Koc. J. Sci. Eng., 8(2): (2025) 174-185 https://doi.org/10.34088/kojose.1632004

Kocaeli University

Kocaeli Journal of Science and Engineering

http://dergipark.org.tr/kojose

Improving the Maintenance and Repair Model of Kocaeli Akçaray Tram System with Data Mining

Melek ERTUĞ YILDIRIM [1], Umut ALTINIŞIK [2,*]

[1] Department of Directorate of Computing, UlaşımPark JSC, Kocaeli, Türkiye, **ORCID:** 0009-0003-6852-1734
[2] Department of Informatics, Kocaeli University, Kocaeli, Türkiye, **ORCID:** 0000-0003-3119-3338

Article Info

Research paper

Received : *February 4, 2025*
Accepted : *July 1, 2025*

Keywords

Data Mining
K-NN
Decision Trees
Deep Learning
Tram Maintenance and Repair

Abstract

Transportation is among the daily routines of many people and a large part of human life is spent in transportation. In a process where transportation has gained so much importance, it is important to regularly perform maintenance and repair of vehicles for comfortable, safe and fast transportation. In this study, it is aimed to improve the maintenance and repair times of Kocaeli Metropolitan Municipality tram operation. Therefore, Maintenance Repair Data Model is designed by applying K-NN (K-nearest neighbors) classification, Decision Trees and Deep Learning algorithms from Data Mining algorithms.Thus, the warranty periods of the trams are extended and the trams operate with maximum performance by looking at the mean time between failures and the length of maintenance work periods.Quantitative findings revealed that the K-NN algorithm achieved the highest performance with an accuracy rate of 87.6%, outperforming the Decision Tree 85.95% and Deep Learning 70.66% models. Moreover, the K-NN model demonstrated the most balanced classification performance, with a precision of 0.716, recall of 0.637, and an F1-score of 0.652. In contrast, the Deep Learning algorithm exhibited lower performance, with an F1-score of only 0.355, indicating its limited effectiveness when applied to structured maintenance data. These results suggest that in cases where the dataset is structured and relatively small in scale, simpler non-parametric models such as K-NN may offer more efficient and practical solutions for predicting maintenance durations.

1. Introduction

Throughout history, human beings have consistently demonstrated a tendency to migrate and relocate, both out of necessity and in pursuit of opportunity. In contemporary individual and social life, a substantial portion of daily activities is comprised of transportation.

Public transportation systems, supported by well-developed infrastructure, are the most efficient solution to the traffic congestion challenges faced by metropolitan areas. As reliance on public transit increases, the number of private vehicles on the road will decline, resulting in reduced travel times. This shift will alleviate current traffic congestion and contribute to significant economic savings. It is vital to promote the use of public transportation and ensure its continuous and efficient operation, which is essential for sustainable urban mobility. Achieving this goal requires public transport systems to be characterised by speed, safety, appropriate service frequency and a high degree of reliability [1].

Urban rail systems, such as trams, require effective maintenance strategies. This is essential to prevent service disruptions and ensure operational reliability. Unexpected failures not only compromise service availability but also lead to increased repair costs and pose significant safety risks [2]. Maintenance in transit systems has historically been reactive or periodic. Conversely, predictive maintenance employs data-driven approaches to anticipate failures before they occur, thereby enabling timely interventions that reduce downtime and associated costs. With the widespread deployment of sensors and the growing availability of operational data, machine learning and data mining techniques have emerged as critical tools for implementing predictive maintenance in the transportation sector [3].

Although tram performance tends to decrease on narrow streets and during periods of heavy traffic, it improves significantly in areas and times with lighter

* Corresponding Author: umuta@kocaeli.edu.tr

traffic conditions. In the long term, trams are considered to represent a healthier, safer, more economical and environmentally sustainable mode of public transportation [4, 5]. In order to guarantee the reliability and dependability of tram services, it is essential to implement maintenance strategies that are not only efficient but also designed to provide consistent comfort and operational performance [6–8].

The rapid evolution of technology has markedly improved the efficiency of data collection and storage processes. Nevertheless, the interpretation and analysis of large-scale datasets remains challenging due to their substantial volume, structural complexity, and heterogeneity. Organisations that transition from traditional methods to data-driven, technologically integrated approaches are more likely to optimize system performance and sustain a competitive advantage. To leverage the full potential of big data, it is imperative to extract actionable insights using analytical techniques that are both accurate and methodologically sound. In the context of public transportation systems characterised by high daily passenger volumes, the assurance of service reliability and availability is contingent upon the efficiency and effectiveness of maintenance and repair operations [9].

Recent research has highlighted the effectiveness of machine learning algorithms in the context of predictive maintenance for railway infrastructure. In particular, interpretable classification models developed using decision trees, random forests, and gradient boosting have been successfully employed to predict maintenance requirements for railway track switches based on existing operational data. These models have been demonstrated to exhibit high levels of accuracy while also providing transparent decision-making support through feature importance analysis, which is especially valuable in transportation systems. Inspired by these advancements, the present study incorporates a decision tree model to classify tram maintenance durations. Furthermore, the K-NN algorithm is employed due to its proven effectiveness in structured maintenance datasets and its ability to deliver reliable results with minimal parameter tuning. In conjunction with a deep learning model, which serves as a benchmark, these algorithms are applied to real-world tram maintenance records. This enables an assessment of their predictive performance and facilitates more efficient maintenance planning in metro rail operations [10-11].

In the tram maintenance and analysis process, evaluating the operability of trams is of critical importance. In this context, the calculation of Mean Time between Failures (MTBF) and Availability Value (AV) serves as key performance indicators for assessing the operational suitability of trams, particularly during the warranty period [12, 13]. Data mining methods are increasingly utilized to analyze maintenance data and establish a foundation for decision support systems [14]. These methods facilitate the determination of optimal maintenance and repair intervals [15, 16] and enable the classification of faults occurring on tram lines [17].

