

مدیریت تولید و عملیات، دوره هفتم، شماره (۱)، پیاپی (۱۲)، بهار و تابستان ۱۳۹۵

دریافت: ۹۲/۲/۱۴ پذیرش: ۹۳/۴/۹

صص: ۱-۲۲

یک رویکرد جامع برای بخش‌بندی بازار و طبقه‌بندی مشتریان با استفاده از روش‌های داده‌کاوی و برنامه‌ریزی خطی

بهرام ایزدی^۱، بهرام رنجبریان^۲، سعیده کتابی^{۳*}، فریا نصیری مفخم^۴

- ۱- دکترای مدیریت بازرگانی، دانشکده علوم اداری و اقتصاد، دانشگاه اصفهان، اصفهان، ایران
- ۲- استاد مدیریت بازاریابی، دانشکده علوم اداری و اقتصاد، دانشگاه اصفهان، اصفهان، ایران
- ۳- دانشیار تحقیق در عملیات، دانشکده علوم اداری و اقتصاد، دانشگاه اصفهان، اصفهان، ایران
- ۴- استادیار دانشکده مهندسی کامپیوتر، گروه فناوری اطلاعات، دانشگاه اصفهان، اصفهان، ایران

چکیده:

با گسترش فناوری و ظهور شرکت‌های الکترونیکی که انباشت داده‌های مشتریان در پایگاه داده را به همراه داشته، جذابیت بخش‌بندی بازار برای پژوهش‌گران فزونی یافته است. زیرا پایگاه داده مشتری یک داشته ارزشمند شرکت‌های الکترونیکی است که اگر به خوبی پردازش شود، می‌تواند شناخت بیشتری از مشتریان ارائه نماید. به همین دلیل روش‌های متعدد اما غیر منسجمی برای بخش‌بندی بازار و طبقه‌بندی مشتریان در علوم مختلف ارائه شده است. این پژوهش، الگوریتم یکپارچه‌ای برای بخش‌بندی بازار و طبقه‌بندی مشتریان ارائه می‌کند که در آن رویکرد خوشه‌بندی تجمیعی مبنای مقایسه عملکرد روش‌های مختلف طبقه‌بندی، به ویژه برنامه‌ریزی خطی، قرار می‌گیرد. این مقایسه نشان می‌دهد روش برنامه‌ریزی خطی که کمتر به آن توجه شده است، عملکرد بهتری از نظر درصد بالاتر انتساب مشتریان به خوشه‌های مشتریان دارد. همچنین برای مواجهه با ناکافی بودن داده‌های مورد نیاز در پایگاه داده، روش دلفی فازی پیشنهاد شده است.

واژه‌های کلیدی: برنامه‌ریزی خطی، بخش‌بندی بازار، پایگاه داده، دلفی فازی، خوشه‌بندی تجمیعی،

طبقه‌بندی مشتریان

۱- مقدمه

روش‌های مرسوم محدود ننماید. به ویژه آن که انبوهی از دیدگاه‌ها، رویکردها و روش‌های متعدد در این حوزه وجود دارد که هر یک از آن‌ها تنها بر بخشی از فرآیند بخش‌بندی و طبقه‌بندی تاکید می‌کنند (لیو و همکاران^۱، ۲۰۱۲، آزادیا و همکاران، ۲۰۱۱، نامور و همکاران، ۲۰۱۱، ما و همکاران^۲، ۲۰۰۹). به همین دلیل، این پژوهش الگوریتمی یکپارچه برای بخش‌بندی بازار و طبقه‌بندی مشتریان ارائه می‌نماید که متناسب با شرایط واقعی بازارهای الکترونیکی، مانند در اختیار نبودن داده‌های مورد نیاز، است و در آن روش‌های مرسوم بخش‌بندی بازار و طبقه‌بندی مشتریان را از منظر بازاریابی و در ارتباط با هم، در علمی چون آمار و یادگیری ماشینی در کنار روش مغفول مانده برنامه‌ریزی خطی مورد بررسی قرار داده و نشان می‌دهد علیرغم کم توجهی به روش برنامه‌ریزی خطی برای طبقه‌بندی، مزایایی چون تفسیر ساده‌تر و کاربردی‌تر نتایج از آن حاصل می‌شود.

مشوق اولیه برای ارائه چنین ایده‌ای، ماهیت غیرپارامتریک برنامه‌ریزی خطی است. به دلیل همین ویژگی، مسایل طبقه‌بندی از قید نیاز به پیش‌شرط‌ها رها می‌شوند، زیرا اغلب روش‌های آماری ارائه شده برای مسایل طبقه‌بندی نیازمند برآورده شدن برخی پیش‌شرط‌ها، مانند نرمال بودن توزیع داده‌ها، هستند که گاهی فراهم کردن آن‌ها امکان‌پذیر نیست. از زمان معرفی این مدل توسط منگسریان^۳ (۱۹۶۵) تاکنون، شکل‌های مختلفی از آن ارائه شده است. در ابتدا مدل‌های ارائه شده محدود به طبقه‌بندی دو گروهی بودند، اما پس از معرفی مدلی توسط فرید و گلور^۴ (۱۹۸۱)، شکل‌هایی از برنامه‌ریزی خطی ارائه شدند که قادر به حل مسایل طبقه‌بندی چند گروهی شدند (کواک و همکاران، ۲۰۰۲)، (رگس‌دیل و

یک گام اساسی در تدوین و اجرای استراتژی بنگاه‌های تجاری از دیرباز بخش‌بندی مشتریان بوده است، زیرا امکان کسب شناخت بیشتر و بهتر مشتریان را فراهم می‌آورد. روش سنتی بخش‌بندی بازار (روش مقدم) بر اساس گمانه‌زنی‌های ذهنی خبرگان از متغیرهای بخش‌بندی انجام می‌شود. در این روش، پژوهش‌گر می‌کوشد شرایط از پیش موجود در گروه مشتریان را حدس بزند تا مشتریان به گروه‌های متمایز تقسیم شوند و از بین آنها گروهی که تناسب بیشتری با امکانات شرکت دارد و با توجه به شرایط رقابتی بیشترین منافع را عاید شرکت می‌کند، شناسایی و هدف‌گیری شود (زیبرنا و زیکار^۵، ۲۰۰۳؛ باینز و همکاران، ۲۰۱۰). اما امروزه با توجه به داده‌های ارزشمندی که در پایگاه داده مشتریان شرکت‌ها، به ویژه در شرکت‌های الکترونیکی به وجود آمده است، شناسایی مشتریانی که بیشترین درآمد را برای شرکت ایجاد می‌کنند، تسهیل شده است. از این رو در بخش‌بندی نوین یا موخر، نوع و تعداد بخش‌ها بر مبنای تجزیه و تحلیل داده‌های رفتاری مشتری تعیین می‌شوند (فوناتسو و هسگوا^۶، ۲۰۱۱). این بخش‌بندی نسبت به بخش‌بندی سنتی از مزایایی چون امکان شناسایی دقیق‌تر بخش‌های با ارزش برخوردار است که بر پایه آن می‌توان استراتژی‌های مناسب نگهداری مشتریان فعلی و جذب مشتریان جدید را تدوین کرد تا هم از ریزش مشتریان^۷ که یکی از نگرانی‌های عمده شرکت‌ها است جلوگیری به عمل آید و هم سهم بازار بیشتری به دست آید که خود یک شاخص مهم مزیت رقابتی است. اما یک الگوریتم جامع، یکپارچه و کاربردی برای بخش‌بندی بازار وجود ندارد که متناسب با شرایط بازارهای الکترونیکی باشد و خود را صرفاً به برخی

است که داده‌های زاید و غیرقابل استفاده‌ای را نیز دربر دارد و لازم است قبل از استفاده پاکسازی یا پیش‌پردازش^{۱۳} شوند. در پیش‌پردازش برخی داده‌های کم اهمیت برای پژوهش از مجموع داده‌ها حذف می‌شوند. این مرحله، فرآیندی چالشی به ویژه برای پایگاه‌های داده بزرگ است (زوکرا^{۱۴} و همکاران، ۲۰۰۰)، زیرا انتخاب داده‌های مناسب نقش مهمی در کیفیت نتایج دارد. در این پژوهش پیش‌پردازش به منظور حذف داده‌های زاید و استخراج داده‌های تازگی، تکرار و مقدار پولی خرید^{۱۵} انجام می‌شود. این تحلیل مبتنی بر اصل پارتو (۸۰٪ سود شرکت توسط ۲۰٪ مشتریان به دست می‌آید) است و به شناسایی سودآورترین مشتریان کمک می‌کند، علاوه بر این که یکی از ساده‌ترین شیوه‌ها برای تعیین ارزش عمر مشتری است. تازگی، آخرین تاریخ خرید مشتری است که هر چه کمتر باشد بهتر است، زیرا با مشتری سروکار داریم که به احتمال بیشتری نسبت به خرید آتی اقدام می‌کند.

تکرار، تعداد دفعاتی است که مشتری نسبت به خرید اقدام کرده است و هر چه بیشتر باشد بهتر است، زیرا مبتنی بر این فرض است که انسان تحت تاثیر عادت است و محصولی که در گذشته چندین بار خریده است را با احتمال بیشتری می‌خرد و ارزش پولی، مقدار پولی است که مشتری در گذشته برای خرید محصول یا خدمت پرداخته است و هر چه بیشتر باشد بهتر است. اما از آنجا که اهمیت هر یک از مقادیر تازگی، تکرار و ارزش پولی در بازارهای مختلف ممکن است متفاوت باشند، می‌توان از مقادیر موزون آن‌ها استفاده کرد. وزن‌ها را

