



# Forecast for Airline Fares Applying Machine Learning Techniques

**1.Palli Anitha 2. Mr.k.Chinna Nagaraju, 3.V Anil Santhosh**

- 1 M.Tech Scholar and student of C.S.E(Artificial Intelligence)., International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh,  
Anithapalli01@gmail.com
- 2 Associate Professor of C.S.E., International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh.
- 3 Professor and HOD of C.S.E., International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh.

## 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. Analysing 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.

**Key words:** Forecasting of Airline Fares, Machine Learning Techniques, Algorithms

## I. 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.

## II. LITERATURE SURVEY

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. : 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. 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 analyse Russian air transportation market and compare the behaviour of prices on local and global flights. Using these data, collected from two independent ticket price information aggregators (Avia Sales 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).

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. 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. 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. 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 semicrouches 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 flights. 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. Our results show that purchase policy guidance using these models can lower the average cost of purchases in the 2 month period prior to a desired departure. The proposed method compares favourably with a deployed commercial web site providing similar purchase policy recommendations. 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 from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifiers [3],[5],[6],[8] are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift [1]. In this paper, we worked with European credit card fraud dataset. With accelerated evolution of the internet as websites, social networks, blogs, online portals, reviews, opinions, recommendations, ratings, and feedback are generated by writers. This writer generated sentiment content can be about books, people, hotels, products, research, events, etc. These sentiments become very beneficial for businesses, governments, and individuals. While this content is meant to be useful, a bulk of this writer generated content require using the text mining techniques and sentiment analysis. But there are several challenges facing the sentiment analysis and evaluation process. These challenges become

obstacles in analysing the accurate meaning of sentiments and detecting the suitable sentiment polarity. Sentiment analysis is the subjective information from text. This paper presents a survey on the sentiment analysis challenges relevant to their approaches and techniques. Recommender systems aim to get the relevant data, based on the user's interests. One of the key problems of the recommender systems is to maintain the dataset and to retrieve the data, which is relevant to the user. A common solution is to track the user's preferences and showing the relevant results, however, it is a complex task in terms of time and space. The user data need to be analysed and learnt using efficient algorithms. To address this problem, we have proposed a method to format the data in the dataset using POS-taggers using NLTK framework. In this paper, we have proposed a user-profile model which uses this tagging mechanism to provide better recommendations compared to the existing state-of-the-art recommender techniques. This paper discusses, trying to accurately assess the price of Bitcoin by looking at different parameters affects the value of Bitcoin. In our work, we focus on understanding and seeing the evolution of Bitcoin daily market, and gaining intuition in the most relevant aspects surrounding the Bitcoin price. In the meantime, market capitalization of publicly traded cryptocurrencies exceeds \$ 230 billion. The most important cryptocurrency, Bitcoin, is used primarily as a digital value store, and its pricing opportunities have been extensively considered. These features are described in more detail in the following paragraph: details of the main Bitcoin, as described in the paper. Bitcoin is the most expensive digital currency in the market. However, Bitcoin prices have been highly volatile, making it difficult to forecast. As a result, the goal of this research is to find the most efficient and accurate model for predicting Bitcoin prices using various machine learning algorithms. Several regression models with scikit-learn and Keras libraries were tested using 1-minute interval trading data from the Bitcoin exchange website bit stamp from January 1, 2012 to January 8, 2018. The best results showed a Mean Squared Error (MSE) as low as 0.00002 and an R- Square (R2) as high as 99.2 percent.

### III. SYSTEM ANALYSIS

Convolutional Neural Networks (CNNs) are often used in deepfake detection structures, and pre-educated Inception Net models may be used. system improvement calls for an expansion of datasets for trying out and training. a few structures consist of temporal consistency checks, audio evaluation, and facial landmarks to enhance detection accuracy. In these kinds of initiatives, open-source frameworks and libraries like PyTorch and TensorFlow are frequently applied. The model may be implemented in a video processing pipeline to discover deepfakes in actual time. because deepfake era techniques are usually converting, it's far imperative to do ongoing research and improvement. The system's overall performance may be better and delicate with using cutting-edge assets and cooperation with issue-count number professionals. 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.

The proposed system for airline fare prediction leverages advanced machine learning algorithms and data analytics to overcome the limitations of traditional rule-based systems. By moving beyond static pricing rules, the system aims to adapt dynamically to changing market conditions and traveller behaviours. A key improvement involves the integration of a comprehensive set of features, including flight schedule, destination, duration, seasonal trends, and special occasions, enabling a more accurate representation of the factors influencing airfare. The system prioritizes personalized predictions by considering individual traveller preferences, booking history, and specific requirements. Real-time updates are implemented to ensure responsiveness to fluctuations in demand, external events, and other dynamic factors. A user-friendly interface empowers travellers with the ability to input preferences, view predictions, and gain insights into the various factors contributing to the forecasted fare. The proposed system seeks to provide a more transparent, adaptive, and user-centric solution for predicting airline fare prices, enhancing the overall experience for travellers and optimizing cost-saving opportunities.

