Scientometrics Research Journal
Scientific Bi-Qualterly Journal of Shahed University/ Article in Press (Not edited)

Ghasem Azadi Ahmadabadi 1*

1. Assistant Professor, Policy evaluation and Monitoring of Science, Technology, and Innovation Department, *National Research Institute for Science Policy*, Tehran, Iran. azadi_gh@yahoo.com ORCID ID :0000-0002-3610-2573 (Corresponding author)

Receive: ../../….
Acceptance: ../../….

Predicting scientific research impacts in biotechnology by machine learning algorithms

Abstract

Purpose: "Research impact" is one of the key concerns of the stakeholders of this field, which can be considered as positive and profitable applications of research in various dimensions such as society, economy, environment, culture, health, etc. The current study aims to measure the relationship of each of the different variables of scientific outputs impacts on each other and also to examine which of the machine algorithms can predict the scientific, social, and economic impact of scientific outputs.

Methodology: The current research is applied in terms of purpose and descriptive in terms of method and has been done by scientometric approach. This study aims to investigate the relationship between the volume of scientific outputs and scientific cooperation as well as the scientific, social, and economic impact of Iran's scientific outputs in biotechnology; and which of the machine learning algorithms are better predictors for measuring the effectiveness of scientific outputs in different dimensions. The research community is Iran's biotechnology scientific outputs, which are indexed in the Scopus database from 2003 to 2024. The data extracted 1 Bahman 1402 equivalent to 21 January 2024. To extract relevant data, a Scival analytical database was used. In this research, Pearson's correlation coefficient and R software package were used to determine the relationship between the studied indicators. Machine learning algorithms such as multiple linear regression, nearest neighbor, decision trees, random forests, and gradient boosting were also used and evaluated as predictive models. To perform tests and algorithms, Python programming language has been used.

Findings: The findings of this study showed that Iran's scientific outputs in this field in the period from 2003 to 2023 had increased 36 times, which is considered extremely high progress. The relationship between international collaboration and indicators such as citations counts, Field-Weighted Citation Impact, Output in Top 10% Citation Percentiles, Patent-Citations Count, Patent-Citations per Scholarly Output, Scholarly Output cited by Patents, Patents Count, Views Count, Output in Top 10% Views Percentiles, Views per Publication, and Field-Weighted Views Impact of the domain were found to be positive and significant. The index of academic collaboration also has a positive and significant relationship with the citation counts, Field-Weighted Citation Impact, Output in the Top 10% Citation Percentiles, Publications in the Top 10% Journal Percentiles by Cite Score Percentile, Patents Count, the Scholarly Output cited by Patents, Patents Count, Views Count and Field-Weighted Views Impact. Academic-government collaboration also has a positive and significant relationship with three indicators citations per publication, Patent-Citations Count, and Patent-Citations per Scholarly Output. In the case of the impact of the studied scientific outputs, citation counts have a positive and significant relationship with the indicators Scholarly Output cited by Patents, Patents Count, Views Count, Views per Publication, and the Field-Weighted Views Impact of the biotechnology scientific outputs. In economic impact, the result indicated that the number of patent citations is one of the

representative indicators of this type of impact, with indicators such as academic Collaboration, international Collaboration, citations counts, citations per Publications, Field-Weighted Citation Impact, Views Count positive and significant relationship in Output in Top 10% Views Percentiles, Views per Publication and Field-Weighted Views Impact the biotechnology scientific outputs. In terms of social impact, it was also concluded that Views Count has a positive and significant relationship with indicators such as citation counts, Field-Weighted Citation Impact, the number of patent citations, Patent-Citations per Scholarly Output, Scholarly Output cited by Patents and Patents Count in the biotechnology field. Based on the obtained results, multivariate linear regression with a higher accuracy score and a lower standard deviation score could better predict the scientific, technological, and social impact of Iran's scientific outputs in biotechnology.

Conclusion: The most important factor affecting the quality of articles, including citations, views, and applicability, is international cooperation, and it is necessary to about measures in this regard. It is suggested to use a variety of indicators during the quantitative and qualitative evaluation of the articles to obtain a clearer picture of the effectiveness of the research. The important point for policymakers in science and technology for scientific development in Iran is that despite the quantitative and qualitative growth of scientific outputs, the country's research system should be directed towards value creation and creating added value, especially in the economic field.

Keywords: Predicting research impact, Machine learning algorithms, Scientific research impact, Economic research impact, Social research impact

ویرایش نشده

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی ـ پژوهشی دانشگاه شاهد/ زودآیند ویرایش نشده

پیش‌بینی تأثیرگذاری پژوهش‌های علمی حوزه زیست‌فناوری با استفاده از الگوریتم‌های یادگیری ماشین

قاسم آزادی احمدآبادی ۱*

۱. استادیار، گروه ارزیابی سیاست‌ها و پایش علم، فناوری و نوآوری، مرکز تحقیقات سیاست علمی کشور، تهران، ایران. کشور azadi_gh@yahoo.com تلفن: ۸۸۰۳۶۱۴۴ -
شناسه ارکید: 0000-0002-3610-2573 (نویسنده مسئول)

دریافت: ۰۰۰/۰۰/۰۰
پذیرش: ۰۰۰/۰۰/۰۰

چکیده

هدف: مطالعه حاضر قصد دارد رابطه هر یک از متغیرهای مختلف اثرگذاری بروندادهای علمی را بر همدیگر مورد سنجش قرار داده و نیز بررسی کند که کدام‌یک از الگوریتم‌های ماشین می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی بروندادهای علمی را پیش‌بینی کنند.

روش‌شناسی: پژوهش حاضر از لحاظ هدف، کاربردی و از نظر روش، توصیفی بوده و با رویکرد علم‌سنجی انجام شده است. جامعه پژوهش، بروندادهای حوزه زیست‌فناوری ایران است که در پایگاه اسکوپوس در بازه ۲۰۰۳-۲۰۲۴ نمایه شده‌اند. به‌منظور استخراج داده‌ها از پایگاه تحلیلی سایول استفاده شد. در این پژوهش از ضریب همبستگی پیرسون و از بسته نرم‌افزاری R به‌منظور تعیین رابطه بین شاخص‌های مورد مطالعه استفاده شد و رگرسیون خطی چندگانه، نزدیک‌ترین همسایه، درخت‌های تصمیم‌گیری، جنگل‌های تصادفی و تقویت گرادیان نیز به‌عنوان مدل‌های پیش‌بینی‌کننده، مورد و ارزیابی قرار گرفت. به‌منظور انجام آزمون‌ها و الگوریتم‌ها از زبان برنامه‌نویسی پایتون استفاده شد.

یافته‌ها: در عرصه اثرگذاری خروجی‌های علمی مورد مطالعه، حجم استنادها با شاخص‌های متعددی رابطه مثبت و معنی‌داری داشته است. در حوزه تأثیرگذاری اقتصادی نیز این نتیجه حاصل شد که تعداد استنادات ثبت اختراع به عنوان یکی از شاخص‌های معرف این نوع تأثیر، با موارد متعددی رابطه مثبت و معنی‌داری داشته است. در مورد تأثیرگذاری اجتماعی نیز تعداد بازدیدها رابطه مثبت و معنی‌داری با بسیاری از شاخص‌ها دارد. براساس نتایج حاصل شده، رگرسیون خطی چندمتغیره با نمره صحت بالاتر و نمره انحراف معیار پایین‌تر، بهتر توانست میزان اثرگذاری علمی، فناورانه و اجتماعی برونداهای علمی ایران در حوزه زیست‌فناوری را پیش‌بینی نماید.

نتیجه‌گیری: مهمترین عامل مؤثر بر کیفیت مقالات ازجمله در بعد استنادها، بازدیدها و کاربردی بودن، همکاری بین‌المللی بوده که در این خصوص لازم است تدابیری اندیشه شود. پیشنهاد می‌شود هنگام ارزیابی کمّی و کیفی مقالات، از شاخص‌های متنوعی استفاده شود تا تصویر شفاف‌تری از اثرگذاری پژوهش‌ها حاصل شود.

واژگان کلیدی: پیش‌بینی تأثیرگذاری پژوهش‌ها، الگوریتم‌های یادگیری ماشین، اثرگذاری علمی پژوهش‌ها، اثرگذاری اقتصادی پژوهش‌ها، اثرگذاری اجتماعی پژوهش‌ها

زودآیند ویرایش نشده

مقدمه و بیان مسئله

یکی از عوامل اصلی مؤثر بر توسعه هر جامعه، کمیت و کیفیت فعالیت‌ها و خروجی‌های علمی و فناورانه پژوهشگران و فناوران آن سرزمین است. به همین میزان، بحث سنجش این تولیدات از زوایای گوناگون نیز از اهمیت بسیار زیادی برخوردار شده است. از ملزومات سیاست‌گذاری کارآمد و درست در حوزه علم و تحلیل حرکت علمی کشور، انجام پژوهش‌هایی به منظور شناسایی وضعیت موجود در جهت شناسایی نقاط قوت و ضعف آنهاست. شاخص‌های علم‌سنجی، مهم‌ترین ابزارهایی که می‌تواند چنین اطلاعاتی را در اختیار برنامه‌ریزان و سیاست‌گذاران قرار دهد (آزادی احمدآبادی، ۱۴۰۳).

با توجه به نقش و اهمیت پژوهش در توسعه و پیشرفت جوامع در برنامه‌های کلان کشورها در جهت نیل به توسعه و پیشرفت، به این امر توجه اساسی شده و سرمایه‌گذاری‌های کلانی در دانشگاه‌ها و مراکز پژوهشی در این رابطه می‌شود. از این جهت، سرمایه‌گذاران و ذی‌نفعان مایل‌اند بدانند پژوهش‌های آنها چه اثرات علمی درون دانشگاهی و چه اثرات اجتماعی، فرهنگی، اقتصادی و سیاسی خارج از دانشگاه دارد (بابااکبری و همکاران، ۱۴۰۰). از سویی دیگر، ارزیابی تأثیر[1] پژوهش‌ها و تعیین میزان تحقق اهداف به‌آسانی ممکن نیست؛ زیرا اثربخشی واقعی پژوهش‌ها غیرخطی و غیرقابل پیش‌بینی هستند. آنچه امروزه اغلب برای ارزیابی اثربخشی پژوهش مورد توجه قرار می‌گیرد معیارهای سنتی ازقبیل تعداد مقالات، ضریب تأثیر مجلات، تعداد گرنت‌ها و تعداد استنادها و غیره هستند (Newson et al. 2015). در واقع، «تأثیر پژوهش» که می‌توان آن را به معنی کاربردهای مثبت و سودآور پژوهش در زمینه‌های مختلف نظیر اجتماع، اقتصاد، محیط‌زیست، فرهنگ، بهداشت و غیره در نظر گرفت از دغدغه‌های کلیدی ذی‌نفعان این حوزه به‌شمار می‌رود (آزادی احمدآبادی و همکاران، ۱۴۰۱).

واقعیت آن است که سنجش تأثیر پژوهش‌ها در ابعاد مختلف به‌سادگی ممکن نیست. آن‌چه که در حال حاضر و به‌طور واقعی در رابطه با تأثیر پژوهش‌های انجام گرفته می‌توان پیگیری کرد خروجی‌های علمی است که در پایگاه‌های استنادی بین‌المللی نمایه شده و امکان پیگیری تأثیرگذاری علمی، اجتماعی و اقتصادی آنها تا حدی امکان‌پذیر شده است.

