Commercial Surveys Bimonthly
دوماهنامه بررسی‌های بازرگانی
Vol.22, No.129, Feb-Mar 2025, 29-48
سال بیست و دوم، شماره ۱۲۹، بهمن و اسفند ۱۴۰۳، ۴۸-۲۹

Anomaly Detection in Inspection: A Hybrid Data Mining and MCDM Approach

https://doi.org/10.22034/bs.2024.2034058.2986

Jafar Pahlevani, MSc. Student of Socio-economic systems engineering, Faculty of Industrial and Systems engineering, Tarbiat Modares University.
Majid Sheikhmohammady*, Associate Professor, Faculty of Industrial and Systems engineering, Tarbiat Modares University.
Babak Teimourpour, Associate Professor, Faculty of Industrial and Systems engineering, Tarbiat Modares University.

Received: 29 June 2024 Revised: 13 Oct 2024 Accepted: 15 Oct 2024

Anomaly Detection / Complaints / Fraud Detection / Infringements / Inspection

Inspection from guilds and market is one of regulation arms, that underpins its stability and leads to infringements reduction, public trust growth in government and justice. Therefore, Simba inspection software data analysis, in order to detect and prevent inspectors' anomaly patterns, can insure the public trust and effectiveness of inspections. Based on a research gap identified in the research area, 1518 performance data were analyzed to detect anomalies. K-means clustering and classifications by decision tree, logistic regression, naïve bayes and support vector machine were employed to detect fraud, of which decision tree and logistic regression were better than others. Then the results synthesized with 243000 inspection report data analysis. In order to enhance practical side of research, data mining and decision science techniques were employed to find the fraudsters. As a result, collective anomaly detected and nine inspectors were identified as fraudsters. Lastly, IT-based solutions like software redesign and managerial tips were mentioned.

Data Availability
The data used or generated in this research are presented in the text of the article.

Conflicts of interest
The authors of this paper declared no conflict of interest regarding the authorship or publication of this article.

* Corresponding Author: msheikhm@modares.ac.ir

دوماهنامه بررسی‌های بازرگانی
Commercial Surveys Bimonthly
سال بیست و دوم، شماره ۱۲۹، بهمن و اسفند ۱۴۰۳، ۴۸-۲۹
Vol.22, No.129, Feb-Mar 2025, 29-48
مقاله پژوهشی
https://doi.org/10.22034/bs.2024.2034058.2986

کشف الگوی ناهنجاری در بازرسی سطح بازار با راه‌کار ترکیبی داده‌کاوی و تصمیم‌گیری چندمعیاره

دریافت: ۱۴۰۳/۰۴/۹ بازنگری: ۱۴۰۳/۰۷/۲۲ پذیرش: ۱۴۰۳/۰۷/۲۴

جعفر پهلوانی [۲]
مجید شیخ محمدی (نویسنده مسئول)[۱] بابک تیمورپور

چکیده

بازرسی از واحدهای اقتصادی یکی از بازوهای تنظیم بازار جهت برقراری آرامش و ثبات در آن است. کارایی فرایند بازرسی به کاهش جرائم، افزایش اعتماد عمومی به حاکمیت و احقاق حقوق مردم منجر خواهد شد. از این رو بررسی صحت عملکرد بازرسان به کمک داده‌های شکل‌گرفته از بازرسی‌های آنان در سامانه یکپارچه مدیریت بازرسی کشور (سیمبا) و کشف رفتارهای متقلبانه و آسیب‌زا در بازرسی‌ها همچون بازرسی‌های صوری، نقشی کلیدی در صیانت از اعتماد شکل‌گرفته در مردم و اثربخشی این فعل ایفا خواهد کرد. با درک وجود خلاء تحقیقاتی در تحلیل و بررسی داده‌های بازرسی به منظور رصد عملکرد بازرسان و کشف الگوهای رفتاری غیرمتعارف آنان در ثبت نتایج بازرسی‌ها، پژوهش حاضر به دنبال شناسایی الگوی بازرسی‌های صوری و به صورت خاص بازرسی‌های صورت‌گرفته حول شکایات و گزارش‌های مردمی و شناسایی بازرسان متخلف است. کشف رفتار غیرمتعارف و متقلبانه بازرسان با تحلیل ۱۵۱۸ ردیف داده‌ی عملکردی بازرسان با به کارگیری الگوریتم‌های کا-میانگین برای خوشه‌بندی و درخت تصمیم، رگرسیون لجستیک، بیز ساده و ماشین بردار پشتیبان به منظور دسته‌بندی دنبال شد. در ادامه ضمن انجام ارزیابی‌های درونی و بیرونی برای سنجش کیفیت نتایج، از روش‌های علم تصمیم با چاشنی نظر خبرگان برای غنی‌سازی بعد کاربردی پژوهش و حصول لیستی از بازرسان متقلب استفاده گردید که نتیجتاً ۹ بازرس به عنوان بازرسان متخلف شناسایی و ناهنجاری از نوع جمعی تشخیص داده شد. در پایان نیز به منظور پیشگیری از بروز رفتارهای متقلبانه در دورزدن سامانهٔ بازرسی، راه‌کارهای سیستمی و پیشنهادهای مدیریتی ارائه گردید.

طبقه‌بندی JEL: M158، M42 ، P46 ، M48

بازرسی / تخلفات صنفی / شکایات مردمی / کشف تقلب / کشف ناهنجاری

۱. دانشجوی کارشناسی ارشد مهندسی صنایع، سیستم‌های کلان، دانشگاه تربیت مدرس pahlevanijafar@modares.ac.ir
۲. دانشیار دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس msheikhm@modares.ac.ir
۳. دانشیار دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس b.teimourpour@madares.ac.ir

۱. مقدمه: طرح مسأله

دولت‌ها رویکردها و ابزارهای مختلفی برای مدیریت بازار دارند. در ایران یکی از ابزارهای مدیریت و تنظیم بازار، بازرسی است. از اقدامات روزانه جهت پایش بازار، بازرسی‌هایی است که در سطح کشور از اصناف و بازاریان صورت می‌گیرد. به عبارت دیگر، بازرسی‌های صنفی، در کنار قیمت‌گذاری محصولات و سیاست‌گذاری‌های تنظیم بازار، به مثابه‌ی بازویی جهت برقرار آرامش و ثبات در سطح بازار است. در معرفی برخی از نهادهای اصلی در امر بازرسی می‌توان به سازمان حمایت مصرف‌کنندگان و تولیدکنندگان، وزارت جهاد کشاورزی، اتاق اصناف و سازمان تعزیرات حکومتی اشاره نمود. هر کدام از این نهادها، بازرسانی در اختیار دارند که به صورت دوره‌ای و در قالب گشت‌های مشترک یا منفک از هم یا همچنین در قالب طرح‌های نظارتی یا پیرو شکایات/گزارش‌های واصله از مردم، بازرسی از اماکن اقتصادی را ترتیب می‌دهند. سپس صورتجلسه‌های کشف تخلف به شعب دادگاه‌های سازمان تعزیرات حکومتی به عنوان مرجع رسیدگی تحویل داده شده و حکم مقتضی توسط آن نهاد صادر می‌شود.

به گفته‌ی فرشاد مقیمی مدیرعامل سازمان صنایع کوچک و شهرک‌های صنعتی ایران، سه میلیون واحد صنفی فعال در کشور وجود دارد (مقیمی، ۱۴۰۲). تا سال ۱۳۹۹ بازرسی‌ها از مجموعه‌ی واحدهای اقتصادی به صورت دستی (کاغذی) و غیر سیستمی (سامانه‌ای) در کشور دنبال می‌شده است. شیوه‌ای که جدا از بهره‌وری پایین و کاهش اعتماد عمومی، می‌توان ضعف‌ها و مشکلات زیر را برای آن برشمرد:

- عدم وجود داده‌های صحیح و دقیق از بازرسی‌ها
- عدم امکان برنامه‌ریزی و تعیین سیاست برای آینده‌ی بازرسی‌ها به کمک داده‌ها به عنوان ابزاری مهم و اساسی
- عدم امکان ارزیابی صحیح و دقیق از عملکردهای فردی و استانی در بازرسی و بروز ناعدالتی اداری
- بروز فساد در بازرسی‌ها و تشریفاتی شدن آن‌ها
- بروز ناعدالتی در رسیدگی به شکایات مردمی از تخلفات اصناف

به منظور افزایش اثربخشی بازرسی‌ها از سال ۱۳۹۹ سامانه یکپارچه مدیریت بازرسی کشور توسط سازمان حمایت مصرف‌کنندگان و تولیدکنندگان، زیرمجموعه وزارت صنعت، معدن و تجارت (صمت)، ایجاد و راه‌اندازی شد. در این سامانه، بازرسی از ثبت گزارش‌ها تا ارسال آن به سازمان تعزیرات و دریافت حکم به صورت الکترونیکی دنبال می‌شود. همچنین سامانه با اتصال برخط به پایگاه‌های اطلاعاتی دیگر و نیز تشکیل پایگاه‌های داده‌ای عظیم از بازرسان، واحدهای اقتصادی، تخلفات، گروه‌های کالایی و... داده‌های خام ارزشمندی را گردآوردی کرده است که کشف دانش از آن بی‌شک بسیار مهم و ارزشمند خواهد بود. بر اساس تجربیات ۳ ساله، داده‌های گردآوری شده تا کنون و به منظور ارتقاء بهره‌وری بازرسی‌ها و ملموس‌ترشدن نتایج آن، این نیاز احساس شد تا الگوی رفتاری متقلبانه‌ی بازرسان در بازرسی و دورزدن فرایند سیستمی آن کشف شود که به بیانی واضح‌تر می‌توان مساله را کشف الگوی رفتاری بازرسین متخلف در ثبت گزارش‌های صوری، شناسایی ویژگی‌های برجسته در کشف این الگو و در نهایت راهجویی سیستمی برای مقابله با آن عنوان کرد. در این مقاله تمرکز بر روی بازرسی‌های پیرو شکایات مردمی بوده است و به این ترتیب داده‌های شکایات مردمی، گزارش‌های بازرسی پیرو آن‌ها و عملکرد بازرسان در حیطه‌ی رسیدگی به گزارش‌ها و شکایات مردمی تحلیل شده‌است.

