Analysis of Risk Factors in The Firework Industries

[1]Prof. Vishal Shinde, [2]Miss. Diksha Jagdish Raut, [3]Miss. Priyanka Mahendra Binnar, [4]Miss. Vallavi Bajirao Jadhav.

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]mailme.vishalshinde@gmail.com, [2]diksharaut2511@gmail.com, [3]priyankabinnar9@gmail.com, [4]vallavi2000@gmail.com

Abstract - Big data, machine learning, and artificial intelligence are also becoming more significant in this era of data mining approaches as the country decided to increase intellectual capacity in its manufacturing operations. Fireworks industry accidents (FlAs), which are the leading cause of mortality, have long been one of the major global hazards and danger issues. Fireworks occurrences are influenced by a variety of factors, including chemicals, facilities, transportation, temperature, friction, etc. Some of which are more crucial than others in determining the accident's seriousness. Risk assessment or worker safety behaviour can offer insights into how to fix the industry's performance problems. The successful identification and forecasting of specific influencing elements among human health and safety predictors can be accomplished with the aid of analytical data mining methods. This method was examined on a real dataset. In this investigation, the decision tree classifier's classification approaches enable the prediction of the risk factor in the fireworks industry among other safety-related criteria [1].

Keywords- Risk Assessment, Decision Tree Classifier, Machine Learning, Artificial Intelligence.

I. INTRODUCTION

Firework industry workers are particularly vulnerable to fires, explosions, and chemical exposure due to the nature of their work. Therefore, it is essential to identify and analyze the risk factors in the firework industry to develop and implement effective safety strategies that prevent accidents and ensure worker safety. This paper discusses the key risk factors in the firework industry and provides insights into the possible strategies for mitigating those risks [1][5].

Using machine learning algorithms such as the random forest and decision tree classifier can provide a more accurate and efficient way of identifying and assessing potential hazards. These algorithms can be trained on historical data and used to predict the likelihood of future hazards allowing for more effective risk management strategies. The random forest algorithm uses multiple decision trees to make predictions. It is well-suited for analyzing complex datasets which is important in the firework industry where there are numerous risk factors to consider. Similarly, the decision tree classifier that can be used to identify and evaluate risk factors in the firework industry [1].

II. AIMS AND OBJECTIVE

a) Aim

The project's goal was to investigate the causes of accidents, analyze the accidents, and suggest preventative measures.

b) Objective

The objective of this analysis was to identify the contributing factors to accidents, to identify the areas of concern and to initiate research to prevent these accidents.

III. LITERATURE SURVEY

Paper 1: Industrial Safety and Well-Being of Fireworks Employees in Sivakasi

AUTHOR: P. Ramya, S.Rubhy.

Particularly in Sivakasi, the fireworks business is particularly thriving in the Virudhunagar District. However, the workers in this field do not receive enough safety precautions to ensure their safety while working. Numerous health problems have been caused by the air quality. Employee safety is of utmost importance in the fireworks industry. Thus, the study's primary goal is to identify safety precautions in the fireworks sector. because a survey was conducted by the researcher in Vijayarangapuram, close to Sivakasi.

 DOI : 10.5281/zenodo.7863705

The researcher collected 30 samples from the population from the 250 employees who made up the overall population. The safety and wellbeing of the employees were the primary focus of this investigation. Most respondents felt they needed extra safety precautions, such as gloves and face masks [2].

Paper 2: Performance Evaluation of Lazy and Decision Tree Classifier: A Data Mining Approach for Global Celebrity's Death Analysis

AUTHOR: V. Kumar, R. Zinovyev, A. Verma, P. Tiwari

Celebrities have a large influence on ordinary people's behaviours through biological, psychological, and social mechanisms. Simultaneously, there is a wide variance in their deaths based on the cause, place, and age. As a result, it is a tough and intriguing endorsement to work on. The goal of this endeavour is to produce a thorough result to comprehend celebrity deaths by investigating incidents that occurred over the decade. The training database was generated using public and open access databases containing 11,200 recorded deaths from 2006 to 2016. Year by year, the work extracts death, the reason of death, age, gender, and location [3][9].

Paper 3: A Survey on Predicting Heart Disease using Data Mining Techniques

AUTHOR: C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh, S. Deepa Rajan

Heart disease is the most dangerous cause of death. It has a significant long-term impairment. This sickness strikes a person so quickly. Medical data is still rich in information but deficient in expertise. As a result, appropriately diagnosing patients in a timely manner is a critical function for medical support. The hospital's reputation suffers as a result of an incorrect diagnosis.

The most pressing biomedical challenge is the accurate detection of cardiac disease. The goal of this study is to create an effective treatment using data mining approaches that can aid in restorative situations. To diagnose heart problems, further data mining classification methods such as decision trees, neural networks, Bayesian classifiers, Support vector machines, Association Rule, and K- closest neighbour classification are utilized [4][8].

IV. EXISTING SYSTEM

The firework safety and risk assessment fisra system and the risk assessment and management of pyrotechnic operations ramp system are two systems that use machine learning algorithms to analyze risk factors in the fireworks industry. The fisra system uses a decision tree classifier to pinpoint elements like the type of fireworks, the venue where the event will take place, and the weather that increase the danger of accidents. The ramp system uses a library of previous firework accident data to train the machine learning algorithms and spot patterns and trends in the data. These systems show the potential of machine learning algorithms for the study of risk variables in the fireworks business [3].

V. COMPARATIVE STUDY

Table.1: Comparative Study

Sr. No	Paper Name	Author/ Publication	Technology	Purpose
1.	Industrial Safety and Well-Being of Fireworks Employees in Sivakasi	P. Ramya, S.Rubhy	Data Mining	This paper reviews the hazards faced by the workers working in match box and fireworks Industry in Sivakasi.
2	Performance Evaluation of Lazy and Decision Tree Classifier: A Data Mining Approach for Global Celebrity's Death Analysis	V. Kumar, R. Zinovyev, A. Verma, P. Tiwari	Decision tree	Celebrities significantly affect the behaviours of regular people through biological, psychological, and social mechanisms. The goal of this research is to investigate the occurrence of celebrity deaths over a ten-year period and produce a comprehensive understanding of them.
3.	A Survey on Predicting Heart Disease using Data Mining Techniques	C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh, S. Deepa Rajan	SVM, KNN	The primary biological problem is the accurate diagnosis of cardiac disease. The goal of this research is to use data mining techniques to create an effective remedy for restorative situations.

VI. PROBLEM STATEMENT

The Random Forest Model, and the Decision Tree Classifier can all be used to enhance good teaching on the hazards of the chemicals people use at work. Knowledge regarding personal hygiene, adequate protective equipment like gloves, apron can be provided and training for usage of newest protective equipment should be offered.

VII. PROPOSED SYSTEM

A proposed system for analyzing risk factors in the firework industry would involve identifying potential risk factors, collecting data related to each factor, analyzing the data to

 DOI : 10.5281/zenodo.7863705

assess the level of risk, and recommending mitigation strategies to reduce the potential for accidents and improve workplace safety. This would require the use of statistical techniques, safety records, and employee training data to identify patterns and correlations between different variables. The system would then provide recommendations for improving safety protocols, enhancing safety training, and implementing new safety technologies [6][1].

VIII. ALGORITHM

The Algorithm for the Analysis of Risk Factors in The Firework Industries:

Step.1: Start
Step.2: Import necessary libraries
import pandas as pd
from sklearn.model_selection
import train_test_split
from sklearn.ensemble
import RandomForestClassifier
from sklearn.tree
import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
Step.3: Load dataset
fireworks_data = pd.read_csv('fireworks_data.csv').
Step.4: Split the data into training and testing sets
X = fireworks_data.drop('Risk', axis=1)
y = fireworks_data['Risk']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
Step.5: Train the decision tree classifier
dtc = DecisionTreeClassifier()
dtc.fit(X_train, y_train)
Step.6: Train the random forest classifier
rfc = RandomForestClassifier(n_estimators=100)
rfc.fit(X_train, y_train)
Step.7: Evaluate the decision tree classifier
y_pred_dtc = dtc.predict(X_test)
dtc_accuracy=accuracy_score(y_test, y_pred_dtc)
dtc_cm=confusion_matrix(y_test, y_pred_dtc)
Step.8: Evaluate the random forest classifier
y_pred_rfc = rfc.predict(X_test)
rfc_accuracy=accuracy_score(y_test, y_pred_rfc)
rfc_cm=confusion_matrix(y_test, y_pred_rfc)
Step.9: Print the results
print('Decision Tree Classifier Results:')
print('Accuracy:', dtc_accuracy)
print('Confusion Matrix:', dtc_cm)
print('\nRandom Forest Classifier Results:')
print('Accuracy:', rfc_accuracy)
print('Confusion Matrix:', rfc_cm)
Step.10: Exit

IX. MATHEMATICAL MODEL

Decision Tree (DT): A supervised learning method called a decision tree can be used to solve classification and regression problems, however it is typically chosen for dealing with classification problems.

Decision trees are used to partition the data into smaller subsets based on the values of the input features, and to make a prediction based on the majority class or average value of the samples in each leaf node. Bootstrap sampling is used to create multiple subsets of the data, which are used to train different decision trees [7].

Random Forest: Random forest is a machine learning algorithm that can be used for classification and regression tasks. In the context of analyzing risk factors in the firework industry, it can be used to identify the most important predictors or risk factors that are associated with accidents or incidents.

The random forest algorithm is based on the idea of creating multiple decision trees, and then combining their predictions to make a final decision.

The random forest algorithm involves several equations, including those used to calculate information gain or impurity, and those used to aggregate the predictions of the decision trees.

Information gain: The information gain is used to measure the reduction in entropy or impurity that results from splitting the data on a particular feature. The equation for information gain is:

$$IG = H(S) - \sum(|Sv|/|S|)\, H(Sv)$$

where:

- IG is the information gain
- H(S) is the entropy or impurity of the parent node S
- Sv is the subset of S that results from splitting on a particular feature v
- |Sv| is the number of samples in Sv
- |S| is the total number of samples in S
- H(Sv) is the entropy or impurity of Sv

Gini impurity: The Gini impurity is another measure of impurity that is commonly used in the random forest algorithm. The equation for Gini impurity is:

$$Gini = 1 - \sum(p\_i)^2$$

where:

- p_i is the proportion of samples in a node that belong to class i

Decision tree prediction: The prediction of a decision tree is based on the majority class or average value of the samples in the leaf node.

 DOI : 10.5281/zenodo.7863705

For classification tasks, the predicted class y is:

y = argmax(∑[y_i == c])

where:

- y_i is the predicted class for sample i
- c is a class label

For regression tasks, the predicted value y is:

y = mean(y_i)

Random forest aggregation: The predictions of the decision trees in a random forest are aggregated using either the mode (for classification tasks) or the mean (for regression tasks) of the individual predictions. The equation for the aggregated prediction is:

y = mode(y_1, y_2, ..., y_n) OR

y = mean(y_1, y_2, ..., y_n)

where:

