

طراحی مدل طبقه‌بندی مشتریان با استفاده از هوش مصنوعی

و ارائه استراتژی‌های بهبود رضایت در خدمات دهی

جعفر جانعمویی^۱، فاطمه دکامینی^۲

تاریخ پذیرش ۱۴۰۳/۰۷/۰۲

تاریخ دریافت: ۱۴۰۳/۰۴/۲۰

چکیده

در چند سال اخیر، با توسعه یادگیری ماشین و فناوری داده کاوی، داده‌های مشتریان به عنصری حیاتی در فرآیند تولید شرکت‌ها تبدیل شده‌اند. با استفاده از رویکردهای داده کاوی بر روی داده‌های مشتریان، سازمان‌ها می‌توانند الگوهای رفتاری مشتریان، نیازها و ارتباطات پنهان داده‌ها را درک کرده و بر اساس این الگوها بهبودی یافته در خدمات به مشتریان و بهینه‌سازی منابع خود داشته باشند. یکی از تکنیک‌های مورد استفاده در این راستا، خوشه‌بندی است که برای گروه‌بندی مشتریان بر اساس ویژگی‌های مختلف آن‌ها به کار می‌رود. هدف اصلی این تحقیق، خوشه‌بندی مشتریان بر اساس شاخص‌های LRFM و در نهایت پیش‌بینی رفتار خرید آن‌ها با استفاده از تکنیک‌های طبقه‌بندی درخت تصمیم (DTC)، پرسپترون چندلایه (MLP) و ماشین بردار پشتیبان (SVM) بود. این مطالعه بر روی تراکنش‌های مشتریان در سامانه پلیس ۱۰+ یک دفتر در تهران در بازه زمانی ۱۴۰۱ تا ۱۴۰۳ انجام شده است. هر تراکنش به یک مشتری نسبت داده می‌شود و مدل‌سازی می‌شود تا رفتار خرید مشتری را مدل کند. با انجام خوشه‌بندی با استفاده از الگوریتم K-means و تعیین مقدار بهینه K، سه خوشه برای مشتریان مشخص شد. همچنین، آزمایش و بررسی طبقه‌بندی‌کننده‌ها نشان داد که طبقه‌بندی‌کننده MLP با یک لایه پنهان و ۸ نورون در این لایه بیشترین دقت را دارد و طبقه‌بندی‌کننده DTC بیشترین سرعت را در میان طبقه‌بندی‌کننده‌های مورد بررسی دارد. بررسی رفتار مشتریان در این خوشه‌ها نشان داد که مشتریان به سه دسته وفادار، بالقوه و از دست رفته تقسیم می‌شوند.

کلمات کلیدی: رفتار خرید مشتری، LRFM، پرسپترون چندلایه، ماشین بردار پشتیبان، درخت

تصمیم‌گیری.

^۱ دانش‌آموخته DBA مدرسه عالی کسب و کار ماهان، تهران، ایران (adnan.elec@gmail.com)

^۲ عضو هیات علمی مدرسه عالی کسب و کار ماهان (f.dekamini@mahanbs.net)

مقدمه

در عصر حاضر، با توسعه فناوری‌های پیشرفته و افزایش رقابت در بازارهای مختلف، تأمین رضایت مشتریان به عنوان یکی از عوامل کلیدی موفقیت و بقا برای سازمان‌ها و شرکت‌ها به مراتب مهم‌تر از گذشته تلقی می‌شود. مشتریان، به عنوان دارندگان حقوق و نقدهای خود، تأثیر مستقیمی بر موفقیت یا شکست کسب‌وکارها دارند. در این زمینه، توسعه و بهره‌گیری از تکنولوژی‌های هوش مصنوعی و یادگیری ماشینی به عنوان ابزارهای قدرتمندی برای بهبود خدمات به مشتریان و ایجاد راهکارهایی جهت ارتقاء رضایت آن‌ها مطرح شده است (باقری، ۲۰۲۲).

در این پژوهش، ما تلاش می‌کنیم تا با استفاده از هوش مصنوعی و الگوریتم‌های یادگیری ماشینی، یک مدل طبقه‌بندی مشتریان طراحی کنیم که به سازمان‌ها کمک کند تا نیازها و ترجیحات مشتریان را بهبود ببخشند. این مدل قادر به تحلیل الگوهای مشتریان و پیش‌بینی نیازهای آینده آن‌ها خواهد بود. علاوه بر این، راهکارهایی ارائه می‌دهیم تا با استفاده از داده‌های به دست آمده از مدل طبقه‌بندی، سازمان‌ها بتوانند خدمات خود را به گونه‌ای بهبود دهند که رضایت مشتریان را به حداکثر برسانند.

با توجه به رقابت شدید در بازار و تنوع گزینه‌های محصولات و خدمات مختلف که به مشتریان ارائه می‌شود، تحلیل رفتار مشتریان و انتخاب بهترین روش بازاریابی بر اساس این تحلیل‌ها، به عنوان یک عامل بسیار مهم برای بقای شرکت‌ها در نظر گرفته می‌شود. تحلیل مشتریان یک فرآیند است که از اطلاعات مرتبط با رفتار مشتریان برای کمک به تصمیم‌گیری‌های کلیدی کسب و کار استفاده می‌کند، از جمله تقسیم‌بندی بازار و پیش‌بینی رفتار خرید مشتریان. از سوی دیگر، شناخت صحیح و دقیق از رفتار مشتریان، بخش حیاتی مدیریت ارتباط با مشتری محسوب می‌شود. مدیریت ارتباط با مشتری با شروع از شناخت مشتریان آغاز می‌شود، که این به معنای تعیین هدف گروهی از افراد که به احتمال زیاد در آینده مشتری شرکت ما خواهند شد یا شناسایی مشتریانی است که برای کسب و کار بهره‌وری ایجاد می‌کنند (خرم، ۲۰۱۷).

همراه با توسعه فناوری‌های اطلاعات و ارتباطی، مفاهیم مدیریت ارتباط با مشتری به ارمغان آورده استفاده گسترده از داده و اطلاعات شده است. این داده‌ها نیاز به داده‌کاوی را به عنوان یکی از ابزارهای استخراج دانش و الگوهای موجود در آن‌ها ایجاد کرده است. روش‌های داده‌کاوی معمولاً در مدیریت ارتباط با مشتری^۳، برای تحلیل رفتار مشتریان، از جمله ترجیحات و تقاضاهای آنان، به کار می‌روند. سازمان‌ها از طریق داده‌کاوی می‌توانند الگوهای رفتاری مشتریان، نیازهای آنان، و

³ CRM (Customer Relationship Management)

ارتباطات پنهان میان داده‌ها را درک کنند. این نگرش به آن‌ها امکان می‌دهد تا منابع خود را بهتر مدیریت کرده و به درستی نیازهای مشتریان را تامین کنند.

یکی از مدل‌های پرطرفدار در زمینه خوشه‌بندی و تحلیل ارزش مشتری، مدل RFM نام دارد. مدل RFM یک ابزار مبتنی بر رفتار مشتریان است که برای تجزیه و تحلیل رفتارهای آنها، با توجه به سوابق خرید گذشته‌شان در پایگاه داده، به کار می‌رود. این مدل شامل سه متغیر است: تازگی (Recency): نمایانگر زمانی است که از آخرین خرید مشتری گذشته است. فراوانی (Frequency): اشاره به تعداد کل خریدهای مشتری در یک دوره زمانی مشخص دارد. ارزش پولی (Monetary Value): میانگین مبلغی که مشتری در یک بازه زمانی خاص صرف کرده است.

این سه متغیر عملکردی مشتری را نمایان می‌کنند و برای خوشه‌بندی مشتریان بر اساس ویژگی‌هایشان به کار می‌روند.

اهداف این پژوهش به شکل زیر می‌باشند:

هدف ۱: تعیین بهترین تعداد خوشه برای بخش‌بندی مشتریان با استفاده از شاخص‌های بدست آمده.

هدف ۲: تحلیل ویژگی‌های مشتریان در هر بخش و بررسی رفتار خرید مجدد آن‌ها با توجه به شاخص‌های مدل پژوهش، و پیشنهاد سیاست‌های مناسب برای این دسته از مشتریان.

هدف ۳: مقایسه کارایی طبقه‌بندی‌کننده‌ها به منظور تشخیص دسته مشتریان با توجه به دقت و سرعت و انتخاب بهترین طبقه‌بندی‌کننده.

با توجه به اهداف پژوهش، سوالات این پژوهش به طور اجمالی در ذیل بیان شده اند:

۱. چه تعداد خوشه مناسب برای بخش‌بندی مشتریان با استفاده از شاخص‌های ماست؟
۲. ویژگی‌هایی که در هر بخش از مشتریان وجود دارد چیستند؟ چگونه رفتار خرید مجدد آن‌ها است و چه سیاست‌هایی می‌توان برای این دسته از مشتریان پیشنهاد داد؟
۳. کدام طبقه‌بندی‌کننده برای تشخیص دسته مشتریان با دقت و سرعت بالاتری عمل می‌کند و چه طبقه‌بندی‌کننده‌ای به عنوان بهترین انتخاب معرفی می‌شود؟

مبانی نظری و پیشینه پژوهش

رفتار مشتریان در اقتصاد

مشتریان در واقعیت، نقش فعالی را در تغییر پویایی بازار بازی می‌کنند. این به معنایی دیگر نشان می‌دهد که بازار به یک مکان تبدیل شده است که در آن مشتریان علاوه بر ایفای نقش در ایجاد

رقابت، ارزش نیز خلق می‌کنند. بنابراین، شرکت‌ها در این دوران نه تنها تمایل دارند ارتقاء ارزش طول عمر مشتریان را داشته باشند، بلکه به دنبال ایجاد مشارکت با مشتریان هستند (گوان⁴، ۲۰۲۱). اهمیت مشارکت مشتری در ایجاد ارزش برای شرکت‌های خدماتی بیشتر تأکید شده است، و نقش مشتری در شرکت‌های خدماتی که محصولات آن‌ها کمتر ناملموس هستند، بسیار مهم است. در دنیای پر رقابت کنونی کسب و کارها، موفقیت خدمات وابسته به تعاملات کامل با مشتریان است تا فعالیت‌های مرتبط با خدمات به چشم مشتری جذابیت ایجاد کند.

رفتار خرید مشتری

رفتار خرید به کلیه اقدامات و تصمیم‌گیری‌هایی اشاره دارد که یک مشتری پیش از خرید محصول یا خدمات موردنظر خود انجام می‌دهد. این عملیات ممکن است به صورت آنلاین در فضای وب و شبکه‌های اجتماعی یا به صورت حضوری در فروشگاه‌های معمولی انجام شود. رفتار خرید شامل اقداماتی مانند جستجو در موتورهای جستجو، مشاوره‌گیری از دیگران، مشاهده پست‌های شبکه‌های اجتماعی و بررسی مختلف دیگر است (صالحی، ۲۰۱۸).

مدیریت ارتباط با مشتری

این رویکرد در نظر می‌گیرد که CRM یک ابزار حیاتی است که با استفاده از فناوری‌های اطلاعاتی، اطلاعات مشتری را تجزیه و تحلیل کرده و آنها را جمع‌آوری می‌کند. سپس این اطلاعات را به اقدامات مشخصی برای مدیران تبدیل می‌کند تا بتوانند به خوبی پاسخگوی تقاضای مشتریان باشند. از این رو، CRM به طور گسترده از تکنیک‌های داده‌کاوی برای دستیابی به روابط مؤثر با مشتری استفاده می‌کند.

مدیریت ارتباط با مشتری، عملاً یک ابزار است که به شناسایی بهتر مشتریانی که سودآوری بیشتری دارند و ارزش رابطه بلندمدت با آن‌ها را دارند، کمک می‌کند. این ابزار ارزشمند به ما در مدیریت ارتباطات مؤثر با مشتریان و افزایش سودآوری در طولانی‌مدت کمک می‌کند (آذر، ۲۰۲۱).

مدل RFM

بخش‌بندی بازار، امروزه یکی از رکن‌های اصلی بازاریابی برای هر کسب‌وکاری است. با این حال، گاهی اوقات شرکت‌ها هنگام دسته‌بندی، دچار اشتباهاتی می‌شوند. یکی از این اشتباهات، اشتباه‌گرفتن دسته‌بندی با جمعیت‌شناسی است. اشتباه دیگر می‌تواند عدم مشخص کردن هدف دسته‌بندی باشد؛ اما بی‌تردید، بزرگ‌ترین اشتباه، دسته‌بندی کردن با روش‌های سنتی از قبیل دموگرافی (جمعیت، جغرافیا و ...) است. دلیل اصلی، پیشرفت فناوری اطلاعات است. این پیشرفت باعث شده که مردم به خدمات، محصولات و امکانات بیشتری دسترسی داشته باشند. در حقیقت،

⁴ Guan

رفتارهای مشتریان و انتظاراتشان در مورد چگونگی، زمان و مکان خرید و استفاده‌ی از محصول، تغییر کرده است. این تغییرات مشتریان، سبب شده تا شرکت‌ها با چالش‌های بزرگی در ریتنشن مارکتینگ مواجه شوند و به طور مداوم تلاش کنند تا خودشان را متناسب با تغییرات مشتریان، تغییر داده و هماهنگ کنند و برای این کار، لازم است تا دسته‌بندی‌های دقیق‌تری از مشتریان خودشان داشته باشند. روش RFM یکی از قدیمی‌ترین و کارآمدترین روش‌های تقسیم‌بندی مشتریان در جهان است. RFM مخفف Recency (به معنای آخرین زمانی که مشتری خرید کرده)، Frequency (به معنای تعداد دفعاتی که مشتری خرید کرده) و Monetary (به معنای مقدار مبلغی که مشتری خرید کرده) است (کیت، ۲۰۱۷).

پیشینه پژوهش

- پژوهشی توسط بریانت در سال ۲۰۱۱ انجام شده است و در آن از مدل RFM و الگوریتم DTC برای خوشه‌بندی و سپس طبقه‌بندی مشتریان فروشگاه ورزشی استفاده شده است. این پژوهش به دقت ۸۱ درصد در طبقه‌بندی مشتریان دست‌یافته است.
- پژوهش دیگری نیز توسط هوآنگ و همکاران در سال ۲۰۲۱ انجام شده و به استفاده از مدل RFM در صنایع خدماتی مختلف برای ایجاد ارزش مشتری پرداخته است. در این پژوهش، تقسیم‌بندی مشتریان در صنایع خدماتی مختلف بر اساس رفتار خریدشان با استفاده از الگوریتم DTC انجام شده است.
- مقاله‌ای دیگر توسط هلت و همکاران در سال ۲۰۲۱ انجام شده و به پیش‌بینی ارزش مشتریان به ازای هر محصول با استفاده از مدل RFM/P و مدل BG/BB می‌پردازد. این پژوهش از ترکیب دیدگاه‌های مشتری و محصول با استفاده از مدل RFM/P برای بهبود پیش‌بینی ارزش مشتریان استفاده می‌کند.
- پژوهش که توسط گوانی و همکاران در سال ۲۰۲۰ انجام شده، به تصویر پرتره اعضای بازار براساس داده‌کاوی می‌پردازد. این پژوهش از مدل LRFMC و الگوریتم‌های K-means و KNN برای تشکیل پرتره اعضا در مراکز خرید با استفاده از محاسبه ارزش عضویت استفاده می‌کند.
- در پژوهشی دیگر که توسط چائو و همکاران در سال ۲۰۲۱ انجام شده، یک رویکرد ترکیبی برای پروفایل‌سازی مشتری در ویدئو در مورد خدمات تقاضا با استفاده از مدل LRFMP و الگوریتم‌های K-means و Apriori ارائه می‌دهد. این پژوهش چهار گروه مشتری اصلی را شامل پرمصرف، کم مصرف، وفادار کم مصرف و بی‌وفا تشخیص می‌دهد.

- در همین راستا یک پژوهش توسط وو و همکاران در سال ۲۰۲۱ انجام شده و به شناسایی ارزش کاربر بر اساس مدل RFM بهبود یافته و الگوریتم ++K-Means برای تجزیه و تحلیل داده‌های پیچیده می‌پردازد. این پژوهش تعیین بهتر ارزش مشتری با استفاده از مدل RFM بهبود یافته و طبقه‌بندی بهتر آن را مورد بررسی قرار می‌دهد.
- نهایتاً پژوهشی توسط بشردوست و همکاران در سال ۱۴۰۱ انجام شده و به تحلیل خوشه‌بندی مشتریان بر اساس مدل WRFM با استفاده از رویکرد داده‌کاوی غیر نظارتی می‌پردازد. این پژوهش تقسیم‌بندی مشتریان محصولات بهداشتی را با استفاده از الگوریتم K-means انجام می‌دهد.
- مطالعه ای دیگر توسط رحیم و همکاران در سال ۱۴۰۰ انجام شده و بر رفتار خرید مجدد مشتریان مبتنی بر مدل RFM تمرکز دارد. در این پژوهش، از مدل‌های MLP، DTC و SVM برای بررسی قصد خرید مشتریان با استفاده از RFM، خوشه‌بندی و سپس طبقه‌بندی مشتریان استفاده شده است.
- پژوهش دیگری توسط چشمی و همکاران در سال ۱۳۹۹ انجام شده و به پیش‌بینی گردش مالی مشتریان با استفاده از شبکه‌های عصبی بازگشتی ارتباطات بی‌سیم و محاسبات سیار می‌پردازد. این پژوهش از مدل جدید پیشنهادی و شبکه عصبی برای انجام تحلیل جامع‌تر از گردش مالی مشتریان استفاده می‌کند.
- پژوهش دیگری توسط جهانیان و همکاران در سال ۱۳۹۷ انجام شده و به به‌کارگیری روش هزینه‌یابی مبتنی بر فعالیت زمان‌مبنا به‌منظور رتبه‌بندی مشتریان سودآور با استفاده از مدل RFM و الگوریتم K-means می‌پردازد.

روش پژوهش

این پژوهش به عنوان یک تحقیق کاربردی توسعه‌ای مطرح می‌شود و طرح آن به صورت پس‌رویدادی انجام شده است. در این نوع تحقیق، از داده‌های موجود در گذشته به عنوان منبع اصلی استفاده می‌شود. این روش معمولاً زمانی مورد استفاده قرار می‌گیرد که داده‌ها از محیطی طبیعی یا واقعه‌ای که بدون مداخله مستقیم پژوهشگر رخ داده است، بدست آمده باشند.

پژوهش حاضر از یک مجموعه داده عمومی استفاده می‌کند که شامل تراکنش‌های یک واحد خدماتی پلیس در یک بازه زمانی یک ساله از سال ۱۴۰۱ تا ۱۴۰۳ می‌شود. هر تراکنش این مجموعه داده شامل ۸ ویژگی است که توضیحات دقیقتری در جدول زیر آورده شده است. این پایگاه داده استاندارد است که در سطح بین‌المللی مورد استفاده قرار می‌گیرد و علاوه بر اطلاعات تراکنش‌ها، اطلاعات خدمات نیز در آن وجود دارد. این مجموعه داده با توجه به تنوع داده‌ها، مناسب برای

تحلیل‌های مشابه در زمینه‌های مختلف است (پکر^۶، ۲۰۱۷). برای تجزیه و تحلیل داده‌ها، از نرم‌افزار متلب بهره‌گیری شده است.

پس از مراحل پاکسازی و نرمال‌سازی داده‌ها، ویژگی‌های LRFM از داده‌ها استخراج می‌شوند و سپس داده‌ها به چندین خوشه تقسیم می‌شوند. سپس برای تشخیص اینکه مشتریان جدید به کدام خوشه تعلق می‌گیرند، از سه الگوریتم طبقه‌بندی‌کننده استفاده می‌شود. این الگوریتم‌ها شامل پرسپترون چندلایه، ماشین بردار پشتیبان و درخت تصمیم هستند (چاو^۷، ۲۰۲۱). فرآیند طبقه‌بندی به این صورت انجام می‌شود که ابتدا با استفاده از بخشی از داده‌های خوشه‌بندی‌شده، الگوریتم‌های طبقه‌بندی آموزش داده می‌شوند. سپس با استفاده از داده‌های تست، دقت عملکرد این الگوریتم‌ها ارزیابی می‌شود. در نهایت، این طبقه‌بندی‌کننده‌ها مشتریان جدید را در یکی از دسته‌ها قرار می‌دهند و می‌توان تعیین کرد که کدام سیاست‌های تشویقی برای هر دسته از مشتریان مناسب است.

جدول ۱. عناصر یک تراکنش و شرح استفاده از آنها

نام عنصر	شرح
User_ID	شماره منحصر به فرد مشتری
Transaction_ID	شماره منحصر به فرد فاکتور
Transaction_Time	تاریخ و ساعت تراکنش
ItemC	شماره منحصر به فرد خدمت
ItemDesc	نام و شرح خدمت
ServicePurchased	تعداد خدمات فروش رفته
ServiceCost	قیمت واحد خدمت

شکل زیر نمایی کلی از الگوی مفهومی پژوهش و مراحل پردازش‌های انجام شده روی داده‌های خام اولیه را نمایش می‌دهد که در ادامه مورد بررسی قرار می‌دهیم.



⁶ Peker

⁷ Chao

شکل ۱. الگوی مفهومی پژوهش

الگوریتم K-Means برای خوشه بندی

خوشه‌بندی مشتریان یکی از مباحث مهم و پرکاربرد در زمینه مدیریت ارتباط با مشتری (CRM) و داده‌کاوی است. در این مفهوم، ما سعی داریم مشتریان را بر اساس ویژگی‌های مشترک و شباهت‌هایی که در داده‌ها و اطلاعات مشتریان دارند، به گروه‌های مختلف تقسیم کنیم. هر گروه یا خوشه از مشتریان شامل مجموعه‌ای از مشتریان با ویژگی‌ها و رفتارهای مشابه است، و از سایر گروه‌ها متفاوت می‌باشد (مینایی، ۲۰۲۱).

الگوریتم SVM

روش SVM یا ماشین بردار پشتیبان (Support Vector Machine) یک الگوریتم یادگیری ماشینی است که برای مسائل دسته‌بندی و رگرسیون استفاده می‌شود. SVM به عنوان یکی از الگوریتم‌های معروف و قدرتمند در زمینه یادگیری ماشینی شناخته می‌شود و در بسیاری از برنامه‌ها مورد استفاده قرار می‌گیرد، از جمله تشخیص تصاویر، تصویربرداری پزشکی، تحلیل متن، معماری و طراحی و غیره (گونیی^۸، ۲۰۲۰).

الگوریتم MLP

شبکه عصبی چند لایه‌ای یا به اختصار MLP (Multilayer Perceptron) یکی از مدل‌های مهم و پرکاربرد در زمینه یادگیری عمیق (Deep Learning) است. این مدل به عنوان یک نوع از شبکه‌های عصبی مصنوعی به تحلیل داده‌های پیچیده و توانایی دسته‌بندی و پیش‌بینی مسائل مختلف از جمله تصویربرداری، پردازش متن، تشخیص الگو و موارد مشابه پرداخته و در بسیاری از برنامه‌ها و حوزه‌ها مورد استفاده قرار می‌گیرد (ویی^۹، ۲۰۱۹).

الگوریتم DTC

الگوریتم درخت تصمیم (Decision Tree Classifier یا DTC) یکی از الگوریتم‌های یادگیری ماشینی است که برای مسائل دسته‌بندی مورد استفاده قرار می‌گیرد. این الگوریتم به شکل یک درخت تصمیم تولید می‌شود که از تصمیم‌ها و شرایط مختلف برای تقسیم داده‌ها به دسته‌های مختلف تشکیل شده است (بشردوست، ۲۰۲۲).

