Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

AIRLINE FARE PREDICTION USING MACHINE LEARNING

Mr. SNVASRK PRASAD, Assistant Professor, J. PAVITHRA, D. SHIVA, S. AJAYREDDY SHUBHANKAR HALDAR

SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY, HYDERABAD, Sheriguda (V), Ibrahimpatnam (M), Rangareddy Dist – 501 510

ABSTRACT

This paper discusses the issue of airfare. A set of characteristics defining a typical flight is chosen for this purpose, with the assumption that these characteristics influence the price of an airline ticket. Flight ticket prices fluctuate depending on different parameters such as flight schedule, destination, and duration, a variety of occasions such as vacations or the holiday season. As a result, having a basic understanding of flight rates before booking a vacation will undoubtedly save many individuals money and time. Analyzing 3 datasets to get insights about the airline fare and the features of the three datasets are applied to the seven different machine learning (ML) models which are used to predict airline ticket prices, and their performance is compared. The goal is to investigate the factors that determine the cost of a flight. The data can then be used to create a system that predicts flight prices

Index : flight schedule, holiday season, money and time, analyzing dataset using ml

1.INTRODUCTION

In today's world, airlines attempt to control flight ticket costs in order to maximize profits. Most people who fly regularly know the best times to buy cheap tickets. However, many customers who are not good at booking tickets fall into the discount trap set by the company, causing them to spend their money. The main goal of airline companies is to make a profit, while the customer is looking for the best purchase. Customers frequently aim to purchase tickets far in advance of the departure date in order to prevent price increases as the departure date approaches. Due to the great complexity of the fare models used by airlines, it is very difficult for a customer to buy an airline ticket at a very low price because the price is constantly fluctuating. Airlines can lower their ticket prices when they need to create a market and when tickets are harder to obtain. These tactics consider a number of financial, marketing, commercial, and social factors that are all linked to ultimate flight pricing. They might be able to get the most profit possible. As a result, costs may be influenced by various factors.

The price model used by airlines is so complex that prices fluctuate constantly, making it very difficult for customers to buy tickets at very low prices. Surveys of customers and airlines have grown steadily over the last two decades. From a customer point of view, it is an important question to establish a low price or a good time to buy a ticket. In this paper, we will be using the collected data from three different sources to build the models using Machine Learning algorithms. Customers can save millions of rupees by using the proposed method to get the information they need to order tickets at the proper moment

2. LITERATURE SURVEY

TITLE: "Robust Dynamic Pricing With Strategic Customers,"

ABSTRACT:

We consider the canonical revenue management (RM) problem wherein a seller must sell an inventory of some product over a finite horizon via an anonymous, posted price mechanism. Unlike typical models in RM, we assume that customers are forward looking. In particular, customers arrive randomly over time and strategize about their times of purchases. The private valuations of these customers decay over time and the customers incur monitoring costs; both the rates of decay and these monitoring costs are private information. This setting has resisted the design of optimal dynamic mechanisms heretofore. Optimal pricing schemes—an almost necessary mechanism format for practical RM considerations— have been similarly elusive.

TITLE: "Airline ticket price and demand prediction: A survey"

ABSTRACT: Nowadays, airline ticket prices can vary dynamically and significantly for the same flight, even for nearby seats within the same cabin. Customers are seeking to get the lowest price while airlines are trying to keep their overall revenue as high as possible and maximize their profit. Airlines use various kinds of computational techniques to increase their revenue such as demand prediction and price discrimination. From the customer side, two kinds of models are proposed by different researchers to save money for customers: models that predict the optimal time to buy a ticket and models that predict the minimum ticket price. In this paper, we present a review of customer side and airlines side prediction models. Our review analysis shows that models on both sides rely on limited set of features such as historical ticket price data, ticket purchase date and departure date. Features extracted from external factors such as social media data and search engine

query are not considered. Therefore, we introduce and discuss the concept of using social media data for ticket/demand prediction.

TITLE: "Data-driven Modeling of Airlines Pricing"

ABSTRACT: The popularity of travelling by airplanes is constantly growing. Much of existing research describe the global flight market. At the same time, Russian air market is characterized by its peculiarities that have to be identified to build proper models of airfare. The objective of this study is to analyze Russian air transportation market and compare the behavior of prices on local and global flights. Using these data, collected from two independent ticket price information aggregators (AviaSales and Sabre) for the period of spring-summer 2015, an empirical data-driven model was built for air prices prediction for different flight directions. We found that the form of price dependency on purchase earliness differs dramatically between local and international flights in two largest Russian cities (Moscow and Saint-Petersburg).

