

کاربرد داده کاوی در حوزه بازاریابی با نگرشی بر الگوی مصارف خوراکی خانوارهای ایرانی

بهروز مینایی^۱، ساراسادات تهامی^۲، سیده معصومه زهرایی^۳

چکیده

رشد سریع فناوری اطلاعات امروزه باعث تغییرات بسیاری در روش‌های بازاریابی سازمان‌ها نموده است و از طرفی با توجه به رقابت فشرده بازار و گزینه‌های متنوعی که از محصولات و خدمات متنوع پیش‌روی مشتریان قرار دارد، تحلیل رفتار مشتریان، به عامل بسیار مهمی برای بقای سازمان‌ها مبدل گشته است. نظر به موارد مذکور، یکی از موثرترین ابزارهای بررسی رفتار مشتریان، داده کاوی است. این مقاله با هدف تمرکز بر رفتار مشتریان ایرانی در میزان مصرف خوراکی‌های مهم، با توجه به میزان درآمد و محل زندگی آنان و در دسترس بودن داده‌های تفصیلی مرکز آمار ایران در طی 3 سال اخیر و طی مراحل فرایند استاندارد داده کاوی (CRISP-DM) انجام گرفته و با بکارگیری تکنیک‌های تحلیل اکتشافی داده‌ها، خوشه‌بندی، قوانین انجمنی و غیره به استخراج دانش کاربردی پرداخته‌ایم. نتایج تحلیل‌های انجام شده به منظور کمک به بررسی ارتباطات و تهیه الگوهای کاربردی در حوزه بازاریابی و تاثیرات آن در مدیریت ارتباط با مشتریان کارخانجات مواد غذایی به کار می‌رود.

کلمات کلیدی

رفتار مشتریان، بازاریابی، داده کاوی، خوشه‌بندی، قوانین انجمنی، CRISP-DM

Minaeibi@cse.msu.edu

Email:

1. استادیار و عضو هیئت علمی دانشگاه علم و صنعت ایران

Stahami@stu.qom.ac.ir

Email: 2. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، گرایش تجارت الکترونیک، دانشگاه قم

Email: M.zahraie@gmail.com

3. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، گرایش تجارت الکترونیک، دانشگاه قم



در سالهای اخیر فرهنگ تجارت به پیشرفتهایی نایل گشته است. از جمله پیشرفتهای مهمی که در این زمینه رخ داده است می توان به تغییر روابط اقتصادی مشتریان اشاره نمود که به تبع آن شرکتها و سازمانها به منظور نظارت بر اینگونه تغییرات نیازمند آرایه راه حلهای بهینه‌ای در این عرصه می‌باشند.

از طرفی ظهور و پیدایش اینترنت در تغییر روشهای بازاریابی نقش بسزایی داشته است. چنانچه اطلاعات آنلاین بیشتر در دسترس قرار گیرد موجب آگاهی و هوشیاری بیشتر مشتریان می‌گردد و آنها همچنین در جریان تمام آنچه آرایسه و پیشنهاد می‌شود قرار می‌گیرند و تقاضای بهترینها را خواهند داشت. برای از عهده برآمدن در چنین شرایطی باید سیستم‌هایی که بتواند به طور دقیق نسبت به مشتریان واکنش نشان دهد به کار رود. جمع آوری آمار مشتریان و داده‌های رفتاری آنها این هدف اصلی و دقیق را ممکن می‌سازد. بنابراین، شرکت‌های امروزی از طریق تجزیه و تحلیل رفتار مشتری به افزایش ارزش مشتری دست یافته‌اند. در واقع امروزه، دیدگاه محصول محوری جای خود را به مشتری محوری داده است.

ابزارها و فناوریهای انبار داده، داده‌کاوی و دیگر تکنیک‌های مدیریت ارتباط با مشتری، روشهایی هستند که فرصتهای جدیدی را برای تجارت فراهم کرده‌اند.

با جمع‌آوری داده‌های مربوط به مشتری و تصمیم‌گیری براساس الگوهای استخراج شده از روابط پنهان میان داده‌ها به وسیله ابزار داده‌کاوی، می توان به خواسته مشتری محوری خود جامه عمل پوشاند. این مقاله مفاهیمی از کاربرد داده‌کاوی در بازاریابی و مدیریت ارتباط با مشتری را مورد بررسی قرار می‌دهد.

داده‌کاوی یکی از شیوه‌های رایج فرآوری داده‌ها در عصر کنونی است که به کمک بسیاری از قلمروها در گسترش دانش آمده است. این فرایند بین رشته‌ای در علم کامپیوتر و مهندسی اطلاعات و ارتباطات جایگاه ویژه‌ای یافته است و مشترکاتی با رشته‌های آمار کاربردی، هوش مصنوعی، سیستم‌های خبره، مهندسی اطلاعات، یادگیری ماشین و شبکه‌های عصبی دارد. بهره‌برداری‌های زیادی از این زمینه در کشور ما هنوز انجام نشده است. اما این روش برای بسیاری از تحلیلگران علوم اقتصادی، ارتباطات، بازاریابی، جامعه‌شناسی، نجوم، تشخیص بیماری و... شناخته شده و صاحب جایگاه است.

در دنیای امروز داده‌ها و اطلاعات بعنوان ثروت سازمانی، محسوب گشته و همواره شرکتها و سازمان‌های بزرگ و موفق دنیا به دنبال استفاده مناسبتر و تجاری‌تر از این منابع مجازی می‌باشند. از سوی دیگر با پیچیده‌تر شدن محیطهای کسب و کار، ماهیت و حجم داده‌های سازمانی بسیار متفاوت گشته و نگاه یکپارچه و مدیریتی به آنها ضروری می‌گردد. یکی از راه‌حلهایی که سازمانهای موفق دنیا در این زمینه اتخاذ می‌نمایند، ایجاد یک سیستم جامع داده‌ای و آماری و بکارگیری صحیح تکنیکهای داده‌کاوی و کشف دانش میباشد [۱]

بنابراین، تحلیل رفتار مشتریان، به عامل بسیار مهمی برای بقای سازمانها مبدل گشته است. نظر به موارد مذکور، یکی از موثرترین ابزارهای بررسی رفتار مشتریان، داده‌کاوی است. این مقاله با هدف تمرکز بر رفتار مشتریان ایرانی در میزان مصرف خوراکی‌های مهم، با توجه به میزان درآمد و محل زندگی آنان و در دسترس بودن داده‌های تفصیلی مرکز آمار ایران در طی 3 سال اخیر و طی مراحل فرایند استاندارد داده‌کاوی (CRISP-DM) انجام گرفته و با بکارگیری تکنیک‌های تحلیل اکتشافی داده‌ها، خوشه‌بندی، قوانین انجمنی و غیره به استخراج دانش کاربردی پرداخته‌ایم. نتایج تحلیل‌های انجام شده به منظور کمک به بررسی ارتباطات و تهیه الگوهای کاربردی در حوزه بازاریابی و تاثیرات آن در مدیریت ارتباط با مشتریان کارخانجات مواد غذایی به کار می‌رود.

در این مقاله با دو فرایند اصلی داده‌کاوی که یکی منجر به تولید الگو و دیگری تولید قانون می‌شود سر و کار داریم و در بخش ۲ و ۳ مفاهیم داده‌کاوی را معرفی می‌نماییم و سپس حوزه کاربردی داده‌کاوی را در داده‌های مرکز آمار ایران با استفاده از متدولوژی کشف دانش CRISP-DM در بخش 4 تشریح می‌نماییم.

در بخشهای بعدی مراحل این متدولوژی اجراء گردیده و دانشهای کشف شده استخراج گردیده و نشان داده می‌شوند.



از اواخر سال ۱۹۸۰ میلادی، انبارهای داده به عنوان نوع متمایزی از پایگاه‌های داده مورد استفاده اغلب سازمانها و شرکت‌های متوسط و بزرگ واقع شدند و در جهت رفع نیاز رو به رشد مدیریت داده‌ها و اطلاعات سازمانی که توسط پایگاه‌های داده سیستم‌های عملیاتی غیر ممکن بود، ساخته شدند. انبار داده به مجموعه‌ای از داده‌ها گفته می‌شود که از منابع مختلف اطلاعاتی سازمان جمع‌آوری، دسته‌بندی و ذخیره می‌شود. در واقع یک انبار داده مخزن اصلی کلیه داده‌های حال و گذشته یک سازمان می‌باشد که برای همیشه جهت انجام عملیات گزارش‌گیری و آنالیز در دسترس مدیران می‌باشد. انبارهای داده حاوی داده‌هایی هستند که به مرور زمان از سیستم‌های عملیاتی آنلاین سازمان (OLTP) استخراج می‌شوند، بنابراین سوابق کلیه اطلاعات و یا بخش عظیمی از آنها را می‌توان در انبار داده‌ها مشاهده نمود. از آنجائیکه انجام عملیات آماری و گزارشات پیچیده دارای بارکاری بسیار سنگینی برای سرورهای پایگاه داده می‌باشند، وجود انبار داده سبب می‌گردد که اینگونه عملیات تأثیری بر فعالیت برنامه‌های کاربردی سازمان (OLTP) نداشته باشد. همانگونه که پایگاه داده سیستم‌های عملیاتی سازمان (برنامه‌های کاربردی) به گونه‌ای طراحی می‌شوند که انجام تغییر و حذف و اضافه داده به سرعت صورت پذیرد، در مقابل انبار داده‌ها دارای معماری ویژه‌ای می‌باشند که موجب تسریع انجام عملیات آماری و گزارش‌گیری می‌شود (OLAP). [۲]

انبارهای داده سیستم‌های عملیاتی سازمان (OLTP) دارای نقاط ضعفی می‌باشند که انبارهای داده آنها را رفع می‌کنند. از جمله:

- بار پردازش گزارشات موجب کندی عملکرد برنامه‌های کاربردی می‌گردد.
- پایگاه‌های داده برنامه‌های کاربردی دارای طراحی مناسبی جهت انجام عملیات آماری و گزارش نیستند.
- بسیاری از سازمانها دارای بیش از یک برنامه کاربردی (منابع اطلاعاتی) می‌باشند، بنابراین تهیه گزارشات در سطح سازمان غیر ممکن می‌شود.
- تهیه گزارشات در سیستم‌های عملیاتی غالباً نیازمند نوشتن برنامه‌های مخصوص می‌باشد که معمولاً کند و پرهزینه هستند.

بنابراین، بسیاری از شرکت‌ها و سازمانها به این باور رسیده‌اند که گردآوری، سازمان دهی و یکپارچه سازی داده‌ها در یک مخزن داده برای مدیریت بهینه و اتخاذ تصمیمات کلان یک ضرورت می‌باشد

به طور کلی ساخت یک انبار داده، شامل مراحل اصلی زیر می‌باشد:

۱-۲- استخراج داده های تراکنشی از پایگاه های داده به یک مخزن واحد

شناخت منابع داده های سازمان و استخراج داده های ارزشمند از آنها یکی از اصلی ترین مراحل ایجاد انبار داده می باشد.

۲-۲- تبدیل داده ها

از آنجائیکه سیستم‌های اطلاعاتی و برنامه‌های کاربردی یک سازمان غالباً توسط افراد و پروژه‌های مختلف به مرور زمان در مواجهه با نیازهای جدید ساخته یا تغییر شکل داده می‌شوند، یکسان سازی آنها امری ضروری می‌باشد. در بسیاری از موارد نیز سیستم‌های اطلاعاتی در بسترهای مختلف پایگاه داده مانند Microsoft SQL Server، Oracle، Sybase، Microsoft Access و غیره طراحی گردیده‌اند. بررسی جداول، برقراری ارتباط بین فیلدها و یک شکل سازی داده‌ها در این مرحله صورت می‌پذیرد.

۳-۲- بارگذاری داده های تبدیل شده به یک پایگاه داده چند بعدی

بر خلاف پایگاه داده سیستم‌های عملیاتی (OLTP) که دارای معماری رابطه‌ای می‌باشند و از اصول نرمالیزه استفاده می‌کنند، طراحی انبار داده به شکلی ویژه بدون بهره‌گیری از اصول نرمالیزاسیون می‌باشد. در انبار داده فیلدها در جاهای مختلفی تکرار می‌شوند و روابط بین جداول کمتر به چشم می‌خورند. علت آن هم افزایش سرعت پردازش اطلاعات هنگام گزارشات و عملیات آماری می‌باشد.



۲-۴- تولید مقادیر از پیش محاسبه شده جهت افزایش سرعت گزارش گیری

مقادیر از پیش محاسبه شده را تراکم نیز می نامند. این مرحله توسط سیستمهایی نظیر Microsoft SQL Server Analysis Services بسیار ساده تر شده است. این تراکمها که در ابعاد مختلف انبار داده ساخته می شوند، موجب می شوند که سرعت انجام عملیات گزارش گیری به شکل محسوسی افزایش یابد. باید توجه داشت که عملیات ساخت این مقادیر بسیار زمان گیر بوده و نیازمند حافظه زیادی بر روی سرور است.

۲-۵- ساخت (یا خرید) یک ابزار گزارش گیری

پس از انجام مراحل فوق، می توان نسبت به ساخت یا خرید یک نرم افزار گزارش گیری تصمیم گیری نمایید. به طور معمول هزینه ساخت یک نرم افزار گزارش گیری، بالاتر از هزینه خرید آن از یک شرکت خارجی می شود.

با وجود اینکه کشور ما در بخش آمارگیری نسبت به سایر کشورهای دنیا ضعیف عمل می نماید، لذا سازمانها و نهادهای مختلف کشور می-بایست تمامی تلاش خویش را در این عرصه به کار بندد تا بتواند داده های آنلاین را به منظور تجزیه و تحلیل سریع در اختیار متخصصین این امر قرار دهد.

۳- داده کاوی

امروزه با حجم عظیمی از داده ها روبرو هستیم. برای استفاده از آنها به ابزارهای کشف دانش نیاز داریم. داده کاوی به عنوان یک توانایی پیشرفته در تحلیل داده و کشف دانش مورد استفاده قرار می گیرد. داده کاوی در علوم (ستاره شناسی، اقتصادی، ارتباطات، بازاریابی، جامعه شناسی،...) در تجارت (تبلیغات، مدیریت ارتباط با مشتری،...) در وب (موتورهای جستجو،...) در مسایل دولتی (فعالیت های ضد تروریستی،...) کاربرد دارد. عبارت داده کاوی شباهت به استخراج زغال سنگ و طلا دارد. داده کاوی نیز اطلاعات را که در انبارهای داده مدفون شده است، استخراج می کند. در واقع هدف از داده کاوی ایجاد مدل هایی برای تصمیم گیری است. [۳]

این مدلها رفتارهای آینده را براساس تحلیلهای گذشته پیش بینی می کنند. به کاربردن داده کاوی به عنوان اهرمی برای آماده سازی داده ها و تکمیل قابلیت های انبار داده، بهترین موقعیت را برای به دست آوردن برتریهای رقابتی ایجاد می کند. سیستم های بانک داده، نقشی کلیدی در سیستم های مدیریت و انبار داده، بازی می کنند. یک سیستم بانک داده، شامل فایل های بانک داده و سیستم های مدیریت بانک داده است.

اغلب تجارتها به تصمیم گیریهای استراتژیک و یا اتخاذ خط مشی های جدید برای خدمت رسانی بهتر به مشتریان نیاز دارند و براساس الگوها بدست آمده تصمیمات لازم اتخاذ می شود. در واقع ابزار داده کاوی، داده را می گیرد و یک تصویر از واقعیت به شکل مدل می سازد، این مدل روابط موجود در داده ها را شرح می دهد.

از نظر فرایندی فعالیتهای داده کاوی به سه طبقه بندی عمومی تقسیم می شوند: [۴]

• اکتشاف

فرایند جستجو در یک بانک داده برای یافتن الگوهای پنهان، بدون داشتن یک فرضیه از پیش تعیین شده درباره اینکه این الگو ممکن است چه باشد.



• مدل پیش بینی

فرایندی که الگوهای کشف شده از بانک داده را می گیرد و آنها را برای پیش بینی آینده به کار می برد.

• تحلیل‌های دادگاهی

به فرایند به کارگیری الگوهای استخراج شده برای یافتن عوامل داده ای نامعقول و متناقض مربوط می شود.

۴- کاربرد داده‌کاوی در مدیریت ارتباط با مشتری

مدیریت ارتباط با مشتری یک فرایند تجاری است که تمام جوانب مشخصه‌های مشتری را آدرس دهی می کند، دانش مشتری را به وجود می آورد، روابط را با مشتری شکل می دهد و برداشت آنها را از محصولات یا خدمات سازمان ایجاد می کند. مدیریت ارتباط با مشتری توسط چهار عنصر از یک چارچوب ساده تعریف شده است: دانش، هدف، فروش و خدمت. مدیریت ارتباط با مشتری با در نظر گرفتن اینکه چه محصولات یا خدماتی، به چه مشتریانی، در چه زمانی و از طریق چه کانالی عرضه شود، بهبود را در پی خواهد داشت. این مدیریت از اجزای مختلفی تشکیل شده است.

پیش از اینکه فرایند آغاز شود، شرکت باید اطلاعات مشتری را در اختیار داشته باشد. این اطلاعات می تواند از داده های داخلی مشتریان و یا از داده‌های منابع خارجی خریداری شده، به دست آید. برای داده‌های داخلی منابع مختلفی وجود دارد مانند پرسشنامه ها و بلاگ ها ، سوابق کارت اعتباری و....

منابع داده خارجی یا بانکهای داده خریداری شده مانند آدرسها، شماره تلفن ها، پروفایل های بازدید از وب سایتها کلیدی برای به دست آوردن دانش بیشتری از مشتری است.

بیشتر شرکتها، بانکهای داده‌ای عظیمی شامل داده های بازاریابی، منابع انسانی و مالی را دارا هستند. بنابراین، سرمایه گذاری در زمینه انبار داده، یکی از اجزای حیاتی در استراتژی مدیریت ارتباط با مشتری است.

پس از تهیه و تخصیص منابع داده، سیستم مدیریت ارتباط با مشتری باید با به کارگیری ابزارهایی مانند داده‌کاوی، داده ها را تجزیه و تحلیل کند. اعم از اینکه شرکت تکنیک‌های آماری سنتی را به کار می برد یا یکی از ابزارهای نرم‌افزاری مانند داده‌کاوی را، کارشناسان نیاز به فهم داده های مشتری و روابط تجاری دارند. بنابراین، داشتن افرادی متخصص که این داده ها را با ابزارهای مربوطه استخراج و به صورت اطلاعات درآورند، مهم است.

داده‌کاوی به مقایسه منابع و میزان مصرف آنها در برنامه ریزی منابع سازمان می پردازد. این روش توسط موضوعاتی مثل تجزیه و تحلیل رقبا، جهت دهی به بازار و استراتژی قیمت گذاری، در بحث رقابت گام می نهد. از طرفی شناخت نوع و میزان واکنش مشتریان به فعالیتهای اخیر سازمان با استفاده از داده‌کاوی، به اتخاذ تصمیمات آگاهانه و نتایج بهتر در زمینه ارائه و پیشنهاد یک نگرش به مشتریان جهت همسو کردن بینش آنان با سازمان منجر می شود و باعث اولویت بندی محصولات و خدمات قابل عرضه به مشتریان شده و بدین ترتیب می توان سطح موجودی کالا و خدمات در سازمان را کنترل نمود.

انتخاب مکان صحیح برای تبلیغ از مواردی است که داده کاوی در آنجا باید به هدف گذاری روی مشتریان پرداخته و به دنبال آنهایی باشد که مشابه مشتریان فعلی هستند. این کار اولاً باعث می شود نرخ پاسخگویی به آنها افزایش یافته و برنامه های ترفیع به طور مناسب به کار گرفته شود. ثانیاً توسط این روش می توان فهمید که سازمان یا واحد بازاریابی برای بهبود فروش، بهتر است تبلیغات خود را در کجا متمرکز کند. برای اجرای چنین برنامه هایی باید شباهت رفتاری افراد جدید را با مشتریان جاری شرکت اندازه گیری نمود. یعنی با استفاده از داده کاوی به گروه بندی رکورد‌های مشتریان بر مبنای طبقه بندیهای پیشین پرداخت. [۵]



۵- داده های مرکز آمار ایران

آمارگیری مرکز آمار ایران هر ساله در کشور انجام می‌شود و به طور متوسط هفتاد هزار خانوار ایرانی پرسشنامه مربوط به این آمارگیری را پر می‌کنند. فیلدهای این پرسشنامه بسیار متنوع می‌باشد و دارای چهار بخش اصلی می‌باشد که خصوصیات اجتماعی اعضای خانوار، مشخصات محل سکونت و تسهیلات و لوازم عمده زندگی، هزینه‌های خوراکی و غیر خوراکی خانوار و درآمدهای خانوار این چهار بخش می‌باشند. بخش سوم این پرسشنامه بسیار مفید است و شامل فیلدهای زیادی می‌باشد که تمام هزینه‌ها را تحت پوشش قرار می‌دهد از جمله هزینه‌های خوراکی، آموزشی، مسافرت، درمان و بهداشت، مسکن، مبلمان و لوازم خانگی، حمل و نقل، ارتباطات، خدمات فرهنگی و تفریحی، غذاهای آماده و هتل و ... هر یک از این موارد خود نیز شامل جزئیات بیشتری می‌باشد.

۵-۱- قسمت‌های پرسشنامه

برای آمارگیری یک پرسشنامه تهیه شده است که در سال‌های مختلف دستخوش تغییراتی شده است البته این تغییرات چندان زیاد نیست و ممکن است در یک سال چند فیلد جدید در یکی از قسمت‌های پرسشنامه اضافه شده باشد ولی قسمت‌های اصلی پرسشنامه ثابت بوده است. ما در ادامه قسمت‌های اصلی این پرسشنامه را شرح می‌دهیم. [۶]

همانطوریکه بیان شد این پرسشنامه دارای چهار قسمت اصلی است. این چهار قسمت به ترتیب:

- خصوصیات اجتماعی اعضای خانوار
- مشخصات محل سکونت و تسهیلات و لوازم عمده زندگی
- هزینه‌های خوراکی و غیر خوراکی خانوار
- درآمدهای خانوار

قسمت اول، شامل فیلدهای در مورد خصوصیات اجتماعی اعضای خانوار است که از آن جمله، جنس، سن، سطح تحصیلات، وضع فعالیت و وضع زناشویی است.

قسمت دوم مشخصات محل سکونت و تسهیلات و لوازم عمده زندگی را شامل می‌شود. از جمله فیلدهای این قسمت می‌توان به نحوه تصرف محل سکونت، تعداد اتاق، نوع اسکلت بنای محل سکونت، مصالح عمده بنای محل سکونت، نوع سوخت عمده مصرفی و تسهیلات عمده مورد استفاده خانوار از جمله: برق، گاز، کولر، حمام، آشپزخانه و اینترنت نام برد. همچنین استفاده از لوازمی از جمله: اتومبیل شخصی، رادیو، دوچرخه، پنکه، اجاق گاز، فریزر و غیره نیز در این قسمت ذکر شده است.

در قسمت سوم که هزینه‌های خانوار در آن ذخیره شده است فیلدهای زیادی وجود دارد و به همین دلیل این قسمت را به ۱۴ بخش تقسیم شده است. این چهارده بخش عبارتند از:



۱. هزینه‌های خوراکی خانوار در ماه گذشته
 ۲. هزینه‌های نوشیدنی‌های طبقه‌بندی نشده، دخانی‌ها و سایر مواد دخانی خانوار در ماه گذشته
 ۳. هزینه‌های پوشاک و کفش خانوار در ماه گذشته
 ۴. هزینه‌های مسکن، آب، فاضلاب، سوخت و روشنایی خانوار در ماه گذشته
 ۵. هزینه‌های مبلمان، لوازم خانگی و نگهداری معمول آن‌ها در ماه گذشته
 ۶. هزینه‌های بهداشتی و درمانی خانوار در ماه گذشته
 ۷. هزینه‌های حمل و نقل خانوار در ماه گذشته
 ۸. هزینه‌های ارتباطات خانوار در ماه گذشته
 ۹. هزینه‌های مربوط به خدمات فرهنگی و تفریحات خانوار در ماه گذشته
 ۱۰. هزینه‌های آموزش و تحصیل خانوار
 ۱۱. هزینه‌های غذاهای آماده، هتل و رستوران خانوار در ماه گذشته
 ۱۲. هزینه‌های کالاها و خدمات متفرقه خانوار در ماه گذشته
 ۱۳. تهیه و فروش کالاهای بادوام منزل، سایر هزینه‌های خانوار و انتقالات در ۱۲ ماه گذشته
 ۱۴. سرمایه‌گذاری خانوار در ۱۲ ماه گذشته
- این چهارده بخش این قسمت هستند که هر یک نیز شامل فیلدهای مربوطه هستند و برای جزئیات بیشتر باید پرسشنامه مراجعه نمود.
- قسمت چهارم درآمدهای خانوار را شامل می‌شود که سه بخش دارد. این سه بخش به ترتیب:

۱. درآمد پولی اعضای شاغل خانوار از مشاغل مزد و حقوق‌گیری
 ۲. درآمد پولی اعضای شاغل خانوار از مشاغل غیر مزد و حقوق‌گیری (آزاد)
 ۳. درآمدهای متفرقه خانوار در ۱۲ ماه گذشته
- در ادامه اطلاعاتی کلی در مورد جدول‌ها و فیلدهای آن‌ها در بخش متدولوژی و مراحل تحقیق ذکر می‌کنیم.

در این تحقیق سعی گردیده است که تمامی توان تکنیکی داده‌کاوی در این حوزه کاربردی و با تبعیت از متدولوژی CRISP-DM به عنوان متدولوژی به کار گرفته شود که در ادامه این متدولوژی و مراحل طی شده تشریح می‌گردد.

۶- متدولوژی و مراحل تحقیق

برای اجرای متدولوژی داده‌کاوی باید از چندین مرحله گذر کرد که بیان‌کننده فرآیند داده‌کاوی است. با وجود اینکه داده‌کاوی باعث کشف الگوها از حجم کثیری از داده می‌شود، توجه به این قضیه کفایت نمی‌کند و می‌بایست با اجرایی کردن اطلاعات و کسب دانش لازم، زمینه‌ای را برای ایجاد ارزش در سازمان فراهم نمود. این موضوع نشان‌دهنده یک چرخه خوب از فرآیند داده‌کاوی می‌باشد که دارای خصوصیت پیوستگی و تکرارپذیری است. چون نتایج داده‌کاوی در طول زمان دچار تغییر گشته و مدلهای به واسطه وجود سرعت زیاد در تغییر داده‌ها بعد از چندی کاربرد خود را از دست می‌دهند.

از آنجایی که هدف اساسی این پروژه دقیقاً به افزایش اثربخشی و کارایی تصمیمات سازمانی از طریق ایجاد بستر پشتیبانی تصمیم‌گیری مدیریت، در قالب مواد اولیه (داده‌ها) صحیح، دقیق و ارزشی و سیستم تولید دانش علمی (داده‌کاوی) از این داده‌ها بر می‌گردد. در نتیجه طبیعی است براساس فرآیند استاندارد داده‌کاوی CRISP میتوان پنج فاز برای این پروژه در نظر گرفت که عبارتند از: [۷]

۱. شناخت کسب و کار
۲. شناخت داده‌ها



۳. آماده سازی داد هها
۴. مدلسازی
۵. ارزیابی و بکارگیری

در ادامه مقاله، اقدامات انجام شده در این تحقیق بر مبنای فازهای روش تحقیق ارائه شده صورت گرفته است.

۱-۶- شناخت فضای کسب و کار

از آنجایی که میزان مصرف خوراکیهای مهم خانوارهای ایرانی برای صنایع غذایی کشور از اهمیت ویژه ای برخوردار می باشد، لذا پس از انتخاب این موضوع و با کمک گرفتن از داده های مرکز آمار ایران به دنبال انجام فرایند داده کاوی و کشف دانش در این زمینه می باشیم.

۲-۶- شناسایی داده ها

ساختار بانک اطلاعاتی مربوط به مصرف خوراکیهای خانوارهای ایرانی مرکز آمار ایران که پایه و اساس این تحقیق را شامل می گردد دارای ۷ فیلد در بخش خوراک خانوار می باشد، که عبارتند از: شماره مربوط به خانوار، کد کالا، طریق تهیه کالا، مقدار کالا به گرم، مقدار کالا به کیلو، قیمت واحد کالا به ریال، ارزش کالای مصرفی به ریال.

اما از آنجایی که میزان مصرف خوراکیها با توجه به درآمد آنان برای ما ملاک بود لذا از داده های بخش درآمد خانوارها نیز استفاده نمودیم. این قسمت دارای دو بخش اصلی است، درآمد افراد حقوق بگیر و درآمد افراد با حقوق آزاد.

هر کدام از این قسستها دارای ۱۶ فیلد می باشد که شامل، شماره خانوار، شماره ردیف اعضای شاغل، کد مربوط به شغل، کد فعالیت اصلی محل کار، میزان ساعت های کار در روز، تعداد روزهای کار در هفته، مجموع درآمد و در آمد خالص می باشد.

همچنین در ادامه به منظور بررسی پارامترهای فوق با توجه به استانهای کشور، با تجزیه و تحلیل کدهای مربوط به خانوارها، کد استانهای مربوطه را استخراج کرده و آنان را نیز که مشتمل بر ۳۰ استان کشور می باشد در نظر گرفتیم.

۳-۶- پیش پردازش داده ها

با استفاده از سیستم مدیریت پایگاه داده اکسس و نرم افزار Clementine این مجموعه داده ها به فرمت مناسب نرم افزار تبدیل و در آن ذخیره گردید. پس از آن برخی از فیلدها از جمله قیمت واحد کالا به ریال، ارزش کالای مصرفی و طریق تهیه از اطلاعات مربوط به خوراکیهای خانوار حذف گردید و سپس خوراکیهای هر خانوار و مقدار مصرف آنها به کیلوگرم محاسبه گردید.

همچنین در قسمت درآمدهای خانوار نیز پس از در نظر گرفتن درآمدهای خالص خانوار که شامل درآمدهای زندگی آنان منهای تمامی هزینه های آنان می باشد، پس از جمع نمودن حقوق اعضای حقوق بگیر و آزاد خانوار به تجزیه و تحلیل پرداختیم.

۴-۶- مدلسازی

در این بخش با توجه به هدف کشف دانش ممکن در مورد مصرف خوراکی های خانوارهای ایرانی با توجه به میزان درآمد آنان روش ها و ابزارهای مختلفی به کار گرفته شده و نتایج و دانشهای گوناگونی نیز استخراج گردید که در ادامه تشریح خواهد گردید:

الف) شناسایی کلی داده ها

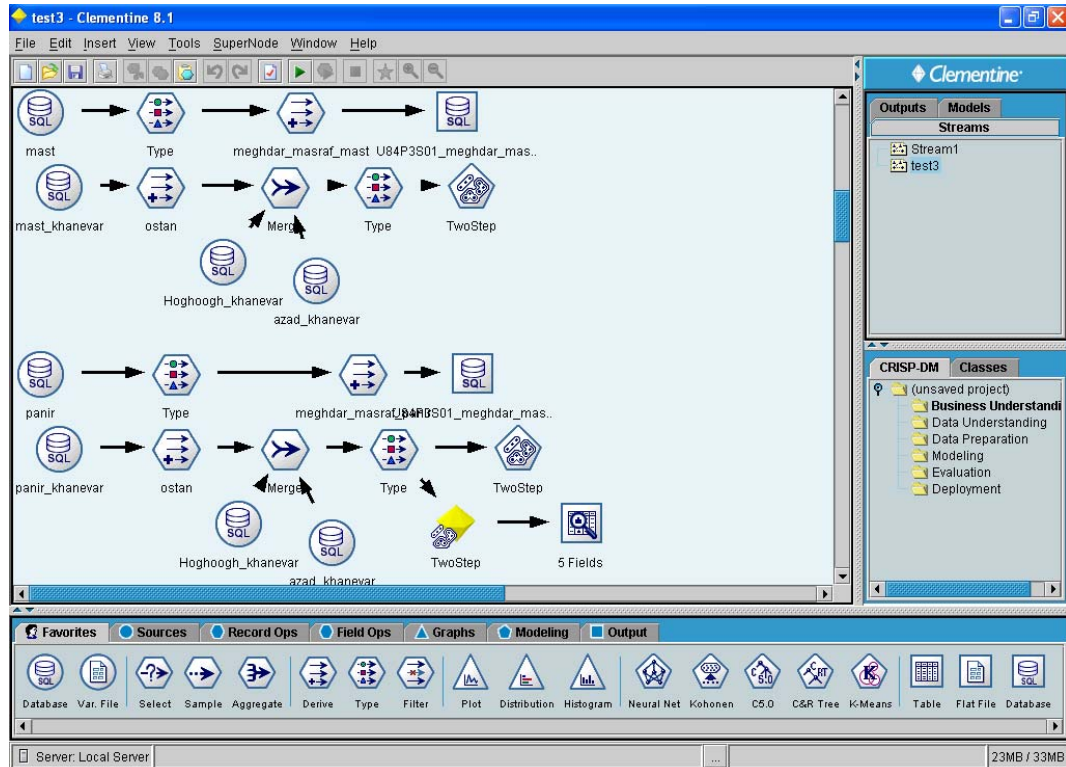
در ابتدا به منظور کم کردن تعداد اقلام مصرفی، آنها را به رده های متعدد زیر شکستیم.

لبنیات، میوه جات، روغن ها، گوشت دام، گوشت پرندگان، گوشت ماهی، قند و شکر، برنج و حبوبات



و به منظور جدا کردن این اقلام، با توجه به جدول مورد نظر در پایگاه داده و کد آن قلم جنس، از دستور SQL استفاده نموده ایم.

سپس از ابزارهای بصری در نرم افزار Clementine استفاده کرده و در نتیجه بر اساس روش کارموسوم در این نرم افزار یک گره منبع اطلاعاتی معرفی و فایبل ذخیره شده مشخص و تنظیمات انجام شد. شکل ۱ شمایی از فرآیند نرم افزار Clementine را نمایش می دهد.



شکل ۱- شمایی از پروژه در نرم افزار Clementine

ب) خوشه بندی

منظور از خوشه بندی، گروه بندی نمونه های مشابه با هم در یک نمونه داده ای می باشد. مسئله اساسی خوشه بندی عبارت است از: توزیع داده ها به K گروه مختلف که نقاط هر گروه با یکدیگر مشابه بوده و داده های گروه های مختلف با یکدیگر نامتشابه باشند.

به طور کلی خوشه بندی را می توان در موارد زیر به کار برد:

- تجزیه و تحلیل شباهت یا عدم شباهت: تجزیه و تحلیل اینکه کدام نقاط داده در یک نمونه به یکدیگر نزدیکند.



- کاهش بعد: داده‌های با ابعاد بالا با یک گروه (خوشه) جایگزین می‌شوند. که این کاربرد بیشتر به عنوان پیش پردازش داده‌ها مورد استفاده قرار می‌گیرد.

تجزیه و تحلیل خوشه‌ای روشی برای گروه‌بندی داده‌ها یا مشاهدات با توجه به شباهت یا درجه نزدیکی آنها است و از طریق آن داده‌ها یا مشاهدات به دسته‌های همگن اما متمایز از یکدیگر تقسیم می‌شوند.

خوشه‌بندی یک تکنیک داده کاوی غیر مستقیم است. یعنی در دیگر روش‌های داده کاوی مثل درخت تصمیم‌گیری و شبکه‌های عصبی، با یک مجموعه آموزشی شروع کرده و به کمک این مجموعه سعی می‌کنیم یک مدل ایجاد نمائیم که داده‌ها را بخش‌بندی کرده و سپس برای یک داده جدید دسته مناسب پیش‌بینی کنیم. اما در روش خوشه‌بندی هیچ دسته‌ای از قبل وجود ندارد و در واقع متغیرها به صورت مستقل و وابسته تقسیم نمی‌شوند. بلکه ما در اینجا فقط به دنبال گروه‌هایی از داده‌ها هستیم که به هم شبیه باشند و با کشف این شباهت‌ها بتوان رفتارها را بهتر شناسایی کرد و بر مبنای این شناخت بهتر تصمیم‌گیری نمود.

در واقع تکنیک‌های خوشه‌بندی برای این انجام می‌شود که مصرف‌کننده نهایی اطلاعات، یک منظر مناسب از اتفاقات که در پایگاه داده‌ها در حال رخ دادن است، داشته باشد.

در این پروژه نیز، با توجه به اینکه میزان ارقام مصرفی از تنوع بسیار زیادی برخوردار بود، لذا با استفاده از خوشه بندی تلاش نمودیم از تنوع آن بکاهیم و بیشتر بر اقلامی تمرکز داشته باشیم که اکثر خانوارهای ایرانی آنان را مصرف می کنند. جدول ۱ نتایج این خوشه بندیها را ارائه میدهد.

جدول ۱- استخراج ارقام مصرفی مهم خانوار با روش خوشه‌بندی

دسته بندی ارقام مصرفی خانوارهای ایرانی	نوع ارقام مصرفی خانوارهای ایرانی
لبنیات	شیر
	ماست
	کره
	پنیر
غلات	برنج
گوشت ماهی	ماهی شمال
	ماهی جنوب
گوشت دام	گوشت گوسفند، بره، بز، بزغاله، گاو، گوساله
	سسیس و کالباس
گوشت پرندگان	مرغ، خروس و جوجه
	روغن حیوانی



دسته بندی اقلام مصرفی خانوارهای ایرانی	نوع اقلام مصرفی خانوارهای ایرانی
روغن	روغن نباتی
میوه	میوه های جالیزی (هندوانه، طالبی، خربزه)
	سیب
	انواع خرما
خشکبار	انواع گردو، فندق، پسته، بادام
قند و شکر	انواع قند و شکر
حبوبات	انواع حبوبات

در مرحله بعدی پس از یافتن اقلام مهم مصرفی خانوارهای ایرانی با استفاده از روش سلسله مراتبی خوشه‌بندی قدم به قدم، بر اساس میزان درآمد و مصرف خوراکیهای مهم خانوارهای ایرانی به کیلو گرم و استانهای محل اقامت آنان، و از میان ۲ الی ۱۵ خوشه، بهترین حالت خوشه بندی در حداکثر ۴ خوشه ایجاد گردید. این خوشه‌ها بر اساس میزان درآمد خانوارهای استان و میزان مصرف خوراکیها متمایز هستند. مشخصات کامل خوشه‌ها در جدول ۲ ارائه گردیده است.

جدول ۲- نتایج حاصل از روش خوشه‌بندی

لبنیات				
نوع	استان	میانگین مصرف (کیلو گرم)	متوسط درآمد حقوق بگیر	متوسط درآمد آزاد
شیر	تهران - اردبیل	۲۷.۱۴۶	۲۷۳۲۸۷۲۳	۴۱۷۱۷۸۷۱
	چهارمحال و بختیاری، همدان، اصفهان	۱۷.۸۴	۲۱۱۸۳۹۰۲	۱۷۶۸۱۸۵۴
	آذربایجان غربی - خراسان رضوی	۱۶	۲۰۵۷۱۴۰۸	۱۹۲۸۳۹۵۷
ماست	تهران - اردبیل	۱۳	۲۶۳۲۷۱۲۳	۳۸۱۵۶۳۶۷
	گیلان - مازندران - آذربایجان غربی - خراسان رضوی	۸.۷۲	۲۰۷۹۸۶۷۸	۱۹۲۱۵۰۶۰



۱۸۶۰۶۵۳۸	۲۲۰۳۹۰۸۶	۱.۶۳	اردبیل - چهارمحال و بختیاری	پنیر
۱۸۶۴۳۸۵۶	۲۰۸۶۲۸۶۲	۱.۵۷	آذربایجان غربی - گلستان	
۴۲۲۹۵۳۰۶	۲۸۳۴۰۷۸۹۷	۲.۱۳۴	تهران - کردستان	
۲۰۹۵۴۷۰۳	۲۲۲۴۷۶۷۸	۰.۷	گلستان - خراسان رضوی	کره
۲۷۶۳۵۱۰۰	۲۲۱۸۹۳۷۸	۰.۶۲۲	تهران - مازندران	
۵۴۲۵۴۳۰۵	۳۵۵۴۸۴۸۱	۱.۴۱۲	اردبیل - آذربایجان غربی	
گوشت				
۴۶۸۵۴۵۵۸	۲۶۹۳۹۹۸۱	۷.۴۴	تهران - مازندران	گوشت دام
۲۰۰۳۱۴۱۲	۲۲۵۸۰۴۹۷	۴.۴۶۱	اردبیل - آذربایجان غربی - همدان - گیلان	
۲۰۲۷۸۴۰۱	۲۳۰۳۵۲۱۵	۴	چهارمحال و بختیاری - خراسان رضوی - گلستان	
۴۶۰۸۲۳۱۶	۲۹۱۸۴۰۲۳	۱.۲۸۲	تهران - هرمزگان - کهگیلویه و بویراحمد - ایلام	سسیس و کالباس
۲۱۲۵۶۳۰۳	۲۲۵۹۸۹۵۷	۱.۰۶	قزوین - گلستان	
۱۹۵۶۵۹۰۷	۲۲۴۹۷۷۸۷	۱.۰۶	گیلان - مازندران	
۳۸۳۶۱۴۶۹	۲۸۳۵۷۷۲۱	۹.۶	تهران - اردبیل	گوشت مرغ و جوجه
۱۹۵۶۵۷۱۴	۱۹۶۲۵۷۷۱	۵.۸۴۵	خراسان رضوی - چهارمحال و بختیاری همدان - آذربایجان غربی	
۱۸۵۶۳۲۶۲	۲۲۵۶۷۶۲۶	۶.۵	اصفهان - گیلان	
۳۸۴۲۹۴۳۲	۲۶۶۴۰۰۰۲	۲.۶۴	خوزستان - کرمان - فارس	گوشت ماهی جنوب
۲۲۶۵۰۳۸۶	۲۶۶۰۲۹۸۳	۵.۵	بوشهر - سیستان و بلوچستان - هرمزگان	
۱۹۳۱۳۳۶۲۵	۲۲۴۱۳۶۳۴	۶	اردبیل - گلستان - گیلان	گوشت ماهی شمال
۳۵۹۳۲۸۰۰	۳۰۶۳۵۵۳۳	۳.۶	تهران - ایلام	

۳۸۴۱۵۱۶۳	۲۸۵۵۹۷۷۰	۳.۵	مازندران - کردستان	
روغن				
۳۸۱۰۵۶۰۸	۲۶۳۹۴۲۵۷	۵	تهران - گیلان	روغن نباتی
۲۲۹۲۷۶۶۹	۲۴۹۹۵۰۷۶	۵.۵	مازندران - قزوین - اصفهان	
۱۹۲۳۰۳۳۴	۱۹۴۸۷۸۵۰	۷	اردبیل - آذربایجان غربی - گلستان - خراسان رضوی	
۳۴۱۲۵۰۵۰	۳۳۲۲۸۲۳۷	۱.۵	مازندران - قم	روغن حیوانی
۲۶۵۷۱۶۹۲	۱۴۴۷۶۶۶۲	۴.۰۳	آذربایجان غربی - سمنان - سیستان و بلوچستان	
میوه				
۳۱۸۹۹۲۰۹	۲۶۱۳۰۲۲۲	۱۰	اردبیل - اصفهان - همدان - کردستان	میوه های جالبزی
۲۳۶۸۲۶۹۶	۲۰۷۴۶۷۶۲	۴.۲	تهران - آذربایجان غربی	
۲۲۴۹۵۸۴۵	۲۴۷۳۵۶۰۹	۴	مازندران - گیلان - چهارمحال و بختیاری	
۱۹۹۹۷۸۸۰	۲۳۴۷۰۳۹۶	۵.۶	مازندران - گلستان	سیب
۴۱۹۷۰۹۲۷	۲۷۵۱۰۷۵۲	۱۰	تهران - اردبیل	
۲۰۶۶۳۹۶۳	۲۰۶۴۶۷۵۸	۵.۷	چهارمحال و بختیاری - همدان - خراسان رضوی - کردستان - کرمانشاه	
۱۱۳۱۰۲۷۲۵	۴۴۵۸۳۸۱۰	۱۰.۵	سیستان و بلوچستان - بوشهر - هرمزگان	خرما
۲۴۶۷۹۴۰۶	۲۴۹۷۷۷۳۰	۲.۵	مازندران - کرمانشاه - همدان - آذربایجان غربی	
۲۱۹۶۲۷۴۶	۲۶۰۸۸۴۹۴	۲.۴	تهران - گلستان - کردستان	
۱۷۰۳۸۱۰۶	۲۱۷۸۲۷۶۶	۲.۳	اردبیل - اصفهان - گیلان	
خشکبار				

۵۰۲۲۶۳۴۲	۳۵۳۵۹۲۲۰	۲۰۷	تهران - اصفهان - ایلام	خشکبار مغزها
۲۵۱۸۲۱۷۶	۲۷۵۰۹۰۸۴	۱۰۲	اردبیل - کردستان - گلستان	
۲۵۷۰۰۳۵۲	۲۸۳۲۶۹۹۹	۱	مازندران - گیلان	
قند و شکر				
۲۴۳۸۹۲۶۵	۲۲۷۳۲۳۰۲	۶	اردبیل - آذربایجان غربی - همدان	قند و شکر
۲۳۵۷۱۹۲۴	۲۲۰۱۱۸۶۳	۴	تهران - مازندران - گلستان	
غلات				
۰۱۷۶۰۳۱۶۲	۲۱۱۰۶۵۳۷	۲۶۰۸۲۷	اردبیل - خراسان رضوی - آذربایجان غربی	برنج
۳۲۷۱۸۴۶۲	۲۴۷۹۴۹۱۴	۳۳۰۲۲۲	تهران - همدان - سیستان و بلوچستان - کرمانشاه	
حبوبات				
۳۵۶۴۶۲۹۹	۲۹۰۳۱۱۵۱	۷۰۶۷۷	خراسان رضوی - کرمان	حبوبات
۲۰۰۳۲۶۴۱	۲۲۳۳۸۴۷۸	۵	مازندران - اصفهان - کردستان	
۱۹۱۳۷۰۷۹	۱۸۲۸۴۴۳۹	۴	تهران - گلستان	
۱۷۶۰۱۲۵۳	۲۰۸۴۱۶۳۳	۵۰۳۵	اردبیل - قزوین - آذربایجان شرقی	
۳۵۶۴۶۲۹۹	۲۹۰۳۱۱۵۱	۷۰۶۷۷	خراسان رضوی - کرمان	

براساس نتایج بدست آمده و با توجه به اینکه متوسط میزان درآمد خانوارها با شغل آزاد ۳۳۲۵۸۲۳۵ ریال و میزان متوسط درآمد خانوارهای حقوق بگیر ۲۸۰۲۳۹۳۰ ریال می باشد جدول فوق به وضوح متوسط درآمد اعضای خانوار و میزان مصرف آنان را تایید کرده و نمایش می دهد.

از آنجایی که هدف ما از این پروژه کمک به بازاریابی محصولات غذایی کشور بود لذا نتایج زیر استخراج گردید:

- از بین استانهای کشور در مجموع در بخش لیبیات، مردم در استانهای تهران، اردبیل و آذربایجان غربی با درآمد متوسط بیشترین میزان مصرف لیبیات را دارند.
- در بخش گوشت دام، مردم در استانهای تهران، مازندران، گیلان و گلستان با درآمد متوسط بیشترین میزان مصرف گوشت دام را دارند.
- در بخش گوشت پرندگان، مردم در استانهای تهران، اردبیل، اصفهان و گیلان با درآمد متوسط بیشترین میزان مصرف گوشت پرندگان را دارند.
- در بخش گوشت ماهی، به دلیل تنوع زیاد ماهی در کشور و به برکت دو دریای بزرگ خزر و خلیج فارس و عمان، تصمیم گرفتیم نوع ماهی ها را به دو دسته ماهی های شمال و جنوب تقسیم نموده و سپس به تجزیه و تحلیل بپردازیم.



لذا در بخش ماهی شمال استانهای گیلان، گلستان و اردبیل بیشترین میزان مصرف ماهی شمال را در سال به خود اختصاص داده اند. و در بخش ماهی جنوب، استانهای سیستان و بلوچستان، هرمزگان و بوشهر بیشترین میزان مصرف ماهی را در میان سایر استانهای کشور برخوردار بودند.

- در بخش روغن استانهای آذربایجان غربی و مازندران بیشترین میزان مصرف را با درآمد متوسط برخوردار بودند. نکته حائز اهمیت در این قسمت این است که مردم استانهای اردبیل- آذربایجان غربی- گلستان- خراسان رضوی مصرف با درآمد کم مصرف روغنشان بیش از سایر استانهای کشور می باشد.
- بیشترین مصرف میوه در استانهای تهران، اردبیل، همدان، کردستان و مازندران در میان مردم با درآمد متوسط قرار دارد.
- میزان خشکبار تقریباً در اکثر استانهای کشور به یک اندازه است ولیکن در تهران، اصفهان و ایلام بیشتر از سایر استانها است.
- مصرف قند و شکر در استانهای اردبیل، آذربایجان غربی و همدان از سایر استانهای دیگر کشور بیشتر می باشد.
- مصرف برنج در استانهای تهران، همدان، سیستان و بلوچستان و کرمانشاه در میان مردم با درآمد متوسط بیشتر از سایر استانها می باشد.
- مصرف حبوبات در استان خراسان رضوی و کرمان بیشتر از سایر استانهای کشور می باشد.

با توجه به نتایج فوق به وضوح درمی یابیم که میانگین میزان مصرف خوراکیهای مهم خانوارها با در نظر گرفتن متوسط درآمدهای آنان در کدام استانها بیشتر قرار دارد. از این رو این اطلاعات علاوه بر کمک به برنامه ریزی بهتر کشور در زمینه مواد خوراکی به شرکتها و سازمانها در امر بازاریابی نیز کمک ویژه ای می نماید. بازاریابان می توانند بر مبنای اطلاعات فوق به برنامه ریزی های صحیح تری دست یابند.

پ) استخراج قوانین انجمنی

یکی از روشهای اساسی استخراج دانش در قالب قواعد منطقی روش قوانین انجمنی است. قوانین انجمنی یعنی استخراج قوانینی برای پیش بینی اتفاق افتادن یک عنصر در تراکنشهای بعدی با توجه به اتفاق افتادن و یا اتفاق نیفتادن عناصر دیگر می باشد. معیارهای ارزیابی این قوانین عبارتند از:

• Support

نسبت تعداد تراکنشهایی که پیش شرایط قوانین استخراج شده را برآورده می کنند به کل تراکنشها .

• Confidence

نسبت تعداد تراکنشهایی که در آنها هم پیش شرطها و هم نتایج قوانین استخراج شده رخ داده اند به کل تراکنشها (تصدیق کننده های قوانین) در این زمینه و با استفاده از الگوریتم GRI به استخراج قوانین پرداختیم. از بین قوانین فوق ما تنها به چند قانون در هریک از گروههای لبنیات، میوه جات، روغن ها، گوشت دام، گوشت پرندگان، گوشت ماهی، قند و شکر، برنج و حبوبات اشاره می کنیم.

۱. بر مبنای قوانین تجمعی در بخش برنج افرادی که درآمد آزاد آنان (کسانی که شغل آزاد دارند) ۳۹۰۹۵۰۰۰۰ ریال که بالاتر از میانگین درآمد آزاد خانوارهای ایرانی است، مصرف برنجشان کمتر از ۲۰ کیلو و در استان همدان قرار دارند و کسانی که مصرف برنجشان کمتر از ۲۰ کیلو و درآمد آزادشان کمتر از ۲۴۵۵۰۰۰۰ ریال به عبارتی در رده خانوارهایی که درآمد آزادشان پایین تر از میانگین دیگر خانوارهاست در استان آذربایجان غربی قرار دارند.



به عبارتی در استان آذربایجان غربی افراد با درآمدهای کمتر مصرف برنجشان بیشتر از سایر استانهاست.

۲. در بخش لبنیات خانوارهایی که مقدار مصرف پنیرشان کمتر از ۰.۴۷۰ کیلوگرم و بیشتر از ۰.۲ کیلوگرم است و مصرف شیر آنان بین ۶.۵ کیلو و ۱۸ کیلو و مصرف کره آنان کمتر از ۰.۸۲ کیلوگرم و مصرف ماستشان کمتر از ۱۱ کیلوگرم است و درآمد افراد حقوق بگیرشان بیشتر از ۱۷۰۵۰۰۰ ریال میباشد در استان مازندران قرار دارند و خانوارهایی که مصرف شیرشان بیشتر از ۵۵.۵ کیلو و کره شان هم بیشتر از ۱۰.۵ کیلو در استان خراسان جنوبی قرار دارند.

۳. خانوارهایی که حقوق آزادشان از ۱۲۰۸۲۵۰۰ ریال کمتر و از ۲۰۵۰۰۰۰ ریال بیشتر و مقدار مصرف سسبیس کالباسشان از ۱ کیلوگرم کمتر و مقدار مصرف گوشتشان ۲ کیلو در استان تهران قرار دارند.

به عبارتی این امر نشان می دهد که افرادی که درآمدشان پایین تر از حد میانگین قرار دارد مصرف سسبیس و کالباسشان نیز کم است.

۴. خانوارهایی که مقدار مصرف روغن حیوانیشان کمتر از ۱.۶ کیلوگرم و مقدار مصرف روغن نباتی کمتر از ۲.۸ کیلوگرم و بیشتر از ۱.۶ کیلوگرم باشد و درآمد آزادشان بیشتر از ۹۹۷۰۰۰۰۰ ریال و کمتر از ۵۶۷۰۷۵۰۰ ریال می باشد، در استان مازندران قرار دارند. به طور کلی افراد در استان مازندران مصرف روغنشان زیاد است. و خانوارهایی که مقدار مصرف روغن حیوانی بیشتر از ۴.۲۵ کیلوگرم و کمتر از ۱۰.۵ کیلو و مقدار مصرف روغن نباتی آنان هم کمتر از ۱۰.۵ کیلو است و درآمد افراد حقوق بگیرشان کمتر از ۳۷۶۵۰۰۰ ریال و درآمد آزادشان بیشتر از ۱۲۸۵۵۰۷۵۲ ریال می باشد در استان آذربایجان غربی قرار دارند. در استان آذربایجان غربی افراد حقوق بگیر با درآمد کمتر بیشتر از روغن استفاده می کنند.

۵. خانوارهایی که درآمد افراد حقوق بگیرشان ۱۰۱۹۵۸۹۲۴ ریال و کمتر ۱۶۴۰۴۳۳۵۲ ریال و درآمد آزادشان هم بیشتر از ۲۴۷۴۰۰۰۰ ریال، مقدار مصرف مرغ و جوجه آنان بیشتر از ۷ کیلو می باشد در استان تهران قرار دارند و خانوارهایی که مقدار مصرف گوشت مرغ و جوجه آنان کمتر از ۰.۲۵۰ کیلوگرم و درآمد حقوق بگیرشان بیشتر از ۴۳۰۴۰۸۰۲ ریال می باشد در استان سیستان و بلوچستان قرار دارند. به عبارتی دیگر، در استان تهران مصرف گوشت و مرغ و جوجه در تمامی اقشار جامعه زیاد است ولی در استان سیستان و بلوچستان در میان مصرف گوشت و مرغ کم است حتی در میان افراد با درآمدهای بالا، این امر به وضوح دیده می شود.

۶. خانوارهایی که درآمد افراد حقوق بگیرشان کمتر از ۲۰۲۰۸۰۰۲ ریال و بیشتر از ۱۳۸۰۹۳۵۶ ریال و درآمد آزادشان بیشتر از ۳۴۲۲۵۰۰۰۰ ریال است مصرف ماهی جنوب کمتر از ۰.۹ کیلوگرم و ماهی شمال کمتر از ۲.۵ کیلو در استان اردبیل قرار دارند به طور کلی مصرف ماهی در میان اقشار کم درآمد استان اردبیل پایین می باشد.

۷. خانوارهایی که درآمد حقوق بگیرشان کمتر از ۳۴۴۴۸۷۹۴ ریال و بیشتر از ۲۳۹۶۹۳۶۵ ریال مقدار مصرف خرماشان بیشتر از ۱۷.۵ کیلو در استان هرمزگان قرار دارند و خانوارهایی که مقدار مصرف سببشان بیشتر از ۳۶.۵ کیلو و مقدار مصرف خرماشان کمتر از ۵.۵ کیلوگرم است و درآمد حقوق بگیرشان بیشتر از ۷۸۱۶۰۰۰۰ ریال و درآمد آزادشان کمتر از ۲۴۲۶۰۰۰۰ ریال و بیشتر از ۱۷۹۲۷۵۰۰ ریال در استان تهران قرار دارند.

به طور کلی مصرف خرما در استان هرمزگان در تمامی اقشار جامعه زیاد است و در استان تهران نیز مصرف خرما کمتر از استان هرمزگان (به طور کلی استانهای جنوبی کشور) است ولی مصرف دیگر میوه ها از جمله سیب در استان تهران در تمامی اقشار جامعه زیاد است.



۸. کسانی که درآمد آزادشان کمتر از ۲۴۷۶۴۲۸.۵ ریال و بیشتر از ۱۲۱۰۰۰۰۰۰ ریال و درآمد حقوق بگیرشان کمتر از ۱۰۳۳۲۰۰ ریال مقدار مصرف خشکبارشان کمتر از ۱.۶ کیلوگرم و بیشتر از ۰.۴ کیلوگرم و در استان گلستان قرار دارند. به عبارت دیگر، مصرف خشکبار در تمامی اقشار جامعه در استان گلستان کمتر از ۲ کیلو است.
۹. خانوارهایی که درآمد حقوق بگیرشان بیشتر از ۱۶۴۰۴۳۳۵۲ ریال و مقدار مصرف قند و شکرشان کمتر از ۱.۵ کیلوگرم در استان اصفهان قرار دارند.
۱۰. خانوارهایی که درآمد حقوقشان بیشتر از ۱۰۸۱۴۸۷۳۶ ریال و کمتر از ۱۶۴۰۴۳۳۵۲ ریال است مقدار مصرف قند و شکرشان ۲.۳۳ کیلو و در استان تهران قرار دارند. به طور کلی در استان تهران در میان اقشار کم درآمد جمع مصرف قند و شکر ۲.۳۳ کیلو می باشد.
۱۱. خانوارهایی که درآمد آزادشان کمتر از ۱۹۶۲۵۰۰۰ ریال و بیشتر از ۶۹۶۲۵۰۰ ریال است و درآمد حقوق بگیرشان بیشتر از ۹۹۸۰۰۰۰ ریال و میزان مصرفشان ۰.۲ کیلو در استان کردستان قرار دارند.

در این قسمت تلاش نمودیم برخی از قوانین را به صورت نمونه نشان دهیم و همانطوریکه قبلا اشاره نمودیم نتایج تحلیل‌های انجام شده به منظور کمک به بررسی ارتباطات و تهیه الگوهای کاربردی در حوزه بازاریابی و تاثیرات آن در مدیریت ارتباط با مشتریان کارخانجات مواد غذایی و خصوصا برنامه‌ریزیهای بهتر کشور می‌تواند به کار رود.

۷- نتیجه گیری

در دنیای کسب و کار، داده‌های مشتریان به عنوان یک سرمایه استراتژیک مطرح است و ضرورت درک مجموعه داده‌های حجیم از مشتریان بر کسی پوشیده نیست. از اینرو حوزه بازاریابی هم به عنوان نبض حرکت سازمان از مهمترین بخش‌هایی است که نیازمند تصمیم‌گیری صحیح بر مبنای دانش حاصل از داده‌های مشتریان دارد، تا بدینوسیله بتواند به علایق و خواسته‌های گوناگون و متفاوت آنها پاسخ مناسب داده و از طرفی بهترین استراتژی‌های را در جهت افزایش بهره‌وری انتخاب نماید. بهترین و قویترین متدولوژی برای تحلیل این حجم بزرگ از داده‌ها و استخراج الگوهای رفتاری مشتریان استفاده از تکنیک داده‌کاوی است. این تحقیق نیز بر اساس متدولوژی CRISP_DM پیگیری و اجرا گردید و در نهایت چندین دانش کاربردی استخراج گردید.

اطلاعات استخراج شده، به وضوح میزان مصرف خوراکیها را در استانهای کشور در میان مردم با درآمد متوسط بیان می کند، بنابراین، این اطلاعات و دانش کاربردی علاوه بر کاربرد در بخش بازاریابی می‌توانند به سیاست‌گذاریهای کشور در امور غذایی مردم نیز کمک شایانی را انجام دهد. دولتمردان ما می‌توانند با استفاده از دانش حاصل از داده‌کاوی برنامه‌ریزیهای اقتصادی بلند مدت کارایی را صورت دهند.

۸- پیشنهاد

آمارگیری مرکز آمار ایران تقریبا هر ساله صورت می‌گیرد. لذا داده کاوی تنها دانش مفیدی است که می‌تواند بهترین اطلاعات کاربردی را استخراج نماید. لذا با استفاده از داده‌های مرکز آمار ایران می‌توانیم به پیشرفت کشورمان در تمامی زمینه‌های به ویژه وضع معیشتی مردم سرعت بیشتری دهیم.



مراجع

1. David Olson, "*Managerial Issues of Enterprise Resource Planning*", Mc Graw Hill, 2003.
2. T. Barron, "*OLAP Goes Online*," InformationWeek, September 20, 1999.
3. Chris Rygielski, "*Data Mining Techniques for Customer Relationship Management*," Technology in society, 2002
4. M. Kantardzic, "*Data Mining; concepts, models, methods, and algorithms*", 1998.
5. Michael J.A.Berry, Gordon S.Linoff, "*Data mining Techniques for Marketing, Sales, and Customer relationship Management*", Second edition, Wiley, 2004

۶. سازمان مدیریت و برنامه ریزی کشور، مرکز آمار ایران، گزارش آمارگیری از هزینه و درآمد خانوارهای ایرانی

7. Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler), "*CRISP-DM 1.0*", 2000

