فصل‌نامه دانش شناسی
(علوم کتابداری و اطلاع رسانی و فناوری اطلاعات)
دانشگاه آزاد اسلامی واحد تهران شمال
سال دوازدهم، شماره ٤٦، پاییز ١٣٩٨، از صفحه ٣٨ الی ٥٠

بررسی وب‌سایت‌های شهرداری تهران و ارایه راه‌کارهای مناسب داده‌کاوی

شایسته شجاعی کاریزکی[1] | سودابه شاپوری[2] | هاجر زارعی[3]

۱- دانشجوی دکتری تخصصی، گروه علم اطلاعات و دانش‌شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران
Shojaee.sh776@gmail.com
۲- استادیار، گروه علم اطلاعات و دانش‌شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران(نویسنده مسئول)
shapoori110@yahoo.com
۳- استادیار، گروه علم اطلاعات و دانش‌شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران hajar_zarei@yahoo.com

تاریخ دریافت: ۹۹/٦/۲۹ تاریخ پذیرش: ۹۹/۷/۳

چکیده

هدف: هدف اصلی این پژوهش، شناسایی و تجزیه و تحلیل انواع داده‌های وب‌سایت شهرداری تهران و ارایه راه‌کارهای مناسب داده‌کاوی است.

روش پژوهش: این پژوهش بنیادی است و از لحاظ ماهیت می‌توان آن را تحلیلی در نظر گرفت. روش گردآوری داده‌های به صورت میدانی بوده‌است و جامعه آماری ۴۷ سایت که از ۲۲۰ دامنه شهرداری تهران انتخاب شده‌است و برای تجزیه و تحلیل از تکنیک‌های داده‌کاوی استفاده شده‌است و منبع گردآوری داده‌ها وب‌سنجی و ابزار مورد استفاده گوگل آنالیتیکس می‌باشد.

یافته‌ها: میزان دقت الگوریتم شبکه عصبی معمولی برابر با ۹۹/۲۵٪ و در نهایت معیار آر.ام.اس.ای الگوریتم شبکه عصبی معمولی برابر با ۰/۱۵۹ است. میزان دقت الگوریتم درخت تصمیم برابر با ۹۹.۸۰٪ ؛ معیار ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۰۰۳ و در نهایت معیار آر.ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۰۴۵است. میزان دقت الگوریتم کی.ان.ان برابر با ۹۹/۸۱٪ و در نهایت معیار آر.ام.اس.ای الگوریتم کی.ان.ان برابر با ۰/۰۳۵ است

نتیجه‌گیری: بر اساس یافته‌های بدست آمده روش دی.بی.اسکن با سایر روش‌های پایه جهت تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران برابر و نسبت به سایر روش‌ها از دقت بالاتری برخوردار است.

واژه‌های کلیدی: وب‌سایت، شهرداری تهران، داده‌کاوی، درخت تصمیم، شبکه عصبی

مقدمه

کشف دانش و داده‌کاوی[1] یک حوزه‌ای جدید، میان رشته‌ای و در حال رشد است که حوزه‌های مختلفی چون پایگاه داده‌ها، آمار، یادگیری ماشینی و سایر زمینه‌های مرتبط را با هم تلفیق کرده تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده‌ها را استخراج نماید. با رشد سریع کامپیوتر و استفاده از آن در دو دهه‌ی اخیر تا حدودی همه سازمان‌ها حجم عظیمی از داده در پایگاه‌های داده خود ذخیره کرده‌اند.

اکثر تحلیل‌گران کسب‌وکار و مدیران اجرایی، خروجی پروژه‌های کلان داده‌ها را برای تجزیه و تحلیل مدیریت و پشتیبانی از تصمیم‌گیری ناکافی می‌دانند. (بندری[2]، ثول[3]، ۲۰۱۶) از سال ۲۰۰۶ تا سال ۲۰۱۱ به طور متوسط حدود ۵۰ درصد از این پروژه‌ها موفق به ارائه خروجی قابل قبول و عملی برای کاربران کسب‌وکار نشدند. هم چنین، درصد پروژه‌های تجزیه و تحلیل داده‌هایی که شکست خورده بودند با رشد چشم‌گیر داده‌ها در سازمان‌ها ادامه یافتند. (خاکی، ۱۳۹۰) مشکل عمده مدیریتی این بود که خروجی‌ها، مدل شناختی بصری، تحلیل‌گران کسب‌وکار و مدیران اجرایی، داده‌های وفق داده شده موجود را تطبیق ندادند. (زیکاری[4] و همکاران، ۲۰۱۵) امروزه فن‌آوری اطلاعات، داده‌کاوی، هوش تجاری و تحلیل داده به کمک صاحبان کسب‌وکار آمده و به آن‌ها این امکان را می‌دهد تا انباره‌های داده خود را تحلیل و بررسی نمایند و اطلاعات ارزشمندی را که سود بیشتری برای سازمان به ارمغان خواهد آورد را به دست آورند و از آن استفاده نمایند.

داده‌کاوی فرایندی تکرارپذیر است که وظیفه آن جستجو برای یافتن اطلاعات ارزشمند از میان حجم زیادی از داده است. با کمک تکنیک‌هایی از قبیل خوشه‌بندی[5]، پیش‌بینی[6] و کشف قوانین وابستگی[7]، اطلاعات ارزشمند و مفید از داده‌ها استخراج خواهد شد و به عبارتی الگوهای میان‌داده که منجر به شناخت الگوی مناسب تصمیم‌یاری مدیران شهری می‌شود، کشف خواهند شد. (ورنر[8]، گهرکه[9]، ناتگنز[10]، ۲۰۱۳) این ابزارها ممکن است مدل‌های آماری، الگوریتم‌های ریاضی و روش‌های یادگیرنده است که کار خود را به صورت خودکار و بر اساس تجربه‌ای که از طریق شبکه‌های عصبی یا درخت‌های تصمیم‌گیری به دست می‌آورند، بهبود می‌بخشند. با گسترش و توسعه شهرنشینی و هم چنین افزایش روزافزون جمعیت، نیاز به ارائه خدمات شهری مناسب شهروندان بیش از گذشته احساس می‌شود. مجموعه بررسی‌ها و تجزیه و تحلیل‌های انجام شده در مورد مجموعه داده‌های شهرداری‌ها و نحوه عملکرد آن‌ها به نظر می‌رسد که محدودیت‌ها و نارسایی‌های عمده‌ای در عملکرد مطلوب شهرداری‌ها وجود دارد. از این رو ارائه مدل‌های کاربردی مناسب جهت پیدا کردن دید مناسب مدیران شهری و برای به حداقل رساندن هزینه‌ها و بهترین خدمت‌رسانی به شهروندان است و می‌تواند کمک شایانی برای برنامه‌ریزی مناسب آینده باشد. با رشد فزاینده‌ی جمعیت در کلان‌شهر تهران و به دنبال آن افزایش مشکلات و مسائل ناشی از شهرنشینی در ابعاد وسیع اقتصادی، اجتماعی، فرهنگی و زیست‌محیطی، لزوم ایجاد و استفاده از بسترهای مدیریت شهری جدید هم‌راه با بروزترین و کاربردی‌ترین فن‌آوری‌ها احساس می‌شود. (امیری و همکاران، ۱۳۸۸) در طول چند سال گذشته افزایش سریع جمع‌آوری و نگهداری حجم اطلاعات وجود داشته‌است و به طور قطع امروزه به شهرهای با ساختمان‌های سر به فلک کشیده و اینترنت پر سرعت نمی‌توان گفت شهر هوشمند و در تعریف حقیقی شهر واقعاً هوشمند باید گفت؛ شهری که در سازمان‌های مختلف خود تصمیم‌گیری و تصمیم‌یاری مدیریت شهری از طریق داده و تجزیه و تحلیل داده انجام می‌گیرد. (فورتانه[11]، ۲۰۱۸) در این پژوهش سعی شده با استفاده از عوامل مختلف مدلی در جهت کشف الگوهای مناسب تصمیم‌گیری مدیران

