Review of Rule Extraction Methods in Data Mining:
A Systematic Review of Studies

Sara Ansari*
Master's Student in Industrial Engineering; Industrial and Systems Engineering Department; Isfahan University of Technology; Isfahan, Iran Email: saraansari@in.iut.ac.ir

Saba Sareminia
PhD in Industrial Engineering; Associate Professor; Industrial and Systems Engineering Department; Isfahan University of Technology; Isfahan, Iran Email: s.sareminia@iut.ac.ir

Iranian Journal of Information Processing and Management

Received: 20, Aug. 2024 **Accepted: 26, May 2025**

Abstract: Extracted rules from data represent one of the most important and practical forms of knowledge representation applied across various domains, including expert systems, decision support, and automated control. Rule extraction methods, valued for their transparency and interpretability, enable researchers to understand underlying patterns and substructural relationships within datasets. Consequently, numerous approaches and algorithms for rule extraction have been developed, each with its own strengths and weaknesses. This paper systematically reviews rule extraction methods in data mining. A total of 678 articles were initially collected from reputable scientific databases, with 19 selected for final analysis after screening. The analysis is conducted in three parts: 1) a critical evaluation of rule extraction strategies and performance metrics, 2) identification of research gaps based on the analysis to propose providing areas of study for future research, and 3) textual content analysis of studies using tools like VOS viewer to identify key concepts and their relationships. The findings of this study can contribute to developing more efficient rule extraction algorithms across diverse domains and pave the way for future research in this field.

Keywords: Systematic Review, Rule Extraction, Data Mining, Rule Optimization, Machine Learning

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 40 | No. 4 | pp. 1147-1178
Summer 2025
https://doi.org/10.22034/jipm.2025.2039221.1737

* Corresponding Author

بررسی روش‌های استخراج قانون در داده‌کاوی: مرور نظام‌مند مطالعات

سارا انصاری
دانشجوی کارشناسی ارشد مهندسی صنایع؛ دانشکده مهندسی صنایع و سیستم‌ها؛ دانشگاه صنعتی اصفهان؛ اصفهان، ایران؛ پدیدآور رابط saraansari@in.iut.ac.ir

صبا صارمی‌نیا
دکتری مهندسی صنایع؛ استادیار؛ دانشکده مهندسی صنایع و سیستم‌ها؛ دانشگاه صنعتی اصفهان؛ اصفهان، ایران s.sareminia@iut.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۳/۰۵/۳۰ | پذیرش: ۱۴۰۴/۰۳/۰۶ | مقاله برای اصلاح به مدت ۳۸ روز نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۴۰ | شمارهٔ ٤ | ص‌ص ۱۱۴۷-۱۱۷۸
تابستان ۱۴۰٤
https://doi.org/10.22034/jipm.2025.2039221.1737

چکیده: قوانین استخراج‌شده از داده‌ها یکی از مهم‌ترین و کاربردی‌ترین اشکال نمایش دانش هستند که در حوزه‌های مختلفی مانند سیستم‌های خبره، پشتیبانی تصمیم و کنترل خودکار کاربرد دارند. روش‌های استخراج قانون به‌دلیل شفافیت و تفسیرپذیری بالا به محققان امکان می‌دهند که الگوها و روابط زیربنایی درون داده‌ها را درک کنند. از این‌رو، رویکردها و الگوریتم‌های متعددی برای استخراج قانون از داده‌ها توسعه یافته که هر یک نقاط ضعف و قوت مربوط به خود را دارند. این مقاله با استفاده از روش مرور نظام‌مند به بررسی روش‌های استخراج قانون در حوزه داده‌کاوی پرداخته است. در این پژوهش ۶۷۸ مقاله از پایگاه‌های علمی معتبر جمع‌آوری شد و پس از غربالگری، ۱۹ مقاله برای تحلیل نهایی انتخاب شدند. تحلیل‌ها در سه بخش انجام شده است: ۱) تحلیل نقادانه راهبردهای استخراج قوانین و معیارهای تحلیل عملکرد آن‌ها، ۲) شناسایی شکاف‌های پژوهشی بر اساس تحلیل‌های صورت‌گرفته با هدف ارائه زمینه‌های مطالعاتی برای تحقیقات آینده، و ۳) تحلیل محتوای متنی پژوهش‌ها با استفاده از ابزارهایی مانند VOS viewer برای شناسایی مفاهیم کلیدی و ارتباطات میان آن‌ها. نتایج این پژوهش می‌تواند به توسعه الگوریتم‌های کارآمدتر استخراج قانون در حوزه‌های مختلف کمک کرده و زمینه‌ساز تحقیقات آینده در این حوزه باشد.

کلیدواژه‌ها: مرور نظام‌مند، استخراج قانون، داده‌کاوی، بهینه‌سازی قوانین، یادگیری ماشین

۱. مقدمه

در کشف دانش از داده‌های ساختاریافته، قوانین استخراج‌شده یکی از بصری‌ترین و پرکاربردترین اشکال نمایش دانش را نشان می‌دهند (Maszczyk, Sikora & Wróbel 2024). روش‌های استخراج قانون به‌عنوان یکی از تکنیک‌های یادگیری ماشین و داده‌کاوی به استخراج الگوهای قابل فهم و قابل تفسیر از داده‌ها پرداخته و به تحلیل و پیش‌بینی رفتارهای آینده کمک می‌کنند. این روش‌ها به‌ویژه برای ایجاد مجموعه قوانین از مقدار زیادی دانش ساختاریافته می‌توانند مفید باشند (Li et al. 2024). به‌طور کلی، استنتاج قوانین به ما این امکان را می‌دهد که از داده‌های پیچیده و بزرگ، قواعدی ساده و قابل فهم استخراج کنیم تا تصمیم‌گیری‌های بهتری داشته باشیم. تفاوت اصلی روش‌های استخراج قانون با روش‌های مبتنی‌بر قانون در این است که اولی بر فرایند تولید قوانین از داده‌ها تمرکز دارد، در حالی که دومی به کاربرد این قوانین در سیستم‌های هوشمند اشاره می‌کند. به‌گفته دیگر، رویکردها و روش‌های استخراج قانون، زیربنای روش‌های مبتنی بر قانون شمرده می‌شوند. سیستم‌های مبتنی بر قانون که از قوانین استخراج‌شده استفاده می‌کنند، برای طیف وسیعی از کاربردها مانند سیستم‌های خبره (Cao et al. 2020)، سیستم‌های پشتیبانی تصمیم (Casalino et al. 2019) و سیستم‌های کنترل خودکار (Ding et al. 2019) استفاده می‌شوند. استفاده از این سیستم‌ها به‌عنوان ابزارهای هوشمند در حوزه‌های مختلف از جمله پزشکی می‌تواند کمک‌کننده باشد. با استفاده از قوانین استخراج‌شده، سیستم‌های مبتنی بر قانون می‌توانند به تشخیص بیماری‌هایی مانند بیماری‌های عروقی (Verma et al. 2016)، و بیماری‌های مزمن کلیه (Ogunleye & Wang 2020) کمک کنند. این سیستم‌ها همین‌طور در حوزه بازاریابی و رفتار مشتری می‌توانند با پیش‌بینی ریزش مشتری (Verma, Srivastava & Negi 2011) تحلیل‌های بازاریابی برجسته‌ای انجام دهند.

قوانین به‌طور معمول، در قالب «اگر-آنگاه» بیان می‌شوند؛ جایی که بخش «اگر» شرایط اعمال قانون را مشخص می‌کند و بخش «آنگاه» اقدام یا نتیجه‌ای را در صورت برآورده شدن شرایط، مشخص می‌سازد (Li et al. 2024). این ساختار برگرفته از نظریه مجموعه‌ها و منطق گزاره‌ای است که پایه و اساس روش‌های استخراج قانون را تشکیل می‌دهند.

هدف اصلی این روش‌ها یافتن مجموعه قوانینی است که محدودیت‌های کیفیت

فرضی مانند دقت (اطمینان)[1] یا پشتیبانی (پوشش)[2] را برآورده کنند؛ ضمن اینکه مجموعه قوانین جامع[3] و دو به ناسازگار[4] تنظیم می‌نمایند. این محدودیت‌ها از مفاهیم آماری و احتمال شرطی نشأت می‌گیرند و به انتخاب بهترین قوانین کمک می‌کنند. معیارهایی نظیر «دقت» و «پوشش» به محققان امکان می‌دهند که قوانین را نه‌تنها بر اساس ارتباط آن‌ها با داده‌ها، بلکه بر اساس قابلیت تعمیم‌پذیری آن‌ها اولویت‌بندی کنند. همچنین پس از آن می‌توان بر اساس جذابیت قانون، قوانین استخراج‌شده را اولویت بندی نمود (Maszczyk, Sikora & Wróbel 2024). در نتیجه، لزوم پژوهش‌هایی در زمینه بهبود دقت پیش‌بینی در داده‌های بزرگ و پیچیده یا توسعه پژوهش‌هایی در زمینه ایجاد الگوریتم‌هایی با تفسیرپذیری بالای قوانین احساس می‌شود (Vale, El-Sharif & Ali 2022). تفسیرپذیری یکی از معیارهای کلیدی برای ارزیابی قوانین استخراج‌شده است که میزان سادگی و قابل فهم بودن قوانین را مشخص می‌کند. این معیار اهمیت بالایی در کاربردهای واقعی دارد، زیرا کاربران و تصمیم‌گیرندگان می‌توانند به‌راحتی الگوها و روابط درون داده‌ها را درک کنند. در حالی که برخی از تکنیک‌های یادگیری ماشین، مانند شبکه‌های عصبی عمیق، در تحلیل داده‌های پیچیده موفق بوده، و به‌دلیل عملکرد به‌صورت جعبه سیاه، فاقد قابلیت توضیح‌پذیری هستند. این تفاوت، روش‌های مبتنی بر قانون را برای کاربردهایی که نیاز به شفافیت در تصمیم‌گیری دارند مانند تحلیل ریسک‌های مالی (Uthayakumar, Vengattaraman & Dhavachelvan 2020)، متمایز می‌کند

این امر می‌تواند به توسعه سیستم‌های کارآمدتر در حوزه‌های مختلف منجر شود. بررسی روش‌های استخراج قانون برای درک اصول اساسی و پیشرفت در این زمینه بسیار مهم است. روش‌های استخراج قانون در داده‌کاوی یکی از ابزارهای کلیدی برای شناسایی الگوهای نهفته و ارائه قواعدی قابل فهم به شمار می‌روند. این روش‌ها با استفاده از نظریه مجموعه‌های فازی و یادگیری استدلالی امکان برخورد با داده‌های نامطمئن و پیچیده را فراهم می‌کنند. مجموعه‌های فازی به‌ویژه، در داده‌هایی که دارای ابهام یا عدم قطعیت هستند، مانند پیش‌بینی‌های آب‌وهوا (Yu et al. 2016)، بسیار مؤثر هستند

1. confidence
2. coverage
3. comprehensive
4. mutually exclusive

با این حال، به‌رغم پیشرفت‌های انجام‌شده، چالش‌هایی نظیر کاهش دقت قوانین در داده‌های نامتعادل (Napierała & Stefanowski 2015)، پیچیدگی بالای الگوریتم‌ها در داده‌های با ابعاد زیاد (Yanjie & Hongbo 2018)، و بی‌توجهی به داده‌های غیرساختاریافته (Chemchem & Drias 2015) همچنان وجود دارند و به‌عنوان موانع اصلی در این حوزه مطرح هستند. این چالش‌ها باعث شده‌اند که توسعه روش‌های ترکیبی، مانند ادغام الگوریتم‌های استخراج قانون با یادگیری ماشین عمیق، مورد توجه قرار گیرد. هدف از این ترکیب، بهره‌گیری از قدرت یادگیری ماشین در تحلیل داده‌های پیچیده و توانایی روش‌های استخراج قانون در تفسیرپذیری است. این چالش‌ها بیانگر شکاف‌های تحقیقاتی در حوزه روش‌های استخراج قانون هستند

