Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 1, April 2025, pp. 517~526
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i1.pp517-526

Advanced tourist arrival forecasting: a synergistic approach using LSTM, Hilbert-Huang transform, and random forest

Harun Mukhtar[1,2], Muhammad Akmal Remli[2,3], Mohd Saberi Mohamad[4], Khairul Nizar Syazwan Wan Salihin Wong[2], Farhan Ridhollah[1], Deprizon[5], Soni[1], Muhammad Lisman[5], Hasanatul Fu'adah Amran[1], Sunanto[1], Edi Ismanto[6]
[1]Faculty of Computer Science, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia
[2]Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Kota Bharu, Malaysia
[3]Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Kota Bharu, Malaysia
[4]Department of Genetics and Genomics, College of Medical and Health Sciences, United Arab Emirates University Abu Dhabi, Abu Dhabi, United Arab Emirates
[5]Faculty of Islamic Studies, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia
[6]Departement of Informatics Education, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia

Article Info

Article history:

Received Jun 27, 2024
Revised Oct 11, 2024
Accepted Oct 30, 2024

Keywords:

Data
Feature selection
Google trends
Hybrid
LSTM
Random forest

ABSTRACT

An advanced synergistic approach for forecasting tourist arrivals is presented, integrating long short-term memory (LSTM), Hilbert-Huang transform (HHT), and random forest (RF). LSTM is leveraged for its capability to capture long-term dependencies in sequential data. Additional data from Google Trends (GT) is processed with HHT for feature extraction, followed by feature selection using the RF algorithm. The combined HHT-RF-LSTM model delivers highly accurate forecasts. Evaluation employs regression analysis with metrics such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean square error (MSE), highlighting the effectiveness of this innovative approach in predicting tourist arrivals. This methodology provides a robust framework for handling limited datasets and improving forecast reliability. By incorporating diverse data sources and advanced preprocessing techniques, the model enhances prediction performance, demonstrating the strong performance of RF in feature selection.

Corresponding Author:

Muhammad Akmal Remli
Faculty of Data Science and Computing, Universiti Malaysia Kelantan
City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia
Email: akmal@umk.edu.my

1. INTRODUCTION

Deep learning (DL) is considered a core technology of the fourth industrial revolution (4IR or Industry 4.0). This technology is a branch of machine learning (ML) and artificial intelligence (AI). DL, derived from artificial neural networks (ANN), has become a hot topic in computing due to its excellent and dynamic capabilities [1]. DL is an advanced ML technique used for extensive data collection, pattern recognition, and prediction [2]. Data from the Indonesian Central Bureau of Statistics (BPS) is often limited, leading to overfitting in DL models. This study addresses this challenge by using feature selection (FS) with the random forest (RF) algorithm to improve accuracy by selecting the most relevant features, reducing data size, and minimizing computational complexity [3]. Additionally, feature selection shortens computation time [4] by removing irrelevant features and retaining only the essential ones [5]. In addition to feature selection, this research also addresses data cleaning. Data cleaning is performed using the Hilbert-Huang transform (HHT) to remove noise from Google Trends (GT) data, as search engine data often contains noise that must be filtered out [6]. Tourism stakeholders recognize that tourism products are fragile and tourist arrivals serve as a benchmark for their utilization. A high number of tourists with limited products can create major issues, making accurate forecasts of tourist arrivals essential for decision-makers [7].

COVID-19 pandemic has adversely affected tourist arrivals in Indonesia. In 2022, it is predicted that tourist arrivals will continue to decline due to the rapid spread of the new Omicron variant [8]. The impact of COVID-19 has not only been felt in Bali and Yogyakarta but also in Riau. In 2020, COVID-19 transmission in Riau was very high [9]. However, by 2023, BPS data indicates that COVID-19 cases in Riau have dropped to zero, providing new hope for the tourism sector. Several previous studies have explored tourism forecasting. For example, Lu *et al* [10] examined daily tourism flows using LSTM combined with convolutional neural networks (CNN). Meanwhile, Wu *et al*. [11] employed a hybrid approach combining SARIMA and LSTM to estimate daily tourist arrivals. Mukhtar *et al*. [12] researched write arrivals using LSTM and HHT with data sourced from GT and others. DL is also applied in other forecasting, such as COVID-19 [13], Room rates [14], Sunspot number time series [15], stock market [16]. Temür *et al*. [17] used DL to predict real estate sales.

2. RELATED WORKS

Lu *et al*. [10] studied daily tourism flows using LSTM with CNN optimized by a genetic algorithm (GA), achieving significant improvements. Wu *et al*. [11] used a hybrid SARIMA-LSTM approach to estimate daily tourist arrivals, outperforming other methods. Mukhtar *et al*. [12] explored tourist arrivals with LSTM and HHT using GT data, achieving satisfactory results. LSTM often overfits due to its need for extensive data [18] used GT to increase data volume, but its noise can reduce accuracy. Höpken *et al.* [19] explored web browsing data, requiring extra cleaning. Li *et al.* [20] applied HHT for data refinement to enhance LSTM forecasting accuracy. Additionally, feature selection (FS) was performed with GT data noise reduction using the RF algorithm for effective classification [21]. Salamanis *et al*. [22] introduced a novel DL approach in tourism, leveraging LSTM for robust predictions aimed at enhancing hotel revenue and operational management by integrating exogenous data. Laaroussi *et al.* [23] affirmed DL as an accurate predictor capable of evaluating non-linear relationships without the limitations of traditional time series models.