1. Data mining
2. Bendre
3 .Thool
4. Zicari
5 .Clustering
6 .classification
7. Association Rules

8 .Werner
9 .Gehrke
10. Nüttgens
11. Fourtané

شهری تهران ارائه گردد. در نهایت هدف شناسایی و تجزیه و تحلیل داده‌های وب‌سایت شهرداری تهران و شناسایی انواع داده‌های مناسب از وب‌سایت‌های وتشخیص الگو برای تصمیم گیری از داده‌های انبوه شهرداری تهران. ضرورت این پژوهش در زمینه شناسایی و معرفی وب‌سایت‌های و استفاده از تکنیک‌ها و فنون مناسب داده‌کاوی، ایجاد نموده است. پیامد تغییر اجتماعی، افزایش مشارکت برای مدیریت در برنامه‌های اجتماعی به منظور حفظ مشارکت شهروندی در جوامع ذینفع از جمله حمایت از رویدادهای اجتماعی و برنامه‌های اجتماعی بوده است.

پژوهش های مشابه با پژوهش حاضر در زیر ارائه می شود:

اسپیلمن[1] و همکاران (۲۰۰۸) در ایالات متحده، در پژوهشی به بررسی و تجزیه و تحلیل اجتماعی مناطق با استفاده از روش‌های داده‌کاوی پرداختند. ترکیب روش‌های داده‌کاوی با سامانه اطلاعات جغرافیایی (جی.آی.اس)[2]، نقشه‌های پیوندی دارای ویژگی‌های اجتماعی و فضای جغرافیایی ایجاد می‌کند. این پژوهش اطلاعات و داده‌های شهر نیویورک مورد مطالعه قرار گرفته‌است، نیویورک به دلیل تنوع بالای نژادی، فرهنگی، قومی و اقتصادی گزینه مناسبی برای تحلیل‌های اجتماعی است. در این پژوهش یک پایگاه داده با ۷۹ ویژگی از ۲۲۱۷ رشته سرشماری از شهر نیویورک مورد استفاده قرار گرفته است. تجزیه و تحلیل با استفاده از این نقشه‌های پیوندی، برخی فرضیه‌ها در مورد شکل عملکردی روابط مکانی که براساس یک‌سری مدل‌سازی‌ها و تکنیک‌های آماری ایجاد شده، را زیر سوال می‌برد. کاندلیری[3] و همکارانش (۲۰۱۴) در ایتالیا، به منظور بهبود مدیریت منابع آب شهری، از روش‌های داده‌کاوی برای شناسایی محل نشت لوله‌های شبکه توزیع آب استفاده نموده‌اند. یکی از مشکلات موجود در شبکه توزیع آب، شبکه‌های لوله‌ای هستند که از نشت مکرر، خرابی و اختلال در سرویس به موقع رنج می‌برند. هزینه‌های پیگیری این مشکلات که شامل هزینه‌های بازرسی، تعمیر و تعویض است، بخش قابل توجهی از هزینه‌ها را تشکیل می‌دهد. در این پژوهش ابتدا داده‌های یک پایگاه داده که شامل جریان‌های نشت شبیه‌سازی شده توسط مدل‌های شبیه‌سازی است، با یک فرآیند تبدیل، به یک گراف تشابه تبدیل شده است. مجموعه داده مورد استفاده در این پژوهش، از اطلاعات مربوط به زیرساخت‌های شبکه توزیع آب و داده‌های مربوط به میزان مصرف مشترکین آب در شهر میلان ایتالیا استفاده شده‌است. لینگ[4] و همکاران (۲۰۱۶)، در پژوهشی برای دسته‌بندی مسیر تغییر محلات از روش‌های داده‌کاوی استفاده نمودند. شناسایی فرآیند تغییر محلات در طول زمان برای مسؤولین شهرداری و سیاست‌گذاران از اهمیت بالایی برخوردار است، این اطلاعات در تدوین سیاست‌ها مؤثر است و چشم‌اندازهای اقتصادی را بهبود می‌بخشد. در این مطالعه از داده‌های جغرافیایی، تاریخی و اقتصادی ۸ شهر مجزا در ایالات متحده استفاده شده‌است، که شامل شهرهای شارلوت[5]، دنور[6]، فونیکس[7]، پورتلند[8]، بوفالو[9]، دیترویت[10]، پیتسبورگ[11] و شیکاگو[12] است. در این مطالعه از پایگاه‌داده‌ای که شامل اطلاعات داده‌های سرشماری سالانه از سال‌های ۱۹۷۰ تا ۲۰۱۰ است، استفاده شده‌است. داده‌ی نهایی مورد استفاده دارای ۱۲ ویژگی در سه دسته‌ی اقتصادی-اجتماعی، مسکن و جمعیت قرار داده شد. آلامینس[13] و همکاران(۲۰۱۸) در اسپانیا، برای پیش‌بینی بحران مالی در شهرداری‌ها، از روش‌های داده‌کاوی استفاده نمودند. بحران مالی شهرداری‌ها، یک پدیده‌ی جهانی است که توجه محققان و مدیران را به دلیل ضرورت تداوم و پایداری مالی خدمات عمومی، به خود جلب کرده‌است. در این پژوهش استفاده از روش‌های داده‌کاوی برای پیش‌بینی بحران‌های مالی شهرداری‌ها پیشنهاد شده‌است. در این مطالعه از مجموعه داده ۱۲۸ شهرداری اسپانیا استفاده شده‌است، که هر کدام جمعیتی بیش از پنجاه هزار نفر دارند. این داده‌ها مربوط به سال ۲۰۱۵

