



به کارگیری تکنیک های داده کاوی در تحلیل رفتار کاربران اینترنت جهت بهبود در مدیریت سرویس دهی اینترنت استان اصفهان

حسن ضیافت^۵

^۵دانشگاه آزاد اسلامی واحد نطنز، گروه کامپیوتر، نطنز، ایران، ziafat@natanziau.ir

چکیده - با افزایش چشمگیر حجم اطلاعات و توسعه وب، نیاز به روش ها و تکنیک هایی که بتوانند امکان دستیابی کارا به داده ها و استخراج اطلاعات از آنها را فراهم کنند، بیش از پیش احساس می شود. در سال های اخیر تکنیک ها و الگوریتم های داده کاوی، محققان علوم مختلف را در رسیدن به اهداف فوق یاری رسانده است. وب کاوی یکی از زمینه های تحقیقاتی است که با به کارگیری تکنیک های داده کاوی به کشف و استخراج خودکار اطلاعات از استناد و سرویس های وب می پردازد. در واقع وب کاوی، فرآیند کشف اطلاعات و دانش ناشناخته و مفید از داده های وب می باشد. از آنجایی که گسترش اینترنت در زندگی امروزی به ابزاری برای رشد و توسعه کشورها تبدیل شده و میزان نفوذ آن در حوزه های مختلف از جمله بانکداری اینترنتی، آموزش مجازی، دولت الکترونیک و ... شاخصی برای رشد و توسعه کشورها در نظر گرفته می شود، می توان از تکنیک های داده کاوی در تحلیل نفوذ اینترنت در زندگی اجتماعی استفاده نمود. با توجه به ضرورت موضوع، طرح آمارگیری از کاربران اینترنت انجام شده است. با این وجود مطالعه ای در باره الگوهای رفتاری کاربران و ویژگی های آنها صورت نپذیرفته است و به علت اینکه روابط و پارامترهای بدست آمده بر حسب شرایط اجتماعی هر جامعه متفاوت خواهند بود، ضرورت انجام این مطالعه در ایران نیز حائز اهمیت است. در این مطالعه داده های بدست آمده از طرح آمارگیری از کاربران اینترنت در سطح استان اصفهان مورد بررسی قرار گرفته و با کمک تکنیک های داده کاوی، کاربران با توجه به سن، جنسیت، سطح تحصیلات، وضعیت اشتغال، محل دسترسی، نوع استفاده، نحوه دسترسی به اینترنت و هزینه های پرداختی برای آن، خوشبندی شده اند.

کلید واژه- داده کاوی، وب کاوی، اینترنت، خوشبندی، درخت تصمیم.

مفید را در میان حجم انبوه داده ها تشخیص دهند و یا اگر قادر

به این کار هم باشند، هزینه عملیات از نظر نیروی انسانی و مادی

- ۱ مقدمه

بسیار بالا است.

از سوی دیگر کاربران معمولاً فرضیه ای را مطرح می کنند و سپس بر اساس گزارشات مشاهده شده به اثبات یا رد فرضیه می پردازنند، در حالی که امروزه نیاز به روش هایی است که اصطلاحاً به کشف دانش بپردازند یعنی با کمترین دخالت کاربر و به صورت خودکار الگوها و رابطه های منطقی را بیان نمایند.

داده کاوی یکی از مهمترین این روشها است که به وسیله آن الگوهای مفید در داده ها با حداقل دخالت کاربران شناخته می شوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می

امروزه با گسترش سیستم های پایگاهی و حجم بالای داده های ذخیره شده در این سیستم ها، نیاز به ابزاری است تا بتوان داده های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد.

با استفاده از پرسش های ساده در SQL و ابزارهای گوناگون گزارش گیری معمولی، می توان اطلاعاتی را در اختیار کاربران قرار داد تا بتوانند به نتیجه گیری در مورد داده ها و روابط منطقی میان آنها بپردازند اما وقتی که حجم داده ها بالا باشد، کاربران هر چند زیر دست و با تجربه باشند نمی توانند الگوهای



طرح آمارگیری از کاربران اینترنت، برای اولین بار در دی ماه 5943 توسط مرکز آمار اجرا شد و اطلاعات کاربران را در دوازده ماه منتهی به زمان آمارگیری جمع‌آوری کرد. تا قبل از این آمارگیری، اطلاعات رسمی از این نرخ وجود نداشت و نرخ‌های اعلام شده برای ضریب نفوذ در سال‌های قبل بر اساس برآوردها و استفاده از فرمول‌های محاسباتی بوده است.

