پژوهشکده پولی و بانکی
بانک مرکزی جمهوری اسلامی ایران
بانک مرکزی جمهوری اسلامی ایران
همایش ملی صنعت لیزینگ و تأمین مالی
— چشم‌انداز صنعت لیزینگ و تنظیم‌گری —
دانشگاه الزهرا انجمن ملی لیزینگ ایران دانشکدگان مدیریت
تهران، بزرگراه حقانی، مرکز همایش‌های بانک مرکزی - یکشنبه ۲۸ بهمن ۱۴۰۳

پیش‌بینی ریسک مشتریان در صنعت لیزینگ با استفاده از تکنیک‌های داده‌کاوی در راستای خوشه بندی

عبدالحسین کرمپور[۱] عبدالحسین زیارتی[۲] سارا سهرابی مطلق فاضلی[۳]

تاریخ ارسال: ۱۴۰۳/۱۱/۰۹

چکیده:

هدف اصلی هدایت پژوهش، ارائه روشی جامع برای طبقه بندی مشتریان بر اساس اعتبار آنها در چارچوب کشورهای در حال توسعه بود. در این تحقیق، جهت بخش بندی صنعت لیزینگ بر اساس ارزش اعتباری مشتری از مدل آر. اف.ام .و برخی تکنیک های داده کاوی و روش های مختلف علمی در قالب فرایندی خاص، بهره گرفته شد . استفاده از جدول تحلیل واریانس جهت تعیین میزان اثرگذاری هر یک از شاخص ها در تفکیک خوشه ها، ارائه تحلیل درون خوشه ای به منظور بررسی دقیق تر ویژگی های مشتریان در خوشه دارای بیشترین مشتری و همچنین بخش بندی نهایی مشتریان بر اساس ارزش اعتباری مشتریان در قالب جدول ارزش اعتبار مشتریان از موارد خاصی است که در این مطالعه از آنها بهره گرفته شده است، درحالی که در مطالعات گذشته به آنها اشاره ای نگردیده است. روش پیشنهادی با ارائه یک مطالعه موردی نشان داد که از پایایی بالایی برخوردار است. همچنین که در مقایسه تطبیقی با دیگر زوش های درختی و خوشه بندی از صحت و دقت بیشتری برخوردار بود. استفاده از سیستم RFM هنوز هم در فرآیندهای طبقه بندی بصورت قاطع کاربرد دارد و می تواند با دیگر روشها بصورت ترکیبی مدلی قویتر را ارائه نماید.

واژه‌های کلیدی: پیش‌بینی ریسک مشتریان، تکنیک‌های داده‌کاوی، خوشه بندی، صنعت لیزینگ.

طبقه‌بندی JEL: بازارهای مالی، سرمایه گذاری مالی، حکمرانی و تامین مالی شرکتی (o16).

[۱] استادیار و عضو هیأت علمی گروه مدیریت بازرگانی دانشکده مدیریت دانشگاه آزاد اسلامی واحد تهران شمال (a.karampour@iau-tnb.ac.ir)

[۲] کارشناس ارشد مدیریت بازرگانی_بازاریابی دانشکدگان فارابی دانشگاه تهران (ah.ziyarati@gmail.com)

[۳] کارشناسی ارشد مدیریت کسب و کار_بازاریابی دانشگاه آزاد اسلامی واحد تهران شمال (sara.sohrab1981@gmail.com)

۱. مقدمه

صورت یک مساله غالباً اساسی تر از حل آن است. حل مساله ممکن است فقط مستلزم مهارت های تجربی و ریاضی باشد، حال آنکه طرح سوالات نو و بررسی مسائل قدیمی از دیدگاهی تازه، نیازمند ذهنی خلاق است و این نکته بیانگر آزمودگی فرد در علم است. هر پژوهش در واقع با قصد پاسخگویی و راه حل یابی برای یک مساله اصلی در قالب یک پرسش آغاز می شود. بدین لحاظ برای اینکه انسجام، هدفمندی و کاربردی بودن پژوهش حفظ شود، بایستی حول مساله اصلی سازماندهی شود (زارعی نژاد، ۱۳۹۲). با در نظر گرفتن مبانی و اهداف خاص تحقیق، سوال اصلی تحقیق بدین صورت مطرح می گردد:

چگونه می توان با استفاده از مدل های داده کاوی یک الگوی بهینه در زمینه طبقه بندی مشتریان بر اساس ریسک اعتباری در صنعت لیزینگ طراحی نمود؟

۲. بدنه اصلی

همانطور که در ابتدا اشاره شد فرآیند شناسایی مشتریان هدف در شرایط رقابتی دارای اهمیت زیادی است. دسته بندی مشتریان بر اساس تکنیک های داده‌کاوی جهت شناسایی مشتریان هدف به کار می رود. هدف اصلی این تحقیق بررسی روش‌های مختلف داده‌کاوی در خوشه‌بندی مشتریان است. درخت تصمیم، رگرسیون لجستیک، شبکه عصبی، نزدیک ترین همسایگی و ...، از جمله تکنیک های داده کاوی هستند که در این رساله مورد تحقیق و بررسی قرار خواهند گرفت. اهداف اختصاصی تحقیق عبارتند از:

هدف اول پژوهش، بررسی و تجزیه و تحلیل مسائل و چالش های درگیر در طبقه بندی مشتریان در کشورهای در حال توسعه است.

هدف دوم بررسی و تجزیه و تحلیل مناسب و همچنین تلفیق روش های داده کاوی، یعنی، RFM و شبکه عصبی که می تواند برای طبقه بندی مشتریان بر اساس ریسک اعتباری استفاده شود.

هدف سوم تدوین و فرموله کردن روش های داده کاوی برای طبقه بندی مشتریان بالقوه بر اساس ریسک اعتباری در چارچوب مباحث نظری.

هدف چهارم. مقایسه روش های داده کاوی ترکیبی با رگرسیون لجستیک.

۴-۳. سؤالات تحقیق

سوال اصلی تحقیق:
چگونه می‌توان در یک شرکت یا سازمان، مشتریان را به لحاظ ریسک اقتصادی با استفاده از الگوریتم های داده کاوی طبقه‌بندی کرد؟

سؤالات فرعی

۱. کدام الگوریتم خوشه بندی صحیح تری را نسبت به سایر الگوریتم های خوشه بندی داده کاوی ارائه می‌کند؟

۲. کدام شاخص مهمترین شاخص در خوشه بندی مشتریان می باشد؟

۳. کدام مدل الگوریتم داده کاوی نسبت به سایر الگوریتم ها از صحت بالاتری برخوردار میباشد؟

۴. آیا بهینه سازی خوشه بندی بر مبنای الگوریتم های هوشمند (ژنتیک، ازدحام ذرات، جستجوی فاخته، رقابت استعماری و...،) از صحت بالاتری برخوردار نیست؟

۳-۵. روش تحقیق

۳-۵-۱ مدل مفهومی پژوهش

اندازه گیری ارزش دوره عمر مشتری (CLV) یکی از ابزارها برای خوشه بندی مشتریان است که به کمک آن می توان با هدف‌گیری مشتریان سودآور و حتی حذف برخی از مشتریان و یا کاهش فعالیت‌های بازاریابی برای بخشی از آنها، سودآوری را تضمین نمود. در این تحقیق از الگوی RFM٤ بولت (۲۰۱۴)، که شامل سه متغیر اصلی تازگی خرید (R)، فراوانی خرید (F) و ارزش مالی خریدهای مشتری (M) است استفاده شده است و اهمیت نسبی سه متغیر مذکور به عنوان دیگر مقادیر لازم با استفاده از روش فرآیند تحلیل سلسله مراتبی AHP تعیین خواهد شد و اهمیت هر یک از این شاخص‌ها در مدل WRFM جهت خوشه بندی و رتبه بندی مشتریان به کار برده خواهد شد. همچنین روش شبکه عصبی نیز در این تحقیق بکار برده می شود.

معمولا تحقیقات طبق معیارهای خاص در طبقه بندی های متفاوتی جای می گیرند. تحقیقات علمی بر اساس هدف تحقیق به سه دسته بنیادی، کاربردی و تحقیق و توسعه تقسیم می شود. همچنین تحقیقات علمی را بر اساس چگونگی به دست آوردن داده های مورد نیاز (طرح تحقیق) می توان به دو دسته تحقیق توصیفی و تحقیق اکتشافی تقسیم نمود.به منظور برقراری رابطه علت-معلولی میان دو یا چند متغیر از طرح های آزمایشی استفاده می شود.