با افزایش تعداد مقالات علمی منتشر شده در سطح جهان، نیاز به ارزیابی و روش‌های نوین سنجش کیفیت مقالات پژوهشی در حال افزایش است (Abrishami & Aliakbary, 2019). براین اساس، سنجه‌های مختلفی به‌منظور سنجش و ارزیابی تولیدات علمی به جامعه علمی معرفی شده که هر کدام، مزایا و معایب خاص خود را دارد و هر روزه برای تکمیل آنها سنجه‌های جدیدتری ایجاد می‌شود. سالیان زیادی است که در حوزهٔ علم‌سنجی در راستای بررسی اثرگذاری علمی پژوهشگران از روش تحلیل استنادی استفاده می‌شود (Bornmann et al. 2013). دگرسنجی نیز یکی از معیارهای نوینی است که بر پایه شاخص‌های مبتنی بر فعالیت در رسانه‌های اجتماعی استوار بوده و به‌عنوان ابزاری برای سنجش تأثیر علمی در محیط وب ظاهر شده است (Wooldridge & King, 2019) که می‌تواند نمایانگر تأثیر فعالیت‌های علمی بر سطوح مختلف جامعه باشد. در کنار این شاخص‌ها برخی ابزارهای تحلیلی که ورودی خود را از پایگاه‌های استنادی دریافت می‌کنند، به‌دنبال شناسایی اثرگذاری پژوهش‌های علمی بر فناوری بوده‌اند. بر این اساس، پایگاه سایول[2]، شاخص جدیدی را مطرح کرده که براساس آن، میزان کاربردی بودن پژوهش‌ها را رصد می‌کند. این کار ازطریق پیگیری میزان استنادهای دریافتی و خروجی به پروانه‌های ثبت اختراع انجام می‌شود. در کنار بررسی و ارزیابی اثر پژوهش‌ها که عمدتاً مبتنی بر گذشته است، پیش‌بینی آن نیز از حوزه‌های جذاب و کاربردی در این زمینه می‌تواند باشد. این پیش‌بینی در سطوح مختلف و با رویکردها و شاخص‌های مختلفی قابل انجام است. با توجه به اهمیت پیگیری تأثیرات مختلف بروندادهای علمی و تعیین ارتباط بین آنها مطالعه حاضر قصد دارد رابطه هر یک از متغیرهای مختلف بر همدیگر را مورد سنجش قرار داده و نیز بررسی کند که کدام‌یک از الگوریتم‌های ماشین می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی بروندادهای علمی را پیش‌بینی کنند. به این منظور، بروندادهای حوزه زیست‌فناوری[3] ایران که در پایگاه اسکوپوس[4] در بازه ۲۰۰۳-۲۰۲۴ نمایه شده‌اند مورد مطالعه قرار گرفته است. درواقع، پرسش اساسی این پژوهش آن است که کدام‌یک از الگوریتم‌های ماشین در مقایسه با یکدیگر می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی پژوهش‌های حوزه زیست‌فناوری ایران را بهتر پیش‌بینی کنند؟

---

[1] .Impact
[2] . Scival
[3] . Biotechnology
[4] . Scopus

پرسش‌های پژوهش

۱. وضعیت بروندادهای حوزه زیست‌فناوری ایران براساس هر کدام از شاخص‌های اثرگذاری علمی، اجتماعی و اقتصادی چگونه است؟

۲. آیا بین تعداد بروندادها و همکاری‌های علمی بر شاخص‌های مربوط به اثرگذاری علمی، اجتماعی و اقتصادی آنها رابطه وجود دارد؟

۳. کدام یک از شاخص‌های اثرگذاری علمی، اجتماعی و اقتصادی پژوهش‌ها با یکدیگر ارتباط دارند؟

۴. کدام‌یک از الگوریتم‌های ماشین در مقایسه با یکدیگر می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی پژوهش‌ها را بهتر پیش‌بینی کنند؟

چارچوب نظری

ارزیابی پژوهش

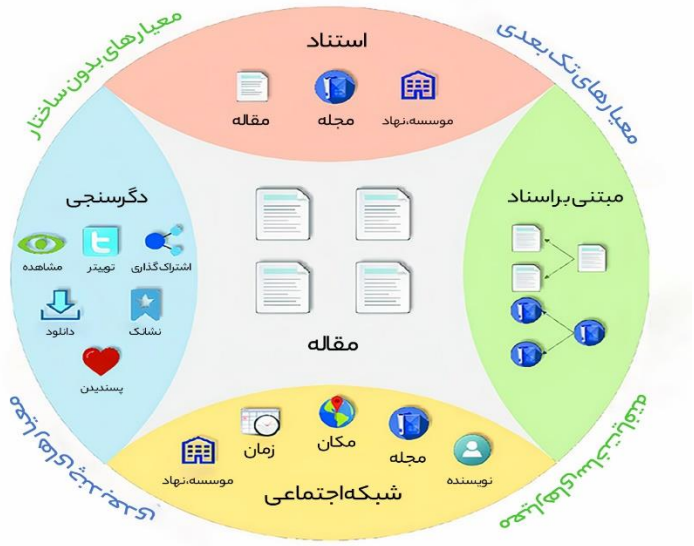
ارزیابی پژوهش، مجموعه‌ای از مفاهیم و روش‌هاست برای تعیین اینکه آیا برتری در این زمینه، حاصل شده و تمرکز آن بر «کیفیت» و «تأثیر» است. چگونگی درک و سنجش تعالی، نحوه تصمیم‌گیری پژوهشگران را شکل می‌دهد و به نوبه خود، انواع دانشی تولید می‌شود که به‌طور بالقوه به نتایج سیاسی، اقتصادی یا اجتماعی تبدیل می‌شود. ارزیابی پژوهش به‌عنوان یک فعالیت فنی یا عملیاتی در نظر گرفته می‌شود که برای به تصویر کشیدن عناصر عینی فرآیند پژوهش، طراحی و پیاده شده و توسط سیاست‌گذاران علمی تشویق و پیگیری می‌شود (Williams & Lewis, 2021). مطالعات علم‌سنجی، مبتنی بر تجزیه و تحلیل کمّی پیشرفت‌های علمی است و در حوزه «نتایج پژوهش»، تلاش می‌کند تأثیر را اندازه‌گیری کند. این ارزیابی (مطابق شکل ۱) می‌تواند در سطوح مختلف و شاخص‌های متفاوت سنجیده شود (آزادی احمدآبادی، زودآیند).

شکل ۱. سطوح مختلف سنجش پژوهش‌ها با رویکرد علم‌سنجی (آزادی احمدآبادی، ۱۴۰۳)

امروزه مجموعه متنوعی از شاخص‌ها به‌منظور ارزیابی کیفیت و اثرگذاری پژوهش‌ها، از معیارهای سنتی تا جدیدترین معیارهای جدید و جایگزین به‌کار گرفته می‌شود. کتاب‌سنجی یکی از این حوزه‌های دیرینه‌ای است که شاخص‌های مبتنی بر استناد را برای سنجش پژوهش‌های علمی به خدمت می‌گیرد. در تحلیل‌های استنادی، سنجش کیفیت مقاله بر پایه استناد اتفاق می‌افتد. اگرچه استناد، سنجه علمی است ولی عدد مطلق نمی‌تواند گویای کیفیت باشد. به همین منظور، استنادها به‌هنجار می‌شوند تا قابل تعمیم به جامعه مورد بررسی باشند. عموماً به مقاله‌های پراستناد به‌عنوان نشانه‌ای از کیفیت فعالیت‌های علمی و در نتیجه، ارزیابی عملکرد پژوهشی توجه خاصی می‌شود. با وجود اینکه استناد از مهم‌ترین و قابل

قبول‌ترین شاخص‌ها برای ارزیابی عملکرد و کارایی علمی است، با محدودیت‌هایی چون وابستگی زمانی، وابستگی به پایگاه‌های گران قیمت و با محدودیت دسترسی همراه است (Guthrie et al. 2013). گام‌های روش‌شناختی در مطالعات ارزیابی پژوهش از اهمیت بالایی برخوردار است. آنینوس این مراحل را این چنین شرح می‌دهد: فرآیند ارزشیابی پژوهش با تعیین سطح علمی آغاز می‌شود. به‌دنبال آن، دوره تجزیه‌وتحلیل، پایگاه‌های داده برای استخراج داده‌ها و انواع قالب‌های انتشار برای تجزیه‌وتحلیل تعیین می‌شود. اگر ارزیابی مربوط به یک گروه پژوهشی (دانشکده) یا یک واحد پژوهشی باشد، باید انتشارات گروه و سازوکار شمارش مناسب، مشخص شود. تجزیه‌وتحلیل استنادی انجام شده و شاخص‌های کتاب‌سنجی در مرحله نهایی، گزارش می‌شود. آنینوس متوجه شد که برخی از شاخص‌ها در ارزیابی پژوهش، مفیدتر از سایرین هستند. این موارد شامل تعداد انتشارات؛ تعداد استنادها؛ میانگین تعداد استنادها در هر نشریه؛ درصد انتشارات بدون استناد و میانگین تعداد استناد به کلیه مقالات منتشر شده توسط بخش یا واحد هستند (Anninos, 2014).

ارزیابی و پیش‌بینی تأثیر مقاله در دهه‌های گذشته توجه زیادی را در عرصه دانشگاهی و علمی به خود جلب کرده است. تغییر رویکردها به این منظور، از یک بعد به چند بعد، از معیارهای غیرساختارمند به معیارهای ساخت‌یافته رخ داده که در شکل شماره ۲ به نمایش درآمده است.



شکل ۲. روش‌های ارزیابی و پیش‌بینی تأثیر مقاله (Bai, et al. 2017)

استنادها به‌طور ضمنی، قضاوت جامعه درباره اهمیت مقاله را رمزگذاری می‌کنند و بنابراین سیگنالی منحصر به فرد برای مطالعه تأثیر علمی ارائه می‌دهند. تلاش‌ها در درک و بازخوانی این سیگنال در مدل‌سازی احتمالی شبکه‌های استنادی و تکثیر معیارهای تأثیر مبتنی بر استناد مانند شاخص هیرش[1] منعکس می‌شود. درحالی‌که این تلاش‌ها بر درک گذشته و حال تمرکز دارند، این سؤال را باز می‌گذارند که آیا می‌توان تأثیر علمی را در آینده پیش‌بینی کرد. این نتایج را می‌توان با استفاده از تکنیک‌های غیرخطی به‌خوبی انجام داد. به‌ویژه، که این روش‌های هوش مصنوعی می‌توانند اندازه‌گیری‌های تأثیر علمی را برای مقالات و نویسندگان، یعنی نرخ استناد و شاخص‌های اچ، با دقت شگفت‌آوری پیش‌بینی کنند (Weihs & Etzioni, 2017). در عین حال، به منظور ارزیابی عینی تأثیر مقاله و پیش‌بینی دقیق تأثیر آن در آینده، تکنیک‌های یادگیری ماشین و داده‌کاوی نقش‌های حیاتی ایفا می‌کنند، مانند استخراج شخصیت‌های مهم شبکه‌های علمی و بهینه‌سازی عملکرد الگوریتم‌ها (Jordan & Mitchell, 2015). این ظرفیت پیش‌بینی می‌تواند به دانشگاه‌ها، دولت‌ها و سرمایه‌گذاران با هشدار دادن به مقالات تأثیرگذار آینده و همچنین به محققان، مؤسسات و حوزه‌هایی که چنین مقالاتی را تولید می‌کنند، کمک کند. پتانسیل ارزشمندی برای چنین ابزار تحلیلی در راستای کمک به توسعه سیاست و تصمیم‌گیری وجود دارد (McNamara et al. 2013).

مدل‌های پیش‌بینی‌کننده با استفاده از الگوریتم‌های یادگیری ماشین

1 . Hircsh Index

عمده‌ترین الگوهای پیش‌بینی‌کننده با استفاده از الگوریتم‌های یادگیری ماشین که در این مطالعه نیز از آنها استفاده شده، به شرح زیر هستند:

۱. درخت تصمیم[1]

درختان تصمیم یکی از محبوب‌ترین الگوریتم‌های یادگیری ماشینی با توجه به درک و سادگی آنها هستند (Wu et al. 2008). از درخت تصمیم به‌عنوان یک رویکرد یادگیری تحت نظارت در آمار، داده‌کاوی و یادگیری ماشین مطرح است. در این موارد، درخت تصمیم طبقه‌بندی یا رگرسیون به‌عنوان یک مدل پیش‌بینی برای نتیجه‌گیری در مورد مجموعه‌ای از مشاهدات استفاده می‌شود. مدل‌های درختی که در آن، متغیر هدف می‌تواند مجموعه‌ای از مقادیر مجزا داشته باشد، درختان طبقه‌بندی نامیده می‌شوند. در این ساختارهای درختی، برگ‌ها نشان‌دهنده برچسب‌های طبقه و شاخه‌ها نشان‌دهنده ترکیبی از ویژگی‌هایی هستند که به آن برچسب‌های طبقه منتهی می‌شوند. درخت‌های تصمیم که در آن متغیر هدف می‌تواند مقادیر پیوسته (معمولاً اعداد واقعی) بگیرد، درخت‌های رگرسیون نامیده می‌شوند. به‌طور کلی‌تر، مفهوم درخت رگرسیون را می‌توان به هر نوع شی مجهز به تفاوت‌های زوجی مانند دنباله‌های طبقه‌بندی بسط داد (Studer et al 2011).

۲. جنگل تصادفی[2]

جنگل تصادفی از الگوریتم‌های محبوب درخت تصمیم‌گیری بوده و یک الگوریتم پیش‌بینی استاندارد در یادگیری ماشین نظارت شده است (Bernard et al. 2012). الگوریتم‌های جنگل تصادفی در مواجهه با داده‌های با ابعاد بالا، سریع، انعطاف‌پذیر و قوی بوده و به‌طور گسترده به‌عنوان یک رویکرد یادگیری ماشینی برای انواع زیادی از وظایف پذیرفته شده‌اند (Ziegler & König, 2014).

۳. رگرسیون[3]

رگرسیون، تکنیکی است که برای دو نظریه استفاده می‌شود: اول، تحلیل‌های رگرسیون معمولاً برای پیش‌بینی استفاده می‌شوند که همپوشانی‌های عمده‌ای با حوزه یادگیری ماشین دارند. دوم، از تحلیل رگرسیون می‌توان در برخی موارد برای تعیین روابط علّی بین متغیرهای مستقل و وابسته بهره برد. نکته مهم این است که رگرسیون‌ها تنها روابط بین یک متغیر وابسته و مجموعه داده ثابت از متغیرهای مختلف را نشان می‌دهند (Maulud & Abdulazeez, 2020). رگرسیون خطی[4] معمولاً در روش‌های پژوهش ریاضی استفاده می‌شود، جایی که امکان اندازه‌گیری اثرات پیش‌بینی‌شده و مدل‌سازی آن‌ها در برابر متغیرهای ورودی چندگانه وجود دارد. این روشی برای ارزیابی و مدل‌سازی داده است که روابط خطی را بین متغیرهای وابسته و مستقل برقرار می‌کند. بنابراین این روش، روابط بین متغیرهای وابسته و متغیرهای مستقل را از تحلیل و یادگیری تا نتایج آموزشی فعلی مدل می‌کند (Wu et al. 2019).

۴. تقویت گرادیان[5]

تقویت گرادیان یک تکنیک یادگیری ماشینی است که مبتنی بر تقویت در یک فضای عملکردی بوده و یک مدل پیش‌بینی را در قالب مجموعه‌ای از مدل‌های پیش‌بینی ضعیف ارائه می‌کند؛ به‌عنوان مثال، مدل‌هایی که مفروضات بسیار کمی در مورد داده‌ها ایجاد می‌کنند، که معمولاً درخت‌های تصمیم ساده هستند. هنگامی که درخت تصمیم یادگیرنده ضعیف باشد، الگوریتم حاصل درختان با گرادیان تقویت شده نامیده می‌شود (Piryonesi & El-Diraby, 2020; Hastie et al. 2009).

۵. قانون k–نزدیک‌ترین همسایه[6]

قانون k–نزدیک‌ترین همسایه نیز از قدیمی‌ترین و ساده‌ترین روش‌ها برای طبقه‌بندی الگوهاست. با این وجود، اغلب نتایج رقابتی را به همراه دارد و در حوزه‌های خاصی، هنگامی که هوشمندانه با دانش قبلی ترکیب می‌شود، به‌طور قابل‌توجهی پیشرفت‌های عمده‌ای را ایجاد کرده است. این قانون، هر نمونه بدون برچسب را با برچسب اکثریت در میان نزدیک‌ترین همسایگان خود در مجموعه آموزشی طبقه‌بندی می‌کند. بنابراین عملکرد آن به‌شدت به شاخص فاصله مورد استفاده برای شناسایی نزدیک‌ترین همسایگان بستگی دارد. در الگو کی–نزدیک‌ترین همسایه برای طبقه‌بندی آماری و رگرسیون، در هر دو حالت کی شامل نزدیک‌ترین مثال آموزشی در فضای داده‌ای می‌باشد و خروجی آن بسته به نوع مورد استفاده در

[1] . Decision Tree
[2] . random forests
[3] . Regression
[4] . Linear Regression
[5] . Gradient boosting
[6] . k-nearest neighbors' algorithm

طبقه‌بندی و رگرسیون متغیر است. در حالت طبقه‌بندی با توجه به مقدار مشخص شده برای کی، به محاسبه فاصله نقطه‌ای که می‌خواهیم برچسب آن را مشخص کنیم با نزدیک‌ترین نقاط می‌پردازد. برای محاسبه این فاصله می‌توان از روش‌های مختلفی استفاده کرد که یکی از مطرح‌ترین این روش‌ها، فاصله اقلیدسی است. با توجه به این که محاسبات این الگوریتم براساس فاصله است نرمال‌سازی داده‌ها می‌تواند به بهبود عملکرد آن کمک کند (Weinberger & Saul, 2009 Hastie & Tibshirani, 1995;).

پیشینه پژوهش

به‌منظور بررسی رابطه بین متغیرهای علم‌سنجی و نیز پیش‌بینی آنها مطالعات مختلفی انجام شده که در ادامه به برخی از مهم‌ترین و مرتبط‌ترین آنها اشاره می‌شود.

بذرافشان و همکاران (۱۴۰۲) به بررسی قدرت پیش‌بینی تعداد استنادات دریافتی مدارک حوزه فیزیک ذرات در پایگاه اسکوپوس ازطریق نمرات دگرسنجی در پایگاه پلام‌ایکس پرداختند. یافته‌های حاصل از بررسی رابطه همبستگی و رگرسیون چندگانه بین تعداد استنادات و نمرات دگرسنجی نشان داد که رابطه قوی بین استنادات پلام‌ایکس و تعداد استنادات اسکوپوس برقرار است. همچنین بین ذکر و یاداشت پلام ایکس مشاهدات پلام‌ایکس با تعداد استنادات دریافتی در اسکوپوس رابطه معناداری وجود دارد. اما بین کاربرد پلام‌ایکس و شبکه‌های اجتماعی پلام‌ایکس رابطه‌ای وجود ندارد.

پژوهش بیرانوند و همکاران (۱۴۰۱) به بررسی امکان پیش‌بینی تعداد استنادات دریافتی تولیدات علمی حوزه وب‌معنایی ازطریق شاخص‌های سایت–اسکور، اس‌ان‌آی‌پی و اس‌جی‌آر نشریات این حوزه پرداخت. نتایج آزمون ضریب همبستگی پیرسون و رگرسیون خطی چندگانه نشان داد که همبستگی بین تعداد استنادات با شاخص اس‌ان‌آی‌پی وجود ندارد. همچنین بین تعداد استنادات با اس‌جی‌آر و تعداد استنادات با سایت‌اسکور، رابطه معناداری وجود ندارد. هیچ‌یک از شاخص‌های سه‌گانه سایت‌اسکور، اس‌ان‌آی‌پی و اس‌جی‌آر توان پیش‌بینی تعداد استنادات دریافتی در پایگاه اسکوپوس را ندارند.

هدف پژوهش ابراهیمی و همکاران (۱۳۹۶) امکان‌سنجی ارائهٔ مدلی برای سنجش میزان تأثیرگذاری مقالات در مرحلهٔ داوری در مجلات علمی بود تا بتوان در فرایند کنترل کیفیت مقالات علمی، میزان استناد به آن در آینده را پیش‌بینی نمود. در این پژوهش، ۳۰۰ مقاله حوزهٔ شیمی بررسی شد. نتایج نشان داد که شاخص «چندنویسندگی و هم‌تألیفی» قوی‌ترین متغیر پیش‌بینی کنندهٔ استناد است که به‌صورت مثبت و معنادار میزان استناد را پیش‌بینی می‌کند. شاخص‌های «طول مقاله»، «تعداد منابع» و«شاخص اچ نویسنده» متغیرهای دیگری بودند که توانستند میزان استناد مقالات را به‌نحو مثبت و معنادار پیش‌بینی کنند.

گو و کرن این فرضیه را آزمون کردند که پیش‌بینی‌های خوبی از تعداد استنادهای بلندمدت را می‌توان از طریق ترکیبی از استنادهای اولیه یک نشریه و ضریب تأثیر مجله میزبان به دست آورد. این آزمایش با استفاده از مدل های رگرسیون خطی روی مجموعه‌ای از ۱۲۳۱۲۸ نشریه وب آو ساینس تالیف دانشمندان ایتالیایی انجام شد. میانگین دقت پیش‌بینی برای بازه‌های زمانی استناد بالای دو سال خوب است، برای انتشارات کم استناد کاهش می‌یابد و در رشته‌های مختلف متفاوت است. همان‌طور که انتظار می‌رفت، نقش ضریب تاثیر در ترکیب تنها پس از دو سال از انتشار ناچیز می شود (Gu & Krenn, 2024).

آلکوهر و همکاران نیز از یک چارچوب یادگیری ماشین برای آزمون توانایی معیارهای جایگزین (دگرسنجه) در پیش‌بینی تأثیر آتی مقالات براساس تعداد استنادها استفاده کردند. مبنای این بررسی، ۷۵۸۸ مقاله از ۱۰ مجله علوم رایانه بود. برای ساخت فضای ویژه برای پیش‌بینی، ۱۴ شاخص دگرسنجی مختلف جمع‌آوری و سه رویکرد انتخاب ویژگی، یعنی آستانه واریانس، همبستگی پیرسون و روش اطلاعات متقابل به‌منظور به حداقل رساندن فضای ویژگی و رتبه‌بندی ویژگی‌ها براساس سهم آنها استفاده شد. برای شناسایی عملکرد طبقه‌بندی این ویژگی‌ها، از سه الگو استفاده شد: درخت تصمیم، جنگل تصادفی و ماشین‌های بردار پشتیبانی. با توجه به داده‌های آزمایشی و براساس عملکرد، جنگل تصادفی نسبت به دیگر طبقه‌بندی‌کننده‌ها، دگرسنجه‌ها می‌تواند استنادهای آینده را پیش‌بینی کند و مفیدترین نشانه‌های دگرسنجی، تعداد رسانه‌های اجتماعی، توییت‌ها، شمارش اخبار، تعداد عکس‌برداری و نمای کامل متن هستند (Alchokr et al. 2023).

ایوب و همکاران در مطالعه خود دو مقوله استناد و نمره دگرسنجی را به‌منظور تعیین میزان مقایسه و ارتباط آنها با یکدیگر بررسی کردند. نمرات دگرسنجی با تعداد استناد مقالات مرتبط با رشته علوم زیستی، زمین‌شناسی و محیط‌زیست، تاریخ و باستان‌شناسی و مطالعات اجتماعی از لیست

۱۰۰ مقاله برتر در پایگاه آلتمتریک در دوره زمانی ۲۰۱۴ تا ۲۰۱۷ مقایسه شد. علاوه بر این، نمرات دگرسنجی مقالات با اس.جی.آر. مجلات مقایسه شد. درنهایت، همبستگی اسپیرمن برای سنجش ارتباط بین متغیرها محاسبه شد. این مطالعه نشان داد که زمین‌شناسی و محیط‌زیست بیشترین همبستگی را بین تعداد استناد و نمره دگرسنجی نشان داد، درحالی‌که برای نمره دگرسنجی و اس.جی.آر. در زمینه مذکور، چنین روندی قابل مشاهده نیست. تاریخ و باستان‌شناسی، همبستگی قوی بین دگرسنجی و نمرات استناد را به استثنای مقالات سال ۲۰۱۶ نشان می‌دهد و تا حدودی روند مشابهی برای امتیاز دگرسنجی و اس.جی.آر. انتشارات مشاهده شد. علوم زیستی همبستگی ضعیفی در هر دو متغیر نشان می‌دهند، درحالی‌که حوزه مطالعات اجتماعی عمدتاً برای هر دو مجموعه متغیر، ارتباط منفی نشان می‌دهند. بنابراین، بجز مقوله مطالعات اجتماعی، سه دسته دیگر عمدتاً همبستگی مثبتی بین دگرسنجی و امتیاز استناد نشریات و همچنین، تا حدی، برای نمره دگرسنجی و اس.جی.آر وجود دارد (Ayoub et al. 2023).

مطالعه آلوهالی و همکاران با هدف بررسی عوامل و بخش‌های مؤثر بر تعداد استناد یک مقاله علمی در زمینه شنوایی‌سنجی انجام شد. این مطالعه، راه‌حل جدیدی را پیشنهاد داده و از یادگیری ماشین و پردازش زبان طبیعی برای پردازش متن انگلیسی استفاده نمود. الگوریتم‌های مختلفی مانند رگرسیون خطی، درخت تصمیم تقویت شده، جنگل تصمیم و شبکه‌های عصبی به این منظور پیاده‌سازی شد. استفاده از رگرسیون شبکه عصبی نشان داد که چکیده مقالات، تأثیر بیشتری بر تعداد استناد مقالات شنوایی‌سنجی دارد. براین اساس، توصیه کردند که از مدل‌های یادگیری ماشینی برای بهبود چکیده مقالات پژوهشی برای دریافت استنادهای بیشتر استفاده شود (Alohali et al. 2022).

در یک پایان‌نامه نیز داده‌های اسکوپوس با هدف تحلیل عملکرد متغیرهای مرتبط با مقاله، مجله و نویسنده به‌عنوان شاخص‌های تأثیر مقالات حوزه بازاریابی و اینکه چگونه می‌توانند مقالات پراستناد را در پنج سال آینده پیش‌بینی کنند، جمع‌آوری شد. برای یافتن معیارهای مرکزیت از تحلیل شبکه اجتماعی استفاده شد و تعداد استنادها یک سال پس از انتشار به‌عنوان تنها متغیر وابسته به زمان گنجانده شد. نتایج نشان داد که استنادهای یک سال بعد، محرک قوی و پیش‌بینی کننده برای استنادهای آینده پس از پنج سال است. تحلیل شبکه هم‌نویسندگی نشان داد که مرکزیت نزدیکی و بینیت نیز محرک استنادهای آینده در حوزه بازاریابی هستند که نشان می‌دهد نزدیک بودن به هسته شبکه و داشتن قدرت واسطه‌ای در این زمینه، مهم است. محققان با استفاده از روش‌های یادگیری ماشین، متوجه شدند که ترکیبی از مؤلفه‌های مرتبط با مقاله، مجله و نویسنده در مقایسه با تنها یک نوع مؤلفه در پیش‌بینی مقالات پراستناد بهتر عمل می‌کنند (Hansen & Torvund, 2022).

در مقاله ژانگ و وو، چارچوب جدید WMR-Rank را برای برای پیش بینی تأثیر آینده مقالات، محققان و مکان ها پیشنهاد می‌کنند. بر اساس شبکه پویا و ناهمگون موجودیت های متعدد، هفت نوع رابطه را از بین آنها استخراج شد. این چارچوب از ویژگی‌های مفیدی از جمله جزئیات دقیق نهادهای مرتبط مانند نویسندگان و مکان‌های برگزاری، آگاهی از زمان برای مقالات منتشر شده و استنادات آن‌ها، متمایز کردن مشارکت چند نویسنده همکار در یک مقاله و غیره پشتیبانی می‌کند. نتایج تجربی نشان داد که رویکرد پیشنهادی به طور قابل توجهی بهتر از رقبای پیشرفته عمل می‌کند (Zhang & Wu, 2020).

میهالا و همکارانش در مطالعه خود ۱۲۶ مقاله در مورد کووید ۱۹ در وب‌آوساینس را برای تعیین اینکه آیا دگرسنجی با تأثیر علمی مقاله مرتبط است یا خیر، تحلیل کردند. روابط و سنجه‌های سنتی و دگرسنجی ازطریق تحلیل مؤلفه‌های اصلی، تحلیل عاملی و همبستگی‌ها بررسی شدند. نتایج نشان داد که تعداد مندلی (خوانده‌ها) استفاده بالقوه آنها برای پیش‌بینی استنادات وب‌آوساینس را نشان می‌دهد، امتیاز توجه دگرسنجی، همبستگی نسبتاً کمی با استنادهای آتی نشان می‌دهد؛ درحالی‌که گوگل اسکولار که با احتیاط استفاده می‌شود، می‌تواند پیش‌بینی‌کننده تعداد استنادهای آینده باشد (Mihaela et al. 2020).

ابریشمی و علی اکبری در مطالعه خود، روش جدیدی را به‌منظور پیش‌بینی استنادهای بلندمدت یک مقاله براساس تعداد استنادهای آن در چند سال اول پس از انتشار، پیشنهاد کردند. آنها به‌منظور آزمون مدل پیش‌بینی تعداد استناد از شبکه عصبی مصنوعی استفاده کردند که یک ابزار یادگیری ماشینی قدرتمند با کاربردهای جدید در بسیاری از حوزه‌ها ازجمله پردازش تصویر و متن است. آنها از تکنیک‌های مدرن یادگیری عمیق (مانند شبکه عصبی بازگشتی و مدل دنباله به دنباله) به‌منظور یادگیری روش پیش‌بینی براساس الگوی توالی استنادها از سال‌های اولیه انتشار مقاله استفاده کردند. ارزیابی‌های جامع آنها نشان داد که روش پیشنهادی با توجه به دقت و توانایی پیش‌بینی استنادات مقالات پراستناد، بهتر از روش‌های پیشرفته پیش‌بینی تعداد استناد است (Abrishami & Aliakbary, 2019).

پژوهشگر دیگری، قدرت نسبی تعداد استنادهای کوتاه مدت، معیارهای کتاب‌سنجی مانند عوامل تأثیر مجلات و رتبه‌بندی مجلات را از نظر پیش‌بینی استنادهای بلندمدت بررسی کرد. وی با استفاده از مجموعه‌ای از مقالات منتشر شده در شصت مجله اقتصادی متوجه شد که استنادهای دریافتی در بازه‌های کوتاه مدت (بین ۱ تا ۲ سال پس از انتشار) در مقایسه با مواردی مانند تأثیر مجلات، پیش‌بینی‌کننده‌های قوی‌تری برای تعداد استنادات بلندمدت یا سایر رتبه‌بندی‌های مجلات هستند. این مقاله بر مقایسه سهم نسبی استنادهای کوتاه مدت، معیارهای کیفیت مجلات مانند عوامل تأثیر مجله و امتیازات رتبه‌بندی مجله و ویژگی‌های مقاله تأکید داشت. از دید این پژوهشگر، این واقعیت که تعداد استنادهای کوتاه‌مدت نسبت به متغیرهای شاخص مجله، پیش‌بینی‌کننده بهتری برای استنادهای بلندمدت ارائه می‌کند، یک نتیجه واقعاً منحصربه‌فرد است (Kosteas, 2018).

تیمیلسینا و همکاران (۲۰۱۶) بر اساس یک شبکه علمی ناهمگون و با ادغام داده‌های کتاب‌سنجی با داده‌های اجتماعی مانند وبلاگ‌ها و اخبار اصلی، تأثیر علمی را پیش‌بینی کرد، که نشان می‌دهد اندازه‌گیری مبتنی بر نمودار می‌تواند به طور منطقی تأثیر محققان مرحله اولیه را پیش‌بینی کند (Timilsina et al. 2016).

برخی محققان نیز مدلی را با هدف پیش‌بینی تعداد استنادهای آینده نشریه پیشنهاد کردند که از دو پیش‌بینی کننده استفاده شد: ضریب تأثیر مجله و تعداد استنادهایی که یک نشریه یک سال پس از انتشار دریافت کرده است. مدل پیشنهادی، مبتنی بر رگرسیون کمّی بود. آنها از این مدل به منظور پیش‌بینی تعداد استنادات در حوزه فیزیک استفاده کردند. تحلیل‌ها نشان داد که هر دو پیش‌بینی‌کننده (یعنی ضریب تأثیر و استنادهای اولیه) به پیش‌بینی دقیق تأثیر استناد درازمدت کمک می‌کنند (Stegehuis et al. 2015).

یو و همکاران (۲۰۱۴) از چهار دسته از ویژگی‌ها، از جمله مقالات، نویسندگان، استنادها و مجلات برای پیش‌بینی استنادهای آینده یک مقاله بر اساس تحلیل رگرسیون گام به گام استفاده کردند. بر اساس شبکه‌های هم‌نویسندگی، یک طبقه‌بندی یادگیری ماشین برای پیش‌بینی اینکه آیا یک نشریه استنادهای بالایی دریافت می‌کند یا خیر، توسعه داده شد (Yu et al. 2014).

یکی دیگر از پژوهش‌ها نیز روش متفاوتی را به‌منظور پیش‌بینی تأثیرات آتی هر مقاله ارائه کرد. ویژگی‌های پیش‌بینی‌کننده براساس موقعیت مقاله در شبکه استناد، با استناد و ارزیابی پژوهش‌های قبلی در مورد انتشار اطلاعات در شبکه‌ها استفاده شد. این روش با داده‌های اسکوپوس از سال ۱۹۹۶–۲۰۱۰، اجرا و ارزیابی شد. تعداد استنادهای مقاله در سال‌های اولیه پس از انتشار، تعداد استنادهای مقاله، میانگین تعداد استنادهای مقالات استناد کننده و استناد شونده و استنادهای بین‌رشته‌ای بیشتر از مقاله و مقالات استناد شده، همبستگی مثبت داشتند. سه الگوریتم رگرسیون خطی، درخت‌های تصمیم‌گیری و جنگل تصادفی برای پیش‌بینی تأثیر آتی مقالات آزمایش شد. نتایج نشان داد که جنگل تصادفی پیش‌بینی‌کننده‌ترین الگوریتم بود، به‌طوری که پیش‌بینی ۱۸ درصدی از سال انتشار مقاله و ۳۴ درصدی پیش‌بینی استناد از دو سال پس از آن به دست آمد (McNamara et al. 2013).

آکلا و همکارانش در پژوهش خود از دگرسنجه‌ها به‌منظور پیش‌بینی استنادات کوتاه‌مدت و بلندمدت مقالات استفاده کردند. این پژوهشگران، مدل‌های طبقه‌بندی و رگرسیون مختلفی را پیاده کرده و تشخیص دادند که شبکه‌های عصبی و مدل‌های مجموعه دارای بهترین عملکرد برای این امر هستند. ایشان دریافتند که خوانندگان مندلی مهم‌ترین عامل در پیش‌بینی استنادهای اولیه بوده و به‌دنبال آن، عوامل دیگری مانند وضعیت علمی خوانندگان (به‌عنوان مثال، دانشجو، فوق دکترا، استاد)، دنبال‌کنندگان در توییتر، طول پست آنلاین، تعداد نویسنده و تعداد ذکر شدن در توییتر، ویکی‌پدیا و در کشورهای مختلف (Akella et al. 2009).

بولان و همکاران (۲۰۰۹) تحلیل مؤلفه اصلی رتبه‌های تولید شده توسط ۳۹ معیار تأثیر علمی موجود و پیشنهادی که براساس داده‌های استناد و گزارش استفاده، محاسبه شده، انجام دادند. نتایج نشان داد که تأثیر علمی یک سازه چندبعدی است که نمی‌توان آن را به‌طور دقیق با هیچ شاخصی اندازه‌گیری کرد، اگرچه برخی از معیارها مناسب‌تر از سایرین هستند. ضریب تأثیر استنادی که معمولاً مورد استفاده است در هسته این ساختار جای نگرفته، بلکه در حاشیه آن قرار دارد و بنابراین باید با احتیاط استفاده شود. تحلیل مؤلفه‌های اصلی نشان داد که ۹۲ درصد از واریانس‌های بین همبستگی‌های رتبه‌بندی مجلات ایجاد شده توسط ۳۷ معیار تأثیر را می‌توان با ۳ مؤلفه اول توضیح داد (Bollen et al. 2009).

از بررسی پیشینه‌ها این نتیجه حاصل می‌شود که پژوهشگران تلاش کرده‌اند با رویکردها و شاخص‌های مختلف، اثرگذاری خروجی‌های علمی را پیش‌بینی کنند. این مطالعات گاهی بر مقالات، متمرکز شده و گاهی نیز مجلات را هدف قرار داده‌اند. عمده تلاش‌ها نیز بر تشخیص رابطه

شاخص‌های استنادی سنتی و دگرسنجه‌ها و توان پیش‌بینی آنها از یکدیگر بوده است. حوزه‌های موضوعی متفاوتی نیز مورد مطالعه قرار گرفته است. پژوهش حاضر نیز تلاش کرد تأثیر شاخص‌های مختلف اثرگذاری بروندادهای علمی حوزه زیست‌فناوری ایران در یک بازه ۲۲ ساله را بر هم بررسی نماید و از الگوریتم‌های متفاوتی برای پیش‌بینی تأثیرات آنها استفاده کند.

روش‌شناسی پژوهش

پژوهش حاضر از لحاظ هدف، کاربردی و از نظر روش، توصیفی بوده و با رویکرد علم‌سنجی انجام شده است و به‌دنبال بررسی همکاری علمی و نیز تأثیرگذاری علمی، اجتماعی و اقتصادی آثار علمی ایران در حوزه زیست‌فناوری است و اینکه کدام‌یک از الگوریتم‌های یادگیری ماشین، پیش‌بینی‌کننده بهتری برای سنجش اثرگذاری بروندادهای علمی در ابعاد مختلف هستند. جامعه پژوهش، بروندادهای حوزه زیست‌فناوری ایران است که در پایگاه اسکوپوس در بازه ۲۰۰۳–۲۰۲۴ نمایه شده‌اند. تاریخ استخراج داده‌ها نیز ۱ بهمن ۱۴۰۲ معادل ۲۱ ژانویه ۲۰۲۴ بود. به‌منظور استخراج داده‌های مربوط از پایگاه تحلیلی سایول استفاده شد. پایگاه سایول که برای اولین بار در سال ۲۰۰۹ معرفی شد، امکان تجزیه و تحلیل‌های استنادی پیشرفته و عمیق را بر مبنای داده‌های پایگاه اسکوپوس فراهم می‌سازد. درواقع، پایگاه استنادی سایول یک نمایه استنادی محسوب نمی‌شود، بلکه یک پایگاه استنادی تجزیه و تحلیل پژوهشی است که همچون این سایتس[1]، امکان مقایسه نتایج بدست آمده را با میانگین‌ها و استانداردهای جهانی فراهم می‌کند. به این ترتیب، با استفاده از سایول می‌توان عملکرد پژوهشی یک پژوهشگر، سازمان، کشور، منطقه، مجله، گروه پژوهشی و مانند آنها را ارزیابی کرد (نوروزی چاکلی، ۱۴۰۲). متغیرهای مورد مطالعه در قالب مؤلفه و شاخص‌های مورد استفاده برای آزمون همبستگی و نیز پیش‌بینی کننده تأثیرات علمی، اقتصادی و اجتماعی به شرح زیر بودند.

جدول ۱. متغیرهای مورد مطالعه

عنوان متغیر
حجم برونداد علمی
همکاری فقط دانشگاهی (٪)
تأثیر همکاری فقط دانشگاهی
تأثیر همکاری بین‌المللی
همکاری دانشگاهی و دولتی (٪)
حجم استنادها
استناد به ازای برونداد
تأثیر استناد وزنی حوزه
بروندادها در ۱۰٪ درصد برتر مجله توسط درصد CiteScore (٪)
بروندادها در ۱۰٪ درصد استنادهای برتر (٪)
تعداد استنادات ثبت اختراع
استنادهای حق ثبت اختراع به ازای هر خروجی علمی
خروجی علمی استناد شده توسط ثبت اختراعات (٪)
تعداد اختراعات استنادکننده به مقالات
تعداد بازدیدها
برونداد در ۱۰٪ درصد بازدیدهای برتر (٪)
بازدید به ازای هر خروجی
تأثیر بازدید وزنی حوزه

[1] . *InCites*

جدول ۲. مؤلفه‌ها و شاخص‌های مرتبط به آنها به‌عنوان پیش‌بینی کننده تأثیر

مؤلفه	عنوان شاخص
تأثیر علمی	حجم استنادها
	استناد به ازای برونداد
	تأثیر استناد وزنی حوزه
	برونداد‌ها در ۱۰٪ درصد برتر مجله توسط درصد CiteScore (٪)
	برونداد‌ها در ۱۰٪ درصد استنادهای برتر (٪)
تأثیر اقتصادی	تعداد استنادات ثبت اختراع
	استنادهای حق ثبت اختراع به ازای هر خروجی علمی
	خروجی علمی استناد شده توسط ثبت اختراعات (٪)
	تعداد اختراعات
تأثیر اجتماعی	تعداد بازدیدها
	برونداد در ۱۰٪ درصد بازدیدهای برتر (٪)
	بازدید به ازای هر خروجی
	تأثیر بازدید وزنی حوزه

شایان ذکر است که منظور از تأثیر اجتماعی، اثری است که خروجی فعالیت‌های علمی در رسانه‌ها و شبکه‌های اجتماعی ایجاد می‌کنند. تأثیر اقتصادی نیز منظور تعاملاتی است که بین تولیدات علمی و پروانه‌های ثبت اختراع شکل می‌گیرد.

در این پژوهش از ضریب همبستگی پیرسون و از بسته نرم‌افزاری R به‌منظور تعیین رابطه بین شاخص‌های مورد مطالعه، استفاده شده است. در زمان بررسی همبستگی بین دو متغیر، اگر هر دو متغیر مورد نظر در مقیاس نسبی و فاصله‌ای باشند، از ضریب همبستگی پیرسون استفاده می‌شود. الگوریتم‌های یادگیری ماشین مانند رگرسیون خطی چندگانه، درخت‌های تصمیم‌گیری، جنگل‌های تصادفی و تقویت گرادیان نیز به‌عنوان مدل‌های پیش‌بینی‌کننده، مورد استفاده و ارزیابی قرار گرفتند. به‌منظور انجام آزمون‌ها و الگوریتم‌ها از زبان برنامه‌نویسی پایتون استفاده شده است. تحلیل‌ها توسط زبان برنامه‌نویسی پایتون انجام شده که برای بررسی و نمایش صحیح آن بایستی از نرم‌افزار jupyter notebook استفاده کرد. داده‌ها برای استفاده توسط الگوریتم‌های یادگیری ماشین و انجام تحلیل‌های مربوطه نیازمند اصلاحات ساختاری بود که این اصلاحات انجام شد. تعداد اندکی از داده‌ها در دسترس نبود و مقداری برای آن گزارش نشده بود که برای انجام تحلیل‌ها بایستی مقادیری به آن ها نسبت داده می‌شد. این کار با استفاده از متد قابل قبول و معتبر knn Imputer انجام شد که این متد با در نظر گرفتن نزدیک‌ترین همسایگی‌ها اقدام به تکمیل داده‌ها می‌کند. در ضمن، فونت داده‌های فارسی، قابل استفاده نبوده و تحلیل‌ها با استفاده از داده‌های انگلیسی انجام شد. در گام بعدی، مدل‌های یادگیری ماشین مختلفی بر روی داده‌ها مدل‌سازی و اجرا شد. مدل‌سازی در نرم‌افزار پایتون به گونه‌ای انجام شده که همه متغیرهای موجود در مجموعه داده، یک بار به‌عنوان مقدار هدف فرض شده‌اند و مدلی ساخته شده که بتواند براساس سایر متغیرها این مقدار را پیش‌بینی کند. برای ارزیابی مدل‌ها نیز از معیار صحت[1] استفاده شد که اساسی‌ترین معیار اندازه‌گیری کیفیت یک دسته بند است. درک آن بسیار ساده است و به راحتی برای مشکل طبقه‌بندی دوحالته و چندحالته مناسب است.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

صحت یعنی نسبت نتایج واقعی به کل موارد بررسی شده. نمره صحت مدل در واقع معرف صحت مدل است که اگر این اعداد در ۱۰۰ضرب شود بیانگر درصد صحت و دقت مدل است. نکته آنکه اعداد منفی، بیانگر دقت بسیار پایین بوده و نشان می‌دهند مدل به‌کار رفته برای پیش‌بینی آن پارامتر خاص، اصلاً مناسب نیست.

[1] . Accuracy

یافته‌های پژوهش

پاسخ به پرسش اول پژوهش: وضعیت بروندادهای حوزه زیست‌فناوری ایران براساس هر کدام یک از شاخص‌های اثرگذاری علمی، اثرگذاری اجتماعی و اثرگذاری اقتصادی چگونه است؟

وضعیت بروندادهای علمی ایران در حوزه زیست‌فناوری براساس داده‌های پایگاه سایول در بازه زمانی ۲۰۰۳ تا ۲۰۲۴ که مبنای تحلیل‌های بعدی قرار گرفت در جدول شماره ۳ آمده است. نکته مهم در این خصوص، سیر صعودی و افزایش ۳۶ برابری حجم بروندادهای علمی ایران در این حوزه تا سال ۲۰۲۳ است که نشان از پیشرفت فوق‌العاده بالا در این فاصله زمانی است. ضمن اینکه انتظار می‌رود کیفیت مقالات نیز همراستا با رشد کمّی آنها رشد کند.

زودآیند ویرایش نشده

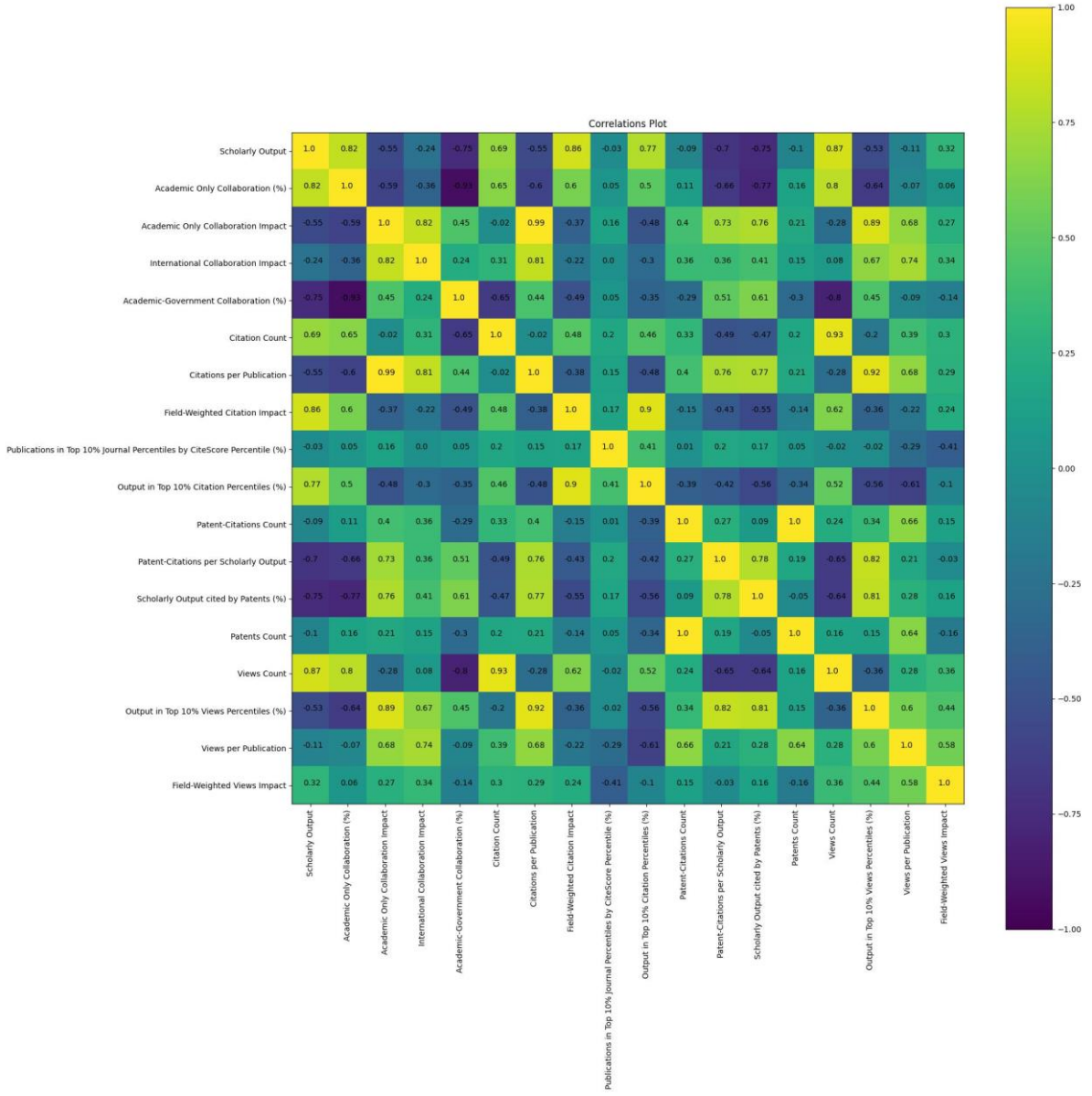
جدول ۳. وضعیت بروندادهای علمی ایران در حوزه زیست‌فناوری براساس داده‌های پایگاه سایول در بازه ۲۰۰۳ تا ۲۰۲۴

	عنوان شاخص	۲۰۰۳	۲۰۰۴	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸	۲۰۰۹	۲۰۱۰	۲۰۱۱	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۵	۲۰۱۶	۲۰۱۷	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۲۱	۲۰۲۲	۲۰۲۳	۲۰۲۴
برونداد علمی	حجم برونداد علمی	۳۵	۴۵	۶۸	۱۰۳	۱۴۷	۲۸۴	۳۵۹	۵۸۹	۷۴۷	۴۶۴	۵۴۶	۷۳۳	۷۷۶	۹۶۵	۹۵۳	۱۱۲۹	۱۱۳۳	۱۱۶۹	۱۱۹۰	۱۲۲۹	۱۲۶۲	۷۴
همکاری	همکاری فقط دانشگاهی (٪)	۷۱،۴	۶۴،۴	۶۰،۳	۷۶،۷	۷۰،۱	۷۷،۱	۷۵،۸	۷۹،۳	۸۲،۵	۷۶،۵	۷۹،۳	۸۰،۹	۸۳،۵	۸۲،۷	۸۲،۶	۸۲،۶	۸۲،۸	۸۴،۵	۸۲،۷	۸۳،۲	۸۳،۲	۷۷
	تأثیر همکاری فقط دانشگاهی	۳۱،۸	۴۰،۹	۳۴،۲	۳۹،۶	۳۷،۱	۲۷،۶	۲۰،۹	۲۲	۱۹،۱	۳۰،۷	۲۸،۲	۲۰،۷	۱۹	۲۶،۳	۲۷	۲۵،۹	۲۲،۸	۱۸،۴	۱۱،۶	۶،۷	۲	۰،۱
	تأثیر همکاری بین‌المللی	۲۶،۵	۵۳،۱	۴۶،۵	۴۱،۴	۳۴،۵	۳۰،۵	۲۶،۹	۳۲،۷	۲۹،۶	۶۴،۵	۲۶،۸	۳۳	۲۹،۶	۴۵،۱	۴۲،۵	۴۲،۴	۳۴،۹	۲۹،۶	۲۰،۵	۱۰،۸	۳،۱	۰،۲
	همکاری دانشگاهی و دولتی (٪)	۱۴،۳	۲۰	۱۹،۱	۱۲،۶	۱۵،۶	۱۲،۳	۱۵،۳	۹،۸	۸	۱۳،۶	۱۲،۵	۱۰	۹،۷	۹،۲	۱۰،۴	۱۰،۵	۱۰،۷	۱۰،۱	۱۱،۵	۱۰،۲	۱۰،۷	۱۴،۹
تأثیر علمی	حجم استنادها	۱۱۱۱	۱۶۸۶	۲۴۶۱	۴۱۷۱	۵۱۷۲	۷۴۵۸	۷۸۷۲	۱۳۱۶ ۷	۱۴۹۹ ۶	۱۳۲۶ ۹	۱۴۲۶ ۶	۱۴۷۴ ۱	۱۴۸۷ ۸	۲۵۱۱ ۲	۲۴۵۴ ۱	۲۹۸۱ ۵	۲۵۲۷ ۳	۲۱۷۹ ۰	۱۴۷۰ ۵	۸۴۸۱	۲۵۵۶	۸
	استناد به ازای برونداد	۳۱،۷	۳۷،۵	۳۶،۲	۴۰،۵	۳۵،۲	۲۶،۳	۲۱،۹	۲۲،۴	۲۰،۱	۲۸،۶	۲۶،۱	۲۰،۱	۱۹،۲	۲۶	۲۵،۸	۲۶،۴	۲۲،۳	۱۸،۶	۱۲،۴	۶،۹	۲	۰،۱
	تأثیر استناد وزنی حوزه	۰،۵۵	۰،۵۹	۰،۵۸	۰،۷۳	۰،۷۴	۰،۶۳	۰،۵۵	۰،۵۸	۰،۵۵	۰،۷۲	۰،۹	۰،۸۸	۰،۷۲	۰،۹۴	۰،۹۳	۱،۱۱	۱،۰۴	۱،۰۷	۱،۱	۱،۱۸	۱،۳۴	۰،۴۴
	برونداد‌ها در ۱۰٪ درصد برتر مجله CiteScore	۱۲،۱	۷،۷	۴،۶	۱۱،۱	۸،۱	۶	۵،۴	۲،۷	۲،۷	۴،۲	۸	۶،۳	۵،۱	۱۲،۲	۹،۵	۱۱،۸	۹	۹،۲	۸،۱	۴،۹	۵،۸	۱۱
	برونداد‌ها در ۱۰٪ درصد استنادهای برتر (٪)	۱۱،۴	۸،۹	۱۳،۲	۱۳،۶	۱۴،۳	۱۱،۶	۸،۶	۸	۸،۸	۱۳،۱	۱۲،۸	۱۱،۳	۱۲	۱۷،۷	۱۸،۳	۲۱،۲	۲۱،۶	۲۲،۲	۲۲،۴	۲۲،۹	۲۰،۵	-
تأثیر اقتصادی	تعداد استنادات ثبت اختراع	۳۳	۳۳	۶۹	۱۷۰	۹۴	۹۰	۷۷	۱۹۵	۱۴۸	۷۴	۲۵۵	۱۵۳	۶۱	۲۵۰	۱۱۲	۷۹	۴۲	۳۶	۲۵	۳	۰	۰
	استنادهای حق ثبت اختراع به ازای هر خروجی	۹۴۲،۹	۷۳۳،۳	۱۰۱۴،۷	۱۶۵۰،۵	۶۳۹،۵	۳۱۶،۹	۲۱۴،۵	۳۳۱،۰	۱۹۸،۱	۱۵۹،۵	۴۶۷	۲۰۸،۷	۷۸،۶	۲۵۹،۱	۱۱۷،۵	۷۰	۳۷،۱	۳۰،۸	۲۱	۲،۴	۰	۰
	خروجی علمی استناد شده توسط ثبت اختراعات	۳۱،۴	۲۰	۱۷،۶	۱۴،۶	۱۸،۴	۱۰،۶	۹،۲	۸،۱	۵،۸	۸	۸،۲	۶،۷	۴،۸	۶،۷	۶،۱	۴،۷	۲،۶	۱،۹	۱،۸	۰،۲	۰	۰
	تعداد اختراعات	۳۳	۳۳	۶۹	۱۷۰	۹۴	۹۰	۷۷	۱۹۵	۱۴۸	۷۴	۲۵۴	۱۵۳	۶۱	۲۴۸	۱۱۰	۷۷	۴۲	۳۶	۲۵	۳	-	-
تأثیر اجتماعی	تعداد بازدیدها	۱۷۳۶	۲۸۰۳	۴۰۱۰	۶۴۵۷	۱۰۰۲ ۳	۱۹۶۷ ۷	۲۵۸۷ ۴	۴۴۸۱ ۶	۵۸۰۵ ۶	۲۸۱۷ ۰	۴۳۲۸ ۴	۴۷۴۵ ۹	۴۹۵۷ ۴	۶۶۴۰ ۳	۶۳۶۰ ۲	۷۵۶۶ ۵	۷۰۵۵ ۲	۵۷۲۱ ۳	۵۱۳۰ ۸	۴۷۰۶ ۴	۳۵۲۷ ۶	۵۳۲
	برونداد در ۱۰٪ درصد بازدیدهای برتر (٪)	۶۵،۷	۶۴،۴	۷۲،۱	۷۰،۹	۵۵،۸	۵۳،۵	۵۱،۳	۵۱،۴	۵۰،۱	۵۱،۳	۵۱،۳	۴۰،۵	۳۹،۶	۴۸،۹	۴۸،۱	۴۷،۳	۴۲،۴	۳۷،۶	۳۷،۶	۳۶،۱	۳۳،۹	۱۳،۵
	بازدید به ازای هر خروجی	۴۹،۶	۶۲،۳	۵۹	۶۲،۷	۶۸،۲	۶۹،۳	۷۲،۱	۷۶،۱	۷۷،۷	۸۲،۳	۷۹،۳	۶۴،۷	۶۳،۹	۶۸،۸	۶۶،۷	۶۷	۶۲،۳	۴۸،۹	۴۳،۱	۳۸،۳	۲۸	۷،۲
	تأثیر بازدید وزنی حوزه	-	۱،۷۴	۱،۵۱	۱،۳۱	۱،۵۱	۱،۶۴	۲،۱۸	۲،۲۷	۱،۶	۱،۶۶	۱،۴۷	۱،۳۸	۱،۳۳	۱،۵۳	۱،۵۶	۱،۷۲	۱،۸۱	۱،۷	۱،۷۷	۱،۶۷	۱،۶۹	۰

پژوهشگاه علوم و فناوری اطلاعات ایران

پاسخ به پرسش دوم پژوهش: آیا بین تعداد بروندادهای و همکاری‌های علمی با شاخص‌های مربوط به اثرگذاری، اجتماعی و اقتصادی رابطه وجود دارد؟

نمودار زیر، میزان همبستگی بین حجم بروندادهای علمی، همکاری، تأثیر علمی، اقتصادی و اجتماعی آنها را به تصویر می‌کشد.



نمودار ۱. نتایج آزمون همبستگی بین حجم بروندادهای علمی، همکاری، تأثیر علمی، تأثیر اقتصادی و تأثیر اجتماعی آنها

داده‌های نمودار شماره ۱ روشن می‌کند که هر کدام از متغیرها با متغیرهای دیگر چه رابطه‌ای با هم دارند. عدد گزارش شده هرچقدر به ۱+ و یا ۱− نزدیک‌تر باشد نشان می‌دهد رابطه بیشتری بین دو متغیر وجود دارد و هر چقدر به صفر نزدیک‌تر باشد نشان می‌دهد که آن دو متغیر رابطه خطی با یکدیگر ندارند. اعداد مثبت نشان می‌دهند که اگر یکی از متغیرها افزایش یابد متغیر دیگری نیز افزایش پیدا می‌کند. اعداد منفی نشان می‌دهند که اگر یکی از متغیرها افزایش پیدا کند متغیر دیگر کاهش می‌یابد. نکته قابل ذکر و بدیهی آن است که هر متغیر با خودش همبستگی کامل یعنی + ۱ دارد. به همین دلیل در قطر ماتریس، تمام اعداد ۱+ هستند.

بر این اساس، حجم بروندادهای علمی ایران در حوزه زیست‌فناوری با حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، بروندادها در ۱۰٪ درصد استنادهای برتر، تعداد بازدیدها و تأثیر بازدید وزنی حوزه رابطه مثبت و معنی‌داری داشته است.

همکاری بین‌المللی با حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰٪ درصد برتر مجلات، تعداد استنادات ثبت اختراع، استنادهای حق ثبت اختراع به ازای هر خروجی علمی، خروجی علمی استناد شده توسط ثبت اختراعات، تعداد اختراعات، تعداد بازدیدها، برونداد در ۱۰٪ درصد بازدیدهای برتر، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه رابطه مثبت و معنی‌داری داشته است. همکاری دولتی دانشگاهی نیز تنها بر سه شاخص استناد به ازای برونداد، تعداد استنادات ثبت اختراع و استنادهای حق ثبت اختراع به ازای هر خروجی علمی رابطه مثبت و معنی‌داری داشته است. همکاری فقط دانشگاهی بر حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰٪ درصد برتر مجلات، بروندادها در ۱۰٪ درصد استنادهای برتر، تعداد استنادات ثبت اختراع، خروجی علمی استناد شده توسط ثبت اختراعات، تعداد اختراعات، تعداد بازدیدها و تأثیر بازدید وزنی حوزه رابطه مثبت و معنی‌داری داشته است.

پاسخ به پرسش سوم پژوهش: کدام یک از شاخص‌های اثرگذاری علمی، اجتماعی و اقتصادی با همدیگر ارتباط دارند؟

در مقوله اثرگذاری علمی بروندادهای مورد مطالعه، حجم استنادها با شاخص‌های تعداد استنادات ثبت اختراع، تعداد اختراعات، تعداد بازدیدها، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری داشته است.

در حوزه اثرگذاری اقتصادی، تعداد استنادات ثبت اختراع با همکاری دانشگاهی، تأثیر همکاری بین‌المللی، حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، تعداد بازدیدها، برونداد در ۱۰٪ درصد بازدیدهای برتر، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری داشته است. استنادات ثبت اختراع به ازای هر برونداد با شاخص‌های استناد به ازای هر برونداد، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری داشته است. تعداد اختراعات با مواردی مانند حجم استنادها، استناد به ازای برونداد، تعداد بازدیدها، برونداد در ۱۰٪ درصد بازدیدهای برتر و بازدید به ازای هر خروجی، رابطه مثبت و معنی‌داری داشته است.

تعداد بازدیدها با شاخص‌هایی نظیر حجم استنادها، تأثیر استناد وزنی حوزه، تعداد استنادات ثبت اختراع، استنادهای حق ثبت اختراع به ازای هر خروجی علمی، خروجی علمی استناد شده توسط ثبت اختراعات و تعداد اختراعات، رابطه مثبت و معنی‌داری داشته است.

پاسخ به پرسش چهارم پژوهش: کدام‌یک از الگوریتم‌های ماشین می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی را پیش‌بینی کنند؟

نتیجه بررسی میزان پیش‌بینی‌کنندگی الگوریتم‌های مورد استفاده در اثرات علمی، فناورانه و اجتماعی بروندادهای علمی در جدول شماره ۴ آمده است.

جدول ۴. میزان پیش‌بینی‌کنندگی اثرات علمی، فناورانه و اجتماعی بروندادهای علمی در الگوریتم‌های مختلف

مدل	شاخص هدف	نمره صحت مدل	نمره انحراف معیار مدل
نزدیک‌ترین همسایه	اثر علمی	۰,۴۴	۰,۵۷
	اثر فناورانه	−۰,۰۷	۰,۹۹
	اثر اجتماعی	۰,۰۸	۰,۸۶
درخت تصمیم	اثر علمی	−۰,۸۸	۰,۲۴
	اثر فناورانه	۰,۴۴	۰,۲۶
	اثر اجتماعی	−۰,۶۶	۱,۷۸
جنگل تصادفی	اثر علمی	−۰,۴۱	۰,۸۶
	اثر فناورانه	−۵۱,۳۰	۶۵,۹۶
	اثر اجتماعی	۰,۶۸	۰,۱۵
	اثر علمی	۰,۷۸	۰,۱۳

زودآیند ویرایش نشده

مدل	شاخص هدف	نمره صحت مدل	نمره انحراف معیار مدل
رگرسیون خطی چندمتغیره	اثر فناورانه	۰٫۵۲	۰٫۲۸
	اثر اجتماعی	۰٫۶۸	۰٫۱۵
تقویت گرادیان	اثر علمی	-۰٫۷۰	۰٫۹۶
	اثر فناورانه	۰٫۹۳	۰٫۰۴
	اثر اجتماعی	۰٫۵۱	۰٫۵۹

در جدول شماره ۴، شاخص هدف، نمره صحت و نمره انحراف معیار ۴ مدل پیاده شده در این مطالعه، درج شده است. لازم به توضیح است که هر چقدر نمره صحت اجرای مدل بالاتر و نمره انحراف معیار حاصل شده پایین‌تر باشد، قدرت پیش‌بینی‌کنندگی الگوی پیاده شده بیشتر خواهد بود. اعداد منفی، بیانگر دقت بسیار پایین بوده و نشان می‌دهد مدل به کار رفته به منظور پیش‌بینی یک متغیر خاص، مناسب نیست. براساس نتایج حاصل شده، رگرسیون خطی چندمتغیره با نمره صحت بالاتر و نمره انحراف معیار پایین‌تر، بهتر توانسته است میزان اثرگذاری علمی، فناورانه و اجتماعی برونداهای علمی ایران در حوزه زیست‌فناوری را پیش‌بینی نماید.

بحث و نتیجه‌گیری

سنجش کمیت و کیفیت فعالیت‌ها و خروجی‌های علمی از زوایای گوناگون از اهمیت بسیار زیادی برخوردار است. با توجه به اهمیت موضوع، این مطالعه با هدف بررسی رابطه هر یک از متغیرهای مختلف اثرگذاری بر همدیگر و نیز ارزیابی الگوریتم‌های ماشین که بتوانند اثرگذاری علمی، اجتماعی و اقتصادی برونداهای علمی را پیش‌بینی کنند، انجام گرفت. به این منظور، برونداهای حوزه زیست‌فناوری نمایه شده ایران در پایگاه اسکوپوس در دوره ۲۰۰۳-۲۰۲۴ از ابعاد مختلف، ارزیابی و تحلیل شد.

یافته‌های این مطالعه نشان داد که برونداهای علمی ایران در این حوزه در بازه زمانی ۲۰۰۳ تا ۲۰۲۳، سیر صعودی داشته و افزایش ۳۶ برابری را تجربه کرده است که پیشرفت فوق‌العاده بالایی در این فاصله زمانی محسوب می‌شود. شکل‌گیری ستاد توسعه زیست‌فناوری در ایران و تصویب سند مربوط به این حوزه و درنتیجه افزایش حمایت‌ها از حوزه زیست‌فناوری می‌تواند ازجمله دلایل این رشد باشد. ضمن اینکه رشد کیفیت مقالات نیز مورد انتظار است. اما ازآنجاکه شاخص‌های موجود علم‌سنجی عمدتاً مبتنی بر استناد بوده و لازمه اثرگذاری در ابعاد مختلف، گذشت زمان بوده، قضاوت در مورد آنها نیازمند توجه به بعد زمان و انجام مطالعه جداگانه‌ای خواهد بود.

براساس داده‌های حاصل شده، حجم برونداهای علمی ایران در حوزه زیست‌فناوری بین حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، برونداد‌ها در ۱۰٪ درصد استنادهای برتر، تعداد بازدیدها و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری برقرار بوده است. از این نظر می‌توان تا حدی عنوان کرد که در کنار رشد کمّی، شاخص‌های سنتی استناد و نیز میزان رؤیت‌پذیری در شبکه‌های علمی نیز ارتقا پیدا کرده‌اند. در عین حال، با افزایش این تعداد، شاخص‌های کاربردی بودن مقالات که مبتنی بر استناد در پروانه‌های ثبت اختراع است، بهبود پیدا نکرده‌اند. درواقع می‌توان گفت شکاف زیادی بین مقالات علمی و پژوهش‌های کاربردی حوزه زیست‌فناوری ایران وجود دارد. شاید بتوان وجود موانع متعدد در مسیر کاربردی کردن پژوهش‌ها، حمایت‌های ناکافی دولتی و عدم وجود انگیزه پژوهشگران برای تجاری‌سازی یافته‌های خود را از علل این امر برشمرد.

رابطه بین همکاری بین‌المللی با شاخص‌هایی ازجمله حجم استنادها، تأثیر استناد وزنی حوزه، برونداد‌ها در ۱۰٪ درصد برتر مجلات، تعداد استنادات ثبت اختراع، استنادهای حق ثبت اختراع به ازای هر خروجی علمی، خروجی علمی استناد شده توسط ثبت اختراعات، تعداد اختراعات، تعداد بازدیدها، برونداد‌ها در ۱۰٪ درصد بازدیدهای برتر، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه مثبت و معنی‌دار تشخیص داده شد. به این ترتیب، این شاخص بر بسیاری از موارد، اثرگذار بوده و درواقع مهمترین عامل مؤثر بر کیفیت مقالات ازجمله در بعد استنادها، بازدیدها و کاربردی بودن، همکاری بین‌المللی بوده و در این مطالعه، هیچ شاخصی به این میزان اهمیت نداشته است. همکاری دولتی دانشگاهی نیز با سه شاخص استناد به ازای برونداد، تعداد استنادات ثبت اختراع و استنادهای حق ثبت اختراع به ازای هر خروجی علمی رابطه مثبت و معنی‌داری داشته است. طبق این یافته می‌توان ادعا کرد با اینکه درصد این نوع

همکاری بالا بوده اما حمایت‌های دولت در مسیر اثرگذاری اقتصادی پژوهش‌های این حوزه، مؤثر افتاده به نحوی که تبادل استناد بیشتری بین مقالاتی که با همکاری نهادهای دولتی منتشر شده و پروانه‌های ثبت اختراع رخ داده است. شاخص همکاری دانشگاهی نیز رابطه مثبت و معنی‌داری با حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰٪ درصد برتر مجلات، بروندادها در ۱۰٪ درصد استنادهای برتر، تعداد استنادات ثبت اختراع، خروجی علمی استناد شده توسط ثبت اختراعات، تعداد اختراعات، تعداد بازدیدها و تأثیر بازدید وزنی حوزه داشته است. به این ترتیب، عمده نقطه اثرگذاری مقالات، در بخش استنادهای علمی و مطالعات سایر پژوهشگران همتراز تبلور یافته است. ضمن اینکه حوزه کاربرد و جامعه را نیز متأثر ساخته است.

همچنین در عرصه اثرگذاری خروجی‌های علمی مورد مطالعه، حجم استنادها با شاخص‌های تعداد استنادات ثبت اختراع، تعداد اختراعات، تعداد بازدیدها، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری داشته است. در این رابطه، محققان با کمک روش‌های یادگیری ماشین، به این نتیجه دست یافتند که ترکیبی از مؤلفه‌های مرتبط با مقاله، مجله و نویسنده در مقایسه با استفاده از تنها یک نوع مؤلفه در پیش‌بینی مقالات پراستناد، بهتر عمل می‌کند (Hansen, 2022). ایوب، امین و وانی در مطالعه خود دو مقوله استناد و نمره دگرسنجی مقالات رشته علوم زیستی، علوم زمین و محیط‌زیست، تاریخ و باستان‌شناسی را به‌منظور تعیین میزان مقایسه و ارتباط آنها با یکدیگر بررسی کردند و به این یافته رسیدند که همبستگی مثبتی بین دگرسنجی و امتیاز استناد و همچنین نمره دگرسنجی و اس.جی.آر وجود دارد (Ayoub et al. 2023). پژوهش بذرافشان و همکاران (۱۴۰۲) با بررسی رابطه همبستگی و رگرسیون چندگانه بین تعداد استنادات و نمرات دگرسنجی نشان داد که رابطه قوی بین استنادات پلام‌ایکس و تعداد استنادات اسکوپوس برقرار است.

در حوزه شاخص‌های مرتبط با تعامل بین بروندادهای علمی با پروانه‌های ثبت اختراع نیز این نتیجه حاصل شد که تعداد استنادات ثبت اختراع به عنوان یکی از شاخص‌های معرف این نوع تأثیر، با مواردی از قبیل همکاری دانشگاهی، همکاری بین‌المللی، حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، تعداد بازدیدها، برونداد در ۱۰٪ درصد بازدیدهای برتر، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری داشته است. درواقع، کاربردی شدن مقالات و پژوهش‌های حوزه زیست‌فناوری ایران زمینه‌ساز تحقق بسیاری از شرایط دیگر خواهد بود ازجمله همکاری بین‌المللی، افزایش حجم استنادها و بهبود تعداد بازدیدها. در همین رابطه، بین استنادات ثبت اختراع به ازای هر برونداد با شاخص‌های استناد به ازای هر برونداد، بازدید به ازای هر خروجی و تأثیر بازدید وزنی حوزه، رابطه مثبت و معنی‌داری برقرار بوده است. همچنین تعداد اختراعات با مواردی مانند حجم استنادها، استناد به ازای برونداد، تعداد بازدیدها، برونداد در ۱۰٪ درصد بازدیدهای برتر و بازدید به ازای هر خروجی، رابطه مثبت و معنی‌داری برقرار نموده است. نکته قابل تأمل برای سیاست‌گذاران عرصه علم و فناوری در کشور این است که برای توسعه علمی ایران، علی‌رغم رشد کمّی و کیفی تولیدات علمی، نظام پژوهشی کشور باید به سمت ارزش‌آفرینی و ایجاد ارزش‌افزوده به‌ویژه در حوزه اقتصادی سوق داده شود.

در مورد شاخص‌های دگرسنجی نیز این نتیجه به دست آمد که تعداد بازدیدها رابطه مثبت و معنی‌داری با شاخص‌هایی نظیر حجم استنادها، تأثیر استناد وزنی حوزه، تعداد استنادات ثبت اختراع، استنادهای حق ثبت اختراع به ازای هر خروجی علمی، خروجی علمی استناد شده توسط ثبت اختراعات و تعداد اختراعات این حوزه دارد. براساس این یافته می‌توان به قدرت و ظرفیت رسانه‌های اجتماعی در راستای زمینه‌سازی به‌منظور ورود آخرین یافته‌های علمی دانشمندان و پژوهشگران به فضای عمومی جامعه اذعان نمود.

یکی دیگر از پرسش‌هایی که این مطالعه درصدد پاسخ به آن بود این بود که کدام‌یک از الگوریتم‌های ماشین می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی را دقیق‌تر پیش‌بینی نمایند. براساس نتایج به دست آمده، الگوی رگرسیون خطی چندمتغیره با نمره صحت بالاتر و نمره انحراف معیار پایین‌تر، بهتر توانست میزان اثرگذاری بروندادهای علمی ایران در حوزه زیست‌فناوری را در ابعاد علمی، فناورانه و اجتماعی پیش‌بینی نماید. آلکوهر و همکاران نیز از چارچوب یادگیری ماشین برای آزمون توانایی دگرسنجه‌ها در پیش‌بینی تعداد استنادها استفاده کردند و نشان دادند از بین الگوهای درخت تصمیم، جنگل تصادفی و ماشین‌های بردار پشتیبانی، جنگل تصادفی نسبت به دیگر طبقه‌بندی‌کننده‌ها، بهتر می‌تواند استنادهای آینده را پیش‌بینی کند (Alchokr et al. 2023). برخی نیز به این نتیجه رسیدند که استفاده از تکنیک‌های مدرن یادگیری عمیق مانند شبکه عصبی بازگشتی و مدل دنباله به دنباله با توجه به دقت و

زودآیند ویرایش نشده

توانایی پیش‌بینی استنادات مقالات پراستناد، بهتر از روش‌های پیشرفته پیش‌بینی تعداد استناد است (Abrishami & Aliakbary, 2019). نتایج مطالعه‌ای دیگر نشان داد که جنگل تصادفی پیش‌بینی‌کننده‌ترین الگوریتم برای پیش‌بینی تأثیر آتی مقالات است (McNamara et al. 2013). در راستای نتایج این مطالعه، برخی محققان نیز به این نتیجه دست یافتند که برای پیش‌بینی توزیع احتمال برای تعداد استنادهای آینده، ضرایب رگرسیون چندگانه برای پیش‌بینی توزیع استنادها مناسب است (Stegehuis et al. 2015).

پیشنهادهای اجرایی پژوهش

با توجه به یافته‌های حاصل از این مطالعه، موارد زیر قابل پیشنهاد است:

- ✓ نظر به اینکه مدل رگرسیون خطی چندمتغیره، قدرت پیش‌بینی‌کنندگی بالاتری نسبت به سایر مدل‌ها داشت، پیشنهاد می‌شود مدیران، سیاست‌گذاران و متخصصان علم‌سنجی که به‌دنبال ارزیابی و آینده‌پژوهی اثرگذاری بروندادهای علمی هستند از این الگو به این منظور استفاده کنند.
- ✓ هنگام ارزیابی کمّی و کیفی مقالات، استفاده از شاخص‌های متفاوت پیشنهاد می‌شود تا تصویر شفاف‌تری از اثرگذاری پژوهش‌ها حاصل شود.
- ✓ با توجه به تأثیر همکاری بین‌المللی بر کیفیت فعالیت‌های پژوهشی، به مدیران و سیاست‌گذاران حوزه علم و فناوری کشور توصیه می‌شود زمینه‌های مشارکت پژوهشگران داخلی با پژوهشگران بین‌المللی را فراهم سازند و مشوق‌هایی را برای همکاری‌های بین‌المللی پیش‌بینی نمایند.
- ✓ به پژوهشگران نیز توصیه می‌شود تا حد ممکن مطالعات خود را با همکاری افراد بین‌المللی پیگیری و اجرا کنند.
- ✓ با در نظر گرفتن تأثیر مثبت تعداد بازدیدهایی که در فضای مجازی اتفاق می‌افتد و بر سایر شاخص‌های کیفیت و اثرگذاری علمی، مؤثر است پیشنهاد می‌شود پژوهشگران در شبکه‌های اجتماعی علمی، فعال باشند و دستاوردها و یافته‌های پژوهش‌های خود را ازطریق این سکوها با سایرین به اشتراک بگذارند تا رؤیت‌پذیری پژوهش‌ها بیشتر شود.
- ✓ با توجه به شکاف زیاد بین مقالات علمی و پژوهش‌های کاربردی حوزه زیست‌فناوری ایران، شناسایی و رفع موانع تجاری‌سازی دستاوردهای علمی، امری ضروری برای مدیران و سیاست‌گذاران این حوزه است.

پیشنهاد برای پژوهش‌های آتی

پیشنهاد می‌شود محورهای زیر، در برنامه مطالعاتی پژوهشگران بعدی قرار گیرد:

- ✓ استفاده از شاخص‌های بکار رفته در مطالعه حاضر برای حوزه‌های علمی متفاوت
- ✓ تکرار این مطالعه در سال‌های بعد و مقایسه نتایج حاصل شده
- ✓ مقایسه آمار و اطلاعات اثرگذاری پژوهش‌های علمی ایران در مقایسه با کشورهای اسلامی، منطقه، خاورمیانه یا جهان
- ✓ استفاده از الگوریتم‌های مورد استفاده در این مطالعه برای پیش‌بینی اثرگذاری پژوهش‌ها در یکی از حوزه‌های علمی
- ✓ استفاده از پایگاه‌های داده متفاوت برای استخراج داده‌ها و تحلیل اثرات مختلف آنها.

تقدیر و تشکر

این مقاله، برگرفته از بخشی از طرح پژوهشی مصوب در مرکز تحقیقات سیاست علمی کشور با عنوان: «ارزیابی تأثیرات بروندادهای علمی: مطالعه موردی حوزه زیست‌فناوری ایران» با شماره ۱۰۴/ص/۱۴۰۲ است. بدینوسیله از مرکز تحقیقات سیاست علمی کشور به سبب حمایت‌های مادی و معنوی در رابطه با این طرح، صمیمانه سپاسگزارم.

فهرست منابع

ابراهیمی، س.، دهقان، م.، و جوکار، ع. (۱۳۹۶). بررسی شاخص‌های پیش‌بینی‌کننده تأثیرگذاری علمی برای افزایش استنادگیری مقالات نشریه‌های علمی. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۲(۳)، ۶۶۱-۶۹۴. https://doi.org/10.35050/JIPM010.2017.033

آزادی احمدآبادی، ق. (۱۴۰۳). *ارزیابی تأثیرات بروندادهای علمی: مطالعه موردی حوزه زیست‌فناوری ایران*. [گزارش طرح پژوهشی]. مرکز تحقیقات سیاست علمی کشور.

آزادی احمدآبادی, ق. (۱۴۰۳). سطوح و شاخص‌های ارزیابی تأثیرات پژوهش براساس تحلیل نظام‌های ارزیابی. ترویج علم, (), -. https://doi.org/10.22034/popsci.2024.424371.1306

آزادی احمدآبادی، ق.، عبدی، س.، و رمضانی، ا. (۱۴۰۱). مطالعه تأثیرات علمی، اقتصادی و اجتماعی پژوهش‌های حوزه محیط‌زیست ایران. *محیط‌زیست و توسعه فرابخشی*، ۷(۷۸)، ۵-۳۸. https://doi.org/10.22034/envj.2023.351434.1217

باباکبری ساری، ا.، قهرمانی، م.، فتحی واجارگاه، ک.، و مؤتمنی، ع. (۱۴۰۰). ارائه الگوی ارزشیابی اثرات پژوهش‌های مدیریتی. *پژوهش‌های مدیریت در ایران*، ۲۱(۱)، ۹۳-۱۱۹. https://dorl.net/dor/20.1001.1.2322200.1396.21.1.4.6

بذرافشان، ا.، بیرانوند، ع.، و شجاعی فرد، ع. (۱۴۰۲). پیش‌بینی تعداد استنادات دریافتی حوزه فیزیک ذرات در اسکوپوس به کمک نمرات دگرسنجی پلام‌ایکس. *فصلنامه بازیابی دانش و نظام‌های معنایی*،()،-. https://doi.org/10.22054/jks.2023.71392.155

بیرانوند، ع.، گلشنی، م.، و دلقندی، ف. (۱۴۰۱). بررسی تأثیر شاخص‌های سایت‌اسکور، اس‌ان‌آی‌پی و اس‌جی‌آر نشریات حوزه وب‌معنایی بر تعداد استنادات دریافتی مقالات. *فصلنامه بازیابی دانش و نظام‌های معنایی*.- ,(), https://doi.org/10.22054/jks.2022.67616.1501

نوروزی چاکلی، عبدالرضا (۱۴۰۲). *آشنایی با علم سنجی (مبانی، مفاهیم، روابط و ریشه‌ها)*. تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)، مرکز تحقیق و توسعه علوم انسانی؛ دانشگاه شاهد، مرکز چاپ و انتشارات. https://samt.ac.ir/fa/book/99/

Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, *13*(2), 485-499. https://doi.org/10.48550/arXiv.1809.04365

Akella, A. P., Alhoori, H., Kondamudi, P. R., Freeman, C., & Zhou, H. (2021). Early indicators of scientific impact: Predicting citations with altmetrics. *Journal of Informetrics*, *15*(2), 101128. https://doi.org/10.1016/j.joi.2020.101128

Alchokr, R., Haider, R., Shakeel, Y., Leich, T., Saake, G., & Krüger, J. (2023). Forecasting Publication's Success Using Machine Learning [Conference presentation]. In *International Workshop on Bibliometric-Enhanced Information (BIR)*. CEUR-WS. org. https://jacobkrueger.github.io/assets/papers/Alchokr2023ForcastingSuccess.pdf

Alohali, Y. A., Fayed, M. S., Mesallam, T., Abdelsamad, Y., Almuhawas, F., & Hagr, A. (2022). A machine learning model to predict citation counts of scientific papers in otology field. *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/2239152

Anninos, L. N. (2014). Research performance evaluation: Some critical thoughts on standard bibliometric indicators. Studies in Higher Education, 39, 1542–1561. https://doi.org/10.1080/03075079.2013.801429

Ayoub, A., Amin, R., & Wani, Z. A. (2023). Exploring the Impact of Altmetrics in Relation to Citation Count and SCImago Journal Rank (SJR). *Journal of Scientometric Research*, *12*(3), 603-608. https://doi.org/10.5530/jscires.12.3.058

Azadi Ahmadabadi, G., Abdi, S., & Ramezani, A. (2022). Studying the Scientific, Economic and Social Effects of Iranian Environmental Researches. *Environment and Interdisciplinary Development*, 7(78), 38-55. https://doi.org/10.22034/envj.2023.351434.1217 [In Persian].

Azadi Ahmadabadi, G. (2024). *Evaluation of the effects of scientific outputs: case study of Iran's biotechnology* [Research project report]. National Research Institute for Science Policy (NRISP). [In Persian].

Azadi Ahmedabadi, G. [In Press]. Levels and indicators of evaluation of research impacts based on the analysis of evaluation systems. *Popularization of Science*. https://doi.org/10.22034/popsci.2024.424371.1306 [In Persian].

Babaakbarisari, A., Ghahremani, M., Fathi vajargah, K., Moatameni, A. (2021). Developing Management Researches Impacts Assessment Model. *Management Research in Iran*, 21(1), 93-119. https://dorl.net/dor/20.1001.1.23222200.1396.21.1.4.6 [In Persian].

Bazrafshan, A., Biranvand, A., & Shojaeifard, A. (2023). Predicting the number of citations received in particle physics Scopus with the help of Plumx-Altmetric scores. *Knowledge Retrieval and Semantic Systems*, (), -. https://doi.org/10.22054/jks.2023.71392.155 [In Persian].

Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., & Xia, F. (2017). An overview on evaluating and predicting scholarly article impact. *Information*, *8*(3), 73. https://doi.org/10.3390/info8030073Bernard, S., Adam, S., & Heutte, L. (2012). Dynamic random forests. *Pattern Recognition Letters*, *33*(12), 1580-1586. https://doi.org/10.1016/j.patrec.2012.04.003

Biranvand, A., Golshani, M., & Delghandi, F. (2022). Investigating the impact of Citescore, SNIP, and SJR indicators of semantic web publications on the number of received citations of articles. *Knowledge Retrieval and Semantic Systems*, (), -. https://doi.org/10.22054/jks.2022.67616.1501 [In Persian].

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS one*, *4*(6), e6022. https://doi.org/10.1371/journal.pone.0006022

Bornmann, L., Leydesdorff, L., & Wang, J. (2013). Which percentile-based approach should be preferred for calculating normalized citation impact values? an empirical comparison of five approaches including a newly developed citation-rank approach (p100). *Journal of Informetrics*, 7(4), 933–944. https://doi.org/10.1016/j.joi.2013.09.003

Ebrahimy, S., Dehghan, M., & Jowkar, A. (2017). Evaluation the predictive indicators of scientific impact to increase the citations of articles in scientific journals. *Iranian Journal of Information Processing and Management*, *32*(3), 661-694. https://doi.org/10.35050/JIPM010.2017.033 [In Persian].

Gu, X., & Krenn, M. (2024). Forecasting high-impact research topics via machine learning on evolving knowledge graphs. *arXiv preprint arXiv:2402.08640*. https://doi.org/10.48550/arXiv.2402.08640

Guthrie, S., Wooding, S., Grant, J., Diepeveen, S., & Wamae, W. (2013). *Developing a research evaluation framework*. RAND Corporation, MG-1217-AAMC, 2013. As of February 20, 2024: https://www.rand.org/pubs/monographs/MG1217.html

Hansen, I. S. (2022). *Predicting the impact of academic articles on marketing research: Using machine learning to predict highly cited marketing articles* [Unpublished master's dissertation] Norwegian School of Economics. https://openaccess.nhh.no/nhh-xmlui/bitstream/handle/11250/3015929/masterthesis.pdf?sequence=1

Hastie, T., & Tibshirani, R. (1995). Discriminant adaptive nearest neighbor classification and regression. *Advances in neural information processing systems*, *8*, 409-415. https://dl.acm.org/doi/10.5555/2998828.2998886

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., & Friedman, J. (2009). Boosting and additive trees. *The elements of statistical learning: data mining, inference, and prediction*, 337-387. https://hastie.su.domains/Papers/ESLII.pdf

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kosteas, V. D. (2018). Predicting long-run citation counts for articles in top economics journals. *Scientometrics*, *115*(3), 1395-1412. https://doi.org/10.1007/s11192-018-2703-0

Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140-147. https://doi.org/10.38094/jastt1457

McNamara, D., Wong, P., Christen, P., & Ng, K. S. (2013). Predicting high impact academic papers using citation network features [Conference presenation]. In *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD* 2013 International Workshops: DMApps, DANTH, QIMIE, BDM, CDA, CloudSD, Gold Coast, QLD, Australia, April 14-17, 2013, Revised Selected Papers 17 (pp. 14-25). Springer Berlin Heidelberg. https://dblp.org/rec/conf/pakdd/2013-w.html

Mihaela, P. A. U. N., Abigaela, B. I. L. B. I. I. E., Paul, B. U. Z. O. I. A. N. U., Anastasia, C. O. S. M. A., Catalina, E. N. E., Anne, H. R. I. S. C. U., ... & Eduard, M. I. L. E. A. (2020). Predicting long-term citation counts in Web of Science: COVID-19 early publications case study. *Romanian Statistical Review*, (4). https://www.revistadestatistica.ro/wp-content/uploads/2020/12/A4-RRS4_2020.pdf

Newson, R., King, L., Rychetnik, L., Bauman, A. E., Redman, S., Milat, A. J., ... & Chapman, S. (2015). A mixed methods study of the factors that influence whether intervention research has policy and practice impacts: perceptions of Australian researchers. *BMJ open*, *5*(7), e008153. https://doi.org/10.1136/bmjopen-2015-008153

Noroozi Chakoli, A. (2023). *Introduction to scientometric (foundations, concepts, relations & origins)*. Tehran: SAMT; Shahed University. https://samt.ac.ir/fa/book/99/ [In Persian]

Piryonesi, S. M., & El-Diraby, T. E. (2020). Data analytics in asset management: Cost-effective prediction of the pavement condition index. *Journal of infrastructure systems*, *26*(1), 04019036. http://dx.doi.org/10.1061/(ASCE)IS.1943-555X.0000512

Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of informetrics*, *9*(3), 642-657. https://doi.org/10.1016/j.joi.2015.06.005

Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological methods & research*, *40*(3), 471-510. https://doi.org/10.1177%2F0049124111415372

Timilsina, M., Davis, B., Taylor, M., & Hayes, C. (2016, August). Towards predicting academic impact from mainstream news and weblogs: A heterogeneous graph-based approach [Conference presentation]. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1388-1389). IEEE. https://ieeexplore.ieee.org/document/7752425

Weihs, L., & Etzioni, O. (2017). Learning to predict citation-based impact measures. [Conference presentation]. In 2017 ACM/IEEE *joint conference on digital libraries* (JCDL), (pp. 1–10). https://dl.acm.org/doi/abs/10.5555/3200334.3200341

Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research*, 10(2). https://jmlr.csail.mit.edu/papers/volume10/weinberger09a/weinberger09a.pdf

Williams, K., & Lewis, J. M. (2021). Understanding, measuring, and encouraging public policy research impact. *Australian Journal of Public Administration*, 80(3), 554-564. https://doi.org/10.1111/1467-8500.12506

Wooldridge, J., & King, M. B. (2019). Altmetric scores: An early indicator of research impact. *Journal of the Association for Information Science and Technology*, 70(3), 271–282. https://doi.org/10.1002/asi.24122

Wu, J., Liu, C., Cui, W., & Zhang, Y. (2019, November). Personalized Collaborative Filtering Recommendation Algorithm based on Linear Regression [Conference presentation]. In *2019 IEEE International Conference on Power Data Science (ICPDS)* (pp. 139-142). IEEE. https://doi.org/10.18280/mmep.060307

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*, 1-37. https://doi.org/10.1007%2Fs10115-007-0114-2

Yu, T., Yu, G., Li, P. Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, 101, 1233-1252. https://doi.org/10.1007/s11192-014-1279-6

Zhang, F., & Wu, S. (2020). Predicting future influence of papers, researchers, and venues in a dynamic academic network. *Journal of Informetrics*, 14(2), 101035. https://doi.org/10.1016/j.joi.2020.101035

Ziegler, A., & König, I. R. (2014). Mining data with random forests: current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(1), 55-63. https://doi.org/10.1002/widm.1114