پژوهشنامهٔ بیمه/ سال سی‌ودوم / شمارهٔ ۱/ بهار ۱۳۹۶/ شمارهٔ مسلسل ۱۲۵/ صفحات ۸۵ تا ۱۰۶
کد شناسایی دیجیتال(DOI): 10.22056/jir.2017.60158.1876

مدل‌سازی پیش‌بینی رویگردانی مشتریان صنعت بیمه (مطالعهٔ موردی)

تاریخ دریافت مقاله:۱۳۹۵/۰۶/۰۵

تاریخ پذیرش مقاله:۱۳۹۵/۱۲/۰۹

علی بقائی[۱]

منیره حسینی[۲]

چکیده

با اشباع بازارها، سازمانها دریافته‌اند که باید حفظ مشتریان، به‌خصوص مشتریان باارزش، در مرکز راهبردهای مدیریتی آنان قرار گیرد چراکه هزینهٔ جذب مشتریان جدید بیشتر از هزینهٔ نگهداری مشتریان موجود است. صنعت بیمه نیز از این امر مستثنی نیست و به دلیل هزینهٔ جابجایی اندک، با مشتریانی روبه‌رو است که تمایل دارند شرکت فعلی خود را تغییر دهند. پژوهش حاضر تأثیر متغیرهای طول ارتباط، تأخر خرید، فرکانس خرید، ارزش مالی، سودآوری و گروه محصولات خریداری‌شده را در ارزش‌گذاری مشتریان مورد مطالعه‌شده بررسی می‌کند. برای این کار از ابزار پیمایشی پرسشنامه به منظور آگاهی از نظرات خبرگان در خصوص متغیرهای مؤثر بر ارزش‌گذاری مشتریان استفاده شده است. نتایج نشان می‌دهد که متغیرهای فرکانس خرید، طول مدت همکاری و تعداد گروه‌های بیمه‌ای خریداری‌شده از منظر خبرگان بیشترین اهمیت را در ارزش‌گذاری مشتریان دارا هستند. سپس با استخراج متغیرها و عوامل مهم بر رویگردانی بیمه‌گذاران، تأثیر و اهمیت این عوامل بر روی رویگردانی مشتریان باارزش بررسی و در ادامه با استفاده از متغیرهای شناسایی‌شده در مرحلهٔ قبل به توسعهٔ مدل پیش‌بینی‌کنندهٔ رویگردانی پرداخته شده است. با مدلهای مختلف (شبکهٔ عصبی، درخت تصمیم، ماشین بردار پشتیبان، و رگرسیون لوژستیک) مدل‌سازی پیش‌بینی انجام و دقت مدلهای ساخته‌شده، ارزیابی شده است. نتایج نشان می‌دهد که مدل درخت تصمیم C5.0 نسبت به سایر مدلها دقت و صحت بالاتری در پیش‌بینی رویگردانی دارد.

---

۱. دانشجوی کارشناسی ارشد، گروه فناوری اطلاعات، دانشکدهٔ مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، a.baghaei@mail.kntu.ac.ir

۲. استادیار گروه فناوری اطلاعات، دانشکدهٔ مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی (نویسندهٔ مسئول)، hosseini@kntu.ac.ir

واژگان کلیدی: *رویگردانی مشتری، تحلیل و پیش‌بینی رویگردانی، ارزش‌گذاری مشتریان، صنعت بیمه.*

۱. مقدمه

امروزه، رقابت علاوه بر کاهش قیمت و تنوع محصولات، از طریق افزایش ارتباط با مشتریان حاصل می‌شود (Ryals, 2002; Teo et al., 2006). سیستمهای مدیریت ارتباط با مشتری، جایگزین استراتژیهای سنتی بازاریابی انبوه[1] شده‌اند و بازاریابی شخصی‌شده[2] را ارائه می‌کنند (Burez and Van den Poel, 2007). مطابق قانون پارتو، ۸۰ درصد سود هر شرکت، تنها توسط ۲۰ درصد از مشتریان حاصل می‌شود؛ لذا شناسایی این گروه از مشتریان جهت مطالعهٔ نیازها و ترجیحاتشان برای تدوین راهبردهای ارتباطی مناسب جهت حفظ، رضایت‌مندی و وفادارسازی آنان ضروری است (Wang et al., 2009).

به منظور شناسایی ارزش مشتریان، نیاز به جمع‌آوری اطلاعات آنها و داشتن ابزار تحلیل داده‌هاست. از سویی، با توجه به بهبود دسترسی به اطلاعات، مشتریان گذرا شده‌اند و دگرگزینی بین رقبا برای آنها آسان‌تر و کم‌هزینه‌تر شده است (Wiersema, 2013). شرکتهای بیمه نیز همانند سایر صنایع در بازارهای خدماتی که با هزینه جابجایی اندک روبه‌رو هستند، با مشکلاتی همچون حفظ مشتریان فعلی مواجه‌اند؛ با رویگردانی مشتریان از شرکت بیمه، بخشی از هزینه‌های از دست دادن مشتری، به شکل سودِ احتمالی نصیب شرکتهای رقیب می‌شود و همچنین بازگویی تجربهٔ تعامل ناموفق از سوی مشتری ناراضی می‌تواند از اثربخشی راهبردهای بازاریابی و جذب مشتریان بکاهد (توکلی و همکاران، ۱۳۸۹). بنابراین حفظ مشتریان فعلی از اهمیت بسیار زیادی برخوردار است و پژوهشی جامع را می‌طلبد و تحلیل مناسب رفتار رویگردانی و پیش‌بینی دقیق آن می‌تواند منجر به کسب سود بیشتر شود.

---

1. Mass Marketing
2. Individual Marketing

۲. مروری بر ادبیات پژوهش

در این بخش، در ابتدا مفاهیم اصلی در حوزهٔ پژوهش را تعریف می‌کنیم و سپس مروری بر ادبیات موضوع و پژوهشهای پیشین در این حوزه خواهیم داشت.

۲-۱. رویگردانی مشتری

جذب مشتریان جدید نیازمند هزینه‌های مختلفی همچون هزینه‌های جستجو، تبلیغات و بازاریابی است (Tsai and Lu, 2009) که این هزینه‌ها چندین برابر تلاش شرکت برای نگهداشت مشتریان موجود است (Sharma and Panigrahi, 2011)؛ از این رو باصرفه‌تر است که تمرکز بیشتر بر روی مشتریان فعلی باشد (Teo et al., 2006).

وفاداری مشتریان در صنایع خدماتی همچون بیمه نسبت به دیگر صنایع اهمیت بیشتری دارد (Kumar and Srivastava, 2013)؛ این صنعت با از دست دادن سهم بازار و مشتریان، با ریسکهای تجاری قابل توجهی روبه‌رو خواهد شد. حفظ مشتریان فعلی، مبتنی بر شناسایی سریع و به موقع مشتریانی است که می‌خواهند شرکت را ترک کنند یا به اصطلاح رویگردان شوند.

