Paradigm

Transport and Telecommunication, 2026, volume 27, no. 1, 11-27
Transport and Telecommunication Institute, Lauvas 2, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2026-0002

ON A HYBRID DECISION SUPPORT FRAMEWORK FOR TRAIN SELECTION IN INDIAN RAILWAYS: AN INTEGRATION OF DOMINANCE-BASED ROUGH SET APPROACHES AND MACHINE LEARNING MODELS

Haresh Kumar Sharma[1], Anupama Singh[2], Olegas Prentkovskis[3], Saibal Majumder[4]

[1]Department of Operations and Decision Sciences, Birla Institute of Management Technology
Greater Noida, India
[1]hareshshrm@gmail.com
[2]Department of Strategic Environmental Management, Birla Institute of Management Technology
Greater Noida, India
[2]anupama.singh@bimtech.ac.in
[3]Department of Mobile Machinery and Railway Transport, Vilnius Gediminas Technical University
Vilnius, Lithuania
[3]olegas.prentkovskis@vilniustech.lt
[4]Department of Computer Science and Engineering (Data Science), Dr. B. C. Roy Engineering College
Durgapur, India
[4]saibal.majumder.1729@gmail.com

The Indian Railway Catering and Tourism Corporation (IRCTC) operates one of the most heavily utilized railway reservation systems globally, reflecting the central role of Indian Railways (IR) as an affordable and essential mode of transportation across the country. However, selecting an appropriate train remains a complex decision-making task for passengers, primarily due to the uncertainty surrounding ticket availability on preferred travel dates. To address this challenge, the present study proposes a hybrid decision support system designed to aid passengers in selecting optimal train options, particularly for tourism-related travel. This research employs a Dominance-Based Rough Set Approach (DRSA) within a Multi-Criteria Decision-Making (MCDM) framework to analyze preference-based data and extract interpretable decision rules in the form of "if...then" statements. These rules assist decision makers in evaluating multiple train-related criteria simultaneously. For comparative purposes, the Classical Rough Set Approach (CRSA) is also implemented to identify the relative advantages and limitations of both rough set methodologies in addressing train selection complexity. In addition, the study integrates machine learning techniques by utilizing two predictive models – Extreme Gradient Boosting (XGBoost) and Support Vector Machine Classifier (SVMC) – to estimate overall train ratings based on user preferences and historical data. Model performance is evaluated using standard classification metrics, including accuracy and precision. By combining MCDM techniques with machine learning algorithms, the proposed hybrid framework enhances the train reservation experience, enabling passengers to make informed, preference-aligned travel decisions through the Indian Railways reservation system.

Keywords: Multiple criteria decision making, dominance-based rough set, data mining, railway reservation system

1. Introduction

The Indian Railways, run by the Ministry of Railways, is among the world's biggest railway networks. Because it goes to nearly every corner of the country, millions of people rely on it every day for transportation. The first passenger train in India travelled the 34 kilometres from Mumbai (formerly Bombay) to Thane in 1853, marking the beginning of Indian Railways. The age of India's railways officially began at this point. The country's rail network quickly expanded throughout time. For the purpose of centralizing and streamlining rail operations, Indian Railways was formed in 1951 as a consequence of the merger and nationalization of numerous private and regional railway companies (Indian Railways, 2025).

The Indian Railway Catering and Tourism Corporation (IRCTC) was formed as a specialized subsidiary of the Indian Railways with the goal of improving passenger services and bringing professionalism to non-core operations like ticketing and catering. The Indian Railways Reservation System (IRRS), marketing rail-based tourism packages, and managing on board and station-based food are all under the purview of this organization.

In addition, there is a wide variety of trains in India, from express trains like the Vande Bharat to slower local trains that carry commodities and passengers. To accommodate passengers of varying financial means, the railway runs a wide range of carriage classes, from AC First Class to General. During festivals and holidays, Indian Railways conducts special trains to handle the increased demand for passengers. Train travel is now more accessible than ever before thanks to reservation systems like Tatkal and internet booking. In order to provide passengers with up-to-the-minute train information, platforms have digital screens and announcement systems. Luxury tourist trains, such as the Palace on Wheels, are also operated by the Indian Railways. Train and station modernization, safety, and punctuality are continuous concerns. Trains facilitate travel and commercial activity in India, making them the backbone of the country's public transportation infrastructure.

A mathematical method utilized to manage ambiguity, incompleteness, and uncertainty in data analysis, Rough Set Theory was first proposed by Zdzisław Pawlak in the early 1980s. The idea behind it is indiscernibility, which involves labeling things as belonging to the same category if there is no way to tell them apart using the criteria that are currently accessible. In order to get close to a target concept, the theory uses two sets of objects: one that definitely belongs to the concept (the lower approximation) and another that might belong (the upper approximation). The region of uncertainty is brought to light by the disparity between these approximations. When it comes to feature selection, data mining, and making decisions, rough sets are your best bet. Rough set theory differs from fuzzy and probabilistic models in that it doesn't necessitate any sort of membership function or probability distribution as input.

Dominance based rough set approach (DRSA) (Greco *et al.*, 2001) is the extension of classical based rough set approach for multi-criteria decision analysis which depends on the dominance concept that would allow it to tackle the inconsistency. DRSA converts the indiscernibility relation by a dominance relation that provides the flexibility to deal with inconsistencies related to criteria and preference-ordered decision classes. The multi-criteria decision proposed for this study depends on the dominance-based Rough Set Approach. Classical rough set approach (CRSA) (Pawlak, 1982) is restricted to sorting problems where the set of attributes (criteria) under consideration contain preference-order pairs. This is necessitated by the inconsistencies generated in violating the dominance principle. In order to deal with such inconsistencies, some methodological changes to the original rough set theory are required. There had been some instances in the past where DRSA has been applied to related problems. Augeri *et al.* (2015) used DRSA model to set the speed limits for vehicles in speed controlled zones. Chakhar and Saad (2012) studied MCDM problems using DRSA approach. Du and Hu (2016) applied DRSA to investigate all the reducts for incomplete ordered information systems. Chakhar *et al.* (2016) implemented DRSA approach to generate rules to study multi-criteria group decision making on different case studies. Majumder *et al.* (2024) developed a hybrid model based on a rough set–based decision support system and machine learning models to evaluate train performance. Li *et al.* (2015) studied different composition and decomposition of granules in DRSA for computation approximations in the corresponding parallel algorithms. Sawicki and Żak (2014) applied DRSA on transportation system by generating certain decision rules based on customer expectation and opinion. Liou and Tzeng (2010) determined DRSA approach to formulate airline service strategies. Chen *et al.* (2019) used fuzzy set theory for optimizing the draft passenger train timetable. Sharma and Kar (2018) employed Classical Rough Set Theory (CRST) to derive "if–then" decision rules from hotel industry datasets. Chang *et al.* (2017) used multilayer genetic algorithm in integrated scheduling in railway. Jiten *et al.* (2016) study bidirectional pedestrian movement in railways station. Givoni and Chen (2017) study airline and railway disintegration in China. Nosheen *et al.* (2022) proposed a DRSA-based parallel technique to compute the approximation set. In this study, the model is applied in practice for classification, prediction, and decision-making. To evaluate the performance of the proposed model, the authors selected 10 different datasets from the source UCI. Nawaz *et al.* (2025) presented a novel framework for improving the efficiency of DRSA by using K-dimensional (K-D) tree partitioning and adaptive recalculation techniques. The proposed model enhances computational performance, especially for large datasets, by optimizing data partitioning and reducing redundant calculations. Chen and Zhu (2024) proposed a dominance-based neighbourhood rough set (DNRS) model to address feature selection for hybrid ordered data (HODs), which include different set-valued criteria. Experimental validation on nine UCI datasets demonstrates the effectiveness and efficiency of the proposed algorithms.

To the best of our knowledge, the literature has not examined the applicability of DRSA within the Indian railway reservation system. This paper examines a case study of a train ticket reservation system in India, utilizing a dominance-based rough set technique to derive specific decision rules. These regulations ultimately determine if an official of Indian Railways enhances service quality of trains. The ranking of trains based on railway officials depends on a few sensible parameters, viz., "Journey duration," "Promptness of train," "Ongoing seat status," "Ticket expense," "Travel span," "Standard speed," "Hygiene," and "Railway security, etc. This paper aims to present a novel extraction method for preference ranking in railway reservation systems, and the results are obtained through the dominance-based rough set approach. The primary objective of this study is to elucidate the preferences of the variables by examining the interaction between Indian Railways and decision-making processes.

The article is structured in the following way: Section *Preliminaries* introduces the fundamental concepts and related properties of Dominance-Based Rough Set Theory. The Section *Case Study* illustrates the application of the proposed model through an example of Indian railway train selection. Finally, the Section Results, Discussion, and Conclusion provide a summary of the study and outline potential directions for future research.

2. Preliminaries

An overview of the dominance-based rough set approach and its related properties is given in this section. Furthermore, we also discuss the two machine learning estimators used in this study.

