پژوهش‌های نوین در تصمیم‌گیری
دوره ۸، شماره ۲، تابستان ۱۴۰۲، صص ۱۲۳-۱۴۸
نوع مقاله: پژوهشی

بخش‌بندی و پیش‌بینی رفتار مشتریان براساس مدل RFM بهبودیافته (LRFMSP)

آمنه خدیور [*۱]، سهیلا مهمان نوازان [۳]

۱. دانشیار، گروه مدیریت، دانشکده علوم اجتماعی و اقتصادی، دانشگاه الزهرا، تهران، ایران
۲. دانشجوی دکتری مدیریت بازاریابی، دانشکده علوم اجتماعی و اقتصادی، دانشگاه الزهرا، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۱/۱۷ تاریخ پذیرش: ۱۴۰۲/۰۴/۱۰

چکیده

در سال‌های اخیر، با توسعه یادگیری ماشین و فناوری کلان داده، داده‌های کاربر به عنصر مهمی در فرآیند تولید شرکت‌ها تبدیل شده‌اند. با به کارگیری رویکردهای داده‌کاوی در داده‌های مشتریان، سازمان‌ها، الگوهای رفتاری مشتریان، نیازهای آن‌ها و ارتباط‌های پنهان داده‌ها را درک می‌کنند و براساس این الگوها بهتر می‌توانند در راستای برآورده ساختن نیاز مشتریان، منابع خود را به کار گیرند. خوشه‌بندی یکی از تکنیک‌های داده‌کاوی است که برای گروه‌بندی مشتریان متناسب با ویژگی‌های مختلف آن‌ها استفاده می‌شود. هدف اصلی این پژوهش خوشه‌بندی مشتریان بر اساس شاخص‌های LRFMSP و در نهایت طبقه‌بندی و پیش‌بینی رفتار خرید آن‌ها با استفاده از تکنیک‌های طبقه‌بندی درخت تصمیم (DTC)، پرسپترون چندلایه (MLP) و ماشین بردار پشتیبان (SVM) است. مطالعه‌ی صورت گرفته بر روی ۳۸۷۴۹۶ تراکنش مشتریان یک فروشگاه خرده‌فروشی در غرب اروپا طی بازه زمانی فوریه ۲۰۱۸ تا فوریه ۲۰۱۹ می‌باشد. هر تراکنش منتسب به مشتری بخشی از رفتار یک فرد است که بر روی مجموعه‌ای از معاملات مدل‌سازی می‌شود تا رفتار خرید مشتری را شکل دهد. انجام خوشه‌بندی K-means++ و تعیین K بهینه منتهی به مشخص شدن سه خوشه برای مشتریان گردید. همچنین آزمایش و بررسی طبقه‌بندی‌کننده‌ها نشان داد که طبقه‌بندی‌کننده MLP با یک لایه پنهان و ۶ نورون در این لایه بیشترین دقت و طبقه‌بندی‌کننده DTC بیشترین سرعت را در بین طبقه‌بندی‌کننده‌های بررسی شده خواهد داشت. بررسی رفتار مشتریان خوشه‌ها نشان داد که مشتریان را می‌توان در سه دسته وفادار، بالقوه و از دست رفته تقسیم‌بندی نمود.

کلیدواژه‌ها: رفتار خرید مشتری، LRFMSP، پرسپترون چندلایه، ماشین بردار پشتیبان، درخت تصمیم‌گیری

---

* نویسنده مسئول مقاله E-mail: a.khadivar@alzahra.ac.ir

۱- مقدمه و بیان مسئله

با توجه به رقابت فشرده در بازار و گزینه‌های متنوعی از محصول‌ها و خدمات که پیش روی مشتریان قرار دارد، تحلیل رفتار مشتریان و انتخاب شیوه مناسب بازاریابی براساس این تحلیل‌ها، عامل بسیار مهمی برای بقای بنگاه‌ها تلقی می‌شود [۱]. تحلیل مشتری فرآیندی است که از داده‌های رفتار مشتری برای کمک به تصمیم‌گیری‌های کلیدی کسب و کار به وسیله تقسیم‌بندی بازار و پیش‌بینی رفتار خرید مشتری استفاده می‌کند [۲]. از سوی دیگر، شناخت صحیح و درست رفتار مشتریان، مهمترین بعد مدیریت ارتباط با مشتری[۱] است. مدیریت ارتباط با مشتری، با شناخت مشتری آغاز می‌شود که عبارت است از هدف‌گیری گروهی از مردم که به احتمال زیاد، در آینده مشتری ما خواهند شد یا شناسایی مشتریانی که برای کسب و کار سودآور باشند [۳].

در کنار مباحث مدیریت ارتباط با مشتری گسترش فناوری‌های اطلاعاتی و ارتباطی باعث بوجود آمدن حجم زیادی از داده و اطلاعات شده است که نیاز به داده‌کاوی[۲] را به عنوان یکی از ابزارهای استخراج دانش و الگوهای موجود در این پایگاه داده‌ها الزامی ساخته است [٤]. تکنیک‌های داده‌کاوی معمولاً در CRM برای کمک به مشاهده رفتار مشتری از جمله ترجیحات مشتری و تقاضاهای مشتری استفاده می‌شود [٥]. سازمان‌ها با به کارگیری رویکردهای داده‌کاوی در تحلیل داده‌های مشتریان، الگوهای رفتاری مشتریان، نیازهای آن‌ها و ارتباط‌های پنهان میان داده‌ها را درک می‌کنند و براساس این الگوها بهتر می‌توانند در راستای برآورده ساختن نیاز مشتریان، منابع خود را به کار گیرند [۳].

در میان تکنیک‌های مختلف داده‌کاوی، روش‌های خوشه‌بندی[۳] به‌ویژه برای وظایف تحلیلی CRM، مانند تقسیم‌بندی مشتریان بر اساس ویژگی‌های مشابه و ارزش‌های مشترک بین آن‌ها، مفید هستند [٦]. خوشه‌بندی در واقع شکستن جمعیت زیاد مشتریان به بخش‌های مختلف است. شناسایی، تجزیه و تحلیل ویژگی‌ها و خوشه‌بندی مشتریان بر اساس ارزشی که برای شرکت دارند، زمینه را برای تخصیص بهینه منابع محدود، استفاده از استراتژی‌های بازاریابی مناسب و در نهایت مدیریت سودآوری در کنار مدیریت ارتباط با مشتری فراهم می‌کند [۲].

---

[۱] CRM
[۲] Data mining
[۳] Clustering

یکی از مدل‌های مطرح در خوشه‌بندی و تحلیل ارزش مشتری، مدل بخش‌بندی موسوم به مدل RFM[۱] است. مدل RFM یک ابزار مبتنی بر رفتار مناسب برای تجزیه و تحلیل رفتارهای مشتری، مطابق با رفتارهای گذشته وی در پایگاه داده است. RFM شامل متغیرهای تازگی، برای ارائه اطلاعات در مورد زمان سپری شده از آخرین خرید مشتری؛ فراوانی، از نظر کل خرید و وفاداری مشتری در یک دوره معین؛ و ارزش پولی، برای اشاره به میانگین مبلغ صرف شده در یک دوره زمانی معین است که به عنوان متغیرهای عملکردی مشتریان برای خوشه‌بندی آن‌ها استفاده می‌شود [۷؛ ۸].

