جلد هشتم، شماره چهام (زمستان ۱۴۰۰): ۶۵-۴۷
http://hii.khu.ac.ir
مقاله اصیل پژوهشی

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

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 «اصالت 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 «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 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.

نوع مقاله: مقاله پژوهشی
دریافت: ۱۴۰۰/۰۹/۲۱
پذیرش: ۱۴۰۰/۱۱/۲۷

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

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

Human Information Interaction. Vol. 8, No. 4 (Winter 2022):47-65.
http://hii.khu.ac.ir
Original Article

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.

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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)

The journal of *Human Information Interaction* is supported by Kharazmi University, Tehran, Iran.
This work is published under CC BY-NC-SA 3.0 licence.

مقدمه

داده‌های باز[1] به داده‌های غیرمحرمانه گفته می‌شود که بدون هیچ محدودیتی در استفاده یا توزیع، در دسترس قرار می‌گیرند (جانسن و همکاران[2]، ۲۰۱۲). داده‌های دولتی باز[3] زیرمجموعه‌ای از داده‌های باز است که به عنوان یک عامل اصلی برای دولت باز در نظر گرفته می‌شود (اتارد و همکاران[4]، ۲۰۱۵). دنیلس و همکاران[5] (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 (دنیلس و لوپز[6]، 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[7]، 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 (هالونن[8]، ۲۰۱۲). داده دولتی باز می‌تواند محرک نوآوری و رشد اقتصادی شود و ارتقای مدل‌های کسب و کار را به دنبال داشته باشد (یو و رابینسون[9]، 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[10] 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[11] (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 رسمی‌داده‌های 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.

نیکیفوروا و امسی براید[12] (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

[1] Open Data
[2] Janssen et al
[3] Open Government Data (OGD)
[4] Attard et al
[5] Danneels et al
[6] Daniels & Lopez
[7] Susha et al
[8] Halonen
[9] Yu & Robinson
[10] Classification
[11] De Juana-Espinosa & Luján-Mora
[12] Nikiforova & McBride

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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[1] (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[2] (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 سیاست‌گذارانی است که مایل به کنترل کیفیت مجموعه داده‌ها در پروژه‌ها یا برنامه‌های داده‌های دولتی باز هستند.

داوس و همکاران[3] (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.

میسوراکا و ویسکوسی[4] (۲۰۱۴) در مورد چارچوب ارزیابی انطباق داده‌های دولت 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 ژیائو[5] (۲۰۲۰) به بررسی چارچوب ارزیابی کیفیت داده‌های دولتی باز پرداخته‌اند. هدف این مطالعه ایجاد یک چارچوب مشترک به عنوان مرجعی 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[6] (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 «میانه»، 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،

[1] Dahbi et al
[2] Vetrò et al
[3] Dawes et al
[4] Misuraca & Viscusi
[5] Zhang & Xiao
[6] Zheng et al

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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 کمّی، 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 بعدهای ارزیابی مختلف و اهمیت هر یک مدّ نظر قرار گیرد. همچنین علاوه بر بررسی وضعیت کشور در این حوزه و خوشه‌بندی[1] 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.



شکل۱. 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[2]، 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 ، ۲۰۱۸؛ هوانگ و همکاران[3]، ۲۰۱۹؛ وترو و همکاران[4]، ۲۰۱۶؛ ساکسنا[5]، ۲۰۱۸): این شاخص تعداد مجموعه داده‌هایی که یک سازمان بصورت باز ارائه کرده است را ارزیابی می‌کند.

ب) تعداد دسته‌های داده (ساکسنا، ۲۰۱۸)

[1] Clustering
[2] Dahbi et al
[3] Huang et al
[4] Vetrò et al
[5] Saxena

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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[1]، 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[2]، 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.

الف) اصالت 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[3]): 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_j داده‌هایی که در j سال پیش منتشر شده است) با استفاده از فرمول ۱ محاسبه می‌شود.

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

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[4]، 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)

[1] Attard et al
[2] Veljković et al
[3] https://index.okfn.org/place/
[4] Susha et al

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

- 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)

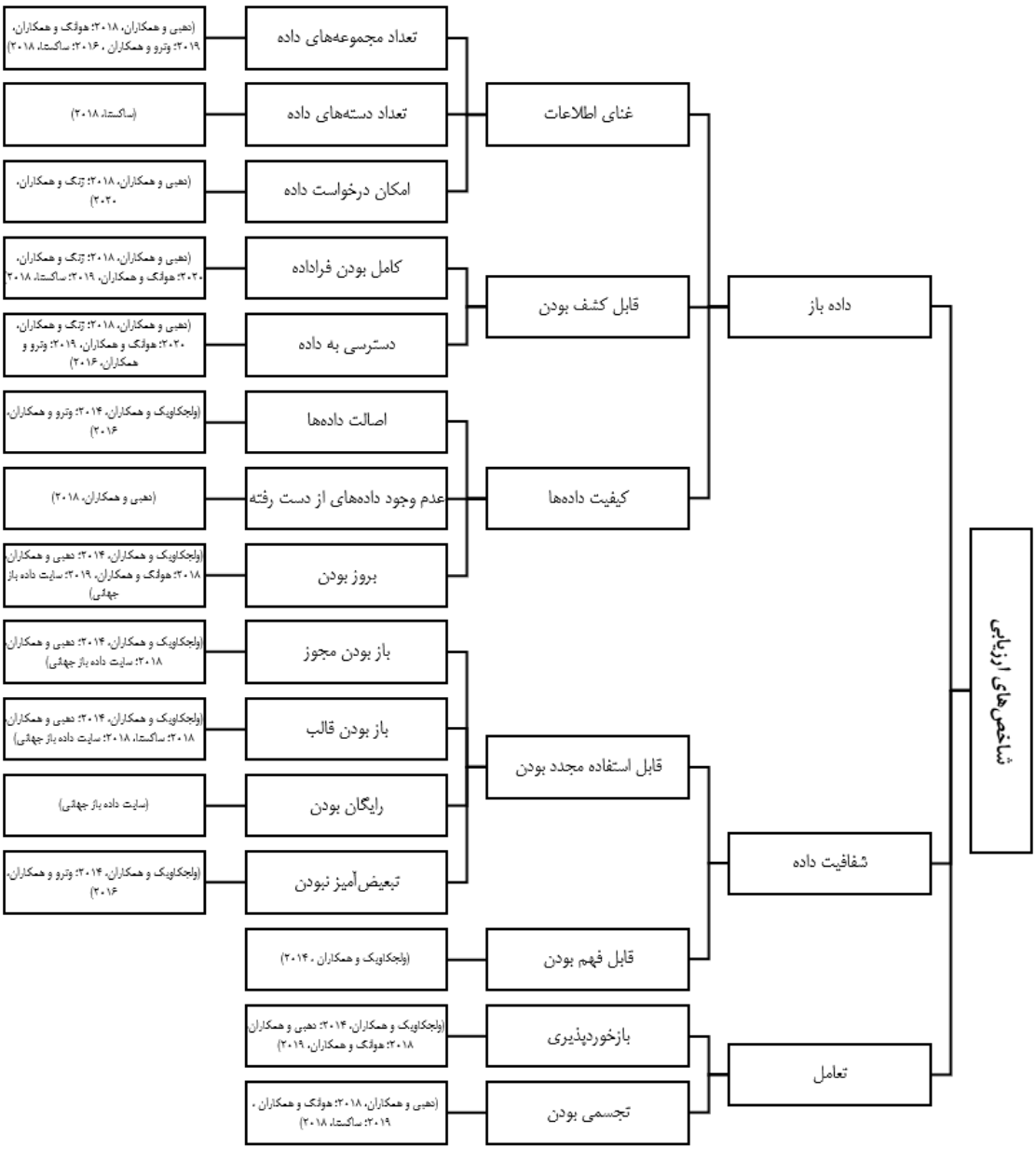
تعامل

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.



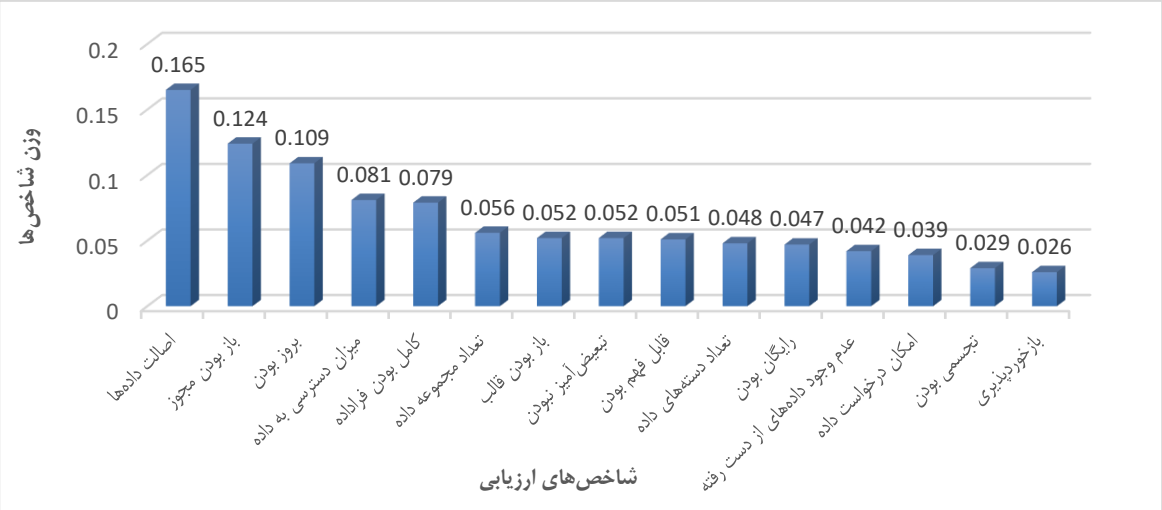
شکل۲. نمودار شاخص‌های ارزیابی بر اساس بعد، شاخص و مرجع

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

محاسبه وزن و اهمیت شاخص‌های ارزیابی

در این بخش از روش میدانی و تکنیک تحلیل سلسله مراتبی[1] 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.



شکل۳. وزن و اهمیت شاخص‌های ارزیابی حاصل از تکنیک تحلیل سلسله مراتبی

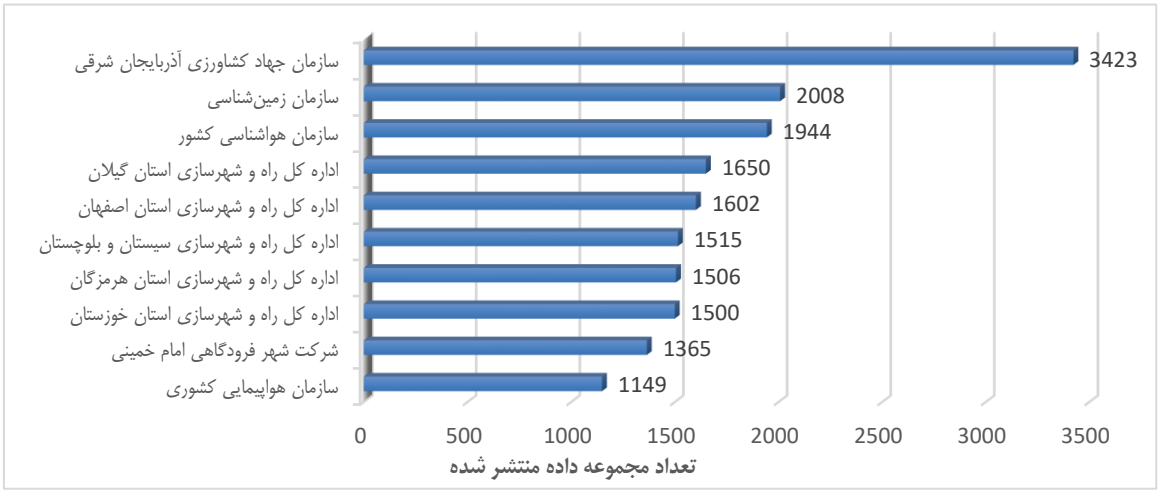
طبق نتایج بدست آمده، سه شاخص «اصالت داده‌ها»، «باز بودن مجوز» و «بروز بودن» به ترتیب سه رتبه نخست از بین شاخص‌های ارزیابی استخراج شده را کسب کرده‌اند.

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[2] و سامانه انتشار و دسترسی آزاد به اطلاعات[3] 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.



شکل۴. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تعداد مجموعه داده منتشر شده

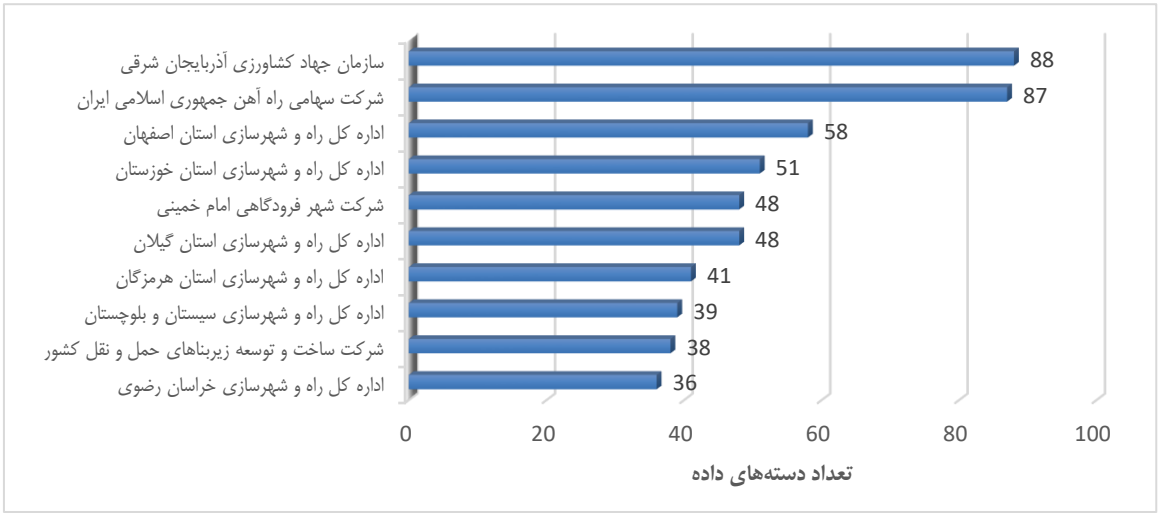
[1] Analytic Hierarchy Process (AHP)
[2] https://data.gov.ir/
[3] https://iranfoia.ir/

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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.

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.

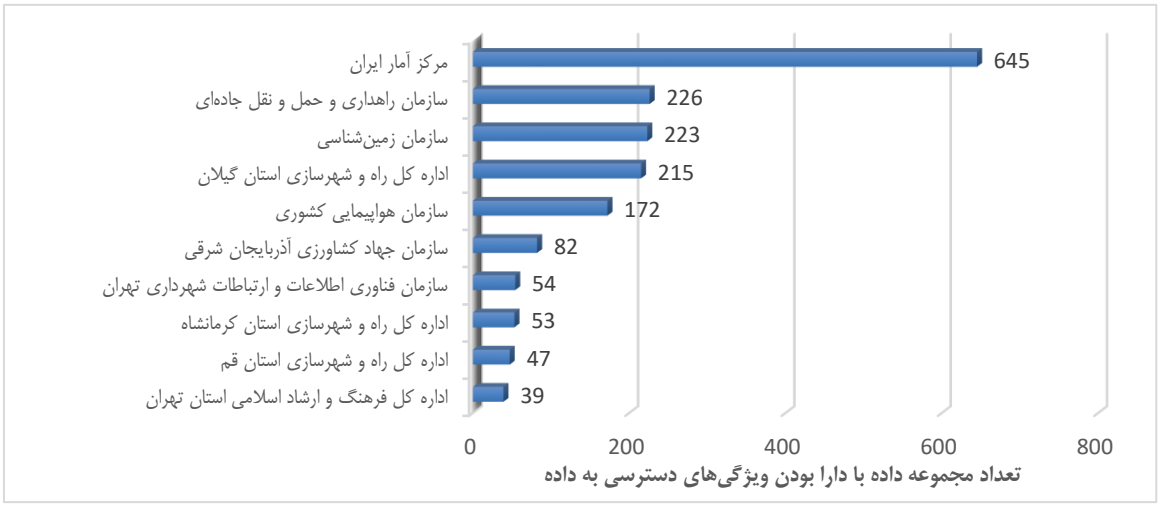


شکل۵. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تعداد دسته‌های داده

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

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.



