

## تحلیل داده محور مشتریان با استفاده از الگوریتم‌های k-means و Apriori در چارچوب روش crisp توسعه یافته

مریم جمشیدی مقدم<sup>۱</sup>، سید رئوف خیامی (نویسنده مسئول)<sup>۲</sup>

<sup>۱</sup> کارشناسی ارشد، آزمایشگاه معماری سازمانی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز

<sup>۲</sup> دانشیار، آزمایشگاه معماری سازمانی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز

### چکیده

امروزه در فضای رقابتی صنعت و کسب و کارها، شناسایی الگوهای رفتاری مشتریان بخصوص مشتریان هدف و بخش‌بندی دقیق آن‌ها نقش بسزایی در ارائه استراتژی‌های بازاریابی و افزایش سودآوری دارد. در این پژوهش، با بهره‌گیری از رویکردهای داده‌کاوی، به تحلیل داده‌های تراکنش مشتریان پرداخته شده است. ابتدا با استفاده از الگوریتم خوشه‌بندی K-Means مبتنی بر شاخص های RFM، مشتریان به چندین خوشه تقسیم‌بندی شدند. سپس با به‌کارگیری الگوریتم Apriori، قوانین انجمنی میان اقلام خریداری شده استخراج گردید تا الگوهای پنهان در رفتار خرید مشتریان شناسایی شود. چارچوب پیشنهادی بر اساس روش توسعه یافته CRISP-DM می‌باشد و بر روی مجموعه داده‌های فروش مشتریان شرکت رامک فارس مورد ارزیابی قرار گرفته است. نتایج نشان می‌دهد که رویکرد ترکیبی ارائه‌شده قادر است بینش‌های ارزشمندی در خصوص ترجیحات خرید مشتریان و فرصت‌های نهفته در فروش متقابل فراهم آورد. این یافته‌ها می‌توانند به‌عنوان ابزاری مؤثر در تصمیم‌گیری‌های بازاریابی هدفمند مورد استفاده قرار گیرند.

**کلمات کلیدی:** داده‌کاوی، خوشه‌بندی، صنعت خرده‌فروشی مواد غذایی، قوانین انجمنی، هوش مصنوعی، الگوریتم Apriori، الگوریتم k-means، RFML، RFMD، RFM

### تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۲/۱۲/۰۸

تاریخ اصلاحات: ۱۴۰۳/۰۵/۱۱

تاریخ پذیرش: ۱۴۰۳/۰۶/۱۰

تاریخ انتشار: ۱۴۰۳/۰۶/۳۰

ایمیل نویسنده مسئول: maryam\_jm75@yahoo.com

### ۱ - مقدمه

باتوجه به رشد سریع و پیشرفت فناوری در صنعت، حجم اطلاعات و داده‌ها به طور چشمگیری افزایش یافته است. این افزایش حجم داده‌ها در کسب‌وکارها، داده‌کاوی را به موضوعی مهم و جالب در ایران و سایر کشورها تبدیل کرده است [۱].

داده‌کاوی علمی است که باتکیه بر هوش مصنوعی و علم داده، به ما امکان می‌دهد از حجم عظیم داده‌ها اطلاعات مفیدی استخراج کنیم،

چراکه داده‌های خام معنی خاصی ندارند و به طور مستقیم دانش کاربردی در اختیار ما نمی‌گذارند.

داده‌کاوی یک فرایند در حال توسعه است که در حوزه‌های مختلفی همچون سلامت و بهداشت، بانک، صنایع، سازمان‌ها و کسب‌وکارها کاربرد دارد [۱].

یکی از حوزه‌های پر کاربرد در داده‌کاوی، صنعت خرده‌فروشی است. این صنعت به دلیل گستردگی خرید و فروش‌ها و تعاملات فراوان با مشتریان، حجم زیادی از داده‌های مربوط به فروش، تاریخچه خرید مشتریان و حمل‌ونقل کالا را تولید می‌کند [۷].

حجم بالای داده‌ها در صنعت خرده‌فروشی، استخراج اطلاعات را بدون استفاده از ابزارها و تکنیک‌های پیشرفته برای انسان بسیار دشوار کرده است. هر شرکتی نیاز دارد مقیاس‌پذیر باشد و توانایی پیش‌بینی رفتار مشتریان آینده را برای افزایش سود خود داشته باشد. اگرچه روش‌های سنتی تحلیل داده همچنان مؤثر هستند، اما این روش‌ها زمان‌بر بوده و به نیروی انسانی زیادی نیاز دارند؛ از این‌رو تکنیک‌های داده‌کاوی می‌توانند جایگزین مناسبی باشند [۷].

امروزه تحول در نحوه جمع‌آوری، سازماندهی و بهره‌برداری از داده‌ها به دلیل گستردگی و حجم بالای آن‌ها، چالش‌های جدیدی را در ذخیره‌سازی و پردازش داده‌ها ایجاد کرده است. در این مسیر، داده‌ها به عنصری پویاتر تبدیل شده‌اند. به‌ویژه در صنعت خرده‌فروشی، داده‌کاوی در تحلیل رفتار مشتریان به ما انواع مشتریان را نشان می‌دهد که مصرف‌کنندگان بیشتر از چه مکان‌هایی خرید می‌کنند، کدام محصولات را ترجیح می‌دهند و کدام کالاها با هم خریداری شده‌اند [۸]. این اطلاعات به خرده‌فروشان کمک می‌کند تا شناخت بهتری از مشتریان و نیازهای آن‌ها داشته باشند و محصولات پر فروش را شناسایی کنند. در نهایت، می‌توان استراتژی‌های مناسبی برای هر گروه از مشتریان تدوین کرد. این فرایند در مدیریت ارتباط با مشتریان به کار می‌رود که هدف آن بهبود ارتباط با مشتری و یافتن راهکارهایی برای افزایش فروش است [۱].