The combination of dominance-based rough set approaches (DRSA) and machine learning in this study follows directly from the structure of the railway reservation problem itself. Passenger decisions are shaped by ordered criteria such as seat status, speed, hygiene, and security, yet these criteria interact in ways that frequently violate strict monotonic assumptions. Reservation availability also changes continuously, introducing inconsistency that cannot be ignored. DRSA addresses this situation more realistically than classical rough sets because it relies on dominance relations rather than exact indiscernibility. This allows preference-ordered attributes to be modelled even when data exhibit contradictions. The results reported in the manuscript show that DRSA produces more stable approximation behaviour for the railway dataset, which aligns with the inherently dynamic nature of the reservation system. An additional advantage is interpretability: the derived decision rules clearly explain how specific combinations of attributes influence reservation availability, which is essential for decision support rather than pure prediction.

Prediction, however, remains a separate concern. Rule-based models alone are not designed to generalize efficiently across large or unseen datasets. For this reason, machine learning models are introduced as a complementary layer. XGBoost and Support Vector Machines capture non-linear relationships among the same criteria and provide a quantitative assessment of predictive performance. Their evaluation using standard classification metrics serves to validate and strengthen the analytical findings obtained through DRSA.

The integration is therefore deliberate rather than redundant. DRSA explains the decision logic in preference-consistent terms, while machine learning evaluates how reliably those decisions can be predicted at scale. Together, they form a decision support framework that balances interpretability with predictive strength, which is necessary for a real-world system such as the Indian Railways Reservation System.

Dominance-based RST:

Earlier research had proposed the dominance-based rough set theory (DRST) (Greco *et al.*, 1998; Greco *et al.*, 2001) has come into existence to overcome the disadvantages of Classical rough set theory (CRST) (Pawlak, 1982) in multi-criteria decision-making models.

Basic concepts of DRST:

In RST, the information table often represents the information of decision objects. This, in turn, becomes a four-tuple information system S=$\langle U, A, V, g\rangle$. The non-empty finite set of objects and the finite set of criteria are respectively U and A, such that r: U→ $V_r, \forall r \in A$, V_r is a domain of criteria r. $g: U \times A \to V$ is the total function expressed such that $g(j, r) \in V_r$, $\forall r \in A$ and $j \in U$. A is divided into the non-empty subset of condition attributes C and a non-empty subset of decision attribute D such that $C \cup D = A$ and $C \cap D = \emptyset$. In this case, the information table is also regarded as a decision table.

In multi-criteria decision making, the domains of condition attributes are ranked according to decreasing and increasing preferences. These attributes are known as criteria. DRSA considers that preference increases with the value of $g(\cdot, r)$ $\forall$ r ∈ C. We assume the set of decision classes of singleton, i.e., $|D| = 1$, the decision attribute $d \in D$ partitions U into decision classes $Cl = \{\mathrm{Cl}_q , \mathrm{q} \in \mathrm{Q}\}$ which are finite in number. Here Q = {1, 2, …, n}, ∀ j ∈ U belongs to the unique class. In addition, we also considered for all $l, m \in Q$ such that if $l > m$ then Cl_l are preferred over Cl_m.

Dominance relation and approximation sets (Greco et al. 1998; Greco et al. 2001):

Suppose P is a subset of condition attributes i.e.$P \subseteq C$. The dominance relation ∇_P related with P is described for every couple of objects j and k so;

$j\nabla_{\mathrm{P}}\mathrm{k} \Leftrightarrow g(j, r) \preccurlyeq g(\mathrm{k}, \mathrm{r}), \forall \mathrm{r} \in \mathrm{P}.$

The symbol "≼" should be changed with "≽" for criteria according to the increasing preference. There are two important sets in dominance relation with every object j ∈ U: (i) P-dominating set $\nabla_P^+(j) = \{k \in U: k\nabla_P j\}$ including objects that dominate j and (ii) the P-dominated set $\nabla_P^-(j) = \{k \in U: j\nabla_P k\}$ Including objects that dominated by j.

In DRSA, the upward union ($Cl_q^{\geqq}$) and downward union ($Cl_q^{\leqq}$) of classes are defined as follows:

$Cl_q^{\geqq} = \bigcup_{m\geq q} Cl_m$,$Cl_q^{\leqq} = \bigcup_{m\leq q} Cl_m$.

The assertion "$j \in Cl_q^{\geqq}$" means "j is a member of at least class Cl_q"and "$j \in Cl_q^{\leqq}$" means "j is a member of at most class Cl_q". The P-lower and P-upper approximation of $Cl_q^{\geqq}$, $q \in Q$ where P⊆ C are defined as below:

$\underline{P}\left(Cl_q^{\geqq}\right) = \{j \in U: \nabla_{\mathrm{P}}^+(j) \subseteq Cl_q^{\geqq}\}$.

$\overline{P}\left(Cl_q^{\geqq}\right) = \bigcup_{j\in Cl_q^{\geqq}} \nabla_{\mathrm{P}}^+(j) = \{\ j \in U: \nabla_{\mathrm{P}}^-(j) \cap Cl_q^{\geqq} \neq \emptyset\}$.

Similarly, the P-lower and P-upper approximation of $Cl_q^{\leqq}$, $q \in Q$ where P⊆ C are defined as below:

$\underline{P}\left(Cl_q^{\leqq}\right) = \{j \in U: \nabla_{\mathrm{P}}^-(j) \subseteq Cl_q^{\leqq}\}$.

$\overline{P}\left(Cl_q^{\leqq}\right) = \bigcup_{j\in Cl_q^{\leqq}} \nabla_{\mathrm{P}}^-(j) = \{\ j \in U: \nabla_{\mathrm{P}}^+(j) \cap Cl_q^{\leqq} \neq \emptyset\}$.

The P- boundary region of $Cl_q^{\geqq}\ and\ Cl_q^{\leqq}$ are described as follows:

$Bnd_P\left(Cl_q^{\geqq}\right) = \overline{P}\left(Cl_q^{\geqq}\right) - \underline{P}\left(Cl_q^{\geqq}\right)$,

$Bnd_P\left(Cl_q^{\leqq}\right) = \overline{P}\left(Cl_q^{\leqq}\right) - \underline{P}\left(Cl_q^{\leqq}\right)$.

The doubtful region objects which can be neither in nor out as members of the class Cl_q.

Example: examine Table 1 and assume that U = {C_1, C_2, C_3, C_4, C_5, C_6, C_7}. With the concept of dominance relation obtain the approximate class $Cl_a^{\leqq}$of overall performance of car at most "average" and the class $Cl_g^{\geqq}$ of overall performance at least "good". For decision attribute w_5, cars are classified into two preference-ordered classes, we have $Cl_a^{\leqq} =\ Cl_a$ and $Cl_g^{\geqq} = Cl_g$. As formerly, C = (w_1, w_2, w_3, w_4) and D = (w_5). In this example, $Cl_a^{\leqq}$ and $Cl_g^{\geqq}$ are preference-ordered classes, whereas, w_1, w_2, w_3, and w_4 are the criteria. We have that

- corresponding to w_1, "high" is the prefer to "medium" followed by "low",
- corresponding to w_2, "good" is the prefer to "average" followed by "bad",
- corresponding to w_3, "big" is the prefer to "medium" followed by "small",
- corresponding to w_4, "light" is the prefer to "dark",
- corresponding to w_5, "good" is the prefer to "average".

From Table 1, we compute the lower and upper approximation and boundaries of classes $Cl_a^{\leqq}$ and $Cl_g^{\geqq}$, i.e., evaluate $\underline{P}(Cl_g^{\geqq})$. By using definition, we have $\underline{P}\left(Cl_g^{\geqq}\right) = \{C \in U: \nabla_{\mathrm{P}}^+(C) \subseteq Cl_g^{\geqq}\}$. From table 1, we attain $Cl_g^{\geqq}$ = $\{\mathrm{C}_1,\ \mathrm{C}_2,\ \mathrm{C}_4,\ \mathrm{C}_6\}$ Applying the P-dominating sets specified in Table 2, $\underline{P}\left(Cl_g^{\geqq}\right)$= $\{\{\mathrm{C}_2,\ \mathrm{C}_4\}\}\}$, since $\nabla_{\mathrm{P}}^+(C_j) \subseteq Cl_g^{\geqq}$, j = 2, 4. Analogously, we obtain other lower and upper approximation classes described in Table 3.

