



Investigating the Accuracy and Validity of Iran's Concrete Bridge Management Method Using Data Mining Approach

Ali Medghalchi¹, Pooria Rashvand², Ali Delnavaz³

1. Department of Civil Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
panam23med@gmail.com

2. Department of Civil Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
poorya_rashvand@qiau.ac.ir

3. Department of Civil Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
A.delnavaz@qiau.ac.ir

Correspondence:

Pooria Rashvand

Department of Civil Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran.

Email: poorya_rashvand@qiau.ac.ir

Abstract: In this research, data mining algorithms were employed to validate Iran's bridge management model. The results indicated that all three data mining methods predicted the worst possible scenario for bridge failure when the daily road traffic volume is less than 8611 vehicles per day. Concrete science experts have identified factors such as high daily traffic rate, high percentage of light heavy vehicles in traffic, heavy snowfall, and number of days with ice on the road per year as the most influential factors in the failure of concrete structures, particularly concrete bridges. Since all three analysis methods yielded almost identical results, it can be concluded that the traditional Iranian model is not valid.

Keywords: bridge management system, data mining, decision tree, validation, Iran, Zanjan..

1. Introduction

Bridge failure has dramatic consequences in every nation's transportation system. In addition to casualties and loss of lives, the disruption in the service results in tremendously adverse effects on economic growth [1], [2], [3]. For example, the failure of the Quebec Bridge in 1907 caused 75 deaths during construction [4], and the failure of the Silver Bridge in 1967 killed 46 people during service [5], [6]. In 2007, Tuojiang Bridge catastrophically collapsed during construction, resulting in 64 deaths and 22 injuries as well as a direct economic loss of about 39.747 million yuan [7]. In 2018, the Italian Morandi's Polcevera Viaduct collapsed during service, resulting in 43 deaths and 9 injuries as well as a direct economic loss of about 100 million yuan [8], [9]. During the period between 1980 and 2012, a total of 1062 bridge failures were reported in the United States, causing huge losses to the nation [10].



The use of a Bridge Management System (BMS) enables administrators to organize hierarchically the assets according to their condition or importance, as well as their needs of maintenance or reparation. The adoption of BMSs provides the asset managers with a sound basis to prioritize investments from both the technical and the economical points of view [11],[12]. The implementation of an inspections-based BMS requires the execution of periodic bridge inspections aimed to identify those bridges which may require maintenance, repair, or refurbishment works. However, the information gathered from inspections constitutes a sound database that can be used to understand how the managed bridges deteriorate under certain environmental, traffic, or durability conditions [13].

There are approximately 350,000 bridges in Iran, with about 7,000 of them located in Zanjan province on suburban roads. Since 2018, the Iran Road Transport Organization has prioritized the implementation of a native bridge management system in line with its goals. In this system, the assessment of bridge health and performance relies heavily on field observations of bridge elements' conditions and the extent of their failures. The structural performance of bridges is ranked using a descriptive method based on coefficients obtained from tables designed according to the condition of bridge elements. A numerical index ranging from 0 to 1000 is used, where a lower index indicates better structural performance. Unfortunately, despite the various features of bridge management systems used globally, including in Iran, there is often confusion among bridge managers when it comes to identifying high-risk situations that bridges may face during their lifespan. The sheer number of bridges, combined with the multitude of physical, climatic, and traffic factors that impact their condition, makes it challenging to establish logical relationships between these factors. Descriptive statistics alone usually fail to provide satisfactory answers in such complex conditions. As a result, governments are constantly seeking validation for the bridge management systems implemented in their respective countries. Given the large number of bridges and the abundance of input and output data, data mining methods with their predictive capabilities and ability to determine relationships between data can be effectively utilized to validate and verify bridge management systems.

For many years, bridge management approaches and condition assessments have been based on long-established manual paperwork and information retained from on-site inspectors and engineers. These approaches have been primarily paper-based and significantly limit the ability to be readily transferred to asset managers or be referred after a few years [14].

In order to address these challenges, this study employs data mining methods to analyze the vast amount of data and prioritize bridges accordingly. Data mining is utilized to extract hidden patterns and insights from the data. By applying data mining techniques, it becomes possible to identify key factors and prioritize bridges based on their risk levels.



Data mining, also known as DM, is a field that involves the extraction of hidden information from large datasets [15]. It is considered both an art and a science [16]). The main objective of data mining is to discover new, unexpected, and valuable patterns in data. This process can be applied to various sources of data, such as text, databases, transaction data, multimedia data, sequences, web data, time series, spatio-temporal data, graphs, social networks, and informational data [17]. The need for data mining arises from the increasing size of databases, which makes it difficult for humans to comprehend and extract meaningful insights from them. Without powerful tools, decision-makers are unable to extract valuable information from the abundance of data available. This situation highlights the challenge of having a wealth of data but a lack of actionable information. In today's highly competitive environment, using data effectively has become a crucial goal for managers to improve the current state of affairs [18].

Decision trees are employed to predict the membership of objects into different categories. The flexibility of this technique has made it a popular choice in data mining. The decision tree methodology consists of two main phases [19]:

A. Creating the primary tree: The decision tree is constructed using the training data set, continuing until each leaf becomes pure (homogeneous).

B. Pruning: The grown tree is pruned in this phase, based on the experimental data set, to enhance the accuracy of the model.

Due to the potential errors in bridge management systems, which can lead to an increased risk of bridge failure and subsequent budget losses, this research aims to investigate and validate the bridge management system in Iran. To achieve this, the decision tree data mining method has been employed on data and datasets. Initially, a database consisting of 384 concrete bridges with a span exceeding 5 meters on the main axis, including physical characteristics and data related to the index of structural failure, was utilized. Additionally, new data on climatic and traffic conditions governing the bridges were incorporated. A data mining model was then developed.

In the subsequent step, the decision tree method in rapid miner software and regression in MINITAB software were utilized to determine the climatic and traffic factors that contributed to the highest number of bridge failures. If Iran's bridge management method proves to be a valid model, it would be expected that bridges with a high failure rate would be located in areas with unfavorable climatic and traffic conditions. Conversely, if discrepancies are found, it may indicate issues within Iran's bridge management system, such as inaccurate determination of inputs, inadequate conversion of the quality status of bridge elements into quantitative figures, or unreliable calculation formulas.

2. Literature Review

Bridge Management System:

Almost exclusively, bridge management programs are based on visual inspections. In special cases, non-destructive evaluation (NDE) tests are performed and although relatively few, there is a growing number of monitoring applications. However, such tests or uses of SHM are generally on an ad-hoc basis to target a specific fault and are not integrated into the overall maintenance and management hierarchy. Despite differences between countries, levels of authority, or ownership classifications, most BMS are organized into modules that perform the functions depicted in Figure (1) which are described in the following sections [20].

