

بررسی وبسایت‌های شهرداری تهران و آرایه راه کارهای مناسب داده کاوی

شایسته شجاعی کاریزکی^۱ | سودابه شاپوری^۲ | هاجر زارعی^۳

۱- دانشجوی دکتری تخصصی، گروه علم اطلاعات و دانش شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران
Shojaee.sh776@gmail.com

۲- استادیار، گروه علم اطلاعات و دانش شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران (نویسنده مسئول)
shapoori110@yahoo.com

۳- استادیار، گروه علم اطلاعات و دانش شناسی، گرایش بازیابی اطلاعات، واحد تنکابن، دانشگاه آزاد اسلامی، ایران
hajar_zarei@yahoo.com

تاریخ دریافت: ۹۹/۶/۲۹ تاریخ پذیرش: ۹۹/۷/۳

چکیده

هدف: هدف اصلی این پژوهش، شناسایی و تجزیه و تحلیل انواع داده‌های وبسایت شهرداری تهران و آرایه راه کارهای مناسب داده کاوی است.

روش پژوهش: این پژوهش بنیادی است و از لحاظ ماهیت می‌توان آن را تحلیلی در نظر گرفت. روش گردآوری داده‌های به صورت میدانی بوده است و جامعه آماری ۴۷ سایت که از ۲۲۰ دامنه شهرداری تهران انتخاب شده است و برای تجزیه و تحلیل از تکنیک‌های داده کاوی استفاده شده است و منبع گردآوری داده‌ها وبسنگی 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 آر.ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۰۰۳ و در نهایت معیار آر.ام.اس.ای الگوریتم درخت تصمیم برابر با ۰/۰۴۵ است. 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

نتیجه گیری: 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، داده کاوی، درخت تصمیم، شبکه عصبی

مقدمه

کشف دانش و داده کاوی^۱ یک حوزه ای جدید، میان رشته ای و در حال رشد است که حوزه های مختلفی چون پایگاه داده ها، آمار، یادگیری ماشینی و سایر زمینه های مرتبط را با هم تلفیق کرده تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده ها را استخراج نماید. با رشد سریع کامپیوتر و استفاده از آن در دو دهه ی اخیر تا حدودی همه سازمان ها حجم عظیمی از داده در پایگاه های داده خود ذخیره کرده اند.

اکثر تحلیل گران کسب و کار و مدیران اجرایی، خروجی پروژه های کلان داده ها را برای تجزیه و تحلیل مدیریت و پشتیبانی از تصمیم گیری ناکافی می دانند. (بندری^۲، ثول^۳، ۲۰۱۶) از سال ۲۰۰۶ تا سال ۲۰۱۱ به طور متوسط حدود ۵۰ درصد از این پروژه ها موفق به ارائه خروجی قابل قبول و عملی برای کاربران کسب و کار نشدند. هم چنین، درصد پروژه های تجزیه و تحلیل داده هایی که شکست خورده بودند با رشد چشم گیر داده ها در سازمان ها ادامه یافتند. (خاکی، ۱۳۹۰) مشکل عمده مدیریتی این بود که خروجی ها، مدل شناختی بصری، تحلیل گران کسب و کار و مدیران اجرایی، داده های وفق داده شده موجود را تطبیق ندادند. (زیکاری^۴ و همکاران، ۲۰۱۵) امروزه فن آوری اطلاعات، داده کاوی، هوش تجاری و تحلیل داده به کمک صاحبان کسب و کار آمده و به آن ها این امکان را می دهد تا انباره های 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.

داده کاوی فرایندی تکرار پذیر است که وظیفه آن جستجو برای یافتن اطلاعات ارزشمند از میان حجم زیادی از داده است. با کمک تکنیک هایی از قبیل خوشه بندی^۵، پیش بینی^۶ و کشف قوانین وابستگی^۷، اطلاعات ارزشمند و مفید از داده ها استخراج

خواهد شد و به عبارتی الگوهای میان داده که منجر به شناخت الگوی مناسب تصمیم یاری مدیران شهری می شود، کشف خواهند شد. (ورنر^۸، گهرکه^۹، ناتگنز^{۱۰}، ۲۰۱۳) این ابزارها ممکن است مدل های آماری، الگوریتم های ریاضی و روش های یادگیرنده است که کار خود را به صورت خودکار و بر اساس تجربه ای که از طریق شبکه های عصبی یا درخت های تصمیم گیری به دست می آورند، بهبود می بخشند. با گسترش و توسعه شهرنشینی و هم چنین افزایش روزافزون جمعیت، نیاز به ارائه خدمات شهری مناسب شهروندان بیش از گذشته احساس می شود. مجموعه بررسی ها و تجزیه و تحلیل های انجام شده در مورد مجموعه داده های شهرداری ها و نحوه عملکرد آن ها به نظر می رسد که محدودیت ها و نارسایی های عمده ای در عملکرد مطلوب شهرداری ها وجود دارد. از این رو ارائه مدل های کاربردی مناسب جهت پیدا کردن دید مناسب مدیران شهری و برای به حداقل رساندن هزینه ها و بهترین خدمت رسانی به شهروندان است و می تواند کمک شایانی برای برنامه ریزی مناسب آینده باشد. با رشد فزاینده ی جمعیت در کلان شهر تهران و به دنبال آن افزایش مشکلات و مسائل ناشی از شهرنشینی در ابعاد وسیع اقتصادی، اجتماعی، فرهنگی و زیست محیطی، لزوم ایجاد و استفاده از بسترهای مدیریت شهری جدید هم راه با بروزترین و کاربردی ترین فن آوری ها احساس می شود. (امیری و همکاران، ۱۳۸۸) در طول چند سال گذشته افزایش سریع جمع آوری و نگهداری حجم اطلاعات وجود داشته است و به طور قطع امروزه به شهرهای با ساختمان های سر به فلک کشیده و اینترنت پر سرعت نمی توان گفت شهر هوشمند و در تعریف حقیقی شهر واقعاً هوشمند باید گفت؛ شهری که در سازمان های مختلف خود تصمیم گیری و تصمیم یاری مدیریت شهری از طریق داده و تجزیه و تحلیل داده انجام می گیرد. (فورتانه^{۱۱}، ۲۰۱۸) در این پژوهش سعی شده با استفاده از عوامل مختلف مدلی در جهت کشف الگوهای مناسب تصمیم گیری 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

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 شده است. لینگ^۴ و همکاران (۲۰۱۶)، در پژوهشی برای دسته‌بندی مسیر تغییر محلات از روش‌های داده‌کاوی استفاده نمودند. شناسایی فرآیند تغییر محلات در طول زمان برای مسئولین 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 شده است، که شامل شهرهای شارلوت^۵، دنور^۶، فونیکس^۷، پورتلند^۸، بوفالو^۹، دیترویت^{۱۰}، پیتمبورگ^{۱۱} و شیکاگو^{۱۲} است. در این مطالعه از پایگاه داده‌ای که شامل اطلاعات داده‌های سرشماری سالانه از سال‌های ۱۹۷۰ تا ۲۰۱۰ است، استفاده شده است. داده‌ی نهایی مورد استفاده دارای ۱۲ ویژگی در سه دسته‌ی اقتصادی-اجتماعی، مسکن و جمعیت قرار داده شد. آلامینوس^{۱۳} 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 کرده است. در این پژوهش استفاده از روش‌های داده‌کاوی برای پیش‌بینی بحران‌های مالی شهرداری‌ها پیشنهاد شده است. در این مطالعه از مجموعه داده ۱۲۸ شهرداری اسپانیا استفاده شده است، که هر کدام جمعیتی بیش از پنجاه هزار نفر دارند. این داده‌ها مربوط به سال ۲۰۱۵

شهری تهران ارائه گردد. در نهایت هدف شناسایی و تجزیه و تحلیل داده‌های وب‌سایت شهرداری تهران و شناسایی انواع داده‌های مناسب از وب‌سایت‌های و تشخیص الگو برای تصمیم‌گیری 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.

پژوهش‌های مشابه با پژوهش حاضر در زیر ارائه می‌شود:

اسپلمن^۱ 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)^۲، نقشه‌های پیوندی دارای ویژگی‌های اجتماعی و فضای جغرافیایی ایجاد می‌کند. این پژوهش اطلاعات و داده‌های شهر نیویورک مورد مطالعه قرار گرفته است، 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^۳ 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

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

1. Spielman
2. Geographic Information System
3. Candelieri

در شناسایی نیازهای شهروندان و بهبود مدیریت خدمات شهری و استفاده نمود. آهنگرکانی و خواسته (۱۳۹۸) در تحلیل مصرف آب شهری (خانگی) شهرستان بابل با استفاده از روش های داده کاوی به مسئله کمبود آب در ایران پرداخته است. در این مطالعه از کاوش قوانین وابستگی و الگوریتم درخت تصمیم به منظور تحلیل داده های موجود در زمینه مصرف آب شهری محله های شهرستان بابل استفاده شده است. با بهره گیری از یک درخت تصمیم، عوامل مرتبط با مصرف، طبقه بندی و میزان مصرف آب بخشی از مشترکین به عنوان داده های تست پیش بینی شده است. با بررسی میزان مصرف آب خانوار، محله های با بیشترین میزان مصرف که محله های ساحلی رودخانه بابل رود هستند، به عنوان محله های پرمصرف شناسایی شده اند. ساختمان ها در این محله ها، ویلایی و دارای تراکم جمعیت پایین تر و مساحت فضای سبز و حیاط بالاتری نسبت به سایر محله ها هستند.

با بررسی و مرور پژوهش های داده کاوی انجام شده در شهرداری ها کاربردهای گوناگون داده کاوی در شهرداری ها مشخص گردید و پایگاه های داده مورد استفاده در این پژوهش ها، بررسی شده است. نتیجه مطالعات نشان می دهد که شهرداری ها مجموعه داده های متنوع و مفیدی دارند که در صورت استفاده از آن ها و شناسایی الگوها و دانش پنهان در آن ها، می توانند کمک قابل توجهی در فرآیند مدیریت شهری و تصمیم گیری ها داشته باشند و در سیاست گذاری های شهرداری و نهایتاً در بسترسازی ایجاد شهر هوشمند مورد استفاده قرار گیرند. قلمرو این مقاله از منظر مکانی-زمانی، در حوزه ی کاربرد داده کاوی در مدیریت سازمان فن آوری اطلاعات و ارتباطات در شهر تهران به علاوه دبیرخانه و مرکز تهران هوشمند است. محورهای ماموریتی و حوزه های کاری سازمان فاوا^۲ عبارت است از: سامانه ها و نرم افزارهای مدیریت شهری، سرویس های شهروندی و خدمات پرداخت، راهبری شهر هوشمند، داده و آمار و رصد شهری، زیرساخت شبکه برق اضطراری و سخت افزار، برنامه ریزی و آموزش و پژوهش و خدمات مدیریت، مرکز داده

است و توسط دیوان محاسبات اسپانیا، که داده های مربوط به حساب های سالیانه شهرداری های اسپانیا را منتشر می کند، جمع آوری شده است. نتایج حاصل از این پژوهش در تصمیم گیری مسئولین و سیاست گذاران بسیار کمک کننده خواهد بود، زیرا این پژوهش عواملی را شناسایی کرده است که تأثیر زیادی روی تداوم خدمات عمومی و اثربخشی اقدامات انجام شده برای تحقق اهداف ثابت در بودجه و پایداری مالی، دارد. مینایی بیدگلی (۱۳۸۸) در پژوهشی با استفاده از داده کاوی در مدیریت ارتباط با شهروند به داده کاوی سامانه ۱۳۷ شهرداری تهران پرداخته است. در مدیریت ارتباط با شهروند، هدف اصلی شناخت نیازها و خواسته های شهروندان و تلاش در جهت بهبود ارائه خدمات و افزایش سطح رضایت مندی ایشان است. شهرداری تهران، سامانه ای با عنوان سیستم تحلیل اطلاعات ۱۳۷ ایجاد نموده که مرکز ساماندهی و رسیدگی به معضلات و مشکلات شهری است و پل ارتباطی بین شهروندان و شهرداری محسوب می شود. اطلاعات حاصل از این سامانه، حاوی مطالب مفیدی در مورد خدمات ارائه شده به شهروندان است و می تواند به عنوان منبعی مهم و مناسب برای مدیریت روابط تحلیلی با شهروند مورد بررسی و کاوش قرار بگیرد. انتظار می رود نتایج به دست آمده، در مدیریت مشکلات شهری و افزایش سطح رضایت مندی شهروندان مثرتر واقع شود. آخوندزاده و همکاران (۱۳۹۲) مدل جدیدی برای مدیریت شهری را با استفاده از روش های داده کاوی پیشنهاد دادند. دولت ها موظفند نیازهای شهروندان را به درستی شناسایی نموده و با پیش بینی آن از بروز مسائل و مشکلات احتمالی جلوگیری کنند. در این راستا، استفاده از تکنیک ها و ابزارهای داده کاوی می تواند بسیار مثرتر واقع شود. سامانه ۱۳۷ 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 می شود. در این مقاله با استفاده از الگوریتم آپریوری^۱ قوانین انجمنی موجود در بازه های زمانی/مکانی رخداد رخداد مشکلات به دست می آید. از قوانین به دست آمده می توان

۲. فاوا عبارتی کلی در برگیرنده تمام فناوری های پیشرفته نحوه ارتباط و انتقال انتقال داده ها در سامانه های ارتباطی است.

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)^۲

نحوه 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^۳ (مربوط به اقدامات هر کاربر)؛
۲. سطح بازدید^۴ (هر بازدید منحصر به فرد)؛
۳. سطح بازدید صفحات^۵ (هر بازدید منحصر به فرد از صفحه)؛
۴. سطح رویداد^۶ (کلیک دکمه، نمایش ویدیو، و غیره) (موسوی نظنزی، ۱۳۹۸).

۲. ژانویه. دانشگاه استنفورد. «Origin of the name "Google"»

سپتامبر دریافت شده در ۱۶ در ۴ ژوئیه ۲۰۱۲ بایگانی شده از اصلی ۲۰۰۴.

۲۰۱۱.

3. User level

4. Session level

5. Pageview level

6. Event level

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 شده است، 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 این که در طرح مقاله حاضر داده‌های مورد استفاده درون پایگاه‌های داده بوده و اینکه براساس مدل‌های خوشه‌بندی و طبقه‌بندی داده‌های 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^۱ است. گوگل آنالیتیکس یک ابزار تحلیلی وب رایگان است که

1. Google Analytics

ارزشمندانشان برخوردار می‌کنند. داده‌های 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 داده‌های وب‌سایت شهرداری تهران ارائه شده است.

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

جدول ۱: مجموعه داده‌های وب‌سایت شهرداری تهران

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

آلودگی هوا	۰/۴۰۴۷۸۵۹۴	7712	426	1
شاخص آلودگی هوا تهران	۰/۳۹۱۷۶۲۳۴	5070	488	1
شاخص آلودگی 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	۰/۲۸۸۸۸۲۵۶	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

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

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 برجسب اولیه به داده‌ها ارائه شده‌است.

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.

شبکه عصبی معمولی برابر با ۰/۰۷۵، معیار ام.اس.ای الگوریتم شبکه عصبی معمولی برابر با ۰/۰۶۴ و در نهایت معیار آرام.اس.ای الگوریتم شبکه عصبی معمولی برابر با ۰/۱۵۹ می‌باشد. الگوریتم درخت تصمیم

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 شبیه‌سازی 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

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.

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

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 شبیه‌سازی الگوریتم کی.ان.ان

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.

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.

Cluster Model

Cluster 0: 1780 items
 Cluster 1: 1790 items
 Cluster 2: 1830 items
 Total number of items: 5400

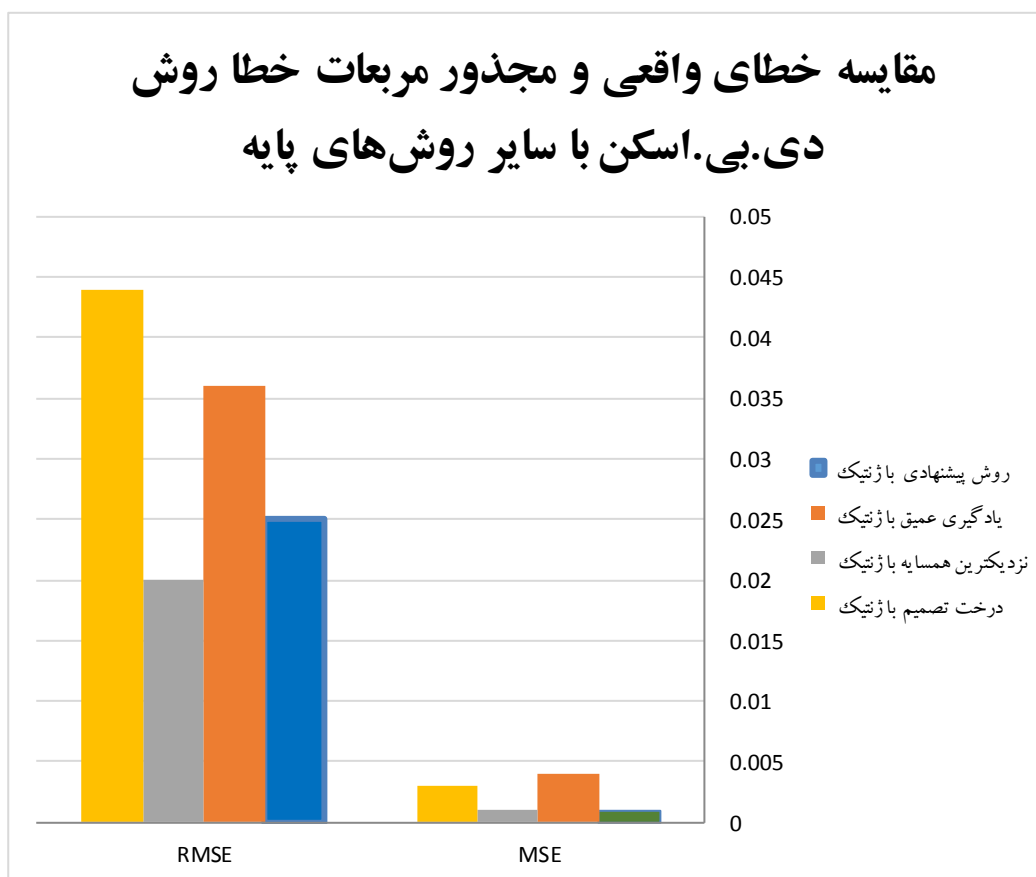
شکل ۴: تعداد خوشه ها در روش دی.بی.اسکن

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 $0.003/0$ و روش درخت تصمیم برابر با $0.003/0$ است. با این تفاسیر میزان مجذور مربعات خطای روش دی.بی.اسکن با روش درخت تصمیم برابر با $0.045/0$ 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.

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 شده است. 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 خوشه‌ها را ۲ و حد بالا برابر با ۵۰ قرار داده است. با اجراهای مختلف الگوریتم خوشه‌بندی ایکس-مینز تعداد 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. بنابراین، بر اساس شماره خوشه‌ها می‌توان تعیین کرد که کلیدواژه جدید ارائه شده



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

تهران و ارائه راهکارهای داده کاوی بهتر عمل کرده و از این استراتژی که ترکیبی از تکنیک های خوشه بندی دی.بی.اسکن و شبکه عصبی عمیق ال.اس.تی.ام است، می توان اعتماد کرد و در شهرداری تهران مورد استفاده قرار داد. یافته ها نشان داد که وبسایت های شهرداری تهران همکاری مطلوبی با یکدیگر نداشته و موازی کاری در آنها بسیار است که این خود موجب اتلاف وقت و انرژی و منابع شهرداری ها شده و موجب اختلال در اجرای تصمیمات مدیران می گردد. از آن جا که این وبسایت ها دارای مطالب مشترک در زمینه های یکسان است، در صورت تبادل لینک، احتمال دیده شدن آنها افزایش می یابد. از آن جا که اطلاعات باید جدید و به روز است، می توان با پشتیبانی و به روزرسانی مرتب امکان مراجعه و تعامل شهروندان با وبسایت های شهرداری را افزایش داد و مراجعه فیزیکی و ازدحام ارباب رجوع به شهرداری ها را کم کرد. مقایسه پیشنهادها بررسی داده های وبسایت های شهرداری تهران با این گستره که تمامی کلیدواژه ها و کلیک ها و بازدیدها را مورد توجه قرار دهد

الگوهای قابل حصول و قابل پیش بینی از طریق راه کارهای داده کاوی در وبسایت های شهرداری تهران کدام است؟ استفاده از الگوهای به دست آمده در فرآیند داده کاوی وبسایت های شهرداری تهران چگونه قابل استفاده است؟ با اعمال الگوریتم خوشه بندی دی.بی.اسکن و خوشه بندی اولیه کلیدواژه های مستخرج از وبسایت های شهرداری تهران یک الگوی اولیه به دست می آید. در واقع خوشه ها ارزش کلیدواژه ها بر اساس میزان امتیاز داده شده، تعداد کلیک های انجام شده بر روی هر کلیدواژه و تعداد مشاهدات آنها را مشخص می کنند. بنابراین براساس الگوی وزندهی با استفاده از الگوریتم خوشه بندی و سپس الگوریتم یادگیری شبکه عصبی عمیق اقدام به شناسایی و تجزیه و تحلیل داده های وبسایت های شهرداری تهران و ارائه راهکارهای داده کاوی می گردد. با توجه به یافته های تحقیق و نتایج به دست آمده مشاهده گردید که دقت، صحت و فراخوانی و سایر یافته های تحقیق نسبت به سایر روش ها جهت تجزیه و تحلیل داده های وبسایت های شهرداری

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

مینایی بیدگلی، ب.؛ آخوندزاده، ا.؛ موسوی، ح. و دیگران (۱۳۸۸). استفاده از داده کاوی در مدیریت ارتباط با شهروندان: 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, (201۶). Evaluation of the maturation of individual Dengue virions with flow virometry. *virology*, 488,20-27

صرفاً در این پژوهش انجام شده است و در پژوهش انجام شده در گذشته در خصوص تجزیه و تحلیل داده‌های ثبت شده 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 بوده است و تاکنون در هیچ پژوهشی انباره عظیم داده‌های وبسایت‌های شهرداری تهران در پایگاه داده‌ای مورد تجزیه و تحلیل و داده کاوی قرار نگرفته است.

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 و سیما، ۸-۹ بهمن، ۲-۱۵.

آهنگرکانی، م.؛ خواسته، ح. (۱۳۹۸). تحلیل مصرف آب شهری (خانگی) شهرستان بابل با استفاده از روش های داده کاوی. *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 الکترونیک، ۶ (۲۲)، ۷-۲۲.

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

دانشگاه استنفورد (۲۰۰۴). *Origin of the name "Google"*.

بایگانی شده از اصلی در ۴ ژوئیه ۲۰۱۲. دریافت شده در ۱۶ سپتامبر

http://graphics.stanford.edu/~dk/google_name_origin.html

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

Shayesteh Shojaei karizaki¹ | Sudabeh Shapoori² | Hajar Zarei³

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