Table 1. An example of an information table for car purchasing

Car	Cost (w_1)	Mileage (w_2)	Space (w_3)	Tint (w_4)	Overall performance (w_5)
C_1	medium	good	medium	dark	good
C_2	high	average	big	light	good
C_3	medium	good	medium	dark	average
C_4	high	average	big	light	good
C_5	medium	bad	small	dark	average
C_6	low	average	small	light	good
C_7	medium	average	big	light	average

Table 2. P-dominating and P-dominated sets

Objects (cars)	P-dominating set	P-dominated set
C_1	$\nabla_P^+(C_1) = \{C_1, C_3\}$	$\nabla_P^-(C_1) = \{C_1, C_3, C_5\}$
C_2	$\nabla_P^+(C_2) = \{C_2, C_4\}$	$\nabla_P^-(C_2) = \{C_2, C_4, C_5, C_6, C_7\}$
C_3	$\nabla_P^+(C_3) = \{C_1, C_3\}$	$\nabla_P^-(C_3) = \{C_1, C_3, C_5\}$
C_4	$\nabla_P^+(C_4) = \{C_2, C_4\}$	$\nabla_P^-(C_4) = \{C_2, C_4, C_5, C_6, C_7\}$
C_5	$\nabla_P^+(C_5) = \{C_1, C_2, C_3, C_4, C_5, C_7\}$	$\nabla_P^-(C_5) = \{C_5\}$
C_6	$\nabla_P^+(C_6) = \{C_2, C_4, C_6, C_7\}$	$\nabla_P^-(C_6) = \{C_6\}$
C_7	$\nabla_P^+(C_7) = \{C_2, C_4, C_7\}$	$\nabla_P^-(C_7) = \{C_5, C_6, C_7\}$

Table 3. Approximation set and boundary regions

Lower approximations	Upper approximations	Boundary
$\underline{P}(Cl_a^{\leq})=\{C_5\}$	$\overline{P}(Cl_a^{\leq}) = \{C_1, C_3, C_5, C_6, C_7\}$	$Bnd_P(Cl_a^{\leq}) = \{C_1, C_3, C_6, C_7\}$
$\underline{P}(Cl_g^{\geq})=\{C_2, C_4\}$	$\overline{P}(Cl_g^{\geq}) = \{C_1, C_2, C_3, C_4, C_6, C_7\}$	$Bnd_P(Cl_g^{\geq})=\{C_1, C_3, C_6, C_7\}$

Accuracy of approximation and quality of approximation in DRST (Slowinski et al., 2000; Greco et al., 2001):

Accuracy of approximation in dominance based rough set approach of upward unions of class $(Cl_q^{\geq})$and downward unions of class $(Cl_q^{\leq})$ are expressed as the ratio:

$$\alpha_P(Cl_q^{\geq}) = \frac{|\underline{P}(Cl_q^{\geq})|}{|\overline{P}(Cl_q^{\geq})|}, \qquad \alpha_P(Cl_q^{\leq}) = \frac{|\underline{P}(Cl_q^{\leq})|}{|\overline{P}(Cl_q^{\leq})|}, \tag{1}$$

The quality of classification (quality of shorting) in DRST can be expressed by ratio measure of partition Cl employing a basis set P:

$$\gamma_P(Cl) = \frac{|U-((\cup_{q\in Q} Bnd_P(Cl_q^{\geq}))\cup(\cup_{q\in Q} Bnd_P(Cl_q^{\leq})))|}{|U|}. \tag{2}$$

A minimal subset P is known as the reduct of Cl, such that $\gamma_P(Cl) = \gamma_C(Cl)$ and represented by $RED_{Cl}(P)$. Moreover, an information table possess one or more reduct. The intersection set of the respective reducts of Cl is known as the core and is represented as $CORE_{Cl}$.

From information Table 1, using Equations (1) and (2), the accuracy of approximation is 0.200 for the decision class at most “average” and 0.3300 for the decision class at least “good”, whereas the quality of classification is equal to 0.4300. According to this example, it has two reduct sets, i.e. {cost, mileage} and {cost, space, tint} and one core attribute {cost}, i.e., $Core_{Cl}(C)$ = {cost}.

Decision rules:

In multi-criteria decision making by DRST (Greco *et al.*, 1998; Greco *et al.*, 2001), condition attributes with preference-ordered scales should be according to increasing and decreasing preferences, and such condition attributes are known as criteria. Slowinski *et al.* (2000) first proposed the concept of decision rules, which were derived based on the preference information contained in the decision table.

The generalized description of actions contained in the information table is induced by the union of the rough approximations of upward and downward classes. These descriptions of actions are represented in terms of “if …., then…..” decision rules.

We consider three decision rules, which are defined as follows.

1. $D_{\geq}$-decision rules, having the following form:
   If $g(j,r_1) \geq z_{r1}$ and $g(j,r_2) \geq z_{r2}$ and ……. $g(j,r_p) \geq z_{rp}$, then $\in Cl_q^{\geq}$,
   where $P = \{r_1, r_2, \dots\dots, r_p\} \subseteq C$, $(z_{r1}, z_{r2}, \dots\dots. z_{rp}) \in V_{r1} \times V_{r2} \times \dots\dots \times V_{rp}$ and $q \in Q$; these decision rules are assisted by objects that belongs to the P-lower approximation of $Cl_q^{\geq}$.
2. $D_{\leq}$-decision rules, having the following form:
   If $g(j,r_1) \leq z_{r1}$ and $g(j,r_2) \leq z_{r2}$ and ……. $g(j,r_p) \leq z_{rp}$, then $\in Cl_q^{\leq}$,
   where $P = \{r_1, r_2, \dots\dots, r_p\} \subseteq C$, $(z_{r1}, z_{r2}, \dots\dots. z_{rp}) \in V_{r1} \times V_{r2} \times \dots\dots \times V_{rp}$ and $q \in Q$; these decision rules are assisted by objects that belongs to the P-lower approximation of $Cl_q^{\leq}$.
3. $D_{\geq\leq}$-decision rules, having the following form:
   If $g(j,r_1) \geq z_{r1}$ and $g(j,r_2) \geq z_{r2}$ and... $g(j,r_n) \geq z_{rn}$ and $g(j,r_{n+1}) \leq z_{rn+1}$ and ….. $g(j,r_p) \leq z_{rp}$, then $j \in Cl_q \cup Cl_{q+1} \cup \dots \cup Cl_m$,

where $O' = \{r_1, r_2, \dots \dots, r_n\} \subseteq C$, $O" = \{r_{n+1}, \dots, r_p\} \subseteq C, P = O' \cup O", O'$ and $O"$ may not be disjoint, $(z_{r1}, z_{r2}, \dots \dots \dots z_{rp}) \in V_{r1} \times V_{r2} \times \dots \dots \times V_{rp}$, $m, q \in Q$ such that $q < m$; these decision rules assisted by objects that belongs to the boundary region of union of classes $Cl_q^{\leq}$ and $Cl_m^{\geq}$.

Suggested methodology:

In order to improve the overall experience for passengers and operational efficiency for the railways, this research will undertake a thorough exploration of the Indian Railways Reservation System. It will aim to identify existing technical and operational issues, assess user satisfaction, and propose potential improvements.

We have covered an Indian railway reservation system that uses crude set theory based on dominance to determine how to choose which passengers to reserve seats for. The purpose of this research is to provide criteria for deciding whether or not to book a reservation for a train ticket in India based on a number of other conditional attributes. In the sections that follow, we will examine each necessary stage of this method.

Research problem and data cluster:

Gaining an awareness of certain issues and establishing goals are the two most important things to do in order to glean valuable information. Reservation availability and other condition features were the focus of the present investigation of the Indian railways. Passenger reviews and data retrieved from the Indian Railways' official websites (https://www.irctc.co.in/nget/ and http://indiarailinfo.com) comprise all the relevant information. Prior to delving into all of the qualities, a thorough analysis was conducted on the crucial aspects pertaining to the Indian railway reservation system (condition and decision). Having familiarity with the reservation system is essential for making informed judgments when reserving train reservations for specific journeys. To guarantee the excellent quality of the forthcoming analysis, the acquired data has been processed accurately.

Dominance-based rough set approach examination and determination:

Several conditional features and a decisional attribute of the Indian railway reservation system are taken into account in this study. The twelve elements of the Indian railway reservation system have been taken into account as characteristics of the condition (Figure 1). The availability of train reservations, on the other hand, has been treated as a standalone decision attribute. The correctness of the approximation, reduction, and decision rules has also been computed using two rough set methodologies: the classical technique and the dominance-based rough set methodology. Additionally, RST-based passenger guidelines for choosing the best train for a given trip have been detailed. A collection of RST tools, 4eMka2, are used to analyze the accuracy of approximation of railways research data. A comparison study on classical and dominance rough set approaches is conducted using proper statistical analysis after describing rules and measuring accuracy of approximation.

Extracting information:

Domain specialists' mature judgement is required to review and inspect the qualified rules in order to confirm the analytical result's pragmatism and efficiency. Applying the rules to design strategies for a passenger's reservation ticket availability requires a thorough examination of any unforeseen events pertaining to the studied data set. It is necessary to periodically evaluate the regulations that were established for the Indian railways reservation scheme. Continuous surveillance and methodical review by domain experts is necessary to establish the validity of the derived rule base of a dynamic system like IRRS.

Machine learning models:

In this section we discuss two machine learning models namely, Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) which are used in our study.

Extreme Gradient Boosting (XGBoost), introduced by Chen and Guestrin (2016), is an enhanced version of the gradient boosting framework that emphasizes computational efficiency and predictive accuracy. It follows an ensemble learning approach in which models are trained sequentially, with each subsequent model aiming to reduce the residual errors of its predecessors. In this framework, decision trees serve as the base learners, and the training process revolves around optimizing a regularized objective function that integrates both a loss function and a complexity penalty. The objective function is formulated as:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k),$$

where $l(y_i, \hat{y}_i)$ represents a differentiable convex loss function (such as logistic loss for classification tasks), and $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2$ acts as a regularization component to control model complexity. Here, TTT denotes the number of leaves in a tree, and www corresponds to the vector of leaf weights.