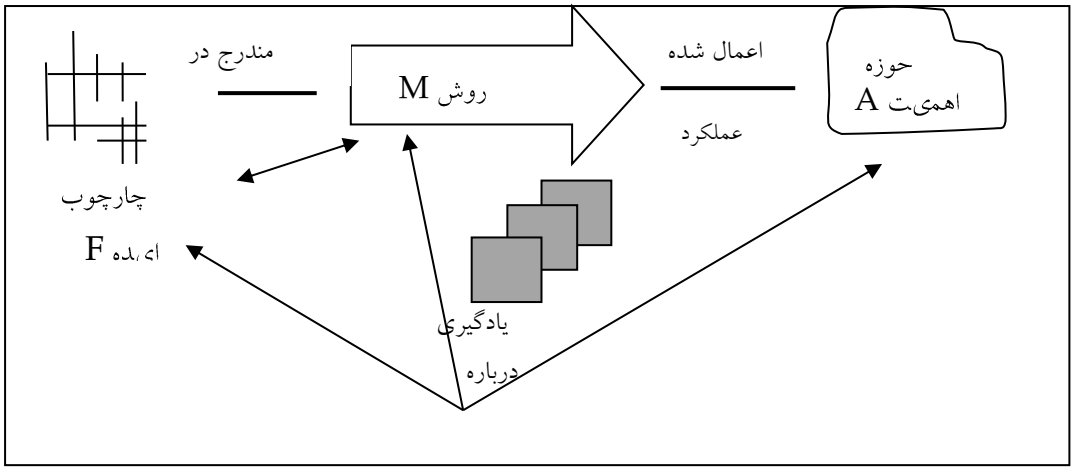
تحقیق توصیفی شامل مجموعه روش هایی است که هدف آنها توصیف کردن شرایط یا پدیده های مورد بررسی است. اجرای تحقیقات توصیفی می تواند صرفا برای شناخت بیشتر شرایط موجود یا یاری دادن به فرآیند تصمیم گیری باشد. تحقیق توصیفی شامل مجموعه روش هایی است که هدف آنها توصیف کردن شرایط یا پدیده های مورد بررسی است. تحقیقات توصیفی خود به انواع : پیمایشی، همبستگی، اقدام پژوهشی، بررسی موردی و پس رویدادی (علی-مقایسه ای) تقسیم می شود.

روش تحقیق حاضر از نظر هدف، کاربردی و از نظر نوع جمع آوری اطلاعات و داده ها، تحقیقی اسنادی است. همچنین برای شناسایی متغیرهای تأثیرگذار در رفتار مشتریان از مصاحبهٔ آزاد با خبرگان صنعت مورد مطالعه استفاده شده است . نوع داده های این پژوهش به صورت عددی(پیوسته)و اسمی(گسسته)است.

بر اساس نظر چکلند و هولول (۱۹۹۸) ، تمام قسمتهای تحقیق بر طبق مدل ساده ای از اجزاء نشان داده شده در شکل ۳-۱ است . اجزاء ضروری این مدل شامل یک چارچوب هوشمندانه ای از ایده های مرتبط است (F)، که خود یک تئوری است که حاوی اطلاعاتی در مورد مسأله " سؤال " تحقیق است (A). این چهارچوب در شیوه ای

1. Recency Frequency Monetary

قرار می گیرد که استفاده از روشهای متنوع تحقیقی(M) و تکنیکهای مناسب این چهارچوب جهت بازرسی نواحی کاربردی را ازتقاء می بخشد . این همان سؤال تحقیق است.



شکل ۱ . عناصر مربوط به هر بخش از پژوهش (چکلند و هالول، ۱۹۹۸)

براساس تجزیه و تحلیل علمی در مورد ریسک اعتباری مشتریان نیاز به یک روش داده کاوی برای حل مسأله، حتمی و ضروری بنظر می‌رسید. بنابراین با مشورت کارشناسان حوزه ارتباط با مشتری و ریسک اعتباری ، به طور تدریجی چهار چوبی از نظرات (F) شکل گرفت و توسعه یافت. در این پژوهش از روش های داده کاوی موسوم به RFM و SOM برای خوشه بندی مشتریان استفاده شد.

۳-۵-۳. مورد کاوی

همان گونه که قبلا نیز ذکر شد ، تمرکز این تحقیق بر روی درک بهتر فرایند های طبقه بندی مشتریان با هدف تقویت کارکرد های کنونی است. مرحله ی بعد شامل آزمایش مدل تلفیقی و پیشنهادی داده کاوی بصورت کلی در یک موقعیت واقعی است، که مستلزم یک روش مورد کاوی است. بر طبق نظر زارعی نژاد (۱۳۹۲) استراتژی های مورد کاوی در بسیاری از موقعیت‌ها برای کمک به اطلاعات شخصی، گروهی، سازمانی، اجتماعی، سیاسی و مسائل مربوطه استفاده می شود. وی گفته است که مورد کاوی شبیه به یک آزمایش است و بسیاری از شرایط که یک آزمایش را تصدیق می کند، این تحقیق را نیز تصدیق می کند. ایشان استفاده از یک تحقیق مورد کاوی را تنها به پنج دلیل ذکر شده در پایین توجیه و تایید کرد که عبارتند از:

۱- زمانی که این تحقیق یک مورد انتقادی را در آزمایش یک تئوری بدون مشکل نشان دهد .

۲- زمانی که این تحقیق یک مورد افراطی و یا یک مورد منحصر به فرد را نشان دهد .

۳- زمانی که مورد مطالعه ای یک مورد شاخص و یا معرف باشد ، و هدفش این باشد که موقعیت ها و شرایط عادی را بدست آورد .

۴- زمانی که مورد تکاملی است، و محقق دارای فرصتی است که پدیده ای را که قبلا جهت تحقیق علمی در دسترس نبوده مشاهده و تجزیه و تحلیل کند.

۵- زمانی که مورد وابسته به جغرافیا باشد، تا بتوان در دو یا چند نقطه ی مختلف آن را بررسی کرد .

در این مرحله صنعت لیزینگ به عنوان یک تحقیق مورد کاوی انتخاب شد تا اجرای عملی مدل داده کاوی پیشنهادی کلی را برای طبقه بندی مشتریان و پیش بینی ریسک آنها را توصیف کند که در آن فرآیند طبقه بندی بصورت یک فعالیت چالشی نمودار خواهد شد. تحقیق مورد کاوی برای طراحی، ترتیب، تحقیق، تایید واقعیت، مصاحبه اداری و کارشناسان زمان زیادی را صرف کرد. برای تحقیق این مسئله لازم بود با طراحان حوزه مدیریت ارتباط با مشتری انتخاب شده که معمولا در طرح‌های انتخابی مدیریت ریسک ماهرترند، تعداد زیادی جلسات مذاکره ای برگزار شود در نهایت این امر کارگاهی را می طلبد، که بواسطه ی آن طراحان مدیریت ارتباط با مشتری قادر به مشارکت باشند و ارتباط و تعامل پیوسته و منسجمی را با نویسنده که به عنوان یک تسهیل کننده عمل میکند، بوجود آورد.

۴-۵-۳. مدل RFM

مدل آر.اف.ام .اولین بار توسط هوگس (۱۹۹۴) معرفی گردید . وی برای تحلیل آر.اف.ام .از رفتار گذشته مشتری که به آسانی قابل پیگیری و دسترسی است، استفاده نمود. این مدل از سه بعد مربوط به داده های مبادلاتی مشتریان، برای تحلیل رفتار آنها استفاده مینماید یعنی تاخیر، فرکانس و پول. مدل RFM یک مدل مبتنی بر رفتار است که برای آنالیز رفتار یک مشتری و سپس پیش بینی او بر اساس رفتارش در بانک اطلاعات استفاده می‌شود(یه و همکاران ۲۰۰۹). در بین متغیرهای RFM شاخص تاخر اغلب مهمترین متغیر شناسایی می شود. با این حال بر اساس مطالعات انجام شده در گذشته متغیرهای RFM در واقع Firm-specific هستند و بر اساس طبیعت محصولات شرکت اهمیت آنها متفاوت است (لومسدن ۲۰۰۸). در این الگو r فاصله زمانی آخرین خرید مشتری تا زمان حال، F تعداد خریدها در یک دوره زمانی مشخص و M مبلغ اسمی خریداری شده در دوره مدنظر تعریف می شود (بولت ۲۰۱۴). این پژوهش از مدل RFM برای تبدیل داده های اولیه به فرم دلخواه برای استفاده در الگوریتم خوشه بندی استفاده کرده است در نتیجه تغییراتی در نحوه محاسبه آنها صورت می گیرد. شاخص تاخیر: تازگی خرید (R)، و ارزش مالی خریدهای مشتری ، شاخص فراوانی خرید (F): تعداد ماههایی که که مشتری تراکنش خرید بیشتری دارد. ارزش مالی خریدهای مشتری: مجموع گردش مشتری را در طول این دوره زمانی در نظر گرفتیم .