رویگردانی دارای تعاریف مختلفی است و تعریفی یکسان و قابل اعمال بر تمام صنایع، وجود ندارد؛ در واقع این مفهوم توسط کارشناسان، با توجه به ویژگیهای محیط صنعت و سازمان تعریف می‌شود (Reinartz, 2000). رویگردانی مشتری، گرایش ذاتی وی برای انصراف از ادامهٔ روابط تجاری با یک شرکت در یک دورهٔ زمانی است (Chandar et al., 2006). با توجه به این تعریف، مشتری رویگردان کسی است که تمامی فعالیتهایش (برای مثال حسابها و دفاتر) را با شرکت قطع کند (Chandar et al., 2006) یا حداقل تکرار خرید وی کمتر از میانگین باشد (Buckinx and Van den Poel, 2005) و ارزش طول عمرش به‌مرور در حال کاهش باشد (Glady et al., 2009).

با هدف توسعهٔ یک برنامهٔ مؤثر حفظ مشتری، مدلهای مورد استفاده باید تا حد ممکن دقیق باشند تا بتوانند مشتریان رویگردان را به‌درستی تشخیص دهند

(Coussement and De Bock, 2013). در غیراین صورت، اگر هزینه‌های انگیزشی برای مشتریانی که قصد رویگردانی ندارند صرف شود، این سیستمها بی‌فایده‌اند (Tsai and Lu, 2009).

۲-۲. داده‌کاوی

به منظور تحلیل و پیش‌بینی رویگردانی، تکنیکهای داده‌کاوی و تحلیل آماری به‌کار گرفته می‌شوند. روشهای داده‌کاوی فراوانی برای ایجاد یک مدل پیش‌بینی کارا و دقیق مطرح شده‌اند (Hung et al., 2006). دو هدف اصلی داده‌کاوی عبارت‌اند از: توصیف و پیش‌بینی (Ngai et al., 2009). در واقع توصیف، بر یافتن الگوهای قابل تفسیر انسانی که داده را توصیف می‌کنند تمرکز دارد و پیش‌بینی درگیر استفاده از تعدادی از متغیرها یا فیلدها در پایگاه‌داده به منظور یافتن مقادیر ناشناخته یا آتی سایر متغیرهای مورد علاقه است. این اهداف می‌تواند با استفاده از روشهای متنوعی مثل دسته‌بندی، خوشه‌بندی، رگرسیون و غیره حاصل شوند (Tsai and Lu, 2009). پیش‌بینی رویگردانی مشتری، به شکل یک مسئلهٔ دسته‌بندی در نظر گرفته می‌شود که هر مشتری در یکی از دو دستهٔ رویگردان یا غیررویگردان دسته‌بندی می‌شود (Zhang et al., 2012).

۲-۳. پژوهشها در حوزهٔ رویگردانی در صنعت بیمه

از جملهٔ پژوهشها، مطالعهٔ موریک و کاپکه[1] (۲۰۰٤) است که به بررسی بیمهٔ عمر در یکی از شرکتهای بیمه در کشور سوئیس پرداخته‌اند. در این پژوهش، محققان با سریهای زمانی، مدلی با بهره‌گیری از مشخصهٔ فراوانی سند/ معکوس فراوانی سند (TF/IDF[2]) ارائه کرده‌اند. در این پژوهش سه رویکرد مختلف ارائه شده است: رویکرد نخست، پیش‌بینی بدون داده‌های زمانی است. رویکرد دوم، پیش‌بینی بر مبنای سریهای زمانی و رویکرد نهایی استفاده از مشخصهٔ TF/IDF و ترکیب داده‌های زمانی

1. Morik and Kopcke
2. Term Frequency–Inverse Document Frequency

با اطلاعات سطح قرارداد و متغیرهای رفتاری مشتریان است. مشخصه‌های جدید تعریف‌شده بر پایهٔ مشخصهٔ TF/IDF دقت، روایی، و پایایی بالاتری نسبت به نتایج قبل دارند.

محمدی و همکاران (۲۰۱٤) به پیش‌بینی رویگردانی مشتریان بیمه با در نظر گرفتن ارزش طول عمر مشتری (CLV[1]) پرداخته‌اند (رشتهٔ بیمهٔ اتومبیل). در این پژوهش در ابتدا با استفاده از ۱۵ ویژگی و بهره‌گیری از الگوریتم خوشه‌بندی K- میانگین[2]، مشتریان در دو خوشهٔ رویگردانان و غیر رویگردانان، خوشه‌بندی شده‌اند و فاکتورهای مهم رفتاری مشتریان در تشخیص این دو خوشه مشخص شده‌اند. سپس با استفاده از ٦ فاکتور مهم، رفتار آیندهٔ مشتریان توسط رگرسیون لوژستیک پیش‌بینی شده است.

حسین‌خانی و همکاران (۱۳۹۳)، در پژوهشی به شناسایی عوامل مؤثر بر رویگردانی مشتریان در صنعت بیمه به روش میدانی و پرسشنامه در یکی از بیمه‌های خصوصی ایران پرداخته‌اند. در این پژوهش متغیرهای رفتاری، دموگرافی و ادراکی لحاظ شده‌اند و برای کاهش تعداد متغیرها از روش تحلیل عاملی استفاده شده است و سپس با روش تحلیل واریانس یک‌طرفه، تفاوت واریانسها در گروه‌های مختلف، تحقیق شده است. نتایج نشان می‌دهد که متغیرهای ادراک و آگاهی مشتریان، بیش از متغیرهای دموگرافی و سطح قرارداد بر رویگردانی مشتریان مؤثرند.

در پژوهش گانتر[3] و همکاران (۲۰۱٤)، با استفاده از متغیرهای اکتشافی پویا و تراکنشهای بیمه‌ای و بهره‌گیری از روش رگرسیون لوژستیک طولی، مدلی جهت پیش‌بینی ریسک رویگردانی ماهیانه در صنعت بیمه ارائه شده است. نویسندگان در این پژوهش از مدل جمعی تعمیم‌یافته (GAM[4]) به‌عنوان یک مرحلهٔ میانی به منظور تشخیص رابطهٔ غیرخطی بین لوجیت و متغیرهای اکتشافی استفاده کرده‌اند و این گونه