1. Spielman
2. Geographic Information System
3. Candelieri

4. Ling
5. Charlotte
6 . Denver
7. Phoenix
8. Portland
9. Buffalo
10. Detroit
11. Pittsburgh
12. Chicago
13. Alaminos

است و توسط دیوان محاسبات اسپانیا، که داده‌های مربوط به حساب‌های سالیانه شهرداری‌های اسپانیا را منتشر می‌کند، جمع‌آوری شده است. نتایج حاصل از این پژوهش در تصمیم‌گیری مسئولین و سیاست‌گذاران بسیار کمک‌کننده خواهد بود، زیرا این پژوهش عواملی را شناسایی کرده‌است که تأثیر زیادی روی تداوم خدمات عمومی و اثربخشی اقدامات انجام‌شده برای تحقق اهداف ثبات در بودجه و پایداری مالی، دارد. مینایی بیدگلی (١٣٨٨) در پژوهشی با استفاده از داده‌کاوی در مدیریت ارتباط با شهروند به داده‌کاوی سامانه ١٣٧ شهرداری تهران پرداخته است. در مدیریت ارتباط با شهروند، هدف اصلی شناخت نیازها و خواسته‌های شهروندان و تلاش در جهت بهبود ارایه خدمات و افزایش سطح رضایت‌مندی ایشان است. شهرداری تهران، سامانه‌ای با عنوان سیستم تحلیل اطلاعات ١٣٧ ایجاد نموده که مرکز ساماندهی و رسیدگی به معضلات و مشکلات شهری است و پل ارتباطی بین شهروندان و شهرداری محسوب می‌شود. اطلاعات حاصل از این سامانه، حاوی مطالب مفیدی در مورد خدمات ارایه شده به شهروندان است و می‌تواند به عنوان منبعی مهم و مناسب برای مدیریت روابط تحلیلی با شهروند مورد بررسی و کاوش قرار بگیرد. انتظار می‌رود نتایج به دست آمده، در مدیریت مشکلات شهری و افزایش سطح رضایت‌مندی شهروندان مثمرثمر واقع شود. آخوندزاده و همکاران (١٣٩٢) مدل جدیدی برای مدیریت شهری را با استفاده از روش‌های داده‌کاوی پیشنهاد دادند. دولت‌ها موظفند نیازهای شهروندان را به درستی شناسایی نموده و با پیش‌بینی آن از بروز مسائل و مشکلات احتمالی جلوگیری کنند. در این راستا، استفاده از تکنیک‌ها و ابزارهای داده‌کاوی می‌تواند بسیار مثمرثمر واقع شود. سامانه ١٣٧ شهرداری تهران یکی از مراکزی است که با هدف ارائه خدمات شهری به شهروندان و رسیدگی سریع به مشکلات شهری تأسیس شده و پل ارتباطی بین شهروندان و شهرداری محسوب می‌شود. در این مقاله با استفاده از الگوریتم آپریوری[1] قوانین انجمنی موجود در بازه‌های زمانی/مکانی رخداد رخداد مشکلات به دست می‌آید. از قوانین به دست آمده می‌توان در شناسایی نیازهای شهروندان و بهبود مدیریت خدمات شهری و استفاده نمود. آهنگرکانی و خواسته (١٣٩٨) در تحلیل مصرف آب شهری (خانگی) شهرستان بابل با استفاده از روش های داده‌کاوی به مسئله کمبود آب در ایران پرداخته است. در این مطالعه از کاوش قوانین وابستگی و الگوریتم درخت تصمیم به منظور تحلیل داده‌های موجود در زمینه مصرف آب شهری محله‌های شهرستان بابل استفاده شده است. با بهره‌گیری از یک درخت تصمیم، عوامل مرتبط با مصرف، طبقه‌بندی و میزان مصرف آب بخشی از مشترکین به عنوان داده‌های تست پیش‌بینی شده است. با بررسی میزان مصرف آب خانوار، محله‌های با بیشترین میزان مصرف که محله‌های ساحلی رودخانه بابل‌رود هستند، به عنوان محله‌های پرمصرف شناسایی شده‌اند. ساختمان‌ها در این محله‌ها، ویلایی و دارای تراکم جمعیت پایین‌تر و مساحت فضای سبز و حیاط بالاتری نسبت به سایر محله‌ها هستند.

با بررسی و مرور پژوهش‌های داده‌کاوی انجام شده در شهرداری‌ها کاربردهای گوناگون داده‌کاوی در شهرداری‌ها مشخص گردید و پایگاه‌های داده مورد استفاده در این پژوهش‌ها، بررسی شده است. نتیجه مطالعات نشان می‌دهد که شهرداری‌ها مجموعه داده‌های متنوع و مفیدی دارند که در صورت استفاده از آن‌ها و شناسایی الگوها و دانش پنهان در آن‌ها، می‌توانند کمک قابل توجهی در فرآیند مدیریت شهری و تصمیم‌گیری‌ها داشته باشند و در سیاست‌گذاری‌های شهرداری و نهایتاً در بسترسازی ایجاد شهر هوشمند مورد استفاده قرار گیرند. قلمرو این مقاله از منظر مکانی-زمانی، در حوزه‌ی کاربرد داده‌کاوی در مدیریت سازمان فن‌آوری اطلاعات و ارتباطات در شهر تهران به علاوه دبیرخانه و مرکز تهران هوشمند است. محورهای ماموریتی و حوزه‌های کاری سازمان فاوا[2] عبارت است از: سامانه‌ها و نرم‌افزارهای مدیریت شهری، سرویس‌های شهروندی و خدمات پرداخت، راهبری شهر هوشمند، داده و آمار و رصد شهری، زیرساخت شبکه برق اضطراری و سخت‌افزار، برنامه‌ریزی و آموزش و پژوهش و خدمات مدیریت، مرکز داده

1. Apriori

٢ . فاوا عبارتی کلی در برگیرنده تمام فناوری‌های پیشرفته نحوه ارتباط و انتقال انتقال داده‌ها در سامانه‌های ارتباطی است.

و امنیت. برقراری ارتباطات فی‌مابین کلیه ساختمان‌های شهرداری تهران از طریق فیبرنوری یا وایرلس، راه‌اندازی اتوماسیون آموزش و اتوماسیون اداری به صورت یکپارچه جهت استفاده کلیه واحدهای شهرداری تهران و اجرای پروژه برق اضطراری جهت تمامی ساختمان‌های شهرداری از اقدامات اساسی صورت گرفته و نقش مهمی که در پژوهش‌های سابق بررسی نگردیده است ولیکن با توجه به تحولات مؤثر و بسیاری که رخ داده پژوهش در این زمینه ضروری به نظر می‌رسد.هدف پژوهش از این مقاله توصیفی-تحلیلی، بررسی مدل داده‌ای یک پروژه داده‌های معمولی در یک محیط اطلاعاتی کلان برای فرصت‌هایی جهت بهبود نمایش اطلاعات است. مدل داده‌ها به عنوان هدف اصلی تحقیق انتخاب شده‌است، زیرا بیان اطلاعات در مدل داده‌ها به پیشرفت و به کارگیری فهم معروف است. فرآیند تصمیم‌گیری اساس مدیریت را در سازمان‌ها تشکیل می‌دهد و اطلاعات صحیح را می‌توان از ملزومات اصلی اخذ تصمیم‌های صحیح در سازمان‌ها به شمار آورد.

روش پژوهش

از آن جا که در مقاله حاضر، هدف توصیف روابط و الگوهای موجود در پایگاه‌های داده سایت‌های شهرداری و یافتن راهی برای کشف الگوی مناسب در آن‌ها بوده و در اصل دست‌کاری و تغییر داده‌ها مورد نیاز نیست؛ بنابراین یک پژوهش تحلیلی خواهد بود. از نظر هدف نیز این پژوهش بنیادی است، زیرا به دنبال تعیین روابط تجربی میان پدیده‌ها بوده و هدف تشخیص الگوها و روابط تحلیلی از داده‌های حاصل از یک محدوده زمانی ۱۳۹۷-۱۳۹۸ است و با توجه به این‌که در طرح مقاله حاضر داده‌های مورد استفاده درون پایگاه‌های داده بوده و اینکه براساس مدل‌های خوشه‌بندی و طبقه‌بندی داده‌کاوی و الگوریتم‌های هوش مصنوعی انجام یافته است، این پژوهش را می‌توان از لحاظ ماهیت تحلیلی در نظر گرفت. منبع گردآوری داده‌ها وب‌سنجی است و ابزار مورد استفاده گوگل آنالیتیکس[1] است. گوگل آنالیتیکس یک ابزار تحلیلی وب رایگان است که توسط گوگل برای کمک به صاحبان مشاغل و سایت‌ها در تحلیل ترافیک وب‌سایت ارائه شده است. حتی اگر "تجزیه و تحلیل وب" بسیار کم و کوچک به نظر می‌رسد، نتایج حاصل از گوگل آنالیتیکس در واقع بسیار زیاد است. دلیل این امر این است که برای اکثر شرکت‌ها، وب‌سایت‌ها به عنوان مرکز اصلی ترافیک دیجیتال قرار دارد. با توجه به اینکه وب‌سایت‌ها قطب اصلی حضور دیجیتالی یک سازمان است، وب‌سایت هر سازمان بهترین راه برای ارائه دیدگاه همه جانبه از اثربخشی همه گروه‌هایی است که برای ارتقاء محصولات و خدمات خود به صورت آنلاین ارائه می‌دهد و گوگل آنالیتیکس ابزاری رایگان است که می‌تواند به سازمان در ردیابی اثربخشی بازاریابی دیجیتال کمک کند. به همین دلیل بیش از ۵۰ میلیون وب‌سایت در سراسر جهان از گوگل آنالیتیکس استفاده می‌کنند. (بایگانی دانشگاه استنفورد، ۲۰۰۴)[2]

نحوه‌ی کار گوگل آنالیتیکس بدین صورت است: گوگل آنالیتیکس چندین خط کد رهگیری را در کد وب‌سایت سازمان قرار می‌دهد. این کد فعالیت‌های مختلف کاربران را هنگام بازدید از وب‌سایت به همراه ویژگی‌های (مانند سن، جنسیت، علایق) آن کاربران ثبت می‌کند. پس از خروج کاربر از وب‌سایت، تمام اطلاعات را به سرور گوگل آنالیتیکس ارسال می‌کند. در مرحله بعد، گوگل آنالیتیکس داده‌های جمع‌آوری شده از وب‌سایت را به چند روش، در درجه اول با چهار سطح جمع می‌کند:

۱. سطح کاربر[3] (مربوط به اقدامات هر کاربر)؛

۲. سطح بازدید[4] (هر بازدید منحصر به فرد)؛

۳. سطح بازدید صفحات[5] (هر بازدید منحصر به فرد از صفحه)؛

۴. سطح رویداد[6] (کلیک دکمه، نمایش ویدیو، و غیره) (موسوی نطنزی، ۱۳۹۸).

1 .Google Analytics

2 «Origin of the name "Google"». .دانشگاه استنفورد ژانویه ۲۰۰۴. بایگانی‌شده ازاصلی در ۱۶ .در ۴ ژوئیه ۲۰۱۲ دریافت‌شده سپتامبر ۲۰۱۱.

3. User level
4. Session level
5. Pageview level
6 .Event level

در کاربردهای آماری، تجزیه و تحلیل داده‌ها را می‌توان به آمار توصیفی، تجزیه و تحلیل داده‌های اکتشافی و تجزیه و تحلیل داده‌های تأییدی تقسیم کرد. روش ای.دی.ای بر کشف ویژگی‌های جدید در داده‌ها و سی.دی.ای بر تأیید یا تکذیب فرضیه‌های موجود تمرکز دارد.

داده‌کاوی در وب‌سایت‌ها داده‌های اطلاعاتی به عنوان یکی از منابع حیاتی سازمان شناخته شده است و بسیاری از سازمان‌ها با اطلاعات و دانش‌سازمانی خود مانند سایر دارایی‌های ارزشمندشان برخورد می‌کنند. داده اطلاعاتی به اطلاعات خام سازمان اطلاق می‌شود و هم‌چنین داده‌های پردازش شده پس از طبقه‌بندی و آنالیز به دانش سازمان تبدیل می‌گردند. (گوبل، گرونوالد، ۱۹۹۹)

وب‌سایت شهرداری تهران

اولین مرحله، ورود داده‌های مسأله به سیستم می‌باشد. در جدول زیر مجموعه داده های وب سایت شهرداری تهران ارائه شده است.

جدول ۱: مجموعه داده های وب سایت شهرداری تهران

تعداد سایت	تعداد کلیک	تعداد مشاهده	امتیاز	کلید واژه
1	507	1812	۱۴/۰۵۰۹۷۶۹۴	تهران من
1	374	2548	۵/۵۲۷۲۴۰۷۲	نقشه تهران
1	260	5746	۲/۶۷۳۹۴۲۶۸	شاخص آلودگی هوای تهران
1	320	3651	۲/۵۶۵۹۰۰۵۲	طرح ترافیک
1	471	2213	۲/۰۰۴۶۵۵۴۵	شاخص آلودگی هوا
1	364	8883	۱/۶۰۶۲۱۱۵۷	آلودگی هوای تهران
1	168	2604	۱/۵۵۸۲۹۹۸۲	Mytehran
1	282	5083	۱/۵۵۶۸۰۹۰۲	طرح ترافیک تهران
1	163	8456	۱/۴۶۴۸۸۶۸۴	آلودگی هوا
1	448	8311	۱/۱۸۸۳۴۳۷۲	شهرداری تهران
1	172	6516	۱/۰۸۲۲۵۴۵۶	کیفیت هوای تهران
1	594	5721	۱/۰۲۷۰۹۶۹۱	my Tehran
1	150	3365	۱/۰۱۷۲۲۶۹۶	tehran man
1	306	7851	۰/۹۳۹۰۳۹۱۸	نقشه مترو تهران
1	327	6792	۰/۸۹۱۲۰۳۸۶	tehran map
1	502	6237	۰/۸۰۱۲۵۵۰۱	مترو تهران
1	599	6355	۰/۷۹۰۱۰۶۱۷	تهران
1	617	7104	۰/۶۵۶۵۶۶۵۳	کنترل کیفیت هوا
1	187	8028	۰/۶۲۲۹۴۰۶۱	Map
1	314	7736	۰/۵۵۱۱۰۶۳	map Tehran
1	207	4771	۰/۵۰۹۸۵۴۶۲	معاینه فنی
1	242	6995	۰/۴۹۴۱۲۶۱	پرداخت عوارض شهرداری
1	317	2698	۰/۴۷۲۵۰۳۸۱	شهروند
1	604	4796	۰/۴۶۷۳۷۰۹۴	خرید طرح ترافیک
1	530	2202	۰/۴۵۷۳۸۷۲۶	عوارض شهرداری
1	397	3634	۰/۴۱۵۶۱۳۶۲	عوارض خودرو

آلودگی هوا	۰/۴۰۴۷۸۵۹۴	7712	426	1
شاخص آلودگی هوا تهران	۰/۳۹۱۷۶۲۳۴	5070	488	1
شاخص الودگی هوا	۰/۳۸۰۳۹۶۰۸	5154	269	1
بهشت زهرا	۰/۳۶۱۵۱۱۲۵	5256	446	1
شهرداری	۰/۳۲۱۶۳۶۲۹	4813	302	1
نقشه	۰/۲۹۹۲۱۰۱۹	770	176	1
نقشه مترو	۰/۲۹۶۱۶۴۹۲	3206	595	1
وضعیت آلودگی هوای تهران	۰/۲۹۰۷۴۶۷۳	7728	553	1
شاخص آلودگی	۰/۰۲۸۸۸۸۲۵۶	5650	411	1
پرونده شهرسازی	۰/۲۱۰۴۱۷۲۴	6143	376	1
سایت تهران من	۰/۲۸۷۷۳۹۹۵	4421	578	1
فروشگاه شهروند	۰/۲۶۶۶۸۸۷۹	8657	428	1
مای تهران	۰/۲۶۰۴۷۲۳۵	7155	599	1
پرداخت عوارض خودرو	۰/۲۵۶۳۷۲۱۷	6578	548	1
آلودگی هوا تهران	۰/۲۵۴۰۶۲۰۷	1073	252	1
نوبت دهی معاینه فنی	۰/۲۵۳۹۰۷۱۲	578	423	1
الودگی هوای تهران	۰/۲۴۵۷۲۹۲۸	5677	250	1
مترو	۰/۲۳۲۴۶۷۳	5324	558	1
سایت بهشت زهرا	۰/۲۱۸۵۶	6843	275	1
سامانه تهران من	۰/۲۱۶۴۵۳۴۶	6353	584	1

همان طور که در جدول ۱ مشاهده می‌شود، داده‌های مربوط به ۴۷ نمونه از سایت‌های شهرداری تهران به عنوان جامعه آماری ارائه شده‌است. داده‌های در دسترس این سایت‌ها عبارتند از: کلید واژه، امتیاز، تعداد مشاهده، تعداد کلیک، نام سایت

به طور کلی، جامعه آماری این پژوهش داده‌های استخراج شده از ۴۷ وب‌سایت شهرداری تهران که از بین ۲۲۰ دامنه انتخاب شده‌است که در سال ۹۷-۱۳۹۸ است. این پژوهش فاقد روش نمونه‌گیری بوده چرا که تحلیل و پشتیبانی تصمیم در مدیریت نیازمند مشارکت تمامی داده‌های مرتبط با آن است. داده‌های جمع‌آوری شده در قالب فایل اکسل ارائه و در نهایت به الگوریتم خوشه‌بندی وارد شده‌است. در مرحله بعد مراحل شبیه‌سازی و اعمال الگوریتم خوشه‌بندی جهت ایجاد برچسب اولیه به داده‌ها ارائه شده‌است.

یافته‌ها

یافته‌های پژوهش الگوریتم شبکه عصبی معمولی، الگوریتم درخت تصمیم، الگوریتم کی.ان.ان و روش دی.بی.اسکن در این پژوهش انجام گرفته است را می‌توان برای وب سایت‌های شهرداری تهران مورد استفاده قرار گیرد.

الگوریتم شبکه عصبی معمولی

در این قسمت به مقایسه یافته‌های شبیه‌سازی الگوریتم شبکه عصبی معمولی پرداخته شده‌است. لازم به ذکر است که در این فاز، شبیه‌سازی بر روی ویژگی اصلی دیتاست صورت می‌گیرد. در شکل زیر نتایج حاصل از شبیه‌سازی الگوریتم شبکه عصبی معمولی نشان داده شده‌است.

همان طور که از شکل ۱ مشاهده می‌گردد، میزان دقت الگوریتم شبکه عصبی معمولی برابر با ۹۹/۲۵٪ می‌باشد. صحت الگوریتم شبکه عصبی معمولی برابر با ۹۸/۴۰٪، فراخوانی الگوریتم شبکه عصبی معمولی برابر با ۹۹/۴۷٪، خطای الگوریتم

شبکه عصبی معمولی برابر با ۰/۷۵٪، معیار ام.اس.ای الگوریتم شبکه عصبی معمولی برابر با ۰/۰۶۴ و در نهایت معیار آر.ام.اس.ای الگوریتم شبکه عصبی معمولی برابر با ۰/۱۵۹ می‌باشد.

الگوریتم درخت تصمیم

در این قسمت به مقایسه یافته‌های شبیه‌سازی الگوریتم درخت تصمیم پرداخته شده‌است. لازم به ذکر است که در این فاز، شبیه‌سازی بر روی ۴۲ ویژگی‌های اصلی دیتاست صورت می‌گیرد. در شکل زیر یافته‌ها حاصل از شبیه‌سازی الگوریتم درخت تصمیم نشان داده شده‌است.

PerformanceVector

```
PerformanceVector:
accuracy: 99.25% +/- 0.29% (mikro: 99.25%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5670    2
1.0:     53      1611
classification_error: 0.75% +/- 0.29% (mikro: 0.75%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5670    2
1.0:     53      1611
weighted_mean_recall: 99.47% +/- 0.18% (mikro: 99.47%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5670    2
1.0:     53      1611
weighted_mean_precision: 98.40% +/- 0.61% (mikro: 98.39%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5670    2
1.0:     53      1611
absolute_error: 0.064 +/- 0.057 (mikro: 0.064 +/- 0.169)
root_mean_squared_error: 0.159 +/- 0.087 (mikro: 0.181 +/- 0.000)
```

شکل ۱: یافته‌های حاصل از شبیه سازی الگوریتم شبکه عصبی معمولی

PerformanceVector

```
PerformanceVector:
accuracy: 99.80% +/- 0.01% (mikro: 99.80%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5720    12
1.0:     3       1601
classification_error: 0.20% +/- 0.01% (mikro: 0.20%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5720    12
1.0:     3       1601
weighted_mean_recall: 99.60% +/- 0.10% (mikro: 99.60%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5720    12
1.0:     3       1601
weighted_mean_precision: 99.80% +/- 0.06% (mikro: 99.80%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5720    12
1.0:     3       1601
absolute_error: 0.003 +/- 0.000 (mikro: 0.003 +/- 0.045)
root_mean_squared_error: 0.045 +/- 0.002 (mikro: 0.045 +/- 0.000)
```

شکل ۲. یافته‌های حاصل از شبیه‌سازی الگوریتم درخت تصمیم

همان طور که از شکل ۲ مشاهده می‌گردد، میزان دقت الگوریتم درخت تصمیم برابر با ۹۹/۸۰٪ می‌باشد. صحت الگوریتم درخت تصمیم برابر با ۹۹/۸۰٪، فراخوانی الگوریتم درخت تصمیم برابر با ۹۹/۶۰٪، خطای الگوریتم درخت تصمیم برابر با ۰/۲۰٪، معیار ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۰۰۳ و در نهایت معیار آر.ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۴۵ می‌باشد

الگوریتم کی. ان. ان

در این قسمت به مقایسه یافته‌های شبیه‌سازی الگوریتم کی.ان.ان پرداخته می‌شود. لازم به ذکر است که در این فاز، شبیه‌سازی بر روی ویژگی‌های اصلی دیتاست صورت می‌گیرد. در شکل زیر یافته‌های حاصل از شبیه‌سازی الگوریتم کی.ان.ان نشان داده شده‌است.

PerformanceVector

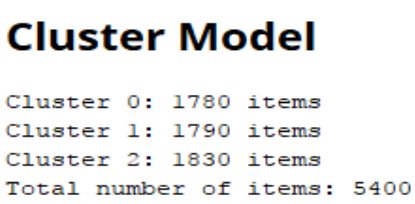
```
PerformanceVector:
accuracy: 99.81% +/- 0.03% (mikro: 99.81%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5723    14
1.0:     0       1599
classification_error: 0.19% +/- 0.03% (mikro: 0.19%)
ConfusionMatrix:
True:    0.0     1.0
0.0:     5723    14
1.0:     0       1599
weighted_mean_recall: 99.57% +/- 0.06% (mikro: 99.57%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5723    14
1.0:     0       1599
weighted_mean_precision: 99.88% +/- 0.02% (mikro: 99.88%), weights: 1, 1
ConfusionMatrix:
True:    0.0     1.0
0.0:     5723    14
1.0:     0       1599
absolute_error: 0.002 +/- 0.000 (mikro: 0.002 +/- 0.035)
root_mean_squared_error: 0.035 +/- 0.004 (mikro: 0.035 +/- 0.000)
```

شکل ۳. یافته‌های حاصل از شبیه سازی الگوریتم کی.ان.ان

همان طور که از شکل ۳ مشاهده می‌گردد، میزان دقت الگوریتم کی.ان.ان برابر با ۹۹/۸۱٪ می‌باشد. صحت الگوریتم کی.ان.ان برابر با ۹۹/۸۸٪، فراخوانی الگوریتم کی.ان.ان برابر با ۹۹/۵۷٪، خطای الگوریتم کی.ان.ان برابر با ۰/۱۹٪، معیار ام.اس.ای الگوریتم کی.ان.ان برابر با ۰/۰۰۲ و در نهایت معیار آر.ام.اس.ای الگوریتم کی.ان.ان برابر با ۰/۰۳۵ است.

روش دی.بی.اسکن

مزیت این روش به نسبت روش‌های دیگری خوشه‌بندی این است که نسبت به شکل داده‌ها حساس نیست و می‌تواند اشکال غیرمنظم را نیز در داده‌ها تشخیص دهد و پس از اعمال الگوریتم خوشه‌بندی دی.بی.اسکن، خروجی ذیل حاصل می‌گردد.



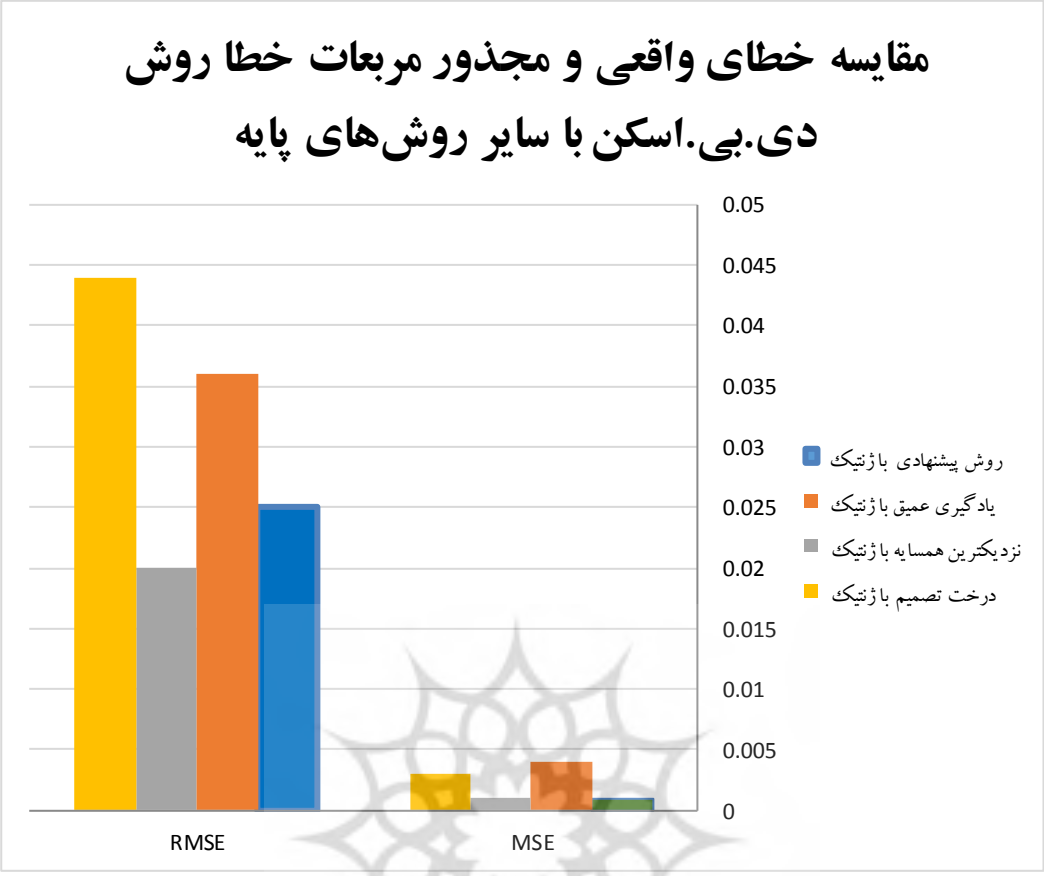
شکل ٤: تعداد خوشه ها در روش دی.بی.اسکن

در این پژوهش سه خوشه به عنوان تعداد خوشه‌های بهینه جهت گروه‌بندی اولیه انتخاب گردید. این سه خوشه مبتنی بر اجرای الگوریتم خوشه‌بندی ایکس-مینز انتخاب شده‌است. الگوریتم ایکس-مینز تعداد حد پایین و حد بالای خوشه‌ها را مشخص نموده و تعداد خوشه بهینه را محاسبه می‌کند. در این مقاله نیز حد پایین خوشه ها را ۲ و حد بالا برابر با ۵۰ قرار داده است. با اجراهای مختلف الگوریتم خوشه‌بندی ایکس-مینز تعداد ۳ خوشه به عنوان خوشه‌های بهینه محاسبه گردید که این تعداد خوشه‌ها به عنوان ورودی الگوریتم خوشه‌بندی دی.بی.اسکن در نظر گرفته است. تعداد ۳ خوشه برای کلیه داده‌ها و کلیدواژه‌های ارائه شده تعیین گردیده است. تعداد نمونه‌های موجود در خوشه ۱ برابر با ۱۷۸۰ نمونه، تعداد نمونه‌های موجود در خوشه ۲ برابر با ۱۷۹۰ نمونه و تعداد نمونه‌های موجود در خوشه ۳ برابر با ۱۸۳۰ نمونه است. وزن خوشه ۱ برابر با ۰.۴۳۳۶، وزن خوشه ۲ برابر با ۰.۲۳۲۶ و وزن خوشه ۳ برابر با ۰.۳۷۴۸ است. بنابر این، بر اساس شماره خوشه‌ها می‌توان تعیین کرد که کلیدواژه جدید ارائه شده برای مدیران با چه اهمیتی برخوردار است. به عنوان نمونه، کلیدواژه "سایت تهران من" که در خوشه ۱ قرار دارد از اهمیت بالاتری نسبت به سایر خوشه‌ها برخوردار است.

بحث و نتیجه گیری

میزان خطای واقعی روش دی.بی.اسکن با سایر روش‌های پایه جهت تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران برابر با ۰/۰۰۳٪و روش درخت تصمیم برابر با ۰/۰۰۳٪ است. با این تفاسیر میزان مجذور مربعات خطای روش دی.بی.اسکن با روش درخت تصمیم برابر با ۰/۰۴۵٪ است. داده‌های استفاده شده به منظور شبیه‌سازی و تست روش دی.بی.اسکن مستخرج از وب‌سایت‌های معرفی شده شهرداری تهران است که این داده‌ها اطلاعاتی را در مورد پسماندها، تصادفات شهری، آلودگی هوا و غیره را ارائه می‌کند. بنابراین می‌توان این گونه استنباط نمود که انواع داده‌های مناسب را استخراج نموده و از آن‌ها برای ارزیابی روش دی.بی.اسکن استفاده کرده‌ایم.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



نمودار ۱. مقایسه خطای واقعی و مجذور مربعات خطا روش دی.بی.اسکن با سایر روش‌های پایه

الگوهای قابل حصول و قابل پیش‌بینی از طریق راه‌کارهای داده‌کاوی در وب‌سایت‌های شهرداری تهران کدام است؟ استفاده از الگوهای به دست آمده در فرآیند داده‌کاوی وب‌سایت‌های شهرداری تهران چگونه قابل استفاده است؟

با اعمال الگوریتم خوشه‌بندی دی.بی.اسکن و خوشه‌بندی اولیه کلیدواژه‌های مستخرج از وب‌سایت‌های شهرداری تهران یک الگوی اولیه به دست می‌آید. در واقع خوشه‌ها ارزش کلیدواژه‌ها بر اساس میزان امتیاز داده شده، تعداد کلیک‌های انجام شده بر روی هر کلیدواژه و تعداد مشاهدات آن‌ها را مشخص می‌کنند. بنابراین براساس الگوی وزن‌دهی با استفاده از الگوریتم خوشه‌بندی و سپس الگوریتم یادگیری شبکه عصبی عمیق اقدام به شناسایی و تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران و ارائه راهکارهای داده‌کاوی می‌گردد. با توجه به یافته‌های تحقیق و نتایج به دست آمده مشاهده گردید که دقت، صحت و فراخوانی و سایر یافته‌های تحقیق نسبت به سایر روش‌ها جهت تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران و ارائه راهکارهای داده‌کاوی بهتر عمل کرده و از این استراتژی که ترکیبی از تکنیک‌های خوشه‌بندی دی.بی.اسکن و شبکه عصبی عمیق ال.اس.تی.ام است، می‌توان اعتماد کرد و در شهرداری تهران مورد استفاده قرار داد. یافته‌ها نشان داد که وب‌سایت‌های شهرداری تهران همکاری مطلوبی با یکدیگر نداشته و موازی‌کاری در آن‌ها بسیار است که این خود موجب اتلاف وقت و انرژی و منابع شهرداری‌ها شده و موجب اخلال در اجرای تصمیمات مدیران می‌گردد. از آن‌جا که این وب‌سایت‌ها دارای مطالب مشترک در زمینه‌های یکسان است، در صورت تبادل لینک، احتمال دیده شدن آن‌ها افزایش می‌یابد. از آن‌جاکه اطلاعات باید جدید و به روز است، می‌توان با پشتیبانی و به‌روزرسانی مرتب امکان مراجعه و تعامل شهروندان با وب‌سایت‌های شهرداری را افزایش داد و مراجعه فیزیکی و ازدحام ارباب‌رجوع به شهرداری‌ها را کم کرد. مقایسه پیشینه‌ها بررسی داده‌های وب‌سایت‌های شهرداری تهران با این گستره که تمامی کلیدواژه‌ها و کلیک‌ها و بازدیدها را مورد توجه قرار دهد

صرفاً در این پژوهش انجام شده‌است و در پژوهش انجام‌شده در گذشته در خصوص تجزیه و تحلیل داده‌های ثبت‌شده در سامانه ۱۳۷ که مبنای وجودی آن دریافت دیدگاه‌ها و نقطه نظرات شهروندان در خصوص مدیریت شهری و هم چنین ایجاد ارتباط بیشتر شهروندان با مدیران ارشد شهرداری بوده‌است و تاکنون در هیچ پژوهشی انباره عظیم داده‌های وب‌سایت‌های شهرداری تهران در پایگاه داده‌ای مورد تجزیه و تحلیل و داده‌کاوی قرار نگرفته‌است.

پیشنهاد پژوهش برگرفته از یافته های پژوهش می تواند شامل:

- ترکیب سایر روش‌های یادگیری ماشین یا تکنیک‌های بهینه‌سازی و انتخاب ویژگی از جمله الگوریتم ژنتیک؛
- الگوریتم گل‌های کلونال به منظور بهبود تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران ؛
- استفاده از الگوریتم‌های بوستینگ به منظور بهبود فرایند تجزیه و تحلیل داده‌های وب‌سایت‌های شهرداری تهران و مقایسه نتایج بدست آمده با یافته‌های این تحقیق.

منابع

آخوندزاده، ا.؛ مینایی، ب.؛ ربیع زاده گنجی، ث. (۱۳۹۲). *کاربرد داده کاوی در ارزیابی مشتریان بانکی*. نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی، تهران، مرکز همایش‌های بین‌المللی صدا وسیما،۸-۹ بهمن، ۲-۱۵.

آهنگرکانی، م.؛ خواسته، ح. (۱۳۹۸). تحلیل مصرف آب شهری (خانگی) شهرستان بابل با استفاده از روش های داده‌کاوی. *اطلاعات جغرافیایی سپهر*، ۲۸ (۱۱۱)، ۵۳ – ۶۹.

امیری، آ.؛ دلجو، غ.؛ قربان یزاده، و. (۱۳۸۸). *عوامل مؤثر بر پذیرش سامانه مدیریت شهری تهران (۱۳۷) توسط شهروندان*. دومین کنفرانس بین المللی شهرداری الکترونیکی،۶ (۲۲)، ۷-۲۲.

خاکی، غ. (۱۳۹۰). روش تحقیق با رویکرد پایان‌نامه‌نویسی. تهران: انتشارات بازتاب.

دانشگاه استنفورد(۲۰۰۴). *"Origin of the name "Google"* . بایگانی‌شده از اصلی در ۴ ژوئیه ۲۰۱۲ .دریافت‌شده در ۱۶ سپتامبر ۲۰۱۱.
http://graphics.stanford.edu/~dk/google_name_origin.html

موسوی نطنزی، م. (۱۳۹۸). گوگل آنالیتیکس چیست و چگونه به آنالیز سایت می‌پردازد؟https://hamyar.co/what-is-google-analytics/ تاریخ دسترس: ۹۹/۵/۵

مینایی بیدگلی، ب.؛ آخوندزاده، ا.؛ موسوی، ح. و دیگران (۱۳۸۸). *استفاده از داده کاوی در مدیریت ارتباط با شهروند: مورد کاوی سامانه ۱۳۷ شهرداری تهران*. دومین کنفرانس بین المللی شهر الکترونیک، تهران، جهاد دانشگاهی، ۲.

Alaminos, D.; Fernández, S. M.; García, F., & et. al. (۲۰۱۸) .*Data mining for municipal financial distress prediction.* Paper presented at the Industrial Conference on Data Mining.

Bendre, M. R., & Thool, V. R. (2016). Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics*, 3(3), 206-239

Candelieri, A., Conti, D., & Archetti, F. (2014). Improving analytics in urban water management: a spectral clustering-based approach for leakage localization. *Procedia-Social and Behavioral Sciences*, 108, 235-248 .

Fourtané, S (2018). Connected Vehicles in Smart Cities: The Future of Transportation (2018) Published by interestingengineering.com on 16 November 2018

Goebel, M., & Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. *ACM SIGKDD explorations newsletter*, 1(1), 20-33 .

Ling, C., & Delmelle, E. C. (2016). Classifying multidimensional trajectories of neighbourhood change: a self-organizing map and k-means approach. *Annals of GIS*, 22(3), 173-186 .

Spielman, S. E., & Thill ,J.-C. (2008). Social area analysis, data mining, and GIS. Computers. *Environment and Urban Systems*, 32(2), 110-122 .

Werner, M.; Gehrke, N., & Nüttgens, M. (2013). Towards Automated Analysis of Business Processes for Financial Audits. Paper presented at the Wirtschaftsinformatik.

Zicari, S.; Arakelyan, A.; Fitzgerald, W, Zaitseva, (2016). Evaluation of the maturation of individual Dengue virions with flow virometry. *virology*, 488,20-27

Quarterly Journal of Knowledge Studies
(Library and Information Science and Information Technology)
Islamic Azad University North Tehran Branch
Vol. 12, No. 46, Autumn 2019

Reviewing the websites of Tehran Municipality and providing appropriate data mining solutions

Shayesteh Shojaei karizaki[1] | Sudabeh Shapoori[2] | Hajar Zarei[3]

1-PhD Student, Department of Knowledge& Information Science, Information Retrieval Orientation, Tonekabon Branch, Islamic Azad University, Iran. Shojaee.sh776@gmail.com
(Corresponding author).
2-Assistant Professor, Department of Knowledge &Information Science, Tonekabon Branch, Islamic Azad shapoori110@yahoo.com University, Iran.
3-Assistant Professor, Department of Knowledge &Information Science, Tonekabon Branch, Islamic Azad University, Iran. hajar_zarei@yahoo.com

Abstract:

Objective: The main purpose of this study is to identify and analyze different types of data on the website of Tehran Municipality and to provide appropriate data mining solutions.

Method: This research is fundamental and in terms of nature it can be considered analytical. The data collection method was field and the statistical population of 47 sites were selected among 220 domains of Tehran Municipality and data mining techniques were used for analysis and the source of data collection is web analytics and tools used by Google Analytics.

Results: The accuracy of the normal neural network algorithm is equal to 99.25% and the RMS standard of the normal neural network algorithm is equal to 0.159. The accuracy of the decision tree algorithm is 99.80% and the MSI criterion of the decision tree algorithm is 0.003 and finally the RMS criterion of the decision tree algorithm is 0.045. The accuracy of the CNN algorithm is equal to 99.81% and finally the RMS criterion of the CNN algorithm is equal to 0.035.

Conclusion: Based on the obtained findings, the DB Scan method is equal to other basic methods for analyzing data of Tehran Municipality websites and has a higher accuracy than other methods.

Keywords: Website Review, Tehran Municipality, Data Mining, Decision Tree, Neural Network