

تاریخ دریافت: ۹۷ / ۱ / ۱۵

تاریخ پذیرش: ۹۷ / ۹ / ۲۵

کاربرد شبکه‌های عصبی مصنوعی در پیش‌بینی شاخص‌های کلان علم و فناوری

ایمان رئیسی وانانی^{۱*}نعیما میرزامومن^۲

چکیده

ارزیابی تحقیق و توسعه و ارتباط بین تولید علم و تکنولوژی در سطح کلان کشورها به دلیل حجم بالای اطلاعات و تغییر و تحولات سریع در این حوزه محدود بوده است. این پژوهش با هدف درک ارتباط و عملکرد توسعه فناوری در رابطه با فعالیت‌های تولید علم در سطح کشورها صورت پذیرفته است که از نوع تحقیقات توصیفی-کاربردی است. هدف ساخت مدلی با استفاده از الگوریتم‌های پیشرفته است که توانایی پیش‌بینی شاخص فناوری را بر مبنای تولید علم در کشورها داشته باشد. همچنین، تأثیر هر یک از شاخص‌های علمی بر شاخص فناوری با استفاده از روش آنالیز حساسیت شبکه عصبی تعیین گردید. روش تحقیق در این پژوهش CRISP-DM بوده و داده‌ها از پایگاه SJCR و سازمان جهانی مالکیت فکری (WIPO) و در بازه زمانی سال‌های ۲۰۰۱ تا ۲۰۱۵ استخراج گردید. طبق نتایج تحقیق، شبکه عصبی نسبت به رگرسیون از دقت و توانایی بیشتری جهت مدل‌سازی برخوردار بود و نتایج تحلیل حساسیت نشان داد که مهم‌ترین پارامتر علم‌سنجی جهت پیش‌بینی شاخص فناوری پارامتر اچ ایندکس و روند ارجاع دهی به مقالات بین‌المللی است. سیاست‌گذاران می‌توانند از نتایج پژوهش جهت شناسایی متغیرهای تأثیرگذار تولید علم که به ایجاد فناوری منجر می‌شوند، بهره‌گیرند.

واژه‌های کلیدی:

شاخص فناوری، شاخص علم‌سنجی، شبکه عصبی مصنوعی، آنالیز حساسیت، داده‌کاوی.

۱. عضو هیئت‌علمی گروه مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران.

* نویسنده عهده‌دار مکاتبات: jsoltanzadeh@yahoo.com

۲. کارشناسی ارشد، مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران.

مقدمه

شاخص‌های علم و نوآوری، عناصر اصلی برای مشخص کردن جایگاه کشورهای جهان از دیدگاه توسعه علم و فناوری و مبنایی برای مقایسه‌های بین‌المللی هستند. اختراعات ثبت شده از برودادهای مهم فرایند تحقیق و توسعه و منبعی منحصر به فرد برای تحلیل فرایند تغییر فناوری هستند. همچنین، تولیدات علمی کشورهای جهان که در پایگاه‌های معتبر علمی مانند اسکوپوس و وب آو ساینس نمایه می‌شوند از معتبرترین راه‌های بررسی توجه به پژوهش در کشورهاست. زمانی که بین صنعت و دانشگاه ارتباط تنگاتنگی وجود داشته باشد پژوهش‌های انجام شده در عرصه عمل قرار می‌گیرند و می‌توانند به رشد کشورها کمک کنند و نتایج حاصل از تولیدات علمی به کار گرفته می‌شود. تولید علم زمینه‌ساز ایجاد فناوری و در نتیجه باعث تولید اشتغال و ثروت در جامعه می‌گردد. امروزه یکی از مهم‌ترین چالش‌های کشورهای، رسیدن به قدرت رقابتی بالاتر در سطح بین‌المللی است. آنچه مشخص است این است که برخورداری از تولید ناخالص داخلی بالا به تنهایی ضامن حفظ و ارتقاء موقعیت رقابتی یک کشور نیست بلکه عوامل دیگری از جمله دستیابی به علم و تکنولوژی در کسب موقعیت رقابتی مؤثرند. کشورها برای بازبینی عملکرد گذشته خود در راستای توسعه ملی و طرح سیاست‌گذاری‌های آینده برای ابعاد مختلف جامعه ناگزیرند بر آوردی واقع‌بینانه از تولید علم و فناوری داشته باشند.

عمده‌ترین شاخص تولید علم در سطح جهان، تعداد مقالات علمی نمایه شده و نیز استنادهای مرتبط با آن است. هر یک از این شاخص‌ها اهمیت خاص خود را در سنجش داده‌های علمی دارد. تعداد مقالات نمایه شده بیانگر رشد کمی تولیدات علمی و استنادهای مرتبط با آن نشانگر اثربخشی مقاله چاپ شده و سطح کیفی آن است (سلطانی نژاد، ۱۳۹۳).

اختراعات از مصادیق بارز مالکیت صنعتی بوده و ثبت آن‌ها در قالب پتنت صورت می‌گیرد. در واقع پتنت عبارت است از یک حق انحصاری که به یک مخترع اعطا می‌گردد. آمار اختراعات معیاری برای اندازه‌گیری منابع سرمایه‌گذاری شده جهت تحقیق و توسعه است. آنچه مسلم است، کیفیت اختراعات لزوماً در یک سطح نیست و تمام اختراعات اهمیت یکسانی از لحاظ نوآوری فناورانه و سود اقتصادی ندارند با این وجود، در سطح مقیاس گسترده مانند کشورها، مقایسه این آمار صورت گرفته و به‌عنوان یک منبع اطلاعاتی مفید جهت اندازه‌گیری نوآوری به کار می‌رود (اکوبو، ۱۹۹۷).

یکی از موضوعات چالش‌برانگیز جهان امروز این است که با وجود غنای داده در زمینه‌های مختلف

با فقر دانش روبرو شده‌ایم؛ بنابراین ضرورت کشف و استخراج سریع و دقیق دانش از پایگاه‌های داده بیشتر از گذشته نمایان شده است و نیاز به طراحی سیستم‌هایی که قادر به اکتشاف سریع اطلاعات مورد نیاز محققان با تأکید بر حداقل مداخله انسانی باشند و روی آوردن به روش‌های تحلیل متناسب با حجم داده‌های حجیم به‌خوبی احساس می‌شود. هدف کلی اکتشاف دانش، استخراج اطلاعات بالقوه مفید و ناشناخته از داده‌ها است (فایاد و همکاران، ۲۰۰۵). برای استخراج نتایج رضایت‌بخش و مناسب به داده‌ها با کیفیت بالا نیاز است (بایکال^۱، ۲۰۱۲). پایگاه داده‌های حاوی داده‌های علم‌سنجی نقش مهمی را در علم مدرن امروزی دارند. تعداد استنادها و دیگر داده‌های استخراج شده از این پایگاه‌ها به‌عنوان شاخص‌هایی برای تأثیر فعالیت‌های تحقیقاتی مورد استفاده قرار می‌گیرند (رادیکی و کاستلانو^۲، ۲۰۱۳).

با توجه به موارد ذکر شده پرسش‌های ذیل مطرح می‌شوند:

- با توجه به اهمیت داده‌ها و منابع مورد استفاده جهت انجام پژوهش، مناسب‌ترین منابع جهت دستیابی به اهداف تحقیق کدام است؟
- چگونه می‌توان با استفاده از قابلیت مدل‌های شبکه‌های عصبی مصنوعی و رگرسیون، شاخص فناوری کشورها را با توجه به شاخص‌های علم‌سنجی پیش‌بینی نمود؟
- کدام شاخص تولید علم در سطح کشورها اهمیت نسبی بیشتری جهت پیش‌بینی فناوری دارد؟

نوآوری این پژوهش در ایجاد ارتباط بین شاخص‌های تولید علم و فناوری با استفاده از تکنیک‌های نوین تحلیل داده‌ها است. نتایج حاصل از تحقیق، در پیش‌بینی شاخص‌های کلان و روند آتی کشورها در زمینه تولید علم و فناوری بسیار کاربردی خواهد بود و نشانگر امکان‌پذیری استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی دقیق روند پیشرفت علم و فناوری کشورها طی سالیان متمادی است. یکی از نکات که باید توسط سیاست‌گذاران مورد توجه قرار گیرد این است که روی چه شاخص‌هایی از تولید علم توجه و سرمایه‌گذاری کنند که تأثیر بیشتری روی شاخص فناوری داشته باشد و زمینه‌ساز ایجاد فناوری در سطح کلان کشورها گردد. با توجه به مطالعات و تحقیقات صورت گرفته جهت پاسخ به سؤالات مذکور، از الگوریتم‌های شبکه عصبی مصنوعی و رگرسیون برای مدل‌سازی ارتباط بین پارامترهای علم‌سنجی و فناوری استفاده گردید و نتایج مورد ارزیابی و مقایسه قرار گرفتند. سپس،

1 . Baykal

2 . Radicchi & Castellano

روش آنالیز حساسیت شبکه عصبی جهت مشخص نمودن پارامترهای علمی که بیشترین تأثیر را در میزان پارامتر فناوری دارند به کار گرفته شد.

پیشینه پژوهش

مالکیت فکری یک ابزار قدرتمند جهت توسعه اقتصادی و تولید ثروت است اما هنوز به صورت بهینه‌ای در تمام کشورها به کار گرفته نشده است، به طوری که برای بسیاری از ساکنان کشورهای در حال توسعه یک مفهوم حقوقی مبهم و گمنام است (ایدریس، ۲۰۰۲). مالکیت فکری به دو دسته مالکیت ادبی و هنری و مالکیت صنعتی تقسیم‌بندی می‌شود. مالکیت ادبی و هنری شامل مواردی چون آثار نوشتاری، تجسمی و صوتی است و مالکیت صنعتی دربرگیرنده مواردی مانند علائم تجاری، اختراعات و جلوگیری از رقابت ناعادلانه است. سازمان جهانی مالکیت فکری^۱ چهاردهمین سازمان تخصصی سازمان ملل متحد است که در جهت نظارت بر اجرای صحیح معاهدات و کنوانسیون‌های بین‌المللی و نیز برخی کارکردهای دیگر فعالیت می‌کند (نقی پور، ۱۳۸۵). تحلیل ثبت اختراعات و نوشتارهای علمی یکی از روش‌های مفید پایش بر تکنولوژی است. کاربرد آن، ردیابی تکنولوژی، تصمیم‌گیری‌های سرمایه‌گذاری و تعیین رقبای جدید است (بهرامی، ۱۳۷۴). همچنین با در نظر گرفتن خروجی تولید علم و مقالات در ژورنال‌ها، کشورها به گروه‌های مختلفی تقسیم می‌شوند (آنگون و سولانا، ۲۰۱۳). سیاست‌گذاران، مدیران دانشگاهی و دانشمندان همواره به دنبال شاخص‌هایی جهت ارزیابی تحقیقات هستند و سؤال اصلی این است که چگونه تأثیر و اهمیت خروجی علم و فناوری اندازه‌گیری می‌شود (سوجیموتو و لاریویر^۲، ۲۰۱۸). علم‌سنجی یک روش سازمان یافته جهت ارزیابی روند گذشته، حال و آینده پیشرفت علمی است (زاینب و اشرف وانی^۳، ۲۰۱۸). روش‌ها و مدل‌های مختلفی جهت درک و ارزیابی سیستم‌های علم و فناوری ایجاد شده است که بر جنبه‌های متفاوت توسعه و عملکرد شاخص‌ها تمرکز دارد (گلانزل و همکاران^۴، ۲۰۱۸). با این وجود، بینش کافی در تفسیر شاخص‌های کلان علم و فناوری وجود ندارد، به طور قطع این شاخص‌ها مسئولیت‌های مختلفی به عهده دارند اما آن‌ها باید در ارزیابی سیستم‌های علم و تکنولوژی در ارتباط قرار گیرند (سیریلی^۵، ۱۹۹۹). سختی‌های تجزیه و تحلیل و تفسیر داده‌های

1 . WIPO

2 . Aneón & Solana

3 . Sugimoto & Lariviere

4 . Zainab & Ashraf Wani

5 . Glänzel et al.

6 . Sirilli

انبوه علم و فناوری منجر به کاهش توجه به آن در تحقیقات شده است.

فناوری یک بعد مهم توسعه و رشد ملی است و عنصری ضروری جهت بقا در محیط رقابتی است (سان ومون^۱، ۲۰۰۴). طبق تحقیقات نارین و همکاران^۲ (۱۹۹۷) یک رابطه در حال رشد بین علم و فناوری وجود دارد. انتقال مستقیم دانش می‌تواند به وسیله انعکاس استناد به مقالات علمی در اختراعات اتفاق بیفتد. پیش‌بینی تکنولوژی تلاشی است برای تصویر کردن قابلیت‌های فناوریانه و پیش‌بینی وقوع اختراعات و توزیع نوآوری‌های تکنولوژی (بهرامی، ۱۳۷۴). پیش‌بینی تکنولوژی اولین مرحله برنامه‌ریزی تکنولوژی است. به عبارت دیگر، هدف از پیش‌بینی تکنولوژی اتخاذ تصمیمات بهتر و برنامه‌ریزی مناسب برای آینده است. با پیش‌بینی تکنولوژی به خوبی می‌توان فرصت‌ها و جهت تغییرات تکنولوژی را شناسایی و تصویر واضح‌تری از آینده ترسیم کرد (پورتر^۳، ۱۹۹۱). در ادامه، تعدادی از تحقیقات صورت گرفته به شرح ذیل ارائه می‌گردند:

- تونیس (۱۹۹۲) معتقد است که در هر پیش‌بینی ابتدا باید منابع اطلاعاتی موجود را شناسایی کرد. او منابع اطلاعاتی را به چهار دسته مفروضات، آگاهی‌ها، داده‌ها و قضاوت‌ها تقسیم‌بندی نمود و بیان داشت که از تکنیک‌های اکتشافی و تجویزی می‌توان برای پیش‌بینی استفاده نمود. خروجی‌ها به چهار دسته کیفی، کمی، زمان و احتمال تقسیم می‌شوند.
- در گزارشی تحت عنوان «خروجی و کارایی علم و تکنولوژی در اروپا» به تولید علم و تکنولوژی در اروپا پرداخته شده است. با توجه به خروجی تکنولوژی نسبت به تولید علم، تولید تکنولوژی در اروپا کمتر رقابتی بوده است. در سال ۲۰۰۷ طبق آمار دفتر ثبت اختراعات اروپا تنها ۴۳ درصد درخواست‌های ثبت اختراع متعلق به اروپا بوده است. واگرایی بین تولید علم و تکنولوژی در اروپا نشانه وجود ضعف در سیستم تحقیق و نوآوری است.
- طبق یک پژوهش در سال ۲۰۱۰، فاکتورهای وابسته به جغرافیای سیاسی و فعالیت‌های علمی کشورها بر اساس داده‌های مقالات در یک دوره سی‌ساله از سال ۱۹۸۰ تا ۲۰۰۹ مورد بررسی قرار گرفتند. در این مقاله از روش‌های کتاب‌سنجی استفاده شده و بر مقیاس گسترده و جنبش‌های تولید دانش در جغرافیای سیاسی تمرکز داشته است. نتایج تحقیق نشان دادند که وقایع مختلف در سیاست‌های علمی کشورهای مختلف اثر می‌گذارند؛ همچنین از دیگر نتایج

1 . Sohn & Moon

2 . Narin et al.

3 . Porter

تحقیق رشد سریع تولید علم در خاورمیانه بوده است (تقریباً ۴ برابر سریع تر از سطح جهانی) که ایران و ترکیه در این زمینه پیشرو بوده‌اند و به طور خاص ایران در دو دهه اخیر شروع به افزایش قابلیت‌های علمی کرده است (ساینس متریکس^۱، ۲۰۱۰).

- در گزارش کمیسیون اروپا (۲۰۱۱) تحت عنوان «تولید و تأثیر علم و تکنولوژی در اروپا»، تولید علم و تکنولوژی و رابطه آن‌ها در اروپا در سال‌های مختلف مورد بررسی قرار گرفته‌اند. در این گزارش بیان شده که در سال ۲۰۰۹ اروپا ۳۳,۴ درصد از مقالات علمی جهان را تولید کرده است در حالی که ظرفیت اتحادیه اروپا برای تولید مقالات با ضریب تأثیر زیاد که نشان‌دهنده کیفیت علمی است پایین‌تر از ایالات متحده است. از سال ۲۰۰۱ اروپا کیفیت علمی خود را ارتقا داده که در این بین کشورهای دانمارک، هلند، ایسلند، بلژیک و سوئیس بالاترین کیفیت مقالات را داشته‌اند.

- سلطانی نژاد (۱۳۹۳) در پژوهشی به تحلیل رابطه میان تولیدات علمی کشورهای برتر جهان و میزان اختراعات ثبت شده در پایگاه اسکوپوس طی سال‌های ۲۰۰۸ تا ۲۰۱۳ پرداخت. در این تحقیق از روش تحقیق همبستگی استفاده شده و تعداد اختراعات ثبت شده از اداره ثبت اختراع و علائم تجاری آمریکا استخراج شده است. نتایج آزمون همبستگی اسپیرمن میان تولیدات علمی کشورها در حوزه‌های موضوعی و میزان اختراعات ثبت شده نشان داد که بین این دو عامل رابطه معنی‌داری وجود دارد.

- جنگ و کیم^۲ (۲۰۱۴) در مقاله‌ای با عنوان «تولید تحقیقات علمی، تکنولوژیکی و علوم زیستی در آسیا» تعداد مقالات ۱۲ کشور را در زمینه علمی و مهندسی در پنج سال گذشته مطالعه کردند. داده‌های مورد استفاده از پایگاه اسکوپوس استخراج شده بود. آن‌ها رشد مقالات، تعداد استنادها به ازای هر مقاله، نرخ همکاری‌های بین‌المللی و تعداد مجلات در هر کشور را مورد مقایسه قرار دادند. نتایج نشان داد که اغلب مقالات تحقیقاتی در کشورهای در حال توسعه آسیایی در زمینه تکنولوژی است؛ بنابراین بیشتر مقالات تحقیقاتی تولید شده در منطقه آسیا به نظر می‌رسد تعداد استناد کمتری را دریافت کنند.

- طبق یک پژوهش، مدلی برای انتقال تکنولوژی با استفاده از تجزیه و تحلیل اختراعات ارائه شده

1 . science-metrix

2 . Jang & Kim

است. این آنالیز جهت آشکار شدن رابطه کمی بین انتقال تکنولوژی و متغیرهای موجود در داده‌ها انجام شده است. داده‌های اختراعات مؤسسات تحقیقاتی ملی، دانشگاه‌های دولتی و خصوصاً، تحقیقات مشارکتی و کشورها جهت ساخت مدل مورد استفاده قرار گرفته‌اند. مدل پیشنهادی با استفاده از SNA، رگرسیون و درخت تصمیم ساخته شده است. طبق نتایج، می‌توان استراتژی‌هایی جهت توسعه فناوری و شاخص‌های اختراعات مربوطه ایجاد کرد (چوی و همکاران^۱، ۲۰۱۵).

- طبق پژوهش مادالانو و همکاران (۲۰۱۶) از تکنیک تحلیل پوششی داده‌ها^۲ (DEA) برای تجزیه و تحلیل و مقایسه کارایی ۲۶ کشور اروپایی استفاده شده است.
- در مقاله‌ای با عنوان «اندازه‌گیری بهره‌وری سیستم‌های ملی R&D: چالش‌های مقایسه ورودی R&D و شاخص‌های خروجی مقالات انتشار یافته» آمار تحقیق و توسعه رسمی برگرفته از OECD با داده‌های مقالات انتشار یافته از سایت وب آو ساینس شامل ۱۸ کشور مورد مقایسه و تجزیه و تحلیل قرار می‌گیرند. در این پژوهش، روش‌هایی که در آن‌ها آمار تحقیق و توسعه کشورها به‌عنوان داده ورودی و مقالات به‌عنوان داده خروجی در نظر گرفته شده‌اند مورد بررسی قرار می‌گیرند و هدف آن رسیدن به نتایجی در رابطه با بهره‌وری یا کارایی تحقیقات ملی و سیستم‌های نوآوری است. در این تحقیق، مسائل احتمالی در این نوع تجزیه و تحلیل در نظر گرفته شده و گام‌هایی را در جهت ارتقاء مقایسه این دو پایگاه داده پیشنهاد می‌دهد (اکسنس و همکاران^۳، ۲۰۱۷).
- در مقاله‌ای با عنوان «شاخص‌های علم، فناوری و نوآوری در گذار» به تجزیه و تحلیل اثرات سیاست‌های علم، تکنولوژی و نوآوری و شاخص‌های مربوط به آن‌ها در ارزیابی عملکرد تحقیق و توسعه در سطوح خرد و کلان پرداخته است. نتایج تحقیق نشان داد که در چهار کشور مورد مطالعه، سرعت هزینه‌های تحقیق و توسعه با سرعت انتشار مقالات علمی در تمام زمینه‌های دانش یکسان نبود. همچنین رشته‌های فیزیک، مواد، مهندسی و بیولوژی رشته‌هایی با بیشترین مقالات علمی در این کشورها بودند (ساندووال رومرو و همکاران^۴، ۲۰۱۸).

1 . Choi

2 . Data Envelopment Analysis

3 . Aksnes et al.

4 . Sandoval-Romero

بررسی ادبیات پژوهش نشان می‌دهد که تحقیقات صورت گرفته برای پیش بینی شاخص فناوری در سطح کشورهای جهان محدود بوده و اغلب از روش‌های تجزیه و تحلیل اختراعات و تجزیه و تحلیل استنادی با اتکا به روش‌های آماری استفاده شده است و پژوهش‌های مرتبط با روش انجام این تحقیق بسیار محدود بوده است. علاوه بر این، پژوهش‌های انجام شده اغلب تنها بر یک جنبه از تجزیه و تحلیل علم و یا فناوری تأکید داشته‌اند و توانایی کافی برای انعکاس فعالیت‌های تولید علم بر توسعه فناوری در سطح کشورها نداشته‌اند. با توجه به داده‌های بسیار زیاد در سطح کشورها و لزوم حداقل مداخله انسانی در تحلیل‌های بین‌المللی، استفاده از تکنیک‌های نوین داده‌کاوی در این پژوهش پیشنهاد شده است.

روش‌شناسی پژوهش

این پژوهش بر اساس هدف، یک تحقیق کاربردی بوده و بر اساس ماهیت و روش گردآوری داده‌ها توصیفی است زیرا محقق به دنبال کشف الگوهای پنهان میان داده‌ها و تشریح این الگوهاست. بدین منظور از روش تحقیق CRISP-DM استفاده شده است که از فازهای شناخت و درک مسئله، شناخت داده، پیش‌پردازش داده‌ها، مدل‌سازی، ارزیابی، تحلیل نتایج و توسعه تشکیل شده است. این روش انجام تحقیق بر مبنای استفاده از یک مدل استاندارد داده‌کاوی است که مراحل اجرای تحقیق با الگوگیری از این مدل به صورت گام به گام مورد بررسی قرار می‌گیرد. این مراحل با تعیین اهداف و شناسایی شاخص‌های مؤثر بر آن آغاز می‌شود. سپس به تناسب نیاز، داده‌ها جمع‌آوری و آماده‌سازی می‌گردد و با استفاده از تکنیک داده‌کاوی مناسب بر روی داده‌ها، مدل‌سازی انجام می‌شود. در فاز بعدی نتایج به دست آمده مورد ارزیابی و اعتبار سنجی قرار می‌گیرند و در نهایت در فاز آخر، نتایج به دست آمده تحلیل می‌گردند. اولین سؤال تحقیق که «مناسب‌ترین منابع جهت دستیابی به اهداف تحقیق کدام است» در بخش شناخت و درک مسئله و شناخت داده مورد بررسی قرار می‌گیرد. در مورد سؤال دوم و سوم تحقیق که «چگونه می‌توان با استفاده از قابلیت مدل‌های شبکه‌های عصبی مصنوعی و رگرسیون، شاخص فناوری کشورها را با توجه به شاخص‌های علم‌سنجی پیش‌بینی نمود و کدام شاخص تولید علم در سطح کشورها اهمیت نسبی بیشتری جهت پیش‌بینی فناوری دارد» در بخش مدل‌سازی و ارزیابی به آن پرداخته می‌شود.

شناخت و درک مسئله

در این تحقیق، ابتدا با مطالعات جامع کتابخانه‌ای، مرور ادبیات تحقیق، بررسی اسناد موجود و مطالعات

و تحلیل‌های انجام شده مرتبط با موضوع تحقیق به شناسایی شاخص‌های مؤثر جهت ارزیابی تولیدات علم و فناوری در مقیاس کشورهای جهان پرداخته شد. طبق بررسی‌های به‌عمل‌آمده مهم‌ترین دفاتر ثبت پتنت در جهان جهت استخراج شاخص فناوری عبارت‌اند از:

- دفتر ثبت اختراعات و علائم تجاری ایالات‌متحده^۱
- دفتر ثبت اختراعات اروپا^۲
- سازمان جهانی مالکیت فکری^۳
- دفتر ثبت اختراعات ژاپن

با توجه به لزوم جامعیت آمار اختراعات در بین همه کشورها سایت سازمان جهانی مالکیت فکری از بین دفاتر انتخاب شد. به‌عبارت‌دیگر، منظور از اختراعات ثبت شده در این تحقیق اختراعاتی هستند که در سازمان جهانی مالکیت فکری (WIPO) نمایه شده‌اند. این داده‌ها از پروفایل اطلاعات هر کشور به‌طور جداگانه استخراج شده است. مأموریت این سازمان ارتقاء، نوآوری و خلاقیت از طریق سیستم مالکیت فکری بین‌المللی مؤثر جهت توسعه اقتصادی، اجتماعی و فرهنگی تمام کشورها است. ایران در سال ۱۳۸۰ به عضویت سازمان جهانی حقوق مالکیت فکری (WIPO) درآمده است. داده‌های شاخص علمی کشورها از پایگاه SJCR استخراج شده است. پایگاه SJCR دربردارنده اطلاعات شاخص‌های علمی کشورها و نشریات از پایگاه اسکوپوس است.

شناخت داده

پس از شناخت مسئله داده‌کاوی، داده‌های مورد نیاز باید از منابع و مکان‌های مختلف جمع‌آوری و سپس در یک انباره داده مجتمع گردند. پایگاه داده‌ای شامل مجموعه متغیرهای مؤثر در ارزیابی علم و فناوری که معیارهای اساسی عملکرد کشورها در این حوزه می‌باشند در اکسل ایجاد گردید. رکوردهای این پایگاه داده کلیه کشورها بوده و فیلدهای آن شاخص‌های علم‌سنجی (چهار فیلد به‌عنوان ورودی مدل) و شاخص فناوری کشورها (یک فیلد به‌عنوان خروجی مدل) است. تعداد اختراعات ثبت شده کشورها به‌عنوان شاخص فناوری در نظر گرفته شده و تعداد مقالات، تعداد مقالات قابل استناد، خوداستنادی، متوسط استنادها به‌ازای هر مقاله و شاخص اچ - ایندکس کشورها به‌عنوان شاخص‌های

1 . United states patent and trademark office

2 . European patent office

3 . World intellectual property organization

علم‌سنجی در نظر گرفته شده‌اند. قلمرو زمانی داده‌ها بازه سال‌های ۲۰۰۱ تا ۲۰۱۵ است.

پیش‌پردازش داده‌ها

این مرحله شامل پاک‌سازی داده‌ها است. هدف این مرحله تأمین ورودی مناسب برای مرحله یادگیری مدل است. در شبکه‌های عصبی، در صورتی که داده‌ها به صورت خام وارد شبکه شود به دلیل تغییرات زیاد داده‌ها، تأثیر متفاوتی بر شبکه می‌گذارد به طوری که برخی از نرون‌ها خیلی زود به حد آتش می‌رسند در حالی که برخی دیگر از نرون‌ها حتی به آستانه فعالیت نیز نمی‌رسند و این باعث خواهد شد که توان پیش‌بینی الگو کم شود (منهاج، ۱۳۸۴). ابتدا داده‌ها و متغیرهایی که در راستای تحقق اهداف پژوهش مناسب بودند انتخاب شدند، سپس با استفاده از فرمول شوارتز ($\mu \pm 3\delta$) نقاط خارج از محدوده شناسایی و رکوردهای حاوی آن از مجموعه داده حذف شدند. داده‌های خارج از محدوده، رکوردهایی هستند که مقادیر فیلدهای آن نسبت به سایر رکوردها بسیار متفاوت است و با اکثریت داده‌ها همبستگی ندارند. در مرحله بعد داده‌ها با استفاده از فرمول کمینه-بیشینه نرمال شدند، در فرایند نرمال‌سازی داده‌ها تبدیل به مجموعه جدیدی می‌شوند که در آن همه مقادیر بین صفر و یک است. دلیل نیاز به نرمال‌سازی داده‌ها این است که تفاوت مقیاس‌ها در متغیرهای مختلف، نتایج داده‌کاوی را در جهات متفاوتی تحت تأثیر قرار می‌دهد. برای کاهش این تأثیر در نتایج داده‌کاوی، نرمال‌سازی داده انجام می‌شود. پس از حذف داده‌های خارج از محدوده، پایگاه داده‌ای شامل تعداد ۱۴۰۴ رکورد با ۵ فیلد به عنوان داده‌های ورودی تشکیل شد و تعداد ۱۴۰۴ رکورد با یک فیلد به عنوان داده خروجی اعمال شد.

مدل‌سازی

در این پژوهش جهت مدل‌سازی از مدل‌سازی پیش‌گویانه استفاده شده است. این قسمت به ساخت مدلی می‌پردازد که متغیر هدف را به صورت تابعی از سایر متغیرها نمایش می‌دهد (تان و همکاران^۱، ۲۰۰۶). در ساخت مدل با توجه به نوع داده‌ها و عملیات مورد نیاز از الگوریتم‌های شبکه عصبی مصنوعی و رگرسیون استفاده شده است. سپس آنالیز حساسیت مدل با استفاده از شبکه‌های عصبی مصنوعی انجام شده است.

الگوریتم شبکه عصبی مصنوعی

یک شبکه عصبی مصنوعی شامل عناصری است که در دو یا چندلایه قرار گرفته‌اند و به طور موازی عمل می‌کنند. این عناصر از سیستم‌های عصبی بیولوژیک (نرون‌ها) الهام گرفته‌اند. همانند طبیعت، عملکرد شبکه توسط ارتباطات بین این عناصر تعیین می‌شود. یک شبکه عصبی می‌تواند برای اجرای یک عملکرد خاص آموزش داده شود و این کار از طریق تنظیم مقادیر بین ارتباطات عناصر (وزن‌ها) صورت می‌پذیرد و در نتیجه، یک ورودی خاص منجر به خروجی هدف می‌شود (هدزیمایارکو و همکاران^۱، ۲۰۱۱).

شبکه‌های عصبی مصنوعی ابزار مناسبی برای رسیدن به راه‌حل بدون دانستن رابطه تحلیلی بین ورودی‌ها و خروجی‌ها هستند (ویلیمک^۲، ۲۰۱۴) و می‌توانند به منظور دسته‌بندی (پیش‌بینی برچسب کلاس رکوردها) یا پیش‌بینی عددی (پیش‌بینی خروجی‌هایی با مقادیر پیوسته) مورد استفاده قرار گیرند (تن و همکاران، ۲۰۰۶).

رگرسیون

پیش‌بینی مقدار یک متغیر پیوسته بر اساس مقادیر سایر متغیرها بر مبنای یک مدل وابستگی خطی یا غیرخطی، رگرسیون گفته می‌شود. روش‌های موجود در رگرسیون بیشتر مبتنی بر ریاضیات آماری هستند (هان و همکاران^۳، ۲۰۱۱).

آنالیز حساسیت مدل

یکی از راه‌های استخراج دانش از شبکه‌های عصبی، کسب اطلاعات در مورد اهمیت نسبی ورودی‌های شبکه با استفاده از آنالیز حساسیت هر یک از متغیرهای ورودی است (دی و همکاران^۴، ۱۹۹۷، پال^۵، ۱۹۹۹).

در این روش شبکه عصبی جهت تعیین اهمیت نسبی داده‌های ورودی اجرا می‌گردد و مشخص می‌کند که خروجی شبکه (مدل) به چه مقدار تحت تأثیر داده‌های ورودی انتخابی قرار می‌گیرد.

1 . Hadzima-Nyarko et al.

2 . Vilimek

3 . Han et al.

4 . De et al.

5 . Pal

این تکنیک اطلاعات مفیدی در رابطه با تأثیر هر یک از داده‌های ورودی مدل بر داده خروجی ارائه می‌دهد. تاکنون روش‌های گوناگونی جهت اندازه‌گیری اهمیت ورودی‌های شبکه عصبی پیشنهاد شده‌اند (سارل^۱، ۲۰۰۰)؛ که دو روش معمول آن عبارت‌اند از اهمیت پیش بینی (PI)^۲ و اهمیت علی (CI)^۳ در روش PI، افزایش خطای ایجاد شده در شبکه عصبی در اثر حذف یک ورودی، معیاری است که نشان‌دهنده اهمیت آن داده در شبکه است. در این تکنیک، یک داده ورودی از شبکه حذف می‌شود و شبکه دوباره آموزش می‌بیند. خطای شبکه با خطای به‌دست‌آمده زمانی که کل داده‌ها مورد استفاده قرار گرفته بودند مقایسه می‌شود. مهم‌ترین داده، داده‌ای است که غیبت آن بیشترین خطا را ایجاد کرده است.

در روش CI داده‌های ورودی دستکاری شده یا تغییر داده می‌شوند و تغییرات خروجی بررسی می‌شوند. در این تکنیک مشتق جزئی خروجی نسبت به ورودی گرفته می‌شود (هاشم^۴، ۱۹۹۲). مشتق جزئی نه تنها به وزن‌ها و بایاس‌ها بلکه به مقادیر متغیرهای ورودی بستگی دارد؛ بنابراین تعمیم دادن روند مقادیر خروجی با در نظر گرفتن یک تغییر در یک داده ورودی مشکل است (مولاس و یامازاکی^۵، ۱۹۹۵). در این تحقیق از روش PI جهت تحلیل حساسیت استفاده شده است.

به دلیل اینکه شبکه عصبی یک مدل غیرخطی است، تنها با یک بار آموزش و مدل‌سازی نمی‌توان اهمیت نسبی متغیرهای ورودی شبکه را تعیین نمود، به این منظور شبکه عصبی ۲۰ بار با وزن دهی‌های اولیه متفاوت آموزش دیده است.

ارزیابی

برای ارزیابی مدل به‌دست‌آمده ابتدا داده‌ها را به دو دسته آموزشی و آزمایشی تقسیم کرده، مدل را با استفاده از داده‌های آموزشی ایجاد و سپس نتایج به‌دست‌آمده را روی داده‌های آزمایشی آزمون می‌کنیم سپس دقت مدل را محاسبه می‌کنیم. البته در اجرای برخی الگوریتم‌ها نیاز است که داده‌ها به سه دسته آموزشی، اعتبارسنجی و آزمایشی تقسیم شوند. در نهایت الگوریتمی که نتیجه اعتبارسنجی آن بهتر از دیگر الگوریتم‌ها است شناسایی می‌شود. برای ارزیابی روش‌های پیش بینی (اختلاف بین

1 . sarle

2 . Predictive importance

3 . Causal importance

4 . Hashem

5 . Molas & Yamazaki

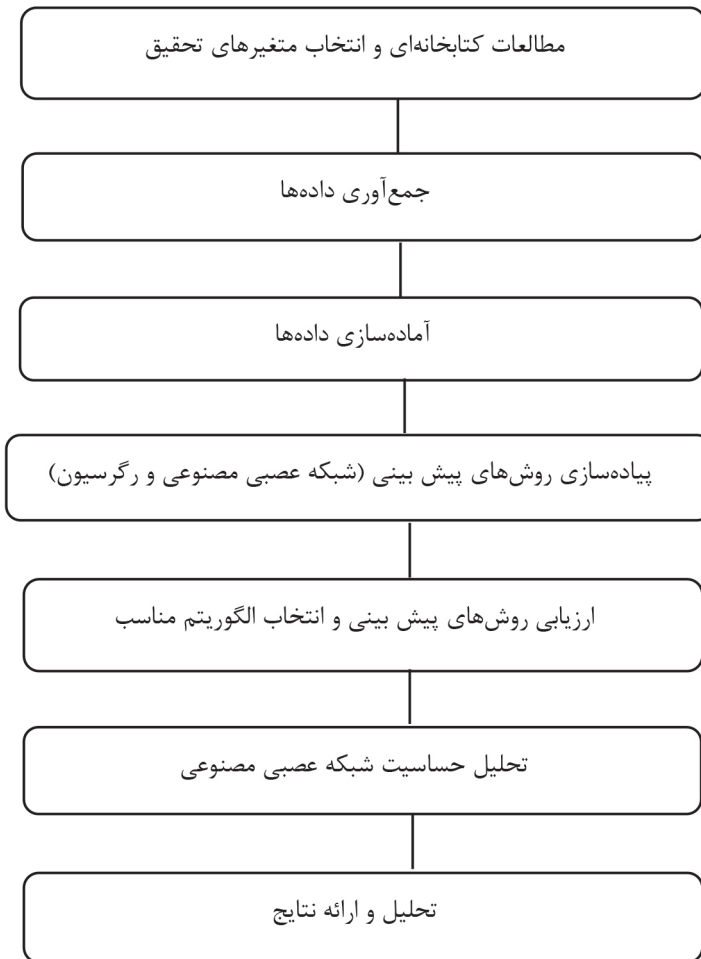
مقدار واقعی و مقدار پیش‌بینی شده) از شاخص میانگین مربعات خطا (MSE)^۱، ریشه دوم مربعات خطا (RMSE)^۲، ضریب همبستگی پیرسون (r)^۳ و ضریب تعیین (R^۲)^۴ استفاده شده است:

- شاخص میانگین مربعات خطا (MSE): لازم است برای بالا بردن صحت روش‌های پیش‌بینی، مقادیر این شاخص تا حد امکان پایین باشد (هن و همکاران، ۲۰۱۱).
- شاخص ریشه دوم مربعات خطا (RMSE): در صورتی که مقدار این شاخص کمتر از ۰٫۱ باشد، الگوریتم مورد نظر مطلوب است.
- شاخص ضریب همبستگی پیرسون (r): مقدار ضریب همبستگی بین -۱ تا +۱ بوده و علامت‌های مثبت و منفی نشان‌دهنده همبستگی مثبت و یا منفی متغیرها است. اگر بین مقدار واقعی و مقدار پیش‌بینی شده همبستگی مثبت قوی وجود داشته باشد r به +۱ نزدیک است و اگر بین آن‌ها همبستگی قوی منفی وجود داشته باشد مقدار r به -۱ نزدیک خواهد بود. هرگاه همبستگی ضعیفی بین متغیرها وجود داشته باشد و یا هیچ‌گونه همبستگی خطی بین متغیرها وجود نداشته باشد r نزدیک به صفر خواهد بود.
- شاخص ضریب تعیین (R^۲): مقدار ضریب تعیین بین صفر و یک است و هر چه به یک نزدیک‌تر باشد نشان‌دهنده برازش بهتر داده‌ها توسط مدل است. مقدار با افزایش مدل رشد می‌نماید و لازم است تعداد مشاهدات و تعداد ویژگی‌ها در مدل تأثیر داده شوند (فیلزموزر^۵، ۲۰۰۸). دقت الگوریتم‌ها در محیط متلب با یکدیگر مقایسه می‌شوند.

تحلیل نتایج و توسعه

در این مرحله، نتایج فرایند داده‌کاوی انجام شده تحلیل و به‌صورت گزارش ارائه می‌گردد. این گزارش ممکن است شامل خلاصه‌ای از روند اجرای پروژه و تجربیات به‌دست‌آمده یا گزارش جامعی از نتایج داده‌کاوی باشد. شکل ۱ چارچوب کلی روش پژوهش را ارائه می‌نماید.

1 . Mean square error
 2 . Root mean square error
 3 . Pearson correlation coefficient
 4 . Coefficient of Determination
 5 . Filzmoser



شکل ۱. چارچوب کلی روش پژوهش

تجزیه و تحلیل یافته‌ها

در اولین مرحله، شبکه‌های عصبی مصنوعی و رگرسیون جهت مدل‌سازی ارتباط بین داده‌های شاخص علمی و شاخص فناوری به کار برده شد، در دومین مرحله آنالیز حساسیت با استفاده از شبکه عصبی انجام شد. نرم‌افزارهای مورد استفاده در این پژوهش، اکسل (جهت نگهداری و پیش‌پردازش داده‌ها) و متلب (جهت اجرای الگوریتم‌های داده‌کاوی) می‌باشند.

شبکه عصبی مصنوعی

در این مرحله داده‌ها با شبکه‌های عصبی مختلف اعم از تک لایه، چندلایه و همچنین توابع انتقال و یادگیری متفاوت و تعداد نرون‌های مختلف در لایه‌های میانی مدل‌سازی شده‌اند. تعداد نرون‌ها در لایه میانی شبکه پرسپترون تأثیر زیادی بر روی عملکرد این شبکه‌ها دارد. تعداد نرون‌های بهینه بر اساس کمترین میزان میانگین مربعات خطا انتخاب شده‌اند. جهت مدل‌سازی با شبکه عصبی، کلیه رکوردها به سه مجموعه آموزشی، اعتبار سنجی و آزمایشی تقسیم‌بندی شده‌اند. تعداد رکوردها در مجموعه آموزشی ۶۰ درصد و در هر کدام از دو مجموعه دیگر ۲۰ درصد کل رکوردها انتخاب شده است. لازم به ذکر است در هر بار اجرا، با تغییر پارامترهای مدل و با توجه به دقت پیش‌بینی خروجی سیستم، مناسب‌ترین پارامترها شناسایی شده‌اند. ساختار شبکه عصبی پیاده‌سازی شده، شبکه چهار لایه ^۱MLP است. با تغییر معماری شبکه و اجرای مکرر شبکه‌های مختلف بهترین معماری با توابع انتقال Tansig در لایه‌های پنهان و Purelin در لایه خروجی به دست آمد. همچنین تابع Trainlm به‌عنوان مناسب‌ترین تابع یادگیری شناخته شد.

شبکه عصبی پیشنهادی برای داده‌ها، دارای معماری چندلایه است، تعداد لایه‌های پنهان، دولایه بوده و تعداد نرون‌ها در لایه اول ۲۰ و در لایه دوم ۲۰ عدد بوده است. لایه خروجی شامل یک نرون و تابع فعال‌سازی آن purelin است. معماری مدل شبکه عصبی بکار گرفته شده در این تحقیق در شکل ۲ ارائه شده است. عملکرد شبکه با شاخص‌های RMSE، MSE و R مورد سنجش قرار گرفت.

در شکل ۳، آموزش شبکه عصبی برای کل داده‌ها ارائه شده است. محور افقی نشان‌دهنده تعداد تکرار آموزش و محور عمودی نشان‌دهنده خطای آموزش است. این نمودار نشان می‌دهد که روند خطای شبکه عصبی بر اساس داده‌های آموزش^۲، اعتبارسنجی^۳ و آزمایش^۴ چگونه کاهش پیدا کرده است که در واقع همان فاصله بین تخمین‌های سیستم و خروجی‌های واقعی است که به ازای هر دوره آموزش^۵، خطا محاسبه شده و در محور عمودی نمایش داده می‌شود. Best Validation Performance نقطه‌ای است که خطای داده‌های اعتبارسنجی و آزمایش در مینیمم قرار گرفته‌اند که این نقطه در تحقیق حاضر به طور خودکار در بخشی قرار گرفته است که بیش برآزش^۶ و کم

1 . Multi-layer perceptron

2 . Train

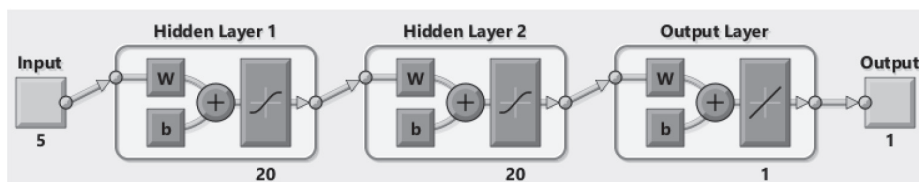
3 . Validation

4 . Test

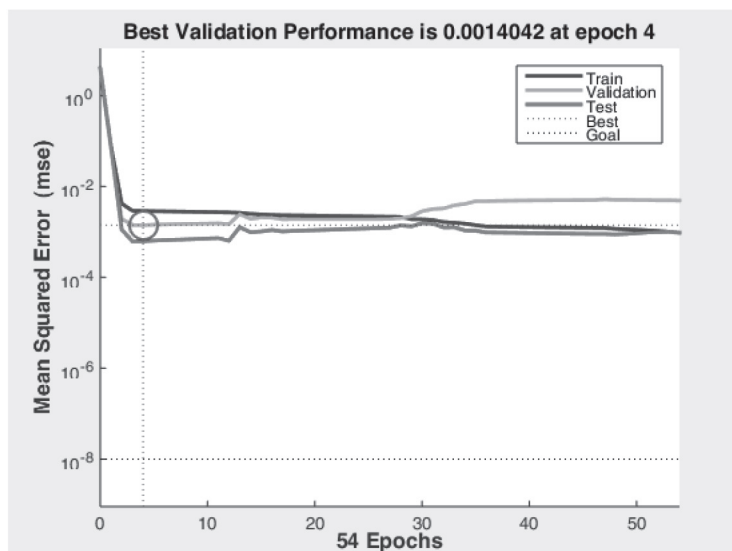
5 . Epoch

6 . Over fit

برازشی^۱ روی ندهد و سه بخش تفکیک شده داده با هم در یک نقطه کمینه ثابت از نظر خطا قرار گیرند. اولین نقطه Best Validation در دوره آموزشی ۴ در نظر گرفته شده است. نقطه بهینه نیز بر مبنای روند کاهش خطاها و بین بازه‌های خطایی به دست آمده است که بیش برآزش و کم برآزشی روی داده روی نداده است و میزان میانگین مجذور خطا برای هر سه بخش داده آموزش، اعتبارسنجی و آزمون روند نسبتاً یکسانی را طی کرده‌اند.