یافته های پژوهش

پردازش اولیه روی داده ها

⁸ Guney

⁹ Wei

شکل اولیه مجموعه داده‌ها به وسیله انواع مختلفی از اطلاعات آلوده شده است که ممکن است به مشتری و موضوعات مرتبط با او نسبت نداشته و در واقعیت می‌تواند مانع استخراج الگوهای رفتاری مفید شود. این مجموعه داده شامل ۱۲۶۴۱۱ تراکنش مختلف می‌باشد، اما بعد از اجرای مراحل پیش‌پردازش مورد نیاز، تنها ۲۹۵۶۰ تراکنش مناسب برای تحلیل‌های آتی باقی مانده‌اند. تراکنش‌های حذف شده نیز در جدول زیر ذکر شده است. مجموعه داده‌های مورد استفاده در آزمایشات پس از انجام پیش‌پردازش‌های مورد نیاز، شامل ۲۷۱۲۴ تراکنش مختلف با ۳۱ کد خدمت یکتا و ۸۷۵۴ شماره مشتری متفاوت می‌باشد. تعداد کل فاکتورها در داده‌های پیش‌پردازش شده ۱۲۴۵۷ مورد است. بیشترین تعداد تراکنش‌ها به کشور انگلستان تعلق دارد و پس از آن، آلمان و فرانسه با تراکنش‌های بیشتری آمده‌اند.

جدول ۲. انواع تراکنش‌های حذف شده

نوع تراکنش	علت حذف
تراکنش تکراری	تراکنش‌هایی که در داده‌ها تکرار شده‌اند
تراکنش برگشتی	تراکنش‌هایی که تعداد آن‌ها منفی بوده و نشان‌دهنده لغو سفارش هستند
تراکنش ناقص	تراکنش‌های که شماره فاکتور یا شماره مشتری ندارند
تراکنش ناهمگون	تراکنش‌هایی که تعداد یا قیمت نامعقول دارند

خوشه‌بندی مشتریان بر اساس LRFM

شش ویژگی کلیدی در تشخیص رفتار مشتری از این شش عنصر استخراج شده و مجموع آن‌ها بردار ویژگی هر مشتری را تشکیل می‌دهد که در جدول زیر ارائه شده است:

جدول ۳. تعریف شاخص‌های LRFM

متغیر	نحوه محاسبه	شرح
تازگی خرید (R)	فاصله زمانی بین زمان آخرین خرید مشتری با تاریخ ۲۸ اسفند ۱۴۰۱ که به عنوان پایان مشاهدات در نظر گرفته شده (برحسب روز)	بیانگر فاصله زمانی بین آخرین خرید مشتری و پایان مشاهدات است. این متغیر هر چه کمتر باشد بهتر است.
فراوانی خرید (F)	شمارش تعداد فاکتورهای منحصر به فرد هر مشتری	بیانگر تعداد فاکتورهای خرید هر مشتری یا دفعات مراجعه وی به خرده‌فروشی است.
ارزش نقدی خرید (M)	مجموع حاصلضرب هزینه هر کالا در تعداد خریداری شده	بیانگر کل مبلغ هزینه شده توسط مشتری در بازه بررسی شده است.

طراحی مدل طبقه‌بندی مشتریان با استفاده از هوش و ...

طول ارتباط مشتری (L)	اختلاف زمانی اولین و آخرین تراکنش مشتری	بیانگر طول مدت زمانی است که مشتری با فروشگاه در ارتباط بوده و طبیعتاً هر چه مقدار بالاتری داشته باشد، وفاداری مشتری را نشان می‌دهد.
----------------------	---	---

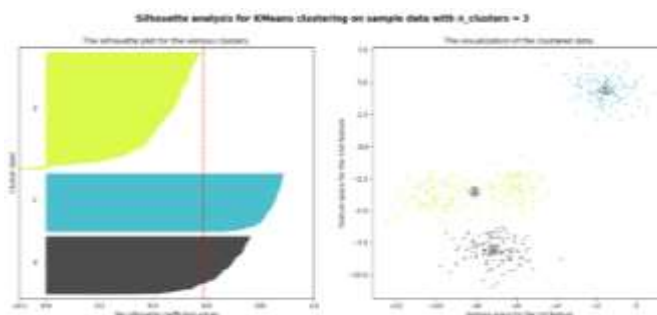
پس از انجام نرمال‌سازی یکنواخت، مقدار بهینه K برای خوشه‌بندی با استفاده از سه روش محاسبه $Silhouette$ ، $DaviesBouldin$ و $CalinskiHarabasz$ تعیین و در جدول زیر نمایش داده شده است.

جدول ۴. تعیین مقدار K بهینه با استفاده از سه روش $DaviesBouldin$ ، $Silhouette$ و

CalinskiHarabasz

K	Silhouette	DaviesBouldin	CalinskiHarabasz
۲	۰,۶۹۲۶	۰,۷۵۷۱	۵۸۹۵,۲
۳	۰,۶۹۹۵	۰,۶۳۴۵	۶۸۹۷,۷
۴	۰,۶۲۸۲	۰,۸۳۴۳	۶۶۳۶,۳
۵	۰,۵۹۰۷	۰,۸۷۸۷	۶۰۴۸,۰
۶	۰,۶۰۲۸	۰,۸۶۶۷	۵۹۵۳,۳
۷	۰,۶۰۵۱	۰,۸۵۷۵	۵۶۹۴,۰
۸	۰,۵۷۵۲	۰,۸۵۸۴	۵۵۹۸۷
۹	۰,۵۲۸۲	۰,۹۶۶۵	۵۳۲۵,۵
۱۰	۰,۵۵۴۵	۰,۹۵۲۵	۵۱۲۸,۵

شکل زیر نمودار $Silhouette$ را برای مقدار $K = 3$ نمایش می‌دهد. نتایج خوشه‌بندی نشان می‌دهند که ۱۵۸۳ مشتری در خوشه اول، ۱۷۶۱ نفر در خوشه دوم و باقی افراد هم در خوشه سوم قرار گرفته‌اند. این توزیع نسبتاً مناسبی از مشتریان را بین سه خوشه نشان می‌دهد. مقایسه ویژگی‌های استخراج شده بین این سه خوشه در جدول زیر ارائه شده است.



