تحلیل و پیش‌بینی سوددهی مشتریان در صنعت بازرگانی لوله و اتصالات با استفاده از الگوریتم‌های داده‌کاوی و یادگیری ماشین

حامد نادری

دانشجوی دکتری مهندسی صنایع، گروه بهینه‌سازی سیستم‌ها ، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

الهام آخوندزاده

استادیار گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

محمد علی رستگار

استادیار، گروه مدیریت سیستم و بهره‌وری ، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

چکیده

پیشرفت‌های تکنولوژیکی و نیاز فزاینده به طرح‌های عمرانی، کشاورزی و ساختمانی مدرن باعث افزایش اهمیت تأمین لوازم و ابزار مورد نیاز در این پروژه‌ها شده است. در این راستا، شرکت‌های بازرگانی فروش لوله و اتصالات نقشی کلیدی در تأمین نیازهای مشتریان ایفا می‌کنند. هدف این پژوهش، خوشه‌بندی مشتریان و پیش‌بینی سودآوری آن‌ها برای یک شرکت بازرگانی در این حوزه است. برای این منظور، از روش‌های داده‌کاوی و الگوریتم‌های یادگیری ماشین از جمله K-means، Decision Tree، KNN و Naive Bayes استفاده شد. در مرحله اول، با استفاده از روش K-means و تکنیک Elbow، تعداد بهینه خوشه‌ها برابر با ۳ تعیین شد. در مرحله بعد، دسته‌بندی و پیش‌بینی سودآوری مشتریان با استفاده از Decision Tree، KNN و Naive Bayes انجام شد. نتایج ارزیابی مدل‌ها نشان داد که الگوریتم Decision Tree با دقت ۹۸.۷۲٪ بهترین عملکرد را داشته است. این پژوهش نشان می‌دهد که با تمرکز بر مشتریان سودآور و ارائه تخفیف‌ها یا امتیازات ویژه به آن‌ها، شرکت می‌تواند سودآوری خود را بهبود بخشد و همچنین برنامه‌های ترغیبی برای سایر مشتریان طراحی کند.

واژگان کلیدی: : خوشه‌بندی، پیش‌بینی، روش کریسپ، Decision Tree، KNN، Naive Bayes

مقدمه

با رشد سریع تکنولوژی و تغییرات مداوم در نیازهای ساخت و ساز، تأمین مواد و ابزارهای با کیفیت به یک دغدغه اصلی برای شرکت‌های بازرگانی تبدیل شده است. در این میان، شرکت‌های بازرگانی لوله و اتصالات به عنوان یکی از ارکان کلیدی تأمین نیازهای پروژه‌های عمرانی، کشاورزی و ساختمانی، نقش بسزایی در موفقیت و پایداری این پروژه‌ها ایفا می‌کنند. این شرکت‌ها با برخورداری از حجم بالای داده‌های مشتریان و تراکنش‌های مالی، به دنبال بهینه‌سازی فرآیندهای تجاری و افزایش سودآوری هستند.

پیش‌بینی سوددهی مشتریان، به عنوان یکی از ابعاد مهم تحلیل داده‌ها، می‌تواند به شرکت‌های بازرگانی کمک کند تا به‌طور هدفمندتر و مؤثرتری به نیازهای مشتریان پاسخ دهند و منابع خود را به بهترین نحو تخصیص دهند. استفاده از الگوریتم‌های داده‌کاوی و یادگیری ماشین در این زمینه، می‌تواند به شناسایی الگوهای خرید و رفتار مشتریان، و در نهایت بهبود تصمیم‌گیری‌های تجاری کمک کند.

یکی از روش‌های مؤثر در این راستا، خوشه‌بندی مشتریان بر اساس ویژگی‌های خرید و سودآوری آنها است. خوشه‌بندی می‌تواند به دسته‌بندی مشتریان به گروه‌های مشابه بر اساس رفتار خرید و ویژگی‌های دیگر آنها کمک کند، که این امر به نوبه خود می‌تواند به بهینه‌سازی استراتژی‌های بازاریابی و فروش منجر شود. این پژوهش به بررسی استفاده از این تکنیک‌های یادگیری ماشین در صنعت بازرگانی لوله و اتصالات می‌پردازد. هدف این است که با استفاده از الگوریتم‌های یادگیری ماشین، مشتریان را خوشه‌بندی کرده و سودآوری آنها را پیش‌بینی کنیم تا شرکت‌های بازرگانی بتوانند با تمرکز بر مشتریان سودآور و طراحی استراتژی‌های هدفمند، به بهبود عملکرد تجاری خود دست یابند.

مبانی نظری

پیش‌بینی سوددهی مشتریان یکی از مباحث کلیدی در تحلیل کسب‌وکار و بازاریابی است که با استفاده از تکنیک‌های داده‌کاوی و یادگیری ماشین بهبود یافته است. این موضوع از اهمیت زیادی در صنعت بازرگانی، به ویژه در بخش‌هایی که نیاز به تحلیل دقیق و بهینه‌سازی منابع دارند، برخوردار است. در این راستا، پژوهش‌های متعددی به بررسی و توسعه روش‌های مختلف پیش‌بینی سودآوری مشتریان پرداخته‌اند. اصطلاحات و تعاریف زیادی در ادبیات تحقیق در خصوص سودآوری مشتری بیان شده است. از جمله این تعاریف، سودآوری مشتری به مشارکت دلاری خالص که توسط مشتری در حق یک سازمان انجام شده باشد اشاره دارد. یکی از اصول بنیادین در تحلیل سودآوری مشتریان، اصل پارتو یا قانون ۲۰/۸۰ است که بیان می‌کند ۸۰ درصد سود از ۲۰ درصد مشتریان حاصل می‌شود (Vilfredo Pareto, 1906). بر اساس این اصل، شناسایی و تمرکز بر مشتریان سودآور می‌تواند به بهینه‌سازی استراتژی‌های کسب‌وکار کمک کند. پژوهش‌ها نشان داده‌اند که تحلیل‌های دقیق بر روی داده‌های مشتریان می‌تواند منجر به شناسایی دقیق این مشتریان شود (Kumar et al., 2010).

در دنیای واقعی، سلایق و عادات خرید مشتریان ناهمگون هستند و ممکن است در طول زمان نوسان داشته باشد. به همین علت، کسب‌وکارها تمایل دارند تا بدانند، بر اساس سودآوری و خریدهای گذشتهٔ هر مشتری در طول زمان، سودآوری او چگونه رشد خواهد نمود. پاسخ به این سؤال‌ها، تأثیر مستقیم بر تخصیص منابع و استراتژی های بازاریابی یک کسب‌وکار خواهد داشت. به طورکلی این نوع از پیش‌بینی‌ها نیازمند درنظرگرفتن چند مورد عمده است، مانند (Chen et al. 2015) یک معیار برای

اندازه‌گیری سودآوری مشتری با داشتن سابقه خرید مشتری و معیار سودآوری تعریف شده، ساخت یک مدل برای توصیف وضعیت سودآوری مشتری و همچنین، پیش‌بینی سودآوری مشتری معیارها و مدل‌های مختلفی در یک چارچوب کسب‌وکار تعریف شده‌اند و بر اساس دغدغه هر کسب‌وکار، انتخاب یا ترکیب می‌شوند. مشکل اصلی که این نوع پیش‌بینی‌ها با آن روبرو هستند معمولا مرتبط با مدل‌سازی است: با درنظرگرفتن سودآوری مشتری به عنوان یک فرایند متغیر در طول زمان، کدام مدل بهترین انتخاب برای ثبت و پیش‌بینی آن را دارد (Chen et al. 2015).

