

Research Article

A Study on Decision-Making of the Indian Railways Reservation System during COVID-19

Haresh Kumar Sharma ¹, Saibal Majumder ², Arindam Biswas ³,
Olegas Prentkovskis ⁴, Samarjit Kar,⁵ and Paulius Skačkauskas ⁴

¹Department of Mathematics, Shree Guru Gobind Singh Tricentenary University, Gurugram, India

²Department of Computer Science and Engineering, NSHM Knowledge Campus, Durgapur, India

³School of Mines and Metallurgy, Kazi Nazrul University (Public University), Asansol, India

⁴Department of Mobile Machinery and Railway Transport, Faculty of Transport Engineering, Vilnius Gediminas Technical University, Vilnius, Lithuania

⁵Department of Mathematics, National Institute of Technology, Durgapur, India

Correspondence should be addressed to Olegas Prentkovskis; olegas.prentkovskis@vilniustech.lt

Received 5 February 2022; Revised 24 April 2022; Accepted 23 May 2022; Published 1 July 2022

Academic Editor: Nagendra R Velaga

Copyright © 2022 Haresh Kumar Sharma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Indian Railways Reservation System (IRRS) is one of the world's busiest reservation systems of railway tickets. Recently, the COVID-19 pandemic situation has severely impacted the Indian Railway's (IR) transportation, which eventually has enforced the IR to alter the passenger reservation system. This research attempts to evaluate and analyse the factors that modify the IRRS. In this research, a rough set-based Data Mining Scaffolding (DMS) has been proposed. Here, the relevant preferential information related to the IRRS is managed by introducing a multi-criteria decision-making (MCDM), where a decision-maker (DM) can make a decision based on several decision rules. The effectiveness of the proposed DMS is explained by gathering realistic data of 26 trains, which run between railway stations of two metro cities of India during the COVID-19 pandemic period.

1. Introduction

Since the inception of the Rough Set Theory (RST) [1, 2], it has emerged as an important mathematical technique and has gradually drawn the attention of researchers and engineers to become an alternative approach of fuzzy set, vague set, etc., in order to tackle vagueness and uncertainty. The RST has wide spread applications in different domains including marketing, banking, data mining, engineering, medicine, and expert systems. The main aim of the RST is the approximation of a set by a pair of two crisp sets called the lower and upper approximation sets. The lower approximation incorporates those classifications that definitely belong to the set and the upper approximation contains those classifications that probably belong to the set. In the RST, data are represented in a tabularized format known as an information table.

As a mathematical tool for data analysis, the RST has found applications in many practical fields like data envelopment analysis (DEA) [3, 4], data mining [5], multi-criteria decision analysis [6, 7], medical diagnosis [8, 9], neural network [10], signal processing [11], etc. The RST [12–14] is a recently developed efficient technique for managing uncertainty. It has been used quite successfully while exploring data dependencies, evaluating attributes' significance, discovering data patterns, reducing redundancies, and recognising and classifying objects. Further, the extraction of rules from databases can also be done using the RST. As far as the decision-making process under the paradigm of the rough set is concerned, Stević et al. [15] proposed the concept of rough numbers and employed the rough best-worst method to determine the weight values of the criteria. Subsequently, Ye et al. [16] proposed a multi-attribute decision-making approach based on fuzzy-rough sets. Moreover, they verified the

feasibility of the proposed method by solving the building shape selection problem present in the UCI database. Furthermore, Hu et al. [17] incorporated different weights in the neighbourhood relation and proposed an innovative approach for attribute reduction of the rough set. The efficiency of the proposed method is analysed by employing it on the benchmark machine learning and biomedical datasets. Besides, some recent studies related to railways transport and its safety [18–23] are also observed in the literature.

The Indian Railways (IR) is a vast and busy network of trains for travel spanning around all corners of India. Everyday a large number of passengers reserve their tickets to avail the services of the IR. Consequently, the Indian Railways Reservation System (IRRS) remains so congested that many passengers do not get their desired seat confirmation even after multiple attempts. The parameters that influence this congestion mostly are the Current Booking Status (CBS) of the trains, Travel Time (TT), Average Speed (AS), etc. These parameters are very dynamic and change continuously with time. Hence, uncertainty is always a major factor while associating with these parameters. To process and represent such uncertainty of vagueness effectively, rough sets can be proved to be efficient in this context.

In this paper, we have considered a case study on the IRRS. We have used rough sets as the data mining tool. The main contributions in this case study are: a Data Mining Scaffolding (DMS) based on the RST has been proposed which can guide passengers efficiently to reserve their train tickets for their journey. The pertinent preferential information of ticket reservation, which is relevant to the IRRS, is then processed effectively by a Multi-Criteria Decision-Making (MCDM) approach where a decision is taken based on several “IF . . . THEN” decision rules. These decision rules are defined by considering various crucial factors regarding ticket reservation in the IRRS.

In this case study, we have defined rules based on 12 different effective parameters using the RST. The generated rule base will actually guide passengers efficiently during the reservation of tickets and will also prove worthwhile to the IRRS as it can also suggest the IR how they can improve the IRRS to serve the passengers.

The rest of the paper is organized as follows. The basic concepts and some related properties of the RST are discussed in Section “Preliminaries.” A rough set-based algorithm for the multi-criteria decision model is presented in Section “Proposed Algorithmic Approach.” The application of the proposed algorithm through a case study of the IRRS is discussed in Section “Case study.” Finally, the epilogue and some future scope of our study are stated in Section “Results, Discussion, and Conclusions.”

2. Preliminaries

This section presents a brief introduction to the RST and their related properties.

