

ارائه مدلی به منظور طبقه‌بندی مشتریان خدمات اینترنتی بر مبنای الگوریتم داده‌کاوی

حامد کاظمی‌پور*

استادیار دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران
hkazemipoor@yahoo.com

فرید نوروزی

دانشگاه آزاد اسلامی، واحد الکترونیکی، تهران، ایران
marketing_norози@yahoo.com

تاریخ دریافت: ۱۳۹۵/۰۹/۲۸

تاریخ اصلاحات: ۱۳۹۵/۱۰/۱۹

تاریخ پذیرش: ۱۳۹۵/۱۰/۲۸

چکیده

امروزه نقش مشتریان بجای پیروی از تولیدکنندگان به هدایت آنها مبدل گشته است به همین دلیل دسته‌بندی مشتریان در هدفمندسازی و سفارشی‌سازی خدمات و اولویت‌بندی محصولات شرکت‌ها بر مبنای میزان سودآوری کمک شایانی می‌کند. در بازار ارائه‌دهندگان خدمات اینترنتی، رقابتی زیادی با هم رقابت می‌کنند زیرا سرمایه‌گذاری بر روی خدمات ارتباطی و اینترنتی سودآور و مفید می‌باشد. لذا برای پیشرفت در بازار مذکور، ارائه سرویس‌های جدید و خلاقیت و نوآوری از موارد ضروری می‌باشد. شرکت‌های ارائه‌دهنده خدمات اینترنتی برای گرفتن حجم بیشتری از بازار باید با شناخت کافی از بازار و مشتریان، به حفظ مشتریان موجود و جذب مشتریان جدید بپردازند. چنین شرکت‌هایی می‌توانند با طبقه‌بندی و خوشه‌بندی مشتریان خود، ضمن شناسایی و حمایت از مشتریان فعال و سودرسان، مشتریان نامتعارف را از چرخه خدمات‌رسانی شرکت خارج نمایند. این تحقیق با استفاده از الگوریتم داده‌کاوی، ضمن کشف و شناسایی چنین مشتریانی، شرکت‌های ارائه‌دهنده خدمات اینترنتی را نیز به اهدافشان نزدیک‌تر می‌کند.

واژگان کلیدی

فناوری اطلاعات و ارتباطات (ICT)^۱؛ سرویس اینترنت پر سرعت؛ پهنای باند^۲؛ خوشه‌بندی^۳؛ رده‌بندی یا طبقه‌بندی یا کلاس‌بندی مشتریان^۴

تحقیقاتی IDC^۵ در این زمینه نشان می‌دهد که از این تعداد بیش از دو میلیارد نفر از ابزارهای موبایل برای دسترسی به اینترنت استفاده خواهند کرد. در این میان بعضی از کشورها با سرعت بیشتری رشد می‌کنند. چین، هندوستان و اندونزی از جمله کشورهای پیشرو در این زمینه هستند و تقریباً نیمی از جمعیت آنها تا پنج سال آینده به اینترنت دسترسی خواهند داشت. از طرفی آفریقا و خاورمیانه دو منطقه‌ای هستند که به ترتیب با ۲۷ و ۲۵ درصد رشد سالانه، به سرعت در حال توسعه هستند. در همین رابطه اپراتورهای کشورهای هند، بنگلادش، ایران، جزایر ایوری، اوگاندا و افغانستان تاکنون سریع‌ترین رشد را داشته‌اند. ایران به‌عنوان یکی از کشورهای مهم خاورمیانه با استقبال گسترده‌ای از خدمات اینترنتی روبرو بوده و ارائه‌دهندگان این خدمات در تکاپوی افزایش سهم بازار از طریق ارائه راه‌حلی جهت جذب مشتریان جدید هستند. در واقع دغدغه این شرکت‌ها هم برحفظ مشتریان فعلی و هم بر جذب مشتریان جدید و بالقوه است. بر این اساس در ایران ADSL^۶ فراگیرترین فناوری باند پهن است و از جمله دلایل گسترش این فناوری در ایران می‌توان به برخورداری کشور از شبکه زوج سیم مسی به‌عنوان اساس و پایه ارائه خدمات DSL، افزایش تقاضا از سوی بازار، پیشرفت‌های فناوری در حوزه تجهیزات مورد

۱- مقدمه

تغییر فضای کسب و کار در هزاره سوم و حرکت سازمان‌ها از محیط حمایتی به سمت محیط رقابتی، موجب توجه بیش از پیش مدیران شرکت‌های ارائه‌دهنده خدمات اینترنتی به حفظ مشتریان قبلی و تلاش در جهت جذب مشتریان جدید گشته است. آمارها حاکی از آن است که جمعیت جهان در ابتدای سال ۲۰۱۶، به حدود ۷/۳ میلیارد نفر رسیده [۱] و همچنین تعداد کاربران جهانی موبایل در همین سال نیز به ۷ میلیارد نفر افزایش پیدا می‌کند که بیش از نیمی از آنها ساکن منطقه آسیا هستند تا این قاره به قطب تلفن همراه جهان تبدیل شود. برآوردهای اتحادیه بین‌المللی مخابرات (ITU)^۵ نشان می‌دهد طبق پیش‌بینی‌های این سازمان پایان سال ۲۰۱۶ میلادی ضریب نفوذ تلفن همراه در جهان ۹۶ درصد خواهد بود که در این میان ضریب نفوذ تلفن همراه در کشورهای توسعه‌یافته ۱۲۸ درصد و در کشورهای در حال توسعه ۸۹ درصد است. علاوه بر این ۴۴ درصد از جمعیت کل جهان یعنی حدود ۳/۲ میلیارد نفر در سال ۲۰۱۶ به اینترنت دسترسی خواهند داشت. جدیدترین بررسی‌های انجام شده توسط مؤسسه

