iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** IV **Month of publication:** April 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.60024

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*

# Airfare Forecasting using Machine Learning to Predict Prices

Ms. A. Hima Bindu[1], Dr. P. Sruthi[2], K. Sai Neeraj Kumar[3], B. Aditya[4], Ch. Saketh[5]
[1]*Assistant Professor, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*
[2] *Head of Department, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*
[3, 4, 5]*Student, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*

***Abstract:*** ***The issue of predicting ticket 111 costs is the focus of this essay. With the assumption that these characteristics have an impact on the cost of an airline ticket, a set of features typical of a normal flight is determined for this purpose. Eight cutting edge machine learning (ML) models using the characteristics are trained to forecast the cost of airline tickets, and the models' output is contrasted with one another. This work examines how the feature set used to represent an airline affects accuracy as well as the prediction accuracy of each model. To train each machine learning model for the trials, a unique dataset including 1814 Aegean Airlines data flights for a particular foreign destination (from Thessaloniki to Stuttgart) was created. Key Words: Airfare price prediction, Machine learning models, Feature dependency, Regression accuracy.***

## I. INTRODUCTION

These days, airline companies employ intricate plans and techniques to determine ticket prices in a flexible manner. Numerous commercial, marketing, financial, and societal elements that are directly related to the final airline rates are being considered by these techniques. Customers find it extremely difficult to acquire an air ticket at the best price because of the dynamic pricing, which is a result of the high complexity of the pricing algorithms used by the airlines. This is the reason why a number of methods that can forecast airfare prices and tell a buyer when is the best time to buy a ticket have been proposed recently. Sophisticated prediction models from the computational intelligence study area known as Machine Learning (ML) are used in most of these techniques. With an accuracy of 75.3% (acc.), Groves and Gini utilized the PLS regression model to maximize the purchase of airline tickets. Using Ripple Down Rule Learner (74.5% accuracy), Logistic Regression (69.9% accuracy), and Linear SVM (69.4% accuracy) machine learning models, Papadakis predicted if the ticket price will decrease in the future. An acceptable performance for cheap tickets many days prior to departure was suggested by Janssen in his linear quantile mixed regression model. In their analysis, Ren, Yang, and Yuan examined the accuracy of four different airline ticket prediction models: Linear Regression (77.06% acc.), NaŹve Bayes (73.06% acc.), SoftMax Regression (76.84% acc.), and SVM (80.6% acc. for two bins). The works listed above only utilized a limited selection of machine learning models. Predicting airline ticket prices globally, with a focus on traditional methods. Nevertheless, as far as the writers are aware, there is still more research to be done on how well the most recent machine learning models do in this regard. The suggested paper's contribution is summed up as follows: For the first time, ticket prices were predicted in Greece. Additionally, features influencing airfare pricing were investigated, and the efficacy of cutting-edge machine learning models used for airfare prediction was examined. This is the format for the remainder of the paper: The second section provides an overview of machine learning and its potential applications to the challenge of predicting airline prices.

## II. OBJECTIVE OF THE PROJECT

Predicting airfare costs is the topic of this essay. Since these characteristics are assumed to have an impact on the cost of an airline ticket, a set of attributes that define a typical flight are determined for this purpose. Eight advanced machine learning (ML) models are trained with the features in order to forecast the cost of airline tickets. The models' output is then compared.This study examines the dependence of accuracy on the feature set used to describe an airline in addition to the forecast accuracy of each model. Each machine learning model in the trials is trained on a unique dataset made up of1814 data flights operated by Aegean Airlines between Thessaloniki and Stuttgart, a specified foreign destination. For a certain class of flying characteristics, the generated experimental findings show that the ML models can handle this regression problem with around 88% accuracy.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*

## III. REALTED WORKLITERATURE REVIEW

*A. "Low-cost airline pricing strategies: The Ryanair case study,"*

We examine the price strategy used by Ryanair, the primarylow-cost airline operating in Europe (2121 44). The best pricing curve for each route is predicted to be 44 based on ayear's worth of fare data for all 22222 of Ryanair's Europeanflights, utilizing a family of hyperbolic price functions.