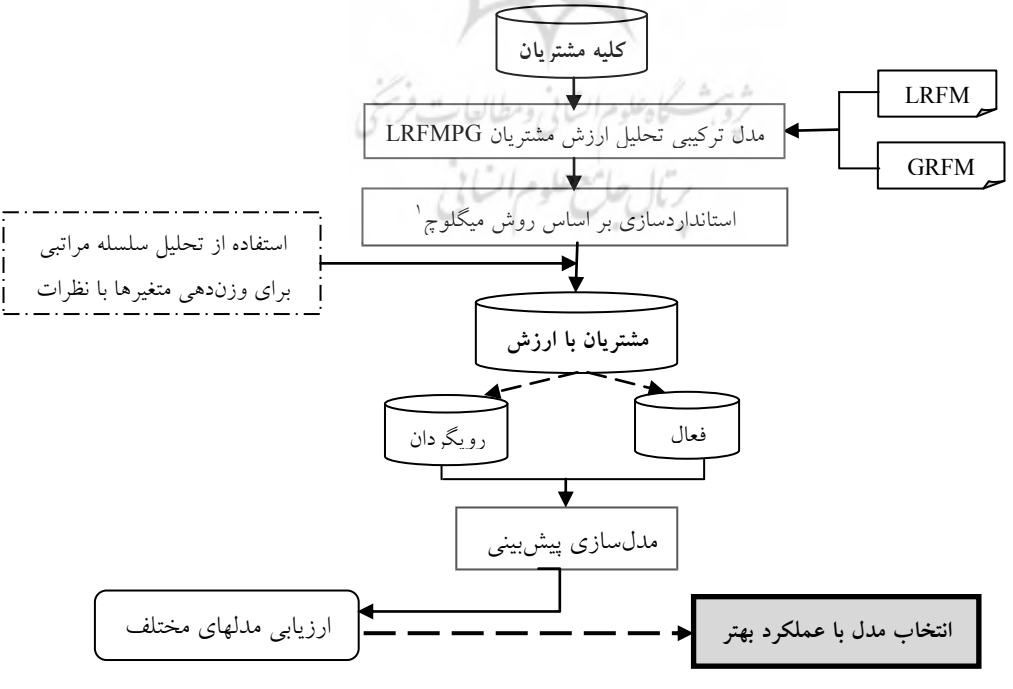
1. Customer Lifetime Value
2. K-Means
3. Günther
4. Generalized Additive Model

بعضی از متغیرهای مؤثر بر رویگردانی دوباره تعریف شده‌اند. سپس با فرض رابطهٔ خطی بین متغیرهای اکتشافی بازتعریف‌شده و لوجیت، مدل‌سازی انجام شده است.

در مطالعهٔ توکلی و همکاران (۱۳۸۹)، کاوش در پایگاه‌های دادهٔ یکی از شرکتهای بیمه‌ای در رشتهٔ آتش‌سوزی انجام شده است. نتایج حاکی از آن است که کد حریق خریداری‌شده که معرف کانال جذب مشتری است، عامل اصلی پیش‌بینی‌کنندهٔ رویگردانی یا ماندگاری مشتری در شرکت است. بعد از آن سابقهٔ خرید و کاربری مکان بیمه‌شده، عوامل پیش‌بینی‌کنندهٔ رویگردانی‌اند. همان‌طور که مشخص است، مطالعات اندکی به حوزهٔ تحلیل و پیش‌بینی رویگردانی در صنعت بیمه، پرداخته‌اند. مقالات انگلیسی حول این موضوع بسیار محدود بوده و مطالعات انجام‌شده در داخل کشور نیز عمدتاً با پرسشنامه و بدون استفاده از داده‌های واقعی صورت گرفته‌اند.

۳. روش تحقیق

طرح تحقیق مطالعهٔ حاضر در شکل ۱ آمده است.

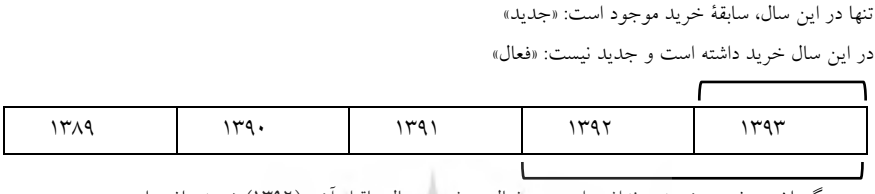


شکل ۱. طرح تحقیق.

مورد مطالعه‌شده، یکی از شرکتهای بیمه‌ای خصوصی ایران است که در این تحقیق، از داده‌های اولیهٔ این شرکت استفاده شده است. به منظور جمع‌آوری داده‌های مورد نیاز، با توجه به ادبیات موضوع، پایگاه‌داده‌های در دسترس و نظرات خبرگان سازمان، مناسب‌ترین داده‌ها و فیلدها برای مدل‌سازی متناسب با اهداف پژوهش، استخراج شده‌اند. داده‌های مورد نیاز، از سه پایگاه‌دادهٔ خسارت، بیمه‌گذار و مشتری، به دست آمده‌اند. بازهٔ زمانی استخراج داده از ۱۳۸۹/۰۱/۰۱ تا ۱۳۹٤/۰۱/۰۱ (۵ سال) بوده است. مجموعاً ۱۹۸۱٤٦ رکورد، مربوط به ۱۰۵٦٤٤ بیمه‌گذار، در اختیار محققان قرار گرفته است. در فاز پیش‌پردازش، ادغام مشتریان، حذف رکوردهای تکراری، و اجماع داده‌ها انجام شد. بر اساس مشاهدات، تعداد زیادی از سطرهای مجموعهٔ داده‌ها (۳٤۳۰۳ نفر)، مربوط به بیمه‌گذاران جدید هستند که در سال ۱۳۹۳ و برای اولین بار وارد شرکت شده‌اند. با توجه به اینکه این دست از مشتریان در ارزش‌گذاری نمی‌توانند به‌عنوان داده‌های مفید استفاده شوند و پیش‌بینی رویگردانی آنها نیز امکان‌پذیر نیست، از مجموعهٔ داده‌ها حذف شدند. پس از حذف مشتریان جدید و انجام فرایند پیش‌پردازش داده‌ها و حذف رکوردهای اشکال‌دار، داده‌های مربوط به ۷۱۳۳۷ نفر مشتری در مجموعهٔ داده‌ها باقی ماند. همچنین چند فیلد به دلیل بی‌تأثیربودن کنار گذاشته شدند.

به منظور برچسب‌گذاری بیمه‌گذاران، کلیهٔ بیمه‌گذارانی که در آخرین سال از مجموعهٔ داده‌ها بیمه‌نامه‌ای خریداری کرده‌اند و جزو مشتریان جدید نیستند، به‌عنوان مشتریان فعال در نظر گرفته شده‌اند؛ چراکه در سال آخر فعالیت داشته‌اند. برای تعریف مشتریان رویگردان، سابقهٔ خرید دوسال اخیر بررسی شده است. اگر مشتری در دو سال اخیر هیچ بیمه‌نامه‌ای خریداری نکرده است، به‌عنوان مشتری رویگردان برچسب‌گذاری شده است. اما اگر تنها سال آخر، خریدی نداشته است و سال قبل از آن خرید کرده است، فعال در نظر گرفته شده است. علت اینکه تنها به سابقهٔ خرید سال آخر توجه نشده است این است که در مجموعهٔ داده‌ها، موارد فراوانی وجود

دارند که مشتری یکسال هیچ خریدی نداشته است اما مجدد به شرکت بازگشته است. این دست از مشتریان به صورت موقت رویگردان شده‌اند. در شکل ۲ پنجرهٔ زمانی تعریف مشتریان رویگردان و فعال آورده شده است. همچنین متغیرهای جدیدی با توجه به اهداف تحقیق تعریف شده‌اند.



شکل ۲. پنجرهٔ زمانی تعریف مشتری جدید، فعال، و رویگردان.

جدول ۱. متغیرهای در نظرگرفته‌شده.

گروه	نام متغیر	توصیف متغیر
مشخصات بیمه‌گذار	State	منطقهٔ جغرافیایی بیمه‌گذار
	Gender	جنسیت بیمه‌گذار (۰ زن و ۱ مرد)
	Age	سن بیمه‌گذار
	Martial	وضعیت تأهل (مجرد و متأهل)
رفتار تراکنشی	G	گروه بیمه‌های خریداری‌شده
	L	تعداد روز بین اولین و آخرین خرید بیمه‌گذار (طول ارتباط)
	R	تعداد روز بین آخرین خرید و تاریخ ۱۳۹۳/۱۲/۲۹ (تازگی)
	F	تعداد کل خریدهای بیمه‌گذار در زمان حضور در شرکت
	M	متوسط پول پرداختی بیمه‌گذار (حق‌بیمه‌ها) در هر سال
	P	مقدار سودآوری بیمه‌گذار در زمان حضور در شرکت
	F/L	فرکانس‌نسبی‌خرید (تقسیم تعداد خرید بر تعداد سال حضور بیمه‌گذار)
	ClaimCount	تعداد دریافت خسارت توسط بیمه‌گذار
	ClaimAve	مقدار متوسط مبلغ خسارت دریافتی بیمه‌گذار در هر سال
	Variation	میانگین تفاضل مبلغ پرداختی در دو سال متوالی
	MIN_TIME	حداقل فاصلهٔ زمانی بین دو خرید متوالی
	MAX_TIME	حداکثر فاصلهٔ زمانی بین دو خرید متوالی
	AVG_TIME	میانگین فاصلهٔ زمانی بین دو خرید متوالی
برچسب	Label	برچسب بیمه‌گذار (فعال یا رویگردان)

پس از مشورت با خبرگان سازمان و در نظر گرفتن مجموعهٔ داده‌های موجود، در مجموع ۱۷ متغیر در دو دستهٔ کلی به منظور تعیین ارزش مشتری و تحلیل و پیش‌بینی رویگردانی در نظر گرفته شد، که در جدول ۱ آورده شده است.

٤. مدل‌سازی

٤-١. ارزش‌گذاری مشتریان

با توجه به اینکه هدف مطالعه، مدل‌سازی هدفمند پیش‌بینی است و مدل‌سازی هدفمند، بر مشتریان با ارزش تمرکز دارد تا منابع سازمان به طور مناسب اختصاص یابد، لذا مدلی جدید برای ارزش‌گذاری مشتریان پیشنهاد شده است.

مدل RFM[1]، از رایج‌ترین و پرکاربردترین مدلها در تعیین ارزش مشتری مبتنی بر رفتار وی است که تعداد دفعات، زمان و مقدار خرید مشتری را در نظر می‌گیرد (Winer, 2001). این تحلیل، برای هر مشتری امتیازی محاسبه می‌کند و مشتریان با امتیاز بالا را به‌عنوان مشتریان سودآور در نظر می‌گیرد. سه پارامتر مورد توجه در این تحلیل، تازگی، فرکانس خرید، و ارزش مالی هستند (Yeh et al., 2009).

مدل RFM توسط محققین مختلف، توسعه داده شده است. مدل پیشنهادی این پژوهش، ترکیبی از دو مدل LRFMP[2] (Chen et al., 2014)، که در آن متغیرهای L (طول مدت ارتباط مشتری با شرکت) و P (سودآوری مشتری) به RFM اضافه شده است، و مدل GRFM[3] که مشخصات گروه‌های کالایی که مشتری خریداری کرده است را به RFM اضافه می‌کند (Chang and Tsai, 2011)، است و مدلی جدید با نام LRFMPG[4] به منظور ارزش‌گذاری مشتریان ایجاد می‌کند که در آن ٦ متغیر در تعیین ارزش مشتریان عبارت‌اند از: طول مدت همکاری (L)، تأخر خرید (R)، فرکانس خرید (F)، ارزش مالی (M)، سودآوری مشتری (P) و گروه بیمه‌های خریداری‌شده

1. Recency, Frequency, Monetary
2. Length, Recency, Frequency, Monetary, Profit
3. Group, Recency,Frequency, Monetary
4. Length, Recency, Frequency,Monetary, Profit, Group

(G). سپس مقادیر LRFMPG محاسبه‌شده، به اعدادی در بازهٔ [۵ ,۱] به صورت استانداردشده نگاشت می‌شوند.

فرایند استانداردسازی با روش ارائه‌شدهٔ میگلوچ (۲۰۰۰)، انجام شده است؛ بدین صورت که در ابتدا کلیهٔ بیمه‌گذاران، به ترتیب صعودی در متغیر تأخر و به ترتیب نزولی ۵ در متغیر دیگر طبقه‌بندی شده‌اند. برای متغیرهای R، M، وP ابتدا مشتریان به ۵ پنجک برابر تقسیم شده‌اند. سپس بر اساس این پنجک‌ها، این متغیرها به اعداد در بازهٔ ۱ تا ۵، نگاشت شده‌اند. بدین صورت که مقدار ۱ نشان‌دهندهٔ ارزش کمتر و مقدار ۵ ارزش بالاتر را نشان می‌دهد. در مورد متغیر F، با توجه به اینکه در مجموعهٔ داده‌ها، تعداد بسیاری از افراد دارای F=1 هستند (بیش از ۲۰ درصد) نمی‌توان از روش فوق استفاده کرد. لذا در زمان نمره‌دهی به این متغیر، خریداران با یک خرید، امتیاز ۱ دریافت می‌کنند. پس از آن، میانگین فرکانسهای خرید باقی‌مانده برای مشتریان با مقدار F>1 محاسبه می‌شود. اگر مجموع تعداد خریدهای یک مشتری در کل کمتر از متوسط محاسبه‌شده باشد، نمرهٔ ۲ را دریافت خواهد کرد. این فرایند دو بار دیگر ادامه می‌یابد تا امتیازها برای کلیهٔ مشتریان محاسبه شود.

روش فوق در مورد متغیر L هم صادق است چراکه در مجموعهٔ داده‌ها تعداد زیادی از مشتریان دارای L=0 هستند؛ بدین معنی که این افراد تنها یک بار خرید داشته‌اند و لذا طول مدت همکاری آنان با شرکت صفر است.

به منظور تعیین وزن (اهمیت) این متغیرها، مطالعهٔ حاضر از فرایند تحلیل سلسله مراتبی(AHP[1]) بهره برده است. بدین صورت که از عوامل مؤثر استخراج‌شده در تعیین ارزش مشتریان، یک ماتریس نظرسنجی تهیه شد. این عوامل، تشکیل‌دهندهٔ سطرها و ستونهای این ماتریس‌اند. از ۹ کارشناس فروش شرکت بیمهٔ مطالعه‌شده در شعب مختلف، برای بررسی اهمیت نسبی متغیرهای مزبور درخواست شد تا با مقایسهٔ زوجی هر یک از

1. Analytic Hierarchy Process

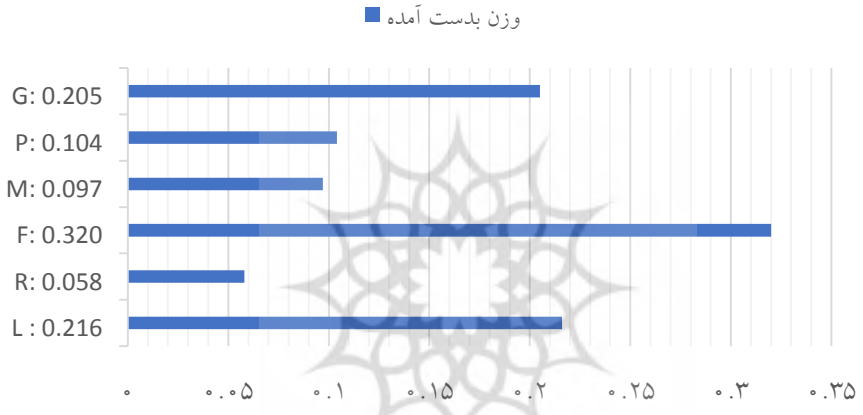
عوامل واقع بر هر سطر ماتریس با تک‌تک عاملهای واقع بر ستونهای ماتریس، اهمیت هر عامل سطری بر عاملهای ستونی را به صورت عددی نمایش دهند.

مثالی از ماتریس مقایسهٔ دوبه‌دوی فرایند تحلیل سلسله مراتبی، برای مدل LRFMPG که توسط یکی از خبرگان تکمیل شده است، در جدول ۲ آورده شده است.

جدول ۲. مثالی از ماتریس مقایسهٔ دوبه‌دو AHP- تکمیل‌شده توسط یکی از خبرگان.

ویژگی	مقایسه از نظر اهمیت									ویژگی
	۱:۹	۱:۷	۱:۵	۱:۳	۱:۱	۳:۱	۵:۱	۷:۱	۹:۱	
تأخر خرید						✓				طول مدت همکاری
فرکانس خرید			✓							
میانگین پول پرداختی					✓					
سودآوری مشتری						✓				
گروه بیمه‌های خریداری شده				✓						
فرکانس خرید				✓						**تاخر خرید**
میانگین پول پرداختی				✓						
سودآوری مشتری	✓									
گروه بیمه‌های خریداری‌شده						✓				
میانگین پول پرداختی							✓			**فرکانس خرید**
سودآوری مشتری					✓					
گروه بیمه‌های خریداری‌شده						✓				
سودآوری مشتری			✓							**میانگین پول پرداختی**
گروه بیمه‌های خریداری‌شده				✓						
گروه بیمه‌های خریداری‌شده								✓		**سودآوری مشتری**

پس از جمع‌آوری پرسشنامه‌ها، به وسیلهٔ ارزیابی AHP، وزن نسبی هر یک از متغیرهای LRFMPG در تعیین ارزش مشتریان (بر اساس اهمیت آنها) به دست آمد که در نمودار شکل ۳ آورده شده است. این نمودار نشان می‌دهد که فرکانس خرید مشتری، دارای بیشترین وزن و اهمیت است و بعد از آن طول مدت همکاری و تعداد گروه بیمه‌های خریداری شده بیشترین وزن را دارا هستند.



شکل ۳. وزن محاسبه‌شده هر یک از متغیرهای LRFMPG با استفاده از AHP.

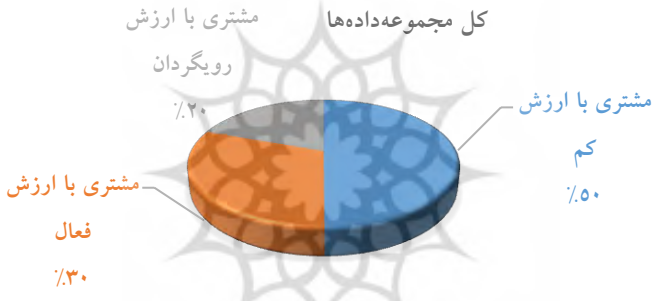
پس از تعیین ضریب تأثیر هر یک از متغیرها، ارزش و امتیاز مجموع هر یک از مشتریان تعیین می‌شود. فرض می‌کنیم که C، مجموعهٔ مشتریان شرکت بیمه بر طبق رابطهٔ $C = c_1 \cup c_2 \cup c_3 \cup \dots \cup c_n$ باشد، امتیازات استانداردشده LRFMPG برای مشتری C_i $(C_i \in C)$ عبارت‌اند از: $SL_i, SR_i, SF_i, SM_i, SP_i, SG_i$.

همچنین فرض می‌کنیم که امتیاز مجموع مشتری C_i با $ST(C_i)$ نشان داده می‌شود، این مقدار از رابطهٔ

$$ST(C_i) = (0.216 \times SL_i) + (0.058 \times SR_i) + (0.320 \times SF_i) + (0.097 \times SM_i) + (0.104 \times SP_i) + (0.205 \times SG_i)$$

به دست می‌آید، که مقادیر ضریب هر متغیر، وزن محاسبه‌شده با روش AHP است. بر این اساس، بالاترین امتیاز مجموع برای یک مشتری ۵ و کمترین امتیاز مجموع، مقدار

۱ است. پس از تعیین امتیاز، به منظور نشان‌دادن تفاوت بین مشتریان با ارزش و کم‌ارزش، تحلیل نمونهٔ تقسیم[1] انجام شده است، بدین صورت که مشتری با امتیاز بیشتر (یا کمتر) از متوسط جمعیت، به‌عنوان مشتری با ارزش (یا کم‌ارزش) برچسب‌گذاری شده است (Zhang et al., 2010). بر این اساس، تقریباً نیمی از مشتریان (۳۵۶۱۷ نفر) که امتیاز مجموع پایینی داشتند و کم‌ارزش بودند، از مجموعهٔ داده‌ها حذف شدند. اطلاعات کلی دربارهٔ دسته‌بندی مشتریان کم‌ارزش و با ارزش (رویگردان و فعال) در شکل ٤ آمده است.



