Archive of SID

نشریه علمی کاوشهای مدیریت بازرگانی
سال یازدهم، شماره بیست و دوم، پاییز و زمستان ۱۳۹۸

پیش‌بینی پاسخ مشتریان در بازاریابی مستقیم با شبکه‌های عصبی چندلایه

مهدی زکی‌پور[۱]، سینا نعمتی‌زاده[۲*]، محمدعلی افشار کاظمی[۳]
[۱] استادیار، گروه مدیریت بازرگانی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران
[۲] دانشیار، گروه مدیریت بازرگانی، دانشکدهٔ مدیریت و حسابداری، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران
[۳] دانشیار، گروه مدیریت صنعتی، دانشکدهٔ مدیریت و حسابداری، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران

تاریخ دریافت: ۱۳۹۷/۰۸/۳۰ تاریخ پذیرش: ۱۳۹۸/۰۳/۲۱

چکیده

هدف پژوهش حاضر شناسایی هر چه دقیق‌تر مشتریان بالقوه جهت مخاطب قرار دادن در برنامه‌های بازاریابی مستقیم است که از دیرباز به عنوان یکی از مسائل مهم و مورد علاقهٔ بازاریابان شیوهٔ مستقیم مطرح بوده است. مهم‌ترین مسئله در این راستا کاوش در مجموعهٔ داده‌های مشتریان است که همواره از عدم توازن بالایی برخوردار می‌باشد. در این پژوهش با ترکیب روش‌های کم نمونه‌گیری و بیش‌نمونه‌گیری تصادفی کلاس اکثریت و اقلیت که در پژوهش‌های گذشته به کرات استفاده شده، با خوشه‌بندی مشتریان و استخراج نمونه‌های متعادل‌تر اقدام به طراحی و توسعه یک الگوریتم پویا و اثربخش در راستای شناسایی و پیش‌بینی مشتریان بالقوه نموده‌ایم. بدین‌منظور از پایگاه دادهٔ مشتریان یک آژانس مسافرتی (بالغ بر ۱۰۰۰۰ رکورد) استفاده شده است. نتایج حاکی از آن است که با استفاده از داده‌های اولیه مشتریان به هیچ وجه نمی‌توان به یک پیش‌بینی قابل اتکا و استفاده دست‌یافت. بکارگیری روش‌های نمونه‌گیری مجدد با استفاده از خوشه‌بندی مشتریان و ترکیب کلاس‌های اقلیت و اکثریت به روش‌های مختلف و مطابق با الگوریتم ابتکاری ارائه شده می‌تواند توان پیش‌بینی طبقه‌بند درخت تصمیم را به طرز شگفت‌انگیزی افزایش داده و در موقعیت‌ها و بازارهای مختلف مورد استفاده قرار گیرد. در نهایت با ترکیب نتایج حاصل از کدهای XML استخراج شده در هر مرحله و معیار «حاصل‌ضرب» می‌توان به شناسایی و رتبه‌بندی مشتریان بالقوه و هدف‌گذاری آنها به شیوه‌ای کارآمد پرداخت.

کلمات کلیدی: ارتقاء توان پیش‌بینی، عدم تعادل کلاس، داده‌کاوی، شبکهٔ عصبی

* نویسنده مسئول: Email: nematizadeh51@yahoo.com

www.SID.ir

Archive of SID

مقدمه

با توجه به هزینه‌های رو به رشد تبلیغات و بازاریابی انبوه، امروزه بازاریاب‌ها به صورتی فزاینده به پست مستقیم و بازاریابی از راه دور روی آورده‌اند و می‌کوشند به جای استفاده از نیروی فروش و تحمل هزینه‌های روزافزون آن، فروش‌ها را بصورت مستقیم انجام دهند (آرمسترانگ[1] و همکاران، ۲۰۱٤). از طرفی افزایش روزافزون رقابت تولیدکنندگان و نیز گسترش بازار مصرف محصولات و خدمات و پیشرفت‌های علمی و فنی، فرصتی فراهم آمده است تا مدیران بازاریابی با استفاده از تسهیلات مخابراتی و امکانات ارتباطی از این وضعیت بهره جسته، با جلب رضایت مشتریان فعلی و بالقوه وضعیتی ایده‌آل برای سازمان فراهم کنند. کاهش هزینه‌های خرید مشتری، صرفه‌جویی در وقت مشتری از راه خدمت‌رسانی سریع‌تر، عرضه محصول طبق خواست مشتری با قیمت مناسب و ... از جمله راه‌های افزایش رضایت مشتری در این زمینه می‌باشد (جیانگ و روزنبلوم[2]، ۲۰۰۵). یکی از ابزارهای مهمی که می‌تواند در این زمینه کمک موثری نماید، استفاده از ابزارهای مخابراتی و اینترنتی است. توسعهٔ فناوری اطلاعات و ارتباطات، تجارت الکترونیک، توسعه وب سایت‌ها و ... باعث ایجاد پایگاه‌های دادهٔ عظیم[3] و کاملاً به روز شده فرایندهای مدیریت و بازاریابی را شدیداً تحت تاثیر قرار داده است که یکی از علایم آن گرایش بیشتر شرکت‌ها به سوی بازاریابی مستقیم است (آکرول و کاتلر[4]، ۱۹۹۹). در نتیجه، ارتباطات بازاریابی به طور فزاینده‌ای تبدیل به ارتباطات فرد به فرد شده و از طریق کانال‌ها و مجاری دیجیتالی بسط یافته است (دینگ[5] و همکاران، ۲۰۱۵). بازاریابی مستقیم به دنبال توسعه روابط مستقیم و مستمر با مشتریان است، چراکه اکثر بازاریابان بر این باورند که هزینه جذب مشتریان جدید، چندین برابر حفظ مشتریان فعلی است، آنکلس[6] و همکاران (۲۰۰۳) و از این جهت به ایجاد و حفظ روابط مستمر با مشتریان فعلی تأکید فراوان دارند. مسئله‌ای که قبل از حفظ مشتریان وفادار قابل توجه و مهم است چگونگی شناسایی و در واقع پیش‌بینی پاسخ آنهاست.

1- Armstrong
2- Jiang & Rosenbloom
3- Big Data
4- Achrol & Kotler
5- Ding
6- Uncles

www.SID.ir

Archive of SID

مدل‌های پیش‌بینی داده‌محور به طور گسترده‌ای برای پیش‌بینی پاسخ مشتری به کمپین‌های بازاریابی مورد استفاده قرار گرفته است چن و همکاران (۲۰۱۵)، ولی آنچه که باعث ضعف این مدل‌ها وخیم‌تر شدن مسئله می‌شود آن است که نرخ پاسخ در شرایط کلی بازاریابی مستقیم اغلب بسیار پایین است. این امر باعث ایجاد یک پایگاه دادهٔ نامتعادل و در نتیجه اختلال در پییش بینی مشتریان می‌شود. اگر راه حل مناسبی برای بر طرف کردن عدم تعادل کلاس بکار برده نشود، الگوریتم‌های طبقه‌بندی که توسط مدل پاسخ استفاده می‌شوند، احتمالاً اکثر مشتریان را به عنوان کسانی که پاسخی ارائه نمی‌کنند در نظر می‌گیرند، که این امر منجر به ایجاد هزینهٔ فرصت از دست رفته خواهد شد. به همین دلیل، مدیریت عدم تعادل کلاس اطلاعات مشتری، به عنوان یک عامل مهم برای موفقیت بازاریابی مستقیم شناخته شده است (هیل[1] و همکاران، ۲۰۰۶؛ لای[2] و همکاران، ۲۰۰۶؛ لینگ[3] و همکاران، ۱۹۹۸). هدف ما در این پژوهش توسعهٔ الگوریتمی ابتکاری برای ترکیب روش‌های مختلف کاهش عدم توازن کلاس مشتریان (مشتریان پاسخ‌گو و غیر پاسخ‌گو) با استفاده از خوشه‌بندی، کم نمونه‌گیری و بیش‌نمونه‌گیری تصادفی مشتریان و نیز استفاده از ترکیب طبقه‌بندهای مختلف به منظور آموزش سیستم‌های خبره و پیش‌بینی پاسخ مشتریان می‌باشد.