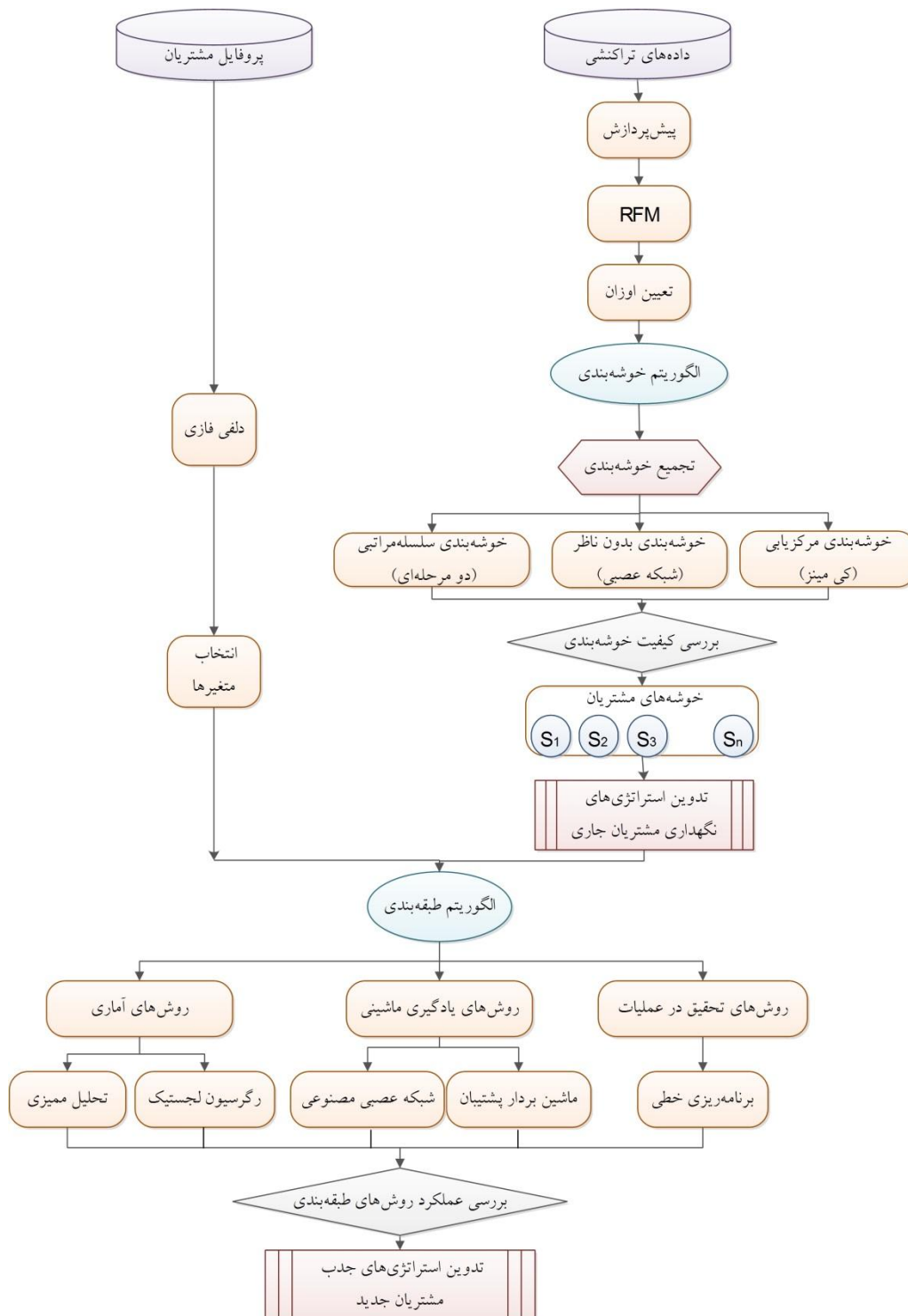
استام^{۱۶}، لام و موی^۹ (۱۹۹۶)، گوچت^{۱۰} و همکاران (۱۹۹۷). اخیرا پای و همکاران (۲۰۱۲) نشان داده‌اند مدل لام و موی (۱۹۹۶) نسبت به روش‌های تحلیل ممیزی، رگرسیون لجستیک، شبکه‌های عصبی و k-امین نزدیک‌ترین همسایگی برتری دارند اما آن‌ها صرفاً به مساله طبقه‌بندی پرداخته و فرض کرده‌اند بخش‌ها از قبل تعیین شده است. یوسف و ربای^{۱۱} (۲۰۰۷) روش برنامه‌ریزی خطی را با روش‌های استاندارد آماری برای طبقه‌بندی مقایسه کرده‌اند و نشان داده‌اند که روش غیرپارامتریک برنامه‌ریزی خطی نسبت به روش‌های پارامتریک آماری مانند تحلیل ممیزی فیشر عملکرد بهتری در طبقه‌بندی از خود نشان داده است. این پژوهش‌گران روش‌های یادگیری ماشینی را در پژوهش خود در نظر نگرفته‌اند. کواک^{۱۲} و همکاران (۲۰۰۲) روش برنامه‌ریزی خطی را در کنار سایر روش‌های طبقه‌بندی مانند روش تحلیل ممیزی و روش نزدیک‌ترین همسایگی برای طبقه‌بندی دو گروهی بیماران مبتلا به ویروس اچ.آی.وی و ایدز به کار برده‌اند و نشان داده‌اند که روش برنامه‌ریزی خطی عملکرد بهتری داشته است. مدل مورد استفاده این پژوهش‌گران مدل استاندارد دو بخشی است و قادر به طبقه‌بندی چندگروهی نیست.

۲- الگوریتم پژوهش

بر مبنای مطالعات انجام شده، الگوریتمی طراحی شد که در شکل (۱) ارائه شده است و در آن دو مسیر برای پردازش داده‌های تراکنشی و داده‌های پروفایل مشتریان وجود دارد.

۲-۱- داده‌های تراکنشی برای خوشه‌بندی مشتریان

این داده‌ها اغلب داده‌های رفتاری نامیده می‌شوند که نشانگر فعالیت‌های واقعی خرید مشتری است و اغلب شامل اقلام متعددی از داده‌های ثبت شده



شکل ۱- الگوریتم پژوهش

تحلیل خوشه‌بندی دومرحله‌ای بر پایه خوشه‌بندی سلسله‌مراتبی و شامل دو مرحله است (چیو، ۲۰۰۱). در مرحله اول که پیش‌خوشه‌بندی نامیده می‌شود، چندین خوشه کوچک شکل می‌گیرد. در مرحله دوم که خوشه‌بندی سلسله‌مراتبی گفته می‌شود، خوشه‌های کوچک مرحله اول مجدداً در هم ادغام و چند خوشه بزرگتر شکل می‌گیرد (پایگاه اینترنتی اس.پس.اس.اس^{۲۴}، ۲۰۱۳).

۲-۱-۲- بررسی کیفیت خوشه‌بندی

به تعداد پژوهش‌های انجام شده در مورد خوشه‌بندی، در مورد کیفیت خوشه‌بندی هم مطالعه صورت گرفته است تا از بیشترین فشردگی حول مرکز هر خوشه و بیشترین جدایی بین مراکز خوشه‌ها اطمینان حاصل شود. روش‌های متعددی مانند روش سیلوئت، انحراف معیار، مجموع مربع خطا و غیره ارائه شده‌اند که به دلیل وجود برخی شواهد از برتری سیلوئت به روش‌های دیگر، در این پژوهش از آن استفاده می‌شود (پتروویچ^{۲۵}، ۲۰۰۶). مقدار سیلوئت هر نقطه، نسبت اندازه شباهت آن نقطه با نقاط درون خوشه‌اش به میزان شباهت با نقاط دیگر خوشه‌ها است. از این تعریف مشخص می‌شود که نمی‌توان سیلوئت را فقط برای یک خوشه به دست آورد. به علاوه اگر تعداد نقاط با تعداد خوشه‌ها برابر باشد، مقدار سیلوئت بیشینه و برابر یک خواهد بود (روزو^{۲۶}، ۱۹۸۶).

۲-۲- داده‌های پروفایل مشتریان

داده‌های پروفایل مشتریان معرف ویژگی‌های مختلف مشتریان از جمله ویژگی‌های دموگرافیک، جغرافیایی و روان‌نگاشتی است که برای ایجاد مدل

می‌توان به طور شهودی یا با استفاده از روش‌هایی مانند فرآیند تحلیل سلسله‌مراتبی^{۱۶} تعیین کرد.

۲-۱-۱- خوشه‌بندی

اکنون داده‌های تازگی، تکرار و ارزش پولی موزون با استفاده از روش‌های کی-مینز^{۱۷}، نقشه‌های خودسازمانده^{۱۸} و خوشه‌بندی دو مرحله‌ای^{۱۹} به صورت تجمیعی خوشه‌بندی^{۲۰} می‌شوند. خوشه‌بندی تجمیعی به شیوه‌های گوناگونی، مانند استفاده از یک روش خوشه‌بندی با شرایط اولیه مختلف به دفعات و استفاده از روش‌های مختلف خوشه‌بندی با شرایط اولیه یکسان، قابل انجام است (کتسیانیتس و پیئتلا^{۲۱}، ۲۰۰۴). در روش کی-مینز نخست تعداد خوشه‌ها (k) توسط کاربر تعیین می‌شود. آنگاه از میان داده‌های n مشتری، k مشتری به عنوان مراکز ابتدایی خوشه‌ها برگزیده شده و بقیه مشتریان برحسب نزدیکی اقلیدسی به این مراکز متناسب می‌شوند. سپس مراکز جدید خوشه‌ها به صورت میانگین مقادیر هر خوشه محاسبه شده و هر مشتری مطابق با فاصله اقلیدسی‌اش به این مراکز جدید متناسب می‌گردد. این فرآیند هنگامی که هیچ انتساب جدیدی وجود نداشته باشد متوقف می‌شود (بلتبرگ و همکاران، ۲۰۰۸). شبکه عصبی خودسازمانده بر اساس ساختار و عملکرد مدل کوهونن^{۲۲} پایه‌ریزی شده است که مدلی بدون نظارت است. دلیل استفاده از این روش برای خوشه‌بندی در این پژوهش، تاکید برخی از پژوهش‌ها بر عملکرد بهتر آن نسبت به روش رایج کی-مینز است (حنفی‌زاده و میرزاده، ۲۰۱۱؛ چیو و همکاران، ۲۰۰۹؛ بلتبرگ^{۲۳} و همکاران، ۲۰۰۸).

