



## کاربرد تکنیک های داده کاوی در مدیریت ریزش مشتری

الهام جمالیان<sup>۱</sup>، رحیم فوکردی<sup>۲</sup> و جلال رضایی نور<sup>۳</sup>

<sup>۱</sup> دانشجوی مقطع کارشناسی ارشد، دانشکده فنی و مهندسی دانشگاه قم، قم، ایران

Elhamjamalian@yahoo.com

<sup>۲</sup> استادیار دانشکده مدیریت دانشگاه قم، قم، ایران

R.Foukerdi@gmail.com

<sup>۳</sup> استادیار دانشکده صنایع دانشگاه قم، قم، ایران

Rezaeenoor@gmail.com

### چکیده

طی سال های اخیر نحوه ارتباط سازمان با مشتریانش به عاملی کلیدی برای ایجاد مزیت رقابتی و مؤلفه های اساسی بقا در بازار تبدیل شده است. مفاهیمی چون جذب، حفظ، وفاداری و رضایت مشتری در بسیاری از سازمان ها نهادینه شده و درک و سنجش آن ها مترادف با موفقیت تعبیر شده است. طبق تحقیقات انجام شده، جذب مشتریان جدید ۵ تا ۶ برابر حفظ مشتریان موجود هزینه بر است [۱]. از این رو، حفظ مشتریان یکی از مهم ترین راهبردهای بازاریابی سازمان ها برای بقا به شمار می رود. از آن جا که نگه داری مشتریان نیز هزینه های خاص خود را دارد و نمی توان این هزینه را برای کلیه مشتریان اعمال کرد، سیستم های مدیریت ریزش مشتری به دنبال شناسایی و حفظ مشتریانی هستند که قصد ترک سازمان را دارند. اخیراً تکنیک ها و روش های داده کاوی متعددی برای پیش بینی ریزش به کار گرفته شده است که از آن جمله می توان به روش های رگرسیون لجستیک، شبکه های عصبی، الگوریتم ژنتیک، درخت تصمیم و ... اشاره کرد. این مقاله برخی تکنیک های داده کاوی مورد استفاده در پیش بینی ریزش مشتری را مرور نموده و سعی دارد با معرفی مهم ترین روش های به کار رفته، به شناخت زمینه های پژوهشی جدید در این حوزه کمک کند.

### کلمات کلیدی

حفظ مشتری، پیش بینی ریزش مشتری، تکنیک های داده کاوی

سازمان ارتقاء می یابد. در این میان، طی سال های اخیر، حفظ مشتری مورد توجه بسیار قرار گرفته است.

#### ۱- مقدمه

مدیریت ریزش مشتری به عنوان یکی از بارزترین راهبردهای حفظ مشتری از اهمیت زیادی برخوردار است. این جزء قلب مدیریت ارتباط با مشتری بوده و از موضوعات اصلی در این زمینه به شمار می رود. رضایت مشتری که به مقایسه توقعات مشتری و ادراک وی از رضایتش باز می گردد، شرط اساسی برای حفظ مشتریان است [۶]. این مرحله از مدیریت ارتباط با مشتری بسیار حیاتی است، زیرا ۱٪ افزایش در نگهداری مشتریان به طور متوسط موجب ۵٪ ارتقا در ارزش سازمان می شود [۷].

تحقیقات بسیاری مدیریت ریزش مشتریان را مورد بررسی قرار داده و از روش های مختلفی برای تعیین دلایل ریزش، پیش بینی ریزش و پیشگیری از آن استفاده نموده اند. در این میان، روش های داده کاوی به شکل گسترده ای مورد استفاده قرار گرفته اند. این مقاله ضمن مرور اجمالی این روش ها، سعی در شناسایی برخی جنبه های جدید از کاربرد داده کاوی در حوزه مدیریت ریزش مشتریان دارد.

مفهوم مدیریت ارتباط با مشتری از دهه ۹۰ میلادی رونق یافته است [۲]. این مفهوم راهبردی سازمانی است که با هدف بهبود در جذب، حفظ، وفاداری، و سودبخشی مشتری به فهم و اثرگذاری روی رفتار وی از طریق ارتباطات معنادار می پردازد [۳]. در تعریفی دیگر، مدیریت ارتباط با مشتری راهبردی جامع تعریف شده که در فرایند جذب، نگهداری و مشارکت مشتریان، به دنبال ایجاد ارزش افزوده برای سازمان و مشتری است. این راهبرد شامل یکپارچه سازی فعالیت های بازاریابی، فروش، ارائه خدمات مشتری و فعالیت های زنجیره تأمین برای افزایش کارایی و بهینگی سیستم است [۴].

مدیریت ارتباط با مشتری دارای چهار جزء اصلی شناسایی، جذب، حفظ، و توسعه مشتری است [۵]. این فرایند با شناسایی مشتری آغاز می شود. سپس شرکت به دنبال جذب مشتری بوده و پس از جذب، به ایجاد رابطه با وی می پردازد. در نهایت، با توسعه این رابطه سود



## ۲- مدیریت ریزش مشتری

رویگردانی یا ریزش مشتری که تحت عناوینی چون فرسایش، واگذاری

سازمان را ترک می کنند. این دسته معمولاً درصد کمی از مشتریان در حال ترک را تشکیل می دهند. دسته دوم گروهی از مشتریان هستند

که به صورت عمدی و با هدف یافتن محصولات و خدمات بهتر و ارزان تر سازمان را ترک کرده و به سوی رقیبان گرایش پیدا می کنند. از آن جاکه نگهداری مشتریان فعلی سازمان هزینه بر بوده و تمام مشتریان سازمان ارزش نگهداری ندارند، سیستم های مدیریت ریزش مشتری تمام مشتریان را در نظر نمی گیرند. در کل، مشتریان مورد نظر این سیستم ها مشتریانی هستند که طبق تقسیم بندی بالا عمداً قصد ترک سازمان و رفتن به سمت رقیبان سازمان را دارند.