با این وجود مطالعات اندکی در خصوص ویژگی‌های کاربران اینترنت و الگوهای رفتاری آنها در ایران انجام شده است. در این مقاله، به تحلیل رفتار کاربران اینترنت در سطح استان اصفهان پرداخته‌ام و با استفاده از اطلاعات استخراج شده از آخرین طرح آمارگیری از کاربران اینترنت، کاربران با رویکرد خوش‌بندی به چهار خوش‌تقسیم شده‌اند، که افراد هر خوش‌هه دارای ویژگی‌های رفتاری شبیه به یکدیگر می‌باشند.

این مقاله شامل بخش‌های زیر است: بخش دوم، به بیان نحوه نمونه‌گیری پرداخته است و در بخش سوم متغیرهای مورد استفاده در این مطالعه توضیح داده شده است و سپس در بخش چهارم از رویکردهای داده کاوی برای بررسی داده‌ها استفاده شده است. نتایج به دست آمده، در بخش پنجم مورد تجزیه و تحلیل قرار گرفته اند و جمع بندی نهایی این مطالعه در بخش ششم ذکر شده است.

2- شیوه نمونه گیری

اطلاعات مورد نیاز برای تشخیص خانوارهای نمونه در هر حوزه در طرح آمارگیری از کاربران اینترنت، از نتایج آمارگیری سال 5945 استخراج شده است. هدف کلی از اجرای این آمارگیری، برآورد ضریب نفوذ اینترنت در کشور به تفکیک مناطق شهری و روستایی است. جامعه‌ی آمارگیری در این طرح، شامل همه‌ی خانوارهای معمولی ساکن، خانوارهای گروهی و خانوارهای مؤسسه‌ای در کل کشور است. زمان آماری طرح بر حسب مورد، زمان آمارگیری یا 51 ماه منتهی به دی ماه 5945 است. در این طرح، مأموران آمارگیری در زمان آمارگیری به خانوارهای نمونه مراجعه کرده و اطلاعات مربوط به خانوار را از طریق مصاحبه‌ی حضوری با اعضای خانوار اخذ و در پرسش نامه‌ی مربوط ثبت کردند. در ضمن پرسشنامه‌ی خانوارهایی که حداقل یکی از اعضای آن در دوره‌ی زمانی یک ساله، از 51 ماه قبل از زمان آمارگیری تا زمان آمارگیری (دی ماه 5945)، از اینترنت استفاده میکردند را به خانوار تحویل داده تا کاربر یا

دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمانها اتخاذ شوند.

در داده کاوی از بخشی از علم آمار به نام تحلیل اکتشافی داده‌ها استفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تاکید می‌شود. علاوه بر این داده کاوی با هوش مصنوعی و یادگیری ماشین نیز ارتباط تنگاتنگی دارد، بنابراین می‌توان گفت در داده کاوی تئوریهای پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار را در هم می‌آمیزند تا زمینه کاربردی فراهم شود.

باید توجه داشت که اصطلاح داده کاوی زمانی به کار برده می‌شود که با حجم بزرگی از داده‌ها، در حد مگا یا ترابایت مواجه باشیم. در تمامی منابع داده کاوی بر این مطلب تاکید شده است. هر چه حجم داده‌ها بیشتر و روابط میان آنها پیچیده‌تر باشد دسترسی به اطلاعات نهفته در میان داده‌ها مشکلتر می‌شود و نقش داده کاوی به عنوان یکی از روشهای کشف دانش، روش‌تر می‌گردد.

اینترنت از پیوند تعداد بی شماری شبکه‌های ارتباطی کامپیوتوری کوچک و بزرگ که حاوی اطلاعات متنوع است، تشکیل شده است. یک فرد متصل به شبکه اینترنت تنها مشاهده‌گر و مرورگر اینترنت نیست بلکه جزئی از این شبکه بوده و می‌توان با آن تبادل اطلاعات نماید. در دنیای امروز اینترنت به عنوان ابزاری برای رشد و توسعه کشورها تبدیل شده و تاثیر آنها هر روز در ابعاد مختلف زندگی بیشتر می‌شود. براساس تعریف ارائه شده توسط مرکز آمار ایران، کاربر اینترنت فردی است که در یک دوره 51 ماهه، در داخل کشور، با اتصال به شبکه اینترنت به هر مدت، شخصاً حداقل یکی از خدمات اینترنتی استفاده کرده باشد. خدمات اینترنتی شامل دریافت اطلاعات، ارائه اطلاعات، پست الکترونیکی، خدمات دولتی، بانکداری اینترنتی، آموزش اینترنتی، کسب و کار اینترنتی، سرگرمی و ... می‌شود. [5]

در سال 1002 okazaki [4] به بررسی الگوهای رفتاری کاربران اینترنت از طریق تلفن همراه در ژاپن پرداخته و با استفاده از رویکرد خوش‌بندی، کاربران را با توجه به پارامترهایی مانند جنسیت، سن، وضع اشتغال، وضعیت تاهل و غیره به چهار خوش‌ تقسیم کرد.



- از خانوارهای مؤسسه ای 400 نفری به بالا، 100 نفر

حداکثر خطای نسبی پذیرفته شده در این طرح برای برآورد هر یک از پارامترهای مورد نظر برابر 4 درصد است. برای این طرح، تعداد 19314 خانوار نمونه، شامل 11541 خانوار معمولی ساکن و گروهی و 395 خانوار مؤسسه‌ای انتخاب شده است.

-3 متغیرهای مورد استفاده

متغیرهایی که در این آمارگیری بکار برده شد عبارتند از:
گروه‌های جنسی و سنی، وضع سواد، دوره یا مدرک تحصیلی، وضع فعالیت و وضع شغلی، تعداد کاربران به تفکیک نوع محل های دسترسی به اینترنت، تعداد کاربران اینترنت به تفکیک نوع دسترسی به اینترنت در منزل، متوسط دفعات دسترسی به اینترنت، متوسط مدت زمان دسترسی به اینترنت، متوسط هزینه پرداختی بابت استفاده از اینترنت، تعداد کاربران اینترنت به تفکیک استفاده از خدمات اینترنت، تعداد کاربران اینترنت به تفکیک میزان رضایت از خانوارها از تلفن ثابت، همراه یا رایانه در منزل، متوسط تعداد رایانه در منزل.

در این مطالعه، تعدادی از متغیرهای فوق مورد بررسی قرار گرفتند، که به منظور دستیابی به دیدگاهی مشترک، توضیحاتی درخصوص برخی از آنها ارائه می‌شود. یکی از پارامترهای مورد بررسی نوع استفاده از اینترنت است که در مورد هر یک توضیح داده می‌شود. دریافت اطلاعات، شامل دریافت اطلاعات درباره کالا یا خدمات، در ارتباط با سلامتی و خدمات پزشکی، سازمان های دولتی و غیر دولتی، جستجو در پایگاه‌های اینترنتی، دسترسی به پایگاه‌های اطلاعاتی، بارگذاری اطلاعات و کسب خبر و ... می‌شود. ارائه اطلاعات، شامل ارائه مطالب در وبلاگها و سایتها و ... می‌شود. پست الکترونیکی، شامل استفاده از پست های الکترونیک (E-Mail) و فرستادن و دریافت اطلاعات از طریق آن است. خدمات دولتی، شامل تکمیل فرم‌های ثبت نام کنکور سراسری، کارت سوخت و ... می‌شود. بانکداری اینترنتی، شامل پرداخت قبوض، اقساط وام‌های بانکی و ... می‌شود. آموزش اینترنتی، شامل تحصیل در دانشگاه‌های مجازی، آموزش Online زبان، بسته‌های نرم افزاری و ... می‌شود. کسب و کار اینترنتی، شامل خرید و فروش و سفارش کالا می‌شود. سرگرمی، شامل استفاده از محیط‌های گفتگو، بازی‌های ویدیویی و کامپیوترا، فیلم و آهنگ یا نرم افزار و ... می‌شود. نحوه

کاربران اینترنت، اطلاعات دقیق را شخصاً وارد پرسشنامه نموده و سپس مأموران آمارگیری در مراجعه‌ی مجدد، پرسشنامه را جمع آوری کردند.

پارامترهای مورد نظر در این طرح عبارتند از:

- نسبت کاربران اینترنت در مناطق شهری هر استان و مناطق روستایی کل کشور
- نسبت خانوارهای کاربر اینترنت در مناطق شهری هر استان و مناطق روستایی کل کشور

بر اساس 1 متغیر کمکی (نسبت سالمدان)، نسبت دانشگاهیان، نسبت محصلان، نسبت سرپرست مرد و نسبت خانوارهای دارای رایانه) که همبستگی بالایی با متغیر نسبت خانوارهای دارای اینترنت داشتند، یک مولفه‌ی اصلی به دست آمد که در استان، همبستگی بیش از 40 درصد با متغیر نسبت خانوارهای دارای اینترنت داشت. با استفاده از این مولفه اصلی و متغیر نسبت خانوارهای دارای اینترنت، حوزه‌های درون هر استان به جز استان قم که فقط یک شهرستان مرکز استان دارد در 4 طبقه، قرار گرفتند. سپس در هر طبقه‌ی استان، حوزه‌های نمونه به روش تصادفی ساده انتخاب شدند. در مرحله‌ی دوم نمونه گیری، درون هر یک از حوزه‌های انتخاب شده، خوش‌های نمونه به روش تصادفی ساده انتخاب شدند. در مرحله‌ی سوم نمونه گیری، داخل هر یک از خوش‌های نمونه، ده خانوار به روش تصادفی ساده انتخاب شدند که از بین آنها پنج خانوار به عنوان نمونه‌ی اصلی و پنج خانوار به عنوان نمونه‌ی جایگزین در نظر گرفته شدند. این خانوارها با روش نمونه گیری تصادفی ساده (بدون جایگذاری) از بین ده خانوار نمونه‌ی مرحله‌ی سوم، انتخاب شدند.

برای نمونه گیری از خانوارهای مؤسسه‌ای در مرحله‌ی اول، خانوارهای نمونه‌ی هر استان به روش PPS متناسب با اندازه‌ی جمعیت خانوارهای مؤسسه‌ای انتخاب شدند. سپس در مرحله‌ی دوم در هر یک از خانوارهای نمونه‌ی مرحله‌ی اول، از فهرست اعضای خانوار که شامل شماره‌ی ردیف افراد است یک نمونه‌ی تصادفی ساده بر اساس جمعیت خانوار به صورت زیر انتخاب شد:

- از خانوارهای مؤسسه‌ای دارای کمتر از 10 نفر، 3 نفر
- از خانوارهای مؤسسه‌ای 10 تا 49 نفری، 5 نفر
- از خانوارهای مؤسسه‌ای 50 تا 99 نفری، 25 نفر
- از خانوارهای مؤسسه‌ای 100 تا 399 نفری، 50 نفر



اعضای دسته های دیگر متفاوتند. تفاوت الگوریتم های مختلف این رویکرد بر اساس تعریف هر یک از آنها از شباهت می باشد.^[3]

الگوریتم درخت تصمیم 48J، یکی از الگوریتم های شناخته شده است. در صورتی که متغیر های مسئله، مانند متغیرهای این مطالعه از نوع کیفی باشند، نرم افزار WEKA ابتدا این متغیرها را به عنوان عددی تبدیل می کند.^[1] در جدول (1) پارامترها به چهار خوش تقسیم شده اند و درصد هر یک از پارامترها را در این جدول نشان داده اند.

دسترسی به اینترنت، شامل Dialup, ADSL, Wireless و سایر روشها مانند lease line WiMax می شود.

- 4 استفاده از داده کاوی و رویکردهای آن در بررسی الگوهای رفتاری کاربران اینترنت

1-1 خوش بندی کاربران اینترنت در سطح استان اصفهان

هدف از خوش بندی، تفکیک مشاهدات به دسته های مجزا بر اساس شباهت ها داده ها به یکدیگر است؛ به طوری که اعضای هر دسته دارای بیشترین شباهت به یکدیگر هستند و با

جدول (1) متغیرهای مورد بررسی

متغیرها	دسته بندی
جنسیت	(5) مرد (1) زن (9) اظهارنشده
سن	(5) کمتر از 50 سال، (1) 55-50 (4) 45-41 (1) 44-90 (4) 15-10 (9) 21 سال به بالا
مدرک تحصیلی	(5) بی سواد (1) ابتدایی، راهنمایی (9) متوسطه (4) عالی (1) سوادآموزی، غیر رسمی و اظهارنشده
وضع فعالیت	(5) شاغل (1) بیکار (9) محصل (4) خانه دار (1) سایر (2) اظهارنشده
محل دسترسی به اینترنت	(5) محل سکونت (1) محل تحصیل (9) محل کار (4) منزل افراد دیگر (1) کتابخانه (2) کافی نت (3) در هر مکان با تلفن همراه (4) در هر مکان با سایر دستگاه های سیار (5) اظهارنشده
تناوب دسترسی به اینترنت	(5) حداقل یک بار در روز (1) حداقل یک بار در هفته ولی نه هر روز (9) حداقل یک بار در ماه ولی نه هر هفته (4) کمتر از یک بار در ماه (1) اظهارنشده
هزینه پرداخت می کند یا نه ؟	(5) بله (1) خیر
متوسط پرداختی برای استفاده از اینترنت در 51 ماه	(5) کمتر از 10000 ریال (1) 10000-55555-500000 (9) 55555-5000000 (4) 45555-1100000 (1) 555555-100000 (2) 5555555-5000000 (3) بیش از 1000000 ریال (5) سایر
دریافت اطلاعات	(5) رضایت کامل (1) رضایت نسبی (9) عدم رضایت
ارائه اطلاعات	(5) عدم استفاده (1) استفاده
پست الکترونیک	(5) رضایت کامل (1) رضایت نسبی (9) عدم رضایت
خدمات دولتی	(5) رضایت کامل (1) رضایت نسبی (9) عدم رضایت
بانکداری اینترنتی	(5) عدم استفاده (1) استفاده
آموزش اینترنتی	(5) عدم استفاده (1) استفاده
کسب و کار اینترنتی	(5) عدم استفاده (1) استفاده
سرگرمی	(5) رضایت کامل (1) رضایت نسبی (9) عدم رضایت
نحوه دسترسی به اینترنت	(5) Wireless (9) ADSL (1) Dial up (5) اظهارنشده

دارند. خوش دوم، شامل 240 کاربر می باشد. تقریبا 31 درصد افراد این گروه مرد و جوان هستند. تحصیلات بیش از 21 درصد افراد این خوش دانشگاهی بوده و در حال تحصیل هستند. حدود 40 درصد از افراد این خوش بیش از یک بار در هفته به اینترنت دسترسی دارند. نوع استفاده این افراد از اینترنت در مقایسه با کل جامعه و سایر گروه ها متمایز است. به عنوان مثال 30 درصد آنها از دریافت اطلاعات از طریق اینترنت ناراضی هستند و 31

جدول (1) نشان می دهد که خوش اول، شامل 5450 کاربر است، که اکثریت آنها در گروه سنی بین 50 تا 55 سال قرار دارند. 41 درصد افراد این خوش دارای سطح تحصیلات دیپلم و پایین تر هستند و 45 درصد آنها شاغل و محصل هستند. اکثریت افراد این خوش بیش از یک بار در روز از اینترنت استفاده کردند. در این گروه 45 درصد افراد از دریافت اطلاعات بهره برده اند که 12 درصد از آنها از این خدمت رضایت نسبی

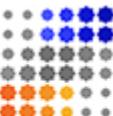


برای بانکداری اینترنتی و آموزش اینترنتی استفاده نموده اند و از این خدمت، رضایت کامل و نسبی دارند. پست الکترونیک و خدمات دولتی سهم بالایی در استفاده از اینترنت این گروه دارند اما بیش از نیمی از آنها از اینترنت به عنوان سرگرمی استفاده نمی کنند.

درصدشان از آموزش مجازی، ۴۰ درصد برای بانکداری از اینترنت بهره می گیرند. خوشه سوم و چهارم از ویژگی های مشترکی برخوردارند. بیش از ۲۵ درصد افراد این دو خوشه، مرد، و بیش از ۵۱ درصد از آنها جوان و دارای تحصیلات دانشگاهی می باشند. اعضای خوشه های سوم و چهارم بیشتر از اینترنت

جدول (۱) نتایج حاصل از خوشه بندی کاربران اینترنتی در سطح استان اصفهان

پارامترها	کل کاربران اصفهان	خوشه اول N=1981	خوشه دوم N=641	خوشه سوم N=911	خوشه چهارم N=851
جنسیت					
• زن	45۱۵	14۱۴	31۱۹۱	11	90۱۹۹
• مرد	14	45۱۱	14۱۱۱	31	25۲۳
• اظهار نشده	0۱۵	0۱۰۱	0۱۵	0	0
سن					
>10 •	50۱۵	91۱۳	1	5۱۲۱	4۱۹۹
19-10 •	41۱۱	41۱۱	35۱۲۰	51۱۱۲	4۵۱۹
29-20 •	11۱۵	51۱۴	11۱۱	5۱۹	9۱۵
44-30 •	51۱	5۱۱	31۱۲	5۱۹۹	11۱۳
64-45 •	4۱۱	0۱۵	5۱۱	0۱۰۵	0
65<= •	1۱۹	0۱۲	5	0۱۰۳	0
مدرک تحصیلی					
• بی سواد	0۱۵	0۱۰۱	0	0۱۰۱	0
• ابتدایی و راهنمایی	50۱۱	10۱۴۴	51۱۵	51۱۱	51۱۱
• متوسطه	40۱۹	91۱۵	11۱۱	40۱۱	40۱۹
• تحصیلات عالی	44۱۱	59۱۵	21۱۵	11۱۴	41۱۹
• سوادآموزی، غیر رسمی و اظهار نشده	51۱۱	59۱۳	0۱۴	51۱۱	44۱۵
وضع فعالیت					
• شاغل	90۱۱	14۱۱	59	50۱۵۱	51۱۱۱
• بیکار	51۱۴	5۱۱۴	0۱۴	0۱۵	5۱۹۱
• محصل	23۱۵۱	21۱۵۴	0۱۴	41۱۹۴	40۱۵
• خانه دار	51۱۱	51۱۱	0	11۱۴	51۱۰
• سایر	0۱۰۴	0۱۰۴	0۱۴	0۱۰۴	0۱۱
• اظهار نشده	0۱۱	0۱۰۱	0	0	0۱۰۴
محل دسترسی به اینترنت					
• محل سکونت	50۱۱	43۱۱	50۱۹	90۱۱	59۱۱
• محل کار	1۱۱۴	5۱۱	5	50۱۹	51۱
• محل تحصیل	91۱۹	1۱۱	59	51۱۹	11۱۳
• منزل افراد دیگر	51۱۹	5۱۱	5۱۱	23۱۵۱	50۱۱
• کتابخانه	11۱۱	1۱۱	1۱۱	41۱۱	11۱۵
• کافی نت	51۱۱	51۱۱	0۱۱	91۱۱	90۱۱
• در هر مکان با تلفن همراه	11۱۲	0۱۱	0۱۱	23۱۵	50۱۵
• در هر مکان با سایر دستگاههای سیار	0۱۱	0۱۱	0۱۱	0۱۱	9۱۱
• اظهار نشده	0۱۱	0۱۱	0۱۱	0۱۱	0۱۱



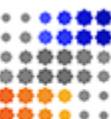


اولین کنفرانس ملی رویکردهای نوین در مهندسی کامپیوتر و بازیابی اطلاعات ایران

باشگاه پژوهشگران جوان و نخبگان دانشگاه آزاد اسلامی واحد رودسر و املش – ۵۱ مهر ۱۳۹۵



					تناولوب دسترسی به اینترنت
414 403 5011 914 0	449 929 509 515 0	14 155 214 5011 0	519 404 112 542 011	294 509 513 50 011	<ul style="list-style-type: none"> • حداقل یک بار در روز • حداقل یک بار در هفته ولی نه هر روز • حداقل ماه یک بار ولی نه هر هفته • کمتر از یک بار در ماه • اظهار نشده
35 15	32 14	443 519	443 559	412 534	هزینه پرداخت می کند؟
					هزینه پرداخت می کند؟
					متوسط پرداختی استفاده از اینترنت در ۵۱ ماه
512 1151 15113 1144 5549 1539 0111	5 2111 9141 15 5031 10111 09	94154 59113 5139 9531 9101 5115 0	59155 1155 94190 11194 55109 50195 012	5011 2115 1412 1219 1012 5144	<ul style="list-style-type: none"> >50000 • 50000-99999 • 10000-249999 • 250000-499999 • 500000-999999 • 1000000-1999999 • <2000000 •
015 412 5015 111	50 433 512 03	411 5211 5 309	50 5311 1011 1119	559 1515 1211 5012	دریافت اطلاعات
312 144	3515 1415	202 954	213 949	214 942	ارائه اطلاعات
					استفاده
					عدم استفاده
511 515 4412 515	99 414 515 111	515 155 313 959	404 9 5511 114	9511 952 93 011	پست الکترونیک
544 0 3511 1	1115 244 112 015	354 504 211 912	359 113 522 514	5015 5014 3214 515	خدمات دولتی
4515 5015	4412 5514	203 959	5014 511	414 542	بانکداری اینترنتی
					استفاده
					عدم استفاده
5111 4111	4111 5414	312 144	443 559	513 419	آموزش اینترنتی
					استفاده
					عدم استفاده
3412 1514	4111 5414	4412 5514	54 2	509 513	کسب و کار
					عدم استفاده
					استفاده
					سرگرمی





4514 14 4514 914	13 9311 12 15	1114 5511 1112 9114	1011 5112 1012 5219	4014 1419 5312 5318	عدم استفاده • رضایت کامل • رضایت نسبی • عدم رضایت
5114 5015 3111 511	419 4419 219 5.1	411 9211 1411 011	9119 2914 013 012	11 2015 319 5012	نحوه دسترسی به اینترنت Dial up • ADSL • Wireless • • سایر

برای اجرای این الگوریتم از نرم افزار WEKA استفاده شد که پس از بررسی های انجام شده تعدادی از این قوانین که بیانگر ویژگی های جالب توجه کاربران اینترنت در استان اصفهان بودند، استخراج گردیدند.

2-4 - درخت تصمیم

درخت تصمیم، از روش های معروف دسته بندی است [2] و به طور گسترده ای در کاربردهای دسته بندی مورد استفاده قرار گرفته است و تغییر و تمدیدات بسیاری داشته است.

یک درخت تصمیم مشتمل از گره های داخلی و پایانی (برگی) می باشد. گره های پایانی با یک برچسب طبقانی مجزا برچسب گذاری شده اند. الگوریتم درخت تصمیم معمولاً به صورت فرآیندی دو مرحله ای است. در مرحله 5 یا مرحله رشد، درخت تصمیمی بیش از حد رشد یافته از داده های آموزشی به دست می آید. معیار انتساب در هر گره داخلی جهت تقسیم کردن مجموعه داده ها به زیرمجموعه هایی که تفکیک پذیری دسته ای بهتری دارند، انتخاب می شود بنابراین خطای دسته بندی را کاهش می دهد. در مرحله 1 مرحله هرس، درخت با استفاده از روش های اکتشافی هرس می شود تا از بیش از اندازه شدن داده ها که مقدمه ای برای خطای دسته بندی در داده های آزمایشی هستند جلوگیری کند.

بطور کلی دسته بندی هایی که توسط درخت تصمیم گیری صورت می گیرد دارای صحت و دقت خوبی می باشند. با این وجود، ممکن است کاربرد موفق آن به اطلاعات موجود وابسته باشد. الگوریتم های درخت تصمیم، برای دسته بندی در بسیاری از کاربردها مثل پزشکی، صنعت، تولید، تجزیه و تحلیل های مالی، نجوم، و بیولوژی مولکولی استفاده شده است. تفاوت ها در الگوریتم های درخت تصمیم گیری شامل چگونگی انتخاب ویژگی ها در تولید درخت و مکانیسم های به کار رفته برای هرس کردن، می باشد.

یکی از شناخته ترین الگوریتم های درخت تصمیم، الگوریتم معروف C4.5 است که از عملیات هرس کردن نیز استفاده می کند. درخت ساخته شده در 0/04 ثانیه ترسیم می شود. در اجرای این الگوریتم میزان دقت اجرا تا 51٪ محاسبه شد.



[6] Zajchowski D. and et al, "Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells", *Cancer Research*, pages 5168–5178, 2001.

[3] بکارگیری تکنیکهای داده کاوی در بررسی ویژگی های کاربران اینترنت- پنجمین کنفرانس داده کاوی ایران، دانشگاه صنعتی امیرکبیر 5950.

عدم بهره گیری از کسب و کار اینترنتی می تواند به علت پایین بودن سطح اطمینان افراد از امنیت سایت های مربوطه باشد.

ارائه خدمات مطلوب در آموزش مجازی نیازمند بهبود سرعت شبکه اینترنت و کاهش هزینه های آن می باشد. می توان امیدوار بود با ارتقای کیفیت در کشور کاربران بتوانند از این خدمت با کیفیت بهتری بهره مند شوند

6- نتیجه گیری

با توجه به نفوذ و گسترش اینترنت در سطح کشور، بررسی نحوه استفاده و ویژگی های کاربران اینترنت یکی از مسائل جالب توجه می باشد، که تاکنون مورد بررسی قرار نگرفته است. این مطالعه، به بررسی ویژگیهای کاربران اصفهانی بر اساس اطلاعات آخرین طرح آمارگیری از کاربران اینترنت پرداخته است. برای یافتن الگوهای رفتاری و ویژگی های کاربران از دو رویکرد کاربردی داده کاوی یعنی خوش بندی و درخت های تصمیم استفاده شده است. نتایج این تحقیق نشان می دهد که به دلیل تلاش های دولت برای گسترش دولت الکترونیک و عدم ارائه حضوری برخی خدمات، کاربران مشارکت بالایی در استفاده از اینترنت برای خدماتی همچون ثبت نام آزمونها، کارت سوخت، یارانه و ... داشته اند. اما گسترش خدماتی نظیر کسب و کار اینترنتی نیازمند تلاش های بیشتری از جانب مسئولین می باشد. کاهش هزینه های سرویس های پرسرعت اینترنتی می تواند به افزایش استفاده از خدماتی مانند بانکداری و آموزش مجازی یاری رساند.

مراجع

[5] مرکز آمار ایران، نشریه نتایج طرح آمارگیری از کاربران اینترنت، 5945

[1] Jiawei Han and Micheline Kamber; Data mining concept and techniques, second edition; Morgan Kaufmann Publishers is an imprint of Elsevier; San Francisco, CA, 2006.

[2] A. Author 1 and B. Author 2, "Title of the conference paper," Proc. Int. Conf. on Power System Reliability. Singapore, pp. 100-105, 1999.

[3] Oliver Pietquin; A framework for unsupervised learning of dialogue strategies; Presses universités de Louvain; Belgium 2004.

[4] Shintaro Okazaki; "What do we know about mobile internet adopters? A cluster analysis"; Information and management, volume 43, 2 -141, 2006

[5] Utgoff P., "Incremental induction of decision trees", Machine Learning, page: 161–186, 1989.