

رویکرد مبتنی بر داده‌کاوی در مدیریت ارتباط با مشتری و بازاریابی

هادی ویسی^۱ - عضو هیئت علمی دانشکده علوم و فنون نوین، دانشگاه تهران
عرفانه غروی - دانشجوی دکتری دانشکده علوم و فنون نوین، دانشگاه تهران

چکیده

حضور در بازار پرتلاطم رقابت، سازمان‌ها را به استفاده از رویکردهای نوین جهت نگهداشت مشتریان جاری، شناسایی و جذب مشتریان جدید سوق داده است. روش‌های داده‌کاوی که قابلیت کشف دانش از انبوه داده‌ها را فراهم می‌کنند، در این موضوع به کمک سازمان‌ها آمده و آنها را در راستای رسیدن به اهداف بازاریابی و مدیریت ارتباط با مشتری^۲ شامل شناسایی، جذب، نگهداشت و ارتقای مشتری یاری می‌رسانند. در این مقاله، پس از معرفی مدیریت ارتباط با مشتری، روش‌های داده‌کاوی به عنوان ابزاری جهت کشف دانش جهت پاسخگویی به نیازهای کسب و کار و استراتژی‌های مدیریت ارتباط با مشتری مورد بررسی قرار گرفته است. به علاوه، مروری بر کاربردهای مختلف داده‌کاوی در مدیریت ارتباط با مشتری در صنایع مختلف شامل خرده‌فروشی، بانکداری و مخابرات صورت گرفته است. پس از آن، از روش‌های داده‌کاوی برای سه کاربرد مختلف به عنوان مطالعات موردی استفاده شد که شامل ۱- بررسی تأثیر عنوان تجاری بر سبب خرید مشتری در مطالعه داده‌های فروشگاه شهروند، ۲- پیش‌بینی سرویس مورد تقاضای آینده مشتری در استفاده از اینترنت و ۳- بررسی روی‌گردانی مشتری از اینترنت هوشمند با استفاده از ارزش طول عمر مشتری است. در مطالعات صورت گرفته از روش‌های مختلف داده‌کاوی مانند Apriori، شبکه عصبی مصنوعی و درخت تصمیم استفاده شده که نتایج حاصل بیانگر کشف قواعد حاکم بر حجم انبوه داده‌ها در این کاربردها بوده است. نتایج بیانگر توانایی بالای این روش‌ها در کشف دانش درونی داده‌های مطالعه شده در سه مطالعه موردی بوده و در بهترین حالت، به دقت ۹۲،۵۹٪ در پیش‌بینی صحیح تقاضای آینده مشتری دست یافته است.

کلمات کلیدی

مدیریت ارتباط با مشتری، بازاریابی، داده‌کاوی، شبکه عصبی مصنوعی، درخت تصمیم، Apriori

۱. نویسنده مسئول: هادی ویسی، h.veisi@ut.ac.ir

2. Customer Relationship Management (CRM)

۱. مقدمه

امروزه صاحبان مشاغل به دنبال رویکردهایی نوین برای پاسخ به تغییرات محیط کسب و کار و نیازمندی‌های جدید مشتریان هستند. مفهوم بازاریابی انبوه که در گذشته کاربرد فراوانی داشت، اکنون با ایده مشتری محوری جایگزین شده است. از طرفی، فزونی و دسترسی به انبوه اطلاعات، مشتریان را آگاه‌تر و ماهرتر ساخته است. بنابراین سازمان‌ها باید از خط‌مشی‌هایی چون مدیریت ارتباط با مشتری، برای جذب و نگهداشت مشتری بهره ببرند. در گذشته رویکرد اولیه برای حفظ مشتری عمدتاً با استفاده از تبلیغات پستی و رسانه‌ها بوده، اما این رویکرد تبلیغاتی به صورت فراگیر، افراد بسیاری شامل مشتریان فعلی و افرادی را که هرگز مشتری ما نمی‌شوند نیز در بر می‌گرفت.

مدیریت ارتباط با مشتری کارآمد تنها در صورتی حاصل می‌شود که سازمان درک درستی از ترجیحات و نیازمندی‌های مشتری داشته باشد. این دانش درست می‌تواند منجر به تدوین استراتژی‌های سودآور گردد. بنابراین سازمان‌ها به دنبال راهی برای شناسایی هرچه بهتر مشتری و درک نیازهای وی می‌باشند. بدین منظور ابزاری قوی، فراتر از بازیابی ساده اطلاعات، جهت تحلیل و بررسی موقعیت مشتری مورد نیاز است. در سال‌های اخیر فناوری‌های تولید و گردآوری داده‌ها به سرعت در حال رشد بوده‌اند اما امروزه مساله پیش‌روی سازمان‌ها، به ویژه در حوزه تصمیم‌گیری‌های مالی، از جمع‌آوری داده‌ها به دستیابی به توان استخراج دانش مفید نهفته در داده‌ها، تغییر سو داده‌است. بسیاری از سازمان‌ها حجم عظیمی از داده‌های ارزشمند در رابطه با مشتریان جاری، مشتریان بالقوه، تأمین‌کنندگان و شرکای تجاری نگهداری می‌کنند، هرچند عدم توانایی سازمان‌ها در به‌کارگیری تکنیک‌های مناسب، آنها را از کشف دانش ارزشمند موجود در این انبوه داده باز می‌دارد. در چنین شرایطی است که باید از رشد فناوری برای استفاده موثر از این دانش بالقوه سود جست (کیم و دیگران، ۲۰۰۶). روش‌ها و ابزارهای داده‌کاوی^۱ تحقق این منظور را ممکن می‌سازند. در واقع، می‌توان مدیریت ارتباط با مشتری و بازاریابی را جزو اولین حوزه‌های کاربرد داده‌کاوی دانست. ابزارهای داده‌کاوی به سؤالاتی پاسخ می‌دهند که در گذشته دستیابی به پاسخ آنها اگر نه غیرممکن، بسیار زمان‌بر بوده‌است. در جدول ۱ مثالی ساده از تفاوت رویکردهای ساده بازیابی اطلاعات و رویکرد مبتنی بر داده‌کاوی آورده شده‌است.

جدول ۱- مثالی از مقایسه رویکردهای سنتی و داده‌کاوی

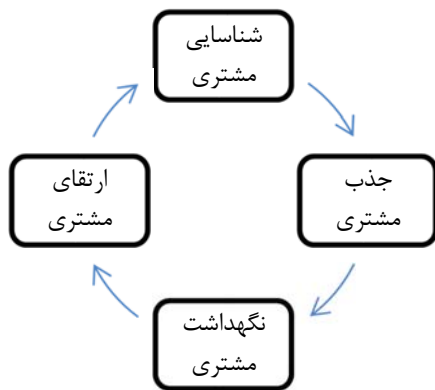
ابزار	سوال	رویکرد
کامپیوترها، دیسکها	میانگین درآمد کلی سازمان در پنج سال گذشته به چه میزان بوده است؟	جمع‌آوری داده
پایگاه داده‌های رابطه‌ای، SQL	فروشگاه‌های پرفروش ایران در بهمن ماه گذشته کدام بوده‌اند؟	دسترسی به داده
الگوریتم‌های پیشرفته، کامپیوترهای چندپردازشی و پایگاه داده‌های بزرگ	در روند فروش فروشگاه‌های شهر تهران در ماه آینده چه چیزی اتفاق می‌افتد؟ چه کالاهایی با یکدیگر خریداری میشوند؟	داده‌کاوی

3. Data Mining

در این مقاله، به بررسی روش‌های مختلف داده‌کاوی با رویکردی بر کاربرد آنها در ابعاد مختلف بازاریابی و مدیریت ارتباط با مشتری پرداخته میشود و همچنین چند مورد از مطالعات موردی در کاربردهای واقعی در ایران ارائه میگردد و نتایج اعمال روش‌های داده‌کاوی در آنها و تحلیل مربوطه بیان می‌شود. ادامه مقاله بدین صورت سازمان‌دهی شده‌است: در بخش ۲ به معرفی مدیریت ارتباط با مشتری و ابعاد آن پرداخته میشود. بخش ۳ مفهوم داده‌کاوی، فازهای یک پروژه مبتنی بر داده‌کاوی و مدل‌های کاربرد داده‌کاوی را توصیف میکند. در بخش ۴ کاربرد داده‌کاوی در مدیریت ارتباط با مشتری در صنایع مختلف بررسی شده و بخش ۵ چند مورد از این کاربردها را به صورت مطالعه موردی در ایران، مطرح و تحلیل میکند.

۲. مدیریت ارتباط با مشتری

مدیریت ارتباط با مشتری یک استراتژی کسب و کار است که هدف اصلی از ایجاد آن در سازمان‌ها، شناسایی مشتریان سودآور و برقراری ارتباط مناسب برای نگهداشت آنان می‌باشد. همچنین مدیریت ارتباط با مشتری، هزینه‌های مربوط به بازاریابی را کاهش داده و باعث بهبود در برنامه‌های وفاداری و نگهداشت مشتریان می‌شود. مدیریت ارتباط با مشتری دارای چهار بعد شناسایی، جذب، نگهداشت و ارتقای مشتری می‌باشد (ناگی و دیگران، ۲۰۰۹). این چهار مرحله، همان‌طور که در شکل ۱ نشان داده شده است، می‌توانند به عنوان یک چرخه برای سیستم مدیریت ارتباط با مشتری در نظر گرفته شوند و درک عمیق‌تر مشتریان برای حداکثرسازی ارزش مشتری برای سازمان در دراز مدت را ممکن می‌سازد. بنابراین تکنیک‌های داده‌کاوی با استخراج یا تشخیص ویژگی‌ها و رفتارهای پنهان مشتری از پایگاه داده‌های بزرگ به دستیابی به این اهداف کمک فراوانی می‌کنند. در ادامه، چهار بعد بیان شده مرور می‌شوند.



شکل ۱- چرخه مدیریت ارتباط با مشتری (ناگی و دیگران، ۲۰۰۹)

۲-۱ شناسایی مشتریان

مدیریت ارتباط با مشتری با شناسایی مشتریان آغاز می‌گردد. این مرحله شامل هدف قرار دادن مشتریانی است که بیشترین سود را برای سازمان به ارمغان می‌آورند. همچنین شناسایی افرادی که به طور بالقوه احتمال دارد

در آینده مشتری سازمان شوند (کراکلاویر و دیگران، ۲۰۰۴). شناسایی مشتریان شامل تحلیل مشتریان هدف و بخش بندی مشتریان است. تحلیل مشتریان هدف، شامل جستجوی بخش سودآور مشتریان از طریق تحلیل ویژگی‌های اصولی آنان می‌باشد، درحالی‌که بخش بندی، مشتریان را به زیربخش-هایی که شامل ویژگی‌های مشترکی هستند، طبقه بندی می‌کند.

۲-۲ جذب مشتریان

پس از شناسایی بخش مشتریان بالقوه، سازمان‌ها می‌توانند تلاش‌ها و منابع را در راستای جذب مشتریان هدف هدایت کنند. یکی از عناصر جذب مشتریان بازاریابی مستقیم است. بازاریابی مستقیم به ترغیب مشتریان از طرق مختلف جهت سفارش می‌پردازد. نامه الکترونیکی مستقیم و توزیع بن تخفیف از انواع بازاریابی مستقیم به شمار می‌روند (ناگی و دیگران، ۲۰۰۹).

۳-۲ نگهداشت مشتری

می‌توان این مرحله را دغدغه اصلی مدیریت ارتباط با مشتری دانست. برآورده کردن انتظارات مشتری در راستای کسب رضایت وی از شروط اصلی نگهداشت مشتری به شمار می‌رود. عناصر اصلی این بخش بازاریابی یک به یک، برنامه‌های وفاداری و مدیریت شکایات را شامل می‌شود. بازاریابی یک به یک به دنبال کشف تغییرات در رفتار مشتری و تغییر استراتژی‌ها، مبتنی بر این رفتار است. برنامه‌های وفاداری، پشتیبانی مشتری را به گونه‌ای که منجر به یک رابطه درازمدت با مشتری گردد، دربرمی‌گیرد. تحلیل روی گردانی، رضایت مندی و کیفیت خدمت بخش‌هایی از برنامه‌های وفاداری به شمار می‌رود (ناگی و دیگران، ۲۰۰۹).

۴-۲ ارتقای مشتری

این مرحله به توسعه روابط با مشتریان شامل بالابردن ارزش تراکنش‌ها و سودآوری مشتری می‌پردازد. این وظایف با تحلیل ارزش طول عمر مشتری، فروش جانبی/فروش بیشتر و تحلیل سبد خرید مشتری محقق می‌شود. تحلیل ارزش طول عمر مشتری، پیش-بینی درآمدی است که از مشتری انتظار می‌رود. فروش جانبی/بیشتر با پیشنهاد خدماتی که با یکدیگر در ارتباط هستند، ارزش مشتری را برای سازمان بالا می‌برد. تحلیل سبد خرید مشتری به دنبال الگوهای متداول در سبد خرید مشتریان است تا تعامل و ارزش مشتریان را حداکثر سازد (ناگی و دیگران، ۲۰۰۹).

۳- داده‌کاوی

داده‌کاوی به معنای استخراج دانش از حجم عظیمی از داده‌ها و همچنین کشف الگوهای پنهان در میان انبوهی از داده‌ها می‌باشد (غضنفری و دیگران، ۲۰۰۸). این رویکرد امکان استخراج روابط معنایی و عمقی را از انبارهای داده فراهم می‌نماید (کراکلاویر و دیگران، ۲۰۰۴).

وقتی حجم عظیمی از داده وجود داشته باشد، داده‌کاوی مفهوم ملموس‌تری می‌یابد. درحقیقت بیشتر الگوریتم‌های داده‌کاوی نیاز به حجم بالایی از داده جهت ساخت و آموزش مدل دارند. هرکدام از روش‌های داده‌کاوی یک یا تعداد بیشتری از انواع مدل‌سازی داده را انجام می‌دهند. انواع مدل‌سازی عبارتند از: پیوستگی^۴،

دسته‌بندی^۵، خوشه‌بندی^۶، پیش‌بینی^۷، رگرسیون^۸، تشخیص توالی^۹ و مصورسازی^{۱۰} که در بخش ۳-۱ مرور می‌شوند. برای هرکدام از مدل‌های بیان شده، روش‌های یادگیری ماشین متعددی وجود دارد که انتخاب آنها بستگی به ویژگی‌های داده و نیاز کسب و کار دارد. برخی از روش‌های پرکاربرد یادگیری ماشین در داده‌کاوی عبارتند از: قواعد تلازمی^{۱۱}، درخت تصمیم^{۱۲}، شبکه عصبی مصنوعی^{۱۳}، K-میانگین^{۱۴}، K-نزدیک‌ترین همسایه^{۱۵}، رگرسیون خطی/لجستیک^{۱۶} و... الگوریتم‌های داده‌کاوی به دو دسته بانظارت و بدون نظارت تقسیم‌بندی می‌شوند که الگوریتم‌های بانظارت با استفاده از داده‌های آموزش که دارای برچسب می‌باشند، مدلی را شکل می‌دهند که در ادامه برای کشف برچسب داده‌های بدون دسته می‌توان از آن بهره برد. در مقابل الگوریتم‌های بدون نظارت داده‌ها را بدون پیش فرض تعیین شده و تنها با توجه به ویژگی‌های آن در چند خوشه جای می‌دهند.

از کاربردهای اولیه و مهم داده‌کاوی، استفاده از آن در مدیریت ارتباط با مشتری است که در آن داده‌کاوی به عنوان یک ابزار بنیادی برای آشکارسازی خصوصیات جمعیت شناختی مشتریان محسوب می‌شود. در بخش ۴ به مرور دقیق‌ترین مساله پرداخته می‌شود.

۱-۳ مدل‌های داده‌کاوی

هرکدام از عناصر بیان شده در بخش ۲ برای مدیریت ارتباط با مشتری می‌توانند با یک یا چند روش داده‌کاوی مدل‌سازی شوند. عمده‌ترین مدل‌هایی که به کار گرفته می‌شوند در ادامه به صورت خلاصه مرور شده‌اند.

۱-۱-۳ پیوستگی

قوانین تلازمی به فرایند کشف وابستگی‌ها میان حجم عظیمی از داده‌ها، و نمایش آنها به فرم عبارت‌های "اگر-آنگاه" اطلاق می‌شود. هدف این قوانین پیدا کردن بیشترین روابط علت و معلولی بین داده‌هاست. این قوانین، از داده و روابط منطقی حاکم بر آنها استنباط می‌شوند که طبعاً احتمالی هستند. پرکاربردترین الگوریتم‌های مورد استفاده در قواعد تلازمی AIS, SETM, Apriori, AprioriTid, AprioriHybrid می‌باشند (غضنفری و دیگران، ۲۰۰۸). هدف پیوستگی، برقراری ارتباط بین عناصری است که با یکدیگر در یک تراکنش گرد آمده‌اند. برنامه‌های فروش جانبی و فروش بیشتر، مثال‌های معمول از به کارگیری پیوستگی می‌باشند. برای نمونه، با استفاده از پیوستگی، سازمان‌های کاتالوگ محور، کاتالوگ‌های فروش محصولات خاص را به مشتریانی ارائه می‌دهند که احتمال خرید آن توسط شخص با توجه به محصولات خریداری شده، از یک مقدار آستانه بیشتر باشد (ناگی و دیگران، ۲۰۰۹).

5. Classification
6. Clustering
7. Prediction
8. Regression
9. Sequence mining
10. Visualization
11. Association rule
12. Decision tree
13. Neural network
14. K-means
15. K-nearest neighbor
16. Logistic/linear regression

۳-۱-۲ دسته بندی

هدف دسته بندی ایجاد مدلی برای پیش بینی رفتار مشتریان بر اساس طبقه بندی رکوردهای داده به دسته های از پیش تعیین شده است. دسته بندی شامل بررسی داده جدید (که دسته آن نامشخص است) و انتصاب آن به یکی از دسته های شناخته شده قبلی می باشد. یک راه برای تشخیص اینکه یک مشتری به یک دسته تعلق دارد یا خیر، محاسبه میزان شباهت آن با دسته مدنظر می-باشد. روش های رایج در این زمینه شامل شبکه عصبی، درخت تصمیم و قواعد اگر-آنگاه هستند. دسته بندی شامل الگوریتم های بانظارت می باشد.

۳-۱-۳ خوشه بندی

روش های خوشه بندی داده ها را به بخش های همگن تقسیم می کنند، به گونه ای که داده های موجود در یک خوشه دارای حداکثر شباهت با یکدیگر بوده و فاصله بین خوشه ها نیز حداکثر باشد. تفاوت خوشه بندی با دسته بندی در این است که خوشه ها دارای عنوان از پیش تعیین شده نمی باشند. بهترین راه برای بازبینی موقعیت مشتریان جاری و بالقوه، بخش بندی مشتریان است. امروزه مهم ترین متغیر جهت بخش بندی مشتریان رفتار خرید آنان است. رفتار خرید مشتری را می توان بر اساس سه اصل اولیه زیر بررسی کرد (شاو و دیگران، ۲۰۰۱)، که البته تعریف این سه متغیر بر اساس ماهیت صنعت می تواند متفاوت باشد.

• تازگی^{۱۷} (R) - به زمان آخرین خرید مشتری باز می گردد. از لحاظ تئوری، مشتریانی که اخیراً به جمع مشتریان سازمان پیوسته اند و از لحاظ زمانی به تازگی خرید کرده اند، مشتریان مناسب تری بوده و بیشتر مستعد پاسخ به پیشنهادات ارائه شده از سوی سازمان می باشند.

• تکرار^{۱۸} (F) - تعداد دفعاتی است که مشتری در یک دوره زمانی مشخص خرید می کند. به راحتی می توان گفت خریداران با تکرار بیشتر مشتریان راضی تری به شمار می روند.

• مقدار پولی^{۱۹} (M) - میزان هزینه ای که مشتری در یک بازه زمانی مشخص پرداخت می کند، را بیان می دارد.

مدل فوق را RFM می نامند. در برخی تحقیقات، نویسندگان از مدل RFM وزن دار^{۲۰} برای محاسبه ارزش طول عمر مشتری بهره می برند. آنها به متغیرهای R و F و M مقادیر متفاوتی به عنوان وزن نسبت می دهند که این وزن ها نمایانگر شدت اهمیت و تأثیر آن متغیر در محاسبه ارزش طول عمر مشتری می باشد. بر اساس ماهیت صنعت، وزن های متفاوتی به پارامترهای RFM اختصاص می یابد. برای مثال، برخی پیشنهاد می کنند که بالاترین وزن به تکرار، به دنبال آن تازگی و کمترین وزن به مقدار پولی تعلق یابد. اما در مقدار پول بیشترین وزن را داشته و تازگی کمترین آن را دارا می باشد (چن و دیگران، ۲۰۰۵). روش فرآیند سلسله مراتبی تحلیل^{۲۱} برای تعیین درجه اهمیت مربوط به متغیرهای RFM به کار می رود.

از طرفی، نباید همه مشتریان با اقدامات فروش و بازاریابی یکسانی هدف بخش بازاریابی سازمان قرار بگیرند. بخش بندی به سازمان کمک می کند تا مشتریان با بیشترین ارزش، پتانسیل های فروش آتی، مشتریان غیر سودآور و علت عدم سودآوری آنها را شناسایی کند. خوشه بندی از روش های بدون ناظر به شمار می رود.

17. Recency
18. Frequency
19. Monetary
20. Weighted RFM
21. Analytic hierarchical processing (AHP)

۳-۱-۴ پیش‌بینی

روش‌های پیش‌بینی، مقادیر آینده را بر اساس الگوهای کشف شده تاکنون تخمین می‌زنند. پیش‌بینی نیاز مشتریان یک مثال معمول از یک مدل پیش‌بینی به شمار می‌رود. روش‌های رایج پیش‌بینی شامل شبکه عصبی و سری‌های زمانی می‌باشد. داده‌کاوی معمولاً با آنچه در طول زمان اتفاق می‌افتد، گره خورده است. تحلیل سری زمانی نیاز به انتخاب فریم زمانی مناسب برای داده دارد و علاوه بر آن، مشخص کردن نقطه آغاز بررسی نیز حائز اهمیت است. مرحله بعد، رسم سری زمانی است چراکه نمودار سری زمانی اطلاعات مفید مختلفی را فراهم می‌کند، هرچند ایده‌ای در رابطه با اینکه آیا تغییرات در طول زمانی مورد انتظار بوده یا نبوده است نمی‌دهد. برای این منظور و تحلیل دقیق‌تر، نیاز به روش‌های بررسی آماری است (غضنفری و دیگران، ۲۰۰۸).

۳-۱-۵ رگرسیون

رگرسیون نوعی روش آماری جهت نگاهت هر داده به یک مقدار واقعی است تا بتوان مقدار یک ویژگی را بر اساس مقدار سایر ویژگی‌ها تخمین زد. کاربردهای رگرسیون شامل برازش منحنی، پیش‌بینی، مدل‌سازی روابط علی و آزمایش فرضیات علمی در رابطه با روابط بین متغیرهاست. روش‌های معمول رگرسیون، رگرسیون خطی و رگرسیون لجستیک می‌باشد.

۳-۱-۶ تشخیص توالی

تشخیص توالی، شناسایی پیوستگی‌ها یا الگوها در طول زمان است. هدف این رویکرد، مدل‌سازی فرآیند ایجاد کننده دنباله و یا استخراج و گزارش انحراف و روند در طول زمان می‌باشد. ابزارهای معمول برای کشف توالی آمار، نظریه مجموعه‌ها و الگوریتم‌های کشف الگوهای متوالی می‌باشد (ناگی، ۲۰۰۹).

۳-۱-۷ مصورسازی

هدف روش‌های مصورسازی، نمایش مناسب داده‌ها به گونه‌ای است که الگوهای پیچیده درون آن استخراج شود. این روش‌ها در ترکیب با سایر مدل‌ها جهت شفاف‌سازی الگوی استخراج شده یا روابط به کار می‌رود. بررسی داده‌ها با استفاده از ابزارهای مصورسازی به ایجاد درک اولیه از داده، جهت انتخاب متغیرها و روش مناسب داده‌کاوی کمک می‌کند.

۳-۲ فازهای داده‌کاوی در مدیریت ارتباط با مشتری

استفاده از یک متدولوژی در بسیاری از پروژه‌های مدیریت ارتباط با مشتری متداول است. یکی از متدولوژی‌های قوی و رایج در این حوزه متدولوژی کریسپ^{۲۳} می‌باشد، که در زیر به بررسی گام‌های آن در یک پروژه داده‌کاوی در حوزه مدیریت ارتباط با مشتری می‌پردازیم.

۳. درک محیط کسب و کار: در این مرحله به شناخت کسب و کار مدنظر پرداخته می‌شود. وضع موجود سنجیده شده و سؤالاتی که داده‌کاوی باید به دنبال پاسخ آنها باشد، مطرح می‌گردد.
۴. شناخت داده‌ها: در این فاز داده‌های خام اولیه موجود در محیط کسب و کار جمع‌آوری و بررسی می‌گردد. در این مرحله تصمیم گرفته می‌شود که مدل‌سازی بر روی کل داده‌ها یا فقط داده‌های نمونه صورت گیرد. شناسایی منابع داده مناسب و جمع‌آوری آنها یکی از فاکتورهای کلیدی موفقیت به شمار می‌رود.
۵. انتخاب و پیش‌پردازش داده‌ها: هدف این مرحله استخراج داده‌های مرتبط برای تحلیل از انباره داده بزرگ‌تر است.

این فاز به پاک‌سازی داده‌ها و آماده‌سازی آن‌ها جهت اطمینان از حصول نتایج درست می‌پردازد. حذف داده‌های مفقود، اطمینان از این‌که مقادیر دارای معنای یکسانی هستند در این مجموعه قرار می‌گیرد. برای مثال یکسان بودن واحدهای پولی و نرمال‌سازی جهت اطمینان از عدم همگراشدن نتایج به متغیرهایی با مقادیر بالا، از کارهای این مرحله است.

۶. مدل‌سازی: با توجه به اهداف مدنظر باید روش یا روش‌های مناسبی جهت یافتن پاسخ سوالات مطرح بکار رود. مدل‌سازی با استفاده از تکنیک‌های داده‌کاوی موجود انجام می‌پذیرد. انواع تکنیک‌های داده‌کاوی بر اساس خروجی‌هایی که تولید می‌کنند دسته‌بندی می‌شوند. چرخه ادامه می‌یابد تا مدلی قابل قبول ارائه گردد. فرآیند انتخاب تکنیک داده‌کاوی بستگی به ماهیت داده و هدف کسب و کار دارد.

۷. ارزیابی نتایج: پس از مدل‌سازی باید بر اساس معیارهای تعیین شده به ارزیابی مدل پرداخته و اقداماتی جهت تغییر و یا بهبود مدل ساخته شده انجام پذیرد.

۸. بهبود: مدل بر اساس اقدامات اصلاحی مراحل قبل توسعه داده شده و بهبود می‌یابد. سپس، نتایج مجدداً مورد ارزیابی قرار می‌گیرد و فرایند بهبود تا رسیدن به پاسخ مطلوب ادامه می‌یابد.

۴- کاربردهای داده‌کاوی در مدیریت ارتباط با مشتری

در زمینه مدیریت ارتباط با مشتری و بازاریابی، داده‌کاوی را می‌توان به عنوان فرآیندی جهت کشف و استفاده مداوم از دانش سودآور حاصل از داده‌های سازمان در نظر گرفت. همچنین داده‌کاوی می‌تواند برای جهت‌دهی درست به تصمیم‌گیری و پیش‌بینی تأثیر تصمیمات اتخاذ شده، به کار گرفته شود. برای مثال، داده‌کاوی سرعت پاسخ سازمان به درخواست مشتریان را با دسته‌بندی آنان به گروه‌ها با ویژگی‌های مختلف افزایش می‌دهد. همچنین می‌تواند احتمال مهاجرت مشتری جاری به سازمان رقیب را محاسبه کند. داده‌کاوی بالاخص برای سازمان‌هایی حائز اهمیت است که هزینه تغییر^{۲۳} تأمین‌کننده برای مشتریان آن بسیار کم است، بنابراین سازمان باید برنامه‌های نگهداشت مشتری را ارتقا دهد. با دانستن این موضوع که چه کسی و به چه علتی و حتی در چه زمانی می‌خواهد سازمان را ترک کند، باید یک برنامه نگهداشت برای مشتریان ارزشمند طرح‌ریزی شود.

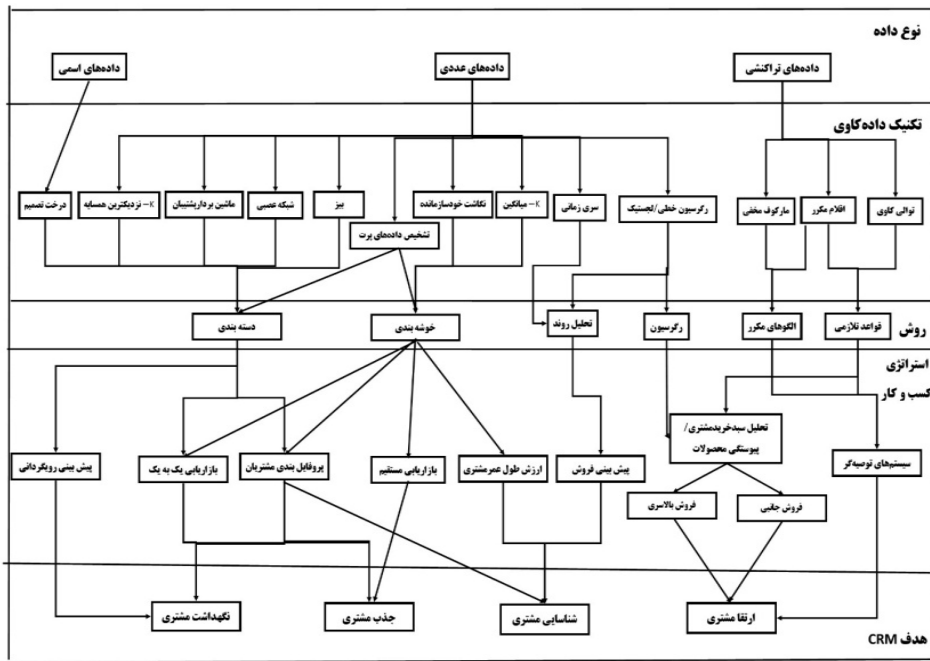
با بررسی تقاضا و رفتار مشتریان می‌توان قوانینی استخراج کرد که خود به عنوان برنامه فروش، استراتژی‌های جدید برای واحد بازاریابی، پیش‌بینی فروش برای سال جدید کارایی فراوانی داشته باشد. فعالیت‌های یک سازمان باید شامل تحویل محصول درست، به مشتری هدف، از طریق کانال درست، در زمان مناسب و با قیمت مناسب باشد. معمولاً ترکیبی از مدل‌های داده‌کاوی برای پشتیبانی و پیش‌بینی تأثیر استراتژی‌های مدیریت ارتباط با مشتری مورد نیاز است. بعد از تحلیل و ارزیابی کامل، اقدامات سازمان ممکن است پاسخ به درخواست‌های ویژه مشتریان، پیش‌بینی تشدید فروش، برخی برنامه‌های بازاریابی جدید، استراتژی‌های جدید برای تبلیغات و دستورالعمل‌هایی برای کارمندان بخش‌های مختلف را شامل شود (راداگریشان و دیگران، ۲۰۱۳).

در مواجهه با مسائل واقعی آنچه قابل مشاهده است تنها داده‌های خام می‌باشد. داده‌کاوی تعیین می‌کند با استفاده از کدام روش از مجموعه روش‌های موجود می‌توان به استراتژی‌های مفید کسب و کار و در نهایت به یکی از اهداف کلیدی مورد نظر دست یافت. شکل ۲ چگونگی به کارگیری روش‌های داده‌کاوی را جهت پاسخگویی به سوالات کسب و کار نشان می‌دهد. همان‌طور که در این شکل پیداست، نوع داده یا ویژگی می‌تواند عددی یا اسمی باشد. ویژگی‌های عددی اکثر خواص اعداد را دارا می‌باشند مانند سن، جرم، طول و ویژگی‌های اسمی

دارای خواص توصیفی هستند مانند رنگ، کد، جنسیت و... داده‌های تراکنشی در واقع چیزی غیر از داده‌هایی مانند فاکتور خرید مشتریان نمی‌باشد. به دلیل کاربرد فراوان این نوع خاص داده، در این بخش به طور جداگانه در نظر گرفته شده است.

تکنیک‌های داده‌کاوی بیانگر الگوریتم‌های متعددی است که هرکدام از آنها زیرمجموعه دسته‌ای از روش‌های داده‌کاوی بیان شده در بخش قبل می‌باشند. هر کدام از روش‌ها هدف خاصی را دنبال می‌کنند، برای مثال خوشه‌بندی داده‌ها را به چند خوشه متمایز تقسیم می‌کند و دسته بندی پس از مرحله آموزش مدل، به پیش‌بینی دسته داده‌های بدون برچسب می‌پردازد.

قواعد و نتایج حاصل از این روش‌ها منجر به کشف یک استراتژی کسب و کار می‌گردد. که این استراتژی‌ها یکی از اهداف CRM را تحقق می‌بخشند. برای مثال با اعمال تکنیک‌های توالی کاوی و یا کشف اقلام مکرر از سری روش‌های قواعد تلازمی، بر روی داده‌های تراکنشی، می‌توان به تحلیل پیوستگی محصولات پرداخته و با فروش جانبی به هدف ارتقا مشتری در سازمان دست یافت. با تکنیک‌های خوشه‌بندی یا دسته‌بندی می‌توان مشتریان را پروفایل بندی کرده و برای هر خوشه با توجه به ویژگی‌های آن استراتژی‌های خاص را مدنظر قرار داد و به مزایای نگهداشت مشتری در سازمان دست یافت. همچنین مشتریانی که پتانسیل مشتری دائم سازمان شدن را دارند، شناسایی کرده و با استراتژی‌های بازاریابی یک به یک به جذب آنها پرداخت. همچنین، می‌توان با تکنیک‌های مختلف بازاریابی، برای نگهداشت مشتریان ارزشمند اقدام نمود.



شکل ۲- بکارگیری روشهای داده‌کاوی برای کشف استراتژی‌های کسب و کار در راستای اهداف CRM

همه کسب وکارها علاقه‌مند به پیش‌بینی رفتار مشتریانش از طریق دانش کسب شده از طریق داده‌کاوی است. در ادامه به بررسی چند مورد از کاربرد داده‌کاوی در مدیریت ارتباط با مشتری می‌پردازیم.

۱-۴ خرده‌فروشی

با استفاده از سیستم پایانه فروش، خرده‌فروشان می‌توانند جزئیات تراکنش‌های هر خرید را ذخیره‌سازی کنند. این امر به درک بهتر طبقه‌های متنوع مشتری کمک می‌کند. برخی از کاربردهای داده‌کاوی در خرده‌فروشی که با پردازش‌های داده‌های تراکنش‌ها حاصل می‌شود، در ادامه آمده است.

- شناسایی مشتریان سودآور و پروفایل کردن آنان: گاهی هدف داده‌کاوی به سادگی توصیف آنچه در پایگاه داده‌های پیچیده رخ می‌دهد می‌باشد، به گونه‌ای که درک از محصولات، افراد یا فرآیندهایی که داده را تولید کرده‌اند بالا ببرد. برای مثال شناسایی ۲۰٪ از مشتریان که سودآور هستند. خوشه‌بندی ابزاری قوی جهت بخش‌بندی مشتریان سودآور و غیرسودآور است. خرده‌فروشان می‌توانند پروفایل مشتریان با رفتار خاص را شناسایی کنند و استراتژی‌های بازاریابی و ارتقا بر اساس پروفایل‌های مدنظر اعمال گردد. برای مثال کسانی که لباس‌های با برچسب طراحان معروف را خریداری می‌کنند، یا کسانی که در حراج‌ها شرکت می‌کنند (رادا کریشنان و دیگران، ۲۰۱۳).

- تحلیل سبد خرید مشتری: این تحلیل به کشف اقلامی که با یکدیگر خریداری می‌شوند می‌انجامد و دانش حاصل از این تحلیل به بهبود انبارداری، استراتژی‌های چیدمان فروشگاه و برنامه‌های ارتقا مشتری کمک می‌کند. دانستن محصولاتی که مشتریان به طور معمول با یکدیگر خریداری می‌کنند، می‌تواند منجر به افزایش سودآوری سازمان از طریق فروش جانبی/ بیشتر شود. عمده‌ترین کاربرد قواعد تلازمی در این بخش متبلور می‌گردد.

- پیش‌بینی فروش: ارزیابی الگوهای مبتنی بر زمان به خرده‌فروشان جهت گرفتن تصمیمات انبارداری بهینه یاری می‌رساند. با استفاده از ابزارهای تحلیل روند مانند تحلیل سری‌های زمانی و رگرسیون می‌توان رفتار آینده مشتری و میزان خرید هر یک از محصولات را پیش‌بینی کرد. همچنین ابزارهای کاوش توالی می‌توانند با توصیه محصولات مورد نیاز آتی به مشتری، سازمان را در راستای هدف نگهداشت یاری رسانند.

- تخمین کارآمدی تبلیغات: با استفاده از بررسی اطلاعات گذشته رفتار مشتریان می‌توان رفتار آینده آنها را در برابر تبلیغات با استفاده از ابزارهای دسته‌بندی مانند شبکه عصبی و یا درخت تصمیم تخمین زد. در اینجا هدف پاسخ به این سوال است که آیا یک گروه از مشتریان تحت تأثیر تبلیغات قرار خواهند گرفت یا خیر.

۲-۴ بانکداری

بانک‌ها می‌توانند از کشف دانش حاکم بر داده‌ها برای کاربردهای متنوعی بهره‌برند، برخی از این کاربردها شامل موارد زیر است:

- بازاریابی: با شناسایی بخش مشتریان سودآور، بانک‌ها می‌توانند این بخش را با برنامه‌های جذب و نگهداشت، توسعه محصول هدف و قیمت‌گذاری اختصاصی مانند پرداخت وام‌های مناسب بهبود بخشند. ابزارهای خوشه‌بندی و همچنین پیش‌بینی برای شناسایی مشتریان خوش حساب و بدحساب در این بخش مورد استفاده قرار می‌گیرند.

- تشخیص کلاهبرداری‌ها: فریب‌ها برای بانک‌ها به طور فزاینده‌ای هزینه‌بر هستند اما با تحلیل تراکنش‌های

گذشته که بعداً به عنوان کلاه برداری شناخته شده است، بانک‌ها می‌توانند الگوهای مرتبط را استخراج نموده و در آینده از تکرار آن جلوگیری نمایند. ابزارهای کارآمد در این بخش، روش‌های پیش‌بینی و تحلیل روند هستند.

۳-۴ صنعت مخابرات

شرکت‌های مخابراتی در سرتاسر جهان با رقابت فزاینده‌ای مواجه‌اند، که این رقابت آنها را وادار به اعمال استراتژی‌های مختلف قیمت‌گذاری با هدف نگهداشت مشتریان جاری و جذب مشتریان جدید می‌کند. کشف دانش در صنعت مخابرات با روش‌های داده‌کاوی شامل موارد زیر است:

• تحلیل جزئیات رکوردهای تماس: صنایع مخابراتی رکوردهای تماس را ذخیره‌سازی می‌کنند که با شناسایی بخش‌های مشتریان با الگوهای استفاده مشابه، شرکت‌ها می‌توانند قیمت‌گذاری و ارتقا خاصی برای یک بخش ویژه ارائه دهند.

• وفاداری مشتری: برخی از مشتریان بطور مداوم بین فراهم‌کنندگان خدمت مختلف جابه‌جا می‌شوند، تا از مزایای استراتژی‌های سازمان رقیب بهره ببرند. شرکت‌ها می‌توانند با استفاده از داده‌کاوی ویژگی‌های مشتریان وفادار و روی‌گردان را استخراج کنند تا با استفاده از پیشنهادات مناسب احتمال روی‌گردانی را کاهش دهند. تکنیک‌های پیش‌بینی و دسته‌بندی در این زمینه کاربرد فراوان دارد.

• هدفمندسازی تبلیغات: عمده‌ترین رویکردها جهت شناسایی مخاطبان اصلی تبلیغات که به پیشنهادها ارائه شده پاسخ مثبت می‌دهند، پروفایل‌بندی و دسته‌بندی مشتریان می‌باشد. این دسته‌بندی می‌تواند بر مبنای گوناگونی بنا نهاده شود، از جمله

o دسته‌بندی براساس ویژگی‌های دموگرافیک مانند سن و جنسیت جهت ارائه خدمات خاص، برای مثال به نوجوانان.

o دسته‌بندی براساس نوع استفاده مانند طول مدت تماس و یا تعداد پیامک‌های ارسال شده، جهت ارائه پیشنهادهایی مانند خدمات مکالمه رایگان یا بسته‌های پیام کوتاه جهت نگهداشت مشتری

• بررسی تأثیر خدمات ارائه شده و همچنین تأثیر متقابل خدمات گوناگون: خدمات ارائه شده به چه میزان بر مصرف مشتری اثرگذار بوده‌اند. در حقیقت می‌توان میزان سودآوری یک خدمت برای سازمان را بررسی نمود و آیا ارائه یک خدمت تأثیری بر ارائه دیگر خدمات خواهد داشت.

• شناسایی فریب‌ها: همانند صنعت بانکداری، صنعت مخابرات نیز مورد تهاجم کلاهبرداری‌های فراوانی است. با تشخیص الگوی متداول رفتار مشتریان می‌توان تغییرناگهانی در مصرف را شناسایی کرد. همچنین مشتریانی که قصد پرداخت ندارند را پیش‌بینی نمود.

• پیش‌بینی خرابی شبکه: به کمک تکنیک‌های تشخیص رویداد و یا پیش‌بینی، می‌توان زمان دقیق افزایش بارکاری و حجم ترافیک را پیش‌بینی نمود. و اقدامات لازم جهت پیشگیری از نابسامانی شبکه را اتخاذ کرد. همچنین همزمانی وقوع رخدادها را مورد بررسی قرار داد.

بدیهی است که داده‌کاوی محدود به کاربردهای اشاره شده در فوق نیست و در صنایع دیگری چون صنعت بیمه برای پیش‌بینی میزان وفاداری مشتری (گیولین و دیگران، مارین، ۲۰۱۲)، کارت‌های اعتباری جهت پیش‌بینی روی‌گردانی مشتریان (نای و دیگران، ۲۰۱۱) و... مورد استفاده قرار می‌گیرد. به صورت عمومی، داده‌کاوی درصنایعی که دارای حجم قابل توجهی داده خام هستند، می‌تواند به منظور پردازش و تبدیل آنها به

دانش و تصمیم مورد استفاده باشد.

۵- مطالعات موردی

همان طور که در بخش قبل مطرح شد، داده‌کاوی در صنایع مختلف و با اهداف متنوعی کاربرد دارد. در این بخش سه نمونه از این کاربردها به عنوان مطالعات موردی انجام شده در دو صنعت مختلف مورد تحلیل و بررسی قرار می‌گیرد و نتایج حاصل ارائه می‌شود. دو صنعت مورد مطالعه در این تحقیق، حوزه خرده‌فروشی با هدف پیش‌بینی تأثیر عنوان تجاری بر سبد خرید مشتری جهت دستیابی به استراتژی‌ها فروش بیشتر، و حوزه خدمات اینترنتی جهت پیش‌بینی روی‌گردانی مشتریان اینترنت هوشمند و پیش‌بینی خدمت‌آتی مورد تقاضای مشتریان خدمت اینترنت بی‌سیم می‌باشد.

۵-۱ کاربرد داده‌کاوی در صنعت خرده‌فروشی

۵-۱-۱ بررسی تأثیر عنوان تجاری بر سبد خرید مشتری

از جمله کارهای مهمی که داده‌کاوی در مدیریت ارتباط با مشتری انجام داده‌است، بررسی تأثیر نشان تجاری در سبد خرید مشتری می‌باشد. مزیت این کاوش در این است که در صورت تأثیر مستقیم عنوان تجاری، می‌توان چیدمان محصولات را جهت رفاه مشتریان براساس عنوان تجاری تغییر داد. همچنین با بسته‌بندی محصولات از یک عنوان تجاری خاص و ارائه محصول مازاد می‌توان به مزایای فروش جانبی دست یافت. در این بررسی به صورت کاربردی داده‌های فروشگاه شهروند (یکی از شعبه‌های تهران) مورد مطالعه قرار گرفت (علیزاده و دیگران، ۱۳۹۱).

• داده‌ها

پایگاه داده اولیه مورد استفاده در این مطالعه، به صورت فاکتور خرید مشتریان فروشگاه شهروند در یکی از شعب تهران طی یک مدت سه ماهه از تاریخ ۸۹/۶/۱ الی ۸۹/۸/۳۰ جمع‌آوری شده است که تعداد کل آن‌ها حدود ۴۳۰,۰۰۰ فاکتور حاوی ۵,۶۶۴,۹۷۶ رکورد می‌باشد. توصیف کالا به عنوان ورودی الگوریتم در نظر گرفته شد، این ویژگی شامل نام کالا، عنوان تجاری و سایر اطلاعات توصیفی می‌باشد. در اینجا توصیف کالا به عنوان یک عنصر برای اعمال الگوریتم در نظر گرفته می‌شود (علیزاده و دیگران، ۱۳۹۱) و نتیجه الگوریتم به صورت قوانینی به شکل $X \rightarrow Y$ خواهد بود که بیانگر این است که خرید کالای X خرید کالای Y را نتیجه می‌دهد.

برای ارزیابی نتایج حاصل، از دو معیار اطمینان^{۲۴} و پشتیبان^{۲۵} استفاده می‌شود. مقدار اطمینان برای قانون $Y \rightarrow X$ ، درصدی از تراکنش‌ها را نشان می‌دهد که مشتریانی که کالای X را خریده‌اند، کالای Y را نیز خریداری کرده‌اند. مقدار پشتیبان برای قانون $Y \rightarrow X$ درصدی از تراکنش‌ها را نشان می‌دهند که شامل هر دو X و Y باشند (چنگ و دیگران، ۱۹۹۶). این مقادیر درجه اعتبار قوانین را مشخص می‌کنند. به طور معمول یک مقدار حداقل برای پشتیبان و یک مقدار حداقل برای معیار اطمینان به عنوان آستانه تعریف می‌شود. قوانینی حائز اهمیت می‌باشند که دارای مقادیر اطمینان و پشتیبان بیش از مقدار آستانه باشند.

24. Confidence
25. Support

• روش داده کاوی

روش مورد استفاده برای کاوش داده‌های خرید مشتریان فروشگاه شهروند در یکی از شعب تهران الگوریتم Apriori است که در بخش ۳-۱-۱ از آن به عنوان یکی از روش‌های استخراج قواعد تلازمی نام برده شد. الگوریتم Apriori در سال ۱۹۹۶ توسط چئونگ مطرح شد (چئونگ و دیگران، ۱۹۹۶). ایده اصلی این الگوریتم این است که مجموعه‌ای از عناصر مکرر نمی‌توانند زیرمجموعه غیرمکرر داشته باشند. یعنی اگر عضوی به یک مجموعه با تعداد تکرار مشخص اضافه کنیم، تعداد تکرار آن افزایش نخواهد یافت. مراحل الگوریتم به شرح زیر است:

ابتدا تمامی مجموعه‌های تک عضوی بررسی می‌گردند، در این مرحله عناصر غیرمکرر حذف می‌شوند. عناصر مکرر باقیمانده با یکدیگر ترکیب شده و مجموعه‌های دو عضوی را می‌سازند. در مرحله بعد، مجموعه‌های دو عضوی مکرر یافت می‌شوند. این الگوریتم ادامه می‌یابد تا جایی که مجموعه‌های n -عضوی ایجاد شده مکرر نباشند. با توجه به اینکه در هر مرحله کاندیدها با اتصال اقلام پرتکرار حاصل از مراحل قبلی ایجاد می‌شود، تعداد کاندیدها جهت بررسی به طرز چشمگیری کاهش می‌یابد و نیازی به جستجوی کل فضای حالت در هر مرحله نمی‌باشد.

• نتایج ارزیابی

پس از اعمال الگوریتم Apriori بر روی دادگان بیان شده، قوانین تلازمی حاکم بر آنها حاصل شده است که برخی از این قوانین که با حداقل میزان پشتیبان ۵٪ و حداقل اطمینان ۱۰٪ بدست آمده، به شرح جدول ۲ است (علیزاده و دیگران، ۱۳۹۱) که در آن دو عدد مقابل قوانین، به ترتیب نشان دهنده میزان پشتیبان و اطمینان آن قانون است.

روشگاه شهروند

بیسکوئیت ساقه طلایی لیوانی مخصوص ۲۰۰ گرمی مینو قاسم ایران	←	ویفر رنگارنگ بسته‌ای مینو (۵۰۸ و ۱۸۰۲۵)
دستمال توال ۴ تایی چشمک پارس کرپ	←	دستمال کاغذی ۲۰۰ برگ ۳+۱ چشمک (۶۰۵۰ و ۱۶۰۳۵)
دستمال کاغذی ۲۰۰ برگ ۱+۳ چشمک	←	پودر ماشین لباسشویی ۵۰۰ گرمی ۴ آنزیم طلایی سافتلن (۵۰۵۲ و ۱۳۰۵۷)
پودر ماشین لباسشویی ۵۰۰ گرمی ۴ آنزیم طلایی سافتلن	←	دستمال کاغذی ۲۰۰ برگ ۱+۳ چشمک (۶۰۵۰ و ۱۳۰۸۲)
بیسکوئیت ساقه طلایی لیوانی مخصوص ۲۰۰ گرمی مینو قاسم ایران	←	پودر ماشین لباسشویی ۵۰۰ گرمی ۴ آنزیم طلایی سافتلن (۵۰۵۲ و ۱۳۰۵۷)
بیسکوئیت ساقه طلایی لیوانی مخصوص ۲۰۰ گرمی مینو قاسم ایران	←	شیر کم چرب پاستوریزه ۱/۵٪ چربی ۹۴۶ سی سی دامداران (۵۰۳۱ و ۱۳۰۳۳)
دستمال توال دو قلو چشمک	←	دستمال کاغذی ۲۰۰ برگ ۱+۳ چشمک (۶۰۵۰ و ۱۳۰۱۱)

در جدول ۲، قانون اول نشان می‌دهد ۵۰۸٪ از کل تراکنش‌ها شامل بیسکوئیت ساقه طلایی لیوانی مینو و ویفر رنگارنگ بسته‌ای مینو می‌باشند. همچنین ۱۸،۲۵۱ از مشتریانی که بیسکوئیت مینو را در لیست اقلام خود قرار داده‌اند، ویفر مینو را نیز خریداری کرده‌اند. همانطور که مشاهده می‌شود، از قوانین حاصل چنین برمی‌آید که محصولات عناوین تجاری مانند مینو و یا چشمک با یکدیگر خریداری می‌شوند اما این محصولات با سایر عناوین تجاری نیز در کنار یکدیگر می‌آیند. بنابراین قوانین حاصل نشان از تأثیر متوسط عنوان تجاری بر خرید مشتریان ایرانی دارد.

۵-۲ کاربرد داده‌کاوی در صنعت ارائه دهندگان خدمات اینترنت

۵-۲-۱ بررسی روی‌گردانی مشتری اینترنت هوشمند با استفاده از ارزش طول عمر مشتری

پیش‌بینی روی‌گردانی یکی از کاربردهای عمده داده‌کاوی در مدیریت ارتباط با مشتری می‌باشد. تشخیص فرد روی‌گردان و زمان روی‌گردانی از چالش‌های اصلی این مقوله به شمار می‌رود. از طریق بررسی روند سری زمانی ارزش مشتری، می‌توان به کشف تمایل یا عدم تمایل مشتری به ترک سازمان پرداخت. مشتری روی‌گردان بر اساس برخی تعاریف مشتری است که ارزش طول عمر وی در طول زمان کاهش یابد. با بررسی روند سری زمانی مصرف، در صورت نزولی بودن این روند، مشتری روی‌گردان شناخته می‌شود.

مورد مطالعه جهت اعتباربخشی این مدل بررسی روند مصرف مشتریان اینترنت هوشمند یکی از ارائه‌دهنده‌های خدمت اینترنت هوشمند در شهر شیراز می‌باشد. با استفاده از اینترنت هوشمند، کاربران اینترنت برخلاف روش‌های سنتی بدون نیاز به نام کاربری و رمزعبور در هر ساعت از شبانه روز می‌توانند به اینترنت متصل شوند و هزینه استفاده بر روی قبض‌های تلفن اضافه می‌شود.

• داده‌ها

پایگاه داده اولیه مربوط به این بررسی، به صورت گزارش اتصال کاربران به شبکه موجود بوده‌است که پس از پردازش آن، ویژگی‌های موردنیاز برای پیش‌بینی از آن استخراج گشت. در این بررسی از مدل RFM که در بخش ۳-۱-۳ بیان شد، جهت محاسبه ارزش طول عمر مشتری بهره برده شد. مدل RFM یکی از قدرتمندترین و در عین حال ساده‌ترین مدل‌ها برای پیاده‌سازی مدیریت ارتباط با مشتری و تخمین ارزش مشتری می‌باشد؛ با این تفاوت که در این بررسی RFM برای هر مشتری به صورت روزانه به دست آمد. محاسبه مقادیر سه پارامتر این مدل به شرح زیر است:

• مقدار پولی (Monetary): هزینه اینترنت مصرفی مشتری به صورت دقیقه‌ای محاسبه می‌شود. مقدار این متغیر از طریق رابطه (۱) محاسبه می‌گردد.

$$(1) \text{ مقدار پولی} = \text{میزان مصرف به دقیقه} \times \text{هزینه هر دقیقه مصرف}$$

• تکرار (Frequency): از آنجا که متغیرها به صورت روزانه محاسبه می‌شوند، فرکانس مصرف از شمارش دفعات اتصال مشتری در طول روز به دست آورده شده‌است.

• تازگی (Recency): تازگی مصرف مشتری فاصله آخرین اتصال مشتری تا اتصال امروز در نظر گرفته شده‌است. به عنوان مثال، عدد ۱ نشان می‌دهد که آخرین اتصال مشتری دیروز صورت گرفته است. هرچه این فاصله کمتر باشد احتمال تکرار اتصال از جانب مشتری بیشتر است.

مقدار این متغیرها باید نرمال شده و با استفاده از رابطه زیر ارزش طول عمر مشتری به طور روزانه محاسبه می‌گردد

$$(2) \text{ CLV} = \text{NR} \times \text{WR} + \text{NF} \times \text{WF} + \text{NM} \times \text{WM}$$

که در آن، NR به مقدار تازگی نرمال شده اشاره دارد، WR وزن متغیر تازگی است، NF به مقدار تکرار نرمال شده اشاره دارد، WF وزن متغیر تکرار است، NM بیانگر مقدار پولی نرمال شده است و WM وزن متغیر مقدار پول می‌باشد.

مقدار CLV در طول زمان سری زمانی مصرف مشتریان را شکل می‌دهد. هدف اصلی داده‌کاوی سری‌های

زمانی، کشف الگوهای موجود در وقایع و داده‌های آن می‌باشد. دستیابی به روند سری زمانی در این مطالعه توسط رگرسیون انجام گرفت. ویژگی‌های بدست آمده از نمودار مصرف مشتری، به علاوه ویژگی‌های دیگری مانند طول مدت مصرف کاربر، مرکز مخابراتی مربوط به شماره تلفن کاربر، نسبت روزهای فعال بودن کاربر به کل دوره مصرف، شبانه یا روزانه بودن اتصال مشتری و حداکثر دوره عدم اتصال مشتری در دوره فعالیت وی به همراه ویژگی دیگری به نام روی گردانی (که روی گردان بودن و یا عدم روی گردانی مشتری را نشان می‌دهد)، مجموعه‌ای از ویژگی‌های مشتری را شکل می‌دهد که این مجموعه می‌تواند به عنوان ورودی انواع مدل‌های پیش‌بینی مورد استفاده قرار گیرد. برای داده‌های آموزش، با توجه به نظر خبره، مشتریانی که در دو ماه آخر بازه زمانی هیچ‌گونه فعالیتی نداشته‌اند، ویژگی روی گردانی آن‌ها مثبت و برای سایر مشتریان منفی در نظر گرفته شد.

• روش داده‌کاوی

در این مطالعه از چهار الگوریتم درخت تصمیم و همچنین یک روش شبکه عصبی برای ساخت مدل جهت پیش‌بینی روی گردانی مشتری استفاده شد.

درخت تصمیم

درخت تصمیم ساختاری مشابه فلوجارت ارائه می‌دهد که در آن نودهای میانی یک تست بر روی یکی از ویژگی‌ها بوده و برگ‌ها یا نودهای پایانی برچسب کلاس هدف را نشان می‌دهند. برای ساخت درخت، معیاری جهت انتخاب ویژگی در نودهای میانی در نظر گرفته می‌شود. ویژگی منتخب باید بگونه‌ای باشد که عناصر را به بهترین وجه به کلاس‌های مجزا تقسیم کند (جیاوی و کامبر، ۲۰۰۱).

در این پژوهش چهار الگوریتم درخت تصمیم مورد استفاده قرار گرفته‌اند که عبارتند از: درخت دسته‌بندی و رگرسیون (C&R)، درخت CHAID، درخت QUEST و درخت C5. درخت دسته‌بندی و رگرسیون یک درخت دودویی را توصیف می‌کند. این الگوریتم و الگوریتم C5.1 یک رویکرد حریصانه را در راستای ایجاد درخت بکار می‌گیرند. در این الگوریتم‌ها درخت با روش تقسیم و غلبه به صورت بازگشتی ساخته می‌شود. بدین معنا که دادگان آموزش بطور بازگشتی به زیرمجموعه‌های کوچکتر تقسیم می‌گردند. الگوریتم درخت تصمیم CHAID که در بازاریابی کاربرد فراوان دارد، از معیار انتخاب ویژگی ای بهره می‌برد که اساس آن بر تست آماری مربع کای برای متغیرهای مستقل استوار است.

از الگوریتم‌های درخت تصمیم قوانینی حاصل می‌شود که با کمک آن می‌توان به صورت واضح به دلیل رخداد برخی نتایج دست یافت. در مقابل، روش شبکه عصبی که به صورت جعبه سیاه عمل کرده و تنها به ارائه نتیجه می‌پردازد.

شبکه عصبی

شبکه‌های عصبی یکی از مدل‌های ریاضی پردازش اطلاعات مبتنی بر عملکرد مغز انسان بوده و در مسائلی که روابط بین متغیرها شکل واضحی ندارد کاربرد فراوانی دارند. شبکه عصبی از تعدادی گره تشکیل شده که پردازش‌های ساده‌ای را انجام می‌دهند. این گره‌ها (نرون‌های عصبی) توسط یال‌هایی که نمایانگر وزن‌های شبکه هستند به یکدیگر متصل می‌شوند. این وزن‌ها حل مسئله توسط شبکه را امکان‌پذیر می‌سازند. در فاز آموزش، مدل مقداردهی و به روزرسانی می‌گردند تا وزن‌های بهینه برای مدل‌سازی مسئله بدست آید. از مدل

آموزش داده شده برای پیش بینی استفاده می‌گردد (فاسیت، ۲۰۱۳). در این مطالعه، از شبکه پرسپترون چندلایه^{۲۶} برای پیش بینی روی گردانی بهره برده شد. این شبکه سه لایه دارای ۱۷ گره ورودی معادل تعداد ویژگی‌ها، ۳ گره در لایه داخلی و ۱ گره خروجی برای تعیین روی گردانی یا عدم روی گردانی می‌باشد.

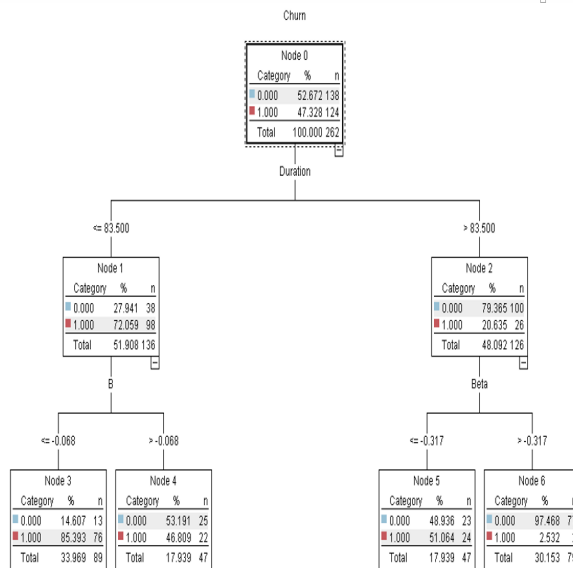
• نتایج ارزیابی

باتوجه به قوانین حاصل از درخت تصمیم، افرادی با ویژگی‌های پایین بودن مدت مصرف، منفی بودن شیب مصرف، پایین بودن نسبت روزهای فعالیت به کل روزهای اشتراک، برای روی گردانی مستعدتر بودند. همچنین قوانین نشان می‌دهد که ساکن بودن در دو مرکز مخابراتی که از سطح اقتصادی پایین تری نسبت به سایر نقاط شهر برخوردار بودند، به دلیل عدم نیاز اینترنت هوشمند به تهیه سخت افزار و زیرساخت، مشتریان را کمتر به سمت روی گردانی سوق می‌دهد.

بخشی از درخت تصمیم حاصل از اجرای الگوریتم تقسیم و رگرسیون (C&R) در شکل ۳ آمده است. متغییر هدف Churn در ریشه درخت ظاهر شده و پس از آن بر اساس اهمیت متغییرها، شکست ریشه به زیرشاخه‌ها صورت می‌گیرد. همچنین در هر گره درخت، درصد مشتریانی که با این ویژگی مصرف سازمان را ترک کرده‌اند، نشان داده شده است. در این شکل دو دسته (Category) از مشتریان روی گردان و غیرروی گردان داریم که به ترتیب با اعداد ۱ و ۰ مشخص شده‌اند، مثلاً در گره Churn درصد مشتریان که سازمان را ترک کرده‌اند، ۴۷،۳۲۸٪ و درصد گروه غیرروی گردان ۵۲،۶۷۲٪ است. در اینجا متغیر Duration که بیانگر کل مدت زمان استفاده مشتری از خدمات سازمان است، دارای بیشترین میزان کسب اطلاعات بوده و به عنوان اولین نقطه شکست در نظر گرفته می‌شود. بدین معنا که این ویژگی به بهترین وجه ممکن نمونه‌ها را به دسته‌های متمایز تقسیم می‌کند. همان‌طور که مشاهده می‌شود نزدیک به ۸۰٪ از مشتریانی که بیش از ۸۳ روز از خدمات سازمان استفاده کرده‌اند، نسبت به سازمان وفادار مانده‌اند. پس از آن متغیرهای شیب (B) و بتا که به ترتیب به معنی شیب نمودار ارزش طول عمر مشتری در طول زمان و میزان نوسانات مصرف مشتری هستند، به عنوان مهم‌ترین متغیرها جهت دسته بندی شناسایی شده‌اند.

از قوانین بدست آمده از درخت تصمیم می‌توان ویژگی‌های مشتریان روی گردان را شناسایی کرده و از روی گردانی آنان جلوگیری کرد. برای مثال، همان‌طور که در شکل ۳ نشان داده شده است، با توجه به قوانین به دست آمده از درخت C&R، اغلب مشتریانی که کمتر از ۸۳ روز (Duration) کاربر اینترنت هوشمند بوده‌اند و ضریب مصرف آنها (B) کمتر از ۰،۰۶۸- (به معنای نزولی بودن سیر مصرف مشتری در طی زمان) باشد، روی گردان خواهند شد. به همین ترتیب از سایر قوانین درخت می‌توان برای دستیابی به ویژگی‌های مشتریان وفادار و روی گردان بهره جست.

الگوهای بدست آمده از مهاجرت مشتریان در طول زمان می‌تواند دیدگاه و دانش مناسبی برای مدیران فراهم آورد تا از آن در جهت کنترل و هدایت مشتری به بخش‌هایی با ارزش بیشتر استفاده و یا حداقل از روی گردانی مشتری جلوگیری کنند (غروی، ۱۳۹۰).



۵-۲-۲ پیش‌بینی سرویس مورد تقاضای آینده مشتری

امروزه کسب و کار از پاسخ صرف به نیازهای مشتری به ایجاد نیازهای جدید در بازار تغییر رویه داده است. سازمان‌ها باید مشتریان را بشناسند و نیازهای آینده آنها را جهت برآورده ساختن آن تخمین بزنند تا بدین صورت از مهاجرتشان به سایر فراهم‌کنندگان خدمت پیشگیری کنند. در این مطالعه، از ویژگی‌های فعلی رفتار و مصرف مشتری برای پیش‌بینی خدمتی که در آینده مورد استفاده قرار خواهد داد، در مورد اینترنت بی‌سیم بهره برده شده است. نوع خدمت اینترنت بی‌سیم بر اساس سرعت و حجم ترافیک و طول مدت مصرف سنجیده می‌شود؛ در اینجا هر کدام از این ویژگی‌های خدمت بطور جداگانه مورد پیش‌بینی قرار گرفت.

• داده‌ها

پایگاه داده اولیه این مطالعه، گزارش اتصال ۱۸۰ کاربر اینترنت بی‌سیم در یک مرکز ارائه دهنده خدمت اینترنت بی‌سیم در شیراز می‌باشد. از این پایگاه داده، برخی ویژگی‌ها جهت پیش‌بینی استخراج شده‌اند که این ویژگی‌های شامل RFM، حجم اطلاعات ورودی، حجم اطلاعات خروجی، طول عمر مشتری، مدت زمان فعال بودن کاربر در حین دوره مصرف، شیب مصرف مشتری و ارزش طول عمر مشتری می‌باشد. در این بررسی، M میزان هزینه خدمت جاری، F تعداد دفعات تکرار اتصال توسط مشتری و R فاصله آخرین اتصال مشتری تا لحظه بررسی را نشان می‌دهد. تمام این مقادیر بصورت نرمال شده مورد استفاده قرار گرفته‌اند.

• روش داده‌کاوی

در این مطالعه، مدل پیش‌بینی مورد استفاده درخت تصمیم بوده که شرح الگوریتم‌های مختلف آن در بخش قبل آمده است. با استفاده از الگوریتم‌های درخت تصمیم با ترکیب متفاوتی از متغیرها، سرعت، حجم ترافیک و طول مدت مصرف سرویس آینده پیش‌بینی می‌گردد. برای این کار، ابتدا متغیرهای RFM به عنوان ورودی الگوریتم‌های مدنظر استفاده شده‌اند. سپس با اضافه کردن هریک از متغیرها، دقت خروجی مدل ارزیابی شد و

در نهایت کل ویژگی‌ها به ورودی داده شد و نتایج حاصل در ادامه آمده است.

• نتایج ارزیابی

مقدار دقت‌های حاصل از اجرای الگوریتم‌های CART و C5 بر روی ترکیب مختلف ویژگی‌ها به ترتیب در جدول‌های شماره ۲ و ۳ نشان داده شده است. از جمله قوانین حاصل از این نتایج بیان می‌دارد، مشتریانی با فرکانس کاری کمتر و همچنین حجم پولی کمتر حجم ترافیکی بیشتری را در سرویس بعدی خود تقاضا کرده‌اند. با مدنظر قرار دادن این نکته که این مشتریان در طول ایام حیات خود فعالیت بیشتری داشته‌اند، با استخراج این قوانین می‌توان به استراتژی‌های فروش بیشتر یا جانبی دست یافت و همچنین تخفیفی در سرویس احتمالی مورد تقاضای آینده مشتری، به منظور نگهداشت وی ارائه داد (غروی و تارخ، ۱۳۹۱).

جدول ۲- دقت پیش‌بینی سرویس مورد تقاضای مشتری با ویژگی‌های مختلف و با روش درخت تصمیم CART

ویژگیها	پیش‌بینی طول زمان سرویس آتی	پیش‌بینی حجم ترافیک سرویس آتی	پیش‌بینی سرعت سرویس آتی
RFM	۹۲،۵۹	۷۷،۷۸	۷۴،۰۷
RFM و میزان فعال بودن کاربر	۸۸،۸۹	۷۷،۷۸	۷۷،۷۸
RFM و ارزش طول عمر مشتری	۹۲،۵۹	۷۷،۷۸	۷۴،۰۷
RFM و طول عمر مشتری	۸۸،۸۹	۸۵،۱۹	۷۴،۰۷
RFM و حجم اطلاعات ورودی و خروجی	۷۷،۷۸	۷۷،۷۸	۸۸،۸۹
RFM و شیب مصرف	۸۸،۸۹	۷۷،۷۸	۷۴،۰۷
تمام ویژگیها	۸۸،۸۹	۸۵،۱۹	۸۵،۱۹

جدول ۳- دقت پیش‌بینی سرویس مورد تقاضای مشتری با ویژگی‌های مختلف و با روش درخت تصمیم C5

ویژگیها	پیش‌بینی طول زمان سرویس آتی	پیش‌بینی حجم ترافیک سرویس آتی
RFM	۶۲،۹۶	۷۷،۷۸
RFM و میزان فعال بودن کاربر	۷۴،۰۷	۷۷،۷۸
RFM و ارزش طول عمر مشتری	۶۲،۹۶	۷۷،۷۸
RFM و طول عمر مشتری	۸۵،۱۹	۸۱،۴۸
RFM و حجم اطلاعات ورودی و خروجی	۶۶،۶۷	۷۷،۷۸

ویژگیها	پیش بینی طول زمان سرویس آتی	پیش بینی حجم ترافیک سرویس آتی
RFM و شیب مصرف	۸۱,۴۸	۸۱,۴۸
تمام ویژگیها	۸۸,۸۹	۸۸,۸۹

۶- نتیجه‌گیری

در این مقاله داده‌کاوی به عنوان ابزاری جهت بررسی و کشف نیازهای مدیریت ارتباط با مشتری و بازاریابی معرفی گردید. هدف مدیریت ارتباط با مشتری شناخت و درک بهتر نیازهای مشتریان برای ارائه خدمات بهتر و در نتیجه نگهداشت مشتریان سودآور برای سازمان است. در اینجا داده‌کاوی با توجه به حجم انبوه داده‌ها، ابزار مناسبی در راستای کشف نیازها و شناسایی مشتریان به شمار می‌رود. داده‌کاوی با روش‌هایی چون پیش‌بینی، دسته‌بندی و خوشه‌بندی که در این مقاله مرور شدند، نیازهای اطلاعاتی مدنظر را برطرف می‌سازد. استفاده از داده‌کاوی جهت شناسایی و ارتباط با مشتری در موارد کاربردی فراوانی مشاهده شده است که تعدادی از آنها شامل خرده‌فروشی، بانکداری و صنعت مخابرات در این مقاله مرور شدند. همچنین، سه مطالعه موردی شامل پیش‌بینی سرویس مورد تقاضای آینده مشتری و بررسی روی‌گردانی مشتری در موضوع اینترنت و بررسی تأثیر عنوان تجاری بر سبد خرید مشتری در خرده‌فروشی مورد بررسی قرار گرفت و داده‌های آنها توسط روش‌های مختلف داده‌کاوی تحلیل شدند. نتایج حاصل از مطالعات موردی بیانگر توان بالای روش‌های داده‌کاوی در استخراج دانش و قواعد حاکم بر حجم انبوه داده‌های مختلف است.

منابع

1. Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 28(4), 773-781.
2. Cheung, D. W., Han, J., Ng, V. T., & Wong, C. (1996). *Maintenance of discovered association rules in large databases: An incremental updating technique*. Paper presented at the Data Engineering, 1996. Proceedings of the Twelfth International Conference on. IEEE, 106-114.
3. Fausett, L. V. (2013). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications* (Veisi, H, Mafakheri K, Bagheri, S, Trans.): Nas. [Persian]
4. Gharavi, E. (2012). *Customer churn management Using customer lifetime value*. (Master Dissertation), Khaje Nasir University of Technology. [Persian]
5. Gharavi, E., & Tarokh, M. (2013). *Predicting customers' future demand using data mining analysis: A case study of wireless communication customer*. Paper presented at the Information and Knowledge Technology (IKT), 2013 5th Conference on.
6. Guillén, M., Nielsen, J. P., Scheike, T. H., & Pérez-Marín, A. M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications*, 39(3), 3551-3558.
7. HAN, J. & MICHELINE, K. (2001). Data mining: concepts and techniques. *San Francisco, CA, itd: Morgan Kaufmann*, 5.

8. Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert systems with applications*, 31(1), 101-107.
9. Kracklauer, A. H., Mills, D. Q., & Seifert, D. (2004). Customer management as the origin of collaborative customer relationship management *Collaborative Customer Relationship Management* (pp. 3-6): Springer.
10. Ghazanfari, M., Alizadeh, S. & Teimourpour, B.. (2008). *Data Mining & Knowledge Discovery*. Tehran: Publication-Iran University of Science and Technology. [Persian]
11. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602. doi: 10.1016/j.eswa.2008.02.021
12. Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12), 15273-15285.
13. Radhakrishnan, B., Shineraj, G., & Anver Muhammed, K. (2013). Application of Data Mining In Marketing. *IJCSN International Journal of Computer Science and Network, ISSN (Online)*, 2277-5420.
14. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision support systems*, 31(1), 127-137.
15. Alizadeh, S., Gharavi, E. & Moradi, M. (2013). *Investigating the effect of brand on customer shopping basket*. Paper presented at the 9th international conference of industrial engineering Khaje Nasir University of Technology. [Persian]