## ۳- داده کاوی و جایگاه آن در مدیریت ریزش

### مشتری

فنون داده کاوی موجب انقلابی بزرگ در عینیت بخشیدن به مفاهیم مدیریت روابط مشتری در کسب و کارهای بزرگ شده اند. قابلیت داده کاوی در عملیاتی نمودن مفاهیم مدیریت ارتباط با مشتری این فناوری را در ردیف اجزای اصلی مدیریت روابط با مشتری و راهبردهای مطرح در آن، نظیر مدیریت ریزش مشتری قرار داده است [۵].

اصلی ترین مرحله در مدیریت ریزش مشتری ایجاد مدل پیش بینی است. مدیریت کارآمد ریزش مشتریان مستلزم ساخت مدل پیش بینی دقیق و کارآمد است. فنون داده کاوی در ارائه مدل های پیش بینی ریزش مشتری به شکل وسیله مورد استفاده قرار گرفته است. از این رو، بخش حاضر با تکیه بر روش ها و تکنیک های داده کاوی مورد استفاده، به طبقه بندی تحقیقات انجام شده می پردازد فهرستی از این روش ها را نشان می دهد. همچنین جدول ۲ روند استفاده از روش های داده کاوی را طی سال های ۲۰۰۳ تا ۲۰۱۳ ارائه می دهد. از دلایل محبوبیت مدل های صدر جدول می توان به سهولت استفاده، عملکرد مناسب، قابلیت استفاده برای انواع مختلف داده ها و مقاومت در برابر داده های مفقود و پرت اشاره کرد.

یا نقص مشتری نیز از آن یاد می شود، اصطلاحی تجاری است که برای از دست رفتن مشتری استفاده می شود. در واقع ریزش مشتری موقعیتی است که مشتریان سازمان تصمیم به ترک آن می گیرند. در تحقیقات گذشته، حرکت مشتری از یک سازمان به سازمان دیگر برای پیدا کردن محصولات و خدمات ارزان تر و بهتر نیز ریزش نامیده شده است. تحت این شرایط، مدیریت ریزش مشتری به معنی مجموعه اقداماتی است که به منظور پیش بینی و جلوگیری از این رفتار صورت می گیرد. در کل، مسأله پیش بینی مشتریان در حال رویگردانی یک مسأله رده بندی دوتایی است که خروجی آن احتمال ترک سازمان توسط مشتری مورد نظر را نشان می دهد [۸ و ۹]. در این مسأله معمولاً مشتری رویگردان را با برچسب یک و مشتری وفادار را با برچسب صفر نشان می دهند.

تعریف ریزش در حوزه های کاربردی مختلف تفاوت دارد [۱۰]. برخی محققین، ریزش را جابجایی مشتری از یک سرویس دهنده به سرویس دهنده دیگر تعریف می کنند [۱۱]. در تعریفی دیگر، ریزش به عنوان گرایش مشتری برای پایان معامله در بازه زمانی معین با شرکتی خاص تعریف شده است [۱۲]. اغلب تعاریف پیشین ریزش، رفتارهای مرتبط با محصول و آستانه تعیین شده با قوانین کسب و کار را مدنظر قرار داده و هنگامی که تراکنش های مشتری کمتر از حد آستانه باشد، وی را ریزشی تلقی می کنند [۱۳]. برخی دیگر مشتریانی که حساب های خود را می بندد ریزشی در نظر می گیرند [۱۴] و برخی مشتریانی که تکرار خریدهایشان کمتر از میانگین و نسبت انحراف استاندارد زمان بین خرید به زمان متوسط بین خرید بالاتر از میانگین باشد را به عنوان رویگردان می دانند [۱۵]. سرانجام [۱۳]، مشتریانی که کمتر از مقدار مشخصی دارایی (پس انداز، وثیقه و انواع دیگر دارایی ها) در بانک داشته باشند را ریزشی معرفی می کند.

به لحاظ فنی، طبق رابطه (۱) ریزش برابر است با نرخ خالص از دست دادن مشتری در مدتی معین [۷ و ۱۶]:

$$(1) \quad \text{ریزش ماهانه} = (C_0 + A_1 - C_1) / C_0$$

در این رابطه،  $C_0$  تعداد مشتریان در شروع دوره،  $C_1$  تعداد مشتریان در پایان دوره،  $A_1$  تعداد مشتریان جدید در دوره تعریف می شود.

