iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42642

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 10 Issue V May 2022- Available at www.ijraset.com

Flight Fare Prediction System Using Machine Learning

Neel Bhosale[1], Pranav Gole[2], Hrutuja Handore[3], Priti Lakade[4], Gajanan Arsalwad[5]
[1, 2, 3, 4, 5]*Department of Information Technology, K. J's Trinity College of Engineering and Research, Pune, India*

Abstract: ***The Flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, duration of flights. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. In this project we majorly targeted to uncover underlying trends of flight prices in India using historical data and also to suggest the best time to buy a flight ticket. The project implements the validations or contradictions towards myths regarding the airline industry, a comparison study among various models in predicting the optimal time to buy the flight ticket and the amount that can be saved if done so. Remarkably, the trends of the prices are highly sensitive to the route, month of departure, day of departure, time of departure, whether the day of departure is a holiday and airline carrier. Highly competitive routes like most business routes (tier 1 to tier 1 cities like Mumbai-Delhi) had a non-decreasing trend where prices increased as days to departure decreased, however other routes (tier 1 to tier 2 cities like Delhi - Guwahati) had a specific time frame where the prices are minimum. Moreover, the data also uncovered two basic categories of airline carriers operating in India – the economical group and the luxurious group, and in most cases, the minimum priced flight was a member of the economical group. The data also validated the fact that, there are certain time-periods of the day where the prices are expected to be maximum. The scope of the project can be extensively extended across the various routes to make significant savings on the purchase of flight prices across the Indian Domestic Airline market.***

Keywords: Flight ticket, Optimal timing, historical data, competitive routes, Indian Domestic Airline market.

I. INTRODUCTION

The flight ticket buying system is to purchase a ticket many days prior to flight take-off so as to stay away from the effect of the most extreme charge. Mostly, aviation routes don't agree this procedure. Plane organizations may diminish the cost at the time, they need to build the market and at the time when the tickets are less accessible. They may maximize the costs. So, the cost may rely upon different factors. To foresee the costs this venture uses AI to exhibit the ways of flight tickets after some time. All organizations have the privilege and opportunity to change its ticket costs at any time. Explorer can set aside cash by booking a ticket at the least costs. People who had travelled by flight frequently are aware of price fluctuations. The airlines use complex policies of Revenue Management for execution of distinctive evaluating systems. The evaluating system as a result changes the charge depending on time, season, and festive days to change the header or footer on successive pages. The ultimate aim of the airways is to earn profit whereas the customer searches for the minimum rate. Customers usually try to buy the ticket well in advance of departure date so as to avoid hike in airfare as date comes closer. But actually, this is not the fact. The customer may wind up by giving more than they ought to for the same seat.

II. LITERATURE SURVEY

1) *K. Tziridis T. Kalampokas G.Papakostas and K. Diamantaras "Airfare price prediction using machine learning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO .2017.8081365L. Li Y. Chen and Z. Li" Yawning detection for monitoring driver fatigue based on two cameras" Proc. 12th Int. IEEE Conf. Intel. Transp. Syst. pp. 1-6 Oct. 2009.*

Proposed study [1] Airfare price prediction using machine learning techniques, For the research work they have used dataset consisting of 1814 data flights of the Aegean Airlines collected and used to train machine learning model. Different number of features were used to train model various to showcase how selection of features can change accuracy of model. They have used various algorithms such as Multilayer Perceptron (MLP), Generalized Regression Neural Network, Extreme Learning Machine (ELM), Random Forest Regression Tree. o Regression Tree, Bagging Regression Tree, Regression SVM (Polynomial and Linear) and Linear Regression (LR) and gained different outputs for each machine learning algorithms. They have tried and trained various types of models with removing and adding different features from the dataset. Followed typical data science life cycle. The best results came from Bagging regression tree.

©IJRASET: All Rights are Reserved

2) *William Groves and Maria Gini "An agent for optimizing airline ticket purchasing" in proceedings of the 2013 international conference on autonomous agents and multi-agent systems.*

In case study [2] by William groves an agent is introduced which is able to optimize purchase timing on behalf of customers. Partial least square regression technique is used to build a model. Initially they have used various techniques for feature selection such as Feature Extraction, Lagged Feature Computation, Regression Model Construction and Optimal Model Selection. Their experiments were designed to estimate real-world costs of using our prediction models. The lag scheme approach works well for many choices of machine learning algorithms, but PLS regression was found to work best for this domain. The improved performance can be attributed to a natural resistance to collinear and irrelevant variables.