XGBoost leverages a second-order Taylor expansion of the loss function to approximate the optimization process more effectively. It further incorporates strategies such as shrinkage (learning rate), feature subsampling, and early stopping to enhance model generalization. Its capacity to process sparse inputs, compatibility with parallel and distributed computation, and consistent performance across diverse structured datasets have established XGBoost as one of the most powerful and widely adopted machine learning algorithms.

Support Vector Machine (SVM), proposed by Cortes and Vapnik (1995), is a supervised machine learning technique widely employed for binary classification tasks, with extensions available for multi-class problems. The fundamental principle behind SVM is to identify the optimal hyperplane that separates data points of different classes in the feature space while maximizing the margin, which is the distance between the hyperplane and the closest data points of each class, referred to as support vectors.

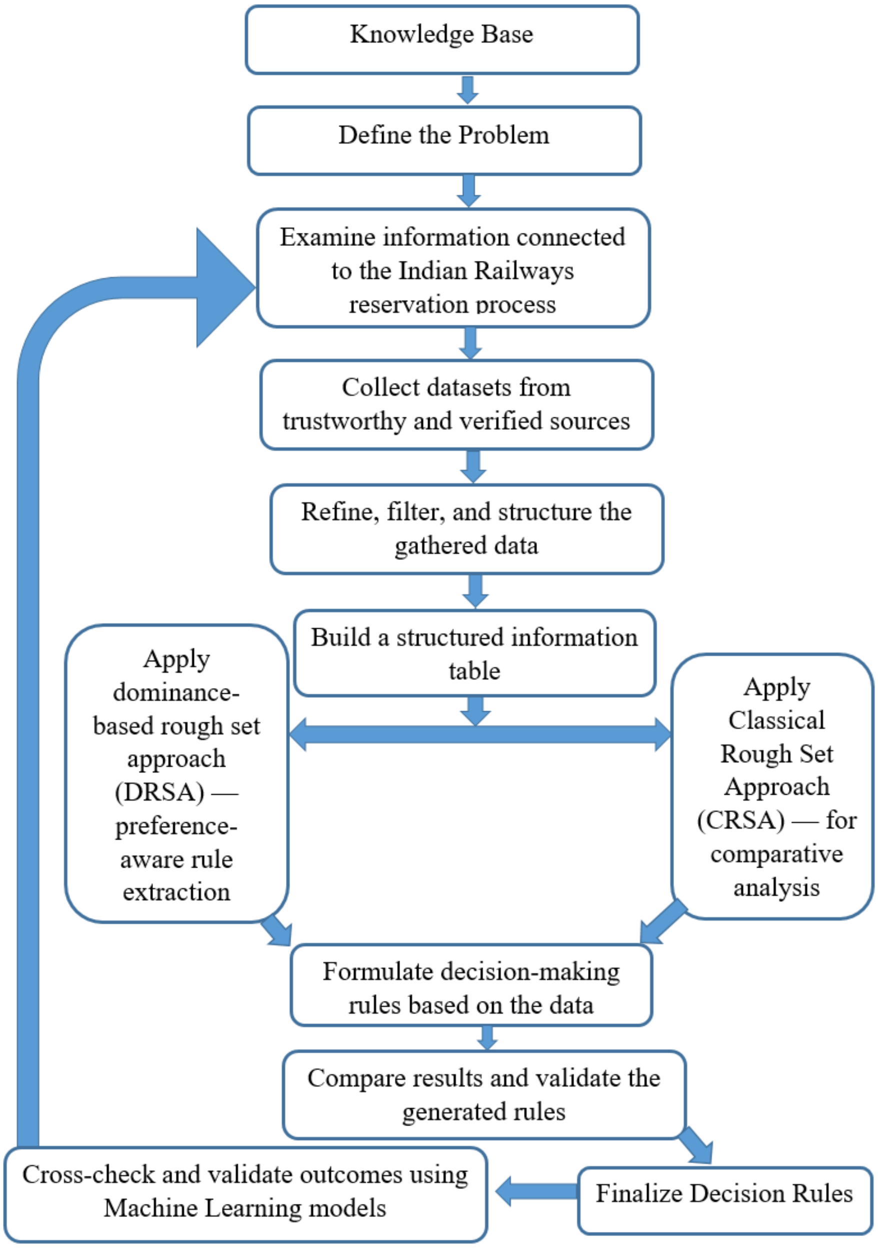

Figure 1. Analysis structure of RST approach for Indian railway reservation system

For linearly separable datasets, the optimization problem is expressed as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \; subject\ to\ y_i(w^T x_i + b) \geq 1, \forall i.$$

When the data are not perfectly separable, SVM employs slack variables ξ_i along with a regularization parameter C, resulting in the soft-margin formulation:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \; C \sum_{i=1}^{n} \xi_i \; subject\ to\ y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0.$$

This formulation balances maximizing the margin with minimizing classification errors, thereby enhancing the model's robustness when handling noisy or overlapping data.

To handle non-linear decision boundaries, SVM applies the *kernel trick*, which transforms the input features into a higher-dimensional space without explicitly computing the transformation. Popular kernels include the Radial Basis Function (RBF), polynomial, and sigmoid kernels. SVMs are highly effective in high-dimensional spaces and are known for their robustness to overfitting, especially when the dataset is clean and the number of features is large. However, they can be computationally expensive and sensitive to the choice of kernel and parameters.

3. Case study

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

Research problem:

The Indian Railways (IR) serves as the backbone of the nation's transport infrastructure and is recognized as Asia's largest and the world's second-largest railway network administered under a single authority. Functioning under the Ministry of Railways, Government of India (as of 2025), it plays a pivotal role in linking communities across the vast expanse of the country. The origins of rail transport in India trace back to 1853, when the first train journey was undertaken between Mumbai and Thane, a landmark event in the nation's history. Subsequently, in 1951, the numerous regional railways were unified and brought under government ownership, forming what is now known as Indian Railways. Over time, IR has expanded into a highly intricate system that not only facilitates passenger mobility but also strengthens commerce and economic development. To improve passenger amenities and services, the Indian Railway Catering and Tourism Corporation (IRCTC) was later established, which manages catering services, online reservations, and tourism packages. Today, Indian Railways stands as a symbol of national integration, affordability, and accessibility, serving millions of passengers daily. To book and manage train tickets across India's extensive railway network, one can use the Indian Railways Reservation System, a sophisticated and all-encompassing system. One of the biggest railway networks in the world, the Indian Railways, manages the reservation system through which customers can purchase tickets for various types of trains. Online ticket booking is available to passengers via the Indian Railways website and other authorized third-party platforms. Travelers now have an easy and accessible way to arrange their trips. When it comes to handling the massive volume of passengers using the railway network, the Indian Railways Reservation System is indispensable. It provides a simple and effective way to plan and schedule train trips all throughout the nation. This study's overarching goal is to investigate and resolve all problems with the Indian Railways Reservation System.

For the present study, data were collected through consultations with railway domain experts from IR and IRCTC, along with feedback provided by passengers. The analysis focuses on 26 major trains operating from Delhi Railway Station. Passenger choices are influenced by decision attributes, which are determined by a set of conditional factors. In this context, the study employs IRRS variables as the key determinants of decision-making. These variables include: (i) “Train setting-out time”, “Journey duration”, “Schedule days”, “Promptness of train”, “Ongoing seat status”, “Ticket expense”, “Travel span” “Travel span”, “Standard speed”, “Hygiene”, “Food grade”, “Railroad photography” and “Railway security”. A detailed description of these criteria and their relevance is presented in Table 4.

Table 4. Ticket availability issues in relation to IRRS factors and associated variables

Criteria	Variable	Explanation
Conditional criteria	train setting-out time	the scheduled time at which a train is set to leave or depart from a station
	journey duration	the total duration taken for a journey or trip from the starting point to the destination
	schedule days	the specific days of train on which service, operation, or activity is in progress or available
	promptness of train	the adherence of a train to its scheduled arrival and departure times, reflecting its timely and reliable performance
	ongoing seat status	the real-time and up-to-date information regarding the availability and reservation status of seats or tickets
	ticket expense	the monetary cost associated with purchasing a ticket for travel
	travel span	the spatial extent or length between the starting point and the endpoint of a journey or route
	standard speed	the overall rate of speed
	hygiene	conditions that promote health, cleanliness, and the prevention of disease by maintaining a clean and sanitary environment
	food grade	products that meet the necessary standards and regulations for safe and suitable use in food processing, handling, and consumption
	railroad photography	hobby or activity of observing, photographing, and documenting trains.
	railway security	the measures and protocols implemented to ensure the secure and risk-free operation of trains, safeguarding passengers
Decision criteria	reservation availability	accessibility of tickets or seats for a particular train

The data that was collected has already been pre-processed so that it may be analyzed and useful insights can be extracted. With some help and guidance from IR experts, the numeric values of reservation system have been transformed into a linguistic format. Such as, conditional criteria, "Train setting-out time" is categorized into three categories; if it is less than 15:00 hour, the time is categories as "day time"; if the Train setting-out time is between 15:00-20:00 hour, the time is categorized as "evening time"; if the Train setting-out time is between 20:00-04:00 hour, the time is categorized as "night train setting-out time". Conditional criteria, "journey duration" is divided into three subgroups; if the journey duration is less than 18 hours, the time duration is identified as "very less duration"; if the journey duration is between 18-24 hour, the time duration becomes "less Journey duration"; if the Journey duration is more than 24 hour, the time duration is characterized by "more journey duration". Conditional criteria, "travel span" is classified into two subsets; if the travel span is less than 1500 kilometre (km), the distance cover is considered as "far travel span"; if the travel span is more than 1500 km, the distance cover is considered as "very far travel span". Conditional criteria, "Ticket expense" is divided into three subgroups; if the ticket expense is less than 1600 Rupees (Rs.), the money expenditure is grouped as "low ticket expense"; if the ticket expense is between 1600 (Rs.)-2500 (Rs.), the money expenditure is determined as "medium ticket expense"; if the fare of ticket is more than 2500 (Rs), the money expenditure is identified as "high ticket expense". Conditional criteria, "Standard speed" is classified into three subcategories; if the standard speed is less than 55 km/hour (h), the train speed was classified as "average standard speed"; if the standard speed is between 55 km/h -70 km/h, the standard speed is classified as "fast standard speed"; if the standard speed is more than 70 km/h, the speed is classified as "very fast standard speed".