3. METHOD

This research develops a forecasting model using the DL method which is optimized with the HHT and RF methods to produce optimal forecasting. The methods used include. Figure 1 explains the process of carrying out this research consisting of three phases, the discussion of which is cleaning data noise, feature selection, and forecasting by improving LSTM-based deep learning.

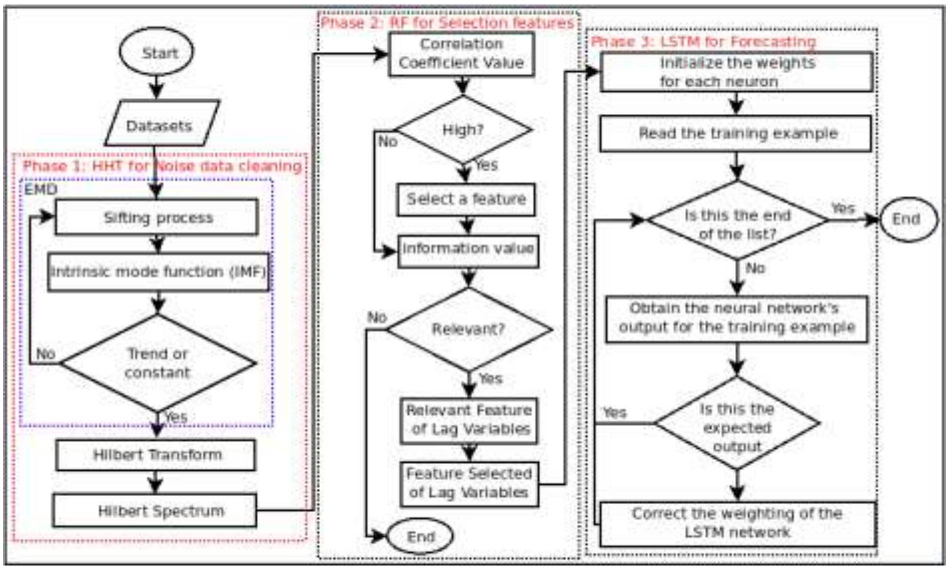

Figure 1. Research framework

3.1. Phase 1: HHT for noise data cleaning

GT can be used if the data cleaning process is carried out correctly. HHT is used to clean data from interference applied in this study [24]. The cleaning step starts from the first one, converting data into signals. Second, make changes to the data. The third is separating the residues, and decoding the signal (EMD). The fourth is doing repetition to produce intrinsic mode function (IMF), and the fifth is setting standard deviation limits to get results in the form of modulation, amplitude and frequency. Algorithm 1 shows the pseudo-code of these steps.

Algorithm 1. Noise processing algorithm [25]

```
Input: data stream X = {x1, x2, ...., xi, ...}; New X; New X1.
Output: Predicted result O
X <-------- x1, x2, ...., xi, ... //Input data
O {} //Output data
New X <-------- y1 + delay
New X1 <-------- New X + New X1 Combination
x(t) <-------- h1
For h1 <-------- x(t) - m1 do
input <-------- h1(k - 1), m1k
target <-------- h1k
end
For r1 <-------- x(t) - h1k do
input <-------- x(t), h1k
target <-------- r1k
end
```

3.2. Phase 2: RF for selection features

The feature selection method using RF is capable of improving classification accuracy [26], thereby making it applicable in forecasting [27]. This study employs five crucial steps in feature selection for forecasting tourist arrivals. First, the time series dataset comprises three components: HHT, resulting from the cleaning process of GT data; BPS, representing actual data; and GT, sourced from the internet. Second, it renders the dataset stationary for analysis. Third, it creates a dataset correlation matrix. Fourth, it calculates the feature importance scores. Fifth, it determines the selected features' impact on the dataset. Algorithm 2 presents the pseudocode outlining the workflow of the feature selection process.

Algorithm 2. Future selection algorithm

```
Input: Xa = {BPS, GT, NC}
Output: predictive features O
Xa
O {}
Create XaS with rST
Select FbC with OV
Convert Xa into Slp
Input F = Ire
Target <--- Lag
End
```

Feature selection is used to select which features are relevant [28]. The feature selection results can increase the accuracy of all experiments [29]. Algorithm 2 describes the workflow of data-driven feature selection. The data used are BPS, GT, and HTT with the results as shown in Algorithm 2. The first step is cleaning the GT from noise (NC). Second, creating a stationary dataset (XaS), Third, removing seasons and trends (rST), Fourth, selecting features based on factors of interest (Slp), Fifth, selecting the most predictive features of the 4 featured features.

3.3. Phase 3: LSTM for forecasting

This concept is utilized in recurrent neural networks (RNN), which have evolved into LSTM networks to achieve superior results [30]. The key advantage of LSTM lies in its ability to handle long-term dependencies within input sequences. This method is particularly suited for forecasting time series data due to its capability with sequential and step-by-step data [31]. LSTM generates two outputs, represented as 0 and 1, known as gates. A gate performs a dot multiplication operation: 1 allows information to pass through, while 0 blocks information. LSTM includes three types of gates: the forget gate selects information to discard, the two input gates update selected information, and the three output gates determine the final results, which are binary values of 0 or 1.

This study combines RF with LSTM to improve predictive results. RF is used to select only relevant features. Step by step is carried out in the order of feature selection, followed by LSTM forecasting. RF function to optimize important parameters [25]. The prediction process is carried out in two main steps: first, selecting features with RF, then second, predicting with LSTM. Algorithm 3 describes the forecasting process flow using LSTM in hybrid with HHT. Hybrid occurs early in the forecasting process. The HHT results are used for forecasting by training 10 times with epochs of 50 to 500, with multiples of 50. The study took the two best epochs at the time of training, both were compared for evaluation.

Algorithm 3. The forecasting process uses Hybrid LSTM with HHT

```
Input: Xb<----{BPS, GT}
Output: Evaluation Result O
```

Clean data are taken from **Algorithm 1**, to **Input**
Output: Predicted result *O* <---- *PrO = BPS, GT, NC;* NC = Noise cleaning.

```
Split PrO with <--- r90 and r10
if r90 then training
    else r10 then test
end if
```

//Building the LSTM model Architecture

Use r90
Train {e50, ..., e500}
If r90 then eTrainB //Choose the best two epochs out of 10 tested epochs
else r10 then eTest
endif
//Make predictions on Prediction with eTest using to best epochs

//Conduct evaluation by analyzing the two best epochs

```
If eTrainA then Ev.Reg
    else eTrainB then Ev.Conf
endif
```

Hybrid LSTM uses HHT and RF for forecasting to make an alternative to optimizing results. This process is carried out in three main steps. First, the data is cleaned. Second, the clean data is selected for its features so that it becomes small. Third, do forecasting using the data that is already good. Algorithm 4 shows the forecasting process using a hybridized LSTM with HHT and RF. A process on Algorithm 4, hybrid HHT, and RF algorithm. HHT is used to clean data from interference. Meanwhile, FR is used to select features. Feature selection produces near-optimal accuracy.

Algorithm 4. The forecasting process uses Hybrid LSTM with HHT and RF

```
Input: BPS, GT, NC, and FS <---- NC = Noise cleaning.
FS = Feature Selection
Output: Evaluation Result O
```

Clean data are taken from **Algorithm 1**, to **Input**
Output: Predicted result *O*

```
Xfs <--- BPS-F, GT-F and NS-F //New data from feature selection
XNew <--- Feature Selection Results from Algorithm 2

Split XNew with <--- r90 and r10
if r90 then training
    else r10 then test
end if
```

//Building the LSTM model Architecture

Use Xnew
Train {e50, ..., e500}
If r90 then eTrainB //Choose the best two epochs out of 10 tested epochs
else r10 then eTest
endif
//Make predictions on Prediction with eTest using to best epochs

//Conduct evaluation by analyzing the two best epochs

```
If eTrainA then Ev.Reg
    else eTrainB then Ev.Conf
endif
```

4. EXPERIMENT

4.1. Dataset and experimental setup

Experiments were conducted to evaluate the proposed algorithm, using data from GT cleaned with HHT as a benchmark. Monthly datasets from 2008 to 2021, totaling 143,312,804 parameters, were divided into 90% training and 10% testing sets. Each experiment was repeated 10 times over epochs ranging from 50 to 500, with epochs 300 and 350 found to be optimal for computational speed and minimal loss. The final evaluation demonstrated that epoch 300 outperformed epoch 350 in accuracy, as measured by mean absolute percentage error (MAPE) and root mean square error (RMSE) [25]. The dataset, available at https://github.com/harunmukhtar/Prediksi-Kunjungan-Wisatawan-Ke-Indonesia/tree/main/data. Includes three columns: BPS (data from the Indonesian Central Bureau of Statistics), HHT (GT data cleaned with HHT), and GT (raw data from GT). Figure 2 shows a comparison graph demonstrating the HHT's effectiveness in noise reduction.

The data from the three columns is then split into three files, namely 1-hht-fit.csv, 2-bps-fit.csv, and 3-gt-fit.csv. The data is stored on https://github.com/harunmukhtar/hht-rf-lstm/tree/main/data/data-mentah. And anyone who needs can reuse. These three data are then subjected to a feature selection process using RF. The data resulting from the feature selection is contained in https://github.com/harunmukhtar/hht-rf-lstm/tree/main/data/hasil-fs. The best data results from each file are put back together into columns. The combined data is used for forecasting tourist arrivals the following year. The combined data used for this forecasting can be stored on https://github.com/harunmukhtar/hht-rf-lstm/blob/main/data/hasil-fs/FS-gab-fit.csv. This data allows for reuse in further research.

4.2. Experimental result

All of our experiments were carried out on a Notebook, 80TV (LENOVO_MT_80TV_BU_ idea_FM_Lenovo ideapad 310-15IKB), CPU: Intel(R) Core (TM) i7-7500U CPU @ 2.70GHz, with RAM capacity: 12GiB. The operating system used is Debian GNU/Linux 11 (bullseye). Experiments were carried out by training ten times the epoch. The training time is 28,643 seconds or less than 8 hours for 2 different files.