۵-۵-۳. روش SOM

برای امتیاز اعتباری و تجزیه و تحلیل امتیازی رفتاری،بسیاری از مطالعات ارائه شده اند که شبکه های عصبی اصلی ترین]نها است که بنظر میرسد بهتر از روش های آماری باشد (هسی ۲۰۰۴). الگوریتم SOM خوشه بندی از نوع شبکه عصبی است که در سال ۱۹۸۱ توسط پژوهشگر فنلاندی کوهنن اختراع شد. این الگوریتم بطور معمول متشکل از دو لایه نورون های ورودی و خروجی است (حسینی، ۲۰۱۳). بطور کلی شبکه عصبی از لایه های نورونی تشکیل شده اند. این نوع نورون ها از طریق ورودی های خود با جهان واقعی در ارتباطند و گروه دیگر از نورون ها نیز از طریق حروجی های خود، جهان بیرون را میسازند (تیسای۲۰۰۹).

برای تعیین بهترین تعداد خوشه ها از روشی بنام SSE استفاده میشود. در این روش نخست مراکز خوشه درنظر گرفته می شود و سپس فاصله نقطه مورد نظر از مراکز خوشه محاسبه میشود. خوشه ای که SSE پایین تری دارد نشان دهند بهترین خوشه بندی است برای این امر از فرمول زیر استفاده میشود:

$$SSE = \sum_{i=1}^{k} \sum_{p \in c} d(P, m_i)^2$$

۶-۵-۳. روش امتیازدهی روش پیشنهادی

ارزش هر مشتری را میتوان بر اساس تازگی، تکرار و ارزش پول بصورت زیر محاسبه نمود (غضنفری ۲۰۱۱):

$$V(c_i) = W^R \times R(c_i) + W^F \times F(c_i) + W^M \times M(c_i)$$

که در آن R(ci)، F(ci) و M(ci) به ترتیب امتیازات مشتری cام با توجه به معیارهای RFM هستند. WR ، WF و WM اهمیت وزنی برای معیارهای RFM را نشان میدهند.

$$WF + WR + WM = 1$$

سودآوری خوشه On با محاسبه میانگین ارزش همه مشتری های خوشه nام حاصل می شود. از این رو میتوان آن را از طریق معادله زیر تعریف نمود (غضنفری ۲۰۱۱).

$$V(O^n) = W^R \times R(O^n) + W^F \times F(O^n) + W^M \times M(O^n)$$

$$R(O^n) = \sum_{C_i \in O^n} R(C_i) \Big/ \|O^n\|$$

$$F(O^n) = \sum_{C_i \in O^n} F(C_i) \Big/ \|O^n\|$$

$$M(O^n) = \sum_{C_i \in O^n} M(C_i) \Big/ \|O^n\|$$

که در آن $R(O^n), F(O^n), M(O^n)$ امتیاز nامین خوشه با توجه به معیارهای RFM است.

۷-۵-۳. مدل رگرسیون لجستیک

روش رگرسیون لجستیک زمانی که متغیر وابسته بوده و دارای دو حالت است مورد استفاده قرار میگیرد. این روش همچنین آزمونهای آماری مختلفی را ارائه داده و توانایی های تشخیص گسترده ای را در خود دارد. یکی از منافع رگرسیون لجستیک مقادیر احتمال برای اختمال وقوع رخدادهای متغیر وابسته است که نمودی عینی به نتایج مدل میدهد. در صورتی که احتمال وقوع بیشتر از حد آستانه که معمولا ۰/۵ در نظر گرفته میشود، پیش بینی شود. در این صورت وقوع پدیده مورد نظر محتمل تلقی شده و در غیر اینصورت غیرمحتمل خواهد بود.

نوعی از مدل رگرسیون است که متغیرهای پیش بین (مستقل) هم در مقیاس کمی و هم در مقیاس مقوله ای می تواند باشد و متغیر وابسته، مقوله ای دو سطحی است. این دو مقوله به گونه ای معمول به عضویت یا عدم عضویت در یک گروه اشاره دارد.

برای توضیح مدل رگرسیون می توان از تابع توزیع تجمعی استفاده نمود. توابع توزیع تجمعی از تغییراتی که مقدار P را در فاصله صفر و یک قرار میدهد به وجود می آید در حالیکه همچنان خواص یکنواختی دارد. فرض کنید که یک توزیع نرمال استاندارد برای بیان احتمال انتخاب شده است (کوسمنت و همکاران، ۲۰۱۴).

$$P(y/x) = \Phi(b'x) = \int_{-\infty}^{b'x} \phi(z)dz$$

که در آن X بردار متغیر توضیحی، P احتمال تجمعی و قوع پیشآمد، b بردار ضریب که در آن $\Phi(Z)$ تابع چگالی نرمال استاندارد است.

این تابع معروف مدل پروبیت است. اگر تابع توزیع لجستیک برای بیان احتمال تجمعی وقوع پیشامد P مورد استفاده قرار گیرد به مدل لجستیک منجر می شود (کوسنت و همکاران ۲۰۱۴). آنگاه:

$$p(y/x) = \phi(b'x) = \int_{-\infty}^{\beta'x} \eta(z)dz = \frac{1}{1+e^{-\beta'x}} = \frac{e^{\beta'x}}{1+e^{\beta'x}}$$

$$1 - p_i = \frac{1}{1+e^{\beta'X}}$$

که در آن $\eta(Z)$ تابع چگالی لجستیک یا لوجیت است. در مدلهای لوجیت و پروبیت متغیر وابسته بصورت صفرویک تعریف می شود. لگاریتم نسبت بخت یا لوجیت از رابطه زیر حاصل می شود.

$$\mathbf{L = \ln\left(\frac{P}{1-P}\right) = B_0 + B_1X_1 + \cdots + B_nX_n}$$

در رگرسیون لجستیک متغیر وابسته یک متغیر صفر و یک است که دو مقدار را به خود اختصاص می دهد. اگر فرض کنیم Y متغیر تصادفی بیاشد که میتواند مقادیر صفر ویک را بخود اختصاص دهد در این صورت احتمال وقوع Y را می توان بشکل رابطه زیر در نظر گرفت:

$$\mathbf{P(Y=1) = P = \frac{e^{B'X}}{1+e^{B'X}} P(Y=0) = (1-P) = \frac{1}{1+e^{B'X}}}$$

که در آن B بردار سطری ضرایب و X بردار ستونی متغیرهای مستقل است.

در حالت کلی برآورد ضرایب متغیرهای مستقل یعنی بردار B از طریق حداکثرسازی رابطه زیر بدست می آید که توسط مشتق گیری نسبت به هریک از ضرایب متغیرهای مستقل و مساوی صفر قرار دادن هر یک از مشتق ها محاسبه می شود.

$$\ln L = l = \sum_{i=1}^{n} y_i \ln\left(\frac{e^{\beta' x}}{1+e^{\beta' x}}\right) + \sum_{i=1}^{n} (1-y_i) \ln\left(\frac{1}{1+e^{\beta' x}}\right)$$

۸-۵-۳. فرآیند تحلیل سلسله مراتبی

روش AHP را توماس ساتی توسعه داد که او کار پیشروی اش را در اوایل دهه ۷۰ آغاز کرد . این یک شیوه تصمیم گیری چند معیاره است که یک مسئله پیچیده را به یک سلسه مراتب که متشکل از عناصر خاص است ، تجزیه می کند . این ابزاری موثر است که هم می تواند ویژگی های قابل لمس و هم ویژگی های نامرئی (غیر قابل لمس) را اداره کند ، خصوصا وقتی در حال بررسی مسائل چند جانبه است. به علت کاربرد وسیع ، سهولت استفاده و به کار گیری آن ، AHP ، در همه جا مورد مطالعه قرار گرفته می شود و در عرصه های گوناگون برای حل مسائل تصمیم گیری از آن استفاده می شود . بعلاوه ، استفاده از مقایسات دو به دو که بر اساس قضاوت های کارشناسان خبره ، برای مقایسه جایگزینهاست ، به فرد تصمیم گیرنده این اجازه را می‌دهد که بر روی مقایسه فقط دو موضوع بپردازد و بررسی و مشاهده را تقریبا آزاد از هر گونه تاثیرات بی ربط انجام دهد. اگر تصمیم گیری توسط گروه صورت بپذیرد ، می توان AHP را به سادگی تنظیم و طراحی کرد تا اینکه بصورت انفرادی انجام شود . راپر — لو ۵ و شارپ ٦ (۱۹۹۰) ، به این موضوع پی بردند که طراحی یک مسئله مانند سلسله مراتب ، کمک مفیدی به درک مسائل و پیشبرد بحث درباره آنها می کند . این پروسه می تواند مسائلی را که قبلا صراحتا توضیح داده نشده اند را فاش کند . علاوه بر این ، درک این پروسه اسان است و افراد تصمیم گیرنده با ان راحت هستند .
این روش در ۴ مرحله بکار می رود (ساتی ۱۹۸۰):

۱) طراحی یک سلسه مراتب که به توصیف مسئله بپردازد .