#### 1. Enhanced Accuracy:

Leveraging advanced machine learning algorithms allows the system to learn from historical data patterns and adapt to evolving market conditions. This results in more accurate predictions of airline fare prices, considering a broader range of influencing factors.

## 2. Comprehensive Feature Set:

The system incorporates a comprehensive set of features, including flight schedule, destination, duration, and various other relevant parameters. This ensures a more holistic analysis of factors affecting airfare, leading to improved prediction accuracy.

## 3. Personalized Predictions:

By considering individual traveller preferences, booking history, and specific requirements, the system provides personalized fare predictions. This level of customization enhances the user experience and enables travellers to make more informed decisions aligned with their unique needs.

## 4. Real-Time Adaptability:

The integration of real-time updates allows the system to adapt swiftly to changes in demand, supply, and external market conditions. This responsiveness ensures that the predictions remain relevant and accurate, even in dynamic and rapidly changing situations.

## 5. User-Friendly Interface:

The system is designed with a user-friendly interface, allowing travellers to input their preferences easily, view predictions, and gain insights into the factors influencing fare prices. This user-centric approach enhances the overall experience, making it more interactive and accessible for users with varying levels of expertise.

## IV. SYSTEM DESIGN

### SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

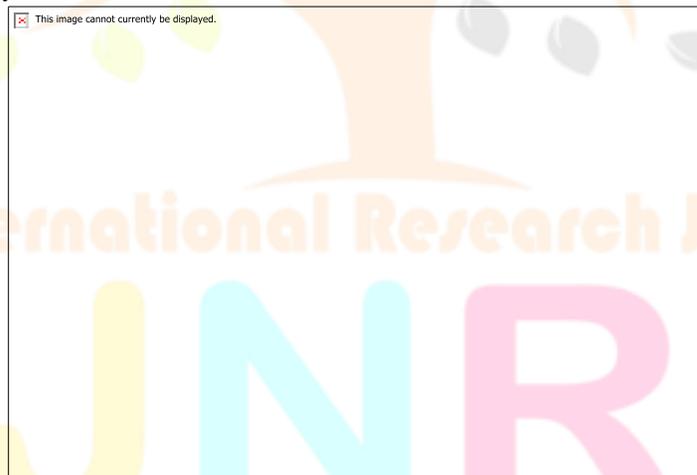


Fig 1. Methodology followed for proposed model

## V. SYSTEM IMPLEMENTATION

### MODULES

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

- KNN Regression.
- Linear Regression.
- Lasso Regression.
- Ridge Regression.
- Decision Tree Regression.
- Stacking Tree Regression.
- Random Forest Regression.

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.

#### A. KNN Regression

A k-neighbour regression analysis gives the average of its k nearest neighbours. 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.

#### B. 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.

#### C. Lasso Regression

Lasso regression takes precedence over other regression approaches for more accurate predictions. This model employs shrinkage. When reduced, the data value is reduced to the centre point, the so-called average value. We recommend a simple, sparse model (that is, a model with few parameters). This type of regression is best suited for models with a large number of multicollinearities, or for automating aspects of the model selection process such as variable selection and parameter deletion. The lasso regression employs the L1 regularization method. Since the function is selected automatically, it is used when there are many functions.

#### D. Ridge Regression

Ridge regression is a data modelling technique used to eliminate multicollinearity. Ridge regression is the most appropriate technique when there are fewer observations than predictor variables. Ridge regression constraint variables have a circular shape when plotted, while the LASSO plot has a diamond shape.

#### E. 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.

#### F. Stacking Regression

Stacking regression is a technique for improving prediction accuracy by creating linear combinations of multiple predictors. Cross-validation data and the least squares method are used to determine the coefficients of the combination under non-negative requirements. In this case, we used Ridge Regressor, Lasso Regressor, and KNN Regressor as regressors, and Decision Tree Regressor as metal regressors.

#### G. 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.

## VI. RESULTS AND DISCUSSION

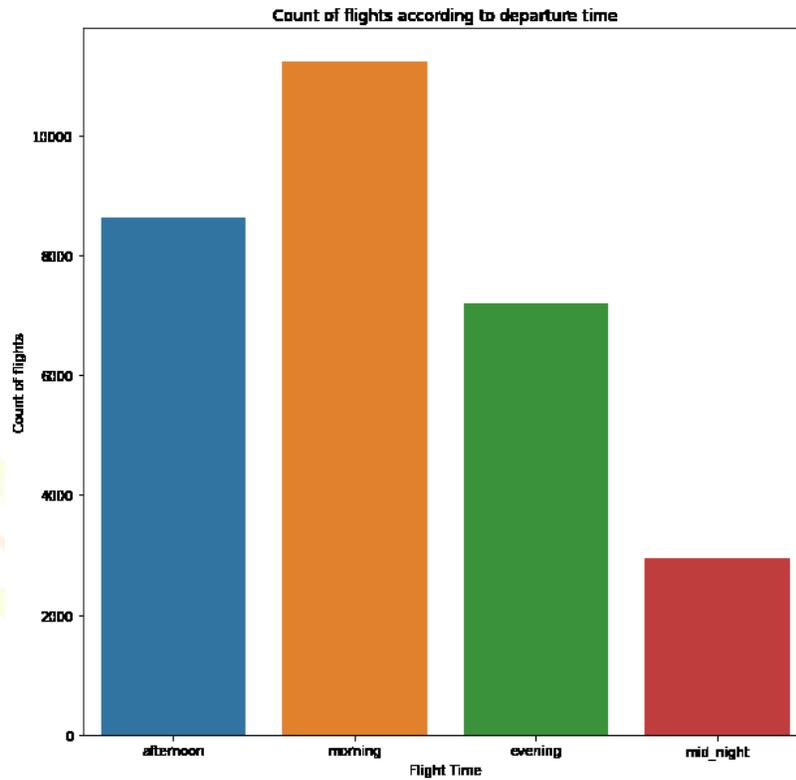
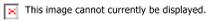


Fig. 4, Number of flights available at different times in a day.

The count plot figure (4) shows that many flights are leaving in the morning, which means the availability of flights in the morning is greater, and the availability of flights in the afternoon and evening is moderate, but the availability of flights in the mid-night is less. This concludes that there are very few flights that travel in the mid-night.

This image cannot currently be displayed.

## COMPARISION OF EVALUATION METRICS

## VII. CONCLUSION AND FUTUREWORK

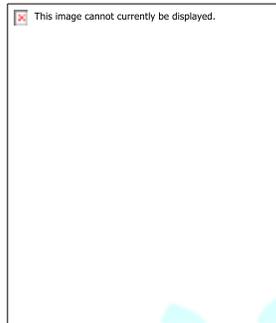
To estimate the dynamic fare of flights, three different datasets from three different sources have been used. Many insights have been found while visualizing the dataset. Seven different machine learning algorithms have been used to build the model. Only limited information can be obtained because data 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.

### REFERENCES :

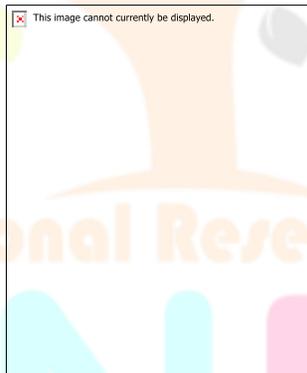
[1] T. Janssen, "A linear quantile mized regression model for prediction of airline ticket prices," in A Treatise on Electricity and Magnetism 3rd ed., vol. 2, 2014, pp. 68-73.

- [2] Yiwei Chen and F. Vivek Farias, "Robust Dynamic Pricing With Strategic Customers," *Mathematics of Operations Research* 43, pp. 1119-1142, 2018.
- [3] Juhar Ahmed Abdella, Nazar Zaki, Khaled Shuaib and Fahad Khan, "Airline ticket price and demand prediction survey.," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 4, pp. 375-391, 2021.
- [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 prediction 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. 19203.
- [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] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," in *5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021.
- [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.
- [12] Manoharan and J. Samuel, "Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm," *Journal of Soft Computing Paradigm (JSCP)* 3, pp. 83-95, 2021.
- [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] Ismail Koc and Emel Arslan, "Dynamic ticket pricing of airlines using variant batch size interpretable multivariable long short-term memory," *Expert Systems with Applications*, vol. 175, 2021.
- [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, 2019.020

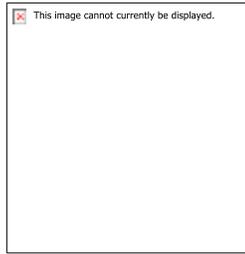
## Biography of authors:



**Palli Anitha** was a M.Tech Scholar and student of student of C.S.E(Artificial Intelligence), International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh **Anitha** is a dedicated research scholar specializing in Data Science, Python and Machine Learning (ML), focusing on innovative approaches to solve complex real-world problems. With a strong academic foundation and a passion for computational technologies.



**k.Chinna Nagaraju** was an Associate Professor of C.S.E., International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh. **k.Chinna Nagaraju** is a dedicated research scholar specializing in Artificial Intelligence (AI) and Machine Learning (ML), focusing on innovative approaches to solve complex real-world problems. Their research interests include developing advanced algorithms for predictive modelling, integrating hybrid ML-DL frameworks, and exploring the ethical and societal impacts of AI systems.



**V Anil Santhosh** was an Assistant Professor and HOD of C.S.E., International School of Technology and Sciences for Women(Autonomous), East Gonagudem, Rajanagaram–Andhra Pradesh. V Anil Santhosh is a dedicated research scholar specializing in Artificial Intelligence (AI) and Machine Learning (ML), focusing on innovative approaches to solve complex real-world problems. Their research interests include developing advanced algorithms for predictive modelling, integrating hybrid ML-DL frameworks, and exploring the ethical and societal impacts of AI systems. Their work primarily focuses on applications in renewable energy forecasting, natural language processing, and computer vision.

