

## ارائه خدمات مناسب به مشتریان بالقوه با استفاده از تکنیک‌های داده کاوی در حوزه بانک‌داری الکترونیک

محمد رضا تقوا\*  
سید مجتبی حسینی بامکان\*\*

### چکیده

طی دو دهه گذشته، دورنمای رقابت در صنعت بانک‌داری به طور قابل توجهی تغییر یافته است. این امر به علت نیروهای هم‌چون قوانین جدید، جهانی سازی، رشد فن‌آوری و تبدیل به محصول شدن خدمات بانک و افزایش قابل توجه تقاضای مشتریان می باشد. بدین جهت بانک‌ها در پی ارائه خدمات مناسب به مشتریان بالقوه این خدمات هستند و ارائه خدمات جدید با تحلیل بهتری همراه می گردد، به گونه‌ای که با تحلیل رفتاری مشتریان بتوان ریسک حاصل از سرمایه‌گذاری را پیش‌بینی کرد و بخش‌بندی مناسبی از نیازمندی‌های مشتریان برای این خدمات به دست آورد. با این‌گونه تحلیل‌ها می‌توان از هزینه‌های اضافی بازاریابی جلوگیری کرده و همچنین ضریب پذیرش خدمات را نیز افزایش داد. به دلیل آنکه محیط تصمیم‌گیری عوامل مختلفی را شامل می‌شود، به مدلی نیاز است تا هم تعداد متغیرهای بیشتری را تحلیل

\* استادیار، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران (مسئول مکاتبات) taghva@yahoo.com

\*\* کارشناسی ارشد مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران

تاریخ پذیرش: ۸۹/۷/۱۸

تاریخ دریافت: ۸۸/۱۲/۲

کرده و هم قابلیت ارائه اطلاعات جهت تصمیم‌گیری را مقدور سازد. در این مقاله ضمن ارائه مدلی جهت تحلیل رفتار مشتریان با استفاده از شبکه‌های خودسازمان دهنده، به کشف قوانین وابستگی بین خدمات بانک و نحوه مناسب ارائه خدمات به مشتریان بالقوه جهت افزایش ضریب پذیرش خدمات پرداخته شده است. واژگان کلیدی: بانک‌داری الکترونیک، شبکه‌های عصبی، قوانین وابستگی، شبکه‌های خودسازمان دهنده

#### مقدمه

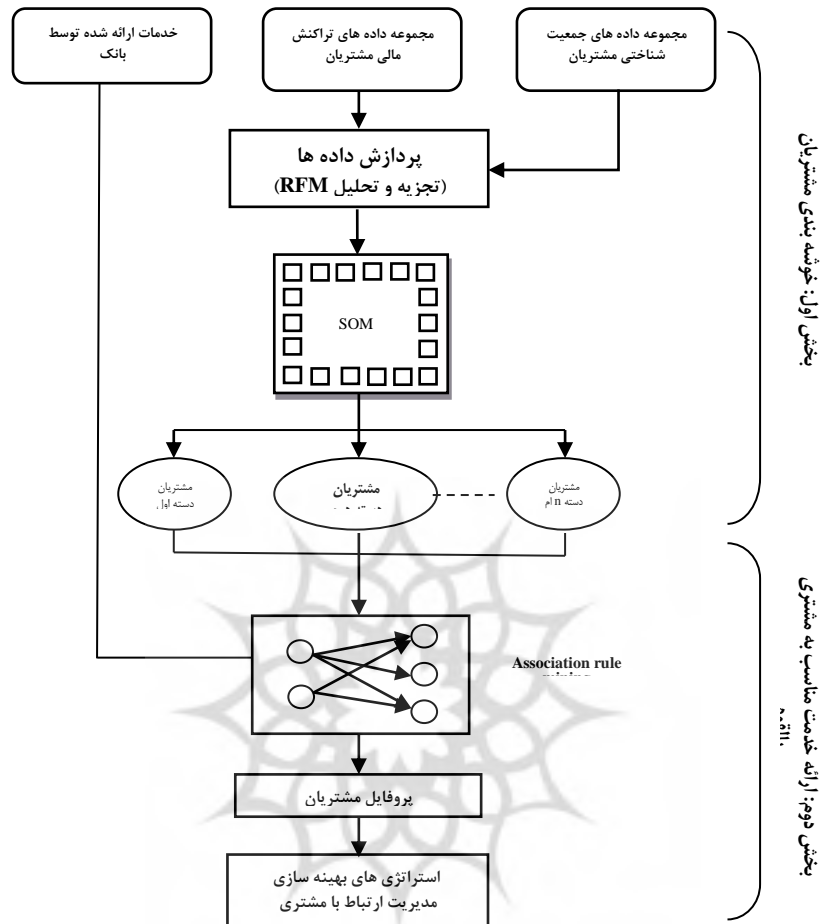
صنعت بانکداری در سرتاسر جهان تحت تغییرات شدید و چشمگیری می‌باشد و با رواج و گسترش بانکداری الکترونیک ثبت اطلاعات تراکنشی راحت‌تر گردیده است و داده‌ها به طور قابل ملاحظه‌ای در حال رشد می‌باشند. با استفاده از داده کاوی برای تجزیه و تحلیل الگوها و روندها، مدیران بانک‌ها می‌توانند با دقت بیشتری پیش‌بینی نمایند که، چه مشتریانی با احتمال بیشتری ارائه محصولات جدید را می‌پذیرند، چه مشتریانی دارای ریسک بیشتری برای عدم پرداخت بدهی‌ها و وام می‌باشند و چگونه می‌توان روابط با مشتریان را سودآورتر نمود. با گسترش ارائه خدمات بانکداری الکترونیکی از جمله، بانک‌داری اینترنتی، ATM<sup>۱</sup>، ارائه کارت‌های اعتباری، انتقال پول به روش الکترونیکی، خریدهای الکترونیکی و غیره باعث شده‌اند داده‌ها با نرخ سریعی تولید و جمع‌آوری شوند. استخراج اطلاعات و دانش و کشف الگوهای پنهان موجود در این حجم عظیم داده‌ها نیازمند تکنیک‌ها و ابزارهای خاص می‌باشد که زمینه‌ساز به کارگیری مبحث داده کاوی در حیطه بانک‌داری الکترونیک را موجب می‌شود. در سالهای اخیر با افزایش دسترسی به داده‌های مشتریان و بهبود قابلیت‌های تحلیل داده‌ها به وسیله روش‌های داده کاوی فعالیت‌های مختلفی جهت تحلیل رفتار مشتریان انجام می‌شود [۹].