در کل، مشتریان در حال ترک سازمان را به دو دسته اختیاری و غیر اختیاری تقسیم می کنند [۱۷]. شناسایی مشتریان اختیاری کار آسانی است زیرا رابطه این دسته با سازمان توسط خود سازمان قطع می شود و علت آن دلایلی چون عدم انجام تعهدات توسط مشتری و مسائلی از این دست است. مشتریانی که به ترک غیراختیاری سازمان می پردازند خود به دو دسته تقسیم می شوند. دسته نخست، آن هایی هستند که غیر عمدی و به واسطه تغییر محل سکونت یا کار،



جدول ۱. کاربرد روش های داده کاوی در ادبیات

تعداد	مراجع	روش داده کاوی
۱۹	[9], [10], [13], [15], [18], [21], [24], [26], [29], [32], [36], [37], [45], [46], [48], [49], [51], [53], [59]	رگرسیون لجستیکی
۱۵	[8], [9], [10], [13], [18], [20], [25], [27], [28], [30], [41], [44], [46], [48], [57]	درخت تصمیم
۱۴	[9], [13], [15], [19], [20], [27], [30], [33], [41], [46], [52], [54], [55], [58]	شبکه عصبی
۹	[15], [20], [21], [24], [26], [32], [35], [37], [46]	جنگل تصادفی
۸	[13], [20], [35], [37], [43], [46], [48], [60]	ماشین بردار پشتیبان <sup>۱</sup>
۶	[24], [31], [39], [40], [47], [52]	آنالیز بقا <sup>۲</sup>
۴	[8], [25], [54], [55]	نقشه خود سازمانده <sup>۳</sup>
۳	[1], [26], [42]	رگرسیون خطی
۲	[8], [27]	k-می‌انگین
۲	[32]	زنجیره مارکوف
۳	[30], [54], [56]	قوانین تلازمی
۱	[13]	آداکاست <sup>۴</sup>
۱	[21]	ماشین تقویت تدریجی <sup>۵</sup>
۱	[8]	k-نمایی‌نده <sup>۶</sup>
۱	[8]	c-می‌انگین فازی <sup>۷</sup>
۱	[19]	امتی‌از z
۱	[50]	تئوری مجموعه نادقیق <sup>۸</sup>
۱	[51]	تیلور-بوتینا <sup>۹</sup>
۱	[51]	سری های زمانی
۱	[9]	تحلیل تفکیک کننده <sup>۱۰</sup>
۱	[41]	شبکه بی‌زین
۱	[26]	جنگل های رگرسیون
۱	[34]	کشف توالی <sup>۱۱</sup>
۱	[38]	مدل جمعی تعمیم یافته <sup>۱۲</sup>

dataacademy.ir

<sup>1</sup> Support Vector Machine (SVM)

<sup>2</sup> Survival analysis

<sup>3</sup> Self-Organizing Map (SOM)

<sup>4</sup> AdaCost

<sup>5</sup> Gradient Boosting Machine

<sup>6</sup> K-Medoids

<sup>7</sup> Fuzzy C-Means

<sup>8</sup> Rough Set Theory

<sup>9</sup> Tailor-Butina

<sup>10</sup> Linear Discriminant Analysis

<sup>11</sup> Sequence Discovery

<sup>12</sup> Generalized additive model



جدول ۲. روند استفاده از روش های داده کاوی طی سالهای ۲۰۱۳-۲۰۰۳

تعداد	سال انتشار											روش داده کاوی
	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸	۲۰۰۹	۲۰۱۰	۲۰۱۱	۲۰۱۲	۲۰۱۳	
۱۹	-	۱	۳	۵	-	۳	۵	-	۱	-	-	رگرسیون لجستیکی
۱۵	۱	۱	۱	۳	۱	۱	۵	-	۱	-	-	درخت تصمیم
۱۴	۱	۲	۲	۲	-	۱	۶	-	-	-	-	شبکه عصبی
۹	-	-	۲	۱	-	۳	۳	-	-	-	-	جنگل تصادفی
۸	-	-	-	-	-	۲	۳	-	۱	۲	-	ماشین بردار پشتیبان
۶	-	۱	۱	۱	-	۲	۱	-	-	-	-	آنالیز بقا
۴	-	۱	-	-	۱	-	۲	-	-	-	-	نقشه خود سازمانده
۳	-	۱	۱	-	۱	-	-	-	-	-	-	رگرسیون خطی
۲	-	-	-	۱	-	-	۱	-	-	-	-	k-می انگی
۲	-	-	-	۱	-	-	-	-	-	-	-	زنجیره مارکوف
۳	۱	۱	-	-	-	-	۱	-	-	-	-	قوانین تلازمی
۱	-	-	-	-	-	-	۱	-	-	-	-	آداکاست
۱	-	-	-	-	-	-	۱	-	-	-	-	ماشین تقویت تدریجی
۱	-	-	-	-	-	-	۱	-	-	-	-	k-نمایی نده
۱	-	-	-	-	-	-	۱	-	-	-	-	c-می انگی فازی
۱	-	-	-	-	-	-	۱	-	-	-	-	امتیاز Z
۱	-	-	-	-	-	-	۱	-	-	-	-	تئوری مجموعه نادقیق
۱	-	-	-	۱	-	-	-	-	-	-	-	تی-لور- بوتی نا
۱	-	-	-	۱	-	-	-	-	-	-	-	سری های زمانی
۱	-	-	-	۱	-	-	-	-	-	-	-	تحلیل تفکیک کننده
۱	-	-	۱	-	-	-	-	-	-	-	-	شبکه بی-زی
۱	-	-	۱	-	-	-	-	-	-	-	-	جنگل های رگرسیون
۱	۱	-	-	-	-	-	-	-	-	-	-	کشف توالی
۱	-	-	-	-	-	-	-	-	-	-	۱	مدل جمعی تعمی م یافته