Theaverage pricing for each route and its length, the number of flights operating on that route, and the percentage of completely booked flights all exhibit positive correlations, according to the data. Fares often drop as the carrier's share of seats at the departure and destination airports rises.

Thereis a negative association between route length and flight frequency and dynamic pricing on the other hand, as competition grows, so do the early ticket savings.

*B. "An airline ticket purchasing schedule that is best predicted using a regression model",*

It is difficult for consumers to choose the best time to buy plane tickets, primarily because they lack the knowledge necessary to predict future changes in pricing. This study proposes a methodology for estimating future prices and analyzing the volatility of pricing.

The suggested approach uses a corpus of past price quotes as a basis to forecast the future estimated minimum price of all available flights on particular routes and dates. We also utilize our technology to forecast flight rates for particular preferred attributes, including flights from a particular airline, nonstop flights exclusively, or multi-segment flights. Customers are able toascertain the expected cost of their preferences by contrastingmodels with various target attributes. Customers find it challenging to determine when is the optimum time to purchase airline tickets, mostly because they lack the knowledge required to anticipate price changes in the future.

A approach for projecting future prices and examining pricing volatility is presented in this paper. The proposed method forecasts the future expected minimum price of all available flights on specific routes and dates based on a corpus of historical price quotes. We also use our algorithm to predict flight prices for specific desired characteristics, such as flights from a specific airline, flights that are only nonstop, or flights that are divided into multiple segments. Customers can compare models with different target features to determine the projected cost of their choices.

*C. "An agent to maximize airline ticket sales,"*

Due to incomplete knowledge, purchasing plane tickets is acommon task where it is challenging for people to save costs. Evaluating how purchase timing translates into changes inpredicted cost is challenging, even with the availability ofhistorical data for inspection (a recent addition to certain tripreservation services).

We introduce an agent that can optimizeclients' purchasing timing in order to solve this issue. We presentdata that show the method can outperform other decision theoretic approaches in this domain, and can performsignificantly closer to the optimal purchase strategy.

*D. "Predicting airline ticket prices using a linear quantile mixed regression model,"*

*That dif*ferent passengers in the same flight class pay significantly different ticket rates for the same service is something that irritates us.

This study compares the goodness offit of four statistical regression models for airline ticket prices. By using this prediction model, travelers can decide with greaterknowledge whether to purchase their ticket now or hold out a little while longer. Our study included a dataset including 126,412 daily observations from Infare [2] of ticket prices for 2,271 distinct flights operating between San Francisco Airport and John F. Kennedy Airport.

Several days prior to leaving, weidentify a model that reasonably captures the data's behavior. Thus, this method may aid future travelers in making a decisionregarding the purchase of a ticket.

## IV. METHODOLOGY<u>PROPOSED SYSTEM</u>

This study investigates the consequences of different aircraft parameters on ticket pricing for the purpose of answer theimportant problem of predicting airfare prices.

After carefully examining our goal is to make sense of the intricate relationship between pricing dynamics and these variables. We assess the accuracy of eight state-of-the-art machine learning models in predicting airline ticket prices through extensive testing. Oddly, our research extends beyond simple accuracy comparisons; it also looks into how the feature set that is used for representation affects the performance of our model.

By utilizing a recently assembled dataset of 1814 Aegean Airlines flight logs on a particular international route, from Thessaloniki to Stuttgart, Important new insights into the predictive capabilities of ML algorithms in the aviation sector are provided by our work. According to our testing results, these models demonstrate a respectable degree of accuracy, with predictions being made with an accuracy of about 88% in certain flight feature configurations.

This shows how machine learning techniquescan increase airfare prediction accuracy, with practical implications for both industry stakeholders and customers.

## V. IMPLEMENTATION

The performance of eight cutting-edge machine learning(ML) models is compared to one another in order to forecastthe cost of airline tickets.