2.1. Data Table and the Indiscernibility Relation of the RST. RST is a mathematical technique to tackle vagueness and uncertainty. It is founded on the assumption that some sort

of information is associated with every object in the universe of the discourse. The theory of rough set can be recognized by means of lower approximation and upper approximation ([1], [24]). In the RST, information related to decision objects is often represented in the form of an information table. This information table is represented as a four-tuple information system $S = \langle U, A, V, g \rangle$, where U is the finite set of objects, A is the finite set of attributes, and $g: U \times A \rightarrow V$ is the information function. For any set of attributes $B \subseteq A$, there exists an equivalence relation $\text{IND}(B)$ such that $\text{IND}(B) = \{x_i, x_j \in U \times U \mid \forall b \in B, b(x_i) = b(x_j)\}$, where $b(x_i)$ represents the value of the attribute b for the element x_i . $\text{IND}(B)$ is known as the indiscernibility relation containing equivalence classes $[x_i]_{\text{IND}(B)}$. A set is known as an elementary set if it contains a set of all indiscernible members with respect to particular attributes.

Given the set of attribute $B \subseteq A$ and X in U , the lower and upper approximation of X are defined as follows:

$$\underline{X}_B = \cup \{x_i \mid [x_i]_{\text{IND}(B)} \subseteq X\}, \quad (1)$$

$$\overline{X}_B = \cup \{x_i \mid [x_i]_{\text{IND}(B)} \cap X \neq \emptyset\}. \quad (2)$$

The boundary region of set X is described as follows:

$$BN_B(X) = \overline{X}_B - \underline{X}_B. \quad (3)$$

The set \underline{X}_B is known as the set of all components, which can be surely classified as a member of X in the knowledge B , whereas \overline{X}_B is the set of elements that can be probably classified as a member of X involving knowledge B .

The boundary region $BN_B(X)$ is the set of objects, which cannot decisively classify into X consisting knowledge B . If there is no boundary region of an exact set, then its lower approximation and upper approximation sets are similar. Otherwise, if there exists a boundary region for the set, then the set X is referred as rough with respect to B . Figure 1 depicts a diagrammatic representation of a rough set.

Table 1 demonstrates a sample of a simple information table of seven objects, in which $U = \{H_1, H_2, H_3, H_4, H_5, H_6, H_7\}$. We can classify a decision table in following manner: $S = (U, A, C, D)$, where U is the universal set of houses, $A = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ is a collection of all attributes. Let us consider C (condition attributes) and D (decision attributes) to be the subsets of the attribute set A . Now, the indiscernibility relations and equivalence classes of the decision table are described as follows:

- (1) $U/\beta_1 = \{(H_2, H_4), (H_1, H_3, H_5, H_7), (H_6)\}$, in view of the objects, H_2 and H_4 have the value β_1 “high,” while the objects $H_1, H_3, H_5,$ and H_7 contribute a similar value “medium” for this attribute; object H_6 contains the value “low” (i.e., $\beta_1(H_6) = \text{low}$).

Similarly, we can compute the other indiscernibility relations:

- (2) $U/\beta_2 = \{(H_1, H_3), (H_2, H_4, H_6, H_7), (H_5)\}$,

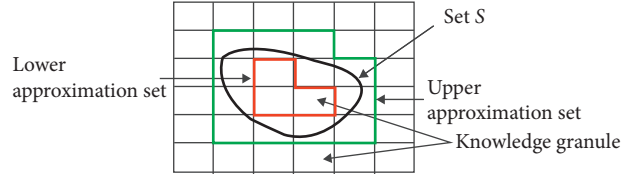


FIGURE 1: Graphical representation of a rough set.

TABLE 1: An example of an information table of house purchasing.

Houses	Price (β_1)	Location (β_2)	Size (β_3)	Colour (β_4)	Purchasing performance (β_5)
H_1	Medium	Good	Medium	Dark	Good
H_2	High	Average	Big	Light	Good
H_3	Medium	Good	Medium	Dark	Average
H_4	High	Average	Big	Light	Good
H_5	Medium	Bad	Small	Dark	Average
H_6	Low	Average	Small	Light	Good
H_7	Medium	Average	Big	Light	Average

$$(3) U/\beta_3 = \{(H_1, H_3), (H_2, H_4, H_7), (H_5, H_6)\},$$

$$(4) U/\beta_4 = \{(H_1, H_3, H_5), (H_2, H_4, H_6, H_7)\},$$

Thus, the indiscernibility relation of the decision attribute β_5 is as follows:

$$(5) U/\beta_5 = \{(H_1, H_2, H_4, H_6), (H_3, H_5, H_7)\},$$

(6) Next, an indiscernibility relation is an equivalence relation that splits the set of objects into equivalence classes. Every equivalence class consists of a set of all similar objects for the provided set of condition attribute C . In this example, $C = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ and $D = \{\beta_5\}$; it is represented as follows: $U/C = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}$ and $U/D = \{(H_1, H_2, H_4, H_6), (H_3, H_5, H_7)\}$.

Now using (1) and (2) we compute the approximate class of the set X of house purchasing performance having “good,” i.e., $X = \{H_1, H_2, H_4, H_6\}$, and house purchasing performance having “average,” i.e., $X = \{H_3, H_5, H_7\}$:

$$(1) \text{ lower approximation of class “good” purchasing performance} = \underline{C}(X_{\{\beta_5=\text{good}\}}) = \{H_2, H_4, H_6\}.$$

$$(2) \text{ upper approximation of class “good” purchasing performance} = \overline{C}(X_{\{\beta_5=\text{good}\}}) = \{H_1, H_2, H_3, H_4, H_6\}.$$

In terms of the RST, objects $H_2, H_4,$ and H_6 come in the lower approximation of X , i.e., these three objects surely belong to the set of houses having “good” purchasing performance. $H_1, H_2, H_3, H_4,$ and H_6 come in the upper approximation of X ; these five objects possibly belong to the set of houses with “good” purchasing performance.