سوالات پژوهش

فلذا در این راستا سوال‌های اصلی پژوهش بدین شرح تدوین شده است؛

آیا داده‌های حاصل از مشتریان می‌تواند مبنای مناسبی برای پیش‌بینی پاسخ و رفتار آتی آنها باشد؟

آیا شبکه‌های عصبی مصنوعی توانایی پیش‌بینی پاسخ مشتریان را دارند؟

اصلاح در سطح داده‌ها، تا چه حد می‌تواند توان پیش‌بینی شبکه‌های عصبی مصنوعی را ارتقاء بخشد؟

1- Hill
2- Lai
3- Ling

www.SID.ir

Archive of SID

مفاهیم نظری و پیشینهٔ عملی پژوهش

بازاریابی مستقیم

در دنیـای رقابتی امروز طراحی محصـــولات و ارائهٔ خدمات مطابق با نیازهای بخش‌های مختلف مشـتریان تبدیل به یک ضرورت اجتناب‌ناپذیر برای سازمان‌ها شده اسـت (کارا و کایناک[1]، ۱۹۹۷). سـازمان‌ها به اهمیت تقسـیم‌بندی بازارها پی برده‌اند و تلاش می‌کننـد تـا با مشـــتریان مختلف به گونه‌ای متفاوت و منطبق با خواســـته‌ها و تمایلات آنها برخورد و تعامل داشـته باشند. مهمترین مسئله‌ای که در این زمینه ظهور و بروز پیـدا می‌کند نبود داده‌های بازار و مشـــتریان نیســـت، بلکه بازاریابان در این حوزه با داده‌های وســـیع و تقریباً نامحدودی روبرو هســـتند. بنابراین چگونگی تحلیل داده‌های موجود مشـــتریان و اســـتخراج دانش نهفته از آنها به منظور شـــناســـایی مشـــتریـان محتمـل و روی‌آوری به ابزارهای بازاریابی مســـتقیم، تبدیل به مهمترین دغدغهٔ بازاریابان شده است (گوبل و گرونوالد[2]، ۱۹۹۹). در این راستا استفاده از فنون و روش‌های داده‌کاوی بیش از پیش راه‌گشـــای بازاریابان می‌باشـــد. در این راســـتا، مدل‌ســـازی پاســـخ[3] (واکنش) یکی از موثرترین ابزارها برای شـرکت‌هایی اسـت که به دنبال برقراری روابط درازمدت با مشتریان خود هستند (سان[4] و همکاران، ۲۰۰۶).

مدل‌سازی پاسخ

مدل‌ســـازی پاســـخ یکی از موثرترین ابزارها برای شـــرکت‌هایی اســـت که به دنبال برقراری روابط درازمدت با مشـتریان خود هسـتند (سان و همکاران، ۲۰۰۶). هدف از مدل‌ســـازی پاســـخ برای بازاریابی مســـتقیم این اســـت که بر اســـاس تاریخچه رفتار مشتری و سایر اطلاعات در دسترس، مشتریانی را که احتمالاً یک محصول رقابتی را خریداری می‌کنند شناسایی کنیم. مدل‌سازی واکنش (پاسخ) یکی از موثرترین ابزارها، برای شرکت‌هایی است که به دنبال برقراری روابط درازمدت با مشتریان خود هستند (بری[5] و همکاران، ۱۹۹۷؛ گونول و هافسـتد[6]، ۲۰۰۶). هدف از مدل‌ســازی واکنش این

1- Kara & Kaynak
2- Goebel & Gruenwald
3- Response Modeling
4- Sun
5- Berry
6- Gönül & Hofstede

www.SID.ir

Archive of SID

است که بر اساس تاریخچه خرید مشتری و سایر اطلاعات در دسترس، مشتریانی را شناسایی کنیم که به احتمال زیاد یک محصول را خریداری می‌کنند. بر اساس پیش بینی‌های مدل، شرکت‌ها تلاش می‌کنند تا خریداران بالقوه بیشتری را برای خرید محصول کمپین شده با استفاده از کانال‌های ارتباطی خودشان مانند تلفن، کاتالوگ ارسال شده یا ایمیل ترغیب و تهییج کنند. یک مدل واکنش به خوبی توسعه یافته می‌تواند به دو طریق به کسب و کار کمک کند. اول اینکه، درآمد کل را افزایش می‌دهد. مشتریان در طی کمپین بازاریابی معمولاً به دو گروه تقسیم می شوند: یک گروه که محصول را خریداری می‌کنند، هرچند که مورد هدف قرار گرفته باشند یا خیر، و گروه دیگری که محصول را نخواهند خرید، و مورد هدف قرار نمی‌گیرند. با یادآوری بموقع آنچه که گروه دوم ممکن است نیاز داشته باشند، ممکن است قانع شوند که کیف پول خود را باز کنند. بنابراین، فروش مازادی که به آن مشتریان انجام می‌شود، سهم آشکار مدل پاسخ است. دوم، مدل واکنش هزینه کل بازاریابی را کاهش می‌دهد. به طور کلی، تبلیغات انبوه بسیار گران است، زیرا متوسط احتمال خرید مشتری بسیار کم است. برخلاف بازاریابی انبوه، مدل پاسخ تنها تلاش برای جذب مشتریانی با احتمال خرید نسبتاً بالا را پیشنهاد می‌کند. بنابراین آن (مدل پاسخ)، موجب صرفه‌جویی در پولی می‌شود که خرج شده تا مشتریانی که علاقه چندانی به خرید محصول نداشته‌اند در معرض پیام‌های تبلیغاتی قرار گیرند (السنر[1] و همکاران، ۲۰۰٤؛ ژانگ و کریشنامورت[2]، ۲۰۰٤).

عدم توازن داده‌ها

مجموعه داده نامتوازن بر اساس تعریف عبارت است از مجموعه داده‌ای که تعداد نمونه‌های متعلق به یک کلاس در آن با تعداد نمونه‌های کلاس دیگر به طور مساوی توزیع نشده باشد (چاولا[3]، ۲۰۰۹). کلاس با تعداد داده‌های بیشتر را کلاس اکثریت و کلاس با داده‌های کمتر را کلاس اقلیت می‌نامند (سان و همکاران، ۲۰۰۹). در الگوریتم‌های طبقه‌بند استاندارد، توزیع کلاس‌ها متوازن در نظر گرفته می‌شود و این دسته از الگوریتم‌ها در مواجهه با مجموعه داده‌های نامتوازن عملکرد مناسبی را از

1- Elsner
2- Zhang & Krishnamurth
3- Chawla

www.SID.ir

Archive of SID

خود ارایه نمی‌دهند؛ چرا که الگوریتم‌های معمول طبقه‌بند به سمت نمونه‌های آموزشی کلاس بزرگ‌تر متمایل می‌شوند که این موضوع باعث افزایش خطا در شناسایی نمونه‌های اقلیت می‌شود. این مساله یکی از چالش‌های پیش رو برای طبقه‌بندی[1] داده‌های نامتوازن محسوب می‌شود و امروزه نظر بسیاری از متخصصان و پژوهشگران حوزه تحلیل داده را به خود جلب کرده است (یانگ و وو[2]، ۲۰۰۶). از روش‌های متنوعی برای حل مساله عدم توازن در علم یادگیری ماشین استفاده می‌شود (گالار[3] و همکاران، ۲۰۱۲). یکی از این روش‌ها، روش‌های بازبینی در سطح الگوریتم[4] است که با تغییر در الگوریتم طبقه‌بندی، به نوعی مساله عدم توازن مرتفع می‌شود (باراندلا[5] و همکاران، ۲۰۰۳). از دیگر روش‌های حل نامتوازن بودن داده‌ها روش‌های مبتنی بر ترکیب طبقه‌بندها است. هدف اصلی روش ترکیب[6] تلاش برای بهبود عملکرد طبقه‌بندی داده‌ها از طریق ترکیب چندین طبقه‌بند است. به‌طوریکه ترکیب چند طبقه‌بند عملکرد بهتری نسبت به یکی از همان طبقه‌بندها خواهد داشت. یان و همکاران با استفاده از این روش و ماشین بردار پشتیبان[7] توانستند مساله پیش‌بینی کلاس اقلیت را بهبود دهند (یان[8] و همکاران، ۲۰۰۳).

روش سوم روش‌های سطح داده[9] است. در این دسته از روش‌ها، توزیع کلاس نامتوازن با نمونه‌گیری مجدد[10] در فضای داده‌ها متوازن می‌شود (ناپیرالا[11] و همکاران، ۲۰۱۰). و نهایتاً روش‌های حساس به هزینه[12] دسته دیگری از روش‌های ارایه شده برای حل عدم توازن در داده‌ها محسوب می‌شود. این دسته از روش‌ها به نوعی از ترکیب روش‌های تغییر در الگوریتم طبقه‌بند و روش‌های سطح داده حاصل می‌شوند (ژانگ و همکاران، ۲۰۰۸).

1- Classification
2-Yang & Wu
3- Galar
4- Algorithm level
5- Barandelaa
6- Ensemble Methodology
7- Support Vector Machine
8- Yan
9- Data Level
10- Re-sampling
11- Napierała
12- Cost-sensitive learning

www.SID.ir

Archive of SID

در این بین، روش‌های سطح داده با دو رویکرد کم‌نمونه‌برداری[1] و بیش‌نمونه‌برداری[2] از روش‌های موثر در متوازن نمودن داده‌ها محسوب می‌شوند (لی[3] و همکاران، ۲۰۱۰). بیش نمونه‌گیری[4] و کم نمونه‌گیری[5] دو دستورالعمل اصلی برای متعادل‌سازی داده است. کم‌نمونه‌گیری، در حالی که تمام موارد کلاس اقلیت را حفظ می‌کند، تعداد نمونه‌های کلاس اکثریت را کاهش می‌دهد. در نتیجه نسبت نهادهای کلاس اقلیت در داده‌های آموزشی افزایش می‌یابد. کم نمونه‌گیری در کاهش زمان آموزش مؤثر است، اما اغلب توزیع کلاس را تحریف می کند، زیرا تعداد زیادی از کلاس اکثریت حذف می‌شوند. نمونه‌گیری تصادفی ساده‌ترین راه برای اجرای کم نمونه‌گیری است. در کم نمونه‌گیری تصادفی، مجموعه‌ای از نمونه‌های کلاس اکثریت به طور تصادفی انتخاب شده و با الگوهای طبقه اقلیت ترکیب می‌شوند. شرینک[6] کوبات[7] و همکاران، (۱۹۹۷) و انتخاب یک طرفه[8] کوبات و ماتوین[9] (۱۹۹۷) از دیگر روش‌های شناخته شده کم نمونه‌برداری می‌باشند.