توزیع نرمال چندمتغیره هستند، اما جهان واقعی کمتر با چنین پیش‌فرض‌هایی سر سازگاری دارد. در نتیجه تخلف از فرض نرمال بودن داده‌ها منجر به سوگیری و انتساب ناصحیح مشتریان به بخش‌ها و سرانجام منجر به استراتژی‌های غیراثربخش بازاریابی می‌شود (کیانگ^{۳۱} و همکاران، ۲۰۰۶). از این رو توجه پژوهش‌گران به توسعه روش‌هایی جلب شده است که محدود به چنین فرض‌هایی نباشند و روش‌های غیرپارامتری طبقه‌بندی مانند برنامه‌ریزی خطی که نیازمند چنین پیش‌فرض‌هایی نیستند امروزه نیز مورد توجه هستند (لام و موی^{۳۲}، ۱۹۹۶). به همین دلیل، در این پژوهش، علاوه بر روش‌های مطرح در آمار مانند تحلیل ممیزی و رگرسیون لجستیک و روش‌های یادگیری ماشینی مانند شبکه‌های عصبی و ماشین بردار پشتیبان، از برنامه‌ریزی خطی نیز برای طبقه‌بندی استفاده می‌شود تا عملکرد روش‌های گفته شده با یکدیگر مقایسه شوند. روش‌های مختلفی، مانند نسبت موفقیت^{۳۳}، ماتریس پریشانی^{۳۴} و محاسبه درصد طبقه‌بندی صحیح، برای بررسی عملکرد روش‌های طبقه‌بندی وجود دارد که در این پژوهش میزان درصد طبقه‌بندی صحیح استفاده شده است.

۲-۲-۲- طبقه‌بندی مشتریان با استفاده از روش

تحلیل ممیزی

در این پژوهش از روش تحلیل ممیزی مرحله‌ای^{۳۵} ماهانولوبیس برای طبقه‌بندی و از معیارهای لاندای ویک و عدد F برای بررسی اعتبار نتایج آن استفاده می‌شود. در روش مرحله‌ای توان پیش‌بینی‌کنندگی ویژگی‌ها به صورت گام به گام

طبقه‌بندی مشتریان ضروری است. اما گاهی پایگاه داده شرکت به دلایلی مانند نداشتن دیدگاهی استراتژیک به بازاریابی الکترونیکی فاقد چنین داده‌هایی است و طبقه‌بندی مشتریان امکان‌پذیر نیست. یک راه حل اصولی تغییر فرآیند ثبت داده‌ها و انتظار برای جمع‌آوری است که طبقه‌بندی مشتریان را به آینده موکول می‌کند. راه حل کوتاه مدت که این پژوهش پیشنهاد می‌دهد استفاده از روش دلفی فازی است که در ضمیمه (۱) معرفی شده است.

۲-۲-۱- طبقه‌بندی مشتریان

هدف از طبقه‌بندی شناسایی رابطه‌ای بین یک متغیر وابسته با مقیاس اسمی و یک بردار ویژگی از مجموعه‌ای از مشاهدات است (پای و همکاران^{۳۷}، ۲۰۱۲). مثلاً یک بانک می‌خواهد مشتری متقاضی اعتبار بانکی را بر حسب مواردی مانند سن، جنس، درآمد و غیره به یکی از دو گروه خوش‌حساب و بدحساب نسبت دهد (یوسف و ربای^{۳۸}، ۲۰۰۷). برای دستیابی به این هدف، نخست مجموعه‌ای از داده‌های اولیه یا آموزش تحلیل می‌شوند و سپس مدلی برای پیش‌بینی با استفاده از ویژگی‌هایی که در داده‌ها وجود دارد، ارائه می‌گردد. اکنون می‌توان چنین مدلی را برای طبقه‌بندی مجموعه داده‌های مستقل آتی هم به کار برد (دیلون^{۳۹} و همکاران، ۱۹۷۹). روش‌های متعددی در ادبیات پژوهش برای طبقه‌بندی ارائه شده است که هر یک نقاط قوت و ضعف خود را دارد. به عنوان مثال روش‌های پارامتری (مانند تحلیل ممیزی خطی فیشر^{۳۰}) مبتنی بر این فرض است که ویژگی‌های مشاهده‌دارای

همبسته بین داده‌های ورودی و مقادیر متناظر هدف را شناسایی و آن را یاد بگیرند و آنگاه این یادگیری را برای پیش‌بینی پیامد داده‌های ورودی جدید به کار برند. مرسوم‌ترین روش برای یادگیری (تنظیم وزن‌ها) الگوریتمی است که به آن انتشار به عقب^{۴۰} گفته می‌شود و در آن خروجی محاسبه شده شبکه و خروجی مورد انتظار مقایسه می‌شود. سپس خطای حاصل برای تغییر پارامترهای شبکه به کار می‌رود. مقادیر پیش‌بینی کننده که در این پژوهش از پروفایل مشتری استخراج می‌شود به عنوان نرون‌های ورودی عمل می‌کنند. عملکرد شبکه‌های عصبی بستگی به تابع فعال‌سازی نرون‌ها دارد و به همین دلیل در این پژوهش عملکرد توابع مختلف بررسی می‌شوند. پژوهش‌گران نتوانسته‌اند برتری مطلق شبکه‌های عصبی را نسبت به روش‌های آماری سنتی برای طبقه‌بندی نشان دهند، مثلاً زهاوی و لوین^{۴۱} (۱۹۹۷) شبکه‌های عصبی را با رگرسیون لجستیک مقایسه کرده‌اند و نشان داده‌اند که نتایج هر دو مدل یکسان بوده اما تفسیر نتایج رگرسیون لجستیک ساده‌تر است.

۲-۲-۵- ماشین بردار پشتیبان

ماشین بردار پشتیبان می‌کوشد خط جداکننده‌ای با بیشترین حاشیه اطمینان بین گروه‌هایی از داده‌هایی که اغلب خطی نیستند، پیدا کند. قبل از آن، داده‌ها را به کمک توابع کرنل به فضای هیلبرت که فضایی با ابعاد خیلی بالاست می‌برد (شیو^{۴۲}، ۲۰۰۴). اما از آنجا که چالش عمده در روش ماشین بردار پشتیبان، انتخاب تابع کرنل مناسب است در این پژوهش برای بررسی عملکرد توابع مختلف

انجام می‌شود و در هر مرحله متغیری وارد محاسبات می‌شود که فاصله ماهانالوبیس بیشتری داشته باشد. برای بررسی معنادار بودن نتایج تحلیل ممیزی از لاندای ویک^{۳۶} و عدد F است. وقتی لاندای ویک برابر با یک باشد، به این معنی است که میانگین گروه‌ها برابر هستند و به خوبی تفکیک نشده‌اند. همچنین اگر F کوچکتر از یک مقدار مشخص باشد، آن متغیر به توان متمایزکنندگی اضافه نمی‌کند و از این رو حذف می‌شود. معمولاً مقدار حدی F را برابر $\frac{3}{84}$ در نظر می‌گیرند (کلکا^{۳۷}، ۱۹۸۰).

۲-۲-۳- رگرسیون لجستیک

روش دوم از گروه روش‌های آماری برای طبقه‌بندی که در این پژوهش استفاده می‌شود، رگرسیون لجستیک است که مانند رگرسیون خطی است، با این تفاوت که متغیر وابسته آن به صورت اسمی و به جای توزیع نرمال از توزیع لجستیک استفاده می‌شود. روش رگرسیون لجستیک با متغیر وابسته‌ی دوتایی سروکار دارد. اما برای استفاده از آن در این پژوهش که در آن سه خوشه مشتریان وجود دارد، از روش کلی‌تر رگرسیون لجستیک چندجمله‌ای^{۳۸} مبتنی بر اصل تخمین بیشترین احتمال استفاده می‌شود.

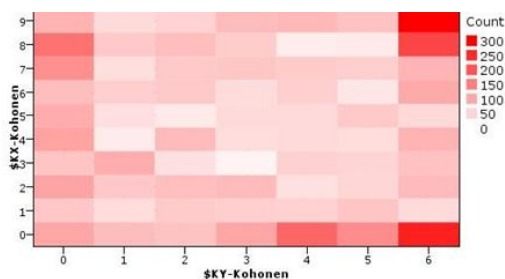
۲-۲-۴- شبکه عصبی مصنوعی

برخی از پژوهش‌های دو دهه اخیر نشان داده‌اند که روش شبکه عصبی ممکن است طبقه‌بندی کننده بهتری نسبت به روش‌های استاندارد آماری باشد (اودوم و شردا^{۳۹}، ۱۹۹۰). شبکه‌های عصبی مصنوعی می‌توانند با استفاده از داده‌های موجود الگوهای

پولی از روش تحلیل سلسله مراتب فازی استفاده شد. بدین منظور، در جلسه‌ای از کارشناسان شرکت خواسته شد تا متغیرهای تازگی، تکرار و ارزش پولی را به صورت زوجی مقایسه کنند و اهمیت هر یک نسبت به دیگری را بر مبنای متغیرهای کلامی «خیلی کم، کم، متوسط، زیاد و خیلی زیاد» بیان کنند. سپس متغیرهای کلامی به اعداد فازی تبدیل شدند. پس از استفاده از معادلات تحلیل سلسله مراتب فازی، وزن‌های گردشده تازگی، تکرار و ارزش پولی به ترتیب $0/3$ ، $0/2$ و $0/5$ به دست آمد.

۳-۱- تجمیع خوشه‌بندی برای بخش‌بندی مشتریان

از آن جا که نمی‌توان صرفاً به نتایج یکی از روش‌های خوشه‌بندی اتکا کرد، برای اطمینان از انجام بهترین خوشه‌بندی از تجمیع روش‌های کی-مینز، شبکه عصبی خودسازمان‌ده و روش خوشه‌بندی دو مرحله‌ای استفاده شد. در ابتدا داده‌های تازگی، تکرار و ارزش پولی موزون با استفاده از روش شبکه عصبی خودسازمان‌ده و با اندازه نقشه خروجی 7×10 و در محیط نرم‌افزار کلمنتاین خوشه‌بندی شد که تصویر خروجی آن در شکل (۲) ارائه شده است.



شکل ۲- تعداد مشتریان به عنوان خروجی نقشه‌های خودسازمان‌ده با ابعاد 7×10 (منبع: یافته‌های پژوهش)

کرنل، از توابع آر.بی.اف.^{۴۳}، چندجمله‌ای، سیگما و خطی استفاده می‌شود. مزیت اصلی روش ماشین بردار پشتیبان این است که هیچ بهینه محلی در آن وجود ندارد و برای داده‌های با ابعاد بزرگ به خوبی کار می‌کند. نقطه ضعف آن انتخاب یک تابع کرنل خوب و مناسب است (ژاکولا^{۴۴}، ۲۰۱۱).