شکل ۲. نمودار $Silhouette$ به ازای $K = 3$

جدول ۵. میانگین مقادیر ویژگی‌ها برای خوشه‌های ایجاد شده

R	F	M	L
---	---	---	---

۰/۱۴۴۱	۰/۰۰۴۸	۰/۰۰۲۷	۰/۱۰۰۷	خوشه اول
۰/۰۹۵۸	۰/۰۳۵۵	۰/۰۱۳۳	۰/۷۳۳۴	خوشه دوم
۰/۶۸۵۶	۰/۰۰۲۴	۰/۰۰۱۷	۰/۰۵۱۹	خوشه سوم
۰/۱۲۰۷	۰/۰۰۷۸	۰/۰۰۳۳	۰/۱۷۰۵	کل داده‌ها

– طبقه‌بندی مشتریان

پرسپترون چندلایه

بر اساس محاسبات انجام شده، تعداد ۸ نورون در لایه میانی به عنوان مقدار مناسب‌تر به نظر می‌آید. افزایش تعداد نورون‌های لایه میانی علاوه بر اینکه تأثیر چندانی بر افزایش دقت ندارد، منجر به افزایش زمان آموزش و محاسبات شبکه می‌شود (برادران، ۲۰۱۵).

جدول ۶. میانگین دقت پیش‌بینی با MLP به ازای تعداد نورون‌های مختلف

تعداد نورون لایه میانی	دقت
۱	٪۸۲٫۹
۲	٪۹۰٫۳۴
۳	٪۹۸٫۰۶
۴	٪۹۹٫۶۹
۵	٪۹۹٫۷۲
۶	٪۹۹٫۷۸
۷	٪۹۹٫۸۱
۸	٪۹۹٫۸۱

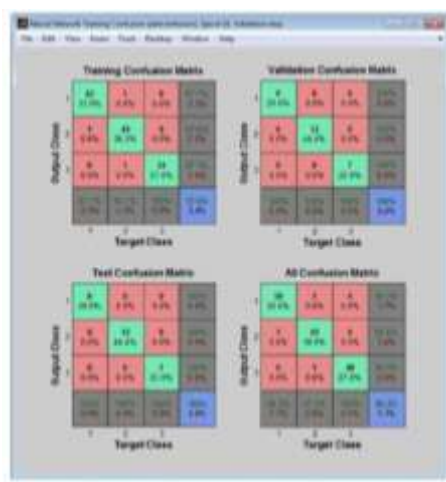
در شکل زیر، ساختار MLP استفاده شده برای پیش‌بینی مشتریان بر اساس ویژگی‌های رفتاری را نمایش می‌دهد. لایه ورودی این MLP شامل ۶ نورون است که تعداد آنها بر اساس تعداد ویژگی‌ها تعیین می‌شود. تعداد نورون‌های لایه میانی با توجه به دقت نتایج پس از ۶ محاسبه تعیین شده و تعداد نورون‌های لایه خروجی نیز با توجه به سه کلاس بودن نتایج به عدد ۳ تعیین شده است.

طراحی مدل طبقه‌بندی مشتریان با استفاده از هوش و ...



شکل ۳. ساختار MLP مورد استفاده برای پیش‌بینی رفتار مشتریان

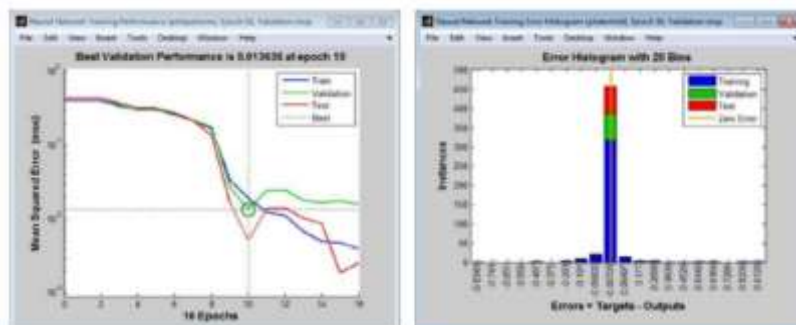
شکل زیر نشان‌دهنده ماتریس Confusion برای یک بار اجرای طراحی انتخابی برای شبکه‌ی عصبی می‌باشد. ۷۰ درصد از داده‌های ورودی برای آموزش، ۱۵ درصد برای تست و ۱۵ درصد برای Validation مورد استفاده قرار گرفته است. تعداد خطای پیش‌بینی برای هر کدام از دسته‌های Train، Validation، Test و مجموع هر سه دسته در جدول قابل مشاهده است. طبق اجرای انجام شده در حالت Validation که بیشترین درصد خطا دیده شده هم درصد خطا تنها حدود ۰,۳ درصد است.



شکل ۴. ماتریس Confusion برای یک نمونه اجرای طراحی انتخابی برای شبکه‌ی عصبی

میزان بهبود کارایی شبکه در هر Epoch در شکل زیر ارائه شده است. در بررسی صورت گرفته آموزش تا جایی ادامه پیدا کرده که در شش Epoch متوالی بهبودی در میزان کارایی شبکه مشاهده

نشود. بنابراین در بررسی نمایش داده شده در شکل زیر پس از ۱۰ دور، میزان کارایی شبکه ثابت شده است.



شکل ۵. میزان بهبود کارایی شبکه در هر Epoch و همگرا شدن نتایج

ماشین بردار پشتیبان

SVM به عنوان یک الگوریتم معتبر در دسته‌بندی داده‌ها مورد استفاده قرار می‌گیرد. این الگوریتم برای داده‌هایی که به طور خطی قابل تفکیک هستند، بسیار کارآمد عمل می‌کند. اما وقتی داده‌ها به صورت خطی قابل تفکیک نیستند، می‌توان از مفهوم کرنل استفاده کرد تا این جداسازی را انجام دهیم. در این پژوهش، به منظور انتخاب مناسب‌ترین تابع کرنل برای SVM، سه نوع کرنل متداول بررسی شده‌اند. این سه کرنل عبارت‌اند از: کرنل خطی، کرنل گوسی (RBF)، کرنل چندجمله‌ای (چشمی^{۱۰}، ۲۰۲۱).

برای انتخاب بهترین تابع کرنل، هر یک از این کرنل‌ها به صورت جداگانه اجرا شده و دقت نتایج برای هر کدام از آنها در ۱۰۰ بار اجرا محاسبه شده است. سپس میانگین دقت به صورت جدولی نمایش داده شده است. این اطلاعات به تحقیق‌گران کمک می‌کند تا تابع کرنل مناسب‌تر را برای مسئله‌ی خود انتخاب کنند.

جدول ۷. میانگین دقت برای هر کرنل در ۱۰۰ بار اجرا

Accuracy	Formula	Kernel Function Name
%۹۹/۵۶	$G(x_j, x_k) = \exp(-\ x_j - x_k\ ^2)$	گوسی
%۹۹/۳۱	$G(x_j, x_k) = x_j' x_k$	خطی
%۹۹/۶۴	$G(x_j, x_k) = (1 + x_j' x_k)^q$	چند جمله ای

درخت تصمیم

در این پژوهش، سه روش مورد ارزیابی قرار گرفته‌اند که به ترتیب درخت CART، یادگیری تصمیم و درخت تصمیم C4.5 هستند. با انجام آزمایشات، به نتیجه رسیده‌ایم که روش ساخت درخت CART بالاترین دقت را با درصد ۹۹،۱۶ ارائه می‌دهد.

¹⁰ Chashmi

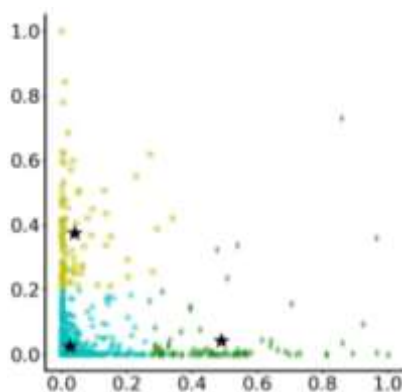
جدول ۸. مقایسه میانگین دقت درخت تصمیم با سه روش **Standard CART**، **Curvature** و

Interaction-Curvature	
Accuracy	Split Predictor Algorithm
%۹۹/۱۶	Standard CART
%۹۸/۵۳	Curvature test
%۹۸/۵۵	Interaction-test

مدل وسپ



شکل ۶. مدل وسپ



شکل ۷. پراکندگی خوشه‌ها در مختصات دفعات خرید (F) نسبت به تازگی خرید (R)

این نمودار نمایش دهنده پراکندگی سه خوشه مشخص شده در مختصات دفعات خرید (F) نسبت به تازگی خرید (R) است. این نمودار به وضوح نشان می‌دهد که مشتریان در سه خوشه مختلف با توجه به الگوی خرید خود جایگاه‌های مختلفی دارند. در جدول زیر، دقت بهترین معماری تعیین شده برای هر یک از طبقه‌بندی‌کننده‌ها مقایسه شده است. در کل، نتایج تحلیل خوشه‌بندی و طبقه‌بندی مشتریان نشان می‌دهد که استفاده از ویژگی‌های مشخص و دقیق می‌تواند در تشخیص الگوهای خرید مشتریان و پیش‌بینی رفتار آینده آن‌ها مؤثر باشد و به مدیران ابزارهای قدرتمندی برای تصمیم‌گیری فراهم کند.

جدول ۹. میزان دقت بهترین معماری تعیین شده برای هر طبقه‌بندی کننده

طبقه‌بندی کننده	معماری مورد انتخاب	دقت
MLP	یک لایه پنهان با ۸ نورون	۹۹,۲۳٪
SVM	کرنل چندجمله‌ای	۹۹,۳۴٪
DTC	Standard CART	۹۹,۱۹٪

نتیجه گیری

در پژوهش حاضر، به بررسی و تحلیل روش‌های مختلفی پرداختیم که برای مدیریت بهتر ارتباط با مشتریان در صنعت خدمات از آن‌ها استفاده می‌شود. از جمله این روش‌ها می‌توان به مدل‌های طبقه‌بندی مشتریان، استفاده از هوش مصنوعی، تجزیه و تحلیل داده‌های خدمات، و بهبود عملکرد خدمات اشاره کرد. یکی از قسمت‌های مهم پژوهش، تحلیل داده‌های موجود در صنعت خدمات پلیس +۱۰ بود. از داده‌کاوی و تحلیل خدمات برای استخراج اطلاعات مفید از داده‌های تراکنش‌های مشتریان استفاده کردیم. در این فرآیند، معیارهای RFM و LRFM را برای تحلیل رفتار مشتریان مورد استفاده قرار دادیم. بر اساس یافته‌های به دست آمده از تحلیل داده‌ها و استفاده از مدل‌های مختلف، به نتیجه رسیدیم که ایجاد مدل‌های طبقه‌بندی مشتریان با استفاده از الگوریتم‌هایی چون K-Means، درخت تصمیم، MLP، و SVM می‌تواند به مدیران کسب‌وکار کمک کند تا مشتریان را به دسته‌های مختلفی تقسیم کنند و سیاست‌های تشویقی مناسب برای هر دسته را اجرا کنند (وو، ۲۰۲۱).

بهترین تعداد خوشه برای بخش‌بندی مشتریان، ویژگی‌های مشتریان در هر بخش، و انتخاب بهترین طبقه‌بندی کننده برای تشخیص دسته مشتریان را بر اساس دقت و سرعت بررسی کردیم. این پژوهش به مدیران و کارشناسان صنعت خدمات ابزارهای قدرتمندی ارائه کرده است تا از داده‌های مشتریان بهره‌برداری کرده و بهبودهای موثرتری در ارتباط با مشتریان و عملکرد کسب‌وکار خود ایجاد کنند. این پایان‌نامه ارتقاء اهمیت داده‌کاوی و هوش مصنوعی در مدیریت ارتباط با مشتریان را تأکید داده و اساسی برای تصمیم‌گیری‌های استراتژیک در این حوزه فراهم کرده است.

همچنین نتایج نشان داد که در میان کرنل‌های مورد بررسی برای طبقه‌بندی کننده SVM، کرنل چندجمله‌ای بهترین نتیجه را با دقت ۹۹,۶۴ درصد در پی دارد، در حالیکه رحیم و همکاران که به طبقه‌بندی رفتار مشتریان با معیار RFM پرداخته بودند، با استفاده از طبقه‌بندی کننده SVM به دقت ۹۴ درصد و یوریسیچ و همکاران با استفاده از معیار RFM و با کمک طبقه‌بندی کننده‌های SVM و DTC به دقت ۹۳,۳۳ رسیدند.

در نهایت نتایج پژوهش نشان داد که طبقه‌بندی مشتریان با معیار **LRFM** و استفاده از درخت تصمیم‌گیری **CART**، دقت ۹۹٫۱۶٪ را در پی دارد. در حالیکه بریانت، ۲۰۱۱ و هوانگ و همکاران ۲۰۲۱ با استفاده از معیار **RFM** و طبقه بندی کننده درخت تصمیم به دقت کمتری رسیده بودند و بریانت میزان دقت طبقه بندی پژوهش خود را ۸۱ درصد گزارش کرده بود.

بنابراین نتایج هر سه طبقه‌بندی‌کننده نشان داد که افزودن ویژگی‌های **S**، **L** و **P** تأثیر مثبتی بر دقت طبقه‌بندی‌کننده خواهد گذاشت. مقایسه سرعت و دقت طبقه‌بندی‌کننده‌ها نشان می‌دهد که درخت تصمیم گرچه دقت کمی پایبندی از روش‌های دیگر دارد، اما به دلیل ساختار ساده‌ای که دارد با سرعت خیلی بالاتری نسبت به دو روش دیگر به پیش‌بینی خوشه مشتریان می‌پردازد.

منابع

- Azar, A., Khodadad Hosseini, S., Safari Kahreh, M. (2021). Customer Lifetime Value (CLV) Analysis for Customer Segmentation and Profitability Management: A Case Study of an Iranian Bank. *Management Research in Iran*, 18(4), 87-110, <https://dorl.net/dor/20.1001.1.2322200.1393.18.4.5.2>, [in Persian].
- Bagheri mazraeh, N., Daneshvar, A., & madanchi zaj, M. (2022). Development a new ensemble learning approach for stock portfolio selection using multiclass SVM and genetic algorithm. *Financial Engineering and Portfolio Management*, 13(50), 282-305, [in Persian].
- Baradaran, V., & Biglari, M. (2015). Customer segmentation in Fast Moving Consumer Goods (FMCG) Industries by using developed RFM model. *Journal of Business Management*, 7(1), 23-42, <https://doi.org/10.22059/jibm.2015.51674>, [in Persian].
- Bashardoust, O., Asgharizadeh, E., & Afshar Kazemi, M. A. (2022). Customers Clustering Analysis Based on WRFM Model Using Non-Supervisory Data Mining Approach (Case study of hygienic and cosmetic products). *Modern Research in Decision Making*, 7(1), 198-223, <https://dorl.net/dor/20.1001.1.24766291.1401.7.1.7.2>, [in Persian].
- Birant, D. (2011). Data mining using RFM analysis. In *Knowledge-oriented applications in data mining*. IntechOpen, <https://doi.org/10.5772/13683>.
- Chao, S. H., Chen, M. K., & Wu, H. H. (2021). An LRFM model to analyze outpatient loyalty from a medical center in taiwan. *SAGE Open*, 11(3), 21582440211031899, <https://doi.org/10.1177/21582440211031899>.
- Chashmi, A. J., Rahmati, V., Rezasoroush, B., Alamoti, M. M., Askari, M., & Khalili, F. H. (2021). Predicting customer turnover using recursive neural networks. *Wireless Communications and Mobile Computing*, 2021, 1-11, <https://doi.org/10.1155/2021/6623052>.
- Guan, J., Tang, C., & Ou, J. (2020). The portrait depiction of the market members based on data mining. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(07), 2059024, <https://doi.org/10.1142/S0218001420590247>.
- Guney, S., Peker, S., & Turhan, C. (2020). A combined approach for customer profiling in video on demand services using clustering and association rule mining. *Ieee Access*, 8, 84326-84335, <https://doi.org/10.1109/ACCESS.2020.2992064>.
- Heldt, R., Silveira, C. S., & Luce, F. B. (2021). Predicting customer value per product: From RFM to RFM/P. *Journal of Business Research*, 127, 444-453, <https://doi.org/10.1016/j.jbusres.2019.05.001>.
- Huang, H. H. (2021). Using RFM model to construct customer value by making segment in different service industries. *Journal of Pervasive Technology*, 1(1), 5-5.

- Jahanyan, S., Mahmoudsalehi, M., & Hosseini, M. (2018). Applying Time-Driven Activity-Based Costing (TDABC) for customer Profitability ranking. *Modern Research in Decision Making*, 3(1), 85-111, [in Persian].
- Kit, T. C., Firdaus, N., & Azmi, M. (2021). Customer profiling for Malaysia online retail industry using K-Means clustering and RM model. *International Journal of Advanced Computer Science and Applications*, 12(1), 106-113, <https://doi.org/10.14569/ijacsa.2021.0120114>.
- Khorram Kashkooli, M., Dehghani, M. (2017). Fault Detection, Identification and Isolation of South Pars Gas Turbine Using a Combined Method Based on the Data Mining Techniques, k-means, PCA and SVM. *TABRIZ JOURNAL OF ELECTRICAL ENGINEERING*, 47(2), 501-515, [in Persian].
- Minaei, B., afsar, A., houshdar mahjoub, R. (2021). Customer credit clustering for Present appropriate facilities. *Management Research in Iran*, 17(4), 1-24, <https://dorl.net/dor/20.1001.1.2322200.1392.17.4.1.1>, [in Persian].
- Peker, S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: a case study. *Marketing Intelligence & Planning*, 35(4), 544-559, <https://doi.org/10.1108/mip-11-2016-0210>.
- Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. *Journal of Retailing and Consumer Services*, 61, 102566, <https://doi.org/10.1016/j.jretconser.2021.102566>.
- Salehi, M., & Farrokhi Pileroud, L. (2018). Profit management prediction using neural network and decision tree. *Financial Accounting and Auditing Research*, 10(37), 1-24, <https://dorl.net/dor/20.1001.1.23830379.1397.10.37.1.1>, [in Persian].
- Wu, J., Shi, L., Yang, L., Li, Y., Tsai, S. B., & Zhang, Y. (2021). User value identification based on improved RFM model and k-means algorithm for complex data analysis. *Wireless Communications and Mobile Computing*, 2021, 1-8, <https://doi.org/10.1155/2021/9982484>.
- Wei, J. T., Lin, S. Y., Yang, Y. Z., & Wu, H. H. (2019). The application of data mining and RFM model in market segmentation of a veterinary hospital. *Journal of Statistics and Management Systems*, 22(6), 1049-1065, <https://doi.org/10.1080/09720510.2019.1565445>