

ORIGINAL RESEARCH PAPER

Data Mining of Occupational Accidents in an Automotive Manufacturing Company for the Identification of Root Causes and the Implementation of Preventive Strategies

Mohammadreza Bahrehbar, Saeed Givehchi*

Faculty of Environment, Health, Safety and Environment Department, University of Tehran, Tehran, Iran

Received: 2026/02/20

Accepted: 2026/07/22

ABSTRACT

Introduction: Contemporary automotive manufacturing enterprises, like other industrial sectors, routinely collect large volumes of data pertaining to productivity, occupational safety, and health. The transformation of these raw data into actionable knowledge has become increasingly critical for organizational decision-making. This study applies data mining techniques to analyze occupational accidents resulting in lost workdays at an Iranian automotive manufacturing company from June 2014 to March 2025, aiming to identify the patterns and determinants of workplace incidents.

Material and Methods: Accident records from the specified period were systematically collated and preprocessed, then subjected to classification and association rule mining using established data mining algorithms. Classifier implementations and experiments were conducted within the Weka software environment. In addition to supervised classification, association rule mining algorithms (Apriori and Predictive Apriori) were employed to discover co-occurring attribute patterns and identify combinations of factors associated with elevated accident risks.

Results: Through comparative evaluation, the support vector machine (SVM) classifier demonstrated the best performance in predicting accident type, severity & causes, achieving an accuracy of 66.08% under cross-validation. However, the decline from a 100% training accuracy indicates the presence of overfitting in the models. Furthermore, Apriori and Predictive Apriori algorithms successfully extracted rules with high performance metrics, demonstrating confidence levels above 83% and predictive accuracies between 97.9% and 98%. Quantitative analysis revealed that the highest correlation with incidents was found in a combination of specific factors: low work experience (0 to 5 years), late shift hours (4 to 8 hours after the shift start), and employment as a body production operator. Secondary contributing variables included shift scheduling, proximate causes of the event, annual production rate, and employee remuneration.

Conclusion: The findings demonstrate that data mining techniques are effective tools for uncovering hidden patterns in occupational accidents within the automotive industry; nevertheless, the observed overfitting in classification models remains a serious limitation, highlighting the need for more balanced datasets. From a practical perspective, the identified association rules enable the design of specialized training programs, optimized rest schedules, and targeted safety interventions specifically for high-risk groups, such as low-experienced workers (0–5 years) operating in body production during the late hours of their shifts (4–8 hours).

Keywords: Occupational accidents, Data mining, Association rule analysis, Occupational health and safety

HOW TO CITE THIS ARTICLE

Bahrehbar M.R., Givehchi S. Data Mining of Occupational Accidents in an Automotive Manufacturing Company for the Identification of Root Causes and the Implementation of Preventive Strategies. *J Health Saf Work.* 2026; 16(2): 448-471.

* Corresponding Author Email: givehchi@ut.ac.ir

Copyright © 2026 The Authors.
Published by Tehran University of Medical Sciences

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

1. INTRODUCTION

Occupational accidents remain one of the most critical challenges confronting industrial organizations, particularly within large-scale manufacturing sectors such as the automotive industry, where complex production systems, intensive human-machine interaction, and stringent productivity demands coexist. Throughout history, the human aspiration for a safe and risk-free working life has been recognized as an integral component of social and organizational development. In contemporary industrial contexts, this aspiration has gained renewed importance due to accelerating globalization, intensified market competition, rapid technological advancement, and increasing pressure on organizations to optimize productivity while minimizing costs.

The automotive industry, as a cornerstone of industrial development, is especially vulnerable to occupational accidents that disrupt production continuity, reduce labor efficiency, and impose substantial direct and indirect costs. Occupational accidents not only result in physical injuries to workers but also lead to work stoppages, regulatory investigations, equipment damage, and prolonged recovery periods, all of which negatively affect overall organizational performance. Moreover, injured employees may experience pain, functional limitations, or temporary or permanent incapacity, further exacerbating productivity losses and increasing organizational burdens.

In such conditions, the systematic analysis of accumulated occupational accident data and the extraction of actionable knowledge from these datasets become essential prerequisites for improving workplace safety and protecting workers' health. Traditional statistical approaches, while valuable, often encounter limitations when applied to complex, high-dimensional, and heterogeneous datasets such as occupational accident records. These datasets typically involve nonlinear relationships and interactions among human, organizational, temporal, and environmental factors that are difficult to capture using conventional methods.

Data mining, as a scientific and analytical discipline, offers powerful tools for discovering hidden patterns, meaningful relationships, and predictive structures within large and complex datasets. By enabling the extraction of implicit and previously unknown knowledge, data mining techniques can significantly enhance

the understanding of occupational accident mechanisms and support evidence-based decision-making in safety management systems. Despite the growing recognition of data mining applications in occupational health and safety research, relatively few studies have focused on root cause analysis of industrial accidents using integrated data mining approaches, particularly within the automotive manufacturing context.

Furthermore, existing studies have often relied solely on classical accident variables, neglecting the potential influence of production-related indicators such as production volume and productivity rates. Addressing this research gap, the present study aims to analyze occupational accidents leading to injury in an automotive manufacturing organization by integrating classical accident data with Annual production and productivity information. The core innovation of this research lies in the combined application of classification techniques and association rule mining to uncover meaningful relationships between production rates and occupational accidents, thereby contributing to the development of predictive and preventive safety strategies.

2. MATERIAL AND METHODS

This study analyzed occupational accidents leading to injury over a twelve-year period from 2014 to 2025. The dataset comprised officially recorded occupational accident cases extracted with the necessary authorizations from the organization's Health, Safety, and Environment (HSE) management system. During the study period, a total of 2,169 occupational accidents were registered, of which 2,167 cases resulted in injury and constituted the primary dataset for analysis.

The collected data included detailed information on accident characteristics, worker attributes, and workplace conditions. In addition to conventional accident-related variables, Annual production and productivity data were incorporated to enable a more comprehensive analysis of the relationship between production pressure and accident occurrence. The integration of these datasets allowed for a multidimensional examination of occupational accidents from human, temporal, organizational, and production-related perspectives.

Data preprocessing was conducted to ensure data quality and analytical validity. This process included data cleaning, handling missing values,

Table 1: The algorithms applied to the dataset and the results obtained.

Algorithm					Test methods for algorithms			
Training set					Cross-validation			
	Correct classified	Kappa	MAE	RMSE	Correct classified	Kappa	MAE	RMSE
ZeroR	1254 (57.86%)	0	0.113	0.2374	1254 (57.86%)	0	0.113	0.2374
NB	1489 (68.71%)	0.4207	0.0796	0.2048	1394 (64.32%)	0.3368	0.0861	0.2189
J48	1867 (86.16%)	0.7653	0.0313	0.1251	1425 (65.76%)	0.3266	0.093	0.2217
kNN	2167 (100%)	1	0.0008	0.0015	1332 (61.46%)	0.163	0.0927	0.2226
MLP	2074 (95.70%)	0.9299	0.0099	0.0804	1236 (57.03%)	0.2791	0.08	0.2665
SVM	1553 (71.66%)	0.4522	0.1501	0.2658	1432 (66.08%)	0.3424	0.1512	0.2677
RF	2167 (100%)	1	0.034	0.0797	1388 (64.05%)	0.2866	0.0931	0.215

J48, J48 Decision Tree; kNN, k Nearest Neighbor; MAE, mean absolute error; MLP, Multilayer Perceptron; NB, Naive Bayes; RF, Random Forest; RMSE, root mean square error; SVM, Support Vector Machines.

and transforming categorical variables into suitable formats for data mining algorithms. Relevant attributes and their subcategories were defined based on the structure of the accident records and the analytical objectives of the study.

For data analysis, the open-source data mining software WEKA version 3.8.6 was employed. WEKA provides a robust environment for implementing and comparing multiple machine learning and data mining algorithms. The analytical framework of the study followed a structured methodology, as illustrated in the study's methodological flowchart, encompassing data preparation, model development, evaluation, and knowledge extraction.

Classification techniques were applied to identify patterns associated with occupational accidents and to evaluate the predictive capability of different models. In parallel, association rule mining was utilized to discover frequent co-occurring factors and to extract interpretable rules that characterize high-risk conditions and groups. Among the association rule mining methods, particular attention was given to algorithms capable of generating rules with high predictive accuracy (Table1).

The combination of classification and association rule mining methods enabled both predictive analysis and descriptive knowledge discovery. While classification models provided insights into the relative importance of different variables in distinguishing accident cases, association rules offered transparent and actionable knowledge regarding the simultaneous presence of multiple risk factors. This integrated analytical approach was designed to support practical decision-making in occupational safety management and to facilitate the development of targeted preventive

interventions.

3. RESULTS AND DISCUSSION

The analysis revealed notable temporal variations in the number of occupational accidents leading to injury over the study period. While fluctuations were observed across different years, the integration of accident data with production and productivity rates highlighted meaningful associations between production intensity and accident occurrence. In particular, years characterized by higher production rates exhibited increased accident frequencies when normalized to production output, underscoring the potential influence of production pressure on occupational safety performance.

The application of data mining techniques demonstrated that combining classification methods with association rule mining yields valuable and actionable insights for identifying high-risk groups and critical working conditions within manufacturing environments. The extracted association rules consistently highlighted specific combinations of human, temporal, and organizational factors that were frequently associated with injury-causing accidents.

One of the most prominent findings was the repeated identification of workers with 0–5 years of job experience as a high-risk group. This result suggests that limited experience, potentially combined with insufficient hazard recognition and task familiarity, plays a significant role in accident occurrence. Additionally, accidents were frequently associated with particular job positions, notably production operators in body shop areas, where intensive manual tasks and close interaction with machinery are common.

Temporal factors also emerged as significant

Table 2: Association rules generated for ALs using the Predictive Apriori algorithm

Rule		AssociaAccuracy		
Attributes of the rule	Number of accidents for rule attributes	Target attribute	Number of accidents for target attribute	Accuracy
1 E = 0-5, WH = 0-5, MP = 24,000-25,999, MW = 6,000-6,999, AC=Laceration injury	23	AL = body shop areas	23	0.9804
2 M = July, E = 0-5, MP = 20,000-21,999	22	AL = body shop areas	22	0.9794
3 M = March, WH = 0-5, WP=Body Shop Operator, AC=Misuse of tools and equipment	22	AL = body shop areas	22	0.9794
4 AO = Toe, S = 8-16, WH = 0-5, WP=Body Shop Operator, AC=Laceration injury	22	AL = body shop areas	22	0.9794
5 E = 0-5, S = 16-24, WH = 0-5, MP = 24,000-25,999, AC=Laceration injury	22	AL = body shop areas	22	0.9794
6 D = Monday, E = 0-5, S = 16-24, WH = 0-5, MW = 5,000-5,999, WP=Body Shop Operator	22	AL = body shop areas	22	0.9794
7 AO = toe, D = Wednesday, S = 16-24, WP=Body Shop Operator	21	AL = body shop areas	21	0.9782
8 M = March, A = 26-30, WP=Body Shop Operator, AC=Laceration injury	21	AL = body shop areas	21	0.9782
9 D = Tuesday, S = 16-24, LWD = 4-28, WP=Body Shop Operator, AC=Laceration injury	21	AL = body shop areas	21	0.9782
10 AO=Waist, E = 0-5, WH = 0-5, WP=Body Shop Operator	20	AL = body shop areas	20	0.9768
11 D = Monday, WH = 0-5, WP=Body Shop Operator, AC=Struck by object	20	AL = body shop areas	20	0.9768
12 D = Saturday, E = 0-5, S = 0-8, WH = 0-5	20	AL = body shop areas	20	0.9768
13 D = Thursday, A = 26-30, S = 8-16, AC=Laceration injury	20	AL = body shop areas	20	0.9768
14 M = April, A = 31-35, E = 0-5, MW = 5,000-5,999	19	AL = body shop areas	19	0.9754
15 M = April, A = 31-35, S = 16-24, WP=Body Shop Operator	19	AL = body shop areas	19	0.9754
16 AO=finger, A = 31-35, WH = 0-5, MP = 20,000-21,999	18	AL = body shop areas	18	0.9736
17 AO = toe, E = 6e10, S = 16e24, AC=Laceration injury	18	AL = body shop areas	18	0.9736
18 AO = toe, WH = 0-5, WP=Body Shop Operator, AC=Struck by object	18	AL = body shop areas	18	0.9736
19 M = April, S = 0-8, MW = 5,000-5,999, WP=Body Shop Operator	18	AL = body shop areas	18	0.9736
20 M = July, A = 31-35, E = 0-5, WH = 0-5	18	AL = body shop areas	18	0.9736

A, age; AC, accident cause; AO, affected organ; D, days; E, experience; ES, education status; LWD, lost working days; M, month; MP, Annually production; MW, Annually wage; S, stage; WH, working hours; WP, worker position.

contributors to accident risk. The extracted rules indicated a higher frequency of accidents during the later hours of work shifts, a pattern that may be attributed to fatigue, reduced attention, or cumulative physical and cognitive load. The concurrence of limited work experience, demanding operational tasks, and specific time periods within the shift formed recurring patterns in the association rules.

