
Air passengers forecasting for Australian airline based on hybrid rough set approach

H. KUMAR SHARMA, K. KUMARI AND S. KAR

Abstract

Accurate and reliable air passenger demand is very important for policy-making and planning by tourism management as well as by airline authorities. Therefore, this article proposed a novel hybrid method based on rough set theory (RST) to construct decision rules for long-term forecasting of air passengers. Level (mean) and trend components are first estimated from the air passengers time series data using DES model in the formulation of the proposed hybrid method. Then the rough set theory is employed to combine the output of DES model and generated decision rules is used to forecasting air passengers. We compare the proposed approach with other time series models using a corrected classified accuracy (CCA) criterion. For the empirical analysis, yearly air transport passenger from 1992 to 2004 is used. Empirical results show that the proposed method is highly accurate with the higher corrected classified accuracy. Also, forecasting accuracy of the proposed method is better than the other time series approaches.

Mathematics Subject Classification 2010:

Keywords: Air passengers, Forecasting, Hybrid approach, Rough set, DES model, Grey model

1. INTRODUCTION

Over the past decade, international tourism demand has become one of the most rapidly growing industries in the world. International air travel demand is often calculated regarding the number of air passenger arrivals (Lim and McAleer, 1999). For the past many years, the Australian foreign tourism demand has been growing rapidly. Moreover, total tourist arrivals have been calculated approximately 2.04 to 2.85 million from 2006 and 2012 by 39.82% growth rate during this year (Australian Bureau of Statistics, 2013a). Also, Brisbane, Melbourne, Perth and Sydney are the four key international airports in Australia by 92.73% of the monthly arrivals between 2006 and 2012, (Australian Bureau of Statistics, and 2013b). Australia has exhibited a 5.8% growth in international tourist arrivals in 2012 (United Nations World Tourism Organisation, 2014).

Consequently, highly accurate and consistent air passenger forecasts play a significant role in policy-making and planning by the private airport and airline management as well as by government tourism authorities and operators. Long-term forecasting of air passenger supplies a key input into decisions of daily operation management, including aircraft scheduling decisions, maintenance planning, advertising and sales campaigns, the opening of new sales offices, etc. More highly reliable and accurate forecasting tools are required to plan effectively. Consequently, this study focuses on the long-term forecasting of air passengers in Australian airline.

Over the past many years different univariate and multivariate time series techniques have been developed in air passenger forecasting (Samagaio and Wolters, 2010), Nguyen et al. (2013), Apergis et al. (2017). They include autoregressive integrated moving average (ARIMA) (Lim and McAleer (1999, 2001, 2002), Cho (2003), Coshall (2006) and Tsui and Balli (2016), naive and smoothing models (Martin and Witt (1989)), Law and Au (1999), Law (2000)) to forecast non-stationary data along with trend component. In particular, Xie et al. (2014) employed a hybrid approach based on decomposition and least squares support vector regression approaches. Pal and Kar (2017) applying forecast technique in the stock exchange case study. The empirical results showed that the forecasting performance was improved when the hybrid approaches of decomposition were implemented. Although, previous commonly used time series models are based on the statistical assumptions and have huge limitations such as large sample, normal distribution, and non-stationary time series data. To overcome the conditions of statistical techniques, modern soft computing method has been introduced such as rough set analysis by Pawlak (1982). Rough set theory has been consistently applied in tourism demand modeling and forecasting for the extraction of decision rules, particularly for the data on the Hong Kong tourism scenario by Law and Au (1998, 2000), Au and Law (2000) and Goh and Law (2003). Also, several variants of the rough set approach have also emerged in the tourism studies by Goh et al. (2008), Lin (2010), Xiaoya and Zhiben (2011) and Celotto et al. (2012). Moreover, Li et al. (2011) analyzed and predicted tourism in Tangshan city of China through the rough set model.

In recent times, various studies on forecasting have introduced a comparative study of rough set and other quantitative time series models. For example, Faustino et al. (2011) present a rough set analysis of electrical charge demand in the United States and the level of the Sapucal river in Brazil. Liou (2016) used the rough set theory to study the airline service quality to Taiwan. Sharma and Kar (2018) applied rough set theory in hotel industry. These authors found that the rough set is more consistent with other forecasting techniques. Most published article on airline industry is primarily based on the statistical models. However, in previous literature the application of hybrid rough set and DES techniques for air passengers forecasting has not been adopted, to the best our knowledge. Consequently, the purpose of this paper is to propose a new approach which applied rough set theory to generate decision rules for the air passengers forecasting. Based on the yearly time series of transportation data for all passengers in Australian airline from 1992 to 2004, empirical analysis is employed to compare the proposed hybrid approaches with other forecasting methods using a corrected classified accuracy.