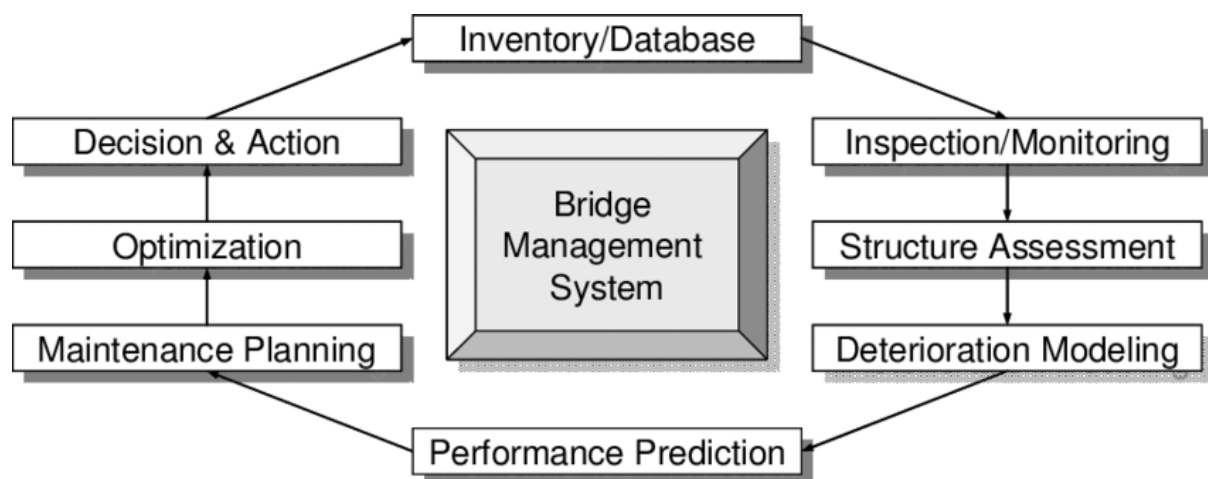


Figure 1 Main modules / functions of a bridge management system.

Failure of Concrete Structures:

Concrete damage can be of the following main types [21]:

1. MECHANICAL

Most mechanical failures occur because of physical impacts such as collisions. Vibrations, overloading and other movements can also result in cracking and loss of entire concrete sections.

2. CHEMICAL

Although an initial chemical reaction is required to create a reliable, load-bearing concrete, other chemical reactions such as ASR and contaminants from soil such as sulphates can weaken the cement matrix.

3. FIRE



The excessive heat generated from fires can alter the properties of concrete. When exposed to extreme temperatures, problems such as expansion and deterioration can occur.

4. STRAY CURRENTS

Stray electrical currents carrying high volts can often act as the catalyst for the corrosion of rebar from the inside. This can lead to spalling, otherwise known as ‘concrete cancer’.

5. CORROSION

When concrete fails to provide adequate protection from extremities such as road salts, seawater and chlorides, the steel rebar will corrode. Cracks, low cover, porosity and low alkalinity are also other causes of inadequate protection.

Failure of Concrete Bridges:

Principal causes can be divided into internal causes and external causes or natural factors and human factors. Internal causes, which include design error, construction mistake, lack of maintenance, material defect, etc., usually require detailed failure investigations, such as material test, an inspection of design, construction, and maintenance documents, structural calculation, etc. [10], [22]. External causes include natural disasters such as earthquake, flood, fire, and wind, as well as extreme loads such as collision and overload. And this kind of cause can be roughly determined through the accident scene investigation [22]. Natural factors are mainly natural disasters, including earthquake, flood, wind, etc., which are not easy to control and prevent. Human factors are mainly associated with human cognitive limitations, carelessness, mismanagement, etc., containing design error, construction mistake, lack of maintenance, material defect, collision, overload, etc. [23], [24], [25]. Design error, construction mistakes, hydraulic, collision, and overload are the top 5 leading causes of bridge failures, resulting in more than 70% of the bridge failures. Causes of bridge failures are closely related to structural type, type of use, material type, and service age [26].

Bridge Management System Method Used in Iran:

The local version of the road bridge management system in the country has been a priority for the Iran Road Maintenance and Transportation Organization since 2018. The implementation of this system involves two stages. In the first stage, the bridge information form is completed and uploaded into a web-based software that has been developed for this purpose. This form includes essential details about the bridges in the system. In the second stage, the system allows for the registration of bridge damages and related information. Users can upload images and additional details about the damages into the system.

The system then utilizes a scoring method (unweighted) to determine the priority of bridges. This scoring is based on factors such as the overall damage index of the bridge structures, the damage index during bridge operation, and the general condition ranking of the bridge. By



implementing this local version of the road bridge management system, the Iran Road Maintenance and Transportation Organization aims to improve the efficiency and effectiveness of bridge maintenance and prioritize necessary repairs based on the condition and impact of damages. The models utilized in this version are as follows (Figure 2):

- ID card model
- Administrative data: organization (3 categories)
- Technical data: geometric data, global coordinates, spans, design (14 categories)
- Main inspection module: conducting the main inspection, conducting the overview, recording the damages, the responsible checklist, evaluating the members, and preparing the report (Scoring the damage index of the entire bridge structures, damage index of the bridge operation, and ranking the overall condition of the bridge)

In this model, the structural indicators of bridges have been determined based on the information collected through the objective observation method. To calculate these indicators, a list of bridge damages has been compiled and coefficients have been assigned to each damage type. The structural index is obtained by summing the values of all the bridge members and multiplying them by their respective coefficients, which take into account factors such as the type and amount of damage, length and height of the bridge, and the complexity of the cross-sectional design. Additionally, an operation index is calculated for damages related to drainage, surface, protection, fence, sidewalk, and cornice. The structural index is calculated separately for other bridge members.

	Day	Completi on data	Bridge Inspection and Technical Identification Form			
	Mon th					
	Year					
Photo		Bridge profile information				
Aerial Photo (Satellite)	Long View of the armpit		Route code		Province	
			Bridge code		Country Area	
			Year of constructi on		(source- destination) Axis	
			GPS-X		Origin of axis	
			GPS-Y		Kilometers	



Received: 16-06-2024

Revised: 12-07-2024

Accepted: 07-08-2024

					Structural Bridge system		Bridge name		
					Type of cross-over complication		Route direction		
Crossover View (Base View and Deck View)	View of the cavern from above			Geometric Bridge information					
					Type of bridge material		Number of opening		
					Free height (m)		Length of largest Span (m)		
					Axle type		Total width of bridge (m)		
					Alternate route (bypass)		Total bridge length (m)		
							Number of traffic lanes		
Corrupted member information									
explanations	Photo number	Damage index	The severity of the injury	Type of injury	Type of materials	Member name	Damage code	Member code	Row
									1
									2
						Structural Damage Index the			



			Ranking Overall Bridge Status		Damage index of operation of the whole bridge

Figure 2 Data entry form

Matlab software:

Matlab is a commercial software developed by MathWorks that can be used for algorithm generation, data mining, and data analysis. It allows the user to call other programs written in different programming languages. Matlab has also developed a toolbox that incorporates various techniques like general algorithms, neural networks, and more. This toolbox enables the user to perform functions such as optimal analysis, statistics, signal processing, image processing, vector analysis, and matrix calculations. It is important to note that raw data preparation is crucial for Matlab as it can impact the calculation efficiency [27]. One of the key components of Matlab is the "Statistics and Machine Learning Toolbox™" [28]. In this section, decision trees, regression trees, and classification trees are used to predict data outcomes. In this article, we employ Matlab regression tree for data mining bridge information.