Approximation by DRSA:

We have used the 4eMka2 software (Poznan University of Technology, 2006a) to reduce the attributes and create decision rules (minimal cover). Every attribute in Table 5 consists of two specific preference values, "Gain" and "None". If the preference of an attribute is set to "Gain" then attribute values of that attribute will follow a preference order relation. If the attribute values do not follow any preference relation, then the attribute value is set to "None".

In our case study, the decision criteria domain *TA* is {poor, average, good, and excellent}. With the decision criteria *TA*, trains are distributed into the following preference-ordered classes: Cl_p = {poor}, Cl_a = {average}, Cl_g ={good}, Cl_e ={excellent}. Consequently, in this example approximate union of classes are:

- $Cl_p^{\leq}$ ← The class of train ticket availability corresponding to at most "poor",

- $Cl_a^{\leq}$ ← The class of train ticket availability corresponding at most "average",
- $Cl_a^{\geq}$ ← The class of train ticket availability corresponding at least "average",
- $Cl_g^{\leq}$ ← The class of train ticket availability corresponding at most "good",
- $Cl_g^{\geq}$ ← The class of train ticket availability corresponding at least "good",
- $Cl_e^{\geq}$ ← The class of train ticket availability corresponding at least "excellent".

Table 5. Description of criteria/attributes linked with the railway dataset

Name of criteria/attribute	Value of criterion/attributes	Value set	Preference
Condition attributes			
Ongoing seat status	WL, RAC, CNF	{WL, RAC, CNF}	gain
Train setting-out time	between 4:01–15:00; between 15:01–20:00; 20:01–4:00	{day, evening, night}	none
Journey duration	less than 18:00 hours; between 18:00–24:00 hours; more than 24:00 hours	{more, less, very less}	gain
Travel span	below1380 (kilometre); above 1380 (kilometre)	{very far, far}	gain
Ticket expense	rs. 2500 and above; rs. 1600-2500; less than rs 1600	{high, medium, low}	gain
Schedule days	weekly; biweekly; frequently; daily	{weekly, biweekly, frequently, daily}	none
Standard speed	below 55; 55-70; 71 and above	{average, fast, very fast}	gain
Hygiene	average; good; excellent	{average, good, excellent}	gain
Promptness of the train	poor; average; good; excellent	{poor, average, good, excellent}	gain
Food grade	poor, average; good; excellent	{poor, average, good, excellent}	gain
Railroad photography	poor; good; excellent	{poor, good, excellent}	gain
Railway security	average; good; excellent	{average, good, excellent}	gain
Decision criterion			
Reservation availability	poor; average; good; excellent	{poor, average, good, excellent}	gain

Table 6. Approximation sets and accuracy of approximation using DRSA

	$(Cl_p^{\leq})$	$(Cl_a^{\leq})$	$(Cl_g^{\leq})$	$(Cl_a^{\geq})$	$(Cl_g^{\geq})$	$(Cl_e^{\geq})$
Lower approximation	1	4	19	25	17	2
Upper approximation	1	9	24	25	22	7
Boundary	0	5	5	0	5	5
Accuracy of approximation	1	0.444	0.790	1	0.770	0.290
The quality of lower approximation: 0.7300						

Decision making using DRSA:

In this paper, we have used the concept of deriving a decision rule from the information in Table 6, which contains preferential information regarding the Indian railways reservation system for the dominance-based rough set approach. The minimal decision rules set obtained from the data are as listed in Table 7.

Table 7 displays the 18 minimum cover rules derived from the railway's dataset. An IF-THEN statement is the most common way to write some decision rules. Here are some examples that show how IF-THEN rules function; IF the departure of the train is at night, AND the Schedule days are daily, AND the Ticket expense is low and higher, THEN the decision criteria for ticket Reservation availability will be excellent.

From Table 7, it is clear that if the Railway security of the train is good or excellent, then ticket Reservation availability will be at least average, with a maximum support of 22 (cf. rule 8). This means Railway security is an important factor for passengers' decisions. In India, passengers prefer trains that have good Railway security. If the Standard speed of trains is very fast and they run biweekly, then the decision criteria will be good or excellent, and their support is 9 (cf. rule 3) and 7 (cf. rule 4), respectively. From rule 14, if the trains run biweekly, then ticket Reservation availability can be either good, average or poor. These decision rules indicate that Railway security, Standard speed and Schedule days are important attributes for passenger decisions. This suggests that most of the passengers reserved their tickets based on the train's Railway security, Standard speed and running days.

Table 7. Exact decision rules of the railway information system

Rule No.	Condition parts	Decision parts	Support
1	If (Ongoing seat status ≥ CNF) & (Hygiene ≥ excellent)	Then (Reservation availability ≥ excellent)	1
2	If (Train setting-out time = night) & (Schedule days= daily) & (Ticket expense≥ low)	Then (Reservation availability ≥ excellent)	1
3	If (speed ≥ very fast)	Then (Reservation availability ≥ good)	9
4	If (Schedule days= biweekly)	Then (Reservation availability ≥ good)	7
5	If (Ongoing seat status ≥ CNF) & (Schedule days= frequently)	Then (Reservation availability ≥ good)	1
6	If (Train setting-out time = night) & (Schedule days= daily)	Then (Reservation availability ≥ good)	2
7	If (Hygiene ≥ excellent) & (Schedule days= daily)	Then (Reservation availability ≥ good)	3
8	If (Railway security ≥ good)	Then (Reservation availability ≥ average)	22
9	If (Train setting-out time = day) & (Schedule days= weekly)	Then (Reservation availability ≥ average)	3
10	If (Train setting-out time = evening) & (Railway security ≤ average)	Then (Reservation availability ≤ poor)	1
11	If (Ongoing seat status ≤ WL) & (Schedule days= weekly)	Then (Reservation availability ≤ average)	2
12	If (Train setting-out time = night) & (Schedule days= frequently)	Then (Reservation availability ≤ average)	1
13	If (Train setting-out time = evening) & (Standard speed ≤ fast)	Then (Reservation availability ≤ average)	2
14	If (Schedule days = biweekly)	Then (Reservation availability ≤ good)	7
15	If (Schedule days = frequently)	Then (Reservation availability ≤ good)	4
16	If (Schedule days= daily) & (Train setting-out time = day)	Then (Reservation availability ≤ good)	3
17	If (Standard speed≤ average) & (Ticket expense≤ medium)	Then (Reservation availability ≤ good)	1
18	If (Train setting-out time = evening) & (food grade ≤ average)	Then (Reservation availability ≤ good)	2

4. Result and discussion

The reservation system for trains is inherently dynamic because the status of a train's seat (CONF, RAC, WL) changes regularly. The imprecision of the attributes processed utilizing rough set techniques is the fundamental cause of the data's dynamic nature. To help passengers with their train berth reservations, we have used classical and dominance-based rough set methods to analyze the data.

We use the following dataset in Table 8 of trains and consider the accuracy of approximation using CRSA and DRSA. It is observed that the accuracy of approximation of the data related to the train reservation system is most of the time lower when the dominance-based rough set approach is used, as compared to the classical rough set approaches. This implies that the dominance-based approach can handle more imprecision of data compared to the classical-based approach. Since in reality, for a data set which is massively dynamic in nature, its corresponding accuracy of approximation should not be as high as we have obtained while using CRSA (software, ROSE2).

