Safety and Health at Work xxx (xxxx) xxx

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

Data Mining Implementations for Determining Root Causes and Precautions of Occupational Accidents in Underground Hard Coal Mining

Bilal Altindis [1], Fatih Bayram [2,*]

[1] TTK Turkish Hard Coal Corporation, Turkiye
[2] Department of Mining Engineering, Afyon Kocatepe University, 03200, Afyonkarahisar, Turkiye

A R T I C L E I N F O

Article history:
Received 26 January 2024
Received in revised form
20 August 2024
Accepted 2 September 2024
Available online xxx

Keywords:
Association rules mining
Data mining
Hard coal production
Occupational accidents
Underground coal mining
WEKA

A B S T R A C T

Background: Nowadays, as in every branch of industry, a large amount of data can be collected in mining, both in productivity and occupational safety. It is increasingly essential to transform this data into useful information for enterprises. Data mining is very useful in processing and extracting useful information from the processed data. This study aims to analyze the data of occupational accidents with injuries between 2010 and 2021 in an underground hard coal mine by data mining.
Methods: The injured accident data for the relevant years were organized and analyzed using data mining algorithms. These algorithms were implemented with the WEKA data mining program, an open-source application.
Results: According to different test methods, k-Nearest Neighborhood and Support Vector Machine algorithms succeeded in classification and prediction. The k Nearest Neighborhood and Support Vector Machine algorithms achieved 100% (training set) and 66% (cross-validation) performance, respectively, according to two different test methods. One of the critical phases of the study is the determination of the attributes and subclasses that are effective in the origin of accidents by association rules mining. Thus, more detailed information was obtained about the root causes of the accidents. A result of Apriori and Predictive Apriori implementations revealed that the root causes of occupational accidents according to the accident locations are the worker experience, the working hours in the shift, and the worker position. In addition, shifts, accident causes, especially monthly production, and monthly wages were also influential.
Conclusions: These results are also in accordance with the actual situation in the enterprise. As a result of the research, practical suggestions were presented for evaluating occupational accidents and taking precautions.

1. Introduction

Underground resources play an essential role in the development of national economies. The production and processing of mines ensure the development of technology and the increase in the level of welfare. The need for mines is inevitable to make life easier and produce the necessary tools and equipment. Therefore, all industry sectors (e.g., chemistry, construction, machinery, electronics, informatics, defense industry) need mines. In addition to being an indispensable sector, mining is in the hazardous group regarding occupational health and safety among the industrial branches.

The activities carried out within the mining are highly hazardous and have a high risk of accidents and death. Injuries, amputations, and deaths can occur in accidents as a result of these high-risk activities. In tackling these risks, analyzing hazardous

Bilal Altindis: https://orcid.org/0000-0002-8347-8725; Fatih Bayram: https://orcid.org/0000-0002-8510-7936
* Corresponding author.
E-mail addresses: kalebent@gmail.com (B. Altindis), bayramfatih@aku.edu.tr (F. Bayram).

2093-7911/$ — see front matter
https://doi.org/10.1016/j.shaw.2024.09.003

Please cite this article as: Altindis B, Bayram F, , Data Mining Implementations for Determining Root Causes and Precautions of Occupational Accidents in Underground Hard Coal Mining, Safety and Health at Work, https://doi.org/10.1016/j.shaw.2024.09.003

conditions and quantitative measurement of risks are essential for prevention and reduction to acceptable levels. Risk measurement activities enable detailed data collection, especially in occupational health and safety. In this case, processing the collected data, extracting useful information, and using them in occupational safety are separate evaluations. The classical statistical methods or data mining methods can make these evaluations. With classical statistical methods, assessing complex data such as occupational accident data may remain constrained. Data mining methods can obtain relationships that classical statistical methods cannot reach. Data mining is a scientific discipline that provides meaningful and usable information, especially from complex and large data. Considering the databases for occupational health and safety, data mining implementations have a significant advantage in extracting the necessary information from data sets. Relationships that cannot be found with classical statistical methods can be obtained with data mining [1].

Occupational health and safety studies in mining have been researched in conjunction with technological developments in the sector. Studies generally focused on related risks, their measurements, and preventions. Ismail et al. [2] conducted a systematic review of research trends in mining accidents. Between 2015 and 2019, 46% of the mining accident studies identified the main causes of mining accidents, 20% focused on preventing mining accidents, and 34% focused on the challenges and impacts of mining accidents. More than in-depth studies on the root causes of mine accidents and what can be done to address them are required. There are also limited studies on occupational health and safety in mining using data mining. Indeed, Niu et al. [3] presented a review of the current research on emerging technologies in the field of accident prevention, current problems, and predictions for the future. In the study, it is stated that there is no general framework in accident prevention research, and there are deficiencies in the interpretation of the research.