RapidMiner Studio Software:

RapidMiner uses a client/server model with the server offered either on-premises or in public or private cloud infrastructures. RapidMiner provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. RapidMiner is written in the Java programming language. RapidMiner provides a GUI to design and execute analytical workflows. Those workflows are called "Processes" in RapidMiner and they consist of multiple "Operators". Each operator performs a single task within the process, and the output of each operator forms the input of the next one. Alternatively, the engine can be called from other programs or used as an API. Individual functions can be called from the command line. RapidMiner provides learning schemes, models and algorithms and can be extended using R and Python scripts [29].

RapidMiner can also use plugins available through the RapidMiner Marketplace. The RapidMiner Marketplace is a platform for developers to create data analysis algorithms and publish them to the community. The RapidMiner Studio Free Edition, which is limited to one logical processor and 10,000 data rows, is available under the AGPL license [30].



The use of this software is increasing in various scientific fields due to its possibilities and features. This article focuses on the application of associative rules and decision tree sections of the software in data mining. The following sections will provide an explanation of each of these features.

Data Mining (DM) & Decision Tree (CART):

Data mining is an iterative process rather than a linear one. Its main objectives are to describe, predict, and prescribe [31]. The concept of classification and regression trees was introduced by Braiman. This method aims to divide data into smaller sections to maximize the separation of information [32]. The decision tree algorithm is a predictive modeling technique that recursively divides input data and can handle predictor variables as well as numerical and categorical targets [33]. In essence, a decision tree is a symbolic learning method that organizes information from a training data set into a hierarchical structure with nodes and branches. All decision tree algorithms define two types of variables: independent variables and dependent variables. The dependent variable, or target variable, takes discrete values in a classification tree and continuous values in a regression tree. To create a decision tree, the data set is divided into training data and test data. The decision tree is built based on the training data, and then the trees are pruned to prevent overfitting. Two examples of common decision tree algorithms are CART C and 5.0 [34].

To use the CART algorithm, a target attribute in the dataset must be specified. The algorithm starts from the root node and divides the training data into two child nodes, which are further divided into grandchild nodes. During tree growth, the CART algorithm searches for questions that separate each parent node into two homogeneous child nodes. Homogeneity means that all records in a node belong to a specific category, in which case the node becomes a leaf. The algorithm looks for features in the dataset that have strong separating properties. As the tree grows, the nodes become more homogeneous and display more information. Tree growth continues until the tree reaches its maximum size and the division operation stops due to lack of data [35].

In the decision tree of the rapidminer software, a new node is formed as the root of the original dataset. Then the gain ratio of each attribute is calculated. The process of each attribute is as follows [36]:

Step 1: First, the entropy of the training data (S) is determined according to equation(1)

$$(1) \text{ Entropy (S)} = \sum_{i=1}^n - p_i \log_2 p_i$$



Where, n refers to the number of different class levels and p_i refers to the proportion of the number of class C_i ($i = 1, 2, n$) data samples to the number of all samples in the training set. The entropy of a sample of data shows how combined the class values are; the minimal value of 0 indicates that the sample is absolutely homogenous, while 1 shows the maximum amount of disorder [37]

Step 2: The conditional entropy of S given F is defined as:

$$(2) \quad \text{Entropy}(S|T) = - \sum_{j=1}^m p'_i \sum_{i=1}^n - p_{ij} \log_2(p_{ij})$$

where, T is the attribute with m different values and p'_i is the proportion of the number of samples whose values of the related feature are t_j to the number of all samples in the training set.

Step 3: Information Gain of attribute T is defined as

$$(3) \quad \text{Gain}(T) = \text{Entropy}(S) - \text{Entropy}(S|T)$$

Where

$$\text{Splite}(T) = \sum_{j=1}^m -p'_j \log_2 p'_j$$

Information gain Ratio can be obtained further as follows:

$$(4) \quad \text{Gain Ratio}(T) = \text{Gain}(T) / \text{Splite}(T)$$

According to the mentioned process, Information Gain Ratio is calculated for decision attributes and the attribute with the largest information gain ratio is placed as a tree node at each depth [38].

In this research, for the decision tree model, the values in Table (1) are defined as limiting the depth and accuracy of the tree:

Table 1- Values limiting the depth and precision of the definition tree

Parameter	Value
Minimal depth	20
Confidence	0.25
Minimal gain	0.01



Minimal leaf size	2
Minimal size for split	4
Number of prepruning alternatives	3

To evaluate the models, the following criteria have been checked:

- Accuracy criterion
- Precision criterion
- Recall criterion

Naive Bayes:

Naive Bayes is a classifier that has high bias and low variance. It applies Bayes' theorem with independent (naive) assumptions. However, it is important to note that this assumption is rarely true in real-world scenarios. Despite this, the Naive Bayes classifier often performs well in practice.

One of the advantages of Naive Bayes classification is that it requires only a small amount of training data to estimate the mean and variance of the variables necessary for classification. Even with a small dataset, it can still build a good model. Additionally, it is simple to use and computationally inexpensive.

The independence assumption greatly simplifies the calculations required to build the Naive Bayes probability model. However, to complete the probability model, assumptions need to be made about the conditional probability distributions for each feature, based on the class. In this case, Gaussian probability densities are used to model attribute data [36].

Rule Induction:

The Rule Induction operator follows a process where it starts with less prevalent classes and iteratively grows and cuts rules until either there are no positive samples remaining or the error rate exceeds 50%. During the growth phase, conditions are greedily added to each rule until it becomes complete. This involves trying every possible value of each feature and selecting the condition with the highest information gain.

In the prune step, each rule's final sequence of predecessors is pruned using the pruning metric $p/(p+n)$. Rule Set learners are often compared to Decision Tree learners. Rule Sets have the advantage of being easy to understand. However, they are also sensitive to noisy data, which can significantly impact their performance [36].



3. Database

Zanjan is a city in northwest Iran. The statistical population of this research includes all concrete road bridges of 5 main axes of Zanjan province with a length of above 5 meters, including 384 bridges.

Data collection method:

In the present study, the databases of Zanjan Road Maintenance and Transportation Organization and the statistical yearbook of the Meteorological Organization for 2019 have been used to collect information according to the type, purpose and subject of study and the characteristics of the statistical population.

Model inputs and outputs:

As mentioned in the previous sections, the measurement of bridge failure is based on the Iranian model used, structural failure and operation indicators, which are obtained based on the observations of inspectors and calculations, and in fact are the data of the model variables. In selecting model inputs, the findings of other researchers such as Lee, Wardhana, Fu, Liu, Zhou and zhang in the field of factors affecting the failure of concrete bridges have been used. Naturally, among the wide range of factors, factors have been used that were possible to prepare and validate data in the statistical population.

Accordingly, three data sets can be used to determine the efficiency of bridges based on the output of structural failure and operation indicators. These three groups include climatic factors such as the number and the amount of rainy and snowy days, number of frosty days, average difference per day, intensity, acidity of rain and snow, salt concentration used in snow sweeping and de-icing operations, a group of traffic factors such as the amount of traffic, the amount of heavy vehicle traffic, the amount of tonnage crossing the bridge, the speed of crossing the bridge, the number of cars colliding with the bridge deck and bases, and maintenance factors such as periods and types of cross-sectional and major repairs, maintenance cost and physical factors such as bridge length, largest span length, bridge width, and bridge height. Unfortunately, there was no information about accidents that occurred on bridges and maintenance records and related costs in the provincial road maintenance and transportation organizations, so the data available in the provincial road accident registration system (Provincial Roads Management Center), intelligent road traffic system and meteorological systems in the area of bridges and important physical characteristics effective in structural failure were extracted and were used as model data after validation and correction. In the following, we will examine the inputs and outputs used in the model.



Meteorological and climatic information:

Input 1. Number of axial snowfalls (the number of times that, according to the Road Management Center reports, there has been snowfall during the year on the mentioned axis section).

Input 2. Number of axial rains (the number of times that, according to the Road Management Center reports, there has been rain during the year on the mentioned axis section).

Input 3. Amount of rainfall (ml) (total rain and snow falls according to the reports of the Road Management Center, during the year on the mentioned axis section)

Input 4. Number of frosty days (number of frosty days in the bridge area according to the reports of meteorological stations near the bridge)

Input 5. Number of days of snowfall (number of days of snowfall in the bridge area according to reports of meteorological stations near the bridge)

Input 6. Average snow/amount of rainfall (ml) (average snow/rainfall in the bridge area according to the reports of meteorological stations near the bridge)

Input 7. Average temperature difference (average temperature difference between day and night in the bridge area according to the reports of meteorological stations near the bridge)

Traffic Information:

Input 8. Average daily traffic (ADT) (average daily round-trip traffic based on the information of traffic counts installed on the axis)

Input 9. Percentage of heavy vehicle (percentage of heavy vehicle traffic including trucks, trailers and buses in relation to the total traffic of vehicles on the axis based on the information of traffic counts installed on the axis)

Input 10. Average speed (km/h) (average speed of vehicles traveling on the axis based on the information of traffic counts installed on the axis)

Model Outputs:

Bridge structural failure index (based on the failure status of each of the bridge elements and the application of member impact coefficients and the degree of failure)

Information Analysis Method:

Since in this study we seek to Investigating the accuracy and validity of Iran's concrete bridge management method, so we used the decision tree method. Using this method, the physical characteristics and atmospheric, climatic and traffic conditions affecting the studied bridges are selected as independent variables and structural failure indicators based on the Iranian BMS



model as dependent variables, then data preparation operation was performed to enter the model. Preparation is based on the features and requirements of RapidMiner Studio 9.0 modeling software and the model used. Obviously, failure to correct the data can lead to errors in the output or wrong prediction of values. Table (2) shows the variables.

Table 2. Model Variables

Numb er of axial snowf alls	Num ber of axial rains	Amo unt of rainf all (ml)	Num ber of frost y days	Num ber of snow fall days	Aver age rainf alls	Average tempera ture differen ce	Aver age daily traffi c (ADT)	Heavy vehicle percent age	Brid ge leng th	Struct ural failure index
X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Index

In this article, information related to 384 bridges is used, which is shown in Table (3).

Table 3- Specifications of model input information

	Cod e	Typ e	Min	Max	Range	Avera ge	Standar d Deviati on	Missin g Value
Number of axial snowfalls	X1	Real	2.00	19.00	17.00	9.27	5.10	0
Number of axial rains	X2	Real	2.00	25.00	23.00	11.41	6.76	0
Amount of rainfall (ml)	X3	Real	277.00	392.00	115.00	353.03	45.96	0
Number of frosty days	X4	Real	80.00	116.00	36.00	103.52	14.27	0
Number of snowfall days	X5	Real	26.00	43.00	17.00	34.52	7.21	0
Average rainfalls	X6	Real	0.76	1.07	0.31	0.97	0.13	0



Average temperature difference		X7	Real	12.40	14.80	2.40	13.71	0.99	0
Average traffic (ADT)	daily	X8	Real	1729.00	15736.00	14007.00	8852.22	5847.55	0
Heavy vehicle percentage	vehicle	X9	Real	9.50	39.00	29.50	26.69	7.08	0
Bridge length		X10	Real	5.00	136.00	131.00	12.66	14.68	0
Structural failure index	failure	Index	Real	0.00	279.94	279.94	23.30	30.09	0

In this research, all methods utilized involved using 70% of the data for training purposes. In order to classify the data, it was necessary to convert the predicted values into qualitative mode. The parameter chosen as a predictive variable was the structural failure index, which was divided into four categories based on BMS values.

In Range 1, the structural damage index ranged from 0 to 30, indicating a favorable condition of the bridge with no need for special action. Range 2 encompassed values between 30 and 200, indicating good performance of the structure but requiring repair and improvement operations. Range 3 included values between 200 and 370, signifying a critical condition that necessitates immediate repair and improvement. Finally, range 4 indicated a very critical condition where the structure's safety is compromised, requiring urgent improvement or even reconstruction. This classification is illustrated in Table (4).

Table 4- Classification of structural damage index

Class	Range1	Range2	Range3	Range4
Amplitude of Structural failure index	[0,30)	[30,200)	[200,370)	[370,1000)
Number of members	290	93	1	0

Upon examining the structural damage index values obtained from the Iranian bridge data bank, it was observed that none of the bridges fell into Range 4, while one bridge was classified under Range 3. Additionally, 93 bridges were categorized under Range 2, and 290 bridges fell into Range 1.

The overall dispersion of the studied bridges, based on the structural index, can be visualized in Figure 3. Furthermore, Figure (4) depicts the location of the selected bridges in Zanjan province, Iran, along with the location of four out of the five selected axes within the west-to-



east corridor of Iran. This corridor is a crucial transit and commercial route with high traffic volume, particularly heavy fleet traffic. Given that the northwest region of Iran experiences a mountainous and cold climate, the findings of this research can be applied to provinces situated in the northwest of Iran by adhering to modeling principles.

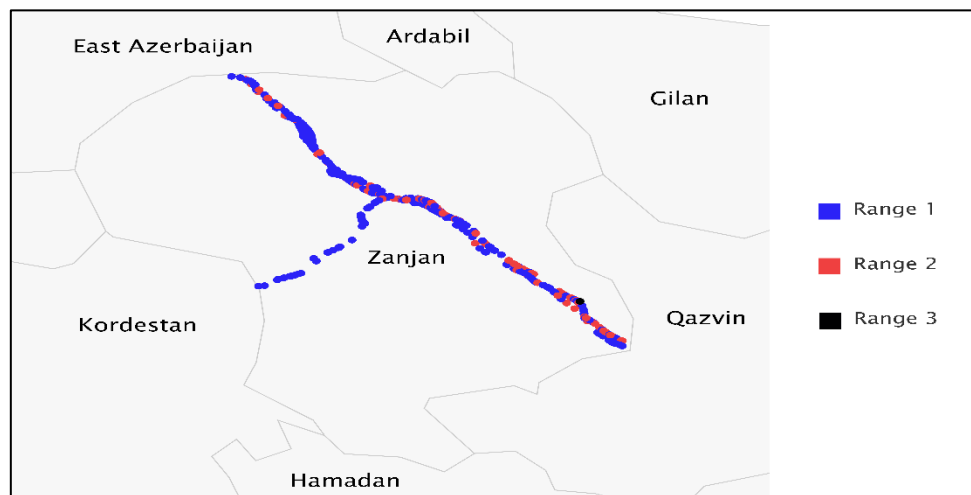


Figure 3- State of bridges in the province based on the span

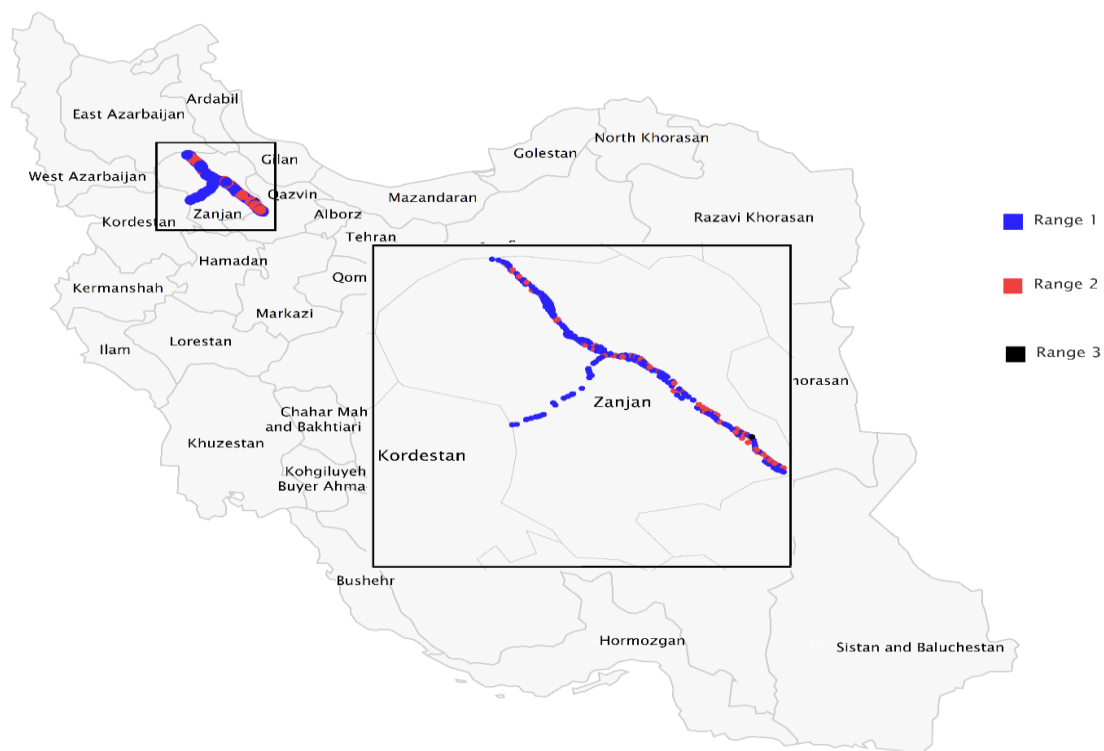


Figure 4- Status of bridges in the country based on span



4. Discussion and results

In this article, we analyze the model and aim to increase the credibility of the research by using three models: decision tree, rule induction, and Naive Bayes. We then compare the results obtained from each model.

Analysis with decision tree model method of Rapid miner software:

The classification model, which utilizes decision tree, association rules, etc., in Rapid miner software, is illustrated in Figure (5). In this model, 70% of the data is used for training, while the remaining 30% is used for model validation. As there is no missing data in the entire statistical dataset of this article, no data replacement operations were performed. However, since there is only one data point in Range 3, none of the models were able to train it.