۲-۲-۶- برنامه‌ریزی خطی برای طبقه‌بندی مشتریان

ایده استفاده از روش‌های برنامه‌ریزی خطی برای طبقه‌بندی را منگسریان^{۴۵} در سال ۱۹۶۵ معرفی کرد و پس از معرفی مدلی مبتنی بر آن توسط فرید و گلور^{۴۶} در سال ۱۹۸۱ موجب علاقمندی پژوهش‌گران به آن گردید، اما هر دو مدل برای حل مسایل طبقه‌بندی دو گروهی بودند. در سال ۱۹۹۶ لام و موی مدلی معرفی کردند که مساله طبقه‌بندی چندگروهی را به صورت چندین مساله دو گروهی حل می‌کند و در ضمیمه (۲) معرفی شده است. پای و همکاران (۲۰۱۲) عملکرد همین مدل را با سایر مدل‌های مرسوم طبقه‌بندی مقایسه کرده و نشان داده‌اند که عملکرد آن تا زمانی که ماهیت خطی مساله حفظ شود بهتر است. از جمله مزایای مهم برنامه‌ریزی خطی گرفتار نشدن در بهینه‌های محلی، و وجود فرآیندی ساده‌تر و قابل درک‌تر است.

۳- تجزیه و تحلیل یافته‌های پژوهش

داده‌های مشتریان اینترنت پرسرعت شرکت ایران‌گیت برای تجزیه و تحلیل مورد استفاده قرار می‌گیرد. براساس الگوریتم پیشنهاد شده ابتدا داده‌های «تازگی، تکرار و ارزش پولی» در محیط نرم‌افزار کلمنتاین استخراج شد. سپس برای تعیین وزن هر یک از متغیرهای تازگی، تکرار و ارزش

یک رویکرد جامع برای بخش‌بندی بازار و طبقه‌بندی مشتریان ۹۱

شناسایی کرده است ولی تعداد مشتریان ارزشمند را کمتر از دو روش قبل و در حدود ۱۲۰۰ مشتری برآورد کرده است. از نظر بازاریابی اتکا صرف به این روش موجب عدم تمرکز بر بخش بیشتری از مشتریان به عنوان مشتریان غیرارزشمند می‌شود.

جدول ۲ - نتایج خوشه‌بندی کی-مینز و نقشه‌های خودسازمان‌دهنده برای سه خوشه

کی-مینز					
خوشه‌ها	تعداد مشتری	میانگین تازگی	میانگین تکرار	میانگین ارزش پولی	میانگین RFM
اول	۲۳۸۷	۴/۸۳۸	۲/۰۶۲	۲/۵۸۲	۳/۱۵۵
دوم	۱۷۷۶	۶/۸۰۶	۷/۰۶۸	۶/۳۰۶	۶/۶۰۹
سوم	۱۱۰۸	۲/۴۰۷	۱/۹۷۷	۷/۸۳۶	۵/۰۳۵
نقشه‌های خودسازمان‌دهنده					
خوشه‌ها	تعداد مشتری	میانگین تازگی	میانگین تکرار	میانگین ارزش پولی	میانگین RFM
اول	۲۴۴۸	۳/۳۹۹	۱/۴۷۲	۳/۸۳۰	۳/۲۲۹
دوم	۱۷۹۹	۶/۸۱۰	۷/۰۱۰	۶/۴۴۴	۶/۶۶۷
سوم	۱۰۲۴	۵/۶۰۱	۳/۳۶۹	۴/۹۵۸	۴/۸۳۳

جدول ۳ - نتایج خوشه‌بندی سلسله مراتبی دو

مرحله‌ای

خوشه‌بندی دو مرحله‌ای					
خوشه‌ها	تعداد مشتری	میانگین تازگی	میانگین تکرار	میانگین ارزش پولی	میانگین RFM
اول	۱۲۷۹	۶/۵۱۴	۸/۰۲۷	۶/۷۲۵	۶/۹۲۲
دوم	۱۷۵۲	۶/۶۲۲	۳/۲۷۵	۴/۴۶۲	۴/۸۷۳
سوم	۸۹۹	۱/۹۳۵	۱/۷۰۴	۸/۰۰۹	۴/۹۲۶
چهارم	۱۳۴۱	۳/۴۵۵	۱/۵۸۹	۱/۸۱۰	۲/۲۵۹

اما برای تعیین کیفیت خوشه‌بندی سه روش فوق، مقادیر سیلوئت در محیط نرم‌افزار متلب محاسبه شد که در جدول (۴) ارائه شده است و مطابق با آن روش کی-مینز کیفیت بهتری دارد و از این جا به بعد مبنای محاسبات بعدی قرار می‌گیرد.

در این شکل هر چه تعداد مشتریان در یک خوشه بیشتر باشد، رنگ خوشه قرمز تندتر است. بنابراین می‌توان ۳ یا ۴ خوشه را به عنوان بهترین تعداد خوشه تعیین کرد. سپس با استفاده از این اطلاعات و کاربرد روش کی-مینز برای تعداد خوشه‌های مختلف (K)، خوشه‌بندی انجام شد و معیار سیلوئت برای تعیین تعداد بهینه خوشه‌ها به کار رفت که نتایج آن در جدول (۱) نشان داده شده است:

جدول ۱ - نتایج معیار سیلوئت برای خوشه‌های مختلف به دست آمده از روش کی-مینز

K	۲	۳	۴	۵	۶
سیلوئت	۰/۴۴۰۸	۰/۶۰۰۵	۰/۴۱۲۳	۰/۵۷۲۱	۰/۵۴۳۳

دیده می‌شود که اگر تعداد خوشه‌ها برابر با ۳ در نظر گرفته شود، بهترین مقدار سیلوئت به دست می‌آید. این امر دلیل دیگری بر وجود ۳ خوشه در میان داده‌ها است. برای اطمینان بیشتر بار دیگر با استفاده از روش خوشه‌بندی نقشه‌های خودسازمان‌دهنده و با تعیین ابعاد نقشه خروجی آن روی ۳×۳ محاسبات تکرار شد. نتایج حاصل از خوشه‌بندی دو روش کی-مینز و نقشه‌های خودسازمان‌دهنده در جدول ۲ نشان داده شده است. مطابق با جدول (۲)، خوشه دوم در هر دو روش با حدود ۱۷۰۰ مشتری دارای بالاترین میانگین تازگی، تکرار و ارزش پولی و ارزشمندترین خوشه است و تقریباً ۳۰ درصد کل مشتریان هستند، که تاحدی مطابق با اصل پارتو یا اصل ۲۰-۸۰ است. در جدول (۳) نتایج روش خوشه‌بندی سلسله مراتبی دو مرحله‌ای نیز ارائه شده است. این روش به صورت خودکار ۴ خوشه را

جدول ۴- مقایسه مقدار سیلوئت سه روش خوشه‌بندی

روش	روش کی- مینز	روش نقش‌های خودسازمانده	روش دو مرحله‌ای
سیلوئت	۰/۶۰۶	۰/۲۷۴۶	۰/۵۴۷۶

۳-۲- مقایسه عملکرد روش‌های طبقه‌بندی کننده

هنگام بررسی داده‌های مشتریان شرکت ایران‌گیت مشخص شد که داده‌های پروفایل مشتریان وجود ندارد. از این رو با استفاده از روش دلفی فازی که در ضمیمه (۱) معرفی شده است، پنج متغیر تحصیلات، هدف از استفاده از اینترنت، درآمد سالیانه، جنس و سن به عنوان پیش‌بینی‌کننده‌های مهم انتخاب شدند و از طریق مصاحبه تلفنی، پنج داده مورد اشاره از ۱۸۰۸ مشتری جمع‌آوری گردید. سپس طبقه‌بندی مشتریان با استفاده از فنون آماری تحلیل ممیزی، رگرسیون لجستیک و فنون یادگیری ماشینی شبکه‌های عصبی و ماشین بردار پشتیبان در کنار برنامه‌ریزی خطی انجام شد که در ادامه نتایج آن ارائه شده است.

۳-۲-۱- تحلیل ممیزی

آزمون شاپیرو ویلک^{۴۷} که آزمونی برای بررسی نرمال بودن داده‌ها است، نشان داد که به جز سن و درآمد، بقیه داده‌های مورد استفاده غیرنرمال هستند. گرچه این شرط اساسی برای استفاده از تحلیل ممیزی وجود ندارد، اما گاهی آزمون‌های آماری بدون توجه به پیش‌شرطها مورد استفاده قرار می‌گیرند و در اینجا نیز برای نشان دادن آن چه از تحلیل ممیزی به دست می‌آید، با چشم پوشی از این

خطا سعی در یافتن توابع ممیزی می‌شود. در ابتدا ۶۵ درصد داده‌ها به عنوان داده‌های آموزشی مورد استفاده قرار می‌گیرد. هدف تحلیل ممیزی، انتخاب مهمترین متغیرها و تعیین توابعی است که تمایز میان خوشه‌ها را نمایان می‌سازد. برای دستیابی به این هدف، از روش تحلیل ممیزی تدریجی استفاده شد. مطابق با آنچه در بخش ۲-۲-۲ گفته شد، در مرحله اول درآمد و در مرحله دوم هدف به عنوان ویژگی‌هایی که بیشترین سهم را در ایجاد تمایز میان گروه‌ها دارند، انتخاب می‌شوند چون در هر مرحله کمترین لاندای ویک و بیشترین عدد F را دارند.

