طراحی الگوریتمی پویا برای کاوش در داده‌ها و پیش‌بینی صحیح پاسخ مشتری

مهدی زکی‌پور[۱]، سینا نعمتی‌زاده [۲،*]، محمدعلی افشار کاظمی[۳]

[۱]دانشجوی دکتری، گروه مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران

[۲]دانشیار، گروه مدیریت بازرگانی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران (عهده دار مکاتبات)

[۳]دانشیار، گروه مدیریت صنعتی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: آذر ۱۳۹۷، اصلاحیه: آذر ۱۳۹۷، پذیرش: دی ۱۳۹۷

چکیده

مسئله شناسایی و پیش‌بینی پاسخ مشتریان جهت مخاطب قرار دادن در برنامه‌های بازاریابی مستقیم از دیرباز به عنوان یکی از مسائل مهم و مورد علاقهٔ بازاریابان شیوهٔ مستقیم مطرح شده است. بازاریابانی که از این رویکردها استفاده می‌کنند با خطر واکنش شدید مصرف‌کنندگانی که بازاریابی مستقیم را تعرض به زندگی خصوصی خود تلقی می‌کنند و از آن رنجیده خاطر می‌شوند روبرو هستند، بنابراین حتی ممکن است شرکت‌هایی را که از این روش‌ها استفاده می‌کنند به کلی تحریم کنند. شبکه‌های عصبی به عنوان ابزاری قدرتمند برای پیش‌بینی پاسخ مشتریان بالقوه شناخته شده‌اند ولی همان‌گونه که در پژوهش‌های گذشته نیز عنوان شده است همانند سایر الگوریتم‌های پیش‌بینی در مواجهه با داده‌های نامتوازن تمایل زیادی به انحراف در نتایج دارند. در این پژوهش به منظور ارتقاء توان شناسایی و پیش‌بینی مشتریان بالقوه توسط شبکه‌های عصبی مصنوعی چندلایه ضمن استفاده از روش‌های متداول متعادل‌سازی داده‌ها که در سایر مقاله‌ها به کرات استفاده شده است، اقدام به ایجاد ترکیب‌های مختلف از مشتریان توسط خوشه‌بندی آنها کرده و در نهایت به معرفی یک الگوریتم ابتکاری و بسیار کارآمد جهت شناسایی و رتبه‌بندی مشتریان بالقوه پرداخته‌ایم. نتایج حاکی از آن است که ایجاد ترکیب بهینه‌ای از داده‌های مشتریان و پیاده‌سازی الگوریتم پویای پژوهش حاضر به طرز قابل توجهی موجب ارتقاء قدرت پیش‌بینی شبکه‌های عصبی مصنوعی شده است.

کلمات کلیدی: طراحی الگوریتم، بازاریابی مستقیم، پاسخ مشتری، بهینه‌سازی پیش‌بینی.

۱- مقدمه

تصمیم‌گیری برای تماس با مشتریان بالقوه شرکت، یک وظیفه مهم در برنامه‌ریزی کمپین‌های بازاریابی است، چرا که پیام‌های بازاریابی اگر به دست مشتریان مناسب برسند، بسیار اثربخش خواهند بود. بازاریابی مستقیم به صورت‌های مختلف و با استفاده از فناوری‌های نوین دیجیتالی و تلفن همراه فرصت‌ها بی‌نظیری را در حوزهٔ بازاریابی در اختیار بازاریابان قرار داده است [۱].

امروزه فعالیت‌های مختلف بازاریابی توسط رسانه‌های مختلف تبلیغاتی، مستقیماً با مصرف‌کنندگان ارتباط برقرار کرده و عمدتاً به دنبال مصرف‌کنندگانی می‌باشد که به احتمال زیاد پاسخ مستقیمی به فعالیت‌های بازاریابی خواهند داد [۲]. استفاده تعاملی از رسانه‌های تبلیغاتی در واقع برای تحریک تغییر رفتار صورت می‌گیرد، به طوری که این رفتار را می‌توان ردیابی، تجزیه و تحلیل و ذخیره نموده و در یک پایگاه داده برای بازیابی آینده مورد استفاده قرار داد که معمولاً برای ایجاد روابط پایدار با مشتری بکار می‌روند [۳]. 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 و از این جهت به ایجاد و حفظ روابط مستمر با مشتریان فعلی تأکید فراوان دارند [۴].

مدل‌های پیش‌بینی داده‌محور به طور گسترده‌ای برای پیش‌بینی پاسخ مشتری به کمپین‌های بازاریابی مورد استفاده قرار گرفته است [۵]. مدل‌سازی پاسخ یکی از موثرترین ابزارها برای شرکت‌هایی است که به دنبال برقراری روابط درازمدت با مشتریان خود هستند. هدف از مدل‌سازی پاسخ برای بازاریابی مستقیم این است که بر اساس تاریخچه رفتار مشتری و سایر اطلاعات در دسترس، مشتریانی را که احتمالاً یک محصول رقابتی را خریداری می‌کنند شناسایی کنیم [۶].

در این راستا یک مسئله نهادینه شده در داده‌های مشتریان، مشکل عدم توازن کلاس است [۷]: نسبت پاسخ دهندگان نسبت به افراد غیر پاسخ دهنده بسیار کوچک است. تعداد مشتریانی که به پیام‌های مستقیم ارسال شده به آنها پاسخ می‌دهند در مقابل مشتریانی که پاسخ نمی‌دهند بسیار اندک می‌باشد. بنابراین خوشه‌بندی مشتریان قبل از تهیه و تنظیم استراتژی‌های بازاریابی مستقیم زیاد معقول و منطقی به نظر نمی‌رسد،

*nematizadeh51@yahoo.com

چرا که ممکن است تعداد کم مشتریان پاسخ‌دهنده به پیام‌های تبلیغاتی را از طریق حذف مشتریانی که در سایر خوشه‌های انتخاب نشده قرار دارند، بار دیگر کاهش دهد [۸].

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

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 استخراج بهترین کدهای XML به منظور پیش‌بینی پاسخ مشتریان می‌باشد.

۲- مفاهیم نظری و پیشینهٔ عملی پژوهش

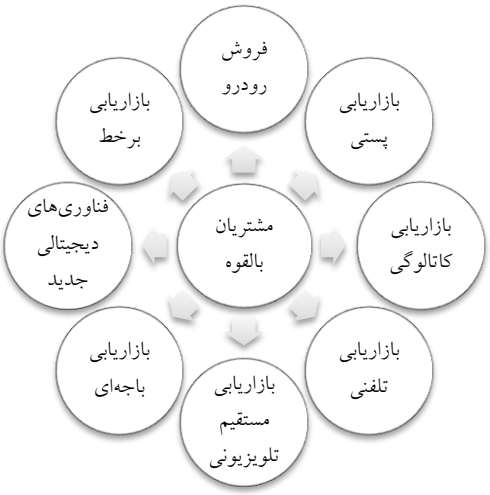
۲-۱- طبیعت در حال تغییر ارتباط با مشتری

شرکت‌ها با توجه به مشتریان هدف خود می‌توانند سطوح مختلفی از ارتباط با مشتری را ایجاد کنند. در سطوح پائین ارتباط که شرکت دارای مشتریان بسیار زیادی با حاشیه سود کم است، برقراری ارتباط در سطح پایه است [۹]. مثلاً شرکت پروکتر اند گمبل[1]، برای دریافت نظرات مصرف‌کنندگان تاید با تک تک مشتریان تماس تلفنی برقرار نمی‌کند. در طرف دیگر طیف، شرکت‌هایی قرار دارند که تعداد کمی مشتری با حاشیه سود بالا دارند. این شرکت‌ها علاقمند به ایجاد رابطه‌ای کامل با تک تک مشتریان هستند. در بین این دو حد، سطوح دیگری از ارتباط با مشتریان وجود دارد [۴]. امروزه روش‌های ارتباط با مشتریان در حال تجربهٔ تغییرات چشمگیری است. در گذشته شرکت‌ها با تمرکز بر بازاریابی انبوه، هر آنچه در توان داشتند را صرف معرفی خود و محصولات خود می‌کردند، اما امروزه شرکت‌ها پس از انتخاب دقیق مشتریان مورد نظر تلاش می‌کنند تا ارتباط بی‌واسطه و بادوام‌تری با آنان برقرار کنند، که در ادبیات بازاریابی تحت عنوان بازاریابی مستقیم مطرح شده است [۱۰].