The primary objective of the paper is to provide accurate and highly reliable forecasting of air transport passengers. The forecasting performance has been evaluated under different criteria. The contribution of the study is that the rough set is firstly applied in exponential smoothing time series modeling approach. The present article can be employed to provide the essential recommendation on tourism policy for the developing countries.

The remaining of the article is structured as follows. Section hybrid approach based on rough set proposed the hybrid method for long-term forecasting of air transport passengers. Section data describe the data. The next section illustrates the empirical analysis related to the study. Section comparison of different models compares the accuracy of different models. Finally, the last section discusses the summary, conclusions and future work.

2. HYBRID APPROACH BASED ON ROUGH SET

In this section, the proposed methodology is discussed. Firstly, DES and rough set techniques are briefly presented. Then the hybrid approach is set up for investigating the decision rules and several steps involved in their implementation are illustrated in details.

2.1. Double exponential smoothing model

Double exponential smoothing (DES) model has been proposed for level and trend estimation process (Holt, 1957). A significant advantage of smoothing model is that the future prediction is obtained using previously estimated the level and trend in time series data with some appropriate smoothing parameters, μ , and ρ , respectively which lie in the range 0 and 1. The estimated level and trend of DES model for time series X_t ($t=1, 2, \dots, T$) is given by:

$$L_t = \mu(X_t) + (1 - \mu)(L_{t-1} + T_{t-1})$$

$$T_t = \rho(L_t - L_{t-1}) + (1 - \rho)T_{t-1}$$

The future value of the series X_t for h period is obtained as

$$\hat{X}_{t+h} = L_t + hT_t$$

2.2. Rough set theory

The rough set is a new mathematical approach to capturing imprecision, vagueness, and uncertainty (Pawlak, 1982). For the evaluation of a vague description of the member RST is the excellent mathematical tool. The adjective vague express the information quality that is uncertainty or ambiguity that chase from information granulation. The indiscernibility relation developed in this manner is a mathematical foundation of the RST; it induces a separation of the universe into pieces of indiscernible (similar) objects, named elementary set. The main concept of the rough set model is based on indiscernible (IND) relation, two approximations set, lower and upper approximation of a set.

Let Y be the non-empty finite set of objects referred to as universe and A be a nonempty finite set of attributes, then $S = (Y, A)$ is called an information system where C, D are two subsets of A , where C and D are condition and decision attribute, respectively. For $S = (Y, A)$ and $P \subseteq A$, $W \subseteq Y$ can be approximated based on the knowledge having in P by assembling the P -lower and P -upper approximation of W , represents by $\underline{P}(W)$ and $\overline{P}(W)$ respectively; where

$$\underline{P}(W) = \{x | [x]_P \subseteq W\}$$

$$\overline{P}(W) = \{x | [x]_P \cap W \neq \emptyset\}$$

A set P is called an elementary set if it having set of all indiscernible(similar) object with respect to particular attributes. $[x]_P$ is an equivalence class that contain x with respect to indiscernibility relation P . The objects in $\underline{P}(W)$ is known as the set of all members of Y which can be surely classified as a member of W in the knowledge P whereas objects in $\overline{P}(W)$ is the set of all elements of Y that can be probably classified as a member of W involving knowledge P . The boundary region of W is expressed as: $BN_P(W) = \overline{P}(W) - \underline{P}(W)$ is the set of a member which cannot decisively classify into W consisting knowledge P . If the boundary region of an exact set is the empty set if lower approximation and upper approximation set are similar. In the adverse case, if the boundary region contains some objects than the set W is referred as rough set concerning P .

The fundamental idea of RST is the creation of decision rules which is based on the IF and THEN logical statement. Decision rules are used to preserve the core semantics of the feature set from the provided information of particular problem which is an additional significant aspect of RST.

2.3. Proposed hybrid model

Since the combination method yields, better results than a single method, the modeling, and forecasting approach with high accuracy is adopted in this study. There are three main steps involved in the proposed hybrid method, i.e., estimation, single

forecast, and combination. After the parameters are estimated by DES model as X_t , L_t and T_t respectively, they are utilized for the making of prediction (\widehat{X}_t). Then the estimated and predicted output is combining using the rough set method. The overall procedures of deriving a hybrid model are as described in Figure 1 as the several steps:

STEP 1: Suppose the original time series is X_t (actual) at time t , and then the level (L_t) and trend (T_t) has been estimated using DES model to establish the forecast of air passengers.

STEP 2: According to the estimated and forecast (\widehat{X}_t) output of DES model, four attributes are formulated to ascertain an information system (data set). Then, the data set are normalized (Mahapatra and Sreekumar, 2010). All attributes have been categorized into three classes (Low, Average and High). The information system (IS) for rough set is as follows:

IS = (X_t , L_t , T_t , \widehat{X}_t), where, X_t , L_t , T_t and \widehat{X}_t are condition and decision attributes, respectively.

STEP 3: In the next stage, an information system is used to derive the decision rules for all passengers.

To create the IF and THEN decision rules, information system has been analyzed by applying RST technique (Cheng et al., 2010). The decision rules are as follows:

IF X_t = [Low, Average, High] AND L_t = [High, Average, Low] AND T_t = [Average, High, Low] THEN \widehat{X}_t = [Low, Average, High].

3. DATA

In this study, time series data is used from tourism demand according to the objective and the availability of the data. Our empirical study uses yearly air transportation data for all passengers (millions) in Australian airline from the period of 1992 to 2004, with a total of 13 observations, as shown in Figure. 2. Data are obtained from <https://www.otexts.org/fpp/7/2>. The R-3.0.3 software is used for the overall empirical analysis of DES and Grey models. The rough set analysis is implemented through Rough Set Data Explorer (ROSE2) software (Predki, 1998). In our data analysis, training dataset (in-sample) is used to estimate the parameters using the time series models. Testing sample is used to evaluate the forecasting performance. In time series modeling, pre-samples are used for making out-of-sample forecasting by using different models.

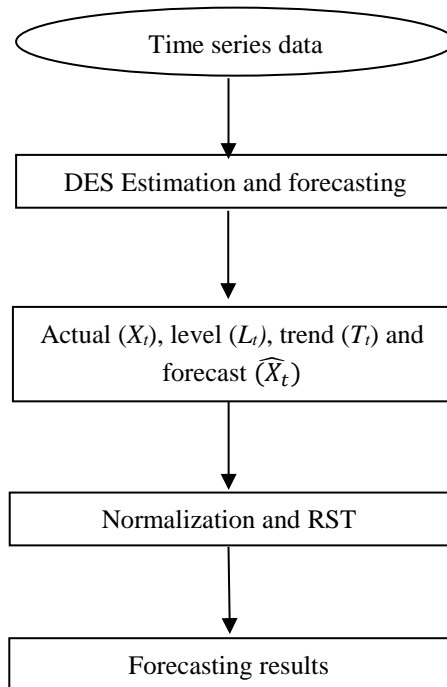


Fig. 1. The framework of DES and hybrid rough set approaches.

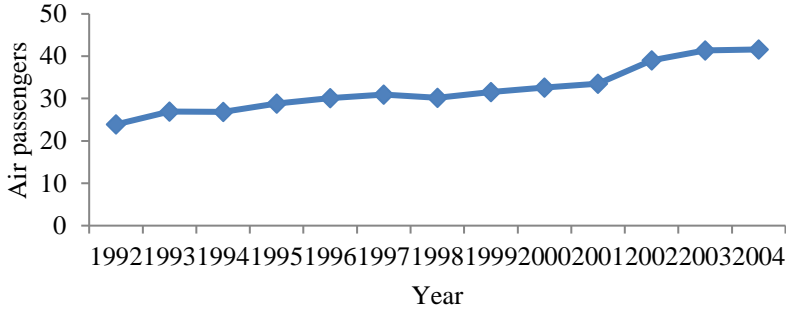


Fig.2. Yearly air passengers in Australia (in millions).

4. EMPIRICAL ANALYSIS

In the hybrid model, DES is first estimated into parameters level (L_t) and trend (T_t) from the air passenger's data (X_t). Then these parameters are used to predict the air passengers (\hat{X}_t) and these actual, estimated and predicted results of the DES model are combined via rough set. In addition, Grey model (GM) is applied for the prediction of air passengers.

In the DES forecasting, the values of μ and ρ parameters are first determined using 10-fold cross-validation grid search method within the range of $[0.8, 0.2]$ and then level (mean) and trend is estimated using these smoothing parameters. The calculated initial level and trend are 41.93, 1.87.

Apart from single DES model, other single time series model, GM is used for forecasting comparison purpose. In the GM model, forecasts are generated by using the solution of linear differential equation (LDE) of order one. The LDE for GM model is given below

$$\frac{dx_t}{dt} + \alpha Y_t = \beta \quad (1)$$

Where, X_t ($t=1,2,...,T$) is a time series data and α and β are parameters. Also, the parameters α and β are calculated by applying the ordinary least squares (OLS) estimation method (Hsu and Wen, 1998; Zhou and Hu, 2008). The final solution of LDE in terms of forecasting equation for GM model for h periods is written as

$$\widehat{X}_t(h) = (1 - e^\alpha) \left(X_1 - \frac{\beta}{\alpha} \right) e^{-\alpha(h-1)}, \quad h=1, 2, \dots$$

Now, the results of estimated parameters, α and β for GM model are -0.04447, 21.82630.

After the DES estimation and forecasting of air passengers, we create an information system via actual (X_t), level (L_t), trend (T_t) and (\widehat{X}_t) as a condition and decision attributes. Then the normalization process is used for the data transformation. Now, the rough set model is used to develop the IF and THEN decision rules, expressed as a rule model (RM) for forecasting of passengers time series data. Four decision rules with their support (S) are described in Table 1. These 4 rules are used to predict air passengers to Australia. R_2 has the highest support. It indicates that R_2 is the strongest rule for the prediction.

Table 1: Decision rules for a hybrid approach

No.	Rules	S
R ₁	IF [$0 \leq \text{actual} \leq 0.3$] then forecast = low	4
R ₂	IF [$0.3 < \text{actual} \leq 0.7$] then forecast = average	6
R ₃	IF [$0.3 < \text{level} \leq 0.7$] and [$0 \leq \text{trend} \leq 0.3$] then forecast = average	5
R ₄	IF [$\text{level} > 0.7$] then forecast = high	2

5. COMPARISON OF DIFFERENT MODELS

To compare the forecasting performance of time series models, we select corrected classified accuracy (CCA) criterion of measuring directional prediction accuracy. CCA is the ratio of corrected classified and the total number of observations (Nassiri and Rezaei, 2012). In general, accuracy is expressed as a percentage of the corrected direction, as follows:

$$CCA = \frac{\sum_{i=1}^n D_i}{n} * 100$$

$$\text{where } D_i = \begin{cases} 1, & (X_t - X_{t-1})(\widehat{X}_t - \widehat{X}_{t-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where, D_i is the corrected direction, n is sample size, X_t is the actual value and \widehat{X}_t is the forecasted value of time series data in a year. Since CCA is the measures of corrected direction between actual and forecasted values. Therefore, the higher CCA denotes highly accurate forecasting performance of time series models.

For forecasting performance, corrected classified accuracy (CCA) for three models are 92.30%, 92.30%, 100%, respectively. According to the CCA results, DES and GM model has equal forecasting performance. Also, the forecasting performance of the hybrid method is highly accurate than DES and GM models.

Empirical analysis of the study reveals that the forecasting accuracy of the hybrid model is superior to that of the individual models. The forecasting reports indicate that the prediction accuracy of the proposed approach is highly accurate to that of the single DES and Grey models. Therefore, our empirical findings reveal that the combine hybrid technique has a better prediction capacity compared with that of the individual time series models. The comparative analysis of actual and predicted values and out-of-sample forecasts from 2005 to 2017 are shown in Figure 3.

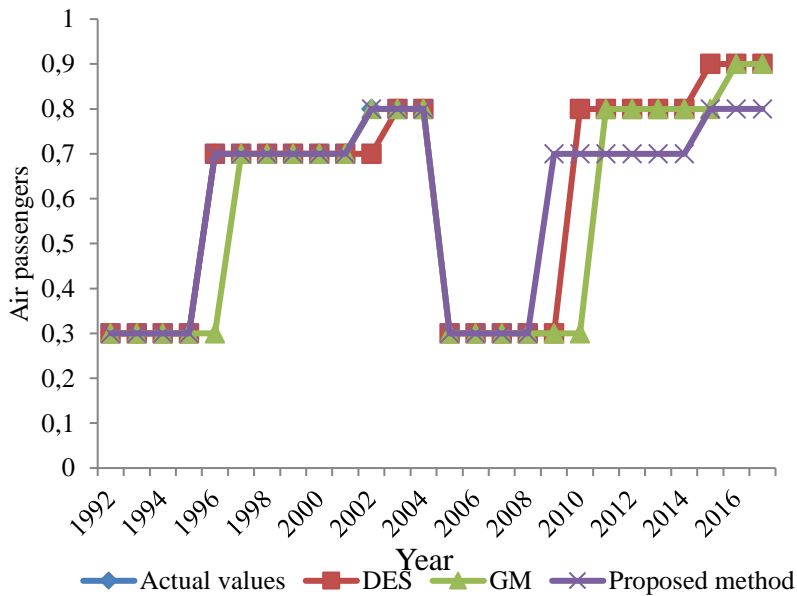


Fig. 3. Out-of-sample forecasts for different models.

6. CONCLUSIONS AND FUTURE WORK

In this study, we proposed a hybrid approach for long-term forecasting of air passenger on the basis of double exponential smoothing (DES) and rough set model. To illustrate the proposed hybrid technique, an empirical analysis has been performed and the proposed technique was compared with other time series methods using the air passengers in Australian airline. It uses corrected classified accuracy (CCA) criterion to compare the performance of hybrid method with other forecasting models.

Accurate and consistent air transport passenger is essential for policy making and planning by tourism management. Generally, tourism management is a significant source to enhance the foreign tourism demand as well as international travel to Australia and its various individual states. Therefore, the principal aim of this work is to evaluate the impact of tourism demand on Australian airline yearly air transport passengers. This work is also contributing that the combination of DES and rough set

model is firstly applied for long-term forecasting of air passenger. The forecasting reports indicate that the prediction accuracy of the proposed approach is superior to that of the single DES and Grey models. Our analysis suggests that the combined hybrid technique is an efficient method for air passengers forecasting. It is important to utilize the quality and nature of air passenger for better forecasting performance. The present article can be employed to provide the essential recommendation on tourism policy for the Australian airline. Furthermore, experimental analysis of air passengers would allow airport management to assign resources effectively for the unique requirements of various visitor classes, such as holiday visitors, visiting family and relatives, business tourist, and traveler guests etc. It is expected to employ other time series and soft computing models for air passengers forecasting for the future research using the data of airline industry.

REFERENCES

- Au, N. and Law, R. (2000). The application of rough sets to sightseeing expenditures, *Journal of Travel Research*, 39(1), 70-77.
- Australian Bureau of Statistics (2013a). Overseas arrivals and departures, Australia. Available at: <http://www.abs.gov.au> (accessed 10 November 2014).
- Australian Bureau of Statistics (2013b). Airport traffic statistics. Available at: <http://www.abs.gov.au> (accessed 10 November 2014).
- Australian Bureau of Statistics (2013c). Gross state product, chain volume measures and current prices. Available at: <http://www.abs.gov.au> (accessed 10 November 2014).
- Apergis N A, Mervar A and Payne JE (2017). Forecasting disaggregated tourist arrivals in Croatia: Evidence from seasonal univariate time series models. *Tourism Economics* 23(1): 78-98.
- Chen, M.-Y., Chen, B.-T. (2014). Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform, *Applied Soft Computing*. 14(B), 156-166
- Cho V (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management* 24(3): 323-330.
- Coshall J (2006). Time series analysis of UK outbound travel by air. *Journal of Travel Research* 44(3): 335-347.
- Cheng, C. H., Chen, T. L. and Wei, L. Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*. 180, 1610-1629.
- Celotto, E., Ellero, A. and Ferretti, P. (2012). Short-medium term tourist services demand Forecasting with rough set theory, *Procedia Economics, and Finance*, 3, 62-67.
- Deng, J. (1982). Control problems of Grey systems. *Systems and Control Letters*. 1(1), 288-294.
- Faustino, P.C., Pinheiro, A.C., Carpinteiro, A.O. and Lima, I., (2011). Time series forecasting through rule-based models obtained via rough sets. *Artif Intell Rev*. 36, 299-310.

-
- Goh, C., Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management*. 24, 511-517.
- Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Pittsburgh PA: Carnegie Institute of Technology.
- Hsu, C. I., Wen, Y. H. (1998). Improved Grey prediction models for the trans-pacific air passenger market. *Transportation Planning and Technology*. 22, 87-107.
- Law, R. and Au, N. (1998). A rough set approach to hotel expenditure decision rules induction, *Journal of Hospitality and Tourism Research*, 22(4), 359-375.
- Law R and Au N (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management* 20(1): 89-97.
- Lee, C.-C., Wan, T.-J., Kuo, C.-Y., Chung, C.-Y. (2007). Modified Grey Model for Estimating Traffic Tunnel Air Quality, *Environmental Monitoring and Assessment*, 132(1-3): 351-364.
- Lim C and McAleer M (1999). A seasonal analysis of Malaysian tourist arrivals to Australia. *Mathematics and Computers in Simulation* 48(6): 573-583.
- Law, R. and Au, N. (2000). Relationship modeling in tourism shopping: a decision rules induction approach, *Tourism Management*, 21(3), 241-249.
- Law R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management* 21(4): 331-340.
- Lim C and McAleer M (2001). Monthly seasonal variations: Asian tourism to Australia. *Annals of Tourism Research* 28(1): 68-82.
- Lim C and McAleer M (2002) Time series forecasts of international travel demand for Australia. *Tourism Management* 23(4): 389-396.
- Lin, L. (2010). Study on the evaluation of tourist satisfaction in tourism destination based on variable precision rough set, *International Conference On Computer Design and Applications*, 1, 124-128.
- Li, J., Feng, L., and Zhou, G. (2011). Travel demand prediction in Tangshan City of China based on rough set, *Springer-Verlag Berlin Heidelberg*, 440-446.
- Liou, J. J. H., Chuang, Y. C., Hsu, C. C. (2016). Improving airline service quality based on rough set theory and flow graphs. *Journal of Industrial and Production Engineering*. 33(2), 123-133.
- Li, J., Feng, L., Zhou, G. (2011). Travel demand prediction in Tangshan City of China based on rough set, *Springer-Verlag Berlin Heidelberg*, 440-446.
- Martin A and Witt F (1989). Forecasting tourism demand: A comparison of the accuracy of several quantitative methods. *International Journal of Forecasting* 5. 1-13.
- Mahapatra, S., Sreekumar, S. S. (2010). Attribute selection in marketing: A rough set approach. *IIMB Management Review*. 22, 16-24.
- Nassiri, H., Rezaei, A. (2012). Air itinerary choice in a low-frequency market: A decision rule approach. *Journal of Air Transport Management*. 18, 34-37.
- Nguyen L T, Shu, HM, Huang, FY and Hsu M. B. (2013). Accurate forecasting models in predicting the inbound tourism demand in Vietnam. *Journal of Statistics and Management Systems* 16(1): 25-43.
- Pal, S. S., & Kar, S. (2017). Time series forecasting using fuzzy transformation and neural network with back propagation learning, *Journal of Intelligent & Fuzzy Systems*, 33 (1): 467 - 477.
- Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Science*. 11, 341- 356.
- Predki, B., Wong, S. K. M., Stefanowski, J., Susmaga, R., Wilk, Sz., (1998). ROSE-software implementation of the rough set theory. In L. Polkowski, A. Skowron (Eds.). *Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence. Berlin: Springer*. 605-608.

Software R, version R-3.0.3. <https://cran.r-project.org/bin/windows/base/old/3.0.3/>

Samagaio, A., Wolters, M. (2010). Comparative analysis of government forecasts for the Lisbon Airport. *Journal of Air Transport Management*. 16, 213-217.

Sharma, H. K., & Kar, S. (2018). Decision Making for Hotel Selection using Rough Set Theory: A case study of Indian Hotels. *International Journal of Applied Engineering Research*, 13(6), 3988-3998.

Tsui W. H. K. and Balli F. (2016). International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure *Tourism Economics* 1-26.

Xie, G., Wang, S., Lai, K. K. (2014). Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. *Journal of Air Transport Management*. 37, 20-26

Xiaoya, H. and Zhiben, J. (2011). Research on econometric model for domestic tourism income based on rough set, *Springer*, 259-266.

Zhou, Z. J., Hu, C. H. (2008). An effective hybrid approach based on Grey and ARMA for forecasting gyro drift. *Chaos, Solitons & Fractals*. 35, 525-529.

Haresh Kumar Sharma

Department of Mathematics, National Institute of Technology Durgapur,
West Bengal 713 209, India

Email: hks.14ma1106@phd.nitdgp.ac.in

Kriti Kumari

Department of Mathematics, National Institute of Technology Durgapur,
West Bengal 713 209, India

Email: kriti.kri89@gmail.com

Samarjit Kar

Department of Mathematics, National Institute of Technology Durgapur,
West Bengal 713 209, India

Email: samarjit.kar@maths.nitdgp.ac.in