3) *J. Santos Dominguez-Menchero, Javier Rivera and Emilio Torres Manzanera "Optimal purchase timing in the airline market".*

In this paper, the researchers have researched the general pattern in airline pricing behaviour and a methodology for analysing different routes and/or carriers. Their purpose is to provide customers with the relevant information they need to decide the best time to purchase a ticket, striking a balance between the desire to save money and any time restraints the buyer may have. Their study shows how non-parametric isotonic regression techniques, as opposed to standard parametric techniques, are particularly useful. Most importantly, we can determine the margin of time consumers may delay their purchase without significant price increase, specify the economic loss for each day the purchase is delayed and detect when it is better to wait until the last day to make a purchase.

4) *Supriya Rajankar, Neha sakhrakar and Omprakash rajankar "Flight fare prediction using machine learning algorithms" International journal of Engineering Research and Technology (IJERT) June 2019.*

Journal by Supriya Rajankar a survey on flight fare prediction using machine learning algorithm uses small dataset consisting of flights between Delhi and Bombay. Algorithms such as K-nearest neighbours (KNN), linear regression, support vector machine (SVM) are applied to gain different outcomes and do research on them. For predicting the flight ticket prices, many algorithms were implemented in machine learning. The algorithms are: Support Vector Machine (SVM), Linear regression, K-Nearest neighbours, Decision tree, Multilayer Perceptron, Gradient Boosting and Random Forest Algorithm. Using python library scikit learn these models have been implemented. The parameters like R-square, MAE and MSE are considered to verify the performance of these models. The best model results were of Decision Tree algorithm.

5) *Tianyi wang, samira Pouyanfar, haiman Tian and Yudong Tao "A Framework for airline price prediction: A machine learning approach"*

In this paper, Tianyi wang, samira Pouyanfar, haiman Tian and Yudong Tao [5] proposed framework where two databases are combined together with macroeconomic data and machine learning algorithms such as support vector machine, XGBoost are used to model the average ticket price based on source and destination pairs. The framework achieves a high prediction accuracy 0.869 with the adjusted R squared performance metrics. They had the result of lowest error rate of 0.92 with the XGBoost Algorithm.

6) *T. Janssen "A linear quantile mixed regression model for prediction of airline ticket prices"*

In this paper, they have predicted the best time to purchase the tickets. They have used various machine learning algorithms such as linear regression, Decision Tree, Random Forest, K-Nearest Neighbour, Multilayer Perceptron (MLP), gradient boosting, support vector machine (SVM). For predictors, they have used Naïve Bayes and Stacked Prediction Model. the research a desired model is implemented using the Linear Quantile Blended Regression methodology for San Francisco–New York course where each day airfares are given by online website. Two features such as number of days for departure and whether departure is on weekend or weekday are considered to develop the model.

7) *Wohlfarth, T.clemencon, S.Roueff "A Data mining approach to travel price forecasting" 10th international conference on machine learning Honolulu 2011.*

In the research paper [7] on Flight fare prediction system by Wohlfarth, T.clemencon, S.Roueff using the technique of yield management in the air travel industry. They have used various data mining techniques. It is the goal of this paper to consider the design of decision-making tools in the context of varying travel prices from the customer's perspective. Terms used in the research are machine techniques/ algorithms mentioned as Clustering.

8) Vinod Kimbhaune, Harshil Donga, Ashutosh Trivedi, Sonam Mahajan and Viraj Mahajan research paper on flight fare prediction system.

In the research paper [7] on Flight fare prediction system by Vinod Kimbhaune, Harshil Donga, Ashutosh Trivedi, Sonam Mahajan and Viraj Mahajan using the various machine learning algorithm approaches i.e., Random Forest, Decision tree and Linear regression are applied on dataset. To determine ideal purchase time for flight ticket. There project aims to develop an application which will predict the flight prices for various flights using machine learning model. The techniques they have used are mentioned as Linear Regression, Decision Tree and random Forest. The performance metrics techniques used are MAE, MSE and RSME. The outcome for their project was not fully accurate but by adding more real time data set will give more accurate results.

9) W. Groves and M. Gini, ―An agent for optimizing airline ticket purchasing,‖ 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), St. Paul, MN, May 06 - 10, 2013, pp. 1341-1342.

