



کاوش خصوصیات مشتریان بانکداری خرد با استفاده از تکنیک‌های داده کاوی

* جمشید ناظمی
* پژمان جعفری
*** حامد هاشمی

پذیرش: ۹۱/۳/۱۱

دریافت: ۹۰/۱۰/۲۶

چکیده

با روند خصوصی سازی در صنعت بانکداری کشور و رشد بی‌سابقه رقابت در تکنولوژی‌های بانکداری، هر روزه بر اهمیت نگهداری از مشتریان کنونی و جذب مشتریان جدید افزوده می‌شود. نگهداری مشتریان یک استراتژی ارزشمند است که سودآوری بلند مدت و موفقیت سازمان‌ها را تضمین می‌کند. در این مطالعه، با استفاده از تکنیک‌های داده کاوی مدلی دو مرحله‌ای برای شناسایی ویژگی‌های گروه‌های مختلف مشتریان سپرده کوتاه مدت بانک معرفی گردیده است. در مرحله نخست، از الگوریتم K-mean جهت شناسایی گروه‌های مشتریان بر مبنای میزان سودآوری آن‌ها بهره گرفته شده است. چون در K-mean تعداد بهینه خوشه‌ها بستگی به مسئله مورد نظر دارد، بنابراین از طریق الگوریتم خوشه‌بندی Twostep تعداد بهینه خوشه‌ها معین می‌گردد. پس از تعیین گروه‌های مختلف مشتریان از لحاظ میزان سودآوری آن‌ها برای بانک، در مرحله دوم، با استفاده از الگوریتم Apriori قواعد وابستگی هر خوشه استنتاج می‌گردد. نتایج حاصل از این کاوش، به بانک‌ها کمک خواهد کرد تا درک

* استادیار، عضو هیات علمی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

** استادیار، عضو هیات علمی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

*** دانش آموخته کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

(EmailToHashemi@gmail.com)

نویسنده مسئول یا طرف مکاتبه: حامد هاشمی

بهتری از انتظارات کنونی و آتی مشتریان داشته باشند. و از این طریق تدوین استراتژی‌های بازاریابی برای جذب و نگهداری از مشتریان بر مبنای سودآوری آنها تسهیل می‌گردد.

واژگان کلیدی:

داده کاوی، مدیریت ارتباط با مشتری، تحلیل RFM، سودآوری مشتری، نگهداری از مشتری

مقدمه

امروزه رقابت بر سر نوآوری و استفاده از فناوری در کسب و کار به مرز اشباع رسیده و رونق خود را از دست داده است. استفاده از سیستم‌های پیشرفته اطلاعاتی و تولید خدمات جدید و نوآورانه نمی‌تواند مزیت رقابتی قابل توجهی را ایجاد نماید و یا حتی بقای سازمان را تضمین کند. در شرایط کنونی سازمان‌ها به این نتیجه رسیده‌اند که می‌بایست تلاش بیشتری در جهت جذب و نگهداری مشتریان صرف نمایند (Chu & et al. 2007). در این بازار رقابتی راضی نگهداشتن مشتریان امری بسیار دشوار است (Hu, 2005). چرا که آن‌ها هر روزه در معرض پیشنهادهایی برابر و یا حتی بهتر از خدمات شما از سوی رقبا می‌باشند و از این رو بسیار باهوش‌تر، پرتوقع‌تر، و حساس‌تر از گذشته شده‌اند (Kotler & Keller, 2008). در شرایط کنونی که افراد با انتخاب‌ها و پیشنهادهای خدماتی متنوعی رو به رو هستند، تصاحب مشتریان جدید فرآیندی دشوار و هزینه‌بر است (Van den Poel & Lariviere, 2004).

تغییرات در رفتار مشتری از قابل پیش‌بینی نبودن سودآوری او و عدم به کارگیری برنامه ریزی بازاریابی موثر و کارا نشأت می‌گیرد. درک و واکنش به تغییرات رفتار مشتری برای بقا در این بازار اشباع شده و رقابتی اجتناب ناپذیر است. امروزه شرکت‌ها فهمیده‌اند که پایگاه داده مشتریان کنونی سرمایه‌ای بسیار ارزشمند است (Athanassopoulos, 2000; Thomas, 2001; Jones & et al. 2000).

آن‌ها دریافته‌اند که نه تنها می‌بایست تلاش بیشتری برای متقاعد کردن مشتریان به تمدید قرارداد کنند، بلکه از مشتریان کنونی نیز باید محافظت شود. تحقیقات نشان می‌دهد که افزایش ۱٪ در نگهداری مشتریان ارزش سازمان را به طور متوسط ۵٪ ارتقا می‌دهد (Lemmens & Croux 2006).

بنابراین نگهداری مشتریان یک استراتژی ارزشمند است که سودآوری بلند مدت و موفقیت سازمان‌ها را تضمین می‌کند.

اغلب این حکم در ارتباط سازمان‌ها با مشتریانشان صادق است که، درصد زیادی از سود سازمان از درصد کمی از مشتریان به دست می‌آید (Niraj & et al. 2001). علاوه بر این، بخش عمده‌ای از پایگاه مشتریان، سودآور نیستند، از این رو پیشنهاد می‌شود که سازمان‌ها فقط بر نگهداری از مشتریانی متمرکز شوند که سهم قابل توجهی در سودآوری سازمان دارند (Ganesh & et al. 2000). همچنین برقراری یک رابطه بلند مدت با همه مشتریان امکان پذیر نیست (Hoekstra & Huizingh, 1999). در پژوهش حاضر، روشی جدید برای سنجش ارزش سودآوری مشتریان حقیقی ارائه گردیده است. که این امر باعث تفکیک مشتریان و کارکرد بهینه کمپین‌های نگهداری از مشتریان خواهد شد. تعیین ارزش مالی مشتریان، تخصیص منابع بازاریابی را برحسب سودآوری هر مشتری برای سازمان تسهیل می‌نماید. این فلسفه نه تنها مبنای مدیریت ارتباط با مشتری است

(Boulding & et al. 2005; Payne & Frow, 2005). بلکه راهی است برای تشخیص این که، چه تاکتیک‌ها و استراتژی‌های بالقوه بازاریابی‌ای می‌تواند بیشترین بازگشت سرمایه را به همراه داشته باشد، و بدین طریق تأثیر مالی این تاکتیک‌ها و استراتژی‌ها قابل اندازه‌گیری خواهند شد (Rust & et al. 2004).

