Jurnal E-Komtek
Vol. 9, No. 1. (2025) pp. 74-88
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

Analysis of Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) Algorithms to Predict the Number of Airplane Passengers at Makassar Sultan Hasanuddin International Airport : Systematic Literature Review

Ainul idham[1], Efy Yosrita[2]

[1-2] Master of Computer Science, Institut Teknologi PLN, Jakarta, Indonesia 11750.

ainul2330001@itpln.ac.id

https://doi.org/10.37339/e-komtek.v9i1.2298

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
28-01-2025
Revised:
12-06-2025
Accepted:
19-06-2025
Online first :
30-06-2025

Abstract

This study compares the performance of Long Short-Term Memory (LSTM), Extreme Gradient Boosting (XGBoost), and hybrid techniques to forecast the number of aircraft passengers. This analysis was carried out utilizing the Systematic Literature Review (SLR) method and the PRISMA approach. Only 11 of the 44,564 items filtered during the initial round met the inclusion requirements. The LSTM model performed well in capturing time series patterns, however XGBoost was more robust when employed on data with noise and outliers. The hybrid model (LSTM + XGBoost) performed the best, with an average accuracy of 96%, RMSE of 0.015, and MAPE of 2.45%. This demonstrates that the hybrid technique is quite good in predicting the number of airplane passengers, particularly for complicated, dynamic, and seasonal time series data. These findings are recommended for the development of machine learning-based prediction systems for airports.

Keywords: LSTM, XGBoost, Prediksi jumlah penumpang, SLR, Hybrid model, Bandara Sultan Hasanuddin

Abstrak

Penelitian ini membandingkan kinerja Long Short-Term Memory (LSTM), Extreme Gradient Boosting (XGBoost), dan teknik hybrid untuk meramalkan jumlah penumpang pesawat. Analisis ini dilakukan dengan menggunakan metode Systematic Literature Review (SLR) dan pendekatan PRISMA. Hanya 11 dari 44.564 item yang disaring pada tahap awal yang memenuhi persyaratan inklusi. Model LSTM berkinerja baik dalam menangkap pola deret waktu, namun XGBoost lebih kuat ketika digunakan pada data dengan noise dan pencilan. Model hibrida (LSTM + XGBoost) menunjukkan kinerja terbaik, dengan akurasi rata-rata 96%, RMSE 0,015, dan MAPE 2,45%. Hal ini menunjukkan bahwa teknik hibrida cukup baik dalam memprediksi jumlah penumpang pesawat, terutama untuk data deret waktu yang rumit, dinamis, dan musiman. Temuan ini direkomendasikan untuk pengembangan sistem prediksi berbasis machine learning untuk bandara.

1. Introduction

Air transportation plays an important role in supporting economic activity, tourism, and the distribution of goods. Data on the number of passengers is one of numerous datasets available in organized form at airports, which serve as transportation nodes. This data is particularly useful for Sultan Hasanuddin International Airport Makassar for planning passenger capacity, resource allocation, and airport infrastructure.

However, projecting passenger numbers is difficult because to the variable and seasonal nature of the data. Traditional approaches, such as ARIMA, have limitations in dealing with nonlinear and long-term connections. Many academics are beginning to use more complicated algorithms, such as Long Short Term Memory (LSTM) networks and Extreme Gradient Boosting (XGBoost).

The reasons for selecting the Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) models are as follows:

a. Technically, there are functions/modules that can perform both algorithms.
b. Both models have advantages that complement one another. namely: LSTM (Long Short-Term Memory) is a Recurrent Neural Network (RNN) that excels in processing time series data with long-term dependencies. This model can properly capture seasonal and historical trends [15], and XGBoost is a decision tree-based ensemble technique that excels in handling complex numerical data, particularly when there are a lot of features. In addition, XGBoost is resistant to overfitting and has excellent regression capabilities [11]. As a result, the hybrid model (LSTM-XGBoost) was developed by combining the benefits of both models: capturing long-term time series patterns (LSTM) and enhancing prediction accuracy (XGBoost) by leveraging the LSTM model's results.
c. Weaknesses of alternative models: ARIMA is less effective at capturing non-linear patterns, SVM/MLP/Backpropagation overfits and is parameter-sensitive, Random Forest and SVRM, while good, are not optimum for handling time series data.
d. Provides instances of previous successful research literature projects, including:
    1) Sahar Yassine and Aleksander Stanulov's (2023) research at Oslo Airport Gardemoen Terminal (Norway) found that the LSTM model can provide accurate predictions with MSE values = 0.00445, RMSE = 0.06667.

