

ارائه یک روش جامع ارزیابی داده‌های دولتی باز با هدف بهبود کیفیت داده‌ها و افزایش اقبال شهروندان

محمد مرادی: دکتری مهندسی فناوری اطلاعات، پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران.

*مجتبی مازوچی: استادیار، پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران. (نویسنده مسئول) Mazoochi@itrc.ac.ir

چکیده

نوع مقاله: مقاله پژوهشی

دریافت: ۱۴۰۰/۰۹/۲۱

پذیرش: ۱۴۰۰/۱۱/۲۷

هدف: هدف پژوهش حاضر، ارائه یک روش ارزیابی داده‌های دولتی باز با در نظر گرفتن ابعاد و شاخص‌های جامع و کامل، محاسبه وزن و اهمیت هر شاخص، بررسی وضعیت کشور در این حوزه، خوشه‌بندی سازمان‌ها و ارائه یک مدل طبقه‌بندی به منظور پیش‌بینی وضعیت سازمان‌ها در ارائه داده‌های دولتی بصورت باز است.

روش پژوهش: این پژوهش بر مبنای هدف، از نوع کاربردی است. از مطالعات کتابخانه‌ای به منظور استخراج ابعاد و شاخص‌های ارزیابی استفاده شده است. جامعه آماری شامل کلیه مقالات مرتبط با شاخص‌های ارزیابی داده‌های دولتی باز است. از روش نمونه‌گیری احتمالی از نوع نمونه‌های تصادفی ساده استفاده شده است و ۱۰ مقاله در این حوزه مورد بررسی قرار گرفته است. روش میدانی و استفاده از تکنیک‌های تصمیم‌گیری چندشاخصه به منظور محاسبه وزن و اهمیت هر شاخص به کار گرفته شده است. پس از استخراج داده‌های سازمان‌ها از دو سامانه کاتالوگ ملی و مجموعه داده‌های باز و کاربردی و سامانه انتشار و دسترسی آزاد به اطلاعات، از تکنیک‌های داده کاوی به منظور خوشه‌بندی و ایجاد یک مدل طبقه‌بندی استفاده شده است.

یافته‌ها: بر اساس مطالعات و بررسی مقالات، ۱۵ شاخص ارزیابی داده‌های دولتی باز شامل «صالت داده‌ها»، «باز بودن مجوز»، «بروز بودن»، «میزان دسترسی به داده»، «کامل بودن فراداده»، «تعداد مجموعه داده»، «باز بودن قالب»، «تبعیض‌آمیز نبودن»، «قابل فهم بودن»، «تعداد دسته‌های داده»، «رایگان بودن»، «عدم وجود داده‌های از دست رفته»، «امکان درخواست داده»، «تجسمی بودن» و «بازخوردپذیری» استخراج شدند. با استفاده از تکنیک تحلیل سلسله مراتبی، وزن شاخص‌ها محاسبه شد که پس از نرمال‌سازی، مجموع وزن ۱۵ شاخص استخراج شده برابر با یک بود. شاخص‌های «صالت داده‌ها» با وزن ۰.۱۶۵، «باز بودن مجوز» با وزن ۰.۱۲۴ و «بروز بودن» با وزن ۰.۱۰۹ به ترتیب رتبه‌های اول تا سوم را در بین ۱۵ شاخص ارزیابی کسب کردند. همچنین بر اساس وزن شاخص‌های ارزیابی بدست آمده و استخراج داده‌های ۳۵۸ سازمان مطابق با ۱۵ شاخص ارزیابی استخراج شده از دو سامانه کاتالوگ ملی و مجموعه داده‌های باز و کاربردی و سامانه انتشار و دسترسی آزاد به اطلاعات به محاسبه وزن سازمان‌ها پرداخته شد که پس از نرمال‌سازی، مجموع وزن‌ها برابر با یک بود. «سازمان جهاد کشاورزی آذربایجان شرقی» با وزن ۰.۰۸۸، «مرکز آمار ایران» با وزن ۰.۰۶۲ و «سازمان زمین‌شناسی» با وزن ۰.۰۵۸ به ترتیب رتبه‌های اول تا سوم در بین ۳۵۸ سازمان و موسسه دولتی بررسی شده بر اساس ترکیب شاخص‌ها و وزن شاخص‌ها بودند.

نتیجه‌گیری: نتایج این پژوهش شامل استخراج جامع شاخص‌های ارزیابی، محاسبه وزن و اهمیت هر شاخص، بررسی وضعیت موجود سازمان‌ها و موسسات دولتی داخل کشور و مدل طبقه‌بندی ایجاد شده می‌تواند به مدیران در شناخت وضع موجود و بهبود آن و در نتیجه افزایش تعامل شهروندان با داده‌های دولتی باز به عنوان نوعی از تعامل انسان و اطلاعات کمک نماید.

کلیدواژه‌ها: داده دولتی باز، شاخص ارزیابی، خوشه‌بندی، طبقه‌بندی

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: حامی مالی نداشته است.

شیوه استناد به این مقاله

APA: Moradi, M., Mazoochi, M., (2022). A comprehensive method for evaluating open government data with the aim of improving data quality and increasing citizens' willingness. *Human Information Interaction*, 8(4);47-65. (Persian)

Vancouver: Moradi, M., Mazoochi, M., A comprehensive method for evaluating open government data with the aim of improving data quality and increasing citizens' willingness. *Human Information Interaction*. 2022;8(4):47-65. (Persian)



انتشار مجله تعامل انسان و اطلاعات با حمایت مالی دانشگاه فوارزمی انجام می‌شود.

انتشار این مقاله به صورت دسترسی آزاد مطابق با CC BY-NC-SA 3.0 صورت گرفته است.

A Comprehensive Method of Evaluating Open Government Data with the Aim of Improving Data Quality and Increasing Citizens' Willingness

Mohammad Moradi: Ph.D., Information Technology, ICT Research Institute, Tehran, Iran.

***Mojtaba Mazoochi:** Assistant Professor, ICT Research Institute, Tehran, Iran. (Corresponding author) Mazoochi@itrc.ac.ir

Received: 12/12/2021

Accepted: 16/02/2022

Abstract

Purpose: The purpose is to present an open government data evaluation method by considering comprehensive and complete dimensions and criteria - calculating the weight and importance of each criterion, examining the country in this area, clustering organizations and presenting a classification model to predict the situation.

Methodology: Library studies was used to extract the dimensions and criteria of evaluation. Population includes articles related to open government data evaluation criteria. Ten articles were reviewed by simple random sampling method. Multiple attribute decision making techniques was used to calculate the weight and importance of each criterion. Data mining techniques was incorporated to cluster and create a classification model.

Findings: By reviewing the articles 15 criteria of open government data evaluation including: Data originality, license openness, up-to-datedness, data access rate, metadata completeness, number of data sets, format openness, non-discriminatory, comprehensible, number of data fields, free, no missing data, data request ability, visual and feedback, were extracted. Using AHP technique, the weights of the criteria were calculated, which after normalization, the total weight of the 15 extracted criteria was equal to one. "Data originality" with a weight of 0.165, "license openness" with a weight of 0.124 and "up-to-datedness" with a weight of 0.109 were ranked first to third among 15 evaluation criteria, respectively. Weight of evaluation criteria obtained and data extraction of 358 organizations in harmony with 15 evaluation criteria, the weight of organizations was calculated. The sum of the weights was equal to one. "East Azerbaijan Agricultural Jihad Organization" with a weight of 0.088, "Statistics Center of Iran" with a weight of 0.062 and "Geological Survey" with a weight of 0.058 were the first to third ranks among 358 organizations and government institutions, respectively, based on the combination of criteria and the weight of criteria.

Conclusion: Evaluation criteria obtained, calculating the weight and importance of each criterion, examining the current situation of government organizations and institutions in the country and the classification model created can help managers to understand the current situation and improve it and thus increase citizens' interaction with open government data as a kind of human information interaction.

Keywords: Open Government Data, Evaluation Criteria, Clustering, Classification.

Conflicts of Interest: None

Funding: None.

How to cite this article

APA : Moradi, M., Mazoochi, M., (2022). A comprehensive method for evaluating open government data with the aim of improving data quality and increasing citizens' willingness. *Human Information Interaction*, 8(4);47-65. (Persian)

Vancouver: Moradi, M., Mazoochi, M., A comprehensive method for evaluating open government data with the aim of improving data quality and increasing citizens' willingness. *Human Information Interaction*. 2022;8(4):47-65. (Persian)



مقدمه

داده‌های باز^۱ به داده‌های غیرمحرمانه گفته می‌شود که بدون هیچ محدودیتی در استفاده یا توزیع، در دسترس قرار می‌گیرند (جانسن و همکاران^۲، ۲۰۱۲). داده‌های دولتی باز^۳ زیرمجموعه‌ای از داده‌های باز است که به عنوان یک عامل اصلی برای دولت باز در نظر گرفته می‌شود (اتارد و همکاران^۴، ۲۰۱۵). دنیلز و همکاران^۵ (۲۰۱۷) داده‌های دولتی باز را به عنوان «خدمات داده، همراه با دسترسی و استفاده مجدد که با هدف اجازه دادن به اشخاص ثالث برای ایجاد ارزش جدید ایجاد شده است» تعریف کردند. داده دولتی باز ابزاری برای توانمندسازی شهروندان و دادن دسترسی و مجوز به آن‌ها برای استفاده از داده‌های تولید شده توسط بخش دولتی است، به طوری که آن‌ها می‌توانند از داده‌ها استفاده کنند، ذخیره کنند، توزیع مجدد کنند و آن‌ها را با سایر منابع داده ادغام کنند (دنیلز و لویز^۶، ۲۰۱۴). علاوه بر این، داده‌های دولتی باز را می‌توان به عنوان داده‌های متعلق به یک نهاد دولتی تعریف کرد که برای استفاده رایگان، استفاده مجدد و توزیع مجدد منتشر می‌شود (سوشا و همکاران^۷، ۲۰۱۵).

ارائه اطلاعات به صورت داده باز، کاهش فساد، کسب اعتماد عمومی و ایجاد جامعه مردم سالار را در پی دارد. داده‌های باز امکان بیشتری برای پایش فعالیت‌های حاکمیتی فراهم می‌کند. به عنوان مثال باعث می‌شود نحوه هزینه کردن بودجه، شفاف شود و اثرات آن مشخص گردد. همچنین شهروندان را به مشارکت بیشتر در نظارت بر حاکمیت تشویق می‌کند. در حوزه شرکتی، داده‌های باز در درجه اول به خود آن نهاد کمک می‌کند تا همه بخش‌های سازمان از وجود داده‌ها باخبر شوند و فعالیت‌های موازی و پرهزینه برای جمع‌آوری داده‌هایی که قبلاً توسط سازمان انجام شده است، صورت نگیرد. داده دولتی باز، این امکان را در اختیار شهروندان قرار می‌دهد تا در فرایندهای تصمیم‌گیری با رویه‌هایی آگاهانه و ساختاریافته مشارکت کنند (هالون^۸، ۲۰۱۲). داده دولتی باز می‌تواند محرک نوآوری و رشد اقتصادی شود و ارتقای مدل‌های کسب و کار را به دنبال داشته باشد (یو و رایینسون^۹، ۲۰۱۱).

با توجه به ضرورت ذکر شده، هدف این پژوهش ارائه یک روش ارزیابی جامع داده‌های دولتی باز، محاسبه وزن و اهمیت هر شاخص ارزیابی، بررسی وضعیت کشور در این حوزه و ارائه یک مدل طبقه‌بندی^{۱۰} به منظور پیش‌بینی وضعیت سازمان‌ها است.