به گفته رینارتز و کومار[۲]، چانگ و تسا[۳] و لی[۴] و همکاران، مدل RFM قادر به تشخیص اینکه آیا مشتریان روابط بلندمدت با یک سازمان دارند یا خیر نیست و بدین ترتیب آن‌ها ایده طول ارتباط با مشتری را پیشنهاد کرده و تأکید کردند که طول ارتباط "L" باید در کنار مدل اصلی RFM دیده شود. این متغیر نشان‌دهنده فاصله زمانی بین اولین و آخرین خرید مشتری در بازه مشاهده شده است. در واقع مدل پایه RFM، مشتریانی را که اخیراً ارزش مالی بالایی برای شرکت ایجاد کرده‌اند و الگوهای خرید کوتاه‌مدتی بیش از میانگین فراوانی خرید در بین مشتریان تکراری دارند، به‌عنوان مشتریان ارزشمند پیشنهاد می‌کند، در حالی که عامل طول ارتباط با شرکت نادیده گرفته می‌شود و باید به مدل RFM افزوده گردد [۲].

برخی مطالعات دیگر انحراف معیار بازه‌های زمانی بین خریدهای یک مشتری (S) را برای شناسایی بهتر ویژگی‌ها و تشخیص میزان منظم بودن مشتریانی که عادات خرید متفاوتی دارند به مدل LRFM افزودند [۲؛ ٦؛ ۹؛ ۱۰]. انحراف معیار بالای بین زمان‌های خرید مشتری نشان‌دهنده‌ی منظم نبودن فواصل بین خرید آن‌هاست و ترکیب آن با دیگر ویژگی‌های LRFM می‌تواند توصیف بهتری از رفتارهای خرید مشتریان و میزان وفادار بودن آن‌ها به دست دهد. برخی دیگر نیز میزان وابستگی مشتری به یک محصول خاص (P) را به مدل LRFM (بدون در نظر گرفتن S) افزوده‌اند [۱ ؛ ۱۱]. این معیار میزان وابستگی مشتری به یک کالای خاص را نشان می‌دهد. در نتیجه در این پژوهش به منظور بررسی جامع تر، مجموعه این معیارها

---

[۱] Recency, Frequency, Monetary
[۲] Reinartz & Kumar
[۳] Chang & Tsay
[۴] Li

(LRFMSP) را برای خوشه‌بندی مشتریان فروشگاه خرده‌فروشی مورد استفاده قرار می‌دهیم.

صنعت خرده‌فروشی، روزانه با حجم انبوهی از اطلاعات خرید مشتریان مواجه است و انجام پژوهشی که بتواند با استفاده از داده‌کاوی و مجموعه کاملی از معیارها به ارائه تقسیم‌بندی بهتری از مشتریان پرداخته و باعث بهبود عملکرد فروش و سودآوری فروشگاه در محیط رقابتی شود، برای مدیران و کارشناسان ضرورت پیدا می‌کند. بنابر مطالب عنوان شده و اهمیت داده‌کاوی پایگاه داده‌های مشتریان به منظور مدیریت بهتر ارتباط با مشتری، این پژوهش به دنبال آن است که بررسی نماید کدام روش داده‌کاوی برای بخش‌بندی مشتریان با هدف بهبود مدیریت ارتباط با مشتری می‌تواند عملکرد بالاتری داشته باشد. در راستای اهداف این پژوهش می‌توان سوال‌های این پژوهش را در قالب ذیل بیان کرد:

۱. بهترین تعداد خوشه برای بخش‌بندی مشتریان با استفاده از شاخص‌های بدست آمده چیست؟

۲. از نقطه نظر شاخص‌های مدل پژوهش مشتریانی که در هر بخش قرار می‌گیرند چه ویژگی‌هایی داشته و رفتار خرید مجدد آن‌ها چگونه بوده و مناسب است که چه سیاستی در قبال آن‌ها اتخاذ شود؟

۳. کدام طبقه‌بندی‌کننده برای تشخیص دسته مشتریان، از دقت و سرعت بالاتری برخوردار است؟

۲- ادبیات و پیشینه پژوهش

۲-۱- مدیریت ارتباط با مشتری

سابقه مدیریت ارتباط با مشتری و پیاده‌سازی آن به صورت امروزی به دهه ۱۹۹۰ برمی‌گردد [٤]. سویفت[۱] CRM را یک رویکرد سازمانی برای درک و تأثیرگذاری بر رفتار مشتری از طریق ارتباطات معنی دار به منظور بهبود جذب، حفظ، وفاداری و سودآوری مشتری تعریف کرده است [۱۲]. CRM به عنوان یک ابزار حیاتی برای بهبود روابط با مشتریان با استفاده از فناوری‌های اطلاعاتی در نظر گرفته می‌شود که اطلاعات مشتری را تجزیه و تحلیل، جمع‌آوری و به اقدامات مشخص برای مدیران تبدیل می‌کند تا بتوانند به خوبی پاسخگوی تقاضای مشتری باشند. از این رو CRM به طور گسترده به تکنیک‌های داده‌کاوی برای دستیابی به روابط خوب

[۱] Swift

با مشتری متکی است [۵]. مدیریت ارتباط با مشتری در واقع یاری‌دهنده‌ای در جهت شناسایی بهتر مشتریانی است که سودآوری بیشتری داشته و ارزش رابطه بلندمدت را دارند [۱۳].

۲-۲- خوشه‌بندی

خوشه‌بندی مشتریان یکی از شاخه‌های داده‌کاوی و از مباحث مطرح در حوزه مدیریت ارتباط با مشتری است [٦]. فرآیند قرار دادن داده‌ها در گروه‌هایی که بیشترین شباهت را در برخی ویژگی‌ها دارند خوشه‌بندی نامیده می‌شود [۱]. هر خوشه شامل مجموعه‌ای از داده‌های مشابه درون خوشه‌ای بوده و متفاوت از داده‌های سایر خوشه‌هاست [۳؛ ٦].

الگوریتم K-means به عنوان یکی از محبوب‌ترین الگوریتم‌های خوشه‌بندی، اولین بار توسط مک کوئین[۱] در سال ۱۹٦۷ مورد استفاده قرار گرفت [۱] و به طور گسترده در زمینه‌های مختلف از جمله داده‌کاوی، تجزیه و تحلیل داده‌های آماری و سایر کاربردهای تجاری استفاده شده است [٦].

K-Means یک روش سریع در تجزیه و تحلیل خوشه‌بندی است، اما دقت و زمان اجرای نتایج خوشه‌بندی تا حد زیادی به محل مرکز خوشه‌بندی اولیه بستگی دارد. برای حل این مشکل که K-means به نقاط اولیه حساس است، آرتور[۲] و همکاران الگوریتم K-means++ را برای بهبود روش انتخاب تصادفی مراکز خوشه‌بندی اولیه K-means، یعنی ایجاد حداکثر فاصله ممکن بین مراکز خوشه‌بندی در هنگام انتخاب مراکز خوشه‌بندی اولیه پیشنهاد کردند [۱]. قبل از اعمال الگوریتم K-means، مقدار k باید مشخص شود [٦]. از آنجایی که الگوریتم K-means++ دقیق و کارآمد است، این مقاله از الگوریتم K-means++ برای طبقه‌بندی مشتریان استفاده می‌کند.

۲-۳- پیشینه پژوهش

در جدول ۱ برخی از پژوهش‌های صورت گرفته در زمینه خوشه‌بندی مشتریان با استفاده از RFM و سایر مدل‌های توسعه یافته آن ارائه گردیده است.

---

[۱] MacQueen

[۲] Arthur

جدول ۱. مروری بر پیشینه پژوهش

نویسندگان	عنوان	معیار	روش	نتیجه
بریانت[۱] (۲۰۱۱) [۱٤]	داده‌کاوی با استفاده از تجزیه و تحلیل RFM	RFM	DTC	خوشه‌بندی و سپس طبقه‌بندی مشتریان فروشگاه ورزشی با دقت ۸۱ درصد
یوریسیچ[۲] و همکاران (۲۰۲۰) [۱۵]	بهینه‌سازی CRM بانک با استفاده از طبقه‌بندی پیشگویانه براساس روش SVM	RFM	K-means، SVM و DTC	خوشه‌بندی و سپس طبقه‌بندی مشتریان بانک با دقت ۹۳,۳۳ درصد
رحیم[۳] و همکاران (۲۰۲۱) [۱٦]	رفتار خرید مجدد مبتنی بر RFM برای طبقه‌بندی و تقسیم‌بندی مشتری	RFM	MLP-DTC-SVM	بررسی قصد خرید مشتریان با محاسبه RFM– خوشه‌بندی و سپس طبقه‌بندی مشتریان
هوآنگ[٤] و همکاران (۲۰۲۱) [۱۷]	استفاده از مدل RFM برای ایجاد ارزش مشتری با ایجاد بخش در صنایع خدماتی مختلف	RFM	DTC	تقسیم‌بندی مشتریان صنایع خدماتی مختلف، بر اساس رفتار خریدشان
هلت و همکاران (۲۰۲۱) [۱۱]	پیش‌بینی ارزش مشتری به ازای هر محصول: از RFM تا RFM/P	RFM/P	BG/BB model	ترکیب دیدگاه‌های مشتری و محصول با استفاده از مدل RFM/P
چائو[۵] و همکاران (۲۰۲۱) [۸]	یک مدل LRFM برای تجزیه و تحلیل وفاداری بیماران سرپایی یک مرکز پزشکی در تایوان	LRFM	تجزیه و تحلیل خوشه‌ای	تقسیم بیماران سرپایی به پنج گروه مختلف (مشتریان اصلی، بالقوه، جدید، از دست رفته و مصرف‌کننده منابع)

[۱] Birant
[۲] Djurisic
[۳] Rahim
[٤] Huang
[۵] Chao

نویسندگان	عنوان	معیار	روش	نتیجه
گوان[1] و همکاران (۲۰۲۰) [۱۸]	تصویر پرتره اعضای بازار براساس داده‌کاوی.	LRFMC	K-means و KNN	تشکیل پرتره اعضا در مراکز خرید با استفاده از محاسبه ارزش عضویت
گانی[2] و همکاران (۲۰۲۰) [٦]	یک رویکرد ترکیبی برای پروفایل‌سازی مشتری در ویدئو در مورد خدمات تقاضا با استفاده از خوشه‌بندی و استخراج قانون انجمن	LRFMP	K-means و Apriori	تقسیم‌بندی مشتریان خدمات ویدیویی و شناسایی چهار گروه مشتری اصلی شامل پرمصرف، کم مصرف، وفادار کم مصرف و بی‌وفا
وو[3] و همکاران (۲۰۲۱) [۱]	شناسایی ارزش کاربر بر اساس مدل RFM بهبود یافته و الگوریتم K-Means++ برای تجزیه و تحلیل داده‌های پیچیده	LRFMP	خوشه‌بندی K-means++	تعیین بهتر ارزش مشتری با استفاده از مدل RFM بهبودیافته و طبقه‌بندی بهتر آن
چشمی[4] و همکاران (۲۰۲۱) [۲]	پیش‌بینی گردش مالی مشتریان با استفاده از شبکه‌های عصبی بازگشتی ارتباطات بی‌سیم و محاسبات سیار	LRFMP	شبکه عصبی بازگشتی	انجام تحلیل جامع‌تر از گردش مالی مشتریان با استفاده از مدل جدید پیشنهادی و شبکه عصبی
جهانیان وهمکاران (۱۳۹۷) [۱۲]	به‌کارگیری روش هزینه‌یابی مبتنی بر فعالیت زمان‌مبنا	RFM	K-means	خوشه‌بندی مشتریان تحت عناوین وفادار امروز، مشتریان رقبا، وفادار فردا و آماده پذیرش پیشنهاد رقبا

[1] Guan
[2] Guney
[3] Wu
[4] Chashmi

نویسندگان	عنوان	معیار	روش	نتیجه
	به‌منظور رتبه‌بندی مشتریان سودآور			
بشردوست و همکاران (۱٤۰۱) [٤]	تحلیل خوشه‌بندی مشتریان براساس مدل WRFM با استفاده از رویکرد داده‌کاوی غیر نظارتی	WRFM	K-means	تقسیم‌بندی مشتریان محصولات بهداشتی و شناسایی ٤ دسته مشتری خاص و کلیدی، طلایی بالقوه، نامطمئن ازدست رفته، نامطمئن جدید

با بررسی پژوهش‌های صورت گرفته، مشخص گردید که اکثر پژوهشگران برای خوشه‌بندی مشتریان از شاخص‌های محدودی استفاده کرده و تعداد کمتری نیز به پیش‌بینی رفتار مشتریان با استفاده از طبقه‌بندی‌کننده‌ها پرداخته‌اند. بنابراین این پژوهش قصد دارد با در نظر گرفتن معیار جامعی چون LRFMSP به خوشه‌بندی مشتریان پرداخته و در نهایت با استفاده از سه طبقه‌بندی‌کننده مختلف به طبقه‌بندی و پیش‌بینی رفتار خرید آنان و مقایسه عملکرد این طبقه‌بندی‌کننده‌ها بپردازد.

۳- روش‌شناسی پژوهش

این پژوهش کاربردی توسعه‌ای و طرح آن از نوع پس رویدادی است. بنابراین با استفاده از اطلاعات موجود در زمان گذشته انجام گرفته است. این روش زمانی مورد استفاده قرار می‌گیرد که داده‌ها از محیطی که به گونه طبیعی وجود داشته یا از واقعه‌ای که بدون دخالت مستقیم پژوهشگر رخ داده فراهم شده باشد.

پژوهش حاضر از یک مجموعه داده[۱] در دسترس عموم، شامل تراکنش‌های یک خرده‌فروشی آنلاین در غرب اروپا طی یک بازه زمانی ۱۳ ماهه، از فوریه ۲۰۱۸ تا فوریه ۲۰۱۹ استفاده شده است. هر تراکنش شامل ۸ عنصر بوده که توضیحات آن در جدول ۲ ارائه شده است. پایگاه داده مورد استفاده شامل داده‌های استانداری بوده که در سطح بین الملل مورد استفاده قرار

[۱] https://www.kaggle.com/datasets/vipin20/transaction-data

گرفته و علاوه بر اطلاعات تراکنش‌ها دربردارنده اطلاعات کالاها نیز است، که کار را برای تحلیل‌های مشابه هموار می‌سازد. جهت تجزیه تحلیل و پردازش داده‌ها از نرم افزار متلب استفاده است.

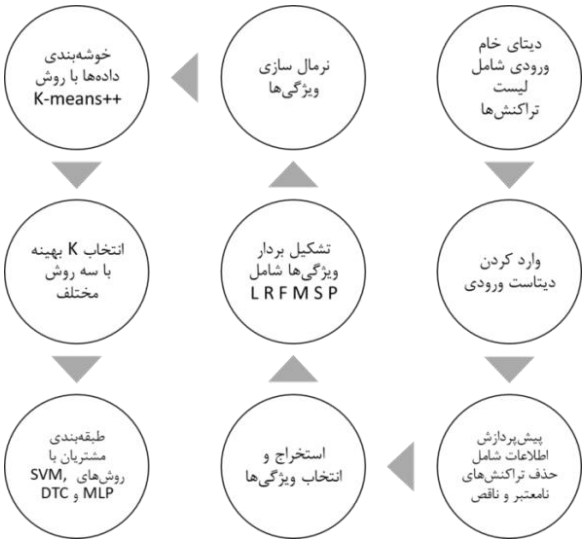
پس از پاکسازی و نرمال‌سازی داده‌های ورودی، ویژگی‌های LRFMSP استخراج شده و سپس به خوشه‌بندی داده‌ها پرداخته می‌شود. در نهایت برای تشخیص اینکه مشتریان جدید در کدام دسته قرار می‌گیرند، از سه طبقه‌بندی‌کننده پرسپترون چندلایه[۱]، ماشین بردار پشتیبان[۲] و درخت تصمیم[۳] استفاده می‌شود. طراحی طبقه‌بندی‌کننده به این نحو انجام می‌پذیرد که ابتدا با بخشی از داده‌های خوشه‌بندی‌شده آموزش داده می‌شود و با استفاده از داده‌های تست، دقت طبقه‌بندی محاسبه می‌گردد. سپس طبقه‌بندی‌کننده می‌تواند تشخیص دهد که هر مشتری در کدام دسته قرار می‌گیرد تا در نهایت تعیین شود که کدام سیاست تشویقی در قبال وی باید صورت پذیرد.

جدول ۲. عناصر یک تراکنش و استفاده از آن‌ها

نام عنصر	شرح
UserId	شماره منحصر به فرد مشتری
TransactionId	شماره منحصر به فرد فاکتور
TransactionTime	تاریخ و ساعت تراکنش
ItemCode	شماره منحصر به فرد کالا
ItemDescription	نام و شرح کالا
NumberOfItemsPurchased	تعداد این کالا در تراکنش
CostPerItem	قیمت واحد کالا
Country	کشور خریدار

شکل ۱ 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.

[۱] Multilayer perceptron (MLP)
[۲] Support vector machines (SVM)
[۳] Decision tree (DT)



شکل۱. الگوی مفهومی پژوهش

۳-۱- پردازش اولیه روی داده‌ها

شکل خام مجموعه داده‌های موجود، با انواع مختلفی از داده‌ها که ممکن است به مشتری نامرتبط باشد و مانع استخراج الگوهای رفتاری شود، آلوده است. اطلاعات ۱۰۸۳۸۱۸ تراکنش در این دیتاست وجود دارد که پس از انجام پیش‌پردازش‌های لازم، تعداد ۳۸۷٤۹٦ تراکنش برای انجام تحلیل‌های بعدی مورد استفاده قرار گرفت. برخی تراکنش‌ها که حذف شده‌اند در جدول ۳ ارائه شده است.

داده‌های مورد استفاده در آزمایش‌ها پس از پیش‌پردازش انجام شده شامل ۳۸۷٤۹٦ تراکنش با ۳۱۸٤ کد محصول منحصر به فرد و ٤۳۱٦ شماره مشتری مختلف است. تعداد کل فاکتورها در داده‌های پیش‌پردازش شده ۱۸۱۹٤ مورد است. بیشتر تراکنش‌ها مربوط به کشور انگلستان بوده و پس از آن آلمان و فرانسه تراکنش‌های بیشتری دارند.

جدول ۳. انواع تراکنش‌های حذف شده

نوع تراکنش	دلیل حذف
تراکنش تکراری	تراکنش‌هایی که در داده‌ها تکرار شده اند
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

۳-۲- خوشه‌بندی مشتریان بر اساس LRFMSP

شش عنصر از هشت عنصر داده‌های تراکنش‌ها (شماره منحصربه‌فرد مشتری، شماره منحصربه‌فرد فاکتور، تاریخ و ساعت تراکنش، شماره منحصربه‌فرد کالا، تعداد این کالا در تراکنش و قیمت واحد کالا) در بررسی صورت گرفته مورد استفاده قرار گرفته است. 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 ٤ ارائه شده است:

جدول ۴. تعریف شاخص‌های LRFMSP

متغیر	نحوه محاسبه	شرح
تازگی خرید (R)	فاصله زمانی بین زمان آخرین خرید مشتری با تاریخ ۱ مارچ ۲۰۱۹ 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.
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.

متغیر	نحوه محاسبه	شرح
انحراف معیار خریدها (S)	انحراف معیار فواصل زمانی بین خریدهای مشتری	بیانگر منظم بودن رفتار خرید مشتری بوده و هر چه مقدار پایین تری داشته باشد، نشان دهنده رفتار خرید دوره‌ای یک مشتری می‌باشد.
تکرار خرید یک کالا (P)	تعداد دفعات خرید یک کالای خاص توسط مشتری (فارغ از تعداد خریداری شده)	بیانگر میزان وابستگی مشتری به یک محصول خاص است.

برای از بین بردن تأثیر تفاوت رنج ویژگی‌های محاسبه شده، داده‌های رفتاری استخراج شده از مشتریان (ماتریس ویژگی‌ها) با اعمال نرمال‌سازی برای هر ستون داده در ماتریس ویژگی‌ها پردازش می‌گردد. متداول‌ترین روش نرمال‌سازی در کارهای مشابه انجام شده، روش نرمال‌سازی Z-Score می‌باشد، که داده‌های مربوط به هر ویژگی را به میانگین صفر و انحراف معیار ۱ نگاشت می‌کند. اما از آن‌جایی که انجام این تبدیل باعث تغییر نسبت بین مقادیر به دست آمده برای ویژگی‌های مشتریان مختلف شده و از طرفی باعث جمع شدن داده‌های یکنواخت (مانند R و L) با پراکندگی بالا حول صفر می‌شود، از دیگر روش معمول نرمال‌سازی متداول یعنی Min-Max یا Range استفاده می‌کنیم. با انجام این نرمال‌سازی داده‌های مربوط به هر ویژگی در بازه بین صفر و یک به صورت یکنواخت نگاشته خواهد شد.

پس از انجام نرمال‌سازی یکنواخت، مقدار K بهینه برای خوشه‌بندی با استفاده از سه روش Silhouette، DaviesBouldin و CalinskiHarabasz محاسبه و در جدول ۵ ارائه شده است.

جدول۵. تعیین مقدار K بهینه با استفاده از سه روش Silhouette، DaviesBouldin و CalinskiHarabasz

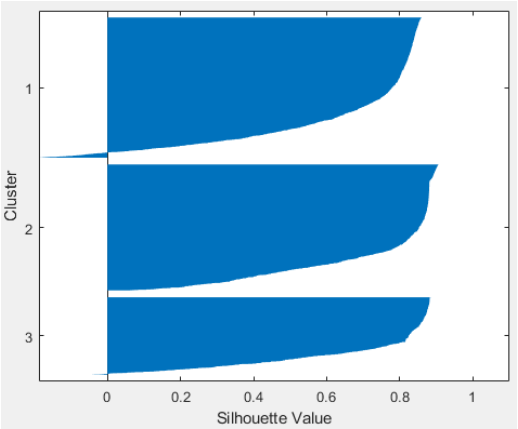
K	Silhouette	DaviesBouldin	CalinskiHarabasz
۲	۰,٦٩٢٦	۰,٧٥٧١	٥٨٩٥,٢
۳	۰,٦٩٩٥	۰,٦٣٤٥	٦٨٩٧,٧
٤	۰,٦٢٨٢	۰,٨٣٤٣	٦٦٣٦,٣
٥	۰,٥٩٠٧	۰,٨٧٨٧	٦٠٤٨,٠

K	Silhouette	DaviesBouldin	CalinskiHarabasz
٦	٠,٦٠٢٨	٠,٨٦٦٧	٥٩٥٣,٣
٧	٠,٦٠٥١	٠,٨٥٧٥	٥٦٩٤,٠
٨	٠,٥٧٥٢	٠,٨٥٨٤	٥٥٩٨٧
٩	٠,٥٢٨٢	٠,٩٦٦٥	٥٣٢٥,٥
١٠	٠,٥٥٤٥	٠,٩٥٢٥	٥١٢٨,٥

در روش‌های Silhouette و CalinskiHarabasz هر چه مقدار ارزیابی شده بالاتر باشد، خوشه‌یابی بهتر است و در روش DaviesBouldin هر چه مقدار کمتر باشد، بهتر است. بنابراین مقدار K بهینه برای خوشه‌بندی یا استفاده از هر سه روش ۳ است. شکل ۲ نمودار Silhouette را برای K = ۳ نمایش می‌دهد.

خوشه‌بندی را با استفاده از روش Kmeans++ و Replicate برابر با ۱۰ انجام می‌دهیم تا با تقریب خوبی از افتادن در نقاط زیر بهینه جلوگیری شود.

نتایج خوشه‌بندی نشان می‌دهد که ۱۵۸۳ مشتری در خوشه اول، ۱۷۶۱ نفر در خوشه دوم و ۹۷۲ نفر هم در خوشه سوم قرار گرفته‌اند که پراکندگی نسبتاً مناسبی را بین مشتریان نشان می‌دهد. مقایسه ویژگی‌های استخراج شده بین خوشه‌ها در جدول ٦ ارائه شده است.



شکل ۲. نمودار Silhoutte به ازای K = ۳

جدول ۶. میانگین مقادیر ویژگی‌ها برای خوشه‌های ایجاد شده

	P	S	L	M	F	R
خوشه اول	۰/۰۱۳٤	۰/۰۲۰٦	۰/۱۰۰۷	۰/۰۰۲۷	۰/۰۰٤۸	۰/۱٤٤۱
خوشه دوم	۰/۰۵۲۹	۰/۱٦۵۹	۰/۷۳۳٤	۰/۰۱۳۳	۰/۰۳۵۵	۰/۰۹۵۸
خوشه سوم	۰/۰۰۸۰	۰/۰۰۹۵	۰/۰۵۱۹	۰/۰۰۱۷	۰/۰۰۲٤	۰/٦۸۵٦
کل داده‌ها	۰/۰۲۸۳	۰/۰۳۷۹	۰/۱۷۰۵	۰/۰۰۳۳	۰/۰۰۷۸	۰/۱۲۰۷

۳-۳- طبقه‌بندی مشتریان

هر کدام از روش‌های طبقه‌بندی نسبت به سایر روش‌ها مزایا و معایبی داشته و برای طبقه‌بندی نوع خاصی از داده‌ها بهتر عمل می‌کند. قدرت ماشین بردار پشتیبان در تشخیص الگوهای پیچیده با استفاده روش مبتنی بر کرنل است و با نگاشت داده‌ها به فضایی با ابعاد بالاتر به راحتی می تواند الگوها را طبقه‌بندی نماید [۱۹]. در ضمن، نسبت به دیگر روش‌ها کمتر دچار مشکل بیش‌برازش داده‌ها می‌شوند [۲۰]. مزیت اصلی رویکرد درخت تصمیم نشان دادن راه‌حل‌ها است. این روش در مقایسه با سایر متدها، به ویژه در شرایطی که فضای نمونه بزرگ است، سریعتر است. علاوه بر این درک آن نیز برای افراد غیرفنی آسان‌تر است. مزیت دیگر آن این است که هر دو نوع داده عددی و رتبه‌ای را می‌تواند طبقه‌بندی کند [۲۱]. از مزایای شبکه‌های عصبی قابلیت یادگیری مدل‌های غیرخطی داده‌هاست و به دلیل پیچیدگی بالای ساختاری، توانایی بالایی در یادگیری مدل‌های پیچیده دارند. همچنین پس از پایان آموزش، با رسیدن داده‌های جدیدتر می‌توان شبکه را با داده‌های جدید مجدداً آموزش داد تا نسبت به این داده‌ها جوابگو باشد [۲۲]. بنابراین با استفاده از هر روش به آموزش طبقه‌بندی‌کننده می‌پردازیم تا با مقایسه سرعت و دقت طبقه‌بندی‌کننده‌ها، بهترین طبقه‌بندی‌کننده را برای این داده‌ها انتخاب کنیم.

۳-۳-۱- پرسپترون چندلایه

در این پژوهش MLP پیشنهادی شامل یک لایه پنهان با ٦ نورون روی لایه پنهان، با تابع فعال‌سازی SCG[۱] برای هر عنصر است. برای محاسبه تعداد نورون‌های لایه میانی برای هر

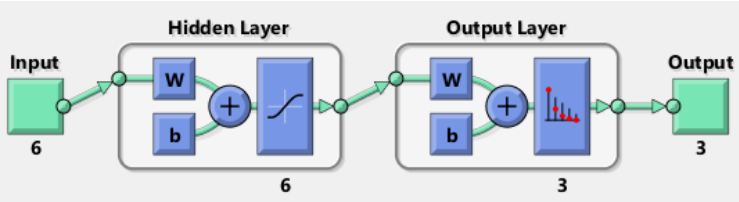
[۱] Scaled conjugate gradient backpropagation

تعداد نورون صد بار محاسبه انجام شده و طبق جدول ۷ میانگین دقت‌ها برابر با ٦ شده است که با توجه به محاسبات انجام شده تعداد ٦ نورون در لایه میانی مناسب‌تر به نظر می‌رسد. افزودن تعداد نورون‌های لایه میانی علاوه بر اینکه تأثیر چندانی روی افزایش دقت ندارد، باعث طولانی‌تر شدن زمان آموزش و محاسبات شبکه می‌گردد.

جدول۷. میانگین دقت پیش‌بینی با MLP به ازای تعداد نورون‌های مختلف

تعداد نورون لایه میانی	دقت
۱	۸۲,۹٪
۲	۹۰,۳٤٪
۳	۹۸,۰٦٪
٤	۹۹,٦۹٪
٥	۹۹,۷۲٪
٦	۹۹,۸۱٪
۷	۹۹,۸۱٪

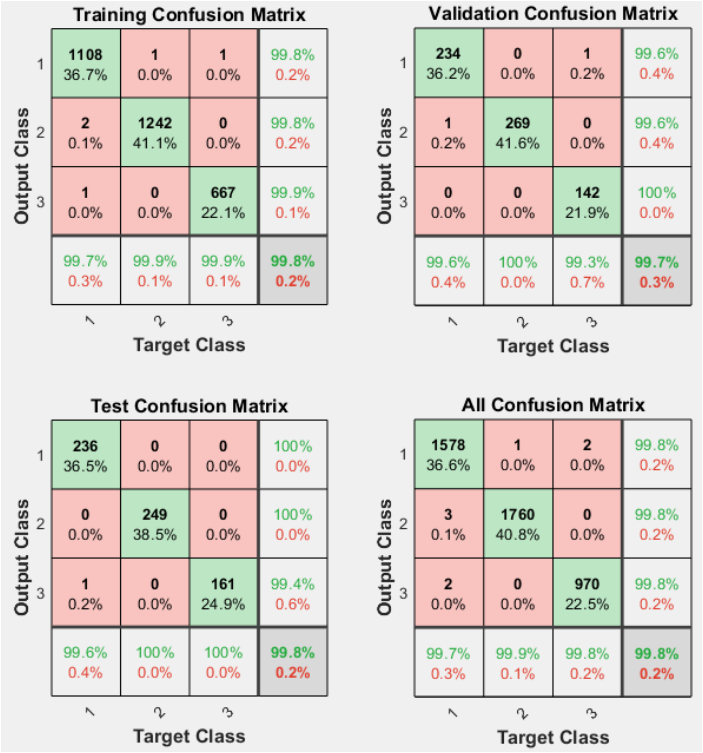
شکل ۳ ساختار MLP مورد استفاده برای پیش‌بینی مشتریان بر اساس ویژگی‌های رفتاری را نشان می‌دهد که لایه ورودی شامل ٦ نورون است که تعداد آن براساس تعداد ویژگی‌ها تعیین می‌گردد. تعداد نورون‌های لایه میانی که با توجه به دقت نتایج ٦ محاسبه گردید و تعداد نورون‌های لایه خروجی با توجه به سه کلاسه بودن نتایج ۳ در نظر گرفته شد.



شکل ۳. ساختار MLP مورد استفاده برای پیش‌بینی رفتار مشتریان

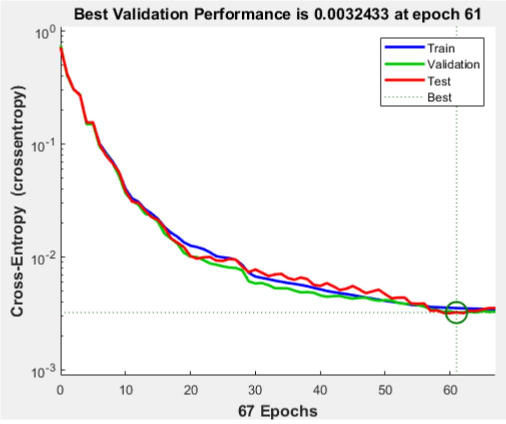
شکل ٤ نشان‌دهنده‌ی ماتریس Confusion برای یک بار اجرای طراحی انتخابی برای شبکه‌ی عصبی می‌باشد. ۷۰ درصد از داده‌های ورودی برای آموزش، ۱۵ درصد برای تست و ۱۵ درصد برای Validation مورد استفاده قرار گرفته است. تعداد خطای پیش‌بینی برای هر کدام

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 درشکل ٦ 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 ٥ پس از ٦١ دور، میزان کارایی شبکه ثابت شده است.



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

۳-۳-۲- ماشین بردار پشتیبان

از SVM برای داده‌های قابل تفکیک خطی می‌توان بهره برد، و برای داده‌هایی که به طور خطی قابل تفکیک نیستند، با استفاده از کرنل [۱] می‌توان این جداسازی را انجام داد [۱۶]. برای تعیین تابع کرنل مورد استفاده در این تحقیق سه کرنل متداول یعنی کرنل خطی، کرنل گوسی [۲] و کرنل چندجمله‌ای مورد بررسی قرار گرفته‌اند و میانگین دقت برای هر کرنل در ۱۰۰ بار اجرا به صورت جدول ۸ به دست آمد:

جدول۸. میانگین دقت برای هر کرنل در ۱۰۰ بار اجرا

Kernel Function Name	Formula	Accuracy
'Gaussian' or 'RBF'	$G(x_j,x_k)=\exp(-\|x_j-x_k\|^2)$	۹۹/۵۶%
'Linear'	$G(x_j,x_k)=x_j'x_k$	۹۹/۳۱%
'Polynomial'	$G(x_j,x_k)=(1+x_j'x_k)^q$	۹۹/۶۴%

[۱] Kernel
[۲] Gaussian kernel

طبق نتایج به دست آمده بیشترین بازدهی با کرنل چندجمله‌ای به دست آمده که برابر با ۹۹,٦٤ درصد می‌باشد.

۳-۳-۳- درخت تصمیم

برای تعیین بهترین روش برای ساخت درخت تصمیم، سه روش مختلف ساخت درخت مقایسه شده که درخت CART بالاترین دقت (۹۹,۱٦ درصد) را نشان می‌دهد. میانگین دقت درخت تصمیم هم پس از ۱۰۰ بار اجرای مختلف (هر بار با انتخاب تصادفی داده آموزش و تست) در جدول ۹ ارائه شده است.

جدول۹. مقایسه میانگین دقت درخت تصمیم با سه روش Standard CART، Curvature و Interaction-Curvature

Split Predictor Algorithm	Description	Accuracy
Standard CART	پیش‌بینی کننده تقسیمی را انتخاب می‌کند که بهره معیار تقسیم را در بین همه تقسیم‌های ممکن همه پیش‌بینی‌ها به حداکثر می‌رساند	۹۹/۱٦%
Curvature test	پیش‌بینی کننده تقسیمی را انتخاب می‌کند که مقدار p آزمون کای اسکوئر استقلال بین هر پیش‌بینی و پاسخ را به حداقل می‌رساند. سرعت تمرین مشابه CART استاندارد است.	۹۸/۰۳%
Interaction-test	پیش‌بینی کننده تقسیمی را انتخاب می‌کند که مقدار p آزمون کای‌دو استقلال بین هر پیش‌بینی‌کننده و پاسخ را به حداقل می‌رساند و مقدار p یک آزمون کای‌دو استقلال بین هر جفت پیش‌بینی و پاسخ را به حداقل می‌رساند.. سرعت تمرین می‌تواند کمتر از CART استاندارد باشد.	۹۸/۰۰%

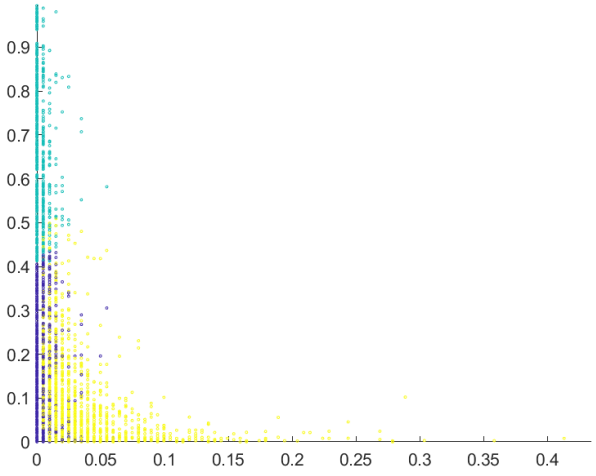
۴- تجزیه و تحلیل نتایج

نتایج خوشه‌بندی نشان می‌دهد که برای خوشه اول مقادیر تمامی ویژگی‌ها متعادل بوده و نزدیک میانگین کل داده‌ها است. برای خوشه دوم مقادیر F، M و L که همگی شاخص‌های مثبتی هستند بالا بوده و شاخص R که شاخصی معکوس است پایین‌تر از سایر خوشه‌هاست. شاخص S که بیانگر میزان انحراف معیار بین خریدهاست، برای خوشه دوم بالاست که نشان

می‌دهد سرمایه‌گذاری‌های تبلیغاتی روی این دسته از کاربران باید برای منظم‌تر کردن بازدیدهای آنان باشد.

شاخص P هم که میزان وابستگی به یک محصول خاص را نشان می‌دهد، برای خوشه دوم بالاتر است و بیانگر این است که این دسته بیشتر از سایرین، محصولات یکسانی را خرید مجدد کرده‌اند. مشتریانی که در خوشه دوم قرار گرفته‌اند، وفادارترین و مفیدترین مشتریان خرده‌فروشی هستند و از این رو می‌توان آن‌ها را **مشتریان وفادار** نامید.

برای خوشه‌ی سوم به جز R که بالاست، بقیه مقادیر پایین‌تر از سایر خوشه‌هاست. این‌ها مشتریانی هستند که در شاخص‌های مثبت، مقادیر پایین و در شاخص‌های معکوس مقدار بالایی دارند و در حقیقت کمترین میزان وابستگی و وفاداری را به این مجموعه دارند و مدت زمان زیادی هم از آخرین خرید آن‌ها می‌گذرد. از این رو این دسته را هم می‌توان **مشتریان از دست رفته** نام نهاد. مشتریان خوشه‌ی اول مشتریانی هستند که می‌توان در تبلیغات و تخفیفات برای وفادارتر کردن و علاقه‌مندتر کردن آن‌ها به مجموعه روی آن‌ها سرمایه‌گذاری نمود. از این رو این مشتریان را می‌توان **مشتریان بالقوه** نامید.

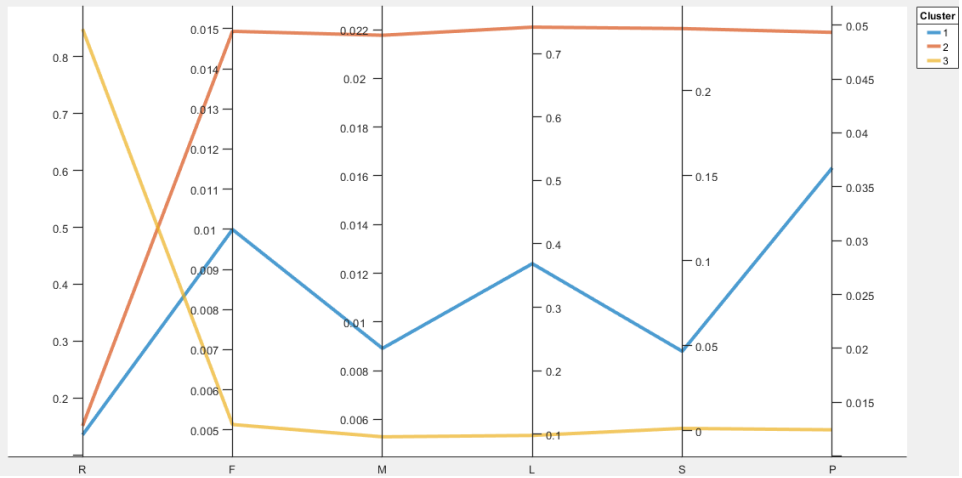


نمودار ۱. پراکندگی خوشه‌ها در مختصات دفعات خرید (F) نسبت به تازگی خرید (R)

آمنه خدیور و همکار ـــ

نمودار ۲. پراکندگی خوشه‌ها در مختصات تازگی خرید (R) به مجموع هزینه انجام شده (M)

برای بررسی بیشتر نتایج خوشه‌بندی، مقادیر نرمال شده‌ی ٦ ویژگی را برای سه مشتری مختلف از خوشه‌های متفاوت در نمودار خطی ۳ ترسیم نموده‌ایم.



نمودار۳. مقادیر نرمال شده‌ی ٦ ویژگی برای سه مشتری مختلف از خوشه‌های متفاوت

مشتری آبی با شناسه ۲۶۱۹۳۳ جزو خوشه اول است و همانطور که در شکل هم پیداست به جز ویژگی R برای بقیه ویژگی‌ها مقدار متعادلی دارد. این مشتری زمان آخرین خرید نزدیکی داشته، ولی نسبت به بقیه ویژگی‌ها رفتارش متعادل بوده است و برای تبدیل شدن به مشتریان وفادار (خوشه دوم) نیاز است که با انجام تبلیغات و برنامه‌های ویژه، وی را به خریدهای بیشتر از فروشگاه ترغیب کرد.

مشتری قرمز با شناسه ۲۵۹۵۳۹ جزو خوشه دوم یا مشتریان وفادار فروشگاه است. این مشتری ٤ مراجعه به فروشگاه داشته و ۲۷٤ روز بین اولین خرید و آخرین خریدش فاصله بوده است. البته انحراف معیار بین زمان‌های خرید وی ۵۷ است که نشان می‌دهد رفتار خرید منظمی نداشته و پس از دو خرید نزدیک به هم، خریدهای بعدی را با فاصله زمانی بیشتری انجام داده است. بنابراین علاوه بر مثبت بودن شاخص‌های این مشتری، باید راهکارهایی هم برای کاسته نشدن علاقه این دسته از مشتریان و حفظ آن‌ها اندیشیده شود.

مشتری نارنجی با شناسه ۲۵۹۹۱۷ جزو خوشه سوم است که ضعیف‌ترین ویژگی‌ها را در بین مشتریان دارند. مشتریان این دسته معمولا یک یا خرید کوچک داشته و رابطه خود را با فروشگاه ادامه نداده‌اند. برخی از این مشتریان ممکن است در جشنواره‌های فروش به فروشگاه آمده باشند و برای وفادار کردن آن‌ها تخفیفات یا تبلیغات بیشتری نیاز خواهد بود.

پس از خوشه‌بندی و برچسب‌گذاری مشتریان، برای تشخیص دسته‌بندی مشتریان جدید از طبقه‌کننده‌های MLP، SVM و DTC استفاده شده است. نتایج حاصل از این طبقه‌بندی‌کننده‌ها نشان می‌دهد که با استخراج این شش ویژگی با دقت بالایی می‌توان دسته‌بندی مشتریان جدید را مشخص کرد. جدول ۱۰ به طور خلاصه دقت بهترین معماری تعیین شده برای هر طبقه‌بندی‌کننده را مقایسه می‌کند.

جدول ۱۰. میزان دقت بهترین معماری تعیین شده برای هر طبقه‌بندی‌کننده

طبقه‌بندی‌کننده	معماری مورد انتخاب	دقت	زمان انجام تست
MLP	یک لایه پنهان با ٦ نورون	۹۹,۸۱٪	۳۵/۵ms
SVM	کرنل چندجمله‌ای	۹۹,٦٤٪	۱۸/۱ms
DTC	Standard CART	۹۹,۱٦٪	۲/۳ms

همانطور که مشاهده می‌شود گرچه دقت درخت تصمیم کمی پایین‌تر از روش‌های دیگر است، اما به دلیل ساختار ساده‌ای که دارد با سرعت خیلی بالاتری نسبت به دو روش دیگر به پیش‌بینی خوشه مشتریان می‌پردازد. هرچند سرعت طبقه‌بندی داده‌های جدید به طور معمول بسیار ناچیز است و در تعیین رفتار خرید مشتریان تأثیرگذار نیست.

۵- بحث و نتیجه‌گیری

هدف اصلی این تحقیق خوشه‌بندی مشتریان بر اساس شاخص‌های LRFMSP و در نهایت طبقه‌بندی و پیش‌بینی رفتار خرید آن‌ها با استفاده از تکنیک‌های طبقه‌بندی پرسپترون چندلایه، ماشین بردار پشتیبان و درخت تصمیم است. نتایج پژوهش نشان داد که استفاده از این شاخص‌ها برای خوشه‌بندی مشتریان منجر به تفکیک‌پذیری بالایی در طبقه‌بندی مشتریان خواهد شد. نتایج خوشه‌بندی نشان داد که تعداد خوشه بهینه برای مشتریان این خرده‌فروشی برابر با سه می‌باشد. پس از خوشه بندی با توجه به شاخص های هر دسته، خوشه ها تحت عناوین وفادار، بالقوه و ازدست رفته برچسب گذاری شدند که باید سیاست‌های متفاوتی را در قبال هر یک از انها اتخاذ نمود. مشتریان وفادار در اکثر شاخصه‌ها امتیازات مثبتی دارند و باید برنامه‌هایی برای وفادار نگه‌داشتن آنان ایجاد نمود. مشتریان بالقوه در اکثر شاخصه‌ها در وضعیت میانی هستند که باید بیشترین میزان برنامه‌های تبلیغاتی و ترویجی را معطوف به آن‌ها نمود. گروه دیگری نیز مشتریان از دست رفته هستند که خرید کمی از این خرده‌فروشی صورت داده‌اند و ارتباط طولانی مدتی هم با مجموعه نداشته‌اند. سیاست‌های تبلیغاتی در قبال این دسته از مشتریان باید نقش یادآور داشته باشد.

در مرحله‌ی بعد برای پیش‌بینی رفتار مشتریان، فراتر از کارهای قبلی انجام شده [۱؛ ٦؛ ۹؛ ۱۰] که به طبقه‌بندی مشتریان نپرداخته بودند، سه روش طبقه‌بندی SVM، DTC و MLP مورد بررسی قرار گرفتند. مقایسه روش‌های مختلف طبقه‌بندی مشتریان با استفاده از معیار LRFMSP نشان می‌دهد که در بین معماری‌های شبکه عصبی بررسی شده شبکه MLP با یک لایه پنهان شامل ٦ نورون به خوبی توانایی طبقه‌بندی مشتریان را داشته و با دقت ۹۹٫۸۱ درصد مشتریان را با بیشترین دقت طبقه‌بندی می‌کند. در حالیکه رحیم و همکاران با استفاده

از معیار RFM به دقت ۹۷ درصد رسیده بودند و چاشمی و همکاران با معیار LRFMS پیش‌بینی را با دقت ۹۰ درصد انجام دادند.

همچنین نتایج نشان داد که در میان کرنل‌های مورد بررسی برای طبقه‌بندی‌کننده SVM، کرنل چندجمله‌ای بهترین نتیجه را با دقت ۹۹,٦٤ درصد در پی دارد، در حالیکه رحیم و همکاران که به طبقه‌بندی رفتار مشتریان با معیار RFM پرداخته بودند، با استفاده از طبقه‌بندی‌کننده SVM به دقت ۹٤ درصد و یوریسیچ و همکاران با استفاده از معیار RFM و با کمک طبقه‌بندی‌کننده‌های SVM و DTC به دقت ۹۳,۳۳ رسیدند.

در نهایت نتایج پژوهش نشان داد که طبقه‌بندی مشتریان با معیار LRFMSP و استفاده از درخت تصمیم‌گیری CART، دقت ۹۹,۱٦٪ 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 L، S و P تأثیر مثبتی بر دقت طبقه‌بندی‌کننده خواهد گذاشت. مقایسه سرعت و دقت طبقه‌بندی‌کننده‌ها نشان می‌دهد که درخت تصمیم گرچه دقت کمی پایینتری 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.

مدیران صنعت خصوصا مدیران فروش و بازاریابی می‌توانند برای بررسی و شناسایی دقیق‌تر رفتار مشتریان‌شان، با استفاده از داده‌های صنعت مربوطه، مدل را به روزرسانی نموده و مطابق با صنعت خود آموزش دهند تا به نتایج دقیق‌تری متناسب با رفتار مشتریان خود دست یابند. به این ترتیب مدیران می‌توانند بهتر و دقیق‌تر مشتریان خود را شناخته و الگوهای رفتاری خرید آن‌ها را درک نمایند. از این رو می‌توانند متناظر با هر مشتری راهبرد بازاریابی متناسب با آن را پیاده کنند.

همچنین به مدیران پیشنهاد می‌شود که با استفاده از مدل پیاده‌سازی شده، تصمیم خرید آینده مشتری را به طور خودکار پیش‌بینی کرده و بر اساس آن در طراحی کمپین‌های تبلیغاتی هدفمند، توسعه محصولات، برنامه‌ریزی‌های کوتاه‌مدت، میان‌مدت و بلندمدت راهکارهای مفید ارائه دهند.

پیشنهاد می‌شود در پژوهش‌های آتی متغیرهای پروفایلی همچون سن، جنسیت، تحصیلات، درآمد افراد و منطقه جغرافیایی فرد یا فروشگاه را در کنار متغیرهای این تحقیق به عنوان پارامترهایی ورودی به مدل جهت خوشه‌بندی مشتریان مورد استفاده قرار گیرد. همچنین پیشنهاد می‌شود که محققان با استفاده از سایر تکنیک‌های داده‌کاوی، مانند یادگیری عمیق به طبقه‌بندی مشتریان بپردازند. علاوه بر این محققان می‌توانند ویژگی‌های رفتاری کیفی مشتریان را با استفاده از روش‌های مبتنی بر منطق فازی به مدل اضافه کنند.

وجود جشنواره‌های فروش، کمپین‌های بازاریابی و فصول مختلف خرید نیز ممکن است بر رفتار خرید مشتریان تأثیر گذار بوده و نتایج را تحت تأثیر قرار دهد. بنابراین پیشنهاد می‌شود که اثر ناشی از این‌گونه موارد با کمک روش‌های شناسایی روند، از داده‌ها حذف شود یا به گونه‌ای این اثر روی داده‌ها تعدیل گردد.

۶- منابع

[1] 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.

[2] 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.

[3] 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].

[4] Bashardoust, O., Asgharizadeh, E., & Afshar Kazemi, M. A. (2022). Customers Clustring 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].

[5] 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.

[6] 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.

[7] 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].

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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].

[13] 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].

[14] Birant, D. (2011). Data mining using RFM analysis. In Knowledge-oriented applications in data mining. IntechOpen, https://doi.org/10.5772/13683.

[15] Djurisic, V., Kascelan, L., Rogic, S., & Melovic, B. (2020). Bank CRM optimization using predictive classification based on the support vector machine method. Applied Artificial Intelligence, 34(12), 941-955, https://doi.org/10.1080/08839514.2020.1790248.

[16] 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.

[17] 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.

[18] 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.

[19] 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].

[20] 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].

[21] Guoxiu, Liang, (2005). A Comparative Study of Three Decision, Tree Algorithms: ID3, Fuzzy ID3 and Probabilistic Fuzzy ID3. Bachelor Thesis, Informatics & Economics Erasmus University, Rotterdam, the Netherlands, PP. 5-7.

[22] 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].