The Kocaeli Akçaray tram system in Turkey is a modern light rail network that has been in operation since 2017. As the system matures, a substantial volume of maintenance and operational data has been accumulated, presenting a valuable opportunity to apply predictive analytics for optimizing maintenance and repair processes. This study utilises historical maintenance records and operational data from the Akçaray system to develop models that predict the likelihood of maintenance being required by a tram or a specific component in the near future. The implementation of proactive maintenance scheduling, predicated on such predictions, has the potential to curtail unanticipated breakdowns, enhance service reliability, and optimise resource utilisation. The average number of passengers utilising the tram system in Kocaeli is 42,786 per day. In order to ensure consistent operational performance, the system requires efficient and data-driven maintenance strategies. The improvement of repair times, the forecasting of potential failures, and the formulation of effective maintenance plans are all essential to the sustaining of high service quality and the minimizing of disruptions. To address these challenges, the present study examines the maintenance personnel's work completion times and applies three data mining techniques: decision trees, deep learning, and K-NN predict and compare maintenance durations.

The primary objectives of this study are to compare multiple data mining techniques in a tram system, and to evaluate these models using robust validation techniques and performance metrics to demonstrate improvements in the accuracy and reliability of maintenance scheduling. In addition, a method was developed to calculate MTBF, which enables the assessment of system performance during the warranty period. By analyzing both MTBF and maintenance durations, the study aims to improve predictive maintenance strategies, extend warranty coverage, and ensure optimal operational efficiency of the tram fleet.

The performance of the decision tree, deep learning, and K-NN models was compared to assess their effectiveness in predicting maintenance requirements. In this context, a maintenance and repair evaluation system is proposed to assess and optimize the warranty periods of tram components, thereby contributing to more reliable, proactive, and cost-effective operations.

The following section describes the process of acquiring the dataset and the design of the proposed system. Section three presents the results obtained from the three data mining methods used in the model along with a comparative analysis of their performance. Finally, section four concludes the study by summarizing the key findings and discussing their implications.

2. Materials and Methods

This study uses the maintenance and repair data of trams operated by UlaşımPark J.C., a subsidiary of the Kocaeli Metropolitan Municipality. In the evaluation phase of the data, three data mining methods were applied, namely Decision Tree, K-NN and Deep Learning. These methods were used to estimate tram maintenance intervals based on the analysis of maintenance and repair records. All data related to tram operations are stored in a Microsoft SQL Server database. Fault and maintenance data were retrieved from the database by writing queries using the Structured Query Language known as SQL.

Furthermore, the RapidMiner software platform was utilised for the purpose of conducting data analysis and constructing predictive models. RapidMiner is a software environment designed to support a wide range of analytical tasks. These include machine learning, pattern recognition, data preparation, clustering, classification, text mining and deep learning. The software offers a user-friendly interface for the construction and evaluation of predictive models, thus rendering it suitable for both researchers and practitioners. The development of this software was initiated by Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer. [18, 19]. RapidMiner's open-source architecture enhances its accessibility and adaptability across diverse research applications. Its intuitive graphical interface for result visualization further supports its broad adoption in academic settings [20].

2.1. Entering the Used Data into the System

The present study utilised 2023 tram maintenance and repair data belonging to Kocaeli Akçaray trams. The mean interval between failures was determined on the basis of the dataset. In addition, three data mining algorithms were implemented to analyze the maintenance and repair data [21]. In the present study, a total of 604 maintenance and repair records from the year 2023 were analyzed. During the evaluation of maintenance durations, average maintenance times were calculated based on specific failure codes.

Maintenance instances with durations exceeding the average were classified as "Long", those equal to the average as "Normal", and those below the average as "Short." These classifications were integrated into the Maintenance and Repair Model process. Subsequently, three different data mining models were employed to estimate maintenance durations based on these categories.

The data used in the RapidMiner software was transferred from the database following the completion of data cleaning, data reduction, and data integration processes. A subset of the dataset, consisting of 604 maintenance and repair records, was imported into the RapidMiner environment. This dataset includes 90 distinct fault types that occurred across 26 different hardware components and were addressed by 15 maintenance personnel. A sample section of this dataset is presented in Table 1.

Table 1. The maintenance repair data

Case	Job Start Date	Start Time	Job End Date	End Time	Time	Worker ID	Fault Name
Long	2023-01-11	09:00	2023-01-11	10:00	60	81	Lcd Display Motherboard Failure
Short	2023-07-10	08:30	2023-07-10	17:30	540	83	Ovality Disorder
Normal	2023-01-12	10:45	2023-01-12	11:15	30	83	Brake Hydraulic Pipe Failure
Short	2023-01-12	10:45	2023-01-12	11:00	15	82	Driver Seat Mechanism Failure
Short	2023-01-12	10:00	2023-01-12	10:30	30	1393	Instructor Seat Malfunction
Long	2023-01-13	09:00	2023-01-13	10:00	60	574	Driver's door Mulfunction
Normal	2023-01-13	09:00	2023-01-13	09:00	10	82	Parking light Malfunction
Short	2023-01-13	09:00	2023-01-13	10:00	60	88	Air conditioner Condenser Fan Failure
Long	2023-01-16	08:28	2023-01-16	09:00	32	88	Camera CCTV Unit Failure
Short	2023-01-16	08:32	2023-01-16	10:00	88	88	Air conditioner Condenser Fan Failure
Long	2023-01-16	15:00	2023-01-16	16:00	60	82	Acceleration Arm moving Malfunction
Short	2023-01-16	10:42	2023-01-16	10:45	3	124	HMI Communication Failure
Short	2023-01-16	08:00	2023-01-16	08:30	30	1393	Driver Seat Mechanism Failure
Long	2023-01-17	08:30	2023-01-17	09:00	30	574	Armrest Mechanical Failure
Short	2023-07-10	08:30	2023-07-10	17:30	540	83	Ovality Disorder

As demonstrated in Table 1, the fault maintenance durations were evaluated and categorised as either long, short, or normal. The table also includes information regarding the personnel who recorded the faults and the corresponding time periods. The necessity to access and analyse the data necessitated navigation across multiple tables within the database, a process that was rendered particularly arduous by how the required data had been stored. In order to address this issue, virtual tables were created using SQL to consolidate the relevant data. The subsequent data refinement procedure was applied to the information extracted from the virtual tables. The process entailed the analysis and correction of missing or inaccurately entered data.

The original dataset included multiple columns related to timestamps, fault descriptions, maintenance duration, worker IDs, and additional metadata. Columns containing extensive missing data (>20%) or irrelevant information were excluded. Missing values in critical columns such as maintenance duration and fault categories were imputed using median values calculated specifically for each fault type. Additionally, data inconsistencies, notably in fault descriptions and component identifiers, were standardized through SQL queries to ensure accurate integration. Consequently, a consolidated Maintenance and Repair Data Virtual Table (MRDVT) was established, containing the refined and structured dataset necessary for subsequent analysis.

Figure 1 illustrates the formation of the MRDVT, which includes key attributes such as operation number, equipment code, equipment name, maintenance code, operation start and end dates, maintenance type, personnel responsible, stock code, and warranty start and end dates.

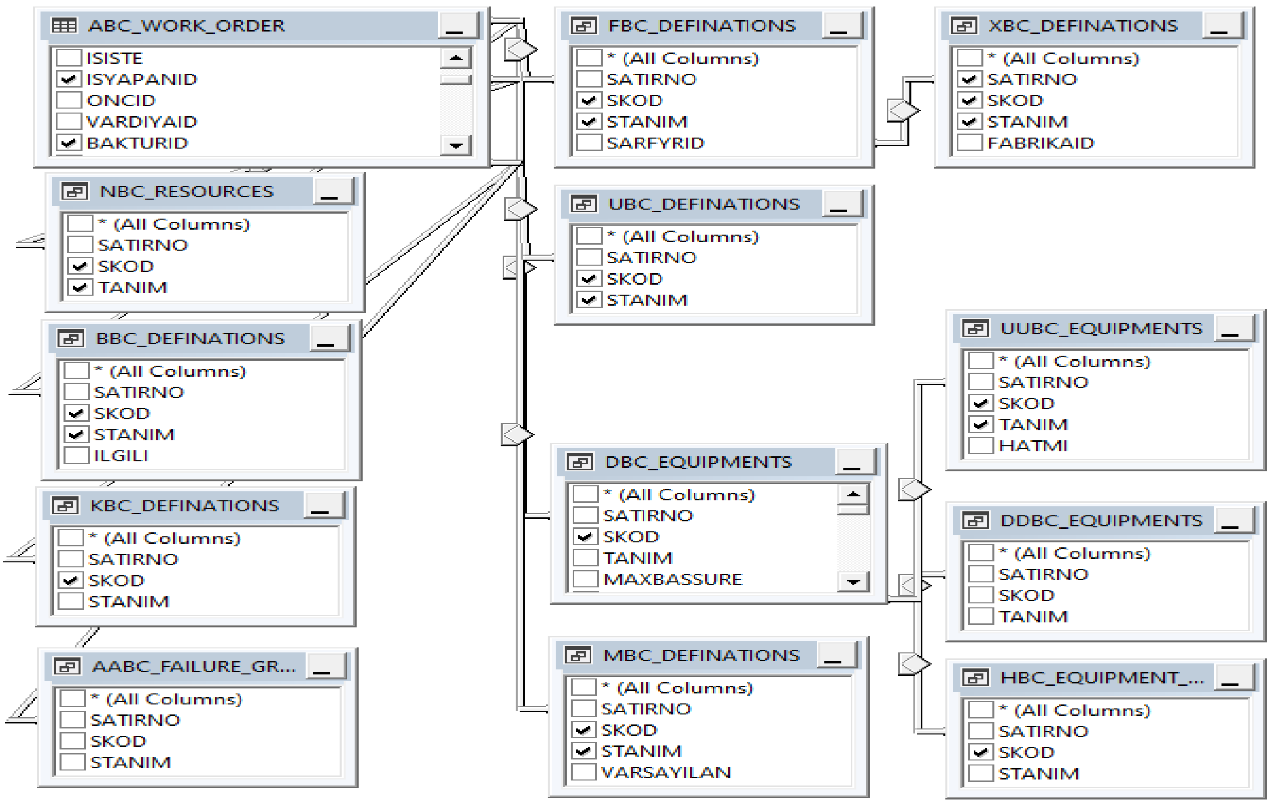

Figure 1. Maintenance and repair data virtual table

2.2. Calculating Tram Operability

Using the data obtained from the MRDVT, key reliability and availability metrics were calculated for the tram system. These are MTBF, Mean Time Between Service Failures (MTBSF), and AV. These metrics reflect the operational availability of the trams during the warranty period. The calculations were based on data collected over a three-month period. MTBF represents the average operating time between failures and is calculated by dividing the total working time (TWT) over the three months by the total number of failures (TNF). The formula for MTBF is presented in Eq. (1).

$$MTBF = TWT / TNF \quad (1)$$

MTBSF provides an estimate of the average time between two consecutive failures of types A and B in repairable systems [22]. This metric is particularly useful for evaluating the reliability of specific failure types that require service intervention.

An MTBSF value greater than 500 is generally considered acceptable for maintenance planning. MTBSF is calculated by dividing TWT over a three-month period by the total number of type A and B failures, denoted as [A/B] TNF. The formula for calculating MTBSF is presented in Eq. (2).

$$MTBSF=TWT / [A/B]\ TNF \quad (2)$$

AV represents the proportion of time that a repairable system remains operational, excluding periods of downtime due to faults. It is calculated by subtracting the total downtime (TD) from TWT and dividing the result by TWT. An AV value exceeding 98% is generally considered acceptable for operational performance. The formula for calculating AV is presented in Eq. (3).

$$AV=TWT–TD / TWT \quad (3)$$

The Tram Usability Trace (TUT) was developed using SQL to calculate MTBF, MTBSF, and AV values based on data obtained from the MRDVT. To ensure consistent and up-to-date calculations, a scheduled work order was established, allowing TUT to operate at regular intervals. The algorithm implemented within TUT for computing MTBF, MTBSF, and AV is illustrated in Figure 2. In Figure 2, the algorithm begins by transferring data from the MRDVT, ordered by equipment code and start date. In this initial stage, any data entries with start dates earlier than the first day of the month are adjusted to align with the corresponding month. In the second stage, once all data are processed, the MTBF, MTBSF, and AV values are automatically calculated for each three-month period and recorded in the U_MTBF table within the database. In the final stage, if the calculated TUT values fall below the predefined thresholds, it is recommended that the warranty periods of the affected trams be extended.

Furthermore, the maintenance durations associated with personnel responsible for tram servicing are reviewed, and relevant parameters are optimized to ensure maximum operational performance of the tram system.

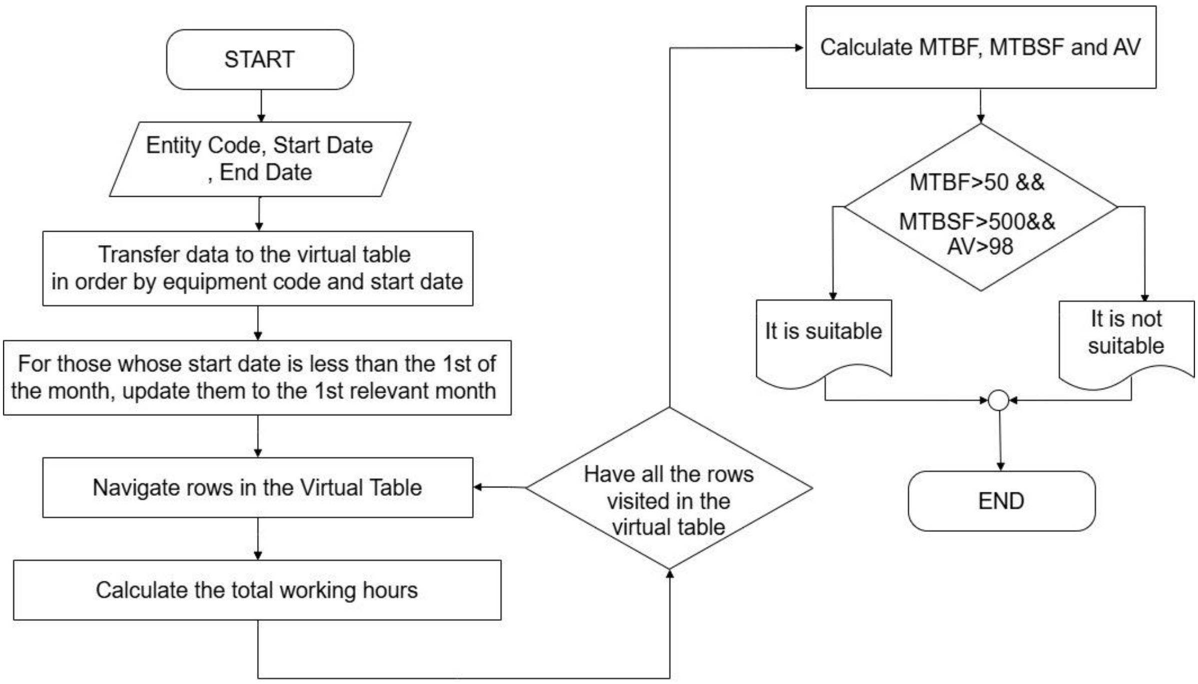

Figure 2. Tram usability trace algorithm

2.3. Maintenance and Repair Model

The Maintenance and Repair Model developed in this study is constructed using three data mining techniques: Decision Tree, K-NN, and Deep Learning. The dataset was then divided into a training set and a test set using the Data Splitting operator, with 60% of the data allocated for training and 40% reserved for testing. After the partitioning process, the model was trained and tested using the selected algorithms. As illustrated in Figure 3, the data retrieved from the MRDVT was integrated into the modelling process. The Set Role operator was applied in order to define the target variable, which was designated as the length of the maintenance period. The Select Attributes operator was then used to specify the input features for the estimation process. The features in question encompassed the fault code, the duration of maintenance, and the average maintenance duration values.

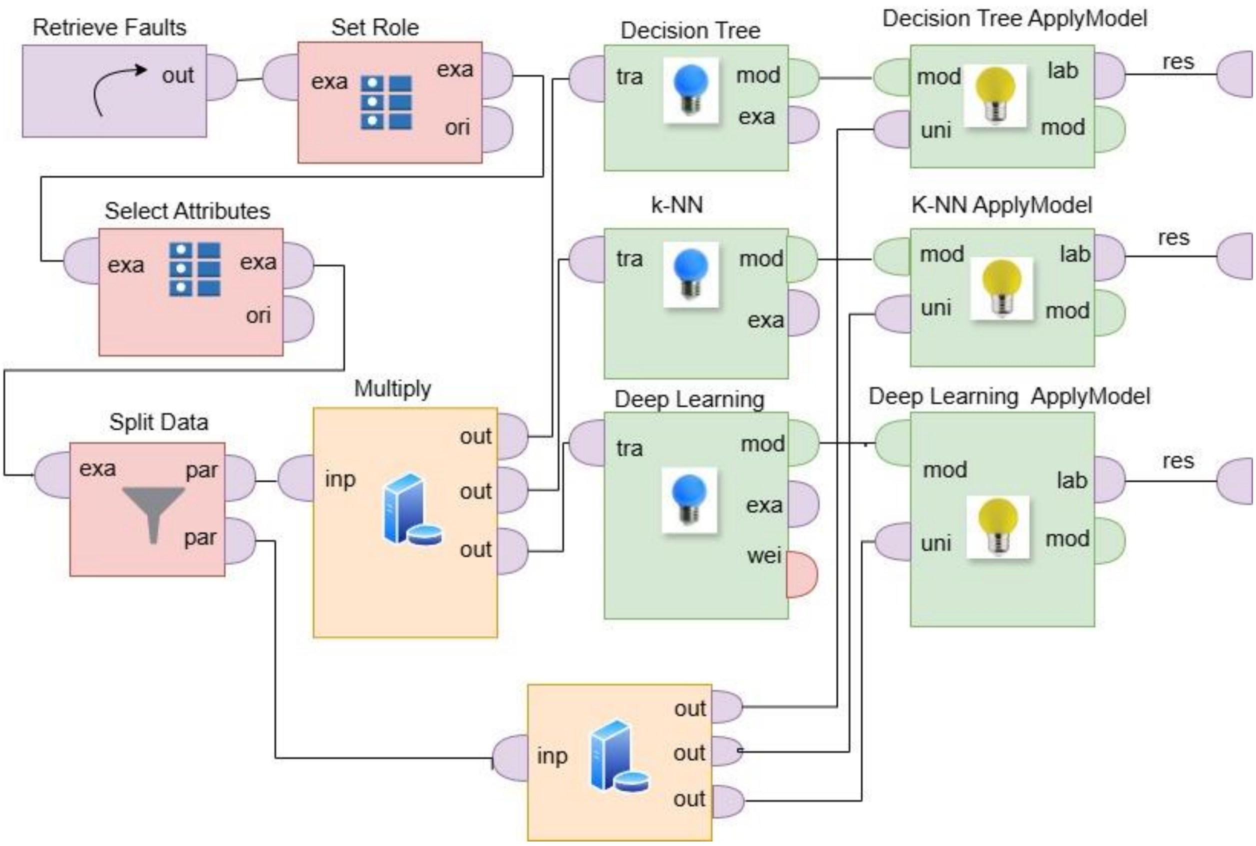

Figure 3. Maintenance and repair model

In this study, a predictive maintenance framework is proposed to with a view to enhancing the efficiency and reliability of tram maintenance planning by incorporating machine learning techniques into the Maintenance Repair and Data Model System (MRDMS). As illustrated in Figure 4, the process begins with the extraction and organization of average maintenance duration data, which serves as the primary input for model development.

The dataset is divided into training and prediction subsets and processed using three supervised learning algorithms: K-NN, DecisionTree, and Deep Learning. Each model is trained on historical maintenance records and applied to classify upcoming maintenance events into one of three predefined categories: The duration of the programme is categorised as short, normal or long. The classification outputs are evaluated using four commonly adopted performance metrics: accuracy, precision, recall, and F1 score.

The resulting performance measures are then used to refine the MRDM process and guide data-informed decision-making. By embedding predictive analytics into the overall maintenance framework, the proposed system not only improves classification accuracy but also supports proactive, data driven maintenance scheduling within metro rail operation.

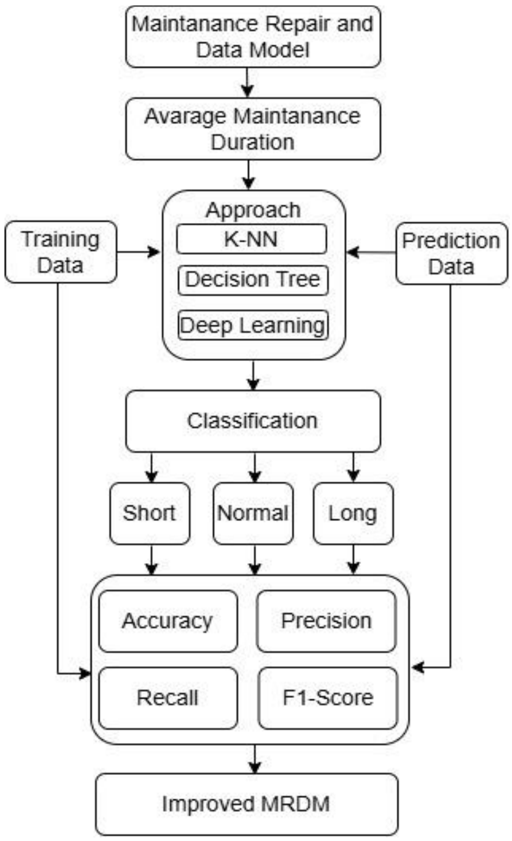

Figure 4. Maintenance repair and data model system

2.3.1. Decision Tree Model

Within the MRDMS process, the Decision Tree model incorporates seven tunable parameters, among which the criterion used to evaluate the quality of splits is particularly influential. Five distinct options are available for this parameter: information gain, which selects the attribute with the highest reduction in entropy; gain ratio, which adjusts information gain by accounting for the intrinsic information of attributes; Gini impurity [23], which selects the split that minimizes impurity within the resulting subsets; accuracy, which directly optimizes classification correctness at each node; and least squares, typically employed in regression tasks to minimize the sum of squared prediction errors. Each criterion reflects a different strategy for balancing partition purity and predictive relevance.

Based on empirical evaluation, the information gain criterion yielded the most accurate results and was therefore adopted in this study. To mitigate the risk of overfitting and maintain model interpretability, the maximum tree depth was constrained to ten levels. Furthermore, pruning was enabled to eliminate branches that failed to improve generalization performance, thus enhancing the robustness of the model across unseen data.

In addition, the minimum number of instances per leaf node was set to 2. This constraint prevents excessive fragmentation of the decision space and helps maintain meaningful terminal nodes. By ensuring that each leaf contains a representative sample, the model strikes a balance between complexity and generalization, reducing the likelihood of high-variance behavior and improving stability during inference.

2.3.2. K-NN Model

In the K-NN method, when a data point is to be classified, it is assigned to the same class as the data point that is closest to it. In a dataset, data points with similar characteristics that are in close proximity to one another are considered neighbors. The fundamental principle of this algorithm is to imitate the behavior of nearby data points. When the behavior of a specific record is to be predicted, the behaviors of ten similar records located near it in the data space can be observed. The average behavior of these ten records is then calculated and used as the predicted value for the given record.

During the training phase, the K-NN model is generally referred to as a lazy learner, as it does not construct an explicit model but rather stores the entire training dataset for reference. With respect to measurement types, textual measurements are applied only to categorical attributes, numerical measurements are used for continuous attributes, and mixed measurements areapplied to datasets containing both types. In the K-NN model implemented using the RapidMiner platform, the parameters were set as follows: the number of neighbors was set to three, the measurement type was numerical, and the distance metric used was Euclidean distance.

The value of k in the K-NN method represents the number of neighboring data points to be considered [24]. To determine the optimal value of k, several values were tested using cross-validation on the training dataset. Specifically, values of k equal to 1, 3, 5, 7, and 9 were evaluated. Model performance improved from k equal to 1 to k equal to 3, after which accuracy declined as the value of k increased further. The highest validation accuracy, approximately 88%, was observed when k was set to 3. This outcome reflects a balance Between bias and variance. A k value of 1 is highly sensitive to noise and may lead to overfitting, as it relies on only one nearest data point, which might be an outlier.

On the other hand, a moderately larger neighborhood, such as k equal to 3, provides a smoothing effect that reduces the influence of outliers while preserving local data structure. For this reason, k equal to 3 was selected as the optimal value for the final model. Additionally, uniform distance weighting was applied, meaning that each neighbors contributed equally to the classification process. Although others trategies, such as distance-based weighting, were tested, they did not yield not able improvements in predictive performance and were therefore excluded from the final configuration.

2.3.3. Deep Learning Model

In this model, the activation function, the number of repetitions of the dataset, the loss function, the number of hidden layers, and the number of neurons in each layer were selected as key parameters [25, 26]. The Tanh function, which is nonlinear and produces output in the range from minus one to one, was chosen as the activation function. This function generates negative outputs for negative inputs and positive outputs for positive inputs, allowing for a more balanced activation. The number of dataset repetitions was set to ten.