Analogously, we can obtain the approximate class of the set X of houses having purchasing performance “average,” i.e., $X = \{H_3, H_5, H_7\}$:

$$(1) \text{ lower approximation of class “average” purchasing performance} = \underline{C}(X_{\{\beta_5=\text{average}\}}) = \{H_5, H_7\};$$

$$(2) \text{ upper approximation of class “average” purchasing performance} = \overline{C}(X_{\{\beta_5=\text{average}\}}) = \{H_1, H_3, H_5, H_7\}.$$

Example 1. Table 1 represents the data table of house purchasing. Table 1 contains seven houses that are related by means of five attributes and every house is described using four condition attributes, $C = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ and one decision attribute, $D = \{\beta_5\}$.

2.2. Accuracy of Approximation and Quality of Approximation in the RST [24]. Inexactness of a category (set) can occur due to the existence of a boundary line region. As the boundary line region of a category increases, the accuracy of the set decreases.

Numerically, we can define the accuracy of approximation in the RST by the following:

$$\alpha_B(X) = \frac{|X_B|}{|\overline{X_B}|}, \quad (4)$$

where $|X|$ represents the cardinality of any set X .

Clearly, $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, X is exact with respect to B ; otherwise, X is rough (ambiguous) with respect to B , when $\alpha_B(X) < 1$.

Let $X = \{X_1, X_2, \dots, X_m\}$ be a partition of universe U , where $X_j, j = 1, 2, \dots, m$, are classes of X and $B \subseteq A$; then, the coefficient

$$\gamma_B(X) = \frac{\sum_{j=1}^m |X_{jB}|}{|U|}, \quad (5)$$

is known as the quality of approximation of the class X by the set of attributes B where $|\cdot|$ represents the cardinality of any set; the quality of classification represents the percentage of all correctly classified objects to the partition of X employing the knowledge B .

If the quality of approximation $\gamma_B(X) = 1$, then X entirely depends on B , otherwise, X is partially dependent on

B , when $0 < \gamma_B(X) < 1$. If $B' \subseteq B$ and $\gamma_B(X) = \gamma_{B'}(X)$, it is called X -reduct of B . Information systems may contain more than one X -reduct. The intersection set of reducts is known as the core of B .

For example, the accuracy of approximation of sets Cl_a and Cl_g corresponds to the class of “average” purchasing performance, and “good” purchasing performance of houses are $\alpha_B(Cl_a)$ and $\alpha_B(Cl_g)$, respectively. The accuracy of approximation of sets Cl_a and Cl_g can be calculated as follows: $\alpha_B(Cl_a) = 2/4 = 0.5000$ and $\alpha_B(Cl_g) = 3/5 = 0.6000$, respectively. The quality of approximation can be computed using formula (4), $5/7 = 0.7143$, since the total number of elements in lower approximation is five and the total number of objects is seven.

2.3. Reduction of Knowledge. By the process of attribute removing, unnecessary attributes are removed from the dataset and a necessary attribute subset for an information system is obtained. This type of attribute subset is known as reduct, and it is a necessary part of the information system.

2.4. Positive Region, Reduct, and Core. Positive region [24] is a very crucial perception of the RST. The C -Positive region of D contains the set of all objects of the universal set U , i.e., those that are assuredly categorized into the group of U/D by attributes from C .

The C -Positive region of D is described as follows:

$$\text{POS}_C(D) = \cup_{X \in U/D} \underline{C} X. \quad (6)$$

Furthermore, an attribute, $b \in C$, is called dispensable in C corresponding to D if $\text{POS}_C(D) = \text{POS}_{(C-\{b\})}(D)$; otherwise, attribute b becomes an indispensable attribute in the information table. $B \subseteq C$ is known a reduct of the attribute set C corresponding to the decision attribute D if and only if B is an indispensable subset of C , such that $\text{POS}_C(D) = \text{POS}_B(D)$.

The core is known as the set of all common reducts of C , which consists of the set of all indispensable attributes of the information table. Moreover, the core also contains the set of more essential immovable parts of information systems.

The core can be denoted as follows:

$$\text{CORE}(C) = \cap \text{RED}(C). \quad (7)$$

Now, following Example 1, the positive region of the decision attribute D with respect to C can be calculated as: where $C = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $D = (\beta_5)$

$U/C = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}$, and $U/D = \{(H_1, H_2, H_4, H_6), (H_3, H_5, H_7)\}$; we have $\text{POS}_C(D) = \underline{C}(X_{\{\beta_5=\text{good}\}}) \cup \underline{C}(X_{\{\beta_5=\text{average}\}})$, where $\underline{C}(X_{\{\beta_5=\text{good}\}}) = \{H_2, H_4, H^6\}$ and $\underline{C}(X_{\{\beta_5=\text{average}\}}) = \{H^5, H^7\}$

Hence, $\text{POS}_C(D) = \{H_2, H_4, H_5, H_6, H_7\}$.

For computing the reduct and the core, first, we calculate the indiscernibility relation for the sequence of attribute sets:

- (1) $U/(C-\{\beta_1\}) = \{(H_1, H_3), (H_2, H_4, H_7), (H_5), (H_6)\}$;
- (2) $U/(C-\{\beta_2\}) = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}$;
- (3) $U/(C-\{\beta_3\}) = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}$;

$$(4) U/(C-\{\beta_4\}) = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}.$$

Thus, the indiscernibility relation of the decision attribute D is as follows:

- (i) $U/D = \{(H_1, H_2, H_4, H_6), (H_3, H_5, H_7)\}$;
- (ii) $U/C = \{(H_1, H_3), (H_2, H_4), (H_5), (H_6), (H_7)\}$.

