



وزارت علوم، تحقیقات و فناوری

دانشگاه علوم و فنون مازندران

پایان نامه

مقطع کارشناسی ارشد

رشته: مهندسی فناوری اطلاعات

عنوان: مدیریت دانش مشتریان بانک مهر اقتصاد با استفاده از تکنیک‌های داده کاوی

اساتید راهنما: دکتر جواد وحیدی، دکتر بابک شیرازی

استاد مشاور: دکتر حسین علیزاده

دانشجو: سیده عطیه علی کیا امیری

(تابستان ۱۳۹۲)

این پایان‌نامه را تقدیم می‌دارم به:

روح پاک پدرم که عالمانه به من آموخت که چگونه در عرصه زندگی، ایستادگی را تجربه نمایم.
و به مادرم، دریای بی‌کران فداکاری و عشق که وجودم برایش همه رنج بود و وجودش برایم
همه مهر.

و خواهرانم که حضورشان همیشه مایه دلگرمی و امید من است.

و سپاس ...

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت‌های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و درود بر محمد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است، و نفرین پیوسته بر دشمنان ایشان تا روز رستاخیز...

و سپاسگزارم از اساتید گرامی و بزرگوار

جناب آقای دکتر جواد وحیدی و جناب آقای دکتر بابک شیرازی که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ نمودند و زحمت راهنمایی این پایان‌نامه را بر عهده گرفتند.

و از استاد گرامی، جناب آقای دکتر حسین علیزاده که زحمت مشاوره این پایان‌نامه را عهده‌دار بودند که بدون مساعدت ایشان، این پروژه به نتیجه مطلوب نمی‌رسید.

و از استاد گرامی، جناب آقای پروفسور ایرج مهدوی که زحمت داوری این پایان‌نامه را متقبل شدند، کمال تشکر و قدردانی را دارم.

چکیده

افزایش روزافزون سطح رقابت در بازار، مدیران و تحلیل‌گران سازمان‌ها را وادار ساخته به دنبال راهکارهایی باشند که مزیت رقابتی را برای سازمان به ارمغان آورند. بر اساس مطالعات موجود، استفاده از دانش مشتری جهت اتخاذ استراتژی‌های لازم برای جلب رضایت مشتری می‌تواند سازمان‌ها را به سمت تحقق این هدف سوق دهد. از سوی دیگر گسترش بهره‌گیری از فناوری‌های بروز در زمینه اطلاعات و ارتباطات خصوصاً در بانک‌ها باعث بر جای ماندن حجم عظیمی از داده‌ها گشته که تحلیل و تصمیم‌گیری بر اساس آن‌ها با روش‌های معمول گزارش‌گیری و روش‌های آماری امکان‌پذیر نمی‌باشد. داده‌کاوی ابزار بروز و قدرتمندی است که در این پایان‌نامه جهت تحلیل داده‌ها به جهت استخراج دانش مشتری پیشنهاد می‌گردد.

هدف از این تحقیق با عنوان «مدیریت دانش مشتریان بانک مهر اقتصاد با استفاده از تکنیک‌های داده‌کاوی»، بخش‌بندی مشتریان بانک مهر اقتصاد، باهدف کشف ویژگی‌های رفتاری مشابه، برای کمک به مدیران این بانک جهت تسهیل اتخاذ استراتژی‌های متناسب با هر بخش و در نتیجه حفظ، تقویت و یا توسعه ارتباط با مشتریان و نهایتاً سودآوری برای این بانک می‌باشد. داده‌های خام مورد نیاز جهت این مطالعه از پایگاه داده‌های بانک مهر اقتصاد استخراج شده است.

کلمات کلیدی فارسی:

مدیریت دانش مشتری ، داده‌کاوی، خوشه‌بندی مشتریان ، کشف دانش.

فهرست مطالب

۱-۱- مقدمه	۲
۲-۱- تعریف مسئله	۳
۳-۱- ضرورت انجام تحقیق	۷
۴-۱- مراحل انجام تحقیق	۸
۵-۱- محدوده تحقیق	۹
۶-۱- اهداف تحقیق	۹
۷-۱- ساختار پایان نامه	۱۰
۱-۲- مقدمه	۱۳
۲-۲- مدیریت دانش	۱۴
۱-۲-۲- دانش چیست؟	۱۵
۲-۲-۲- هرم دانش	۱۵
۳-۲-۲- انواع دانش	۱۶
۱-۳-۲-۲- دانش صریح	۱۶
۲-۳-۲-۲- دانش ضمنی	۱۶
۴-۲-۲- مدیریت دانش چیست؟	۱۷
۵-۲-۲- استراتژی های مدیریت دانش	۱۸
۱-۵-۲-۲- استراتژی اجتماعی سازی (تبدیل دانش پنهان به پنهان)	۱۹
۲-۵-۲-۲- استراتژی برونی سازی (پنهان به آشکار)	۱۹

۲۰	۲-۲-۵-۳- استراتژی ترکیب سازی (آشکار به آشکار)
۲۰	۲-۲-۵-۴- استراتژی درونی سازی (آشکار به پنهان)
۲۰	۲-۲-۶- معایب عدم بهره‌گیری از دانش در سازمان
۲۱	۲-۲-۷- اهداف مدیریت دانش
۲۱	۲-۲-۸- مدل‌های مدیریت دانش
۲۳	۲-۳- مدیریت دانش مشتری
۲۴	۲-۳-۱- انواع دانش مشتری
۲۸	۲-۳-۲- مدل مدیریت دانش مشتری
۲۹	۲-۴- مدیریت ارتباط با مشتری
۳۲	۲-۴-۱- مدیریت ارتباط مشتریان در نظام بانکی
۳۳	۲-۴-۲- مدیریت ارتباط با مشتری: اهداف، مزایا و چالش‌ها
۳۴	۲-۵- مقایسه مفاهیم CKM و KM و CRM
۳۷	۲-۶- تاریخچه‌ای از بانک و بانکداری
۳۸	۲-۷- سیر تحول فناوری اطلاعات در صنعت بانکداری
۳۸	۲-۷-۱- دوره اول: اتوماسیون پشت باجه
۳۸	۲-۷-۲- دوره دوم: اتوماسیون جلوی باجه
۳۸	۲-۷-۳- دوره سوم: اتصال مشتریان به حساب‌هایشان
۳۹	۲-۷-۴- دوره چهارم: یکپارچه‌سازی سیستم‌ها و مرتبط کردن مشتریان با تمامی عملیات بانکی
۳۹	۲-۷-۵- بانکداری الکترونیک

۴۰	۸-۲- داده کاوی
۴۰	۸-۲-۱- مقایسه روش های آماری و داده کاوی
۴۲	۸-۲-۲- مفهوم داده کاوی
۴۴	۸-۲-۳- داده کاوی و کشف دانش
۴۵	۸-۲-۴- فرایند داده کاوی
۵۱	۸-۲-۵- معرفی روش های داده کاوی
۵۳	۸-۲-۵-۱- دسته بندی
۵۳	۸-۲-۵-۲- درخت تصمیم
۵۵	۸-۲-۵-۳- شبکه های عصبی
۵۶	۸-۲-۵-۴- پیش بینی
۵۶	۸-۲-۵-۵- خوشه بندی
۵۷	۸-۲-۵-۵- انواع خوشه بندی
۵۹	۸-۲-۵-۵-۲- معیارهای ارزیابی در خوشه بندی
۶۰	۸-۲-۵-۶- تحلیل انحراف
۶۱	۸-۲-۵-۷- قواعد وابستگی (انجمنی)
۶۱	۸-۲-۵-۸- تحلیل توالی
۶۲	۸-۲-۶- نرم افزار داده کاوی
۶۳	۸-۲-۷- کاربردهای داده کاوی
۶۳	۸-۲-۷-۱- داده کاوی در صنعت بانکداری

۶۵	۹-۲- پیشینه تحقیق
۶۶	۱-۹-۲- کاربرد داده‌کاوی در بخش‌بندی و مدل‌سازی رفتاری مشتریان در صنعت بانکداری
۶۸	۲-۹-۲- کاربرد داده‌کاوی در ارزیابی اعتبار مشتریان
۶۹	۳-۹-۲- کاربرد داده‌کاوی در زمینه کشف تقلب
۶۹	۴-۹-۲- کاربرد داده‌کاوی در تحلیل روی‌گردانی مشتری
۷۴	۱۰-۲- جمع‌بندی مطالب فصل
۷۷	۱-۳- مقدمه
۷۷	۲-۳- روش پیشنهادی
۷۷	۱-۲-۳- چارچوب تحقیق
۷۹	۲-۲-۳- انتخاب متغیرها
۸۰	۳-۲-۳- آماده‌سازی و پیش‌پردازش داده‌ها
۸۱	۱-۳-۲-۳- نرمال‌سازی داده‌ها
۸۱	۴-۲-۳- تعیین تعداد بهینه خوشه‌ها
۸۲	۵-۲-۳- خوشه‌بندی
۸۳	۱-۵-۲-۳- انواع خوشه‌بندی
۸۴	۲-۵-۲-۳- خوشه‌بندی به روش K-Means
۸۵	۱-۱-۵-۲-۳- مزایای استفاده از الگوریتم خوشه‌بندی K-Means
۸۵	۲-۱-۵-۲-۳- محدودیت‌های الگوریتم K-Means
۸۶	۲-۵-۲-۳- خوشه‌بندی به روش WK-Means

۸۷	A-H-Means خوشه‌بندی به روش
۸۸	 ۶-۲-۳- ارزیابی خوشه‌ها به روش مجموع مربع خطاها و انتخاب بهترین روش
۹۰	 ۷-۲-۳- به‌کارگیری دانش حاصل از خوشه‌بندی
۹۰	 ۳-۳- روش‌های جمع‌آوری اطلاعات
۹۰	 ۴-۳- جمع‌بندی مطالب فصل
۹۲	 فصل چهارم
۹۲	 ۱-۴- مقدمه
۹۳	 ۲-۴- معرفی بانک مهر اقتصاد
۹۴	 ۳-۴- موضوع و فعالیت مؤسسه
۹۴	 ۴-۴- محاسبات تحقیق
۹۵	 ۱-۴-۴- گام انتخاب و جمع‌آوری متغیرهای ورودی
۹۶	 ۲-۴-۴- گام آماده‌سازی و پیش‌پردازش داده‌ها
۹۷	 ۳-۴-۴- گام تعیین تعداد بهینه خوشه‌ها
۹۷	 ۴-۴-۴- گام خوشه‌بندی داده‌ها
۹۸	 K-Means خوشه‌بندی به روش
۱۰۰	 WK-Means خوشه‌بندی به روش
۱۰۰	 A-H-Means خوشه‌بندی به روش
۱۰۱	 ۵-۴-۴- ارزیابی خوشه‌ها به روش مجموع مربع خطاها و انتخاب بهترین روش
۱۰۲	 ۶-۴-۴- گام به‌کارگیری دانش حاصل از خوشه‌بندی

۱۰۴.....	۴-۵- نتایج تحقیق.....
۱۰۶.....	۴-۶- جمع‌بندی مطالب فصل.....
۱۰۷.....	فصل پنجم.....
۱۰۸.....	۵-۱- مقدمه.....
۱۰۸.....	۵-۲- خلاصه تحقیق.....
۱۰۹.....	۵-۳- نتیجه‌گیری.....
۱۱۰.....	۵-۴- زمینه‌های پیشنهادی، راهکارها و پیشنهادات جهت پژوهش‌های آتی.....

فهرست جداول

جدول ۱-۲ انواع مختلف تبدیلات دانش [۵۱ و ۵۵]	۱۹
جدول ۲-۲ مقایسه مفاهیم مدیریت دانش، مدیریت ارتباط با مشتری و مدیریت دانش مشتری [۳۸]	۳۵
جدول ۳-۲ مقایسه روش‌های تحلیل آماری و داده‌کاوی [۱۱ به نقل از ۱۴]	۴۱
جدول ۴-۲ فعالیت‌های مربوط به فازهای CRISP-DM و خروجی هر فعالیت [۵۰]	۵۰
جدول ۵-۲ نمونه داده‌های مورد نیاز در یک مسئله مدل‌سازی به روش دسته‌بندی [۱]	۵۴
جدول ۶-۲ معیارهای محاسبه شباهت در خوشه‌بندی	۵۹
جدول ۷-۲ معیارهای محاسبه فاصله در خوشه‌بندی	۶۰
جدول ۸-۲ پژوهش‌های انجام‌گرفته در زمینه کاربرد داده‌کاوی در صنعت بانکداری	۷۱
جدول ۱-۳ متغیرهای تحقیق	۸۰
جدول ۱-۴ نمونه ده‌تایی از داده‌های مربوط به مشتریان بانک مهر اقتصاد	۹۵
جدول ۲-۴ متغیرهای نرمال شده	۹۶
جدول ۳-۴ وزن نسبی متغیرهای تحقیق	۱۰۰
جدول ۵-۴ مقادیر مجموع مربع خطاها در الگوریتم‌های مختلف خوشه‌بندی	۱۰۱
جدول ۶-۴ دسته‌بندی مشتریان بر مبنای ویژگی‌های رفتاری مشابه [۱]	۱۰۳
جدول ۷-۴ اطلاعات مربوط به خوشه‌بندی مشتریان بانک مهر اقتصاد به روش K-Means	۱۰۴

فهرست تصاویر و نمودارها

شکل ۱-۲	سلسله مراتب دانش	۱۶
شکل ۲-۲	دانش صریح فقط بخش کوچکی از دانش را تشکیل می دهد [۵۱].	۱۷
شکل ۳-۲	مدل مدیریت دانش پروبست و رمهارد [۴].	۲۲
شکل ۴-۲	چارچوب خوشه انگور جهت نوع شناسی دانش مشتری	۲۶
شکل ۵-۲	مدل مدیریت دانش مشتری [۱۷ به نقل از ۳۷]	۲۸
شکل ۷-۲	گام های فرایند تولید دانش از پایگاه داده ها [۳۱, ۵۳]	۴۴
شکل ۸-۲	متدولوژی فرآیند استاندارد میان صنعتی داده کاوی (CRISP-DM) [۵۰]	۴۷
شکل ۹-۲	دسته بندی کلی عملکردهای داده کاوی	۵۲
شکل ۱۱-۲	نیروهای رقابتی پورتر [۱۱ به نقل از ۶]	۶۴
شکل ۱-۳	چارچوب تحقیق	۷۸
شکل ۱-۴	خوشه اول، الگوریتم K-Means	۹۸
شکل ۲-۴	خوشه دوم، الگوریتم K-Means	۹۸
شکل ۳-۴	خوشه سوم، الگوریتم K-Means	۹۹
شکل ۴-۴	خوشه چهارم، الگوریتم K-Means	۹۹
شکل ۵-۴	خوشه پنجم، الگوریتم K-Means	۹۹

فصل اول

مقدمه و کلیات تحقیق

در سال‌های اخیر دانش^۱ به عنوان منبعی ارزشمند در کنار منابعی چون کار، زمین، سرمایه قرار گرفت و به عنوان موتور تولیدکننده درآمد و یک دارایی مهم و راهبردی برای سازمان شناخته شد. از طرفی به دلیل وجود رقابت شدید در بین کسب و کارهای امروزی از جمله صنعت بانکداری، مشتری و توجه به جایگاه او و ارتباط با او اهمیت ویژه‌ای یافته است. بنابراین مبحث بسیار مهم مدیریت دانش مشتری^۲ مطرح گردید که امروزه مطالعات بسیار زیادی را به خود اختصاص داده است. مدیریت دانش مشتری با استفاده از راهکارهای مختلف مدیریت دانش نظیر روش‌های داده‌کاوی^۳ زمینه بسیار خوبی را جهت استفاده مفید از گنجینه گران‌بهای دانش مشتری فراهم می‌آورد.

از سوی دیگر در عصر حاضر بهره‌گیری از فناوری‌های نوین اطلاعات و ارتباطات در عرصه‌های مختلف کسب‌وکار به امری گریزناپذیر مبدل گشته است. به طور خاص صنعت بانکداری از جمله صناعی است که به‌کارگیری فناوری‌های روز دنیا در این صنعت می‌تواند مزیت رقابتی انکارناپذیری را برای آن ایجاد نماید. بنابراین این صنعت نیز از بهره‌گیری از بروزترین فناوری‌ها مستثنا نبوده و مواردی مانند بانکداری الکترونیک، سی‌ستم‌های یکپارچه بانکداری^۴، دستگاه‌های خودپرداز، کارت‌های اعتباری، پایانه‌های خرید الکترونیک و... از مصادیق این امر می‌باشد.

ورود فناوری‌های جدید به سازمان سبب افزایش چشمگیر سرعت تولید اطلاعات و در نتیجه بر جای ماندن حجم عظیمی از داده‌ها گشته است. از آنجا که این داده‌ها اغلب حجیم و وسیع می‌باشند، معمولاً به صورت خام قابل استفاده نیستند، بلکه دانش موجود در آن‌ها باید استخراج گردد. با این وجود که ارزش این داده‌ها بر کسی پوشیده نیست، حجم بسیار بالای داده‌های ارزشمند موجود، تحلیل و بهره‌گیری از آن‌ها را به امری چالش‌برانگیز مبدل ساخته است. چرا که تحلیل به واسطه روش‌های گزارش‌گیری سنتی در این مقیاس امکان‌پذیر نیست و روش‌های آماری موجود نیز از ظرفیت کافی جهت تحلیل این داده‌ها برخوردار نمی‌باشند. بنابراین باید به دنبال راهکاری بود که با غلبه بر این محدودیت، امکان انجام مطالعاتی جامع‌تر با نتایج دقیق‌تر و درصد خطای پایین‌تر را

^۱ Knowledge

^۲ Customer Knowledge Management (CKM)

^۳ Data Mining

^۴ Core Banking

ممکن سازد. این مسئله گواهی بر لزوم به‌کارگیری روش‌های نوین تحلیل داده‌ها جهت حصول دانش، نظیر روش داده‌کاوی خواهد بود.

دانش قابل توجهی که در زمان استفاده از خدمت یا مصرف کالا توسط مشتری، بین مشتری و سازمان تبادل می‌شود، به عنوان منبعی مهم برای سازمان شناخته می‌شود و کسب و بهره‌برداری از آن به یک مزیت رقابتی در سازمان‌ها تبدیل شده است.

مدیریت دانش مشتری دربردارنده فرایندهایی است که با شناسایی و اکتساب اطلاعات مشتری و نیز ایجاد و بهره‌برداری از دانش مشتریان، مربوط است [۹]. چنین اطلاعاتی در ماورای محدوده‌های خارجی سازمان قرار دارند و دانشی که از آن‌ها استخراج می‌شود موجب ایجاد ارزش برای سازمان و مشتریان آن خواهد شد [۳۲]. در این تحقیق مسئولیت کشف دانش بر عهده الگوریتم‌های داده‌کاوی خواهد بود. در ادامه از این دانش به عنوان راهنما در مسیر اتخاذ استراتژی‌های سازمان، بهره‌گیری می‌شود.

۱-۲- تعریف مسئله

بااهمیت یافتن مشتری در عرصه پرتلاطم رقابت میان کسب‌وکارهای مباحثی همچون مدیریت دانش مشتری و مدیریت ارتباط با مشتری مطالعات و تحقیقات فراوانی را به خود اختصاص داده‌اند.

مدیریت دانش، کسب دانش درست، برای افراد مناسب، در زمان صحیح و مکان مناسب است، به گونه‌ای که آنان بتوانند برای دستیابی به هدف‌های سازمان، بهترین استفاده را از دانش ببرند.

در تعریفی دیگر مدیریت دانش فرایند کشف، کسب، توسعه و ایجاد، تسهیم، نگهداری، ارزیابی و به‌کارگیری دانش مناسب در زمان مناسب توسط فرد مناسب در سازمان، که از طریق ایجاد پیوند مناسب بین منابع انسانی، فناوری اطلاعات و ایجاد ساختاری مناسب برای دستیابی به اهداف سازمانی صورت می‌پذیرد، تعریف شده است.

مدیریت ارتباط با مشتری^۱ از جمله راهکارهایی است که در سال‌های اخیر با افزایش روزافزون سطح رقابت در بازار به عنوان سلاحی ارزشمند در جهت افزایش وفاداری مشتری و جلب رضایت او و با هدف به ارمغان آوردن مزیت رقابتی بالاتر برای سازمان، از سوی سازمان‌ها بکار گرفته شده

^۱ Customer Relationship Management

است. امروزه بیشتر روش‌های مدیریت ارتباط با مشتری مبتنی بر فناوری اطلاعات می‌باشند و مسلماً برای رسیدن به مدیریت مؤثر ارتباط با مشتری ناگزیر از مدیریت دانش مشتری خواهیم بود.

مدیریت ارتباط با مشتری در برگیرنده مجموعه‌ای از فرایندهاست که سازمان‌ها را قادر می‌سازد تا از استراتژی‌های کسب‌وکار در جهت ایجاد روابط بلندمدت و سودآور با مشتریان خاص پشتیبانی نمایند [۴۶].

در حقیقت CRM یک فناوری پیشرفته در جهت دستیابی به قله‌های اطلاعات مشتری است [G] و شرکت‌ها از آن به عنوان ابزاری در جهت افزایش رضایتمندی مشتری استفاده می‌کنند. مدیریت ارتباط با مشتری به عنوان فعالیتی جهت گسترش و نگاه‌داری مشتریان سازمان‌ها به طور گسترده‌ای مورد توجه قرار گرفته است و ابزارهای آن افزایش رضایت مشتری و وفاداری اوست. همچنین مدیریت دانش KM همچون مدیریت روابط با مشتری بر جمع آوری منابعی تأکید دارد که از فعالیت‌های تجاری در جهت رسیدن به توانایی رقابت‌پذیری حمایت می‌کند [۳۷] برای بهبود روابط با مشتری، خدمات‌رسانی به روشی که مورد دلخواه اوست، ضروری است. از این رو به مدیریت دانش مشتری احتیاج است [۱۷].

امروزه حجم بالای پایگاه‌های داده و پراکندگی و عدم به‌کارگیری راهکارهای مناسب جهت تحلیل این داده‌ها مطالعه و تصمیم‌گیری بهینه پیرامون ارتباط با مشتریان را با مشکل مواجه نموده است.

به طور ویژه بانک‌ها سازمان‌هایی هستند که با مشتریان تعامل مستقیم دارند و عنصر مشتری در این سازمان‌ها اهمیت ویژه‌ای دارد. بدیهی است جهت پیشرو بودن در عرصه رقابت توجه به جایگاه مشتری و اختصاص خدمات ویژه با تشخیص نیازمندی‌های آن‌ها و ارائه خدمات درست به آن‌ها محقق خواهد شد. تحلیل منابع باارزش داده‌ای در رابطه با مشتریان کنونی بانک‌ها و مشتریان بالقوه و... مقدمه‌ای جهت تحقق این مهم خواهد بود.

با ورود فناوری‌های جدید به سازمان‌ها و افزایش سرعت تولید اطلاعات و در نتیجه بر جای ماندن حجم عظیمی از داده‌ها و دشواری استفاده از این حجم وسیع، معمولاً امکان استفاده از این داده‌ها به صورت خام وجود ندارد، بلکه دانش موجود در آن‌ها باید استخراج گردد. همچنین تحلیل این داده‌ها به واسطه روش‌های گزارش‌گیری سنتی در این مقیاس امکان‌پذیر نیست و روش‌های

آماري موجود نيز از ظرفيت كافي جهت تحليل اين داده‌ها برخوردار نمي‌باشند. داده‌كاوي راهكاري است كه با غلبه بر اين محدوديت، امكان انجام مطالعاتي جامع‌تر با نتايج دقيق‌تر و درصد خطاي پايين‌تر را ممكن سازد.

داده‌ها^۱ كميت‌هاي عددي يا خصيصة‌اي هستند كه در نتيجه مشاهده يا آزمايش حاصل شده‌اند. داده‌هاي دسته‌بندي شده اطلاعات^۲ را تشكيل مي‌دهند. اطلاعات از ترتيب، تركيب و شبكه شدن دانش را ايجاد مي‌نمايد. دانش، اطلاعات سازمان‌يافته، تحليل يافته و يا تلخيص شده براي افزايش درك، آگاهي يا تشخيص مي‌باشد.

داده‌كاوي ابزاري بروز، قدرتمند و وسيع است كه مي‌تواند جهت تحليل حجم عظيم داده بكار گرفته شود. داده‌كاوي به بهره‌گيري از ابزارهاي تجزيه و تحليل داده‌ها به منظور كشف الگوها و روابط معتبري كه تا كنون ناشناخته بوده‌اند اطلاق مي‌شود. اين ابزار با كاوش در بين داده‌هاي موجود و استخراج الگوها و روابط موجود در پايگاه داده‌ها، موجب تسهيل مطالعات و اتخاذ تصميمات خواهد شد. همچنين به كمك امكانات موجود در اين ابزار مي‌توان حجم داده‌ها را كاهش داد و داده‌هاي اضافي را حذف نمود. استفاده از روش‌هاي مختلف داده‌كاوي مي‌تواند جهت كشف دانش و الگوهاي موجود در حجم عظيم داده‌ها (براي مثال به طور خاص داده‌هاي جمعيت شناختي مشتريان بانك نظير جنسيت، سن، وضعيت تاهل، تحصيلات، شغل و غيره، يا داده‌هاي مربوط به تراكنش‌هاي مالي مشتريان و يا سرويس‌هاي ارائه شده توسط بانك) استفاده شود. اين الگوها مي‌توانند از سوي مديران جهت اتخاذ تصميمات مقتضي در ارتباط با مشتري، چگونگي برخورد با شرايط متفاوت كسب‌وكار، ارائه خدمات خاص و غيره بكار گرفته شود.

تعاريف متفاوتي از داده‌كاوي وجود دارد ولي تعريفی كه در اكثر مراجع به اشتراك ذكر شده عبارت است از «استخراج اطلاعات و دانش و كشف الگوهاي پنهان از پايگاه داده‌هاي بسيار بزرگ و پيچيده». داده‌كاوي يك متدولوژي بسيار قوي و با پتانسيل بالا مي‌باشد كه به سازمان‌ها كمك مي‌كند كه بر مهم‌ترين اطلاعات از مخزن داده‌هاي خود تمرکز نمايند [۵۲].

^۱ Data

^۲ Information

داده‌کاوی را می‌توان به عنوان ابزاری جهت کشف دانش از پایگاه‌های داده تعریف نمود. داده‌کاوی یک گام در فرایند کشف دانش از پایگاه داده^۱ است که با استفاده از الگوریتم‌های کشف و تحلیل داده‌ها تعداد خاصی الگو یا مدل را از روی داده‌ها تولید می‌کند.

یک الگو زمانی می‌تواند به عنوان دانش تلقی شود که :

۱. به سادگی برای افراد قابل فهم باشد.

۲. اعتبار آن از یک حد آستانه‌ای^۲ پایین‌تر نباشد.

۳. دانش جدیدی متناسب با اهداف تعیین شده سازمان ارائه دهد.

۴. کاربردی باشد. [۱]

امروزه کاربردهای مختلفی برای داده‌کاوی در صنعت بانکداری شناخته شده است.

شهرابی و همکاران در [۱] کاربردهایی نظیر مدیریت ارتباط با مشتریان از طریق بخش‌بندی آن‌ها و استخراج قواعد رفتاری آن‌ها جهت هدایت فعالیت‌های بازاریابی بانک، مدل‌سازی پاسخ بازار به تصمیمات کلان بانک و شناسایی اثر هر یک از سیاست‌های بانک بر تقاضا و امکان شبیه‌سازی تعاملی سیاست‌ها و تصمیمات قبل از اجرا، ارزیابی ریسک اعتباری مشتریان متقاضی دریافت وام، محاسبه کارایی شعب و پیش‌بینی سری‌های زمانی مالی و ... کشف الگوهای پنهان در داده‌های حاصل از تعاملات با مشتریان و تحلیل این الگوها را به عنوان مثال‌هایی از کاربرد ابزار داده‌کاوی در بانک ذکر کرده است.

بانک‌ها می‌توانند جهت افزایش سودآوری از طریق مشتریان تراز اول خود، برای یک محصول جدید بانکی، مشتریان بالقوه خود را از طریق داده‌کاوی شناسایی نموده و آن‌ها را به سمت استفاده از این محصول سوق دهند و بدین ترتیب در زمان و هزینه صرفه‌جویی کنند.

پیاده‌سازی سیستم هوشمند ارزیابی اعتبار مشتری به کمک داده‌کاوی از کاربردهای دیگر این ابزار جهت مدیریت دانش مشتریان در بانک است. در این کاربرد می‌توان با استفاده از اطلاعات گروهی از وام‌گیرندگان سابق (مانند بعضی اطلاعات شخصی وام‌گیرنده مانند سن و جنسیت و

^۱ Knowledge Discovery in Databases (KDD)

^۲ Threshold

وضعیت تأهل و درآمد و...، تعداد تراکش ماهیانه مشتری با بانک، سایر بدهی‌های بانکی، نوع ضمانت و...) و وضعیت آن‌ها از نظر بازپرداخت وام به مدلی دست یافت که مشتریان متقاضی وام را در دسته‌هایی مانند مشتری خوش حساب، مشتری بدحساب و ... دسته‌بندی نمود. با حصول این دانش در مورد مشتری متقاضی اعتبار تصمیم‌گیری در مورد پرداخت یا عدم پرداخت اعتبار بسیار ساده‌تر خواهد بود.

برخی از مشکلات نبود مدیریت دانش می‌توان عدم اولویت‌بندی و استفاده از انواع دانش، ایجاد شکاف دانشی در صورت جدا شدن کارشناسان از سازمان، احتکار دانش به جای ازدیاد آن، استفاده محدود از دانش‌های موجود، نبود مستندات در خصوص تجربیات به دست آمده از پروژه‌ها و کارها و نبود ساختار مناسب برای تسهیم سریع دانش را نام برد.

۱-۳- ضرورت انجام تحقیق

افزایش روزافزون سطح رقابت در بازار، مدیران و تحلیل‌گران سازمان‌ها را وادار ساخته به دنبال راهکارهایی باشند که مزیت رقابتی را برای سازمان به ارمغان آورند. بر اساس مطالعات موجود، استفاده از دانش مشتری جهت اتخاذ استراتژی‌های لازم برای جلب رضایت مشتری می‌تواند سازمان‌ها را به سمت تحقق این هدف سوق دهد. از سوی دیگر گسترش بهره‌گیری از فناوری‌های بروز در زمینه اطلاعات و ارتباطات خصوصاً در بانک‌ها باعث بر جای ماندن حجم عظیمی از داده‌ها گشته که تحلیل و تصمیم‌گیری بر اساس آن‌ها با روش‌های معمول گزارش‌گیری و روش‌های آماری امکان‌پذیر نمی‌باشد. داده‌کاوی ابزار بروز و قدرتمندی است که در این پایان‌نامه جهت تحلیل داده‌ها به جهت استخراج دانش مشتری پیشنهاد می‌گردد.

امروزه تحقیقات فراوانی در زمینه داده‌کاوی و به‌کارگیری آن در صنعت بانکداری دنیا و کشورمان در حال انجام است. با این وجود زمینه‌های بالقوه بسیاری در به‌کارگیری این دانش در بانک‌ها و مؤسسات مالی و اعتباری در کشور وجود دارد. بدیهی است پیش‌دستی در بهره‌گیری از این ظرفیت‌ها می‌تواند مزیت رقابتی را برای بانک یا مؤسسه مالی و اعتباری مربوطه به ارمغان آورده و ارزش افزوده بیشتری را نصیب بانک یا مؤسسه مورد نظر نماید. علاوه بر این ایجاد زمینه‌هایی جهت آشنایی کارشناسان و متخصصان امور بانکی با مباحث داده‌کاوی و کاربردهای آن در صنعت بانکداری ضروری به نظر می‌رسد.

بدین ترتیب مدیران بانک‌ها می‌توانند با یافتن الگوهای حاکم بر روند تغییرات بازار، استراتژی‌های لازم را در زمان نیاز اتخاذ نمایند و یا با یافتن خصوصیات مشتریان و دسته‌بندی آن‌ها الگوی مناسبی جهت جذب مشتری و جلب رضایت آن‌ها کشف نموده و در نتیجه سودآوری را برای سازمان به ارمغان آورند.

۱-۴- مراحل انجام تحقیق

برای استخراج دانش مشتری توسط داده‌کاوی در سازمان هدف، الگوی کشف دانش از پایگاه داده [۵۳] انتخاب شده است. گام‌های این الگو به صورت زیر است:

ابتدا باید شناسایی دامنه فعالیت‌های کسب‌وکار صورت گرفته و دسترسی به داده‌ها مقدور گردد. داده‌ها در اینجا می‌تواند مجموعه‌ای از حقایق، الگوها و یا خبرگی باشد.

در مرحله بعد با توجه به اهداف تعیین‌شده مورد نظر ذینفعان سازمان (مدیران، کارمندان، سرپرستان شعب، کاربران، خبرگان و مشاورین فناوری اطلاعات)، رکوردهای مورد نیاز توسط داده‌کاو از میان حجم عظیم داده‌های سازمان مورد نظر انتخاب می‌شوند. در این مرحله حجم انبوه داده‌ها به داده‌های هدف کاهش داده می‌شود.

برای آماده‌سازی بخش داده‌های انتخابی باید عملیات پیش‌پردازشی روی آن‌ها صورت گیرد. این عملیات شامل جمع‌آوری اطلاعات لازم برای مقداردهی به مقادیر رکوردهای از قلم افتاده، تشخیص داده‌های متفاوت و هماهنگی آن‌ها با بقیه، جمع‌آوری اطلاعات لازم برای اصلاح یا حذف داده‌های نامناسب^۱ می‌باشد. محصول این مرحله داده‌های پیش‌پردازش شده خواهد بود.

گام بعد گزینش داده‌ها با هدف کاهش حجم آن‌ها و یافتن ویژگی‌هایی از داده‌ها با توجه به هدف تعیین شده برای کسب‌وکار خواهد بود. با کاهش حجم داده‌ها یا روش‌های تغییرات آن‌ها حجم داده‌ها می‌تواند به مقدار مفید و مؤثری کاهش یابد.

مرحله بعد اعمال یک یا چند الگوریتم (مانند کلاسیفایرها و یا شبکه‌های عصبی و ...) از بین الگوریتم‌های موجود و یا ابتکاری داده‌کاوی بر روی داده‌هاست. حاصل این مرحله یک یا چند مدل و یا الگو خواهد بود.

^۱ noise

در گام بعد با تحلیل، ارزیابی و تفسیر الگوهای معنادار کشف شده، دانش استخراج می‌گردد. دانش استخراج شده می‌تواند به صورت قواعدی مستند شود.

۱-۵- محدوده تحقیق

در این تحقیق، تجزیه و تحلیل بر روی داده‌های خام پایگاه داده‌های شعب بانک مهر اقتصاد استان مازندران صورت خواهد پذیرفت. جامعه آماری مورد مطالعه کل داده‌های ذخیره شده در پایگاه داده‌های این شعب می‌باشد.

مبنای این تحقیق، مطالعات کتابخانه‌ای گسترده‌ای شامل بررسی پایان‌نامه‌ها و تحقیقات موجود و مرتبط با موضوع، مطالعه مقالات و کتب لاتین و فارسی مرتبط و همچنین استفاده از اینترنت می‌باشد.

۱-۶- اهداف تحقیق

دانش در سازمان به طور مداوم در حال گردش است. این دانش ایجاد، جذب، تحلیل و توزیع می‌شود و همواره در تمام فعالیت‌های سازمان و خصوصاً فعالیت‌های مرتبط با مشتری مورد استفاده قرار می‌گیرد. مدیریت دانش مشتری در سازمان به منظور تبدیل دانش تئوری مشتری به دانش قابل بهره‌برداری جهت پشتیبانی اتخاذ تصمیمات آتی در سازمان بکار گرفته می‌شود.

در بسیاری از منابع هدف از مدیریت دانش، تبدیل دانش ضمنی به دانش صریح و انتشار مؤثر آن در سازمان بیان شده است.

مدیریت دانش عملی کلی به منظور مدیریت فرآیندهای خلق، ذخیره و نگهداری و به اشتراک گذاردن دانش است؛ که به طور عمومی باید شامل شناسایی وضعیت موجود، تشخیص، وضوح نیازها و خواسته‌ها و بهبود فرآیندهای مورد اثر باشد و به تبع آن پروژه‌های مدیریت دانش پروژه‌های بهبود بخشی هستند.

مدیریت دانش به خودی خود امری فناوری محور نیست اما به‌کارگیری فناوری در تحقق کاراتر این چرخه ضروری به نظر می‌رسد، بنابراین در این پایان‌نامه داده‌کاوی به عنوان ابزار تحلیل داده‌ها و کشف دانش بکار گرفته خواهد شد.

در این پایان‌نامه هدف بر این است که طبق الگوی تشریح شده با به‌کارگیری نرم‌افزارهای موجود در زمینه داده‌کاوی و به کمک ابزارهایی از آن نظیر دسته‌بندی و خوشه‌بندی و... به تحلیل

داده‌های حاصل از تعامل سازمان (بانک) با مشتریان (برای مثال داده‌های جمعیت شناختی مشتریان بانک نظیر جنسیت، سن، وضعیت تأهل، تحصیلات، شغل و غیره، یا داده‌های مربوط به تراکنش‌های مالی مشتریان و یا سرویس‌های ارائه شده توسط بانک) به دسته‌بندی مشتریان با حداکثر اطمینان و کمترین خطای ممکن پرداخته شود. از آن جا که مشتریان بانک خصوصیات و نیازهای متفاوتی دارند، با دسته‌بندی مشتریان و تحلیل دسته‌ها توسط داده‌کاو و نظرات افراد خبره در سازمان، دانش حاصل استخراج می‌شود. مشتریان هر دسته به طور معمول خصوصیات مشابهی دارند و می‌توان با اتخاذ استراتژی مناسب هر دسته به نیازهای آنان رسیدگی نمود و در حصول هر چه بیشتر رضایت مشتری پیشرو بود. شناسایی مشخصه‌های هر گروه می‌تواند در موارد زیر بکار گرفته شوند:

- عرضه‌ی محصول طراحی شده متناسب با نیازهای هر گروه
- عرضه محرک‌های مناسب برای توسعه ارتباط مشتری با بانک
- تولید محصول جدید متناسب با ویژگی هر گروه

برای مثال با تعیین مشتریان وفادار بانک که معمولاً سودآوری زیادی برای بانک دارند می‌توان به آنان خدمات تشویقی ارائه نمود و یا خدمات تازه‌تری به آنان پیشنهاد داد و در حفظ این دسته از مشتریان کوشید. از طرفی به کمک مدل‌های حاصل از داده‌کاوی می‌توان اطلاعات مربوط به هر مشتری دلخواه را به مدل داده و تعیین نمود که مشتری در کدام دسته قرار می‌گیرد تا خدمات متناسب‌تری به آنان ارائه گردد.

۱-۷- ساختار پایان‌نامه

ساختار این پایان‌نامه به صورت زیر است:

در فصل دوم، مفاهیم پایه‌ای مورد استفاده در این تحقیق شرح داده خواهند شد. این مفاهیم شامل موضوعاتی چون مدیریت دانش، مدیریت دانش مشتری، مدیریت ارتباط با مشتری و داده‌کاوی و خلاصه‌ای از روش‌های موجود در زمینه داده‌کاوی می‌باشد، در ادامه پیشینه کارهای مرتبط انجام شده در این زمینه معرفی خواهند شد. سپس در فصل سوم روش پیشنهادی و مؤلفه‌های آن به همراه مدل‌های مورد استفاده بیان خواهند شد. در فصل چهارم جزئیات مربوط به گام‌های اصلی بیان شده و فرایندهای داخلی آن‌ها معرفی و روش پیشنهادی بر روی داده‌های واقعی مربوط به مشتریان بانک

اقتصاد مهر شعب استان مازندران اعمال می‌گردد. در فصل پنجم نتیجه‌های ممکن استخراج شده و راهکارها و پیشنهاداتی برای تحقیقات آتی، ارائه خواهند شد.

فصل دوم

ادبیات و پیشینه تحقیق

افزایش روزافزون سطح رقابت در میان کسب‌وکارهای امروزی سبب گشته سازمان‌ها جهت ربودن گوی سبقت از رقیبان، به دنبال رویکردهای نوینی باشند. بنابراین مطالعات و تحقیقات گسترده‌ای جهت کسب مزیت رقابتی توسط سازمان‌ها صورت گرفت. در این میان عامل مشتری از اهمیت خاصی برخوردار است و مطالعات ویژه‌ای به این مبحث مهم اختصاص یافته است.

صنعت بانکداری از جمله صناعی است که در آن مشتریان، به دلیل ارتباطات مستقیم با سازمان اهمیت ویژه‌ای دارند. از طرفی رواج کسب‌وکار الکترونیک در سازمان‌ها و بانک‌ها و مؤسسات مالی و اعتباری سبب برجای ماندن داده‌های غنی حاصل از تعاملات و ارتباطات مشتری با سازمان شده که این داده‌ها در صورت بهره‌برداری درست و استخراج دانش و الگو از میان آن‌ها به گنجینه‌ای غیر قابل توصیف بدل خواهند شد.

از طرفی، به طور خاص در صنعت بانکداری، به نظر می‌رسد مسائلی چون مشکلات موجود در نظام اقتصادی کشور، افزایش میزان عرضه بر تقاضا در صنعت بانکداری، دولتی بودن بسیاری از بانک‌های کشور و عرضه خدمات بانکی مشابه و... توجه به مشتری و جایگاه او در نظام بانکی را با بی‌مهری مواجه نموده است. وجود این مسائل خود می‌تواند فرصتی باشد تا بانک با بهره‌گیری از ابزارهای مدیریت دانش مشتری، نیازها و توقعات مشتریان را زودتر از سایر رقبای پیش‌بینی نموده و گوی سبقت را از رقیبان برآید [۱۱].

مدیریت دانش مشتری ابزاری است که سعی دارد با تمرکز بر چگونگی پیشرفت و توسعه سازمان به کمک دانش کسب شده از روی داده‌های حاصل از روابط با مشتریان و کسب مشتریان جدید و تأمین رضایت مشتریان ارزش افزوده را برای سازمان به ارمغان آورد.

در این تحقیق ابزار داده‌کاوی جهت کشف دانش پنهان در میان حجم عظیم داده‌های مربوط به مشتریان انتخاب شده است. این ابزار با وجود روش‌های متنوع موجود در آن و امکان تحلیل انبوهی از داده‌ها، محدودیت‌های روش‌های سنتی گزارش‌گیری و روش‌های آماری را برطرف نموده است. امروزه کاربردهای جالب و وسیع این ابزار در اکثر علوم و به طور خاص صنعت بانکداری مورد توجه محققان و متخصصین امر قرار گرفته است.

این فصل از بخش‌های مدیریت دانش، مدیریت ارتباط با مشتری، مدیریت دانش مشتری، تاریخچه بانکداری و داده‌کاوی، مفاهیم و کاربردهای آن تشکیل شده است.

در بخش مدیریت دانش به تفسیر هرم دانش، ارائه تعاریفی از مدیریت دانش و معایب عدم استفاده از مدیریت دانش در سازمان خواهیم پرداخت. در ادامه یک مدل مدیریت دانش را تفسیر نموده و استراتژی‌های موجود در این زمینه مطرح خواهند شد.

در بخش مدیریت ارتباط با مشتری ابتدا تعاریف و فلسفه CRM بیان شده و سپس به طور خاص به مدیریت ارتباط با مشتریان در نظام بانکی پرداخته می‌شود.

بخش مدیریت دانش مشتری تعاریف این مفهوم و مقایسه آن با مدیریت دانش و مدیریت ارتباط با مشتری بیان می‌شود.

در بخش تاریخچه بانکداری تاریخچه بانک و بانکداری و سیر تحول خدمات بانکی مطرح خواهد شد.

در بخش داده‌کاوی به تعاریف و مفاهیم و تفسیر مختصری از الگوریتم‌های موجود در زمینه علم داده‌کاوی خواهیم پرداخت. در ادامه به نمونه‌هایی از کاربردهای داده‌کاوی به خصوص در صنعت بانکداری اشاره شده و در نهایت پیشنهادی تحقیق بررسی خواهد شد.

۲-۲- مدیریت دانش^۱

دانش در قرن اخیر رشد بسیار چشمگیری داشته است به طوری که گفته می‌شود ۸۰٪ یافته‌های فناوری و دانش ۹۰٪ کل دانش‌ها و اطلاعات فنی جهان در قرن اخیر تولید شده است.

سرعت پیشرفت دانش بشر به گونه‌ای بوده که دانشمندان عصر حاضر را عصر دانش نامیده‌اند. تفکر و تحقیق خود به عنوان شغل شناخته شد و اطلاعات به عنوان سرمایه مطرح شد. محصول کار بشر به عنوان متفکر و یا محقق بر روی مواد اولیه اطلاعات، دانش نام گرفت.

^۱ Knowledge Management (KM)

هرچند شاید دانش از نظر گستره و کاربرد آن در همه سازمان‌ها بحثی جدید نباشد، اما دانش به عنوان دارایی معنوی سازمان و منبعی جهت کسب مزیت رقابتی^۱ امری است که امروزه شدیداً مورد توجه قرار گرفته است.

۲-۲-۱- دانش چیست؟

دانش، مخلوطی از تجارب، کمیت‌ها و اطلاعات منظم است که قالبی برای ارزیابی، یکپارچگی و هماهنگی تجربه‌ها و اطلاعات ایجاد می‌کند. دانش، درک و فهمی است که از طریق تجربه، استدلال، درک مستقیم و یادگیری حاصل می‌شود [۱۶].

امروزه دانش به یکی از منابع مهم استراتژیک سازمان بدل شده است. توانایی خلق دانش و تداوم یادگیری از دانش حاصل می‌تواند سود و رقابت‌پذیری را برای سازمان به ارمغان آورد زیرا دانش خلاقانه امروز، هسته اصلی دانش فردا خواهد بود [۹ و ۴۴].

دانش، اطلاعات آمیخته شده با تجربه، زمینه کاری و تفسیر می‌باشد [۴۸]. دانش موتور تولیدکننده درآمد و یک دارایی مهم و استراتژیک برای سازمان است.

تمایز بین داده، اطلاعات و دانش ساده نیست اما معانی آن از دیدگاه کاربرد می‌تواند به تفاوت بین آن‌ها منجر شود. در کل، داده به حقایق خام^۲، اطلاعات به مجموعه یا ترکیب سازمان دهی شده‌ای از داده‌ها و دانش به اطلاعات معنی‌دار اطلاق می‌شود.

داده و اطلاعات بر پایه سازماندهی‌شان مشخص می‌شوند و اختلاف اطلاعات و دانش بر پایه تفسیر آن‌هاست [۴۵].

۲-۲-۲- هرم دانش

هرم دانش، نمایانگر سلسله‌مراتبی از جریان داده تا خرد می‌باشد. هرچند بحث‌های متفاوتی در مورد جزئیات سلسله‌مراتب دانش موجود است، اما در کل جریان کلی از سوی داده به سمت خرد وجود دارد. دانش را می‌توان با عمل و تصمیماتی که به سبب آن اتخاذ می‌شود ارزیابی نمود [۱۶].

^۱ Competitive Advantage

^۲ Knowledge

^۳ Raw Facts



شکل ۲-۱ سلسله مراتب دانش

داده^۱ها: ارائه، ثبت، ذخیره و نگهداری نهادهایی را بر عهده دارند.

اطلاعات^۲: دانستن آن و برآیندی است از پردازش داده‌ها مثل سازماندهی، مرتب کردن و غیره.

دانش: به صورت دانستن چگونگی و برآمدی است که از اعمال پردازش اطلاعات تعریف می‌شود.

خرد^۳ (معرفت): تقریباً دانستن «زمانی که» و یا «اگر» است. دانش از طریق فعالیت‌هایی مثل کشف، استنتاج، ارزش، تجربه و غیره خرد را معنا می‌بخشد.

۲-۲-۳- انواع دانش

۲-۲-۳-۱- دانش صریح^۴

دانشی که قابل رمزگذاری و کدگذاری بوده و در نتیجه می‌توان آن را به سادگی مخابره، پردازش و منتقل و در پایگاه داده‌ها ذخیره کرد. این نوع از دانش را می‌توان فرم داد و یک فرمول علمی و یا کتابچه راهنما بین افراد سازمان منتشر کرد. دستورالعمل‌ها، مقررات، قوانین، رویه‌های انجام کار، آیین‌نامه‌ها، شرح جزئیات و... که به صورت رسمی در بین افراد سازمان به آسانی قابل انتقال هستند همه دانش تصریحی به حساب می‌آیند [۵۱].

۲-۲-۳-۲- دانش ضمنی^۵

^۱ Data

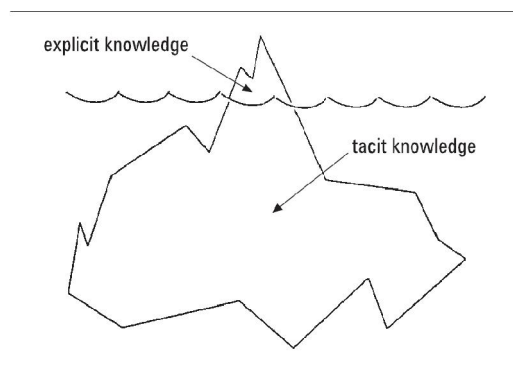
^۲ Information

^۳ Wisdom

^۴ Explicit Knowledge

^۵ Implicit Knowledge

دانش ضمنی شخصی بوده و فرموله کردن آن بسیار مشکل است. این نوع از دانش که از طریق تسهیم تجربیات با مشاهده و تقلید اکتساب می‌شود، ریشه در اعمال، رویه‌ها، تعهدات، ارزش‌ها و احساسات افراد داشته، قابل کدگذاری نبوده، از طریق یک زبان مخابره نمی‌شود [۵۱].



شکل ۲-۲ دانش صریح فقط بخش کوچکی از دانش را تشکیل می‌دهد [۵۱].

۲-۲-۴- مدیریت دانش چیست؟

مدیریت به معنای فرایند به‌کارگیری مؤثر و کارآمد منابع انسانی و مادی بر مبنای یک نظام ارزشی پذیرفته شده است و از طریق برنامه‌ریزی، سازماندهی، بسیج منابع و امکانات، هدایت و کنترل عملیات برای دستیابی به اهداف معین تصمیماتی صورت می‌پذیرد.

مدیریت دانش بر فرایندها و چگونگی کشف و به اشتراک گذاری دانسته‌های یک سازمان و یا ذینفعان خارجی آن سازمان تمرکز دارد. توانایی به اشتراک گذاری به‌روش‌های^۱ داخلی برای کارایی کلی سازمان اهمیت بسیاری دارد [۵۴].

مدیریت دانش برای مدیران مبحثی آشناست و تحقیقات زیادی در این زمینه صورت گرفته است. بسیاری از تحقیقات اثربخشی^۲ مفاهیم مدیریت دانش و نتایج به‌کارگیری از آن در زمینه‌های کاربردی در سازمان را اثبات کرده‌اند. از نمونه‌های این نتایج توسعه محصول جدید^۳ در سازمان می‌باشد [۲۷].

^۱ Best Practices

^۲ Effectiveness

^۳ New Product Development (NPD)

اهمیت مدیریت دانش از آن جهت است که برخورد هوشمندانه با منبع دانش، عاملی مؤثر و مهم جهت موفقیت است. در [۳۸] فایده مدیریت دانش ایجاد رضایت مشتری با ارائه خدمات و محصولات بهتر از سوی سازمان بیان شده است.

اگر مدیریت دانش را به عنوان ابزار کار با اطلاعات در نظر بگیریم، اغلب به دانش به عنوان مسئله‌ای که با اطلاعات سروکار دارد توجه می‌شود که با خلق، مدیریت و استخراج اطلاعات سروکار دارد [۳۶]. در ادبیات تحقیقات دانشجویی مدیریت دانش در گام‌های جداگانه اما مرتبط تعریف شده است. گام اول فرایند کسب^۱ اطلاعات است. گام دوم، ذخیره‌سازی و سازماندهی منطقی اطلاعات ورودی است. مدیریت دانش درباره کسب و ذخیره دانش کارکنان و در دسترس گذاشتن اطلاعات برای سایر کارمندان سازمان است. این هدف معمولاً با استفاده از فناوری‌های گوناگونی نظیر اینترنت و پایگاه داده و ترجمه دانش از دانش صریح به دانش ضمنی (درونی سازی) محقق می‌شود [۴۹] و [۴۱].

زمانی که اطلاعات در پایگاه داده‌های متفاوت ذخیره می‌شود گام سوم آغاز می‌شود. در این مرحله، اطلاعات ذخیره شده باید در دسترس اکثر کارکنان سازمان قرار گیرد. دانش توزیع شده باید در زمان صحیح، کاربران نهایی صحیح و در جایی که نیاز بود در دسترس قرار گیرد. گام نهایی که به‌کارگیری اطلاعات است در مورد بهره‌برداری از اطلاعات است. این فرایند با ارتباطات بین افراد و اشتراک و توزیع دانش با صحبت کردن یا ارتباطات اجتماعی با دیگران و یا با تبادل اطلاعات به فرم دیجیتال یا آنالوگ آغاز می‌شود [۳۶].

۲-۲-۵- استراتژی‌های مدیریت دانش

نوناکا و تاکاشی در [۵۵] بیان می‌کنند دانش به دو دسته آشکار و پنهان تقسیم می‌شود که دانش آشکار (صریح) می‌تواند در قالب واژگان و اعداد بیان و به شکل داده‌ها، فرمول، مشخصات، دستورالعمل‌ها و نظایر آن ثبت شود. این نوع دانش می‌تواند به آسانی کدگذاری شود و به سادگی بین افراد به طور رسمی و نظام‌مند منتقل شود. از طرف دیگر دانش پنهان (ضمنی) در ذهن افراد موجود است و به طور عمیق در اقدامات، تجارب، ارزش‌ها و مطلوب‌های افراد می‌باشد. مدل‌های ذهنی، شهود و تصورات در این حوزه از دانش قرار می‌گیرند. نوناکا و تاکاشی چهار نوع استراتژی خلق و

^۱ acquisition

تبدیل دانش‌های صریح و ضمنی در سازمان‌ها را ارائه نموده‌اند که عبارتند از: اجتماعی‌سازی، برونی‌سازی، ترکیب‌سازی، درونی‌سازی.

جدول ۲-۱ انواع مختلف تبدیلات دانش [۵۱ و ۵۵]

	به دانش آشکار	به دانش ضمنی
از دانش ضمنی	۲- برونی‌سازی	۱- اجتماعی‌سازی
از دانش آشکار	۳- ترکیب‌سازی	۴- درونی‌سازی

۲-۲-۵-۱- استراتژی اجتماعی‌سازی (تبدیل دانش پنهان به پنهان)

در این استراتژی دانش پنهان انتقال می‌یابد و مجدداً به دانش پنهان تبدیل می‌گردد. در استراتژی اجتماعی‌سازی، افراد تجارب و مدل‌های ذهنی خود را جهت بهبود دانش، با دیگران به اشتراک می‌گذارند. این فرایند شامل دستیابی افراد به درک متقابل از طریق تعاملات اجتماعی چهره به چهره، تسهیم دیدگاه‌ها، هم‌اندیشی، تعاملات حمایت‌گری و غیره می‌باشد. برای مثال می‌توان به روابط پیشرفته همکاران، آموزش ضمن کار، روش آزمون و خطا، تقلید از دیگران، جلسات طوفان مغزی، تمرین و آموزش، تبادل عقاید، گفتگوی بسیار، اقدامات تشویقی و... اشاره نمود.

۲-۲-۵-۲- استراتژی برونی‌سازی (پنهان به آشکار)

در این نوع استراتژی، دانش پنهان از طریق فرایند برونی‌سازی به دانش آشکار تبدیل می‌گردد. این فرایند به افراد اجازه می‌دهد که به صورت انفرادی مفاهیم دانش ضمنی را تدوین نموده و با دیگران به اشتراک بگذارند و دانش جدید به وجود آورند. به عبارت دیگر، دانش شخصی یا ضمنی، به صورت استعاره‌ها، تشبیه‌ها، فرضیات و مدل‌ها به دانش صریح تبدیل می‌شود. شخص، زمانی که از تبادلات و توجهات جمعی جهت این فرایند طراحی استفاده می‌کند، اغلب به برونی‌کردن اقدام می‌کند. در این استراتژی اصول مدیریت محتوا به منظور آرشیو کردن، به‌روزرسانی و بازیابی دانش آشکار شده مورد نیاز خواهند بود. نوناکا و تاکاشی از برونی‌کردن به عنوان فرایندی کلیدی در تبدیل دانش یاد می‌کنند. زیرا در اینجاست که دانش ضمنی تجدید می‌شود و طراحی‌های آشکار ظهور می‌کند.

۲-۲-۵-۳- استراتژی ترکیب‌سازی (آشکار به آشکار)

در استراتژی ترکیب‌سازی دانش تجزیه شده موجود، به اشتراک گذاشته شده و ترکیب و تفسیر می‌گردد. در این حالت دانش صریح به دانش صریح پیچیده‌تری تبدیل می‌شود، به عبارت دیگر، عقاید در این مرحله با یک سیستم دانش (مانند پایگاه داده) ترکیب می‌شوند. افراد به تبادل دانش می‌پردازند و این دانش از طریق مستندات، جلسات، مکالمات تلفنی و تبادل اطلاعات از طریق فناوری‌ها و ابزارهایی مانند شبکه‌های کامپیوتری ترکیب می‌شوند. دانش جدید نیز می‌تواند از طریق ساختاردهی مجدد اطلاعات کنونی به وسیله ذخیره‌سازی، افزودن، ترکیب و طبقه‌بندی دانش صریح ایجاد می‌شود. ترکیب، نوعی از ایجاد دانش است که در مقوله آموزش و تحصیلات گنجانده می‌شود. نمونه‌هایی از استراتژی ترکیب‌سازی، سیستم‌های دانش و اطلاعات، تهیه گزارشات مروری، تحلیل روند و خلاصه مدیریتی می‌باشند.

۲-۲-۵-۴- استراتژی درونی‌سازی (آشکار به پنهان)

استراتژی درونی‌سازی شیوه‌ای است که دانش آشکار از طریق تفسیر دانش، درونی می‌شود و به دانش ضمنی تبدیل می‌گردد. این امر می‌تواند از طریق یادگیری حین عمل صورت گیرد و دانش مستندسازی شده، نقش حیاتی را در این فرایند بازی می‌کند. درونی‌سازی زمانی رخ می‌دهد که کارگران جدید دانش، یک پروژه را با مطالعه بایگانی‌های آن، احیاء کنند. همچنین درونی‌سازی را زمانی می‌توان مشاهده کرد که مدیران یا خبرگان با تجربه سخنرانی کنند و یا نویسندگان تصمیم به نوشتن بیوگرافی از یک کار آفرین یا مؤسسه داشته باشند. به محض درونی شدن، دانش‌های جدید مورد استفاده کارکنانی قرار می‌گیرد که آن را در پایگاه دانش پنهان موجود خود توسعه داده و سازماندهی مجدد کرده‌اند [۵۱].

۲-۲-۶- معایب عدم بهره‌گیری از دانش در سازمان

برخی از مشکلات نبود مدیریت دانش را می‌توان به صورت زیر نام برد:

- عدم اولویت‌بندی و استفاده از انواع دانش
- ایجاد شکاف دانشی در صورت جدا شدن کارشناسان از سازمان
- احتکار دانش به جای ازدیاد آن
- استفاده محدود از دانش‌های موجود

- نبود مستندات در خصوص تجربیات به دست آمده از پروژه‌ها و کارها
- نبود ساختار مناسب برای تسهیم سریع دانش.

۲-۲-۷- اهداف مدیریت دانش

از اهداف مدیریت دانش می‌توان به موارد زیر اشاره نمود:

- ایجاد ساختار راهبری مدیریت دانش؛
- مدیریت دانش به عنوان یک فعالیت اصلی؛
- ایجاد نرم‌افزاری برای انتقال دانش؛
- ایجاد چارچوبی برای اندازه‌گیری دانش‌ها و دستاوردهای اجرای فرآیندها؛
- ایجاد انجمن‌های خبرگی.

اهمیت مدیریت دانش در بین کسب‌وکارهای پیشرفته امروزی به خوبی شناخته شده است. به طوری که هر ساله ارزیابی جهانی سازمان دانشی برتر^۱ از سوی شبکه جهانی دانش^۲ برگزار می‌گردد و بسیاری از شرکت‌های معتبر و پیشرو در عرصه رقابت از کشورهای مختلف جهان به طور مستمر در این ارزیابی شرکت نموده و مجموعه اقدامات خود در حوزه مدیریت دانش محور را بر اساس صد و پنجاه شاخص طرح شده مورد ارزیابی قرار داده و از گزارش‌های تحلیلی جامع و نیز تجربیات سازمان‌های برتر دنیا در این ارزیابی بهره‌مند می‌شوند.

۲-۲-۸- مدل‌های مدیریت دانش

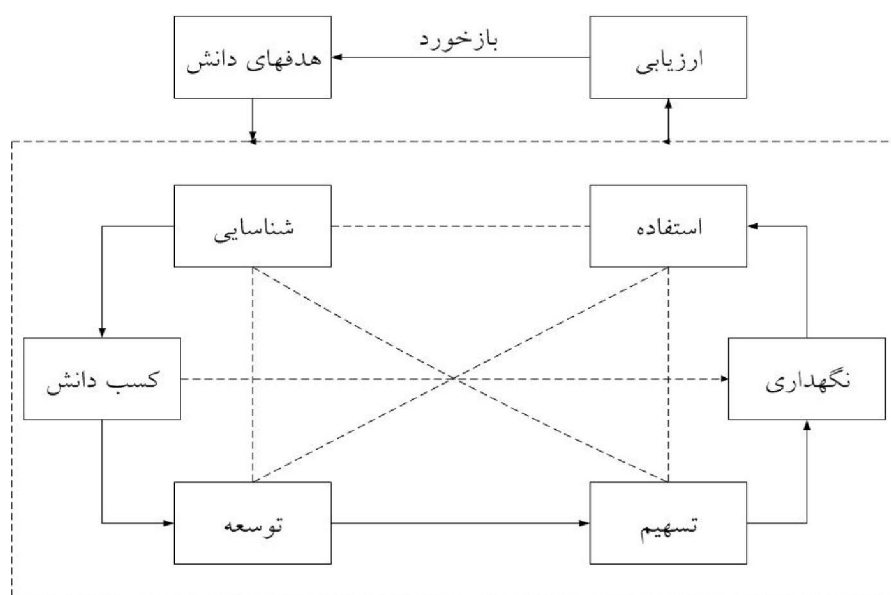
در سال‌های اخیر محققان بسیاری در زمینه مدیریت دانش کار کرده‌اند و مدل‌های بسیاری در این زمینه ارائه شده است. این مدل‌ها معمولاً اصل و محتوای مشابهی دارند اما اصطلاحات و ترتیب فازها در آن‌ها ممکن است متفاوت باشد. نکته‌ای که در تمامی این مدل‌ها به چشم می‌خورد تأکید بر استفاده و به‌کارگیری دانش است و مراحل دیگر به عنوان مقدمه‌ای در ایجاد بسترهای لازم و مناسب برای به‌کارگیری دانش هستند [۹]. این مدل‌ها عمدتاً دو تا هشت مرحله دارند. مدل نوناکا و تاکاچی (۱۹۹۵)، مدل هفت سی (۱۹۹۶)، مدل بک من (۱۹۹۹)، مدل هیسینگ (۲۰۰۰)، مدل سنگ بنای

^۱ MAKE

^۲ Know network

مدیریت دانش^۱ (۲۰۰۲) و مدل مارک م.مک الروی (۲۰۰۲) و ... از مدل‌های ارائه شده در زمینه مدیریت دانش هستند [۴ و ۹].

از آنجا که مدل پایه‌های ساختمان دانش پربست نسبت به سایر مدل‌های دانش کاربردی‌تر و جامع‌تر می‌باشد، در ادامه شرح مختصری از این مدل خواهیم داشت.



شکل ۲-۳ مدل مدیریت دانش پروبست و رمهارد [۴]

این مدل شامل هشت جزء است که مدیریت دانش در آن به صورت سیکلی جریان دارد. این مدل شامل دو سیکل درونی و بیرونی است و سیکل درونی شامل مراحل: کشف (شناسایی)، کسب، توسعه، تسهیم، کاربرد و نگهداری از دانش می‌باشد. سیکل بیرونی شامل بلوک‌های اهداف دانش و ارزیابی دانش است و کامل کننده این دو سیکل بازخورد است [۴ و ۹].

تعیین اهداف دانش: اهداف مدیریت دانش باید از اهداف اصلی سازمان در دو سطح استراتژیک و عملیاتی نشئت می‌گیرند.

شناسایی دانش: طرح سؤال آیا می‌دانیم که چه می‌دانیم مرحله آغازین شناسایی دانش است. شناسایی منابع دانشی هم بیرون از مرزهای سازمان و هم درون سازمان صورت پذیرد.

^۱ این مدل توسط پروبست و رمهارد (Probst & Raub & Romhard) ارائه شده است.

توسعه دانش: دانش باید از منابع شناسایی شده بازار داخلی و خارجی و از منابعی نظیر مشتریان، رقبا، همکاران و ... جمع آوری گردد.

توسعه دانش: باید با دانش کسب شده سازمان را توسعه داد. این توسعه می‌تواند شامل هرگونه تغییر یا نوآوری در ارائه خدمات و محصولات به مشتریان باشد.

تسهیم دانش: دانش کسب شده باید در زمان و مکان مناسب بدست افراد مناسب برسد. چگونگی انتقال دانش از سطح فردی به سطح گروهی و سازمانی در این فاز مورد بررسی قرار می‌گیرد.

استفاده از دانش: در این فاز موانع موجود بر سر راه دانش باید شناسایی و برطرف گردند تا دانش در سازمان به صورت عملی و مفید در ارائه خدمات و محصولات بکار آید.

نگهداری دانش: در این فاز هدف حفظ، ذخیره و به‌روزرسانی دانش خواهد بود.

ارزیابی: در فاز ارزیابی نحوه رسیدن به اهداف تعیین شده و استفاده از نتایج به عنوان بازخورد در این فاز صورت می‌گیرد [۴].

۲-۳- مدیریت دانش مشتری^۱

امروزه در زمان استفاده از خدمت یا مصرف کالا توسط مشتری دانش قابل توجهی تبادل می‌شود. این دانش به عنوان منبعی مهم برای سازمان و کسب و بهره‌برداری از آن به یک مزیت رقابتی در سازمان‌ها تبدیل شده است.

مدیریت دانش مشتری دربردارنده فرایندهایی است که با شناسایی و اکتساب اطلاعات مشتری و نیز ایجاد و بهره‌برداری از دانش مشتریان، مربوط است [۹]. چنین اطلاعاتی در ماورای محدوده‌های خارجی سازمان قرار دارند و دانشی که از آن‌ها استخراج می‌شود موجب ایجاد ارزش برای سازمان و مشتریان آن خواهد شد [۳۲].

در فرایند ایجاد دانش، مشتری به عنوان شریک سازمان شناخته می‌شود و از آنجا که هر دو هدف مشترکی در ذهن دارند (خدمات و محصولات بهتر)، هر دو طرف می‌توانند شرکای خوبی در فرایند ایجاد دانش باشند.

^۱ Customer Knowledge Management (CKM)

مدیریت دانش مشتری سرمایه مهمی برای انواع سازمان‌هاست و اساس ارتقای ارزش مشتری است [۴۲]. با تأکید بر دانش به عنوان فاکتور رقابتی مهم در اقتصاد جهانی، سازمان‌ها باید به دنبال یک عنصر کلیدی باشند: دانش مشتری [۲۷].

دانش مشتری نوعی از دانش (یا داده و اطلاعاتی است که قابلیت تحلیل و تفسیر و نهایتاً تبدیل به دانش را دارند) در حوزه ارتباط با مشتری است که قابلیت تأثیرگذاری مستقیم یا غیر مستقیم بر عملکرد سازمانی را دارا می‌باشد [۴۳]. دانش مشتری اطلاعاتی است که جهت برقراری ارتباط قوی‌تر با مشتری مورد نیاز است [۴۷]. بهره‌گیری از دانش مشتری مزایایی را هم برای سازمان و هم برای مشتری به دنبال خواهد داشت. از جمله این مزایا می‌توان به بهبود محصولات و خدمات ارائه شده به مشتری، افزایش فروش، جلب رضایت مشتری، حفظ مشتری و ... اشاره نمود [۳۴].

مدیریت دانش مشتری و مدیریت دانش به دنبال اهداف یکسانند: ارتقای سطح رضایت مشتری. نتایج حاصل از این تلاش دانش‌گرا مدیریت دانش مشتری و یا مدیریت ارتباط با مشتری نامیده می‌شود [۲۷].

در ادبیات تحقیقات مدیریت دانش، مدیریت دانش مشتری، به عنوان به‌کارگیری انواع دانش شامل دانش برای مشتری^۱ (مثلاً اطلاعات یک محصول)، دانش از طرف مشتری^۲ (مثلاً ایده‌هایی برای بهبود محصول) و دانش درباره مشتری^۳ (مثلاً نیازها و انتظارات مشتریان) برای ارتقاء قابلیت‌های ارتباط با مشتریان شرکت است [۱۵].

۲-۳-۱- انواع دانش مشتری

دانش مشتری در ارتباطات بین مشتری و سازمان انواع متفاوت دارد که عبارتند از:

دانش برای مشتری، دانش از طرف مشتری، دانش درباره مشتری.

دانش درباره مشتری: نوعی از دانش است که به سازمان کمک می‌کند مشتریانش را بشناسد و به طور مؤثری در مورد آن‌ها هدف‌گذاری کند. سه نوع دانش درباره مشتری وجود دارد: جمعیت‌شناختی، روان‌شناختی و رفتاری [۳۶].

^۱ Knowledge For Customer

^۲ Knowledge From Customer

^۳ Knowledge About Customer

دانش درباره مشتری می‌تواند یکی موارد زیر باشد [۷ به نقل از ۴۲]:

- دانش و اطلاعات درباره مشتریان فعلی (شامل دانش و اطلاعات تاریخی و یا دانش و اطلاعات مربوط به نیازمندی مشتریان).
- دانش و اطلاعات در مورد مشتریان بالقوه (شامل دانش و اطلاعات مربوط به علاقمندی و ترجیحات آن‌ها).

دانش برای مشتری: این نوع دانش برای ارتقاء تجربیات مشتریان در مورد محصولات و خدمات شرکت کاربرد دارد [۷ به نقل از ۴۲]. این دانش دانشی است که از طرف سازمان به مشتری ارائه می‌شود تا مشتری شناخت بهتری از سازمان، محصولات و خدمات سازمان کسب نماید. دانش برای مشتری شامل اطلاعاتی در مورد محصولات، بازارها و تأمین‌کنندگان است که بر روی درک مشتری از کیفیت خدمات تأثیرگذار خواهد بود [۳۳].

انواع دانش برای مشتری عبارتند از:

- دانش و اطلاعات محصولات و خدمات سازمان (شامل دانش و اطلاعات مشتریان فعلی و بالقوه).
- دانش و اطلاعات عمومی شرکت (شامل اطلاعات معرف و محصولات دانشی سازمان).
- دانش و اطلاعات محیطی شرکت (مانند دانش و اطلاعات مربوط به بازار و محصولات و خدمات رقبا) [۷ به نقل از ۴۲].

دانش از طرف مشتری: واضح است که تمرکز سازمان باید بر روی تحویل آنچه مشتریان نیاز دارند باشد نه توسعه محصولات و خدماتی که مشتریان نیاز و تمایلی به آن ندارند، سازمان نیازمند این خواهد بود که نظرات و اطلاعات مشتریان را در مورد نیازهای آتی و نوآوری‌های مدنظر سازمان بداند تا محصولات و خدمات مورد نیاز آن‌ها توسعه داده شود [۷]. اما مشتریان متوجه نیازهای خود نخواهند شد مگر اینکه محصولات نوآورانه به آن‌ها پیشنهاد شود و سازمان به آن‌ها کمک کند که محصولات یا خدمات را با در نظر گرفتن نیازها و ترجیحات خود شخصی‌سازی کنند [۳۹].

- گرایش‌ها و رفتارهای مشتریان چیستند؟
- چگونه می‌توان قدرت روابط با مشتری را اندازه گرفت؟
- چگونه می‌توان برای قابلیت سوددهی و ارزش‌آفرینی مشتری برنامه‌ریزی کرد و آن را افزایش داد؟
- چگونه می‌توان مشتریان را به گروه‌هایی تقسیم کرد تا تحلیل‌های بازاریابی و ارتباطات با آن‌ها راحت‌تر شود؟
- چه کسانی مشتریان بلندمدت ما هستند؟ (۵ سال و بیشتر)
- رفتار مشتریان در ۵ سال آینده چگونه تغییر خواهد کرد؟ [۱۷]

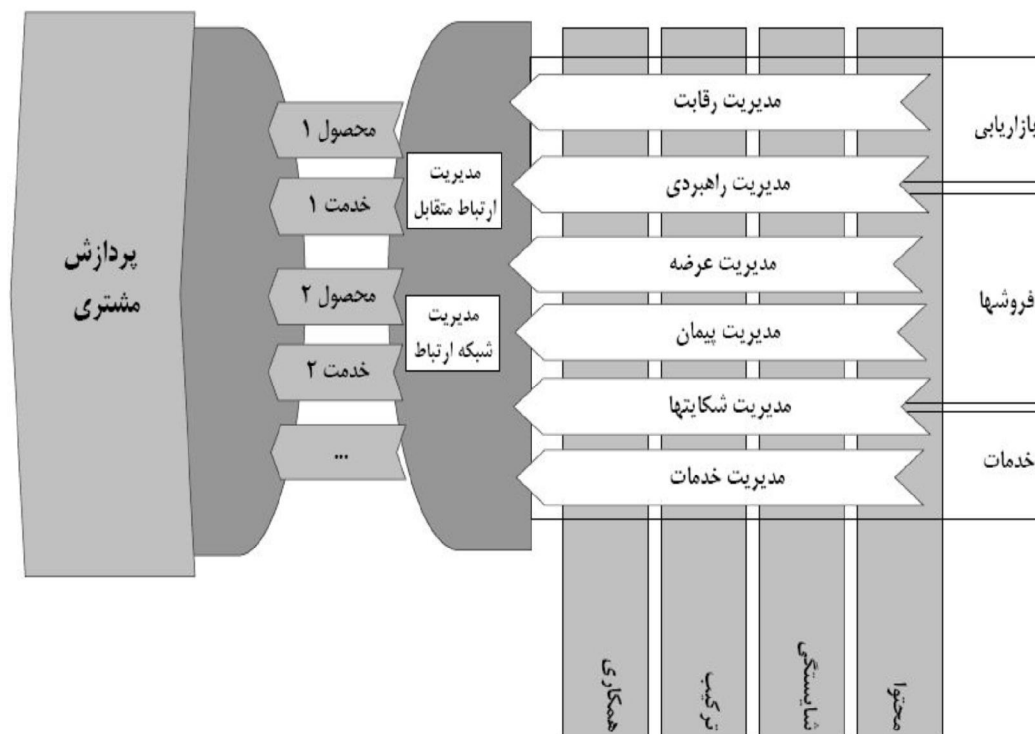
همچنین Rowly و همکاران در [۴۲] راهکارهای زیر را جهت دستیابی به دانش مشتری پیشنهاد می‌دهند و تأکید می‌کنند که برای این راهکارها باید در نظر داشت که اطلاعات حاصل از هر کدام از آن‌ها به روش‌های تحلیلی خاصی نیاز دارند و هر کدام مشخصاتی مخصوص به خود دارند. این راهکارها عبارتند از:

- مشتری در ازای درخواست اطلاعات پاسخ دهد. این روش می‌تواند به عنوان بازخوردی از نحوه خدمات باشد یا قسمتی از یک سیستم مبتنی بر اینترنت و ... باشد.
- مشتری آنلاین اطلاعاتی را در مورد خود ارائه دهد. در این شیوه می‌توان با بازبینی رونوشت صحبت‌ها به اطلاعات دست یافت.
- مشتری اطلاعات را در ازای دریافت یک محصول ارائه دهد. مثلاً فروشنده می‌تواند نام، آدرس و ایمیل یا دیگر اطلاعات مشتری را ذخیره کند.
- اطلاعات از مسیر جستجوی مشتری دریافت شود. می‌توان از طریق دنبال کردن آدرس IP یک مشتری به مسیر جستجوهای او و علایق او دست یافت.
- فروشنده با استفاده از کوکی‌ها^۱ فعالیت‌های مشتری را دنبال کند، از این طریق می‌توان ورود و خروج مشتری به وب‌سایت را تعقیب نمود و متوجه شد که او به چه قسمت‌هایی علاقه بیشتری دارد [۱۷].

^۱ Cookies

۲-۳-۲- مدل مدیریت دانش مشتری

شکل ۲-۵ یک مدل را برای مدیریت دانش مشتری نمایش می‌دهد. طبق [۳۷] مدل فرایند CKM یکپارچگی دو مفهوم CRM و KM را مد نظر دارد. ادغام دو مفهوم CRM و KM در مدل CKM می‌تواند منافع به‌کارگیری هر یک را ارتقاء و ریسک به‌کارگیری هر یک را کاهش دهد [۹].



شکل ۲-۵ مدل مدیریت دانش مشتری [۱۷ به نقل از ۳۷]

در این مدل وجه محتوا راهنمایی است برای جواب دادن به این سؤال که چه موقع باید دانش از اشخاص جدا شود و به صورت بسته‌های اطلاعات خلاصه‌سازی شود. خلاصه‌سازی اجازه می‌دهد که ذخیره و توزیع دانش از نظر هزینه با کارایی بیشتری انجام شود.

وجه شایستگی نیروی کار به این سؤال پاسخ می‌دهد که نیروی کار باید چه دانشی داشته باشند تا پیشرفت رقابتی را برای فرآیندهای تجاری فراهم کند. این امر می‌تواند با آموزش نیروی کار موجود یا استخدام نیروی کار جدید انجام پذیرد.

وجه همکاری بر این سؤال تأکید دارد چگونه می‌توان از دانش در میان گروه‌هایی همچون تیم پروژه یا چرخه‌های کیفیت به صورت مؤثر حمایت کرد.

وجه ترکیب بر توزیع دانش در داخل سازمان به صورتی که از نظر هزینه‌ای به صرفه باشد تأکید می‌ورزد. این عمل به کمک بسته‌های اطلاعاتی انجام می‌گیرد.

و جوه ۴ گانه دانش به مدیریت دانش اجازه می‌دهند که بر اساس این ابعاد و مشخصات که به طور خاص بر عملکرد فرآیندها تأثیر خواهند گذاشت، عمل کند.

در سطوح بالای مدیریت یک شرکت باید مشخص باشد که چگونه مدیریت دانش مشتری می‌تواند اهداف و فرآیندهای تجاری آن‌ها را حمایت کند. در این صورت می‌توان از این موارد به عنوان راهنمایی برای طراحی فرآیندهای مدیریت دانش مشتری و شاخص‌های عملکرد آن استفاده کرد [۱۷].

در واقع مدل مدیریت دانش مشتری راهکارهای گوناگونی را برای یافتن مشخصات دانش مناسب از افق دید تجاری دنبال می‌کند. در عمل مشاهده می‌شود که مدیریت روابط با مشتری و مدیریت دانش هم‌افزایی قابل توجهی دارند. هنگامی که مدیریت دانش به عنوان یک خدمت‌دهنده برای مدیریت روابط مشتری عمل می‌کند، استقلال بین این دو حفظ می‌شود و از هم‌افزایی بین این دو راهکار منفعت‌های مشترکی به وجود می‌آید [۱۷].

۲-۴- مدیریت ارتباط با مشتری^۱

افزایش سرعت تحولات و ناپایداری عرصه رقابت در عصر حاضر که در نتیجه تغییرات فناوری و فشردگی فضای رقابت، در عرصه کسب‌وکار به خصوص در سازمان‌هایی چون بانک‌ها و بیمه‌ها حاصل شده است، نیاز بازار به چاره‌اندیشی جهت کسب مزیت رقابتی بیشتر را نمایان می‌سازد.

تغییرات سریع سلايق مشتریان و به دنبال آن ورود محصولات جدید بهبود یافته مطابق این سلايق موجب کاهش طول عمر مفید ارائه محصولات و خدمات می‌شود. بنابراین در این شرایط استفاده از سیستم‌هایی همچون مدیریت ارتباط با مشتری نه تنها یک مزیت رقابتی بلکه یک ضرورت برای سازمان محسوب می‌شود. مدیریت ارتباط با مشتری یک ابزار مدیریتی جدید است که ارتباطات شرکت با مشتریان را با به‌کارگیری فناوری اطلاعات به‌روز (مانند تحلیل داده‌های برخط، داده‌کاوی و مدیریت

^۱Customer Relationship Management (CRM)

پایگاه داده) مدیریت می‌کند. هدف مدیریت ارتباط با مشتری افزایش رضایت و وفاداری مشتری است [۴۰].

مدیریت ارتباط با مشتری کلیه فرایندها و فناوری‌هایی است که سازمان برای شناسایی، انتخاب، ترغیب، گسترش، حفظ و خدمت به مشتریان خود بکار می‌گیرد.

مدیریت ارتباط با مشتری سازمان را قادر می‌سازد تا از دانش مشتری جهت تغییر، توسعه یا تجدید خدمات و محصولات به مشتری استفاده نمایند و از این مسیر سود نهایی حاصل برای سازمان را افزایش دهند [۹].

سیستم‌های مدیریت ارتباط با مشتری در حقیقت راهبردی برای جمع‌آوری نیازها و رفتارهای تجاری مشتریان جهت ایجاد روابط قوی‌تر با آنهاست و از سه بخش اصلی مشتری و روابط و مدیریت تشکیل شده است. منظور از مشتری مصرف‌کننده نهایی است. منظور از روابط ایجاد مشتریان وفادارتر و سودمندتر از طریق ارتباط یادگیرنده می‌باشد؛ و در نهایت نقش مدیریت در این سیستم عبارت است از خلاقیت و هدایت یک فرایند کسب‌وکار مشتری‌مدار و قرار دادن مشتری در مرکز فرایندها و تجارب سازمان [۱۰].

در تعریفی دیگر مدیریت ارتباط با مشتری شامل مجموعه‌ای از فرایندها و سیستم‌هاست که جهت ایجاد روابط بلندمدت سودآور با مشتریان، استراتژی‌های کسب‌وکار پیشنهاد می‌کند [۴۶].

داده‌های مشتریان و ابزارهای تکنولوژی اطلاعات زمینه‌ای را برای ایجاد ارتباط موفق با مشتری ایجاد می‌کنند. علاوه بر این سرعت روبه رشد اینترنت و تکنولوژی‌های وابسته باعث افزایش فرصت بازاریابی شده است و راه‌های ارتباط بین سازمان و مشتریان را تغییر داده است [۲۵].

بازاریابی ارتباطی، سودآوری مشتری^۱، ارزش دوره حیات مشتری^۲ و رضایت^۳ و حفظ^۴ مشتری از طریق فرایندهای کسب‌وکار مبانی فلسفی مدیریت ارتباط با مشتری را تشکیل می‌دهند [۱۱].

^۱ Customer Profitability

^۲ Lifetime Value

^۳ Satisfaction

^۴ Retention

مشتریان در ترجیحات و عادات خرید خود با یکدیگر متفاوت‌اند. این مسئله را می‌توان علت ابداع مدیریت ارتباط با مشتری دانست. اگر همه مشتریان مانند یکدیگر بودند نیاز چندانی برای مدیریت ارتباط با مشتری و درک انگیزه‌های مشتریان و سودآوری آن‌ها احساس نمی‌شد و اقدامات سازمان به راحتی جهت حداکثرسازی ارزش سازماندهی می‌شد [۳۹].

مدیریت ارتباط با مشتری رویکردی سازمانی برای درک و تأثیرگذاری بر رفتار مشتری از طریق ارتباط معنی‌دار جهت بهبود شناسایی، جذب، حفظ و توسعه مشتری و وفاداری و سودآوری او تعریف می‌شود [۲۵].

بر اساس [۲۵] مدیریت ارتباط مشتری شامل چهار بعد است:

- شناسایی مشتریان: این مرحله فاز آغازین مدیریت ارتباط با مشتری است که طی آن افرادی را که احتمال زیاد وجود دارد که مشتری ما شوند یا برای سازمان سودآور باشند مورد هدف قرار می‌دهد. شناسایی مشتریان در این مرحله، از طریق تجزیه و تحلیل ویژگی‌های اساسی آن‌ها و بخش‌بندی مشتریان بر اساس خصوصیات مشابه و سپس تصمیم‌گیری در مورد هر گروه صورت می‌پذیرد [۱۴].
- جذب مشتری: در پی شناسایی مشتریان بالقوه، منابعی از سازمان جهت دعوت این مشتریان به استفاده از خدمات یا محصولات سازمان تخصیص داده می‌شود.
- حفظ مشتری: این مسئله بحث اصلی در فرایند مدیریت ارتباط با مشتری است و رضایت مشتری شرط اساسی بقای ارتباط او با سازمان است. این هدف معمولاً با روش‌هایی چون بازاریابی فرد به فرد و تشخیص و پیش‌بینی رفتار مشتری و فعالیت‌های پشتیبانی روابط بلندمدت مشتریان حاصل خواهد شد [۱۴].
- توسعه مشتری: این فاز شامل توسعه کثرت تراکنش‌ها، ارزش تراکنش‌ها و سودآوری مشتری می‌باشد که با فروش محصولات بیشتر به یک مشتری و یا فروش یک محصول به مشتریان بیشتر انجام می‌پذیرد [۱۴].

هدف کلی مدیریت دانش مشتری ایجاد درک عمیق‌تر مشتریان برای حداکثر کردن ارزش مشتریان در بلندمدت است. تجزیه و تحلیل و درک رفتارها و ویژگی‌های مشتریان اساس توسعه یک استراتژی CRM رقابتی می‌باشد [۱۱].

هدف CRM افزایش فرصت‌های کسب‌وکار به روش‌های زیر است:

- بهبود فرایند ارتباط با مشتریان واقعی
- ارائه محصولات صحیح از طریق کانال‌های صحیح به هر مشتری
- ارائه محصولات صحیح در زمان صحیح به هر مشتری [۱۴].

۲-۴-۱- مدیریت ارتباط مشتریان در نظام بانکی

در سال‌های اخیر با اهمیت یافتن مشتری از یک سو و تحولات صورت گرفته در زمینه فناوری اطلاعات از سوی دیگر شاهد پیدایش مفاهیمی چون مدیریت ارتباط با مشتری به عنوان رویکرد مهمی از کسب‌وکار بوده‌ایم. این مفهوم بر این معنا دلالت دارد که هر مشتری محصولات و خدمات خاصی را می‌طلبد و از این‌رو بهتر است به جای بازاریابی انبوه، مشتریان را تک‌تک بازاریابی کنیم. برای این منظور با استفاده از دانش مشتری نظیر سابقه خرید و نیازهای جدید او کالاها و خدمات در مسیر مورد نظر وی سازماندهی شود [۱۰].

بازاریابی از ابزارهایی است که می‌تواند منجر به رشد درآمدهای بانک شود. بنا به گفته مدیران بزرگ بانک‌های معروف دنیا فقط ۵٪ از مشتریان بیش از ۸۵٪ سودآوری بانک‌ها را تشکیل می‌دهند. تحقیقات نشان می‌دهد که سودآوری بانک‌ها با افزایش مشتریان ارزنده و تراز اول و ایجاد رضایت مؤثر در آن‌ها، به طور چشمگیری افزایش می‌یابد [۱۰].

در بازار رقابت تنگاتنگ کنونی در عرصه صنعت بانکداری با توجه به تأسیس تعداد زیادی از بانک‌های خصوصی، بانک‌ها را ملزم می‌سازد که برای ربودن گوی سبقت، با ارتباطات مشتری‌مدار، توجه خود را به مشتری و حفظ او افزایش دهند. چرا که امروز مشتریان فرصت‌های بیشتری برای مقایسه و انتخاب خدمات پیش رو دارند [۱۰]. مشتریان ارزش زیادی برای تکنولوژی و سرعت و تخصص فنی بانک‌ها قائل‌اند. اینکه بانک بتواند سرویس‌های مورد نظر مشتری را به خوبی ارائه نماید از عناصر مهم موفقیت بانک است. بنابراین بانک‌ها نیازمند کسب اطلاعات کافی از مشتری و تحلیل این اطلاعات جهت کشف دانش نهفته در این اطلاعات و درک علایق و خواسته‌های مشتری جهت توسعه روابط و ارائه خدمات و محصولات متناسب و مطلوب هر مشتری می‌باشند [۱۰].

CRM به معنی مدیریت ارتباط با مشتری است و نه بازاریابی ارتباط با مشتری و مدیریت مفهومی گسترده است که مستلزم رویکردی سازمانی در تمام سطوح کسب و کار و همه بخش‌های تولید، منابع انسانی، مدیریت، خدمات، فروش و تحقیق و توسعه است.

امروزه در سازمان‌هایی نظیر بانک‌ها برخلاف گذشته یافتن مشتریان جدید تنها هدف بازاریابی نیست و تنها رضایتمندی مشتریان برای ماندگاری آن‌ها کافی نبوده و باید در عین رضایتمندی از وفاداری آن‌ها نیز اطمینان حاصل نمود. در این ارتباط با هدف برقراری روابط بلندمدت و متقابل با گروه‌های ذی‌نفع بانک و بالاخص مشتریان است، به نحوی که مشتریان بیشتری حفظ شوند که ماحصل این موضوع سهم بیشتر بازار و در نتیجه سودآوری بیشتر برای بانک است [۱۰ به نقل از ۳].

۲-۴-۲- مدیریت ارتباط با مشتری: اهداف، مزایا و چالش‌ها

اهداف اصلی مدیریت دانش مشتری در مؤسسات مالی و اعتباری و بانک‌ها در [۱۰] چنین ذکر شده است:

- شناسایی ارزش‌های خاص هر بخش از بازار مشتریان
- ارائه ارزش‌های دلخواه مشتریان به شیوه مورد درخواست آن‌ها برای دریافت اطلاعات
- تقسیم بخش‌های مختلف بازار و بهبود فرایند ارتباط با مشتریان هدف
- افزایش درآمد حاصل از محل کارمزد ارائه خدمات
- افزایش رضایتمندی و وفاداری مشتریان
- بهینه‌سازی کانال‌های خدمت‌دهی به مشتریان
- جذب مشتریان جدید با توجه به تجربیات کسب شده در خصوص مشتریان سابق
- کسب نظرات و علاقمندی‌های مشتریان به منظور بهینه‌سازی استراتژی و فرایندهای عملیات [۱۰].

از مزایای CRM می‌توان به

- شناخت عمیق‌تر مشتری
- شناسایی مشتریان کلیدی و ارائه خدمات خاص به آن‌ها
- پاسخگویی سریع و با دقت بیشتر به درخواست مشتریان
- فراهم نمودن شرایط مراجعه مجدد مشتریان

- کاهش هزینه‌های تبلیغاتی
- افزایش فرصت‌های بازاریابی و فروش
- افزایش سطح رضایت مشتری
- افزایش سودآوری
- دریافت بازخورد از مشتری و توسعه خدمات و محصولات جاری بر اساس نیازها و خواسته‌های مشتری [۱۰].

از چالش‌های اجرای CRM می‌توان به

- هزینه‌بر و وقت‌گیر بودن اجرای CRM
- عدم پذیرش اجرای CRM از سوی برخی مدیران، کارکنان و مشتریان
- نبود زیرساخت‌های مناسب جهت اجرا
- موقعیت مکانی شعبه که ممکن است امکان دسترسی به شعبه را دشوار سازد
- امکان ارائه خدمات خاص با کیفیت بالاتر از سوی بانک‌های رقیب
- وجود مشتریان وفادار به رقیب [۱۰].

۲-۵- مقایسه مفاهیم CKM و KM و CRM

هرچند در نگاه اول ممکن است CKM یا KM نام دیگری برای CRM به نظر برسد اما مدیران دانش مشتری نیازمند رویکردی متفاوت هستند و به انجام یکسری فعالیت متفاوت بر روی تعدادی متغیر کلیدی نیاز دارند [۳۸].

CKM حوزه‌ای از مدیریت است که در آن رویه‌های KM برای حمایت از تبادل دانش مشتری در سازمان و بین سازمان و مشتریان آن بکار گرفته می‌شود. جایی که دانش مشتری برای CRM، برای بهبود فرایندهای CRM از قبیل خدمت به مشتری، حفظ مشتری و سودمندی ارتباط با مشتری مورد استفاده قرار می‌گیرد.

مدیران دانش مشتری بیش از آن که به دانش در مورد مشتری اهمیت دهند به دانش از طرف مشتری علاقه دارند [۳۸]. در نتیجه شرکت‌های پیشرو جستجوی دانش از طریق روابط مستقیم با مشتری را با جستجوی دانش از طریق نمایندگان فروش خود همراه کرده‌اند. مدیریت دانش مشتری با مدیریت دانش سنتی در اهدافی که دنبال می‌کنند تفاوت دارند. مدیریت دانش سنتی به دنبال به دست

آوردن کارایی است در حالی که مدیریت دانش مشتری در پی نوآوری و رشد است. مدیریت دانش مشتری به دنبال موقعیت‌هایی است که مشتریان را به عنوان همکاری در پدید آوردن ارزش برای سازمان مشارکت دهد [۱۷].

مثل معروفی در دنیای مدیریت روابط با مشتری بیان می‌دارد که نگاه‌داری مشتری بهتر از به دست آوردن اوست. ولی متأسفانه در دنیای کنونی که رقبا به سرعت می‌توانند همانندسازی کنند، نگاه‌داری مشتری بسیار سخت است. در نتیجه مدیران دانش مشتری کمتر با ارقام نگاه‌داری مشتری روبرو هستند. در عوض آن‌ها باید بر چگونگی رشد و همکاری بین مشتری جدید و شرکت در یک محیط فعال متمرکز شوند [۱۷].

جدول ۲-۲ مقایسه مفاهیم مدیریت دانش، مدیریت ارتباط با مشتری و مدیریت دانش مشتری [۳۸]

KM	CRM	CKM	
کارمندان، تیم‌ها و شرکای تجاری	پایگاه داده‌های مربوط به مشتری	تجارب، خلاقیت‌ها و عدم رضایت از محصولات و خدمات از سوی مشتری	مکان‌های جستجوی دانش
اگر فقط می‌دانستیم که چه می‌دانیم.	حفظ مشتری بسیار آسان‌تر از بدست آوردن مشتری است.	اگر فقط می‌دانستیم که مشتریان چه می‌دانند.	قاعده کلی
اشتراک دانش در مورد مشتری در بین کارمندان	کاوش دانش در مورد مشتریان در پایگاه داده‌های مشتری	کسب، اشتراک و گسترش دانش مشتری-تجارب شخصی یا گروهی در کاربردها و رفتارهای رقابتی-راهکارهای مربوط به آینده	اهداف
منفعل یا غیر فعال و دریافت کننده محصولات	درگیر و مرتبط با وفاداری به محصولات و خدمات	فعال، شریک دانش	نقش مشتری
کارمند	مشتری	مشتری	دریافت کننده محرک‌ها یا مشوق‌ها
تشویق کارکنان برای تسهیم دانش خود با همکاران	برقراری روابط پایدار با مشتریان	تبدیل مشتری از حالت دریافت کننده صرف خدمات و محصولات به شریک فعال در خلق ارزش	نقش سازمان

اهداف سازمانی	مشارکت با مشتریان جهت ایجاد ارزش مشترک	رشد مشتری محور مشتری و حفظ مشتری	اثربخشی و افزایش سرعت و پرهیز از تولید مجدد اطلاعات
منافع	موفقیت مشتری و نوآوری و یادگیری سازمانی	حفظ مشتری	رضایت مشتری
معیارهای سازمانی	کارایی با رشد و نوآوری در برابر رقبا و مشارکت برای کامیابی مشتری	کارایی بر حسب رضایت و وفاداری مشتری	کارایی در مقابل بودجه

مدیران دانش مشتری بیش از آنکه به دانش در مورد مشتری اهمیت دهند به دانش از طرف مشتری علاقه دارند. در نتیجه شرکت‌های پیشرو جستجوی دانش از طریق روابط مستقیم با مشتری را با جستجوی دانش از طریق نمایندگان فروش خود همراه کرده‌اند. مدیریت دانش مشتری با مدیریت دانش سنتی در اهدافی که دنبال می‌کنند تفاوت دارند. مدیریت دانش سنتی به دنبال به دست آوردن کارایی است، در حالی که مدیریت دانش مشتری در پی نوآوری و رشد است. مدیریت دانش مشتری به دنبال موقعیت‌هایی است که مشتریان را به عنوان همکاری در پدید آوردن ارزش برای سازمان مشارکت دهد [۳۸].

به طور خلاصه تفاوت‌های اساسی بین مفاهیم CKM و KM و CRM را می‌توان به صورت زیر عنوان نمود:

- مدیران دانش مشتری اولین و مهم‌ترین تمرکزشان بر روی دانش از طرف مشتری است در صورتی که تمرکز بر دانش درباره مشتری ویژگی شاخص مدیریت ارتباط با مشتری است [۳۸].
- در مدیریت دانش سنتی، دانش آن‌چنان در سازمان نقش اساسی یافت که به منبع ارزش‌افزای کلیدی در سازمان تبدیل شد. این دانش هنوز هم در سازمان به اشتراک گذاری و بسط داده شده و در بین مدیران و کارکنان سازمان‌ها بکار می‌رود. با این تفاوت که سعی شده مهم‌ترین شریک سازمان یعنی مشتری را در فرایندهای خلق ارزش در سازمان مشارکت دهد. در مدیریت دانش تمرکز بر این مسئله است که «اگر ما می‌دانستیم که چه می‌دانیم». اما در مدیریت دانش مشتری سازمان به دنبال این است که

بداند «مشتری چه می‌داند» بنابراین مبنای کار در مدیریت دانش مشتری «اگر ما می‌دانستیم مشتریانمان چه می‌دانند» می‌باشد [۳۸].

• مدیران دانش مشتری کمتر نگران حفظ مشتری هستند، در عوض بر روی چگونگی توسعه و پیشرفت سازمان با بدست آوردن مشتریان جدید و ارتباط ارزش‌افزا با آنها تمرکز دارند.

۲-۶- تاریخچه‌ای از بانک و بانکداری

کلمه بانک از واژه ایتالیایی بانکا به معنی نیمکت مشتق شده است. از آنجایی که صرافان ایتالیایی ابتدا در پشت نیمکت عملیات صرافی را انجام می‌دادند. بعدها نیز مؤسساتی که این گونه معاملات را انجام می‌دادند به نام بانک معروف گردیدند [۱۱] به نقل از [۱۲].

بانکداری در ایران از قرون وسطی تا اوایل قرن نوزدهم منحصر به فعالیت‌های صرافی بود. صرافی‌های بزرگی در تبریز، مشهد، تهران، اصفهان، شیراز و بوشهر یعنی مراکز تجاری عمده آن دوره وجود داشتند. در این دوره هیچ مؤسسه دولتی یا بانک خارجی در کشور فعالیت نداشت و نقل و انتقال وجوه در داخل یا در خارج توسط صرافان انجام می‌گرفت که فعالیت اکثر آنها تا قبل از سال ۱۳۰۰ هجری شمسی متوقف گردید.

بانکداری نوین در ایران با تأسیس برخی بانک‌های خارجی شکل گرفت، در سال ۱۲۶۷ هجری شمسی یک بانک انگلیسی به نام بانک جدید شرق در بعضی از شهرهای ایران مانند تهران، تبریز، رشت، مشهد، اصفهان، شیراز و بوشهر شعبه‌هایی تأسیس نمود که پس از مدتی تأسیسات آن به بانک شاهنشاهی واگذار گردید [۱۱].

اولین بانک ایرانی در سال ۱۳۰۴ به نام «بانک پهلوی قشون»، از محل وجوه بازنشستگی نظامیان و برای سامان دادن به امور مالی ارتش به وجود آمد. سپس نام آن به «بانک سپه» تغییر کرد. فعالیت‌های بانکی این بانک در آغاز تأسیس محدود بود.

با وجود عدم منع قانونی تأسیس بانک در ایران و با وجود توسعه بانک‌های ملی و سپه و بانک‌های خارجی، هیچ بانک خصوصی با سرمایه ایرانی تا بعد از جنگ جهانی در کشور تأسیس نشد و در نتیجه فعالیت‌های بانکی در سراسر کشور، در دست چند بانک باقی ماند. در تیرماه ۱۳۲۸ دولت به منظور کمک مالی به واحدهای تولیدی خصوصی، «بانک برنامه» را ایجاد کرد. سپس «بانک بازرگانی ایران»

به عنوان اولین بانک خصوصی به صورت شرکت سهامی در بهمن ماه ۱۳۲۸ تاسیس شد و در سال ۱۳۲۹ فعالیت‌های بانکی خود را آغاز کرد [۵].

۲-۷- سیر تحول فناوری اطلاعات در صنعت بانکداری

فناوری بانک‌ها مشتمل بر فناوری پردازش، ثبت، نگهداری، تغذیه و تبادل اطلاعات مشتریان است. این فناوری طی دوره‌های چهارگانه‌های زیر به تکامل رسیده است [۱۱].

۲-۷-۱ دوره اول: اتوماسیون پشت باجه

این نخستین دوره کاربرد کامپیوتر در صنعت بانکداری بود. با استفاده از کامپیوترهای مرکزی، اطلاعات و اسناد کاغذی تولید شده در شعب به صورت بسته بندی شده به مرکز ارسال و پردازش شبانه انجام می‌شود. در این دوره کاربرد اصلی کامپیوتر محدود به ثبت دفاتر و تبدیل کاغذ به فایل‌های کامپیوتری است.

۲-۷-۲ دوره دوم: اتوماسیون جلوی باجه

این دوره از زمانی آغاز می‌شود که کارمند شعبه در حضور مشتری عملیات بانکی را به صورت الکترونیکی ثبت و دنبال می‌کند. از اواخر دهه ۱۹۷۰ امکان انتقال لحظه‌ای از طریق به‌کارگیری ترمینال‌ها در جلوی باجه فراهم آمد. این ترمینال‌ها که به ظاهر شبیه به کامپیوترهای شخصی امروزی بودند، از طریق خطوط مخابراتی به کامپیوترهای بزرگ مرکزی متصل می‌شدند و امکان انتقال اطلاعات به صورت مؤثر در بین شبکه‌های بزرگ کامپیوتری و ترمینال‌های ورودی و خروجی به وجود آمد.

۲-۷-۳ دوره سوم: اتصال مشتریان به حساب‌هایشان

در این دوره که از اواسط دهه ۸۰ آغاز شد امکان دسترسی مشتریان به حساب‌هایشان فراهم می‌گردد. یعنی مشتری از طریق تلفن یا مراجعه به دستگاه خودپرداز و استفاده از کارت هوشمند یا کارت مغناطیسی یا کامپیوتر شخصی به حسابش دسترسی پیدا می‌کند و ضمن انجام عملیات دریافت و پرداخت، نقل و انتقال وجه را به صورت الکترونیکی انجام می‌دهد.

از مهم‌ترین ویژگی‌های دوره سوم که آن را از دوره‌های قبل متمایز می‌سازد عبارتند از: توسعه جزیره ای سی ستم‌های مکانیزه در جلوی باجه و پشت باجه و همچنین توسعه سی ستم‌های ارتباطی مشتریان با حساب‌هایشان مثل ATM و تلفن بانک و فاکس بانک. در دوره سوم هنوز نیروی انسانی در

ارائه خدمات مؤثر است و بخشی از نیروی انسانی وظیفه ایجاد هماهنگی بین سی ستم‌های جزیره ای و نیازهای مختلف مشتریان را به عهده دارد.

می‌توان گفت در این دوره عملیات بانکی نیمه الکترونیکی شده است. دو مشکل اساسی در این دوره به چشم می‌خورد، یکی عدم یکپارچگی سی ستم‌های مکانیزه، جزیره ای بودن آن‌ها و ناهماهنگی آن‌ها برای ارائه خدمات به مشتریان و دیگری عدم وجود و تکامل خطوط مخابراتی مطمئن و پروتکل‌های ارتباطی مربوط به متصل نمودن مشتری به شعب و شعب بانک‌ها به یکدیگر.

۲-۷-۴- دوره چهارم: یکپارچه‌سازی سیستم‌ها و مرتبط کردن مشتریان با تمامی عملیات بانکی

آخرین دوره زمانی آغاز می‌شود که همه نتایج بدست آمده از دوره‌های قبل به طور کامل به سیستم عملیات الکترونیکی انتقال یابند تا هم بانک و هم مشتریان بتوانند به صورت دقیق و منظم اطلاعات مورد نیازشان را کسب نمایند. دوره چهارم به یکی از دو وجه زیر تاکید دارد [۱۲]:

- تلاش برای استاندارد سازی نرم افزاری و سخت افزاری در سی ستم‌های کامپیوتری موجود برای رسیدن به یک سیستم یکپارچه.
- تلاش برای تاسیس سی ستم‌های یکپارچه صرف نظر از سی ستم‌های جزیره‌ای که قبلاً به وجود آمده است.

۲-۷-۵- بانکداری الکترونیک

بانکداری الکترونیک دارای سطوح مختلفی می‌باشد و به تناسب هر کدام می‌توان تعریف خاصی ارائه نمود.

آنچه که در تمامی سطوح می‌توان مشاهده نمود استفاده از سی ستم‌های نرم‌افزار و سخت‌افزار رایانه‌ای می‌باشد. هر چقدر به سطوح بالاتر یعنی بانکداری الکترونیک کامل حرکت نماییم عملیات دستی کمتر، سی ستم‌های رایانه‌ای متمرکزتر، شبکه قابل دسترسی گسترده‌تر، محدودیت زمانی و مکانی کمتر و در نهایت امنیت اطلاعات بانکی بیشتر خواهد بود [۱۱ به نقل از ۱۲].

بانکداری الکترونیک، اساساً به فراهم آوردن امکان دسترسی مشتریان به خدمات بانکی با استفاده از واسطه‌های ایمن و بدون حضور فیزیکی اطلاق می‌شود. بنا بر تعریف دیگری بانکداری الکترونیک به ایجاد محصولات و خدمات با بهای کم، از طریق کانال‌های الکترونیکی اطلاق می‌شود. این محصولات

و خدمات می‌تواند شامل صورتحساب، وام، مدیریت سپرده‌ها، پرداخت‌های الکترونیکی و ایجاد محصولات و خدمات پرداخت‌های الکترونیکی همانند پول الکترونیکی باشند [۱۱ به نقل از ۱۲].

۲-۸- داده‌کاوی

سرعت روزافزون انباشت داده‌های حاصل از پردازش و تعاملات و تبادلات و ارتباطات موجود در فضای کسب‌وکار، به خصوص با وجود پیشرفت سیستم‌های فناوری اطلاعات و ارتباطات^۱، مجموعه وسیعی از داده‌ها را پدید آورده است. طبق آخرین تحقیقات محققان، ظرفیت ذخیره‌سازی داده‌ها در سراسر دنیا هر نه ماه دو برابر و داده‌های در دسترس نیز هر پنج سال دو برابر می‌شوند. همچنین دانشمندان پیش‌بینی کرده‌اند پایگاه‌های داده هر ۵ سال یک بار به طور کامل فراموش شده و داده‌های موجود در آن‌ها فقط در انبارهای داده ذخیره می‌شوند.

هرچند در نگاه اول ممکن است به نظر برسد این داده‌ها بی‌فایده هستند، اما این داده‌ها می‌توانند در صورت وجود ابزارهای مناسب، حکم معادن طلا را برای سازمان‌ها، کسب‌وکارها و شرکت‌ها داشته باشند. با این حال که ما از نظر داده‌ای غنی هستیم اما از نظر دانشی بسیار فقیریم [۳۱]. اکثر سازمان‌ها با فقر دانش مواجه‌اند و می‌توان ادعا نمود این مسئله به سبب عدم شناخت سازمان‌ها از ابزارهایی نظیر داده‌کاوی است که برخلاف ابزارهای گزارش‌گیری و آمارگیری قدیم می‌توانند حجم عظیمی از داده‌ها را تحلیل نموده و الگوهای پنهان مفیدی از دل این داده‌ها آشکار نمایند.

داده‌کاوی به کشف الگوهای پنهان و استخراج دانش و اطلاعات از پایگاه داده‌های بسیار بزرگ اشاره دارد.

داده‌کاوی الگوریتم‌هایی چون دسته‌بندی^۲، خوشه‌بندی^۳، قوانین انجمنی^۴ و ... را بر روی مجموعه‌ای از داده‌ها^۵ اعمال نموده و بعد از تحلیل محتویات داده‌ای مدل‌هایی را به عنوان خروجی تولید می‌نماید. روش‌های آماری اساس کار داده‌کاوی را تشکیل می‌دهند.

۲-۸-۱- مقایسه روش‌های آماری و داده‌کاوی

^۱ Information Communication Technology (ICT)

^۲ Classification

^۳ Clustering

^۴ Association Rules

^۵ Data set

با پیشرفت روزافزون علوم در گذر زمان روش‌های قدیمی موجود جای خود را به روش‌های نوین می‌سپارند تا محدودیت‌های روش‌های گذشته را برطرف نمایند. علم آمار که یکی از شاخه‌های علوم ریاضی است نیز از این امر مستثنی نبوده و دست‌خوش تغییرات بسیاری گشته است. آمار، علمی است که به جمع‌آوری، توضیح و تفسیر داده‌ها می‌پردازد. اشتراک تکنیک‌های آماری و داده‌کاوی بیشتر در تخمین و پیش‌بینی است. همچنین از آزمون‌های آماری در ارزیابی نتایج داده‌کاوی نیز استفاده می‌شود. از این منظر می‌توان روش‌های آماری را اساس کار داده‌کاوی دانست.

داده‌کاوی به صورت یک محصول قابل خریداری نمی‌باشد، بلکه یک رشته علمی و فرآیندی است که باید به صورت یک پروژه پیاده‌سازی شود. داده‌ها اغلب حجیم می‌باشند و به تنهایی قابل استفاده نیستند، بلکه دانش نهفته در داده‌ها قابل استفاده می‌باشد. بنابراین بهره‌گیری از قدرت فرآیند داده‌کاوی جهت شناسایی الگوها و مدل‌ها و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نهفته در داده‌ها و نهایتاً تبدیل داده به اطلاعات، روز به روز ضروری‌تر می‌شود [۱۱].

پایه و اساس داده‌کاوی به دو مقوله آمار و هوش مصنوعی تقسیم شده است، که روش‌های هوش مصنوعی به عنوان روش‌های یادگیری ماشین در نظر گرفته می‌شوند. تفاوت اساسی بین روش‌های آماری و روش‌های یادگیری ماشین بر اساس فرض‌ها و یا طبیعت داده‌هایی که پردازش می‌شوند.

تفاوت‌های موجود بین روش‌های آنالیز آماری و روش‌های داده‌کاوی در جدول ۲-۳ نشان داده شده است.

جدول ۲-۳ مقایسه روش‌های تحلیل آماری و داده‌کاوی [۱۱ به نقل از ۱۴]

روش	تحلیل آماری	داده‌کاوی
فرضیه	آمارشناسان همیشه با یک فرضیه شروع بکار می‌کنند.	داده‌کاوی به فرضیه احتیاجی ندارد.
نوع داده‌ها	از داده‌های عددی استفاده می‌کند.	ابزارهای داده‌کاوی می‌توانند از انواع مختلف داده و نه فقط داده عددی استفاده کنند.
ایجاد روابط	آمارشناسان باید رابطه‌هایی ایجاد کنند که با فرضیه ارتباط دارد.	الگوریتم‌های داده‌کاوی به صورت اتوماتیک روابط را ایجاد می‌کنند.

صحت داده‌ها	آن‌ها می‌توانند داده‌های نابجا و نادرست را در طول تحلیل مشخص کنند.	داده‌کاوی به داده‌های صحیح و درست نیاز دارد.
قابلیت تفسیر	آن‌ها می‌توانند نتایج کار خود را تفسیر و برای مدیران بیان کنند.	نتایج داده‌کاوی نسبتاً پیچیده است و نیاز به متخصصانی جهت تحلیل و بیان نتایج به مدیران دارد.

۲-۸-۲- مفهوم داده‌کاوی

برای داده‌کاوی تعاریف بسیار زیادی ارائه شده است که البته بسیاری از این تعاریف بسیار نزدیک به یکدیگرند. در بعضی منابع داده‌کاوی را با اصطلاحاتی نظیر کشف دانش معادل قرار داده‌اند که باید این‌گونه اصلاح شود، داده‌کاوی یک گام اساسی در فرایند کشف دانش است.

اصطلاحات و عبارات بسیار دیگری نیز معادل با داده‌کاوی بکار رفته‌اند که معانی مشابه و گاهی متفاوت دارند مانند، کاوش دانش از پایگاه داده، استخراج دانش^۱، تحلیل الگوی داده^۲، لایروبی داده^۳ و باستان‌شناسی^۴ [۱۱].

داده‌کاوی یک گام از فرایند کشف دانش از پایگاه داده است و به الگوریتم‌هایی که برای استخراج الگو از داده‌ها بکار می‌رود، گفته می‌شود. اطلاعات حاصل می‌تواند به عنوان ورودی برای تشکیل مدل دسته‌بندی و یا پیش‌بینی، یا بهبود یک مدل موجود بکار رود [۵۳]. داده‌کاوی جستجو و یا کاوش^۵ دانش (الگوهای دلخواه) از مقادیر زیاد داده‌های موجود است [۳۱].

در تعریفی دیگر داده‌کاوی به این صورت تعریف شده است: به‌کارگیری الگوریتم‌های یادگیری ماشین^۶ برای یافتن الگوهای نهفته از ارتباطات موجود در بین عناصر داده‌ای موجود در یک مجموعه داده بسیار بزرگ که می‌تواند منجر به افزایش سود شود [۲۶].

^۱ Knowledge extraction

^۲ Data/Pattern Analysis

^۳ Data Dredging

^۴ Data Archaeology

^۵ Mining

^۶ Machine Learning

داده‌کاوی یک متدولوژی بسیار قوی و با پتانسیل بالا می‌باشد که به سازمان‌ها کمک می‌کند که بر مهم‌ترین اطلاعات از مخزن داده‌های خود تمرکز نمایند [۵۲].

داده‌کاوی به فرایند استخراج اطلاعات نهفته، قابل فهم، قابل تعقیب از پایگاه داده‌های بزرگ و استفاده از آن‌ها در تصمیم‌گیری‌های تجاری مهم اطلاق می‌شود [۱۳].

فرایند داده‌کاوی را می‌توان به عنوان سیستمی تصمیم‌یار تلقی نمود که سازمان و مدیران آن را قادر می‌سازد از داده‌های خام به گنجینه‌های دانشی دست یابند و از این دانش در مسیر اتخاذ استراتژی‌های سازمانی بهره گیرند.

اگر هرم داده‌ها را به صورت زیر در نظر بگیریم:



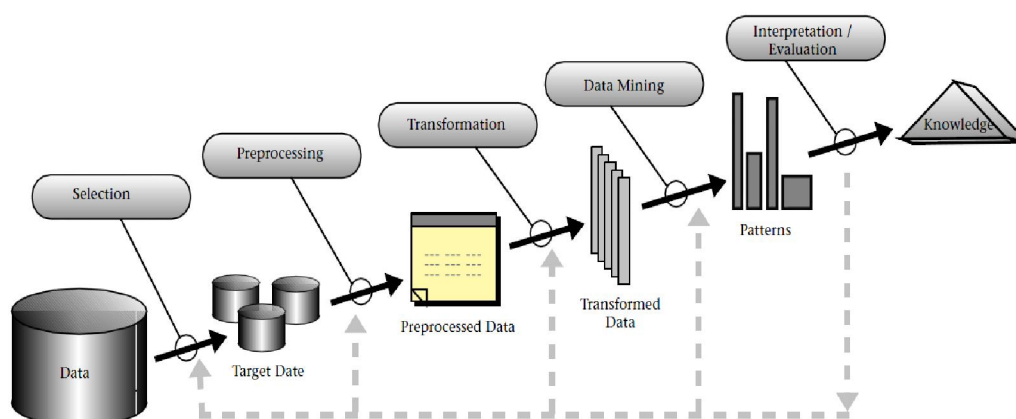
شکل ۲-۶ سیر تکامل در هرم دانش [۱۱]

با آگاهی از تعریف داده‌کاوی و نگاهی به هرم بالا بهتر می‌توان به نقش داده‌کاوی در سازمان‌ها پی برد. داده‌کاوی سبب می‌شود که سازمان‌ها از سطح داده به سطوح بالاتر دانش و الگوهای ناشناخته برسند. الگوهای استخراج شده می‌توانند رابطه‌ای بین ویژگی‌ها و مشخصات سیستم مانند نوع تقاضا و نوع مشتری، پیش‌بینی‌های آینده بر اساس مشخصات سیستم، قوانین (اگر-آنگاه) بین متغیرهای سیستم، دسته‌بندی‌ها و خوشه‌بندی‌های اشیاء و رکوردهای شبیه به هم در یک سیستم و غیره باشند [۱۱] به نقل از [۵۲].

۲-۸-۳- داده کاوی و کشف دانش

داده کاوی را می توان به عنوان ابزاری جهت کشف دانش از پایگاه های داده تعریف نمود. داده کاوی یک گام در فرایند کشف دانش از پایگاه داده^۱ است که با استفاده از الگوریتم های کشف و تحلیل داده ها تعداد خاصی الگو یا مدل را از روی داده ها تولید می کند.

الگوی زیر مراحل استخراج دانش مشتری به کمک داده کاوی در سازمان را نمایش می دهد.



شکل ۲-۷ گام های فرایند تولید دانش از پایگاه داده ها [۵۳ و ۳۱]

این الگو گام های فرایند کشف دانش از پایگاه داده را در [۵۳] به صورت زیر نشان داده است: چنانچه از نمایش الگوی فوق برداشت می شود، ابتدا باید شناسایی دامنه فعالیت های کسب و کار صورت گرفته و دسترسی به داده ها مقدور گردد. داده ها در اینجا می تواند مجموعه ای از حقایق، الگوها و یا خبرگی باشد.

در مرحله بعد با توجه به اهداف تعیین شده مورد نظر ذینفعان سازمان (مدیران، کارمندان، سرپرستان شعب، کاربران، خبرگان و مشاورین فناوری اطلاعات)، رکوردهای مورد نیاز توسط داده کاو از میان حجم عظیم داده های سازمان مورد نظر انتخاب می شوند. در این مرحله حجم انبوه داده ها به داده های هدف کاهش داده می شود.

^۱ Knowledge Discovery in Databases (KDD)

برای آماده‌سازی بخش داده‌های انتخابی باید عملیات پیش‌پردازشی روی آن‌ها صورت گیرد. این عملیات شامل جمع‌آوری اطلاعات لازم برای مقداردهی به مقادیر رکوردهای از قلم افتاده، تشخیص داده‌های متفاوت و هماهنگی آن‌ها با بقیه، جمع‌آوری اطلاعات لازم برای اصلاح یا حذف داده‌های نامناسب^۱ می‌باشد. محصول این مرحله داده‌های پیش‌پردازش شده خواهد بود.

گام بعد گزینش داده‌ها با هدف کاهش حجم آن‌ها و یافتن ویژگی‌هایی از داده‌ها با توجه به هدف تعیین شده برای کسب‌وکار خواهد بود. با کاهش حجم داده‌ها یا روش‌های تغییرات آن‌ها حجم داده‌ها می‌تواند به مقدار مفید و مؤثری کاهش یابد.

مرحله بعد اعمال یک یا چند الگوریتم (مانند کلاسیفایرها^۲ و یا شبکه‌های عصبی^۳ و ...) از بین الگوریتم‌های موجود و یا ابتکاری داده‌کاوی بر روی داده‌هاست. حاصل این مرحله یک یا چند مدل و یا الگو خواهد بود.

در گام بعد با تحلیل، ارزیابی و تفسیر الگوهای معنادار کشف شده، دانش استخراج می‌گردد. دانش استخراج شده می‌تواند به صورت قواعدی مستند شود.

یک الگو زمانی می‌تواند به عنوان دانش تلقی شود که :

۵. به سادگی برای افراد قابل فهم باشد.

۶. اعتبار آن از یک حد آستانه‌ای^۴ پایین‌تر نباشد.

۷. دانش جدیدی متناسب با اهداف تعیین شده سازمان ارائه دهد.

۸. کاربردی باشد [۱].

۲-۸-۴- فرایند داده‌کاوی

داده‌کاوی، فرایندی خلاقانه است که دانش و مهارت‌های گوناگونی را می‌طلبد. وجود یک استاندارد جهت انجام این فرایند به تبدیل مشکلات کسب‌وکار به مسائل داده‌کاوی، گزینش روش‌های

^۱ noise

^۲ Classifier

^۳ Neural Network

^۴ Threshold

داده‌کاوی و تبدیل داده‌ها متناسب با نوع کسب‌وکار، معنا دادن به ارزیابی اثربخشی و مستندسازی نتایج اهمیت فراوان دارد.

چرخه ^۱CRISP-DM فرایند استاندارد^۲ است که با هدف ارائه چارچوبی برای اجرای پروژه‌های داده‌کاوی بزرگ، یا هزینه کمتر، قابلیت اطمینان بیشتر، قابلیت مدیریت بیشتر و سریع‌تر توسعه یافت [۵۰].

مدل مرجع CRISP-DM نمای کلی از چرخه عمر^۳ یک پروژه داده‌کاوی را نشان می‌دهد. این مدل شامل فازهای یک پروژه، کارهای لازم برای انجام هر فاز و خروجی‌های هر فاز است.

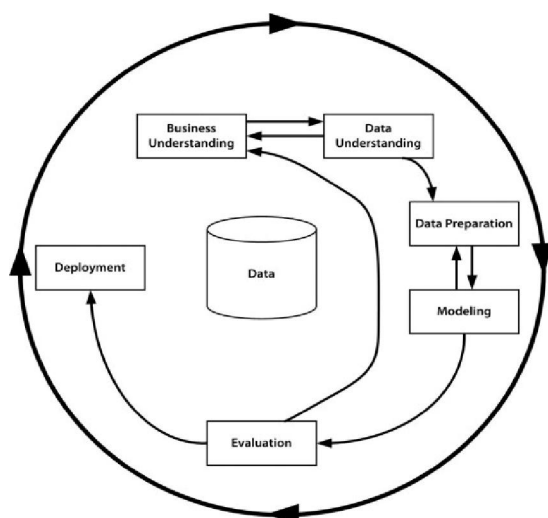
چرخه عمر یک پروژه داده‌کاوی چنانچه در شکل ۲-۸ نشان داده شده است به شش فاز شکسته می‌شود. توالی فازها اجباری نیست بلکه این ترتیب مرسوم‌ترین، مهم‌ترین و پر تکرارترین وابستگی بین فازها را نشان می‌دهد. در مورد پروژه‌های خاص این ترتیب بستگی به این دارد که خروجی کدام فاز و یا عمل خاص نیازمند اجرای فاز بعدی است [۵۰].

حلقه بیرونی مربوط به طبیعت پروژه‌های داده‌کاوی است. بدین معنی که پروژه داده‌کاوی با تولید یک راهکار خاص در زمان مشخص پایان نمی‌یابد بلکه با اجرای یک دور فرایند داده‌کاوی در سازمان و اجرای آن راهکار در سازمان، ممکن است نتایجی حاصل شود که سؤالات جدیدی از کسب‌وکار را برانگیزد و این نیازمند تکرار دو و یا چند باره چرخه خواهد بود.

^۱ Cross Industry Standard Process for Data Mining

^۲ این مدل فرایند توسط یک کنسرسیوم از کاربران و تأمین کنندگان مانند AG، SPSS، NCR و OHRA توسعه داده شد و تاحدودی تحت حمایت کمیسیون اروپا (ESPRIT) بوده است.

^۳ Life Cycle



شکل ۸-۲ متدولوژی فرآیند استاندارد میان صنعتی داده‌کاوی (CRISP-DM) [۵۰]

مدل فرایند CRISP-DM شامل فازهای زیر می‌باشد:

- گام درک موقعیت کسب‌وکار^۱
- گام درک داده‌ها^۲
- گام پیش‌پردازش داده‌ها^۳
- گام مدل‌سازی^۴
- گام ارزیابی^۵
- گام به‌کارگیری^۶

در ادامه هر یک از این گام‌ها به طور خلاصه توضیح داده می‌شود.

درک موقعیت کسب و کار: این گام، بخش آغازین مدل CRISP-DM می‌باشد که بر روی درک و تعیین اهداف^۷ و نیازمندی‌ها از منظر کسب‌وکار و تبدیل این دانش به تعریف یک مسئله داده‌کاوی و طراحی طرح اولیه پروژه برای دستیابی به این اهداف تمرکز دارد. در این گام باید متخصص داده‌کاو از

^۱ Business understanding

^۲ Data Understanding

^۳ Data Preparation

^۴ Modeling

^۵ Evaluation

^۶ Deployment

^۷ objectives

توان و تجربه متخصص کسب و کار بهره‌مند گردد [۵۰]. البته در یک پروژه موفق داده‌کاوی این مشارکت در گام اولیه متوقف نخواهد شد، بلکه در سراسر فرآیند داده‌کاوی ادامه خواهد داشت [۵۳].

درک داده‌ها: گام درک داده‌ها با جمع‌آوری داده‌های اولیه آغاز می‌شود و با فعالیت‌هایی جهت آشنایی با داده‌ها، شناسایی مشکلات موجود در کیفیت داده‌ها، اکتشاف داده‌ها و یا تشخیص زیرمجموعه‌های جالب جهت تشکیل فرضیه‌ها برای اطلاعات مخفی پیگیری می‌شود. ارتباط نزدیکی بین دو گام درک موقعیت کسب و کار و درک داده‌ها وجود دارد. ساختار مسئله داده‌کاوی و طرح پروژه به درک حداقلی از داده‌های در دسترس نیازمند است [۵۰].

در کل می‌توان گفت گام درک داده‌ها بر اساس سه گام جمع‌آوری داده‌ها، به‌کارگیری تحلیل اکتشافی برای دستیابی به یک دید اولیه و ارزیابی کیفیت داده‌ها استوار است [۱۱].

پیش‌پردازش داده‌ها: فاز آماده‌سازی داده‌ها کلیه فعالیت‌های مربوط به ساخت مجموعه داده^۱ (داده-ای که می‌تواند به عنوان ورودی ابزار مدل‌سازی مورد استفاده قرار گیرد) از داده‌های خام اولیه را در بر می‌گیرد. پس از گردآوری داده‌ها باید خطاهای احتمالی موجود در آن‌ها را از بین برده و داده‌ها را تمیز نمود [۳۱]. فعالیت‌های مربوط به آماده‌سازی داده‌ها شامل ساخت جدول، ثبت داده‌ها و انتخاب ویژگی^۲، پاک‌سازی داده‌ها^۳، تشکیل ویژگی‌های جدید و تغییر شکل داده‌ها متناسب با ابزارهای مدل‌سازی می-باشد [۵۰].

اهمیت پیش‌پردازش داده‌ها به این دلیل است که اکثر داده‌هایی که در پایگاه‌های داده‌ای وجود دارند و پیش‌پردازش نشده‌اند، ناقص و نویزدار هستند و دارای مشکلات احتمالی زیر می‌باشند:

مقادیر دور افتاده^۴ یا حدی، مقادیر گم شده^۵، صفات تکراری^۶، داده‌هایی که در فرم مناسب برای مدل‌سازی نیستند و داده‌هایی که با عقل سلیم جور در نمی‌آیند. برای این منظور در داده‌کاوی روش‌های مختلف تمیز کردن و تغییر شکل داده‌ها ارائه گردیده است [۱۱].

^۱ Dataset

^۲ attribute

^۳ Data cleaning

^۴ Outliers

^۵ Missing Values

^۶ Redundant Attribute

مدل‌سازی: در این گام، روش‌های متفاوت مدل‌سازی انتخاب و بکار گرفته می‌شود، و پارامترهای آن‌ها به صورت مقادیر بهینه تنظیم می‌شود. بعضی از روش‌ها ساختار داده‌ای خاصی را می‌طلبند. بین فازهای آماده‌سازی و مدل‌سازی داده‌ها ارتباط نزدیکی برقرار است. اغلب مشکلات داده‌ها حین مدل‌سازی درک می‌شود و یا ایده‌هایی برای ساخت داده‌های جدید به ذهن می‌رسد [۵۰].

این گام شامل بخش‌های زیر است [۱۱]:

- انتخاب و استفاده از تکنیک مدل‌سازی مناسب
- دست‌کاری و تنظیم مدل برای دستیابی به نتایج بهینه
- در صورت نیاز برگشت به گام پیش‌پردازش

ارزیابی: در پروژه‌ای که بر اساس یک یا چند مدل ساخته شده است، پیش از رسیدن به گام به‌کارگیری، مدل‌ها باید به طور دقیق ارزیابی شوند و گام‌های اجرایی ساخت مدل‌ها بازبینی شوند تا از حصول اهداف کسب و کار اطمینان حاصل شود [۵۰]. مدل‌های داده‌کاوی باید به فرآیند تصمیم‌گیری کمک کنند [۳۱]. پس مدل زمانی مفید است که تفسیر پذیر باشد زیرا انسان‌ها مایل به استفاده از اصول پیچیده در فرایند تصمیم‌گیری جعبه سیاه مانند خود نیستند [۱۱].

به‌کارگیری: عموماً ساخت مدل پایان کار پروژه نیست. معمولاً دانش حاصل از این چرخه باید به صورتی سازماندهی و ارائه شود که مشتری نهایی بتواند از آن استفاده نماید. بسته به نیازمندی‌ها، فاز به‌کارگیری می‌تواند به سادگی یک گزارش و یا به پیچیدگی اجرای یک فرایند داده‌کاوی تکرارپذیر^۱ باشد. در بسیاری از موارد کاربر یک تحلیلگر داده نیست که گام‌های استقرار را درک نماید. بنابراین نکات لازم باید برای او توضیح و تبیین شود [۵۰].

هر کدام از گام‌های مدل فرایند CRISP-DM که پیش از این تفسیر شد شامل کارهای^۲ مختلفی است که کارهای مربوط به هر گام و خروجی آن به طور خلاصه در جدول ۲-۴ آمده است.

^۱ Repeatable

^۲ Tasks

جدول ۲-۴ فعالیت‌های مربوط به فازهای CRISP-DM و خروجی هر فعالیت [۵۰]

درک کسب-وکار	درک داده‌ها	آماده‌سازی داده‌ها	مدل‌سازی	ارزیابی	به‌کارگیری
- تعیین اهداف کسب‌وکار درک اهداف کسب و کار و معیارهای موفقیت^۱ آن - ارزیابی وضعیت موجودی منابع، نیازمندی‌ها، فرضیات، محدودیت‌ها، ریسک‌ها و احتمالات، مجموعه اصطلاحات^۲ و هزینه‌ها و فایده‌ها - تعیین اهداف داده‌کاوی اهداف داده‌کاوی، عوامل موفقیت داده‌کاوی - تولید طرح پروژه طرح پروژه تخمین اولیه ابزارها و تکنیک‌ها	- جمع‌آوری داده‌های اولیه گزارش جمع‌آوری داده‌های اولیه - توصیف داده‌ها گزارش توصیف داده‌ها - مرور^۳ داده‌ها گزارش مرور داده‌ها - ارزیابی کیفیت داده‌ها گزارش کیفیت داده‌ها	- جمع‌آوری داده‌ها توصیف مجموعه داده - انتخاب داده منطق انتخاب/حذف^۴ داده‌ها - پاک‌سازی داده‌ها گزارش پاک‌سازی داده‌ها - ساختاردهی داده‌ها تفکیک ویژگی‌ها تولید رکوردها - یکپارچگی داده‌ها داده‌های یکپارچه - سازماندهی داده‌ها داده‌های مجدداً سازماندهی شده	- انتخاب روش مدل‌سازی روش مدل‌سازی فرضیات مدل‌سازی - تولید تست طرح تست - ساخت مدل تنظیمات پارامترهای مدل شرح مدل - ارزیابی مدل ارزیابی مدل پارامترهای اصلاح شده	- ارزیابی نتایج ارزیابی داده‌ها کاوش نتایج بر مبنای معیارهای موفقیت کسب-وکار مدل‌های بهبود یافته - بازیابی فرایند بازیابی فرایند - تعیین اقدامات بعدی لیستی از اقدامات و تصمیمات ممکن	- به‌کارگیری طرح طرح توسعه نظارت و نگهداری طرح طرح نگهداری و نظارت - گزارش نهایی پروژه گزارش نهایی ارائه نهایی - بازیابی پروژه مستندات تجربی

^۱ Success Criteria

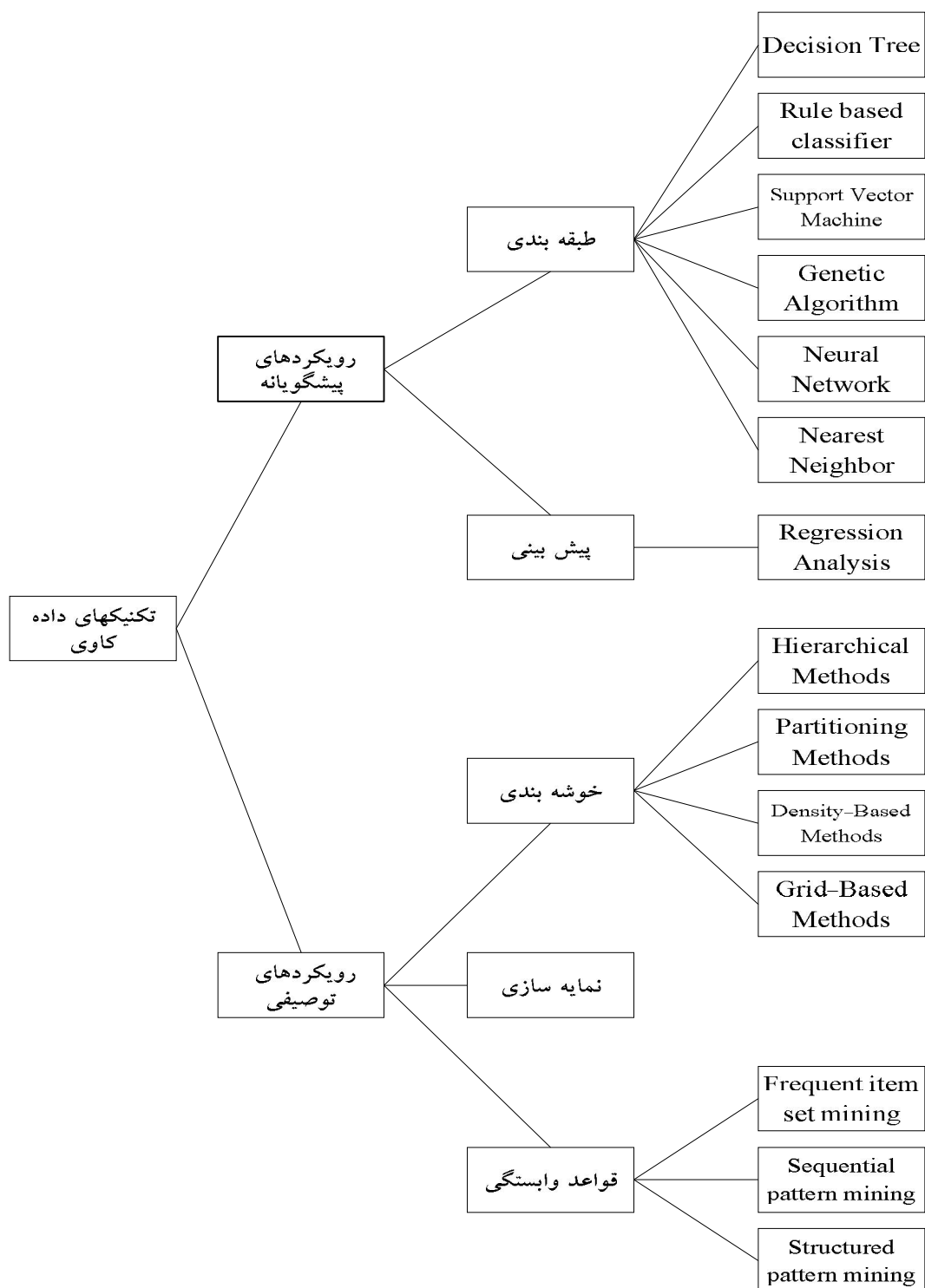
^۲ Terminology

^۳ Explore

^۴ Inclusion/Exclusion

۲-۸-۵- معرفی روش‌های داده‌کاوی

روش‌های داده‌کاوی در یک تقسیم‌بندی کلی به دو دسته روش‌های توصیفی و روش‌های پیش‌بینانه تقسیم می‌شوند. روش‌های توصیفی به خواص عمومی داده‌ها می‌پردازند. بدین ترتیب که الگوهایی قابل تفسیر توسط انسان از داده‌ها استخراج می‌کنند. در رویکرد پیش‌بینانه هدف پیش‌بینی رفتارهای آینده است. در این روش‌ها از چند متغیر جهت پیش‌بینی مقادیر آینده استفاده می‌شود. تقسیم‌بندی برخی روش‌های داده‌کاوی در شکل ۲-۹ نشان داده شده است.



شکل ۹-۲ دسته‌بندی کلی عملکردهای داده کاوی

در ادامه برخی از این روش‌ها به طور مختصر معرفی خواهند شد.

۲-۸-۵-۱- دسته‌بندی^۱

این روش از رایج‌ترین، محبوب‌ترین و قابل درک‌ترین روش‌های داده‌کاوی می‌باشد. در این روش هر یک از نمونه‌ها به یکی از گروه‌ها یا دسته‌ها تخصیص می‌یابد. در روش دسته‌بندی هر نمونه با تعدادی خصوصیت^۲ که یکی از این خصوصیت‌ها صفت کلاس مربوط به آن نمونه است، شناخته می‌شود. برای دسته‌بندی نمونه‌ها جهت استخراج مدل، وجود صفت کلاس برای هر نمونه الزامی است. هدف از دسته‌بندی، یافتن مدلی بر اساس خصوصیت قابل پیش‌بینی یا کلاس هر نمونه به عنوان تابعی از سایر ویژگی‌های آن نمونه می‌باشد [۱]. این روش به انتساب نمونه‌ها به دسته‌ها بر اساس صفتی قابل پیش‌بینی اشاره دارد [۲].

مطابق شکل ۲-۹ از الگوریتم‌های شاخص این روش می‌توان درخت تصمیم‌گیری^۳، شبکه‌های عصبی^۴ و ناوی بیز^۵ را نام برد. این الگوریتم‌ها معمولاً به یک مقدار هدف نیاز دارند تا هر داده را به یکی از مجموعه کلاس‌های از پیش تعریف شده منتسب کنند. به چنین الگوریتم‌هایی که برای یادگیری نیاز به مقدار هدف دارند الگوریتم‌های با ناظر^۶ گویند.

دسته‌بندی در مسائل تجاری مانند مدیریت ریسک، تبلیغات هدفمند^۷ و تحلیل روی‌گردانی^۸ مشتری کاربرد دارد.

۲-۸-۵-۲- درخت تصمیم

مهم‌ترین مدلی که در این رویکرد استفاده می‌شود، درخت تصمیم‌گیری می‌باشد. در این روش درختی ساخته می‌شود و در هر گره از آن آزمونی بر روی یک از ویژگی‌ها انجام می‌شود و داده با

^۱ Classification

^۲ Attribute or variable, field, characteristic, feature, or observation

^۳ Decision tree

^۴ Neural Network

^۵ Naïve Bayes

^۶ Supervised Algorithm

^۷ Ad targeting

^۸ Churn Analysis

توجه به مقدار مشخصه-های خود در راستای یکی از فرزندان گره هدایت می‌شود، تا جایی که داده به برگ برسد. هر برگ نشان دهنده یک دسته می‌باشد [۱۱].

جدول ۵-۲ نمونه‌ای از یک مسئله دسته‌بندی را نشان می‌دهد. X_1 تا X_m ویژگی‌هایی هستند که به کمک آن‌ها کلاس (C_1 یا C_2) هر یک از نمونه‌های U_1 تا U_n مشخص می‌شود. a_{ij} مقدار مربوط به ویژگی j از نمونه i است.

جدول ۵-۲ نمونه داده‌های مورد نیاز در یک مسئله مدل‌سازی به روش دسته‌بندی [۱]

	X_1	X_2	. . .	X_m	Y
U_1	a_{11}	a_{12}	. . .	a_{1m}	C_2
U_2	a_{21}	a_{22}	. . .	a_{2m}	C_1
.	.	.	. . .	.	.
.	.	.	. . .	.	.
.	.	.	. . .	.	.
U_n	a_{n1}	a_{n2}	. . .	a_{nm}	C_2

از دلایل استفاده از درخت تصمیم می‌توان به موارد زیر اشاره نمود:

۱. پس از تشکیل درخت تصمیم پیچیدگی محاسباتی به صورت لگاریتمی شده و کاهش می‌یابد^۱.
 ۲. می‌توان قواعدی به صورت اگر-آنگاه از درخت تصمیم استخراج و از آن به صورت جداگانه استفاده نمود [۱۱]. درخت تصمیم برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد و نتایج بدست آمده از درخت در قالب یک سری قوانین توضیح داده می‌شود.
 ۳. در درخت تصمیم برخلاف شبکه‌های عصبی لزومی ندارد داده‌ها به صورت عددی باشند.
- از ویژگی‌های درخت تصمیم می‌توان به عدم حذف داده‌ها در زمان دسته‌بندی، استفاده آسان، درک ساده مدل حاصل با وجود پیچیدگی کار الگوریتم ساخت درخت تصمیم و ... اشاره نمود [۳۱].

^۱ این الگوریتم از مرتبه $n \log n$ است.

۲-۸-۵-۳- شبکه‌های عصبی

شبکه عصبی مصنوعی یک سیستم پردازش اطلاعات است که ویژگی‌های عملکردی مشابه شبکه عصبی بیولوژیکی دارد. شبکه‌های عصبی مصنوعی به عنوان مدل‌های ریاضی تعمیم یافته از نرون‌های بیولوژیکی یا شناخت انسان بر اساس مفروضات زیر توسعه یافته است:

- پردازش اطلاعات در واحدهای بسیار ساده‌ای که نرون نامیده می‌شوند رخ می‌دهد.
 - سیگنال‌های نرون‌ها از طریق اتصالات انتقال می‌یابد.
 - هر ارتباط یک وزن دارد که در یک شبکه عصبی متداول، در سیگنال ورودی ضرب می‌شود.
 - هر نرون یک تابع فعال سازی دارد (که عمدتاً غیر خطی است) که ورودی شبکه جمع و سیگنال‌های ورودی وزن را به خروجی تبدیل می‌کند.
- می‌توان هر شبکه عصبی را با ویژگی‌های زیر تشریح نمود:
- الگوی ارتباطات بین وزن‌ها (که معماری شبکه نامیده می‌شود)
 - روش تعیین وزن‌ها در ارتباطات (که یادگیری یا آموزش نامیده می‌شود)
 - تابع فعال سازی

عمدتاً شبکه‌های مختلف در جنبه‌های فوق متفاوت هستند و بدین طریق از هم متمایز می‌شوند.

هر شبکه عصبی از تعداد زیادی واحد ساده پردازش کننده که نرون‌ها، واحدها، سلول‌ها و یا گره‌ها نامیده می‌شوند، تشکیل شده است. هر نرون با نرون‌های دیگر به وسیله اتصالات مستقیم مرتبط است که هر اتصال با یک وزن نمایندگی می‌شود. در واقع وزن‌ها اطلاعاتی را که برای حل مسئله مورد استفاده قرار می‌گیرد بازنمایی می‌کند.

هر نرون یک وضعیت درونی دارد که فعال‌سازی^۲ یا سطح فعالیت^۳ نامیده می‌شود و تابعی از ورودی‌هایی است که دریافت می‌کند. هر نرون فعالیت خود را به عنوان یک سیگنال به سایر نرون‌ها

^۱ Neuron

^۲ activation

^۳ activity level

ارسال می‌کند. البته لازم به توجه است که هر نرون تنها یک سیگنال در هر لحظه از زمان قادر است ارسال کند [۱۱].

۲-۸-۵-۴- پیش‌بینی^۱

این تکنیک از روش‌های مهم و قدیمی داده‌کاوی است که می‌تواند جهت پاسخگویی به مسائلی از قبیل میزان جذب سرمایه‌گذاری ماه آینده بانک، پیش‌بینی ارزش سهام در روز آینده و ... مورد استفاده قرار گیرد.

در پیش‌بینی مجموعه‌ای از اطلاعات به عنوان ورودی به الگوریتم داده می‌شود. خروجی یک معادله ریاضی متناسب با این ورودی خواهد بود. برای پیش‌بینی یک متغیر در آینده داده‌ها را به فرمول ایجاد شده می‌دهیم و حاصل پیش‌بینی مورد نظر خواهد بود.

تفاوت این روش با روش‌های دسته‌بندی در این است که در روش دسته‌بندی متغیر هدف، مقادیری گسسته دارد اما در مورد روش پیش‌بینی مقادیر هدف پیوسته‌اند [۱].

از مدل‌های رایج در زمینه پیش‌بینی می‌توان به رگرسیون^۲ اشاره نمود. رگرسیون خطی و منطقی از پرکاربردترین روش‌های رگرسیون‌اند. رگرسیون خطی رویکردی برای مدل‌سازی رابطه بین متغیرهای اسکار Y و متغیرهای پیش‌بینی کننده X است.

امروزه برای مدل‌سازی روابط غیرخطی و پیچیده بین متغیرهای ورودی خروجی از روش‌هایی مانند شبکه عصبی استفاده می‌شود.

۲-۸-۵-۵- خوشه‌بندی^۳

خوشه‌بندی یا گروه‌بندی، تقسیم اقلام موجود در یک مجموعه داده است که به طور طبیعی با هم شباهت دارند. داده‌هایی که با این معیار به صورت خوشه‌هایی تفکیک می‌گردند، با داده‌های موجود در خوشه‌ای که در آن قرار می‌گیرند، بیشترین شباهت را دارند؛ و با داده‌های موجود در سایر خوشه‌ها متفاوت‌اند.

^۱ Prediction

^۲ Regression

^۳ Clustering

در خوشه‌بندی موضوعات زیر مورد توجه است:

- چه تعداد از خوشه‌ها می‌تواند دانش نهفته در داده‌ها را کشف نماید؟ مسئله تعداد خوشه‌ها معمولاً به صورت جداگانه مورد بررسی قرار می‌گیرد.
- معیارهای شباهت و تفاوت داده‌ها چیست؟ این معیارها خود به واسطه روش‌های مختلفی محاسبه می‌گردد، اما در بیشتر روش‌های خوشه‌بندی موجود از معیار فاصله فضایی دو داده از یکدیگر، استفاده می‌شود. فاصله فضایی می‌تواند با روش‌های مختلفی چون فاصله اقلیدسی^۱، فاصله مینکوفسکی^۲ و یا فاصله مانهاتان^۳ محاسبه شود.
- بعد از تعیین تعداد خوشه‌ها و معیار شباهت یا فاصله داده‌ها، داده‌ها با استفاده از چه روشی در تعداد خوشه‌های معین جای گیرند [۱۱].

خوشه‌بندی یک الگوریتم بدون ناظر^۴ در داده‌کاوی است، زیرا هیچ صفتی منفردی برای هدایت فرایند Training استفاده نمی‌شود و همه صفات ورودی ارزش یکسان دارند [۲].

تفاوت این روش با طبقه‌بندی در این است که در طبقه‌بندی هر داده به یک طبقه (کلاس) از پیشین مشخص شده تخصیص می‌یابد ولی در خوشه‌بندی هیچ اطلاعی از کلاس‌های موجود درون داده‌ها وجود ندارد و به عبارتی خود خوشه‌ها نیز از داده‌ها استخراج می‌شوند.

به طور خاص در صنعت بانکداری از الگوریتم‌های خوشه‌بندی در بخش‌بندی مشتریان به گروه‌های مشخص بر طبق الگوهای رفتاری آن‌ها کاربرد دارد. شناسایی گروه‌های مشتریان بسیار برای بانک حائز اهمیت بوده و می‌تواند در تصمیم‌گیری‌های کنونی و اتخاذ راهبردهای میان‌مدت و بلندمدت بانک راهگشا باشد.

۲-۸-۵-۵- انواع خوشه‌بندی

امروزه الگوریتم‌های متنوعی در زمینه خوشه‌بندی معرفی شده‌اند. این الگوریتم‌ها به طور کلی به سه دسته تقسیم می‌شوند [۱]:

^۱ Euclidean distance

^۲ Minkowski

^۳ Manhattan distance

^۴ Unsupervised

۱- خوشه‌بندی سلسله‌مراتبی^۱

در این نوع خوشه‌بندی خوشه‌ها به صورت متوالی به دو شیوه انباشتی^۲ و یا تقسیم‌کنندگی^۳ توسعه می‌یابند. در روش انباشتی، هر یک از نقاط به عنوان خوشه در نظر گرفته شده و سپس خوشه‌های مشابه باهم ادغام می‌گردند. در شیوه تقسیم‌کنندگی در ابتدا کل داده‌ها به عنوان یک خوشه در نظر گرفته شده و تقسیمات متوالی تا رسیدن به تعداد مناسب خوشه‌ها ادامه می‌یابد.

۲- خوشه‌بندی مبتنی بر مدل^۴

در این روش یک مدل احتمالی مشخص برای داده‌ها در نظر گرفته شده و سپس پارامترها برآورد می‌شوند. در این گروه از الگوریتم‌ها یک مدل چگالی آمیخته مطرح می‌شود و فرض می‌شود که داده‌ها از مخلوط شدن تعدادی منبع داده به وجود آمده‌اند. هر یک از این منابع یک خوشه بالقوه در نظر گرفته می‌شود.

۳- خوشه‌بندی مبتنی بر بخش‌بندی^۵

نام دیگر این روش خوشه‌بندی بر مبنای تابع هدف است که در آن اساس کار فرمول‌بندی تابع هدف می‌باشد. تابع هدف حاصل باید طبیعت مسئله را به خوبی نشان دهد تا بتوان از طریق کمینه‌سازی آن، ساختار معنی‌داری (خوشه‌ها) را در داده‌های مفروض آشکار ساخت. معروف‌ترین و ساده‌ترین الگوریتم خوشه‌بندی مبتنی بر بخش‌بندی، الگوریتم K-means است. این الگوریتم به خاطر سادگی اجرا، سادگی برنامه و کارایی آن استفاده فراوانی دارد [۱]. عملکرد کلی این روش به این صورت است که هدف ایجاد K خوشه است، بدین ترتیب که عناصر درون هر خوشه نسبت به میانگین رکوردهای آن خوشه که مرکز نامیده می‌شود بیشترین شباهت و با مراکز دیگر خوشه‌ها بیشترین عدم شباهت را داشته باشند.

^۱ Hierarchical Clustering

^۲ Divisive

^۳ Agglomerative

^۴ Model-based Clustering

^۵ Partitioning

۲-۸-۵-۵-۲- معیارهای ارزیابی در خوشه‌بندی

تکنیک‌های خوشه‌بندی برای تخصیص هر داده به خوشه مربوطه بر اساس یکی از معیارهای شباهت^۱ و یا عدم شباهت^۲ (یا فاصله) عمل می‌کنند. داده، در صورتی که بیشترین شباهت را با داده‌های یک خوشه و یا بیشترین فاصله را با خوشه‌های دیگر داشته باشد، به خوشه متناسب می‌شود.

زمانی که مشاهدات به گونه‌ای خوشه‌بندی شوند که هر خوشه در حداکثر تراکم باشد و حداکثر فاصله را با دیگر خوشه‌ها داشته باشند، خوشه‌بندی خوبی انجام گرفته است [۱۱].

معیار شباهت، معیاری عددی است که میزان شباهت دو شیء داده را نسبت به یکدیگر نشان می‌دهد. هر چقدر دو شیء به یکدیگر شبیه‌تر باشند این معیار مقدار بالاتری دارد و معمولاً مقداری بین ۰ و ۱ دارد. معیارهای ارزیابی شباهت در جدول ۲-۶ نشان داده شده است.

n: تعداد ابعاد (ویژگی‌های) ورودی، p و q: اشیاء داده ورودی، pk و qk: kامین ویژگی یا شیء داده ورودی می‌باشند.

جدول ۲-۶ معیارهای محاسبه شباهت در خوشه‌بندی

نام معیار	فرمول محاسبه
کسین ^۳	$\frac{\cos(d_1 \cdot d_2)}{\ d_1\ \ d_2\ }$ d _۱ و d _۲ دو بردار متن می‌باشند
جاکارد ^۴	$T(p, q) = \frac{p \cdot q}{\ p\ ^2 + \ q\ ^2 - p \cdot q}$

معیار فاصله نیز مقداری عددی است که میزان تفاوت دو شیء داده را نشان می‌دهد. هر چقدر دو شیء داده به یکدیگر شبیه‌تر باشند این معیار مقدار پایین‌تری خواهد داشت. معیارهای ارزیابی فاصله در جدول ۲-۷ ارائه شده است.

^۱ Similarity

^۲ Dissimilarity

^۳ Cosine

^۴ Jaccard

جدول ۷-۲ معیارهای محاسبه فاصله در خوشه‌بندی

نام معیار	فرمول محاسبه
فاصله اقلیدسی	$dist = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$
مینکوسکی	$dist = \left(\sum_{k=1}^n p_k - q_k ^r \right)^{\frac{1}{r}}$
سیتی بلاک ^۱ یا مانهاتان	$dist = \left(\sum_{k=1}^n p_k - q_k ^r \right)^{\frac{1}{r}} \quad \text{if } r=1$
ماهالانوبیس ^۲	$mahalanobis(p, q) = (p - q) \sum^{-1} (p - q)^T$

۲-۸-۵-۶- تحلیل انحراف^۳

این روش جهت یافتن داده‌هایی که بسیار متفاوت از داده‌های دیگر عمل می‌کنند یا تغییرات چشمگیری نسبت به رفتار مشاهده شده قبلی دارند. این داده‌ها، داده‌ها پرت^۴ نام دارند. از کاربردهای این روش می‌توان به کشف تقلب در کارت‌های اعتباری اشاره نمود جایی که شناسایی رفتار غیر نرمال از بین میلیون‌ها تراکنش کاری دشوار و تا حدودی ناممکن می‌نماید. کشف نفوذ در شبکه و تشخیص کلاهبرداری در معاملات از دیگر کاربردهای این روش می‌باشد. این الگوریتم در ساده‌ترین حالت با استفاده از سطح زیر منحنی نرمال و در نظر گرفتن مقدار معینی از انحراف معیار نمونه‌های متفاوت با سایرین را شناسایی می‌نماید [۱].

یک سوءاستفاده مالی در بانک می‌تواند مدیران بانک را متحمل هزینه‌های زیادی نماید. با به‌کارگیری این الگوریتم می‌توان به کشف رفتارهای مشکوک مشتریان پرداخت و در نتیجه هزینه‌های ناشی از این‌گونه سوءاستفاده‌ها را کاهش داد.

^۱ City block

^۲ Mahalanobis

^۳ Deviation Analysis

^۴ Outlier

۲-۸-۵-۷- قواعد وابستگی (انجمنی)^۱

این روش برای یافتن الگوهای پر تکرار، وابستگی‌ها، ارتباطات یا ساختارهای علی موجود در میان مجموعه‌ای از عناصر و یا اشیاء در پایگاه داده‌های تراکنشی و یا پایگاه داده‌های رابطه‌ای و دیگر مخازن اطلاعات به کار می‌رود. مثلاً اینکه چه نوع خدمات بانکی اغلب به صورت همزمان توسط یک مشتری مطالبه می‌شود می‌تواند به صورت یک قاعده از پایگاه داده‌های بانک استخراج شود [۳۱].

قواعد وابستگی برای بیان حقایق موجود در مجموعه‌ای از داده‌ها به کار می‌روند. بدین ترتیب که اگر بعضی وقایع رخ دهند آن‌گاه وقایع دیگری نیز رخ خواهند داد. اگر یک قاعده وابستگی را با $X \Rightarrow Y$ نشان دهیم (یعنی در صورتی که X اتفاق بیفتد Y هم اتفاق می‌افتد)، مسئله قواعد وابستگی در اصل یافتن قواعدی است که از حداقل پشتیبانی^۲ برخوردار بوده و به اندازه کافی به آن‌ها اعتماد^۳ داشته باشیم.

۲-۸-۵-۸- تحلیل توالی^۴

تحلیل توالی برای یافتن الگوهای موجود در بین یک مجموعه داده است. تحلیل توالی و قواعد انجمنی هر دو شامل مجموعه‌ای از حالات، تراکنش‌ها و یا ارتباطات هستند. تفاوت این دو روش در این است که تقدم و تأخر رخدادها در تحلیل توالی بسیار حائز اهمیت است و توالی انتقال‌های بین حالت‌های مختلف تحلیل می‌شود. در صورتی که در قواعد انجمنی تراکنش‌های مختلف یک مشتری ارزش یکسان دارند. مثلاً در تحلیل توالی اینکه یک مشتری بانک از خدمات انتقال وجه بعد از واریز سود بهره گیرد یا پیش از واریز سود اهمیت دارد. اما در کشف قواعد انجمنی اینکه کدام فعالیت زودتر یا دیرتر انجام شده اهمیتی ندارد و دو تراکنش واریز سود و برداشت وجه مجموعه اقلام یکسانی را می‌سازند.

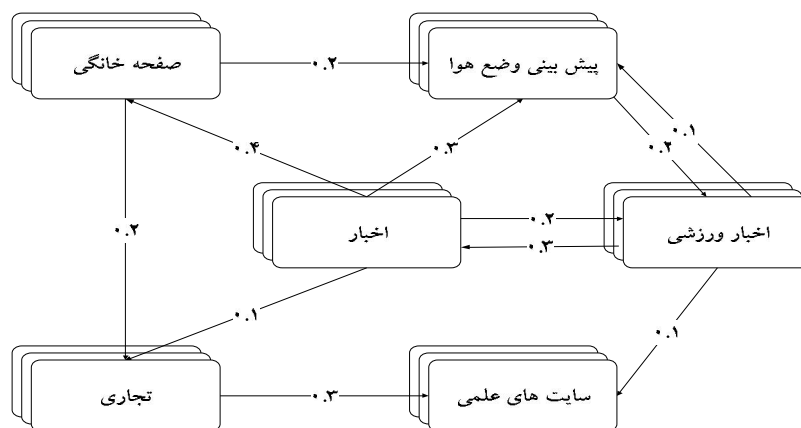
شکل ۲-۱۰ توالی از بازدیدهای وب را نشان می‌دهد. هر گره شامل یک گروه URL است. هر خط جهت دار انتقال بین این گره‌ها را نشان می‌دهد و وزن هر انتقال نشان دهنده احتمال انتقال بین این گره‌هاست [۱].

^۱ Association Rules

^۲ Support

^۳ Confidence

^۴ Sequence Analysis



شکل ۲-۱۰ نمونه‌ای از توالی بازدیدها در وب

۲-۸-۶- نرم افزار داده کاوی

ابزارهای داده کاوی که جهت مدیریت دانش مشتری انتخاب می‌شوند باید قادر باشند تا اطلاعات ضروری را از میان انبوه داده‌های در دسترس استخراج کنند. برای حصول این منظور ابزارهای داده کاوی باید خصوصیات زیر را داشته باشند:

- محیط کاربر پسند.
- بهره‌وری^۱ بالا
- قابلیت انجام عملیات اساسی لازم
- هزینه اجرای نسبتاً پایین [۳۰].

در حال حاضر چندین نرم افزار جهت داده کاوی توسط شرکت‌های معتبر نرم افزاری عرضه شده است. نرم افزار مورد استفاده در این تحقیق ۱۲,۰ SPSS Clementine می‌باشد. SPSS از شرکت‌های عمده آماری است که حاوی تعدادی از محصولات داده کاوی می‌باشد. SPSS توسط شرکت بریتانیایی ISL در اواخر سال ۱۹۹۸ ارائه شد و بسته داده کاوی Clementine را ارائه نمود.

^۱ Efficiency

Clementine از اولین نرم افزارهایی بود که به مفهوم جریان داده‌کاوی پرداخت و به کاربر این امکان را داد که کارهایی چون پاک‌سازی داده‌ها، تبدیل داده‌ها و آموزش مدل را در همان محیط گردش کار انجام دهند. این نرم‌افزار شامل ابزارهایی برای مدیریت چرخه پروژه داده‌کاوی دارا می‌باشد.

۲-۸-۷- کاربردهای داده‌کاوی

امروزه کاربردهای بسیار زیادی از داده‌کاوی در بسیاری از زمینه‌ها از جمله بازاریابی، علوم پزشکی، شرکت‌های بیمه و مخابرات، شرکت‌های تبلیغاتی و همه سازمان‌هایی که به علت تراکنش‌های بالا و سروکار داشتن با مشتریان، دارای پایگاه داده‌های بزرگ می‌باشند، شناخته شده است.

این علم می‌تواند پیش‌بینی‌هایی با دقت بالا برای سازمان‌ها انجام دهد، همچنین به مدیران و متخصصان و کارشناسان کمک می‌کند که بتوانند آینده شرکت و کسب و کار خود را با دقت‌های بسیار بالا پیش‌بینی کنند، داده‌کاوی به مراکز صنعتی کمک می‌کند که درک و بینش صحیحی از نحوه کارکرد دستگاه‌های صنعتی پیچیده خود بدست آورند. در واقع داده‌کاوی ابزاری بسیار کارا برای استفاده از داده‌هایی است که تولید می‌شوند اما استفاده‌ای از آن‌ها نمی‌شود.

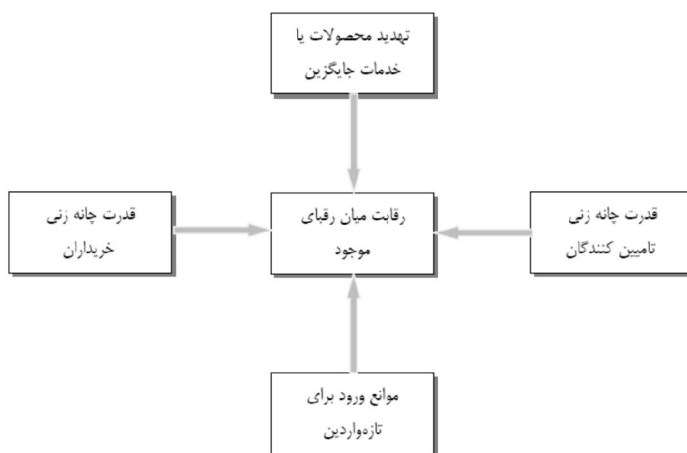
داده‌کاوی علمی است که باعث تغییر و تحول در هر صنعت و کسب و کاری می‌شود و اکنون بسیاری از شرکت‌های معتبر در سرتاسر دنیا توانسته‌اند با استفاده از روش‌های داده‌کاوی سود سالیانه خود را به اندازه چشم‌گیری افزایش دهند تا جایی که این شرکت‌ها در کنار واحدهایی مانند تحقیق و توسعه، تکنولوژی اطلاعات، برنامه‌ریزی و غیره، اقدام به تأسیس واحدهای داده‌کاوی نموده‌اند.

همچنین کاربردهای زیادی از داده‌کاوی در زمینه صنعت بانکداری شناخته و بکار گرفته شده است. از این میان می‌توان به کاربرد این ابزار در بازاریابی و مدیریت ارتباط با مشتریان (نظیر پروفایل سازی جهت جذب، حفظ و توسعه مشتری، مراقبت از مشتری، تحلیل روی‌گردانی مشتری)، کاربرد داده‌کاوی در کشف تقلب و سوءاستفاده‌های مالی (مانند فساد مالی، سوءاستفاده از دارایی‌ها و تقلب در گزارش‌گری مالی)، کاربرد داده‌کاوی در پیش‌بینی عملکرد بانک، داده‌کاوی جهت ارزیابی اعتبار مشتریان بانک و داده‌کاوی جهت ارزیابی عملکرد بانک‌ها اشاره نمود [۱].

۲-۸-۷-۱- داده‌کاوی در صنعت بانکداری

بخش بانکداری طبق الگوی پورتر برای فعالیت در محیط رقابتی امروز با پنج نیروی رقابتی مواجه است که یکی از آن‌ها قدرت چانه زنی مشتری است. افزایش رقابت بین بانک‌ها، تغییر قوانین و معرفی

فناوری‌های جدید و خصوصاً زیرساخت اینترنتی سبب افزایش آگاهی مشتریان به همه امور شده است و در نتیجه قدرت چانه زنی مشتری را بیشتر کرده است، به طوری که مشتری قادر است در کمترین زمان به بانک دیگری رجوع کند [۱۱].



شکل ۲-۱۱ نیروهای رقابتی پورتر [۱۱ به نقل از ۶]

صنعت بانکداری در جهان تحت تغییرات شدید در طریقه انجام کسب و کار می‌باشند. بانک‌های پیشرو از ابزارهای داده‌کاوی برای بخش‌بندی، تعیین سودمندی، دسته‌بندی اعتبار، پیش‌بینی قصور در بازپرداخت‌ها، بازاریابی، تشخیص تراکشن‌های متقلبانه و غیره استفاده می‌کنند. این ابزار به عنوان یک ابزار رقابتی در بانک شناخته شده است [۱۱].

داده یکی از با ارزش‌ترین دارایی‌های شرکت‌ها می‌باشد، اما فقط در صورتی که بدانیم چگونه دانش در آن‌را آشکار کنیم. داده‌کاوی امکان استخراج دانش موجود در داده‌های تاریخی و پیش‌بینی پیامدهای موقعیت‌های آینده را در اختیار می‌گذارد. داده‌کاوی ابزار ارزشمندی می‌باشد که با کمک آن یک سازمان می‌تواند با شناسایی اطلاعات مفید بالقوه از مقدار اطلاعات جمع‌آوری شده، مزیت واضحی نسبت به رقبایش کسب نماید [۱۱].

کاهش هزینه‌ای ذخیره‌سازی داده‌ها و افزایش راحتی در ذخیره‌سازی داده‌ها، توسعه الگوریتم‌های قوی و مؤثر یادگیری ماشین برای پردازش داده‌ها و کاهش هزینه قدرت محاسباتی از عواملی است که باعث گسترش و علاقه به داده‌کاوی گشته است [۱۱].

۹-۲- پیشینه تحقیق

از آنجاکه هوشمندی سازمانی بر کشف دانش به واسطه بهره‌گیری از منابع داده‌ای تأکید داشته و این امکان جهت پشتیبانی اتخاذ تصمیمات در سازمان‌ها کاربرد دارد، جای تعجب نیست که اخیراً مسائلی چون مدیریت دانش مشتری، مدیریت ارتباط با مشتری، داده‌کاوی و ترکیب این زمینه‌ها مطالعات بسیاری را به خود اختصاص داده‌اند. در این میان بانک‌ها به دلیل ماهیت ارتباط خود با مشتریان و حجم زیاد تراکنش‌ها و داده‌های برجای مانده از آن‌ها توجه ویژه‌ای را می‌طلبند.

مدهوشی و همکاران در [۲۱] اثر مدیریت دانش مشتری (CKM) بر مدیریت ارتباط با مشتری (CRM) را توسط ابزارهای آماری بررسی و تحلیل نمودند. در این مطالعه برای جمع آوری داده‌ها، پرسشنامه‌ای بین ۱۴۴۰ نفر از مدیران، معاونان و کارشناسان شعب بانک ملی استان مازندران توزیع شد. داده‌های حاصل توسط نرم‌افزارهای SPSS و LISREL مورد تجزیه و تحلیل قرار گرفت. تست نتایج آزمون با روش آلفای کرونباخ، ضریب ثابت ۹۴,۷۸ را نشان داد. نتایج آزمون کولموگروف-اسمیرنوف نشان داد که متغیرهای CKM و CRM نرمال هستند. به منظور تست اولویت اجزای CKM و CRM آزمون کروسکال-والیس را اجرا شد. سپس با استفاده از تست رگرسیون همبستگی، ارتباط بین CKM و CRM مورد مطالعه قرار گرفت. در نتیجه آن‌ها نشان دادند که ارتباط معنی‌داری بین CKM و CRM وجود دارد. در نهایت تحلیل برازش منحنی ریشه متوسط خطای مربعی کمتر ۰,۱۰ و شاخص‌های مناسب بودن برازش بیشتر از ۰,۹۰ را نشان داد که این نتایج حاکی از تأثیر CKM بر CRM در بانک ملی ایران، شعب مازندران بوده است.

همان‌طور که پیش از این اشاره شد تا به امروز زمینه‌های بسیاری از کاربرد داده‌کاوی در صنعت بانکداری شناخته شده و این زمینه‌ها مطالعات و تحقیقات زیادی را به خود تخصیص داده‌اند. از جمله این کاربردها می‌توان به کاربرد داده‌کاوی در بخش‌بندی مشتریان و مدل‌سازی رفتاری آن‌ها با اهداف پیش‌بینی ارزش طول عمر مشتری (CLV) یا ارائه سرویس مناسب به مشتریان بالقوه در مسیر اجرای CRM، کاربرد داده‌کاوی در رتبه‌بندی اعتبار مشتریان متقاضی وام، کاربرد داده‌کاوی در زمینه تحلیل روی‌گردانی مشتریان بانک، کاربرد داده‌کاوی در زمینه کشف تقلب و سوءاستفاده‌های مالی، کاربرد داده‌کاوی در زمینه پیش‌بینی عملکرد بانک و کاربرد داده‌کاوی در ارزیابی عملکرد بانک‌ها و ... اشاره نمود [۱].

از این رو در بخش پیشینه تحقیق، بررسی مطالعات صورت گرفته به تفکیک کاربرد مطرح خواهد شد.

۲-۹-۱- کاربرد داده کاوی در بخش بندی و مدل سازی رفتاری مشتریان در صنعت بانکداری

با در نظر گرفتن این اصل که برای موفقیت در کسب و کار درک کامل مشتریان ضروری است و از آنجا که مشتریان سازمان نیازها و روحیات متفاوتی دارند می توان گفت استفاده از استراتژی هایی چون بازاریابی مستقیم جهت ارائه و معرفی خدمات و محصولات جدید سازمان به مشتریان چندان کارآمد نخواهد بود. از این رو تکنیک های رتبه بندی مشتریان به مدیران امکان خواهد داد تا با هر مشتری بر اساس علایق و خواسته های او برخورد نموده و از ویژگی های مربوط به هر دسته جهت اتخاذ استراتژی های متناسب بهره گیرد.

مؤمنی و همکاران در [۸] به لزوم ارتباط با مشتری در بانک ها پرداخته اند. در این مقاله اشاره شده که طبق قاعده معروف ۲۰/۸۰ پارتو معمولاً ۲۰ درصد مشتریان، ۸۰ درصد سودآوری سازمان را منجر می شوند. بنابراین دسته بندی مشتریان می تواند ابزار شناسایی این گروه طلایی و ارائه خدمات بیشتر به آنها بوده و سود بیشتر سازمان را برای سازمان به ارمغان آورد. در ادامه به کاربردهای داده کاوی در صنعت بانکداری اشاره شده و به طور خاص داده کاوی جهت امتیازدهی و برآورد اعتبار مشتریان پرداخته شده است. در این مطالعه مشتریان از نظر اعتبار به چهار دسته تقسیم شده و مشتریان متقاضی اعتبار جدید نیز بر مبنای مدل حاصل دسته بندی می شوند.

هسیه در [۳۵] به تحقیقی تحت عنوان «مدل یکپارچه داده کاوی و رتبه بندی رفتاری برای تجزیه و تحلیل رفتار مشتریان بانک» پرداخته است. این تحقیق مدلی ادغامی از داده کاوی و مدل رتبه بندی مشتریان جهت مدیریت کارت های اعتباری مشتریان در بانک ارائه می دهد. در این تحقیق جهت ساخت مدلی برای شناسایی الگوی رفتاری مشتریان از شبکه های عصبی و روش خود سازمان دهنده^۱، استفاده گردیده است. در ابتدا مشتریان بانک با استفاده از تکنیک های SOM خوشه بندی گردیده و سپس با استفاده از قوانین وابستگی متغیرهای رفتاری که بیشترین ارتباط را با یکدیگر دارند شناسایی گردیده اند. این تحقیق مشتریان بانک را به سه گروه اصلی تقسیم می نماید. این مطالعه نشان می دهد که شناسایی

^۱ Self-Organizing map

مشخصه‌های مشتریان با استفاده از مدل رتبه‌بندی رفتاری مفید می‌باشد و توسعه استراتژی‌های بازاریابی را تسهیل می‌نماید.

حسینی در [۱۱] از تکنیک‌های هوشمند داده‌کاوی جهت مدیریت ارتباط با مشتریان بانک پارسیان بهره گرفته است. او در پایان‌نامه خود با ترکیب دو دسته از متغیرها، ۳۰ متغیر برای بررسی الگوی رفتاری مشتریان شناسایی نموده و با استفاده از شبکه عصبی کوهنن ماتریس نرون‌های خروجی شبکه ۳*۵ را تشخیص داده و مشتریان را در ۱۳ خوشه تقسیم نمود. در ادامه این تحقیق ارائه سرویس‌های بانکی مناسب به مشتریان بالقوه با استفاده از تکنیک‌های داده‌کاوی را مورد بررسی قرار داده و نتایج بدست آمده حاکی از آنست که از ۲۱ سرویس در نظر گرفته شده بانک پارسیان که شامل اعلام موجودی، دریافت صورت حساب، پرداخت قبض، برداشت وجه، انتقال وجه، عملیات مربوط به چک از کانال‌های مختلف بانکی شامل دستگاه ATM، اینترنت، موبایل، تلفن و پایانه‌های خرید POS می‌باشد، ۱۲ سرویس در گروه مورد نظر بیشترین استفاده را داشته است. کشف این مسئله که کدام دسته از مشتریان سودمندی لازم را برای نگهداری دارند و جلوگیری از هزینه‌های اضافی تبلیغات به صورت گسترده با ارائه سرویس‌های مناسب به مشتریان بالقوه، از نتایج این بررسی اعلام شده است.

مقاله [۲۳] به خوشه‌بندی مشتریان با خصوصیات رفتاری مشابه (RFM) با هدف تخمین وفاداری مشتری جهت تسهیل اتخاذ استراتژی و پیشنهاد سرویس‌های متناسب هر گروه با ترکیب روش‌های داده‌کاوی و ارزش طول عمر مشتری^۱ (CLV)، پرداخته است. در این تحقیق از داده‌های جمعیت شناختی و داده‌های مربوط به تراکنش‌های مالی مشتریان (RFM) مربوط به شرکت ایرانی ساپکو به عنوان ورودی استفاده شد. این مقاله رویه جدیدی را بر مبنای مدل RFM توسعه یافته که شامل یک پارامتر اضافی است معرفی می‌نماید. در این رویه، روش RFM وزن دار^۲ با الگوریتم K-Means در داده‌کاوی ترکیب شده و از روش دیویس-بولدین^۳ برای محاسبه مقدار بهینه K استفاده شده است و نهایتاً مشتریان بر مبنای وفاداری به محصولات شرکت ساپکو ایران دسته‌بندی شدند. نتایج حاصل نشان داد قابلیت سازمان برای جلب وفاداری مشتریان خود با استراتژی‌هایی که با توجه به این دسته‌بندی

^۱ Customer Lifetime Value (CLV)

^۲ Weighted Recency-Frequency-Monetary (WRFM)

^۳ Davies-Bouldin Index

بکار برد در مقایسه با روش معمول انتخاب تصادفی که در اکثر سازمان‌ها در ایران بکار می‌رود، افزایش چشمگیری داشته است.

۲-۹-۲- کاربرد داده‌کاوی در ارزیابی اعتبار مشتریان

عبدو و همکاران در [۲۹] به مطالعه داده‌های مشتریان یک بانک مصری که از این بانک وام گرفته بودند پرداختند. در این مطالعه مشتریان بانک از لحاظ اعتبار دسته‌بندی شدند. الگوریتم‌های بکار رفته در این مقاله شبکه‌های عصبی احتمالی و چند لایه و رگرسیون لجیت و ورودی‌های این الگوریتم‌ها اطلاعات شخصی مشتری شامل سن، درآمد ماهانه، جنسیت، وضعیت تأهل، وضعیت مسکن، مقدار وام، مدت بازپرداخت و... بوده است. در این تحقیق عملکرد شبکه‌های عصبی احتمالی و چند لایه با روش‌های مرسوم مانند آنالیز تفکیکی، رگرسیون منطقی و آنالیز بر اساس حداقل انحراف از میزان متوسط مقایسه شده است. نتایج حاصل بیانگر این مسئله است که رگرسیون لجیت با ۸۸٪ پیش‌بینی صحیح عملکرد بهتری نسبت به روش‌های مرسوم که نام برده شد، داشته است؛ و در مقایسه کلی شبکه عصبی با ۹۶٪ دقت بهترین عملکرد را داشته است.

لی و همکاران در [۲۲] یک روش ترکیبی جدید برای انتخاب داده‌های ورودی دسته‌بندی مشتریان بانک جهت ارزیابی اعتبار آنان معرفی کرده‌اند. در این مقاله رویکردی ترکیبی که حاصل ترکیب رویکردهای انتخاب ویژگی موجود (آنالیز ترکیبی خطی، تئوری مجموعه‌های ناهموار، درخت تصمیم و روش Fscore) با روش ماشین بردار پشتیبان^۱ است برای انتخاب متغیرها معرفی شده و در ادامه مقایسه‌ای بین رویکردهای ترکیبی بر پایه SVM به منظور انتخاب ویژگی‌ها صورت گرفته است. در این تحقیق از داده‌های معروف مربوط به کارت‌های اعتباری UCI مربوط به کشور آلمان و استرالیا استفاده شد. تمرکز مقاله بر یافتن مرتبط‌ترین ویژگی‌ها با اهداف منظور است و بیان می‌شود که مسئله مهم نه فقط کاهش حجم متغیرهای ورودی بلکه حذف نویزهای ورودی نیز می‌باشد. نتایج حاصل نشان می‌دهند که عملکرد روش‌ها به صورت ترکیبی در انتخاب ویژگی‌های ورودی بسیار بهتر از عملکرد هر یک از این روش‌ها به طور منفرد است.

[۱۸] ایجاد یک ارتباط محکم و بلند مدت و متمرکز با مشتریان را هسته اصلی CRM، و درک درست از مشتری را اساس افزایش ارزش طول عمر مشتری (CLV) دانسته است. در این مقاله، هدف

^۱ Support Vector Machine (SVM)

از بخش‌بندی مشتری، ایجاد گروه‌های سودآور و در حال رشد بر پایه ویژگی‌های مشترک آن‌هاست که سازمان را قادر می‌سازد تا هر گروه را با پیشنهاداتی خاص مورد هدف قرار دهد. این امر بدون استفاده از روش‌ها و راهکارهای هوشمند برای تحلیل داده‌ها میسر نیست. تمرکز این مقاله بر روی بخش‌بندی استراتژی محور مشتریان سازمان در جهت تلاش برای به حداکثر رساندن پتانسیل مشتری که با اهمیت‌ترین منبع در کسب و کار است، می‌باشد. این مقاله بر بخش‌بندی اعتبار مشتریان در صنعت بانکداری تمرکز دارد و در مطالعه موردی خود از شبکه‌های عصبی چند لایه با بازخورد رو به جلو^۱ برای بخش‌بندی مشتریان به دو دسته استفاده نموده است: مشتریانی که با پرداخت مشکل دارند یا با پرداخت مشکلی ندارند.

۲-۹-۳- کاربرد داده‌کاوی در زمینه کشف تقلب

چن و همکاران در [۲۴] به دسته‌بندی مشتریان جهت کشف تقلب و سوءاستفاده‌های مالی و شناسایی بعضی از عوامل کلیدی برای تشخیص تقلب خریداران و فروشندگان در معاملات آنلاین توسط کارت اعتباری پرداخته‌اند. به این منظور از داده‌های مربوط به یک دوره سه ماهه مزایده‌ات آنلاین فروش لپ تاپ استفاده شده است. تکنیک مورد استفاده در این تحقیق رگرسیون لجستیک بوده که برخی از ویژگی‌های مربوط به داده‌های اولیه ورودی آن مبلغ اولیه پیشنهادی، مدت زمان انجام مزایده، سود مزایده، میزان اعتبار فروشنده و قیمت بهایی مزایده بوده است که اعتبار فروشنده به عنوان یک عامل مهم در تشخیص تقلب استفاده شده است. در نهایت دقت مدل پیشنهادی ۹۱٪ محاسبه شده که دقت نسبتاً قابل قبولی می‌باشد.

۲-۹-۴- کاربرد داده‌کاوی در تحلیل روی‌گردانی مشتری^۲

مسئله حفظ مشتری و افزایش وفاداری او به سازمان از مسائلی است که هسته اصلی بحث ارتباط با مشتری را تشکیل می‌دهد. با تحلیل داده‌های بر جای مانده از مشتریانی که سازمان را ترک نموده‌اند، قوانین و الگوهایی حاصل می‌شود که می‌توان به کمک آن‌ها مشتریانی که احتمال می‌رود در آینده نزدیک سازمان را ترک کنند و به سوی رقیب بروند شناسایی نمود. بدین ترتیب مدیران می‌توانند با اتخاذ تصمیماتی جهت بهبود ارتباط با این‌گونه مشتریان مانع از روی‌گردانی آنان گردند.

^۱ Multilayer feed forward Neural Network

^۲ Customer Churn Analysis

در [۱۹] از الگوریتم‌های داده‌کاوی جهت ساخت مدلی به منظور پیش‌بینی روی‌گردانی مشتریانی که از کارت اعتباری استفاده می‌کنند استفاده شده است. این تحقیق بر روی پایگاه داده مشتریان یک بانک چینی صورت گرفته و از چهار دسته متغیر اطلاعات مشتری، اطلاعات کارت اعتباری، داده‌های مربوط به ریسک مشتری و اطلاعات مربوط به تراکنش‌ها استفاده شده که در مجموع شامل ۱۳۵ متغیر می‌باشد. از بین این متغیرها ۹۵ متغیر با توجه به همبستگی بین آن‌ها برای انجام مراحل بعدی انتخاب شدند. در این تحقیق مشتری روی‌گردان فردی تعریف شده که در طول دوره مشاهده^۱ دوازده ماهه هیچ تعاملی را با بانک نداشته است. در این تحقیق الگوریتم‌های رگرسیون و درخت تصمیم جهت دسته‌بندی مشتریان انتخاب شده و نتایج تحقیق نشان داد که عملکرد مدل‌های رگرسیونی کمی بهتر از عملکرد درخت‌های تصمیم بوده است. در این تحقیق یک فاصله یک ساله به عنوان دوره ارزیابی عملکرد^۲ در نظر گرفته شد.

[۲۸] تأکید می‌کند که روی‌گردانی مشتری در بانک‌های چینی مورد تأکید زیادی قرار گرفته است. این بانک‌ها تلاش می‌کنند با منابع محدودی که در اختیار دارند به حداکثر رضایت مشتریان دست یابند. در این مقاله از یک روش یادگیری جدید به نام جنگل‌های تصادفی متوازن بهبود یافته^۳ (IBRF) جهت تعدیل مشکل بحث برانگیز عدم توازن در توزیع داده‌ها در تحلیل روی‌گردانی مشتری استفاده شده است. در این مطالعه اثربخشی رویکرد جنگل تصادفی استاندارد و روش‌های مختلف نمونه‌برداری نیز در پیش‌بینی روی‌گردانی مشتری نیز بررسی شده و با عملکرد IBRF مورد مقایسه قرار گرفت. این روش بر روی داده‌های واقعی مربوط به روی‌گردانی مشتری یک بانک چینی اعمال شده و مشخص شد که این روش به طور قابل توجهی دقت پیش‌بینی را در مقایسه با سایر الگوریتم‌ها مانند شبکه‌های عصبی مصنوعی، درخت‌های تصمیم و ماشین بردار پشتیبان بهتر عمل کرده است. این روش در مقایسه با سایر الگوریتم‌های جنگل تصادفی مانند الگوریتم جنگل تصادفی متوازن^۴ و الگوریتم جنگل تصادفی وزن‌دار^۵ بهتر عمل کرده است.

^۱ Observation Period

^۲ Performance evaluation period

^۳ Improved Balanced Random Forests (IBRF)

^۴ Balanced Random Forests

^۵ Weighted Balanced Random Forests

[۲۰] بیان می‌کند که از منظر اقتصاد و مدیریت ریسک، درک ویژگی‌های مشتری جهت حفظ مشتری ضروری به نظر می‌رسد و لازم است سازمان‌ها بین مشتریان معتبر و مشتریان بد تمایز قائل شوند. این مقاله تئوری مجموعه‌های ناهموار^۱ را که یک روش تصمیم‌گیری بر اساس قوانین است، برای استخراج قوانین مرتبط با روی‌گردانی مشتری بکار می‌گیرد. سپس از گراف شبکه جریان^۲ که یک رویکرد وابسته به مسیر است برای کشف قوانین و متغیرهای تصمیم‌گیری استفاده می‌نماید و در نهایت ارتباط بین قوانین و انواع مختلف روگردانی را نشان می‌دهد. این تحقیق بر روی نمونه‌ای ۲۱۰۰۰ نفره از مشتریان صورت گرفته و آن‌ها را به سه دسته مشتریان پایدار^۳، روی‌گردانی داوطلبانه و روی‌گردان غیر داوطلبانه تقسیم می‌کند. متغیرهای ورودی این تحقیق را متغیرهای جمعیت‌شناختی، روان‌شناختی^۴ و تراکنش‌های مربوط به این گروه از مشتریان تشکیل می‌دهند. نتایج نشان داد که این مدل ترکیبی پیش‌گویی خوبی از روی‌گردانی مشتری داشته و اطلاعات مفیدی برای تصمیم‌گیران در تدوین استراتژی‌های بازاریابی فراهم می‌کند.

جدول ۲-۸ پژوهش‌های انجام‌گرفته در زمینه کاربرد داده‌کاوی در صنعت بانکداری

شماره مرجع	هدف	روش‌های مورد استفاده	ورودی‌ها
کاربرد داده‌کاوی در بخش‌بندی و مدل‌سازی رفتاری مشتریان در صنعت بانکداری			
[۲۳]	بخش‌بندی مشتریان جهت تخمین وفاداری مشتری	داده‌های جمعیت‌شناختی، داده‌های وزن‌دار مربوط به تراکنش‌های مالی (WRFM) مشتریان شرکت ایرانی ساپکو	خوشه‌بندی با استفاده از روش SOM و K-means
[۸]	دسته‌بندی مشتریان جهت شناسایی مشتریان طلایی طبق اصل ۸۰/۲۰ پارتو	اطلاعات مربوط به مشتریان وام‌گیرنده از بانک	شبکه‌های عصبی
[۳۵]	شناسایی الگوی رفتاری مشتریان	متغیرهای جمعیت‌شناختی، متغیرهای RFM مشتریان	شبکه‌های عصبی، خوشه‌بندی توسط تکنیک‌های خود سازمان دهنده (SOM) و

^۱ Rough Set Theory

^۲ Flow Network Graph

^۳ Survival

^۴ Psychographic

الگوریتم Apriori برای شناسایی وابستگی متغیرهای رفتاری	بانک و لیستی از خدمات قابل ارائه توسط بانک		
شبکه‌های عصبی، خوشه‌بندی توسط تکنیک‌های خود سازمان دهنده (SOM) و الگوریتم Apriori برای کشف وابستگی بین خدمات ارائه شده توسط بانک با مشتریان.	متغیرهای جمعیت‌شناختی، متغیرهای RFM مشتریان بانک و لیستی از خدمات ارائه شده توسط بانک. مورد مطالعه: بانک پارسیان	بخش‌بندی مشتریان جهت مدیریت ارتباط با آنان و کشف سرویس‌های مرتبط با هر بخش	[۱۱]
کاربرد داده‌کاوی در ارزیابی اعتبار مشتریان			
رویکردی ترکیبی که حاصل ترکیب رویکردهای انتخاب ویژگی موجود (آنالیز ترکیبی خطی، تئوری مجموعه‌های ناهموار، درخت تصمیم و روش Fscore) با روش ماشین بردار پشتیبان	داده‌های مربوط به کارت اعتباری UCI مربوط به کشور آلمان و استرالیا	دسته‌بندی مشتریان بانک جهت ارزیابی اعتبار آنها	[۲۲]
شبکه‌های عصبی چند لایه با بازخورد رو به جلو	مجموعه داده‌های اعتباری مشتریان	بخش‌بندی اعتبار مشتریان در صنعت بانکداری به دو دسته: مشتریانی که با پرداخت مشکل دارند یا با پرداخت مشکلی ندارند	[۱۸]
الگوریتم شبکه‌های عصبی احتمالی و چند لایه و رگرسیون لاجیت	اطلاعات شخصی مشتری (همچون سن، درآمد ماهانه، جنسیت، وضعیت مسکن و...) و اطلاعات وام (شامل مقدار وام و مدت بازپرداخت وام و...)	دسته‌بندی مشتریان از لحاظ اعتبار بازپرداخت وام	[۲۹]
کاربرد داده‌کاوی در زمینه کشف تقلب			
رگرسیون لاجیت	داده‌های مربوط به یک دوره سه ماهه مزایادات	دسته‌بندی مشتریان جهت کشف تقلب و	[۲۴]

	سوءاستفاده‌های خریداران و فروشندگان در معاملات آنلاین توسط کارت اعتباری	آنلاین فروش لپ تاپ (شامل ویژگی‌های مبلغ اولیه پیشنهادی، مدت زمان انجام مزایده، سود مزایده، میزان اعتبار فروشنده و قیمت بهایی مزایده)	
کاربرد داده‌کاوی در تحلیل روی گردانی مشتری			
[۱۹]	پیش‌بینی روی گردانی مشتریان از کارت اعتباری	پایگاه داده ۱۲ ماهه مشتریان یک بانک چینی (شامل ۹۵ متغیر از ۱۳۵ متغیر مربوط به اطلاعات مشتری، اطلاعات کارت اعتباری، داده‌های مربوط به ریسک مشتری و اطلاعات مربوط به تراکنش‌ها)	الگوریتم‌های رگرسیون و درخت تصمیم جهت دسته‌بندی مشتریان
[۲۸]	روی گردانی مشتری در بانک‌های چینی جهت دستیابی به حداکثر رضایت مشتریان	داده‌های واقعی مربوط به روی گردانی مشتری یک بانک چینی	جنگل‌های تصادفی متوازن بهبود یافته ^۱ (IBRF)
[۲۰]	استخراج قوانین مرتبط با روی گردانی	نمونه‌ای ۲۱۰۰۰ نفره از مشتریان (شامل متغیرهای جمعیت‌شناختی، روان‌شناختی ^۲ و تراکنش‌های مربوط به این گروه از مشتریان)	تئوری مجموعه‌های نا هموار و گراف شبکه جریان

نگای در [۲۵] به مرور و دسته بندی ادبیات در زمینه کاربردهای داده‌کاوی در مدیریت ارتباط با مشتری پرداخته است. در این تحقیق ضمن بیان ضرورت کاربردهای تکنیک‌های داده‌کاوی در مدیریت

^۱ Improved Balanced Random Forests (IBRF)

^۲ Psychographic

ارتباط با مشتری به دسته‌بندی تکنیک‌ها و مقایسه روش‌های داده‌کاوی در این زمینه پرداخته است. در انتها چارچوبی جهت انجام تحقیقات آتی در زمینه کاربردهای داده‌کاوی در حوزه مدیریت ارتباط با مشتری ارائه شده است.

۲-۱۰- جمع‌بندی مطالب فصل

در این فصل ابتدا به تعاریف و مفاهیم پایه مرتبط با موضوع تحقیق پرداخته شد و مفاهیمی چون مدیریت دانش، مدیریت دانش مشتری، مدیریت ارتباط با مشتری و داده‌کاوی معرفی شدند. در نهایت پیشنهادی جهت تحقیق مورد بررسی قرار گرفت.

به طور خلاصه می‌توان گفت با تبدیل مشتری به یک دارایی مهم و ارزشمند سازمان و تعیین دانش به عنوان عامل پویایی و بقای سازمان، مباحثی چون مدیریت دانش مشتری در میان محققان جایگاه ویژه‌ای یافت. مدیریت دانش می‌تواند مزایایی چون تسهیم و اشتراک دانش و صرفه‌جویی در زمان و ... را به دنبال داشته باشد. بهره‌گیری از این مزایا منجر به حفظ بقای سازمان در محیط متلاطم و پر رقابت کسب و کار امروز خواهد بود.

مدیریت ارتباط مشتری سنتی ابزاری جهت ارتباط با مشتریان و تلاش جهت حفظ وفاداری آن‌ها بود. اما خلاً بزرگ آن عدم بهره‌مندی از مدیریت دانش مشتریان بود. اما زمانی که مدیریت ارتباط مشتری، مدیریت دانش را جهت کسب دانش مشتری بکار گیرد فرایند مدیریت دانش مشتری خواهیم داشت. ابزار حاصل کمک خواهد کرد ارتباط سازمان با مشتریان از حالت منفعل خارج شده و با مشتری به عنوان شریک دانشی سازمان ارتباط برقرار نماید.

مدیریت دانش مشتری با ترکیب هر دو رویکرد فناوری محور^۱ و داده‌گرا^۲ در بحث مدیریت ارتباط با مشتری و رویکرد فرد گرا^۳ در بحث مدیریت دانش با هدف بهره‌گیری از پتانسیل هم‌افزای هر دوی آن‌هاست [۴۱]. نتیجه این مسئله می‌تواند توصیف شیواتر دانش برای مشتری، دانش درباره مشتری و دانش از سوی مشتری باشد. بنابراین این امکان وجود خواهد داشت که محصولات و خدمات

^۱ Technology driven

^۲ Data oriented

^۳ People oriented

جدید به گروه مناسب از مشتریان تحویل شود. بنابراین ریسک شکست محصول جدید کاهش خواهد یافت [۲۷].

از آنجا که به طور مشخص، در نظام بانکی، مشتری سودمندترین و مهم‌ترین رکن این سازمان می‌باشد، تمرکز و مطالعه بر اطلاعات و رفتار مشتریان و کشف دانش موجود و بهره‌گیری از دانش کشف شده جهت اتخاذ استراتژی‌های کسب‌وکار می‌تواند در مسیر جذب و حفظ مشتریان و افزایش حس رضایت آن‌ها مؤثر باشد.

در این مسیر ابزار داده‌کاوی جهت کشف دانش مشتریان گردید. فرایند داده‌کاوی به سازمان‌ها کمک می‌کند تا انبوه داده‌های مشتریان را تحلیل نمایند و اطلاعات مفید را از آن‌ها استخراج نموده و مزیت رقابتی را نسبت به دیگران کسب نمایند [۳۰].

لازم به ذکر است جزییات دقیق روش انتخابی جهت داده‌کاوی در فصل‌های آینده به تفصیل مورد بحث قرار خواهد گرفت.

فصل سوم

روش تحقیق

۳-۱- مقدمه

تحقیق از نظر لغوی به معنای بررسی و پیدا کردن حقیقت است و از منظر علمی عبارتست از تلاش و انجام یک فعالیت منظم و هدف دار برای رسیدن به حقیقت، پاسخ به سؤال، دستیابی به دانش و آگاهی بیشتر در مورد یک پدیده و یا یک مسئله به منظور چاره جویی با استفاده از مراحل و روش علمی. به عبارت دیگر تحقیق تجزیه و تحلیل و ثبت عینی و سیستماتیک مشاهدات کنترل شده که ممکن است به پروراندن قوانین کلی، اصول یا نظریه‌هایی بینجامد و به پیش بینی و یا احتمالاً به کنترل نهایی رویدادها منتج شود.

در فصل پیش رو چارچوب کلی روش پیشنهادی تشریح می‌گردد و سعی شده با طرح این چارچوب بر پایه‌ی گام‌های اساسی، اهداف تحقیق که همان بخش‌بندی مشتریان و استخراج استراتژی‌های متناسب با هر بخش و به‌کارگیری استراتژی‌ها و دانش حاصل در چرخه مدیریت دانشی سازمان می‌باشد، حاصل گردد.

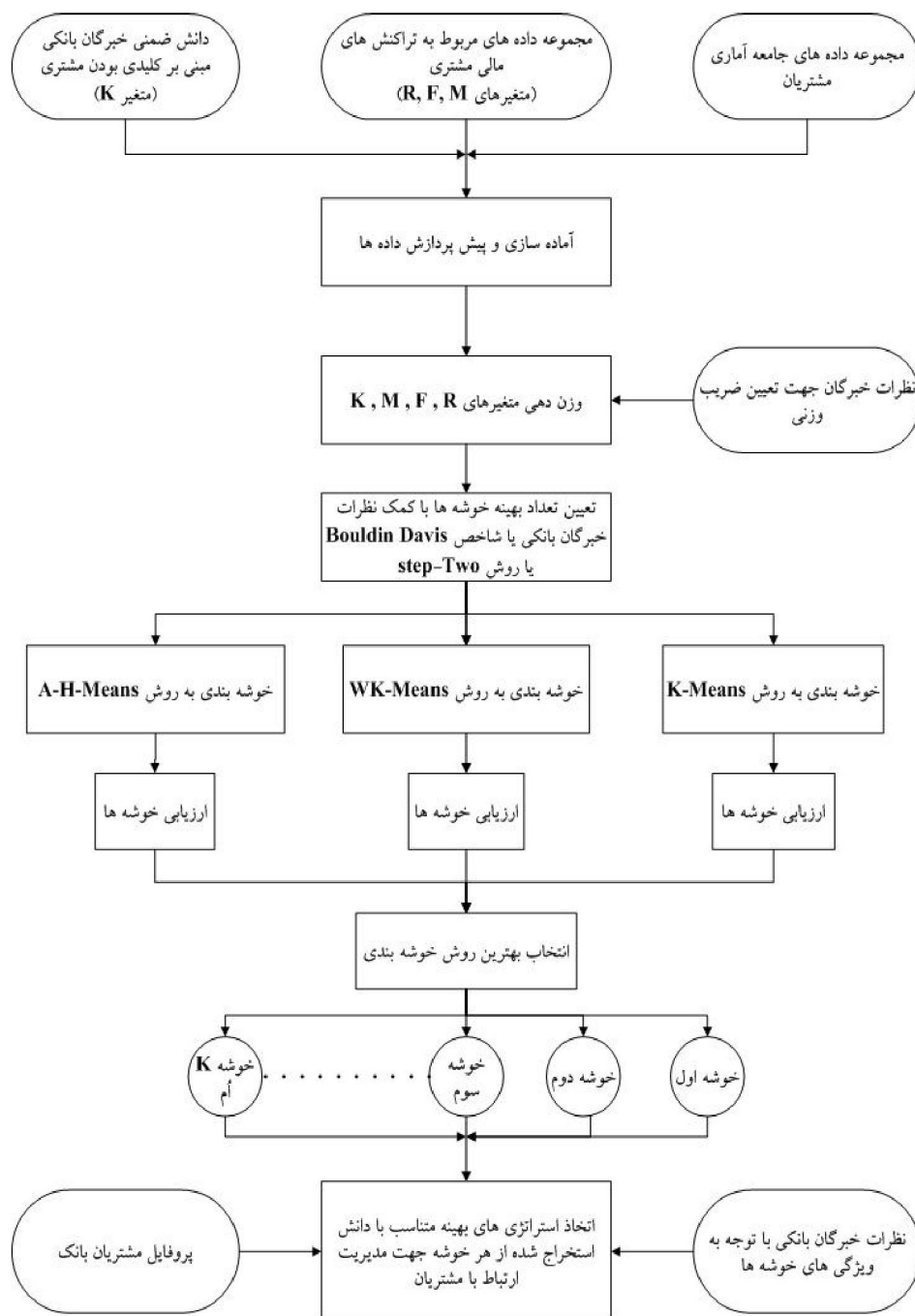
در این مسیر برخی روش‌های موجود در زمینه خوشه‌بندی مبتنی بر بخش‌بندی مانند K-Means، WK-Means، A-Harmonic means و معیارهای ارزیابی و اعتبارسنجی خوشه‌ها همچون شاخص مجموع مربع خطاها و روش Min-Max برای نرمال سازی داده‌ها تشریح خواهند شد.

۳-۲- روش پیشنهادی

هدف از این تحقیق کشف دانش نهفته در داده‌های خام مشتریان بانک اقتصاد مهر با استفاده از روش‌های داده‌کاوی و به‌کارگیری آن جهت اتخاذ استراتژی‌هایی برای مدیریت دانش مشتریان است که این مهم از تلفیق ابزارهای دانشی با رویکردهای مربوط به مدیریت ارتباط با مشتریان حاصل خواهد شد.

۳-۲-۱- چارچوب تحقیق

در شکل ۳-۱ چارچوب کلی این تحقیق نشان داده شده است.



شکل ۳-۱ چارچوب تحقیق

۳-۲-۲- انتخاب متغیرها

شناسایی متغیرها از گام‌های اساسی در مسیر تحقیق است. با در نظر گرفتن اهداف تحقیق، مطالعه تحقیقات انجام شده در حوزه بخش‌بندی رفتاری مشتریان بانک و بررسی رویکردهای سازمانی بانک مهر اقتصاد، سه دسته متغیر زیر انتخاب شده است:

- داده‌های جمعیت آماری مشتریان: این متغیرها شامل جنسیت، سن، تحصیلات، وضعیت تأهل، شغل و تاریخ تولد، محل منطقه بانکی و ... می‌باشد.
 - تراکنش‌های مالی مشتریان: تراکنش‌های مالی مشتری می‌تواند شاخص مناسبی برای بررسی الگوی رفتاری مشتری باشد و تحقیقات زیادی در این زمینه صورت گرفته است. در این میان متغیرهایی چون تراکنش اخیر مشتری^۱، متغیر پولی^۲، فراوانی^۳ (که به اصطلاح با سر نام RFM شناخته می‌شوند)، در مطالعات داده‌کاوی کاربرد گسترده‌ای یافته‌اند. این متغیرها در پایگاه اطلاعاتی بانک وجود دارند و از طریق پردازش داده‌های تراکنشی مشتریان محاسبه می‌شوند که R، فاصله میان زمان آخرین تراکنش مشتری تا زمان مورد ارزیابی، F، تعداد خریدهای مشتریان در یک بازه زمانی خاص و M، میانگین مبالغ تراکنش‌های مشتری در یک بازه زمانی خاص می‌باشند.
 - دانش ضمنی خبرگان بانکی: بانک مهر اقتصاد برخی از مشتریان خود را به عنوان مشتریان کلیدی^۴ می‌شناسد. مشتریان کلیدی بانکی در هر استان بنا بر نظر خبرگان بانکی با ویژگی‌های خاصی تعریف می‌شوند. منشأ این ویژگی‌ها دانش ضمنی خبرگان بانکی در هر استان می‌باشد. در استان مازندران، این متغیر بر اساس ویژگی مانده حساب^۵ مشتری تعیین می‌گردد. در این تحقیق این متغیر با نام K تعریف می‌گردد.
- این متغیرها در جدول ۳-۱ نمایش داده شده است.

^۱ Recency

^۲ Monetary

^۳ Frequency

^۴ Key Customer

^۵ Balance

جدول ۳-۱ متغیرهای تحقیق

ردیف	متغیر	شرح
۱	ID	شماره مشتری
۲	Sex	جنسیت
۳	Age	سن
۴	Education	تحصیلات
۵	Marital Status	وضعیت تأهل
۶	Job	شغل
۷	Location	محل منطقه بانکی
۸	R	فاصله زمانی میان آخرین تراکنش مشتری تا زمان گزارش گیری
۹	F	تعداد تراکنش های مشتریان در بازه زمانی یک ساله تا زمان گزارش گیری
۱۰	M	میانگین مبلغ تراکنش های مشتری در بازه زمانی یک ساله تا زمان گزارش گیری
۱۱	K	مشتری جزء مشتریان کلیدی بانک می باشد یا خیر

۳-۲-۳- آماده سازی و پیش پردازش داده ها

در این تحقیق برای جمع آوری داده ها از داده های موجود در پایگاه اطلاعاتی بانک مهر اقتصاد استفاده شده است.

به طور معمول ویژگی های نمونه های مورد مطالعه در یک تحقیق داده کاوی دارای بازه تغییرات یکسان نمی باشند. برای مثال اگر سن و متغیر پولی را به عنوان دو ویژگی یک مشتری در نظر بگیریم، این ویژگی ها دارای بازه تغییرات متفاوتی هستند. همچنین مقادیر به دست آمده برای هر ویژگی نیز به واحد استفاده شده برای اندازه گیری آن ویژگی ارتباط مستقیم دارد برای مثال اگر برای اندازه گیری سن از واحد سال یا روز استفاده شود، مقادیر در بازه های مختلفی به دست خواهند آمد. در این صورت، ویژگی های با مقادیر بزرگ تأثیر بیشتری بر تابع هدف می گذارند که لزوماً به معنی مهم تر بودن آنها در

الگوریتم بخش‌بندی نیست و بنابراین اثری نامطلوب تلقی می‌گردد. برای رفع این مشکل از نرمال سازی به روش Min-Max استفاده شده است.

۳-۲-۱- نرمال سازی داده‌ها

در این روش، با یک تبدیل خطی یا غیرخطی، داده‌ها را در بازه‌ای که توسط کاربر انتخاب می‌شود قرار می‌گیرند. این بازه در کاربردهای داده‌کاوی معمولاً $[-1, 1]$ یا $[0, 1]$ می‌باشد. برای قرارگیری داده‌ها در بازه بین صفر و یک می‌توان از فرمول ۳-۱ استفاده نمود [۳۱].

$$X^* = \frac{X - \text{Min}(X)}{\text{Range}(X)} \quad (3-1)$$

در فرمول بالا X نشان‌دهنده مقدار متغیر است.

۳-۲-۲- تعیین تعداد بهینه خوشه‌ها

یکی از مهم‌ترین مسایل در خوشه‌بندی انتخاب تعداد خوشه‌های مناسب می‌باشد. تعداد خوشه‌ای مناسب می‌باشد که:

نمونه‌های موجود در یک خوشه تا حد امکان شبیه به یکدیگر باشند.

نمونه‌های متعلق به خوشه‌های متفاوت تا حد امکان با یکدیگر نامشابه باشند.

عبارات فوق را بدین صورت نیز بیان می‌کنند که خوشه‌ها باید بیشینه فشردگی داشته باشند و تا حد امکان جدایی آن‌ها نیز زیاد باشد. برای یک خوشه‌بندی مناسب هر دو معیار باهم باید ارضا شوند چرا که اگر تنها معیار فشردگی مورد استفاده قرار گیرد در آن صورت هر داده می‌تواند به صورت یک خوشه در نظر گرفته شود چرا که هیچ خوشه‌ای فشردتر از خوشه‌ای با یک داده نیست و اگر تنها معیار جدایی در نظر گرفته شود در آن صورت بهترین خوشه‌بندی این است که کل داده‌ها را یک خوشه بگیریم با این توضیح که فاصله هر خوشه از خودش صفر است. بنابراین باید از ترکیب دو معیار فوق استفاده شود.

به منظور تعیین تعداد بهینه خوشه‌ها می‌توان از روش‌هایی چون شاخص Davis-Bouldin، روش Two Step، تعیین تعداد بهینه خوشه‌ها به کمک روش‌های مبتنی بر گراف و یا نظرات خبرگان استفاده نمود.

۳-۲-۵- خوشه‌بندی

خوشه‌بندی یا گروه‌بندی، تقسیم اقلام موجود در یک مجموعه داده است که به طور طبیعی باهم شباهت دارند. داده‌هایی که با این معیار به صورت خوشه‌هایی تفکیک می‌گردند، با داده‌های موجود در خوشه‌ای که در آن قرار می‌گیرند، بیش‌ترین شباهت را دارند؛ و با داده‌های موجود در سایر خوشه‌ها متفاوت‌اند.

در خوشه‌بندی موضوعات زیر مورد توجه است:

- چه تعداد از خوشه‌ها می‌تواند دانش نهفته در داده‌ها را کشف نماید؟ مسئله تعداد خوشه‌ها معمولاً به صورت جداگانه مورد بررسی قرار می‌گیرد.
- معیارهای شباهت و تفاوت داده‌ها چیست؟ این معیارها خود به واسطه روش‌های مختلفی محاسبه می‌گردد، اما در بیشتر روش‌های خوشه‌بندی موجود از معیار فاصله فضایی دو داده از یکدیگر، استفاده می‌شود. فاصله فضایی می‌تواند با روش‌های مختلفی چون فاصله اقلیدسی^۱، فاصله مینکوفسکی^۲ و یا فاصله مانهاتان^۳ محاسبه شود.
- بعد از تعیین تعداد خوشه‌ها و معیار شباهت یا فاصله داده‌ها، داده‌ها با استفاده از چه روشی در تعداد خوشه‌های معین جای گیرند [۱۱].

خوشه‌بندی یک الگوریتم بدون ناظر^۴ در داده‌کاوی است، زیرا هیچ صفتی منفردی برای هدایت فرایند Training استفاده نمی‌شود و همه صفات ورودی ارزش یکسان دارند [۲].

تفاوت این روش با طبقه‌بندی در این است که در طبقه‌بندی هر داده به یک طبقه (کلاس) از پیش مشخص‌شده تخصیص می‌یابد ولی در خوشه‌بندی هیچ اطلاعی از کلاس‌های موجود درون داده‌ها وجود ندارد و به عبارتی خود خوشه‌ها نیز از داده‌ها استخراج می‌شوند.

با مطالعه و بررسی روش‌های داده‌کاوی موجود جهت خوشه‌بندی داده‌ها، الگوریتم‌های K-Means، WK-Means و A-H-Means برای اعمال بر روی مجموعه داده‌ها انتخاب شده‌اند.

^۱ Euclidean distance

^۲ Minkowski

^۳ Manhattan distance

^۴ Unsupervised

۳-۲-۵-۱- انواع خوشه‌بندی

امروزه الگوریتم‌های متنوعی در زمینه خوشه‌بندی معرفی شده‌اند. این الگوریتم‌ها به طور کلی به سه دسته تقسیم می‌شوند [۱]:

۱- خوشه‌بندی سلسله‌مراتبی^۱

در این نوع خوشه‌بندی خوشه‌ها به صورت متوالی به دو شیوه انباشتی^۲ و یا تقسیم‌کنندگی^۳ توسعه می‌یابند. در روش انباشتی، هر یک از نقاط به عنوان خوشه در نظر گرفته شده و سپس خوشه‌های مشابه باهم ادغام می‌گردند. در شیوه تقسیم‌کنندگی در ابتدا کل داده‌ها به عنوان یک خوشه در نظر گرفته شده و تقسیمات متوالی تا رسیدن به تعداد مناسب خوشه‌ها ادامه می‌یابد.

۲- خوشه‌بندی مبتنی بر مدل^۴

در این روش یک مدل احتمالی مشخص برای داده‌ها در نظر گرفته شده و سپس پارامترها برآورد می‌شوند. در این گروه از الگوریتم‌ها یک مدل چگالی آمیخته مطرح می‌شود و فرض می‌شود که داده‌ها از مخلوط شدن تعدادی منبع داده به وجود آمده‌اند. هر یک از این منابع یک خوشه بالقوه در نظر گرفته می‌شود.

۳- خوشه‌بندی مبتنی بر بخش‌بندی^۵

نام دیگر این روش خوشه‌بندی بر مبنای تابع هدف است که در آن اساس کار فرمول‌بندی تابع هدف است. تابع هدف حاصل باید طبیعت مسئله را به خوبی نشان دهد تا بتوان از طریق کمینه سازی آن، ساختار معنی‌داری (خوشه‌ها) را در داده‌های مفروض آشکار ساخت. معروف‌ترین و ساده‌ترین الگوریتم خوشه‌بندی مبتنی بر بخش‌بندی، الگوریتم K-means است. این الگوریتم به خاطر سادگی اجرا، سادگی برنامه و کارایی آن استفاده فراوانی دارد [۱]. عملکرد کلی این روش به این صورت است که هدف ایجاد K خوشه است، بدین ترتیب که عناصر درون هر خوشه نسبت به میانگین رکوردهای

^۱ Hierarchical Clustering

^۲ Divisive

^۳ Agglomerative

^۴ Model-based Clustering

^۵ Partitioning

آن خوشه که مرکز نامیده می‌شود بیش‌ترین شباهت و با مراکز دیگر خوشه‌ها بیش‌ترین عدم شباهت را داشته باشند.

۳-۲-۵-۲ خوشه‌بندی به روش K-Means

این روش، یک روش خوشه‌بندی مبتنی بر بخش‌بندی است که در آن هر خوشه به یک مرکز وابسته است. هر نقطه بسته فاصله خود با هر یک از مراکز، به خوشه‌ای که نزدیک‌ترین فاصله را با مرکز آن دارد مرکز تخصیص می‌یابد. تعداد خوشه‌ها که همان تعداد مراکز است باید از قبل تعیین شده باشد. الگوریتم پایه این روش بسیار ساده است.

۱- الگوریتم با انتخاب K مرکز آغاز می‌شود. این مراکز می‌تواند به طور حدسی یا تصادفی انتخاب گردد.

۲- محاسبه مقدار تابع هدف به صورت زیر است:

$$KM(X, C) = \sum_{i=1}^n \min_{j \in \{1 \dots k\}} \|x_i - c_j\|^2 \quad (2-3)$$

۳- برای هر داده x_i ، عضویت $m(c_j|x_i)$ به ازای هر مرکز c_j و وزن مربوط به آن $w(x_i)$ محاسبه می‌شود.

تابع عضویت به صورت زیر محاسبه می‌شود:

$$m_{KM}(c_i|x_i) = \begin{cases} 1 & ; \text{if } l = \operatorname{argmin}_j \|x_i - c_j\|^2 \\ 0 & ; \text{otherwise} \end{cases} \quad (3-3)$$

تابع وزن عضویت به صورت زیر محاسبه می‌شود:

$$w_{KM}(x_i) = 1 \quad (4-3)$$

۴- برای هر دسته، مجدداً مرکز c_j جدید به صورت زیر محاسبه می‌شود:

$$c_j = \frac{\sum_{i=1}^n m(c_j|x_i) w(x_i) x_i}{\sum_{i=1}^n m(c_j|x_i) w(x_i)} \quad (5-3)$$

۵- گام‌های ۳ و ۴ را تا همگرا شدن دسته‌ها تکرار نمایید [۳۱].

۱. Select K points as the initial centroids.
 ۲. repeat
 ۳. From K clusters by assigning all points to the closest centroid.
 ۴. Recompute the centroid of each cluster.
 ۵. until the centroids don't change
-

۳-۲-۵-۱-۱- مزایای استفاده از الگوریتم خوشه‌بندی K-Means

- در صورت زیاد بودن تعداد متغیرها، این الگوریتم نسبت به روش سلسله‌مراتبی دارای سرعت محاسبات بالاتر است (در صورتی که K کوچک باشد).
- الگوریتم K-Means نسبت به روش سلسله‌مراتبی خوشه‌های کمتری تولید می‌نماید.

۳-۲-۵-۲- محدودیت‌های الگوریتم K-Means

- K-Means زمانی که خوشه‌ها از لحاظ اندازه و چگالی متفاوت و اشکال غیر کروی داشته باشند محدودیت‌هایی خواهد داشت. همچنین این الگوریتم زمانی که داده‌ها حاوی مقادیر پرت باشند با مشکل مواجه خواهد شد.
- علی‌رغم اینکه خاتمه‌پذیری الگوریتم بالا تضمین شده است ولی جواب نهایی آن واحد نبوده و همواره جوابی بهینه نیست. به طور کلی روش ساده بالا دارای مشکلات زیر است:
- جواب نهایی به انتخاب خوشه‌های اولیه وابستگی دارد.
 - روالی مشخص برای محاسبه اولیه مراکز خوشه‌ها وجود ندارد.
 - اگر در تکراری از الگوریتم تعداد داده‌های متعلق به خوشه‌ای صفر شد راهی برای تغییر و بهبود ادامه روش وجود ندارد.
 - در این روش فرض شده است که تعداد خوشه‌ها از ابتدا مشخص است. اما معمولاً در کاربردهای زیادی تعداد خوشه‌ها مشخص نیست [۳۱].

۳-۲-۵-۲- خوشه‌بندی به روش WK-Means^۱

در خوشه‌بندی با الگوریتم K-Means ارزش همه داده‌ها یکسان در نظر گرفته می‌شود که این مسئله یکی از نقاط ضعف این الگوریتم شناخته می‌شود. در روش WK-Means سعی شده با وزن دهی متغیرها بر اساس اهمیت نسبی آن‌ها این ضعف را تعدیل نماید.

در الگوریتم W-K-Means به داده‌ها بر اساس اهمیت وزن داده می‌شود یعنی اگر داده‌هایمان $x_1, x_2, \dots, x_n$ باشد به ترتیب وزن $w_1, w_2, \dots, w_n$ می‌گیرند. این رویکرد می‌تواند تا حد زیادی با مقدار اولیه وزن، کیفیت خوشه‌ای را تحت تأثیر قرار دهد. پس از آماده سازی داده‌ها، وزن داده‌ها برای ارائه اطلاعات بیشتر برای الگوریتم W-K-Means در جهت بهبود دقت طراحی می‌شوند.

مراحل انجام کار در این الگوریتم مانند الگوریتم K-Means است. تنها تفاوت آن‌ها در محاسبه میانگین می‌باشد. در الگوریتم W-K-Means میانگین به صورت زیر محاسبه می‌گردد.

$$\bar{X} = \frac{\sum_{i=1}^k W_i * X_i}{n} \quad (۳-۶)$$

که در آن X_i داده‌های ورودی، K ، تعداد خوشه‌ها و W_i وزن هر یک از این داده‌ها می‌باشد. با این فرمول مرکز خوشه‌ها مشخص می‌شود و داده‌ها در خوشه‌ای قرار می‌گیرد که کمتری فاصله را تا مرکز خوشه‌ها دارد، بقیه مراحل مانند الگوریتم K-Means تکرار می‌شود.

الگوریتم خوشه‌بندی به روش w-K-means

۱. Compute $X_i * W_i$ for all points
 ۲. Select K points as the initial centroids.
 ۳. repeat
 ۴. From K clusters by assigning all points to the closest centroid
 ۵. Recompute the centroid of each cluster (the new centroid is calculated by the formula: $\bar{x} = \frac{\sum_{i=1}^k W_i * X_i}{n}$).
 ۶. until the centroids don't change
-

^۱ Weighted K-Means

۳-۲-۵-۳- خوشه‌بندی به روش^۱ A-H-Means

خوشه‌بندی K-Means یک روش ساده و سریع است که به دلیل پیاده‌سازی آسان و تعداد تکرار کم، عموماً مورد استفاده قرار می‌گیرد. الگوریتم K-Means در تلاش برای یافتن مراکز خوشه‌های $(c_1, c_2, \dots, c_k)$ به گونه‌ای عمل می‌کند که مجموع مربعات فاصله‌ی هر نقطه x_i تا نزدیک‌ترین مرکز خوشه (c_j) کمترین شود. وابستگی کارایی K-Means روی مقداردهی اولیه مراکز، یک مشکل اصلی این الگوریتم می‌باشد. در این الگوریتم ارتباطی قوی بین نقاط داده و نزدیک‌ترین مراکز خوشه برقرار شده و باعث می‌شود مراکز خوشه‌ها از محدوده‌ی تراکم محلی داده‌ها خارج نشوند. روش K-harmonic means این مشکل عمده را از طریق جایگزینی کمترین فاصله یک نقطه از مراکز که در K-Means استفاده می‌شود با میانگین هارمونیک^۲ فاصله هر نقطه تا تمامی مراکز برطرف می‌کند. میانگین هارمونیک یک امتیاز مناسبی را به هر نقطه‌ی داده بر اساس نزدیکی آن به هر مرکز می‌دهد که این امر را به عنوان یک ویژگی میانگین هارمونیک در نظر می‌گیرند.

روش A-H-Means، یک روش میانگین‌گیری بسط یافته است که به ازای مقادیر مختلف Q سایر روش‌های محاسبه میانگین را نیز نتیجه خواهد داد.

نمادهای زیر برای فرمول‌بندی الگوریتم A-harmonic means استفاده می‌شود:

$X = \{x_1, x_2, \dots, x_n\}$: داده‌ای که باید خوشه‌بندی شود.

$C = \{c_1, c_2, \dots, c_k\}$: مجموعه مراکز خوشه‌ها.

Q: متغیر توان که می‌تواند از بازه R انتخاب می‌شود.

الگوریتم پایه برای خوشه‌بندی A-harmonic means مشابه الگوریتم K-Means می‌باشد با این تفاوت که در این روش میانگین توسط فرمول ----- محاسبه می‌گردد.

$$AHM(X, C) = \frac{\sum_{i=1}^n \frac{\|x_i - c_j\|^{Q+1}}{j \in \{1 \dots k\}}}{\sum_{i=1}^n \frac{\|x_i - c_j\|^Q}{j \in \{1 \dots k\}}} \quad (7-3)$$

^۱ Anti-Harmonic means

^۲ harmonic mean

در فرمول بالا در صورتی که $Q=0$ فرمول $\frac{\sum_{i=1}^n \|x_i - c_j\|}{j \in \{1 \dots k\}}$ که همان میانگین حسابی است حاصل خواهد شد و اگر $Q=-1$ میانگین هارمونیک را نتیجه خواهد داد.

الگوریتم خوشه‌بندی به روش A-H-means

۱. Select K points as the initial centroids.
۲. for $Q=-1$ to 5
۳. repeat
۴. From K clusters by assigning all points to the closest centroid
۵. Recompute the centroid of each cluster (the new centroid is calculated by the formula: $AHM(X, C) = \frac{\sum_{i=1}^n \|x_i - c_j\|^{Q+1}}{\sum_{i=1}^n \|x_i - c_j\|^Q}$).
۶. until the centroids don't change

۳-۲-۶- ارزیابی خوشه‌ها به روش مجموع مربع خطاها^۱ و انتخاب بهترین روش

پس از اینکه خوشه‌بندی با روش‌های K-Means، WK-Means و A-H-Means انجام گرفت لازم است نتایج حاصل از هر یک از سه روش مقایسه گردد تا از نتایج بهترین الگوریتم استفاده گردد. هدف از اعتبارسنجی خوشه‌ها یافتن خوشه‌هایی است که بهترین تناسب را با داده‌های مورد نظر داشته باشند. برای سنجش همگنی درون خوشه‌ای و ناهمگنی بین خوشه‌ای تعاریف متفاوتی بر اساس فاصله اعضا وجود دارد که در زیر به تفصیل بررسی گردیده است. سپس با در نظر گرفتن این تعاریف معیارهایی برای سنجش کیفیت خوشه‌بندی تعریف شده است. یکی از رایج‌ترین معیارهای ارزیابی خوشه‌ها مجموع مربع خطاهاست. روش محاسبه SSE به صورت زیر است:

• برای هر نقطه خطا فاصله آن تا نزدیک‌ترین خوشه است.

• برای محاسبه SSE، مربع فاصله هر نقطه را حساب کرده و باهم جمع می‌نماییم.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \sum_{j=1}^n (m_j^i - x_j)^2 \quad (۸-۳)$$

^۱ Sum of Squared Error (SSE)

نمادهای بکار رفته در رابطه ۳-۸ به شرح زیر می باشد:

x : یک نقطه در خوشه C_i

x_j : آمین ویژگی x

m_j^i : آمین ویژگی نقطه نماینده خوشه C_i

می توان نشان داد که m_j^i ، با مرکز خوشه متناسب است.

- از بین دو خوشه، خوشه ای که SSE کمتری دارد انتخاب می شود. یک راه ساده جهت کاهش SSE، افزایش تعداد خوشه ها (K) است. که البته خوشه بندی خوب خوشه بندی ای است که با K کوچک تر بتواند SSE کمتری از یک خوشه بندی ضعیف با K بالاتر داشته باشد [۳۱].

کیفیت خوشه ها و میزان خطای خوشه بندی در این روش تا حد زیادی به انتخاب مراکز اولیه وابسته است.

در مورد انتخاب مراکز اولیه مشکلاتی وجود دارد. در صورتی که K خوشه واقعی داشته باشیم شانس انتخاب یک مرکز از هر خوشه بسیار کم است و هر چقدر K بزرگ تر باشد این احتمال کمتر است. اگر خوشه ها اندازه های مساوی و برابر n داشته باشند، در این صورت این احتمال از فرمول ۳-۹ محاسبه خواهد شد:

$$p = \frac{\text{number of ways to select one centroid from each cluster}}{\text{number of ways to select } K \text{ centroids}} \quad (9-3)$$

$$= \frac{K! n^K}{(Kn)^K} = \frac{K!}{K^K}$$

برای مثال اگر $K=10$ باشد این احتمال برابر است با: $0.00036 = 10!/10^{10}$. گاهی اوقات نقاط مرکز خود را در مسیر درست اصلاح می کنند و گاهی خیر.

پیشنهاداتی برای حل مشکل مراکز اولیه:

- اجرای چند باره: این روش گاهی اوقات می تواند نتیجه بخش باشد اما همیشه این طور نخواهد بود.

- نمونه برداری و استفاده از خوشه بندی سلسله مراتبی برای تشخیص نقاط مرکزی.
- انتخاب بیش از K مرکز اولیه در بازه ی گسترده تر و سپس انتخاب بهترین نقاط از بین نقاط انتخابی.
- انجام عملیات پس پردازش [۳۱]

۳-۲-۷- به کارگیری دانش حاصل از خوشه بندی

پس از بخش بندی مشتریان می توان با بررسی خوشه های حاصل و با کمک نظرات خبرگان بانک و با استفاده از پرو فایل مشتریان بانک مهر اقتصاد می توان به مطالعه و استخراج دانش این خوشه ها پرداخت و یا به رتبه بندی خوشه های مشتریان اقدام نمود. دانش حاصل از این مرحله راهنمای بسیار خوبی جهت اتخاذ راهبردهای بهینه متناسب با هر خوشه برای مدیریت هر چه بهتر ارتباط با مشتریان خواهد بود.

به طور معمول مشتریانی که در یک خوشه قرار می گیرند دارای ویژگی های رفتاری مشابهی خواهند بود.

۳-۳- روش های جمع آوری اطلاعات

منابع مورد استفاده در این تحقیق شامل بررسی تحقیقات و پایان نامه های موجود و مرتبط با موضوع، مطالعه مقالات و کتب فارسی و انگلیسی، جستجوی موضوعات مرتبط در پایگاه های داده علمی و اینترنت می باشد. در این تحقیق برای جمع آوری داده ها از داده های موجود در پایگاه اطلاعاتی بانک مهر اقتصاد استفاده شده است.

۳-۴- جمع بندی مطالب فصل

در این فصل چارچوب کلی طرح پیشنهادی مطرح شده و جزئیات گام ها و روش های مورد استفاده در هر گام مورد بررسی قرار گرفت. برای مثال روش های WK-Means, K-Means و Anti Harmonic Means جهت خوشه بندی داده ها انتخاب شده و معیار SSE به عنوان ملاکی جهت مقایسه نتایج حاصل از اجرای هر یک از این الگوریتم ها بر روی مجموعه داده های مورد نظر خواهد استفاده خواهد شد. هر چه میزان SSE خوشه ها کمتر باشد نشان دهنده عملکرد بهتر الگوریتم خواهد بود.

در نهایت، دانش موجود در خوشه‌های حاصل از بهترین روش انتخابی، در مسیر مطالعه و اخذ تصمیماتی جهت مدیریت هر چه بهتر ارتباط با مشتری، به کمک مدیران سازمان خواهد شتافت.

فصل چهارم

محاسبات و یافته‌های تحقیق

۴-۱- مقدمه

جایگاه مشتری در فضای رقابتی کسب و کار بر کسی پوشیده نیست. به ویژه در نظام بانکی مشتری، نقش بسیار مؤثری در مسیر نیل به اهداف سازمان ایفا می‌کند. این نقش در بانک‌های خصوصی بسیار مورد توجه قرار گرفته و این بانک‌ها مطالعات و فعالیت‌های ویژه‌ای را در این زمینه اختصاص داده‌اند. در این تحقیق سعی شده تا با استفاده از روش‌های داده‌کاوی دانش نهفته در اطلاعات موجود در پایگاه داده‌های حاصل از تعاملات بانک و مشتری را استخراج نموده و از آن در جهت اخذ راهبردهای مدیریت ارتباط با مشتری بهره‌گیری نماییم.

در فصل سوم چارچوب کلی روش پیشنهادی و الگوریتم‌های انتخابی برای پیاده‌سازی گام‌های این روش تشریح گردید.

در این فصل بر طبق چارچوب ارائه شده در فصل قبل، سعی شده تا با اعمال الگوریتم‌ها بر روی پایگاه داده‌های مشتریان بانک مهر اقتصاد، در جهت حصول اهداف تحقیق که همان بخش‌بندی مشتریان و استخراج راهبردهای متناسب با هر بخش و به‌کارگیری راهبردها و دانش حاصل در چرخه مدیریت دانشی سازمان می‌باشد، گام برداریم.

در این تحقیق روش Min-Max برای نرمال سازی داده‌ها، روش‌های K-Means، WK-Means، A-Harmonic means جهت خوشه‌بندی داده‌ها و شاخص مجموع مربع خطاها برای ارزیابی و اعتبارسنجی خوشه‌ها به کار گرفته خواهند شد.

مورد مطالعه در این تحقیق مشتریان بانک مهر اقتصاد می‌باشند.

۴-۲- معرفی بانک مهر اقتصاد

موسسه مالی و اعتباری مهر (موسسه قرض الحسنه بسیجیان سابق) با ایجاد زمینه‌های مساعد جهت رشد اقتصادی جامعه و ارتقاء سطح رفاه اجتماعی و رفع نیاز آحاد ملت به ویژه بسیجیان در تاریخ ۱۳۷۲/۹/۲۳ با سرمایه‌ای به ارزش ده میلیون ریال آغاز به کار کرد و از شروع فعالیت با توجه به نیاز مبرم کشور به تحول در زمینه سیستم بانکداری سنتی، ارائه خدمات بانکداری الکترونیکی را سرلوحه فعالیت بانکی و پولی خود قرار داده است. موسسه مالی و اعتباری مهر اولین موسسه مالی و پولی در کشور در زمینه ارائه سرویس‌های بانکداری مدرن، چون اینترنت بانک، موبایل بانک، دستگاه

خودپرداز^۱، تلفن بانک، خدمات پایانه‌های فروش^۲، کارت هوشمند و ارایه سیستم‌های برخط^۳ بانکی و پولی می‌باشد.

۴-۳- موضوع و فعالیت بانک

موسسه مالی و اعتباری مهر با عنایت به مسوولیت اجتماعی خود و حضور موثر در عرصه‌های اقتصادی (پولی و بانکی) ضمن توسعه شعب خود در اقصی نقاط کشور (نزدیک به هفتصد شعبه) خدمات گوناگونی در حمایت از پروژه‌های مولد و اشتغال‌زا، ازدواج، درمان، تحصیل و ... ارایه نموده و همچنین در توسعه خدمات الکترونیک گام‌های اساسی برداشته است. همچنین خدمات دیگری شامل افتتاح انواع سپرده‌های سرمایه‌گذاری و سود مناسب، اعطای تسهیلات در بخش‌های صنعتی و تولیدی، مسکن، خرید خودرو و لوازم خانگی و خدمات اعتباری به مشتریان خود ارایه می‌نماید. موسسه مالی و اعتباری مهر در راستای توسعه فناوری اطلاعات اقداماتی از قبیل راه‌اندازی سایت مرکزی، ایجاد بستر فناوری اطلاعات برای هفتصد شعبه تحت پوشش، ایجاد بانکداری یکپارچه^۴ با ظرفیت بالا و ایمن، راه‌اندازی سایت اطلاع رسانی و تجهیز دویست و شانزده شعبه به دستگاه خودپرداز را انجام داده است. سیستم جامع اتوماسیون بانکی موسسه در حال حاضر بزرگترین شعبه اتوماسیون در بین بانک‌های خصوصی محسوب شده و قابلیت‌های فراوانی از قبیل: اینترنت بانک، موبایل بانک، اس ام اس بانک، تلفن بانک، ایمیل بانک و پایانه‌های فروش را دارا می‌باشد.

از دیگر قابلیت‌های موسسه رویکرد بانکداری اسلامی است و تمامی معاملات بانکی موسسه به منظور تطابق با قوانین شرع مقدس اسلام مورد نظارت بوده و تاییدیه‌های شرعی مربوطه به صورت دوره‌ای اخذ می‌گردد.

۴-۴- محاسبات تحقیق

در این فصل گام‌های مربوط به چارچوب طرح شده در فصل سوم این تحقیق، در مطالعه موردی بانک مهر اقتصاد به کار گرفته خواهد شد.

^۱ ATM

^۲ POS

^۳ Online

^۴ Core Banking

۴-۴-۱- گام انتخاب و جمع آوری متغیرهای ورودی

در حوزه بخش‌بندی رفتاری مشتریان بانک و بررسی رویکردهای سازمانی بانک مهر اقتصاد، سه دسته متغیر داده‌های جمعیت شناختی مشتریان شامل جنسیت، سن، تحصیلات، وضعیت تأهل، شغل و تاریخ تولد، محل منطقه بانکی و ... داده‌های مربوط به تراکنش‌های مالی مشتریان نظیر تراکنش اخیر مشتری (R)، متغیر پولی (F)، فراوانی (M) و فاکتور (K) که نشان‌دهنده کلیدی بودن یا نبودن مشتری بانک است و در بانک مهر اقتصاد مشتریان کلیدی بانکی در هر استان بنا بر نظر خبرگان بانکی با ویژگی‌های خاصی تعریف می‌گردند، انتخاب شده است.

در جدول ۴-۱ نمونه‌ای ده‌تایی از این داده‌های خام که نشان‌دهنده اطلاعات مربوط به مشتریان بانک مهر اقتصاد می‌باشد، نمایش داده شده است.

جدول ۴-۱ نمونه ده‌تایی از داده‌های مربوط به مشتریان بانک مهر اقتصاد

ID	Sex	Age	Education	Marital Status	Job	Location	R	F	M	K
۱	مرد	۵۴	کارشناسی	متاهل	فرهنگی	بابل	۱۳۵	۲۶	۳۴۰۰۰۰	خیر
۲	مرد	۶۹	سیکل	متاهل	آزاد	بابل	۱۵	۲	۶۹۹۵۲۱	خیر
۳	مرد	۲۹	دیپلم	متاهل	کشاورز	بابل	۳۱۱	۷	۲۹۹۳۱۹۹۲۵	خیر
۴	مرد	۲۳	کارشناسی ارشد	مجرد	آزاد	بابل	۴۸	۲۹	۸۲۸۰۱۰۰۰۰	خیر
۵	مرد	۴۱	دیپلم	متاهل	آزاد	بابل	۲۰	۳۹	۲۳۴۵۴۳۱۱۵	خیر
۶	زن	۳۴	کارشناسی	متاهل	کارمند	بابل	۶۵	۱۱	۶۶۳۰۹۷۲۸	خیر
۷	مرد	۴۱	دیپلم	متاهل	آزاد	بابل	۱۸	۱۵۰	۲۲۴۶۸۰۷۳۱۷۶	خیر
۸	مرد	۴۴	دیپلم	متاهل	آزاد	بابل	۷	۲۱	۵۲۰۵۲۱۳۹	خیر
۹	مرد	۴۳	کارشناسی	متاهل	نمایشگاه اتومبیل	بابل	۲	۸۰	۵۱۲۹۶۶۵۰۱۵	بله
۱۰	زن	۶۷	بی‌سواد	متاهل	خانه‌دار	بابل	۳۴	۶	۸۵۶۶۲۵۴۳	خیر

در این تحقیق به جهت حفظ اطلاعات مشتریان بانکی شماره شناسایی مشتریان با شماره‌های قراردادی شماره‌گذاری شده است.

۴-۴-۲- گام آماده‌سازی و پیش‌پردازش داده‌ها

داده‌های این تحقیق از داده‌های موجود در پایگاه اطلاعاتی بانک مهر اقتصاد جمع آوری شده است. از آنجایی که داده‌های خام انتخاب مناسبی برای اعمال الگوریتم‌ها نیستند نیاز است عملیاتی بر روی داده‌ها جهت آماده‌سازی آن‌ها برای ورود به الگوریتم‌ها صورت گیرد. از جمله این عملیات نرمال داده‌ها می‌باشد.

از میان یازده متغیر شناسایی شده جهت مطالعه رفتار مشتریان، متغیرهای تراکنش اخیر مشتری و متغیر کلیدی بودن مشتری به عنوان ورودی الگوریتم‌های خوشه‌بندی در نظر گرفته می‌شوند. اما از آنجایی که ویژگی‌های انتخابی از نمونه‌های مورد مطالعه در این تحقیق دارای بازه تغییرات یکسان نمی‌باشند، برای جلوگیری از تأثیر متفاوت ویژگی‌ها (از آنجایی که متغیرهای با مقادیر بزرگ تأثیر بیشتری بر تابع هدف می‌گذارند که لزوماً به معنی مهم‌تر بودن آن‌ها در الگوریتم بخش‌بندی نیست) از نرمال سازی به روش Min-Max که پیش از این شرح داده شد استفاده شده است.

به طور مثال جدول ۴-۲ یک نمونه ده‌تایی از متغیرهای نرمال شده R ، F ، M و K از داده‌های مورد مطالعه در این تحقیق را نشان می‌دهد.

جدول ۴-۲ متغیرهای نرمال شده

ID	R	F	M	K
۴۱۱	۰,۰۳۵۲۹۴	۰	۰,۳۹۴۷۲۲۸۵۲	۰
۴۱۲	۰,۸	۰,۶۸۹۶۵۵۱۷۲	۰,۴۶۶۰۱۰۷۱	۰
۴۱۳	۰,۲۰۵۸۸۲	۰	۰,۷۷۰۶۱۶۲۹۵	۰
۴۱۴	۰,۱	۰,۶۲۰۶۸۹۶۵۵	۰,۹۰۱۹۳۰۱۶	۰
۴۱۵	۰,۰۷۰۵۸۸	۰	۰,۲۱۳۲۶۵۹۱۱	۰
۴۱۶	۰,۳۵۸۸۲۴	۰,۸۱۰۳۴۴۸۲۸	۰,۸۱۶۷۸۵۷۲۶	۰
۴۱۷	۰,۳۶۴۷۰۶	۰,۲۴۱۳۷۹۳۱	۰,۹۱۹۶۲۴۹۹۷	۱
۴۱۸	۰,۴۲۳۵۲۹	۰	۰,۹۸۰۸۳۹۷۲۹	۰

۰	۰,۱۱۱۲۶۲۰۵	۰,۳۷۹۳۱۰۳۴۵	۰,۴۹۴۱۱۸	۴۱۹
۰	۰,۶۱۱۸۹۳۰۶۵	۰,۴۳۱۰۳۴۴۸۳	۰,۴۴۷۰۵۹	۴۲۰

۴-۴-۳- گام تعیین تعداد بهینه خوشه‌ها

تعداد بهینه خوشه‌ها می‌تواند با توجه به اهداف تحقیق و دیدگاه‌های مختلف خبرگان متفاوت باشد. در این تحقیق با استفاده از مجموعه داده‌های مورد استفاده و بررسی نظرات خبرگان بانک مهر اقتصاد بهترین گزینه برای تعیین تعداد مناسب خوشه‌ها، جهت تفکیک رفتاری مشتریان این بانک عدد ۵ به دست آمد. این عدد به عنوان یک ورودی در الگوریتم‌های خوشه‌بندی مورد نظر بکار گرفته خواهد شد.

۴-۴-۴- گام خوشه‌بندی داده‌ها

خوشه‌بندی به معنای تقسیم موجودیت‌ها در گروه‌هاست به طوری که موجودیت‌هایی که در یک گروه قرار می‌گیرند بیش‌ترین شباهت را با هم‌گروه‌های خود و بیش‌ترین فاصله را با موجودیت‌های دیگر گروه‌ها داشته باشند.

به طور خاص در صنعت بانکداری از الگوریتم‌های خوشه‌بندی در بخش‌بندی مشتریان به گروه‌های مشخص بر طبق الگوهای رفتاری آن‌ها کاربرد دارد. شناسایی گروه‌های مشتریان بسیار برای بانک حائز اهمیت بوده و می‌تواند در تصمیم‌گیری‌های کنونی و اتخاذ راهبردهای میان‌مدت و بلندمدت بانک راهگشا باشد.

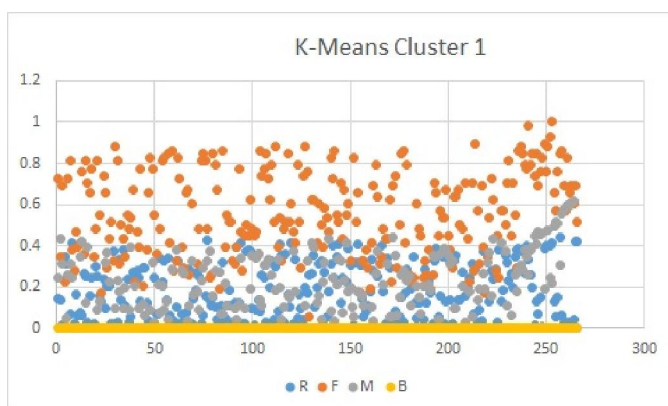
همان طور که در فصل سوم بیان شد برای تفکیک مشتریان در پنج بخش، سه الگوریتم K-Means، WK-Means و A-H-Means بکار گرفته خواهند شد. کدهای مربوط به این الگوریتم‌ها به زبان C#.net نوشته شده است.

ورودی این الگوریتم‌ها فایل اکسل حاوی نمونه داده‌های مربوط به هزار مشتری بانک مهر اقتصاد استان مازندران، شهرستان بابل می‌باشد. این الگوریتم‌ها بر روی چهار متغیر R، F، M و K از این نمونه‌ها اعمال می‌شود.

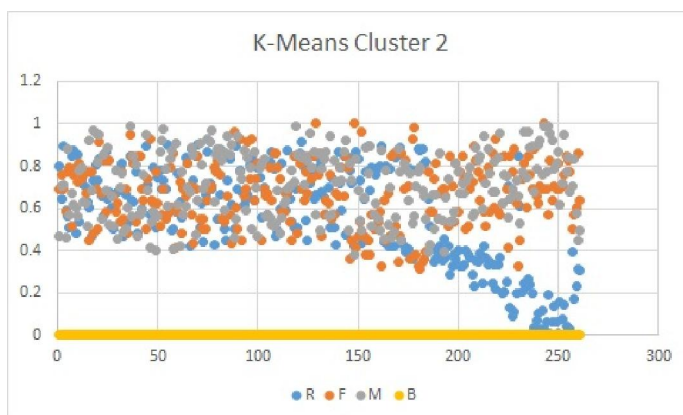
۴-۴-۱- خوشه‌بندی به روش K-Means

هدف از اجرای این الگوریتم که یکی از الگوریتم‌های متنوع خوشه‌بندی می‌باشد، تقسیم نمونه‌ها در K خوشه است. K مرکز اولیه به صورت تصادفی از بین داده‌های ورودی انتخاب می‌شوند. سپس فاصله (در این تحقیق فاصله اقلیدسی) هر داده با هر یک از K مرکز محاسبه شده و داده به خوشه‌ای که کمترین فاصله را با مرکز آن داراست تخصیص می‌یابد. بعد از تخصیص تمام نقاط به K مرکز، میانگین هر خوشه به عنوان مرکز جدید محاسبه شده و محاسبه فاصله عناصر و تخصیص آن‌ها به مراکز جدید ادامه می‌یابد تا جایی که هیچ جابجایی در عناصر خوشه‌ها رخ ندهد.

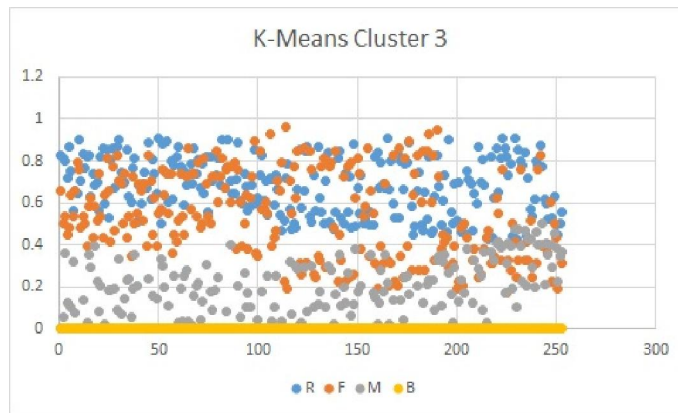
شکل ۴-۱ تا ۴-۵ نمودارهای مربوط به پنج خوشه خروجی الگوریتم K-Means را نشان می‌دهد.



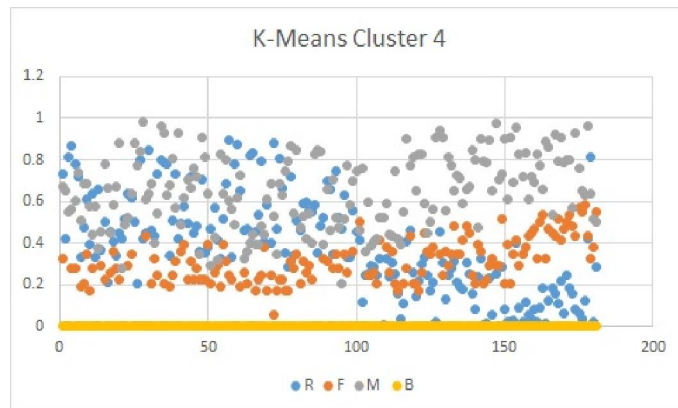
شکل ۴-۱ خوشه اول، الگوریتم K-Means



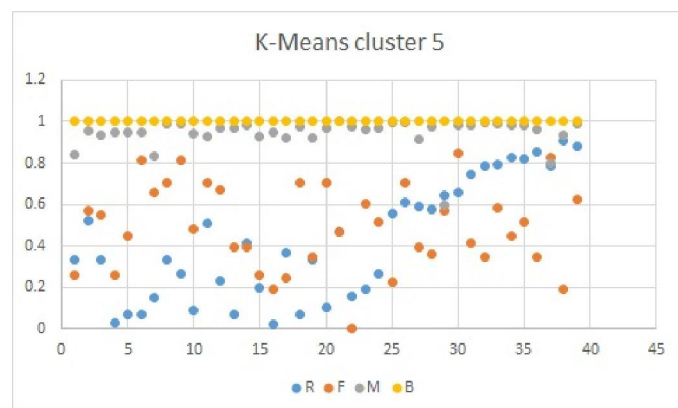
شکل ۴-۲: خوشه دوم، الگوریتم K-Means



شکل ۳-۴ خوشه سوم، الگوریتم K-Means



شکل ۴-۴ خوشه چهارم، الگوریتم K-Means



شکل ۵-۴ خوشه پنجم، الگوریتم K-Means

۴-۴-۲- خوشه‌بندی به روش WK-Means^۱

در خوشه‌بندی به روش WK-Means، اهمیت متغیرهای شرکت‌کننده در خوشه‌بندی متفاوت در نظر گرفته می‌شود و به هر یک از متغیرها وزن تخصیص می‌یابد.

این وزن‌ها بر اساس اهمیت نسبی متغیرها در نظر گرفته می‌شود. به این منظور در این تحقیق این وزن‌ها بر اساس نظر خبرگان بانکی بانک مهر اقتصاد محاسبه شده است. این وزن‌ها در جدول ---- آمده است.

جدول ۴-۳ وزن نسبی متغیرهای تحقیق

متغیر	R	F	M	K
ضریب وزنی	۰,۱۲	۰,۲۲	۰,۳۲	۰,۳۴

بعد از تعیین وزن نسبی و محاسبه حاصل ضرب آن در متغیرها، الگوریتم K-Means بر روی متغیرها اعمال می‌گردد.

در ادامه قطعه کد مربوط به خوشه‌بندی به روش‌های K-Means و WK-Means آورده شده است. قابل ذکر است در خوشه‌بندی به روش K-Means وزن چهار متغیر برابر در نظر رفته می‌شود.

۴-۴-۳- خوشه‌بندی به روش A-H-Means^۲

A-H-Means یکی دیگر از الگوریتم‌های پیشنهادی جهت خوشه‌بندی داده‌های این تحقیق است که جزئیات این روش در فصل سوم به تفصیل مطرح شد. در این کاربرد $Q \in [-۵, ۵]$ انتخاب شده است که در این صورت $Q=۰$ میانگین حسابی و اگر $Q=-۱$ میانگین هارمونیک را نتیجه خواهد داد. کدهای مربوط به این الگوریتم‌ها در پیوست ۱ آمده است.

^۱ Weighted K-Means

^۲ Anti-Harmonic means

۴-۴-۵- ارزیابی خوشه‌ها به روش مجموع مربع خطاها^۱ و انتخاب بهترین روش

در این گام لازم است مجموع مربع خطاهای خوشه‌های حاصل از روش‌های K-Means، WK-Means و A-H-Means محاسبه شود تا روشی که کمترین مجموع مربع خطاها را داراست به عنوان روش منتخب گزیده شود و خوشه‌های حاصل از روش مذکور جهت استخراج دانش در گام بعد مورد مطالعه قرار گیرند.

با توجه مقادیر SSE حاصل در جدول ۴-۵ در این مرحله روش K-Means و معادل آن A-H-Means بهترین تناسب را با داده‌های مورد نظر داشته‌اند.

جدول ۴-۵ مقادیر مجموع مربع خطاها در الگوریتم‌های مختلف خوشه‌بندی

الگوریتم	SSE
K-Means	۹۹,۹۸۴۳۶۲۴۱۷۱۵۲۸
W-K-Means	۱۰۱,۷۰۳۹۶۵۷۰۸۰۲
A-H-Means	Q= -۵
	۵۷۸,۸۴۶۳۳۹۶۵۸۴۷۸
	Q= -۴
	۴۵۱,۹۹۸۶۵۶۶۲۳۸۵۲
	Q= -۳
	۴۱۳,۶۳۳۹۵۱۸۷۲۵۳۹
	Q= -۲
	۴۵۲,۳۰۱۲۶۹۲۱۰۵۸۲
	Q= -۱
	۴۲۹,۵۳۴۷۹۴۰۲۰۳۱۲
W-A-H-Means	Q= ۰
	۹۹,۹۸۴۳۶۲۴۱۷۱۵۲۴
	Q= ۱
	۲۶۳,۰۳۲۹۵۴۲۹۷۷۴۹
	Q= ۲
	۳۸۷,۲۹۶۸۱۵۵۰۰۴۳۴
	Q= ۳
	۴۶۱,۳۷۱۴۰۲۵۱۶۸۴
	Q= ۴
	۵۲۰,۴۳۱۷۶۶۰۳۰۳۳۶
W-A-H-Means	Q= ۵
	۵۶۳,۳۴۳۱۹۸۱۹۹۵۶۸
W-A-H-Means	Q= -۵
	۹۰۹,۹۹۴۹۰۵۸۷۴۲۰۴
W-A-H-Means	Q= -۴
	۹۰۸,۵۴۶۵۳۹۳۵۸۳۳۱

^۱ Sum of Squared Error (SSE) Q=

۸۱۲,۹۸۷۱۲۳۵۵۴۹۵۹	Q= -۳
۶۷۵,۲۴۴۸۶۷۰۷۲۶۶۹	Q= -۲
۴۳۲,۹۹۱۶۳۸۷۲۲۰۷	Q= -۱
۱۰۱,۷۰۳۹۶۵۷۰۸۰۲	Q= ۰
۲۱۳,۱۰۹۵۸۹۶۱۰۲۹۲	Q= ۱
۳۹۴,۶۳۱۴۰۸۱۶۹۲۲۳	Q= ۲
۴۶۵,۴۹۳۸۳۳۴۹۷۸۷۴	Q= ۳
۵۲۰,۳۴۱۰۷۲۵۲۰۵۹۴	Q= ۴
۵۶۷,۳۹۸۳۲۶۶۳۳۸۲۲	Q= ۵

۴-۴-۶- گام به کارگیری دانش حاصل از خوشه‌بندی

پس از اینکه مناسب‌ترین الگوریتم برای خوشه‌بندی داده‌های مربوط به مشتریان بانک مهر اقتصاد شناسایی شده و مشتریان در پنج خوشه بخش‌بندی شدند نوبت به بهره‌گیری از دانش حاصل از این خوشه‌ها می‌رسد. در این مرحله با حضور خبرگان بانکی و با مراجعه به پروفایل مشتریان بانک می‌توان به بررسی و تحلیل خوشه‌ها پرداخت و از آنجا که به طور معمول مشتریانی که در یک خوشه قرار می‌گیرند دارای ویژگی‌های رفتاری مشابهی خواهند بود، دانش استخراج شده می‌تواند راهنمای بسیار خوبی جهت اتخاذ راهبردهای بهینه متناسب با هر خوشه برای مدیریت هر چه بهتر ارتباط با مشتریان باشد.

جدول ۴-۶ که مطابق نظر خبرگان بانک تنظیم شده است می‌تواند برای بررسی خوشه‌های مشتری مورد استفاده قرار گیرد.

جدول ۴-۶ دسته‌بندی مشتریان بر مبنای ویژگی‌های رفتاری مشابه [۱]

وضعیت مشتریان	M	F	R	ردیف
مشتریانی ارزشمند با احتمال فرار بالا	↑	↑	↑	۱
تعداد گردش بالا اما اخیراً گردش حسابی نداشته‌اند.	↓	↑	↑	۲
این گروه از مشتریان در بالاترین رتبه قرار دارند.	↑	↑	↓	۳
مشتریانی نسبتاً ارزشمند ولی با فعالیت حساب پایین (به طور معمول مشتریان جدید)	↑	↓	↓	۴
مشتریانی وفادار با مانده حساب پایین (مشتریان قدیمی‌تر)	↓	↑	↓	۵
مشتریانی با متغیر پولی بالا اما فعالیت حساب پایین (وفاداری کمتر)	↑	↓	↑	۶
مشتریان جدید یا مشتریانی که اخیراً به بانک برگشته‌اند.	↓	↓	↓	۷
تنها در صورت افزایش فعالیت حساب می‌توانند در رتبه‌های بالاتر قرار گیرند.	↓	↓	↑	۸

به عنوان مثال در رابطه با مشتریان ردیف ۷ مشتریانی هستند که به تازگی مشتری بانک شده‌اند یا بعد از مدتی نسبتاً طولانی مجدد به بانک مراجعه نموده‌اند. این مشتریان در صورتی می‌توانند تبدیل به مشتریان ارزشمند تر شوند که مانده حساب و تعداد گردش‌های حساب آنان افزایش یابد. بانک باید مشتریانی که سعی در برقراری ارتباط بیشتر با بانک و استفاده بیشتر از خدمات بانک را دارند شناسایی

نموده و تنها گروهی که قابل تبدیل شدن به مشتریانی ارزشمندتر هستند را حفظ نماید. چرا که با توجه به تعداد بسیار مشتریان این خوشه، نگهداری همه این مشتریان می‌تواند بانک را متحمل هزینه بالایی نماید.

بانک می‌تواند با معرفی سایر خدمات خود خوشه ۸ را به فعال‌تر شدن حسابشان تشویق نماید. مشتریان خوشه ۱ بالاترین ارزش را نزد بانک دارند اما به دلیل بالا بودن مدت زمان گذشته از آخرین تراکنش حسابشان با بانک احتمال فرار این گروه از مشتریان به سمت بانک‌های دیگر وجود دارد. بانک باید با ارائه خدمت رسانی سریع‌تر و بهتر به این دسته از مشتریان، این احتمال را کاهش دهند [۱].

جدول ۴-۷ بخشی از اطلاعات حاصل از خوشه‌بندی نمونه هزارتایی از داده‌های مربوط به مشتریان بانک مهر اقتصاد را نشان می‌دهد.

جدول ۴-۷ اطلاعات مربوط به خوشه‌بندی مشتریان بانک مهر اقتصاد به روش K-Means

میانگین متغیرها در خوشه				تعداد عناصر	
K	M	F	R	خوشه	
۰	۰.۱۷۱۹۷۳	۰.۵۳۸۳۷۲	۰.۱۸۶۲	۲۶۶	خوشه اول
۰	۰.۷۱۱۵۲۹	۰.۶۶۹۶۳۹	۰.۵۵۹۲۷	۲۶۱	خوشه دوم
۰	۰.۱۴۹۷۷۶	۰.۵۲۲۳۵۲	۰.۶۸۰۲۴	۲۵۳	خوشه سوم
۰	۰.۶۴۹۲۷۲	۰.۲۵۹۷۶۴	۰.۳۸۷۰۳	۱۸۱	خوشه چهارم
۱	۰.۹۴۴۴۷۵	۰.۴۹۱۱۵۸	۰.۴۲۶۰۹	۳۹	خوشه پنجم

۴-۵- نتایج تحقیق

با مقایسه جدول ۴-۷ با جدول راهنمای ۴-۶ می‌توان دریافت که خوشه اول که مقدار متغیر R آن پایین (یعنی اخیراً به بانک مراجعه داشته‌اند)، متغیر F یعنی تعداد تراکنش اخیر آن‌ها طی یک سال گذشته آن نسبتاً بالا و متغیر M یعنی میانگین مبلغ تراکنش‌های آن‌ها در طی یک سال گذشته نسبتاً پایین است، مشتریانی هستند وفادار با مانده حساب پایین (مشتریان قدیمی‌تر) و بانک در مواجهه با این گروه

باید به دنبال جواب این سؤال باشد که چطور می‌توان این گروه از مشتریان را که وفاداری نسبتاً بالایی به بانک دارند به سرمایه‌گذاری بیشتر و افزایش موجودی و فعالیت حساب از طرق مختلف در بانک ترغیب نمود. ارائه تسهیلات، می‌تواند پیشنهادی مناسب جهت نگهداشت مشتریان این گروه باشد.

خوشه دوم که مقدار R ، F و M نسبتاً بالایی دارند، از آنجاییکه میانگین مبلغ تراکنش‌ها و تعداد تراکنش بالایی دارند، مشتریانی ارزشمند برای بانک محسوب می‌شوند. اما از آنجاکه فاصله زمانی آخرین تراکنش آن‌ها با بانک عدد به نسبت بزرگی است، احتمال فرار این مشتریان بالاست. بانک باید با ارائه خدمت رسانی سریع‌تر و بهتر به این دسته از مشتریان، این احتمال را کاهش داده و جهت جلب اعتماد مجدد آن‌ها تلاش نماید.

در مورد مشتریان خوشه سوم می‌توان گفت این مشتریان با وجود تعداد گردش بالا در یک سال گذشته، اخیراً گردش حسابی نداشته‌اند. در مورد مشتریان این خوشه و خوشه دوم می‌توان گفت باید به روش‌هایی مانند معرفی طرح‌های جدید به این مشتریان سعی در جذب مجدد آن‌ها نمود. البته مشتریان خوشه دوم به دلیل داشتن میانگین مبلغ بالاتر نسبت به مشتریان این خوشه از ارزش بالاتری برخوردارند.

مشتریان خوشه سوم مشتریانی نسبتاً ارزشمند ولی با فعالیت حساب پایین و به طور معمول مشتریان جدید هستند که میزان نسبتاً بالای تبادل نقدینگی با بانک را دارا می‌باشند. بانک می‌تواند با معرفی کامل کلیه خدمات بانکی ارائه شده به روش‌های مختلف مانند ارائه کاتالوگ، ارسال پیامک، ایمیل و ... مشتری جدید را در انتخاب خدمات متناسب و در نتیجه بقای ارتباط با این بانک یاری نماید.

مقدار میانگین ۱ مربوط به متغیر K در خوشه پنجم نشان می‌دهد کلیه مشتریان این خوشه جزء مشتریان کلیدی بانک می‌باشد. با گسترش دامنه مطالعه به سایر متغیرهای مربوط به این مشتریان می‌توان به گزاره‌های دانشی ارزشمندی دست یافت. مثلاً اکثریت افراد این خوشه مرد، متأهل و دارای شغل آزاد می‌باشند. این دانش در مورد ارزشمندترین مشتریان بانک می‌تواند در زمینه اخذ تصمیماتی جهت افزایش روزافزون رضایت این دسته از مشتریان و ممانعت از رویگردانی آن‌ها از این بانک بسیار کمک کننده باشد. یکی از این رویکردها که می‌تواند در ارتقاء رضایت این مشتریان تأثیرگذار باشد، ارسال پیامک‌هایی خاص این گروه‌ها و معرفی خدمات جدید به قصد افزایش بازدهی حساب‌های آنان در

نزد این بانک می‌باشد. توسعه خدمات جدید متناسب با مطالبات این گروه می‌تواند موجب بقای ارتباط آنها با بانک بوده و قطعاً رضایت مشتری در دراز مدت سودآوری زیادی را برای بانک به همراه خواهد داشت.

۴-۶- جمع‌بندی مطالب فصل

در این تحقیق یک نمونه هزارتایی از داده‌های مربوط به تراکنش‌های اخیر مشتریان با بانک مهر اقتصاد از پایگاه داده‌های این بانک موجود استخراج شده و با انجام عملیات پیش‌پردازشی مانند نرمال، جدول داده‌ها، حاوی مقادیر نرمال شده متغیرهای F ، R ، M و K در قالب جدول اکسل، به عنوان ورودی الگوریتم‌های K -Means، WK -Means و Anti Harmonic Means آماده‌سازی شد. سپس به کمک معیار SSE ، خوشه‌های به دست آمده از روش K -Means که دارای کمترین مقدار SSE بودند، به عنوان مناسب‌ترین خوشه‌ها شناسایی و به گام نهایی یعنی گام تحلیل خوشه‌ها، تحویل شد. و در نهایت این خوشه‌ها با کمک نظرات خبرگان بانکی مورد تحلیل قرار می‌گیرد. دانش حاصل می‌تواند پشتیبانی قدرتمند برای مدیران سازمان جهت اتخاذ راهبردهایی جهت بهبود ارتباط با مشتری باشد.

فصل پنجم

نتیجه‌گیری و پیشنهادها

۵-۱- مقدمه

هرچند ممکن است استخراج دانش از حجم عظیم داده‌های برجای مانده از تعاملات سازمان‌ها و به خصوص تعاملات بانک‌ها با مشتریان، در نگاه اول بسیار سخت و گاهی ناممکن به نظر برسد، اما با بهره‌گیری از ابزارهای نوین تحلیل داده‌ها همچون داده‌کاوی و استفاده از دانش‌های روز دنیا همچون مباحث مدیریت دانش و مدیریت ارتباط با مشتری می‌توان این داده‌ها را به گنجینه‌ای گران‌بها تبدیل نمود. این گنجینه همان دانش استخراج‌شده از طریق کند و کاو این داده‌هاست.

می‌توان فن‌های داده‌کاوی را ابزارهای پشتیبان تصمیم دانست که مدیران سازمان را قادر می‌سازد که به جای تأثیرپذیری از محیط، بر بازار و عوامل آن تأثیرگذار باشند.

رتبه‌بندی مشتریان با بخش‌بندی آن‌ها سعی در کشف الگوهای رفتاری مشابه حاکم بر آن بخش دارد. کشف این الگوها می‌تواند در مسیر اتخاذ تصمیمات راهبردی سازمان بکار گرفته شود.

۵-۲- خلاصه تحقیق

هدف از این تحقیق با عنوان «مدیریت دانش مشتریان بانک مهر اقتصاد با استفاده از تکنیک‌های داده‌کاوی»، بخش‌بندی مشتریان بانک مهر اقتصاد باهدف کشف ویژگی‌های رفتاری مشابه جهت تسهیل اتخاذ استراتژی‌های متناسب با هر بخش و در نتیجه حفظ، تقویت و یا توسعه ارتباط با مشتریان این بانک می‌باشد. داده‌های خام مورد نیاز جهت این مطالعه از پایگاه داده‌های بانک مهر اقتصاد استخراج شده است.

چارچوب کلی این تحقیق در هفت گام تنظیم شده است. در گام انتخاب متغیرها، با توجه به اهداف تحقیق و به جهت کاهش حجم داده‌ها، متغیرهای تأثیرگذار بر این مطالعه شناسایی شد و از این میان سه دسته متغیر داده‌های جمعیت آماری مشتریان، متغیرهای تراکنش‌های مالی مشتریان (R)، فاصله میان زمان آخرین تراکنش مشتری تا زمان مورد ارزیابی، F ، تعداد خریدهای مشتریان در یک بازه زمانی خاص و M ، میانگین مبالغ تراکنش‌های مشتری در یک بازه زمانی خاص) و متغیر K مربوط به کلیدی بودن مشتری که با نظر خبرگان بانکی تعیین می‌گردد، انتخاب شده است. در گام آماده‌سازی داده‌ها عملیات پیش‌پردازشی نظیر حذف داده‌های نامناسب و نرمال سازی، جهت

آماده‌سازی برای ورود به الگوریتم انجام می‌گیرد. این داده‌ها در یک مجموعه با هزار نمونه و در قالب جدول اکسل تنظیم شدند. در ادامه تعداد خوشه‌ها به منظور تفکیک مشتریان با استفاده از نظرات خبرگان بانکی، پنج دسته تعیین گردید. در گام بعد الگوریتم‌های خوشه‌بندی WK-Means, K-Means و A-H-Means بر روی پایگاه داده‌ها اعمال شد. در نهایت با شناسایی خوشه‌های مشتریانی که بیشترین شباهت را در الگوهای رفتاری خود دارند، دانشی حاصل خواهد شد که می‌تواند در جهت اتخاذ استراتژی مناسب جهت جذب و نگهداری مشتری و بهبود کارآیی مالی او، توسط مدیران سازمان بکار گرفته شود.

۵-۳- نتیجه‌گیری

با بکارگیری تکنیک‌های هوشمند داده‌کاوی مشتریان بانک مهر اقتصاد در پنج دسته و با توجه به بیشترین شباهت در الگوی رفتاری گروه‌بندی شدند.

خوشه‌بندی مشتریان بر مبنای چارچوب ارائه شده در این تحقیق به ما کمک می‌کند تا مشتریانی که دارای الگوهای رفتاری مشابه بر حسب متغیرهای تراکنشی مشتریان و ممتغیرهای جمعیت آماری هستند را در گروه‌های واحدی در نظر گرفته و بر مبنای ویژگی‌های اکثریت مشتریان هر گروه، سرویس‌های متناسب با این ویژگی‌ها را به روش‌های متفاوت به مشتریان پیشنهاد نمود.

می‌توان با بررسی خوشه‌ها، خوشه‌هایی که بیش‌ترین تعداد مشتریان کلیدی در آن‌ها قرار گرفته‌اند شناسایی نمود. با این امکان مشتریان مستعد کلیدی شدن شناخته می‌شوند و با خدمت رسانی هر چه بیشتر به این گروه‌ها منجر به حفظ و ادامه ارتباط مؤثر آن‌ها با بانک گردید.

با بررسی خوشه‌هایی که بیش‌ترین مشتریان کلیدی را در خود جای داده‌اند و شناسایی اکثریت مشتریان در این خوشه‌ها از منظر نوع شغل، محل منطقه بانکی و جنسیت و ... و با نظر کارشناسان و خبرگان بانکی می‌توان خدمات بانکی ویژه و متناسب با این گروه‌ها را طراحی نموده و از طریق پیامک و یا به طور رودررو به این افراد معرفی نمود.

۴-۵- زمینه‌های پیشنهادی، راهکارها و پیشنهادات جهت پژوهش‌های آتی

امروزه تحقیقات فراوانی در زمینه داده‌کاوی و بکارگیری آن در صنعت بانکداری دنیا و کشورمان در حال انجام است. با این وجود زمینه‌های بالقوه بسیاری در بکارگیری این دانش در بانک‌ها و مؤسسات مالی و اعتباری در کشور وجود دارد. بدیهی است پیشدستی در بهره‌گیری از این ظرفیت‌ها می‌تواند مزیت رقابتی را برای بانک یا مؤسسه مالی و اعتباری مربوطه به ارمغان آورده و ارزش افزوده بیشتری را نصیب بانک یا مؤسسه مورد نظر نماید. علاوه بر این ایجاد زمینه‌هایی جهت آشنایی کارشناسان و متخصصان امور بانکی با مباحث داده‌کاوی و کاربردهای آن در صنعت بانکداری ضروری به نظر می‌رسد.

با توجه به کاربردهای وسیع شناخته‌شده از داده‌کاوی در صنعت بانکداری، به‌کارگیری این ابزار در سایر زمینه‌ها مانند کشف تقلب، ارزیابی اعتبار مشتریان و پیش‌بینی توانمندی مشتری در بازپرداخت اقساط تسهیلات، ارزیابی عملکرد بانک و ... در این بانک پیشنهاد می‌گردد.

در حال حاضر مشتریان کلیدی بانک مهر اقتصاد توسط نظر خبرگان این بانک و صرفاً با استفاده از چند معیار محدود انتخاب می‌گردند. استفاده از الگوریتم‌های داده‌کاوی مانند قواعد همبستگی و استخراج قوانینی جهت شناسایی مشتریان کلیدی بانک می‌تواند معیار خوبی جهت تکمیل نظر خبرگان در جهت انتخاب این دسته از مشتریان باشد.

برای بهبود نتایج حاصل از خوشه‌بندی، در گام تعیین تعداد بهینه خوشه‌ها استفاده از روش‌های Davis Bouldin، Two-Step و یا روش‌های موجود در زمینه تعداد بهینه خوشه‌ها در خوشه‌بندی‌های فازی، می‌تواند نتیجه نهایی را بهبود بخشد.

ارائه سرویس مناسب به مشتریان بالقوه با استفاده از الگوریتم Apriori جهت شناسایی هر چه دقیق‌تر رابطه میان گروه‌های مختلف مشتریان و سرویس‌های مورد علاقه آن‌ها و در نتیجه تخصیص امکانات متناسب جهت کاهش هزینه‌های مربوط به تبلیغات بانک از زمینه‌هایی است که برای تکمیل این تحقیق پیشنهاد می‌گردد.

در گام تعیین تعداد بهینه خوشه‌ها می‌توان از روش‌های Two-Step، Davis Bouldin و یا

روش‌های موجود در زمینه تعداد بهینه خوشه‌ها در خوشه‌بندی‌های فازی استفاده نمود.

پیوست ۱: کد الگوریتم‌ها

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Text;
using System.Windows.Forms;

namespace Read_Excel_Files
{
    public partial class Form2 : Form
    {
        DataTable dt;
        DataTable[] dtp=new DataTable[5];
        double Epsilon = 0.000000000008;
        public Form2()
        {
            InitializeComponent();
        }

        private void button1_Click(object sender, EventArgs e)
        {
            try
            {
                dt = new DataTable("MyTable");
                for (int dt_index=0; dt_index<5; dt_index++)
                    dtp[dt_index] = new DataTable("MyTable");
                OpenFileDialog openFileDialog1 = new OpenFileDialog();
                // Set filter options and filter index.
                openFileDialog1.Filter = "Text Files (.xlsx)|*.xlsx|All Files
                openFileDialog1.FilterIndex = 1;

                openFileDialog1.Multiselect = false;

                // Call the ShowDialog method to show the dialog box.
                // Process input if the user clicked OK.
                if (openFileDialog1.ShowDialog() == DialogResult.OK)
                {
                    Excel tmp_excel = new Excel();
                    tmp_excel.excel_init(openFileDialog1.FileName);
                    int[] length = tmp_excel.Excel_Length("Sheet1");
                    string Start = "A";
                    string Finish = Make_Title(length[1]);
                    int i = 1;
                    Array tmp_array = tmp_excel.GetRange(Start + i.ToString() +
                    ":" + Finish + i.ToString(), "sheet1");
                    string[] arrays = ConvertToStringArray(tmp_array);
                    string field_type = "Double";
                    int col_index = 0;
                    foreach (string tmp_item in arrays)
                    {

```

```

        if (tmp_item.Trim() == string.Empty)
        {
            break;
        }
        field_type = "Double";
        if ((col_index == 1) || (col_index == 3))
            field_type = "String";
        addColumn(ref dt, tmp_item.Trim(), field_type);
        for (int dt_index = 0; dt_index < 5; dt_index++)
            addColumn(ref dtp[dt_index], tmp_item.Trim(),
field_type);

        col_index++;
    }
    for (i = 2; i <= length[0]; i++)
    {
        tmp_array = tmp_exel.GetRange(Start + i.ToString() + ":"
+ Finish + i.ToString(), "sheet1");
        arrays = ConvertToStringArray(tmp_array);
        if (arrays[0].Trim() == string.Empty)
        {
            break;
        }
        DataRow dtr = dtp[(i-2) % 5].NewRow();
        DataRow dtr2 = dt.NewRow();
        col_index = 0;
        foreach (string tmp_item in arrays)
        {
            if (tmp_item.Trim() == string.Empty)
            {
                break;
            }
            if ((col_index == 1) || (col_index == 3))
            {
                dtr[col_index] = tmp_item;
                dtr2[col_index++] = tmp_item;
            }
            else
            {
                dtr[col_index] = Convert.ToDouble(tmp_item);
                dtr2[col_index++] = Convert.ToDouble(tmp_item);
            }
        }
        dtp[(i - 2) % 5].Rows.Add(dtr);
        dt.Rows.Add(dtr2);
    }
    tmp_exel.excel_close();
    Converts();
}
}
catch (Exception ex)
{
    MessageBox.Show(ex.Message);
}
}

```

```

string[] ConvertToStringArray(System.Array values)
{
    // create a new string array
    string[] theArray = new string[values.Length];

    // loop through the 2-D System.Array and populate the 1-D String
    for (int i = 1; i <= values.Length; i++)
    {
        if (values.GetValue(1, i) == null)
            theArray[i - 1] = "";
        else
            theArray[i - 1] = (string)values.GetValue(1, i).ToString();
    }

    return theArray;
}

private void Form2_Load(object sender, EventArgs e)
{
}

public static string Reverse(string s)
{
    char[] charArray = s.ToCharArray();
    Array.Reverse(charArray);
    return new string(charArray);
}

public string Make_Title(int in_Length)
{
    dataGridView1.Rows.Clear();
    dataGridView1.Columns.Clear();
    string temp = "A";
    for (int i = 0; i < in_Length; i++)
    {
        dataGridView1.Columns.Add(temp, temp);
        temp = Reverse(temp);
        string t_temp = "";
        int Carry = 1;
        foreach (char ch in temp)
        {
            char t_ch = (char)(ch + Carry);
            Carry = 0;
            if (t_ch > 'Z')
            {
                t_ch = 'A';
                Carry = 1;
            }
            t_temp = t_ch.ToString() + t_temp;
        }
        if (Carry == 1)
        {
            t_temp = "A" + t_temp;
        }
    }
}

```

```

        }
        temp = t_temp;
    }
    return temp;
}

private void dataGridView1_CellContentClick(object sender,
DataGridViewCellEventArgs e)
{
}

public void addColumn(ref DataTable table, string columnName, string
in_type)
{
    DataColumn auxColumn = new DataColumn(columnName,
Type.GetType("System." + in_type));
    //auxColumn.DefaultValue = defaultValue;
    table.Columns.Add(auxColumn);
}

private void Converts()
{
    bool change = true;
    double wight_R = ConvertToDouble(textBox_R.Text);
    double wight_F = ConvertToDouble(textBox_F.Text);
    double wight_M = ConvertToDouble(textBox_M.Text);
    double wight_B = ConvertToDouble(textBox_B.Text);
    while (change)
    {
        change = false;
        double[] Sum_R = new double[5];
        double[] Sum_F = new double[5];
        double[] Sum_M = new double[5];
        double[] Sum_B = new double[5];
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            DataRow[] resultp1 = dtp[dt_index].Select();
            double sum_R = 0;
            double sum_F = 0;
            double sum_M = 0;
            double sum_B = 0;
            int Row_Count = 0;
            foreach (DataRow row in resultp1)
            {
                sum_R +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]);
                sum_F +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]);
                sum_M +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]);
                sum_B +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]);
                Row_Count++;
            }
            sum_R = Row_Count == 0 ? 0 : sum_R / Row_Count;

```

```

        sum_F = Row_Count == 0 ? 0 : sum_F / Row_Count;
        sum_M = Row_Count == 0 ? 0 : sum_M / Row_Count;
        sum_B = Row_Count == 0 ? 0 : sum_B / Row_Count;
        Sum_R[dt_index] = sum_R;
        Sum_F[dt_index] = sum_F;
        Sum_M[dt_index] = sum_M;
        Sum_B[dt_index] = sum_B;
    }

//=====
    DataRow[] resultp = dt.Select();
    double tmp_R = 0;
    double tmp_F = 0;
    double tmp_M = 0;
    double tmp_B = 0;
    double[] distance_mean = new double[5];

    foreach (DataRow row in resultp)
    {
        tmp_R =
Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]));
        tmp_F =
Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]));
        tmp_M =
Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]));
        tmp_B =
Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]));
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            distance_mean[dt_index] = Math.Sqrt((wight_R *
Math.Pow((tmp_R - Sum_R[dt_index]), 2)) + (wight_F * Math.Pow((tmp_F -
Sum_F[dt_index]), 2)) + (wight_M * Math.Pow((tmp_M - Sum_M[dt_index]), 2)) +
(wight_B * Math.Pow((tmp_B - Sum_B[dt_index]), 2))));
        }

        int tmp_min_index = IndexOfMin(distance_mean);

        DataRow[] dr = dtp[tmp_min_index].Select("ID = " +
row["ID"].ToString());
        if (dr.Length == 0)
        {
            change = true;
            for (int dt_index = 0; dt_index < 5; dt_index++)
            {
                DataRow[] dr2 = dtp[dt_index].Select("ID = " +
row["ID"].ToString());
                foreach (DataRow r in dr2)
                    r.Delete();
            }
            dtp[tmp_min_index].ImportRow(row);
        }
    }
    ExportToExcel("output"+" SSE="+SSEValue().ToString(), dtp, "Group");
    MessageBox.Show("date convert ok");
}

```

```

private void ConvertsH(int in_Q)
{
    bool change = true;
    double wight_R = ConvertToDouble(textBox_R.Text);
    double wight_F = ConvertToDouble(textBox_F.Text);
    double wight_M = ConvertToDouble(textBox_M.Text);
    double wight_B = ConvertToDouble(textBox_B.Text);
    while (change)
    {
        change = false;
        double[] Sum_R = new double[5];
        double[] Sum_F = new double[5];
        double[] Sum_M = new double[5];
        double[] Sum_B = new double[5];
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            DataRow[] resultp1 = dtp[dt_index].Select();
            double sum_R = 0;
            double sum_F = 0;
            double sum_M = 0;
            double sum_B = 0;
            double Row_Count_R = 0;
            double Row_Count_F = 0;
            double Row_Count_M = 0;
            double Row_Count_B = 0;
            foreach (DataRow row in resultp1)
            {
                sum_R +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]) + Epsilon), in_Q
+ 1);
                sum_F +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]) + Epsilon), in_Q
+ 1);
                sum_M +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]) + Epsilon), in_Q
+ 1);
                sum_B +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]) + Epsilon), in_Q
+ 1);
                Row_Count_R +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]) + Epsilon),
in_Q);
                Row_Count_F +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]) + Epsilon),
in_Q);
                Row_Count_M +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]) + Epsilon),
in_Q);
                Row_Count_B +=
Math.Pow((Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]) + Epsilon),
in_Q);
            }
            sum_R = Row_Count_R == 0 ? 0 : sum_R / Row_Count_R;
            sum_F = Row_Count_F == 0 ? 0 : sum_F / Row_Count_F;
            sum_M = Row_Count_M == 0 ? 0 : sum_M / Row_Count_M;

```

```

        sum_B = Row_Count_B == 0 ? 0 : sum_B / Row_Count_B;
        Sum_R[dt_index] = sum_R;
        Sum_F[dt_index] = sum_F;
        Sum_M[dt_index] = sum_M;
        Sum_B[dt_index] = sum_B;
    }

//=====
    DataRow[] resultp = dt.Select();
    double tmp_R = 0;
    double tmp_F = 0;
    double tmp_M = 0;
    double tmp_B = 0;
    double[] distance_mean = new double[5];

    foreach (DataRow row in resultp)
    {
        tmp_R =
Convert.ToDouble(row[ row.Table.Columns.IndexOf("R")]);
        tmp_F =
Convert.ToDouble(row[ row.Table.Columns.IndexOf("F")]);
        tmp_M =
Convert.ToDouble(row[ row.Table.Columns.IndexOf("M")]);
        tmp_B =
Convert.ToDouble(row[ row.Table.Columns.IndexOf("B")]);
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            distance_mean[dt_index] = Math.Sqrt((wight_R *
Math.Pow((tmp_R - Sum_R[dt_index]), 2)) + (wight_F * Math.Pow((tmp_F -
Sum_F[dt_index]), 2)) + (wight_M * Math.Pow((tmp_M - Sum_M[dt_index]), 2)) +
(wight_B * Math.Pow((tmp_B - Sum_B[dt_index]), 2))));
        }

        int tmp_min_index = IndexOfMin(distance_mean);

        DataRow[] dr = dtp[tmp_min_index].Select("ID = " +
row["ID"].ToString());
        if (dr.Length == 0)
        {
            change = true;
            for (int dt_index = 0; dt_index < 5; dt_index++)
            {
                DataRow[] dr2 = dtp[dt_index].Select("ID = " +
row["ID"].ToString());

                foreach (DataRow r in dr2)
                    r.Delete();
            }
            dtp[tmp_min_index].ImportRow(row);
        }
    }

    ExportToExcel("output"+" Q="+in_Q.ToString()+"
SSE="+SSEValue(in_Q).ToString(), dtp, "Group");
    //MessageBox.Show("date convert ok");
}

```

```

public static int IndexOfMin(double[] source)
{
    if (source == null)
        throw new ArgumentNullException("source");

    double minValue = double.MaxValue;
    int minIndex = -1;
    int index = -1;

    foreach (double num in source)
    {
        index++;

        if (num <= minValue)
        {
            minValue = num;
            minIndex = index;
        }
    }

    if (index == -1)
        throw new InvalidOperationException("Sequence was empty");

    return minIndex;
}

private void ExportToExcel(string fileName, DataTable[]
dtExcelTable, string in_sheetname) //Exports the given dataGridView to Excel with
the given fileName
{
    Microsoft.Office.Interop.Excel.Application xlApp;
    Microsoft.Office.Interop.Excel.Workbook xlWorkBook;
    Microsoft.Office.Interop.Excel.Worksheet xlWorkSheet;
    object misValue = System.Reflection.Missing.Value;

    xlApp = new Microsoft.Office.Interop.Excel.Application();
    xlWorkBook = xlApp.Workbooks.Add(misValue);

    for (int dt_index = 0; dt_index < dtExcelTable.Length; dt_index++)
    {
        xlWorkSheet =
(Microsoft.Office.Interop.Excel.Worksheet)xlWorkBook.Worksheets.Add(Type.Missing,
Type.Missing, Type.Missing, Type.Missing);
        int i = 0;
        int j = 0;
        xlWorkSheet.Name = in_sheetname + dt_index.ToString();

        for (i = 0; i < dtExcelTable[dt_index].Columns.Count; i++)
        {
            xlWorkSheet.Cells[1, i + 1] =
dtExcelTable[dt_index].Columns[i].ColumnName;
        }

        for (i = 0; i < dtExcelTable[dt_index].Rows.Count; i++)
        {

```

```

        for (j = 0; j < dtExcelTable[dt_index].Columns.Count; j++)
        {
            xlWorkSheet.Cells[i + 2, j + 1] =
dtExcelTable[dt_index].Rows[i][j];
        }
    }

    try
    {
        SaveFileDialog saveFileDialog1 = new SaveFileDialog();
        saveFileDialog1.Filter = "Excel File|*.xls";
        saveFileDialog1.Title = "Save File";
        saveFileDialog1.FileName = fileName;
        saveFileDialog1.ShowDialog();

        // If the file name is not an empty string open it for saving.
        if (saveFileDialog1.FileName != "")
        {
            xlWorkBook.SaveAs(saveFileDialog1.FileName + ".xls",
Microsoft.Office.Interop.Excel.XlFileFormat.xlWorkbookNormal, misValue, misValue,
misValue, misValue,
Microsoft.Office.Interop.Excel.XlSaveAsAccessMode.xlExclusive, misValue,
misValue, misValue, misValue, misValue);
            xlWorkBook.Close(true, misValue, misValue);
            xlApp.Quit();
        }

    }
    catch (Exception ex)
    {
        MessageBox.Show(ex.Message);
    }
}

private double SSEValue(int in_Q)
{
    try
    {
        double[] Sum_R = new double[5];
        double[] Sum_F = new double[5];
        double[] Sum_M = new double[5];
        double[] Sum_B = new double[5];
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            DataRow[] resultp1 = dtp[dt_index].Select();
            double sum_R = 0;
            double sum_F = 0;
            double sum_M = 0;
            double sum_B = 0;
            double Row_Count_R = 0;
            double Row_Count_F = 0;
            double Row_Count_M = 0;
            double Row_Count_B = 0;

```

```

        foreach (DataRow row in resultp1)
        {
            sum_R +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]) + Epsilon, in_Q +
1);
            sum_F +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]) + Epsilon, in_Q +
1);
            sum_M +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]) + Epsilon, in_Q +
1);
            sum_B +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]) + Epsilon, in_Q +
1);
            Row_Count_R +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]) + Epsilon, in_Q);
            Row_Count_F +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]) + Epsilon, in_Q);
            Row_Count_M +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]) + Epsilon, in_Q);
            Row_Count_B +=
Math.Pow(Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]) + Epsilon, in_Q);
        }
        sum_R = Row_Count_R == 0 ? 0 : sum_R / Row_Count_R;
        sum_F = Row_Count_F == 0 ? 0 : sum_F / Row_Count_F;
        sum_M = Row_Count_M == 0 ? 0 : sum_M / Row_Count_M;
        sum_B = Row_Count_B == 0 ? 0 : sum_B / Row_Count_B;
        Sum_R[dt_index] = sum_R;
        Sum_F[dt_index] = sum_F;
        Sum_M[dt_index] = sum_M;
        Sum_B[dt_index] = sum_B;
    }

//=====
    double SSE = 0;
    for (int dt_index = 0; dt_index < 5; dt_index++)
    {
        DataRow[] dr = dtp[dt_index].Select();
        double tmp_R = 0;
        double tmp_F = 0;
        double tmp_M = 0;
        double tmp_B = 0;

        foreach (DataRow row in dr)
        {
            tmp_R =
Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]);
            tmp_F =
Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]);
            tmp_M =
Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]);
            tmp_B =
Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]);
            SSE += Math.Pow((Sum_R[dt_index] - tmp_R), 2);
            SSE += Math.Pow((Sum_F[dt_index] - tmp_F), 2);
            SSE += Math.Pow((Sum_M[dt_index] - tmp_M), 2);
        }
    }

```

```

        SSE += Math.Pow((Sum_B[dt_index] - tmp_B), 2);
    }
    }
    return SSE;
}
catch (Exception ex)
{
    throw ex;
}
}

private double SSEValue()
{
    try
    {
        double[] Sum_R = new double[5];
        double[] Sum_F = new double[5];
        double[] Sum_M = new double[5];
        double[] Sum_B = new double[5];
        for (int dt_index = 0; dt_index < 5; dt_index++)
        {
            DataRow[] resultp1 = dtp[dt_index].Select();
            double sum_R = 0;
            double sum_F = 0;
            double sum_M = 0;
            double sum_B = 0;
            int Row_Count = 0;
            foreach (DataRow row in resultp1)
            {
                sum_R +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]);
                sum_F +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]);
                sum_M +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]);
                sum_B +=
Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]);
                Row_Count++;
            }
            sum_R = Row_Count == 0 ? 0 : sum_R / Row_Count;
            sum_F = Row_Count == 0 ? 0 : sum_F / Row_Count;
            sum_M = Row_Count == 0 ? 0 : sum_M / Row_Count;
            sum_B = Row_Count == 0 ? 0 : sum_B / Row_Count;
            Sum_R[dt_index] = sum_R;
            Sum_F[dt_index] = sum_F;
            Sum_M[dt_index] = sum_M;
            Sum_B[dt_index] = sum_B;
        }
    }

//=====
    double SSE = 0;
    for (int dt_index = 0; dt_index < 5; dt_index++)
    {
        DataRow[] dr = dtp[dt_index].Select();
        double tmp_R = 0;
        double tmp_F = 0;

```

```

        double tmp_M = 0;
        double tmp_B = 0;

        foreach (DataRow row in dr)
        {
            tmp_R =
Convert.ToDouble(row[row.Table.Columns.IndexOf("R")]);
            tmp_F =
Convert.ToDouble(row[row.Table.Columns.IndexOf("F")]);
            tmp_M =
Convert.ToDouble(row[row.Table.Columns.IndexOf("M")]);
            tmp_B =
Convert.ToDouble(row[row.Table.Columns.IndexOf("B")]);
            SSE += Math.Pow((Sum_R[dt_index] - tmp_R), 2);
            SSE += Math.Pow((Sum_F[dt_index] - tmp_F), 2);
            SSE += Math.Pow((Sum_M[dt_index] - tmp_M), 2);
            SSE += Math.Pow((Sum_B[dt_index] - tmp_B), 2);
        }
        return SSE;
    }
    catch (Exception ex)
    {
        throw ex;
    }
}

private double ConvertToDouble(string in_Number)
{
    try
    {
        return Convert.ToDouble(in_Number);
    }
    catch
    {
        return 1;
    }
}

private void button2_Click(object sender, EventArgs e)
{
    try
    {
        dt = new DataTable("MyTable");
        for (int dt_index=0; dt_index<5; dt_index++)
            dtp[dt_index] = new DataTable("MyTable");
        OpenFileDialog openFileDialog1 = new OpenFileDialog();
        // Set filter options and filter index.
        openFileDialog1.Filter = "Text Files (*.xlsx)|*.xlsx|All Files
(*.*)|*.*";

        openFileDialog1.FilterIndex = 1;

        openFileDialog1.Multiselect = false;

        // Call the ShowDialog method to show the dialog box.
        // Process input if the user clicked OK.
    }
    catch
    {
    }
}

```

```

if (openFileDialog1.ShowDialog() == DialogResult.OK)
{
    Excel tmp_excel = new Excel();
    tmp_excel.excel_init(openFileDialog1.FileName);
    int[] length = tmp_excel.Excel_Length("Sheet1");
    string Start = "A";
    string Finish = Make_Title(length[1]);
    int i = 1;
    Array tmp_array = tmp_excel.GetRange(Start + i.ToString() +
":" + Finish + i.ToString(), "sheet1");
    string[] arrays = ConvertToStringArray(tmp_array);
    string field_type = "Double";
    int col_index = 0;
    foreach (string tmp_item in arrays)
    {
        if (tmp_item.Trim() == string.Empty)
        {
            break;
        }
        field_type = "Double";
        if ((col_index == 1) || (col_index == 3))
            field_type = "String";
        addColumn(ref dt, tmp_item.Trim(), field_type);
        for (int dt_index = 0; dt_index < 5; dt_index++)
            addColumn(ref dtp[dt_index], tmp_item.Trim(),
field_type);

        col_index++;
    }
    for (int Q=-5;Q<=5;Q++)
    {
        dt.Rows.Clear();
        for (int dt_index = 0; dt_index < 5; dt_index++)
            dtp[dt_index].Rows.Clear();
        for (i = 2; i <= length[0]; i++)
        {
            tmp_array = tmp_excel.GetRange(Start + i.ToString() +
":" + Finish + i.ToString(), "sheet1");
            arrays = ConvertToStringArray(tmp_array);
            if (arrays[0].Trim() == string.Empty)
            {
                break;
            }
            DataRow dtr = dtp[(i-2) % 5].NewRow();
            DataRow dtr2 = dt.NewRow();
            col_index = 0;
            foreach (string tmp_item in arrays)
            {
                if (tmp_item.Trim() == string.Empty)
                {
                    break;
                }
                if ((col_index == 1) || (col_index == 3))
                {
                    dtr[col_index] = tmp_item;
                    dtr2[col_index++] = tmp_item;

```

```

        }
        else
        {
            dtr[col_index] = Convert.ToDouble(tmp_item);
            dtr2[col_index++] =
Convert.ToDouble(tmp_item);
        }
    }
    dtp[(i - 2) % 5].Rows.Add(dtr);
    dt.Rows.Add(dtr2);
}
ConvertsH(Q);
}
tmp_exel.excel_close();
MessageBox.Show("date convert ok");
}
}
catch (Exception ex)
{
    MessageBox.Show(ex.Message);
}
}
}
}
}

```

منابع و مأخذ

- [۱] شهرابی، جمال، هداوندی، اسماعیل، "داده‌کاوی در صنعت بانکداری"، انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر، ۱۳۹۰، صص ۲۵-۷۰.
- [۲] مینائی، بهروز، نصیری، مهدی، حسنی، دانیال، شناسا، ابراهیم، آموزش گام به گام داده‌کاوی با Clementine، انتشارات گروه مهندسی-پژوهشی ساحر، چاپ اول، پاییز ۱۳۹۰.
- [۳] الهی، شعبان، حیدری، بهمن، مدیریت ارتباط با مشتری، تهران، شرکت چاپ و نشر بازرگانی، ۱۳۸۷.
- [۴] افراز، ع.، مدیریت دانش (مفاهیم، مدل‌ها، اندازه‌گیری و پیاده‌سازی)، چاپ اول، انتشارات دانشگاه صنعتی امیرکبیر، ۱۳۸۴.
- [۵] تقوی، مهدی، اقتصاد کلان، انتشارات دانشگاه پیام نور، صص ۱۷۶-۱۸۵.
- [۶] علی احمدی، علیرضا. ابراهیمی، مهدی. سلیمانی ملکان، حجت، برنامه ریزی استراتژیک فناوری اطلاعات و ارتباطات، انتشارات تولید دانش، تهران، ۱۳۸۳.
- [۷] شامی زنجانی، مهدی، نجف‌لو، فاطمه، ارائه چارچوبی مفهومی برای نوع‌شناسی دانش مشتری: یک مطالعه موردی، مدیریت فناوری اطلاعات، دوره ۳، شماره ۹، صص ۱۶۳-۱۸۹، زمستان ۱۳۹۰.
- [۸] مؤمنی، حمیدرضا و همکاران، "استفاده از تکنیک داده‌کاوی در مدیریت ارتباط با مشتریان در نظام بانکی"، کنفرانس ملی فناوری اطلاعات و جهاد اقتصادی، دانشگاه سلمان فارسی، کازرون، اسفند ۱۳۹۰.
- [۹] برزین پور، فرناز، اخوان، پیمان، زارع مقدم، زینب، ارائه چارچوب مفهومی عوامل و مؤلفه‌های مدیریت دانش مشتری در بانک ملی، نشریه علمی پژوهشی مدیریت فردا، سال هشتم، شماره ۲۴، پاییز ۱۳۸۹.
- [۱۰] شهرکی، علیرضا، چهکندی، محسن، ملاشاهی، علیرضا، بررسی سیستم مدیریت ارتباط با مشتری در نظام بانکی ایران، ۲۰۱۰.

[۱۱] حسینی بامکان، سید مجتبی، پایان نامه کارشناسی ارشد، بکارگیری تکنیک‌های داده کاوی جهت بهبود مدیریت ارتباط با مشتری در صنعت بانکداری (مطالعه موردی: بانک پارسیان)، استاد راهنما، دکتر محمدرضا تقوا، دانشکده حسابداری و مدیریت، دانشگاه علامه طباطبائی، ۱۳۸۸.

[۱۲] ترابی، مصطفی، عوامل مؤثر بر بکارگیری خدمات موبایل بانک توسط مشتریان مورد مطالعه بانک تجارت، پایان نامه کارشناسی ارشد، به راهنمای، دکتر تقوی فرد، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، ۱۳۷۸.

[۱۳] ناوی، مانیا، شناسایی مولفه های تشکیل دهنده بار با استفاده از تکنیک داده کاوی، پایان نامه کارشناسی ارشد، استاد راهنما، دکتر محسن پارسا مقدم، دانشکده فنی و مهندسی، دانشگاه تربیت مدرس، ۱۳۸۷.

[۱۴] حسین زاده، لیلا، دسته بندی مشتریان هدف در صنعت بیمه با استفاده از داده کاوی، پایان نامه کارشناسی ارشد، استاد راهنما، دکتر شعبان الهی، دانشکده انسانی، دانشگاه تربیت مدرس، ۱۳۸۶.

[۱۵] جمالی فیروز آبادی، کامران، دارنده، احسان، ابراهیمی، سید بابک، ارائه چارچوب مدیریت دانش مشتری (CKM) بر مبنای استقرار همزمان KM و CRM در سازمان، مجموعه مقالات مهندسی صنایع و نرم افزار، ۱۳۸۵.

[۱۶] نوروزیان، م، کاربرد مدیریت دانش در بخش دولتی، مجله تدبیر، شماره ۱۷۶، ۱۳۸۵.

[۱۷] همایونی، سید مهدی، محتشمی، الهام، اسفندیاری، علیرضا، مدیریت دانش مشتری، کنفرانس بین‌المللی مدیریت بازاریابی، شهریور ۱۳۸۵، صص ۳-۱۲.

[۱۸] Bosjank, z. , "Credit users segmentation for improved customer relationship management in banking", Applied Computational Intelligence and Informatics (SACI), IEEE , ۲۰۱۱, pp. ۳۷۹ - ۳۸۴.

[۱۹] Nie, Guangli, et al. "Credit card churn forecasting by logistic regression and decision tree." Expert Systems with Applications ۳۸,۱۲ (۲۰۱۱) : ۱۵۲۷۳-۱۵۲۸۵.

[۲۰] Lin, Chiun-Sin, Gwo-Hshiung Tzeng, and Yang-Chieh Chin. "Combined rough set theory and flow network graph to predict customer churn in credit card accounts." *Expert Systems with Applications* ۳۸,۱ (۲۰۱۱) : ۸-۱۵.

[۲۱] Madhoushi, Mehrdad and et al., "Survey of Customer Knowledge Management Impact on Customer Relationship Management: (Iranian study)", *International Journal of Business and Social Science*, Vol. ۲, No. ۲۰, November ۲۰۱۱.

[۲۲] Lee, Byungtae and et al., "Empirical analysis of online auction fraud: Credit card phantom transactions", *Expert Systems with Applications*, vol ۳۷, pp. ۲۹۹۱-۲۹۹۹, ۲۰۱۰.

[۲۳] Hosseini, Seyed Mohammad Seyed, Anahita Maleki, and Mohammad Reza Gholamian. "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty." *Expert Systems with Applications* ۳۷,۷ (۲۰۱۰): pp. ۵۲۵۹-۵۲۶۷.

[۲۴] Chen, Fei-Long and et al., "Combination of feature selection approaches with SVM in credit scoring", *Expert Systems with Applications*, vol ۳۷, pp. ۷۹۰۲-۷۹۰۹, ۲۰۱۰.

[۲۵] Ngai, Eric WT, Li Xiu, and D. C. K. Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert Systems with Applications* ۳۶,۲ (۲۰۰۹): ۲۵۹۲-۲۶۰۲.

[۲۶] Nisbet, Robert, John Fletcher Elder, and Gary Miner. *Handbook of statistical analysis and data mining applications*. Academic Press, ۲۰۰۹.

[۲۷] Rezvani, Zeinab, New Product Development Based on Customer Knowledge Management, Master Thesis, Luleå University of Technology, pp.۲۵-۳۵, ۲۰۰۹.

[۲۸] Xie, Yaya, et al. "Customer churn prediction using improved balanced random forests." *Expert Systems with Applications* ۳۶,۳ (۲۰۰۹) : ۵۷۷۵-۵۷۷۹.

[۲۹] Abdou, Hussein et al., "Neural nets versus conventional techniques in credit scoring in Egyptian banking", *Expert Systems with Applications*, vol ۳۵, pp. ۱۲۷۵-۱۲۹۲, ۲۰۰۸.

[۳۰] Ranjan, Jayanthi, A Review of Data Mining Tools In Customer Relationship Management ,*Journal of Knowledge Management Practice*, Vol. ۹, No. ۱, March ۲۰۰۸.

[31] Han, Jiawei, and Micheline Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[32] Paquette, Scott. "Customer knowledge management." *Encyclopedia of Knowledge Management* (2006): 90.

[33] Rollins, Minna, and Aino Halinen. "Customer knowledge management competence: towards a theoretical framework." *System Sciences*, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on. IEEE, 2005.

[34] Feng, Tian-Xue, and Jin-Xin Tian. "Customer knowledge management and condition analysis of successful CKM implementation." *Machine Learning and Cybernetics*, 2005. *Proceedings of 2005 International Conference on*. Vol. 1. IEEE, 2005.

[35] Hsieh, Nan-Chen. "An integrated data mining and behavioral scoring model for analyzing bank customers." *Expert Systems with Applications* 27,1 (2004): 623-633.

[36] Bose, Ranjit, and Vijayan Sugumaran. "Application of knowledge management technology in customer relationship management." *Knowledge and process management* 10,1 (2003): 3-17.

[37] Gebert, Henning, et al. "Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts [1]." *Journal of knowledge management* 7,5 (2003): 107-123.

[38] Gibbert, Michael, Marius Leibold, and Gilbert Probst. "Five styles of customer knowledge management, and how smart companies use them to create value." *European Management Journal* 20, no 5, pp 559-569, 2002.

[39] Bose, Ranjit. "Customer relationship management: key components for IT success." *industrial management & data systems* 102,2 (2002): 89-97.

[40] Dyche, Jill. *The CRM handbook: a business guide to customer relationship management*. Addison-Wesley Professional, 2002.

[୧୧] Garcia-Murillo, Martha, and Hala Annabi. "Customer knowledge management." *Journal of the Operational Research Society*, pp. ୮୭୦-୮୮୧, ୨୦୦୨.

[୧୨] Rowley J. Eight Enhancing Questions for Customer Knowledge Management in e-Business. *Journal of Knowledge Management* ୨୦୦୨; ୬(୦): ୦୦୦-୦୧୧.

[୧୩] Rowley, Jennifer E. "Reflections on customer knowledge management in e-business." *Qualitative Market Research: An International Journal* ୫,୧ (୨୦୦୨): ୨୮୮-୨୮୦.

[୧୪] Zack, Michael H. "Developing a knowledge strategy." *The strategic management of intellectual capital and organizational knowledge* (୨୦୦୨): ୩୮-୨୦୦.

[୧୫] Bhatt, Ganesh D. "Knowledge management in organizations: examining the interaction between technologies, techniques, and people." *Journal of knowledge management* ୫,୧, pp. ୮୮-୩୦, ୨୦୦୧.

[୧୬] Ling, Raymond, and David C. Yen. "Customer relationship management: An analysis framework and implementation strategies." *Journal of Computer Information Systems* ୧୧,୩ (୨୦୦୧): ୮୨-୧୩.

[୧୭] Davenport, Thomas H., Jeanne G. Harris, and Ajay K. Kohli. "How do they know their customers so well?." *MIT Sloan Management Review* ୧୨,୨ (୨୦୦୧): ୮୩-୩୩.

[୧୮] Davenport, Thomas H., and Laurence Prusak. *Working knowledge: How organizations manage what they know*. Harvard Business Press, ୨୦୦୦.

[୧୯] Mårtensson, Maria. "A critical review of knowledge management as a management tool." *Journal of knowledge management* ୧, no ୩ (୨୦୦୦): ୨୦୧-୨୧୮.

[୨୦] Wirth, Rüdiger, and Jochen Hipp. "CRISP-DM: Towards a standard process model for data mining." *Proceedings of the ୧୮th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. ୨୦୦୦.

[୨୧] uit Beijerse, Roelof P. "Questions in knowledge management: defining and conceptualizing a phenomenon." *Journal of Knowledge Management* ୩,୨ (୧୯୯୯): ୧୧-୧୧୦.

[٥٢] Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." *Knowledge and data Engineering, IEEE Transactions on* ٨,٦ (١٩٩٦): ٨٦٦-٨٨٣.

[٥٣] Fayyad, Usama and et al., "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, Vol. ١٧, Number ٣, ١٩٩٦.

[٥٤] Szulanski, Gabriel. "Exploring internal stickiness: Impediments to the transfer of best practice within the firm." *Strategic management journal* ١٧ (١٩٩٦): ٢٧-٤٣.

[٥٥] Nonaka, Ikujiro, and Hirotaka Takeuchi. *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press, USA, ١٩٩٥.

Summary

Increasing level of competition between the markets, made the managers and organization analysts to seek solutions that bring a competitive advantage for organizations. According to literature, using customer knowledge to adopt strategies, for customer satisfaction, direct organizations toward this goal. On the other, expand taking advantage of update technologies in information and communication, in particular, in banks cause to staying large amounts of data that analyze and make decisions based on their with conventional methods of reporting and statistical methods, is not possible. Data mining is a powerful and update tool that is proposed for data analysis to extract customer knowledge in this thesis.

The purpose of this study that titled as “Knowledge management of the Mehr Eghtesad bank customers using data mining techniques”, is segmentation of the customers of the Mehr Eghtesad bank, aimed to discover the similar behavioral characteristics, for help the managers in banks to facilitate the adoption of appropriate strategies for each cluster and consequently preserve, strengthen and develop relationships with customers and ultimately profitability for the bank. Raw data for this study were extracted from the database of Mehr Eghtesad bank.

Key words:

Customer Knowledge Management, Data Mining, Knowledge Discovery, Customers Classification.