شرکت‌ها باید درک عمیق‌تری از داده‌های مشتریان در تمام ابعاد داشته باشند و همچنین به شناسایی شباهت‌ها و تفاوت‌های بین مشتریان، پیش‌بینی رفتارهای آن‌ها، ارائه پیشنهادهای بهتر، و ایجاد فرصت‌های مناسب برای تعامل بیشتر مشتری با شرکت بپردازند. این

### ۲-۳: الگوریتم Apriori

این تکنیک یکی از الگوریتم‌های پرکاربرد قواعد انجمنی است و می‌تواند با شناسایی و استخراج آیت‌مست‌های پرتکرار ارتباط بین فروش محصولات و رفتار مشتریان را بررسی کند [۵]. این الگوریتم دارای چند پارامتر ارزشیابی است که در این تحقیق به بررسی دو پارامتر support و confidence پرداخته شده است.

### ۳- روش‌شناسی پژوهش

#### ۳-۱: روش کریسپ

روش کریسپ<sup>۱</sup> یکی از مدل‌های فرایندی است که هدف آن قابل‌مدیریت‌تر کردن پروژه‌های داده‌کاوی است [۱۵] و شامل ۶ مرحله زیر می‌باشد:

۱- درک کسب‌وکار<sup>۲</sup>، ۲- درک داده‌ها<sup>۳</sup>، ۳- آماده‌سازی داده‌ها<sup>۴</sup>،

۴- مدل‌سازی<sup>۵</sup>، ۵- ارزیابی<sup>۶</sup>، ۶- استقرار<sup>۷</sup>

این روش، انجام فرایند داده‌کاوی را قابل‌اعتمادتر، قابل‌مدیریت‌تر و سریع‌تر می‌کند [۲] اما یکی از نقاط ضعف این روش این است که هیچ فاز یا زیر فازی باهدف الگوبرداری از تحقیقات و کارهای انجام شده مشابه وجود ندارد که باتوجه‌به پروژه‌های مشابه در آزمایشگاه معماری سازمانی دانشگاه صنعتی شیراز [۲] مرحله BenchMarking به این روش اضافه گردید [۱] که در شکل ۱ آمده است.

استفاده از این روش توسعه یافته نقاط قوت و ضعف مدل‌های قبلی را تحلیل کرده، اهداف و انتظارات پروژه را با استفاده از دانش و تجربیات پیشین برای جهت دهی بهتر به پروژه و تعیین مسیرهای تحقیق براساس بررسی‌های انجام شده مشخص می‌کند.

در مدل توسعه یافته این متد، پروژه با بررسی و تحلیل داده‌ها و نتایج پیشین آغاز می‌شود و با تمرکز بر تجربیات مستند و تحلیل تحقیقات گذشته می‌تواند از خطاهای تکراری جلوگیری کرده، فرایند را تسریع کند و منجر به انتخاب بهینه الگوریتم‌ها و تکنیک‌ها شود.

اضافه کردن مرحله BenchMarking اطلاعات و داده‌های ارزشمندی را به مراحل بعدی ارائه می‌دهد.

اقدامات برای بهبود نرخ حفظ مشتری و طراحی برنامه‌های وفاداری از اهمیت ویژه‌ای برای مدیران بازاریابی برخوردار است [۹]. باتوجه‌به اهمیت زیاد سوددهی در صنعت خرده‌فروشی در این بازار رقابتی، هدف این مطالعه تحلیل داده‌های فروش و مشتریان و شناخت انواع مشتریان با استفاده از تکنیک‌های هوش مصنوعی و داده‌کاوی است تا بدین‌طریق بتوان به مدیران جهت تصمیم‌گیری دقیق‌تر و مؤثرتر کمک کرد.

### ۲- ادبیات موضوع

برای تحلیل داده‌های مشتریان و پیدا کردن ارتباط بین کالاهای خریداری شده توسط هر مشتری، از تکنیک‌ها و مدل‌های مختلفی استفاده شده که در ادامه به توضیح آن‌ها پرداخته شده است.

#### ۲-۱: خوشه بندی

یکی از تکنیک‌های قدرتمند در خوشه‌بندی الگوریتم k-means است. این الگوریتم جز روش‌های یادگیری بدون نظارت است که برای تجزیه و تحلیل و خوشه‌بندی داده‌های مشتریان استفاده می‌شود [۱۰] و برای شناخت رفتار خرید مصرف کنندگان و انواع مشتریان هدف، داده‌ها خوشه‌بندی می‌شوند.

همچنین برای تخمین ارزش هر مشتری از مدل‌های RFM استفاده می‌شود. در مدل RFM، R به معنی تازگی خرید، F به معنی تعداد خرید و M به معنی مقدار پولی که مشتری در یک دوره زمانی هزینه می‌کند و RFM‌های توسعه یافته همچون RFMD که D به معنی تنوع کالاهای خریداری شده توسط مشتری [۱۱] و RFML که L به معنی طول مدت ارتباط مشتری با شرکت [۱۲] است. این مدل‌ها با محاسبه پارامترهایشان، تقسیم بندی بازار را با استفاده از تکنیک‌ها و الگوریتم‌های خوشه‌بندی انجام می‌دهند [۱۳].