۲) طرح ریزی ماتریس های برای مقایسات دو به دوی میان سطوح (رده های) متوالی

۳) ارائه اولویت ها یا مقادیر نسبی عناصر در هر سطح از سلسله مراتب که از بردارهای ویژه استفاده می کنند .

۴) ترکیب مقادیر نسبی سطوح مختلف که از طریق مرحله سوم بدست می‌آیند تا یک نمره کلی از جایگزین های تعمیم ایجاد کنند .

۱-۸-۵-۳. مقایسات دو به دو (زوجی)

پس از طرح ریزی سلسله مراتبی که مسئله را شرح می دهد ، مرحله دوم مرحله اندازه گیری و جمع آوری داده هاست که شامل اداره مقایسات دو به دو است . دو نوع مقایسه وجود دارد که انسان ها بوجود می آورند : مقایسه مطلق و نسبی . اندازه گیری مطلق (معروف به امتیاز بندی) برای دسته بندی گزینه ها بر حسب معیارها یا درجه بندی معیار بکار برده می شود ؛ برای مثال : عالی ، خیلی خوب ، خوب ، متوسط ، زیر متوسط، بد ، خیلی بد . در مقایسات مطلق ،گزینه ها با یک نمونه استاندارد یا یک خط مبنا که در حافظه فرد وجود دارد و از طریق تجزیه رشد یافته است تطبیق می شوند . اندازه گیری نسبی در مقایسات دو به دویی دو عنصر نسبت به یک ویژگی

5- Roper – Lowe
6- Sharp

مشترک بکار برده می شود . در مقایسات نسبی ، گزینه ها بصورت دو تا دو تا براساس یک ویژگی مشترک مقایسه می شوند . AHP با هر دو نوع از مقایسه ها برای گرفتن مقیاس نسبی اندازه گیری استفاده شده است (ساعتی ، ۱۹۸۰) .

مقایسه دو به دو ، فرایند جمع اوری داده ای ورودی عناصر تصمیم گیری است ، طوریکه بتوان اولویت های مقیاس نسبتی را استنتاج کرد . این موضوع در مشکلات (مسائل) تعمیم گیری حائز اهمیت است ، تا جائیکه انتخاب یک گزینه خاص بطور فوری ، معمولا غیر ممکن است . به ندرت گزینه ای وجود دارد که از لحاظ تمام معیارها برگزیده دارای ارجحیت باشد . عملا ، این عادتی که گزینه های خاص از لحاظ برخی معیارها نسبت به بقیه بهتر هستند در حالیکه بقیه گزینه ها بر حسب معیارهای باقی مانده ارجح تر در نظر گرفته می شوند . این مشکل از طریق مقایسات دو به دویی انتخاب گزینه ها حل می شود.

اصطلاحات کلامی مقیاس اصولی ساتی که دارای شدت ۹ امتیازی مطلق است و در جدول ۳-۲ نشان داده شده است. برای برارود کردن نسبت ها همانند اعداد استفاده می شود ، یعنی اهمیت نسبی عناصر را در هر سطحی از سلسله مراتب با عدد نشان می دهد.

جدول ۱ . اصطلاحات کلامی مقایسات زوجی

ردیف	قضاوت	مقدار عددی
۱	کاملا مهم تر و مطلوب تر	۹
۲	ترجیح یا اهمیت خیلی قوی	۷
۳	ترجیح یا اهمیت قوی	۵
۴	کمی مهم تر یا کمی مطلوب تر	۳
۵	اهمیت یکسان و مطلوبیت برابر	۱
۶	ترجیحات بین فواصل	۲ و۴و۶و۸

مقدار معکوس نیز به مقایسه معکوس نسبت داده می شود، یعنی aij = 1/aij .

ماتریس مقایسات زوجی بصورت معادله (۱) نگاره می شود. که در آن aij نشان دهنده قضاوت های کلامی ست.

$$A=\begin{bmatrix} a_{11} & a_{12} & .... & a_{1n} \\ a_{21} & a_{22} & .... & a_{2n} \\ .... & & .... & .... \\ a_{n1} & a_{n2} & .... & a_{nn} \end{bmatrix} \qquad (۱)$$

در صورتی که در معادله (۱) رابطه ی $a_{jk}=a_{ik}/a_{ij}\ \ i,j,k=1,...,n$ برای تمام مقایسات برقرار باشد، ماتریس A سازگار است(ساتی، ۱۹۸۰).

هنگامی که تمام مقاسیات دو به دو در هر سطح انجام شدند و ماتریس های مقایسه بوجود آمدند ، مقیاس اولویت های نسبی از طرف مقایسات زوجی کسب می شود . از این رو ، بجای تعیین دو عدد w_i و w_j که نسبت w_i/w_j را شکل می دهد ، یک عدد مطلق مفرد از مقیاس بنیادی (که بیان می کند چند بار بزرگتر بر کوچکتر تسلط می یابد) برای تقریب نسبت $(w_i/w_j)/1$ معین شده است . نزدیک ترین تقریب عدد صحیح نسبت به w_i/w_j است . بنابراین ، فرایند مقایسه دو به دو از اندازه واقعی برای عنصرهایی که تحت مقایسه قرار گرفتند استفاده می کند . یعنی ، وزن های $w=(w_1,w_2,...,w_n)$ که قبلا بدست آمده اند به سمت ماتریس معکوس A که در معادله (۲) نشان داده شده است، برده می شود (ساتی ، ۱۹۸۰):

$$A=\begin{bmatrix} w_1/w_1 & w_1/w_2 & .... & w_1/w_n \\ w_2/w_1 & w_2/w_2 & .... & w_2/w_n \\ .... & .... & .... & .... \\ w_n/w_1 & w_n/w_2 & .... & w_n/w_n \end{bmatrix} \qquad (۲-۳)$$

از آنجا که $a_{ij}=w_i/w_j$ و همچنین $a_{ij}.a_{jk}=\frac{w_i}{w_j}.\frac{w_j}{w_k}=\frac{w_i}{w_k}=a_{ik}$ ، بنابراین بمنظور اثبات مقادیر معکوس خواهیم داشت:

$$a_{ji}=\frac{w_j}{w_i}=\frac{1}{w_i/w_j}=\frac{1}{a_{ij}}$$

تصور کلی ساتی که در بالا بیان شد ، منجر به تقریب حکم ماتریس A شد که در معادله (۳-۱) با ماتریس نسبت های داده شده در معادله (۲) نشان داده می شود، که عناصرشان نسبت های مقادیر w_i/w_j هر عنصر (n)

نسبت به بقیه عناصر هستند. ورودی های درون ماتریس، سلطه برتری عنصر مسیر سطری را نسبت به عنصر مسیر ستونی بیان می کند.

۳. جمع‌بندی و نتیجه‌گیری

در این بخش به توصیف روش پژوهش، اهداف، سوالات و روندتکاملی تحقیق پرداختیم. یک تحقیق زمانی پایایی مناسب را خواهد داشت که از نظر روش شناسی منطق مناسبی داشته باشد. در ادامه روش های داده کاوی استفاده شده در تحقیق مورد بحث قرار گرفتند. روش RFM یکی از متداولترین روشهای داده کاوی است که برای پیش بینی بر اساس سه متغیر تاخر، تناوب و ارزش پول، بکار می رود. روش SOM شبکه عصب نیز ارز روشهای داده کاوی بر پایه تشکیل نورون های عصبی با روابط پیچیده است که دنیای خارج مسایل ارتباط برقرار می کند. از طرفی رگرسیون لجستیک یکی دیگر از روش های داده کاوی است که بر پایه آمار و احتمالات و متغیر های وابسته عمل مینماید. در این تحقیق این روشها مورد استفاده قرار گرفتند و مورد مقایسه واقع شدند. برای اتفاقات سازمانی، نمی توان توجیه های ساده یافت و برای آنها به سادگی علتهایی پیدا کرد. بعلاوه غالباً یافتن سرچشمه ها مقدور نیست زیرا منشأها معمولاً در فاصله ای بسیار دورتر از نتایج و علایم قرار دارند و نتایج و علایم بواسطه حلقه های تشدید کننده انحراف از معیار با اصل خود تفاوتی فاحش و غیر قابل تصور می یابند. پژوهشگر پس از اینکه روش تحقیق خود را مشخص کرد و با استفاده از ابزارهای مناسب، داده های مورد نیاز را برای آزمون فرضیه های خود جمع آوری کرد، نوبت آن فرا می رسد که با بهره گیری از تکنیکهای داده پردازی مناسبی که با روش تحقیق، نوع متغیرها، ... سازگاری دارد، داده های جمع آوری شده را دسته بندی و تجزیه و تحلیل نماید و در نهایت فرضیه ها ی سوالاتی را که تا این مرحله او را در تحقیق هدایت کرده اند در بوته آزمون قرار دهد و تکلیف آنها را روشن کند و سرانجام بتواند پاسخی (راه حلی) برای پرسشی که تحقیق (تلاشی سیستماتیک برای بدست آوردن آن بود) بیابد. (خاکی، ص ۳۰۴-۳۰۳ : ۱۳۸۲)

۲-۴. آماده سازی و پیش پردازش داده ها

در این پژوهش از داده های تراکنشی مشتریان مربوط به صنعت لیزینگ استفاده شد. در مجموع تعداد ۱۲۳۲ مشتری در فاصله زمانی ۱۴۰۳-۱۴۰۲ جمع آوری شد. با توجه به محدودیت هایی که در تحویل داده های دموگرافیک وجود داشت، تنها داده های تراکنشی در اختیار قرار گرفتند. مرحله آماده سازی از مهمترین و پیچیده ترین مراحل داده کاوی است. مراحل بکار رفته شامل فرآیند پاکسازی و کاهش بعد می باشد. در نهایت تعداد ۱۶۰ مشتری برای انجام عملیات اعتبار سنجی باقی ماندند. با مدل RFM داده های آماده برای استفاده در عملیات عتبار سنجی شدند.