۲-۲- بازاریابی مستقیم به عنوان یکی از ابزارهای نوین در کسب و کار امروز

بازاریابی مستقیم نوعی از بازاریابی است که در آن مستقیماً با مشتری مشکوک[2] ارتباط برقرار می‌شود. مشتری مشکوک به کسی گفته می‌شود که امکان فروش خدمات و یا محصول مورد نظر به وی وجود دارد [۱۱]. بازاریابی مستقیم به بازاریاب اجازه می‌دهد که پاسخ‌های مستقیم بیشتری را از مشتری دریافت کرده، بازار هدف را به گونهٔ بهتری نشانه‌گیری کند و محصول را بدون قرار گرفتن در فرایند عریض و طویل و پرهزینه کانالهای سنتی توزیع، به فروش رساند [۱۲]. بازاریابی مستقیم یکی از اجزای پنجگانه ابزارهای ترویج[3] است که به همراه تبلیغات[4]، پیشبرد فروش[5]، روابط عمومی[6] و فروش شخصی[7]، به منظور برقراری ارتباط با مخاطبان هدف مورد استفاده قرار می‌گیرد [۱۳]. شرکت‌ها و سازمان‌های بی‌شماری از کوچکترین خرده فروش‌ها و سازمان‌های مدنی و خیریه گرفته، تا بزرگترین کمپانی‌های بین‌المللی از امکانات بازاریابی مستقیم برای تبلیغ محصولات و خدمات خود استفاده می‌کنند [۱۴]. در این روش، افرادی که بازاریابی می‌شوند مستقیماً از بازار هدف انتخاب می‌شوند. نقطه مقابل بازاریابی مستقیم، بازاریابی غیرمستقیم نام گرفته‌است. در این روش از بازاریابی، به صورت غیرمستقیم سعی در جذب مشتری می‌شود؛ مانند تبلیغات بیلبوردی که در فضاهای عمومی انجام می‌شود [۱۵].

امروزه بازاریاب‌های فعال در حوزهٔ بازاریابی مستقیم، با استفاده از قدرت رایانه و داده‌های بسیار زیاد، می‌توانند از بین مشتریان خود بهترین مشتریان احتمالی را شناسایی کرده، و در این راه از روش تحلیل هزینه و منفعت[8]، استفاده کنند [۱۶]. 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 [۱۷]. اشکال عمده بازاریابی مستقیم در شکل (۱) نشان داده شده است.



شکل (۱): اشکال عمدهٔ بازاریابی مستقیم

منبع: [۱۳]

۲-۳- اهمیت داده‌کاوی و کاربرد آن در بازاریابی نوین

تجزیه و تحلیل عظیم‌داده‌ها[9] انقلابی در چهرهٔ پشتیبانی تصمیم‌گیری[10]

1- Procter & Gamble
2- Suspicious customer

3- Promotion
4- Advertising
5- Promotion
6- Publicity
7- Personal Sales
8- Cost-benefit Analysis
9- Big Data
10- Decision Support

بوجود آورده است [۷]. شک و تردید نسبت به تصمیم‌گیری‌های رسمی در میان مدیران شایع و همه‌گیر شده است [۱۸]، با این وجود امروزه ما شاهد علاقهٔ بی‌سابقه‌ای به مدل‌های حمایت از تصمیم‌گیری کمی هستیم. مقادیر بسیار زیادی از داده‌های در دسترس، الگوریتم‌های قدرتمند استخراج الگوها، و استفادهٔ آسان از سیستم‌های نرم‌افزاری، روند این توسعه را تحرک کرده و بهبود در تصمیم‌گیری مدیران را وعده می‌دهد. به عنوان مثال، بر اساس یک تحقیق در میان مدیران اجرایی شرکت‌ها چنین تخمین زده شده است که افزایش سیستم تحلیل بازاریابی به طور متوسط با ۸٪ افزایش در بازده دارایی‌ها همراه است [۱۹]. به شیوه‌ای مشابه [۲۰] به این نتیجه رسیده است که استفاده از تکنولوژی‌های عظیم داده با افزایش ۱ تا ۳ درصد در بهره‌وری شرکت‌ها همراه بوده است.

تکنیک و مدل‌های تصمیم‌گیری برای موفقیت در پیاده‌سازی فرایندهای بازاریابی بسیار حیاتی هستند [۲۱]. داده‌کاوی بخش مهمی از فرایند مدیریت ارتباط با مشتری[1] برای تجزیه و تحلیل جریان داده‌های بزرگ و به دست آوردن بینش در رفتار، نیازها و ترجیحات آنان می‌باشد. امروزه داده‌کاوی کاربرد وسیعی در زمینهٔ تقسیم‌بندی، طبقه‌بندی و خوشه‌بندی مشتریان و نیز استفاده از دانش بدست آمده از این فرایند، در راستای بهره‌گیری از ابزارهای بازاریابی مستقیم دارد [۱۱]. چنین دانشی طراحی فرایندهای تجاری مشتری‌محور و همچنین بازاریابی مستقیم و فعالیت‌های خدماتی را با استفاده از ایجاد وفاداری در مشتری و حفظ رقابت در بازارهای مصرف کنندهٔ اشباع شده، تسهیل می‌کند [۲۲].

بر اساس تحقیقات گذشته، شبکه‌های عصبی کارایی بسیار بالایی برای تخمین[2] و تقریب[3] از خود نشان داده‌اند. شبکه‌های عصبی مصنوعی[4] نوعی مدل‌سازی ساده‌انگارانه از دستگاه‌های عصبی واقعی می‌باشد که کاربرد فراوانی در حل مسائل مختلف دارند. حوزهٔ کاربرد این شبکه‌ها آن‌چنان گسترده است که از کاربردهای طبقه‌بندی گرفته تا کاربردهایی نظیر تخمین، آشکارسازی و ... را شامل می‌شود. شواهد عملی نشان می‌دهند که شبکه‌های عصبی مصنوعی عموماً عملکرد بهتری نسبت به مدل‌های سری‌های زمانی کلاسیک و مدل‌های رگرسیون چند متغیره در پیش‌بینی دارند [۲۳].