- y_1, y_2, ..., y_n are the predictions of the individual decision trees

X. SYSTEM ARCHITECTURE

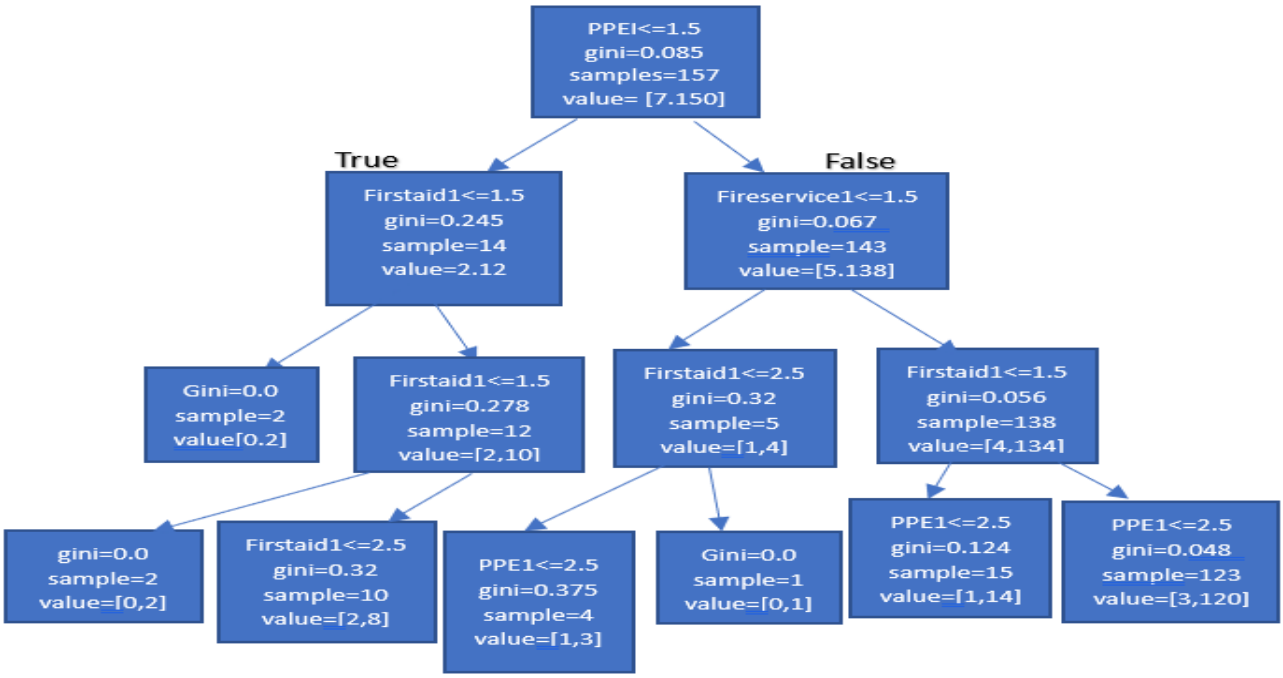

Fig.1: System Architecture

In above fig, a decision tree model is created to show the relationships between the different risk factors and their potential outcomes, allowing the system to visualize and analyze the various factors and their interactions. The system then creates a random forest model by combining multiple decision trees, which helps to improve the accuracy of the risk analysis by reducing the effects of overfitting and increasing the robustness of the model.

XI. ADVANTAGES

- Risk analysis can help to identify potential hazards associated with firework manufacturing, handling, and storage.
- Development of safety measures can help minimize or eliminate the identified hazards.
- Reducing accidents and injuries can result in a safer workplace and reduced workers' compensation claims.
- Compliance with safety regulations and standards can help avoid legal liabilities and penalties.
- Improved reputation can help to maintain customer loyalty and increase business profitability.

 DOI : 10.5281/zenodo.7863705

XII. DESIGN DETAILS

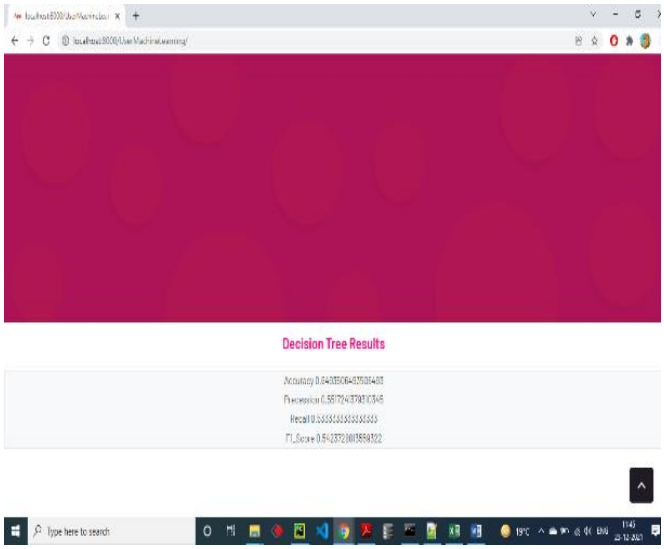

Fig.2: Result

Classification Algorithm	Accuracy	Precision	Recall	F1-measure
Decision Tree	93.32%	93.34%	93.32%	93.22%
Random Forest	95.67%	94.10%	93.32%	95.88%

Fig .3: Testing Performance on Train Set

Classification Algorithm	Accuracy	Precision	Recall	F1-measure
Decision Tree	93.42%	93.34%	92.32%	91.52%
Random Forest	94.66%	92.10%	93.10%	95.80%

Fig.4: Testing Performance on Validation Dataset

XIII. CONCLUSION

Thus, we have tried to implement the paper "Analysis of risk factors in the Firework Industries: Using Decision Tree Classifier", N. Indumathi, R. Ramalakshmi and V. Ajith, ICACITE, 2021 and the conclusion is as follows: The results obtained from the analysis shows that the most critical risk factors are the type of firework, the storage conditions, and the location of the storage area. The decision tree and random forest models identified these factors as having the highest predictive power in determining the risk of accidents in the firework industry. Moreover, the results indicate that the random forest classifier performed better than the decision tree classifier, as it achieved a higher accuracy and F1 score. This suggests that the random forest model is better suited for predicting the risk of firework accidents based on multiple risk factors. By implementing this measures, firework manufacturers can reduce the risk of accidents and promote a safe working environment for their employees.

REFERENCE

[1] N. Indumathi, R. Ramalakshmi and V. Ajith, "Analysis of risk factors in the Firework Industries: Using Decision Tree Classifier", International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 811-814.

[2] P. Ramya, S.Rubhy, (2019), Industrial Safety and Well-Being of Fireworks Employees in Sivakasi, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9, Issue-2S2, Page:322-324.

[3] V. Kumar, R. Zinovyev, A. Verma, P. Tiwari, (2018), Performance Evaluation of Lazy and Decision Tree Classifier: A Data Mining Approach for Global Celebrity's Death Analysis, International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, El Salvador, pp. 1-6, doi: 10.1109/RICE.2018.8509045.

[4]C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh, S. Deepa Rajan, (2018), A Survey on Predicting Heart Disease using Data Mining Techniques, Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, India, pp. 253-255, doi: 10.1109/ICEDSS.2018.8544333.

[5] Porchelvil, R. S., & Devi, P. J. (2015). Regression Model for the People working in Fire Work Industry- Virudhunagar District, International Journal of Scientific and Research Publications, Volume 5, Issue 4.

[6] Kabra R.R., Bichkar R.S. (2011). Performance Prediction of Engineering Students using Decision Trees, International Journal of Computer Applications (0975- 8887), Vol-36-No.11.

[7] Yeon, Y. K., Han, J. G., & Ryu, K. H. (2010). Landslide susceptibility mapping in Injae, Korea, using a decision tree. Engineering Geology, 116(3-4), 274-283.

[8] Ermini, L., Catani, F., & Casagli, N. (2005). Neural artificial networks applied to landslide susceptibility assessment. geomorphology, 66(1-4), 327-343.

[9] Han J., Kamber M., and Pie J. (2006). Data Mining Concepts and Techniques. 2nd edition, Morgan Kaufmann Publishers.

 DOI : 10.5281/zenodo.7863705