هدف این پژوهش، ارائه یک مرور نظام‌مند از روش‌های استخراج قانون در حوزه داده‌کاوی، تحلیل ویژگی‌ها و شناسایی شکاف‌های تحقیقاتی برای هدایت مطالعات آینده است. دامنه این پژوهش معطوف به بررسی رویکردهای مختلف استخراج قانون است. این مطالعه به‌دنبال ارائه بینشی جامع از وضعیت کنونی پژوهش‌ها و تبیین زمینه‌های کمتر کاوش‌شده، مانند کاربرد روش‌های استخراج قانون در داده‌های غیرساختاریافته است. با بررسی ادبیات پژوهش این حوزه، محققان می‌توانند درک عمیق‌تری از ویژگی‌ها، محدودیت‌ها و کاربردهای روش‌های استخراج قانون به‌دست آورند. این فرایند به شناسایی روندهای پژوهشی، مانند تمرکز بر تفسیرپذیری یا استفاده از رویکردهای ترکیبی، و چالش‌های موجود، مانند عملکرد در داده‌های پیچیده، کمک می‌کند و با تحلیل نظام‌مند روش‌های موجود برای استخراج قوانین از داده‌ها، به بررسی نقادانه رویکردهای توسعه‌یافته پرداخته است. هدف اصلی این تحقیق، شناسایی رویکردهای کنونی، درک بهتر تکنیک‌ها و کاربردهای آن‌ها و تبیین شکاف‌های تحقیقاتی به‌منظور ارائه زمینه‌های مطالعاتی برای تحقیقات آینده است

۲. روش پژوهش

هدف این پژوهش یک مطالعه مروری نظام‌مند با رویکرد تحلیلی و انتقادی است که با بهره‌گیری از تحلیل‌های آماری و محتوایی، به بررسی پژوهش‌های موجود و شناسایی شکاف‌های تحقیقاتی در حوزه روش‌های استخراج قانون در داده‌کاوی می‌پردازد. این مطالعه با استفاده از چارچوب مرور نظام‌مند، به تحلیل دقیق مقالات علمی منتخب

پرداخته و ضمن بررسی روش‌های مختلف استخراج قانون و کاربردهای آن‌ها، در پی تبیین وضعیت کنونی پژوهش‌ها و ارائه بینشی برای جهت‌دهی تحقیقات آینده است

مرور نظام‌مند این پژوهش بر اساس چارچوب «پریسما»[1] انجام شده است، که مجموعه‌ای از راهنماها و اصول برای اطمینان از شفافیت و ساختارمندی فرایند جست‌وجو، غربالگری، و انتخاب مقالات است. بر اساس این چارچوب فرایند انجام مرور نظام‌مند ادبیات پژوهش مشتمل بر تبیین سؤال پژوهش، راهبرد جست‌وجو و غربالگری مطالعات، بررسی تحلیلی مقالات و ارائه نتایج است. این روش شمای کلی از معیارهای قیاسی را نمایان کرده و به ارزیابی و بهینه‌سازی مقیاس‌ها کمک می‌کند (Moher et al. 2009) و با ارائه خلاصه‌ای جامع از مطالعات مورد بررسی می‌تواند در ارائه راهنمایی‌های خاص برای گزارش‌دهی، تسهیل شفافیت گزارش، و تکرارپذیری بررسی‌ها کمک کند (McInnes et al. 2018; Siddaway, Wood & Hedges 2019).

این بخش در چهار زیرمجموعه سازماندهی شده که عبارت است از: ۱) **تبیین سؤال پژوهش**، که سؤالات اصلی هدایت‌کننده مطالعه را مشخص می‌کند، ۲) **راهبرد جست‌وجو و غربال**، که فرایند جمع‌آوری مقالات از پایگاه‌های علمی را شرح می‌دهد، ۳) **ارزیابی و انتخاب مقالات**، که مراحل غربالگری و انتخاب مقالات نهایی را توضیح می‌دهد، و ۴) **بررسی و تحلیل مطالعات**، که روش‌های تحلیل آماری و متنی مقالات منتخب را معرفی می‌کند. جزئیات هر مرحله در ادامه ارائه شده است

تبیین سؤال پژوهش

این پژوهش با هدف پاسخ به سؤالات زیر طرح‌ریزی شده است:

۱. اصلی‌ترین روش‌های استخراج قانون از داده و معیار عملکردی مقایسه قوانین در ادبیات پژوهش چیست؟ و کدام روش بیشتر از بقیه مورد اقبال قرار گرفته است؟
۲. مشخصات و محدودیت‌های داده‌های ورودی راهبردهای توسعه قانون را چگونه تحت‌الشعاع قرار می‌دهد؟
۳. شکاف پژوهشی در فرایند توسعه و استخراج قانون چیست؟

راهبرد جست‌وجو و غربال

مرور نظام‌مند با استفاده از پایگاه‌های اطلاعاتی معتبر جهانی برای جمع‌آوری مقالات

1. preferred reporting items for systematic reviews and meta-analyses (PRISMA)

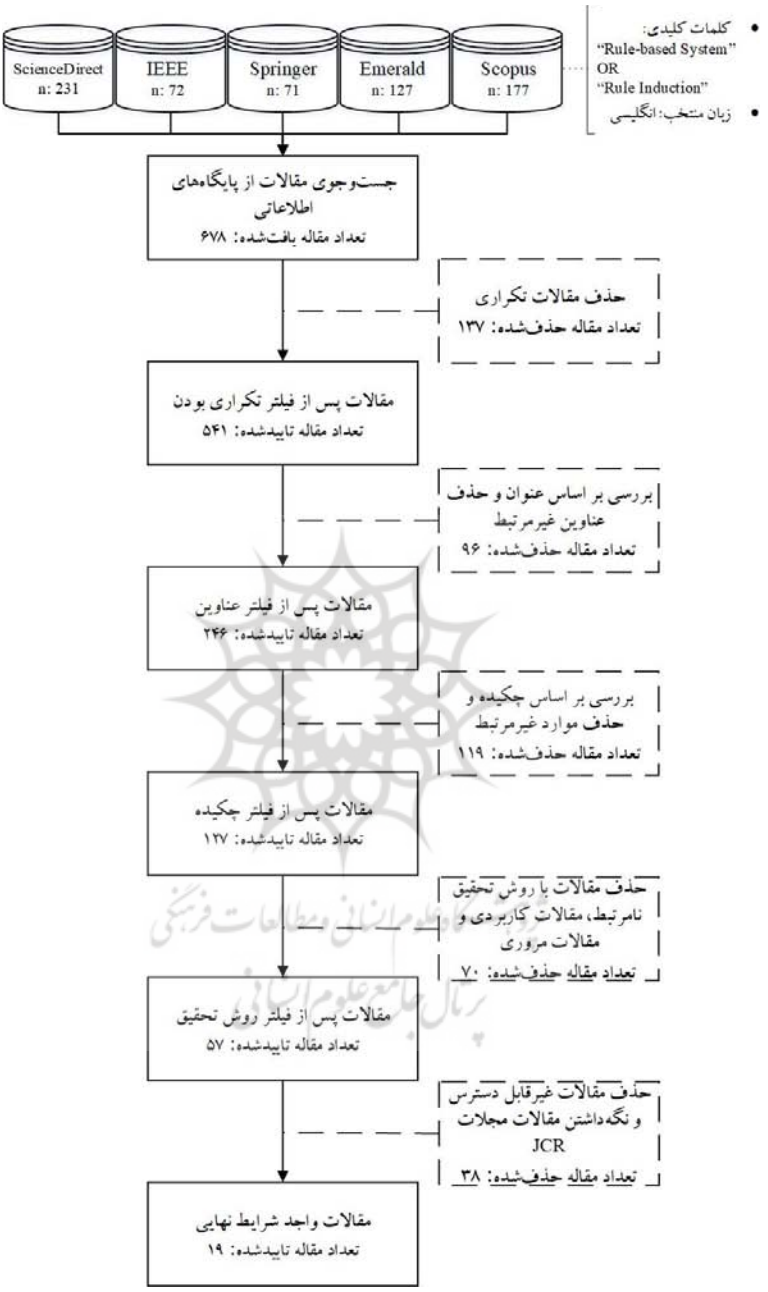
مرتبط با «الگوریتم‌های استخراج قانون» انجام شده است. پایگاه داده‌های مورد استفاده شامل ScienceDirect، IEEE، SPRINGER، Emerald و Scopus بودند. در این پژوهش از کلمات کلیدی «rule-based system» و «rule induction» استفاده شده است. فرایند جست‌وجو بر اساس چارچوب «پریسما» در چهار مرحله اصلی شناسایی، غربالگری، بررسی واجد شرایط بودن، و انتخاب نهایی انجام شد. در مرحله شناسایی، جست‌وجو در پایگاه‌های اطلاعاتی با کلمات کلیدی منتخب در عنوان مقاله، چکیده و کلمات کلیدی صورت گرفته است. محدودیت‌های اعمال‌شده در جست‌وجو عبارت‌اند از: انتخاب زبان انگلیسی برای مطالعات، و بازه زمانی ۲۰۱۵ تا ۲۰۲۴ برای انتشار. محدودیت دیگری اعمال نشده است. بر این اساس، ۶۷۸ مقاله در مرحله اول جست‌وجو شناسایی شد. جزئیات مراحل بعدی در بخش ارزیابی و انتخاب مقالات ارائه شده است

ارزیابی و انتخاب مقالات

ارزیابی و انتخاب مقالات در این پژوهش بر اساس چارچوب «پریسما» در شش مرحله منسجم انجام شد. در **مرحله نخست**، با توجه به محدودیت‌های مشخص‌شده، تعداد ۶۷۸ مقاله از پایگاه‌های اطلاعاتی ScienceDirect، IEEE، Springer، Emerald و Scopus جمع‌آوری گردید. در **مرحله دوم**، مقالات تکراری که در چند پایگاه اطلاعاتی موجود بودند، حذف شدند و در این فرایند ۱۳۷ مقاله کنار گذاشته شد و در نتیجه، ۵۴۱ مقاله باقی ماند. در **مرحله سوم**، با بررسی عناوین مقالات، آن‌هایی که به حوزه داده‌کاوی و موضوع پژوهش مرتبط نبودند، حذف گردید و بدین‌ترتیب، ۲۹۵ مقاله حذف و تعداد مقالات به ۲۴۶ کاهش یافت. در **مرحله چهارم**، چکیده مقالات بررسی شد و مقالاتی که ارتباط کافی با موضوع پژوهش نداشتند، کنار گذاشته شدند و در نتیجه، ۱۱۹ مقاله حذف و ۱۲۷ مقاله باقی ماند. در **مرحله پنجم**، مقالات با روش‌های تحقیق نامرتبط، از جمله مقالاتی که فاقد نوآوری در فرایند توسعه الگوریتم بودند؛ مانند مقالات کاربردی (مقالاتی که صرفاً به استفاده از رویکردهای جاری در یک مورد مطالعه پرداخته‌اند)، و مروری و همایشی حذف شدند و ۷۰ مقاله در این مرحله کنار گذاشته شد و تعداد مقالات به ۵۷ رسید. سرانجام، در **گام ششم**، مقالات غیرقابل دسترس حذف شدند و تنها مقالاتی که در مجلات JCR منتشر شده بودند، برای تضمین کیفیت و اعتبار باقی ماندند. در این مرحله، ۳۸ مقاله دیگر حذف و ۱۹ مقاله نهایی انتخاب شدند. فرایند کامل مرور نظام‌مند شامل تعداد مقالات واردشده و حذف‌شده در هر مرحله، در شکل ۱، نمایش داده شده است

بررسی و تحلیل مطالعات

در این گام با توجه به سؤالات پژوهش، ضمن مطالعه دقیق پژوهش‌های منتخب، انواع رویکردها و راهبردهای استخراج قوانین از داده‌ها، شناسایی و تحلیل شده است. معیارهای بررسی عملکرد قوانین شناسایی و راهبردهای معرفی‌شده بر اساس آن‌ها مورد مقایسه قرار گرفته‌اند. مطالعات منتخب بر اساس محدودیت‌های داده‌ای و نوع داده مورد مطالعه، تحلیل شده و چالش‌های این حوزه معرفی شده است. در ادامه، با توجه به جایگاه هر پژوهش بر اساس معیارهای معرفی‌شده، شکاف تحقیقاتی جاری شناسایی و پیشنهادات پژوهش‌های آتی ارائه شده است. در پایان، مطالعات منتخب با هدف شناسایی الگوی پنهان بین اصطلاحات، مفاهیم و دسته‌بندی موضوعی آن‌ها، با استفاده از ابزار VOS Viewer (ابزار تحلیل متن و محتوا) مورد بررسی قرار گرفته‌اند



شکل ۱. فرایند مرور نظام‌مند پژوهش

۳. تجزیه‌وتحلیل یافته‌ها

پس از مرور نظام‌مند پایگاه‌های اطلاعاتی مورد نظر و انتخاب مقالات منتخب در حوزه «روش‌های استخراج قانون در داده‌کاوی»، در این بخش به تحلیل یافته‌های حاصل از مطالعات پرداخته شده است. این تحلیل در سه بخش اصلی انجام شده است ۱) تحلیل نقادانه راهبردهای استخراج قوانین و معیارهای تحلیل عملکرد آن‌ها، ۲) شناسایی شکاف‌های پژوهشی بر اساس تحلیل‌های صورت‌گرفته با ارائه زمینه‌های مطالعاتی برای تحقیقات آینده، و ۳) تحلیل محتوای متنی مقالات منتخب برای شناسایی مفاهیم کلیدی و ارتباطات میان آن‌ها.

- ابتدا برای سازماندهی تحلیل، ویژگی‌های کلیدی هر پژوهش شامل الگوریتم پیشنهادی، نوآوری‌ها و تعداد ارجاعات (با هدف شناسایی رویکردها با توجه بیشتر در پژوهش‌های آتی)، در جدول ۱، ارائه شده است.

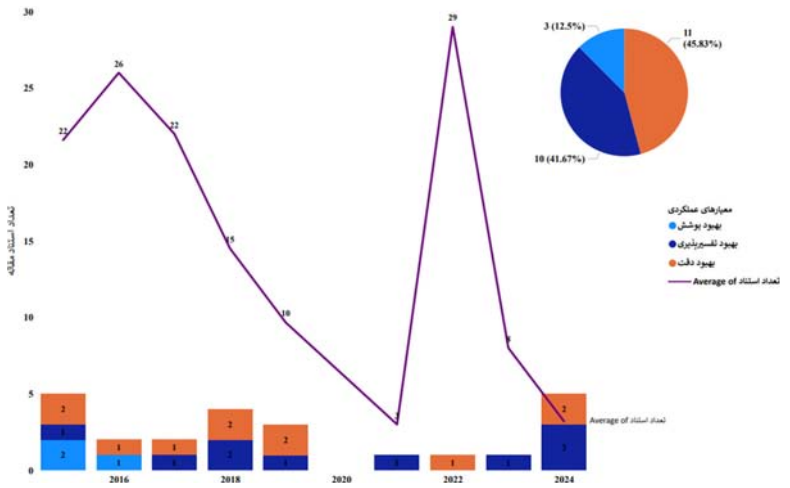
جدول ۱. بررسی مقالات منتخب

ردیف	پژوهش	عنوان پژوهش	الگوریتم پیشنهادی	نوآوری	تعداد ارجاع
۱	Lin, Huang & Che (2015)	Rule induction for hierarchical attributes using a rough set for the selection of a green fleet	HDRIA	◇ بررسی ویژگی‌های تصمیم سلسله‌مراتبی ◇ کاهش پیچیدگی محاسباتی ◇ بهبود دقت و پوشش قوانین ایجادشده	۹
۲	Napierała & Stefanowski (2015)	Addressing imbalanced data with argument-based rule learning	ABMODLEM	◇ بهبود یادگیری قوانین از کلاس‌های اقلیت ◇ بهبود دقت قوانین ایجادشده	۳۱
۳	Chemchem & Drias (2015)	From data mining to knowledge mining: Application to intelligent agents	MIA	◇ تسریع امکان کشف و استنتاج دانش جدید ◇ بهبود تفسیرپذیری قوانین استخراج‌شده	۲۵
۴	Ibarguren et al. (2015)	Coverage-based resampling: Building robust consolidated decision trees	ارائه یک روش نمونه‌برداری و اعمال بر الگوریتم CTC	◇ پوشش‌دهی مناسب زیرنمونه‌ها	۳۴

ردیف	پژوهش	عنوان پژوهش	الگوریتم پیشنهادی	نوآوری	تعداد ارجاع
۵	Asadi & Shahrabi (2016)	ACORI: A novel ACO algorithm for rule induction	ACORI	◇ بهبود عملکرد دسته‌بندی ◇ بهبود دقت قوانین ایجادشده	۳۰
۶	Thabtah, Qabajeh & Chiclana (2016)	Constrained dynamic rule induction learning	eDRI	◇ کاهش فضای جست‌وجو ◇ کاهش تعداد قوانین تولیدشده ◇ افزایش پوشش داده‌ها توسط هر قانون	۲۲
۷	Asadi & Shahrabi (2017)	Complexity-based parallel rule induction for multiclass classification	ERIMC	◇ ارائه روش یادگیری موازی قوانین ◇ بهینه‌سازی ترتیب قوانین ◇ بهبود دقت و تفسیرپذیری قوانین استخراج‌شده	۲۲
۸	Yanjie & Hongbo (2018)	Application of biclustering algorithm to extract rules from labeled data	دوخوشه‌بندی	◇ تولید قوانین قابل تفسیر با ابعاد بالا	۱
۹	Liu & Cocea 2018))	Induction of classification rules by gini-index based rule generation	GIBRG	◇ افزایش دقت طبقه‌بندی ◇ کاهش پیچیدگی طبقه‌بندی کننده‌های مبتنی بر قانون ◇ بهبود تفسیرپذیری قوانین ایجادشده	۲۸
۱۰	Ibarguren et al. (2018)	UnPART: PART without the ‘partial’ condition of it	UnPART_CHD UnPART_C45	◇ کاهش پیچیدگی طبقه‌بندی کننده ◇ بهبود مدل طبقه‌بندی ◇ بهبود دقت قوانین استخراج‌شده	۱
۱۱	Rajab (2019)	New associative classification method based on rule pruning for classification of datasets	APR	◇ بهبود دقت طبقه‌بندی ◇ کاهش تعداد و افزایش تفسیرپذیری قوانین	۴
۱۲	Castellanos-Garzón, Costa & Corchado (2019)	An evolutionary framework for machine learning applied to medical data	RIM-GP	◇ بهبود دقت و تفسیرپذیری ◇ قابلیت تعمیم قوانین	۲۱

ردیف	پژوهش	عنوان پژوهش	الگوریتم پیشنهادی	نوآوری	تعداد ارجاع
۱۳	Breskvar & Džerosk (2021)	Multi-target regression rules with Random Output Selections	FIRE-ROS	◇ بهبود تفسیرپذیری قوانین تولیدشده	۳
۱۴	Fu et al. (2022)	Extended Belief Rule-Based System with Accurate Rule Weights and Efficient Rule Activation for Diagnosis of Thyroid Nodules	AP-EBRB	◇ بهبود دقت و کارایی	۲۹
۱۵	Ma et al. (2023)	A novel rule generation and activation method for extended belief rule-based system based on improved decision tree	Distributed EBRB	◇ بهبود قابلیت تفسیر ◇ مدیریت عدم قطعیت و پیچیدگی‌های سیستم‌های مبتنی بر قانون	۸
۱۶	Fu et al. (2024)	A novel extended rule-based system based on K-Nearest Neighbor graph	Graph-EBRB	◇ شناسایی روابط پیچیده بین داده‌ها ◇ بهبود دقت پیش‌بینی	۱
۱۷	Barut arut, Yildirim & Tatar (2024)	An intelligent and interpretable rule-based metaheuristic approach to task scheduling in cloud systems	IRMTS	◇ انعطاف‌پذیری الگوریتم ◇ بهبود تفسیرپذیری	۱۱
۱۸	Li et al. (2024)	A belief rule-based classification system using fuzzy unordered rule induction algorithm	BRB- FURIA	◇ بهبود عملکرد طبقه‌بندی ◇ کاهش تعداد قواعد ◇ بهبود تفسیرپذیری	۰
۱۹	Hong, Lee & Sim (2024)	Concise rule induction algorithm based on one-sided maximum decision tree approach	OSM Concise algorithm Random forest	◇ افزایش دقت پیش‌بینی ◇ بهبود تفسیرپذیری	۲

بـر اسـاس تحلیـل مقـالات منتخـب مشـخص می‌شـود کـه در حـوزه اسـتخراج قانـون، تمرکـز عمـده پژوهش‌هـا بـر بهبـود معیارهـای عملکـردی ماننـد دقـت، پوشـش و تفسیرپذیـری اسـت. اگرچـه ایـن اهـداف به‌طـور معمـول، در روش‌هـای یادگیـری ماشیـن رایـج هسـتند، تحلیـل مقـالات نشـان‌دهنده تغییـرات قابـل توجـه در اولویـت‌دهی بـه ایـن معیارهـا در زمینـه داده‌کاوی اسـت. بـا توجـه بـه شـکل ۲، افزایـش توجـه بـه تفسیرپذیـری و کاهـش پیچیـدگی قوانیـن به‌عنـوان رونـدی برجسـته در مقـالات مشـاهده شـده اسـت کـه نشـان‌دهنده تلاش بـرای سـازگاری بـا داده‌هـای پیچیـده و چالش‌هـای واقعـی اسـت. همچنیـن، بـا بـررسی آمـاری سـال انتشـار مقـالات منتخـب، گسـتردگی پژوهش‌هـا در ۱۰ سـال اخیـر قابـل مشـاهده اسـت. در سـال ۲۰۱۵، تعـداد قابـل توجهی از مقـالات منتخـب در ایـن حـوزه منتشـر شـده اسـت. ایـن مـورد نشـان‌دهنده توجـه ویـژه و فعالیـت پژوهـشی گسـترده در زمینـه روش‌هـای اسـتخراج قانـون در آن زمـان اسـت. ایـن پژوهش‌هـا بـا بـررسی معیارهـای عملکـردی، سـعی در افزایـش دقـت قوانیـن تولیدشـده داشـته‌اند. پـس از یـک دوره کاهـشی، دوبـاره در سـال ۲۰۲۴، تعـداد مقـالات منتشرشـده افزایـش یافتـه اسـت. پژوهش‌هـای اخیـر در زمینـه افزایـش تفسیرپذیـری قوانیـن توسـعه یافته‌انـد. ایـن رونـد نشـان می‌دهـد کـه بـا افزایـش پیچیـدگی و حجـم زیـاد داده‌هـای امـروزی، ایـن روش‌هـا چالش‌هـای بیشـتری را تجربـه کـرده اسـت و ایـن امـر لـزوم توسـعه روش‌هـای پیشـرفته‌تر و مقیاس‌پذیرتـر را برجسـته می‌کنـد. افـزون بـر ایـن، شـکل ۲، نشـان می‌دهـد کـه اگرچـه تعـداد ارجاعـات بـه مقـالات قدیمی‌تـر بیشـتر بـوده، امـا در سـال‌های اخیـر، تحقیقـات در ایـن زمینـه همچنـان ادامـه دارد و تمرکـز بیشـتر بـر بهبـود دقـت و تفسیرپذیـری روش‌هـای اسـتخراج قانـون در داده‌کاوی اسـت



شکل ۲. تحلیل تعداد ارجاعات و معیارهای عملکردی مقالات منتخب

معیارهای تحلیل رویکردهای استخراج قانون

پس از بررسی‌های انجام‌شده در رویکرد توسعه الگوریتم‌های استخراج قانون، معیارهای «مشخصات داده ورودی، محدودیت نوع داده ورودی، اندازه داده، راهبرد توسعه قانون و عملکرد قوانین» از اهمیت ویژه‌ای برخوردار است. بنابراین، در این پژوهش ضمن بررسی پژوهش‌ها، مدل‌های توسعه‌یافته از منظر این معیارها مورد بررسی قرار گرفته‌اند

الف. مشخصات و محدودیت داده‌های ورودی

ماهیت داده‌های ورودی الگوریتم‌های استخراج قانون در کیفیت قوانین تولیدشده اثرگذار است. به‌طور کلی، داده‌های ورودی برای استخراج قوانین به دو دسته ساختاریافته و غیرساختاریافته تقسیم می‌شوند. **داده‌های ساختاریافته** شامل اطلاعاتی هستند که از یک طرح یا الگوی خاص پیروی می‌کنند (Berrone et al. 2022). این نوع داده‌ها اغلب در قالب یک بانک اطلاعاتی رابطه‌ای و به‌صورت مشخصه‌های اسمی، عددی و باینری در پایگاه‌های داده ثبت و ضبط می‌گردد. بر اساس بررسی‌های انجام‌شده، ۸۹ درصد مطالعات این پژوهش از این نوع داده‌ها برای استخراج قوانین استفاده کرده‌اند. **داده‌های غیرساختاریافته** فاقد ساختار ازپیش‌تعریف‌شده هستند و می‌توانند شامل متن، تصویر، صدا و ... باشند (Khrupovych & Borysova 2021). استخراج قوانین از این نوع داده‌ها چالش‌برانگیزتر است. در مطالعات بررسی‌شده، تنها Chemchem & Drias (2015) و

Ibarguren et al. (2018) استخراج قوانین از داده‌های غیرساختاریافته را پشتیبانی می‌کنند.

همچنین، استخراج قوانین ممکن است از مجموعه داده‌های ناقص، نامتعادل، ابعاد بالا و یا دارای چند ویژگی هدف[1] انجام شود. هر یک از ویژگی‌ها می‌تواند چالشی برای تولید قانون ایجاد کند. یادگیری قوانین از داده‌های نامتعادل، که در آن یکی از کلاس‌ها دارای تعداد نمونه کمتری (کلاس کمینه) نسبت به کلاس‌های دیگر (کلاس بیشینه) است، از چالش‌های حوزه‌های مختلفی مانند کشف تقلب (Whiting et al. 2012)، پزشکی و غیره محسوب می‌شود. چنین وضعیتی بیشتر در دنیای واقعی مشاهده می‌شود و روش‌های طبقه‌بندی‌کنندهٔ استاندارد نیز بیشتر در طبقه‌بندی کلاس‌های بیشینه تمرکز دارند. در نتیجه، نمونه‌های کلاس کمینه به‌طور معمول، به‌اشتباه طبقه‌بندی می‌شوند (Napierała & Stefanowski 2015). به‌طور خاص پژوهش‌هایی مانند Napierała & Stefanowski (2015) و Ibarguren et al. (2015) با ارائه الگوریتم‌های پیشنهادی استخراج قوانین از داده‌های نامتعادل به بهبود این چالش کمک کرده‌اند

در داده‌هایی با ابعاد و تنوع بالا، وجود تعداد زیادی ویژگی نامرتبط، از چالش‌های تولید قانون به شمار می‌آید (Yanjie & Hongbo 2018). از چالش‌های ایجادشده می‌توان به کاهش دقت مدل‌های تحلیلی و افزایش پیچیدگی در فرایند مدیریت داده‌ها اشاره نمود. پژوهش‌های Yanjie & Hongbo (2018) و Castellanos-Garzón, Costa & Corchado (2019) با ارائه الگوریتم‌های دوخوشه‌بندی و RIM-GP به‌ترتیب سعی در بهبود استخراج قوانین از داده‌هایی با ابعاد بالا داشته‌اند

همچنین در بسیاری از مجموعه داده‌ها، چند ویژگی هدف وجود دارد و ساده‌ترین روش مدیریت این مجموعه داده‌ها، پیش‌بینی این ویژگی‌های هدف به‌صورت جداگانه است. اما اگر ویژگی‌ها به هم مرتبط باشند، دقت پیش‌بینی مدل کاهش پیدا می‌کند (Breskvar & Džeroski 2021). در پژوهش Breskvar & Džeroski (2021) با ارائه الگوریتم FIRE-ROS تلاش بر حل این چالش است

ب. اندازه داده

اندازه داده یکی از معیارهای مهم تحلیل در روش‌های استخراج قانون در داده‌کاوی است و نقش به‌سزایی در عملکرد الگوریتم‌های مورد استفاده دارد. این تقسیم‌بندی به‌طور عمده

1. label attribute

بـه عوامـلی ماننـد تعـداد مشـاهدات، تعـداد ویژگی‌هـا و پیچیـدگی سـاختار داده بسـتگی دارد. داده‌هـا از نظـر انـدازه بـه سـه دسـته کوچـک، متوسـط و بـزرگ تقسیـم می‌شـوند کـه هـر یـک نیازمنـد رویکردهـا و راهبردهـای خـاصی بـرای پـردازش و اسـتخراج قوانیـن اسـت

سـهم زیـادی از مطالعـات بـر روی داده‌هـای بـزرگ تـا متوسـط تعلـق دارد. در حـدود ۸۴ درصـد پژوهش‌هـا بـا در نظرگرفتـن کمتریـن هزینـه محاسـباتی و منابـع پـردازشی، سـعی داشـته‌اند تـا داده‌هـای حجیـم را تحلیـل کننـد. ایـن پژوهش‌هـا به‌طـور معمـول، از الگوریتم‌هـای بهینه‌شـده‌ای اسـتفاده کرده‌انـد کـه کارایی بـالایی در تحلیـل داده‌هـای حجیـم دارنـد و پیچیـدگی پـردازش را کاهـش می‌دهنـد

در مطالعـات Napierała & Stefanowski (2015) و Fu et al. (2024) بـه توسـعه الگوریتـم در داده‌هـای بـا انـدازه کوچـک تـا متوسـط پرداختـه شـده اسـت. ایـن پژوهش‌هـا از الگوریتـمی اسـتفاده کرده‌انـد کـه بـر بهبـود دقـت و طبقه‌بنـدی داده‌هـای محـدود تمرکـز دارد. در مطالعـه Yanjie & Hongbo (2018) از داده‌هـای جامـع اسـتفاده شـده اسـت کـه ترکیـبی از داده‌هـای کوچـک، متوسـط و بـزرگ را دربرمی گیـرد. ایـن الگوریتـم به‌صـورت ویـژه بـرای داده‌هـایی بـا ابعـاد بـالا و بهبـود تفسیرپذیـری توسـعه داده شـده اسـت. امـا بـا وجـود ایـن می‌تـوان گفـت کـه یـکی از چالش‌هـای کلیـدی در توسـعه الگوریتم‌هـای اسـتخراج قانـون، همگـرایی بـا داده‌هـای چندمنظـوره و ناهمگـن[1] اسـت. در بسـیاری از مـوارد، داده‌هـا از منابـع مختلـف و بـا سـاختارهای متفـاوت جمـع‌آوری می‌شـوند؛ ماننـد داده‌هـای سـاخت‌یافته، بـدون سـاختار، و داده‌هـای چندرسـانه‌ای. ایـن تنـوع در قالـب، سـاختار، و منبـع داده‌هـا موجـب می‌شـود کـه الگوریتم‌هـای جـاری اسـتخراج قانـون نتواننـد به‌صـورت مؤثـر و کارآمـد بـر ایـن نـوع داده‌هـا اعمـال شـوند

ه. راهبرد توسعه قانون

راهبـرد توسـعه قوانیـن را می‌تـوان بـا رویکردهـای مختلـفی ماننـد «تقسیـم و تسـخیر»[2]، «جـدا و تسـخیر»[3]، «الگوریتم‌هـای حریصانـه»[4]، «یادگیـری مبتـنی بـر اسـتدلال»، «پوشـش مبتـنی بـر نمونه‌گیـری مجـدد»، «یادگیـری ماشیـن ترکیـبی»[5] و غیـره به‌دسـت آورد

1. heterogeneous and multimodal data
2. divide and conquer
3. separate and conquer
4. greedy algorithms
5. ensemble machine learning

رویکـرد «تقسیـم و تسـخیر» همچنیـن به‌عنـوان اسـتقرای بـالا بـه پاییـن درخـت تصمیـم شـناخته می‌شـود. در ایـن روش، مجموعـه‌ای از قوانیـن به‌صـورت یـک درخـت تصمیـم تولیـد می‌شـود. الگوریتم‌هـای معـروفی ماننـد ID3 (Quinlan 1987)، C4.5 (Quinlan 2014) و CART (Breiman 2017) از ایـن رویکـرد اسـتفاده می‌کننـد. در هـر مرحلـه از یادگیـری، یـک ویـژگی انتخـاب می‌شـود و داده‌هـا بـر اسـاس مقادیـر آن ویـژگی تقسیـم می‌شـوند. حـدود ۵۲ درصـد از مطالعـات بررسی‌شـده بـا اسـتفاده از رویکـرد «تقسیـم و تسـخیر» بـه اسـتخراج قوانیـن پرداخته‌انـد.

رویکـرد «جـدا و تسـخیر» به‌عنـوان روش پوشـشی نیـز شـناخته می‌شـود. در ایـن روش، قوانیـن بـه صـورت متـوالی و مسـتقیم از داده‌هـای آمـوزشی یـاد گرفتـه می‌شـوند. پـس از یادگیـری هـر قانـون، نمونه‌هـای پوشـش داده از آن قانـون از مجموعـه آمـوزش حـذف می‌شـوند و یادگیـری قانـون بعـدی بـر روی مجموعـه کوچک‌تـر ادامـه می‌یابـد. الگوریتم‌هـایی ماننـد «پریسـم»[1]، (Cendrowska 1987) از ایـن رویکـرد اسـتفاده می‌کننـد. در پژوهش‌هـای (Asadi & Shahrabi (2016، Asadi & Shahrabi (2017) و Liu & Cocea (2018) از ایـن رویکـرد جهـت اسـتخراج قوانیـن اسـتفاده شـده اسـت. مهم‌تریـن چالـش رویکـرد «تقسیـم و تسـخیر» ایـن اسـت کـه تمرکـز بـر تحلیـل بخش‌هـای کوچـک و مجـزا ممکـن اسـت منجربـه از دسـت رفتـن تصویـر کامـل و درک روابـط و الگوهـای بلندمـدت در داده‌هـا شـود. ایـن رویکـرد ممکـن اسـت قوانیـن ناسـازگار تولیـد کنـد و نتوانـد سـاختار جامـع و کامـل داده‌هـا را نشـان دهـد. در نتیجـه، همبسـتگی‌های بلندمـدت و الگوهـای پیچیـده به‌خـوبی آشـکار نمی‌شـوند و تحلیـل کلی از داده‌هـا کاهـش می‌یابـد

در پژوهـش Kozlov et al. (2019) مفهـوم «سیاسـت حریـص» به‌عنـوان یـک متاالگوریتـم معـرفی شـده اسـت کـه به‌تدریـج یـک راه‌حـل را شـکل می‌دهـد. ایـن موضـوع نمایانگـر اصـل بنیـادی «الگوریتـم حریـص» اسـت؛ جایی کـه بـدون توجـه بـه تصویـر کلی، انتخاب‌هـایی انجـام می‌شـود کـه در لحظـهٔ حاضـر بهتریـن بـه نظـر می‌رسـند. اسـتفاده از «الگوریتم‌هـای حریصانـه» بـرای اسـتخراج قوانیـن در پژوهـش Thabtah, Qabajeh & Chiclana (2016) مشـاهده شـده اسـت. مهم‌تریـن چالـش «الگوریتم‌هـای حریصانـه» در اسـتخراج قوانیـن در داده‌کاوی، بیشـتر مربـوط بـه خطـر کمینه‌سـازی محـلی اسـت. ایـن الگوریتم‌هـا در فراینـد جسـت‌وجو، بـا انتخاب‌هـای سـادهٔ نخسـتین روبـه‌رو می‌شـوند کـه ممکـن اسـت منجربـه یافتـن قوانیـن محـلی و

1. Prism

غفلت از الگوهای بهتر و جامع‌تر شود. به‌گفته دیگر، این رویکرد ممکن است در سطح اولیه بر راه‌حل‌های بهینه محلی تمرکز کند و از کشف قوانین بهینه‌تر و معتبرتر در سطح کلی بازماند. در نتیجه، کیفیت و جامعیت قوانین استخراج‌شده تحت تأثیر قرار می‌گیرد و ممکن است نتایج ناپایدار یا ناقص ارائه دهد

یادگیری مبتنی بر استدلال رویکردی آموزشی است که بر استفاده از استدلال منطقی و استنتاج برای کسب دانش تأکید دارد. این رویکرد در تضاد با رویکردهای دیگر، با هدف توسعه مهارت‌های تفکر انتقادی و ظرفیت یادگیری مستقل از طریق تجزیه‌وتحلیل منطقی عمل می‌کند (Ignatiev et al., 2021). این رویکرد در پژوهش‌های Napierała & Stefanowski (2015) و Barut, Yildirim & Tatar (2024) برای تولید قوانین به‌کار رفته است. مهم‌ترین چالش استفاده از یادگیری مبتنی بر استدلال در استخراج قانون، کاهش شدید دقت، و قابلیت اطمینان قوانین در داده‌های نویزی و ناهمگن است. در این نوع داده‌ها استنتاج‌های منطقی ممکن است نادرست باشد و قوانین فاقد دقت و اعتبار ایجاد کند. همچنین، اجرای این روش‌ها نیازمند منابع و زمان زیادی است

راهبرد پوشش مبتنی بر نمونه‌گیری مجدد، یک روش نمونه‌گیری است که بر اساس توزیع کلاس نمونه‌های آموزشی عمل می‌کند. این رویکرد به‌دنبال تضمین حداقل نمایندگی همه کلاس‌ها هنگام نمونه‌گیری مجدد است (Ibarguren et al. 2015). به‌گفته دیگر، هدف این راهبرد اطمینان از این است که هر کلاس در نمونه‌های انتخاب‌شده به‌اندازه کافی نمایندگی داشته باشد. این رویکرد تولید قانون تنها در پژوهش Ibarguren et al. (2015) مورد استفاده قرار گرفته است. مهم‌ترین چالش این رویکرد در استخراج قانون، دشواری در تضمین کامل بودن و نمایندگی نمونه‌هاست. یعنی ممکن است نمونه‌گیری‌های مجدد نتایج کافی و جامع ارائه ندهند و برخی الگوهای مهم از دست بروند؛ یا نمونه‌ها به‌اندازه کافی متنوع نباشند و قوانین کامل و دقیق استخراج نشود. این موضوع می‌تواند باعث کاهش دقت و قابلیت اعتماد به قوانین استخراج‌شده شود

یادگیری ترکیبی یک روش مؤثر در یادگیری ماشین است که با ترکیب چندین مدل به بهبود دقت پیش‌بینی و افزایش استحکام نتایج کمک می‌کند. این تکنیک از مزایای الگوریتم‌های مختلفی مانند درخت‌های تصمیم، شبکه‌های عصبی و طبقه‌بندی‌کننده‌های آماری بهره می‌برد تا به‌تنهایی خروجی‌ای دقیق‌تر و قابل اعتمادتر نسبت به هر یک

از مدل‌ها ارائه دهد (Zhang & Ma 2012). در مطالعات Fu et al. (2024) و Breskvar & Džeroski (2021) به استفاده از روش یادگیری ماشین ترکیبی به استخراج قانون پرداخته شده است. این دسته از رویکردها ضمن اینکه دارای کارایی قابل توجهی در دقت و قابلیت اطمینان قوانین استخراج شده است، در استخراج قوانین از داده‌های متنوع و پیچیده نیز بهتر از سایر رویکردها عمل می‌کند؛ اما به‌دلیل ترکیب مدل‌ها در استخراج قوانین، دارای پیچیدگی و هزینه محاسباتی بالاتری است و در برخی از موارد قابلیت تفسیرپذیری قوانین نیز به‌واسطهٔ مدل منتخب تحت‌الشعاع قرار می‌گیرد

ح. معیارهای عملکردی

معیارهای ارزیابی را می‌توان به‌عنوان ابزار اندازه‌گیری توصیف کرد که عملکرد طبقه‌بندی‌کننده را اندازه‌گیری می‌کند (Hossin & Sulaiman 2015). در پژوهش‌های مورد بررسی از معیارهای عملکردی گوناگونی برای ارزیابی الگوریتم‌های پیشنهادی استفاده شده است.

از طریق دقت، کیفیتِ راه‌حلِ تولیدشده بر اساس درصد پیش‌بینی‌های صحیح در کل نمونه‌ها ارزیابی می‌شود (ibid). حدود ۴۶ درصد پژوهش‌های بررسی‌شده سعی در بهبود دقت قوانین تولیدشده داشته‌اند

در حالی که دقت، تعداد مواردی را که به‌درستی طبقه‌بندی شده‌اند، اندازه‌گیری می‌کند، پوشش به‌طور خاص بر میزان کاربرد این قانون تمرکز می‌کند (Pham & Afify 2005). پوشش یک قانون به‌عنوان تعداد موارد مثبتی که قانون به‌درستی از مجموعه داده شناسایی می‌کند، تعریف می‌شود. مطالعات Ibarguren et al.، Lin, Huang & Che (2015) (2015) و Thabtah, Qabajeh & Chiclana (2016) به بهبود معیار پوشش در الگوریتم‌های استخراج قانون پرداخته‌اند

تفسیرپذیری یک قانون با تعداد شرایط موجود در قاعده و کیفیت قاعده سنجیده می‌شود. هرچه تعداد شرایط موجود در قانون کمتر باشد، به احتمال زیاد قابل تفسیر است (Hong, Lee & Sim 2024). کیفیت یک قانون به‌طور معمول، با متعادل کردن پوشش آن و دقت آن ارزیابی می‌شود. هدف الگوریتم‌ها بهینه‌سازی یک معیار کیفیت قانون است که هر دو جنبه را دربرمی‌گیرد و تضمین می‌کند که قوانین تولیدشده نه‌تنها دقیق هستند، بلکه قابل اعتماد و قابل استفاده برای بخش قابل توجهی از مجموعه داده‌ها نیز هستند (Maszczyk, Sikora & Wróbel 2024). پژوهش‌های Chemchem & Drias (2015)،

Rajab ،Liu & Cocea (2018) ،Yanjie & Hongbo (2018) ،Asadi & Shahrabi (2017) و Ma et al., (2023) ،Hong, Lee & Sim (2024) ،Breskvar & Džeroski (2021) ،(2019) Barut, Yildirim & Tatar (2024) به بهبود تفسیرپذیری قوانین تولیدشده توسط الگوریتم‌ها پرداخته‌اند

با توجه به موارد ذکرشده می‌توان استنتاج کرد که تمرکز بر دقت در استخراج قوانین ممکن است باعث کاهش پوشش و تفسیرپذیری قوانین شود، زیرا قوانینِ بسیار خاص و دقیق‌شده سبب کاهش دامنه پوشش نمونه‌ها و کاهش قابلیت تفسیرپذیری آن‌ها خواهد شد. برعکس، تأکید بر پوشش‌دهی ممکن است باعث تولید قوانین عمومی‌تر شود که دقت آن‌ها کاهش یابد. همچنین، اگر بخواهیم قوانین بسیار ساده و تفسیرپذیر ایجاد کنیم، ممکن است دقت و پوشش کافی نداشته باشند و نتوانند داده‌های پیچیده یا متنوع را به‌خوبی توضیح دهند. در نتیجه، دستیابی به تعادل بین دقت، پوشش و تفسیرپذیری یکی از چالش‌های اصلی الگوریتم‌های استخراج قانون است، زیرا بیشتر رویکردها به‌دلیل مکانیزم اجرایی خود، تنها بر یک یا دو معیار تمرکز دارند

در این راستا، رویکردهای مختلف تأثیر متفاوتی بر این تعادل دارند. رویکرد «تقسیم و تسخیر» با تولید قوانین سلسله‌مراتبی، بیشتر بر تفسیرپذیری و سادگی تمرکز دارد، اما ممکن است پوشش کافی در داده‌های پیچیده فراهم نکند. «الگوریتم‌های حریصانه» به‌طور عمده، بر بهبود دقت تأکید دارند که ممکن است پوشش‌دهی را تحت‌الشعاع قرار دهند. «یادگیری مبتنی بر استدلال» با اولویت‌دهی به شفافیت و قابلیت تفسیر قوانین عمل می‌کند، اما در مواجهه با داده‌های مخدوش یا پیچیده محدودیت‌هایی دارد. «پوشش مبتنی‌بر نمونه‌گیری مجدد» به‌طور عمده هدفش افزایش پوشش نمونه‌هاست، اما ممکن است دقت یا تفسیرپذیری قوانین کاهش یابد. سرانجام اینکه «یادگیری ماشین ترکیبی» تلاش می‌کند تعادلی بین تمام معیارها برقرار کند، اما در عمل اغلب به دقت و کارایی تمرکز دارد و ممکن است تعادل کامل بین معیارها را محقق نسازد

شناسایی شکاف‌های پژوهشی

هدف از شناسایی شکاف پژوهشی در حوزه استخراج قوانین از داده، شناسایی زمینه‌هایی است که در آن زمینه‌ها کمتر کار شده است. این بررسی ضمن ارائه درک بهتری از وضعیت فعلی، چشم‌اندازی برای جهت‌گیری‌های آینده در این حوزه فراهم می‌آوَرَد. جهت تسهیل شناسایی شکاف پژوهشی، ۱۹ مطالعه مورد بررسی بر اساس معیارهای

«داده‌های ورودی، مشخصات داده‌های ورودی، اندازه داده، راهبرد عملکردی و معیار ارزیابی عملکردی» در جدول ۲، خلاصه شده است

جدول ۲. خلاصه ساختارمند پژوهش‌های صورت‌گرفته در حوزه استخراج قوانین از داده

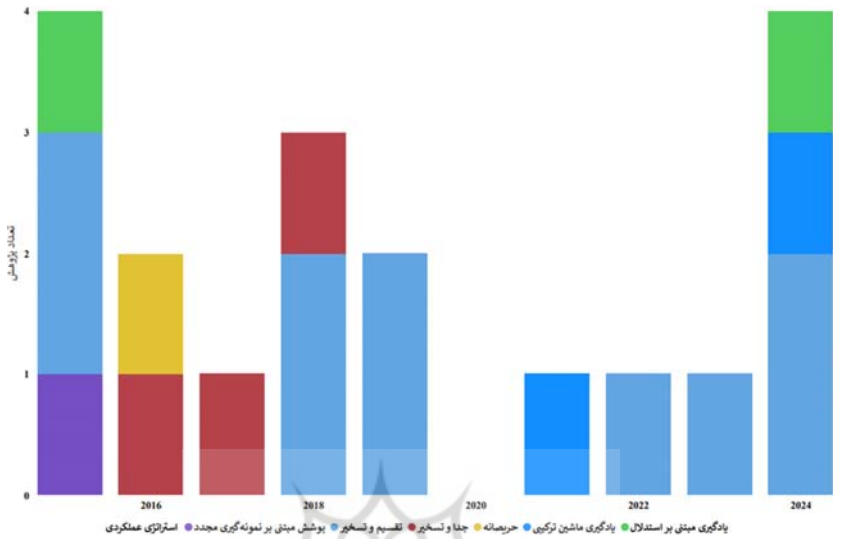
ردیف	پژوهش	نوع داده‌های ورودی	مشخصات داده‌های ورودی	اندازه داده	راهبرد عملکردی	معیارهای عملکردی
۱	Lin, Huang & Che (2015)	ساختاریافته	داده‌های نامعین	بزرگ و متوسط	تقسیم و تسخیر	بهبود دقت و پوشش
۲	Napierała & Stefanowski (2015)	ساختاریافته	داده‌های نامتعادل	کوچک و متوسط	یادگیری مبتنی بر استدلال	بهبود دقت
۳	Chemchem & Drias (2015)	بدون محدودیت	–	بزرگ	تقسیم و تسخیر	بهبود تفسیرپذیری
۴	Ibarguren et al. (2015)	ساختاریافته	داده‌های نامتعادل	بزرگ	پوشش مبتنی بر نمونه‌گیری مجدد	بهبود پوشش
۵	Asadi & Shahrabi (2016)	ساختاریافته	–	بزرگ	جدا و تسخیر	بهبود دقت
۶	habtah, Qabajeh & Chiclana (2016)	ساختاریافته	–	بزرگ و متوسط	حریصانه	بهبود پوشش
۷	Asadi & Shahrabi (2017)	ساختاریافته	–	بزرگ	جدا و تسخیر	بهبود دقت و تفسیرپذیری
۸	Yanjie & Hongbo (2018)	ساختاریافته	داده‌هایی با ابعاد بالا	جامع	تقسیم و تسخیر	بهبود تفسیرپذیری
۹	Liu & Cocea (2018)	ساختاریافته	–	بزرگ و متوسط	جدا و تسخیر	بهبود دقت و تفسیرپذیری
۱۰	Ibarguren et al. (2018)	بدون محدودیت	–	بزرگ و متوسط	تقسیم و تسخیر	بهبود دقت
۱۱	Rajab (2019)	ساختاریافته	–	بزرگ و متوسط	تقسیم و تسخیر	بهبود دقت و تفسیرپذیری
۱۲	Castellanos-Garzón, Costa & Corchado (2019)	ساختاریافته	داده‌هایی با ابعاد بالا	بزرگ	تقسیم و تسخیر	بهبود دقت