استراتژی کنونی بانک‌ها ارائه محصول و خدمت مناسب، به مشتری حائز شرایط می‌باشد [۹]. اما اجرای عملی و اثر بخش این اهداف در عمل چندان ساده نیست.

1. Automated teller machine

چیزی که باعث مشکل می‌شود این است که خدمات متنوع بوده و فعالیت‌ها تحت مجموعه‌ای پیچیده‌ای از محدودیت‌های کسب و کار عمل می‌کنند [۱۱]. به دلیل آنکه محیط تصمیم‌گیری از عوامل مختلفی برخوردار است نیازمند مدلی هستیم تا هم تعداد متغیرهای بیشتر را تحلیل کرده و هم قابلیت ارائه اطلاعات جهت تصمیم‌گیری را مقدور سازد. در این مقاله سعی شده است تا طبق چارچوبی تحقیق که در شکل ۱ ارائه گردیده است در ابتدا مشتریان بانک را با استفاده از شبکه‌های خود سازمان دهنده به خوشه‌های همگن دسته‌بندی نموده و سپس به کشف روابط بین خدمات مورد استفاده هر خوشه پردازیم تا با ارائه محصول مناسب به مشتری بالقوه آن ضریب پذیرش محصولات را افزایش دهیم.





شکل ۱. چارچوب تحقیق [۱۳]

تاکنون تحقیقات مختلفی در زمینه کاربرد داده کاوی در صنعت بانک داری انجام شده است. هسیه<sup>۱</sup> در سال ۲۰۰۴ با استفاده از تکنیک SOM، مدل یکپارچه از داده کاوی و رتبه بندی رفتاری برای تجزیه و تحلیل رفتار مشتریان بانک ارائه داده است. نگای در سال ۲۰۰۹ به مرور و دسته بندی ادبیات در زمینه کاربردهای داده کاوی در مدیریت ارتباط با مشتری پرداخته است [۱۷]. هورموزی و گیلز در سال ۲۰۰۴، داده کاوی را اسلحه ای قدرتمند برای بانک و صنایع خرده فروشی معرفی

1. Hsieh

می‌کنند [۱۱]. آگلیس و چریستوداکیس در سال ۲۰۰۳ با استفاده از قوانین وابستگی به ارائه مدلی جهت پیش بینی برای خدمات بانکداری الکترونیک پرداخته اند [۵] و همچنین حسین جواهری در سال ۲۰۰۸ مدلی پاسخ دهنده<sup>۱</sup> در بازاریابی مستقیم جهت پایان نامه کارشناسی ارشد ارائه داده است [۱۲].

با مطالعه پیشینه تحقیق، مشخص می‌شود که تحقیق حاضر چارچوبی جامع تر جهت خوشه بندی مشتریان با در نظر گرفتن دو دسته از متغیرهای جمعیت شناختی و داده‌های تراکنشی مشتریان ارائه می‌دهد. در مرحله بعد با استفاده از قوانین وابستگی خدمات مورد علاقه خوشه‌ها به دست آمده و زمینه به کارگیری استراتژی مناسب جهت ارائه خدمت مناسب به مشتری بالقوه آن فراهم شده است. در این تحقیق سعی شده است تا به دو پرسش زیر پاسخ داده شود.

۱. ارائه خدمات بانکی مناسب به مشتریان بالقوه با استفاده از تکنیک‌های داده کاوی چگونه می‌باشد؟
۲. طبقه بندی مشتریان بانک بر اساس مدل های هوشمند داده کاوی چگونه است؟

## ادبیات تحقیق

### داده کاوی در صنعت بانکداری

تعاریف متفاوتی از داده کاوی وجود دارد ولی تعریفی که در اکثر مراجع به اشتراک ذکر شده عبارت است از "استخراج اطلاعات و دانش و کشف الگوهای پنهان از یک پایگاه داده‌های بسیار بزرگ و پیچیده" [۶]. داده کاوی یک روش شناسی بسیار قوی و با پتانسیل بالا می‌باشد که به سازمان‌ها کمک می‌کند که بر مهمترین اطلاعات از مخزن داده‌های خود تمرکز نمایند [۸]. با توجه به این امر که داده کاوی فرآیند اکتشاف مدل های گوناگون، خلاصه‌ها و ارزشهای نشأت گرفته از مجموعه خاصی از داده است [۱۰]. برای اجرای چنین فرآیندی باید از روش شناسی خاصی استفاده کرد. در این راستا روش شناسی فرآیند استاندارد میان

صنعتی داده کاوی<sup>۱</sup> به وسیله تحلیل نمایندگی های دایملر کرایسلر ایجاد شد [۷]. با توجه به این روش، یک پروژه داده کاوی چرخه عمری متشکل از شش گام دارد. این گام ها به صورت مستمر و تکراری در تمام فرآیند داده کاوی به کار گرفته می شوند. گام های روش شناسی داده کاوی CRISP عبارتند از [۷]:

گام درک موقعیت کسب و کار

گام درک داده ها<sup>۲</sup>

گام پیش پردازش داده ها<sup>۳</sup>

گام مدل سازی

گام ارزیابی مدل و استنتاج مدل

گام به کارگیری

داده کاوی نه تنها موجب بهبود تصمیم گیری در سراسر سازمان از طریق استخراج اطلاعات مهم از داده های موجود و جستجو روابط و الگوها از داده های جمع آوری شده توسط سازمان خواهد شد، بلکه می تواند موجب کاهش داده های اضافی شود. یکی از گسترده ترین زمینه های کاربرد داده کاوی برای صنعت بانکداری، مدیریت ارتباط با مشتری می باشد. جهت مدیریت مؤثر ارتباط با مشتری، بانک ها می توانند با تجزیه و تحلیل پایگاه داده مشتریان و ارائه خدماتی که مشتریان واقعاً خواهان آن هستند، مبالغ قابل توجهی را برای تبلیغات و ارائه خدمات صرفه جویی نمایند. بنابراین نیاز است تا بازاریابان بر مشتریان خود از طریق یادگیری بیشتر درباره آنها تمرکز نمایند.