حسابها - که برای ارزیابی مدل پیش بینی ریزش کارت های اعتباری مناسب تر است- را در بر می گیرد، مورد ارزیابی قرار گرفته اند. نتایج حاکی از آن است که رگرسیون لجستیک کمی بهتر از درخت تصمیم عمل می کند. در واقع، هدف این تحقیق ارائه روشی جدید نیست، بلکه با اجرا و درک مدل سعی در بررسی دلایل ریزش برای انتخاب متغیرهای مناسب مورد بررسی قرار دهد. طبق این مطالعه اطلاعات دموگرافیک در پیش بینی نقش کمی داشته و اطلاعات کارت و تراکنش برای ایجاد مدل مناسب بوده است.

روش دیگری که در پیشینه به میزان قابل توجهی بکار رفته است، شبکه عصبی است. به عنوان مثال، در تحقیق [۱۹] دو مدل شبکه عصبی مبتنی بر الگوریتم ژنتیک برای پیش بینی ریزش مشتریان در صنعت خدمات مخابرات سیار ارائه شده است. اولین مدل از معیار کراس آنترویی (حداکثر شباهت لگاریتمی) برای پیش بینی ریزش استفاده کرده و مدل دوم سعی در بهینه سازی مستقیم صحت پیش بینی دارد. این تحقیق روش شبکه های عصبی را با استفاده از الگوریتم ژنتیک به منظور انتخاب اوزان بهینه شبکه بهبود داده است.

تابع هدف در مدل اول مبتنی بر کراس آنترویی و در دیگری دقت مدل است. برای بهینه سازی این مدل ها از اعتبارسنجی کراس بهره

چنان که مشاهده می شود، رگرسیون لجستیک بیشترین کاربرد را در پیشینه داشته است. به عنوان مثال، [۱۸] برای پیش بینی ریزش مشتریان بانک، ابتدا متغیرهای مستقل را به گونه ای که متغیرهای هر دسته با هم مرتبط باشند، به دسته های مجزا تقسیم کرده است. سپس هر گروه به عنوان ورودی روش ADTrees، که بهبود یافته روش درخت تصمیم است، استفاده شده و خروجی تمام مدل های ADTrees به عنوان ورودی Logit انتخاب شده اند. همچنین از روشی مشابه RFE برای انتخاب زیر مجموعه بهینه ویژگی ها استفاده شده است. این مدل با TreeNet و بر اساس ROC مقایسه و عملکرد مشابهی را نشان داده است.

مدل پرکاربرد دیگر درخت تصمیم است. در تحقیق [۱۰]، از درخت تصمیم و رگرسیون لجستیک برای ایجاد یک مدل پیش بینی ریزش با استفاده از داده های واقعی کارت های اعتباری یک بانک چینی استفاده شده است. همچنین ترکیب چهار دسته متغیر: اطلاعات مشتری، اطلاعات کارت، اطلاعات ریسک و اطلاعات تراکنش ها مورد بررسی قرار گرفته و به جای پرسش نامه از داده های پایگاه داده استفاده شده است. در نهایت، این الگوریتم ها علاوه بر دقت، با شاخص عدم رده بندی درست که دو نوع خطا و حساسیت اقتصادی نسبت به



## دومین کنفرانس ملی کامپیوتر

معاونت آموزشی، پژوهشی و فرهنگی سازمان سما  
دانشکده فنی و حرفه ای سما سنندج، ۶ و ۷ آذر ۱۳۹۲

می گیرند، سپس نتایج این رده بندها با هم ترکیب شده و به عنوان نتیجه به خروجی داده می شوند. در این تحقیق نمونه گیری با استفاده از نمونه گیری ساده انجام شده و نتایج حاصل از بکارگیری مدل روی سه پایگاه داده با روش های مختلف نظیر حداکثر سود، شاخص اچ، AUC، لیفت دهک اول، حساسیت و ... مورد ارزیابی قرار گرفته اند و همچنین از روش اعتبارسنجی یکی در میان برای اعتبارسنجی مدل استفاده شده است.

آنالیز بقا نیز در تعداد قابل توجهی از مقالات به کار رفته است. به عنوان مثال، [۲۴] به دنبال تمیز دو نوع ریزش با نام های تجاری و مالی در خدمات قراردادی است. ریزش تجاری به این معنی است که مشتریان با یک تصمیم گیری برنامه ریزی شده قرارداد خود را تمدید نمی کنند و ریزش مالی به ریزشی اشاره دارد که به خاطر عدم توانایی مشتری در تقبل هزینه های مالی رخ می دهد. نوع ریزش هر مشتری را می توان از پایگاه داده ی شرکت استخراج کرد. در این تحقیق برای پیش بینی ریزش از دو مدل تحلیل بقا و جنگل های تصادفی به عنوان مدل های پویا و ایستا بر روی دو گروه ریزش کننده و نیز کل مشتریان بهره گرفته شده است. نتایج به دست آمده بیانگر سادگی پیش بینی ریزش مالی نسبت به ریزش تجاری، تأثیر چشمگیرتر متغیرهای رفتاری بر ریزش مالی و تأثیر بسیار بیشتر اقدامات پیش گیرانه در مورد گروه ریزش کنندگان تجاری بر نرخ حفظ شرکت است.

