Scientometrics Research Journal

Scientific Bi-Quarterly of Shahed University
Vol. 11, No. 1, Spring & Summer 2025 (Serial 21)

Research Article

DOI: 10.22070/rsci.2024.18868.1719

Predicting Scientific Research Impacts in Biotechnology by Machine Learning Algorithms

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.

Email: azadi_gh@yahoo.com

Received: 22/08/2024

Revised: 28/10/2024

Accepted: 12/11/2024

Early online access: 19/11/2024

Published: 01/04/2025

Abstract

Purpose: Research impact is a key concern for stakeholders, as it reflects the beneficial and profitable applications of research across multiple dimensions, including society, economy, environment, culture, and health. This study aims to analyze the interrelationships among variables influencing scientific outputs and to identify the most effective machine learning algorithms for predicting their scientific, social, and economic impacts.

Methodology: The current research is applied in purpose and descriptive in method, utilizing a scientometric approach. This study aims to explore the relationship between the quantity of scientific publications and scientific cooperation, as well as the scientific, social, and economic impact of Iran's scientific contributions in the field of biotechnology. Additionally, it seeks to determine which machine learning algorithms are more effective in predicting the impact of scientific outputs across various dimensions. The research focuses on Iran's biotechnology scientific outputs indexed in the Scopus database from 2003 to 2024. The data, extracted on 21 January 2024, were analyzed using the SciVal analytical database. In this research, Pearson's correlation coefficient and the R software package were used to examine the relationships between the studied indicators. Machine learning algorithms, including multiple linear regression, nearest neighbors, decision trees, random forests, and gradient boosting, were applied and evaluated as predictive models. The Python programming language was employed to conduct tests and implement these algorithms.

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. A positive and significant relationship was observed between international collaboration and various impact indicators, including citation 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,

Scientometrics Research Journal

Scientific Bi-Quarterly of Shahed University
Vol. 11, No. 1, Spring & Summer 2025 (Serial 21)

Research Article

DOI: 10.22070/rsci.2024.18868.1719

Ghasem Azadi Ahmadabadi [1*]

Received: *22/08/2024*

Revised: *28/10/2024*

Accepted: *12/11/2024*

Early online access: *19/11/2024*

Published: *01/04/2025*

Views per Publication, and Field-Weighted Views Impact. The index of academic collaboration also shows a positive and significant relationship with several indicators, including 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, Scholarly Output cited by Patents, Views Count, and Field-Weighted Views Impact. Academic-government collaboration is also positively and significantly correlated 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 are positively and significantly associated with several indicators, including Scholarly Output cited by Patents, Patents Count, Views Count, Views per Publication, and the Field-Weighted Views Impact of the field of biotechnology. In terms of economic impact, the result indicated that the number of patent citations is a key representative indicator. This indicator showed positive and significant relationships with other metrics, including academic Collaboration, international Collaboration, citation 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 in the field of biotechnology. Regarding social impact, the analysis revealed that Views Count is positively and significantly correlated with several indicators, including 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 biotechnology. Based on the results, multivariate linear regression demonstrated higher accuracy and a lower standard deviation, making it a more effective model for predicting the scientific, technological, and social impact of Iran's scientific outputs in biotechnology.

Conclusion: International cooperation is the most significant factor influencing the quality of articles, including metrics such as citations, views, and applicability. Therefore, it is crucial to explore and implement strategies to enhance such collaborations. To better assess the effectiveness of research, it is recommended to employ a diverse set of indicators during both quantitative and qualitative evaluations of articles. For policymakers in science and technology, the primary focus should be on value creation and generating added value, particularly in the economic sector. This is essential despite the observed quantitative and qualitative growth in Iran's scientific outputs.

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

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی دانشگاه شاهد/ دوره ۱۱/ شماره ۱/ بهار و تابستان ۱۴۰۴ (پیاپی ۲۱)

مقاله پژوهشی

DOI: 10.22070/rsci.2024.18868.1719

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

قاسم آزادی احمدآبادی [1*]

۱. استادیار، علم اطلاعات و دانش‌شناسی، گروه ارزیابی سیاست‌ها و پایش علم، فناوری و نوآوری، مرکز تحقیقات سیاست علمی کشور، تهران، ایران.

Email: azadi_gh@yahoo.com

صفحه ۲۴-۱

دریافت: ۱۴۰۳/۰۶/۰۱

بازنگری: ۱۴۰۳/۰۸/۰۷

پذیرش: ۱۴۰۳/۰۸/۲۲

زودآیند: ۱۴۰۳/۰۸/۲۹

انتشار: ۱۴۰۴/۰۱/۱۲

چکیده

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

روش‌شناسی: پژوهش حاضر ازنظر هدف، کاربردی و ازلحاظ روش، توصیفی بوده و با رویکرد علم‌سنجی انجام‌شده است. جامعه پژوهش، بروندادهای حوزهٔ زیست‌فناوری ایران است که در پایگاه اسکوپوس در بازهٔ ۲۰۰۳-۲۰۲۴ نمایه شده‌اند. برای 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 موردمطالعه استفاده شد و رگرسیون خطی چندگانه، نزدیک‌ترین همسایه، درخت‌های تصمیم‌گیری، جنگل‌های تصادفی و تقویت گرادیان نیز به‌عنوان مدل‌های پیش‌بینی کننده، مورد و ارزیابی قرار گرفت. به‌منظور انجام آزمون‌ها و الگوریتم‌ها از زبان برنامه‌نویسی پایتون استفاده شد.

یافته‌ها: در عرصه اثرگذاری خروجی‌های علمی موردمطالعه، حجم استنادها با شاخص‌های متعددی رابطه مثبت و معنی‌داری داشته است. در حوزهٔ تأثیرگذاری اقتصادی نیز این نتیجه حاصل شد که تعداد استنادات ثبت اختراع به‌عنوان 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. بر اساس نتایج حاصل‌شده، رگرسیون خطی چند متغیره با نمره صحت بالاتر و نمره انحراف معیار پایین‌تر، بهتر توانست میزان اثرگذاری علمی، فناورانه و اجتماعی بروندادهای علمی ایران در حوزهٔ زیست‌فناوری را پیش‌بینی کند.

نتیجه‌گیری: مهم‌ترین عامل مؤثر بر کیفیت مقالات ازجمله در بُعد استنادها، بازدیدها و کاربردی بودن، همکاری بین‌المللی بوده که در این زمینهٔ لازم است تدابیری اندیشه شود. پیشنهاد می‌شود هنگام ارزیابی کمّی و کیفی مقالات، از شاخص‌های متنوعی استفاده گردد 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.

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، موضوع سنجش این تولیدات از زوایای گوناگون نیز اهمیت بسیاری یافته است. از ملزومات سیاست‌گذاری کارآمد و درست در حوزهٔ علم و تحلیل حرکت علمی کشور، انجام پژوهش‌هایی برای شناسایی وضعیت موجود به‌منظور شناسایی نقاط قوت و ضعف آن‌هاست. شاخص‌های علم‌سنجی، مهم‌ترین ابزارهایی هستند که می‌تواند چنین اطلاعاتی را در اختیار برنامه‌ریزان و سیاست‌گذاران قرار دهند (آزادی احمدآبادی، ۱٤۰۳).

با توجه به نقش و اهمیت پژوهش در توسعه و پیشرفت جوامع در برنامه‌های کلان کشورها به‌منظور نیل به گسترش و پیشرفت، به این موضوع توجه اساسی شده و سرمایه‌گذاری‌های کلانی در دانشگاه‌ها و مراکز پژوهشی در این رابطه می‌شود. ازاین‌جهت، 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، ۱٤۰۰). از سویی دیگر، ارزیابی تأثیر[1] پژوهش‌ها و تعیین میزان تحقق اهداف به‌آسانی ممکن نیست؛ زیرا اثربخشی واقعی پژوهش‌ها غیرخطی و غیرقابل‌پیش‌بینی هستند. آنچه امروزه برای ارزیابی اثربخشی پژوهش موردتوجه قرار می‌گیرد معیارهای سنتی از قبیل تعداد مقالات، ضریب تأثیر نشریات، تعداد گرنت‌ها و تعداد استنادها و غیره هستند (Newson et al., 2015). درواقع، «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 به شمار می‌رود (آزادی احمدآبادی و همکاران، ۱٤۰۱).