And then using the positive region concept, we try to find the set of all indispensable attributes:

- (1) $\text{POS}_{(C-\{\beta_1\})}(D) \neq \text{POS}_C(D)$, then the attribute β_1 is called indispensable;
- (2) $\text{POS}_{(C-\{\beta_2\})}(D) = \text{POS}_C(D)$, then the attribute β_2 is called dispensable;
- (3) $\text{POS}_{(C-\{\beta_3\})}(D) = \text{POS}_C(D)$, then the attribute β_3 is called dispensable;
- (4) $\text{POS}_{(C-\{\beta_4\})}(D) = \text{POS}_C(D)$, then the attribute β_4 is called dispensable.

Thus, the attribute $\{\beta_1\}$ is the core of this example; we can say this attribute is most important attribute for our dataset.

Core is the most crucial part of the condition attribute C ; thus, in the decision table, attributes “price” is necessary for decision construction.

2.5. Decision-Making Using the RST. The initial dataset can be reduced without excluding any necessary attributes, which are represented in the reduct set. The minimal rule set created [24, 25] from the minimized information table possess the following steps from the RST:

Decision rules are adhered to simplify the structure of the decision table listed by following the required steps:

- (1) construction of an information table of the dataset;
- (2) calculate the lower and upper approximations of the dataset;
- (3) reduct the computation of the condition attributes, which is equivalent to the elimination of some columns of the decision table;
- (4) remove superfluous attribute value;
- (5) determine the core attribute of the attribute set A and find the minimal subset for all decision attributes.

A decision rule in an information system can be obtained by the following expression, containing decision rules of the type “IF... THEN”; thus,

IF $g(x, b_1) = v_{b_1} \wedge g(x, b_2) = v_{b_2} \wedge \dots \longrightarrow$
THEN $g(x, d) = v_d$, for each $b_i \in C, d \in D$ and, $v_{b_i}, v_d \in V$, for every $x \in U$.

3. Proposed Algorithmic Approach

This study discusses a reservation system of the IR based on the RST for the selection rules of train berths. In this study, we generate rules to determine the availability (decision attribute) of tickets (berth) in the IR depending on other condition attributes. All required steps of this approach are interpreted in the following subsections (Figure 2).

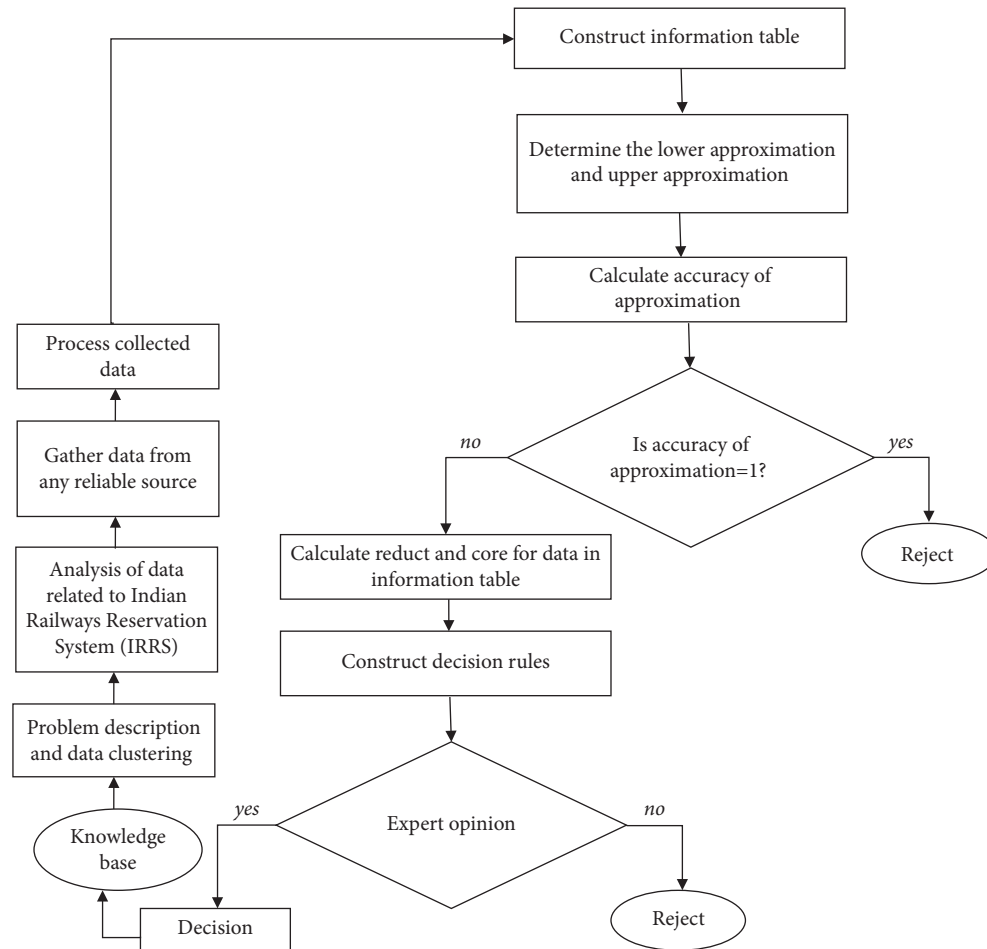


FIGURE 2: Analysis structure of the RST approach for the IRRS.

3.1. Problem Description and Data Cluster. The two most critical steps required to extract useful information are to develop the understanding of particular problems and setting their objectives. This present study of the IR has been performed for train ticket availability and other condition attributes. All useful information has been collected from the official website of the IR (<https://www.irctc.co.in>, <http://indiarailinfo.com>) and passengers' feedback. The important factors of the IRRS were critically analysed before discussing all the attributes (condition and decision). Making efficient decisions requires the knowledge of the reservation system before booking train tickets for a particular journey. The collected data have been processed precisely to ensure the high quality of subsequent analysis.

