

ارئه الگویی مبتنی بر یادگیری قانون وابستگی (Apriori) و تکنیک های داده کاوی جهت پیش بینی تراکنش های آینده مشتریان، مطالعه موردی شرکت ایران کیش

افسانه تقی زاده اردکانی

¹ کارشناس ارشد، دانشکده امنیت، گروه هوش مصنوعی و رباتیک، تهران،

Taghizadedafsaneh@gmail.com

چکیده

این تحقیق از جهت هدف، کاربردی و به لحاظ روش، توصیفی می باشد و در آن از تکنیک های داده کاوی، به عنوان ابزاری برای ایجاد یک الگویی برای دسته بندی مشتریان و پیش بینی خریدهای مشتریان دسته ی متوسط جهت ارائه پیشنهادهایی به آن ها با این هدف که اعضای این دسته را به سمت انجام تراکنش های بیشتر سوق دهد می پردازد. در این پروژه داده های مربوط به پرداخت قبض و خرید شارژ شامل دوره سه ماهه در سال جاری از شرکت ایران کیش مورد استفاده قرار می گیرد. بررسی این دادگان و خوشه بندی مشتریان در سه خوشه مشتریان خوب، مشتریان متوسط و بد می باشد که با توجه به اطلاعات بدست آمده از این شرکت مبنای دسته بندی مشتریان تعداد تراکنش های آن ها می باشد. پس از آن به پیش بینی خریدهای آتی مشتریانی که در دسته مشتریان متوسط قرار گرفته اند می پردازیم. پیش بینی تراکنش های مشتریان با توجه به سوابق تراکنش آن ها با استفاده از الگوریتم آپریوری می باشد، کاربرد الگوریتم آپریوری در این پروژه یافتن وابستگی بین تراکنش هایی می باشد که مشتری در گذشته انجام داده است، و با توجه به این وابستگی ها تراکنش های بعدی مشتریان پیش بینی می گردد.

در این تحقیق از الگوریتم های خوشه بندی k-means و الگوریتم Apriori برای پیاده سازی الگو استفاده شده است.

کلمات کلیدی

خوشه بندی مشتریان، داده کاوی، الگوریتم آپریوری، مشتریان خوب، مشتریان متوسط، مشتریان بد، تعداد تراکنش.

۱- مقدمه

شدند، سپس رتبه بندی بر اساس طول عمر مشتری صورت گرفت، و کشف الگوهای پنهان بین داده های استفاده از خدمات بانکداری الکترونیکی با استفاده از الگوریتم C5 انجام گرفت، که نتایج حاصل میتوانست به عنوان راهکاری در بازاریابی الکترونیکی به کار گرفته شود. با توجه به اهمیت بانکداری الکترونیکی به ویژه اینترنت بانک، شناخت نیاز مشتریان برای استراتژی بازاریابی کمک شایانی به افزایش کاربرد اینترنت بانک در جوامع امروزی خواهد کرد، ما در این پژوهش مشتریان اینترنت بانک ایران کیش را مورد مطالعه قرار داده و با بررسی تراکنش های آن ها، خوشه بندی را انجام می دهیم تا بتوانیم مشتریان خوب و وفادار را شناسایی کرده و از طرفی استراتژی لازم را برای افزایش تعداد این مشتریان خوشه خوب پیاده سازی کنیم.

۲- داده کاوی (Data Mining) چیست؟

نگاهی به ترجمه تحت اللفظی داده کاوی، به ما در درک بهتر این واژه کمک می کند. Mine به معنای استخراج از منابع نهفته و با ارزش زمین اتلاق میشود. پیوند این کلمه با کلمه داده، جستجوی عمیق جهت پیدا کردن اطلاعات اضافی مفید که قبلاً نهفته بودند، از داده های قابل دسترس حجیم، را پیشنهاد میکند. داده کاوی یک رشته نسبتاً جدید علمی میباشد که از انجام تحقیقات در رشته های آمار، یادگیری ماشینی، علوم کامپیوتر خصوصاً مدیریت پایگاه داده ها شکل گرفته است.

تعاریف متنوعی از داده کاوی در مراجع مختلف و توسط افراد مختلف ارائه شده است. از جمله:

- داده کاوی یک فرآیند شناخت الگوهای معتبر، جدید، ذاتاً مفید و قابل فهم از داده ها می باشد.
- داده کاوی (کشف دانش از پایگاه دادهها نامیده میشود) نشانگر فرآیند جالب استخراج دانش از قبل ناشناخته (الگو) از داده است.
- فرآیند کشف الگوهای مفید از داده ها را داده کاوی می گویند.
- فرآیند انتخاب، کاوش و مدلبندی داده های حجیم، جهت کشف روابط نهفته با هدف به دست آوردن نتایج واضح و مفید، برای مالک پایگاه داده ها را، داده کاوی گویند.

۱-۲- فرآیند داده کاوی (CRISP)

موفق ترین پروژه های داده کاوی، در چارچوب فرآیند استاندارد اجرا می شود که توسط یک تیم کاری در شرکت SPSS در قالب پروژه ای به نام فرآیند اجرای داده کاوی استاندارد صنعتی CRISP-DM Cross ارائه شده است. بر طبق CRISP-DM یک پروژه داده کاوی معین، شامل چرخه حیات شش مرحله ای است (شکل ۱).

در مرحله درک تجارت، اهداف تجاری، وضعیت، اهداف داده کاوی تعریف می شود و طرح پروژه تولید می گردد. در مرحله فهم داده، به جمع آوری و توصیف داده های اولیه، کشف داده، بررسی کیفیت داده پرداخته می شود. در مرحله آماده سازی داده، داده ها انتخاب و پاک سازی می شود و سپس به ساخت داده، جمع آوری داده و در نهایت طراحی ظاهر داده می پردازیم. در الگو سازی، تکنیک الگو سازی انتخاب می شود و طرح آزمون تولید می گردد و در نهایت الگو ساخته و مورد ارزیابی قرار می گیرد. مراحل

همزمان با الکترونیکی شدن تجارت و کسب و کار، بانکداری نیز به شکل الکترونیکی درآمده تا بتواند به مشتریان خدمات مالی و بانکی برخط ارائه کند. توسعه و پیشرفت بانک ها در زمینه بانکداری الکترونیک که یکی از پیش شرط ها و ابزارهای اساسی مورد نیاز تجارت الکترونیک به شمار می رود، بسیار ضروری و حیاتی است، اما آنچه بیش از همه این ها اهمیت دارد، عمومی کردن بانکداری الکترونیک و تلاش در جهت استقبال مشتریان از خدمات الکترونیکی بانک ها است، بانک های سراسر جهان، روزانه حجم زیادی از داده های الکترونیکی را در پایگاه داده ی متمرکز نگهداری می کنند. برای سازمان ها غیرممکن است که با حجم زیاد داده ها به تجزیه و تحلیل و بازاریابی اطلاعات مفید بپردازند در این راستا می بایستی رفتار مشتریان در زمینه خدمات بانکداری الکترونیک به دقت بررسی شود تا بانک ها بتوانند بر اساس آن ها الگوهای مناسب را شناخته و به ترویج بانکداری الکترونیک اقدام کنند.

دلیل حجم گسترده و بسیار زیاد داده های موجود در بانک و همچنین عدم امکان ذخیره سازی آنها و نیاز به پردازش و استخراج پویای اطلاعات و دانش نهفته ها در داده ها بحث داده کاوی جریان داده ها در بانک طی سالهای اخیر بسیار مورد توجه قرار گرفته است. ارتباط با مشتری یکی از مهمترین عوامل سازنده مخصوصاً در سازمانهای خدماتی می باشد از جمله این سازمانها بانک ها و موسسات مالی است که مدیریت ارتباط با مشتری در آنها نقش کلیدی را دارا می باشد. در بانکداری نوین، بانکها برای شناسایی مشتریان و نیازهای آنها دست به بازاریابی براساس پایگاه داده میزنند.

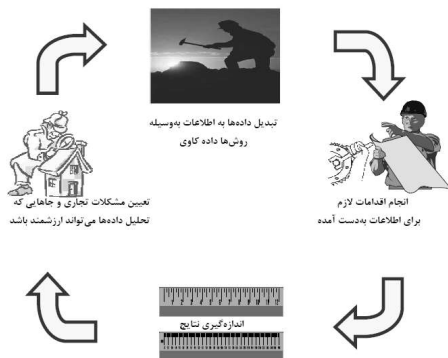
به عنوان مثالی از بازاریابی بر اساس پایگاه داده در تحقیق «شناسایی و کشف الگوهای خرید مشتریان لوازم خانگی از طریق داده کاوی» سعی شده است تا روابط پنهان میان داده های حاصل از فروش محصولات خانگی با استفاده از تکنیک قوانین انجمنی کشف شود. نتایج حاصل از این روش کمک شایان توجهی در شناخت رفتار خرید مشتریان لوازم خانگی، طراحی و تدوین برنامه های بازاریابی جدید در حوزه لوازم خانگی، هدفمند ساختن بودجه های تبلیغاتی، امکان سنجی بازار برای تولید و عرضه محصول جدید و بررسی و شناخت قوت و ضعف محصولات و خدمات خواهد نمود. در تحقیقی دیگر با عنوان "طبقه بندی مشتریان اینترنت بانک با کمک الگوریتم های داده کاوی" یک مدل برای طبقه بندی مشتریان با بهره گیری از خدمات اینترنت بانک ارائه شده است. این مدل بر اساس استاندارد CRISP-DM انجام گرفته و داده های مورد نیاز از پایگاه داده مشتریان اینترنت بانک سینا استخراج شده است، در این روش با استفاده از درخت تصمیم به طبقه بندی مشتریان بر اساس سن، تحصیلات، شغل، شهر و جنسیت پرداخته شده است و این نتیجه حاصل شده است که شاخص شغل و تحصیلات مهمترین شاخص ها در بخش بندی مشتریان است.

خدایور و همکاران در پژوهشی برای خوشه بندی مشتریان شرکت کدبانو از دو معیار ارزش گذاری RFM و LRFM استفاده کرد که در نهایت تعداد خوشه های بهینه را به ۸ عدد تعیین کردند.

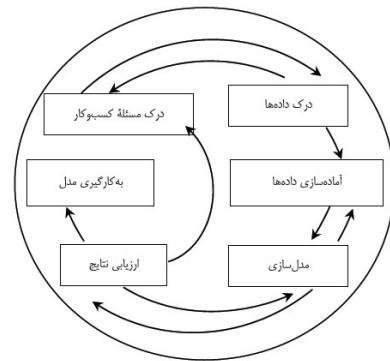
تقوی فرد و همکاران به بخش بندی خوشه ای مشتریان بانک صادرات پرداختند، نمونه مشتریان بر اساس تحلیل RFM با استفاده از الگوریتم Twostep خوشه بندی و به این ترتیب مشتریان به چهار بخش تفکیک

ارزیابی بدین شرح می باشد: ارزیابی نتیجه ها، بازبینی فرآیند، تعیین گام بعدی، توسعه، طرح ریزی برای توسعه، طرح ریزی برای نگهداری و نظارت، تهیه گزارش پایانی، بازبینی پروژه.

ارزیابی بدین شرح می باشد: ارزیابی نتیجه ها، بازبینی فرآیند، تعیین گام بعدی، توسعه، طرح ریزی برای توسعه، طرح ریزی برای نگهداری و نظارت، تهیه گزارش پایانی، بازبینی پروژه.



شکل ۲: حلقه داده کاوی



شکل ۱: فرآیند کریسپ

۲-۳- داده کاوی و کشف دانش از پایگاه داده

۲-۲- حلقه داده کاوی چیست؟

تاریخچه کشف دانش از پایگاه های اطلاعاتی قدمت چندانی ندارد و امروزه به داده کاوی مشهور است. اصطلاح کشف دانش برای نخستین بار در دهه ۱۹۹۰ مطرح شد و توجه پژوهشگران را به سمت الگوریتم های داده کاوی معطوف کرد. هدف داده کاوی، کشف دانش جدید، معتبر و قابل پیگیری با استفاده از ابزار های هوش مصنوعی و آماری در حجم بالایی از داده ها است. در (شکل ۳) مراحل کشف دانش از پایگاه داده ها نشان داده شده است:

چهار قدم اصلی حلقه داده کاوی عبارتند از: (شکل ۲)

- انتخاب داده ها: داده های مربوط به تجزیه و تحلیل و تصمیم گیری از داده های دیگر جدا می شوند.
- پیش پردازش اطلاعات: پردازش، تمیز کردن و یکپارچ سازی داده ها انجام می گیرد.
- تبدیل داده ها: داده های انتخاب شده به شکل مناسبی برای روش داده کاوی تبدیل می شوند.
- داده کاوی: در این مرحله از روش های هوشمندانه ای برای استخراج الگوهای بالقوه مفید استفاده می شود و در مورد این روش ها تصمیم گیری می شود.
- تفسیر و ارزیابی: در این مرحله، الگوهای جالب توجه نشان دهنده ی دانش، بر اساس اقدامات انجام شده شناخته می شوند و دانش کشف شده در اختیار کاربر قرار می گیرد.

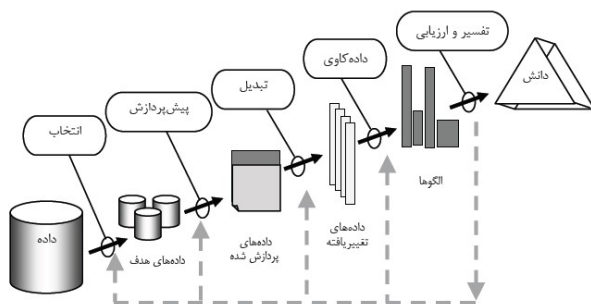
- صورت مساله و مشکل تجاری را تعریف کنیم.
- از داده کاوی استفاده کنیم تا داده ها را به اطلاعات و دانش تبدیل کنیم.
- اقدامات لازم را در مورد اطلاعات کسب شده انجام دهیم.
- نتایج را اندازه گیری کنیم.

همانطور که از این مراحل پیداست کلید موفقیت داده کاوی در آمیختن آن با فرآیند تجارت و کسب و کار می باشد. این مراحل بسیار به هم وابسته می باشند و نتایج هر مرحله مانند یک فرآیند تولید چند مرحله ای یک ورودی برای مرحله بعد می باشد یعنی هر مرحله به نتیجه و خروجی مرحله قبل وابسته است.

به دست آوردن نتایج واضح و مفید، برای مالک پایگاه دادهها راه، داده کاوی گویند. اما تعریفی که در اکثر مراجع به اشتراک ذکر شده عبارت است از: « استخراج اطلاعات و دانش و کشف الگوهای پنهان از یک پایگاه داده ی بسیار بزرگ و پیچیده ». داده کاوی یک متدولوژی بسیار قوی و با پتانسیل بالا می باشد که به سازمانها کمک می کند که بر روی مهمترین اطلاعات از مخزن داده های خود تمرکز نمایند.

داده کاوی فرآیندی است که از ابزارهای تحلیلی گوناگونی برای کشف الگوها و روابط بین داده ها استفاده میکند که ممکن است برای اعتبار بخشیدن به پیشبینی استفاده شود.

داده کاوی کمک میکند تا سازمانها با کاوش بر روی داده های یک سیستم، الگوها، روندها و رفتارهای آینده را کشف و پیشبینی کرده و بهتر تصمیم بگیرند. داده کاوی با استفاده از تحلیل وقایع گذشته یک تحلیل اتوماتیک و پیش بینانه ارایه می نماید و به سوالاتی جواب میدهد که پاسخ آنها در گذشته ممکن نبوده و یا به زمان زیادی نیاز داشت. ابزارهای داده کاوی



شکل ۳: مراحل کشف دانش از پایگاه داده ها

دو هدف اصلی داده کاوی، پیشگویی و توصیف است.

- داده کاوی پیشگویی کننده، مدلی را از سیستم ارائه می دهد که این مدل را مجموعه ای از داده های مشخص، پیش بینی می کنند. هدف کلی آن ایجاد الگویی برای طبقه بندی، پیش بینی و تخمین داده ها است.
- داده کاوی توصیفی، اطلاعات جدید و غیر بدیهی را بر اساس مجموعه ای از داده های موجود ارائه می دهد و هدف کلی آن درک و شناخت سیستم های تجزیه و تحلیل شده با استفاده از الگوها و روابط موجود است.

شایان ذکر است که داده کاوی از ترکیب چندین رشته نشئت می گیرد. آمار، یادگیری ماشین، روش های بهینه سازی، روش های تشخیص و شناخت الگو، بانک اطلاعاتی، تجسم سازی، شبکه های عصبی، مدل های ریاضی، بازیابی اطلاعات، الگوریتم ژنتیک و هوش مصنوعی فونونی هستند که داده کاوی از آنها بهره می برد.

۳- گردآوری اطلاعات

با آگاهی از چگونگی و ویژگی های اطلاعات مورد نیاز هر شرکت، این پرسش پیش می آید که اطلاعات سودمند را چگونه میتوان به صورت موثر و کارآمد گردآوری نمود. اطلاعات با خود هر دو مورد سودمندی و هزینه سازی را به همراه می آورد. ممکن است در گردآوری داده ها زیاده روی شود. شرکت هایی هستند که در اطلاعات غرق شده ولی همچنان تشنه دانش و آگاهی هستند. میان داده ها، اطلاعات، دانش و خرد تفاوت گسترده و آشکاری است. تا هنگامی که داده به اطلاعات پرورده، و سپس به دانش دست آخر به خرد و شناخت از بازار تبدیل نشوند، هدر رفته و بی مصرف هستند.

این مرحله با جمع آوری اولیه داده ها شروع می شود و به توصیف داده ها و تعیین کیفیت آنها می پردازد. در این پژوهش برای ارزیابی اولیه از داده ها، ابتدا پایگاه های داده ی مشتریان حقیقی را که در خود ذخیره کرده است را، مورد بررسی قرار می دهیم و سپس با نمونه هایی از داده های موجود در پایگاه های داده، نحوه اخذ داده ها و اطلاعات مورد نیاز از سیستم های عملیاتی رایانه ای، سیستم رتبه بندی، نحوه ذخیره سازی داده ها و کسب اطلاعات از آنها، آشنایی کامل حاصل می شود. پیش از جمع آوری و انتخاب داده ها برای شروع عملیات پیش پردازش، ابتدا کیفیت داده های موجود و دسترسی به آنها مورد سنجش قرار می گیرد. این ارزیابی کیفیت به این دلیل انجام می گیرد که در داده ها، خصیصه های اصلی تأثیرگذار در رفتار مشتریان وجود داشته باشد و داده های گم شده یا ناقص در میان داده ها، میزان قابل قبولی داشته باشند.

۱-۳- آماده سازی داده ها

مرحله آماده سازی داده ها، شامل کلیه فعالیت هایی است که برای ساختن مجموعه داده های نهایی (داده هایی که برای مدل سازی آماده شده اند) از داده های خام اولیه به کار می رود. هرچه کیفیت این آماده سازی بهتر باشد، مدل سازی نیز بهتر خواهد بود. وظایف آماده سازی داده ها در چند

دوره انجام می گیرد و هیچ ترتیب از پیش تعریف شده ای ندارد. این وظایف شامل انتخاب جداول، رکوردها و خصیصه ها و همچنین انتقال و پاک سازی داده برای مدل سازی است. در این مرحله، داده های مرتبط با مشتریان از پایگاه داده هایی که در مرحله قبل شناسایی شدند، استخراج و ثبت شدند. سپس داده ها در یک پایگاه داده جامع و یکپارچه (البته پالایش نشده) به منزله یک پایگاه داده رابطه ای قرار گرفتند. سپس داده ها پالایش شده و ساختار مورد نظر برای مدل سازی روی آنها اعمال می شود. برای پاک سازی و پیش پردازش داده ها، دو عملیات کاهش داده و اعمال تغییرات در شکل داده ها، روی پایگاه داده رابطه ای می گیرد.

۴- خوشه بندی

تحلیل خوشه ای که اصطلاح آن نخستین بار توسط تریان در سال ۱۹۳۹ به کار برده شد، یک تکنیک طبقه بندی برای تشکیل گروه های همگون در مجموعاً پیچیده ای از داده ها است داده ها بر اساس اصل حداکثر کردن شباهت داخل گروه ها و حداقل کردن شباهت بین گروه ها، خوشه بندی می شوند. خوشه بندی یک روش متداول توصیفی است که در جستجوی تشخیص تعداد محدودی خوشه، برای توصیف داده ها است. بوسیله خوشه بندی می توانیم نحوه توزیع الگوها را مشخص نماییم. این تکنیک با روش های طبقه بندی Classification متفاوت است. طبقه بندی با تعداد معلوم و مشخصی از گروه ها ارتباط دارد، و هدف آن اختصاص مشاهده های جدید به یکی از این گروه ها است. ما تحلیل خوشه ای یک تکنیک ابتدائی است که به هیچ مفروضه ای درباره تعداد یا ساختار گروه ها متکی نیست. در تحلیل خوشه ای، عضویت گروهی برای همه مشاهده ها نامعلوم، و حتی تعداد گروه ها نیز نامشخص است. تقسیم مشتریان با ویژگی های معمول به گروه های کوچکتر می تواند مرجع مفیدی برای تصمیم گیری باشد. تاکنون الگوریتم K-means پر استفاده ترین تکنیک خوشه بندی برای گروه بندی مشتریان بوده است در تحقیق های مشابه خوشه بندی، از شاخص RFM استفاده شده است. در بسیاری از روش های ارائه شده برای خوشه بندی، مقدار k (تعداد خوشه ها) باید توسط کاربر تعیین شود. در این روش ها جستجو برای یافتن بهترین تعداد خوشه بر اساس معیاری که قبلاً تعیین شده انجام می شود. لیو در سال ۱۹۶۸ نشان داد که تعداد حالت های مختلف برای خوشه بندی n شیء به k خوشه برابر است با:

$$NW(n, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^n \quad (1)$$

از عبارت فوق به راحتی می توان فهمید که حتی اگر k از قبل تعیین شده باشد، پیدا کردن بهترین خوشه بندی کار ساده ای نیست. اما کاربردهای زیادی وجود دارد که در آنها مقدار k از قبل تعیین نشده است و کاربر نیز دید روشنی از تعداد خوشه هایی که در داده ها می توان مشخص کرد، ندارد. بنابراین باید از معیاری برای تعیین تعداد خوشه بهینه استفاده شود. متداول ترین شاخص های مورد استفاده برای ارزیابی خوشه ها، شاخص دان و شاخص مجموع خطای مربعی SSE می باشند که در این تحقیق از این معیارها برای خوشه بندی استفاده شده است.

۵- قواعد انجمنی

نهایتاً کشف دانش از این الگوها و ارائه قواعد استنتاج شده از این الگوها، به منظور کمک به بانکداری الکترونیکی برای ارائه پیشنهادهای مناسب به مشتریان رده متوسط و جذب آن ها برای قرار گرفتن در دسته مشتریان خوب است.

۷- راه حل پیشنهادی

داده های مربوط به سه ماه اول سال ۲۰۱۷ میلادی از مشتریان شرکت ایران کیش جمع آوری شده و پس از انجام عملیات پیش پردازش و حذف داده های غیرضروری که برای انجام تحقیق مناسب نیستند، داده ها برای عملیات لازم آماده می شوند. پس از آن این داده ها که شامل ۵۷۵۷۷ رکورد از مشتریان این شرکت است توسط الگوریتم خوشه بندی برای تعیین اینکه هر مشتری بر اساس تراکنش هایی که در این سه ماه انجام داده است در کدام خوشه قرار می گیرد، خوشه بندی می گردد. که ما برای این کار از روش خوشه بندی k-means استفاده می کنیم. روش کار بدین صورت است که نقاطی به عنوان مراکز خوشه ها محاسبه می شوند این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند سپس هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد نسبت داده می شود. الگوریتم k-means یکی از مشهورترین الگوریتم های خوشه های بندی داده می باشد که به دلیل پیاده سازی آسان و سرعت عملکرد، محبوبیت زیادی یافته است. پس از مشخص شدن هر خوشه و مشتریانی که متعلق به آن خوشه هستند، هدف بعدی ما مشتریان دسته متوسط هستند تا بتوانیم با کمترین سرمایه آن ها را تشویق به انجام تراکنش های بیشتر کنیم. برای این کار تراکنش هایی که این دسته از مشتریان انجام داده اند را مورد بررسی قرار داده و آن ها برای پیش بینی توسط الگوریتم Apriori آماده سازی می کنیم.

۸- نتیجه گیری

در این مقاله از بین ۷۵۵۷۷ نفر مشتری که از پایگاه داده شرکت ایران کیش در طول سه ماه اول سال ۲۰۱۷ میلادی دریافت شده بود پس از انجام عملیات خوشه بندی توسط الگوریتم k-means و تعیین سه خوشه، مشتریان ما به خوشه مشتریان خوب شامل مشتریان که در طول این سه ماه میانگین تراکنش آن ها برابر با ۱۴۴.۲ بود و خوشه مشتریان متوسط دارای میانگین تراکنش برابر با ۳۷.۸۹ و در نهایت خوشه مشتریان بد که میانگین تعداد تراکنش آن ها برابر با ۲۶ بود تقسیم شد. نتایج حاصل از خوشه بندی حاکی از آن است که مشتریان خوشه بد، مشتریانی هستند که به صورت دائم تراکنش های خود را با شرکت ایران کیش انجام نمی دهند و می توان گفت که به صورت اتفاقی تراکنش را انجام داده اند و جز مشتریان وفادار محسوب نمی شوند. از طرفی مشتریان متوسط مشتریانی هستند وفاداری کمتری نسبت به مشتریان خوب دارند ولی می توان با سیاست های بازاریابی آن ها را به سمت خوشه مشتریان خوب جلب کرد. برای این کار تراکنش های مشتریان خوشه متوسط که شامل ۱۹۷۱ عضو هست را در طول این سه ماه مورد مطالعه قرار داده و توسط الگوریتم اپریوری به پیش بینی تراکنش های آینده آن ها پرداختیم تا بانک بتواند با سیاست هایی مانند ارسال پیشنهادهای ویژه برای مشتری توسط پیامک یا ایمیل آن ها را جذب کند.

یکی از مهمترین بخشهای داده کاوی، کشف قوانین وابستگی در پایگاه داده است، الگوریتم وابستگی نوعی آنالیز پیوندی است که برای تشخیص رفتار یک رویداد و یا یک پروسه خاص استفاده می شود. این الگوریتم از دو بخش شرط و نتیجه تشکیل شده. بخش شرط یک آیتم خاص از اطلاعات را کشف و در بخش نتیجه آیتمی دیگر از اطلاعات که وابسته به شرط است را پیدا می کند. در واقع شرحی از دو کلمه ی "اگر" و "پس"، برای کشف رابطه های ناشناخته میان اطلاعات است. برای مثال، می توان گفت که "اگر" شخصی چکش بخرد "پس" به احتمال ۸۰ درصد میخ هم خریداری می کند.

یادگیری قانون وابستگی در داده کاوی، یک متد مناسب برای یافتن روابط جذاب بین متغیرهای موجود در پایگاه داده های بزرگ است. بیاتسکی-شاپیرو چگونگی تحلیل و ارائه قوانین قوی یافته شده را در پایگاه های داده با استفاده از معیارهای متفاوت جذابیت توضیح می دهد. بر مبنای مفهوم قوانین قوی، راکش اگرول و همکارانش قوانین وابستگی را برای کشف قاعده های موجود بین محصولات در داده های تراکنشی با مقیاس بالا معرفی می کنند. به عنوان مثال، قانون وابستگی در داده های فروش یک سوپرمارکت، نشان می دهد در صورتی که یک مشتری پیاز (onions) و سیب زمینی (potatoes) را در سبد خرید خود قرار داده است، احتمالاً او مایل به خرید گوشت همبرگر نیز خواهد بود. چنین اطلاعاتی می تواند به عنوان مبنای تصمیماتی برای برخی از فعالیت های فروشگاهی همچون ارائه مناسب تخفیف برای محصولات یا قراردادن مناسب محصولات در کنار هم، مورد استفاده قرار گیرد. علاوه بر مثال فوق که در مورد تحلیل سبد خرید مطرح شد، امروزه یادگیری قانون وابستگی در کاربردهای متفاوت همچون مصرف کاوی وب، تشخیص نفوذ، و بیوانفورماتیک مورد استفاده قرار می گیرد.

الگوریتم وابستگی بسیار مفید است چراکه توسط آن می توانیم رفتار مشتریان را تجزیه و تحلیل و پیش بینی کنیم. بعلاوه این الگوریتم نقش مهمی در تحلیل اطلاعات سبد خرید و چیدمان محصولات دارد که در نتیجه سبب می شود تا کالاها را به صورت هدف دار به مشتریان پیشنهاد دهیم.

۶- طرح مسئله

آیا تا کنون به این موضوع فکر کرده ایم که فروشگاه های بزرگ اینترنتی در mail های خود به مشتریان از چه تبلیغاتی استفاده می کنند؟ و آیا این تبلیغات برای همه مشتریان یکسان است؟

رتبه بندی مشتریان یکی از دغدغه های بانک ها و سایر موسسات مالی و اعتباری به منظور اختصاص منابع و تعیین استراتژی های کلان در حوزه بازاریابی بوده و هست، سودمندترین مشتریان به وسیله ی ابزارهای امتیاز دهی مشتریان شناسایی شده و مورد توجه خاص قرار می گیرد. بنابراین یافتن روش ها و مدل هایی برای بخش بندی مشتریان، برای سازمان ضروری به نظر می رسد. بنابراین بانک داری الکترونیک نیازمند آن است که مشتریان خود را با روش هایی شناسایی کرده و برای هربخش از مشتریان تصمیمات لازم را گرفته تا بتواند سطح مشتریان وفادار خود را افزایش دهد.

هدف از این تحقیق، کاوش در پایگاه داده ی شرکت ایران کیش، با استفاده از روش های داده کاوی، به منظور کشف الگوهای حاکم بر خرید مشتریان و

نتایج حاصل از این الگو نشان می دهد افرادی که عملیات خرید شارژ را با این اپلیکیشن انجام داده اند با اطمینان ۹۷ درصد عملیات پرداخت قبض خود را از این درگاه پرداخت می کنند. و افرادی که عملیات پرداخت قبض را با این سامانه انجام داده اند با اطمینان ۹۲ درصد برای خرید شارژ اقدام خواهند کرد. بنابراین مشتریان این دسته تمایل بیشتری به خرید شارژ و پرداخت قبض نسبت به عملیات های پرداخت دارند، و زمانی که مشتریان این دسته اقدام به پرداخت شارژ کردند می توان با سیاست های نامبرده پیشنهادات ویژه ای را برای آن ها در نظر گرفت.

منابع

- [۱] خدیور _آمنه ، حامدی _پروشات ، راهبردهای بازرگانی دانشور ، ارائه یک الگو ترکیبی داده کاوی با استفاده از قواعد انجمنی و خوشه بندی برای تعیین استراتژی تخفیف دهی، مطالعه موردی شرکت پخش پگاه ، دو فصلنامه علمی _ پژوهشی دانشگاه شاهد ، سال بیست و دوم- دوره جدید ، شماره ۵، بهار و تابستان ۱۳۹۶
- [۲] تقوی فرد _ محمدتقی ، خواجهوند _ سمانه ، پژوهشی کاوشهای مدیریت بازرگانی، بخش بندی خوشه ای مشتریان بانک ، دو فصلنامه علمی _ کاوشهای مدیریت بازرگانی، شماره ۱، صص ۶۴-۳۹، بهار و تابستان ۱۳۹۲
- [۳] مولایی _منیژه ، پارسا _سودابه پیش بینی رفتار مشتریان با استفاده از تکنیک شبکه های عصبی مصنوعی ، سال دوم _ شماره ۳ (پیاپی: ۱۰) جلد ۱ _ خرداد ۱۳۹۵
- [۴] خدیور _ آمنه ، رزمی _ زهرا ، حامدی _ پروشات ، فصلنامه علمی _پژوهشی تحقیقات بازاریابی نوین ، خوشه بندی مشتریان به منظور اصلاح استراتژی قیمت گذاری و تدوین استراتژی تخفیف دهی ، مطالعه موردی شرکت کدبانو ، سال سوم _ شماره سوم _شماره پیاپی(۱۰) ، پاییز ۱۳۹۲
- [۵] اکبریور شیرازی _ محسن ، توپچی _ حسن ، داده کاوی : مفاهیم، روش ها و کاربردها ، دانشگاه خواجه نصیرالدین طوسی
- [۶] [عز] ادفر _ رضا ، نظافتی _ نوید ، یوسفی اصل _ سعید ، دانشگاه مدیریت دانشگاه تهران ،مدیریت فناوری اطلاعات ، طبقه بندی مشتریان اینترنت بانک با کمک الگوریتم های داده کاوی ، دوره ۶، شماره ۱ ، بهار ۱۳۹۳ ، صص ۷۱-۹۰
- [۷] [۷] فضلی مقصودی _ حسن ، مومنی _ حسین ، مقایسه و بررسی الگوریتمهای داده کاوی درخت تصمیم و ماشین بردار پشتیبان برای تشخیص ، هشتمین همایش بین المللی پیشرفت علم و فناوری مشهد

[8]Harikumar_Sandhya·Divya Usha Dilipkumar·Apriori Algorithm for Association Rule Mining in High Dimensional Data , Department of Computer Science and Engineering , Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University, India

[9] Ching-Hsue Cheng, You-Shyang Chen , Classifying the segmentation of customer value via RFM model and RS theory , Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan