



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 11    Issue: IV    Month of publication: April 2023**

**DOI: <https://doi.org/10.22214/ijraset.2023.49913>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Prediction on RAC Tickets in Indian Railways

Anusha R<sup>1</sup>, Dr. T. A. Albinaa<sup>2</sup>

<sup>1</sup>PG Research Scholar, <sup>2</sup>Assistant Professor Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu

**Abstract:** The “PREDICTION ON RAC TICKETS IN INDIAN RAILWAYS” is based on the Rail transport. It is one of the prominent modes of transportation with reference to goods as well as passenger’s movement in any Economy. Indian Railways also played important role in Economic development of our country and will be at key position in future too. Despite of lots of changes and development in Indian Railways with reference to passenger transportation i.e., increase in number of Railways services, trains, types of trains, facilities inside and outside the trains especially on platforms, ticket reservation system and its procedure via technological advancement and digitalization, information to the passengers regarding their ticket status, and so on, still we find certain basic missing link in the services of Indian Railways and due to which it suffers not only revenue losses but many more. An attempt is being made to address that basic deficiency of Indian Railways reservation system especially with reference to wait listed reservation ticket with the passengers enroute and no information about actual position of the waitlisted ticket enroute in case of vacant seat after departure from the origin point and II charting done. Thus, in this project solution for better convenience of enroute passengers especially with waitlisted or RAC tickets and prevent TTE’s for making illegal money and save revenue loss to the railway through efficiently implementation of suggested ICT within existing systems.

**Keywords:** Ticket Reservation System, wait listed reservation ticket, Save revenue loss

## I. INTRODUCTION

A significant change has been seen in Indian railways since its inception in both the segment i.e., Freight and passenger transportation. Although, Indian Railways still must go long way with some more structural changes to increase its share in Freight as well as passenger movement as compared to road transportation, whatever has been done so far by our concerned authorities for the development of Indian railway system is incredible.

Those developments can be seen in terms of route length, number of coaches, number of railway services, electrification, dedicated corridor, modern infrastructure supported by better and efficient digitalization, etc. Further, it has learnt that there are about 1200 trains that carry about 20 lakhs or more passengers every day across the country through more than 10,000 services by those trains. It is also learnt that during those services especially with reference to passenger movements, in each service there are around 40-50 seats vacant due to various reasons, and given our current system of Indian Railways in order to check onboard passengers and if there are any vacant seat enroute due to no passenger, only TTE are having that information and given this asymmetry those passengers who are travelling with either waitlisted or RAC tickets and who really deserve the vacant seat given their waiting ticket status, don’t get the seat and against this TTE sale those vacant seats to the other passengers who give them hefty amount. This is how inconvenience to passengers and revenue loss to railways occurs. Thus, given this back drop this project looks at the issue and try to come up with appropriate solution which helps Railway to serve better with more comfort and convenience to the passengers.

## II. RELATED WORK

Indian Railways initially had a different way of reserving tickets to passengers. As days went on many advancements came into practice.

But that could also not lead in preventing revenue loss by vacant seat, mid station booking calculation and others. The papers that are published so far includes explains about the ticket reservation through online mode. Machine Learning algorithms are used to check the enroute waitlisted passengers and ticket reservation system.

Taking this to the next level, this paper works on checking the probability of point to point and mid station booked tickets. By using Logistic Regression, Random Forest, K-Means algorithm, the accuracy score of these are compared and the one best algorithm can be used to check the probability of each train and allocate percentage for point to point and mid station bookings which will increase the revenue to the Indian Railways.

### III. PROBLEM DESCRIPTION

Indian Railways has grown significantly since its inception in 1853. Today, we can see lots of paradigm shifts, and structural changes in Indian Railways with lots of developments and improvements with reference to facilities and availability with top class infrastructure too. Currently, on an average, Indian Railways carried 22.24 million passengers and 3.04 million tonnes of freight each day. Along with all these facilities the Railways authorities trying their best to reduce asymmetries with reference to passenger ticket reservation system through various means and ways. However, despite of all those improvements, still there exists some lacuna which creates asymmetric in the current system. This project entitled “PREDICTION ON RAC TICKETS IN INDIAN RAILWAYS” aims to understand Indian Railways and its ticket reservation system, evaluate the status of enroute wait listed tickets given current scenario and response of TTEs, explore the possibilities of solution to the problem of enroute waitlisted and RAC tickets and some other benefits too, suggest policy guidelines and recommendations based on study.

### IV. METHODOLOGY

A real time dataset with source station, destination station, mid stations counting from station 1, station 2, station 3, station 4, station 5, total reserved seats in the train, total seats available in the train, the price of each ticket, and the mid station booked tickets. An analysis is being made with the dataset to understand the reservation probability of tickets in each station and by which the vacant seats can be given to the wait listed passengers and other deserving passengers.

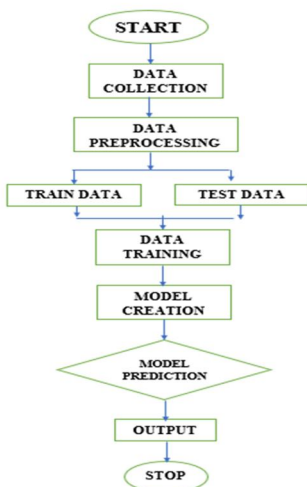
The Machine Learning algorithms are used to build the model and predict the future reservations from the existing one. The accuracy of these algorithms in these are analyzed and the one with high accuracy is suggested for reservation systems.

Train_Series	Train_No	Train_Name	Departure_Time	Arrival_Time	SourceStation	Destination_Station	total_Seats	Reserved_TicketsPP	Reservation_Ticket_price	Station1
1	12679	Intercity Express Coimbatore - Chennai	5.12	14.15	Chennai	Coimbatore	200	90	450	Arakonam
2	12007	Chennai Bangalore Express	6.12	10.25	Chennai	Bangalore	300	100	500	Tiruvallur
3	12657	Chennai Bangalore Fast Express	6.15	10.25	Chennai	Bangalore	350	100	600	Tiruvallur
4	12027	Shatabdi Chennai blr fast exp	7.10	10.12	Chennai	Bangalore	200	70	1000	Tiruvallur
5	12607	Lalbah express	11.45	14.26	Chennai	Bangalore	300	150	600	Tiruvallur

Figure: Dataset Description

### V. DATA PREPROCESSING

The data are cleaned and processed before sending it into a model. The null values are checked and either removed or replaced. In this dataset, there were null values present. Those null values were replaced with the unknown. This dataset contains both categorical and numerical values. Both the numerical and categorical values will be converted into 0 and 1 for the modeling process. Initially, the needed columns are converted to list for easy analysis. Later all the values are converted into numbers.



## VI. PROCESS FLOW

As the analysis works best on list, the data is converted to list. For general understanding of the data filtering of the needed columns is done. From the entire dataset the source station is chosen one by one, name of the trains and the number of tickets booked are analysed. The same way the destination station is also changed one by one, to check the availability of seats for point-to-point passengers.

Logistic regression is used to train the model and predict the future tickets of source, destination and all the mid stations. The accuracy score of logistic is also checked. Next algorithm that is being used is K-Means which vectorizes the data and forms clusters. The top clusters are found. The accuracy score of these is also calculated. This is followed by a Random Forest classifier. The model is built and the classification report is pulled out. The confusion report is also generated for understanding the dataset.

By comparing the accuracy score of all the three algorithms, the one best can be suggested for the betterment of the reservation system.

## VII. MODEL EVALUATION

Logistic Regression, Random Forest Classifier, and K-Means Clustering are implemented for training and testing the model. For choosing a model we split our dataset into training and testing. Here data is split into a 3:1 ratio which means training data has 70 percent and testing data has 30 percent. In this split process, performing based on the `train_test_split` model. This train and test data are assigned to various variables and the model is built. By testing the model, the future prediction of bookings in several stations is easily understood. For the K-Means, the number of clusters needed are formed and the top among the clusters are found, indicating the most used trains and stations.

### A. Logistic Regression

Logistic regression is a predictive analysis. Logistic regression describes data and explains the relationship between a dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. Another important consideration is the model fit when selecting the model for the logistic regression analysis. Split our dataset into train and test sets and fit the dataset to the Logistic regression model.

The probabilities of ticket reservation from point-to-point and for mid stations of each train are calculated.

### B. Random Forest Classifier

Random Forest is a flexible, easy-to-use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks). In random forest classification, multiple decision trees are created using different random subsets of the data and features. Each decision tree is like an expert, providing its opinion on how to classify the data. Predictions are made by calculating the prediction for each decision tree, then taking the most popular result. Classification report gives a better understanding of the data.

### C. K-MEANS Classifier

If our data is labeled, we can still use K-Means, even though it's an unsupervised algorithm. We only need to adjust the training process. Since now we do have the ground truth, we can measure the quality of clustering via the actual labels. Even more so, we don't have to guess the number of clusters—we set it to the number of classes. Once we identify the clusters, we classify a new object by its proximity to the centroids.

In classification, accuracy is an important evaluation parameter. Accuracy is the proportion of the total number of predictions that were correct. It can be obtained by the sum of true positive and true negative instances divided by 100. And Precision is the fraction of true positive and predicted yes instances.

The formula for Accuracy and Precision are given below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{100}$$

The following table shows the implemented algorithm and its accuracy arrived.

Algorithm	Accuracy
Logistic Regression	0.6
Random Forest Classifier	1.0
K-Means Classifier	0.9

The accuracy score of the three algorithms is evaluated and the prediction model is built using the CSV file by defining the object function and reading the file.

```
#CURRENT STATUS
#accuracy for Station3_Booked_Tickets
acc_logistic=metrics.accuracy_score(test_Y1,predicted_value_1)
acc_logistic

0.07142857142857142
```

Figure: Accuracy score of Logistic Regression

```
#accuracy for Status
acc_randomforest=metrics.accuracy_score(y_test,rfc_pred)
acc_randomforest

1.0
```

Figure: Accuracy score of Random Forest Classifier

### VIII. PREDICTED RESULTS

The actual bookings were trained for each station and the future bookings of several stations were also predicted. Some stations had more count than the actual bookings and some stations had very less future count when compared to actual bookings.

The point-to-point bookings of each train are also analyzed, and the counts are found.

Visualization of the stations and their booked tickets is done to have an better understanding of the bookings of each train.

The final analysis of the probability of tickets that can be given for the point-to-point travelers is also found for all the trains in the dataset.

Results of Prediction of Station3\_Booked\_Ticket

Future Prediction	Actual Booked Ticket
87	20
48	5
65	20
60	20
10	20
...	...
51	30
92	20
57	20
89	20
55	20

Figure: Future prediction of station 3 bookings

Results of Prediction of Station2\_Booked

predicted_value	Actual Booking
87	34
48	15
65	34
60	34
10	34
...	...
51	32
92	34
57	10
89	34
55	34

Figure: Future prediction of station 2 bookings

Station1\_Booked\_Tickets

Train_Name	Station1_Booked_Tickets
Chennai Bangalore Express	15
Lalbah express	15
Chennai Bangalore Fast Express	11
Intercity Express Coimbatore - Chennai	11
Shatabdi Chennai blr fast exp	11
Bharthiyar Tirunelveli express	4
Blr sppt Express	4
Chearan Express	4
South Mountain exp	4
Tirunelveli express	4

Figure: Station 1 bookings

Destination\_Station

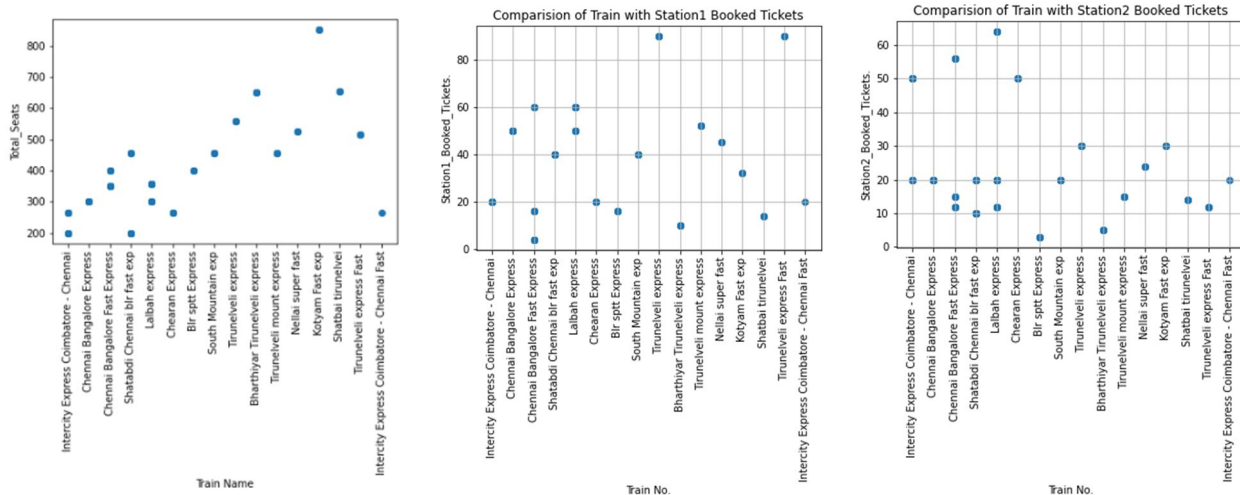
SourceStation	Destination_Station
Chennai	35
Bangalore	32
Namakkal	24
Coimbatore	9

Figure: Destination bookings

```
print(classification_report(y_test,rfc_pred))
```

	precision	recall	f1-score	support
200	1.00	1.00	1.00	7
265	1.00	1.00	1.00	3
300	1.00	1.00	1.00	9
350	1.00	1.00	1.00	3
356	1.00	1.00	1.00	2
456	1.00	1.00	1.00	6
514	1.00	1.00	1.00	1
526	1.00	1.00	1.00	2
560	1.00	1.00	1.00	2
650	1.00	1.00	1.00	3
654	1.00	1.00	1.00	1
852	1.00	1.00	1.00	1
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Figure: K-Means report



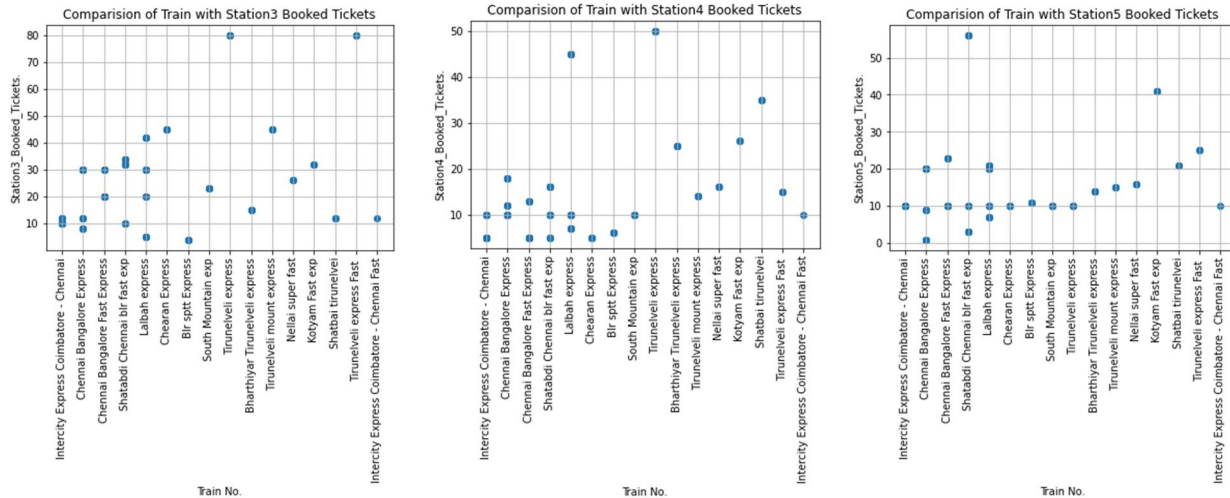


Figure: Station wise bookings

145695	----- Chennai Bangalore Express	----- 300	----- 100	----- 121	----- 40.33333333333333
334323	----- Blr sptt Express	----- 400	----- 100	----- 40	----- 10.0
121212	----- South Mountain exp	----- 456	----- 70	----- 103	----- 22.587719298245617
169851	----- Lalbah express	----- 356	----- 150	----- 180	----- 50.56179775280899
189893	----- Tirunelveli express	----- 560	----- 310	----- 260	----- 46.42857142857143
898192	----- Bharthiyar Tirunelveli express	----- 650	----- 400	----- 69	----- 10.615384615384615
129393	----- Tirunelveli mount express	----- 456	----- 200	----- 141	----- 30.92105263157895
898131	----- Nellai super fast	----- 526	----- 300	----- 127	----- 24.14448669201521
712111	----- Kotyam Fast exp	----- 852	----- 310	----- 161	----- 18.896713615023476
512101	----- Shatbai tirunelveli	----- 654	----- 258	----- 96	----- 14.678899082568808
165786	----- Tirunelveli express Fast	----- 514	----- 278	----- 222	----- 43.190661478599225
25689	----- Chearan Express	----- 265	----- 45	----- 130	----- 49.056603773584904
14568	----- Chennai Bangalore Express	----- 300	----- 100	----- 120	----- 40.0
54782	----- Chennai Bangalore Fast Express	----- 400	----- 100	----- 76	----- 19.0
87452	----- Shatabdi Chennai blr fast exp	----- 456	----- 70	----- 154	----- 33.771929824561404
69852	----- Lalbah express	----- 356	----- 150	----- 130	----- 36.51685393258427
12678	----- Intercity Express Coimbatore - Chennai	----- 200	----- 90	----- 95	----- 47.5
12008	----- Chennai Bangalore Express	----- 300	----- 100	----- 89	----- 29.666666666666668
12656	----- Chennai Bangalore Fast Express	----- 350	----- 100	----- 113	----- 32.285714285714285
12026	----- Shatabdi Chennai blr fast exp	----- 200	----- 70	----- 107	----- 53.5
12608	----- Lalbah express	----- 300	----- 150	----- 143	----- 47.666666666666667
25688	----- Intercity Express Coimbatore - Chennai	----- 265	----- 45	----- 72	----- 27.169811320754718
14569	----- Chennai Bangalore Express	----- 300	----- 100	----- 121	----- 40.33333333333333
54781	----- Blr sptt Express	----- 400	----- 100	----- 40	----- 10.0
87451	----- South Mountain exp	----- 456	----- 70	----- 103	----- 22.587719298245617
69851	----- Lalbah express	----- 356	----- 150	----- 180	----- 50.56179775280899
89893	----- Tirunelveli express	----- 560	----- 310	----- 260	----- 46.42857142857143
89892	----- Bharthiyar Tirunelveli express	----- 650	----- 400	----- 69	----- 10.615384615384615
29393	----- Tirunelveli mount express	----- 456	----- 200	----- 141	----- 30.92105263157895
89813	----- Nellai super fast	----- 526	----- 300	----- 127	----- 24.14448669201521
71211	----- Kotyam Fast exp	----- 852	----- 310	----- 161	----- 18.896713615023476
51210	----- Shatbai tirunelveli	----- 654	----- 258	----- 96	----- 14.678899082568808
65786	----- Tirunelveli express Fast	----- 514	----- 278	----- 222	----- 43.190661478599225

Figure: Probabilities of Source to Destination Travelling passengers in each train

### IX. CONCLUSION

The study concludes that despite of many developments in Indian Railway there still exist one or the other deficiency which may take some time to overcome. In this connection, due to manual ticket checking system by TTEs enroute, even today there exist information asymmetry resulting into havoc situation amongst waitlisted passengers. Thus, this study suggest that railway should develop algorithms enabled system (as suggested in paper) to tackle this situation which will possibly resulting into win-win situation for railways in saving the revenue and for passengers in waitlisted ticket bookings. Thus, in this regard a detailed study is called for. Revenue management (RM) is a collection of techniques for increasing income by controlling the availability and the price of products. Actively applied in airline and hotel businesses, it can be just as successfully implemented for rail operators. Though research in this area is still limited and implementations are derived from other industries, there are successful examples and existing solutions that provide revenue optimization for passenger rail.



### JOURNALS

- [1] Dinesh Mohan. Intelligent Transportation Systems (ITS) and the Transportation System, in Information Technology and Communications Resources for Sustainable Development, [Ed. Ashok Jhunjhunwala], in Encyclopaedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, 2008.
- [2] M. Parihar, "Use of Intelligent Transport System (ITS): A Solution to the Problems of Road Goods Transport (Trucking) Industry in India", Supply Chain Pulse, vol. 8 (1) June 2017.
- [3] N. Parmar, et.al "Intelligent Transportation System", International Journal of Scientific Research and Development, vol. 5 (9) 2017.
- [4] Paul fraga- Lamas, Tiago etc. "Towards Internet of Smart Trains: A Review on Industrial IOT-connected Tailways", Sensors, mdpi Journal, DOI:10.3390/s17061457, published 21 June 2017.
- [5] Ohyun Jo etc. "IoT for Smart Railways feasibility and Applications", IEEE Internet of Things Journal, ISSN:2237-4662, DOI: 10.1109/JIOT.2017.2749401 ISSN:2237-4662
- [6] Daniel T Fokum and Victor S Frost. A survey on methods for broadband internet access on trains. Communications Surveys & Tutorials, IEEE, 12(2):171–185, 2010.
- [7] E Masson and M. Berbineau, " Chapter 2: Railway Applications required Broadband Wireless Applications" Studies in Systems, Decision and Control 82, DOI: 10.1007/978-3-319-47202-7-2.
- [8] William Stallings, Chapter 17, "Data and Computer Communications", Eighth edition, Pearson Prentice Hall @2007, ISBN 0-13-24310-9. Technical Reports
- [9] Rajnish Kumar, "IOT and Indian Railways" National Academy of Indian Railways, Vadodara, India.
- [10] Sandeep Patalay "Railway Signaling Using Wireless Sensor Networks" from the desk of Sandeep Patlay CMC Systems Ltd., P.No 1-25





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)