Table 8. Accuracy of approximation

Train No.	CRSA				DRSA					
	Poor	Average	Good	Excellent	At most poor	At most average	At least average	At most good	At least good	At least excellent
100	1	0.8529	0.8000	0.8095	1	0.8200	1	0.9500	0.8800	0.8100
125	1	0.8214	0.8571	0.8519	0.8800	0.8400	0.9800	0.9400	0.9200	0.7900
150	1	0.8750	0.8767	0.8621	0.8900	0.8800	0.9900	0.9400	0.9300	0.7400
175	1	0.8529	0.9063	0.8788	0.8300	0.8400	0.9700	0.9600	0.9300	0.8200
200	1	0.9091	0.9043	0.8947	0.9100	0.9000	0.9900	0.9600	0.9400	0.8200
225	1	0.9074	0.9204	0.9111	0.8300	0.8500	0.9800	0.9600	0.9200	0.8300
250	1	0.9180	0.9291	0.9091	0.8000	0.9000	0.9700	0.9800	0.9500	0.9100
275	1	0.9306	0.9291	0.9149	0.8300	0.8800	0.9700	0.9700	0.9300	0.8400
300	1	0.9219	0.9416	0.9231	0.8600	0.9100	0.9800	0.9800	0.9500	0.8900
325	1	0.9419	0.9464	0.9167	0.8100	0.9200	0.9800	0.9700	0.9500	0.8400
350	1	0.9457	0.9503	0.9273	0.8500	0.8800	0.9800	0.9600	0.9300	0.8100
375	1	0.9405	0.9505	0.9437	0.8600	0.8800	0.9800	0.9600	0.9300	0.8500
400	1	0.9500	0.9579	0.9310	0.9000	0.9100	0.9900	0.9700	0.9500	0.8100
425	1	0.9515	0.9577	0.9545	0.8800	0.8600	0.9900	0.9700	0.9300	0.8700
450	1	0.9505	0.9581	0.9500	0.7400	0.7800	0.9500	0.9400	0.8700	0.7500
475	1	0.9576	0.9615	0.9529	0.7000	0.8900	0.9600	0.9700	0.9400	0.8700
500	1	0.9580	0.9641	0.9437	0.7000	0.8600	0.9500	0.9700	0.9200	0.8000
525	1	0.9554	0.9658	0.9596	0.7100	0.8300	0.9600	0.9600	0.9200	0.8400
550	1	0.9632	0.9669	0.9592	0.7000	0.8500	0.9600	0.9400	0.9200	0.7400
575	1	0.9677	0.9667	0.9604	0.7400	0.8800	0.9700	0.9700	0.9300	0.8500
600	1	0.9669	0.9674	0.9652	0.7100	0.8200	0.9600	0.9300	0.9000	0.7400
625	1	0.9653	0.9713	0.9579	0.7900	0.8300	0.9700	0.9400	0.9000	0.7200
650	1	0.9701	0.9697	0.9661	0.6900	0.8100	0.9500	0.9200	0.8800	0.6900
675	1	0.9693	0.9734	0.9636	0.5500	0.7800	0.9300	0.9500	0.8700	0.7700
700	1	0.9682	0.9740	0.9658	0.7700	0.7800	0.9600	0.9400	0.8700	0.7100
725	1	0.9689	0.9741	0.9704	0.6100	0.7900	0.9300	0.9500	0.8800	0.7800
750	1	0.9725	0.9752	0.9699	0.7600	0.8400	0.9700	0.9300	0.9100	0.7200
775	1	0.9728	0.9763	0.9706	0.7200	0.8400	0.9600	0.9500	0.9100	0.7600
800	1	0.9711	0.9772	0.9730	0.600	0.8400	0.9400	0.9400	0.9200	0.7400
825	1	0.9714	0.9783	0.9732	0.7200	0.8300	0.9600	0.9400	0.9100	0.7400
850	1	0.9756	0.9777	0.9771	0.6800	0.8600	0.9600	0.9500	0.9300	0.8300
875	1	0.9746	0.9795	0.9726	0.6000	0.8500	0.9400	0.9700	0.9200	0.8600
900	1	0.9744	0.9795	0.9748	0.6700	0.8400	0.9400	0.9500	0.9100	0.8000
925	1	0.9755	0.9811	0.9724	0.6300	0.8200	0.9400	0.9600	0.9000	0.7900
950	1	0.9763	0.9813	0.9730	0.6900	0.8100	0.9500	0.9500	0.9000	0.7700
975	1	0.9747	0.9816	0.9760	0.6500	0.8000	0.9300	0.9400	0.8900	0.7500
1000	1	0.9773	0.9815	0.9793	0.5900	0.7600	0.9400	0.9200	0.8800	0.7200

The graphical illustration of the experiment is elucidated in six figures. Figure 2 and Figure 3 show the accuracy of the approximation of the ticket Reservation availability situation of passengers in different trains by using the classical and dominance-based rough set approach. Figure 4 and Figure 5 depict the corresponding box plots for the accuracy of approximation related to ticket Reservation availability in trains by the passenger.

Figure 4 and Figure 5 give an impression of the dispersion of data, whereby the interquartile range (IQR), around the median, is less for the dominance-based approach when compared to the classical rough set approach. In Figure 5, we have also observed that there are no outliers in the boxplot. This means the accuracy of approximation is quite consistent for all the trains when the dominance-based approach is used. Figure 6 and Figure 7 show a normal probability plot (also called a normal quantile-quantile plot) to decide whether it is reasonable to consider the accuracy measure derived using CRSA and DRSA of the railway dataset, which is sampled from a population that follows a normal distribution.

The case study presented in this article analyses the decision taken by the passengers (decision makers) regarding the reservation of tickets for different trains. These decisions are influenced by the decision rules, which are generated from the decision table using CRSA and DRSA. By applying both the rough set approaches, it has been observed that in the case of DRSA, the decision of the passengers is much easier compared to their decision when the classical-based rough set approach is considered.