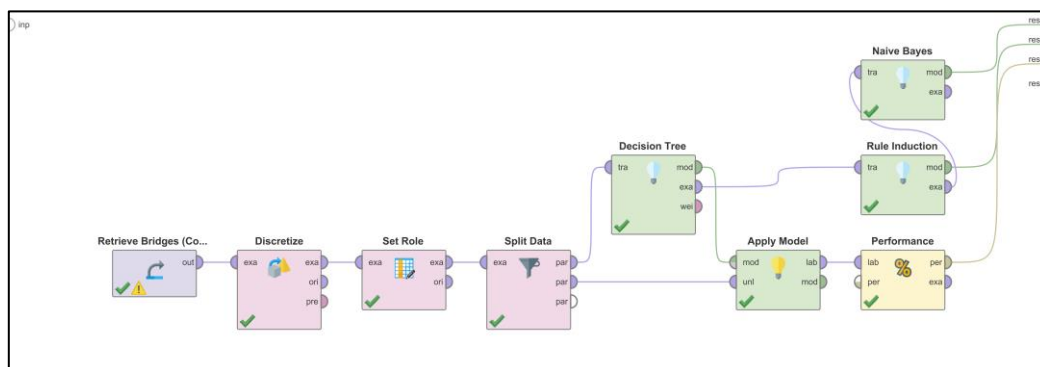


Figure 5 - Data mining model of Zanzibar bridges management system

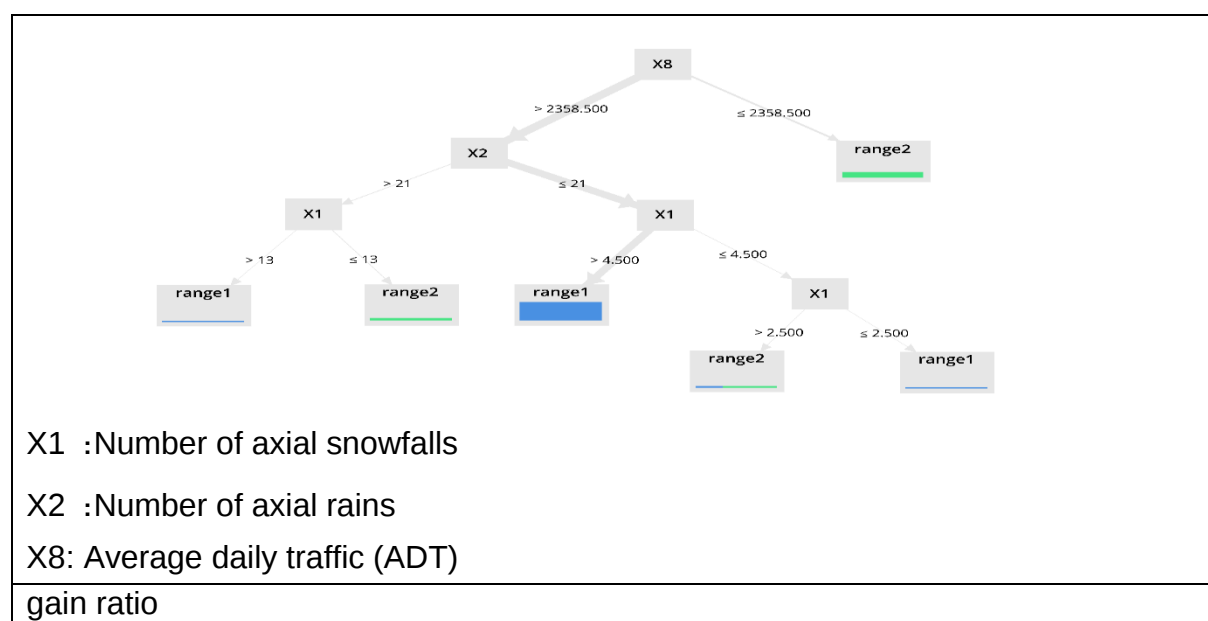


Figure 6 - The decision tree created based on the research model



The decision tree, represented in Figure (6), identifies X1, X2, and X8 as significant attributes for estimating the structural failure index. Notably, the worst failure index (Range 2) occurs when X8, representing the average daily traffic, is less than 2358.5 vehicles per day.

The descriptive result of the decision tree is as follows: (Figure 7)

X8 > 2358.500
X2 > 21
X1 > 13: range1 {range1=3, range2=0, range3=0, range4=0}
X1 ≤ 13: range2 {range1=0, range2=16, range3=0, range4=0}
X2 ≤ 21
X1 > 4.500: range1 {range1=188, range2=0, range3=0, range4=0}
X1 ≤ 4.500
X1 > 2.500: range2 {range1=2, range2=4, range3=0, range4=0}
X1 ≤ 2.500: range1 {range1=8, range2=0, range3=0, range4=0}
X8 ≤ 2358.500: range2 {range1=0, range2=48, range3=0, range4=0}

Figure 7 - Descriptive result of the decision tree

After modeling, it is essential to evaluate the results to improve the model's performance and usability. The accuracy and precision of a classification method on the training dataset are calculated as the percentage of correctly classified observations from the training set. The evaluation is based on the test data, also known as the error rate. The misclassification rate can also be derived from the accuracy index [39].

The performance of this model in predicting the failure index is presented in Table (5).

	true range1	true range2	true range3	true range4	class precision
pred. range1	89	0	0	0	100.00%
pred. range2	0	25	1	0	96.15%
pred. range3	0	0	0	0	0.00%



pred. range4	0	0	0	0	0.00%
class recall	100.00%	100.00%	0.00%	0.00%	
gain ratio					

Table 5- Prediction of failure index

In this article, the correctness index is used to evaluate the performance of the algorithms employed. As mentioned earlier, 70% of the statistical population is used for training, while 30% is used for validation. According to the model results (Table 5), the error rate is 3.75%, and the accuracy is 96.15%, indicating an acceptable validity of the model.

Next, we examine the data using the Rule Induction method. The descriptive result of the Rule Induction model is shown in Figure (7).

According to this model, the variables:

Average daily traffic (X8), Number of snowfall days (X5), Heavy vehicle percentage (X9), Amount of rainfall (X3) and Number of axial snowfalls (X1)

The variables have shown the greatest impact on predicting the structural failure index. Due to the lack of training data for Range 3, this model couldn't provide a correct prediction for this range. However, it performed well in predicting Range 1 and Range 2. Similar to the previous model, 70% of the statistical population is used for training, and 30% is used for validation. The performance result of the model for predicting the test data is presented in Table (6), indicating an accuracy of 98.89%.

if X8 > 14340.500 then range1 (79 / 0 / 0 / 0)
if X5 ≤ 27 then range1 (40 / 0 / 0 / 0)
if X3 > 388.500 then range1 (32 / 0 / 0 / 0)
if X8 ≤ 2358.500 then range2 (0 / 48 / 0 / 0)
if X9 > 22.250 then range1 (36 / 0 / 0 / 0)
if X1 > 8 then range2 (0 / 16 / 0 / 0)
if X1 ≤ 2.500 then range1 (8 / 0 / 0 / 0)
else range1 (5 / 3 / 0 / 0)
correct: 264 out of 267 training examples.
Rule Induction Model

Figure 7- Descriptive result of Rule Induction model



	true range1	true range2	true range3	true range4	class precision
pred. range1	89	1	0	0	98.89%
pred. range2	0	24	1	0	96.00%
pred. range3	0	0	0	0	0.00%
pred. range4	0	0	0	0	0.00%
class recall	100.00%	96.00%	0.00%	0.00%	

Table 6- The result of the Rule Induction model performance

We further analyze the data using the Naive Bayes model classification method (Figure 9). The performance result of this method can be seen in Table (7), showing an accuracy of 89.89%, which is considered acceptable.

Class range1 (0.747)
10 distributions
Class range2 (0.253)
10 distributions
Class range3 (0.000)
10 distributions
Class range4 (0.000)
10 distributions
Distribution model for label attribute Index based on Naive Bayes classifier

Figure 9- Descriptive result of Naive Bayes model classification method

	true range1	true range2	true range3	true range4	class precision
pred. range1	80	9	0	0	89.89%
pred. range2	9	16	1	0	61.54%
pred. range3	0	0	0	0	0.00%
pred. range4	0	0	0	0	0.00%
class recall	89.89%	64.00%	0.00%	0.00%	

Table 7- Performance result of Naive Bayes model



Received: 16-06-2024

Revised: 12-07-2024

Accepted: 07-08-2024

To summarize the different forecasting methods used in Rapidminer software, Table (8) presents their accuracy based on the criteria of accuracy, class precision, and class recall. According to the accuracy criterion, the decision tree achieves the highest accuracy of 99.13% in predicting the structural failure index.

criterion	Decision tree	Rule Induction	Naive Bayes
	(gain ratio)		
accuracy	99.13%	98.26%	83.48%
class precision	96.15%	96.00%	61.54%
class recall	100.00%	96.00%	64.00%

Table8-comparison of the results of three methods: Naive Bayes, Rule Induction, gain ratio

Analysis with regression decision tree model of MATLAB software:

The MATLAB software regression tree was utilized to assess the percentage effect of variables on the prediction of structural damage index. To do this, the data was prepared and the input and target data were defined in the software. The Randperm command was used to randomly select 70% of the data as training data and 30% as test data. The fit tree command was then employed to build the regression decision tree model, which is presented in Figure (10).