This is the extended version of the research paper [3] exploited Partial Least Square Regression (PLSR) for building up a model. The information was gathered from major travel adventure booking sites from 22 February 2011 to 23 June 2011. Extra information was additionally gathered and are utilized to check the correlations of the exhibitions of the last model. Janssen.

10) Viet Hoang Vu, Quang Tran Minh and Phu H. Phung,‖An Airfare Prediction Model for Developing Markets‖, IEEE paper 2018.

In this paper, they have proposed a new model that can help the buyer to predict the price trends without official information from the airlines. Their findings demonstrated that the proposed model can predict the trends as well as actual airfare's changes up to the departure dates using public airfare data available online despite the missing of many key features like the number of unsold seats on flights. They have also identified the features that have the strongest impacts on the airfare changes. They proposed a ticket purchasing time improvement model subject to a significant pre-processing known as macked point processors, data mining frameworks (course of action and grouping) and quantifiable examination system. This framework is proposed to change various added value arrangements into included added value arrangement heading which can support to solo gathering estimation. This value heading is packed into get together reliant on near evaluating conduct. Headway model measure the value change plans. A tree-based analysis used to pick the best planning gathering and a short time later looking at the progression model.

11) Wohlfarth, T. Clemencon, S.Roueff.-A Dat mining approach to travel price forecastingl, 10 th international conference on machine learning Honolulu 2011.

In this paper we learned that a large body of data-mining techniques have been developed over the last two decades for the purpose of increasing profitability of airline companies. The mathematical optimization strategies put in place resulted in price discrimination, similar seats in a same flight being often bought at different prices, depending on the time of the transaction, the provider, etc. Itis the goal of this paper to consider the design of decision-making tools in the context of varying travel prices from the customer’s perspective. Based on vast streams of heterogeneous historical data collected through the internet, we describe here two approaches to forecasting travel price changes at a given horizon, taking as input variables a list of descriptive characteristics of the flight, together with possible features of the past evolution of the related price series. Though heterogeneous in many respects (e.g., sampling, scale), the collection of historical prices series is here represented in a unified manner, by marked point processes (MPP). State-of-the-art supervised learning algorithms, possibly combined with a preliminary clustering stage, grouping flights whose related price series exhibit similar behaviour, can be next used in order to help the customer to decide when to purchase her/his ticket.

12) Dominguez-Menchero, J.Santo, Reviera,

optimal purchase timing in airline markets. This paper presents general patterns in airline pricing behaviour and a methodology for analysing different routes and/or carriers. The purpose is to provide customers with the relevant information they need to decide the best time to purchase a ticket, striking a balance between the desire to save money and any time restraints the buyer may have. The study shows how non-parametric isotonic regression techniques, as opposed to standard parametric techniques, are particularly useful. Most importantly, we can determine the margin of time consumers may delay their purchase without significant price increase, specify the economic loss for each day the purchase is delayed and detect when it is better to wait until the last day to make a purchase.

13) medium.com/analytics-vidhya/mae-mse-rmse-coefficient-ofdetermination adjusted-r-squared-which-metric-is
Better article on performance metrics. In this paper we learned that the objective of Linear Regression is to find a line that minimizes the prediction error of all the data points. The essential step in any machine learning model is to evaluate the accuracy of the model. The Mean Squared Error, Mean absolute error, Root Mean Squared Error, and R-Squared or Coefficient of determination metrics are used to evaluate the performance of the model in regression analysis. However, RMSE is widely used than MSE to evaluate the performance of the regression model with other random models as it has the same units as the dependent variable (Y-axis). The RMSE tells how well a regression model can predict the value of a response variable in absolute terms while R- Squared tells how well the predictor variables can explain the variation in the response variable.

14) www.keboola.com/blog/random-forest-regression article on random forest
Random forest is both a supervised learning algorithm and an ensemble algorithm. It is supervised in the sense that during training, it learns the mappings between inputs and outputs. Ensemble algorithms combine multiple other machine learning algorithms, in order to make more accurate predictions than any underlying algorithm could on its own. In the case of random forest, it ensembles multiple decision trees into its final decision. Random forest can be used on both regression tasks (predict continuous outputs, such as price) or classification tasks (predict categorical or discrete outputs). The way in which you use random forest regression in practice depends on how much you know about the entire data science process. We recommend that beginners start by modelling data on datasets that have already been collected and cleaned, while experienced data scientists can scale their operations by choosing the right software for the task at hand.

15) https://towardsdatascience.com/machine-learning-basics-decisiontreeregression-1d73ea003fda article on decision tree regression.
In this paper we learned Decision Tree is one of the most commonly used, practical approaches for supervised learning. It can be used to solve both Regression and Classification tasks with the latter being put more into practical application. It is a tree-structured classifier with three types of nodes. The Root Node is the initial node which represents the entire sample and may get split further into further nodes. The Interior Nodes represent the features of a data set and the branches represent the decision rules. Finally, the Leaf Nodes represent the outcome. This algorithm is very useful for solving decision-related problems. Decision trees have an advantage that it is easy to understand, lesser data cleaning is required, non-linearity does not affect the model's performance and the number of hyper-parameters to be tuned is almost null.

16) O. Etzioni, R. Tuchinda, C. A. Knoblock, and A. Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price.
This paper reported on a pilot study in "price mining" over the web. We gathered airfare data from the web and showed that it is feasible to predict price changes for flights based on historical fare data. Despite the complex algorithms used by the airlines, and the absence of information on key variables such as the number of seats available on a flight, our data mining algorithms performed surprisingly well. Most notably, our Hamlet data mining method achieved 61.8% of the possible savings by appropriately timing ticket purchases. Our algorithms were drawn from statistics (time series methods), computational finance (reinforcement learning) and classical machine learning (Ripper rule learning). Each algorithm was tailored to the problem at hand (e.g., we devised an appropriate reward function for reinforcement learning), and the algorithms were combined using a variant of stacking to improve their predictive accuracy.

17) Manolis Papadakis. Predicting Airfare Prices.
Airlines implement dynamic pricing for their tickets, and base their pricing decisions on demand estimation models. The reason for such a complicated system is that each flight only has a set number of seats to sell, so airlines have to regulate demand. In the case where demand is expected to exceed capacity, the airline may increase prices, to decrease the rate at which seats fill. On the other hand, a seat that goes unsold represents a loss of revenue, and selling that seat for any price above the service cost for a single passenger would have been a more preferable scenario. The purpose of this project was to study how airline ticket prices change over time, extract the factors that influence these fluctuations, and describe how they're correlated (essentially guess the models that air carriers use to price their tickets).

©IJRASET: All Rights are Reserved

18) Groves and Gini, 2011. A Regression Model for Predicting Optimal Purchase Timing for Airline Tickets.

Optimal timing for airline ticket purchasing from the consumer perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. This paper presents a model for computing expected future prices and reasoning about the risk of price changes. The proposed model is used to predict the future expected minimum price of all available flights on specific routes and dates based on a corpus of historical price quotes. Also, we apply our model to predict prices of flights with specific desirable properties such as flights from a specific airline, non-stop only flights, or multi-segment flight. By comparing models with different target properties, buyers can determine the likely cost of their preferences. We present the expected costs of various preferences for two high-volume routes. Performance of the prediction models presented is achieved by including instances of time- delayed features, by imposing a class hierarchy among the raw features based on feature similarity, and by pruning the classes of features used in prediction based on in-situ performance.

19) Modelling of United States Airline Fares - Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DBIB), Krishna Rama Murthy, 2006.

Prediction of airline fares within the United States including Alaska & Hawaii is required for transportation mode choice modelling in impact analysis of new modes such as NASA Small Airplane Transportation System (SATS). Developing an aggregate cost model i.e., a generic fare model' of the disaggregated airline fares is required to measure the cost of air travel. In this thesis, the ratio of average fare to distance i.e., fare per mile and average fare is used as a measure of this cost model. The thesis initially determines the Fare Class categories to be used for Coach and Business class for the analysis. The thesis then develops a series of generic fare models; using round trip distance travelled as an independent variable. The thesis also develops a set of models to estimate average fare for any origin and destination pair in the US. The factors considered by these models are: the round-trip distance travelled between the origin (o) and destination (d), the type of fare class chosen by the traveller.

20) B.S. Everitt: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge (3rd edition, 2006). ISBN 0-521-69027-7.

This is university book by University of Cambridge, England which tells us about lots of maths and algorithms to be used.