سودآوری مشتری یکی از فاکتورهای تمیز دادن مشتری با ارزش از مشتری بدون ارزش است. سودآوری مشتری اشاره دارد به سهم دلار خالصی که توسط مشتری به یک سازمان داده می‌شود. در ادبیات چندین اصطلاح به وفاداری مشتری اشاره می‌کند مانند ارزش طول عمر، ارزش طول عمر مشتری، ارزش مشتری، ارزش ارتباط مشتری و سهم مشتری تحلیل سودآوری زمانی بر روی سطح شرکت، محصول یا برند متمرکز بود. سودآوری در سطح مشتری تا زمان دسترسی به پایگاه داده‌های بزرگ مشتری حاوی تاریخچه رفتار خرید مشتری، عملی نشد (Feng et al. 2016).

همچنین، تحقیقات اخیر به بررسی تکنیک‌های داده‌کاوی مانند K-means، درخت تصمیم و KNN در پیش‌بینی سودآوری مشتریان پرداخته‌اند. به‌عنوان مثال، پژوهشگرانی مانند (Smith & Kote, 2002) از روش‌های یادگیری ماشین و الگوریتم‌های خوشه‌بندی برای پیش‌بینی فروش چند کالایی در صنعت خرده‌فروشی استفاده کردند و نشان دادند که این تکنیک‌ها می‌توانند به بهبود عملکرد پیش‌بینی فروش کمک کنند.

آدوماویسیوس و کوون (۲۰۱۴) نیز بر اهمیت پیش‌بینی دقیق فروش کالاها تأکید کرده و بیان کردند که این پیش‌بینی می‌تواند به بهبود کارایی انبار و کاهش کمبود موجودی کمک کند. آن‌ها به این نتیجه رسیدند که تحلیل دقیق داده‌ها و استفاده از الگوریتم‌های یادگیری ماشین می‌تواند به بهینه‌سازی فرآیندهای تجاری کمک کند (Adomavicius & Kwon, 2014).

چن و همکاران (۲۰۱۵)، بر اساس سریهای زمانی RFM اقدام به پیش‌بینی سودآوری مشتری نموده‌اند. تیم آن‌ها با رویکرد سیستم داینامیک، بر اساس رکوردهای تراکنش‌های قبلی مشتریان، سری‌های زمانی بر اساس امتیاز RFM، با استفاده از تحلیل خوشه‌بندی، تولید نموده و ازآن‌ها برای توصیف و اندازه‌گیری سودآوری مشتریان استفاده می‌نمایند. این تحقیق بر رو ی تراکنش‌های واقعی که از یک خرده‌فروشی آنلاین مستقر در انگلیس اخذ شده، انجام شده است. نوع شبکه عصبی مورداستفاده آنها MFNN بوده و نتایج تجربی ایشان، عملکرد خوبی را برای رویکرد پیشنهادی، نشان داده است. همچنین راست و همکاران (۲۰۱۱) نیز دراین خصوص تلاش‌هایی نموده‌اند تا در ادامه تلاش‌های محققان گذشته، با استفاده از رکوردهای خریدهای گذشته مشتری روش‌های بهتری جهت پیش‌بینی سودآوری در طول زمان پیشنهاد نمایند. این تحقیق بر روی داده‌های برگرفته از یک شرکت هایتک در زمینه B2B انجام و مدلی بر روی آن تقریب می‌زند، سپس با استفاده از داده‌های کنار گذاشته شده، صحت مدل را ارزیابی می‌کند. این تحقیق نشان می‌دهد مدلی که بر اساس شبیه‌سازی آینده مشتری است، بهبود زیادی نسبت به برون‌یابی ساده از متوسط سودآوری که در گذشته بدین منظور استفاده شده، تأمین می‌کند و پیش‌بینی می‌کند با استفاده از مدل شبیه‌سازی برای انتخاب مشتری، نرخ بازگشت سرمایه‌گذاری بازاریابی تا ٪۵۸ افزایش یابد.

برخی از کارشناسان خاطرنشان کردند که فروش کالا یکی از شاخص‌هایی است که بنگاه‌ها به آن اهمیت زیادی می‌دهند. این می‌تواند به فروشندگان کمک کند تا استراتژی‌های مناسب برای فروش فعلی را برای به حداکثر رساندن سود تدوین کنند (هاشمی نژادو همکاران، Smith & Kote, 2014 & Hasheminejad et al. 2023). برخی از محققان یک مدل پیش‌بینی بر اساس خوشه‌بندی k-means و الگوریتم رگرسیون یادگیری ماشین برای پیش‌بینی فروش چند کالایی در صنعت خرده‌فروشی پیشنهاد کرده‌اند (Smith and Cote, 2022 and Takahashi & Goto, 2022). علاوه بر این، برخی از کارشناسان معتقدند که پیش‌بینی دقیق فروش کالا می‌تواند کارایی انبار را بهبود بخشد، مصرف مواد خام را کاهش دهد، کمبود موجودی را کاهش دهد و تقاضای بازار را بهتر برآورده کند(Chen et al. 2015). این پیشینه پژوهش به بررسی و تحلیل استفاده از تکنیک‌های

مختلف داده‌کاوی و یادگیری ماشین در پیش‌بینی سودآوری مشتریان می‌پردازد و به پژوهشگران کمک می‌کند تا زمینه‌های مرتبط را به طور جامع و دقیق بررسی کنند.
با پیشرفت سریع تکنولوژی و تغییرات قابل توجه در شیوه‌های زندگی، نیاز به طرح‌های عمرانی، کشاورزی و ساختمانی مدرن و پیشرفته به‌طور فزاینده‌ای افزایش یافته است. این پروژه‌ها نیاز به لوازم و ابزارهای لوله و اتصالات با کیفیت و مدرن را به‌دنبال دارد. تأمین این نیازها برای شرکت‌های ساخت و ساز اهمیت ویژه‌ای پیدا کرده است. شرکت‌های بازرگانی در این زمینه به‌طور فعال اقدام به تأمین و توزیع انواع مختلف لوله‌ها و اتصالات کرده‌اند، از جمله لوله‌های ساختمانی و کشاورزی، اتصالات و ابزارآلات مختلف، که همگی از تولیدکنندگان معتبر و شناخته‌شده تأمین می‌شود.

دسته‌بندی مشتریان و تجزیه و تحلیل رفتار آن‌ها در صنعت بازرگانی و فروش اهمیت ویژه‌ای دارد، زیرا این اطلاعات به صاحبان صنایع کمک می‌کند تا استراتژی‌های فروش و بازاریابی خود را بهینه‌سازی کنند. در این راستا، استفاده از تکنیک‌های داده‌کاوی برای دسته‌بندی مشتریان بر اساس مدل RFML (Recency، Frequency، Monetary و Length) امری ضروری است. این مدل به تحلیل رفتار مشتریان از جنبه‌های مختلفی مانند زمان خرید، تعداد خرید، مبلغ خرید و طول ارتباط مشتری می‌پردازد.