به فراخور داده‌های موجود، دو رویکرد (۱) تحلیل داده‌های پرونده‌های بازرسی و (۲) تحلیل داده‌های عملکردی بازرسان مطرح می‌شود. به رویکرد اول در مقاله‌ی (پهلوانی و همکاران، ۱۴۰۳) پرداخته شد. در این مقاله رویکرد دوم و ترکیب آن با رویکرد اول دنبال می‌شود.

در ادامه مرور ادبیات، روش تحقیق و نتایج تحقیق مشتمل بر خوشه‌بندی و کشف خوشه‌ی بازرسان متقلب ارائه می‌گردد. در بخش پایانی نیز، به راه‌کارهای مدیریتی و سیستمی به منظور مقابله با تقلب اشاره خواهد شد.

۲. مبانی نظری و پیشینهٔ پژوهش

داده‌کاوی علم اکتشاف و تحلیل عظیم داده‌ها به منظور یافتن الگوها و قواعد از آن‌ها است. روش‌های داده‌کاوی معمولاً به دو دسته‌ی نظارت‌شده و نظارت‌نشده تقسیم می‌شوند. خوشه‌بندی روشی نظارت‌نشده و مبتنی بر حداکثر کردن شباهت داخل هر خوشه و حداقل‌کردن آن بین خوشه‌ها است. بخش اعظمی از تحلیل داده‌ها در دنیای امروز معطوف به داده‌های با رفتار عادی و منتظره است. در این پژوهش اما با توجه به هدفی که توسط نویسندگان آن دنبال می‌شود، تمرکز بر داده‌های ناهنجار است. ناهنجاری‌ها اقلیتی از داده هستند که رفتاری غیرمنتظره و خارج از روند عادی داده‌های دیگر از خود نشان می‌دهند (کندولا و همکاران، ۲۰۰۹). در ادبیات داده‌کاوی داده‌های پرت بیشتر از داده‌های ناهنجار به کاربرده می‌شود و غالباً این دو دسته یکی گرفته می‌شوند حال آنکه تفاوت‌هایی با یکدیگر دارند. تفاوت اصلی داده‌ی پرت با داده‌ی ناهنجار، آن است که دادهٔ پرت رخدادی غیرعادی و قابل انتظار است؛ اما، دادهٔ ناهنجار رخدادی غیرعادی ولی غیرقابل انتظار است (آلا و کایلان آداری، ۲۰۱۹). همانطور که در سطور پیشین اشاره شد، مقاله حاضر به بررسی داده‌های عملکردی بازرسان در کشف تقلب پرداخته و سپس با ترکیب نتایج آن با نتایج رویکرد نخست (پهلوانی و همکاران، ۱۴۰۳)، تحلیلی جامع از وجود تقلب و ناهنجاری در سامانه بازرسی به همراه راه‌کارهای سیستمی و مدیریتی به منظور رفع و پیشگیری ارائه می‌نماید. به بیانی دیگر، در مقالهٔ رویکرد اول شکایات مردمی با هدف شناسایی شکایات صوری رسیدگی‌شده، خوشه‌بندی شده و خوشهٔ ناهنجار مشخص شد. در رویکرد دوم - که رویکرد مقاله حاضر است - تمرکز بر روی عملکرد بازرسان دخیل در رسیدگی به کل شکایات و تشخیص خوشه‌ی ناهنجار یا متقلب بازرسان و سپس تلفیق نتایج دو رویکرد برای تدقیق نتیجه و احصاء بازرسان متقلب است. بدین منظور تحقیقات متعدد مرتبط با شناسایی تقلب در عملکرد بازرسان بر اساس داده‌های سامانه و لاگ کارکردن آن‌ها با سامانه بررسی و غربال گردید که نهایتاً ۱۴ مقاله انتخاب شد.

جدول ۱- مقالات کشف ناهنجاری و تقلب در عملکرد بازرسان

مقاله	هدف یا نوآوری	نتایج
(سُه و همکاران، ۲۰۱۹)	توسعه یک چارچوب از نمایه کارکنان جهت تشخیص ناهنجاری در عملکرد انسان-دستگاه آنان	داده‌های بیومتریک کارکنان از طریق الگوریتم کشف ناهنجاری مورد بررسی قرار گرفته و می‌تواند ۵ نفر نخست از نظر بیشترین ناهنجاری را ارائه نماید.
(آکپینار و همکاران، ۲۰۲۱)	کشف ناهنجاری عملکردی کارکنان دورکار به کمک نرم‌افزار هوشمند	نرم‌افزار "رادار هوشمند" می‌تواند با داده‌هایی که از عملکرد کارکنان ذخیره می‌کند، با دقت ۸۹ درصد، قشر نامتعارف آنان را شناسایی کند.
(زوفک و همکاران، ۲۰۲۱)	توسعه مدلی نظارت‌نشده برای کشف ناهنجاری‌ها در سوابق کار با سامانه	مدل توسعه یافته بر پایه الگورتیم اُپتیکس با امتیاز fscore ۸۳ درصد می‌تواند ناهنجاری‌های سیستم را شناسایی کند.
(رنجان و کومار، ۲۰۲۲)	ارائه تکنیکی جهت تفکیک کاربر خوب از بد	مدل توسعه داده شده با جنگل تصادفی با دقت ۶۰ تا ۷۰ درصد موفق به تشخیص کاربران با رفتارهای ناهنجار شد.

مقاله	هدف یا نوآوری	نتایج
(دُماشووا و همکاران، ۲۰۲۲)	توسعه مدلی جهت شناسایی رفتارهای مجرمانه کارکنان در فرایندهای مالی سازمان	الگوریتم بگینگ متعادل‌شده با دقت ۹۸ درصد، بهتر از سایر الگوریتم‌ها می‌تواند رفتارهای مجرمانه را شناسایی نماید.
(آمرگا هیچاواریا و شفیق، ۲۰۲۲)	توسعه مدلی جهت پیش‌بینی رفتار کاربران یک سامانه آموزش از راه دور به کمک داده‌های کاربری کاربران (اعم از کلیک و...)	مدل توسعه یافته مبتنی بر روش شبکه بیزی می‌تواند رفتارهای کاربران یک سامانه یادگیری از راه دور را کسب fscore ۰٫۶۵ شناسایی نماید.
(جین و همکاران، ۲۰۲۲)	توسعه چارچوبی جهت شناسایی داده‌های ناهنجار در شبکه	مدل توسعه یافته بر پایه روش دسته‌بندی بردار پشتیبان، با امتیاز fscore ۹۳ درصد قابلیت شناسایی ناهنجاری‌های سیستم را داراست.
(هان و همکاران، ۲۰۲۳)	چارچوبی جهت شناسایی داده‌های ناهنجار به کمک الگوریتم‌های نایوبیز، کا-نزدیک‌ترین همسایه، جنگل تصادفی، خودرمزگذار و خودرمزگذار عمیق با حافظه اضافه‌شده	از بین الگوریتم‌های مد نظر، الگوریتم خودرمزگذار عمیق با حافظه اضافه‌شده با بهترین امتیاز fscore یعنی ۱٫۰۰، نسبت به سایرین عملکرد بهتری در شناسایی داده‌های ناهنجار داشت. همچنین استفاده از تکنیک‌های آداسین و اِسموت برای بهبود عملکرد مدل‌ها موفقیت آمیز بود.
(رامیرز و همکاران، ۲۰۲۳)	توسعه مدل شناسایی رفتارهای عادی و مجرمانه	مدل ابداعی با دقتی بالغ بر ۷۵٪ بر پایه‌ی دو درخت تصمیم، که می‌تواند داده‌های عادی را از ناهنجار تمیز دهد.
(خان و همکاران، ۲۰۲۳)	ارائه راه‌کاری جهت شناسایی ناهنجاری‌ها در سامانه از طریق سوابق عملکرد در آن	مدل با کسب امتیاز بین ۷۱ درصد تا ۸۶ درصد معیار fscore، قابلیت شناسایی ناهنجاری‌ها در عملکرد سامانه را از طریق سوابق آن داراست.
(لاندر و همکاران، ۲۰۲۳)	مقاله مروری	مقاله مروری با رویکرد جامعیت ادبیات مطالعه شد.
(لی و همکاران، ۲۰۲۳)	جهت شناسایی رفتارهای ناهنجار در سامانه	مدل توسعه یافته موسوم به لانوبرت نسبت به سایر مدل‌های نظارت نشده‌ی موجود از قابلیت بالاتری برای کشف ناهنجاری‌های سوابق عملکرد در سامانه برخوردار است.
(ایانی و ماسکیاری، ۲۰۲۳)	توسعه مدلی جهت شناسایی تهدیدات امنیت سایبری به واسطه سوابق کار با سامانه	مدل توسعه داده شده قابلیت کار با داده‌های بسیار حجیم را دارد به شکلی که نقطه ضعف سایر مدل‌ها مبنی بر گران و پرهزینه بودن را ندارد. همچنین مدل یادشده به شکل نظارت نشده و بدون دخالت انسان در تعیین تعداد خوشه‌ها، اقدام به خوشه‌بندی داده‌ها می‌نمایند.
(راچراچه و همکاران، ۲۰۲۳)	توسعه رویکردی نوین در استفاده از یادگیری ماشین برای پایش ترافیک برخط برای تشخیص رفتارهای ناهنجار کاربران شبکه بدون نیاز به نرم‌افزار	استفاده از روش‌های دسته‌بندی درخت تصمیم، جنگل تصادفی، شبکه عصبی پیشرو، ال‌اِس‌تی‌اِم، شبکه‌های عصبی پیچشی، یادگیری گروهی برای کلاس‌بندی نمایه‌های کاربری و سپس تحلیل نمایه‌ها با روش‌های ماشین بردار پشتیبان تک‌کلاسه، جنگل انزوا، کوواریانس استوار

اکثریت پژوهش‌ها از روش‌های نظارت‌نشده برای تحلیل داده‌های رویدادنگاری‌شده استفاده کرده‌اند. علت استفاده از روش‌های نظارت‌نشده در این حوزه این است که داده‌هایی که برای تحلیل استفاده می‌شود، رخدادنگاره‌ها (لاگ‌ها)، داده‌های خام و دسته اول، هستند و نیاز است تا حتماً خوشه‌بندی روی آن‌ها صورت گیرد. در بین الگوریتم‌ها کا-میانگین، درخت تصمیم، جنگل ایزوله، جنگل تصادفی عادی و متعادل شده و خانواده‌ی ماشین بردار اعم از عادی و تک‌کلاسه هر کدام سه بار در مقالات استفاده شده‌اند که در رتبه نخست قرار می‌گیرند. سپس روش دی‌بی‌اِسکن و رگرسیون لجستیک بیشتر مورد استقبال قرار گرفتند. لاندر و همکاران در بررسی ۶۲ مقاله‌ی کشف ناهنجاری داده‌ها، برخی از چالش‌های کشف الگوهای غیرمتعارف را بیان کردند که عدم پایداری، نامتعادل بودن و درهم‌آمیختگی یا ترکیب داده‌ها از آن‌ها هستند. همچنین در ۳۶ مورد از ۶۲ مورد پژوهش مورد بررسی آنان، از روش شبکه عصبی بازگشتی استفاده کرده بودند که اکثریت قابل توجه این ۳۶ مورد به صورت ویژه از روش اِل‌اِس‌تی‌اِم بهره برده بودند. سپس شبکه عصبی پیچشی و در نهایت خودرمزگذارها در شمار پرتکرارترین روش‌های به کاررفته برای کشف ناهنجارها بودند. همچنین از میان انواع ناهنجاری‌های یادشده، اکثریت مقالات، ناهنجاری‌های متوالی (ناهنجاری‌هایی که حاصل تغییر رویه‌ها است) را مورد بررسی قرار داده بودند (لاندر و همکاران، ۲۰۲۳).