۲-۴- *پیشینه عملی*

همان‌گونه که در پژوهش‌های گذشته به کرات اشاره شده، از آنجائیکه متوسط احتمال خرید مشتری بسیار کم است تبلیغات انبوه برای شرکت‌ها بسیار گران تمام می‌شود. برخلاف بازاریابی انبوه، مدل پاسخ تلاش برای جذب مشتریانی با احتمال خرید نسبتاً بالا را پیشنهاد می‌کند. بنابراین، هزینه‌ای که به منظور در معرض پیام قراردادن مشتریانی که علاقه کمتری به خرید محصول دارند صرفه‌جویی خواهد شد و در نهایت، همراه با درآمد افزایش یافته و هزینه کاهش یافته، سود خالص شرکت افزایش می‌یابد [۲۴]. مطالعات گذشته نشان داده است که در حین اینکه افزایش میزان پاسخ کار آسانی نمی‌باشد، تاثیر آن کاملاً باور نکردنی است. به عنوان مثال، [۲۵] به این نکته اشاره‌کرده‌اند که حتی بهبود اندکی در میزان پاسخ می‌تواند نتیجه کلی یک کمپین پستی مستقیم را از شکست به موفقیت تغییر دهد. [۲۶] نشان دادند که چگونه بهبود کوچکی در میزان پاسخ می‌تواند منجر به ایجاد سودهای اضافی عظیم شود. در نمونهٔ آنها، تنها ۱٪ افزایش در نرخ پاسخگویی برای یک شرکت سفارش پستی، ۵۰۰۰۰۰ یورو سود اضافی به همراه داشت. [۲۷] گزارش دادند که برای یک بانک خرده فروشی تنها ۰/۷٪ افزایش در نرخ پاسخ‌دهی درآمد کل را سه برابر کرد و درآمد را به‌ازای هر پاسخگو تا ۲۰٪ افزایش داد. [۶] اشاره کردند که بهبود میزان پاسخ نه تنها می‌تواند سود را افزایش دهد، بلکه موجب تقویت وفاداری مشتری نیز می‌شود، زیرا مشتریانی که به طور صحیح مورد هدف قرار گرفته‌اند به احتمال بیشتری راضی می‌شوند و در بلند مدت به همراه شرکت باقی می‌مانند. با تشویق اثر مثبت قابل ملاحظه‌ای که از موفقیت مدل‌های پاسخ بدست آمده است، تعداد زیادی از مطالعات با هدف افزایش میزان پاسخ از طریق بهبود الگوریتم‌های پیش‌بینی استفاده شده در مدل‌سازی پاسخ انجام شده است.

روحانی و همکاران (۱۳۹۷)، در مقاله‌ای با عنوان «پیش‌بینی رفتار مشتری در صنعت خرده‌فروشی کالاهای تند مصرف» به پیش‌بینی رفتار مشتری در صنعت خرده‌فروشی کالاهای تند مصرف[5] با هدف افزایش کمّی و کیفی فروش در شرکت گلپخش اول پرداخته شده است. این تحقیق از لحاظ هدف در دسته پژوهش‌های کاربردی، و از نظر نحوه گردآوری داده‌ها، بصورت پیمایشی طرح‌ریزی شده است. مراحل اجرای تحقیق مبتنی بر فرایند CRISP-DM[6] است که از مدل خوشه‌بندی RFMCL[7] و تکنیک‌های دسته‌بندی و پیش‌بینی رگرسیونی استفاده کرده و در نهایت از روش پیشنهاددهی مشارکتی بهره برده است. یافته‌های تحقیق حاکی از آن است که با پیاده‌سازی این مدل در شرکت، برنامه‌ریزی تولید پیش‌فعالانه شده و فرایند فروش از ویزیتوری به «فروش آنی» تغییر می‌یابد که این تغییر، صرفه‌جویی شایان توجهی در حمل و نقل و هزینه‌های پرسنلی فروش به ارمغان خواهد آورد.

۳- روش‌شناسی پژوهش

۳-۱- *نمونهٔ اولیه پژوهش و آماده‌سازی داده‌ها*

نمونه مورد بررسی حاوی اطلاعات ۱۰۰۰۰ نفر از مشتریان یک آژانس مسافرتی است. از بین این مشتریان، افرادی به عنوان نمونه انتخاب و یک کمپین تبلیغاتی را دریافت کرده‌اند که تعداد آنها ۱۹۱۷ نفر می‌باشد. از بین ۱۹۱۷ نفر، ۹۷ نفر به کمپین دریافت شده پاسخ داده و ۱۸۲۰ نفر هیچ پاسخی ارائه نکرده‌اند. هدف ما پیش‌بینی پاسخ سایر مشتریان (۸۰۸۳ نفر) و در واقع شناسایی مشتریان بالقوه در پایگاه دادهٔ موجود می‌باشد.

۳-۲- *مدل‌سازی پاسخ مشتریان و مواجهه با عدم توازن داده‌ها*

مدل‌سازی پاسخ (واکنش) مشتریان یکی از موثرترین ابزارها برای

1- Customer Relationship Management
2- Estimation
3- Approximation
4- Artificial Neural Network
5- Fast-Moving Consumer Goods (FMCG)
6- Cross-Industry Standard Process For Data Mining
7- Recency, Frequency, Monetary Value Clustering

شرکت‌هایی است که به دنبال برقراری روابط درازمدت با مشتریان خود هستند [۶]. هدف از مدل‌سازی پاسخ برای بازاریابی مستقیم این است که بر اساس تاریخچهٔ رفتار مشتری و سایر اطلاعات در دسترس، مشتریانی که احتمالاً یک محصول رقابتی را خریداری می‌کنند شناسایی کنیم. بر خلاف استراتژی بازاریابی انبوه[1]، مدل‌های واکنش به خوبی توسعه یافته که برای هدف‌گیری مشتریان خاص مورد استفاده قرار گرفته‌اند، می‌توانند سود شرکت‌ها را نه تنها از طریق افزایش درآمد، بلکه بوسیله کاهش هزینه‌های بازاریابی افزایش دهند.

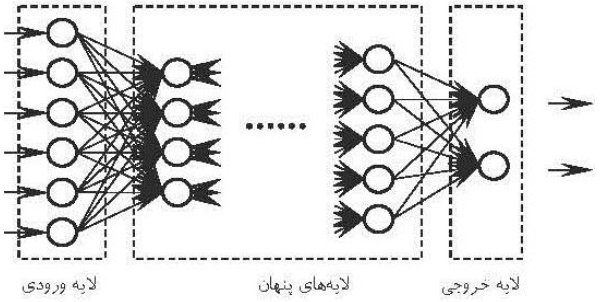
در این راستا یک مسئله نهادینه شده در داده‌های مشتری که تأثیر سوء در مدل‌سازی پاسخ مشتری می‌گذارد، مشکل عدم توازن کلاس است [۲۸]: نسبت پاسخ دهندگان نسبت به افراد غیر پاسخ دهنده بسیار کوچک است. در حین طبقه بندی، عدم تعادل کلاس زمانی اتفاق می‌افتد که بروز یک کلاس به شدت از تعداد سایر کلاس‌ها فراتر رود [۲۹]. عدم تعادل کلاس معمولاً کارایی الگوریتم‌های طبقه‌بندی را کاهش می‌دهد. اکثر الگوریتم‌های طبقه‌بندی، برای تولید مدل‌های پایداری که یک طبقه‌بندی بدون بایاس را ارائه دهند نیازمند به نمونه‌های کافی از همه کلاس‌ها هستند [۳۰]. اگر تعداد اعضاء یک کلاس به میزان قابل توجهی از کلاس‌های دیگر فراتر رود، نتایج طبقه بندی به سمت طبقه اکثریت متمایل می‌شوند. در خصوص پایگاه‌های داده مشتری که برای مدل‌سازی پاسخ استفاده می‌شود، معمولاً تعداد افرادی که پاسخ نداده‌اند به طور فاحشی از پاسخ‌دهندگان بیشتر است.