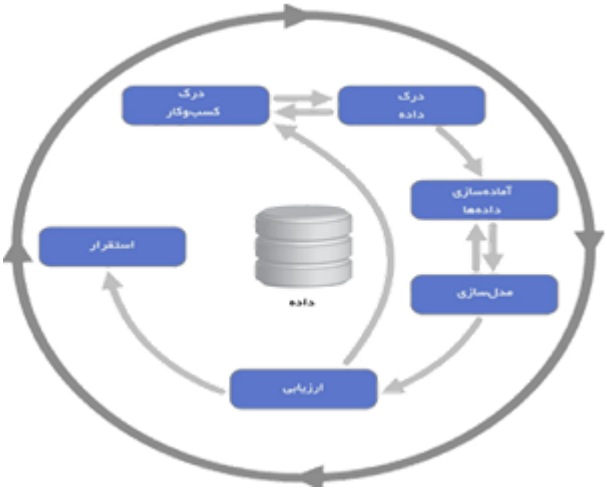
هدف این پژوهش خوشه‌بندی مشتریان و پیش‌بینی سودآوری آن‌ها برای یک شرکت بازرگانی فعال در زمینه لوله و اتصالات است. از آنجایی که در داده‌های موجود ستون مشخصی برای سودآوری مشتریان وجود ندارد، نیاز به استفاده از الگوریتم‌های یادگیری نظارت شده از جمله KNN، Decision Tree و Naive Bayes برای پیش‌بینی سودآوری مشتریان احساس می‌شود.
اهداف اصلی پژوهش به شرح ذیل است:

- خوشه‌بندی مشتریان شرکت بازرگانی: شناسایی و طبقه‌بندی مشتریان به گروه‌های مشابه بر اساس رفتار خرید و ویژگی‌های مختلف.
- دسته‌بندی و پیش‌بینی سودآوری مشتریان: استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی میزان سودآوری هر مشتری و بهینه‌سازی استراتژی‌های فروش بر اساس این پیش‌بینی‌ها.

روش تحقیق

روش کریسپ[1] (CRISP) فرایندهای استاندارد صنعت متقابل برای داده‌کاوی است. در واقع روش‌های تحلیل متفاوتی برای اجرای پروژه‌های داده‌کاوی وجود دارد. این فرآیند دارای شش مرحله اصلی است، این شش مرحله از درک نیازهای اصلی کسب و کار شروع می‌شود و در نهایت به ارائه راه‌کاری برای آن ختم می‌شود. به‌نظر می‌رسد که این مراحل به دنبال یکدیگر انجام می‌شوند اما در عمل رفت و برگشت‌های زیادی بین مراحل وجود دارند. الگوریتم کریسپ را در شکل ۱ قابل مشاهده است.

[1] Cross Industry Standard Process for Data Mining



شکل (۱): مدل فرآیندهای استاندارد صنعت متقابل برای داده‌کاوی

روش کریسپ در قالب شش مرحله انجام می‌شود که به شرح ذیل است:

فهم تجاری[۲]: این مرحله شامل گردآوری الزامات و مصاحبه با مدیران ارشد و خبرگان برای تعیین اهدافی بالاتر از کار با داده‌ها می‌شود.

درک داده‌ها[۳]: مرحله درک داده شامل نگاه نزدیک‌تر به دردسترس بودن داده برای داده‌کاوی است. این مرحله شامل گردآوری داده‌های اولیه، توصیف داده، کشف داده و تغییر کیفیت داده می‌شود.

آماده‌سازی داده‌ها[۴]: آماده‌سازی داده یکی از مهم‌ترین و اغلب زمان‌برترین جوانب پروژه‌های داده‌کاوی است و شامل انتخاب داده، پاک‌سازی داده، ساختاربندی داده جدید و ادغام داده است.

مدل‌سازی[۵]: در این مرحله برای آموزش مدل، از داده‌های پردازش شده استفاده می‌شود. بسته به اینکه مساله شما چه نوع مساله‌ای است، باید از تکنیک‌ها و روش‌های مدل‌سازی متناسب با اهداف کسب‌وکار خود استفاده کنید.

ارزیابی[۶]: در این مرحله، ارزیابی نتایج، فرایند بازبینی و تعیین مراحل بعدی انجام شده است.

توسعه[۷]: نتایج به دست آمده توسعه یافته و برای بهبود عملکرد سازمان به کار گرفته می‌شوند.

یک متخصص علم داده باید کسب‌وکاری را که قرار است بر روی آن پروژه داده‌کاوی انجام دهد را به خوبی درک کند. فهم کامل و درست کسب‌وکار، ذهن او را برای کار بر روی پروژه آماده کرده و به او این امکان را می‌دهد تا با شناخت بیشتر و کامل‌تری وارد مراحل بعدی شود. در مرحله فهم کسب‌وکار باید مواردی مانند ابعاد مختلف آن کسب‌وکار، محدودیت‌ها، شرایط موجود، اهداف کسب‌وکار و سایر عوامل مرتبط را بررسی کنید. یک متخصص در این مرحله می‌تواند با افزایش دانش و شناخت خود، تا حدودی به کسب‌وکار موجود مسلط شود.

مجموعه داده مورد استفاده دارای ۱۰ ویژگی[۸] است که در ادامه توضیحات لازم مربوط به هرکدام از ویژگی‌ها در جدول ۱ ارائه شده است.

---

[2] Business Understanding
[3] Data Understanding
[4] Data Preparation
[5] Modeling
[6] Evaluation
[7] Deployment
[8] Feature

جدول (۱): ویژگی‌های مجموعه داده شرکت بازرگانی مورد مطالعه

ردیف	ویژگی	توضیح	واحد
۱.	کد کالا	کد کالایی که توسط شرکت بازرگانی به فروش می‌رسد.	عددی
۲.	نوع کالا	عنوان کالایی که به فروش می‌رسد.	غیرعددی
۳.	تاریخ سفارش	تاریخ سفارش کالا توسط مشتری به صورت روز، ماه و سال قابل نمایش است.	عددی
۴.	کد مشتری	نمایش هر مشتری با کد مخصوص به خود مشتری	عددی
۵.	نام مشتری	اسم هر مشتری خریدار کالا	غیرعددی
۶.	مقدار سفارش	مقداری که مشتری در هر مراجعه خریداری می‌کند.	عددی
۷.	Recency	تعداد روزهایی که از آخرین خرید می‌گذرد.	عددی
۸.	Frequency	تعداد دفعات مراجعه به شرکت بازرگانی برای خرید.	عددی
۹.	Monetary	میزان سودآوری و خرید مشتری برای شرکت بازرگانی	عددی
۱۰.	Length	تعداد روزهای که از اولین مراجعه مشتری گذشته است.	عددی

این مجموعه داده ۱۰۵۱ رکورد و ۱۰ ویژگی دارد، که در هیچ کدام از ویژگی‌ها مقدار داده از دست رفته[9] و داده تکراری[10] وجود ندارد. داده‌های پرت[11] با دو روش نمودار جعبه‌ای[12] و Z-Score شناسایی شد و این داده‌های پرت حذف شد. در این مجموعه داده ویژگی‌های نوع کالا و نام مشتری به دلیل اینکه در نتایج نهایی تاثیری ایجاد نمی‌کنند، بنابراین قابلیت حذف شدن دارند و از مجموعه داده حذف می‌شود و با حذف این دو ویژگی ابعاد مسئله کاهش پیدا می‌کند و محاسبات سریع‌تر انجام می‌شود. در جدول ۲ مجموعه داده شرکت بازرگانی قابل مشاهده است.