۳. روش‌شناسی تحقیق

روش تحقیق این پژوهش از نوع اسنادی و با تکیه بر روش‌های داده‌کاوی جهت کشف دانش مورد نظر است. به علت آنکه این تحقیق از روش داده‌کاوی انجام می‌پذیرد، پژوهش فاقد فرضیه است. شایان ذکر است داده‌های این پژوهش از نوع داده‌های ثانویه‌ی آرشیوی بوده و به صورت کامل مورد بررسی قرار خواهند گرفت. فرایند کشف دانش محوریت انجام داده‌کاوی بوده است.

در شناسایی الگوی ناهنجاری در بازرسی‌های پیرو شکایات مردمی، اساساً دو رویکرد مطرح است که در مقدمه بدان اشاره شد. در این پژوهش ماحصل ترکیب دو رویکرد ارائه شده‌است. فرایند کلی تحقیق به این صورت است که پس از اخذ داده‌های عملکردی بازرسان فعال در امر بازرسی پیرو شکایات مردمی، عمل پاک‌سازی انجام و ویژگی‌های مفید برای خوشه‌بندی با کا-میانگین و تشخیص خوشهٔ ناهنجار، تعیین شد. در ادامه برای دسته‌بندی از روش‌های ماشین بردار پشتیبان، درخت تصمیم، رگرسیون لجستیک و گوسین نایوبیز استفاده شد. اما آنچه قوام بیشتری به نتایج این پژوهش می‌بخشد، ارزیابی نتایج آن با شواهد در دسترس است چراکه یکی از محصولات تحقیق حاضر، احصاء لیستی از بازرسان متخلف است. به علت حساسیت این کار لازم است تا علاوه بر ارزیابی نتایج کمی، دانش کشف‌شده از تحلیل گزارش‌های بازرسی نیز دخیل شود. به این منظور در گام بعدی نتایج داده‌کاوی ارزیابی شدند. در نهایت نیز پیشنهادهای اصلاحی ارائه شد. نمودار (۱) فرایند کلی تحقیق را به تصویر می‌کشد.

نمودار۱- فرایند کلی تحقیق

پیش‌پردازش داده‌ها

داده‌های بازرسی کل کشور از تاریخ ۱۴۰۱/۰۲/۲۴ تا ۱۴۰۲/۰۲/۱۷ که در سامانه یکپارچه مدیریت بازرسی گردآوری‌شده بود، مورد استفاده قرار گرفت. در مرحله پیش‌پردازش، ۴۸۴ داده اصلا مربوط به شکایات نبوده از این رو حذف شدند. همچنین اقدامگر یا مسئول ۱۵۲۰۴ شکایت در داده‌ی خام، خالی بود. با بررسی دقیق و استفاده از خروجی‌های اکسل متعدد، مسئول یا بازرس حقیقی آن‌ها کشف و جایگزین گردید. علی القاعده انتظار می‌رود که اگر شکایتی به بازرسی محول شد، فرم بازرسی که منعکس‌کننده نتیجه بازرسی حول آن شکایت است نیز به نام وی ثبت شده باشد. در یک مرحله از بررسی داده‌ها مشاهده شد که برای ۱۱۹۳۹ ردیف شکایت، تناقضی بین بازرسی که شکایت به آن محول شده و بازرسی که بازرسی انجام داده، وجود دارد. داده درست این مقوله نیز احصاء و جایگزین گردید. در یک مورد دیگر، برای ۱۳۳۶۹ اختلاف زمانی تاریخ ثبت فرم بازرسی پیرو شکایت از تاریخ تخصیص شکایت به بازرس، جایگزین اختلاف زمانی تاریخ عملیات بازرسی پیرو شکایت از تاریخ تخصیص شکایت به بازرس شد. علت آن بود که تاریخ عملیات بازرسی یا به عبارتی دیگر تاریخ بازرسی به صورت دستی قابل انتخاب است. از این رو در حدود ۱۳ هزار فقره بازرسی، تاریخ بازرسی زودتر از تاریخ تخصیص شکایت به بازرس برای بازرسی حول آن بوده است و این سبب منفی شدن اعداد یکی از مهم‌ترین ویژگی‌های تحلیل برای پرونده‌ها شده بود. از این رو، برای آن‌ها اختلاف زمانی تاریخ ثبت فرم بازرسی پیرو شکایت از تاریخ تخصیص شکایت به بازرس، جایگزین شد. البته توضیح دقیق و مفصل تمامی پاک‌سازی‌ها و جایگزین‌کردن داده‌ها در این مقال نمی‌گنجد؛ اما، مواردی به صورت خلاصه ذکر گردید.

ویژگی‌های تحلیل

پس از پاکسازی داده‌ها، انتخاب ویژگی‌ها برای تحلیل در دستور کار قرار گرفت. یکی از پیچیدگی‌های پژوهش حاضر، عدم وجود پژوهش مشابه به علت تفاوت مدل حکمرانی اقتصادی کشورها است. بدین صورت که اساساً مفهوم بازرسی در بازار به شکل موجود در ایران، در بسیاری از کشورهای دیگر موضوعیت ندارد؛ لذا، الگوگیری از منابع علمی با بن‌بست مواجه می‌شود. هیچ یک از پژوهش‌ها و مقالات پیشین مرورشده، در تعیین ویژگی‌ها راهگشا واقع نشده و ویژگی‌های مربوط به گزارش‌های بازرسی حول شکایات بر اساس نظرات خبرگان و فهم کسب و کار، استخراج گردید. در ادامه با کمک نظرات خبرگان، تعمیم ویژگی‌های گزارش‌های حول شکایات برای عملکرد بازرسان و نیز برداشت از مقالهٔ (دُماشووا و همکاران، ۲۰۲۲) که مشخصاً به ویژگی‌های تحلیل در خود اشاره نموده، ویژگی‌های زیر برای بازرسان طرح گردید:

- میانگین فاصله زمانی تاریخ عملیات بازرسی پیرو شکایت از تاریخ محول کردن شکایت به بازرس (بر حسب روز)
- میانگین تعداد بازرسان همکارگزیده در بازرسی‌هایش
- سرانه انجام عملیات بازرسی حول شکایت در یک روز
- درصد عدم کشف تخلف در شکایات (درصد ثبت نتیجه عدم احراز تخلف برای بازرسی شکایت)
- تعداد استفاده از اولین دلیل عدم احراز در بازرسی‌ها
- تعداد استفاده از دومین دلیل عدم احراز در بازرسی‌ها
- تعداد استفاده از سومین دلیل عدم احراز در بازرسی‌ها
- تعداد استفاده از چهارمین دلیل عدم احراز در بازرسی‌ها
- تعداد استفاده از پنجمین دلیل عدم احراز در بازرسی‌ها
- تعداد استفاده از ششمین دلیل عدم احراز در بازرسی‌ها

- میزان سابقه عضویت در سامانه
- سازمان بازرس
- استان محل خدمت بازرس

روش خوشه‌بندی

خوشه‌بندی با روش کا-میانگین صورت گرفت که در پردازش داده‌های حجیم، سریع و موفق است.

کا-میانگین

الگوریتم کا-میانگین به عنوان پرکاربردترین الگوریتم خوشه‌بندی تفکیکی مشخص شده است. این الگوریتم اولین بار توسط مک کوییین در سال ۱۹۶۷ ارائه شد. این الگوریتم در عین سهولت، یک الگوریتم اصلی و پایه برای تعداد زیادی از الگوریتم‌های خوشه‌بندی دیگر (به طور مثال خوشه‌بندی فازی) در نظر گرفته می‌شود. این الگوریتم، به عنوان یک الگوریتم منحصر به فرد و مسطح معرفی می‌شود. در نوع ساده‌ای از این الگوریتم نخست به میزان خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شود. در ادامه، داده‌ها بر اساس میزان شباهت به یکی از این خوشه‌ها منسوب می‌گردد. به این نحو خوشه‌های جدیدی به‌دست می‌آید. با تکرار همین مراحل می‌توان در هر تکرار با گرفتن میانگین از داده‌ها مراکز جدیدی برای آن‌ها تعیین نمود و مجددا داده‌ها را به خوشه‌های جدید منسوب کرد. این روال تا موقعی ادامه می‌یابد که دیگر تغییری در داده‌ها ایجاد نشود. مراحل اجرایی الگوریتم کا-میانگین در این پژوهش به شرح ذیل می‌باشد.

۱- نخست دوازده نقطه به عنوان نقاط مراکز خوشه‌ها تعیین می‌شوند.

۲- هر نمونه داده به خوشه‌ای که مرکز آن خوشه کوتاه‌ترین فاصله تا آن داده را دارد، منسوب می‌گردد.