3.2. RST Examination and Determination. The current study considers various condition attributes and a decision attribute of the IRRS. The 12 factors of the IRRS have been considered as condition attributes, whereas train ticket availability has been considered as a single decision attribute. Furthermore, the accuracy of approximation, reduct, and decision rules have been calculated by approaching rough set methodologies. Moreover, rules based on the RST have been described for passengers with a view towards selecting the

most appropriate train for their journey. The accuracy of approximation of randomly generated data is analysed by a set of RST tools like ROSE2.

3.3. Information Extraction. To verify the pragmatism and efficiency of the analysed result, the qualified rules must be examined and inspected by the mature judgement of domain experts. Any unexpected situations regarding the considered dataset need to be examined properly before the rules are applied to develop strategies for the availability of a train berth for a passenger at the time of reservation. The rules developed for an IRRS must be reviewed at regular intervals. In order to establish the validity of the derived rule base of a dynamic system like the IRRS, the rule base should be subjected to continuous surveillance and systematic examination by domain experts.

4. Case Study

This section presents the application of the proposed approach using the data from the IR.

4.1. Research Problem and Data Collection [26]. The IR is the premier transport organization of India, and it is Asia's

largest and World's second largest rail network under a single management service. The IR has been owned and operated by the Government of India through the Ministry of Railways. Back in 1853, Railways were first introduced in India between Mumbai and Thane. Thereafter in 1951, the various constituent units of railways were nationalised as a single unit, i.e., the IR. Here, the current study includes the Delhi to Mumbai multi-gauge long distance rail network, which is one of the oldest routes with the maximum number of passenger trains running through the route. For passenger amenities, the IR created the Indian Railway Catering and Tourism Corporation (IRCTC), which handles the catering, tourism, and the IRRS. The present study is focused on the IRRS for selected 26 best passenger trains between the metro cities Delhi and Mumbai. From the passenger point of view, it is one of the critical factors to reserve a confirm ticket to a suitable train running between the two metro cities during COVID-19. For the passengers, the possibility of getting a confirmed ticket is maximised by selecting certain related attributes of the IRRS. Therefore, the data related to the berth reservation of passenger trains have been considered one month before the date of journey. To understand the relationship between the IRRS factors and different IRRS variables affecting the decision of the passengers, data have been listed from the IRCTC and the IR for 26 best passenger trains on the route from Delhi to Mumbai.

The information for the present study has been collected with the help of various domain experts from the IR, IRCTC, and also from passengers' feedback. Passenger decision is controlled by the decision attribute, which is governed by conditional attributes. Conditional attributes were further classified into IRRS factors, which have been derived from the IRRS variables. Hence, the driving force for decision-making in this study is the IRRS variables, which include (i) "Departure time of train," "Travel time," "Running days," and "Punctuality of train" (based on the interview of the deputy chief controller/dispatcher); (ii) "Current booking status," "Fare of ticket," and "Ticket availability" (based on the information provided by the IR senior divisional commercial staff and zonal officer); (iii) "Distance from source to destination" and "Average Speed" (based on the Research Designs & Standards Organisation (RDSO) guideline and dispatcher); (iv) "Cleanliness," "Food quality," and "Railfanning" (based on a discussion with private contractors for hire by the IRCTC); (v) "Safety of train" (Railway police force (RPF) guidelines).

All considered attributes, factors, and variables with explanation can be seen in Table 2.

Collected data have been pre-processed to convert it into the most suitable format for analysing and consequently deriving meaningful information (Table 3).

4.2. Analysis of the Case Study. The knowledge is mapped from 13 important features (criteria) of the IR to the 26 trains arranged in Table 3 for the rough set analysis. Out of these 13 attributes, 12 are considered as the condition attributes and the remaining one is recognised as a decision attribute (ticket availability). Ticket availability is again

divided into four decision classes represented as poor, average, good, and excellent. The following Table 4 shows the approximation of sets and the accuracy of approximation.

The quality of lower approximation: 0.8462.

4.3. Decision Rules Using the RST. The decision rule of the initial dataset (Table 5) is obtained using objects supporting a certain decision rule.

Certain decision rules can be expressed in the form of IF-THEN form. Here are some example to illustrate the IF-THEN rules:

- (1) IF the departure time is evening, AND safety is average, THEN the decision ticket availability will be poor,
- (2) IF the fare of ticket is medium, AND the train's speed is average, THEN the decision ticket availability will be good.

We can see from Table 5 that if the departure time is evening and the train's safety is average, then ticket availability is poor, which eventually means that there is a very less chance of getting a ticket for that particular train. Rule 4, if the running days of trains are biweekly, then ticket availability is good. According to rule 2, if the distance is far, train speed is fast, and punctuality of train is excellent, then ticket availability is average

4.4. Machine Learning Implementation. Using the attribute reduction technique of the rough set, we have observed that the attributes Punctuality of train (PT) and Railfanning (RF) are eliminated from the list of all the 12 attributes. For comparison purposes, we have employed Recursive Feature Elimination with Cross-Validation (RFECV) as one of the machine learning (ML) techniques on the same dataset. Subsequently, we observed that the identical attributes, PT and RF, are also eliminated from the dataset. The feature importance of the remaining 10 attributes is presented in Table 6. Here, we observe that Safety of train (SF) and Ticket availability (TA) are selected as the two most important attributes with the maximum feature importance score as determined by the RFECV when applied on the classifiers. In this study, we have also considered two classifiers: Random Forest Classifier (RFC) [27] and Extra Trees Classifier (ETC) [28], and analysed the predictive capability of the classifiers with respect to the cross-validation score as well as six performance metrics: (i) accuracy, (ii) precision, (iii) recall, (iv) $f1$ -score, (v) Hamming loss, and (vi) Matthews correlation coefficient. The cross-validation score is determined by considering the *StratifiedKfold* cross-validation technique, where K is set to 10. We have considered *RandomizedSearchCV* to optimize the hyper-parameters (cf. Table 7), which are the same for RFC ETC. For both the classifiers, *random_state* is set to 9. For the experimental study, we have used *Jupyter* notebook server 6.4.2 and *Python* 3.9.6 to implement the ML techniques on our dataset.

Subsequently, we split the dataset into training and testing datasets with 80% and 20% of data samples,

TABLE 2: Ticket availability issue collaborated with the listed IRRS factors associated with certain IRRS variables.

Attribute	IRRS factor	IRRS factor classification	IRRS variables	Explanation
Conditional attribute	Reservation	Time	(DT) departure time of train	Time at which particular train departed from the source station
Conditional attribute	Reservation	Time	(TT) travel time	Actual traveling time of particular train
Conditional attribute	Reservation	Time	(RD) running days	Days in which particular train run
Conditional attribute	Reservation	Time	(PT) punctuality of train	Running performance of particular train
Conditional attribute	Reservation	Booking particulates	(CBS) current booking status	Present scenario of train berth
Conditional attribute	Reservation	Booking particulates	(FT) fare of ticket	Fixed amount of money paid for particular journey
Conditional attribute	Reservation	Distance	(DST) distance from source to destination	Total distance covered by train from different source (Delhi) to different destination (Mumbai)
Conditional attribute	Reservation	Distance	(AS) average speed	Train mean speed
Conditional attribute	Passenger amenities	Facility and security	(CN) cleanliness	Cleaning status of a particular train as per IR guidelines
Conditional attribute	Passenger amenities	Facility and security	(FD) food quality	Passengers' health acceptable standard characteristics of food
Conditional attribute	Passenger amenities	Facility and security	(RF) railfanning	Train transport recreational ability access by the passengers
Conditional attribute	Passenger amenities	Facility and security	(SF) safety of train	All measures of passengers' safety and security in particular train
Decision attribute	Decision making	—	(TA) ticket availability	Average availability in particular train of seat/berth

TABLE 3: Information table after discussions with an expert of the IR.

Train no.	CBS*	DT	TT	DST	FT	RD	AS	CN	PT	FD	RF	SF	TA
1	WL	Day	More	Very far	Medium	Daily	Fast	Average	Average	Good	Good	Good	Average
2	CNF	Day	Less	Very far	Low	Biweekly	Fast	Good	Good	Good	Good	Good	Good
3	WL	Day	Less	Far	Low	Daily	Fast	Excellent	Good	Good	Excellent	Excellent	Good
4	WL	Day	More	Very far	Medium	Daily	Fast	Average	Average	Good	Good	Good	Good
5	CNF	Day	Less	Very far	Low	Frequently	Fast	Good	Excellent	Good	Excellent	Good	Good
6	RAC	Day	Less	Very far	Low	Weekly	Fast	Excellent	Good	Good	Good	Excellent	Average
7	CNF	Evening	Very less	Far	Low	Weekly	Very fast	Average	Good	Average	Excellent	Good	Good
8	RAC	day	Less	Very far	Low	Weekly	Fast	Good	Good	Average	Good	Average	Excellent
9	WL	Evening	Less	Very far	Low	Weekly	Fast	Good	Good	Average	Good	Average	Poor
10	RAC	Evening	Very less	Very far	High	Daily	Very fast	Excellent	Excellent	Good	Excellent	Excellent	Excellent
11	WL	Evening	Less	Far	Low	Daily	Fast	Good	Excellent	Good	Excellent	Good	Average
12	WL	Night	More	Very far	Medium	Daily	Average	Good	Good	Average	Excellent	Good	Good
13	CNF	Night	Less	Far	Low	Biweekly	Very fast	Good	Excellent	Average	Excellent	Good	Good
14	WL	Night	More	Far	Low	Daily	Average	Average	Average	Average	Poor	Average	Excellent
15	RAC	Evening	Very less	Far	High	Daily	Very fast	Excellent	Excellent	Good	Excellent	Excellent	Good
16	WL	Day	Very less	Very far	High	Frequently	Very fast	Good	Excellent	Good	Excellent	Excellent	Good

TABLE 3: Continued.

Train no.	CBS*	DT	TT	DST	FT	RD	AS	CN	PT	FD	RF	SF	TA
17	RAC	Night	Less	Far	Low	Frequently	Fast	Good	Excellent	Good	Excellent	Good	Average
18	CNF	Night	Very less	Very far	High	Biweekly	Very fast	Good	Good	Average	Good	Good	Good
19	RAC	Night	Less	Very far	Medium	Biweekly	Fast	Good	Good	Good	Excellent	Excellent	Good
20	WL	Day	Less	Far	Low	Biweekly	Fast	Good	Good	Excellent	Excellent	Good	Good
21	WL	Day	Less	Very far	Low	Biweekly	Fast	Good	Good	Average	Excellent	Excellent	Good
22	CNF	Evening	Less	Very far	Medium	Weekly	Very fast	Excellent	Poor	Good	Good	Good	Excellent
23	WL	Night	Less	Far	Low	Weekly	Fast	Excellent	Excellent	Excellent	Excellent	Excellent	Average
24	RAC	Day	Less	Very far	Low	Weekly	Fast	Good	Good	Average	Good	Average	Good
25	WL	Day	Very less	Very far	High	Biweekly	Very fast	Excellent	Excellent	Excellent	Excellent	Excellent	Good
26	RAC	Evening	Very less	Far	Low	Frequently	Very fast	Good	Excellent	Good	Good	Good	Good

*WL → Waiting List, CNF → Confirmed, RAC → Reservation Against Cancellation.