پرکاربردترین روش توصیفی مورد استفاده، نقشه خود سازمانده است. به عنوان مثال در تحقیق [۲۵]، که در شرکتی مخابراتی صورت گرفته است، سعی در ارائه مدلی برای پیش بینی احتمال ریزش مشتریان و پیشنهاد راهکاری برای جلوگیری از ریزش با توجه به خوشه مشتری و احتمال ریزش پیش بینی شده بوده است. برای این منظور از روش C5.0 و GHSOM استفاده شده است. همچنین اصلاحاتی در روش دوم صورت گرفته و با بهره گیری از اعتبارسنجی ده بخشی کراس و نرخ بهره، C5.0 بهبود داده شده است.

روش رگرسیون خطی نیز در برخی موارد مورد استفاده قرار گرفته است. به عنوان مثال، [۲۶] برای پیش بینی ریزش، خرید بعدی و تغییرات سود مشتریان بانک، از مدل های جنگل های تصادفی، جنگل های رگرسیون، رگرسیون خطی و رگرسیون لجستیک استفاده کرده و مدل های دودویی را با شاخص AUC و مدل های پیوسته را با میانگین انحراف مطلق ارزیابی کرده است. مدل های جنگل های تصادفی و جنگل های رگرسیون برتر بوده اند.

دومین روش توصیفی مورد استفاده در تحقیقات پیشین، k میانگین بوده است. به عنوان مثال، [۲۷] بر اساس ویژگی های دموگرافیک، اطلاعات صورت حساب، وضعیت قرارداد، جزئیات تماس ها و سابقه تغییر نوع خدمات به ارائه دو راهکار متفاوت برای تعیین احتمال ریزش مشترکین یک شرکت مخابراتی پرداخته است. در راهکار اول، مشتریان بر اساس ارزش، وفاداری و اعتبار و با استفاده از k میانگین به ۵ خوشه تقسیم شده و سپس در هر خوشه یک درخت

برده شده است. برای اثبات برتری مدل های پیشنهادی بر روش های آماری، بر اساس شاخص های دقت، AUC و لیفت دهک اول با روش امتیاز z مقایسه شده اند و در نهایت مدل مبتنی بر دقت برترین مدل شناخته شده است.

جنگل های تصادفی روش دیگری است که در این زمینه مورد توجه محققان بوده است. برای مثال، [۲۰] از روش یادگیری جدیدی به نام جنگل های تصادفی متوازن بهبود یافته (IBRF) برای پیش بینی ریزش مشتری در محیط بانکی استفاده کرده است. این روش به گونه ای است که بهترین ویژگی ها به صورت تکرار شونده با تغییر توزیع کلاس ها و با قرار دادن جریمه سنگین تر برای دسته بندی غلط اعضای کلاس اقلیت تعیین می شوند. در واقع با ترکیب روش یادگیری حساس به هزینه جنگل های تصادفی موزون و روش نمونه گیری افزایشی جنگل های تصادفی متوازن مدل جنگل های تصادفی متوازن بهبود یافته ایجاد شده است. نتایج حاکی از آن است که این روش در مقایسه با روش هایی نظیر شبکه های عصبی مصنوعی، درختان تصمیم، ماشین بردار پشتیبان و سایر جنگل های تصادفی مانند جنگل های تصادفی متوازن و موزون به طور چشمگیری دقت را افزایش داده است. برای ارزیابی این روش از نرخ صحت و نمودار شاخص لیفت استفاده شده است، که IBRF در مقایسه با شبکه عصبی مصنوعی، درخت تصمیم و ماشین بردار پشتیبان عملکرد بهتری داشته است. همچنین نتایج پیش بینی این روش بهتر از سایر جنگل های تصادفی مانند جنگل های تصادفی متوازن و موزون بوده است. در مقاله ای دیگر، [۲۱]، که به دنبال راه حلی برای مشکل عدم توازن کلاس ها در پیش بینی ریزش مشتریان است، از دو روش نمونه گیری ساده و پیشرفته و آموزش حساس به هزینه به همراه روش تقویت تدریجی پیشنهاد و آزمایش شده است. برای آزمون پیشنهادات از مدل جنگل های تصادفی، جنگل های تصادفی موزون، رگرسیون لجستیک و ماشین تقویت تدریجی به همراه تکنیک ارزیابی عرضی ده بخشی استفاده شده است. نتایج بیانگر اثر مثبت نمونه گیری و روش های تقویتی بر شاخص AUC مدل ها است و مدل جنگل های تصادفی موزون به عنوان مدل برتر شناخته شده است [۲۲].