۳- پس از نسبت دادن تمام داده‌ها به یکی از خوشه‌ها، برای هر خوشه یک نقطه جدید به عنوان مرکز در نظر گرفته می‌شود. (میانگین نقاط متعلق به هر خوشه)

۴- فعالیت ۲ و ۳ تکرار می‌شوند تا وقتی که دیگر هیچ تغییری در مرا کز خوشه‌ها ایجاد نشود (حریفی و همکاران، ۲۰۲۲).

روش‌های دسته‌بندی داده‌ها

برای دسته‌بندی از چهار روش ماشین بردار پشتیبان، درخت تصمیم، رگرسیون لجستیک و گوسین نایوبیز استفاده شد.

ماشین بردار پشتیبان

ماشین بردار پشتیبان از روش‌های یادگیری ماشین است که بر مبنای یادگیری آماری در دهه ۹۰ میلادی توسط وپنیک و همکارانش ارائه گردیده است. ماشین بردار پشتیبان روشی مبتنی بر برنامه‌ریزی غیرخطی و یک طبقه‌بندی دو وجهی است. این روش سعی دارد تا در مورد دو طبقه، ابر صفحه‌ای ایجاد نماید که فاصله هر طبقه تا ابر صفحه حداکثر باشد. داده‌های نقطه‌ای که به ابرصفحه نزدیک‌ترند، برای اندازه‌گیری فاصله به کار می‌روند. بنابراین داده‌های نقطه‌ای بردارهای پشتیبان محسوب می‌شوند. این روش دارای خواصی همچون (۱) طراحی طبقه‌بندی کننده‌ای با حداکثر تعمیم (۲) رسیدن به نقطه بهینه کلی تابع و (۳) تعیین خودکار ساختار و توپولوژی بهینه برای طبقه‌بندی کننده می‌باشد (منصورفر، غیور و لطفی، ۱۳۹۴).

درخت تصمیم

ساختار درخت تصمیم‌گیری یک ساختار درختی شبیه فلوچارت است که در این ساختار هر گره داخلی آزمونی را بر روی یک ویژگی مشخص می‌کند و هر شاخه خارج شده از این گره، دستاورد این آزمون را نشان می‌دهد و گره‌های برگ، دسته‌ها یا توزیع دسته‌ها را ارائه می‌نمایند. بالاترین گره در درخت گره ریشه است. درخت تصمیم‌گیری یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد. درخت تصمیم‌گیری برخلاف شبکه‌های عصبی

به تولید قانون می‌پردازد. در ساختار درخت تصمیم‌گیری پیش‌بینی به‌دست آمده از درخت در قالب یک سری قوانین توضیح داده می‌شود در حالی که در شبکه‌های عصبی تنها نتیجه پیش‌بینی بیان می‌شود و چگونگی به‌دست‌آمدن آن نتایج در خود شبکه پنهان باقی می‌ماند. همچنین در درخت تصمیم‌گیری بر خلاف شبکه‌های عصبی ضرورتی وجود ندارد که داده‌ها لزوما بصورت عددی باشند. در برخی موارد تنها درستی دسته‌بندی و صحت پیش‌بینی حاصل از آن مهم است و لزوماً ارائه توضیحی برای پیش‌بینی انجام شده، نیاز نیست. در مورد خصوصیات درخت تصمیم‌گیری به موارد زیر می‌توان اشاره نمود:

روش درخت تصمیم‌گیری در تقسیم‌بندی داده‌ها به گروه‌های مختلف به‌گونه‌ای است که هیچ داده‌ای حذف نمی‌شود (تعداد داده‌ها در گروه مادر با مجموع داده‌ها در شاخه‌های درخت ایجاد شده برابر هستند).

استفاده از تکنیک درخت تصمیم‌گیری آسان می‌باشد.

درک مدل ایجاد شده توسط درخت تصمیم‌گیری آسان می‌باشد. به عبارت دیگر با وجود اینکه فهمیدن روش کار الگوریتم‌های پدیدآورنده درخت ممکن است چندان ساده نباشند ولی فهمیدن نتایج به‌دست آمده از آنها آسان می‌باشد.

دسته‌بندی‌هایی که توسط درخت تصمیم‌گیری ایجاد می‌شوند، از روی شباهت داده‌های ذخیره‌شده در پارامترهای پیش‌بینی‌کننده، قابل انجام می‌باشد (غضنفری، علیزاده و تیمورپور، ۱۳۸۴).

رگرسیون لجستیک

از بعضی الگوریتم‌های رگرسیون می‌توان برای دسته‌بندی استفاده کرد. رگرسیون لجستیک، الگورتیمی است که احتمال متعلق بودن یک نمونه به یک دسته و طبقه‌ی خاص را تخمین می‌زند. اگر احتمال تخمین زده شده، بیشتر از پنجاه درصد باشد، مدل نمونه جزء دسته مثبت و در غیر این صورت جزء دسته منفی قرار می‌گیرد.

(۱) معادله دسته‌بندی رگرسیون لجستیک

$$\hat{y} = \begin{cases} 0 & if \quad \hat{P} < 0 \\ 1 & if \quad \hat{P} \geq 0 \end{cases}$$

این الگوریتم مانند رگرسیون خطی، مجموع وزن‌دار ویژگی‌های ورودی را به همراه یک مقدار خطا، محاسبه می‌کند. الگوریتم به جای آنکه مثل رگرسیون خطی، نتیجه نهایی را نشان دهد، شکل لجستیک نتیجه را نشان می‌دهد (تیمورپور و یادگاری و فاتحی پیکانی، ۱۳۹۹). نشان‌دهندهٔ خروجی پیش‌بینی شده است که در اینجا مقدار آن ۰ یا ۱ خواهد بود. نشان‌دهندهٔ احتمال تعلق داشتن نمونه به دسته مثبت است.

گوسین نایوبیز یا بیز ساده

این روش دسته‌بندی برای داده‌های با حجم بالا دارای دقت و سرعت بالایی است. در رده‌بندی بیز ساده فرض بر این است که احتمال رخداد یک صفت روی احتمال سایر صفت‌ها، بی‌تأثیر است. در تئوری بیز استخراج احتمال پسین با استفاده از احتمال پیشین امکان‌پذیر است (گودرزی و جنت‌بابایی، ۱۳۹۵).

روش وزن‌دهی فار برای ترکیب علم تصمیم

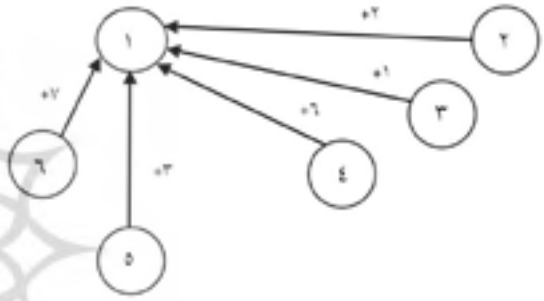
داده‌کاوی و کشف خوشه(ها)ی پرت، دامنه‌ی جست‌وجو برای احصاء لیستی از بازرسان متخلف را محدود می‌سازد. در این قسمت، علم تصمیم و مشخصاً تحلیل سلسله مراتبی برای دستیابی به نتیجهٔ غایی به کمک می‌آید. به منظور وزن‌دهی معیارها برای استفاده در تحلیل سلسله مراتبی، از روش فار استفاده گردید. این روش نخستین بار در سال ۲۰۱۱ توسط گینویسیوس (گینویسیوس، ۲۰۱۱) مطرح شد. یکی از امتیازات این روش، کاربرد آن در مسائلی است که معیارهای زیادی دارند و کار قیاس و رتبه‌بندی آن‌ها مشکل است. در این روش فرض بر آن است که معیارها بر

روی هم اثرگذارند و گرافی از رابطه مابین آن‌ها وجود دارد. اگر مسأله m معیار داشته باشد، آنگاه تعداد روابط بین معیارها (R) به طریق زیر خواهد بود.

(۲) فرمول محاسبه تعداد روابط بین معیارها

$$R = \frac{m(m-1)}{2}$$

همچنین وضعیت اهمیت روابط نسبت به هم مطابق نمودار (۲) زیر نشان داده می‌شود. شایان ذکر است که ابتدا باید مهم‌ترین معیار انتخاب گردد و سپس سایر معیارها نسبت به آن عددگذاری شود. طیف نمره‌دهی از ۱ تا ۱۰ است که ۱ بیانگر عدم برتری تا ۱۰ که بیانگر برتری مطلق است.



نمودار ۲- رابطهٔ بین اولین (مهم‌ترین) معیار و سایر معیارها

فلش از ۶ به سمت ۱ با ۷+ امتیاز به این معناست که معیار اول از معیار ششم، ۷ برابر مهم‌تر است. سپس برای تمامی معیارها، محاسبات زیر گام به گام انجام می‌شود.

فرمول محاسبه aij (تفاوت بین دو معیار i و j)

$$a_{ij}=a_{1j}-a_{1i},\ \ i,j=2,3,...,m;\ \ a_{ji}=-a_{ij}\ (i\neq j;\ i\neq 1, j\neq 1)$$

این فرمول تفاوت بین اثر معیار j و اثر معیار i نسبت به معیار مرجع یا معیار اصلی را نشان می‌دهد.

(۳) فرمول محاسبه Pi یا تاثیر اولیه معیار i

$$P_i = P_1 - m\,a_{1i}$$

(۴) فرمول محاسبه P (مجموع تاثیر معیارها) و PS (مجموع تاثیر سیستم):

$$P = S(m-1)$$

$$PS = m \cdot P = mS\,(m-1)$$

که در آن S، حداکثر امتیازی است که در ارزیابی داده شده است (در طیف ۱ تا ۱۰).