روش‌های نمونه‌گیری مجدد از مجموعه داده نامتوازن اصلی یکی از متداول‌ترین روش‌ها برای حل مساله عدم توازن به شمار می‌آید. بنابراین در این پژوهش پس از بررسی داده‌های اولیه و ملاحظهٔ ناتوانی الگوریتم‌های پیش‌بینی برای این داده‌ها از روش‌های مختلف متوازن سازی داده‌ها به منظور تقویت قدرت پیش‌بینی شبکه عصبی مصنوعی برای مشتریان استفاده کرده و در نهایت با بررسی راه حل‌های پیشنهادی در این زمینه، اقدام به ارایهٔ الگوریتمی کارا در راستای پیش‌بینی پاسخ مشتریان بوسیله شبکه‌های عصبی مصنوعی خواهیم نموده.

۳-۳- *ابزار تجزیه و تحلیل داده‌ها*

شبکه‌های عصبی مصنوعی مدل‌هایی ریاضی می‌باشند که الهام گرفته از سیستم عصبی و مغز انسان می‌باشند. برای تجزیه و تحلیل داده‌ها در این پژوهش از ساختار شبکه عصبی استفاده شده است که معمولاً یک شبکه یا گراف چند لایه با ارتباطات بین لایه‌ها است. در هر لایه یک یا چندین واحد محاسباتی به نام گره یا نرون مصنوعی وجود دارد که در حقیقت الگویی ساده از نرونهای عصبی مغز انسان هستند. نقش نرون‌ها در شبکه عصبی، پردازش اطلاعات است و این امر در شبکه‌های عصبی مصنوعی به وسیله یک پردازشگر ریاضی که همان تابع فعال‌سازی است، انجام می‌شود. همانگونه که در شکل (۲) نشان داده شده است، شبکه عصبی توسط الگوی ارتباطی بین لایه‌های مختلف شبکه، تعداد نرون‌ها، تعداد لایه‌ها، الگوریتم یادگیری و تابع عملیاتی نرون، شناسایی و تعریف می‌شود.

1- Mass Marketing



شکل (۲): ساختار کلی شبکه‌های عصبی چند لایه

شبکه عصبی یاد می‌گیرد که مسئله‌ای را حل کند و در واقع برنامه‌ریزی قبلی نمی‌شود. در جریان اصلاح مکرر وزن‌ها، یک شبکه آموزش می‌بیند. با تکرار فرایند یادگیری، شبکه مقادیر صحیح وزن‌ها را شناسایی کرده و خطا را کاهش می‌دهد.

۴-۳- *معیارهای مورد استفاده جهت ارزیابی مدل‌ها*

معیارهای ارزیابی نقشی اساسی در مسائل طبقه‌بندی ایفا می‌کنند. برای ارزیابی کارایی خروجی شبکه عصبی به عنوان یک طبقه‌بند[2]، ساده‌ترین روش تحلیل براساس ماتریس اغتشاش[3] (ماتریس درهم ریختگی) است. پیکرهٔ اصلی ماتریس درهم ریختگی یا ماتریس اغتشاش در جدول (۱) ارایه شده است؛

جدول (۱): ماتریس درهم ریختگی (اغتشاش)

کلاس‌های واقعی	کلاس‌های پیش‌بینی شده	
	غیر پاسخ‌گو	پاسخ‌گو
غیر پاسخ‌گو	TN	FP
پاسخ‌گو	FN	TP

رایج‌ترین معیار مستخرج از این ماتریس، معیار دقت[4] 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 ارزیابی استفاده شده در پژوهش حاضر به شرح زیر می‌باشد:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (۱)$$

$$Sensivity = \frac{TP}{TP + FN} \quad (۲)$$

$$Specificity = \frac{TN}{TN + FP} \quad (۳)$$

$$Precision = \frac{TP}{TP + FP} \quad (۴)$$

$$G - Mean = \sqrt{Precision.Recall} \quad (۵)$$

$$F - Measure = \frac{2.Recall.Precision}{Recall + Precision} \quad (۶)$$

از معیار AUC[5] نیز که نشان دهنده سطح زیر منحنی مشخصهٔ عملکرد سیستم (ROC[6]) می‌باشد استفاده شده است. هر چه مقدار این عدد

2- Classifier
3- Confusion Matrix
4- Accuracy
5- Area Under Curve (AUC)
6- Receiver Operating Characteristic (ROC)

مربوط به یک طبقه‌بند بزرگتر (نزدیک به یک) باشد کارایی نهایی مطلوب‌تر ارزیابی می‌شود.

در نهایت معیار حاصل‌ضرب[1] برابر با ضرب هفت معیار معرفی شده فوق می‌باشد ما برای اولین بار و بصورت کاملاً ابتکاری در این مقاله به معرفی آن پرداخته‌ایم؛

$$\text{Multiple} = \text{Accuracy} \times \text{Sensitivity} \times \text{Specificity} \times \text{Precision} \times \text{G-mean} \times \text{F-measure} \times \text{AUC} \quad (۷)$$

از آنجا که این معیار حاصل‌ضرب معیارهای معروف مورد استفاده فوق می‌باشد می‌تواند به عنوان ملاک و معیاری مفید جهت جمع‌بندی کفایت مدل مد نظر قرار گیرد.

۳-۵- روش پیاده‌سازی و جنبهٔ نوآوری پژوهش

در این پژوهش ابتدا مشتریان را در گروه‌های مختلفی خوشه‌بندی کردیم. پس از بررسی نتایج (همان‌گونه که در جدول (۲) مشاهده می‌شود) مشخص شد که مشتریان پاسخ‌گو (۹۷ نفر) در تمامی خوشه‌های شناسایی شده، به نسبت حجم آنها قرار گرفته‌اند. بنابراین خوشه‌بندی مشتریان ابزار مناسبی برای شناسایی مشتریان محتمل نخواهد بود.

از طرفی استفاده از شبکه‌های عصبی مصنوعی با داده‌های اولیهٔ در دسترس، در پیش بینی مشتریان محتمل کاملاً ناتوان بوده‌اند. بطوریکه عدم تعادل کلاس مشتریان پاسخ‌گو و غیرپاسخ‌گو باعث شده بود تا خروجی مدل، تمامی مشتریان را به عنوان مشتریان غیرپاسخ‌گو شناسایی کند.

از طرفی این ساده لوحانه خواهد بود که مشتریان را بر اساس پاسخ دادن یا ندادن به کمپین ارائه شده طبقه‌بندی کنیم. زیرا مشتریان غیر پاسخ‌گو هم قطعاً از خصوصیات و ویژگی‌های متمایز کننده‌ای برخوردارند. جنبهٔ نوآوری پژوهش حاضر بر این خواهد بود که پس از بررسی ضعف مدل شبکهٔ عصبی مصنوعی در پیش‌بینی پاسخ مشتریان با داده‌های اولیه، به منظور ارتقاء توان پیش‌بینی پاسخ مشتریان، پس از خوشه‌بندی مشتریان غیر پاسخ‌گو اقدام به تشکیل زیرمجموعه‌هایی از خوشه‌ها و بخش‌های مختلف شناسایی شده با مشتریان پاسخ‌گو نموده و در نهایت، اقدام به ارائه الگویی مؤثر جهت شناسایی مشتریان بالقوه و پیاده‌سازی ابزارهای بازاریابی مستقیم خواهیم نمود.