عدم رعایت شاخص‌های مورد نیاز در ارائه داده‌ها بصورت باز، کاهش تعامل شهروندان با داده‌های دولتی باز را در پی خواهد داشت. لذا این پژوهش با استخراج شاخص‌ها و وزن و اهمیت هر شاخص به افزایش کیفیت داده‌های دولتی باز و در نتیجه افزایش تعامل شهروندان با داده‌های دولتی باز به عنوان نوعی از تعامل انسان و اطلاعات کمک خواهد کرد.

پیشینه پژوهش

دجوانا اسپینوسا و لوجان مورا^{۱۱} (۲۰۲۰) به ارزیابی پورتال‌های داده دولتی باز در اتحادیه اروپا از سال ۲۰۱۵ تا ۲۰۱۷ پرداخته‌اند. این مطالعه، داده‌های جمع‌آوری شده پورتال‌های داده‌های دولتی باز در ۲۸ کشور اتحادیه اروپا را ارائه می‌دهد. چندین پارامتر و شاخص مشاهده شده در طی یک دوره ۳ ساله در پورتال‌های رسمی داده‌های ملی باز، برای ایجاد این مجموعه داده شناسایی و ضبط شده است. داده‌ها به صورت دستی از منابع موجود اطلاعات عمومی و پورتال‌های رسمی داده‌های دولتی باز که به صورت آزاد در وب موجود است، به دست آمده است. در این پژوهش، شاخص‌های «وجود پیوند از پورتال داده دولتی باز به سایت منبع ارائه دهنده مجموعه داده»، «وجود پلاگین‌های شبکه‌های اجتماعی» به منظور بحث در مورد تجارب کاربران در استفاده از پورتال داده دولتی باز، «پشتیبانی از فرمت‌های مختلف مجموعه داده‌ها» و «امکان جستجو و فیلتر داده‌ها» به عنوان شاخص‌های ارزیابی پورتال‌های داده‌های دولتی باز پیشنهاد شده‌اند.

نیکفوروا و امسی براید^{۱۲} (۲۰۲۰) به پژوهشی با عنوان «قابلیت استفاده از پورتال داده‌های دولتی باز: تحلیل کاربر محور ۴۱ پورتال داده‌های دولتی باز» پرداخته‌اند. با تأیید اهمیت قابلیت استفاده از پورتال برای فرآیند استفاده مجدد از داده‌ها، این مطالعه با طرح دو سوال به توضیح برخی از بینش‌های اولیه کمک می‌کند: «چگونه می‌توان قابلیت استفاده از پورتال‌های داده دولتی باز را ارزیابی کرد و آن‌ها را در زمینه‌های مختلف مقایسه کرد؟» و «معمولاً جنبه‌های کاربردی موجود در پورتال‌های داده باز دولتی چه مواردی هستند؟». برای پاسخ به این سوالات تحقیق، مجموعه‌ای از ۴۱ پورتال داده دولتی باز برای تجزیه و تحلیل قابلیت استفاده با توجه به بازخورد ۴۰ کاربر انتخاب شده است. بر اساس نتایج این پژوهش، عدم وجود تعامل بین کاربران با پورتال‌های داده‌های دولتی باز در مواردی مانند ارائه بازخورد یا

⁷ Susha et al

⁸ Halonen

⁹ Yu & Robinson

¹⁰ Classification

¹¹ De Juana-Espinosa & Luján-Mora

¹² Nikiforova & McBride

¹ Open Data

² Janssen et al

³ Open Government Data (OGD)

⁴ Attard et al

⁵ Danneels et al

⁶ Daniels & Lopez

میسوراکا و ویسکوسی^۴ (۲۰۱۴) در مورد چارچوب ارزیابی انطباق داده‌های دولتی باز مبتنی بر کیفیت بحث کرده‌اند. این معیارها شامل سه بعد کیفیت مختلف هستند: کامل بودن، دقیق بودن و به موقع بودن. هاریسون و همکاران (۲۰۱۲) بر ارزیابی کیفیت فراداده متمرکز شده‌اند.

ژنگ و ژیاو^۵ (۲۰۲۰) به بررسی چارچوب ارزیابی کیفیت داده‌های دولتی باز پرداخته‌اند. هدف این مطالعه ایجاد یک چارچوب مشترک به عنوان مرجع برای ارزیابی کیفیت داده‌های دولتی باز می‌باشد. در این پژوهش به تلفیق ۱۰ مطالعه کیفی در یک چارچوب مرجع مشترک برای ارزیابی کیفیت داده‌های دولتی باز پرداخته شده است. بر اساس یک تحلیل هفت مرحله‌ای، یک چارچوب مرجع مشترک برای ارزیابی کیفیت داده‌های دولتی باز ارائه شده است که شامل شش شاخص دقت، دسترسی، کامل بودن، به موقع بودن، ثبات و قابل درک بودن می‌باشد.

ژنگ و همکاران^۶ (۲۰۲۰) در پژوهشی با عنوان «ارزیابی داده‌های دولتی جهانی باز: روش‌ها و وضعیت» ابتدا با مقایسه و تحلیل منظم چارچوب‌ها، شاخص‌ها و روش‌های آن‌ها، هفت روش ارزیابی داده دولتی باز را بررسی می‌کنند. بر اساس این تجزیه و تحلیل، چارچوبی برای ارزیابی عملکرد داده‌های دولتی باز برای همه کشورهای عضو سازمان ملل ایجاد شده است. طبق نتایج، بیشتر برنامه‌های ارزیابی فعلی بر داده و بنیاد تمرکز دارند و توجه کمتری به بسترهای نرم افزاری، استفاده و تأثیر دارند. این مطالعه نشان می‌دهد که در سال ۲۰۱۸، ۳۴ کشور (۱۸٪) نمره «بسیار بالا»، ۴۰ کشور (۲۱٪) نمره «بالا»، ۴۳ کشور (۲۲٪) نمره «میان»، در حالی که ۷۶ کشور (۳۹٪) امتیاز «کم» را دریافت کرده‌اند.

ابوالمعالی و همکاران (۱۳۹۹) به ارائه الگویی برای شناسایی پیش‌ران‌های کاربست مدل حکومت باز در دستگاه‌های اجرایی پرداختند. نویسندگان بیان می‌کنند که انتشار داده‌ها به تنهایی کافی نیست. انتشار داده‌های بی معنا هم فاقد کارایی لازم برای شهروندان است. لذا در این پژوهش از روش کتابخانه‌ای و مصاحبه با خبرگان برای شناسایی ابعاد موثر در بهبود کیفیت استفاده شده است. استخراج بُعد فناوری شامل شناسایی سریع‌تر نواقص سیستم، بُعد دولت به مثابه پلتفرم شامل خدمت‌رسانی بهتر با وب ۲.۰ و اپلیکیشن‌های دولتی، بُعد مشارکت واقعی مردم شامل بازخوردگیری، جذب ایده‌های بکر، سازوکارهای شکایت،

درخواست مجموعه داده از مشکلات اصلی پورتال‌های داده‌های دولتی باز است. بنابراین دولت‌ها باید روی توسعه اکوسیستم‌های داده‌های دولتی باز و افزایش قابلیت تعامل در این پورتال‌ها تمرکز نمایند.

دهبی و همکاران^۱ (۲۰۱۸) به پژوهشی با عنوان «مدل ارزیابی پورتال‌های داده دولتی باز» پرداخته‌اند. نویسندگان در این پژوهش یک مدل ارزیابی برای پورتال‌های داده‌های دولتی باز بر اساس پنج بعد اصلی که تأثیر زیادی بر کاربرد آن‌ها دارند مشخص می‌کنند. ابعاد مشخص شده عبارتند از: غنای اطلاعاتی که با انطباق پورتال با نیازهای کاربر از نظر محتوا سروکار دارد؛ قابلیت کشف که با ابزارها و مکانیزم‌هایی که باعث افزایش دسترسی به داده‌ها در پورتال می‌شود، مرتبط است؛ قابلیت استفاده مجدد، که با باز بودن داده‌های منتشر شده در پورتال و امکان استفاده مجدد از آن‌ها سروکار دارد؛ تعامل، که با باز بودن پورتال برای بازخورد کاربر، همکاری و تعامل با داده‌های منتشر شده مرتبط است؛ و کیفیت داده‌ها که با کیفیت داده‌های منتشر شده در پورتال، سروکار دارد. مدل ارزیابی پیشنهادی برای انجام ارزیابی چهار پورتال داده‌های دولتی باز ملی استفاده شده است. وترو و همکاران^۲ (۲۰۱۶) رویکردی را برای اندازه‌گیری کیفیت مجموعه داده‌های دولتی باز ارائه می‌دهند. آن‌ها از کاربران حرفه‌ای داده‌های دولتی باز در مورد چالش‌های موجود در کیفیت داده‌های دولتی باز سوال کردند. به طور کلی آن‌ها پیشنهاد می‌کنند که مجموعه داده‌ها از نظر کامل بودن، دقت، قابل ردیابی، قابل فهم بودن، انطباق و انقضاء ارزیابی شوند. به عبارت دیگر، مجموعه داده ایده آل باید شامل داده‌های کامل و دقیق باشد، در قالب قابل خواندن توسط ماشین باشد، دارای فراداده باشد، بروزرسانی شود، دقیق باشد و از نظر منبع قابل پیگیری باشد. این یک چارچوب کاملاً کامل و مفید برای سیاست‌گذاری است که مایل به کنترل کیفیت مجموعه داده‌ها در پروژه‌ها یا برنامه‌های داده‌های دولتی باز هستند.

داوس و همکاران^۳ (۲۰۱۶) چارچوبی را برای ارزیابی کیفیت پورتال‌های داده باز در سطح ملی معرفی کردند و مجموعه‌ای از معیارها را برای ارزیابی مشکلات کیفیت داده در پورتال‌های داده‌های دولتی باز ارائه دادند. این معیارها بر روی ۱۲ پورتال اعمال شد و چندین بعد کیفیت داده معرفی شد. این ابعاد شامل وجود استاندارد در قالب‌های داده، وجود فراداده، قابلیت خواندن توسط ماشین و بروز بودن داده‌ها بود.

⁴ Misuraca & Viscusi

⁵ Zhang & Xiao

⁶ Zheng et al

¹ Dahbi et al

² Vetrò et al

³ Dawes et al

تجدید نظر و پاسخگویی آنلاین و بُعد شفافیت شامل رشد شبکه‌های اجتماعی از نتایج این پژوهش بوده است.

مومن کاشانی و همکاران (۱۳۹۹) به ارائه مدل سنجش میزان تمایل خطمشی‌گذاران به دولت باز در ایران پرداختند. هدف این پژوهش ارائه مدلی برای سنجش میزان تمایل خطمشی‌گذاران به دولت باز و بررسی تفاوت سطح تمایل به دولت باز در گروه‌های مختلف خطمشی‌گذاران است. در مرحله کیفی، روش فراترکیب و در مرحله کمی، انجام پیمایش و تحلیل داده‌های پرسشنامه از خطمشی‌گذاران به کار گرفته شده است. نویسندگان بیان می‌کنند تمایل به دولت باز سازه‌ای چندبُعدی است که شامل ابعاد شفافیت، مشارکت شهروندان، پاسخگویی دولت و پشتیبانی از نوآوری است. بر اساس نتایج این پژوهش، میانگین میزان تمایل خطمشی‌گذاران به دولت باز در گروه سنی کم‌تر از ۳۰ سال، بالاتر است. همچنین بین میانگین میزان تمایل به دولت باز بین مردان و زنان، تفاوت آماری معناداری وجود ندارد.