شکل۶. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی میزان دسترسی به داده

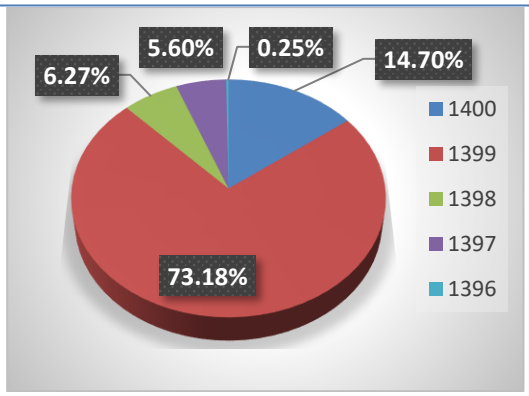
[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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.

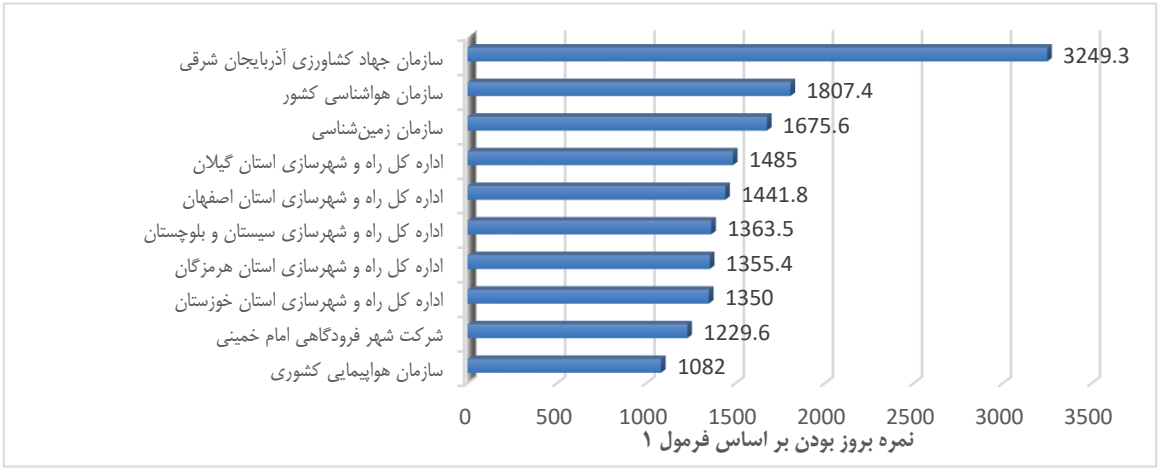
وضعیت سازمان‌ها و موسسات دولتی در ارائه داده‌ها بصورت باز بر اساس شاخص ارزیابی بروز بودن

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.



شکل۷. 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.

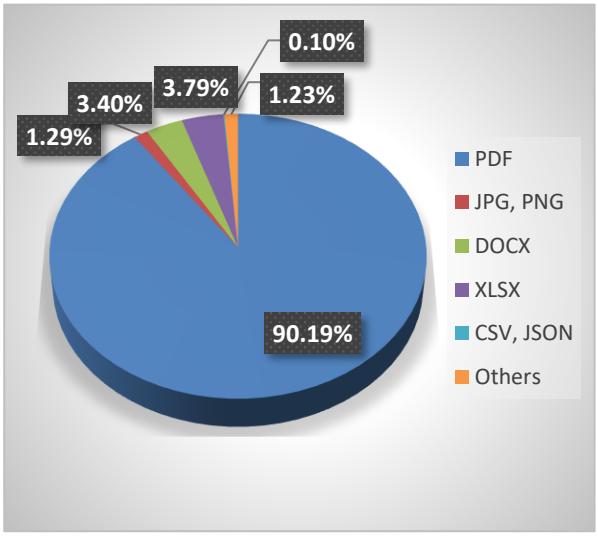


شکل۸. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی بروز بودن

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

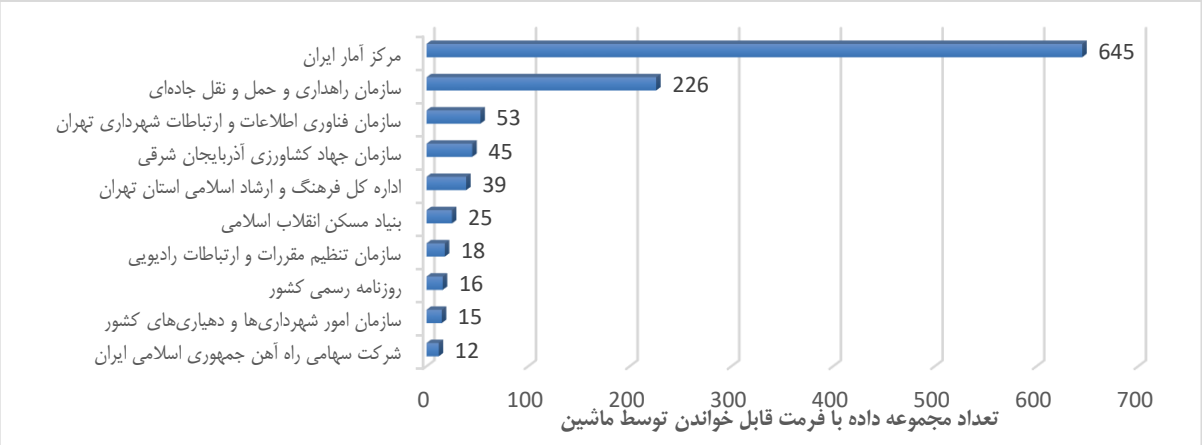
همان‌طور که پیش‌تر بیان شد، شاخص باز بودن قالب، قابل خواندن بودن مجموعه داده توسط ماشین را ارزیابی می‌کند. برخی فرمت‌های داده مانند PDF، JPG، PNG قابل خواندن توسط ماشین نیست. اما برخی فرمت‌ها مانند JSON قابلیت خواندن توسط ماشین را دارا هستند. در این بخش ابتدا تعداد مجموعه داده بر اساس فرمت‌های مختلف،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



شکل۹. درصد مجموعه داده‌های منتشر شده بر اساس فرمت‌های مختلف

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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

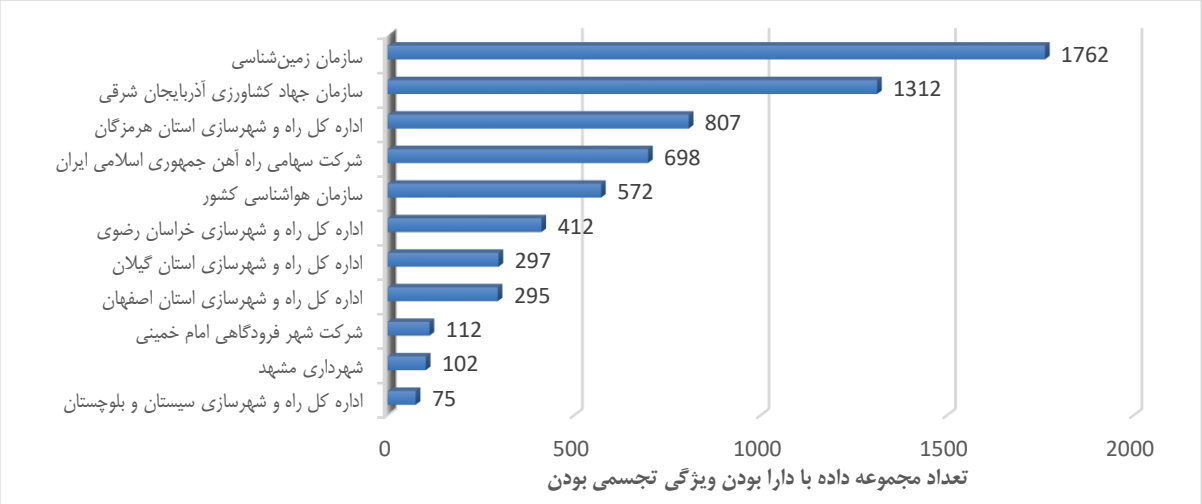


شکل۱۰. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی باز بودن قالب

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.

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.



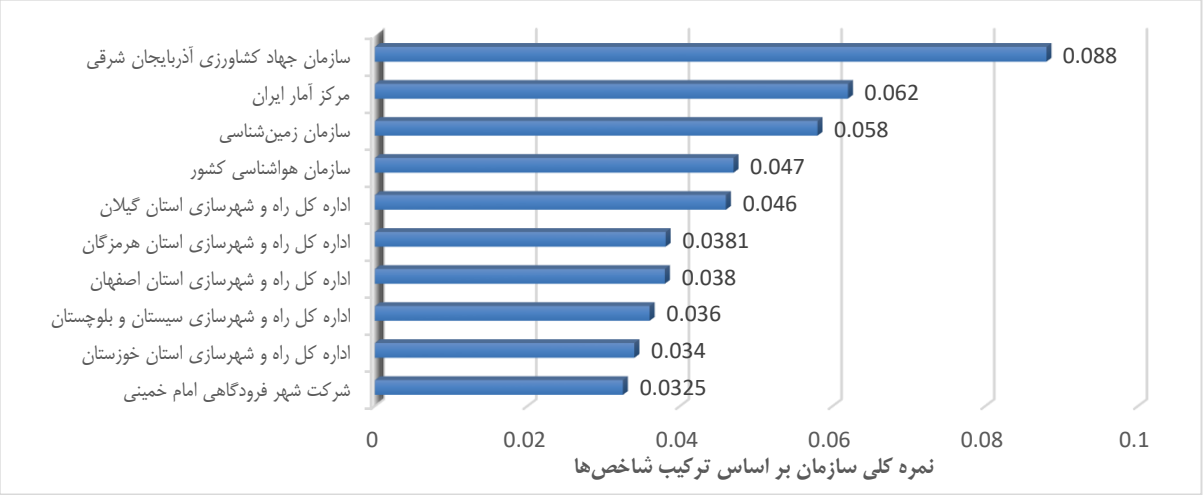
شکل۱۱. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس شاخص ارزیابی تجسمی بودن

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

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.

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]



شکل۱۲. ده رتبه نخست سازمان‌ها و موسسات دولتی بر اساس ترکیب شاخص‌های ارزیابی و وزن شاخص‌ها

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.

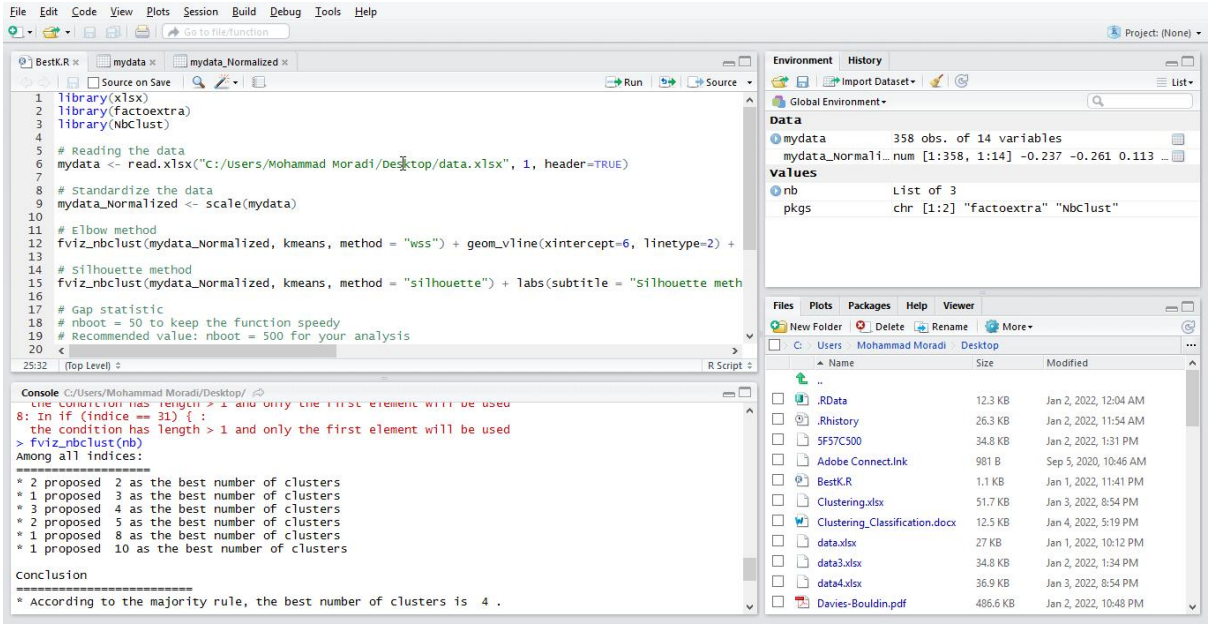
اعمال تکنیک‌های داده‌کاوی بر روی داده‌های استخراج شده

خوشه‌بندی

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

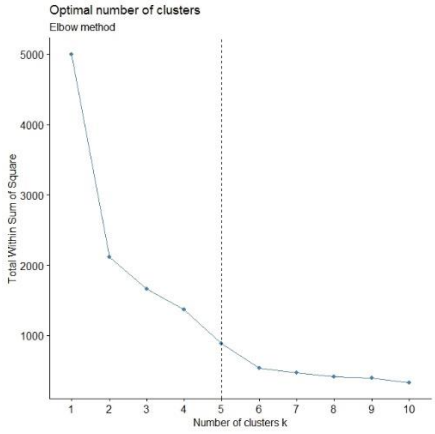
به منظور شناسایی تعداد خوشه‌های بهینه از کدنویسی در R استفاده شد. از ۱۰ معیار شامل silhouette، gap statistic، cindex، hubert، mcclain، frey، ptbiserial، ball، beale و duda استفاده شد. شکل ۱۳ نمایی از کد و برنامه R را نشان می‌دهد.



شکل۱۳. نمایی از کد و محیط برنامه R در تعیین تعداد بهینه خوشه‌ها

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

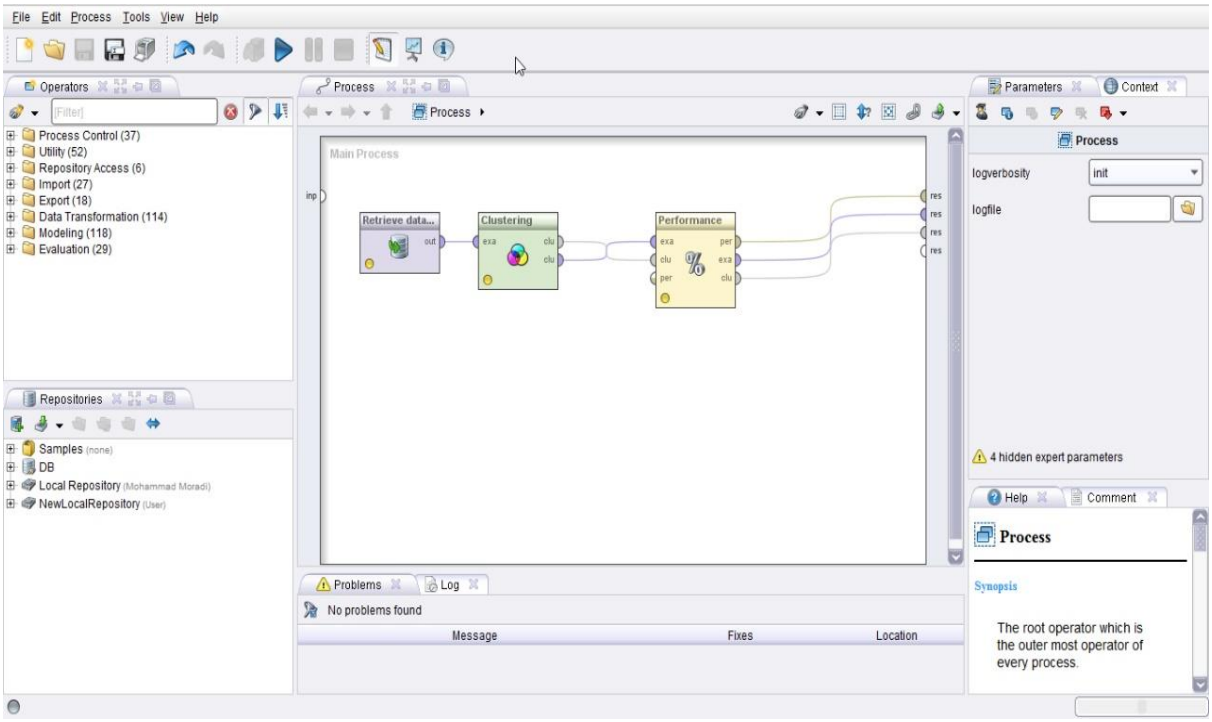
همان‌طور که از خروجی کنسول شکل ۱۳ مشخص است، بیشتر معیارها، ۴ خوشه را به عنوان تعداد خوشه بهینه انتخاب کردند. در توضیح نحوه تعیین تعداد خوشه بهینه در هر روش می‌توان به شکل ۱۴، که روش Elbow را نشان می‌دهد و از مجموع مربعات خطا[1] استفاده شده است، اشاره کرد. در این روش از خوشه ۵ به بعد تغییرات محسوسی در کاهش مربعات خطا 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.



شکل۱۴. روش Elbow در تعیین تعداد خوشه بهینه

الگوریتم خوشه‌بندی

در این بخش به اعمال الگوریتم خوشه‌بندی بر روی داده‌های استخراج شده پرداخته شده است. به دلیل حساس بودن K-means به داده‌های پرت[2] از الگوریتم K-medoids استفاده شده است. بدین منظور نرم‌افزار رپیدماینر[3] مورد استفاده قرار گرفته است. شکل ۱۵، نمایی از پیاده‌سازی خوشه‌بندی در نرم‌افزار رپیدماینر را نشان می‌دهد.

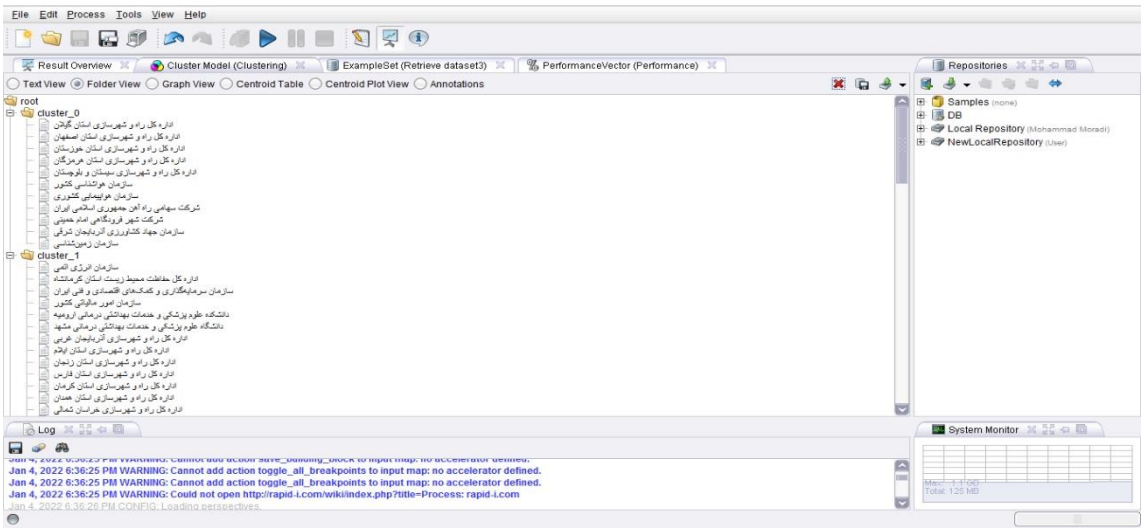


شکل۱۵. نمایی از پیاده‌سازی خوشه‌بندی در نرم‌افزار رپیدماینر

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

[1] Total Within Sum of Square
[2] Outlier
[3] Rapid Miner

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]



شکل۱۶. خروجی خوشه‌بندی در نرم‌افزار رپیدماینر

طبقه‌بندی

در این بخش به ایجاد یک مدل طبقه‌بندی بر اساس خوشه‌های مشخص شده در بخش پیشین پرداخته شده است. با استفاده از مدل طبقه‌بندی می‌توانیم به پیش‌بینی وضعیت سازمان‌ها و موسسات دولتی بپردازیم.

برچسب‌ها

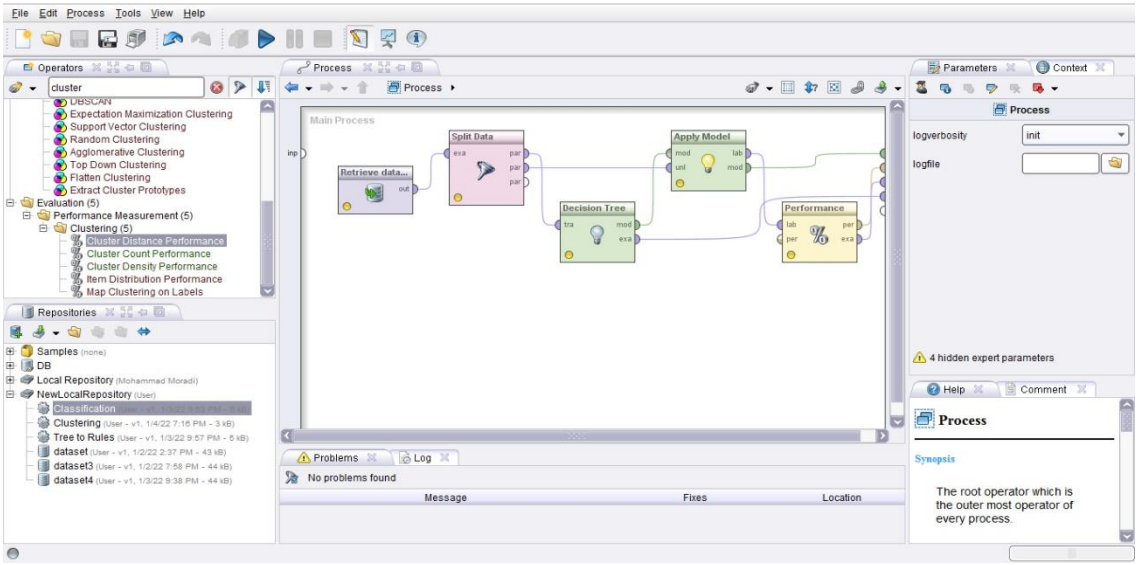
برچسب‌های در نظر گرفته شده و تعداد سازمان‌ها و موسسات دارای هر برچسب در جدول ۱ مشخص شده است.

جدول۱. 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[1] با استفاده از نرم‌افزار رپیدماینر پرداخته شد. شکل ۱۷، نمایی از پیاده‌سازی را نشان می‌دهد.



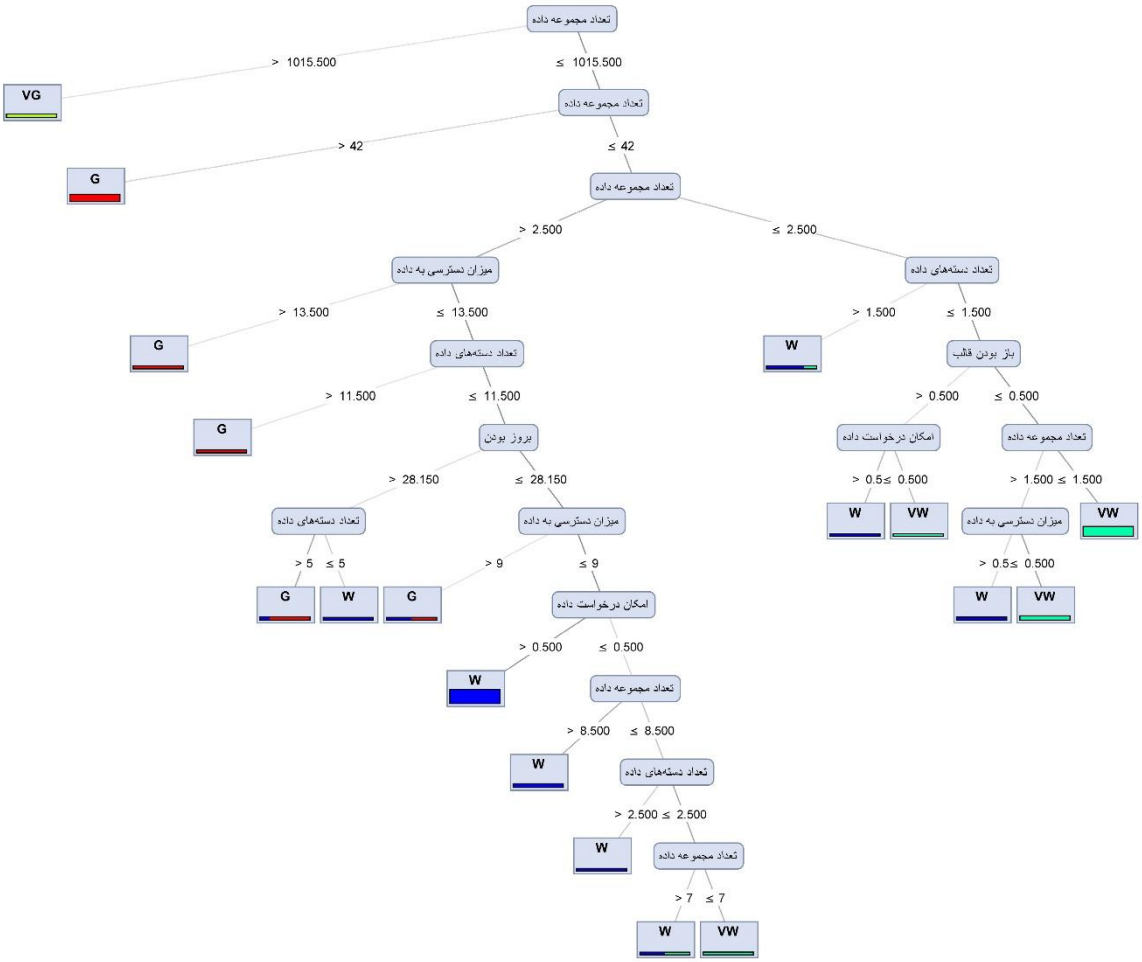
شکل۱۷. نمایی از پیاده‌سازی مدل طبقه‌بندی درخت تصمیم در نرم‌افزار رپیدماینر

[1] Decision Tree

تعامل انسان و اطلاعات
جلد هشتم، شماره چهارم (زمستان ۱۴۰۰): ۶۵-۴۷
http://hii.khu.ac.ir

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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.



شکل۱۸. درخت تصمیم

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.

ارزیابی مدل طبقه‌بندی

در این بخش به ارزیابی مدل طبقه‌بندی ایجاد شده پرداخته شده است. شکل ۱۹، ماتریس درهم‌ریختگی[1] درخت تصمیم ایجاد شده را نشان می‌دهد.

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%	

شکل۱۹. ماتریس درهم‌ریختگی درخت تصمیم ایجاد شده

[1] Confusion Matrix

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

صحت[1] مدل دسته‌بندی ایجاد شده برابر ۹۶.۲۶ ٪ است که نشان از خوب بودن مدل دارد. همچنین مقادیر سایر معیارها مانند بازیابی[2] و دقت[3] 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. متاسفانه 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 ابوالمعالی 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 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 «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.

[1] Accuracy
[2] Recall
[3] Precision

[Downloaded from hii.khu.ac.ir on 2024-03-28]
[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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 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 منافعی وجود ندارد.

منبع حمایت‌کننده

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

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., Vidiasova, 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

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

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).

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]

[Downloaded from hii.khu.ac.ir on 2024-03-28]

[DOR: 20.1001.1.24237418.1400.8.4.6.1]