الجمهـــوريـــة الجزائـــريـــة الديمقـــراطيـــة الشـعبيـــة
People's Democratic Republic of Algeria
وزارة التــــعلـــيــــم العــــــالـــــي والبــحـــــث العــلـمـــــي
Ministry of Higher Education and Scientific Research

جامعة وهران 2 محمد بن أ حمد
University of Oran 2 Mohamed Ben Ahmed

معهد الصيانة و الأمن الصناعي
Institute of Maintenance and Industrial Safety

Department of Industrial Safety and Environment

THESIS

For the Requirements of a Master's Degree

Field: Industrial Safety and Hygiene
Specialty: Safety, Prevention and Intervention

Topic

Safety Management with Machine Learning Models

Submitted By:

SOUIYOU Fares and TEKOUK Mohamed

Before the Jury Composed of:

Full Name	Grade	Institution	Role
ROUANE SERRIK Mehdi	MCB	University of Oran 2	**Chairman**
ZOUAIRI Saim	MAA	University of Oran 2	**Supervisor**
TITAH Mouloud	MCB	University of Oran 2	**Examiner**

Academic Year 2022/2023

Acknowledgements

بِسْمِ اللهِ الرَّحْمَٰنِ الرَّحِيمِ، الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ، وَالصَّلَاةُ وَالسَّلَامُ عَلَىٰ مُحَمَّدٍ عَبْدِهِ وَرَسُولِهِ

First and foremost, we must acknowledge our limitless thanks to Allah, the Ever-magnificent, the Ever-Thankful, for His help and bless by giving us the opportunity, courage and sufficient energy to carry out and complete this entire thesis for the attainment of our Master's degree.

We sincerely appreciate the guidance, support, and advice provided by our supervisor, ***ZOUAIRI Saim****, throughout the research process.*

We would like to extend our heartfelt appreciation to the jury members, ***Mr. ROUANE SERRIK Mehdi****, and* ***Mr. TITAH Mouloud****.*

Furthermore, we would like to express our heartfelt appreciation to our families and loved ones for their unwavering support, prayers, and understanding. Their belief in our abilities and continuous encouragement have been a source of motivation and strength throughout this academic journey.

We extend our gratitude to our friends and colleagues for their support, intellectual discussions, and encouragement.

Lastly, we acknowledge and appreciate our own dedication, hard work, and perseverance. This thesis is the result of our efforts, late nights, and continuous learning. We are proud of the accomplishments we have achieved and the knowledge we have gained throughout this research and our academic journey.

May Allah bless everyone who has played a part in our journey and grant us success in all our future endeavors.

Dedications

Our great parents, who never stop giving of themselves in countless ways to us.

To all our family, the symbol of love and giving,

***"Souiyou"** and **"Tekouk"**,*

Our friends who encourage and support us, all the people in our life who touch our heart

Those eagerly awaiting the realization of this success,

We dedicate this work.

ABSTRACT

How can machine learning enhance safety management? This thesis explores the application of machine learning models to enhance safety management practices across diverse industries. The specific focus is on the prediction of incident severity using a comprehensive dataset and the evaluation of various algorithms, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression. Through rigorous analysis, the Random Forest model emerges as the top performer in accurately classifying incident severity.

This research provides valuable insights into the integration of machine learning into safety management, highlighting its potential benefits and associated considerations. By addressing these challenges, we can fully harness the power of machine learning, enabling informed decision-making in safety management and fostering safer work environments. The findings of this study contribute to future research and advancements in the utilization of machine learning for improved safety management practices.

Keywords: Safety management, Risk assessment, Machine learning, Incident severity.

الملخص

كيف يمكن للتعلم الآلي تعزيز إدارة السلامة؟ تستكشف هذه المذكرة تطبيق نماذج التعلم الآلي لتعزيز ممارسات إدارة السلامة في مختلف الصناعات. ينصب التركيز على التنبؤ بخطورة الحادث باستخدام مجموعة بيانات شاملة وتقييم الخوارزميات المختلفة ، بما في ذلك Support Vector Machines (SVM) و K-Nearest Neighbors (KNN) و Random Forest و Logistic Regression. من خلال التحليل الدقيق ، يظهر نموذج Random Forest أفضل أداء في التصنيف الدقيق لشدة الحادث.

يوفر هذا البحث رؤى قيمة حول دمج التعلم الآلي في إدارة السلامة ، مع تسليط الضوء على فوائده المحتملة والاعتبارات المرتبطة به. من خلال مواجهة هذه التحديات ، يمكننا الاستفادة بشكل كامل من قوة التعلم الآلي ، وتمكين اتخاذ قرارات مستنيرة في إدارة السلامة وتعزيز بيئات العمل الأكثر أمانًا. تساهم نتائج هذه الدراسة في الأبحاث المستقبلية والتقدم في استخدام التعلم الآلي لتحسين ممارسات إدارة السلامة.

الكلمات المفتاحية: إدارة السلامة ، تقييم المخاطر ، التعلم الآلي ، خطورة الحوادث.

TABLE OF CONTENTS

CHAPTER 3 – MACHINE LEARNING MODELS

CHAPTER 4 – APPLICATION OF MACHINE LEARNING: PREDICT THE SEVERITY OF INCIDENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under the Curve
CRISP-DM	CRoss-Industry Standard Process for Data Mining
CVL01	Construction project identifier
DNN	Deep Neural Network
E	Experience
EDA	Exploratory Data Analysis
GBM	Gradient Boosting Machines
GFA	Genetic Function Approximation
ISO	International Organization for Standardization
KDD	Knowledge Discovery Databases
K-NN	K-Nearest Neighbors
LR	Logistic Regression
LSA	Latent Semantic Analysis
MAR	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perception
MLR	Multiple Linear Regression
MSE	Mean Squared Error
NLP	Natural Language Processing
NSC	National Safety Council
OSHA	Occupational Safety and Health Administration
P	Performance
PSO	Particle Swarm Optimization
QSAR/QSPR	Quantitative Structure-Activity Relationship/Quantitative Structure-Property Relationship
RF	Random Forest
RFC	Random Forest Classifier
RMSE	Root Mean Squared Error

ROC	Receiver Operating Characteristic
SEMMA	Sample, Explore, Modify, Model and Assess
SGTB	Stochastic Gradient Tree Boosting
SMS	Safety Management System
SVC	Support Vector Machine Classifier
SVM	Support Vector Machine
T	Task
VEC	Vector Error Correction
WHS	Workplace Health and Safety

GENERAL INTRODUCTION

In recent years, there has been an increasing interest in leveraging the power of machine learning models to enhance safety management practices. The ability of these models to analyze vast amounts of data, identify patterns, and make accurate predictions has opened up new possibilities for improving incident prevention, risk assessment, and decision-making in various industries. This thesis focuses on the application of machine learning techniques in safety management, specifically in predicting the severity of incidents.

Safety management plays a vital role in ensuring the well-being of workers and minimizing risks in the workplace. Traditional approaches to safety management often rely on manual processes, which can be time-consuming and prone to human error. With the advent of artificial intelligence and machine learning, there is an opportunity to transform safety management practices by automating processes, extracting valuable insights from data, and enabling proactive decision-making.

The objective of this thesis is to explore the potential of machine learning models in predicting incidents severity, providing safety managers with valuable information to prioritize resources, implement appropriate risk mitigation strategies, and improve overall safety outcomes. By leveraging historical incident records and associated attributes, these models can learn from patterns and relationships to make accurate predictions, assisting safety management professionals in making informed decisions.

To achieve this objective, we have conducted an extensive literature review, examining the current state of research and practice in utilizing machine learning algorithms for safety management. We have investigated various applications of machine learning in risk assessment, explored different implementation approaches across industries, and identified the benefits and challenges associated with integrating machine learning into safety management workflows.

Drawing on this knowledge, we have developed a comprehensive methodology for predicting the severity of incident classes using machine learning models. We have carefully selected and analyzed a relevant dataset, applying exploratory data analysis techniques to gain insights into the data's characteristics. Pre-processing techniques have been employed to clean and transform the data, ensuring its quality and suitability for model training.

To evaluate the performance of different machine learning algorithms, we have implemented and compared several state-of-the-art models, including support vector machines, k-nearest neighbors, random forest classifier, logistic regression, and multilayer perceptron. Evaluation metrics have been employed to assess the models' predictive performance, enabling us to identify the most effective approach for incident severity classification.

Through our research, we have identified both the strengths and limitations of machine learning models in safety management. While these models demonstrate promising results in accurately predicting incident severity, challenges such as data availability and quality, class imbalance, feature selection, and model interpretability need to be addressed for their widespread adoption.

The findings of this thesis contribute to the growing body of knowledge on the application of machine learning in safety management. By harnessing the capabilities of artificial intelligence and machine learning algorithms, safety management professionals can gain valuable insights, improve incident prevention strategies, and enhance decision-making processes. Our work highlights the potential for machine learning models to revolutionize safety management practices and pave the way for safer work environments.

In the subsequent chapters, we will delve into the foundational concepts of safety management, explore different types of machine learning algorithms, detail our methodology for incident severity prediction, present our experimental results, discuss the limitations and challenges encountered, and conclude with recommendations for future research and implementation in safety management. By the end of this thesis, we aim to provide a comprehensive understanding of the potential and implications of machine learning in enhancing safety management practices.

CHAPTER 1 – LITERATURE REVIEW

1.1 INTRODUCTION

The field of safety management has seen significant advancements with the integration of Machine Learning models. Machine Learning techniques have been increasingly applied to various domains, including risk assessment, chemical health and safety, banking risk management, and construction safety. These applications have shown promising results in improving safety practices, predicting risks, and enhancing overall safety performance. This chapter presents a literature review that explores the utilization of Machine Learning models in safety management. It examines relevant studies, methodologies employed, key findings, and their implications for safety management practices. By reviewing the existing literature, this chapter aims to provide a comprehensive understanding of the current state of Machine Learning in safety management and identify areas for further research and improvement.

1.2 USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN SAFETY MANAGEMENT ENHANCEMENT

A comprehensive study conducted by the National Safety Council has revealed the alarming reality of workplace safety issues, with an estimated annual loss of a staggering US $163 billion. Each year, over 4 million individuals seek medical attention due to work-related injuries, while the year 2020 witnessed the tragic deaths of over 4,000 people in workplace accidents that the NSC believes could have been prevented with an improved safety management system.[1]

These statistics demand urgent attention, not just from the health and safety team but also from business leaders, decision-makers, and managers who must actively engage in addressing safety risks. Although there is an increasing awareness regarding the importance of a safer work environment, many safety managers and HR professionals struggle to identify risks and implement adequate preventive measures to avoid incidents.[2]

To accurately predict the occurrence of such incidents and proactively implement corrective and preventive actions, these professionals require access to data. It has become evident that reducing incidents is not solely achieved by focusing on major injuries or accidents but also by documenting and responding to safety observations and near-misses.[1]

Recognizing this imperative, the European Agency for Safety and Health at Work published a policy briefing in 2021 emphasizing the need for health and safety teams to embrace emerging technologies such as Artificial Intelligence (AI) and Machine Learning (ML) to enhance safety processes. These advanced technologies can revolutionize safety management by providing insights, predictions, and automated systems that enable proactive measures to mitigate risks.[1]

By leveraging AI and ML, organizations can analyze vast amounts of data, including incident reports, near-miss data, and safety observations, to identify patterns, detect potential risks, and generate accurate predictions. This empowers safety managers and HR professionals to take timely and targeted actions, preventing accidents and injuries before they occur. Additionally, AI and ML algorithms can continuously learn and adapt from new data, improving the accuracy and effectiveness of safety management systems over time.[1], [2]

1.3 LEVERAGING MACHINE LEARNING TECHNIQUES FOR SAFETY MANAGEMENT (RELATED WORKS)

Machine Learning in Risk Assessment

The article of ***Nicola et al*** [3] suggests a risk assessment approach based on Machine Learning, specifically a deep neural network (DNN) model, which was developed and tested. The study focused on a drive-off scenario involving an Oil & Gas drilling rig and showed reasonable accuracy and suitability of DNN for risk assessment. However, the authors caution that model selection and customization should be carefully carried out. The article highlights the importance of continuous risk assessment, improvement in learning past lessons, and the definition of techniques to process relevant data, which are to be coupled with adequate capability to deal with unexpected events and provide the right support to enable risk management.

DNN and MLR models were trained on risk data and evaluated using metrics for precision, accuracy, and recall. DNN model had higher precision and accuracy but more false negatives than MLR model. MLR model had higher recall.

Machine Learning and Deep Learning in Chemical Health and Safety

The results from the research of **Jiao et al** [4] provide information on the use of Machine Learning and deep learning in chemical health and safety. The articles and books discuss the fundamentals and applications of Machine Learning in chemical safety and health-related model development, including predicting material/molecular properties, chemical discovery, and process conditions. The interdisciplinary studies combining Machine Learning and chemical health and safety have demonstrated their advantages in identifying trends and prediction assistance, which can greatly save manpower, material resources, and financial resources. The search results also provide a systematic review of techniques and applications of Machine Learning and deep learning in chemical health and safety, as well as recent applications of Machine Learning and artificial intelligence approaches in different areas of toxicology.

Methodologies used are QSAR/QSPR model development with regression algorithms (MLR, PSO, GFA, SVM, ANN), integration of ML algorithms and dispersion models, and fault detection and diagnostics using neural networks. QSAR/QSPR models have high prediction accuracy, SVM and ANN are extensively used. ML algorithms, particularly ANNs, improve consequence prediction accuracy, including gas dispersion and source terms estimation.

Machine Learning in Banking Risk Management

The article of **Leo at al** [5] discusses the use of Machine Learning in banking risk management. The article provides a literature review on the topic, covering theoretical background, risk management at banks, and Machine Learning. The authors suggest that Machine Learning can improve insight into client preferences, risk management, fraud detection, and credit underwriting. However, poorly managed machine-learning models could expose a bank to reputational, ethical, regulatory, and financial risks. The article recommends that banks refine their model governance and model validation practices to combat these risks.

ML models used are Classification algorithms, Clustering analysis, Bayesian networks, Decision Trees, and Neural Networks. Classification algorithms like SVM have been effective in credit scoring and predicting default probabilities. ML techniques outperform traditional statistical methods in classification and prediction. ML models are still vulnerable to biases and issues similar to traditional statistical methods.

Machine Learning in Construction Safety

The article of **Cao and Goh** [6] describes how temporal analysis techniques can be applied to improve the safety management of construction data. Various time series (TS) methods were adopted for identifying the leading indicators or predictors of construction accidents. The data set used herein was obtained from a large construction company that is based in Singapore and contains safety inspection scores, accident cases, and project-related data collected from 2008 to 2015. Five projects with complete and sufficient data for temporal analysis were selected from the data set. The filtered data set contained 23 potential leading indicators, predictors or input variables of accidents. TS analyses were used to identify suitable accident predictors for each of the five projects. Subsequently, the selected input variables were used to develop three different TS models for predicting accident occurrences, and the vector error correction model was found to be the best model. It had the lowest root mean squared error value for three of the five projects analyzed. This study provides insights into how construction companies can utilize TS data analysis to identify projects with high risk of accidents.

The study employed time series (TS) approaches to analyze construction safety and forecast the number of accidents for each project. The ANN (Artificial Neural Network) model is the only machine learning approach used. The VEC model was the most accurate for accident forecasting. The ARIMA model only considers past accidents and lacks consideration of other variables. The Artificial Neural Network (ANN) model had lower RMSE than VEC and ARIMA for project CVL01.

1.4 APPLICATIONS OF ML METHODS FOR RISK ASSESSMENT

Risk assessment is a crucial aspect of safety management in various industries, including manufacturing, construction, healthcare, mining, energy, oil and gas, and many others. It involves identifying potential hazards, analyzing their likelihood and consequences, and determining the tolerances for such events. In complex systems engineering, sophisticated risk assessments are often made within safety engineering and reliability engineering when it concerns threats to life, natural environment, or machine functioning. The agriculture, nuclear, aerospace, oil, railroad, and military industries have a significant need for risk assessment to ensure safety and improve outcomes. Machine Learning methods have become increasingly popular in engineering risk assessment as they often aid the risk identification phase during risk assessments.

The article *of* ***Hegde and Rokseth*** [7]**,** aims to present a structured review of publications utilizing Machine Learning methods to aid in engineering risk assessment. The study aims to identify the most commonly used Machine Learning techniques for engineering risk assessment and highlight areas for future research. The study also provides valuable insights into the types of data used, data acquisition approaches, types of machine learning implementation, the integration of machine learning across risk assessment phases, and the utilization of machine learning methods in various industries. All of these aspects will be discussed in the following sections:

1.4.1 Data Analysis and Acquisition

1.4.1.1 Exploring Textual and Numerical Data Analysis

This study explores the application of Machine Learning techniques in analyzing textual and numerical data to enhance risk analysis and safety measures in various industries. Textual data analysis involves techniques such as NLP, LSA, and ML models like RF and SGTB to extract meaningful information from safety incident narratives. It enables the identification of contributing factors, injury classification, and standardized coding. Numerical data analysis utilizes algorithms such as ensemble, clustering, and classification models to predict crash severity, assess risk factors, and analyze safety performance in domains like transportation and construction. The findings emphasize the potential of Machine Learning in uncovering hidden patterns and improving safety measures through data-driven approaches.

1.4.1.2 Types Of Data Acquisition Approaches

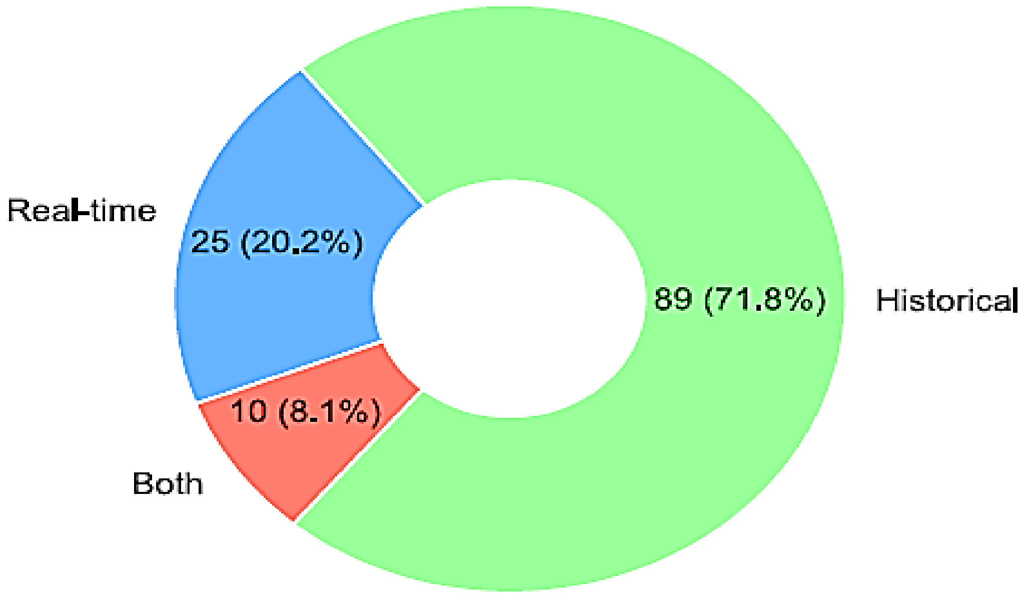


Figure 1-1: Input Data Acquisition Approaches Utilized to Build the Model

Figure 1-1 illustrates the various approaches to data acquisition discussed in the literature. Here's what each approach means:

Historically Available Data: This approach involves using existing data that has been collected and recorded in the past. It typically refers to datasets that were collected and made available for research purposes or were previously generated for other applications. In the context of the study mentioned, 89 articles utilized historically available data to develop machine learning-based risk assessment models. This means that researchers relied on pre-existing datasets to train their models and make predictions about risk assessment.

Real-Time Sensor Data: This approach involves collecting data in real-time from various sensors or monitoring devices. These sensors could be embedded in physical systems or environments and provide continuous or periodic measurements of specific parameters. In the study mentioned, 25 articles used real-time sensor data as inputs for their machine learning-based risk assessment models. This means that researchers collected data directly from sensors to feed into their models, enabling them to make risk assessments based on current and up-to-date information.

Combination of Historically Available Datasets and Real-Time Sensor Data: This approach involves utilizing both historically available data and real-time sensor data together. In the mentioned study, 10 articles (review articles) describe using a combination of these two types of data for developing machine learning-based risk assessment models. By leveraging historical data along with real-time sensor data, researchers can create more comprehensive models that capture both the historical trends and the current state of the system being analyzed. This approach allows for a more holistic understanding of risk assessment.

1.4.2 Types Of Implementations

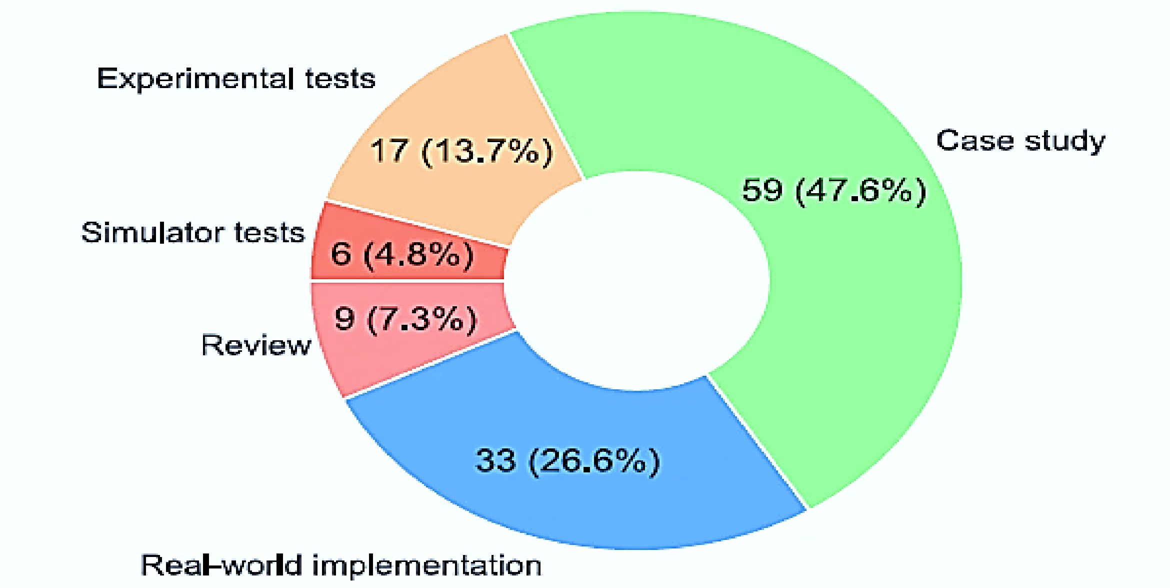


Figure 1-2: Researches Classified by Type of Implementation.

The researches are further categorized based on the type of implementation. Here's what each type of implementation means:

Case Studies: Case studies refer to 59 articles that have used real-world scenarios or specific examples to validate the proposed machine learning models for risk assessment. These articles likely presented and analyzed actual cases or situations where the machine learning models were applied and assessed their effectiveness in addressing risk-related challenges.

Real-World Implementation: This type includes 33 articles that describe the development of machine learning-based risk assessment tools or applications specifically designed for real-world use. These articles likely focused on creating practical solutions that can be implemented in real-world settings, such as industries, organizations, or systems, to assess and manage risks effectively using machine learning techniques.

Experimental Tests: This type comprises 17 articles that have validated their machine learning models for risk assessment through experimental setups. These articles likely conducted controlled experiments in laboratory or controlled environments to evaluate the performance and accuracy of their models. The experimental tests could involve simulated scenarios or controlled data collection to assess the models' capabilities and measure their effectiveness in risk assessment tasks.

Simulator Tests: This type consists of 6 articles that have used a simulator environment to test the developed machine learning models for risk assessment. These articles likely utilized computer-based simulations to create virtual environments that mimic real-world conditions or systems. The models were then tested within these simulated environments to evaluate their performance and assess their suitability for risk assessment applications.

1.4.3 Machine Learning Methods Used in Risk Assessment

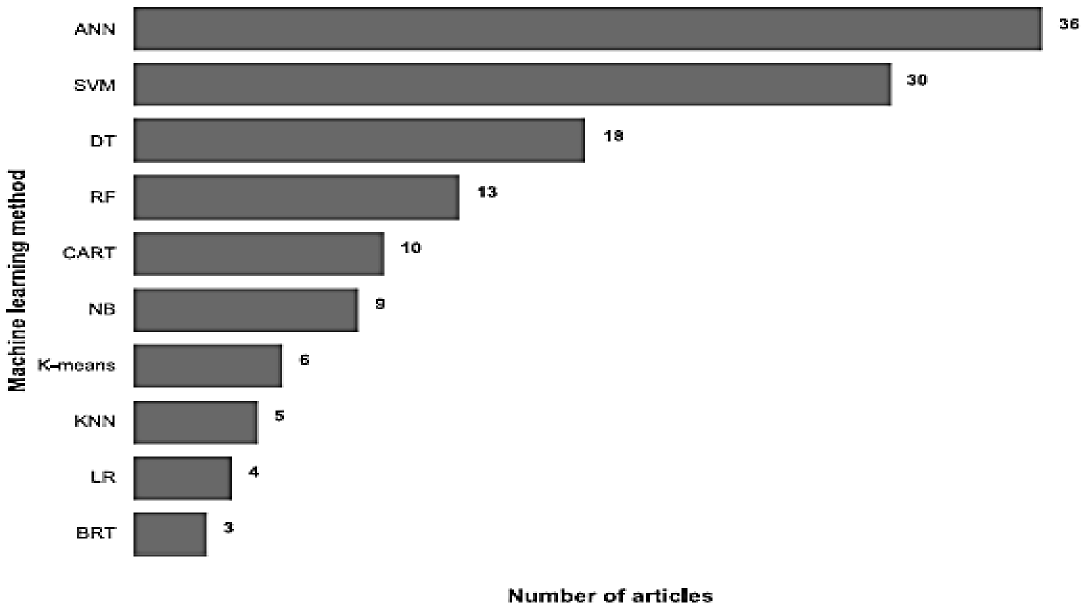


Figure 1-3: Frequently Used Machine Learning Methods in Risk Assessment.

Figure 1-3 presents the top ten learning algorithms commonly employed for risk assessment. Among the reviewed researches, artificial neural networks (ANNs) are utilized for risk assessment in 36 different researches, making it the most frequently used algorithm. Support vector machines (SVMs) closely follow with 30 applications, and decision trees (DTs) are used in 18 applications. Overall, a total of 68 different Machine Learning methods are applied across the selected literature to perform risk assessment.

Artificial Neural Networks (ANNs) are the most frequently used Machine Learning method in risk assessment. ANNs are favored due to their ability to establish meaningful relationships between input and output variables using non-linear mathematical equations. ANNs offer advantages such as not requiring extensive statistical training and being capable of detecting all possible interactions between input and output variables. These factors contribute to their popularity in the selected literature. However, ANNs have disadvantages that should not be overlooked. They are prone to overfitting, meaning their relationships may be specific to a particular dataset. Additionally, ANNs are considered "black box" methods, unable to explicitly identify causal relationships between input and output variables. Trusting the output of such methods can be challenging.

Overall, the conclusions highlight the dominance of ANNs in risk assessment, the limitations of these methods, and the challenges encountered in conducting a comprehensive literature review due to terminology variations and metadata issues.

1.4.4 Machine Learning Applications in Risk Assessment Phases

Figure 1-4 displays a Venn diagram that depicts the classifications of the researches. Among them, 34 articles primarily concentrate on utilizing Machine Learning for risk identification. 7 researches center around using Machine Learning for risk analysis, while only 1 article pertains to the use of Machine Learning for risk evaluation activities. Furthermore, 19 researches focus on all three phases of risk assessment, namely risk identification, analysis, and evaluation.

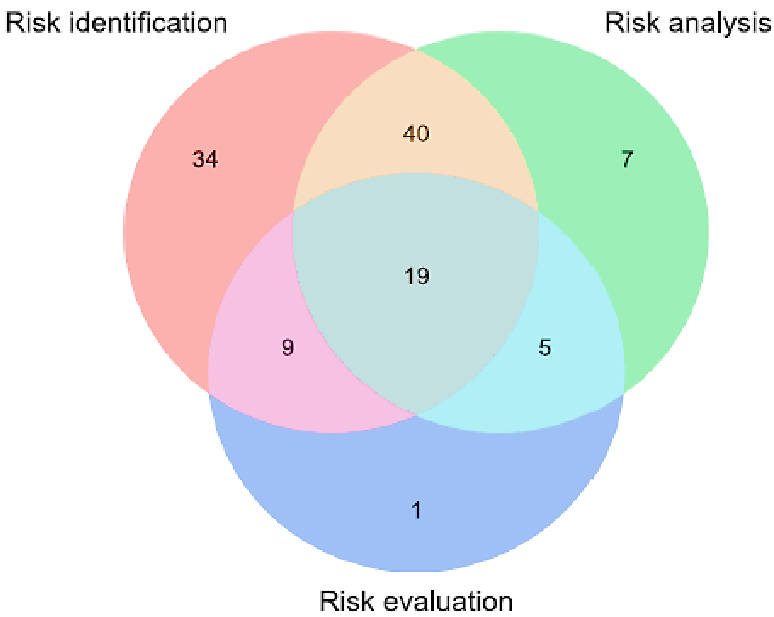


Figure 1-4: Classification of Researches with Respect to The Three Risk Assessment Phases

The analysis presented in ***Figure 1-4*** reveals an imbalance in the focus of Machine Learning applications within risk assessment. Specifically, there is a disproportionate emphasis on the risk identification and analysis phases, with limited attention given to the risk evaluation phase.

Surprisingly, out of the 124 articles reviewed, only one article ***(Curiel-Ramirez et al., 2018)***[8] focuses solely on risk evaluation. This article demonstrates how a vision-based system can assist in evaluating an ideal steering angle for an autonomous road vehicle. This indicates a lack of dedicated research and development in the area of risk evaluation using Machine Learning methods.

It can be argued that the risk identification phase is relatively easier to address compared to the applications requiring risk analysis or evaluation. An example provided is the work of ***(Zhou et al. (2018)***)[9], where a vision-based Machine Learning algorithm is developed to identify critical parts during the inspection of railway locomotives. In this case, the main objective is to identify the object of interest rather than taking evident actions or making evaluative judgments based on the identified risks.

These findings highlight a potential gap in the application of Machine Learning within risk assessment, with a need for more emphasis on developing methods and models for risk evaluation. While Machine Learning algorithms can effectively aid in risk identification and analysis, further research is required to enhance their capabilities in supporting risk evaluation processes, enabling organizations to make informed decisions and take appropriate actions based on the assessed risks.

1.4.5 Machine Learning Applications in Various Industries

Machine Learning techniques utilized in various industries for conducting risk assessment. It is evident from the findings that the aviation industry has implemented Machine Learning in 6 applications, healthcare industry in 3 applications, workplace safety industry in 2 applications, oil and gas industry in 2 applications, and the energy industry in 2 applications. The automotive industry takes the lead with 29 applications, while the construction industry follows closely with 20 applications of Machine Learning for risk assessments. Furthermore, a total of 17 diverse industries have ventured into utilizing Machine Learning for risk assessment purposes.

The review also identified 10 researches that propose generic methods applicable across industries, without specific industry limitations. The study reveals a wide array of Machine Learning algorithms being employed to address risk assessment-based problems.

1.4.6 Types Of Implementations of Machine Learning in Various Industries

Table 1-1: Type of Implementation in Each Industry.

Industry	Types of implementations				
	Case study	Real-world implementation	Experimental tests	Simulator tests	Review
Automotive	6	12	3	5	3
Aviation	4	1	1	0	0
Construction	10	6	4	0	0
Energy	2	0	0	0	0
Healthcare	2	0	1	0	0
Oil & Gaz	1	0	1	0	0
Workplace	2	0	0	0	0

Table 1-1 expands upon the analysis conducted in ***Figure 1-2*** by examining the Research dataset categorized by industry. In ***Table 1-1***, the darker green shaded cells indicate a higher number of contributing researches, while the red shaded cells indicate a lower number of contributions. The findings reveal that researches pertaining to the automotive industry frequently validate their models through real-world implementations. Furthermore, simulator-based implementations are exclusively utilized by researches focusing on the automotive industry. Conversely, researches associated with the construction industry validate their findings using case studies, real-world implementations, or experimental tests.

While the industries of aviation, healthcare, oil and gas, and workplace safety focusing on case studies to validate their findings.

1.4.7 Input Data Used to Build Machine Learning Model in Each Industry

Table 1-2: Collection of Input Data Used to Build Machine Learning Models for Risk Assessment in Different Industries.

Industry	Input data used in literature to train the Machine Learning models for risk assessment
Automotive	'merging traffic data from a work zone in Singapore', 'textual accident description', 'car sensor data', 'steering angle', 'vehicle sensors', 'injury severity', 'gender', 'seatbelt', 'cause of crash', 'location type', 'lighting condition', 'weather conditions', 'road surface condition', 'occurrence', 'shoulder type', 'accident location', 'environmental conditions', 'primary cause', 'injury levels of occupants', 'number of lanes', 'horizontal curvature', 'vertical grade', 'temperature', 'humidity', 'precipitation', 'wind speed', 'cloudiness', 'accident data', 'roadway characteristics',...
Construction	'Output from monte carlo simulations', 'textual injury reports', 'workplace hazard information', 'accident and incident data from interviews', 'incident reports', 'safety events', 'accident data', 'video', 'motion detection', 'predictions from mathematical models', 'results from field inspections', 'concrete thickness', 'employee status', 'employee company', 'occupation', 'occupation group', 'seniority', 'work demands', 'body part discomfort', 'psychosocial needs', 'work rhythm', 'work extension', 'work life balance', 'workplace risk assessment', 'accident data', 'augmented images',…
Aviation	'Aviation incident reports', 'aviation accident data', 'incident reports', 'flight-data records',…
Healthcare	'Physiological signals', 'accelerometer reading', 'gyroscope reading', 'magnetometer reading',…
Workplace safety	'Textual accident narratives'
Oil & gas	'Accident data', 'underwater vehicle sensors'
Environmental engineering	'Process measurements', 'process parameters', 'source risk index', 'air risk index', 'water risk index', 'target vulnerability'
Energy	'Generated operating points', 'voltage signals'

Table 1-2 provides a collection of input data used to build Machine Learning models for risk assessment in different Industries.

The type of input data used in Machine Learning-based risk assessment can vary based on the industry or application. This suggests that different data sources are utilized by authors when developing their Machine Learning models.

The availability and quality of data sources can significantly impact the effectiveness and applicability of the proposed Machine Learning models. For instance, in the case of predicting road accident severity, the developed model's performance may heavily rely on the quality and availability of the input data source. If the data source is limited or of substandard quality, it may not be feasible for other researchers to adopt or reuse the proposed method.

The outcome of a Machine Learning model is highly dependent on the availability of the input data source. This dependency can hinder the adoption, repeatability, and benchmarking of proposed methods in the literature.

1.5 CONCLUSION

Safety management is of utmost importance in various domains to prevent accidents, mitigate risks, and ensure the well-being of individuals. With the advancements in Machine Learning and data-driven approaches, there is a significant opportunity to enhance safety practices and improve overall safety outcomes. By leveraging large datasets, sophisticated algorithms, and predictive models, Machine Learning can help identify potential hazards, predict risks, and develop proactive strategies to prevent incidents. The integration of Machine Learning into safety management processes enables organizations to analyze vast amounts of data, identify patterns, and make data-driven decisions that can have a profound impact on safety outcomes. This combination of human expertise and Machine Learning capabilities empowers safety professionals to implement proactive measures, anticipate risks, and foster a culture of safety within their organizations. Ultimately, the adoption of Machine Learning in safety management holds great potential for achieving higher levels of safety, reducing accidents, and protecting the well-being of individuals in various domains.

CHAPTER 2 – SAFETY MANAGEMENT

2.1 INTRODUCTION

This chapter focuses on safety management and its integral components, including risk assessment, safety concepts, and the integration of Machine Learning (ML) in risk assessment. We begin by defining safety management and highlighting its importance in various industries. We then explore key safety concepts such as hazard identification, incident reporting and investigation, and safety culture. Moving on to risk assessment, we discuss its process, methods, and limitations. Lastly, we examine how ML can enhance risk assessment by utilizing advanced data analysis and predictive modeling techniques. Through this chapter, readers will gain a comprehensive understanding of safety management and its key components, setting the stage for exploring ML's role in risk assessment.

2.2 FOUNDATIONS OF SAFETY MANAGEMENT

2.2.1 What Is Safety Management?

Safety management is a crucial organizational function that involves the systematic Application of a set of principles, frameworks, processes, and measures to prevent accidents, Injuries, and other adverse consequences that may arise from using a product or service. Its Main purpose is to assist managers in designing and implementing operational systems in a Way that ensures safety and minimizes risks.

Safety management involves the prediction of potential system deficiencies before Errors occur, and professional analysis of safety occurrences to identify and correct any Deficiencies. It requires a systematic approach to managing safety, which includes the Necessary organizational structures, policies, and procedures to ensure that all safety risks Have been identified, assessed, and satisfactorily mitigated. The ultimate goal of safety Management is to ensure the safety of all individuals involved in the operation, while also Enhancing productivity and minimizing liabilities.[10]

2.2.2 The Importance of Safety Management

Safety management is crucial for creating a work environment that is free from injuries and accidents. It not only attracts employees but also enhances their satisfaction and productivity. Employers have the responsibility to protect their employees and ensure a safe working environment. Workplace safety is essential for the well-being of both employees and employers. It helps prevent human casualties, provides assurance to employees, and addresses occupational safety and health risks.[11], [12]

2.2.3 Benefits of Safety Management

Implementing effective safety measures brings several benefits. A safe and healthy work environment promotes productivity, reduces operational costs, and improves the quality of products and services. It enhances the wellness of employees, leading to better health and efficiency. A safe workplace results in fewer accidents, less downtime for investigations, and reduced costs for worker's compensation. It also prevents damage to industrial equipment, minimizing expenses and increasing profits. Furthermore, a focus on employee safety boosts confidence and comfort, reduces absenteeism, and improves employee focus and task performance.[12]

2.2.4 Base Components Required by OSHA for Safety Management

The Occupational Safety and Health Administration (OSHA) has established federal standards for process safety management in various industries. To ensure safety management, OSHA requires businesses to adhere to certain minimum elements, including:

- Developing and maintaining safety information on workplace processes and chemical hazards.
- Conducting industry risk assessments, identifying potential sources of hazardous releases, and evaluating the effects on employee safety.
- Consulting with experts and representatives to develop risk assessments and accident prevention plans.
- Establishing response systems for risk assessment findings, focusing on mitigation, prevention, and emergency response.
- Regularly reviewing the workplace risk assessment and response system.
- Developing standard operating procedures, including working restrictions and safety practices.
- Providing written operating information and training to employees, emphasizing hazards and safety practices.
- Ensuring that contract workers and contractors receive adequate information and training.
- Training employees and contractors in emergency response procedures.
- Establishing a quality assurance program for parts and maintenance materials.
- Ensuring compliance with process safety management regulations to protect employees from preventable harm.[11]

2.3 SAFETY CONCEPTS

2.3.1 Hazard Identification

One of the fundamental causes of workplace injuries, illnesses, and incidents is the failure to identify or acknowledge existing or potential hazards. A proactive and continuous process of hazard identification is a crucial component of an effective safety and health program.[13]

2.3.1.1 What is Hazard Identification?

Hazard identification is an integral part of the broader risk assessment process, which aims to determine if a situation, item, or entity carries the potential to cause harm. It involves:

- Identifying hazards and risk factors that have the potential to cause harm (hazard identification).

- Analyzing and evaluating the level of risk associated with each identified hazard (risk analysis and risk evaluation).
- Determining appropriate measures to eliminate the hazard or control the risk, especially when complete elimination is not feasible (risk control).

The ultimate objective of hazard identification is to identify and document potential hazards within the workplace. Conducting the inspection as a team, involving individuals familiar with the work area as well as those who are not, can provide a comprehensive perspective, leveraging both experience and fresh eyes. By actively identifying hazards, organizations can take proactive steps to mitigate risks and create a safer work environment.[13], [14]

2.3.1.2 Process of Hazard Identification

The process of hazard identification involves the following action items:

1. **Collect Existing Information About Workplace Hazards**
   Gather and review available information, such as equipment manuals, safety data sheets, incident reports, workers' compensation records, and input from workers.

2. **Inspect The Workplace for Safety Hazards**
   Conduct regular inspections of all operations, equipment, work areas, and facilities, involving workers in the inspection process. Document the inspections and address identified hazards promptly.

3. **Identify Health Hazards**
   Identify chemical, physical, biological, and ergonomic hazards that workers may be exposed to. Review safety data sheets, product labels, exposure monitoring results, medical records, and conduct quantitative exposure assessments.

4. **Conduct Incident Investigations**
   Thoroughly investigate workplace incidents, injuries, illnesses, and near misses to identify root causes and prevent future occurrences. Develop a clear plan, involve a trained investigative team, analyze root causes, and communicate the investigation results.

5. **Identify Hazards Associated with Emergency and Nonroutine Situations**
   Identify foreseeable emergency scenarios and nonroutine tasks and develop plans and procedures to respond to hazards in such situations. Consider fires, chemical releases, startups, maintenance activities, and other potential emergencies.

6. **Characterize the Nature of Identified Hazards, Identify Interim Control Measures, and Prioritize Hazards for Control**
   Assess the hazards identified, understand potential outcomes and worker exposures, implement interim control measures to protect workers temporarily, and prioritize hazards based on severity, likelihood, and number of exposed workers.[13]

2.3.2 Incident Reporting and Investigation

Under clause ***10.2*** of ***ISO 45001***, companies need to establish, implement and maintain a process for the investigation and reporting of incidents. Taking corrective action, the company can be better equipped to react to future incidents and manage them effectively.[15]

2.3.2.1 What is Incident Reporting and Investigation?

Incident reporting and investigation involve the examination and analysis of worksite incidents, including fatalities, injuries, illnesses, and close calls. This process provides employers and workers with an opportunity to identify hazards within their operations and deficiencies in their safety and health programs. It focuses on understanding the root causes of incidents rather than assigning blame, aiming to implement corrective actions that prevent future occurrences. By involving a collaborative effort between managers and employees, incident reporting and investigation foster workplace morale and productivity by showcasing an employer’s commitment to a safe and healthy work environment. It goes beyond immediate causes, delving into underlying factors and systemic changes required for prevention.[16], [17]

2.3.2.2 Importance of Incident Reporting and Investigation

Incident reporting and investigation play a vital role in maintaining workplace safety and preventing future incidents. An incident is an unplanned event that results in injury, near misses, or dangerous occurrences. Investigating incidents involves a comprehensive examination of all contributing factors and causes to uncover the chain of events leading to the incident. Causes are typically categorized as direct, indirect, and root causes.[16], [17]

Direct causes are the immediate actions triggered by the root cause. Indirect causes are the hidden or latent factors that contribute to the incident. Root causes represent the fundamental reason behind the event and may not be immediately apparent. Understanding these causes is essential for implementing effective preventive measures.[16], [17]

By conducting thorough incident reporting and investigation, organizations can foster a culture of safety, address deficiencies, and identify areas for improvement in their safety protocols. This proactive approach promotes continuous learning, risk mitigation, and the overall well-being of workers. It allows for the implementation of necessary changes to prevent similar incidents in the future, safeguarding the workforce and enhancing the overall safety and productivity of the workplace.[16], [17]

2.3.2.3 Process of Incident Reporting and Investigation

To prepare for an incident investigation and subsequent report, we will need several documents. Once we have determined the level of investigation needed, our procedure should follow the following steps:

1. **Form an Investigation Team**
- Allocate resources and ensure team competency.
2. **Determine Your Incident Statement**
- Agree on investigation format and formulate a comprehensive incident statement.
3. **Gather the Information**
- Use a systematic approach and checklist.
- Address critical questions and collect witness statements, photographs, measurements, and review documents.
4. **Analyze the Evidence Gathered**
- Consider physical, human, and administrative factors.
- Evaluate equipment condition, training, supervision, policies, and procedures.
5. **Conduct Root Cause Analysis**
- Identify contributing factors in sequential order.
- Determine immediate, underlying, and root causes.
6. **Write an Investigation Report**
- Include essential information such as date, time, place, and investigator's name.
- Describe the incident and its causes.
- Document resources used and corrective actions taken.
- Provide a concise, clear, factual report.[15]

2.3.3 Safety Culture

A safety culture refers to the organizational culture that emphasizes the importance of safety beliefs, values, and attitudes, which are shared by the majority of individuals within the company or workplace. It can be described as "the way we do things around here." When a positive safety culture is established, it can lead to improved workplace health and safety (WHS) and overall organizational performance.[18]

Organizational safety culture encompasses various aspects, often summarized as "how people in the company behave when no one is looking." It comprises several subcultures that contribute to the overall safety culture of the organization.[19]

2.3.3.1 Subcultures Within Safety Culture

1. **Flexible culture**

Similar to the ability to anticipate system requirements, a flexible culture allows for adapting and reconfiguring the system to handle the pressures it faces.

2. Reporting culture

Similar to the ability to monitor, individuals within the organization feel entirely comfortable reporting threats, errors, undesirable conditions, and near-misses. These reports serve as crucial information for managers to assess the system's resilience levels and areas where resilience may be compromised.

3. Informed culture

Similar to the ability to respond, all components of the system possess the necessary knowledge of their roles and expectations within the larger system.

4. Learning culture

Similar to the ability to learn, the system can reconfigure itself based on lessons learned from past experiences.

5. Just culture

People working in the system trust that they will be treated fairly in case of unsafe acts. The acceptance of the New View of error is key to establishing a successful just culture.[19]

2.3.3.2 Crucial Culture Actions for Fostering Positive Safety Culture

The Construction Safety Competency Framework has identified nine crucial culture actions that are essential for fostering a positive safety culture. These actions can be easily implemented by companies of any size and often require minimal financial investment. Below are the detailed descriptions of each of the nine culture actions.[18]

Culture action 1: Communicate company values

Clearly communicate and reinforce the safety values and beliefs of the organization through various channels, such as policy statements, posters, and regular communication methods.

Culture action 2: Demonstrate leadership

Lead by example and consistently convey the importance of work health and safety through active engagement, wearing personal protective equipment, conducting inspections and risk assessments, and facilitating toolbox talks.

Culture action 3: Clarify required and expected behaviors

Clearly communicate the specific behaviors and actions expected of employees, addressing inappropriate behaviors and reinforcing appropriate ones.

Culture action 4: Personalize safety outcomes

Emphasize the personal impact of risks, injuries, and fatalities, and highlight the importance of individual responsibility for safety and health. Make safety relevant and emotional for individuals.

Culture action 5: Develop positive safety attitudes

Foster attitudes and beliefs that support safe behavior by encouraging employees to challenge unsafe behaviors, recognize positive attitudes, and establish shared values centered around safety.

Culture action 6: Engage and own safety responsibilities and accountabilities

Increase individual involvement and ownership in the safety management process by promoting understanding, proactive hazard identification, and risk management. Foster engagement through employee input, relationship-building, and trust.

Culture action 7: Increase hazard/risk awareness and preventive behaviors

Enhance individuals' understanding of work health and safety outcomes associated with their decisions and actions. Facilitate meaningful two-way communication to generate preventive behaviors among employees and contractors.

Culture action 8: Improve understanding and effective implementation of safety management systems

Increase knowledge of hazard management and safety processes by involving senior management in the actual work, promoting identification of unsafe practices, continuous improvement, problem-solving, and adherence to safety standards.

Culture action 9: Monitor, review, and reflect on personal effectiveness

Continuously gather feedback on the effectiveness of culture actions and safety-related behaviors through various sources of information. Use this feedback to fine-tune and improve safety leadership and the implementation of other culture actions. Foster positive relationships with managers and the workforce to promote a supportive safety culture.[18]

2.4 SAFETY MANAGEMENT SYSTEM (SMS)

A Safety Management System (SMS) is a program designed to promote safety and Reduce risk to employees and the public in all operations[20]. It is a formalized organization-wide approach to helping a business achieve an acceptable level of workplace safety[21]. The Main goal of safety and health programs is to prevent workplace injuries, illnesses, and Deaths, as well as the suffering and financial hardship these events can cause for workers, Their families, and employers.[22]

The four components of a safety management system are management commitment and Employee involvement, worksite analysis, hazard prevention and control, and safety and Health training[23]. Safety management systems are implemented in various industries, Including aviation, to manage safety risk in the workplace[24]. Effective safety management Systems prevent injuries, process failures, and improve the long-term profitability of Organizations.[20]

Figure 2-1: The Four Components of a Safety Management System

2.4.1 Management Commitment and Employee Involvement

This element involves the active participation and support of management in establishing a safety policy, allocating resources, and engaging employees in the development and implementation of safety programs. It emphasizes the importance of creating a safety culture throughout the organization.[25]

2.4.2 Worksite Analysis

This element focuses on identifying and assessing hazards in the workplace. It includes evaluating physical, chemical, and biological hazards that may pose risks to employees' safety and health. Worksite analysis helps in understanding the potential risks and taking appropriate measures to address them.[20]

2.4.3 Hazard Prevention and Control

This element encompasses the development and implementation of procedures and policies to eliminate or control hazards in the workplace. It involves implementing engineering controls, administrative controls, and personal protective equipment to mitigate risks and create a safer working environment.[26]

2.4.4 Safety And Health Training

This element emphasizes providing employees with the necessary knowledge and skills to identify and control hazards. Training programs help raise awareness about safety practices, procedures, and regulations, empowering employees to actively contribute to a safer workplace.[21].

Effective safety management systems are designed to manage safety risk in the workplace and can be created to fit any business type and/or industry sector.[27]

2.5 RISK ASSESSMENT (AS STATED IN ISO 31000)

2.5.1 Overview

Risk assessment is the overall process of identifying, analyzing, and evaluating risk. It is important that risk assessment is conducted systematically, iteratively, and collaboratively, relying on the knowledge and opinions of stakeholders. The best available information should be used, supplemented if necessary by further investigation.[28]

2.5.2 Risk Assessment Phases

2.5.2.1 Risk Identification

The purpose of risk identification is to search for, recognize, and describe risks that may help or hinder an organization in achieving its objectives. It is essential that the information used for risk identification is relevant, appropriate, and up to date.

The organization can use a range of techniques to identify uncertainties that may impact one or more objectives.

The following factors and their relationships should be considered:

- Tangible and intangible risk sources.
- Causes and events.
- Threats and opportunities.
- Vulnerabilities and capabilities.
- Changes in the external and internal context.
- Indicators of emerging risks.
- Nature and value of assets and resources.
- Consequences and their impact on objectives.
- Limitations of knowledge and reliability of information.
- Time-related factors.
- Biases, assumptions, and beliefs of involved individuals.

The organization should identify risks, whether their sources are under its control or not. It should be noted that there may be multiple types of outcomes with various tangible or intangible consequences.[28]

2.5.2.2 Risk Analysis

The purpose of risk analysis is to understand the nature of the risk and its characteristics, including the level of risk, if applicable. Risk analysis involves detailed consideration of uncertainties, risk sources, consequences, likelihood, events, scenarios, control measures, and their effectiveness. An event can have multiple causes and consequences and affect multiple objectives.

Risk analysis can be conducted at different levels of detail and complexity depending on the purpose of the analysis, availability and reliability of information, and available resources. Analysis techniques can be qualitative, quantitative, or a combination thereof, depending on the circumstances and intended use. Risk analysis should take into account factors such as:

- Likelihood of events and consequences.
- Nature and significance of consequences.
- Complexity and interconnections.
- Time-related factors and volatility.
- Effectiveness of existing control measures.
- Levels of sensitivity and confidence.

Risk analysis can be influenced by differences in opinions, biases, risk perceptions, and judgments. Additional influences include the quality of information used, assumptions and exclusions made, any limitations of techniques, and how they are implemented. These influences should be considered, documented, and communicated to decision-makers.

Extremely uncertain events can be difficult to quantify. This can be problematic when analyzing events with severe consequences. In such cases, the use of a combination of techniques generally leads to a deeper understanding.

Risk analysis provides data for assessing risk, making decisions on whether or not to treat it, and how to do so, as well as selecting the most effective treatment strategy and methods. The results provide information for decision-making when choices need to be made and options involve different types and levels of risk.[28]

2.5.2.3 Risk Evaluation

The purpose of risk evaluation is to lead to more informed decisions. Risk evaluation involves comparing the results of risk analysis to established risk criteria to determine if further action is required. This can result in the decision to:

- Take no further action.
- Explore risk treatment options.
- Conduct further analysis to better understand the risk.
- Maintain existing risk control measures.
- Review objectives.

Decisions should consider a broader context and the actual and perceived consequences for external and internal stakeholders. The outcome of risk evaluation should be recorded, communicated, and validated at appropriate levels within the organization.[28]

2.5.3 Purpose And Benefits

As stated in ***ISO 31010***, the purpose of risk assessment is to provide evidence-based information and analysis that can support decision-making processes regarding the treatment of specific risks and the selection of options. In other words, risk assessment aims to gather and analyze data about risks in order to provide a foundation for informed decision-making.

There are several benefits associated with performing risk assessment, as outlined in ***ISO 31010*** [29]:

Understanding the risk and its potential impact upon objectives: Risk assessment helps stakeholders gain a clear understanding of the risks they face and the potential consequences these risks may have on their objectives. This understanding is crucial for developing effective risk management strategies.

Providing information for decision makers: Risk assessment provides decision makers with valuable information regarding the risks involved in various options or courses of action. This information enables them to make well-informed decisions that consider the potential risks and benefits associated with each option.

Contributing to the understanding of risks for treatment option selection: Risk assessment helps in the evaluation and comparison of different treatment options for managing risks. By assessing and quantifying risks, decision makers can select the most appropriate treatment option that effectively mitigates or controls the identified risks.

Identifying important contributors to risks and weak links in systems and organizations: Risk assessment enables the identification of key factors or contributors that significantly influence the occurrence or severity of risks. It also helps identify weak links or vulnerabilities in systems and organizations, allowing for targeted risk management efforts to strengthen these areas.

Comparing risks in alternative systems, technologies, or approaches: Risk assessment allows for the comparison and evaluation of risks associated with different systems, technologies, or approaches. This comparison assists in selecting the option that presents the most favorable risk profile.

Communicating risks and uncertainties: Risk assessment provides a structured framework for communicating risks and uncertainties to stakeholders. This promotes transparency and facilitates effective risk communication among decision makers, risk owners, and other relevant parties.

Assisting with establishing priorities: Risk assessment helps in prioritizing risks based on their likelihood and potential impact. This prioritization ensures that resources and efforts are allocated to address the most significant risks first.

Contributing towards incident prevention based upon post-incident investigation:

By analyzing risks and their underlying causes, risk assessment aids in identifying areas for improvement and implementing preventive measures. It helps organizations learn from past incidents and reduce the likelihood of similar incidents occurring in the future.

Selecting different forms of risk treatment: Risk assessment provides the necessary information to support the selection of appropriate risk treatment strategies. It assists in determining whether risks should be avoided, mitigated, transferred, or accepted based on their assessment results.

Meeting regulatory requirements: Risk assessment helps organizations meet regulatory and legal obligations by providing a systematic approach to identify, assess, and manage risks. It ensures compliance with relevant standards, regulations, and industry best practices.

Providing information to evaluate risk acceptance against pre-defined criteria:

Risk assessment offers a basis for evaluating whether a risk should be accepted based on predefined criteria. This evaluation helps decision makers determine whether the potential benefits outweigh the associated risks and make informed decisions about risk acceptance.

Assessing risks for end-of-life disposal: Risk assessment plays a role in evaluating the risks associated with the disposal of products, systems, or assets at the end of their life cycle. It helps identify and manage potential hazards and environmental impacts during the disposal process.

2.6 UNLOCKING SAFETY POTENTIAL: MACHINE LEARNING IN SAFETY MANAGEMENT

Machine Learning provides numerous advantages in the realm of safety management. Below are a few notable benefits that come with its application:

Risk prediction and prevention: Machine Learning algorithms can analyze large volumes of data to identify patterns and predict potential safety risks. By analyzing historical data, Machine Learning models can help identify factors and conditions that contribute to safety incidents, allowing organizations to take proactive measures to prevent accidents or mitigate risks.[3], [30], [31]

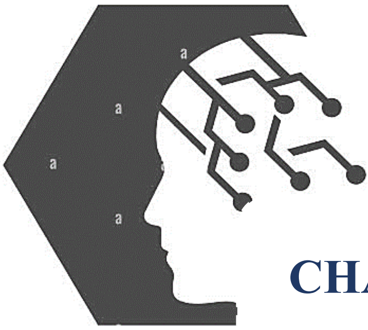
Real-time monitoring and alerting: Machine Learning algorithms can monitor safety-related data in real-time, such as sensor data, employee behavior, or environmental conditions. This enables organizations to detect anomalies or unsafe conditions promptly. By providing real-time alerts, Machine Learning systems can help prevent accidents or trigger immediate responses to mitigate risks.[31]–[33]

Improved incident response: Machine Learning algorithms can assist in incident response by analyzing data collected during safety incidents. These algorithms can identify patterns, root causes, or contributing factors to incidents, enabling organizations to develop more effective response strategies. By learning from past incidents, Machine Learning models can provide valuable insights and recommendations to prevent similar incidents in the future.[2], [34]

Enhanced safety training: Machine Learning algorithms can analyze data on employee behavior, near misses, or incidents to identify areas where additional safety training is needed. By identifying patterns and trends, Machine Learning can help organizations develop targeted training programs to address specific safety risks or improve employee safety awareness.[35], [36]

Data-driven decision-making: Machine Learning enables organizations to make data-driven decisions when it comes to safety management. By analyzing large and complex datasets, Machine Learning models can identify hidden correlations, identify high-risk areas, or recommend interventions to improve safety. This helps organizations prioritize safety efforts, allocate resources effectively, and continuously improve safety performance.[31], [37]

Predictive maintenance: Machine Learning can be applied to equipment and machinery to predict maintenance needs and detect potential failures. By monitoring sensor data, historical maintenance records, and other relevant information, Machine Learning models can identify signs of equipment malfunction or deterioration. This allows organizations to schedule proactive maintenance, reducing the risk of accidents or incidents caused by equipment failure.[38]–[40]



CHAPTER 3 – MACHINE LEARNING MODELS

3.1 INTRODUCTION

In recent years, machine learning has emerged as a popular and widely used technology, and its applications have been extended to various fields, including healthcare, finance, retail, and more. Machine learning is the science and art of programming computers so they can learn from data. In this chapter, we provide an overview of machine learning and machine learning models, their types, and their applications.

3.2 WHAT IS MACHINE LEARNING

At its core, ML is a subset of artificial intelligence that involves the development of algorithms and models that enable computers to learn from data without being explicitly programmed.

Here is a slightly more general definition:

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.[41]

And a more engineering-oriented one:

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.[41]

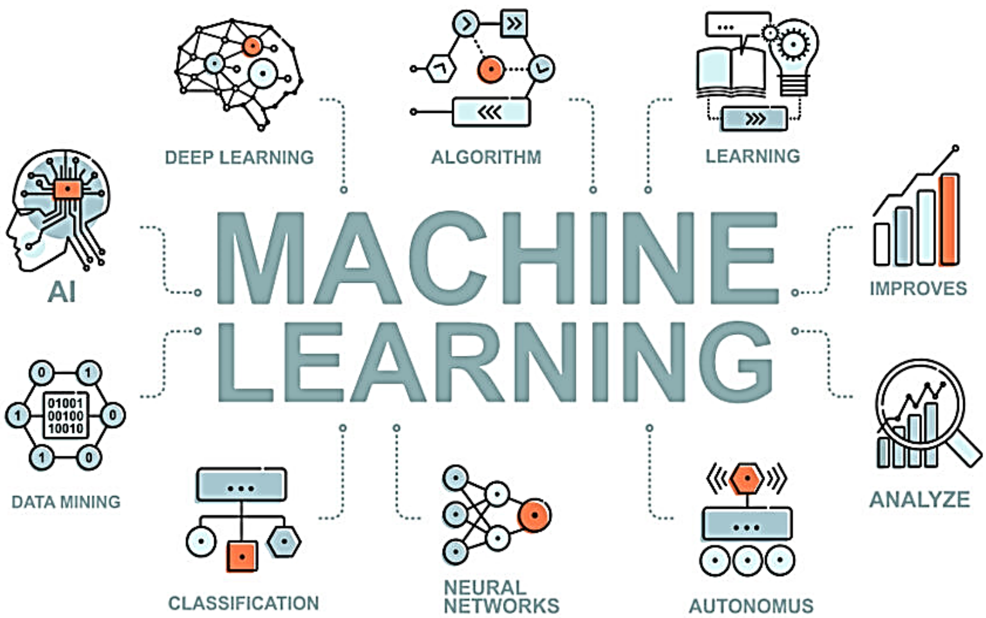


Figure 3-1: Machine Learning

3.3 TYPES OF MACHINE LEARNING

There are three primary categories for ML systems, which are determined by the level and nature of supervision they receive during their training. These categories include supervised learning, unsupervised learning, learning, and Reinforcement Learning.

3.3.1 Supervised Learning

Supervised learning is a common type of machine learning where the computer is trained using labeled or classified data to accurately predict the outcome for new data sets. This technique involves tasks like classification (e.g., determining spam emails) and regression (e.g., predicting customer spending). The main difference between supervised learning and other types of machine learning is that the desired outcome or target is known beforehand. As a result, the computer can learn how to achieve the desired result by adjusting its parameters until it achieves a high level of accuracy.[41], [42]

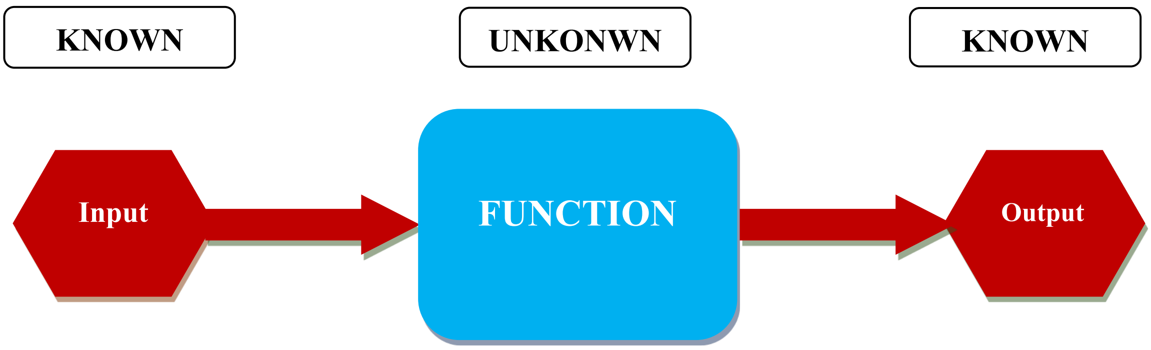


Figure 3-2: The Input and Output Value is Already Known, and the Machine Learning Algorithm Learns the Mapping Function

Supervised machine learning algorithms require external assistance and involve dividing the input dataset into train and test datasets. The train dataset has an output variable that needs to be predicted or classified. The algorithms learn patterns from the training dataset and apply them to the test dataset for prediction or classification.[43]

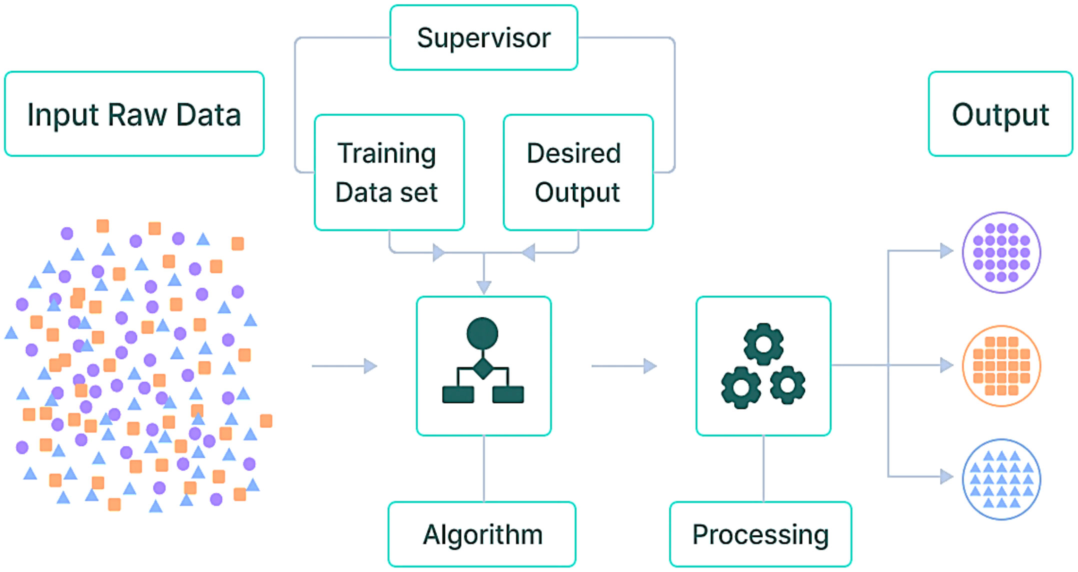


Figure 3-3: Supervised Learning Process

3.3.1.1 Supervised Learning Types

There are two main areas where supervised machine learning comes in handy, classification problems and regression problems.

Classification

Classification is a machine learning process that predicts an output, which can be either the actual or the probability of an event/class, and the number of classes to be predicted can be two or more. Through analysis of historical data, the algorithm must learn to recognize patterns in the relevant input for each class and accurately predict future occurrences of the event or class.[44]

For a simple example, consider how the shapes in the following graph can be differentiated and classified as "circles" and "triangles".

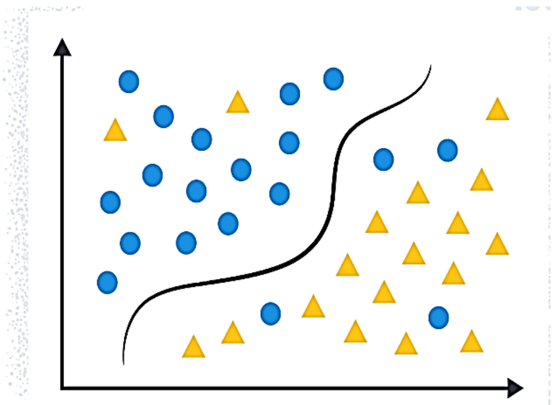


Figure 3-4: Classification Example

Regression

Regression is related to continuous data (value functions). In Regression, the predicted output values are real numbers. It deals with problems such as predicting the price of a car using predictors such as (mileage, age, and brand). To train the system, many examples of cars are provided, including both their predictors and labels (prices).[45]

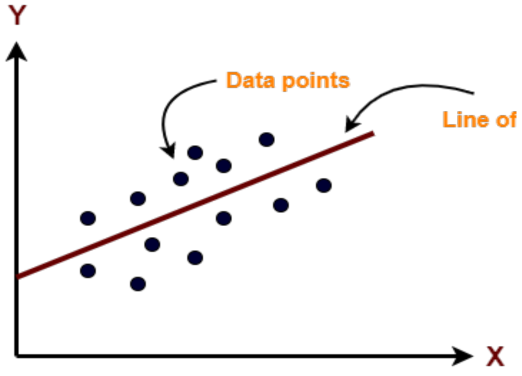


Figure 3-5: Regression Example

3.3.2 Unsupervised Learning

Unsupervised learning is a type of machine learning where the system is trained with unlabeled data, without the need for a teacher or expert intervention to provide guidance on the output. The aim is to find regularities and patterns in the input dataset, group them into specific classes or events, and uncover the dataset's underlying structure.[44], [46]

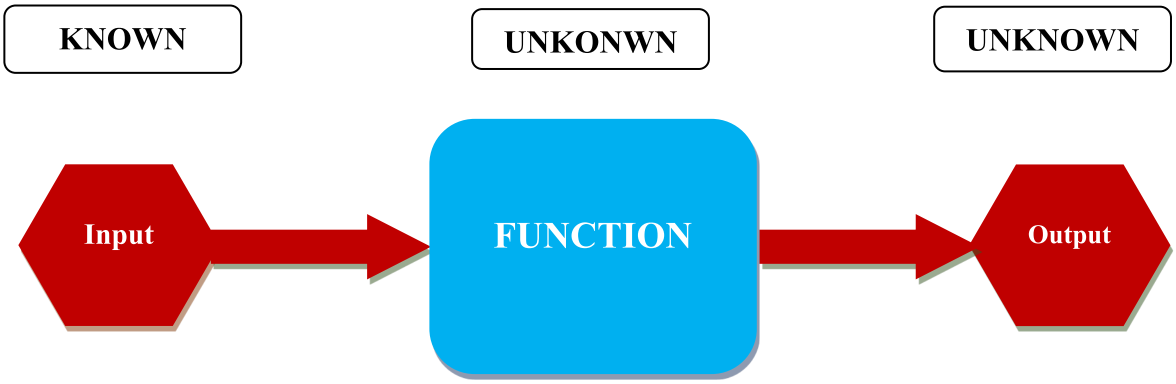


Figure 3-6: Only the Input Data Fed to the Model Without Any Corresponding Output Data

The algorithms used in unsupervised learning learn features from the data, and when new data is introduced, they use the previously learned features to recognize the class of the data. Unsupervised learning is mainly used for clustering and feature reduction. It differs from supervised learning in that it does not require any target outcomes to achieve, and instead, the computer must figure out the desired result by itself. Some common applications of unsupervised learning include creating detailed marketing or business strategies based on the data's patterns and groupings.[43]

Unsupervised learning models can perform more complex tasks than supervised learning models, but they are also more unpredictable. Unsupervised learning is used when there is a lot of data, but the desired output class/event is unknown, and the computer is given unlabeled data without instructions.[41], [44]

3.3.2.1 Unsupervised Learning Types

There are different types of Unsupervised Learning techniques**,** including:

Clustering

Clustering is an Unsupervised Learning technique that involves identifying patterns within a dataset based on similarities or differences in attributes like shape, size, or color. The resulting patterns can be used to group data items or create clusters. There are different types of clustering algorithms, including exclusive, overlapping, hierarchical, and probabilistic.[45]

The primary objective of clustering is to group related items in a given dataset, without prior knowledge of the classes. This technique can be applied in various scenarios, such as grouping similar news articles or customers based on their profiles.[44]

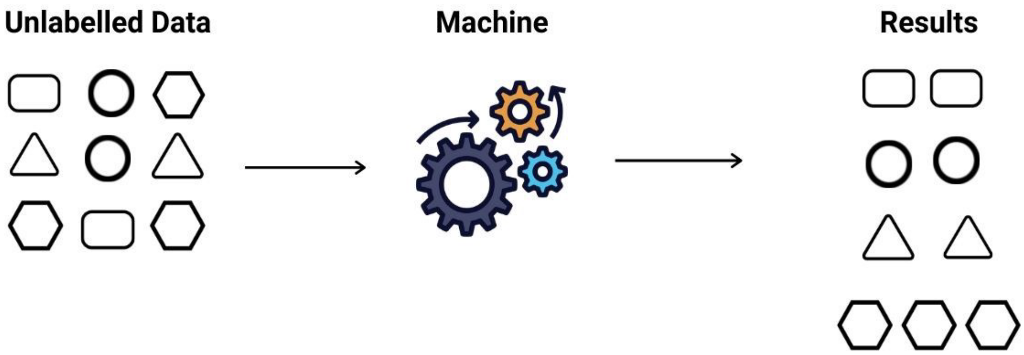


Figure 3-7: Unsupervised Learning Process

Dimensionality reduction

Dimensionality Reduction is a technique used to reduce the number of dimensions in a dataset, making it easier to visualize and analyze relationships among attributes. By eliminating irrelevant features, Dimensionality Reduction helps extract the most important ones, which are key to understanding and building effective Machine Learning models. The process of reducing dimensions also helps simplify a large input dataset by mapping it to a lower dimensional space, making analysis less computationally intensive. This technique is also known as feature extraction, which merges several correlated features into one to retain as much relevant information as possible.[44], [45], [47]

Visualization

Visualization algorithms are unsupervised learning techniques that create 2D or 3D visual representations of complex and unlabeled data to help humans understand how the data is organized and identify patterns. These algorithms aim to preserve the structure of the data, such as avoiding overlapping of separate clusters, in the visual representation.[41]

Visualization is a process of presenting data in a visual form, such as graphs or charts, to make it easier to interpret for humans. Although visualization algorithms can be used for dimensionality reduction, their primary focus is on creating a visually interpretable representation of the data.[41]

Association Rule Learning

Association rule learning is a common unsupervised task that involves exploring large amounts of data to discover interesting relationships between attributes. Specifically, it focuses on finding the dependencies between different data items, which can then be mapped to benefit a particular purpose or objective.[45]

In the field of unsupervised learning, association learning involves finding the relationship of one data item to another data item. These dependencies can then be used to map and understand patterns and behaviors that can ultimately benefit the task at hand. For example, suppose you own a supermarket. Running an association rule on your sales logs may reveal that people who purchase barbecue sauce and potato chips also tend to buy steak. Thus, you may want to place these items close to each other.[41], [47], [48]

3.3.3 Reinforcement Learning

Reinforcement learning is a subfield of machine learning that involves training an agent to take actions in an environment to achieve a goal. The agent observes the state of the environment, selects an action, and receives feedback in the form of rewards or penalties. The goal of the agent is to learn a policy that maps environmental states to actions that maximize the cumulative reward over time.[41], [46]

Unlike supervised and unsupervised learning, reinforcement learning is focused on optimizing a specific outcome rather than predicting a target outcome. The goodness of policies is assessed based on past good action sequences, which the agent uses to improve its decision-making process.[42]

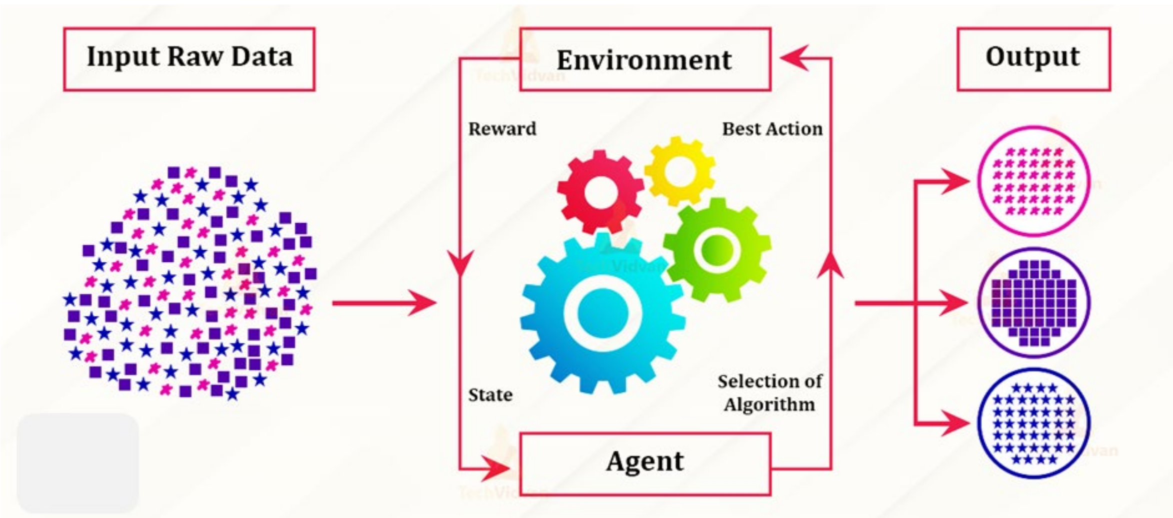


Figure 3-8: Reinforcement Learning Process

The key concepts in reinforcement learning are the agent, the environment, the action, and the reward. The agent is the learning system that interacts with the environment by selecting and performing actions. The environment is the external world in which the agent operates. The action is the decision made by the agent based on the observed state of the environment. The reward is the feedback given to the agent to reinforce its behavior. Through these components, the agent learns to take actions that maximize the cumulative reward over time.[47]

Figure 3-9: Reinforcement Learning

3.4 THE DATA-DRIVEN APPROACH

The data-driven approach is a powerful methodology that involves several key components. At its core, the approach emphasizes the importance of data and seeks to leverage this data to train machine learning models that are accurate, effective, and relevant to the problem domain. the key components of the data-driven approach are:

Data collection and preprocessing:

The data-driven approach to machine learning begins with collecting data from a variety of sources, which can include databases, APIs, web scraping, and sensor networks, among others. Once the data is collected, it must be cleaned and preprocessed to ensure that it is in a format suitable for machine learning. This involves removing duplicates, dealing with missing values, and transforming the data into a format that can be easily used by machine learning algorithms. Data preprocessing may also involve feature scaling and normalization, as well as feature selection, which involves identifying and removing irrelevant or redundant features.

Feature engineering:

Feature engineering is the process of selecting and transforming relevant features from the preprocessed data that can be used by machine learning models to make accurate predictions. Feature engineering may involve domain expertise and creativity to identify the most important features for a particular problem domain. It can also involve techniques such as dimensionality reduction, which reduces the number of features in the dataset while retaining as much relevant information as possible.

Model selection and training:

The selection and training of machine learning models is a critical component of the data-driven approach. Different machine learning algorithms may be more or less suitable for a particular problem domain, and selecting the most appropriate algorithm for a given task can have a significant impact on the accuracy and effectiveness of the resulting model. Once the algorithm is selected, the model is trained on the preprocessed data, and the best parameters are selected to optimize the model's performance. The model is then evaluated on a separate test set to ensure that it is generalizing well to new data.

Model evaluation:

Model evaluation involves testing the trained model on a separate test set to ensure that it is generalizing well to new data. This involves using appropriate metrics to measure the accuracy and effectiveness of the model, such as precision, recall, and F1-score. Adjustments may need to be made to the model or data preprocessing approach based on the evaluation results.

Deployment and maintenance:

Once the model is trained and evaluated, it needs to be deployed in a production environment. This involves selecting appropriate hardware and software platforms, integrating the model into existing systems, and monitoring the model's performance over time to ensure that it remains accurate and effective. Maintenance may involve retraining the model periodically, as well as updating the data preprocessing pipeline to ensure that the model continues to receive high-quality data.[49], [50]

3.4.1 Frameworks For Building Machine Learning Systems

Data mining process

Is an essential component of the data-driven approach, allowing organizations to extract valuable insights from their data. A process involves the use of various techniques to identify patterns and relationships in data. These patterns are then used to develop predictive models that can be used to make informed decisions. There are several models available to guide the data mining process, including:

- Knowledge Discovery Databases (KDD) process model
- CRoss Industrial Standard Process for Data Mining (CRISP – DM)
- Sample, Explore, Modify, Model and Assess (SEMMA)

Data mining is critical to the success of a data-driven approach in ML. By using data mining techniques to extract valuable insights, organizations can make informed decisions that can lead to increased efficiency, profitability, and competitiveness.[44]

3.4.1.1 Knowledge Discovery Databases (KDD)

Knowledge Discovery Databases (KDD) refers to the process of extracting useful knowledge and information from large and complex datasets. It involves various techniques such as data mining, machine learning, and statistical analysis to identify patterns, relationships, and trends in the data that can be used to make informed decisions.

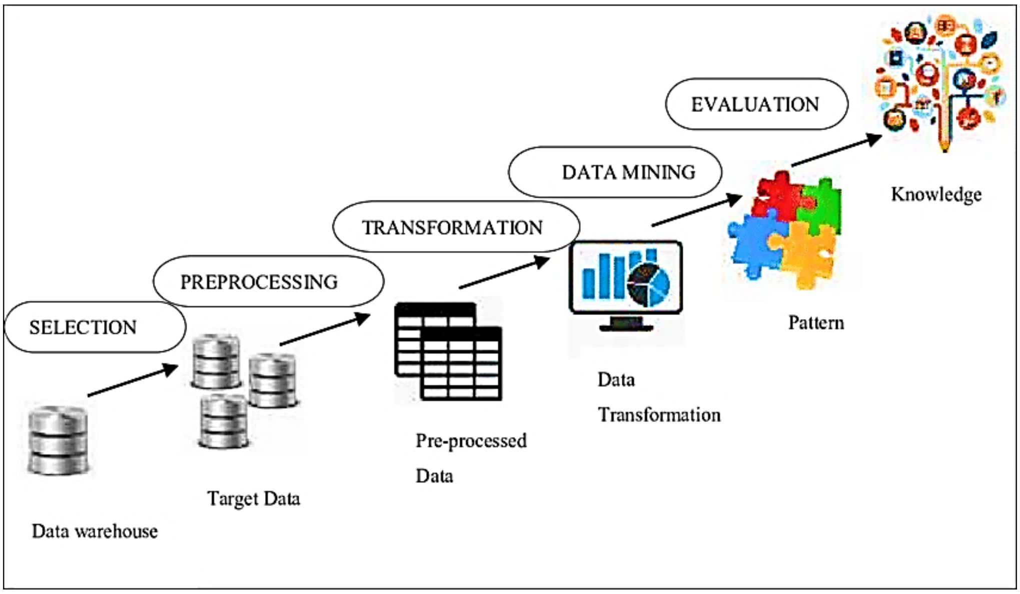


Figure 3-10: KDD Data Mining Process Flow

The KDD process involves several stages including:

Selection: The first step, which involves identifying the target data and retrieving a relevant subset of variables and data samples from potentially diverse and heterogeneous sources.

Preprocessing: Is the next step and is necessary to address issues such as missing values, errors, and inconsistencies in the data. Techniques such as outlier treatment, noise removal, and handling of missing data are used to ensure that the data is of high quality and suitable for analysis. Prior domain knowledge can be used to remove inconsistencies and duplicates from the data.

Transformation: Is the process of preparing the data for analysis by finding useful features to represent the data, reducing dimensionality, and applying various techniques such as smoothing, aggregation, generalization, normalization, feature construction, data reduction, and compression.

Data Mining: Involves using machine learning algorithms to extract patterns from the data, such as exploratory or predictive models and cluster analysis. The selection of appropriate methods for pattern search is crucial to ensure reliable and valid outputs.

Interpretation/Evaluation: Involves making the mined patterns understandable to the user through summarization, visualization, and interpretation. Patterns are local structures that make statements about restricted regions of the space spanned by the variables, while models are global structures that make statements about any point in measurement space. It is important to evaluate the results and ensure that they align with the original goals of the analysis.[44]

3.4.1.2 CRoss Industrial Standard Process for Data Mining

The CRoss-Industry Standard Process for Data Mining (CRISP-DM) is a widely recognized and popular methodology for data mining, it is a non-proprietary and open process that outlines a comprehensive framework for data mining projects.

The CRISP-DM model consists of six phases, as follows:

Phase 1: Business Understanding: Is the first step of the framework, where the main objective is to gain a comprehensive understanding of the project's goals and expectations from a business perspective. The objectives are then transformed into a data mining or machine learning problem definition. A plan of action is developed around data requirements, input from business owners, and the design of outcome performance evaluation metrics.

Phase 2: Data Understanding: Is where initial data collection takes place, based on the requirements identified in the previous phase. The data is analyzed to determine its relevance to the project's objectives, identify any gaps or quality issues, and generate hypotheses.

The results are presented to the business in an iterative manner to gain more clarity and understanding of the project objectives.

Phase 3: Data Preparation: Is where the collected data is cleaned and made ready for the model building phase. This involves tasks such as filling in missing data, identifying important features, applying transformations, and creating new features as necessary. This phase is critical because the accuracy of the resulting model depends on the quality of the data fed into it.

Phase 4: Modeling: Involves the application of various machine learning algorithms to the clean dataset. The parameters of each model are tuned to optimize their performance, and the resulting model performance is recorded.

Phase 5: Evaluation: Involves benchmarking the performance of the various models identified as having high accuracy against data that was not used in training. The results are then verified against the business requirements identified in Phase 1, with subject matter experts from the business involved to ensure accuracy and usability.

Phase 6: Deployment: Is the final phase of the framework, where the focus is on the usability of the model output. The final model is implemented, and users are trained on how to interpret and use the model output to make informed business decisions. Periodic model training and prediction times are scheduled based on business requirements.[44]

3.4.1.3 Sample, Explore, Modify, Model and Assess (SEMMA)

SEMMA is a framework for building machine learning models that is incorporated in SAS Enterprise Miner, a product by SAS Institute Inc. It consists of five sequential steps: Sample, Explore, Modify, Model, and Assess.

Sample: involves selecting a representative subset of the large dataset provided for building the model. This subset should contain sufficient information to retrieve and should be divided into training and validation data. This step is essential to building the model efficiently, especially when computational power is limited.

Explore: involves analyzing the selected data to understand the data gaps and relationships between variables. Two key activities in this phase are univariate and multivariate analysis, and data visualization is heavily used to help understand the data better.

Modify: involves cleaning variables where necessary, creating new derived features by applying business logic to existing features based on the requirement, and transforming variables if necessary. The outcome of this phase is a clean dataset that can be passed to the machine learning algorithm to build the model.

Model: involves applying various modeling or data mining techniques to the preprocessed data to benchmark their performance against desired outcomes. This step helps to choose the best model for the given problem.

Assess: involves evaluating the model's performance against test data that was not used in model training. This step ensures that the model is reliable and useful for business purposes.[44]

3.5 MACHINE LEARNING ALGORITHMS

3.5.1 Decision Trees

Decision Trees are versatile machine learning algorithms that can be used for both classification and regression tasks, and are capable of fitting complex datasets. They learn a hierarchy of if/else questions that lead to a decision.

Decision Trees work by splitting the population into two or more homogeneous sets based on the most significant attributes/independent variables, which can be categorical or continuous.[51]

It is structured as a tree-like diagram consisting of nodes and branches. The diagram starts with a **root node**, which is the initial decision point. The root node splits into multiple **branches**, with each branch representing a feature or attribute of the dataset. The **internal nodes** of the tree represent the features that are used to make decisions, and each internal node splits into further branches based on the possible outcomes of that feature. This process continues until the final nodes of the tree are reached, which are known as **leaf nodes** and they represent the final outcomes of the decision process.[52]

To simplify this concept in a diagram, we can label decision nodes (root and internal nodes) as "Questions" and leaf nodes as "Answers".

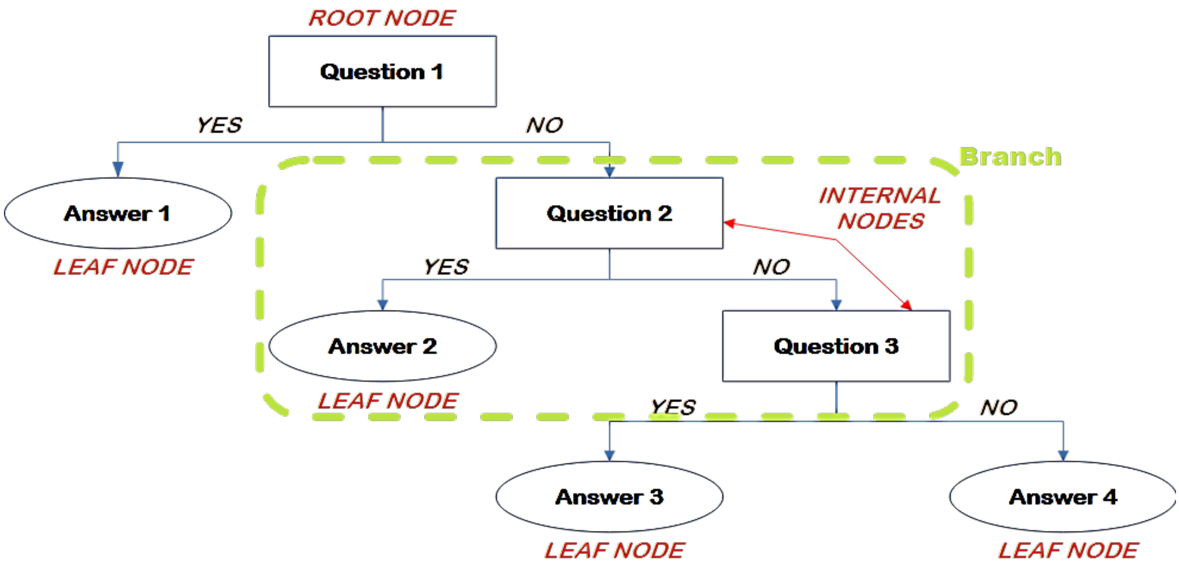


Figure 3-11: Decision Tree Structure

3.5.1.1 How Does Decision Trees Work?

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree.[53]

3.5.1.2 Why Decision Trees?

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.[53]

3.5.2 Random Forests

Random Forest is a popular supervised learning algorithm that can be used for both classification and regression problems in machine learning. It is based on the concept of ensemble learning, which involves combining multiple classifiers to solve complex problems and improve the model's performance.

Random Forest is essentially a classifier that contains a number of decision trees on various subsets of the given dataset. Instead of relying on a single decision tree, the algorithm takes the prediction from each tree and uses majority voting to predict the final output. The more trees in the forest, the better the accuracy, and overfitting is prevented.[54]

To improve the predictive accuracy of the model, Random Forest generates multiple decision trees for subsets of the dataset and finds the average. Typically, a Random Forest should contain 64-128 trees. The algorithm is fast and can handle missing and incorrect data efficiently.[52]

To classify a new dataset or object, each tree in the Random Forest gives a classification result. The final output is then determined by majority voting among the predictions of all the trees.[52]

3.5.2.1 How Does Random Forests Work?

Random Forest algorithm works in two phases. In the first phase, the algorithm creates a random forest by combining N decision trees. In the second phase, the algorithm makes predictions for each tree created in the first phase.

The working process of the Random Forest algorithm can be explained in the following steps and diagram:

Step 1: Randomly select K data points from the training set.

Step 2: Build decision trees associated with the selected data points (subsets).

Step 3: Decide on the number of decision trees (N) to build.

Step 4: Repeat **steps 1 & 2** to create N decision trees.

Step 5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that receives the majority votes.

By following these steps, the Random Forest algorithm improves the predictive accuracy of the model, prevents overfitting, and handles missing or incorrect data efficiently.[54]

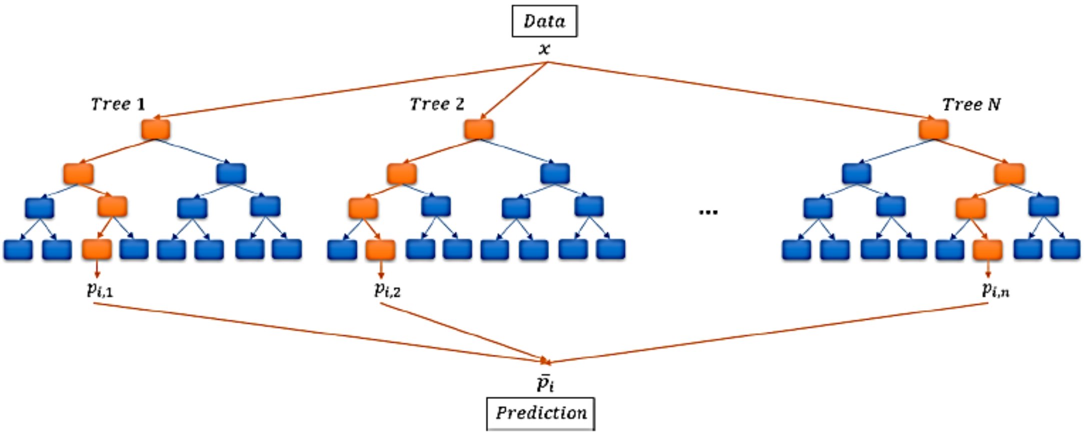


Figure 3-12: Random Forests Structure

3.5.2.2 Why Random Forests?

Here are some reasons why the Random Forest algorithm is preferred:

- It has a shorter training time compared to other algorithms.
- It provides high accuracy predictions, even with large datasets, and can handle missing data effectively.[54]

3.5.3 Support Vector Machines (SVMs)

A Support Vector Machine (SVM) is a very powerful and versatile Machine Learning model, capable of performing linear or nonlinear classification, regression, and even outlier detection. It is one of the most popular models in Machine Learning, and anyone interested in Machine Learning should have it in their toolbox. SVMs are particularly well suited for classification of complex but small- or medium-sized datasets.[41]

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.[55]

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.[55]

Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

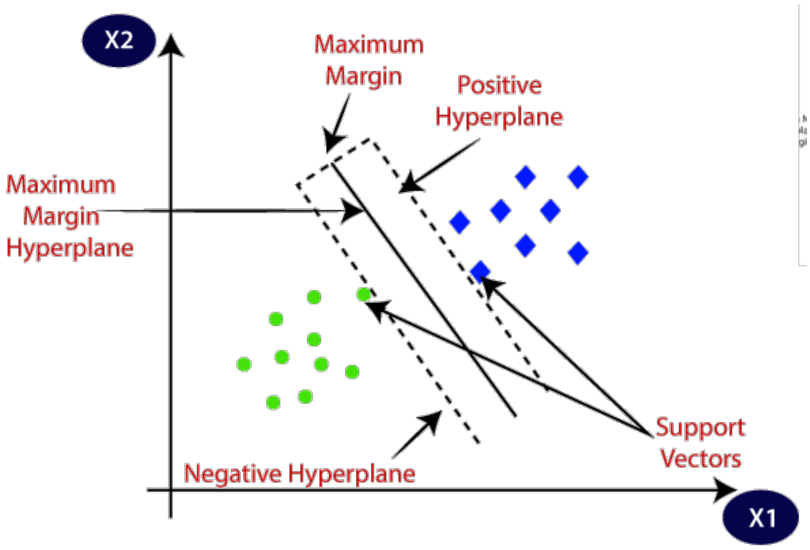


Figure 3-13: Support Vector Machines

3.5.3.1 Types of Support Vector Machines

Linear SVM: Is used when the dataset can be separated into two classes using a straight line. This separation is known as linearly separable data. The classifier used for linearly separable data is called a Linear SVM classifier. Linear SVM is commonly used in image processing, text classification, and bioinformatics.[55]

Non-Linear SVM: Is used when the dataset cannot be classified using a straight line. Such data is termed as non-linear data. Non-linear SVM classifiers use kernel functions to transform the data from the input space to a higher-dimensional feature space where the data can be linearly separated. Some commonly used kernel functions are polynomial, radial basis function, and sigmoid. Non-linear SVM is used in a variety of applications such as image recognition, text analysis, and gene expression analysis.[55]

3.5.3.2 Hyperplane and Support Vectors in the SVM Algorithm

Hyperplane

A hyperplane in SVM is the best decision boundary to classify data points in n-dimensional space. The hyperplane's dimensions depend on the number of features in the dataset, with 2 features resulting in a straight line and 3 features resulting in a 2D plane. The hyperplane created has a maximum margin, which is the maximum distance between the data points.[55]

Support Vectors

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.[55]

3.5.3.3 How Does Support Vector Machines Works?

Linear SVM

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair (x1, x2) of coordinates in either green or blue. Consider the **(Image 1)** of **Figure 3-14**.

So, as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the **(Image 2)** of **Figure 3-14**.

Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a hyperplane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to maximize this margin. The hyperplane with maximum margin is called the optimal hyperplane. **(Image 3)** of **Figure 3-14**.[55]

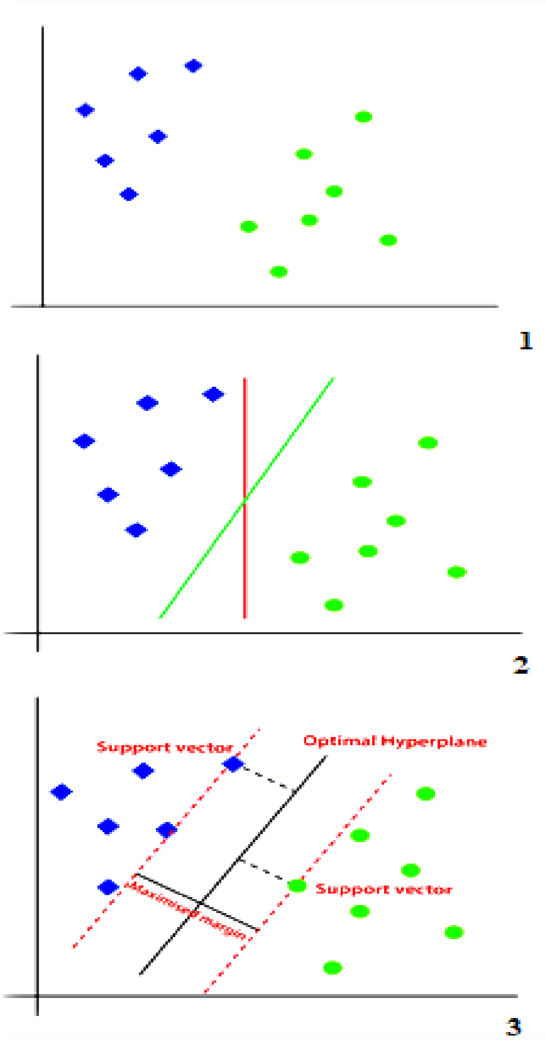


Figure 3-14: Linear SVM Workflow

Non-Linear SVM

If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the **(Image 1)** of **Figure 3-15**.

So, to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third-dimension z. It can be calculated as:

$$Z = x^2 + y^2$$

By adding the third dimension, the sample space will become as **(Image 2)** of **Figure 3-15**.

So now, SVM will divide the datasets into classes in the following way. Consider the **(Image 3)** of **Figure 3-15**.

Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with z=1, then it will become as **(Image 4)** of **Figure 3-15**.

Hence, we get a circumference of radius 1 in case of non-linear data.[55]

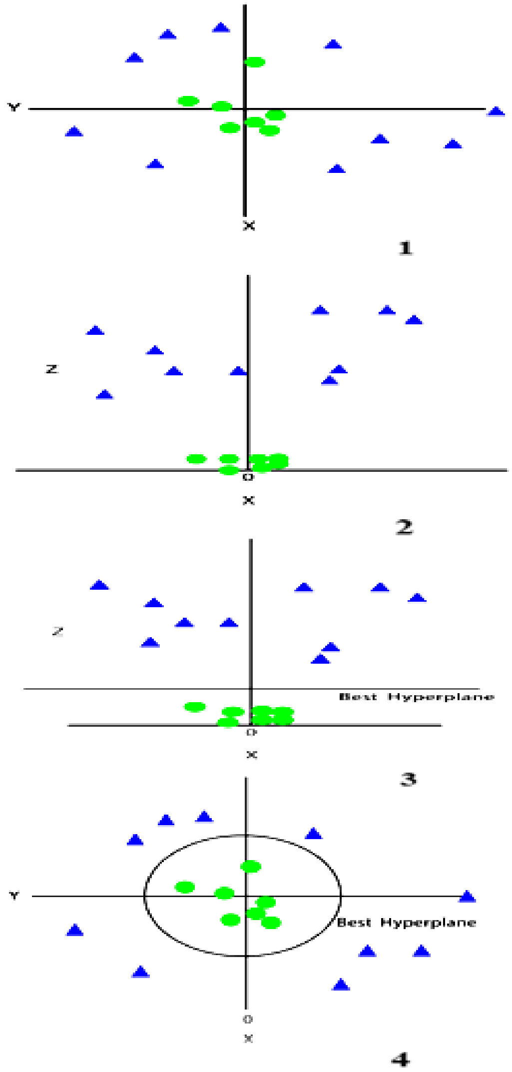


Figure 3-15: Non-Linear SVM Workflow

3.5.3.4 Why Support Vector Machines?

- Support vector machines (SVMs) are important for developing predictive models, as they are easy to understand and use, and can process linear and non-linear data using kernels.
- SVMs have a wide range of applications in various domains where higher dimensional spaces are considered.
- Developing SVM models requires selecting the right kernel, tuning hyper-parameters, and investing time and resources in the training phase.[56]

3.5.4 K-Nearest Neighbors

K-Nearest Neighbor (K-NN) is a simple supervised learning algorithm that is widely used in machine learning. The K-NN algorithm works on the assumption that new data points are similar to the available data points and can be put into a category that is most similar to the available categories. In other words, it classifies new data points by comparing them to existing data points, based on their similarity. The algorithm stores all the available data and classifies new data points based on their similarity to the existing data. This makes it easy to classify new data points as they appear.[57]

K-NN can be used for both classification and regression problems, but it is mostly used for classification problems. The algorithm is non-parametric, which means that it does not make any assumptions about the underlying data. It is also known as a lazy learner algorithm because it does not learn from the training set immediately. Instead, it stores the training set and performs actions on it at the time of classification.[57]

During the training phase, the K-NN algorithm stores the entire dataset, and when it receives new data, it classifies it into the category that is most similar to the new data point. The similarity between the new data point and the existing data points is measured using a distance metric such as Euclidean distance or Manhattan distance. The K-NN algorithm then assigns the new data point to the category that has the highest number of neighboring points. The number of neighbors, known as the **"K"** value, can be set manually, and the optimal value for **"K"** depends on the dataset and the problem being solved.[57]

3.5.4.1 How Does K-Nearest Neighbors Work?

Step 1: Select the number **"K"** of the neighbors that we want to consider when making a prediction for a new data point.

- There is no particular way to determine the best value for **"K"**, so we need to try some values to find the best out of them. The most preferred value for **"K"** is 5.
- A very low value for **"K"** such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for **"K"** are good, but it may find some difficulties.

Step 2: Calculate the Euclidean distance between the new data point and all the existing data points in the training set. Euclidean distance is a measure of the distance between two points in a Euclidean space.

Step 3: Select the **"K"** data points with the smallest Euclidean distance from the new data point.

Step 4: Count the number of data points in each category (i.e., the number of neighbors that belong to each class) among the K nearest neighbors.

Step 5: Assign the new data point to the category that has the highest number of neighbors among the K nearest neighbors. This is also known as the majority vote.

Step 6: After performing the above steps, our K-NN model is ready to make predictions for new data points based on the labeled data in the training set.[57]

Suppose we have a new data point and we need to put it in the required category. Consider the **(Image 1)** of **Figure 3-16**.

Firstly, we will choose the number of neighbors, so we will choose the K=5.

Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as: **(Image 2)** of **Figure 3-16**.

By calculating the Euclidean distance, we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the **(Image 3)** of **Figure 3-16**.

As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.[57]

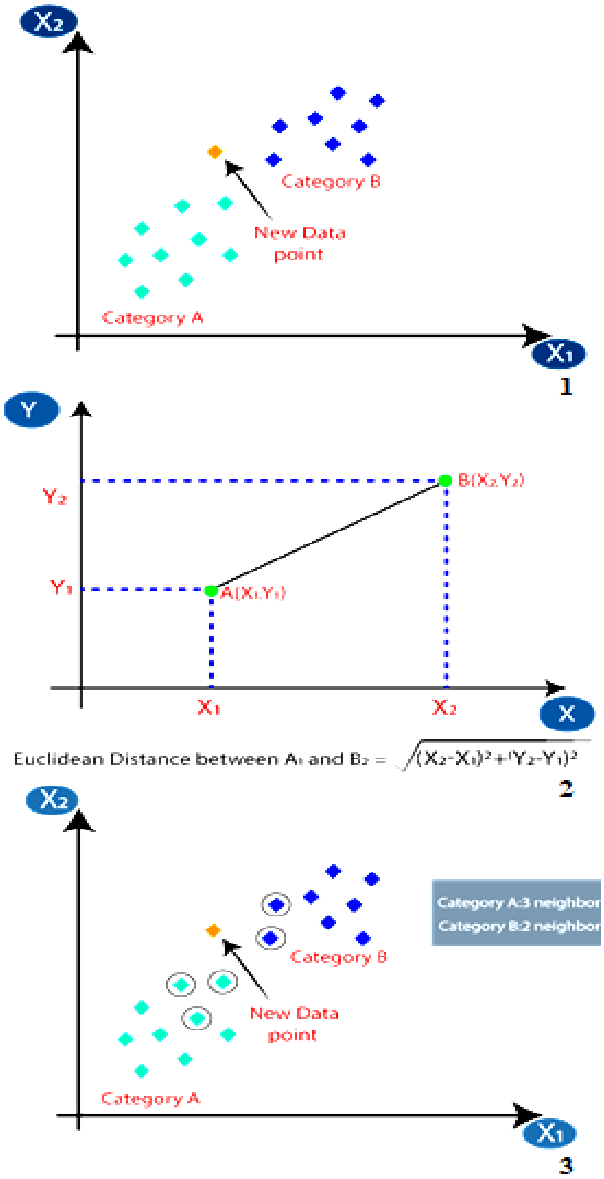


Figure 3-16: K-NN Workflow

3.5.4.2 Why K-Nearest Neighbors?

- Simple algorithm and hence easy to interpret the prediction.
- Non parametric, so makes no assumption about the underlying data pattern.
- Used for both classification and Regression.
- Training step is much faster for nearest neighbor compared to other machine learning algorithms.[58]

3.6 HOW AI AND ML CAN BE USED TO ENHANCE SAFETY MANAGEMENT PROCESSES AND WORKFLOWS

The utilization of AI and ML in safety management processes and workflows can greatly enhance their effectiveness. There are various ways in which AI can be applied, including predictive analytics, automatic risk prioritization, simplification of complicated workflows, next-best actions after audit findings, better analytics in safety management reviews, identification of similar safety events, reduction of redundant safety event records, and automatic recognition of the type of event being recorded.

Predictive Analytics: Which involves analyzing past safety management data and metrics to predict future incidents. Another approach is through automatic risk prioritization, where AI can be used to prioritize safety risks by running analytics on similar data from the past, once safety observations, near-misses or incidents are reported.

Automatic Risk Prioritization: Once safety observations, near-misses or incidents are reported, AI can be used to prioritize which safety risk to tackle with urgency. This can be done by running analytics on similar data (***CQ.AI*** has a framework/model for similarity identification) from the past.

Simplify Complicated Workflows: Some safety processes can be simplified by using intelligent automation. For instance, let us say someone from the ***H&S*** team is conducting a safety inspection and filling in a document. Is it possible to take in certain data points from this safety inspection process into the safety observations module automatically? Then, we're one step closer to conducting a risk assessment and pushing it to ***CAPA***.

Next-Best Actions after Audit Findings: Once a safety audit is complete, a good AI model can suggest the "next best steps" by spotting similar audit findings from the past.

Better Analytics in Safety Management Reviews: By using next-generation dashboards, interactive charts, and a tower of data, the process of running management review meetings can be made more efficient (thanks to predictive insights, risk prioritization, etc.).

Identify Similar Safety Events: ***Compliance Quest EHS*** solution can automatically identify similar safety events and determine root cause reoccurrence.

Reduce Number of Redundant Safety Event Records: Sometimes, the same safety event gets recorded by multiple people creating redundant records. By using intelligent automation, redundant records are automatically deleted, thus saving time and increasing the productivity of the safety team.

Automatically Recognize Type of Event Being Recorded: Using AI and ML, it is possible to classify and record the type of event being recorded in the self-reporting portal. For instance, if a safety event is being entered in the **MyCQ** portal, based on the description the tool suggests whether it is a complaint record, safety event record, or a change request. While these are simple value additions, it saves critical time and effort, while also making it easy to filter events by type. It also helps eliminate manual data entry errors.[1]

3.7 LIMITATION AND CHALLENGES OF MACHINE LEARNING

Insufficient Quantity of Training Data

One of the primary challenges of machine learning is the lack of sufficient data for training models. Without a significant amount of training data, machine learning algorithms may not be able to learn the underlying patterns and relationships in the data. This can lead to poor model performance and inaccurate predictions.[41]

Nonrepresentative Training Data

Another challenge is ensuring that the training data is representative of the data that the model will encounter in the real world. If the training data is biased or unrepresentative, the model may make inaccurate predictions or fail to generalize to new data.

Poor-Quality Data

Poor-quality data can also be a significant challenge for machine learning. Data that contains errors, outliers, or missing values can negatively impact the accuracy of the model. It is important to ensure that the data used for training is clean and of high quality.

Irrelevant Features

The presence of irrelevant features in the data can also pose a challenge for machine learning. Irrelevant features can lead to overfitting, where the model fits the noise in the data rather than the underlying patterns. Feature selection and feature engineering techniques can be used to address this challenge.

Overfitting the Training Data

Overfitting occurs when the model is too complex and fits the training data too closely, resulting in poor generalization to new data. This can occur when the model has too many parameters relative to the amount of training data or when the model is too flexible. Regularization techniques such as L1 and L2 regularization can be used to address overfitting.

Underfitting the Training Data

Underfitting occurs when the model is too simple and cannot capture the underlying patterns in the data. This can occur when the model has too few parameters or is too rigid. Adding more features, increasing the complexity of the model, or using a more powerful algorithm can help to address underfitting.[41]

CHAPTER 4 – APPLICATION OF MACHINE LEARNING: PREDICT THE SEVERITY OF INCIDENTS

4.1 INTRODUCTION

In this chapter, we present our work on predicting the incidents severity, our model leverages the power of Machine Learning algorithms to predict incidents severity based on relevant features extracted from incident and accidents data. By analyzing historical incident records and their associated attributes, the model learns patterns and relationships to make accurate predictions. In this study, we focused on utilizing supervised Machine Learning algorithms to classify incidents as either Fatal or Non-Fatal based on a set of relevant features. By employing binary classification techniques.

4.2 OVERVIEW OF OUR WORK

To achieve our objective, we employed several popular supervised Machine Learning algorithms, including Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Multilayer Perception (MLP) and K-Nearest Neighbor (KNN). Each algorithm offers distinct characteristics and capabilities, enabling us to explore different modeling approaches and evaluate their performance in predicting the severity of incidents accurately. Then, we will compare the results and the performance of these models to know which one achieve our objective accurately.

4.3 TOOLS AND SOFTWARE USED

4.3.1 Python

Python is a high-level programming language that was created by Guido Van Rossum in 1990 [59]. It is an easy language to learn due to its simple syntax. Python is interpreted and object-oriented, with dynamic semantics. It has built-in data structures and supports rapid application development and script writing. Python emphasizes code readability and modularity, and supports modules and packages for code reuse. It has unique features that make it easy to debug, including a source level debugger and a simpler approach to debugging with print statements. Python is available in binary or source form and can be distributed for free [60].

4.3.2 Kaggle

Kaggle is the largest platform for competitive data science, with over 500,000 competitors, and offers a great opportunity to learn about the latest techniques and share knowledge. To win Kaggle competitions, it is important to use an agile sprint and iteration approach with versatile libraries such as Scikit-learn, Vowpal Wabbit, XGBoost, and Keras, among others.[61].

4.3.3 Python Libraries / Dependencies

Python has a wide range of libraries that are used in Machine Learning. These libraries provide a variety of tools and functions that make it easy to perform Machine Learning tasks. Here are some of the most popular Python libraries for Machine Learning:

NumPy: Is a popular Python library for large multi-dimensional array and matrix processing. It provides a variety of mathematical functions that are useful for Machine Learning.

Scikit-learn: Scikit-learn is one of the most popular Machine Learning libraries for classical Machine Learning algorithms. It is built on top of NumPy and SciPy and provides a variety of tools for data mining and data analysis.

TensorFlow: TensorFlow is an open-source Python library that specializes in deep learning. It provides a variety of tools for building and training deep neural networks.

Keras: Is a high-level neural networks API that is written in Python. It is built on top of TensorFlow and provides a simple and easy-to-use interface for building and training deep neural networks.

PyTorch: Is another popular deep learning library that is written in Python. It provides a variety of tools for building and training deep neural networks.

Matplotlib: Matplotlib is a popular Python library for data visualization. It provides a variety of tools for creating charts, graphs, and other visualizations that are useful for Machine Learning.

Pandas: Is a popular Python library for data manipulation and analysis. It provides a variety of tools for reading, writing, and manipulating data that are useful for Machine Learning.[62]

Figure 4-1: Python Libraries

4.3.4 Google Collaboratory.

Google Colab is a document that allows you to write, run, and share Python code within your browser. It is a version of the popular Jupyter Notebook within the Google suite of tools. Jupyter Notebooks (and therefore Google Colab) allow you to create a document containing executable code along with text, images, HTML, LaTeX, etc. which is then stored in your Google Drive and shareable to peers and colleagues for editing, commenting, and viewing.[63]

4.4 METHODOLOGY

In the following sections, we will provide a detailed overview of the dataset used, the preprocessing steps applied, the feature selections process, and the evaluation metrics used to assess the performance of the model. Additionally, we will present and discuss the result obtained from each model, highlighting their strength and limitations in the context of safety prediction.

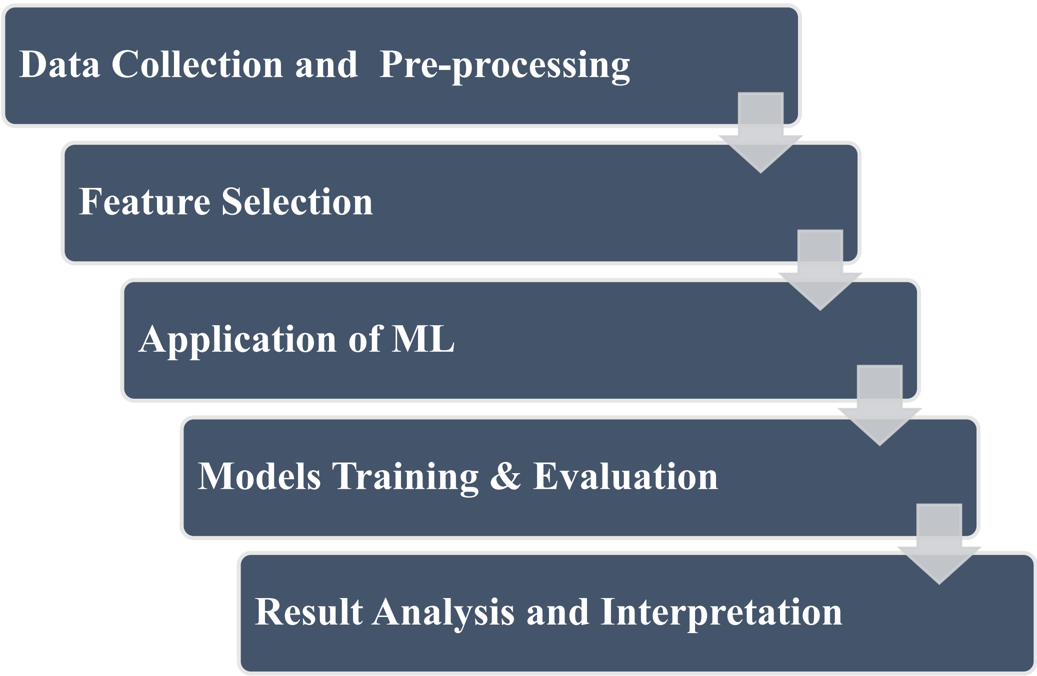


Figure 4-2: Methodology

4.5 DATA COLLECTION

4.5.1 Source

For our study, we utilized the OSHA construction dataset obtained from the Kaggle platform as the primary source of data [64]

4.5.2 Data Introduction

A comprehensive dataset of incident reports was obtained from the Kaggle platform, which encompasses a wide range of incident and accident attributes. The dataset captures abstracts that describe the accidents and injuries sustained by construction workers within the timeframe of 2015-2017.

The dataset consists of a combination of structured and unstructured data elements. While the unstructured text abstracts provide detailed narratives of the incidents, the dataset also includes structured data associated with these abstracts.

The table provided below presents a summary of the data information available in the dataset. It outlines the various columns along with their descriptions and data types. This information is essential for understanding the dataset and its contents. Each column represents a specific attribute or characteristic of the events recorded in the dataset.

Table 4-1: Summary of the Dataset Columns, their Descriptions, and Corresponding Data Types.

Columns Name	Description	Dtype
summary_nr	Summary number	int64
Event Date	Date of the event	datetime64[ns]
Abstract Text	Textual abstract	object
Event Description	Description of the event	object
Event Keywords	Keywords related to the event	object
con_end	Construction end	object
Construction End Use	End use of the construction	object
build_stor	Building stories	Int64
Building Stories	Number of building stories	object
proj_cost	Project cost	object
Project Cost	Cost of the project	object
proj_type	Project type	object
Project Type	Type of the project	object
Degree of Injury	Degree of injury	int64
nature_of_inj	Nature of injury	int64
Nature of Injury	Description of the injury	object
part_of_body	Part of body	int64
Part of Body	Body part affected	object
event_type	Event type	Int64
Event type	Type of the event	object
evn_factor	Environmental factor	Int64
Environmental Factor	Factor related to the event	object
hum_factor	Human factor	Int64
Human Factor	Factor related to humans	object
task_assigned	Task assigned	int64
Task Assigned	Description of the task	object
hazsub	Hazardous substance	object

Columns Name	Description	Dtype
fat_cause	Fatal cause	int64
fall_ht	Fall height	int64
Event Year	Year of the event	int64
Event Month	Month of the event	int64

4.5.3 Exploratory Data Analysis (EDA)

EDA is all about understanding our data by employing summarizing and visualizing techniques.

4.5.3.1 Visualizing The Data

The line plot **(Figure4-3)** of incidents over time provides a comprehensive view of the trend and patterns of accidents and incidents recorded from 2015 to 2017. The graph reveals an overall increasing trend in the number of incidents, indicating a rise in the frequency of reported accidents and incidents during this period. Within each year, there are fluctuations, suggesting some level of seasonality or periodicity. Notably, there is a peak in incidents around mid-year, followed by a decrease towards the end of the year. Comparing the years, 2017 has the highest number of incidents, followed by 2016 and then 2015. This indicates a potential escalation in the reporting or occurrence of incidents over time.

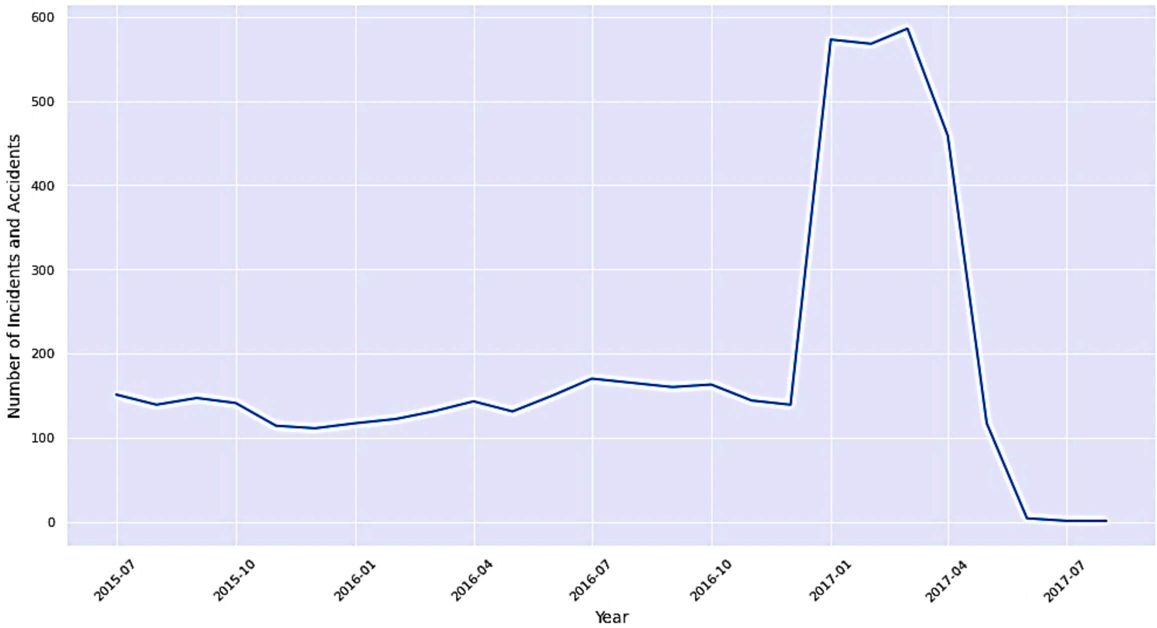


Figure 4-3: Distribution of Accidents and Incidents Over Time

The heatmap visualization (**Figure 4-4**) provides a comprehensive understanding of the distribution and temporal trends of incidents and accidents over a three-year period. This graphical analysis offers valuable insights into the dynamics of incidents, particularly when examining their occurrence by month and year.

The heatmap illustrates a consistent trend where the highest number of incidents is reported in the year 2017, followed by 2016 and 2015. This observation indicates a potential escalation in incidents over the examined timeframe. Moreover, exploring within-year variations reveals distinct fluctuations in incident frequency across different months.

An intriguing pattern emerges as we assess the monthly distribution of incidents. Notably, certain months exhibit relatively higher incident counts compared to others. For instance, January and December of specific years, along with other months in different years, consistently experience elevated incident numbers. This pattern may suggest a link between incidents and transitional phases, such as the beginning and end of a calendar year. These periods often coincide with shifts in activities, seasonal changes, or work routines, which could contribute to elevated incident rates. However, unraveling the underlying factors necessitates deeper investigation, considering elements like industry-specific trends, seasonal influences, workforce dynamics, or reporting patterns.

While the heatmap offers valuable insights into incident distribution, uncovering the nuances of this pattern demands a multi-faceted analysis, enriched by contextual information and domain expertise

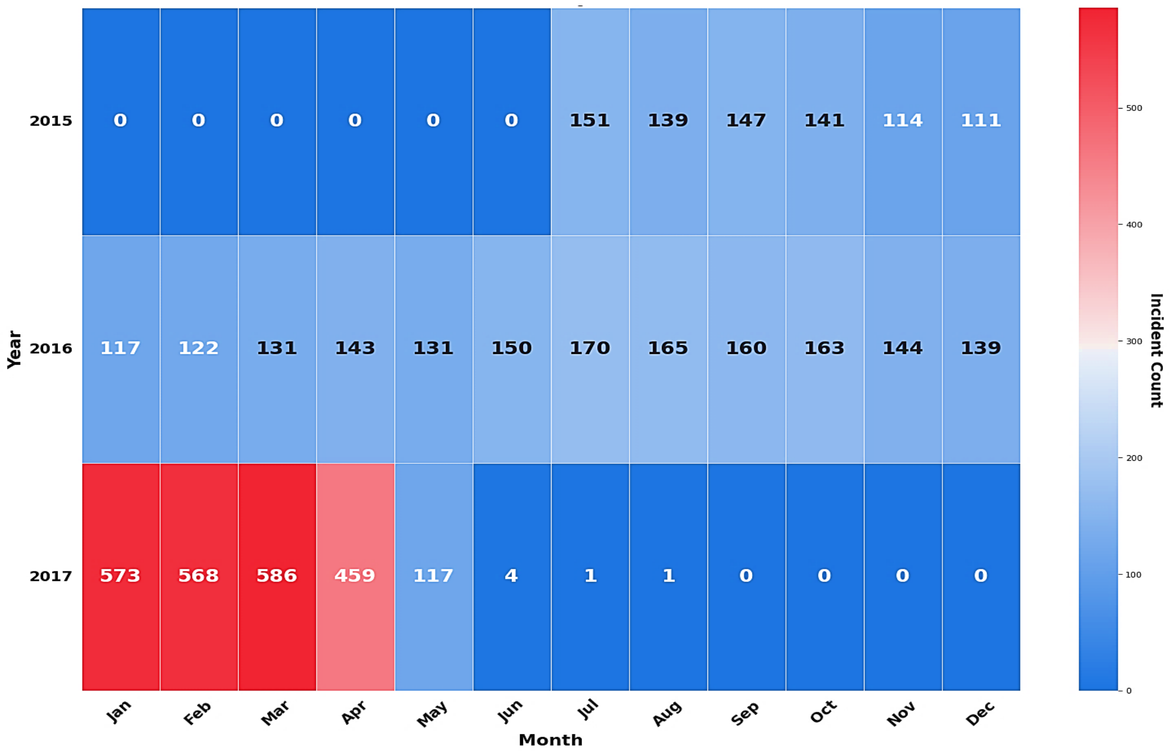


Figure 4-4: Distribution of Accidents and Incidents by Year and Month

Now let's do the visualization and analysis of some different factors that we see it have relation to the severity of incidents, provides valuable insights for understanding the patterns and risks associated with workplace incidents.

Figure 4-5, The visualization illustrates the impact of different types of injuries on their severity in construction sites. It showcases a range of injuries including serious falls/strikes, fractures, amputations/crushings, lacerations, head trauma, and bruising/contusions among others. By examining the total degree of injury for each category, we find an average of approximately 373 injuries per type. Among them, serious falls/strikes emerge as the most severe, accounting for a total degree of injury of 1683. On the other end of the spectrum, freezer burns, illnesses, and falls from elevation exhibit the lowest severity, each with a single incident and a degree of injury of 1.

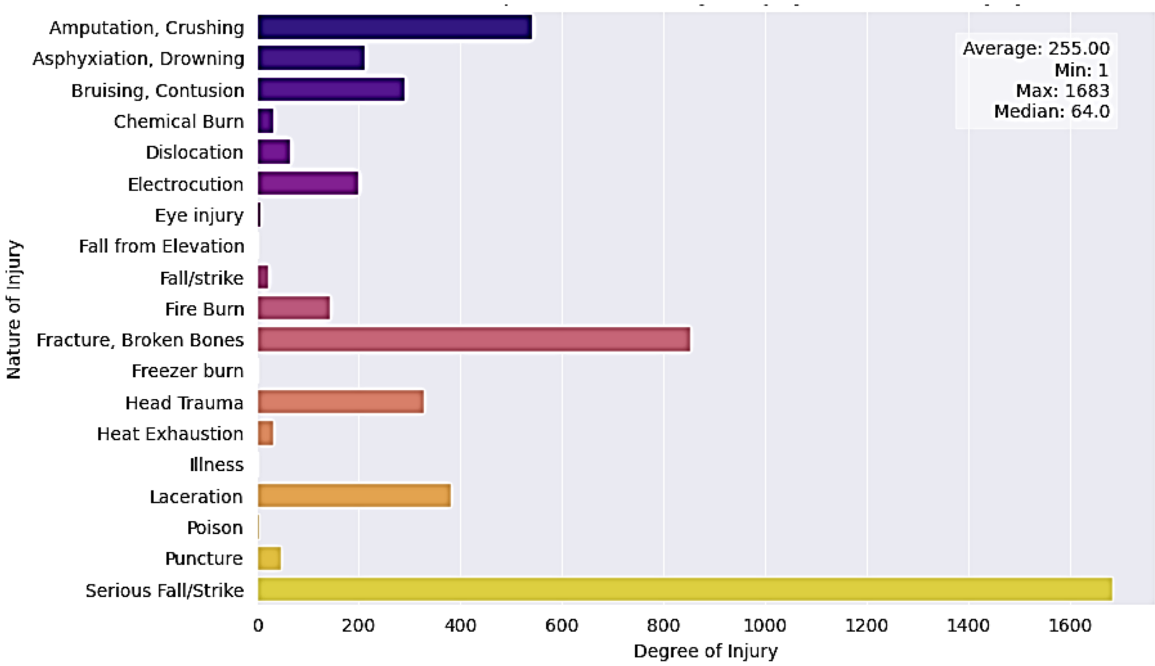


Figure 4-5: Correlation Between Severity of Injury and Nature of Injury

This analysis sheds light on the relationship between injury type and severity in construction sites, emphasizing the significance of addressing and preventing specific types of injuries that contribute to higher degrees of severity. Prominent examples include serious falls/strikes, fractures, and Amputations/ Crushings. By comprehending these associations, safety measures can be enhanced to mitigate the frequency and severity of injuries within construction environments.

Figure 4-6, Upon analyzing the Environmental Factor column, we observe that the "Other" category represents a significant portion of the data. However, due to its lack of specific information and classification, it poses a challenge for our objective of predicting incident severity. This ambiguity hinders our ability to determine the individual impact of these unidentified factors on the degree of injury.

To address this challenge and ensure valuable insights, we propose two potential approaches. Firstly, we can investigate the instances categorized as "Other" to identify subcategories that provide more specific and meaningful information. By doing so, we can gain a clearer understanding of how these factors relate to the severity of injuries. This approach would enable us to refine our analysis and make more accurate predictions. Alternatively, if it is not possible to subdivide the "Other" category effectively, we can consider excluding it from our analysis. By focusing solely on the known and well-defined environmental factors, we can ensure the relevance and reliability of our findings. This exclusion would allow us to prioritize actionable insights that contribute directly to predicting incident severity.

In summary, ensuring the quality and relevance of the data is crucial for achieving our objective. By refining the categorization or excluding the "Other" category, we can enhance the accuracy and effectiveness of our analysis in predicting incident severity based on the identified environmental factors.

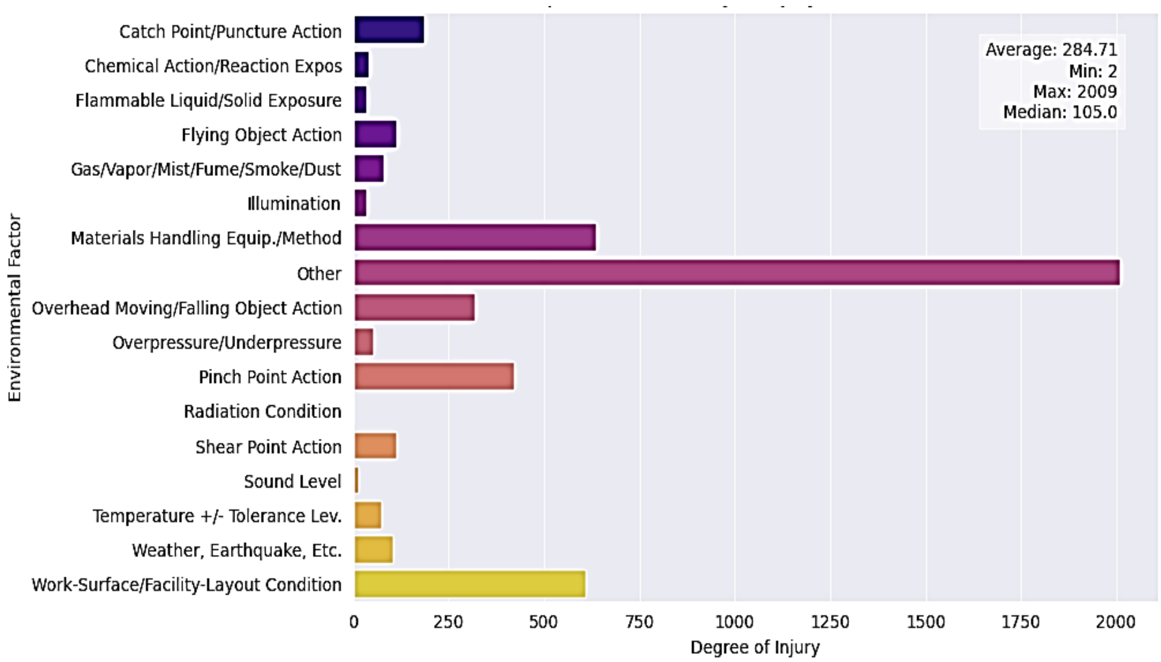


Figure 4-6: Correlation Between Severity of Injury and Environmental Factor

Figure 4-7, The bar plot provides valuable insights into the distribution of injuries across different body parts within the construction site context. The analysis reveals that the head is the most frequently affected body part, with 1023 reported injuries, emphasizing its vulnerability to hazards in construction environments.

Additionally, injuries involving the whole body and fingers are prevalent, with 555 and 550 reported cases, respectively. This suggests that injuries affecting multiple body parts or specific hand-related activities are common in construction work. Conversely, certain body parts, such as the kidney, exhibit a low incidence of injuries, indicating their relative safety in construction site incidents.

Understanding the distribution of injuries across different body parts allows for prioritizing safety measures and protocols. Focusing on protecting vulnerable areas, such as the head, whole body, and fingers, can effectively reduce the severity of injuries and enhance the well-being of construction workers. By addressing specific body parts prone to harm, safety measures can be tailored to mitigate risks and improve overall workplace safety.

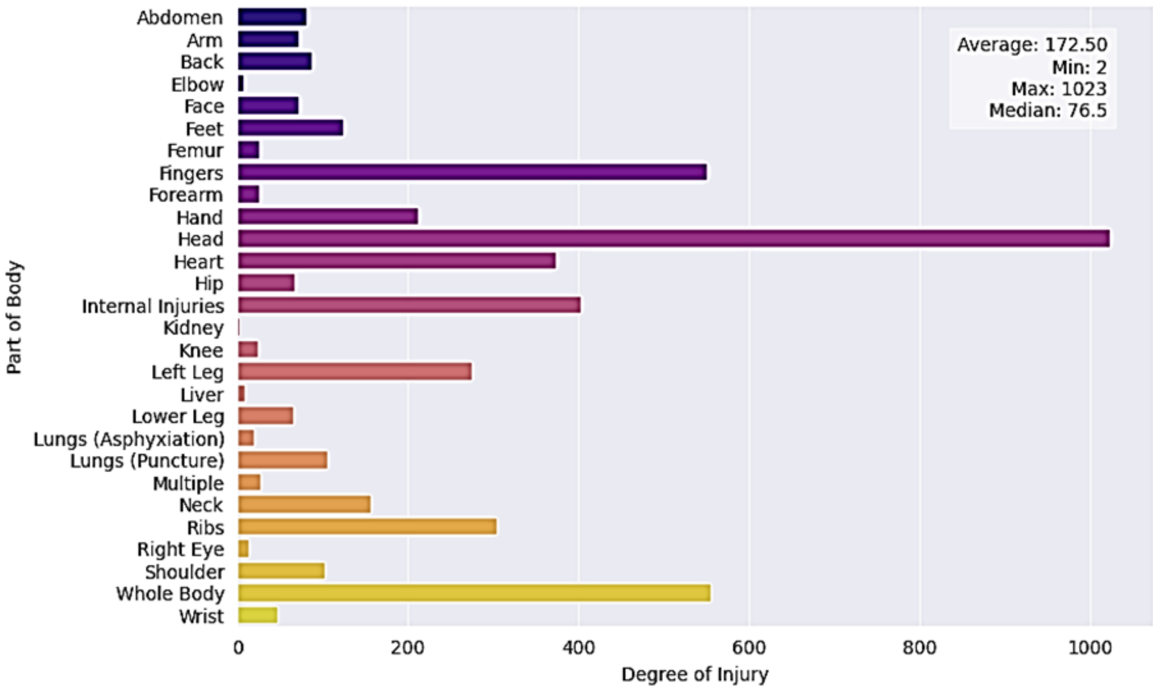


Figure 4-7: Correlation Between Severity of Injury and Part of the Body

Figure 4-8, The bar plot visualizes the relationship between the severity of injury and different event types in the context of construction sites. The severity, measured by the degree of injury, ranges from 23 to 1683. The data reveals that falls from elevation, being struck by objects, and being caught in or between objects are the event types with the highest degrees of injury, indicating a greater potential for severe harm. These event types have a significant impact on the overall severity of injuries in construction sites.

On average, the severity of injuries across all event types is approximately 341.79, with some cases resulting in minimal harm (e.g., degree of injury = 23) and others causing more serious consequences (e.g., degree of injury = 1683). The median degree of injury, representing the middle value, is 199, further highlighting the variability in injury severity.

Given these findings, it is crucial to prioritize risk mitigation strategies and implement effective safety measures, particularly for falls from elevation, being struck by objects, and being caught in or between objects. By addressing the specific challenges associated with these event types, construction sites can reduce the occurrence and severity of injuries, creating safer working environments for workers.

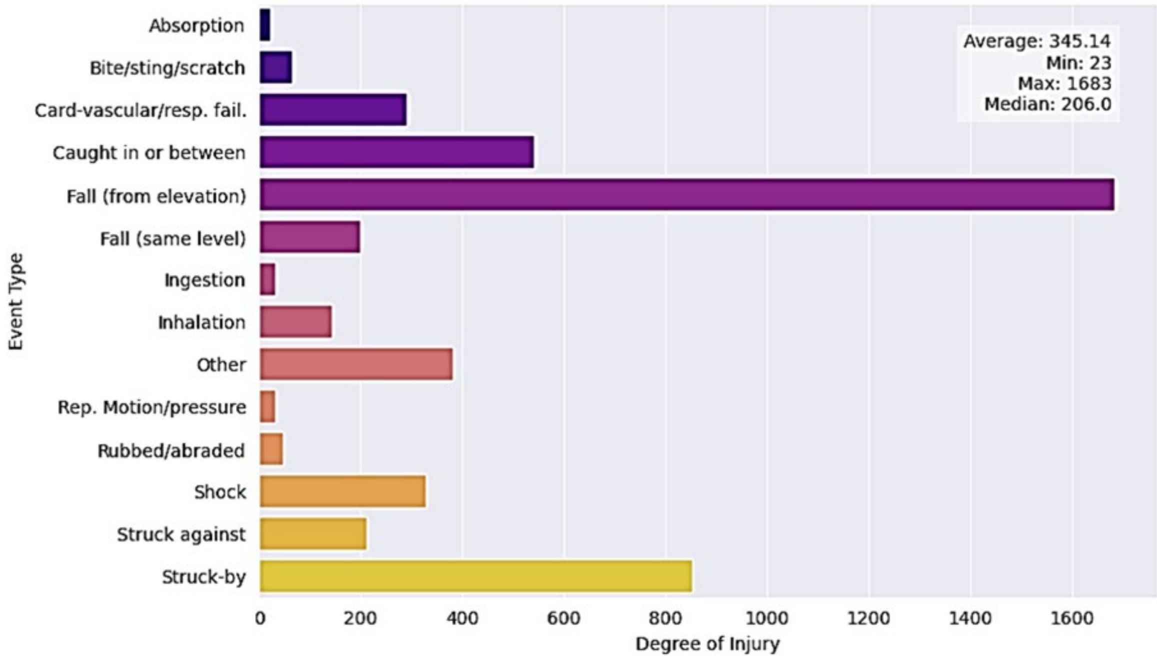


Figure 4-8: Correlation Between Severity of Injury and Event Type

```
#The Shape of the Dataset
osha.shape
```

(4847, 29)

The frequency distribution for the 'Degree of Injury' categories is as follows:

```
osha['Degree of Injury'].value_counts()
```

- **Fatal:** 2964 incidents
- **Non fatal:** 1883 incidents

This indicates that the majority of incidents in the dataset resulted in fatal injuries, while a smaller portion resulted in nonfatal injuries.

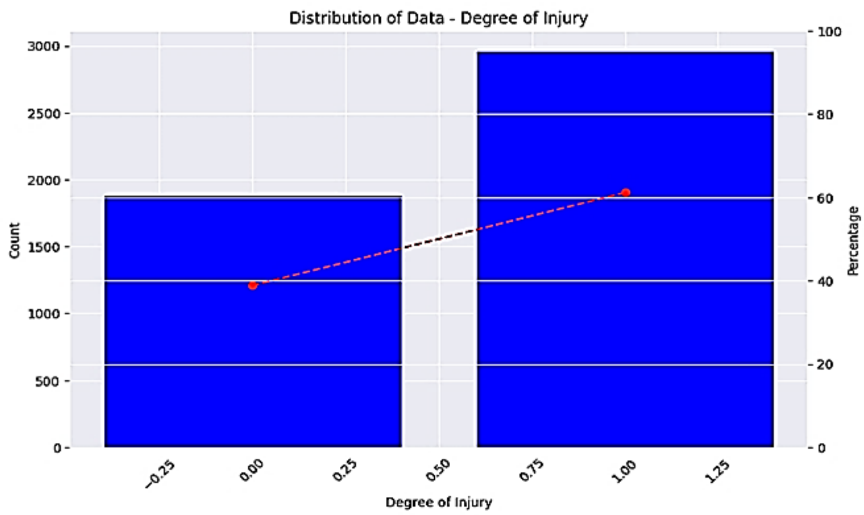


Figure 4-9: Distribution of Data - Degree of Injury

```
## Summary statistics
print('\nDescription of the Data :')
print(osha.describe())
```

The description provides statistics such as count, mean, standard deviation, minimum, 25th percentile, 50th percentile (median), and maximum values for each column. These statistics provide insights into the distribution and variation of the data within each column.

Table 4-2: Summary Statistics Output

	count	mean	std	min	25%	50%	75%	max
summary_nr	4.85E+03	2.21E+08	5.59E+04	2.21E+08	2.21E+08	2.21E+08	2.21E+08	**2.21E+08**
build_stor	4847	0.4473	3.0117	0	0	0	0	**139**
nature_of_inj	4847	11.8525	7.6343	0	5	12	21	**22**
part_of_body	4847	13.6293	7.8389	0	10	13	19	**31**
event_type	4847	5.1937	4.6083	0	2	5	6	**14**
evn_factor	4847	11.9837	6.1986	0	7	13	18	**18**
hum_factor	4847	9.457	6.1614	0	1	13	14	**20**
task_assigned	4847	1.3712	0.4832	1	1	1	2	**2**
fat_cause	4847	3.0656	7.2476	0	0	0	0	**30**
fall_ht	4847	0	0	0	0	0	0	**0**

```
#Checking for Missing values
print('\nMising values Counts : ')
print(osha.isnull().sum())
```

We have some missing values in the following columns:

- Nature of Injury: 2 missing values.
- Part of Body: 2 missing values.
- Event type: 2 missing values.
- Environmental Factor: 7 missing values.
- Human Factor: 7 missing values.

4.6 PRE-PROCESSING THE DATA

Is the process of involving a series of steps to transform and prepare the data for further analysis. These steps include one-hot encoding, imputation of missing values, feature scaling, and feature engineering.

One-Hot Encoding: Is used to convert categorical variables, such as the 'Environmental Factor' and 'Event type' columns, into binary features. This process creates new columns for each unique category, where a value of 1 indicates the presence of that category and 0 indicates its absence. This transformation allows the categorical data to be represented numerically and facilitates the inclusion of these variables in analytical models.

```
encoder = OneHotEncoder(sparse=False)
```

Imputation: Is employed to handle missing values in the dataset. Like in the 'Environmental Factor' column, is imputed using the most frequent value strategy, which replaces missing values with the category that appears most frequently in the column. This ensures that the imputed values align with the majority of the data and reduces the impact of missing values on subsequent analysis.

```
imputer = SimpleImputer(strategy='most_frequent')
```

Feature Scaling: Is applied to ensure that the numerical features in the dataset have a consistent scale. Such as the 'Environmental Factor' column after imputation, which scales the data to have zero mean and unit variance. This step is important to avoid biases that can arise from features with different scales and to improve the performance of certain Machine Learning algorithms.

```
scaler = StandardScaler()
```

Feature Engineering: Such as the 'Part of Body' column to create a new column called 'BodyPartGroup'. This process categorizes body parts into groups based on their anatomical regions. The 'BodyPartGroup' column is then one-hot encoded to represent each group as binary features. This feature engineering step enhances the dataset by introducing additional meaningful information and capturing potential patterns related to different body part groups.

These preprocessing techniques facilitate the standardization of data, effective handling of missing values, conversion of categorical variables into numerical forms, and derivation of meaningful additional features. These steps ensure that the dataset is well-prepared for more precise and dependable analysis and modeling endeavors.

4.7 FEATURES SELECTION

4.7.1 Extract and Separating Features and Target

```
#Extarct our Target
X = osha[[''nature_of_inj', 'part_of_body', 'event_type', 'evn_factor', 'hum_factor',
'task_assigned', 'fat_cause',…']] #we have now 105 columns as features
Y = osha['Degree of Injury']
```

First, we extract the target (features) columns from our dataset. These columns includer 'nature_of_inj', 'part_of_body', 'event_type', 'evn_factor', 'hum_factor', 'task_assigned', 'fat_cause', and several environmental factors such as 'Environmental Factor_Catch Point/Puncture Action', 'Environmental Factor_Chemical Action/Reaction Expos', 'Environmental Factor_Flammable Liquid/Solid Exposure', and others.

Additionally, we extract columns related to different types of injuries such as 'Absorption', 'Bite/sting/scratch', 'Card-vascular/resp. fail.', 'Caught in or between', 'Fall (from elevation)', 'Fall (same level)', 'Ingestion', 'Inhalation', 'Other', 'Rep. Motion/pressure', 'Rubbed/abraded', 'Shock', 'Struck against', 'Struck-by', and body part groupings like

'BodyPartGroup_Extremities', 'BodyPartGroup_Head and Neck', 'BodyPartGroup_Lower Body', 'BodyPartGroup_Other', 'BodyPartGroup_Upper Body'.

Furthermore, we include additional sets of environmental factors and body part groupings, as well as columns related to the average degree of injury. We assign the extracted columns to the variable X, which represents the input features for our analysis.

Next, we assign the column 'Degree of Injury' from our dataset to the variable Y, which represents the target variable we want to predict or analyze. By extracting these specific columns, we aim to focus on relevant information related to the nature of injuries, body parts affected, environmental factors, the degree of injury etc...

4.8 APPLICATION OF MACHINE LEARNING

We trained each algorithm on the preprocessed dataset using appropriate training and validation strategies. This involved splitting the data into training and validation sets, and then fitting the model on the training data while evaluating its performance on the validation data.

4.8.1 Evaluation Metrics for Comparing Machine Learning Algorithms

When evaluating Machine Learning algorithms, we rely on various performance metrics to assess their effectiveness. These metrics provide valuable insights into the algorithm's performance and allow us to compare different models. Let's explore some commonly used evaluation metrics:

Confusion Matrix: The confusion matrix is a tabular representation utilized to evaluate the effectiveness of a classification model.[44]

		Predicted	
		FALSE	TRUE
Actual	FALSE	**TN = 2**	**FP =1**
	TRUE	**FN = 0**	**TP = 6**

Figure 4-10: Confusion Matrix

- True Negatives (TN): Actual FALSE, which was predicted as
- FALSE
- False Positives (FP): Actual FALSE, which was predicted as TRUE
- (Type I error)
- False Negatives (FN): Actual TRUE, which was predicted as FALSE
- (Type II error)
- True Positives (TP): Actual TRUE, which was predicted as TRUE

Ideally a good model should have high TN and TP and less of Type I & II errors. Table 4-5 describes the key metrics derived out of the confusion matrix to understand the classification model performance.[44]

Table 4-3: Classification Performance Matrices

Metric	Description	Formula
Accuracy	What % of predictions were correct?	(TP+TN)/(TP+TN+FP+FN)
Misclassification Rate	What % of prediction is wrong?	(FP+FN)/(TP+TN+FP+FN)
True Positive Rate OR Sensitivity OR Recall (completeness)	What % of positive cases did model catch?	TP/(FN+TP)
False Positive Rate	What % of 'No' were predicted as 'Yes'?	FP/(FP+TN)
Specificity	What % of 'No' were predicted as 'No'?	TN/(TN+FP)
Precision (exactness)	What % of positive predictions were correct?	TP/(TP+FP)
F1 score	Weighted average of precision and recall	2*((precision * recall) / (precision + recall))

ROC Curve: A ROC curve is one more important metric, and it's a most commonly used way to visualize the performance of a binary classifier, and AUC is believed to be one of the best ways to summarize performance in a single number. AUC indicates that the probability of a randomly selected positive example will be scored higher by the classifier than a randomly selected negative example. In case of multiple models with nearly the same accuracy, we can pick the one that gives a higher AUC.[44]

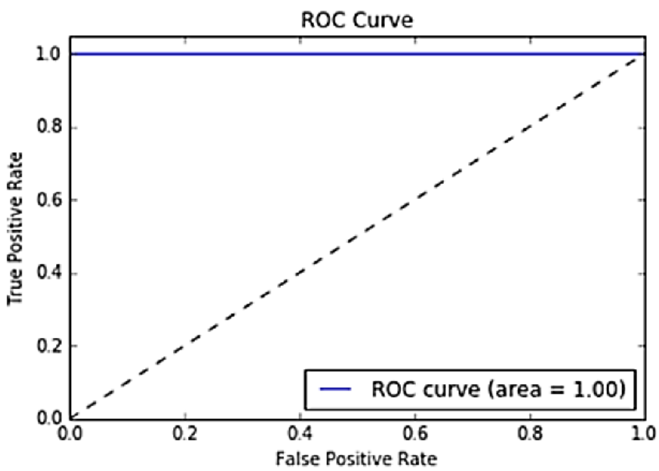


Figure 4-11: ROC Curve

Mean Squared Error (MSE): Measures the average squared difference between the predicted and actual values. It gives us an idea of how close the predicted values are to the actual values. A lower MSE indicates better accuracy and a better fit of the model to the data.[65]

Mean Absolute Error (MAE): Calculates the average absolute difference between the predicted and actual values. It provides a measure of the average magnitude of the errors. Like MSE, a lower MAE indicates better accuracy and a better fit of the model.[65]

R-squared (R2) Score: Represents the proportion of the variance in the dependent variable that can be explained by the independent variables. It ranges from 0 to 1, with 1 indicating a perfect fit. A higher R2 score signifies that the model captures a larger portion of the target variable's variability.[65]

Classification Report: Is used for evaluating classification models. It provides metrics such as precision, recall, and F1-score for each class, along with the overall accuracy. Precision measures the proportion of correctly predicted positive instances, recall calculates the proportion of actual positive instances correctly identified, and the F1-score combines both precision and recall.[66]

By utilizing these evaluation metrics, we can quantitatively assess the performance of Machine Learning algorithms. They enable us to compare different models and make informed decisions about which algorithm is the most suitable for our specific task.

4.8.2 Splitting The Data

```
#Split the data into training & testing dataset
X_train, X_test, Y_train, Y_test = train_test_split(X_scaled, Y, test_size=0.3,
random_state=2)
```

First, we split the dataset into training and testing datasets using the `**train_test_split**` function from the scikit-learn library.

We pass the scaled input features `**X_scaled**` and the target variable `Y` to the function. The `**test_size**` parameter is set to 0.3, indicating that we want to allocate 30% of the data for testing, while the remaining 70% will be used for training.

Additionally, we set the `**random_state**` parameter to 2 to ensure reproducibility. This random state value is used to shuffle the data before splitting it, so by setting it to a specific value, we can obtain the same train-test split each time we run the code.

The function returns four sets of data: `**X_train**`, `**X_test**`, `**Y_train**`, and `**Y_test**`.

- `**X_train**` contains the training data for the input features, which is 70% of the scaled data.
- `**X_test**` contains the testing data for the input features, which is 30% of the scaled data.
- `**Y_train**` contains the training data for the target variable, which corresponds to the degree of injury. It is 70% of the target variable data.

- `**Y_test**` contains the testing data for the target variable, corresponding to the degree of injury. It is 30% of the target variable data.

This split allows us to train a model using the training data and evaluate its performance on unseen data using the testing data.

4.8.3 Support Vector Machine Classifier (SVC)

We chose the SVM algorithm because of its capability to handle non-linear relationships, high-dimensional data, and outliers. With SVMs, we can effectively perform binary classification tasks, which aligns with our goal of predicting incident severity. Additionally, SVMs have been successfully applied in various domains, showcasing their reliability and versatility.

4.8.3.1 Create and Train the Support Vector Machine Classifier (SVC) Algorithm

```
from sklearn.svm import SVC
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.metrics import classification_report
from sklearn import metrics
from sklearn.metrics import roc_auc_score, roc_curve
#Create the SVM classifier
clf2 = SVC(kernel='linear', C=1.0, random_state=0)
#Train the classifier
clf2.fit(X_train, Y_train)
# Make predictions on the test
Y_pred_clf2 = clf2.predict(X_test)
#Generate evaluation metrics on the training set
train_accuracy = metrics.accuracy_score(Y_train, clf2.predict(X_train))
train_confusion_matrix = metrics.confusion_matrix(Y_train, clf2.predict(X_train))
train_classification_report = metrics.classification_report(Y_train,
clf2.predict(X_train))
#Generate evaluation metrics on the testing set
test_accuracy = metrics.accuracy_score(Y_test, clf2.predict(X_test))
test_confusion_matrix = metrics.confusion_matrix(Y_test, clf2.predict(X_test))
test_classification_report = metrics.classification_report(Y_test,
clf2.predict(X_test))
# Calculate AUC and ROC for training data
train_probs = clf2.decision_function(X_train)
train_auc = roc_auc_score(Y_train, train_probs)
train_fpr, train_tpr, _ = roc_curve(Y_train, train_probs)
# Calculate AUC and ROC for testing data
test_probs = clf2.decision_function(X_test)
test_auc = roc_auc_score(Y_test, test_probs)
test_fpr, test_tpr, _ = roc_curve(Y_test, test_probs)
# Calculate evaluation metrics
mse_clf2 = mean_squared_error(Y_test, Y_pred_clf2)
mae_clf2 = mean_absolute_error(Y_test, Y_pred_clf2)
r2_clf2 = r2_score(Y_test, Y_pred_clf2)
```

Explanation:

In this code, we are performing classification using Support Vector Classifier (SVC). We import the necessary modules and functions from scikit-learn to calculate evaluation metrics and create classification reports. First, we create an SVM classifier by instantiating the SVC class with the desired parameters. In this case, we choose a linear kernel and set the regularization parameter C to 1.0. Next, we train the classifier using the **fit()** method by passing the training data **X_train** and corresponding labels **Y_train**.

To evaluate the model's performance, we make predictions on the test set (**X_test**) using the **predict()** method, and store the predicted labels in **Y_pred_clf2**. We then generate evaluation metrics for both the training and testing sets. The **accuracy_score()** function from the **metrics** module is used to calculate the accuracy of the predictions. The **confusion_matrix()** function is used to compute the confusion matrix, which provides information about true positives(TP), true negatives(TN), false positives(FP), and false negatives(FN). The **classification_report()** function generates a report with precision, recall, and F1-score for each class.

Additionally, we calculate the Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) for both the training and testing sets. The **decision_function()** method of the SVM classifier returns the decision scores, which are then used to compute the AUC using the **roc_auc_score()** function. The **roc_curve()** function generates the false positive rate (FPR) and true positive rate (TPR) for creating the ROC curve.

Finally, we calculate the Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2) score using the respective functions from the **metrics** module. These metrics provide insights into the accuracy and performance of the SVM classifier in predicting the incident severity. In summary, this code demonstrates the training and evaluation process for an SVM classifier.

4.8.3.2 Support Vector Machine Classifier Performance and Results

Table 4-4: SVM Classifier Performance

Metric	Training Set	Testing Set
AUC	0.9763	0.9706
Mean Squared Error	-	0.0845
Mean Absolute Error	-	0.0845
R-squared (R2) Score	-	0.6503
Accuracy	0.9304	0.9155
Confusion Matrix		
- True Negative (TN)	1169	524
- False Positive (FP)	119	71
- False Negative (FN)	117	52
- True Positive (TP)	1987	808
Classification Report		

Metric	Training Set	Testing Set
- Precision (Class 0)	0.9100	0.91
- Precision (Class 1)	0.9400	0.92
- Recall (Class 0)	0.9100	0.88
- Recall (Class 1)	0.9400	0.94
- F1-score (Class 0)	0.9100	0.89
- F1-score (Class 1)	0.9400	0.93
Support	3392	1455

In the SVC algorithm, we achieved high performance in both the training and testing sets. The AUC values of 0.9763 for the training set and 0.9706 for the testing set indicate that the model has excellent discriminative ability in distinguishing between the two classes. The high AUC scores suggest that the model's predicted probabilities are well-calibrated.

When evaluating the model's accuracy, we obtained an accuracy of 0.9304 on the training set and 0.9155 on the testing set. This means that the model correctly predicted the class labels for a significant portion of the instances in both sets. Looking at the confusion matrix, we can see that the model achieved a high number of true negatives (TN) with 1169 in the training set and 524 in the testing set. The false positive (FP) and false negative (FN) values are relatively low, indicating that the model made fewer incorrect predictions.

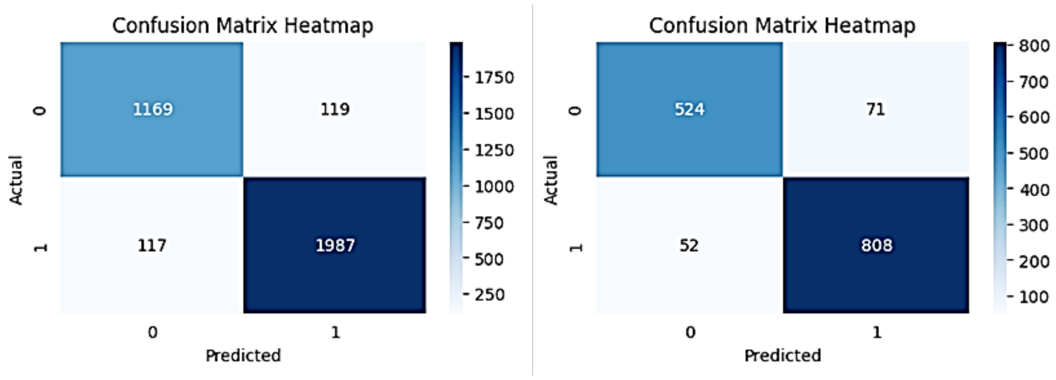


Figure 4-12: SVM Classifier Confusion Matrix Heatmap

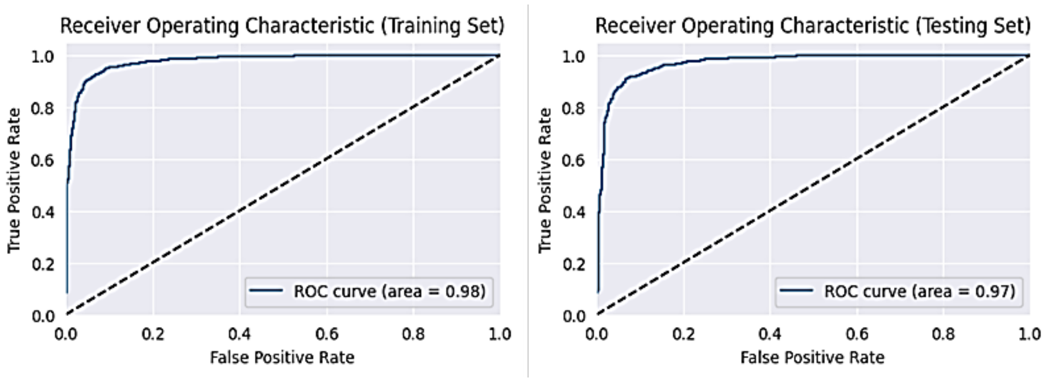


Figure 4-13: Plot ROC Curve - SVM Classifier

Examining the precision, recall, and F1-score metrics, we observe consistently high values for both classes (0 and 1) in both sets. The precision values of 0.9100 for class 0 and 0.9400 for class 1 indicate that the model has a high ability to correctly identify positive instances. The recall values of 0.9100 for class 0 and 0.9400 for class 1 indicate that the model effectively captures a high proportion of the actual positive instances. The F1-score values reflect a balanced combination of precision and recall.

The support values indicate the number of instances for each class in the dataset, with 3392 instances in the training set and 1455 instances in the testing set.

In general, the (SVM) Classifier algorithm demonstrates strong performance in terms of accuracy, AUC, precision, recall, and F1-score. It effectively classifies instances into the correct classes and shows a high level of predictive power. The low mean squared error, mean absolute error, and the relatively high R-squared (R2) score further indicate the model's good fit to the data and its ability to explain the variance in the target variable.

Note: The following algorithms will follow a similar methodology to the Support Vector Machine (SVM) in terms of calculating evaluation metrics and generating classification reports. However, we will focus on explaining the creation of the algorithm part for each. In summary, while the calculation of evaluation metrics and generation of classification reports remain consistent across models

4.8.4 K-Nearest Neighbors (KNN) Classifier

We chose the K-nearest Neighbors (KNN) algorithm for our work because it offers simplicity, effectiveness, and the ability to capture complex relationships in the data. KNN is a non-parametric algorithm that works based on the similarity of instances and does not make assumptions about data distribution. It is suitable for classification tasks and can handle both numerical and categorical features. The KNN algorithm is robust to outliers and can capture nonlinear relationships. Considering the characteristics of our dataset, including diverse feature types and potential complex patterns, we found KNN to be a valuable choice.

4.8.4.1 Create and Train the K-Nearest Neighbors (KNN) Classifier

```python
from sklearn.neighbors import KNeighborsClassifier
#Creat and train the KNN model :
KNN = KNeighborsClassifier(n_neighbors=5, p=2, metric='minkowski')
KNN.fit(X_train, Y_train)
```

Explanation:

First, we import the necessary class from scikit-learn: KNeighbors-Classifier. We create an instance of the classifier, specifying the desired parameters such as the number (n_neighbors=5) This specifies that the algorithm will consider the 5 nearest neighbors when making predictions. It is a commonly used value, but it can be adjusted based on the specific problem and dataset. p=2: This parameter indicates the power parameter for the Minkowski distance metric. In this case, p=2 corresponds to using the Euclidean distance metric. Other distance metrics can be used by modifying this parameter. metric='minkowski': This parameter defines the distance metric used to measure the similarity between instances. In our case, we choose the Minkowski distance, which generalizes other distance metrics like Euclidean and Manhattan. Next, we train the KN Classifier algorithm on the training data by calling the fit() function and passing the training features (X_train) and labels (Y_train).

To evaluate the performance of the model, we generate various evaluation metrics on both the training and testing sets just like in the SVM model.

4.8.4.2 K-Nearest Neighbors (KNN) Classifier Performance and Results

Table 4-5: KNN Classifier Performance

Metric	Training Set	Testing Set
AUC	0.9771	0.9374
Mean Squared Error	-	0.1251
Mean Absolute Error	-	0.1251
R-squared (R2) Score	-	0.4825
Accuracy	0.9172	0.8749
Confusion Matrix		
- True Negative (TN)	1141	495
- False Positive (FP)	147	100
- False Negative (FN)	134	82
- True Positive (TP)	1970	778
Classification Report		
- Precision (Class 0)	0.8900	0.86
- Precision (Class 1)	0.9300	0.89
- Recall (Class 0)	0.8900	0.83
- Recall (Class 1)	0.9400	0.9
- F1-score (Class 0)	0.8900	0.84
- F1-score (Class 1)	0.9300	0.89
Support	3392	1455

In the KNN model, we obtained reasonably good performance on both the training and testing sets. The AUC values of 0.9771 for the training set and 0.9374 for the testing set indicate that the model has good discriminatory power in distinguishing between the two classes, although it is slightly lower than the SVM model. The accuracy of the KNN model was 0.9172 on the training set and 0.8749 on the testing set. This means that the model correctly predicted the class labels for a significant portion of the instances in both sets, although the accuracy is slightly lower than the SVM model. Looking at the confusion matrix, we can see that the KNN model achieved a high number of (TN) with 1141 in the training set and 495 in the testing set. However, there are relatively higher (FP) and (FN) values compared to the SVM model.

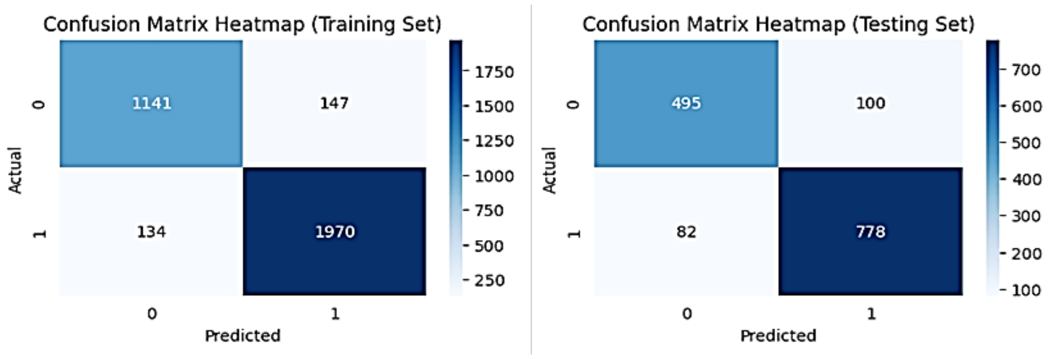


Figure 4-14: KNN Classifier Confusion Matrix Heatmap

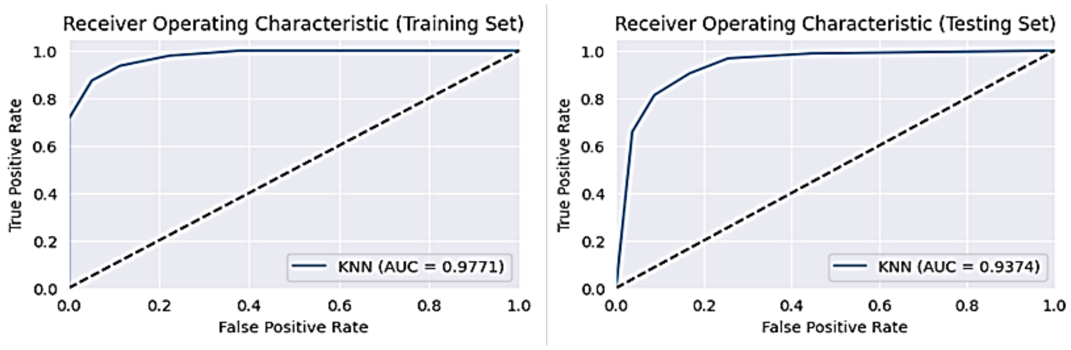


Figure 4-15: Plot ROC Curve - KNN Classifier

Examining the precision, recall, and F1-score metrics, we observe relatively high values for both classes (0 and 1) in both sets. The precision values of 0.8900 for class 0 and 0.9300 for class 1 in the training set indicate that the model has a good ability to correctly identify positive instances. The recall values of 0.8900 for class 0 and 0.9400 for class 1 in the training set indicate that the model captures a good proportion of the actual positive instances. The F1-score values reflect a balanced combination of precision and recall.

The KNN model shows higher MSE, MAE, and lower (R2) score compared to the SVM model, suggesting that it may not fit the data as well.

Overall, the KNN model demonstrates reasonable performance with respect to accuracy, precision, recall, and F1-score. However, it shows lower discriminative ability (AUC) and lower predictive power compared to the SVM model. The model's performance may benefit from further refinement or exploration of alternative algorithms.

4.8.5 Random Forest Classifier (RFC)

We chose the (RFC) because it is an ensemble learning method that combines multiple decision trees, allowing us to capture complex relationships and reduce the risk of overfitting. Additionally, RFC is robust to outliers and provides insights into feature importance, making it suitable for our analysis. Its ability to handle categorical variables without explicit encoding adds further convenience.

4.8.5.1 Create and Train the (RF) Classifier

```
from sklearn.ensemble import RandomForestClassifier
# Create a Random Forest classifier (RFC) model
rfc_model = RandomForestClassifier(n_estimators=100, random_state=42)
rfc_model.fit(X_train, Y_train)
```

Explanation:

To create the RFC we import the necessary class from scikit-learn: RandomForestclassifier. Create an instance of the RFC with 100 estimators (number of decision trees) and set the random_state parameter to ensure reproducibility of results. Then, we fit the model to the training data using the fit() method. And then do the calculation.

4.8.5.2 Random Forest Classifier (RFC) Performance and Results

Table 4-6: RF Classifier Performance

Metric	Training Set	Testing Set
AUC	1	0.9705
Mean Squared Error (MSE)	-	0.0708
Mean Absolute Error (MAE)	-	0.0708
R-Squared (R2) Score	-	0.7071
Accuracy	1	0.9292
Confusion Matrix		
- True Negative (TN)	1288	538
- False Positive (FP)	0	57
- False Negative (FN)	0	46
- True Positive (TP)	2104	814
Classification Report		
- Precision (Class 0)	1.0000	0.92
- Precision (Class 1)	1.0000	0.93
- Recall (Class 0)	1.0000	0.9
- Recall (Class 1)	1.0000	0.95
- F1-Score (Class 0)	1.0000	0.91
- F1-Score (Class 1)	1.0000	0.94
Support	3392	1455

In terms of AUC, the model achieved a perfect score of 1.0 on the training set, indicating excellent discriminatory power in distinguishing between the two classes. On the testing set, the AUC was 0.9705, which suggests that the model performs well in classifying instances correctly. The (MSE) and (MAE) were both low, indicating that the model's predictions have minimal deviation from the true values. The (R2) score of 0.7071 suggests that approximately 70.71% of the variance in the target variable can be explained by the model, indicating a reasonable fit to the data. The model achieved high accuracy on both the training set (1.0) and the testing set (0.9292), indicating that it correctly classified a significant proportion of the instances. The confusion matrix provides that the model correctly classified 1288 true negatives and 2104 true positives in the training set, as well as 538 true negatives and 814 true positives in the testing set. However, it made 57 false positive errors and 46 false negative errors in the testing set. The classification report provides further insights into the precision, recall, and F1-score for each class. In both the training and testing sets, the model achieved perfect precision for class 0, indicating that it rarely misclassified instances as class 0. The precision for class 1 was also high, with values of 1.0 in the training set and 0.93 in the testing set, suggesting that the model correctly identified a large portion of class 1 instances. The recall (sensitivity) values indicate that the model identified nearly all instances of class 0, with a recall of 1.0 in the training set and 0.9 in the testing set. The recall for class 1 was also high, with values of 1.0 in the training set and 0.95 in the testing set. The F1-scores, which consider both precision and recall, were also high for both classes in both sets.

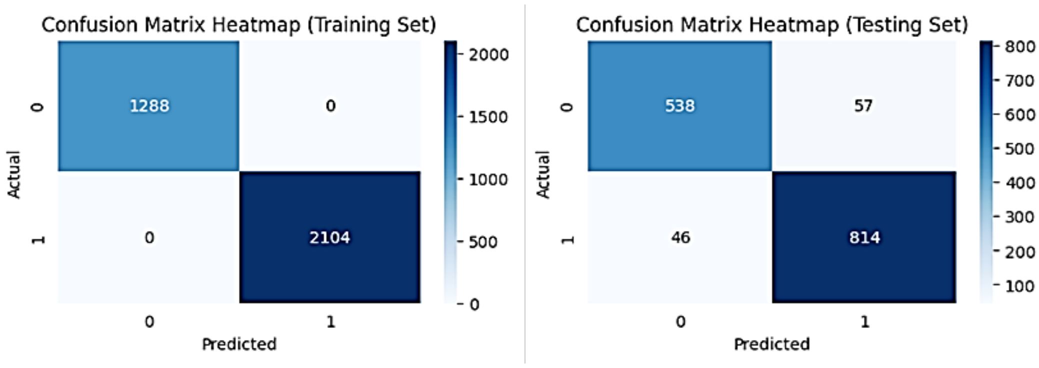


Figure 4-16: RF Classifier Confusion Matrix Heatmap

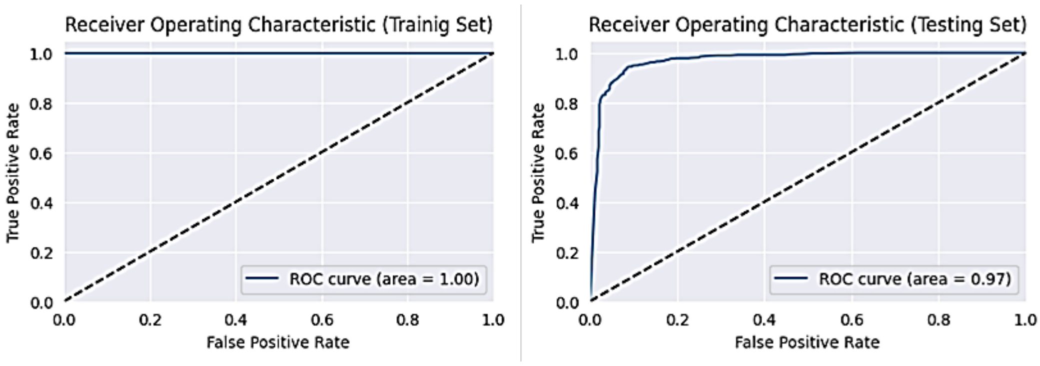


Figure 4-17: Plot ROC Curve - RF Classifier

In summary, the Random Forest Classifier demonstrated exceptional performance across various metrics, including AUC, accuracy, precision, recall, and F1-score. It effectively distinguished between the two classes, providing reliable predictions for the given dataset.

4.8.6 The Logistic Regression (LR)

We chose the (LR) model for our work due to its well-established nature, interpretability, and suitability for our dataset's characteristics. LR allows us to understand the impact of features on incident severity and handle a combination of numerical and categorical variables. Its computational efficiency and scalability make it a practical choice for large datasets.

4.8.6.1 Create and Train the Logistic Regression (LR)

```python
# Creat and train the Logistic Regression (LR) model
from sklearn.linear_model import LogisticRegression

lr_model = LogisticRegression()
lr_model.fit(X_train, Y_train)
```

Explanation:

In this code snippet, we are creating and training a (LR) model, start by importing the LogisticRegression class from the sklearn.linear_model module. Next, we instantiate an instance of the LogisticRegression class and assign it to the variable lr_model. This object will serve as our LR model. Then train the LR model using the fit() method, passing in the training data X_train and the corresponding labels Y_train. This step involves finding the optimal coefficients for the LR model that best fit the training data and allow us to make accurate predictions. By executing this code, we are effectively training the LR model on our dataset.

4.8.6.2 Logistic Regression (LR) Performance and Results

Table 4-7: LR Performance

Metric	Training Set	Testing Set
AUC	0.9768	0.9707
Mean Squared Error (MSE)	-	0.0866
Mean Absolute Error (MAE)	-	0.0866
R-Squared (R2) Score	-	0.6417
Accuracy	0.9287	0.9134
Confusion Matrix		
- True Negative (TN)	1170	525
- False Positive (FP)	118	70
- False Negative (FN)	124	56
- True Positive (TP)	1980	804
Classification Report		
- Precision (Class 0)	0.9000	0.9
- Precision (Class 1)	0.9400	0.92
- Recall (Class 0)	0.9100	0.88
- Recall (Class 1)	0.9400	0.93

Metric	Training Set	Testing Set
- F1-Score (Class 0)	0.9100	0.89
- F1-Score (Class 1)	0.9400	0.93
Support	3392	1455

The LR model achieved promising results, with an AUC of 0.9768 on the training set and 0.9707 on the testing set. This indicates that the model has good discriminatory power in distinguishing between different severity levels. Exhibited high accuracy, with an accuracy of 0.9287 on the training set and 0.9134 on the testing set. This suggests that the model is effective in correctly classifying the severity of incidents in both datasets.

Examining the confusion matrix, we see that the LR model achieved a high number of true negatives (TN) and true positives (TP), indicating its ability to correctly predict both low and high incidents severity. The model's precision, recall, and F1-scores are also consistently high for both classes, demonstrating its balanced performance in classifying incidents of high severity levels.

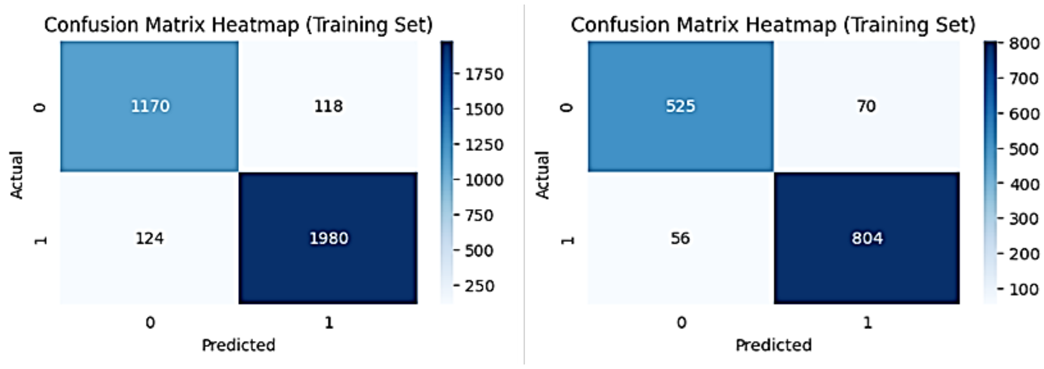


Figure 4-18: LR Confusion Matrix Heatmap

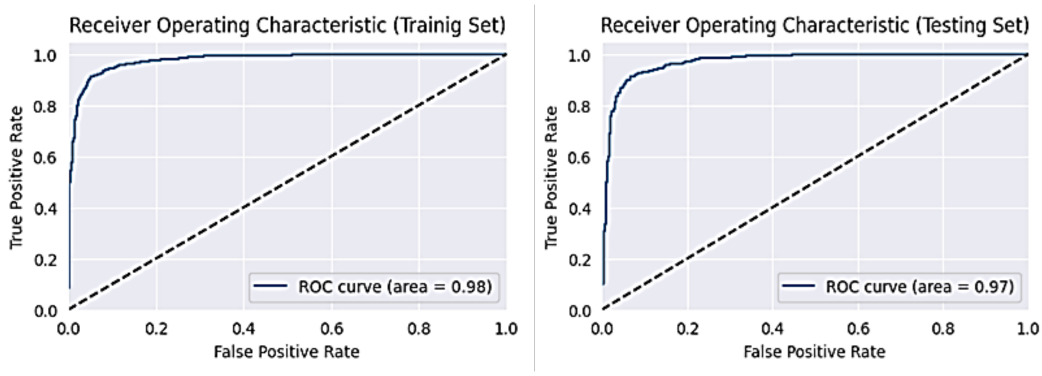


Figure 4-19: Plot ROC Curve - LR

The LR model demonstrates strong predictive capabilities in identifying incident severity. Its accuracy, AUC, and robust performance across the various evaluation metrics.

4.8.7 Multilayers Perception (MLP)

4.8.7.1 Create and Train the Multilayers Perception (MLP) Classifier

```
1. from sklearn.neural_network import MLPClassifier
2. #Creat and train the (MLP) model :
3. MLP_model = MLPClassifier(hidden_layer_sizes=(256, 128, 64), activation='relu',
random_state=42)
4. MLP_model.fit(X_train,Y_train)
```

Explanation:

We incorporated a Multilayer Perceptron MLP Classifier model from the scikit-learn library. The MLP model is a type of artificial neural network that consists of multiple layers of interconnected nodes or neurons. It is trained using backpropagation to learn complex patterns and relationships within the data. For our MLP model, we specified a hidden layer architecture with three layers, having 256, 128, and 64 neurons respectively. The activation function used in each neuron is the rectified linear unit (ReLU), which helps introduce non-linearity into the model and enables it to learn complex decision boundaries. During the training process, we fitted the MLP model to the training data using the fit method. This process involved adjusting the weights and biases of the network to minimize the difference between the predicted labels and the true labels in the training set.

4.8.7.2 Multilayers Perception (MLP) Classifier Performance and Results

Table 4-8: MLP Classifier Performance

Metric	Training Set	Testing Set
AUC	0.9998	0.9593
Mean Squared Error	-	0.1038
Mean Absolute Error	-	0.1038
R-Squared (R2) Score	-	0.5706
Accuracy	0.9938	0.8962
Confusion Matrix		
- True Negative (TN)	1275	514
- False Positive (FP)	13	81
- False Negative (FN)	8	70
- True Positive (TP)	2096	790
Classification Report		
- Precision (Class 0)	0.99	0.88
- Precision (Class 1)	0.99	0.91
- Recall (Class 0)	0.99	0.86
- Recall (Class 1)	1.00	0.92
- F1-Score (Class 0)	0.99	0.87
- F1-Score (Class 1)	1.00	0.91
Support	3392	1455

The model achieved high accuracy, correctly predicting the severity of incidents in the training set with an accuracy of 99.38% and in the testing set with an accuracy of 89.62%. The (MAE) was 0.1038 for both the training and testing sets. This indicates that, on average, the predicted incident severity values deviate from the actual values by approximately 0.1038. The AUC values, were 0.9998 for the training set and 0.9593 for the testing set. These values suggest that the model has excellent discriminatory power and can effectively differentiate between incident severity levels. Looking at the confusion matrix, we can see that the model made a small number of false predictions. In the training set, it correctly identified 1275 incidents as non-severe (true negatives) and 2096 incidents as severe (true positives). In the testing set, it correctly identified 514 non-severe incidents and 790 severe incidents. The classification report provides additional insights into the model's performance. It shows high precision, recall, and F1-score values for both classes in both the training and testing sets, indicating a balanced performance between identifying non-severe and severe incidents.

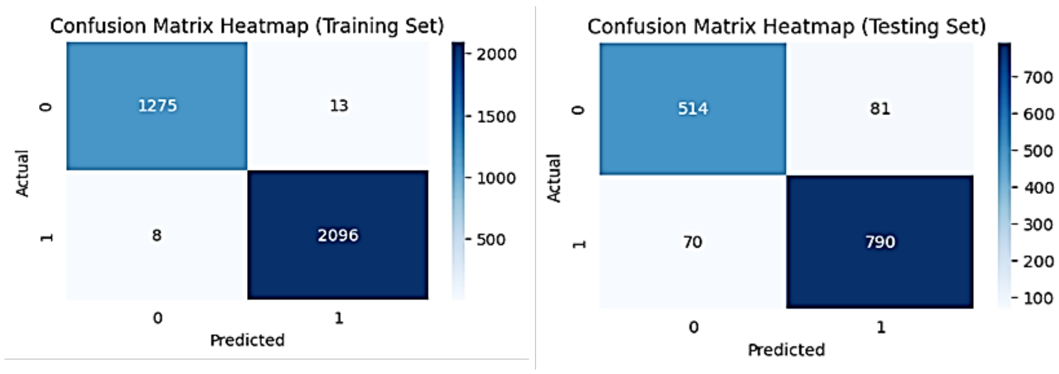


Figure 4-20: MLP Classifier Confusion Matrix Heatmap

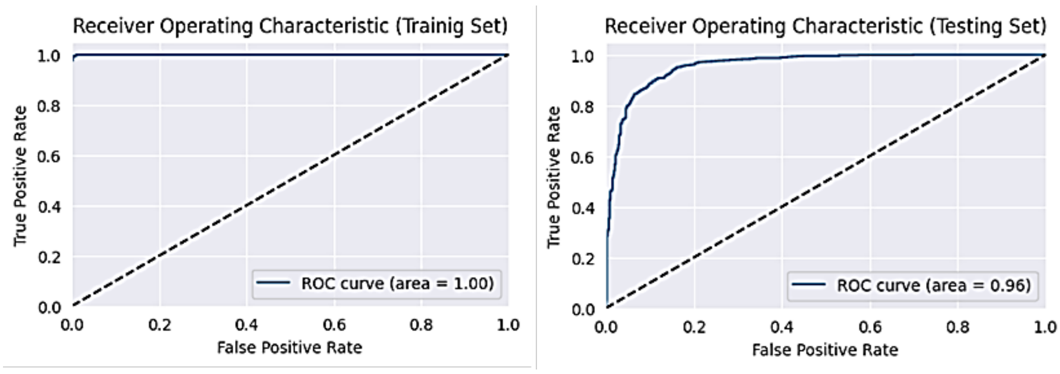


Figure 4-21: Plot ROC Curve - MLP Classifier

The MLP model demonstrates strong predictive capabilities for incident severity classification, with high accuracy and robust performance metrics.

4.9 DISCUSSION

Our study delves into the realm of incident severity classification in workplace safety, where we evaluated the prediction capabilities of five machine learning models: Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Multi-Layer Perceptron (MLP). In this discussion, we'll delve into the insights we gleaned from this evaluation and their practical implications for improving incident severity prediction.

Model Performance and Comparisons

Our evaluation brought to light a diverse range of model performances across various metrics. The Random Forest model stood out, displaying strong alignment between precision, recall, and F1-score. Its excellent Area Under the Curve (AUC) value and minimal errors underscore its robustness in classifying incidents effectively.

The SVM, KNN, and Logistic Regression models also exhibited respectable performances, each with unique strengths. SVM demonstrated good accuracy and AUC, signaling its capability in discerning incident severity. KNN, though trailing slightly in AUC, showcased competitive recall, hinting at its proficiency in detecting true positives. Logistic Regression, while overall sound, could benefit from enhanced precision and recall, making it apt for more balanced prediction needs. The MLP, while promising, struck a trade-off between precision and recall when compared to the Random Forest. This suggests its applicability in scenarios valuing a balanced prediction approach.

Making Sense of the Results

Understanding the models' divergent performances requires a closer look at their underlying mechanisms. The Random Forest's ensemble design empowers it to grasp intricate data relationships, which explains its superior predictive ability. SVM's adeptness in creating effective decision boundaries contributes to its well-rounded performance. KNN's reliance on nearby data instances influences its recall, and the linear nature of Logistic Regression shapes its strengths and limitations

Practical Significance and Future Avenues

Our findings hold significant implications for real-world application. The Random Forest's predictive accuracy can inform strategic decisions on risk management and resource allocation. Meanwhile, the SVM, KNN, and Logistic Regression models provide alternative choices suited to specific application needs.

As we move forward, opportunities for deeper exploration emerge. Exploring ensemble methods, model interpretability techniques, and enhancing the dataset with more diverse features could further elevate predictive accuracy.

In closing, our evaluation marks a milestone in advancing incident severity prediction. By grasping the unique strengths and trade-offs of each model, both practitioners and researchers can tailor their choices to specific application contexts, ultimately fostering safer and more secure work environments.

4.10 LIMITATION AND CHALLENGES

Our study has illuminated certain limitations and challenges inherent to our research, which is essential for a comprehensive understanding of our results and their implications. These considerations offer valuable insights into the complexity of incident severity classification and highlight areas for further refinement and investigation.

The availability and quality of data surfaced as a significant consideration. While we employed rigorous data collection methods, the accuracy and reliability of our incident severity classification inherently depend on the comprehensiveness and accuracy of the training data.

Class imbalance, a common challenge in predictive modeling, also appeared in our study. Despite implementing sophisticated sampling techniques and evaluation metrics, this issue underscores the intricacies of modeling real-world scenarios with varied class distributions.

Selecting and engineering meaningful features posed another challenge. While we meticulously curated our feature set, the richness of contextual information and domain-specific variables might still contain untapped potential that future studies could explore.

Interpreting results, particularly with the Random Forest model, posed challenges due to its ensemble nature. Balancing predictive power with interpretability remains an ongoing consideration, especially in scenarios where model transparency is vital.

Lastly, we acknowledge the scope limitation of our study. Our findings are rooted in a specific dataset and context, which underlines the need for diversified datasets and broader context exploration in future research.

4.11 FURTHER DIRECTIONS AND FUTURE IMPLICATIONS

Beyond the Evaluation of Algorithm Performance: Exploring New Frontiers

Our research serves as a stepping stone to multiple uncharted territories in the realm of incident severity classification. While our current study evaluated algorithm performance, it unveils promising avenues for future exploration that can significantly amplify the impact of our findings.

Enhancing Model Accuracy and Interpretability through Feature Refinement

One direction for advancement lies in refining our feature selection techniques. This refinement can concurrently bolster model accuracy and interpretability, elevating our ability to both predict and comprehend the factors influencing incident severity.

Mitigating Class Imbalance: From Theory to Application

The challenge of class imbalance in predictive modeling presents an exciting avenue for further investigation. Strategies such as oversampling, under-sampling, and ensemble methods could enhance the model's proficiency in accurately predicting incident severity, making strides towards a more equitable classification approach.

Unleashing the Power of Ensemble Models

Delving into ensemble models, which amalgamate the strengths of multiple algorithms, promises a higher echelon of predictive performance. This avenue holds immense potential to push the boundaries of incident severity classification and deepen our understanding of intricate relationships within the data.

Emphasizing Interpretability for Informed Decision-Making

Ensuring interpretability in model outcomes is a vital trajectory. It aligns with the necessity for transparent decision-making in safety management, allowing stakeholders to comprehend the rationale behind the model's predictions and make well-informed choices.

Real-Time Implementation: Bridging Theory and Practice

Translating our developed Machine Learning models into real-time systems stands as a pivotal milestone for practical implementation. By seamlessly integrating these models, we bridge the gap between theoretical advancements and their tangible impact on enhancing workplace safety.

Broadening Horizons: Evaluating Cross-Domain Effectiveness

Validating the efficacy and robustness of our models across diverse datasets is essential. This cross-domain evaluation ensures that our findings extend beyond our current scope, facilitating the broad applicability of Machine Learning in incident severity classification.

In pursuing these multifaceted directions, we pave the way for the advancement and efficacy of Machine Learning applications in incident severity classification. This journey holds promise for cultivating informed decision-making, fortifying risk mitigation strategies, and ultimately fostering safer and more secure work environments.

4.12 CONCLUSION

In this chapter, we explored the application of Machine Learning models for predicting the severity of incidents in safety management. After evaluating different algorithms, including SVM, KNN, Random Forest, Logistic Regression, and MLP, we found that the Random Forest model gives us the best performance in classifying incidents as fatal or non-fatal based on the relevant features we used. It demonstrated high accuracy, precision, recall, and F1-score, making it a valuable tool for incident severity classification. It has the potential to enhance decision-making, resource allocation, and risk mitigation strategies across various industries.

However, we also identified several limitations and challenges that need to be addressed. These include data availability and quality, class imbalance, feature selection, model interpretability, and the generalizability of the findings. By overcoming these challenges, we can improve the reliability and applicability of incident severity classification models.

Moving forward, we plan to explore further avenues to enhance the performance and practicality of the models. This includes investigating interpretability techniques, refining feature engineering processes, and optimizing the models for better accuracy and efficiency.

Overall, our study contributes valuable insights into the use of Machine Learning in safety management and incident severity prediction. By leveraging advanced algorithms and data-driven approaches, organizations can make informed decisions, mitigate risks, and foster safer working environments. Continued research and advancements in Machine Learning will drive further improvements in safety management practices and contribute to the prevention of incidents across industries.

GENERAL CONCLUSION

In this thesis, we have conducted an in-depth investigation into the application of advanced machine learning techniques for predicting the severity of incident classes in the domain of safety management. Our goal was to explore how artificial intelligence and machine learning algorithms can be leveraged to enhance safety management processes and provide more accurate predictions.

Through a comprehensive literature review, we have examined the current state of research and practice in using AI and machine learning for safety management. We have identified various applications of machine learning in risk assessment and explored different implementation approaches across industries. By analyzing the existing body of knowledge, we have gained insights into the potential benefits and challenges associated with integrating machine learning into safety management workflows.

In the chapter dedicated to safety management, we have laid the foundation by discussing the fundamental concepts and components of effective safety management systems. We have highlighted the importance of hazard identification, incident reporting and investigation, and fostering a positive safety culture. Furthermore, we have emphasized the significance of risk assessment and its various phases, including risk identification, analysis, and evaluation.

Delving into the realm of machine learning, we have provided a comprehensive overview of different types of machine learning algorithms. We have explained the principles behind supervised learning, unsupervised learning, and reinforcement learning, showcasing their unique capabilities and applications. Additionally, we have discussed the data-driven approach and presented different frameworks for building machine learning systems, allowing for a structured and systematic implementation of machine learning models in safety management.

In our practical work, we have employed a robust methodology to predict the severity of incident classes using machine learning algorithms. We have carefully selected the OSHA construction dataset from the Kaggle platform as our primary data source. Through exploratory data analysis, we have gained valuable insights into the dataset, enabling us to understand the patterns and distributions of the variables.

To ensure data quality and model performance, we have applied pre-processing techniques to clean and transform the data. Feature selection methods have been employed to identify the most relevant attributes for predicting incident severity. By considering both textual and numerical data analysis techniques, we have maximized the information extracted from the dataset.

Through rigorous experimentation, we have trained and evaluated several state-of-the-art machine learning algorithms for incident severity classification. Support vector machines, k-nearest neighbors, random forest classifier, logistic regression, and multilayer perceptron models have been implemented and compared. We have utilized appropriate evaluation metrics to assess the performance of each algorithm, considering accuracy, precision, recall, and F1-score.

Our findings demonstrate the effectiveness and potential of machine learning in predicting the severity of incident classes. The random forest classifier emerged as the top performer, showcasing its ability to capture complex relationships and make accurate predictions. However, we also acknowledge the limitations and challenges associated with machine learning, such as the availability and quality of training data, class imbalance, feature selection, and model interpretability.

To overcome these challenges, future research should focus on addressing the scarcity of training data through data augmentation techniques and leveraging domain knowledge for better feature selection. Techniques for handling class imbalance should be explored to ensure unbiased and reliable model predictions. Additionally, interpretability methods can be employed to provide insights into the decision-making process of the machine learning models, especially in safety management domains where transparency and explainability are crucial.

In conclusion, our advanced research contributes to the growing body of knowledge on applying machine learning in safety management. By harnessing the power of AI and machine learning algorithms, we have demonstrated the potential for improved incident severity prediction. Our work provides valuable insights into the strengths and limitations of different machine learning algorithms and highlights the need for further advancements and considerations in data quality, model interpretability, and addressing class imbalance. Ultimately, by embracing and refining these machine learning techniques, safety management professionals can make more informed decisions, allocate resources efficiently, and mitigate risks effectively, leading to safer work environments and improved outcomes.

BIBLIOGRAPHY

[1] CQ, ‘Leverage AI and ML to Improve Safety Management’, *ComplianceQuest QHSE Solutions*, Aug. 12, 2022. https://www.compliancequest.com/blog/leverage-ai-ml-to-improve-safety-management/ (accessed May 15, 2023).

[2] C. Gundlapalli, ‘Council Post: How To Leverage AI/ML For Predictive Incident Management’, *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2022/09/19/how-to-leverage-aiml-for-predictive-incident-management/ (accessed May 15, 2023).

[3] N. Paltrinieri, L. Comfort, and G. Reniers, ‘Learning about risk: Machine learning for risk assessment’, *Saf. Sci.*, vol. 118, pp. 475–486, Oct. 2019, doi: 10.1016/j.ssci.2019.06.001.

[4] Z. Jiao, P. Hu, H. Xu, and Q. Wang, ‘Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications’, *ACS Chem. Health Saf.*, vol. 27, no. 6, pp. 316–334, Nov. 2020, doi: 10.1021/acs.chas.0c00075.

[5] M. Leo, S. Sharma, and K. Maddulety, ‘Machine Learning in Banking Risk Management: A Literature Review’, *Risks*, vol. 7, no. 1, p. 29, Mar. 2019, doi: 10.3390/risks7010029.

[6] H. Cao and Y. M. Goh, ‘Analyzing construction safety through time series methods’, *Front. Eng. Manag.*, vol. 6, no. 2, pp. 262–274, Jun. 2019, doi: 10.1007/s42524-019-0015-6.

[7] J. Hegde and B. Rokseth, ‘Applications of machine learning methods for engineering risk assessment – A review’, *Saf. Sci.*, vol. 122, p. 104492, Feb. 2020, doi: 10.1016/j.ssci.2019.09.015.

[8] L. A. Curiel-Ramirez, R. A. Ramirez-Mendoza, G. Carrera, J. Izquierdo-Reyes, and M. R. Bustamante-Bello, ‘Towards of a modular framework for semi-autonomous driving assistance systems’, *Int. J. Interact. Des. Manuf. IJIDeM*, vol. 13, no. 1, pp. 111–120, Mar. 2019, doi: 10.1007/s12008-018-0465-9.

[9] F. Zhou, Y. Song, L. Liu, and D. Zheng, ‘Automated visual inspection of target parts for train safety based on deep learning’, *IET Intell. Transp. Syst.*, vol. 12, no. 6, pp. 550–555, Aug. 2018, doi: 10.1049/iet-its.2016.0338.

[10] ‘Safety Management | SKYbrary Aviation Safety’. https://www.skybrary.aero/articles/safety-management (accessed May 07, 2023).

[11] B. SAFETY, ‘The Importance of Safety Management’, *BRITE SAFETY*. https://britesafety.com/blogs/news/the-importance-of-safety-management (accessed May 20, 2023).

[12] ‘Workplace Safety: Importance, Benefits and Ways to Create a Safe Workplace’, *Nurture an Engaged and Satisfied Workforce | Vantage Circle HR Blog*, Apr. 03, 2019. https://blog.vantagecircle.com/workplace-safety/ (accessed May 20, 2023).

[13] ‘Safety Management - Hazard Identification and Assessment | Occupational Safety and Health Administration’. https://www.osha.gov/safety-management/hazard-Identification#ai1 (accessed May 20, 2023).

[14] C. C. for O. H. and S. Government of Canada, 'CCOHS: Hazard and Risk - Hazard Identification', Apr. 05, 2023. https://www.ccohs.ca/oshanswers/hsprograms/hazard/hazard_identification.html (accessed May 20, 2023).

[15] R. Keen, 'Incident Reporting & Investigation Procedure Explained [ISO 45001, with Procedure]', *ISO 9001 Checklist*, Aug. 30, 2021. https://www.iso-9001-checklist.co.uk/ISO-45001/incident-reporting-investigation-procedure.htm (accessed May 20, 2023).

[16] 'Incident Investigation - Overview | Occupational Safety and Health Administration'. https://www.osha.gov/incident-investigation (accessed May 20, 2023).

[17] 'Incident Investigations'. https://www.safemanitoba.com/topics/Pages/Incident-Investigations.aspx (accessed May 20, 2023).

[18] S. Foster, 'Understanding safety culture', *Br. J. Nurs.*, vol. 30, no. 13, pp. 831–831, Jul. 2021, doi: 10.12968/bjon.2021.30.13.831.

[19] D. Moriarty, *Practical human factors for pilots*. London: Academic Press, 2015.

[20] 'The Ultimate Guide to Safety Management Systems', *Safesite*, Apr. 26, 2020. https://safesitehq.com/safety-management-systems/ (accessed May 15, 2023).

[21] 'Safety Management System - The Beginner's Guide | Connecteam'. https://connecteam.com/safety-management-system/ (accessed May 15, 2023).

[22] 'Safety Management - A safe workplace is sound business | Occupational Safety and Health Administration'. https://www.osha.gov/safety-management (accessed May 15, 2023).

[23] National Research Council (U.S.), National Research Council (U.S.), and National Academies Press (U.S.), Eds., *The owner's role in project risk management*. Washington, DC: National Academies Press, 2005.

[24] 'Safety management system', *Wikipedia*. Feb. 07, 2023. Accessed: May 15, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Safety_management_system&oldid=1137921911

[25] A. Teller, 'What Are the Key Elements and Components of a Safety Management System?', *StrongArm Technologies*, Jan. 10, 2023. https://www.strongarmtech.com/blog-posts/elements-of-a-safety-management-system/ (accessed May 15, 2023).

[26] '4 Essential Components Of A Safety Management System | Rapid Global'. https://www.rapidglobal.com/knowledge-centre/what-is-a-safety-management-system/ (accessed May 15, 2023).

[27] 'What is Safety? Safety Management Systems for Workplace & Food Quality | ASQ'. https://asq.org/quality-resources/safety (accessed May 15, 2023).

[28] 'ISO 31000 2018 Risk management - Guidelines.pdf'.

[29] 'ISO 31010 2009 Risk management Risk assessment techniques.pdf'.

[30] A. J.-P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, 'Application of machine learning to construction injury prediction', *Autom. Constr.*, vol. 69, pp. 102–114, Sep. 2016, doi: 10.1016/j.autcon.2016.05.016.

[31] 'How Machine Learning Helps Improve Fleet Safety', *Robotics & Automation News*, Mar. 16, 2023. https://roboticsandautomationnews.com/2023/03/16/how-machine-learning-helps-improve-fleet-safety/65938/ (accessed May 18, 2023).

[32] S. Oladele, 'A Comprehensive Guide on How to Monitor Your Models in Production', *neptune.ai*, Aug. 11, 2022. https://neptune.ai/blog/how-to-monitor-your-models-in-production-guide (accessed May 18, 2023).

[33] 'Predict OH&S Incidents using AI and ML -Predictive Analytics', *AI Consulting Group*, Apr. 04, 2022. https://www.aiconsultinggroup.com.au/predict-ohs-incidents-with-artificial-intelligence-and-machine-learning/ (accessed May 18, 2023).

[34] F. Recal and T. Demirel, 'Comparison of machine learning methods in predicting binary and multi-class occupational accident severity', *J. Intell. Fuzzy Syst.*, vol. 40, no. 6, pp. 10981–10998, Jun. 2021, doi: 10.3233/JIFS-202099.

[35] 'Machine Learning Safety Training for Autonomous Vehicles', *UL Solutions*. https://www.ul.com/services/machine-learning-safety-training-autonomous-vehicles (accessed May 18, 2023).

[36] T. V. Chandni and A. Babu, 'IMPLEMENTATION OF SAFETY MANAGEMENT USING MACHINE LEARNING APPROACH', vol. 10, no. 10, 2022.

[37] B. Wang, C. Wu, L. Huang, and L. Kang, 'Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers', *J. Clean. Prod.*, vol. 210, Nov. 2018, doi: 10.1016/j.jclepro.2018.11.181.

[38] C. Wodecki, 'Predictive Maintenance & Machine Learning: A Complete Guide', *LLumin*, Oct. 13, 2022. https://llumin.com/predictive-maintenance-machine-learning-a-complete-guide-llu/ (accessed May 18, 2023).

[39] 'Machine Learning Techniques for Predictive Maintenance', *InfoQ*. https://www.infoq.com/articles/machine-learning-techniques-predictive-maintenance/ (accessed May 18, 2023).

[40] A. Kane, A. Kore, A. Khandale, S. Nigade, and P. P. Joshi, 'Predictive Maintenance using Machine Learning'.

[41] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*, First edition. Beijing ; Boston: O'Reilly Media, 2017.

[42] 'Machine Learning | Types | Benefits'. https://adservio.fr/post/machine-learning-types-benefits (accessed Mar. 15, 2023).

[43] B. Mahesh, 'Machine Learning Algorithms - A Review', vol. 9, no. 1, 2018.

[44] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps*. Berkeley, CA: Apress, 2017. doi: 10.1007/978-1-4842-2866-1.

[45] 'Supervised vs. Unsupervised Learning [Differences & Examples]'. https://www.v7labs.com/blog/supervised-vs-unsupervised-learning (accessed Mar. 15, 2023).

[46] E. Alpaydin, *Introduction to machine learning*, 2nd ed. in Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2010.

[47] R. Raj, 'Supervised, Unsupervised and Semi-supervised Learning'. https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning (accessed Mar. 18, 2023).

[48] 'What is Unsupervised Learning? | Supervised vs. Unsupervised', *Edureka*, Nov. 20, 2019. https://www.edureka.co/blog/unsupervised-learning/ (accessed Mar. 21, 2023).

[49] 'Life cycle of Machine Learning - Javatpoint'. https://www.javatpoint.com/machine-learning-life-cycle (accessed May 07, 2023).

[50] 'Applied Machine Learning Life Cycle for Computer Vision Tasks | by pixolution | Becoming Human: Artificial Intelligence Magazine'. https://becominghuman.ai/applied-machine-learning-life-cycle-for-computer-vision-tasks-65f54bf77c37?gi=6aba0cea3ac8 (accessed May 07, 2023).

[51] S. Ray, 'Commonly used Machine Learning Algorithms (with Python and R Codes)', *Analytics Vidhya*, Sep. 08, 2017. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/ (accessed Apr. 08, 2023).

[52] 'Machine Learning Algorithms - Javatpoint', *www.javatpoint.com*. https://www.javatpoint.com/machine-learning-algorithms (accessed Apr. 08, 2023).

[53] 'Decision Tree Algorithm in Machine Learning - Javatpoint', *www.javatpoint.com*. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm (accessed Apr. 08, 2023).

[54] 'Machine Learning Random Forest Algorithm - Javatpoint', *www.javatpoint.com*. https://www.javatpoint.com/machine-learning-random-forest-algorithm (accessed Apr. 08, 2023).

[55] 'Support Vector Machine (SVM) Algorithm - Javatpoint', *www.javatpoint.com*. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm (accessed May 05, 2023).

[56] 'What Is a Support Vector Machine? Working, Types, and Examples', *Spiceworks*. https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/ (accessed May 05, 2023).

[57] 'K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint', *www.javatpoint.com*. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning (accessed May 05, 2023).

[58] R. Khandelwal, 'K-Nearest Neighbors(KNN)', *Medium*, Nov. 16, 2018. https://medium.datadriveninvestor.com/k-nearest-neighbors-knn-7b4bd0128da7 (accessed May 06, 2023).

[59] 'Python Presentation'. https://www.slideshare.net/narendra.sisodiya/python-presentation-presentation (accessed Jun. 08, 2023).

[60] 'What is Python? Executive Summary', *Python.org*. https://www.python.org/doc/essays/blurb/ (accessed Jun. 08, 2023).

[61] 'Kaggle presentation'. https://www.slideshare.net/HJvanVeen/kaggle-presentation (accessed Jun. 08, 2023).

[62] C. D. Costa, ‘Best Python Libraries for Machine Learning and Deep Learning’, *Medium*, Mar. 24, 2020. https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c (accessed Jun. 08, 2023).

[63] ‘Google-Colab-Introduction.pdf’. Accessed: Jun. 08, 2023. [Online]. Available: https://mcgrawect.princeton.edu/guides/Google-Colab-Introduction.pdf

[64] ‘OSHA Accident and Injury Data’. https://www.kaggle.com/datasets/ruqaiyaship/osha-accident-and-injury-data-1517 (accessed Jun. 09, 2023).

[65] ‘Performance Metrics in Machine Learning - Javatpoint’. https://www.javatpoint.com/performance-metrics-in-machine-learning (accessed Jun. 17, 2023).

[66] ‘Classification Report in Machine Learning | Aman Kharwal’. https://thecleverprogrammer.com/2021/07/07/classification-report-in-machine-learning/ (accessed Jun. 17, 2023).