21) Bishop: Pattern Recognition and Machine Learning, Springer, ISBNO-387-31073-8.

The dramatic growth in practical applications for machine learning over the last ten years has been many important developments in the algorithms and techniques. This completely new textbook represents these recent developments while providing a comprehensive introduction to the fields of pattern recognition and machine learning. No previous knowledge of pattern recognition or machine learning concepts is assumed and some experience in the use of probabilities would be helpful though not essential as the book includes a self-contained introduction to basic probability theory. The book is suitable for courses on machine learning, statistics, computer science, signal processing, computer vision, data mining. Solutions for a subset of the exercises are available from the book web site, while solutions for the remainder can be obtained by instructors from the publisher. The book is helpful by a great deal of additional material, and the reader is to visit the inspire book web site for the latest information. while new models based on kernels have had a significant impact on both algorithms and applications.

22) E Bachis and C. A. Piga Low-cost airlines and price dispersion. International Journal of Industrial Organization, In Press, Corrected Proof, 2011.

The following represents the tactics of online pricing in which different airlines announce fares for same flights at same time but in different currencies that causes violation of; Law Of One Price The survey reveals that different airlines post different fares for less competitive routes with more heterogeneous demands. The temporal persistence of intra-firm fare dispersion suggests that it is an equilibrium phenomenon engendered by the airlines' need to manage stochastic demand conditions for a specific flight.

23) P.P. Belobaba. Ariline yield management. An overview of seat inventory control. Transportation Science, 21(2):63, 1987.

The topic of seat inventory control of airlines yield management is examined by practical aspects of the problem. A current survey on the following topic represents that rather than systematicanalys is seat inventory control is generally depends upon human judgement. The past work on this topic has pointed on simplification of problems and largescale optimization of models. It replenishes that there is a need for the practical solution approaches that incorporate the quantitative decision tools. 24. Y. Levin, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. Management

Science, 55(1):32-46, 2009 The rapid growth in Internet sales channels and point-of-sale technologies has given many firms a new capability for revenue management (RM). They can now monitor demand for their products in the real time and adjust prices dynamically in response to changes in demands patterns for example, many online airline booking systems allow consumers to choose preferred seats from the remaining seats on given flight. Experienced consumers may now behave strategically by timing their purchases to anticipated periods of lower price. If a reasonable approximation of the effects of competitor response can be captured by a price-sensitive demand model. The effects of competition between firms on their pricing policies this requires some form of dynamic differentiated products model that also captures. Thus, another important type of strategic interaction that needs to be captured is consumer choice it is the most important thing all over the main strategic how consumers choose among different products. The model provides insights about equilibrium price dynamics under different levels of competition, asymmetry between firms, and multiple market segments with varying properties. We demonstrate that strategic behaviour by consumers can have serious impacts on revenues if firms ignore that behaviour in their dynamic pricing policies. Moreover, ideal equilibrium responses to consumer strategic behaviour can recover only a portion of the lost revenues. A key conclusion is that firms may benefit more from limiting the information available to consumers than from allowing full information and responding to the resulting strategic behaviour in an optimal fashion.

24) *B Smith, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. Management at American arilines, Interfaces, vol.22, pp. 8-31, 1992.*

Critical to an airline's operation is the effective use of its reservations inventory. In early 1960's American Airlines began to research in managing revenue from this inventory. American Airlines DesicionTechnologies developed series of OR models that decreases large problems into three much smaller and great more subproblems caused because of the problem of size and difficulty like discount allocation, traffic management and overlooking etc. The end products of solutions of subproblems are combined together to determine final inventory levels. American Airlines roughly calculates the benefit at $1.4 billion over the last three years and awaits an annual revenue contribution of over $500 million to continue into the future.

25) *T. Janseen, "A linear quantile mixed regression model prediction of airline ticket prices," Bachelor Thesis, Radbound University, 2014.*

The airline implements diverse pricing of flight tickets. According to all surveys, the fares of flight tickets changes during morning and evening time Also in days of festivals ans holidays. There are various factors that affects the fares of flight tickets. The seller has all of the information about airlines fares but for buyers it is hard to predict as they have limited information. Considering the aspects like number of days for departure, departure time and time of day which gives best time to buy the flight tickets. The paper reports about the factors which influence the airfare prices and how they are related to the changes. And by using all this feature build a system which supports buyers to decide whether to buy a ticket or not.