Among the association rule mining approaches,

the Predictive Apriori algorithm produced rules with notably high predictive accuracy. These rules provided a clear framework for prioritizing preventive measures by identifying conditions under which accidents are most likely to occur. However, the study also acknowledged several limitations, including the potential for overfitting in training datasets, reliance on default algorithm parameters, and the need for external validation to enhance the generalizability of the findings (Table2).

Despite these limitations, the results underscore the practical value of data mining in occupational safety research. By revealing complex and non-obvious relationships among accident-related variables, data mining techniques complement traditional analytical approaches and enable a deeper understanding of accident causation mechanisms. The integration of production data further enriches this understanding by situating occupational accidents within their broader organizational and operational context.

4. CONCLUSIONS

This study demonstrated that the application of data mining techniques to occupational accident data, particularly when integrated with production and productivity indicators, can effectively uncover meaningful relationships among human, temporal, organizational, and operational factors influencing accident occurrence. The combined use of classification methods and association rule mining provided a comprehensive analytical framework capable of generating both predictive insights and interpretable knowledge.

The identification of high-risk groups—such as workers with limited job experience, specific

operational roles, and critical time periods within work shifts—offers a concrete basis for designing targeted preventive interventions. The extracted association rules, especially those generated by the Predictive Apriori algorithm, can inform the development of focused training programs, revisions to shift scheduling practices, and the prioritization of safety inspections in high-risk production areas.

Overall, the findings confirm that data mining represents a powerful and practical approach for learning from occupational accidents and for supporting proactive safety management strategies. By leveraging existing organizational data resources, industrial organizations can move beyond reactive accident analysis toward predictive and preventive safety practices. The study provides a clear methodological pathway for future research and practical applications aimed at enhancing occupational health and safety performance in complex manufacturing environments.

5. AUTHORS' DECLARATION

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

داده‌کاوی ناشی از کار در یک شرکت خودروسازی به منظور تعیین علل ریشه‌ای و اجرای راهکارهای پیشگیرانه

محمدرضا بهره‌بر، سعید گیوه‌چی*

دانشکده محیط‌زیست، گروه ایمنی، بهداشت و محیط‌زیست، دانشگاه تهران، تهران، ایران.

تاریخ دریافت: ۱۴۰۴/۱۲/۱، تاریخ پذیرش: ۱۴۰۵/۴/۳۱

چکیده

مقدمه: امروزه حجم وسیعی از داده‌های حوزه بهره‌وری، ایمنی و سلامت شغلی در صنعت خودروسازی همانند سایر شاخه‌های صنعتی جمع‌آوری می‌شود. تبدیل این داده‌ها به اطلاعات مفید برای بنگاه‌های اقتصادی روزبه‌روز حیاتی‌تر شده است. داده‌کاوی نقش قابل‌توجهی در پردازش و استخراج بینش‌های ارزشمند از این مجموعه داده‌ها ایفا می‌کند. هدف این مطالعه، تحلیل داده‌های مربوط به حوادث شغلی منجر به روزهای از دست رفته در یک شرکت تولیدی خودروساز ایرانی در بازه زمانی ۱۳۹۲ الی ۱۴۰۳ با استفاده از تکنیک‌های داده‌کاوی می‌باشد.

روش کار: این مطالعه به تحلیل حوادث شغلی منجر به آسیب در بازه زمانی ۱۳۹۲ تا ۱۴۰۳ پرداخت. داده‌های حوادث با اخذ مجوز از مدیریت HSE استخراج و پس از حذف دو حادثه منجر به مرگ، ۲۱۶۷ حادثه منجر به آسیب مورد بررسی قرار گرفت. داده‌ها با استفاده از نرم‌افزار داده‌کاوی WEKA نسخه ۳,۸,۶ پردازش و بر اساس ۱۴ صفت اصلی نظیر متغیرهای فردی، شغلی، زمانی، علل حادثه، محل وقوع، تولید و دست‌مزد طبقه‌بندی شدند. در مجموع، این صفات به ۱۰۸ زیررده تفکیک و برای تحلیل الگوهای حوادث شغلی استفاده شدند.

یافته‌ها: بر اساس روش‌های آزمون مختلف، الگوریتم‌های k -نزدیک‌ترین همسایه (k -NN) و ماشین بردار پشتیبان (SVM) قادر به طبقه‌بندی داده‌ها و پیش‌بینی نوع، شدت و علل حوادث بودند. اگرچه این الگوریتم‌ها در مجموعه آموزش دقت ۱۰۰٪ را نشان دادند، اما کاهش این دقت به ۶۶٪ در اعتبارسنجی متقابل، حاکی از وقوع پدیده بیش‌برازش (Overfitting) در مدل‌هاست. در بخش کاوش الگوهای همبستگی، الگوریتم‌های Apriori و Predictive Apriori موفق به استخراج قواعدی با شاخص‌های عملکردی بالا شدند؛ به‌طوری‌که قواعد استخراج‌شده دارای سطح اطمینان بالای ۸۳٪ و دقت پیش‌بینی بین ۹۷,۹٪ تا ۹۸٪ بودند. کمی‌سازی این قواعد نشان داد که بیشترین همبستگی با وقوع حوادث، در ترکیب عواملی چون سابقه کار پایین (۰ تا ۵ سال)، ساعات پایانی شیفت (۴ تا ۸ ساعت پس از شروع شیفت) و اشتغال در پست اپراتور بدنه‌سازی نهفته است.

نتیجه‌گیری: یافته‌ها نشان داد که تلفیق روش‌های داده‌کاوی، به‌ویژه طبقه‌بندی و کاوش الگوهای همبستگی، امکان شناسایی الگوهای پنهان و عوامل ریشه‌ای حوادث شغلی در محیط‌های تولیدی را فراهم می‌سازد. با این حال، بروز بیش‌برازش در مدل‌های طبقه‌بندی، یک محدودیت جدی پژوهش است که بر ضرورت ارزیابی تعمیم‌پذیری مدل‌ها و استفاده از داده‌های متعادل‌تر تأکید دارد. قواعد استخراج‌شده، به‌ویژه توسط Predictive Apriori، امکان طراحی برنامه‌های آموزشی ویژه، تنظیم توالی استراحت و انجام مداخلات ایمنی هدفمند برای گروه پرخطر، یعنی کارگران کم‌سابقه (صفر تا ۵ سال) و اپراتورهای تولید بدنه در ساعات پایانی شیفت (۴ تا ۸ ساعت پس از شروع کار) را فراهم می‌آورد.

کلمات کلیدی: حوادث ناشی از کار، داده‌کاوی، تحلیل الگوهای همبستگی، ایمنی و سلامت شغلی

مقدمه

در طول تاریخ، تمایل به داشتن زندگی امن و بدون خطر یکی از خواسته‌های مشترک انسان‌ها بوده است، چرا که ایمنی و امنیت به عنوان بخشی جدایی‌ناپذیر از وجود انسانی تلقی می‌شود (۱). با شتاب روزافزون جهانی شدن اقتصاد، صنعت با چالش‌هایی بی‌سابقه از زمان اولین انقلاب صنعتی روبرو شده است. پیشرفت فناوری اطلاعات و جهانی شدن اقتصاد، فشارهای محیطی زیادی را بر صنعت تحمیل کرده‌اند، از جمله بازارهای رکود، تغییرات در تقاضای مشتریان و رقابت با شرکت‌هایی که نیروی کار ارزان تری دارند (۲). در چنین شرایطی، شرکت‌هایی که بتوانند محصولات و خدمات نوآورانه عرضه کنند، تنوع ایجاد کنند و نیازهای ذی‌نفعان خود را با کمترین منابع پاسخ دهند، در مسیر موفقیت قرار خواهند گرفت (۳).

یکی از عواملی که باعث اختلال در سیستم تولید می‌شود، وقوع حوادث شغلی در صنعت خودروسازی است. این حوادث، به دلیل توقف کار و انجام بررسی‌های لازم توسط سازمان‌ها و مقامات قانونی، تولید و بهره‌وری را تحت تأثیر قرار می‌دهند (۴، ۵). علاوه بر این، زمانی که کارکنان به علت درد یا جراحت ناشی از حادثه قادر به انجام وظایف عادی خود نیستند (۶) و یا تجهیزات و زیرساخت‌ها دچار خسارت شده و نیاز به تعمیرات پرهزینه دارند، زمان مفید کار نیز از دست می‌رود (۷، ۴-۱۰).

در صنایع تولیدی، وقوع حوادث شغلی با پیامدهای گسترده‌ای همراه است. این حوادث نه تنها سلامت و جان نیروی کار را به خطر می‌اندازند، بلکه هزینه‌های مالی، تبعات حقوقی و آسیب به اعتبار سازمان را نیز در پی دارند (۱۱). از منظر جهانی، گزارش‌های سازمان بین‌المللی کار نشان می‌دهد که سالانه حدود ۲.۹ میلیون نفر در اثر بیماری‌ها و حوادث شغلی جان خود را از دست می‌دهند که از این میان، حدود ۳۳۰ هزار نفر مربوط به حوادث شغلی است (۱۲). همچنین هر سال نزدیک به ۳۹۵ میلیون حادثه غیرمرگبار در محیط‌های کاری رخ می‌دهد. این حوادث سالانه هزینه‌ای در حدود ۳ تریلیون دلار آمریکا، معادل نزدیک به ۳.۹ درصد تولید ناخالص داخلی

جهان، به اقتصاد جهانی تحمیل می‌کنند (۱۳).

در ایران نیز حوادث شغلی سهم قابل توجهی از آسیب‌های ناشی از کار را به خود اختصاص داده‌اند. بر اساس سالنامه آماری وزارت تعاون، کار و رفاه اجتماعی در سال ۱۴۰۱، حدود ۳۶.۵ درصد از کل حوادث شغلی کشور در صنایع تولیدی رخ داده است. همچنین گزارش پزشکی قانونی در سال ۱۴۰۲ از وقوع حدود ۲۱۱۵ مورد فوت و ۲۷۳۷۷ مورد مصدومیت ناشی از حوادث کار خبر می‌دهد. افزون بر این، آمارها نشان می‌دهد که به طور متوسط روزانه بین ۵ تا ۷ نفر در کشور بر اثر حوادث کار جان خود را از دست می‌دهند (۱۴).

در میان صنایع تولیدی، صنعت خودروسازی به دلیل ماهیت فرایندها، تجهیزات و شرایط کاری، از جمله صنایع پرریسک به شمار می‌رود (۱۱). در این صنعت، علاوه بر حوادث منجر به مرگ و مصدومیت، انواع حوادث فردی نیز به طور مکرر رخ می‌دهد. برای مثال، در ایالات متحده، شاخص صدمات غیرمرگبار در بخش تولید خودرو در سال ۲۰۱۸ برابر با ۶.۳ مورد به ازای هر ۱۰۰ کارگر تمام وقت بوده است که بیش از دو برابر نرخ کلی در سایر بخش‌های صنعتی است. این داده‌ها نشان می‌دهد که صنعت خودروسازی به طور ذاتی نیازمند نظام‌های دقیق و اثربخش مدیریت ایمنی است (۱۵). در شرکت خودروساز مورد مطالعه نیز نرخ بروز حوادث شغلی طی سال‌های ۱۳۹۲ تا ۱۴۰۳ بین ۲.۵۵ تا ۴.۱۷ گزارش شده است که بیانگر روندی افزایشی در وقوع حوادث ناشی از کار است (۱۱).

در این شرایط، پردازش داده‌های تجمع یافته، استخراج بینش‌های مفید و بهره‌گیری از آن‌ها به منظور ارتقای سطح ایمنی محیط کار و حفظ سلامت کارکنان از جمله تکالیف ضروری است. این ارزیابی‌ها می‌توانند با استفاده از روش‌های آماری مرسوم یا تکنیک‌های داده‌کاوی انجام شوند (۱۱). در بسیاری از موارد، رویکردهای آماری کلاسیک در تحلیل مجموعه داده‌های پیچیده مانند داده‌های حوادث شغلی با محدودیت مواجه‌اند، درحالی‌که روش‌های داده‌کاوی توانایی کشف روابط پیچیده‌ای را

اعمال الگوریتم Apriori بر داده‌های مخاطرات پنهان در صنعت معادن، راهکارهایی جهت مدیریت و پیشگیری ارائه دادند که به کاهش قابل توجه خسارات ناشی از حوادث صنعتی منجر شده است (۲۰).

در همین حال، تحلیل دقیق علل ریشه‌ای و آماری حوادث نقش محوری در استراتژی‌های پیشگیرانه ایفا می‌کند. Mevsim (۲۱) با استفاده از تحلیل درخت خطا نشان داد که عواملی نظیر نقص سیستم‌های تهویه و تجهیزات معیوب، از علل اصلی آتش‌سوزی‌ها و انفجارها هستند؛ نتایج این مطالعه حاکی از آن است که اگرچه احتمال وقوع حادثه در یک دوره ۹۰ ماهه قطعی است، اما اقدامات اصلاحی هوشمند می‌تواند این دوره را به ۱۸۰ ماه افزایش دهد. همچنین Erdogan با بهره‌گیری از تکنیک‌های تحلیل واریانس، شدت حوادث شغلی را بر اساس جراحات و تلفات ارزیابی و سطوح ریسک را تعیین نمود (۲۲). در این سیاق، Sanmiquel با تحلیل داده‌های بخش خودروسازی اسپانیا، داده‌کاوی را ابزاری ارزشمند برای شناسایی علل ریشه‌ای حوادث معرفی کرد (۲۳) و Ak مدلی مبتنی بر سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) توسعه داد که در مقایسه با روش‌های سنتی، کارایی و نتایج بهتری را نشان داده است (۲۴).