جدول ۵- مقادیر لاندای ویک و F برای انتخاب مهم‌ترین متغیرها در روش تحلیل ممیزی تدریجی پیش‌رونده

مرحله	متغیر	لاندای ویک	عدد F برای ورود
اول	سن	۰/۹۶۸	۱۹/۴۴۸
	جنس	۰/۹۹۹	۰/۸۵۴
	تحصیلات	۰/۹۷۹	۱۲/۳۹۶
	هدف	۰/۹۵۱	۲۹/۵۵۳
	درآمد سالیانه	۰/۸۶۶	۸۹/۸۲۲
دوم	سن	۰/۸۶۵	۰/۱۸۶
	جنس	۰/۸۶۱	۳/۴۳۵
	تحصیلات	۰/۸۴۸	۱۲/۲۷۵
	هدف	۰/۸۲۱	۳۱/۴۵۳
سوم	سن	۰/۸۲۱	۰/۱۳۲
	جنس	۰/۸۱۸	۲/۲۵۱
	تحصیلات	۰/۸۱۹	۱/۴۰۴

به این ترتیب سه تابع ممیزی که ۱۰۰ درصد واریانس ویژگی‌ها را دربر می‌گیرند، ارائه می‌شود که به صورت زیر است:

$$F_1 = -5.506 + 0.074 \times income + 3.374 \times purpose$$

$$F_2 = -4.803 + 0.091 \times income + 2.862 \times purpose$$

$$F_3 = -7.319 + 0.132 \times income + 3.531 \times purpose$$

یک رویکرد جامع برای بخش‌بندی بازار و طبقه‌بندی مشتریان | ۱۱

$$0.9714 \times [Purpose=1] + 0.7218 \times [Purpose=2] + 0.0562 \times Income - 0.8118$$

$$F3 = -4.096 + 0.4978 \times [Sex=0] - 0.5183 \times [Education=0] + 1.695 \times [Education=1] + 1.972 \times [Education=2] + 1.753 \times [Education=3] + 1.54 \times [Education=4] - 20.76 \times [Purpose=0] - 0.2584 \times [Purpose=1] - 0.1181 \times [Purpose=2] + 0.1047 \times Income$$

توابع رگرسیون شامل سطوح مختلف متغیرها است. به عنوان مثال، در تابع دوم ضریب رگرسیونی مشتریان با تحصیلات زیر دیپلم (کد صفر) برابر با ۰/۵۱۸۳- است، در حالی که ضریب رگرسیونی مشتریان با تحصیلات لیسانس (کد ۳) برابر با ۱/۷۵۳+ است. به عبارت دیگر مشتریان با تحصیلات لیسانس در خوشه دوم دارای بالاترین ضریب رگرسیونی هستند. اینک برای بررسی میزان دقت پیش‌بینی این توابع رگرسیونی، آن‌ها را با استفاده از داده‌های آزمایشی به کار می‌بریم که نتایج آن در جدول (۷) نشان داده شده است. همچنان که مشاهده می‌شود، میزان انتساب صحیح مشاهدات به خوشه اول ۵۸/۵ درصد، به خوشه دوم ۵۰ درصد و به خوشه سوم ۵۰/۶ درصد و عملکرد یا دقت پیش‌بینی کلی مدل ۵۳/۱ درصد است که اندکی بیش از دقت مدل تحلیل ممیزی است.

با در اختیار داشتن توابع ممیزی که با استفاده از داده‌های آموزشی به دست آمده‌اند و برای بررسی میزان دقت مدل، از داده‌های آزمایشی استفاده می‌شود تا میزان طبقه‌بندی صحیح مدل به دست آید. نتایج در جدول (۶) ارائه شده است. همچنانکه مشاهده می‌شود دقت پیش‌بینی تحلیل ممیزی در انتساب صحیح مشاهدات به خوشه اول ۵۲/۲ درصد، به خوشه دوم ۵۰/۸ درصد و به خوشه سوم ۴۶ درصد است که به طور میانگین برابر با ۴۹/۷ درصد است. به عبارت دیگر کمتر از نصف پیش‌بینی‌های ارائه شده از رفتار آتی مشتریان مقرون به صحت است.

جدول ۶- دقت پیش‌بینی تحلیل ممیزی بر مبنای داده‌های آموزشی

نتایج طبقه‌بندی						
کل	عضویت پیش‌بینی شده در			خوشه	تعداد	اعتبار سنجی
	۳	۲	۱			
۳۹۵	۴۴	۱۴۵	۲۰۶	۱		
۳۹۲	۹۴	۱۹۹	۹۹	۲		
۳۷۴	۱۷۲	۸۹	۱۱۳	۳		
۱۰۰	۱۱/۱	۳۶/۷	۵۲/۲	۱	درصد	ی
۱۰۰	۲۴/۰	۵۰/۸	۲۵/۳	۲		
۱۰۰	۴۶/۰	۲۳/۸	۳۰/۲	۳		

۳-۲-۲- رگرسیون لجستیک

با در نظر گرفتن خوشه اول به عنوان خوشه مرجع، مدل رگرسیون لجستیک با استفاده از تابع کرنل RBF و داده‌های آموزشی مورد استفاده قرار گرفت تا توابع رگرسیون به دست آید. دو تابع رگرسیون به دست آمده به شرح زیر است:

$$F2 = 0.0879 \times [Sex=0] - 2.836 \times [Education=0] - 0.4651 \times [Education=1] - 0.9607 \times [Education=2] - 0.8236 \times [Education=3] - 0.981 \times [Education=4] - 20.26 \times [Purpose=0] +$$

جدول ۷- نتایج طبقه‌بندی رگرسیون لجستیک با استفاده از تابع RBF

نتایج طبقه‌بندی					
کل	عضویت پیش‌بینی شده			خوشه	
	۳	۲	۱		
۳۹۵	۴۳	۱۲۱	۲۳۱	۱	تعداد
۳۹۲	۹۴	۱۹۶	۱۰۲	۲	
۳۷۴	۱۸۹	۷۲	۱۱۳	۳	
۱۰۰	۱۰/۹	۳۰/۶	۵۸/۵	۱	درصد
۱۰۰	۲۴	۵۰	۲۶	۲	
۱۰۰	۵۰/۶	۱۹/۲	۳۰/۲	۳	

می‌دهد (پایگاه اینترنتی اس.پی.اس، ۲۰۱۳). در جدول (۸) دیده می‌شود که میزان دقت پیش‌بینی تابع هرس فراگیر که دارای پیچیدگی و لایه‌های پنهان بیشتری است، بالاتر است. اما پیچیدگی و لایه‌های پنهان بیشتر، شبکه عصبی را مانند یک جعبه سیاه می‌نماید که دشواری تحلیل نتایج را در پی دارد.

جدول ۸- ساختار و عملکرد توابع مختلف شبکه عصبی در طبقه‌بندی

تابع فعال‌سازی	نرون‌های لایه ورودی	نرون‌های لایه پنهان اول	نرون‌های لایه پنهان دوم	نرون‌های لایه پنهان سوم	نرون‌های لایه خروجی	دقت پیش‌بینی شده
RBFN	۱۳	۲۰	-	-	۳	۴۶/۸
پویا	۱۳	۸	۶	-	۳	۵۲/۶
هرس فراگیر	۱۳	۲۶	۱۶	-	۳	۷۰/۶
چندگانه	۱۳	۱۲	۱۲	۱۱	۳	۶۴/۴

۳-۲-۳- شبکه عصبی مصنوعی

برای بررسی عملکرد توابع مختلف شبکه عصبی، توابع پویا^{۴۸}، چندگانه^{۴۹}، هرس فراگیر^{۵۰} و تابع پایه شعاعی^{۵۱} مورد بررسی قرار می‌گیرند. در روش پویا یا دینامیک در طی فرآیند آموزش به لایه‌های پنهان افزوده یا از آن‌ها کاسته می‌شود. در روش چندگانه، در طی فرآیند آموزش به صورت همزمان چندین روش به کار گرفته می‌شود. در انتها مدلی که دارای پایین‌ترین خطا است، به عنوان مدل نهایی ارائه می‌شود. روش شبکه تابع پایه شعاعی از روشی مانند خوشه‌بندی کی-مینز استفاده می‌کند تا داده‌ها را بر مبنای مقادیر گروه‌های هدف تقسیم کند. همچنین در روش هرس فراگیر^{۵۲}، ابتدا شبکه بزرگی ایجاد می‌شود و سپس ضعیف‌ترین واحدها در لایه‌های پنهان و ورودی در طی فرآیند آموزش حذف می‌شوند. این روش از همه روش‌های گفته شده زمان‌برتر است، اما بهترین جواب را به دست

۳-۲-۳- ماشین بردار پشتیبان

چهار تابع کرنل مختلف برای بررسی عملکرد آن‌ها مورد استفاده قرار می‌گیرد. معیار توقف در نرم‌افزار کلمنتاین روی ۰/۰۰۱ که مقدار قراردادی است تنظیم شد. هنگامی که از تابع کرنل RBF استفاده شد، مقدار گامای آن روی ۱/۲ تنظیم شد. این مقدار معمولاً بین ۳/k و ۶/k است که k تعداد خوشه‌های تعیین شده است. چون در اینجا تعداد خوشه‌ها ۳ است، مقدار آن بین ۱ و ۲ مجاز است. افزایش این مقدار، دقت طبقه‌بندی برای داده‌های آموزشی را افزایش می‌دهد، اما ممکن است منجر به دقت بیش از حد^{۵۳} شود. هنگامی که از تابع کرنل چندجمله‌ای یا سیگما استفاده شد، مقدار گاما به

یک رویکرد جامع برای بخش‌بندی بازار و طبقه‌بندی مشتریان ۱۳۱

محیط متلب نوشته شد، انتخاب گردید. سپس مدل (۵) برای محاسبه وزن ویژگی‌های مشتریان در سه بخشی که قبلاً به دست آمده است، مورد استفاده قرار گرفت که نتایج آن در جدول (۱۰) ارائه شده است. خوشه‌های ۱ و ۲، ۱ و ۳، و ۲ و ۳ با یکدیگر مقایسه و وزن ویژگی‌ها که تعیین کننده میزان اهمیت آنها در تفکیک دو گروه است به دست آمده است. به عنوان مثال، وزن ویژگی هدف در تفکیک بین خوشه‌های ۱ و ۲ برابر با ۰/۴۷۸ است که نسبت به وزن سایر ویژگی‌ها در همین مقایسه بیشتر است.

جدول ۱۰- وزن ویژگی‌های مشتریان در مقایسه

بخش‌ها

تایع	داده‌های آموزشی		داده‌های آزمایشی	
	تعداد	درصد	تعداد	درصد
آر.بی.اف	۶۷۵	۵۸	۳۶۳	۵۶
چندجمله‌ای	۸۴۱	۷۲	۴۱۳	۶۴
سیگما	۴۴۲	۳۸	۲۳۹	۳۷
خطی	۵۹۵	۵۱	۳۵۳	۵۴
میانگین	۶۳۸	۵۵	۳۴۲	۵۳

بررسی وزن‌های به دست آمده بین خوشه‌های ۱ و ۳ و خوشه‌های ۲ و ۳ نیز نشانگر اهمیت هدف مشتریان در ایجاد تمایز میان آنها است. این ویژگی بیشترین تاثیر را در تفکیک خوشه‌های ۲ و ۳ داشته است. در مرحله دوم خطوط برش میان سه خوشه مطابق با مدل (۶) محاسبه گردید که در جدول (۱۱) نتایج آن آمده است.