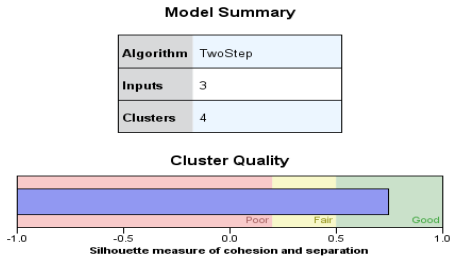
در این مرحله مشخص گردید که الگوریتم خوشه‌بندی به روش دو مرحله‌ای توانسته است با سه متغیر جنسیت، وضعیت ازدواج و سن، بهترین خوشه‌بندی را برای نمونهٔ مورد بررسی ارائه دهد. شکل (۳) نشان دهنده معیار سیلوئت[2] (ضریب نمیرخ) در این مرحله می‌باشد:

شایان ذکر است؛

- خوشهٔ اول شامل ۴۸۰ نفر و معادل ۲۶/۴٪ مشتریان غیر پاسخ‌گو می‌باشد. آنها عمدتاً زنان مجرد با میانگین سن ۴۰ می‌باشند.

1- Multiple
2- Silhouette Criteria

- خوشه دوم شامل ۴۱۲ نفر و معادل ۲۲/۶٪ مشتریان غیر پاسخ‌گو می‌باشد. آنها زنان متأهل با میانگین سن ۴۸ هستند.
- خوشه سوم شامل ۴۷۴ نفر و معادل ۲۶٪ مشتریان غیر پاسخ‌گو می‌باشد. آنها مردان مجرد با میانگین سن ۳۷ هستند.
- خوشه چهارم شامل ۴۵۴ نفر و معادل ۲۴/۹٪ مشتریان غیر پاسخ‌گو می‌باشد. آنها مردان متأهل با میانگین سن ۵۰ هستند.



شکل (۳): معیار سیلوئت برای خوشه‌بندی مشتریان غیرپاسخ‌گو

در این مرحله مشتریان غیر پاسخ‌گو (۱۸۲۰ نفر) را با مشتریانی که پاسخ داده‌اند (۹۷ نفر) به چند شیوه ترکیب می‌نماییم. یادآوری می‌کنیم که خوشه‌بندی مشتریان غیر پاسخ‌گو به 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 چند راه حل ابتکاری به شرح زیر تدوین و طراحی شده است؛

NS[3]؛ پیش‌بینی مدل بدون در نظر گرفتن هر گونه روش نمونه‌برداری مجدد به عمل خواهد آمد.

RUS[4]؛ استفاده از روش کم نمونه‌گیری تصادفی کلاس اکثریت.

C-RUS-01؛ تعداد مشتریان پاسخ‌گو را تقسیم بر تعداد خوشه‌ها نموده و در واقع عدد حاصل تعداد مشتریان انتخاب شده از هر خوشه از مشتریان غیر پاسخ‌گو می‌باشد.

C-ROS-RUS-02؛ ابتدا مشتریان پاسخ‌گو را ضربدر تعداد خوشه‌های شناسایی شده از مشتریان غیر پاسخ‌گو نماییم (در اینجا ۳۸۸=۴×۹۷) 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، بدین ترتیب از خوشهٔ اول ۱۰۲ نفر، از خوشهٔ دوم ۸۸ نفر، از خوشهٔ سوم ۱۰۱ نفر و از خوشهٔ چهارم ۹۷ نفر (جمعاً ۳۸۸ نفر) به عنوان نمونه انتخاب خواهند شد.

- C-RUS-03؛ 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.

3- No-sampling
4- Random under-sampling

جدول (۲): توزیع مشتریان پاسخ‌گو در خوشه‌های چهارگانه شناسایی شده از کل مشتریان

خوشه‌های شناسایی شده		تعداد کل مشتریان	نسبت اعضاء خوشه به کل مشتریان	مشتریان در پاسخ به آزمون (کمپین) ارائه شده		نسبت پاسخ‌دهندگان به غیر پاسخ‌دهندگان در هر خوشه	نسبت پاسخ‌دهندگان به کل مشتریان	نسبت پاسخ‌دهندگان هر خوشه به کل مشتریان پاسخگو
				غیر پاسخ‌دهندگان	پاسخ‌دهندگان			
گروه خوشه‌ها	۱	٤۰۰۰	٪٤۰	۷۰۵	۳٦	٪۵/۱	٪۰/۹	٪۳۷/۱
	۲	۲٤۲٤	٪۲٤	٤۵٦	۳۵	٪۷/٦	٪۱/٤	٪۳٦/۱
	۳	۱۹۰۹	٪۱۹/۱	۳٤٤	۱٦	٪٤/٦	٪۰/۸	٪۱٦/۵
	٤	۱٦٦۷	٪۱٦/۷	۳۱۵	۱۰	٪۳/۱	٪۰/٦	٪۱۰/۳
جمع		۱۰۰۰۰	٪۱۰۰	۱۸۲۰	۹۷			٪۱۰۰

جدول (۳): نمونه‌ها و مدل‌های مورد استفاده جهت پیش‌بینی پاسخ مشتریان

ردیف	عنوان روش	اختصار استفاده شده	ترکیب مشتریان					حجم نمونه
			C4	C3	C2	C1	R	
۱	بدون نمونه‌گیری مجدد	NS	1820				97	1917
۲	کم نمونه‌گیری تصادفی کلاس اکثریت	RUS	97				97	194
۳	روش ابتکاری اول	C-RUS-01	24	24	24	24	97	193
٤	روش ابتکاری دوم	C-ROS-RUS-02	97	101	88	102	388	776
۵	روش ابتکاری سوم	C-RUS-03	24	25	22	26	97	194
٦	بیش نمونه‌گیری تصادفی کلاس اقلیت	ROS	1820				1820	3640
۷	روش ابتکاری چهارم	C-ROS-RUS-04	412	412	412	412	1648	3296

- ROS[1]؛ استفاده از روش بیش نمونه‌گیری تصادفی کلاس اقلیت.
- C-ROS-RUS-04؛ 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.

نتایج حاصل از نمونه‌برداری به روش‌های NS، RUS و ROS و راه حل‌های ابتکاری ترکیبی آنها در جدول (۳) نشان داده شده است.

در نهایت پس از پیاده‌سازی مدل‌های پیش‌بینی با توجه به هر یک از نمونه‌های فوق، اقدام به محاسبه معیار ضریب و مشتریان بالقوه با توجه به کدهای XML برای سایر مشتریان (۸۰۸۳ نفر) که اطلاعاتی از پاسخ آنها نداریم نموده و به منظور تحقق هرچه بهتر این هدف، الگوریتمی را به شیوهٔ کاملاً ابتکاری طراحی و در قسمت «پیشنهادات کاربردی» پژوهش معرفی خواهیم نمود.

٤- تجزیه و تحلیل داده‌ها

نتایج حاصل از پیش‌بینی شبکه عصبی مصنوعی با توجه به هفت نمونهٔ معرفی شده بصورت خلاصه به شرح جدول (۴) می‌باشد. اعداد ارائه شده در جدول مرتبط با معیارهای ارزیابی هر مدل و نمونهٔ مرتبط با آن می‌باشد. هدف اصلی این تحقیق تهیهٔ نمونه‌های مناسب و طراحی الگوریتمی کارا به منظور استفاده بهینه از مدل‌های پیش‌بینی و نیل به نرخ بالایی از پاسخ مشتری می‌باشد.