در این حوزه بانک ها در پی ارائه خدمات مناسب، به مشتریان بالقوه این خدمات و خدمات می باشند و ارائه خدمات جدید با تحلیل بهتری همراه گردد به گونه ای که با تحلیل رفتاری مشتریان بتوان ریسک حاصل از سرمایه گذاری را پیش بینی نمود و بخش بندی مناسبی از نیازمندیهای مشتریان برای این خدمات به دست آورد. با این گونه تحلیل ها می توان از هزینه های اضافی بازاریابی جلوگیری کرده و در ضمن

1. Cross Industry standard process (CRISP)

2. Data understanding phase

3. Data Preprocessing

ضریب پذیرش خدمات را نیز افزایش داد.

### شبکه های خود سازمان دهنده<sup>۱</sup> (SOM)

از نظر مفهومی خوشه‌بندی یعنی گروه بندی یک سری موجودی در گروه های مختلف، به طوری که این گروه‌ها نشان دهنده مفهوم یا معنی خاصی باشند و یا به عبارت ساده‌تر به یکدیگر شبیه باشند. از دیدگاهی علمی‌تر خوشه‌بندی یعنی دسته‌بندی داده‌ها به  $k$  گروه مختلف به طوری که داده‌هایی که در یک دسته قرار می‌گیرند به یکدیگر شبیه باشند و داده‌های دسته‌های مختلف با یکدیگر تفاوت داشته باشند. به طور کلی می‌توان تکنیکهای خوشه‌بندی را در دو بخش آماری و فراگیری ماشین تفکیک کرد. دو شیوه معمول خوشه بندی آماری مشاهدات، سلسله مراتبی و تجزیه‌ای می‌باشند. روش های فراگیری ماشین عموماً رویه‌های خودکاری هستند که بر اساس عملیاتی منطقی یا دودویی طراحی شده اند. این مدل‌ها نظام یا حالتی را از نمونه‌ها یاد می‌گیرند که آنها را قادر می‌سازد تا با مسائل پیچیده تر با داده‌های کافی به طور مقبول رویارو شوند. تاکنون توجه زیادی معطوف رویکرد شبکه عصبی، خصوصاً شبکه های خودسازمان دهنده شده است. شبکه‌های خود سازمانده به طور گسترده‌ای در داده کاوی، نمایش فضاهای پیچیده و خوشه‌بندی فضاهای با ابعاد بالا و به طور خاص در پردازش تصویر، کنترل فرآیند، مدیریت پروژه، تحلیل های مالی و تشخیص های صنعتی و پزشکی مورد استفاده قرار می‌گیرد [۱۸]. این شبکه از انواع شبکه‌های عصبی با قابلیت یادگیری بدون ناظر<sup>۲</sup> می‌باشد که در تحلیل فضاهای پیچیده، توانایی زیادی دارد. این مدل از شبکه های عصبی اولین بار توسط کوهنن<sup>۳</sup> در ۱۹۸۱ معرفی گردید [۱۴]. قابلیت این شبکه در ارائه خروجی شبکه در قالب نقشه های گرافیکی گویا و قابل فهم برای مدیران سازمان‌ها، توانایی نمایش روابط خطی و غیرخطی بین متغیرها، قدرت بالا در دسته‌بندی دادگان، عدم حساسیت شبکه به تعداد دادگان تعلیم و حساسیت کم این

1. Self-Organizing Map(SOM)  
2. Unsupervised  
3. Kohonen

نوع شبکه‌ها به وجود نويز در دادگان تعليم اين ابزار را به ابزاري كارآمد جهت استفاده در مدل سازي تبديل نموده است.

### يادگيري در شبكه‌هاي عصبی

از بين تمامی خواص شبكه عصبی، خاصیت يادگيري از اهمیت ویژه‌ای برخوردار است. شبكه‌هاي عصبی به عنوان سیستم‌هاي يادگير دارای اين توانایی هستند که از گذشته، تجربه و محیط بیاموزند و رفتار خود را در حین يادگيري بهبود بخشند. در حالت کلی دو نوع يادگيري با ناظر<sup>۱</sup> و بدون ناظر<sup>۲</sup> وجود دارد.

#### يادگيري با ناظر

در يادگيري با ناظر به قانون يادگيري مجموعه‌ای از زوج داده‌هاي به نام «داده‌هاي يادگيري»  $\{(p^i, t^i); i=1,2,\dots,L\}$  می‌دهند که در آن  $p^i$  ورودی به شبكه و  $t^i$  خروجی مطلوب شبكه برای ورودی  $p^i$  است. پس از اعمال ورودی به شبكه عصبی در خروجی شبكه  $a^i$  با  $t^i$  مقایسه شده و سپس خطای يادگيري محاسبه و از آن جهت تنظیم پارامترهاي شبكه استفاده می‌شود.

#### يادگيري بدون ناظر

در يادگيري بدون ناظر یا يادگيري خود سازمانده، پارامترهاي شبكه تنها توسط پاسخ سیستم اصلاح و تنظیم می‌شوند. به عبارتی دیگر تنها اطلاعات دریافتی از محیط به شبكه را بردارهاي ورودی تشکیل می‌دهند و بردار جواب مطلوب به شبكه اعمال نمی‌شود [۴].

#### الگوریتم کوهنن

نگاشت خودسازمان دهنده کوهنن از نوع يادگيري بدون ناظر می‌باشد. الگوریتم آموزشی گره‌هاي شبكه را به صورت همسایه‌هاي محلی منظم می‌کند، به طوری که بتواند ویژگی‌هاي داده‌هاي ورودی را طبقه بندی کند. نقشه توپوگرافی شبكه به طور

1. Supervised learning  
2. unsupervised learning



خود کار با مقایسه دوره‌ای ورودی‌ها هر گره با برداری که توسط آن گره در خطوط ارتباطی آن ذخیره شده است شکل می‌گیرد. هیچ جواب مطلوبی برای ورودی‌های آموزشی تعریف نمی‌شود.

## الگوریتم شبکه کوهنن [۱۹]، [۱]

۱. مقادیر اولیه را تعیین کنید.

فرض کنید  $w_{ij}(t)$  ( $0 \leq i \leq n-1$ ) مقدار ضرایب وزنی از ورودی  $i$  به گره خروجی  $j$  در زمان  $t$  باشد. میزان اولیه ضرایب وزنی از  $n$  ورودی به  $j$  گره خروجی را با مقادیر کوچک تصادفی معین کنید. شعاع همسایگی گره های خروجی  $j$ ،  $N_j(0)$ ، را در ابتدا بزرگ تعیین کنید.

۲. ورودیها را به شبکه عرضه کنید.

بردار ورودی  $x_0(t), x_1(t), x_2(t), \dots, x_n(t)$  عرضه کنید.  $x_i(t)$  به معنای مقدار ورودی گره  $i$  در زمان  $t$  است.

۳. فاصله‌ها را محاسبه کنید.

فاصله  $d_j$  بین بردار ورودی و بردار خروجی هر گره  $j$  توسط فرمول ۱ محاسبه می‌شود:

$$d_j = \sum_{i=0}^{n-1} ((x_i(t) - w_{ij}(t))^2)$$

فرمول ۱: محاسبه فاصله فاصله  $d_j$  بین بردار ورودی و بردار خروجی هر گره  $j$

۴. کوتاه ترین فاصله را انتخاب کنید.

گره خروجی دارای کوتاه ترین فاصله را با علامت  $j^*$  مشخص کنید.

۵. ضرایب وزنی را اصلاح کنید.

ضرایب وزنی گره  $j^*$  و همسایگان آن گره که در فاصله همسایگی  $N_{j^*}(t)$  قرار دارند را اصلاح کنید. ضرایب وزنی جدید با استفاده از فرمول ۲ محاسبه می‌شود:

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

فرمول ۲: محاسبه ضرایب وزنی گره  $j^*$  و همسایگان آن گره که در فاصله همسایگی  $N_{j^*}(t)$

برای تمامی گره های  $j$  که در فاصله  $N_{j^*}(t)$  قرار دارند و ( $0 \leq i \leq n-1$ ) می‌باشد.

عبارت  $\eta(t)$  ضریب بهره است ( $0 < \eta(t) < 1$ ) که به تدریج در طول زمان کاهش می‌یابد.

بنابراین ضریب بهره ضرایب وزنی به تدریج کند می‌شود. اندازه شعاع همسایگی  $N_{j^*}(t)$  نیز به

تدریج کاهش می‌یابد و بدین صورت محدوده بیش ترین فعالیت به تدریج موضعی می‌گردد.

۶. با رفتن به مرحله ۲ الگوریتم تکرار شود.

## قوانین وابستگی

قوانین وابستگی داده کاوی که ابتدا توسط آگراول و همکاران (۱۹۹۳) توسعه داده شدند، یکی از معروفترین تکنیکهای داده کاوی بوده و کاربردهای وسیعی در کسب و کار سنتی و کسب و کارهای الکترونیکی پیدا کرده‌اند. فروش متقابل، طراحی کاتالوگ، تعیین چیدمان فروشگاه و بخش‌بندی مشتریان و همچنین بهبود و شخصی‌سازی صفحات وب سایت نمونه‌هایی از کاربرد قوانین وابستگی داده کاوی در کسب و کارهای سنتی و الکترونیکی می‌باشند [8].

اگر  $A \subset J, B \subset J, \text{ and } A \cap B = S$  آنگاه  $A \Rightarrow B$  یک قانون وابستگی است. پشتیبان این قانون در  $D$  به صورت درصد تراکشنهایی تعریف می‌شود که  $A \cup B$  را شامل شوند. آن را می‌توان معادل احتمال اجتماع  $A$  و  $B$  دانست، یعنی  $P(A \cup B)$ . میزان اطمینان به این قانون را به صورت درصد تراکشنهایی در  $D$  است که علاوه بر شمول  $A$ ،  $B$  را نیز شامل شوند و معادل احتمال وقوع  $B$  است به شرط وقوع  $A$ ، یعنی  $P(B|A)$ .

اگر  $A$  و  $B$  دو مجموعه آیت باشند،  $A \Rightarrow B$  یک قانون وابستگی می‌باشد. در این صورت معیارهای قوانین وابستگی عبارتند از: پشتیبانی (Support): نشان دهنده تعداد دفعاتی است که قانون در داده‌های تراکشنی ظاهر شده است.

حداقل حمایت: حداقل پشتیبانی قانون مورد نظر که از قبل توسط کاربر تعیین می‌شود.

$A \Rightarrow B$  با  $\text{Support}=50\%$ ، یعنی  $A, B$  در  $50\%$  تراکشنها با هم خریداری می‌شوند.

اطمینان (Confidence): نشان دهنده تعداد دفعاتی است که اگر قسمت اول قانون اتفاق بیفتد، قسمت دوم نیز اتفاق می‌افتد. اطمینان صحت قانون را نشان می‌دهد.  $\text{Confidence}=85\%$ ، یعنی در  $85\%$  موارد، کسی که  $A$  را خرید می‌کند،  $B$  را هم خرید می‌کند.

حداقل اطمینان: حداقل اطمینان قانون مذکور که از قبل توسط کاربر تعیین می‌شود.

بهبود (Lift): معیاری برای ارزیابی اینکه وجود یک قانون به چه میزان بهتر از نبودن آن است و طبق فرمول شماره ۳ محاسبه می‌گردد.

$$\text{Lift} = \frac{P(Y/X)}{P(Y)} = \frac{P(X \cap Y)}{P(x).P(Y)} \text{Lift} = \text{Confidence} (X \Rightarrow Y) / \text{Support} (Y)$$

فرمول ۳. معیار بهبود

اگر بهبود از عدد یک تجاوز کند، هم زمان اتفاق افتادن  $X, Y$  بیشتر از مستقل رخ دادن آنها است. همچنین اگر برابر با عدد ۵ باشد، یعنی اگر  $X$  در سبد باشد،  $Y$  ۵ بار بیشتر در آن سبد قرار می‌گیرد، در مقایسه با زمانی که  $X$  در سبد نباشد.

$Lift > 1$ : جفت کالاهایی که از نظر آماری مستقل اند و قانون استفاده شود بهتر است.

$Lift < 1$ : بهتر است از قانون استفاده نشود و بر اساس حدس الگوی خرید را تعیین نمود [۲].

### تجزیه و تحلیل روش تحقیق

در این مقاله با هدف کشف قوانین وابستگی بین خدمات بانک و ارائه خدمات بانکی مناسب به مشتریان بالقوه آنها از داده‌های موجود در پایگاه اطلاعاتی بانک پارسیان استفاده شده است. بر این اساس سعی شده تا با کاوش این اطلاعات و بررسی الگوی رفتاری گذشته مشتریان، روابط میان متغیرها شناسایی و دانش جدیدی استخراج گردد. در تحقیق حاضر با توجه به بررسی ادبیات موضوع از سه دسته متغیر استفاده شده است که در جدول شماره ۱ آورده شده است [۱۱] و [۱۳] داده‌های جمعیت شناختی: این متغیرها شامل جنسیت، سن، تحصیلات، وضعیت تأهل، شغل و ... می‌باشد. خدمات ارائه شده توسط بانک: خدمات مورد مطالعه در این تحقیق در حوزه بانک داری الکترونیک می‌باشد. تراکنش های مالی مشتریان: این متغیر شامل مبالغ برداشت یا واریز مشتری در بازه زمانی مورد تحقیق است.

جدول ۱. متغیرهای تحقیق

ردیف	متغیرها	توضیح
۱	ID	شماره اختصاص داده شده به هر مشتری
۲	age	سن
۳	birth-L	محل تولد
۴	Sex	جنسیت
۵	education	تحصیلات
۶	Job	شغل
۷	R	فاصله میان زمان آخرین تراکنش مشتری تا زمان مورد ارزیابی
۸	F	تعداد خریدهای مشتریان در یک بازه زمانی خاص
۹	M	میانگین مبالغ تراکنش های مشتری در یک بازه زمانی خاص
۱۰	ATM-BV	استفاده از خودپرداز برای برداشت پول
۱۱	ATM-EM	استفاده از خودپرداز برای اعلام موجودی
۱۲	ATM-EV	استفاده از خودپرداز برای انتقال وجه
۱۳	ATM-SH	استفاده از خودپرداز برای دریافت صورت حساب
۱۴	ATM-PG	استفاده از خودپرداز برای پرداخت قبض
۱۵	Int-Kh	استفاده از اینترنت برای خرید کالا
۱۶	Int-EM	استفاده از اینترنت برای اعلام موجودی
۱۷	Int-EV	استفاده از اینترنت برای انتقال وجه
۱۸	Int-SH	استفاده از اینترنت برای دریافت صورت حساب
۱۹	Int-Chek	استفاده از اینترنت برای فعالیت های مربوط به چک
۲۰	Int-PG	استفاده از اینترنت برای پرداخت قبض
۲۱	Tell-EM	استفاده از اینترنت برای اعلام موجودی
۲۲	Tell-EV	استفاده از اینترنت برای انتقال وجه
۲۳	Tell-PG	استفاده از اینترنت برای پرداخت قبض
۲۴	Tell-Chek	استفاده از اینترنت برای فعالیت های مربوط به چک
۲۵	Mobile-EM	استفاده از موبایل برای اعلام موجودی
۲۶	Mobile-PG	استفاده از موبایل برای پرداخت قبض
۲۷	Mobile-SH	استفاده از موبایل برای دریافت صورت حساب
۲۸	POS-Kh	استفاده از سیستم POS برای پرداخت مبالغ خرید
۲۹	POS-EM	استفاده از سیستم POS برای اعلام موجودی
۳۰	POS-SH	استفاده از سیستم POS برای دریافت صورت حساب

متغیرهای تعریف شده در جدول ۱ برای خوشه‌بندی مشتریان با استفاده از شبکه خودسازمان دهنده استفاده شده است. قبل از اینکه این متغیرها وارد شبکه شوند باید نرمال‌سازی شوند. برای نرمال‌سازی داده‌ها در این تحقیق از نرمال‌سازی max-min استفاده شده است، این روش بر فاصله میان نقاط حداقل و حداکثر تأکید دارد و مقیاس‌سازی را بر اساس تفاوت این دو مقدار انجام می‌دهد و طبق فرمول شماره ۴ محاسبه می‌شود [۴].

$$X^* = \frac{X_i - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

فرمول ۴. نرمال‌سازی max-min

### خوشه‌بندی مشتریان

همان‌گونه که در قسمت ۲-۲ به طور مفصل توضیح داده شد، برای خوشه‌بندی مشتریان از تکنیک SOM استفاده گردید. متغیرهای رفتاری یکی از متغیرهای بسیار پر استفاده در تحلیل رفتار مشتریان هستند. از این متغیرها به عنوان ورودی شبکه عصبی خودسازمان دهنده استفاده شده است. در شبکه‌های عصبی باید ابتدا متغیرهای ورودی نرمال‌سازی شوند. در انتهای این بخش مشتریان بر اساس ویژگی‌های مشترک خود در گروه‌های مختلفی تقسیم‌بندی می‌شود. از این گروه‌ها می‌توان برای شناسایی ترجیحات مشتریان و خصوصیات مشترک میان مشتریان استفاده کرد.

برای انجام خوشه‌بندی از نرم افزار SPSS Celementine و ماژول نگاشت خود سازمان دهنده کوهن استفاده شده است. خروجی حاصل از شبکه به صورت یک ماتریس ۵×۳ می‌باشد طبق جدول شماره ۲ که مشتریان بر اساس میانگین متغیرهای به دست آمده برای هر کدام از قسمتهای ماتریس خوشه‌بندی می‌شوند و بر اساس شباهت خود در این معیارها در خوشه مشابه قرار می‌گیرند. خوشه‌بندی مشتریان در این قسمت به ما کمک می‌کنند تا خصوصیات مشتریان مشابه را ارزیابی کرده و ترجیحات هر گروه را مشخص سازیم.

جدول ۲. خوشه‌های مشتریان حاصل از شبکه‌های SOM

\$KY- Kohonen	2	R:0.453 F:0.5 M:0.446 Records: 981	R:0.415 F:0.439 M:0.466 Records: 189	R:0.392 F:0.523 M:0.464 Records: 354	R: 0.354 F:0.556 M:0.504 Records: 181	R: 0.482 F: 0.613 M: 0.423 Records: 357
	1	R:0.535 F:0.457 M:0.467 Records: 142		R:0.514 F:0.325 M:0.5 Records: 91	R: 0.642 F: 0.307 M: 0.537 Records: 30	R:0.419 F: 0.57 M:0.493 Records: 27
	0	R: 0.54 F: 0.481 M: 0.431 Records: 829	R:0.526 F:0.329 M:0.485 Records: 133	R:0.444 F:0.492 M:0.445 Records: 823		R: 0.449 F: 0.503 M: 0.45 Records: 863
	0	1	2	3	4	

\$KX- Kohonen

نتایج حاصل از خوشه‌بندی مشتریان بانک با توجه به نگاشت خودسازمان دهنده کوهنن که در جدول فوق نشان داده شده است که مشتریان را ۱۳ خوشه دسته‌بندی نموده است. در ادامه تحقیق به بررسی الگوی رفتاری و کشف قوانین وابستگی در هر خوشه می‌پردازیم. بدین منظور بزرگترین خوشه که هم مشتریان بیشتری دارد و به لحاظ متغیرهای مالی و فراوانی نیز میانگین بالاتری دارند انتخاب شده است. تعداد بالاتر این مشتریان و میانگین بالاتر متغیر RFM، این گروه از مشتریان را با اهمیت تر ساخته است. بنابراین خوشه  $X=0$  &  $Y=2$  با تعداد مشتری ۹۸۱ و متغیرهای  $R:0.453$ ،  $F:0.5$ ،  $M:0.446$  به عنوان خوشه مورد بررسی در نظر گرفته شده است.

### کشف قوانین وابستگی

نتایج حاصل از خوشه‌بندی مشتریان به عنوان ورودی این مرحله برای شناسایی خدمات مورد علاقه هر خوشه استفاده شده است. برای شناسایی رابطه میان گروه‌های مختلف مشتریان و خدمات مورد علاقه آنها از الگوریتم Apriori استفاده شده است [۱۵]. در این مرحله قبل از تحلیل مشتریان خوشه انتخابی، مشتریان این خوشه به دو گروه داده‌های آموزشی و آزمون تقسیم شده‌اند. گروه آزمایش برای

ارزیابی صحت روابط به دست آمده و بررسی نتایج پیشنهادها استفاده می‌شود. در این خوشه ۸۰ درصد از مشتریان به عنوان داده‌های آموزش و ۲۰ درصد آنها نیز به عنوان تست برای ارزیابی مدل انتخاب شده‌اند. با استفاده از داده‌های آموزشی، برای شناسایی رابطه میان خدمات مورد استفاده در خوشه انتخابی از الگوریتم Apriori با حداقل درجه پشتیبانی ۲۰ درصد و حداقل اطمینان ۷۰ درصد استفاده شده است. تعدادی از قوانین ایجاد شده در این خوشه در جدول ۳ آورده شده است، قوانین بدست آمده نشان دهنده رابطه میان خدمات مختلف در خوشه مورد بررسی است.

جدول ۳. قوانین ایجاد شده از خدمات مورد استفاده خوشه مورد نظر

Row	Consequent	Antecedent	Support %	Confidence %
1	POS-Kh	ATM-BV and Int-SH and ATM-EM	11.51	68.70
2	POS-Kh	ATM-BV and Tell-PG and Int-SH	11.41	68.42
3	Int-SH	ATM-PG and ATM-BV	8.21	68.29
4	Mobile-SH	Int-Kh	8.91	66.29
5	ATM-BV	Int-PG and POS-Kh	10.61	66.04
6	POS-Kh	ATM-BV and Int-SH and ATM-EM	13.21	65.91
7	Tell-PG	Int-PG	11.71	65.81
8	POS-Kh	ATM-BV	11.71	65.81
9	Tell-PG	Int-SH and ATM-EM	13.11	65.65
10	Tell-PG	POS-Kh and ATM-EM	12.51	65.60
11	Int-Kh	ATM-PG and ATM-BV and Tell-EM	9.31	65.59
12	ATM-EV	Int-PG and Tell-EM	8.71	65.52
13	ATM-BV	POS-Kh and Int-SH	11.21	65.18
14	ATM-BV	Int-PG and POS-Kh and Int-SH	11.91	64.71
15	ATM-BV	Int-PG and POS-Kh and Int-SH	10.21	64.71

با توجه به قوانین ایجاد شده مشخص شد که این گروه از مشتریان از ۱۲ خدمت بانکی که در زیر آمده است بیشتر از سایر خدمات استفاده می‌نمایند.

ATM-BV: استفاده از خود پرداز برای برداشت پول

ATM-EM: استفاده از خود پرداز برای اعلام موجودی

ATM-PG: استفاده از خودپرداز برای پرداخت قبض

ATM-EV: استفاده از خود پرداز برای انتقال وجه

Mobile-EM: استفاده از موبایل برای اعلام موجودی  
Mobile-SH: استفاده از موبایل برای دریافت صورت حساب  
Tell-PG: استفاده از تلفن برای پرداخت قبض  
Tell-EM: استفاده از تلفن برای اعلام موجودی  
POS-KH: استفاده از سیستم POS برای پرداخت مبالغ خرید  
Int-SH: استفاده از اینترنت برای دریافت صورت حساب  
Int-KH: استفاده از اینترنت برای خرید کالا  
Int-PG: استفاده از اینترنت برای پرداخت قبض

#### ارائه خدمت مناسب به مشتری بالقوه

با شناسایی خدمات مورد علاقه مشتریان در هر خوشه می‌توان این خدمات را به مشتریان با خصوصیات مشابه پیشنهاد کرد. در قسمت قبل، خدمات مورد علاقه خوشه مورد نظر از مشتریان شناسایی شدند. در ادامه می‌بایست این خدمات با خدمات مورد استفاده توسط مشتری همان خوشه مقایسه گردد. اگر مشتری از خدمت مورد نظر استفاده نکرده باشد در این مرحله می‌توان این خدمت را به وی پیشنهاد نماییم. با تحلیل اختلاف میان محصولات مورد علاقه هر خوشه و محصولات مورد استفاده توسط هر مشتری در آن خوشه، خدماتی را که می‌توان به وی پیشنهاد نماییم شناسایی می‌گردد. به علت زیاد بودن تعداد مشتریان این خوشه و حجم زیاد جدول کلی سعی شده است تا در جدول ۴ برخی از مشتریان تحلیل و خدمات مورد نیاز آنها شناسایی گردد. با استفاده از حروف T و F نشان داده شده است که آیا مشتری مورد نظر از این خدمت استفاده می‌کند یا خیر، بدین صورت که برای خدمات مورد استفاده مشتری حرف T در نظر گرفته شده است. در جدول شماره ۴ در صورتی که مشتری از خدمات مورد نظر استفاده می‌نماید با حرف T مشخص گردیده است و از حرف F جهت نمایش عدم استفاده آن خدمت استفاده شده است. در سطر مربوط به ارائه پیشنهاد به مشتری در صورتی که خدمتی با حرف F می‌باشد با نقطه مشخص گردیده است که این خدمت باید به مشتری پیشنهاد گردد.



Services	ID	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real	5	T	T	F	F	T	F	F	T	F	T	T
Offer				سے			سے			سے		
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		T	T	T	T	F	F	F	T	F	T	
Offer												
Services	34	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real		T	T	T	T	T	T	F	F	F	F	T
Offer											سے	
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		T	F	T	T	T	T	T	F	T	T	
Offer												
Services	49	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real		T	F	T	F	T	F	F	T	T	T	T
Offer			سے				سے					
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		F	T	T	T	F	T	T	T	F	T	
Offer												
Services	53	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real		F	T	T	F	F	F	F	F	T	F	T
Offer						سے	سے					
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		T	F	F	T	T	T	F	T	F	F	
Offer												
Services	850	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real		F	F	T	F	F	T	F	F	T	T	T
Offer			سے			سے						
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		T	T	F	F	F	F	F	T	T	F	
Offer												
Services	1294	ATM-BV	ATM-EM	ATM-EV	ATM-SH	ATM-PG	Int-Kh	Int-EM	Int-EV	Int-SH	Int-Chek	Int-PG
Real		F	T	F	T	F	F	F	F	T	F	F
Offer						سے	سے					سے
Services		Tell-EM	Tell-EV	Tell-PG	Tell-Chek	Mobile-EM	Mobile-PG	Mobile-SH	POS-Kh	POS-EM	POS-SH	
Real		T	T	T	F	T	T	T	T	T	T	
Offer												

جدول ۴. ارائه خدمت مناسب به مشتری بالقوه آن

## ارزیابی صحت پیشنهاد

برای ارزیابی صحت پیشنهادها از متغیر آماری MAE بر روی داده‌های آزمون و آموزش استفاده شده است. میانگین قدرمطلق خطا (MAE) یکی از معیارهایی است که در تحقیقات مختلفی استفاده شده است [۱۶] و در معادله زیر نحوه محاسبه این معیار ارائه گردیده است. گروه آزمون آن دسته از مشتریان می‌باشند که پس از خوشه‌بندی و انتخاب خوشه مورد نظر وارد مرحله شناسایی خدمات مورد علاقه نشده‌اند، بنابراین می‌توان نتایج حاصل از کشف قوانین وابستگی در خوشه مورد نظر را با عملکرد واقعی این گروه مقایسه نمود. برای محاسبه میانگین قدرمطلق خطا، در ابتدا براساس خدمات شناسایی شده برای این خوشه از مشتریان در داده‌های آموزش، پیشنهادها سیستم برای این گروه نیز به دست آمده است. در ادامه با مقایسه اختلاف میان خدمات استفاده شده و پیشنهادها سیستم، متغیر آماری MAE فرمول ۵ برای هر دو گروه آموزش و آزمون محاسبه شده است.

$$Ei = \frac{1}{n} \sum_{i=1}^N |Pi - Ti|$$

فرمول ۵: متغیر آماری MAE

در این معیار Pi پیشنهادات سیستم و Ti داده‌های واقعی می‌باشد.

جدول ۵. مقایسه میانگین قدر مطلق خطا در گروه آزمون و آموزش

معیار	گروه آموزش	گروه تست
میانگین قدر مطلق خطا	۰,۴۹۳۸۸۵	۰,۴۹۲۳۰۸

نتایج به دست آمده در جدول ۵ نشان دهنده اختلاف ناچیز پیشنهادها در دو گروه آزمون و آموزش است. بر این اساس می‌توان استدلال کرد که چارچوب طراحی شده برای ارائه خدمت مناسب به مشتریان بالقوه آن از عملکرد مطلوبی برخوردار است.

## نتیجه گیری

در نهایت با به کارگیری تکنیک های هوشمند داده کاوی به این سوال که طبقه بندی مشتریان بانک براساس روش های هوشمند چگونه است بدین طریق پاسخ داده شد. با ترکیب دو دسته از متغیرها به طور کلی ۳۰ متغیر برای بررسی الگوی رفتاری مشتریان شناسایی شد و با استفاده از شبکه عصبی کوهنن ماتریس نرونهای خروجی شبکه ۵×۳ تشخیص داد شده و مشتریان در ۱۳ خوشه دسته بندی شدند. این طبقات هر کدام با توجه به بیشترین شباهت در الگوی رفتاری دسته بندی گردیدند. از میان خوشه های ایجاد شده یکی از خوشه ها با مختصات  $(X=0, Y=2)$  خوشه با تعداد مشتری ۹۸۱ و متغیرهای  $M:0.446, F:0.5, R:0.453$  برای انجام تحلیل بیشتر و تشخیص قوانین انجمنی بین خدمات این گروه انتخاب گردید. در این چارچوب ارائه شده و تحقیق حاضر تعداد زیادی متغیر مورد استفاده گردید و این نکته از جمله ویژگی است که باعث اتکاء بیشتر به نتایج تحقیق گردیده است.

در مورد پاسخ به سوال دیگر تحقیق که بیان شده است " ارائه خدمات بانکی مناسب به مشتریان بالقوه با استفاده از تکنیک های داده کاوی چگونه می باشد؟" نتایج به دست آمده حاکی از آن است که از ۲۱ خدمت در نظر گرفته شده که شامل "اعلام موجودی، دریافت صورت حساب، پرداخت قبض، برداشت وجه، انتقال وجه، عملیات مربوط به چک" از کانالهای مختلف بانکی که شامل " دستگاه ATM، اینترنت، موبایل، تلفن و پایانه های خرید POS" می باشد، ۱۲ خدمت در گروه مورد نظر بیشترین استفاده را داشته و قوانین ایجاد شده در قسمت قبلی نشان دهنده این موضوع می باشد.

با در نظر گرفتن این مطلب که جذب مشتری جدید بسیار گران تر از حفظ مشتری موجود می باشد و از طرفی نگهداری و جلب رضایت تمام مشتریان هزینه و سرمایه گذاری نیاز دارد و گاهی مقرون به صرفه نیست. بنابراین مسئله مهم برای بانک ها این است که، کدام از مشتریان سودمندی لازم را برای نگهداری دارند و با ارائه خدمات مناسب به مشتریان بالقوه آن از هزینه های اضافی تبلیغات به صورت گسترده جلوگیری نمود که در این تحقیق به طور کامل به این موضوع پرداخته شد.

## منابع و مآخذ

۱. آر. بیل، تی. جکسون، "آشنایی با شبکه‌های عصبی، ترجمه دکتر محمود البرزی، دانشگاه صنعتی شریف، موسسه انتشارات علمی، چاپ اول (۱۳۸۰)
۲. شهرابی، جمال. نیستانی، ریحانه صوفی. " تجزیه و تحلیل سبد خرید مشتریان با استفاده از قوانین وابستگی در فروشگاه‌های زنجیره‌ای شهروند" ، دومین کنفرانس داده کاوی، دانشگاه امیرکبیر (۱۳۸۷)
۳. کانتاردزیک، مه‌مد، "داده کاوی"، مترجم؛ علیخانزاده، امیر، انتشارات علوم رایانه، تهران، (۱۳۸۵)
۴. من‌هاج، محمد باقر، مبانی شبکه‌های عصبی (هوش محاسباتی) جلد اول، تهران: مرکز نشر پرفسور، (۱۳۸۴)
5. Aggelis ,Vasilis., Christodoulakis, Dimitris ., "Association Rules and Predictive Models for e-Banking Services", In 1st Balkan Conference on Informatics, 2003
6. Berry, M. a. L., Gordon "Mastering Data Mining: The Art and Science of Customer Relationship" (1997).
7. Chapman, J. C., T. Khabaza, T. Reinartz, R. Wirth: **The CRISP-DM Process Model**, (<http://www.crisp-dm.org>), 1999.
8. Chen, M.S., Han, J., Yu, P.S. Data mining: **An overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering**, 8, 866-883. (1996).
9. Cohen, M.D, **Exploiting response models-optimizing cross-sell and up-sell opportunities in banking**, Information Systems ,volume 29. 327° 341. 2003
10. Hand, D. J. **Construction and Assessment of Classification Rules**. Chichester, UK: John Wiley and Sons. (1997).
11. Hormozi. Amir M, Giles .Stacy, **Data mining : A competitive weapon for banking and retail industries** ,Information systems management, volume 31.214-225. 2004
12. Hossein Javaheri, Sadaf., "Response Modeling in Direct Marketing, A Data mining Based Approach for Target Selection", Master Thesis, Department Industrial Engineering, Tarbiat Modares University. (2008)
13. Hsieh. Nan-Chen, **An integrated data mining and behavioral scoring model for analyzing bank customer**, Expert Systems with Applications, Volume 27. 623° 633. 2004
14. Kohonen, T., **Self-organization and associative memory**, New York: Springer, 1988
15. Liao,Sh and Pushao,Y with Wang,H and Chen,A.."The adoption

- of virtual banking: an empirical study". International Journal of Information Management, 19 (1999)
16. Min. sung-hwan, Han. Ingoo, **Detection of the customer time-variant pattern for improving recommender systems**, Expert Systems with Applications, 28, 189° 199. 2005
  17. Ngai, E.W.T, Xiu, Li, Chau, D.C.K., **"Application of data mining techniques in customer relationship management:A literature review and classification"**, Expert Systems with Applications 36 (2009) 2592° 2602.(2009)
  18. Olsan. D, Shi. Y., **"Introduction to business data mining"**, McGraw Hill International Edition, 2007
  19. Ye, Nong., **"The handbook of data Mining"**, lawrence eribaum associates, publishers, London, (2003)