در صنعت مالی برای شناسایی ارزش مشتری، بیشتر از مدل ارزش دوره عمر مشتری^۱ (CLV) به منظور مدل سازی استفاده شده است

(Helgesen 2007; van Raaij et al. 2003; Kaplan & Anderson 2007; Farajian & Mohammadi 2010).

حال آن که در سایر صنایع به ویژه صنایع تولیدی از مدل تحلیل سودآوری مشتری^۲ (CPA) بهره گرفته شده (Holm & et al. 2011). در پژوهش حاضر از مدل تحلیل سودآوری مشتری (CPA) ابداعی برای دسته‌بندی مشتریان استفاده شده که این روش ضمن برخورداری از تمامی مزایای ارزش دوره عمر مشتری (CLV) به بانک‌ها

1- Customer Lifetime Value

2- Customer Profitability Analysis

به فرد^۴، برنامه‌های وفاداری^۵ و مدیریت شکایات^۶ است. بازاریابی فرد به فرد به بازاریابی شخصی سازی شده‌ای ارجاع دارد که با تحلیل، هدایت می‌شود و می‌تواند تغییراتی که ممکن است در رفتار مشتری اتفاق بیافتد را کشف و پیش‌بینی نماید (Chen & et al. 2005).

یک مشتری ارزشمند معمولاً پویا است، برای همین رابطه با او در طول زمان متغیر می‌باشد. پس بنابراین، نقش اصلی مدیریت ارتباط با مشتری (CRM) درک این روابط می‌باشد. که این درک از طریق کنکاش درباره چرخه حیات مشتری، یا طول عمر مشتری که به مراحل مختلف رابطه با مشتری و کسب و کار ارجاع دارد، حاصل می‌گردد (Olafsson & et al. 2008).

در واقع مدیریت ارتباط با مشتری از نوع سنتی خود که بازاریابی کلی^۷ بوده است، تبدیل به رویکرد جدید بازاریابی تعاملی و یا یک به یک شده است. فرآیند مشتری مداری مستقل از نوع خود، شامل ۲ مرحله اصلی است: ابتدا پی بردن به ماهیت مشتری و سپس تصمیم‌گیری در مورد مناسب‌ترین عمل و اجرای آن. این فرآیند ۲ مرحله‌ای دارای سه جز می‌باشد که جز اول یادگیری و انطباق پذیری است، ابتدا اطلاعات پیش‌گویی را از پایگاه داده‌های بزرگ پیش‌بینی می‌کند سپس مشتریان ارزشمند را شناسایی کرده، ترجیحات آن را ثبت و رفتار آینده آن‌ها را پیش‌بینی می‌شود. جز دوم نگاه به جلو^۸ نام دارد. که در حقیقت نگاهی است به آینده فعالیت‌های ارتباط با مشتری کنونی و نتایجی که در زمینه بازاریابی در پی خواهد داشت این کار برای تصمیم‌گیری فعالیت‌های جایگزین ضروری است. جز سوم بهینه سازی می‌باشد که جزیی یکپارچه در فعالیت‌های مدیریت ارتباط با مشتری محسوب می‌شود. بهینه سازی به معنای ایجاد تعادل بین هزینه‌های کوتاه مدت و سود دراز مدتی که در پی فعالیت‌هایی ارتباط با مشتری ایجاد می‌شود، است (Sun & et al. 2006).

کمک خواهد کرد تا شناسایی دقیقی از ارزش کنونی سازمان خود داشته باشند.

تمرکز این پژوهش بر داده کاوی است، که ترکیبی از رویکرد اکتشافی و تحلیل تقویتی می‌باشد. هدف از این مطالعه، ارائه مدلی برای شناسایی قواعد و ویژگی‌های مشتریان، بر اساس درجه سودآوری آن‌ها برای بانک می‌باشد. در بخش‌های آتی، به مرور برخی از مفاهیم پایه و پژوهش‌های پیشین خواهیم پرداخت. سپس چارچوب پیشنهادی دو مرحله‌ای برای کاوش رفتار مشتری ارائه می‌گردد.

مروری بر مفاهیم اولیه

مدیریت ارتباط با مشتری

گرایش به سوی مدیریت ارتباط با مشتری^۱ (CRM) از سال ۱۹۹۰ شروع به رشد کرد. توسعه روابط با تک‌تک مشتریان می‌تواند در نهایت موجب افزایش وفاداری و نگهداری و همچنین سودآوری مشتری گردد (Ngai, 2005). داده‌های مشتری و ابزارهای تکنولوژی اطلاعات^۲ پایه و اساس موفقیت هر سیستم مدیریت ارتباط با مشتری هستند. به علاوه رشد سریع اینترنت و تکنولوژی‌های وابسته به شدت فرصت‌های بازاریابی را افزایش داده است و مسیر ارتباطی بین مشتریان و شرکت را به گونه‌ای مدیریت پذیر تبدیل کرده است (Ngai & et al. 2009). مروری بر ادبیات از سال‌های ۲۰۰۰ تا ۲۰۰۶ نشان می‌دهد که ۵۴ مقاله از ۸۷ مقاله (۶۲٪) منتشر شده در زمینه کاربرد داده کاوی در مدیریت ارتباط با مشتری، با تمرکز بر موضوع نگهداری از مشتری بوده است (Ngai & et al. 2009). نگهداری از مشتری^۳ در حقیقت قلب یک سیستم مدیریت ارتباط با مشتری است. رضایت مشتریان، در واقع مقایسه‌ای بین درک آن‌ها از ارزش کسب شده توسط محصول و انتظاراتی که داشته‌اند حاصل می‌شود (Kracklauer & et al. 2004).

پس بنابراین اجزاء نگهداری از مشتری شامل بازاریابی فرد

4- One-to-one Marketing
5- Loyalty Programs
6- Complaints Management
7- Mass Marketing
8- Forward Looking

1- Customer Relationship Management
2- Information Technology
3- Customer Retention

تحلیل سودآوری مشتری در مقایسه با ارزش دوره عمر مشتری^۱

سودآوری مشتری به صورت "تفاوت بین درآمدهای کسب شده از مشتری و هزینه‌های مرتبط با رابطه با مشتری در خلال یک دوره خاص" تعریف شده است (Pfeifer & et al. 2005).

ولی ارزش دوره عمر مشتری (CLV) به طور مفهومی به صورت "ارزش کنونی همه عایدی‌های کسب شده آتی از یک مشتری در خلال عمر ارتباطی او با سازمان" تعریف شده (Gupta & Lehmann, 2006).

بنابراین برخلاف CLV که ارزش مشتری را با تمرکز بر سود رسانی‌های آتی می‌سنجد، تحلیل سودآوری مشتری بر مبنای سودهای عایدی از مبادلات گذشته است.

پیش‌بینی رفتار مشتری، ارزش مشتری، رضایت مشتری، و وفاداری مشتری مثال‌هایی از اطلاعاتی هستند که می‌توان از داده‌های ذخیره شده در پایگاه‌های شرکت‌ها استخراج کرد (Hadden & et al. 2007). تکنیک‌های داده کاوی می‌توانند تغییرات در رفتار مشتری را کشف و پیش‌بینی کنند. این امکان فرآیند ایجاد یک رابطه بلند مدت و مستحکم را بین شرکت و مشتریانش تسهیل می‌کند. به علاوه، شناخت رفتار مشتریان می‌تواند مدیران را در برنامه‌ریزی اثربخش‌تر کمپین‌های بازاریابی مستقیم یاری دهد و خلاصه کلام این که، داده کاوی تکنولوژی بسیار مهمی برای کمک به مدیران در بازاریابی مستقیم می‌باشد (Wang & Hong, 2006; Chen & et al. 2005).

از متغیرهای RFM معمولاً برای مدل سازی ارزش دراز مدت مشتری (CLV) استفاده می‌شود. (Yada & et al. 2005)، از مزیت‌های به کارگیری این مدل، کاهش پیچیدگی تحلیل ارزش مشتری به واسطه استفاده از کمترین معیار (۳بعدی) به عنوان ویژگی‌های خوشه‌بندی است (Kaymak, 2001).

این تکنیک رفتار مشتری را تشخیص می‌دهد و ویژگی‌های رفتار مشتری را به وسیله سه متغیر ذیل به نمایش می‌گذارد:

۱- تازگی^۲: تازگی آخرین خرید، فاصله بین مدت زمانی که آخرین خرید صورت گرفته را نمایش می‌دهد.

۲- تناوب^۳: تناوب خرید، به تعداد مبادلات در یک دوره خاص ارجاع دارد.

۳- ارزش پولی^۴: ارزش پولی خرید، به حجم مبادلات مالی مالی در یک دوره خاص ارجاع دارد.

برخی از پژوهشگران برای طبقه بندی ارزش مشتری مدلی بر اساس محاسبه فاصله بین مرکز خوشه و نقطه صفر ارزش را، به منزله وفاداری بیشتر مشتری، پیشنهاد داده‌اند (Cheng & Chen, 2009).

همچنین، ها و دیگران^۵ (2002)، از RFM برای ساخت سیستم مدیریت ارتباط با مشتری پویا استفاده نموده‌اند و سیستم خود را بر اساس تحلیل این متغیرها و گروه‌بندی مشتریان بر اساس آن‌ها پیاده سازی کرده‌اند (Ha & et al. 2002).

آخرین باری که مشتری خرید کرده است تاثیر بسزایی در تشخیص رفتار او می‌گذارد و تعداد دفعاتی که او خرید کرده نیز، می‌تواند ما را در تشخیص این که آیا او قصد رفتن دارد یا خیر کمک کند، از طرفی هر چه مشتری در سازمان بیشتر خرج کند ارزش او بیشتر بوده و سرمایه گذاری برای حفظ او در سازمان مقرون به صرفه‌تر می‌باشد (Hadden & et al. 2007).

با ظهور روش هزینه‌یابی بر مبنای عملیات^۶ (ABC)، که در این روش هزینه‌های منابع را از تجمیع هزینه‌های هر فعالیت و هزینه‌های جانبی به دست می‌آورد، چارچوب جدیدی ارائه گردید که می‌توانست طیف وسیع‌تری از هزینه‌ها و سرمایه‌های هر مشتری را معین گرداند. در پژوهش‌های اخیر استفاده از تکنیک ABC برای تحلیل سودآوری مشتری رواج یافته است (Gleaves & et al. 2008; McManus & Guilding, 2008).

2- Recency
3- Frequency
4- Monetary
5- Ha & et al.
6- Activity-Based Costing

1- Customer Profitability Analysis vs. Customer Lifetime Value

مدل پیشنهادی دو مرحله‌ای تحلیل رفتار مشتری

همان طور که در شکل ۱ نشان داده شده است در مطالعه حاضر، از یک رویکرد دو مرحله‌ای برای تحلیل رفتار مشتری به منظور استخراج دانش از داده‌های مشتریان بانک و رفتارهای تراکنشی آن‌ها، استفاده شده است.

الگوریتم‌های خوشه‌بندی سعی در تفکیک مجموعه داده‌های ورودی به داخل خوشه‌ها یا زیر گروه‌های متجانس وابسته را دارند که شباهت رکوردهای داخل خوشه‌ها ماکزیمم و خارج خوشه‌ها مینیمم است (Larose, 2006). در الگوریتم K-mean هیچ قانون کلی برای تعداد بهینه خوشه‌ها وجود ندارد و تعداد خوشه‌ها بستگی به مشکل و مسئله مورد نظر دارد (Ahola & Rinta, 2001). برای مشخص نمودن تعداد خوشه‌های مطلوب به طور اتوماتیک، از روش خوشه‌بندی TwoStep استفاده می‌کنیم. این روش در دو مرحله خوشه‌بندی را انجام می‌دهد. در مرحله نخست، رکوردها به تعداد زیادی از خوشه‌ها تقسیم می‌شوند و سپس در مرحله بعد، خوشه‌های منتج شده از گام اول به تعداد بهینه شکل داده می‌شود. خروجی این الگوریتم تعداد بهینه خوشه‌ها را معین می‌کند.

پس از تعیین تعداد بهینه خوشه‌ها، از الگوریتم K-mean برای تقسیم مشتریان به گروه‌های متجانس استفاده شده است. این روش یکی از پرکاربردترین و ساده‌ترین روش‌های خوشه‌بندی است که برای تفکیک n شی در k دسته استفاده می‌شود. در این روش هر شی به دسته‌هایی اختصاص می‌یابد که فاصله میانگین آن دسته تا شی مورد نظر کم‌ترین اندازه ممکن باشد. تکنیک خوشه‌بندی K-mean علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی K-center و K-median) محسوب می‌شود. برای این الگوریتم شکل‌های مختلفی بیان شده است. ولی همه آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

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

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

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

۱- در ابتدا K نقطه به عنوان نقاط مراکز خوشه‌ها انتخاب می‌شوند.

۲- هر نمونه داده به خوشه‌ای که مرکز آن خوشه کم‌ترین فاصله تا آن داده را داراست، نسبت داده می‌شود.