با توجه به نتایج معیارهای ارزیابی در جدول (۴)، می‌توان فهمید، نمونهٔ اولیه (NS) یا به عبارتی دیگر، استفاده از مدل شبکهٔ عصبی بدون نمونه برداری مجدد به هیچ وجه توانایی پیش‌بینی مشتریان بالقوه را نخواهد داشت. معیارهای ارزیابی و معیار ضریب صفر حاکی از عدم توانایی مدل در این مرحله داشته و خروجی XML تمامی مشتریان (۸۰۸۳ نفر) را به عنوان مشتریان غیر پاسخ‌گو شناخته است! این در حالی است که سایر روش‌های نمونه‌گیری مجدد نتایج شگفت انگیز و خروجی‌های XML بهینه‌ای در اختیار قرار داده است. نمونه‌ای از خروجی‌های موصوف در پیوست (۱) ارائه شده است.

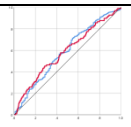
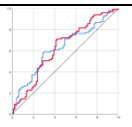
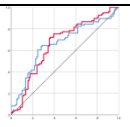
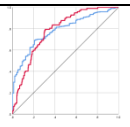
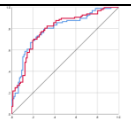
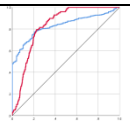
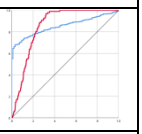
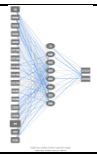
۵- نتیجه‌گیری

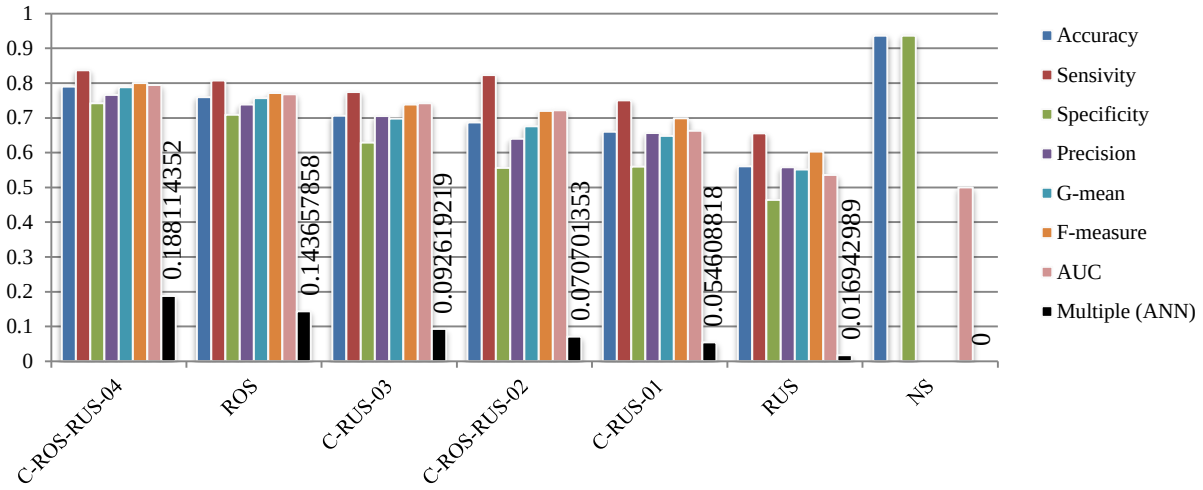
۵-۱- نتایج پژوهش

با توجه به نمودار میله‌ای ارائه شده در شکل (۴) می‌توان نتیجه گرفت، در بین روش‌های مورد استفاده برای نمونه‌گیری مجدد داده‌ها روش RUS ضعیف‌ترین پیش‌بینی و روش C-ROS-RUS-04 قوی‌ترین پیش‌بینی را از آن خود کرده‌اند. با اندکی مداقه در شیوه‌های نمونه‌گیری این دو روش و سایر روش‌های ابتکاری می‌توانیم به نتایج درخور توجی دست پیدا کنیم. اول اینکه روش RUS با ساده‌ترین شیوه توانسته از عدم توازن کلاس را به راحتی از بین برده و معیار ضریب را از صفر به ۰/۰۱ ارتقاء دهد. مهمتر آنکه تفاوت معیار ضریب برای RUS و C-RUS-01 حدود پنج برابر است. این درحالی است حجم هر دو نمونه برابر با ۱۹۴ نفر بوده و هر دوی این نمونه‌ها به روش کم نمونه‌گیری تصادفی کلاس اکثریت عدم توازن را از بین برده‌اند. ولی آنچه که شایان ذکر است توجه به خوشه‌های مختلف مشتریان در روش C-RUS-01 می‌باشد.

1- Random over-sampling

جدول (٤): نتایج حاصل از معیارهای ارزیابی نمونه‌های مختلف اتخاذ شده و مدل‌های مختلف پیش‌بینی

معیارهای ارزیابی مدل		NS	RUS	C-RUS-01	C-ROS-RUS-02	C-RUS-03	ROS	C-ROS-RUS-04
Accuracy		0.936	0.561	0.660	0.687	0.706	0.759	0.790
Sensitivity		0	0.655	0.750	0.823	0.774	0.807	0.837
Specificity		0.936	0.464	0.560	0.556	0.629	0.709	0.742
Precision		0	0.558	0.656	0.640	0.705	0.738	0.766
G-mean		0	0.551	0.648	0.676	0.698	0.757	0.788
F-measure		0	0.603	0.699	0.720	0.738	0.771	0.800
ROC	AUC	0.500	0.536	0.663	0.722	0.742	0.768	0.794
	Std. Error	0.037	0.041	0.039	0.019	0.036	0.008	0.008
	Sig	1.000	0.385	0.000	0.000	0.000	0.000	0.000
	Curve							
Multiple		0	0.016	0.054	0.070	0.092	0.143	0.188
Artificial Neural Network Diagram With One Hidden Layer								
Hidden Nodes		7	4	4	5	8	6	11



شکل (٤): معیار ضریب برای نمونه‌های مختلف پژوهش

در نمونهٔ C-RUS-01 تعداد ۹۷ نفر از مشتریان غیر پاسخ‌گو به شیوه‌ای انتخاب شده‌اند که از هر خوشه تعداد ۲۴ عضو شرکت داشته باشند در صورتی که در نمونهٔ RUS این کار بطور تصادفی انجام شده و به عضویت اعضاء در خوشه‌های مختلف توجهی نشده است. همین عامل باعث کاهش معیار ضریب تا پنج برابر شده است! بنابراین تهیهٔ ترکیبی از مشتریان غیرپاسخ‌گو که نمایندهٔ تمام خصوصیات نمونهٔ مورد بررسی باشد می‌تواند عامل مهمی در ارتقاء توان پیش‌بینی باشد.

نکتهٔ مهم دیگر تفاوت در معیار ضریب برای نمونهٔ C-RUS-01 و C-RUS-03 تا حدود دو برابر است. با توجه به حجم و شیوهٔ نمونه‌گیری در این دو روش مشاهده می‌شود که هر دو حجم برابر داشته و از روش کم نمونه‌گیری تصادفی کلاس اکثریت استفاده نموده‌اند. ولی آنچه که موجب افزایش معیارهای ارزیابی در نمونهٔ C-RUS-03 شده است می‌تواند نشأت گرفته از توجه محقق به حجم خوشه‌ها در نمونه‌های اتخاذ شده باشد. این مسئله بار دیگر اهمیت خوشه‌بندی مشتریان غیرپاسخ‌گو و انتخاب بهینهٔ آنها به منظور کاهش عدم توازن کلاس و نیز تشکیل ترکیبی از مشتریانی که دارای توزیع و ویژگی‌های کل جامعه باشند را بیش از پیش مشخص می‌کند.