This study examines how the accuracy of each model depends on the feature set that is used to describe an airline in addition to the forecast accuracy of each model. 1814 Aegean Airlines flights for a particular overseas destination (from Thessaloniki to Stuttgart) comprised a unique dataset used in the studies.

### *A. Modules*

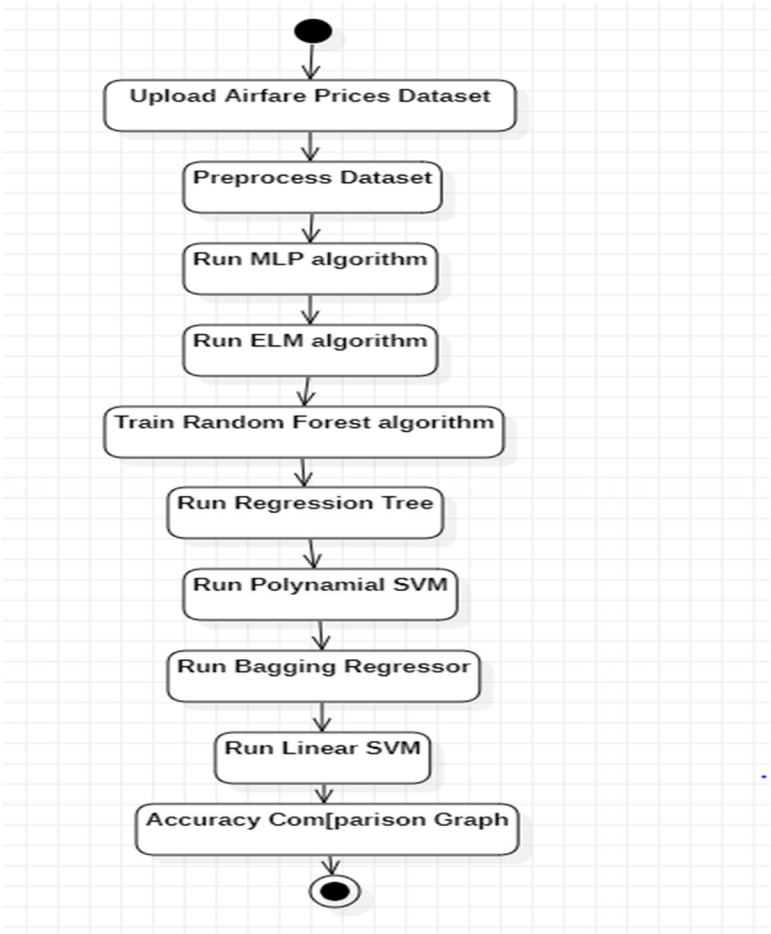
*1)* Upload Airfare Prices Dataset
*2)* Pre-process Dataset
*3)* Run MLP Algorithm
*4)* Accuracy Comparison Graph.