(۵) فرمول محاسبه P_i^f یا تاثیر نهایی معیار i

$$P_i + P = P_i^f$$

(۶) فرمول محاسبه وزن (wi یعنی وزن معیار i):

$$w_i = \frac{P_i^f}{P_S} = \frac{P_1 - m a_{1i} + S(m-1)}{mS(m-1)}$$

روش تحلیل سلسله مراتبی

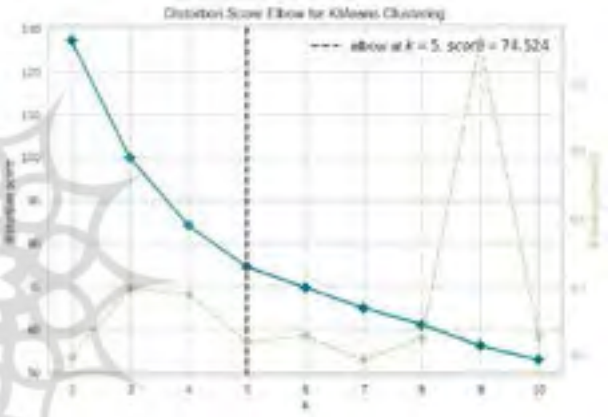
این روش توسط فردی به نام ساعتی، در دهه ۱۹۷۰ پیشنهاد شد و به عنوان یک ابزار مؤثر در ساختاردهی و مدل‌سازی مسائل چند معیاره مورد استفاده قرار می‌گیرد چراکه این فرایند تصمیم‌گیرندگان را یاری می‌کند تا اولویت‌ها را براساس اهداف، دانش و تجربه خود تنظیم نمایند (سمیعی اشکذری و همکاران، ۲۰۱۸). در این روش برای هر گزینه وزن معیار ضرب در امتیاز معیار برای آن گزینه شده و بدین ترتیب همه گزینه‌ها امتیازدهی می‌شوند. نهایتاً گزینه با امتیاز بهتر انتخاب می‌شود. در ادامه یافته‌های پژوهش و نتایج به‌کارگیری تکنیک‌ها، ارائه می‌گردد.

۴. یافته‌ها

نتایج خوشه‌بندی داده‌های بازرسان

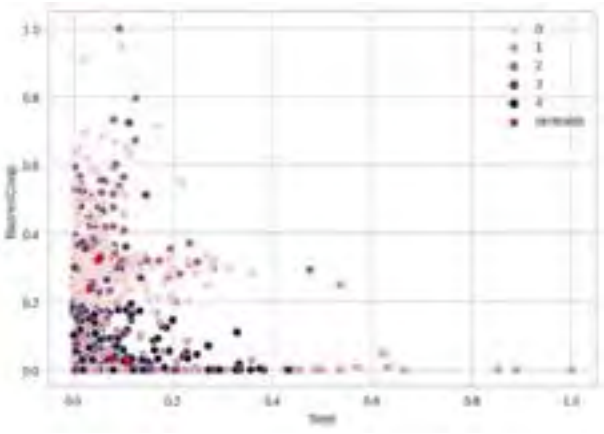
تعداد ۹۱۵۳ بازرس در امر بازرسی از بازار فعالیت داشتند که ۱۵۱۸ نفر که حداقل ۱۰ شکایت مردمی را رسیدگی کرده بودند، به عنوان نمونه آماری انتخاب شدند. یازده ویژگی (۱) میانگین فاصله زمانی تاریخ عملیات بازرسی پیرو شکایت از تاریخ محول کردن شکایت به بازرس، (۲) میانگین تعداد بازرسان همکارگزیده در بازرسی‌هایش، (۳) سرانه انجام عملیات بازرسی حول شکایت در یک روز، (۴) درصد عدم کشف تخلف در شکایات، (۵) تعداد استفاده از اولین دلیل عدم احراز در بازرسی‌ها، (۶) تعداد استفاده از دومین دلیل عدم احراز در بازرسی‌ها، (۷) تعداد استفاده از سومین دلیل عدم احراز در بازرسی‌ها، (۸) تعداد استفاده از چهارمین

دلیل عدم احراز در بازرسی‌ها، (۹) تعداد استفاده از پنجمین دلیل عدم احراز در بازرسی‌ها، (۱۰) تعداد استفاده از ششمین دلیل عدم احراز در بازرسی‌ها و (۱۱) میزان سابقه عضویت در سامانه برای خوشه‌بندی استفاده شدند. علت گزینش برخی ویژگی‌ها از بین تمام آن‌ها برای خوشه‌بندی آن است که ویژگی‌های اصلی در خوشه‌بندی تحت‌تأثیر برخی ویژگی‌های غیراصلی قرار می‌گرفتند و نتیجه تحلیل را تحت الشعاع قرار می‌دادند. بنابراین ابتدا یکبار تمام ویژگی‌ها برای خوشه‌بندی به کاررفته و پس از مشاهده نتایج، ۱۱ ویژگی منتخب به کار برده شد. نمودار (۳) نتایج معیار سیلوئت را نمایش می‌دهد.



نمودار ۳- نمودار سیلوئت برای انتخاب تعداد خوشهٔ داده‌های بازرسان

در ادامه نتیجه تقسیم داده‌های عملکردی بازرسان به پنج خوشه، تشریح می‌گردد.



نمودار ۴- خوشه‌بندی داده‌های عملکردی بازرسان

خوشه ۱ به عنوان خوشه پرت کشف شد. میانگین همکارگزینی برای بازرسی‌ها کمتر از سایر خوشه‌ها، میانگین مدت زمانی رسیدگی به شکایت بازرسان این خوشه، بیش از سایر خوشه‌ها، متوسط سرانه بازرسی منجر به عدم احراز تخلف در یک روز و نیز درصد ثبت نتیجه عدم احراز برای شکایات بیش از سایر خوشه‌هاست. همچنین متوسط سابقه عضویت بازرسان در سامانه از کمترین سابقه‌هاست.

جدول ۲- جزئیات اطلاعات ۵ خوشه بازرسان

شماره خوشه	فراوانی بازرس در خوشه	میانگین زمان رسیدگی به شکایت	میانگین همکارگزینی	متوسط سرانه	متوسط درصد عدم احراز کردن شکایت بازرسان در خوشه	متوسط سابقه عضویت بازرسان خوشه
۰	۴۶۹	۱۴/۵	۲	۳	۰/۸	۳/۲
۱	۲۴۹	۳۳	۱	۷/۵	۰/۹	۱/۲
۲	۱۴۳	۱۶/۱	۲	۴/۱	۰/۸	۱/۱
۳	۲۶۴	۱۰/۴	۱/۷	۲/۳	۰/۶	۳/۲
۴	۳۹۳	۲۲/۵	۱	۵/۱	۰/۹	۳/۲
کل	۱۵۱۸	۱۹	۱/۵	۴/۳	۰/۸	۲/۷

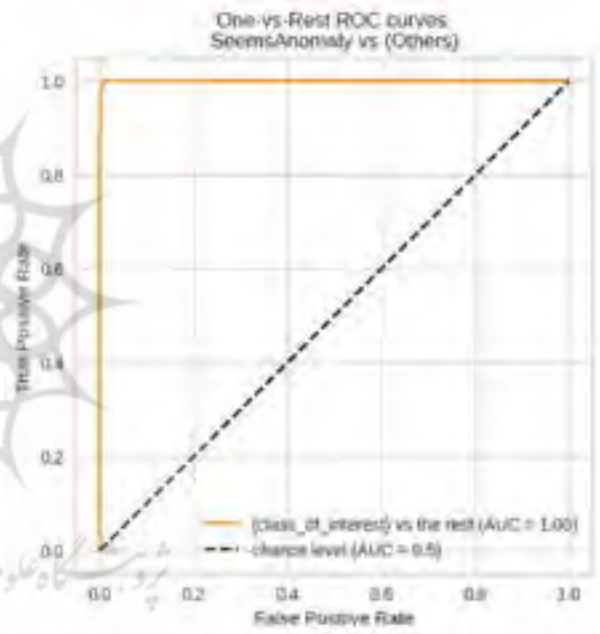
نتایج دسته‌بندی بازرسان

برای دسته‌بندی از چهار روش استفاده شد و بهترین نتایج برای درخت تصمیم بود.

جدول ۳- نتایج امتیاز f-score برای دسته‌بندی داده‌های بازرسان

روش دسته‌بندی	درخت تصمیم	رگرسیون لجستیک	بیز ساده	ماشین بردار پشتیبان
نتایج امتیاز f-score	۰/۹۸	۰/۹۶	۰/۹۶	۰/۹۴

نمودار (۵) منحنی مشخصه عملکرد بر اساس دسته ۱ یا همان خوشه‌ی پرت را نشان می‌دهد.



نمودار ۵- منحنی مشخصه عملکرد تحلیل پرونده‌های بازرسی

تفسیر و ارزیابی الگوهای ناهنجاری

در این قسمت ارزیابی‌های درونی و بیرونی مدل به خوانندگان عرضه می‌گردد. منظور از ارزیابی درونی، کنکاش در داخل داده‌ها و دو الگوی مستخرج جهت کشف ریشه‌هایی از اشتراکات و ارتباطات است. منظور از ارزیابی خارجی یا بیرونی، جست‌وجوی شواهدی اسنادی از وجود تخلف یا تقلب در بازرسی و انطباق آن با نتایج حاصله است.

ارزیابی درونی الگو

همانطور که پیشتر اشاره شد، پژوهش در دو رویکرد داده‌های بازرسی و داده‌های عملکردی دنبال‌شده و دو خوشه پرت از بازرسان و پرونده‌ها حاصل شد. در ارزیابی درونی دو خوشه پرت با یکدیگر تطابق داده شده تا اگر پرونده‌ای وجود دارد که بازرس آن در خوشه بازرسان مظنون است، پیدا شود. به بیانی دیگر یک پرونده ممکن است چند بازرس داشته باشد و یک بازرس هم ممکن است چندین پرونده ثبت کرده باشد. با تطابق دو خوشه، پرونده‌ها و بازرسان نظیر به نظیر با یکدیگر تطابق داده شده تا اشتراک آن‌ها حاصل و مورد تحلیل قرار گیرد. نتیجهٔ بررسی دو خوشه حاکی از آن بود که بازرس ۷۰ درصد پرونده‌ها و شکایاتِ خوشهٔ پرت در خوشهٔ پرت بازرسان یافت شد و نیز ۳۰ درصد بازرسان مظنون، در لیست ۱۲۸۴۳ تایی شکایات، پرونده داشتند.