پیشینه عملی پژوهش

مطالعات گذشته نشان داده است در حالی که افزایش میزان پاسخ کار آسانی نیست، تاثیر آن کاملاً خارق‌العاده است. به عنوان مثال، کوئینن[10] و همکاران (۲۰۰۰) اشاره کرده‌اند که حتی بهبود کمی از میزان پاسخ می‌تواند نتیجه کلی یک کمپین پستی مستقیم از شکست به موفقیت تغییر دهد. بیزنس[11] و همکاران (۲۰۰۲) نشان دادند که چگونه بهبود کوچکی در نرخ پاسخ ممکن است سود عظیمی را به همراه داشته باشد. در مثال آنها، تنها ۱ درصد افزایش در نرخ پاسخگویی برای یک شرکت سفارش پستی، ۵۰۰۰۰۰ یورو اضافه درآمد به‌همراه داشت. نات[12] و همکاران (۲۰۰۲) گزارش

1- Under Sampling
2- Over Sampling
3- Li
4- Over Sampling
5- Under Sampling
6- Shrink
7 Kubat
8- One-Sided Selection (OSS)
9- Kubat & Matwin
10- Coenen
11- Baesens
12- Knott

www.SID.ir

Archive of SID

کردند که برای یک بانک خرده فروشی تنها ۷ درصد افزایش پاسخ کل، درآمد را به میزان سه برابر و درآمد هر پاسخ‌گو را به میزان ۲۰ درصد افزایش داد. سان و همکاران (۲۰۰۶) اشاره کرد که بهبود میزان پاسخ نه تنها می‌تواند سود را افزایش دهد، بلکه موجب تقویت وفاداری مشتری نیز می‌شود، زیرا مشتریانی که به طور صحیحی هدف قرار گرفته‌اند به احتمال بیشتری راضی می‌شوند و مدت زمان طولانی‌تری همراه شرکت می‌ماند. در این بخش از مقاله پیشینه پژوهش بصورت خلاصه در جدول ۱ ارائه می‌گردد.

جدول ۱- پیشینه پژوهش

محقق	سال	عنوان تحقیق	هدف	یافته‌ها
رضائی نوایی و همکاران	۲۰۱۷	به‌کارگیری و ارزیابی تکنیک‌های داده‌کاوی جهت پیش‌بینی رویگردانی مشتری در صنعت بیمه	بررسی علل و پیش‌بینی رویگردانی مشتریان در صنعت بیمه	یافته‌های تحقیق نشان می‌دهد که روش ماشین بردار پشتیبان از عملکرد بالاتری نسبت به سایر روش‌ها برخوردار است. در مدل پیشنهادی مبتنی بر این روش، مشخصه‌های سابقه خرید، نحوه آشنایی با سازمان و تمایل به خرید، به عنوان مشخصه‌های اصلی پیش‌بینی‌کننده رویگردانی مشتری شناسایی شدند.
قاسم نیا عربی و صفایی قادیکلایی	۱۳۹۷	مقایسه عملکرد مدل‌های کلاسیک و هوش مصنوعی در پیش‌بینی وضعیت اعتباری مشتریان بانک	سنجش کارایی مدل رگرسیون لجستیک و شبکه عصبی مصنوعی در تشخیص وضعیت اعتباری مشتریان	مدل‌های آماری با مفروضات کلاسیک نظیر خطی بودن روابط متغیرها، نمی‌توانند ریسک اعتباری مشتریان را به درستی ارزیابی نماید؛ از این رو بکارگیری یا تلفیق تکنیک‌های هوش مصنوعی در این مساله ضرورتا توصیه می‌شود
حری و همکاران	۱۳۹٤	پیش‌بینی رتبه اعتباری مشتریان بانک‌ها با استفاده از الگوریتم فراابتکاری و هیبریدی چند معیاره	پالایش، جذب و نگهداشت بهترین سرمایه‌گذاران، مشاوران، مشتریان و قرض‌گیرندگان	سه شاخص "سن"، "سابقه ارتباط با بانک (مدت حساب)" و "میزان اعتبار" برای پیاده‌سازی مدل ترکیبی عصبی فازی انتخاب گردید و نتایج بیانگر آن می‌باشد که ۸۹/٦۷ درصد از مواقع این سیستم می‌تواند تخمین درستی نسبت به رتبه اعتباری مشتریان ارائه دهد.
بصیری	۱۳۸٦	کاربرد تکنیک داده‌کاوی در مدیریت روابط مشتری	بررسی تلفیق مدیریت ارتباط با مشتری و داده کاوی در جهت پاسخ‌گویی موثر به نیاز مشتری	ابزارهای داده‌کاوی به سئوالاتی از کسب و کار پاسخ می‌دهد که در گذشته پیگیری آنها بسیار وقت‌گیر بوده است. با این حال توانمندی‌های موجود در داده‌کاوی، مدیریت روابط مشتری را به نحو بهتری ممکن می‌سازد
کنگ و همکاران	۲۰۱۲	مدل‌سازی پاسخ بهبودیافته بر اساس خوشه‌بندی کم‌نمونه‌گیری و ترکیب	بررسی تأثیر روش‌های مختلف متعادل‌سازی داده‌ها	خوشه‌بندی، کم‌نمونه گیری و ترکیب به وضوح باعث بهبود ضریب تصحیح متعادل[1] (**BCR**) شده است. این موفقیت با افزایش نرخ پاسخ واقعی[2] (**TR**)، با بالا نگه‌داشتن نرخ عدم پاسخ واقعی[3] (**TN**)، حاصل شده است. دوم، **CUE** مدل پاسخ را پایدار تر از روش های متعادل کننده دیگر ساخته است.

1- Balanced Correction Rate (BCR)
2- True Respondents (TR)
3- True Non-respondents (TN)

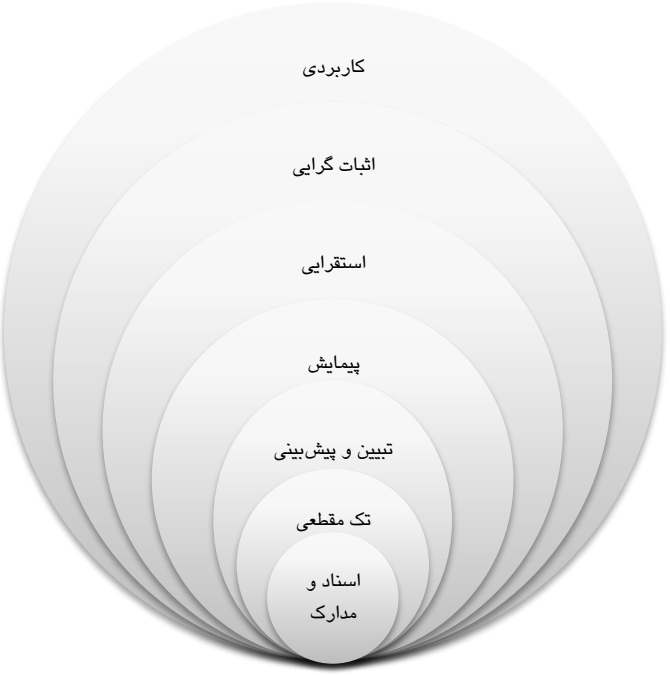
www.SID.ir

Archive of SID

روش‌شناسی پژوهش

نقطه آغازین و روند کلی پژوهش

سوال اساسی و پایه ای در این پژوهش آن بوده است که " آیا داده‌های حاصل از مشتریان می‌تواند مبنای مناسبی برای پیش‌بینی پاسخ و رفتار آتی آنها باشد؟" نظر و رویکرد محقق آن بوده است که اساس پاسخ آتی مشتریان وابستگی زیادی به پاسخ گذشته آنها دارد. اما سوالی که از پی این رویکرد نشئت می گیرد آن است که چرا تکیه بر این داده‌ها نمی‌تواند مبنای مناسبی برای پیش‌بینی‌های آتی باشد؟ و آیا روش‌هایی برای بهره‌برداری بهینه‌تر از این منابع جهت نیل به مقصود که همانا دست‌یابی به توان پیش‌بینی پاسخ مشتریان است وجود دارد.؟جواب به این پرسش‌های کلیدی فرآیندی را خواستار بود که به اختصار در شکل ۱ نمایش داده شده است.