### *B. Modules Description*

*1)* *Upload Airfare Prices Dataset:* Using this module we willupload dataset to application.
*2)* *Pre-process Dataset:* Using this module, we will remove missing values and then encode non-numeric string to numeric data by using Label Encoder class as machinelearning accept only numeric data and above dataset containssome numeric and non-numeric data so by applying Preprocessing will convert that into numeric.
*3)* *Run MLP Algorithm:* Using this module we will split dataset into train and test and then trained algorithm by using TRAIN data and then apply test data to calculate its prediction accuracy.Similarly, we will run all algorithms and compute their accuracy.
*4)* *Accuracy Comparison Graph:* Using this module, an accuracycomparison graph comparing the relative performances of each approach will be made. The names of the algorithms will be displayed on the graph's x-axis, and each algorithm's accuracy will be shown on the y-axis. Regression Tree is notably found to exhibit superior accuracy on a consistent basis in contrast to the other models the study looked at, suggesting that it is a useful model for predicting airline prices.By evaluating the performance of eight state-of-the-art machine learning models on a recently created dataset containing 1814 Aegean Airlines flight records—with a focus on an international route that connects Thessaloniki and Stuttgart—the study addresses the challenge of predicting airfare prices. Choosing a set of features that represent various aspects of a flight allows you to first hypothesize how a set of attributes may affect the price of a ticket. Subsequently, these features undergo preprocessing in order to handle missing values and encode non-numeric data, ensuring their use by machine learning algorithms. The methodology consists of four main modules: preprocessing, model training with the MLP algorithm, uploading datasets, and comparing the prediction accuracies of the various models. Regression tasks are well- suited for the models chosen for comparison, as they are often found in the literature. The dataset is split into training and testing sets as part of the experimental design, and model performance is evaluated using suitable evaluation criteria. Finding patterns in model correctness and feature set dependencies is emphasized during the analysis and presentation of the results. To aid with interpretation, visual aids like accuracy comparison graphs are usedThe discussion section examines the implications of the findings, considering how they might be applied to airfare forecasting as well as potential lines of inquiry. The study concludes by highlighting the need of precise airfare forecast, outlining future research possibilities, and highlighting the research's accomplishments.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*

## VI. FLOW CHART




To illustrate the flow from one action to another, it is essentially a flowchart. One may make use of the activity as a system operation. As a result, the control flow is transferred across operations. This flow may occur concurrently, forked, or sequentially.

## VII. RESULTS AND SCREENSHOTS

1) *Testing and Implementation:* Among the most crucialproject jobs, implementation is one where vigilance is required because all project-related activities will be highlyparticipatory. The implementation stage is critical to building a successful system and guaranteeing users' confidence in the newsystem's feasibility and effectiveness. Using sample data, each program is tested separately during development to ensure that itlinks to other programs in the manner described in the programspecification. The user is satisfied with the testing done on the computer system and its surroundings.
2) *Implementation:* Compared to system design, the implementation phase is less imaginative. Its main concerns are file conversion and user training. It's possible that the system needs a lot of user training. As a result of programming, the system's starting parameters should change. The user can quickly and simply become familiar with the many capabilities by using the simple operating technique.The user can acquire the variousreports using either the dot matrix or inkjet printer that isprovided to them. It’s fairly simple to put the suggested systeminto place. Implementation, in general, refers to the process of turning a newly created or updated system design into a functional one.
3) *Testing:* Testing is the process of preparing test data, which is then used to test each module separately and, subsequently, to validate the fields. To make sure that every component of the system property functions as a whole, system testing is then carried out. It is important to select test data that can withstand every scenario.


Fig 6.1 User Interface

| ML ALGORITHM | ACCURACY | EXECUTION TIME |
|---|---|---|
| MLP | 0.4136 | 0.086 sec |
| ELM | 0.9353 | 0.144 sec |
| RANDOM FOREST | 0.9328 | 0.136 sec |
| REGRESSION TREE | 0.9932 | 0.002 sec |
| BAGGING REGRESSOR | 0.9199 | 0.024 sec |
| POLYNOMIAL SVM | 0.7071 | 0.012 sec |
| LINEAR SVM | 0.3785 | 0.005 sec |
| LINEAR REGRESSION | 0.3958 | 0.021 sec |