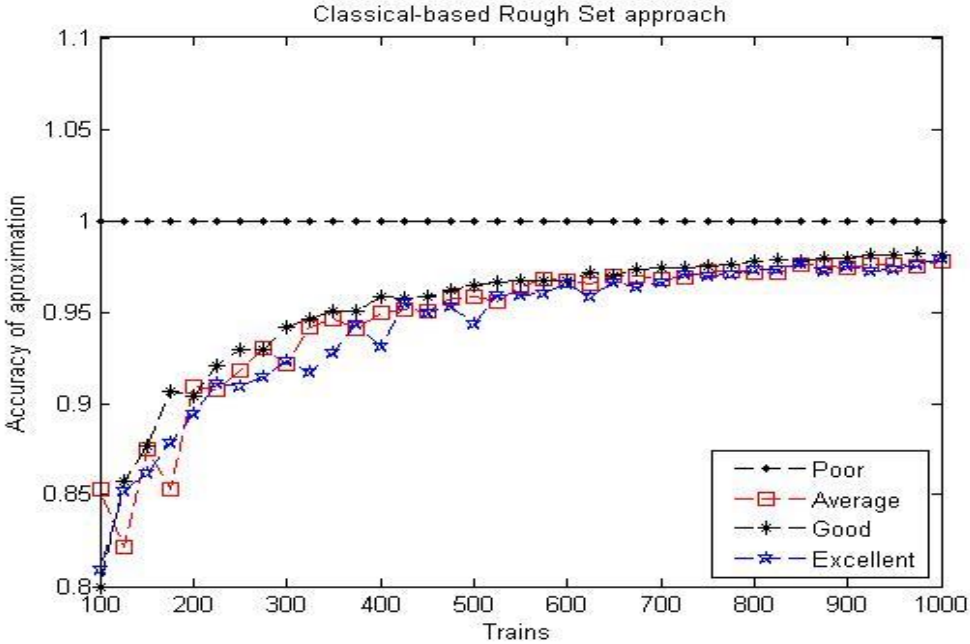

Figure 2. Accuracy of approximation for railway reservation system by CRSA

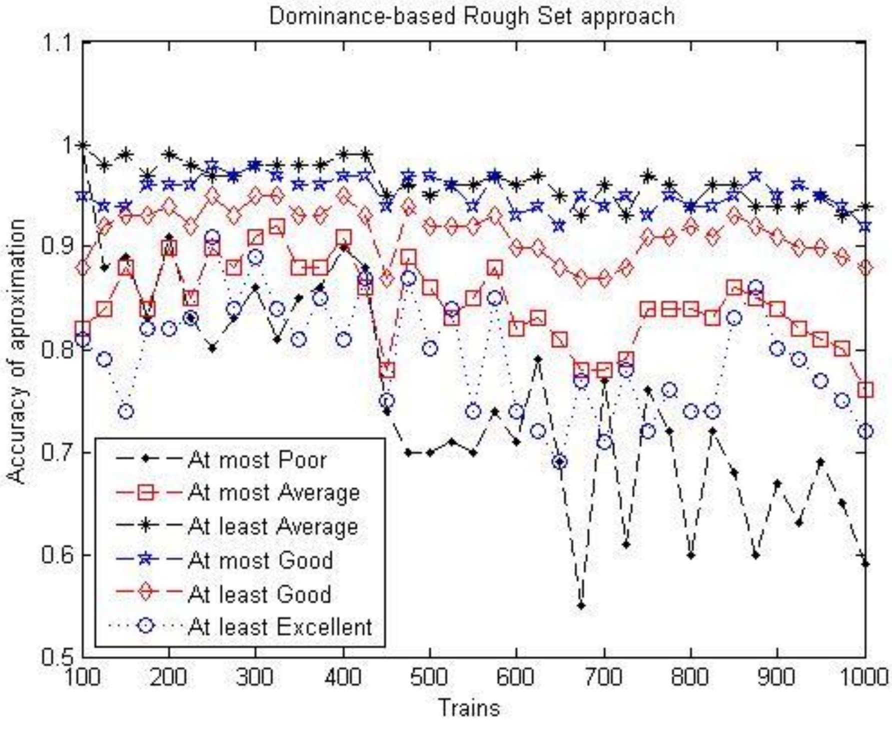

Figure 3. Accuracy of approximation for railway reservation system by DRSA

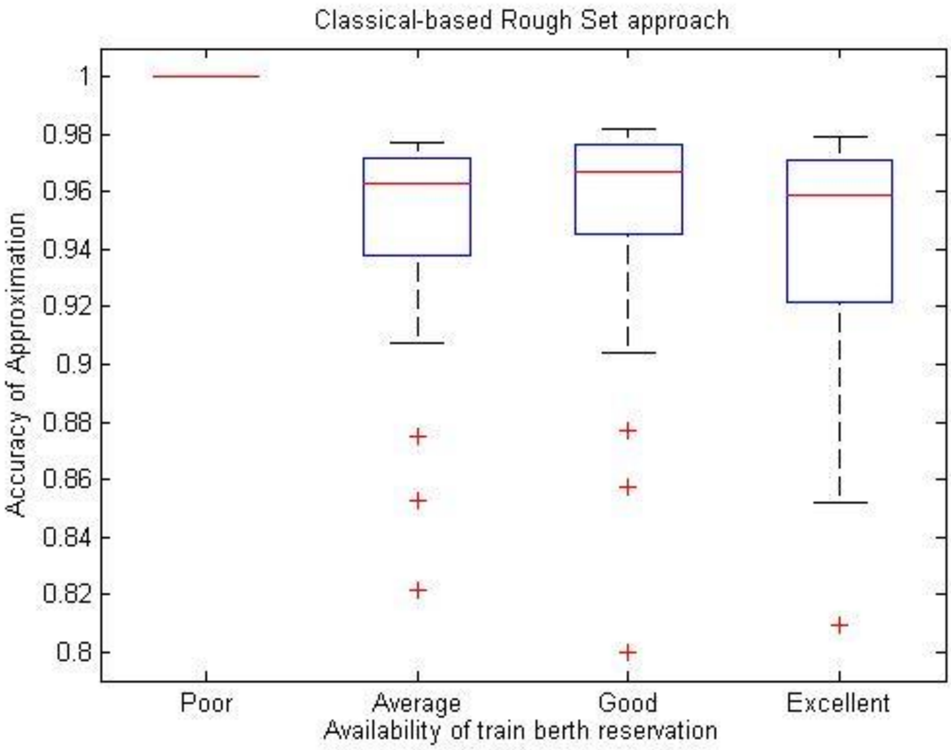

Figure 4. Box plot of accuracy of approximation for the railway reservation system by CRSA

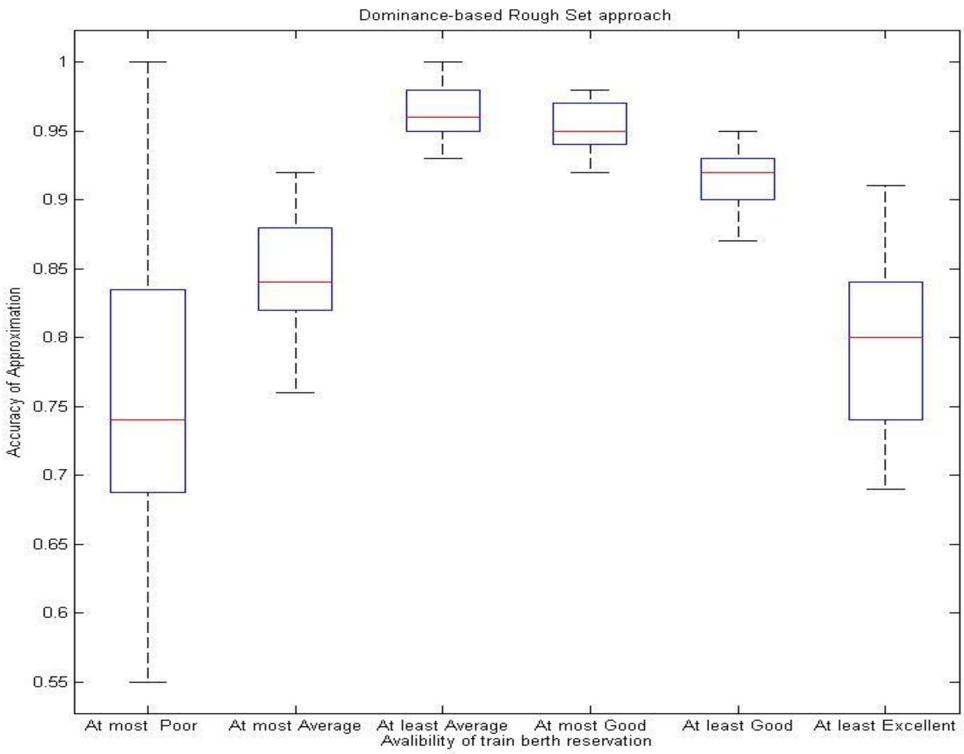

Figure 5. Box plot of accuracy of approximation for railway reservation system by DRSA

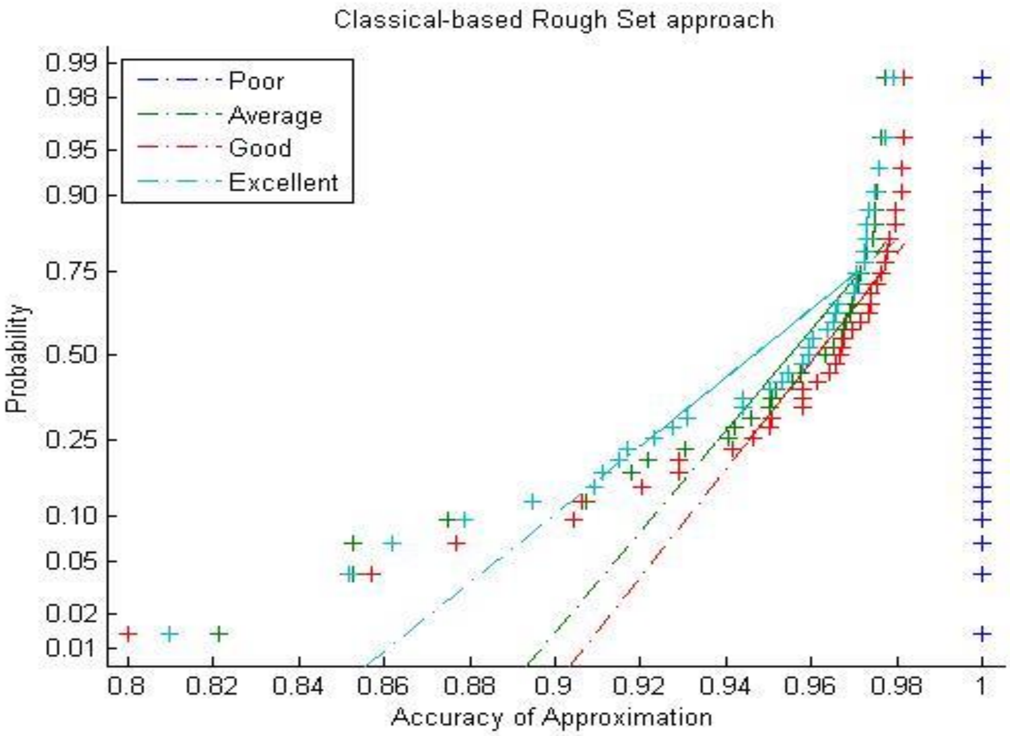

Figure 6. Normal probability plot for the railway reservation system by CRSA

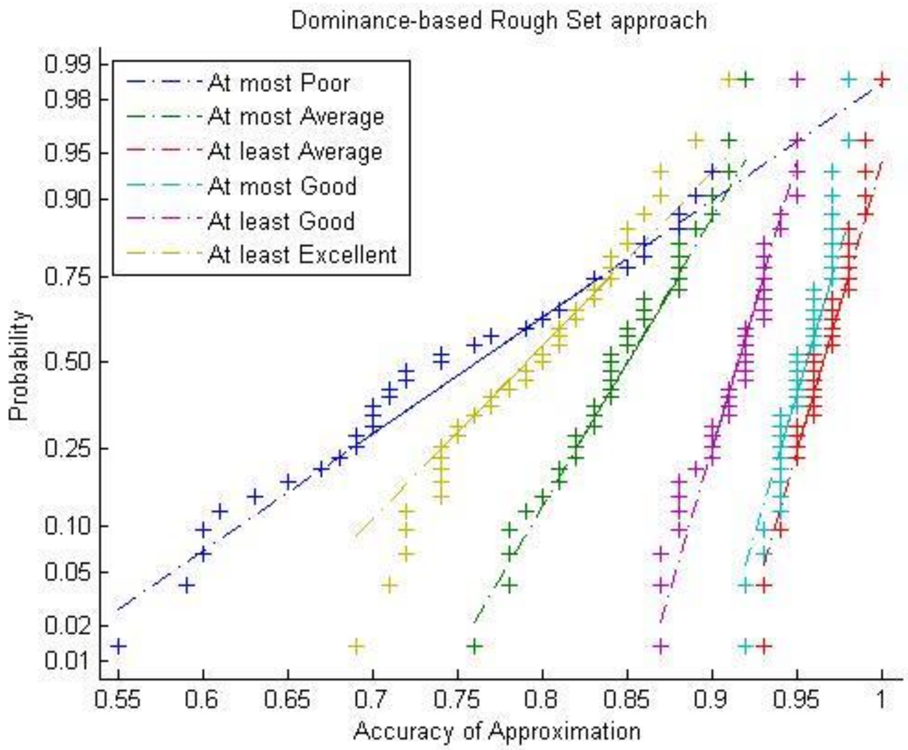

Figure 7. Normal probability plot for the railway reservation system by DRSA

We have considered a large-scale data set to show the efficiency of the proposed data mining technique for the number of trains ranging from 100 through 1000 using Matlab R2013b, ROSE2 and 4eMka2. The data of all the conditional attributes are generated randomly for all the trains. The rules are generated considering the tuples of the conditional attributes. Furthermore, in this section we discuss two machine learning models namely, Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) which are used to predict the availability of the reservation of the trains in the dataset. Subsequently, we compare the performance of the two classifiers on the dataset by considering the 10-fold cross-validation and have observed that for all the four performance measures, i.e., accuracy, precision, recall and F1-score, SVM emerges as the better classifier. The corresponding result is reported in Table 9, where the superior values are highlighted as bold. Her it is to be mentioned that, we have used ANACONDA distribution 2025.06 for measuring the performance metrics of the machine learning estimators.