شکل ۱- پیاز پژوهش

با توجه به شکل ۱ مشخص که جهت‌گیری پژوهش کاربردی، مبانی فلسفی پژوهش اثبات‌گرایی، رویگرد پژوهش استقرایی، نوع پژوهش پیمایشی، هدف اصلی پژوهش تبیین و پیش‌بینی و شیوه‌های گردآوری اطلاعات کتابخانه‌ای می‌باشد.

www.SID.ir

Archive of SID

نمونهٔ اولیه پژوهش

به منظور دسترسی به اطلاعات تعداد زیادی از مشتریان، در این پژوهش از پایگاه دادهٔ مشتریان یک آژانس مسافرتی (بالغ بر ۱۰۰۰۰ رکورد) استفاده شده است. روش پیاده‌سازی پژوهش بر این اساس بوده است که از بین مشتریان مورد بررسی، افرادی به عنوان نمونه انتخاب و یک کمپین تبلیغاتی را دریافت کرده‌اند که تعداد آنها ۱۹۱۷ نفر می‌باشد و در نهایت، ۹۷ نفر به کمپین دریافت شده پاسخ داده‌اند. اگر مشتریان را در گروه‌های مختلفی خوشه‌بندی کنیم، مشاهده خواهیم کرد که مشتریان پاسخ‌گو در تمامی خوشه‌های شناسایی شده، به نسبت حجم آنها قرار گرفته‌اند. بنابراین استفادهٔ سنتی از روش‌های خوشه‌بندی مشتریان روش مناسبی برای شناسایی مشتریان محتمل نخواهد بود. بنابراین خوشه‌بندی مشتریان به منظور تهیه و تنظیم استراتژی‌های بازاریابی مستقیم زیاد معقول و منقطی به نظر نمی‌رسد، چرا که ممکن است تعداد مشتریان پاسخ‌دهنده به پیام‌های تبلیغاتی را از طریق حذف مشتریانی که در سایر خوشه‌های انتخاب نشده قرار دارند، بار دیگر کاهش دهد! به منظور فائق آمدن بر این مسئله، ما اقدام به توسعه و معرفی الگوریتمی اثربخش نموده‌ایم؛ به نحوی که پس از گروه‌بندی بازار هدف به دو گروه از مشتریانی که به فعالیت‌های بازاریابی مستقیم شرکت پاسخ داده و نداده‌اند، اقدام به ایجاد دسته‌هایی مرکب از خوشه‌های مختلف مشتریان غیر پاسخ‌گو و پاسخ‌گو به شیوه‌هایی که در ذیل توضیح داده شده است نموده و سپس پاسخ مشتریان را پیش‌بینی و در نهایت عملکرد مدل را مورد بررسی و تحلیل قرار خواهیم داد.

روش‌های متعادل‌سازی داده‌ها، با استفاده از استراتژی‌های مختلف نمونه‌گیری برای هر کلاس یک مجموعه داده جدید آموزشی ایجاد می‌کند که در آن همه کلاس‌ها به خوبی متعادل شده‌اند. آنها نسبت به روش‌های اصلاح الگوریتم دارای یک مزیت فراگیر هستند. از آنجا که روش‌های متعادل‌سازی داده‌ها به طور مستقل از الگوریتم‌های طبقه‌بندی کار می‌کنند، می‌توانند با هر نوع طبقه‌بندی کننده‌ای ترکیب شوند، در حالی که اصلاحات الگوریتم فقط با طبقه‌بندی کننده خاصی که برای آن طراحی شده است به خوبی کار می‌کنند.

در این مرحله مشتریان غیر پاسخ‌گو را با مشتریانی که پاسخ داده‌اند (۹۷ نفر) به چند شیوه ترکیب می‌نماییم. یادآوری می‌کنیم که خوشه‌بندی مشتریان غیر پاسخ‌گو به

www.SID.ir

Archive of SID

منظور مد نظر قرار دادن مشتریان (نمایندگانی) از تمامی خوشه‌های متفاوت مشتریان می‌باشد. ولی سوالی که در اینجا مطرح می‌شود این است که از هر خوشه چه تعداد مشتری (و به چه شیوه‌ای) انتخاب شده و چگونه با مشتریان پاسخ‌گو ترکیب شوند؟ بدین‌منظور می‌توان چندین راه حل ابتکاری به شرح زیر در نظر گرفت؛

به منظور مورد توجه قرار دادن حجم خوشه‌ها، تعداد اعضای انتخاب شده از هر خوشه برابر با نسبت حجم خوشه به کل مشتریان غیر پاسخ‌گو باشد. بدین ترتیب اعضاء نمونه انتخاب شده از مشتریان غیر پاسخ‌گو متناسب با حجم خوشهٔ عضویت آنها خواهد بود و نیز عدم تعادل کلاس هم از بین خواهد رفت. این روش را (C-RUS-01) نامگذاری کرده و در تحلیل‌های مقاله مورد استفاده قرار خواهیم داد.

از ترکیب روش‌های خوشه‌بندی، RUS و ROS استفاده نماییم. در واقع ما با معرفی این روش قصد در پوشش ضعف هر یک از روش‌های اشاره شدهٔ پیشین داریم. روش پیشنهادی ما بدین صورت است که (در راستای استفاده از مزیت ROS) ابتدا مشتریان پاسخ‌گو را ضربدر تعداد خوشه‌های شناسایی شده از مشتریان غیر پاسخ‌گو نماییم (در اینجا ۳۸۸=٤×۹۷) و به همان تعداد از مشتریان غیر پاسخ‌گو اقدام به نمونه‌گیری نماییم. نمونه‌گیری از مشتریان غیر پاسخ‌گو متشکل از کلیه خوشه‌ها و به نسبت حجم آنها خواهد بود. این روش را (C-ROS-RUS-02) نامگذاری کرده و در تحلیل‌های مقاله مورد استفاده قرار خواهیم داد.

از ترکیب روش‌های خوشه‌بندی RUS و ROS استفاده می‌نماییم، بدین‌نحو که کوچکترین خوشهٔ شناسایی شده را ملاک قرار می‌دهیم. از سایر خوشه‌ها به اندازهٔ کوچکترین خوشهٔ شناسایی شده نمونه‌گیری کرده و اندازه کلاس اقلیت را به اندازه جمع نمونه‌های اخذ شده بسط می‌دهیم. این روش را (C-ROS-RUS-03) نامگذاری کرده و در تحلیل‌های مقاله مورد استفاده قرار خواهیم داد.

معیارهای مورد استفاده جهت ارزیابی مدل‌ها

برای ارزیابی کارایی طبقه‌بندها، ساده‌ترین روش تحلیل براساس ماتریس اغتشاش[1] (ماتریس درهم ریختگی) است؛ این ماتریس از ابزارهای مناسب برای بررسی میزان موفقیت و کارایی سیستم‌های طبقه‌بندی محسوب می‌شود و در واقع نمونه‌هایی را که

www.SID.ir

1- Confusion matrix

Archive of SID

برای هر طبقه به درستی و به اشتباه تشخیص داده شده‌اند گزارش می‌کندکه در آن؛ TP[1] (مثبت حقیقی): تعداد نمونه های طبقهٔ مثبت که به درستی طبقه‌بندی شده‌اند، TN[2] (منفی حقیقی): تعداد نمونه‌های طبقهٔ منفی که به درستی طبقه‌بندی شده‌اند، FP[3] (مثبت کاذب): تعداد نمونه‌های طبقه منفی است که به اشتباه مثبت پیش‌بینی شده‌اند و FN[4] (منفی کاذب): تعداد نمونه‌های طبقه مثبت است که به اشتباه منفی پیش‌بینی شده‌اند، می‌باشد. رایج‌ترین معیار مستخرج از این ماتریس، معیار دقت[5] است. معیار دقت، بیان می‌کند دو مقدار TP و TN مهمترین مقادیری هستند که در یک مسئله دو دسته‌ای باید بیشینه شوند. رابطه سایر معیارهای استفاده شده در پژوهش حاضر به شرح زیر می‌باشد:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad Sensivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP} \qquad Precision = \frac{TP}{TP+FP}$$

$$G-Mean = \sqrt{Precision.Recall} \qquad F-Measure = \frac{2.Recall.Precision}{Recall+Precision}$$

از معیار AUC[6] نیز که نشان دهنده سطح زیر منحنی مشخصهٔ عملکرد سیستم (ROC[7]) می‌باشد استفاده شده است. هر چه مقدار این عدد مربوط به یک دسته‌بند بزرگتر (نزدیک به یک) باشد کارایی نهایی دسته‌بند مطلوب‌تر ارزیابی می‌شود. در این پژوهش ما بصورتی ابتکاری اقدام به محاسبه معیار جدیدی نموده‌ایم که از حاصلضرب هفت معیار فوق بدست می‌آید، بدین‌معنی که اگر تمامی معیارهای پیش‌گفته مقدار مناسب و مطلوبی را ارائه دهند حاصلضرب آنها نیز مقداری بالاتری خواهد شد. این معیار را تحت عنوان «حاصلضرب[1]» به شرح زیر معرفی نموده‌ایم؛

Multiple = (Accuracy × Sensitivity × Specificity × Precision × G-mean × F-measure × AUC)

و در نهایت از معیار «حاصلضرب» به عنوان نشانگری برای مطلوبیت پیش‌بینی طبقه‌بند درخت تصمیم با هر نمونهٔ تولید شده استفاده خواهیم کرد.