GT data allegedly contains a lot of noise. HHT is used to clean the noise so the GT can be used. The initial step of cleaning this data is by converting data into signals. The decomposition results in four IMF functions. The series of all component functions and their residues. EMD adaptively decomposes the signal into IMF and residual components. Each IMF function contains a different characteristic time scale. the frequency of all IMF functions decreases alternately and does not overlap while the residuals are closed for monotonic functions. The horizontal axis represents the time span, and the vertical axis shows the function value of each IMF. Figure 3 is the result of the final process of this work.

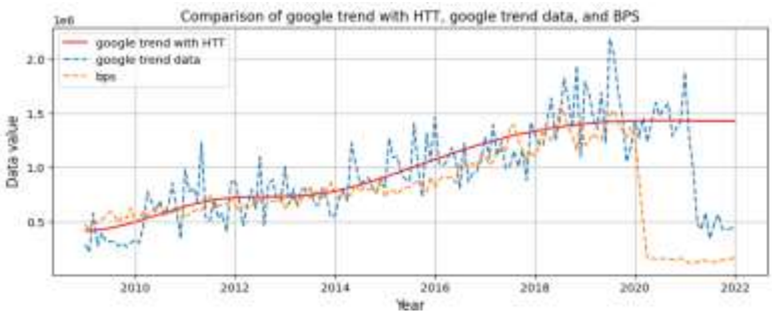

Figure 2. Raw comparison graph and data after cleaning

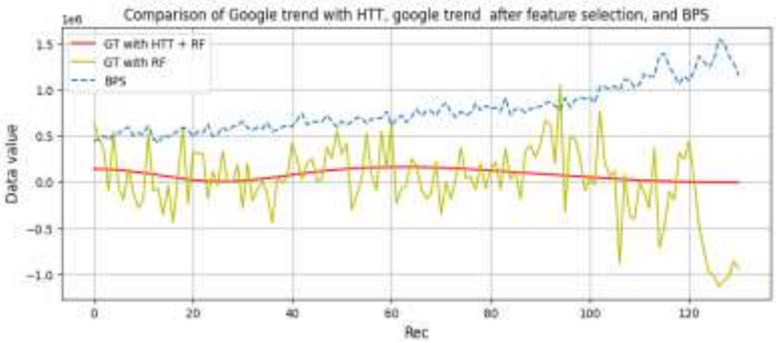

Figure 3. Graph Comparison of google trend with HTT, google trend data, and BPS

The following experiment is to do feature selection using the RF algorithm. Feature selection starts with stationary to adjust the time series seasonally. Empty data is also trimmed to produce new data. Feature selection is carried out in 5 steps to produce the best data features. The first step involves making the dataset stationary by removing seasonality and trends over a 12-month period. The second step selects features based on their correlation with the output variable (autocorrelation), with significant correlations shown in blue. A lag score of 1 indicates a 100% positive correlation.

The third step transforms the dataset into a supervised learning problem using lagged observations as input and the current observation as output, using 12-month lag values. The fourth step evaluates feature importance to determine the significance of input features over 12 lag observations. The fifth step automates the selection of variable lag features using recursive feature elimination (RFE), which builds a predictive model, assigns feature weights, and removes less significant features.

This experiment hybridizes LSTM with HHT for forecasting. Two critical stages were carried out, namely, building the architectural forecasting model with LSTM. The architecture built is as shown in Table 1. Table 1 shows two architectures, the first is using BPS and GT data. GT data is cleaned using HHT. Total parameters are 143,312,804, trainable parameters are 80,504, and non-trainable parameters are

143,232,300. Both architectures for feature-selected data. After selecting the data, the total parameters are 16,394,804, the Trainable parameters are still 80,504 and the non-trainable parameters are 16,314,300. Feature selection using RF managed to reduce parameters by 126,918,000.

Second, carry out the training process with twenty epochs in multiples of 25, such as epochs 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, and 500. Two experiments are observed in tabular form for HHT-LSTM and HHT-RF-LSTM. Based on training with a ratio of 90% training and 10% testing calculated from each epoch. The lowest time, the longest time, as well as the lowest loss, and the highest loss are evaluated at each epoch. Based on the training conducted, epoch 300 was identified as the optimal epoch. Table 2 illustrates that for RF-LSTM, the minimum computing time was 0 s 4 ms, the maximum was 3 s 4 ms, the lowest loss value was 8.02, and the highest was 92.97. On the other hand, HHT-RF-LSTM recorded a minimum time of 0 s 4 ms and a maximum of 2 s 4 ms, with the lowest loss value at 8.81 and the highest at 33.50.

Table 1. Comparison of parameter values before and after feature selection

Layer (type)	Before feature selection		After feature selection	
	Output shape	Parameters	Output shape	Parameters
Embedding	Output Shape	Parameters	None, 2, 100	16314300
LSTM	None, 2, 100	143232300	None, 2, 100	80400
Dense	None, 2, 100	80400	None, 2, 1	101
Flatten	None, 2, 1	101	None, 2	0
Dense_1	None, 2	0	None, 1	3
Total parameters		143,312,804		16,394,804
Trainable parameters		80,504		80,504
Non-trainable parameters		143,232,300		16,314,300

Table 2. Experimental results with 20 epochs for HHT – RF – LSTM

Algorithm	Epoch	Lowes time	Highest time	Lowest loss	Highest loss	Epoch	Lowes time	Highest time	Lowest loss	Highest loss
RF-LSTM	25	0 s 4 ms	3 s 4 ms	11.69	99.99	275	0 s 4 ms	2 s 4 ms	8.55	42.79
HHT-RF-LSTM		0 s 4 ms	2 s 4 ms	11.54	69.43		0 s 4 ms	2 s 6 ms	9.64	99.99
RF-LSTM	50	0 s 4 ms	2 s 4 ms	9.14	99.99	300	0 s 4 ms	3 s 4 ms	8.02	92.97
HHT-RF-LSTM		0 s 4 ms	2 s 4 ms	13.26	80.63		0 s 4 ms	2 s 4 ms	8.81	33.50
		0 s 4 ms	...	...	...		...	...	...	...
RF-LSTM	225	0 s 4 ms	2 s 4 ms	8.92	99.99	475	0 s 4 ms	2 s 4 ms	6.99	99.99
HHT-RF-LSTM		0 s 4 ms	2 s 5 ms	8.95	99.99		0 s 4 ms	2 s 6 ms	7.09	99.99
RF-LSTM	250	0 s 4 ms	2 s 4 ms	8.92	99.99	500	0 s 4 ms	3 s 6 ms	6.87	99.99
HHT-RF-LSTM		0 s 4 ms	2 s 4 ms	8.95	99.99		0 s 4 ms	2 s 4 ms	8.13	99.99

4.3. Evaluation

The actual value and the predicted value may be different. This difference is called prediction error. Analyzing the prediction error is very necessary to determine the accuracy of the prediction model used. A reliable model should generate predicted values that closely approximate actual values. To assess the performance of the prediction model, RMSE [32], MAPE, MSE, and MAE were selected [33], [34]. These metrics are employed to compare and determine the model's efficacy in achieving optimal results. Previous research [25]. The best MAPE and RMSE values were 157.00 and 217628.75 for LSTM, and 93.71 and 123882.20 for HHT-LSTM, respectively. Since the best MAPE value was different from expectations, a re-examination was conducted as shown in Table 3.

Table 3. Evaluation for regression

Algorithm	RMSE	MAE	MAPE	MSE
LSTM	1015057.72	655338.00	4.78	1.03
HHT-LSTM	495974.29	324665.65	2.43	2.45
RF-LSTM	553233.33	461693.15	0.37	0.06
HHT-RF-LSTM	**548645.33**	**462669.36**	**0.36**	**3.01**

Based on Table 4, it can be seen that the LSTM improved with HTT and has a smaller difference for all evaluation results. HHT-RF-LSTM is so good that it is recommended for use in DL-based forecasting. A confusion matrix approach can be used to evaluate forecasting results properly [35]. The standard evaluation tool used is the confusion matrix [36].

Table 4. Evaluation results with confusion matrix

Algorithm	Accuracy	Precision	Recall	F1-score
LSTM	50.00	100.00	50.00	66.67
HHT-LSTM	91.67	88.89	72.73	80.00
RF-LSTM	61.54	72.73	80.00	76.19
HHT-RF-LSTM	**91.31**	**91.67**	**100**	**95.65**

5. DISCUSSION

The experiment results show that the accuracy of forecasting with LSTM using GT data is quite good. However, a data cleaning process from noise is required. Data cleaning can be done with various algorithms such as HHT [25], [37], Kalman filter (KF) [38]. KF is not discussed in this research, this research focuses on HHT. Based on research on experiments that have been carried out, data from GT is very good as an explanatory variable. As is known, LSTM requires large data while statistical data cannot fulfill this [38]. HHT is very good for improving GT data but still has many features that cannot be trained. This research found that parameters that could not be trained reached 143,232,300. Many features that cannot be trained will affect whether the data is good or bad. Based on this consideration, feature selection is required. Feature selection is needed to reduce the number of parameters that cannot be trained. This research uses RF for feature selection. RF is very effectively used to improve parameters that cannot be trained. Based on the research conducted, it can be seen that the parameters that cannot be trained have been reduced to 16,314,300. As a result of reducing features that cannot be trained, it can improve accuracy very well.

6. CONCLUSION

This research integrates three algorithms to achieve near-optimal optimization. In addition to algorithm optimization, this research focuses on new techniques for utilizing GT data sourced from search engines. Incorporating this new data addresses the challenges posed by the limited volume of tourist visit data in tourism forecasting. This addition is expected to improve forecasting techniques, optimization, and accuracy. Based on this research, GT data is very suitable for estimation. Even though it causes quite a lot of noise, this can be overcome effectively by using HHT. The HHT algorithm can be relied on to denoise tourist arrival data. Feature selection is very effective in improving accuracy and training time per epoch. LSTM is a powerful deep-learning algorithm, showing excellent performance. The confusion matrix remains a classic and powerful tool for measuring accuracy. Future research will combine these two indicators and concurrently adjust parameters.

REFERENCES

[1] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, Nov. 2021, doi: 10.1007/s42979-021-00815-1.

[2] N. A. Zanury, M. A. Remli, H. K. Adli, and K. N. S. W. S. Wong, "Recent developments of deep learning in future smart cities: a review," in *Intelligent Systems Reference Library*, vol. 121, 2022, pp. 199–212. doi: 10.1007/978-3-030-97516-6_11.

[3] X. Li, "A random forest based learning framework for tourism demand forecasting with search queries," *Travel and Tourism Research Association: Advancing Tourism Research Globally*. p. 6, 2016.

[4] S. Kamolov, "Feature selection: state-of-the-art survey," *Annals of Mathematics and Computer Science*, vol. 4, pp. 48–54, 2021, [Online]. Available: https://annalsmcs.org/index.php/amcs/article/view/42

[5] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A review of feature selection methods for machine learning-based disease risk prediction," *Frontiers in Bioinformatics*, vol. 2, Jun. 2022, doi: 10.3389/fbinf.2022.927312.

[6] X. Li, Q. Wu, G. Peng, and B. Lv, "Tourism forecasting by search engine data with noise-processing," *African Journal of Business Management*, vol. 10, no. 6, pp. 114–130, 2016, doi: 10.5897/ajbm2015.7945.

[7] M. L. Shen, H. H. Liu, Y. H. Lien, C. F. Lee, and C. H. Yang, "Hybrid approach for forecasting tourist arrivals," *ACM International Conference Proceeding Series*, vol. Part F147956, pp. 392–396, 2019, doi: 10.1145/3316615.3316628.

[8] I. G. N. M. Jaya and N. Sunengsih, "Forecasting for the arrival ofinternational tourists after two years of the COVID-19 pandemic in Indonesia," *International Journal of Applied Research in Social Sciences*, vol. 4, no. 1, pp. 1–8, Feb. 2022, doi: 10.51594/ijarss.v4i1.297.

[9] H. Mukhtar, R. M. Taufiq, I. Herwinanda, D. Winarso, and R. Hayami, "Forecasting COVID-19 time series data using the long short-term memory (LSTM)," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 211–217, 2022, doi: 10.14569/IJACSA.2022.0131026.

[10] W. Lu, H. Rui, C. Liang, L. Jiang, S. Zhao, and K. Li, "A method based on GA-CNN-LSTM for daily tourist flow prediction at scenic spots," *Entropy*, vol. 22, no. 3, 2020, doi: 10.3390/e22030261.

[11] D. C. W. Wu, L. Ji, K. He, and K. F. G. Tso, “Forecasting tourist daily arrivals with a hybrid Sarima–LSTM approach,” *Journal of Hospitality and Tourism Research*, vol. 45, no. 1, pp. 52–67, 2021, doi: 10.1177/1096348020934046.

[12] H. Mukhtar, M. A. Remli, K. N. S. W. S. Wong, E. Fuad, J. Siregar, and Y. Rizki, “Forecasting tourist arrivals with partial time series data using Long-Short Term Memory,” *Engineering and Technology Quarterly Reviews*, vol. 6, no. 1, pp. 56–64, 2023, doi: 10.5281/zenodo.7970542.

[13] S. Polyzos, A. Samitas, and A. E. Spyridou, “Tourism demand and the COVID-19 pandemic: an LSTM approach,” *Tourism Recreation Research*, vol. 46, no. 2, pp. 175–187, Apr. 2021, doi: 10.1080/02508281.2020.1777053.

[14] T. Zheng, S. Liu, Z. Chen, Y. Qiao, and R. Law, “Forecasting daily room rates on the basis of an LSTM model in difficult times of hong kong: evidence from online distribution channels on the hotel industry,” *Sustainability (Switzerland)*, vol. 12, no. 18, 2020, doi: 10.3390/SU12187334.

[15] T. Lee, “EMD and LSTM hybrid deep learning model for predicting sunspot number time series with a cyclic Pattern,” *Solar Physics*, vol. 295, no. 6, 2020, doi: 10.1007/s11207-020-01653-9.

[16] A. Moghar and M. Hamiche, “Stock market prediction using LSTM recurrent neural network,” *Procedia Computer Science*, vol. 170, pp. 1168–1173, 2020, doi: 10.1016/j.procs.2020.03.049.

[17] A. Soy Temür, M. Akgün, and G. Temür, “Predicting housing sales in turkey using arima, LSTM and hybrid models,” *Journal of Business Economics and Management*, vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[18] K. Volchek, A. Liu, H. Song, and D. Buhalis, “Forecasting tourist arrivals at attractions: search engine empowered methodologies,” *Tourism Economics*, vol. 25, no. 3, pp. 425–447, 2019, doi: 10.1177/1354816618811558.

[19] W. Höpken, T. Eberle, M. Fuchs, and M. Lexhagen, “Search engine traffic as input for predicting tourist arrivals,” in *Information and Communication Technologies in Tourism 2018*, vol. 1, Cham: Springer International Publishing, 2018, pp. 381–393. doi: 10.1007/978-3-319-72923-7_29.

[20] C. Li, P. Ge, Z. Liu, and W. Zheng, “Forecasting tourist arrivals using denoising and potential factors,” *Annals of Tourism Research*, vol. 83, p. 102943, Jul. 2020, doi: 10.1016/j.annals.2020.102943.

[21] N. H. Z. Abidin *et al.*, “Improving intelligent personality prediction using Myers-Briggs type indicator and random forest classifier,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 192–199, 2020, doi: 10.14569/IJACSA.2020.0111125.

[22] A. Salamanis, G. Xanthopoulou, D. Kehagias, and D. Tzovaras, “LSTM-based deep learning models for long-term tourism demand forecasting,” *Electronics (Switzerland)*, vol. 11, no. 22, 2022, doi: 10.3390/electronics11223681.

[23] H. Laaroussi, F. Guerouate, and M. Sbihi, “Deep learning framework for forecasting tourism demand,” in *2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, Nov. 2020, pp. 1–4. doi: 10.1109/ICTMOD49425.2020.9380612.

[24] S. R. Qin and Y. M. Zhong, “A new envelope algorithm of Hilbert-Huang transform,” *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 1941–1952, 2006, doi: 10.1016/j.ymssp.2005.07.002.

[25] H. Mukhtar, M. A. Remli, K. N. S. W. S. Wong, and M. S. Mohamad, “Deep learning with processing algorithms for forecasting tourist arrivals,” *TEM Journal*, vol. 12, no. 3, pp. 1742–1753, 2023, doi: 10.18421/TEM123-57.

[26] L. Wang, S. Jiang, and S. Jiang, “A feature selection method via analysis of relevance, redundancy, and interaction,” *Expert Systems with Applications*, vol. 183, p. 115365, Nov. 2021, doi: 10.1016/j.eswa.2021.115365.

[27] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, “Feature selection using random forest classifier for predicting prostate cancer,” *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, p. 052031, Jun. 2019, doi: 10.1088/1757-899X/546/5/052031.

[28] I.-A. Kang, S. N. Njimbouom, and J.-D. Kim, “Optimal feature selection-based dental caries prediction model using machine learning for decision support system,” *Bioengineering*, vol. 10, no. 2, p. 245, Feb. 2023, doi: 10.3390/bioengineering10020245.

[29] A. A. Rizal, S. Soraya, and M. Tajuddin, “Sequence to sequence analysis with long short term memory for tourist arrivals prediction,” *Journal of Physics: Conference Series*, vol. 1211, no. 1, p. 012024, Apr. 2019, doi: 10.1088/1742-6596/1211/1/012024.

[30] L. Peng, L. Wang, X.-Y. Ai, and Y.-R. Zeng, “Forecasting tourist arrivals via random forest and long short-term memory,” *Cognitive Computation*, vol. 13, no. 1, pp. 125–138, Jan. 2021, doi: 10.1007/s12559-020-09747-z.

[31] H. Nan, “Apply RF-LSTM to predicting future share price,” *SHS Web of Conferences*, vol. 170, p. 02012, Jun. 2023, doi: 10.1051/shsconf/202317002012.

[32] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, “Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches,” *Energies*, vol. 11, no. 7, p. 1636, Jun. 2018, doi: 10.3390/en11071636.

[33] G. Xie, Y. Qian, and S. Wang, “Forecasting Chinese cruise tourism demand with big data: an optimized machine learning approach,” *Tourism Management*, vol. 82, p. 104208, Feb. 2021, doi: 10.1016/j.tourman.2020.104208.

[34] Y. Han, C. Wang, Y. Ren, S. Wang, H. Zheng, and G. Chen, “Short-term prediction of bus passenger flow based on a hybrid optimized LSTM network,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 9, p. 366, Aug. 2019, doi: 10.3390/ijgi8090366.

[35] D. Bowes, T. Hall, and D. Gray, “Comparing the performance of fault prediction models which report multiple performance measures,” in *Proceedings of the 8th International Conference on Predictive Models in Software Engineering*, Sep. 2012, pp. 109–118. doi: 10.1145/2365324.2365338.

[36] M. Farsi, “Application of ensemble RNN deep neural network to the fall detection through IoT environment,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 199–211, Feb. 2021, doi: 10.1016/j.aej.2020.06.056.

[37] H. Sun, Q. Si, N. Chen, and S. Yuan, “HHT-based feature extraction of pump operation instability under cavitation conditions through motor current signal analysis,” *Mechanical Systems and Signal Processing*, vol. 139, 2020, doi: 10.1016/j.ymssp.2019.106613.

[38] C. H. Nkwayep, S. Bowong, J. J. Tewa, and J. Kurths, “Short-term forecasts of the COVID-19 pandemic: a study case of Cameroon,” *Chaos, Solitons & Fractals*, vol. 140, p. 110106, Nov. 2020, doi: 10.1016/j.chaos.2020.110106.

BIOGRAPHIES OF AUTHORS

Harun Mukhtar is an associate professor at the Universitas Muhammadiyah Riau specializing in the informatics engineering program. S.Kom, Bachelor's degree from STMIK-AMIK Riau, and M.Kom, master's degree from UPI YPTK Padang. He is also pursuing a doctorate at Universiti Malaysia Kelantan in the field of artificial intelligence. His research interests include the development of new techniques for effective data processing and analysis to address real-world challenges. He is also a member of several professional organizations, contributing to ongoing professional development and collaboration with other experts in his field. For academic or research questions, he can be contacted at email: harun.mukhtar@umri.ac.id.

Muhammad Akmal Remli joins Institute for Artificial Intelligence and Big Data (AIBIG), Universiti Malaysia Kelantan (UMK) as a fellow researcher in early 2020 and now he is AIBIG's director. He is also a senior lecturer at Faculty of Data Science and Computing, U K. He received a master and a Ph.D. degree in computer science from Universiti Teknologi Malaysia in 2014 and 2018 before joining Universiti Malaysia Pahang from 2018 until 2020. In 2016, he worked at The Bioinformatics, Intelligent Systems and Educational Technology (BISITE) Research Group at University of Salamanca, Spain as research attachment and was working in cancer bioinformatics. His main research interests are artificial intelligence, data science, business intelligence, and computational systems biology. He can be contacted at email: akmal@umk.edu.my.

Dr. Mohd Saberi Mohamad is a professor of Artificial Intelligence and Health Data Science. He is now the director of the Health Data Science Lab in the Department of Genetics and Genomics (CMSH-UAEU). His research interests include artificial intelligence, bioinformatics, data science, and computational biology. Before joining CMHS-UAEU, he served at several universities in Malaysia as the director of the Institute for Artificial Intelligence and Big Data, head of the Artificial Intelligence and Bioinformatics Research Group, founder of the Department of Data Science, Manager of Information Technology, and deputy director (academic) for Centre of Computing and Informatics. He can be contacted at email: saberi@uaeu.ac.ae.

Khairul Nizar Syazwan Wan Salihin Wong received his Bachelor of Engineering (Mechatronics) from UIAM and master's degree in Electrical - Mechatronics and Automatic Control from Universiti Teknologi Malaysia (UTM) in 2012 and 2015. Previously, he was an instrumentation and control engineer at an international oil and gas industry for 5 years. Currently, he is a lecturer at Faculty of Data Science and Computing (FSDK), UMK. His current research interest such as internet-of-things (IoT) system, computer vision, artificial intelligence, and instrumentation. He can be contacted at email: nizar.w@umk.edu.my.

Farhan Ridhollah is a student in the informatics engineering program at Universitas Muhammadiyah Riau, supervised by Assoc. Prof. Harun Mukhtar. A graduate of SMK Bina Profesi Pekanbaru, he has strong technical skills, with experience in independent studies on AI and data analysis. He is a Certified Associate Data Scientist (CADS) by BNSP and actively assists in research projects related to data science and AI. He can be contacted at email: 200401121@student.umri.ac.id.

Dr. Deprizon born in Gunung Sahilan, Riau, he holds a bachelor's in Islamic Education, a Master's in Islamic Education Management, and a doctorate in Islamic Education from UIN Suska Riau. He is a lecturer at Universitas Muhammadiyah Riau (UMRI). His works include the "LSIK-UMRI Worship Practice Guide, Teacher's Handbook, Hifzil-Quran Method 'Ibroh Robbaniyyah, and Integrated Quality Management in Education and School Quality Accreditation". He can be contacted at email: deprizon@umri.ac.id.

Soni joins Faculty of Computer Science, Universitas Muhammadiyah Riau. He is also a senior lecturer and now he is deputy dean. He received a bachelor degree in Informatics Engineering Department from STMIK AMIK Riau, Indonesia. And a master degree in Computer Science from Islamic University of Indonesia. His main research interests are data science, artificial intelligence, machine learning, and digital forensic. He can be contacted at email: soni@umri.ac.id.

Muhammad Lisman is a lecturer at the Universitas Muhammadiyah Riau who is currently pursuing a Ph.D. at a University in Malaysia. His undergraduate education is from the University of Muhammadiyah Yogyakarta, and his master's degree is from the Islamic University of Indonesia. His expertise or interest is in digital business. During his career as a lecturer, he has published nationally accredited and Scopus (Springer) indexed journals and has attended several international conferences. He also collaborate on national and international research with the same field of expertise. He can be contacted at email: muhammadlisman@umri.ac.id.

Hasanatul Fu'adah Amran is a lecturer in the informatics engineering program at Universitas Muhammadiyah Riau, having completed both her bachelor's and master's degrees in Mathematics Education at Universitas Ahmad Dahlan Yogyakarta, the latter in 2019. She teaches applied mathematics, discrete mathematics, and algorithms and programming, and was also a guest lecturer at Universitas Pahlawan (2019-2020). Her research interests include applied mathematics, algorithms, and AI. She can be contacted at email: hasanatul@umri.ac.id.

Sunanto is a senior lecture at Universitas Muhammadiyah Riau focus field is embedded system using artificial intelligent and machine learning. Research was public in smart parking using biometric and image recognitions. Education background diploma at polytechnic TEDC Bandung in hardware computer, bachelor degree at STMIK AMIK Riau and master degree at University Putra Indonesia Padang. Now we studied doctor of philosophy computer science at University Sultan Zainal Abidin (UniSZa) in Malaysia Terengganu. Research implementation in precision agriculture using embedded system. He can be contacted at email: sunanto@umri.ac.id.

Edi Ismanto completed education bachelor's degree in the Informatics Engineering Department, State Islamic University of Sultan Syarif Kasim Riau. And master's degree in master of Computer Science at Putra Indonesia University Padang. Now working as a lecturer in the Department of Informatics, University Muhammadiyah of Riau. With research interests in the field of machine learning algorithms and AI. He can be contacted at email: edi.ismanto@umri.ac.id.