به عنوان جمع‌بندی و نکتهٔ آخر می‌توان گفت روش C-ROS-RUS-04 بهترین نتایج پیش‌بینی را از آن خود کرده است و علت آن را می‌توان در استفادهٔ این نمونه بطور همزمان از روش‌های کم نمونه‌گیری کلاس اکثریت و بیش نمونه‌گیری کلاس اقلیت و توجه به خوشه‌های مختلف شناسایی شده از مشتریان غیر پاسخ‌گو دانست. به عنوان نتیجه‌ای دیگر

می‌توان گفت ظواهر امر نشان می‌دهد شبکه‌های عصبی با نمونه‌های بزرگتر پیش‌بینی بهتر و قابل اطمینان‌تری به ارمغان آورده‌اند به شرط آنکه نمونه‌های تهیه شده به خوبی متعادل شده و ترکیب خوبی از مشتریان را در خود جای داده باشند.

۵-۲- *پیشنهادات کاربردی*

از نتایج حاصل از این تحقیق می‌توان به عنوان راهکاری جهت تدوین استراتژی‌های بازاریابی مستقیم و پیشنهاد محصول و خدمات برای هر یک از گروه‌های مشتریان بهره برد.

الگوریتم پیشنهادی (شکل ۵)، نهایتاً منجر به استخراج هفت کد XML (که نمونه‌ای از آن در پیوست شماره ۱ آمده است) و نیز معیارهای ارزیابی مدل و محاسبه معیار «حاصل‌ضرب» در هر مرحله خواهد شد؛ لذا به عنوان اصلی‌ترین و کاربردی‌ترین پیشنهاد پژوهش می‌توان گفت:

امتیاز هر مشتری می‌تواند در تمامی نمونه‌های هفت‌گانه محاسبه و در فیلد جداگانه‌ای وارد شود. به عنوان مثال اگر یک مشتری در تمامی نمونه‌ها به عنوان یک مشتری بالقوه شناخته شود امتیاز وی برابر خواهد بود با :(۰/۵۶۳ = ۰/۱۸۸ + ۰/۱۴۳ + ۰/۰۹۲ + ۰/۰۷ + ۰/۰۵۴ + ۰/۰۱۶).

سپس می‌توان مشتریان را با توجه به امتیاز کسب شده در طبقه‌های متفاوتی قرار داد. با توجه به استراتژی بازاریاب، نوع کالا، و منابع در دسترس می‌توان از ابزارها و ارتباطات مستقیم بازاریابی برای مشتریانی که در طیف‌های مختلف امتیازی قرار دارند استفاده کرد. طبقه‌بندی مشتریان در چهار طبقه می‌تواند به عنوان یک طبقه‌بندی پیشنهادی مدنظر قرار گیرد.

به عنوان پیشنهادات کلی می‌توان بیان نمود مشتریانی که کمتر از ۵۰٪ میانگین امتیازها را کسب کرده‌اند نمی‌توانند به عنوان مشتریان بالقوه در نظر گرفته شود. بنابراین چارک سوم و چهارم از مشتریان بهتر است نادیده گرفته شود.

پیشنهاد می‌شود طرز تعامل با دو چارک اول و دوم نیز متفاوت باشد. مشتریان چارک اول به احتمال زیاد جزو مشتریانی هستند که تمایل بیشتری به خرید دارند. پیشنهاد می‌شود تبلیغات تهاجمی زیادی در خصوص آنها انجام نگیرد. بلکه شرکت ارتباطات محکم خود را پس از خرید و به منظور کاهش هرچه بیشتر تعارض پس از خرید با آنها انجام دهد.

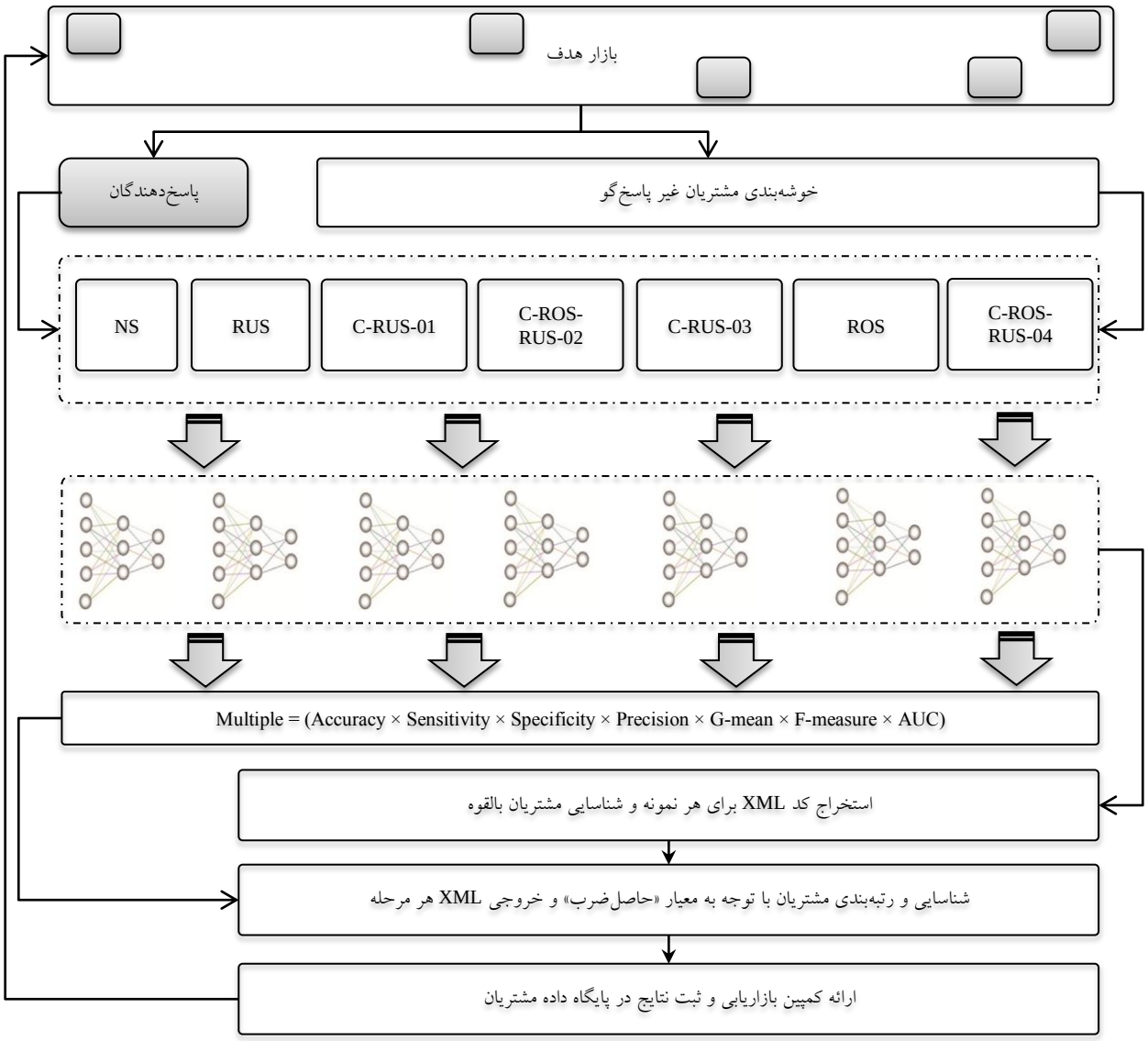
چارک دوم مشتریان محتملی هستند که ارزش هرگونه کمپین بازاریابی را دارند. می‌توان گفت ماحصل پژوهش حاضر مشتریان واقع شده در این چارک می‌باشند زیرا بر خلاق مشتریان بسیار محتمل که خودبه‌خود خریدار کالا یا محصول ما هستند و بر خلاف مشتریانی که به هیچ عنوان تمایل به خرید نشان نخواهند داد، آماده پاسخ‌گویی به کمپین‌های بازاریابی هستند، اگر و فقط به درستی انتخاب شوند.