جدول (۲): مجموعه داده شرکت بازرگانی

	کد کالا	نوع کالا	سال مالی	ماه	روز	کد مشتری	نام مشتری	مقدار	R	F	M	L
0	333000100	نیوفلکس دوسر 110 * 2000	1400	1	11	958	مقدسی(لوله گستر)	1.0	NaN	NaN	NaN	NaN
1	333000101	نیوفلکس دوسر 110 * 1000	1400	1	11	958	مقدسی(لوله گستر)	2.0	NaN	NaN	NaN	NaN
2	333000102	نیوفلکس یکسر 110 * 500	1400	1	11	958	مقدسی(لوله گستر)	2.0	NaN	NaN	NaN	NaN
3	333000103	نیوفلکس سیفون یک تکه با درپوش 63	1400	1	11	958	مقدسی(لوله گستر)	1.0	NaN	NaN	NaN	NaN
4	333000104	نیوفلکس زانو 110	1400	1	11	958	مقدسی(لوله گستر)	2.0	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
1045	333000127	نیوپایپ لوله (PEX-AL-PEX) 16	1400	2	26	1255	آریس	600.0	NaN	NaN	NaN	NaN
1046	333000739	دانو لوله UPVC 4.7*63	1400	2	26	1256	مولایی	1.0	NaN	NaN	NaN	NaN
1047	333000740	لوله 32 هیدرویول(16بار)upvc	1400	2	26	1256	مولایی	1.0	NaN	NaN	NaN	NaN
1048	333000693	دانو لوله UPVC 3.7*50	1400	2	26	1256	مولایی	8.0	NaN	NaN	NaN	NaN
1049	333000741	دانو زانو UPVC 90*63	1400	2	26	1256	مولایی	2.0	NaN	NaN	NaN	NaN

خصیصه‌های مدل RFML که در مجموعه داده استفاده شده است با توجه به مفهوم هرکدام از خصیصه‌ها، محاسبه می‌شود. هرکدام از خصیصه‌هایی که به زمان بستگی دارد، زمان مبدا کامپیوتری که استفاده شده است را در نظر می‌گیرد. باتوجه به اینکه در مجموعه داده استفاده شده به علت محرمانه بودن میزان قیمت هر یک از کالاها، از قیمت کالاها استفاده نشده است و

[9] Missing value
[10] Duplicated
[11] Outlier
[12] Box Plot

در محاسبات قیمت کالاها را یکسان و برابر هم فرض شده است. در ادامه به علت یکسان نبودن ابعاد داده‌ها در مدل RFML، هریک از خصیصه‌های مدل نرمال‌سازی می‌شود و در جدول ۳ قابل مشاهده است.

جدول (۳): خصیصه‌های مدل برای نرمال‌سازی

ردیف	خصیصه	نماد خصیصه	نماد نرمال‌سازی خصیصه
.۱	Recency	R	R1
.۲	Frequency	F	F1
.۳	Monetary	M	M1
.٤	Length	L	L1

ویژگی‌هایی (نوع کالا و نام مشتری) که در نتایج نهایی تاثیری نداشته حذف می‌شود و نرمال‌سازی خصیصه‌ها محاسبه شده است که در جدول ۴ قابل نمایش است.

جدول (۴): نرمال‌سازی خصیصه‌ها و حذف ویژگی‌ها بدون تاثیر

	کد کالا	سال مالی	ماه	روز	کد مشتری	مقدار	R	F	M	L	R1	F1	M1	L1
0	0	1400	1	14	0	120.0	511.0	2.0	180.0	511.0	0.915851	0.111111	0.033028	0.915851
1	69	1400	2	4	1	120.0	490.0	13.0	2015.0	509.0	0.955102	0.722222	0.369725	0.919450
2	71	1400	2	7	2	72.0	487.0	2.0	172.0	511.0	0.960986	0.111111	0.031560	0.915851
3	3	1400	1	18	3	152.0	507.0	1.0	152.0	507.0	0.923077	0.055556	0.027890	0.923077
4	63	1400	2	1	4	90.0	493.0	5.0	482.0	507.0	0.949290	0.277778	0.088440	0.923077
5	59	1400	1	31	5	100.0	494.0	5.0	750.0	507.0	0.947368	0.277778	0.137615	0.923077
6	64	1400	2	1	6	80.0	493.0	4.0	388.5	503.0	0.949290	0.222222	0.071284	0.930417
7	17	1400	1	22	7	220.0	503.0	13.0	3255.0	503.0	0.930417	0.722222	0.597248	0.930417
8	30	1400	1	23	8	130.0	502.0	4.0	500.0	502.0	0.932271	0.222222	0.091743	0.932271
9	75	1400	2	22	9	60.0	472.0	6.0	732.0	504.0	0.991525	0.333333	0.134312	0.928571
10	38	1400	1	24	10	60.0	501.0	1.0	60.0	501.0	0.934132	0.055556	0.011009	0.934132
11	0	1400	2	7	11	800.0	487.0	18.0	5450.0	501.0	0.960986	1.000000	1.000000	0.934132
12	51	1400	1	25	12	80.0	500.0	3.0	430.0	500.0	0.936000	0.166667	0.078899	0.936000
13	3	1400	1	25	13	80.0	500.0	3.0	324.0	514.0	0.936000	0.166667	0.059450	0.910506
14	0	1400	1	26	14	120.0	499.0	1.0	120.0	499.0	0.937876	0.055556	0.022018	0.937876
15	54	1400	1	28	15	125.0	497.0	3.0	315.0	497.0	0.941650	0.166667	0.057798	0.941650
16	60	1400	1	31	16	74.0	494.0	4.0	357.0	496.0	0.947368	0.222222	0.065505	0.943548
17	65	1400	2	6	17	90.0	488.0	2.0	150.0	493.0	0.959016	0.111111	0.027523	0.949290
18	73	1400	2	12	18	70.0	482.0	1.0	70.0	482.0	0.970954	0.055556	0.012844	0.970954
19	76	1400	2	25	19	80.0	469.0	9.0	1510.0	469.0	0.997868	0.500000	0.277064	0.997868

یافته‌ها

مدل‌سازی در ۲ مرحله انجام شده است، در مرحله اول خوشه‌بندی[۱۳] انجام می‌شود که از روش K-means استفاده شده است و در مرحله دوم دسته‌بندی[۱۴] و پیش‌بینی[۱۵] انجام می‌شود که از روش‌های Decision Tree، KNN و Gaussian استفاده شده است. ایده اصلی روش‌های خوشه‌بندی مبتنی بر تقسیم مانند k-means به دست آوردن تعداد خوشه‌ها به نحوی است که مجموع فواصل درون خوشه‌ای داده‌ها (یا مجموع مربعات فواصل درون خوشه‌ای) حداقل شود. مجموع فواصل درون خوشه‌ای داده‌ها، میزان فشردگی خوشه‌بندی انجام شده را نشان می‌دهد و هدف حداقل‌سازی فواصل درون خوشه‌ای است. روش آرنج[۱۶] مجموع فواصل درون خوشه‌ای داده‌ها را به عنوان تابعی از تعداد خوشه‌ها در نظر می‌گیرد. به این ترتیب تعداد خوشه‌ها به نحوی انتخاب می‌شوند که افزودن یک خوشه دیگر، بهبودی در حداقل‌سازی WSS ایجاد نکند. تعداد بهینه خوشه‌ها طبق الگوریتم زیر به‌دست می‌آید:

۱) اجرای الگوریتم خوشه‌بندی مانند k-means برای مقادیر متفاوت k (به‌طور مثال با در نظرگرفتن مقدار k در بازه ۱ تا ۱۰)
۲) محاسبه مقدار WSS برای هر مقدار k
۳) رسم مقدار WSS برحسب مقادیر مختلف k
۴) نقطه زانوئی نمودار رسم شده، تعداد بهینه خوشه‌ها را نشان می‌دهد.

در مرحله اول مدل‌سازی با استفاده از روش K-means خوشه‌بندی انجام می‌شود. در این مرحله با استفاده از نظر خبره برای تعداد ۱ الی ۲۰ خوشه اجرا می‌شود و با استفاده از روش آرنج تعداد بهینه خوشه‌ها بدست می‌آید. تشخیص نقطه آرنج در منحنی همیشه ساده نیست و ممکن است اشتباه تشخیص داده شود بنابراین با استفاده از کتابخانه kneed و استفاده از دستور ذیل در شکل ۲ نقطه آرنج تعیین می‌شود.

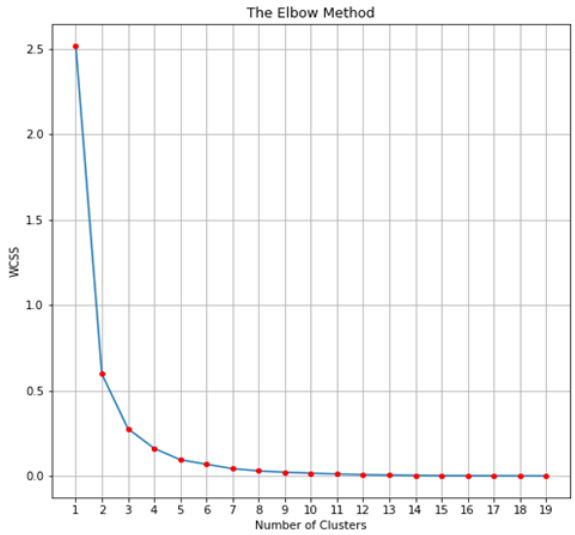
```
kl = KneeLocator(range(1, 20), wcss, curve="convex", direction="decreasing")

kl.elbow
```

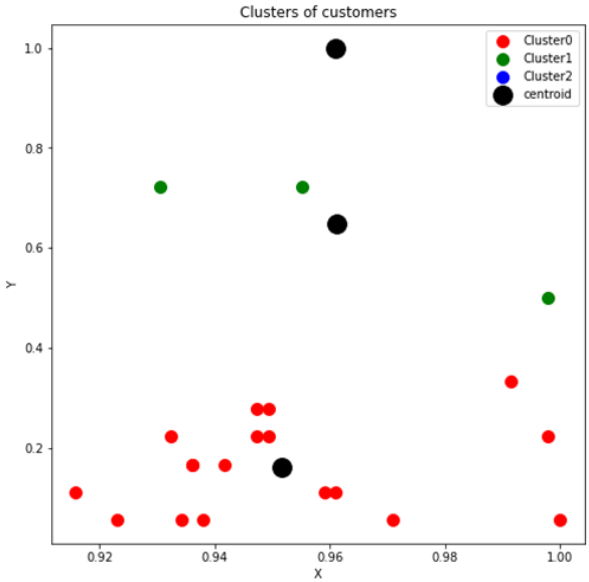
شکل (۲): کد پایتون تعیین نقطه آرنج

نقطه ۳ به عنوان نقطه آرنج انتخاب شد و از بین نقاط ۱ و ۲ و ۳ نقطه ۳ کمترین میزان wcss[17] را دارد و تعداد خوشه‌های بهینه برابر با ۳ انتخاب می‌شود، در ادامه عملیات خوشه‌بندی را با استفاده از خصیصه‌های نرمال شده مدل RFML انجام می‌شود. بنابراین در شکل ۳ با استفاده از روش Elbow تعداد بهینه خوشه‌ها را نشان می‌دهد و در شکل ۴ پراکندگی خوشه‌ها مشهود است.

[13] Clustering
[14] Classification
[15] Prediction
[16] Elbow
[17] Within Cluster Sum of Squares



شکل (۳): نمودار روش Elbow



شکل (۴): پراکندگی خوشه‌ها

در بهترین نوع خوشه‌بندی داده‌های داخل یک خوشه کمترین میزان فاصله و بیشترین میزان شباهت را دارد و داده‌های بین دو خوشه بیشترین میزان فاصله و کمترین میزان تشابه را دارد، بنابراین در شکل ۴ بهترین خوشه با رنگ قرمز نشان داده شده است. در این مرحله، دسته‌بندی با استفاده از ۳ روش انجام شده است.

درخت تصمیم[18] (DT)، روشی در یادگیری ماشین برای ساختاربندی (یا شکل‌دهی یا سازماندهی) به الگوریتم است. یک الگوریتم درخت تصمیم برای تقسیم ویژگی‌های «مجموعه داده[۱۹]» از طریق «تابع هزینه[۲۰]» مورد استفاده قرار می‌گیرد. این الگوریتم قبل از انجام بهینه‌سازی و حذف شاخه‌های اضافه، به گونه‌ای رشد می‌کند که دارای ویژگی‌های نامرتبط با مسئله است؛ به همین دلیل، عملیات «هرس کردن[۲۱]» برای حذف این شاخه‌های اضافه در آن انجام می‌شود. در الگوریتم درخت تصمیم،

[18] Decision Tree
[19] Data Set
[20] Cost Function
[21] Pruning

پارامترهایی ازجمله عمق درخت تصمیم را نیز می‌توان تنظیم کرد تا از «بیش‌برازش[۲۲]» یا «پیچیدگی بیش از حد درخت[۲۳]» تا جای امکان جلوگیری شود.

انواع بسیاری از درخت‌های تصمیم در یادگیری ماشین برای مسئله‌های دسته‌بندی اشیا براساس ویژگی‌های آموزش داده شده، استفاده می‌شوند. همچنین این روش می‌تواند در مسائل رگرسیون[۲۴] یا روشی برای پیش‌بینی نتایج پیوسته داده‌های دیده نشده، مورد استفاده قرار بگیرد. مزیت اصلی استفاده از یادگیری ماشین در پروژه‌ها، سادگی آن است؛ زیرا با استفاده از آن، فرآیند تصمیم‌گیری به راحتی قابل تجسم و درک خواهد شد. با این حال، در مسائل یادگیری ماشین با افزایش تعداد شاخه‌های درخت تصمیم، ممکن است درک و استفاده از آن به دلیل پیچیدگی بیش از حد درخت، چالش برانگیز شود؛ بنابراین، هرس کردن درخت در چنین شرایطی بسیار ضروری به‌نظر می‌رسد.

به یک ستون جهت پیش‌بینی برای سودآور بودن مشتری، میزان سودآوری و اطلاعات کیفی دیگر نیاز داریم بنابراین در مجموعه داده یک ستون تحت عنوان S ایجاد می‌شود. طبق توضیحات قبلی با استفاده از حاصل جمع خصیصه‌های مدل RFML که عددی بین ۰ تا ۴ است، می‌توان میزان سودآور بودن مشتری را تعیین کرد. خروجی مدل RFML در جدول ۵ قابل مشاهده است.

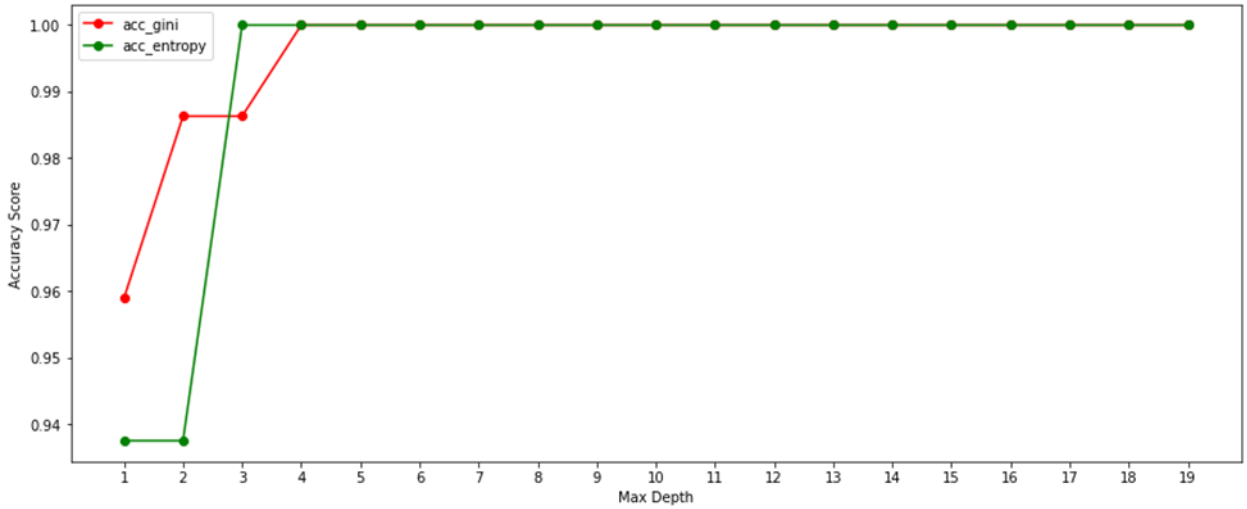
- ۳ تا ۴ مشتری عالی
- ۲ تا ۳ مشتری خوب
- ۱ تا ۲ مشتری متوسط
- ۰ تا ۱ مشتری ضعیف

جدول (۵): خروجی مدل RFML

	کد کالا	سال مالی	ماه	روز	کد مشتری	مقدار	R	F	M	L	R1	F1	M1	L1	R+F+M+L	S
0	0	1400	1	14	0	120.0	511.0	2.0	180.0	511.0	0.915851	0.111111	0.033028	0.915851	1.975841	متوسط
1	69	1400	2	4	1	120.0	490.0	13.0	2015.0	509.0	0.955102	0.722222	0.369725	0.919450	2.966499	خوب
2	71	1400	2	7	2	72.0	487.0	2.0	172.0	511.0	0.960986	0.111111	0.031560	0.915851	2.019508	خوب
3	3	1400	1	18	3	152.0	507.0	1.0	152.0	507.0	0.923077	0.055556	0.027890	0.923077	1.929599	متوسط
4	63	1400	2	1	4	90.0	493.0	5.0	482.0	507.0	0.949290	0.277778	0.088440	0.923077	2.238585	خوب
5	59	1400	1	31	5	100.0	494.0	5.0	750.0	507.0	0.947368	0.277778	0.137615	0.923077	2.285838	خوب
6	64	1400	2	1	6	80.0	493.0	4.0	388.5	503.0	0.949290	0.222222	0.071284	0.930417	2.173214	خوب
7	17	1400	1	22	7	220.0	503.0	13.0	3255.0	503.0	0.930417	0.722222	0.597248	0.930417	3.180305	عالی
8	30	1400	1	23	8	130.0	502.0	4.0	500.0	502.0	0.932271	0.222222	0.091743	0.932271	2.178507	خوب
9	75	1400	2	22	9	60.0	472.0	6.0	732.0	504.0	0.991525	0.333333	0.134312	0.928571	2.387742	خوب
10	38	1400	1	24	10	60.0	501.0	1.0	60.0	501.0	0.934132	0.055556	0.011009	0.934132	1.934828	متوسط
11	0	1400	2	7	11	800.0	487.0	18.0	5450.0	501.0	0.960986	1.000000	1.000000	0.934132	3.895117	عالی
12	51	1400	1	25	12	80.0	500.0	3.0	430.0	500.0	0.936000	0.166667	0.078899	0.936000	2.117566	خوب
13	3	1400	1	25	13	80.0	500.0	3.0	324.0	514.0	0.936000	0.166667	0.059450	0.910506	2.072622	خوب
14	0	1400	1	26	14	120.0	499.0	1.0	120.0	499.0	0.937876	0.055556	0.022018	0.937876	1.953325	متوسط
15	54	1400	1	28	15	125.0	497.0	3.0	315.0	497.0	0.941650	0.166667	0.057798	0.941650	2.107765	خوب
16	60	1400	1	31	16	74.0	494.0	4.0	357.0	496.0	0.947368	0.222222	0.065505	0.943548	2.178644	خوب
17	65	1400	2	6	17	90.0	488.0	2.0	150.0	493.0	0.959016	0.111111	0.027523	0.949290	2.046941	خوب
18	73	1400	2	12	18	70.0	482.0	1.0	70.0	482.0	0.970954	0.055556	0.012844	0.970954	2.010308	خوب
19	76	1400	2	25	19	80.0	469.0	9.0	1510.0	469.0	0.997868	0.500000	0.277064	0.997868	2.772800	خوب