1- True Positive
2- True Negative
3- False Positive
4- False Negative
5- Accuracy
6- Area Under Curve (AUC)
7- Receiver Operating Characteristic (ROC)

www.SID.ir

Archive of SID

تجزیه و تحلیل داده‌ها

نتایج حاصل از پیش‌بینی شبکه عصبی مصنوعی با توجه به شش نمونه و هشت معیار ارزیابی معرفی شده بصورت خلاصه به شرح جدول (۲) می‌باشد. اعداد ارائه شده در جدول مرتبط با معیارهای ارزیابی هر مدل و نمونهٔ مرتبط با آن می‌باشد. هدف اصلی این تحقیق تهیهٔ نمونه‌ها و طراحی الگوریتمی کارا به منظور استفاده بهینه از مدل‌های پیش‌بینی و نیل به نرخ بالایی از پاسخ مشتری می‌باشد.

جدول ۲- نتایج حاصل معیارهای ارزیابی طبقه‌بندی مشتریان با توجه به روش‌های مختلف نمونه‌گیری

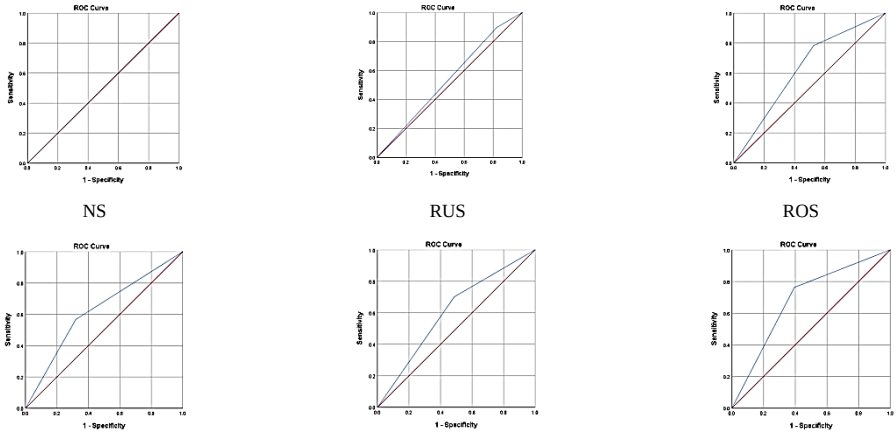
معیارهای ارزیابی مدل		NS	RUS	ROS	C-RUS-01	C-ROS-RUS-02	C-ROS-RUS-03
دقت		۰/۹٤٤	۰/٥۳٦	۰/٦۲۸	۰/٦۲۳	۰/٦۰٦	۰/٦۸٤
حساسیت[1]		۰	۰/۸۹٦	۰/۷۸۲	۰/٥٦۰	۰/۷۰۱	۰/۷٦۲
ویژگی[2]		۱	۰/۱۷٥	۰/٤۷٤	۰/٦۸۰	۰/٥۱۲	۰/٦۰٥
صحت[3]		۰	۰/٥۲۰	۰/٥۹۸	۰/٦۳۹	۰/٥۹۰	۰/٦٥۹
میانگین هندسی[4]		۰	۰/۱٥۷	۰/٦۰۹	۰/٦۲۱	۰/٥۹۹	۰/٦۷۹
معیار F[5]		۰	۰/٦٥۹	۰/٦۷۸	۰/٦۰۱	۰/٦٤۰	۰/۷۰۷
معیار راک	سطح زیر منحنی	۰/٥۰۰	۰/٥۳٦	۰/٦۲۹	۰/٦۲٤	۰/٦۰۷	۰/٦۸٤
	انحراف استاندارد	۰/۰۳۰	۰/۰٤۱	۰/۰۰۹	۰/۰٤۰	۰/۰۲۰	۰/۰۰۹
	معنی‌داری	۱/۰۰۰	۰/۳۸٥	۰/۰۰۰	۰/۰۰۳	۰/۰۰۰	۰/۰۰۰
حاصل‌ضرب		۰	۰/۰۰۲	۰/۰۳٦	۰/۰۳٥	۰/۰۲۹	۰/۰٦۸

همان‌گونه که از جدول (۲) می‌توان فهمید، نمونهٔ اولیه (NS) به هیچ وجه توانایی پیش‌بینی مشتریان بالقوه را نخواهد داشت. بیشترین مقدار معیار «حاصل‌ضرب» با نمونهٔ C-ROS-RUS-03 بدست آمده است. بنابراین در این مرحله این نمونه، به عنوان نمونه‌ای کارآمد جهت آموزش طبقه‌بند درخت تصمیم درنظر گرفته می‌شود. خروجی کدهای XML جهت پیش‌بینی پاسخ مشتریان (۸۰۸۳ نفر) پیوست (۱) ارائه شده است. منحنی‌های راک (ROC) با توجه به شش نمونهٔ مورد بررسی بصورت زیر ترسیم شده است:

1- Sensitivity
2- Specificity
3- Precision
4- G-mean
5- F-measure

www.SID.ir

Archive of SID



شکل ۲- منحنی‌های راک مربوط به نمونه‌های مورد بررسی

با توجه به معیارهای ارائه شده و سطح زیر منحنی نمودارهای راک مشخص است که نمونهٔ C-ROS-RUS-03 بهترین پیش‌بینی را با طبقه‌بند درخت تصمیم بعمل آورده است.

جمع‌بندی و نتیجه‌گیری

نتایج پژوهش

هدف از بازاریابی مستقیم کاهش هزینه‌های زیاد تبلیغات عمومی در رسانه‌های پربیننده و از طرفی پرهزینه است. از سوی دیگر عدم ارائه پاسخ ایده‌آل از مشتریانی که به شیوه بازاریابی مستقیم مورد هدف بازاریابان بوده‌اند، مشکل اساسی و عمدهٔ بازاریابی مستقیم است. تحقیقات زیادی نشان داده است که ارسال پیام‌های ارتباطی و تبلیغاتی به مشتریانی که به هر دلیل تمایل یا نیاز به محصولات شرکت را ندارند موجب ایجاد ذهنیت بد در مشتری و یا حتی تحریم شرکت توسط آنها خواهد شد (آنتونیسن[1]، ۲۰۰۸). بازاریابی مستقیم همیشه و در مورد همهٔ محصولات کاربرد اثربخشی ندارد. به عنوان مثال همانگونه که در پژوهش حسینی و همکاران (۱۳۹۲) عنوان شده است فروش حضوری، پیشبرد فروش و تبلیغات به ترتیب مؤثرتری عوامل در افزایش فروش رنگ مو بوده، ولی بازاریابی مستقیم تفثیر معناداری بر اف ایش فروش ای محصول ندارد و چه بسا این مسئله به خاطر نوع محصول و درگیری

www.SID.ir

1- Anthonissen

Archive of SID

پائین ذهنی مصرف کننده به هنگام تصمیم‌گیری در مورد محصولات پر مصرف و کم‌هزینه است. بنابراین اولین گام در بازاریابی مستقیم، شناسایی محصولاتی است که بازاریابی مستقیم در خصوص آنها روش اثربخشی برای متقاعد کردن مشتری باشد. سپس شناسایی مشتریان بالقوه و هدفگذاری آنان هدف اصلی می‌باشد ولی بزرگترین مشکل در این راستا عدم توازن پایگاه دادهٔ مشتریان و بالطبع ناتوانی مدل‌های پیش‌بینی و طبقه‌بندی مشتریان در انجام این مهم است.

با توجه به نمودارهای راک شش‌گانه که شکل ۲ نشان داده شده است و نیز سطوح زیر منحنی مرتبط با هر نمودار که در جدول ۲ ارائه شده است می‌توان پی‌برد که پیش‌بینی صحیح پاسخ مشتریان با نمونه‌های اولیه و اصلی به هیچ‌وجه میسر نمی‌باشد. همچنین تفاوتهای فاحش بین نمودارهای و سطح زیر منحنی‌های راک بیانگر این مطلب است که با اصلاح داده‌های ورودی مدل‌های پیش‌بینی می‌توان دقت و حساسیت پیش‌بینی‌های بعمل آمده را از صفر تا حدود ۷۰٪ ارتقاء داد که مقدار بسیار چشم‌گیری است. در پژوهش حاضر خوشه‌بندی مشتریان غیرپاسخ‌گو، و ترکیب آنها با روش‌های مختلف با مشتریان پاسخ‌گو، به عنوان ابزاری در جهت متعادل‌سازی پایگاه دادهٔ مشتریان معرفی شد. بدیهی است استفاده از شیوه‌های مختلف خوشه‌بندی، متغیرهای مختلف برای انجام این کار، تعداد خوشه‌های شناسایی یا تعیین شده و روش‌های مختلف ترکیب اعضاء خوشه‌ها به طرق مختلف و بسیار متنوعی قابل اجراست که به جهت جلوگیری از تفصیل مقاله از ارایهٔ آنها خودداری نمودیم. هدف پژوهش حاضر دست‌یابی به بهینه‌ترین پیش‌بینی برای یک جامعهٔ خاص یا مورد بررسی در این مقاله نمی‌باشد، بلکه هدف اصلی ارایهٔ یک الگو و الگوریتمی ابتکاری است که بصورتی پویا برنامه‌ریزی شده و بهترین روش ترکیب و نمونه‌برداری را جهت پیش‌بینی مشتریان در هر جامعه یا بازاری انجام دهد. بطور خلاصه، لزوم ارزیابی مدل‌های مختلف خوشه‌بندی و تحلیل خوشه‌های بدست آمده از هر روش، انتخاب و ارزیابی روش‌های مختلف ترکیب خوشه‌ها با مشتریان پاسخ‌دهنده، و در نهایت استفاده از ابزارهای مختلف پیش‌بینی از قبیل رگرسیون لجستیک، شبکه عصبی مصنوعی، درخت تصمیم و ... از جملهٔ گام‌های مهم الگوریتم پیشنهاد شده توسط محقق می‌باشد که در بخش پیشنهادات کابردی نشان داده شده است.