TITLE: "Airfare prices predictiono using machine learning techniques,"

ABSTRACT: This paper deals with the problem of airfare prices prediction. For this purpose a set of features characterizing a typical flight is decided, supposing that these features affect the price of an air ticket. The features are applied to eight state of the art machine learning (ML) models, used to predict the air tickets prices, and the performance of the models is compared to each other. Along with the prediction accuracy of each model, this paper studies the dependency of the accuracy on the feature set used to represent an airfare. For the experiments a novel dataset consisting of 1814 data flights of the Aegean Airlines for a specific international destination (from Thessaloniki to Stuttgart) is constructed and used to train each ML model. The derived experimental results reveal that the ML models are able to handle this regression problem with almost 88% accuracy, for a certain type of flight features.

TITLE: "A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter,"

ABSTRACT: Decision-making under uncertainty is one of the major issues faced by recent computer-aided solutions and applications. Bayesian prediction techniques come handy in such areas of research. In this paper, we have tried to predict flight fares using Kalman filter which is a famous Bayesian estimation technique. This approach presents an algorithm based on the linear model of the Kalman Filter. This model predicts the fare of a flight based on the input provided from an

observation of previous fares. The observed data is given as input in the form of a matrix as required to the linear model, and an estimated fare for a specific upcoming flight is calculated.

TITLE: "A regression model for predicting optimal purchase timing for airline tickets,"

ABSTRACT: Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principallybecause buyers have insufficient information for reasoning about future price movements. This paperpresents 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 specificroutes and dates based on a corpus of historical price quotes. Also, we apply our model to predict pricesof flights with specific desirable properties such as flights from a specific airline, non-stop only flights, ormultisegment flights. By comparing models with different target properties, buyers can determine thelikely cost of their preferences. We present the expected costs of various preferences for two high-volumeroutes. Performance of the prediction models presented is achieved by including instances of time-delayedfeatures, by imposing a class hierarchy among the raw features based on feature similarity, and by pruningthe classes of features used in prediction based on in-situ performance. Our results show that purchasepolicy guidance using these models can lower the average cost of purchases in the 2 month period priorto a desired departure. The proposed method compares favorably with a deployed commercial web siteproviding similar purchase policy recommendations.

TITLE: "Credit Card Fraud Detection Using Machine Learning,"

ABSTRACT:

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions. The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy [1], to aggregate the transaction made by the cardholders.

3. PROBLEM STATEMENT

Airlines can lower their ticket prices when they need to create a market and when tickets are harder to obtain. These tactics consider a number of financial, marketing, commercial, and social factors that are all linked to ultimate flight pricing. They might be able to get the most profit possible. As a result, costs may be influenced by various factors. The price model used by airlines is so complex that prices fluctuate constantly, making it very difficult for customers to buy tickets at very low prices. Surveys of customers and airlines have grown steadily over the last two decades. Regression machine learning models for airline ticket price prediction have been developed by [4]. Data from 1814 flights on a single international route was used in the development of this model, including departure and arrival times, bag allowance, and the number of free baggage allowances per flight. They used eight different regression machine learning models, which are Extreme Learning Machine (ELM), Multilayer Perceptron (MLP), Generalized Regression Neural Network, Random Forest Regression Tree, Regression Tree, Linear Regression (LR), Regression SVM (Polynomial and Linear), Bagging Regression Tree. The model produced the following performance results: The Bagging Regression is accurate to 87.42% and 85.91% accuracy for Random Forest Regression Tree.