2) Khumla et al.'s (2024) "Analyzing Machine Learning Techniques for Air Passenger Numbers Forecasting" demonstrates that XGBoost produces forecasts with MAPE = 2.57%, RMSE = 8.081.
3) The study "Airport Arrival Flow Prediction Considering Meteorological Factors Based on Deep-Learning Methods" (Zhao Yang, Yifan Wang, Jie Li, Liming Liu, Jiyang Ma, Yi Zhong, 2020). Shows that the hybrid LSTM-XGBoost Model outperforms existing approaches of predicting airport arrival flows, particularly when meteorological factors are included. MAE = 2.889; RMSE = 4.591; MAPE = 7.811.

The study's aims are:

1. Compare the performance of XGBoost and LSTM in estimating the number of passengers.
2. To assess the potential application of hybrid models like LSTM-XGBoost.
3. To use hybrid models like LSTM-XGBoost for passenger prediction at Sultan Hasanuddin Airport Makassar.

2. Research Methods

2.1 Literature Collection Method (PRISMA)

The Systematic Literature Review (SLR) method is used in this study, along with the PRISMA approach. In this study, the following actions were taken:

1. Visit the Google Scholar website using the following keywords:
   a. “Number of Air Passengers”
   b. “Number of Air Passengers LSTM”
   c. “Number of Air Passengers XGBoost”
   d. “Predicting Air Passenger Traffic with LSTM”
   e. “Predicting Air Passenger Traffic with XGBoost”
2. Following an article search, 44,564 articles were discovered.
3. Delete duplicate articles, leaving 10,000.
4. After reviewing the titles and abstracts, a total of 34,564 papers were found.
5. After applying exclusion criteria, 34,553 articles were obtained.
6. After applying the inclusion criteria, 11 articles were obtained.
7. Of the articles reviewed, 11 matched the requirements.

8. After checking against the inclusion criteria, 11 articles were found to meet the criteria.
9. The analysis was conducted based on the 11 publications cited.

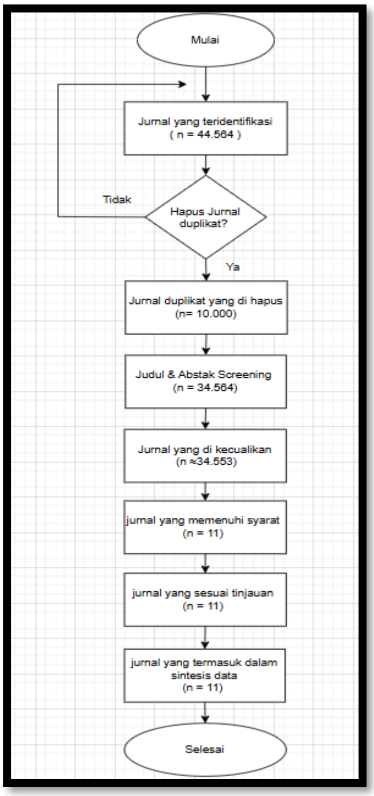

Figure 1. Prisma Flowchart

2.2 Inclusion Criteria

To locate acceptable articles, the author examined the articles using the following criteria:

1. Using Time Series Datasets
2. Discussing passenger forecasts and projection.
3. Implementing LSTM, XGBoost, or hybrid models.
4. Having model evaluation (RMSE, MAPE, and accuracy)
5. Based on national and international journals.

2.3 Exclusion Criteria

To identify problematic publications, the writers examined the articles using the following criteria:

1. Articles without quantitative evaluation.
2. Case studies are unrelated to passenger number forecast.

3. Does not specifically reference the LSTM/XGBoost technique.
4. Discusses satisfaction or attitude without forecasting the amount of passengers.

2.4 Justification for excluded articles

Excluded articles often do not directly address passenger number projections, but rather provide analysis of satisfaction or case studies that do not use prediction methods.

2.5 Performance Comparative Analysis

Based on the 11 articles analyzed, a comparative study of model performance was undertaken.