۳-۴. وزن شاخص ها

سه شاخص تولیدشده برای استفاده در الگوریتم خوشه بندی بر اساس روش AHP تعیین وزن می شوند. این هدف می توانند با مد نظر گرفتن معیارهای تازگی خرید (R)، و ارزش مالی خریدهای مشتری ، شاخص فراوانی خرید (F): در ارتباط با خوش بندی مشتریان کسب شوند. یکی از مراحل اصلی روش تحلیل سلسله مراتبی، مقایسات زوجی است. این مرحله یکی از نقاط قوت اصلی مدل AHP است. این مدل نسبت مقیاس اولویت ها را به درستی استخراج می کند، برعکس رویکردهای سنتی ارزیابی اوزان که توجیه آنها نیز سخت است. از

متخصصین خواسته شد تا با استفاده از اعداد مقیاس اصلی قضاوت خود را بیان کنند: ۱ = برابر، ۳ = نسبتا مهمتر، ۵ = به شدت مهمتر، ۷ = خیلی به شدت مهمتر است و ۹ = فوق العاده مهمتر است (ساتی، ۱۹۸۰). نتایج جمع اوری شده، سپس، در یک ماتریس متقارن، جهت تشکیل ماتریس های قضاوت مقایسه زوجی مربوطه، وارد شدند. ذکر این نکته ضروری است که AHP در مقایسه با ANP به تعداد مقایسه های زوجی کمتری نیاز دارد.

پس از بدست آوردن ماتریس مقایسه زوجیبا استفاده از بردارهای ویژه‌ی نرمال‌سازی شده‌ی حاصل که دربرگیرنده‌ی اولویت‌های نسبی هستند، ارائه می‌شوند، یعنی اهمیت نسبی عناصر سطوح مشابه سلسله‌مراتب نسبت به عنصر سطح بالاتر در یک ویژگی مشترک مورد مقایسه قرار می‌گیرند. در این فصل، روش بردار ویژه، که بردار ویژه‌ی یک ماتریس دوسویه را تخمین می‌زند، به شکلی که در زیر آمده، جهت تخمین بردارهای اولویت نسبی مورد استفاده قرار می‌گیرد. با توجه به این مساله وزن شاخص ها طبق ماتریس جدول ۲ محاسبه شد.

جدول ۲ . ماتریس مقایسه زوجی

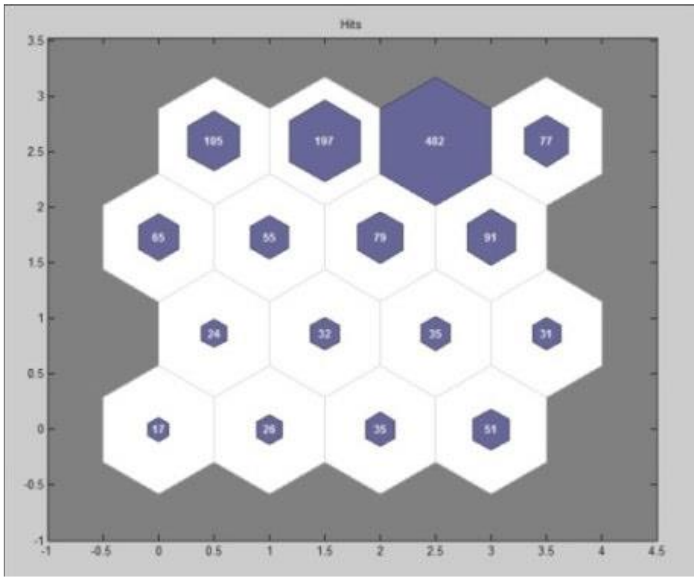
	R	M	F	وزن
R تازگی خرید	1	6	4	0.69
M ارزش مالی خریدهای مشتری	1/6	1	1/3	0.09
Fفراوانی خرید	1/4	3	1	0.22

بنابراین وزن شاخص تازگی خرید ۶۹%، ارزش مالی مشتری ۹% و فراوانی خرید ۲۲% از مجموع عوامل RFM را دارا هستند.

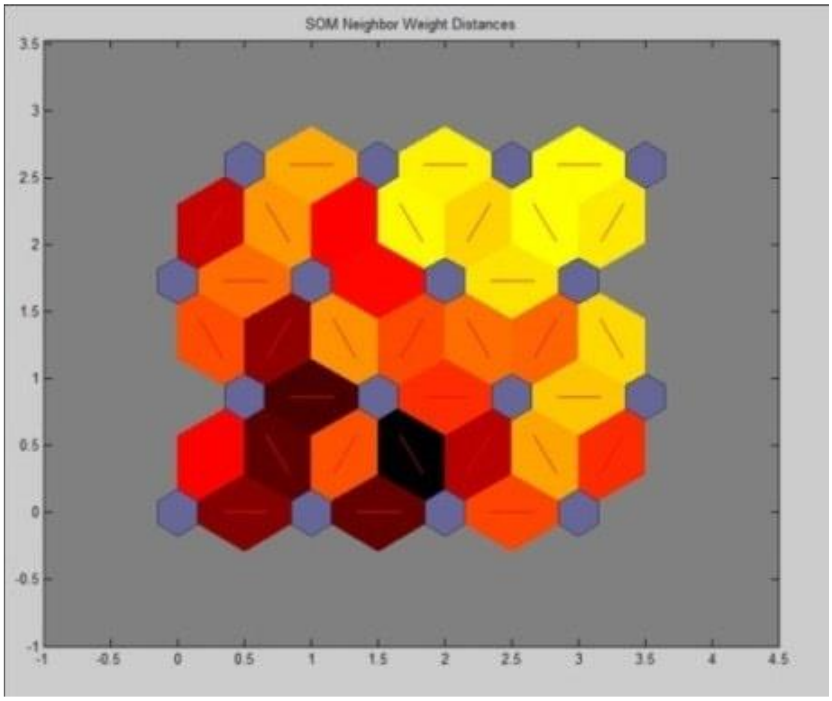
۴-۴. خوشه بندی SOM بر اطلاعات RFM

در این مطالعه در نخست با بکارگیری الگوریتم SOM و شبکه ای در ابعاد چهار در چهار و نورون های شش ضلعی تولید شد. هریک از این سلول های عصبی از راه وزن های سیناپسی که در طول یادگیری به بردار ورودی متصل است، تنظیم شدند. فاز اول برآورد ناهموار است که برای تولید الگوهای داده های ناخالص استفاده شد. فاز دوم فاز تنظیم بود که برای تنظیم نقشه شبکه به مدل ویژگی های خوب از داده مور د استفاده قرار گرفت (۱۵).

انجام الگوریتم خوشه بندی به این علت است که بتوان گروه هایی از مشتریان را برای رتبه بندی شناسایی نمود. نتایج اجرای الگوریتم SOM در نرم افزار متلب در شکل ۲ و ۳ قابل مشاهده است. در شکل ۲ تعداد مشتریانی را که در هر نورون تقسیم شده اند نمایش میدهد. شکل ۳ نشان دهنده میزان فاصله بین نورون‌های همسایه از یکدیگر است که هرچقدر میزان فاصله بین نورون‌های همسایه از یکدیگر بیشتر باشد، با رنگ تیره تر نشان داده می‌شود و هرچقدر فاصله کمتر باشد با رنگ های روشن تر نمایش داده می شود.



شکل ۲ . تعداد اعضای نورون ها



شکل ۳ . میزان فاصله نورون‌ها

در جدول ۳ نتایج محاسبات SSE بر اساس تعداد خوشه ها ی مختلف را مشاهده می کنید. در نتیجه می توان گفت که ۸ خوشه اساسی در این شبکه می توان یافت.