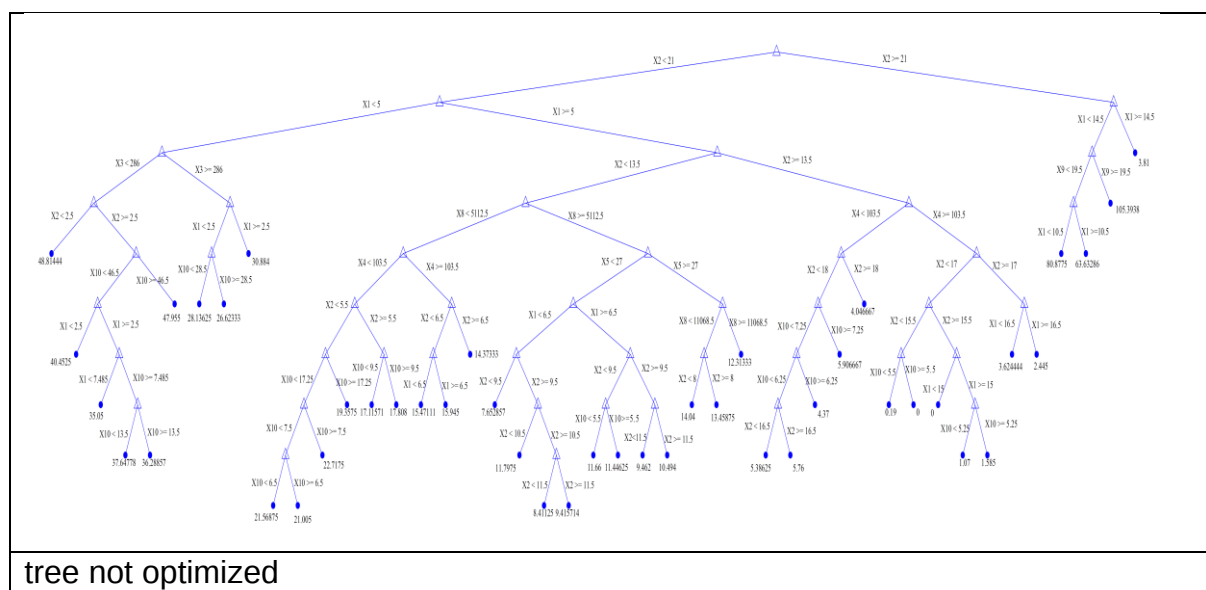


Figure 10- Regression decision tree

<pre> graph TD Root(()) -- "x2 < 21" --> Node1(()) Root -- "x2 >= 21" --> Node2(()) Node1 -- "x1 < 5" --> Node3(()) Node1 -- "x1 >= 5" --> Node4(()) Node2 -- "x8 < 8611" --> Leaf1((94.9492)) Node2 -- "x8 >= 8611" --> Leaf2((36.0727)) Node3 -- "x9 < 24.25" --> Leaf3((28.7112)) Node3 -- "x9 >= 24.25" --> Node5(()) Node4 -- "x2 < 13.5" --> Node6(()) Node4 -- "x2 >= 13.5" --> Node7(()) Node5 -- "x1 < 2.5" --> Leaf4((47.6336)) Node5 -- "x1 >= 2.5" --> Leaf5((37.873)) Node6 -- "x8 < 5112.5" --> Node8(()) Node6 -- "x8 >= 5112.5" --> Node9(()) Node7 -- "x4 < 103.5" --> Leaf6((5.41833)) Node7 -- "x4 >= 103.5" --> Leaf7((1.33906)) Node8 -- "x3 < 388.5" --> Leaf8((15.5368)) Node8 -- "x3 >= 388.5" --> Leaf9((19.8037)) Node9 -- "x5 < 27" --> Leaf10((9.75283)) Node9 -- "x5 >= 27" --> Leaf11((13.4256)) </pre>	X1: Number of axial snowfalls X2: Number of axial rains X3: Amount of rainfall X4: Number of frosty days X5: Number of snowfall days X6: Average rainfalls X7: Average temperature difference X8: Average daily traffic X9: Heavy vehicle percentage X10: Bridge length
tree optimized	

The mean square error for the training data (Resubstituting Error) was found to be 11.25 for the non-optimized decision tree and 0.95 for the optimized decision tree (Table 9). Additionally, the mean squared error per new data (mean squared errors) in the optimized tree model was 121.49, while it was 26.41 for the optimized tree model, indicating the improved accuracy of the optimized model.



	Regression Tree	Optimized Regression Tree
Resubstitution Error	11.2561	0.9572
mean squared errors	121.4969	26.4121

Table 9- The performance result of the decision tree regression model

5. Conclusion and Recommendations:

A bridge management system aims to assess the condition of bridges by examining component failures and quantifying qualitative data obtained from objective observations. Based on the degree of failure, bridges are prioritized, and a maintenance plan is formulated accordingly. However, like any human-made system, errors and mistakes are inevitable, and system validation is essential to instill confidence in transport infrastructure maintenance managers. In this study, we sought to validate the Iranian model of concrete bridge management using data analysis and data mining techniques. Factors such as traffic frequency (modeled by traffic volume) and load during the service period (modeled by the percentage of heavy vehicles) have been proven to significantly impact the service life of concrete bridges, along with atmospheric and environmental factors such as snowfall, rain, frost, and temperature variation. Therefore, it is expected that bridges identified as having the most damage based on the Iranian model are indeed subjected to conditions that potentially lead to damage in concrete structures.

Based on the findings of this study and the decision tree analysis using gain ratio and regression methods on the Iranian concrete bridge management model, incorporating traffic and climate factors, it was observed that with high accuracy, the highest index of structural failure occurs on roads with annual traffic volumes below 2358 (using the gain ratio method) and 8911 vehicles per day (using the regression method). This discrepancy raises doubts either about the findings of concrete technology researchers or the validity of the Iranian model.

Various reasons can account for this contradiction, including expert errors in assessing bridge damage and its extent, discrepancies in matching observations with quantification tables, inaccuracies in quantification table values, formulation errors in damage calculation, and finally, errors in the approach to prioritizing bridges. While the Iranian model has provided a form of bridge prioritization, the bridges identified as being in poor condition according to this method are not necessarily subjected to unfavorable traffic and climatic conditions. More specifically, the predictions resulting from data mining on traffic and climatic factors influencing bridges do not align with the results of the Iranian model. Consequently, if budget allocation is based on bridges in a deteriorated state, there is no guarantee that this program will effectively improve the condition of the bridge network in a region, considering the specific traffic and climate factors of that area.



Recommendations:

Based on the intriguing results of this research, future studies are recommended to develop a model that predicts the impact of climatic, meteorological, and traffic factors on the failure of different bridge elements. This would help identify the most vulnerable bridge elements to these factors.