www.SID.ir

Archive of SID

پیشنهادات کاربردی

ما در این تحقیق به دنبال اثبات توان یا ضعف شبکه‌های عصبی به عنوان ابزاری برای پیش‌بینی مشتریان بالقوه و یا معرفی یک روش نمونه‌گیری خاص برای تحقق بهترین نتیجه نبوده‌ایم، بلکه هدف ما از این بررسی نیل به الگوریتمی کارآمد و پویا در جهت پیش‌بینی مشتریان بالقوه است. حال آنکه این پیش‌بینی می‌تواند با مدل‌های مختلف طبقه‌بندی و یا نمونه‌های مختلف به عمل آید ولی آنچه مهم است توان بالا و کارآمدی این الگوریتم پویا در پیش‌بینی است برای مدیران و فعالان حوزه بازاریابی مستقیم بسیار شگفت‌انگیز خواهد بود.

در این الگوریتم پیشنهادی و کاربردی، ما ابتدا مشتریان را به دو گروه مشتریان پاسخ‌گو و مشتریان غیرپاسخ‌گو طبقه‌بندی نموده‌ایم. سپس مشتریان غیر پاسخ‌گو را خوشه‌بندی کرده و جهت کاهش عدم توازن کلاس آنها را به شیوه‌های مختلف با مشتریان پاسخ‌گو ترکیب می‌نماییم. نمونه‌های حاصل با استفاده از مدل‌های شبکه‌های عصبی مصنوعی مورد آزمون قرار گرفته و معیارهای ارزیابی مدل و مهم‌تر از آن معیار «ضریب» در هر مرحله ثبت و ضبط می‌شود. سپس اقدام به استخراج کدهای XML مربوط به هر مدل نموده و سایر مشتریان را مورد پیش‌بینی قرار می‌دهیم. هر مشتری که به عنوان مشتری پاسخ‌گو در هر مرحله شناسایی می‌شود یک امتیاز گرفته، و امتیاز مربوطه ضرب‌در معیار «ضریب» مربوط به مدل پیش‌بینی می‌شود. در پژوهش حاضر در بهترین حالت، یک مشتری می‌تواند در تمامی نمونه‌های هفت‌گانه (که بسیار به ندرت اتفاق می‌افتد) به عنوان مشتری بالقوه شناسایی شود و یا می‌تواند در شش نمونه، پنج نمونه و ... به عنوان مشتری بالقوه شناخته شود. بنابراین جمع ضرایب مشتری نوعی در هر نمونه‌ای که در آن به عنوان یک مشتری بالقوه شناخته شده است می‌تواند امتیاز نهایی آن مشتری را تشکیل دهد. در نهایت پیشنهاد می‌شود تمامی مشتریان موجود در پایگاه داده مشتریان با توجه به امتیازات کسب شده رتبه‌بندی شده و مورد هدف بازاریابان شیوهٔ مستقیم قرار گیرند. فلوچارت الگوریتم توضیح داده شده در شکل (۳) ترسیم شده است؛

پیشنهاد برای تحقیقات آتی

دو زمینهٔ تحقیقات آتی به محققان علاقمند پیشنهاد می‌شود؛

www.SID.ir

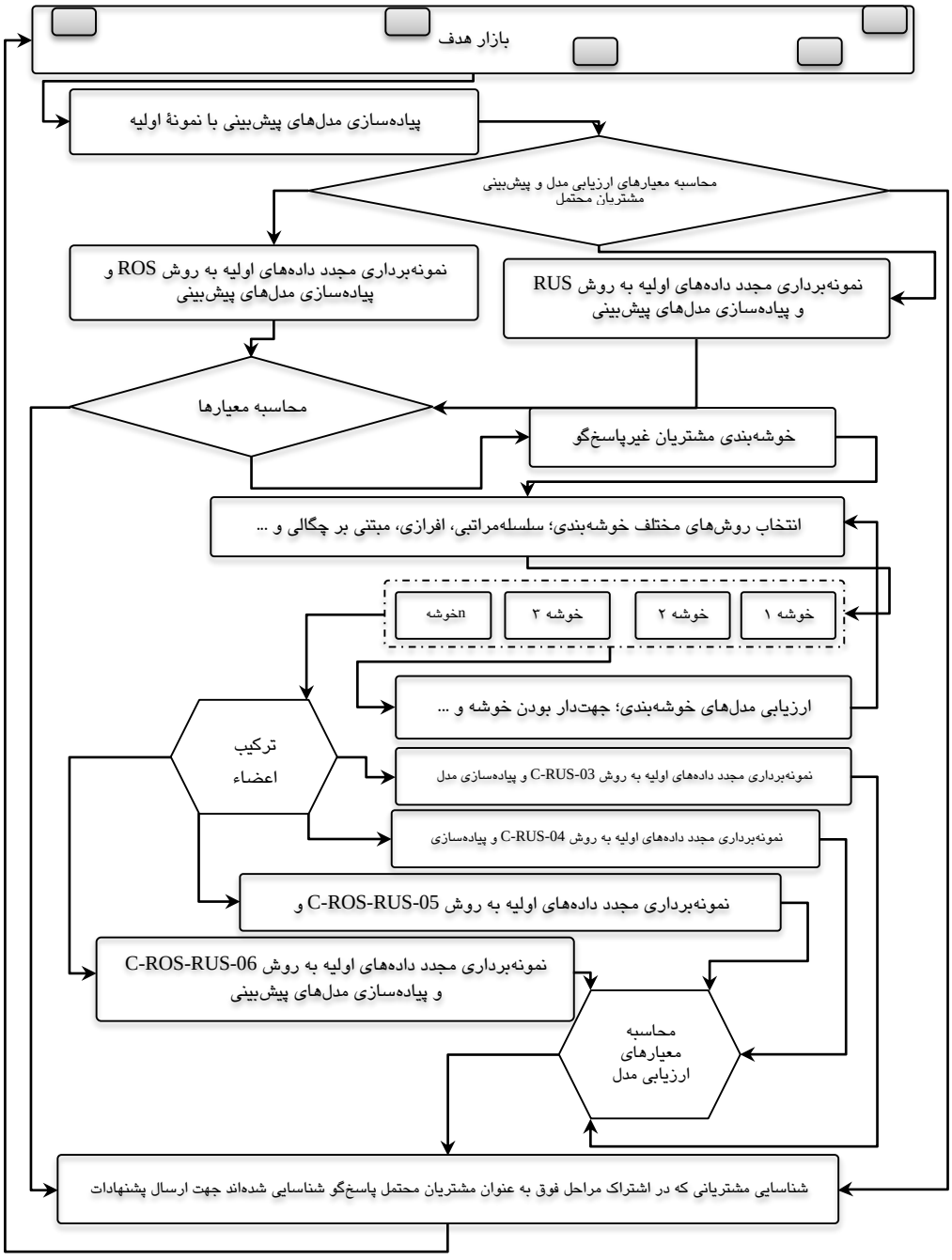
Archive of SID

با توجه به اینکه شبکه‌های عصبی مصنوعی ابزاری قدرتمند در طبقه‌بندی و شناسایی مشتریان و نیز شناسایی متغیرهای مهم در راستای طبقه‌بندی آنها می‌باشند، پیشنهاد می‌شود از نتایج حاصل از شبکه‌های عصبی در زمینهٔ پیش‌بینی مشتریان و شناسایی متغیرهای اثربخش به عنوان بازخوردی جهت تهیه نمونه‌های ورودی استفاده شود. بدین نحو که متغیرهای تأثیرگذار در مدل‌های پیش‌بینی شبکهٔ عصبی به عنوان متغیرهای مهم در زمینهٔ خوشه‌بندی مشتریان غیرپاسخ‌گو در نظر گرفته شده و با توجه به این متغیرها، یکبار دیگر اقدام به خوشه‌بندی و ترکیب مشتریان غیرپاسخ‌گو با مشتریان پاسخ‌گو (برای هر روش نمونه‌گیری مجدد) بصورت مجزا نمایند.