۳- پس از تعلق تمام داده‌ها به یکی از خوشه‌ها، برای هر خوشه یک نقطه جدید به عنوان مرکز محاسبه می‌شود. (میانگین نقاط متعلق به هر خوشه)

۴- مراحل ۲ و ۳ تکرار می‌شوند تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود.

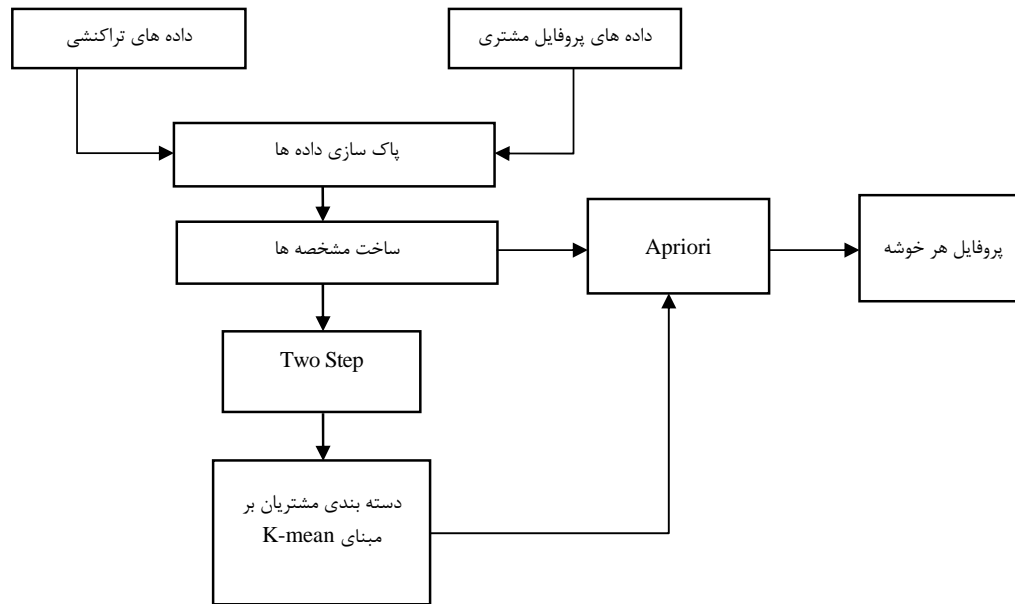
وقتی که K-mean گروه‌های سودآور مشتریان را مشخص کرد، الگوریتم Apriori برای مشخص کردن ویژگی‌های گروه‌های مشتریان به کار گرفته می‌شود. استخراج قوانین وابستگی نوعی عملیات داده کاوی می‌باشد که به جستجو برای یافتن ارتباط بین ویژگی‌ها در مجموعه داده‌ها می‌پردازد. به عبارت دیگر، تحلیل وابستگی، مطالعه ویژگی‌ها یا خصوصیتی است که با یکدیگر همراه بوده و به دنبال استخراج قواعد از میان خصوصیات می‌باشد. این روش به دنبال استخراج قواعد به منظور کمی کردن ارتباط بین دو یا چند خصوصیت است. قواعد وابستگی به شکل اگر و آن گاه به همراه دو معیار پشتیبان^۱ و اطمینان^۲ تعریف

1- Support

2- Confidence

گروه از مشتریان کمک می‌کند و می‌تواند به عنوان ابزاری برای مدیریت استراتژیک بازاریابی بانکی به کار گرفته شود.

می‌شوند. شناسایی ارتباط بین سودآوری و سایر خصوصیات یک مشتری، به ایجاد پروفایلی برای تشریح ویژگی‌های هر



شکل ۱: مدل پیشنهادی برای تحلیل رفتار مشتری

نتایج تجربی

ویژگی‌های شماره حساب، تاریخ، میزان تراکنش آن روز، تعداد دفعات تراکنش به تفکیک ترمینال در آن روز، مانده حساب در پایان آن روز است. روزهایی که مشتری هیچ تراکنشی نداشته از جدول حذف شده‌اند.

پاک سازی داده‌ها

پیش از استخراج دانش، داده‌ها می‌بایست برای انجام داده کاوی، آماده سازی شوند. گام آماده سازی داده‌ها عبارت است از: انتخاب داده‌ها، پاکسازی داده‌ها، مجتمع کردن آن‌ها و قالب بندی داده‌ها. جمع آوری و محافظت از داده‌ها گام بسیار مهمی می‌باشد. اصولاً چون قالب و نوع داده‌ها در طول زمان تغییر می‌کند، ممکن است قالب بسیاری از داده‌های موجود متفاوت باشد. همچنین به علت این که داده‌ها از منابع مختلف داخلی و خارجی جمع آوری شده و یکپارچه می‌شوند، باز هم ممکن است قالب داده‌ها با هم

در این مطالعه، از پایگاه داده حدود ۲۲۷۰۰۰ مشتری یکی از بانک‌های خصوصی ایران در بازه زمانی ۱۳۸۷/۱/۱ تا ۱۳۸۸/۱/۱ استفاده شده است. لازم به ذکر است که مشتریان حقوقی از مجموعه داده مورد مطالعه حذف شده‌اند. زیرا رفتارهای صورت گرفته از جانب مشتریان حقوقی اغلب جنبه بازخوردی و عقلایی ندارد، بلکه به سبب برخی از روابط سیاسی یا ارتباط بین مدیران سازمان و بانک می‌باشد. همچنین، از بین انواع سپرده‌های مشتریان (شامل حساب جاری، دراز مدت، و ...) تنها سپرده‌های کوتاه مدت هر مشتری مورد مطالعه قرار گرفته است. و نیز هر مشتری - سپرده، به عنوان موجودیتی منحصر بالفرد نگریسته شده و برای سنجش میزان سودآوری هر مشتری مجموع سودآوری هر مشتری - سپرده در نظر گرفته شده است. رکوردهای مربوط به جزئیات تراکنش روزانه هر مشتری ابتدا در جدولی گردآوری گردید. این جدول شامل

طور گسترده در دسترس بوده و به عنوان پیش‌بینی کنندگان توانمند و کارآمد شناخته شده‌اند

(Schmittlein & Peterson, 1994; Buckinx & et al. 2004; Buckinx & Van den Poel, 2005). در ادامه به شرح هر یک از این متغیرها خواهیم پرداخت.

متغیرهای دموگرافیکی: در این مطالعه از معیارهای گروه سنی مشتری، جنسیت مشتری، وضعیت تاهل، میزان تحصیلات، کد محل تولد، و متغیری برای نمایش اولین بانکی که مشتری در آن حساب افتتاح کرده استفاده شده است. همچنین سایر ویژگی‌های دموگرافیک چون شغل، موقعیت جغرافیایی، تعداد فرزندان، درآمد خانوار، ایمیل، تلفن تماس و غیره نیز به دلیل نقصان و کیفیت پایین داده‌های ذخیره شده قابلیت استفاده نداشتند.

فاصله زمانی بین خرید^۵: در پایگاه مشتریان مورد مطالعه، معیاری به نام تازگی که حاصل میانگین فواصل زمانی تراکنش‌های یک سپرده در یک ماه می‌باشد، به کارگرفته شده است. این ویژگی از نوع نسبی می‌باشد.

شاخص پولی^۶: این شاخص حاصل مجموع گردش ماهیانه حساب در ماه است. در واقع از این طریق رفتار تراکنشی مشتری در ماه‌های گذشته را می‌سنجیم. در این مطالعه از معیاری به نام شاخص پولی که حاصل مجموع مبلغ تراکنش‌های ماهیانه یک سپرده است، به عنوان یکی از متغیرهای کشف قواعد بهره گرفته شده است.

تناوب خرید^۷: تناوب خرید معادل است با تعداد دفعات تراکنش مشتری از حساب. در این مطالعه از معیاری به نام تناوب برای نشان دادن مجموع تعداد تراکنش‌های ماهیانه یک حساب استفاده شده است.

طول ارتباط مشتری با سازمان^۸: طول ارتباط مشتری

یکسان نبوده و یا حتی برخی از داده‌های قبلی از بین رفته و دور ریخته شده باشند و تنها بخش‌هایی از داده‌ها موجود باشد. در داده کاوی اهمیت داده‌های قدیمی به هیچ وجه کمتر از داده‌های جدید نمی‌باشد. داده‌های موجود در پایگاه داده سازمان معمولاً کامل نبوده و مقادیر گم شده در آن قابل توجه می‌باشند. کامل کردن پایگاه داده از پیش پردازش‌هایی است که می‌بایست در مرحله‌ی آماده سازی انجام داد. اگر چه در برخی مواقع رکوردهایی که دارای مقادیر گم شده هستند را می‌توان به طور کلی حذف نمود، اما در بسیاری از موارد جایگزینی صحیح موارد گم شده امری ضروری است (Neslin & et al. 2006).

در این مطالعه، رکوردهای مربوط به سن مشتریان دارای برخی اغتشاشات بود. برای مثال در برخی موارد کاربر به جای وارد کردن تاریخ تولد مشتری، سن او را درج کرده است. مواردی از این دست اصلاح گردیدند. و در برخی موارد مانند داده‌های نامفهوم سن مشتری رکورد مورد نظر در طبقه‌ی ۲۱ که برای داده‌های نامفهوم است طبقه بندی شد. داده‌های مربوط به سن افراد به ۲۱ رده که هر رده نمایانگر ۵ سال سن است طبقه بندی شدند. در مواردی مشاهده شد که وضعیت تاهل مشتریان در پایگاه درج نشده بود، در این موارد ارزشی برابر ۰ برای نمایش ناشناخته بودن درج گردید. همچنین در مورد مشخصه تحصیلات نیز مقداری برابر ۰ برای ناشناخته بودن در نظر گرفته شد.

به طور کلی، در ادبیات چهار گروه متغیر برای پیش‌بینی رفتار مشتری معرفی شده است. این متغیرها عبارتند از: متغیرهای رفتاری مشتری^۱، متغیرهای مربوط به ویژگی‌های دموگرافیکی مشتری^۲، متغیرهای مربوط به ادراک و آگاهی مشتری^۳ و متغیرهای کلان محیطی^۴. داده‌های موجود در این مطالعه شامل اطلاعات رفتاری و اطلاعات دموگرافیکی مشتریان است. در این مطالعه با استفاده از رفتار مشاهده شده مشتری در گذشته و اطلاعات دموگرافیکی او، متغیرهایی به منظور پیش‌بینی رفتار ایجاد شده است. این متغیرها به

5- Inter purchase Times
6- Monetary Indicator
7- Frequency
8- Length of Relationship

1- Customer Behavior
2- Customer Demographics
3- Customer Perceptions
4- Macro Environment Variable

دسته از مشتریان جریان پیوسته‌های از سود را برای سازمان فراهم کرده و در واقع پیوستگی سودآوری را تضمین می‌کنند. برای محاسبه این معیار از تکنیک هزینه بر مبنای فعالیت^۲ بهره گرفته شده است (Cooper & Kaplan 1988).

$$CP = \frac{\text{میزان موجودی} \times \text{فاصله زمانی بین آخرین تراکنش‌ها}}{\sqrt{\sum \text{هزینه تمام شده هر ترمینال خدمت دهنده} \times \text{کد ترمینال خدمت دهنده}}}$$

رابطه‌ی ۲: نحوه‌ی محاسبه ارزش مشتری

یکی دیگر از متغیرهایی به کار گرفته شده متغیر Rank می‌باشد. این متغیر نمایانگر خوشه‌ای است که آن مشتری به آن تعلق دارد. بنابراین مقدرهای این متغیر پس از خوشه‌بندی مشتریان بر اساس ارزش‌هایشان مشخص می‌گردد. در جدول ۱ ویژگی‌های هر یک از متغیرهای به کار گرفته شده در این پژوهش نمایش داده شده است.

جدول ۱: متغیرهای^۲ پیش‌بینی مورد استفاده در این مطالعه

| نام متغیر | نام لاتین | نوع داده | تعریف |
|---------------------|------------|----------|--|
| ماه | Month | Set | ماه‌های سال (فروردین = ۱، ...، اسفند = ۱۲) |
| تازگی | Recency | Range | میانگین فواصل زمانی تراکنش‌های یک سپرده در یک ماه می‌باشد |
| شاخص پولی | Monetary | Range | مجموع گردش ماهیانه (TurnOver) حساب در ماه است. |
| کد شعبه افتتاح حساب | Branch | Set | کد شعبه‌ای که مشتری نخستین بار افتتاح حساب کرده است. |
| کد محل تولد | Birthplace | Set | کد شهر محل تولد |
| سن | AgeGroup | Set | سن مشتری (۱، ۲، ...، ۲۱) |
| سن عضویت | EnblAge | Set | مدت زمان عضویت مشتری در بانک به سال (۱، ۲، ...، ۵) |
| جنسیت | SexCode | Flag | ۱=مرد و ۰=زن |
| وضعیت تاهل | Marital | Set | ۰=نامشخص، ۱=مجرد و ۲=متاهل |
| میزان تحصیلات | Education | Set | تحصیلات در ۱۰ سطح طبقه بندی شده است |
| شماره خوشه | RANK | Set | هر مشتری به یکی از ۳ خوشه‌ی طلایی، نقره‌ای، برنزی تعلق دارد. |
| تناوب | Freq | Range | تعداد دفعات تراکنش در ماه |
| کد ترمینال | Type | Set | ۱۱=اینترنتی، ۱۲=ATM، ۱۳=موبایل بانک، ۱۴=تلفن بانک، ۱۵=حضور |

2- Activity-Based Costing

3- Attributes

1- Customer Profitability

طلایی، نقره‌ای، برنزی بر اساس میزان سودآوری آن‌ها در واحد زمان دسته‌بندی می‌کنند. که در این طبقه بندی مشتریان طلایی میزان سودآوری بیشتری نسبت به مشتریان نقره‌ای و برنزی به همراه دارند (Verhoef & Langerak 2002).

در گام بعد مجموعه داده را با استفاده از الگوریتم K-mean خوشه بندی می‌نماییم. نتایج حاصل از این مرحله در جدول ۲ نمایش داده شده است.

پس از پیش پردازش و ساخت متغیرهای مورد نیاز، همان طور که گفته شد با استفاده از الگوریتم Twostep تعداد بهینه خوشه‌ها را تعیین می‌کنیم. در این مطالعه الگوریتم ۳ خوشه متفاوت را تعیین نمود. که از این پس با نام‌های مشتریان طلایی، نقره‌ای، و برنزی آن‌ها را می‌شناسیم. در ادبیات موضوع خوشه‌های طلایی، نقره‌ای، و برنزی اصطلاحی رایج و شناخته شده است. برخی از شرکت‌ها مشتریان را به گروه‌های مختلفی با نام‌های مشتریان

جدول ۲: نتایج خوشه‌بندی و نحوه‌ی توزیع داده‌ها در خوشه‌های مختلف

| شماره خوشه | اسم | میانگین CP | انحراف معیار | درصد | تعداد |
|------------|---------|------------|--------------|-------|--------|
| خوشه ۱ | برنزی | ۶/۴ E۹ | ۱.۸ E۱۰ | ۹۹/۲۹ | ۲۲۵۰۲۶ |
| خوشه ۲ | نقره‌ای | ۲/۷ E۱۱ | ۲.۱ E۱۱ | ۰/۶۸ | ۱۵۴۷ |
| خوشه ۳ | طلایی | ۱/۱ E۱۳ | ۱.۱ E۱۲ | ۰/۰۲ | ۵۵ |

مشتری محاسبه می‌شود و سپس با استفاده از آن به عنوان متغیر هدف کشف وابستگی انجام می‌گیرد یعنی خروجی فاز اول ورودی فاز ثانویه است. بنابراین در گام بعد می‌بایست با استفاده از الگوریتم Apriori قواعد وابستگی معین گردید. در این الگوریتم متغیر شماره خوشه به عنوان متغیر هدف تعیین گردید و وابستگی سایر متغیرها با این متغیر سنجیده شد. از آن جایی که تعداد مشتریان طلایی و نقره‌ای محدود هستند، قواعد منتج شده برای این دو خوشه بسیار کم می‌باشد و این روش بیشتر به شناخت مشتریان برنزی کمک می‌کند. برخی از قواعد حاصل در جدول ۳ نمایش داده شده است.

نتایج حاصل از خوشه‌بندی نشان می‌دهد تعداد زیادی از مشتریان به خوشه برنزی‌ها تعلق دارند. این موضوع به این دلیل است که تعداد زیادی از مشتریان دارای ویژگی‌های ترانکشی مشابهی هستند. همان طور که توضیح داده شد، دلیل انتخاب نام برنزی برای خوشه ۱ این است که این مشتریان چون همیشه باقی مانده حساب کوچکی دارند منفعت زیادی به بانک نمی‌رسانند. اگر سودآوری کسب شده از خوشه طلایی را محاسبه کنیم، مشخص می‌شود ۲۳٪ از سودآوری بانک از ۵۵ مشتری بوده است. در ابتدا هیچ تقسیم بندی بین مشتریان وجود ندارد، پس از انجام خوشه‌بندی متغیر شماره خوشه (Rank) برای هر

جدول ۳: قواعد به دست آمده از داده‌های یکی از بانک‌های خصوصی

| اطمینان | پشتیبان | مقدم | تالی |
|---------|---------|--|----------|
| ٪ ۹۴/۸۸ | ٪ ۱۵/۶۳ | AgeGroup = ۶ and SexCode = ۱ | RANK = ۱ |
| ٪ ۹۴/۸۶ | ٪ ۱۰/۱۹ | AgeGroup = ۶ and SexCode = ۱ and Monetary < ۷۹۰۳۶ and Monetary > ۲۵۱۳ | RANK = ۱ |
| ٪ ۸۷/۸ | ٪ ۱۴/۸۲ | AgeGroup = ۶ and Martial = ۱ | RANK = ۱ |
| ٪ ۷۳/۲۷ | ٪ ۵/۴۷ | Branch = ۲۲۵ and Recency = ۹ and SexCode = ۱ | RANK = ۱ |
| ٪ ۸۹ | ٪ ۱۱/۸ | Recency = ۱۳ and Birthplace = ۱۰۴ and SexCode = ۱ | RANK = ۱ |
| ٪ ۸۷ | ٪ ۶/۳ | Freq = ۵ and Type = ۱۵ | RANK = ۱ |
| ٪ ۸۳ | ٪ ۰/۰۱ | AgeGroup = ۱۰ and SexCode = ۱ and Education = ۶ | RANK = ۳ |
| ٪ ۷۶/۲۹ | ٪ ۰/۰۲ | AgeGroup = ۱۱ and SexCode = ۱ and Recency = ۴/۲ and Freq > ۶۴ | RANK = ۲ |
| ٪ ۸۵ | ٪ ۲/۹ | EnblAge = ۲ and SexCode = ۱ | RANK = ۱ |
| ٪ ۷۵/۱ | ٪ ۱/۵ | Freq < ۹ and Type = ۵ | RANK = ۱ |
| ٪ ۷۷/۳ | ٪ ۶/۳ | Birthplace = ۱۰۴ and EnblAge = ۶ | RANK = ۱ |
| ٪ ۷۳/۵ | ٪ ۰/۰۳ | Recency < ۶/۳ and AgeGroup = ۱ and Freq > ۴۲ | RANK = ۲ |
| ٪ ۸۳ | ٪ ۲۵/۶ | Branch = ۲۲۵ and SexCode = ۱ and Monetary > ۳۱۸۷ and Freq < ۸ | RANK = ۱ |
| ٪ ۸۱/۱ | ٪ ۴/۶ | Birthplace = ۱۰۴ and EnblAge = ۳ | RANK = ۱ |
| ٪ ۹۳/۶ | ٪ ۴/۷ | EnblAge < ۳ and SexCode = ۱ | RANK = ۱ |
| ٪ ۸۷/۹ | ٪ ۱۸/۱ | Freq < ۸ and AgeGroup < ۹ and Recency > ۹/۵ | RANK = ۱ |
| ٪ ۷۴ | ٪ ۱۳/۷ | Recency = ۱۴/۵ and Freq > ۷ | RANK = ۱ |

نتیجه‌گیری

یکی از مهم‌ترین مولفه‌های بقا در بازارهای پویا درک و انطباق با تغییرات رخ داده در رفتار مشتریان است. در مطالعه حاضر مدلی دو مرحله‌ای جدیدی برای شناسایی مشخصات مشتریان بر مبنای ویژگی‌های مشترک بین اعضا هر خوشه ارائه گردید. این مدل با استفاده از داده‌های واقعی یکی از بانک‌های خصوصی ایران مورد مطالعه

تجربی قرار گرفت. در مدل دو مرحله‌ای ارائه شده، با استفاده از روش هزینه بر مبنای فعالیت^۱، برای نخستین بار معیار جدیدی برای سنجش سودآوری مشتریان به منظور خوشه‌بندی آنها در صنایع مالی معرفی گردید. در ادبیات بیشتر از این روش برای سنجش سودآوری در صنایع تولیدی بهره گرفته شده است. در واقع این معیار

1- Activity-Based Costing

رویکردی ترکیبی از مدل RFM و روش هزینه بر مبنای فعالیت می‌باشد. همچنین نتایج تجربی به دست آمده، نشان می‌دهد که تعداد اندکی از مشتریان (۵۵ نفر) سود فراوانی را به همراه دارند (حدود ۲۳٪). برای محاسبه این موضوع می‌بایست سود کسب شده از خوشه مشتریان طلایی را بر سود کسب شده از همه مشتریان تقسیم کنیم. به علاوه نتایج تجربی حاکی از این است که ویژگی‌هایی همچون تناوب تراکنش، تازگی بهتر می‌تواند نمایانگر سودآوری آتی مشتری باشد.

از محدودیت‌های این مطالعه منحصر بودن نتایج حاصل به صنعت بانکداری خرد و به طور خاص سپرده‌های کوتاه مدت است. تعمیم مدل موجود به سایر انواع سپرده، یکی از فرصت‌های مطالعاتی آتی است. در این مطالعه فقط از داده‌های رفتاری و دموگرافیکی مشتریان برای شناسایی عوامل موثر بهره گرفته شده است. توسعه این مشخصه‌ها و استفاده از مشخصه ادراک مشتری می‌تواند عملکرد مدل حاضر را افزایش دهد. در این مطالعه داده‌های موجود در سال ۸۷ برای تعیین گروه هدف مورد استفاده قرار گرفته است. تاثیر طول این بازه زمانی بر نتایج حاصل نامشخص است. در صورتی که داده‌های بیشتری در دسترس قرار گیرد، امکان بررسی تاثیر طول بازه زمانی مشاهدات بر نتایج نیز فراهم خواهد شد. در این مطالعه فقط به ارائه مدلی برای شناسایی ویژگی‌های مشتریان بانکداری خرد به تفکیک سودآوری پرداخته شد. شناخت این ویژگی‌ها، نقطه آغازین فرآیند مدیریت کمپین نگهداری و جذب مشتریان است. بر همین اساس بانک‌ها با بهره‌گیری از این مدل می‌توانند مشتریان با ارزش خود را شناسایی نموده. و سپس حرکت مشتریان را از خوشه‌ای به خوشه دیگر در طی ماه‌ها رصد نمایند، تا از ریزش آن‌ها جلوگیری کنند. به علاوه با شناخت ویژگی‌های هر گروه از مشتریان می‌توان ارزش مشتریان را در فاز جذب پیش بینی نمود. این نتایج می‌تواند به واحد بازاریابی جهت ارائه برنامه‌های بازاریابی برای هر خوشه کمک کند.

منابع و مأخذ

- Ahola, J. & Rinta-Runsala. (2001). "Data mining case studies in customer profiling", Research report TTE1-2001-29, VTT Information Technology.
- Athanassopoulos, A. D. (2000). "Customer satisfaction cues to support market segmentation and explain switching behavior", *Journal of Business Research*, 47 (3), pp. 191-207.
- Boulding, W., Staelin, R., Ehret, M. & Johnston, W. J. (2005). "A customer relationship management roadmap: What is known, potential pitfalls, and where to go", *Journal of Marketing*, 69 (4), pp. 155-166.
- Buckinx, W. & Van den Poel, D. (2005). "Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research*, 164 (1), pp. 252-268.
- Buckinx, W., Moons, E., Van den Poel, D. & Wets, G. (2004). "Customer-adapted coupon targeting using feature selection", *Expert Systems with Applications*, 26 (4), pp. 509-518.
- Chen, M. C, Chiu, A. L. & Chang, H. H. (2005). "Mining changes in customer behavior in retail marketing", *Expert Systems with Applications*, 28 (4), pp. 773-781.
- Cheng, C. H. & Chen, Y. S. (2009). "Classifying the segmentation of customer value via RFM model and RS theory", *Expert Systems with Applications*, 36 (3), pp. 4176-4184.
- Chu, B. H, Tsai, M. S. & Ho, C. S. (2007). "Toward a hybrid data mining model for customer retention", *Knowledge-Based Systems*, 20 (8), pp. 703-718.
- Farajian, M. A. & Mohammadi, S. (2010). "Mining the Banking Customer Behavior Using Clustering and Association Rules Methods", *International Journal of Industrial Engineering*, 21 (4), pp. 239-245.
- Ganesh, J., Arnold, M. J. & Reynolds, K. E. (2000). "Understanding the customer base of service providers: an examination of the differences between switchers and stayers", *Journal of Marketing*, 64 (3), pp. 65-87.
- Gleaves, R., Burton, J., Kitshoff, J., Bates, K. & Whittington, M. (2008). "Accounting is from Mars, marketing is from Venus: establishing common ground for the concept of customer profitability", *Journal of Marketing Management*, 24 (7-8), pp. 825-845.
- Gupta, S. G. & Lehmann, D. R. (2006). "Customer lifetime value and firm valuation", *Journal of Relationship Marketing*, 5 (2-3), pp. 87-110.
- Ha, S. H, Bae, S. M. & Park, S. C. (2002). "Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case", *Computers and Industrial Engineering*, 43 (4), pp. 801-820.
- Hadden, J., Tiwari, A., Roy, A. R. & Ruta, D. (2007). "Computer assisted customer churn management: State-of-the-art and future trends", *Computers and Operations Research*, 34(10), pp. 2902-2917.

- Helgesen, O. (2007). "Customer accounting and customer profitability analysis for the order handling industry--A managerial accounting approach", *Industrial marketing management*, 36 (6), pp. 757-769.
- Hoekstra, J. C, & Huizingh, E. K. R. E. (1999). "The lifetime value concept in customer-based marketing", *Journal of Market-Focused Management*, 3 (3), pp. 257-274.
- Holm, M., Kumar, V. & Rohde, C. (2011). "Measuring customer profitability in complex environments: an interdisciplinary contingency framework", *Journal of the Academy of Marketing Science*, pp. 1-15.
- Hu, X. (2005). "A data mining approach for retailing bank customer attrition analysis", *Applied Intelligence*, 22 (1), pp. 47-60.
- Jones, M. A., Mothersbaugh, D. L. & Beatty, S. E. (2000). "Switching barriers and repurchase intentions in services", *Journal of Retailing*, 76 (2), pp. 259-274.
- Kaplan, R. S. & Anderson, S. R. (2007). "Time-driven activity-based costing: a simpler and more powerful path to higher profits", *Harvard Business Press*.
- Kaymak, U. (2001). "Fuzzy target selection using RFM variables", *IFSA World Congress and 20th NAFIPS International Conference, Joint 9th, Vol. 2*, pp.1038-1043.
- Kotler, P. & Keller, K. (2008). "Marketing management: Analysis, planning, and control, prentice-Hall, New Jersey."
- Kracklauer, A. H., Mills, D. Q. & Seifert, D. (2004). "Customer management as the origin of collaborative customer relationship management", *Collaborative customer relationship management-taking CRM to the next level 3-6*.
- Larose, D. (2006). "Data mining methods and models", *Hoboken NJ: Wiley-Inter science*.
- Lemmens, A. & Croux, C. (2006). "Bagging and boosting classification trees to predict churn." *Journal of Marketing Research*, 43 (2), pp. 276-286.
- McManus, L. & Guilding, C. (2008). "Exploring the potential of customer accounting: a synthesis of the accounting and marketing literatures", *Journal of Marketing Management*, 24 (7-8), pp. 771-795.
- Neslin, S. A. & Gupta, S. (2006). "Defection detection: Improving predictive accuracy of customer churn models", *Journal of Marketing Research*, 43 (2), pp. 204-211.
- Ngai, E. W. T. (2005). "Customer relationship management research (1992-2002): An academic literature review and classification." *Marketing Intelligence and Planning*, 23 (6), pp.582-605.
- Ngai, E. W.T, Xiu, L. & Chau, D. C. K. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, 36 (2), pp. 2592-2602.
- Niraj, R., Gupta, M. & Narasimhan, C. (2001). "Customer profitability in a supply chain", *The Journal of Marketing*, 65 (3), pp. 1-16.

- Olafsson, S., Li, X. & Wu, S. (2008). "Operations research and data mining", *European Journal of Operational Research*, 187 (3), pp. 1429-1448.
- Payne, A. & Frow, P. (2005). "A strategic framework for customer relationship management", *Journal of Marketing*, 69 (4), pp. 167-176.
- Pfeifer, P. E., Haskins, M. E. & Conroy, R. M. (2005). "Customer lifetime value, customer profitability and the treatment of acquisition spending", *Journal of Managerial Issues*, 17 (1), pp. 11-25.
- Van den Poel, D. & Lariviere, B. (2004). "Customer attrition analysis for financial services using proportional hazard models", *European Journal of Operational Research*, 157 (1), pp. 196-217.
- Van Raaij, E. M., Vernooij, M. J. A. & Van Triest, S. (2003). "The implementation of customer profitability analysis: A case study", *Industrial Marketing Management*, 32 (7), pp. 573-583.
- Rust, R. T., Ambler, T., Carpenter, G. S., Kumar, V. & Srivastava, R. K. (2004). "Measuring marketing productivity: Current knowledge and future directions", *Journal of marketing*, 68 (4), pp. 76-89.
- Schmittlein, D. C. & Peterson, R. A. (1994). "Customer base analysis: An industrial purchase process application", *Marketing Science*, 13 (1), pp. 41-67.
- Sun, B., Li, S. & Zhou, C. (2006). "'Adaptive' learning and 'proactive' customer relationship management", *Journal of Interactive Marketing*, 20 (34), pp. 82-96.
- Thomas, J. S. (2001). "A methodology for linking customer acquisition to customer retention", *Journal of Marketing Research*, 38 (2), pp. 262-268.
- Verhoef, P. C. & Langerak, F. (2002). "Eleven misconceptions about customer relationship management", *Business Strategy Review*, 13 (4), pp. 70-76.
- Wang, H. F. & Hong, W. K. (2006). "Managing customer profitability in a competitive market by continuous data mining", *Industrial Marketing Management*, 35 (6), pp. 715-723.
- Yada, K. (2005). "Data mining oriented CRM systems based on MUSASHI: C-MUSASHI", *Active Mining*, 103 (304), pp. 259-276.

Mining the retail banking customers characteristics using data mining techniques

J. Nazemi, Ph.D.

P. Jafari, Ph.D.

* H. Hashemi

Abstract

Deregulation within the banking industry's and unprecedented growth competition in new technologies, every day the importance of keeping current customers and attract new customers are added. This study presents a two-step model to identify characteristics of different groups of bank customers, based on their profitability. The new criteria introduced to analyze the profitability of each customer. And then, this criterion has been used for clustering customers based on their profitability. Because in k-mean algorithm there is not a general rule for the optimal number of clusters and the number of clusters depend on the problem, therefore firstly with applied by Two-step clustering algorithm determine the optimal number of clusters. In this study, Customers clustering to 3 groups as golden, silver, and bronze customers. Then, by using K-mean algorithm different groups of customers are identified. And with help of Apriori algorithm association rules of each cluster is inference. The result of this exploration, help to banks have better understanding of current and future customer expectations. And through this may be facilitated develop of marketing strategies to attract and retain profitable customers.

Keywords:

Clustering, Customer Profitability Analysis, Data Mining, Customer Relationship Management, RFM Analysis, K-mean Algorithm, Apriori Algorithm, Customer Retention, Customer Lifetime Value

* **Corresponding Author:** EmailToHashemi@gmail.com