#### ۲-۲: قوانین انجمنی

یکی دیگر از تکنیک‌های داده‌کاوی کشف قوانین انجمنی است که این کار اصطلاحاً با نام تجزیه و تحلیل سبد خرید بازار متداول است. تجزیه و تحلیل سبد بازار یکی از روش‌های کلیدی است که توسط خرده‌فروشان برای کشف ارتباط بین اقلام مورد استفاده قرار می‌گیرد، به عبارت دیگر به کسب‌وکارها اجازه می‌دهد تا روابط بین اقلامی که مشتریان می‌خرند را شناسایی کنند [۱۴].

جهت بهبود خدمات مشتری، استراتژی‌های فروش، توزیع مؤثر [۱۳] و بسیاری موارد دیگر است.

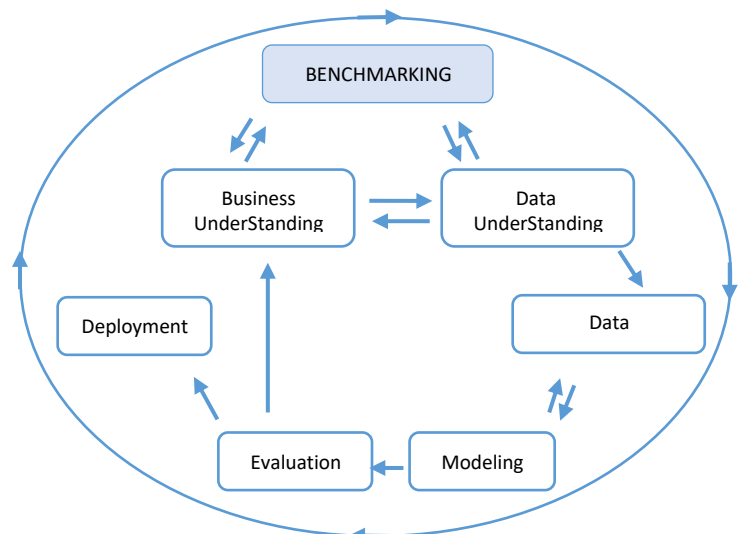
محدوده وسیعی از تحقیقات برای روش‌های هوش مصنوعی و تکنیک‌ها داده‌کاوی در حوزه صنعت شناسایی شده است.

تحقیقی [۴] در سال ۲۰۱۸ بر اساس استاندارد جهانی CRISP-DM با الگوریتم k-means و همچنین با استفاده از مدل RFM به تحلیل رفتار مشتریان پرداخته است که در نتیجه این مقاله مشخص شده شناسایی مشتریان کلیدی و تمرکز بر آن‌ها، به کسب‌وکارها این امکان را می‌دهد که زمان و منابع کمتری را صرف مشتریانی با سود کمتر کنند و در نتیجه، کارایی و بازدهی خود را افزایش دهند.

در رقابت تجاری امروز شرکت‌ها به دنبال حفظ مشتریان برتر خود و نیز فروش بیشتر محصولاتشان به مشتریان فعلی هستند، بنابر این موضوع برای شناسایی نیازها و خواسته‌های مشتریان با رفتار مشابه تحقیقات بسیاری انجام گرفته که هدف آنها اندازه‌گیری ارزش طول عمر مشتری و دسته‌بندی آنها بود [9]. به دلیل اینکه یکی از دغدغه‌های شرکت‌های بزرگ، نگرانی برای ریزش و حفظ مشتریان است، لذا برای درک نیازهای گروه‌های هدف در شرکت‌ها، به دسته‌بندی داده‌ها پرداخته می‌شود [17]. برای این منظور در تحقیقات سال‌های اخیر از مدل RFM یا RFMهای توسعه‌یافته که با الگوریتم خوشه‌بندی k-means ترکیب شده، استفاده شده است که این روش به شرکت‌ها کمک می‌کند تا به شناخت نیازها و ترجیحات مصرف‌کنندگان خود رسیده و درک بهتری از بازار هدف خود به دست آورند.

به‌منظور ارزش‌گذاری مشتریان به روشی دیگر در مقاله [12] پیشنهاد می‌گردد که بر اساس مدل RFML بر روی داده‌های ۶ ماهه و ۷۰۰۲ داده تراکنش مشتری با الگوریتم c-means فازی خوشه‌بندی انجام شود و در نتیجه با بررسی خوشه‌ها بتوان مشتریان سودآور را شناسایی کرد.

باهدف افزایش وفاداری مشتریان و افزایش درآمد از طریق بهینه‌سازی قرارگیری محصولات و همچنین شناسایی الگوهای خرید مصرف‌کنندگان، مطالعه‌ای [6] بر روی داده‌های فروش یک خرده‌فروشی در کنیا با استفاده از الگوریتم Apriori انجام شد و در نهایت ۴۲ قانون انجمنی برای روابط بین محصولات استخراج شدند تا با بررسی و تجزیه تحلیل‌ها به خرده‌فروشان جهت درک بهتر



شکل ۱: روش کریسپ توسعه‌یافته [۳]

#### ۴- طرح مسئله و هدف تحقیق

باتوجه به اینکه هزینه جذب مشتریان جدید به طور قابل توجهی بیشتر از حفظ مشتریان فعلی است، یکی از اصلی‌ترین دغدغه‌های کسب‌وکارها این است که چگونه محصولات بیشتری را به مشتریان کنونی بفروشند، برای این منظور از تکنیک خوشه‌بندی مشتریان بر اساس معیارهای مختلف و گروه‌بندی آن‌ها برای درک بهتر رفتار و نیازهایشان استفاده می‌شود، این دسته‌بندی کمک می‌کند تا برای هر گروه مشتریان، یک استراتژی بازاریابی یا مدیریتی خاص و منحصربه‌فرد تدوین [۱۶] و الگوهای مناسب جهت رتبه‌بندی و خوشه‌بندی مشتریان کشف شود.