همان‌طور که در پیشینه پژوهش ملاحظه می‌شود، هر یک از پژوهش‌های انجام شده بر بعد ارزیابی خاصی تمرکز کرده‌اند. همچنین در پژوهش‌های صورت گرفته، وزن و اهمیت هر یک از ابعاد و شاخص‌های ارزیابی مشخص نشده‌اند. در این پژوهش سعی شده است بدهای ارزیابی مختلف و اهمیت هر یک مد نظر قرار گیرد. همچنین علاوه بر بررسی وضعیت کشور در این حوزه و خوشه‌بندی^۱ سازمان‌ها، به ارائه یک مدل طبقه‌بندی نیز پرداخته شده است.

روش پژوهش

در این پژوهش، نوع تحقیق بر مبنای هدف از نوع تحقیقات کاربردی است. در ابتدا از مطالعات کتابخانه‌ای به منظور استخراج ابعاد و شاخص‌های ارزیابی استفاده شده است. جامعه آماری شامل کلیه مقالات مرتبط با شاخص‌های ارزیابی داده‌های دولتی باز است. از روش نمونه‌گیری احتمالی از نوع نمونه‌های تصادفی ساده استفاده شده است و ۱۰ مقاله در این حوزه مورد بررسی قرار گرفته است. سپس روش میدانی و استفاده از تکنیک‌های تصمیم‌گیری چندشاخصه به منظور محاسبه وزن و اهمیت هر شاخص به کار گرفته شده است. در فاز بعدی، سازمان‌ها، شامل تمامی سازمان‌ها و موسسات دولتی حاضر در دو سامانه کاتالوگ ملی و مجموعه داده‌های باز و کاربردی و سامانه انتشار و دسترسی آزاد به اطلاعات که به ارائه داده‌ها بصورت باز می‌پردازند مورد بررسی قرار گرفته‌اند و داده‌های مرتبط با هر

شاخص ارزیابی، استخراج شده است. تعداد سازمان‌ها و موسسات دولتی بررسی شده، ۳۵۸ مورد بوده است. سپس سازمان‌ها بر اساس شاخص‌های ارزیابی و وزن شاخص‌ها، رتبه‌بندی شده‌اند. در نهایت از تکنیک‌های داده کاوی به منظور خوشه‌بندی و ایجاد یک مدل طبقه‌بندی استفاده شده است. نمودار فرایند تحقیق در شکل ۱ نشان داده شده است.



شکل ۱. نمودار فرایند تحقیق

یافته‌های پژوهش

استخراج ابعاد و شاخص‌های ارزیابی

در این بخش، شاخص‌های ارزیابی با سه بخش داده باز، شفافیت داده و تعامل پس از مطالعات کتابخانه‌ای انجام شده، بیان شده است. مراجع مرتبط با هر بعد و شاخص ارزیابی نیز مشخص شده است.

داده باز

۱- غنای اطلاعات (دهبی و همکاران^۲، ۲۰۱۸): غنای اطلاعات، رفع نیازهای کاربر را از نظر مقدار داده بررسی می‌کند. در ادامه شاخص‌های مرتبط با بعد غنای اطلاعات بیان شده است.

الف) تعداد مجموعه داده (دهبی و همکاران^۲، ۲۰۱۸؛ هوانگ و همکاران^۳، ۲۰۱۹؛ وترو و همکاران^۴، ۲۰۱۶؛ ساکسنا^۵، ۲۰۱۸): این شاخص تعداد مجموعه داده‌هایی که یک سازمان بصورت باز ارائه کرده است را ارزیابی می‌کند.

ب) تعداد دسته‌های داده (ساکسنا، ۲۰۱۸)

^۴ Vetrò et al

^۵ Saxena

^۱ Clustering

^۲ Dahbi et al

^۳ Huang et al

الف) اصلات داده‌ها (ولجکاویک و همکاران، ۲۰۱۴؛ وترو و همکاران، ۲۰۱۶)

ب) عدم وجود داده‌های از دست رفته (دهبی و همکاران، ۲۰۱۸)
ج) بروز بودن (ولجکاویک و همکاران، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸؛ هوانگ و همکاران، ۲۰۱۹؛ سایت داده باز جهانی^۳)
این شاخص، بروز بودن داده‌ها را برای مجموعه داده‌های منتشر شده ارزیابی می‌کند که براساس فراداده زمانی به ویژه تاریخ ایجاد مجموعه داده و تاریخ به روزرسانی مجموعه داده‌ها استوار است. نمره هر سازمان O_i بر اساس داده‌هایی که در پنج سال اخیر منتشر کرده است (T_i داده‌هایی که در j سال پیش منتشر شده است) با استفاده از فرمول ۱ محاسبه می‌شود.

$$O_i = \sum_{j=0}^4 (1 - \frac{j}{10}) \times T_j \quad \text{فرمول ۱}$$

شفافیت داده

۱- قابل استفاده مجدد بودن (ولجکاویک و همکاران، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸)؛ ارزش داده‌های دولتی باز فقط پس از استفاده مجدد از آن تحقق می‌یابد (سوشا و همکاران^۴، ۲۰۱۵). زمانی داده دولتی باز قابل استفاده مجدد در نظر گرفته می‌شود که داده تحت مجوز باز منتشر شود و اجازه دسترسی، استفاده مجدد و توزیع مجدد داده را بدون محدودیت صادر کند. همچنین بایستی در قالب الکترونیکی منتشر شود و قابل خواندن توسط ماشین باشد. قابلیت استفاده مجدد همچنین با ویژگی‌هایی سروکار دارد که راهی آسان برای استفاده مجدد از داده‌ها مانند برنامه‌ها و API (رابط برنامه نویسی برنامه) فراهم می‌کند. در ادامه شاخص‌های مرتبط با بعد قابل استفاده مجدد بودن بیان شده است.

الف) باز بودن مجوز (ولجکاویک و همکاران، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸؛ سایت داده باز جهانی)؛ این شاخص، باز بودن مجوز مجموعه داده برای استفاده مجدد را ارزیابی می‌کند.
ب) باز بودن قالب (ولجکاویک و همکاران، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸؛ ساکسنا، ۲۰۱۸؛ سایت داده باز جهانی)؛ این شاخص باز بودن قالب منابع را ارزیابی می‌کند. برای هر مجموعه داده D_n ، نمره FOI_n بر اساس قالب منبع، به صورت زیر تخصیص می‌یابد:

- اگر قالب غیرقابل خواندن توسط ماشین است: $FOI_n = 0$ (به عنوان مثال PDF)

ج) امکان درخواست داده (دهبی و همکاران، ۲۰۱۸؛ ژنگ و همکاران، ۲۰۲۰)؛ این شاخص میزان باز بودن نسبت به درخواست کاربران را ارزیابی می‌کند. به عبارت دیگر امکان اجازه به کاربران برای درخواست مجموعه داده‌های جدید را بررسی می‌کند. با توجه به امکان وجود درخواست داده، دو مقدار ۱ یا ۰ تعلق داده می‌شود.

۲- قابل کشف بودن (دهبی و همکاران، ۲۰۱۸)؛ قابلیت کشف با ابزارها و مکانیزمی که دسترسی و پیمایش داده‌ها را افزایش می‌دهند، سروکار دارد. به عبارت دیگر کاربران باید بتوانند مجموعه داده‌های مربوطه را به روشی ساده و کارآمد جستجو و دسترسی پیدا کنند. اگر فراداده ارائه نشود، این امر محقق نمی‌شود. فراداده، درک بهتری از اهمیت داده و ساختار داده را فراهم می‌کند و به کاربران کمک می‌کند تا به داده‌های مربوط به نیاز خود دسترسی پیدا کنند (اتارد و همکاران^۵، ۲۰۱۵). ارزیابی بعد قابل کشف بودن مستلزم ارزیابی کامل فراداده توصیفی و وجود ویژگی‌های دسترسی به داده است. در ادامه شاخص‌های مرتبط با بعد قابل کشف بودن بیان شده است.

الف) کامل بودن فراداده (دهبی و همکاران، ۲۰۱۸؛ ژنگ و همکاران، ۲۰۲۰؛ هوانگ و همکاران، ۲۰۱۹؛ ساکسنا، ۲۰۱۸)؛ این شاخص کامل بودن فراداده توصیفی را ارزیابی می‌کند. برای هر مجموعه داده، کامل بودن فیلدهای فراداده توصیفی داده ارزیابی می‌شود. این فیلدها عبارت‌اند از: عنوان، توضیحات، برچسب‌ها، ناشر و غیره.

ب) میزان دسترسی به داده (دهبی و همکاران، ۲۰۱۸؛ ژنگ و همکاران، ۲۰۲۰؛ هوانگ و همکاران، ۲۰۱۹؛ وترو و همکاران، ۲۰۱۶)؛ این شاخص وجود ویژگی‌هایی را ارزیابی می‌کند که کشف داده‌ها را افزایش می‌دهد، به ویژه وجود سه ویژگی: جستجو، مرتب‌سازی و فیلتر، که مقادیر خود را در محدوده [۰، ۱] با توجه به وجود این ویژگی‌ها دریافت می‌کنند.

۳- کیفیت داده‌ها (ولجکاویک و همکاران^۶، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸)؛ این بعد با کیفیت داده‌ها سروکار دارد. موارد از دست رفته یا فاقد داده، مانع استفاده و استفاده مجدد از داده‌ها می‌شود و تأثیر زیادی بر کیفیت برنامه‌هایی دارد که از داده‌ها مجدداً استفاده می‌کنند. در ادامه شاخص‌های مرتبط با بعد کیفیت داده‌ها بیان شده است.

³ <https://index.okfn.org/place/>

⁴ Susha et al

⁵ Attard et al

⁶ Veljković et al

- اگر قالب قابل خواندن توسط ماشین است: $FOI_n = 1$ (به عنوان مثال JSON یا CSV)

(ج) رایگان بودن (سایت داده باز جهانی)

(د) تبعیض آمیز نبودن (ولجکاویک و همکاران، ۲۰۱۴؛ وترو و همکاران، ۲۰۱۶): امکان دسترسی و استفاده مجدد از داده‌ها برای تمامی افراد حقیقی و حقوقی یکسان باشد.

۲- قابل فهم بودن (ولجکاویک و همکاران، ۲۰۱۴)

تعامل

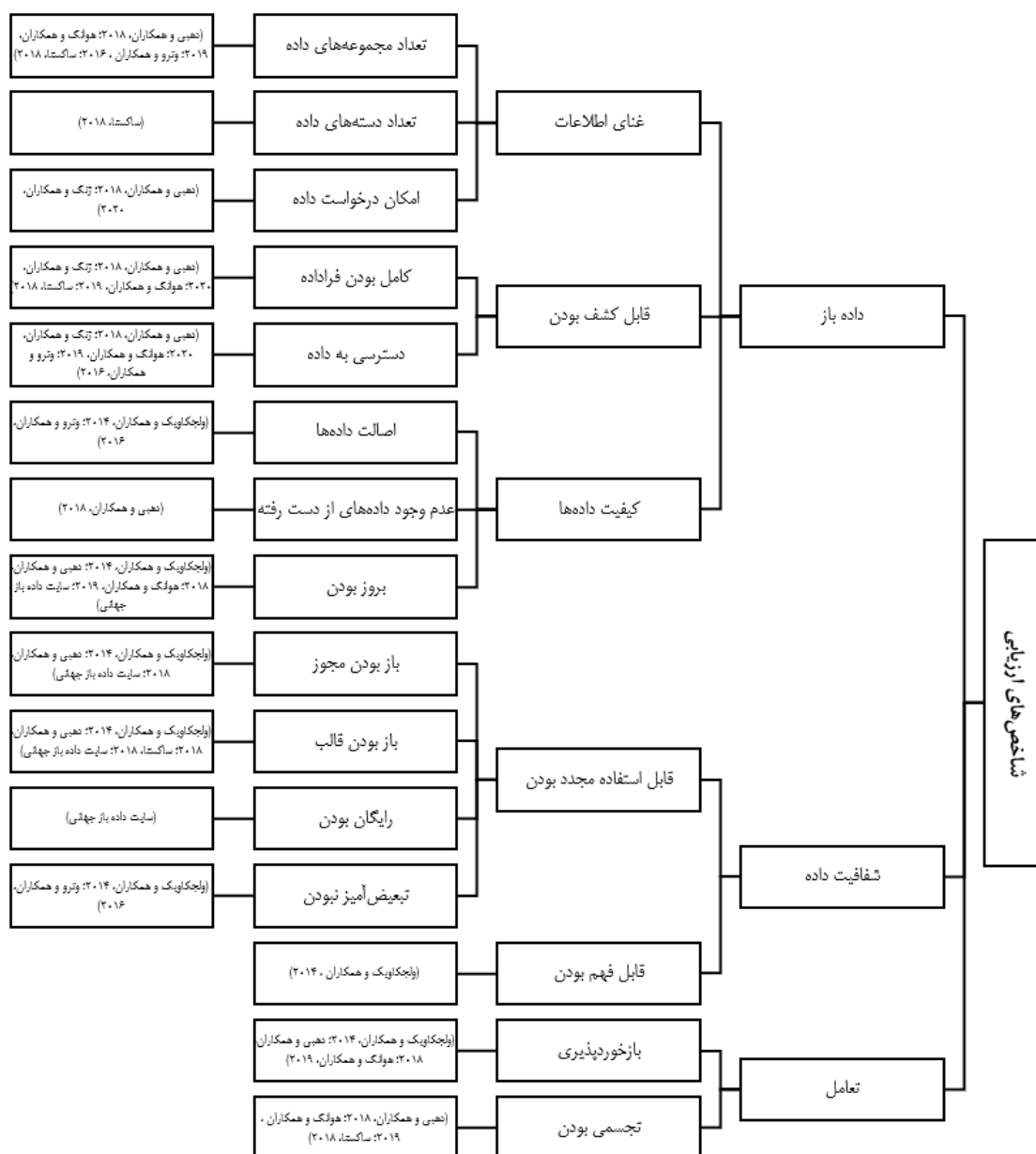
۱- بازخوردپذیری (ولجکاویک و همکاران، ۲۰۱۴؛ دهبی و همکاران، ۲۰۱۸؛ هوانگ و همکاران، ۲۰۱۹): این شاخص وجود ویژگی‌های مربوط به همکاری، بازخورد و ارزیابی را بررسی

می‌کند و وجود سه امکان را مورد سنجش قرار می‌دهد: اظهار نظر در مورد مجموعه داده‌ها، رتبه‌بندی مجموعه داده‌ها، بازخورد درباره پورتال.

۲- تجسمی بودن (دهبی و همکاران، ۲۰۱۸؛ هوانگ و همکاران، ۲۰۱۹؛ ساکسنا، ۲۰۱۸): این شاخص وجود ابزارها و ویژگی‌های تجسم مانند نقشه‌ها، نمودارها یا برنامه‌ها را برای تجسم و تعامل با داده‌ها ارزیابی می‌کند.

نمودار شاخص‌های ارزیابی

شکل ۲، نمودار شاخص‌های ارزیابی بیان شده را بر اساس بعد، شاخص و مرجع نشان می‌دهد.

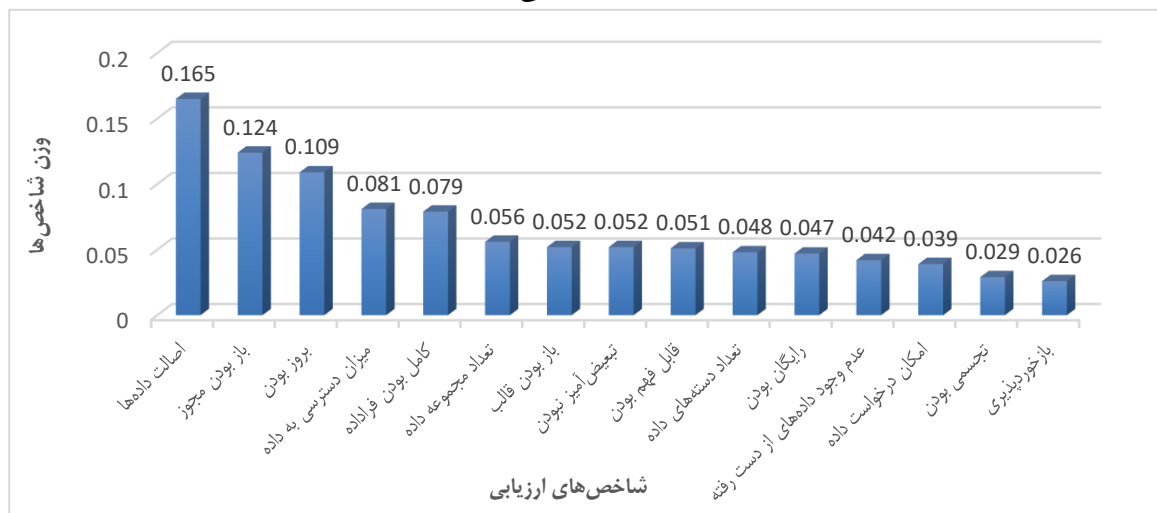


شکل ۲. نمودار شاخص‌های ارزیابی بر اساس بعد، شاخص و مرجع

محاسبه وزن و اهمیت شاخص‌های ارزیابی

در این بخش از روش میدانی و تکنیک تحلیل سلسله مراتبی^۱ برای محاسبه وزن و اهمیت شاخص‌های ارزیابی استفاده شده است. پرسشنامه‌ای طراحی شد و از خبرگان خواسته شد،

شاخص‌ها را بصورت زوجی با یکدیگر مقایسه کنند. شکل ۳ نتایج حاصل از تکنیک تحلیل سلسله مراتبی بر روی شاخص‌های ارزیابی را نشان می‌دهد. لازم به ذکر است نرخ ناسازگاری ۰.۰۰۷ بود که چون کمتر از ۰.۱ است، پایایی پرسشنامه‌ها را تایید می‌کند.



شکل ۳. وزن و اهمیت شاخص‌های ارزیابی حاصل از تکنیک تحلیل سلسله مراتبی

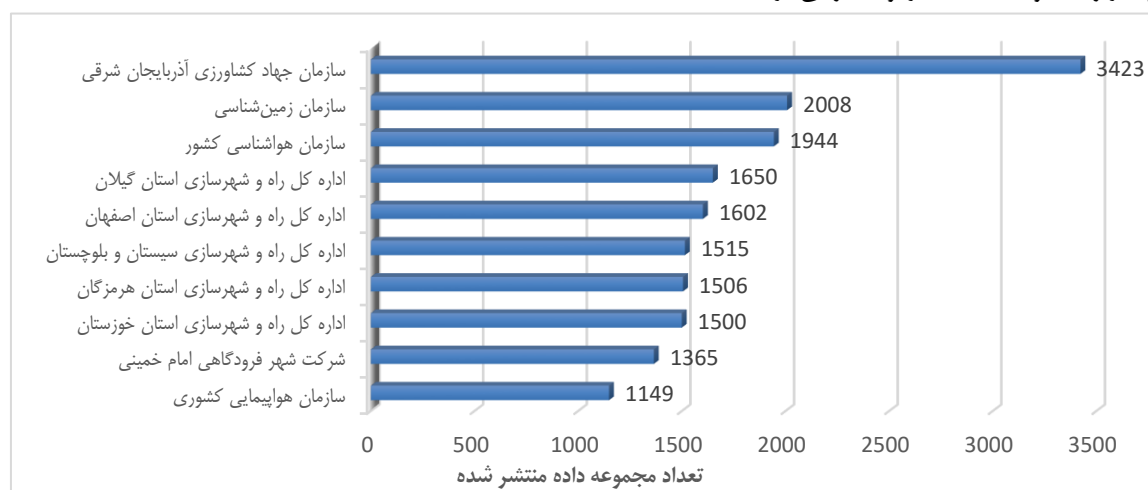
استفاده شده است. سازمان‌ها و موسسات دولتی که از این دو سامانه مورد بررسی قرار گرفتند، ۳۵۸ مورد بوده است. وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس شاخص ارزیابی تعداد مجموعه داده

شکل ۴، ده رتبه نخست از بین ۳۵۸ سازمان و موسسه دولتی بر اساس تعداد مجموعه داده منتشر شده را نشان می‌دهد.

طبق نتایج بدست آمده، سه شاخص «اصلت داده‌ها»، «باز بودن مجوز» و «بروز بودن» به ترتیب سه رتبه نخست از بین شاخص‌های ارزیابی استخراج شده را کسب کرده‌اند.

بررسی وضعیت سازمان‌ها و موسسات دولتی بر اساس شاخص‌های ارزیابی استخراج شده

به منظور تحلیل وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز از دو سامانه کاتالوگ ملی و مجموعه داده‌های باز و کاربردی^۲ و سامانه انتشار و دسترسی آزاد به اطلاعات^۳



شکل ۴. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تعداد مجموعه داده منتشر شده

³ <https://iranfoia.ir/>

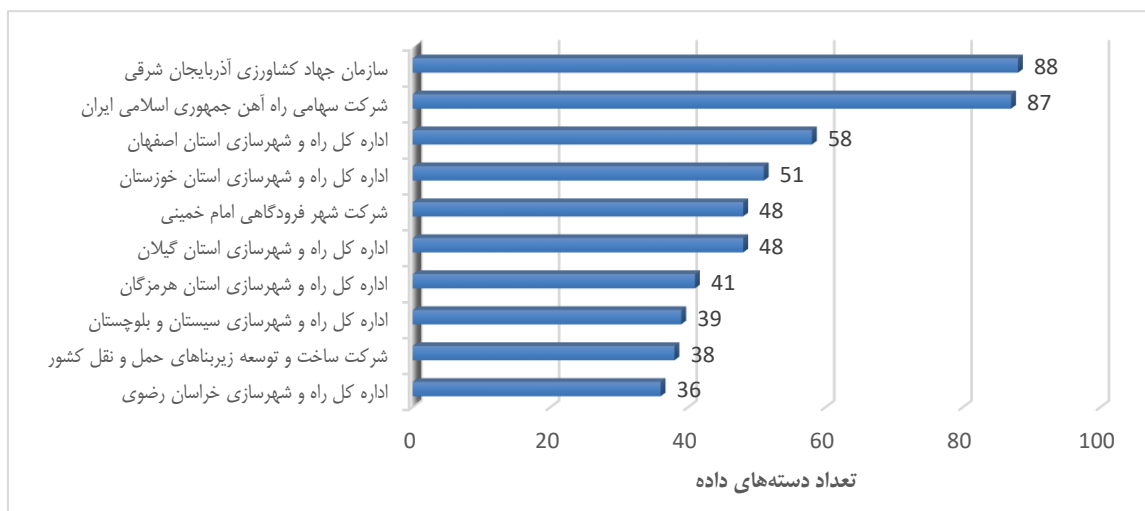
¹ Analytic Hierarchy Process (AHP)

² <https://data.gov.ir/>

وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس شاخص ارزیابی تعداد دسته‌های داده

ده رتبه نخست از بین ۳۵۸ سازمان و موسسه دولتی بررسی شده بر اساس شاخص تعداد دسته‌های داده در شکل ۵ نشان داده شده است.

طبق نتایج شکل ۴، «سازمان جهاد کشاورزی آذربایجان شرقی» با انتشار ۳۴۲۳ مجموعه داده، رتبه نخست از بین ۳۵۸ سازمان بررسی شده در شاخص ارزیابی «تعداد مجموعه داده منتشر شده» را دارا می‌باشد.

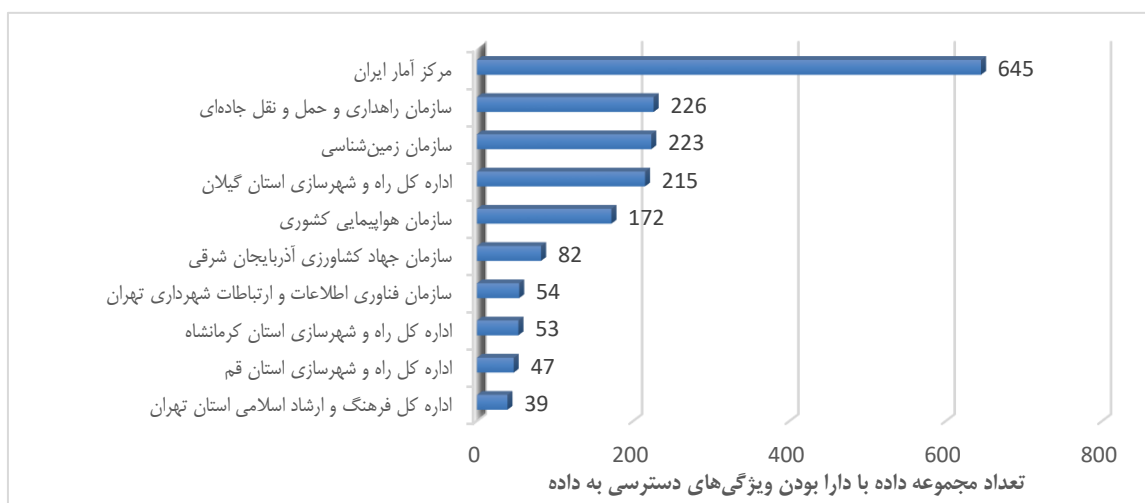


شکل ۵. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تعداد دسته‌های داده

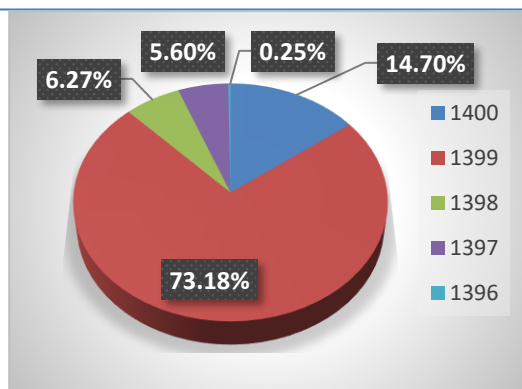
همان‌طور که پیش‌تر بیان شد شاخص دسترسی به داده، وجود ویژگی‌هایی را ارزیابی می‌کند که کشف داده‌ها را افزایش می‌دهد، به ویژه وجود سه ویژگی: جستجو، مرتب‌سازی و فیلتر. در این بخش تعداد مجموعه داده‌هایی که در آن‌ها امکان جستجو، مرتب‌سازی و فیلتر وجود دارد، برای هر سازمان استخراج شده است. شکل ۶ ده رتبه نخست از ۳۵۸ سازمان و موسسه بررسی شده برای شاخص دسترسی به داده را نشان می‌دهد.

همان‌طور که از شکل ۵ مشخص است، «سازمان جهاد کشاورزی آذربایجان شرقی» با تعداد ۸۸ دسته داده، رتبه نخست از بین ۳۵۸ سازمان بررسی شده در شاخص ارزیابی «تعداد دسته‌های داده» را بدست آورده است.

وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس شاخص ارزیابی میزان دسترسی به داده



شکل ۶. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی میزان دسترسی به داده



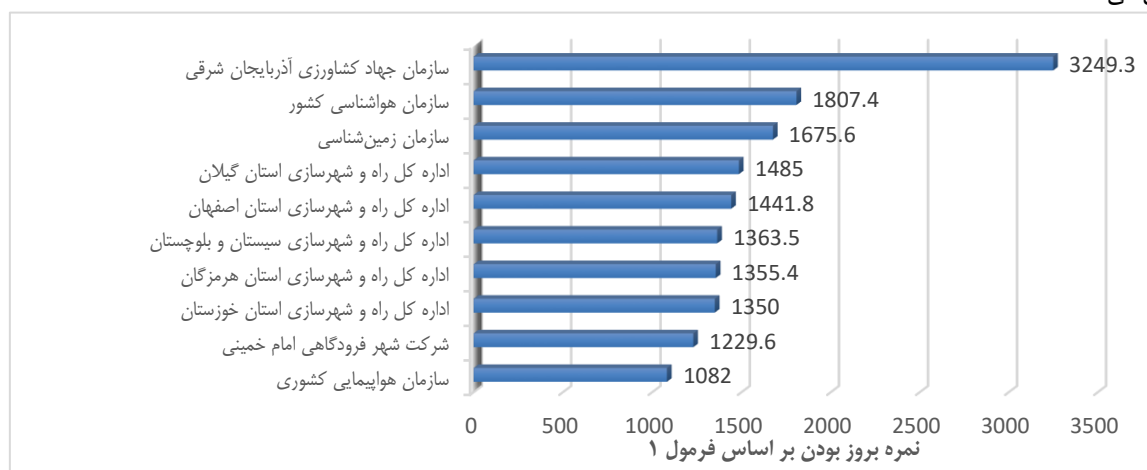
شکل ۷. درصد مجموعه داده‌های منتشر شده به تفکیک سال همان‌طور که در شکل ۷ مشاهده می‌شود حدود ۷۳ درصد داده‌ها در سال ۱۳۹۹ منتشر شده‌اند. همچنین شکل ۸، ده رتبه نخست از بین ۳۵۸ سازمان و موسسه بررسی شده بر اساس نمره‌دهی فرمول ۱ برای شاخص بروز بودن را نشان می‌دهد.

طبق نتایج شکل ۶، «مرکز آمار ایران» با ارائه ۶۴۵ مجموعه داده که دارای ویژگی‌های ذکر شده دسترسی به داده را دارند، رتبه نخست را کسب کرده است.

وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها

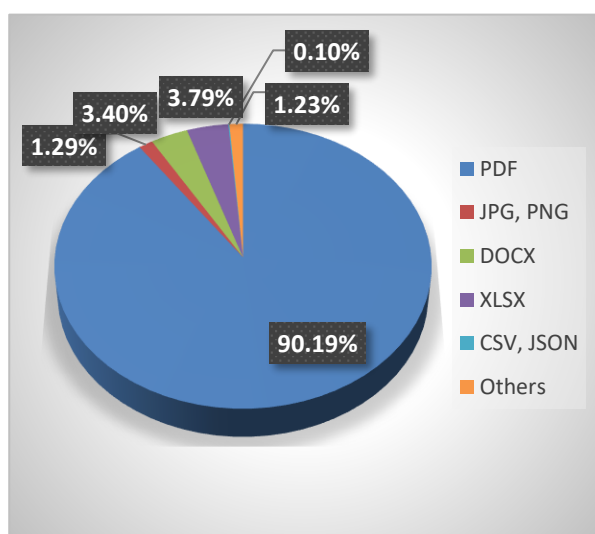
بصورت باز بر اساس شاخص ارزیابی بروز بودن

همان‌طور که در توضیح شاخص بروز بودن بیان شد، تنها مجموعه داده‌هایی که در پنج سال اخیر منتشر شده‌اند، در نظر گرفته شده است. در این بخش در ابتدا میزان مجموعه داده‌های منتشر شده به تفکیک سال برای ۳۵۸ سازمان و موسسه دولتی بررسی شده است. سپس طبق فرمول ۱، سازمان‌ها بر اساس میزان بروز بودن محاسبه و رتبه‌بندی شده‌اند. نمودار دایره‌ای شکل ۷، درصد مجموعه داده‌های منتشر شده به تفکیک سال را نشان می‌دهد.



شکل ۸. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی بروز بودن

مجموعه داده‌های منتشر شده بر اساس فرمت‌های مختلف را نشان می‌دهد.



شکل ۹. درصد مجموعه داده‌های منتشر شده بر اساس فرمت‌های مختلف

طبق نتایج شکل ۸، «سازمان جهاد کشاورزی آذربایجان شرقی» بیشترین مجموعه داده بروز را منتشر کرده است. «سازمان هواشناسی کشور» و «سازمان زمین‌شناسی» رتبه‌های بعدی را کسب کرده‌اند.

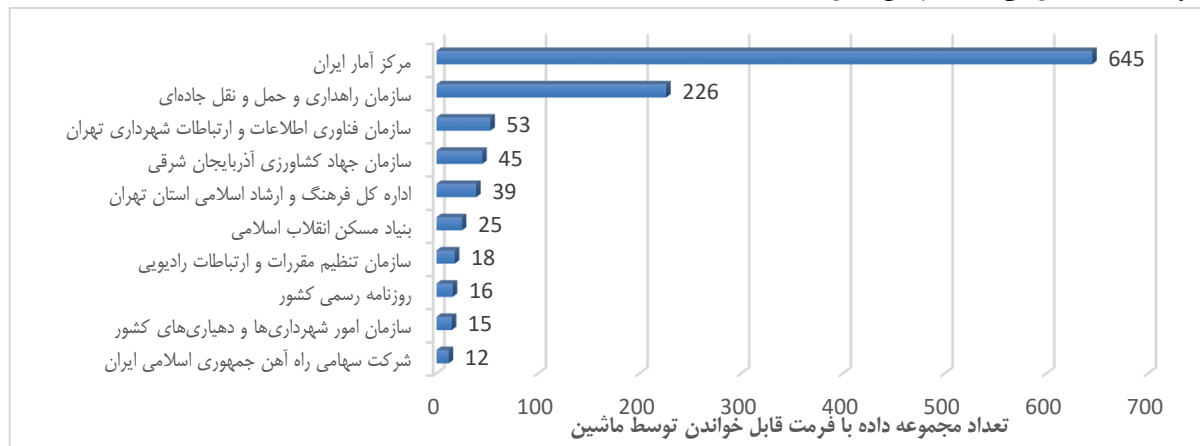
وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها

بصورت باز بر اساس شاخص باز بودن قالب

همان‌طور که پیش‌تر بیان شد، شاخص باز بودن قالب، قابل خواندن بودن مجموعه داده توسط ماشین را ارزیابی می‌کند. برخی فرمت‌های داده مانند PDF، JPG، PNG قابل خواندن توسط ماشین نیست، اما برخی فرمت‌ها مانند JSON قابلیت خواندن توسط ماشین را دارا هستند. در این بخش ابتدا تعداد مجموعه داده بر اساس فرمت‌های مختلف، مشخص شده‌اند. سپس به رتبه‌بندی سازمان‌ها و موسسات دولتی بر اساس شاخص باز بودن قالب پرداخته شده است. نمودار دایره‌ای شکل ۹، درصد

از بین ۳۵۸ سازمان و موسسه دولتی بررسی شده را بر اساس شاخص ارزیابی باز بودن قالب نشان می‌دهد.

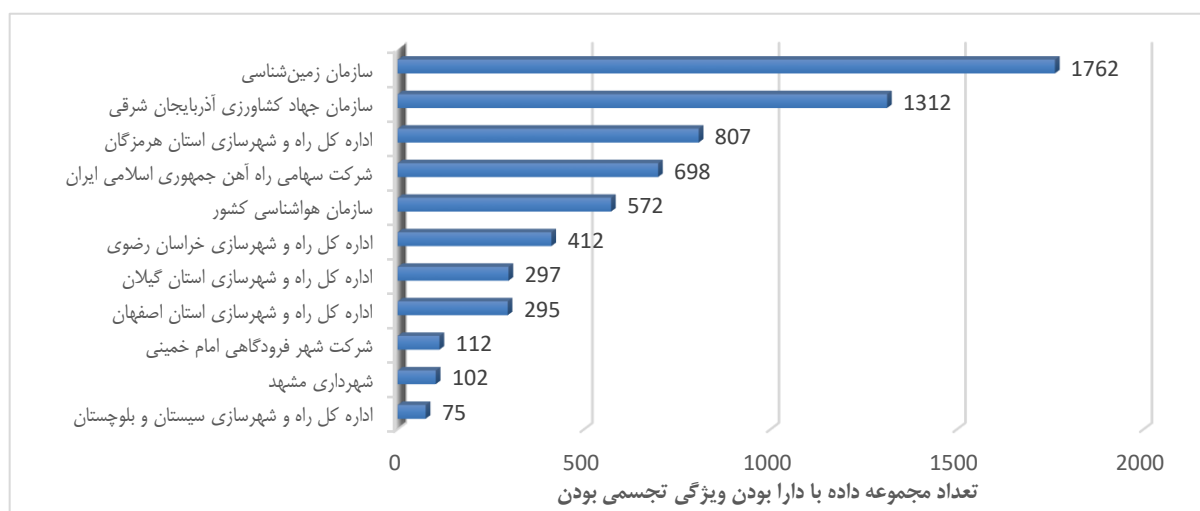
همان‌طور که در شکل ۹ ملاحظه می‌شود، فرمت PDF با دارا بودن ۹۰ درصد، بیشترین فرمت استفاده شده در ارائه داده‌ها بصورت باز را تشکیل می‌دهد. همچنین شکل ۱۰، ده رتبه نخست



شکل ۱۰. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی باز بودن قالب

وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس شاخص تجسمی بودن
شکل ۱۱، ده رتبه نخست سازمان‌ها و موسسات دولتی از بین ۳۵۸ سازمان بررسی شده را بر اساس شاخص ارزیابی تجسمی بودن نشان می‌دهد.

طبق نتایج شکل ۱۰، «مرکز آمار ایران» با ارائه ۶۴۵ مجموعه داده قابل خواندن توسط ماشین، رتبه نخست را کسب کرده است. «سازمان راهداری و حمل و نقل جاده‌ای» با ارائه ۲۲۶ مجموعه داده با فرمت قابل خواندن توسط ماشین رتبه دوم را دارا می‌باشد.

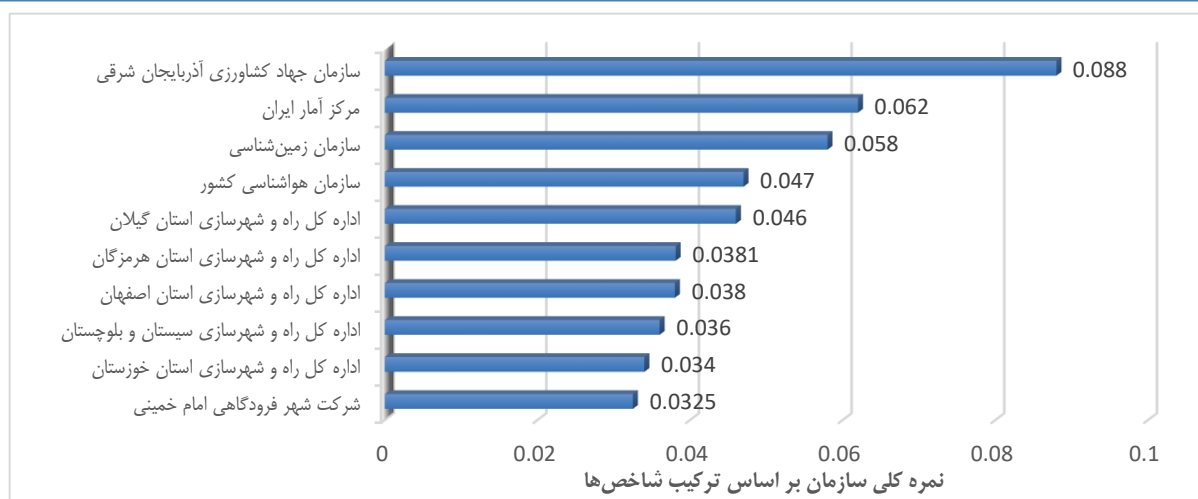


شکل ۱۱. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تجسمی بودن

در بخش‌های پیشین، وضعیت سازمان‌ها و موسسات دولتی بر اساس هر شاخص بصورت جداگانه ارزیابی شد. در این بخش به بررسی سازمان‌ها در ارائه داده‌ها بصورت باز بر اساس ترکیب تمامی شاخص‌های ارزیابی ذکر شده و وزن شاخص‌ها پرداخته شده است. شکل ۱۲، ده رتبه نخست را نشان می‌دهد.

طبق نتایج شکل ۱۱، «سازمان زمین‌شناسی» با ارائه ۱۷۶۲ مجموعه داده دارای ویژگی تجسمی بودن، رتبه نخست را در این شاخص ارزیابی کسب کرده است.

وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس ترکیب شاخص‌های ارزیابی و وزن شاخص‌ها



شکل ۱۲. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس ترکیب شاخص‌های ارزیابی و وزن شاخص‌ها

داده‌های استخراج شده در مراحل پیشین استفاده شده است. برای این کار در ابتدا نیاز است تعداد خوشه‌های بهینه مشخص شود. بدین منظور از کدنویسی در R استفاده شده است که در ادامه توضیح داده شده است.

تعیین تعداد خوشه‌های بهینه

به منظور شناسایی تعداد خوشه‌های بهینه از کدنویسی در R استفاده شد. از ۱۰ معیار شامل silhouette, gap statistic, beale, ball, ptbserial, frey, mcclain, hubert, cindex و duda استفاده شد. شکل ۱۳ نمایی از کد و برنامه R را نشان می‌دهد.

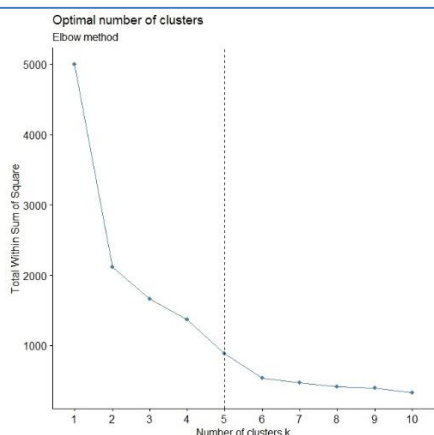
طبق نتایج شکل ۱۲ و بر اساس ترکیب شاخص‌های ارزیابی و در نظر گرفتن وزن و اهمیت هر شاخص، «سازمان جهاد کشاورزی آذربایجان شرقی»، «مرکز آمار ایران» و «سازمان زمین‌شناسی» سه رتبه نخست از بین ۳۵۸ سازمان بررسی شده را در ارائه داده‌ها بصورت باز دارا می‌باشند.

اعمال تکنیک‌های داده‌کاوی بر روی داده‌های استخراج شده خوشه‌بندی

به منظور شناسایی سازمان‌ها و موسسات دولتی با وضعیت مشابه، ایجاد برچسب و ایجاد یک مدل طبقه‌بندی، از خوشه‌بندی

```

File Edit Code View Plots Session Build Debug Tools Help
mydata.R mydata mydata.Normalized
1 library(xlsx)
2 library(factoextra)
3 library(Nbclust)
4
5 # Reading the data
6 mydata <- read.xlsx("C:/Users/Mohammad Moradi/Desktop/data.xlsx", 1, header=TRUE)
7
8 # Standardize the data
9 mydata.Normalized <- scale(mydata)
10
11 # Elbow method
12 fviz_nbclust(mydata.Normalized, kmeans, method = "wss") + geom_vline(xintercept=6, linetype=2) +
13
14 # Silhouette method
15 fviz_nbclust(mydata.Normalized, kmeans, method = "silhouette") + labs(subtitle = "Silhouette meth
16
17 # Gap statistic
18 # nboot = 50 to keep the function speedy
19 # Recommended value: nboot = 500 for your analysis
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
238
```



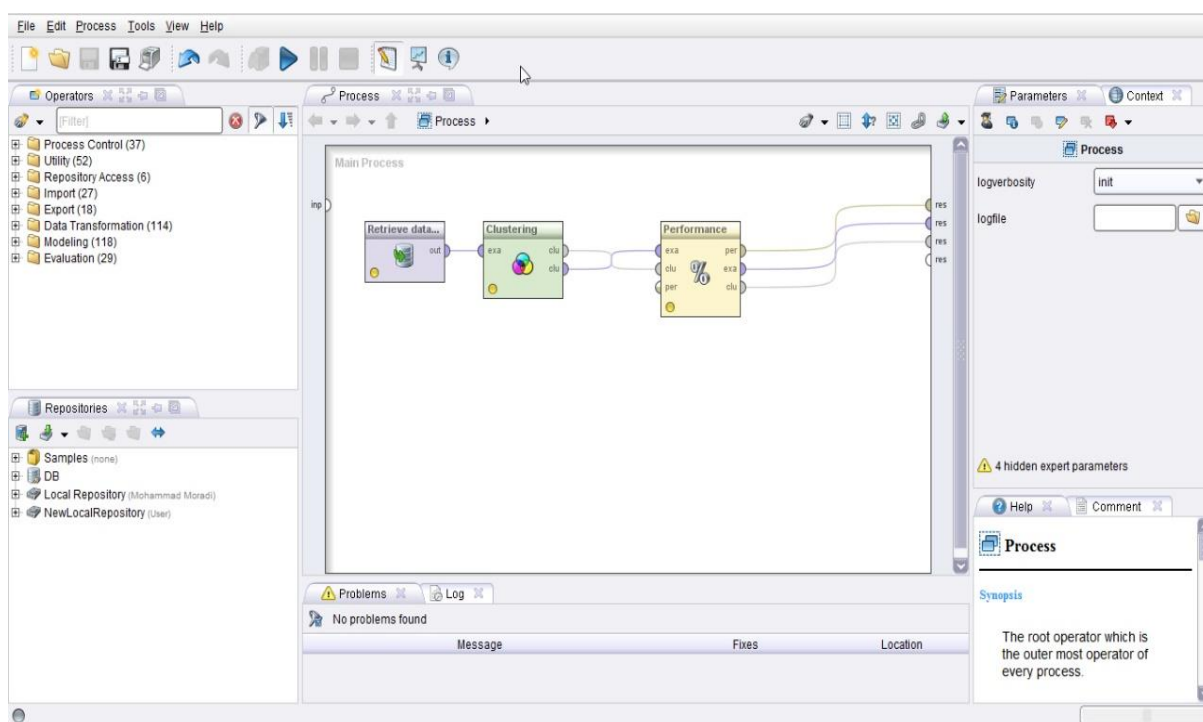
شکل ۱۴. روش Elbow در تعیین تعداد خوشه بهینه

همان‌طور که از خروجی کنسول شکل ۱۳ مشخص است، بیشتر معیارها، ۴ خوشه را به عنوان تعداد خوشه بهینه انتخاب کردند. در توضیح نحوه تعیین تعداد خوشه بهینه در هر روش می‌توان به شکل ۱۴، که روش Elbow را نشان می‌دهد و از مجموع مربعات خط^۱ استفاده شده است، اشاره کرد. در این روش از خوشه ۵ به بعد تغییرات محسوسی در کاهش مربعات خط دیده نمی‌شود و این روش، ۵ خوشه را به عنوان تعداد خوشه بهینه اعلام کرده است. بر همین اساس برای سایر معیارها، تعداد خوشه بهینه مشخص شده است.

means به داده‌های پرت^۲ از الگوریتم K-medoids استفاده شده است. بدین منظور نرم‌افزار ریپیدماینر^۳ مورد استفاده قرار گرفته است. شکل ۱۵، نمایی از پیاده‌سازی خوشه‌بندی در نرم‌افزار ریپیدماینر را نشان می‌دهد.

الگوریتم خوشه‌بندی

در این بخش به اعمال الگوریتم خوشه‌بندی بر روی داده‌های استخراج شده پرداخته شده است. به دلیل حساس بودن K-



شکل ۱۵. نمایی از پیاده‌سازی خوشه‌بندی در نرم‌افزار ریپیدماینر

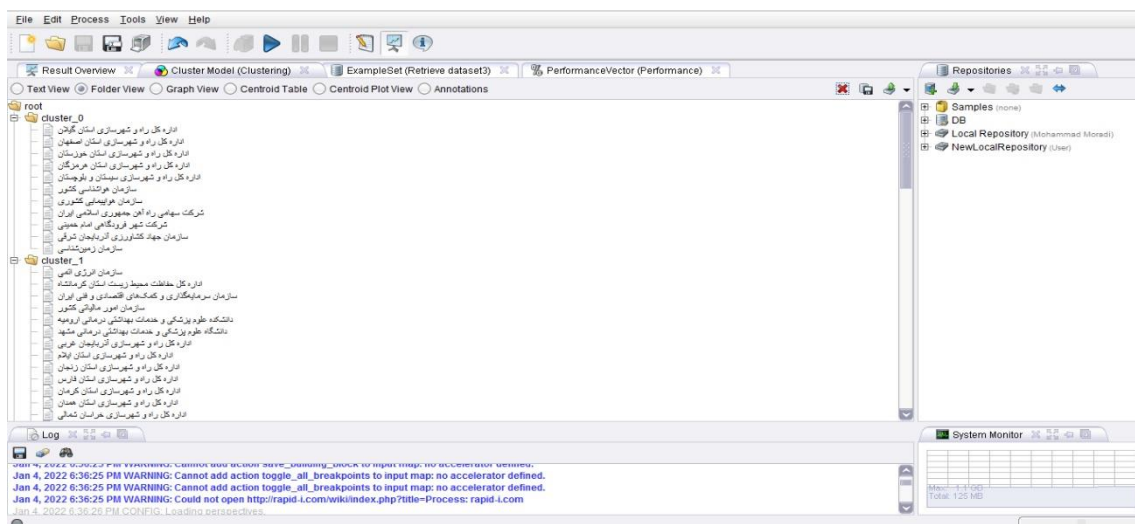
کسب کرده بودند در خوشه ۰ قرار گرفته‌اند. خوشه‌های دیگر به ترتیب رتبه‌های بعدی را تشکیل می‌دهند.

شکل ۱۶ خروجی خوشه‌بندی را نشان می‌دهد. همان‌طور که مشاهده می‌شود تقریباً سازمان‌ها و موسساتی که رتبه‌های نخست با استفاده از تکنیک‌های تصمیم‌گیری چندشاخصه را

^۳ Rapid Miner

^۱ Total Within Sum of Square

^۲ Outlier



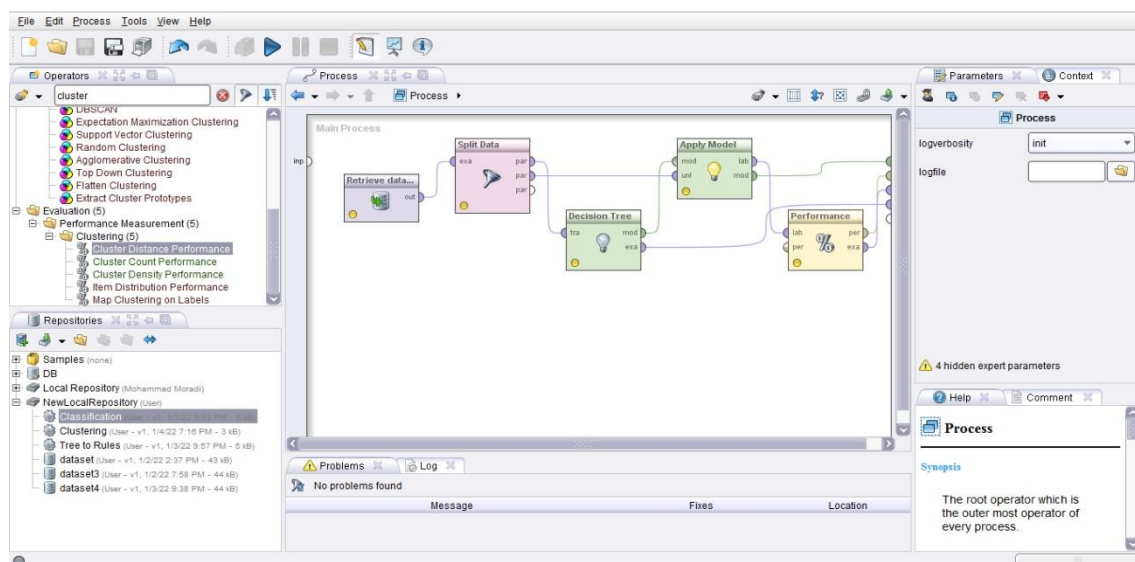
شکل ۱۶. خروجی خوشه‌بندی در نرم‌افزار ریپیدماینر

جدول ۱. برچسب‌ها و تعداد سازمان‌های دارای هر برچسب

برچسب	تعریف برچسب	تعداد سازمان‌ها با برچسب مشخص شده
Very Good (VG)	بسیار مطلوب	۱۲
Good (G)	مطلوب	۶۶
Weak (W)	نامطلوب	۱۶۸
Very Weak (VW)	بسیار نامطلوب	۱۱۲

مدل طبقه‌بندی

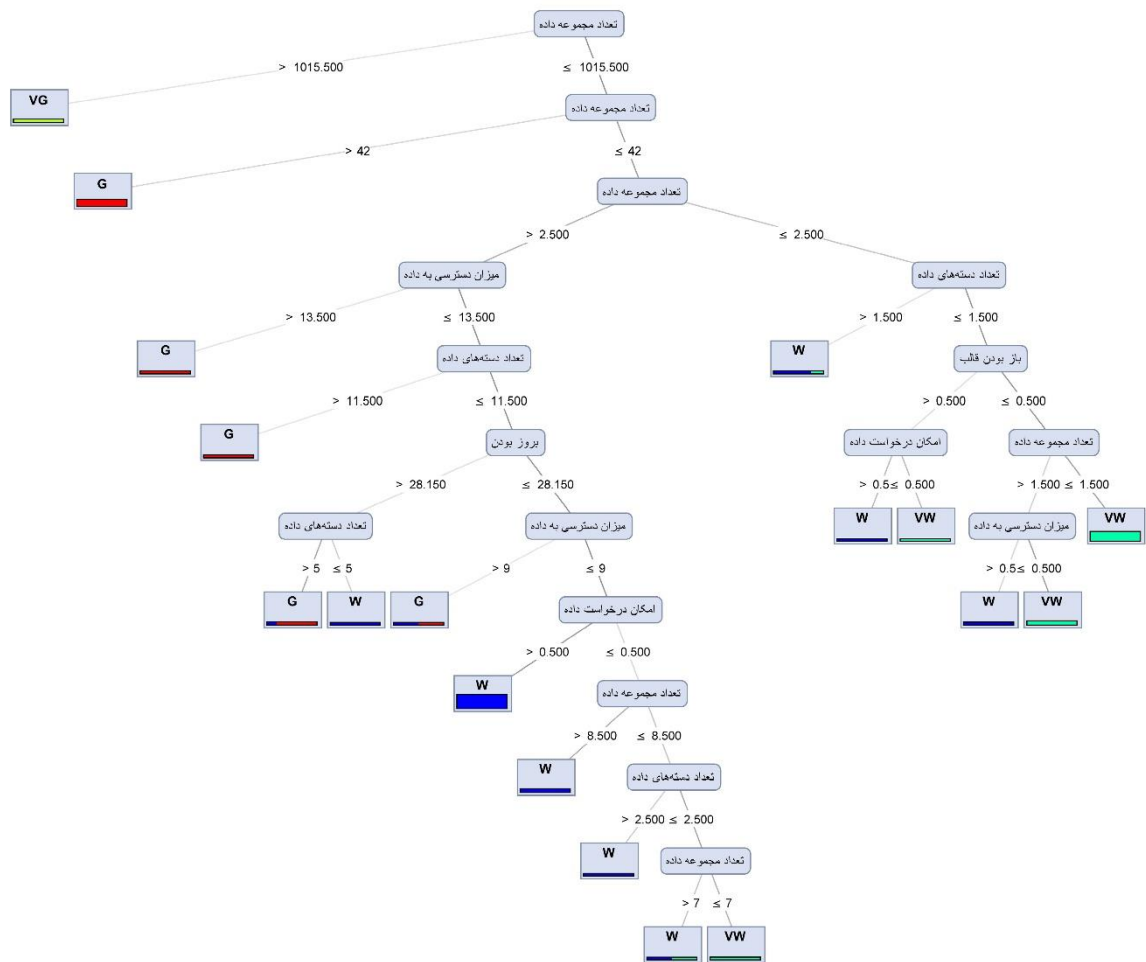
پس از برچسب‌گذاری به ایجاد مدل طبقه‌بندی درخت تصمیم^۱ با استفاده از نرم‌افزار ریپیدماینر پرداخته شد. شکل ۱۷، نمایی از پیاده‌سازی را نشان می‌دهد.



شکل ۱۷. نمایی از پیاده‌سازی مدل طبقه‌بندی درخت تصمیم در نرم‌افزار ریپیدماینر

^۱ Decision Tree

در این دسته‌بندی از ۷۰ درصد داده‌ها به عنوان داده‌های آموزش استفاده شد و ۳۰ درصد باقی مانده به عنوان داده‌های تست در نظر گرفته شد. شکل ۱۸، درخت تصمیم ایجاد شده را نشان می‌دهد.



شکل ۱۸. درخت تصمیم

ارزیابی مدل طبقه‌بندی

در این بخش به ارزیابی مدل طبقه‌بندی ایجاد شده پرداخته شده است. شکل ۱۹، ماتریس درهم‌ریختگی^۱ درخت تصمیم ایجاد شده را نشان می‌دهد.

با استفاده از مدل طبقه‌بندی ایجاد شده در شکل ۱۸ می‌توانیم به پیش‌بینی وضعیت سازمان‌ها بپردازیم. به عنوان مثال اگر تعداد مجموعه داده منتشر شده یک سازمان کوچک‌تر از ۳ و تعداد دسته‌های داده آن سازمان بزرگ‌تر از ۱ باشد، دارای برچسب W بوده وضعیت آن نامطلوب خواهد بود. همین‌طور می‌توان قوانین دیگر را از درخت تصمیم ایجاد شده استخراج کرد.

accuracy: 96.26%					
	true W	true VW	true VG	true G	class precision
pred. W	49	1	0	1	96.08%
pred. VW	0	35	0	0	100.00%
pred. VG	0	0	3	0	100.00%
pred. G	1	0	1	16	88.89%
class recall	98.00%	97.22%	75.00%	94.12%	

شکل ۱۹. ماتریس درهم‌ریختگی درخت تصمیم ایجاد شده

^۱ Confusion Matrix

فرا داده»، «تعداد مجموعه داده»، «باز بودن قالب»، «تبعض آمیز نبودن»، «قابل فهم بودن»، «تعداد دسته‌های داده»، «رایگان بودن»، «عدم وجود داده‌های از دست رفته»، «امکان درخواست داده»، «تجسمی بودن» و «بازخوردپذیری» استخراج شدند. با استفاده از تکنیک تحلیل سلسله مراتبی، وزن شاخص‌ها محاسبه شد که پس از نرمال‌سازی، مجموع وزن ۱۵ شاخص استخراج شده برابر با یک بود. شاخص‌های «اصالت داده‌ها» با وزن ۰.۱۶۵، «باز بودن مجوز» با وزن ۰.۱۲۴ و «بروز بودن» با وزن ۰.۱۰۹ به ترتیب رتبه‌های اول تا سوم را در بین ۱۵ شاخص ارزیابی کسب کردند. همچنین به استخراج داده‌های سازمان‌ها و موسسات دولتی دو سامانه کاتالوگ ملی و مجموعه داده‌های باز و کاربردی و سامانه انتشار و دسترسی آزاد به اطلاعات پرداخته شد که شامل ۳۵۸ مورد بود. «سازمان جهاد کشاورزی آذربایجان شرقی» با انتشار ۳۴۲۳ مجموعه داده، رتبه نخست از بین ۳۵۸ سازمان بررسی شده در شاخص ارزیابی «تعداد مجموعه داده منتشر شده» را کسب کرد. همچنین این سازمان با تعداد ۸۸ دسته داده، رتبه نخست از بین ۳۵۸ سازمان بررسی شده در شاخص ارزیابی «تعداد دسته‌های داده» را بدست آورد. «مرکز آمار ایران» با دارا بودن ۶۴۵ مجموعه داده که امکان جستجو، مرتب‌سازی و فیلتر را دارند، رتبه نخست را در شاخص ارزیابی «میزان دسترسی به داده» کسب کرد. بر اساس شاخص ارزیابی «بروز بودن»، «سازمان جهاد کشاورزی آذربایجان شرقی» بیشترین مجموعه داده بروز را منتشر کرده است. «سازمان هواشناسی کشور» و «سازمان زمین‌شناسی» رتبه‌های بعدی را کسب کرده‌اند. «مرکز آمار ایران» با ارائه ۶۴۵ مجموعه داده قابل خواندن توسط ماشین، رتبه نخست را در شاخص ارزیابی «باز بودن قالب» کسب کرده است. «سازمان راهداری و حمل و نقل جاده‌ای» با ارائه ۲۲۶ مجموعه داده با فرمت قابل خواندن توسط ماشین رتبه دوم را دارا می‌باشد. «سازمان زمین‌شناسی» با ارائه ۱۷۶۲ مجموعه داده دارای ویژگی تجسمی بودن، رتبه نخست را در این شاخص ارزیابی کسب کرده است. همچنین بر اساس ترکیب شاخص‌های ارزیابی و در نظر گرفتن وزن و اهمیت هر شاخص، «سازمان جهاد کشاورزی آذربایجان شرقی» با وزن ۰.۰۸۸، «مرکز آمار ایران» با وزن ۰.۰۶۲ و «سازمان زمین‌شناسی» با وزن ۰.۰۵۸ به ترتیب رتبه‌های اول تا سوم در بین ۳۵۸ سازمان و موسسه دولتی بررسی شده را در ارائه داده‌ها بصورت باز کسب کردند.

صحت^۱ مدل دسته‌بندی ایجاد شده برابر ۹۶.۲۶٪ است که نشان از خوب بودن مدل دارد. همچنین مقادیر سایر معیارها مانند بازیابی^۲ و دقت^۳ در شکل ۱۹ نشان داده شده است.

بحث و نتیجه‌گیری

علی رغم این که برخی سازمان‌ها و موسسات دولتی به ارائه داده‌های خود بصورت باز پرداختند، تعامل شهروندان با داده‌های دولتی باز مطلوب نبوده است. این مسئله می‌تواند ناشی از عوامل مختلفی مانند بروز نبودن داده‌ها، عدم استفاده از ابزارهای تجسمی و غیره باشد. لذا نیاز است شاخص‌های ارزیابی داده‌های دولتی باز به منظور بهبود کیفیت داده‌ها و در نتیجه افزایش تعامل شهروندان با داده‌های دولتی باز مشخص شوند. متأسفانه پژوهش‌های صورت گرفته هر یک فقط بر بعد خاصی از داده‌های دولتی باز متمرکز شده‌اند و یک مجموعه شاخص ارزیابی جامع و کامل وجود ندارد. همچنین میزان اهمیت هر بعد و شاخص ارزیابی در نظر گرفته نشده است و وضعیت کشور در این حوزه بررسی نشده است. از طرفی ارائه یک مدل طبقه‌بندی می‌تواند به پیش‌بینی وضعیت سازمان‌ها در ارائه داده‌هایشان بصورت باز کمک کند. به عنوان مثال در پژوهش دجوانا اسپینوسا و لوجان مورا (۲۰۲۰)، «وجود پیوند از پورتال داده دولتی باز به سایت منبع ارائه دهنده مجموعه داده»، «وجود پلاگین‌های شبکه‌های اجتماعی»، «پشتیبانی از فرمت‌های مختلف مجموعه داده‌ها» و «امکان جستجو و فیلتر داده‌ها» به عنوان شاخص‌های ارزیابی پورتال‌های داده‌های دولتی باز پیشنهاد شده‌اند. در این پژوهش، شاخص‌های حوزه‌های کیفیت داده‌ها و فراداده مانند تعداد مجموعه‌های داده، تعداد دسته‌های داده، وجود فراداده توصیفی، امکان درخواست داده و غیره در نظر گرفته نشده است. همچنین در پژوهش صورت گرفته توسط دهبی و همکاران (۲۰۱۸)، ابعاد ارزیابی شامل غنای اطلاعاتی، قابلیت کشف، قابلیت استفاده مجدد، تعامل و کیفیت داده‌ها، معرفی شده‌اند. اما وزن و اهمیت هر یک از ابعاد و شاخص‌های ارزیابی مشخص نشده‌اند. در پژوهش ابوالعالی و همکاران (۱۳۹۹) به استخراج ابعاد فناوری، دولت به مثابه پلتفرم، مشارکت و شفافیت پرداخته شده است. اما وضعیت کشور بر اساس ابعاد استخراج شده، بررسی نشده است. همچنین به تجزیه و تحلیل داده‌ها، خوشه‌بندی و ارائه مدل طبقه‌بندی پرداخته نشده است.

در این پژوهش، بر اساس مطالعات و بررسی مقالات، ۱۵ شاخص ارزیابی داده‌های دولتی باز شامل «اصالت داده‌ها»، «باز بودن مجوز»، «بروز بودن»، «میزان دسترسی به داده»، «کامل بودن

³ Precision

¹ Accuracy

² Recall

وضعیت آن نامطلوب خواهد بود. مواردی که در این پژوهش به آن‌ها پرداخته شد شامل استخراج شاخص‌های ارزیابی، محاسبه وزن و اهمیت هر شاخص، بررسی وضعیت موجود سازمان‌ها و موسسات دولتی داخل کشور و مدل طبقه‌بندی ایجاد شده، می‌تواند به مدیران در شناخت وضع موجود و بهبود آن و در نتیجه افزایش تعامل شهروندان با داده‌های دولتی باز کمک نماید.

تقدیر و تشکر

بدین وسیله از کلیه افرادی که در انجام پژوهش حاضر همکاری نمودند، تشکر و قدردانی به عمل می‌آید.

تعارض منافع

نویسندگان، اعلام می‌دارند در رابطه با انتشار مقاله ارائه‌شده، هیچ‌گونه تعارض منافی وجود ندارد.

منبع حمایت‌کننده

پژوهش حاضر، پژوهشی مستقل و بدون دریافت هرگونه حمایتی انجام شده است.

در فاز بعد به خوشه‌بندی سازمان‌ها و موسسات دولتی به منظور قرار گرفتن سازمان‌های مشابه در خوشه‌های یکسان، پرداخته شد. تقریباً سازمان‌ها و موسساتی که رتبه‌های نخست با استفاده از تکنیک‌های تصمیم‌گیری چندشاخصه را کسب کرده بودند در خوشه ۰ قرار گرفتند. خوشه‌های دیگر به ترتیب رتبه‌های بعدی را تشکیل دادند. برای هر خوشه، برچسبی بر اساس ویژگی خوشه‌ها تخصیص یافت. خوشه ۰ که شامل سازمان‌های با رتبه‌های ۱ تا ۱۲ بودند، برچسب «بسیار مطلوب» را دریافت کردند. همین‌طور خوشه ۱ با ۶۶ سازمان برچسب «مطلوب»، خوشه ۲ با ۱۶۸ سازمان، برچسب «نامطلوب» و خوشه ۳ با ۱۱۲ سازمان برچسب «نامطلوب» را دریافت کردند. پس از برچسب‌گذاری به ایجاد مدل طبقه‌بندی درخت تصمیم پرداخته شد. با استفاده از مدل طبقه‌بندی ایجاد شده می‌توان به پیش‌بینی وضعیت سازمان‌ها پرداخت. به عنوان مثال اگر تعداد مجموعه داده منتشر شده یک سازمان کوچک‌تر از ۳ و تعداد دسته‌های داده آن سازمان بزرگ‌تر از ۱ باشد، دارای برچسب W بوده و

References

- Aboalmaali, F.S., Daneshfard, K., & Pourezzat, A.A. (2020). A Pattern to Recognition of Triggering Element of Open Government Implementation in Iran's Public Organizations (Case Study: Ministry of Interior). *Journal of Public Administration*, 12(1), 145-174. (in Persian)
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government information quarterly*, 32(4), 399-418.
- Dahbi, K. Y., Lamharhar, H., & Chiadmi, D. (2018, October). Toward an evaluation model for open government data portals. In *International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning* (pp. 502-511). Springer, Cham.
- Daniels, M. S. F., & Lopez, R. (2014, August). Automatic generation of roadmaps for open data. In *Electronic Government and Electronic Participation: Joint Proceedings of Ongoing Research, Posters, Workshop and Projects of IFIP EGOV 2014 and EPart 2014* (Vol. 21, p. 95). IOS Press.
- Danneels, L., Viaene, S., & Van den Bergh, J. (2017). Open data platforms: Discussing alternative knowledge epistemologies. *Government Information Quarterly*, 34(3), 365-378.
- Dawes, S. S., Vidasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15-27.
- de Juana-Espinosa, S., & Luján-Mora, S. (2020). Open government data portals in the European Union: A dataset from 2015 to 2017. *Data in brief*, 29, 105156.
- Halonen, A. (2012). Being open about data. Analysis of the UK Open Data Policies and Applicability of Open Data. London: Finnish Institute in London.
- Huang, R., Wang, C., Zhang, X., Wu, D., & Xie, Q. (2019). Design, develop and evaluate an open government data platform: a user-centred approach. *The Electronic Library*.
- Janssen, K. (2012). Open government data and the right to information: Opportunities and obstacles. *The Journal of Community Informatics*, 8(2).
- Misuraca, G., & Viscusi, G. (2014). Is open data enough?: E-governance challenges for open government. *International Journal of Electronic Government Research (IJEGR)*, 10(1), 18-34.
- Momen Kashani, N., Rahnavard, F., Mortazavi, M., Shirazi, M. (2020). A model for measuring the willingness of policy makers to open government in Iran. *Public Administration Perspaective*, 11(2), 35-64. (In Persian)
- Nikiforova, A., & McBride, K. (2021). Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*, 58, 101539.
- Saxena, S. (2018). Open government data (OGD) in six Middle East countries: An evaluation of the national open data portals. *Digital Policy, Regulation and Governance*.
- Susha, I., Grönlund, Å., & Janssen, M. (2015). Driving factors of service innovation using open

- government data: An exploratory study of entrepreneurs in two countries. *Information polity*, 20(1), 19-34.
- Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278-290.
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.
- Yu, H., & Robinson, D. G. (2011). The new ambiguity of open government. *UCLA L. Rev. Discourse*, 59, 178.
- Zhang, H., & Xiao, J. (2020). Quality assessment framework for open government data: Meta-synthesis of qualitative research, 2009-2019. *The Electronic Library*.
- Zheng, L., Kwok, W. M., Aquaro, V., Qi, X., & Lyu, W. (2020, September). Evaluating global open government data: methods and status. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 381-391).

