

ارائه‌ی مدلی مبتنی بر داده‌کاوی برای پیش‌بینی جذب مشتری با استفاده از درخت تصمیم در «مدیریت ارتباط با مشتری»

ابوالفضل کاظمی* (استادیار)

محمداسماعیل بابائی (کارشناس ارشد)

دانشکده‌ی مهندسی صنایع و مکانیک، دانشگاه آزاد اسلامی واحد قزوین

مهندسی صنایع و مدیریت شریف، زمستان ۱۳۹۲ (دوره ۱ - شماره ۲، ص. ۲۳-۱۵)

در دنیای رقابتی و کیفیت‌گرای امروز، جذب مشتری از اهمیت زیادی برخوردار است. از این رو، «مدیریت روابط با مشتری» به عنوان هسته‌ی اصلی استراتژی سازمان در چهار بعد: شناسایی، جذب، نگهداری و رضایت مشتری به ایفای نقش می‌پردازد. سازمان‌ها با تجزیه و تحلیل چرخه‌ی زندگی مشتری به افزایش ارزش مشتری دست یافته‌اند. این ادبیات با کاربرد عملی داده‌کاوی در شناسایی مشتریان بالقوه، سعی دارد که معیارهای شناسایی این مشتریان را در محیط رقابتی کسب و کارشان تعیین، و سازوکاری برای بالفعل شدن آنان ارائه دهد. در این مطالعه، با استفاده از ابزار درخت تصمیم معیارهای اصلی و زیرمعیارهایی را شناسایی و سپس میزان اهمیت آن‌ها را تعیین می‌کنیم. از این طریق به سازمان این امکان داده می‌شود که در هر مراجعه فرایند فروش را به سمتی سوق دهد که منجر به خرید از سوی مراجعه‌کننده شود.

واژگان کلیدی: داده‌کاوی، درخت تصمیم، مدیریت ارتباط با مشتری.

۱. مقدمه

ارائه‌ی خدمات و محصولاتی با کیفیت بالاتر و مطابق با نیازهای خاص هر مشتری تحت فشارند.

امروزه جذب مشتری از اهمیت زیادی برخوردار است در محیط رقابتی کسب و کار امروز توانایی تشخیص مشتریان سودآور و توسعه‌ی مستمر روابط موجود یکی از عوامل کلیدی رقابت برای سازمان است. مدیریت ارتباط با مشتری (CRM) به سازمان‌ها کمک می‌کند تا ارزش مشتریان را دریابند و سودآورترین مشتریان خود را شناسایی کنند و روابط کیفی خود را با آن‌ها توسعه دهند. بررسی دقیق سودآوری هر مشتری و همچنین شناسایی سودآورترین مشتریان از عوامل مهم در موفقیت CRM است.^[۱]

مدیریت ارتباط با مشتری (CRM) یک راهکار کلیدی تجاری است که در آن سازمان با تمرکز بر نیازهای مشتریانش، یک رویکرد مشتری‌مدار را در کل سازمان توسعه می‌دهد.^[۱] یک فرایند کامل تجاری، شامل جذب مشتری، افزایش مشتری و حفظ مشتری است.^[۲] استخراج داده‌ها روش جدید کمک به سازمان‌ها در شناسایی الگوها و تمایلات مشتریان و بهینه‌سازی روابط با آنان است. این امر یکی از مهم‌ترین ابزارهای شناخته شده در مدیریت ارتباط با مشتری است.^[۲]

داده‌کاوی و کشف دانش در پایگاه داده‌ها از جمله موضوع‌هایی است که همزمان با ایجاد و استفاده از پایگاه داده‌ها در اوایل دهه‌ی ۸۰ برای جست‌وجوی دانش در داده‌ها شکل گرفت. یکی از مهم‌ترین وظایف سازمان شناسایی مشتری و سرمایه‌گذاری روی مشتریان بالقوه است. اگر سازمان به‌درستی تشخیص دهد که

امروزه سازمان‌های تولیدی با توجه به نقش تولید در نظام رقابتی بازار، محیطی کاملاً تغییر یافته را تجربه می‌کنند. در مدل‌های کسب و کار دهه‌های گذشته: به دلیل بهره‌گیری از روش‌های تولید سنتی عرضه کم‌تر از تقاضا بود، تولیدکننده‌سالاری وجود داشت، هدف تولیدکنندگان تنها تولید هرچه بیشتر بدون توجه به افزایش کیفیت بود، تنوع بسیار محدودی در محصولات دیده می‌شد، هزینه مهم‌ترین عامل رقابت بود، و چرخه‌ی زندگی محصول زیاد بود و محصولات تولیدی مدت‌زمان بیشتری به‌روز بودند. پیشرفت سیستم‌های یکپارچه‌ی مدیریت تولید و به‌کارگیری روش جدید مدیریت، انتخاب استراتژی‌های مناسب تولیدی، پیشرفت سامانه‌های الکترونیکی با توسعه‌ی زیرساخت‌های فناوری اطلاعات و ارتباطات (ICT)، تحول شگرفی در دنیای صنعتی به وجود آورد. در مدل‌های کسب و کار امروزی با استفاده از شیوه‌های جدید تولید و توجه به مشتریان به‌عنوان تنها مرکز توجه مراکز تولیدی و با هدف کسب حداکثر رضایت آنان، عرضه بیشتر از تقاضا شده است. دیگر تولیدکننده‌سالاری وجود ندارد بلکه مشتری‌سالاری است و این مشتری است که تعیین می‌کند چه کالایی باید تولید شود، به عبارتی هدف کمیّت تولید نیست، بلکه کیفیت مهم است. چرخه‌ی زندگی محصول کوتاه شده و تنوع بسیار زیادی در محصولات به چشم می‌خورد؛ نهایتاً پاسخ‌گویی به‌عنوان مهم‌ترین عامل رقابت مطرح است. سازمان‌ها اینک برای

* نویسنده مسئول

تاریخ دریافت: ۱۳۸۹/۱۰/۲۹، اصلاحیه ۱۳۹۰/۷/۳۰، پذیرش ۱۳۹۰/۱۰/۱۰

چه مشتریان بالقوه‌ی به مشتریان بالفعل تبدیل می‌شوند به راحتی آن‌ها را شناسایی، جذب و نگه‌داری می‌کند و توسعه می‌دهد.^[۲]

در حوزه‌ی مدیریت ارتباط با مشتری با استفاده از مدل‌های داده‌کاوی کارهای زیادی ارائه شده، اما تاکنون مدلی ارائه نشده که به کمک آن بتوان با توجه به شناسایی متغیرهای مؤثر در تبدیل مشتریان بالقوه به بالفعل در مرحله‌ی جذب مشتری، توان به نتیجه رساندن فرایند فروش (انجام خرید از سوی مراجعه‌کننده) را افزایش دهد. مدل‌های ارائه شده بیشتر مربوط به ارائه‌ی راهکار بعد از دریافت خدمات توسط مشتری از سازمان است، ولی در این نوشتار مدلی ارائه خواهد شد که با توجه به شناسایی عوامل مؤثر در افزایش فروش، برای جذب مراجعه‌کنندگان و ارائه‌ی خدمات به آن‌ها و نگه‌داری و توسعه‌ی آنان تلاش می‌کند؛ یعنی از این طریق سعی می‌شود مراجعه‌کنندگان به مشتریان وفادار تبدیل شوند و سرمایه‌گذاری‌های سازمان بر آن دسته از مشتریان که طبق مدل خرید خواهند کرد، بیشتر خواهد بود. این مدل از اتلاف هزینه‌های سازمان که برای توسعه‌ی بازار و کسب سهم بیشتر از بازار قابل کسب است، جلوگیری می‌کند.

۲. مرور ادبیات

شاید بتوان لوول (۱۹۸۳) را اولین شخصی دانست که گزارشی در مورد داده‌کاوی تحت عنوان «شبیه‌سازی فعالیت داده‌کاوی» ارائه کرد. هم‌زمان، پژوهش‌گران و متخصصین علوم رایانه، آمار، هوش مصنوعی، یادگیری ماشین و غیره نیز به پژوهش در این زمینه و حوزه‌های مرتبط با آن پرداختند. پژوهش جدی روی موضوع داده‌کاوی از اوایل دهه‌ی ۹۰ شروع شد. پژوهش‌ها و مطالعه‌های زیادی در این زمینه صورت گرفت، همچنین سمینارها، دوره‌های آموزشی و کنفرانس‌های زیادی نیز برگزار شد.^[۳]

امروزه با استفاده از ابزارهای استخراج داده‌ها برای کمک‌کردن به مدیریت ارتباط با مشتری (CRM)، از روش‌هایی مثل درخت تصمیم (DT)^۲، شبکه‌های عصبی مصنوعی (ANN)^۳، الگوریتم ژنتیک (GA)^۴، قوانین پیوستگی (AR)^۵ و غیره در برخی رشته‌ها مثل مهندسی، علوم، مالی، تجارت برای حل مشکلات مربوط به مشتریان استفاده شده است. درخت تصمیم ساختاری شبیه نمودار آویخته است که هر ساقه‌ی داخلی آن حاکی از یک آزمایش، هر شاخه‌اش نشان‌گر خروجی آزمایش، و برگ‌های ساقه‌ها نشان‌گر درجه یا پراکنندگی درجات‌اند. شبکه‌ی عصبی مصنوعی شامل یک سری سازه‌های پردازش‌گر متصل به هم (نورون‌ها) است که با استفاده از مدل آماری و مدل محاسبه‌ی یا ابزارهای مدل داده‌های آماری غیرخطی می‌تواند به پردازش اطلاعات بپردازد و روابط ورودی و خروجی پیشرفته را نشان دهد. الگوریتم ژنتیک، که اولین بار در سال ۱۹۷۰ توسط جان هولند در دانشگاه میشیگان معرفی شد، الگوریتم جویای حل مشکلات رایانه‌ی براساس انتخاب و تکامل پردازش معمولی است. در سال ۲۰۰۵، محققین در مطالعات خود نظریه‌ی RS و مدل RFM را تشریح کردند و مزایای این دو روش را در تقسیم‌بندی مشتریان به همراه مطالعه‌ی موردی بیان کردند. در نظریه‌ی RS هیچ‌گونه داده‌ی اولیه و اضافی برای پارامترهای ضروری نیست؛ این نظریه با داده‌های اولیه‌ی ناصحیح عملیاتی می‌شود. این نظریه از داده‌های کمی و کیفی به طور همزمان می‌توان بهره می‌برد و منجر به تولید قوانینی قابل فهم خواهد شد. اما تقسیم‌بندی مشتریان در روش RFM مبتنی بر سه ویژگی است: جدیدبودن آخرین خرید (R)، تکرار خرید (F) و ارزش پولی خرید (M).^[۵]

در سال ۲۰۰۴ پژوهش‌گران ضمن اشاره به مفهوم پایگاه داده‌های کشف دانش

(K.D.D) چندین شیوه‌ی ریاضی خوشه‌بندی فازی مبتنی بر الگوریتم فازی ارائه کردند. آنان با تمرکز بر کشف دانش و استخراج آن از پایگاه داده‌های بزرگ نشان دادند که روش‌های فازی برای خوشه‌بندی مناسب است.^[۶]

در همان سال محققین دیگری ضمن اشاره به اهمیت والای جذب و نگه‌داری مشتریان سودآور برای سازمان‌ها، محاسبه‌ی ارزش مشتریان را دشوار دانستند. آن‌ها با استفاده از LTV چارچوبی برای تحلیل ارزش مشتری ارائه دادند که طی آن ارزش مشتری را سه گروه ارزش جاری، ارزش بالقوه، ارزش وفاداری جای گرفت و بر همین اساس ارزش مشتری را محاسبه کردند. آن‌ها تعریف کامل و ادبیات جامعی از مدل‌های LTV بیان کرده‌اند و مدل LTV جدیدی ارائه داده‌اند.^[۷]

در سال ۲۰۰۸، ناگی و همکاران در بازه زمانی سال ۲۰۰۰ تا ۲۰۰۶ با در نظر گرفتن چهار بعد مدیریت ارتباط با مشتری (شناسایی مشتریان، جذب مشتری، نگه‌داری مشتری و توسعه و تعالی مشتری) و ابعاد داده‌کاوی (طبقه‌بندی، خوشه‌بندی، پیش‌بینی، رگرسیون و غیره) به طبقه‌بندی مطالعات انجام شده پرداختند.^[۸] در همین سال مدلی برای دسته‌بندی ارزش مشتریان براساس خصیصه‌ی RFM و الگوریتم میانگین K که براساس تقسیم‌بندی ارزش میانه مشتریان در هر گروه عمل می‌کند، بیان شد (مدل RFM به عنوان خصیصه‌ی ورودی در نظر گرفته شد) و سپس ارزش کمی برای دسته‌بندی میانگین K محاسبه شد. همچنین یک روش ابداعی چهارمرحله‌ی ارائه شده است.^[۸]

در تحقیق دیگری که در سال ۲۰۰۸ انجام شد، ادبیات موضوعی مربوط به روش درخت تصمیم و خوشه‌بندی و مباحث عمومی مربوط به نظریه‌ی گراف مورد مطالعه قرار گرفت و الگوریتمی مبتنی بر درخت تصمیم برای کمینه‌کردن هزینه ارائه شد.^[۹]

در سال ۲۰۰۹ نیز پژوهشگران در تحقیقات خود ضمن معرفی e-CRM الگوریتمی بر پایه‌ی شبکه‌های عصبی برای خوشه‌بندی فازی^۷ ارائه کردند. آن‌ها این الگوریتم را FC-CRM نامیدند. توسعه و ارائه‌ی مدل پیشنهادی آن‌ها در e-CRM مورد استفاده قرار گرفته است. همچنین آن‌ها خوشه‌بندی پویا را از طریق روش شبکه‌های عصبی به عنوان راه حلی برای طبقه‌بندی مشتریان مطرح کردند. مدل آن‌ها دارای قابلیت یادگیری (همچون شبکه‌های عصبی) است.^[۱۰]

در پایان این بخش باید به این نکته اشاره کرد که تحقیقات نظری زیادی در حوزه‌ی مدیریت دانش و داده‌کاوی انجام شده است (به دلیل رویکرد دانش‌محور عصر کنونی)، اما برای پیش‌بینی جذب مشتری و شناسایی مشتریان بالقوه متأسفانه به دلیل عدم جافتادن موضوعات دانش‌محور در سازمان‌ها و مشکلات کاربردی کردن این سیستم‌ها و ترس از عدم موفقیت در کاربرد این نوع سیستم‌ها تحقیقات زیادی در این مورد انجام نشده است.

۳. متدولوژی تحقیق

براساس گام‌های داده‌کاوی، مراحل طی شده برای ارائه‌ی مدل پیش‌بینی در بازه زمانی ۱۸ ماهه (سال ۸۷ و نیمه‌ی اول سال ۸۸) در سازمان مورد مطالعه (تولیدکننده‌ی مبل‌مان اداری با ۳۵ سال سابقه در این عرصه) عبارت است از (شکل ۱):

- جمع‌آوری داده،
- آماده‌سازی و پاک‌سازی داده‌ها،
- کد کردن داده‌ها،
- تعیین متغیر هدف،

و نمایندگی)، نحوه‌ی درخواست محصول (تلفنی، پس از نمایش محصول)، کارشناس فروش (کارشناس شماره ۱، کارشناس شماره ۲)، نوع محصولات (پارتیشن تک‌جداره، پارتیشن دوجداره، متعلقات پارتیشن، میز مدیریتی، میز کارشناسی، میز کارمندی، میز کنفرانس، قفسه، صندوقی وارداتی، صندوقی داخلی، بغل مبلی / جلوبمبلی / نیمه‌ست، مواد اولیه و سایر خدمات)، کارشناس طراحی (کارشناس شماره ۱، کارشناس شماره ۲، کارشناس شماره ۳، کارشناس شماره ۴)، نتیجه‌ی مراجعه (خرید، انصراف)، دلایل عدم خرید، موعد تحویل (LT)، عوامل اجرایی نصب، مبالغ پیش فاکتور و فاکتور، تخفیفات، هزینه‌های حمل و نصب.

۲. جمع‌آوری داده‌ها طبق بند ۱ از طریق فرم‌های مورد نیاز، دستورالعمل‌ها و آیین‌نامه‌های اجرایی.

۲.۳. آماده‌سازی و پاک‌سازی داده‌ها

در این مرحله با توجه به داده‌های به دست آمده از بند ۱.۳. که در نرم‌افزار Access جمع‌آوری شده، نسبت به آماده‌سازی و پاک‌سازی داده‌هایی که به‌طور کامل و صحیح در اختیار نبود اقدام شد.

۳.۳. کدکردن داده‌ها

پس از آماده‌سازی و پاک‌سازی، داده‌ها به‌صورتی رمزگذاری شدند که در نرم‌افزار Excel قابل کدکردن باشند.

۴.۳. تعیین متغیر هدف

از آنجا که در فرایند فروش، بهره‌گیری یا عدم بهره‌گیری از خدمات سازمان در نتیجه‌ی مراجعه نقش دارد، متغیر «نتیجه‌ی مراجعه» به‌عنوان متغیر هدف در نظر گرفته شده است.

۵.۳. تعیین متغیرهای مدل پیش‌بینی

برای این منظور از دو روش -- استفاده از نرم‌افزار Clementine و استفاده از پرسش‌نامه -- با اهدافی که در ادامه به آن‌ها اشاره خواهد شد استفاده شده است:

۱.۵.۳. مدل‌سازی به‌کمک درخت تصمیم، مقایسه و انتخاب درخت بهینه و تعیین متغیرهای مدل

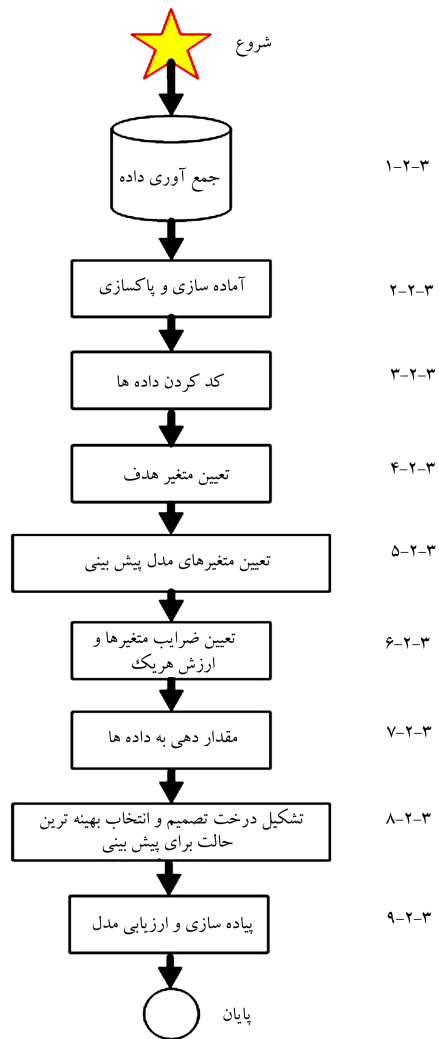
در این مرحله، به‌منظور شناسایی متغیرهای مدل برای داده‌های سال ۸۷ و نیمه‌ی اول سال ۸۸، از داده‌های تهیه شده در بند ۳.۳. توسط چهار مدل CHAD، CRT، QUEST و C5.0 -- از مدل‌های معروف در تولید درخت تصمیم (با رویکرد پیش‌بینی) هستند -- در نرم‌افزار Clementine استفاده می‌شود (شکل ۲).

الف) داده‌های سال ۸۷

با قراردادن داده‌های سال ۸۷ (۴۲۴ داده در ۳۳۹۲ فیلد) در مدل‌سازی یادشده در بند ۱.۵.۳، به مدل C5.0 ۸۳٫۹۶ درصد اهمیت به‌عنوان بهینه‌ترین متغیرها و درخت تصمیم دست یافتیم (جدول ۱).

ب) داده‌های نیمه‌ی اول سال ۸۸

با قرار دادن داده‌های نیمه‌ی اول سال ۸۸ (۲۰۸ داده در ۱۶۶۴ فیلد) در



شکل ۱. متدولوژی تحقیق.

-- تعیین متغیر مدل پیش‌بینی،

-- تعیین ضرایب متغیرها و ارزش هر یک،

-- مقداردهی به داده‌ها،

-- تشکیل درخت تصمیم و انتخاب بهینه‌ترین حالت برای پیش‌بینی،

-- پیاده‌سازی و ارزیابی مدل.

۱.۳. جمع‌آوری داده

این مرحله در دو گام اجرا شد:

۱. شناسایی متغیرها و زیرمتغیرهایی که داده‌هاشان براساس مراجعه‌ی تلفنی یا حضوری قابل جمع‌آوری است، پس از تعامل با خبرگان سازمان مورد نظر (تولیدکننده‌ی میلمان اداری با ۳۵ سال سابقه). این متغیرها عبارت‌اند از:

جنسیت مراجعه‌کننده (مرد یا زن)، نام شرکت / فروشگاه / دفتر / محل کار، تاریخ مراجعه، آدرس، تلفن، موبایل، فاکس، پست الکترونیک، سبب کسب و کار (کوچک، بزرگ، متوسط)، نحوه‌ی آشنایی (خرید قبلی، کانالوک و CD، مطبوعات، پرسنل، نمایشگاه، سایت، نامه‌ی تبلیغاتی، بیلبورد، معرف

جدول ۳. میزان اهمیت متغیرها در خروجی پرسش‌نامه.

متغیر	میزان اهمیت
نحوه‌ی درخواست محصول	۰/۰۸
کارشناس فروش	۰/۴۵
نحوه‌ی آشنایی	۰/۲۸
نوع محصول	۰/۱۹

جدول ۴. خلاصه میزان اهمیت متغیر در هر خروجی.

متغیر	جدول ۱	جدول ۲	جدول ۳
نحوه‌ی درخواست محصول	۰/۴۱۱	۰/۱۹۶	۰/۰۸
کارشناس فروش	۰/۱۳۳	۰/۱۳۲	۰/۴۵
نحوه‌ی آشنایی	۰/۲۷۱	۰/۴۲۲	۰/۲۸
نوع محصول	۰/۱۸۵	۰/۲۵	۰/۱۹
میزان درستی	۸۳/۹۶	۸۴/۶۲	۸۴

متغیر نام با متغیر زام.

$$\sqrt[n]{a_{ij1} a_{ij2} \dots a_{ijn}} \quad (1)$$

$$n = 1, \dots, 14 \quad i = 1, \dots, 4 \rightarrow$$

$$j = 1, \dots, 4 \downarrow$$

$$if \ i = j \Rightarrow a_{ijn} = 1 \quad (2)$$

۶.۳. تعیین ضرایب متغیرها و ارزش هر زیر متغیر

۱.۶.۳. تشریح مدل

تابع هدف پیش‌بینی و متغیرهای آن چنین تعریف می‌شود:

$$Z = A(MR_i) + B(CS_i) + C(MF_i) + D\left(\sum_{j=1}^{13} KP_j\right)_i \quad (3)$$

MR_i نحوه‌ی درخواست مشتری نام، CS_i کارشناس فروش مشتری نام، MF_i نحوه‌ی آشنایی مشتری نام، $\left(\sum_{j=1}^{13} KP_j\right)_i$ نوع محصولات درخواستی توسط مشتری i .

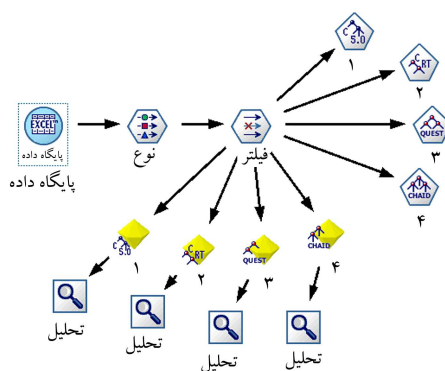
۲.۶.۳. تعیین ضرایب A, B, C و D

برای مشخص کردن این چهار ضریب از خروجی‌های پرسش‌نامه‌های پیوست الف و ب استفاده شده است (جدول ۴). برای به دست آوردن ارزش هر ضریب از رابطه‌ی ۴ استفاده شده است:

$$v(x_j) = \frac{\sum_{i=1}^4 w_i x_{ij}}{\sum_{i=1}^4 w_i} \quad (4)$$

که در آن، w_i میزان درستی خروجی نام $i = 1, \dots, 3$ اهمیت متغیر زام در خروجی نام، $i = 1, \dots, 4$ و $v(x_j)$ ارزش متغیر زام هستند. با جایگذاری اعداد جدول ۴ در رابطه‌ی ۴، رابطه‌ی ۵ به دست می‌آید:

$$Z = 0.223MR_i + 0.24CS_i + 0.32MF + 0.21\left(\sum_{j=1}^{13} KP_j\right)_i \quad (5)$$



شکل ۲. مدل‌سازی برای تعیین متغیرهای مدل.

جدول ۱. میزان اهمیت متغیرها در خروجی $CS_{0.05}$ برای داده‌های سال ۸۷.

متغیر	میزان اهمیت
نحوه‌ی درخواست محصول	۰/۴۱۱
کارشناس فروش	۰/۱۳۳
نحوه‌ی آشنایی	۰/۲۷۱
نوع محصول	۰/۱۸۵

جدول ۲. میزان اهمیت متغیرها در خروجی $CS_{0.05}$ برای داده‌های نیمه اول سال ۸۸.

متغیر	میزان اهمیت
نحوه‌ی درخواست محصول	۰/۱۹۶
کارشناس فروش	۰/۱۳۲
نحوه‌ی آشنایی	۰/۴۲۲
نوع محصول	۰/۲۵

مدل‌سازی بند ۳.۵.۱، به مدل $CS_{0.05}$ با اهمیت ۸۴/۶۲ به عنوان بهینه‌ترین متغیرها و درخت تصمیم دست یافتیم (جدول ۲).

۲.۵.۳. استفاده از پرسش‌نامه برای جمع‌آوری نظر خبرگان

الف) پرسش‌نامه تحقیقاتی شماره ۱

این پرسش‌نامه (پیوست الف) با هدف شناسایی متغیرهای اساسی در تبدیل مشتریان بالقوه به بالفعل در اختیار افراد قرار گرفت. خروجی این بررسی با آلفای کرونباخ ۰/۷۴، عبارت بود از متغیرهای:

۱. کارشناس فروش
۲. نحوه‌ی آشنایی
۳. نحوه‌ی درخواست محصول
۴. نوع محصول

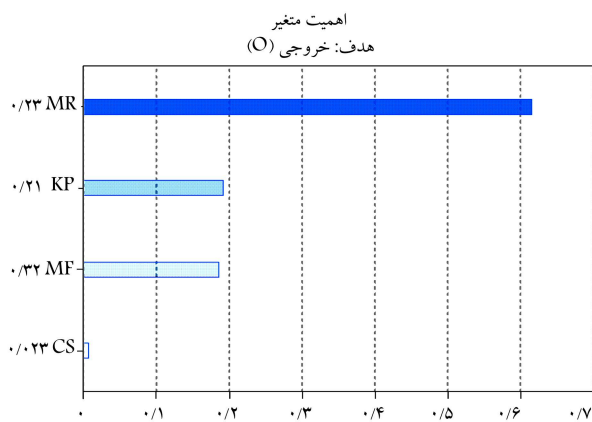
ب) پرسش‌نامه تحقیقاتی شماره ۲

این پرسش‌نامه (پیوست ب) با هدف تعیین میزان اهمیت هریک از ۴ متغیر خروجی از پرسش‌نامه تحقیقاتی شماره ۱، در اختیار افراد مطرح سازمان قرار گرفت.

با استفاده از تحلیل سلسله‌مراتبی گروهی (GAHP) و نیز رابطه‌ی ۱، جدول مقایسات زوجی خلاصه شده و جدول ۳ به دست آمده است. در این مرحله از تحلیل، میزان آلفای کرونباخ برابر ۰/۸۴ است. (a_{ijn}) نظر نفر m در مقایسه‌ی

جدول ۵. ارزش زیرمتغیرها بعد از طی مراحل شش گانه.

متغیر	زیرمتغیر	کد	ارزش زیرمتغیر
MF _i	خرید قبلی	F ₁	۰/۳
	کاتالوگ و CD	F ₂	۰/۲
	مطبوعات	F ₃	۰
	پرسنل	F ₄	۰/۰۱
	نمایشگاه	F ₅	۰/۰۳
	سایت	F ₆	۰/۰۲
	نامه تبلیغاتی	F ₇	۰
	بیلبورد	F ₈	۰/۱۲
	معرف	F ₉	۰/۰۷
	نماینده	F ₁₀	۰/۰۶
MR _i	تلفنی	R ₁	۰/۲
	حضور	R ₂	۰/۴۲
CS _i	شماره ۱	F	۰/۳
	شماره ۲	M	۰/۳۲
KP _j	پارتیشن تک جداره	P ₁	۰/۰۹
	پارتیشن دوجداره	P ₂	۰/۰۶
	متعلقات پارتیشن	P ₃	۰/۱۲
	میز مدیریتی	P ₄	۰/۱
	میز کارشناسی	P ₅	۰/۱۱
	میز کارمندی	P ₆	۰/۰۷
	میز کنفرانس	P ₇	۰/۱۶
	قفسه	P ₈	۰/۱۱
	صندلی وارداتی	P ₉	۰/۰۹
	صندلی داخلی	P ₁₀	۰/۰۸
	بغل مبلی/جلو مبلی/نیم ست	P ₁₁	۰/۰۹
	مواد اولیه	P ₁₂	۰/۲۳
	سایر خدمات	P ₁₃	۰/۴



شکل ۳. اهمیت متغیرها در خروجی CHAID براساس چهار متغیر.

۳.۶.۳. تعیین ارزش هر یک از زیر متغیرها

با توجه به تعیین ضرایب ۴ متغیر اصلی، این متغیرها، مطابق آنچه که در قسمت الف بند ۱.۳.۱، آمد، خود دارای زیرمتغیرند. مراحل انجام شده برای محاسبه ی ارزش هر یک از این زیرمتغیرها عبارت است از:

- تعیین تعداد کارهایی که از طریق نام صورت پذیرفته است (در بازه زمانی مورد نظر)
- محاسبه ی نسبت تعداد کارهایی که از طریق نام صورت پذیرفته است به کل کارها (در بازه زمانی مورد نظر)
- تعیین تعداد کارهایی که از طریق نام منجر به خرید از سوی مشتری است (در بازه زمانی مورد نظر)
- محاسبه ی نسبت تعداد کارهایی که از طریق نام منجر به خرید از سوی مشتری است به کل کارهایی که از طریق نام صورت پذیرفته است (در بازه زمانی مورد نظر)

۵. محاسبه ی امید ریاضی از طریق حاصل ضرب بندهای ۲ و ۴

۶. میانگین گیری از دو عدد محاسبه شده از طریق بند ۵ برای سال ۸۷ و نیمه ی اول سال ۸۸، به طور جداگانه به منظور کسب ارزش هر زیرمتغیر:

$$v(F) = \frac{(A.a)_{87} + (A.a)_{88}}{2} \quad (6)$$

با طی مراحل شش گانه برای هر زیرمتغیر، جدول ۵ به دست می آید.

۷.۳. مقداردهی به مدل به دست آمده

با توجه به خروجی مراحل ۲.۶.۳ و ۳.۶.۳ برای داده های سال ۸۷ و نیمه اول سال ۸۸، تابع هدف رابطه ی ۵ محاسبه، و یک پایگاه داده ترکیبی از داده های هجده ماهه در نرم افزار Excel تشکیل شد (۶۳۲ داده در ۱۷۰۶۴ فیلد).

۸.۳. تشکیل درخت تصمیم و استخراج درخت بهینه ی پیش بینی

مشتریان

برای استخراج درخت بهینه ی پیش بینی می توان از دو رویکرد زیر استفاده کرد:

۱. مدل سازی براساس ارزش محاسبه شده ی هر یک از چهار متغیر MR_i (۰/۲۳)، CS_i (۰/۲۴)، MF (۰/۳۲) و KP_j ($\sum_{j=1}^{13} KP_j$) (۰/۲۱) در هر مراجعه؛

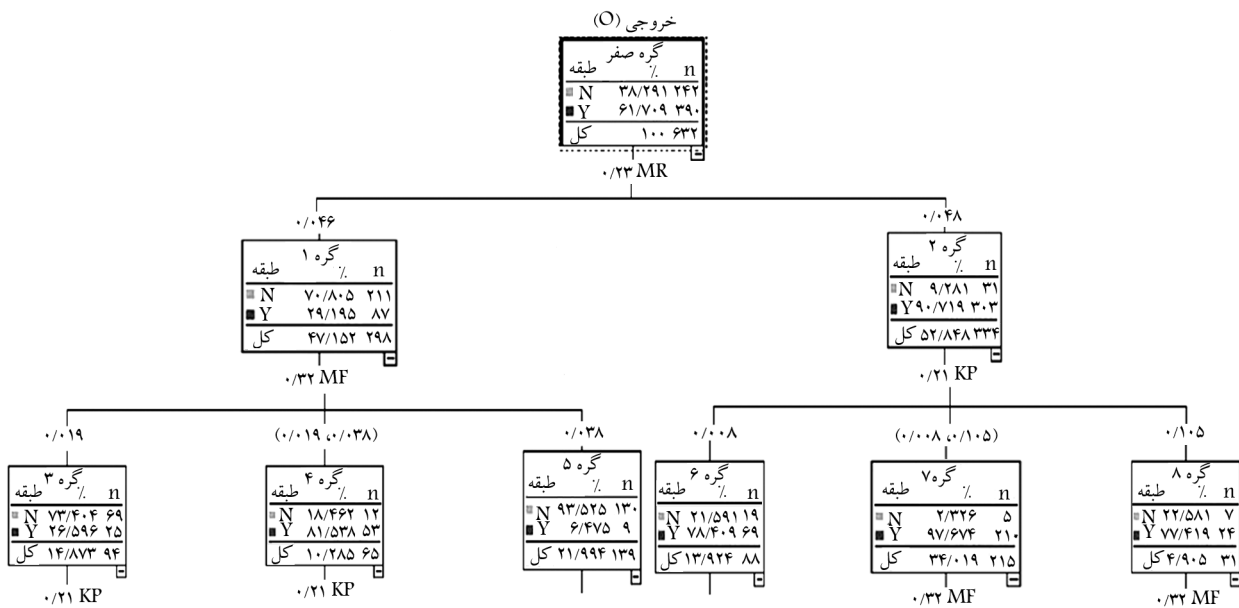
۲. مدل سازی براساس ارزش محاسبه شده ی Z در هر مراجعه.

۱.۸.۳. مدل سازی براساس ارزش محاسبه شده ی هر یک از چهار متغیر در هر مراجعه و نتیجه ی حاصل از آن

با قرار دادن داده های بند ۷.۳ براساس فرض ۱.۸.۳ در مدل سازی تشریح شده در بند ۱.۵.۳. مدل سازی CHAID با ۹۲/۲۵ درصد اهمیت، به بهترین حالت رسیدیم (شکل ۳ و ۴، جدول ۶).

۲.۸.۳. مدل سازی براساس ارزش محاسبه شده ی تابع هدف در هر مراجعه و نتیجه ی آن

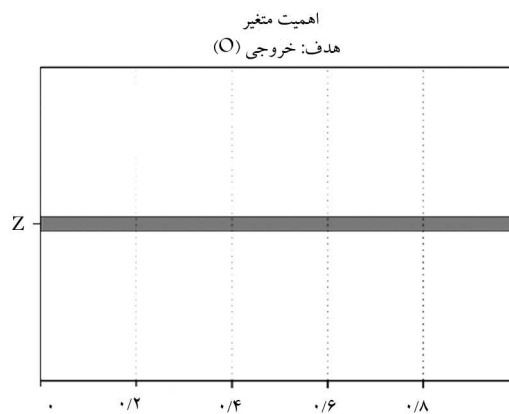
با قرار دادن داده های بند ۷.۳ براساس فرض ۲.۸.۳ در مدل سازی تشریح شده در بند ۱.۵.۳. مدل سازی $C50\%$ بهینه ترین حالت را ارائه می کند (شکل ۵ و ۶).



شکل ۴. قسمتی از درخت تصمیم متغیرهای با اهمیت CHAID براساس چهار متغیر.

جدول ۶. میزان اهمیت متغیرها در خروجی CHAID براساس چهار متغیر.

میزان اهمیت	متغیر
۰٫۶۱۵	نحوه‌ی درخواست محصول
۰٫۰۰۷	کارشناس فروش
۰٫۱۸۶	نحوه‌ی آشنایی
۰٫۱۹۲	نوع محصول



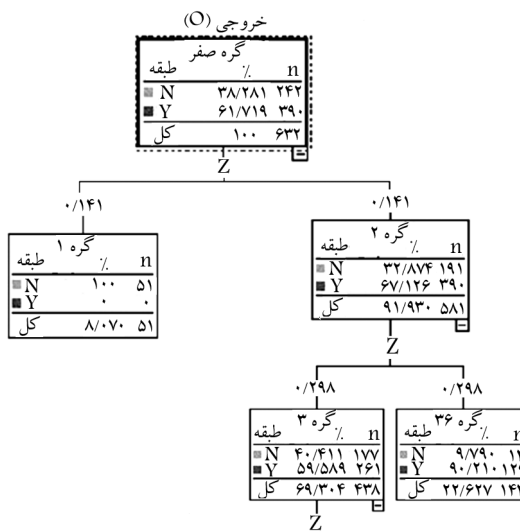
شکل ۵. اهمیت متغیرها در خروجی C5.0 براساس چهار متغیر.

۹.۳. پیاده‌سازی و ارزیابی مدل

خروجی اطلاعات بندهای ۱.۸.۳ و ۲.۸.۳. در اختیار پرسنل سازمان قرار گرفت و مقرر شد در بازه زمانی ۲۵ تا ۳۰ مهر سال ۸۸ از مدل استفاده کنند. نتایج حاصل از ۱۷ مراجعه حاکی از آن است که این مدل به درستی نتیجه‌ی ۱۶ مراجعه را پیش‌بینی کرده که معادل ۹۴٪ است (۱۱ مورد فروش و ۵ مورد انصراف، یکی از دلایل انصراف به دلیل عدم حضور کارشناس فروش شماره ۱ در سازمان برای ارتباط با مراجعه‌کننده، و یک مورد به علت ناتوانی کارشناس فروش از متقاعد کردن مشتری برای مراجعه‌ی حضوری بود).

۴. نتیجه‌گیری

تلفیق CRM و داده‌کاوی باعث پاسخ‌گویی مؤثر به نیاز مشتری، بهینه‌سازی بازدهی سرمایه و بهره‌وری نیروی انسانی، ارتقاء کیفیت در محصولات و در نهایت موجب پاسخ‌گویی سریع به تغییرات محیطی می‌شود. از این رو شرکت‌هایی که بقاء خویش را در بازار شدیداً رقابتی امروز تعقیب می‌کنند و به دنبال کسب مزیت رقابتی و افزایش سهم خود در آن هستند، لازم است به تغییر نگرش از محصول‌محوری به مشتری‌محوری، جمع‌آوری داده‌های مشتریان و تشکیل انبار داده‌ی مشتریان، آموزش کارکنان درخصوص استفاده از سیستم‌های سخت‌افزاری و نرم‌افزاری و استفاده‌ی بهینه از داده‌ها، فراهم آوردن امکانات فنی مورد نیاز، شناخت نیازهای گروه‌های



شکل ۶. قسمتی از درخت تصمیم متغیرهای با اهمیت C5.0 براساس متغیر Z.

استفاده از اطلاعات خروجی بند ۱.۸.۳. تمایل بیشتری از خود نشان می‌دهد. کاربرد دیگر درخت تصمیم «استخراج قانون» است (پرسنل فروش در تماس تلفنی با مراجعه‌کننده برای ایجاد تمایل دروی به منظور بازدید از نمایشگاه سازمان تلاش کنند؛ کارشناس فروش شماره ۱ را به مراجعه‌کنندگان قبلی، و کارشناس فروش شماره ۲ را به مراجعه‌کنندگان جدید تخصیص دهیم؛ بر مراجعه‌کنندگانی که از روش‌های F_1 ، F_2 ، F_3 و F_4 مراجعه می‌کنند سرمایه‌گذاری بیشتری انجام شود؛ مسئول ارتباط در نحوه تخصیص مراجعه‌کنندگان به کارشناسان فروش از درخت تصمیم پیروی کند؛ با توجه به این که در سید خرید مشتریان برخی از محصولات به محصولات دیگر وابسته‌اند، توجه کارشناسان فروش به قوانین وابستگی محصولات ضروری است).

مختلف مشتریان و ارتقاء کیفیت محصولات و خدمات مورد نظر آن‌ها توجه ویژه داشته باشند.

در سازمان مورد تحقیق، درخت تصمیم ابزار مناسبی برای شناسایی و دسته‌بندی عوامل برای تبدیل مشتریان بالقوه به بالفعل است. متغیرهای «نحوه آشنایی»، «نوع محصول»، «کارشناس فروش»، «نحوه درخواست» چهار متغیری هستند که تأثیر بیشتری در تبدیل مشتریان بالقوه به بالفعل دارند. هر چه معیارهای بیشتری در ساخت درخت تصمیم داشته باشیم، شناسایی مشتریان راحت‌تر صورت می‌گیرد. درختی که بر اساس الگوریتم $C5.0$ به دست آمد، به نتایج میدانی بند ۱.۵.۳. نزدیک‌تر است و بهتر عمل می‌کند، در پیاده‌سازی مدل پیش‌بینی نهایی کارکنان در

پانویس‌ها

1. information & communication technology
 2. decision trees
 3. artificial neural networks
 4. genetic algorithms
 5. association rules
 6. knowledge discovery database
 7. fuzzy clustering
۸. در اینجا دو بازه زمانی سال ۸۷ و نیمه‌ی اول سال ۸۸ مدنظر است.

منابع (References)

1. Cheng, C.H. and Chen, Y.S. "Classifying the segmentation of customer value via RFM model and RS theory", *Expert System with Applications*, **36**, pp. 4176-4184 (2008).
2. Yong, S. and Yoonseong, K. "Searching customer patterns of mobile service using clustering and quantitative association rule", *Expert system for application*, **34**, pp. 1070-1077 (2008).
3. Rygielski, C., Wang, J. and Yen, C. "Data mining techniques for customer relation management", *Technology in Society*, **24**, pp. 483-502 (2002).
4. Stefanou, C. and Sarmaniotis, C. "CRM and customer-centric Knowledge management: an empirical research", *Technology in Society*, **9**, pp. 917-634 (2004).
5. Lee, J.H. and Park, S.C. "Intelligent profitable customers segmentation system based on business intelligence tools", *Expert System for Application*, **29**, pp. 145-152 (2005).
6. Baensens, B., Verstreeten, G. and Poel, D. "Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers", *European Journal of Operation Research*, **156**, pp. 508-523 (2004).
7. Hwang, H., Jung, T. and Suh, E. "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry", *Expert Systems with Applications*, pp. 181-188 (2004).
8. Ngai, E. and Li, X. "Application of data mining techniques in customer relation management: A literature review and classification", *Expert Systems with Application*, **36**, pp. 2592-2602 (2008).
9. Korach, E. and Stern, M. "The complete optimal stars-clustering-tree problem", *Discrete Applied Mathematics*, **156**, pp. 444-450 (2008).
10. Mahdavi, I., Shirazi, B. and Cho, N. "Designing evolving user profile in e-CRM with dynamic clustering of Web documents", *Data & Knowledge Engineering*, **65**, pp. 355-372 (2008).

پرسش‌نامه‌ی تحقیقاتی شماره ۱

از آنجا که تأمین‌کنندگان و فروشندگان با شناخت مشتریان بالقوه می‌توانند ضمن سرمایه‌گذاری بر روی آن‌ها، با فعالیت‌هایی چون بازاریابی و تبلیغات آن‌ها را جذب و به مشتریان بالفعل خود تبدیل کنند، لزوم شناخت این دسته از مشتریان بالقوه ضروری به نظر می‌رسد. به همین دلیل شناسایی معیارهای لازم برای شناخت مشتریان بالقوه هدف تحقیق ماست. از این رو در این پرسش‌نامه درصدد محاسبه‌ی میزان اهمیت چند معیار اولیه‌ی هستیم که می‌تواند ما را در شناخت این مشتریان یاری کند. معیارهای دیگری که مدنظر شماست قید فرمایید. شما باید با توجه به میزان اهمیت هر معیار یکی از گزینه‌های «بی‌اهمیت»، «کمی مهم»، «مهم»، «خیلی مهم» را انتخاب کنید.

مشخصات پرسش‌شونده			
نام:	نام خانوادگی:	تحصیلات:	رشته:
شغل:	تاریخ تحویل پرسش‌نامه:	تاریخ عودت پرسش‌نامه:	

معیارهای پیش‌بینی					
ردیف	نام معیارها	بی‌اهمیت	کمی مهم	مهم	خیلی مهم
۱	سایز کسب و کار				
۲	نحوه‌ی درخواست محصول				
۳	میزان تمایل به مشترک شدن				
۴	داشتن برنامه خرید				
۵	نتیجه مراجعه				
۶	رونق بازار				
۷	برند بودن محصول				
۸	هدف مشتری				
۹	سابقه خرید از سایر تأمین‌کنندگان				
۱۰	زمان انتظار				
۱۱	نحوه‌ی آشنایی				
۱۲	میزان وفاداری				
۱۳	میزان سرمایه				
۱۴	نوع کسب و کار				
۱۵	ظاهر مشتری				
۱۶	ماه خرید				
۱۷	میزان سرمایه				
۱۸	خرید قبلی				
۱۹	نوع محصولات				
۲۰	کارشناس فروش				
۲۱	کارشناس طراحی				

پرسش نامه تحقیقاتی شماره ۲

نام خانوادگی: نام: رشته: شغل:
 تحصیلات: تاریخ تحویل پرسش نامه: تاریخ عودت پرسش نامه:

شناسایی مشتریان بالقوه و بالفعل از اهمیت زیادی برای هر سازمان برخوردار است. سازمان ما درصدد شناسایی بهتر مشتریان خود است. برای پیش‌بینی مشتری و این که آیا مشتری بالقوه به یک مشتری بالفعل تبدیل می‌شود به یکسری معیار نیازمندیم. ما با مطالعات میدانی فاز ۱ و مشخص کردن معیارهای مهم‌تر به این نتیجه رسیدیم که معیارهای ساینز نحوه‌ی درخواست، کارشناس فروش، نحوه‌ی آشنایی و نوع محصول درخواستی از اهمیت بیشتری برخوردارند. هدف این پرسش‌نامه مقایسه‌ی میزان اهمیت هر کدام از این معیارها در مقایسه با یکدیگر است. تا وزن نهایی هر معیار در مدل نهایی به دست آید. شما ابتدا باید میزان اهمیت هر معیار را با استفاده از طیف ۱ تا ۹ لیکرت در جدول ۱ مشخص، و سپس با استفاده از مقایسات زوجی در جدول شماره ۲ این معیارها را با هم مقایسه کنید.

C۱: نحوه‌ی درخواست (نحوه‌ی درخواست یک مشتری می‌تواند به صورت حضوری یا تلفنی باشد).

C۲: کارشناس فروش (کارشناسان فروش دو نفر می‌باشند).

C۳: نحوه‌ی آشنایی (نحوه‌ی آشنایی می‌تواند به یکی از طرق: خرید قبلی، معرف، نمایشگاه بین المللی، کاتالوگ و CD، پرسنل، نامه‌های تبلیغاتی، سایت، بیلبورد، نمایندگی و رسانه باشد).

C۴: نوع محصول درخواستی شامل پارتیشن، فایل، میز مدیریت و... می‌باشد.

جدول ۱.ب

بی اهمیت = ۱	کمی مهم = ۳	نسبتاً مهم = ۵	مهم = ۷	بسیار مهم = ۹					
۱	۲	۳	۴	۵	۶	۷	۸	۹	C۱
۱	۲	۳	۴	۵	۶	۷	۸	۹	C۲
۱	۲	۳	۴	۵	۶	۷	۸	۹	C۳
۱	۲	۳	۴	۵	۶	۷	۸	۹	C۴

در هر یک از خانه‌های خالی جدول یکی از اعداد ۱ تا ۹ را به منزله اهمیت آن سطر مرقوم فرمایید.

جدول ۲.ب

معیارهای پیش‌بینی	نحوه‌ی درخواست	کارشناس فروش	نحوه‌ی آشنایی	نوع محصولات درخواستی
نحوه‌ی درخواست	۱			
کارشناس فروش		۱		
نحوه‌ی آشنایی			۱	
نوع محصولات درخواستی				۱

PRESENTING A MODEL BASED ON DATA MINING TO FORECAST CUSTOMER ATTRACTION USING DECISION TREE IN CUSTOMER RELATIONSHIP MANAGEMENT

A. Kazemi(corresponding author)

abkaazemi@qiau.ac.ir

M. E. Babaei

m.babaei@gmail.com

**Dept. of Industrial and Mechanical Engineering
Islamic Azad University**

Sharif Industrial Engineering and Management Journal

Volume 29, Issue 2, Page 15-23, Original Article

© Sharif University of Technology

- Received 19 January 2011; received in revised form 22 October 2011; accepted 31 December 2011.

Abstract

With the rapid change in the business competitive environment, enterprise resource integration and innovative issues of business operation have gradually become the most important issues for businesses. Furthermore, many organizations have implemented novel information technology and developed innovative application systems to enhance their competitive advantages. CRM systems can help organizations to gain potential new customers. Customer relationship management is a multi-perspective business paradigm which aims to maximize the benefits gained from relationships with customers. Today, in the quality-based and competitive world, which is known as knowledge time, customer attraction is very important. In line with the slogan, “the customer is always right”, customer relation management is at the core of an organizational strategy. It plays an important role in four aspects: customer identification, customer attraction, customer retaining, and customer satisfaction. By analysis of customer life cycles, commercial organizations have obtained increases in customer value. Data store and data mining tools, and other customer relation management methods, have provided new opportunities for business. Data mining (DM) methodology has been of tremendous assistance to researchers in extracting hidden knowledge and information inherited in their data. This paper, by the practical use of data mining in identification of potential customers, tries to help organizations to determine identification criteria of potential customers in the competitive environment of

their business. It also presents mechanisms for identification of potential customers who have the ability to become real customers. In this paper, using a decision tree tool, we identify main criteria and determine their importance levels. We also consider that each main criterion consists of several sub-criteria, and we determine their importance in turning potential customers into real. We allow organizations to sell the process to each attendant in a direction which results in attendant (future customer) purchases, considering the criteria and sub-criteria identified.

Key Words: Data mining (DM), decision tree (DT), customer relationship management (CRM).