در این پژوهش ما به دنبال یافتن مشتریانی بودیم که به احتمال زیاد در خصوص یک پیشنهاد خاص ارائه شده از سوی شرکت پاسخ خواهد داد. پیشنهاد ارائه شده در خصوص یک محصول یا کالا به مشتری ارائه شده است و یا اینکه می‌تواند بصورت یک تخفیف برای طیف خاصی از کالاها یا خدمات ارائه گردد. به هر صورت مسئله‌ای که گاها مشاهده شده است این است که همه مشتریانی که خرید می‌کنند و حتی خرید خود را تکرار کرده و جزو مشتریان وفادار شرکت به حساب می‌آیند ممکن است سودآور نباشند. در مقابل برخی از مشتریان بی‌وفا هم ممکن است سودآور باشند. سوالی که مطرح می‌شود این است که «پس شرکت کدامیک از مشتریان را باید حفظ کرده و کدامیک از آنها را حذف کند؟» تکلیف کسانی که گه‌گاه زیاد خرج می‌کنند یا دائماً کمی خرج می‌کنند چیست؟ بنابراین پیشنهاد می‌شود در پژوهش‌های آتی ضمن تلاش در جهت تشخیص مشتریان محتمل خرید و یا به عبارتی شناخت مشتریان وفادار اقدام به شناسایی مشتریان ارزشمند نماییم که معمولاً ترکیبی از وفاداری و سودمندی را در کارنامه خودشان دارند.

www.SID.ir

Archive of SID



شکل ۳- فلوچارت الگوریتم پیشنهادی پژوهش

www.SID.ir

Archive of SID

با توجه به الگوریتم معرفی شده در بخش پیشنهادات کاربردی، مشخص است که ادامهٔ چرخهٔ طراحی شده در هر مرحله منجر به پیش‌بینی مشتریان و در نهایت مشاهدهٔ عملکرد واقعی آنان خواهد شد. اطلاعات ثبت شده در هر مرحله منجر به تولید شاخص‌های جدید خواهد شد که به عنوان خریدها و پیش‌بینی خریدهای مشتریان ثبت می‌گردد. فلذا پیشنهاد می‌شود پژوهشگران آتی با تهیهٔ چنین بانک اطلاعاتی اقدام به تحلیل RFM مشتریان نمایند.

پیشنهاد می‌شود در پژوهش‌های آتی به حداکثرسازی سود کمپین توجه شود. به عبارتی دیگر پژوهش حاضر صرفاً در جستجوی مشتریان محتمل به خرید یا پاسخ بود و به عبارتی هدف در این پژوهش یافتن مشتریان پاسخ‌گو بود. ولی شاید شرایطی وجود داشته باشد که صرف خرید مشتری برای شرکت سود زیادی به ارمغان نیاورد. شرایطی که منجر به حداکثرسازی هزینه‌ها می‌شود. بنابراین نوع و هزینه کمپین انتخاب شده نیز باید مدنظر قرار گیرد. از طرف دیگر برخی از مشتریان ممکن است با هدف متضرر نمودن شرکت اقدام به خرید نمایند! مشتریانی که مترصد تخفیفات عمده بوده و به محض دریافت چنین پیشنهادهایی در حجم کلان خرید کرده و بعد از اتمام تخفیف اجناس خریداری شده را فروخته یا به خود شرکت پس می‌دهند! بنابراین شناخت مشتریان پاسخ‌گو و تحلیل بیشتر رفتار آنها می‌تواند پیشنهاد دیگری برای توسعهٔ تحقیقات آتی باشد.

محدودیت‌های پژوهش

یکی از مهمترین محدودیت‌هایی که در انجام این پژوهش بدان توجهی نشده است، ناهمگونی ارزش‌های مشتریان مختلف می‌باشد. در مطالعهٔ مطالعات پیشین نیز این مسئله مشهود است، بطوریکه محققان مختلف نیز به مجموعه‌ای از عوامل توجه نموده‌اند که تأثیر مستقیمی در پاسخ به پیشنهاد ارائه شده دارند. با این حال پاسخ ارائه شده، همیشه در بین همه مشتریان مختلف با ارزش‌های مختلف یکسان نخواهد بود. با توجه به اینکه ارزش‌ها به عنوان متغیری مهم و اثربخش در بین مشتریان مختلف متفاوت است، مهم است که در پژوهش‌های آتی، ساز و کاری به منظور شناخت ارزش‌های مشتری انجام و سپس پاسخ‌مشتریان با ارزش‌های مختلف به کمپین‌های مختلف ارائه شده مدنظر قرار گیرد.

www.SID.ir

Archive of SID

آنچه به عنوان بزرگترین محدودیت در راستای انجام این پژوهش مشاهده شد عدم در اختیار قرار داشتن داده‌های کافی، جامع و صحیح از مصرف‌کنندگان است. بسیاری از شرکت‌ها نیز از ارایهٔ داده‌های مشتریان خود امتناع کرده و آن‌را نوعی تعرض به حقوق مصرف‌کنندگان می‌دانند. این شرکت‌ها فارغ از علم به اهمیت هدفگذاری صحیح مشتریان هنوز به تبلیغات گسترده و عمومی پرداخته و با بدست آوردن موفقیت‌های نسبی، خود را بی‌نیاز از پیاده‌سازی و اجرای پژوهش‌های مشابه برای بهبود عملکرد بازاریابی شرکت می‌بینند.

منابع و مآخذ

1. Basiri, Mehdi, "Application of Data Mining Technique in Customer Relationship Management", National Conference on E-Commerce, Ministry of Industry, Mine and Trade, Iranian E-Commerce Association, 2007, Volume 4.
2. Hari, Sadegh, Mehdi Mahdavi, & Kaveh. (2015). "Designing a Model for Predicting Credit Rating of Bank Customers Using Multi-Criteria Hyper-Algorithm and Ant-Fuzzy-Colony Neural Network Algorithm (Case Study of Tehran Bank Post Branches)". Management Research in Iran, 19 (1), 91-116.
3. Hosseini. Mir Hassan, the Durban. Behnam, trustee, Maryam. (1392). "Assessing the Role of Promotion Mix in Increasing Sales of Cosmetics". Journal of Business Management Research, 5 (9), 21-38.
4. Rezaei Navaei, Samira, Kousha, Hamid Reza, "Application and Evaluation of Data Mining Techniques to Predict Customer Turnover in the Insurance Industry", International Journal of Industrial Engineering and Production Management, Volume (27), Number (4), Vol (1-2017), Pages (635-653.(
5. Qasim Nia Arab. N, Safa Qadiklaee. A. (1397). "Comparison of Performance of Classic and Artificial Intelligence Models in Predicting Bank Clients' Credit Status". Journal of Business Management Research, 10 (20), 51-69.
6. Anthonissen, P. (2008). Crisis communication: Practical PR strategies for reputation management & company survival: Kogan Page Publishers.
7. Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. European Journal of Operational Research, 138(1), 191-211 .
8. Barandelaa, R., Sanchezb, J., & Garcia, V. (2003). Strategies for learning in class imbalance problems .
9. Berry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support: John Wiley & Sons, Inc.
10. Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview Data mining and knowledge discovery handbook (pp. 875-886): Springer.
11. Chen, Z.-Y., Fan, Z.-P., & Sun, M. (2015). Behavior-aware user response modeling in social media: Learning from diverse heterogeneous data. European Journal of Operational Research, 241(2), 422-434 .
12. Coenen, F., Swinnen, G., Vanhoof, K., & Wets, G. (2000). The improvement of response modeling: combining rule-induction and case-based reasoning. Expert Systems with Applications, 18(4), 307-313.
13. Ding, A. W., Li, S., & Chatterjee, P. (2015). Learning user real-time intent for