3.1 DISADVANTAGES:

Increased Dependency: Visually-impaired individuals may become more dependent on others for assistance in identifying objects and navigating their environment of the limiting their independence and autonomy. Safety Risks: Without an object detection and recognition system, visually-impaired individuals may be more prone to accidents and injuries due to obstacles and hazards that they are unable to detect

4. PROPOSED SYSTEM & IT'S ADVANTAGES:

The proposed system aims to address the issue of airfare by analysing a set of characteristics that define a typical flight, assuming that these features significantly influence the price of an airline ticket. The fluctuation in flight ticket prices is attributed to various parameters, including flight schedule, destination, duration, and occasions such as vacations or holiday seasons. Data Collection: Gather a dataset comprising historical flight information, including departure and arrival locations, dates, times, airlines, ticket prices, and other relevant features. This dataset should cover a wide range of routes, airlines, and time periods to capture diverse patterns. This involves cleaning the data, handling missing values, encoding categorical variables, and possibly feature scaling or normalization. Create new features or transform existing ones that might better represent the

relationships between the input variables and the target variable (fare). For example, you might extract features such as day of the week, time of the day, distance between departure and arrival locations, and any seasonal trends. Choose appropriate machine learning algorithms for regression tasks. Common choices include linear regression, decision trees, random forests, gradient boosting methods (like XG Boost or Light GBM), and neural networks. Split the dataset into training and testing sets. Train the selected model(s) on the training data.

4.1 ADVANTAGES:

Improved Accuracy: Machine learning models can analyze vast amounts of historical data and complex patterns to make more accurate fare predictions compared to traditional methods. This can help both airlines and travelers make better-informed decisions regarding ticket prices. Dynamic Pricing: Airlines can leverage machine learning models to implement dynamic pricing strategies, adjusting fares in real-time based on factors such as demand, time until departure, competitor pricing, and seat availability. This flexibility can maximize revenue for airlines while offering competitive prices to travelers. Personalized Pricing: Machine learning algorithms can analyze individual traveler preferences, booking history, and browsing behavior to offer personalized fare recommendations. This can enhance customer satisfaction and increase loyalty by providing tailored pricing options.

5. IMPLEMENTATION

Machine learning introduces several techniques for predicting aircraft ticket pricing. Algorithms that we have used include:

• Linear Regression.

• K-Neighbor Regression.

• Support Vector Machine.

• Decision Tree.

• Random Forest.

These models have been implemented using the sci-kit learn python library. In order to verify the performance of these models, parameters such as R-square, MAE, MSE, and RMSE are used.

KNN Regression

A k-neighbor regression analysis gives the average of its k nearest neighbors. Like SVM, this is a non-parametric approach. The results are obtained using only a few values to get the best value. KNN is a supervised classification technique used as a regressor. It adds a new data point to the class. Since no assumptions are made, it is not parametric. It calculates the distance between each training example and a new data set. The model selects K elements from the data set that are near the new data point. The distance is calculated using the Euclidean distance, the Manhattan distance or the Hamilton distance.

Linear Regression

Linear regression is a supervised learning (ML) technique. It performs regression tasks. It is a linear model, assuming that there is a linear relationship between the input variable (x) and a single output variable (y). Y can be calculated by linear inclusion of input variables, especially (x). Because our data set contains many independent features that prices may depend on, we will use multiple linear regression (MLR) to estimate the relationship between two or more independent variables and a dependent variable.

Decision Tree Regression

A decision tree is a tree structure used to build regression or classification models. In addition, a decision tree is generated for each data set that is reduced in size. This generates solutions and leaf nodes. The decision tree selects independent variables from the dataset as decision nodes for making a decision. When test data is entered into the model, the result is determined by looking at which segment the data point belongs to. And the decision tree will output the average of all data points in the subsection of the section that the data point belongs to.

Random Forest Regression

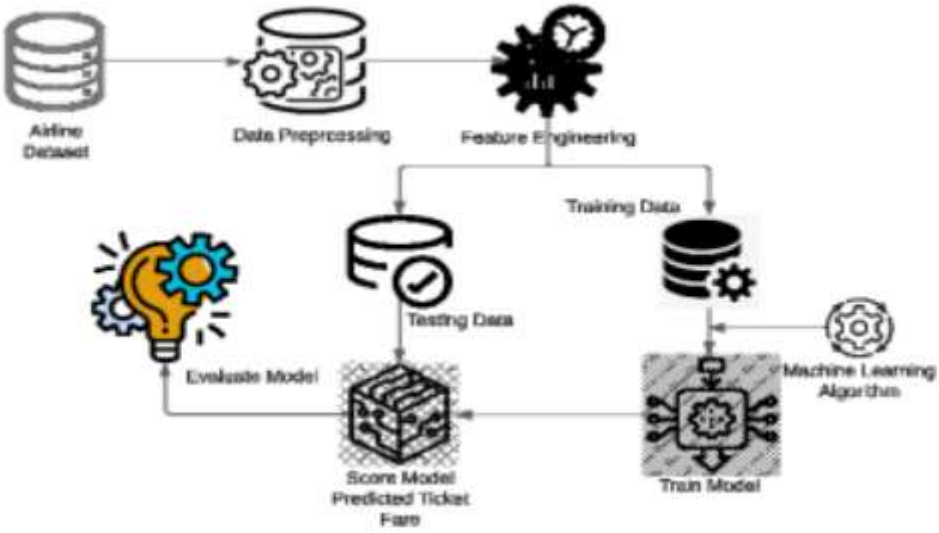
The random forest algorithm combines less accurate models to create more accurate models. It combines the base model with another model to create a larger model. The features are scanned and passed on to the trees without replacement in order to generate strongly uncorrelated decision trees. It is necessary to have a lower correlation between trees in order to choose the best split. The main principle that distinguishes the random forest from the decision tree is the aggregated uncorrelated trees. A random forest is an ensemble learning technique in which the training model uses a variety

of learning algorithms that are then combined to produce a final predicted result. When the output of the random forest model is examined, a random number of features and data sets will average the predicted values, which falls within the bagging area of ensemble learning.

Support Vector Machine

A support vector machine (SVM) is a supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space. There are two approaches to calculating the margin, or the maximum distance between classes, which are hard-margin classification and soft-margin classification

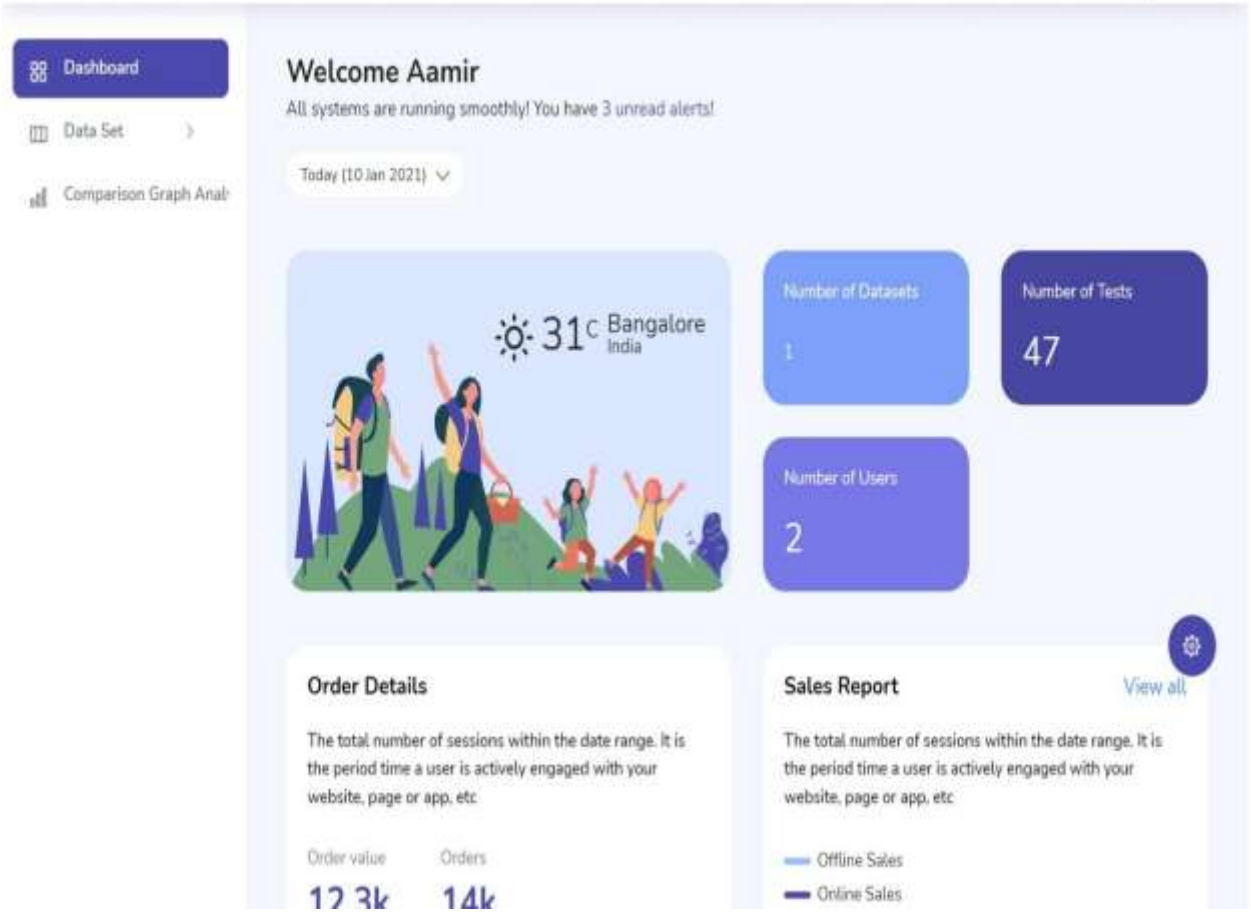
6. SYSTEM ARCHITECTURE

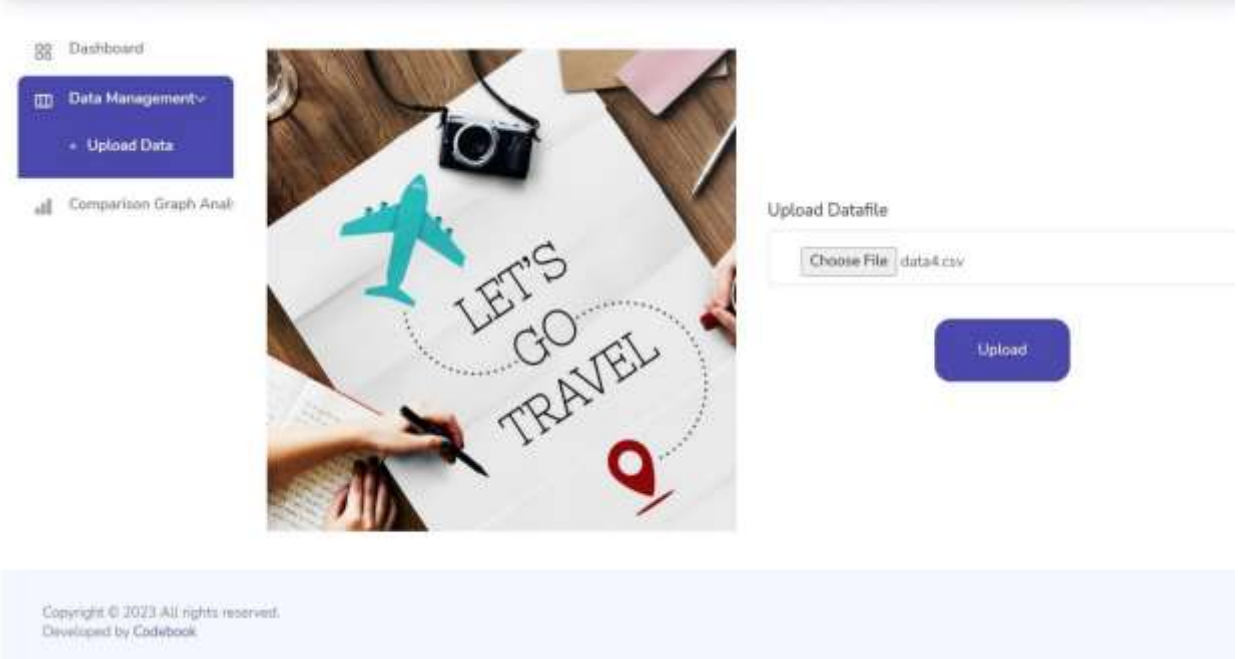


7. EXPECTED RESULTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024





Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

8. CONCLUSION

To estimate the dynamic fare of flights, three different datasets from three different sources have been used. Many insights have been found while visualizing the dataset. Seven different machine learning algorithms have been used to build the model. Only limited information can be obtained because data is acquired from websites that sell flight tickets. The correctness of the model is determined by the evaluation metrics table I values obtained from the procedure. The Random Forest Regressor outperformed the other algorithms with good accuracy. So, Random Forest Regressor works fine for predicting the airline fare price. If more data, such as actual seat availability, could be obtained in the future, the anticipated results would be more accurate. Prediction-based services are currently employed in a variety of sectors, including stock price predictor programs used by stock brokers and services like Zestimate, which provides an estimate of housing values. As a result, in the

aviation business, a service like this is required to assist clients in reserving tickets. There have been numerous studies conducted on this topic using various methodologies, and additional research is required to increase the accuracy of prediction utilizing various algorithms. To acquire more reliable findings, more accurate data with greater features might be employed

9 REFERENCES

[1] T. Janssen, "A linear quantile mized regression model for prediction of airline ticket prices," in A Treatise on Electricity and Magnetism 3rd ed., vol. 2, 2014, pp. 68- 73.

[2] Yiwei Chen and F. Vivek Farias, " Robust Dynamic Pricing With Strategic Customers," Mathematics of Operations Research 43, pp. 1119-1142, 2018.

[3] Prasad, D. C. G. V. N., Bhargavram, K., & Guptha, K. G. (2015). Challenging Security Issues of Mobile Cloud Computing. *IJRDO -Journal of Computer Science Engineering*, *1*(7), 33-44. https://doi.org/10.53555/cse.v1i7.931

[4] Lantseva, Anastasia, Mukhina, Ksenia, Nikishova, Anna, Ivanov, Sergey, Knyazkov and Konstantin, "Data-driven Modeling of Airlines Pricing," Procedia Computer Science, vol. 66, pp. 267-276, 2015.

[5] K. Tziridis, T. Kalampokas, G. A. Papakostas and K. I. Diamantaras, "Airfare prices predictiono using machine learning techniques," in 25th European Signal Processing Conference (EUSIPCO). Kos 2017, 2017.

[6] A. Boruah, K. Baruah, B. Das, M. Das and N. Gohain, "A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter," in Progress in Advanced Computing and Intelligent Engineering, Singapore, 2019, pp. 191-203.

[7] William Groves and Maria Gini, "A regression model for predicting optimal purchase timing for airline tickets.," Technical report, University of Minnesota, Minneapolis, USA, Report number 11-025, 2011.

[8] Narasimha Chary, CH.GVN Prasad. Humanoid Ai Robot: A Member of Our Next the Generation Family. Journal of Advancements in Robotics. 2024; 11(1): 20–24p.

[9] R. R. Subramanian, N. Akshith, G. N. Murthy, M. Vikas, S. Amara and K. Balaji, "A Survey on Sentiment Analysis," in 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, 2021.

[10] S. Amara and R. R. Subramanian, " Collaborating personalized recommender system and content-based recommender system using TextCorpus," in 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020.

[11] Andi and Hari Kirshnan, "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model," Journal of Soft Computing Paradigm 3, pp. 205- 217, 2021. 69

[12] CHOLLETI, N., & HIRWARKAR, T. (2020). BIOMEDICAL DATA ANALYSIS IN PREDICTING AND IDENTIFICATION CANCER DISEASE USING DUO-MINING. *Advances in Mathematics: Scientific Journal*, *9*, 3487-3495.

[13] V. Suma and Shavige Malleshwara Hills, "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics," Journal of Soft Computing Paradigm (JSCP) 2, pp. 101-110, 2020.

[14] W. K. Michael and A. G. Thomas, "A Framework for the Evaluation of Statistical Prediction Models," CHEST, vol. 158, no. 1, pp. S29-S38, 2020.

[15] L. Yuling and L. Zhichao, "Design and implementation of ticket price forecasting system," in AIP Conference Proceedings, 2018.

[16] Elizaveta Stavinova, Petr Chunaev and Klavdiya Bochenina, "Forecasting railway ticket dynamic price with Google Trends open data," Procedia Computer Science, vol. 193, pp. 333-342, 2021.

[17] S. Deepa, A. Alli, Sheetac and S. Gokila, "Machine learning regression model for material synthesis prices prediction in agriculture," in materialstoday, 2021.

[18] S. Matthew and Lewis, "Identifying airline price discrimination and the effect of competition," International Journal of Industrial Organization, vol. 78, 2021.

[19] R. Shobarani, R. Sharmila, M. N. Kathiravan, A. A. Pandian, C. Narasimha Chary and K. Vigneshwaran, "Melanoma Malignancy Prognosis Using Deep Transfer Learning," *2023*

International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169528

[20] Rian Mehta, Stephen Rice, John Deaton and Scott R. Winter, "Creating a prediction model of passenger preference between low cost and legacy airlines," Transportation Research Interdisciplinary Perspectives, vol. 3, 2019.