جدول ۳. نرخ پارامتر SSE به ازای تعداد خوشه در SOM

تعداد خوشه	SSE
۸	۱۴/۶۲
۹	۱۸/۶۱
۱۱	۱۵/۷۶
۱۰	۱۹/۵۰

بر اساس تعداد خوشه تعیین شده، مشتریان بر اساس شاخص های آر. اف.ام .خوشه بندی شدند که نتایج این خوشه بندی در جدول ۴ آورده شده است.

جدول ۴. نتایج خوشه بندی مشتریان

خوشه	تعداد مشتریان
۱	۱
۲	۱
۳	۲
۴	۸
۵	۲
۶	۱۴۰
۷	۲
۸	۴
جمع	۱۶۰

۵-۴. تحلیل واریانس معیارهای RFM

بدلیل آنکه بدانیم کدام فاکتور تاثیر بیشتری بر جداسازی خوشه ها داشته است باید از تحلیل واریانس استفاده می کردیم. با توجه به جدول تحلیل واریانس جدول ۵ مشخص میگردد که شاخص ارزش پولی خریدهای مشتری (M') با بزرگترین مقدار آماره F یعنی ۸۷۷/۲۵ در سطح خطای کمتر از ۱ درصد بیشترین نقش را در جداسازی خوشه ها از یکدیگر داشته است. همچنین شاخص تازگی خرید (R') با کوچکترین مقدار F یعنی ۲/۲۷ در سطح خطای کمتر از ۰/۰۲۹ کمترین نقش را در جداسازی خوشه ها داشت.

جدول ۵. تحلیل واریانس خوشه ها

شاخص ها	خوشه ها		آماره F	Sig.
	میانگین مربعات	درجه آزادی		
تازگی خرید	۰/۰۰۱	۷	۲/۲۷	۰/۰۲۹
شاخص فراوانی خرید	۰/۰۵	۷	۲۸۱/۸۱	۰
ارزش پولی خریدهای مشتری M')	۰/۱۳۹	۷	۸۷۷/۲۵	۰

۶-۴. طبقه بندی مشتریان

در ادامه، متوسط ارزش هر یک از شاخص های تازگی خرید و ارزش مالی خریدهای مشتری، شاخص فراوانی خرید در هر خوشه با تقسیم مجموع ارزش شاخص ها در هر خوشه به تعداد مشتریان آن خوشه تعیین گردید که در جدول ۶ آورده شده است.

جدول ۶. متوسط ارزش شاخص های RMF در خوشه‌ها

خوشه	متوسط ارزش تازگی خرید	متوسط ارزش مالی خرید مشتری	متوسط ارزش فراوانی خرید
۱	۰/۰۷۱	۰/۰۷۳۵	۰/۵۹۹
۲	۰/۰۷۱۲	۰/۱۳۵۶	۰/۳۷۸۳
۳	۰/۰۷۱	۰/۲۶۵۳	۰/۳۲۰۴
۴	۰/۰۶۴۳	۰/۱۰۶	۰/۰۲۵۴
۵	۰/۰۷۱۵	۰/۰۸۷۹	۰/۲۵۵
۶	۰/۰۵۲	۰/۰۰۸۵	۰/۰۰۶۳
۷	۰/۰۷۱	۰/۲۳۶	۰/۱۵۱۳
۸	۰/۰۷۲	۰/۰۹۹	۰/۱۲۱۶

با توجه به خروجیهای مراحل قبل، مقایسه متوسط ارزش شاخصهای آر.اف.ام .در هر خوشه با متوسط ارزش این شاخصها در کل داده ها و همچنین رتبه بندی خوشه ها با توجه به مقدار متوسط ارزش شاخص های آر.اف.ام.، تحلیل خوشه‌ها در قالب جدول ۷ صورت گرفت.

جدول ۷. تحلیل خوشه ها

تحلیل	خوشه ها							
	۱	۲	۳	۴	۵	۶	۷	۸
وضعیت متوسط ارزش شاخص ها	↑	↑	↑	↑	↑	↓	↑	↑
رتبه متوسط ارزش شاخص MM	۱	۲	۳	۷	۴	۸	۵	۶
رتبه متوسط ارزش شاخص MF	۷	۳	۱	۲	۶	۸	۲	۵
رتبه متوسط ارزش شاخص MR	۱	۵	۱	۵	۵	۸	۱	۴

در این جدول، ابتدا در قسمت وضعیت شاخص ها، به مقایسه متوسط ارزش هر یک از شاخص های .RFM در هر خوشه با متوسط ارزش این شاخصها در کل داده ها پرداخته شده است که این مقایسه مشخص می‌نماید که متوسط ارزش هر یک از شاخصهای آر.اف.ام .در هر خوشه نسبت به متوسط ارزش این شاخصها در کل داده‌ها در چه وضعیتی قرار دارد. برای هر یک از شاخصها، در وضعیتی که متوسط ارزش شاخص در یک خوشه بیشتر از متوسط ارزش آن شاخص در کل دادهها باشد، این وضعیت با علامت(↑) (وضعیت مطلوب) و در صورتی که متوسط ارزش شاخصی در یک خوشه کمتر از متوسط ارزش آن در کل داده ها باشد، این (وضعیت با علامت(↓) نشان داده است (وضعیت نامطلوب) در قسمت بعدی جدول ۷ جهت بررسی دقیق تر خوشه ها، رتبه متوسط ارزش شاخصها در هر خوشه در مقایسه با خوشه های دیگر آورده شده است. برای مثال رتبه مربوطه در خوشه ۱ و سطر تازگی خرید مشخص میکند که این خوشه از نظر مقدار متوسط ارزش شاخص های تازگی خرید در رتبه X قرار دارد. پس از انجام تحلیل خوشه ای، ریسک اعتباری مشتری در هر خوشه بر اساس رابطه زیر محاسبه گردید و در نهایت نیز مشتریان بر اساس ارزش اعتباری در خوشه های ایجاد شده در قالب ارزش اعتبار مشتری جدول ۸ نشان داده شده است، بخش بندی شدند.

$$RV = M_{R''} + M_{F''} + M_{M''}$$

جدول ۸. ارزش رتبه بندی شده مشتریان

خوشه	تعداد	ارزش اعتباری RV	رتبه بندی
۱	۱	۰/۷۹۶	۱
۲	۱	۰/۵۹۷	۳
۳	۲	۰/۶۷۸	۲
۴	۸	۰/۱۹۹	۷
۵	۲	۰/۴۳۶	۵
۶	۱۴۰	۰/۱۶۵	۸
۷	۲	۰/۴۶۰	۴
۸	۴	۰/۲۹۲	۶

برای سنجش کارایی مدل ارائه شده صحت نتایج این مدل با الگوریتم های طبقه بندی دیگری چون شبکهٔ عصبی، درخت تصمیم CHAID و رگرسیون لجستیک مورد مقایسه قرار گرفت که نتایج در جدول ۹ قابل مشاهده است.

جدول ۹ . مقایسه تطبیقی دقت الگوریتم‌ها

الگوریتم	RFM پیشنهادی	شبکه عصبی	CHAID	رگرسیون لجستیک
دقت	۷۴%/۹۴	۷۴%/۳۳	۷۳%/۹۹	۶۸%/۳۹

همان طور که در جدول ۹ مشاهده می شود، آر.اف.ام پیشنهادی به میزان ۰/۶۱ درصد دارای دقت بالاتری نسبت به بقیه روش ها می باشد. این الگوریتم از پیچیدگی کمتری برخوردار اس که بدلیل گستردگی آن، نیازی به فرم های درختی زیاد نیست و این یک مزیت بشمار می رود. الگوریتم استخراج شده قابلیت طبقه بندی رکوردهای جدید مشتریان را داشته و می توان به میزان صحت آن در تشخیص وضعیت رفتار مشتریان جدید اطمینان کرد .هنگامی که اطلاعات مشتری جدید وارد سیستم می شود، آن را از ابتدای خوشه بندی بر اساس ویژگی های بیوگرافیک هدایت می کنیم تا زمانی که به یک گرهٔ برگ برسیم، در آنجا ردهٔ مربوطه را به مشتری نسبت داده و می توانیم پیش بینی کنیم مشتری در کدام طبقهٔ قرار می گیرد.

۷-۴. نتیجه گیری

در بازارهای رقابتی امروزی، با گرایش شرکتها به سمت مشتری مداری، مدیریت ارتباط با مشتری نیز به سمت پیچیدگی های خاصی گرایش پیدا کرده است .طبق مطالعات گذشته تخمین زده شده است که هزینه های جذب مشتریان جدید، پنج برابر هزینه های حفظ مشتریان موجود خواهد بود (کوتلر ۱۹۹۴). از سویی دیگر، تمرکز شرکتهای امروزی تنها بر فروش کالاهایشان نیست، آنها در پی خلق و حفظ مشتریان سودآور هستند. در این مبحث پر پیچ و خم طبقه بندی مشتریان اهمیت بسزایی دارد. در این فصل داده ها بر اساس روش تحقیق

پیشنهادی RFM و SOM تحلیل شدند و خوشه بندی مشتریان و در نهایت رتبه بندی آنها بر اساس اعتبار مشتریان انجام شد. این تحلیل بر روی ۱۶۰ مشتری صنعت لیزینگ مورد مطالعه انجام شد و نتایج بسیار ارزنده ای را در بر داشت. در آخر نیز مدل پیشنهادی با دیگر روش های دده کاوی مقایسه شد که مشخص شد دقت این روش پیشنهادی به اندازه ۰/۶۱ درصد بیش از سایر روش ها بوده است.

۲-۵. چگونگی پردازش اهداف پژوهش

هدف اصلی هدایت پژوهش، ارائه روشی جامع برای طبقه بندی مشتریان بر اساس اعتبار آنها در چارچوب کشورهای در حال توسعه بود. هدف اول پژوهش، بررسی و تجزیه و تحلیل مسائل و چالش های درگیر در طبقه بندی مشتریان شرکت ها در کشورهای در حال توسعه است. دو روش پژوهش در دستیابی به این هدف مورد استفاده قرار گرفت — بررسی ادبیات سنتی و گذشته و تعامل با کارشناسان عرصه مدیریت ارتباط با مشتری، دو روش پژوهشی هستند که در دستیابی به این هدف مورد استفاده قرار گرفت. به دنبال بحث ارائه شده در ادامه بررسی های صورت گرفته در زمینه طبقه بندی مشتریان مشخص کرد که انواع پژوهش های صورت گرفته طبق مدل های کیفی و کمی هستند. این تحقیق در ردیف مدل های کیفی قرار می گیرد. مدل های کیفی، مدل های توصیفی هستند که بر پایه قوانین ریاضیاتی و فرمول ها نیستند. این مدل ها سیستم های موجود را با توجه به روابط اتفاقی، پایه ای، ترکیبی و یا تنظیم شده میان موضوعات (اشیا) و وقایع به تصویر می کشند(کلانسی، ۱۹۸۹). این تحقیقات از حوزه مدیریت ارتباط با مشتری آغاز شده و در نهایت خوشه بندی مشتریان بر اساس اعتبار آنها ارائه شد. با این حال، رتبه بندی مشتریان در کشورهای در حال توسعه بعنوان یک مسئله پیچیده در مسائل اقتصادی و مدیریتی در نظر گرفته می شود. علاوه بر این، طبقه بندی مشتریان در زیرساخت های اعتبار ی خواستار تحقیق و تفحص از تلفیق عوامل فنی و غیر فنی مؤثر بر خوشه بندی گروهها به عنوان یک هدف برای اولین بار راه اندازی شد. تحقیقات نشان می دهد که اکثر پژوهش ها در مدل های کیفی مربوط به استفاده از مدل های تصمیم گیری و روش های ژنتیک و شبکه های عصبی در جهت طبقه‌بندی مشتریان و استراتژیک راهبردی بوده است.

براساس مطالعات صورت گرفته هیچ روش طبقه بندی جامعی برای ارزیابی اعتبار مشتریان وجود ندارد که از نظر تحلیلی بتواند به ماهیت پیچیده سیستم ارزیابی اعتبار مشتریان صنعت لیزینگ رسیدگی کند. تحقیقات اولیه در این مطالعه نشان داد که روش های داده کاوی نقش هدایت کلیدی را در توسعه مدل های خوشه بندی مشتریان ایفا می کند. این نتیجه گیری در نهایت منجر به نیاز یک رویکرد چند مرحله برای طبقه بندی مشتریان، در صنعت مورد مطالعه شد. بنابراین لازم بود که روش ها اکتشافی مبنا قرار گیرند تا برای مقابله با مسائل پیچیده مرتبط با ارزیابی و طبقه بندی مشتریان مناسب باشد.

هدف دوم بررسی و تجزیه و تحلیل مناسب و همچنین تلفیق روش های داده کاوی، یعنی، RFM و شبکه عصبی، که می تواند برای طبقه بندی مشتریان بر اساس ریسک اعتباری استفاده شود. تمرکز بر تحقیق و تجزیه و تحلیل روش های مناسب و تکنیک های زمینه ای داده کاوی، می‌تواند برای بهبود روند خوشه بندی جاری مشتریان مورد استفاده قرار گیرد. به منظور دستیابی به این هدف، لازم بود که برای انجام این پژوهش، روش های داده کاوی بازنگری شود. در واقع تا جایی که امکان داشت تکنیک های اتخاذ شده مناسب برای طبقه بندی مشتریان اعتباری در این تحقیق، استفاده شد.

هدف سوم تدوین و فرموله کردن روش های داده کاوی برای طبقه بندی مشتریان بالقوه بر اساس ریسک اعتباری در چارچوب مباحث نظری است. تکنیک های شرح داده شده در این تحقیق از طریق ماهیت تحلیلی خود با ارائه درک و بینش عمیق تری نسبت به مسائل مرتبط با طبقه بندی مشتریان شرکت داشته اند. در درجه اول، مدل کلی RFM تدوین شد که در آن کارشناسان مدیریت ارتباط با مشتری دعوت شده بودند تا قضاوت های خود را ارائه دهند. در این پژوهش از داده های تراکنشی مشتریان مربوط به صنعت لیزینگ استفاده شد. در مجموع تعداد ۱۲۳۲ مشتری در فاصله زمانی ۱۴۰۳-۱۴۰۲ جمع آوری شد. با توجه به محدودیت هایی که در تحویل داده های دموگرافیک وجود داشت، تنها داده های تراکنشی در اختیار قرار گرفتند. مرحله آماده سازی از مهمترین و پیچیده ترین مراحل داده کاوی است. مراحل بکار رفته شامل فرآیند پاکسازی و کاهش بعد می باشد. در نهایت تعداد ۱۶۰ مشتری برای انجام عملیات اعتبار سنجی باقی ماندند. با مدل RFM داده های آماده برای استفاده در عملیات عتبار سنجی شدند. در مرحله بعد، در مرحله تحلیل، مدل مذکور با استفاده از روش AHP ، وزن نهایی هریک از گزینه ها تعیین شد. با این حال، سادگی ساختار سلسله مراتبی و روابط یک سویه ی خطی در میان معیارها و زیر معیارها در روش AHP و پنهان کردن مسائل مهمی، مانند وابستگی متقابل میان عوامل کیفی و تعامل میان سطوح تصمیم گیری و پردازش بیش از حد آسان در این مدل سبب پیدایش مشکلاتی در مسأله می شود. در مرحله بعد روش SOM برای خوشه بندی استفاده شد. آزمون اعتبار سنجی مدل فوق را تایید نمود. بنابراین می توان آن را به عنوان مدل عمومی و مرجع در سایر تکنیک های خوشه بندی مشتریان به کار برد.

هدف چهارم. مقایسه روش های داده کاوی ترکیبی با رگرسیون لجستیک بود. بر این اساس مدل پیشنهادی با دیگر روشهای داده کاوی مورد مقایسه قرار گرفت. نتایج نشان داد که آر.اف.ام پیشنهادی به میزان ۰/۶۱ درصد دارای دقت بالاتری نسبت به بقیه روش ها می باشد. این الگوریتم از پیچیدگی کمتری برخوردار اس که بدلیل گستردگی آن، نیازی به فرم های درختی زیاد نیست و این یک مزیت بشمار می رود. الگوریتم استخراج شده قابلیت طبقه بندی رکوردهای جدید مشتریان را داشته و می توان به میزان صحت آن در تشخیص وضعیت رفتار مشتریان جدید اطمینان کرد .هنگامی که اطلاعات مشتری جدید وارد سیستم می شود، آن را از ابتدای خوشه بندی بر اساس ویژگی های بیوگرافیک هدایت می کنیم تا زمانی که به یک گرهٔ برگ برسیم، در آنجا ردهٔ مربوطه را به مشتری نسبت داده و می توانیم پیش بینی کنیم مشتری در کدام طبقهٔ قرار می گیرد.

۳-۵. پردازش سوالات تحقیق

سوال اول: کدام الگوریتم خوشه بندی صحیح تری را نسبت به سایر الگوریتم های خوشه بندی داده کاوی ارائه می کند؟.

در این تحقیق با ارائه یک مطالعه موردی اثبات شد که روش پیشنهادی RFM و شبکه مصنوعی SOM بهترین الگوریتم برای اجرای خوشه بندی مشتریان صنعت لیزینگ بود.

