مهندسی صنایع و مدیریت شریف، (تابستان ۱۳۹۳)
دوره‌ی ۱ - ۳۰، شماره‌ی ۱/۲، ص. ۸۱-۸۸

Original Article

به‌کارگیری هوش تجاری و مدیریت ارتباط با مشتری در شهرداری الکترونیک (مطالعه‌ی موردی سازمان فنّاوری اطلاعات و ارتباطات شهرداری اصفهان، داده‌های بخش عوارض و نوسازی)

شهیده ناصری اصفهانی (دانشجوی کارشناسی ارشد)
هدی صفائی‌پور (دانشجوی کارشناسی ارشد)
محمدجعفر تارخ* (دانشیار)
دانشکده‌ی مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی

امروزه با رشد سیستم‌های اطلاعاتی، سازمان‌ها به‌منظور هوشمندی هرچه بیشتر کسب وکار خود، به استفاده‌ی بهینه از اطلاعات و تبدیل آن به دانش سودمند روی آورده‌اند. شهرداری‌ها نیز به‌دلیل بهره‌مندی از حجم زیادی از داده‌ها و پیچیدگی ارتباط‌شان با شهروندان، بستر مناسبی برای تحقیقات هوش تجاری به حساب می‌آید. در این تحقیق با استفاده از ابزارهای کاوش داده (داده‌کاوی)، برای فهم الگوهای پنهانِ رفتاریِ شهروندان در حوزه‌ی عوارض و نوسازی و عوامل مؤثر در خوش‌حسابی و بدحسابی آنان تلاش شده است. بدین ترتیب گام‌های اولیه برای رسیدن به هوشمندی کسب‌وکار در یک مطالعه‌ی موردی، مورد بررسی قرار گرفته و با استفاده از الگوریتم k-means، با در نظرگرفتن ویژگی‌های متعدد ملک و مالک، مشتری‌های سازمان عوارض و نوسازی در پنج سطح از خوش‌حسابی طبقه‌بندی شده‌اند. بدین‌وسیله مؤثرترین عوامل خوش‌حسابی برای سرمایه‌گذاری‌های آتی، شناسایی و امکان پیش‌بینی رفتار شهروندان در این زمینه فراهم شده است.

واژگان کلیدی: شهرداری الکترونیک، هوش تجاری، مدیریت ارتباط با مشتری، داده‌کاوی.

nasery_isfahani@yahoo.com
h_safaiepour@yahoo.com
mjtarokh@kntu.ac.ir

۱. مقدمه

«هوش تجاری (BI)[۱]» به تکنیک‌های مبتنی بر رایانه برای کشف، آشکارسازی و تحلیل داده‌های کسب‌وکار اطلاق می‌شود و هدف آن، پشتیبانی از تصمیم‌گیری‌های بهتر کسب‌وکار است. بنابراین سیستم هوش تجاری را می‌توان یک سیستم پشتیبان تصمیم‌گیری[۲] دانست. هوش تجاری روشی استراتژیک برای هدف‌گذاری، ردیابی، ایجاد ارتباطات به‌صورت راهبردی است که طی آن علائم ضعیف اما مرتبط، به اطلاعاتِ عملیاتیِ مبتنی بر آن تصمیمات راهبردی تبدیل می‌شود.[۱]

امروزه مشتری‌محوری از موضوعات مهم مورد بحث است و سیستم‌های هوش تجاری و مدیریت ارتباط با مشتری[۳] در ایجاد و توسعه‌ی مزیت رقابتی نقش اساسی دارند. استفاده از فنّاوری اطلاعات (IT) نیز راه‌های جدیدی برای سازمان‌ها به‌منظور بهره‌برداری از پتانسیل عظیم ارتباطات با مشتری که قبلاً مورد توجه نبوده، ایجاد کرده‌اند. با رشد رقابت‌ها از حالت سنتی به‌صورت برخط، راضی نگه داشتن مشتریان و افزایش وفاداری آنها از استراتژی‌های مهم برای موفقیت کاری محسوب می‌شوند. برای پیشرفت ارتباطات مشتری، ابزارهای هوش تجاری با تمرکز روی فرایندهایی مثل تحقیقات بازاریابی، فرایند تصمیم‌گیری، سرویس‌های مشتری و مشارکت مشتری در سرویس‌ها و محصولات، به‌کمک مدیریت ارتباط با مشتری آمده‌اند.[۲]

در شهرداری‌ها نیز در دهه‌ی اخیر، داشتن زیرساخت‌های توانمند فنّاوری اطلاعات ــ ازجمله سیستم‌های اطلاعاتی ــ از چالش‌های موجود است و چنان‌که شاهدیم، طی دوره‌های مختلف مدیران به دنبال کسب آگاهی مناسب درخصوص عملیات موجود در سیستم خود هستند.

سیر تحول سیستم‌های اطلاعاتی در شهرداری، با تولد سیستم‌های درآمد، نوسازی، حسابداری، حقوق و دستمزد در دهه‌ی هفتاد در کشور شکل گرفت و با ظهور سیستم‌های یکپارچه شهرسازی و مالی اداری به اوج خود رسید و تا شکل‌گیری سیستم‌های هوش تجاری در دوره‌ی کنونی پیش رفت، به‌شکلی که امروزه به‌ندرت شهرداریِ فاقد ابتدایی‌ترین سیستم‌های اطلاعاتی را می‌بینیم. ولی مهم‌ترین ویژگی دوره‌ی کنونی، تمایل مدیران و سیاست‌گذاران شهرداری‌ها به کنترل و هدایت لحظه به

---

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۰/۱/۲۲، اصلاحیه ۱۳۹۰/۷/۲۴، پذیرش ۱۳۹۰/۹/۵.

لحظه و صحیح فرایندهای سیستم خود و تولد نرم‌افزارها و سخت‌افزارهای مناسب این کار است.[۳]

اگرچه سیستم‌های عملیاتی همچون نوسازی، عوامل حیاتی برای شهرداری هستند متأسفانه این سیستم‌ها، علی‌رغم این که تراکنش‌ها و تعاملات داده‌ها را در پیشخوان‌های شهرداری به‌خوبی اجرا می‌کنند، در برابر تحلیل داده‌ها با مشکلات عدیده‌یی مواجه می‌شوند که عمدتاً ناشی از عدم به کارگیری تکنولوژی مناسب در این زمینه است.

به‌همین دلیل، در این نوشتار ضمن مطالعه‌ی موردی داده‌های سازمان فنّاوری اطلاعات و ارتباطات شهرداری اصفهان در بخش عوارض و نوسازی، برای به کارگیری تکنیک‌های داده‌کاوی در هوش تجاری به‌منظور بهبود رابطه‌ی شهروندان با شهرداری در این بخش و افزایش سودآوری این حوزه شده است.