Fig 6.2 Algorithms Accuracy Table

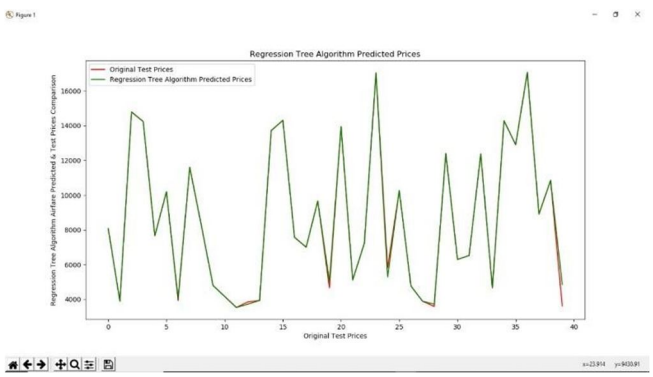



Fig 6.3 Regression Tree Graph

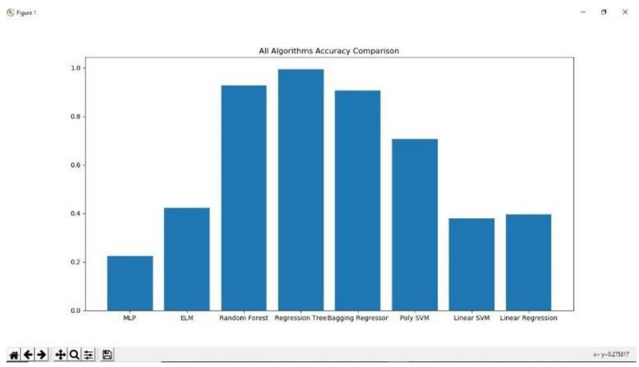



Fig 6.4 Accuracy Comparison Graph

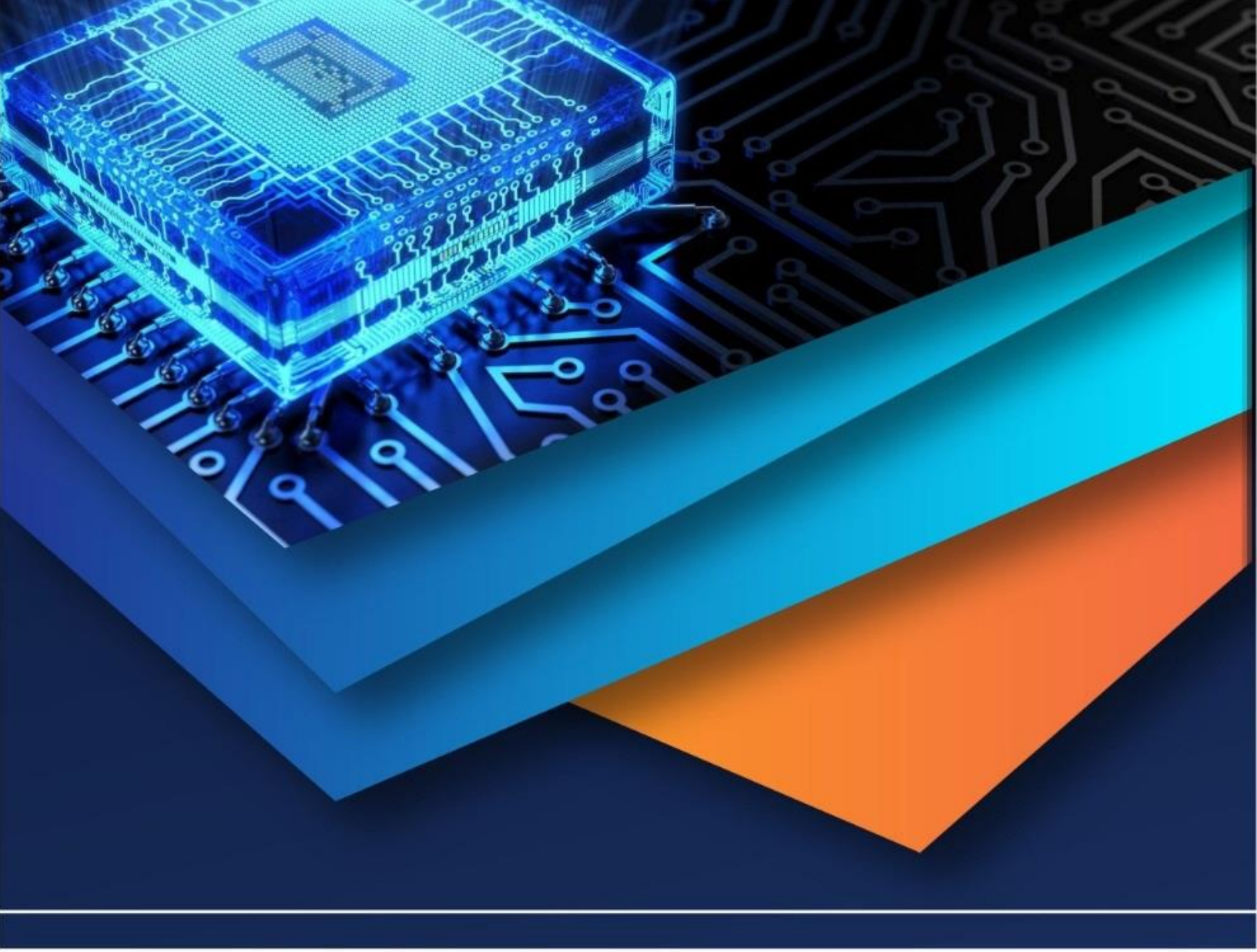
## VIII. CONCLUSION

The preliminary research is covered in this article under "Airfare Forecasting". We demonstrated that it is possible toforecast flight prices based on past price data by gathering flight ticket data of one Greek airline (Aegean Airlines) from the internet. Based on experimental results, machine learning models are a useful tool for predicting airline ticket costs. Data collection and feature selection are also critical components of airfare forecasting, from which we havederived insightful findings.

Which features have the biggestimpact on the prediction of airline tickets was determined bythe experiments. There are additional features that can increase the prediction's accuracy in addition to the ones that were chosen. This work could be expanded to forecast the cost of plane tickets throughout the whole map in the future. Although more research using larger datasets of airline tickets is required, this preliminary pilot study shows how machine learning models can help customers buy tickets at the highestpossible price.

## REFERENCES

[1] Sruthi. p, Premkumar, L.” Attribute-based storage supporting secure de-duplication of encrypted data in cloud”, International Journal of Recent Technology and Engineering,2019, 7(6), pp. 418–421.
