

## فصل اول

### مقدمه

شناسایی مشتریان ناراضی و دادن انگیزه ماندن به آنها، یکی از روش‌های مدیریت نارضایتی مشتری می‌باشد. تحقیقات زیادی درخصوص بدست آوردن مشتری انجام شده است که ثابت کرده‌اند بدست آوردن مشتری بسیار گران‌تر از نگهداری او می‌باشد [Reinartz and Kumar, 2003]. هدف از این مطالعه آن است که با به حداقل رساندن میزان ترک مشتریان، سود شرکت‌ها، به حداکثر رسانده شود.

### 1-1- شرح مسئله پژوهشی

مدیریت ارتباط مشتری<sup>1</sup>، از مجموعه‌ای از پروسس‌ها تشکیل شده است و استراتژی‌هایی را برای کسب و کار پشتیبانی می‌کند [E. W. T. Ngai *et al*, 2009]. این استراتژی‌ها، برای برقراری ارتباطی سودمند و طولانی مدت با مشتریان مناسب می‌باشند. پایه و اساس هر استراتژی را داده‌های مشتری و ابزارهای تکنولوژی اطلاعات تشکیل داده‌اند [E. W. T. Ngai *et al*, 2005]. مدیریت ارتباط مشتری، روشی برای برقراری ارتباط بهتر با مشتری، برای بدست آوردن، نگهداری، افزایش وفاداری، سودآوری و در نتیجه افزایش دوره زندگی او می‌باشد [J. W. Kincaid, 2003]. به طور کلی، مدیریت ارتباط مشتری، بازاریابی، فروش و خدمات را برای افزایش ارزش دوره زندگی مشتریان تکمیل می‌کند و تلاش می‌کند مشتریان را نگهداری کند [A. Parvatiyar and J. N. Sheth, 2001].

---

<sup>1</sup> CRM (Customer relationship management)

بازاریابی همراه با رشد سریع اینترنت، افزایش یافته است و راه ارتباطی بین مشتریان و شرکت‌ها را تغییر داده است. در گذشته، تمرکز شرکت‌ها روی فروش محصولات و سرویس‌ها بود، اما امروزه تمرکز آن‌ها بیشتر روی مشتریان است [A. Berson et al, 2000]. شرکت‌ها نیاز به راه‌هایی برای بدست آوردن سهم بازار و زیاد کردن آن، در حین کاهش هزینه دارند. در نتیجه، شرکت‌های موجود باید در ارتباط با مشتریانشان تجدید نظر کنند.

مدیریت ارتباط مشتری به دو قسمت عملیاتی و تحلیلی تقسیم می‌شود [Z. He et al, 2004]. مدیریت ارتباط مشتری به صورت عملیاتی، مربوط به خودکارسازی پروسه‌های کسب و کار می‌باشد، در صورتی که مدیریت ارتباط مشتری به صورت تحلیلی، به تحلیل ویژگی‌ها و رفتار مشتری می‌پردازد تا روش‌هایی را برای مدیریت روابط خود با مشتری پشتیبانی کند و به یک سازمان در تشخیص بهتر و اختصاص کارا تر منابع به مشتریان سودمندتر کمک کند.

چهار بعد مدیریت ارتباط مشتری عبارتند از شناسایی مشتری، جذب مشتری، نگهداری مشتری و در نهایت، افزایش تعداد مشتری. اصلی‌ترین بخش مدیریت ارتباط مشتری، نگهداری مشتری می‌باشد که از بخش‌های زیر تشکیل شده است:

۱. بازاریابی یک به یک<sup>۱</sup>: مبارزات بازاریابی شخصی که با تحلیل، کشف و پیش‌بینی تغییر رفتار مشتری، پشتیبانی می‌شود.

۲. برنامه وفاداری<sup>۲</sup>: فعالیت‌های پشتیبانی یا مبارزاتی برای برقراری ارتباط طولانی مدت با مشتریان، از قبیل تحلیل نارضایتی<sup>۳</sup>، بررسی رضایت مشتری، بررسی کیفیت خدمات و نمره‌های اعتباری<sup>۴</sup> مشتریان می‌باشد.

مشتریانی که خرید آن‌ها کاهش یافته است، به دو قسمت تقسیم می‌شوند: مشتریان اختیاری و مشتریان غیراختیاری [J. Hadden et al, 2007]. مشتریان غیراختیاری آن‌هایی می‌باشند که شرکت به دلایل متفاوت، سرویس‌دهی به آن‌ها را لغو می‌کند و به راحتی قابل شناسایی می‌باشند. مشتریان اختیاری آن‌هایی می‌باشند که خود ارتباطشان با شرکت را به پایان می‌رسانند و به راحتی قابل شناسایی نمی‌باشند. این دسته از مشتریان نیز به دو دسته تقسیم می‌شوند: تصادفی و عمدی. از جمله دلایل تصادفی، تغییر شرایط مالی مشتریان یا جابجایی آن‌ها به مکانی دیگر می‌باشد که به دلیل تغییر شرایط محیطی رخ می‌دهد. این گروه درصد کمی از این دسته را تشکیل می‌دهند. رفتن مشتریان از روی عمد به سمت رقبا، همان مسئله‌ایست که مدیران سعی می‌کنند راهی برای آن بیابند. این نارضایتی زمانی رخ می‌دهد که مشتری سفارشش را به شرکت‌های رقیب می‌-

<sup>1</sup> One-to-one marketing

<sup>2</sup> Loyalty program

<sup>3</sup> churn

<sup>4</sup> credit scoring

دهد. این نارضایتی می‌تواند به دلایل فنی، اقتصادی، و غیره رخ دهد. دلایل فنی، مانند زمانی که مشتری با رقبای روبرو می‌شود که جدیدترین محصولات را ارائه می‌کنند در حالی که این شرکت نمی‌تواند آن‌ها را تهیه کند. دلایل اقتصادی، مانند زمانی که رقبای محصولات این شرکت را با قیمت پایین‌تر به فروش می‌رسانند. دلایل دیگر می‌تواند عوامل کیفیتی، پوشش ناچیز، یا داشتن تجربه بد با مراکز تلفنی مربوطه و غیره باشد که باید این دلایل را شناسایی کرد [H. Kim and C. Yoon, 2004].

تلاش‌های مدیریت نارضایتی به دلایل زیر نباید برای کل مشتریان باشد: اول آن که همه مشتریان ارزش نگهداری ندارند و دوم آن که نگهداری مشتریان هزینه‌بر است و تلاش برای نگهداری مشتریانی که قصد رفتن به سمت رقبای ندارند، باعث از دست دادن منابع می‌شود.

شرکت‌ها نیاز دارند مشتریانشان را بشناسند و آن‌ها را راضی نگهدارند [D. Liu and Y. Shih, 2005]. این مسئله باعث رقابت سازمان‌ها برای توسعه روش‌های بازاریابی، برای شناسایی نیازهای مشتریان، افزایش رضایت و نگهداری آن‌ها شده است [D. Liu and Y. Shih, 2005]. برای شرکت‌هایی که رقبای زیادی دارند، فروش بیشتر نمی‌تواند در طولانی مدت سودمند باشد و نیاز به روش‌های کاراتری برای افزایش درآمد می‌باشد [G. Canning, 1982]. نتایج نشان می‌دهد که ۵۳ درصد از دلایل نارضایتی مشتریان، عواملی جز هزینه می‌باشد. بنابراین به روش‌های قویتری برای نگهداری آن‌ها نیاز است [B. H Chu et al, 2007].

یک راه برای مدیریت نارضایتی مشتری، شناسایی مشتریان ناراضی و دادن انگیزه ماندن به آن‌ها می‌باشد. در نتیجه، سیستم‌های اطلاعاتی، تصمیم‌گیرندگان بازاریابی را برای بازاریابی مشتریان مناسب، پشتیبانی می‌کنند. ثابت شده است که نگهداری و رضایت مشتریان موجود، از جذب مشتریان جدید با احتمال نارضایتی، سودمندتر است. شرکت‌ها باید بتوانند با به حداقل رساندن میزان نارضایتی مشتریانشان، آن‌ها را نگهداری کنند و سود شرکت خود را به حداکثر برسانند [J. Hadden et al, 2007].

## 1-2- اهداف تحقیق

اهمیت این تحقیق در یافتن مشتریان ناراضی و نگهداری آن‌ها می‌باشد. هدف از این تحقیق آن است که علاوه بر افزایش ارزش دوره زندگی مشتریان، از کاهش سود آن‌ها جلوگیری کند. شرکت‌ها باید از رفتن مشتریانشان به سمت رقبای جلوگیری کنند. روش‌هایی که برای نگهداری مشتریان استفاده می‌شود، هزینه‌بر است، بنابراین نیاز نیست که برای نگهداری همه مشتریان هزینه صرف کرد. باید بتوان مشتریان ناراضی را شناسایی کرد و از آن‌ها نگهداری کرد. با استفاده از روش‌های داده کاوی، می‌توان پیش‌بینی دقیق‌تری از نارضایتی مشتریان داشته و سود شرکت‌ها، کارخانجات، کسب و کار، تجارتخانه‌ها و غیره را افزایش داد.

### 3-1- طرح پیشنهادی برای نگهداری مشتریان

در این تحقیق، ابتدا ارزش دوره زندگی مشتریان در طول چهار ماه محاسبه می‌شود. سپس مشتریان به چهار گروه تقسیم‌بندی می‌شوند. گروه اول مشتریان کاملاً وفادار، گروه دوم مشتریان نسبتاً وفادار، گروه سوم مشتریان نسبتاً ناراضی و گروه چهارم مشتریان کاملاً ناراضی می‌باشند.

پس از تقسیم‌بندی مشتریان به گروه‌های متفاوت، سیاست‌هایی برای نگهداری آن‌ها اعمال می‌شود. برای این منظور، کل فعالیت‌های انجام شده برای نگهداری مشتری را در رویدادهای ثبت شده، وارد کرده و دوباره ارزش دوره زندگی آن‌ها را بدست می‌آورد. سپس فرآیند مربوط به مشتریانی که ارزش دوره زندگی آن‌ها بیشتر از ۱۰۱٪ افزایش یافته است، را در نظر گرفته و فرآیند مربوط به آن‌ها را با استفاده از الگوریتم  $\alpha$  از فرآیند کاوی کاوش می‌کند. در نتیجه برای هر یک از گروه‌ها بهترین روش برای نگهداری آن‌ها مشخص می‌شود. در نهایت، با ورود هر مشتری جدید، احتمال فرار او پیش بینی می‌شود و با توجه به گروهی که در آن قرار می‌گیرد، تکنیکی برای نگهداری او استفاده می‌شود.

### 4-1- کاربرد تحقیق در جامعه کنونی

در این تحقیق مشتریان ناراضی شناسایی و از آن‌ها نگهداری می‌شود. به طور کلی، همه فروشگاه‌های بزرگ، کارخانجات، شرکت‌ها و سایر مؤسسات بزرگ، می‌توانند این تحقیق را برای نگهداری مشتریان ناراضی بکار گیرند. با استفاده از روش C5.0 از داده کاوی، می‌توان مشتریان ناراضی را شناسایی کرد. الگوریتم آلفا در فرآیند کاوی می‌تواند برای شناسایی بهترین فرآیند برای نگهداری مشتریان، بکار گرفته شود.

کاوش بهترین فرآیند برای نگهداری مشتریان را می‌توان برای مؤسساتی بکاربرد که برای نگهداری مشتریان، از تکنیک‌های متفاوتی استفاده می‌کنند. این مؤسسات باید بتوانند این تکنیک‌ها را در فایل ثبت تراکنش وارد کنند. این روش، برای شرکت‌های بزرگی مانند شرکت ایرانسل، می‌تواند از کارایی بالایی برخوردار باشد. این مسئله به این دلیل می‌باشد که شرکت‌هایی مانند ایرانسل، تکنیک‌های زیادی برای نگهداری مشتریان استفاده می‌کنند، اما نمی‌دانند کدام یک از این تکنیک‌ها از کارایی بالاتری برخوردار است. چنین شرکت‌هایی می‌توانند پس از شناسایی مشتریان ناراضی، از آن‌ها نگهداری کنند.

این تحقیق، از داده‌های مربوط به فروشگاه زنجیره‌ای رفاه استفاده می‌کند. ابتدا داده‌های مربوط به بیش از هزار مشتری را در نظر گرفته، و مشتریانی که داده‌های آن‌ها ناقص می‌باشد را حذف می‌کند. از نتیجه فیلتر کردن این

مجموعه داده، داده‌های مربوط به ۷۰۰ مشتری را مورد ارزیابی قرار می‌دهد. این مشتریان در طول شش ماه مورد ارزیابی قرار گرفتند و در نهایت پس از شش ماه، بهترین فرآیندها برای نگهداری مشتریان، استخراج گردیدند. پس از کاوش بهترین فرآیندها برای نگهداری مشتریان متفاوت، این نتایج مورد تست واقع شدند. درآمد مربوط به این فروشگاه، در طول دو ماه ارزیابی شد. نتایج نهایی، حاکی از بهبود ۷۶٪ در درآمد مربوط به این شرکت بود.

## 5-1- نتیجه‌گیری

در این تحقیق مشتریان ناراضی شناسایی و از آن‌ها نگهداری می‌شود. هدف از این تحقیق آن است که علاوه بر افزایش ارزش دوره زندگی مشتریان، از کاهش سود آن‌ها و خرید آن‌ها از رقبا جلوگیری کند. با استفاده از روشهای داده کاوی، می‌توان پیش‌بینی دقیق‌تری از نارضایتی مشتریان داشته و سود شرکت‌ها، کارخانجات، کسب و کار، تجارتخانه‌ها و غیره را افزایش داد. به طور کلی، همه فروشگاه‌های بزرگ، کارخانجات، شرکت‌ها و سایر مؤسسات بزرگ، می‌توانند این تحقیق را برای نگهداری مشتریان ناراضی بکار گیرند.

## 6-1- ساختار پایان‌نامه

جزئیات بیشتر در مورد نگهداری مشتری، در فصل‌های بعدی قید می‌شود. در فصل دوم، تاریخچه کارهای انجام شده روی شناسایی مشتریان ناراضی، بررسی شده است. موضوعات مرتبط با تحقیق در فصل سوم مورد بررسی قرار گرفته است. در فصل چهارم، طرح پیشنهادی برای نگهداری مشتری بررسی شده است. فصل پنجم، نتایج به دست آمده و تجزیه و تحلیل آن‌ها را مورد بررسی قرار داده است. فصل ششم، نتیجه‌گیری می‌باشد که شامل جمع‌بندی کلی نتایج، ارزیابی روش نسبت به کارهای انجام شده و پیشنهادات می‌باشد. منابع استفاده شده در این تحقیق، در فصل آخر ذکر شده است.

## فصل دوم

### تاریخچه

در این فصل، علاوه بر بررسی تاریخچه تحقیقات انجام شده روی شناسایی مشتریان ناراضی، ارتباط موضوع تحقیق با کارهای قبلی و اهمیت و هدف آن نیز مورد بررسی قرار می‌گیرد. تحقیقات انجام شده، شامل تکنیک-هایی می‌باشد که تاکنون برای شناسایی مشتریان ناراضی استفاده شده است. یکی از این تکنیک‌ها، داده‌کاوی می‌باشد که محققان متفاوت، تکنیک‌های متفاوت داده‌کاوی را مورد بررسی قرار داده‌اند. روش دیگری که برای شناسایی این مشتریان بکار گرفته شده است، اندازه‌گیری ارزش دوره زندگی آن‌ها می‌باشد که با محاسبه این مقدار، این مشتریان را شناسایی می‌کنند. تحقیقات دیگری برای کاوش فرآیندها از ثبت تراکنش، انجام شده است. در این فصل، این تحقیقات مورد بررسی قرار می‌گیرند.

### 2-1- تحقیقات انجام شده برای شناسایی مشتریان ناراضی

شناسایی مشتریان ناراضی و نگهداری آن‌ها، یکی از روش‌های مدیریت نارضایتی مشتری، می‌باشد. هدف از این مطالعه آن است که با به حداقل رساندن میزان نارضایتی، سود شرکت‌ها به حداکثر رسانده شود. روش‌های متفاوتی برای پیش‌بینی احتمال ترک مشتریان وجود دارد که مشهورترین

مدل‌ها، شبکه عصبی<sup>۱</sup>، پشتیبانی از دستگاہ بردار<sup>۲</sup> و درخت تصمیم‌گیری می‌باشند که تاریخچه‌ی بررسی شده در این پایان‌نامه، بیشتر روش‌های داده‌کاوی را مورد تحلیل و بررسی قرار داده است. یکی از روش‌هایی که برای تحلیل نارضایتی مشتریان مورد استفاده قرار می‌گیرد، به ترتیب مراحل زیر را طی می‌کند: شناسایی بهترین داده‌ها<sup>۳</sup>، معناشناسی داده‌ها<sup>۴</sup>، انتخاب ویژگی<sup>۵</sup>، توسعه و پیش‌بینی مدل<sup>۶</sup>، و در نهایت معتبرسازی نتایج<sup>۷</sup> [J. Hadden et al, 2007]. انتخاب ویژگی دارای سه ویژگی مهم خواهد بود. اول آن‌که با کاهش تعداد شاخه‌ها به مجموعه‌ای کوچکتر از اطلاعات، کارایی یک طبقه افزایش خواهد یافت. دوم آن‌که زمان پردازش کاهش خواهد یافت. سوم آن‌که با کاهش ابعاد پایگاه داده، بهتر می‌توان از روشهایی مانند شبکه عصبی استفاده کرد. به طور کلی، شبکه عصبی، نمی‌تواند عمل پیش‌بینی را به خوبی انجام دهد. در نتیجه، مدل نارضایتی به خوبی ساخته نمی‌شود و علت نارضایتی این مشتریان به خوبی پیش‌بینی نمی‌شود.

یکی دیگر از روش‌های بکار برده شده در پیش‌بینی رفتار مشتریان، بکاربردن مدل آر-اف-ام یا ضرایب تازگی، تکرار و ارزش پولی<sup>۸</sup> می‌باشد [D. Liu and Y. Shih, 2005]. ضریب تازگی، نشان‌دهنده زمان آخرین خرید مشتری می‌باشد که هر چه این عدد کمتر باشد، این مشتری وفادارتر به حساب می‌آید. ضریب تکرار، نشان‌دهنده تعداد خریدهایی است که مشتری انجام داده است که هر چه این عدد بزرگتر باشد، نشان‌دهنده وفاداری بیشتر می‌باشد. ضریب ارزش پولی، نشان‌دهنده کل پولی می‌باشد که یک مشتری در یک دوره مشخص در یک شرکت خرج کرده است که بالاتر بودن این عدد نیز نشان‌دهنده وفاداری بیشتر می‌باشد. اما چون این ضرایب تنها گذشته مشتری را مورد ارزیابی قرار می‌دهند، نمی‌توانند به صورت کامل، نارضایتی مشتریان را پیش‌بینی کنند.

در مطالعه دیگری با ارزیابی سه عامل از سرانجام مشتری، احتمال نارضایتی او را پیش‌بینی می‌کند که این عوامل عبارتند از [B. Larivie`re and D. V. Poel, 2005]:

<sup>1</sup> Neutral network

<sup>2</sup> Support vector machines

<sup>3</sup> Identification of the best data

<sup>4</sup> Data semantics

<sup>5</sup> Feature selection

<sup>6</sup> Development of predictive model

<sup>7</sup> Validation of results

<sup>8</sup> RFM: recency, frequency, monetary

۱. خرید بعدی: آیا مشتری محصول بعدی را خریداری می‌کند؟

۲. نارضایتی: تصمیم‌گیری مشتری در پایان دادن خرید یک محصول چیست؟

۳. ارزیابی سودمندی: سیر تکاملی سود مشتری و کاهش آن را مورد ارزیابی قرار می‌دهد.

برای ارزیابی این عوامل، مجموعه گسترده‌ای از متغیرها، از جمله رفتار گذشته مشتری، ناهمگنی مشتریان موجود و متغیرهایی مربوط به فروشندگانه‌ها<sup>۱</sup>، با استفاده از روش جنگل‌های تصادفی<sup>۲</sup>، ارزیابی شده است. نتایج نشان می‌دهد که مجموعه یکسانی از متغیرها، تأثیر متفاوتی روی خرید و سودآوری خواهند داشت. از عوامل مؤثر در خرید تکراری و ارزیابی سودمندی مشتری، خرید گذشته او می‌باشد. از نتایج تحلیل خروجی مشتریان متفاوت به صورت همزمان، یافتن وابستگی متغیرهای خروجی به یکدیگر است، که جنگل‌های تصادفی این متغیرهای وابسته را به خوبی محاسبه می‌کنند. این روش به دلیل آن که متغیرهای زیادی را برای ارزیابی در نظر می‌گیرد، مشتریان را به گروه‌های زیادی تقسیم‌بندی می‌کند و بسیاری از مشتریان با ویژگی‌های مشابه، در گروه‌های متفاوت قرار می‌گیرند.

با مقایسه انواع روش‌های داده‌کاوی، می‌توان به صورت دوره‌ای به هر مشتری متعهد، نمره‌ای با عنوان احتمال نارضایتی اختصاص داد [Sh. Y. Hung et al, 2006]. نتایج نشان می‌دهد که دو روش درخت تصمیم‌گیری و شبکه‌عصبی می‌توانند با استفاده از اطلاعات ثبت شده از مشتریان، اطلاعات صورت حساب، موقعیت سرویس-دهی، جزئیات مکالمات، و تغییرات سرویس‌دهی، پیش‌بینی دقیقی از نارضایتی مشتریان داشته باشند. در این روش، ابتدا با استفاده از روش خوشه‌بندی<sup>۳</sup>، مشتریان به پنج دسته تقسیم می‌شوند و سپس یک درخت تصمیم-گیری<sup>۴</sup> در هر دسته استفاده می‌شود تا مشتریان ناراضی شناسایی شوند. روش ارائه شده، با افزایش پیچیدگی کسب و کار جوابگو نمی‌باشد و به روش‌های پیچیده تری برای تقسیم‌بندی مشتریان نیاز است.

تحقیق دیگری از یک ساختار هیبرید<sup>۵</sup> استفاده می‌کند به طوری که علاوه بر پیش‌بینی احتمال نارضایتی، سیاست-هایی را برای نگهداری آن‌ها بکار می‌گیرد [B. H. Chu et al, 2007]. این تحقیق از دو بخش یادگیری<sup>۶</sup> و استفاده<sup>۷</sup> تشکیل شده است. در بخش یادگیری، مدل‌های سیاستی و نارضایتی ساخته می‌شوند. ابتدا از روش C5.0 استفاده می‌کند تا احتمال نارضایتی مشتریان را پیش‌بینی کند. سپس الگوهای پنهان شده در پایگاه داده

<sup>1</sup> Salesperson

<sup>2</sup> Random forest

<sup>3</sup> Clustering

<sup>4</sup> Decision Tree

<sup>5</sup> hybridize

<sup>6</sup> learning

<sup>7</sup> usage



مشترک، مورد جستجو قرار می‌گیرند و از خوشه‌بندی برای شکل‌دادن یک مدل نارضایتی استفاده می‌شود. این روش از یک مدل سیاستی برای بکار بردن صفات شناسایی شده در مدل نارضایتی استفاده می‌کند تا این افراد را به گروه‌های مجزا تقسیم کند. سپس برای هر گروه یک مدل سیاستی را توسعه می‌دهد تا در روش استفاده، آن سیاست را بکار گرفته و از حرکت آن‌ها به سمت رقبا جلوگیری کند. در این روش، مشتریان ناراضی به خوبی شناسایی می‌شوند، اما تکنیک مناسبی برای نگهداری آن‌ها استفاده نمی‌شود.

برای ساخت مدل‌های نارضایتی کارا تر برای شناسایی مشتریان، می‌توان از پشتیبانی از دستگاه بردار استفاده کرد [K. Coussement and D. V. Poel, J. 2008]. این روش زمانی که برای داده‌های بازاریابی شلوغ بکار گرفته می‌شود، کارایی خوبی به همراه خواهد داشت. با نگاشت ورودی‌های غیر خطی به فضای ویژه چند بعدی، پشتیبانی از دستگاه بردار، مسائل پیچیده را به توابع مشخص ساده‌تری تقسیم‌بندی می‌کند. اما این روش فقط با بهترین پارامترها، دقیق عمل می‌کند و در صورتی نتایج بهتری به همراه خواهد داشت که پارامترهای آن به درستی انتخاب شده باشند.

تجزیه و تحلیل عوامل<sup>1</sup> و متغیر سازگاری از سلطه راف گاید دستگاه بر اساس روش<sup>2</sup>، روش دیگری برای مدیریت نارضایتی می‌باشد [J. Liou, 2009]. روش اشاره شده، یک روش داده‌کاوی می‌باشد که اطلاعات پنهان شده در پایگاه داده‌های بزرگ را استخراج می‌کند و روش مفیدی برای کاهش داده‌ها به صورت کمی و کیفی ارائه می‌دهد. از مشخصات ویژه این روش، آن است که با توجه به تغییر محیط و فرهنگ محلی، به آسانی قابل تغییر و گسترش به سایر محیط‌ها می‌باشد. مزیت این روش بر روش‌های دیگر آن است که برای صفات، اولویت در نظر نمی‌گیرد و بر خلاف روش‌های قبلی به خوبی با ناسازگاری روبرو می‌شود.

الگوریتم ژنتیک<sup>3</sup> بر مبنای مدل‌های شبکه عصبی، روشی دیگری است که برای پیش‌بینی نارضایتی مشتریان استفاده می‌شود [C. P. Pendharkar, 2009]. این الگوریتم تلاش می‌کند تا دقت پیش‌بینی نارضایتی را به حداکثر برساند. نتایج بدست آمده، حاکی از عملکرد بهتر الگوریتم ژنتیک در حین گران‌تر بودن آن دارد. همچنین نشان داده شده است که تعداد ندهای پنهان در شبکه عصبی نقش مهمی در پیش‌بینی کارایی خواهند داشت. انتخاب پویای ورودی‌ها و ندهای پنهان در ارزیابی این روش تأثیر بسزایی خواهند داشت، در صورتی که

<sup>1</sup> Factor analysis

<sup>2</sup> Variable consistency dominance-based rough set approach (VC-DRSA)

<sup>3</sup> Genetic algorithm

این مقاله از این ندها به صورت ثابت استفاده می‌کند. استفاده ثابت از این ندها باعث از دست دادن کارایی آن خواهد شد.

با ترکیب صدای مشتریان<sup>۱</sup> از طریق ایمیل‌های مرکز مکالمه و یک مدل پیش‌بینی نارضایتی سنتی، مدل بهتری با کارایی بالاتر ارائه می‌شود [K. Coussement and D. V. Poel, A. 2008]. در این روش اطلاعات متنی با فرمت آزاد را از ایمیل مشتریان، با اطلاعات مشتق شده از پایگاه داده یکی می‌کند. اطلاعاتی که از ایمیل مشتریان استخراج می‌شود، ساختاریافته نمی‌باشند، که در این مقاله برای تبدیل آن‌ها به فرم ساختاریافته، از فضای برداری استفاده شده است. در این روش بسیاری از ویژگی‌های مشتریان در نظر گرفته نمی‌شود و کارایی لازم را نخواهد داشت.

روش عدم تعادل مسئله<sup>۲</sup> در داده‌کاوای روش دیگری می‌باشد که برای پیش‌بینی نارضایتی مشتریان مورد استفاده قرار می‌گیرد [J. Burez and D. V. Poel, 2009]. در حقیقت هر یک از داده‌هایی که از پایگاه داده استخراج می‌کنیم، نواقصی خواهند داشت. در این مطالعه، روش‌های نمونه‌گذاری توسعه یافته و تصادفی، گرادیان تصادفی<sup>۳</sup> و جنگل‌های تصادفی وزن‌دار پیاده‌سازی شده‌اند. نتایج نشان می‌دهد که نمونه‌گذاری می‌تواند باعث افزایش دقت پیش‌بینی شود. همچنین زمانی که از روش‌های نمونه‌گذاری پیچیده و نمونه‌گذاری توسعه یافته مکعبی<sup>۴</sup> در این روش استفاده می‌کنیم، افزایشی در پیش‌بینی کارایی نخواهیم داشت. برخلاف جنگل‌های تصادفی وزن دار که کارایی را افزایش می‌دهند، این روش برخلاف پیچیدگی زیادی که به همراه دارد، کارایی را به خوبی افزایش نمی‌دهد.

مقالات دیگر با اندازه‌گیری ارزش دوره زندگی مشتریان، به نگهداری آن‌ها پرداخته‌اند. به دلیل آن که بدست‌آوردن مشتریان جدید ساده نیست، بهتر است ارزش دوره زندگی مشتریان موجود را بالا ببریم. یکی از این محققان، یک روش پیشنهادی را برای توسعه ارزش دوره زندگی مشتریان نشان داده است که مشتریان را تشویق به خریداری موارد بیشتری از سرویس‌های آماده شده می‌کند [T. Iwata et al, 2008]. در این مقاله الگوهای خرید تکراری را از بین مشتریان با ارزشمندتر انتخاب می‌کنند و آن‌ها را به مشتریان جدیدی که از آن الگوها

<sup>1</sup> Voice of customers(VOC)

<sup>2</sup> Class imbalance

<sup>3</sup> Gradient boosting

<sup>4</sup> CUBE

پیروی می کنند، پیشنهاد می کنند. این مقاله برای یافتن الگوهای خرید، از روش های تجزیه و تحلیل بقا<sup>۱</sup> استفاده می کند. در نتیجه، هر چه ارزش دوره زندگی مشتریان بالاتر، رضایت مشتریان بیشتر می باشد.

می توان برای تحلیل ارزش مشتریان و تقسیم بندی آنها بر اساس ارزش زندگیشان، الگوریتم مناسبی ارائه داد [S.Y. Kim et al, 2006]. در این تحقیق، مشتریان به دو روش تقسیم بندی می شوند. در روش اول پس از تقسیم بندی مشتریان به صورت نزولی ارزش ها، محصولات مشهورتر با قیمت پایین تر را به مشتریان سودمندتر سفارش می دهد. در روش دوم، سه عامل ارزش جاری، ارزش پتانسیل، و وفاداری مشتریان را در نظر می گیرد و سپس از داده کاوی برای محاسبه وفاداری مشتری استفاده می کند.

مقاله دیگری از داده کاوی برای محاسبه ارزش خالص دوره زندگی مشتریان در بازاریابی صنعت موبایل استفاده کرده است [K. Gelbrich<sup>1</sup> and R. Nakhaeizadeh<sup>2</sup>, 2000]. در این مقاله، به جای ارزیابی مشتریان جاری، مشتریان آینده را در نظر می گیرد. زمان پایان رابطه یک مشتری با یک شرکت، زمان مرگ او در نظر گرفته شده است و این را به عنوان هدفی برای تلاش برای نگهداری مشتری در نظر گرفته است. در این مقاله سه مورد توسط داده کاوی محاسبه شده است که عبارتند از: پیش بینی تکرار خرید، پیش بینی قبول قیمت و پیش بینی نرخ تنزیل پول.

در تحقیق دیگری به جای تمرکز روی کالاها، روی مشتریان متمرکز می شود و مشتریانی که ارزش دوره زندگی آنها در حال کاهش است را شناسایی می کند [N. Glady et al, A. 2009]. به دلیل آن که مشتریان متفاوت، عوامل کاهش وفاداری متفاوتی دارند، برای این محاسبات نیاز به اندازه گیری های دقیق تری می باشد. این مقاله با اندازه گیری تغییرات ارزش دوره زندگی مشتریان، این افراد را شناسایی می کند.

در مقاله دیگری مدل هایی برای افزایش تأثیر صلیب فروش<sup>۲</sup> ارزیابی شده است [A. Knott et al, 2002]. در این مقاله با پیش بینی محصول بعدی که مشتری خریداری می کند، علاوه بر این که سود و دقت انجام عملیات را افزایش می دهد، از پیش بینی محصولات غیرمشهور جلوگیری می کند.

مقاله دیگری معادله ارزیابی ارزش دوره زندگی مشتریان را به صورت پویا ارزیابی کرده است، به طوری که بتوان هر متغیر دلخواه را به آن اضافه کرد [M. Crowder et al, 2007]. عاملی که در این مقاله اضافه شده است، تغییر پذیری افراد می باشد. در مدل های قبلی تنها رفتار مشتری در نظر گرفته می شد، اما در این مدل، علاوه بر رفتار

<sup>1</sup> Survival analysis

<sup>2</sup> Cross-selling

مشتریان، هر تاثیر تصادفی که بخواهیم را در نظر می‌گیرد. به طور کلی این مدل به صورتی طراحی شده است که یک شرکت بتواند، با توجه به اطلاعات ثبت شده آن مشتری، هر عامل دیگری را به ارزیابی ارزش دوره زندگی مشتریان بیفزاید و معادله جدیدتری را در نظر بگیرد. این تحقیق معادله ارزیابی ارزش مشتریان را پویا فرض کرده است، در صورتی که پویا در نظر گرفتن این متغیر، تنها باعث افزایش پیچیدگی خواهد شد.

مقاله دیگری از تحلیل پیوسته<sup>۱</sup> برای طراحی بهینه یک فروشگاه تجملاتی با عنوان تجاری خوش گذران، استفاده کرده است، تا بتواند ارزش دوره زندگی مشتریان را به حداکثر برساند [K. Gyeongmu et al, 2009]. در این محاسبات، ملیت نیز اضافه شده است. در نتیجه، در فروشگاه‌های بزرگ که افراد با ملیت‌های متفاوت وارد می‌شوند، ارزش دوره زندگی مشتریان، به بهترین نحو محاسبه می‌شود و تعداد رجوع‌کنندگان تا ده سال آینده پیش‌بینی می‌شود.

در مقاله دیگری مدل نارضایتی مشتریان، با ترکیب زنجیره مارکو، رده بندی و درخت بازگشتی طرح‌ریزی می‌شود [M. Haenlein et al, 2007]. مدل ارائه شده می‌تواند با درآمدهای گسسته همانند جریان پیوسته آن روبرو شود. علاوه بر آن به جای آن که روی تک تک مشتریان در نظر گرفته شود، روی گروه‌های همگن عمل می‌کند. این روش قابلیت فهم بالایی دارد.

مقاله دیگری ارزیابی ارزش دوره زندگی مشتریان را با استفاده از زنجیره مارکو<sup>۲</sup> و شش عامل دیگر انجام داده است [Ma. Ming et al, 2008]. این عوامل در این جا عنوان شده‌اند. عامل اول زمان خرید مشتری می‌باشد که بر حسب خواست مشتری یا پیشنهاد فروشنده انجام می‌شود.

عامل دوم، موقعیت رفتار مشتری می‌باشد که یا از او نگهداری شده است، یا به سمت رقبا مهاجرت کرده است. عامل سوم، به هزینه نگهداری مشتری در یک دوره اشاره می‌کند.

قانون پاسخ، عامل چهارم می‌باشد که توسط زنجیره مارکو، ترتیب زمان پاسخ مشتری را برای بررسی نگهداری او نشان می‌دهد. عامل پنجم، قانون سود خالص می‌باشد که ترتیب سود خالص بدست آمده از یک مشتری در هر پاسخ او به شرکت را بررسی می‌کند. عامل ششم، راه حل قطع ارتباط می‌باشد، که ارتباطش را با مشتریانی که سود آنها در حال کاهش است قطع می‌کند.

<sup>1</sup> Conjoint analysis

<sup>2</sup> Markov

در این روش ارزش دوره زندگی مشتریان به صورت سیستماتیک بررسی شده است و در نتیجه توانسته است روی زمان اتمام ارتباط با مشتری و هزینه بازاریابی تصمیم‌گیری کند، اما تلاشی برای نگهداری مشتریان ناراضی انجام نداده است.

مقاله دیگری مشتریان ارزشمندتر را نگهداری می‌کند و ارزش خاص دوره زندگی این مشتریان را افزایش می‌دهد [Ch. Shen and H. M. Chuang, 2009]. برای آن‌که بتواند مشتریان را تقسیم‌بندی کند و مشتریان ارزشمند را شناسایی کند، از مدل آر-اف-ام استفاده می‌کند. سپس با استفاده از روش‌های داده‌کاوی، از داده‌های مشتری و تراکنش‌های پایگاه داده استفاده می‌کند تا رقاب ایجاد کند. هدف این مقاله، ارزیابی ارزش دوره زندگی مشتریان و توسعه آن‌ها با استفاده از بازاریابی می‌باشد.

برای محاسبه ارزش دوره زندگی مشتریان، می‌توان به پیش‌بینی تعداد تراکنش‌های آینده یک مشتری و سود آن‌ها پرداخت [N. Glady et al, M. 2009]. در این مقاله با استفاده از مدل توزیع دو جمله‌ای منفی یا پاراتو<sup>۱</sup>، تعداد تراکنش‌های آینده یک مشتری پیش‌بینی می‌شود. سپس علاوه بر تعداد تراکنش‌ها، سود هر تراکنش محاسبه خواهد شد. برای این‌که دقت پیش‌بینی ارزش دوره زندگی مشتریان را افزایش دهد، از وابستگی بین تعداد تراکنش‌ها و سود هر تراکنش استفاده می‌کند، چون در نظر گرفتن مستقل آن‌ها ممکن است کارایی را کاهش دهد.

تکنیک‌های زیادی برای نگهداری مشتریان وجود دارد، اما بهترین تکنیک برای نگهداری آن‌ها مشخص نیست. تاکنون تحقیق‌های زیادی با استفاده از داده‌کاوی روی نگهداری مشتری انجام شده است، اما هیچ‌کدام نتوانسته‌اند روشی ارائه دهند که از بین تکنیک‌های موجود، مناسب‌ترین تکنیک انتخاب شود. می‌توان تراکنش‌های یک سیستم را در فایل ثبت تراکنش، وارد نمود و از الگوریتم آلفا در فرآیند کاوی برای استخراج اطلاعات پنهان شده در پایگاه داده استفاده نمود [A.K.A. de Medeiros et al, 2004].

## 2-2- ارتباط موضوع تحقیق و تاریخچه

بیشتر کارهایی که تاکنون انجام شده است، تمرکز روی جلوگیری از نارضایتی مشتری داشته‌اند، یا تلاش‌هایی برای اندازه‌گیری ارزش دوره زندگی او داشته‌اند. کارهای دیگری نیز انجام شده است که علاوه بر شناسایی مشتریان ناراضی، تکنیک‌هایی برای نگهداری آن‌ها اعمال کرده‌اند. مقاله‌های متفاوت، تکنیک‌های متفاوتی از

<sup>1</sup> Pareto/Negative Binomial Distribution(NBD)

داده کاوی را مورد استفاده قرار داده‌اند. در این تحقیق ابتدا با اندازه‌گیری ارزش دوره زندگی مشتریان، آن‌ها را گروه‌بندی می‌کند و سپس با ترکیب آن با داده کاوی، مشتریان ناراضی را شناسایی می‌کند. در نهایت، روش‌هایی برای نگهداری آن‌ها با استفاده از فرآیند کاوی اعمال می‌شود. تاکنون بیشتر روش‌هایی که برای نگهداری مشتری استفاده شده است، خوشه‌بندی یا گروه‌بندی می‌باشد، اما در این‌جا از فرآیند کاوی برای نخستین بار استفاده شده است تا تکنیک‌های مناسب برای نگهداری مشتریان کاوش شود.

### 2-3- اهمیت موضوع و هدف آن

سالیانه شرکت‌ها میلیون‌ها دلار سرمایه خود را به خاطر عدم مدیریت مناسب در روابط خود با مشتریان، از دست می‌دهند. کارشناسان در حال ارزیابی روش‌هایی برای شناسایی مشتریان ناراضی و نگهداری آن‌ها می‌باشند. در این تحقیق، علاوه بر شناسایی مشتریان ناراضی توسط داده کاوی، تکنیک‌هایی، برای نگهداری آن‌ها توسط فرآیند کاوی اعمال شده است.

### 2-4- نتیجه‌گیری

شرکت‌ها سالیانه میلیون‌ها دلار سرمایه خود را به دلیل عدم مدیریت روابط خود با مشتریان، از دست می‌دهند. یکی از روش‌هایی که می‌توان برای مدیریت مناسب روابط مطرح کرد، شناسایی مشتریان ناراضی و نگهداری آن‌ها می‌باشد. همان‌طور که در تاریخچه اشاره شد، روش‌های داده کاوی، برای پیش‌بینی مشتریان ناراضی و نگهداری آن‌ها مناسب می‌باشند. تکنیک‌هایی که تاکنون برای نگهداری مشتری استفاده شده است، فقط شامل سیاست‌هایی می‌باشند که پس از گروه‌بندی آن‌ها توسط خوشه‌بندی، برای نگهداری آن‌ها استفاده شده است. در این تحقیق علاوه بر شناسایی این مشتریان توسط درخت تصمیم‌گیری، بهترین تکنیک‌ها توسط الگوریتم آلفا از فرآیند کاوی، استخراج شده است.

## فصل سوم

### موضوعات مرتبط با تحقیق

در این فصل، موضوعات مرتبط با تحقیق مورد بررسی قرار می‌گیرد. در این تحقیق، ابتدا ارزش دوره زندگی مشتریان اندازه‌گیری شده است. سپس از درخت تصمیم‌گیری که یکی از روشهای داده‌کاوی می‌باشد، برای ساخت مدل نارضایتی استفاده شده است. روش C5.0، یکی از روشهای توسعه‌یافته در داده‌کاوی، می‌باشد که در این تحقیق بکار رفته است. در مرحله بعدی فرآیند کاوی برای یافتن بهترین تکنیک برای نگهداری مشتری مورد استفاده قرار گرفته است. از الگوریتم‌های معروف در فرآیند کاوی می‌توان به الگوریتم آلفا اشاره کرد، که در این تحقیق از این الگوریتم استفاده می‌کند تا بهترین روش برای نگهداری مشتریان را کاوش کند. در بخش ۱-۳، ارزش خالص دوره زندگی مشتری محاسبه می‌شود. در بخش ۲-۳، داده‌کاوی مورد بررسی قرار می‌گیرد. در بخش ۳-۳، فرآیند کاوی بررسی می‌شود.

### 3-1- ارزش خالص دوره زندگی مشتری<sup>1</sup>

ارزش خالص دوره زندگی مشتری، ارزش خالص کنونی پول نقد به‌دست‌آمده در یک دوره زمانی مشخص می‌باشد و دقیقاً نشان می‌دهد که هر مشتری در شرایط متفاوت چه

---

<sup>1</sup> CLV: Customer Lifetime Value

ارزشی خواهد داشت و چه هزینه‌ای برای آن صرف می‌شود [G. Sunil et al, 2006]. اصطلاح ارزش مشتری، ارزش خالص او را نشان می‌دهد. اصطلاح دوره زندگی، به مدت ارتباط بین خریدار و فروشنده برمی‌گردد. بنابراین برای اندازه‌گیری ارزش خالص دوره زندگی مشتری، نیاز به ارزیابی پایان یک رابطه است. به طور کلی سود مشتری در آینده با در نظر گرفتن کاهش ارزش پول در آینده، بررسی می‌شود. اصطلاح بازاریابی، پیشه‌ای برای یافتن مشتریان و قانع کردن آن‌ها به خرید بیشتر می‌باشد. با استفاده از ارزش خالص دوره زندگی مشتری، می‌توان پی برد که کدام مشتریان سزاوار توجه بیشتری می‌باشند و از دست دادن آن‌ها گرانتر است. برای محاسبه ارزش خالص دوره زندگی مشتری، هزینه بدست آوردن، خدمت کردن، و نگهداری هر مشتری را از درآمد تولید شده توسط آن مشتری کم می‌کنند. برای مثال اگر هزینه بدست آوردن یک مشتری \$۵۰ باشد و ارزش خالص دوره زندگی مشتری آن \$۶۰ باشد، آن مشتری سودمند است. بنابراین بدست آوردن مشتریانی شبیه به آن مفید است. هزینه واقعی بدست آوردن یک مشتری عبارتست از هر آنچه یک بازاریاب برای بدست آوردن آن‌ها صرف می‌کند تا از آن‌ها پاسخ مثبت دریافت کند.

عوامل مؤثر در محاسبه ارزش خالص دوره زندگی مشتری عبارتند از: میزان تغییر جهت مشتری به سمت رقبا<sup>۱</sup>، میزان تخفیف<sup>۲</sup>، هزینه نگهداری، دوره زمانی در نظر گرفته شده<sup>۳</sup>، درآمد هر دوره<sup>۴</sup> و درصد سود (خالص یا ناخالص) حاصل از فروش.

محاسبه ارزش خالص دوره زندگی مشتری به دو قسمت تقسیم می‌شود:

۱: ارزش واقعی آن که عبارتست از ارزش قابل پیش‌بینی کنونی، بر مبنای آنچه در حال حاضر از مشتری می‌داند.

۲: ارزش بالفعل که عبارتست از ارزشی که با شناخت بیشتر مشتری به آن دست می‌آید.

به طور کلی برای آن که بتوان ارزش دوره زندگی مشتری را بدست آورد، باید بتوان کل هزینه‌ای که برای بدست آوردن و نگهداری آن صرف شده است را محاسبه کرد. ارزش بدست آوردن مشتریان، شامل هر هزینه‌ای از جمله تبلیغات و غیره می‌باشد. هزینه نگهداری او شامل هر گونه تلاشی که برای

<sup>1</sup> churn

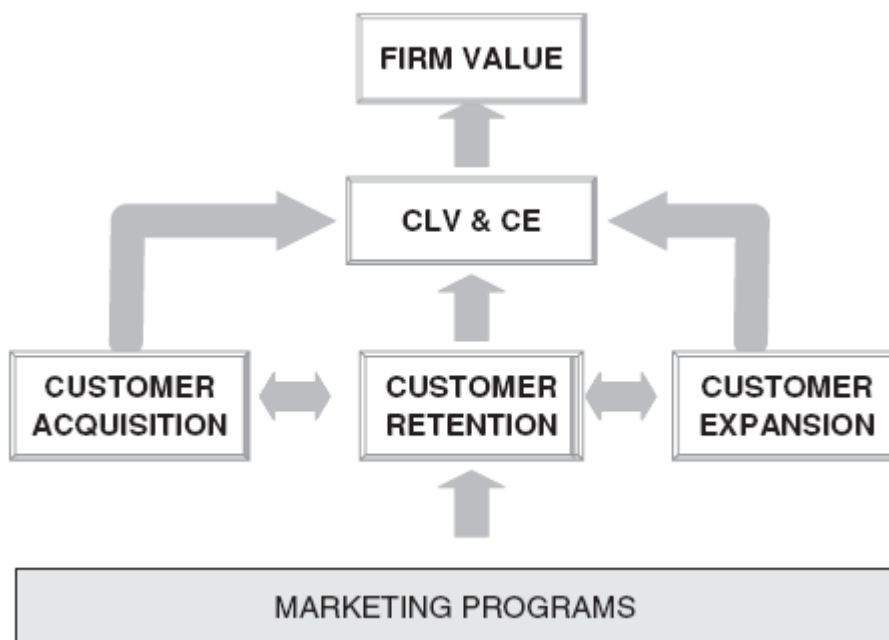
<sup>2</sup> Discount rate

<sup>3</sup> Period

<sup>4</sup> Periodic Revenue



نگهداری او انجام داده‌است، از جمله دادن تخفیف، دادن هدیه، دادن اعتماد، پس گرفتن اجناس در صورت نداشتن کیفیت لازم و مواردی شبیه به آن می‌باشد. پس از محاسبه همه این موارد، به محاسبه ارزش دوره زندگی مشتری با فرمولهای ارائه شده پرداخته می‌شود.



شکل 1-3 ساختار ارزش دوره زندگی مشتری

همان‌طور که شکل ۱-۳ نشان می‌دهد [G. Sunil et al, 2006]، برنامه‌های بازاریابی روی سه عامل بدست آوردن، نگهداری، و گسترش مشتری بررسی‌هایی انجام داده‌اند و محققان متفاوت هر یک اتصال بین این عوامل را بررسی کرده‌اند.

### 2-3- داده کاوی<sup>۱</sup>

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره شده در این سیستم‌ها، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده در پایگاه داده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد [M. Kamber and J. Han, 2000]. با استفاده از پرسش‌های ساده در زبان پرس و جوی ساختاریافته<sup>۲</sup> و ابزارهای گوناگون گزارش‌گیری معمولی، می‌توان اطلاعاتی را در اختیار کاربران قرار داد تا بتوانند به نتیجه‌گیری در مورد داده‌ها و روابط منطقی میان آن‌ها بپردازند. اما وقتی که حجم داده‌ها بالا باشد، کاربران هر چند زیر دست

<sup>۱</sup> Data Mining

<sup>۲</sup> SQL

و با تجربه باشند، نمی‌توانند الگوهای مفید را در میان حجم انبوه داده‌ها تشخیص دهند و یا اگر قادر به این کار هم باشند، هزینه عملیات از نظر نیروی انسانی و مادی بسیار بالا است. علاوه بر این، با وجود سیستم‌های یکپارچه اطلاعاتی، سیستم‌های یکپارچه بانکی و تجارت الکترونیک، لحظه به لحظه به حجم داده‌ها در پایگاه داده‌ها اضافه شده و باعث به وجود آمدن انبارهای عظیمی از داده‌ها شده است، به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده‌ها را بیش از پیش نمایان کرده است.

داده‌کاوی به عنوان یک راه حل برای این مسائل مطرح می‌باشد. در یک تعریف غیر رسمی، داده‌کاوی فرآیندی است خودکار، برای استخراج الگوهایی که دانش را بازنمایی می‌کنند، که این دانش به صورت ضمنی در پایگاه داده‌های عظیم، انباره داده و دیگر مخازن بزرگ اطلاعات، ذخیره شده است.

در داده‌کاوی از بخشی از علم آمار به نام تحلیل اکتشافی داده‌ها استفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تأکید می‌شود. علاوه بر این داده‌کاوی با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، بنابراین می‌توان گفت در داده‌کاوی تئوری‌های پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می‌آمیزند تا زمینه کاربردی آن فراهم شود.

### 3-2-1- ساختار سیستم داده کاوی

پس از پیش‌پردازش داده‌ها، نتایج در پایگاه داده یا انبار داده جدیدی ذخیره می‌شود. ورودی این ساختار، مخازن بزرگ داده و خروجی آن اطلاعات مورد نیاز کاربر است و پایان کار به عنوان کشف دانش از پایگاه داده‌ها تلقی می‌شود [M. Kamber and J. Han, 2000].

در حالت کلی یک سیستم داده‌کاوی می‌تواند شامل مولفه‌های زیر باشد:

الف) پایگاه داده: برای ذخیره‌سازی داده‌های حجیم به کار می‌روند.

ب) خادم<sup>۱</sup>: مسئول بازخوانی داده‌های مربوط به کاوش بر اساس اهداف داده‌کاوی است.

ج) پایگاه دانش<sup>۲</sup>: محدوده‌ای از دانش است که جهت هدایت فرآیند کاوش یا ارزیابی جذابیت الگوهای حاصل به کار می‌رود. این اطلاعات می‌تواند شامل دانستنی‌هایی مانند عقاید کاربر یا مفاهیم رده بندی شده جهت سازماندهی خصیصه‌ها، آستانه‌های تعیین جذابیت و غیره باشد.

<sup>۱</sup> -Server

<sup>۲</sup> - Knowledge Base

د) موتور داده کاوی<sup>۱</sup>: بصورت گسترده از روش های آماری و یادگیری ماشین برای توصیف، پیوند دادن، رده بندی، تحلیل خوشه ای، پیش بینی و غیره استفاده می شود و هسته اصلی سیستم داده کاوی را تشکیل می دهد.

ه) واحد ارزیابی الگوها: با استفاده از معیارهای جذابیت الگو و آستانه های جذابیت و همچنین ارتباط متقابل با واحدهای دیگر داده کاوی، کاوش را به سمت تمرکز بر استخراج الگوهای جذاب هدایت می کند.

و) رابط گرافیکی کاربر<sup>۲</sup>: یک رابط گرافیکی بین کاربران و سیستم داده کاوی، ارتباط برقرار می کند.

با توجه به ساختار فوق، انجام داده کاوی مستلزم روشن شدن موارد زیر است:

- مجموعه داده هایی که باید کاوش شوند، کدامند؟

- چه نوع اطلاعاتی باید کاوش شوند؟

- از چه پس زمینه ای اطلاعاتی باید در این جا استفاده کرد؟

- چه معیارهایی می توانند برای ارزیابی الگو به کار روند؟

روشن شدن موارد بالا به کاربر اجازه می دهد که متقابلاً با سیستم داده کاوی در طول کاوش رابطه برقرار کند.

برای یکپارچه ساختن همه ای این موارد می توان یک زبان داده کاوی<sup>۳</sup> طراحی کرد. هر کاربری بسته به هدف خود، به نوعی به سمت داده کاوی برای تحلیل داده ها تمایل دارد. هر یک از موارد بالا در زیر به اختصار شرح داده شده اند.

۱. داده های کاری: منظور از داده های کاری، مجموعه ای از داده ها است که عملیات کاوش روی آنها

صورت می گیرد.

۲. نوع اطلاعاتی که باید استخراج شوند: نوع اطلاعات مورد نیاز، به هدف فرآیند داده کاوی وابسته است.

اطلاعات را به پنج دسته کلی توصیف کردن، پیوند دادن، رده بندی، پیش بینی و خوشه بندی تقسیم می کند. بر حسب این که هدف، استخراج کدام نوع از اطلاعات باشد، نیاز به توابع داده کاوی متفاوتی می باشد. علاوه بر

<sup>۱</sup> - Datamining Engine

<sup>۲</sup> - GUI(Geraphical User Interface)

<sup>۳</sup> - Data mining Query(DMQ)

مشخص ساختن نوع دانشی که باید جستجو شود، کاربر می‌تواند قالب‌هایی تهیه کند که همه‌ی الگوهای استخراجی با آن هم‌خوانی داشته باشد.

### 3-2-2- وظایف داده کاوی

مهمترین وظایف داده کاوی عبارتند از [M. Kamber and J. Han, 2000]:

۱. تحلیل کاوشگرانه داده‌ها<sup>۱</sup>: رسم جداول و نمودارها و شکل‌های گرافیکی، بدون این که ایده‌ای واضح از آن‌چه جستجو می‌کنیم داشته باشیم، ما را در تحلیل داده‌ها کمک می‌کند.
۲. مدل‌بندی توصیفی: هدف شرح همه داده‌ها یا فرآیند تولید داده‌ها می‌باشد، تکنیک‌های آن مانند برآورد توزیع، تحلیل خوشه‌ای و غیره می‌باشند.
۳. مدل‌بندی پیشگو (رده‌بندی رگرسیونی): هدف ساختن یک مدل است که بتواند مقادیر متغیر پاسخ را با استفاده از متغیرهای دیگر پیشگویی کند. اگر متغیر پاسخ رسته‌ای باشد آن را رده‌بندی و اگر متغیر پاسخ پیوسته یا کمی باشد، رگرسیون گویند. تفاوت اصلی بین مدل‌های پیشگو و توصیفی، در این است که مدل‌های پیشگو دارای یک متغیر یکتای هدف می‌باشند اما مدل‌های توصیفی این گونه نیستند. تکنیک‌های آن عبارت از درخت-های تصمیم، شبکه‌های عصبی، رگرسیون خطی و غیره می‌باشد.
۴. کشف الگوها و قوانین: هدف تعیین الگوها می‌باشد.

### 3-2-3- کاربردهای داده کاوی

داده کاوی به عنوان ده علم برتر که منجر به ایجاد تحول در عصر تکنولوژی می‌شود، در تمام زمینه‌ها کاربرد دارد و اصولاً هر جایی که داده وجود داشته باشد، داده کاوی نیز معنا می‌یابد [M. Kamber and J. Han, 2000]. به کمک ابزارهای داده کاوی، می‌توان مقادیر متغیرهایی را پیش‌بینی کرد و رفتار آن متغیرها را در آینده پیش‌بینی و توصیف نمود. این ابزارها در فرآیندهای تصمیم‌گیری متکی بر اطلاعات و دانش می‌باشند و کاربرد فراوان دارند و فعالیت‌های تجاری نوین و مدرن امروزه به شدت بر آن متکی هستند. فرآیندهای داده کاوی را نباید با روش‌های متداول آنالیز داده و اطلاعات و سیستم‌های تصمیم‌گیری معمولی یکی دانست. به کمک روش‌های داده کاوی می‌توان به پرسشهای زیادی پاسخ گفت که به طور سنتی عملاً امکان آن وجود ندارد. بسیاری از شرکت‌ها و موسسات در حال حاضر نیز دارای حجم انبوهی از اطلاعات هستند. تکنیک‌های داده کاوی به طور

<sup>1</sup> - Exploratory Data Analysis(EDA)