

A Comparative Study of Statistical, Machine Learning, and Hybrid Forecasting Models for Airline Passenger Time Series

Mona Mahmoud Abo El Nasr^{1*}, Sana Hafdhi², Wesam M. Amer³, Safia M. Ezzat⁴

¹Department of Applied Statistics and Insurance, Faculty of Commerce, Mansoura University, Mansoura 35516, Egypt; ²Department of Management Information System, College of Business Administration in Yanbu, Taibah University, Yanbu 42353, Saudi Arabia; ³Department of Statistics and Insurance, Faculty of Management Sciences, Sadat Academy for Management Sciences, Cairo 2222, Egypt; ⁴Statistics Dept, Faculty of Commerce, Al-Azhar University, Girls' Branch, Cairo, Egypt

ABSTRACT

This study presents a comparative analysis of six forecasting techniques ETS, ANN, ETS-ANN, SNR, LSTM, and K-NN using the well-known Air Passengers dataset, which consists of monthly measurements of U.S. air travel passengers across a twelve-year span (1949–1960). The analysis examines both the theoretical foundations and empirical performance of statistical, machine learning, and hybrid forecasting models. ETS demonstrates strong interpretability and effective modeling of trend and seasonality, while ANN and LSTM provide enhanced capabilities for capturing nonlinear patterns and long-term dependencies. The hybrid ETS-ANN model is designed to combine these strengths by modeling residual structures left unexplained by ETS, although its empirical performance depends on the presence of exploitable nonlinear patterns. K-NN offers a simple instance-based forecasting approach, whereas the SNR-based method incorporates noise-sensitive estimation. To identify the best-performing model, five evaluation criteria, ME, MAE, RMSE, MAPE, and MASE, were applied to forecast values for the year 1960. The results indicate that the ANN model consistently achieved the lowest error across all metrics, establishing it as the most accurate method in this study. In contrast, the SNR model exhibited the highest error values and performed poorly relative to the other approaches. Overall, the findings highlight the superiority of ANN for forecasting monthly airline passenger data and support the effectiveness of nonlinear machine learning techniques in time-series prediction.

ARTICLE HISTORY

Received March 25, 2026

Accepted April 27, 2026

KEYWORDS

Forecasting Models,
Hybrid ETS-ANN Model,
ANN Model,
SNR Model,
Air Passengers

1. Introduction

Forecasting plays a crucial role in demand management, as it offers estimates of future demand that serve as the foundation for various business decisions. Accurate forecasting of civil aviation activity plays a vital role in the strategic planning efforts of governments, airports, airlines, and other stakeholders. Reliable projections of future passenger volumes are particularly important for airport authorities as they guide decisions on infrastructure investments and capacity expansion. For airline companies, precise traffic forecasts are essential for optimizing financial planning, adjusting flight schedules, and refining pricing strategies. Over the years, numerous studies have been conducted to model and predict air traffic flows at airports worldwide, reflecting the ongoing importance and complexity of this task (Al-Rukaibi & Al-Mutairi, 2013; Al-Sultan et al., 2021; BaFail, 2004; Bougas, 2013; Grubb & Mason, 2001; Jin et al., 2020; Xie et al., 2014). Forecasting air traffic involves modeling complex patterns influenced by seasonality, trends, and nonlinear behaviors. Various statistical and machine learning techniques have emerged to handle these challenges. This analysis focuses on six key forecasting methodologies: Exponential Smoothing State Space (ETS), Artificial Neural Networks (ANN), a combined ETS-ANN architecture, Signal-to-Noise Ratio (SNR), Long Short-Term Memory networks (LSTM), and K-Nearest Neighbors (KNN) (Cule & De Iorio, 2012).

ETS modeling serves as a popular statistical technique for time series prediction, particularly appreciated for capturing trend and seasonal elements within nonstationary data (Hyndman & Athanasopoulos, 2018; Hyndman et al., 2002). Pegels (Pegels, 1969) provided the original ETS classification, later broadened by Gardner Jr (1985) and formalized within a state space structure by Hyndman et al. (2002) and Taylor (2003). This mathematical formulation permits exponential smoothing

procedures, including their nonlinear counterparts, to be characterized as optimal forecasts obtained from innovations state space models (Hyndman et al., 2005; Ord et al., 1997). Univariate ETS approaches have been adapted into multivariate (vector) frameworks, enabling the utilization of inter-series relationships to enhance model fit and predictive accuracy (De Silva et al., 2010). The penalized ETS framework proposed in Shi (2022) enhances coherence and predictive accuracy for mortality rate forecasting.

Time series prediction challenges have constituted a fundamental focus in machine learning research. Long Short-Term Memory (LSTM) networks, developed by Hochreiter and Schmidhuber (1997), represent a particularly effective solution. Based on recurrent neural network (RNN) architecture, these models overcome the vanishing and exploding gradient problems that limit conventional RNNs (Li et al., 2024). Consequently, LSTMs have demonstrated exceptional performance in diverse fields such as speech recognition (Zhao et al., 2024), handwriting analysis, and generative image modeling, driving substantial advances in these areas. In financial applications, LSTM networks have demonstrated strong predictive capabilities, making them widely adopted for time series forecasting in capital markets (Xu et al., 2022).

Artificial Neural Networks (ANNs), despite initial theoretical limitations, have also proven highly effective. First conceptualized by Rosenblatt in 1958, ANNs initially struggled with basic logical problems like XOR, which delayed their practical use. The introduction of backpropagation in 1986 marked a turning point, enabling efficient training of multilayer networks and accelerating progress in the field (Yuan et al., 2024). Consequently, ANNs demonstrated superior capabilities in diverse domains such as natural language processing and computer vision. Today, ANNs are recognized as versatile tools for financial forecasting (Tan et al., 2024). Functionally, ANNs simulate the human brain's neural structure for information processing (Sherstinsky, 2020). Each network consists of interconnected neurons, basic units with input and output ports whose varying connection patterns form different architectures ANNs excel in forecasting due to advantages including universal approximation capability (Hornik et al., 1989), data-driven methodologies (Fan et al., 2024), and capacity for nonlinear pattern modeling. However, standalone neural networks often struggle to capture both linear and nonlinear dynamics effectively (Zhang, 2003). Consequently, hybrid models integrating linear and nonlinear methods have emerged to overcome this limitation (Zhang, 2003). These hybrids consistently outperform individual models in forecasting accuracy. Though state space frameworks enable exponential smoothing techniques, including nonlinear variants (ETS), to achieve optimal forecasting performance (Hyndman et al., 2008), hybrid applications of these methods have been underinvestigated. Panigrahi and Behera (2017) tackled this research gap through their ETS-ANN approach, which fuses linear and nonlinear exponential smoothing models derived from innovation state space with neural network architectures. Ampountolas (2023) elaborates on this methodology, which delivers superior forecasting performance through exploiting the combined advantages of both ETS and ANN frameworks (Ritter & Muñoz-Carpena, 2013).

The Signal-to-Noise Ratio (SNR) is a forecasting technique that enhances prediction accuracy by distinguishing meaningful patterns, referred to as signals, from random fluctuations or irrelevant variations within the data (Harris, 2022). In recent years, supervised learning algorithms have been introduced for SNR estimation, demonstrating notable advancements in performance compared to conventional approaches. Suhadi et al. (2010) introduced a data-driven method that employs two trained neural networks to estimate the a priori SNR. Papadopoulos et al. (2017) developed a channel-adapted deep neural network (DNN) that leverages energy ratio features and i-vectors to construct a DNN model capable of estimating global SNR under both known and unknown channel conditions. A two-stage framework is proposed for global SNR estimation, in which the initial stage generates noise residuals derived from a speech enhancement model. In the second stage, the noise residuals are fed into a DNN to predict the global SNR. A higher SNR value suggests that the underlying trend within the time series is more prominent relative to noise, thereby improving the overall reliability of forecasting outcomes (Hyndman & Athanasopoulos, 2018). In real-world applications, SNR-based filtering or transformation is commonly utilized as a preprocessing step to enhance model performance, particularly when handling noisy or highly volatile datasets (Shumway & Stoffer, 2017). K-Nearest Neighbors (KNN) is among the most widely adopted prospective statistical classification algorithms. It operates by categorizing objects according to their closest training instances within the feature space (Lipar et al., 2015; Papadaniil & Hadjileontiadis, 2013). The KNN method is a nonparametric forecasting technique that predicts future values based on the similarity between historical patterns (Altman, 1992).

In time series forecasting, KNN identifies the most similar past observations to the current state and averages their subsequent values to generate forecasts. Due to its simplicity and adaptability, KNN has been widely applied in demand and financial forecasting, often serving as a baseline or complementary model in hybrid approaches (Ahmed et al., 2010). This study aims to evaluate and compare the predictive performance of six models spanning statistical, machine learning, and hybrid approaches ETS, ANN, ETS-ANN, SNR, LSTM, and KNN using the Air Passengers time series. The models are evaluated based on their ability to capture trend, seasonality, and nonlinear dynamics, with forecast accuracy assessed using five error measures: ME, MAE, RMSE, MAPE, and MASE. By analyzing out-of-sample forecasts for the year 1960, the study seeks to determine the most accurate and practically applicable forecasting method for predicting monthly airline passenger demand. The findings provide empirical evidence on the relative effectiveness of nonlinear machine learning techniques compared to traditional and hybrid time-series models.

2. Forecasting Model

2.1 Exponential Smoothing State Space (ETS)

ETS models are a class of forecasting methods that break down a time series into three core components: Error, Trend, and Seasonality (ETS). This section synthesizes the mathematical foundations, estimation techniques, and applications of ETS models, emphasizing their flexibility and accuracy. It presents the unified state space framework proposed introduced by [Hyndman et al. \(2008\)](#), derives key equations, and discusses model selection. ETS models extend classical exponential smoothing ([Holt, 1957](#); [Winters, 1960](#)) by embedding them within a probabilistic state space structure. The ETS framework classifies y_t each component into distinct types: the Error component is categorized as either additive (A) or multiplicative (M); the Trend component is specified as none (N), additive (A), additive damped (Ad), multiplicative (M), or multiplicative damped (Md); and the Seasonality component is defined as none (N), additive (A), or multiplicative (M).

[Hyndman et al. \(2008\)](#) classify 30 distinct ETS model specifications based on different combinations of these components.

Let:

$l_t = \text{level}$, $b_t = \text{trend}$, $s_t = \text{Seasonal component}$, $m = \text{Seasonal period}$, $\varepsilon_t = \text{white noise}$

The general notation is ETS (Error, Trend, Seasonality). For example, ETS (A, A, A) denotes additive error, additive trend, and additive seasonality.

Among the most widely recognized ETS configurations is:

- ETS (A, N, N), which corresponds to Simple Exponential Smoothing

$$\hat{y}_{t/t-1} = l_{t-1}$$

$$l_t = l_{t-1} + \alpha \varepsilon_t$$

$$\varepsilon_t = y_t - \hat{y}_{t/t-1}$$

- ETS (A, A, N): Holt's Linear Trend Model

$$\hat{y}_{t/t-1} = l_{t-1} + b_{t-1}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t$$

$$b_t = b_{t-1} + \beta \varepsilon_t$$

- ETS (A, A, A): Additive Holt-Winters Model

$$\hat{y}_{t/t-1} = l_{t-1} + b_{t-1} + s_{t-m}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t$$

$$b_t = b_{t-1} + \beta \varepsilon_t$$

$$s_t = s_{t-m} + \gamma \varepsilon_t$$

- ETS (M, A, M): Multiplicative Error and Seasonality

$$\hat{y}_{t/t-1} = (l_{t-1} + b_{t-1}) \times s_{t-m}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$$

$$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m}}$$

$$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{l_{t-1} + b_{t-1}}$$

- Damped Trend Model ([Gardner Jr & McKenzie, 1985](#)) For ETS (A, Ad, N):

$$b_t = \varphi b_{t-1} + \beta \varepsilon_t, \quad \varphi \in [0, 1]$$

2.1.1 State Space Representation

The general ETS state space model can be expressed as ([Hyndman et al., 2008](#)):

$$y_t = w(x_{t-1}) + r(x_{t-1})\varepsilon_t$$

$$x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t$$

Where the state vector is:

$$x_t = l_t, b_t, s_t, s_{t-1}, s_{t-m+1}$$

2.1.2 Estimation

ETS parameters are typically estimated via maximum likelihood:

$$L(\theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log \left(\sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2} \right)$$

2.1.3 Strengths and Limitations

2.1.3.1 Advantages

- Accommodates diverse trends and seasonal structures
- Robust performance in short-term forecasting
- Enables probabilistic forecasts

2.1.3.2 Limitations

- Assumes linear trends unless damped
- Multiplicative models are sensitive to zero or near-zero values.

2.1.4 Applications

ETS models are widely applied in retail sales forecasting (ETS (M, A, M)), energy demand forecasting (ETS (A, Ad, A)), and inventory management (Croston-type variants). Overall, ETS provides a coherent and flexible forecasting framework that often outperforms ad-hoc exponential smoothing approaches (Bandara et al., 2020).

2.2 Artificial Neural Networks (ANNs)

ANNs are powerful data-driven models widely used for function approximation and forecasting. One of their primary advantages lies in their universal approximation capability, enabling them to model complex and nonlinear relationships without requiring explicit model assumptions (Ma, 2020).

In temporal data prediction, the most commonly employed ANN architecture is the single hidden-layer feedforward neural network, which is composed of an input layer, one hidden layer, and an output layer interconnected through acyclic links (Zhang et al., 1998). The association between inputs and outputs can be mathematically represented as:

$$y_t = w_0 + \sum_{j=1}^q w_j g(w_{0,j} + \sum_{i=1}^p w_{ij} y_{t-i}) + \varepsilon_t$$

Where $w_i (i = 0, 1, 2, \dots, q)$ and $w_{ij} (i = 0, 1, 2, \dots, p; j = 0, 1, 2, \dots, q)$ are model parameters called the connection weight, p is the number of input nodes, and q is the number of hidden nodes, y_t is the output variable, and $y_{t-1}, y_{t-2}, \dots, y_{t-n}$ These are the input variables.

2.3 ETS-ANN Hybrid Model

The ETS-ANN hybrid framework integrates the advantages of both statistical and machine learning models. ETS captures linear trend and seasonal patterns, while ANN models the nonlinear structure present in the residuals. This hybrid methodology frequently yields superior forecasting performance compared to standalone models, particularly for series exhibiting both strong structure and nonlinear behavior (Petropoulos et al., 2014).

Sequential Hybrid Strategy (Smyl, 2020):

- Stage 1: ETS Modeling
Fit an optimal ETS model to the original series y_t , and obtain residuals:

$$e_t = y_t - \hat{y}_t|ETS$$

- Stage 2: ANN Modeling of Residuals:

Train an ANN using:

- 1- Lagged residuals ($e_{t-1}, e_{t-2}, \dots, e_{t-p}$)
- 2- Lagged original series values ($y_{t-1}, y_{t-2}, \dots, y_{t-q}$)
- 3- Exogenous variables ($X_{1,t}, X_{2,t}, \dots, X_{k,t}$), if available and relevant.
- 4- Time-based features (e.g., hour, day of week, month) are especially useful for complex seasonality. The ANN

learns a function $f(\cdot)$ Such that (Taieb & Atiya, 2015):

$$\hat{e}_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-p}, y_{t-1}, y_{t-2}, \dots, y_{t-q}, X_{1,t}, X_{2,t}, \dots, X_{k,t}, \text{Time Features}) + \varepsilon_t$$

where ε_t , is the ANN model error (ideally white noise). As inputs (Taieb & Atiya, 2015):

Stage 3: Hybrid Forecasting:

Generate

- ETS forecast $\hat{y}_{t+h}|ETS$ for horizon h .

- ANN forecast $\hat{e}_{t+h}$

The final hybrid forecast is the sum (Zhang, 2003):

$$\hat{y}_{t+h|Hybrid} = \hat{y}_{t+h|ETS} + \hat{e}_{t+h}$$

2.3.1 Advantages

- 1- Higher accuracy than standalone ETS or ANN.
- 2- Captures both linear/seasonal and non-linear structures.
- 3- Handles complex residual behavior.
- 4- Supports exogenous predictors.
- 5- Decomposes information cleanly.

2.3.2 Challenges and Considerations

- 1- Model selection for ETS and ANN is complex.
- 2- ANNs risk overfitting, requiring careful validation.
- 3- Increased computational cost.
- 4- Multi-step forecasting for residuals is non-trivial.
- 5- Reduced interpretability due to the ANN component
- 6- Error propagation from ETS to the ANN inputs.

2.4 Signal-to-Noise Ratio (SNR)

"SNR evaluates system fidelity by quantifying the ratio of signal variance to noise variance. Despite its widespread application in physical and communication systems, accurately estimating SNR poses significant challenges in non-Gaussian or non-additive environments (Czanner et al., 2008).

SNR is characterized as the proportion of signal power relative to noise power. It can also be expressed in decibels using $10\log_{10}(SNR)$. This metric is widely applied to evaluate system fidelity and compare performance across different systems. A higher SNR indicates that the signal is less distorted by noise. Because of its importance as a measure of signal quality, methods for SNR estimation have been extensively studied across various applications (Zheng et al., 2024).

SNR estimation is typically carried out using the received signal. In standard communication systems, the transmitter sends a signal that propagates through the channel before reaching the receiver. The received signal can be expressed as (Zheng et al., 2024):

$$y(n) = h(n) \otimes b(n)e^{j(2\pi\Delta f n + \theta)} + w(n), n \in [0, N - 1]$$

where $b(n)$ is the transmitted signal, $h(n)$, is the channel response, Δf is the carrier frequency offset, θ is the random phase shift, $w(n)$ is the noise, N is the signal length, and $\otimes$ denotes the convolution operation.

Traditional methods often struggle to accurately separate the true signal from noise and to estimate SNR reliably, primarily due to channel propagation effects and limited knowledge of noise characteristics. Given the received signal $y(n)$, which includes additive noise, the noise-free component referred to as the effective signal can be written as:

$$x(n) = h(n) \otimes b(n)e^{j(2\pi\Delta f n + \theta)}$$

The signal power is defined as:

$$P_x = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2$$

And the noise power is:

$$P_w = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |w(n)|^2$$

Thus, the SNR of $y(n)$, is:

$$SNR = \frac{P_x}{P_w}$$

SNR is typically expressed in decibels (dB) as:

$$SNR_{dB} \triangleq p = 10 \log_{10} \left(\frac{p_x}{p_w} \right)$$

The SNR estimation problem consists of estimating the SNR of the received signal $y(n)$, which can be formulated as:

$$\hat{p} = \max_p \Pr\{p|y(n)\}$$

Where $\hat{p}$ is the estimated SNR and $\Pr\{.\}$ denotes probability.

2.5 The Long Short-Term Memory (LSTM)

The LSTM network is a sophisticated neural architecture that builds upon the conventional Recurrent Neural Network (RNN) framework. First developed by [Hochreiter and Schmidhuber \(1997\)](#), LSTM is part of the wider RNN model family and is composed of recurring modules that operate in a manner analogous to sequential neural units ([Greff et al., 2016](#)). The architecture was specifically designed to overcome the limitations of conventional RNNs, most notably the vanishing gradient problem. This issue arises when gradient values diminish as they propagate backward through time, preventing adequate weight updates and hindering the model's ability to learn long-term dependencies ([HOERL & KENNARD, 1970](#)).

In contrast to conventional RNNs, LSTM incorporates a cell state mechanism that allows the model to preserve information over extended sequences, enabling it to effectively learn and capture long-term dependencies. This mechanism allows LSTM to generalize well and perform effectively on both large and small datasets. Its ability to detect complex, non-linear patterns makes it particularly suitable for improving prediction accuracy in time-dependent data ([Purnomo et al., 2021](#)).

Because of its design, LSTM performs exceptionally well on sequential or temporal tasks, especially those involving seasonality, trends, and other temporal structures. It has been extensively utilized across various domains, including financial temporal data prediction, speech recognition, and production forecasting. As an illustration, in the modeling of Crude Palm Oil (CPO) production, LSTM demonstrates the ability to account for both seasonal fluctuations and long-term dependencies, resulting in more consistent and precise predictions. These properties make LSTM a robust tool for solving complex and dynamic problems where traditional approaches frequently fail, thereby supporting improved decision-making and strategic planning ([Ulama et al., 2025](#)).

The LSTM model controls the flow of information through three fundamental gates: the input gate, the forget gate, and the output gate. These gates collectively govern which information is retained, modified, or discarded within the cell state. Figure 2 depicts the internal structure of an LSTM cell, which constitutes the core unit responsible for processing sequential data in temporal data prediction applications ([Ladjal et al., 2025](#)).

2.5.1 Input Gate

The input gate regulates the extent to which new information is incorporated into the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

Where i_t is the input gate, W_i is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_i , is the bias.

2.5.2 Forget Gate

The forget gate determines the portion of previous information retained:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Where f_t , is the forget gate, W_f is the weight matrix, and b_f , is the bias term.

2.5.3 Cell State Update

The updated cell state is computed as:

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

Where C_t is the cell state at time t, C_{t-1} is the previous cell state, and W_c , is the weight matrix for the cell state

update.

2.5.4 Output Gate

The output gate determines the information transmitted to the subsequent time step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \text{ and } h_t = o_t * \tanh C_t$$

Where h_t is the new hidden state, influenced by the output of the cell state C_t , and o_t , is the output gate.

2.6 K-Nearest Neighbors (K-NN)

K-NN is a non-parametric approach for classification and regression that falls under the category of instance-based learning, in which the algorithm retains the training dataset rather than building an explicit generalized model. K-NN is often considered a typical example of a lazy learner, not due to simplicity, but because it does not build a discriminative function during training. Instead, it memorizes training samples and performs computations only when making predictions (Chen & Beaulieu, 2007). Lazy learning is a form of instance-based learning characterized by negligible computational cost during the training phase (Raschka, 2015).

K-NN is an oversaw machine learning technique used to predict the class or value of an unknown data point based on its distance from labeled samples in the training dataset. The algorithm identifies the k nearest points to the query instance and determines the prediction either by majority voting (classification) or by averaging the target values (regression). Due to its simplicity and ease of implementation, K-NN is widely applied in straightforward classification and regression tasks.

Once the value of k is selected, predictions can be generated using the K-NN model. In regression problems, the K-NN prediction is calculated as the mean of the outcomes of the k closest neighbors (Imandoust & Bolandraftar, 2013):

$$y = \frac{1}{k} \sum_{i=1}^k y_i$$

Where y_i represents the output value of the i^{th} neighbor and y is the predicted value for the query point.

To improve prediction robustness, a weight can be assigned to each nearest neighbor based on its distance from the query point. A commonly used weighting scheme is:

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^k \exp(-D(x, p_i))}$$

Where $(D(x, p_i))$, is the distance between the query point x and the i^{th} neighbor p_i .

These weights satisfy the condition:

$$\sum_{i=1}^k W(X_0, X_i) = 1$$

Thus, in regression tasks, the weighted prediction becomes:

$$y = \sum_{i=1}^k W(X_0, X_i) y_i$$

Or classification tasks, the predicted class is the one associated with the maximum weighted sum.

Moreover, when $K > 1$, it becomes straightforward to estimate the standard deviation of predictions in regression problems, which provides a measure of prediction uncertainty:

$$errbar = \pm \sqrt{\frac{1}{K-1} \sum_{i=1}^k (y - y_i)^2}$$

3. Model Selection Criterion

In model selection, we typically consider a set of models $M_1, \dots, M_m$, where usually $m > 2$. These models may have a nested structure, where one model is a special case of another, or they may be non-nested. Instead of performing multiple hypothesis tests between pairs of models each time rejecting or accepting one, it is more appropriate to employ a criterion that selects the most suitable model overall. The purpose of model selection is to prevent overfitting and improve forecasting performance.

To avoid overfitting, several statistical criteria have been proposed for evaluating model accuracy. One of the earliest measures is the Mean Squared Error (MSE) (Wackerly et al., 2008), defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Willmott and Matsuura (2005) introduced a different criterion, the Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i)|$$

Momani et al. (2024) used another criterion, the Mean Absolute Percentage Error (MAPE), calculated as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|(y_i - \hat{y}_i)|}{y_i}, y_i \neq 0$$

Waller (2003) introduced the Mean Error (ME) criterion:

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

Khan and Hildreth (2003) described another criterion, the Mean Percentage Error (MPE):

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{y}_i)}{y_i}, y_i \neq 0$$

4. Numerical Analysis

This study relied on the monthly totals of U.S. airline passengers from 1949 to 1960. The dataset was obtained from a built-in dataset in R called Air Passengers, which is widely used in time-series analysis. The same dataset is also available on Kaggle at: <https://www.kaggle.com/datasets/brmik07/air-passengers-dataset>. Researchers often use it in forecasting models, seasonal fluctuations, smoothing, and decomposition, analyzing patterns, trends.

Table 1 presents the descriptive statistics for the Air Passengers dataset, including the number of observations, mean, standard deviation, minimum and maximum values, first quartile (Q1), third quartile (Q3), skewness, and kurtosis. The dataset contains 144 monthly observations.

The average number of passengers per month is approximately 280, with a standard deviation of about 120. The highest recorded value is 622 passengers in July 1960, while the lowest value is 104 passengers in November 1949. The first quartile (Q1) is 180 passengers, indicating that 25% of the monthly observations fall below this value. The median (Q2) is 265 passengers, and the third quartile (Q3) is 360 passengers, meaning that 75% of the observations fall below this level.

Table 1: Descriptive Statistics of Air Passengers Data

Observations	Mean	SD	Median	Min	Max	Q1	Q3	Skewness	Kurtosis
144	280.30	119.97	265.50	104	622	180	360.50	0.58	2.61

The data displayed evident seasonality, with air passenger numbers rising during the summer months and declining throughout the winter months. The series demonstrated stationarity at the level, as confirmed by the outcomes of both the ADF and KPSS tests. The Augmented Dickey-Fuller (ADF) test returned a p-value of 0.01, indicating rejection of the null hypothesis of a unit root. Likewise, the KPSS test produced a p-value of 0.01, reinforcing the conclusion of stationarity. These results are illustrated in Figure 1.

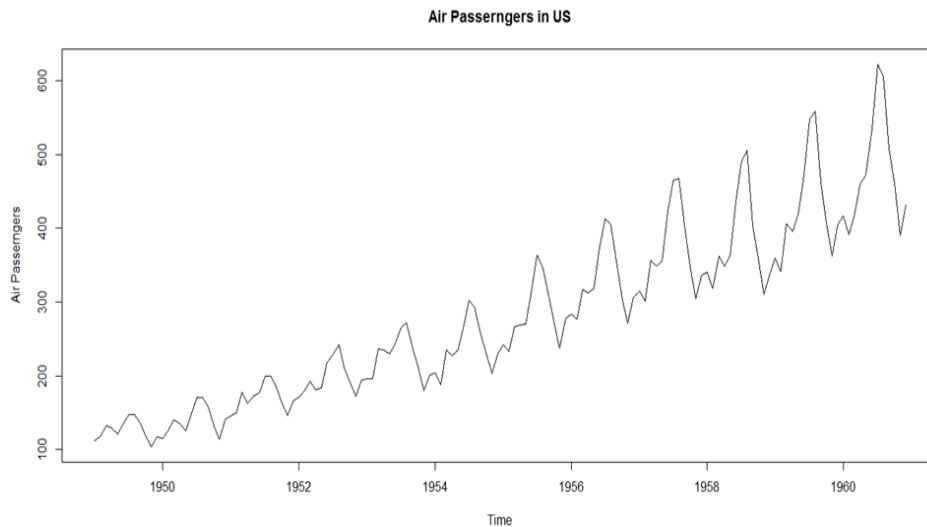


Figure 1: Air Passengers Data

To compare the six forecasting techniques (ETS, ANN, ETS-ANN, SNR, LSTM, and K-NN), the dataset was partitioned into two segments. The first segment functioned as the training set, encompassing the period from 1949 to 1959, while the second segment served as the test set, comprising the data for 1960.

Table 3 summarizes the predicted values generated by all models. The outcomes were derived using R software, version 4.2.2.

Table 2: Forecast of all Models

	Actual	ETS	ANN	ETS-ANN	SNR	LSTM	KNN
Jan 1960	417	411.9	398.5	411.9	1014.0	410.7	204
Feb 1960	391	407.0	381.4	407.0	1001.9	405.7	277
Mar 1960	419	467.3	437.1	467.3	1150.5	441.5	233
Apr 1960	461	450.7	435.9	450.7	1109.6	434.4	270
May 1960	472	451.5	458.4	451.5	1111.6	438.5	234
Jun 1960	535	513.2	512.3	513.2	1263.5	476.3	270
Jul 1960	622	569.9	598.1	569.9	1403.0	511.6	347
Aug 1960	606	567.6	618.9	567.6	1397.3	512.8	347
Sep 1960	508	496.1	518.9	496.1	1221.4	468.6	259
Oct 1960	461	432.2	455.3	432.2	1064.0	433.8	274
Nov 1960	390	376.6	406.6	376.6	927.2	403.5	203
Dec 1960	432	424.8	440.2	424.8	1045.7	431.3	278

To determine the most accurate forecasting model, each predicted value was compared with its corresponding actual observation.

For example, in January 1960, the actual number of U.S. airline passengers was 417, and the closest forecast was produced by both the ETS and ETS-ANN models, each estimating approximately 412, resulting in an error of 5 passengers.

During March 1960, the ANN model yielded the most accurate prediction, estimating 437 passengers against the actual recorded figure of 419.

During December 1960, the LSTM model attained the greatest level of accuracy, recording a deviation of merely 1 passenger, which rendered it the top-performing model for that particular month.

The forecasting results for all six models are illustrated in the following figures, from Figure 2 to Figure 8.

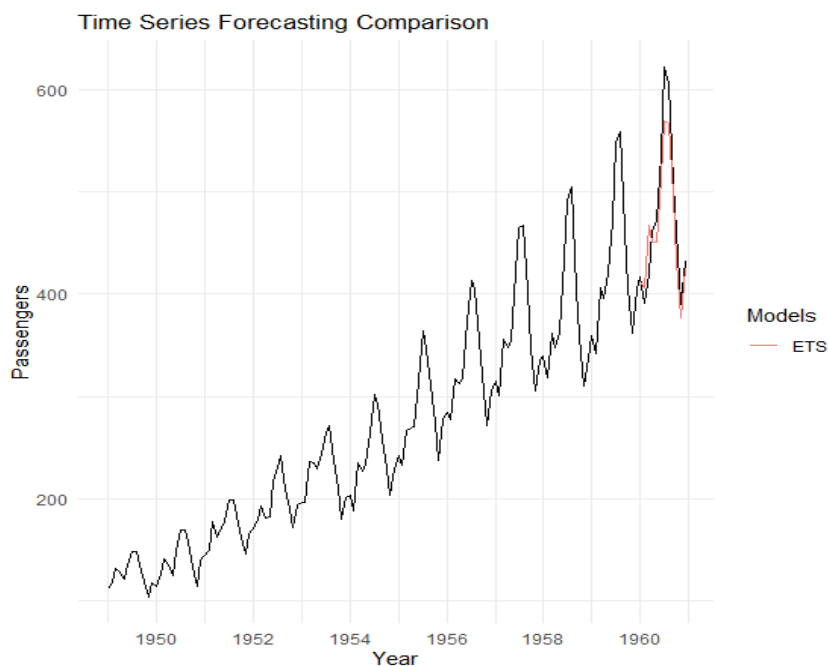


Figure 2: ETS Forecasting of Air Passengers Data

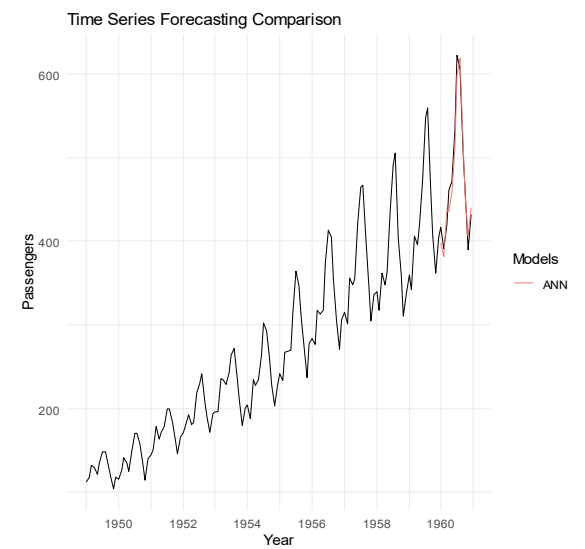


Figure 3: ANN Forecasting of Air Passengers Data

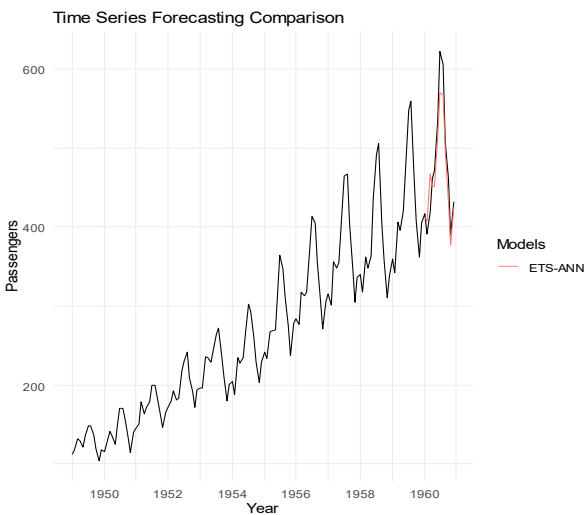


Figure 4: ETS-ANN Forecasting of Air Passengers Data

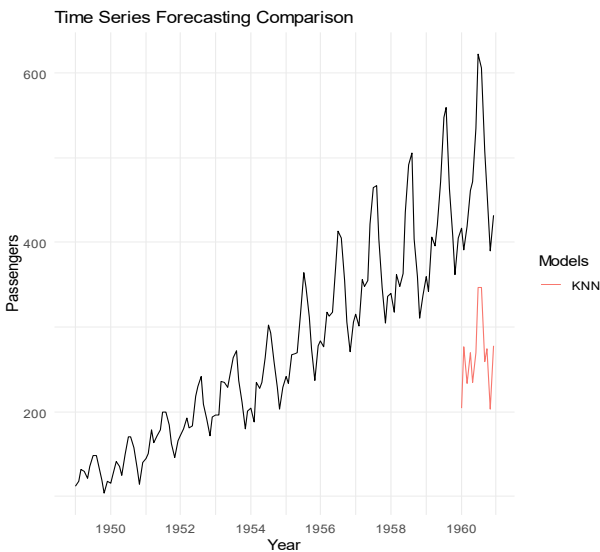


Figure 5: KNN Forecasting of Air Passengers Data

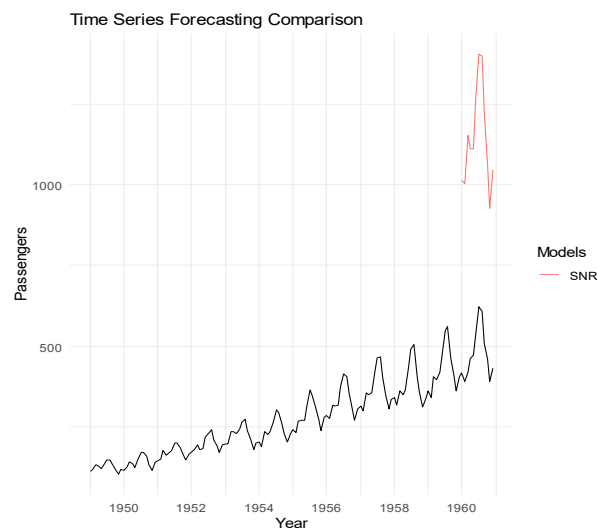


Figure 6: SNR Forecasting of Air Passengers Data

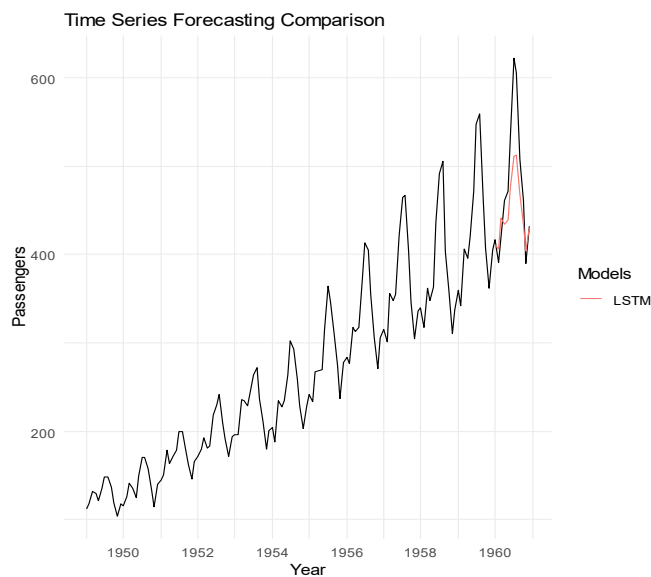


Figure 7: LSTM Forecasting of Air Passengers Data

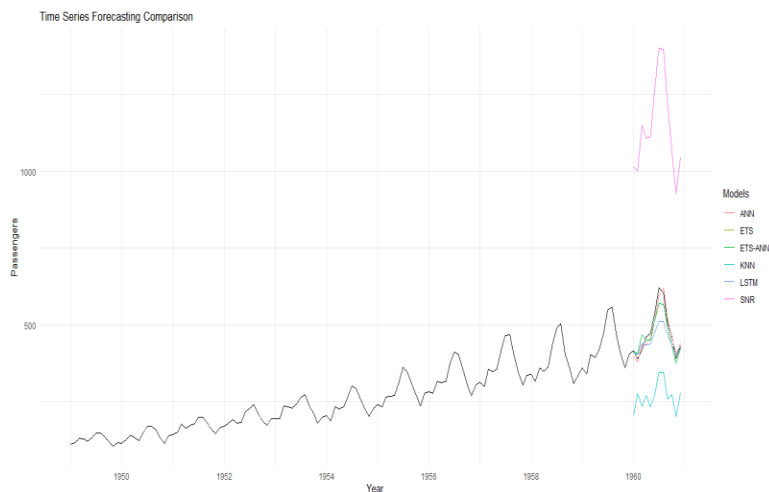


Figure 8: Forecasting of Air Passengers Data

To guarantee the reliability of the study, five evaluation criteria were employed to determine the most accurate forecasting model. Table 3 displays the outcomes of each criterion across all forecasting models, derived from the errors produced by each respective model.

Table 3: Model Selection Criteria

Model	ME	MAE	RMSE	MAPE	MASE
ETS	12.08	22.80	27.40	4.66	0.47
ANN	4.38	15.48	16.64	3.27	0.32
ETS-ANN	12.08	22.80	27.40	4.66	0.47
KNN	209.83	209.83	214.99	43.85	4.35
SNR	-666.31	666.31	670.77	141.13	13.80
LSTM	28.77	37.22	49.48	7.07	0.77

Table 4 provides a comparative assessment of forecasting models utilizing multiple accuracy measures, encompassing ME, MAE, RMSE, MAPE, and MASE. These criteria collectively evaluate both the magnitude and scale-adjusted performance of each model. Among all examined models, the Artificial Neural Network (ANN) demonstrates the most superior overall forecasting performance, attaining the lowest values across MAE, RMSE, MAPE, and MASE. This indicates superior accuracy and robustness relative to the benchmark model, highlighting ANN's ability to capture nonlinear patterns in the data effectively.

The ETS and ETS-ANN hybrid models yield identical performance measures across all criteria. This suggests that, in this application, the hybridization process did not contribute additional predictive gains beyond those already captured by the ETS component. This outcome may indicate that the residual structure left by the ETS model does not contain substantial nonlinear patterns exploitable by the ANN.

The LSTM model demonstrates moderate forecasting accuracy, outperforming ETS-based approaches but remaining inferior to the ANN model across all error metrics. While LSTM is capable of learning long-term dependencies, its performance may be affected by data size, parameter tuning, or model complexity.

In contrast, KNN and SNR-based models perform poorly, as evidenced by substantially higher error values across all criteria. The SNR model, in particular, shows extreme forecast deviations, indicating its unsuitability for this forecasting task.

Overall, the results indicate that ANN provides the most accurate and reliable forecasts, followed by LSTM, while ETS-based and distance-based methods show limited effectiveness for the analyzed dataset. Figures (9) through (13) further illustrate that the error values of the ANN model are closest to zero, confirming its superiority. In contrast, the SNR model performed the worst, showing the highest error values among all forecasting models.

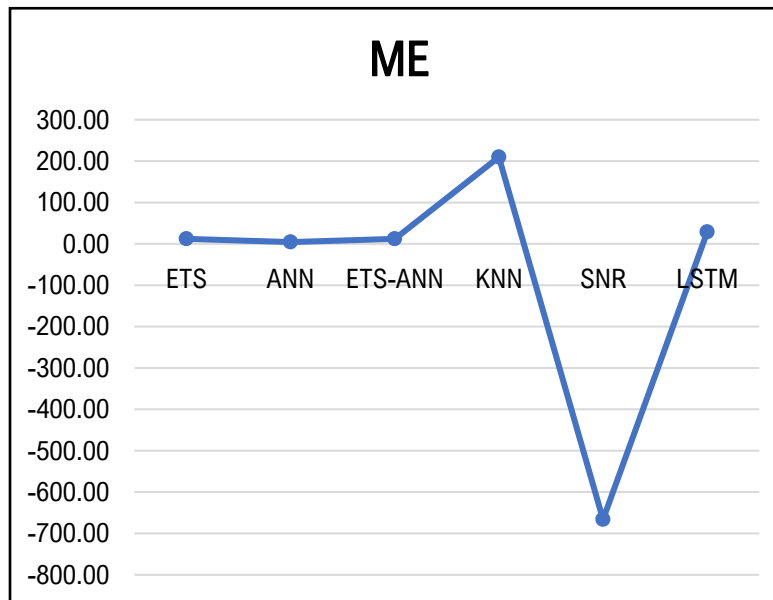


Figure 9: ME Criteria of Air Passengers Forecasting Data

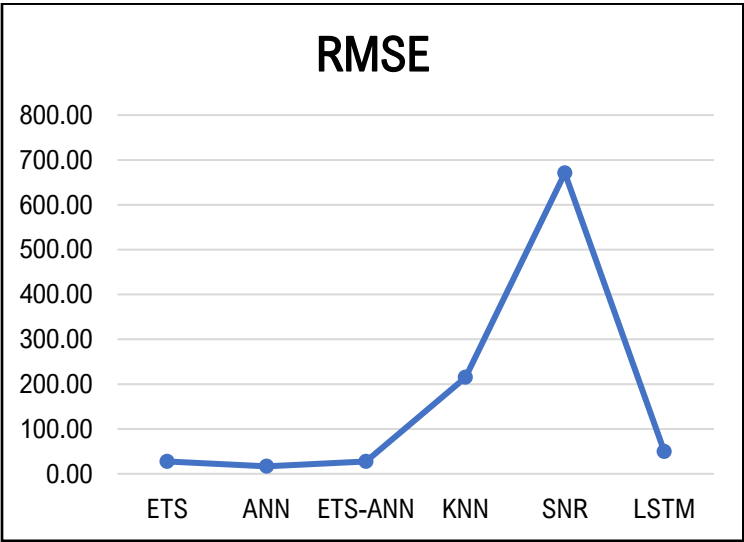


Figure 10: RMSE Criteria of Air Passengers Forecasting Data

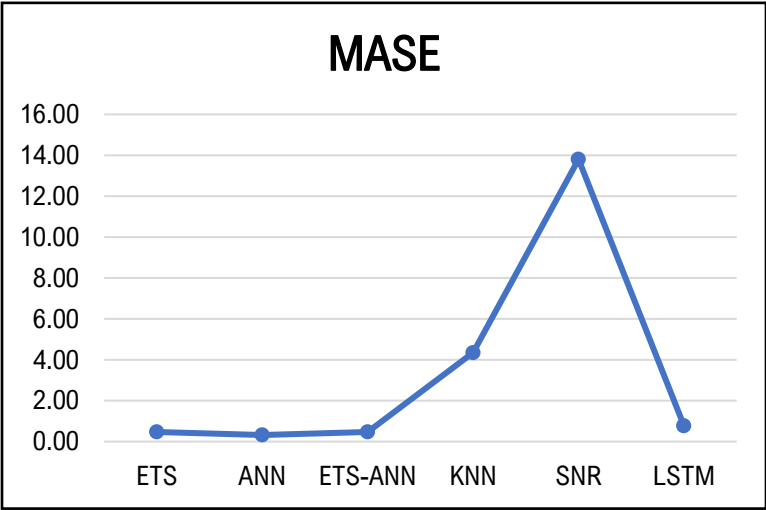


Figure 11: MSE Criteria of Air Passengers Forecasting Data

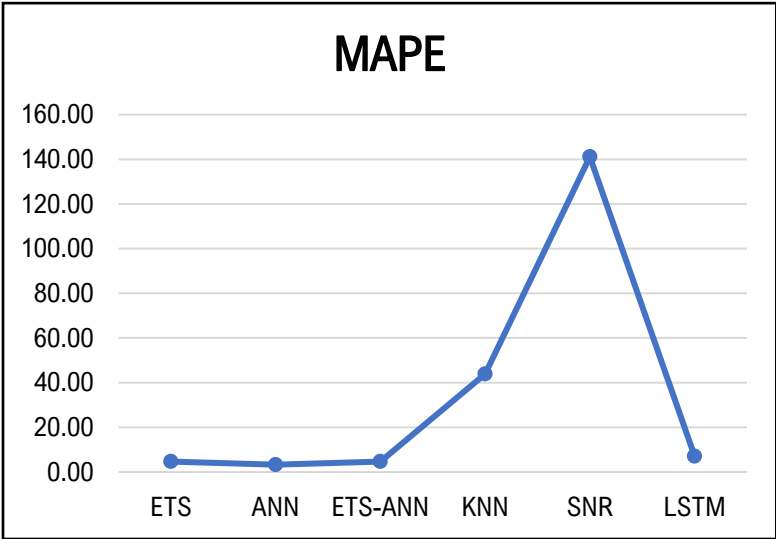


Figure 12: MAPE Criteria of Air Passengers Forecasting Data

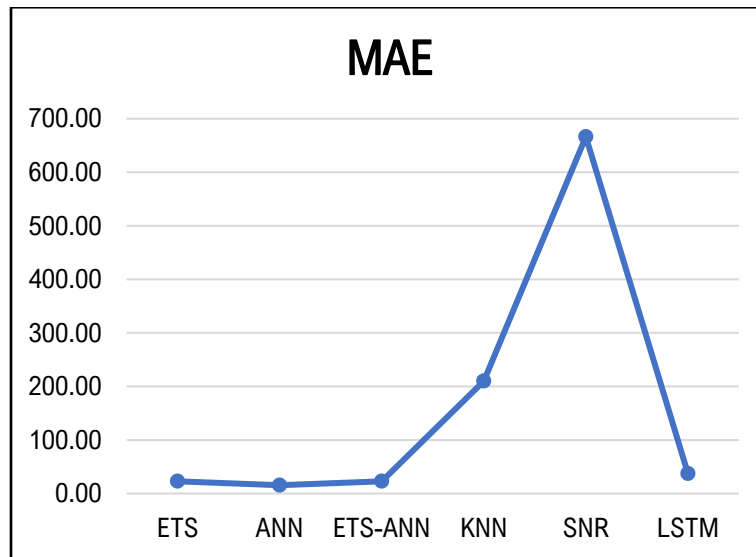


Figure 13: MAE Criteria of Air Passengers Forecasting Data

All evaluation criteria indicate that the ANN model ranks first, achieving the minimum values for ME, MAE, RMSE, MAPE, and MASE. The ETS and ETS-ANN models occupy the second rank, as they produce identical values across the same criteria. The LSTM model ranks third, while the SNR model ranks last, showing the highest values for all five error measures.

5. Conclusion

This comparative analysis examined the theoretical foundations, strengths, and limitations of several widely used forecasting methodologies. Exponential Smoothing State Space (ETS) models demonstrate strong interpretability and efficiently account for trend and seasonal components in univariate time series; nevertheless, their capacity to represent complex nonlinear patterns remains inherently constrained. By contrast, Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models provide greater adaptability in identifying nonlinear relationships and long-term temporal dependencies, respectively. Despite these benefits, such models generally demand larger datasets, considerable computational resources, and frequently exhibit diminished interpretability.

Hybrid forecasting approaches, such as the ETS-ANN model, integrate statistical and machine learning techniques to mitigate the individual weaknesses of each method. By allowing machine learning components to model residual structures not captured by statistical models, hybrid methods can potentially enhance forecasting accuracy. The K-Nearest Neighbors (K-NN) approach provides a simple, training-free forecasting mechanism; however, its performance is constrained by the curse of dimensionality and its limited capability to extrapolate beyond the observed data range. Similarly, while the Signal-to-Noise Ratio (SNR)-based method offers robustness to noise through the use of ϵ -insensitive loss functions, its effectiveness is highly dependent on appropriate kernel selection and careful hyperparameter tuning.

In this study, the Air Passengers dataset was employed to evaluate the forecasting performance of six different models using five error-based accuracy criteria. The empirical findings reveal that the ANN model persistently surpassed all competing approaches across every evaluation metric, highlighting its exceptional predictive capability for the dataset under examination. Conversely, the SNR-based model exhibited the weakest performance, yielding the highest error values. These results emphasize the significance of choosing forecasting models that strike an appropriate balance between flexibility, accuracy, and practical suitability in accordance with the inherent characteristics of the data.

6. Data Availability

The data in this study were obtained from an open online repository from <https://www.kaggle.com/datasets/abhishekmamidi/air-passengers>

7. References

Ahmed, N., Atiya, A., Gayar, N. E., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time

- Series Forecasting. *Econometric Reviews*, 29(5-6), 594-621.
- Al-Rukaibi, F., & Al-Mutairi, N. (2013). Forecasting air travel demand of Kuwait: A comparison study by using regression vs. artificial intelligence. *Journal of Engineering Research*, 1(1), 113-143.
- Al-Sultan, A., Al-Rubkhi, A., Alsaber, A., & Pan, J. (2021). Forecasting air passenger traffic volume: evaluating time series models in long-term forecasting of Kuwait air passenger data. *Advances and Applications in Statistics*, 70(1), 69-89.
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- Ampountolas, A. (2023). Comparative analysis of machine learning, hybrid, and deep learning forecasting models: evidence from European financial markets and bitcoins. *Forecasting*, 5(2), 472-486.
- BaFail, A. O. (2004). APPLYING DATA MINING TECHNIQUES TO FORECAST NUMBER OF AIRLINE PASSENGERS IN SAUDI ARABIA (DOMESTIC AND INTERNATIONAL TRAVELS). *Journal of Air Transportation*, 9(1).
- Bandara, K., Bergmeir, C., & Hewamalage, H. (2020). LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4), 1586-1599.
- Bougas, C. (2013). Forecasting air passenger traffic flows in Canada: an evaluation of time series models and combination methods.
- Chen, Y., & Beaulieu, N. C. (2007). Maximum likelihood estimation of SNR using digitally modulated signals. *IEEE Transactions on Wireless Communications*, 6(1), 210-219.
- Cule, E., & De Iorio, M. (2012). A semi-automatic method to guide the choice of ridge parameter in ridge regression. *arXiv preprint arXiv:1205.0686*.
- Czanner, G., Sarma, S. V., Eden, U. T., & Brown, E. N. (2008). A signal-to-noise ratio estimator for generalized linear model systems. Proceedings of the world congress on engineering,
- De Silva, A., Hyndman, R. J., & Snyder, R. (2010). The vector innovations structural time series framework: a simple approach to multivariate forecasting. *Statistical Modelling*, 10(4), 353-374.
- Fan, X., Tao, C., & Zhao, J. (2024). Advanced Stock Price Prediction with xLSTM-Based Models: Improving Long-Term Forecasting. 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCM),
- Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1), 1-28.
- Gardner Jr, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31(10), 1237-1246.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232.
- Grubb, H., & Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. *International Journal of Forecasting*, 17(1), 71-82.
- Harris, F. J. (2022). *Multirate signal processing for communication systems*. River Publishers.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- HOERL, A. E., & KENNARD, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
- Holt, C. (1957). Forecasting seasonals and trends by exponentially weighted averages (ONR Memorandum No. 52). *Carnegie Institute of Technology, Pittsburgh USA*, 10.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.
- Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of forecasting*, 24(1), 17-37.

- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439-454.
- Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International journal of engineering research and applications*, 3(5), 605-610.
- Jin, F., Li, Y., Sun, S., & Li, H. (2020). Forecasting air passenger demand with a new hybrid ensemble approach. *Journal of Air Transport Management*, 83, 101744.
- Khan, A., & Hildreth, W. B. (2003). Case studies in public budgeting and financial management.
- Ladjal, B., Nadour, M., Bechouat, M., Hadroug, N., Sedraoui, M., Rabehi, A., Guermoui, M., & Agajie, T. F. (2025). Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria. *Scientific Reports*, 15(1), 15404.
- Li, X., Ma, Y., Huang, Y., Wang, X., Lin, Y., & Zhang, C. (2024). Integrated optimization of large language models: Synergizing data utilization and compression techniques. *Preprints*. 10.20944/preprints202409.0662.v1.
- Lipar, P., Čudina, M., Šteblaj, P., & Prezelj, J. (2015). Automatic recognition of machinery noise in the working environment. *Strojniški vestnik-Journal of Mechanical Engineering*, 61(12), 698-708.
- Ma, Q. (2020). Comparison of ARIMA, ANN and LSTM for stock price prediction. E3S web of conferences,
- Momani, M. A., Tashtush, S. A., Shahrour, R. J., & AlSafari, A. M. (2024). Modeling of Long-Term Load Forecast in Jordan Based on Statistical Techniques. *Journal of Electrical and Computer Engineering*, 2024(1), 8255513.
- Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, 92(440), 1621-1629.
- Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS-ANN model for time series forecasting. *Engineering applications of artificial intelligence*, 66, 49-59.
- Papadaniil, C. D., & Hadjileontiadis, L. J. (2013). Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. *IEEE Journal of Biomedical and Health Informatics*, 18(4), 1138-1152.
- Papadopoulos, P., Travadi, R., & Narayanan, S. S. (2017). Global SNR Estimation of Speech Signals for Unknown Noise Conditions Using Noise Adapted Non-Linear Regression. *Interspeech*,
- Pegels, C. C. (1969). Exponential forecasting: some new variations. *Management Science*, 311-315.
- Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for Courses' in demand forecasting. *European Journal of Operational Research*, 237(1), 152-163.
- Purnomo, H., Suyono, H., & Hasanah, R. N. (2021). Peramalan Beban Jangka Pendek Sistem Kelistrikan Kota Batu Menggunakan Deep Learning Long Short-Term Memory. *Transmisi: Jurnal Ilmiah Teknik Elektro*, 23(3), 97-102.
- Raschka, S. (2015). Python Machine Learning.
- Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of hydrology*, 480, 33-45.
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica d: Nonlinear phenomena*, 404, 132306.
- Shi, Y. (2022). Forecasting mortality rates with the penalized exponential smoothing state space model. *Journal of the Operational Research Society*, 73(5), 955-968.
- Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. In *Time series analysis and its applications: with R examples* (pp. 75-163). Springer.
- Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1), 75-85.
- Suhadi, S., Last, C., & Fingscheidt, T. (2010). A data-driven approach to a priori SNR estimation. *IEEE transactions on audio, speech, and language processing*, 19(1), 186-195.

- Taieb, S. B., & Atiya, A. F. (2015). A bias and variance analysis for multistep-ahead time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1), 62-76.
- Tan, L., Liu, S., Gao, J., Liu, X., Chu, L., & Jiang, H. (2024). Enhanced self-checkout system for retail based on improved YOLOv10. *Journal of Imaging*, 10(10), 248.
- Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4), 715-725.
- Ulama, B. S. S., Putra, R. A., Hibatullah, F., Habibi, M. R., & Nafis, M. A. (2025). Comparison of Artificial Neural Network and Long Short-Term Memory for Modelling Crude Palm Oil Production in Indonesia. *International Journal of Advanced Computer Science & Applications*, 16(1).
- Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications* (Vol. 7). Thomson Brooks/Cole Belmont, CA.
- Waller, D. L. (2003). Operations management: a supply chain approach.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324-342.
- 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.
- Xu, Y., Sushmit, A., Lyu, Q., Li, Y., Cao, X., Maltz, J. S., Wang, G., & Yu, H. (2022). Cardiac CT motion artifact grading via semi-automatic labeling and vessel tracking using synthetic image-augmented training data. *Journal of X-Ray Science and Technology*, 30(3), 433-445.
- Yuan, Y., Huang, Y., Ma, Y., Li, X., Li, Z., Shi, Y., & Zhou, H. (2024). Rhyme-aware Chinese lyric generator based on GPT. Fourth International Conference on Advanced Algorithms and Neural Networks (AANN 2024),
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- Zhao, Y., Gao, H., Cao, J., & Yang, S. (2024). Utilizing large language models for information extraction from real estate transactions. *arXiv preprint arXiv:2404.18043*.
- Zheng, S., Chen, S., Chen, T., Yang, Z., Zhao, Z., & Yang, X. (2024). Deep learning-based SNR estimation. *IEEE Open Journal of the Communications Society*, 5, 4778-4796.