جدول ۱۱- خطوط برش بین سه بخش مشتریان

خطوط برش میان خوشه‌ها		
خوشه ۱	خوشه ۲	خوشه ۳
		خوشه ۱
		خوشه ۲
		خوشه ۳

صورت قراردادی ۲ انتخاب شد. افزایش این مقدار نیز می‌تواند دقت طبقه‌بندی را افزایش دهد، گرچه مانند قبل ممکن است به دقت بیش از حد منجر شود. نتایج در جدول (۹) ارائه شده است که نمایانگر برتری تابع کرنل چند جمله‌ای نسبت به سه تابع کرنل دیگر است. گرچه تابع کرنل چند جمله‌ای با ۷۲ درصد طبقه‌بندی صحیح، نتایج بهتری برای داده‌های آموزشی ارائه کرده است، اما میزان طبقه‌بندی صحیح آن برای داده‌های آزمایشی، ۶۴ درصد است. این امر نشان دهنده عدم ثبات یا پایداری مدل است.

جدول ۹- میزان طبقه‌بندی صحیح توابع کرنل مختلف ماشین بردار پشتیبان

	وزن ویژگی‌ها			
	سن	جنس	تحصیل	هدف
خوشه ۱	۰/۲۵۷	۰/۰۹۳۷	-۰/۱۱۷۲	۰/۴۷۸۴
خوشه ۲	۰/۰۴۰	۰/۰۸۶۳	-۰/۰۱۶۸	-۰/۱۲۶۰
خوشه ۳	۰/۰۱۶۴	۰/۰۱۲۸	۰/۲۶۲۶	-۰/۶۶۱۸

۳-۲-۴- برنامه‌ریزی خطی چندگروهی

تاکید اصلی این پژوهش بر توانایی برنامه‌ریزی خطی در انجام کاراتر طبقه‌بندی دست کم برای پایگاه‌های داده کم تا متوسط است. در ابتدا داده‌های جمعیت‌شناسی مربوط به ۱۸۰۸ مشتری نمونه شرکت ایران‌گیت، به صورت تصادفی به دو نمونه آموزشی (۶۵٪) و آزمایشی (۳۵٪) تقسیم شد.

۳-۲-۴-۱- مرحله اول: آموزش مدل برنامه‌ریزی

خطی

۶۵ درصد داده‌ها که متعلق به ۱۱۷۰ مشتری است به صورت تصادفی توسط نرم‌افزاری که به منظور حل مدل برنامه‌ریزی خطی چندگروهی در

همچنان که مشاهده می‌شود، به طور میانگین دقت طبقه‌بندی مدل برنامه‌ریزی خطی بالاتر از روش‌های استفاده شده است. یک نگرانی در مورد استفاده از برنامه‌ریزی خطی، مدت زمان بالای انجام محاسبات نسبت به سایر مدل‌ها است. اما امروزه با توجه به توانایی و سرعت بالای کامپیوترهای شخصی این نگرانی تا حد زیادی بر طرف شده است. زمان انجام محاسبات در مورد ۱۸۰۸ داده مورد استفاده این پژوهش با استفاده از برنامه‌ریزی خطی در حدود ۲ دقیقه بیشتر از سایر مدل‌ها بود.

۴- نتیجه‌گیری

مساله فقدان یک الگوریتم یکپارچه برای بخش‌بندی بازار و طبقه‌بندی مشتریان که به ویژه متناسب با شرایط بازارهای الکترونیکی باشد و صرفاً محدود به یکی دو روش مرسوم نباشد، انگیزه انجام این پژوهش بوده است. در این راستا، برای بخش‌بندی بازار صرفاً به یک روش اتکا نشده و روش‌های غالب کی-مینز، نقشه‌های خودسازمانده و خوشه‌بندی دو مرحله‌ای در قالب رویکردی تجمیعی مورد استفاده قرار گرفته تا نسبت به دقت و صحت نتایج اطمینان حاصل شود. به علاوه، برای اطمینان بیشتر معیار سیلوئت برای بررسی صحت و اعتبار نتایج مورد استفاده قرار گرفت. آن‌گاه به مساله‌ای که اغلب به آن توجه نمی‌شود و آن فقدان داده‌های مورد نیاز در پایگاه داده مشتریان است، توجه گردید و روش دلفی فازی برای جبران این کمبود پیشنهاد شد. سپس بر پایه نتایج حاصل از خوشه‌بندی و داده‌های گردآوری شده از طریق مصاحبه تلفنی، مشتریان با استفاده از روش‌های

۳-۲-۴- مرحله دوم: آزمون مدل برنامه‌ریزی خطی در این قسمت با استفاده از وزن‌های محاسبه شده ویژگی‌ها در خوشه‌های سه‌گانه و خطوط برش بین آن‌ها، نمرات طبقه‌بندی ۳۵ درصد باقیمانده مشتریان محاسبه می‌شود. نتایج دقت طبقه‌بندی با استفاده از داده‌های آموزشی در جدول (۱۲) و برای داده‌های آزمایشی در جدول (۱۳) ارائه شده است. همچنان که مشاهده می‌شود دقت مدل پس از محاسبات اولیه و برای داده‌های آزمایشی بیشتر و به طور میانگین برابر با ۶۹ درصد است. به عبارت دیگر با ۶۹ درصد اطمینان می‌توان احتمال تعلق مشتری جدید به یکی از ۳ بخش بازار را گمانه‌زنی کرد. جدول (۱۴) دقت روش‌هایی که در این پژوهش برای طبقه‌بندی استفاده شد را نشان می‌دهد.

جدول ۱۲- دقت طبقه‌بندی با استفاده از داده‌های آموزشی

مقایسه خوشه‌ها	داده‌های آموزشی (٪۶۵)
۱ و ۲	٪۶۵
۱ و ۳	٪۶۹
۲ و ۳	٪۶۳
میانگین	٪۶۶

جدول ۱۳- دقت طبقه‌بندی با استفاده از داده‌های آزمایشی

مقایسه خوشه‌ها	داده‌های آزمایشی (٪۳۵)
۱ و ۲	٪۶۸
۱ و ۳	٪۷۳
۲ و ۳	٪۶۵
میانگین	٪۶۹

جدول ۱۴- مقایسه دقت مدل‌های طبقه‌بندی کننده

مدل	داده‌های آموزشی	داده‌های آزمایشی
شبکه عصبی	٪۵۸	٪۶۰
تحلیل ممیزی	٪۵۰	٪۵۳
رگرسیون لجستیک	٪۵۳	٪۵۵
ماشین بردار پشتیبان	٪۵۵	٪۵۳
برنامه‌ریزی خطی	٪۶۶	٪۶۹

محدودیت دوم پژوهش آن است که نتایج به دست آمده در مورد یک شرکت مورد مطالعه است. حتی این نتایج با افزایش داده‌ها در پایگاه داده مشتریان همین شرکت نیز قطعا دستخوش تغییر خواهند شد. زیرا اساسا استخراج دانش از داده‌ها امری پویا است و از این رو ضروری است که شرکت به تناوب محاسبات را تکرار کند تا نتایج به روز رسانی شوند.

یک محدودیت پژوهش‌گر، بی‌میلی شرکت‌های الکترونیکی برای در اختیار قرار دادن داده‌های مورد نیاز بود. دلایل متعددی برای این بی‌ رغبتی وجود دارد که از جمله مهم‌ترین آن‌ها می‌توان به ضرورت در اختیار قرار دادن داده‌های مالی یا ارزش پولی مشتریان اشاره کرد. شرکت‌ها به دلیل نگرانی از نحوه استفاده از داده‌ها و پیامدهای احتمالی مانند امور مالیاتی از ارائه آن خودداری می‌کنند. یک دلیل دیگر آن است که لازم است تمامی داده‌های مشتریان از قبیل آدرس‌های تماس و اطلاعات خرید آن‌ها در اختیار پژوهش‌گر قرار گیرد و آن‌ها نگران این هستند که این داده‌ها در اختیار رقبا قرار گیرد. علت سوم بی‌میلی آن‌ها این است که شرکت‌ها اصولا با مباحث بخش‌بندی و طبقه‌بندی ناآشنا هستند و از مزایای آن آگاهی ندارند.

۴-۲- پیشنهادت به پژوهش‌گران آینده

پیشنهاد می‌گردد سایر روش‌های بخش‌بندی و طبقه‌بندی مطرح در ادبیات پژوهش در کنار روش برنامه‌ریزی خطی مورد استفاده و مقایسه قرار گیرند. همچنین پیشنهاد می‌گردد علاوه بر معیار سیلوئت، سایر معیارهای بررسی کیفیت بخش‌بندی نیز استفاده

آماري تحليل ممیزی و رگرسیون لجستیک و همچنین روش‌های یادگیری ماشینی شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان، طبقه‌بندی شدند. اما روش‌های گفته شده به دلیل درگیر بودن متغیرهای بسیار زیاد اغلب به صورت جعبه سیاه عمل می‌کنند و از این رو تفسیر نتایج آن‌ها به ویژه از منظر بازاریابی دشوار است، علاوه بر آن که گاه نیاز به وجود پیش‌شرط‌های خاص دارند و گاه دچار کاستی‌هایی مانند گرفتاری در بهینه‌های محلی هستند. از این رو این پژوهش نشان داد که روش طبقه‌بندی برنامه‌ریزی خطی نه تنها می‌تواند نتایج دقیق‌تری از روش‌های مرسوم گفته شده ارائه نماید، بلکه به دلیل سادگی، شفافیت و وجود متغیرهای شناخته شده و اندک، به ویژه برای بازاریابان، نتایج روشن‌تر و قابل تفسیرتری در بر دارد.

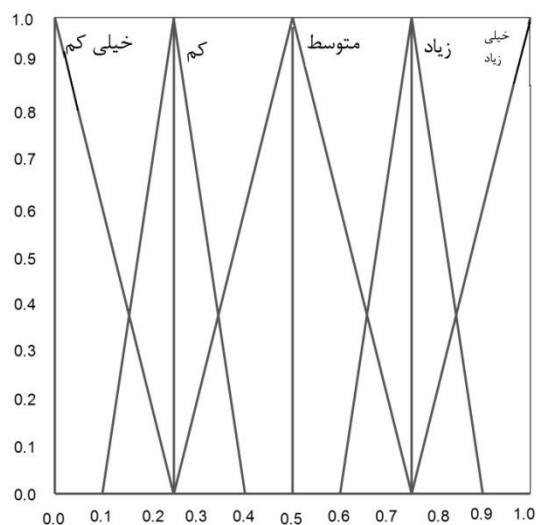
۴-۱- محدودیت‌های پژوهش و پژوهش‌گر

نخستین محدودیت پژوهش این است که هیچ روش یگانه‌ای برای بخش‌بندی و طبقه‌بندی وجود ندارد. زیرا بخش‌بندی نوعی داده‌کاوی اکتشافی است که در حوزه‌های مختلف و مرتبط به هم مانند آمار، یادگیری ماشینی، تشخیص الگو، بازاریابی اطلاعات و غیره مورد بحث و بررسی قرار گرفته است. طبقه‌بندی نیز در علوم چون آمار و ریاضی، شیمی، زیست‌شناسی، پزشکی، جامعه‌شناسی، سینما و غیره به کار رفته است. وجود همین روش‌های بسیار متنوع و فراوان برای دو مساله بخش‌بندی و طبقه‌بندی در علوم مختلف است که بررسی تک تک آن‌ها را امکان‌پذیر نمی‌سازد و از این رو کوشش شده است که رهیافت‌های اساسی استفاده شوند.

- Industrial Engineering and Engineering Management, Singapor IEEE*, 16-20.
- Baines, P., Fill, C., & Page, K. (2010). *Market Segmentation and Posintioning Marketing, Second Edition* (pp. 227-244). USA: Oxford University Press.
- Blattberg, R. C., Kim, B., & Neslin, S. A. (2008). *Database marketing: analyzing and managing customers*. New York: Springer.
- Chiu, C. Y., Chen, Y. F., Kuo, I. T., & Ku, H. C. (2009). An intelligent market segmentation system using k-means and particle swarm optimization. *Expert Systems with Applications*, 36, 4558–4565.
- Chiu, T., Fang, D., Chen, j., Wang, y., & Jeris., C. (2001). *A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment*. Paper presented at the international conference on knowledge discovery and data mining.
- Dillon, W., Calantone, R., & Worthing, P. (1979). The new product problem: An approach for investigating product failures. *Management Science*, 25(12), 1184-1196.
- Freed, N., & Glover, F. (1981). Simple but powerful goal programming models for discriminant problems. *European Journal of Operational Research* 7, 44-66.
- Funatsu, K., & Hasegawa, K. (2011). *Knowledge-Oriented Applications in Data Mining*. Rijeka, Croatia: InTech.
- Gochet, W., Stam, A., Srinivasan, V., Chen., S. (1997). "Multi-group discriminant analysis using linear programming". *Operations Research*, 45(2), 213-225.
- Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and fuzzy delphi method – ADSL market of a telecommunication company *Expert Systems with Applications*, 38, 198-205.
- Jakkula, V. (2011). *Tutorial on Support Vector Machine (SVM)* Unpublished
- شود. به شرکت‌ها نیز پیشنهاد می‌گردد در لحظه ثبت نام مشتریان در سایت شرکت، با انجام تمهیدات تشویقی لازم تا جایی که امکان دارد و حریم شخصی افراد مورد خدشه قرار نگیرد، داده‌های مربوط به آن‌ها درخواست شود.
- ### منابع
- حسنقلی‌پور، طهمورث؛ میری، مهدی؛ مروتی شریف‌آبادی، علی (۱۳۸۶)، تقسیم بازار با استفاده از شبکه عصبی مصنوعی، مطالعه موردی فرآورده‌های گوشتی (سوسیس)، فصل‌نامه مدرس علوم انسانی، ویژه‌نامه مدیریت، ۵۹-۸۰
- شهرابی، جمال؛ نامدار علی‌آبادی، عباس؛ غلامی، رمضان (۱۳۸۹)، دسته‌بندی و شناسایی مشتریان هدف برای به کارگیری استراتژی بازاریابی مستقیم با یکپارچه‌سازی داده‌کاوی و منطق فازی، چهارمین کنفرانس داده‌کاوی ایران، دانشگاه صنعتی شریف، تهران، ایران
- عباسی، المیرا؛ صالح اولیاء، محمد؛ علیزاده، سمیه؛ خادمی زارع، حسن (۱۳۸۹)، استفاده از داده‌کاوی برای تعیین استراتژی‌های بازاریابی سیستم فروش اینترنتی بلیط قطار، چهارمین کنفرانس داده‌کاوی ایران، دانشگاه صنعتی شریف، تهران
- غضنفری، مهدی؛ ملک‌محمدی، سمیرا؛ علیزاده، سمیه؛ فتح‌الهه، مهدی (۱۳۸۹)، بخش‌بندی مشتریان در صادرات پوشاک بر پایه الگوریتم بخش‌بندی، فصل‌نامه پژوهشنامه بازرگانی، شماره ۵۶، ۵۹-۸۶
- Azadnia, A. H., Mat Saman, M.Z., Wong, K.Y and Hemdi, A.R, (2011), "Integration model of fuzzy C-Means clustering algorithm and TOPSIS method for customer lifetime value assessment",

- method for intelligent customer segmentation", *International Conference on Intelligent Systems, Modelling and Simulation, IEEE*, pp. 215-219.
- Odom, M., & Sharda, R. (1990). *A neural network model for bankruptcy prediction*. Paper presented at the Proceedings of the international joint conference on neural networks, Alamitos, CA.
- Pai, D. R., Lawrence, K. D., Klimberg, R. K., & Lawrence, S. M. (2012). Experimental comparison of parametric, non-parametric, and hybrid multigroup classification. *Expert Systems with Applications*, 39, 8593-8603.
- Petrović, S. (2006). *A comparison between the Silhouette index and the Davies-Bouldin index in labelling IDS clusters*. Paper presented at the 11th Nordic Workshop on Secure IT-systems, NORDSEC, Linkoping, Sweden.
- Ragsdale, C., Stam., A. (1991). Mathematical programming formulations for the discriminant problem. *Decision Sciences*, 22, 296-307.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*(20), 53-65.
- Shawe, J., & Cristianini, T. (2004). *Kernel Methods for Pattern Analysis*: Cambridge University Press.
- spss. (2013). from www.spss.com
- Youssef, S., & Rebai, A. (2007). Comparison between statistical approaches and linear programming for resolving classification problem. *International Mathematical Forum*, 63, 3125 - 3141.
- Zahavi, J., & Levin, N. (1997). Issues and problems in applying neural computing to target marketing. *Journal of Direct Marketing*, 11(4), 63-75.
- Ziberna, A., & Zabkar, V. (2003). Application of End-Users Market Segmentation using Statistical Methods. In A. Ferligoj, Mrvar, manuscript, School of EECS, Washington State University.
- Kaufmann, A. and Gupta M. (1991) "Introduction to fuzzy arithmetic". Amsterdam: Van Nostrand Reinhold Company Inc.
- Kiang, M. Y., Hu, M. Y., & Fisher, D. M. (2006). An extended self-organizing map network for market segmentation—a telecommunication example. *Decision Support Systems* 42, 36-47.
- Klecka, W. (1980). *Discriminant Analysis*: Sage.
- Kotsiantis, S., & Pintelas, P. (2004). Recent Advances in Clustering: A Brief Survey. *WSEAS Transactions on Information Science and Applications*, 1, 73--81.
- Kwak, N. K., Kim, S.H., Lee, C.W. and Choi, T.S (2002). "An application of linear programming discriminant analysis to classifying and predicting the symptomatic status of HIV/AIDS Patients." *Journal of Medical Systems* 26(5): 427-438.
- Lam, K., Choo, E., & Moy, J. (1996). Improved Linear Programming Formulations for the Multi-Group Discriminant Problem. *Journal of the Operational Research Society*, 47(12), 1526-1529.
- Liu, Y., Kiang, M., & Brusco, M. (2012). A unified framework for market segmentation and its applications. *Expert Systems with Applications*, Article in Press.
- Ma, B., Li, F., Wang, G. and Ran, L., (2009), "Targeting valuable customers within a retail reward program database by RFMG Model", *International Conference on Intelligent Systems, Modelling and Simulation, IEEE*, Beijing
- Mangasarian, O. (1965). Linear and nonlinear separation of patterns by linear programming. *Journal of Operations Research*, 13, 444-452.
- Namvar, M., Gholamian, M. and KhakAbi, S., (2010), "A two phase clustering

به عنوان مثال در مورد درآمد، ۱ نفر اهمیت آن را خیلی کم، ۱ نفر آن را کم، ۲ نفر آن را متوسط، ۴ نفر آن را زیاد و ۲ نفر آن را خیلی زیاد ارزیابی کرده‌اند. در مرحله بعد نظرات کارشناسان در قالب متغیرهای کلامی با توجه به مقیاس نشان داده شده در شکل (۳) و جدول (۱۶) به اعداد فازی مثلثی تبدیل شد.



شکل ۳- مقیاس متغیرهای کلامی

(کافمن و گوپتا، ۱۹۹۱)

جدول ۱۶- اعداد فازی مثلثی متناظر با متغیرهای کلامی

(کافمن و گوپتا، ۱۹۹۱)

اعداد فازی مثلثی متغیرهای کلامی	
متغیر کلامی	عدد فازی مثلثی
خیلی زیاد	(۰/۰۰، ۰/۲۵، ۱/۰۰)
زیاد	(۰/۱۵، ۰/۱۵، ۰/۷۵)
متوسط	(۰/۲۵، ۰/۲۵، ۰/۵۰)
کم	(۰/۱۵، ۰/۱۵، ۰/۲۵)
خیلی کم	(۰/۲۵، ۰/۰۰، ۰/۰۰)

A (Ed.), *Developments in Applied Statistics* (pp. 243- 263). Ljubljana: FDV.

Zucker, R., & Kietz, J. U. (2000). *How to preprocess large databases. In Data Mining, Decision Support, Meta-learning and ILP*. Lyon, France: Forum for Practical Problem Presentation and Prospective Solutions.

ضمیمه (۱): انتخاب و تعیین متغیرهای مورد نیاز برای

طبقه‌بندی مشتریان از طریق دلفی فازی

در ابتدا با بررسی ادبیات پژوهش، هشت ویژگی سن، جنس، شغل، تحصیلات، درآمد، هدف از استفاده از اینترنت، محل سکونت و سبک زندگی به عنوان متغیرهایی که می‌توانند به عنوان پیش‌بینی‌کننده^۵ انتخاب شوند، تهیه شد. سپس جدولی به صورت جدول (۱۵) تهیه و از ۱۰ نفر از دست‌اندرکاران شرکت ایران‌گیت خواسته شد تا نظر خود را در مورد اهمیت ویژگی مشتریان بر توصیف آن‌ها به صورت خیلی کم، کم، متوسط، زیاد و خیلی زیاد بیان کنند. جدول (۱۵) در ضمن بیانگر تعداد نظرات کارکنان شرکت نیز هست.

جدول ۱۵- نظرات ۱۰ نفر از کارکنان شرکت ایران‌گیت

در مورد اهمیت هر ویژگی در توصیف مشتریان

عامل	میزان موافقت			
	خیلی کم	خیلی زیاد	متوسط	کم
۱ محل سکونت	۳	۰	۲	۵
۲ سبک زندگی	۴	۰	۲	۴
۳ تحصیلات	۰	۴	۱	۵
۴ درآمد	۱	۴	۲	۲
۵ هدف	۰	۲	۲	۵
۶ سن	۳	۲	۱	۳
۷ جنس	۲	۲	۳	۳
۸ شغل	۳	۱	۲	۴

ضمیمه (۲): مدل برنامه‌ریزی خطی چند گروهی برای طبقه‌بندی

اگر تعداد کل مشتریان n باشد که در m گروه از قبل خوشه‌بندی شده‌اند، بنابراین $n = n_1 + n_2 + \dots + n_m$ که n_1 تعداد مشاهدات در گروه اول، n_2 تعداد مشاهدات در گروه دوم و الی آخر است. اگر G_k گروه k ام با n_k مشتری باشد، $k=1, \dots, m$ ، آنگاه میانگین متغیر j ام در گروه k ام به صورت زیر می‌باشد (لام و موی، ۱۹۹۶):

$$\bar{x}_j(k) = \sum_{i \in G_k} x_{ij} / n_k \quad (۴)$$

به علاوه q را تعداد متغیر یا ویژگی دموگرافیکی مشتریان فرض می‌کنیم که قبلاً از طریق روش تحلیل سلسله مراتب فازی و کسب نظر خبرگان تعیین شده است. از آنجا که لازم است هر جفت مشتری را دو به دو مقایسه کرد، برای هر جفت مشتری (u, v) که $u=1, \dots, m-1$ و $v=u+1, \dots, m$ مدل کمینه‌سازی انحرافات مرحله اول به صورت زیر نوشته می‌شود:

$$\begin{aligned} \min \quad & \sum_{i \in G_u, i \in G_v} d_i \\ \text{St} : \quad & \sum_{j=1}^q w_j (x_{ij} - \bar{x}_j(u)) + d_i \geq 0, \quad \forall i \in G_u \\ & \sum_{j=1}^q w_j (x_{ij} - \bar{x}_j(v)) - d_i \leq 0, \quad \forall i \in G_v \\ & \sum_{j=1}^q w_j (\bar{x}_j(u) - \bar{x}_j(v)) \geq 1, \\ & d_i \geq 0 \end{aligned} \quad (۵)$$

که در آن w_j ($j=1, \dots, q$) متغیر تصمیم (آزاد در علامت) وزن یا سهم ویژگی j ام مشتریان در تفکیک گروه‌ها از یکدیگر، d_i متغیر نامنفی برای $i \in G_u \cup G_v$ میزان انحراف یک مشاهده تکی از نمره برش یا جدایی و ضرایب x_{ij} مقدار ویژگی j ام برای مشتری i ام است. تابع هدف مجموع تمام انحرافات مشاهدات را کمینه می‌سازد که همانند مدل کمینه‌سازی مجموع

جدول ۱۷- نتایج محاسبات دلفی فازی

عامل	میانگین (m, α, β)			میانگین
	m	α	β	
۱ سن	۰/۲۰۰	۰/۱۵۰	۰/۳۷۵	۰/۳۸۷
۲ جنس	۰/۱۵۰	۰/۲۰۰	۰/۶۲۵	۰/۶۱۲
۳ تحصیلات	۰/۰۸۵	۰/۲۱۰	۰/۸۵۰	۰/۸۱۹
۴ درآمد	۰/۱۵۰	۰/۱۷۵	۰/۶۲۵	۰/۶۱۹
۵ هدف	۰/۰۹۵	۰/۲۲۰	۰/۷۷۵	۰/۷۴۴
۶ محل	۰/۱۸۰	۰/۱۰۵	۰/۱۷۵	۰/۱۹۴
۷ سبک	۰/۲۰۰	۰/۰۶۰	۰/۱۰۰	۰/۱۳۵
۸ شغل	۰/۱۲۵	۰/۱۲۵	۰/۲۷۵	۰/۲۷۵

اینک اگر نظر خبره i ام در مورد یک ویژگی با رابطه (۱) نشان داده شود، میانگین فازی نظر خبرگان در مورد آن ویژگی به صورت رابطه (۲) خواهد بود. به علاوه برای به دست آوردن میانگین فازی‌زدایی شده از رابطه (۳) که به رابطه مینکوسکی معروف است، استفاده می‌شود.

$$A_i = (m_i, \alpha_i, \beta_i) \quad (۱)$$

$$A_{ave} = \left(\frac{1}{n} \sum_{i=1}^n m_i, \frac{1}{n} \sum_{i=1}^n \alpha_i, \frac{1}{n} \sum_{i=1}^n \beta_i \right) \quad (۲)$$

$$\chi = m + \frac{\beta - \alpha}{4} \quad (۳)$$

نتایج محاسبات در جدول (۱۷) نشان داده شده است. مطابق با این نتایج به ترتیب ویژگی‌های تحصیلات، هدف، درآمد، جنس، سن، شغل، محل سکونت و سبک زندگی قرار دارند. از بین این ویژگی‌ها، پنج ویژگی اول که دارای بالاترین اهمیت هستند انتخاب و از طریق مصاحبه تلفنی داده‌های ۱۸۰۸ مشتری از ۵۲۷۱ مشتری جمع‌آوری گردید. به این ترتیب با در اختیار داشتن داده‌های لازم، امکان تهیه الگوی طبقه‌بندی فراهم می‌گردد.

$$\min \sum_{i \in G_u, i \in G_v}^n d_i \quad (۸)$$

St :

$$\sum_{j=1}^q w_j (x_{ij} - \text{median}_j(u)) + d_i \geq 0, \quad \forall i \in G_u$$

$$\sum_{j=1}^q w_j (x_{ij} - \text{median}_j(v)) - d_i \leq 0, \quad \forall i \in G_v$$

$$\sum_{j=1}^q w_j (\bar{x}_j(u) - \text{median}_j(k)) \geq 1,$$

پی نوشت

- 1 - Ziberna and Zabkar
- 2 - Funatsu and Hasegawa
- 3 - customer churn
- 4 - Liu et al
- 5 - Ma et al
- 6 - Mangasarian
- 7 - Freed and Glover
- 8 - Ragsdale and Stam
- 9 - Lam and Moy
- 10 - Gochet
- 11 - Yousef and Rebai
- 12 - Kwak
- 13 - Preprocessing
- 14 - Zucker
- 15 - Recency- Frequency- Monetary (RFM)
- 16 - Analytical Hierarchical Process
- 17 - K-means
- 18 - Self-Organizing-Map (SOM)
- 19 - Two Step Clustering
- 20 - Cluster ensemble
- 21 - Kotsiantis and Pintelas
- 22 - Kohonen
- 23 - Blattberg
- 24 - SPSS
- 25 - Petrović
- 26 - Rousseeuw
- 27 - Pai
- 28 - Yousef and Rebai
- 29 - Dillon
- 30 - Fisher's Linear Discriminant Analysis (FLDA)
- 31 - Kiang et al
- 32 - Lam and Moy
- 33 - Hit ratio
- 34 - Confusion matrix
- 35 - Stepwise
- 36 - Wilks' lambda

انحرافات^{۵۵} است. محدودیت‌های اول و دوم با کمینه‌سازی انحراف d_i (متعلق به گروه u یا v) باعث می‌شوند تا نمرات طبقه‌بندی مشتریان در گروه k تا حد ممکن به میانگین نمره طبقه‌بندی گروه نزدیک شوند. محدودیت سوم که محدودیت نرمالیزه نامیده می‌شود برای پرهیز از ارائه مقادیر بدیهی (صفر) به وزن‌های تمایزی مشتریان است. اینک در مرحله دوم با در اختیار داشتن وزن (w_j) ویژگی‌های جفت گروه (u, v) در تفکیک گروه‌ها، مقادیر جدا کننده یا برش مشاهدات (C_{uv}) با حل برنامه‌ریزی خطی زیر به دست می‌آید:

(۶)

$$\min \left(\sum_{i \in G_u} \sum_{u=1}^{m-1} \sum_{v=u+1}^m d_{iuv} + \sum_{i \in G_v} \sum_{v=1}^{m-1} \sum_{u=v+1}^m d_{iuv} \right)$$

St :

$$S_{iuv} + d_{iuv} \geq C_{uv}, \text{ for } u=1, \dots, m-1, v=u+1, \dots, m, i \in G_u$$

$$S_{iuv} - d_{iuv} \leq C_{uv}, \text{ for } u=1, \dots, m-1, v=u+1, \dots, m, i \in G_v$$

که تمام نمره‌های برش C_{uv} آزاد در علامتد و همچنین $d_{iuv} \geq 0$. بنابراین، نمره‌های طبقه‌بندی (S_i) از رابطه زیر به دست می‌آید:

$$S_i = \sum_{j=1}^q w_j x_{ij}, \text{ for } i \in G_k \quad (۷)$$

با داشتن نمره‌های طبقه‌بندی هر مشتری و مقادیر نمره‌های برش گروه‌ها، می‌توان به راحتی هر یک را به گروه‌های از پیش تعیین شده متناسب ساخت. به علاوه می‌توان هر مشتری جدیدی که در بانک اطلاعاتی شرکت دارای تراکنش نیست را نیز بر پایه ویژگی‌های دموگرافیکی‌اش به یکی از گروه‌ها نسبت داد. یک توسعه دیگر مدل لام و موی (۱۹۹۶)، استفاده از میانه به جای میانگین داده‌ها است. زیرا ثابت شده است که برای توزیع‌های غیرنرمال، میانه آماره توصیفی مناسب‌تری نسبت به میانگین است (پای و همکاران، ۲۰۱۲).. از این رو در این پژوهش نیز به جای میانگین از میانه استفاده می‌شود که در مدل (۸) نشان داده شده است.

- 37-Klecka
- 38 - multinomial logistic regression (MLR)
- 39 - Odom & Sharda
- 40 - back propagation
- 41 - Zahavi and Levin
- 42 - Shawe
- 43 - Radial Basis Function
- 44 - Jakkula
- 45 - Mangasarian
- 46 - Freed and Glover
- 47 - Shapiro Wilk
- 48 - Dynamic
- 49 - Multiple
- 50 - Exhaustive prune
- 51 - Radial basis function network
- 52 - Exhaustive prune
- 53 - Over-fitting
- 54 - Predictor
- 55 - Minimization of the Sum of Deviation