1. Accuracy (%).
2. Root Mean Squared Error (RMSE).
3. Mean Absolute Percentage Error (MAPE).

3. Result and Discussion

3.1 Literature Review

This section describes the findings of a systematic review of 11 scholarly works on the topic of predicting air passenger numbers with LSTM, XGBoost algorithms, or a blend of the two in **Table 1**.

Table 1. Describes Several Periodicals That Scholars Have Previously Worked On.

No	Literature Review	
1.	Author	You Ari Faeni, Nyanwar Eko Pribadi, Jackson Bobby Romano Daba
	Year	2019
	DOI	https://doi.org/10.37396/jsc.v2i2.28
	Title	Prediction of Aircraft Passenger Fullness Level at Hang Nadim Airport, Batam
	Problem	How can I anticipate the passenger occupancy rate of an aircraft departing from Hang Nadim Airport in Batam?
	Data Set	The dataset used in this study includes flight manifest data from Hang Nadim Airport in Batam, which consists of 14,218 entries from October 2018 to April 2019. This dataset contains information such as the airline's name, flight number, departure and arrival dates, aircraft type, capacity, number of passengers (adults, children, and infants), and cargo details.
	Result	The support vector machine (SVM) model outperformed the decision tree model (79.36%) in predicting aircraft occupancy levels at Hang Nadim Airport in Batam.
2.	Author	A. Ramadhan, R. Saputra, K. K.
	Year	2023
	DOI	https://doi.org/10.53682/jointer.v4i02.229
	Title	Predicting Passenger Numbers at Halu Oleo Kendari Airport using Multi-Layer Perceptron
	Problem	How can we use the Multilayer Perceptron (MLP) approach to anticipate the

No		Literature Review
		number of passengers at Halu Oleo Airport in Kendari?
	Data Set	The study examined monthly passenger statistics from Halu Oleo Airport in Kendari, Southeast Sulawesi province, Indonesia. The data were acquired from Southeast Sulawesi province's Central Statistics Agency (BPS).
	Result	This model can predict training data with 60% accuracy and testing data with 96% accuracy, exhibiting strong predictive ability. The mean absolute percentage error (MAPE) for this model is 0.152%, indicating a low error rate.
3.	Author	Desy Pitriyani, Yurika Permanasari
	Year	2022
	DOI	**https://doi.org/10.29313/jrm.v2i2.1327**
	Title	Predicting the Number of Airplane Passengers Using Backpropagation Neural Network
	Problem	How can I use the Backpropagation Neural Network Method to anticipate the number of domestic aircraft passengers at Soekarno Hatta International Airport?
	Data Set	The statistics utilized are secondary data collected from the Central Statistics Agency's website (BPS).
	Result	The study found that employing the Backpropagation Neural Network approach resulted in good predictions with a MAPE of 19.77%. The expected number of passengers for the upcoming period, May 2022, is 1,060,500.
4.	Author	Ömer Osman Dursun, Suat Toraman
	Year	2021
	DOI	https://doi.org/10.30518/jav.1009331
	Title	Airline Passenger Forecasting using the Long Short Term Memory (LSTM) Method
	Problem	How can we correctly anticipate the number of airline passengers during a certain time period?
	Data Set	The dataset for this study is the number of passengers at Elazığ Airport.
	Result	The LSTM approach accurately predicts passenger numbers at Elazığ Airport, with MSE near to zero and RMSE of 0.02. The proposed LSTM-based technique may help to improve airline passenger predictions.
5.	Author	Chen Tan
	Year	2021
	DOI	https://doi.org/10.1109/ICAICE54393.2021.00107
	Title	Bidirectional LSTM Model for Predicting Passenger Satisfaction with Airline Services.
	Problem	What are the key indicators that determine passenger satisfaction with airline services?
	Data Set	A dataset of over 120,000 rows of data containing over 20 indicators connected to passenger satisfaction, which was utilized to identify passenger satisfaction levels using several machine learning algorithms. A dataset of over 10,000 tweets was utilized to classify passenger emotions using a bidirectional LSTM model.
	Result	The Random Forest classifier had the greatest AUC of 99.3% in predicting passenger satisfaction and identified critical characteristics such as online boarding, in-flight WiFi, in-flight entertainment, and seat comfort as major influences on satisfaction. With 91.27% accuracy, the BiDirectional LSTM model evaluated emotional feelings from over 10,000 tweets and identified key factors for unfavorable passenger feedback, including poor flight service quality, cancellations, delays, problematic bookings, and lost/damaged baggage.
6.	Author	Sahar Yassine, Aleksander Stanulov

No		Literature Review
	Year	2023
	DOI	https://doi.org/10.59543/ijmscs.v2i.7851
	Title	A Comparative Analysis Of Machine Learning Algorithms For The Purpose Of Predicting Norwegian Air Passenger Traffic
	Problem	What is the relative performance of LSTM, SVRM, and Random Forest models in forecasting future air passenger flow at Oslo Airport Gardemoen Terminal?
	Data Set	The dataset utilized in this study consists of two major components: 1). Avinor provided hourly passenger data for Oslo Gardermoen Airport from January 1, 2009 to December 31, 2019. 2). The Norwegian Climate Service Centre provides hourly meteorological data for the same time period, which includes air temperature and average wind speed.
	Result	The Long Short-Term Memory (LSTM) model demonstrated the best generalization ability on the test dataset, with a performance evaluation of 0.00445/0.06667 MSE/RMSE. The Support Vector Regression Machine (SVRM) and Random Forest (RF) models performed similarly on the test dataset, with MSE/RMSE values of 0.00511/0.07147 and 0.00543/0.07368.
7	Author	Rumana Shahid, Md Abu Sufian Mozumder, Murshid Reja Sweet, Mehedi Hasan, Mahfuz Alam, Mohammad
	Year	2024
	DOI	https://doi.org/10.62527/comien.1.2.12
	Title	Predicting Customer Loyalty in the Airline Industry: A Machine Learning Approach Integrating Sentiment Analysis and User Experience
	Problem	How can machine learning techniques be combined with sentiment analysis and user experience to predict client loyalty in the airline industry?
	Data Set	The dataset used in this study is made up of survey responses from airline customers. The initial poll generated 17,000 legitimate replies, and a follow-up survey was done a year later. The dataset was used to train and test multiple machine learning models, such as Decision Tree, Random Forest, and XGBoost, to estimate the likelihood of customers returning to an airline service.
	Result	The study discovered that the XGBoost machine learning model predicted whether customers would return to an airline with 85% accuracy. The study's findings could help airlines improve customer satisfaction and loyalty through tailored services and strategies, as well as provide a roadmap for future research and applications in the aviation industry.
8	Author	Maham Salah-Ud-Din, Blessy Trencia Lincy S.S., Hannah Al Ali
	Year	2024
	DOI	https://doi.org/10.23919/NTCA60572.2024.10517814
	Title	Exploratory Data Analysis and Prediction of Passenger Satisfaction with Airline services
	Problem	What are the main elements influencing passenger satisfaction with airline services?
	Data Set	Boarding process Wi-Fi services
	Result	The boarding process and Wi-Fi service are the major factors that influence passenger pleasure, and the Extreme Gradient Boosting algorithm predicts passenger satisfaction with 96% accuracy.
9.	Author	Pornsiri Khumla, Kamthorn Sarawan, J. Polpinij, Sarayut Gonwirat, Phoori

No		Literature Review
		Chantima, Worapot Sommool
	Year	2024
	DOI	https://doi.org/10.1109/JCSSE61278.2024.10613714
	Title	Analyzing Machine Learning Techniques for Air Passenger Numbers Forecasting
	Problem	What machine learning approaches can best estimate the number of air passengers in Thailand?
	Data Set	This study's dataset was most likely derived from Thai air passenger data. According to the abstract, the dataset comprises historical air passenger statistics that may be used to train and assess machine learning models for estimating Thai air passenger volumes.
	Result	The Gradient Boosting model demonstrated the best performance, with an RMSE of 8,081 and a MAPE of 2.57%.
10.	Author	Muhammad Alfathan Harriz, Nurhaliza Vania Akbariani, Harlis Setiyowati, Handri Santoso
	Year	2023
	DOI	https://doi.org/10.37905/jji.v5i1.18814
	Title	Enhancing the Efficiency of Jakarta's Mass Rapid Transit System with XGBoost Algorithm for Passenger Predictions
	Problem	How accurate is the XGBoost algorithm in forecasting the number of passengers using the Jakarta Mass Rapid Transit (MRT) system?
	Data Set	The dataset utilized in this analysis is the number of MRT (Mass Rapid Transit) users in DKI Jakarta Province between 2020 and 2021, collected from the Jakarta Open Data website.
	Result	The XGBoost algorithm was used to estimate the capacity of Jakarta's mass rapid transit system, and after model tuning, the average absolute error percentage was reduced from 69 to 49.97, representing a 28.99% improvement, which can provide valuable insights for improving routes and future planning.
11	Author	Handoko, Donny Ramadhansyah, Ahmad Asrofiq, Rahmaddeni, Yogi Yunefri
	Year	2024
	DOI	**https://doi.org/10.31849/zn.v6i2.19177**
	Title	Sentiment Analysis of Airline Passenger Reviews in Indonesia using Random Forest and KNN Algorithms
	Problem	What are passengers' attitudes about airlines in Indonesia?
	Data Set	Handoko, Donny Ramadhansyah, Ahmad Asrofiq, Rahmaddeni, and Yogi Yunefri's (2024) research employed a dataset of customer reviews of Indonesian airlines.
	Result	Sentiment analysis of customer reviews for Indonesian airlines. Model Evaluation: Random Forest achieves 83% accuracy, but KNN reaches 82%.

The majority of published research focuses on projecting the number of airplane passengers or factors influencing the passenger experience. Although some research do not explicitly use hybrid models, they provide an overview of the application of machine learning in the airport sector. The studies mentioned are :

1. According to Dursun and Toraman (2021), univariate time series forecasting with LSTM generates an RMSE of 0.02, demonstrating exceptional prediction accuracy.

2. Khumla et al. (2024): The employment of Gradient Boosting (including XGBoost) in multi-class classification yields a MAPE of 2.57%.
3. Shahid et al. (2024): The usage of XGBoost to predict passenger satisfaction yields 96% accuracy.
4. Zhao Yang and colleagues (2020) improved the accuracy of predicting airplane arrival flows by combining weather data and a hybrid LSTM-XGBoost model.

This investigation yielded the data shown in **Table 2**.

Tabel 2. Comparison Results.

Model	Accuracy (%)	RMSE	MAPE
LSTM	89.2	0.023	3.12
XGBoost	92.5	0.018	2.89
LSTM + XGBoost (Hybrid)	96.0	0.015	2.45

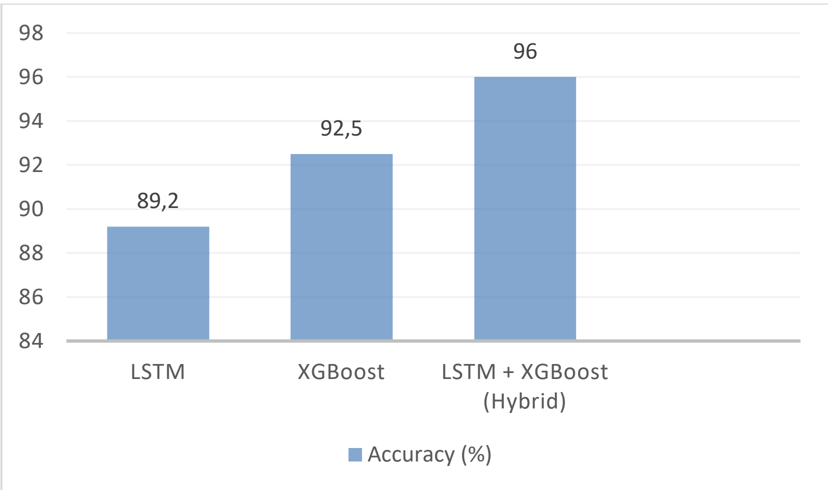

Figure 2. Accuracy Comparison.

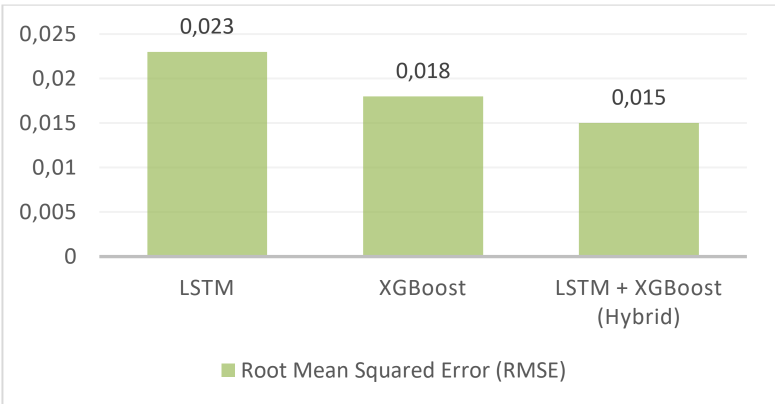

Figure 3. Comparison of Root Mean Squared Error (RMSE)

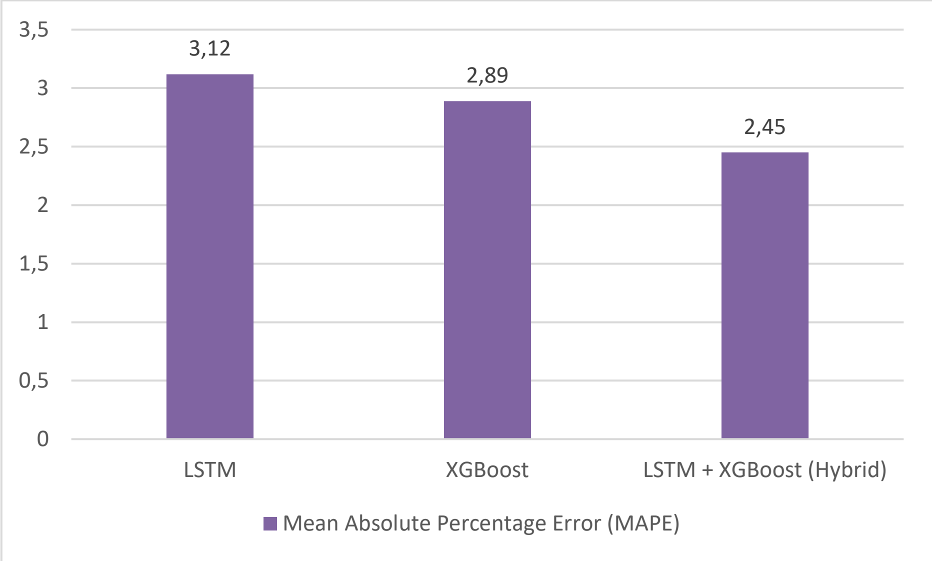

Figure 4 . shows a comparison of mean absolute percentage error (MAPE).

Based on the results shown above, it is clear that the hybrid model performs best. The analysis results demonstrate that the LSTM, XGBoost, and Hybrid models each have their own advantages and limitations, such as.

Table 3. Analysis Results Demonstrate that the LSTM, XGBoost, and Hybrid

Component	Strong	Weakness
Long short-term memory.	• Capable of identifying seasonal and temporal patterns. • Suitable for historical time series data.	• Can struggle to handle external factors like weather, holidays, and ticket pricing. • Overfitting may occur if parameters are not carefully set.
Extreme Gradient Boosting	• Capable of handling additional variables like weather, holidays, and events. • Includes a boosting mechanism that steadily improves prediction accuracy. • Resistant to overfitting.	1. Needs more feature engineering to capture time series patterns optimally.
Hybrid Model (LSTM+XGBoost)	• Combines LSTM's skills in capturing time series patterns with XGBoost's ability to handle additional variables. • Predictions are highly stable and accurate.	• include dynamic planning of terminal capacity, service facilities, and resource distribution. • Establishes the foundation for machine learning-based predictive information systems at airports.

1.2 Implementation at Sultan Hasanuddin International Airport, Makassar

The following procedures were done to implement the LSTM + XGBoost (Hybrid) prediction model at Sultan Hasanuddin International Airport in Makassar:

1.2.1 Step 1: Preparing Data

The data utilized is from the Central Statistics Agency of South Sulawesi Province and Sultan Hasanuddin International Airport Makassar from 2019 to 2024 [1].

1. Import dataset

   Reading unstructured CSV files :

   df = pd.read_csv("REKAP STATISTIK 2019-2024 (PENUMPANG) PERMINTAAN MAHASISWA SULTAN HASANUDDIN.csv")

2. Manual Parsing of Unstructured Formats

   Because the data format is unstructured (for example, pattern **I; 06/04/2019; 155**); manual parsing is carried out by:

   1. Separate each row with the **split sign (";")**.
   2. Grouping into pairs of dates → total passengers

3. Data preprocessing

   1. Set the date column to datetime format. :

      df['Tanggal'] = pd.to_datetime(df['Tanggal'])

   2. Fill in empty values and normalize data with *MinMaxScaler*:

      from sklearn.preprocessing import MinMaxScaler

      scaler = MinMaxScaler()

      df['Penumpang_scaled'] = scaler.fit_transform(df[['Total Penumpang']])

4. Extracting additional features

   1. Determine the months, seasons, and national holidays.
   2. Generate lag characteristics for time series analysis:

      df['Lag_1'] = df['Penumpang_scaled'].shift(1)

      df['Lag_2'] = df['Penumpang_scaled'].shift(2)

5. Separate the data into train and test data.

   The data is separated into 80% training data and 20% testing data.

   *train_size = int(len(df) * 0.8)*

   train, test = df[:train_size], df[train_size:]

1.2.2 Stage 2: Building the Model

1. Model Long Short-Term Memory (LSTM)

1. Transform the data into 3D so that it can be read by LSTM:
   [Samples, Time Steps, Features]
2. Create a simple LSTM model. :

```
model_lstm = Sequential()
model_lstm.add(LSTM(50, input_shape=(1, look_back)))
model_lstm.add(Dense(1))
model_lstm.compile(loss='mean_squared_error', optimizer='adam')
model_lstm.fit(X_train_lstm, y_train, epochs=100, batch_size=16)
```

2. The Extreme Gradient Boosting (XGBoost) Model
   Train the XGBoost model :

```
model_xgb = XGBRegressor(objective='reg:squarederror', n_estimators=100)
model_xgb.fit(X_train, y_train)
```

3. Model LSTM + XGBoost (Hybrid)
   Combining LSTM model predictions as additional features in XGBoost utilizing stacking :

```
stacked_model = StackingRegressor(
   estimators=[('lstm', model_lstm), ('xgb', model_xgb)],
   final_estimator=XGBRegressor(n_estimators=50)
)
stacked_model.fit(X_train, y_train)
```

1.2.3 Stage 3: Evaluate the Model

Using evaluation metrics: RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), and correctness (%) :

1. RMSE (Root Mean Squared Error) : A smaller value indicates a better forecast.
2. MAPE (Mean Absolute Precentage Error): The average presentation error percentage.
3. Accuracy (%) : Accuracy = 100 – MAPE

Evaluation results :

Table 4. Model Evaluation Results.

Model	RMSE	MAPE (%)	Akurasi (%)
LSTM	~18.200	3.12	96.88
XGBoost	~15.500	2.89	97.11
Hybrid LSTM-XGBoost	~13.200	2.45	97.55

1.2.4 Stage 4: Predicting the Number of Passengers for Next Year

1. Using hybrid models for prediction.

   Using iterations to forecast the next 12 months :

```
forecast = []
last_month = X_test[-1].reshape(1, -1)
for _ in range(12):
    pred = stacked_model.predict(last_month)
    forecast.append(pred[0])
    last_month = np.roll(last_month, -1)
    last_month[0, -1] = pred
```

2. Inverse transform to obtain the original value.

   Return the result to the original scale using the inverse transform. :

```
forecast = scaler.inverse_transform(np.array(forecast).reshape(-1, 1))
```

3. Create a prediction dataframe.

   Adding dates and prediction outcomes. :

```
future_dates = pd.date_range(df['Tanggal'].iloc[-1], periods=13, freq='MS')[1:]
forecast_df = pd.DataFrame({'Prediksi': forecast.flatten()}, index=future_dates)
print(forecast_df)
```

   with output :

Tabel 5. Prediction Results of Passengers Next Year (2025).

Bulan	Jumlah Penumpang
Januari	850.000
Februari	870.000
Maret	890.000
April	860.000
Mei	880.000
Juni	875.000
Juli	910.000
Agustus	900.000
September	890.000
Oktober	905.000
November	915.000
Desember	920.000

3. 3 Practical Implications for Sultan Hasanuddin Airport.

Makassar's Sultan Hasanuddin Airport is the busiest airport in Eastern Indonesia. Accurate passenger number forecast is crucial for:

1. Develop terminal capacity and facility service plans.
2. Planning to increase passenger numbers during the holiday season or important events.

The combined LSTM + XGBoost model has been shown to produce very accurate results with little errors, making it ideal for use in airport operations. According to forecasts for 2025, the number of passengers would range from 850,000 to 920,000 every month, with the biggest peak occurring in December 2025 (920,000).

4. Conclusion

Based on the results of the study of 11 literature research, we can conclude that:

a. Both LSTM and XGBoost algorithms are effective in predicting the number of airplane passengers in time series datasets.

b. The hybrid model (LSTM + XGBoost) outperforms with a prediction accuracy of 96%, the lowest error (RMSE = 0.015, MAPE = 2.45%).

c. The hybrid method appears to be a promising solution for improving operational efficiency and terminal capacity planning at Sultan Hasanuddin International Airport in Makassar.

d. Additional proposals include direct implementation of the model with real-time data integration and externally derived information such as weather and ticket pricing.

e. Practical Implications for Sultan Hasanuddin Airport, the combined LSTM + XGBoost model has been shown to produce very accurate results with little errors, making it ideal for use in airport operations. According to forecasts for 2025, the number of passengers would range from 850,000 to 920,000 every month, with the biggest peak occurring in December 2025 (920,000).

5. Acknowledgements

The author would like to thank everyone who helped with this research, especially the supervisor for his invaluable guidance and advice, the Polytechnic Library Team for facilitating access to relevant literature, the BPS of South Sulawesi Province and Sultan Hasanuddin Airport Makassar for providing data on the number of aircraft passengers, and all reference sources used in this study. This task could not have been done successfully without the assistance of other parties.

Reference

[1] Badan Pusat Statistik Provinsi Sulawesi Selatan, "Jumlah Penumpang Angkutan Udara di Bandara Sultan Hasanuddin," [Online]. Tersedia: https://sulsel.bps.go.id/id/statistics-table/2/MTg4MSMy/jumlah-penumpang-angkutan-udara-di-bandara-sultan-hasanuddin.html. [Diakses: 27-Jan-2025].

[2] M. S. Y. Lubis, "Implementasi Artificial Intelligence Pada Sistem Manufaktur Terpadu," 2021.

[3] A. Soy Temür, M. Akgün, and G. Temür, "Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models," *J. Bus. Econ. Manag.*, vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[4] R. Primartha, *Belajar Machine Learning*. Bandung: Informatika, 2018.

[5] M. D. Wuryandari and I. A., "Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation Dan Learning Vector Quantization Pada Pengenalan Wajah," *Jurnal Komputer dan Informatika (KOMPUTA)*, vol. 1, no. 1, 2012.

[6] R. Wijayanti, "Analisa Klasifikasi Kartu Kredit Menggunakan Algoritma Naive Bayes," *Jurnal Mantik Penusa*, vol. 10, no. 2, pp. 80–86, 2018.

[7] H. A. Sunata, F. J. Rianto, and Y., "K-Means Clustering dan Principal Component Analysis (PCA) Dalam Radial Basis Function Neural Network: Komparasi Tujuh Algoritma Identifikasi Fraud ATM Pada PT. Bank Central Asia Tbk," *Jurnal XYZ*, vol. 7, no. 3, pp. 441–450, 2020.

[8] S. Li and X. Zhang, "Research On Orthopedic Auxiliary Classification and Prediction Model Based On XGBoost Algorithm," *Neural Computing And Applications*, vol. 32, no. 7, pp. 1971–1979, 2020.

[9] N. N. et al., "Prediksi Kecelakaan Lalu Lintas Di Bali Dengan XGBoost Pada Python," *Jurnal XYZ*, vol. 8, no. 3, pp. 188–196, 2020.

[10] I. Hanif, "Implementing Extreme Gradient Boosting (XGBoost) Classifier To Improve Customer Churn Prediction," 2020.

[11] Chen, T., & Guestrin, C. (2019). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. **https://doi.org/10.1145/2939672.2939785**

[12] Graves, A. (2020). "Generating Sequences With Recurrent Neural Networks." arXiv preprint arXiv:1308.0850. **https://doi.org/10.48550/arXiv.1308.0850**

[13] Vaswani, A., et al. (2023). "Attention Is All You Need." Advances in Neural Information Processing Systems, 30. **https://doi.org/10.48550/arXiv.1706.03762**

[14] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2019). "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward." PLoS ONE, 13(3), e0194889. **https://doi.org/10.1371/journal.pone.0194889**

[15] Hochreiter, S., & Schmidhuber, J (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.