26) B kotsiants, "Decision trees: a recent overview, "Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.

Decision tree techniques are widely used to build classification models and these are easy to understand and closely resembles human reasoning. The paper emphasizes on basic decision tree issues and current reasearch points. Decision trees are sequential models, which combine together a sequence of simple tests and each test compares a nominal attribute against a set of possible values or a numeric attribute against a threshold value. Many programs have been developed that perform automatic induction (creation) of decision trees but they require a set of labelled instances. This article will cover the major theoretical issues, instructing or guiding the researcher in interesting research directions and giving the suggestions of possible bias combinations that have yet to be explored.

27) L. Breiman, "Random forests,"Machine Learning, vol. 45, pp. 5-32, 2001.

Random forests are a combination of tree predictors such that each of the tree depends upon values of a random vector sampled independently and with the same distribution for all trees in the forest. It is structurally defined as A random forest is a classifier consisting of a collection tree structured classifiers {h(x,Θk), k=1, ...} where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. The generalization error of a forest of tree classifiers depends upon the strength of each tree present in the forest and the correlation between them. Using a random selection of characteristics to split each node yields error rates that compare favorably to Adaboost but are more robust with respect to noise. Correlation, strength and internal estimates monitor error and these are used to represent the response to ascending the number of

characters used in the splitting. Internal estimates are also used to measure the variable importance. The following aspects and ideas are also applied or applicable to regression.

28) S. Haykin, Neural Networks-A Comprehensive foundation. Prentice Hall, 2 nd edition 1999.

This article covers different topics such areas as: Reinforcement learning/neurodynamic programming, support vector machines, dynamically driven current works. It exposes the reader to the many factors of neural networks and helps them explore the technology capabilities and potential applications, the detailed analysis of back-propagation learning and multi-layer perceptron. It gives ideas about the intricacies of the learning process of essential component for understanding neural networks. Considering the recurrent networks, such as Boltzmann machines, Hopfield fireworks and mean field theory machines and also the modular networks, temporal processing, and neurodynamic integrates the computer experiments throughout, giving the opportunity to review how the neural networks are designed and performed in practices. The article examines the use of neural networks as an engineering tool for signal processing applications. The aim is threefold: 1. To articulate a new philosophy in the approach 2. To statistical signal processing using neural networks to describe three case studies using real-life data that is non linearity, nonstationarity, and non gaussianity 3. To discuss mutual information as a criterion for designing unsupervised neural networks. 30.H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, "Support vector regression machines, "Advances in neural information processing systems, voi. 9, pp. 155-161, 1997.

29) In 1992 Vapnik and coworkers proposed a supervised algorithm for theclassification that has since evolved into what are now known as Support Vector Machines (SVMs):

A class of algorithms for classification, regression and other applications that replenishes the current state of the art in the field. Among the key innovations of this method were the explicit use of convex optimization, statistical learning theory, and kernel functions. A new regression technique based on Vapnik;s concept of support vectors is introduced. It compares the support vector regression (SVR) with a committee regression technique (bagging) based on regression trees and ridge regression done in feature space. On the basis of these experiments, it is awaited that SVR will have advantages in high dimensionality space because SVR optimization doesn't depend on the dimensionality of the input space.

III.PROPOSED SYSTEM

Following is the basic proposed system:

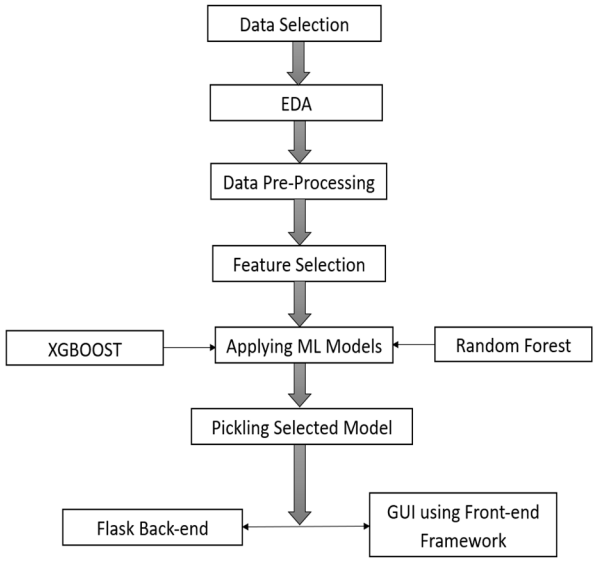


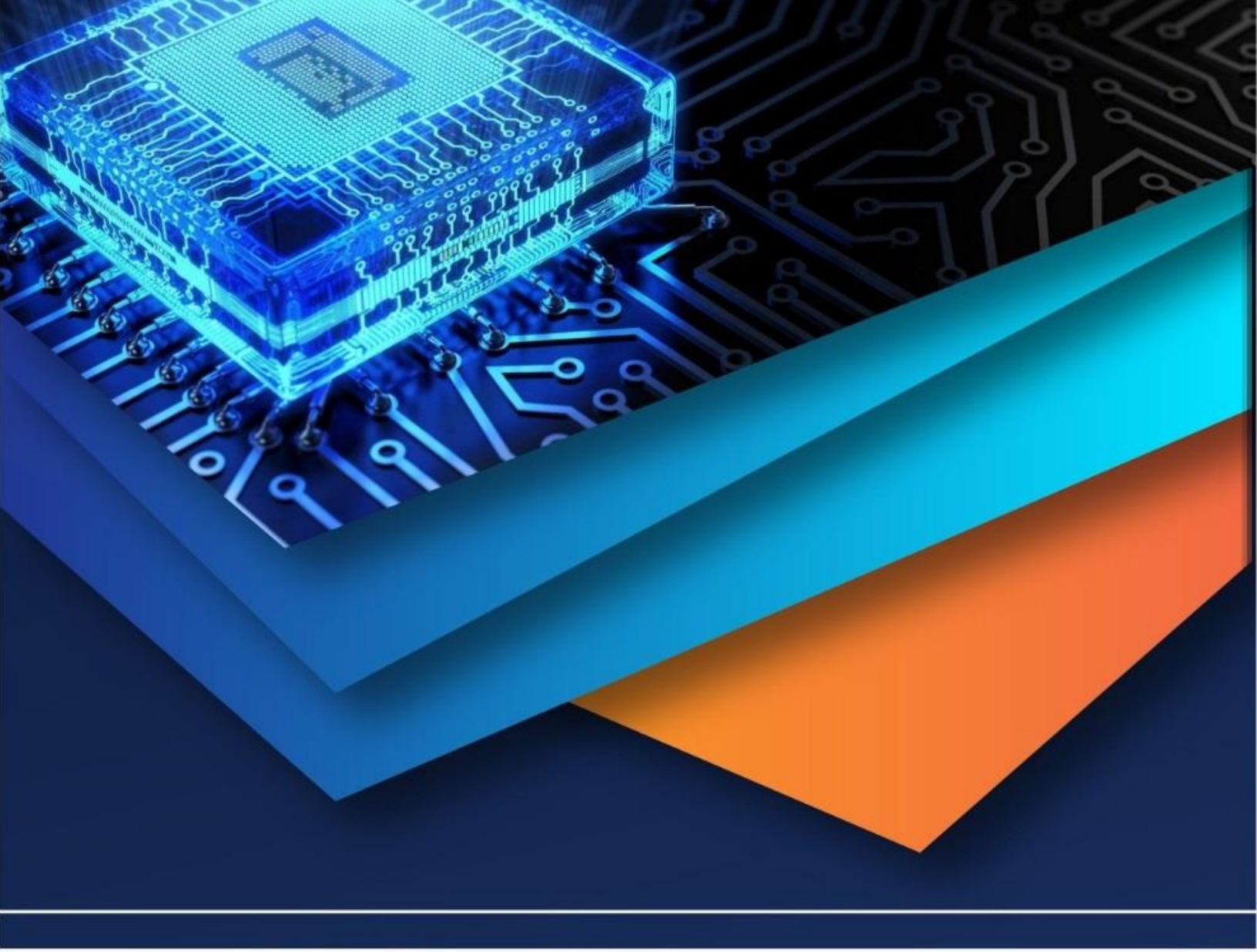
Fig. 1 Proposed System Diagram

IV. CONCLUSIONS

Machine Learning algorithms are applied on the dataset to predict the dynamic fare of flights. This gives the predicted values of flight fare to get a flight ticket at minimum cost. Data is collected from the websites which sell the flight tickets so only limited information can be accessed. The values of R-squared obtained from the algorithm give the accuracy of the model. In the future, if more data could be accessed such as the current availability of seats, the predicted results will be more accurate. Finally, we have created the entire process of predicting an airline ticket and given a proof of our predictions based on the previous trends with our prediction.

REFERENCES

[1] K. Tziridis T. Kalampokas G.Papakostas and K. Diamantaras "Airfare price prediction using machine learning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO .2017.8081365L. Li Y. Chen and Z. Li" Yawning detection for monitoring driver fatigue based on two cameras" Proc. 12th Int. IEEE Conf. Intel. Transp. Syst. pp. 1-6 Oct. 2009.
[2] William Groves and Maria Gini "An agent for optimizing airline ticket purchasing" in proceedings of the 2013 international conference on autonomous agents and multi-agent systems.
[3] J. Santos Dominguez-Menchero, Javier Rivera and Emilio TorresManzanera "Optimal purchase timing in the airline market".
[4] Supriya Rajankar, Neha sakhrakar and Omprakash rajankar "Flight fare prediction using machine learning algorithms" International journal of Engineering Research and Technology (IJERT) June 2019.
[5] Tianyi wang, samira Pouyanfar, haiman Tian and Yudong Tao "A Framework for airline price prediction: A machine learning approach"
[6] T. Janssen "A linear quantile mixed regression model for prediction of airline ticket prices"
[7] Wohlfarth, T.clemencon, S.Roueff "A Dat mining approach to travel price forecasting" 10th international conference on machine learning Honolulu 2011.
[8] Vinod Kimbhaune, Harshil Donga, Ashutosh Trivedi, Sonam Mahajan and Viraj Mahajan research paper on flight fare prediction system.
[9] W. Groves and M. Gini, ―An agent for optimizing airline ticket purchasing, ‖ 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), St. Paul, MN, May 06 - 10, 2013, pp. 1341-1342.
[10] Viet Hoang Vu, Quang Tran Minh and Phu H. Phung,‖An Airfare Prediction Model for Developing Markets‖, IEEE paper 2018.
[11] Wohlfarth, T. Clemencon, S.Roueff, ―A Dat mining approach to travel price forecasting‖, 10 th international conference on machine learning Honolulu 2011.
[12] Dominguez-Menchero, J.Santo, Reviera, ‖optimal purchase timing in airline markets‖ ,2014
[13] medium.com/analytics-vidhya/mae-mse-rmse-coefficient-ofdetermination-adjusted-r-squared-which-metric-is bettercd0326a5697e article on performance metrics
[14] www.keboola.com/blog/random-forest-regression article on random forest
[15] https://towardsdatascience.com/machine-learning-basics-decisiontree-regression-1d73ea003fda article on decision tree regression.
[16] O. Etzioni, R. Tuchinda, C. A. Knoblock, and A. Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price.
[17] Manolis Papadakis. Predicting Airfare Prices.
[18] Groves and Gini, 2011. A Regression Model for Predicting Optimal Purchase TimingFor Airline Tickets.
[19] Modeling of United States Airline Fares – Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DB1B), Krishna Rama-Murthy, 2006.
[20] B. S. Everitt: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge (3rd edition, 2006). ISBN 0-521-69027-7.
[21] Bishop: Pattern Recognition and Machine Learning, Springer, ISBN 0-387-31073-8.
[22] E. Bachis and C. A. Piga. Low-cost airlines and online price dispersion. International Journal of Industrial Organization, In Press, Corrected Proof, 2011.
[23] P. P. Belobaba. Airline yield management. an overview of seat inventory control. Transportation Science, 21(2):63, 1987.
[24] Y. Levin, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. Management Science, 55(1):32–46, 2009
[25] B. Smith, J. Leimkuhler, R. Darrow, and Samuels, ―Yield managementat american airlines,‖Interfaces, vol.22, pp. 8–31, 1992.
[26] T. Janssen, ―A linear quantile mixed regression model for prediction of airline ticket prices,‖ Bachelor Thesis, Radboud University, 2014.
[27] S.B. Kotsiantis, ―Decision trees: a recent overview,‖ Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.
[28] L. Breiman, ―Random forests, ‖ Machine Learning, vol. 45, pp. 5-32, 2001.
[29] S. Haykin, Neural Networks – A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999.
[30] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, ‖Support vector regression machines, ‖ Advances in neural information processing systems, vol. 9, pp. 155-161, 1997.



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)