www.SID.ir

Archive of SID
optimal dynamic web page transformation. Information Systems Research, 26(2), 339-359 .
14. Elsner, R., Krafft, M., & Huchzermeier, A. (2004). Optimizing Rhenania's direct marketing business through dynamic multilevel modeling (DMLM) in a multicatalog-brand environment. Marketing Science, 192-206 .
15. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4), 463-484 .
16. Gönül, F. F., & Hofstede, F. T. (2006). How to compute optimal catalog mailing decisions. Marketing Science, 25(1), 65-74 .
17. Kara, A., & Kaynak, E. (1997). Markets of a single customer: exploiting conceptual developments in market segmentation. European journal of marketing, 31(11/12), 873-895 .
18. Knott, A., Hayes, A., & Neslin, S. A. (2002). Next product to buy models for cross selling applications. Journal of interactive Marketing, 16(3), 59-75 .
19. Kubat, M., Holte, R., & Matwin, S. (1997). Learning when negative examples abound. Paper presented at the European Conference on Machine Learning.
20. Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced training sets: one-sided selection. Paper presented at the Icml.
21. Li, D.-C., Liu, C.-W., & Hu, S. C. (2010). A learning method for the class imbalance problem with medical data sets. Computers in biology and medicine, 40(5), 509-518 .
22. Napierała, K., Stefanowski, J., & Wilk, S. (2010). Learning from imbalanced data in presence of noisy and borderline examples. Paper presented at the International Conference on Rough Sets and Current Trends in Computing.
23. Sun, B., Li, S., & Zhou, C. (2006). "Adaptive" learning and "proactive" customer relationship management. Journal of Interactive Marketing, 20(3-4), 82-96 .
24. Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence, 23(04), 687-719 .
25. Uncles, M. D., Dowling, G. R., & Hammond, K. (2003). Customer loyalty and customer loyalty programs. Journal of consumer marketing, 20(4), 294-316 .
26. Yan, R., Liu, Y., Jin, R., & Hauptmann, A. (2003). On predicting rare classes with SVM ensembles in scene classification. Paper presented at the Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on.
27. Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. International Journal of Information Technology & Decision Making, 5(04), 597-604 .
28. Zhang, J., & Krishnamurthi, L. (2004). Customizing promotions in online stores. Marketing Science, 23(4), 561-578 .
29. Zhang, S., Liu, L., Zhu, X., & Zhang, C. (2008). A strategy for attributes selection in cost-sensitive decision trees induction. Paper presented at the Computer and Information Technology Workshops, 2008. CIT Workshops 2008. IEEE 8th International Conference on.
30. Armstrong, G., Adam, S., Denize, S., & Kotler, P. (2014). Principles of marketing: Pearson Australia.
31. Jiang, P., & Rosenbloom, B. (2005). Customer intention to return online: price perception, attribute-level performance, and satisfaction unfolding over time. European journal of marketing, 39(1/2), 150-174 .
32. Achrol, R. S., & Kotler, P. (1999). Marketing in the network economy. The Journal of Marketing, 146-163 .
33. Hill, S., Provost, F., & Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science, 21(2), 256-276 .
34. Lai, Y.-T., Wang, K., Ling, D., Shi, H., & Zhang, J. (2006). Direct marketing when there
www.SID.ir

Archive of SID
are voluntary buyers. Paper presented at the Data Mining, 2006. ICDM'06. Sixth International Conference on.
35. Ling, C. X., & Li, C. (1998). Data mining for direct marketing: Problems and solutions. Paper presented at the Kdd.
36. Kara, A., & Kaynak, E. (1997). Markets of a single customer: exploiting conceptual developments in market segmentation. European journal of marketing, 31(11/12), 873-895 .
37. Goebel, M., & Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. ACM SIGKDD explorations newsletter, 1(1), 20-33.
www.SID.ir

Archive of SID

پیوست (۱)؛ نمونه‌ای از کدهای XML حاصل از پیش‌بینی شبکهٔ عصبی مصنوعی


```
<?xml version="1.0"?>
<PMML xsi:schemaLocation="http://www.dmg.org/PMML-4_3 pmml-4-3.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.dmg.org/PMML-4_3"
version="4.3">
  - <Header copyright="Copyright (c) IBM Corp. 1989, 2017.">
        <Application version="25.0.0.0" name="IBM SPSS Statistics"/>
        <Timestamp/>
    </Header>
  - <DataDictionary numberOfFields="8">
        <DataField name="Children" dataType="double" isCyclic="0" optype="continuous"
            displayName=""/>
        <DataField name="Gender" dataType="double" isCyclic="0" optype="categorical" displayName=""/>
      - <DataField name="Education" dataType="double" isCyclic="0" optype="ordinal" displayName="">
            <Value property="valid" displayValue="Some high school or less" value="1"/>
            <Value property="valid" displayValue="High school" value="2"/>
            <Value property="valid" displayValue="Some college" value="3"/>
            <Value property="valid" displayValue="College" value="4"/>
            <Value property="valid" displayValue="Post-graduate" value="5"/>
        </DataField>
        <DataField name="Age" dataType="double" isCyclic="0" optype="continuous" displayName=""/>
        <DataField name="Reside" dataType="double" isCyclic="0" optype="continuous"
            displayName="Years at current residence"/>
      - <DataField name="Married" dataType="double" isCyclic="0" optype="categorical" displayName="">
            <Value property="valid" displayValue="No" value="0"/>
            <Value property="valid" displayValue="Yes" value="1"/>
        </DataField>
      - <DataField name="Income" dataType="double" isCyclic="0" optype="ordinal"
        displayName="Income category (thousands) ">
            <Value property="valid" displayValue="<25" value="1"/>
            <Value property="valid" displayValue="25-49" value="2"/>
            <Value property="valid" displayValue="50-74" value="3"/>
            <Value property="valid" displayValue="75+" value="4"/>
        </DataField>
      - <DataField name="Responded" dataType="double" isCyclic="0" optype="categorical"
        displayName="Responded to test offer">
            <Value property="valid" displayValue="No" value="0"/>
            <Value property="valid" displayValue="Yes" value="1"/>
        </DataField>
    </DataDictionary>
  - <TreeModel functionName="classification" algorithmName="CRT">
      - <Extension extender="spss.com">
            <X-risk value="0.463917525773196"/>
            <X-seOfRisk value="0.0358043117150723"/>
        </Extension>
      - <MiningSchema>
            <MiningField name="Income" usageType="active"/>
            <MiningField name="Responded" usageType="predicted"/>
        </MiningSchema>
      - <ModelStats>
          - <UnivariateStats field="Income">
                <Counts invalidFreq="0" missingFreq="0" totalFreq="194"/>
            </UnivariateStats>
        </ModelStats>
      - <Node id="0" recordCount="194" score="0">
          - <Extension>
              - <X-Node>
                    <X-NodeStats improvement="0.00543357728986443"/>
```


www.SID.ir

Archive of SID


```
            </X-Node>
        </Extension>
        <True/>
    - <ScoreDistribution value="0" recordCount="97">
        <Extension name="probability" value="0.5" extender="spss.com"/>
      </ScoreDistribution>
    - <ScoreDistribution value="1" recordCount="97">
        <Extension name="probability" value="0.5" extender="spss.com"/>
      </ScoreDistribution>
    - <Node id="1" recordCount="27" score="0">
      - <CompoundPredicate booleanOperator="surrogate">
        - <CompoundPredicate booleanOperator="and">
            <True/>
            <SimplePredicate value="1" field="Income" operator="lessOrEqual"/>
          </CompoundPredicate>
          <False/>
        </CompoundPredicate>
      - <ScoreDistribution value="0" recordCount="17">
          <Extension name="probability" value="0.62962962962963" extender="spss.com"/>
        </ScoreDistribution>
      - <ScoreDistribution value="1" recordCount="10">
          <Extension name="probability" value="0.37037037037037" extender="spss.com"/>
        </ScoreDistribution>
      </Node>
    - <Node id="2" recordCount="167" score="1">
      - <CompoundPredicate booleanOperator="surrogate">
        - <CompoundPredicate booleanOperator="and">
            <SimplePredicate value="1" field="Income" operator="greaterThan"/>
            <True/>
          </CompoundPredicate>
          <True/>
        </CompoundPredicate>
      - <ScoreDistribution value="0" recordCount="80">
          <Extension name="probability" value="0.479041916167665" extender="spss.com"/>
        </ScoreDistribution>
      - <ScoreDistribution value="1" recordCount="87">
          <Extension name="probability" value="0.520958083832335" extender="spss.com"/>
        </ScoreDistribution>
      </Node>
  </Node>
- <Extension>
  - <X-TreeModel>
    - <X-PredictorImportanceList>
        <X-PredictorImportance importance="100" predictorName="Income"/>
      </X-PredictorImportanceList>
    - <X-Priors>
        <X-Prior-Value value="0.5" targetCategory="0"/>
        <X-Prior-Value value="0.5" targetCategory="1"/>
      </X-Priors>
    </X-TreeModel>
  </Extension>
</TreeModel>
</PMML>
```


www.SID.ir

Archive of SID

Improvement of customer response prediction in direct marketing by neural networks

Mehdi Zakipour[1], Sina Nematizadeh[2*], Mohamdali Afsharkezemi[3]

[1]Department of Business Management, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[2]Associate Professor, Faculty of Management and Accounting, Islamic Azad University, Central Tehran Branch, Tehran, Iran
[3]Associate Professor, Faculty of Management and Accounting, Islamic Azad University, Central Tehran Branch, Tehran, Iran

Received: 21-11-2018 Accepted: 11-06-2019

Abstract

The purpose of this study is to identify the potential customers to address in direct marketing programs, which has been regarded as one of the most important issues for direct marketers. The most important matter is the customer's data set, which is always highly imbalanced. In this study, by combining the random under-sampling and over-sampling of the majority and minority classes that have been used frequently in past studies, we have designed and developed a dynamic and effective algorithm to identify and predict potential customers by clustering customers and extracting more balanced samples. For this purpose, a travel agency database (of over 10,000 records) has been used. The results indicate that customer's raw data cannot make a reliable prediction. On the other hand, Re-sampling methods using customer clustering and combining of minority and majority classes according to the proposed algorithm dramatically increases the prediction power of the decision tree and can be used in different situations and markets. Finally, by combining the results of the extracted XML codes and "multiple" criterion in each step, we can identify and rank potential customers and target them in an efficient way.

Keywords: Prediction optimization, Class imbalance, Data mining, Neural network

www.SID.ir

(*-Corresponding Author Email: nematizadeh51@yahoo.com)