In this study, a deep learning model was employed, consisting of two hidden layers positioned between the input and output layers, each containing 50 neurons. This architectural choice was guided by both practical constraints and the complexity of the task. The dataset includes 604 instances, which limits the applicability of deeper architectures due to the increased risk of overfitting and computational inefficiency. Given the moderate complexity of the problem and the absence of highly abstract feature hierarchies, a two-layer structure was deemed sufficient to capture non-linear relationships. Additionally, the model was trained using the cross-entropy loss function, which is commonly utilized in classification tasks. Cross-entropy loss quantifies the divergence between predicted and actual values, increasing as the discrepancy grows; an ideal loss value of zero indicates a perfect prediction. This combination of architecture and loss function ensured an appropriate balance between model complexity, generalization ability, and classification performance.

3. Results

In the study, 40% of the 604 maintenance and repair data were used as test data. Within this subset,, 162 were classified as short, 64 as long, and 16 as normal in terms of repair time. As shown in Figure 5, performance operators were added to the Maintenance and Repair Model process to compare the effectiveness of the Decision Tree, K-NN and Deep Learning data mining methods.

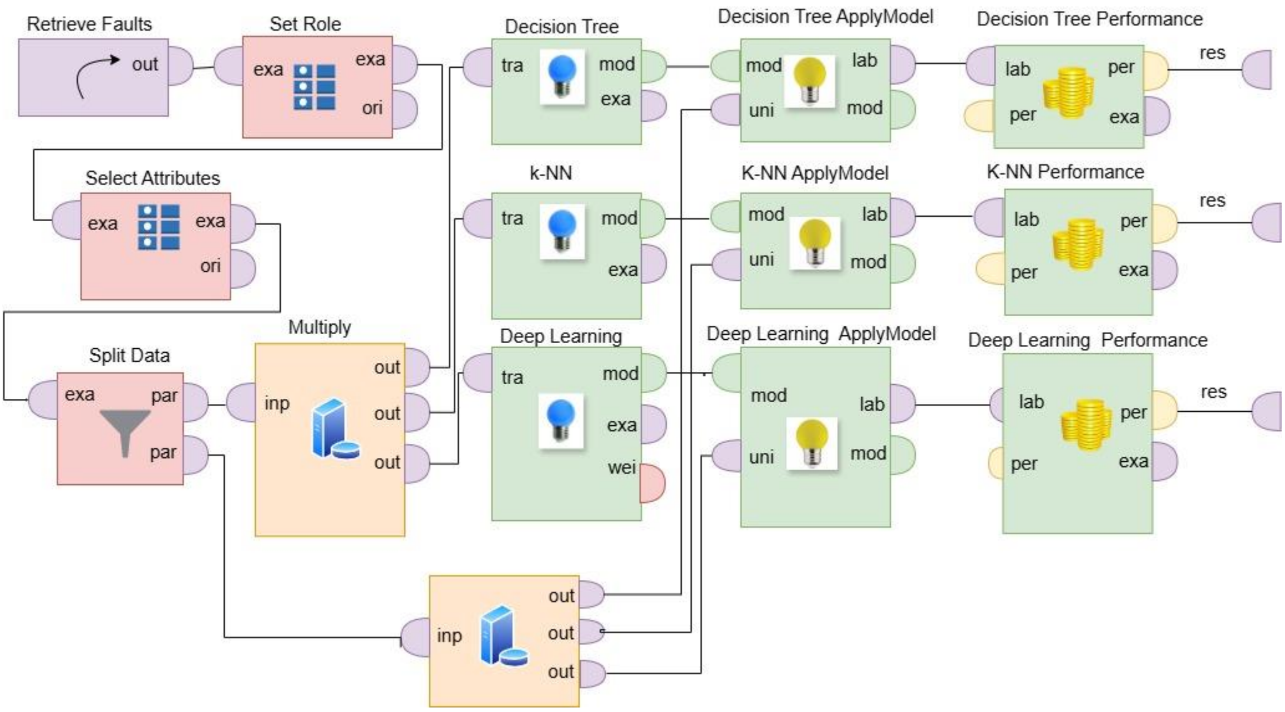

Figure 5. Maintenance and repair performance model

3.1. Decision Tree Model Performance

When the Decision Tree method was used to estimate the average maintenance period data, a success rate of 85.95% was achieved, which was higher than the Deep Learning Data Mining method and lower than the K-NN Data Mining method. To make a comparison with the other two methods, the same amount of training and prediction data is applied to the same process and the average maintenance time results are presented in Table 2.

As shown in Table 2, for the 162 instances that were actually classified as Short, the model correctly predicted 158 as Short, resulting in an accuracy rate of 97.53% for this category. Among the remaining cases, 4 were incorrectly classified as Long, and none were classified as Normal. For the 64 instances that were actually as Long, the model correctly predicted 50, while misclassifying 14 as Short and none as Normal, achieving an accuracy rate of 78.12%. In the case of the 16 instances that were truly Normal, the model misclassified 8 as Short and 8 as Long, with no correct predictions made, resulting in 0% accuracy for this category. Overall, taking both correct and incorrect classifications into account, the K-NN model achieved an average prediction accuracy of 85.95%.

Table 2. The confusion matrix of the decision tree model

Actual \ Predicted	Short	Long	Normal	Total
Short	158	4	0	162
Long	14	50	0	64
Normal	8	8	0	16
Total	180	62	0	242

3.2. K-NN Model Performance

When the K-NN model was used to predict the average maintenance time, a higher performance was achieved compared to the Deep Learning and Decision Tree Data Mining methods with a success rate of 87.60%. To make a comparison with other methods, the same process was estimated by looking at the 3 nearest neighbors with the K-NN model using the same amount of training and prediction data. Table 3 shows the results obtained according to the average maintenance time.

As shown in Table 3, for the 162 instances that were actually classified as Short, the model correctly predicted 158 as Short, resulting in an accuracy rate of 97.53% for this category. Among the remaining predictions, 2 were incorrectly classified as Long and 2 as Normal. For the 64 instances that were actually Long, the model predicted 52 correctly, while misclassifying 11 as Short and 1 as Normal, yielding an accuracy rate of 81.25% for the Long category. Regarding the 16 instances that were actually Normal, the model predicted 2 correctly, while misclassifying 7 as Short and 7 as Long, resulting in a prediction accuracy of 12.50% for this category.

Overall, considering both accurate and inaccurate predictions across all categories, the K-NN model achieved an average classification accuracy of 87.60%.