TABLE 4: Approximation sets and accuracy of approximation.

	Poor	Average	Good	Excellent
Lower approximation	1	4	14	3
Upper approximation	1	6	18	5
Boundary	0	2	4	2
Accuracy of approximation	1	0.6667	0.7778	0.6000

TABLE 5: Certain decision rules of the railway information system.

Rule no.	Condition parts	Decision parts	Support
1	If (departure time = evening) & (safety = average)	Then (availability = poor)	1
2	If (distance = far) & (speed = fast) & (punctuality = excellent)	Then (availability = average)	3
3	If (booking status = RAC) & (running days = weekly) & (cleanliness = excellent)	Then (availability = average)	1
4	If (running days = biweekly)	Then (availability = good)	7
5	If (running days = frequently) & (speed = very fast)	Then (availability = good)	2
6	If (distance = far) & (running days = daily) & (safety = excellent)	Then (availability = good)	2
7	If (booking status = CNF) & (railfanning = excellent)	Then (availability = good)	3
8	If (fare = medium) & (speed = average)	Then (availability = good)	1
9	If (distance = very far) & (speed = very fast) & (cleanliness = excellent) & (food = good)	Then (availability = excellent)	2
10	If (railfanning = poor)	Then (availability = excellent)	1

TABLE 6: Feature importance score of the 10 selected attributes.

Feature	Importance score
DT	0.06
TT	0.13
DST	0.09
FT	0.06
RD	0.09
AS	0.09
CN	0.06
FD	0.12
SF	0.15
TA	0.15

TABLE 7: Hyper-parameters of RFC ETC.

Hyperparameter	Value
n_estimators	127
min_samples_split	10
min_samples_leaf	2
max_features	“auto”
max_depth	36
bootstrap	False

respectively. Once the classifiers are trained on the training dataset, we observe that all the data samples of the testing set are predicted correctly. This fact is well observed by studying

TABLE 8: Cross-validation score and the six performance metrics of both RFC ETC.

Performance indicators	RFC	ETC
Accuracy	1.0	1.0
Precision	1.0	1.0
Recall	1.0	1.0
$f1$ -score	1.0	1.0
Hamming loss	0.0	0.0
Matthews correlation coefficient	1.0	1.0
Cross-validation score	0.9817	0.9953

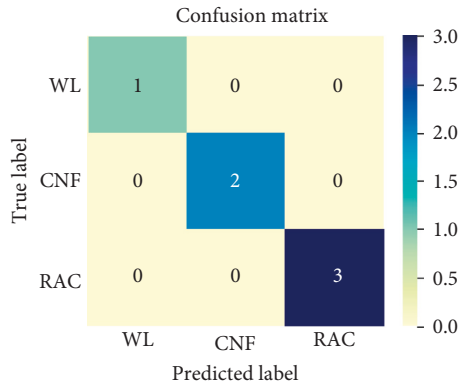


FIGURE 3: Confusion matrix of both RFC ETC.

the six performance metrics reported in Table 8 and the confusion matrix depicted in Figure 3. For both the classifiers, the values of all the performance metrics are the same and the confusion matrices are identical. However, while determining the cross-validation score with 10-fold stratified cross-validation, it is observed that ETC outperforms RFC with a better cross-validation score. This is also listed in Table 8 with the better cross-validation score highlighted as bold.

5. Results, Discussion, and Conclusions

Due to the COVID-19 lockdown, numerous passenger trains of the IR were suspended. Such a scenario has created a chaotic situation for the passengers who wish to travel for their urgency. Hence, to normalize the troublesome situation to some extent, the government of India allowed some important trains for significant routes. Delhi to Mumbai is essentially an important route, as Mumbai is the financial capital of India. Furthermore, the route is financially profitable for the IR. As a matter of fact, in this study, we have focused on 26 best trains that were functional after COVID-19 lockdown in the country.

The railway reservation system is dynamic in nature since the status of the train berth, i.e., CNF, RAC, WL, alters quite frequently. This dynamic nature of the data is basically dependent on imprecise attributes, which have been processed in this study by employing rough set approaches. We have analysed the data using the RST, which will actually guide the passenger while reserving a train berth.

In this paper, we have presented a MCDM problem based on the IRRS. In this problem, a DM (passenger) has to decide whether to reserve a train berth for availing the journey on a particular train between stations, based on 12 conditional attributes of the IRRS. The analysis has been performed on real datasets corresponding to the IR. According to passenger decisions, we have analysed a set of attributes, which are necessarily critical as far as the reservation of the tickets in a particular train is concerned. In this context, we have used the attribute reduction technique of the rough set. This approach is also essential to analyse a suitable train for the journey. For comparative analysis, we have also used machine learning approaches by considering two machine learning estimators, RFC ETC.

In contrast to the traditional technique, the proposed approach provides a technique to help passengers select the convenient train that suits their journey.

As a future research interest, the extension of the proposed data mining technique of the decision support system can prove to be beneficial to its proprietary. Further, the natural language interpretation of the decision rules can be understood in a better way. Moreover, the development of some suitable metrics in order to remove redundant and contradictory decision rules is also considered as our future research interest.