ارزیابی بیرونی الگو

اولین و مهم‌ترین نکته‌ای که لازم است مورد نظر قرار گیرد آن است که داده‌کاوی و کشف خوشهٔ ناهنجار به معنای آن نیست که تمام افراد در خوشهٔ پرت، قطعاً متخلف بودند و یا تمام پرونده‌های آن قطعاً صوری ثبت شده‌اند بلکه به این معناست که اگر قرار بر جست‌وجوی متخلفان باشد، داده‌کاوی کمک می‌کند تا دامنهٔ جست‌وجو محدودتر و دقیق‌تر گردد. در ارزیابی بیرونی الگو، هدف جست‌وجوی شواهد و اسنادی است که به ما در صحت‌سنجی خوشهٔ مظنونان کمک کند. منظور از اسناد و شواهد، مصاحبه یا نامه و مدارک دیگر است. در بررسی‌های صورت گرفته از نامه‌های با مضمون عملکرد بازرسان و ارائه گزارش از آن‌ها یک نامه یافت شد که تاریخ آن به حدود یکسال پیش از آغاز داده‌کاوی پژوهش مربوط می‌شد و طی آن اسامی ۲۱ نفر از بازرسانی که عملکرد نامناسبی در بازرسی پیرو شکایات داشتند، اشاره شده بود. بررسی اسامی ۲۱ نفر و نیز خوشهٔ مظنون تحقیق، حاکی از وجود ۷۶ درصد بازرسان

اشاره‌شدهٔ نامه، در خوشهٔ مظنونان تحقیق حاضر بود. البته شایان ذکر است که مبنای کشف بازرسان در نامهٔ مذکور، خبرگی و اندک تحلیل‌های آماری بوده است که با پژوهش مبسوط حاضر با ویژگی‌ها و الگوریتم‌های دقیق به کاررفته اصلا قابل قیاس نیست؛ اما به هر روی این انطباق نشان از صحت مسیر پژوهش می‌دهد. یک مورد دیگر از شواهد به این شرح بود که طی تماس تلفنی یکی از استان‌ها، سخن از عملکرد بازرسی به میان آمد که عملکردش از جانب خود استان محل شک بود. پس از اخذ کدملی آن بازرس و جست‌وجوی آن داخل لیست بازرسان خوشهٔ پرت، نام وی یافت شد.

ترکیب روش‌های تصمیم‌گیری و داده‌کاوی جهت شناسایی بازرسان متخلف

همانطور که در سطور پیشین اشاره شد، خوشه‌ی پرت، دامنهٔ محدودی از داده‌ها است که امکان دستیابی به بازرسان متخلف و گروه هدف را تدقیق و تسریع می‌سازد. در ادامه این مهم که چطور بتوان لیستی از بازرسان متخلف را بر پایهٔ داده‌کاوی و حسب نظر کارشناسی و خبرگانی به دست آورد، شرح داده خواهد شد. طبیعتاً نمی‌توان تمام افراد خوشهٔ مظنون را متخلف دانست از این رو به منظور شناسایی بخشی از بازرسان خوشهٔ پرت به عنوان بازرسان شدیداً مظنون به تخلف، سه اقدام صورت می‌گیرد. ابتدا با استفاده از نظر خبرگان معیارهای عملکردی بازرسان از نظر اهمیت در کشف تخلف با روش رابطهٔ عوامل یا فار مورد وزن‌دهی قرار گرفته و سپس ۲۴۹ بازرس خوشه پرت مجدد خوشه‌بندی شده و خوشه‌ها با روش تحلیل سلسله مراتبی مقایسه می‌شوند. هر خوشه‌ای که بیشترین امتیاز را در تحلیل سلسله مراتبی کسب کند، با احتمال بسیار بالایی خوشه متخلفین است. معیارهای منتخب جهت وزن‌دهی خبرگان و نتیجتاً ارزیابی بازرسان به منظور شناسایی گونهٔ متخلف آنان عبارتند از:

۱. میانگین فاصله زمانی تاریخ عملیات بازرسی پیرو شکایت از تاریخ محول کردن شکایت به بازرس (بر حسب روز)
۲. میانگین تعداد بازرسان همکارگزیده در بازرسی‌هایش
۳. سرانه انجام عملیات بازرسی حول شکایت در یک روز
۴. درصد عدم کشف تخلف در شکایات (درصد ثبت نتیجه عدم احراز تخلف برای شکایت)
۵. فراوانی اولین دلیل عدم احراز تخلف ثبت کرده
۶. فراوانی دومین دلیل عدم احراز تخلف ثبت کرده
۷. فراوانی سومین دلیل عدم احراز تخلف ثبت کرده
۸. فراوانی چهارمین دلیل عدم احراز تخلف ثبت کرده
۹. فراوانی پنجمین دلیل عدم احراز تخلف ثبت کرده
۱۰. فراوانی ششمین دلیل عدم احراز تخلف ثبت کرده
۱۱. میزان سال سابقه عضویت در سامانه

کثرت معیارها، لزوم جامع‌نگری بین آن‌ها و لحاظ نمودن اثر مستقیم و نیز غیرمستقیم آن‌ها بر هم، سبب انتخاب روش فار برای وزن‌دهی و قیاس منطقی معیارها شد چراکه سایر روش‌ها نظیر بهترین و بدترین قادر به پوشش موارد فوق به خوبی روش فار نیستند. در ادامه محاسبات و اوزان معیارها مطابق با مسأله تحقیق ارائه می‌گردد.

جدول ۴- جدول معیارها و محاسبات وزن‌دهی

معیار	معیار											P_i	P_i^f	وزن
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱			
۱		+۶	+۴	+۴	+۸	+۷	+۷	+۵	+۵	+۷	+۹	+۶۲	۱۵۲	۰٫۱۵
۲	-۶		-۱	-۱	+۱	-۲	-۲	-۶	-۶	-۲	+۱	-۲۴	۶۶	۰٫۰۷
۳	-۴	+۱		+۲	+۵	+۳	+۳	-۴	-۴	-۲	+۲	+۲	۹۲	۰٫۱
۴	-۴	+۱	-۲		+۶	+۱	+۱	-۱	-۱	-۱	+۲	+۲	۹۲	۰٫۱
۵	-۸	-۱	-۵	-۶		-۳	-۴	-۶	-۶	-۵	+۲	-۴۲	۴۸	۰٫۰۵
۶	-۷	+۲	-۳	-۱	+۳		-۱	-۲	-۲	-۲	+۲	-۱۱	۷۹	۰٫۰۸
۷	-۷	+۲	-۳	-۱	+۴	+۱		-۲	-۲	-۱	+۲	-۷	۸۳	۰٫۰۸
۸	-۵	+۶	+۴	+۱	+۶	+۲	+۲		+۴	+۱	+۲	+۲۳	۱۱۳	۰٫۱۱

معیار	معیار											P_i	P_i^f	وزن
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱			
۹	-۵	+۶	+۴	+۱	+۶	+۲	+۲	-۴		+۴	+۲	+۱۸	۱۰۸	۰٫۱۱
۱۰	-۷	+۲	+۲	+۱	+۵	+۲	+۱	-۱	-۴		+۲	+۳	۹۳	۰٫۰۹
۱۱	-۹	-۱	-۲	-۲	-۲	-۲	-۲	-۲	-۲	-۲		-۲۶	۶۴	۰٫۰۶
جمع	-۶۲	+۲۴	-۲	-۲	+۴۲	+۱۱	+۷	-۲۳	-۱۸	-۳	+۲۶	۰	۹۹۰	۱

مقدار S که بیانگر بالاترین امتیاز در طیف ۱۰ تایی است، عدد ۹ و با احتساب عدد ۱۱ برای m به عنوان تعداد معیارها، مقدار P معادل ۹۰ و PS معادل ۹۹۰ خواهد بود.

خوشه‌بندی بازرسان خوشه پرت

بازرسان خوشه پرت داده‌های عملکردی بازرسان (خوشه ۱) بر اساس ۱۱ معیار به ۵ خوشه تقسیم شدند.

جدول ۵- نتیجه خوشه‌بندی ۲۴۹ بازرس

خوشه	تعداد	معیار۱	معیار۲	معیار۳	معیار۴	معیار۵	معیار۶	معیار۷	معیار۸	معیار۹	معیار۱۰	معیار۱۱
۰	۶۸	۲۷/۷۹	۱/۰۴	۴/۸۹	۰/۹۶	۵۲/۱۰	۲/۵۰	۰/۷۵	۸/۶۵	۱۷/۳۸	۰/۱۹	۰/۸۸
۱	۱۶	۱۷/۱۰	۱/۰۳	۲/۷۴	۰/۷۱	۵۳/۰۶	۲/۵۰	۰/۷۵	۴/۱۳	۸/۰۶	۰/۰۶	۱/۲۱
۲	۹	۳۳/۳۸	۱/۰۳	۵۹/۶۶	۰/۹۸	۸۲۳/۳۳	۲۵۵/۴۴	۷۷/۱۱	۱۳۰۲	۲۳۰۵/۸	۱۴/۲۲	۱/۴۱
۳	۳۵	۱۷/۵۷	۱/۴۲	۴/۵۹	۰/۹۵	۱۰۹/۲۶	۱۱/۷۴	۱۲/۲۰	۱۱/۶۹	۲۷/۰۹	۱/۰۳	۱/۳۱
۴	۱۲۱	۴۲/۴۵	۱/۰۱	۶/۶۷	۰/۹۷	۵۲/۲۷	۴۵/۱۱	۸/۹۸	۳۲/۲۶	۳۱/۰۲	۲/۰۲	۱/۴۸
کل	۲۴۹	۳۲/۹۹	۱/۰۸	۷/۵۶	۰/۹۵	۸۸/۱۶	۳۳/۶۵	۹/۱۲	۶۷/۰۱	۱۰۷/۴۹	۱/۷۰	۱/۲۷

در ادامه هر خوشه بر اساس معیارها و اوزان با یکدیگر مقایسه خواهند شد.

نتیجه تحلیل سلسله مراتبی خوشه‌های بازرسان خوشه پرت

طی فرایند تحلیل سلسله مراتبی ۵ خوشه با ۱۱ معیار، نتیجتاً خوشه دوم با ۴۷۷ امتیاز بالاتر از سایر خوشه‌ها که دربردارندهٔ ۹ بازرس است، به عنوان خوشه متخلفین شناسایی گردید.

کاوش در وضعیت عملکرد ۹ بازرس متخلف

برای روشن‌تر شدن موضوع و تحلیل، کاوش در عملکرد ۹ بازرس از میان ۱۵۱۸ بازرس می‌تواند مفید باشد. در ادامه وضعیت عملکرد بازرسان مذکور در قالب نمودارها به رشته‌ی تحریر درآمده است. براساس دستورعمل‌های بازرسی، گروه‌های بازرسی باید متشکل از حداقل ۲ نفر باشد.

شش بازرس از ۹ نفر، تمام بازرسی‌های خود را تکی انجام داده‌اند که خلاف دستورعمل بوده و همچنین زمینه‌ساز فساد خواهد بود. همچنین کمترین آن‌ها ۹۱٫۸ درصد و تعداد زیادی نیز ۱۰۰ درصد شکایات خود را با نتیجه عدم احراز تخلف ثبت نتیجه کردند. یعنی اکثریت قابل توجه ۹ نفر، در هیچ یک از شکایاتی که به آنان محول شده بوده، برای حتی یک شکایت نیز پرونده‌ی تخلفی تشکیل ندادند. به بیانی آماری‌تر آن‌ها جمعاً ۴۳ هزار شکایت از ۲۴۴ هزار شکایت (معادل یک ششم کل شکایات مردمی) را با نتیجه عدم احراز تخلف، مختومه کردند و ۲۰۷۵۲ شکایت معادل تقریبا نیمی از این ۴۳ هزار شکایت، با دلیل «تعطیل یا تخلیه بودن واحد اقتصادی مورد شکایت» مختومه شدند که بسیار واهی به نظر می‌رسد.

۵. نتیجه‌گیری و توصیه‌های سیاستی

آنچه در بندهای پیشین گذشت، خوشه‌بندی بازرسان بر اساس داده‌های عملکردی و تمرکز بر خوشه‌ی پرت به منظور شناسایی بازرسان متخلف به کمک راه‌کارهای علم تصمیم بود. در این قسمت و پس از رسیدن به نتایج پژوهش، به فراخور تحلیل انجام‌شده و دانش کشف‌شده از داده‌کاوی و مشتقّات آن، نتیجه‌گیری و توصیه‌ها ارائه می‌گردد.

چالش‌های مدیریتی

چالش‌های مدیریتی، مشکلاتی هستند که ناشی از عدم مدیریت بازرسان برای رسیدگی به شکایات می‌باشد. به ازای هر چالش اداری سعی می‌شود تا راه‌کار سیستمی مقتضی نیز معرفی گردد؛ اما، برخی نظیر عدم استفاده از جی‌پی‌اس در بازرسی‌ها که بشود حرکت بازرسان را ردیابی کرد و صحت و سقم ادعای بازرسی آنان را مورد کاوش قرار داد، از حیطهٔ فنی خارج است.

ثبت اپراتوری نتیجه شکایات

بر اساس مشاهدات و بررسی‌های صورت‌گرفته حین تحقیق، در سازمان‌های متولی رسیدگی به شکایات، بازرسان -در میدان- بازرسی کرده و نتایج را کتباً به اپراتورها می‌دهند تا با حساب کاربری آنان نتایج بازرسی را ثبت کنند. این رویه، رویه‌ای غیرسیستمی است که علت عمدهٔ آن، فقدان رایانه به ازای هر بازرس و نیز ضعف سواد در کار با آن است. نبود فضا برای استقرار رایانه و نبود بودجه برای خرید آن، مشکلی مدیریتی بوده و در خصوص ضعف علمی، تنها راه، آموزش و یادگیری است.

راه‌کار سیستمی رفع چالش ثبت اپراتوری نتیجه شکایات

با فرض تخصیص رایانه به اندازه کافی، اگر حساب کاربری مشتمل بر کدملی و رمز عبور از جانب بازرس به اپراتور داده می‌شود تا فرم‌های وی را ثبت کند، می‌توان به صورت رندمی هنگام ورود به حساب کاربری و ثبت فرم، پیامک حاوی رمز یکبار مصرف جهت احراز هویت به شماره تماس بازرس صادر نموده و به واسطهٔ آن حساب‌های کاربری اپراتوری را شناسایی و مسدود نمود. البته روش دیگری نیز می‌توان مطرح نمود به این صورت که سمت خاصی برای اپراتورها تعریف گردد که طی آن بازرس از اپراتور در گزارش‌گیری‌ها و عملکردسنجی‌ها تمیز داده شود.

مشکلات و چالش‌های فنی - تحلیلی

چالش‌های تحلیلی، مشکلاتی هستند که از ضعف یا خلل در تحلیل و طراحی سامانه نشأت می‌گیرند. به طور مثال لازم است تا فرایندی سیستمی جهت اعتبارسنجی صحت و سقم ادعای بازرس مبنی بر مصالحه و رضایت شاکی از متشاکی طراحی گردد تا در صورت انتخاب گزینه «رضایت شاکی» توسط بازرس، به شکل سیستمی ارزیابی شود که آیا واقعاً شاکی رضایت داده یا ادعای بازرس دروغین بوده است. در ادامه به چالش‌ها اشاره شده و به ازای هر چالش، راه‌کار رفع آن نیز بیان می‌گردد.

تناقض برخی دلایل عدم احراز تخلف با انواع بازرسی

به طور مثال دلیل «عدم کفایت مستندات شاکی» با نوع بازرسی «نوبه‌ای» در تناقض است.

راه‌کاری سیستمی رفع چالش تناقض دلایل عدم احراز با انواع بازرسی

لازم است تا رابطه‌ای بین دلایل عدم احراز تخلف و انواع بازرسی ایجاد شود و در صورت انتخاب هر نوع بازرسی توسط کاربر، دلیل عدم احراز منطقی و مربوط به آن نوع بازرسی، قابل انتخاب باشد.

موازی‌کاری بازرس و ارزیاب شکایت در انتخاب گزینه «عدم کفایت مستندات شکایت»

ارزیاب شکایت، وظیفه بررسی کلی شکایت را دارد و در صورت صلاحدید می‌تواند شکایت را به بازرس برای بازرسی

ارجاع دهد. در صورتی که شکایتی مردود باشد یا نیاز به ویرایش و تکمیل اطلاعات توسط شهروند را داشته باشد، توسط ارزیاب به شهروند عودت داده شده و وی می‌تواند ویرایش کند. نظیر همین دلیل در ثبت نتیجه بازرسی توسط بازرس نیز وجود دارد. به زبانی دیگر، بازرس می‌تواند شکایتی را به دلیل عدم کفایت مستنداتش مختومه کند حال آنکه در مرحلهٔ ارزیابی شکایت، باید این مهم تشخیص داده می‌شد تا به شهروند جهت افزودن مستندات ارسال گردد.

راه‌کاری سیستمی رفع چالش موازی‌کاری بازرس و ارزیاب شکایت در انتخاب گزینه «عدم کفایت مستندات شکایت»

این گزینه باید در لیست دلایل عدم احراز تخلف فرم، غیرفعال شود تا بازرس در صورت چنین مشکلی، شکایت را به ارزیاب عودت داده و وی نیز به شهروند جهت افزودن مستندات عودت دهد.

عدم وجود رخدادنگار (لاگ) مفید از سوابق عملکرد کارکنان در سامانه

یکی از مشکلات و چالش‌های پژوهش حاضر عدم وجود لاگ یا سابقهٔ کارکردن کاربران با سامانه بود. به طور مثال اگر لاگ آخرین مرتبهٔ مشاهده شکایت قبل از ثبت فرم وجود داشت، میسر می‌بود تا با محاسبه فاصله زمانی آخرین مرتبه مشاهده شکایت از زمان انجام بازرسی به این مهم پی‌برده شود که چقدر بوده است. در صورتی‌که این مقدار بسیار کم باشد، می‌تواند این ظن را به وجود آورد که بازرس بلافاصله پس از مشاهده شکایت، فرم گزارش بازرسی آن را به شکل صوری ثبت می‌کند. امروزه در سامانه‌ها و پایگاه‌های اینترنتی استفادهٔ بسیار زیادی از لاگ‌ها می‌شود. به نظر می‌رسد سامانه بازرسی باید از این حیث تقویت شود.

راه‌کار سیستمی رفع مشکل عدم وجود لاگ مفید در سامانه

لازم است تا با الهام از نتایج تحقیق حاضر و نیز مصاحبه با خبرگان و افراد فنی و تحلیل‌گر، لیستی از لاگ‌های مفید که به شناسایی بازرسان متخلف کمک می‌کند، تهیه و به ستون داده‌های پایگاه دادهٔ سامانه اضافه شود.

چالش رسوبی‌شدن شکایت در کارتابل ارزیاب و بازرس

یکی از دلایل عدم رسیدگی درست و ثبت صوری نتایج بازرسی شکایات، رسوبی‌شدن آن‌ها است. به طور مثال شکایتی پس از یک سال از زمان ثبت آن توسط شهروند، به بازرس می‌رسد. در طی این مدت شاید قیمت‌ها به سبب نرخ تورم یا دلایل دیگر تغییر کرده و از این رو بازرس با شکایتی مواجه است که الان موضوعیت ندارد. گرچه بازرس می‌تواند شکایت را بر حسب زمان دریافت آن گزارش رسیدگی کند ولی از کارایی آن کاسته خواهد شد. همچنین ممکن است شکایتی پس از یک سال به تازگی ارزیابی شده و به بازرس جهت بازرسی ارجاع شود. این مورد نیز همانند مثال اول است.

راه‌کار سیستمی رفع چالش رسوبی‌شدن شکایات در کارتابل ارزیاب‌ها و بازرسان

لازم است در صورتی که تعداد شکایات در کارتابل ارزیاب جهت ارزیابی یا بازرس جهت بازرسی، از یک تعداد مشخصی بیشتر شد، سامانه اخطاری به کاربر ارسال کند و چنانچه توجه نشد، برای بار دوم به وی و رئیس مافوقش پیامک برود تا مسئولین امر از خطر رسوبی‌شدن شکایات مطلع شوند.

محدودیت‌های تحقیق

یکی از محدودیت‌های تحقیق که در سطور پیشین به رشتهٔ تحریر درآمد، نبود لاگ مناسب برای کشف دقیق‌تر ناهنجاری‌ها بود. از دیگر محدودیت‌ها می‌توان به نبود داده‌ی سابقه بازرسی بازرسان اشاره نمود. داده‌ی سابقه

بازرسان در امر بازرسی دست‌نیافتنی عنوان می‌شود و در این پژوهش صرفاً به داده‌ی سابقه عضویت بازرسان از زمان ثبت‌نام آن‌ها در سامانه بسنده شده است. در صورت وجود داده‌ی مذکور، نتایج می‌تواند دقیق‌تر باشد. تغییر چندباره‌ی سازمان و استان محل خدمت برخی بازرسان نیز از جمله محدودیت‌هایی بود که سبب خلل در تحلیل داده عملکرد یا نقص در آن می‌شد.

پیشنهادها برای مطالعات آتی

مهم‌ترین پیشنهاد، بررسی بروز مجدد ناهنجاری‌های مطروح در پژوهش، پس از به‌کار بستن راه‌کارهای سیستمی پژوهش است. لازم است تا پس از اصلاح و بازطراحی سامانه بر اساس راه‌کارها، مجدداً تحلیل صورت گیرد و نتایج مقایسه شود. همچنین این عمل پس از به کارگیری ویژگی‌های جدید برآمده از لاگ‌های اضافه‌شده نیز صورت گیرد.

رتبه‌بندی استان‌ها از نظر تعداد بازرسی‌های صوری و بازرسان متخلف دومین پیشنهاد نویسندگان می‌باشد. موضوع دیگر ترکیب نتایج تحلیل داده‌ی گراف همکاری بازرسان دخیل در رسیدگی به شکایت با نتایج این پژوهش به منظور شناسایی جمعی (گروهی) و نه فردیِ ناهنجاری است. در صورت توسعه‌ی داده‌های عملکردی نظیر داده‌ی مسیر حرکت بازرسان در بازرسی روزمره توسط مکان‌یاب، توسعه‌ی روشی نوین برای دسته‌بندی یا ترکیب روش‌های جدید با یکدیگر می‌تواند پیشنهادی برای مطالعات آتی یا توسعه‌ی پژوهش حاضر باشد.

دسترسی به داده‌ها

داده‌های استفاده شده از پایگاه اینترنتی داده‌های بانک جهانی و سازمان جهانی تجارت قابل دسترسی است. داده‌های تولید شده در متن مقاله ارائه شده است.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌کنند که هیچ‌گونه تضاد منافعی در رابطه با نویسندگی یا انتشار این مقاله ندارند.

منابع

پهلوانی، ج.، شیخ محمدی، م.، و تیمورپور، ب. (۱۴۰۳). کشف الگوی ناهنجاری در بازرسی‌های پیرو شکایات مردمی با راه‌کارهای داده‌کاوی. نوزدهمین کنفرانس بین‌المللی مهندسی صنایع.

تیمورپور، ب.، یادگاری، وحید و فاتحی پیکانی، م. (۱۳۹۹) داده‌کاوی با پایتون؛ به همراه تحلیل شبکه‌های اجتماعی. Olomrayaneh.

حریفی، ر.، نعیمی صدیق، ع.، و مظفری، م. (۲۰۲۲). بخش‌بندی مشتریان سازمان بنادر و دریانوردی با به کارگیری شبکه عصبی خودسازمانده و الگوریتم K-Means. بررسی‌های بازرگانی، ۲۰(۱۱۴)، ۱۳۵-۱۵۴.

https://doi.org/10.22034/bs.2022.254024

سمیعی اشکذری، م.، سلیمانی، غ. و شمس، ک. (۲۰۱۸). تبیین و تحلیل ارزیابی استراتژی‌های ورود به بازارهای خارجی (مطالعه موردی: خوشه کاشی و سرامیک استان یزد). بررسی‌های بازرگانی، ۱۶(شماره ۸۸-۸۹)، ۶۳-۷۸.

https://barresybazargani.itsr.ir/article_34479.html

غضنفری، م.، علیزاده، س. و تیمورپور، ب. (۱۳۸۴) داده‌کاوی و کشف دانش. دانشگاه علم و صنعت ایران.

گودرزی، آتوسا و جنت‌بابایی، سجاد (۱۳۹۵) «ارزیابی الگوریتم‌های درخت تصمیم، بیز ساده و رگرسیون لجستیک در کشف تقلبات بیمه اتومبیل»، پژوهش‌های بیمه ای، ۲(۱)، pp. ۶۱-۸۰.

Available at: https://www.noormags.ir/view/fa/articlepage/1428946.

مقیمی، ف. (۱۴۰۲). فعالیت ۳ میلیون واحد صنفی در کشور.

https://www.yjc.ir/00a6IX

منصورفر، غ.، غیور، ف. و لطفی، ب. (۱۳۹۴) «توانایی ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی»، پژوهش‌های تجربی حسابداری، ۵(۱۷)، pp. ۱۷۷-۱۹۵.

Available at: https://doi.org/https://doi.org/10.22051/jera.2015.646.

Akpinar, M., Adak, M. F., & Guvenc, G. (2021). SVM-based anomaly detection in remote working: Intelligent software SmartRadar. Applied Soft Computing, 109, 107457. https://doi.org/https://doi.org/10.1016/j.asoc.2021.107457

Alla, S., & Kalyan Adari, S. (2019). Beginning Anomaly Detection Using Python-Based Deep Learning. Apre. https://doi.org/https://doi.org/10.1007/978-1-4842-5177-5

Amezaga Hechavarria, A., & Shafiq, M. O. (2022). A modified attention mechanism powered by Bayesian Network for user activity analysis and prediction. Data & Knowledge Engineering, 140, 102034. https://doi.org/https://doi.org/10.1016/j.datak.2022.102034

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Comput. Surv., 41. https://doi.org/10.1145/1541880.1541882

Domashova, J., Kripak, E., & Pisarchik, E. (2022). Application of machine learning methods to identify suspicious actions of employees related to violation of the procedures of a credit institution. Procedia Computer Science, 213, 110-118. https://doi.org/https://doi.org/10.1016/j.procs.2022.11.045

Ghazanfari, M., Alizadeh, S., & Teimourpour, B. (2005). *Data mining and knowledge discovery*. Iran University of Science and Technology. [In Persian]

Ginevičius, R. (2011). A New Determining Method for the Criteria Weights in multicriteria Evaluation. International Journal of Information Technology & Decision Making (IJITDM), 10, 1067-1095. https://doi.org/10.1142/S0219622011004713

Goodarzi, A., & Janatbabaei, S. (2016). Evaluation of decision tree, naive Bayes, and logistic regression algorithms in detecting auto insurance fraud. Insurance Research, 2*(1), 61-80. https://www.noormags.ir/view/fa/articlepage/1428946 [In Persian]

Han, R., Kim, K., Choi, B., & Jeong, Y. (2023). A Study on Detection of Malicious Behavior Based on Host Process Data Using Machine Learning. In Applied Sciences (Vol. 13, Issue 7). https://doi.org/10.3390/app13074097

Harifi, R., Naimi-Sadigh, A., & Mozafari, M. (2022). Customer Segmentation of the Ports and Maritime Organization using Self Organization Map and K-Means algorithm. Commercial Reviews, 20*(114), 135-154. https://doi.org/10.22034/bs.2022.254024 [In Persian]

Ianni, M., & Masciari, E. (2023). SCOUT: Security by computing OUTliers on activity logs. Computers & Security, 132, 103355. https://doi.org/https://doi.org/10.1016/j.cose.2023.103355

Jain, M., Kaur, G., & Saxena, V. (2022). A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. Expert Systems with Applications, 193, 116510. https://doi.org/https://doi.org/10.1016/j.eswa.2022.116510

Khan, S., Parkinson, S., & Murphy, C. (2023). Context-based irregular activity detection in event logs for forensic investigations: An itemset mining approach. Expert Systems with Applications, 233, 120991. https://doi.org/https://doi.org/10.1016/j.eswa.2023.120991

Landauer, M., Onder, S., Skopik, F., & Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey. Machine Learning with Applications, 12, 100470. https://doi.org/https://doi.org/10.1016/j.mlwa.2023.100470

Lee, Y., Kim, J., & Kang, P. (2023). LAnoBERT: System log anomaly detection based on BERT masked language model. Applied Soft Computing, 146, 110689. https://doi.org/https://doi.org/10.1016/j.asoc.2023.110689

Mansourfar, Gh., Ghayour, F., & Lotfi, B. (2015). The Ability of Support Vector Machine (SVM) in Financial Distress Prediction. *Empirical Research in Accounting, 5*(17), 177-195. https://doi.org/10.22051/jera.2015.646 [In Persian]

Moghimi, F. (2023). Activity of 3 million trade units in the country. https://www.yjc.ir/00a6IX [In Persian]

Pahlavani, J., Sheikhmohammadi, M., & Teimourpour, B. (2024). Anomaly detection in inspections on public complaints: a data mining approach. 19th International Conference on Industrial Engineering. [In Persian]

Racherache, B., Shirani, P., Soeanu, A., & Debbabi, M. (2023). CPID: Insider threat detection using profiling and cyber-persona identification. Computers & Security, 132, 103350. https://doi.org/https://doi.org/10.1016/j.cose.2023.103350

Ramírez, J. M., Díez, F., Rojo, P., Mancuso, V., & Fernández-Anta, A. (2023). Explainable machine learning for performance anomaly detection and classification in mobile networks. Computer

Communications, 200, 113-131. https://doi.org/https://doi.org/10.1016/j.comcom.2023.01.003

Ranjan, R., & Kumar, S. S. (2022). User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user. High-Confidence Computing, 2(1), 100034. https://doi.org/https://doi.org/10.1016/j.hcc.2021.100034

Samiei Ashkezari, M., Soleimani, Gh., & Shams, K. (2018). Assessing the Effectiveness of Foreign Market Entry Strategies (Case Study: Ceramic and Tile Cluster of Yazd). *Commercial Reviews, 16*(88-89), 63-78. https://barresybazargani.itsr.ir/article_34479.html [In Persian]

Soh, C., Yu, S., Narayanan, A., Duraisamy, S., & Chen, L. (2019). Employee profiling via aspect-based sentiment and network for insider threats detection. Expert Systems with Applications, 135, 351-361. https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.043

Teimourpour, B., Yadegari, V., & Fatehi Pikani, M. (2020). Data mining with Python; Along with social network analysis. Olomrayaneh. [In Persian]

Zeufack, V., Kim, D., Seo, D., & Lee, A. (2021). An unsupervised anomaly detection framework for detecting anomalies in real time through network system's log files analysis. High-Confidence Computing, 1(2), 100030. https://doi.org/https://doi.org/10.1016/j.hcc.2021.100030