پس از آنالیز جنگل های تصادفی، ماشین بردار پشتیبان پرکاربردترین روش مورد استفاده در مدیریت ریزش است. به عنوان مثال، در مقاله [۲۳]، روش جدیدی برای پیش بینی مستقیم از داده های رفتاری ارائه شده که در آن نیازی به تبدیل و ایستا نمودن داده ها نیست. در این روش که ماشین بردار پشتیبان چند هسته ای سلسله مراتبی نام دارد، یک الگوریتم آموزش سه مرحله ای پیاده سازی و آزمون می شود. به این ترتیب، با تخمین ضرایب متغیرهای ایستا و رفتاری طی فرآیند آموزش بدون تبدیل داده های رفتاری یک رده بند ارائه می گردد. بدین معنی که به جای ایستا نمودن داده های رفتاری از طریق تجمیع یا سه تایی سازی، مستقیماً به یک رده بند سری زمانی وارد می شوند و داده های ایستا نیز ورودی یک رده بند استاندارد قرار





سازمان دارد. الگوریتم‌های مختلفی برای پیش‌بینی ریزش مشتری مورد استفاده قرار گرفته‌اند. از معروف‌ترین آنها شبکه‌های عصبی، درخت‌های تصمیم، ماشین بردار پشتیبان، رگرسیون لجستیک و الگوریتم ژنتیک است. با این وجود، پیش‌بینی ریزش مشتریان هنوز هم به عنوان یکی از مباحث داغ در تحقیقات بازار به حساب می‌آید.

در این مقاله تعدادی از مقالات منتشر شده در حوزه مدیریت ریزش مشتری از حیث مدل‌های داده‌کاوی استفاده شده در آن‌ها مورد بررسی قرار گرفت. در تحقیقات آینده می‌توان این بررسی را از جهات دیگری مانند روش‌های به کار رفته در سایر فازهای کشف دانش مانند نمونه‌گیری، انتخاب یا کاهش ویژگی و... انجام داد.

## مراجع

- [1] A. Eshghi, D. Houghton, H. Topi; "Determinants of customer loyalty in the wireless telecommunications industry", Telecommunications Policy, vol. 31, pp. 93-106, 2007.
- [2] E. Ngai; "Customer relationship management research(1992-2002)", Marketing Intelligence and planning, pp. 582-605, 2005.
- [3] R. Swift; Accelerating customer relationships: Using CRM and relationship technologie, Prentice Hall Professional, 2001.
- [4] Arvatiyar and J. Sheth; " CRM: emerging practice, process, and discipline", Journal of Economic and social Research, pp. 1-34, 2001.
- [6] A. H. Kracklauer, D. Mills, D. Seifert; Collaborative Customer Relationship Management; Taking CRM to the Next Level, Springer, 2004.
- [5] E. Ngai, L. Xiu, and D. Chau; "Application of data mining techniques in CRM: A literature review and classification", Expert system and applications, pp. 2592-2602, 2009.
- [7] J. Basiri, F. Taghiyareh, B. Moshiri; "A Hybrid Approach to Predict Churn", In Proceedings of APSCC'2010, Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific , pp. 485 - 491, 2010.
- [8] I. Bose, X. Chen; "Hybrid Models Using Unsupervised Clustering for Prediction of Customer", Journal of Organizational Computing and Electronic Commerce, vol.19, pp. 133-151, 2009.
- [9] A. Lemmens, and C. Croux; "Bagging and boosting classification trees to predict churn", Journal of Marketing Research, pp. 276-286, 2006.
- [10] G. Nie, W. Rowe, L. Zhang, Y. Tian, Y. Shi; "Credit card churn forecasting by logistic regression and decision tree", Expert Systems with Applications, pp. 15273-15285, 2011.
- [11] J. Lu; "Predicting customer churn in the telecommunications industry - An application of survival analysis modeling using SA", Sprint Communications Company, 2002.
- [12] S. Neslin, S. Gupta, W. Kamakura, J. Lu and C. Mason; "Defection detection: Improving predictive accuracy of customer churn models", Journal of Marketing Research, pp. 204-211, 2006.
- [13] N. Glady, B. Baesens, C. Croux; "Modeling churn using customer lifetime value", European Journal of Operational Research, vol. 197, pp. 402-411, 2009.
- [14] D. Van den Poel, B. Larivière; "Customer attrition analysis for financial services using proportional hazard models", European Journal of Operational Research, pp. 196-217, 2004.
- [15] W. Buckinx, D. Van den Poel; "Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual fmcc retail setting. European Journal of Operational Research, pp. 252-268, 2005.
- [16] C. Geppert; "Customer Churn Management: Retaining High-Margin Customers with Customer Relationship Management Techniques", USA: Eine Studie der KPMG LLP, 2002.

تصمیم برای پیش‌بینی ریزش ایجاد شده است. در راهکار دوم روش‌های C5.0 و شبکه عصبی مصنوعی جداگانه و مستقیماً برای پیش‌بینی احتمال ریزش مشتریان به کار رفته‌اند.

زنجیره مارکوف نیز در برخی مقالات برای پیش‌بینی ریزش مورد استفاده قرار گرفته است. به عنوان مثال، [۲۸] از سه روش رگرسیون لجستیک ترکیب شده با زنجیره مارکوف، جنگل‌های تصادفی و رگرسیون لجستیک ساده برای پیش‌بینی ریزش استفاده کرده و در نهایت عملکرد آن‌ها را با استفاده از شاخص دقت، لیفت و AUC سنجیده است. در نهایت روش ترکیبی رگرسیون لجستیک با زنجیره مارکوف دقت بالایی داشته و روش جنگل‌های تصادفی دارای شاخص‌های لیفت و AUC بهتری بوده است.

