AI and Machine Learning-Powered Predictive Maintenance for Structures and Mechanical Assets Using BIM

Mutaz W. Qafisheh[1]*, [2]Muwafak A. AbuZeineh
[1] The Department of Civil Engineering, College of Engineering and Technology, Palestine Polytechnic University, Hebron, Palestine.
[2]Ishbilia Consultancy, Hebron, Palestine.
**Corresponding author: mutazq@ppu.edu*

***Abstract*—While executing various management and maintenance techniques, relevant operational efficacy, output optimization, and the seamless operation of professional and residential areas are increasingly important. However, conventional maintenance approaches face substantial challenges due to rising costs associated with time-based preventative maintenance programs and the consequences of reactive repairs. This investigation investigates the use of Building Information Modelling (BIM) maintenance repositories as a foundation for creating machine learning tools for predictive maintenance. Our findings show that machine learning models can predict project vulnerabilities and equipment faults, allowing maintenance administrators to ensure that facilities continue to operate safely. The findings of this study support the use of such analytical methodologies in related disciplines. With a focus on cost-effectiveness, this research recommends examining various database options, from specialised repositories customised for particular industries and asset categories to comprehensive Computerised Maintenance Management Systems (CMMS). The research investigation also emphasises the increasing importance of mobile accessibility, the incorporation of AI and advanced analytical features, the growth of open-source distribution models, and the critical importance of carefully selecting solutions that meet the specific needs of various organisations and industry-specific imperatives. This study seeks to provide operators with a complete and analytical overview of the available options to assist them in making well-informed judgements about which maintenance database solution best meets their specific operational demands. The research results showed that the average accuracy for the investigated datasets and machine learning models reaches around 90 percent. The study concludes that machine learning has strong predictive maintenance capabilities.**

I. INTRODUCTION

BIM can potentially revolutionize the conventional architecture, engineering, and construction sectors. Accordingly, BIM is mainly utilized to establish a comprehensive digital representation from the abstracted three-dimensional models [1]. Engineering projects and systems are fully represented digitally, giving stakeholders a holistic perspective and enormous information about their operational and physical properties. BIM offers an integrated platform that integrates construction design and visualization in accordance with planning, cost estimation, execution, and maintenance. In other words, it encompasses the entire task life cycle. Therefore, as the potential of BIM is increasingly recognized, its value proposition is greatly expanded into the entire project life-cycle [2,3]. Combining Building Information Modeling with machine learning models was investigated to see how it could improve the prediction of machine and construction asset failures. The goal was to proactively prevent production breakdowns and service interruptions by identifying potential issues before they occur. The research focused on the benefits and enhancements of this integrated approach.

Plevris and Papazafeiropoulos, (2024) investigated the impact of AI on structural health monitoring in the context of safety and maintenance. The research highlighted AI's potential to improve the efficiency, safety, and sustainability of infrastructure systems. Lawal et al., (2025) demonstrate the potential of the integration between the Convolutional Neural Networks (CNNs) and the Internet of Things (IoT), showing that this integration could improve predictive maintenance via advanced notifications. It also reduces the need for centralised information archiving. Ahmad et.al, (2025) examined how AI and machine learning integration with innovative infrastructure and electrical systems improves conventional engineering by improving efficiency, accuracy, and sustainability. The study emphasises that AI will continue to advance through innovative future structures and worldwide power systems. Similarly, Ahmaed et.al, (2025) outline the primary challenges associated with using reinforcement learning and IoT in predictive maintenance systems, including data quality, computational constraints, privacy concerns, and interface with legacy systems [1,2,3,4]. Ohalete et.al, (2023) review the advances of predictive maintenance in the oil and gas industry, focusing on the integration and influence of AI and Data Science. The study indicates that AI's ongoing evolution will significantly impact the future of maintenance methods in the oil and gas industry. In addition, Guidotti et.al, (2025) emphasise the paradigm shift in manufacturing, which is driving industries to adopt innovative technology for more effective decision-making. Predictive maintenance is a critical component of this revolution, playing a significant role in this transition by employing supervised machine learning techniques to forecast equipment failures, optimise maintenance schedules, and improve operational efficiency. The paper examines major trends in ML adoption and provides insights into future research directions, emphasising the need for open datasets, explainable AI, and cross-domain generalisation. Predictive maintenance has recently become an essential approach for guaranteeing the best possible performance and durability of construction assets due to the emergence of Artificial Intelligence (AI), the Internet of Things IoT that is aligned with machine learning, and deep learning integrated with the wealth of information that BIM can

Corresponding author: Mutaz Qafisheh is with the Palestine Polytechnic University, Hebron, Palestine (phone: 00972599820855; e-mail: Mutazq@ppu.edu).

Muwafak AbuZeineh., is with ISHBILIA Consultancy, Hebron, Palestine. He is now a part time lecturer with Palestine Polytechnic University, (e-mail: muwafakadel@gmail.com).

provide. Therefore, machine learning and deep learning might assist managers and stakeholders in anticipating possible equipment or asset breakdowns under this approach [4]. Predictive maintenance is consequently essential for reducing equipment failure and extending the lifespan of construction assets. Accordingly, it mitigated the limitations inherent in conventional maintenance paradigms (corrective and preventive, conditional, predetermined maintenance) – characterized by temporal inefficiencies, elevated operational risks stemming from equipment or asset malfunctions, and substantial economic detriments arising from protracted repair durations – [5,6,7]. Consequently, the predictive maintenance protocols are attracting the attention of many scholars these days [8, 9,10].

Due to the rapid advancements in relevant scientific domains, such as the Internet of Things, artificial intelligence (AI) [11,12,13,14], BIM specifications and standards [15], and machine and deep learning algorithms [16,17], the predictive maintenance strategy is still evolving. Consequently, this study examines how effectively various machine algorithms anticipate the likelihood of construction and equipment failures. The research's main findings could be valuable to stakeholders and operational assessments of mechanical systems and structures. By utilizing open-source libraries, the project applies artificial intelligence (AI) and machine learning (ML) models to facilitate predictive maintenance. The outcomes of this project could prove valuable to stakeholders and for operational assessments of mechanical systems and structures. By utilizing open-source libraries, the project applies artificial intelligence (AI) and machine learning (ML) models to facilitate predictive maintenance.

II. METHODOLOGY AND DATA

A. Maintenance and BIM Data

A centralized digital repository related to an organization's maintenance activities is known as a maintenance database. Its main goal is to convert natural maintenance data into valuable insights that increase productivity, reduce operational interruptions, and prolong the life of costly assets. From the days of handwritten and paper-based systems to the unconventional digital database solutions that are currently widely accessible, maintenance record-keeping has undergone significant evolution. The way that firms handle maintenance management has changed dramatically due to this technical novelty. The advantages of using a maintenance database are numerous and include a notable decrease in expensive equipment and facility downtime as a result of proactive maintenance strategies, a significant increase in operational efficiency through streamlined workflows and automated processes, improved asset life-cycle management through thorough tracking and historical data analysis; a stronger devotion to industry-specific and regulatory compliance requirements; and the ability to make data-driven choices. In order to give complete guidance to businesses looking to improve their maintenance operations [18,19].

The research highlighted the importance of training ML models with high-quality data on specific maintenance needs. Accordingly, it is crucial to establish several ML predictive models to fulfill the maintenance requirements. Consequently, tailored ML predictive maintenance models are required; customized models are required for Heating, Ventilation, and Air Conditioning (HVAC), Mechanical, Electrical, structural elements, and other construction assets. Subsequently, suitable data is needed to train the tailored ML predicting maintenance models, which have many records sufficient to train, validate, and evaluate the model's performance. Moreover, the research highlighted leveraging features (columns), which provide machine learning with several data types; for instance, in mechanical assets, features related to temperature, oil level, resistance observations, pressure, torque, etc., are essential for well-model establishments. Afridi et.al, (2021) investigated the artificial intelligence-based prognostic maintenance of renewable energy systems. Discussing methods, obstacles, and potential research directions, the analysis revealed that ingenious algorithms and procedures have been adopted to reduce equipment and plant downtimes. Efforts are being undertaken to create robust prognostic maintenance systems that detect defects before they arise. To achieve this goal, advanced Data Analytics and Artificial Intelligence (AI) algorithms are applied to boost the overall efficiency of these prognostic maintenance systems. Shaban et.al, (2025) examine how maintenance is integrated into HVAC systems and its importance in the industry. Maintenance makes use of sophisticated technology like artificial intelligence and IoT sensing. Similarly, Sikdar et.al, (2024) present a detailed study of innovative HVAC systems' accomplishments, problems, and future prospects by reviewing 125 peer-reviewed articles using the PRISMA technique. The Sikdar findings enable researchers, industry stakeholders, and policymakers to create scalable, cost-effective, and energy-efficient HVAC solutions. Similarly, Suryadarma and Ai, (2020) examine previous research on predictive maintenance in Supervisory Control and Data Acquisition (SCADA) based industries. In the field of predictive maintenance research, many approaches for predicting machine damage or time to failure have been presented and used in a variety of sectors [5,6,7,8].

The current research investigated two BIM databases to train, validate, and evaluate the proposed machine learning models. The first database related to several infrastructure projects, including tunnels, dams, and buildings in the USA. The first database has several features such as type of project, planned and actual duration, schedule deviation, vibration level, crack width, environmental indicators, energy consumption, material usage, lab hours, safety risk score, the percentage of completion, and final the project risk level with three possible values (high, low, and medium). The database is available to download through the following link https://www.kaggle.com/datasets/ziya07/bim-ai-integrated-dataset?resource=download. The first database, which contains 1,000 entities and 28 features, is focused on large-scale infrastructure projects such as bridges, dams, tunnels, and other megastructures. The data underwent a crucial preparation phase, including cleaning and feature scaling with the Sklearn standard scaler. Categorical features were encoded to facilitate model training. The database was then partitioned, with 35% allocated for training and the remaining portion used for testing and validation.

The second investigated database is related to mechanical equipment; several researchers have investigated this database. In addition, the database includes several features such as air temperature, process temperature, rotational speed, torque, wear, and machine failures, which are associated as aggregated features from the five failure indicators donates as (TWF, HDF, PWF, OSF, and RNF) where each term of the prementioned abbreviations denotes the Temperature Tool Wear Failure, Heat Dissipation Failure, Power Failure, Overstrain Failure, and Random Failure, respectively [20]. The second database, which is significantly larger with 10,000 entities and 14 features, relates to various pieces of mechanical equipment. Like the first database, it was subjected to cleaning, scaling with the Sklearn standard scaler, and encoding categorical values. The same data split ratio was applied, with 35% designated for model training and the remainder for testing and validation.

B. *Methodology*

A series of work packages was developed to effectively execute the research endeavor, each designed to address a specific study stage.

A. *Work Package 1 (Data Collection and Integration)*

- Gather thorough BIM data, including specifics about the systems, materials, and construction components.
- Construction management systems incorporate real-time sensor data, including vibration, temperature, humidity, and energy usage.
- Develop a robust data pipeline to efficiently collect, process, and store the combined dataset.

B. *Work Package 2 (AI/ML Model Development)*

- Utilize BIM and AI capabilities to develop and train predictive maintenance models.
- To find the best strategy, investigate different machine learning algorithms, including anomaly detection, deep learning, and time-series analysis.
- Use the asset management and knowledge to direct model development to guarantee precise forecasts.

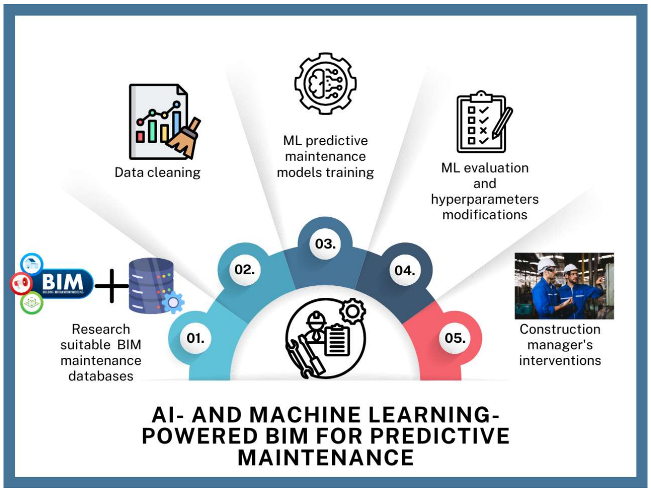
C. *Work Package 3 (Model Deployment and Validation)*

- Install the real-time learned models so they may be analyzed and predicted in real time.
- Monitor model performance and adjust algorithms in response to feedback and real-world data.
- To guarantee the precision and dependability of the predictive maintenance system, thorough testing and validation must be carried out.

Figure 1 shows the overview of the utilized methodology in the current study.

Figure 1. *Methodology overview*

The study applied a consistent three-work-package methodology to analyze two distinct databases. Both datasets underwent identical preprocessing steps to prepare them for machine learning model training and evaluation.



For data visualization, several powerful Python libraries are available. Matplotlib offers fine-grained control over plots, allowing for detailed customization. Developers often turn to Seaborn to create more aesthetically pleasing and statistically informative visualizations with less code. Another versatile option is Plotly, which supports creating various interactive charts, including scatter plots, heatmaps, and geographic maps. Beyond visualization, other key libraries handle data manipulation and machine learning. NumPy is fundamental for numerical operations, efficiently supporting large, multi-dimensional arrays. Building on this, Pandas is essential for data analysis, offering easy-to-use, high-performance data structures for manipulating and analyzing datasets. For machine learning, scikit-learn provides a comprehensive collection of supervised and unsupervised algorithms and tools for model evaluation and data preprocessing. In deep learning, Keras serves as a user-friendly API built on top of TensorFlow, simplifying the construction of neural network models. Finally, the built-in pickle library is used for serializing and deserializing Python objects, making it easy to save and load models or other data structures[9,10,11,12,13,14,15,16,17].

The performance of the machine learning models was comprehensively evaluated using both overall accuracy and the F1 score. However, for clarity and accessibility, the study primarily presents the accuracy metric to illustrate the models' performance.

In order to illustrate the research methodology, Figures 2 and 3 demonstrate the applied methodology specifically to the investigated database. However, Figure 1 represents the overview of the research methodology.

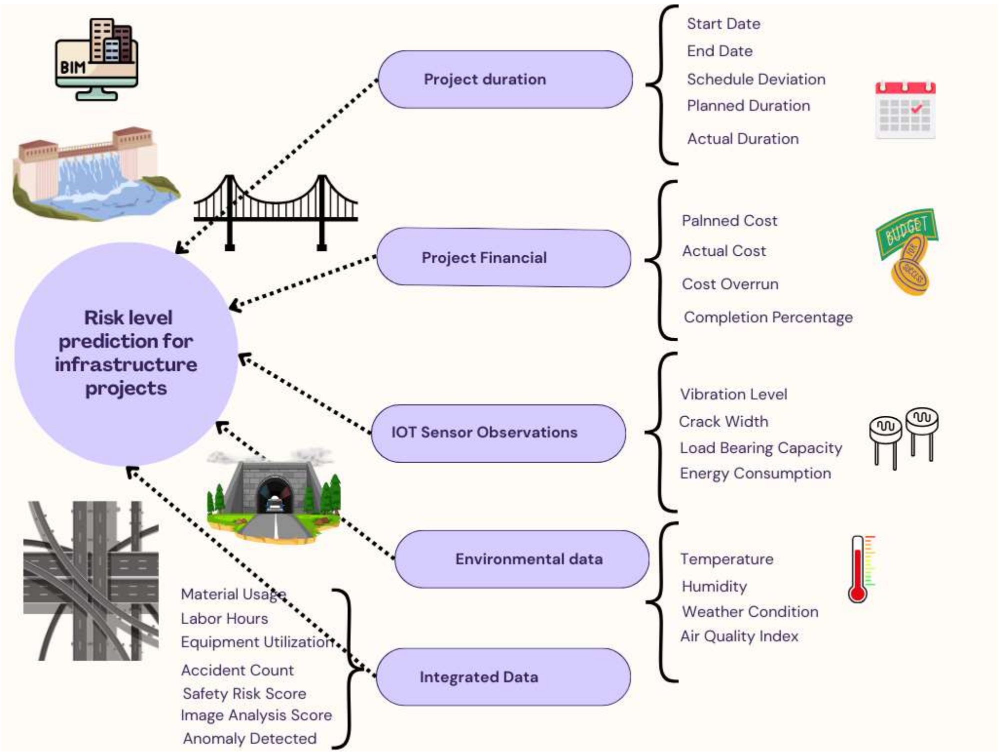


Figure 2. *Civil project risk level prediction methodology*

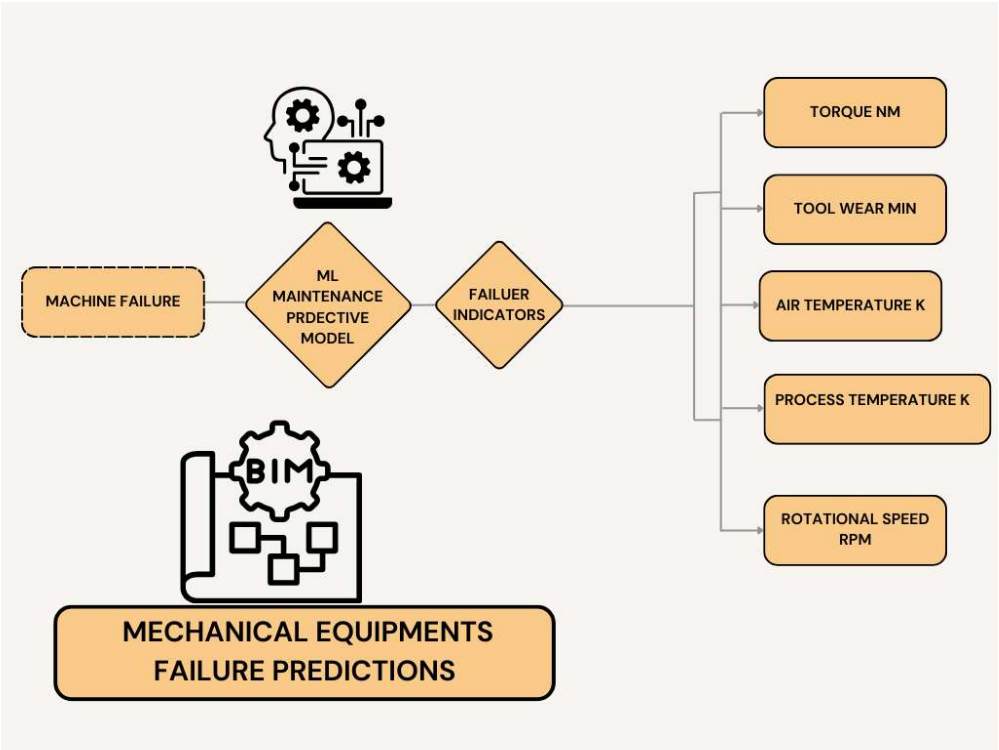
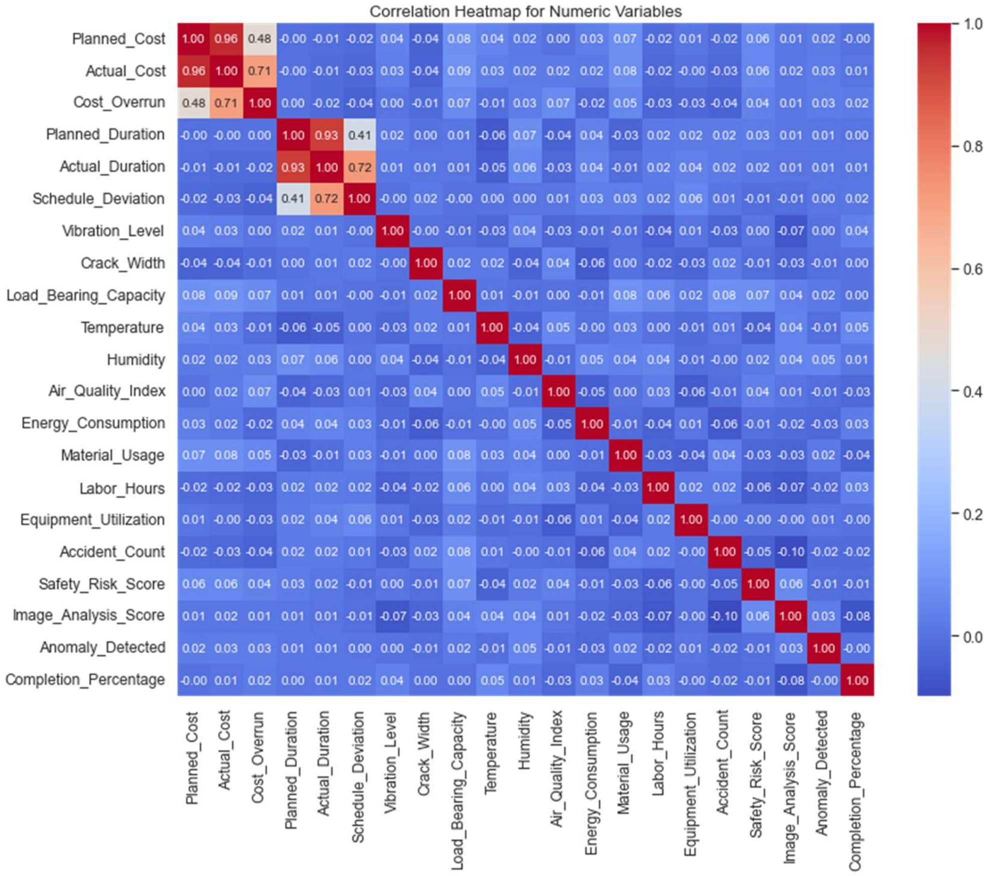


Figure 3. *Mechanical equipment failures prediction methodology*

I. RESULTS AND DISCUSSIONS

This section presents the study's significant findings and discussions regarding the ability of integrated machine learning approaches and BIM data as predictive maintenance tools. The section also illustrates the investigation carried out on the research in order to validate, train, and evaluate the machine learning algorithms. Accordingly, in the current section, the research presents several findings regarding the data exploration, correlation, and data anomalies. The first database investigates the current endeavor to establish a machine learning predictive maintenance model. Accordingly, the count of each project on the research database demonstrates a more or less equal distribution, with around 180 projects for each type of infrastructure project.

In addition to the histogram analysis, we investigated the correlation matrix in this research. Accordingly, this matrix presents the correlation between the research variables or features. In other words, it shows if there is a positive, negative, or neutral relation between variables; accordingly, its value varies from 1 to -1, where 1 represents a strong positive relation. At the same time, -1 denotes the negative relation between the research features. However, zero values represent no relation between variables. Figure 4 shows the heat map figures created to investigate the feature relation. It can be seen from the figure that there is a strong positive relationship among some variables, namely, planned, actual cost, and cost overrun. Similarly, the same finding was found for planned, actual duration, and schedule deviation.

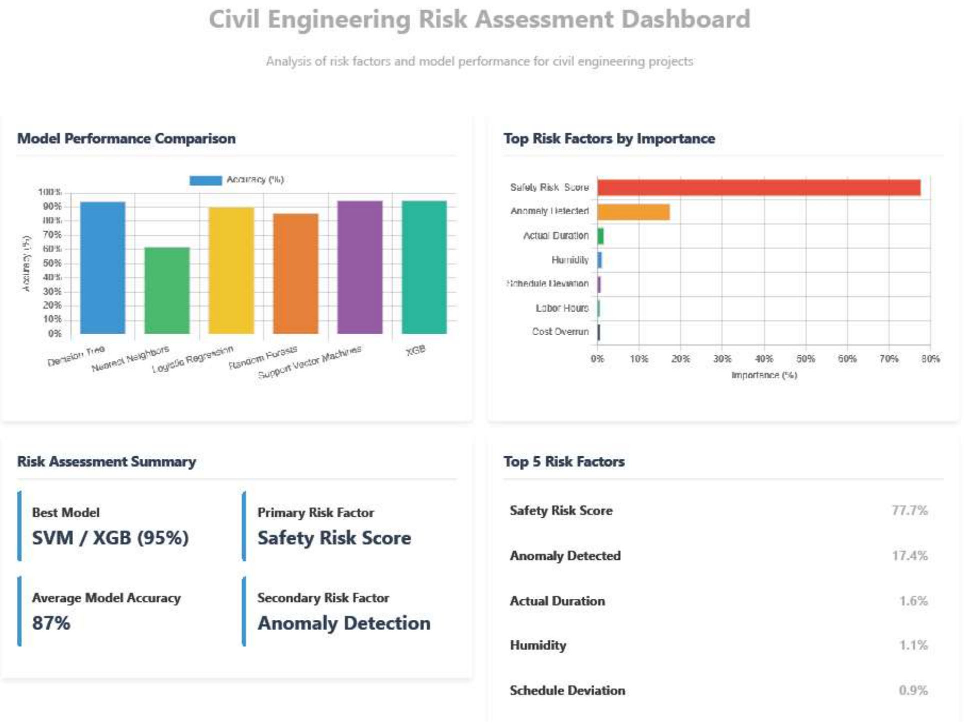


	Planned_Cost	Actual_Cost	Cost_Overrun	Planned_Duration	Actual_Duration	Schedule_Deviation	Vibration_Level	Crack_Width	Load_Bearing_Capacity	Temperature	Humidity	Air_Quality_Index	Energy_Consumption	Material_Usage	Labor_Hours	Equipment_Utilization	Accident_Count	Safety_Risk_Score	Image_Analysis_Score	Anomaly_Detected	Completion_Percentage
Planned_Cost	1.00	0.96	0.48	-0.00	-0.01	-0.02	0.04	-0.04	0.08	0.04	0.02	0.00	0.03	0.07	-0.02	0.01	-0.02	0.06	0.01	0.02	-0.00
Actual_Cost	0.96	1.00	0.71	-0.00	-0.01	-0.03	0.03	-0.04	0.09	0.03	0.02	0.02	0.02	0.08	-0.02	-0.00	-0.03	0.06	0.02	0.03	0.01
Cost_Overrun	0.48	0.71	1.00	0.00	-0.02	-0.04	0.00	-0.01	0.07	-0.01	0.03	0.07	-0.02	0.05	-0.03	-0.03	-0.04	0.04	0.01	0.03	0.02
Planned_Duration	-0.00	-0.00	0.00	1.00	0.93	0.41	0.02	0.00	0.01	-0.06	0.07	-0.04	0.04	-0.03	0.02	0.02	0.02	0.03	0.01	0.01	0.00
Actual_Duration	-0.01	-0.01	-0.02	0.93	1.00	0.72	0.01	0.01	0.01	-0.05	0.06	-0.03	0.04	-0.01	0.02	0.04	0.02	0.02	0.01	0.01	0.01
Schedule_Deviation	-0.02	-0.03	-0.04	0.41	0.72	1.00	-0.00	0.02	-0.00	0.00	0.00	0.01	0.03	0.03	0.02	0.06	0.01	-0.01	-0.01	0.00	0.02
Vibration_Level	0.04	0.03	0.00	0.02	0.01	-0.00	1.00	-0.00	-0.01	-0.03	0.04	-0.03	-0.01	-0.01	-0.04	0.01	-0.03	0.00	-0.07	0.00	0.04
Crack_Width	-0.04	-0.04	-0.01	0.00	0.01	0.02	-0.00	1.00	0.02	0.02	-0.04	0.04	-0.06	0.00	-0.02	-0.03	0.02	-0.01	-0.03	-0.01	0.00
Load_Bearing_Capacity	0.08	0.09	0.07	0.01	0.01	-0.00	-0.01	0.02	1.00	0.01	-0.01	0.00	-0.01	0.08	0.06	0.02	0.08	0.07	0.04	0.02	0.00
Temperature	0.04	0.03	-0.01	-0.06	-0.05	0.00	-0.03	0.02	0.01	1.00	-0.04	0.05	-0.00	0.03	0.00	-0.01	0.01	-0.04	0.04	-0.01	0.05
Humidity	0.02	0.02	0.03	0.07	0.06	0.00	0.04	-0.04	-0.01	-0.04	1.00	-0.01	0.05	0.04	0.04	-0.01	-0.00	0.02	0.04	0.05	0.01
Air_Quality_Index	0.00	0.02	0.07	-0.04	-0.03	0.01	-0.03	0.04	0.00	0.05	-0.01	1.00	-0.05	0.00	0.03	-0.06	-0.01	0.04	0.01	-0.01	-0.03
Energy_Consumption	0.03	0.02	-0.02	0.04	0.04	0.03	-0.01	-0.06	-0.01	-0.00	0.05	-0.05	1.00	-0.01	-0.04	0.01	-0.06	-0.01	-0.02	-0.03	0.03
Material_Usage	0.07	0.08	0.05	-0.03	-0.01	0.03	-0.01	0.00	0.08	0.03	0.04	0.00	-0.01	1.00	-0.03	-0.04	0.04	-0.03	-0.03	0.02	-0.04
Labor_Hours	-0.02	-0.02	-0.03	0.02	0.02	0.02	-0.04	-0.02	0.06	0.00	0.04	0.03	-0.04	-0.03	1.00	0.02	0.02	-0.06	-0.07	-0.02	0.03
Equipment_Utilization	0.01	-0.00	-0.03	0.02	0.04	0.06	0.01	-0.03	0.02	-0.01	-0.01	-0.06	0.01	-0.04	0.02	1.00	-0.00	-0.00	-0.00	0.01	-0.00
Accident_Count	-0.02	-0.03	-0.04	0.02	0.02	0.01	-0.03	0.02	0.08	0.01	-0.00	-0.01	-0.06	0.04	0.02	-0.00	1.00	-0.05	-0.10	-0.02	-0.02
Safety_Risk_Score	0.06	0.06	0.04	0.03	0.02	-0.01	0.00	-0.01	0.07	-0.04	0.02	0.04	-0.01	-0.03	-0.06	-0.00	-0.05	1.00	0.06	-0.01	-0.01
Image_Analysis_Score	0.01	0.02	0.01	0.01	0.01	-0.01	-0.07	-0.03	0.04	0.04	0.04	0.01	-0.02	-0.03	-0.07	-0.00	-0.10	0.06	1.00	0.03	-0.08
Anomaly_Detected	0.02	0.03	0.03	0.01	0.01	0.00	0.00	-0.01	0.02	-0.01	0.05	-0.01	-0.03	0.02	-0.02	0.01	-0.02	-0.01	0.03	1.00	-0.00
Completion_Percentage	-0.00	0.01	0.02	0.00	0.01	0.02	0.04	0.00	0.00	0.05	0.01	-0.03	0.03	-0.04	0.03	-0.00	-0.02	-0.01	-0.08	-0.00	1.00

Figure 4. *correlation matrix heatmap*

This research has progressed to a stage where investigating feature importance is paramount for elucidating the contribution of specific variables that significantly influence project risk. Consequently, identifying these salient features is crucial, necessitating heightened consideration of these variables during subsequent data acquisition processes. Our analysis reveals that the safety risk score exhibits a predominant importance of approximately 78% relative to other features. However, additional variables warrant consideration, including anomaly detection at 17% and marginal importance of 1% each for humidity, schedule deviation, labor hours, and cost overrun.

Subsequently, this research investigated the performance of several machine learning models, specifically decision trees, k-nearest neighbors, logistic regression, random forests, support vector machines (SVM), and XGBoost. The results indicate that the SVM model demonstrated superior performance compared to the other models. This outcome may be attributed to the model's capacity for complex pattern recognition, often necessitating intensive parameter tuning. The SVM model achieved an accuracy of 95%, significantly exceeding the average model accuracy of 87%. Figure 4 visually presents the model accuracies and feature importance in the upper section, while the lower section illustrates the risk assessment summary and identifies the top five risk factors.



Risk Assessment Summary

Best Model
SVM / XGB (95%)

Primary Risk Factor
Safety Risk Score

Average Model Accuracy
87%

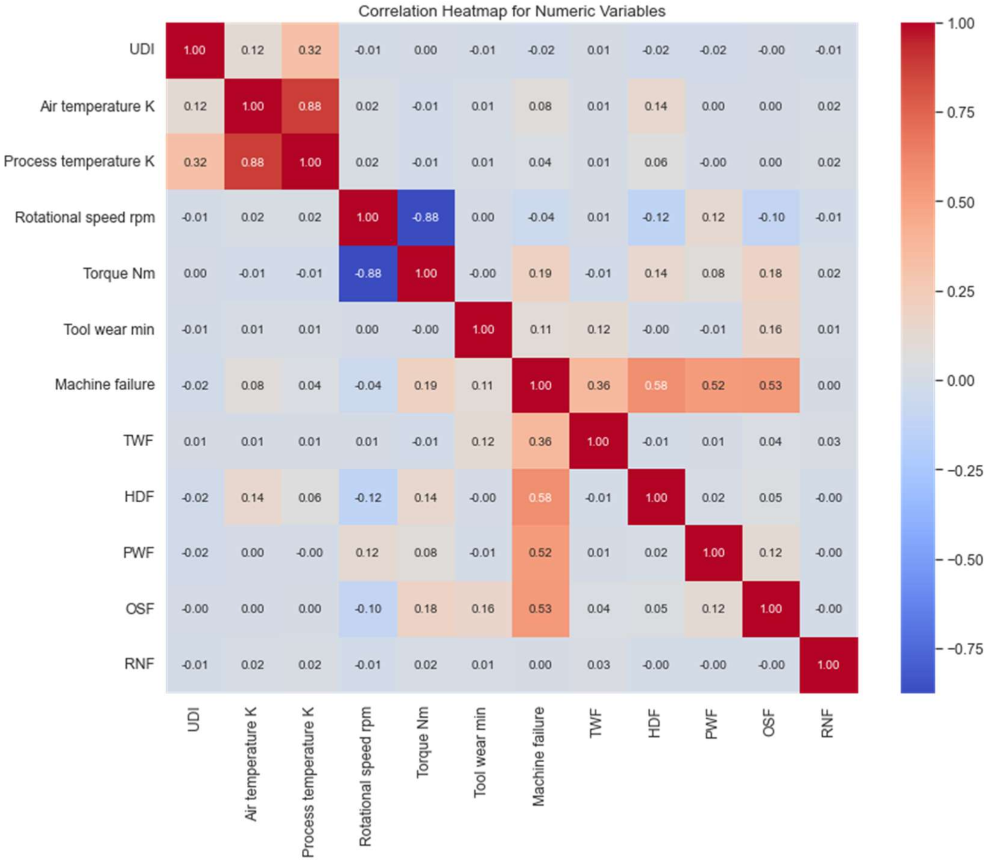
Secondary Risk Factor
Anomaly Detection

Top 5 Risk Factors

Safety Risk Score	77.7%
Anomaly Detected	17.4%
Actual Duration	1.6%
Humidity	1.1%
Schedule Deviation	0.9%

Figure 5 Results summary regarding civil project risk predictions

The second database investigates the current endeavor to establish a machine learning predictive maintenance model for mechanical equipment. Accordingly, the database shows the count of each equipment type on the research database; with around 6000 equipment for equipment type L and 3000 equipment for M type, the minor quantities are related to the H type with 1000 equipment. Moreover, it is essential to investigate the data distribution; the histogram is a well-known statistical tool that details how data is derived from sensors. For example, it shows if the data is normally distributed or has anomalies; these anomalies are essential as they are present when the data goes out of the normal limits, which can relate to construction asset failures. In addition to the histogram analysis, we investigated the correlation matrix in this research. Accordingly, this matrix presents the correlation between the research variables or features. In other words, it shows if there is a positive, negative, or neutral relation between variables; accordingly, its value varies from 1 to -1, where 1 represents a strong positive relation. At the same time, -1 denotes the negative relation between the research features. However, zero values represent no relation between variables. Figure 5 shows the heat map figures created to investigate the feature relation. It can be seen from the figure that there is a strong positive relationship among some variables, namely, air temperature, process temperature, and machine failures associated with TWF, HDF, PWF, OSF, and RNF, which are the terms that represent the machine failure indicators. Inversely, a different finding was found for rational speed and torque.

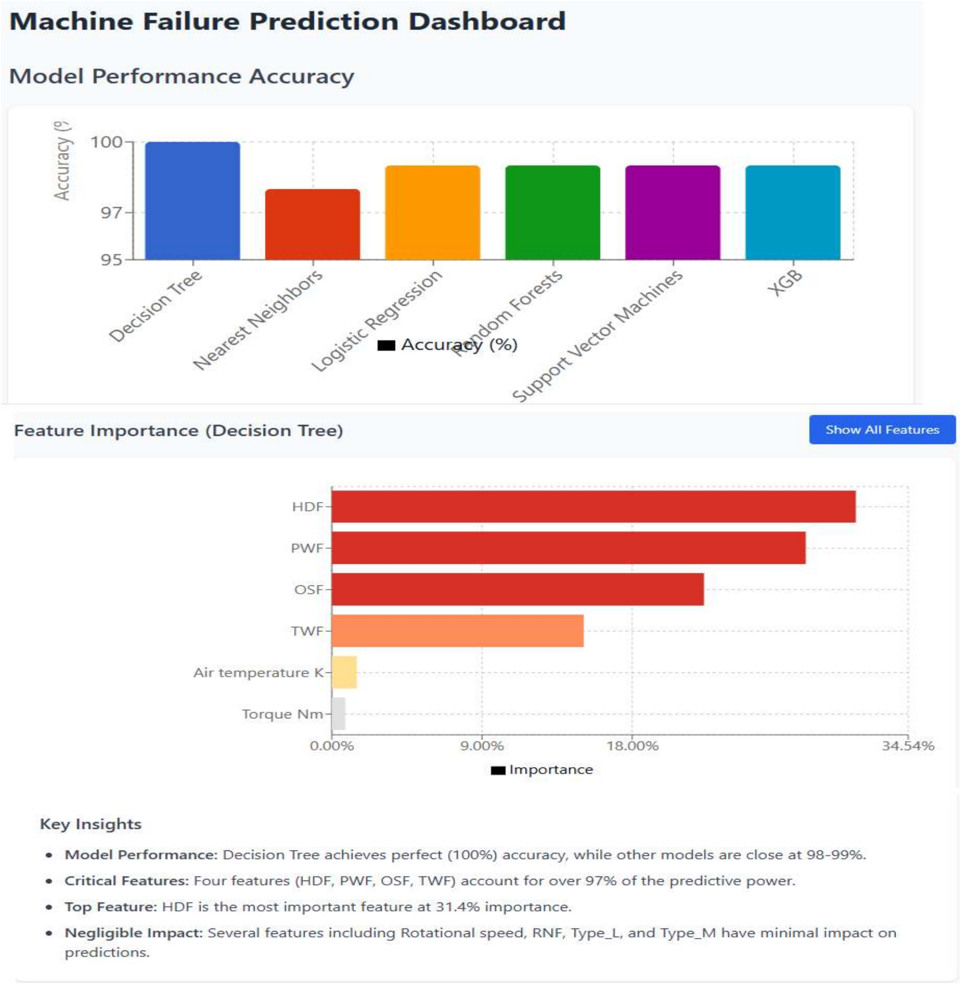


	UDI	Air temperature K	Process temperature K	Rotational speed rpm	Torque Nm	Tool wear min	Machine failure	TWF	HDF	PWF	OSF	RNF
UDI	1.00	0.12	0.32	-0.01	0.00	-0.01	-0.02	0.01	-0.02	-0.02	-0.00	-0.01
Air temperature K	0.12	1.00	0.88	0.02	-0.01	0.01	0.08	0.01	0.14	0.00	0.00	0.02
Process temperature K	0.32	0.88	1.00	0.02	-0.01	0.01	0.04	0.01	0.06	-0.00	0.00	0.02
Rotational speed rpm	-0.01	0.02	0.02	1.00	-0.88	0.00	-0.04	0.01	-0.12	0.12	-0.10	-0.01
Torque Nm	0.00	-0.01	-0.01	-0.88	1.00	-0.00	0.19	-0.01	0.14	0.08	0.18	0.02
Tool wear min	-0.01	0.01	0.01	0.00	-0.00	1.00	0.11	0.12	-0.00	-0.01	0.16	0.01
Machine failure	-0.02	0.08	0.04	-0.04	0.19	0.11	1.00	0.36	0.58	0.52	0.53	0.00
TWF	0.01	0.01	0.01	0.01	-0.01	0.12	0.36	1.00	-0.01	0.01	0.04	0.03
HDF	-0.02	0.14	0.06	-0.12	0.14	-0.00	0.58	-0.01	1.00	0.02	0.05	-0.00
PWF	-0.02	0.00	-0.00	0.12	0.08	-0.01	0.52	0.01	0.02	1.00	0.12	-0.00
OSF	-0.00	0.00	0.00	-0.10	0.18	0.16	0.53	0.04	0.05	0.12	1.00	-0.00
RNF	-0.01	0.02	0.02	-0.01	0.02	0.01	0.00	0.03	-0.00	-0.00	-0.00	1.00

Figure 6: The correlation matrix heatmap

This research has progressed to a stage where investigating feature importance is paramount for elucidating the contribution of specific variables that significantly influence equipment failures. Consequently, identifying these salient features is crucial, necessitating heightened consideration of these variables during subsequent data acquisition processes. Our analysis reveals that the HDF features exhibit a predominant importance of approximately 32% relative to other features. However, additional variables warrant consideration, including PWF, OSF, and TWF, with importance varying between 28% and 15%.

Subsequently, this research investigated the performance of several machine learning models, specifically decision trees, k-nearest neighbors, logistic regression, random forests, support vector machines (SVM), and XGBoost. The results indicate that the decision tree model demonstrated superior performance compared to the other models. This outcome may be attributed to the model's capacity for complex pattern recognition, often necessitating intensive parameter tuning. The decision tree model achieved an accuracy of 100%, significantly exceeding the average model accuracy of 99%. Figure 6 visually presents the model accuracies and feature importance in the upper section, while the lower section illustrates the risk assessment summary and identifies the top five important factors.



Key Insights

- **Model Performance:** Decision Tree achieves perfect (100%) accuracy, while other models are close at 98-99%.
- **Critical Features:** Four features (HDF, PWF, OSF, TWF) account for over 97% of the predictive power.
- **Top Feature:** HDF is the most important feature at 31.4% importance.
- **Negligible Impact:** Several features including Rotational speed, RNF, Type_L, and Type_M have minimal impact on predictions.

Figure 7 Results summary regarding machine failure predictions

I. CONCLUSION

The culmination of this research endeavor marks the development of an integrated Building Information Modeling (BIM) and machine learning framework designed for predictive maintenance. Findings robustly demonstrate the capacity of the developed machine learning model to forecast both project risks and potential equipment failures within construction projects. The investigation focused on analyzing two distinct BIM-based maintenance databases, underscoring the critical prerequisite of substantial historical data for the adequate validation, training, and rigorous performance assessment of the implemented machine learning algorithms. Notably, the study revealed a disparity in predictive accuracy between the two datasets: the mechanical equipment database exhibited higher predictability, with the performance of all evaluated machine learning models approaching 100% accuracy. In contrast to the infrastructure projects database. A significant challenge encountered during the research was the identification of suitable and sufficiently comprehensive BIM maintenance databases for establishing predictive maintenance capabilities. In conclusion, this research posits the considerable value of the proposed approach, highlighting its potential to significantly mitigate maintenance expenditures through proactive failure prediction and risk assessment.

Future research will focus on expanding the scope and sophistication of the current methodology. We plan to incorporate more diverse datasets from a broader range of buildings and systems to improve the model's generalizability. A significant next step is the real-time integration of IoT data, which will allow for continuous monitoring and more dynamic predictive analysis. We also intend to explore advanced deep learning models, such as Long Short-Term Memory (LSTM) networks for time-series forecasting and Graph Neural Networks (GNNs) to model complex interdependencies within building systems. To address the black box nature of these advanced models, we will incorporate explainability tools like SHAP and LIME, providing greater transparency and trust in the maintenance recommendations. Ultimately, the long-term goal is to directly integrate this predictive maintenance framework into a live Building Information Modeling (BIM) system for pilot testing, demonstrating its practical application and effectiveness in a real-world environment.

REFERENCES

[1] M. I. Khan, B. Ahmad, M. Shoaib, S. Anwar, and Z. Ali, “Artificial intelligence and machine learning applications in smart infrastructure and electrical systems,” *The Critical Review of Social Sciences Studies*, vol. 3, no. 1, pp. 2005–2016, 2025.

[2] O. O. Lawal, N. O. Nawari, and O. Lawal, “AI-enabled cognitive predictive maintenance of urban assets using city information modeling—Systematic review,” *Buildings*, vol. 15, no. 5, p. 690, 2025.

[3] V. Plevris and G. Papazafeiropoulos, “AI in structural health monitoring for infrastructure maintenance and safety,” *Infrastructures (Basel)*, vol. 9, no. 12, p. 225, 2024.

[4] U. J. Ahmad, S. A. Raza, and N. A. Channar, “REINFORCEMENT LEARNING FOR IOT-BASED PREDICTIVE MAINTENANCE: APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS,” *Spectrum of Engineering Sciences*, vol. 3, no. 7, pp. 876–887, 2025.

[5] D. Guidotti, L. Pandolfo, and L. Pulina, “A Systematic Literature Review of Supervised Machine Learning Techniques for Predictive Maintenance in Industry 4.0,” 2025. doi: 10.1109/ACCESS.2025.3578686.

[6] I. A. Shaban, H. Salem, A. Y. Abdullah, H. M. A. Q. Al Ameri, and M. M. Alnahdi, “Maintenance 4.0 for HVAC systems: Addressing implementation challenges and research gaps,” *Smart Cities*, vol. 8, no. 2, p. 66, 2025.

[7] M. S. H. Sikdar, M. S. Hossain, S. M. Mobin, A. Chowdhury, and S. M. Y. Bhuiyan, “ADVANCEMENTS IN SMART AND ENERGY-EFFICIENT HVAC SYSTEMS: A PRISMA-BASED SYSTEMATIC REVIEW,” *American Journal of Scholarly Research and Innovation*, vol. 3, no. 01, pp. 1–19, 2024.

[8] E. H. E. Suryadarma and T. J. Ai, “Predictive Maintenance in SCADA-Based Industries: A literature review,” *International Journal of Industrial Engineering and Engineering Management*, vol. 2, no. 1, pp. 57–70, 2020.

[9] Matplotlib, “Matplotlib: Python plotting — Matplotlib 3.1.2 documentation.” Accessed: Jan. 15, 2020. [Online]. Available: https://matplotlib.org/

[10] Matplotlib, “Matplotlib — Visualization with Python.” Accessed: Mar. 02, 2023. [Online]. Available: https://matplotlib.org/

[11] M. Waskom, “seaborn: statistical data visualization,” *J Open Source Softw*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/JOSS.03021.

[12] Plotly, “Plotly Python Graphing Library.” Accessed: Mar. 02, 2023. [Online]. Available: https://plotly.com/python/

[13] NumPy, “NumPy.” Accessed: Mar. 02, 2023. [Online]. Available: https://numpy.org/

[14] Pandas, “pandas - Python Data Analysis Library.” Accessed: Mar. 02, 2023. [Online]. Available: https://pandas.pydata.org/

[15] Keras, “Keras: the Python deep learning API.” Accessed: Mar. 02, 2023. [Online]. Available: https://keras.io/

[16] SciKit-Learn, “scikit-learn: machine learning in Python — scikit-learn 0.22.1 documentation.” Accessed: Jan. 15, 2020. [Online]. Available: https://scikit-learn.org/stable/index.html

[17] Scikit-learn, “scikit-learn: machine learning in Python — scikit-learn 1.2.1 documentation.” Accessed: Mar. 02, 2023. [Online]. Available: https://scikit-learn.org/stable/index.html