پس از سال ۲۰۲۰، رویکردهای پژوهشی به سمت استفاده از هوش مصنوعی و مدل‌های ترکیبی تغییر جهت دادند. Gerassis ضمن بررسی کاربرد داده‌کاوی در تحلیل حوادث، بر لزوم تجمیع داده‌های آماری جهت دستیابی به سطوح ایمنی هدفمند تأکید ورزید (۲۵). همزمان، Iphar و Çukurluöz روشی مبتنی بر منطق فازی را معرفی کردند که از دانش خبرگان برای شناسایی ریسک‌ها بهره می‌برد (۲۶)، در حالی که Aliabadi از شبکه‌های بیزی برای کمی‌سازی تأثیر خطاهای انسانی و عوامل سازمانی استفاده نمود (۲۷). علاوه بر این، Yedla با استفاده از یادگیری ماشین، متغیرهایی همچون زمان شیفت و تجربه کارکنان را به‌عنوان پیش‌بینی‌کننده‌های روزهای کاری از دست‌رفته مورد تحلیل قرار داد (۲۸). مطالعات نوین‌تر، بر مدل‌سازی‌های پیچیده جهت

دارند که روش‌های آماری سنتی قادر به آشکارسازی آن‌ها نیستند. داده‌کاوی یک حوزه علمی است که استخراج اطلاعات معنادار و قابل عمل از مجموعه داده‌های پیچیده و حجیم را ممکن می‌سازد (۱۶).

تحقیقات حوزه ایمنی و بهداشت حرفه‌ای در صنعت خودروسازی در موازات پیشرفت‌های فناورانه نیازمند توسعه است. مطالعات این حوزه عمدتاً بر شناسایی ریسک‌های مرتبط، توسعه روش‌های اندازه‌گیری و پیاده‌سازی استراتژی‌های پیشگیرانه تمرکز داشته‌اند. در برخی مرورهای نظام‌مند، روندهای پژوهشی در خصوص حوادث شغلی در بخش خودروسازی بررسی شده و یافته‌های آن‌ها حاکی از آن است که بخش عمده‌ای از مطالعات به شناسایی علل اصلی وقوع حوادث پرداخته‌اند، درحالی‌که سایر مطالعات بر استراتژی‌های پیشگیری از حوادث و ارزیابی چالش‌ها و پیامدهای حوادث شغلی متمرکز بوده‌اند. با این وجود، نیاز مبرمی به تحقیقات عمیق‌تر در زمینه علل ریشه‌ای حوادث و تدوین راهبردهای مؤثر کاهش مخاطرات احساس می‌شود (۱۷).

کاربرد داده‌کاوی در حوزه ایمنی و بهداشت حرفه‌ای در صنعت خودروسازی نسبتاً کمتر مورد کاوش قرار گرفته است. برخی مطالعات به بررسی ادغام فناوری‌های نوین در پیشگیری از حوادث، چالش‌های موجود و چشم‌اندازهای آتی این حوزه پرداخته‌اند. این بررسی‌ها فقدان چارچوب استاندارد برای پژوهش‌های پیشگیری از حوادث و کاستی‌های مهم در تفسیر نتایج تجربی را برجسته کرده‌اند که ضرورت توسعه روش‌های نوآورانه از جمله داده‌کاوی را برای تحلیل دقیق‌تر حوادث و ارتقای پروتکل‌های ایمنی در محیط‌های صنعتی نشان می‌دهد (۱۸).

در راستای ارتقای سیستم‌های ایمنی و بهداشت حرفه‌ای، بهره‌گیری از فناوری‌های نوین تحلیل داده امری اجتناب‌ناپذیر است. Dessureault و همکاران در سال ۲۰۰۷ با تمرکز بر کاربرد داده‌کاوی، پایگاه‌های داده مرتبط را به‌روزرسانی نمودند تا زمینه لازم برای تحلیل دقیق حوادث و ریسک‌های ایمنی توسط نرم‌افزارهای مدرن فراهم گردد (۱۹). متعاقباً، Li و Shuangyue با

و اعمال فشار زمانی بر کارکنان و سرپرستان، احتمال بروز شرایط نایمن و وقوع حادثه را افزایش دهد. بنابراین، بررسی هم‌زمان داده‌های تولید، بهره‌وری و حوادث می‌تواند به شناسایی الگوهای پنهان و سازوکارهای سازمانی مؤثر بر ایمنی کمک کند. بر این اساس، پرسش اصلی پژوهش این است که آیا شاخص‌های تولید و بهره‌وری با الگوهای وقوع حوادث در صنعت خودروسازی رابطه معناداری دارند یا خیر، و کدام ترکیب از این متغیرها بیشترین نقش را در تبیین ریسک حادثه ایفا می‌کند؟

روش کار

روش شناسی

این مطالعه، حوادث شغلی منجر به آسیب که منجر به روزهای از دست‌رفته کرده است را در بازه زمانی ۱۳۹۲ تا ۱۴۰۳ تحلیل نموده است. داده‌های گردآوری‌شده حوزه حوادث شغلی منجر به آسیب و محل وقوع حوادث با استفاده از روش‌های داده‌کاوی مرتبط شد. به منظور تحلیل داده‌ها از نرم‌افزار داده‌کاوی دارای کُد منبع در دسترس WEKA نسخه ۳,۸,۶ استفاده شد. شکل ۱ متدولوژی مطالعه را نشان می‌دهد. جدول ۱ اطلاعات مربوط به تعداد حوادث رخ‌داده در این سال‌ها را ارائه می‌دهد. اطلاعات تفصیلی این حوادث با مجوزهای لازم از مدیریت HSE آن سازمان استخراج شد. در مجموع ۲۱۶۹ حادثه شغلی در سال‌های مورد تحلیل رخ داده است. از این تعداد، ۲۱۶۷ حادثه منجر به آسیب بودند که داده‌های اصلی این مطالعه را تشکیل می‌دهند.

داده‌های مربوط به حوادث شغلی از طریق یک سامانه ثبت ساختاریافته و جامع جمع‌آوری می‌گردد. در این سامانه، علاوه بر اطلاعات هویتی و جایگاه شغلی فرد آسیب‌دیده، مؤلفه‌های دقیق زمانی و مکانی (از جمله تاریخ، ساعت وقوع، ایستگاه کاری و وضعیت شیفت یا اضافه‌کاری) ثبت می‌شود. همچنین، پیامدهای بالینی شامل نوع جراحت و عضو آسیب‌دیده به همراه مداخلات اورژانسی و درمانی انجام‌شده به دقت مستند می‌گردند. به منظور تحلیل ریشه‌ای حوادث، مواردی نظیر وضعیت

پیش‌بینی دقیق پیامدها متمرکز شده‌اند. Opokua و Opoku سیستم استنتاج عصبی-فازی تطبیقی را برای پیش‌بینی نرخ بقای کارکنان در محیط‌های پرخطر توسعه دادند (۲۹) و Sanmiquel مدلی گرافیکی برای پیش‌بینی فراوانی حوادث ارائه نمود (۳۰). از سوی دیگر، You از الگوریتم SVM برای ارزیابی سطح ریسک‌های شیمیایی بهره برد (۳۱). به‌طور مشابه، Li ترکیبی از استخراج متن و شبکه‌های بیزی را برای شناسایی علل ریشه‌ای مانند مدیریت ناکافی به کار گرفت (۳۲). نهایتاً، Bayraktar با تحلیل گسترده حوادث و استفاده از روش‌های Life Table و Kaplan-Meier، قابلیت اعتماد تحلیل‌های بقا در بررسی حوادث شغلی را اثبات نمود (۳۳) و Baraza نیز تأثیر متغیرهای متعدد بر نرخ شدت حوادث را تبیین کرد (۳۴). اگرچه مطالعات متعددی به تحلیل علل، پیامدها و پیشگیری از حوادث در صنعت خودروسازی پرداخته‌اند، اما هنوز پژوهش‌های محدودی به‌صورت یکپارچه داده‌های عملیاتی سازمان، به‌ویژه شاخص‌های تولید و بهره‌وری، را همراه با داده‌های حوادث تحلیل کرده‌اند تا رابطه بین فشار تولید و الگوهای بروز حادثه آشکار شود.

با توجه به افزایش حجم داده‌های ایمنی و بهداشت حرفه‌ای، پردازش این داده‌ها برای استخراج بینش‌های قابل اجرا ضروری به‌نظر می‌رسد. درحالی‌که نقش داده‌کاوی در این زمینه انکارناپذیر است، مطالعات محدودی بر تحلیل علل ریشه‌ای حوادث صنعتی با استفاده از تکنیک‌های داده‌کاوی و ارائه راهکارهای پیشگیرانه متمرکز شده‌اند. این مطالعه در پی پر کردن خلأ در تحلیل داده‌ای در صنعت خودروسازی، توسعه روش‌های تحلیلی پیشرفته و طراحی راهبردهای پیش‌بینانه جهت پیشگیری از حوادث است. نوآوری کلیدی این تحقیق استفاده از داده‌های سالیانه تولید و بهره‌وری در کنار داده‌های کلاسیک حوادث به منظور آشکارسازی روابط میان نرخ‌های تولید و حوادث صنعتی می‌باشد.

بر مبنای نظریه فشار تولید، افزایش اهداف تولید و الزامات بهره‌وری می‌تواند از طریق تشدید سرعت کار، افزایش خستگی، کاهش فرصت‌های توقف و نگهداری،



شکل ۱: روش‌شناسی مطالعه.

جدول ۱: تعداد حوادث شغلی منجر به آسیب به تفکیک سال

سال	تعداد حوادث شغلی منجر به آسیب	تعداد حوادث شغلی منجر به آسب به ازای نرخ تولید
۱۳۹۲	۲۶۷	۹۲.۸۲
۱۳۹۳	۲۵۱	۸۶.۵۸
۱۳۹۴	۱۹۴	۷۷.۵۴
۱۳۹۵	۲۲۴	۸۵.۷۴
۱۳۹۶	۱۳۶	۶۱.۱۶
۱۳۹۷	۱۵۱	۷۷.۵۴
۱۳۹۸	۱۳۷	۸۰.۲۰
۱۳۹۹	۱۳۳	۱۱۴.۲۲
۱۴۰۰	۱۱۲	۱۱۰.۹۴
۱۴۰۱	۱۹۰	۱۴۵.۴۴
۱۴۰۲	۱۶۴	۱۴۱.۲۸
۱۴۰۳	۲۰۸	۱۶۵.۸۶

پیش‌بینی ویژگی‌های مرتبط با حوادث ثبت‌شده، از جمله محل وقوع، علت مستقیم و سایر ویژگی‌های حادثه مورد استفاده قرار گرفته‌اند. از این‌رو، نتایج مطالعه باید در چارچوب طبقه‌بندی الگوهای حوادث تفسیر شوند و نه پیش‌بینی احتمال وقوع حادثه در جمعیت کارکنان.

پردازش داده‌های حادثه

داده‌های خام حوادث شغلی پردازش و طبقه‌بندی شدند تا پارامترهای ورودی برای تحلیل‌های آماری آماده شوند. کلیه داده‌های حوادث مربوط به دوازده سال از پایگاه‌داده استخراج و در ابتدا حوادث مرگبار شناسایی

استفاده از تجهیزات حفاظت فردی، شرح فعالیت حین حادثه، ابزار مورد استفاده و علل بروز رویداد مورد ارزیابی قرار می‌گیرد. در نهایت، پیامدهای سازمانی رویداد مانند تعداد روزهای کاری از دست‌رفته و ارجاع پرونده به کمیته پیشگیری جهت اتخاذ اقدامات اصلاحی در این سامانه ثبت و لحاظ می‌شود. لازم به ذکر است که مجموعه داده این پژوهش صرفاً شامل سوابق حوادث ثبت‌شده بوده و هیچ نمونه‌ای از شرایط فاقد حادثه (Non-accident cases) در پایگاه داده وجود نداشت. بنابراین، مدل‌های طبقه‌بندی توسعه‌یافته در این مطالعه با هدف پیش‌بینی وقوع حادثه طراحی نشده‌اند، بلکه برای طبقه‌بندی و