Table 9. Comparisons of the performance metrics for XGBoost and SVM classifiers

Performance metrics	XGBoost	SVM
Accuracy	0.75	**0.89**
Precision	0.77	**0.81**
Recall	0.73	**0.75**
F1-score	0.72	**0.78**

In addition, the common test platform used for the simulation purpose of DRS includes Intel(R) Core(TM) 2 Duo CPUE8400@ 3.00 GHz, 3003 MHz's, 2 Core(s), 2 Logical Processor(s) and 4.00 GB RAM.

5. Conclusions and future scope

This study has presented a Dominance-based Rough Set Approach (DRSA) to support decision-making in selecting the most suitable train for a journey, based on multiple significant criteria. In the proposed framework, a passenger – acting as the decision-maker – evaluates various conditional attributes associated with Indian Railways to determine whether or not to reserve a berth on a particular train. The DRSA-based model offers a systematic and explainable method to identify the most appropriate train among multiple alternatives.

The results demonstrate the practical utility of this model for both passengers and railway authorities. Specifically, it enables: (i) identification of key criteria requiring service improvement, (ii) strategic planning to help maximize Indian Railways' profitability, (iii) detection of high-traffic routes requiring infrastructure or scheduling adjustments, and (iv) assessment of the impact of passenger feedback on service enhancements, using optimization models to quantify the benefits of improved travel conditions.

A comprehensive comparative analysis was conducted on real-world Indian Railways data, focusing on ticket availability and passenger preferences. In addition to DRSA, two machine learning models – XGBoost and Support Vector Machines (SVM) – were applied to validate the findings and enhance the robustness of the analysis. Attribute reduction techniques further helped identify the most influential factors affecting passengers' reservation decisions. Unlike traditional methods, the proposed DRSA model provides an interpretable and adaptable approach to help passengers make informed travel choices, while also offering actionable insights for railway service planning.

Looking ahead, we plan to extend this data-driven approach into a full-fledged decision support system that can be utilized by Indian Railways and its stakeholders. One of our key goals is to translate decision rules into natural language for better user interpretability. Additionally, we aim to integrate appropriate metrics to refine the decision-making process by eliminating redundant or contradictory rules. Such advancements will not only enhance the model's usability but also contribute meaningfully to the ongoing improvement of passenger services and operational efficiency in Indian Railways.

Declaration of Generative AI and AI-assisted technologies in the writing process:

- During the preparation of this manuscript the author(s) did not used Generative AI and AI-assisted technologies and take(s) full responsibility for this declaration.

References

1. Augeri, M. G., Cozzo, P., Greco, S. (2015) Dominance based rough set approach: An application case study for setting speed limits for vehicles in speed controlled zones. *Knowledge Based Systems*, 89, 288–300. DOI: 10.1007/11681960_3.

2. Chakhar, S., Ishizaka, A., Labib, A., Saad, I. (2016) Dominance-based rough set approach for group decisions. *European Journal of Operational Research*, 251, 206–224. DOI:10.1016/j.ejor.2015.10.060.
3. Chakhar, S., Saad, I. (2012) Dominance based rough set approach for groups in multicriteria classification problem. *Decision Support Systems*, 54, 372–380. DOI:10.1016/j.ejor.2015.10.060.
4. Chang, Y., Zhu, X., Yan, B., Wang, L. (2017) Integrated scheduling of handling operations in railway container terminals. *Transportation Letters*, 11(4), 402–412.
5. Chen, D., Ni, S., Xu, C., Jiang, X. (2019) Optimizing the draft passenger train timetable based on node importance in a railway network. *Transportation Letters,* 11(1), 20–32. DOI:10.1080/19427867.2016.1271523.
6. Chen, J., Zhu, P. (2024) Feature selection of dominance-based neighborhood rough set approach for processing hybrid ordered data. *International Journal of Approximate Reasoning*, 167, 109134. DOI:10.1016/j.ijar.2024.109134.
7. Chen, T., Guestrin, C. (2016) XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, August 2016. New York: Association for Computing Machinery, pp. 785–794.
8. Cortes, C., Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20(3), 273–297. DOI:10.1007/BF00994018.
9. Du, W. S., Hu, B. Q. (2016) Dominance-based rough set approach to incomplete ordered information systems. *Information Sciences*, 346–347, 106–129. DOI:10.1016/j.ins.2016.01.098.
10. Givoni, M., Chen, X. (2017) Airline and railway disintegration in China: The case of Shanghai Hongqiao Integrated Transport Hub. *Transportation Letters*, 9(4), 202–214. DOI:10.1080/19427867.2016.1252877.
11. Greco, S., Matarazzo, B., Slowinski, R. (1998) A new rough set approach to multicriteria and multiattribute classification. In: *Proceedings of Rough Sets and Current Trends in Computing Conference RSCTC*, Warsaw, June 1998. Berlin Heidelberg: Springer-Verlag, pp. 60–67.
12. Greco, S., Matarazzo, B., Slowinski, R. (1999) The use of rough sets and fuzzy sets in MCDM. *International Series in Operations Research & Management Science, 21.* DOI:10.1007/978-1-4615-5025-9_14.
13. Greco, S., Matarazzo, B., Slowinski, R. (2001) Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129, 1–47. DOI:10.1016/S0377-2217(00)00167-3.
14. Greco, S., Matarazzo, B., Slowinski, R., Stefanowski, J. (2001) An algorithm for induction of decision rules consistent with dominance principle. In: *Proceedings of Rough Sets and Current Trends in Computing Conference RSCTC*, Baniff, October 2000. Berlin Heidelberg: Springer, pp. 304-313.
15. Indian Railways. (2025) *Indian Railways Information System (IRIS)*. Available at: https://indiarailinfo.com/train.
16. Jiten, S., Gaurang, J., Purnima, P., Shriniwas, A. (2017) Effect of directional distribution on stairway capacity at a suburban railway station. *Transportation Letters*, 9(2), 70–80.
17. Li, S., Li, T., Zhang, Z., Chen, H., Zhan, J. (2015) Parallel computing of approximations in dominance-based rough set approach. *Knowledge Based Systems,* 87, 102–111.
18. Liou, J. H., Tzeng, G. H. (2010) A dominance-based rough set approach to customer behaviour in the airline market. *Information Sciences,* 180, 2230–2238. DOI:10.1016/j.ins.2010.01.025.
19. Majumder, S., Singh, A., Singh, A., Karpenko, M., Sharma, H. K., Mukhopadhyay, S. (2024) On the analytical study of the service quality of Indian Railways under soft-computing paradigm. *Transport*, 39(1), 54-63. DOI:10.3846/transport.2024.21385.
20. Nawaz, U., Saeed, Z., Atif, K. (2025) A novel framework for efficient dominance-based rough set approximations using K-dimensional (KD) tree partitioning and adaptive recalculations techniques. *Engineering Applications of Artificial Intelligence*, 154, 110993. DOI:10.1016/j.engappai.2025.110993
21. Nosheen, F., Qamar, U., Raza, M. S. (2022) A parallel rule-based approach to compute rough approximations of dominance based rough set theory. *Engineering Applications of Artificial Intelligence*, 115, 105285. DOI:10.1016/j.engappai.2022.105285.
22. Pawlak, Z. (1982) Rough sets. *International Journal of Computer and Information Science*, 11, 341–356.
23. Poznan University of Technology. (2006a) *4eMka2 software*. Available at: https://fcds.cs.put.poznan.pl/IDSS/software/4emka2.htm.
24. Poznan University of Technology. (2006b) *ROSE2 software*. Available at: https://fcds.cs.put.poznan.pl/IDSS/software/rose.htm.
25. Sawicki, P., Zak, J. (2014) The application of dominance based rough set theory to evaluation of transportation systems. *Procedia – Social and Behavioral Sciences*, 111, 1238-1248.
26. Sharma, H. K., Kar, S. (2018). Decision making for hotel selection using rough set theory: A case study of Indian hotels. *International Journal of Applied Engineering Research*, 13(6), 3988-3998.