سوال دوم: کدام شاخص مهمترین شاخص در خوشه بندی مشتریان می باشد؟
پاسخ: بر اساس روش AHP استفاده شده در این پژوهش مشخص شد که وزن شاخص تازگی خرید ۶۹%، ارزش مالی مشتری ۹% و فراوانی خرید ۲۲% از مجموع عوامل RFM را دارا هستند. بنابراین می توان گفت که شاخص تازگی خرید با ۶۹% مهمترین شاخص در خوشه بندی مشتریان بود.

سوال سوم: کدام مدل الگوریتم داده کاوی نسبت به سایر الگوریتم ها از صحت بالاتری برخوردار میباشد؟
پاسخ: بر اساس مقایسه تطبیقی مشخص شد که مدل پیشنهادی RFM و SOM از صحت بالتری برخوردار هستند طوریکه آر.اف.ام پیشنهادی به میزان ۰/۶۱ درصد دارای دقت بالاتری نسبت به بقیه روش ها می باشد. این الگوریتم از پیچیدگی کمتری برخوردار اس که بدلیل گستردگی آن، نیازی به فرم های درختی زیاد نیست و این یک مزیت بشمار می رود.

سوال چهارم: آیا بهینه سازی خوشه بندی بر مبنای الگوریتم های هوشمند (ژنتیک، ازدحام ذرات، جستجوی فاخته، رقابت استعماری و...،) از صحت بالاتری برخوردار نیست؟
پاسخ: بنظر می رسد در پیرو سوال قبل روش های ذکر شده نتوانند صحت بالاتری نسبت به روش پیشنهادی داشته باشند چراکه نتیجه مقایسات تطبیقی این موضوع را نشان داده است.

۴. پیشنهادهای کاربردی

با توجه به اینکه مشتریان تمامی خوشه ها به غیر از خوشه ششم، از نظر شاخص های آر.اف.ام .در وضعیت مطلوبی قرار دارند، لذا پیشنهاد می گردد که به منظور حفظ این مشتریان، شرکت با برقراری ارتباطات و تعاملات بیشتر با آنها، سعی در تبدیل وفاداری رفتاری این مشتریان به وفاداری نگرشی نماید . اگرچه مشتریان خوشه های ۳ و ۷، ارزش اعتباری بالایی ندارند، اما فراوانی خرید آنها در مقایسه با خوشه های دیگر، بسیار بالا است. شرکت می تواند با در نظر گرفتن تخفیف های حجمی خاص برای مشتریان این خوشه ها، بر حجم خرید این مشتریان افزوده و از این طریق، تعداد دفعات خرید آنها و متعاقب آن، قسمتی از هزینه های تحمیل شده را کاهش دهد. با توجه به اینکه خوشه ششم در پایین ترین سطح اعتباری قرار دارند اما احتمال ارتقای این مشتریان به سطوح بالای هرم ارزش مشتری، بسیار بالا می باشد و شرکت می تواند با در نظر گرفتن مزایایی خاص برای خرید آنها، وفاداری رفتاری را در بین آنها ترویج دهد.
پیشنهاد میگردد که در مطالعات آتی، از مجموعه داده وسیع تری به لحاظ قلمروی زمانی استفاده گردد که به طور قطع منتج به ایجاد نتایج قویتر و گستره وسیع تری از دانش کاربردی پیرامون ویژگیهای رفتاری مشتریان خواهد شد .از طرفی، برشهای زمانی متناوب از پایگاه داده شرکت، امکان اجرای پویای داده کاوی بر مبنای مدل آر.اف.ام . جهت تعیین ارزش اعتبار مشتریان را فراهم خواهد آورد که نتایج آن از این لحاظ که روند تغییرات در رفتار مشتریان را منعکس میکند، میتواند نقش شایانی در اقدامات بازاریابی و بهبود مدیریت ارتباط با مشتری ایفا نماید.همچنین استفاده از روش های تلفیقی دیگر همانند تکنیک های تصمیم گیری چند معیاره مانند ANP میتواند کمک شایانی برای بهبود این تحقیق باشد. لزوم فازی بودن امتیازات در RFM میتواند مسیری برای توسعه مدل مذکور در شرایط عدم قطعیت باشد.

در این تحقیق، جهت بخش بندی صنعت لیزینگ بر اساس ارزش اعتباری مشتری از مدل آر. اف.ام .و برخی تکنیک های داده کاوی و روش های مختلف علمی در قالب فرایندی خاص، بهره گرفته شد . استفاده از جدول تحلیل واریانس جهت تعیین میزان اثرگذاری هر یک از شاخص ها در تفکیک خوشه ها، ارائه تحلیل درون خوشه ای به منظور بررسی دقیق تر ویژگی های مشتریان در خوشه دارای بیشترین مشتری و همچنین بخش بندی نهایی مشتریان بر اساس ارزش اعتباری مشتریان در قالب جدول ارزش اعتبار مشتریان از موارد خاصی است که در این

مطالعه از آنها بهره گرفته شده است، درحالی که در مطالعات گذشته به آنها اشاره ای نگردیده است. روش پیشنهادی با ارائه یک مطالعه موردی نشان داد که از پاییایی بالایی برخوردار است. همچنین که در مقایسه تطبیقی با دیگر زوش های درختی و خوشه بندی از صحت و دقت بیشتری برخوردار بود. استفاده از سیستم RFM هنوز هم در فرآیندهای طبقه بندی بصورت قاطع کاربرد دارد و می تواند با دیگر روشها بصورت ترکیبی مدلی قویتر را ارائه نماید.

۵. فهرست منابع

حافظنیا، م، ر(۱۳۸۲). مقدمه‌ای بر روش تحقیق در علوم انسانی، انتشارات سمت، تهران، چاپ هشتم

زارعی نژاد، محسن، ۱۳۹۲ . ارزیابی و انتخاب تامین کننده ثالث لجستیک معکوس با رویکرد تلفیقیANP و IFG-MCDM . پایان نامه کارشناسی ارشد. دانشگاه آزاد واحد تهران جنوب.

خاکی ، غلامرضا " روش تحقیق با رویکرد پایان نامه نویسی " (تهران ، انتشرات بازتاب ۱۳۸۲)

Hughes, Arthur M. (2023), Strategic database marketing, Chicago: Probus publishing.

Coussement, K., Van den Bossche, F. A., & De Bock, K. W. (2024). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, *67*(1), 2751-2758.

Saaty, T. L. (2022). The analytic hierarchy process: planning, priority setting, resources allocation. *New York: McGraw*.

Kotler, P. (2023), Marketing management: Analysis, planning, implementation, and control, New Jersey: Prentice-Hall.

Yeh, C., Yang, K. & Ting, T.; "Knowledge Discovery on RFM Model Using Bernoulli Sequence", *Expert Systems with Applications*, Vol. 36, pp. 5866–5871, 2019.

Lumsden SA, Beldona S, Morison AM. Customer value in an all-inclusive travel vacation club: An application of the RFM framework. J. Hosp. Leisure Mark., 16(3): 2, pp. 70-285, 2024

N. C. Hsieh, “An integrated data mining and behavioral scoring model for analyzing bank customers”, *Expert Systems with Applications*, 27, pp. 623–633, 2024

S. Y. Husseini, & et al., “Segmenting and Profiling Green Consumers with Use of Self Organizing Maps”, *Journal of Management Research in Iran*, Volume 17, Number 2, 2023.

Tsai, Lu. ,”Customer churn prediction by hybrid neural networks”, *Expert Systems with Applications*, Vol. 36, pp. 12547-12553, 2019.

Ghazanfari, M. et al, “Customer segmentation export edible fruits”, *Quarterly Journal of Commerce*, No. 55, 151 – 181, 2022

predicting customers ' risk in the leasing industry using data mining techniques to clustering

Authors[7]
Abdolhossien karampour
Abdolhossien ziyarati
Sara sohrabi motlagh fazeli

Abstract

The main objective of this research is to present a comprehensive method for classifying customers based on their validity in developing countries.In this research, for classifying the leasing industry based on customer's credit worthiness RFM.and some data mining techniques and different scientific methods have been used in the formof a specific process. the use of variance analysis table to determine the effect of each indicator in cluster segregation, presenting in-cluster analysis in order to examine the more accurate characteristics of customers in cluster with customer loyalty and customer's final segmentation based on customer credit worthiness in the form of customer credit worthiness is one of the special cases that have been used in this study. the proposed method by presenting a case study showed that it has a highfrequency. also, in comparison with other tree methods and clustering, it was more accurate. the use of RFM is still widely used in classification processes and can be combined with other methods in a hybrid model.

Keywords: predicting customers ' risk,
leasing industry,data mining techniques, clustering.

JEL Classification O16 Financial Markets ● Saving and Capital Investment ● Corporate Finance and Governance

[7]assistant professor and member of the faculty of department of business administration of islamic azad university , north Tehran branch.