

۱-۲- مقدمه

در سال‌های اخیر فرهنگ تجارت به پیشرفت‌هایی نایل گشته است. مطابق با آن روابط اقتصادی مشتریان با شیوه‌های بنیادی و اساسی در حال تغییر بوده است. شرکت‌ها به منظور نظارت بر این گونه تغییرات، نیازمند ارائه راه حل‌ها هستند. ظهور و پیدایش اینترنت در تغییر جهت مرکز توجه بازاریابی نقش بسزایی داشته است. چنانچه اطلاعات بر خط^۱ بیشتر در دسترس قرار گیرد موجب آگاهی و هوشیاری بیشتر مشتریان می‌گردد. آنان در جریان تمام آنچه ارایه و پیشنهاد می‌شود قرار می‌گیرند و تقاضای بهترین‌ها را دارند. برای از عهده برآمدن در چنین شرایطی باید سیستم‌هایی که بتوانند به طور دقیق نسبت به مشتریان واکنش نشان دهند، به کار روند. جمع-آوری آمار مشتریان و داده‌های رفتاری آنان این هدف اصلی و دقیق را ممکن می‌سازد. این نوع هدف‌گیری به یک برنامه ریزی عالی هنگام ایجاد یک رقابت سخت و به مشخص کردن مشتریان بالقوه، هنگام عرضه محصولات جدید کمک می‌کند [۱].

مدیریت ارتباط با مشتری، بر مبادله‌ی ارزش بین مشتری و سازمان بنا شده و بر ارزش ایجاد شده در این ارتباط تاکید می‌کند تا بر این تلاش سازمان‌ها برای توسعه‌ی ارتباط بلندمدت با مشتریان، بر مبنای ارزش برای هر دو طرف، از اهداف اصلی مدیریت ارتباط با مشتری است. به عبارت دیگر هدف مدیریت ارتباط با مشتری، ارائه مزایایی از طریق مبادله دو جانبه و عمل به وعده‌هاست [۲].

بازاریابی ارتباط یک راهبرد کسب و کار است که با پیشرفت فناوری تقویت می‌شود و از طریق آن شرکت‌ها به ایجاد ارتباطات سودمند بر پایه‌ی بهینه‌سازی ارزش دریافتی و ادراکی مشتریان

می‌پردازند. اصطلاح بازاریابی ارتباط به نام‌های مختلفی شناخته می‌شود: بازاریابی پایگاه داده‌ها بازاریابی شخص به شخص، بازاریابی ابقا^۱ و بازاریابی وفاداری^۲ یا وابستگی^۳ [۲].

مدیریت روابط مشتری ترکیبی از مردم، فرآیندها و تکنولوژی است که به دنبال درک نیازهای مشتریان شرکت می‌باشد و در واقع به دنبال ارائه‌ی دیدگاهی جامع در مدیریت روابط از طریق تمرکز بر حفظ و توسعه‌ی روابط مشتری است. مدیریت ارتباط با مشتری نتیجه‌ی پیشرفت در زمینه‌ی فناوری اطلاعات و نیز تغییرات سازمانی حول فرآیندهای مشتری محوری می‌باشد. شرکت‌هایی که به طور موفقیت‌آمیزی مدیریت ارتباط با مشتری را پیاده کرده‌اند، نتایج بسیار خوبی در زمینه‌ی کسب وفاداری مشتریان و سوددهی دراز مدت آنان بدست آورده‌اند؛ اگر چه بخش اعظمی از مدیریت ارتباط با مشتری مبتنی بر تکنولوژی می‌باشد. با این حال اعتقاد بر اینکه فقط یک راه حل تکنولوژی است، مردود می‌باشد [۲۲].

مدیریت ارتباط با مشتری مستلزم حرکت به سمت مشتری مداری و تعریف استراتژی بازار از دیدگاه برون سازمانی است. این روند تحت عنوان دوره‌ی عمر مشتری^۳ مطرح می‌شود که شامل جذب مشتریان جدید، افزایش ارزش مشتری و حفظ مشتریان خود می‌باشد [۲۳].

مدیریت ارتباط با مشتری یک مفهوم دو مرحله‌ای است که وظیفه‌ی مرحله اول کسب مهارت در توجه ویژه به مشتری می‌باشد و این به معنی انتقال از محصول محوری و تمرکز بر نیازهای مشتری بیش از توجه به ویژگی‌های محصول است. شرکت‌ها در مرحله‌ی دوم در حال پیگیری اصول مرحله‌ی اول هستند. در این مرحله شرکت‌ها نایستی به مزیت‌های موجود اکتفا کنند؛ بلکه باید

1. Restention Marketing

2. Afinity Marketing

3 . Life cycle

توسعه‌ی مشتری محوریشان را از طریق تلفیق مدیریت ارتباط با مشتری با کل زنجیره‌ی مشتری، گسترش تکنولوژی و در نهایت نوآوری مداوم در ارزش‌های مد نظر مشتریان پیش برند [۲۴].

مدیریت ارتباط با مشتری یک فرایند تجاری است که تمام جوانب مشخصه‌های مشتری را آدرس دهی می‌کند، دانش مشتری را به وجود می‌آورد، روابط را با مشتری شکل می‌دهد و برداشت آنان را از محصولات یا خدمات سازمان ایجاد می‌کند. مدیریت ارتباط با مشتری توسط چهار عنصر از یک چارچوب ساده تعریف شده است: دانش، هدف، فروش و خدمت [۲۵].

رشد روز افزون تکنولوژی، مشتریان به راحتی می‌توانند نیازهای خود را از طرق مختلف رفع نمایند. در این شرایط تمام دغدغه شرکت‌ها نه تنها جذب مشتریان جدید، بلکه نگه داشتن و راضی نمودن مشتریان قدیم می‌باشد. در این میان استراتژی‌های تجاری شرکت‌ها امروزه بیشتر تکنولوژی مشتری - محور^۱ با خدمات شخصی سازی شده^۲ می‌باشد. 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 آن آغاز شود، شرکت باید اطلاعات مشتری را در اختیار داشته باشد. این اطلاعات می‌تواند از داده‌های داخلی مشتریان و یا از داده‌های منابع خارجی خریداری شده، به دست آید. برای داده‌های داخلی منابع مختلفی وجود دارد مانند پرسشنامه‌ها و بلاگ‌ها، سوابق کارت اعتباری و.... . منابع داده خارجی یا بانک‌های داده خریداری شده مانند آدرس‌ها، شماره تلفن‌ها، پروفایل‌های بازدید از وب سایت‌ها کلیدی برای به دست آوردن دانش بیشتری از مشتری است [۲۶].

بیشتر شرکت‌ها، بانک‌های داده‌ای عظیمی شامل داده‌های بازاریابی، منابع انسانی و مالی را دارا هستند. بنابراین، سرمایه‌گذاری در زمینه انبار داده، یکی از اجزای حیاتی در استراتژی مدیریت ارتباط با مشتری است [۲۷].

با اجرای مدیریت ارتباط با مشتری، تغییرات در سازمان و عملیات‌های شرکت با بهبود اجرا و مزیت رقابتی ایجاد می‌شود. بیشترین بهبودهایی را که می‌توان تخمین زد، موارد ذیل می‌باشند [۲۸]:

۱- رضایتمندی^۱ بیشتر مشتری

۲) وفاداری و چسبندگی بیشتر مشتری

۳) مدیریت برای افزایش تعداد مشتری

۴) بهبود و تمایل به رابطه با مشتری

۵) شناخت بخش‌های مشتری

۶) افزایش اثر بخشی خدمت رسانی به مشتری

۷) کاهش هزینه‌ها

۸) افزایش اطلاعات در دسترس و به موقع بازاریابی و فروش در خصوص نیازمندی، انتظارات

و تخمین مشتری

پس از تهیه و تخصیص منابع داده، سیستم مدیریت ارتباط با مشتری باید با به کار گیری ابزارهایی مانند داده‌کاوی، داده‌ها را تجزیه و تحلیل کند. اعم از اینکه شرکت تکنیک‌های آماری سستی

را به کار می‌برد یا یکی از ابزارهای نرم‌افزاری مانند داده‌کاوی را، کارشناسان نیاز به فهم داده‌های مشتری و روابط تجاری دارند. بنابراین، داشتن افرادی متخصص که این داده‌ها را با ابزارهای مربوطه استخراج و به صورت اطلاعات درآورند، مهم است [۱].

۱-۱-۲- مفاهیم مدیریت ارتباط با مشتری

مدیریت ارتباط با مشتریان عبارت است از مجموعه گام‌هایی که به منظور ایجاد، توسعه، نگهداری و بهینه سازی روابط طولانی مدت و ارزشمند بین مشتریان و سازمان برداشته می‌شود. همچنین بخشی از استراتژی یک سازمان جهت شناسایی مشتریان، راضی نگه داشتن آنان و تبدیل آنان به مشتری دائمی می‌باشد که به کمک آن می‌توانیم درباره‌ی نیازها و رفتارهای مشتریان بیشتر بیاموزیم و روابط قوی‌تر، دوستانه‌تر و سودمندتری با آنان داشته باشیم. در واقع داشتن رابطه‌ی خوب با مشتری، قلب هر کسب و کار سالم و موفق است. با این حال مدیریت ارتباط با مشتری تکنولوژی نیست بلکه اساساً یک فرآیند است و فناوری اطلاعات در این فرآیند نقش تسهیل کننده را دارد. ایده اصلی نهفته در آن، ترکیب شدن توانایی‌های فناوری اطلاعات و نیروی انسانی برای رسیدن به بینشی عمیق، درباره‌ی خواسته‌ها و ارزش مشتریان است. مدیریت ارتباط با مشتری از سه واژه‌ی مهم تشکیل شده است:

✓ مشتری: منظور از مشتری، مصرف کننده نهایی است که در روابط ارزش آفرین، نقش حمایت کننده را دارا می‌باشد. از نظر بسیاری از سازمان‌ها، مشتریان خوب کسانی هستند که سهم زیادی در سودآوری داشته اما در دوران حاضر نمی‌توان سودآوری و درآمد زایی را تنها معیار تعریف مشتریان دانست.

✓ ارتباط: ارتباط مجموعه‌ای از مهارت‌هاست. اما مهم‌ترین آنان درک نقطه نظرهای طرف مقابل و تفهیم نقطه نظرهای خویش است. رابطه میان یک سازمان و مشتریانش یک رابطه‌ی مداوم، دوجانبه، فعل و انفعالی، بسیار با ارزش و سودمند می‌باشد. این رابطه می‌تواند کوتاه و یا بلند مدت، مداوم و یا گسسته، به دفعات و یا یکباره باشد و مشتریان می‌توانند از راه‌های مختلف با دست اندر کاران شرکت ارتباط برقرار کنند.

✓ مدیریت: مدیریت، علم و هنر برنامه ریزی (پیش بینی)، سازماندهی (تقسیم کار و تفکیک وظایف)، هدایت و رهبری (راهنمایی کردن زیر دستان و توانایی نفوذ کردن در دل ایشان)، نظارت و کنترل (ناظر بودن بر اعمال ایشان و اصلاح اشتباهاتشان) و نهایتاً ایجاد هماهنگی (برقراری نگرش سیستمی بین کارکنان) برای رسیدن به هدف یا اهداف از پیش تعیین شده، است. البته تمام مباحث فوق روی منابع انسانی صورت می‌گیرد که مهم‌ترین و ارزنده‌ترین دارایی هر سازمان منابع انسانی (کارکنان) هستند و سپس منابع دیگر نظیر دارایی‌های فیزیکی، مادی و حتی اطلاعات قرار می‌گیرند.

۲-۱-۲- تعاریف مدیریت ارتباط با مشتری

مدیریت ارتباط با مشتری شامل مجموعه‌ای از فرایندها و سیستم‌های توانا می‌باشند که استراتژی کسب و کار را برای ساختن روابط بلند مدت و سودآور با مشتریان حمایت می‌کند. داده‌های مشتریان و ابزارهای تکنولوژی اطلاعات مبنایی را برای هر مدیریت ارتباط با مشتری موفق تشکیل می‌دهند. بعلاوه سرعت رو به رشد اینترنت و تکنولوژی‌های وابسته باعث افزایش فرصت‌هایی برای بازاریابی شده است و راه‌های ارتباطی بین شرکت‌ها و مشتریانش را تغییر داده است [۲۹].

برای مدیریت ارتباط با مشتری، تعاریف متعددی ارائه شده و هر کدام از تعاریف از زاویه متفاوتی به آن نگرسته‌اند، به طوری که کاتلر آن را انقلابی از تمرکز به مبادله، به تمرکز بر رابطه می‌داند. لئونارد بری^۱ آن را راهبرد جذب، نگهداری و ارتقای روابط با مشتریان تعریف می‌کند. برکویتز و همکاران آن را تلاش سازمان برای توسعه ارتباط بلندمدت و دو سویه یا اثربخشی هزینه تعریف می‌کنند. بویر و تانر^۲ تواناسازی خریداران و فروشندگان برای برنامه‌ریزی مشترک، کار دایم برای اطمینان از آینده را اساس تعریف خود قرار می‌دهند، مورگان و هانت^۳ بازاریابی روابط را به همه فعالیت‌هایی که به سمت تعیین، توسعه و نگهداری مبادلات موفق ارتباط حرکت می‌کنند، اطلاق می‌کنند. رالف و کاتمن^۴، درجه‌ی نزدیکی خریدار و فروشنده را مد نظر قرار داده، درجه‌ی نزدیکی را شامل مدت زمان رابطه، درجه اعتماد بین طرف‌های روابط، مقدار و تواتر خرید و یا ارتباط بین بخش‌ها می‌دانند. مک کورت^۵ بازاریابی روابط را فلسفه‌ای فرض می‌کند که در آن سازمان‌ها خواهان شکل دادن یک رابطه‌ی بلند مدت با مشتریان هستند در تعریفی دیگر، مدیریت ارتباط با مشتری با روش شناسایی، جذب و ایجاد مشتریان می‌داند. هدف مدیریت ارتباط با مشتری دستیابی به مزیت رقابتی مدیریت مشتری و در نهایت افزایش سطح سود می‌باشد [۳۰].

مدیریت ارتباط با مشتری روش خودکارسازی وظایف پس‌خوان فروش، بازاریابی و خدمات به مشتریان است که زمینه‌ی تقویت و حفظ ارتباط با مشتری را فراهم می‌سازد [۳۱]. مدیریت ارتباط با مشتری فرآیندی است که با استفاده از تسهیلات فراهم شده توسط فناوری اطلاعات درصدد ایجاد ارزش برای مشتریان و سازمان از طریق همکاری دو جانبه است [۳].

-
- 1 . Berry
 - 2 . Bowyer & Tanerr
 - 3 . Morgan & Hant
 - 4 . Ralf & Kattman
 - 5 .Kourt

اگر چه مدیریت ارتباط با مشتری به طور گسترده به عنوان یک روش مهم کسب و کار شناخته شده است، اما یک تعریف پذیرفته شده‌ی جهانی در مورد آن وجود ندارد [۳۲]. نگای^۱ در سال ۲۰۰۵، مدیریت ارتباط با مشتری را به عنوان «رویکردی سازمانی برای فهمیدن و تأثیرگذاری بر رفتار مشتری از طریق ارتباط معنی دار به منظور بهبود جذب مشتری، حفظ مشتری، وفاداری مشتری و سودآوری مشتری» تعریف می‌کند. همچنین کین کید^۲ مدیریت ارتباط با مشتری را به عنوان «استفاده استراتژیک از اطلاعات، فرایندها، تکنولوژی و افراد برای مدیریت کردن روابط مشتری با سازمان» نگریسته است. این تعاریف بر اهمیت در نظر گرفتن مدیریت ارتباط با مشتری به عنوان فرآیند جامع جذب و حفظ مشتری جهت شبیه سازی ارزش مشتری برای سازمان تاکید دارند [۳].

۳-۱-۲- چرخه‌ی زندگی مشتری

واژه‌ی چرخه‌ی زندگی مشتری به مراحل در ارتباط بین مشتری و تجارت برمی‌گردد و آگاهی نسبت به آن موجب سودآوری مشتری می‌شود. عموماً چهار مرحله در چرخه زندگی مشتری وجود دارد [۲۵و۱]:

- ۱) مشتری‌های بالقوه : افرادی که هنوز مشتری نیستند ولی در هدف بازار قرار دارند؛
- ۲) مشتری‌هایی که عکس‌العمل نشان می‌دهند: مشتریان بالقوه یا احتمالی که به یک محصول یا خدمت علاقه و واکنش نشان می‌دهند.
- ۳) مشتری‌های بالفعل: افرادی که در حال حاضر محصول یا خدمتی را به کار می‌برند.

1 . Ngai

2 . Kincaid

۴) مشتری‌های سابق: این‌گونه افراد مشتریان مناسبی نیستند چرا که مدت زیادی در هدف فروش قرار ندارند و یا خریدشان را به سمت محصولات رقیب برده‌اند.

۴-۱-۲- ابعاد ارتباط با مشتری

مدیریت ارتباط با مشتری شامل چهار بعد می‌باشد [۴]:

۱) شناسایی مشتریان

۲) جذب مشتریان

۳) حفظ مشتریان

۴) توسعه مشتریان

۴-۱-۲-۱- شناسایی مشتریان

مدیریت ارتباط با مشتری با شناسایی مشتریان شروع می‌شود. این فاز شامل هدف قرار دادن جمعیتی است که احتمال اینکه مشتری شوند، زیاد باشد یا برای سازمان سودآور می‌باشند. عناصر فاز شناسایی شامل تجزیه و تحلیل مشتریان هدف و بخش بندی مشتریان می‌باشد. تجزیه و تحلیل مشتریان هدف شامل جستجوی بخش‌های سودآور مشتریان از طریق ویژگی‌های اساسی مشتریان است. در حالی که بخش بندی مشتریان شامل تقسیم کل مشتریان به گروه‌ها و بخش‌های کوچک‌تر مشتریان است که شامل مشتریان نسبتاً مشابه در هر گروه می‌باشند [۳۲].

۲-۴-۱-۲- جذب مشتری

این فاز بعد از فاز شناسایی صورت می‌گیرد. بعد از شناسایی مشتریان بالقوه، سازمان‌ها می‌توانند با تلاش و تخصیص منابع مستقیم مشتریان هدف را جذب نمایند. یک عنصر جذب مشتری، بازاریابی مستقیم می‌باشد. بازاریابی مستقیم شامل فرآیند ارتقاء است که مشتریان را برای سفارش از طریق کانال‌های مختلف برمی‌انگیزاند [۳۲].

۳-۴-۱-۲- حفظ مشتری

فاز حفظ مشتری، بحث اصلی مدیریت ارتباط با مشتری است. رضایت مشتری، اشاره به انتظارات مشتری با ادراک رضایتمندی می‌باشد، شرط اساسی برای حفظ مشتری می‌باشد. عناصر حفظ مشتری شامل بازاریابی فرد به فرد (بازاریابی شخصی که با تجزیه و تحلیل پشتیبانی می‌شود)، تشخیص و پیش‌بینی تغییر در مشتری، برنامه‌های وفاداری (شامل: فعالیت‌های پشتیبانی با هدف روابط بلند مدت با مشتریان، رتبه‌بندی اعتباری، کیفیت خدمات یا رضایتمندی و سایر برنامه‌های وفاداری) می‌باشد [۳۲].

۴-۴-۱-۲- توسعه‌ی مشتریان

این فاز شامل توسعه‌ی کثرت تراکنش‌ها، ارزش تراکنش‌ها و سودآوری مشتریان می‌باشد. عناصر توسعه مشتریان شامل تجزیه و تحلیل ارزش دوره‌ی حیات مشتری و فروش Cross and up (اشاره به فعالیت‌های ارتقاء که با هدف افزودن تعداد خدمات مرتبط که یک مشتری در شرکت اتخاذ می‌شود)، تجزیه و تحلیل سبد بازار با هدف حداکثر کردن کثرت تراکنش‌های مشتری و آشکار

کردن روندها در رفتار مشتریان می‌باشد [۳۲]. این چهار بعد را می‌توان به عنوان حلقه بسته مدیریت ارتباط با مشتری در نظر گرفت [۴].

۵-۱-۲- دلایل استفاده از مدیریت ارتباط با مشتری

مدیریت ارتباط با مشتری یک فرآیند تجاری است که تمام جوانب مشخصه‌های مشتری را آدرس دهی می‌کند، دانش مشتری را به وجود می‌آورد، روابط را با مشتری شکل می‌دهد و برداشت آنان را از محصولات یا خدمات سازمان ایجاد می‌کند. مدیریت ارتباط با مشتری با در نظر گرفتن این که چه محصولات یا خدماتی، به چه مشتریانی، در چه زمانی و از چه طریقی عرضه شود، بهبود را در پی خواهد داشت. این مدیریت از اجزای مختلفی تشکیل شده است که پیش از اینکه فرایند آن آغاز شود، شرکت باید اطلاعات مشتری را در اختیار داشته باشد. این اطلاعات می‌تواند از داده‌های داخلی مشتریان و یا از داده‌های منابع خارجی خریداری شده، به دست آید. برای داده‌های داخلی منابع مختلفی وجود دارد (مانند پرسشنامه‌ها و بلاگ‌ها، سوابق کارت اعتباری و ...) منابع داده خارجی یا بانک‌های داده خریداری شده (مانند آدرس‌ها، شماره تلفن‌ها، پروفایل‌های بازدید از وب سایت‌ها)، کلیدی برای به دست آوردن دانش بیشتری از مشتری است. از دیگر علل استفاده از مدیریت ارتباط با مشتری، می‌توان به: تمرکز بر مستحکم کردن ارتباط نزدیک‌تر و عمیق‌تر با مشتریان، سودمندی بیشتر مشتریان فعلی نسبت به مشتریان جدید، تجزیه و تحلیل اطلاعات مشتری برای تصمیم‌گیری تجاری، ارتباط موثر مشتری بر مبنای داده‌های تبدیل شده به اطلاعات، بازاریابی اینترنتی، بازاریابی فرد به فرد و بازاریابی از طریق پایگاه داده، اشاره نمود [۳۴].

۲-۲- داده‌کاوی^۱

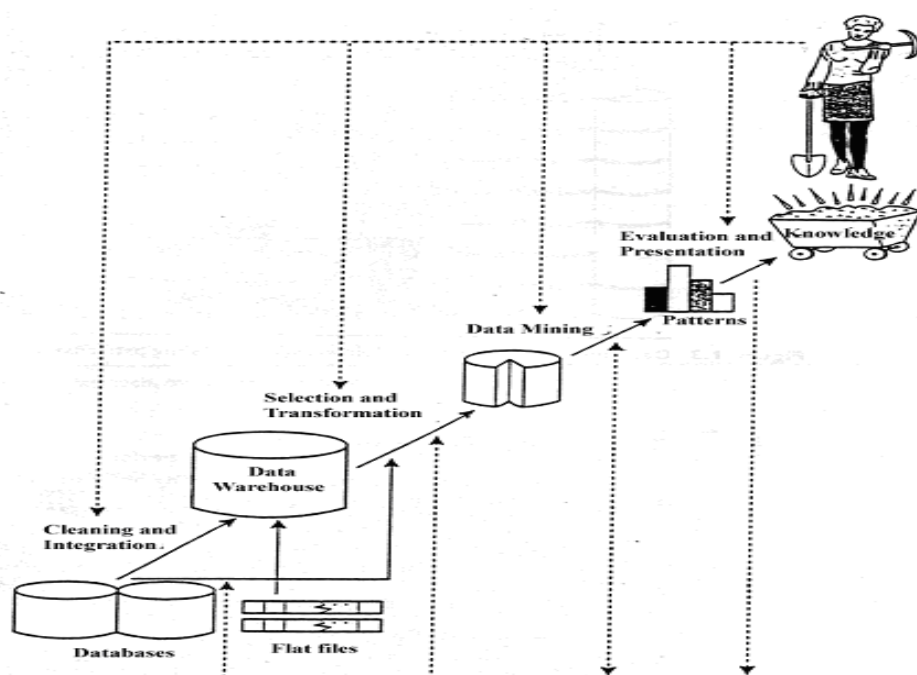
در دو دهه‌ی قبل توانایی‌های فنی بشر برای تولید و جمع‌آوری داده‌ها به سرعت افزایش یافته است. عواملی نظیر استفاده گسترده از بارکد برای تولیدات تجاری، به خدمت گرفتن کامپیوتر در کسب و کار، علوم، خدمات دولتی و پیشرفت در وسائل جمع‌آوری داده، از اسکن کردن متون و تصاویر تا سیستم‌های سنجش از دور ماهواره‌ای، در این تغییرات نقش مهمی دارند.

به طور کلی استفاده همگانی از وب و اینترنت به عنوان یک سیستم اطلاع‌رسانی جهانی ما را مواجه با حجم زیادی از داده و اطلاعات می‌کند. این رشد انفجاری در داده‌های ذخیره شده، نیاز مبرم به وجود تکنولوژی‌های جدید و ابزارهای خودکاری را ایجاد کرده که به صورت هوشمند به انسان یاری رسانند تا این حجم زیاد داده را به اطلاعات و دانش تبدیل کند؛ 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^۲ و دیگر مخازن بزرگ اطلاعات، ذخیره شده است. داده‌کاوی به طور هم‌زمان از چندین رشته علمی، نظیر: تکنولوژی پایگاه داده، هوش مصنوعی، یادگیری ماشین، شبکه‌های عصبی، آمار، شناسایی الگو، سیستم‌های مبتنی بر دانش^۳، حصول دانش^۴، بازیابی اطلاعات^۵، محاسبات سرعت بالا^۶ و بازنمایی بصری داده^۷ بهره می‌برد. داده‌کاوی در اواخر دهه‌ی ۱۹۸۰ پدیدار گشته، در دهه‌ی ۱۹۹۰ گام‌های بلندی در این شاخه از علم برداشته شده و انتظار می‌رود در این قرن به رشد و پیشرفت خود ادامه دهد.

-
- 1 . Data Mining
 - 2 . Data warehouses
 - 3 . Knowledge-based system
 - 4 . Knowledge-acquisition
 - 5 . Information retrieval
 - 6 . High-performance computing
 - 7 . Data visualization

واژه‌های «داده کاوی» و «کشف دانش در پایگاه داده»^۱ اغلب به صورت مترادف یکدیگر مورد استفاده قرار می‌گیرند. کشف دانش به عنوان یک فرآیند در شکل ۱-۲ نشان داده شده است.

کشف دانش در پایگاه داده فرآیند شناسایی درست، ساده، مفید، و نهایتاً الگوها و مدل‌های قابل فهم در داده‌ها می‌باشد. داده کاوی، مرحله‌ای از فرایند کشف دانش می‌باشد و شامل الگوریتم‌های مخصوص داده کاوی است، به طوری که، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده کشف می‌کند [۳۵].



شکل ۱-۲: داده کاوی به عنوان یک مرحله از فرآیند کشف دانش [۳۶].

به بیان ساده‌تر، داده کاوی به فرایند استخراج دانش ناشناخته، درست و بالقوه مفید از داده اطلاق می‌شود. به عبارت دیگر داده کاوی گونه‌ای از تکنیک‌ها برای شناسایی اطلاعات و یا دانش تصمیم‌گیری از قطعات داده می‌باشد، به نحوی که با استخراج آن‌ها، در حوزه‌های تصمیم‌گیری، پیش‌بینی، پیشگویی، و تخمین مورد استفاده قرار گیرند. داده‌ها اغلب حجیم، اما بدون ارزش

می‌باشند، داده به تنهایی قابل استفاده نیست، بلکه دانش نهفته در داده‌ها قابل استفاده می‌باشد. به این دلیل اغلب به داده کاوی، تحلیل داده‌ای ثانویه^۱ گفته می‌شود.

داده‌کاوی استخراج و تحلیل مقدار زیادی داده به منظور کشف قوانین و الگوهای معنی دار در آن‌هاست. هدف اصلی داده کاوی، استخراج الگوهای از داده‌ها، افزایش ارزش اصلی آنان و انتقال داده‌ها به صورت دانش است. کشف دانش در پایگاه داده فرایند شناسایی درست، ساده، مفید و نهایتاً الگوها و مدل‌های قابل فهم در داده‌ها می‌باشد. داده‌کاوی، مرحله‌ای از فرایند کشف دانش می‌باشد و شامل الگوریتم‌های مخصوص داده‌کاوی است؛ به طوری که، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده کشف می‌کند [۳۵]. استخراج دانش در پایگاه داده^۲ (*KDD*)، به عنوان روالی برای شناسایی الگوهای معتبر، جدید، بالقوه مفید، و سرانجام قابل فهم در داده‌ها، تعریف شده است. روال سراسری شامل تبدیل داده سطح-پایین به دانش سطح-بالاست [۳۷].

۱-۲-۲- تعریف داده کاوی

داده‌کاوی دارای تعاریف متنوعی می‌باشد. این تعاریف به مقدار زیادی به پیش زمینه‌ها و نقطه نظرهای افراد بستگی دارد. در اینجا به ارائه برخی از این تعاریف می‌پردازیم:

(۱) داده‌کاوی استفاده از روش‌های آماری با استفاده از کامپیوتر برای کشف اطلاعات مفید

است [۳۸].

1 . Secondary Data Analysis

2 . Knowledge discovery in databases

(۲) داده کاوی فرآیند کشف الگوهای مفید از داده است [۳۹].

(۳) داده کاوی به فرآیند استخراج اطلاعات نهفته، قابل فهم، قابل تعقیب از پایگاه داده‌های بزرگ و استفاده از آن در تصمیم گیری‌های تجاری مهم اطلاق می‌شود [۴۰].

(۴) داده کاوی فرآیند پشتیبانی تصمیم است که در آن از پایگاه داده‌های بزرگ برای کشف الگوهای ناشناخته و غیر منتظره از اطلاعات، استفاده شده است [۴۰].

(۵) داده کاوی یک فرآیند شناخت الگوهای معتبر، جدید، ذاتاً مفید و قابل فهم از داده‌ها می‌باشد [۳۸].

۲-۲-۲- علل گرایش به داده کاوی

برخی از علل گرایش به داده کاوی به شرح ذیل می‌باشد [۴۱]:

- (۱) حجم زیاد داده در دسترس: در دهه‌ی اخیر، قیمت سخت افزار به ویژه فضای دیسک سخت، بسیار کاهش یافته است و به دنبال آن، تشکیلات اقتصادی مقدار زیادی از داده‌ها را از کاربردهای زیادی گردآوری کرده‌اند. با این انفجار داده‌ها، تشکیلات اقتصادی می‌خواهند که الگوهای پنهان در این داده‌ها را برای هدایت استراتژی‌های تجارت خود بکار گیرند. داده کاوی هنگامی بیشترین معنی را پیدا می‌کند که داده‌های زیادی وجود داشته باشد. اغلب الگوریتم‌های داده کاوی نیازمند میزان زیادی از داده‌ها هستند تا مدل‌هایی را ترتیب دهند که بعداً برای دسته بندی، تخمین، پیش بینی یا سایر کارکردهای داده کاوی مورد استفاده قرار گیرند.

۲) افزایش رقابت: رقابت به علت وجود بازارهای مدرن و کانال‌های توزیع مثل اینترنت و ارتباطات راه دور، به طور فزاینده‌ای در حال افزایش است. تشکیلات اقتصادی با رقابت‌های جهان وب مواجهند و کلید موفقیت در تجارت، حفظ مشتریان کنونی و به دست آوردن مشتریان جدید است. داده کاوی، تکنولوژی‌هایی دارد که اجازه می‌دهد که تشکیلات تجاری فاکتورهایی را برای مواجهه با این زمینه‌ها تحلیل کند.

۳) آماده بودن تکنولوژی آن: داده‌کاوی قبلاً فقط در حوزه‌ی آکادمیک قرار داشت، اما در حال حاضر بسیاری از این تکنولوژی‌ها کامل شده‌اند و برای اعمال در صنعت آماده‌اند. الگوریتم‌ها، بسیار دقیق‌تر و کاراتر شده‌اند و می‌توانند به طور فزاینده‌ای داده‌های پیچیده را مدیریت کنند. بعلاوه رابط برنامه نویسی کاربردهای داده کاوی^۱ اکنون استاندارد شده‌اند، که به توسعه دهندگان این امکان را می‌دهند که کاربردهای داده‌کاوی بهتری بسازند.

علاقه به مدیریت روابط با مشتریان فراوان است؛ در طیف وسیعی از صنایع، شرکت‌ها به این بینش رسیده‌اند که مشتریان برای سازمان حیاتی هستند و اطلاعات درباره آن مشتری‌ها یکی از دارایی‌های اساسی سازمان می‌باشد. اطلاعاتی که شرکت‌ها درباره مشتریان‌شان دارند نه تنها برای خودشان بلکه برای دیگران هم ارزشمند است. اطلاعات یک محصول است. یک شرکت کارت اعتباری چیزهایی می‌داند که شرکت‌های خطوط هوایی دوست دارند بدانند یعنی چه کسی بلیط‌های پرواز متعددی می‌خرد. گوگل می‌داند مردم در وب دنبال چه چیزی هستند و از این شناخت با فروش لینک‌هایی با پشتیبان مالی بهره می‌برد. در واقع هر شرکتی که داده‌های با ارزش جمع‌آوری می‌کند در موقعیت یک واسطه‌ی اطلاعات قرار دارد [۴۱].

۳-۲-۲- مراحل کشف دانش

کشف دانش دارای مراحل تکراری زیر است [۳۵]:

- ۱) پاک‌سازی داده‌ها^۱ (از بین بردن اغتشاش و ناسازگاری داده‌ها).
- ۲) یکپارچه‌سازی داده‌ها^۲ (چندین منبع داده ترکیب می‌شوند).
- ۳) انتخاب داده‌ها^۳ (داده‌های مرتبط با آنالیز از پایگاه داده بازیابی می‌شوند).
- ۴) تبدیل کردن داده‌ها^۴ (تبدیل داده‌ها به فرمی که مناسب برای داده کاوی باشد مثل خلاصه سازی^۵ و همسان سازی^۶).
- ۵) داده کاوی (فرایند اصلی که روال‌های هوشمند برای استخراج الگوها از داده‌ها به کار گرفته می‌شوند).
- ۶) ارزیابی الگو^۷ (برای مشخص کردن الگوهای صحیح و مورد نظر به وسیله معیارهای اندازه گیری).
- ۷) ارائه دانش^۸ (یعنی نمایش بصری، تکنیک‌های بازنمایی دانش برای ارائه دانش کشف شده به کاربر استفاده می‌شود).

-
- 1 . Data cleaning
 - 2 . Data integration
 - 3 . Data selection
 - 4 . Data transformation
 - 5 . Summary
 - 6 . Aggregation
 - 7 . Pattern evaluation
 - 8 . Knowledge presentation

۴-۲-۲- کاربردهای داده کاوی

داده کاوی در بسیاری از شاخه‌ها همچون بازاریابی امور مالی، بانکداری، تولید، پزشکی، مدیریت ارتباط با مشتری، رد یابی، پیش بینی خرابی‌ها و آموزش سازمانی کاربرد دارد. برخی از کاربردهای داده کاوی به طور خلاصه به شرح ذیل است.

❖ کاربردهای معمول تجاری: از قبیل تحلیل و مدیریت بازار، تحلیل سبد بازار، بازاریابی، فهم

رفتار مشتری، تحلیل و مدیریت ریسک

❖ مدیریت و کشف فریب: کشف فریب تلفنی، کشف فریب بیمه‌ای و اتومبیل، کشف حقه‌های

کارت اعتباری، کشف تراکنش‌های مشکوک مالی (پولشویی)

❖ متن کاوی: پالایش متن (نامه‌های الکترونیکی، گروه‌های خبری، جستجوی مقالات و

موضوعات خاص و غیره)

❖ پزشکی: کشف ارتباط علامت و بیماری، تحلیل آرایه‌های DNA و تصاویر پزشکی

❖ ورزش: آمارهای ورزشی

❖ وب کاوی: پیشنهاد صفحات مرتبط، بهبود ماشین‌های جستجوگر یا شخصی سازی حرکت

در وب سایت.

۵-۲-۲- مسائل کسب و کار برای داده‌کاوی

همان‌طور که در قسمت قبلی اشاره شد، تکنیک‌های داده‌کاوی می‌تواند برای کاربردهای

بسیاری بکار رود، در ادامه تعدادی از مسائل کسب و کار معمولی را که می‌توان با داده‌کاوی برای

آنان پاسخی یافت، شرح داده می‌شود [۴۱]:

(۱) تحلیل روی گردانی^۱: کدام مشتریان بیشتر تمایل دارند به سمت رقیب ما بروند؟ امروزه صنایع مخابرات، بانکداری و بیمه، بیش از دیگران در معرض این رقابت‌ها هستند. به طور متوسط، هر مشترک جدید تلفن همراه، برای شرکت تلفن، هزینه‌ای بالغ بر ۲۰۰ دلار در بازار سرمایه گذاری دارد. هر کسب و کاری می‌خواهد مشتریان بیشتری را کسب کند. تحلیل روی گردانی می‌تواند به مدیران بازاریابی کمک کند تا دلایل روی گردانی مشتری را درک کند و روابط مشتری را بهبود دهد و وفاداری مشتری را افزایش دهد.

(۲) فروش متقاطع^۲: مشتریان چه محصولاتی را دوست دارند بخرند؟ فروش متقاطع برای خرده فروشان یک چالش تجاری بزرگ است. بسیاری از خرده فروشان، به ویژه خرده فروشان بر خط، برای افزایش فروش خود از این ویژگی استفاده می‌کنند. برای مثال، اگر شما به یک کتاب‌فروشی بر خط مثل سایت آمازون^۳ برای خرید یک کتاب، بروید، حتماً متوجه شده‌اید که وب سایت مجموعه‌ای از پیشنهادات درباره کتاب‌های مرتبط را به شما پیشنهاد می‌کند. این پیشنهادات را می‌توان از تحلیل داده کاوی گرفت.

(۳) کشف تقلب: آیا این ادعای بیمه، کلاهبرداری است؟ شرکت‌های بیمه، روزانه هزاران دعوی بیمه دارند. برای آنان مهم است که درباره هر مورد تحقیق و بررسی کنند. داده‌کاوی می‌تواند برای شناسایی دعاوی‌ای که بیشتر نادرست هستند، کمک کند.

(۴) مدیریت ریسک: آیا باید این درخواست وام برای این مشتری تصویب شود؟ این یک سؤال بسیار رایج در سناریوهای بانکی است. تکنیک‌های داده‌کاوی می‌توانند برای رتبه بندی سطح ریسک یک مشتری، بکار روند و به مدیر در گرفتن یک تصمیم مناسب برای هر کاربرد، کمک کنند.

1 . Churn Analysis
2 . Cross selling
3 . Amazon.com

۵) **قطعه‌بندی مشتریان:** چه کسی مشتری من است؟ قطعه‌بندی مشتریان به مدیران بازاریابی کمک می‌کند که تفاوت‌های پروفایل‌های مشتریان را درک کنند و عمل بازاری مناسبی را بر مبنای هر بخش، انجام دهند.

۶) **تبلیغات هدف‌دار:** چه بنر تبلیغی باید برای یک بازدید کننده‌ی خاص، نمایش داده شود؟ فروشندگان وب و سایت‌های پورتال تمایل دارند که محتوای خود را برای مشتریان‌شان سفارشی کنند. با استفاده از الگوهای ناوبری مشتری یا خرید بر خط، این سایت‌ها می‌توانند راه حل‌های داده‌کاوی را برای نمایش تبلیغات هدف دار برای مشتریان‌شان، بکار ببرند.

۷) **پیش‌بینی فروش:** من چه نمونه‌هایی را در این فروشگاه در این هفته خواهم فروخت؟ تکنیک‌های پیش‌بینی داده‌کاوی می‌تواند برای پاسخ به انواع این پرسش‌های مرتبط با زمان، بکار روند.

۶-۲-۲- فرآیند داده‌کاوی در مدیریت ارتباط با مشتری

از نقطه نظر ادبیات، چارچوب CRM می‌تواند به عملیاتی و تجزیه و تحلیلی دسته‌بندی شود. CRM عملیاتی به خودکار کردن فرآیندهای کسب و کار اشاره دارد، در حالی که CRM تجزیه و تحلیلی می‌تواند به سازمان جهت تشخیص بهتر و اختصاص مؤثرتر منابع به گروه‌های سودآور مشتریان کمک کند. تکنیک‌های داده‌کاوی، ابزارهای مشهور تجزیه و تحلیل داده‌های مشتریان در چارچوب CRM تجزیه و تحلیلی می‌باشد. اکثر سازمان‌ها داده‌های ارزشمندی درباره‌ی مشتریان جاری‌شان، مشتریان بالقوه، تأمین‌کنندگان و شرکای تجاری‌شان جمع‌آوری و ذخیره کرده‌اند. به هر حال عدم توانایی برای کشف اطلاعات ارزشمند پنهان در داده‌ها، مانع حرکت سازمان‌ها از داده به دانش سودمند و ارزشمند شده است. سیستم مدیریت ارتباط با مشتری باید با به کار گیری ابزارهای

موجود مانند داده‌کاوی، داده‌ها را تجزیه و تحلیل کند. اعم از اینکه شرکت تکنیک‌های آماری سستی را به کار می‌برد یا یکی از ابزارهای نرم افزاری داده‌کاوی را، کارشناسان نیاز به فهم داده‌های مشتری و روابط تجاری دارند. بنابراین، داشتن افرادی متخصص که این داده‌ها را با ابزارهای مربوطه استخراج و به صورت اطلاعات درآورند، مهم است [۴۲]. ابزارهای داده‌کاوی می‌تواند به این سازمان‌ها جهت کشف دانش پنهان در حجم داده‌ها کمک نماید.

هدف کلی CRM ایجاد درک عمیق‌تر مشتریان برای حداکثر کردن ارزش مشتریان در بلند مدت می‌باشد. تجزیه و تحلیل و درک رفتارها و ویژگی‌های مشتریان بر اساس توسعه‌ی یک استراتژی CRM رقابتی می‌باشد که جهت جذب و حفظ مشتریان بالقوه و حداکثر کردن ارزش مشتریان می‌باشد. ابزارهای مناسب داده‌کاوی که برای استخراج و شناسایی دانش و اطلاعات مفید از حجم عظیم داده‌های مشتریان می‌باشند، یکی از ابزارهای حمایتی برای ساختن تصمیمات مختلف CRM می‌باشد؛ بنابراین تکنیک‌های داده‌کاوی می‌تواند به انجام یک چنین اهدافی برای استخراج یا تشخیص رفتار و ویژگی‌های مشتریان از پایگاه داده کمک نمایند. جنبه‌ی عمومی داده‌کاوی شامل ساختن مدل از داده‌ها می‌باشد. هر یک از تکنیک‌های داده‌کاوی می‌تواند شامل مدل سازی‌های ذیل باشد [۴۲]:

(۱) همبستگی

(۲) دسته بندی^۱

(۳) خوشه بندی^۲

(۴) پیش بینی

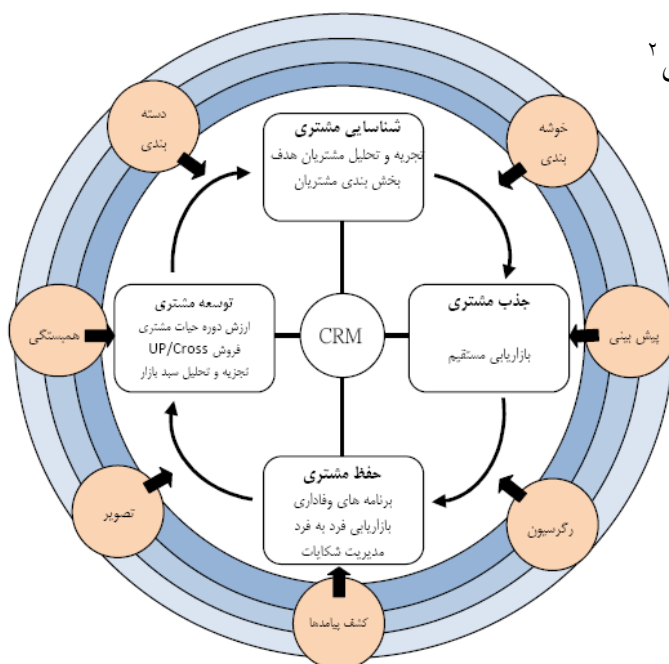
1.Classification

2.Clustering

(۵) رگرسیون

(۶) کشف پیامدها^۱

(۷) نمایه سازی^۲



شکل ۲-۲: چارچوب دسته بندی تکنیک‌های داده کاوی در CRM [۴۲].

۳-۲- شبکه‌های عصبی مصنوعی^۳

مغز انسان یک سیستم پردازش اطلاعات با ساختاری موازی، از 10^{11} نرون^۴ به هم مرتبط با تعداد 10^{16} ارتباط، تشکیل شده است. نرون‌ها ساده‌ترین واحد ساختاری سیستم‌های عصبی هستند. بافت‌هایی که عصب نامیده می‌شوند، اجتماعی از نرون‌ها هستند که اطلاعات و پیام‌ها را از یک قسمت بدن به قسمت دیگر بدن منتقل می‌کنند [۵].

1.Sequence discovery

2.Visualization

3. Artificial Neural Networks (ANN)

4. Neuron

گرچه همه‌ی نرون‌ها کارکرد یکسانی دارند، ولی اندازه و شکل آنان بستگی به محل استقرارشان در سیستم عصبی دارد. با وجود این همه تنوع، بیشتر نرون‌ها از سه قسمت اساسی تشکیل شده- اند [۵]:

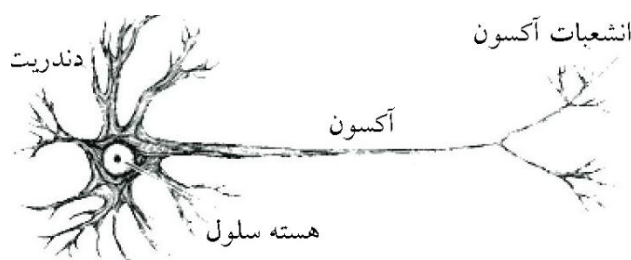
(۱) بدنه‌ی سلول (که شامل هسته و قسمت‌های حفاظتی دیگر می‌باشد).

(۲) دندریت^۱

(۳) آکسون^۲

که دو قسمت آخر عناصر ارتباطی نرون را تشکیل می‌دهند [۵]. شکل ۲-۳ ساختمان سلول

عصبی را نشان می‌دهد.



شکل ۲-۳: نواحی اصلی یک سلول عصبی [۵].

دندریت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی عمل کرده و سیگنال‌ها را به هسته‌ی

سلول منتقل می‌کنند. بدنه‌ی سلول، انرژی لازم برای فعالیت نرون را فراهم نموده و بر روی سیگنال-

های دریافتی عمل می‌کند، که با یک «عمل ساده جمع و مقایسه با یک سطح آستانه» مدل می‌گردد.

اکسون طول بیشتری نسبت به دندریت داشته اما از سطحی هموارتر و تعداد شاخه‌های کمتری

برخوردار می‌باشد. وظیفه‌ی آکسون انتقال سیگنال‌های الکتروشیمیایی دریافتی از هسته‌ی سلول به

دندریت‌های نرون دیگر است. محل تلاقی یک آکسون از یک نرون به دندریت‌های یک نرون دیگر

1. Dendrite

2. Axon

را سیناپس^۱ می‌نامند. سیناپس‌ها واحدهای ساختاری کوچک تابعی هستند که ارتباط بین نرون‌ها را برقرار می‌سازند [۵].

پیام‌های عصبی تنها به صورت یک طرفه حرکت می‌کنند: از دندریت به بدنه‌ی سلول و سپس به آکسون. زمانی که پیام به انتهای آکسون می‌رسد، فیبرهای ماهیچه‌ای، سلول‌های غددی و دیگر نرون‌هایی که به آکسون متصل هستند را، تحریک می‌کند [۵].

۱-۳-۲- کاربردهای شبکه‌های عصبی

با عنایت به این که شبکه‌های عصبی از دو ویژگی اساسی یادگیری یا نگاشت پذیری بر اساس ارائه‌ی داده‌های تجربی (قدرت و توانایی تعمیم‌پذیری) و ساختار پذیری موازی برخوردار می‌باشند؛ این شبکه‌ها برای مسائل کنترل، علی‌الخصوص سیستم‌های پیچیده که مدل سازی این سیستم‌ها یا میسر نیست و یا به سختی انجام می‌شود، بسیار مناسب می‌باشند. به طور خلاصه می‌توان جمع بندی زیر را از کاربردهای شبکه‌های عصبی ارائه نمود [۵]:

❖ طبقه بندی، شناسایی و تشخیص الگو

❖ پردازش سیگنال

❖ پیش بینی سری‌های زمانی

❖ مدل سازی و کنترل

❖ بهینه سازی

❖ سیستم‌های خبره و فازی

❖ مسائل مالی، بیمه، امنیتی، بازار بورس و وسایل سرگرم کننده

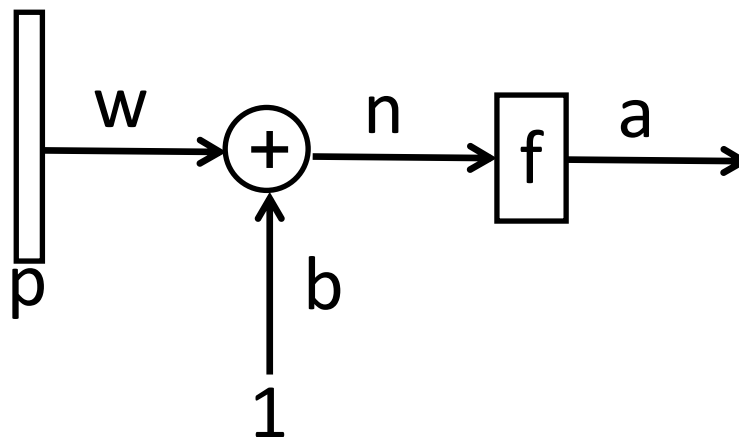
❖ ساخت وسایل صنعتی، پزشکی و امور حمل و نقل

۲-۳-۲ مدل ریاضی نرون

نرون کوچک‌ترین واحد پردازشگر اطلاعات است، که اساس عملکرد شبکه‌های عصبی را

تشکیل می‌دهد. شکل ۲-۴ ساختار یک نرون تک ورودی را نشان می‌دهد که در آن اسکالرها p و

a به ترتیب ورودی و خروجی می‌باشند [۵].



شکل ۲-۴: مدل نرون تک ورودی [۵].

میزان تأثیر p بر a به وسیله‌ی مقدار اسکالر w تعیین می‌شود. ورودی دیگر که مقدار ثابت ۱

است، در جمله‌ی بایاس b ضرب شده و سپس با wp جمع می‌گردد که این مقدار، ورودی خالص n

برای تابع محرک (یا تابع تبدیل) f خواهد بود و بدین صورت خروجی نرون با معادله‌ی زیر تعریف

می‌شود [۵]:

$$a = f(wp + b)$$

در مقایسه این مدل تک ورودی با یک نرون بیولوژیکی، w معادل شدت سیناپس، مجموعه‌ی جمع کننده و تابع محرک معادل هسته‌ی سلول و a معادل سیگنال گذرنده از آکسون خواهد بود. نکته‌ای که باید به آن توجه شود، اهمیت و تأثیر جمله‌ی بایاس b است. این جمله را می‌توان مانند وزنه w در نظر گرفت، با این تصور که میزان تأثیر ورودی ثابت 1 را روی نرون منعکس می‌سازد [۵].

۴-۳-۲-فرآیند یادگیری

به فرآیند یادگیری نیاز است چون اطلاعات (ارتباط ورودی و خروجی) کاملاً مشخص نیستند. تجربه‌ها در طول مسیر زمان حاصل می‌شوند و به عبارت دیگر هیچ کس آینده‌ی خود را تجربه نکرده است. میزان یادگیری ما به درجه‌ی کامل بودن اطلاعات قبلی ما بستگی دارد. در حالت کلی دو نوع یادگیری موجود است: یادگیری با ناظر^۱ و یادگیری بدون ناظر^۲ [۵].

در یادگیری با ناظر فرض بر این است که در هر مرحله‌ی تکرار الگوریتم یادگیری، جواب مطلوب سیستم یادگیرنده از قبل آماده است و به عبارتی الگوریتم یادگیری به جواب واقعی و مطلوب دسترسی دارد. برای مثال فرض کنید که سیستم یادگیر می‌خواهد نگاشت $y = x^2$ را بیاموزد. اگر به سیستم مقدار 0.5 را بدهیم جواب مطلوب 0.25 را به دست می‌دهد. به طور کلی جوابی را که سیستم یادگیر با وضعیت قبلی پارامترهایش می‌دهد، به عنوان جواب واقعی در نظر بگیرد. در این جا الگوریتم یادگیرنده که پارامترهای سیستم یادگیرنده را تنظیم می‌کند، هم به جواب مطلوب 0.25 و هم به جواب واقعی دسترسی داشته و خطای یادگیری که همان خطای بین مقدار مطلوب و واقعی است، دسترسی خواهد داشت [۵].

1 . Suervised Learning

2 . Unsupervised Learning

در یادگیری بدون ناظر جواب مطلوب برای سیستم یادگیرنده، موجود نیست. بنابراین خطای یادگیری جهت بهبود رفتار سیستم یادگیرنده، در دسترس نمی‌باشد. یادگیری بدون ناظر را خود یادگیری^۱ نیز می‌نامند. در نگاه اول به نظر می‌رسد که یادگیری بدون ناظر بی فایده و غیر ممکن باشد و به نظر نمی‌آید که این سیستم به تنهایی بتواند مشخصه‌هایی که موجب گروه‌بندی می‌شوند را، بیابد و ویژگی‌های دیگر را نادیده انگارد. اما حقیقت این است که این گونه سیستم‌های یادگیری می‌آموزند که الگوهای ورودی را به تعداد متناهی از گروه‌ها تقسیم بندی کنند. در فرآیند یادگیری سه مورد ذیل باید به ترتیب انجام شوند [۵]:

- (۱) سیستم یادگیرنده توسط محیط تحریک شود.
- (۲) قانون یادگیری با رجوع به نتیجه‌ی تحریک، پارامترهای سیستم یادگیری را تغییر دهد.
- (۳) سیستم یادگیرنده به خاطر تغییراتی که در ساختار داخلی آن اتفاق افتاده است، پاسخ مناسب‌تری به محیط بدهد.

۵-۳-۲- شبکه‌های عصبی به عنوان سیستم‌های دینامیکی آموزش پذیر

سیستم‌های عصبی به عنوان سیستم‌های یادگیری دارای این توانایی هستند که از گذشته، تجربه و محیط بیاموزند و رفتار خود را در حین یادگیری بهبود بخشند. بهبود در یادگیری در طول زمان باید بر اساس معیاری سنجیده شود [۵].

معیار بهبود، همان هدف سیستم یادگیر را مدل می‌کند. در این جا قانون یادگیری توسط روابط بازگشتی، عموماً به صورت معادلات تفاضلی بیان می‌شود. به این روابط بازگشتی، قوانین یادگیری

می‌گویند. قانون یادگیری روندی است که توسط آن ماتریس وزن‌ها (w) و بردارهای بایاس (b)^۱، شبکه‌ی عصبی تنظیم می‌شوند. هدف قانون یادگیری آموزش شبکه عصبی جهت انجام کار مشخصی است و به عبارتی دیگر شبکه‌های عصبی در خلال آموزش پس از تکرار الگوریتم یادگیری از محیط، شرایط و هدف کار خود بیشتر مطلع می‌گردند [۵].

۴-۲- خوشه‌بندی

خوشه بندی به عنوان یکی از فعالیت‌های داده کاوی می‌باشد و به گروه بندی تراکشن‌ها و مشاهدات یا حالت‌ها و یا کلاس‌های مشابه می‌پردازد. همچنین یک خوشه مجموعه‌ای از رکوردها است که به هم شبیه می‌باشند و از رکوردهای بیرون خوشه تفاوت دارند. در خوشه بندی متغیر هدف وجود ندارد و به طبقه بندی تخمین و پیش گوئی مقدار متغیر هدف نمی‌پردازد [۴۳].

اوریت^۲ چندین تعریف برای یک خوشه بیان کرده است [۶]:

- ❖ هر خوشه، مجموعه‌ای از نمونه‌های متشابه است و نمونه‌های خوشه‌های مختلف نا متشابه‌اند.
 - ❖ خوشه، اجتماعی از نقاط در فضای نمونه است به طوری که فاصله‌ی بین هر دو نقطه در یک خوشه از فاصله بین دو نقطه‌ای که در یک خوشه نیستند کمتر می‌باشد.
 - ❖ خوشه را می‌توان ناحیه‌ای همبند از فضای چند بعدی شامل نقاطی نسبتاً زیاد توصیف کرد.
- به این ترتیب یک خوشه، یک سری داده‌های مشابه می‌باشد که همانند یک گروه واحد رفتار می‌کنند. لازم به ذکر است خوشه‌بندی همان کلاسه‌بندی است، با این تفاوت که کلاس‌ها از پیش

۱. زمانی که ورودی‌ها بیشتر از یک مورد می‌باشند، w به صورت ماتریس و b به صورت بردار خواهد بود.

تعریف‌شده و معین نمی‌باشند و عمل گروه‌بندی داده‌ها بدون نظارت انجام می‌گیرد. یعنی در خوشه بندی متغیر هدف وجود ندارد و به طبقه بندی تخمین و پیش گوئی مقدار متغیر هدف نمی‌پردازد [۴۴].

بر خلاف کلاسه‌بندی و پیشگویی که اشیا داده‌ها را بر اساس کلاس‌ها تحلیل می‌کنند، خوشه‌بندی اشیا داده‌ها را بدون در نظر گرفتن برچسب‌های کلاس، تحلیل و آنالیز می‌نماید. عمدتاً برچسب کلاس‌ها در داده‌های آموزشی به آسانی مشخص نیست زیرا این کلاس‌ها شناخته شده نمی‌باشند. خوشه‌بندی گاهی برای تعیین و تولید چنین برچسب‌هایی بکار می‌رود. اشیای خوشه‌بندی شده بر اساس اصل حداکثر شباهت بین اعضای هر کلاس و حداقل شباهت بین کلاس‌های مختلف گروه‌بندی می‌شوند، یعنی خوشه‌ها به گونه‌ای تنظیم می‌شوند که اشیای داخل هر خوشه بیشترین شباهت را با یکدیگر داشته باشند. هر خوشه به عنوان یک کلاس می‌باشد که قوانین از آن مشتق می‌شوند. ضمناً خوشه‌بندی می‌تواند امکان طبقه‌بندی تشکیلات^۱ را فراهم کند، یعنی سازماندهی مذکور، به صورت سلسله مراتبی از کلاس‌هاست که هر کلاس شامل مشخصه‌های مشابه یکدیگر می‌باشد.

تکنیک‌های خوشه بندی با این هدف شروع بکار می‌کنند که مجموعه‌ای از رکوردها را به چند گروه تقسیم کرده و رکوردهای مشابه را در خوشه‌های یکسان قرار دهند و بدین ترتیب چندین زیر مجموعه داده‌ای را ایجاد می‌نماید [۴۴].

نمونه‌ها در یک گروه مشابه قرار می‌گیرند و بدین ترتیب نمونه‌هایی که در گروه‌های متفاوت قرار دارند مشابه نیستند. مثلاً یک خوشه ممکن است دسته‌ای از مشتریان با سابقه خرید یکسان،

تراکنش‌های مشابه و دیگر فاکتورهای همانند و مشابه یکدیگر باشد. نمونه‌ها در تکنیک خوشه بندی به عنوان یک بردار در یک فضای چند بعدی یا حتی یک نقطه می‌توانند خود را نشان دهند [۴۴].

۱-۴-۲- تعریف فرآیند خوشه‌بندی

خوشه‌بندی فرآیند تقسیم یک سری از داده‌ها یا اشیا به زیر کلاس‌هایی به نام خوشه می‌باشد که توسط این رویه فهم ساختار داده‌ها و گروه‌بندی آنان ساده می‌گردد. هر خوشه، شامل یک سری داده‌های مشابه می‌باشد که به صورت یک گروه رفتار می‌کنند. می‌توان خوشه‌بندی را به صورت کلاسه‌بندی تعریف کرد، با این تفاوت که کلاس‌ها و برچسب آنان از پیش تعریف شده نبوده، عمل کلاسه‌بندی بدون نظارت انجام می‌گیرد.

۲-۴-۲- کیفیت خوشه‌بندی

کیفیت خوشه‌بندی بر اصول زیر متکی است [۴۴]:

❖ داده‌های داخل یک کلاس بیشترین تشابه و داده‌های کلاس‌های متمایز بیشترین تفاوت را دارا باشند.

❖ استفاده از معیارهای صحیح یافتن تشابهات در متدها و پیاده‌سازی آن روش‌ها.

❖ توانایی استخراج برخی الگوهای مخفی موجود در داده‌ها.

❖ چگونگی تعریف و نمایش خوشه‌های انتخابی.

در داده‌کاوی و در زمینه‌ی خوشه‌بندی می‌بایست ملاحظات زیر را منظور کرد [۴۴]:

❖ توسعه‌پذیری

❖ امکان استخراج خوشه‌ها با هر شکل دلخواه

❖ حداقل دانش مورد نیاز در تصمیم‌گیری پارامترهای داخلی

❖ عدم آسیب‌پذیری در مواجهه با خطا

❖ عدم حساسیت به ترتیب رکوردهای ورودی و یا عدم تأثیر مخرب رکوردهای آتی بر

رکوردهای فعلی

❖ پشتیبانی انواع مختلف خصیصه‌ها

❖ پشتیبانی ابعاد بزرگ داده‌ها

۳-۴-۲- خوشه بندی در مدیریت ارتباط با مشتری

انسان به طور فطری تمایل به گروه بندی اشیاء و مفاهیم بر اساس شباهت یا تفاوت (فاصله) دارد. گروه بندی در ادبیات داده کاوی خوشه بندی یا خوشه یابی نامیده می‌شود. روش خوشه بندی برای توصیف گروه‌های مختلف در مجموعه داده به کار می‌رود و بر خلاف دسته بندی خوشه‌ها از قبل مشخص نیستند. معمولاً ابتدا خوشه بندی انجام شده و بعد خوشه‌ها به عنوان نام دسته‌ها برای دسته بندی به کار می‌روند. پس از خوشه بندی، برای تعبیر خوشه‌ها نماینده هر خوشه را در نظر می‌گیرند. نماینده می‌تواند یکی از مشاهدات میانه خوشه و یا میانگین همه مشاهدات یک خوشه در نظر گرفته شود. با توجه به تفاوت و شباهت نماینده هر خوشه به نماینده کل داده‌ها از نظر مشخصات می‌توان هر خوشه را تعبیر کرد. در مدیریت ارتباط با مشتری گروه بندی از آن جهت اهمیت دارد که ایجاد پروفایلی برای مشتریان مختلف را فراهم می‌کند و امکان برنامه ریزی استراتژیک روی گروه‌های

مشتریان را می‌دهد و از طرف دیگر امکان انواع گروه بندی‌های دیگر مانند محصولات مشابه را ایجاد می‌کند. فرق مهم خوشه بندی هدف و دسته بندی این است که مقوله از پیش تعریف شده برای خوشه بندی هدف وجود ندارد [۷].

۴-۲-۴- هدف از خوشه بندی

هدف از خوشه بندی این است که داده‌های موجود را به چند گروه تقسیم کنند و در این تقسیم بندی داده‌های گروه‌های مختلف باید حداکثر تفاوت ممکن را به هم داشته باشند و داده‌های موجود در یک گروه باید بسیار به هم شبیه باشند [۳۵].

در خوشه بندی، گروه‌ها از قبل مشخص نمی‌باشند و همچنین معلوم نیست که بر حسب کدام خصوصیات گروه بندی صورت می‌گیرد. در نتیجه پس از انجام خوشه بندی باید یک فرد خبره خوشه‌های ایجاد شده را تفسیر کند و در بعضی مواقع لازم است که پس از بررسی خوشه‌ها بعضی از پارامترهایی که در خوشه بندی در نظر گرفته شده‌اند ولی بی ربط بوده یا اهمیت چندانی ندارند حذف شده و جریان خوشه بندی از اول صورت گیرد [۳۵]. پس از اینکه داده‌ها به چند گروه منطقی و توجیه پذیر تقسیم شدند از این تقسیم بندی می‌توان برای کسب اطلاعات در مورد داده‌ها یا تقسیم داده‌های جدید استفاده کرد [۳۹].

به عنوان مثال می‌توان از خوشه بندی برای داده‌هایی که با سایر داده‌ها تفاوت چشمگیر دارد استفاده نمود. مثلاً یکسری از مشتریان همگی خریدی بالای ۱۰۰ دلار در ماه دارند به غیر از یکی. یا تعدادی فروشگاه که همگی بجز دو فروشگاه دارای سودآوری ۱۰۰ درصد در سال گذشته برای

محصول بوده‌اند. پس از تعیین این داده‌های متفاوت می‌توان ریشه را بررسی نمود. از مهم‌ترین الگوریتم‌هایی که برای خوشه بندی استفاده می‌شوند می‌توان کوهونن و الگوریتم کا-میانگین^۱ را نام برد [۳۹].

۵-۴-۲- برخی از الگوریتم‌های خوشه بندی

در این قسمت به بررسی دو الگوریتم خوشه بندی، کا-میانگین و کوهونن پرداخته می‌شود.

۱-۵-۴-۲- الگوریتم کا-میانگین

در کشور ایالات متحده اخیراً مطالعه‌ای برای چگونگی طراحی لباس‌های افراد نظامی صورت گرفته است و هدف از این کار کاهش اندازه‌های مختلف لباس‌ها در انبار و تعلق لباس‌های متناسب با هیکل افراد به آنان می‌باشد. برای این کار یک پایگاه داده از ۳۰۰۰ نفر از اشکال بدن افراد تشکیل شد و بیش از ۱۰۰ اندازه گیری برای هر نفر صورت گرفت و از یکی از روش‌های خوشه بندی به نام کا-میانگین استفاده شد [۸].

اولین کار مشخص کردن تعداد خوشه‌ها (K) در الگوریتم کا-میانگین است. سپس K مقدار به عنوان مقادیر اولیه یا مرکز ثقل یا میانگین خوشه‌ها^۲ تعیین می‌شوند. هر مرکز، ترکیبی از مقادیر اندازه گیری‌ها می‌باشد. یک مرکز می‌تواند یکی از نقاط واقعی باشد مثلاً اندازه گیری‌های واقعی یکی از افراد نظامی باشد ولی این مطلب یک الزام نمی‌باشد. سپس هر رکورد و یا داده با توجه به اینکه به کدام مرکز نزدیک‌تر است به یک خوشه تخصیص داده می‌شود. سپس مرکز (یا میانگین) خوشه‌های

1 . K-means

2 . Seed

جدید به عنوان مراکز جدید تعیین می‌شود. تا زمانی که مراکز خوشه‌ها در حال تغییر می‌باشد بعضی نقاط ممکن است از یک خوشه به خوشه دیگر بروند. در واقع این نواحی خوشه‌ها است که تغییر می‌کند نه خود نقاط. بعد از تعدادی تکرار مرکز خوشه‌ها دیگر تغییر نمی‌کند و ثابت می‌ماند و این اندازه‌ها همان اندازه‌های لباس‌های جدید نظامی است [۹].

هر سابقه در مقایسه با یکسری خوشه‌ها که به وسیله مراکزشان مشخص شده‌اند در یک خوشه قرار می‌گیرد و به خوشه‌ای تعلق می‌گیرد که به آن نزدیک‌تر است و این باعث تغییر ارزش خوشه‌ای می‌شود که شئی جدید به آن تعلق گرفته است. این کار آنقدر انجام می‌شود و داده‌ها آنقدر به خوشه‌ها تعلق می‌گیرند تا جواب بهینه حاصل شود (داده‌هایی که به خوشه‌ها تخصیص داده شده‌اند نیز دوباره به خوشه‌ای که به آنان نزدیک است تخصیص داده می‌شوند، زیرا از آنجایی که مرکز و ارزش خوشه در حال تغییر است ممکن است دیگر داده به مرکز جدید خوشه مورد نظر نزدیک نباشد). در خوشه بندی باید عوامل متعددی را تحلیل نماییم به عنوان مثال اگر به دنبال خوشه‌هایی در بین مغازه‌های سوپر مارکت باشیم، از جمله عوامل مهم عبارتند از: میزان مراجعه در هر ماه، میزان خرید در هر مراجعه، میزان خرید از هر نوع محصول، مدت زمان باقی ماندن در مغازه و روش پول دادن [۹].

«روش کامیانگین علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش روشی انحصاری و مسطح محسوب می‌شود» [۴۶]. برای این الگوریتم شکل‌های مختلفی بیان شده است ولی همه‌ی آنان دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

❖ بدست آوردن نقاطی به عنوان مراکز خوشه‌ها این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.

❖ نسبت دادن هر نمونه داده به یک خوشه که آن داده کم‌ترین فاصله تا مرکز آن خوشه را دارا باشد.

در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز نقاطی به صورت تصادفی انتخاب می‌شود. سپس داده‌ها با توجه به میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آنان محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود [۴۷].

علی‌رغم اینکه خاتمه‌پذیری الگوریتم بالا تضمین شده است ولی جواب نهایی آن واحد نبوده و همواره جوابی بهینه نمی‌باشد. به طور کلی روش ساده مذکور دارای مشکلات ذیل است [۴۷]:

- ❖ جواب نهایی به انتخاب خوشه‌های اولیه وابستگی دارد.
- ❖ روالی مشخص برای محاسبه‌ی اولیه‌ی مراکز خوشه‌ها وجود ندارد.
- ❖ اگر در تکراری از الگوریتم تعداد داده‌های متعلق به خوشه‌ای صفر شد راهی برای تغییر و بهبود ادامه‌ی روش وجود ندارد.
- ❖ در این روش فرض شده است که تعداد خوشه‌ها از ابتدا مشخص است. اما معمولاً در کاربردهای زیادی تعداد خوشه‌ها مشخص نمی‌باشد.

۲-۵-۲- شبکه‌ی عصبی کوهونن^۱

شبکه‌های عصبی با قابلیت جالب توجهی که در استنتاج معانی از داده‌های پیچیده یا مبهم دارند، می‌توانند برای استخراج الگوها و شناسایی روش‌هایی که آگاهی از آنان برای انسان و دیگر تکنیک‌های کامپیوتری بسیار پیچیده و دشوار است به کار گرفته شوند [۴].

شبکه‌های عصبی نسبت به کامپیوترهای معمولی مسیر متفاوتی را برای حل مسئله طی می‌کنند. کامپیوترهای معمولی از یک مسیر الگوریتمی استفاده می‌کنند، به این معنی که کامپیوتر یک مجموعه از دستورالعمل‌ها را به قصد حل مسئله پی می‌گیرد. اما اگر کامپیوترها می‌توانستند کارهایی را انجام دهند که ما دقیقاً نمی‌دانیم چگونه انجام دهیم خیلی پر فایده بودند. امتیاز یک شبکه عصبی این است که خود کشف می‌کند چگونه مسئله را حل کند؛ یعنی عملکرد آن غیرقابل پیش بینی است. از یک شبکه عصبی متخصص می‌توان برای برآورد وضعیت‌های دلخواه جدید استفاده کرد [۴].

یکی از کاربردهای شبکه‌های عصبی تشخیص الگو است. الگو شناسی با استفاده از شبکه‌های عصبی که از پیش تغذیه شده‌اند و مطابق با موضوع پرورش یافته‌اند، قابل اجرا است. وقتی از شبکه استفاده می‌شود، خودش الگوهای ورودی را شناسایی می‌کند و سعی می‌کند تا خروجی متناسب با آن الگو را ارائه دهد. توان شبکه‌های عصبی زمانی خودش را نشان می‌دهد که الگویی به عنوان ورودی به آن داده شود که از قبل خروجی مرتبط با آن را نداشته، در این وضعیت خروجی متناسب است با یک الگوی ورودی که داری حداقل اختلاف با الگوی داده شده است [۳۵].

شبکه‌های عصبی اطلاعات را به روشی مشابه با کاری که مغز انسان انجام می‌دهد، پردازش می‌کنند. آنان از تعداد زیادی از عناصر پردازشی (سلول عصبی) که بهم پیوسته‌اند، تشکیل شده‌اند که

این عناصر به صورت موازی با هم برای حل یک مسئله مشخص کار می‌کنند. شبکه‌های عصبی با مثال کار می‌کنند و مثال‌ها می‌بایست با دقت انتخاب شوند در غیر این صورت زمان سودمند، تلف می‌شود و یا حتی بدتر از این، شبکه ممکن است نادرست کار کند. امتیاز شبکه عصبی این است که خودش کشف می‌کند که چگونه مسئله را حل کند، عملکرد آن غیر قابل پیش‌گویی است [۳۵].

کوهونن تعریف بسیار جامعی در مورد شبکه‌های عصبی ارائه کرده است:

شبکه‌هایی متشکل از مقادیر فراوان اما متناهی از عناصر ساده، موازی و به هم پیوسته و متصل که دارای سازمان سلسله‌مراتبی هستند و هدف کل نظام برخورد و روبرویی با مسائل دنیای واقعی، از طریق به کار بستن روش‌های مورد استفاده در سلول‌های عصبی بیولوژیکی است [۳۵].

به زبان ساده‌تر شبکه‌های عصبی نظام‌هایی هستند که در طی فرایند آموزش از طریق وظایف نمونه در محیط آزمایشگاهی تجربه و مهارت انجام آن وظایف را به نحو احسن کسب می‌کنند [۳۵].

شبکه‌ی کوهونن یک شبکه‌ی عصبی بدون ناظر است که در آن داده‌های ورودی شبیه به هم در یک فرآیند یادگیری-رقابتی در یک خوشه قرار می‌گیرند. شبکه عصبی شامل مراحل آموزش و یادگیری به طریق زیر است:

- ❖ مرحله‌ی ۱- تعیین مقادیر اولیه وزن‌ها: وزن‌هایی که از N ورودی به M خروجی رفته‌اند مساوی با مقادیر کوچکی قرار می‌گیرند و شعاع اولیه‌ی همسایگی اولیه، تعیین می‌گردد.
- ❖ مرحله‌ی ۲- اعمال ورودی جدید

❖ مرحله‌ی ۳- محاسبه‌ی فاصله با تمامی گره‌ها (نرون‌ها): فاصله‌ی d_j بین ورودی i و گره خروجی j را با استفاده از رابطه‌ی زیر محاسبه می‌گردد:

$$d_j = \sum_{i=0}^{N-1} [x_i(t) - W_{ij}(t)]^2$$

که $x_i(t)$ ورودی سلول عصبی i در زمان t و $W_{ij}(t)$ (سیناپس) وزن از گره ورودی i به گره خروجی j در زمان t است.

❖ مرحله‌ی ۴- انتخاب گره خروجی دارای کمترین فاصله:

گره‌ی j^* به عنوان گره خروجی دارای کمترین d_j انتخاب می‌گردد.

❖ مرحله‌ی ۵- وزن‌های گره j^* و همه‌ی همسایه‌هایش با ورودی تطبیق داده می‌شود.

وزن‌های گره j^* و همه‌ی همسایه‌هایش را که با $NEj^*(t)$ تعریف می‌شود با ورودی تطبیق داده و وزن‌های جدید به صورت زیر تغییر می‌یابد:

$$W_{ij}(t+1) = W_{ij}(t) + \alpha(t)(x_i(t) - W_{ij}(t))$$

برای هر $NEj^*(t)$ و $0 \leq i \leq N-1$ و $\alpha(t)$ ضریب یادگیری می‌باشد که با گذشت زمان کاهش می‌یابد.

❖ مرحله‌ی ۶- تکرار عملیات با رفتن به مرحله‌ی (۲) تا زمانی که نرخ یادگیری و یا وزن‌ها تغییر نیابند.

۵-۲- پیشینه تحقیق

داده کاوی و کشف دانش در پایگاه داده‌ها از جمله موضوع‌هایی هستند که هم‌زمان، با ایجاد و استفاده از پایگاه داده‌ها در اوایل دهه ۸۰ برای جستجوی دانش در داده‌ها شکل گرفت. از آن تاریخ به بعد، پژوهشگران و متخصصان علوم رایانه، آمار، هوش مصنوعی، یادگیری ماشین و ... نیز به پژوهش در این زمینه و زمینه‌های مرتبط با آن پرداخته‌اند [۱۰]. در ادامه پژوهش‌های مرتبط با موضوع تحقیق به تفکیک داخلی و خارجی مورد بررسی قرار می‌گیرند.

۱-۵-۲- پیشینه‌ی خارجی

پژوهش جدی روی موضوع داده کاوی از اوایل دهه ۹۰ شروع شد. پژوهش‌ها و مطالعه‌های زیادی در این زمینه صورت گرفته، همچنین سمینارها، دوره‌های آموزشی و کنفرانس‌هایی نیز برگزار شده است. نتایج پایه‌های نظری داده کاوی در تعدادی از مقاله‌های پژوهشی آورده شده است. مثلاً سال ۱۹۹۱ پیاتسکی-شاپیرو^۱ «کشف دانش از پایگاه داده: کشف، تجزیه و تحلیل و ارائه‌ی قوانین قوی» را ارائه نمود [۵۲]. سال ۱۹۹۵ هافمن و نش استفاده از داده کاوی و داده انبار^۲ توسط بانک‌های آمریکا را بررسی نموده و بیان کردند که چگونه این سیستم‌ها برای بانک‌های آمریکا قدرت رقابت بیشتری ایجاد می‌کنند [۵۳]. چت فیلد مشکلات ایجاد شده توسط داده کاوی را بررسی نمود و همچنین مقاله‌ای تحت عنوان «مدل‌های خطی غیر دقیق داده کاوی و استنباط آماری» ارائه نمود [۵۴]. هندری و مورگان نیز دیدگاه اقتصاد سنجی روی داده کاوی را تهیه کرد [۵۵]. در این سال انجمن داده کاوی هم‌زمان با اولین کنفرانس بین‌المللی «کشف دانش و داده کاوی» شروع به کار کرد. این

1 . Piatetsky-shapiro
2 . Data warehouse

کنفرانس توسعه یافته چهار دوره آموزشی بین‌المللی در پایگاه‌های داده در سال ۱۹۸۹ تا ۱۹۹۴ بود. انجمن مذکور، یک سازمان علمی به نام ACM- SIGKDD را ایجاد نمود [۵۶]. سال ۱۹۹۶ ایمیلنسکی^۱ و منیلا^۲ دیدگاهی از داده کاوی به عنوان «پرس و جو کننده از پایگاه‌های استنتاجی»^۳ را پیشنهاد کردند. فایاد، پیاتسکی و شاپیرو، اودور اسامی پیشرفت‌های کشف دانش و داده کاوی را عنوان کردند. در سال ۱۹۹۷ منیلا خلاصه‌ای از مطالعه روی اساس داده کاوی ارایه نمود [۵۷]. باربارا و همکاران نیز دیدگاه کاهش داده‌ها روی داده کاوی را در گزارش کاهش داده‌های نیوجرسی ارایه نمودند [۵۸]. همچنین می‌توان برای کاربرد داده کاوی در مدیریت مالی می‌توان، تحلیل داده‌های مالی و مدل سازی مالی بنینگاه و چاچ کز^۴ را ملاحظه کرد [۵۹]. فریدمن نیز مقاله‌ای در ارتباط با مفهوم آمار و داده کاوی ارایه نمود [۶۰]. سال ۱۹۹۸ هند^۵ مقاله‌ای تحت عنوان «داده کاوی: آمار یا بیشتر؟» ارائه نمود [۶۱]. کلینبرگ^۶ و همکاران دیدگاه اقتصاد سنجی روی داده کاوی و عملکرد داده کاوی به عنوان یک مسئله‌ی بهینه را ارایه نمودند [۶۲]. در این سال نیز کنفرانس‌های ناحیه‌ای و بین‌المللی در مورد داده کاوی برگزار شد که از جمله می‌توان به کنفرانس آسیا و اقیانوسیه درباره‌ی کشف دانش و داده کاوی اشاره کرد. سال ۲۰۰۱ هند و همکاران و اسمیت بحث‌های مقایسه‌ای بین آمار و داده کاوی را ارایه کردند. سری و استاوا، کولی، رش پاند و تن استفاده از وب در کاوش داده‌ها و کاربردهای آن را ارایه کردند [۶۳]. سال ۲۰۰۲ کلاديو کانورسانو و همکاران «مدل آمیخته چندگانه جمع پذیر تعمیم یافته» برای داده کاوی را بررسی نمودند [۶۴]. پائلو و پاسرینو، «داده کاوی ساختارهای پیوند برای مدل رفتار مصرف کننده» را ارایه نمودند [۱۰].

1 . Imielinski

2 . Manila

3 . Inductive databases

4 . Benninga, Czaczkas

5 . Hand

6 . Kleinberg

یامان و همکاران در تحقیق خود با عنوان «خوشه بندی مشتریان فروشگاه‌های مواد غذایی بر اساس الگوهای بهینه سازی» در استانبول ترکیه، با ارائه‌ی یک مدل ریاضی اقدام به خوشه بندی مشتریان در پنج خوشه نموده‌اند [۴۸].

وست و همکاران اقدام به بررسی ارتباط بین محصولات بر اساس وفاداری مشتریان و خوشه بندی بر اساس تعداد مراجعه به سوپرمارکت و الگوهای هزینه نمودند. در این تحقیق ابتدا تعداد خوشه‌های بهینه با استفاده از الگوریتم شبکه‌ی عصبی کوهونن^۱، ۵ خوشه انتخاب و در ادامه با استفاده از الگوریتم k - میانگین^۲ خوشه بندی صورت گرفته است [۴۹].

۲-۵-۲- پیشینه‌ی داخلی

در کار تحقیقی تحت عنوان «استفاده از داده کاوی برای ارائه مدل رفتاری آژانس‌های فروش بلیط در زنجیره تأمین شرکت‌های هواپیمایی»، محمد رضا داوری و همکاران، برای رفع مشکل پرداخت ثابت (درصد ثابتی از قیمت بلیط) کارمزد به آژانس‌های فروشنده بلیط از طرف شرکت‌های هواپیمایی با استفاده از داده کاوی پرداخته شد. به این منظور ایشان، داده‌های پنج سال اخیر موجود در سیستم فروش بلیط، با استفاده از تکنیک خوشه بندی به بخش‌های چهارگانه که هر کدام شامل تعدادی شهر با شرایط مشابهی می‌باشند، تقسیم کردند. سپس در هر گروه بر اساس مشخصه هر آژانس (سال‌های پنج‌گانه فروش بلیط، ماه‌های دوازده‌گانه شمسی هر سال، درصد فروش بلیط هر دفتر فروش در هر ماه نسبت به فروش کل آن گروه و پیشینه رزرو بلیط در هر دفعه فروش توسط هر آژانس در ماه) خوشه بندی را انجام داده و به دانش الگوهای رفتاری آژانس‌های هواپیمایی در

1. Kohonen
2. K-means

فروش بلیط دست پیدا کرده و بر مبنای آن مدلی پویا و جدید برای پرداخت کار بها به آژانس‌ها ارائه نمودند [۱۱].

همچنین شکور نیاز و حاج علی اکبری در تحقیق خویش تحت عنوان «خوشه بندی داده‌های آماری دانشجویان دانشگاه علم و صنعت و استخراج نمایه ساز توفیقی برای دانشجویان موفق»، با بکار گرفتن فرآیند داده کاوی از پایگاه داده موجود و گروه بندی اطلاعات آماری دانشجویان با استفاده از زمینه‌ی کاری، خوشه بندی بر اساس مشخصه‌های ورودی، به استخراج نمایه سازهای توصیفی از خوشه‌ها پرداخته و در نهایت ویژگی‌های دانشجویان موفق و معیارهای موفقیت‌های آموزش را تعیین نموده‌اند. در تحقیق فوق محدوده داده‌های مورد بحث فایل صفحه گسترده اکسل مربوط به اطلاعات آماری دانشجویان ورودی ۸۲ تا ۸۵ برای سال‌های ۸۲ تا ۸۵ می‌باشد. در این تحقیق بعد از پیاده سازی مراحل داده کاوی و الگوریتم‌های آماری و کامپیانگین، با توجه به معیارهایی خاص، دانشجویان را از لحاظ موفقیت یا پیشرفت، آینده مناسب شغلی خوشه بندی نموده‌اند [۱۲].

شهرابی و صوفی نیستانی در پژوهشی با عنوان «تجزیه و تحلیل میزان و نحوه خرید مشتریان فروشگاه‌های زنجیره‌ای شهروند در مناطق مختلف با استفاده از داده کاوی» به تحلیل سبد خرید مشتریان با استفاده از تکنیک داده کاوی و با استفاده از قوانین وابستگی پرداختند. آنان در تحقیق خود پس از جمع آوری اطلاعات داده‌های مورد نیاز در مورد سبد خرید مشتریان از طریق پرسشنامه، تمام داده‌ها، یکپارچه و جهت تجزیه و تحلیل آماده نمودند. سپس بعد از تعیین عوامل تأثیر گذار در سبد خرید مشتریان، مشتری سه فروشگاه زنجیره‌ای در تهران با استفاده از الگوریتم کامپیانگین و با توجه به عوامل بدست آمده، خوشه بندی شده و سپس قوانین وابستگی در هر خوشه را با استفاده از

الگوریتم علی و معلولی استخراج نمودند. آنان پس از مشخص شدن خوشه‌های هر فروشگاه با استفاده از الگوریتم علی و معلولی، قوانین وابستگی موجود در هر خوشه استخراج و مورد تجزیه و تحلیل قرار دادند. آنان در نهایت مشتریان هر فروشگاه را به ۵ خوشه تقسیم بندی نموده و در هر خوشه نیز قوانین وابستگی استخراج کرده و تحلیل نمودند [۱۳].

شهرابی و صوفیه نیز در تحقیق خود تحت عنوان «استفاده از داده کاوی برای پیش بینی مشتریان بد حساب در مخابرات» به استفاده از چند تکنیک داده کاوی از جمله درخت تصمیم و شبکه عصبی برای شناسایی و پیش بینی مشتریان بد حساب پرداختند. در تحقیق فوق پایگاه داده متشکل از ۳۵۰۰۰ مشترک تلفن ثابت تهران در نظر گرفته شده که شامل کاربری‌های مختلف اعم از خانگی، تجاری، اداری و غیره می‌باشند. با توجه به حجم بالای داده به ۴۰۵۶ مشتری کاهش دادند در مراحل بعد آنان بدون برچسب خوش حسابی و بد حسابی با استفاده از ساختار ترکیبی متشکل از دو تکنیک کامیانگین و SOM، در گروه‌هایی خوشه بندی شدند. در گام آخر مجموعه داده‌ها در دو دسته خوش حساب و بد حساب دسته بندی شدند. این کار را با استفاده دو تکنیک درخت تصمیم و شبکه عصبی انجام داده‌اند و دسته بندی توسط هردو تکنیک به طور جداگانه انجام داده و نتایج آن دو را مقایسه نمودند [۱۴].

کوپایی و مینایی در کار تحقیقی خویش به بررسی روشی برای رتبه بندی تغییرات در الگوی رفتاری مشتریان فروشگاه‌های زنجیره‌ای پرداخته‌اند. آنان ابتدا اثرات تغییرات در معیارهایی مانند پیش‌بینی ارزش طول عمر مشتری و بازگشت سرمایه فعالیت‌های ارتباط با مشتری را محاسبه نموده و در ادامه تغییرات رفتار مشتری با استفاده از فرآیند تحلیل سلسله مراتبی فازی رتبه‌بندی کرده‌اند. در این مطالعه از انبار داده‌ی فروشگاه زنجیره‌ای مواد غذایی موندریان برای آزمایش روش استفاده شده

است. این انبار داده حاوی داده‌های مشتریان و فروش دو سال فروشگاه‌های زنجیره‌های موندریان است [۱۵].

شیب و همکاران در تحقیق به بررسی «کاربرد تکنیک‌های داده کاوی در شناخت رفتار مشتری در تجارت سیار» پرداختند. آنان در تحقیق خویش به ارائه الگوریتمی به منظور ارائه N محصولی که لازم است به کاربر پیشنهاد دهد، فازهای سه‌گانه را پیشنهاد نمودند. در فاز یک به دسته بندی محصولات پرداختند و در فاز دو به شکل ماتریس علاقه اقدام نمودند و در فاز سه به خوشه بندی کاربران مشابه پرداختند. آنان در فاز سه برای خوشه بندی کاربران مشابه از روش کامیانگین استفاده نمودند [۱۶].

ایرجی و مینایی نیز به «کاربرد داده کاوی زمانی جهت بهبود مدیریت فروش» پرداختند. آنان در تحقیق خویش یک مسئله مشکل و اساسی یعنی تئوری عمومی تحلیل داده کاوی زمانی را مورد بحث قرار داده و جواب‌هایی برای آن ارائه نمودند. همچنین مرور کلی از تکنیک‌های داده کاوی زمانی برای این نوع مسائل ارائه گردیده و نتایج مربوط به آنالیز آماری تکنیک‌ها را مورد بحث قرار دادند [۱۷].

مقیینی و همکاران در تحقیقی با عنوان «طبقه بندی ایستگاه‌های مسافری راه آهن جمهوری اسلامی ایران» با استفاده از تکنیک خوشه بندی پرداختند. آنان هدف از تحقیق خویش را بررسی ایستگاه‌های مسافری راه آهن جمهوری اسلامی ایران و دسته بندی آنان بر اساس اهمیت ایستگاه‌ها عنوان نمودند. در دسته بندی ایستگاه‌ها بر اساس اهمیت ایستگاه، شاخص‌هایی چون میانگین تعداد مسافر روزانه، مرکز استان بودن، ایستگاه مبدأ و مقصد بودن در نظر گرفته و با استفاده از الگوریتم

مورچه‌ها خوشه بندی انجام داده و نتایج حاصل را با الگوریتم کامیانگین مقایسه نمودند. در تحقیق فوق دقت بالای خوشه بندی انجام شده به روش کامیانگین مورد تاکید قرار گرفته است [۱۸].

احمدوند و همکاران با استفاده از تکنیک‌های داده کاوی به «استفاده از داده کاوی در مدیریت ارتباط با شهروند: مورد کاوی سامانه ۱۳۷ شهرداری تهران» پرداختند. آنان با استفاده از قواعد انجمنی و الگوریتم اپریوری^۱ اطلاعات مربوط به رضایتمندی شهروندان از خدمات شهری ارائه شده را مورد تجزیه و تحلیل قرار داده و ارتباط بین مشکلات، زمان بروز، واحد اجرایی مربوطه و میزان رضایت فعلی شهروندان از نحوه رفع آن مشکل به تفکیک مناطق مختلف شهری را، مورد بررسی قرار دادند. همچنین آنان ارتباط بین پارامترهای رضایتمندی شهروندان از کیفیت و سرعت خدمات ارائه شده مورد توجه قرار داده و عوامل و شرایط تأثیر گذار بر رضایتمندی و نارضایتی شهروندان مشخص نموده‌اند [۱۹].

غضنفری و همکاران در تحقیق خویش با عنوان «بخش بندی کشورها در بازرگانی خارجی مورد مطالعاتی: سازمان توسعه تجارت ایران» به استفاده از داده کاوی در حوزه‌ی تجارت پرداخته‌اند. آنان در تحقیق خویش به کمک ابزارهای داده کاوی شباهت رفتار صادرات کشور ایران به کشورهای دیگر را سنجیده و تابع سنجش شباهت رفتار کشورها بر اساس مفاهیم قواعد انجمنی تعریف نمودند و بر اساس تابع مذکور الگوریتم کامیانگین خوشه بندی صورت گرفت [۲۰].

۶-۲- تمایز پژوهش

درصد بالایی از تحقیقات انجام شده در رابطه با خوشه بندی، با استفاده از الگوریتم k - میانگین (علی‌الخصوص در داخل کشور) می‌باشد. اما با توجه به این موضوع که این الگوریتم توانایی یافتن تعداد خوشه‌ی بهینه را ندارد، بنابراین یا باید تعداد خوشه‌ها از قبل تعیین شود و یا با استفاده از الگوریتم‌های دیگر این تعداد را مشخص گردد که این موضوع نیز ممکن است منجر به بهینگی خوشه بندی نگردد. در این تحقیق سعی گردید که از الگوریتم شبکه‌ی عصبی کوهونن که جزو الگوریتم‌های خود سازمانده (SOM)^۱ بوده و قادر به یافتن تعداد خوشه‌های بهینه می‌باشد، استفاده گردد.