شکل ۲. معماری مدل شبکه عصبی



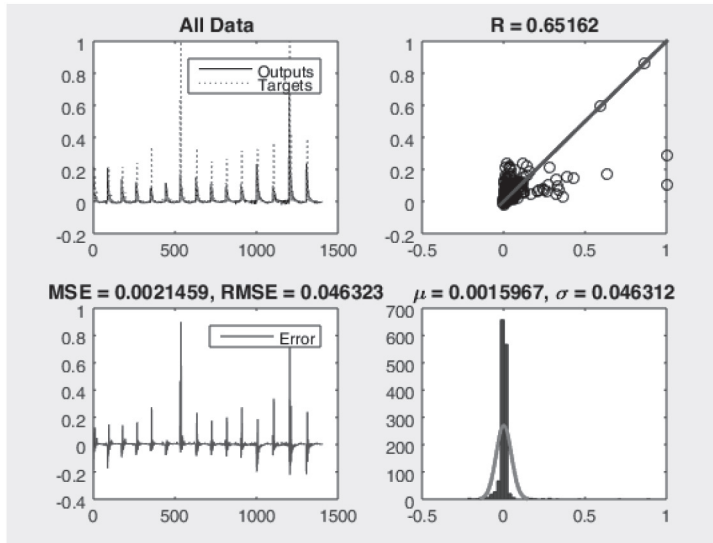
شکل ۳. آموزش شبکه عصبی کل داده‌ها

در شکل ۴، شکل ۵، شکل ۶ و شکل ۷ اطلاعاتی نظیر RMSE، MSE، هیستوگرام خطا^۲، رگرسیون، میانگین و انحراف معیار به تفکیک کل داده‌ها، داده‌های آزمایش، داده‌های آموزش و داده‌های ارزیابی

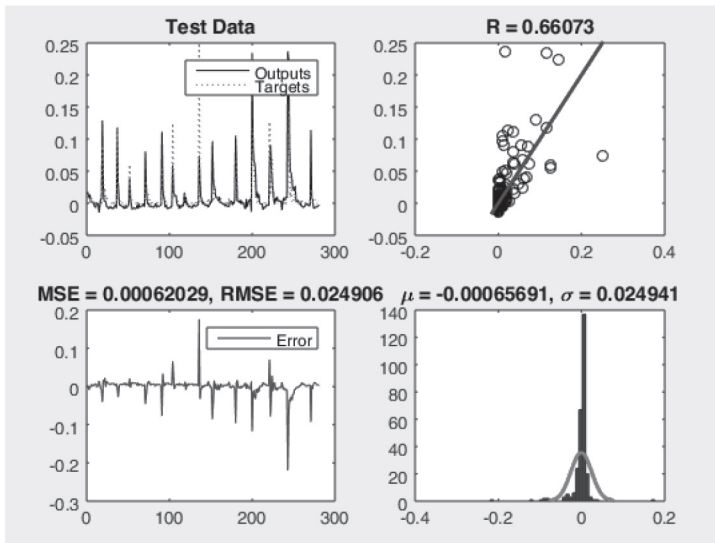
1 . Under fit

2 . Error histogram

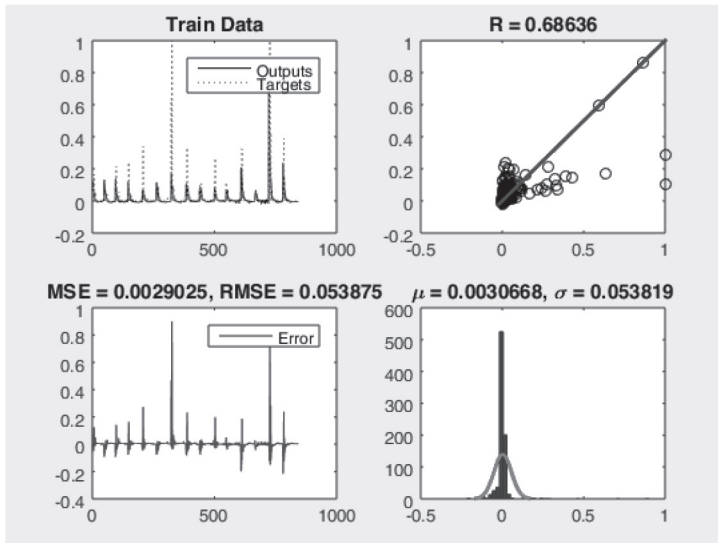
ارائه شده است.



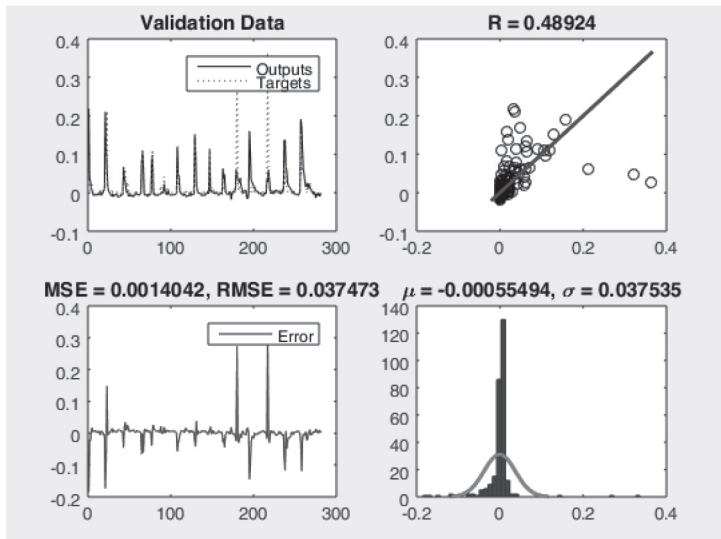
شکل ۴. نتایج آموزش شبکه عصبی (کل داده‌ها)



شکل ۵. نتایج آموزش شبکه عصبی (داده‌های آزمایش)



شکل ۶. نتایج آموزش شبکه عصبی (داده‌های آموزش)



شکل ۷. نتایج آموزش شبکه عصبی (داده‌های اعتبارسنجی)

در هر کدام از شکل‌های فوق، نمودار سمت راست پایین، خروجی تخمین سیستم نسبت به توزیع

نرمال را نشان می‌دهد. اختلاف مقادیر خطا با نقطه میانی نمودار نشان‌دهنده میزان فاصله خطا با عدد صفر است. هر چه نمودار توزیع خطا به نمودار توزیع نرمال نزدیک‌تر باشد نشان‌دهنده نزدیک‌تر بودن عدد خطا به عدد صفر و دقت بیشتر سیستم در تخمین مقادیر است. نمودار سمت راست بالا، نمودار رگرسیون مدل بین خروجی‌های سیستم و خروجی‌های واقعی است، ضریب همبستگی R^1 نشان می‌دهد به چه مقدار تغییرات خروجی شبکه عصبی توسط خروجی واقعی (هدف) توضیح داده شده است. هر چقدر مقدار این ضریب به یک نزدیک‌تر باشد همبستگی بیشتری بین خروجی‌های مدل و خروجی‌های هدف وجود دارد (در همبستگی کامل $R=1$ است). نمودار سمت چپ بالا، تخمین سیستم و داده‌های واقعی مربوط به داده‌های اعتبارسنجی را نشان می‌دهد. نمودار سمت چپ پایین، میزان MSE ، $RMSE$ و روند افزایش و کاهش خطا را در طی دوره آموزش، اعتبارسنجی و آزمایش نشان می‌دهد. نتایج آموزش شبکه عصبی به طور خلاصه در جدول ۱ ارائه شده است.

جدول ۱: نتایج آموزش شبکه عصبی MLP

R	RMSE	MSE	
۰.۶۵	۰.۰۴۶	۰.۰۰۲۱	کل داده‌ها
۰.۶۸	۰.۰۵۳	۰.۰۰۲۹	داده‌های آموزش
۰.۶۶	۰.۰۲۴	۰.۰۰۰۶۲	داده‌های آزمون
۰.۴۸	۰.۰۳۷	۰.۰۰۱۴	داده‌های اعتبارسنجی

با توجه به مقادیر کم MSE و $RMSE$ به دست آمده و همچنین ضریب همبستگی مناسب بین خروجی مدل و خروجی هدف، شبکه‌های عصبی به طور مطلوبی قادر به مدل‌سازی داده‌ها هستند.

رگرسیون

نتایج رگرسیون داده‌های پیش‌بینی کننده (تعداد مقالات، مقالات قابل استناد، خود استنادی، تعداد استنادها به ازای هر مقاله و شاخص اچ ایندکس) در برابر عامل پیش‌بینی (اختراعات) در جدول ۲ ارائه شده است. معادله رگرسیون به دست آمده به صورت ذیل است:

(۱)

$$\text{Patents} = -0.0022 - 2.58 \text{ Documents} + 2.92 \text{ Citable documents} + 0.32 \text{ H-index} - 0.363 \text{ Self citation} - 0.15 \text{ Citation per documents}$$

جدول ۲: نتایج رگرسیون

P	T	SE Coef	Coef	
۰,۳۹۷۵۵	-۰,۸۴۶۲۷	۰,۰۰۲۵۶۳	-۰,۰۰۲۱۶۹	Intercept
۰,۰۰۰۴۵۳۳۹	-۳,۵۱۵۲	۰,۷۳۳۷	-۲,۵۷۹۱	تعداد مقالات
۰,۰۰۰۰۰۴۳	۴,۶۱۱۱	۰,۶۳۴۰۵	۲,۹۲۳۷	مقالات قابل استناد
۰,۰۰۵۴۸۲۹	۲,۷۸۱۵	۰,۱۳۰۴۴	۰,۳۶۲۸۳	خود استنادی
۰,۰۸۷۸۲۷	-۱,۷۰۸۲	۰,۰۰۸۶۸۰۶	-۰,۰۱۴۸۲۸	استنادها به ازای هر مقاله
۰,۱۸۶۶۷	۱,۳۲۱۱	۰,۰۲۴۰۱۸	۰,۰۳۱۷۳۱	اچ ایندکس

جهت اعتبارسنجی مدل، کلیه رکوردها به دو دسته آموزشی و آزمایشی تقسیم‌بندی شدند. تعداد رکوردها در مجموعه آموزشی ۸۰ درصد و در مجموعه آزمایشی ۲۰ درصد انتخاب شده است. جهت بهینه‌سازی پارامترها از روش معادلات نرمال استفاده شده است. نتایج در جدول ۳ ارائه شده است.