6. Internet Data Center
7. Asymmetric Digital Subscriber Line

* نویسنده مسئول
1. Information and Communications Technology
2. Bandwidth
3. Clustering
4. Classification
5. International Telecommunication Union

را فراهم می‌کند. در ایران اکثر ارگان‌ها و سازمان‌ها و مراکز بزرگ فرهنگی مانند دانشگاه‌ها دارای اتصال فیبرنوری هستند و استفاده از این فناوری برای مناطق مسکونی نیز شروع شده و در حال رشد است [۲]. در این پژوهش به دنبال پاسخ به این سؤالات هستیم که:

۱. چگونه مشتریان خدمات اینترنتی را خوشه‌بندی نماییم؟
 ۲. چگونه مشخصات خوشه را تعیین نماییم؟
 ۳. چگونه و با چه روش‌هایی می‌توان مشتریان را در خوشه‌های ایجاد شده طبقه‌بندی کرد؟
 ۴. چگونه گروه‌های نامتعارف مشتریان را مشخص نماییم؟
- در گذشته تحقیقات و پژوهش‌هایی پیرامون طبقه‌بندی و خوشه‌بندی مشتریان توسط برخی از شرکت‌ها و مؤسسات انجام گردیده که نمونه‌ای از آنها در جدول شماره ۱ بیان شده است.

جدول ۱- پیشینه تحقیق (پژوهش‌های سیستم توصیه‌گر پالایش مشارکتی)

محقق (سال)	موضوع	مدل / متغیرهای اصلی	جامعه آماری و نمونه	روش گردآوری و تحلیل داده	یافته‌های تحقیق
۱۳۹۳	ارائه سیستم پیشنهاددهنده رشته تحصیلی در حوزه آموزش الکترونیکی مبتنی بر تکنیک‌های داده‌کاوی (مورد مطالعه دانشکده مدیریت پردیس مرکزی و دانشکده مدیریت پردیس قم دانشگاه تهران)	سیستم توصیه‌گر پالایشگر مشارکتی با استفاده از ترکیب خوشه‌بندی و طبقه‌بندی	داده‌های مرکز آموزش الکترونیکی دانشگاه تهران	داده‌های دانشجویان الکترونیکی	اطلاعات فراگیران و رشته تحصیلی در هر خوشه نگاشت ایجاد کرده و در نتیجه به ارائه پیشنهاد رشته تحصیلی به فراگیران پرداخت
۱۳۹۱	ارائه مدلی برای طراحی و توسعه یک سیستم پیشنهاددهنده محصول بر مبنای بخش‌بندی مشتریان و با بهره‌گیری از تکنیک‌های داده‌کاوی	سیستم توصیه‌گر مبتنی بر خوشه‌بندی و سپس درک قواعد انجمنی	داده‌های سایت فروشگاه آنلاین	داده‌های خرید مشتریان الکترونیکی	ارائه سیستم توصیه‌گر برای خرید مشتریان در فروشگاه‌های آنلاین
۱۳۹۳	طراحی سیستم پیشنهاددهنده عطر براساس سلیقه مشتریان بر مبنای ترکیبی از روش‌های داده‌کاوی	سیستم توصیه‌گر مبتنی بر خوشه‌بندی و سپس طبقه‌بندی با روش آموزش شبکه عصبی	جمع‌آوری داده‌های ترکیبات ساختاری شماری از عطرهای موجود در بازار دنیا	داده‌های پایگاه‌های داده جهانی	ارائه سیستم توصیه‌گر توصیه‌گر عطر به مشتریان
۱۳۹۴	خوشه‌بندی و تحلیل رفتار مشتریان دریافت‌کننده خدمات اینترنت: مطالعه موردی در یکی از شرکت‌های ارائه‌دهنده خدمات اینترنت	خوشه‌بندی kmeans	اطلاعات مشتریان یک شرکت خدمات اینترنتی	داده مشتریان	خوشه‌بندی مشتریان

در اینجا به تعریف برخی واژگان و اصطلاحات فنی و تخصصی مربوط به این مقاله می‌پردازیم:

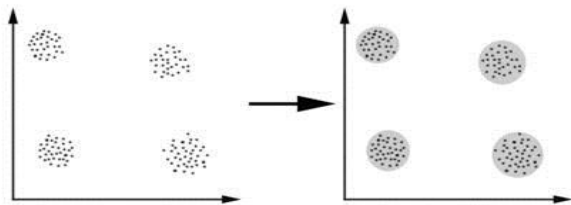
فن‌آوری اطلاعات و ارتباطات (ICT) (Information and Communication Technology): هرگونه تولید، خدمات، موضوع‌ها و عوامل مربوط به دریافت، گردآوری، ذخیره‌سازی، پردازش، بازیابی، انتقال محلی، کاربرد و مدیریت داده‌ها و اطلاعات در سیستم‌های کامپیوتری و تلفیق آن با مخابرات به صورت کاربردی و زمینه‌ای را فناوری اطلاعات و ارتباطات گویند.

سرویس اینترنت پر سرعت: خدمتی است که برخی شرکت‌ها بر مبنای آن وظیفه اتصال به اینترنت از طریق خطوط تلفن را انجام می‌دهند. **پهنای باند:** حداکثر انتقال داده در زمان است که واحد آن بایت بر ثانیه است.

نیاز در ارائه سرویس و در نتیجه کاهش قیمت نهایی برای مشترک، افزایش رقابت میان ارائه‌دهندگان سرویس و کاهش تعرفه‌ها اشاره کرد. بعد از ADSL بیش‌ترین سهم از آن Wimax^A است. این فناوری در بسیاری از کشورها با استقبال چندانی روبرو نشده است اما در کشورهای منطقه خاورمیانه به دلیل کمبود زیرساخت‌های شبکه سایر فناوری‌ها، رشد، این فناوری فراتر از سایر مناطق بوده است. در ایران نیز با شروع به کار اپراتورهای Wimax در انتهای سال ۱۳۸۹، روند افزایش تعداد مشترکین باند وسیع ثابت رشد قابل توجهی داشته است. از جمله سایر فناوری‌های باند وسیع ثابت، فناوری اینترنت ماهواره‌ای و فناوری اینترنت بی‌سیم است که کمتر مورد استقبال خانواده‌ها در ایران قرار گرفته و بیشتر توسط مشترکین تجاری مورد استفاده قرار می‌گیرد. فیبر نوری نیز از جمله فناوری‌های باند وسیع ثابت است که در سال‌های اخیر بسیار مورد استقبال قرار گرفته و با دارا بودن قابلیت ارائه پهنای باند بالا، امکان ارائه خدمات مختلف به کاربران

مدیران با استفاده از نتایج به دست آمده در این تحقیقات، توانستند به تصمیم‌گیری درست پیرامون موضوع ارائه خدمات چنین شرکت‌هایی بپردازند. از جمله می‌توان به مؤسسات مالی و بانک‌ها اشاره نمود که این سیستم به بانک‌ها کمک می‌کند تا با شناسایی انواع مشتریان تسهیلاتی خود، تصمیم‌های مرتبط با هر گروه را به منظور ایجاد حداکثر سود و کاهش هزینه‌های وصول مطالبات اتخاذ نمایند و رویکرد متناسبی را برای هر مشتری در پیش گیرند [۳] و [۴]. به نظر می‌رسد این پژوهش بتواند ضمن کشف و شناسایی مشتریان خوب شرکت‌های ارائه‌دهنده خدمات اینترنتی، کمک شایانی به مدیران در تصمیم‌گیری و حمایت از مشتریان، همچنین نزدیک کردن شرکت به اهداف بلندمدت بنماید.

بدون نظارت در نظر گرفت. به‌عنوان یک تعریف ساده، خوشه‌بندی فرایند سازماندهی کردن تعدادی شی در قالب گروه‌هایی است که اعضا آن به نحوی با یکدیگر مشابهت داشته باشند [۵]. خوشه‌بندی فازی یکی از روش‌های متداول داده‌کاوی است. فازی بودن این الگوریتم خوشه‌بندی سبب می‌شود در مواقعی که نتوان مرز مشخص و معلومی بین خوشه‌ها اعمال نمود، یک داده در یک زمان متعلق به چندین زیرمجموعه باشد. هدف از خوشه‌بندی، دسته‌بندی اشیا براساس میزان مشابهت بین آنها و سازماندهی گروهی خوشه‌هاست [۶]. اصلی‌ترین ظرفیت خوشه‌بندی، شناسایی ساختار نهفته در داده‌هاست. در شکل ۱ نمونه‌ای از اعمال خوشه‌بندی روی مجموعه‌ای از داده‌ها مشخص شده است که از معیار فاصله (Distance) به‌عنوان عدم شباهت (Dissimilarity) بین داده‌ها استفاده شده است [۷].



شکل ۱- اعمال خوشه‌بندی روی مجموعه‌ای از داده‌ها

تکنیک‌های خوشه‌بندی از طریق شناسایی گروه‌هایی از کاربران با اولویت‌های مشابه و تقسیم‌بندی گروه‌هایی با اولویت‌های بسیار متفاوت کار می‌کنند. برخلاف طبقه‌بندی، در خوشه‌بندی برچسب کلاس هر گروه نامعلوم است. در مقابل، طبقه‌بندی به صورت تخصیص داده به گروه‌هایی تعریف شده است. به‌طور خلاصه، یک روش خوشه‌بندی خوب، خوشه‌هایی را با کیفیت بالا تولید می‌کند، به‌طوری‌که درون هر کلاس بیشترین شباهت و بین کلاس‌ها کمترین شباهت وجود داشته باشد. در نهایت، خوب بودن یک خوشه، وابسته به نظر کاربر است [۱۲].

۲-۲-۱- روش‌های خوشه‌بندی

خوشه‌بندی را می‌توان از چندین جنبه تقسیم‌بندی کرد:

۲-۲-۱-۱- خوشه‌بندی انحصاری (Exclusive Hard Clustering) و

خوشه‌بندی با همپوشی (Overlapping Soft Clustering)

در روش خوشه‌بندی انحصاری پس از خوشه‌بندی، هر داده دقیقاً به یک خوشه تعلق می‌گیرد مانند روش خوشه‌بندی K-Means. ولی در خوشه‌بندی با همپوشی پس از خوشه‌بندی به هر داده یک درجه تعلق به ازاء هر خوشه نسبت داده می‌شود. به عبارتی یک داده می‌تواند با نسبت‌های متفاوتی به چندین خوشه تعلق داشته باشد. نمونه‌ای از آن خوشه‌بندی فازی است.

خوشه‌بندی: خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت می‌باشد و فرایند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود. تقسیم خوشه‌ها را به چند روش می‌توان انجام داد که از جمله معروف‌ترین آنها می‌توان به خوشه‌بندی انحصاری (Exclusive Hard Clustering)، خوشه‌بندی با همپوشی (Overlapping Soft Clustering)، خوشه‌بندی سلسله مراتبی (Hierarchical) و خوشه‌بندی مسطح (Flat) اشاره نمود.

رده‌بندی یا طبقه‌بندی یا کلاس‌بندی: فرایندی برای پیدا کردن مدلی است که مفاهیم یا کلاس‌های داده‌ای را تعریف می‌نماید و متمایز می‌کند، با این هدف که بتوان از این مدل برای پیش‌بینی کلاس یا اشیای که برچسب کلاس آنها (متغیر هدف) ناشناخته می‌باشد استفاده نمود. این مدل استخراج شده بر پایه آنالیز مجموعه داده‌های آموزشی ساخته می‌شود.

۲- چهارچوب‌ها و مبانی نظری

۲-۱-۲- داده‌کاوی

داده‌کاوی به استخراج دانش از مقادیر زیادی از داده‌ها، اشاره می‌کند. داده‌ها ممکن است شامل داده‌های مکانی، داده‌های چند رسانه‌ای، داده‌های سری زمانی، داده متنی و داده‌های وب باشند. داده‌کاوی شامل فرایند استخراج الگوهای جالب، کوچک اما با اهمیت، ضمنی، ناشناخته و بالقوه مفید و یا دانش از مقدار بسیار عظیمی از داده‌ها می‌باشد. این فرایند در بردارنده مجموعه‌ای از فعالیت‌ها برای پیدا کردن الگوهای جدید، ضمنی و یا غیرمنتظره در داده‌ها، یا الگوهای غیرمعمول در آنها است. داده‌کاوی می‌تواند به سازمان‌ها در راستای درک بهتر کسب و کار خود، ارائه خدمات بهتر به مشتریان و افزایش اثربخشی سازمان در بلندمدت کمک کند [۱۱]. تعریف ساده داده‌کاوی به صورت کاربرد الگوریتم‌های خاص برای کشف اطلاعات مفید از بین مقادیر زیادی داده است و هدف آن کشف دانش جالب توجه از یک پایگاه داده، یک انبار داده و یا برخی دیگر از واحدهای بزرگ ذخیره‌سازی اطلاعات است. از دیدگاه فنی، داده‌کاوی یک روش جمع‌آوری و فهرست‌بندی اطلاعات را ترکیب می‌کند و سپس به تولید دانش شبیه به قانون از بین مقادیر زیاد داده می‌پردازد. رایج‌ترین توابع مدل‌سازی در الگوریتم‌های داده‌کاوی عبارتند از: طبقه‌بندی، رگرسیون، خوشه‌بندی، قوانین انجمنی، ایجاد قانون، خلاصه‌سازی، مدل‌سازی وابستگی (تحلیل لینک) و تحلیل دنباله (کشف توالی).

۲-۲- خوشه‌بندی

خوشه به مجموعه‌ای از داده‌ها گفته می‌شود که به هم شباهت داشته باشند. در خوشه‌بندی، سعی می‌شود تا داده‌ها به خوشه‌هایی تقسیم شوند که شباهت بین داده‌های درون هر خوشه بیشینه و شباهت بین خوشه‌های کمینه شود. خوشه‌بندی را می‌توان به‌عنوان مهم‌ترین مسأله در یادگیری

- نسبت‌دادن 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 این روال می‌توان در هر تکرار، با میانگین‌گیری از داده‌ها مراکز جدیدی برای آنها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدیدی نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود.

الگوریتم زیر الگوریتم پایه برای این روش محسوب می‌شود:

۱. در ابتدا K نقطه به‌عنوان نقاط مراکز خوشه‌ها انتخاب می‌شوند.
۲. هر 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 به‌عنوان مرکز محاسبه می‌شود. (میانگین نقاط متعلق به هر خوشه)
۴. مراحل ۲ و ۳ تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود تکرار می‌شوند [۹].

۳-۲-۳ طبقه‌بندی

طبقه‌بندی یکی از رایج‌ترین روش‌های کاربردی داده‌کاوی است که مجموعه‌ای از نمونه‌های از پیش طبقه‌بندی شده را برای توسعه جامعه بزرگ مورد استفاده قرار می‌دهد. برنامه‌های کاربردی تشخیص تقلب و ریسک اعتباری، به خوبی یا این نوع از تجزیه و تحلیل تناسب دارند.

۲-۳-۱-۱ صحت‌سنجی طبقه‌بندی

الگوریتم‌های زیادی برای صحت‌سنجی طبقه‌بندی وجود دارند روش ریزابلاس یکی از این روش‌هاست که ابتدا داده‌هایی را به‌عنوان داده‌های تست انتخاب می‌کند و آنقدر این فرایند را ادامه می‌دهد تا داده‌ها و داده‌های تطابق به هیچ‌وجه مستقل نباشند [۱۴].

۲-۳-۲ خوشه‌بندی در مقابل طبقه‌بندی

در طبقه‌بندی هر داده به یک طبقه (کلاس) از پیش مشخص شده تخصیص می‌یابد ولی در خوشه‌بندی هیچ اطلاعی از کلاس‌های موجود درون داده‌ها وجود ندارد و به عبارتی خود خوشه‌ها نیز از داده‌ها استخراج می‌شوند. در شکل ۲ تفاوت بین خوشه‌بندی و طبقه‌بندی بهتر نشان داده شده است. در طبقه‌بندی با استفاده از یک سری اطلاعات اولیه داده‌ها به دسته‌های معلومی نسبت داده می‌شوند. در خوشه‌بندی داده‌ها با توجه به الگوریتم انتخاب شده به خوشه‌هایی نسبت داده می‌شوند [۱۳].

۲-۱-۲-۲ خوشه‌بندی سلسله مراتبی (Hierarchical) و خوشه‌بندی مسطح (Flat)

در روش خوشه‌بندی سلسله مراتبی، به خوشه‌های نهایی براساس میزان عمومیت آنها ساختاری سلسله مراتبی نسبت داده می‌شود. مانند روش Single Link، ولی در خوشه‌بندی مسطح تمامی خوشه‌های نهایی دارای یک میزان عمومیت هستند مانند K-Means. به ساختار سلسله مراتبی حاصل از روش‌های خوشه‌بندی سلسله مراتبی دندوگرام گفته می‌شود. با توجه با اینکه روش‌های خوشه‌بندی سلسله مراتبی اطلاعات بیشتر و دقیق‌تری تولید می‌کنند برای تحلیل داده‌های با جزئیات پیشنهاد می‌شوند ولی از طرفی چون پیچیدگی محاسباتی بالایی دارند برای مجموعه داده‌های بزرگ، روش‌های خوشه‌بندی مسطح پیشنهاد می‌شوند [۱۳].

۲-۱-۲-۳ روش‌های خوشه‌بندی غیر سلسله مراتبی

روش‌های غیر سلسله مراتبی برای دسته‌بندی کردن اقلام به جای متغیرها به مجموعه‌ای از k خوشه طراحی شده است. یکی از روش‌های خوشه‌بندی غیر سلسله مراتبی، روش تقسیم‌بندی k -Medoids است. الگوریتم به کاررفته در این روش، براساس جست و جو برای یافتن k عنصر نماینده از بین مجموعه داده‌هاست که این k ، عنصر نماینده را Medoids می‌نامند. در حقیقت میدوئیدها نمایندگان خوشه‌ها هستند که باید بتوانند به خوبی ساختار داده‌ها را نشان دهند. نمایندگان هر خوشه کمترین مقدار متوسط عدم تشابه را با بقیه اعضای خوشه دارند. پس از مشخص شدن میدوئیدها k خوشه به این صورت ساخته می‌شود که هر عنصر به خوشه‌ای می‌رود که کم‌ترین فاصله را با نماینده آن خوشه داشته باشد.

پس از اجرای روش‌های خوشه‌بندی و حاصل شدن خوشه‌ها، می‌توان یک نمایش گرافیکی برای خوشه‌بندی به روش PAM داشت که آن را Silhouette یا سایه‌نما می‌نامیم. می‌توان به ازای هر خوشه، نمودار سایه نما داشت و با کنارهم قرار دادن آن‌ها کیفیت خوشه‌بندی‌ها را با یکدیگر مقایسه کرد. محور افقی این نمودار نشان‌دهنده عرض خوشه‌ها و محور عمودی، نحوه تعلق گرفتن عناصر به خوشه‌ها را نشان می‌دهد [۸].

۲-۱-۲-۴ روش خوشه‌بندی انحصاری و مسطح و غیر سلسله مراتبی

K-Means

این روش انحصاری و مسطح علی‌رغم سادگی یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر محسوب شده و الگوریتمی بسیار ساده، قابل فهم و بطور منطقی قابل مقیاس‌بندی است و می‌توان آن را به سادگی اصلاح کرد تا با سناریوهای مختلف مانند یادگیری شبه مشاوره یا داده‌های جاری سروکار داشته باشد. این الگوریتم به شکل‌های مختلفی بیان شده است ولی همه آنها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

- بدست آوردن نقاطی به‌عنوان مراکز خوشه‌ها این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.

و ویژگی‌های خوشه‌ها مشخص می‌گردد و با روش‌های گوناگون طبقه‌بندی انجام می‌گیرد. در پایان صحت طبقه‌بندی‌ها ارزیابی شده و کم‌ترین خطا به‌عنوان روش بهینه انتخاب می‌گردد.

۴- توصیف متغیرها

برای ورود و پاکسازی داده‌ها مراحل زیر در نظر گرفته شد:

- جنسیت مشتری: داده‌های جنسیت به ترتیب ۱ مرد و ۲ زن کدگذاری گردید.
- سن مشتری: سن مشتریان بدون تغییر وارد سامانه گردید.
- حجم مورد استفاده: براساس مصرف در ۳ گروه مصرفی از زیر ۱۰ گیگابایت تا بالای ۳۵ گیگابایت تقسیم‌بندی گردید که مرکز هر طبقه وارد نرم‌افزار گردید.
- تحصیلات مشتری بر مبنای:
 - ✓ دانش آموز و زیر دیپلم تا دارای مدرک دیپلم کد ۱
 - ✓ از دانشجویان ترم اول دانشگاه تا مشتریان دارای مدارک کاردانی و کارشناسی کد ۲
 - ✓ از دانشجویان ترم اول کارشناسی‌ارشد تا مشتریان دارای مدارک دکترا و بالاتر کد ۳
- تعداد سال استفاده از خدمت: بدون تغییر وارد نرم‌افزار گردید.

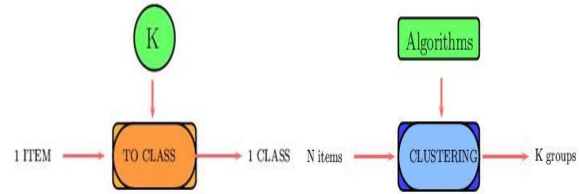
با توجه به متنوع بودن نوع فیلدها مثل مبلغ، تعداد و ... داده‌ها [۱۰] قبل از ورود به نرم‌افزار با استفاده از روش Z-score نرمال‌سازی گردید تا مقادیر به صورت یکپارچه مورد تحلیل قرار گیرند. سپس برای تعیین روش و تعداد خوشه‌ها، از ابزار ارزیابی خوشه در نرم‌افزار متلب که امکان بررسی روش‌های خوشه‌بندی مختلف شامل خوشه‌بندی سلسله مراتبی و میانگین k را دارد استفاده گردید و روش‌های خوشه‌بندی kmeans و سلسله مراتبی با روش سیلوئت سنجیده شده و نتایج مشابه در روش kmeans و Ward بدست آمد. اما همانطوریکه قبلاً توضیح داده شد با توجه به سرعت بهتر در kmeans همچنین تعداد داده‌های بسیار بالا و به منظور صرفه‌جویی در وقت و هزینه، این روش به‌عنوان روش اصلی انتخاب گردید.

evakmeans.CriterionValues							
84	85	86	87	88	89	90	
0.9983	0.9986	0.9990	0.9997	0.9999	1	1	

شکل ۳- نتیجه سیلوئت روش kmeans

evalinkage.CriterionValues							
83	84	85	86	87	88	89	90
0.9978	0.9984	0.9990	0.9993	0.9997	0.9999	1	NaN

شکل ۴- نتایج روش Ward



شکل ۲- تفاوت بین خوشه‌بندی و طبقه‌بندی

۳- روش تمقیق

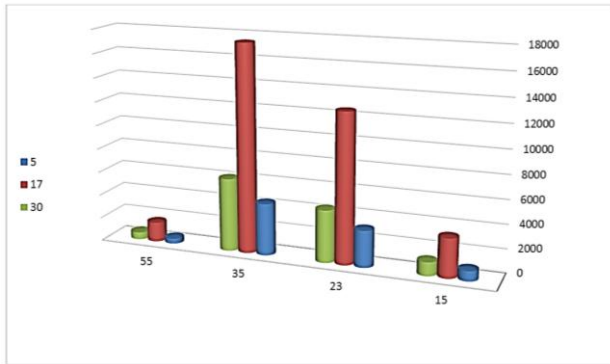
روش داده‌کاوی برخلاف روش‌های آماره براساس اثبات فرضیه‌ها عمل نمی‌کند و این پژوهش فرضیه آماری ندارد و مفروض است که روش داده‌کاوی می‌تواند معیار مناسبی برای شناسایی خوشه‌های داده‌ها باشد. نوع تحقیق کاربردی است، زیرا براساس تعریف در تحقیق کاربردی هدف، به دست آوردن درک یا دانش لازم برای تعیین ابزاری است که به‌وسیله آن نیازی مشخص و شناخته شده برطرف گردد در این نوع تحقیقات هدف کشف دانش تازه‌ای است که کاربرد مشخصی را درباره فرایندی در واقعیت دنبال می‌کند. به عبارت دقیق‌تر تحقیق کاربردی تلاشی برای پاسخ‌دادن به یک معضل و مشکل علمی است که در دنیای واقعی وجود دارد. این پژوهش به منظور هدف کاربردی و از منظر روش به دلیل بررسی وضعیت موجود و استفاده از داده‌های حقیقی و توصیف آن‌ها توصیفی و از منظر نوع داده‌ها کمی می‌باشد.

مراحل تحقیق عبارتند از:

- تحلیل مسأله موجود: شناسایی داده‌ها و کدگذاری و پاک‌سازی جهت استفاده در نرم‌افزار
- پیشنهاد روش‌ها و سیستم‌های مورد نیاز: مطالعه تحقیقات و پیشینه فعالیت‌های عملیاتی و تحقیقاتی و آزمودن روش‌های مختلف مبتنی بر خوشه‌بندی و طبقه‌بندی
- اجرای سیستم‌ها: بررسی نتایج و مشخصات هر خوشه
- ارزیابی و اعتبارسنجی: تأیید صحت طبقه‌بندی و خوشه‌بندی به منظور اجرای این پژوهش، داده‌های مشتریان کلیه شرکت‌های ارائه‌دهنده خدمات اینترنتی کشور [۱۰] که شامل بر ۴۲ میلیون رکورد بوده و نسبتی از آن یعنی نسبت یک به ده هزار که مشتمل بر ۴۲۰۰ رکورد در نرم‌افزار متلب وارد شد. سپس فیلدهای مناسب که ویژگی‌های لازم از جمله یکتا بودن را داشتند جهت اجرای خوشه‌بندی به شرح زیر انتخاب گردید:

- ✓ جنس مشتری
- ✓ سن مشتری
- ✓ تحصیلات مشتری
- ✓ شروع استفاده از سیستم
- ✓ میزان حجم مصرفی

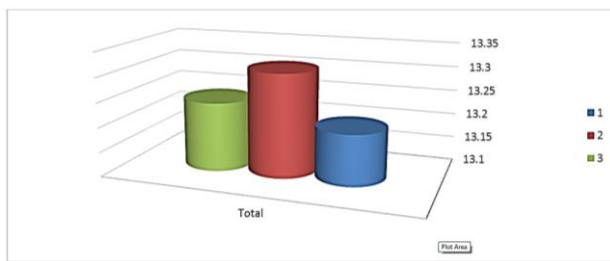
جمع‌آوری و پاکسازی شده و فیلدهای نامناسب حذف می‌گردند. پس از آن با استفاده از الگوریتم‌های ارزیابی خوشه‌ها مانند سیلوئت نوع خوشه‌بندی و تعداد خوشه‌های مناسب انتخاب و نتایج تجزیه تحلیل شده



نمودار ۴- مقایسه سن و حجم مصرفی

۴-۵- نسبت تحصیلات به حجم مصرف

بر اساس این نمودار بیش‌ترین حجم مصرفی متعلق به مشتریان از دانشجویان ترم اول دانشگاه تا مشتریان دارای مدارک کاردانی و کارشناسی و کمترین حجم مصرف نیز متعلق به مشتریان دانش‌آموز و زیر دیپلم و یا دارای مدرک دیپلم می‌باشد و مشتریان از دانشجوی ترم اول کارشناسی‌ارشد تا مشتریان دارای مدارک بالاتر نیز در میانه میدان قرار دارند.



نمودار ۵- نسبت تحصیلات به مصرف

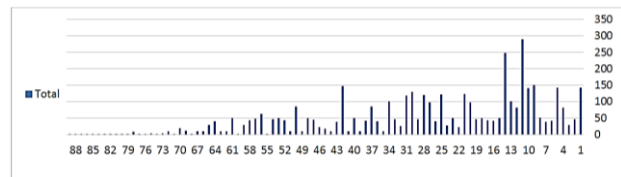
۴-۵- نتایج و تحلیل

در این پژوهش مشتریان مورد نظر شناسایی و خوشه‌بندی شدند و با طبقه‌بندی، امکان پیش‌بینی داده‌های جدید بوجود آمد. این سیستم آمادگی پیش‌بینی خوشه‌ها را داشته و امکان کمک به مدیران مربوط جهت اتخاذ تصمیم‌های فوری برای مشتریان خدمات اینترنتی را دارد. همچنین می‌تواند نوع مشتری را پیش‌بینی کرده و سایر روش‌های داده‌کاوی را نیز مورد آزمون و بررسی قرار دهد لذا به منظور طبقه‌بندی مشتریان خدمات اینترنتی پس از جمع‌آوری داده‌ها با استفاده از روش‌های ارزیابی روش kmeans با ۸۹ خوشه به‌عنوان دقیق‌ترین روش خوشه‌بندی انتخاب گردید. سپس به منظور ایجاد امکان پیش‌بینی خوشه مشتریان جدید، طبقه‌بندی داده‌ها با روش‌های گوناگون انجام شد و روش نزدیک‌ترین همسایه انتخاب گردید. خوشه‌های ۸۹ گانه از منظر ویژگی‌های مختلف مقایسه گردیده و با توجه به اینکه صفاتی مانند تحصیلات و مصرف مشتریان، این خوشه‌ها را از هم متمایز می‌نماید بر اساس این ویژگی‌ها، می‌توان سیاست‌های خاصی را به‌منظور بهبود ارائه خدمات به

همانطوریکه در شکل‌های ۳ و ۴ ملاحظه می‌شود در روش‌های kmeans و Ward با درجه سیلوئت ۱، بهترین تعداد خوشه ممکن ۸۹ می‌باشد.

۴-۱- تعداد اعضای هر خوشه

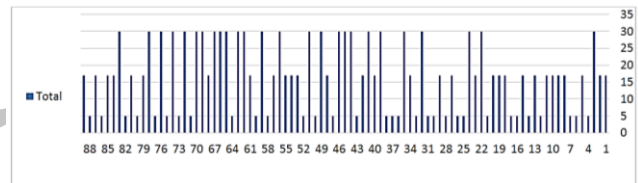
نمودار زیر مشخص‌کننده تعداد اعضای هر خوشه بوده و بر مبنای عدد خوشه به دست آمده، امکان انجام تحلیل‌های بسیاری وجود دارد که روی هر یک از فیلدهای ورودی داده و با استفاده از روش‌هایی مانند محاسبه میانگین، میانه، واریانس و انحراف‌معیار قابل بررسی است و همانطوریکه ملاحظه می‌شود در خوشه ۱۱ بیشترین تعداد مشتریان خدمات اینترنتی وجود دارد.



نمودار ۱- جمعیت در هر طبقه

۴-۲- میانگین مصرف هر طبقه

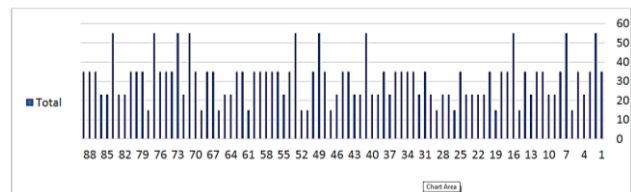
میانگین حجم مصرفی کاربران در هر یک از طبقات به شرح نمودار زیر بدست آمده است. بر این اساس میانگین به ترتیب ۵، ۱۷ و ۳۰ کیگابایت در نظر گرفته شده است.



نمودار ۲- میانگین حجم مصرفی در هر طبقه

۴-۳- میانگین سن هر طبقه

میانگین سن هر طبقه به شرح زیر به دست آمده است:



نمودار ۳- میانگین سن در هر طبقه

۴-۴- رابطه سن و حجم مصرف

بر اساس این نمودار می‌توان نتیجه گرفت که بطور متوسط در رده سنی ۴۰-۳۰ سال بیش‌ترین حجم مصرف اینترنت وجود دارد. در واقع در این رده سنی در حجم‌های مختلف ۵، ۱۷ و ۳۰ بالاترین تعداد مصرف‌کنندگان از خدمات اینترنتی وجود دارد. همانطوریکه در نمودار مشخص است رده سنی بالای ۵۰ سال نیز کم‌ترین استفاده را از خدمات اینترنتی دارند.

- Access Provider شرکت‌های ارائه‌دهنده خدمات عمومی انتقال داده ماهواره‌ای).
بهار ۱۳۹۵ و گزارشات عملکردی اپراتورها.
- 11- A Purpose of Data Mining in Banking Sector. International Journal 2015.
- 12- Chen C-C, Chen A-P. (2007). Using data mining technology to provide a service in the digital library. The Electronic Library ,711-724.
- 13- Keller F. (May 2013). Clustering. International Journal of Advanced Research in Computer Science and Software Engineering Computer University Saarlandes, Tutorial Slides. Vol. 3, Issue 5,861-864.
- 14- Edward B.classification of fault-prone software modules(DECEMBER2002),IEEE TRANSACTIONS ON RELIABILITY, VOL. 51, NO. 4, 455-462.
- 15- Mehta JS, Gawande A.(2015).A Purpose of Data Mining in Banking Sector. International Journal.
- 16- DeOña R & J(2015)Analysis of transit quality of service through segmentation and classification tree techniques. Transportmetrica A: Transport Science. 365-387.

مشتریان هر خوشه ارائه نمود. برای مثال بیش‌ترین مصرف‌کنندگان در خوشه ۱۱ شامل دانشجویان ترم اول دانشگاه تا مشتریان دارای مدارک کاردانی و کارشناسی هستند که ویژگی‌هایی مانند سن بالاتر از ۲۵ سال نیز دارند و این گروه می‌توانند مورد سیاست‌های تشویقی از قبیل بسته‌های رایگان اینترنتی و یا افزایش حجم مصرفی رایگان و یا برداشتن محدودیت زمانی قرار گیرند.

در پایان به محققان توصیه می‌گردد این پژوهش را با پیشنهادات زیر برای مطالعات آینده تکمیل نمایند:

- شناسایی فعالیت بازاریابی مناسب هر خوشه
- استفاده از روش‌های فازی
- کسب ویژگی‌های جدید در داده‌های در دسترس
- استفاده از داده‌های جدید شرکت‌های خدمات اینترنتی به‌منظور شناسایی الگوهای بیشتر
- طراحی نرم‌افزار جامع خوشه‌بندی مشتریان
- پیاده‌سازی سیستم و بررسی تأثیر استفاده از خوشه‌ها در فعالیت‌های شرکت‌ها و تأثیر آن بر حجم مصرف و مدت استفاده از خدمات
- استفاده از داده‌های ترکیبی نظیر داده‌های روان‌شناختی، رفتاری و یا جغرافیایی

۶- مراجع

- ۱- خبرگزاری اسپوتنیک روسیه "به نقل از دفتر سرشماری نفوس آمریکا"
- ۲- سازمان فناوری اطلاعات ایران. تهران. ۱۳۹۴
- ۳- زارع، فائزه: "طراحی سیستم توصیه‌گر به منظور بهینه‌سازی و مدیریت تسهیلات بانکی بر مبنای الگوریتم‌های خوشه‌بندی و طبقه‌بندی تسهیلات". مجله پژوهش‌های نوین در تصمیم‌گیری: دوره ۱، شماره ۲، (تابستان ۱۳۹۵). ۵۳-۷۶.
- ۴- نژاد فلاطوری مقدم، طه: "بخش‌بندی مشتریان بانک و تعیین استراتژی ارتباط با مشتری با استفاده از داده‌کاوی". پایان‌نامه کارشناسی‌ارشد، دانشکده اقتصاد و مدیریت، دانشگاه تربیت مدرس، ۱۳۹۰.
- ۵- طاهرپور، حمید و دیگران: "بررسی کارایی تولید محصولات زیربخش‌های زراعت و باغبانی در استان خراسان رضوی: کاربرد روش‌های ترکیبی شبکه عصبی مصنوعی الگوریتم و خوشه‌بندی فازی". مجله دانش کشاورزی و تولید پایدار: جلد ۲۰/۲ شماره ۴، ۱۳۸۹.
- ۶- تجریشی، مسعود: "پهنه‌بندی آلودگی رودخانه‌ها توسط تکنیک تحلیل طبقه‌بندی فازی". سازمان مدیریت منابع آب: نشریه ۲۶۱، ۱۳۸۱.
- ۷- مختاری حسن‌آباد، وحید: "خوشه‌بندی داده‌های جریانی با استفاده از موازی‌سازی الگوریتم‌های ترکیبی". پایان‌نامه کارشناسی‌ارشد. دانشگاه آزاد قزوین. زمستان ۸۶.
- ۸- علوی‌مجد، حمید و دیگران: "مقایسه نتایج خوشه‌بندی سلسله مراتبی و غیر سلسله مراتبی پروتئین‌های مرتبط با سرطان‌های مری، معده و کلون براساس تشابهات تفسیر هستی‌شناسی ژنی" مجله دانشور: دوره ۱۷، شماره ۸۸ (شهریور ۱۳۸۹). ۱۷-۳۰.
- ۹- یقینی، مسعود و دیگران: "یک روش ترکیبی خوشه‌بندی مبتنی بر الگوریتم ژنتیک با استفاده از عملگرهای جدید متغیر" نشریه بین‌المللی مهندسی صنایع و مدیریت تولید: دوره ۲۳، شماره ۱، (خرداد ۱۳۹۱). ۱۲۲-۱۲۸.
- ۱۰- داده‌های مرکز آمار ایران، سازمان تنظیم مقررات و ارتباطات رادیویی. سال ۱۳۹۴، شرکت‌های (Private Access Provider) PAP شرکت‌های ارائه‌دهنده خدمات عمومی انتقال داده کابلی اینترنت پرسرعت (ADSL)، شرکت‌های (SAP) Satellite