جدول ۲: صفات و زیرطبقه‌های آن‌ها

صفت	زیرطبقه‌ها
عضو آسیب‌دیده	بازو، تنه، صورت، انگشت دست، سر، زانو، پا، انگشت پا، کمر.
ماه	فروردین، اردیبهشت، خرداد، تیر، مرداد، شهریور، مهر، آبان، آذر، دی، بهمن، اسفند.
روز	شنبه، یکشنبه، دوشنبه، سه‌شنبه، چهارشنبه، پنجشنبه، جمعه.
سن	۱۸-۲۵، ۲۶-۳۰، ۳۱-۳۵، ۳۶-۴۰، ۴۱-۴۵، ۴۵-۴۹
سابقه کاری	۰-۵، ۶-۱۰، ۱۱-۱۵، ۱۶-۲۰، ۲۱-۲۵، ۲۶-۳۰
وضعیت تحصیلی	دیپلم، فوق دیپلم، لیسانس، فوق لیسانس، دکتری.
شیفت	۰-۸، ۸-۱۶، ۱۶-۲۴
ساعات کاری	۱ دقیقه الی ۱ ساعت کار (۱-۰)، ۱ ساعت الی ۴ ساعت کار (۴-۱)، ۴ ساعت الی ۸ ساعت کار (۸-۴)
روزهای کاری از دست‌رفته (۲۳)	۰، ۱-۳، ۴-۲۸، ۲۹-۱۸۰، ۱۸۱-۳۶۵
تولید سالیانه (تعداد خودرو)	۷۹۹۹۹، ۸۰۰۰۰-۹۹۹۹۹، ۱۰۰۰۰۰-۱۱۹۹۹۹، ۱۲۰۰۰۰-۱۳۹۹۹۹، ۱۴۰۰۰۰-۱۵۹۹۹۹، ۱۶۰۰۰۰-۱۷۹۹۹۹، ۱۸۰۰۰۰-۱۹۹۹۹۹، ۲۰۰۰۰۰-۲۱۹۹۹۹
دستمزد سالیانه (دلار)	۳۹۹۹۹-۴۹۹۹۹، ۵۰۰۰۰-۵۹۹۹۹، ۶۰۰۰۰-۶۹۹۹۹، ۷۰۰۰۰-۷۹۹۹۹
پست شغلی	اپراتور سالن‌های مونتاژ، اپراتور سالن‌های رنگ، اپراتور سالن‌های بدنه سازی، اپراتور سالن‌های موتورسازی، اپراتور سالن‌های پرس، اپراتور خدمات فنی، کارشناس کیفیت، کارشناس آزمایشگاه‌ها، کارشناس ایمنی و بهداشت صنعتی، اپراتورهای حمل و نقل، سرپرستان تولیدی، سرگروه‌های تولیدی، کارشناسان برنامه ریزی، اپراتورهای خدماتی.
علت مستقیم حادثه	خرابی تجهیزات، استفاده نامناسب از ابزار یا ماشین‌آلات، لغزش و سقوط، برخورد با اجسام متحرک، برخورد با اجسام تیز و برنده، جابجایی نامناسب مواد سنگین، قرارگیری در معرض مواد شیمیایی خطرناک، خطرات الکتریکی، حوادث مرتبط با گرما، خطای انسانی یا آموزش ناکافی. (براساس دسته بندی علل مستقیم حوادث در شرکت مورد مطالعه)
مکان وقوع حادثه	سالن‌های مونتاژ، سالن‌های رنگ، سالن‌های بدنه سازی، سالن‌های موتورسازی، سالن‌های پرس، محوطه عمومی داخلی شرکت، محوطه خارجی شرکت، آزمایشگاه‌ها، محیط‌های اداری، کانال‌های تاسیساتی، رستوران و آشپزخانه صنعتی، کارگاه‌های تاسیساتی و تعمیراتی، سیرکلاسیون‌های رنگ.

سالیانه و حقوق سالیانه. ۱۴ صفت انتخاب‌شده برای بررسی حوادث به ۱۰۸ زیررده تقسیم شدند. جدول ۲ صفت‌ها و زیررده‌های انتخاب‌شده برای حوادث را نشان می‌دهد.

صفت‌ها و زیررده‌های جدول ۲ را از لحاظ تعداد حوادث مورد تحلیل قرار گرفته و مشخص گردید که رایج‌ترین اندام‌های آسیب‌دیده انگشتان دست و انگشتان پا بوده و حوادث تقریباً در هر ماه سال به‌طور مشابه رخ داده‌اند. از نظر روزهای هفته، حوادث در روزهای دوشنبه، سه‌شنبه، چهارشنبه، پنجشنبه و جمعه تقریباً برابرند. گروه سنی ۳۱-۳۵ بیشترین تعداد حوادث را به خود اختصاص داده‌اند. کارگرانی با سابقه کار ۵-۰ سال بیشترین در معرض حادثه قرار دارند. از نظر وضعیت تحصیلی، دارندگان تحصیلات دیپلم بیشترین تعداد حوادث را داشته‌اند. با توجه به شیفت وقوع حوادث و ساعات کاری داخل شیفت،

شدند. در بازه زمانی مورد تحقیق، تنها دو حادثه منجر به مرگ در مجموعه رخ داده است (درسال‌های ۱۳۹۲ و ۱۳۹۳). پس از حذف حوادث مرگبار، ۲۱۶۷ حادثه منجر به آسیب مورد تحلیل قرار گرفتند.

حوادث منجر به آسیب با استفاده از ۱۴ متغیر اصلی (که در WEKA به‌عنوان ویژگی (attributes) تعریف شده‌اند) گروه‌بندی شدند. این ویژگی‌ها با ارزیابی مطالعات پیشین (۲۱-۲۳)، تجربه‌های گذشته شرکت و شرایط عملیاتی تعیین شدند. این ۱۴ ویژگی حیاتی شامل متغیرهایی هستند که مستقیماً با حوادث مرتبط‌اند (اندام آسیب‌دیده، ماه وقوع حادثه، روز هفته وقوع حادثه، سن کارگر آسیب‌دیده، سابقه کار کارگر آسیب‌دیده، وضعیت تحصیلات کارگر آسیب‌دیده، شیفت، ساعات کاری در شیفت، تعداد روزهای ازدست‌رفته، شغل کارگر، علت مستقیم حادثه و مکان وقوع حادثه) و همچنین تولید

کاوش الگوهای همبستگی و وابستگی روشی از داده‌کاوی است که به دنبال کشف الگوها و قواعد اگر - آنگاه (if $\rightarrow$ then) درون یک مجموعه داده است؛ یعنی می‌کوشد ترکیباتی از ویژگی‌ها را پیدا کند که هم‌زمان با فراوانی مشخصی رخ می‌دهند و سپس آن ترکیب‌ها را به صورت قواعد قابل فهم گزارش می‌کند (۱۱).

اجرای روش‌های داده‌کاوی بر روی داده‌های حوادث

فرآیندهای طبقه‌بندی با استفاده از الگوریتم‌های مناسب داده‌کاوی بر روی پایگاه داده حوادث منجر به آسیب انجام شد. در فرآیند طبقه‌بندی دو هدف اصلی دنبال شد: نخست تعیین الگوریتمی که بهترین بازنمایی را از کل داده‌ها ارائه می‌دهد و دوم سنجش توانایی الگوریتم‌های داده‌کاوی در پیش‌بینی پارامتر مورد نظر. در مدل‌سازی‌های این مرحله، مجموعه آموزش (training set) و اعتبارسنجی (cross-validation) به عنوان روش‌های آزمون انتخاب شدند. علاوه بر این روش‌های آزمون، آمارهای Kappa، میانگین قدرمطلق خطا (MAE) و ریشه میانگین مربعات خطا (RMSE) برای تعیین دقت طبقه‌بندی الگوریتم‌ها مورد استفاده قرار گرفتند. دقت الگوریتم‌ها بر مبنای درصد داده‌هایی که به درستی در کل مجموعه داده طبقه‌بندی شده‌اند تعیین شد. تنظیمات پیش‌فرض الگوریتم‌های داده‌کاوی اعمال شده تغییر داده نشدند؛ تنها جهت بهبود عملکرد برخی پارامترها تعدیلاتی کوچک صورت گرفت تا نتایج بهینه‌تری حاصل شود.

در این پژوهش، الگوریتم ZeroR به عنوان مدل مبنا به کار گرفته شد تا معیاری ساده برای مقایسه عملکرد سایر الگوریتم‌ها فراهم شود. این الگوریتم بدون در نظر گرفتن ویژگی‌های ورودی، تنها بر اساس فراوان‌ترین کلاس یا میانگین متغیر هدف پیش‌بینی انجام می‌دهد و به عنوان شاخص پایه برای ارزیابی میزان بهبود دقت سایر مدل‌ها مورد استفاده قرار می‌گیرد (۳۶). همچنین، الگوریتم Naive Bayes به عنوان یک طبقه‌بند احتمالاتی مبتنی بر قضیه بیز و فرض استقلال ویژگی‌ها (۳۷)، الگوریتم J48

بیشترین تعداد حوادث در شیف ۸-۱۶ و در بازه ۴-۸ ساعت از شیف رخ داده است. تعداد حوادث در ساعات اولیه شیف (۱-۰ ساعت) نسبتاً کم است. بیشترین تعداد روزهای از دست رفته مربوط به بازه ۴-۲۸ روز است. تعداد حوادث بر اساس تولید سالانه و حقوق سالانه نوسان دارد. از نظر جایگاه شغلی، اپراتور سالن‌های بدنه سازی بیشترین تعداد حوادث را داشته‌اند. به طور مشابه، بر اساس علت حادثه، برخورد با اجسام تیز و برنده شایع‌ترین علت حوادث هستند.

متغیرهای تولید سالیانه و دستمزد سالیانه به عنوان شاخص‌های سطح سازمان و در مقیاس سال استخراج شدند و به صورت بازه‌های طبقه‌ای وارد تحلیل شدند. از آنجا که داده‌های حوادث در سطوح زمانی روز، ماه، شیف و ساعت ثبت شده بودند، این دو متغیر به عنوان متغیرهای زمینه‌ای سالانه در کنار داده‌های حادثه استفاده شدند و تفسیر آن‌ها نیز در همان سطح سالانه انجام شد. به منظور همگن‌سازی مقیاس متغیرها و جلوگیری از اثر اختلاف واحد اندازه‌گیری، مقادیر تولید سالیانه و دستمزد سالیانه پس از تخصیص به سال وقوع هر حادثه، به بازه‌های طبقه‌بندی شده تبدیل شدند. در نتیجه، این متغیرها به صورت متغیرهای اسمی (Categorical Attributes) وارد الگوریتم‌های داده‌کاوی شدند و نیازی به نرمال‌سازی عددی متداول در الگوریتم‌های مبتنی بر فاصله وجود نداشت.

یافته‌ها

مطالعات داده‌کاوی در دو مرحله انجام شد. در مرحله اول، الگوریتم‌های مختلف برای طبقه‌بندی و پیش‌بینی داده‌های حادثه به کار گرفته شدند. هدف در این مرحله آشکار ساختن توانمندی الگوریتم‌های داده‌کاوی در طبقه‌بندی و پیش‌بینی رخداد‌های حادثه‌ساز بود. در مرحله دوم، کاوش الگوهای همبستگی و وابستگی (association rule mining) بر روی داده‌های حوادث اعمال شد. هدف در این مرحله کشف روابط پنهان و معتبر میان متغیرها به صورت قواعد بود (۳۵).

جدول ۳: الگوریتم‌های اعمال شده روی مجموعه داده و نتایج به دست آمده

الگوریتم	روش های آزمون الگوریتم ها							
	Cross-validation				Training set			
	RMSE	MAE	Kappa	Correct classified	RMSE	MAE	Kappa	Correct classified
ZeroR	۰.۲۳	۰.۱۱	۰	۱۲۵۴ (/۵۷.۸۶)	۰.۲۳	۰.۱۱	۰	۱۲۴۵ (/۵۷.۸۶)
NB	۰.۲۱	۰.۰۸۶	۰.۳۳	۱۳۹۴ (/۶۴.۳۲)	۰.۲	۰.۰۸	۰.۴۲	۱۴۸۹ (/۶۸.۷۱)
J48	۰.۲۲	۰.۰۹۳	۰.۳۲	۱۴۲۵ (/۶۵.۷۶)	۰.۱۲	۰.۰۳	۰.۷۶	۱۸۶۷ (/۸۶.۱۶)
kNN	۰.۲۲	۰.۰۹۲	۰.۱۶	۱۳۳۲ (/۶۱.۴۶)	۰.۰۰۱	۰.۰۰۰۸	۱	۲۱۶۷ (/۱۰۰)
MLP	۰.۲۶	۰.۰۸	۰.۲۷	۱۲۳۶ (/۵۷.۰۳)	۰.۰۸	۰.۰۰۹	۰.۹۲	۲۰۷۴ (/۹۵.۷)
SVM	۰.۲۶	۰.۱۵	۰.۳۴	۱۴۳۲ (/۶۶.۰۸)	۰.۲۶	۰.۱۵	۰.۴۵	۱۵۵۳ (/۷۱.۶۶)
RF	۰.۲۱	۰.۰۹	۰.۲۸	۱۳۸۸ (/۶۴.۰۵)	۰.۰۸	۰.۰۳۴	۱	۲۱۶۷ (/۱۰۰)

J48 درخت تصمیم / k-k : نزدیک ترین همسایه / MAE: میانگین خطای مطلق / MLP: پرسپترون چندلایه / NB: بیز ساده (نایو بیز) / RF: جنگل تصادفی / RMSE: ریشه میانگین مربعات خطا (یا جذر میانگین مربعات خطا) / SVM: ماشین های بردار پشتیبان

است وجود داده های نویزی را کنترل کند. مقادیر پرت و داده های افراطی در مجموعه با استفاده از فیلتر بازه میان چارکی (Interquartile Range — IQR) در WEKA شناسایی می شوند. در فرایند پالایش، در صورت شناسایی مقادیر غیرمعنی دار، این مقادیر از مجموعه حذف و داده های نویزی پاک سازی می گردند. هنگام اعمال فیلتر IQR بر مجموعه داده های مورد استفاده در این مرحله از مطالعه، هیچ نتیجه مثبتی از نظر وجود مقادیر پرت یا افراطی به دست نیامد (تعداد مقادیر پرت و افراطی برابر با ۰ بود). این امر نشان می دهد که داده های بی معنی در مجموعه وجود ندارد. پارامترهای عملکردی دیگر نیز تحلیل شدند؛ مقادیر آمار کاپا، MAE و RMSE نشان دهنده برتری الگوریتم های kNN و RF بودند.