Table 3. The **c**onfusion matrix of the k-nn model

Actual \ Predicted	Short	Long	Normal	Total
Short	158	2	2	162
Long	11	52	1	64
Normal	7	7	2	16
Total	176	61	5	242

3.3. Deep Learning Model Performance

When the deep learning model performed to estimate the average maintenance period data, it demonstrated lower performance compared to the K-NN and Decision Tree data mining methods, with a success rate of 70.66%.

As shown in Table 4, when there are 162 actual instances classified as Short, the model correctly predicted all 162 of them as Short. Therefore, the prediction accuracy for the Short category was 100%. Additionally, for these Short instances, the number of predictions labeled as Long was 0, and the number of predictions labeled as Normal was also 0.

In the case where 64 instances were actually classified as Long, the model predicted 55 as Short, 9 as Long, and 0 as Normal. Consequently, the correct prediction rate for the Long category was 14.06%. Similarly, for the 16 instances that were truly classified as Normal, the model incorrectly predicted all 16 as Short, and none as Long or Normal, resulting in a 0% accuracy for the Normal category.

Considering both accurate and inaccurate predictions across all categories, the Deep Learning model achieved an overall prediction accuracy of 70.66%.

Table 4. The confusion matrix of the deep learning model

Actual \ Predicted	Short	Long	Normal	Total
Short	162	0	0	162
Long	55	9	0	64
Normal	16	0	0	16
Total	233	9	0	242

3.4. Performance Comparison of Models

The results showing the performance of the Deep Learning, K-NN, and Decision Tree models, based on variations in the amount of training and test data in the Maintenance Repair and Data Model process, are presented in Table 5.

Table 5. The performances with different training/testing data amounts

Training Data / Test Data	Guess Rate		
	Deep Learning	Decision Tree	K-NN
%70/%30	70.33	83.52	81.87
%60/%40	70.66	85.95	87.60
%50/%50	71.19	76.49	80.46
%40/%60	68.78	80.66	81.49

Upon reviewing the data presented in Table 5, it is evident that the most optimal results were obtained when the dataset was divided into 60% training data and 40% test data. When the average maintenance duration data is analyzed using Deep Learning, K-NN, and Decision Tree methods, applied under the same training and prediction data conditions, clear performance differences emerge. The Deep Learning model, despite its reputation for handling complex, nonlinear patterns, achieves only 70.66% accuracy, significantly lagging behind K-NN (87.60%) and Decision Tree (85.95%).

This suggests that in this specific maintenance prediction context, the complexity of Deep Learning architectures does not translate into better overall correctness. Among the evaluated methods K-NN model attained the highest accuracy at 87.60%. These results indicate that, among the methods evaluated, the K-NN model demonstrated the best overall performance.

An analysis of the training and testing performance revealed minimal discrepancies, indicating a good balance between bias and variance, particularly for K-NN and Decision Tree methods. Cross-validation confirmed stable performance across folds, suggesting that overfitting was effectively avoided. Conversely, the shallow neural network's lower accuracy points toward possible underfitting due to limited complexity and relatively few parameters. Future investigations may explore deeper network architectures or expanded datasets to enhance neural network performance.

As shown in Table 6, a deeper analysis of the models'performance across multiplee valuation metrics reveals that K-NN consistently outperforms both Deep Learning and Decision Tree methods.

Table 6. The comparative performance metrics of k-nn, deep learning and decision tree model

Method\ Metric	**Accuracy**	**Precision**	**Recall**	**F1-Score**
K-NN	0.876	0.716	0.637	0.652
Deep Learning	0.706	0.565	0.380	0.355
Decision Tree	0.860	0.561	0.585	0.572

The K-NN model achieves the highest accuracy at 0.876, indicating its strong overall predictive performance. The Decision Tree model follows closely with an accuracy of 0.860, while the Deep Learning model performs notably lower, with an accuracy of 0.706. These results suggest that K-NN is the most reliable model in terms of correctly predicting outcomes across the dataset.

Precision, which measures the proportion of correctly identified positive predictions, is highest for K-NN at 0.716, compared to 0.565 for Deep Learning and 0.561 for the Decision Tree. This indicates that K-NN produces fewer false positives and is more dependable when predicting positive maintenance events.

In terms of recall, which assesses how well each model captures all actual positive cases, K-NN again outperforms with 0.637, compared to Decision Tree (0.585) and Deep Learning (0.380). The poor recall of Deep Learning highlights its difficulty in identifying relevant maintenance cases, likely missing critical predictions [27].

Finally, considering the F1-score, which balances precision and recall, K-NN achieves the highest value at 0.652, followed by Decision Tree at 0.572, while Deep Learning falls drastically behind with a score of 0.355.

These quantitative findings collectively demonstrate that K-NN not only excels in raw accuracy but also offers the best balance between minimizing false alarms and capturing true maintenance needs.

Consequently, for the MRDMS, K-NN emerges as the most effective approach, while the Deep Learning model requires further optimization or may simply be unsuitable given the characteristics of the available data.

Table 7 provides a comparative overview of recent predictive maintenance studies, emphasizing the performance of the K-NN algorithm relative to alternative methodologies across various application domains. Esteban et al. (2022) demonstrated approximately 85% accuracy using K-NN in structured predictive maintenance datasets, highlighting its robustness and effectiveness in scenarios with limited complexity. Similarly, Mahale et al. (2025) reported an accuracy of 88% for vehicle maintenance, underscoring that K-NN outperformed neural network models largely due to its simplicity, computational efficiency, and reduced risk of overfitting, especially in datasets of moderate size.

Our study complements these findings by achieving an accuracy of 87.6%, reinforcing the suitability and efficiency of K-NN for structured tram maintenance datasets. Notably, Le-Nguyen et al. (2023) used a Decision Tree model for railway maintenance data, achieving slightly lower accuracy (83%) despite offering good interpretability. This comparison underlines that while Decision Trees provide transparency and ease of understanding, K-NN remains preferable in terms of accuracy and balanced predictive performance for structured, relatively small datasets typical of maintenance prediction tasks. Consequently, our results further confirm the advantage of K-NN for reliable, practical predictive maintenance implementations.

Table 7. The comparative analysis of predictive maintenance studies

Refs.	**Model**	**Acc. (%)**	**Dataset Type**	**Results**
Esteban et al. (2022) [22]	K-NN	~85	Predictive Maint.	Effective for small and structured datasets
Mahale et al. (2025) [2]	K-NN	88	Vehicle Maint.	Outperformed neural networks due to simplicity
Le-Nguyen et al. (2023) [3]	Decision Tree	83	Railway Maint.	Good interpretability but lower accuracy than K-NN
Our work	K-NN	87.6	Tram Maint.	High accuracy and balanced performance in structured datasets

4. Conclusion

The results of the study indicated that the Deep Learning model attained a performance of 70.66%, the Decision Tree model achieved 85.95%, and the K-NN model achieved 87.60%. Among the three models, the K-NN model demonstrated the highest performance in estimating maintenance and repair durations.

According to this model, the system that automatically calculates the operating times between two failures and the operability rate of the trams has been examined and the warranty period has been extended. The fact that the work can be audited has brought seriousness to the work and increased the performance of the personnel. Additionally, the system has supported managerial decision-making through reliable data-driven insights.

In conclusion, the experimental results demonstrate that the K-NN algorithm achieved the highest classification performance, with an accuracy of 87.60% and F1-score of 0.652, outperforming both Decision Tree and Deep Learning models. These findings suggest that traditional non-parametric methods may offer more robust performance in structured maintenance datasets, thus challenging the assumption that complex deep architectures are always superior. The proposed model is intended to provide a practical decision support tool for improving maintenance planning in urban tram systems.

Future work will aim to expand the evaluation scope to include the maintenance performance of all transportation vehicles, analysis of trip completion times, and personnel performance assessment.

Declaration of Ethical Standards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Conflict of Interest

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

Acknowledgements

The authors would like to thank Kocaeli Metropolitan Municipality Transportation Park Joint Stock Company for her contributions.

References

[1] Aydın F. M., 2021. Comparison of traditional and rubber tyred tram (translohr) systems and investigation of applicability in Erzincan, Msc. Thesis, Erzincan Binali Yıldırım University, Institute of Science.

[2] Mahale Y., Kolhar S. and More A. S., 2025. A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions, Discover Applied Sciences,**7**(4), 243.

[3] Le-Nguyen M. H., Turgis F., Fayemi P.-E., Bifet, A., 2023. Real-time learning for real-time data: Online machine learning for predictive maintenance of railway systems, Transportation Research Procedia*,* **72***,* pp.171–178.

[4] Ahac M., Lakusic S., 2017. Track Gauge Degradation Modelling on Small Urban Rail Networks: Zagreb Tram System Case Study, Urban Transport Systems.

[5] Arvidsson N., Browne M., 2013. A review of the success and failure of tram systems to carry urban freight: the implications for a low emission intermodal solution using electric vehicles on trams, European Transport-TransportiEuropei, (54), pp. 1-18.

[6] Liden T., 2015. Railway infrastructure maintenance – a survey of planning problems and conducted research, Transportation Research Procedia, **10**, pp. 574-583.

[7] Ahac M., Lakusic S., 2015. Tram track maintenance planning by gauge degradationmodelling, Transport, **30**, pp. 430-436.

[8] Yousefikia M., Moridpour S., Setunge S., and Mazloumi E., 2014. Modeling degradation of tracks for maintenance planning on a tram line, Journal of Traffic and Logistics Engineering, **2**(2), pp. 86–91.

[9] MajidiParast S., Monemi R. N., and Gelareh S., 2025. A graph convolutional network for optimal intelligent predictive maintenance of railway tracks, Decision Analytics Journal, **14**, 100542.

[10] Bukhsh Z. A., Saeed A., Stipanovic I. and Doree A. G., 2019. Predictive maintenance using tree-based classification techniques: A case of railway switches, Transportation Research Part C: Emerging Technologies, **101**, 35-54.

[11] Nair V., Premalatha M., and Braveen, M., 2024. Enhancing metro rail efficiency: A predictive maintenance approach leveraging machine learning and deep learning technologies.

[12] Kowalski M., Magott J., Nowakowski T., Werbińska-Wojciechowska S., 2014.Exact and approximation methods for dependability assessment of tram systems with time window, European Journal of Operational Research, **235**(3), pp. 671-686.

[13] Carrese S., Ottone G., 2006. A model for the management of a tram fleet, European Journal of Operational Research, **175**(3), pp. 1628-1651.

[14] Gürbüz F., Turna F., 2018. Rule extraction for tram faults via data mining for safe transportation, Transportation research part A: policy and practice, **116**, pp. 568-579.

[15] Gökgöz K. B., 2015. Medical equipment maintenance decision model with data mining, Msc. Thesis, Başkent University, Institute of Social Science.

[16] Kiefer A., Schilde M., Doerner K. F., 2018. Scheduling of maintenance work of a large-scale tramway network, European Journal of Operational Research, **270**(3), pp. 1158-1170.

[17] Turna F., 2011. Rule extraction for tram faults via data mining, Msc. Thesis, Erciyes University, Institute of Science.

[18] YadavA. K., Malik H., Chandel S. S., 2015. Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76 sites in Northwestern India, Renewable and Sustainable Energy Reviews, **52**, pp. 1093-1106.

[19] Sudarsono B. G., Leo M. I., Santoso A., Hendrawan F., 2021. Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner, JBASE-Journal of Business and Audit Information Systems, **4**(1).

[20] Balta S., 1998. Analysis of electric consumption data by data mining methods and determining the right tariff, Msc. Thesis, Sakarya University, Institute of Science.

[21] Yıldırım M. E., 2024. Analysis of maintenance and repair processes of Kocaeli Akçaray transportation system using data mining, Msc. Thesis, Kocaeli University, Institute of Science.

[22] Esteban A., Zafra A., and Ventura, S., 2022. Data mining in predictive maintenance systems: A taxonomy and systematic review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **12**(5), e1471.

[23] Dorfman R., 1979. A formula for the Gini coefficient. The review of economics and ststistics, pp. 146-149,

[24] Zhang S., Li X., Zong M., Zhu X., Cheng D., 2017. Learning k for knn classification, ACM Transactions on Intelligent Systems and Technology, **8**(3), pp. 1-19.

[25] Lei C., 2021. Deep Learning Methods and Applications.In: Deep Learning and Practice with MindSpore, Cognitive Intelligence and Robotics, Springer.

[26] Ahmad J., Muhammad K., Lloret J. and Baik S. W., 2018. Efficient Conversion of Deep Features to Compact Binary Codes Using Fourier Decomposition for Multimedia Big Data, IEEE Transactions on Industrial Informatics, **14**(7), pp. 3205-3215.

[27] Chen, L, 2021.Deep learning and practice with mindspore. Springer Nature.