شکل ٤. فراوانی مشتریان کم ارزش و با ارزش (رویگردان و فعال).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

٤-۲. مدل‌سازی پیش‌بینی و ارزیابی

در این بخش قصد داریم مدل‌سازی پیش‌بینی رویگردانی را انجام دهیم. گام اول، تقسیم داده به ۲ بخش آموزش و آزمون است (Luo et al. 2007). در این تحقیق، در ابتدا مشتریان رویگردان و فعال از یکدیگر جدا شدند، سپس به صورت تصادفی، ۷۰ درصد از هریک از این دو مجموعهٔ داده انتخاب شدند و به منظور ایجاد بخش آموزش با یکدیگر ترکیب شدند. سپس ۳۰ درصد باقی‌مانده از دو مجموعهٔ داده مشتریان رویگردان

1. Split Sample Analysis

و فعال، برای داده‌های آزمایشی مدل با یکدیگر ترکیب شدند (Li and Deng, 2012). همچنین به منظور دستیابی به نتایج دقیق‌تر، از ارزیابی متقابل ۱۰باره[1] استفاده شده است.

٤-۲-۱. مدل‌سازی و ارزیابی- شبکهٔ عصبی

در ابتدا، با استفاده از شبکهٔ عصبی و به ازای مقادیر مختلف پارامترهای آلفا (فاکتور تکانه[2]) و اِتا (نرخ یادگیری)، مدلهای مختلف ساخته و بر اساس صحت مدلهای ساخته‌شده، مقادیر بهینهٔ آلفا و اِتا ۰/۹ و ۰/۳ در نظر گرفته شد. همچنین چندین مدل مختلف با تعداد متفاوت لایهٔ پنهان (۱ تا ۳) و نرون (۲ تا ۵ نرون در هر لایه) ساخته شد.

معیارهای ارزیابی دقت، صحت، بازخوانی و معیار F1 برخی از مدلهای ساخته‌شده در جدول ۳ آورده شده است.

جدول ۳. ارزیابی مدلهای ساخته‌شده با شبکهٔ عصبی.

تکنیک	لایهٔ‌پنهان	دقت	بازخوانی	صحت	F_1
شبکهٔ عصبی MLP[3] (پس از انتشار)	۱ لایهٔ پنهان و ۲ نرون	٪۸۱/٦۰	٪۷٤/۸۱	٪۷٤/۳۵	٪۷۸/۰۸
	۱ لایهٔ پنهان و ۳ نرون	**٪۷۸/۲۱**	**٪۸۱/۲۸**	**٪۷٤/۷۷**	**٪۷۹/۷۱**
	۱ لایهٔ پنهان و ٤ نرون	٪۷۳/۵٦	٪۸۲/٤۹	٪۷۱/۲۵	٪۷۷/۷۹
	۲ لایهٔ پنهان و ٤ نرون	٪۷٦/۵۹	٪۸۳/۰۸	٪۷٤/۲۰	٪۷۹/۷۰
	۳ لایهٔ پنهان و ۹ نرون	٪۸٤/۳۵	٪۵۱/٦۲	٪٦٤/٦٦	٪٦٤/۰٤
شبکهٔ عصبی RBF[4]	---	٪۷٦/۹۵	٪۷٤/۰۵	٪۷۰/٦۵	٪۷۵/٤۷

همان‌طور که از جدول ۳ پیداست، شبکهٔ عصبی MLP با تعداد یک لایهٔ پنهان و سه نرون، بهترین عملکرد پیش‌بینی را داشته است.

1. 10 Fold Cross-Validation
2. Momentum Factor
3. Multilayer perceptron
4. Radial basis function

٤-۲-۲. مدل‌سازی و ارزیابی- سایر مدلها

در این تحقیق، علاوه بر شبکهٔ عصبی، با روشهای درخت تصمیم، ماشین بردار پشتیبان و رگرسیون لوژستیک نیز مدل‌سازی انجام شده است. همچنین به منظور افزایش دقت مدل‌سازیهای انجام‌شده، الگوریتمهای درخت تصمیم CHAID[1] و C5.0 به صورت تقویت‌شده[2] نیز استفاده شده‌اند. عملکرد مدلهای ساخته‌شده با معیارهای ارزیابی مختلف در جدول ٤ آورده شده است.

جدول ٤. ارزیابی مدلهای ساخته‌شده با درخت تصمیم، ماشین بردار پشتیبان و رگرسیون لوژستیک.

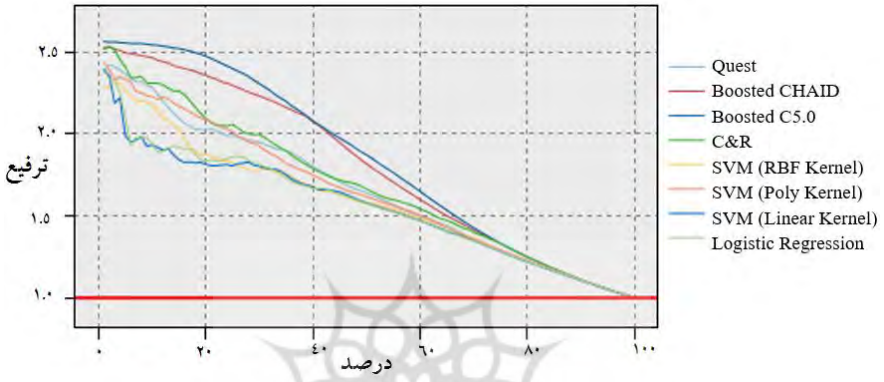
روش	الگوریتم	دقت	بازخوانی	صحت	F_1
درخت تصمیم	QUEST	٪۸۰/٤۳	٪۸٤/۳۳	٪۷۷/۹٤	٪۸۲/۳۳
	CHAID	٪۸۵/۱۷	٪۸۵/۲۱	٪۸۱/۹٤	٪۸۵/۱۸
	Boosted CHAID	٪۸۸/۳۸	٪۸۸/۵۹	٪۸۵/۹٤	٪۸۸/٤۸
	C5.0	٪۸٦/٤٤	٪۸٦/۸۳	٪۸۳/۵٦	٪۸٦/۵۳
	Boosted C5.0	**٪۸۷/۳۸**	٪۹۰/۱۷	**٪۸٦/۰٦**	**٪۸۸/۷۵**
	C&R	٪۸۰/۵۰	٪۸۲/٦۵	٪۷۷/۲۲	٪۸۱/۵٦
ماشین بردار پشتیبان	RBF Kernel	٪۷٦/۷۳	٪۸۲/۱۸	٪۷۳/۹٤	٪۷۹/٤۱
	Poly Kernel	٪۸۱/٦۹	٪۷۵/۸۳	٪۷٤/۹۰	٪۷۸/٦۵
	Linear Kernel	٪۷٦/۲۵	٪۸۳/۲۰	٪۷۳/۹۵	٪۷۹/۵۷
رگرسیون لوژستیک	---	٪۷۸/۲۰	٪۷۷/۱۳	٪۷۲/۹٤	٪۷۷/٦٦

٤-۲-۳. تحلیل نتایج و بحث

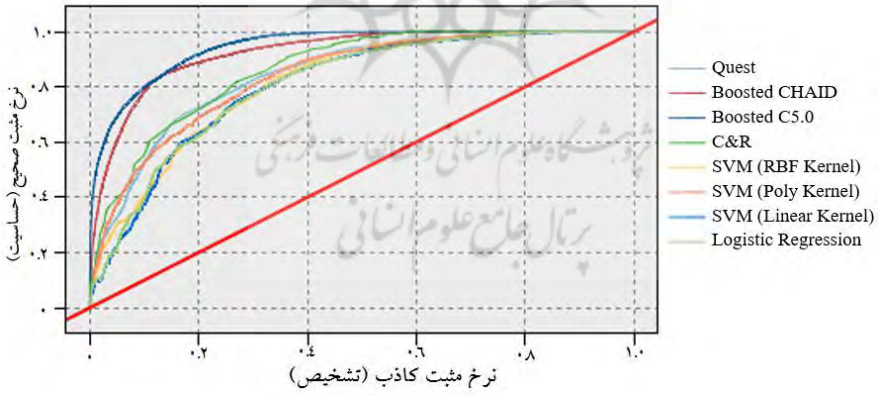
آنچه از جدولهای ۳ و ٤ برمی‌آید این است که درخت تصمیم C5.0 نسبت به سایر روشها عملکرد بهتری در پیش‌بینی رویگردانی داشته است و در کل الگوریتمهای درخت تصمیم، به طور قابل توجهی نسبت به سایر روشها برتری داشته‌اند.

1. Chi-square Automatic Interaction Detector
2. Boosted

در شکلهای ۵ و ٦، نمودارهای ترفیع[1] و مشخصهٔ عملکرد سیستم (ROC[2]) برای الگوریتمها آورده شده است.



شکل ۵. نمودار ترفیع.

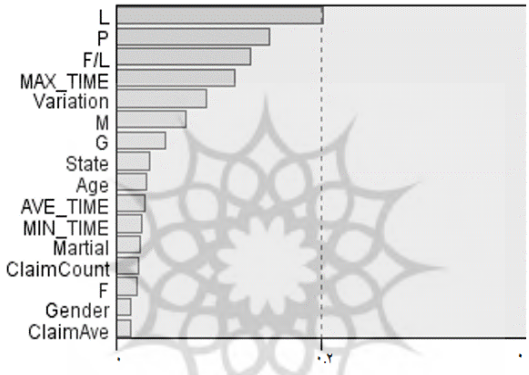


شکل ٦. نمودار ROC.

استنتاج درخت تصمیم‌گیری می‌تواند برای توصیف گروه مشتریان رویگردان استفاده شود و نتایج حاصله از آن را می‌توان در مدیریت ارتباط با مشتری استفاده

1. Lift
2. Receiver Operating Characteristic

کرد. همچنین از منظر اهمیت متغیرها بر فعال یا رویگردان‌بودن مشتریان، با درخت تصمیم C5.0 بررسی انجام شده است. بر این اساس، متغیر طول‌مدت ارتباط مشتری، تأثیرگذارترین متغیر است. بعد از آن، سودآوری مشتری، فرکانس خرید نسبی و حداکثر فاصلهٔ بین دو خرید متوالی، به‌ترتیب متغیرهای با اهمیت بوده‌اند. شکل ۷ ترتیب اهمیت متغیرها را در پیش‌بینی انجام‌شده با استفاده از C5.0 نشان می‌دهد.



شکل ۷. اهمیت متغیرها در پیش‌بینی با استفاده از درخت C5.0.

بر اساس قوانین استخراج‌شده و همچنین متغیرهای تأثیرگذار بر پیش‌بینی، آن دست از مشتریان که طول مدت ارتباط کمتری با شرکت دارند و به تعبیری به تازگی وارد شرکت شده‌اند، مشتریانی که سودآوری (تفاضل پرداختهای مالی و مقدار کل خسارت دریافتی) کمتر در طی حضور در شرکت دارند و مقدار خرید نسبی پایین‌تری دارند، احتمال بالاتری دارد که رویگردان شوند.

همچنین آن دست از مشتریان که حداکثر فاصلهٔ بین دو خرید متوالی برای آنها بیشتر است و مقدار رشد خریدشان (تفاضل خرید دو سال متوالی) رو به کاهش است، جزو مشتریان در خطر رویگردانی‌اند. متغیرهای فرکانس خرید و تعداد و مقدار دریافت خسارت، تأثیر اندکی بر رویگردانی داشتند. از سویی برخلاف آنچه تصور می‌شد، متغیرهای مربوط به مشخصات مشتری (پروفایل) تأثیر بسیار کمی بر

رویگردانی دارند و به جز منطقه جغرافیایی مشتری، سایر متغیرها جزو کم‌تأثیرترین متغیرها بوده‌اند.

۵. نتیجه‌گیری و پیشنهادها

مدل پیش‌بینی رویگردانی می‌تواند به‌عنوان ابزار هشداردهندهٔ اولیه برای کسب‌وکارها و استخراج‌کنندهٔ فاکتورهای حیاتی مرتبط با رویگردانی مشتری استفاده شود و امکان فراهم‌کردن دانش مفید افزوده برای پشتیبانی از تصمیم‌گیری را فراهم آورد. در این مطالعه، از یک‌سو به بررسی متغیرهای تأثیرگذار در رویگردانی مشتریان با ارزش در یک شرکت بیمه در ایران پرداخته شد و از سوی دیگر پیش‌بینی رویگردانی مشتریان با ارزش شرکت انجام شد. برای این منظور در ابتدا متغیرهای مهم و تأثیرگذار در ارزش‌گذاری مشتریان و همچنین پیش‌بینی رویگردانی، با بررسی ادبیات موضوع و نظرات خبرگان شناسایی شدند. سپس مشتریان ارزشمند، بر اساس نمرات ترکیبی که با استفاده از متغیرهای وزن‌دار LRFMPG محاسبه شد، مشخص گشتند. سپس مدلهای پیش‌بینی مختلف با استفاده از الگوریتم‌های متفاوت شبکه‌های عصبی مصنوعی (MLP و RBF) با چندین لایهٔ پنهان و نرون، درخت تصمیم (QUEST[1]، CHAID، C&R[2] و C5.0)، ماشین بردار پشتیبان (با هسته‌های مختلف) و رگرسیون لوژستیک ساخته شد و مقایسهٔ بین عملکرد پیش‌بینی مدلهای مختلف صورت گرفت.

