



## ارزیابی مشتریان در زنجیره تأمین و به‌کارگیری تکنیک‌های داده‌کاوی در دسته‌بندی آن‌ها

تاریخ دریافت: ۱۳۹۱/۰۹/۰۲

تاریخ پذیرش: ۱۳۹۱/۰۹/۲۳

### چکیده

در دنیای امروز دسته‌بندی مشتریان سازمان از جنبه رفتاری، از اهمیت بالایی برخوردار است. تجزیه و تحلیل و درک رفتار مشتریان از اصول توسعه راهبردهای مراکز تولیدی - خدماتی می‌باشد به طوری که باعث جذب مشتریان بالقوه و نیز حفظ مشتریان فعلی و حداکثرسازی ارزش برای آن‌ها می‌شود. سازمان‌ها می‌توانند از طریق داده‌کاوی با استخراج اطلاعات پنهانی از پایگاه‌های داده‌ای بزرگ، مشتریان ارزشمند خود را تعیین و رفتار آینده‌شان را پیش‌بینی کنند. در این مقاله قصد داریم با استفاده از روش‌های داده‌کاوی به ارائه چارچوبی ساده و جامع پیرامون مدیریت ارتباط با مشتری و رابطه آن با زنجیره تأمین بپردازیم. ابتدا رفتار مشتریان را بر اساس معیار RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار در خرید) بررسی می‌نماییم و با توجه به تراکنش‌های انجام شده، مشتریان وفادار را مشخص می‌کنیم. سپس با استفاده از روش‌های دسته‌بندی نظیر: رگرسیون لجستیک، درخت تصمیم، شبکه عصبی و... که در تحقیقات مختلف مورد استفاده قرار گرفته‌اند، به شناسایی ویژگی‌های مشتریان می‌پردازیم تا سازمان به راحتی بتواند مشتریان وفادار را شناسایی نماید. نتایج این پژوهش نشان داد که تکنیک درخت تصمیم از لحاظ شاخص اطمینان، دارای عملکرد بهتری نسبت به سایر تکنیک‌های ذکر شده می‌باشد.

**واژه‌های کلیدی:** داده‌کاوی، دسته‌بندی، رفتارسنجی، زنجیره تأمین، مشتریان وفادار، شبکه عصبی.

## مقدمه

مشتریان، مهم‌ترین منابع یک سازمان هستند. بنابراین بسیار حیاتی است که علاوه بر مشتریان قدیم، مشتریان جدیدی را نیز به دست آورده و ارزش بالایی از آن‌ها را برای خود نگهداری نمایند. داده کاوی، ابزار مفیدی برای مدیران دانش است تا بتوانند کشف دانش را با تحلیل آن تلفیق کنند؛ تلفیقی که اغلب خود منجر به ایجاد دانش می‌شود، زیرا با در دسترس داشتن حجم زیادی از داده‌های مشتریان، می‌توان با استفاده از ابزارهای داده کاوی، فرصت‌های حیاتی برای ایجاد مزیت رقابتی در سازمان ایجاد نمود. برای مثال، وال مارت (بزرگ‌ترین فروشگاه زنجیره‌ای آمریکا) یک پایگاه داده‌ای با حجم ۴۳ ترابایت دارد که بزرگ‌تر از پایگاه داده‌ای جمع‌آوری خدمات مالیات بر درآمد می‌باشد (گرین، ۱۹۹۹). از سوی دیگر بسیاری از سازمان‌ها دریافته‌اند که دانش نهفته در این پایگاه عظیم داده‌ای، برای بهبود و گسترش سیستم‌های پشتیبان تصمیم‌گیری سازمان، نقش کلیدی دارد. بویژه این که در این زمینه، دانش مشتریان نیز برای تصمیم‌گیری درباره عملکرد زنجیره تأمین، یک متغیر حیاتی به شمار می‌رود. متأسفانه حجم زیادی از این دانش پنهان، بی‌استفاده مانده و به علاوه، رقابت شدید و افزایش فرصت انتخاب برای مشتریان، فشار جدیدی را بر تصمیم‌گیران سازمان وارد نموده است و به همین دلیل نیاز به مدیریت ارتباط با مشتری و نیز مدیریت دانش مشتریان در یک بازه زمانی طولانی بیش از پیش آشکار گردیده است.

برای تحقق این هدف، بسیاری از سازمان‌ها در اندیشه جمع‌آوری حجم قابل توجهی از پایگاه داده‌ای وسیع مرتبط با مشتریان خود هستند، تا بتوانند با به کارگیری داده‌ها و تجزیه و تحلیل آن‌ها، فرصت‌های پیش‌رو را شناسایی کرده و در جهت توسعه راهبردها به کار برند. از این‌رو، به جای هدف قرار دادن تمام مشتریان به صورت یکسان، یا فراهم نمودن پیشنهادهای مشابه برای ایشان، سازمان‌ها می‌توانند فقط مشتریانی را که نیازها و رفتارشان به طور یقین سودمند می‌باشد را انتخاب نمایند. اخیراً سازمان‌ها برای تبلیغ محصولات جدید و یا ارائه خدمات به مشتریان، از بازاریابی گروهی استفاده می‌کنند. آن‌ها در این روش، از یک ارتباط مستقیم با مشتریان برای ارائه ارزش‌ها و سرویس‌های تازه استفاده نمی‌کنند، بلکه از طریق

رسانه‌هایی نظیر ایمیل و دیگر رسانه‌های پیشرفته، یک پیام ارتباطی گسترده برای ایشان ارسال می‌نمایند.

از سوی دیگر همزمان با افزایش رقابت میان سازمان‌ها و تنوع محصولات آن‌ها، مشتریان فرصت انتخاب بیش‌تری را خواهند داشت و لذا سازمان‌ها تلاش می‌کنند تا با تشخیص نیازمندی‌های آن‌ها و متقابلاً ارائه خدمات بهتر، باعث افزایش وفاداری آن‌ها شوند. بر این اساس، تحلیل رفتار مشتریان برای افزایش میزان وفاداری آن‌ها از اهمیت بالایی برخوردار می‌باشد.

همچنین ارائه راهبرد مناسب در تعامل با مشتریان، علاوه بر افزایش وفاداری، می‌تواند کارایی مالی آن‌ها را نیز افزایش دهد که این امر برای سودآوری سازمان‌ها از جایگاه ویژه‌ای برخوردار است و با استفاده از این الگوها می‌توان راهبرد پیشگیرانه مناسبی را برای این‌گونه الگوهای رفتاری مشخص ساخت. از این‌رو داشتن مدلی مناسب برای تحلیل و بررسی الگوهای رفتاری مشتریان که توانایی پیش‌بینی و طبقه‌بندی آن‌ها را داشته باشند از اهمیت خاصی برخوردار می‌باشد.

### پیشینه تحقیق

در این زمینه کوهنن و همکاران (۲۰۰۰)، مدل ترکیبی الگوریتم درخت تصمیم و استدلال را در شناسایی مشتریان ارائه کردند و مالتوس (۲۰۰۱)، به ارزیابی مدل‌ها در بازاریابی مستقیم پرداخته و با مقایسه نتایج، نشان داد که مدل رگرسیون خطی مناسب‌تر است. کلمبو و جیانگ (۲۰۰۰)، مدل RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار) را برای دسته‌بندی مشتریان ارائه کردند و چیوو (۲۰۰۲)، نسبت به طراحی مدل ترکیبی الگوریتم ژنتیک و استدلال مبتنی بر بهبود دسته‌بندی مشتریان اقدام نمود.

ها و همکاران (۲۰۰۵) نیز مدل پاسخ بازاریابی مستقیم مشتریان را بر اساس شبکه عصبی پیشنهاد کردند و به ضریب دقت و صحت بالای ۹۵ درصد دست یافتند. بوس و چن (۲۰۰۹)، مدل‌های بازاریابی را معرفی و بررسی کرده و گوینراجان و چاندراسکاران (۲۰۱۰)، عملکرد مدلی بر اساس الگوریتم نزدیک‌ترین همسایگی را در خصوص بازاریابی مشتریان

ارزیابی کردند. کوی و همکاران (۲۰۰۸) مدل جدیدی در شناسایی مشتریان ارائه داده و عملکرد آن را بر اساس انتخاب ویژگی‌های مورد نظر ارزیابی نمودند. سالی هاردیچ (۲۰۰۸)، طی مقاله‌ای دیدگاه خود را نسبت به بازاریابی مستقیم ارائه داد و پریزی و وان دن پوئل (۲۰۰۵)، عملکرد مدل‌های بازاریابی را بهبود دادند. کرون و همکاران (۲۰۰۶)، تأثیر نحوه پردازش داده‌ها را در عملیات داده کاوی در مدل‌های بازاریابی مورد سنجش قرار دادند و هسی (۲۰۰۴)، مدل تحلیل رفتار مشتریان بانک را بر اساس مدل RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار) ارائه نمود. هاگتون و اولابی (۱۹۹۷)، از دو الگوریتم متفاوت درخت تصمیم برای ایجاد مدل پاسخ شناسایی مشتری مستقیم استفاده کردند و پس از مقایسه نتایج نشان دادند که اختلاف مهمی بین دو الگوریتم وجود ندارد. در نهایت چینگ و چن (۲۰۰۹)، نیز با استفاده از متغیرهای مدل RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار) به دسته‌بندی مشتریان مبادرت ورزیدند. در جدول شماره ۱ خلاصه مدل‌های بررسی شده در ادبیات تحقیق آورده شده است.

نام نویسنده	سال انجام تحقیق	روش به کار گرفته شده در بررسی رفتار مشتریان
هاگتون و اولابی	۱۹۹۷	درخت تصمیم
کلمبو و جیان	۲۰۰۰	RFM
کوهونن و همکاران	۲۰۰۰	درخت تصمیم
مالتوس	۲۰۰۱	رگرسیون
چیوو	۲۰۰۲	الگوریتم ژنتیک
هسی	۲۰۰۴	RFM
ها و همکاران	۲۰۰۵	شبکه عصبی
چینگ و چن	۲۰۰۹	RFM
گویندراجان و چاندراسکاران	۲۰۱۰	نزدیک‌ترین همسایگی

جدول ۱- نمایش خلاصه مدل‌های بررسی شده در مرور ادبیات

## مدیریت زنجیره تأمین

در فضای تجاری دنیای امروز، سازمان‌ها و مؤسسات باید محصولات متنوع را با توجه به درخواست مشتریان در دسترس ایشان قرار دهند. در شرایط حاضر، خواست مشتری در زمینه کیفیت بالا و نیز خدمت‌رسانی سریع، موجب افزایش فشارهایی شده که در گذشته وجود نداشته است. در نتیجه، سازمان‌ها نمی‌توانند به تنهایی از عهده رفع تمامی مسائل برآیند و به ناچار با بهره‌گیری از انواع فن‌آوری‌ها و علوم مدیریتی و ارتباطات سازمانی، برای شناسایی مزایای رقابتی خویش اقدام به طراحی نقشه راه می‌کنند.

از مهم‌ترین علوم مدیریتی که به عنوان یکی از مبانی زیرساخت پیاده‌سازی کسب و کار الکترونیک در دنیا مطرح است، مدیریت زنجیره تأمین (SCM) است. این دانش در طول قرن بیستم میلادی ظهور و شکل یافته و مباحث سودمندی را مطرح کرده است، به گونه‌ای که بنگاه اقتصادی موفق شود در زمان بسیار کمتری محصول خود را به بازار عرضه کرده و هزینه‌های اتلافی را کاهش دهد و متقابلاً مشتریان نیز بتوانند کالاها و خدمات قابل اطمینان و سریعی را با کیفیت بالا و حداقل هزینه دریافت کنند.

در بازار رقابتی موجود، بنگاه‌های اقتصادی و تولیدی علاوه بر پرداختن به سازمان و منابع داخلی، خود را نیازمند به مدیریت و نظارت بر منابع و ارکان مرتبط خارج از سازمان یافته‌اند. علت این امر در واقع دستیابی به مزیت رقابتی با هدف کسب سهم بیشتری از بازار است. بر این اساس فعالیت‌هایی نظیر برنامه‌ریزی عرضه و تقاضا، تهیه مواد، تولید و برنامه‌ریزی محصول، نگهداری کالا، کنترل موجودی، توزیع، تحویل و خدمات مشتریان که قبلاً همگی در سطح شرکت انجام می‌شدند، اکنون به سطح زنجیره تأمین انتقال پیدا می‌کنند. نباید فراموش کرد که مسئله کلیدی در یک زنجیره تأمین، مدیریت و کنترل هماهنگ تمامی این فعالیت‌ها است. لذا مدیریت زنجیره تأمین، فرآیند یکپارچه‌سازی فعالیت‌های زنجیره تأمین و نیز جریان‌های اطلاعاتی مرتبط با آن، از طریق بهبود و هماهنگ‌سازی فعالیت‌ها می‌باشد. بر این اساس، با بهره‌گیری از نتایج مباحث علمی مزبور، سازمان‌ها قادر خواهند بود روابط تجاری خود را با همکاران و شرکای تجاری نظیر تأمین‌کنندگان مواد

اولیه، توزیع‌کنندگان محصولات و پیمان‌کاران حمل و نقل کالا به منظور بهینه‌سازی تبادل اطلاعات توسعه دهند.

## داده کاوی

داده کاوی به مفهوم استخراج دانش از حجم عظیمی از داده‌ها و همچنین کشف الگوهای مناسبی در این میان است که نوعاً در داده‌های مذکور پنهان می‌باشند. از داده کاوی می‌توان در انجام عملیاتی نظیر دسته‌بندی، پیش‌بینی، تخمین و خوشه‌بندی داده‌ها استفاده کرد که این تکنیک‌ها، امکان استخراج روابط معنایی و عمقی را فراهم می‌نمایند.

عملیات داده کاوی، به سازمان‌ها کمک می‌کند تا با کاوش بر روی داده‌ها، برخی الگوها، روندها و رفتارهای آینده را کشف و پیش‌بینی کرده و موجبات تصمیم‌گیری مناسب‌تر را فراهم نمایند. در این عملیات با استفاده از تحلیل وقایع گذشته، نسبت به پیش‌بینی و تحلیل اتوماتیک اقدام گردیده و به سؤالاتی جواب می‌دهد که پاسخ به آن‌ها در گذشته ممکن نبوده و یا به زمان زیادی نیاز داشته است (هوور و پرز، ۲۰۰۰). مؤسسات مخابراتی، بانک‌ها، شرکت‌های بازاریابی و تبلیغاتی و کلیه سازمان‌هایی که از بانک‌های اطلاعاتی بزرگی برخوردارند جملگی می‌توانند از داده کاوی به عنوان یک ابزار قوی برای رشد و پیشرفت در بازار استفاده نمایند (بری، ۱۹۹۷).

از کاربردهای دیگر داده کاوی، مدیریت ارتباط با مشتریان است که به عنوان یک ابزار بنیادین برای آشکارسازی خصوصیات جمعیت شناختی مشتریان به کار می‌رود. از جمله کارهای مهم قابل انجام در بستر داده کاوی در مدیریت ارتباط با مشتری، بررسی تأثیر نشان تجاری در خرید مشتریان می‌باشد.

از فنون داده کاوی، برای دستیابی به دامنه وسیعی از اهداف گوناگون استفاده می‌شود که در ذیل به مواردی از کاربردهای آن اشاره می‌گردد:

- شناسایی مشتریان سودآور؛
- پیش‌بینی رفتار خرید مشتریان؛
- بهینه‌سازی سهم سبد خرید مشتریان؛



- تخمین کارآمدی تبلیغات؛

- برآورد و اولویت‌بندی ریسک اعتباری مشتریان.

باید توجه داشت که کشف و استخراج دانش و یادگیری، یک فرآیند تکرارپذیر است که تا زمان تأیید صلاحیت مدل، ادامه دارد و شامل مراحل زیر است:

۱. **اهداف داده کاوی** (اهداف تصمیم‌گیری): باید قبل از انجام عملیات داده کاوی، اهداف مورد نظر را شناسایی و تعیین نموده و دانش اولیه را جمع‌آوری کرد، زیرا اگر قبل از شناسایی اهداف بخواهیم تکنیک‌های داده کاوی را به کار گیریم، به الگوهای بی‌معنا و نامشخصی خواهیم رسید.

۲. **جمع‌آوری داده‌ها**: در این مرحله تصمیم گرفته می‌شود که عملیات کاوش صرفاً بر روی نمونه‌هایی از داده‌ها صورت بگیرد یا کل پایگاه بررسی شود.

۳. **پیش پردازش داده‌ها**: این مرحله دارای بخش‌های گوناگونی به صورت زیر است:

• **انتخاب داده‌ها**: در این قسمت، باید داده‌ها انتخاب و بررسی شوند که آیا برای هدف مورد نظر، مناسب هستند یا خیر؛

• **پاک‌سازی داده‌ها**: اکنون داده‌ها پس از انتخاب، باید آماده‌سازی شوند. لذا داده‌های پرت (اضافه)، متعادل‌سازی می‌شوند، یعنی نقاط پرت حذف شده و یا مقدار رکورد میانگین معادل را به جای آن قرار می‌دهند؛

• **تبدیل و انتقال داده‌ها**: در این قسمت، بر روی داده‌ها تغییراتی انجام می‌شود، مثل اصلاح فرمت، تاریخ، واحد پول و... و همچنین تبدیل واحدها، مقیاس‌ها و داده‌های عددی به متغیر و بالعکس.

۴. **داده کاوی**: در این قسمت، وظایف داده کاوی که در پیش گفته شد بر روی داده‌ها صورت می‌گیرد.

۵. **تصویری کردن**: از آنجا که انسان‌ها در پردازش و استفاده اطلاعات تصویری موفق‌تر هستند و یک شکل تصویری می‌تواند در مدت چند ثانیه اطلاعات زیادی را به آن‌ها انتقال بدهد، در این مرحله به بررسی داده‌ها با استفاده از تصویری کردن آن‌ها اقدام می‌شود. البته

این کار می‌تواند هم پیش از عملیات کاوش (به منظور محسوس شدن داده‌ها و تعیین متغیرها و شاخص‌های مناسب) و هم بعد از آن (به منظور بررسی الگوهای به دست آمده) مورد استفاده قرار گیرد.

۶. **ارزیابی مدل:** پس از مراحل بالا، مدل حاصله، ارزیابی شده و صلاحیت آن تأیید می‌گردد. در داده کاوی، اطمینان از این که دانش به دست آمده از داده کاوی قابل اعتماد است یا خیر، بسیار مهم می‌باشد.

۷. **کشف و استخراج دانش:** پس از تعیین صلاحیت مدل‌های حاصله، اطلاعات به دست آمده از مدل‌ها توسط خبرگان به دانش مفید و قابل توجه تبدیل می‌شود.

### دسته‌بندی

عملیات دسته‌بندی یکی از معروف‌ترین و پرکاربردترین فعالیت‌هایی است که در داده کاوی مورد استفاده قرار می‌گیرد. هدف از دسته‌بندی، سازمان‌دهی و گروه‌بندی داده‌ها در دسته‌های مشخص می‌باشد. در این روش، ابتدا بر اساس توزیع داده‌ها، مدلی ایجاد می‌شود که از آن بعداً برای دسته‌بندی داده‌ها در گروه‌های مورد نظر استفاده می‌شود. با استفاده از روش دسته‌بندی، می‌توان برای مقادیر اسمی و گسسته پیش‌گویی انجام داد. در روش دسته‌بندی، عنوان طبقات مشخص هستند و به همین دلیل دسته‌بندی با نظارت نامیده می‌شود. پس از انجام دسته‌بندی، می‌توان مقدار هر شاخص را بر اساس مقادیر شاخص‌های دیگر، پیش‌بینی نمود (بری، لیئف، ۲۰۰۴).

روش‌هایی که در دسته‌بندی داده‌ها به کار گرفته می‌شوند عبارتند از:

- درخت تصمیم؛
- شبکه‌های عصبی؛
- الگوریتم ژنتیک؛
- رگرسیون لجستیک؛
- نزدیک‌ترین همسایگی.



پس از به‌کارگیری این روش‌ها می‌توان بر اساس نتایج حاصله و نیز تست داده‌ها، یک سری قوانین «اگر - آنگاه» را ارائه کرد (پرسون، اسمیت، ۲۰۰۸).

## روش تحقیق

در این بخش به بررسی تکنیک‌های دسته‌بندی در داده کاوی، که در این پژوهش از آن‌ها استفاده شده، می‌پردازیم:

### ۱. درخت تصمیم

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی است. ساختار درخت تصمیم، شبیه به یک ساختار فلوچارت (نمودار جریان کار) است. بالاترین گره، گره ریشه است و گره‌های برگ، دسته‌ها یا توزیع دسته‌ها را نشان می‌دهند. در این ساختار هر گره داخلی، آزمونی را بر روی یک ویژگی مشخص انجام داده و هر شاخه خارج شده از این گره، دستاورد این آزمون را نشان می‌دهد. درخت تصمیم برخلاف شبکه‌های عصبی، به تولید قانون می‌پردازد، یعنی پیش‌بینی خود را در قالب تولید یک سری قوانین توضیح می‌دهد، در حالی که در شبکه‌های عصبی تنها پیش‌بینی نهایی بیان می‌شود و چگونگی آن در خود شبکه باقی می‌ماند (کورپرین، ۱۹۹۹).