[22] Overfitting
[23] Overly Complex Tree
[24] Regression

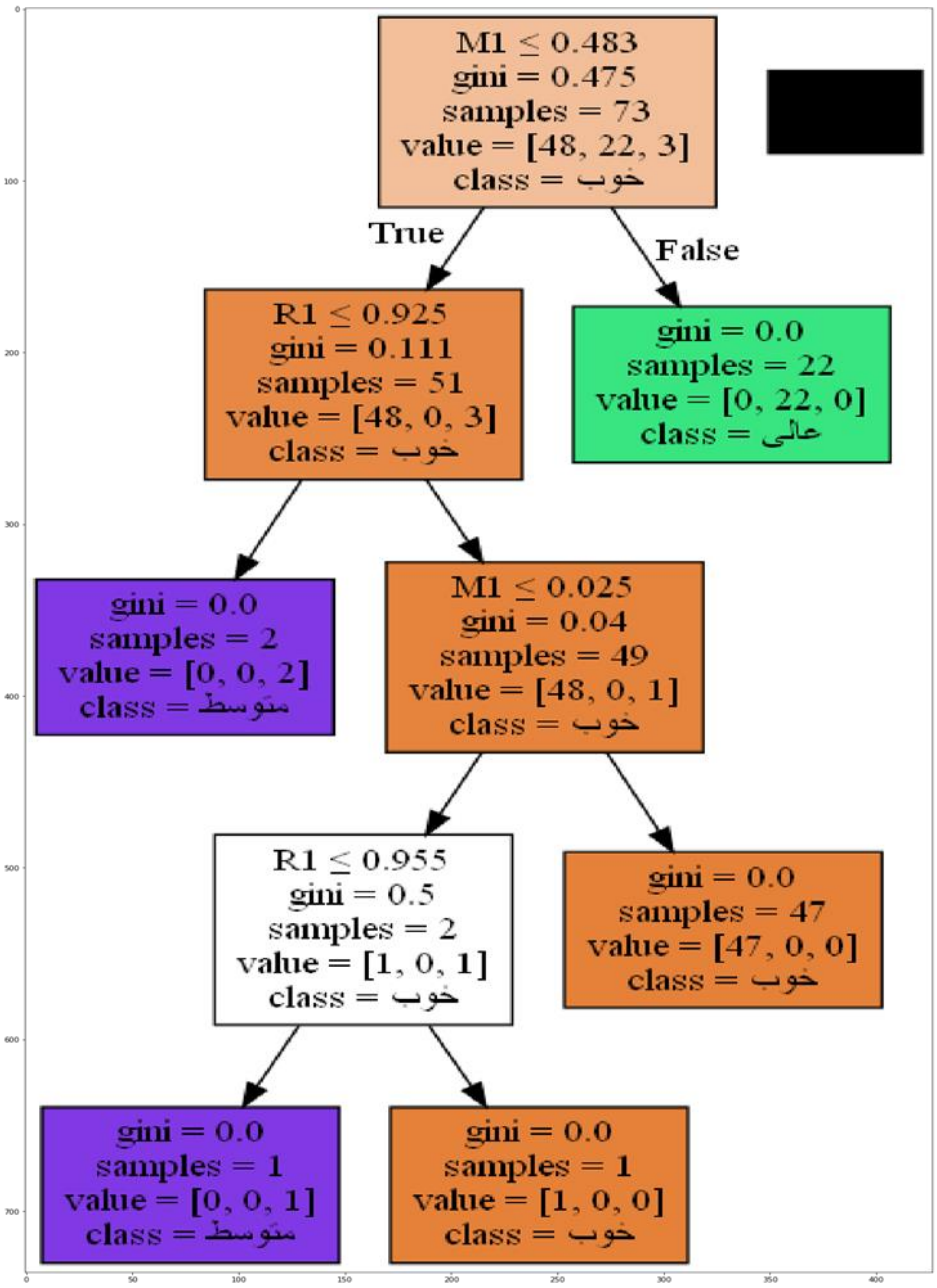
در این مدل‌سازی از شاخص‌های Gini و Entropy برای بدست آوردن یک درخت مطلوب استفاده شده است. هرچقدر مقدار این دو شاخص کمتر باشد بهتر است و بهترین حالت مقدار صفر است، در این حالت بهترین برگ انتخاب می‌شود. باتوجه به شکل ۵ مشخص است از سطح[۲۵] ۴ به بعد مقدار شاخص‌های Gini و Entropy برابر است. بنابراین در این مسئله هر دو شاخص خوب هست و در ادامه به‌صورت انتخابی در کد نویسی از شاخص Gini استفاده شده است.



شکل (۵): تعیین شاخص Gini و Entropy

در ادامه مدل‌سازی درخت تصمیم برحسب شاخص Gini در ۴ سطح رسم شده است. هر برگی که مقدار Gini برابر با صفر داست به عنوان بهترین برگ انتخاب می‌شود. با توجه به شکل ۶، ۵ شاخه وجود دارد که در انتهای هر شاخه شاخص Gini برابر با صفراست. بنابراین به دنبال بهترین شاخه هستیم. بهترین شاخه، شاخه‌ای است که بیشترین تاثیر را بر نتایج داشته باشد بنابراین با توجه نتایج در کد نویسی که در شکل ۶ قابل مشاهده است، از بین ویژگی‌های استاندارد شده R1، F1، M1 و L1 ویژگی M1 به عنوان ویژگی مهم و تاثیرگذار انتخاب می‌شود. بنابراین شاخه‌ای که به برگ سبز رنگ منتهی می‌شود به عنوان شاخه بهینه انتخاب می‌شود. نتایج مربوط به درخت تصمیم در شکل ۷ نشان داده شده است. جهت اطمینان از عملکرد مدل پیش‌بینی از ۲ روش دیگر در ادامه استفاده شده است.

[25] Max depth



شکل (۷): درخت تصمیم

روش K نزدیک‌ترین همسایه[۲۶] (KNN) یک روش یادگیری موردی است و از جمله ساده‌ترین الگوریتم‌های یادگیری ماشین است که به روش K همسایه نزدیک نیز معروف است. در این الگوریتم یک نمونه با رای اکثریت از همسایه‌ها دسته‌بندی می‌شود و این نمونه در عمومی‌ترین کلاس بین k همسایه نزدیک تعیین می‌شود. K یک مقدار مثبت صحیح و عموماً کوچک است. اگر k=1باشد نمونه به سادگی در کلاس همسایگان نزدیک به خود تعیین می‌شود. فرد بودن مقدار k مفید است چون با این کار جلوی آراء برابر گرفته می‌شود. روش k همسایه نزدیک، برای بسیاری از روش‌ها کاربرد دارد، زیرا اثربخش، غیرپارامتریک و دارای پیاده‌سازی راحت است. با این حال زمان دسته‌بندی آن طولانی است و یافتن مقدار k بهینه مشکل است. بهترین انتخاب

[26] K-Nearest Neighbors

از k، وابسته به داده‌ها است به طور کلی مقدار بزرگ از k اثر نویز روی دسته‌بندی را کاهش می‌دهد، اما مرز بین کلاس‌ها کمتر متمایز می‌شود.

دسته‌بندی‌کننده بیز ساده[۲۷] در یادگیری ماشین به گروهی از دسته‌بندی‌کنندههای ساده بر پایه احتمالات گفته می‌شود که با فرض استقلال متغیرهای تصادفی و براساس قضیه بیز ساخته می‌شوند. به‌طور ساده روش بیز روشی برای دسته‌بندی پدیده‌ها، بر پایه احتمال وقوع یا عدم وقوع یک پدیده‌است.

جهت انجام پیش‌بینی نیاز است داده‌ها به دو دسته آموزش[۲۸] و تست[۲۹] تقسیم شود. مقدار اندازه‌گیری تست برای سه روش مورد استفاده KNN، DT و Naive Bayes برابر ۳۰ درصد فرض شده است. مجموعه داده مورد استفاده به نسبت داده‌های تمیزی بودند، بنابراین مقدار test accuracy و train accuracy برای هر سه روش مورد استفاده مقداری قابل قبول بود و نتایج در شکل ۸ قابل مشاهده است.

train accuracy :

Score	Model
100.00	Naive Bayes
100.00	Decision Tree
98.63	KNN

test accuracy :

Score	Model
0.96875	KNN
0.96875	Naive Bayes
0.96875	Decision Tree

شکل (۸): نتایج داده‌های آموزش و تست

در مرحله ارزیابی مدل برای بررسی دقت مدل از شاخص Accuracy استفاده شده است که نتایج در جدول ۶ قابل مشاهده است.