ردیف	پژوهش	نوع داده‌های ورودی	مشخصات داده‌های ورودی	اندازه داده	راهبرد عملکردی	معیارهای عملکردی
۱۳	Breskvar & Džeroski (2021)	ساختاریافته	داده‌هایی با چند ویژگی هدف	بزرگ و متوسط	یادگیری ماشین ترکیبی	بهبود تفسیرپذیری
۱۴	Fu et al. (2022)	ساختاریافته	–	بزرگ و متوسط	تقسیم و تسخیر	بهبود دقت
۱۵	Ma et al., (2023)	ساختاریافته	–	بزرگ و متوسط	تقسیم و تسخیر	بهبود تفسیرپذیری
۱۶	Fu et al. (2024)	ساختار یافته	–	کوچک و متوسط	یادگیری ماشین ترکیبی	بهبود دقت
۱۷	Barut, Yildirim & Tatar (2024)	ساختاریافته	–	بزرگ و متوسط	یادگیری مبتنی بر استدلال	بهبود تفسیرپذیری
۱۸	Li et al. (2024)	ساختاریافته	–	بزرگ و متوسط	تقسیم و تسخیر	بهبود تفسیرپذیری
۱۹	Hong, Lee & Sim (2024)	ساختاریافته	–	بزرگ	تقسیم و تسخیر	بهبود دقت و تفسیرپذیری

بررسی جدول ۲، نشان می‌دهد که بیشترین تمرکز تحقیقات بر روی داده‌های ساختاریافته با حدود ۸۹ درصد جامعه آماری بوده است. با این حال، هنوز تحقیقات جامعی در زمینه کاربرد این روش‌ها بر روی داده‌های پیچیده و ساختارنایافته صورت نگرفته است. این شکاف می‌تواند به‌عنوان یک فرصت تحقیقاتی برای آینده مورد بررسی قرار گیرد. همچنین، از چالش‌های دیگر در این حوزه، عدم توجه کافی به توسعه الگوریتم‌هایی است که بتوانند داده‌های جامع را پوشش دهند. این نوع داده‌ها می‌توانند نمایانگر سناریوهای واقعی و متنوع‌تر باشند. چنین داده‌هایی امکان مقایسه دقیق الگوریتم‌های مختلف را از منظر عملکرد در شرایط متفاوت فراهم می‌کنند

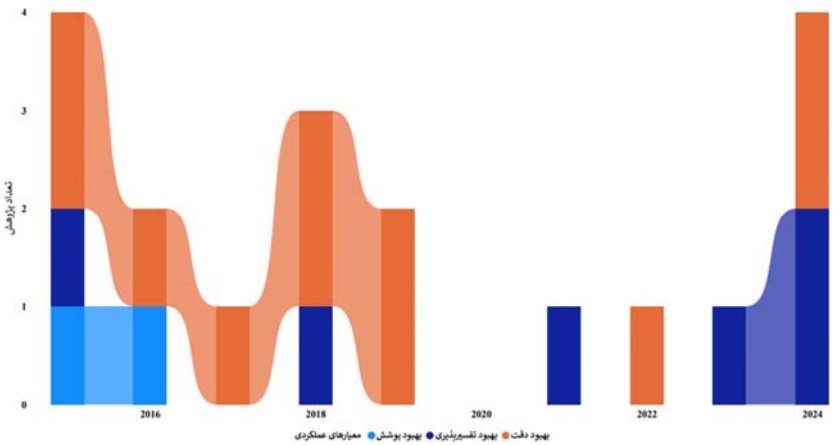
همان‌طور که در شکل ۳، مشاهده می‌شود، راهبردهای استخراج قانون در طول زمان دچار تحول شده‌اند. از سال ۲۰۱۵ به بعد، روش‌های متنوعی مورد استفاده قرار گرفته‌اند، اما به‌تازگی تمایل به استفاده از راهبرد «تقسیم و تسخیر» افزایش یافته است. معرفی روش‌های یادگیری ماشین ترکیبی برای تولید قانون در سال ۲۰۲۱، نشان‌دهنده جست‌وجوی رویکردهای جدید است. توسعه پژوهش‌هایی که بتوانند عملکرد الگوریتم‌های مختلف را در داده‌های جامع و با ویژگی‌های متنوع بررسی کنند، می‌توانند به شناسایی نقاط قوت و

ضعف این راهبردها کمک شایانی کند



شکل ۳. بررسی روند راهبرد عملکردی پژوهش‌ها

شکل ۴، روند تغییر اهمیت معیارهای عملکردی در پژوهش‌های اخیر را نشان می‌دهد. این نمودار تأکید می‌کند که در سال‌های اخیر تمرکز تحقیقات از بهبود دقت مدل‌ها به سمت بهبود تفسیرپذیری آن‌ها تغییر یافته است. به بیان دیگر، پژوهشگران بیشتر به‌دنبال توسعه مدل‌هایی هستند که بتوان آن‌ها را به‌راحتی تفسیر کرد و قوانین استخراج‌شده توسط این مدل‌ها قابل فهم باشند. به‌رغم افزایش اهمیت تفسیرپذیری، هنوز مطالعات کافی بر روی داده‌های پیچیده و واقعی که تفسیرپذیری آن‌ها دشوار است، انجام نشده است



شکل ٤. بررسی روند معیارهای عملکردی پژوهش‌ها

به‌طور خلاصه، می‌توان مهم‌ترین شکاف‌های پژوهشی در الگوریتم‌های استخراج قانون را ناتوانی در مدیریت همزمان سه معیار اساسی عملکرد (دقت، پوشش و تفسیرپذیری) و عدم توجه به داده‌های با تنوع بالا ذکر کرد. به‌طوری که بیشتر رویکردها، یا بر یکی تمرکز بسیار دارند و دیگری را فدای آن می‌کنند، یا در حوزه نمونه‌کاری، مقیاس‌پذیری و تطبیق با داده‌های چندمنظوره، ضعف نشان می‌دهند. همچنین، نبودِ رویکردهای جامع که بتوانند در کنار حفظ سادگی قوانینی قابل فهم، کارایی بالا، پوشش وسیع و تطبیق با داده‌های نامتناهی و ناهمگن را همزمان تضمین کنند، این فاصله تحقیقاتی مهم را برجسته می‌سازد. بنابراین، طراحی الگوریتم‌هایی که نه‌تنها در ابعاد و تنوع داده‌ها مؤثر باشند، بلکه بتوانند معیارهای عملکرد مختلف را در کنار راهبردهای متفاوت، متعادل و همزمان برآورده کنند، یکی از چالش‌های اصلی در این حوزه است.

تحلیل محتوای متنی پژوهش‌ها

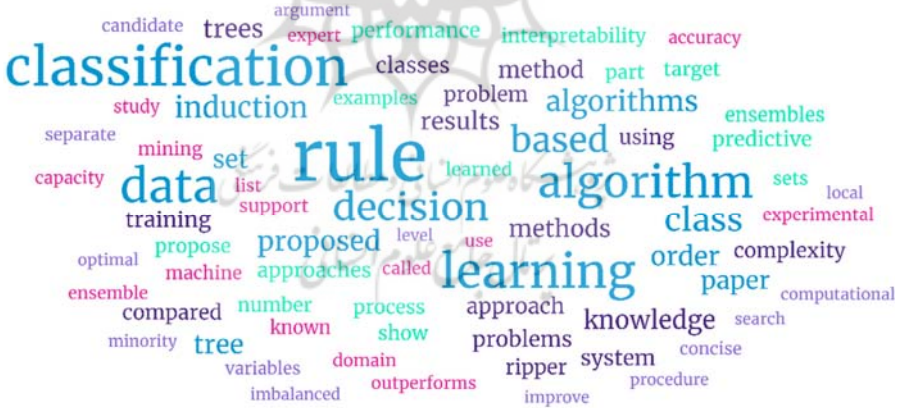
هدف از این بخش، استفاده از روش‌های تحلیل متن و محتوا با هدف شناسایی الگوی پنهان بین اصطلاحات، مفاهیم و دسته‌بندی موضوعی آن‌ها در مطالعات بررسی‌شده است. برای انجام این تحلیل، ابتدا متن چکیده‌های مقالات منتخب جمع‌آوری شد. سپس، فرایند متن‌کاوی با پیش‌پردازش داده‌های متنی آغاز گردید. در این مرحله، متن‌ها توکن‌سازی[1]

1. tokenize

شده و پس از آن، کلمات توقف[1] که فاقد ارزش معنایی قابل توجه در تحلیل هستند، حذف شدند تا تمرکز بر کلمات کلیدی و معنادار حفظ شود. این تحلیل در چهار گام اصلی انجام شد: ۱) استخراج ابرکلمات از چکیده‌های مقالات، ۲) استخراج ابرکلمات از عناوین مقالات، ۳) خوشه‌بندی محتوای چکیده‌ها، و ۴) خوشه‌بندی محتوای کامل مقالات. این گام‌ها امکان شناسایی مفاهیم کلیدی، روابط میان آن‌ها و چارچوب‌های نظری و کاربردی در حوزه استخراج قانون را فراهم کردند

گام اول: تحلیل چکیده‌های مقالات

در گام اول، چکیده‌های مقالات منتخب با استفاده از فرایند متن‌کاوی تحلیل شدند. داده‌های متنی پیش‌پردازش‌شده به‌صورت ابرکلمات مبتنی بر فراوانی تکرار کلمات نمایش داده شدند. در شکل ۵، طیف گسترده‌ای از مفاهیم و روش‌ها در حوزه استخراج قانون نمایان شده است. برجسته‌ترین کلمات «طبقه‌بندی»، «قانون» و «یادگیری» هستند که بیانگر اهمیت این مفاهیم در فرایند پردازش داده است. افزون بر این، وجود کلماتی همچون «دقت»، «کارایی» و «تفسیرپذیری» نشان می‌دهد که پژوهشگران به‌دنبال بهبود عملکرد و قابلیت کاربرد الگوریتم‌های استخراج قانون هستند.