Dessureault et al. [4] investigated the potential for data mining applications using mines' occupational health and safety data. They accordingly modernized the US National Institute for Occupational Safety and Health databases. Thus, data mining software can use modern databases. Shuangyue and Li [5] used the Apriori algorithm on coal mine hidden hazard data and suggested hidden hazard management and prevention. Hazard management and prevention suggestions are reported to be of considerable practical importance in reducing accident loss. Mevsim [6] aimed to determine the leading causes of grizu explosions in underground coal mining using the fault tree analysis technique. The study showed that mechanical ignition devices, ventilation system defects, sparks, electrical ignition devices, and methane degassing significantly affected grizu explosions. It was seen that the probability of a grizu explosion occurring was 100% in 108 months, but if the mechanical ignition device was eliminated, the likelihood of 100% increased to 255 months. Erdogan [7] conducted a study on occupational accidents in the Turkish Hard Coal Corporation between 2000 and 2014 and evaluated occupational accident statistics using the analysis of variance technique. Injuries and workforce losses resulting from occupational accidents are also considered accident severity, and lower and upper-risk levels are determined. Sanmiquel et al. [8] analyzed occupational accidents in mining with data mining techniques using data from the Spanish Ministry of Labor and Social Security between 2005 and 2015. Data mining techniques were selected as a valuable tool to find the root cause of accidents. Ak [9] applied an adaptive neuro-fuzzy inference system-based model for risk assessment in an underground metal mine. The results were better than the traditional methods used for the mines studied.

After 2020, significant contributions have been demonstrated with the studies carried out. Gerassis et al. [10] discussed the applicability of artificial intelligence applications and data mining methods in investigating occupational accidents in mining enterprises in Spain. The study emphasized that to increase occupational health and safety to the targeted level, all statistical information in enterprises should be collected in databases, and these data should be examined using data mining techniques. Iphar and Çukurluöz [11] introduced a fuzzy logic-based safety assessment method to improve the risk assessment process to overcome the uncertainties encountered in the classical decision matrix risk assessment method. This way, high-risk situations and operations in mechanized underground coal mines were identified in linguistic forms with expert knowledge and engineering judgment. Aliabadi et al. [12] quantitatively investigated the effects of human and organizational parameters in mining accidents using the Bayesian network. The analysis revealed that skill-based errors, routine violations, environmental factors, and planned inappropriate operations are essential in mining accidents. Yedla et al. [13] used machine learning techniques for lost working day prediction. The study's predictors, such as shift start time, accident time, and mining experience, are essential in predicting lost working days. Opokua and Opoku [14] developed an adaptive neuro-fuzzy inference system to predict workers' survival rates in underground mines that are falling. Sanmiquel et al. [15] analyzed the types and deviations of accidents in the Spanish mining sector between 2009 and 2018. A graphical evaluation was made by considering the number of accidents that occurred, and a model predicting the number of accidents based on lost working days was also presented. You et al. [16] proposed an approach to assess coal mine gas risk levels. The study used the Support Vector Machines (SVM) algorithm to predict coal mine gas accidents. Li et al. [17] proposed a method of combining text mining, association rule mining, and Bayesian networks for identifying coal mine safety risk factors. The lack of management, education, and supervision was the root cause of coal mine accidents. Bayraktar et al. [18] analyzed and compared the data of 13,653 accidents in Karadon Hard Coal Enterprise between 2000 and 2011 using Life Table and Kaplan−Meier methods. It demonstrated the reliability of using survival analysis methods to investigate occupational accidents in the mining sector. Baraza et al. [19] analyzed the effect of twelve variables related to occupational accidents in mining on the accident severity rate. The variables analyzed (age, gender, nationality, length of service, economic activity, company size, accident location, etc.) were found to be related to the accident severity rate.

As in all fields of mining, there is an increasingly large amount of data collected on occupational health and safety. It is essential to process and evaluate the collected data to extract useful information for the mining sector. The function of data mining in this case is indisputable. There are limited studies in the literature in which occupational accidents in the mining sector are examined with data mining methods, the root causes of accidents are analyzed, and precautions are discussed. This study will help overcome the lack of data analysis for underground hard coal mining, develop data analysis methods, and plan proactive strategies against occupational accidents. Moreover, in this study, it is crucial to use monthly production and monthly wage data in addition to classical accident data (such as worker experience, worker age, affected organ, shift hours, accident cause, and accident location) to reveal the relationship between production and accidents. Therefore, this study mainly aims to analyze the accidents in underground hard coal mining with data mining and to present the relationships and association rules of the accident parameters determined for the study with the accident locations.

2. Material and methods

2.1. Amasra Hard Coal Enterprise

Amasra Hard Coal Enterprise, operating under the General Directorate of Turkish Hard Coal Corporation, is in the Amasra district of Bartın province. Fig. 1 gives the location map of Amasra (Bartın-Turkiye). The coal site has an area of approximately 49 km^2. Coal seams have an average seam thickness of 0—4 m, and strata are at $\leq$40 slope. The production method is usually the retreat longwall mining. Semi-mechanized and fully mechanized excavation systems are used in this production method. As of 2022, there are 542 workers in the enterprise [1].

2.2. Methodology

This study analyzed occupational accidents with injuries in Amasra Hard Coal Enterprise between 2010 and 2021. The data collected on occupational accidents with injuries and the accident locations were associated using data mining methods. The study used WEKA 3.8.6 version open-source data mining software for the analysis. Fig. 2 shows the study methodology.

2.2.1. Data collection

The dataset used in the study includes occupational accidents with injuries in Amasra Hard Coal Enterprise between 2010 and 2021. Table 1 provides information on the number of accidents that occurred during these years. Detailed information on these accidents was obtained from the electronic database of the Turkish Hard Coal Corporation with the necessary permissions. A total of 2169 occupational accidents occurred in the analyzed years. Two thousand one hundred sixty-seven of these accidents were injury accidents, the primary data for this study.

2.2.2. Processing of accident data

Raw occupational accident data were processed and classified to prepare input parameters for statistical analysis. All accident data for twelve years covering 2010-2021 were extracted from the

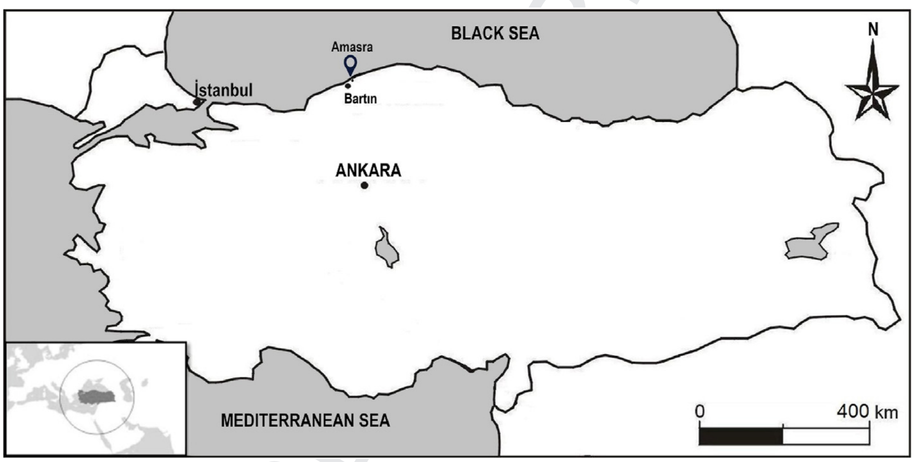

Fig. 1. Location map of Amasra.

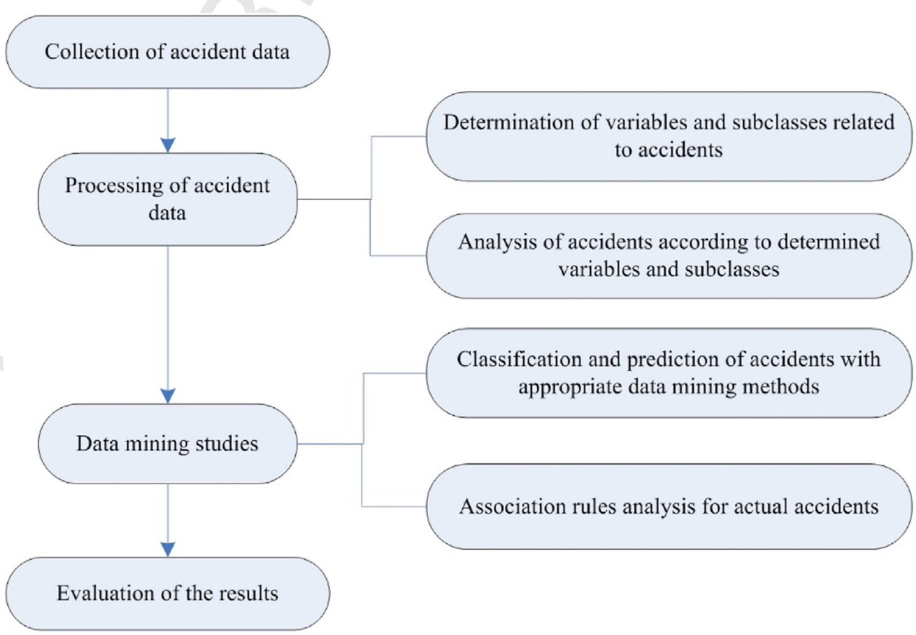

Fig. 2. Methodology of study.

Please cite this article as: Altindis B, Bayram F, , Data Mining Implementations for Determining Root Causes and Precautions of Occupational Accidents in Underground Hard Coal Mining, Safety and Health at Work, https://doi.org/10.1016/j.shaw.2024.09.003

Table 1
Number of occupational accidents with injuries by year [1].

