IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 3, May 2022

Flight Fare Prediction using Random Forest Algorithm

Prof. Ms. Archana Dirgule[1], Shubham Agarwal[2], Ram Agrawal[3], Neha Singh[4], Kiran Adsul[5]

Faculty, Department of Computer Engineering[1]

Students, Department of Computer Engineering[2,3,4,5]

Sinhgad College of Engineering, Pune, Maharashtra, India

Abstract: *Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillance broadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition flight schedule, and airport information. The designed prediction tasks contain different classification tasks and a regression task. Experimental results show that long short-term memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem.*

Keywords: Flight delay prediction, ADS-B, machine learning, LSTM neural network, random forest.

I. Introduction

Everyday consumers buy cloud services as it is quite advantageous in several directions. The choice of service is much more preferred for a customer is sometimes difficult to recognize, because the purchasing decision is not only based on price but also a number of factors that might be considered before going to make the final commitment. There are several factors that are actually affecting the purchase behavior of customers. These include cultural, social, personal decision elements, etc. All of these have effected directly or indirectly on buying behavior.

Several machine learning algorithms have been proposed to analyze customer behavior. Although customers do not follow any predefined rules before deciding to purchase any service, we can predict which is the most probable service customer might buy [1]. And for this, we first need to identify the buying pattern of other customers, and if a new customer's buying pattern matches the previous ones; it can be able to predict the decision of the new customer [2-3]. And if the purchase decision can be identified beforehand, companies can able to provide a better customer experience by recommending his/her preferred services.

II. Literature Survey

M Gopal et al: Different from various machine learning methods targeting at direct fare or price prediction, we constructed a state predictor of class seats by applying a Naïve Bayes algorithm based on Multinomial Event Model on the core flight reservations inventory big data, to tell the probability of class availability within the next several hours or days. Four fundamental models and one integrated model are developed to propose an optimal decision to the airfare search engine layer, which makes the engine be capable of forecasting a smart buy-or-wait suggestion to customers. In our experimental route from SHA to TYO, the integrated model reaches an average of 95.42% accuracy.

Dipika H. Zala et al: This paper proposes a novel application based on two public data sources in the domain of air transportation: the Airline Origin and Destination Survey (DB1B) and the Air Carrier Statistics database (T-100). The proposed framework combines the two databases, together with macroeconomic data, and uses machine learning algorithms to model the quarterly average ticket price based on different origin and destination pairs, as known as the market segment. The framework achieves a high prediction accuracy with 0.869 adjusted R squared score on the testing dataset.

Bahl et al: This paper describes 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 behavior, can be next used in order to help the customer to decide when to purchase her/his ticket.

P. S. Maya Gopal et al: In this paper, The Bourta algorithm was used to select the most essential features for crop yield prediction. Crop area, canal length, number of open wells, number of tube wells, number of tanks, maximum temperature, average temperature, nitrogen, phopers, and potash fertilizers, sun radiation, and seed rate have all been identified as essential factors. These characteristics are fed into the MLR model to determine accuracy. By incorporating these features into the model, it was able to reach an accuracy of 84%.

K.Rajini et al: In this paper, By merging the IB1 learner and the A1DE updatable learner, a novel ensemble of heterogeneous incremental classifiers has been presented. The methodology was created primarily to accurately predict ART outcomes. For the ART data set used in the study, the suggested model accomplishes this goal with a ROC area of 94.1. Several ART data sets and several test choices are used to evaluate the model's efficacy. For the same data set, the proposed ensemble outperforms the others.

To see if the suggested model is generalizable to other domains, it is compared to additional benchmark data sets and shown to perform better.

J.-Y. Hsieh et al: In this Paper, we tried to solve the prediction problem by using the Machine Learning as well as Neural Network methods. The classification model was built based on the climate and historical data. This classifier's projected results have a 72 percent accuracy rate. Increasing the number of microclimate parameters can improve prediction accuracy, according to this study.

J. Camargo et al: In this work, the importance of feature selection in raising accuracy levels to above 95% was a crucial conclusion of this study. On average, accuracy scores were 3.8 percent lower without feature removal. Feature selection offers the added benefit of allowing for faster real-time computations with fewer features. Chow-Liu trees were discovered to be a feature selection method that allows for the rapid selection of a strong collection of features with a minimal number of iterations and comparable accuracy to that of a forward selection strategy.

R.Rajasheker Pullanagari et al: In this paper, the viability of employing RF–RFE to combine hyperspectral, topography, and soil data to obtain pasture quality parameters of heterogeneous pasture was investigated in this work. Because many environmental and management factors influence pasture quality, our findings revealed that combining hyperspectral data with commonly accessible environmental characteristics (elevation, slope angle, and soil type) enhanced prediction accuracy when compared to hyperspectral data alone.This finding also revealed that by selecting the most sensitive factors throughout the spectrum and environmental data, RF-RFE significantly improved estimations of pasture quality (RPD = 2.11–2.35). Elevation, slope, and soil type were found to be key determinants in predicting CP, while the same variables were found to be significant in predicting ME, with the exception of elevation.

F. Balducci et al: The research presented in this paper introduces practical, low-cost, and simple-to-implement tasks that can help an agricultural company increase productivity while also deepening the study of the smart farm model; technological progress in a field that requires control and optimization can truly contribute to conserving natural resources, adhering to business and international laws, meeting consumer needs, and pursuing economic profits. Machine learning and more traditional statistical techniques were used to utilize the three separate data sources, with a special focus on the IoT sensors dataset. The first task demonstrates that a neural network model can forecast total apple and pear crops on the Istat dataset with a success rate of nearly 90%, while the second task demonstrates that polynomial predictive and regression models are better suited for the CNR scientific data due to the nature of the dataset.

M. Lango et al: Two sorts of methodological contributions are presented in our work. To begin, we propose incorporating a random set of attributes into this ensemble. This approach enhances G-mean and sensitivity measurements for greater dimensional complex datasets, according to experiments. We've also noticed that, unlike the original RBBag, it increases the diversity of component classifiers and that employing a larger number of components improves classification. Second, for several imbalanced classes, we have presented an extension of Roughly Balanced Bagging that uses the multinomial distribution to estimate cardinalities of class examples in bootstrap samples. The undersampling version of our proposed

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 3, May 2022

Impact Factor: 6.252

Multiclass RBBag enhances G-mean and is better than the oversampling variation and simpler multi-class classifiers, according to testing with synthetic and real datasets

III. Methodology/Mathematical Model

Let us consider S as a system for crop yield prediction system.

S=

INPUT: Identify the inputs

F= f1, f2, f3 ....., FN— F as set of functions to execute commands.

I= i1, i2, i3—I sets of inputs to the function set

O= o1, o2, o3.—O Set of outputs from the function sets,

S= I, F, O

I = Input

O = Output

F = Functions implemented to get the output

Space Complexity:

The space complexity depends on Presentation and visualization of discovered patterns.

More the storage of data more is the space complexity.

Time Complexity:

Check No. of patterns available in the datasets= n

If (n(1)) then retrieving of information can be time consuming. So the time complexity of this algorithm is O (nˆn).

= Failures and Success conditions.

Failures:

1. Huge database can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

Success:

1. Search the required information from available in Datasets.
2. User gets result very fast according to their needs.

IV. Conclusion

Researchers from the marketing and CRM fields make a lot of significant contributions to customer purchase behavior prediction for traditional business. In this proposed algorithm, it has been investigated several key factors that have an impact on the purchase decision making of customers in the cloud environment, including different cloud service policies of customers based on their interest. Furthermore, by exploiting previous customer's history and the locations they belong to, this proposed method has been used to quantify the strength of these factors.

There exists an association-initiated cloud service purchase and it can be exploited to predict the needs of customers. Experiments in this paper favor our point of view. Accordingly, experimental results show that associations between categories of cloud services can significantly improve the predictive performance. A Customer Purchase Behavior Prediction Model has been developed to predict the cloud service a customer will buy in the future. The experimental results prove that this approach is feasible and provides real-time analysis of customer behavior. Experimental results demonstrate that online advertisements play important role in customer buying decisions as by analyzing the online activities (ex. Advertisements visited by customers), the prediction of purchase behavior becomes quite easy.

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 3, May 2022

REFERENCES

[1]. J. Stavins, "Price discrimination in the airline market: The effect of market concentration," Review of Economics and Statistics, vol. 83, no. 1, pp. 200–202, 2001.

[2]. B. Mantin and B. Koo, "Dynamic price dispersion in airline markets," Transportation Research Part E: Logistics and Transportation Review, vol. 45, no. 6, pp. 1020–1029, 2009.

[3]. P. Malighetti, S. Paleari, and R. Redondi, "Has ryanair's pricing strategy changed over time? an empirical analysis of its 2006–2007 flights," Tourism Management, vol. 31, no. 1, pp. 36–44, 2010.

[4]. T. H. Oum, A. Zhang, and Y. Zhang, "Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets," Journal of Transport Economics and Policy, vol. 7, no. 2, pp. 171–192, 1993.

[5]. B. Burger and M. Fuchs, "Dynamic pricing – A future airline business model," Journal of Revenue and Pricing Management, vol. 4, no. 1, pp. 39–53, 2005.

[6]. T. M. Vowles, "Airfare pricing determinants in hub-to-hub markets," Journal of Transport Geography, vol. 14, no. 1, pp. 15–22, 2006.

[7]. K. Rama-Murthy, "Modeling of united states airline fares– using the official airline guide (OAG) and airline origin and destination survey (DB1B)," Ph.D. dissertation, Virginia Tech, 2006.

[8]. B. Derudder and F. Witlox, "An appraisal of the use of airline data in assessing the world city network: a research note on data," Urban Studies, vol. 42, no. 13, pp. 2371–2388, 2005.

[9]. A. Mottini and R. Acuna-Agost, "Deep choice model using pointer networks for airline itinerary prediction," in the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017, pp. 1575–1583.

[10]. K. Tziridis, T. Kalampokas, G. A. Papakostas, and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," in the 25th IEEE European signal processing conference, 2017, pp. 1036–1039.

[11]. Y. Chen, J. Cao, S. Feng, and Y. Tan, "An ensemble learning based approach for building airfare forecast service," in the IEEE international conference on big data, 2015, pp. 964– 969.

[12]. T. Liu, J. Cao, Y. Tan, and Q. Xiao, "ACER: An adaptive context-aware ensemble regression model for airfare price prediction," in the international conference on progress in informatics and computing, 2017, pp. 312–317.

[13]. V. Pai, "On the factors that affect airline flight frequency and aircraft size," Journal of Air Transport Management, vol. 16, no. 4, pp. 169–177, 2010.

[14]. M. S. Ryerson and H. Kim, "Integrating airline operational practices into passenger airline hub definition," Journal of Transport Geography, vol. 31, pp. 84– 93, 2013.

[15]. H. Baik, A. A. Trani, N. Hinze, H. Swingle, S. A