قوانین تلازمی، آداکست، ماشین تقویت تدریجی، K نماینده، بیرج، میانگین فازی، امتیاز z، تئوری مجموعه‌ی تقریبی، تیلور-بوتینا، سری‌های زمانی، تحلیل ممیز خطی، شبکه بیزین، جنگل‌های رگرسیون، توزیع‌های آماری، کنترل فرآیند آماری و کشف توالی برخی دیگر از تکنیک‌های مورد استفاده برای پیش‌بینی به شمار می‌روند.

همان‌طور که گفته شد، رویکردهای مختلفی این مشکلات را مد نظر قرار داده‌اند که خود با محدودیت‌هایی مواجه هستند. به عنوان مثال، هر چند از الگوریتم‌های مبتنی بر درخت تصمیم برای رتبه‌بندی استفاده می‌شود، ممکن است برخی برگ‌های درخت احتمال کلاس مشابهی داشته باشند. به علاوه، این روش نسبت به نویز آسیب‌پذیر است. شبکه‌های عصبی نیز به دنبال یافتن راه‌حل‌های نزدیک به بهینه هستند و هنگامی که تعداد پارامترهای مدل افزایش یابند، دچار تطبیق بیش از حد می‌شوند. اگر چه الگوریتم‌های ژنتیک مدل‌های پیش‌بینی دقیقی را ارائه می‌دهند، قادر به شفاف‌سازی احتمال وقوع این پیش‌بینی‌ها نیستند. سرانجام، روش‌هایی نظیر ماشین بردار پشتیبان و الگوهای ترتیبی نیز معمولاً به بهترین نتایج منتهی نمی‌شوند.

تحت این شرایط، به الگوریتمی کاربردی با قابلیت اجرایی بالا برای پیش‌بینی ریزش مشتریان نیاز است، از این رو، تلاش برای افزایش دقت مدل‌های پیش‌بینی ریزش ادامه دارد.

## ۴- نتیجه‌گیری

طی سال‌های اخیر حفظ مشتریان تبدیل به موضوعی مهم برای کسب و کارهای مختلف شده است. چرا که جذب مشتریان جدید بسیار پرهزینه‌تر از حفظ مشتریان کنونی است. اهمیت مدیریت ریزش مشتری در بسیاری از مطالعات اخیر مورد تأکید قرار گرفته و از آن به عنوان بخشی مهم از سیستم مدیریت ارتباط با مشتری یاد شده است. این مسئله حکایت از نقش مدیریت ریزش مشتری در تضمین بقای



- Expert systems with applications, vol. 36, pp. 6127–6134, 2009.
- [38] K. Coussement, K. W.; “Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning”, *Journal of Business Research*, 2013.
- [39] P. S. Fader, B. G. S. Hardie, K. L. Lee; “ ‘Counting Your Customers’ the Easy Way: An Alternative to the Pareto/NBD Model”, *Journal of Marketing Science*, vol. 24, No. 2, pp. 275–284, 2005.
- [40] F. F. Gönül, F. T. Hofstede; “ How to Compute Optimal Catalog Mailing Decisions”, *Journal of Marketing Science*, vol. 25, No. 1, pp. 65–74, 2006.
- [41] X. Hu; “A Data Mining Approach for Retailing Bank Customer Attrition Analysis”, *Applied Intelligence*, vol.25, pp. 47–60, Springer, 2005.
- [42] M. Kim, M. Park and D. Jeong ; “The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services”, *Telecommunications Policy*, vol.28, pp. 145–159, 2004.
- [43] N. Kim, K-H. Jung, Y-S. Kim, J. Lee; “Uniformly subsampled ensemble (USE) for churn management: Theory and implementation”, *Expert Systems with Applications*, pp. 11839–11845, 2012.
- [44] S. Y. Kim, T. S. Jung, E. H. Suh, H. S. Hwang ; “Customer segmentation and strategy development based on customer lifetime value: A case study”. *Expert Systems with Applications*, vol. 31, pp. 101–107, 2006.
- [45] R. Kivetz, O. Urminsky, Y. Zheng; “The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention”, *Journal of Marketing Research*, XLIII, pp. 39–58, 2006.
- [46] D. A. Kumar, V. Ravi; “Predicting credit card customer churn in banks using data mining”, *Int. J.Data Analysis Techniques and Strategies*, vol. 1, pp. 4-28, 2008.
- [47] B. Larivière, D. Van den Poel; “Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services”, *Expert Systems with Applications*, vol. 27, pp. 277–285, 2004.
- [48] S. Lessmann, S. Voß; “A reference model for customer-centric data mining with support vector machines” *European Journal of Operational Research*, vol. 199(2), pp. 526-530, 2009.
- [49] M. Lewis; “ Research Note: A Dynamic Programming Approach to Customer Relationship Pricing”, *Management Science*, vol.51, pp.986–994, 2005.
- [50] Liou, J.H James; “A novel decision rules approach for customer relationship management of the airline market”, *Expert Systems with Applications*, vol.36 (3), pp. 4374–4381, 2009.
- [51] A. Prinzie T and D. Van den Poel; “Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM”, *Decision Support Systems*, vol.42, pp.508–526, 2006.
- [52] N. Sarlija, M. Bencic, M. Z. Susac ; “Comparison procedure of predicting the time to default in behavioural scoring”, *Expert Systems with Applications*, vol. 36, pp. 8778–8788, 2009.
- [53] D. Seo, C. Ranganathan, Y. Babad; “Two-level model of customer retention in the US mobile telecommunications service market”, *Telecommunications Policy*, vol. 32, pp.182–196, 2008.
- [54] H. S. Song, J. K. Kim, Y. B. Cho and S. H. Kim; “A Personalized Defection Detection and Prevention Procedure based on the Self-Organizing Map and Association Rule Mining: Applied to Online Game Site”, *Artificial Intelligence Review*, vol. 21, pp. 161–184, 2004.
- [55] C. F. Tsai, Y. H. Lu ; “Customer churn prediction by hybrid neural networks”, *Expert Systems with Applications*, vol.36, pp. 12547–12553, 2009.
- [56] C-F. Tsai, M-Y. Chen; “Variable selection by association rules for customer churn prediction of multimedia on demand”, *Expert Systems with Applications*, 2009.
- [57] Y.Wang, D-An. Chiang, M-H. Hsu, C-J. Lin, I-L. Lin; “A recommender system to avoid customer churn: A case study”, *Expert Systems with Applications*, pp. 8071-8075, 2009.
- [58] L. Yan, R. H. Wolniewicz, R. Dodier; “Predicting Customer Behavior in Telecommunications”, *IEEE Intelligent Systems*, IEEE Computer Society, 2004.
- [17] Bucklin, Z. Jamal and RE; “Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach”, *Journal of Interactive Marketing, Special Issue: CRM Meets Data Mining*, vol. 20(3-4), pp. 16–29, 2006.
- [18] J. Qi, L. Zhang, Y. Liu, L. Li, Y. Shen, L. Liang and H. Li; “ADTreesLogit model for customer churn prediction”, *Annals of Operations Research*, springer, vol. 168, pp. 247–265, 2009, Springer.
- [19] P.C. Pendharkar; “ Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services”, *Expert Systems with Applications*, vol. 36, pp. 6714-6720, 2009.
- [20] Y. Xie, X. Li, E.W.T. Ngai and W. Ying; “Customer churn prediction using improved balanced random forests”, *Expert Systems with Applications*, vol. 36, pp. 5445–5449, 2009.
- [21] J. Burez and D. Van den Poel; “Handling class imbalance in customer churn prediction” *Expert systems with applications*, vol.36, pp. 4626–4636, 2009.
- [22] S. KhakAbi, M.R. Gholamian, M. Namvar “ Data Mining Applications in Customer Churn Management”, In *Proceedings of 2010 International Conference on Intelligent Systems*, pp. 220-225, 2010.
- [23] Z-Y. Chen, Z-P. Fan, M. Sun; “A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data”, *European Journal of Operational Research*, pp.461–472, 2012.
- [24] J. Burez and D. Van den Poel ; “Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department”, *Expert Systems with Applications*, vol. 35, pp. 497–514, 2008.
- [25] B-H Chu, M-Sh. Tsai and Ch-S. Ho; “Toward a hybrid data mining model for customer retention”, *Knowledge-Based Systems*, vol. 20, pp. 703–718, 2007.
- [26] B. Larivière and D. Van den Poel ; “Predicting customer retention and profitability by using random forests and regression forests techniques”. *Expert Systems with Applications*, vol.29, pp. 472–484, 2005.
- [27] S-Y. Hung, David C. Y. and H-Y. Wang; “Applying data mining to telecom churn management”, *Expert Systems with Applications*, vol. 31, pp. 515–524, 2006.
- [28] M. Pettersson; “SPC with Applications to Churn Management”, *International Journal of quality and reliability management*, vol. 20, pp.397–406, 2004.
- [29] J-H. Ahna, S-P. Hana and Y-S. Lee; “ Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry”, *Telecommunications Policy*, vol. 30, pp.552–568, 2006.
- [30] W. H. Au, K. C. C. Chan, X. Yao; “A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction”, *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 7, pp. 532-545, 2003.
- [31] S. Borle, S. S. Singh, D. C. Jain; “Customer Lifetime Value Measurement”, *Journal of Management Science*, vol. 54, pp. 100–112, 2008.
- [32] J. Burez, D. Van den Poel; “CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services”, *Expert Systems with Applications*, vol.31, pp. 101–107, 2006.
- [33] S. E. Chang, Y. T. Jang ; “Assessing Customer Satisfaction in a V-commerce Environment”, *Journal of Organizational Computing and Electronic Commerce*, vol. 19, pp. 30-49, 2009.
- [34] Ding-An Chiang, Yi-Fan Wang, Shao-Lun Lee and Cheng-Jung Lin; “ Goal-oriented sequential pattern for network banking churn analysis”, *Expert Systems with Applications*, vol. 25, pp. 293–302, 2003.
- [35] K. Coussement and Dirk Van den Poel; “ Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques”, *Expert systems with applications*, 34, 313–327, 2008.
- [36] K. Coussement and D. Van den Poel ; “ Integrating the voice of customers through call center emails into a decision support system for churn prediction”, *Information & Management*, vol.49, pp.164–174, 2008.
- [37] K. Coussement, K.W. De Bock; “Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers”,



[59] Yoon, H.S. Kim ; “Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market”, Telecommunications Policy, vol. 28, pp. 751–765, 2004.

[60] Xi. Yu, Sh. Guo, J. Guo and X. Huang; “An extended support vector machine forecasting framework for customer churn in e-commerce”. Expert Systems with Applications , pp.1425–1430, 2011.

dataacademy.ir