جدول (۶): نتایج ارزیابی

	Accuracy
Decision Tree	0.9872
KNN	0.9785
Naive Bayes	0.9667

باتوجه به نتایج حاصل شده، می‌توان پیشنهاد کرد در صورت تغییر نکردن تعداد مشتریان، بهتر است مشتریان در ۳ خوشه قرار بگیرند و اگر محدودیت‌های دسترسی به داده‌های محرمانه برداشته شود تعداد خوشه‌های بهینه تغییر می‌کند. با توجه به پیش-بینی‌های که انجام گرفته برای سودآور بودن مشتریان در درخت تصمیم مشخص است از ۷۳ مشتری مراجعه کننده به شرکت بازرگانی تعداد ۴۸ مشتری سودآور است بنابراین تعداد بیشتر مشتری‌ها در این کلاس قرار می‌گیرد، پیشنهاد می‌شود شرکت بازرگانی به مشتریانی که در کلاس سودآور قرار می‌گیرد تخفیف یا امتیازهایی داده شود که شرکت بتواند این مشتریان را حفظ کند و با ارائه برنامه‌های مشتریانی که در کلاس‌های دیگر قرار می‌گیرد را ترغیب کند که از این شرکت خریداری کنند.

بحث و نتیجه‌گیری

نتایج این پژوهش تأثیر قابل توجهی از استفاده از تکنیک‌های داده‌کاوی و الگوریتم‌های یادگیری ماشین برای خوشه‌بندی مشتریان و پیش‌بینی سودآوری در شرکت‌های بازرگانی فروش لوله و اتصالات را نشان می‌دهد. تحلیل داده‌ها با استفاده از روش

[27] Naive Bayes
[28] train
[29] test

K-means و تکنیک Elbow برای تعیین تعداد بهینه خوشه‌ها، منجر به شناسایی سه خوشه اصلی شد که ویژگی‌های متمایز هر کدام به وضوح مشخص شد. در مرحله پیش‌بینی سودآوری، الگوریتم درخت تصمیم با دقت ۹۸.۷۲٪ بهترین عملکرد را نشان داد، که نشان‌دهنده قدرت و دقت بالای این روش در دسته‌بندی و پیش‌بینی سودآوری مشتریان است. از این نتایج می‌توان نتیجه‌گیری کرد که استفاده از الگوریتم‌های داده‌کاوی و یادگیری ماشین می‌تواند به بهبود فرآیندهای تجزیه و تحلیل مشتریان و بهینه‌سازی استراتژی‌های فروش کمک کند. به‌ویژه، شناسایی و تمرکز بر مشتریان سودآور و ارائه امتیازات ویژه به این گروه می‌تواند منجر به افزایش وفاداری مشتریان و بهبود نتایج مالی شرکت شود.

به عنوان پیشنهادهای آتی می‌توان در شرایطی که داده‌های محرمانه و کامل‌تری در دسترس باشد، بررسی تعداد بیشتر خوشه‌ها و تحلیل دقیق‌تر هر خوشه می‌تواند به شناسایی جزئیات بیشتری درباره رفتار مشتریان و نیازهای آنان کمک کند. اضافه کردن ویژگی‌های بیشتر به مجموعه داده‌ها، مانند جزئیات مالی و رفتار خرید دقیق‌تر، می‌تواند به بهبود دقت مدل‌های پیش‌بینی و خوشه‌بندی کمک کند. با پیشرفت‌های مداوم در الگوریتم‌های یادگیری ماشین، ارزیابی مدل‌های جدید و مقایسه آن‌ها با روش‌های فعلی می‌تواند به شناسایی بهترین الگوریتم برای پیش‌بینی سودآوری مشتریان کمک کند. با توجه به این پیشنهادات، شرکت می‌تواند به بهینه‌سازی فرآیندهای فروش و افزایش سودآوری خود ادامه دهد و به این ترتیب، موقعیت خود را در بازار رقابتی تقویت کند.

منابع

Hasheminejad, S.A., Shabaab, M. & Javadinarab, N. Developing Cluster-Based Adaptive Network Fuzzy Inference System Tuned by Particle Swarm Optimization to Forecast Annual Automotive Sales: A Case Study in Iran Market. *Int. J. Fuzzy Syst.* **24**, 2719–2728 (2023). https://doi.org/10.1007/s40815-022-01263-6

Adomavicius G, Kwon Y (2014) Optimization-based approaches for maximizing aggregate recommendation diversity. *INFORMS J. Comput.* 26(2):351–369.

Smith, M. A., & Côté, M. J. (2022). Predictive analytics improves sales forecasts for a pop-up retailer. *INFORMS Journal on Applied Analytics*, *52*(4), 379-389.

Takahashi, K., & Goto, Y. (2022). Embedding-Based Potential Sales Forecasting of Bread Product. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *26*(2), 236-246.

Schmidt A, Kabir MWU, Hoque MT. Machine Learning Based Restaurant Sales Forecasting. *Machine Learning and Knowledge Extraction*. 2022; 4(1):105-130. https://doi.org/10.3390/make4010006

Chen, D., Guo, K., & Ubakanma, G. (2015). Predicting customer profitability over time based on RFM time series. International Journal of Business Forecasting and Marketing Intelligence, 2(1), 1. https://doi.org/10.1504/ijbfmi.2015.075325

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers and Industrial Engineering, 101, 554–564. https://doi.org/10.1016/j.cie.2016.09.011

Rust, R. T., Kumar, V., & Venkatesan, R. (2011). Will the frog change into a prince? Predicting future customer profitability. International Journal of Research in Marketing, 28(4), 281–294. https://doi.org/10.1016/j.ijresmar.2011.05.003

Pareto, V. (1906). *Manual of Political Economy*. New York: A.M. Kelley.

Kumar, V., Rajan, B., & Shankar, S. (2010). *Customer Lifetime Value: Imperatives for Marketing*. Journal of Marketing, 74(3), 130-147.

Chen, J., Wang, X., & Liu, Y. (2015). *Customer Profitability Analysis and Prediction Using Dynamic Systems*. Journal of Business Research, 68(9), 1915-1922.

Rast, A., Lee, H., & Park, J. (2011). *Predicting Customer Profitability with Simulation Models*. IEEE Transactions on Knowledge and Data Engineering, 23(4), 545-558.

Smith, A., & Kote, S. (2022). *Predicting Multi-Item Sales in Retail Using Data Mining Techniques*. Data Mining and Knowledge Discovery, 36(1), 155-170.

Adomavicius, G., & Kwon, K. (2014). *Enhancing Retail Inventory Management with Predictive Analytics*. Information Systems Research, 25(2), 297-312.

Analyzing and predicting the profitability of customers in the pipe and fittings business industry using data mining and machine learning algorithms

Hamed Naderi

Ph. D. student of Industrial Engineering, Department of Industrial Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.

Elham Akhondzadeh

Assistant Prof, Department of Information Technology Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.

Mohammad Ali Rastegar

Assistant Prof, Department of System and Productivity Management, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.

Abstract

Technological advances and the increasing need for civil, agricultural and modern construction projects have increased the importance of providing the equipment and tools needed in these projects. In this regard, commercial companies selling pipes and fittings play a key role in meeting the needs of customers. The purpose of this research is to cluster customers and predict their profitability for a commercial company in this field. For this purpose, data mining methods and machine learning algorithms such as K-means, Decision Tree, KNN and Naive Bayes were used. In the first step, using the K-means method and the Elbow technique, the optimal number of clusters was determined to be 3. In the next step, customer profitability classification and prediction was done using Decision Tree, KNN and Naive Bayes. The evaluation results of the models showed that the Decision Tree algorithm had the best performance with an accuracy of 98.72%. This research shows that by focusing on profitable customers and offering discounts or special privileges to them, the company can improve its profitability and also design incentive programs for other customers.

Keywords: Clustering, Prediction, CRISP method, Decision Tree, KNN, Naive Bayes.