Data Availability

The data used to support the findings of this study are available from the first/corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Haresh Kumar Sharma and Saibal Majumder conceptualized the study. Haresh Kumar Sharma developed the methodology; Saibal Majumder and Arindam Biswas dealt with the software; Haresh Kumar Sharma and Saibal Majumder validated the results; Olegas Prentkovskis and Paulius Skačkauskas were responsible for the formal analysis; Haresh Kumar Sharma and Olegas Prentkovskis dealt with the investigation; Paulius Skačkauskas collected the resources; Haresh Kumar Sharma and Saibal Majumder curated the data; Haresh Kumar Sharma prepared and wrote the original draft; Haresh Kumar Sharma, Saibal Majumder, Olegas Prentkovskis, Paulius Skačkauskas, and Samarjit Kar reviewed and edited the manuscript; Haresh Kumar Sharma and Arindam Biswas were responsible for visualization; Olegas Prentkovskis and Paulius Skačkauskas supervised the study; All the authors have read and agreed to the published version of the manuscript.

References

- [1] Z. A. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

- [2] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information Sciences*, vol. 177, no. 1, pp. 3–27, 2007.
- [3] M. Shafiee and N. Shams-e-alam, "Supply chain performance evaluation with rough data envelopment analysis," in *Proceedings of the International Conference on Business and Economics Research*, vol. 11ACSIT Press, Kuala Lumpur, Malaysia, October 2011.
- [4] J. Xu, B. Li, and D. Wu, "Rough data envelopment analysis and its application to supply chain performance evaluation," *International Journal of Production Economics*, vol. 122, no. 2, pp. 628–638, 2009.
- [5] T. Lin, Y. Yao, and L. Zadeh, *Data Mining, Rough Sets, and Granular Computing*, Springer-Verlag, Berlin, Germany, 2002.
- [6] S. Greco, B. Matarazzo, and R. Slowinski, "A new rough set approach to multicriteria and multiattribute classification," in *Rough Sets and Current Trends in Computing Rough Sets and Current Trends in Computing*, L. Polkowski and A. Skowron, Eds., Springer-Verlag, Berlin, Germany, pp. 60–67, 1998.
- [7] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets methodology for sorting problems in presence of multiple attributes and criteria," *European Journal of Operational Research*, vol. 138, no. 2, pp. 247–259, 2002.
- [8] S. Hirano and S. Tsumoto, "Rough representation of a region of interest in medical images," *International Journal of Approximate Reasoning*, vol. 40, no. 1-2, pp. 23–34, 2005.
- [9] S. Tsumoto, "Medical differential diagnosis from the viewpoint of rough sets," *Information Sciences 2007*, pp. 28–34, 2007.
- [10] P. Lingras, *Rough Neural Networks, International Conference on Information Processing and Management of Uncertainty*, pp. 1445–1450, Granada, Spain, 1996.
- [11] S. Xiao and E. M.-K. Lai, "A Rough Programming Approach to Power-Aware VLIW Instruction Scheduling for Digital Signal Processors," in *Proceedings of the (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, Philadelphia, PA, USA, March 2005.
- [12] G. Liu and W. Zhu, "The algebraic structures of generalized rough set theory," *Information Sciences*, vol. 178, no. 21, pp. 4105–4113, 2008.
- [13] J. Liang, J. Wang, and Y. Qian, "A new measure of uncertainty based on knowledge granulation for rough sets," *Information Sciences*, vol. 179, no. 4, pp. 458–470, 2009.
- [14] L. Jian, S. Liu, and Y. Lin, *Hybrid Rough Sets and Applications in Uncertain Decision-Making*, CRC Press, Taylor and Francis Group, London, UK, 2011.
- [15] Ž. Stević, D. Pamučar, E. Kazimieras Zavadskas, G. Čirović, and O. Prentkoviškis, "The selection of wagons for the internal transport of a logistics company: a novel approach based on rough bwm and rough saw methods," *Symmetry*, vol. 9, no. 11, p. 264, 2017.
- [16] J. Ye, J. Zhan, and Z. Xu, "A novel multi-attribute decision-making method based on fuzzy rough sets," *Computers & Industrial Engineering*, vol. 155, Article ID 107136, 2021.
- [17] M. Hu, E. C. C. Tsang, Y. Guo, D. Chen, and W. Xu, "A novel approach to attribute reduction based on weighted neighborhood rough sets," *Knowledge-Based Systems*, vol. 220, Article ID 106908, 2021.
- [18] Y. Guo and B. Dong, "Railway and trade in modern China: evidence from the 1930s," *China Economic Review*, vol. 69, Article ID 101661, 2021.
- [19] S.-H. Rao, "Transportation synthetic sustainability indices: a case of Taiwan intercity railway transport," *Ecological Indicators*, vol. 127, Article ID 107753, 2021.
- [20] M. Vojtek, J. Matuska, J. Siroky, J. Kugler, and M. Kendra, "Possibilities of railway safety improvement on regional lines," *Transportation Research Procedia*, vol. 53, pp. 8–15, 2021.
- [21] E. Cascetta, A. Carteni, I. Henke, and F. Pagliara, "Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post evaluation and future perspectives," *Transportation Research Part A: Policy and Practice*, vol. 139, pp. 412–428, 2020.
- [22] K. Papoušková, "The concept of railway transport in the Czech republic," *Transportation Research Procedia*, vol. 53, pp. 154–158, 2021.
- [23] R. Zhang, L. Li, and W. Jian, "Reliability analysis on railway transport chain," *International Journal of Transportation Science and Technology*, vol. 8, no. 2, pp. 192–201, 2019.
- [24] Z. Pawlak, *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.
- [25] J. W. Grzymala-Busse, "LERS-A system for learning from examples based on rough sets," in *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, R. kki, Ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 3–18, 1992.
- [26] L.-F. Chen and C.-T. Tsai, "Data mining framework based on rough set theory to improve location selection decisions: a case study of a restaurant chain," *Tourism Management*, vol. 53, pp. 197–206, 2016.
- [27] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [28] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized Trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.