داده های استخراج شده از سامانه ثبت حوادث ابتدا از نظر کامل بودن، سازگاری و وجود مقادیر گم شده بررسی شدند. متغیرهای عددی و طبقه ای به صورت جداگانه پایش شدند و پیش از اجرای الگوریتم های داده کاوی، داده ها به قالب تحلیلی مناسب تبدیل گردیدند. برای شناسایی و حذف نقاط پرت، فیلتر IQR در محیط

به عنوان نسخه توسعه یافته درخت تصمیم (۳۸)، الگوریتم kNN به عنوان روش مبتنی بر نزدیک ترین همسایگی (۳۹)، شبکه عصبی مصنوعی چندلایه (MLP) به عنوان مدلی مبتنی بر یادگیری غیرخطی و الگوریتم پس انتشار (۴۰)، ماشین بردار پشتیبان (SVM) به عنوان طبقه بند مبتنی بر بیشینه سازی حاشیه تفکیک کلاس ها (۴۱)، و جنگل تصادفی (RF) به عنوان روش ensemble مبتنی بر ترکیب چندین درخت تصمیم (۴۲) در مجموعه داده اعمال شدند. نتایج حاصل از اجرای الگوریتم های مذکور در جدول ۳ ارائه شده اند.

در روش مجموعه آموزش، الگوریتم های kNN و RF نتایج بسیار خوبی نشان دادند و به دقت ۱۰۰٪ دست یافتند. در داده کاوی، هنگامی که دقت ۱۰۰٪ مشاهده می شود، اولین گمانه معمولاً یادگیری بیش از حد و بیش برآزش (overlearning/overfitting) الگوریتم ها است؛ با این حال، هر نتیجه ۱۰۰٪ الزاماً مشکوک نیست. در صورتی که داده های نویزی (noisy data) در مجموعه وجود نداشته باشد که موجب بیش برآزش شود، این نتیجه می تواند معتبر باشد. برای اطمینان، نرم افزار WEKA قادر

WEKA به‌کار گرفته شد؛ بر اساس نتایج این مرحله، مورد پرت معناداری در مجموعه داده شناسایی نشد. علاوه بر این، به‌منظور اطمینان از مناسب‌بودن داده‌ها برای تحلیل با الگوریتم‌های طبقه‌بندی و کاوش الگو، پیش‌پردازش داده‌ها با هدف کاهش خطاهای ناشی از ثبت ناقص اطلاعات و حفظ کیفیت داده‌ها انجام شد.

در میان الگوریتم‌های مورد استفاده در روش اعتبارسنجی متقاطع، الگوریتم SVM بهترین عملکرد را ارائه داد و 0.66 ٪ از مجموعه داده را به‌درستی طبقه‌بندی نمود. از نظر آمارهای دقت (Kappa, MAE, RMSE)، الگوریتم SVM نیز نسبت به سایر الگوریتم‌های به‌کاررفته عملکرد بهتری نشان داد.

کاوش الگوهای همبستگی و وابستگی برای داده‌های حوادث

در کاوش الگوهای همبستگی و وابستگی، الگوریتم‌های Apriori و Predictive Apriori پیاده‌سازی شدند. توزیع ویژگی‌ها و زیررده‌ها برای هر دو پیاده‌سازی تحلیل شد و بدین‌وسیله دانش عمیق‌تری در مورد منشأ حوادث استخراج گردید.

الگوریتم Apriori پرشمارترین ویژگی‌های مجموعه داده را می‌یابد و با استفاده از آن‌ها الگوهای همبستگی و وابستگی تولید می‌کند. هنگام تولید قواعد توسط این الگوریتم، معیار اعتماد (confidence) تعیین‌شده در پیاده‌سازی لحاظ می‌گردد و قواعد تولیدشده بر اساس این معیار رتبه‌بندی می‌شوند. جدول ۴ بیست قاعده اول تولیدشده برای مکان وقوع حادثه توسط الگوریتم Apriori را ارائه نموده و الگوهای همبستگی و وابستگی را به‌صورت نزولی بر اساس مقدار معیار اعتماد فهرست می‌کند. با لحاظ تعداد حوادث، این قواعد بیشترین سازگاری ویژگی‌ها را با مکان وقوع حادثه جمع‌آوری نموده‌اند. برای مثال، در قاعده ۱۱ معیار اعتماد بالای 0.83 ٪ محتمل است که یک اپراتور تولید بدنه با $5-0$ سال سابقه در سالن بدنه سازی در اثر برخورد با اجسام تیز و برنده بین ساعات ۴-۸ شیفت خود دچار آسیب گردد. به‌طور مشابه، مطابق قاعده

۲، اپراتور تولید بدنه با $5-0$ سال سابقه که بین ساعات ۴-۸ شیفت در سالن بدنه سازی دچار حادثه می‌شود، در 82 ٪ موارد ۴-۲۸ روز از زمان کاری خود را به‌دلیل حادثه از دست می‌دهد. بنابراین از طریق الگوهای همبستگی و وابستگی تولیدشده می‌توان تحلیل‌های عمیق و اطلاعات مهمی برای بهره‌بردار مجموعه استخراج نمود. در قواعد استخراج‌شده به‌وسیله روش Apriori مشاهده شد که پرتکرارترین ویژگی‌ها عبارتند از: سابقه ($5-0$)، ساعات کاری داخل شیفت (۴-۸) و موقعیت شغلی (اپراتور تولید بدنه). در تمام قواعد، سالن بدنه سازی به‌عنوان مکان وقوع حادثه لحاظ شده است. توزیع ویژگی‌ها و زیررده‌ها در ۲۰ الگو همبستگی در جدول ۵ نشان داده شده است. در جدول ۵ فراوانی ویژگی‌های سابقه کاری، ساعت کاری داخل شیفت و پست شغلی که بیشترین تطابق را با مکان وقوع حادثه دارند به‌ترتیب برابر با 65 ٪، 70 ٪ و 75 ٪ است. فراوانی سایر ویژگی‌ها در قواعد بین ۰ تا 35 ٪ متغیر است. زیررده‌های مرتبط با پرتکرارترین ویژگی‌ها عبارت‌اند از: $5-0$ سال سابقه، ۴-۸ ساعت کار در شیفت و پست شغلی اپراتور تولید بدنه. نتیجه‌گیری شد که کارگران با سابقه $5-0$ سال با احتمال بیشتری در معرض حوادث قرار دارند؛ حوادث عمدتاً بین ساعات ۴ تا ۸ رخ می‌دهند که ساعات پایانی شیفت محسوب می‌شوند و اپراتور سالن‌های بدنه سازی بیشترین فراوانی حوادث را تجربه می‌کنند.

با استفاده از الگوریتم Predictive Apriori، قواعدی با دقت پیش‌بینی بالاتر و بر اساس معیار دقت پیش‌بینی تولید می‌شوند. از آن‌جا که این الگوریتم همه قواعد را اسکن نمی‌کند، نتایج سریع‌تر و قواعد قابل‌اعتمادتری تولید می‌نماید (۴۳). جدول ۶ بیست قاعده برتر تولیدشده توسط الگوریتم Predictive Apriori را نشان می‌دهد. برای نمونه، در قاعده ۱ مشاهده می‌شود که در صورت وقوع برخورد با اجسام تیز و برنده بین ساعات ۴ و ۸ شیفت در سالن بدنه سازی و تحت شرایطی که تولید سالیانه برابر با $24,000-25,999$ عدد خودرو و حقوق سالیانه برابر با $6,999-6,999$ باشد، در 98 ٪ موارد سابقه کارگر مصدوم

جدول ۴: الگوهای وابستگی و همبستگی تولید شده برای محل وقوع حوادث با استفاده از الگوریتم Apriori

الگو	الگوهای وابستگی و همبستگی			اطمینان
	ویژگی های الگو	تعداد حوادث مربوط به ویژگی های الگو	هدف ویژگی	تعداد حوادث مربوط به ویژگی های الگو
۱	$WH = E, WH = WP, AC =$ تولید بدنه, بریدگی	۲۸۳	$AL =$ سالن بدنه سازی	۲۳۴
۲	$WH = E, WH = WP, AC =$ تولید بدنه	۳۳۲	$AL =$ سالن بدنه سازی	۲۷۱
۳	$WH = E, WH = WP, AC =$ تولید بدنه	۳۳۹	$AL =$ سالن بدنه سازی	۲۷۶
۴	$WH = E, WH = WP, AC =$ تولید بدنه	۳۴۵	$AL =$ سالن بدنه سازی	۲۷۸
۵	$WH = E, WH = WP, AC =$ تولید بدنه	۲۷۵	$AL =$ سالن بدنه سازی	۲۲۱
۶	$WH = E, WH = WP, AC =$ تولید بدنه	۲۷۱	$AL =$ سالن بدنه سازی	۲۱۷
۷	$WH = E, WH = WP, AC =$ تولید بدنه	۵۷۷	$AL =$ سالن بدنه سازی	۴۶۲
۸	$WH = E, WH = WP, AC =$ تولید بدنه	۲۹۱	$AL =$ سالن بدنه سازی	۲۳۳
۹	$WH = E, WH = WP, AC =$ تولید بدنه	۳۰۰	$AL =$ سالن بدنه سازی	۲۴۰
۱۰	$WH = E, WH = WP, AC =$ تولید بدنه	۴۵۸	$AL =$ سالن بدنه سازی	۳۶۵
۱۱	$WH = E, WH = WP, AC =$ تولید بدنه	۳۱۴	$AL =$ سالن بدنه سازی	۲۴۸
۱۲	$WH = E, WH = WP, AC =$ تولید بدنه	۲۸۹	$AL =$ سالن بدنه سازی	۲۲۷
۱۳	$WH = E, WH = WP, AC =$ تولید بدنه	۲۸۸	$AL =$ سالن بدنه سازی	۲۲۶
۱۴	$WH = E, WH = WP, AC =$ تولید بدنه	۴۱۷	$AL =$ سالن بدنه سازی	۳۲۳
۱۵	$WH = E, WH = WP, AC =$ تولید بدنه	۳۴۸	$AL =$ سالن بدنه سازی	۲۶۸
۱۶	$WH = E, WH = WP, AC =$ تولید بدنه	۳۷۲	$AL =$ سالن بدنه سازی	۲۸۶
۱۷	$WH = E, WH = WP, AC =$ تولید بدنه	۳۷۸	$AL =$ سالن بدنه سازی	۲۹۰
۱۸	$WH = E, WH = WP, AC =$ تولید بدنه	۶۸۶	$AL =$ سالن بدنه سازی	۵۲۵
۱۹	$WH = E, WH = WP, AC =$ تولید بدنه	۳۳۲	$AL =$ سالن بدنه سازی	۲۵۴
۲۰	$WH = E, WH = WP, AC =$ تولید بدنه	۲۸۵	$AL =$ سالن بدنه سازی	۲۱۸

A: سن؛ AC: علت حادثه؛ AL: محل حادثه؛ E: سابقه کار؛ LWD: روزهای کاری از دست رفته؛ MW: دستمزد سالیانه؛ WH: ساعات کاری؛ WP: موقعیت شغلی.

جدول ۵: توزیع ویژگی‌ها و زیرکلاس‌ها در الگوهای همبستگی و وابستگی تولیدشده با استفاده از الگوریتم Apriori

ویژگی	تعداد الگو یافت شده	فراوانی	زیرکلاس (تعداد الگو یافت شده)
WP	۱۵	۷۵	کارگر تولید (۱۵)
WH	۱۴	۷۰	۴-۸ (۱۴)
E	۱۳	۶۵	۰-۵ (۱۳)
AC	۷	۳۵	کارشناس مهندسی (۷)
LWD	۶	۳۰	۴-۲۸ (۶)
A	۴	۲۰	۲۶-۳۰ و ۳۱-۳۵ (۱)
ES	۳	۱۵	دیپلم (۳)
MW	۲	۵	۵۰۰۰-۵۹۹۹ (۲)
S	۱	-	۰-۸ (۱)
AO	-	-	
M	-	-	
D	-	-	
MP	-	-	

A: سن؛ AC: علت حادثه؛ AO: عضو آسیب دیده؛ D: روز؛ E: سابقه کار؛ ES: وضعیت تحصیلات
LWD: روزهای کاری از دست رفته؛ M: ماه؛ MP: تولید سالیانه؛ MW: دستمزد سالیانه؛ S: مرحله
WH: ساعات کاری؛ WP: موقعیت شغلی.

Apriori نشان می‌دهد که: کارگران با ۵-۰ سال سابقه با احتمال بیشتری دچار حادثه می‌شوند؛ شیفت‌های ۱۶-۲۴ از اهمیت ویژه‌ای در وقوع حوادث برخوردارند؛ حوادث عموماً بین ساعات ۴-۸ (ساعات پایانی شیفت‌ها) رخ می‌دهند؛ اپراتور سالن‌های بدنه سازی بیشترین برخورد با حوادث را دارند و برخورد با اجسام تیز و برنده به‌عنوان علت برجسته حوادث مشاهده می‌شوند. علاوه بر این، شایان توجه است که تولید سالیانه و حقوق سالیانه نیز در میان علل ریشه‌ای حوادث مطرح شده‌اند.

بحث

در این پژوهش، به منظور شناسایی الگوهای پنهان و علل ریشه‌ای حوادث شغلی در یک شرکت تولیدی خودروساز ایرانی، از ترکیب تکنیک‌های داده‌کاوی در دو فاز طبقه‌بندی و کاوش قواعد همبستگی استفاده شد. ارزیابی مدل‌های پیش‌بینی نشان داد که الگوریتم ماشین بردار پشتیبان (SVM) با دقت ۶۶ درصد در مرحله اعتبارسنجی متقاطع، عملکرد بهینه‌ای در پیش‌بینی حوادث دارد. در بخش کاوش قواعد، خروجی

در بازه ۵-۰ سال قرار دارد. به‌طور مشابه، مطابق قاعده ۲، هنگامی که تولید در بازه ۲۰,۰۰۰-۲۱,۹۹۹ در ماه تیر تحقق یابد، احتمال این که کارگری با ۵-۰ سال سابقه در یک حادثه سالن بدنه سازی مصدوم شود برابر با ۹۷,۹٪ است. بنابراین، استنتاج‌های دقیق‌تری را می‌توان با روش Predictive Apriori به‌دست آورد.

در الگوهای همبستگی و وابستگی تولیدشده برای محل وقوع حادثه به‌وسیله Predictive Apriori، توزیع ویژگی‌ها یکنواخت‌تر است؛ ویژگی‌های سابقه، شیفت، ساعات کاری داخل شیفت، پست شغلی و علت حادثه در اغلب قواعد شایع‌ترند. فراوانی ویژگی‌های سابقه، شیفت، ساعات کاری داخل شیفت، پست شغلی و علت حادثه که بیشترین همخوانی را با محل وقوع حادثه دارند به‌ترتیب برابر با ۴۵٪، ۵۰٪، ۵۵٪، ۵۵٪ و ۵۰٪ است (جدول ۷). درصد سایر ویژگی‌ها در قواعد بین ۰ تا ۳۵٪ متغیر است. با تحلیل زیررده‌های مرتبط با پرتکرارترین ویژگی‌ها روشن شد که زیررده‌های ۵-۰ سال سابقه، شیفت ۱۶-۲۴، ساعات کاری داخل شیفت ۴-۸، پست شغلی اپراتور تولید بدنه و علت حادثه برخورد با اجسام تیز و برنده برجسته‌اند. تحلیل مبتنی بر روش Predictive

جدول ۶: الگوهای وابستگی برای محل های وقوع حوادث به کمک الگوریتم Predictive Apriori

شماره قانون	متغیرهای تشکیل دهنده الگو	تعداد حوادث برای صفات قانون	هدف ویژگی	تعداد حوادث برای هدف	دقت
۱	$E = ۵-۰, WH = ۸-۴, MP = ۲۵۹۹۹-۲۴۰۰۰, MW = ۶۹۹۹-۶۰۰۰, AC =$ بریدگی	۲۳	سالن بدنه سازی	۲۳	۰.۹۸۰۴
۲	$M =$ تیر, $E = ۵-۰, MP = ۲۱۹۹۹-۲۰۰۰$	۲۲	سالن بدنه سازی	۲۲	۰.۹۷۹۴
۳	اپراتور تولید $WH = ۸-۴, WP =$ اسفند $M =$ استفاده نامناسب از ابزار یا ماشین آلات $AC =$ بدنه	۲۲	سالن بدنه سازی	۲۲	۰.۹۷۹۴
۴	$AO =$ انگشت دست $S = ۱۶-۸, WH = ۸-۴, WP =$ بریدگی $AC =$ اپراتور تولید بدنه	۲۲	سالن بدنه سازی	۲۲	۰.۹۷۹۴
۵	$E = ۵-۰, S = ۲۴-۱۶, WH = ۸-۴, MP = ۲۴-۲۵۹۹۹, AC =$ بریدگی	۲۲	سالن بدنه سازی	۲۲	۰.۹۷۹۴
۶	$D =$ دوشنبه $E = ۵-۰, S = ۲۴-۱۶, WH = ۸-۴, MW = ۵۹۹۹-۵۰۰۰, WP =$ اپراتور تولید بدنه	۲۲	سالن بدنه سازی	۲۲	۰.۹۷۹۴
۷	$AO = ۲۴-۱۶, S =$ چهارشنبه $D =$ انگشت دست $WP =$ اپراتور تولید بدنه	۲۱	سالن بدنه سازی	۲۱	۰.۹۷۸۲
۸	اپراتور تولید بدنه $WP = ۳۰-۲۶, A =$ اسفند $M =$ بریدگی $AC =$	۲۱	سالن بدنه سازی	۲۱	۰.۹۷۸۲
۹	$D =$ سه شنبه $S = ۲۴-۱۶, LWD = ۲۸-۴, WP =$ بریدگی $AC =$ اپراتور تولید بدنه	۲۱	سالن بدنه سازی	۲۱	۰.۹۷۸۲
۱۰	$AO =$ دست $E = ۵-۰, WH = ۸-۴, WP =$ اپراتور تولید بدنه	۲۰	سالن بدنه سازی	۲۰	۰.۹۷۶۸
۱۱	اپراتور تولید $WH = ۸-۴, WP =$ دوشنبه $D =$ برخورد با اجسام $AC =$ بدنه	۲۰	سالن بدنه سازی	۲۰	۰.۹۷۶۸
۱۲	$D =$ شنبه $E = ۵-۰, S = ۸-۰, WH = ۸-۴$	۲۰	سالن بدنه سازی	۲۰	۰.۹۷۶۸
۱۳	$D =$ پنجشنبه $A = ۳۰-۲۶, S = ۱۶-۸, AC =$ بریدگی	۲۰	سالن بدنه سازی	۲۰	۰.۹۷۶۸
۱۴	$M =$ فروردین $A = ۳۵-۳۱, E = ۵-۰, MW = ۵۹۹۹-۵۰۰۰$	۱۹	سالن بدنه سازی	۱۹	۰.۹۷۵۴
۱۵	$M =$ فروردین $A = ۳۵-۳۱, S = ۲۴-۱۶, WP =$ اپراتور تولید بدنه	۱۹	سالن بدنه سازی	۱۹	۰.۹۷۵۴
۱۶	$AO =$ انگشت پا $A = ۳۵-۳۱, WH = ۸-۴, MP = ۲۱۹۹۹-۲۰۰۰$	۱۸	سالن بدنه سازی	۱۸	۰.۹۷۳۶
۱۷	$AO =$ انگشت دست $E = ۱۰-۶, S = ۲۴-۱۶, AC =$ بریدگی	۱۸	سالن بدنه سازی	۱۸	۰.۹۷۳۶
۱۸	اپراتور $WH = ۸-۴, WP =$ انگشت دست $AO =$ برخورد با اجسام $AC =$ تولید بدنه	۱۸	سالن بدنه سازی	۱۸	۰.۹۷۳۶
۱۹	$M =$ فروردین $S = ۸-۰, MW = ۵۹۹۹-۵۰۰۰, WP =$ اپراتور تولید بدنه	۱۸	سالن بدنه سازی	۱۸	۰.۹۷۳۶
۲۰	$M =$ تیر $A = ۳۵-۳۱, E = ۵-۰, WH = ۸-۴$	۱۸	سالن بدنه سازی	۱۸	۰.۹۷۳۶

A: سن؛ AC: علت حادثه؛ AO: عضو آسیب دیده؛ D: روز؛ E: سابقه کار؛ ES: وضعیت تحصیلات؛ LWD: روزهای کاری از دست رفته؛ M: ماه؛ MP: تولید سالیانه؛ MW: دستمزد سالیانه؛ S: مرحله؛ WH: ساعات کاری؛ WP: موقعیت شغلی.

تولید در سالن بدنه سازی». همچنین، ظهور متغیرهایی نظیر میزان تولید و حقوق سالیانه در قواعد مستخرج، نشان دهنده تأثیر مستقیم فشارهای تولیدی و اقتصادی بر رخداد حوادث است. یکی از کلیدی ترین نتایج این مطالعه، آسیب پذیری

الگوریتم های Apriori و Predictive Apriori حاکی از آن بود که وقوع حوادث بیشترین ارتباط را با سه عامل کلیدی دارد: «سابقه کاری پایین کارکنان (۰ تا ۵ سال)»، «ساعات پایانی شیفت کاری (بازه زمانی ۴ تا ۸ ساعت پس از شروع شیفت)» و «پست شغلی اپراتور

جدول ۷: توزیع صفات و زیرطبقه‌ها در الگوهای همبستگی و وابستگی با Predictive Apriori

ویژگی	الگو	فراوانی	زیرکلاس (تعداد قواعد یافت‌شده)
WH	۱۱	۵۵	۸-۴ (۱۱)
WP	۱۱	۵۵	اپراتور تولید بدنه (۱۱)
S	۱۰	۵۰	۸-۰ (۲)، ۱۶-۸ (۲)، ۲۴-۱۶ (۶)
AC	۱۰	۵۰	بریدگی (۷)، استفاده نامناسب از ابزار یا ماشین‌آلات (۱)، برخورد با اجسام (۲)
E	۹	۴۵	۵-۰ (۸)، ۱۰-۶ (۱)
M	۷	۳۵	اسفند (۲)، فروردین (۳)، تیر (۲)
AO	۶	۳۰	انگشت دست (۴)، دست (۱)، انگشت پا (۱)
D	۶	۳۰	دوشنبه (۲)، سه شنبه (۱)، چهارشنبه (۱)، پنجشنبه (۱)، شنبه (۱)
A	۶	۳۰	۳۰-۲۶ (۲)، ۳۵-۳۱ (۴)
MP	۴	۲۰	۲۰۰۰-۲۱۹۹۹ (۲)، ۲۴۰۰۰-۲۵۹۹۹ (۲)
MW	۴	۲۰	۵۰۰۰-۵۹۹۹ (۳)، ۶۰۰۰-۶۹۹۹ (۱)
LWD	۱	۵	۴-۲۸ (۱)
ES	-	-	-

A: سن؛ AC: علت حادثه؛ AO: عضو آسیب‌دیده؛ D: روز؛ E: سابقه کار؛ ES: وضعیت تحصیلات؛ LWD: روزهای کاری از دست‌رفته؛ M: ماه؛ MP: تولید سالیانه؛ MW: دستمزد سالیانه؛ S: مرحله؛ WH: ساعات کاری؛ WP: موقعیت شغلی.

کار، افت طبیعی تمرکز و توان بدنی کارگر با فشار برای رسیدن به اهداف تولید روزانه تلاقی پیدا می‌کند. این شرایط، به‌خصوص در محیط‌های پیچیده و پرمحرک مانند سالن بدنه که نیازمند هوشیاری دائمی است، احتمال خطای انسانی و اتخاذ تصمیمات پرخطر (مانند نادیده گرفتن یک مرحله ایمنی برای افزایش سرعت) را به شدت افزایش می‌دهد. بنابراین، به نظر می‌رسد طراحی نامناسب ایستگاه کاری یا نقص در آموزش به‌تنهایی عامل اصلی نیست، بلکه این تعامل میان خستگی و فشار تولید است که شرایط را برای وقوع حادثه مهیا می‌سازد.

تفسیر این یافته‌ها نشان می‌دهد که نیروی کار با سابقه صفر تا پنج سال، به دلیل عدم برخورداری از تجربه عملی کافی و آشنایی کمتر با مخاطرات محیط‌های صنعتی، بیشترین پتانسیل آسیب‌پذیری را داراست. این موضوع ضرورت تمرکز بر آموزش‌های بدو استخدام را برجسته می‌سازد. در مقایسه با پیشینه تحقیق، این یافته با مطالعه Sanmiquel (۲۳) که داده‌کاوی را ابزاری ارزشمند برای شناسایی علل ریشه‌ای حوادث در صنعت خودروسازی اسپانیا معرفی کرده بود، همخوانی کامل دارد و نشان می‌دهد که در صنعت خودروسازی (چه در

بالای کارگران با سابقه کاری کمتر از پنج سال است. این یافته همسو با بسیاری از مطالعات پیشین (۲۸) است که کم‌تجربگی را به‌عنوان یک عامل خطر مهم معرفی کرده‌اند. با این حال، تحلیل عمیق‌تر این موضوع صرفاً به «کمبود تجربه» محدود نمی‌شود. همان‌طور که در داده‌های ما نیز مشهود است، این گروه از کارگران به دلیل آشنایی کمتر با مخاطرات پنهان و پیچیده‌ی محیط صنعتی و همچنین عدم تسلط کافی بر رویه‌های ایمنی در شرایط فشار تولید، بیشتر در معرض خطر قرار می‌گیرند. این امر نشان می‌دهد که برنامه‌های آموزشی بدو استخدام اگرچه ضروری هستند، اما کافی نیستند و لازم است سیستم‌های مربیگری (Mentorship) و نظارت مستمر در سال‌های اولیه خدمت برای این گروه از کارکنان طراحی و اجرا گردد تا شکاف بین دانش تئوریک ایمنی و تجربه عملی پر شود.

یافته دیگر این پژوهش، تمرکز بالای حوادث در بازه زمانی ۴ تا ۸ ساعت پس از شروع شیفت کاری، به‌ویژه در سالن بدنه، است. این نتیجه صرفاً یک همبستگی آماری نیست، بلکه به یک عامل ریشه‌ای مهم اشاره دارد: خستگی تجمعی ناشی از فشار تولید. در ساعات پایانی

بود. همچنین، در پژوهش Aliabadi (۲۷) بر نقش عوامل انسانی و سازمانی در حوادث تأکید شده و در مطالعه Sanmiquel (۳۰) نیز داده کاوی در تحلیل حوادث شغلی، از جمله در صنعت خودروسازی اسپانیا، ابزاری ارزشمند برای شناسایی علل ریشه‌ای معرفی شده است. در همین راستا، مطالعه حاضر با افزودن داده‌های تولید سالانه و بهره‌وری به داده‌های کلاسیک حوادث، دامنه تحلیل را از سطح متغیرهای صرفاً حادثه‌محور فراتر برده و پیوند میان فشارهای تولیدی و بروز حادثه را آشکار کرده است.

نتایج این پژوهش بر این اصل موجود در ادبیات تحقیق صحه می‌گذارد که ادغام داده‌های عملیاتی و تولیدی با داده‌های حوادث، برای درک عمیق‌تر ریسک‌ها الزامی است. بر پایه این یافته‌ها، طراحی مداخلات پیشگیرانه هدفمند از جمله بازنگری در شیفت‌بندی‌ها جهت کاهش خستگی مفرد، اولویت‌بندی بازرسی‌های ایمنی در سالن بدنه‌سازی، و پیاده‌سازی برنامه‌های آموزشی و نظارتی ویژه برای اپراتورهای تازه‌کار (با سابقه زیر ۵ سال) در صنایع خودروسازی اکیداً پیشنهاد می‌گردد.

تفسیر یافته‌ها باید با تأمل درباره محدودیت‌های روش‌شناختی همراه باشد. نخست اینکه دقت صددرصد در مجموعه آموزش می‌تواند ناشی از ساختار داده‌ها یا توزیع ویژگی‌ها باشد و هرچند حذف مقادیر پرت گزارش شده است، نیاز به آزمون‌های بیشتر برای تضمین عدم وجود الگوهای مخفی یادگیرنده بیش‌ازحد باقی است. دوم اینکه استفاده محدود از تنظیمات پیش‌فرض الگوریتم‌ها مانع از بررسی کامل فضای پارامتری و یافتن پیکربندی‌های مقاوم‌تر شده است. سوم، داده‌های مورد استفاده به ماهیت و دامنه یک یا چند واحد تولیدی مشخص وابسته‌اند؛ از این رو تعمیم نتایج به سایر کارخانه‌ها یا صنایع نیازمند اعتبارسنجی بیرونی است. نهایتاً، متغیرهایی چون خستگی تجمعی، شبه حوادث و جزئیات زمان‌بندی دقیق (وابستگی‌های زمانی و سلسله‌مراتبی) در این پایگاه کمتر مورد توجه قرار گرفته‌اند و افزودن این ابعاد می‌تواند دقت و کارکردهای پیشگیرانه تحلیل‌ها را به‌طور قابل‌توجهی افزایش دهد.

ایران و چه در سایر کشورها، عوامل انسانی و کمبود تجربه از محرک‌های اصلی بروز حوادث به شمار می‌روند. از سوی دیگر، رخداد مکرر حوادث در ساعات ۴ تا ۸ داخلی شیفت، بیانگر آن است که افت تمرکز و بروز خستگی در ساعات پایانی کار، در کنار فشارهای ناشی از حجم تولید، ضریب خطای اپراتورها را به شدت افزایش می‌دهد. در مطالعات پیشین، محققانی نظیر Mevsim (۲۱) با استفاده از تحلیل درخت خطا نشان دادند که نقص تجهیزات (نظیر تهویه معیوب) از علل اصلی حوادث است؛ با این حال، نتایج مطالعه حاضر ثابت می‌کند که در سالن‌های پیچیده‌ای نظیر بدنه‌سازی، فراتر از مشکلات تجهیزاتی، این «خستگی نیروی کار» و «فشارهای تولیدی» است که نقش کلیدی‌تری در بروز حوادث ایفا می‌کند. این نتایج همچنین با یافته‌های Erdogan (۲۲) که با روش تحلیل واریانس (ANOVA) شدت حوادث را بر اساس جراحات ارزیابی کرده بود، از منظر لزوم پایش مستمر ایستگاه‌های پرخطر مطابقت دارد.

کاربست موفق الگوریتم‌های کاوش قواعد در این محیط صنعتی، هم‌راستا با دستاوردهای Shuangyue (۲۰) است که با استفاده از الگوریتم Apriori در صنعت معدن، راهکارهایی اثربخش برای پیشگیری از حوادث ارائه دادند. عملکرد برتر این تکنیک‌های نوین در قیاس با روش‌های سنتی، همچنین همسو با ادعای Ak (۲۴) پیرامون مزیت سیستم‌های استنتاج عصبی-فازی و تأکیدات Dessureault (۱۹) بر ضرورت استفاده از داده‌کاوی در به‌روزرسانی پایگاه‌های ایمنی است. با این وجود، کسب دقت یادگیری ۱۰۰ درصدی در فاز آموزش برخی از مدل‌های این پژوهش، لزوم احتیاط نسبت به پدیده بیش‌برازش (Overfitting) و بررسی دقیق تعمیم‌پذیری الگوها را ایجاب می‌نماید.

مقایسه نتایج این پژوهش با مطالعات پیشین نشان می‌دهد که یافته‌های حاضر با جهت‌گیری کلی ادبیات موضوع همسو است. برای نمونه، در مطالعه Yedla (۲۸) نیز زمان شیفت و تجربه کارکنان به‌عنوان پیش‌بین‌های پیامدهای ایمنی و روزهای از دست‌رفته از کار مطرح شده

بحث شد، می‌تواند منجر به شتاب‌زدگی و کاهش توجه به ایمنی شود. از سوی دیگر، ارتباط دستمزد با حوادث می‌تواند از طریق نظریه‌های مدیریتی و اقتصاد کار تبیین شود. دستمزد پایین یا نامتناسب با سختی کار، می‌تواند منجر به کاهش انگیزه، بی‌توجهی به دستورالعمل‌ها و فرسودگی شغلی شود. کارگری که از نظر مالی تأمین نیست، ممکن است تمرکز ذهنی کافی برای رعایت دقیق اصول ایمنی نداشته باشد یا برای افزایش درآمد، به انجام کار اضافی و پذیرش ریسک‌های بیشتر تن دهد. این فرضیه‌ها نشان می‌دهد که سیستم‌های جبران خدمت و تعیین اهداف تولیدی در یک سازمان، نمی‌توانند مستقل از استراتژی‌های مدیریت ایمنی آن باشند و باید به‌صورت یکپارچه دیده شوند. هرچند حضور متغیر تولید سالیانه در قواعد استخراج‌شده بیانگر آن است که شاخص‌های تولید می‌توانند با الگوهای وقوع حوادث همراه باشند، اما این مطالعه با ماهیت داده‌کاوی توصیفی طراحی شده و هدف آن کمی‌سازی شدت این رابطه یا بررسی شکل تابعی آن نبوده است. بنابراین، نتایج حاضر صرفاً وجود هم‌رخدادی (Association) میان برخی سطوح تولید و الگوهای حوادث را نشان می‌دهد و نباید به‌عنوان شواهدی از رابطه علی یا خطی تفسیر شود. پیشنهاد می‌شود در مطالعات آینده با استفاده از مدل‌های رگرسیونی، تحلیل حساسیت و داده‌های طولی، میزان و نوع اثر شاخص‌های تولید بر احتمال وقوع حوادث به‌صورت کمی ارزیابی شود. پیامدهای کاربردی این نتایج برای مدیریت ایمنی محیط کار روشن و مستقیم است: قواعد استخراج‌شده امکان هدف‌گیری گروه‌های پرخطر (نظیر کارگران با سابقه کم)، بازه‌های زمانی حساس (ساعات پایانی شیفت یا شیفت‌های خاص مانند ۱۶-۲۴)، و بخش‌های محیطی مشخص (سالن بدنه‌سازی) را برای مداخلات پیشگیرانه فراهم می‌آورند. علاوه بر این، ظهور متغیرهای اقتصادی-تولیدی (میزان تولید سالیانه، بازه حقوق) به عنوان عوامل همبستگی‌دار با وقوع حوادث حاکی از ضرورت توجه مدیریتی فراتر از راهکارهای صرفاً فنی- از قبیل تعدیل برنامه تولید، بازنگری در بارکاری، و سیاست‌های

از مهم‌ترین محدودیت‌های این پژوهش، دستیابی الگوریتم‌های k-NN و Random Forest به دقت ۱۰۰ درصد در مجموعه آموزش است. اگرچه بررسی داده‌ها با استفاده از فیلتر IQR نشان داد که مجموعه داده فاقد مقادیر پرت آماری معنادار است، با این حال عدم وجود نقاط پرت لزوماً به معنای نبود بیش‌برازش در مدل‌های یادگیری ماشین نیست. در واقع، فیلتر IQR تنها قادر به شناسایی مقادیر افراطی از منظر آماری بوده و نمی‌تواند وجود الگوهای تصادفی، وابستگی‌های خاص مجموعه داده یا نویزهای سیستماتیک را که ممکن است موجب یادگیری بیش از حد مدل شوند، رد کند. از سوی دیگر، اختلاف قابل توجه میان عملکرد این الگوریتم‌ها در مجموعه آموزش و نتایج حاصل از اعتبارسنجی متقاطع، احتمال محدود بودن قابلیت تعمیم مدل‌ها به داده‌های جدید را مطرح می‌سازد. بنابراین، دقت‌های بسیار بالا در مجموعه آموزش باید با احتیاط تفسیر شوند و به عنوان نشانه‌ای قطعی از توان پیش‌بینی مدل تلقی نگردند. در مطالعات آتی، استفاده از روش‌های اعتبارسنجی سخت‌گیرانه‌تر، مجموعه‌های داده مستقل برای آزمون خارجی، تنظیم نظام‌مند فرآیندها و تحلیل حساسیت مدل‌ها می‌تواند به ارزیابی دقیق‌تر پایداری و قابلیت تعمیم نتایج کمک نماید. همچنین ارزیابی مدل‌ها صرفاً بر پایه مجموعه آموزش و اعتبارسنجی متقاطع در محیط WEKA انجام شد و داده‌های مستقل بیرونی برای آزمون تعمیم‌پذیری مدل‌ها در دسترس نبود. بنابراین، هرچند نتایج به‌دست‌آمده توانایی الگوریتم‌ها را در شناسایی الگوهای موجود در داده‌های این مطالعه نشان می‌دهد، اما برای اطمینان بیشتر از قابلیت تعمیم آن‌ها به داده‌های جدید، استفاده از مجموعه داده مستقل یا به‌کارگیری روش hold-out با تفکیک زمانی در پژوهش‌های آتی ضروری است.

نوآوری این پژوهش در شناسایی ارتباط میان حجم تولید و دستمزد با میزان حوادث بود. هرچند مقاله صرفاً این ارتباط را گزارش کرده است، اما تحلیل عمیق‌تر آن ضروری است. فشار برای افزایش تولید، همان‌طور که

انگیزشی/حقوقی- برای کاهش ریسک است. در سطح فنی، ترکیب روش‌های داده‌کاوی الگوها با مدل‌های پیش‌بینی (مانند شبکه‌های بیزی، ANFIS، یا روش‌های ترکیبی یادگیری ماشین) می‌تواند چارچوب تصمیم‌گیری مقتدرانه‌تری برای برنامه‌ریزی بازرسی‌ها، آموزش‌های هدفمند و تخصیص منابع حفاظتی ایجاد نماید.

نتیجه‌گیری

یافته‌های این پژوهش نشان داد که به‌کارگیری روش‌های داده‌کاوی، به‌ویژه ترکیب طبقه‌بندی و کاوش الگوهای همبستگی، می‌تواند برای شناسایی الگوهای پنهان و عوامل مؤثر بر وقوع حوادث شغلی در محیط‌های تولیدی بسیار کارآمد باشد. هم‌سویی نتایج حاصل از دو رویکرد تحلیلی نشان می‌دهد که بروز حادثه در این شرکت صرفاً به یک عامل منفرد وابسته نیست، بلکه از برهم‌کنش هم‌زمان ویژگی‌های انسانی، شغلی، زمانی و تا حدی سازمانی ناشی می‌شود. در مرحله طبقه‌بندی، الگوریتم‌های kNN و RF در مجموعه آموزش به دقت ۱۰۰ درصد دست یافتند، اما افت دقت در اعتبارسنجی متقاطع و عملکرد بهتر SVM با دقت ۶۶٫۰۸ درصد بیانگر آن است که بخشی از نتایج مدل‌های طبقه‌بندی می‌تواند تحت تأثیر بیش‌برازش قرار گرفته باشد. بنابراین، هرچند این مدل‌ها توانایی بالایی در بازنمایی داده‌های موجود داشته‌اند، تعمیم‌پذیری آن‌ها به داده‌های جدید باید با احتیاط تفسیر شود. در مرحله کاوش قواعد نیز الگوریتم‌های Apriori و به‌ویژه Apriori Predictive توانستند الگوهایی با دقت و اطمینان بالا استخراج کنند که در آن‌ها سه مؤلفه اصلی بیش از سایر عوامل برجسته بودند: سابقه کار پایین (۰ تا ۵ سال)، ساعات پایانی شیفت (۴ تا ۸ ساعت پس از شروع کار) و اشتغال در پست اپراتور تولید بدنه. افزون بر این، حضور متغیرهای تولید سالیانه و حقوق سالیانه در برخی قواعد نشان می‌دهد که فشارهای تولیدی و شرایط اقتصادی-سازمانی نیز در کنار عوامل فردی و شغلی در شکل‌گیری الگوی حوادث نقش دارند. از منظر کاربردی، نتایج این پژوهش نشان می‌دهد

که پیشگیری از حوادث باید به‌طور هدفمند بر گروه‌های پرخطر متمرکز شود. بر این اساس، آموزش‌های بدو استخدام و بازآموزی برای کارکنان با سابقه ۰ تا ۵ سال باید با تأکید بر شناسایی خطرات محیط کار، روش صحیح استفاده از ابزار و ماشین‌آلات، کار ایمن در مجاورت اجسام تیز و برنده، و گزارش‌دهی سریع شرایط ناایمن طراحی شود. همچنین، با توجه به تمرکز بیشتر حوادث در ساعات پایانی شیفت، بازنگری در شیفت‌بندی به‌نحوی که استراحت‌های کوتاه و مؤثر در ساعات میانی و پایانی شیفت پیش‌بینی شود، و نیز کاهش کار مداوم در بازه ۴ تا ۸ ساعت پس از شروع کار، می‌تواند در کاهش خستگی و خطاهای انسانی مؤثر باشد. از سوی دیگر، در سالن بدنه‌سازی باید مداخلات مهندسی و کنترلی با اولویت بیشتری اجرا شوند؛ از جمله بهبود حفاظ‌گذاری و ایمن‌سازی تجهیزات، سامان‌دهی بهتر مسیرهای کاری و جابجایی مواد، کاهش تماس مستقیم با اجسام تیز و برنده، ارتقای نظم و پاکیزگی محیط کار، و افزایش بازرسی‌های ایمنی در ایستگاه‌های پرخطر. این اقدامات، در کنار نظارت هدفمند بر اپراتورهای بدنه‌سازی، می‌تواند به کاهش احتمال وقوع حادثه کمک کند.

با وجود این، تفسیر نتایج باید با توجه به محدودیت‌های پژوهش انجام شود. نخست، مشاهده دقت بسیار بالا در مجموعه آموزش و افت آن در اعتبارسنجی متقاطع، احتمال بیش‌برازش را مطرح می‌کند و نشان می‌دهد که بخشی از الگوهای شناسایی‌شده ممکن است وابسته به ساختار همین داده‌ها باشند. دوم، تحلیل حاضر بر داده‌های مربوط به یک شرکت تولیدی مشخص استوار است و بنابراین تعمیم مستقیم آن به سایر محیط‌های صنعتی بدون اعتبارسنجی بیرونی، با اطمینان کامل امکان‌پذیر نیست. سوم، هرچند داده‌های پژوهش از متغیرهای متعددی تشکیل شده‌اند، اما برخی ابعاد مهم مانند خستگی تجمعی، شبه‌حوادث، و جزئیات دقیق‌تر ترتیب زمانی رویدادها به‌طور مستقیم در مدل وارد نشده‌اند؛ از این‌رو، امکان دارد بخشی از سازوکارهای واقعی بروز حادثه در این تحلیل بازنمایی نشده باشد.

باشند. در مجموع، این مطالعه نشان می‌دهد که داده‌کاوی می‌تواند با آشکارسازی روابط میان عوامل انسانی، زمانی، شغلی و تولیدی، مبنای مناسبی برای تصمیم‌گیری‌های پیشگیرانه در محیط‌های صنعتی فراهم کند؛ با این حال، بهره‌برداری مؤثر از این نتایج مستلزم توجه هم‌زمان به محدودیت‌های مدل‌ها و طراحی مداخلات عملی، مشخص و قابل اجرا است.

تضاد منافع

نویسندگان اظهار می‌دارند که هیچ‌گونه منافع مالی رقابتی یا روابط شخصی شناخته‌شده‌ای که بتواند بر کار گزارش‌شده در این مقاله تأثیر گذاشته باشد، ندارند.

REFERENCES

- Ormazdin M. Explanation and assessment of contextual factors for the implementation of knowledge management in the National Iranian Oil Products Distribution Company, Tehran Region (Dissertation). Tehran: Faculty of Management and Accounting, Shahid Beheshti University; 2007.
- Friedman, T. L. (2005). *The World is Flat: A Brief History of the Twenty-First Century*. Farrar, Straus and Giroux.
- Didla, S., Mearns, K., & Flin, R. (2009). "safety citizenship behaviour: a proactive approach to risk management". *Journal of risk Research*, 12(3-4), 475-483. <https://doi.org/10.1080/13669870903041433>.
- Rohani, J., Johari, M., Hamid, W., Atan, H., Adeyemi, A., & Udin, A. (2015). "occupational accident indirect cost model validation using confirmatory factor analysis". *Procedia Manufacturing*, 2, 291-295. <https://doi.org/10.1016/j.promfg.2015.07.051>.
- Sun, L., Paez, O., Lee, D., Salem, S., & Daraiseh, N. (2006). "estimating the uninsured costs of work-related accidents, part i: a systematic review". *Theoretical Issues in Ergonomics Science*, 7(3), 227-245. <https://doi.org/10.1080/14639220500090521>.
- Fordyce, T., Leonhard, M., Watson, H., Mezei, G., Vergara, X., & Krishen, L. (2016). "an analysis of fatal and non-fatal injuries and injury severity factors among electric power industry workers". *American Journal of Industrial Medicine*, 59(11), 948-958. <https://doi.org/10.1002/ajim.22621>.
- Misiurek, K. and Misiurek, B. (2017). "methodology of improving occupational safety in the construction industry on the basis of the twi program". *Safety Science*, 92, 225-231. <https://doi.org/10.1016/j.ssci.2016.10.017>.
- Panuwatwanich, K., Al-Haadir, S., & Stewart, R. A. (2016). "Influence of safety motivation and climate on safety behavior and outcomes: Evidence from the Saudi Arabian construction industry". *International Journal of Occupational Safety and Ergonomics*, 23(1), 60-75.
- Hinze, J., & Godfrey, R. (2003). An evaluation of safety performance measure for construction projects. *Journal of Construction Research*, 4(1), 5-15.
- Hrymak, V., & Perezgonzalez, J. (2007). The costs and effects of workplace accidents: 20 Case studies from Ireland the costs and effects of workplace accidents A report for the Health and Safety Authority by. A report for Health and Safety Authority.
- Bahrehbar M, Givchchi S. Examining the Effect of Hazard Prevention and Control on Safety Outcomes in the Automotive Industry: The Mediating Role of Safety Behaviors and the Moderating Role of Safety Leadership. *Risk Anal.* 2025 Sep;45(11):3643-3665. <https://doi.org/10.1111/risa.70095>.
- International Labour Organization. (n.d.). Nearly 3 million people die of work-related accidents and diseases. Retrieved May 29, 2026, from <https://www.ilo.org/resource/news/nearly-3-million-people-die-work->

- related-accidents-and-diseases.
13. Eurostat. (n.d.). Accidents at work statistics. Retrieved May 29, 2026, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics.
 14. Ministry of Cooperatives, Labour, and Social Welfare. (2022). Statistical yearbook of the Ministry of Cooperatives, Labour, and Social Welfare.
 15. U.S. Bureau of Labor Statistics. (2020). Motor vehicle manufacturing nonfatal injury and illness rate twice as high as private industry. <https://www.bls.gov/opub/ted/2020/motor-vehicle-manufacturing-nonfatal-injury-and-illness-rate-twice-as-high-as-private-industry.htm>.
 16. Altindis B. Investigation of occupational accidents in Amasra Hard Coal Enterprise) master's thesis. (Afyonkarahisar). Afyon Kocatepe University; 2023. 95 p (in Turkish). [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref1](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref1).
 17. Ismail SN, Ramli A, Aziz HA. Research trends in mining accidents study: asystematic literature review. *Saf Sci* 2021;143:105438. <https://doi.org/10.1016/j.ssci.2021.105438>.
 18. Niu Y, Fan Y, Ju X. Critical review on data-driven approaches for learning from accidents: comparative analysis and future research. *Saf Sci* 2024;171:106381. <https://doi.org/10.1016/j.ssci.2023.106381>.
 19. Dessureault S, Sinuhaji A, Coleman P. Data mining mine safety data. *Min Eng* 2007;59(8):64e70. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref4](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref4).
 20. Shuangyue L, Li P. Analysis of coal mine hidden danger correlation based on improved apriori algorithm. In: Fourth international conference on intelligent systems design and engineering applications (ISDEA) 2013. p. 112e6. 2013 Nov 6-7; Zhangjiajie, Hunan Province, China. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref5](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref5).
 21. Mevsim R. Risk assessment by fault tree analysis of methane explosions in Turkish Hard Coal Enterprises Underground Mines (master's thesis). (Ankara). Middle East Technical University; 2016. 125 p. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref6](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref6).
 22. Erdo_gan HH. A quantitative risk assessment methodology for occupational accidents in underground coal mines: a case of Turkish Hard Coal Enterprises (dissertation). (Ankara). Middle East Technical University; 2016. 278 p. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref7](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref7).
 23. Sanmiquel L, Bascompta M, Rossell JM, Anticoi HF, Guash E. Analysis of occupational accidents in underground and surface mining in Spain using data-mining techniques. *Int J Environ Res Public Health* 2018;15:462. <https://doi.org/10.3390/ijerph15030462>.
 24. Ak MF. A neuro-fuzzy-based multi-criteria risk evaluation approach: a case study of underground mining. In: Al-Turjman F, editor. Artificial intelligence in IoT, transactions on computational science and computational intelligence. Cham: Springer; 2019. p. 167e205. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref9](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref9).
 25. Gerassis S, Saavedra Á, Taboada J, Alonso E, Bastante FG. Differentiating between fatal and non-fatal mining accidents using artificial intelligence techniques. *Int J Min Reclam Environ* 2020;34(10):687e99. <https://doi.org/10.1080/17480930.2019.1700008>.
 26. Iphar M, Çukurluöz AK. Fuzzy risk assessment for mechanized underground coal mines in Turkey. *Int J Occup Saf Ergon* 2020;26(2):256e71. <https://doi.org/10.1080/10803548.2018.1426804>.
 27. Aliabadi MM, Aghaei H, Kalatpour O, Soltanian AR, Nikravesh A. Analysis of human and organizational factors that influence mining accidents based on Bayesian network. *Int J Occup Saf Ergon* 2020;26(4):670e7. <https://doi.org/10.1080/10803548.2018.1455411>.
 28. Yedla A, Kakhki FD, Jannesari A. Predictive modeling for occupational safety outcomes and days away from work analysis in mining operations. *Int J Environ Res Public Health* 2020;17:7054. <https://doi.org/10.3390/ijerph17197054>.
 29. Opokua M, Opoku SK. An adaptive neuro-fuzzy inference system for predicting survivability rate in underground mining accident. *Conf Proc ICAST 2020* 2021;6(1):250. 265. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref14](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref14).
 30. Sanmiquel L, Bascompta M, Rossell JM, Anticoi H. Analysis of occupational accidents in the Spanish mining sector in the period 2009e2018. *Int J Environ Res Public Health* 2021;18:13122. <https://doi.org/10.3390/ijerph182413122>.
 31. Li S, You M, Li D, Liu J. Identifying coal mine safety production risk factors by employing text mining

- and Bayesian network techniques. *Process Saf Environ* 2022;162:1067e81. <https://doi.org/10.1016/j.psep.2022.04.054>.
32. Li S, Li D, Xu S, You M. Applications of artificial intelligence for coal mine gas risk assessment. *Saf Sci* 2021;143:105420. <https://doi.org/10.1016/j.ssci.2021.105420>.
33. Bayraktar B, Uyguçgil H, Konuk A. Investigation of occupational accidents in mining with survival analysis. *Min Metall Explor* 2023;40:1827e38. <https://doi.org/10.1007/s42461-023-00810-5>.
34. Baraza X, Cugueró-Escofet N, Rodríguez-Elizalde R. Statistical analysis of the severity of occupational accidents in the mining sector. *J Saf Res* 2023;86:364e75. <https://doi.org/10.1016/j.jsr.2023.07.015>.
35. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: *Proceedings of the 20th international conference on very large data bases 1994*. p. 487e99. Santiago, Chile. [http://refhub.elsevier.com/S2093-7911\(24\)00069-6/sref20](http://refhub.elsevier.com/S2093-7911(24)00069-6/sref20).
36. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
37. John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338–345). Morgan Kaufmann.
38. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
39. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
40. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>.
41. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>.
42. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
43. Scheffer T. Finding association rules that trade support optimally against confidence. In: De Raedt L, Siebes A, editors. *Principles of data mining and knowledge discovery*. Berlin: Springer; 2001. p. 424e35. https://doi.org/10.1007/3-540-44794-6_35.