در بخش اول این تحقیق، جایگاه هوش تجاری در شهرداری‌ها و گام‌های پیاده‌سازی آن مورد بررسی می‌شود. در بخش دوم داده‌های مربوط به مطالعه‌ی موردی به تحلیل، و مراحل ذکر شده در بخش قبل انجام گام به گام اجرا می‌شود. و نهایتاً در قسمت پایانی، نتایج حاصل از تحلیل و پیاده‌سازی هوش تجاری برای داده‌های عوارض و نوسازی جمع‌بندی، و زمینه‌ی تحقیقات آینده معرفی می‌شود.

۲. هوش تجاری در شهرداری

هوش تجاری ابزاری است که سازمان برای جمع‌آوری، مدیریت و تحلیل داده‌های ساخت یافته و غیر ساخت یافته، از طریق بهره‌مندی از ابزارهای به‌روز فنّاوری اطلاعات، استفاده می‌کند. هوش تجاری از قسمت‌های اصلی داده‌های جمع‌آوری شده در سازمان بهره می‌گیرد و داده‌ها را به اطلاعات و دانش تبدیل می‌کند تا مانع فرضیات غلط و ناآگاهانه‌ی سازمان در تصمیماتش شود.[۴و۵]

هوش تجاری یعنی در اختیار قراردادن اطلاعات مناسب به افراد مناسب در زمان مناسب برای اخذ تصمیم مناسب؛ درواقع نیاز به هوش تجاری در سازمان هنگامی احساس می‌شود که:

- سازمان نیازهای اطلاعاتی ضروری و حساس خود را تشخیص نمی‌دهد یا نمی‌شناسد.
- اطلاعات و داده‌هایی که از برخی منابع بیرونی می‌رسند به‌طور بهینه جمع‌آوری نشده‌اند.
- اطلاعات و دانش پرسنل سازمان به‌صورت بهینه مورد استفاده قرار نمی‌گیرد.
- حجم اطلاعاتی که باید ذخیره، دسته‌بندی، پردازش و تحلیل شوند خیلی زیاد است.
- از اطلاعات و داده‌های موجود در سازمان درست استفاده نمی‌شود.

اینک با رشد صنعت فنّاوری اطلاعات و کسب تجربیات گوناگون در حوزه‌ی جمع‌آوری، ذخیره‌سازی و بازیابی اطلاعات، موضوع «معنا بخشیدن به داده‌ها» و تسهیل فرایند تصمیم‌سازی، هم‌زمان در مرکز توجه کارشناسان فنّاوری اطلاعات و متخصصان علم مدیریت و کسب و کار قرار گرفته است. از راه حل «پایگاه داده‌ی تحلیلی و هوش تجاری» می‌توان برای دست‌یابی به این هدف کمک گرفت. به‌عبارت دیگر هوش تجاری را می‌توان همانند یک پالایشگاه داده تصور کرد که گردش داده در آن موجب ارزش افزوده برای شهرداری خواهد شد.[۶]

۱.۲. هوش تجاری و مدیریت ارتباط با مشتری

سازمان‌ها همزمان با توسعه‌ی ارتباطات‌شان با مشتریان از ابزارهای هوش تجاری برای تحلیل بهتر این رابطه استفاده می‌کنند. هوش تجاری ضمن تسهیل یکپارچگی، تحلیل و همچنین دسترسی به حجم زیادی از داده‌ها، به فرایند تصمیم‌گیری کسب و کار کمک می‌کند. ابزارهای اصلی هوش تجاری عبارت‌اند از: پردازش تحلیلی به‌هنگام (OLAP)[۴] و داده‌کاوی (DM)[۵]. به‌کمک ابزار پردازش تحلیلی به‌هنگام با پشتیبانی از تحلیل چندبعدی، می‌توان داده‌ها را در انبارهای داده‌یی عظیم با ابعاد متفاوتی که در پایگاه داده‌های معمولی به‌راحتی قابل دست‌یابی نیست، مشاهده کرد. به کمک فنّاوریِ داده‌کاوی نیز جست‌وجو در حجم زیادی از داده‌ها به‌منظور کشف الگوهای معنادار از رفتار مشتری (مانند خوش‌حسابی و بدحسابی)، و در نهایت هوشمندی کسب و کار برای تولید فرصت‌های جدید ممکن می‌شود.[۲و۱۶]

مدل‌های پیشین تقسیم‌بندی مشتریان، مبتنی بر ویژگی‌های آماری مشتریان بود.[۷] این مدل‌ها منجر به نتایج بسیار ساده و ضعیفی برای محیط کسب و کار پیچیده‌ی فعلی می‌شوند. به‌همین دلیل روش‌های جدیدی توسعه یافته‌اند که مبتنی بر داده‌های رفتاری و تراکنشی مشتریان هستند؛ برای مثال داده‌های مرتبط با زمان مراجعه‌ی مشتری، تعداد دفعات مراجعه، تاریخچه‌ی رفتاری وی و موارد دیگر.

در محیط کسب و کار رقابتی فعلی، توانایی شناخت مشتریان سودآور، ایجاد وفاداری بلندمدت در ارتباط با آنها و توسعه‌ی مستمر ارتباطات و تعاملات فعلی، فاکتورهای رقابتی کلیدی برای سازمان هستند. برای کسب این فاکتورها، شرکت‌ها، در حوزه‌های متنوعی از صنعت، مدیریت ارتباط با مشتری را به‌عنوان یک استراتژی پیشرو در نظرگرفته‌اند. درواقع یک راهکار پایه برای مدیریت ارتباط با مشتری، متمرکز شدن روی منابع محدود و فعالیت‌های بازاریابی روی مشتریان با بیشترین ارزش است.[۱۴] در تحقیق پیرس و رگرس[۱۵] مشتریان براساس ارزش‌شان برای سازمان در سه دسته: مشتریان با بیشترین ارزش[۶]، مشتریان با بیشترین رشد[۷] و مشتریان زیر صفر[۸] (از لحاظ ارزش) معرفی می‌شوند که می‌توان با استفاده از مدیریت ارتباط با مشتری، و شناسایی مشتریان دسته اول ارتباط عمیقی با آنها برقرارکرد. این ارتباط عمیق و قوی در نهایت منجر به افزایش وفاداری و سودآوری مشتریان می‌شود.

در تحقیق ان‌گای،[۱۶] یک دوره ادبیات تحقیق روی کاربرد تکنیک‌های داده‌کاوی در مدیریت ارتباط با مشتری انجام شده که در آن حدود یکصد مقاله بر مبنای ابعاد و عناصر CRM و تکنیک‌های داده‌کاوی دسته‌بندی شده‌اند. براساس تحقیق مذکور، تعداد زیادی از مطالعات انجام‌شده در این زمینه با استفاده از خوشه‌بندی بوده است که نتایج آن در جدول ۱ قابل مشاهده است.

۲.۲. گام‌های پیاده‌سازی هوش تجاری در شهرداری

برای پیاده‌سازی هوش تجاری در شهرداری پنج مرحله باید طی کرد که در ادامه به ترتیب به این مراحل اشاره شده است.

۱.۲.۲. شناسایی منابع داده

منابع اطلاعاتی بسیاری در شهرداری‌ها وجود دارد. گام نخست شناسایی منابع داده‌یی در شهرداری‌هاست. سوابق املاک، اصناف، پروژه‌ها، طرح‌ها، قراردادها، پژوهش‌ها، اسناد مالی، دستورالعمل‌ها و نقشه‌های طرح‌های تفصیلی و اجرایی، کتابچه‌های ضوابط و مقررات و بخشنامه‌ها بخشی از داده‌های کاغذی در شهرداری‌ها هستند. همچنین داده‌های موجود در سیستم‌های عملیاتی که به‌صورت مکانیزه در حال سرویس‌دهی به شهروندان هستند نیز منابع بسیار مهم داده‌یی در شهرداری هستند.

جدول ۱. دوره‌ی ادبیات تحقیق در زمینه‌ی به‌کارگیری تکنیک‌های داده‌کاوی برای CRM با استفاده از خوشه‌بندی.[۱۶]

ابعاد CRM	عناصر CRM	تکنیک‌های خوشه‌بندی داده	مراجع
مشخص کردن مشتری	آنالیز مشتری هدف	Self-organizing map	[۲۲]
	بخش بندی	k-means	[۱۷]
		Data envelopment analysis, self organizing map & decision tree	[۱۸]
		Pattern based cluster	[۱۹]
		Self-organizing map	[۲۰و۲۱]
جذب مشتری	بازاریابی مستقیم	Outlier detection	[۲۳]
نگهداری مشتری	مدیریت شکایات	Self-organizing map	[۲۴]
	برنامه‌ی وفاداری	Attribute oriented induction	[۲۵]
	بازاریابی نفر به نفر	Association rules	[۲۶]
		K-nearest neighbor	[۲۷]
		Neural network and genetic algorithm	[۲۸]
		Self-organizing map	[۲۹]
		Association rules	[۳۰]
		Association rules and online	[۳۱]
		Analytical mining neural network	[۳۲]
توسعه مشتری	ارزش دوره عمر	Neural network	[۳۳]
	ارتباط مشتری	Survival analysis	[۳۴]

این منابع در اولین گام باید به‌خوبی شناسایی و مورد بررسی قرار بگیرد و منابع مورد نیاز برای پیاده‌سازی هوش تجاری به‌دقت انتخاب شوند.[۳۵]

۲.۲.۲. استخراج، تبدیل و بارگذاری داده‌ها

به‌منظور استخراج داده‌ها از بایگانی‌های شهرداری، باید تیم‌های مورد نیاز سازمان‌دهی و آموزش لازم به آنها داده شود. سپس اطلاعات استخراج شده به بانک‌های اطلاعاتی وارد شود و نهایتاً به‌منظور پالایش داده‌های جمع‌آوری شده کنترل‌های لازم به عمل آید و اعتبار داده‌های جمع‌آوری شده مورد بررسی و تأیید قرار گیرد.[۸و۳۵]

۳.۲.۲. تجمیع داده

داده‌های جمع‌آوری شده از منابع مختلف ــ اعم از سیستم‌های عملیاتی و منابع کاغذی و سایر منابع ــ باید دسته‌بندی، و در یک پایگاه داده‌یی تحلیلی جمع‌آوری و سازمان‌دهی شوند.

۴.۲.۲. تحلیل داده‌های جمع‌آوری شده

در این مرحله با استفاده از فناوری‌های داده‌کاوی و پردازش‌های تحلیلیِ برخط، انواع تحلیل‌های موضوعی و مدل‌سازی‌ها روی اطلاعات موجود در پایگاه اطلاعاتیِ به وجود آمده، انجام می‌شود.[۸]

۵.۲.۲. ارائه اطلاعات

تحلیل‌های انجام‌شده با استفاده از فناوری‌های داشبورد[۹]، کارت‌های امتیاز متوازن و ابزارهای گزارش‌گیری به ساده‌ترین شکل در اختیار کارشناسان، مدیران و سیاست‌گذاران شهرداری قرار خواهد گرفت. [۸]

۳. مطالعه‌ی موردی: داده‌های عوارض و نوسازی شهرداری اصفهان

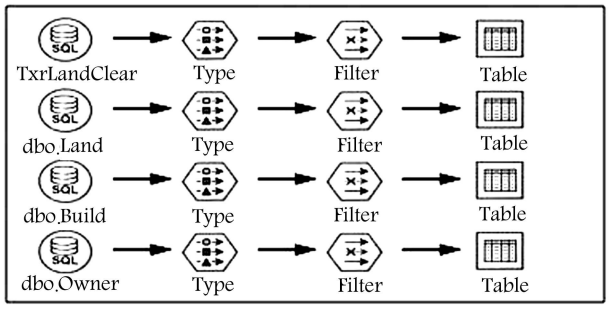
۱.۳. شناسایی منابع داده (شناخت سیستم و کسب اطلاعات)

حوزه‌ی مورد مطالعه در این تحقیق داده‌های بخش عوارض و نوسازی از شهرداری اصفهان، مرکز فنّاوری اطلاعات و ارتباطات (فاوا) است که تمامی داده‌های مرتبط با بخش‌های مختلف شهرداری در آن جمع‌آوری می‌شود، و این سازمان «امین اطلاعاتی» شهرداری است. طی سالیان متمادی ــ از سال ۷۱ تاکنون ــ دو حوزه‌ی «عوارض نوسازی» و «عوارض کسب و پیشه» دارای حجم زیاد داده بوده، و منابع اصلی درآمد شهرداری نیز از طریق همین عوارض تأمین می‌شود. طبق گفته‌ی مسئول مرتبط، در حال حاضر شهرداری علاقه زیادی به افزایش سوددهی این حوزه‌ها دارد.

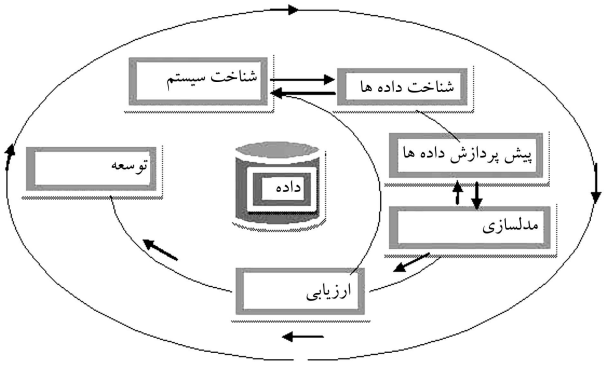
داده‌های مورد استفاده، پایگاه داده‌های متشکل از ۳ میلیون رکورد، شامل جداول مربوط به ویژگی املاک و ویژگی نوسازی و ارتباطات است که برای فهم سرآیندها و آشنایی با نوع و مفهوم داده‌ها با طراح پایگاه داده سیستم مصاحبه کردیم.

۲.۳. استخراج و تجمیع داده‌ها

در این مرحله داده‌های مورد نیاز از پایگاه داده‌ها را استخراج کردیم. برای این کار از جداول پایگاه داده‌ها[۱۰] که در آن روابط بین جداول ترسیم شده بود، استفاده کردیم و از بین حدود ۲۵ جدول ۶ مورد از آنها به‌عنوان جداول اصلی شناخته شدند. سپس در ادامه با بررسی بیشتر این جداول، ۴ جدول به‌عنوان جداول فرعی، و دو جدولِ «صورت‌حساب پرداخت عوارض[۱۱]» و «پرداخت عوارض زمین[۱۲]» به‌عنوان جداول اصلی برای بررسی در مرحله‌ی بعد و انجام عملیات داده‌کاوی انتخاب شدند (شکل ۱).



شکل ۱. جداول انتخابی.



شکل ۲. متدولوژی crisp.[۱۱]

سپس با استفاده از نرم‌افزار[۹] SQL Server و با نوشتن پرسش‌های[۱۳] مختلف، ارتباط بین فیلدهای جداول بررسی شد، و از این طریق تسلط بر داده‌ها برای تحلیل بهتر آنها با استفاده از نرم‌افزار[۱۰] Clementine افزایش یافت.

۳.۳. تحلیل داده‌های جمع‌آوری شده

در این مرحله از متدولوژی crisp برای تحلیل داده‌ها استفاده می‌شود.[۱۱] متدلوژی crisp دارای شش مرحله است که این مراحل در شکل ۲ مشاهده می‌شود.[۱۱] درادامه، پیاده‌سازی این روش برای این پروژه، با کمک نرم‌افزار Clementine بررسی می‌شود.

۱.۳.۳. شناخت سیستم

این گام قبلاً در بخش ۱.۳. مورد بررسی قرار گرفت.

۲.۳.۳. شناخت و بررسی داده‌ها

تعریف مسئله. پیش‌بینی مورد نظر در این تحقیق مربوط به تشخیص مشتری‌های خوش‌حساب و بدحساب سازمان عوارض و نوسازی و ارتباط میزان خوش‌حسابی با دیگر ویژگی‌های ملک یا مالک است.

بررسی داده‌ها و مفهوم آنها. در جدول ۲ نمونه توضیحات مختصر درمورد تعدادی از فیلدهای انتخاب شده از جداول بیان شده است.

۳.۳.۳. پیش پردازش داده‌ها

این مرحله شامل انتخاب داده‌ها، پاک‌سازی داده‌ها، انتخاب ویژگی‌های مرتبط با داده‌کاوی، تجمیع داده‌ها و قالب‌بندی داده‌هاست.[۳۶]

مراحل حذف داده‌های اضافه. مسائل داده‌کاوی ممکن است شامل صدها یا هزاران فیلد باشد که به‌عنوان گزاره‌هایی بالقوه برای استنتاج مورد استفاده قرار می‌گیرند. به‌همین دلیل، ممکن است زمان و نیروی زیادی صرف بررسی این فیلدها

جدول ۲. تعدادی از فیلدهای انتخابی.

نام فیلد	توضیح فیلد
StartDate	زمان صدور برگه عوارض
paidDat	زمان پرداخت عوارض
payAmount	مقدار پرداختی لازم برای عوارض
payStat	وضعیت پرداخت
AmozeshAmnt	سهم آموزش و پرورش
DebitAmnt	مقدار اعتباری

و متغیرهای مربوطه، به‌منظور استفاده از آنها شود. برای حل این مشکل، به‌عنوان نمونه در این تحقیق، ابتدا با نوشتن پرسش‌های متناسب در محیط[۱۴] SQL، تعداد داده‌ها به پانصدهزارکاهش داده شد؛ زیرا زمان پردازش برای داده‌های قبلی بالا بود. همچنین از انتخاب ویژگی‌ها[۱۵] و فیلترینگ[۱۶][۳۶]، برای شناسایی فیلدهایی که از بیشترین اهمیت برای تحلیل برخوردارند و حذف داده‌های اضافی استفاده شد. انتخاب ویژگی‌ها شامل سه گام است:

الف) غربال داده‌ها[۱۷]: حذف گزاره‌ها، فیلدها و نمونه‌های کم‌اهمیت یا مشکل‌دار مانند فیلدهایی با تعداد زیاد مقادیر مفقوده[۱۸] یا با تنوع خیلی زیاد یا خیلی کم مقادیرکه منجر به بی استفاده شدن آنها می‌شود.

ب) اولویت‌دهی[۱۹]: مرتب‌سازی فیلدهای باقی مانده و انتصاب رتبه به هریک متناسب با اهمیت‌شان.

ج) گزینش[۲۰]: انتخاب زیرمجموعه‌یی از فیلدها برای استفاده در مدل مورد نظر، به‌عنوان نمونه با نگه داشتن مهمترین آنها و حذف بقیه.

انتخاب ویژگی‌ها باعث ساده‌سازی و سرعت دادن به فرایند پردازش می‌شود، زیرا می‌توان برمبنای نتایج به دست آمده از آن، داده‌های اضافی جداول را با استفاده از یکی از روش‌های کاهش داده مثل روش فیلترینگ کاهش داد.

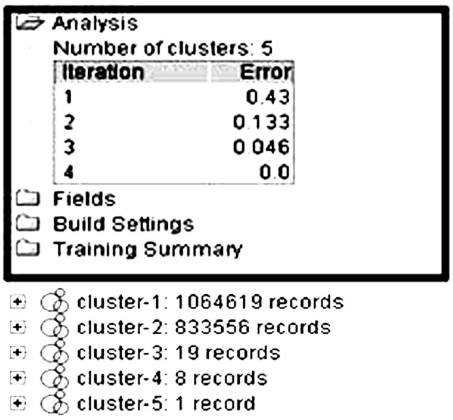
۴.۳.۳. مدل‌سازی

یکی از روش‌های مدل‌سازی داده‌ها، خوشه‌بندی است. خوشه‌بندی فرایند گروه‌بندی مجموعه‌یی از اشیاء در گروه‌های مشابه است. هر «خوشه» مجموعه‌یی از اشیاء داده‌یی است که به اعضای درون خوشه‌ی خود شبیه‌اند و با اعضای دیگر خوشه‌ها متفاوت‌اند. k-means یکی از الگوریتم‌های خوشه‌بندی شناخته شده است که نسبت به نقطه‌ی آغازین برای خوشه‌بندی به k خوشه‌ی اولیه بسیار حساس است.[۱۲]

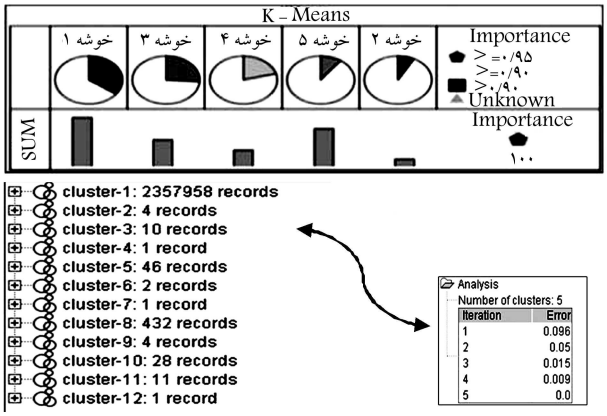
در این مرحله، از خوشه‌بندی داده‌ها برای رسیدن به اهداف پیش‌بینی شده در تعریف مسئله استفاده کردیم. بدین‌منظور، الگوریتم k-means را در چندین مرحله بر روی ویژگی‌های انتخاب شده از جدول‌های اصلی یعنی دو جدول حاصل از مرحله‌ی استخراج و تجمیع داده به کار بردیم. در ادامه شرح مختصری از این اجراها و روند تحلیل روی آنها بیان می‌شود.

اجرای اول. ابتدا روی داده‌های به دست آمده از مرحله‌ی «حذف داده‌های اضافی»، مدل‌سازی k-means انجام شد. این عملیات نتایجی با خطای بالا داشت. نمونه‌ی اجرا برای جدول پرداخت عوارض زمین، در شکل ۳ قابل مشاهده است. چنان که ملاحظه می‌شود تقریباً تمامی داده‌ها (بیش از ۹۹٪) در خوشه‌های ۱ و ۲ قرار گرفته‌اند.

اجرای دوم. در این مرحله الگوریتم را برای ۱۲ =k امتحان کردیم تا تقسیم بیشتری روی خوشه‌ها انجام بگیرد ولی باز هم خوشه‌ی اول بیشترین تجمع داده را داشت و نتیجه‌ی قابل قبولی به دست نیامد.



شکل ۳. اجرای اول الگوریتم k-means.



شکل ۴. اجرای سوم الگوریتم k-means.

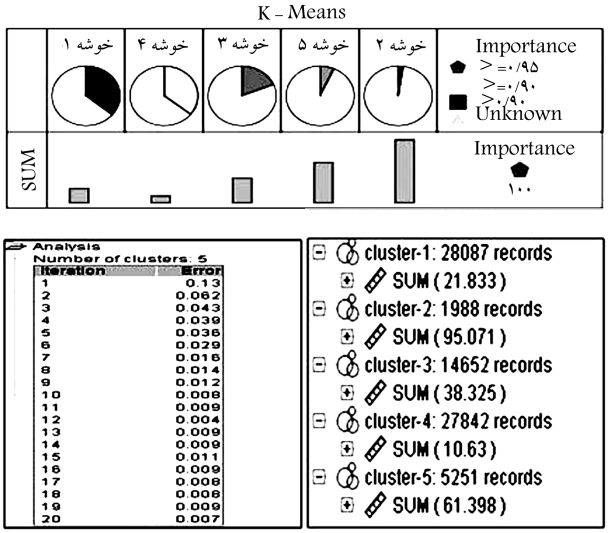
اجرای سوم. با توجه به نتایج به دست آمده در دو اجرای پیشین، در این مرحله تصمیم گرفتیم با اجرای تعدادی پرسش، داده‌ی مناسب‌تری برای مدل‌سازی به دست آوریم. بدین‌منظور با بازگشت به مرحله‌ی پیش‌پردازش داده‌ها، داده‌ی مورد نظر را با استفاده از عملیات ریاضی و طی چند عملیات روی چندین جدول به دست آوردیم، و حاصل را در جدولی جدید در محیط Clementine فراخوانی کردیم. این مقدار جدید تعداد ماه‌هایی است که در این فاصله مشتری عوارض خود را پرداخت کرده است. سپس روی این مقدار مجدداً مدل‌سازی k-mean را انجام دادیم (شکل ۴)؛ نتایج خطای کمی داشت. در این مرحله باید از شاخصی برای مطمئن شدن از کیفیت خوشه‌بندی استفاده کنیم. در ادامه این شاخص معرفی خواهد شد.

شاخص دان. روش‌های اندازه‌گیری متعددی برای اعتبارسنجی خوشه‌بندی وجود دارد که «شاخص دان» یکی از آنهاست. شاخص دان بیانگر نسبت کم‌ترین فاصله‌ی برون‌خوشه‌یی به بیشترین فاصله‌ی درون‌خوشه‌یی است:

$$(D = d_{\min}/d_{\max}) \qquad (۱)$$

که در آن $d_{\min}$ بیانگر کوچک‌ترین فاصله بین دو شیء از دو خوشه‌ی متفاوت است، و $d_{\max}$ بیانگر بزرگ‌ترین فاصله بین دو شیء از یک خوشه است. شاخص دان محدود به بازه $[0, \infty]$ است و باید بیشینه شود.[۱۳]

محاسبه‌ی شاخص دان. برای آخرین خوشه‌بندی انجام‌شده شاخص دان را محاسبه می‌کنیم. کوچک‌ترین فاصله‌ی بین‌خوشه‌یی مربوط به فاصله‌ی بین خوشه‌های ۴



شکل ۵. اجرای چهارم الگوریتم k-mean.

و ۲ (یعنی ۶٫۸ واحد) است. بزرگ‌ترین واحد درون‌خوشه‌یی مربوط به خوشه‌ی سوم با قطر ۱۰۱ است. بنابراین شاخص دان برابر است با:

$$D = \frac{۶٫۸}{۱۰۱} = ۰٫۰۶۷$$

با دسته‌بندی مشتری‌ها در دو دسته خوش‌حساب و بدحساب با استفاده از مقادیر sum در شکل ۴، در این حالت حدود ۳۳ درصد از مشتری‌ها خوش‌حساب و ۶۷ درصد بدحساب شدند، ولی چنان که در این بخش مشاهده کردید، نتایج شاخص دان در این حالت قابل قبول نبود. در نتیجه دسته‌بندی هم مورد پذیرش قرار نگرفت.

اجرای چهارم. بار دیگر الگوریتم k-mean را بر روی مقدار sum اجرا می‌کنیم (شکل ۵).

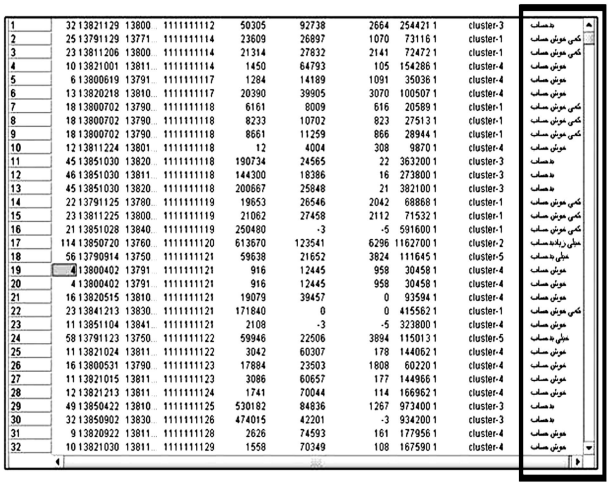
در این اجرا، شاهد نتایجی متفاوت در مقایسه با اجرای قبلی (شکل ۴) هستیم که نشان می‌دهد این اجرا نیز دارای خطای کمی است. حال شاخص دان را محاسبه می‌کنیم:

کوچک‌ترین فاصله‌ی برون‌خوشه‌یی برابر است با ۱۱٫۲ و بزرگ‌ترین فاصله‌ی درون‌خوشه‌یی برابر است با ۲۷. بنابراین مقدار شاخص دان برابر خواهد بود با:

$$D = \frac{۱۱٫۲}{۲۷} = ۰٫۴۱$$

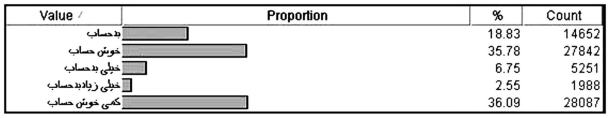
مقدار محاسبه‌شده برای شاخص در فرمول بالا، مؤید قابل قبول بودن این نتیجه است.

در این مرحله، با استفاده از متد derive در Clementine، مشتری‌ها را به ۵ دسته‌ی مختلف، با انتساب ۵ خوشه‌ی به دست آمده در مدل‌سازی، و با توجه به میزان خوش‌حسابی و بدحسابی‌شان در یک ستون جدید تقسیم‌بندی کردیم (شکل ۶)؛ در این مرحله به هدف اصلی، یعنی اختصاص برچسب به مشتری‌ها، نایل آمدیم. سپس از یک نمودار میله‌یی برای به دست آوردن خلاصه‌یی از نتایج استفاده کردیم (شکل ۷). نتایج گویای آن است که تعداد مشتری‌های خوش‌حساب ــ همان‌هایی که در خوشه‌ی ۴ (شکل ۵) با مرکز خوشه با مقدار میانگین ۱۰ =sum قرارگرفته بودند ــ حدود ۳۵ درصد کل مشتری‌ها هستند که این نتیجه‌یی رضایت‌بخش برای شهرداری خواهد بود. البته مشتری‌های کمی خوش‌حساب



1	32	13821129	13800	1111111112	50305	92738	2664	2544211	cluster-3	بدحساب
2	25	13791129	13771	1111111114	23609	26897	1070	731161	cluster-1	کمی خوش حساب
3	23	13811206	13800	1111111114	21314	27832	2141	724721	cluster-1	کمی خوش حساب
4	10	13821001	13811	1111111114	1450	64793	105	1542861	cluster-4	خوش حساب
5	6	13800619	13791	1111111117	1284	14189	1091	350361	cluster-4	خوش حساب
6	13	13820218	13810	1111111117	20390	39905	3070	1005071	cluster-4	خوش حساب
7	18	13800702	13790	1111111118	6161	8009	616	205891	cluster-1	کمی خوش حساب
8	18	13800702	13790	1111111118	8233	10702	823	275131	cluster-1	کمی خوش حساب
9	18	13800702	13790	1111111118	8661	11259	866	289441	cluster-1	کمی خوش حساب
10	12	13811224	13801	1111111118	12	4004	308	98701	cluster-4	خوش حساب
11	45	13851030	13820	1111111118	190734	24565	22	3632001	cluster-3	بدحساب
12	46	13851030	13811	1111111118	144300	18386	16	2738001	cluster-3	بدحساب
13	45	13851030	13820	1111111118	200667	25848	21	3821001	cluster-3	بدحساب
14	22	13791125	13780	1111111119	19653	26546	2042	688681	cluster-1	کمی خوش حساب
15	23	13811225	13800	1111111119	21062	27458	2112	715321	cluster-1	کمی خوش حساب
16	21	13851028	13840	1111111119	250480	-3	-5	5916001	cluster-1	کمی خوش حساب
17	114	13850720	13760	1111111120	613670	123541	6296	11627001	cluster-2	خیلی زیادبدحساب
18	56	13790914	13750	1111111121	59638	21652	3824	1116451	cluster-5	خیلی بدحساب
19	4	13800402	13791	1111111121	916	12445	958	304581	cluster-4	خوش حساب
20	4	13800402	13791	1111111121	916	12445	958	304581	cluster-4	خوش حساب
21	16	13820515	13810	1111111121	19079	39457	0	935941	cluster-4	خوش حساب
22	23	13841213	13830	1111111121	171840	0	0	4155621	cluster-1	کمی خوش حساب
23	11	13851104	13841	1111111121	2108	-3	-5	3238001	cluster-4	خوش حساب
24	58	13791123	13750	1111111122	59946	22506	3894	1150131	cluster-5	خیلی بدحساب
25	11	13821024	13811	1111111122	3042	60307	178	1440621	cluster-4	خوش حساب
26	16	13800531	13790	1111111123	17884	23503	1808	602201	cluster-4	خوش حساب
27	11	13821015	13811	1111111123	3086	60657	177	1449661	cluster-4	خوش حساب
28	12	13821213	13811	1111111124	1741	70044	114	1669621	cluster-4	خوش حساب
29	49	13850422	13810	1111111125	530182	84836	1267	9734001	cluster-3	بدحساب
30	32	13850902	13830	1111111126	474015	42201	-3	9342001	cluster-3	بدحساب
31	9	13820922	13811	1111111128	2626	74593	161	1779561	cluster-4	خوش حساب
32	10	13821030	13811	1111111129	1558	70349	108	1675901	cluster-4	خوش حساب

شکل ۶. ستون‌های جدید ایجاد شده پس ازاجرای الگوریتم k-means.



Value	Proportion	%	Count
بدحساب		18.83	14652
خوش حساب		35.78	27842
خیلی بدحساب		6.75	5251
خیلی زیادبدحساب		2.55	1988
کمی خوش حساب		36.09	28087

شکل ۷. پنج دسته مشتری و تعداد آنها درهر خوشه.

ـ آنها که فاصله‌ی تعداد ماه‌های پرداختی‌شان بین یک تا دو سال بود (۲ =sum) ـ حدود ۳۶ درصد مشتری‌ها را شامل شدند و بقیه که کم‌تر از ۳۰ درصد مشتری‌ها را شامل می‌شوند، مشتری‌های بدحساب، خیلی بدحساب و خیلی زیاد بدحساب‌اند که می‌توانند به‌عنوان لیست سیاه برای شهرداری محسوب شوند.

۵.۳.۳. ارزیابی

در این مرحله از متدولوژی crisp، باید به ارزیابی نتایج بپردازیم.

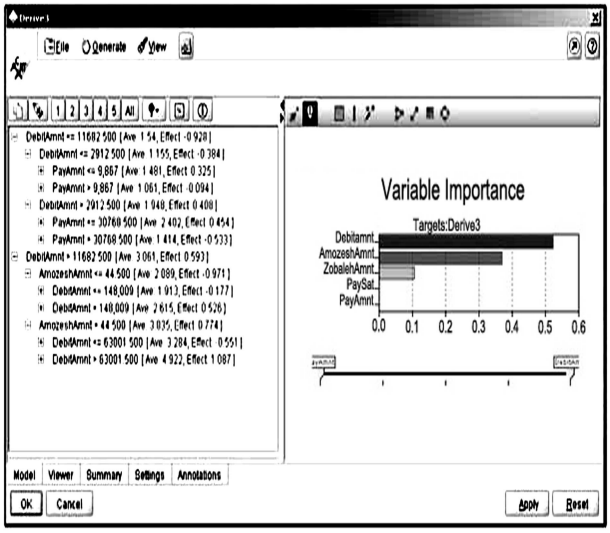
برای ارزیابی نتایج به دست آمده از مدل‌سازی، از دو الگوریتم شبکه‌ی عصبی[۲۱] و درخت تصمیم استفاده کردیم. الگوریتم شبکه‌ی عصبی یک مدل پیشگویی شبکه‌ی عصبی است و الگوریتم درخت تصمیم[۲۲] یک درخت تصمیم پیش‌بینی می‌سازد که درآن از دو الگوریتم خوشه‌بندی و رگرسیون استفاده می‌شود.

طبق نتایج به دست آمده (شکل ۸) در درخت تصمیم، از بین متغیرهای داده شده به این الگوریتم، اهمیت متغیرهای DebitAmnt و AmozeshAmnt که توضیح آنها در جدول ۱ آمده از همه بیشتر شناخته شده و درخت به دست آمده نیز براساس همین اولویت متغیرها سطح‌بندی شده است.

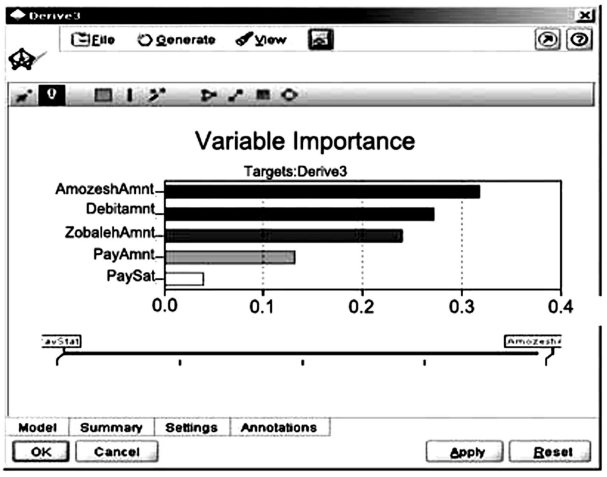
همچنین براساس الگوریتم شبکه‌ی عصبی که درشکل ۹ نتیجه‌ی آن قابل مشاهده است این ترتیب اولویت متغیرها کمی متفاوت است؛ ابتدا AmozeshAmnt، سپس DebitAmnt، و بعد ZobalehAmnt دارای اولویت‌اند. متغیرهای PayAmnt و PayStat نیز در اولویت‌های بعدی قرار گرفته‌اند.

۴.۳. ارائه‌ی اطلاعات

درادامه گام‌های پیاده‌سازی هوش تجاری در شهرداری، پس ازانجام مراحل شناسایی منابع داده، استخراج و تجمیع داده و تحلیل داده‌های جمع‌آوری شده، در بخش‌های قبل این اطلاعات را می‌توان در قالب گزارش‌هایی در نمایش‌گر (داشبورد) سیستم شهرداری قرار داد.



شکل ۸. نتایج اجرای الگوریتم درخت تصمیم.



شکل ۹. نتایج اجرای الگوریتم شبکه عصبی.

۴. مشکلات تحقیق

از جمله مشکلاتی که در این تحقیق با آن روبه‌رو بودیم می‌توان اشاره کرد به:

۱. محدود بودن تحقیقات درمورد کاربرد مدیریت ارتباط با مشتری و هوش تجاری درکنار یکدیگر.

۲. وجود موانعی برای دسترسی به داده‌های شهرداری به‌علت محرمانه‌بودن داده‌ها.

۳. زمان‌بر بودن پیش‌پردازش داده‌ها (به‌علت عدم قابلیت تبدیل داده‌های تاریخی به‌صورت روز، ماه، و سال مجزا و...).

۴. حجم زیاد داده‌ها و زمان‌بر بودن عملیات پردازش توسط نرم‌افزارهای تحلیلی.

۵. نتیجه‌گیری

در شهرداری‌ها با رشد فنّاوری اطلاعات، حرکت به سوی تدارک زیرساخت‌های توانمند فنّاوری اطلاعات نظیر سیستم‌های اطلاعاتی، از چالش‌های راهبردی است و طی دوره‌های مختلف مدیران با پیاده‌سازی سیستم‌های اطلاعاتی در شهرداری

در جهت کسب آگاهی مناسب در زمینه‌ی روند موجود در سیستم خود می‌کوشند. در این نوشتار نیز در راستای نیل به این هدف بهره‌گیری از راهکارهای هوش تجاری و مدیریت ارتباط با مشتری در قالب پایگاه داده تحلیلی، در شهرداری الکترونیک بررسی شد و با پیاده‌سازی گام‌های جمع‌آوری، ذخیره، پاک‌سازی، تجمیع، تحلیل (با استفاده از گام‌های متدولوژی crisp) و بازیابی داده، برای داده‌های مطالعه‌ی موردی ــ یعنی «اطلاعات بخش عوارض و نوسازی مرکز فنّاوری اطلاعات و ارتباطات شهرداری اصفهان» ــ مشتری‌های این سازمان با استفاده از الگوریتم مدل‌سازی k-means در پنج دسته‌ی مختلف خوش‌حساب، کمی خوش‌حساب، بدحساب، خیلی بدحساب و خیلی زیاد بدحساب دسته‌بندی و برچسب‌گذاری شد. سپس تأثیرگذارترین ویژگی‌ها بر خوش‌حسابی و بدحسابی مشتری شناسایی شد و امکان پیش‌بینی رفتار مشتریان در این زمینه با توجه به ویژگی‌های ملک و مالک فراهم و با استفاده از الگوریتم‌های درخت تصمیم و شبکه‌ی عصبی، تحلیل شد. نتایج حاصل از دو روش تحلیل کاملاً مشابه نبودند اما هر دو روش دو عامل مشابه را به‌عنوان اصلی‌ترین عوامل معرفی کردند.

این تحقیق داده‌های مربوط به چند ناحیه از نواحی شهر را مورد بررسی قرار داد و در گام بعد، توسعه‌ی این تحلیل برای دیگر نواحی و بررسی دقیق‌تر تأثیر محل ملک بر خوش‌حسابی شهروند بررسی خواهد شد. همچنین الگوریتم‌های دیگر مانند Apriori، C۵٫۰ یا ابزارهای نوین فازی و آنالیز RFM روی داده‌ها را می‌توان به‌منظور کسب نتایج دقیق‌تر در تحقیقات دیگر به کار گرفت. علاوه بر این گسترش کاربرد هوش تجاری و داده‌کاوی برای دیگر حوزه‌های اطلاعاتی شهرداری و مطالعه‌ی موردی دیگر بخش‌ها برای دست‌یابی به مزیت رقابتی بیشتر و پایدارتر از دیگر زمینه‌های تحقیق در آینده است.

پانوشت‌ها

1. business intelligence (BI)
2. decision support system (DSS)
3. customer relationship management (CRM)
4. online analytical processing
5. data mining
6. most-valuable customers
7. most-growable customers
8. below-zero customers
9. Dashboard
10. entity-relationship diagram (ERD)
11. txrLandBill
12. txrLandPaid
13. query
14. با استفاده از دستور TOP
15. feature selection
16. Filtering برای مثال در جدول پرداخت عوارض زمین از ۳۳ رکورد موجود، ۱۸ رکورد آن فیلتر شده است.
17. screening
18. missing value
19. ranking
20. selection
21. neural net
22. C&R tree

منابع (References)

1. Rouibah, K. and Ould-ali S. "PUZZLE: A concept and prototype for linking business intelligence to business strategy", *Strategic Information Systems*, **11**(2), pp. 133-152 (2002).
2. Phan, D. and Vogel, D. "A model of customer relationship management and business intelligence systems for catalogue and online retailers", *Information & Management*, **47**, pp. 69-77 (2010).
3. Dorado, A. "MIEEE: Advanced business intelligence", ICE–The TECH Conference (2008).
4. Wang, Z., *Business Intelligence*, Taiwan, DrMaster Culture Limited Company (2005).
5. Lin, Y., Tsai, K., Shiang, W., Kuo T. and Tsai, C. "Research on using ANP to establish a performance assessment model for business intelligence systems", *Expert Systems with Applications*, **36**, pp. 4135-4146 (2009).
6. Vince, L., *ASEAN SW Group: Introduction to Business Intelligence*, IBM (2007).
7. Griffin, J. "Customer segmentation: Divide and prosper", IQ Magazine, Cisco Systems (March-April 2003).
8. Dubé, E., Dard, T. and Bédard, Y. "Building geospatial business intelligence solutions with free and open source components", FOSS4G (2007).
9. SQL Server 2005, www.microsoft.com.
10. Celementine 12, www.spss.com.
11. Chapman, P., Clinton, J. and Reinartz, T. "CRISP-DM 1.0 step-by-step data mining guide", pp-35-62 (2000).
12. Han, J. and Kamber, M., *Data mining Concepts and Techniques*, San Francisco, Morgan Kaufmann Publishers (2001).
13. Dunn, J. "Well separated clusters and optimal fuzzy partitions", *Journal of Cybernetics*, **4**, pp. 95-104 (1974).
14. Park, S.C. "Intelligent profitable customers segmentation system based on business intelligence tools", *Expert Systems with Applications*, **29**, pp. 145-152 (2005).
15. Peppers, D. Rogers, M., "The One to One Manager: Real -World Lessons in Customer Relationship Management" Currency/Doubleday, New York (2002).
16. Ngai, E.W.T., Xiu, L. and Chau, D.C.K. "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, **36**, pp. 2592-2602 (2009).
17. Dennis, C., Marsland, D. and Cockett, T. "Data mining for shopping centrescustomer knowledge management framework", *Journal of Knowledge Management*, **5**, pp. 368-374 (2001).

18. Lee, J.H. and Park, S.C. "Intelligent profitable customer segmentation system based on business intelligence tools", *Expert Systems with Applications*, **29**, pp. 145-152 (2005).

19. Yang, Y. and Padmanabhan, B. "A hierarchical pattern-based clustering algorithm for grouping web transactions", *IEEE Transaction on Knowledge and Data Engineering*, **17**, pp. 1300-1304 (2005).

20. Bae, S.M., Park, S.C. and Ha, S.H. "Fuzzy web ad selector based on web usage mining", *IEEE Intelligent Systems,* pp. 62-69 (2003).

21. Verdú, S.V., Garcı'a, M.O., Senabre, C., Marı'n, A.G., and Franco, F.J.G. "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps", *IEEE Transactions on Power Systems,* **21**, pp. 1672-1682 (2006).

22. Lee, S.C., Suh, Y.H., Kim, J.K., and Lee, K.J. "A cross-national market segmentation of online game industry using SOM", *Expert Systems with Applications*, **27**, pp. 559-570 (2004).

23. He, Z., Xu, X., Huang, J.Z., and Deng, S. "Mining class outliers: Concepts, algorithms and applications in CRM", *Expert Systems with Applications*, **27**, pp. 681-697 (2004).

24. Bae, S.M., Ha, S.H., and Park, S.C. "A web-based system for analyzing the voices of call center customers in the service industry", *Expert Systems with Applications*, **28**, pp. 29-41 (2005).

25. Li, S.T., Shue, L.Y. and Lee, S.F. "Enabling customer relationship management in ISP services through mining usage patterns", *Expert Systems with Applications*, **30**, pp. 621-632 (2006).

26. Liao, S.H. and Chen, Y.J. "Mining customer knowledge for electronic catalog marketing", *Expert Systems with Applications,* **27**, pp. 521-532 (2004).

27. Cho, Y.H. and Kim, J.K. "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce", *Expert Systems with Applications*, **26**, pp. 233-246 (2004).

28. Kuo, R.J., Liao, J.L. and Tu, C. "Integration of art2 neural network and genetic kmeans algorithm for analyzing web browsing paths in electronic commerce", *Decision Support Systems*, **40**, pp. 355-374 (2005).

29. Min, S.H. and Han, I. "Detection of the customer time-variant pattern for improving recommender systems", *Expert Systems with Applications*, **28**, pp. 189-199 (2005).

30. Song, H.S., Kim, J.K. and Kim, S.H. "Mining the change of customer behavior in an internet shopping mall", *Expert Systems with Applications*, **21**, pp. 157-168 (2001).

31. Kwan, I.S.Y., Fong, J. and Wong, H.K. "An e-customer behavior model with online analytical mining for internet marketing planning", *Decision Support Systems*, **41**, pp. 189-204 (2005).

32. Chang, S.E., Changchien, S.W. and Huang, R.H. "Assessing users' productspecific knowledge for personalization in electronic commerce", *Expert Systems with Applications*, **30**, pp. 682-693 (2006).

33. Drew, J.H., Mani, D.R., Betz, A.L. and Datta, P. "Targeting customers with statistical and data-mining techniques", *Journal of Service Research*, **3**, pp. 205-220 (2001).

34. Rosset, S., Neumann, E., Eick, U. and Vatnik, N. "Customer lifetime value models for decision support", *Data Mining and Knowledge Discovery*, **7**, pp. 321-339 (2003).

35. R & D unit of Safarayaneh, Automating the processes and Knowledge Base creation in e-municipalices (in Persian)(2005).

36. Ghazanfari, M., Alizadeh, S., Teimurpur, B. "Data Mining and Knowledge Discovery" ,publication of Iran University of Science and Technology (in Persian)(2008).