Year	Number of occupational accidents with injuries	Annual raw production (tons)	Number of occupational accidents with injuries per 100,000 tons of production
2010	267	287630	92.82
2011	251	289880	86.58
2012	194	250206	77.54
2013	224	261235	85.74
2014	136	222349	61.16
2015	151	194746	77.54
2016	137	170819	80.20
2017	133	116437	114.22
2018	112	100953	110.94
2019	190	130640	145.44
2020	164	116080	141.28
2021	208	125406	165.86

database, and primarily fatal accidents were identified. As of the years covered by the research, only two fatal accidents occurred in the enterprise, one each in 2012 and 2013. With the removal of deadly accidents, 2167 injury accidents were analyzed.

Injury accidents were grouped using 14 essential variables (defined as attributes in WEKA). These attributes were determined by evaluating the enterprise's previous experiences and operating conditions. These 14 critical attributes include variables directly related to the accidents (affected organ, month of the accident, day of the week of the accident, age of the injured worker, experience of the injured worker, education status of the injured worker, the shift, the working hours of the shift, the number of lost days, worker position, the accident cause, and the accident location) as well as the monthly production and monthly wages. The 14 attributes selected for accident investigation are divided into 108 subclasses. Table 2 shows the attributes and their subclasses selected for the accidents.

When the attributes and subclasses in Table 2 are analyzed regarding the number of accidents, the most common affected organs are the fingers and toes. Accidents occurred in approximately the same numbers every month of the year. The accidents are roughly the same on days (Monday, Tuesday, Wednesday, Thursday and Friday). The age group with the highest number of accidents is 31–35. Workers with 0–5 years of experience are the most exposed to accidents. Regarding education status , primary school graduates had the highest number of accidents. Considering the shift in which the accidents occurred and the working hours (WH) within the shift, the highest number of accidents happened in the 8–16 shift and 4–8 hours of the shift, respectively. The number of accidents in the first hours of shifts (0–1 hour) is relatively low. The highest number of lost working days accidents occurred in the 4–28 days range. The number of accidents fluctuates according to monthly production (MP) and monthly wage (MW). Regarding the worker position (WP), production miners have the highest number of accidents. Similarly, according to the accident cause (AC), underground fallings are the most common cause of accidents.

3. Results

Data mining studies were carried out in two stages. In the first stage, various algorithms were applied to classify and predict accident data. The aim here is to reveal the ability of data mining algorithms to classify and predict accident data. In the second stage,

Table 2
Attributes and their subclasses.

Attribute	Definition of attribute	Subclasses
Affected organ (AO)	Indicates the organ of the worker was damaged as a result of the accident.	Arm, body, face, hand finger, head, knee, leg, toe, waist
Month (M)	Identifies the month in which the accident occurred.	January, February, March, April, May, June, July, August, September, October, November, December
Day (D)	Indicates the day of the week of the accident.	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday
Age (A)	Identifies the age of the injured worker at the date of the accident.	18–25, 26–30, 31–35, 36–40, 41–45, ≥45
Experience (E)	Gives the years of experience of the injured worker at the date of the accident.	0–5, 6–10, 11–15, 16–20, 21–25, ≥26
Education status (ES)	Gives the education status of the injured worker.	Primary, secondary, high school, associate, undergraduate
Shift (S)	Identifies the shift time in which the accident occurred.	0–8, 8–16, 16–24
Working hours (WH)	Shows the time of the accident during the shift.	0–1, 1–4, 4–8
Lost working days (LWD)	Indicates the number of days the worker cannot work due to the accident.	0, 1–3, 4–28, 29–180, >180
Monthly production (MP)	Shows the coal production in tons in the month of the accident.	≤7,999, 8,000–9,999, 10,000–11,999, 12,000–13,999, 14,000–15,999, 16,000–17,999, 18,000–19,999, 20,000–21,999, 22,000–23,999, 24,000–25,999, ≥26,000
Monthly wage (MW)	Indicates the total daily wages for production in the month the accident occurred.	≤3,999, 4,000–4,999, 5,000–5,999, 6,000–6,999, ≥7,000
Worker position (WP)	Refers to the primary position of the injured worker.	Blasting, chief miner, drilling, gallery repair, general service, haulage, machine operator, maintenance, mechanization miner, preparation miner, production miner, shift engineer, shift supervisor, technician, washer
Accident cause (AC)	Refers to the leading cause of the accident.	Technical usage of surface equipment, surface hit-bump, surface material control and handling, surface slip-fall-sprain, surface other, technical usage of underground equipment, underground falling, underground hit-bump, underground material control and handling, underground slip-fall-sprain, underground other
Accident location (AL)	Indicates the location of the accident.	Surface pit area, surface rotary dumper, surface shaft, surface washery, surface workshop, underground face, underground gallery, underground production shaft, underground rotary dumper, underground shaft, underground ventilation shaft

Please cite this article as: Altindis B, Bayram F, , Data Mining Implementations for Determining Root Causes and Precautions of Occupational Accidents in Underground Hard Coal Mining, Safety and Health at Work, https://doi.org/10.1016/j.shaw.2024.09.003

association rule mining was applied to accident data. The aim here is to reveal the hidden and exciting relationships between variables in the form of rules.

3.1. Implementation of data mining methods on accident data

The classification processes were performed using appropriate data mining algorithms in the database of injury accident data for 2010–2021. There are two primary purposes in the classification process. The first is to determine the algorithm that best represents all the data. The other is to assess the ability of the data mining algorithms to predict the desired parameter.

Training set and cross-validation were selected as the test methods in modeling studies in this stage. In addition to these test methods, Kappa, mean absolute error (MAE), and root mean square error (RMSE) statistics were used to determine the classification accuracy of the algorithms.

The accuracy of the algorithms is determined by the percentage of correctly classified data in the whole data set. No changes were made to the default settings of the applied data mining algorithms. Considering only the performance, some modifications were made to some of the parameters to obtain better results. The first algorithm applied in the modeling studies is the ZeroR algorithm. This algorithm gives a fundamental result in the modeling studies. It is a crucial algorithm for evaluating the accuracy of other algorithms and checking for increases in the level of relationship. The accuracy of modeling studies in data mining can be measured by how far the accuracy of the ZeroR implementation, i.e., the correlation coefficient obtained as a result of this algorithm, is improved. Other algorithms applied to the dataset are Naive Bayes, J48 Decision Tree (J48), k Nearest Neighbor (kNN), Artificial Neural Networks (MLP -Multilayer Perceptron), Support SVM and Random Forest (RF) algorithms. Table 3 shows the algorithms applied to the dataset and the results obtained.

In the training set method, kNN and RF algorithms showed excellent results and achieved 100% accuracy. In data mining, the first thing that comes to mind in such cases (100% accuracy) is overlearning and overfitting the algorithms. However, not every 100% success should be considered as such. If there is no noisy data in the data set that would cause overlearning and overfitting, this result may not be something to be suspicious about. For this, the WEKA program can check whether the data set contains noisy data. Outliers and extreme values in the data set are identified using the interquartile range filter in the WEKA program. In the filtering process, if positive results are obtained regarding these values, these values are removed from the data set, and noisy data is cleaned. When the interquartile range filter was applied to the data set used in this study phase, no positive results were obtained for either value (the number of outliers and extreme values is 0). This shows that there are no meaningless data in the data set. Other performance parameters were also analyzed. The Kappa statistic, MAE, and RMSE values indicate that the kNN and RF algorithms are the most successful.

Among the algorithms applied in the cross-validation test method, the SVM algorithm produced the best result by correctly classifying 66.08% of the dataset. Regarding accuracy statistics (Kappa, MAE, and RMSE), the SVM algorithm also performed better than the other applied algorithms.

3.2. Association rule mining for accident data

In association rule mining, Apriori and Predictive Apriori algorithms were implemented. The distributions of attributes and subclasses were analyzed for both algorithm implementations. Thus, more profound knowledge about the accident's origin was uncovered.

The Apriori algorithm finds the dataset's most frequent attributes and generates association rules with these attributes [20]. While generating the rules with the algorithm, the confidence criterion determined in the implementation is considered. The association rules generated according to this confidence criterion are ranked. Table 4 presents the first 20 rules generated for accident locations by the Apriori algorithm.

Table 4 lists the association rules starting with the highest confidence criterion. Considering the number of accidents, these association rules gathered the most compatible attributes with the accident location (AL). For example, in Rule 1, it is possible with high confidence (83%) that a production miner with 0–5 years of experience in the underground face will be injured in an underground falling between 4–8 hours of his shift. Again, according to Rule 2, a production miner with 0–5 years of experience who has an accident between 4–8 hours of his shift in the underground face suffers 4–28 days of lost working days due to the accident. This association is realized at a rate of 82%. Therefore, crucial deep analysis and information can be obtained for the enterprise from the association rules generated. In the association rules generated by the Apriori method, it was observed that the most common attributes were experience (0–5), WH (4–8), and WP (production miner). In all of the rules, underground face is included as the accident location (AL). Table 5 shows the distribution of attributes and subclasses in the 20 association rules.

In Table 5, the frequency of experience, WH, and WP, which are the attributes that match the highest with AL, in the 20 association rules, is 65, 70, and 75%, respectively. The frequency of other attributes in the rules varies between 0 and 35%. The subclasses of the most relevant attributes are 0–5 years experience, 4–8 WH, and

Table 3
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.

Please cite this article as: Altindis B, Bayram F, , Data Mining Implementations for Determining Root Causes and Precautions of Occupational Accidents in Underground Hard Coal Mining, Safety and Health at Work, https://doi.org/10.1016/j.shaw.2024.09.003

Table 4
Association rules generated for ALs using the Apriori algorithm.

Rule	Association rules				Confidence
	Attributes of the rule	Number of accidents for rule attributes	Target attribute	Number of accidents for target attribute	
1	E = 0–5, WH = 4–8, WP=production miner, AC=underground falling	283	AL = underground face	234	0.83
2	E = 0–5, WH = 4–8, LWD = 4–28, WP=production miner	332	AL = underground face	271	0.82
3	E = 0–5, ES=primary, WH = 4–8, WP=production miner	339	AL = underground face	276	0.81
4	E = 0–5, ES=primary, WH = 4–8	345	AL = underground face	278	0.81
5	S = 0–8, WH = 4–8, WP=production miner	275	AL = underground face	221	0.80
6	E = 0–5, WH = 4–8, MW = 5000––5999, WP=production miner	271	AL = underground face	217	0.80
7	E = 0–5, WH = 4–8, WP=production miner	577	AL = underground face	462	0.80
8	A = 26–30, E = 0–5, WH = 4–8, WP=production miner	291	AL = underground face	233	0.80
9	A = 26–30, WH = 4–8, WP=production miner	300	AL = underground face	240	0.80
10	E = 0–5, WP=production miner, AC=underground falling	458	AL = underground face	365	0.80
11	E = 0–5, WH = 4–8, AC=underground falling	314	AL = underground face	248	0.79
12	E = 0–5, WH = 4–8, MW = 5,000–5,999	289	AL = underground face	227	0.79
13	E = 0–5, LWD = 4–28, AC=underground falling	288	AL = underground face	226	0.78
14	WH = 4–8, WP=production miner, AC=underground falling	417	AL = underground face	323	0.77
15	E = 0-5, ES=primary, LWD = 4–28, WP=production miner	348	AL = underground face	268	0.77
16	E = 0–5, WH = 4–8, LWD = 4–28	372	AL = underground face	286	0.77
17	LWD = 4–28, WP=production miner, AC=underground falling	378	AL = Underground Face	290	0.77
18	WP=production miner, AC=underground falling	686	AL = underground face	525	0.77
19	A = 31–35, WH = 4–8, WP=production miner	332	AL = underground face	254	0.77
20	A = 26–30, LWD = 4–28, WP=production miner	285	AL = underground face	218	0.76

A, age; AC, accident cause; AL, accident location; E, experience; LWD, lost working days; MW, monthly wage; WH, working hours; WP, worker position.

WP as production miner. It was found that workers with 0–5 years of experience are more likely to have accidents than others; accidents usually occur between 4 and 8 hours, which are the last hours of shifts, and production miners have accidents.

With the Predictive Apriori algorithm, rules are obtained with high accuracy using the predictive accuracy criterion [21]. Since not all rules are scanned compared with the other implementation, the algorithm gives faster results and generates more reliable rules. Table 6 shows the top 20 association rules generated with the Predictive Apriori algorithm.

The Predictive Apriori (Table 6) algorithm provides sharper inferences than the other method. For example, in Rule 1, in the case of an underground falling between 4 and 8 hours of the shift on underground face under the conditions where MP is 24,000–25,999 tons and MW is 6,000–6,999, 98% of the time, the experience of the injured worker is 0–5 years. Likewise, according to Rule 2, when 20000–21999 tons of MP is realized in July, the probability of a worker with 0–5 years of experience being injured in an underground face accident is 97.9%. More precise associations can be more accurately inferred with the Predictive Apriori method.

Table 5
Attribute and subclass distributions in association rules generated with the Apriori algorithm.

Attribute	Number of association rules found	Frequency (%)	Subclass (number of rules found)
WP	15	75	Production miner (15)
WH	14	70	4–8 (14)
E	13	65	0–5 (13)
AC	7	35	Underground falling (7)
LWD	6	30	4–28 (6)
A	4	20	26–30 (3), 31–35 (1)
ES	3	15	Primary (3)
MW	2	10	5,000–5,999 (2)
S	1	5	0–8 (1)
AO	-	-	
M	-	-	
D	-	-	
MP	-	-	

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

The attributes show a more uniform distribution within the association rules generated for AL by the Predictive Apriori method. Experience, shift, WH, WP, and AC attributes are more common in most rules. The AL was considered an underground face in all of the rules. Table 7 shows attributes and subclass distribution in the 20 association rules.

In 20 association rules, the frequencies of the attributes experience, shift, WH, WP, and AC matching the AL attribute the most were 45, 50, 55, 55, 55, and 50%, respectively (Table 7). The percentages of other attributes in the rules vary between 0 and 35%. When the subclasses of the most matched attributes are analyzed, the subclasses of 0–5 years of experience, 16–24 of shift, WH as 4–8 hours of the relevant shift, production miner as WP, and underground falling as AC are prominent. In the analysis performed with the Predictive Apriori method, it has been determined that workers with 0–5 years of experience are more likely to come across accidents than others, especially 16–24 shifts are essential in terms of accidents, accidents generally occur between 4–8 hours, which are the last hours of the shifts, production miners come across the most accidents, and underground fallings are prominent in an accident cause. In addition, it is noteworthy that MP and MW are among the root causes of accidents.

3.3. Discussions

This study analyzed 2167 injury accidents in Amasra Hard Coal Enterprise using various data mining methods between 2010 and

Table 6
Association rules generated for ALs using the Predictive Apriori algorithm

Rule	Association rules				Accuracy
	Attributes of the rule	Number of accidents for rule attributes	Target attribute	Number of accidents for target attribute	
1	E = 0–5, WH = 4–8, MP = 24,000–25,999, MW = 6,000–6,999, AC=underground falling	23	AL = underground face	23	0.9804
2	M = July, E = 0–5, MP = 20,000–21,999	22	AL = underground face	22	0.9794
3	M = March, WH = 4–8, WP=production miner, AC=underground material control and handling	22	AL = underground face	22	0.9794
4	AO = Toe, S = 8–16, WH = 4–8, WP=production miner, AC=underground falling	22	AL = underground face	22	0.9794
5	E = 0–5, S = 16–24, WH = 4–8, MP = 24,000–25,999, AC=underground falling	22	AL = underground face	22	0.9794
6	D = Monday, E = 0–5, S = 16–24, WH = 4–8, MW = 5,000–5,999, WP=Production Miner	22	AL = underground face	22	0.9794
7	AO = toe, D = Wednesday, S = 16–24, WP=production miner	21	AL = underground face	21	0.9782
8	M = March, A = 26–30, WP=production miner, AC=underground falling	21	AL = underground face	21	0.9782
9	D = Tuesday, S = 16–24, LWD = 4–28, WP=production miner, AC=underground falling	21	AL = underground face	21	0.9782
10	AO=Waist, E = 0–5, WH = 4–8, WP=production miner	20	AL = underground face	20	0.9768
11	D = Monday, WH = 4–8, WP=production miner, AC=underground hit-bump	20	AL = underground face	20	0.9768
12	D = Saturday, E = 0–5, S = 0–8, WH = 4–8	20	AL = underground face	20	0.9768
13	D = Thursday, A = 26–30, S = 8–16, AC=underground falling	20	AL = underground face	20	0.9768
14	M = April, A = 31–35, E = 0–5, MW = 5,000–5,999	19	AL = underground face	19	0.9754
15	M = April, A = 31–35, S = 16–24, WP=production miner	19	AL = underground face	19	0.9754
16	AO=finger, A = 31–35, WH = 4–8, MP = 20,000–21,999	18	AL = underground face	18	0.9736
17	AO = toe, E = 6–10, S = 16–24, AC=underground falling	18	AL = underground face	18	0.9736
18	AO = toe, WH = 4–8, WP=production miner, AC=underground hit-bump	18	AL = underground face	18	0.9736
19	M = April, S = 0–8, MW = 5,000–5,999, WP=production miner	18	AL = underground face	18	0.9736
20	M = July, A = 31–35, E = 0–5, WH = 4–8	18	AL = underground face	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, monthly production; MW, monthly wage; S, stage; WH, working hours; WP, worker position.

2021. Practical data mining algorithms were used to determine the ALs, and association rule mining revealed the main causes of accidents. The presence of 14 different attributes specified for the accidents considered in the study was determined when the accidents occurred.

The study includes the accidents in Amasra Coal Enterprise operating in the Zonguldak Coal Basin in Turkiye. Therefore, the results obtained here may differ from those in other countries. However, the results are valuable since the risks and hazards in underground hard coal mining are similar in every quarry.

In data mining based on the classification and prediction of accidents, the kNN algorithm showed the best result according to the training set test method (100% success) and the SVM algorithm (66% success) according to the cross-validation test method. Mainly, kNN and SVM algorithms can be used in deep analysis of occupational accidents in mining.

The association rule mining applied to accident data with Apriori and Predictive Apriori methods investigated the coexistence of AL attribute and other attributes. Association rules were generated with high confidence and accuracy. While the Apriori method generated rules with a 76–83% confidence interval, the Predictive Apriori method generated rules with 97.3–98% accuracy.

Experience, WH, and WP attributes are prominent in association rules generated for the AL target attributes by the Apriori method. Experience, Shift, WH, WP, and AC attributes gain importance in

Table 7
Attribute and subclass distributions in association rules generated with the Predictive Apriori algorithm

Attribute	Number of association rules found	Frequency (%)	Subclass (number of rules found)
WH	11	55	4–8 (11)
WP	11	55	Production miner (11)
S	10	50	0–8 (2), 8–16 (2), 16–24 (6)
AC	10	50	Underground falling (7), underground material control and handling (1), underground hit-bump (2)
E	9	45	0–5 (8), 6–10 (1)
M	7	35	March (2), April (3), July (2)
AO	6	30	Toe (4), waist (1), hand finger (1)
D	6	30	Monday (2), Tuesday (1), Wednesday (1), Thursday (1), Saturday (1)
A	6	30	26–30 (2), 31–35 (4)
MP	4	20	20,000–21,999 (2), 24,000–25,999 (2)
MW	4	20	5,000–5,999 (3), 6,000–6,999 (1)
LWD	1	5	4–28 (1)
ES	-	-	

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

association rules generated for AL target attributes by the Predictive Apriori method.

The Apriori algorithm is the most basic association analysis method for finding frequently repeated items. Association rules created with Apriori consist of two parameters: support and confidence. If the support and confidence values are above the threshold values defined by the user, association rules are created. Predictive Apriori combines confidence and support into a single measure of prediction accuracy and finds the n best association rules sequentially. Like Apriori, it ultimately searches for rules with more than one condition but differs in that these conditions are ORed rather than ANDed. The frequency of attributes or subclasses may vary in both methods' first 20 association rules.

4. Conclusions

As a result of the study, several recommendations are presented for evaluating and preventing underground mining occupational accidents. Many different parameters are effective in mining occupational accidents. In the present conditions, using data mining methods to analyze mining occupational accidents and put forward prevention is vital.

When the occurrence of accidents is evaluated, 4–8 hours of the shift appear to be the most matched attribute subclass with a rate of 70% in the Apriori method and 55% in the Predictive Apriori method. This suggests that workers may be fatigued between the 4th and 8th hours of the shift, or their concentration may be impaired due to the anxiety of completing the assigned work. The shift supervisors should closely supervise workers, especially between these hours. The work to be done in the shift should be planned to be completed within the working hours. These plans should be organized, considering any previous or current shift disruptions.

Another attribute that most matches with the AL in the association rules is the experience of the worker. In most accidents, less experienced workers are involved, especially workers with 0–5 years of experience . Therefore, it is recommended that more qualitative and specialized training should be provided to workers with less experience, and workers with less experience should be partnered with more experienced workers and work under their supervision.

Underground falling is the subclass with the highest match with AL as an AC. Compared with conventional production, mechanized systems support the roof more safely. These analyses show that using fully mechanized systems as the production method and shield supports as the support system will benefit coal mining.

It has been observed that MP and MW attributes are also included in the association rules. It is recommended that it would be beneficial to carry out the productions made by the work plan by Q2 applying the occupational health and safety principles as a priority.

CRediT authorship contribution statement

Bilal Altindis: Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. **Fatih Bayram:** Data curation, Investigation, Methodology, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are very grateful to the General Directorate of Turkish Hard Coal Corporation Amasra Hard Coal Enterprise for its support (permission) during this research.

References

[1] Altindis B. Investigation of occupational accidents in Amasra Hard Coal Enterprise [master's thesis]. [Afyonkarahisar]. Afyon Kocatepe University; 2023. 95 p [in Turkish].
[2] Ismail SN, Ramli A, Aziz HA. Research trends in mining accidents study: a systematic literature review. Saf Sci 2021;143:105438. https://doi.org/10.1016/j.ssci.2021.105438.
[3] 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.
[4] Dessureault S, Sinuhaji A, Coleman P. Data mining mine safety data. Min Eng 2007;59(8):64–70.
[5] 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. 112–6. 2013 Nov 6-7; Zhangjiajie, Hunan Province, China.
[6] 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.
[7] Erdoğan 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.
[8] Sanmiquel L, Bascompta M, Rossell JM, et al. 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. Q3
[9] 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. 167–205.
[10] Gerassis S, Saavedra Á, Taboada J, et al. Differentiating between fatal and non-fatal mining accidents using artificial intelligence techniques. Int J Min Reclam Environ 2020;34(10):687–99. https://doi.org/10.1080/17480930.2019.1700008.
[11] Iphar M, Çukurlüöz AK. Fuzzy risk assessment for mechanized underground coal mines in Turkey. Int J Occup Saf Ergon 2020;26(2):256–71. https://doi.org/10.1080/10803548.2018.1426804.
[12] 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):670–7. https://doi.org/10.1080/10803548.2018.1455411.
[13] 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.
[14] 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.
[15] Sanmiquel L, Bascompta M, Rossell JM, Anticoi H. Analysis of occupational accidents in the Spanish mining sector in the period 2009–2018. Int J Environ Res Public Health 2021;18:13122. https://doi.org/10.3390/ijerph182413122.
[16] 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;162:1067–81. https://doi.org/10.1016/j.psep.2022.04.054.
[17] 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.
[18] Bayraktar B, Uyguçgil H, Konuk A. Investigation of occupational accidents in mining with survival analysis. Min Metall Explor 2023;40:1827–38. https://doi.org/10.1007/s42461-023-00810-5.
[19] Baraza X, Cugueró-Escofet N, Rodríguez-Elizalde R. Statistical analysis of the severity of occupational accidents in the mining sector. J Saf Res 2023;86:364–75. https://doi.org/10.1016/j.jsr.2023.07.015.
[20] 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. 487–99. Santiago, Chile.
[21] 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. 424–35. https://doi.org/10.1007/3-540-44794-6_35.