هدف این پژوهش، کشف قوانین بین کالاهای خریداری شده توسط مشتریان است که برای این منظور طبق روش CRISP-DM توسعه یافته پیش رفته و بر اساس الگوبرداری از تحقیقات پیشینی که در این حوزه انجام گرفته ابتدا با استفاده از روش‌های هوش مصنوعی داده‌های مشتریان را خوشه‌بندی کرده و سپس با قوانین انجمنی روابط بین کالاهای خریداری شده توسط مصرف‌کنندگان در هر خوشه استخراج می‌شود تا از این طریق، مدیران شرکت‌ها بتوانند با انتخاب استراتژی‌های کارآمدتر، عملکرد خود را در جهت مدیریت بهتر ارتباط با مشتری بهبود بخشند.

#### ۵- پیشینه تحقیقات

هدف داده‌کاوی در صنعت خرده‌فروشی شناسایی مشتریان بالقوه، شناسایی محصولات پرفروش و ترجیحات مصرف‌کنندگان، شناسایی رابطه‌ی بین فروش محصولات با یکدیگر و سپس ارائه راهکارهایی

ابتدا داده‌ها با استفاده از مدل RFM ، RFMD و RFML با تکنیک k-means خوشه‌بندی شده‌اند، سپس با بهره‌گیری از تکنیک‌های قوانین انجمنی، قوانین و الگوهای پنهان در هر خوشه استخراج شده است به این صورت که روابط بین کالاهای خریداری شده توسط مشتریان در هر خوشه به دست آورده می‌شود.

#### ۷- خوشه‌بندی داده‌ها با مدل RFM و RFMD های توسعه یافته و الگوریتم k-means

در مرحله اول بر روی داده‌های کل سال برطبق روش کریسپ توسعه‌یافته پیش‌رفته و بر اساس مدل RFM با تکنیک k-means خوشه‌بندی انجام شده و در نهایت نام‌گذاری پیشنهادی برای هر خوشه مطرح گردید.

برای انتخاب تعداد بهینه خوشه‌ها از روش Elbow و محاسبه مقدار Silhouette استفاده شد که نتایج تعداد مناسب خوشه‌ها با مدل RFM در جدول ۱ آمده است:

(جدول ۱-): تعیین تعداد مناسب خوشه‌ها در مدل RFM

برای داده‌های کل سال

تعداد خوشه‌ها (K)	مقدار Silhouette برای هر خوشه‌بندی
K = ۳	۰/۹۰۵
K = ۴	۰/۵۱۹
K = ۵	۰/۵۴۸
K = ۶	۰/۵۵۷
K = ۷	۰/۴۷۱

باتوجه به نتایج به دست آمده بهترین مقدار Silhouette برای خوشه‌بندی با مدل RFM ، برای تعداد ۳ خوشه و سپس ۶ خوشه است به همین دلیل نام‌گذاری خوشه‌ها را برای تعداد  $k=3$  و  $k=6$  انجام می‌شود که نتیجه نام‌گذاری پیشنهادی برای ۳ خوشه در جدول ۲ و برای ۶ خوشه در جدول ۳ آمده است.

(جدول ۲-): نام‌گذاری خوشه‌بندی RFM برای داده‌های کل

سال با  $k=3$

R	F	M	نام‌گذاری خوشه‌ها
∇	∧	∇	مشتریانی که غالباً و به تعداد بیشتر خرید می‌کنند
∧	∇	∇	مشتریان روی گردان یا نیاز به توجه
∇	∧	∧	قهرمانان

الگوهای خرید مشتریان کمک کنند و استراتژی‌های بهتری برای مدیریت چیدمان محصولات اتخاذ کنند.

برای شناسایی بهتر رفتار مشتریان تحقیقی دیگر [۱۴] بر روی ۹۸۳۵ رکورد تراکنش یک شرکت خرده‌فروشی با الگوریتم Apriori انجام شد و در نتیجه مشخص شد که یافتن این روابط به خرده‌فروشان کمک می‌کند تا با در نظر گرفتن اقلامی که اغلب با هم خریداری می‌شوند، می‌توان استراتژی‌های فروش را بسته به هر نوع مشتری توسعه داد.

همچنین در تحلیل داده‌های مربوط به فروشگاه‌های خرده‌فروشی و شناسایی الگوهای پرتکرار در خرید، در تحقیقی دیگر [18] سبد خرید هر مشتری<sup>۸</sup> بر روی داده‌های ۵۲۲ هزار مشتری بررسی شد و با مقایسه دو الگوریتم Apriori و FP-Growth بهینه‌ترین روش برای استخراج آیت‌های پرتکرار و تولید قوانین ارتباطی کشف گردید تا منجر به کمک به کسب‌وکارها جهت شناسایی رفتار مشتریان و اجرای استراتژی‌های بازاریابی هدفمند شوند.

در مطالعه‌ای [۱۹] دیگر به منظور بهبود کارایی در مورد ارائه بهتر محصولات به مصرف‌کنندگان با به کارگیری الگوریتم Apriori روابط بین کالاهایی که با هم در ارتباط هستند و نیز اقلامی که بیشترین فروش را دارند مشخص شد و از این اطلاعات برای تعیین استراتژی‌های فروش استفاده گردید.

برای پیش‌بینی فروش در مقاله [۲۰] با الگوریتم Apriori الگوها و آیت‌های پرتکرار با محاسبه دو مقدار support و confidence در داده‌های مشتریان شناسایی شده و منجر به بهبود استراتژی‌های کسب و کار و نگهداشت مشتریان می‌شوند.

#### ۶- مراحل کار و پایگاه داده

باتوجه به اینکه هدف این پژوهش موضوع امکان‌سنجی دسته‌بندی مشتریان مبتنی بر تحلیل داده است و در این راستا از تکنیک‌های هوش مصنوعی برای تحلیل و دسته‌بندی داده‌های فروش و مشتریان استفاده می‌شود، می‌توان تحلیل‌های مختلفی را بر روی پایگاه داده‌های فروش انجام داد و مشتریان را خوشه‌بندی کرد.

این تحلیل بر روی ۳۰۷۰ مشتری از دیتاست شرکت فرآورده‌های لبنی رامک شیراز انجام شده و بر اساس روش توسعه‌یافته کریسپ پیش‌رفته است.

<sup>8</sup> Market Basket Analysis

### ۸- تحلیل سبد خرید مشتریان

باتوجه به اینکه کار این تحقیق بر روی امکان‌سنجی تکنیک‌های هوش مصنوعی با داده‌کاوی در خرده‌فروشی است، به‌جز بحث مربوط به دسته‌بندی مشتریان، در مرحله بعد از یکی دیگر از انواع داده‌کاوی توصیفی به نام Association Rules استفاده شد تا بررسی شود آیا می‌توان ارتباط بین کالاهای خریداری شده توسط مصرف‌کنندگان را به دست آورد که مبنای این روش بر این اساس است که وقتی مشتری یک کالایی را خرید، چقدر احتمال دارد که کالای دیگر را همراه با کالای اول خریداری کند.

ایده دوم روی یکی از الگوریتم‌های Association Rules به نام Apriori انجام شد. این الگوریتم مجموعه‌های یکی، دوتایی و سه‌تایی از اقلام را پیدا می‌کند که چند دفعه با هم تکرار شده‌اند و ارتباطات و قوانین پرتکرار را انتخاب می‌کند.

در تحلیل قوانین به‌دست‌آمده توسط الگوریتم Apriori یک‌سری قانون به دست آورده شد که شاخص انتخاب قوانین مهم‌تر و معتبرتر دو پارامتر به نام support و confidence است.

به‌منظور کشف قوانین و ارتباطات کالاها، بر روی خوشه‌های به‌دست‌آمده از داده‌های کل سال تحلیل و بررسی‌هایی انجام شد.

برای انجام تحلیل سبد خرید، نیاز به سیستم پردازشی قدرتمندتری است، به همین دلیل از سیستمی با هارد ۱T SSD، RAM 32G، Core i7، و باقابلیت GPU استفاده شد.

تحلیل Apriori را بر روی سه خوشه به‌دست‌آمده و همچنین بر روی داده‌های کل سال نیز انجام شد که طبیعتاً این نتیجه حاصل می‌گردد که قوانین به‌دست‌آمده برای هر خوشه، قوانین مناسب‌تر و کاربردی‌تر برای تصمیم‌گیری در مورد آن خوشه‌ها است تا اینکه قوانین به‌صورت کلی به دست آورده شوند.

نتایج قوانین و ارتباطات به‌دست‌آمده برای ۳ خوشه از داده‌های کل سال در جداول ۷ و ۸ و ۹ آمده که به‌عنوان نمونه، قوانین با بیشترین support و confidence در جداول قرار گرفته است.

(جدول ۷): قوانین انجمنی روی خوشه صفر برای  $k=3$  در

#### داده‌های کل سال

antecedents		consequents	supp	conf
shir_nim_946	→	shir_por_946	۰/۴۰۲	۰/۸۹
shir_por_946	→	shir_nim_946	۰/۴۰۲	۰/۷۳
khame_kaka_100	→	khame_asali_100	۰/۱۴۹	۰/۵۶
khame_asali_100	→	khame_kaka_100	۰/۱۴۹	۰/۶۶۷

### (جدول ۳): نام‌گذاری خوشه‌بندی RFM برای داده‌های کل

سال با  $k=6$

R	F	M	نام‌گذاری خوشه‌ها
∇	∇	∇	مشتریان تازه‌وارد
∇	∧	∧	مشتریان وفادار
∇	∧	∧	قهرمانان
∧	∇	∇	مشتریان در معرض خطر
∇	∧	∇	مشتریانی که می‌توانند به مشتریان وفادار تبدیل شوند
∧	∇	∇	مشتریان نیاز به توجه

برای نام‌گذاری پیشنهادی میانگین R و F و M در هر خوشه محاسبه شده و با میانگین R و F و M در کل داده‌ها مقایسه می‌شود و سپس باتوجه به بیشتر یا کمتر بودن مقادیر R و F و M نسبت به میانگین کلی، برای هر خوشه نام‌گذاری پیشنهاد می‌شود. در ادامه برای اطمینان از اینکه آیا دسته‌بندی داده‌های مشتریان با مدل RFM خوشه‌های همبسته‌تری تولید می‌کند یا نه با مدل RFMD و RFML نیز خوشه‌بندی انجام گرفت تا مقایسه شود اضافه کردن پارامتر D و L چه تأثیری نسبت به خوشه‌بندی با مدل RFM دارد. نتایج خوشه‌بندی با مدل RFMD در ۷ دسته کلی کالا و همچنین در ۸۲ نوع کالا در جداول ۴ و ۵ آمده است:

### (جدول ۴): تعیین تعداد مناسب خوشه‌ها در مدل RFMD

برای داده‌های کل سال بر روی ۷ دسته کالا

مقدار Silhouette برای هر خوشه‌بندی	تعداد خوشه‌ها (K)
۰/۵۱	$K = 9$
۰/۵۲	$K = 10$

### (جدول ۵): تعیین تعداد مناسب خوشه‌ها در مدل RFMD

برای داده‌های کل سال بر روی ۸۲ دسته کالا

مقدار Silhouette برای هر خوشه‌بندی	تعداد خوشه‌ها (K)
۰/۴۵	$K = 3$
۰/۴۶	$K = 4$

همچنین نتیجه خوشه‌بندی با مدل RFML در جدول ۶ آمده است:

### (جدول ۶): تعیین تعداد مناسب خوشه‌ها در مدل RFML

مقدار Silhouette برای هر خوشه‌بندی	تعداد خوشه‌ها (K)
۰/۵۳	$K = 3$
۰/۴۲	$K = 4$

باتوجه به مقایسه مقدار Silhouette در مدل‌های RFM و RFMD و RFML این نتیجه حاصل می‌شود که خوشه‌بندی با مدل RFM بر روی داده‌هایی که ما در اختیار داریم، خوشه‌های هم‌بسته‌تری را تولید می‌کند، لذا در مرحله بعد قوانین انجمنی بر روی خوشه‌های به‌دست‌آمده از مدل RFM انجام می‌شود.

اما قوانین دیگر در خوشه‌ها نشان می‌دهند مصرف‌کنندگان رفتار متفاوتی در سبد خرید خود داشته‌اند که با تحلیل و ارزیابی قوانین و ارتباطات دیگر در هر دسته می‌توان فرصت‌های بازاریابی و ترجیحات و عادات مصرف‌کنندگان را کشف کرد و به مدیران صنایع این امکان داده می‌شود تا الگوها و قوانین فروش را شناسایی و تصمیمات آگاهانه‌ای در مورد بسته‌بندی یا جای‌گذاری محصولات و استراتژی‌های بازاریابی هدفمند برای انواع مشتریان هدف اتخاذ کنند تا در نهایت منجر به بهبود مدیریت ارتباط با مشتریان شود.

همچنین از نتایج به‌دست‌آمده از این پژوهش می‌توان برای چیدمان محصولات و نیز سیستم‌های پیشنهاددهنده و یا در توسعه نرم‌افزار ویزیتورها استفاده کرد به طوری که با توجه به تقاضایی که مشتری می‌دهد، پیشنهادهای مؤثرتری به آن داده شود؛ همچنین در تعیین تخفیف برای مشتریان و تعیین اینکه چه سیاستی را برای چه گروه از مشتریان می‌توان اعمال کرد، استفاده نمود که به تبع موجب مؤثرتر شدن سیاست‌ها و صرفه‌جویی در هزینه‌ها برای اهداف تبلیغاتی می‌شود که این مسئله یکی از مهم‌ترین دغدغه‌های صنایع و شرکت‌های بزرگ است.

به‌طور کلی می‌توان گفت تحلیل داده‌های فروش، خوشه‌بندی و استفاده از قوانین انجمنی راهکار مناسبی جهت شناسایی بهتر مشتریان هدف و نیازهای آنها، پیش برد اهداف و افزایش سود برای شرکت‌ها را در بر دارد و این نتایج را می‌توان در اختیار مدیران قرارداد تا در بحث‌های انتخاب استراتژی‌های مناسب برای افزایش فروش و برنامه‌ریزی برای توزیع کمک‌کننده باشد.

#### سپاس‌گزاری

با تشکر و تقدیر ویژه از مدیریت محترم شرکت لبنی رامک که در انجام این پژوهش ما را حمایت و همراهی کردند.

#### ۱۰- منابع

- [1] M. JamshidiMoghaddam and S. R. Khayami, "Identifying performance improvement solutions in the retail industry using data mining," *The 8th Conference on Advances in Iranian Organizational Architecture, Tarbiat Modares University, Tehran, Nov. 2024* (InPersian).
- [2] Mohammad Mohammadinia, "Designing a roadmap for data mining projects for healthcare centers," *Shiraz University of Technology, Shiraz, 2022* (InPersian).
- [3] M. JamshidiMoghaddam, "Identifying performance improvement solutions in the retail industry using data mining," *Shiraz University of Technology, Shiraz, 2024* (InPersian).
- [4] Reza Samizadeh and Roxana Taheri, "Evaluating customer loyalty in a customer relationship management (CRM) structure using k-means clustering and RFM analysis method

#### (جدول ۸-): قوانین انجمنی روی خوشه یک برای $k=3$ در

##### داده‌های کل سال

antecedents		consequents	supp	conf
shir_esl_por_946	→	shir_esl_nim_946	۰/۶۱۸	۰/۸۴۳
shir_esl_nim_946	→	shir_esl_por_946	۰/۶۱۸	۰/۹۲۵
shir_esl_por_946	→	shir_kakao_200	۰/۲۷۷	۰/۳۷۸
shir_kakao_200	→	shir_esl_por_946	۰/۲۷۷	۰/۷۹۲

#### (جدول ۹-): قوانین انجمنی روی خوشه دو برای $k=3$ در

##### داده‌های کل سال

antecedents		consequents	supp	conf
shir_por_946	→	shir_nim_946	۰/۸۴	۰/۹۶
shir_nim_946	→	shir_por_946	۰/۸۴	۰/۹۷
mastm_khame_900	→	shir_por_946	۰/۷۱	۰/۸۸
shir_por_946	→	mastm_khame_900	۰/۷۱	۰/۸۱

در نتیجه قوانین و الگوهای به‌دست‌آمده از قوانین انجمنی، می‌توان ارتباط بین اقلام و کالاهایی که مشتریان خرید می‌کنند را یافت که همان‌طور که مطرح شده بود مبنای آن بر این اساس است که چقدر احتمال دارد وقتی مشتری یک کالایی را می‌خرد، کالای دوم را نیز همراه با کالای اول خریداری کند که به تبع قوانین با support و confidence بیشتر، قوانین بهتری برای تصمیم‌گیری جهت انتخاب راهکارهای مناسب برای افزایش فروش است.

#### ۹- نتیجه‌گیری

در این مطالعه از هوش مصنوعی و تکنیک‌های داده‌کاوی برای شناسایی انواع مشتریان، حفظ و نگهداشت مشتریان فعلی شرکت‌ها استفاده شده است که به بهبود فرایند تصمیم‌گیری و اتخاذ اقدامات مناسب برای مشتریان هدف کمک می‌کند.

در این پژوهش با استفاده از روش کریسپ توسعه یافته، مدل RFM و RFM های توسعه یافته و الگوریتم k-means داده‌های مشتریان خوشه‌بندی و در نهایت مشاهده شد که مدل RFM خوشه‌های هم‌بسته‌تری را بر روی داده‌هایی که در اختیار داریم تولید می‌کند.

با توجه به اینکه هدف این پژوهش بررسی امکان‌سنجی استفاده از هوش مصنوعی در دسته‌بندی و تحلیل رفتار مشتریان بود، در ادامه تحقیق با استفاده از تکنیک قوانین انجمنی، ارتباط بین کالاهای خریداری شده توسط مشتریان در هر خوشه به دست آورده شد.

با بررسی قوانین به‌دست‌آمده از هر خوشه، رفتار خرید مشتریان تحلیل و قوانین پرتکرار خرید در تمام دسته‌ها شناسایی شد که خرید شیر پرچرب و نیم چرب ظاهراً رفتار خرید پرتکرار در هر خوشه است،

in the retail industry,” *Malaysian Journal of Computing (MJoC)*, vol. 9, no. 1, pp. 1746–1758, 2024.

[19] A. Fauzi, A. B. H. Yanto, and N. Indriyani, “Implementation of Data Mining of Organic Vegetable Sales With Apriori Algorithm,” *Jurnal Teknologi dan Open Source*, vol. 6, no. 1, pp. 98–109, 2023.

[20] N. S. Poli and A. S. Sikder, “Predictive Analysis of Sales Using the Apriori Algorithm: A Comprehensive Study on Sales Forecasting and Business Strategies in the Retail Industry.: Predictive Analysis of Sales Using the Apriori Algorithm,” *International Journal of Imminent Science & Technology*, vol. 1, no. 1, pp. 1–16, 2023.



مریم جمشیدی مقدم، کارشناس ارشد مهندسی نرم‌افزار از دانشگاه صنعتی شیراز، در حوزه تحلیل داده، داده کاوی و فناوری اطلاعات فعالیت دارد.

ایشان در حال حاضر به عنوان تحلیل‌گر داده مشغول به کار است و سابقه همکاری در پروژه‌های

کاربردی با رویکرد پاسخ‌گویی به نیازهای صنعت و جامعه را در کارنامه دارد. زمینه‌های تخصصی وی شامل داده‌کاوی، تحلیل سیستم‌های اطلاعاتی، و طراحی وب بوده و مقالاتی نیز در حوزه سلامت و معماری سازمانی ارائه کرده‌اند.

نشانه رایانامه ایشان عبارتند از:

Maryam\_jm75@yahoo.com



سیدرئوف خیامی در حال حاضر دانشیار و مدیر آزمایشگاه تحقیقاتی معماری سازمانی دانشگاه صنعتی شیراز می‌باشد. ایشان تحصیلات خود را در تمامی مقاطع در دانشگاه شیراز با رتبه ممتاز گذرانده و از سال ۱۳۷۶ تاکنون در دانشگاه های

و موسسات آموزش عالی کشور تدریس و همکاری داشته است. حوزه های تخصصی و پژوهشی ایشان معماری سازمانی فناوری اطلاعات، مدیریت فناوری اطلاعات، داده کاوی هوش مصنوعی کاربردی با هدف پاسخگویی به نیازهای جامعه و صنعت کشور می باشد. بطوریکه اکثر فعالیت های کلاس، پروژه های کارشناسی و پایان نامه ها ارشد و رساله های دکتری تحت نظارت ایشان، مبتنی بر نیازهای جامعه و صنعت انجام شده و همکاری علمی و پژوهشی با مجموعه های مردم نهاد و خیریه ها در این خصوص مورد توجه ویژه قرار گرفته است. ایشان سابقه ای پر بار در مشاوره، طراحی و پیاده سازی پروژه های فناوری اطلاعات در صنعت دارند و تاکنون در چندین دوره به عنوان استاد نمونه و پژوهشگر برتر دانشگاه نیز معرفی شده اند.

نشانه رایانامه ایشان عبارتند از:

khayami@sutech.ac.ir

and evaluating clustering validity using the Davis-Bolden index,” 2018(InPersian).

[5] N. S. Poli and A. S. Sikder, “Predictive Analysis of Sales Using the Apriori Algorithm: A Comprehensive Study on Sales Forecasting and Business Strategies in the Retail Industry.: Predictive Analysis of Sales Using the Apriori Algorithm,” *International Journal of Imminent Science & Technology*, Vol. 1, no. 1, pp. 1–16, 2023..

[6] E. J. Omol, D. A. Onyango, L. W. Mburu, and P. A. Abuonji, “Apriori algorithm and market basket analysis to uncover consumer buying patterns: Case of a Kenyan supermarket,” *Buana Information Technology and Computer Sciences (BIT and CS)*, vol. 5, no. 2, pp. 51–63, 2024.

[7] S. K. Manthri and H. P. Chilakalapudi, “Case studies on data mining in market analysis,” *International Journal of Engineering Trends and Technology, Special Issue*, pp. 180–184, 2017.

[8] Ginni, “What is the role of data mining in the retail industry?,” *tutorialspoint*. [Online]. Available: <https://www.tutorialspoint.com/what-is-the-role-of-data-mining-in-the-retail-industry>

[9] O. Dogan, E. Ayçin, and Z. Bulut, “Customer segmentation by using RFM model and clustering methods: a case study in retail industry,” *International Journal of Contemporary Economics and Administrative Sciences*, vol. 8, 2018.

[10] G. H. Kiani, “Determining profitable products in the retail market with consideration of cash limitation and exhibition periods,” *Journal of Retailing and Consumer Services*, vol. 55, p. 102079, 2020.

[11] M. Y. Smali and H. Hachimi, “New RFM-D classification model for improving customer analysis and response prediction,” *Ain Shams Engineering Journal*, p. 102254, 2023.

[12] M. Husnah and R. A. Vinarti, “Customer Segmentation Analysis Using LRFM Based Product and Brand Dimensions,” presented at the 2023 2nd International Conference for Innovation in Technology (INOCON), IEEE, 2023, pp. 1–6.

[13] S. Yousefi, H. Shabanpour, and R. Farzipoor Saen, “Sustainable clustering of customers using capacitive artificial neural networks: a case study in Pegah Distribution Company,” *RAIRO-Operations Research (2804-7303)*, vol. 55, no. 1, 2021.

[14] A. Jadhav, A. Jadhav, and D. R. D. Jadhav, “Association Rule Mining in Retail: Exploring Market Basket Analysis with Apriori Algorithm,” Available at SSRN 4461121, 2023.

[15] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, “The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks,” *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 28–43, 2024.

[16] A. Ullah et al., “Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation Using Recency, Frequency, Monetary, and Time,” *Sensors*, vol. 23, no. 6, p. 3180, 2023.

[17] I. Maryani, D. Riana, R. D. Astuti, A. Ishaq, and E. A. Pratama, “Customer segmentation based on RFM model and clustering techniques with K-means algorithm,” presented at the 2018 Third International Conference on Informatics and Computing (ICIC), IEEE, 2018, pp. 1–6.

[18] A. A. Hashad, K. W. Khaw, A. Alnoor, and X. Y. Chew, “Exploratory analysis with association rule mining algorithms

**روش ارجاع:** م.جمشیدی مقدم، ر. خیامی، تحلیل داده محور مشتریان با استفاده از الگوریتم‌های k-means و Apriori در چارچوب روش Crisp توسعه یافته، دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال هفتم، شماره ۱، شماره پیاپی ۱۳، صفحه ۹۸ تا ۱۰۵، سال ۱۴۰۳.

**How to cite:** M. Jamshidi, R. Khayami, Customer data analysis leveraging k-means and Apriori algorithms, developed based on the CRISP-DM framework. . Journal of Distributed Computing and Systems (JDACS), Vol 7, Issue 1, Pages 98-105, 2024.

### Customer data analysis leveraging k-means and Apriori algorithms, developed based on the CRISP-DM framework

M.JamshidiMoghaddam<sup>1</sup>, R.Khayami<sup>2</sup>

<sup>1</sup>Shiraz Uneversity Of Technology, Shiraz, Iran

<sup>2</sup>Shiraz Uneversity Of Technology, Shiraz, Iran

In today's highly competitive business and industrial landscape, identifying customer behavior patterns—particularly those of target customers—and accurately segmenting them plays a crucial role in developing effective marketing strategies and increasing profitability. This study employs data mining techniques to analyze customer transaction data. Initially, customers were segmented into several clusters using the K-Means clustering algorithm based on RFM (Recency, Frequency, Monetary) metrics. Subsequently, the Apriori algorithm was applied to extract association rules among purchased items in order to uncover hidden patterns in customer purchasing behavior. The proposed framework is based on an extended CRISP-DM methodology and was evaluated using customer sales data from Ramek Company. The results demonstrate that the proposed hybrid approach can provide valuable insights into customer purchase preferences and reveal hidden opportunities for cross-selling. These findings can serve as an effective tool for informed, targeted marketing decisions.