۵-۳- *محدودیت‌های پژوهش*

محدودیت‌های ذیل به عنوان محدودیت‌های پژوهشی در تحقیق حاضر مطرح شده و می‌توانند به عنوان پیشنهادی برای انجام پژوهش‌های آتی مدنظر قرار گیرند؛

ورود مشتریان جدید و تغییر در داده‌های مشتریان قبلی و قدیمی و به روز نشدن پایگاه داده‌ها از محدودیت‌های این پژوهش می‌باشد. چرا که پیش‌بینی آینده با داده‌های قدیمی و یا متحول شدهٔ مشتریان با اریب بسیار زیادی همراه خواهد بود.

در این پژوهش به جهت جلوگیری از گسترش بیش از حد ابعاد و زوایای تحت بررسی که یکی از ویژگی‌های مهم برای رسیدن به نتایج هدفمند در پژوهش‌های علمی می‌باشد، از ورود متغیرهای اجتماعی – اقتصادی در طی مدت انجام پژوهش خودداری شده است. فلذا به عنوان یکی از محدودیت‌های پژوهش قابل طرح بوده و نیز می‌توان به عنوان پیشنهادی برای پژوهش‌های آتی نیز در نظر گرفته شود.



شکل (۵): فلوچارت الگوریتم پیشنهادی پژوهش

منابع و مأخذ

[۱] روحانی، سعید، غضنفری، حامد، سهرابی، بابک، (۱۳۹۷)، **پیشبینی رفتار مشتری در صنعت خردهفروشی کالاهای تند مصرف (مورد مطالعه: شرکت گلپخش اول)**، فصلنامه مدیریت بازرگانی، (۱۰)۳، (۶۲۳ -۶۴۲).

[1] Andrews, J.C., Shimp, T.A., (2017), **Advertising, Promotion, and other Aspects of Integrated Marketing Communications**, Nelson Education.

[2] Lancaster, G., Massingham, L., (2017), **Direct Marketing, in Essentials of Marketing Management**, Routledge. p. 289-307.

[3] McCarthy, J., McCarthy, E., (2005), **Method, System and Personalized Web Content Manager Responsive to Browser Viewers' Psychological Preferences, Behavioral Responses and Physiological Stress Indicators**, Google Patents.

[4] Pansari, A., Kumar, V., (2017), **Customer Engagement: the Construct, Antecedents, and Consequences,** Journal of the Academy of Marketing Science, 45(3): p. 294-311.

[5] 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): p. 422-434.

[6] Sun, B., Li, S., Zhou, C., (2006), **"Adaptive" Learning and "Proactive" Customer Relationship Management**, Journal of interactive Marketing, 20(3-4): p. 82-96.

[7] Gupta, M., George, J.F., (2016), **Toward the development of a big data analytics capability**, Information & Management,53(8): p. 1049-1064.

[8] Kang, P., Cho, S., MacLachlan, D.L., (2012), **Improved Response Modeling Based on Clustering, Under-sampling, and Ensemble**, Expert Systems with Applications, 39(8): p. 6738-6753.

[9] Cooper, R., (2017), **Target Costing and Value Engineering**, Routledge.

[10] Means, G., (2017), **The Modern Corporation and Private Property**,Routledge.

[11] Linoff, G.S., Berry, M.J., (2011), **Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management**,John Wiley & Sons.

[12] Wilkinson, T.J., McAlister, A., Widmier, S., (2007), **Reaching the International Consumer: An Assessment of the International Direct Marketing environment**, Direct Marketing: An International Journal, 1(1): p. 17-37.

[13] Kotler, P., Armstrong, G., (2010), **Principles of Marketing**, Pearson education.

[14] Thomas, A.R., (2007), **The End of Mass Marketing: or, Why all Successful Marketing is Now Direct Marketing**, Direct Marketing: an International Journal, 1(1): p. 6-16.

[15] Tuten, T.L., (2008), **Advertising 2.0: Social Media Marketing in a Web 2.0 World: Social Media Marketing in a Web 2.0 World**,ABC-CLIO.

[16] Rygielski, C., Wang, J.-C., Yen, D.C., (2002), **Data Mining Techniques for Customer Relationship Management**. Technology in society,24(4): p. 483-502.

[17] Jiang, P., B. 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): p. 150-174.

[18] Lilien, G.L., (2011), **Bridging the Academic–practitioner Divide in Marketing Decision Models**, Journal of Marketing, 75(4): p. 196-210.

[19] Germann, F., Lilien, G.L., Rangaswamy, A., (2013), **Performance Implications of Deploying Marketing Analytics**, International Journal of Research in Marketing,30(2): p. 114-128.

[20] Tambe, P., (2014), **Big Data Investment, Skills, and Firm Value**, Management Science,60(6): p. 1452-1469.

[21] Olson, D.L. Chae, B., (2012), **Direct Marketing Decision Support Through Predictive Customer Response Modeling**, Decision Support Systems,54(1): p. 443-451.

[22] Lessmann, S., VoB, S., (2009), **A Reference Model for Customer-Centric Data Mining with Support Vector Machines**, European Journal of Operational Research,199(2): p. 520-530.

[23] Russell, S.J., Norvig, P., (2016), **Artificial Intelligence: a Modern Approach**,Malaysia; Pearson Education Limited.

[24] Gönül, F.F., Hofstede, F.T., (2006), **How to Compute Optimal Catalog Mailing Decisions**, Marketing Science,25(1): p. 65-74.

[25] Coenen, F., et al., (2000), **The Improvement of Response Modeling: Combining Rule-induction and Case-based Reasoning**, Expert Systems with Applications,18(4): p. 307-313.

[26] Baesens, B., et al., (2002), **Bayesian Neural Network Learning for Repeat Purchase Modelling in Direct Marketing**, European Journal of Operational Research,138(1): p. 191-211.

[27] Knott, A., Hayes, A., Neslin, S.A.,(2002), **Next-product-to-buy Models for Cross-selling Applications**, Journal of interactive Marketing, 16(3): p. 59-75.

[28] Kang, P., Cho, S., (2006), **EUS SVMs: Ensemble of Under-sampled SVMs for data Imbalance Problems**, in Neural Information Processing,Springer.

[29] LóPez, V., et al., (2013), **A Hierarchical Genetic Fuzzy System Based on Genetic Programming for Addressing Classification with Highly Imbalanced and Borderline Data-sets**, Knowledge-Based Systems, 38: p. 85-104.

[30] Williams, D.P., Myers, V., Silvious, M.S., (2009), **Mine Classification with Imbalanced Data**, IEEE Geoscience and Remote Sensing Letters, 6(3): p. 528-532.