[2] Shirisha, N., Bhaskar, T., Kiran, A., Alankruthi, K.”Restaurant Recommender System Based on Sentiment Analysis”,2023 International Conference on Computer Communication and Informatics, ICCCI 2023, 2023.
[3] Sasi Bhanu, J., Kamesh, D.B.K., Durga Bhavani, B., Saidulu, G.” An Architecture on Drome Agriculture IoT Using Machine Learning”, Cognitive Science and Technology 2023, Part F1493, pp. 635–641.
[4] Y. Ambica, Dr N. Subhash Chandra MRI brain segmentation using correlation based on adaptively regularized kernel-based fuzzy C-means clustering Int. J. Advanced Intelligence Paradigms, Vol. 19, No. 2, 2021 ‘
[5] Reddy, A.R., Narendar, Ch., Srinu, K., ...Reddy, R.V., Sruthi, P.” COVID 19 Patient Recognition and Prevention Utilizing Machine Learning and CNN Model Techniques”,5th IEEE International Conference on Cybernetics, Cognition and Machine Learning Applications, ICCCMLA 2023, 2023, pp.677–686
[6] Sruthi, P. Singamsetty, P.L.” Enhance the efficiency in multi packet routing for neighbor discovery”, Journal of Advanced Research in Dynamical and Control Systems, 2017, 9(Special issue 14), pp. 2117–2130
[7] Korra, S., Bhaskar, T., Ramana, N., Bhukya, S., Rajender, N.” An Efficient Guided Backpropagation Approach for Detection of Plant Diseases Using Deep Learning Models”, International Journal of Intelligent Systems and Applicationsin Engineering 2024, 12(10s), pp. 52–64.
[8] Bigul, S.D., Prakash, A., Bhanu, J.S.” Futuristic evaluation of CoVID-19 spread using transfer learning: A post vaccination scenario”, AIP Conference, 2021, 2358, 080009.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY