد فراوانی رفتار رویگردانی در گروه‌های مختلف مشتریان در این سه متغیر، تفاوت معناداری با یکدیگر ندارد.

از بررسی درخت حاصل از مدل (جدول ۴)، نیز نتایج زیر قابل مشاهده است.  
- گره صفر درخت (گره ریشه) بیانگر نرخ خروج حدود ۸۰ درصدی مشتریان از شرکت در دوره مطالعه است.

- گره یک درخت (اولین برگ) بیانگر اولین قاعده مدل است: " قریب به ۱۰۰ درصد مشتریان بیمه آتش سوزی که از مسیر طرح‌های همگانی (خانه به خانه) جذب شرکت شده‌اند، پس از انقضای مدت قرار داد، سازمان را ترک می‌کنند".

- گره دو درخت بیانگر این است که نرخ خروج مشتریانی که از مسیر انعقاد قرارداد بیمه حریق عمومی (عادی) با شرکت در تعامل قرار گرفته‌اند در مقایسه با گروه قبل تقریباً به نصف کاهش می‌یابد.

- مقایسه گره‌های ۳ و ۴ درخت نشان می‌دهد در این رده از مشتریان نیز، نرخ خروج در گروه مشتریانی که یکبار سابقه خرید از شرکت را دارند در مقایسه با گروهی که بیمه حریق عمومی را برای اولین بار می‌خرند، به نصف کاهش می‌یابد.

- مقایسه برگ‌های ۵ و ۶ درخت، نشان می‌دهد که در رده مشتریانی که دارای سابقه خرید از شرکت بوده و از مسیر بیمه عمومی جذب شرکت می‌شوند، بیمه‌گذارانی که برای بیمه منازل خود

اقدام می‌کنند، بیش از آنان که محل کسب خود را بیمه کرده‌اند در معرض خطر از دست رفتن می‌باشند.

- از مقایسه برگه‌های ۷ و ۸ درخت ملاحظه می‌شود، مشتریانی که به تقاضای بانک اعطا کننده تسهیلات ناچار به بیمه ملک در رهن بانک شده‌اند، پس از رفع این اجبار معمولاً شرکت را ترک می‌کنند.

ذکر این نکته لازم است که کاربرد این نتایج در فرایند تصمیم‌گیری مدیران بازاریابی باید با توجه به میزان مطلوبیت ضریب صحت به دست آمده برای مدل در ماتریس ارزیابی آن (۸۲٪) در نگاه مدیران این سازمان صورت بگیرد.

### پیشنهادات کاربردی

ارزش یک مطالعه داده کاوی در قابلیت تأویل خروجی‌های آن به راهبردها و برنامه‌های اجرایی کسب و کار نهفته است. به این منظور پیشنهادات زیر که مبتنی بر دستاوردهای مدل این مطالعه هستند، برای جایگذاری در بسته برنامه‌های بازاریابی دانش محور شرکت مطرح می‌شود.

۱. مدیران سازمان می‌توانند با استفاده از این یافته تحقیق که "کانال جذب مشتری" عامل اصلی تعیین‌کننده ماندگاری مشتری در شرکت است، تخصیص منابع به برنامه‌های حفظ و وفادارسازی مشتریان را به صورت هدفمند درآورند؛ به این معنی که گروه‌های مشتریان جذب شده از کانالهای طرحهای همگانی و بانکها را در اولویت اجرای طرحها و برنامه‌های ترفیعی حفظ مشتری قرار دهند.

۲. بی‌تفاوت بودن رفتار رویگردانی نسبت به متغیر "ارزش ریالی مورد بیمه" (که در واقع بازتاب میزان سود مشتری برای شرکت است) نشان می‌دهد مشتریان موجد درآمد بالاتر برای سازمان با نرخی برابر با گروه مقابل سازمان را ترک می‌کنند. این نتیجه به طور قطع مطلوب سازمان نیست و لازم است مدیران تدابیر خاصی برای حفظ مشتریان سود آورتر سازمان اتخاذ کنند.

۳. فرار گرفتن ۶۰ درصد از کل مشاهدات، در رده مشتریانی که از مسیر طرح همگانی به سازمان پیوسته‌اند، نشان از موفقیت شرکت در طراحی برنامه‌های جذب مشتری دارد. اما ریزش کامل این مشتریان در دور بعد، نشان می‌دهد سازمان نتوانسته از این فرصت برای تثبیت یک جریان سود دائمی بهره ببرد. به نظر می‌رسد مدیران شرکت باید در جهت گیریهای راهبردی از جذب به حفظ مشتری معطوف شوند.

۴. از آنجا که نرخ رویگردانی در رده مشتریانی که دارای سابقه خرید خدمت بیمه هستند کمتر است، پیشنهاد می‌شود برنامه‌های وفادارسازی بیشتر با تأکید بر مشتریان جدید انجام پذیرد.

۵. با توجه به اینکه نرخ رویگردانی در بیمه حریق مکانهای تجاری کمتر از مسکونی می‌باشد، شرکت می‌تواند منابع خود در طراحی برنامه‌های ترفیعی را به سمت حفظ مشتریان متقاضی بیمه حریق مسکونی هدایت کند.

### محدودیت‌های تحقیق

از آنجا که نتایج تحقیقات داده کاوی از پایگاههای داده یک سازمان خاص به دست می‌آیند، قابلیت تعمیم به دیگر سازمانها را نخواهند داشت. برش زمانی داده‌ها در این تحقیق یک مقطع ۱۸ ماهه بوده و الگوهای حاصل از مدل، بر پایه داده‌های ثبت شده از تعامل شرکت و مشتری در این دوره به دست آمده است. بدیهی است هر نوع تغییر در خط مشی فروش شرکت در طول زمان، چنانچه ساختار داده ثبت شده را تغییر دهد می‌تواند الگوهای این مقطع را نامعتبر کند.

### پیشنهاداتی برای تحقیقات آتی

در برش ۱۸ ماهه تحقیق حاضر از پایگاههای داده شرکت بیمه، حداکثر سه تکرار خرید برای هر مشتری منعکس شده است. چنانچه مدیران شرکت مایل به استفاده از الگوهای دقیق‌تر رویگردانی یا وفاداری مشتریان باشند، می‌توان پژوهش را در قلمرو زمانی وسیعتری به انجام رسانید.

در صورتی که مدیران شرکت تمایل به مطالعه روندهای تغییر در عوامل ایجاد رویگردانی مشتریان داشته باشند، می‌توان با برداشت برشهای زمانی متناوب از پایگاههای داده شرکت و اجرای مکرر فرایند داده کاوی این تغییرات را از حیث ماهیت و شدت مطالعه کرده و از نتایج آن در تدوین چشم انداز و برنامه استراتژیک شرکت به نحو مقتضی بهره گرفت. چنانچه ساختار رکوردهای پایگاههای داده مربوط به رشته بیمه‌های مختلف شرکت به نحوی باشد که امکان تلفیق داده‌ها و تشکیل پروفایل مشتری میسر شود، مدیران سازمان می‌توانند از فرایند داده کاوی برای پیش‌بینی تأثیر رویگردانی مشتری از یک خدمت بیمه‌ای بر رفتار وی در سایر خدمات بیمه‌ای، استفاده کنند.

## منابع

۱. شهرابی، جمال (۱۳۸۶). *داده کاوی جلد اول*. تهران: جهاد دانشگاهی واحد صنعتی امیرکبیر.
2. Buckinx, W., & Van den Poel, D.(2005), "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting". *European Journal of Operation Research*, 164(1), 252-268.
3. Burez, J., & Van den Poel, D.(2007), "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services". *Expert System with Applications*, 32, 277-288.
4. Chandar, M., Laha, A., & Krishna, P. (2006), "Modeling churn behavior of bank customers using predictive data mining techniques". In National conference on soft computing techniques for engineering applications (SCT-2006), March 24-26, 2006.
5. Chen, M. Y., Tsai, C.F.(2010), "Variable selection by association rules for customer churn prediction of multimedia on demand". *Expert System with Applications*, 37, 2006-2015.
6. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P.(1996), "From data mining to knowledge discovery in databases". *AI Magazine*, 17(3), 37-54.
7. Gladys, N., Baesens, B., & Croux, C. (2009), "Modeling churn using customer lifetime value". *European Journal of Operation Research*, 197, 402-411.
8. Hung, S. Y., Yen, D., Wang, H. Y. (2006), "Applying data mining to telecom churn management". *Expert System with Applications*, 31, 515-524.
9. Liou, J.H. (2008), "A novel decision rules approach for relationship management of the airline market". *Expert System with Applications*, 36, 4374- 4381.
10. Ngai, E.W. T., Xiu, L., & Chau, D. C. K.(2008), "Application of data mining techniques in customer relationship management: A literature review and classification". *Expert System with Applications*, 36, 2592-2602.
11. Reichheld, F.F., & Sasser, W. E (1990), "Zero defections: quality comes to service". *Harvard Business Review*, 68, 105-111.
12. Rygielski, C., Wang, J.C., & Yen, D. C. (2002), "Data mining techniques for customer relationship management". *Technology in Society*, 24, 483-502.
13. Shaw, M.J., Subramaniam, C., Gek, W.T., & Welge, M.E.(2001), "Knowledge management and data mining for marketing". *Journal of Decision Support System*, 31, 127-137.
14. Tsai, C. F., Lu, Y. H. (2009), "Customer churn prediction by hybrid neural network". *Expert System with Applications*, 36, 12547-12553.
15. Van den Poel, D., & Lariviere, B. (2004), "Customer attrition analysis for financial services using proportional hazard models". *European Journal of Operation Research*, in press.
16. Wang, Y. F., Ding-An, C., Mei-Hua, H., Cheng-Jung, L., & I-Long, L. (2009), "A recommender system to avoid customer churn: A case study". *Expert System with Applications*, 36, 8071-8075.