References:

- [1] Cook, W.& Barr, P.J.& Halling, M.W. (2015) Bridge failure rate. *Journal of Performance of Constructed Facilities*, 29 (3), Article 04014080.
- [2] Diaz, E.E.M.& Moreno, F.N.& Mohammadi, J. (2009). Investigation of common causes of bridge collapse in Colombia. *Practice Periodical on Structural Design and Construction*, 14 (4), pp.194-200.
- [3] Smith, D.W. (1976) Bridge failures. *Proceedings of the Institution of Civil Engineers*, 60 (3), pp.367-382.
- [4] Pearson, C.& Delatte, N. (2006) Collapse of the Quebec bridge, 1907. *Journal of Performance of Constructed Facilities*, 20 (1), pp.84-91.
- [5] Harik, I.E.& Shaaban, A.M.& Gesund, H. & et al. (1990) United States bridge failures, 1951–1988. *Journal of Performance of Constructed Facilities*, 4 (4), pp.272-277
- [6] Lichtenstein, A.G. (1993) The silver bridge collapse recounted. *Journal of Performance of Constructed Facilities*, 7 (4), pp.249-261.
- [7] Peng, W.& Shen, J.& Tang, X. & et al. (2019) Review, analysis, and insights on recent typical bridge failures. *China Journal of Highway and Transport*, 32 (12), pp.132-144.
- [8] Morgese, M.& Ansari, F.& Domaneschi, M.& et al. (2020) Post-collapse analysis of Morandi's Polcevera viaduct in Genoa Italy. *Journal of Civil Structural Health Monitoring*, 10 (1), pp.69-85.
- [9] Nuti, C.& Briseghella, B.& Chen, A.& et al. (2020) Relevant outcomes from the history of Polcevera viaduct in Genova, from design to nowadays failure. *Journal of Civil Structural Health Monitoring*, 10 (1), pp.87-107.
- [10] Lee, G.C.& Mohan, S.B.& Huang, C.& et al. (2013) A Study of U.S. Bridge Failures (1980-2012). State University of New York at Buffalo, New York, USA
- [11] Calvert, G.& Neves, L.& Andrews, J.& Hamer, M. (2020) Multi-defect modelling of bridge deterioration using truncated inspection records. *Reliability Eng. Stability Safety*, 200, p.106962, 10.1016/j.ress.2020.106962.
- [12] Alonso, Medina P.& León, González F.J.& Todisco, L. (2022) Data-driven prediction of long-term deterioration of RC bridges. *Constr Build Mater*, 317, p.125790, 10.1016/j.conbuildmat.2021.125790.



- [13] Frangopol, D.M.& Dong, Y.& Sabatino, S. (2017) Bridge life-cycle performance and cost: analysis, prediction, optimisation and decision-making. Struct Infrastruct Eng, 13 (10), pp.1239-1257, 10.1080/15732479.2016.1267772.
- [14] Rashidi, M.& Hoshyar, A.N.& Smith, L.& Bijan, S.& Siddique, R. (2020) A comprehensive taxonomy for structure and material deficiencies, preventions and remedies of timber bridges. J. Build. Eng. 34, 101624. <https://doi.org/10.1016/j.jobbe.2020.101624>.
- [15] Shoba, G.& Shobha, G. (2014) Water Quality prediction using data mining techniques: A survey. International Journal of Engineering and Computer Science.
- [16] Rajakumari, S.B.& Nalini, C. (2016) Identification of lead contaminant in river water quality data. Journal of Chemical and Pharmaceutical Sciences.
- [17] Krishnamoorthy, M.& Karthikeyan, R. (2022) Pattern mining algorithms for data streams using item set. Measurement: Sensors, Volume 24, December, 100421.
- [18] Shahrabi, J. (2013) Data Mining. Tehran: Jahad-e Daneshgahi Publication, (in Persian)
- [19] Cichosz, P. (2014) Data Mining Algorithms: Explained Using R. New York: John Wiley & Sons, USA.
- [20] Messerve, T.B. (2009) Integration of Structural Health Monitoring into the Design, Assessment, and Management of Civil Infrastructure. Dissertation, University Pavia Italy, DOI:10.13140/2.1.3919.2968. <https://www.researchgate.net/publication/265466992>.
- [21] Delatte, N. (2009) Failure, distress and repair of concrete Structures. Woodhead Publishing Limited and CRC Press LLC, Canada.
- [22] Wardhana, K.& Hadipriono, F.C. (2003) Analysis of recent bridge failures in the United States. Journal of Performance of Constructed Facilities, 17 (3), pp.144-150.
- [23] Fu, Z.& Ji, B.& Cheng, M.& et al. (2012) Statistical analysis of the causes of bridge collapse in China. Sixth Congress on Forensic Engineering: Gateway to a Safer Tomorrow, San Francisco, USA, DOI:10.1061/9780784412640.009.
- [24] Liu, M. (2013) Analysis of Bridge Accident. Dissertation, Southwest Jiaotong University, Chengdu.
- [25] Zhou, J.& Zhen, D. (2017) Safety of highway bridges in China. Strategic Study of CAE, 19 (6), pp.27-37.
- [26] Zhang, G.& Liu, Y.& Liu, J.& Lan, S. (2022). Causes and statistical characteristics of bridge failures: A review. Journal of Traffic and Transportation Engineering (English Edition), Volume 9, Issue 3, June, pp.388-406.
- [27] Yau, N.J.& Chuang, Y.H. (2015) Analyzing Taiwan Bridge Management System for decision making in bridge maintenance: A big data approach. In 2015 10th International Joint Conference on Software Technologies (ICSOFT), Vol. 1, pp. 1-6, IEEE.
- [28] Yau, N.J.& Liao, H.K. (2016) Establishing a Decision Support Module for bridge maintenance in Taiwan. In Advanced Materials and Structural Engineering: Proceedings



- of the International Conference on Advanced Materials and Engineering Structural Technology (ICAMEST 2015), April 25-26, Qingdao, China, pp.371, CRC Press.
- [29] Norris, D. (2013) RapidMiner - a potential game changer. Bloor Research, UK.
- [30] Ohri, A. (2011) Interview with Rapid-I Ingo Mierswa and Simon Fischer. KDnuggets, UK.
- [31] Dominguez, X.& Prado, A.& Arboleya, P.& Terzija, V. (2023) Evolution of knowledge mining from data in power systems: The Big Data Analytics breakthrough. Electric Power Systems Research, Volume 218, May, 109193.
- [32] Gordon, L. (2013) Using classification and regression trees (CART) in SAS® enterprise miner TM for applications in public health. SAS Global Forum 2013, <https://support.sas.com/resources/papers/proceedings13/089-2013.pdf>.
- [33] Loh, W.Y. (2011) Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(1), pp.14 – 23, DOI:10.1002/widm.8.
- [34] Nie, G.& Rowe, W, & Zhang, L.& Tian, Y.& Shi, Y. (2011) Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications, Volume 38, Issue 12.
- [35] Sharma, H.& Kumar, S. (2016) A survey on decision tree algorithms of classification in data mining, International Journal of Science and Research.
- [36] Pang, S.L.& Gong, J.Z. (2009) C5.0 Classification Algorithm and Its Application on Individual Credit Score for Banks. Systems Engineering-Theory & Practice, 29, pp. 94-104.
- [37] Lantz, B. (2019) Machine learning with R: expert techniques for predictive modeling. Packt Publishing, UK.
- [38] Matae, B.& Moghadas nejad, F.& Zakeri, h. (2021) Pavement maintenance and rehabilitation optimization based on cloud decision tree. International Journal of Pavement Research and Technology volume 14, pp.740–750.
- [39] Han, J.& Kamber, M.& Pei, J. (2006) chapter 1: introduction: Data Mining: Concepts and Techniques. Morgan Kaufman Publisher, USA.