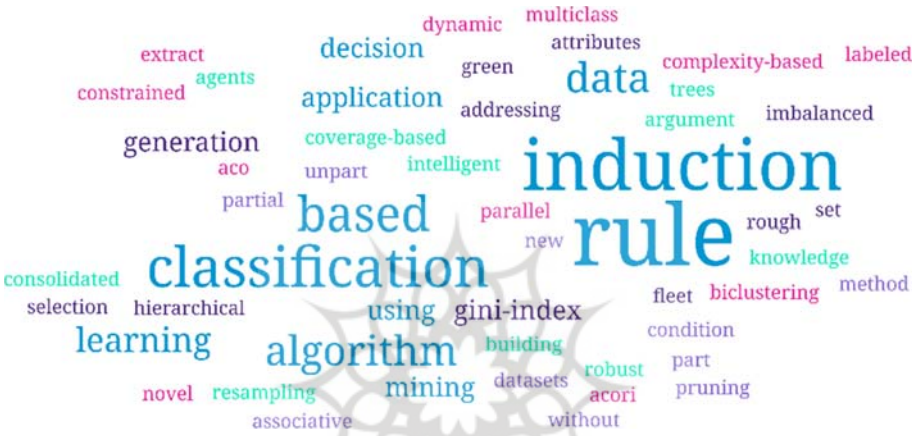


شکل ۵. ابرکلمات استخراج‌شده از چکیده‌های مقالات منتخب

1. stop words

گام دوم: تحلیل عناوین مقالات

در گام دوم، عناوین مقالات منتخب با بهره‌گیری از فرایند متن‌کاوی بررسی شدند. پس از پیش‌پردازش و حذف کلمات توقف، داده‌های متنی به‌صورت ابرکلمات مبتنی بر فراوانی تکرار کلمات ارائه شدند. شکل ۶، ابرکلمات استخراج‌شده از عناوین را نشان می‌دهد که نمای کلی‌تری از موضوعات اصلی پژوهش‌ها ارائه می‌کند. کلمات «استنتاج قانون» و «طبقه‌بندی» برجستگی بیشتری دارند و بیانگر تمرکز اصلی مطالعات بر این مفاهیم هستند.



شکل ٦. ابرکلمات استخراج‌شده از عناوین مقالات منتخب

گام سوم: خوشه‌بندی محتوای چکیده‌ها

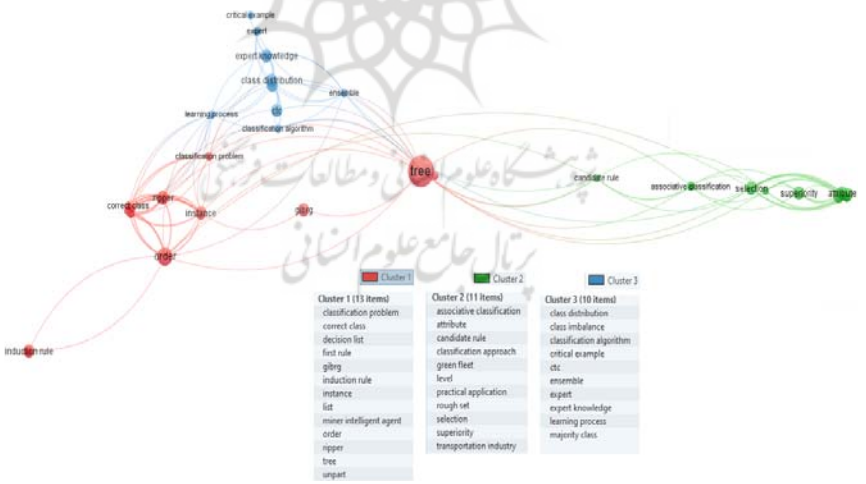
در گام سوم، برای بررسی جامع روش‌های استخراج قوانین، چکیده‌های مقالات منتخب با استفاده از تحلیل متن‌کاوی خوشه‌بندی شدند. این فرایند سه خوشه اصلی از کلمات کلیدی را شناسایی کرد که در شکل ۷، نمایش داده شده‌اند

خوشه اول بر توصیف و شناسایی مشخصات داده‌ها تمرکز دارد که به درک ساختار و ماهیت داده‌های مورد بررسی کمک می‌کند. این خوشه شامل مفاهیمی چون «مشکل طبقه‌بندی»، «نمونه‌ها»، و «لیست تصمیم» است. «لیست تصمیم» مدلی است که شامل مجموعه‌ای از قوانین یا شرایطی است که منجر به یک تصمیم یا طبقه‌بندی خاص می‌شود. هر قانون در فهرست به‌طور معمول، بر اساس مقادیر مشخصه‌های خاص است و فهرست به‌طور متوالی ارزیابی می‌شود تا به یک تصمیم برسد. تفسیر لیست‌های تصمیم‌گیری ساده است و می‌تواند برای پیش‌بینی‌های مبتنی بر داده‌های تاریخی مؤثر باشد (Herbst et al. 2015)

خوشه دوم به معیارهای ارزیابی و انتخاب قوانین استخراج‌شده اختصاص یافته است. مفاهیمی چون «قانون کاندید»، «برتری» و «انتخاب» در این خوشه قرار دارند. «قانون کاندید» یک قانون پیشنهادی است که از تجزیه‌وتحلیل داده‌ها ایجاد می‌شود و هنوز اعتبارسنجی نشده است (Agrawal & Srikant 1994). «برتری» به اثربخشی نسبی یک قانون نسبت به قانون دیگر اشاره دارد که اغلب با معیارهای عملکردی سنجیده می‌شوند (Mining 2006). فرایند «انتخاب» مرتبط‌ترین و مؤثرترین قوانین از میان مجموعه‌ای از قوانین نامزد بر اساس معیارهای ارزیابی خاص است. این فرایند اغلب شامل ارزیابی هر قانون نامزد در برابر معیارهایی مانند تفسیرپذیری است (Hastie 2009)

خوشه سوم شامل مفاهیمی همچون «الگوریتم طبقه‌بندی»، «فرایند یادگیری» و «CTC»[1] است. این خوشه به رویکردها و روش‌های استخراج قانون می‌پردازد. الگوریتم ساختمان درخت تلفیقی (CTC) به‌عنوان یک الگوریتم برای حل یک مسئله طبقه‌بندی است که شامل درجه بالایی از عدم تعادل کلاس تعریف شده است (Ibarguren et al. 2015)

این سه خوشه در مجموع، چارچوبی جامع برای درک و اجرای فرایند استخراج قوانین در پروژه‌های داده‌کاوی ارائه می‌دهند.



شکل ۷. خوشه‌بندی محتوای متنی چکیده مقالات منتخب

1. consolidated tree construction (CTC)

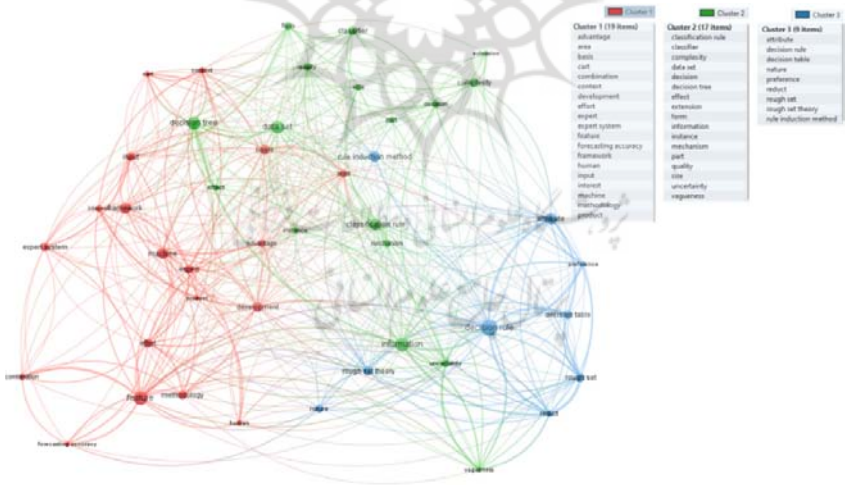
گام چهارم: خوشه‌بندی محتوای کامل مقالات

در گام چهارم، محتوای کامل مقالات منتخب با استفاده از متن‌کاوی خوشه‌بندی شد و سه خوشه اصلی در شکل ۸ نمایش داده شد.

خوشه اول بر وجوه کاربردی و عملیاتی تمرکز دارد. این خوشه شامل مفاهیمی چون «مزیت»، «سیستم خبره»، «روش‌شناسی» و «دقت پیش‌بینی» است. محور اصلی این خوشه، ارزیابی و پیاده‌سازی روش‌های استخراج قوانین در محیط‌های واقعی و عملی است.

خوشه دوم بر جنبه‌های فنی و الگوریتمی متمرکز است. مفاهیم کلیدی این خوشه شامل «قوانین طبقه‌بندی»، «پیچیدگی»، «مجموعه داده» و «درخت تصمیم» است. این خوشه به بررسی و توسعه تکنیک‌ها و روش‌های خاص در فرایند استخراج قوانین می‌پردازد.

خوشه سوم بر مبانی نظری و مفهومی تأکید دارد. این خوشه شامل مفاهیمی چون «ویژگی»، «قانون تصمیم‌گیری»، «ترجیح» و «روش القای قانون» است. تمرکز این خوشه بر اصول بنیادی و چارچوب‌های نظری در پس زمینه استخراج قوانین است



شکل ۸. خوشه‌بندی محتوای متنی مقالات منتخب

ارتباطات متقابل بین خوشه‌ها، که توسط خطوط ارتباطی در نمودار نمایش داده شده، نشان‌دهنده تعاملات و همپوشانی‌های قابل توجه بین این حوزه‌هاست. این امر بر اهمیت اتخاذ رویکردی چندبعدی و یکپارچه در پژوهش‌های این زمینه تأکید می‌کند.

٤. نتیجه‌گیری

در این پژوهش، با بهره‌گیری از روش مرور نظام‌مند مبتنی بر چارچوب «پریسما» و تحلیل‌های متنی و محتوایی با ابزار VOS viewer، تصویری جامع از وضعیت کنونی پژوهش‌ها در حوزه استخراج قانون در داده‌کاوی ارائه شد. تحلیل نظام‌مند ۱۹ مقاله منتخب از میان ۶۷۸ مقاله اولیه که از پایگاه‌های معتبر علمی جمع‌آوری شده بودند، نشان داد که بیشتر مطالعات بر داده‌های ساختاریافته متمرکز بوده و بهینه‌سازی معیارهای دقت و تفسیرپذیری را هدف اصلی خود قرار داده‌اند. این نتایج بر نقش حیاتی قوانین شفاف در کاربردهایی مانند سیستم‌های خبره، پشتیبانی تصمیم و تحلیل رفتار مشتری تأکید دارند

یکی از کلیدی‌ترین شکاف‌ها، عدم توجه کافی به استخراج قوانین از داده‌های غیرساختاریافته، مانند داده‌های متنی، تصویری و صوتی است. این نوع داده‌ها در حوزه‌های واقعی مانند پزشکی و بازاریابی کاربرد گسترده‌ای دارند و نیازمند الگوریتم‌های جدیدی برای پردازش پیچیدگی‌های ذاتی آن‌ها هستند.

همچنین، ناتوانی در برقراری تعادل بین سه معیار اصلی عملکرد (دقت، پوشش و تفسیرپذیری) به‌عنوان یک چالش اساسی شناسایی شد. تحلیل‌های این پژوهش نشان داد که بیشتر روش‌ها بر یک یا دو معیار تمرکز دارند. این امر کارایی قوانین را در سناریوهای متنوع محدود می‌کند. افزون بر این، کمبود پژوهش‌های جامع در زمینه داده‌های ناهمگن و چندمنظوره که نمایانگر سناریوهای واقعی هستند، شکاف دیگری است که توسعه الگوریتم‌های مقیاس‌پذیر را ضروری می‌سازد. این شکاف‌ها همراه با نقاط قوت و ضعف روش‌های استخراج قانون نسبت به سایر روش‌های یادگیری ماشین نشان‌دهندهٔ نیاز به توسعه الگوریتم‌هایی است که بتوانند تعادل بهتری بین دقت، تفسیرپذیری و کارایی در داده‌های پیچیده برقرار کنند. برای عملیاتی‌تر شدن پیشنهادات، توسعه الگوریتم‌های ترکیبی که از تکنیک‌های پیشرفته یادگیری ماشین مانند یادگیری عمیق و شبکه‌های عصبی برای پردازش داده‌های غیرساختاریافته بهره می‌برند، می‌تواند مسیری مؤثر باشد. این الگوریتم‌ها می‌توانند دقت را در داده‌های پیچیده افزایش دهند در حالی که ساختارهای «اگر-آنگاه» شفافیت و تفسیرپذیری را حفظ می‌کنند. همچنین، ابزارهای بصری و تعاملی نیز می‌توانند قوانین را برای کاربران غیرفنی قابل فهم کرده و پذیرش این روش‌ها را در صنایع مختلف تقویت کنند. این پیشنهادات، با هدف افزایش کارایی و کاربردپذیری قوانین، جهت‌گیری آینده پژوهش‌ها را به سمت نوآوری‌های یکپارچه هدایت می‌کنند

خوشه‌بندی متنی چکیده‌ها، عناوین و محتوای کامل مقالات، چارچوبی چندوجهی از مفاهیم کلیدی ارائه داد. خوشه‌بندی چکیده‌ها سه خوشه اصلی را شناسایی کرد ۱) مشخصات داده‌ها، ۲) معیارهای ارزیابی قوانین، و ۳) روش‌های الگوریتمی. در تحلیل محتوای کامل مقالات سه خوشه دیگر پدیدار شدند: ۱) کاربردهای عملیاتی، ۲) جنبه‌های فنی، و ۳) مبانی نظری. این خوشه‌ها ناشی از تمرکز پژوهش‌ها بر ساختار داده‌ها، معیارهای ارزیابی و پیوند نظریه با عمل، تعامل قوی بین ابعاد نظری، فنی و کاربردی را نشان می‌دهند. این ساختار چندوجهی بر ضرورت رویکردهای یکپارچه برای غلبه بر محدودیت‌هایی مانند پردازش داده‌های بدون ساختار تأکید دارد

سرانجام اینکه پژوهش حاضر نشان داد که رویکردهای نوین در زمینه استخراج قانون از داده‌کاوی به‌خوبی توانسته‌اند مشکلات موجود را تا حدودی حل کنند؛ اما نیاز به پژوهش‌های عمیق‌تر و گسترده‌تر برای دستیابی به الگوریتم‌های کارآمدتر و قابل اعتمادتر همچنان احساس می‌شود.

References

Agrawal, R., & R. Srikant. 1994. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB (pp. 487-499).

Asadi, S., & J. Shahrabi. 2016. ACORI: A novel ACO algorithm for rule induction. *Knowledge-Based Systems* 97: 175-187.

_____. 2017. Complexity-based parallel rule induction for multiclass classification. *Information Sciences* 380: 53-73.

Barut, C., G. Yildirim & Y. Tatar. 2024. An intelligent and interpretable rule-based metaheuristic approach to task scheduling in cloud systems. *Knowledge-Based Systems* 284: 111241.

Berrone, S., F. Della Santa, A. Mastropietro, S. Pieraccini, & F. Vaccarino, F. 2022. Graph-informed neural networks for regressions on graph-structured data. *Mathematics* 10 (5): 786.

Breskvar, M., & S. Džeroski. 2021. Multi-target regression rules with Random Output Selections. *IEEE Access* 9: 10509-10522.

Cao, Y., Z. Zhou, C. Hu, W. He & S. Tang. 2020. On the interpretability of belief rule-based expert systems. *IEEE Transactions on Fuzzy Systems* 29 (11): 3489-3503.

Casalino, G., G. Castellano, C. Castiello, V. Pasquadibisceglie & G. Zaza. 2019. A fuzzy rule-based decision support system for cardiovascular risk assessment. Fuzzy Logic and Applications: 12th International Workshop, WILF 2018, Genoa, Italy, September 6–7, 2018, Revised Selected Papers,

Castellanos-Garzón, J. A., E. Costa & J. M. Corchado. 2019. An evolutionary framework for machine learning applied to medical data. *Knowledge-Based Systems* 185: 104982.

Cendrowska, J. 1987. PRISM: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies* 27 (4): 349-370.

Chemchem, A., & H. Drias. 2015. From data mining to knowledge mining: Application to intelligent agents. *Expert Systems with Applications* 42 (3): 1436-1445.

Ding, J., L. Li, H. Peng & Y. Zhang. 2019. A rule-based cooperative merging strategy for connected and automated vehicles. *IEEE Transactions on Intelligent Transportation Systems* 21 (8): 3436-3446.

Fu, C., B. Hou, M. Xue, L. Chang & W. Liu. 2022. Extended belief rule-based system with accurate rule weights and efficient rule activation for diagnosis of thyroid nodules. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53 (1): 251-263.

Fu, Y.-G., X.-Y. Lin, G.-C. Fang, J. Li, H.-Y. Cai, X.-T. Gong & Y.-M Wang. 2024. A novel extended rule-based system based on K-Nearest Neighbor graph. *Information Sciences* 662: 120158.

Hastie, T. 2009. The elements of statistical learning: data mining, inference, and prediction. New York: Springer.

Herbst, K., S. Juvekar, T. Bhattacharjee, M. Bangha, N. Patharia, T. Tei, B. Gilbert & O. Sankoh. 2015. The INDEPTH Data Repository: An International Resource for Longitudinal Population and Health Data From Health and Demographic Surveillance Systems. *J Empir Res Hum Res Ethics*, *10* (3), 324-333. https://doi.org/10.1177/1556264615594600

Hong, J.-S., J. Lee & M. K. Sim. 2024. Concise rule induction algorithm based on one-sided maximum decision tree approach. *Expert Systems with Applications* 237: 121365.

Hossin, M., & M. N. Sulaiman. 2015. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process* 5 (2): 1.

Ibarguren, I., J. M. Pérez, J. Muguerza, I. Gurrutxaga, O. & Arbelaitz. 2015. Coverage-based resampling: Building robust consolidated decision trees. *Knowledge-Based Systems* 79: 51-67.

_____. 2018. UnPART: PART without the 'partial'condition of it. *Information Sciences* 465: 505-522.

Ie, K. S., & T. M. Borysova. 2021. Vykorystannia shtuchnoho intelektu pry marketynhovomu analizi nestrukturovanykh danykh [Use of artificial intelligence in marketing analysis of unstructured data]. *Marketynh i tsyfrovi tekhnolohii* 1: 17-26.

Ignatiev, A., J. Marques-Silva, N. Narodytska & P. J. Stuckey. 2021. Reasoning-based learning of interpretable ML models. International Joint Conference on Artificial Intelligence 2021. Association for the Advancement of Artificial Intelligence (AAAI).

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel & A. Stamatakis. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics 35* (21): 4453-4455. https://doi.org/10.1093/bioinformatics/btz305

Li, Y., I. J. Pérez, F. J. Cabrerizo, H. Garg & J. A. Morente-Molinera. 2024. A belief rule-based classification system using fuzzy unordered rule induction algorithm. *Information Sciences* 667: 120462.

Lin, S.-H., C.-C. Huang & Z.-X Che. 2015. Rule induction for hierarchical attributes using a rough set for the selection of a green fleet. *Applied Soft Computing* 37: 456-466.

Liu, H., & M. Cocea. 2018. Induction of classification rules by gini-index based rule generation. *Information Sciences* 436: 227-246.

Ma, J., A. Zhang, F. Gao, W. Bi & C. Tang. 2023. A novel rule generation and activation method for extended belief rule-based system based on improved decision tree. *Applied Intelligence* 53 (7): 7355-7368.

Maszczyk, C., M. Sikora & L. Wróbel. 2024. Classification, Regression, and Survival Rule Induction with Complex and M-of-N Elementary Conditions. *Machine Learning and Knowledge Extraction* 6 (1): 554-579.

McInnes, M. D. F., D. Moher, B. D. Thombs, T. A. McGrath, P. M. Bossuyt & a. t. P.-D Group. 2018. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 319 (4): 388-396. https://doi.org/10.1001/jama.2017.19163

Mining, W. I. D. 2006. Data mining: Concepts and techniques. *Morgan Kaufinann* 10 (4): 559-569.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman & T. PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151 (4): 264-269.

Napierała, K., & J. Stefanowski. 2015. Addressing imbalanced data with argument based rule learning. *Expert Systems with Applications* 42 (24): 9468-9481.

Ogunleye, A., & Q. G. Wang. 2020. XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 17 (6): 2131-2140. https://doi.org/10.1109/TCBB.2019.2911071

Pham, D. T., & A. Afify. 2005. Rules-6: a simple rule induction algorithm for supporting decision making. 31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005.?

Quinlan, J. R. 1987. Simplifying decision trees. *International Journal of Man-Machine Studies* 27 (3): 221-234. https://doi.org/https://doi.org/10.1016/S0020-7373(87)80053-6

_____. 2014. *C4. 5: programs for machine learning*.?: Elsevier.

Rajab, K. D. 2019. New associative classification method based on rule pruning for classification of datasets. *Ieee Access* 7: 157783-157795.

Siddaway, A. P., A. M. Wood & L. V. Hedges. 2019. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology*, *70* (Volume 70, 2019), 747-770. https://doi.org/https://doi.org/10.1146/annurev-psych-010418-102803

Thabtah, F., I. Qabajeh, F. & Chiclana. 2016. Constrained dynamic rule induction learning. *Expert Systems with Applications* 63: 74-85.

Uthayakumar, J., T. Vengattaraman & P. Dhavachelvan. 2020. Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: An application of bankruptcy prediction and credit risk analysis. *Journal of King Saud University - Computer and Information Sciences*, 32 (6): 647-657. https://doi.org/https://doi.org/10.1016/j.jksuci.2017.10.007

Vale, D., A. El-Sharif & M. Ali. 2022. Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law. *AI and Ethics* 2 (4): 815-826. https://doi.org/10.1007/s43681-022-00142-y

Verbeke, W., D. Martens, C. Mues & B. Baesens. 2011. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications* 38 (3): 2354-2364. https://doi.org/https://doi.org/10.1016/j.eswa.2010.08.023

Verma, L., S. Srivastava & P. C. Negi. 2016. A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data. *Journal of Medical Systems* 40 (7): 178. https://doi.org/10.1007/s10916-016-0536-z

Whiting, D. G., J. V. Hansen, J. B. McDonald, C. Albrecht & W. S. Albrecht. 2012. MACHINE LEARNING METHODS FOR DETECTING PATTERNS OF MANAGEMENT FRAUD. *Computational Intelligence* 28 (4): 505-527. https://doi.org/https://doi.org/10.1111/j.1467-8640.2012.00425.x

Yanjie, Z., & S. Hongbo. 2018. Application of biclustering algorithm to extract rules from labeled data. *International Journal of Crowd Science* 2 (2): 86-98.

Yu, J., H. Lee, Y. Jeong & S. Kim. 2016. Identifying chaff echoes in weather radar data using tree-initialized fuzzy rule-based classifier. 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).?

Zhang, C., & Y. Ma. 2012. *Ensemble machine learning* (Vol. 144). New York: Springer.

سارا انصاری

دانشجوی کارشناسی ارشد مهندسی صنایع، گرایش مدل‌سازی سیستم‌ها و تحلیل داده‌ها در دانشگاه صنعتی اصفهان است.

داده‌کاوی، توسعه آنتولوژی و توسعه سیستم‌های اطلاعاتی از جمله علایق پژوهشی وی است.

صبا صارمی‌نیا

دارای مدرک دکتری مهندسی صنایع-مدیریت فناوری اطلاعات گرایش هوشمندی کسب‌وکار از دانشگاه تربیت مدرس است. ایشان هم‌اکنون عضو هیئت علمی و استادیار دانشکده مهندسی صنایع و سیستم‌ها دانشگاه صنعتی اصفهان است.

داده‌کاوی، توسعه آنتولوژی و نمایش دانش، توسعه سیستم‌های اطلاعاتی، هوشمندسازی کسب‌وکار و یادگیری الکترونیک از جمله علایق پژوهشی وی است