جدول ۳: نتایج آموزش مدل با رگرسیون

R2	RMSE	
۰,۸۳۶۴	۰,۰۴۲	داده‌های آموزش
۰,۸۶۷۳	۰,۰۴۸	داده‌های آزمون

با توجه به مقادیر کم RMSE و مقدار ضریب تعیین که نزدیک به عدد یک است رگرسیون الگوریتم مناسبی برای مدل‌سازی است.

مقایسه نتایج مدل‌سازی با شبکه عصبی و رگرسیون

با توجه به شاخص RMSE داده‌های آزمون که در شبکه عصبی مصنوعی ۰,۰۲۴ و در رگرسیون ۰,۰۴۸ است، شبکه‌های عصبی مصنوعی از قابلیت پیش‌بینی بهتری نسبت به رگرسیون برخوردار

هستند و می‌توانند جهت مدل‌سازی شاخص فناوری در سطح کشورهای جهان به کار روند.

آنالیز حساسیت

در این مرحله، شبکه عصبی با پنج ورودی و یک خروجی ۲۰ بار با وزن دهی‌های اولیه متفاوت آموزش می‌بیند. عملکرد شبکه با میانگین شاخص MSE برای کل داده‌ها به ازای تمام دفعات آموزش اندازه‌گیری می‌شود. این خطای شبکه به‌عنوان مبنایی در نظر گرفته می‌شود که بقیه خطاها با آن مقایسه می‌شود. جهت تعیین اهمیت پیش‌بینی (PI) پنج فیلد ورودی تعداد ۶ ترکیب مختلف را می‌توان در نظر گرفت. برای هر ترکیب از داده‌های ورودی شبکه عصبی ۲۰ بار آموزش می‌بیند. ترکیب‌های مختلف ورودی‌ها در جدول ۴ نشان داده شده است.

جدول ۴: پارامترهای ترکیب‌های ورودی

ترکیب ورودی (IC)	پارامترهای ورودی
IC۱	مقالات قابل استناد - خود استنادی - استناد به ازای هر مقاله - اچ ایندکس
IC۲	تعداد مقالات - خوداستنادی - استناد به ازای هر مقاله - اچ ایندکس
IC۳	تعداد مقالات - مقالات قابل استناد - استناد به ازای هر مقاله - اچ ایندکس
IC۴	تعداد مقالات - مقالات قابل استناد - خوداستنادی - اچ ایندکس
IC۵	تعداد مقالات - مقالات قابل استناد - خوداستنادی - استناد به ازای هر مقاله
IC۶	تعداد مقالات - مقالات قابل استناد - خوداستنادی - استناد به ازای هر مقاله - اچ ایندکس

خطای شبکه به دست آمده برای هر ترکیب از ورودی‌ها در جدول ۵ آمده است.

با توجه به جدول زمانی که تمام متغیرها به‌عنوان ورودی انتخاب می‌شوند کمترین خطا مشاهده می‌گردد. در مقایسه سایر ترکیب‌ها، با توجه به نتایج به دست آمده ترتیب میزان خطای شبکه هر ترکیب ورودی از بیشترین خطا به کمترین خطا به ترتیب ذیل است:

$$IC5 > IC4 > IC1 > IC2 > IC3$$

IC۳ کمترین خطا را دارد که نشان‌دهنده این است که پارامتر خود استنادی کمترین اهمیت را بین داده‌های ورودی دارد. پارامتر بعدی که کمترین خطا را بعد از IC۳ دارد IC۲ است که نشان

می‌دهد پارامتر تعداد مقالات قابل استناد اهمیت بیشتری نسبت به خود استنادی دارد اما نسبت به سایر عوامل اهمیت کمتری دارد، بیشترین خطای ایجاد شده مربوط به ترکیب IC۵ است که پارامتر اچ ایندکس از آن حذف شده است و این امر نشان می‌دهد که اچ ایندکس اهمیت بیشتری نسبت به سایر عوامل داشته که حذف آن خطای بیشتری به شبکه داده است. ترتیب اهمیت پارامترها به شرح ذیل است:

اچ ایندکس < متوسط استنادها به ازای هر مقاله < تعداد مقالات < مقالات قابل استناد < خود استنادی

جدول ۵: خطای شبکه برای ترکیب‌های ورودی

میانگین MSE	ترکیب ورودی (IC)
۰.۰۰۲۲۸۳	IC۱
۰.۰۰۲۱۹۵	IC۲
۰.۰۰۲۱۳۲	IC۳
۰.۰۰۲۳۴۵	IC۴
۰.۰۰۲۴۰۰	IC۵
۰.۰۰۲۰۴۵	IC۶

پاسخ به سؤالات تحقیق

- با توجه به نتایج به دست آمده از تجزیه و تحلیل یافته‌ها به سؤالات تحقیق پاسخ داده می‌شود:
- با توجه به اهمیت داده‌ها و منابع مورد استفاده جهت انجام پژوهش، مناسب‌ترین منابع جهت دستیابی به اهداف تحقیق کدام است؟
- با توجه به لزوم استفاده از داده‌های کلان در سطح کشورها و نیاز به جامع و استاندارد بودن داده‌ها، مناسب‌ترین منبع، پایگاه داده‌ای است که معتبر، بین‌المللی و به‌روز باشد که در این پژوهش از پایگاه‌های داده SJCR اسکوپوس و سازمان جهانی مالکیت فکری (WIPO) استفاده شد که مورد تأیید کلیه کشورهای جهان بوده و معیارهای اساسی عملکرد کشورها از لحاظ تولید علم و فناوری می‌باشند.
- چگونه می‌توان با استفاده از قابلیت مدل‌های شبکه‌های عصبی مصنوعی و رگرسیون، شاخص

فناوری کشورها را با توجه به شاخص‌های علم‌سنجی پیش‌بینی نمود؟ در این پژوهش، از دو مدل شبکه‌های عصبی مصنوعی و رگرسیون استفاده گردید. در هر دو مدل داده‌های خروجی، شاخص فناوری کشورها و داده‌های ورودی، شاخص‌های علم‌سنجی کشورها بودند. با توجه به معیارهای اعتبارسنجی استاندارد مشخص گردید که هر دو مدل شبکه‌های عصبی مصنوعی و رگرسیون قادرند به نحو مطلوبی داده‌های خروجی را بر اساس داده‌های ورودی پیش‌بینی نمایند و در مقام مقایسه این دو مدل، مدل شبکه عصبی مصنوعی با شاخص RMSE داده‌های آزمون به میزان 0.024 نسبت به رگرسیون با شاخص RMSE داده‌های آزمون 0.048 از قابلیت پیش‌بینی مناسب‌تری نسبت به رگرسیون برخوردار بود و به‌عنوان الگوریتم دقیق‌تر، انتخاب گردید. در آینده، امکان توسعه این الگوریتم به الگوریتم‌های ترکیبی مانند سیستم‌های استنتاج فازی عصبی انطباقی در ترکیب با الگوریتم‌های تکاملی نیز وجود خواهد داشت.

- کدام شاخص تولید علم در سطح کشورها اهمیت نسبی بیشتری جهت پیش‌بینی فناوری دارد؟ با توجه به نتایج تحلیل حساسیت شبکه‌های عصبی، شاخص اچ‌اچ‌اچ نسبت به سایر شاخص‌های تولید علم اهمیت بیشتری جهت پیش‌بینی فناوری داشته است چراکه حذف آن خطای بیشتری را در شبکه عصبی ایجاد می‌کند و حساسیت روند توسعه علمی کشورها به استاندارد مؤثر و کاربرد دانش از طرف دیگر کشورها به یافته‌های علمی کاملاً مشاهده می‌گردد. همچنین سایر شاخص‌ها به ترتیب اهمیت بدین صورت شناسایی شدند: شاخص متوسط استنادها به ازای هر مقاله، تعداد مقالات، مقالات قابل استناد و خوداستنادی.

نتیجه‌گیری

در این پژوهش رویکردی نوین جهت ارزیابی تولید تکنولوژی با توجه به فاکتورهای کمیت و کیفیت تولید علم در سطح کلان کشورها ارائه گردید. جهت نیل به این هدف، پایگاه داده‌ای متشکل از شاخص‌های استاندارد و بین‌المللی تولید علم و فناوری ایجاد گردید و الگوریتم‌های شبکه عصبی مصنوعی و رگرسیون جهت مدل‌سازی داده‌ها به کار گرفته شد. نتایج به دست آمده تحلیل و ارزیابی گردید و شاخص‌های اعتبارسنجی نشان دادند که دقت الگوریتم‌های بکار گرفته شده در سطح قابل قبولی بوده و در مقام مقایسه، شبکه‌های عصبی مصنوعی دقت بیشتری را ارائه داده است. در مرحله بعد

با استفاده از شبکه عصبی، آنالیز حساسیت جهت مشخص نمودن کیفیت و اثر هر یک از پارامترهای ورودی به شبکه (شاخص‌های علم‌سنجی) در پیش‌بینی شاخص فناوری انجام شد. این روش با استفاده از ترکیبات مختلف پارامترهای ورودی به شبکه و آموزش مجدد شبکه صورت پذیرفت. برای هر ترکیب از ورودی‌ها، شبکه عصبی ۲۰ بار با وزن دهی‌های اولیه متفاوت آموزش دید و میانگین خطای شبکه (MSE) تمام دفعات آموزش به‌عنوان خطای مدل شبکه عصبی در نظر گرفته شد. بر اساس شاخص اهمیت پیش‌بینی (PI) در تحلیل حساسیت، پارامتر اچ‌اچ‌اچ‌اچ‌اچ با اهمیت‌ترین، شاخص متوسط استنادها به ازای هر مقاله در درجه دوم اهمیت و پس از آن شاخص تعداد مقالات و مقالات قابل استناد قرار گرفته‌اند. خود استنادی کم‌اهمیت‌ترین پارامتر در پیش‌بینی شاخص فناوری شناسایی شد. نوآوری استفاده از الگوریتم‌های یادگیری ماشین بدین‌صورت بر روی داده‌های کلان‌علم و فناوری در سطح بین‌المللی و مقالات معتبر محدود بوده و اخیراً استفاده از آن‌ها آغاز شده است. در این پژوهش محقق ابتدا الگوریتم‌های شبکه عصبی مصنوعی و رگرسیون را بر روی داده‌ها اعمال نموده و هدف رسیدن به دقت قابل قبول جهت پیش‌بینی شاخص فناوری کشورها با توجه به داده‌های تولید علم بوده است. رسیدن به نتایج ارزیابی بسیار مناسب مدل، حاکی از توانایی الگوریتم‌های شبکه‌های عصبی جهت ایجاد ارتباط بین عناصر تولید علم با عنصر تولید فناوری در سطح کلان بوده است. سپس با استفاده از تحلیل حساسیت شبکه عصبی مشخص شد که چه شاخص‌هایی از تولید علم تأثیر بیشتری بر تولید فناوری دارند. با روندی که وجود دارد به نقطه بهتری از لحاظ تولید فناوری نخواهیم رسید مگر اینکه به متغیرهای حساس و تأثیرگذار تأکید و توجه بیشتری داشته باشیم. در واقع شاخص‌هایی از تولید علم اهمیت بیشتری دارند که زمینه‌ساز ایجاد فناوری باشند. نتایج حاصل از تحقیق، در پیش‌بینی شاخص‌های کلان و روند آتی کشورها در زمینه تولید علم و فناوری بسیار کاربردی خواهد بود و نشانگر امکان‌پذیری استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی دقیق روند پیشرفت علم و فناوری کشورها طی سالیان متمادی است. محقق در این پژوهش با محدودیت‌هایی همچون عدم دسترسی به داده‌های سایر پایگاه‌ها و منابع اطلاعاتی، دامنه زمانی محدود داده‌ها و محدود بودن متغیرهای پژوهش در دسترس در سطح کشورها مواجه بوده است. همچنین پیشینه تحقیقات معادل این پژوهش و استفاده از این نوع الگوریتم‌های یادگیری ماشین بر روی داده‌های موجود بسیار محدود بوده است. شناسایی و استفاده از فیله‌های اطلاعاتی جدید در پایگاه داده و همچنین استفاده از روش‌های نوروفازی و نوروفازی ژنتیک جهت پژوهش‌های آتی پیشنهاد می‌گردد.

منابع

- بهرامی، م. (۱۳۷۴). *تکنولوژی‌های آینده، شناسایی و پیش‌بینی*. تهران: انتشارات خضرا.
- خلیل، ط. (۱۳۸۱). *مدیریت تکنولوژی، رمز موفقیت در رقابت و خلق ثروت*. (س. باقری، مترجم) انتشارات پیام متن وابسته به مرکز تکنولوژی نیرو.
- ذوالفقار نسب، س. (۱۳۸۳). *تاریخچه ارزیابی علم و فناوری و ظهور شاخص‌های علمی با تأکید بر کشور آمریکا، هیئت نظارت و ارزیابی فرهنگی و علمی*.
- سلطانی نژاد، ع. (۱۳۹۳). *تحلیل رابطه میان تولیدات علمی کشورهای برتر جهان و میزان اختراعات ثبت شده در پایگاه استنادی اسکوپوس طی سال‌های ۲۰۰۸-۲۰۱۳*. پایان‌نامه کارشناسی ارشد رشته علم اطلاعات و دانش‌شناسی، دانشکده ادبیات و علوم انسانی دانشگاه شهید باهنر کرمان.
- گودرزی، م. (۱۳۸۲). *بررسی وضعیت نظام مالکیت فکری ایران و ارائه راهکارهای بهبود آن در جهت توسعه تکنولوژیکی کشور*. تهران: پایان‌نامه کارشناسی ارشد رشته مدیریت تکنولوژی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی.
- منہاج، م. (۱۳۸۴). *مبانی شبکه‌های عصبی*. تهران: دانشگاه صنعتی امیرکبیر.
- نقی پور، پ. (۱۳۸۵). *اولویت‌بندی کشورها جهت ثبت پتنت‌ها در رابطه با تکنولوژی*. تهران: پایان‌نامه کارشناسی ارشد، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی.
- نواز شریف، م. (۱۳۶۷). *مدیریت انتقال تکنولوژی و توسعه*، ترجمه رشید اصلانی. تهران: سازمان برنامه‌وبودجه.
- Aksnes, ., Sivertesen, ., Leeuwen, ., & Wendt, K. (2017). Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators. *Science and Public Policy*, 44(2), 246–258.
- Anegón, F. & Solana, V. (2013). Worldwide Topology of the Scientific Subject Profile: A Macro Approach in the Country Level. *PLOS ONE*, 8(12).
- Baykal, N. (2012). Predicting the disease of Alzheimer CAD with SNP biomarkers and clinical data based decision support system using data mining classification approaches. *Middle East Technical university*.
- Choi, J. Jang, D. Jun, S. & Park, S. (2015). A Predictive Model of Technology Transfer Using Patent Analysis. *Sustainability*, 16175–16195.
- De, R. Pal, N. & Pal, S. (1997). Feature analysis: neural network and fuzzy set theoretic approaches. *Pattern recognition*, 30(10), 1579-1590.
- European commission. (2011). *Retrieved from European commission*.

- Filzmoser, P. (2008). *Linear and nonlinear methods for regression and classification and applications in R*. Wien: Wiedner Hauptstr.
- Glänzel, W. Moed, H. Schmoch, U. & Thelwall, M. (2018). *Springer Handbook of Science and Technology Indicators*. Springer.
- Hadzima-Nyarko, M. Nyarko, E. & Moric, D. (2011). A neural network based modeling and sensitivity analysis of damage ratio coefficient. *Expert systems with applications*, 13405-13413.
- Han, J. Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed). Morgan Kaufmann Publishers.
- Hashem, S. (1992). Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. *Proceedings of the 1992 international joint conferences on neural networks*, Baltimore, MD (Vol. 1, pp. 419–424). IEEE Press.
- Idris, K. (2002). *Intellectual Property, A power tool for economic growth*. WIPO publication.
- Jang, H. & Kim, H. (2014). Research output of science, technology and bioscience publications in Asia. *Science editing*, 1(2), 62-70.
- Madaleno, M. Moutinho, V. & Robaina, M. (2016). Economic and environmental assessment: EU cross-country efficiency ranking analysis. *energy procedia*, 106, 134-154.
- Molas, G. & Yamazaki, F. (1995). Neural networks for quick earthquake damage estimation. *Earthquake Engineering and Structural Dynamics*, 24, 505–516.
- Narin, F. Hamilton, K. & Olivastro, D. (1997). The increasing linkage between U.S technology and public science. *Research Policy*, 26, 317-330.
- Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*. *OECD Science, Technology and Industry Working Papers*.
- Pal, N. (1999). Soft computing for feature analysis. *Fuzzy sets and systems*, 103, 201-221.
- Porter, A. Roper, A. Manson, T. Rossini, F. Banks, J. & Wiederhdt, B. (1991). *Forecasting and management of technology*. USA: Wiley.
- Radicchi, F. & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*. doi: 10.1007/s11192-013-1027-3
- Sandoval-Romero, V. Mongeon, P. & Lariviere, V. (2018). Science, technology and

innovation indicators in transition. *International conference on science and technology indicators*. Leiden, The Netherlands.

- sarle, W. (2000). *How to measure importance of inputs?* Retrieved from <ftp://ftp.sas.com/pub/neural/importance.html>
- science-metrix. (2010). *30 years in science: Secular movements in knowledge creation*. Retrieved from science-metrix: <http://www.science-metrix.com/30years-Paper.pdf>
- SCImago. (2007, July 21). <http://www.scimagojr.com>. Retrieved from SJR — SCImago Journal & Country Rank.
- Sirilli, G. (1999). Innovation indicators in science and technology evaluation. *Scientometrics*, 45, 439-443.
- Sohn, Y. & Moon, T. (2004). Decision Tree based on data envelopment analysis for effective technology commercialization. *Expert Systems with Applications*, 26, 279–284.
- Sugimoto, C. & Lariviere, V. (2018). *Measuring Research. What Everyone Needs to Know*. New York: Oxford University Press.
- Tan, P.N. Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. Pearson.
- Twiss, B. (1992). Forecasting for technologist and engineers.
- VILIMEK, M. (2014). An artificial neural network approach and sensitivity analysis in predicting skeletal muscle forces. *Acta of Bioengineering and Biomechanics*, 16. doi: 10.5277/abb140314
- Zainab, T. & Ashraf Wani, Z. (2018). *Advancement and Application of Scientometric Indicators for Evaluation of Research Content*. IGI Global.