الگوریتم درخت تصمیم با انتخاب آزمون شروع می‌شود که بهترین جداسازی را برای دسته‌ها انجام می‌دهد. مهم‌ترین هدف از انجام دسته‌بندی، به دست آوردن مدلی برای پیش‌بینی است. بدین منظور از مجموعه‌ای از داده‌ها به نام داده‌های آموزشی<sup>۱</sup> (ساخت مدل)، که مجموعه‌ای از متغیرها و رکوردها است، استفاده می‌شود. در مراحل بعدی، همین کار برای گره‌های پایین‌تر با داده‌های کم‌تری صورت می‌گیرد تا بهترین قانون‌ها حاصل شود و سرانجام درخت مزبور آن‌قدر بزرگ می‌شود تا دیگر نتوان جداسازی بهتری برای داده‌های گره انجام داد (عادل آذر و همکاران، ۱۳۸۹).

اکنون و در این مرحله باید اثر بخشی درخت ایجاد شده اندازه‌گیری شود. برای این کار از مجموعه‌ای از داده‌های آزمایشی ۱ استفاده می‌شود که متفاوت با داده‌های اولیه‌ای است که درخت را ایجاد کرده‌اند. معیاری که اندازه‌گیری می‌شود نیز عبارت است از درصد داده‌هایی که به صورت صحیح دسته‌بندی می‌شوند و دسته پیش‌بینی شده با دسته واقعی آن‌ها یکسان است. به طور کلی مزایای استفاده از شیوه درخت تصمیم نسبت به سایر تکنیک‌های داده کاوی به شرح زیر است:

(الف) نسبت به سایر مدل‌های دسته‌بندی زودتر محاسبه می‌شود؛

(ب) دقت آن نسبت به سایر مدل‌ها بیشتر است؛

(ج) برای یادگیری، ساده و آسان‌تر می‌باشد؛

(د) قانون‌های بدست آمده در آن راحت‌تر درک می‌گردد.

آزمون‌های انجام شده در هر یک از الگوریتم‌های شناخته شده درخت تصمیم، مختلف بوده و انتخاب شاخه‌ها و جداسازی در هر یک به روش متفاوتی انجام می‌گیرد (عادل آذر و همکاران، ۱۳۸۹).

## ۲. شبکه‌های عصبی

به دلیل توسعه شبکه‌های عصبی در سال‌های اخیر، به عنوان یک ابزار داده کاوی متداول شده‌اند. این شبکه‌ها، ابزارهایی هستند که در موارد پیش‌بینی، رده‌بندی، خوشه‌بندی و تخمین کاربرد دارند. جذابیت شبکه‌های عصبی در این است که از طریق مدل‌سازی ارتباطات عصبی مغز انسان در رایانه‌های دیجیتال، پلی به عنوان رفع فاصله ایجاد می‌کنند (بیگاس، ۱۹۹۶).

شبکه‌های عصبی، روش محاسباتی متفاوتی نسبت به روش‌های متداول دارند. محاسباتی که با روش‌های معمولی انجام می‌شود، از نوع برنامه‌ریزی شده بوده و در آن‌جا از الگوریتم‌ها و مجموعه‌هایی از قواعد، برای حل مسئله استفاده می‌شود. در این روش‌ها اگر الگوریتمی در دست نبوده یا پیچیده باشد، راه حلی برای حل مسئله وجود نخواهد داشت، اما در محاسبات

اجرا شده توسط شبکه عصبی، به الگوریتم و مجموعه قواعد نیاز نیست. شبکه‌های عصبی به همراه درخت تصمیم، در زمره نسل جدید تکنیک‌های داده کاوی به شمار می‌آیند که در دو دهه اخیر توسعه زیادی یافته‌اند. از این تکنیک‌ها می‌توان هم برای کشف و استخراج دانش از یک پایگاه داده‌ای و هم برای ایجاد مدل‌های پیش‌بینی کننده استفاده نمود (امبرکت، آرسینیگاس و همکاران، ۲۰۰۳).

### ۳. رگرسیون لجستیک

رگرسیون لجستیک همانند رگرسیون خطی، به ترکیب خطی ویژگی‌ها و خصوصیات بستگی دارد. تابع لجستیک، ابزاری برای تبدیل ترکیب خطی، درون یک فاصله صفر و یک می‌باشد (یی، ۲۰۰۳). رگرسیون لجستیک شامل دو گام است:

**گام اول:** تخمین احتمال مشخصه‌های هر گروه تشکیل شده؛

**گام دوم:** برش نقاط با این احتمالات، برای طبقه‌بندی هر یک از مشخصه‌ها در یک گروه. متغیرهای مدل با روشی تحت عنوان (بیش‌ترین احتمال تکرار موفق یک فرآیند) تخمین زده می‌شوند. به طور کلی، رگرسیون لجستیک مشابهت‌های رفتاری را برای آزمون تشخیصی نشان می‌دهد (پالمر، ۲۰۱۱). این تکنیک یکی از روش‌های آماری برای تحلیل داده‌های دودویی است و در زمینه داده کاوی و یادگیری ماشین مورد استفاده قرار می‌گیرد. به کمک این روش می‌توان به طبقه‌بندی داده‌ها و پیش‌بینی اطمینان از خروجی دودویی و... پرداخت (جاکولا، جورد، ۱۹۹۷).

### جمع‌آوری داده‌ها

به طور معمول، دو نوع داده مربوط به مشخصات مشتری و خریدار وجود دارد که در مدل‌های کمی برای بازاریابی مستقیم استفاده می‌شود:

**نوع اول اطلاعات:** شامل اطلاعات جغرافیایی، جمعیتی، سبک زندگی، خصوصیات اجتماعی و گرافیکی مشتریان است (بولت، ۱۹۹۶).

- اطلاعات جمعیتی: سن، جنس، اندازه خانواده، شیوه زندگی مشتریان و... است؛

- داده‌های سبک زندگی: عادات مشتریان، علایق تفریحی و... می‌باشد؛

- داده‌های جغرافیایی: موقعیت منازل، اداره‌ها و کسب و کار مشتریان است.

این نوع از اطلاعات، داده‌های خارجی نامیده می‌شوند.

**نوع دوم اطلاعات:** شامل رفتار تعاملی مشتریان با بازاریابان است. اطلاعات در مورد رفتار مشتریان شامل سوابق معامله مشتریان و کسب بازخورد از آنان می‌باشد. متغیرهای RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار در خرید) اغلب از سوابق معاملاتی مشتری و خریدار استخراج می‌شوند (جونکر، پی یرسما، ۱۹۹۹؛ مالتوس، ۲۰۰۶).

**تأخیر در خرید (R):** از آخرین تاریخ و زمان مراجعه مشتری چه مدتی سپری شده است؟ با پاسخ به این سؤال، بازاریاب می‌تواند دلایل عدم مراجعه یک مشتری یا گروهی از مشتریان وفادار فروشگاه را بیابد و روش‌های مناسبی را برای جذب دوباره آن‌ها به کار بندد. در بیشتر موارد ممکن است دلیل این امر (از دست دادن مشتری)، مراجعه آن‌ها به سازمان دیگری باشد.

**حجم خرید (F):** مشتری در یک خرید عادی از سازمان شما چه مقدار خرید می‌کند؟ این اطلاعات کمک می‌کند تا بازاریاب، منابع تبلیغاتی مناسب را به آن دسته از مشتریانی که بیش‌تر خرج می‌کنند، اختصاص دهد.

**میزان خرید (M):** آیا اغلب مشتری برای خرید محصولات مورد نیاز خود به سازمان شما می‌آید؟

میزان خرید مبلغ پولی است که ارزش خرید مشتری در طول یک دوره زمانی معین و یا آخرین معامله خرید مشتری را نشان می‌دهد.

**اطلاعات RFM** (تاریخ آخرین خرید، حجم مالی خرید، تکرار در خرید) به تخمین این احتمال که مشتریان محصول خاصی را خریداری خواهند کرد، کمک می‌کند. پژوهشگران نیز از طریق انتقال بازخوردهای مشتریان و خریداران به بازاریابان، به کشف رفتار مشتریان در برابر محصولات یا خدمات سازمان کمک می‌کنند.

در این بخش کارایی چند روش دسته‌بندی به منظور ارزیابی مشتریان بررسی شده است. در روش‌های دسته‌بندی، نیازمند طراحی مدل هستیم تا با استفاده از آن بتوانیم رفتار مشتریان آتی را بسنجیم. با ایجاد این مدل اعتبار سنجی، داده‌های قدیمی را به منظور تعیین خصوصیات رفتاری مشتریان، تحلیل می‌کنیم.

مجموعه‌ای از داده‌ها، به منظور آموزش و آزمایش روش‌های دسته‌بندی به کار گرفته شده است. در این پژوهش از مجموعه داده‌های مشتریان که مربوط به یک مرکز خرید فروشگاه می‌باشد، استفاده شده است. معیارهایی که در مدل مورد بررسی قرار گرفته، متشکل از ۱۲ ویژگی مربوط به مشتریان می‌باشد که برخی از داده‌های آن عددی و برخی نمادین است. در جدول ۲ هر یک از این خصوصیات نشان داده شده است.

ردیف	معیار	نوع معیار
۱	جنسیت	نمادین
۲	وضعیت تأهل	نمادین
۳	تحصیلات	نمادین
۴	سن	عددی
۵	شغل	نمادین
۶	درآمد سالانه	عددی
۷	منزل مسکونی	نمادین
۸	استفاده از وام بانکی	نمادین
۹	وضعیت خودرو شخصی	نمادین
۱۰	نحوه آشنایی (تبلیغات)	نمادین
۱۱	مقدار خرید کالا	عددی
۱۲	زمان خرید کالا	عددی
۱۳	هزینه پرداخت شده	عددی
۱۴	نوع محصول	نمادین
۱۵	شناسه مشتری	نمادین

جدول ۲- ویژگی‌های مشتریان

در گام نخست با استفاده از فرآیند تحلیل سلسله مراتبی و استفاده از نظر خبرگان، میزان اهمیت هر یک از شاخص‌های مورد بررسی RFM (آخرین بار خرید، حجم مالی خرید، تکرار در خرید) تعیین گردیده که در جدول ۳ نشان داده شده است.

شاخص‌ها		وزن	اولویت
پولی	MONETARY	٪۵۷	۱
تکرار در خرید	FREQUENCY	٪۳۱	۲
آخرین زمان مراجعه	RECENCY	٪۱۲	۳
کل	TOTAL	٪۱۰۰	۴

جدول ۳- میزان وزن شاخص‌ها

تشکیل پروفایلی از مشتریان جهت شناسایی رفتار (وفاداری و عدم وفاداری) میزان وفاداری مشتریان با کیفیت و تنوع کالا و خدمات ارائه شده به ایشان رابطه مستقیمی دارد. در ادامه، رفتار مشتریان و راهبرد سازمان‌های مرتبط با آن‌ها شرح داده می‌شود.

### - مشتریانی با الگوی: $R \downarrow F \uparrow M \uparrow$

با توجه به این که هر سه متغیر برای این دسته، وضعیتی مطلوب از نظر (وفاداری و ارزش برای مراکز) دارند، می‌توان گفت دسته‌هایی با این الگو، جزء ارزشمندترین مشتریان هستند. بنابراین، این دسته از مشتریان، مهم‌ترین دسته از نظر الگوی وفاداری و ارزش برای سازمان‌ها محسوب می‌شوند و باید هزینه‌های بیش‌تری صرف آن‌ها گردیده و همیشه در جلب رضایت بیش‌تر آنان تلاش شود. با توجه به تعداد کم این دسته از مشتریان، سازمان‌ها باید با ارائه خدمات ویژه در راستای وفادارسازی آنان اقدام کنند تا در دوره‌های بعدی به گروه مشتریانی با این الگو افزوده شود.

اکنون با استفاده از اولویت‌بندی‌های بالا، مشتریان سازمان بر طبق RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار در خرید) اولویت‌بندی می‌شوند که در این صورت در جدول زیر (شماره ۴) خواهیم داشت:

امتیازبندی (وزن‌دهی) مشتریان	بخش‌های مختلف ممکن	ردیف
(وفادارترین) ۱۰۰٪	$R \downarrow F \uparrow M \uparrow$	۱
۸۸٪	$R \uparrow F \uparrow M \uparrow$	۲
۶۹٪	$R \downarrow F \downarrow M \uparrow$	۳
۵۷٪	$R \uparrow F \downarrow M \uparrow$	۴
۴۳٪	$R \downarrow F \uparrow M \downarrow$	۵
۳۱٪	$R \uparrow F \uparrow M \downarrow$	۶
۱۲٪	$R \downarrow F \downarrow M \downarrow$	۷
(غیر وفادارترین) ۰٪	$R \uparrow F \downarrow M \downarrow$	۸

جدول ۴- نحوه امتیازبندی مشتریان

با توجه به جدول امتیازبندی، مشتریان ردیف یک تا چهار، جزء مشتریان وفادار محسوب می‌شوند و باقی جزء دسته غیر وفادار طبقه‌بندی می‌شوند.

مجموعه داده‌های ذکر شده مشتمل بر تراکنش‌های ۱۰۲۳ مورد از مشتریان فروشگاه زنجیره‌ای می‌باشد که سوابق آن‌ها در یک دوره زمانی ثبت شده است. ۷۰ درصد از داده‌های موجود برای ساخت مدل و ۳۰ درصد برای تست مدل و ارزیابی عملکرد آن به کار رفته است.

تخصیص داده‌ها به دو دسته داده‌های ساخت مدل و آزمایش آن به صورت کاملاً تصادفی انجام شده است. همان‌طور که در بخش قبلی ذکر شد، از سه روش دسته‌بندی طبق جدول ۵ برای پیش‌بینی وضعیت رفتاری مشتری استفاده می‌شود.

پارامترها	روش دسته‌بندی
-	رگرسیون لجستیک
نوع درخت: (CHAID) / لایه های زیرین درخت: ۷ / آلفای جدا سازنده: ۰.۲ / آلفای ادغام: ۰.۵ / اپسیلون همگرایی ۰.۰۰۱ / حداکثر تکرار برای همگرایی: ۱۰۰	درخت تصمیم
نوع شبکه عصبی: پرسپترون چند لایه‌ای / گره‌های لایه ورودی: ۱۲ / تعداد لایه‌های مخفی یک: ۲ / تعداد گره های لایه مخفی دو: ۳ / تعداد گره‌های لایه خروجی: ۱	شبکه عصبی

جدول ۵- متغیرهای مؤثر در روش‌های دسته‌بندی

رگرسیون لجستیک، روابط بین متغیرها در تابع را خطی فرض نموده و خروجی‌های دو بخشی می‌دهد که به دلیل وجود این رابطه (خطی) می‌توان گفت که عدم صحت در آن وجود دارد.

در روش‌های درخت تصمیم، شبکه عصبی به دلیل دارا بودن فرض رابطه غیر خطی بین متغیرها، قابلیت بهتری برای دسته‌بندی مشتریان دارند.

مدل	میانگین دقت دسته‌بندی	میانگین خطای دسته‌بندی
رگرسیون لجستیک	٪۷۰.۱	٪۲۸.۹
درخت تصمیم	٪۷۵.۴۷	٪۲۴.۳۳
شبکه عصبی	٪۷۱	٪۲۹

جدول ۶- مقایسه میانگین دقت و خطای مدل‌های دسته‌بندی

با توجه به جدول بالا، مشاهده می‌شود با افزایش میانگین دقت دسته بندی، میانگین خطای کل در دسته‌بندی‌ها کاهش می‌یابد. از این رو با افزایش دقت و کاهش خطا در رفتارسنجی مشتریان می‌توان اطمینان وفادار بودن مشتریان را در سازمان افزایش داده و هزینه‌های مالی را به طور قابل ملاحظه‌ای کاهش داد.



## نتیجه‌گیری

علی‌رغم این‌که داده کاوی در زمینه‌های مختلفی همچون کشفِ خطا (تقلب)، پیش‌بینی قیمت سهام، تشخیص طبی و یافته‌های علمی مورد استفاده واقع می‌شود اما کاربرد آن در سیستم‌های پشتیبان تصمیم‌گیری، خصوصاً در زمینه مدیریت زنجیره تأمین منحصر به فرد است. در محیط تجاری مشتری مدار کنونی، باور و عقیده هر سازمانی، اعم از مالی و صنعتی چنین است که فهم و درک عمیق‌تری از نیازهای مشتریان به منظور استفاده در تصمیم‌گیری‌های مهم خویش داشته باشند و این امر، جز با استفاده از داده کاوی، استخراج دانش و مدیریت آن حاصل نمی‌گردد. در راستای نیل به هدف مذکور، در این مقاله نشان داده شد که چگونه می‌توان داده کاوی را به عنوان یک تکنیک کارآمد برای استخراج دانش در یک چارچوب یکپارچه به کار گرفت. همچنین در این چارچوب مشخص شد چگونه می‌توان با دسترسی به حجم زیادی از داده‌ها، که ممکن است به کمک تکنیک‌های پیشرفته اطلاعاتی به دست آمده باشد، آن‌ها را تحلیل و مدیریت کرد.

دسته‌بندی مشتریان به گروه‌های مختلف، باعث بهبود کیفیت محصولات می‌گردد؛ همچنین این امکان را برای تصمیم‌گیران به وجود می‌آورد که تقسیم‌بندی بازار را به طور واضح‌تری تعیین کرده و راهبردهای بازاریابی را بهبود بخشید. از آنجا که مشتریان با هم برابر نیستند، هدف داده کاوی این است که مؤسسات بدانند تمایل مشتریان چگونه است تا بتوانند وفاداری آن‌ها را جلب کرده و پیرو آن به افزایش میزان درآمد و کاهش هزینه‌ها دست یابند. شناخت بهتر مشتریان و درک تفاوت بین آن‌ها، به سازمان‌ها کمک می‌کند در ارائه محصولات جدید و تخصیص منابع به مشتریان مطلوب‌تر (وفادارتر) کارایی بیش‌تری داشته باشند. در این راستا مقاله حاضر مدلی را برای شناسایی مشتریان وفادار و دسته‌بندی آن‌ها ارائه کرده است.

در این پژوهش، برای رتبه‌بندی مشتریان، ابتدا رفتار آن‌ها را بر اساس مدل RFM (تاریخ آخرین خرید، حجم مالی خرید، تکرار در خرید) مشخص کردیم و سپس با استفاده از نظرات کارشناسان، وزن مربوط به هر عامل را به دست آورده و با توجه به سابقه تراکنش‌های مشتریان، آن‌ها را به دو دسته وفادار و غیر وفادار تقسیم نمودیم. این تحقیق با

هدف ارائه الگویی جهت تحلیل رفتار و ارزش مشتریان به انجام رسید و در کنار این هدف اصلی، سه روش برای دسته‌بندی مشتریان از نظر سطح وفاداری آن‌ها مورد استفاده قرار گرفت که درخت تصمیم به عنوان بهترین و مطمئن‌ترین روش انتخاب شد.

## منابع

۱. آذر، عادل و احمدی، پرویز، (۱۳۸۹)، «طراحی مدل انتخاب نیروی انسانی با رویکرد داده کاوی»، نشریه فن‌آوری اطلاعات، شماره ۲.

۲. حسین‌زاده، لیلا، (۱۳۸۶)، «دسته‌بندی مشتریان هدف در صنعت بیمه با استفاده از داده کاوی»، پایان‌نامه کارشناسی ارشد (مدیریت فن‌آوری اطلاعات).
۳. شهرابی، جمال، (۱۳۸۶)، داده کاوی، تهران، جهاد دانشگاهی دانشگاه صنعتی امیرکبیر.
4. H. Abdi.(2007). " Discriminant correspondence analysis", Salkind (Ed.): Encyclopedia of Measurement and Statistics. Thousand Oaks (CA).Sage.pp. 270-- 275.
5. A Froot,&Kenneth & S. Scharfstein&David& C. Stein& Jeremy.(1993 );" Risk Management: Coordinating Corporate Investment and Financing Policies".The Journal of Finance. Vol. 1, No. 2, pp. 1629- 1658.
6. Basel Committee on Banking Supervision.(2003). "Risk Management Principle for Electronic Banking", Bank for international settlement.
8. D.J. Bartholomew, & M.Knott, (1998)."Latent Variable Models and Factor Analysis". Oxford University Press, London.
9. M.Berry, L.Gordon .(1997)."Mastering Data Mining: The Art and Science of Customer Relationship".
10. A.Berson.,(2000). "Building data mining applications for CRM".
11. J. Bigus (1996). "Data Mining with Neural Networks ".
12. M. A.Boyacioglu, Y.Kara, & O. K.Baykan, (2009).“Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund(SDIF) transferred banks in Turkey”. Expert systems with applications ,
13. R.Burbidge, M.Trotter, B.Buxton, & S.Holden.(2001)“Drug design by machine learning: Support vector machine for pharmaceutical data analysis”. Computer &chemistry.
14. L. M.Camarinha.( 1998). “Application of machine learning in water distribution networks”.Intelligent data analysis.
15. J. H.Chen, L.R. Su.Yang, , & J. Z.Lin. ( 2010). “A rule extraction based approach in predicting derivative use for financial risk hedging by construction companies”. *Expert systems with applications*.
16. J.Han, & M.Kamber, Data Mining.(2006). “Concepts and Techniques”.an Francisco, U.S.A, Morgan Kaufman Publishers.
17. A. Hormozi , "Data Mining.(2004).“A Competitive weapon for banking and retail industries”, information systemsmanagement.

18. Y. S.Kim, & S. Y.Sohn (2004).“Managing loan customers using misclassification patterns of credit scoring model.Expert Systems with Applications”.26, 567-5730.
19. T.Kohonen. (1988). “Self-organization and associative memory” , New York.Springer.
20. R.Malhotra, & D. K. Malhotra,(2005).“Evaluating consumer loans using neural networks”. Omega, 31(2), 83-96.
21. E.W.T, Xiu Ngai, , D.C.K.Li, Chau,(2009).“Application of data mining techniques in customer relationship management”.A literature review and cl.
22. S.Yang Zhao.(2008). “ Expert Systems with Applications journal'.GA,70003.