واقعیت این است که سنجش تأثیر پژوهش‌ها در ابعاد مختلف به‌سادگی ممکن نیست. آنچه در حال حاضر و به‌طور واقعی در رابطه با تأثیر پژوهش‌های انجام‌گرفته 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 منتشرشده در سطح جهان، نیاز به ارزیابی و روش‌های نوین سنجش کیفیت مقالات پژوهشی در حال افزایش است (Abrishami & Aliakbary, 2019). بر این اساس، سنجه‌های مختلفی به‌منظور سنجش و ارزیابی تولیدات علمی به جامعه علمی معرفی‌شده که هرکدام مزایا و معایب خاص خود را دارند و هرروزه 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). دگر سنجی 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 ظاهرشده 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، به دنبال شناسایی اثرگذاری پژوهش‌های علمی بر فناوری بوده‌اند. بر این اساس، پایگاه سایول[2]، شاخص جدیدی را مطرح کرده که بر اساس آن، میزان کاربردی بودن پژوهش‌ها را رصد

1 . Impact
2 . Scival

می‌کند. این کار از طریق پیگیری میزان استنادهای دریافتی و خروجی به پروانه‌های ثبت اختراع انجام می‌شود. در کنار بررسی و ارزیابی اثر پژوهش‌ها که بیشتر مبتنی برگذشته است، پیش‌بینی آن نیز می‌تواند از حوزه‌های جذاب و کاربردی در این زمینه 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 رابطهٔ هر یک از متغیرهای مختلف بر همدیگر را موردسنجش 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[1] ایران که در پایگاه اسکوپوس[2] در بازهٔ ۲۰۰۳-۲۰۲٤ نمایه شده‌اند موردمطالعه قرارگرفته است. درواقع، پرسش اساسی پژوهش حاضر این است که کدام‌یک از الگوریتم‌های ماشین در مقایسه با یکدیگر می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی پژوهش‌های حوزهٔ زیست‌فناوری ایران را بهتر پیش‌بینی کنند؟

پرسش‌های پژوهش

۱. وضعیت بروندادهای حوزهٔ زیست‌فناوری ایران بر اساس هرکدام 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؟

٤. 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، حاصل‌شده و تمرکز آن روی «کیفیت» و «تأثیر» است. چگونگی درک و سنجش تعالی، نحوهٔ تصمیم‌گیری پژوهشگران را شکل می‌دهد و به‌سهم خود، انواع دانشی تولید می‌شود که به‌طور بالقوه به نتایج سیاسی، اقتصادی یا اجتماعی منجر می‌گردد. 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 تجزیه‌وتحلیل کمّی پیشرفت‌های علمی است و در حوزهٔ «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).

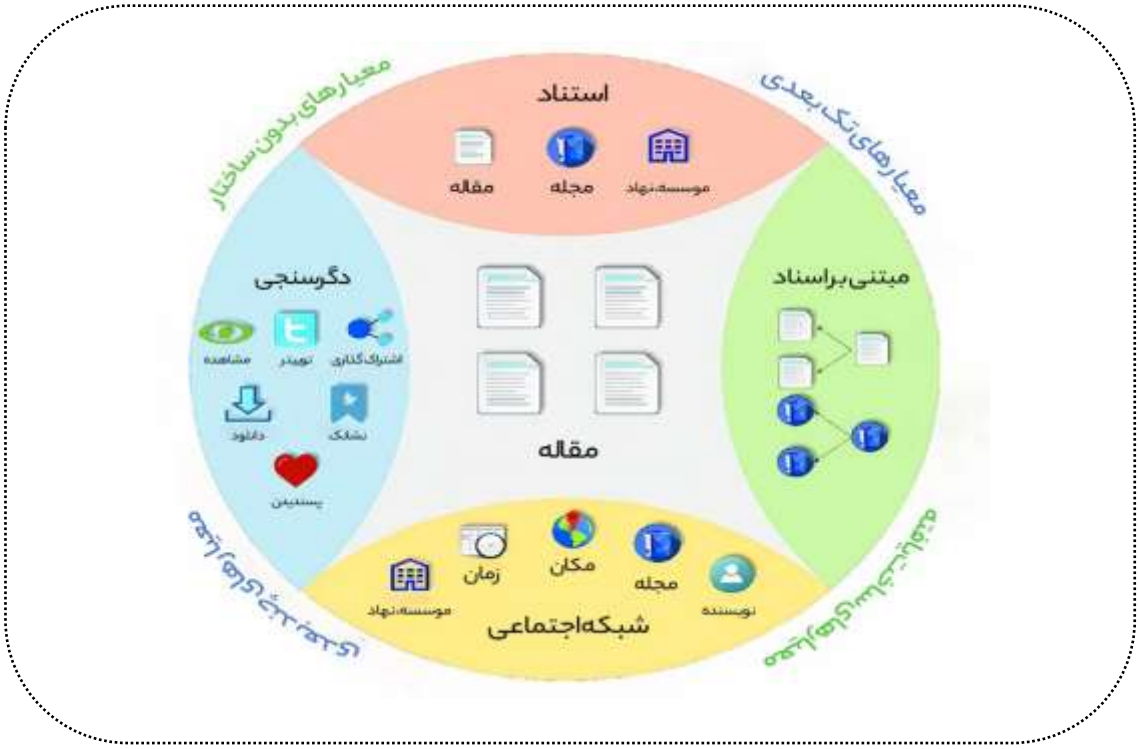
امروزه مجموعهٔ متنوعی از شاخص‌ها به‌منظور ارزیابی کیفیت و اثرگذاری پژوهش‌ها، از معیارهای سنتی تا جدیدترین معیارهای جدید و جایگزین به کار گرفته می‌شود. کتاب‌سنجی یکی از این حوزه‌های دیرینه‌ای است که شاخص‌های مبتنی بر استناد را برای سنجش پژوهش‌های علمی به خدمت می‌گیرد. در تحلیل‌های استنادی، سنجش کیفیت

1 . Biotechnology
2 . Scopus

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

مقاله بر پایه استناد اتفاق می‌افتد. اگرچه استناد، سنجه علمی است ولی عدد مطلق نمی‌تواند گویای کیفیت باشد. به همین منظور، استنادها به هنجار می‌شوند تا قابل‌تعمیم به جامعه موردبررسی باشند. عموماً به مقاله‌های پر استناد به‌عنوان نشانه‌ای از کیفیت فعالیت‌های علمی و درنتیجه، ارزیابی عملکرد پژوهشی توجه خاصی می‌شود. باوجوداینکه استناد از مهم‌ترین و قابل‌قبول‌ترین شاخص‌ها برای ارزیابی عملکرد و کارایی علمی است، اما با محدودیت‌هایی همچون وابستگی زمانی، وابستگی به پایگاه‌های گران‌قیمت و با محدودیت دسترسی همراه است (Guthrie et al., 2013). گام‌های روش‌شناختی در مطالعات ارزیابی پژوهش از اهمیت بسیاری برخوردار است. آنینوس این مراحل را این‌چنین شرح می‌دهد: فرآیند ارزشیابی پژوهش با تعیین سطح علمی آغاز می‌شود. به دنبال آن، دورهٔ 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 انجام‌شده و شاخص‌های کتاب‌سنجی در مرحلهٔ 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 منتشرشده 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 بُعد به چند بُعد، از معیارهای غیرساختارمند به معیارهای ساخت‌یافته رخ‌داده که در شکل 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، بنابراین سیگنالی منحصربه‌فرد برای مطالعهٔ 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 گذشته و حال تمرکز دارند، همچنان این سؤال باقی است 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). درعین‌حال، به‌منظور 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 پیش‌بینی کننده با استفاده از الگوریتم‌های یادگیری ماشین

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

1 . Hircsh Index

۱. درخت تصمیم[1]

درختان تصمیم یکی از محبوب‌ترین الگوریتم‌های یادگیری ماشینی با توجه به درک و سادگی آن‌ها هستند (Wu et al., 2008). درخت تصمیم همچون 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 نشان‌دهندهٔ 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 نظارت‌شده 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 پذیرفته‌شده‌اند (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 حوزهٔ 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 روش برای ارزیابی و مدل‌سازی داده است که روابط خطی را بین متغیرهای وابسته و مستقل برقرار می‌کند؛ پس این روش، روابط بین متغیرهای وابسته و متغیرهای مستقل را از تحلیل و یادگیری تا نتایج آموزشی فعلی مدل می‌کند (Wu et al., 2019).

٤. 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؛ به‌عنوان‌مثال، مدل‌هایی که مفروضات بسیار کمی در مورد داده‌ها ایجاد می‌کنند که معمولاً درخت‌های تصمیم ساده هستند. هنگامی‌که درخت تصمیم یادگیرنده ضعیف باشد، الگوریتم حاصل درختان با گرادیان تقویت‌شده نامیده می‌شود (Piryonesi & El-Diraby, 2020; Hastie et al., 2009).

1 . Decision Tree
2 . random forests
3 . Regression
4 . Linear Regression
5 . Gradient boosting

۵. قانون k- نزدیک‌ترین همسایه[1]

قانون k- نزدیک‌ترین همسایه نیز از قدیمی‌ترین و ساده‌ترین روش‌ها برای طبقه‌بندی الگوهاست. بااین‌وجود، اغلب نتایج رقابتی به همراه دارد و در حوزه‌های خاص، هنگامی‌که 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 مورداستفاده برای شناسایی نزدیک‌ترین همسایگان بستگی دارد. در الگوی k- نزدیک‌ترین همسایه برای طبقه‌بندی آماری و رگرسیون، در هر دو حالت k شامل نزدیک‌ترین مثال آموزشی در فضای داده‌ای است و خروجی آن بسته به نوع مورداستفاده در طبقه‌بندی و رگرسیون متغیر است. در حالت طبقه‌بندی با توجه به مقدار مشخص‌شده برای k، 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 این‌که محاسبات این الگوریتم بر اساس 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 انجام‌شده که در ادامه به برخی از مهم‌ترین و مرتبط‌ترین آن‌ها اشاره می‌شود.

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

در پژوهش بیرانوند و همکاران (۱۴۰۱) امکان پیش‌بینی تعداد استنادات دریافتی تولیدات علمی حوزهٔ وب‌معنایی از طریق شاخص‌های سایت-اسکور، اس‌ان‌آی‌پی و اس‌جی‌آر نشریات این حوزه مورد بررسی قرار گرفته است. نتایج آزمون ضریب همبستگی پیرسون و رگرسیون خطی چندگانه نشان داد بین تعداد استنادات با شاخص اس‌ان‌آی‌پی همبستگی وجود ندارد. همچنین بین تعداد استنادات با اس‌جی‌آر و تعداد استنادات با سایت اسکور، رابطه معناداری وجود ندارد. هیچ‌یک از شاخص‌های سه‌گانه سایت اسکور، 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 نشریات علمی بود تا بتوان در فرایند کنترل کیفیت مقالات علمی، میزان استناد به آن در آینده را پیش‌بینی کرد. در این پژوهش، ۳۰۰ مقاله حوزهٔ شیمی بررسی شد و نتایج نشان داد که شاخص «چند نویسندگی و هم تألیفی» قوی‌ترین متغیر پیش‌بینی کنندهٔ استناد است که به‌صورت مثبت و معنادار میزان استناد را پیش‌بینی می‌کند.

1 . k-nearest neighbors' algorithm

شاخص‌های «طول مقاله»، «تعداد منابع» و «شاخص اچ نویسنده» متغیرهای دیگری بودند که توانستند میزان استناد مقالات را به نحو 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 مدل‌های رگرسیون خطی روی مجموعه‌ای از ۱۲۳۱۲۸ پایگاه وب‌آوساینس تألیف 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 تأثیر در ترکیب تنها پس از دو سال از انتشار ناچیز می‌شود (Gu & Krenn, 2024).

آلکوهر و همکاران نیز از یک چارچوب یادگیری ماشین برای آزمون توانایی معیارهای جایگزین (دگر سنجه) در پیش‌بینی تأثیر آتی مقالات بر اساس تعداد استنادها استفاده کردند. مبنای این بررسی، ۷۵۸۸ مقاله از ۱۰ مجله علوم رایانه بود. برای ساخت فضای ویژه برای پیش‌بینی، ۱٤ شاخص دگر سنجی 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 بر اساس 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 بر اساس عملکرد، جنگل تصادفی نسبت به دیگر طبقه‌بندی کننده‌ها، دگر سنجه‌ها می‌تواند استنادهای آینده را پیش‌بینی کند و مفیدترین نشانه‌های دگر سنجی، 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 نمرهٔ دگر سنجی را به‌منظور تعیین میزان مقایسه و ارتباط آن‌ها با یکدیگر بررسی کردند. نمرات دگر سنجی با تعداد استناد مقالات مرتبط با رشته علوم زیستی، زمین‌شناسی و محیط زیست، تاریخ و باستان‌شناسی و مطالعات اجتماعی از لیست ۱۰۰ مقاله برتر در پایگاه آلتمتریک در دورهٔ زمانی ۲۰۱٤ تا ۲۰۱۷ مقایسه شد. علاوه بر این، نمرات دگر سنجی 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 نمرهٔ دگر سنجی نشان داد، درحالی‌که برای نمرهٔ دگر سنجی و اس.جی.آر. درزمینهٔ مذکور، چنین روندی قابل‌مشاهده نیست. تاریخ و باستان‌شناسی، همبستگی قوی بین دگر سنجی و نمرات استناد را به‌استثنای مقالات سال ۲۰۱٦ نشان می‌دهد و تا حدودی روند مشابهی برای امتیاز دگر سنجی و اس.جی.آر. انتشارات مشاهده شد. علوم زیستی همبستگی ضعیفی در هر دو متغیر نشان می‌دهند، درحالی‌که حوزهٔ مطالعات اجتماعی عمدتاً برای هر دو مجموعه متغیر، ارتباط منفی نشان می‌دهند. بنابراین، به‌جز مقوله مطالعات اجتماعی، سه دسته دیگر عمدتاً همبستگی مثبتی بین دگر سنجی و امتیاز استناد نشریات و همچنین تا حدی، برای نمرهٔ دگر سنجی 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 درزمینهٔ شنوایی‌سنجی انجام شد. این مطالعه، راه‌حل جدیدی را پیشنهاد داده و از یادگیری ماشین و پردازش زبان طبیعی برای پردازش متن انگلیسی استفاده کرد. الگوریتم‌های مختلفی مانند رگرسیون خطی، درخت تصمیم تقویت‌شده، 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. بر این 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، نشریه و نویسنده به‌عنوان شاخص‌های تأثیر مقالات حوزهٔ بازاریابی و اینکه چگونه می‌توانند مقالات پر استناد 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 هم نویسندگی نشان داد که مرکزیت نزدیکی و بینیت نیز محرک استنادهای آینده در حوزهٔ 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، نشریه و نویسنده در مقایسه با تنها یک نوع مؤلفه در پیش‌بینی مقالات پر استناد بهتر عمل می‌کنند (Hansen & Torvund, 2022).

در مقاله ژانگ و وو، چارچوب جدید WMR-Rank را برای پیش‌بینی تأثیر آینده مقالات، محققان و مکان‌ها پیشنهاد می‌کنند. بر اساس شبکهٔ پویا و ناهمگون موجودیت‌های متعدد، هفت نوع رابطه از بین آن‌ها استخراج شد. این چارچوب از ویژگی‌های مفیدی ازجمله جزئیات دقیق نهادهای مرتبط مانند نویسندگان و مکان‌های برگزاری، آگاهی از زمان برای مقالات منتشرشده و استنادات آن‌ها، متمایز کردن مشارکت چند نویسندهٔ همکار در یک مقاله و غیره پشتیبانی می‌کند. نتایج تجربی نشان داد که رویکرد پیشنهادی به‌طور قابل‌توجهی بهتر از رقبای پیشرفته عمل می‌کند (Zhang & Wu, 2020).

میهالا و همکارانش در مطالعه خود ۱۲۶ مقاله در مورد کووید ۱۹ در پایگاه وب‌آوساینس را برای تعیین اینکه آیا دگر سنجی با تأثیر علمی مقاله مرتبط است یا خیر، تحلیل کردند. روابط و سنجه‌های سنتی و دگر سنجی از طریق 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 دگر سنجی، همبستگی نسبتاً کمی با استنادهای آتی نشان می‌دهد؛ درحالی‌که گوگل اسکولار که با احتیاط استفاده می‌شود، می‌تواند پیش‌بینی کنندهٔ تعداد استنادهای آینده باشد (Mihaela et al., 2020).

ابریشمی و علی اکبری در مطالعهٔ خود، روش جدیدی را به‌منظور پیش‌بینی استنادهای بلندمدت یک مقاله بر اساس 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 شبکهٔ عصبی بازگشتی و مدل دنباله به دنباله) به‌منظور یادگیری روش پیش‌بینی بر اساس الگوی توالی استنادها از سال‌های اولیه انتشار مقاله استفاده کردند. ارزیابی‌های جامع آن‌ها نشان داد که روش پیشنهادی با توجه به‌دقت و توانایی پیش‌بینی استنادات مقالات پر استناد، بهتر از روش‌های پیشرفته پیش‌بینی تعداد استناد است (Abrishami & Aliakbary, 2019).

پژوهشگر دیگری، قدرت نسبی تعداد استنادهای کوتاه‌مدت، معیارهای کتاب‌سنجی مانند عوامل تأثیر و رتبه‌بندی نشریات را ازنظر پیش‌بینی استنادهای بلندمدت بررسی کرد. وی با استفاده از مجموعه‌ای از مقالات منتشرشده در شصت نشریه اقتصادی دریافت استنادهای دریافتی در بازه‌های کوتاه‌مدت (بین ۱ تا ۲ سال پس از انتشار) در مقایسه

با مواردی مانند تأثیر نشریات، پیش‌بینی کننده‌های قوی‌تری برای تعداد استنادات بلندمدت یا سایر رتبه‌بندی‌های نشریات هستند. این مقاله بر مقایسه سهم نسبی استنادهای کوتاه‌مدت، معیارهای کیفیت نشریات مانند عوامل تأثیر نشریه و امتیازات رتبه‌بندی نشریه و ویژگی‌های مقاله تأکید داشت. از دید این پژوهشگر، این واقعیت که تعداد استنادهای کوتاه‌مدت نسبت به متغیرهای شاخص نشریه، پیش‌بینی کننده بهتری برای استنادهای بلندمدت ارائه می‌کند، یک نتیجه واقعاً منحصربه‌فرد است (Kosteas, 2018).

تیمیلسینا و همکاران 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 به‌طور 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 به‌منظور پیش‌بینی تعداد استنادات در حوزهٔ فیزیک استفاده کردند. تحلیل‌ها نشان داد که هر دو پیش‌بینی کننده (یعنی ضریب تأثیر و استنادهای اولیه) به پیش‌بینی دقیق تأثیر استناد درازمدت کمک می‌کنند (Stegehuis et al., 2015).

بر اساس رگرسیون چندگانه، استگهویس و همکاران (Stegehuis et al., 2015) مدلی برای پیش‌بینی توزیع احتمال به‌منظور استنادهای آینده یک مقاله پیشنهاد کردند و دو ویژگی کلیدی را در نظر گرفتند: استنادهای اولیه و ضریب تأثیر نشریه.

یو و همکاران از چهار دسته از ویژگی‌ها، ازجمله مقالات، نویسندگان، استنادها و نشریات برای پیش‌بینی استنادهای آینده یک مقاله بر اساس تحلیل رگرسیون گام‌به‌گام استفاده کردند. بر اساس شبکه‌های هم نویسندگی، 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 پیش‌بینی کننده بر اساس 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، به‌طوری‌که پیش‌بینی ۱۸ درصدی از سال انتشار مقاله و ۳٤ درصدی پیش‌بینی استناد از دو سال پس‌ازآن به دست آمد (McNamara et al., 2013).

آکلا و همکارانش در پژوهش خود از دگر سنجه‌ها 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 موضوع هستند. آن‌ها دریافتند که خوانندگان مندلی مهم‌ترین عامل در پیش‌بینی استنادهای اولیه بوده و به دنبال آن، عوامل دیگری مانند وضعیت علمی خوانندگان (به‌عنوان‌مثال، دانشجو، فوق دکترا، استاد)، دنبال کنندگان در توییتر، طول پست آنلاین، تعداد نویسنده و تعداد به اشتراک گذاشتن در توییتر، ویکی‌پدیا و در کشورهای مختلف (Akella et al., 2009).

بولان و همکاران تحلیل مؤلفه اصلی رتبه‌های تولیدشده توسط ۳۹ معیار تأثیر علمی موجود و پیشنهادی که بر اساس داده‌های استناد و گزارش استفاده، محاسبه‌شده، انجام دادند. نتایج نشان داد که تأثیر علمی یک سازه چند بُعدی است و نمی‌توان آن را به‌طور دقیق با هیچ شاخصی اندازه‌گیری کرد، اگرچه برخی از معیارها مناسب‌تر از سایرین هستند. ضریب تأثیر استنادی که معمولاً مورداستفاده است در هسته این ساختار جای نگرفته، بلکه در حاشیه آن قرار دارد و بنابراین باید بااحتیاط استفاده شود. تحلیل مؤلفه‌های اصلی نشان داد که ۹۲ درصد از واریانس‌های بین همبستگی‌های رتبه‌بندی نشریات ایجادشده 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 روی مقالات، متمرکزشده و گاهی نیز نشریات را هدف قرار داده‌اند. عمده تلاش‌ها نیز بر تشخیص رابطهٔ شاخص‌های استنادی سنتی و دگر سنجه‌ها و توان پیش‌بینی آن‌ها از یکدیگر بوده است. حوزه‌های موضوعی متفاوتی نیز موردمطالعه قرارگرفته است. پژوهش حاضر نیز تلاش کرد تأثیر شاخص‌های مختلف اثرگذاری بروندادهای علمی حوزهٔ زیست‌فناوری ایران در یک بازهٔ ۲۲ ساله را بر هم بررسی کند و از الگوریتم‌های متفاوتی برای پیش‌بینی تأثیرات آن‌ها استفاده کند.

روش‌شناسی پژوهش

پژوهش حاضر از نظر هدف، کاربردی و از لحاظ روش، توصیفی بوده که با رویکرد علم‌سنجی انجام‌شده است و به دنبال بررسی همکاری علمی و نیز تأثیرگذاری علمی، اجتماعی و اقتصادی آثار علمی ایران در حوزهٔ زیست‌فناوری است. از آنجایی که این پژوهش در نظر دارد نشان دهد که کدام‌یک از الگوریتم‌های یادگیری ماشین، پیش‌بینی کنندهٔ بهتری برای سنجش اثرگذاری بروندادهای علمی در ابعاد مختلف هستند، جامعهٔ آن، شامل بروندادهای حوزهٔ زیست‌فناوری ایران است که در پایگاه اسکوپوس در بازهٔ ۲۰۰۳–۲۰۲٤ نمایه شده‌اند. تاریخ استخراج داده‌ها نیز ۱ بهمن ۱٤۰۲ معادل ۲۱ ژانویه ۲۰۲٤ بود. به‌منظور استخراج داده‌های مربوط از پایگاه تحلیلی سایول استفاده شد. پایگاه سایول که برای اولین بار در سال ۲۰۰۹ معرفی شد، امکان تجزیه‌وتحلیل‌های 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 تجزیه‌وتحلیل پژوهشی است که همچون این سایتس[1]، امکان مقایسه نتایج به‌دست‌آمده را با میانگین‌ها و استانداردهای جهانی فراهم می‌سازد. به‌این‌ترتیب، با استفاده از سایول می‌توان عملکرد پژوهشی یک پژوهشگر، سازمان، کشور، منطقه، نشریه، گروه پژوهشی و مانند آن‌ها را ارزیابی کرد (نوروزی چاکلی، ۱٤۰۲). متغیرهای موردمطالعه در قالب مؤلفه و شاخص‌های مورداستفاده برای آزمون همبستگی و نیز پیش‌بینی کنندهٔ تأثیرات علمی، اقتصادی و اجتماعی به شرح زیر هستند:

جدول ۱. متغیرهای موردمطالعه

عنوان متغیر
حجم برونداد علمی
همکاری فقط دانشگاهی (درصد)
تأثیر همکاری فقط دانشگاهی
تأثیر همکاری بین‌المللی

1 . InCites

ادامه جدول ۱. متغیرهای موردمطالعه

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

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 (درصد)
	برونداد‌ها در ۱۰ درصد استنادهای برتر (درصد)
تأثیر اقتصادی	تعداد استنادات ثبت اختراع
	استنادهای حق ثبت اختراع به ازای هر خروجی علمی
	خروجی علمی استناد شده توسط ثبت اختراعات (درصد)
	تعداد اختراعات
تأثیر اجتماعی	تعداد بازدیدها
	برونداد در ۱۰ درصد بازدیدهای برتر (درصد)
	بازدید به ازای هر خروجی
	تأثیر بازدید وزنی حوزه

گفتنی 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 رابطهٔ بین شاخص‌های موردمطالعه، استفاده‌شده است. در زمان بررسی همبستگی بین دو متغیر، اگر هر دو متغیر موردنظر 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 پیش‌بینی کننده، مورداستفاده و ارزیابی قرار گرفتند. به‌منظور انجام آزمون‌ها و الگوریتم‌ها از زبان برنامه‌نویسی پایتون استفاده‌شده است. تحلیل‌ها توسط زبان برنامه‌نویسی پایتون انجام‌شده که برای بررسی و نمایش صحیح آن باید 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 نبودند و مقداری برای آن گزارش نشده بود که برای انجام تحلیل‌ها باید مقادیری به آن‌ها نسبت داده می‌شد. این کار با استفاده از متد قابل‌قبول و معتبر knn Imputer انجام شد که این متد با در نظر گرفتن نزدیک‌ترین همسایگی‌ها اقدام به تکمیل داده‌ها می‌کند. همچنین، فونت داده‌های فارسی، قابل‌استفاده نبوده و تحلیل‌ها با استفاده از داده‌های انگلیسی انجام شد. در گام بعدی، مدل‌های یادگیری ماشین مختلفی روی داده‌ها مدل‌سازی و اجرا شد. مدل‌سازی در نرم‌افزار پایتون به‌گونه‌ای انجام‌شده که همه متغیرهای موجود در مجموعه داده، یک بار به‌عنوان مقدار هدف فرض شده‌اند و مدلی ساخته‌شده که بتواند بر اساس 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 و به‌راحتی برای مشکل طبقه‌بندی دوحالته و چندحالته مناسب است.

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

صحت یعنی نسبت نتایج واقعی به‌کل موارد بررسی‌شده. نمرهٔ صحت مدل درواقع معرف صحت مدل است که اگر این اعداد در ۱۰۰ ضرب شود بیانگر درصد صحت و دقت مدل است. نکته آنکه اعداد منفی، بیانگر دقت بسیار پایین بوده و نشان می‌دهند مدل به‌کاررفته برای پیش‌بینی آن پارامتر خاص، اصلاً مناسب نیست.

یافته‌های پژوهش

پاسخ به پرسش اول پژوهش: وضعیت بروندادهای حوزهٔ زیست‌فناوری ایران بر اساس هرکدام از شاخص‌های اثرگذاری علمی، اثرگذاری اجتماعی و اثرگذاری اقتصادی چگونه است؟

وضعیت بروندادهای علمی ایران در حوزهٔ زیست‌فناوری بر اساس داده‌های پایگاه سایول در بازهٔ زمانی ۲۰۰۳ – ۲۰۲٤ که مبنای تحلیل‌های بعدی قرار گرفت در جدول ۳ آمده است. نکته مهم در این زمینه، سیر صعودی و افزایش ۳٦ برابری حجم بروندادهای علمی ایران در این حوزه تا سال ۲۰۲۳ است که نشان از پیشرفت فوق‌العاده در این فاصله زمانی است. ضمن اینکه انتظار می‌رود کیفیت مقالات نیز هم‌راستا با رشد کمّی آن‌ها رشد کند.

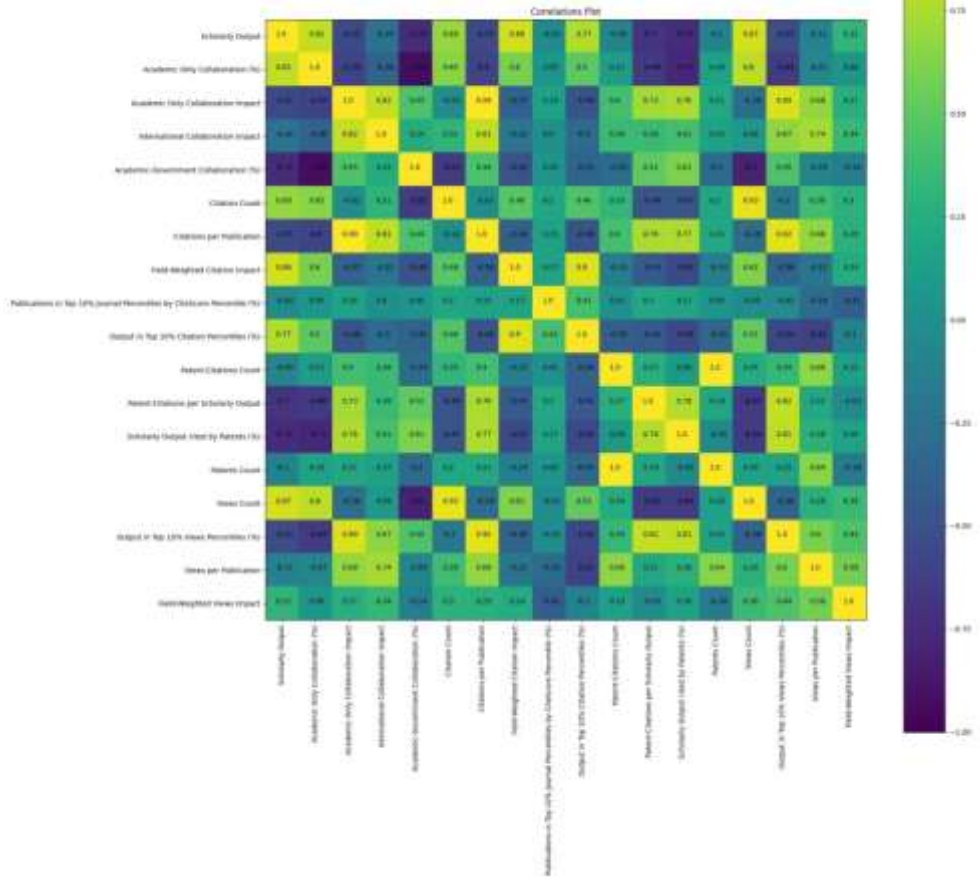
1 . Accuracy

جدول ۳. وضعیت بروندادهای علمی ایران در حوزهٔ زیست‌فناوری بر اساس داده‌های پایگاه سایول در بازه ۲۰۰۳ -۲۰۲٤

	عنوان شاخص	۲۰۰۳	۲۰۰٤	۲۰۰۵	۲۰۰۶	۲۰۰۷	۲۰۰۸	۲۰۰۹	۲۰۱۰	۲۰۱۱	۲۰۱۲	۲۰۱۳	۲۰۱٤	۲۰۱۵	۲۰۱۶	۲۰۱۷	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۲۱	۲۰۲۲	۲۰۲۳	۲۰۲٤
بروندادهای علمی	حجم برونداد علمی	۳۵	٤۵	٦۸	۱۰۳	۱٤۷	۲۸٤	۳۵۹	۵۸۹	۷٤۷	٤٦٤	۵٤٦	۷۳۳	۷۷٦	۹٦۵	۹۵۳	۱۱۲۹	۱۱۳۳	۱۱٦۹	۱۱۹۰	۱۲۲۹	۱۲٦۲	۷٤
همکاری	همکاری فقط دانشگاهی (درصد)	۷۱.٤	٦٤.٤	٦۰.۳	۷٦.۷	۷۰.۱	۷۷.۱	۷۵.۸	۷۹.۳	۸۲.۵	۷٦.۵	۷۹.۳	۸۰.۹	۸۳.۵	۸۲.۷	۸۲.٦۸۲	۸۲.۸۸۲	۸۲.۵٤	۸۲.۷۸٤	۸۳.۲۸۳	۸۳.۲	۷۷
	تأثیر همکاری فقط دانشگاهی	۸	۳۱	٤۰.۹	۲٤.۳	۳۷.۱	۳۹.٦	۲۷.٦	۲۰.۹	۱۹.۱	۲۲	۲۸.۲	۳۰.۷	۱۹	۲٦.۳	۲۷	۲۵.۹	۲۲.۸	۱۸.٤	۱۱.٦	٦.۷	۲	۰.۱
	تأثیر همکاری بین‌المللی	۲٦.۵	۵۳.۱	٤٦.۵	٤۱.٤	۳٤.۵	۳۰.۵	۲٦.۹	۲۹.٦	۳۲.۷	٦٤.۵	۲٦.۸	۳۳	۲۹.٦	٤۵.۱	٤۲.۵	٤۲.٤	٤۲.۹	۲۹.٦	۲۰.۵	۱۰.۸	۳.۱	۰.۲
	همکاری دانشگاهی و دولتی (درصد)	۱٤.۳	۲۰	۱۹.۱	۱۲.٦	۱۵.٦	۱۲.۳	۱۵.۳	۹.۸	۸	۱۳.٦	۱۲.۵	۱۰	۹.۷	۹.۲	۱۰.٤	۱۰.۵	۱۰.۷	۱۰.۱	۱۰.۵	۱۱.۲	۱۰.۷	۱٤.۹
تأثیر علمی	حجم استنادها	۱۱۱۱	۱٦۸٦	۲٤٦۱	٤۱۷۱	۵۱۷۲	۷٤۵۸	۷۸۷۲	۱۳۱٦۷	۱٤۹۹٦	۱۳۲٦۹	۱٤۲٦٦	۱٤۸۷۸	۱٤۷٤۱	۲۵۱۱۲	۲٤۵٤۱	۲۹۸۱۵	۲۵۲۷۳	۲۱۷۹۰	۱٤۷۰۵	۸٤۸۱	۲۵۵٦	۸
	استناد به ازای برونداد	۳۱.۷	۳۷.۵	۳٦.۲	٤۰.۵	۲٦.۳	۳۵.۲	۲۱.۹	۲۲.٤	۲٦.۱	۲۸.٦	۲۰.۱	۱۹.۲	۲٦	۲۵.۸	۲٦.٤	۲۲.۳	۱۸.٦	۱۲.٤	٦.۹	۲	۰.۱		
	تأثیر استناد وزنی حوزه	۰.۵۵	۰.۵۹	۰.۵۸	۰.۷۳	۰.۷٤	۰.٦۳	۰.۵۵	۰.۵۸	۰.۵۵	۰.۷۲	۰.۸۸	۰.۹	۰.۷۲	۰.۹٤	۰.۹۳	۱.۱۱	۱.۰٤	۱.۰۷	۱.۱	۱.۱۸	۱.۳٤	۰.٤٤	
	بروندادها در ۱۰درصد برتر مجله CiteScore	۱۲.۱	۷.۷	٤.٦	۱۱.۱	۸.۱	٦	۵.٤	۲.۷	۲.۷	٤.۲	۸	٦.۳	۵.۱	۱۲.۲	۹.۵	۱۱.۸	۹	۹.۲	۸.۱	٤.۹	۵.۸	۱۱	
	بروندادها در ۱۰درصد استنادهای برتر (درصد)	۱۱.٤	۸.۹	۱۳.۲	۱۳.٦	۱٤.۳	۱۱.٦	۸.٦	۸	۸.۸	۱۳.۱	۱۲.۸	۱۱.۳	۱۲	۱۷.۷	۱۸.۳	۲۱.۲	۲۱.٦	۲۲.۲	۲۲.٤	۲۲.۹	۲۰.۵	-	
تأثیر اقتصادی	تعداد استنادات ثبت اختراع	۳۳	۳۳	٦۹	۱۷۰	۹٤	۹۰	۷۷	۱۹۵	۱٤۸	۷٤	۲۵۵	۱۵۳	٦۱	۲۵۰	۱۱۲	۷۹	٤۲	۳٦	۲۵	۳	۰	۰
	استنادهای حق ثبت اختراع به ازای هر خروجی	۹٤۲.۹	۷۳۳.۳	۱۰۱٤.۷	۱٦۵۰.۵	٦۳۹.۵	۳۱٦.۹	۲۱٤.۵	۳۳۱.۱	۱۹۸.۱	۱۵۹.۵	٤٦۷	۲۰۸.۷	۷۸.٦	۲۵۹.۱	۱۱۷.۵	۷۰	۳۷.۱	۳۰.۸	۲۱	۲.٤	۰	۰	
	خروجی علمی استناد شده توسط ثبت اختراعات	۳۱.٤	۲۰	۱۷.٦	۱٤.٦	۱۸.٤	۱۰.٦	۹.۲	۸.۱	۵.۸	۸	۸.۲	٦.۷	٤.۸	٦.۷	٦.۱	٤.۷	۲.٦	۱.۹	۱.۸	۰.۲	۰	۰	
	تعداد اختراعات	۳۳	۳۳	٦۹	۱۷۰	۹٤	۹۰	۷۷	۱۹۵	۱٤۸	۷٤	۲۵٤	۱۵۳	٦۱	۲٤۸	۱۱۰	۷۷	٤۲	۳٦	۲۵	۳	-	-	
تأثیر اجتماعی	تعداد بازدیدها	۱۷۳٦	۲۸۰۳	٤۰۱۰	٦٤۵۷	۱۰۰۲۳	۱۹٦۷۷	۲۵۸۷٤	٤٤۸۱٦	۵۸۰۵٦	۳۸۱۷۰	٤۳۲۸٤	٤۷٤۵۹	٤۹۵۷٤	٦٦٤۰۳	٦۳٦۰۲	۷۵٦٦۵	۷۰۵۵۲	۵۷۲۱۳	۵۱۳۰۸	٤۷۰٦٤	۳۵۲۷٦	۵۳۲
	برونداد در ۱۰درصد بازدیدهای برتر (درصد)	٦۵.۷	٦٤.٤	۷۲.۱	۷۰.۹	۵۵.۸	۵۳.۵	۵۱.۳	۵۰.۱	۵۱.٤	۵۱.۳	۵۱.۳	٤۰.۵	۳۹.٦	٤۸.۹	٤۸.۱	٤۷.۳	٤۲.٤	۳۷.٦	۳٦.۱	۳۳.۹	۱۳.۵		
	بازدید به ازای هر خروجی	٤۹.٦	٦۲.۳	۵۹	٦۲.۷	٦۸.۲	٦۹.۳	۷۲.۱	۷۷٦.۱	۷۷.۷	۸۲.۳	۷۹.۳	٦٤.۷	٦۳.۹	٦۸.۸	٦٦.۷	٦۷	٦۲.۳	٤۸.۹	٤۳.۱	۳۸.۳	۲۸	۷.۲	
	تأثیر بازدید وزنی حوزه	-	۱.۷٤	۱.۵۱	۱.۳۱	۱.۵۱	۱.٦٤	۱.٦	۱.٦٦	۱.٤۷	۱.۳۸	۱.۳۳	۱.۵۳	۱.۵٦	۱.۷۲	۱.۸۱	۱.۷	۱.۷۷	۱.٦۷	۱.٦۹	۰			

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 نشان می‌دهد:.



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

داده‌های نمودار ۱ نشان می‌دهد هرکدام از متغیرها با متغیرهای دیگر چه رابطه‌ای باهم دارند. عدد گزارش‌شده هرچقدر به ۱+ و یا ۱- نزدیک‌تر باشد گویای این است که رابطه بیشتری بین دو متغیر وجود دارد و هرچقدر به صفر نزدیک‌تر باشد نشان می‌دهد که آن دو متغیر رابطه خطی با یکدیگر ندارند. اعداد مثبت نشان می‌دهند اگر یکی از متغیرها افزایش یابد متغیر دیگری نیز افزایش پیدا می‌کند. اعداد منفی نشان می‌دهند اگر یکی از متغیرها افزایش پیدا کند متغیر دیگر کاهش می‌یابد. بدیهی است 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 حوزهٔ زیست‌فناوری با حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، بروندادها در ۱۰درصد استنادهای برتر، تعداد بازدیدها و تأثیر بازدید وزنی حوزه، رابطهٔ مثبت و معنی‌داری داشته است.

همکاری بین‌المللی با حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰درصد برتر نشریات، 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 ۱۰درصد 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 ۱۰درصد برتر نشریات، بروندادها در ۱۰درصد 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: کدام‌یک از شاخص‌های اثرگذاری علمی، اجتماعی و اقتصادی با یکدیگر ارتباط دارند؟

در مقوله اثرگذاری علمی بروندادهای موردمطالعه، 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 ۱۰درصد 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 ۱۰درصد 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 مورداستفاده در اثرات علمی، فناورانه و اجتماعی بروندادهای علمی در جدول ٤ آمده است.

در جدول ٤، شاخص هدف، نمرهٔ صحت و نمرهٔ انحراف معیار ٤ مدل پیاده شده در این مطالعه، آمده است. گفتنی است هرچقدر نمرهٔ صحت اجرای مدل بالاتر و نمرهٔ انحراف معیار حاصل‌شده پایین‌تر باشد، قدرت پیش‌بینی‌کنندگی الگوی پیاده شده بیشتر خواهد بود. اعداد منفی، بیانگر دقت بسیار پایین بوده و نشان می‌دهد مدل به‌کاررفته به‌منظور پیش‌بینی یک متغیر خاص، مناسب نیست. بر اساس نتایج حاصل‌شده، رگرسیون خطی چند متغیره با نمرهٔ صحت بالاتر و نمرهٔ انحراف معیار پایین‌تر، بهتر توانسته است میزان اثرگذاری علمی، فناورانه و اجتماعی بروندادهای علمی ایران در حوزهٔ زیست‌فناوری را پیش‌بینی کند.

جدول ٤. 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	۰.٤٤	۰.۵۷
	اثر فناورانه	−۰.۰۷	۰.۹۹
	اثر اجتماعی	۰.۰۸	۰.۸٦
درخت تصمیم	اثر علمی	−۰.۸۸	۰.۲٤
	اثر فناورانه	۰.٤٤	۰.۲٦
	اثر اجتماعی	−۰.٦٦	۱.۷۸
جنگل تصادفی	اثر علمی	−۰.٤۱	۰.۸٦
	اثر فناورانه	−۵۱.۳۰	٦۵.۹٦
	اثر اجتماعی	۰.٦۸	۰.۱۵
رگرسیون خطی چند متغیره	اثر علمی	۰.۷۸	۰.۱۳
	اثر فناورانه	۰.۵۲	۰.۲۸
	اثر اجتماعی	۰.٦۸	۰.۱۵
تقویت گرادیان	اثر علمی	−۰.۷۰	۰.۹٦
	اثر فناورانه	۰.۹۳	۰.۰٤
	اثر اجتماعی	۰.۵۱	۰.۵۹

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 رابطهٔ هر یک از متغیرهای مختلف اثرگذاری بر یکدیگر و نیز ارزیابی الگوریتم‌های ماشین که بتوانند اثرگذاری علمی، اجتماعی و اقتصادی بروندادهای علمی را پیش‌بینی کنند، انجام گرفت. به این منظور، بروندادهای حوزهٔ زیست‌فناوری نمایه شده ایران در پایگاه اسکوپوس در دورهٔ ۲۰۰۳–۲۰۲٤ از ابعاد مختلف، ارزیابی و تحلیل شد.

یافته‌های این مطالعه نشان داد که بروندادهای علمی ایران در این حوزه در بازه زمانی ۲۰۰۳–۲۰۲۳، سیر صعودی داشته و افزایش ۳٦ برابری را تجربه کرده است که پیشرفت فوق‌العاده‌ای در این فاصله زمانی محسوب می‌شود. شکل‌گیری ستاد توسعه زیست‌فناوری در ایران و تصویب سند مربوط به این حوزه و درنتیجه افزایش حمایت‌ها از حوزهٔ 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 بُعد زمان و انجام مطالعه جداگانه‌ای خواهد بود.

بر اساس داده‌های حاصل‌شده، حجم بروندادهای علمی ایران در حوزهٔ زیست‌فناوری بین حجم استنادها، استناد به ازای برونداد، تأثیر استناد وزنی حوزه، بروندادها در ۱۰درصد استنادهای برتر، تعداد بازدیدها و تأثیر بازدید وزنی حوزه، رابطهٔ مثبت و معنی‌داری برقرار بوده است. از این نظر می‌توان تااندازه‌ای گفت در کنار رشد کمّی، شاخص‌های

سنتی استناد و نیز میزان رؤیت پذیری در شبکه‌های علمی نیز ارتقا پیداکرده‌اند. درعین‌حال، 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 حوزهٔ زیست‌فناوری ایران وجود دارد. شاید بتوان وجود موانع متعدد در مسیر کاربردی کردن پژوهش‌ها، حمایت‌های ناکافی دولتی و عدم انگیزهٔ پژوهشگران برای تجاری‌سازی یافته‌های خود را از دلایل این موضوع برشمرد.

رابطه بین همکاری بین‌المللی با شاخص‌هایی ازجمله حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰ درصد برتر نشریات، 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. به‌این‌ترتیب، این شاخص بر بسیاری از موارد، اثرگذار بوده و درواقع مهم‌ترین عامل مؤثر بر کیفیت مقالات ازجمله در بُعد 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 بااینکه درصد این نوع همکاری بالا بوده اما حمایت‌های دولت در مسیر اثرگذاری اقتصادی پژوهش‌های این حوزه، مؤثر افتاده به‌نحوی‌که تبادل استناد بیشتری بین مقالاتی که با همکاری نهادهای دولتی منتشرشده و پروانه‌های ثبت اختراع رخ‌داده است. شاخص همکاری دانشگاهی نیز رابطه مثبت و معنی‌داری با حجم استنادها، تأثیر استناد وزنی حوزه، بروندادها در ۱۰ درصد برتر نشریات، 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. به‌این‌ترتیب، عمدهٔ نقطه اثرگذاری مقالات، در بخش استنادهای علمی و مطالعات سایر پژوهشگران هم‌تراز تبلوریافته است. ضمن اینکه حوزه کاربرد و جامعه را نیز متأثر ساخته است.

همچنین در عرصه اثرگذاری خروجی‌های علمی موردمطالعه، 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 پر استناد، بهتر عمل می‌کند (Hansen, 2022). ایوب، امین و وانی در مطالعه خود دو مقولهٔ استناد و نمرهٔ دگر سنجی 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 نتیجه رسیدند که همبستگی مثبتی بین دگر سنجی و امتیاز استناد و همچنین نمرهٔ دگر سنجی 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 دگر سنجی نشان داد که رابطه قوی بین استنادات پلام‌ایکس و تعداد استنادات اسکوپوس برقرار است.

در حوزهٔ شاخص‌های مرتبط با تعامل بین بروندادهای علمی با پروانه‌های ثبت اختراع نیز این نتیجه حاصل شد که تعداد استنادات ثبت اختراع به‌عنوان 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، رابطهٔ مثبت و معنی‌داری داشته است. درواقع، کاربردی شدن مقالات و پژوهش‌های حوزهٔ 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، رابطهٔ مثبت و معنی‌داری برقرار کرده است. نکتهٔ قابل‌تأمل برای سیاست‌گذاران عرصه علم و فناوری در کشور این است که برای توسعهٔ علمی ایران، علی‌رغم رشد کمّی و کیفی تولیدات علمی، نظام پژوهشی کشور باید به سمت ارزش‌آفرینی و ایجاد ارزش‌افزوده به‌ویژه در حوزهٔ اقتصادی سوق داده شود.

در مورد شاخص‌های دگر سنجی نیز این نتیجه به دست آمد که تعداد بازدیدها رابطهٔ 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. بر اساس 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 کرد.

یکی دیگر از پرسش‌هایی که این مطالعه درصدد پاسخ به آن بود این است که کدام‌یک از الگوریتم‌های ماشین می‌توانند اثرگذاری علمی، اجتماعی و اقتصادی را دقیق‌تر پیش‌بینی کنند. بر اساس نتایج به‌دست‌آمده، الگوی رگرسیون خطی چند متغیره با نمرهٔ صحت بالاتر و نمره انحراف معیار پایین‌تر، بهتر توانست میزان اثرگذاری بروندادهای علمی ایران در حوزهٔ زیست‌فناوری را در ابعاد علمی، فناورانه و اجتماعی پیش‌بینی کند. آلکوهر و همکاران نیز از چارچوب یادگیری ماشین برای آزمون توانایی دگر سنجه‌ها 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 طبقه‌بندی کننده‌ها، 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 به‌دقت و توانایی پیش‌بینی استنادات مقالات پر استناد، 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 این مطالعه، موارد زیر قابل پیشنهاد است:

- نظر به اینکه مدل رگرسیون خطی چند متغیره، قدرت پیش‌بینی‌کنندگی بالاتری نسبت به سایر مدل‌ها داشت، پیشنهاد می‌شود مدیران، سیاست‌گذاران و متخصصان علم‌سنجی که به دنبال 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 حوزهٔ علم و فناوری کشور توصیه می‌شود زمینه‌های مشارکت پژوهشگران داخلی با پژوهشگران بین‌المللی را فراهم سازند و مشوق‌هایی را برای همکاری‌های بین‌المللی پیش‌بینی کنند.
- 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 بازدیدهای فضای مجازی در سایر شاخص‌های کیفیت و اثرگذاری علمی، پیشنهاد می‌شود پژوهشگران در شبکه‌های اجتماعی علمی فعال باشند و دستاوردها و یافته‌های پژوهش‌های خود را از طریق این سکوها با سایرین به اشتراک بگذارند تا رؤیت پذیری پژوهش‌ها بیشتر شود.
- با توجه به شکاف زیاد بین مقالات علمی و پژوهش‌های کاربردی حوزهٔ 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 برنامهٔ مطالعاتی پژوهشگران بعدی قرار گیرد:

- استفاده از شاخص‌های به کار رفته در مطالعهٔ حاضر برای حوزه‌های علمی متفاوت.
- تکرار این مطالعه در سال‌های بعد و مقایسه نتایج حاصل‌شده.
- مقایسه آمار و اطلاعات اثرگذاری پژوهش‌های علمی ایران در مقایسه با کشورهای اسلامی، منطقه، خاورمیانه یا جهان.
- استفاده از الگوریتم‌های مورداستفاده در این مطالعه برای پیش‌بینی اثرگذاری پژوهش‌ها دریکی 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 حوزهٔ 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 پیش‌بینی کننده تأثیرگذاری علمی برای افزایش استناد گیری 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.https://nrisp.ac.ir/%D8%A7%D8%B1%D

آزادی احمدآبادی، ق. (۱۴۰۳). سطوح و شاخص‌های ارزیابی تأثیرات پژوهش بر اساس تحلیل نظام‌های ارزیابی. *ترویج علم*. ۱۵(۱)، ۷۶-۱۰۳. https://doi.org/10.22034/popsci.2024.424371.1306

آزادی احمدآبادی، ق.، عبدی، س.، و رمضانی، ا. (۱٤۰۱). مطالعه تأثیرات علمی، اقتصادی و اجتماعی پژوهش‌های حوزه محیط‌زیست ایران. *محیط‌زیست و توسعه فرا بخشی*، ۷(۷۸)، ۵-۳۸.
https://doi.org/10.22034/envj.2023.351434.1217

باباکبری ساری، ا.، قهرمانی، م.، فتحی واجارگاه، ک.، و مؤتمنی، ع. (۱۳۹٦). ارائه الگوی ارزشیابی اثرات پژوهش‌های مدیریتی. *پژوهش‌های مدیریت در ایران*، ۲۱(۱)، ۹۳-۱۱۹. https://mri.modares.ac.ir/article_418.html

بذرافشان، ا.، بیرانوند، ع.، و شجاعی فرد، ع. (۱٤۰۲). پیش‌بینی تعداد استنادات دریافتی حوزه فیزیک ذرات در اسکوپوس به کمک نمرات دگر سنجی پلام‌ایکس. *فصلنامه بازیابی دانش و نظام‌های معنایی*، [انتشار آنلاین از ۱٦ خرداد]. https://doi.org/10.22054/jks.2023.71392.1551

بیرانوند، ع.، گلشنی، م.، و دلقندی، ف. (۱٤۰۱). بررسی تأثیر شاخص‌های سایت اسکور، اس‌ان‌آی‌پی و اس‌جی‌آر نشریات حوزه وب‌معنایی بر تعداد استنادات دریافتی مقالات. *فصلنامه بازیابی دانش و نظام‌های معنایی*، ۱۲(٤۲).
https://doi.org/10.22054/jks.2022.67616.1501

نوروزی چاکلی، ع. (۱٤۰۲). *آشنایی با علم‌سنجی (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. *arXiv*. 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 Prediction Models [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*, 1-12. https://doi.org/10.1155/2022/2239152

Anninos, L. N. (2013). Research performance evaluation: Some critical thoughts on standard bibliometric indicators. Studies in Higher Education, 39,(9) 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). https://nrisp.ac.ir/%D8%A7%D8%B1%/ **[In Persian].**

Azadi, G. (2024). Evaluation research impacts: levels and indicators. *Journal of the Popularization of Science, 15(1), 76-103.* 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://mri.modares.ac.ir/article_418.html **[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*, [Available online from 6 June]. https://doi.org/10.22054/jks.2023.71392.1551 **[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/info8030073

Bernard, 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*, 12(42). 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 *(Version 2). arXiv*. https://doi.org/10.48550/arXiv.2402.08640

Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S., & Grant, J. (2013). Measuring research, A guide to research evaluation frameworks and tools, RAND Corporation, MG-1217-AAMC, 2013. https://www.rand.org/pubs/monographs/MG1217.html

Hansen, I. S., & Torvund, M. (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. (2009). The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics, 337-387. https://www.sas.upenn.edu/~fdiebold/NoHesitations/BookAdvanced.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*(2), 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://link.springer.com/chapter/10.1007/978-3-642-40319-4_2

Paun, M., Abigaela, B., Paul, B., Anastasia, C., Catalina, E., Anne, H., Nicoleta, I., & Eduard, M. (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., Schroeder, J., Cohen, G., & 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/en/book/3376/introduction-to-scientometric **[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). 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 & amp; research*, *40*(3), 471-510. https://doi.org/10.1177%2F0049124111415372

Timilsina, M., Davis, B., Taylor, M., & Hayes, C. (2016). 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), (p 49–58). 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*, 244-207. 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. (2018). 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

Xiong, C., Sun, H., Pan, D., & Li, Y. (2019). Personalized Collaborative Filtering Recommendation Algorithm based on Linear Regression [Conference presentation]. In *2019 IEEE International Conference on Power Data Science (ICPDS)* (pp. 363–368). IEEE. https://doi.org/10.18280/mmep.060307

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.H., Steibach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and information systems*, *14(1)*, 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(2), 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. (2013). 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