برای ساخت مدل، تأثیر ۱۷ متغیر در دو دستهٔ کلی (مشخصات مشتری و رفتار تراکنشی) مورد بررسی قرار گرفت. بر این اساس مدل درخت تصمیم‌گیری با الگوریتم C5.0 به صورت تقویت‌شده نسبت به سایر مدلها در پیش‌بینی رویگردانی مشتری عملکرد بهتری داشت. ۵ متغیر مهم و تأثیرگذار بر پیش‌بینی به‌ترتیب عبارت‌اند از طول مدت ارتباط مشتری، سودآوری مشتری، فرکانس خرید نسبی، حداکثر فاصلهٔ بین دو خرید متوالی و میانگین تفاضل پرداخت دو سال متوالی.

1. Quick Unbiased Efficient Statistical Trees
2. Classification and Regression

نتایج حاکی از آن است که تمامی متغیرهای LRFMPG تأثیر بالایی بر رویگردانی ندارند و تأثیر هر یک متفاوت است. همچنین متغیرهای مربوط به مشخصات مشتری (پروفایل) در پیش‌بینی رویگردانی بسیار کم اهمیت هستند.

وجوه مختلفی می‌تواند در آینده مدنظر قرار گرفته شوند. در ابتدا اینکه مجموعهٔ دادهٔ استفاده‌شده، تنها توسط یک شرکت بیمه فراهم شده است. داده‌های شرکتهای بیمهٔ متعدد می‌تواند جمع‌آوری شوند و به منظور بالابردن تعمیم‌پذیری مدل LRFMPG، مقایسه شوند. از طرفی با توجه به اینکه تعدادی از متغیرهای در نظر گرفته شده در مدل ارائه‌شده، برای تحلیل ارزش مشتری و پیش‌بینی رویگردانی اهمیت کمی دارند، لذا مطالعات آینده می‌توانند مدلهای دیگر تحلیل ارزش مشتری را به کار گیرند. همچنین می‌توان از الگوریتمهای فرا ابتکاری به منظور پیش‌بینی رویگردانی استفاده کرد تا دقت مدل پیش‌بینی را بالا برد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

۱. توکلی، ا.، مرتضوی، س.، کاهانی، م. و حسینی، ز.، ۱۳۸۹. به کار گیری فرایند داده کاوی برای پیش‌بینی الگوهای رویگردانی مشتری در بیمه. *چشم‌انداز مدیریت بازرگانی*، دورهٔ چهارم، شمارهٔ ۳۷، ٤۱-۵۵.

۲. حسین‌خانی، ن، حسینی مطلق، م. و خاکزار بفروئی م.، ۱۳۹۳. شناسایی عوامل مؤثر بر رویگردانی مشتریان در صنعت بیمه. *هفتمین همایش بین‌المللی بیمه و توسعه*.

3. Buckinx, W. and Van den Poel, D., 2005. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), pp. 252-268.

4. Burez, J. and 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 Systems with Applications*, 32(2), pp. 277–288.

5. Chandar, M., Laha, A. and Krishna, P., 2006. Modeling churn behavior of bank customers using predictive data mining techniques. *National conference on soft computing techniques for engineering applications (SCT-2006)*, pp. 24-26.

6. Chang, H.C. and Tsai, H.P., 2011. Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38(12), pp. 14499-14513.

7. Chen, K., Hu, Y.H. and Hsieh, Y.C., 2014. Predicting customer churn from valuable B2B customers in the logistics industry: A case study. *Information Systems and e-Business Management*, pp. 1-20.

8. Coussement, K. and De Bock, K.W., 2013. Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, 66(9), pp. 1629–1636.

9. Glady, N., Baesens, B. and Croux, C., 2009. Modeling churn using customer lifetime value. *European Journal of Operational Research*, Vol. 197, pp. 402–411.

10. Günther, C.C., Tvete, I.F., Aas, K., Sandnes, G.I. and Borgan, Ø., 2014. Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, 1, pp. 587–598.

11. Hung, S.Y., Yen, D. and Wang, H.Y., 2006. Applying data mining to telecom churn management. *Expert System with Applications*, 31(3), pp. 515-524.

12. Kumar, R.K. and Srivastava, M., 2013. The antecedents of customer loyalty: An empirical investigation in life insurance context. *Journal of Competitiveness*, 5(2).

13. Li, G. and Deng, X., 2012 . Customer churn prediction of China telecom based on cluster analysis and decision tree algorithm. *Emerging Research in Artificial Intelligence and Computational Intelligence*, International Conference, AICI 2012, Springer Berlin Heidelberg, China, October 26-28, pp. 319-327.

14. Luo, B., Shao, P. and Liu, J., 2007 . Customer churn prediction based on the decision tree in personal handyphone system service. *International Conference on Service Systems and Service Management*, pp. 1–5.

15. Miglautsch, J.R., 2000. Thoughts on RFM scoring . *The Journal of Database Marketing*, 8(1), pp. 67-72.

16. Mohammadi, V.D., Albadvi, A. and Teymorpur, B., 2014. Predicting customer churn using CLV in insurance industry. *Shiraz Journal of System Managemen*, 2(1), Ser.5, pp. 39-49.

17. Morik, K. and Kopcke, H., 2004. Analysing customer churn in insurance data–a case study. *Knowledge Discovery in Databases: PKDD,* Springer, pp. 325-336.

18. Ngai, E.W.T., Xiu, L. and Chau, D.C.K., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), pp. 2592–2602.

19. Reinartz, W.J., 2000. On the profitability of long life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, Vol. 64, pp. 17-35.

20. Ryals, L., 2002. Are your customers worth more than money?. *Journal of Retailing and Consumer Services*, 9, pp. 241–251.

21. Sharma, A. and 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.

22. Teo, T.S.H., Devadoss, P. and Pan, S.L., 2006. Towards a holistic perspective of customer relationship management (CRM) implementation:

A case study of the Housing and Development Board, Singapore. *Decision Support Systems*, 42(3), pp. 1613–1627.

23. Tsai, C.F., and Lu, Y.H., 2009. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), pp. 12547–12553.

24. Wang, W.F., Chiang, D. A., Hsu, M. H., Lin, C. J. and Lin, I. L., 2009. A recommender system to avoid customer churn: A case study. *Expert Systems with Applications*, 36(4), pp. 8071–8075.

25. Wiersema, F., 2013. The B2B Agenda: The current state of B2B marketing and a look ahead . *Industrial Marketing Management*, 4(42), pp. 470-488.

26. Winer, R.S., 2001. A framework for customer relationship management. *California management review*, 43(4), pp. 89-105.

27. Yeh, I.C., Yang, K.J. and Ting, T.M., 2009. Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36(3), pp. 5866-5871.

28. Zhang J.Q., Dixit A.A. and Friedmann R.R., 2010. Customer loyalty and lifetime value: an empirical investigation of consumer packaged goods . *J Mark Theory Pract*, 18(2), pp. 127–140.

29. Zhang, X., Zhu, J., Xu, S. and Wan, Y., 2012. Predicting customer churn through interpersonal influence. *Knowledge-Based Systems*, Vol.28, pp. 97–104.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی