به نام خداوند بخشنده ومهربان

© TemplatesWise.com

داده کاوی و کاربرد آن دربانکداری به عنوان شاخصی ازاقتصاد

- استاد راهنما:سرکار خانم دکتر رابوکی
- گردآوری وارائه:آرزوهمایی فصیح
- نیمسال اول سال تحصیلی 93-92

داده کاوی علمی برای رشد اقتصادی

برخی از فروشندگان می دانند که خریدار بعد از خرید خود به دنبال چه کالایی می رود. برای مثال پس از فروش تلویزیون، مشتری به دنبال خرید میز تلویزیون جدید و تزئینات جانبی می رود.این مثالی ساده است. در مواقع پیچیده دیگرنمی توان با استفاده از حدس و تجربه فهمید که که خریدار به دنبال کدام کالا می رود، در مواقع مشابه چه رفتاری از سوی یک فرد رخ می دهد ؟مجریان بیشتر در چه زمینه ای فعالیت می کنند ؟ و بسیاری از این موارد. اما در پس این موارد داده کاوی وجود دارد که می تواند از بسیاری از موارد آگاهی یافته و براساس آن تصمیمات درست و بموقع گرفت و سود سرشاری را نصیب خود کرد.

این نوع استفاده از داده کاوی می‌تواند فروشگاه‌ها را دربرگزاری هوشمندانه فستیوال‌های فروش و نحوه ارایه اجناس به مشتریان کمک کند. در یک شرکت بزرگ تکثیر وعرضه فیلم های سینمایی در امریکا، براساس دانش داده کاوی روابط مشتریان و هنرپیشگان سینمایی و گروه های مختلف مشتریان با سبک های فیلم ها مشخص شد.بعد از تعیین این رابطه که برروی حجم زیادی از داده های از جمله فیلم های خریداری شده، گروه سنی خریداران، نتایج نظرسنجی و دیگر مسایل انجام شد این شرکت توانست به طور کاملا هوشمندانه مشتریان بالقوه فیلم های سینمایی خود را براساس علاقه مشتریان به هنرپیشه های مختلف و سبک های سینمایی شناسایی و فروش خود را چند برابر کند.

داده کاوی فرایند کشف اطلاعات مفید از منابع داده حجیم و بزرگ است. داده کاوی را جستجوی الگو در پایگاه داده های بزرگ معرفی کرده اند.داده کاوی را شاید بتوان کاوش دانش از درون داده ها نیز نامید.تکنیکهای داده کاوی عموما برای استخراج دانش و الگو از مجموعه داده های بزرگ و به منظور یادگیری و پیش بینی شرایط جدید به کار می روند.این روش ها در زمینه هایی که داده های زیادی تولید می کنند وپردازش این داده ها به صورت دستی ممکن نیست مورد استفاده قرار می گیرند. داده کاوی جمع آوری اطلاعات و تحقیق روی آن ها نیست. در واقع علم داده کاوی رشته ای میان رشته است که در علوم کامپیوتر، آنالیز آماری و دیگر علوم مشترک است اما خود رشته ای مستقل است. داده کاوی در اقتصاد توسط تجهیزات خاصی انجام می شود که عملیات کاوش براساس تجزیه و تحلیل داده ها به صورت مکرر انجام می شود.

تفاوت داده کاوی با روش های آنالیز آماری

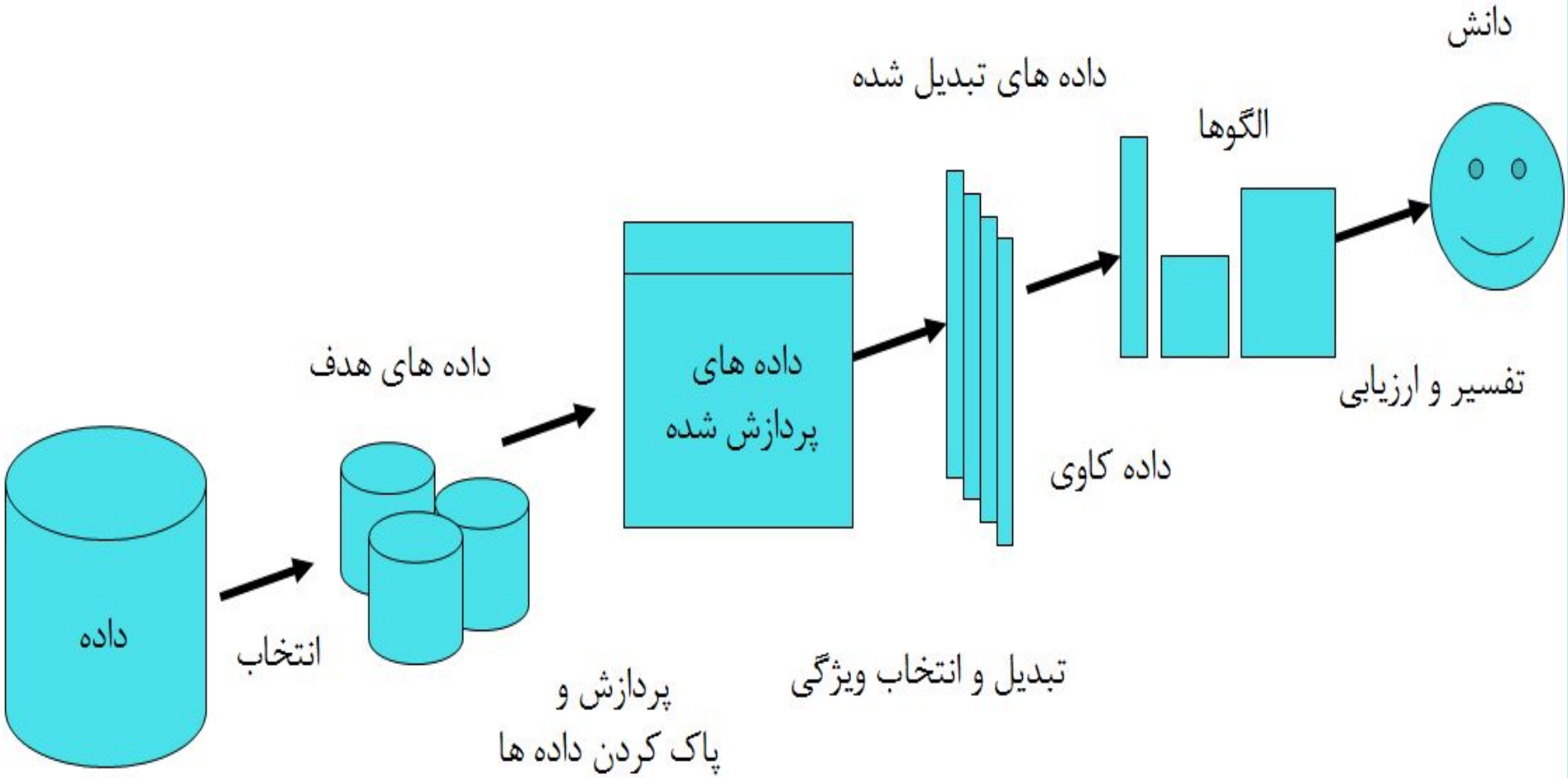
درباره تفاوت داده کاوی با روش های آنالیز آماری می توان گفت آمار شناسان همیشه با یک فرضیه کار خود را آغاز می کنند درحالی که یکمتخصص داده کاوی به فرضیه نیازی ندارد.کارشناسان آنالیز آماری می توانند داده های نابجا و نادرست را در طول آنالیز مشخص کنند و نتایج کار خود را تفسیر کنند ولی این درحالی است که یک کارشناس داده کاوی به داده های صحیح و درست نیازمند است و به دلیل پیچیدگی بیش از حد نیاز به متخصصانی برای تفسیر دارد.

مراحل فرآیند داده کاوی

هدف داد ه کاوی، تجزیه و تحلیل اکتشافی داده ها، کشف الگوها و قواعد و الگوریتم ها، مدل سازی پیش بینانه وجست وجوی انحرافات است. برای انجام این هدف، فرآینددادہ کاوی درمراحل مختلف انجام می شود که عبارت است از

1. **شناسایی هدف و فهم حوزه کاربرد که مشخص می کند که چه کاری، در چه حوزه ای انجام خواهد شد.**
2. **انتخاب داده ها یعنی تعیین اهداف برای تجزیه وتحلیل و کشف آن**

3. آماده سازی داده ها شامل تمیزسازی داده ها
4. اتخاذ بهترین روش داده کاوی برای دستیابی به اهداف
5. اجرای داده کاوی یعنی به کارگیری الگوریتم
6. ارزیابی و اعتبارسنجی یافته ها
7. استفاده از نتایج و تثبیت وتحکیم دانش کشف شده
8. تصمیم گیری براساس دانش کشف شده


دانش
داده های تبدیل شده
الگوها
داده های هدف
داده های
پردازش شده
تفسیر و ارزیابی
داده کاوی
داده
انتخاب
پردازش و
پاک کردن داده ها
تبدیل و انتخاب ویژگی

کاربرد داده کاوی

امروزه داده کاوی در حوزه های بسیارمتنوع و متفاوت استفاده می شود:

حوزه علم:

شیمی، فیزیک، داروسازی، تجزیه و تحلیل تصاویر پزشکی، تعیین نوع رفتار با بیماران و پیش گویی میزان موفقیت های اعمال پزشکی، تعیین میزان موفقیت روش های درمانی در برخورد با بیمارهای سخت، بیوشیمی، حسگرهای کنترل در اقمار مصنوعی،بیوعلم شامل توصیف ژن ها و تقسیم بندی گروه پروتئینها و توسعه داروها.

خرده فروشی:

تجزیه و تحلیل سبد خرید بازار، تعیین الگوهای خرید مشتریان.

بانکداری:

پیش بینی الگوهای کلاهبرداری از طریق کارت های اعتباری، تعیین میزان استفاده ازکارتهای اعتباری براساس گروههای اجتماعی

فروش و بازاریابی:

تجزیه و تحلیل سهام و سرمایه، تعیین مشتریان وفادار، مدیریت ریسک و پیش بینی فروش بیمه و تجزیه و تحلیل دعاوی و پیش گویی میزان جریمه بیمه نامه های جدید توسط مشتریان

دیگر حوزه ها: ورزش وسرگرمی ، فضانوردی و...

موانع و چالش های داده کاوی

- فقدان داده برای پشتیبانی تجزیه و تحلیل وقدرت محدود محاسبه برای به دست آوردن محاسبات ریاضی موردنیاز الگوریتم های داده کاوی
- عدم مطلوبیت و جذابیت بیش ترالگوها
- خطر وجود داده های آلوده و کسب نتایج کاملا غلط
- تمرکز بیش از حد بر الگوریتم ها
- هزینه نسبتا گران سرمایه گذاری در همه ی حوزه ها
- عدم پوشش کامل همه حوزه ها

فنون داده کاوی

فنون داده کاوی در یک نگاه کلی به دومنظور به کار می روند: توصیف و پیش بینی. هدف از به کارگیری فنون پیش بینی کننده، پیش بینی ارزش یک ویژگی خاص بر اساس سایر ویژگی هاست. ویژگی مورد پیش بینی هدف نامیده شده و وابسته به سایر ویژگی هاست و ویژگی هایی که کمک به پیش بینی کنند متغیرهای توضیحی و مستقل هستند اماهدف از بکارگیری فنون توصیفی استخراج الگو است به نحوی که ارتباط بین لایه های زیرین داده ها را خلاصه سازی کند.فنون پیش بینی شامل دسته بندی ، رگرسیون و غیره می شود و فنون توصیفی شامل خوشه بندی ، تشخیص ناهنجاری و مواردی از این دست است تکنیک های داده کاوی مورد استفاده در این تحقیق شامل خوشه بندی و دسته بندی است.

خوشه بندی

خوشه بندی عملی است که موجب تقسیم بندی جمعیتی ناهمگن به تعدادی خوشه همگن می شود . هنگامی که یک فرایند خوشه بندی آغاز می شود تعداد ، شکل و ویژگی های خوشه ها مشخص نیست و از آنجا که هیچ دانش قبلی از خوشه ها وجود ندارد، فن خوشه بندی یک تکنیک بدون ناظر نامیده می شود. از این فن معمولاً برای ساختن بخش ها و خوشه هایی استفاده می شود که برای

مدل های پیش بینی را به دو گروه اصلی دسته بندی و رگرسیون تقسیم بندی می کنند: دسته بندی برای پیش بینی متغیرهای هدف گسسته به کار می رود، درحالی که کاربرد رگرسیون برای پیش بینی متغیرهای هدف پیوسته است هدف اصلی هر دو روش، کاهش خطا بین مقدارپیش بینی شده و مقدار واقعی متغیر هدف است.

دسته بندی

دسته بندی از مدل های پیش بینی در داده کاوی است و به زبان ساده به معنای اختصاص یک شیء داده به یکی از چند دسته یا کلاس از پیش تعریف شده است دسته بندی فرایندی است برای یافتن مدل یاتابعی که دسته های مجزایی را با هدف پیش بینی دسته شیئی که برچسب دسته یا کلاس آن مشخص نیست، تعریف کند .دسته بندی از روشهای با ناظر در داده کاوی است.

این بدان معناست که هدف از انجام دسته بندی کام

لاًمشخص است و متغیر یا متغیرهایی وجود دارند که ارزش

درخت تصمیم

درخت های تصمیم از مدل های دسته بندی هستند. هر درخت تصمیم از تعدادی گره و یال تشکیل شده است درخت های تصمیم بر اساس داده های گذشته ساخته می شوند و هرگاه داده جدیدی با تمام ویژگی هایش (تمام ویژگی ها به جز دسته مربوطه اش)وارد شود، گره های داخلی برروی ویژگی های ورودی آزمون انجام داده و نتیجه آزمون بر روی یال ها می رود و گره های برگ نیز دسته پیش بینی شده برای داده جدید را ارائه خواهند کرد. در ساخته شدن درخت و تشکیل هر گره الگوریتم درخت تصمیم به دنبال انتخاب بهترین ویژگی برای شکستن درخت به دو یا چند زیر درخت است.

فرایند تحلیل سلسله مراتبی (AHP)

برای تعیین میزان اهمیت هریک از عوامل (وزن دهی) میتوان از AHP • که یک فن ریاضی برای تصمیم گیری چند معیاره است استفاده کرد. این فن در سه مرحله انجام می گیرد:

1) **ساختن درخت سلسله مراتبی:** گسترش درختی که درآن آرمان کلی وشرایط تصمیم گیری مشخص شده باشد.

2) مقایسات زوجی پارامترها با پرسش از افراد خبره

- ملاحظات زوجی اساس کار AHP است. این مقایسه ها بصورت دو به دو انجام می شود. روش AHP از اعداد 1 تا 9 برای درجه بندی مقایسه ها استفاده می کند.

3) ارزیابی سازگاری بین مقایسات زوجی

نرخ ناسازگاری معیاری است برای سنجش اعتبار پاسخ های پرسش شوندگان

- I.I = (λmax – n) / (n)
- I.R = I.I / I.I.R
- λmax :حداکثر مقدار ویژه ماتریس
- n بعد ماتریس
- I.I :شاخص ناسازگاری
- I.R :نرخ ناسازگاری
- I.I.R :شاخص ماتریس nدر n که به ازای هر n مقدارمشخصی دارد.

داده کاوی

و کاربرد آن

در بانکداری

بخش بندی مشتریان بانک صادرات ایران
با استفاده از داده کاوی

امروزه یکی از چالش های بزرگ سازمان های مشتری محور، شناخت مشتریان، ایجاد تمایز بین گروه ها ی مختلف مشتریان و رتبه بندی آنهاست. در گذشته تفکیک مشتریان به گروه های مختلف با رویکرد بخش بندی بر اساس نیاز مشتری صورت می گرفت. اما امروزه ارزش مشتری به عنوان عامل قابل اندازه گیری می تواند دربخش بندی مشتریان به کار می رود.

سازمان ها علاوه بر ارتباط با مشتریان نیازمند استفاده از مقیاسی برای سنجش میزان ارزش و اهمیت مشتریان مختلف هستند. این مقیاس درصورتی فراهم خواهد شد که سازمان بتواند با استفاده از ابزار مناسب به میزان ارزش مشتریان خود دست یافته و به تجزیه و تحلیل آن بپردازد.

شناخت گروه های مختلف مشتریان و ایجاد ارتباط اثربخش با آنها به گونه ای که بتوان منافع اقتصادی سازمان رادر آینده تضمین نمود، مسئله ای مهم در کسب و کار امروز است.

جذب مشتریان سودآور ونگهداری و هم چنین حفظ مشتریان ارزشمند قدیمی هر دو دارای اهمیت هستند که جزبا شناسایی دقیق ویژگی های آنها امکان پذیر نمی باشد.

یکی از راه های کسب دانش و شناخت در مورد مشتریان، بخش بندی آنها به گروه های متجانس وبررسی ویژگی های هر بخش است. این کار توازنی بین عدم شناخت مشتریان و شناخت تک به تک آنهاست.

رتبه بندی مشتریان یکی از دغدغه های بانک ها و سایر مؤسسات مالی و اعتباری به منظور اختصاص منابع و تبیین راهبردهای کلان به خصوص در حوزه بازاریابی بودهست. سودمندترین مشتریان بوسیله ابزارهای امتیازبندی مشتریان نظیر مدلهای ارزش طول عمر مشتری شناسایی شده و مورد توجه خاص قرار می گیرند.بنابراین یافتن روشها و مدلهایی برای بخش بندی مشتریان بر اساس میزان ارزش مشتری برای سازمان ضروری به نظر می رسد.جهت تحقق این امر، این امر، بانک ها نیازمندابزاری هستند تا بتوانند مشتریان ارزشمند خود را شناسایی کنند. در این میان شاخص

ارزش طول عمر مشتری می تواند به عنوان ملاک ارزش گذاری مشتریان، چارچوب مناسبی برای این کار باشد.

مدیریت ارتباط با مشتری

تمام شرکت ها فعالیت خود را صرفاً برای مشتری انجام می دهند وبدیهی است نتیجه نهایی متعلق به مشتری است. مدیریت ارتباط با مشتری راهبرد کسب وکاری است جهت بهینه سازی سوددهی، درآمدزایی و رضایت مشتری که بر اساس مبانی زیر طراحی می گردد:

1)ساماندهی ارائه خدمات بر اساس نیازهای مشتری

2)بالا بردن سطح رضایت مشتریان مطابق اصول مشتری مداری

3)پیاده سازی فرایندهای مشتری محور

مدیریت ارتباط با مشتری در واقع فرایندی است جهت گردآوری و یکپارچه سازی اطلاعات به منظور بهره برداری مؤثر و هدفمند از آنها که با جمع آوری اطلاعات مشتریان در یک پایگاه داده امکان طبقه بندی مشتریان را بر اساس راهبردها و معیارهای سازمان فراهم می سازد.

مفهوم ارزش طول عمر مشتری(CLV)

ارزش چرخه عمر مشتری شامل محاسبه جمع ترکیب مالی است. یعنی درآمدمنهای هزینه های یک مشتری (درآمد خالص حاصل شده از مشتری) در طول ارتباطش با سازمان. ارزش طول عمر مشتری روشی است که به حفظ مشتری کمک خواهدکرد. این روش می گوید چه اطلاعاتی از مشتری در دسترس است. سازمان ها با استفاده از این اطلاعات می توانند سودآوری مشتریان خود را پیش بینی و در مورد آنهاتصمیمات آگاهانه ای اتخاذ کنند.

مدلRFM

یکی ازمتدهای معروف و کارا در تحلیل ارزش مشتری است و نقطه قوت آن در این است که خصوصیات مشتریان را با تعداد معیار کمتر (تنها سه بعد) به کمک متدهای خوشه بندی استخراج می کند مدل براساس سه فاکتور شکل گرفته است:

تازگی R و تعداد دفعات Fو ارزش مالیM

تازگی : فاصله زمانی از آخرین مراجعه (خرید، تراکنش مالی).

تعداد دفعات: تعداد دفعات مراجعه (خرید، تراکنش مالی)در یک بازه زمانی مشخص.

ارزش مالی : پول پرداخته شده در یک بازه زمانی معین

چارچوب کلی پروژه

فاز (1) شناخت کسب و کار و داده

فاز (2) آماده سازی و پیش پردازش داده ها

فاز (3) مدلسازی

فاز (4) ارزیابی

فاز (1) شناخت کسب و کار و داده

این فاز شامل پنج مرحله زیر است:

شناخت کسب و کار: سازمان مورد بررسی بانک صادرات ایران (شعب تهران) بوده وحوزه انتخابی در این تحقیق بانکداری خرد است. بانکداری خرد شامل فعالیت های معمولی است که در شعب انجام می گیرد، مانند افتتاح سپرده، واریز به سپرده، برداشت از سپرده و ... . مشتریان این حوزه از بانکداری بزرگترین دسته مشتریان را در بین حوزه های مختلف به خود اختصاص می دهند و اکثر فعالیت های بانک از جمله فعالیت های بازاریابی، تولید خدمات و محصولات جدید و سایر فعالیت ها بر اساس این گروه از مشتریان صورت می گیرد.

بررسی مشکل کسب و کار:

«مسئله رتبه بندی مشتریان»

ایجاد سیستم بخش بندی و رتبه بندی مشتریان بر اساس عوامل مؤثر بر میزان ارزش آنها می تواندراهگشای تبیین برخی از فعالیت های بازاریابی و ارتباط بامشتریان باشد. در این پروژه از مفهوم ارزش طول عمر مشتری و عوامل مؤثر بر ارزش آفرینی مشتری به عنوان عوامل تأثیرگذار بر بخش بندی مشتریان استفاده خواهد شد.

تعیین اهداف تحقیق: هدف از این پروژه، ایجاد مدلی برای بخش بندی و رتبه بندی مشتریان است. برای ایجاد یک سیستم رتبه بندی نیاز به استفاده از تکنیک های بخش بندی در داده کاوی می باشد

جمع آوری داده های اولیه: اطلاعات درخواست شده شامل اطلاعات مربوط به30000 مشتری در طول یک سال بود (ابتدای فروردین ماه 1389 تا ابتدای فروردین ماه 1390). مشتریان مورد نظر باید قبل از ابتدای بازه زمانی به بانک پیوسته باشند واطلاعات مشتریانی که در طول یک سال مذکوربه بانک پیوسته اند مد نظر نیست. اطلاعات مورد نظر شامل دو دسته کلی می شود:

اطلاعات رفتاری (تراکنشی) مشتریان:

این اطلاعات شامل تاریخ آخرین مراجعه(تراکنش)، تعداد تراکنش ها در طول بازه زمانی یک ساله و ارزش پولی تراکنش ها میباشد.

اطلاعات مربوط به خدمات ارائه شده به مشتری از طرف بانک:

این اطلاعات نوع خدمات مورد استفاده هر مشتری را نشان می دهد.

درخواست و دریافت داده ها: در این مرحله طبق اطلاعات درخواستی در قبل، جدولی با این عناوین در اختیار تحقیق قرار گرفت: مشتریان نمونه، تاریخ آخرین تراکنش، تعداد تراکنش ها در بازه زمانی مورد نظر، مجموع مانده تمام سپرده های مشتری در انتهای بازه زمانی، تعداد حساب کوتاه مدت ، بلندمدت، جاری، قرض الحسنه، گنجینه و تعداد خدمات مورد استفاده برای هر مشتری.

فاز (2): آماده سازی و پیش پردازش داده ها

این بخش شامل چهار گام است:

گام اول: حذف داده های ناقص، بی کیفیت و مغشوش.

در این گام برخی ازرکوردهای موجود که دارای اطلاعات مفقوده بودند یا با سایر اطلاعات سازگاری نداشتند حذف شدند.

گام دوم: استخراج داده و ایجاد انباره داده ها. هدف از این گام ایجاد یک انباره داده یکپارچه از مشتریان است. یک انباره داده مخزنی از اطلاعات جمع آوری شده از منابع مختلفی است که به شکل های مختلفی ذخیره شده و ساختار یافته اند.

گام سوم: مقیاس بندی داده های تراکنشی.

دراین گام داده های تراکنشی مقیاس بندی می شوند.

این کاربه کمک متد تحلیل ; RFM که در نرم افزار وجود دارد صورت می گیرد.

گام چهارم: نرمال سازی اطلاعات.

در این گام نرمالسازی بر روی داده های تراکنشی به روش Min-max انجام می گیرد.

فاز (3):مدل سازی

این فاز شامل پنج گام می باشد :

گام اول: مشتریان بر اساس پارامترهای تحلیل RFM تطبیق یافته خوشه بندی می شوند.

گام دوم: بر اساس متدAHP وبا نظر افراد خبره وزن پارامترهای مؤثر در خوشه بندی را محاسبه می کنیم.

گام سوم: محاسبه ارزش طول عمر مشتری برای هر یک از خوشه های موجود و رتبه بندی خوشه ها بر اساس ارزش طول عمر آنها.

گام چهارم: استفاده از تکنیک نمونه برداری ترکیبی برای مواجهه با مسئله عدم توازن بین خوشه ها.

گام پنجم: کشف الگوهای پنهان بین ارزش مشتریان هر خوشه و ویژگی های رفتاراستفاده از خدمات بانک مانند تعداد و نوع سپرده های مشتریان.

فاز (4): ارزیابی

در این مرحله ارزیابی مدلسازی انجام شده با نظر افراد خبره در واحدهای سازمان روش ها، ارتباط مشتریان، مدیریت اطلاعات و کارشناسی داده کاوی صورت گرفت.

اجرای مطالعه موردی

در جدول 1 نوع اطلاعات دریافت شده و توضیحات مربوط به هر یک آورده شده.

نوع اطلاعات	ویژگی‌ها	توضیح
اطلاعات تراکنشی (پارامترهای RFM)	رتبه تازگی (R)	عاملی است متناسب با تاریخ آخرین تراکنش
	تعداد تراکنش (F)	تعداد تراکنش‌های مالی در طول یک سال
	عامل مالی (M)	مجموع مانده تمام سپرده‌های مشتری در پایان یک سال
اطلاعات استفاده از خدمات	حساب جاری	در صورتی که مشتری از این محصول استفاده کند مقدار ۱ و در غیر اینصورت مقدار صفر می گیرد
	حساب کوتاه مدت	در صورتی که مشتری از این محصول استفاده کند مقدار ۱ و در غیر اینصورت مقدار صفر می گیرد
	حساب بلند مدت	در صورتی که مشتری از این محصول استفاده کند مقدار ۱ و در غیر اینصورت مقدار صفر می گیرد
	حساب قرض الحسنه	در صورتی که مشتری از این محصول استفاده کند مقدار ۱ و در غیر اینصورت مقدار صفر می گیرد
	حساب گنجینه	در صورتی که مشتری از این محصول استفاده کند مقدار ۱ و در غیر اینصورت مقدار صفر می گیرد
	تعداد خدمات	تعداد خدمات مورد استفاده هر مشتری

جدول ۱. دسته بندی اطلاعات مشتری

نتایج حاصل از مقیاس بندی پارامترهای RFM در جدول (۲) آورده شده است.

امتیاز	مقیاس	R	F	M
امتیاز ۱	خیلی پایین (VL)	[۸۹/۱/۱,۸۹/۱۲/۲۲]	[۱,٤۷)	[۰,٦۱۹٥٤۸)
امتیاز ۲	پایین (L)	(۸۹/۱۲/۲۲,۸۹/۱۲/۲٦]	[٤۷,۱۹۸)	[٦۱۹٥٤۸,۷۳۰٥٥٦٦)
امتیاز ۳	متوسط (M)	(۸۹/۱۲/۲٦,۸۹/۱۲/۲۷]	[۱۹۸,۳۹۸)	[۷۳۰٥٥٦٦,۲۸٤٦۷٤۰۹)
امتیاز ٤	بالا (H)	(۸۹/۱۲/۲۷,۸۹/۱۲/۲۸]	[۳۹۸,۱۳۲٥)	[۲۸٤٦۷٤۰۹,۱٦٥۸٥۱۱۷۲)
امتیاز ٥	خیلی بالا (VH)	(۸۹/۱۲/۲۸,۸۹/۱۲/۲۹]	[۱۳۲٥,۱۰۹۳۲)	[۱٦٥۸٥۱۱۷۲,۳۸٦٤۸۸۷٦۹۹٤)

جدول ۲. مقیاس بندی پارامترهای RFM

در ابتدا 30000 مشتری به عنوان مشتریان نمونه معرفی شدند و با حذف اطلاعات ناقص 29،579 مشتری باقی ماندند که کار مدل سازی بر روی این تعداد انجام گرفت. برای خوشه بندی مشتریان بر اساس فاکتورهای مؤثر بر ارزش طول عمر مشتری، آنها را با استفاده ازالگوریتم two step خوشه بندی کردیم . میانگین مقادیر M,R,F وتحلیل RFM هریک از خوشه ها در جدول 3 آمده است.

تحلیل RFM	M	F	R	تعداد اعضا	خوشه
L-M-H	۷۱۶۷۴۶۵۸	۲۷۶	۸۹/۱۲/۲۴	۱۷۹۴۸	C_1
VH-VH-VH	۵۳۵۶۱۸۶۳۹۵	۳۷۷۰	۸۹/۱۲/۲۹	۱۲۰۵	C_2
VL-VL-VL	۵۴۷۸۳۷۰	۳۱	۸۹/۵/۲۵	۴۳۴۴	C_3
H-VH-VH	۳۷۲۲۲۳۰۴۸	۱۷۳۷	۸۹/۱۲/۲۸	۶۰۸۲	C_4

جدول ۳. خوشه بندی مشتریان بر اساس پارامترهای RFM

امتیاز RFM هر خوشه را همان طور که در زیرمشاهده می شود به صورت مجموعی از امتیازات هر یک از پارامتر های آن محاسبه می کنیم:

F Score ci + M Score ci + R Score ci = RFM Score ci

که درآن

- RFM Score ci =امتیاز RFM خوشه Ci
- R Score ci =امتیاز پارامتر R در خوشه Ci
- FScore ci =امتیاز پارامتر F در خوشه Ci
- M Score ci =امتیاز پارامتر M در خوشه Ci

نتایج حاصل از امتیاز دهی RFM

خوشه	درصد اعضا	امتیاز R	امتیاز F	امتیاز M	امتیاز RFM
C1	٦٠/٦	٢	٣	٤	٩
C2	٤/١	٥	٥	٥	١٥
C3	١٤/٧	١	١	١	٣
C4	٢٠/٦	٤	٥	٥	١٤

جدول ٤. امتیاز RFM خوشه‌ها

در گام بعدی امتیاز CLV (ارزش طول عمر مشتری) خوشه ها را به شکلی که در رابطه ی زیر(رابطه ی زیر) آمده محاسبه می کنیم:

$$CLV_{ci} = NR_{ci} \times WR_{ci} + NF_{ci} \times WF_{ci} + NM_{ci} \times WM_{ci}$$

NRci مقدار نرمال شده ی پارامتر R در خوشه ی ciو WRci وزن پارامتر R است . NFci مقدار نرمال شده پارامتر F در خوشه ciو WFci وزن پارامتر Fاست . NMci مقدار نرمال شده ی پارامتر M در خوشه ci ودرنهایت WMci وزن پارامتر Mاست.

در این تحقیق از روش نرمالسازی Min-Max که یک انتقال خطی روی داده های اصلی ایجاد می کند استفاده شده است.

چنانچه MaxA و MinA به ترتیب بیشترین وکمترین مقدار ارزش برای ویژگی A باشند این روش نرمال سازی ارزش v پارامتر Aرا به ارزش v´ تبدیل می کند.

$$v' = \frac{v - min_A}{max_A - min_A}$$

از سوی دیگر برای محاسبات رابطه(الف) نیاز به محاسبه وزن پارامترهای RFM خواهد بود. برای این کار ازمتدAHP و نظر افراد خبره سازمان درواحد داده کاوی، سازمان روش ها و مدیریت اطلاعات استفاده شد. بر این اساس وزن های بدست آمده برای پارامترهای مورد نظر در جدول 5 آمده است.

پارامتر	R	F	M
وزن	۰/۱۳۳	۰/۳۱۲۱	۰/۵۵۴۹

جدول ۵. وزن پارامترهای RFM

در جدول 6 مقادیر نرمال شده هریک از پارامترهای RFM, ارزش طول عمر ورتبه خوشه ها بر اساس ارزش CLV آنها آمده است.

خوشه	R نرمال	F نرمال	M نرمال	CLV	رتبه CLV	بخش (سگمنت)
C_1	٠/٩٩٦	٠/٠٢٥	٠/٠٠٢	٠/١٤١٤	٣	٣
C_2	١	٠/٣٤٥	٠/١٣٩	٠/٣١٧٨	١	١
C_3	٠/٣٩٤	٠/٠٠٣	٠	٠/٠٥٣٣	٤	٤
C_4	٠/٩٩٩	٠/١٥٩	٠/٠١	٠/١٨٨٠	٢	٢

جدول ٦. محاسبه CLV و رتبه‌بندی مشتریان

کشف الگوهای پنهان بین بخش بندی مبتنی برارزش مشتریان و ویژگی های استفاده از خدمات بانک

برای کشف الگوهای پنهان بین بخش بندی مبتنی بر ارزش مشتری به عنوان فیلدهدف و اطلاعات استفاده از خدمات بانک به عنوان پیشگویی کننده، می توان الگوریتم هایی را که منجر به ایجاد قوانین می شوند مورد استفاده قرار داد. به علت پدیده عدم توازن بین چهار سگمنت و اینکه سگمنت 3 حدود 60 درصد از کل اعضارا به خود اختصاص داده است، اکثر قوانین به این سگمنت منتهی می شوند. در اینجا ازتکنیک مبتنی بر داده نمونه برداری ترکیبی استفاده شده است تا مشکل نادیده گرفته شدن برخی از الگوها و اطلاعات تا حدی از بین برود. پس از متعادل کردن دسته ها، بااجرای الگوریتم دسته بندی بر روی داده های استفاده از خدمات بانکداری به عنوان عوامل پیش گویی کننده، می توان دسته مشتریان را که همان شماره سگمنت است پیش گویی کرد.

نتیجه گیری و ارزیابی

در این تحقیق سعی شد تا چارچوبی جهت بخش بندی مشتریان بانک صادرات ایران بر اساس عوامل مؤثر برارزش طول عمر مشتری ارائه شود. به این منظور مجموعه ای از تراکنش های مربوط به 30000 مشتری بانک صادرات (شعب تهران) در حوزه بانکداری خرد و در بازه زمانی اول فروردین ماه سال 1389 تا اول فروردین ماه سال 1390 دریافت شد. داده های دریافت شده به دو دسته داده های تراکنشی وداده های استفاده از خدمات تقسیم شدند. داده های تراکنشی شامل تاریخ آخرین تراکنش مشتریان نمونه R,
تعداد تراکنش های هر مشتری در بازه زمانی یک ساله F,

و مجموع مانده تمام سپرده های هر مشتری در انتهای این بازه زمانی بوده M وداده های استفاده از خدمات مربوط به تعداد و نوع سپرده های مشتریان نمونه بوده است.پس از بخش بندی مشتریان بر اساس داده های تراکنشی (M,F,R)با ادغام مشخصات استفاده از خدمات بانکداری و نتایج حاصل از بخش بندی مشتریان بر اساس داده های تراکنشی، سعی شد تا ویژگی های چهار دسته "مشتریان طلایی"، "مشتریان ارزشمند و وفادار"، "مشتریان با احتمال رویگردانی" و "مشتریان کم ارزش" کشف گردد. این کار به کمک تکنیک درخت تصمیم انجام گرفت و

الگوریتم C5.0 به علت دقت بالاتر نتایج در ارزیابی، به عنوان بهترین الگوریتم برای کشف الگوهای پنهان بین بخش های مختلف مشتریان از نظر میزان ارزش و ویژگی های استفاده ازخدمات و محصولات بانکی

انتخاب شدنتیجه نهایی بخش بندی مشتریان، دست یافتن به چهار بخش است. به هریک ازبخش ها با همکاری افراد خبره بانک در واحد داده کاوی و تحلیل اطلاعات، نامی اختصاص یافته و تحلیلی بر روی آنها صورت گرفته است.

بخش اول (مشتریان طلایی):

این بخش از مشتریان ارزشمندترین مشتریان بانک را تشکیل می دهند و امتیاز تمام فاکتورهای تحلیل RFM انطباق یافته شان "خیلی بالا"است. مشتریان خوشه دوم که میانگین ارزش طول عمر آنها 0/3178 است و 4/1% ازکل مشتریان را شامل می شوند این بخش را تشکیل می دهند. خدمات بانک به این مشتریان می تواند تنها محدود به فعالیت های بانکداری خرد نشود. در واقع این مشتریان پتانسیل آن را دارند که بانک برایشان برنامه های ویژه ای داشته باشد و خدمات بانکداری اختصاصی به آنها ارائه کند. طبق نتایج حاصل از کشف الگوهای پنهان بین بخش بندی مبتنی بر ارزش مشتریان و ویژگی های استفاده از خدمات بانکی، مشتریانی که دارای حساب جاری و حساب کوتاه مدت بصورت توأمان هستند، جزء این دسته از مشتریان می باشند.

بخش دوم (مشتریان وفادار و ارزشمند):

مشتریان خوشه چهارم با 20/6درصد ازکل مشتریان، این سگمنت را تشکیل می دهند. امتیاز فاکتور مالی این بخش از مشتریان"خیلی بالا" است ولی محدوده آن کمتر از مشتریان سگمنت اول است. امتیاز رتبه تازگی آنها نیز "بالا" می باشد. ارزش طول عمر این مشتریان 0/1880است. طبق نتایج حاصل از کشف الگوهای پنهان بین بخش های مشتریان و استفاده از خدمات ومحصولات بانک، پیش بینی شده مشتریانی که تعداد خدمات مورد استفاده آنها دو یاسه محصول است که یکی از آنها حساب کوتاه مدت می باشد، به این دسته تعلق دارند. برای پیوستن اعضای این گروه به مشتریان پرارزش تر بخش اول می توان حساب جاری را به آنها پیشنهاد کرد.

بخش سوم (مشتریان با احتمال رویگردانی):

مشتریان خوشه یک با 60 درصد ازکل مشتریان، وارزش طول عمر 0/1414 این سگمنت را تشکیل دهند. مشتریان این بخش فاکتور تازگی قابل توجهی ندارند ("پایین")، اما امتیاز عامل F آنها " متوسط " و وامتیاز فاکتور مالی Mنیز "بالا" است.پایین بودن آخرین تراکنش مشتریان می تواند دو دلیل داشته باشد: این مشتریان اخیراً به رقبا پیوسته اند و یا اینکه تنها دارای حساب های بلند مدت و پس انداز قرض الحسنه هستند. برای پر ارزش شدن این گروه از مشتریان باید با توسعه محصولات مورد علاقه آنها سعی در ارزشمند کردنشان داشته باشیم.

طبق نتایج حاصل از کشف الگوهای پنهان بین بخش های مشتریان و ویژگی های استفاده از خدمات و محصولات بانک، مشتریانی که از دو محصول حساب بلند مدت و پس انداز قرض الحسنه استفاده می کنند به بخش سوم تعلق دارند. این سپرده ها ازنظر قیمت نیز جزء سپرده های گران قیمت بانک هستند.

بخش چهارم (مشتریان کم ارزش):

این بخش شامل خوشه سوم می شود که14/7درصد ازمشتریان را به خود اختصاص داده است. میانگین ارزش طول عمر این بخش 0/0533بوده و رتبه تمام فاکتورهای مربوط به تحلیل RFM آنها "خیلی پایین" است. طبق نتایج حاصل از کشف الگوهای پنهان بین بخش های مشتریان و ویژگی های استفاده از خدمات و محصولات بانک، این مشتریان اصولاً تنها از یک محصول بانک استفاده می کنند که آن نیز عموماً حساب پس انداز قرض الحسنه است.

کاربرد های این تحقیق

این تحقیق در تبیین راهبردهای بازاریابی و تدوین برنامه های مدیریت ارتباط با مشتری بر اساس تفکیک گروهی مشتریان، مثمر ثمر می باشد. از جمله راهبردهایی که می توان بر اساس نتایج این تحقیق پیشنهاد کرد می توان به موارد زیر اشاره نمود:

- طراحی خدمات و محصولات متعدد، متنوع و متفاوت برای مشتریان گروه های مختلف به خصوص مشتریان "ارزشمند و وفادار".

- تبیین برنامه های نگهداشت "مشتریان طلایی" که حجم عمده ای از منابع بانک را تأمین می کنند.
- تخصصی کردن و تفکیک بازاریابی اختصاصی از بانکداری خرد و تبیین برنامه های بانکداری اختصاصی برای گروه "مشتریان طلایی" جهت مدیریت دارایی مشتریان به نحوی که بیشترین بازدهی را برای مشتریان در پی داشته باشد.

- استفاده از نتایج تحقیق برای اعطای تسهیلات، به این صورت که رتبه ارزشی گروهی هر یک از مشتریان
- متقاضی تسهیلات، به عنوان یکی از عوامل ورودی برای سنجش اعتبار مشتری جهت اعطای تسهیلات مورد توجه قرار گیرد.

منابع

1)فصلنامه علمی پژوهشی مطالعات مدیریت (بهبود و تحول) سال بیست و دوم شماره 76 ، بهار و تابستان 91 ,صفحات 179 تا 200

2)فصل نامه انجمن صنایع شوینده،بهداشتی و آرایشی ایران سال دهم، شماره 40 ، بهارسال 1391

3)سایت های مختلف اینترنتی