AIMS Mathematics, 10(5): 12761–12807
DOI: 10.3934/math.2025575
Received: 17 February 2025
Revised: 24 April 2025
Accepted: 14 May 2025
Published: 30 May 2025

https://www.aimspress.com/journal/Math

*Research article*

# A new Log-Lomax distribution, properties, stock price, and heart attack predictions using machine learning techniques

**Aliyu Ismail Ishaq[1,*], Abdullahi Ubale Usman[2], Hana N. Alqifari[3], Amani Almohaimeed[3,*], Hanita Daud[4], Sani Isah Abba[5], and Ahmad Abubakar Suleiman[4]**

[1] Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

[2] Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria

[3] Department of Statistics and Operation Research, College of Science, Qassim University, Saudi Arabia

[4] Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia

[5] Department of Civil Engineering, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia

* **Correspondence:** Email: ama.almohaimeed@qu.edu.sa, binishaq05@gmail.com; Tel: +2348060471748.

**Abstract:** In this study, we introduced the log-Lomax distribution, a more versatile probabilistic model for capturing various statistical properties. The study was divided into two sections: Modeling stock price exchange rates with the proposed log-Lomax distribution and incorporating log-Lomax features into machine learning models for prediction. In the modeling section, we introduced the log-Lomax distribution, which employed a logarithmic transformation with an exponent parameter. The model was left- and right-skewed, monotonic, inverted, and bathtub-shaped. Some properties were obtained, and several parameter estimation techniques were evaluated using a simulation study. The model was applied to two Nigerian stock exchange rate datasets: Naira-to-Euro and Naira-to-Riyal, as well as the Worcester heart attack patient dataset. The prediction section used insights from modeling methods and machine learning workflows to improve accuracy and reduce overfitting. The predictions were evaluated in two ways: With raw data and features derived from the log-Lomax model. Employing log-Lomax model features, Random Forest, and XGBoost achieved 99.87% accuracy in the Euro dataset, respectively. Random Forest and XGBoost had accuracy rates of 98.67% and 99.33% on the Riyal dataset, respectively, and 91.25% and 88.75% on the heart attack dataset. Random Forests and XGBoost are the preferred models, as they consistently provide the best prediction performance and stability mix across datasets.

**Keywords:** machine learning; artificial intelligence; statistical model; lomax distribution; stability; log-Lomax distribution; stock market
**Mathematics Subject Classification:** 60E05, 62F10, 62H12

## 1. Introduction

Financial and medical data, such as stock market records and patient health indicators, are complex, volatile, and require accurate predictive modeling. The stock market is an essential component of the global economy, serving as a barometer of economic health and offering opportunities for investment, capital raising, and wealth generation. However, stock prices are impacted by complex influences such as economic indicators, investor sentiment, global events, and company performance [1]. These dynamic and often volatile market conditions make accurate prediction of stock price trends a challenging yet vital task. As stock markets are characterized by irregular fluctuations, unpredictability, and sometimes extreme shifts, effective modeling techniques are essential to capture these complexities [2]. Similarly, in the medical field, particularly regarding heart attack prediction, patient health data frequently exhibits high dimensionality and skewness, necessitating reliable techniques to handle uncertainty and discover patterns [3].

In recent years, machine learning techniques have become highly valuable tools for analyzing and predicting complex patterns in financial and medical data. These models may handle the complexities and non-linear features and identify hidden patterns, which makes them very effective for stock price predictions [4]. Unlike traditional statistical methods, machine learning algorithms do not rely on strict assumptions about the distribution of the data, making them ideal for handling volatile and unpredictable stock and heart attack data. [5]. Algorithms, including support vector machines and decision trees [6], are often utilized for predicting outcomes, identifying market patterns, and making predictions.

Despite their strengths, machine learning models have considerable challenges in making accurate predictions. Stock markets are inherently volatile, with prices frequently showing non-stationary, irregular behavior. Market returns are rarely symmetrically distributed, with heavy tails, skewness, and abrupt fluctuations that challenge modeling efforts [7]. Medical data, such as heart attack datasets, are often characterized by high dimensionality, skewness, and the presence of nonlinear relationships among clinical variables [3]. Traditional machine learning algorithms, particularly those based on basic assumptions like Gaussian distributions, struggle to adapt to the dynamic and unpredictable nature. These challenges contribute to problems such as overfitting, poor generalization, and the difficulty of recording severe events [8].

Probability distributions are essential in data analysis since they provide a fundamental framework for modeling many forms of data encountered in practical fields. When examining real-world data, choosing the best suitable probability distribution is crucial to improve the sensitivity, effectiveness, and efficiency of statistical tests. This selection enhances the precision of statistical investigations and ensures a greater alignment between the data and the selected distribution. Therefore, it is vital to possess an extensive knowledge of the optimal distribution for a particular dataset to provide significant analyses and interpretations.

In recent years, the Lomax distribution, initially introduced in [9], has emerged as a versatile distribution widely applicable across various fields in applied sciences. We encompass the works of [10] for life testing applicability, [11] for queue service discipline applications, [12] for personal wealth applications, and [13] for internet traffic applications, to name many. Bryson [14] describes the Lomax distribution as a polynomial substitute to the Rayleigh, exponential, and Weibull distributions.

Its heavy tails and simple corresponding probability functions are what make it so popular. Khalaf et al. [15] noted that the Lomax distribution is a widely applicable model for analyzing various failure patterns in real-world scenarios, making it an effective tool in reliability and survival analysis. Pokharel

et al. [16] emphasized the importance of selecting a suitable distribution for stock index returns, as it works as a foundation for assessing investment performance. They argued that effective decision-making in the financial realm relies heavily on empirical data analysis.

The log-Lomax distribution broadens the lifetime Lomax distribution and offers a potential solution to various prediction problems by modeling asymmetric, skewed data and variable hazard functions. Its ability to capture extreme values and heavy tails, as opposed to traditional models that presume normality, enables it to handle data's volatility and skewness better, enhancing prediction accuracy during periods of instability.

Integrating the log-Lomax distribution into machine learning algorithms has numerous appealing benefits. First, it increases modeling versatility by accurately simulating skewed and heavy-tailed distributions, which are common in stock price and heart attack data movements with frequent severe variations. Second, the log-Lomax distribution improves risk management by capturing extreme events and tail risks, making it useful for managing abrupt price shifts, which can have serious effects in various scenarios. Furthermore, integrating this distribution into machine learning models enhances predictive accuracy by enabling predictions to better reflect the underlying distributional features of stock prices, hence improving model performance in terms of both accuracy and generalization.

The probability density function (pdf) that defines the Lomax distribution is as follows:

$$g(y;\beta,\theta) = \frac{\beta}{\theta}\left(1+\frac{y}{\theta}\right)^{-(\beta+1)}, \qquad y>0, \tag{1.1}$$

where the shape and scale parameters are represented by $\beta > 0$ and $\theta > 0$, respectively.

Nonetheless, despite its heavy tails, the Lomax distribution has some drawbacks, particularly in terms of flexibility, which may make it inadequate for a thorough analysis. As a result, considerable resources have been employed towards generalizing these distributions in a manner that enhances their applicability. Among the famous of these generalizations, it is worth mentioning the Nadarajah-Haghighi Lomax distribution by [17], the McDonald Lomax distribution by [18], the exponential Lomax distribution by [19], the Weighted Lomax distribution by [20], the Gumbel Lomax distribution by [21], the Maxwell Lomax distribution by [22], the gamma Lomax distribution by [23], and the Odd Lomax-Lomax distribution by [24].

In recent years, the process of developing new distributions has piqued the interest of many researchers and made significant contributions to the study of lifetimes, which is critical in a variety of scientific and technological domains, including finance, biology, environmental sciences, economics, physics, and hydrology. A variety of distributions have been introduced for this purpose, each adopting a different methodology. For example, the Maxwell-Weibull distribution was introduced by [25], the modified Kies Frechet distribution defined by [26], the features and applications of the new unit distribution by [27], the novel alpha power Frechet distribution by [28], the novel alpha power Gumbel distribution by [29], the F-Weibull distribution by [30], the Exponentiated Weibull Burr Type X distribution by [31], a new Inverse Rayleigh distribution by [32], the odd beta prime Burr X distribution by [33], the Kavya-Manoharan Kumaraswamy distribution by [34], the Kumaraswamy-generated family by [35], the odd beta prime logistic distribution by [36], the generalized Topp-Leone-G power series class by [37], the Maxwell-Kumaraswamy distribution by [38], the OBP-Weibull distribution by [39], the Topp-Leone Burr Type X distribution by [40], the OBP-Gumbel distribution by [41], the Maxwell-Log Logistic distribution by [42], among others.

To our knowledge, this is the first study in the literature to investigate the log-Lomax distribution. Our investigation aims to explore whether the log-Lomax distribution is suitable for showcasing real-

world data. In this study, we present the log-Lomax distribution as a unique probabilistic model for stock price exchange rate and heart attack modeling and prediction. This distribution is derived through a logarithmic transformation applied to the classical Lomax distribution, providing enhanced flexibility and adaptability. The log-Lomax distribution is capable of capturing a wider range of statistical behaviors, including skewness, heavy tails, and nonlinear patterns, which are prevalent in financial and medical data. By utilizing this transformation technique, the log-Lomax distribution overcomes the limitations of traditional models and proves to be a more effective tool for modeling complex data characteristics. This study makes several important contributions. First, while the classical Lomax distribution is limited to modeling only certain types of data behaviors, the log-Lomax distribution extends its flexibility by accommodating both left- and right-skewed data, as well as various kinds of hazard functions, including monotonic, bathtub-shaped, as well as inverted bathtub-shaped curves. This versatility makes the log-Lomax distribution highly suitable for modeling diverse datasets, including stock price exchange rates and heart attack. Second, applying the log-Lomax distribution to real-world data shows that it performs better in capturing complex data patterns than other Lomax-based models. Third, the researchers use the log-Lomax distribution in machine learning workflows to improve prediction accuracy. Machine learning algorithms like Random Forest and XGBoost, using features from the log-Lomax model, perform nearly perfectly in prediction results, surpassing traditional models that use raw data.

We introduce the log-Lomax distribution, a novel statistical model developed by applying a logarithmic process similar to how the log-normal distribution is derived. We are the first to explore this new distribution, aiming to determine its usefulness for representing real-world data. We present the log-Lomax as a useful tool for analyzing and forecasting stock price exchange rates and heart attack data. By transforming the basic Lomax distribution, we develop a more adaptable model that can capture a broader range of statistical behaviors found in financial and medical data, such as skewness, heavy tails, and complex patterns. This transformation enables the log-Lomax to overcome limitations of conventional models, making it more effective for analyzing complicated data properties.

This study makes significant contributions by presenting the log-Lomax distribution as a versatile model capable of fitting both left- and right-leaning data, along with various hazard function types, such as increasing, decreasing, U-shaped, and inverted U-shaped patterns, which the conventional Lomax distribution cannot accommodate. Applying the log-Lomax model to real exchange rate and medical data demonstrates its superior ability to represent complex data patterns. Furthermore, incorporating features derived from the log-Lomax model into machine learning techniques like Random Forest and XGBoost greatly improves the accuracy of stock price trend predictions, outperforming models relying on raw data. This research establishes the log-Lomax distribution as a valuable tool for statistical modeling and enhancing predictive capabilities in financial and medical data analysis.

The remaining sections of this work are organized as follows. In Section 2, we explain the log-Lomax distribution, as well as density and hazard functions. In Section 3, we delve into the log-Lomax distribution's properties, such as quantile function, moments, order statistics, and moments of order statistics. In Section 4, we discuss parameter estimates and a simulation study using various approaches. In Section 5, we present simulation results for MLE, LS, and WLSE behaviors. A real data application is conducted in Section 6. In Section 7, we investigate the use of machine learning to predict stock market exchange and heart attack data. In Section 8, we outline the study's findings and limitations. Finally, Section 9 has concluding remarks.

## 2. Materials and methods

In this section, we describe the derivation of the log-Lomax (LLomax) distribution via a log transformation with an exponent parameter, using the following relationships:

$$x^{\alpha} = \log(y), \qquad y > 0;\ \alpha > 0, \tag{2.1}$$

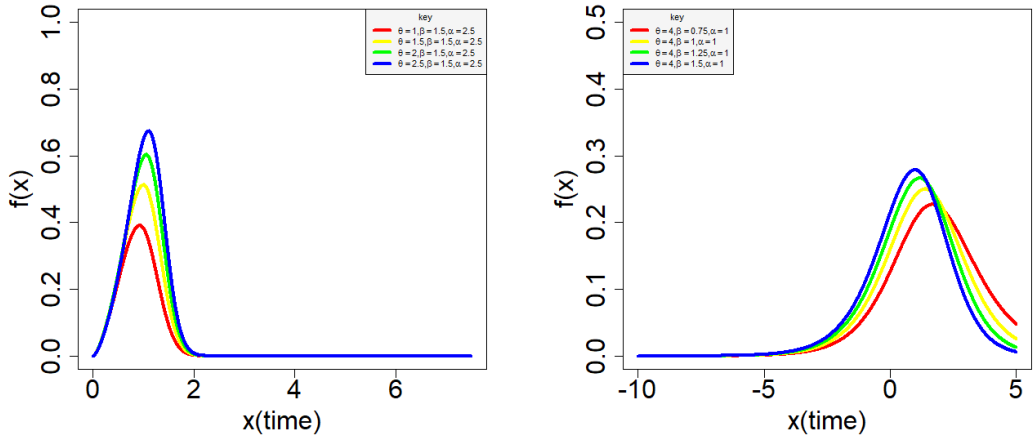
and

$$f(x;\beta,\theta) = f(y;\beta,\theta)\left|\frac{dy}{dx}\right|, \tag{2.2}$$

where $\alpha > 0$ is an exponent parameter, $f(y;\beta,\theta)$ is defined in Eq (1.1), $y = e^{x^{\alpha}}$, and $\frac{dy}{dx} = \alpha x^{\alpha-1}e^{x^{\alpha}}$ is derivative of Eq (2.1). Thus, the proposed LLomax distribution can be obtained from the relations presented in Eqs (1.1), (2.1), and (2.2) as:

$$\begin{aligned} f(x;\beta,\theta,\alpha) &= \frac{\beta}{\theta}\left(1+\frac{e^{x^{\alpha}}}{\theta}\right)^{-(\beta+1)} \times \alpha x^{\alpha-1}e^{x^{\alpha}} \\ &= \frac{\alpha\beta}{\theta}x^{\alpha-1}e^{x^{\alpha}}\left(1+\frac{e^{x^{\alpha}}}{\theta}\right)^{-(\beta+1)}, \qquad x \in \Re. \end{aligned} \tag{2.3}$$

Thus, a random variable $X$ with a pdf specified in Eq(2.3) can be expressed as $X \sim LLo\max(\beta,\theta,\alpha)$. The pdf explained in Eq (2.3) can be represented as $X \sim LLo\max(\beta,\theta,\alpha)$. For convenience, the dependence on the vector of parameters in Eq (2.3) can be dropped, writing $f(x;\beta,\theta,\alpha) = f(x)$. Figure 1 shows the pdf forms of the LLomax distribution, which can be either (a) right- or (b) left-skewed and used to mimic skewed datasets.




**(a)** **(b)**

**Figure 1.** Log-Lomax density plots for various parameter values.

### 2.1. Model validity check

A critical feature of a legitimate statistical distribution is that its pdf integrates to one (1) over the domain. This property ensures that the total probability under the pdf equals one, signifying that a random variable described by this distribution is certain to fall within its defined range. If a proposed pdf

satisfies Eq (2.4), it can be considered a legitimate statistical distribution. Upon satisfying Eq (2.4), we can conclude that the pdf in Eq (2.3) has a legitimate statistical distribution. Mathematically, this can be explained according to:

$$\int_{-\infty}^{\infty} f(x;\beta,\theta,\alpha)dx = 1. \tag{2.4}$$

To demonstrate this, the pdf from Eq (2.3) might represented as:

$$\int_{-\infty}^{\infty} f(x;\beta,\theta,\alpha)dx = \frac{\alpha\beta}{\theta}\int_{-\infty}^{\infty} x^{\alpha-1}e^{x^{\alpha}}\left(1+\frac{e^{x^{\alpha}}}{\theta}\right)^{-(\beta+1)} dx. \tag{2.5}$$

Letting,

$$m = \frac{e^{x^{\alpha}}}{\theta}, \qquad dx = \frac{\theta}{\alpha x^{\alpha-1}e^{x^{\alpha}}}dm. \tag{2.6}$$

Substituting Eq (2.6) to Eq (2.5), we can get

$$\int_{-\infty}^{\infty} f(x;\beta,\theta,\alpha)dx = \beta\int_{0}^{\infty}(1+m)^{-(\beta+1)}\,dm. \tag{2.7}$$

Assume we let

$$\frac{1}{A} = (1+m), \qquad dm = -(1+m)^2\,dA. \tag{2.8}$$

Finally, Eq (2.7) might be described according to:

$$\int_{-\infty}^{\infty} f(x;\beta,\theta,\alpha)dx = \beta\int_{0}^{1} A^{\beta-1}dA = 1. \tag{2.9}$$

This proved Eq (2.4), confirming that the pdf in Eq (2.3) represents a legitimate statistical distribution.

### *2.2. Distribution function*

The distribution function (cdf) of the LLomax distribution can be derived from Eq (2.3) by the relations:

$$F(x;\beta,\theta,\alpha) = \frac{\alpha\beta}{\theta}\int_{-\infty}^{x} t^{\alpha-1}e^{t^{\alpha}}\left(1+\frac{e^{t^{\alpha}}}{\theta}\right)^{-(\beta+1)} dt. \tag{2.10}$$

Applying Eq (2.6) to Eq (2.10) gives

$$F(x;\beta,\theta,\alpha) = \beta\int_{0}^{\frac{e^{x^{\alpha}}}{\theta}}(1+m)^{-(\beta+1)}\,dm. \tag{2.11}$$

Considering Eq (2.8), then the cdf in Eq (2.11) is expressed as:

$$F(x;\beta,\theta,\alpha) = \beta\int_{\frac{1}{1+\frac{e^{x^{\alpha}}}{\theta}}}^{1} A^{\beta-1}dA. \tag{2.12}$$

Solving for Eq (2.12) yields the cdf for the LLomax distribution as:

$$F(x;\beta,\theta,\alpha) = F(x) = 1 - \left(1 + \frac{e^{x^{\alpha}}}{\theta}\right)^{-\beta}, \quad x \in \Re. \tag{2.13}$$

### 2.3. *Survival and hazard functions*

The survival function (sf) of the LLomax distribution might be computed according to Eq (2.13) as:

$$S(x;\beta,\theta,\alpha) = \left(1 + \frac{e^{x^{\alpha}}}{\theta}\right)^{-\beta}. \quad x \in \Re \tag{2.14}$$

The hazard function (hf) can be derived from Eqs (2.3) and (2.14) as:

$$h(x;\beta,\theta,\alpha) = \frac{\alpha\beta}{\theta} x^{\alpha-1} e^{x^{\alpha}} \left(1 + \frac{e^{x^{\alpha}}}{\theta}\right)^{-1}. \quad x \in \Re \tag{2.15}$$

Figure 2(a–d) illustrates the diverse hazard function shapes achievable with the LLomax distribution. These shapes include (a) increasing, (b) upside-down bathtub, (c) decreasing, and (d) bathtub functions.

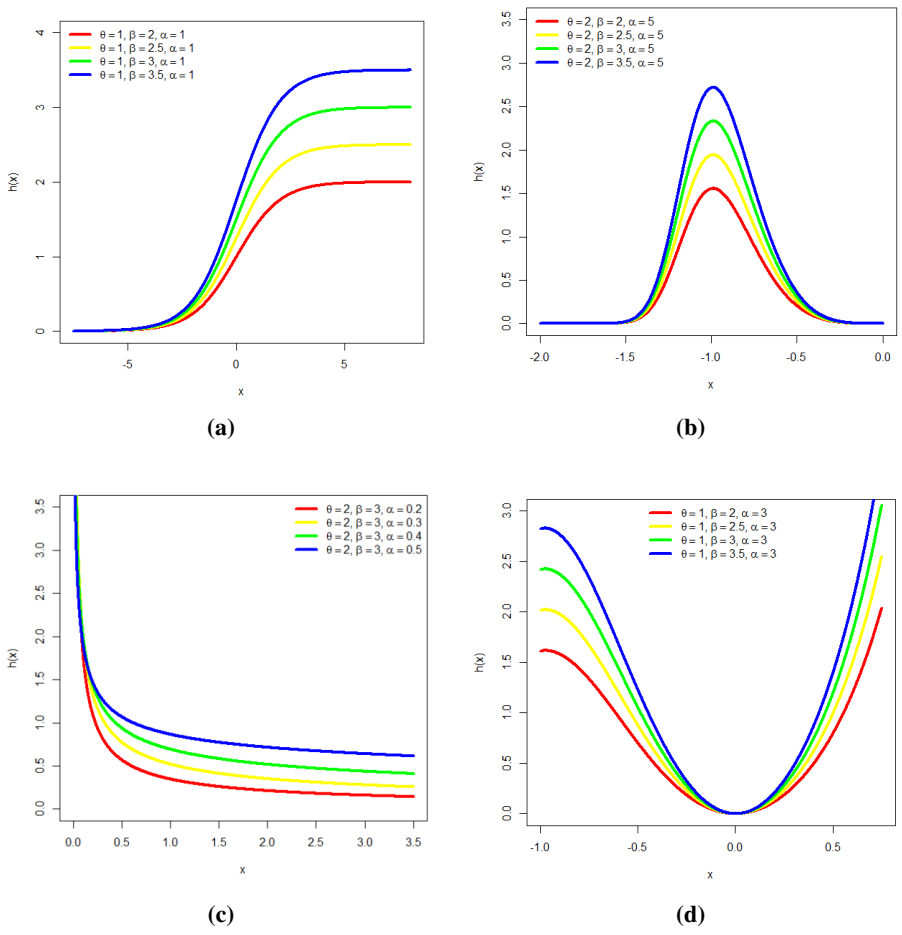



**Figure 2.** Log-Lomax hazard plots for parameter values.

### 2.4. Mixture representations for the pdf of the LLomax Distribution

The proposed distribution's pdf can be given as a mixture representation investigated in [43] by the following expressions:

Consider the expression from the binomial expansion for real-valued $\varepsilon$ as:

$$(1+z)^{-\varepsilon} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(\varepsilon + j)}{j!\Gamma(\varepsilon)} z^j. \tag{2.16}$$

Applying Eq (2.16) into Eq (2.3), we can obtain

$$\begin{aligned} f(x) &= \frac{\alpha\beta}{\theta} x^{\alpha-1} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(\beta + j + 1)}{\theta^j j!\Gamma(\beta + j)} e^{x^{\alpha}(1+j)} \\ &= \alpha\beta x^{\alpha-1} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma\left(\beta - k_j\right)}{\theta^{-k_j} j!\Gamma(\beta + j)} e^{-x^{\alpha}k_j} \\ &= x^{\alpha-1} \sum_{j=0}^{\infty} A_j e^{-x^{\alpha}k_j}, \end{aligned} \tag{2.17}$$

where

$$k_j = -(1+j)$$

and

$$A_j = \alpha\beta \frac{(-1)^j \theta^{k_j} \Gamma\left(\beta - k_j\right)}{j!\Gamma(\beta + j)}.$$

## 3. Properties of the LLomax distribution

In the following section, we explore some properties of the LLomax distribution, which involves the moments, quantile function, order statistics, and moments of order statistics.

### 3.1. Quantile function

The quantile function (qf), also known as the inverse cdf, percentile function, or percent-point function, is an essential instrument in probability and statistics. The qf is the inverse of the cdf, but only if it is continuous and strictly increasing. Quantile functions are useful for a variety of applications, including:

i. Hypothesis testing involves identifying the crucial values for statistical tests.
ii. Monte Carlo simulations are used to generate random variables based on a given probability distribution.

The qf of the LLomax distribution is obtained by inverting Eq (2.13) as:

$$u = 1 - \left(1 + \frac{e^{x^{\alpha}}}{\theta}\right)^{-\beta}, \tag{3.1}$$

where $u$ represents a uniform random variable over the (0, 1) interval. Consequently, Eq (3.1) can be reduced to:

$$\left(1+\frac{e^{x^{\alpha}}}{\theta}\right)^{-\beta}=1-u. \tag{3.2}$$

This implies that

$$\frac{e^{x^{\alpha}}}{\theta}=(1-u)^{-\frac{1}{\beta}}-1. \tag{3.3}$$

Simplifying for Eq (3.3) resulted in:

$$e^{x^{\alpha}}=\theta\left((1-u)^{-\frac{1}{\beta}}-1\right). \tag{3.4}$$

Taking the logarithm to both sides of Eq (3.4), we received

$$x^{\alpha}=\log\left(\theta\left((1-u)^{-\frac{1}{\beta}}-1\right)\right). \tag{3.5}$$

It follows that the qf is derived from Eq (3.5) as:

$$x=\left\{\log\left(\theta\varpi\right)\right\}^{\frac{1}{\alpha}}, \tag{3.6}$$

where

$$\varpi=(1-u)^{-\frac{1}{\beta}}-1,$$

which is the qf of the LLomax distribution.

### *3.2. Moments*

Suppose $X\sim LLomax\left(\beta,\theta,\alpha\right)$, then the moments of the LLomax distribution is derived employing

$$\begin{aligned}\mu_{r}^{'}&=\int_{-\infty}^{\infty}x^{r}f\left(x;\beta,\theta,\alpha\right)dx\\&=\sum_{j=0}^{\infty}A_{j}\int_{-\infty}^{\infty}x^{\alpha+r-1}e^{-x^{\alpha}k_{j}}dx.\end{aligned} \tag{3.7}$$

Let us define,

$$w=x^{\alpha}k_{j},\quad\Rightarrow\quad dx=\frac{dm}{\alpha x^{\alpha-1}k_{j}}. \tag{3.8}$$

Inserting Eq (3.8) into Eq (3.7) yields

$$\begin{aligned}\mu_{r}^{'}&=\frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_{j}}{k_{j}}\int_{0}^{\infty}x^{r}e^{-w}dw\\&=\frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_{j}}{k_{j}^{1+\frac{r}{\alpha}}}\int_{0}^{\infty}w^{\frac{r}{\alpha}}e^{-w}dw\\&=\frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_{j}}{k_{j}^{1+\frac{r}{\alpha}}}\Gamma\left(1+\frac{r}{\alpha}\right),\end{aligned} \tag{3.9}$$

which is the moments of the LLomax distribution. The first, second, third, and fourth moments are obtained by setting $r = 1, 2, 4$ and 4 into Eq (3.9) as:

$$\mu_1^{'} = \mu = \frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{1}{\alpha}}}\Gamma\left(1+\frac{1}{\alpha}\right), \tag{3.10}$$

$$\mu_2^{'} = \frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{2}{\alpha}}}\Gamma\left(1+\frac{2}{\alpha}\right), \tag{3.11}$$

$$\mu_3^{'} = \frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{3}{\alpha}}}\Gamma\left(1+\frac{3}{\alpha}\right), \tag{3.12}$$

and

$$\mu_4^{'} = \frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{4}{\alpha}}}\Gamma\left(1+\frac{4}{\alpha}\right). \tag{3.13}$$

Therefore, the mean is provided in Eq (3.10), the variance, skewness, and kurtosis are presented as follows:

$$\begin{aligned}\text{variance} &= \mu_2^{'} - \mu_2 \\ &= \frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{2}{\alpha}}}\Gamma\left(1+\frac{2}{\alpha}\right) - \left(\frac{2}{\alpha}\sum_{j=0}^{\infty}\frac{A_j}{k_j^{1+\frac{1}{\alpha}}}\Gamma\left(1+\frac{1}{\alpha}\right)\right)^2,\end{aligned} \tag{3.14}$$

$$\text{skewness} = \frac{\mu_3^{'} - 3\mu\mu_2^{'} + 2\mu^3}{\sigma^3}, \tag{3.15}$$

and

$$\text{kurtosis} = \frac{\mu_4^{'} - 4\mu\mu_3^{'} + 6\mu^2\mu_2^{'} - 3\mu^4}{\sigma^4}, \tag{3.16}$$

where

$$\sigma = \sqrt{\text{variance}}.$$

Tables 1 and 2 present the mean, variance, skewness, and kurtosis for the LLomax distribution, providing insights into its distributional characteristics. Table 1 highlights the distribution's response to changes in the exponent parameter $\alpha$, focusing on fixed values of $\theta$ and $\beta$. The results show that as $\alpha$ increases, the mean of the distribution rises, signaling an increase in central tendency. Furthermore, there are moderate changes in variance, indicating that $\alpha$ influences the spread of the distribution as well. For example, when $\alpha$ shifts from 1 to 1.6 with $\theta = 0.02$ and $\beta = 0.1$, the mean increases from 2.30532 to 2.60325, while the variance remains relatively stable. Skewness and kurtosis also change significantly with $\alpha$, with skewness decreasing and kurtosis becoming less pronounced as $\alpha$ increases.

Table 2 shows the influence of the parameter $\theta$ on the LLomax distribution's moments. Variations in $\theta$ produce substantial changes in mean and variance, though the effect on variance appears more moderate compared to other parameters. For example, when $\theta = 0.15$, the mean increases from 0.83243 to 0.99275 as $\alpha$ increases from 0.1 to 0.4, while variance remains relatively stable. Skewness and kurtosis shift considerably as $\theta$ and $\alpha$ change, with the distribution capable of modeling both symmetric and highly skewed scenarios, as well as light-tailed and heavy-tailed data.

**Table 1.** Mean, Variance, Skewness, and Kurtosis of the LLomax distribution.

| $\theta$ | $\beta$ | $\alpha$ | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| 0.02 | 0.1 | 1 | 2.30532 | 2.04627 | 0.16747 | -1.15897 |
| 0.02 | 0.1 | 1.2 | 2.45960 | 1.92381 | 0.07083 | -1.13514 |
| 0.02 | 0.1 | 1.4 | 2.55844 | 1.79377 | 0.02075 | -1.08748 |
| 0.02 | 0.1 | 1.6 | 2.60325 | 1.66179 | 0.01395 | -1.02860 |
| 0.02 | 0.2 | 1 | 2.10458 | 1.97400 | 0.33800 | -1.03883 |
| 0.02 | 0.2 | 1.2 | 2.19482 | 1.83660 | 0.29149 | -1.00563 |
| 0.02 | 0.2 | 1.4 | 2.21680 | 1.67063 | 0.30035 | -0.91522 |
| 0.02 | 0.2 | 1.6 | 2.17553 | 1.47325 | 0.35309 | -0.75954 |
| 0.02 | 0.3 | 1 | 1.91315 | 1.86183 | 0.50607 | -0.83943 |
| 0.02 | 0.3 | 1.2 | 1.94980 | 1.68706 | 0.50001 | -0.74920 |
| 0.02 | 0.3 | 1.4 | 1.91754 | 1.46439 | 0.54464 | -0.55781 |
| 0.02 | 0.3 | 1.6 | 1.83355 | 1.20394 | 0.60797 | -0.27921 |
| 0.04 | 0.1 | 1 | 2.31216 | 2.04313 | 0.16236 | -1.15909 |
| 0.04 | 0.1 | 1.2 | 2.46575 | 1.91953 | 0.06663 | -1.13364 |
| 0.04 | 0.1 | 1.4 | 2.56382 | 1.78911 | 0.01751 | -1.08531 |
| 0.04 | 0.1 | 1.6 | 2.60790 | 1.65726 | 0.01161 | -1.02641 |
| 0.04 | 0.2 | 1 | 2.11229 | 1.97236 | 0.33192 | -1.04188 |
| 0.04 | 0.2 | 1.2 | 2.20195 | 1.83392 | 0.28626 | -1.00712 |
| 0.04 | 0.2 | 1.4 | 2.22323 | 1.66765 | 0.29604 | -0.91600 |
| 0.04 | 0.2 | 1.6 | 2.18126 | 1.47046 | 0.34965 | -0.76002 |
| 0.04 | 0.3 | 1 | 1.92158 | 1.86194 | 0.49897 | -0.84632 |
| 0.04 | 0.3 | 1.2 | 1.95767 | 1.68628 | 0.49371 | -0.75478 |
| 0.04 | 0.3 | 1.4 | 1.92466 | 1.46334 | 0.53922 | -0.56272 |
| 0.04 | 0.3 | 1.6 | 1.83989 | 1.20291 | 0.60342 | -0.28340 |
| 0.06 | 0.1 | 1 | 2.31888 | 2.04003 | 0.15735 | -1.15915 |
| 0.06 | 0.1 | 1.2 | 2.47180 | 1.91530 | 0.06250 | -1.13211 |
| 0.06 | 0.1 | 1.4 | 2.56909 | 1.78451 | 0.01433 | -1.08314 |
| 0.06 | 0.1 | 1.6 | 2.61246 | 1.65282 | 0.00932 | -1.02424 |
| 0.06 | 0.2 | 1 | 2.11986 | 1.97072 | 0.32596 | -1.04478 |
| 0.06 | 0.2 | 1.2 | 2.20895 | 1.83126 | 0.28114 | -1.00851 |
| 0.06 | 0.2 | 1.4 | 2.22953 | 1.66470 | 0.29181 | -0.91670 |
| 0.06 | 0.2 | 1.6 | 2.18689 | 1.46771 | 0.34628 | -0.76044 |
| 0.06 | 0.3 | 1 | 1.92987 | 1.86199 | 0.49201 | -0.85295 |
| 0.06 | 0.3 | 1.2 | 1.96540 | 1.68548 | 0.48754 | -0.76015 |
| 0.06 | 0.3 | 1.4 | 1.93166 | 1.46227 | 0.53392 | -0.56745 |
| 0.06 | 0.3 | 1.6 | 1.84610 | 1.20188 | 0.59898 | -0.28744 |

**Table 2.** Mean, Variance, Skewness and Kurtosis of the LLomax distribution.

| $\theta$ | $\beta$ | $\alpha$ | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| 0.05 | 1 | 0.1 | 0.82091 | 1.38376 | 1.77698 | 2.29457 |
| 0.05 | 1 | 0.2 | 0.86098 | 1.41075 | 1.70545 | 2.03482 |
| 0.05 | 1 | 0.3 | 0.90854 | 1.42762 | 1.62739 | 1.77764 |
| 0.05 | 1 | 0.4 | 0.95431 | 1.42179 | 1.55922 | 1.58488 |
| 0.05 | 2 | 0.1 | 0.67594 | 1.15222 | 2.11514 | 3.84929 |
| 0.05 | 2 | 0.2 | 0.60605 | 0.98936 | 2.33028 | 5.07374 |
| 0.05 | 2 | 0.3 | 0.56669 | 0.85687 | 2.47734 | 6.09744 |
| 0.05 | 2 | 0.4 | 0.54334 | 0.73878 | 2.57378 | 6.96770 |
| 0.05 | 3 | 0.1 | 0.54950 | 0.93208 | 2.50274 | 5.98524 |
| 0.05 | 3 | 0.2 | 0.41659 | 0.64030 | 3.10226 | 10.25750 |
| 0.05 | 3 | 0.3 | 0.34855 | 0.45524 | 3.54222 | 14.43377 |
| 0.05 | 3 | 0.4 | 0.31520 | 0.33460 | 3.77417 | 17.61270 |
| 0.1 | 1 | 0.1 | 0.82677 | 1.39226 | 1.76508 | 2.24541 |
| 0.1 | 1 | 0.2 | 0.87224 | 1.42645 | 1.68388 | 1.94922 |
| 0.1 | 1 | 0.3 | 0.92442 | 1.44861 | 1.59886 | 1.66900 |
| 0.1 | 1 | 0.4 | 0.97385 | 1.44609 | 1.52604 | 1.46277 |
| 0.1 | 2 | 0.1 | 0.68380 | 1.16488 | 2.09456 | 3.74720 |
| 0.1 | 2 | 0.2 | 0.61950 | 1.01153 | 2.28931 | 4.84666 |
| 0.1 | 2 | 0.3 | 0.58374 | 0.88447 | 2.42008 | 5.75235 |
| 0.1 | 2 | 0.4 | 0.56237 | 0.76811 | 2.50593 | 6.52873 |
| 0.1 | 3 | 0.1 | 0.55860 | 0.94787 | 2.47104 | 5.79746 |
| 0.1 | 3 | 0.2 | 0.42981 | 0.66365 | 3.03318 | 9.73485 |
| 0.1 | 3 | 0.3 | 0.36310 | 0.47937 | 3.44262 | 13.53066 |
| 0.1 | 3 | 0.4 | 0.32980 | 0.35608 | 3.66019 | 16.43226 |
| 0.15 | 1 | 0.1 | 0.83243 | 1.40044 | 1.75368 | 2.19867 |
| 0.15 | 1 | 0.2 | 0.88312 | 1.44151 | 1.66338 | 1.86884 |
| 0.15 | 1 | 0.3 | 0.93977 | 1.46871 | 1.57184 | 1.56790 |
| 0.15 | 1 | 0.4 | 0.99275 | 1.46932 | 1.49468 | 1.34983 |
| 0.15 | 2 | 0.1 | 0.69141 | 1.17709 | 2.07492 | 3.65076 |
| 0.15 | 2 | 0.2 | 0.63259 | 1.03299 | 2.25060 | 4.63592 |
| 0.15 | 2 | 0.3 | 0.60040 | 0.91132 | 2.36633 | 5.43604 |
| 0.15 | 2 | 0.4 | 0.58103 | 0.79678 | 2.44246 | 6.12929 |
| 0.15 | 3 | 0.1 | 0.56744 | 0.96315 | 2.44090 | 5.62122 |
| 0.15 | 3 | 0.2 | 0.44280 | 0.68655 | 2.96831 | 9.25541 |
| 0.15 | 3 | 0.3 | 0.37751 | 0.50334 | 3.34979 | 12.71424 |
| 0.15 | 3 | 0.4 | 0.34433 | 0.37763 | 3.55420 | 15.37147 |

### 3.3. Order statistics

The order statistics can be explained as:

$$f_{i,\eta}(x) = \frac{\eta f(x)}{(i-1)!(\eta-i)!} F^{i-1}(x)[1-F(x)]^{\eta-i}, \tag{3.17}$$

where $f(x)$ and $F(x)$ are presented in Eqs (2.3) and (2.13), respectively. Consider the binomial expansion for real-valued $\varepsilon > 0$, then Eq (3.17) may be presented as:

$$\begin{aligned} f_{i,\eta}(x) &= \frac{\eta f(x)}{(i-1)!(\eta-i)!} \sum_{m_1=0}^{\infty} (-1)^{m_1} \binom{\eta-i}{m_1} z^{m_1} F^{i+m_1-1}(x) \\ &= \frac{\eta f(x)}{(i-1)!(\eta-i)!} \sum_{m_1=0}^{\infty} (-1)^{m_1} \binom{\eta-i}{m_1} \left\{1-\left(1+\frac{e^{x^\alpha}}{\theta}\right)^{-\beta}\right\}^{i+m_1-1} \\ &= x^{\alpha-1} \sum_{m_1,m_2,m_3=0}^{\infty} \Delta_{m_1,m_2,m_3} e^{-x^\alpha k_{m_3}}, \end{aligned} \tag{3.18}$$

where

$$\Delta_{m_1,m_2,m_3} = \binom{i+m_1-1}{m_2} \binom{\eta-i}{m_1} \frac{\alpha\beta\eta(-1)^{m_1+m_2+m_3}}{(i-1)!(\eta-i)!} \frac{\Gamma(\beta(1+m_2)+m_3+1)}{\theta^{-k_{m_3}} m_3! \Gamma(\beta(1+m_2)+1)}.$$

and $k_{m_3} = -(1+m_3)$.

### 3.4. Moments of order statistics

To obtain the moments of order statistics, we can employ the pdf described in Eq (3.18) as:

$$\begin{aligned} E\left(X_{i,\eta}^r\right) &= \int_{-\infty}^{\infty} x^r f_{i,\eta}(x) dx \\ &= \sum_{m_1,m_2,m_3=0}^{\infty} \Delta_{m_1,m_2,m_3} \int_0^{\infty} x^{r+\alpha-1} e^{-x^\alpha k_{m_3}} dx \\ &= \frac{2}{\alpha} \sum_{m_1,m_2,m_3=0}^{\infty} \frac{\Delta_{m_1,m_2,m_3}}{k_{m_3}^{1+\frac{r}{\alpha}}} \Gamma\left(1+\frac{r}{\alpha}\right). \end{aligned} \tag{3.19}$$

## 4. Parameter estimation

In this section, we use the maximum likelihood (ML), least squares (LS), and weighted least squares (WLS) methodologies to estimate the unknown parameters of the proposed LLomax distribution. Let the random variables $X_1, \ldots, X_n$ be independently and identically distributed, taken from the sample sizes n with observed values $x_i$. Assume that $x_{(1)}, \ldots, x_{(n)}$ are the ordered observations derived from the LLomax distribution with the cdf $F(x|\theta,\beta,\alpha)$. The LS and WLS for the parameters $\theta$, $\beta$, and $\alpha$ are minimized as follows:

$$\ell_{LS}(\theta,\beta,\alpha) = \sum_{i=1}^{n} \left\{F(x_{(i)}|\theta,\beta,\alpha) - \frac{i}{n+1}\right\}^2, \tag{4.1}$$

and

$$\ell_{WLS}(\theta,\beta,\alpha) = \sum_{i=1}^{n}\left\{\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F(x_{(i)}|\theta,\beta,\alpha) - \frac{i}{n+1}\right]^2\right\}. \tag{4.2}$$

To estimate using ML approach, the likelihood function $\ell$ of Eq (2.3) can be represented as:

$$\ell = \left(\frac{\alpha\beta}{\theta}\right)^n \prod_{i=1}^{n} x_i^{\alpha-1} e^{x_i^{\alpha}} \left(1 + \frac{e^{x_i^{\alpha}}}{\theta}\right)^{-(\beta+1)}. \tag{4.3}$$

The log-likelihood function of Eq (4.3) is computed as:

$$L = n\log(\alpha) + n\log(\beta) - n\log(\theta) + (\alpha-1)\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}(x_i^{\alpha}) - (\beta+1)\sum_{i=1}^{n}\log\left(1 + \frac{e^{x_i^{\alpha}}}{\theta}\right). \tag{4.4}$$

To compute the log-likelihood function of Eq (4.3), follow these steps:

$$\frac{\partial L}{\partial\theta} = -\frac{n}{\theta} + (\beta+1)\sum_{i=1}^{n}\left(\frac{e^{x_i^{\alpha}}}{\theta\left(\theta + e^{x_i^{\alpha}}\right)}\right), \tag{4.5}$$

$$\frac{\partial L}{\partial\beta} = \frac{n}{\beta} - \sum_{i=1}^{n}\log\left(\frac{\theta + e^{x_i^{\alpha}}}{\theta}\right) \tag{4.6}$$

and

$$\frac{\partial L}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}(x_i^{\alpha})\log(x_i) - (\beta+1)\sum_{i=1}^{n}\left(\frac{x_i^{\alpha}e^{x_i^{\alpha}}\log(x_i)}{\theta + e^{x_i^{\alpha}}}\right). \tag{4.7}$$

Equating (4.5)–(4.7) gives an estimate for the LLomax distribution.

We utilized R's built-in functions in the parameter estimation and integration processes, such as optim and integrate. These functions entail iterative algorithms that can be conceptualized as series approximations. We rely on the default convergence criteria implemented within these R functions to ensure convergence and numerical stability. For example, the optim function, used for maximum likelihood estimation, terminates when successive iterations produce parameter estimates within a predefined tolerance. Similarly, the integrate function, used for numerical integration, employs adaptive quadrature methods that refine the integral estimate until a specified accuracy is achieved.

The default convergence criteria in these R functions are designed to provide reliable results for a wide range of applications. While a comprehensive analysis of the specific convergence properties of these algorithms for the log-Lomax distribution is beyond the scope of this paper, we believe that the established methods and default tolerances used in R provide sufficient assurance of the validity of our numerical results. Future work could explore a more detailed analysis of the convergence properties of the log-Lomax distribution under various parameter settings and stricter tolerance levels.

## 5. Simulation

In the following section, we address a simulation study that evaluates the efficacy of the LLomax distribution parameters using the LSE, WLSE, and MLE techniques. A simulation study is conducted using the quantile function specified in Eq (3.6) for different sample sizes $n = 5$, 10, 20, 30, and 50, and the parameter values of $\beta = 1$, $\theta = 1$, and $\alpha = 1$. This simulation is run 1,000 times, yielding the

means of the estimations (estimates) as well as the mean squared errors (MSE). Tables 3–5 show the simulation results for the proposed LLomax distribution using three estimation methods: LS, WLS, and MLE. Table 3 shows simulation results for $\beta = 1$, $\theta = 1$, and $\alpha = 1$.

Table 3 shows the estimated parameters $\theta$, $\beta$, and $\alpha$, together with their corresponding MSEs, using LSE, WLSE, and MLE approaches. This table shows that for $n = 5$, the parameters $\theta$, $\beta$, and $\alpha$ have the lowest MSE values compared to other estimation methods. Similarly, estimating the parameter $\alpha$ using the MLE approach yielded the lowest MSE value. However, as the sample size increased, the estimates of each parameter approached true parameter values, and the MSE of each estimate fell, eventually approaching zero. Hence, we can conclude that the estimations using MLE give the least MLE compared to other methods of analysis, such as LSE and WLSE approaches.

Table 4 displays the simulation findings of the LLomax distribution when $\theta = \alpha = 1$ and $\beta = 1.5$. At $n = 5$, estimating the parameters $\theta$ and $\beta$ employing the WLSE technique provided the least MSE values, unlike the LSE and MLE methods. Using the MLE approach to estimate $\alpha$ resulted in the lowest MSE value. When the sample size becomes larger, the parameter estimates shrink, and the estimates of the parameters decrease. The MSE of the parameters $\beta$ and $\alpha$ using LSE and WLSE techniques gave the least MSE and hence offered the best estimate.

Table 5 displays the estimated parameters $\theta$, $\beta$, and $\alpha$, together with their corresponding MSE values. This table reveals that as the sample size grows, the estimates of the parameters $\theta$, $\beta$, and $\alpha$ using the three estimation methods decrease and converge to zero. The parameter $\theta$ can be estimated more accurately and efficiently use the MLE technique. Similarly, estimating parameter $\beta$ using the WLS approach and parameter $\alpha$ using three estimation methods yield the best results. This revealed that both estimation methods performed well, although the WLSE and MLE approaches are more appropriate for estimating the parameters of the LLomax distribution.

**Table 3.** The estimates and MSE using LS, WLS, and MLE when $\beta = 1$.

| sample size (n) | parameter | LSE | | WLSE | | MLE | |
|---|---|---|---|---|---|---|---|
| | | estimate | MSE | estimate | MSE | estimate | MSE |
| 5 | $\theta = 1$ | 1.0962 | 1.4489 | 1.0569 | 0.0240 | 1.1544 | 1.2111 |
| | $\beta = 1$ | 1.0543 | 0.5256 | 1.0318 | 0.0148 | 1.0734 | 0.2151 |
| | $\alpha = 1$ | 1.0378 | 0.1029 | 1.0170 | 0.0282 | 1.0134 | 0.0140 |
| 10 | $\theta = 1$ | 1.0466 | 0.0056 | 1.0455 | 0.0053 | 1.0503 | 0.0281 |
| | $\beta = 1$ | 1.0435 | 0.0053 | 1.0424 | 0.0052 | 1.0467 | 0.0045 |
| | $\alpha = 1$ | 1.0030 | 0.0035 | 1.0022 | 0.0018 | 1.0011 | 0.0005 |
| 20 | $\theta = 1$ | 1.0438 | 0.0042 | 1.0434 | 0.0041 | 1.0419 | 0.0039 |
| | $\beta = 1$ | 1.0415 | 0.0039 | 1.0418 | 0.0039 | 1.0448 | 0.0041 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 30 | $\theta = 1$ | 1.0422 | 0.0040 | 1.0419 | 0.0040 | 1.0406 | 0.0037 |
| | $\beta = 1$ | 1.0427 | 0.0039 | 1.0426 | 0.0039 | 1.0437 | 0.0040 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 50 | $\theta = 1$ | 1.0410 | 0.0038 | 1.0409 | 0.0038 | 1.0399 | 0.0036 |
| | $\beta = 1$ | 1.0398 | 0.0036 | 1.0400 | 0.0036 | 1.0410 | 0.0036 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

**Table 4.** The estimates and MSE using LS, WLS, and MLE when $\beta = 1.5$.

| sample size (n) | parameter | LSE | | WLSE | | MLE | |
|---|---|---|---|---|---|---|---|
| | | estimate | MSE | estimate | MSE | estimate | MSE |
| 5 | $\theta = 1$ | 1.0708 | 0.0132 | 1.0708 | 0.0129 | 1.1288 | 1.1305 |
| | $\beta = 1.5$ | 1.5671 | 0.0146 | 1.5620 | 0.0125 | 1.5910 | 0.1321 |
| | $\alpha = 1$ | 1.0049 | 0.0066 | 1.0012 | 0.0009 | 0.9993 | 0.0002 |
| 10 | $\theta = 1$ | 1.0642 | 0.0091 | 1.0636 | 0.0091 | 1.0624 | 0.0087 |
| | $\beta = 1.5$ | 1.5679 | 0.0097 | 1.5662 | 0.0095 | 1.5719 | 0.0101 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 20 | $\theta = 1$ | 1.0609 | 0.0084 | 1.0605 | 0.0084 | 1.0603 | 0.0082 |
| | $\beta = 1.5$ | 1.5622 | 0.0087 | 1.5627 | 0.0088 | 1.5689 | 0.0094 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 30 | $\theta = 1$ | 1.0584 | 0.0081 | 1.0579 | 0.0080 | 1.0573 | 0.0076 |
| | $\beta = 1.5$ | 1.5640 | 0.0088 | 1.5641 | 0.0088 | 1.5670 | 0.0090 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 50 | $\theta = 1$ | 1.0540 | 0.0071 | 1.0540 | 0.0071 | 1.0541 | 0.0069 |
| | $\beta = 1.5$ | 1.5596 | 0.0080 | 1.5601 | 0.0080 | 1.5629 | 0.0082 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

**Table 5.** The estimates and MSE using LS, WLS, and MLE when $\beta = 2$.

| sample size (n) | parameter | LSE | | WLSE | | MLE | |
|---|---|---|---|---|---|---|---|
| | | estimate | MSE | estimate | MSE | estimate | MSE |
| 5 | $\theta = 1$ | 1.0870 | 0.0166 | 1.0873 | 0.0167 | 1.0841 | 0.0159 |
| | $\beta = 2$ | 2.0961 | 0.0185 | 2.0878 | 0.0170 | 2.1016 | 0.0193 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 10 | $\theta = 1$ | 1.0809 | 0.0150 | 1.0807 | 0.0150 | 1.0798 | 0.0145 |
| | $\beta = 2$ | 2.0906 | 0.0173 | 2.0882 | 0.0168 | 2.0980 | 0.0183 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 20 | $\theta = 1$ | 1.0755 | 0.0136 | 1.0752 | 0.0136 | 1.0751 | 0.0132 |
| | $\beta = 2$ | 2.0829 | 0.0155 | 2.0835 | 0.0157 | 2.0936 | 0.0170 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 30 | $\theta = 1$ | 1.0716 | 0.0128 | 1.0705 | 0.0125 | 1.0702 | 0.0120 |
| | $\beta = 2$ | 2.0856 | 0.0157 | 2.0854 | 0.0157 | 2.0909 | 0.0163 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| 50 | $\theta = 1$ | 1.0630 | 0.0104 | 1.0629 | 0.0104 | 1.0636 | 0.0102 |
| | $\beta = 2$ | 2.0796 | 0.0143 | 2.0801 | 0.0142 | 2.0852 | 0.0147 |
| | $\alpha = 1$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

## 6. Application to datasets

We explore two real-life datasets: Nigerian stock price exchange rates and the Worcester heart attack study. Nigerian stock price exchange rate data include the Naira-to-Euro exchange rate from January 2020 to January 2024 and the Naira-to-Riyal exchange rate from January 2024 to

September 2024. The data was obtained from the Central Bank of Nigeria at https://www.cbn.ng/rates. The Worcester heart attack dataset is collected from the University of Massachusetts Medical School's Department of Cardiology. The dataset includes information collected across 13 separate one-year intervals from 1975 to 2001, covering all MI admissions in Worcester-area hospitals as investigated in [44] and [45], respectively.

### *6.1. Competing distributions*

In this section, we present information criteria metrics that include Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC), as well as competing distributions, preferably, the heavy-tailed distributions, including the Lomax distribution by [9], Exponentiated-Lomax (ELomax) distribution by [46], Power-Lomax (PLomax) distribution by [47], Topp Leone-Lomax (TLomax) distribution by [48], Odd Lomax-Lomax (OLLomax) distribution by [24], Maxwell-Lomax (MLomax) distribution by [49], Rayleigh-Lomax (RLomax) distribution by [50], Generalized Pareto (GPareto) distribution by [51], Generalized Beta-II (GBeta-II) distribution by [52], and Dagum distribution by [53]. The statistical distribution featuring the lowest score on these criteria and the highest logarithmic likelihood (L) should be chosen as the best fit for the dataset.

Tables 6 and 7 compare the findings of the proposed distribution with those of the competitor distributions using exchange rate datasets using L, AIC, CAIC, BIC, and HQIC, respectively.

Tables 6–8 show the LLomax distribution results for Naira-to-Euro, Naira-to-Riyal, and Worcester heart attack datasets. The LLomax distribution has the lowest AIC, CAIC, BIC, and HQIC values, as well as the highest log-likelihood value among competing distributions, accounting for the Naira-to-Euro, Naira-to-Riyal, and Worcester heart attack datasets. This demonstrates that the LLomax distribution is best suited for modeling these datasets.

**Table 6.** Goodness-of-fit metrics for the LLomax distribution compared to other models using the Naira-to-Euro exchange rate dataset.

| Model | Estimates | | | L | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|---|---|---|
| LLomax | $\beta$ =1.20 | $\theta$=14990.00 | $\alpha$ =0.36 | -6473.16 | 12952.32 | 12952.35 | 12967.02 | 12957.91 |
| OLLomax | $\beta$ =3.01 | $\theta$=4783.58 | $a$=1.20 | -7150.86 | 14307.72 | 14307.75 | 14322.42 | 14313.31 |
| PLomax | $\beta$=2.26 | $\theta$= 7495.00 | $b$=1.28 | -7131.88 | 14269.75 | 14269.77 | 14284.44 | 14275.34 |
| ELomax | $\beta$=0.150 | $\theta$=0.01 | $c$=0.15 | -8966.21 | 17938.42 | 17938.44 | 17953.11 | 17944.01 |
| TLomax | $\beta$=0.07 | $\theta$=0.01 | $d$=0.15 | -8814.50 | 17635.00 | 17635.03 | 17649.70 | 17640.59 |
| MLomax | $\beta$=0.04 | $\theta$=0.02 | $e$=4.53 | -8814.50 | 17635.00 | 17635.03 | 17649.70 | 17640.59 |
| RLomax | $\beta$=1.52 | $\theta$=16.94 | $f$=158.79 | -6498.38 | 13002.75 | 13002.77 | 13017.44 | 13008.34 |
| GPareto | $\tau$=0.01 | $\mu$=55.78 | $\sigma$=1.18 | -40517.58 | 81041.16 | 81041.19 | 81055.57 | 81046.66 |
| GBeta-II | $\psi$=101.90 | p=0.94 | q=0.01 | -8825.71 | 17657.42 | 17657.45 | 17671.83 | 17662.92 |
| Dagum | $\gamma$=548.12 | $\lambda$=2.17 | $\eta$=949.60 | -6636.59 | 13279.19 | 13279.21 | 13293.88 | 13284.78 |
| Lomax | $\beta$=10310.00 | $\theta$=5337000.00 | | -7176.90 | 14357.79 | 14357.80 | 14367.59 | 14361.52 |

**Table 7.** Goodness-of-fit metrics for the LLomax distribution compared to other models using the Naira-to-Riyal exchange rate dataset.

| Model | Estimates | | | L | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|---|---|---|
| LLomax | $\beta$ =2.75 | $\theta$=32570.00 | $\alpha$ =0.38 | -1092.56 | 2191.12 | 2191.25 | 2200.76 | 2195.03 |
| OLLomax | $\beta$ =3.12 | $\theta$=36753.22 | $a$=1.31 | -1256.67 | 2519.34 | 2519.47 | 2528.98 | 2523.25 |
| PLomax | $\beta$=3.28 | $\theta$= 8342.00 | $b$=1.35 | -1251.46 | 2508.92 | 2509.06 | 2518.57 | 2512.83 |
| ELomax | $\beta$=0.76 | $\theta$=0.02 | $c$=2.49 | -1366.30 | 2738.60 | 2738.73 | 2748.24 | 2742.51 |
| TLomax | $\beta$=0.15 | $\theta$=0.01 | $d$=0.15 | -1609.28 | 3224.55 | 3224.68 | 3234.20 | 3228.46 |
| MLomax | $\beta$=0.03 | $\theta$=0.01 | $e$=4.75 | -1746.34 | 3498.69 | 3498.82 | 3508.33 | 3502.60 |
| RLomax | $\beta$=0.053 | $\theta$=0.018 | $f$=3.15 | -1881.23 | 3768.46 | 3768.59 | 3778.10 | 3772.37 |
| GPareto | $\tau$=2.77 | $\mu$=227.72 | $\sigma$=0.23 | -1538.36 | 3082.72 | 3082.853 | 3092.365 | 3086.629 |
| GBeta-II | $\psi$=69.74 | p=1.65 | q=0.02 | -1580.412 | 3166.824 | 3166.957 | 3176.469 | 3170.733 |
| Dagum | $\gamma$=3607.00 | $\lambda$=2.82 | $\eta$=4121.00 | -1114.24 | 2234.47 | 2234.61 | 2244.12 | 2238.38 |
| Lomax | $\beta$=1902.00 | $\theta$=719200.00 | | -1277.79 | 2559.59 | 2559.66 | 2566.02 | 2562.196 |

**Table 8.** Goodness-of-fit metrics for the LLomax distribution compared to other models using the heart attack dataset.

| Model | Estimates | | | L | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|---|---|---|
| LLomax | $\beta$ =0.01 | $\theta$=9.46.30 | $\alpha$ =3.18 | -194.98 | 395.97 | 396.22 | 403.78 | 399.13 |
| OLLomax | $\beta$ =3.07 | $\theta$=2394.14 | $a$=3.28 | -214.54 | 435.09 | 435.33 | 442.89 | 438.24 |
| PLomax | $\beta$=3.24 | $\theta$= 2156.66 | $b$=3.41 | -201.69 | 409.38 | 409.63 | 417.19 | 412.54 |
| ELomax | $\beta$=266.10 | $\theta$0.01 | $c$=23.55 | -206.63 | 419.26 | 419.51 | 427.07 | 422.42 |
| TLomax | $\beta$=174.90 | $\theta$=0.01 | $d$=23.37 | -206.56 | 419.10 | 419.35 | 426.92 | 422.26 |
| MLomax | $\beta$=0.01 | $\theta$=0.01 | $e$=2.36 | -310.53 | 627.06 | 627.31 | 634.88 | 630.22 |
| RLomax | $\beta$=0.16 | $\theta$=0.05 | $f$=3.15 | -472.42 | 950.83 | 951.08 | 958.65 | 954.00 |
| GPareto | $\tau$=0.00 | $\mu$=2.98 | $\sigma$=1.20 | -341.16 | 688.33 | 688.58 | 696.14 | 691.49 |
| GBeta-II | $\psi$=0.34 | p=287.50 | q=191.20 | -201.87 | 409.74 | 409.99 | 417.55 | 412.90 |
| Dagum | $\gamma$=10340.00 | $\lambda$=5.02 | $\eta$=1.37 | -197.50 | 401.00 | 401.25 | 408.82 | 404.17 |
| Lomax | $\beta$=36950.00 | $\theta$=253000.00 | | -292.60 | 589.20 | 589.33 | 594.41 | 591.31 |

Figures 3–8 support the findings of Tables 6–8, demonstrating that the suggested log-lomax distribution may outperform other competing distributions for the Naira-to-Euro, Naira-to-Riyal, and Worcester heart attack datasets.

Table 9 shows the numerical differences between the LLomax distribution and existing heavy-tailed distributions in various data applications. The table covers the most significant variations in support, goodness-of-fit metrics, and numerical applications. The LLomax distribution can accommodate both positive and negative data, making it useful for statistical modeling tasks.

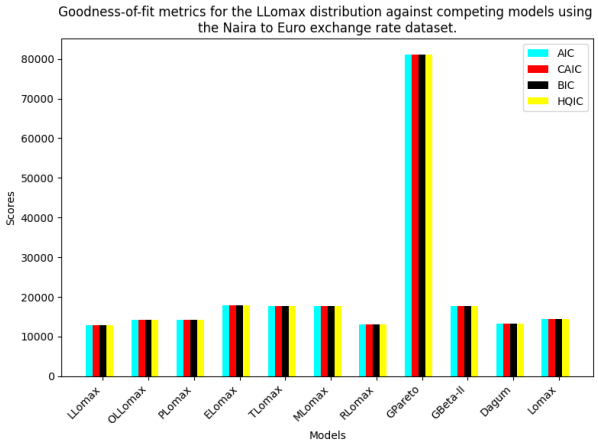



**Figure 3.** Goodness-of-fit metrics for the Log-Lomax distribution against competing models using the Naira to Euro exchange rate dataset.

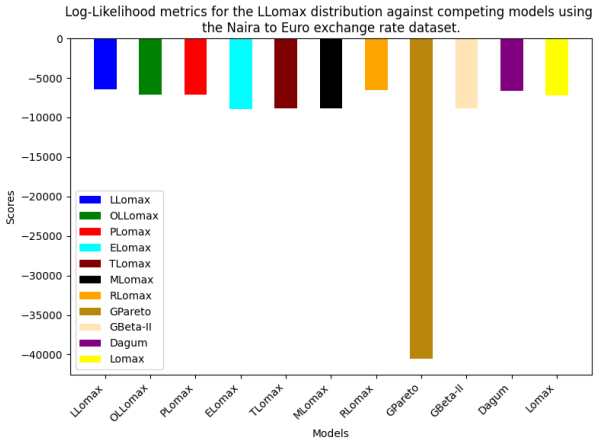



**Figure 4.** Log-Likelihood metrics for the Log-Lomax distribution against competing models using the Naira to Euro exchange rate dataset.

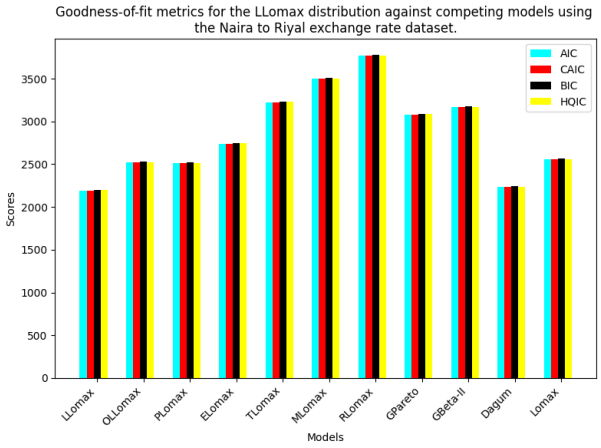



**Figure 5.** Goodness-of-fit metrics for the Log-Lomax distribution against competing models using the Naira to Riyal exchange rate dataset.

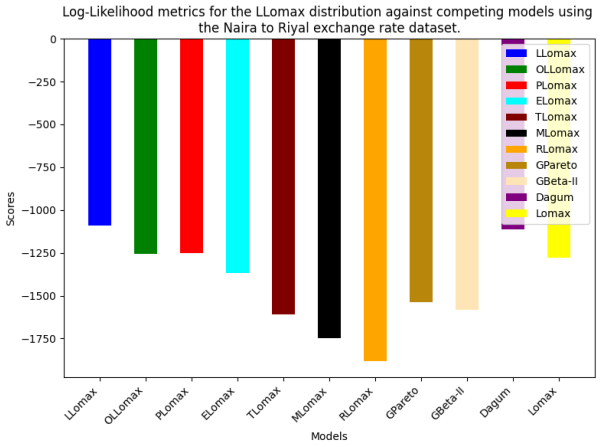



**Figure 6.** Log-Likelihood metrics for the Log-Lomax distribution against competing models using the Naira to Riyal exchange rate dataset.

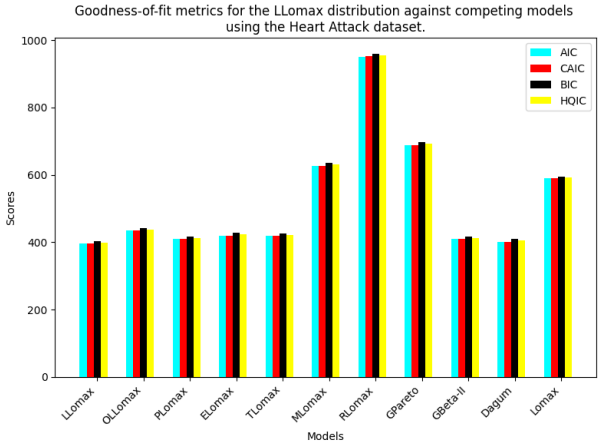



**Figure 7.** Goodness-of-fit metrics for the Log-Lomax distribution against competing models using the Heart Attack dataset.

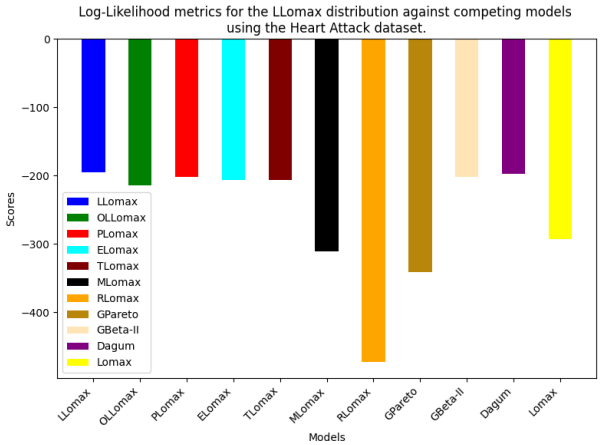



**Figure 8.** Log-Likelihood metrics for the Log-Lomax distribution against competing models using the Heart Attack dataset.

**Table 9.** Heavy-tailed distributions with their numerical findings.

| Model | Support | Goodness-of-fit (AIC) | | Applications |
|---|---|---|---|---|
| Lomax [54] | $(0, \infty)$ | 209.90 (Lomax) | 213.90 (Gamma) | Airborne Communication |
| | | 212.90 (Weibull) | | |
| OLLomax [24] | $(0, \infty)$ | 825.00 (OLLomax) | 831.67 (Lomax) | Bladder Cancer |
| | | 828.08 (WLomax) | 826.14 (ELomax) | |
| | | 825.25 (TLomax) | | |
| PLomax [47] | $(0, \infty)$ | 831.67 (OLLomax) | 825.48 (PLomax) | Bladder Cancer |
| | | 835.95 (ELomax) | 829.62 (WLomax) | |
| TLomax [48] | $(0, \infty)$ | 194.14 (OLLomax) | 185.27 (PLomax) | Airborne Communication |
| | | 88.54 (TLomax) | 185.52 (TLomax) | |
| MLomax [49] | $(0, \infty)$ | -32.34 (MLomax) | -3.93 (ELomax) | Milk Production |
| | | 0.47 (MOLomax) | | |
| | | 829.12 (MLomax) | 832.09 (ELomax) | Bladder Cancer |
| | | 830.67 (MOLomax) | | |
| LLomax [Proposed] | $(-\infty, \infty)$ | 2191.12 (Lomax) | 2519.34 (OLLomax) | Naira-Euro |
| | | 2738.60 (ELomax) | 81041.16 (GPareto) | |
| | | 2508.92 (PLomax) | 17657.42 (GBeta-II) | |
| | | 2191.12 (OLLomax) | 2508.92 (PLomax) | Naira-Riyal |
| | | 3082.72 (GPareto) | 3166.82 (GBeta-II) | |
| | | 2738.60 (ELomax) | | |
| | | 395.97 (OLLomax) | 409.38 (PLomax) | Heart Attack |
| | | 688.33 (GPareto) | 409.74 (GBeta-II) | |
| | | 419.26 (ELomax) | | |

Other heavy-tailed distributions are listed below: Weibull-Lomax (WLomax) and Marshall-Olkin-Lomax (MOLomax).

## 7. Prediction using machine learning

### *7.1. Data preprocessing*

Data preprocessing is important in achieving accurate prediction. We consider the descriptive statistics to guide this research in obtaining a quality dataset for accurate predictions. Table 10 presents descriptive statistics relating to the Euro dataset, which offers valuable information for machine learning predictions. The daily structure of the data is confirmed by the time feature, which has a mean of 15.75 and a range of 1 to 31, with a spread of 8.75, suggesting a reasonable distribution over the month. This feature helps improve the model's predictive accuracy by capturing time series trends or seasonality. The presence of outliers may have an impact on the model's performance, as seen by the large range of sales values from 329.53 to 1475.34. Handling these outliers may enhance the model's robustness, and utilizing

quartiles to visualize the skewness in the sales distribution can help guide normalization and other data preprocessing techniques. The minimal class imbalance is shown by the binary sell direction feature, which has a balanced mean close to 0.5 that supports more equal predictions in model classification.

**Table 10.** Descriptive statistics of the Euro dataset.

| | Time(days) | Sell | Sell Direction |
|---|---|---|---|
| Observation | 990 | 990 | 990 |
| Mean | 15.75 | 517.67 | 0.50 |
| Standard deviation | 8.72 | 167.55 | 0.50 |
| Minimum | 1.00 | 329.53 | 0.00 |
| $1^{st}$ Quartile | 8.00 | 438.83 | 0.00 |
| $2^{nd}$ Quartile | 16.00 | 462.30 | 0.50 |
| $3^{rd}$ Quartile | 23.00 | 493.01 | 1.00 |
| Maximum | 31.00 | 1475.34 | 1.00 |

Similarly, Table 11 of the Riyal dataset contains 184 times (days), the sell price, and sell direction observations. With a mean of 15.63 and a standard deviation (SD) of 8.87, and the time feature has a range of 1 to 31 days, suggesting a generally uniform distribution throughout the month. The distribution of the sale price feature, which has a mean of 381.55 and an SD of 58.32, exhibits notable variability. To enhance prediction accuracy, data transformation techniques may be beneficial given its slightly negatively skewed distribution. The sell direction feature indicates a well-balanced dataset, lowering the possibility of bias and enhancing machine learning prediction reliability, with a mean and SD of 0.50.

**Table 11.** Descriptive Statistics of the Riyal Dataset.

| | Time (days) | Sell | Sell Direction |
|---|---|---|---|
| Observation | 184.00 | 184.00 | 184 |
| Mean | 15.63 | 381.55 | 0.50 |
| Standard deviation | 8.87 | 58.32 | 0.50 |
| Minimum | 1.00 | 227.64 | 0.00 |
| $1^{st}$ Quartile | 8.00 | 368.07 | 0.00 |
| $2^{nd}$ Quartile | 16.00 | 400.48 | 0.50 |
| $3^{rd}$ Quartile | 23.00 | 423.47 | 1.00 |
| Maximum | 31.00 | 444.36 | 1.00 |

According to Table 12, most patients are 68 years of age or older, the average follow-up time for heart attack patients is approximately 1505 days, the data set is balanced in terms of follow-up category and vital status (dead or alive), and hospital stays range from 1 to 56 days.

**Table 12.** Descriptive statistics of the heart attack dataset.

| | Follow up Time | Length of Stay | Vital Status | Age | Gender | BMI | Follow up Category |
|---|---|---|---|---|---|---|---|
| Observation | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Mean | 1505.40 | 6.84 | 0.51 | 68.52 | 0.35 | 27.04 | 0.50 |
| Standard deviation | 858.89 | 5.92 | 0.50 | 14.43 | 0.48 | 4.86 | 0.50 |
| Minimum | 6.00 | 1.00 | 0.00 | 32.00 | 0.00 | 27.19 | 0.50 |
| $1^{st}$ Quartile | 715.00 | 4.00 | 0.00 | 59.75 | 0.00 | 23.54 | 0.00 |
| $2^{nd}$ Quartile | 1877.50 | 5.00 | 1.00 | 71.00 | 0.00 | 27.19 | 0.50 |
| $3^{rd}$ Quartile | 2076.50 | 7.00 | 1.00 | 80.25 | 1.00 | 30.35 | 1.00 |
| Maximum | 2719.00 | 56.00 | 1.00 | 92.00 | 1.00 | 39.94 | 1.00 |

### *7.2. Feature engineering*

The feature selection enables us to extract and engineer relevant and influential features to predict market stock prices. The target features for Euro data are engineered from a selling feature, which categorizes prices into "low"and "high"by considering the median price value of 462.3 as the threshold value. Below the threshold value is considered a low category, and above the threshold is regarded as a high category. The median price value of 400.48 for the Riyal dataset is considered the threshold value to classify the price as low price"and "high price".

The correlation study provides important information for machine learning prediction for both the Euro and Riyal datasets presented in Figures 9 and 10, respectively. The near-zero correlation between the time feature and the sell_direction (0.00 for Euro, -0.02 for Riyal) demonstrates that linear time-based patterns have a minimal direct influence on price prediction for either currency. The absence of a linear relationship, on the other hand, implies that the time-based feature's predictive value is found in non-linear patterns that more advanced models can identify, which does not lessen the feature's potential significance. In contrast with the Riyal dataset, the Euro expresses a moderate correlation (0.55) between sell price and sell direction, suggesting distinct linear correlations between current prices and their categorization as high or low for the two currencies.

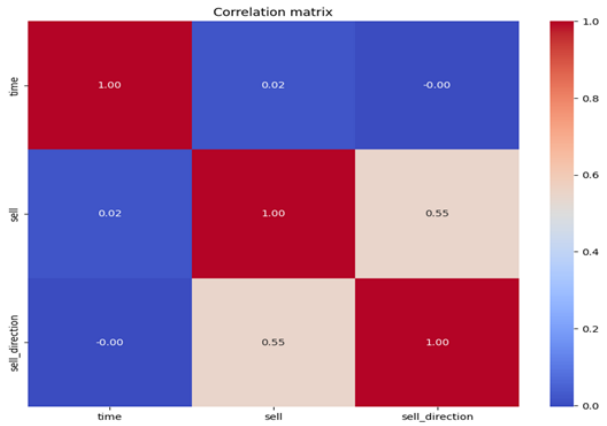



**Figure 9.** Heatmap of the correlation matrix of Euro dataset.

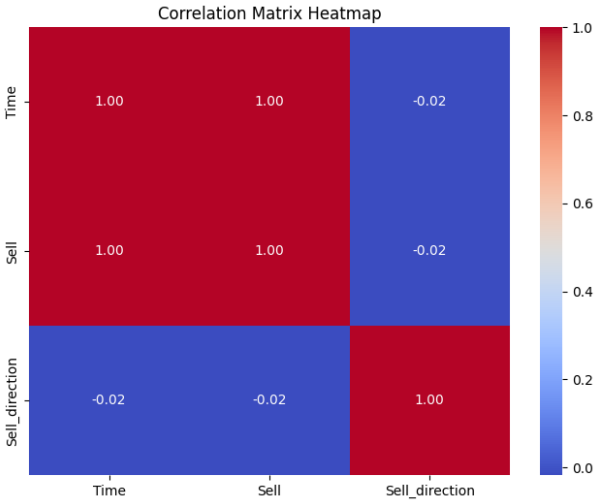



**Figure 10.** Heatmap of the correlation matrix of the Riyal dataset.

These correlation patterns contribute to the understanding of the prediction results' discrepancy in performance between linear and non-linear models. The fact that time and sale features for both currencies have low correlation (0.02) highlights the need for effective prediction models to identify interactions between features and hidden non-linear correlations that correlation analysis alone cannot disclose. Getting an understanding of this is essential to building strong prediction models that are flexible enough to adjust to the particularities of various currency markets.

Figure 11 shows that there is a strong correlation between the follow-up category and follow-up time, 0.79, and a moderate correlation with length of stay of 0.21 and BMI of 0.23. On the other hand, there is an inverse relationship between follow-up category and vital status of -0.70, age of -0.25, and gender of -0.19, indicating that longer follow-up times affect survival.

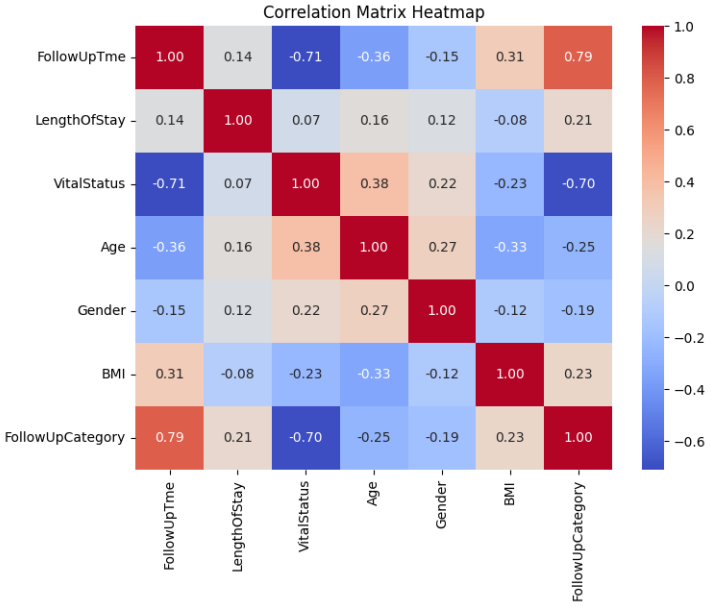



**Figure 11.** Heatmap of the correlation matrix of the Heart Attack dataset.

### 7.3. Model selection

The following models are chosen to predict market prices based on their advantages.

#### 7.3.1. Logistics regression

A popular statistical technique for classification and prediction is logistic regression. This is a form of regression that is useful for forecasting a categorical result based on the presence of one or more continuous or categorical features [55]. The goal of logistic regression is to use data on independent features to estimate the likelihood that an event will occur. Since the result signifies the likelihood, the dependent variable falls between 0 and 1. The odds, which show the ratio of the chance of success to the chance of failure, are transformed logistically. This research employed logistics regression due to its peculiarity in handling model overfitting by using L2 regularization. The mathematical representation of the logistic function is given below:

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}, \tag{7.1}$$

where $\beta_1$ is the weight given to input features $x$ and $\beta_0$ is the bias term.

#### 7.3.2. Support vector machine

Support vector machine (SVM), being among the foremost well-known supervised learning algorithms, is used for classification and regression tasks; however, in this study, it is just employed for stock price classification. SVM selects the odd characteristics that support the hyperplane, also known as support vectors, from which the algorithm derives its name [56]. The SVM algorithm identifies the best line or decision limit for partitioning an n-dimensional space into binary classes, enabling us to conveniently place fresh data points in the right classification [56]. A hyperplane represents the optimal decision limit. In this study, SVM is employed to deal with non-linearity. The mathematical representation of SVM is provided in:

$$\omega x + b = 0, \tag{7.2}$$

where $\omega$ is the weight attached to different input features $x$ and $b$ is the bias term.

#### 7.3.3. K-nearest neighbor

K-nearest neighbor (KNN) is a straightforward but efficient technique for predicting stock price. KNN makes predictions based on the closest training feature space; the feature vectors and class labels of the training samples are stored during the algorithm's training phase. The test sample whose class is unknown has the same features generated for it as before during the real prediction step [57]. The k nearest sample is chosen after calculating the distances between each newly stored vector and the original vector. The new point will be expected to be in the class with the highest number in the set. The distance between points is measured using Euclidean distance as given in:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{7.3}$$

#### 7.3.4. Random forest

This is an ensemble learning method that uses a large number of decision trees to generate predictions. Each tree is generated from a random subset of the data, whereas the total forecast is

derived by averaging the forecasts of all trees [58]. This method reduces overfitting and increases model accuracy by merging results from numerous different models. The mathematical formulation of Random Forest is more advanced than that of standard regression models, yet the underlying idea can be described as:

$$Y = \frac{1}{N}\sum_{i=1}^{n} f(x), \tag{7.4}$$

where $Y$ represents the projected value, N specifies the number of trees in the forest, and $f(x)$ represents the forecast given by the $i^{th}$ tree for the input x. During training, a total of 100 estimators are assigned.

#### 7.3.5. XGBoost

XGBoost is a highly effective and extensible machine learning algorithm for tree boosting, widely utilized in stock price prediction. XGBoost is built to be extremely portable, adaptable, and efficient [59]. Boosting tasks involve merging some low-accuracy weak classifiers to produce a stronger classifier that performs better in classification. Gradient Boosting Machines are named after the loss function's gradient direction, which defines the weak learner for each step.

### *7.4. Proposed procedure for integrating log-Lomax distribution into machine learning model*

#### 7.4.1. Input features

Two datasets are used, each containing Time and Sell as features and Sell_direction as the target variable. The time and sell features were transformed using the log-Lomax model.

#### 7.4.2. Log-Lomax distribution

The pdf and cdf for the log-Lomax distribution are employed and applied to integrate the probability into the machine learning model.

#### 7.4.3. Parameter initialization

The parameters $\beta$, $\theta$, and $\alpha$ were optimized via maximum likelihood estimation. In this study, the parameters needed to integrate the log-lomax model into machine learning are available in Tables 6–8 for the Euro, Riyal, and Heart Attack datasets, respectively, and is provided in Table 13.

**Table 13.** Parameter initialization for the Euro, Riyal, and Worcester Heart Attack datasets.

| Parameter | Euro | Riyal | Heart Attack |
|---|---|---|---|
| $\beta$ | 1.20 | 2.75 | 0.01 |
| $\theta$ | 14,990.00 | 32,570.00 | 9,46.30 |
| $\alpha$ | 0.36 | 0.38 | 3.18 |

#### 7.4.4. Feature transformation

The log-Lomax cdf is used to transform the Time and Sell features into new features. These transformed attributes are used to generate a new dataset for machine learning models. Figure 12 depicts the processes involved in incorporating log-lomax features into a machine learning model.

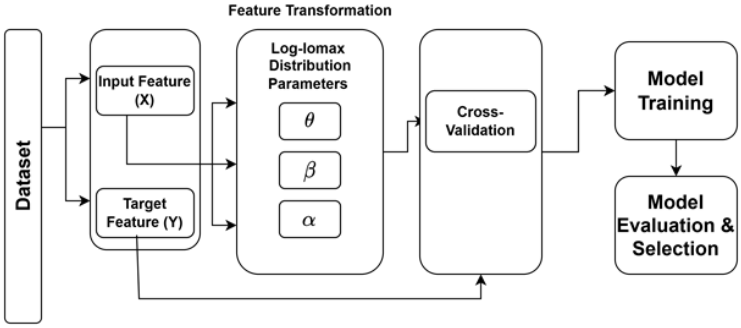



**Figure 12.** Process flow for feature transformation.

#### 7.4.5. Data split

The transformed dataset is divided into training and testing sets in an 80:20 ratio.

#### 7.4.6. Model training

Machine learning algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost are trained on the transformed dataset.

#### 7.4.7. Performance metrics

Model performance is assessed using metrics such as Accuracy, Recall, and F1 Score. Additional evaluation criteria include Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent AIC (CAIC), Hannan–Quinn Information Criterion (HQIC), and Log-Likelihood (LL).

- **Confusion matrix:** The confusion matrix is a method of evaluating machine learning models [60]. A confusion matrix is a $2 \times 2$ matrix with true labels displayed in its rows and predicted labels displayed in its columns, as seen in Table 14. The metrics are derived from the confusion matrix [61].

**Table 14.** Confusion Matrix.

| Predicted Values | True Values | |
|---|---|---|
| | Positive (1) | Negative (0) |
| Positive (1) | True Positive (TP) | False Positive (FP) |
| Negative (0) | False Negative (FN) | True Negative (TN) |

- **Accuracy:** This measures the rate of correct predictions on the overall prediction performance of the model and it is mathematically given as:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}. \tag{7.5}$$

- **Recall:** This measures the rate of predicting positive instances, it is given as:

$$Recall = \frac{TP}{TP + FN}. \tag{7.6}$$

- **F1-Score:** This is the harmonic mean of precision and recall, which can be expressed as:

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall}. \tag{7.7}$$

- **Log-likelihood:** These measure how well a model fits a given dataset by taking the natural logarithm of the likelihood function. This metric shows how likely it is to observe the data given the parameters of the model; a higher log-likelihood value indicates a better model fit because it indicates a higher probability of observing the data under the assumptions of the model.
- **AIC:** This prediction error estimator evaluates the relative quality of statistical models for each given data collection. Given a collection of models that match the data, AIC determines the quality of each model compared to the others. As a result, the strategy that has the lowest AIC is the most effective for model selection.
- **BIC:** This method determines the optimal statistical model from a set of alternative models by balancing the goodness of fit with the model's complexity, favoring models with a lower BIC value. Essentially, it penalizes models with too many parameters compared to the amount of data available.
- **CAIC:** This is a model selection criterion used to assess alternative statistical models, like the more commonly recognized AIC (Akaike Information Criterion), but with an additional penalty for model complexity that is intended to be more accurate, especially with large sample sizes. Essentially, CAIC aims to continuously select the "true" model as the sample size grows; a lower CAIC yields the best model section.
- **HQIC:** This model selection criterion, like the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), is used to assess various statistical models. A lower HQIC value shows better model fit while penalizing the number of parameters in the model; in effect, it serves as a compromise between the lower penalty for AIC and the higher penalty for BIC for model complexity.

#### 7.4.8. Model selection

In this models are compared based on their Accuracy, Recall, F1 Score, and LL, with higher values indicating better performance. Additionally, models with lower AIC, BIC, CAIC, and HQIC values are considered better fits.

### *7.5. Machine learning results*

In this section, we detail the results obtained from the training of the machine learning models. The training is conducted using a Python environment with a device having a Processor of 2.30 GHz and RAM of 16.0 GB. To prevent overfitting and ensure robustness, we split the data into 80/20 with 10-fold cross-validation. The capacity of tree-based models to capture complex patterns created by the log-Lomax transformation is shown in their high accuracy. Although such performance is outstanding, we use careful interpretation and verify dependability using several evaluation measures like the F1-score and careful feature engineering.

#### 7.5.1. Euro dataset prediction results

The machine learning training is conducted in two scenarios. The first scenario utilizes simply raw data, while the second scenario employs machine learning training with log-Lomax integrated features.

The performance of several machine learning models with raw features (before administered log-lomax distribution) for stock prediction is displayed in Table 15 and Figure 13. KNN ranks first in Accuracy (0.9873), Recall (0.9797), and F1 Score (0.9871). SVM and Logistic Regression both perform well, but KNN exceeds both in recall. XGBoost excels at recall (0.9977) but has significantly lower accuracy. Random Forest performs poorly, scoring the lowest across all metrics, demonstrating insufficient prediction ability.

**Table 15.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Euro exchange rate dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.9469 | 0.8956 | 0.9439 |
| Random Forest | 0.4823 | 0.4000 | 0.2589 |
| SVM | 0.9545 | 0.9101 | 0.9521 |
| KNN | 0.9873 | 0.9797 | 0.9871 |
| XGBoost | 0.9405 | 0.9977 | 0.9578 |

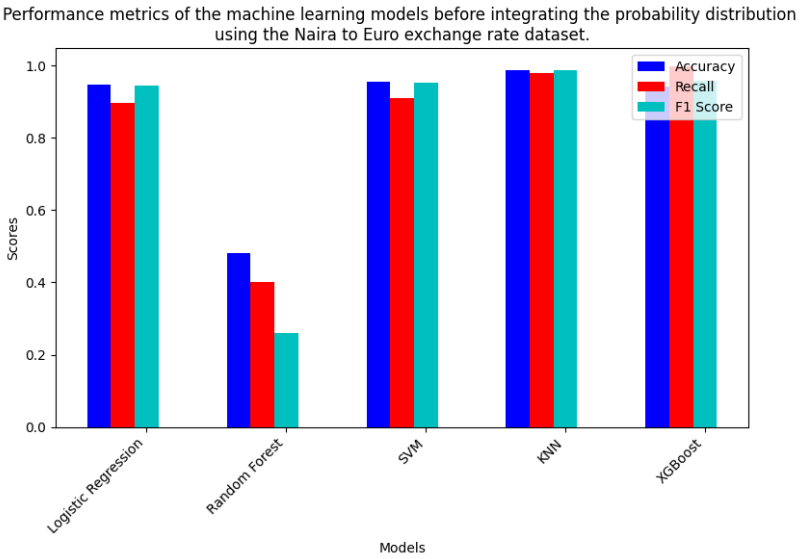



**Figure 13.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Euro exchange rate dataset.

Before the inclusion of probability distributions, as given in Table 16, Random Forest and KNN demonstrate the best model fit, as evidenced by their low AIC, BIC, and HQIC values in the training results for the Euro dataset. AIC and BIC values are greater for Logistic Regression and SVM, which demonstrates worse performance. Even though XGBoost provides a strong fit with a low log-likelihood, its higher AIC and BIC values suggest that it might be overfitting. Here, the most successful models are KNN and Random Forest. Figures 14 and 15 confirm these findings provided in Table 16.

**Table 16.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Euro exchange rate dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -89.0585 | 188.5132 | 189.1560 | 203.2327 | 199.3300 |
| Random Forest | -8.8308 | 53.6616 | 54.3660 | 141.8202 | 87.1843 |
| SVM | -70.5043 | 171.0085 | 171.5014 | 244.4741 | 198.9442 |
| KNN | -8.7147 | 33.4294 | 33.5762 | 72.6110 | 48.3284 |
| XGBoost | -6.6836 | 91.3672 | 94.6515 | 282.3777 | 164.0000 |

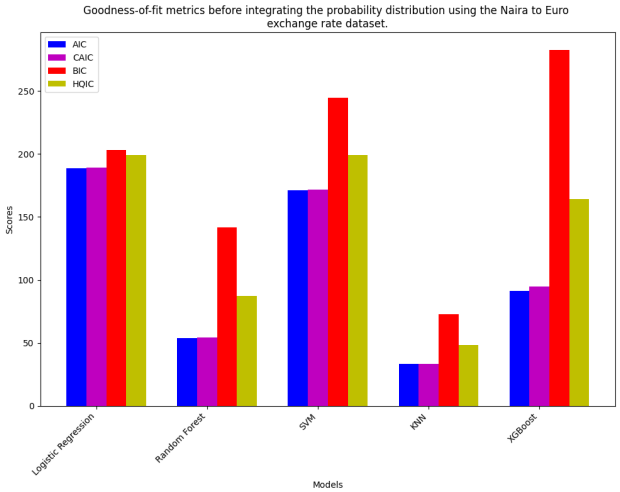



**Figure 14.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Euro exchange rate dataset.

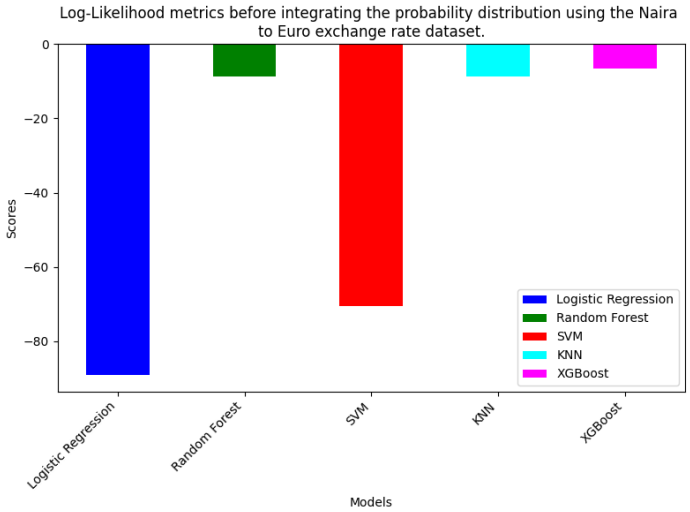



**Figure 15.** Log-Likelihood metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Euro exchange rate dataset.

Table 17 and Figure 16 show the performance of machine learning models using integrated probability distributions. The table demonstrates that all models perform quite well, with Random Forest and XGBoost having the highest accuracy and F1 scores (0.9987 and 0.9988, respectively). KNN and Logistic Regression also perform well, while SVM achieves slightly lower but still good results. Overall, the results show that the models, notably Random Forest and XGBoost, provide very reliable predictions.

**Table 17.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Euro exchange rate dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.9874 | 0.9749 | 0.987 |
| Random Forest | 0.9987 | 0.9977 | 0.9988 |
| SVM | 0.9646 | 0.9297 | 0.9628 |
| KNN | 0.9949 | 0.9952 | 0.9950 |
| XGBoost | 0.9987 | 0.9977 | 0.9988 |

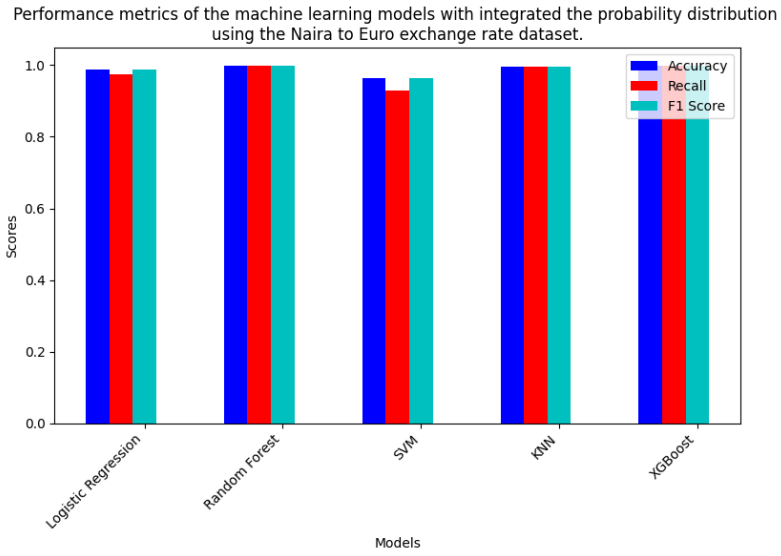



**Figure 16.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Euro exchange rate dataset.

Table 18 shows that Random Forest has the lowest AIC, BIC, CAIC, and HQIC, as well as the highest LL values, indicating an optimal mix of model fit and complexity. XGBoost and KNN are also effective with relatively low complexity measures. In comparison, Logistic Regression and SVM have higher values, implying that they may be less effective at capturing the data's underlying patterns.

Overall, using the log-Lomax distribution improves model performance, with Random Forest emerging as the most successful solution for the Euro dataset. Figures 17 and 18 confirm the findings in Table 18.

**Table 18.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Euro exchange rate dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -86.3698 | 178.7396 | 178.7639 | 193.4327 | 184.3300 |
| Random Forest | -1.8573 | 9.7146 | 9.7390 | 24.4077 | 15.3020 |
| SVM | -32.5719 | 71.1438 | 71.1682 | 85.8369 | 76.7310 |
| KNN | -4.6507 | 11.5871 | 15.3257 | 29.9945 | 20.8890 |
| XGBoost | -3.6886 | 13.3771 | 13.4015 | 28.0703 | 18.9640 |

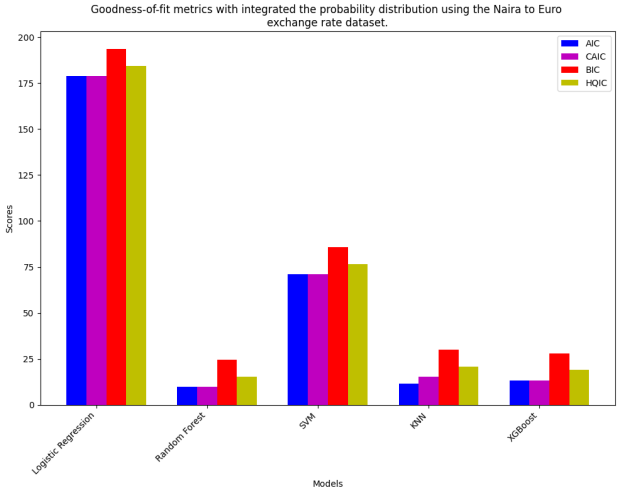



**Figure 17.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Euro exchange rate dataset.

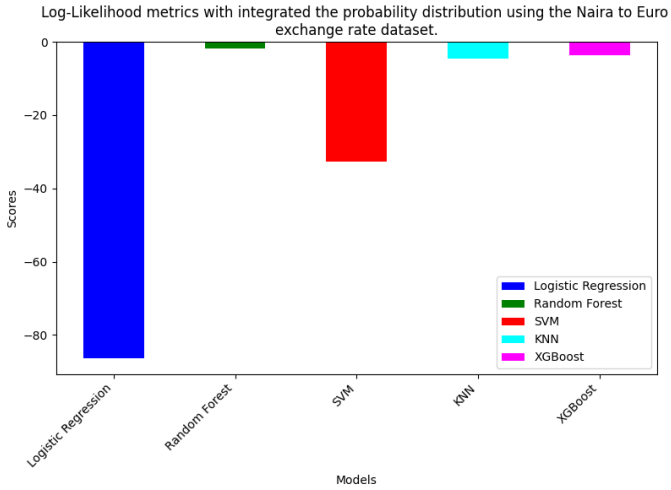



**Figure 18.** Log-Likelihood metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Euro exchange rate dataset.

In summary, integrating the log-Lomax distribution into machine learning models improves their

capacity to address skewness and heavy tails in stock price predictions, resulting in higher predicted accuracy and model fit. In practice, it offers a more adaptable probabilistic approach that balances model complexity and performance trade-offs, as indicated by reduced AIC and BIC values.

### 7.5.2. Riyal Dataset Prediction Results

The machine learning results for the raw features from the Riyal dataset reveal different model performances as shown in Table 19 and Figure 19. Before integrating the log-Lomax distribution, XGBoost produces an accuracy of 0.9933 and an F1 score of 0.9889, followed closely by Random Forest. KNN also performs well, while Logistic Regression demonstrated modest accuracy but high recall, suggesting its strength in detecting positive situations. SVM had the worst performance, showing the need for further optimization to increase predictive accuracy.

**Table 19.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.5652 | 0.8606 | 0.6565 |
| Random Forest | 0.9867 | 0.9800 | 0.9822 |
| SVM | 0.3876 | 0.5000 | 0.2808 |
| KNN | 0.9795 | 0.9589 | 0.9777 |
| XGBoost | 0.9933 | 0.9800 | 0.9889 |

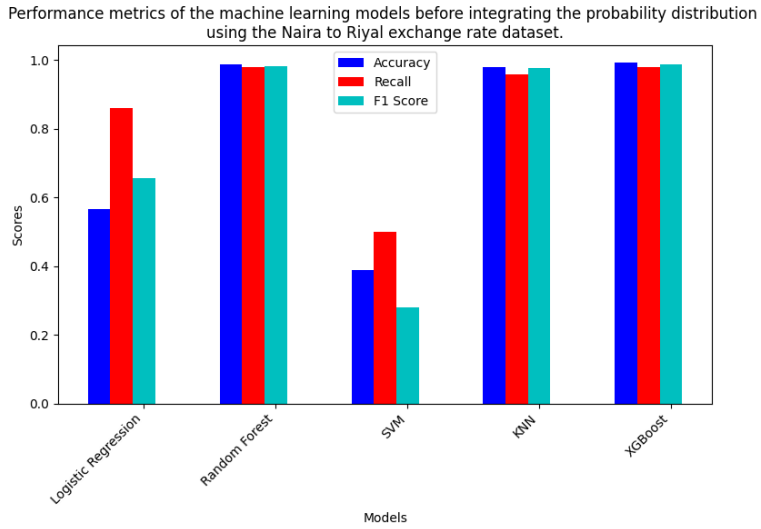



**Figure 19.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

The Riyal dataset training results are in Table 20; Figures 20 and 21, correspondingly. These results show that the most successful models are KNN and Random Forest, with the lowest AIC, BIC, and HQIC values, indicating better model fit. Conversely, the AIC and BIC values of Logistic Regression, SVM, and XGBoost were higher, suggesting less than ideal performance. The most promising models for this dataset are KNN and Random Forest.

**Table 20.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -99.4235 | 228.8470 | 231.7042 | 277.0711 | 248.39 |
| Random Forest | -9.7609 | 55.5218 | 59.6673 | 113.3907 | 78.977 |
| SVM | -102.4849 | 234.9698 | 237.8269 | 283.1938 | 254.5200 |
| KNN | -11.1638 | 38.3276 | 39.1505 | 64.0471 | 48.7520 |
| XGBoost | -31.9151 | 141.8302 | 163.4968 | 267.2127 | 192.6500 |

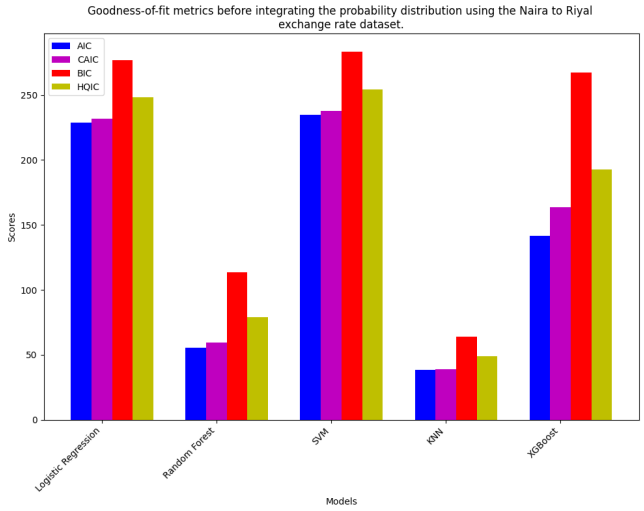



**Figure 20.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

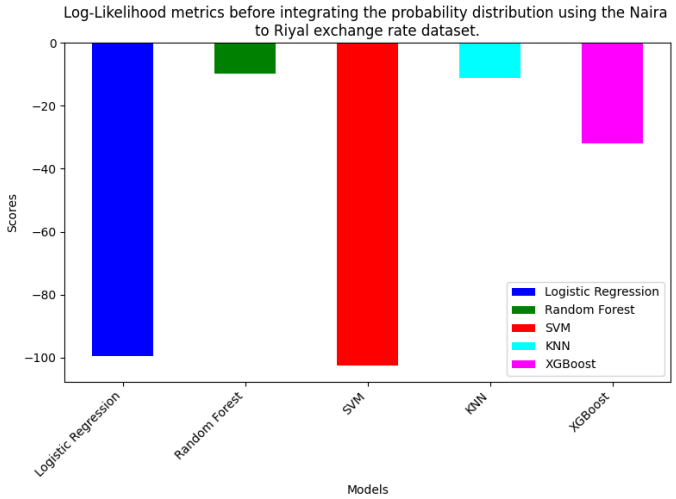



**Figure 21.** Log-Likelihood metrics of the machine learning models before integrating the Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

The improvement is demonstrated in several models by the findings of integrated log-Lomax probability distributions in Table 21 and Figure 22. After integrating the log-Lomax distribution, Random Forest and XGBoost maintain the highest performers, with XGBoost achieving the highest accuracy and F1 score. KNN produces good results, while Logistic Regression improves recall but has lower accuracy. SVM continues to underperform, with low accuracy and F1 scores.

**Table 21.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.6043 | 0.9000 | 0.6526 |
| Random Forest | 0.9867 | 0.9800 | 0.9933 |
| SVM | 0.3876 | 0.5000 | 0.2808 |
| KNN | 0.9457 | 0.9143 | 0.9386 |
| XGBoost | 0.9933 | 0.9800 | 0.9889 |

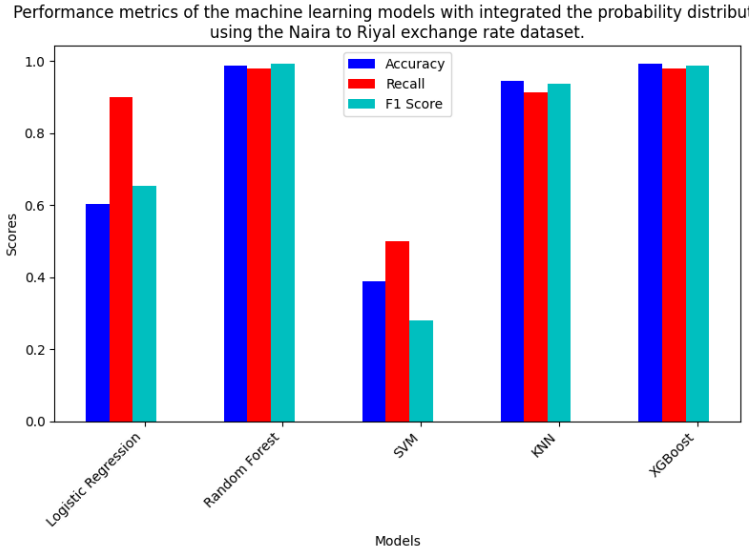



**Figure 22.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

After integrating the log-Lomax distribution with the Riyal dataset as contained in Table 22, and Figures 23 and 24 show that Random Forest and XGBoost, respectively, show the best performance, as evidenced by low AIC, CAIC, BIC, and HQIC values, indicating a good fit with the data and attainable model complexity. KNN follows with slightly higher values while being efficient. Logistic Regression has higher AIC and other criterion values, indicating a less effective fit than the other models. SVM has the lowest performance, with the highest values across all criteria, indicating poor fit and increased complexity. Overall, Random Forest and XGBoost are the most effective models after incorporating log-Lomax.

**Table 22.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -98.0401 | 202.0801 | 202.2135 | 211.7249 | 205.9900 |
| Random Forest | -1.9069 | 9.8137 | 9.9471 | 19.4586 | 13.7230 |
| SVM | -102.8778 | 211.7556 | 211.8889 | 221.4004 | 215.6600 |
| KNN | -4.8520 | 15.7041 | 15.8374 | 25.3489 | 19.6130 |
| XGBoost | -1.9384 | 9.8769 | 10.0102 | 19.5217 | 13.7860 |

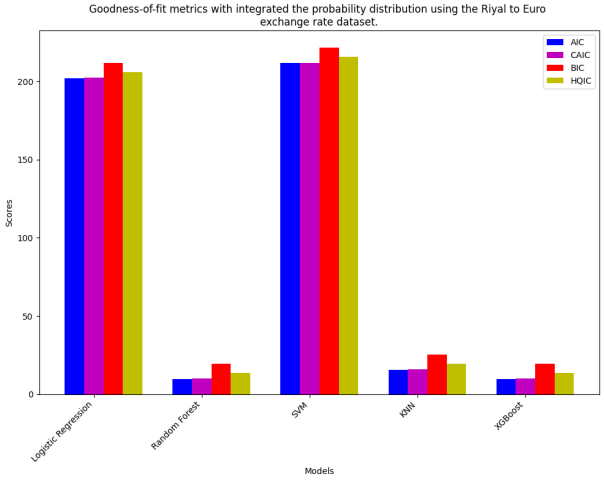



**Figure 23.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

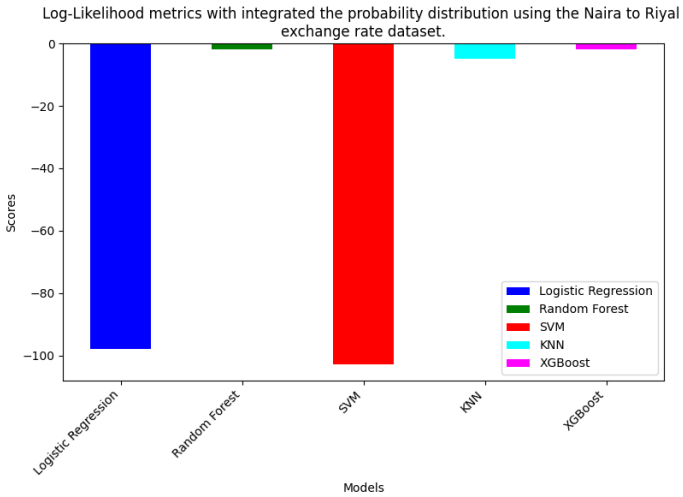



**Figure 24.** Log-Likelihood metrics of the machine learning models with integrated Log-Lomax distribution using the Naira to Riyal exchange rate dataset.

In conclusion, using the log-Lomax probability distribution in model training for the Riyal dataset dramatically improves model performance by better capturing the skewed and heavy-tailed character of stock price prediction, resulting in greater accuracy, recall, and F1 scores. Technically, this increases model fit, as evidenced by lower AIC, BIC, and other criteria, while enabling models like Random Forest and XGBoost to better capture the underlying data distribution, resulting in more accurate forecasts and robust risk management techniques.

### 7.5.3. Heart Attack Dataset Prediction Results

Table 23 shows the machine learning training results with raw features, which show that XGBoost and Random Forest are the best models for classification in this medical dataset, outperforming other models in terms of accuracy, recall, and F1 score. Figure 25 supports the findings in Table 23.

**Table 23.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using the Heart Attack dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.6750 | 0.7350 | 0.6950 |
| Random Forest | 0.8875 | 0.8750 | 0.8834 |
| SVM | 0.6500 | 0.7150 | 0.6812 |
| KNN | 0.6375 | 0.7550 | 0.6960 |
| XGBoost | 0.9000 | 0.8750 | 0.9044 |

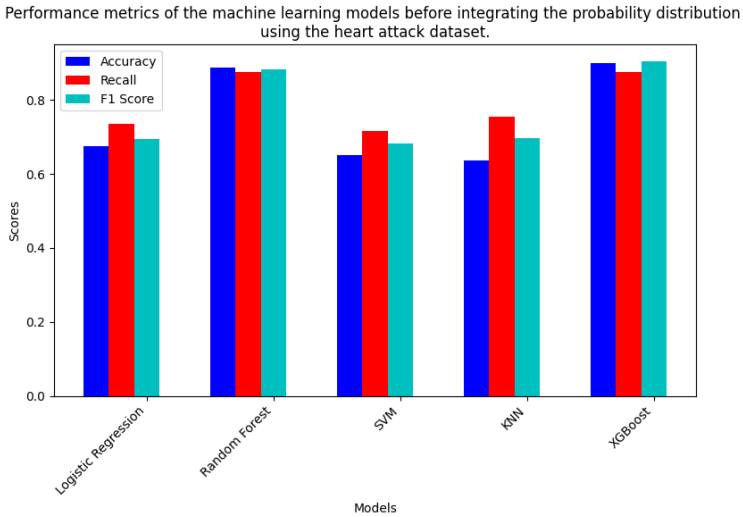



**Figure 25.** Performance metrics of the machine learning models before integrating the Log-Lomax distribution using Heart Attack dataset.

Table 24 shows that Random Forest and KNN provided the best model fit, with low AIC, BIC, and HQIC values in training outcomes for the Heart Attack dataset. Logistic Regression and SVM have higher AIC and BIC values, implying poor performance. Thus, Random Forest and KNN are the best models for this dataset. Figures 26 and 27 support the findings of Table 24.

**Table 24.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Heart Attack dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -43.7429 | 117.4986 | 123.2129 | 156.5762 | 133.3140 |
| Random Forest | -16.1388 | 70.2775 | 79.7775 | 99.7758 | 90.3104 |
| SVM | -47.2296 | 124.4591 | 130.1734 | 163.5367 | 140.2745 |
| KNN | -31.4157 | 78.8315 | 80.4139 | 99.6728 | 87.2663 |
| XGBoost | -5.7721 | 89.5441 | 141.5441 | 191.1457 | 130.6641 |

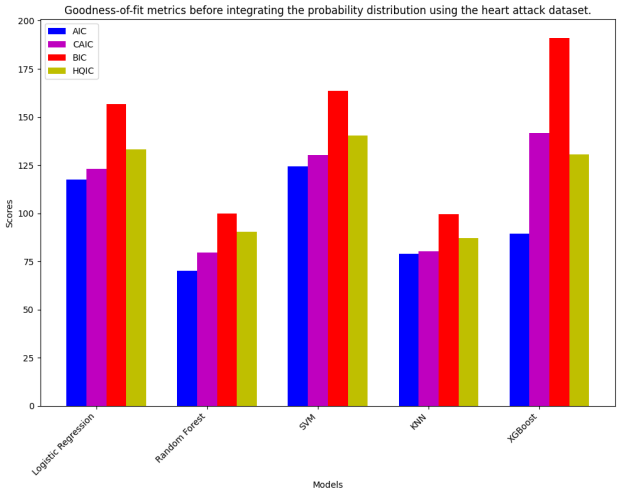



**Figure 26.** Goodness-of-fit metrics of the machine learning models before integrating the Log-Lomax distribution using the Heart Attack dataset.

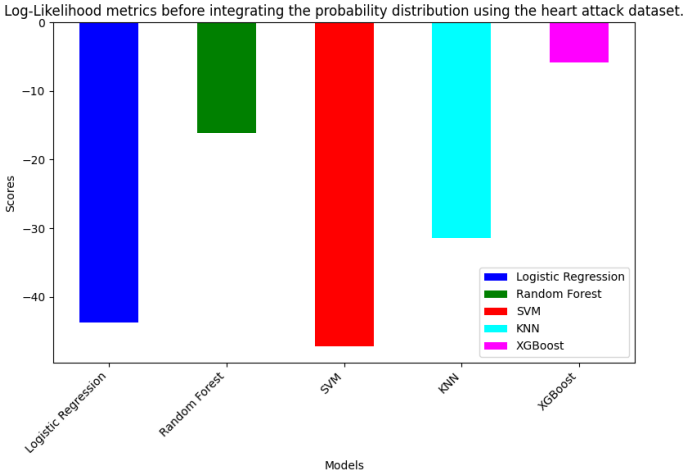



**Figure 27.** Log-Likelihood metrics of the machine learning models before integrating the Log-Lomax distribution using the Heart Attack dataset.

The use of log-Lomax probability features improve model performance, as shown in Table 25 and Figure 28. The table shows that the Random Forest had the highest accuracy (91.25%) and F1 score (91.51%), followed by XGBoost, while KNN, SVM, and Logistic Regression have intermediate success scores.

**Table 25.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Heart Attack dataset.

| Model | Accuracy | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.7000 | 0.8050 | 0.7416 |
| Random Forest | 0.9125 | 0.9150 | 0.9151 |
| SVM | 0.7000 | 0.8400 | 0.7515 |
| KNN | 0.5000 | 0.4033 | 0.4685 |
| XGBoost | 0.8875 | 0.8750 | 0.8953 |

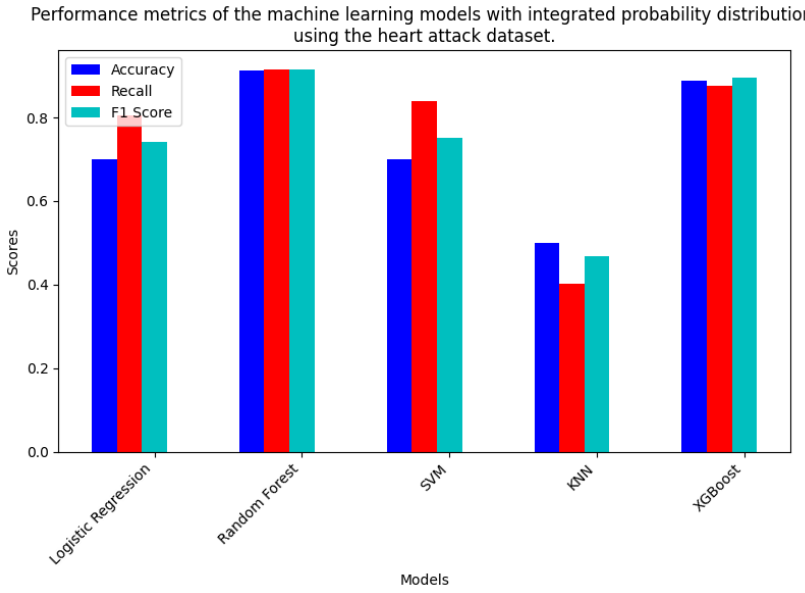



**Figure 28.** Performance metrics of the machine learning models with integrated Log-Lomax distribution using the Heart Attack dataset.

Table 26 indicates that Random Forest has the lowest AIC, BIC, CAIC, and HQIC, as well as the greatest LL values, indicating optimal model fit and complexity. Logistic Regression and SVM are also useful for low-complexity measures. Overall, utilizing the log-Lomax distribution enhances model performance, with Random Forest outperforming other strategies on the Heart Attack dataset. Figures 29 and 30 support the findings in Table 26.

Overall, Random Forest and XGBoost have the highest accuracy and recall, whereas the medical dataset analysis demonstrates that including log-Lomax probability features improves model performance. While SVM and Logistic Regression provide interpretable alternatives, these models are highly predictive, making them suitable for medical decision-making.

**Table 26.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Herat Attack dataset.

| Model | Log-Lik. | AIC | CAIC | BIC | HQIC |
|---|---|---|---|---|---|
| Logistic Regression | -37.3917 | 104.7835 | 110.4978 | 143.8610 | 120.5989 |
| Random Forest | -7.8741 | 53.7483 | 63.2483 | 103.2465 | 73.7811 |
| SVM | -43.6632 | 117.3263 | 123.0406 | 156.4039 | 133.1417 |
| KNN | -23.5670 | 63.1340 | 64.7164 | 83.9754 | 71.5689 |
| XGBoost | -5.5461 | 89.0921 | 141.0921 | 190.6938 | 130.2121 |

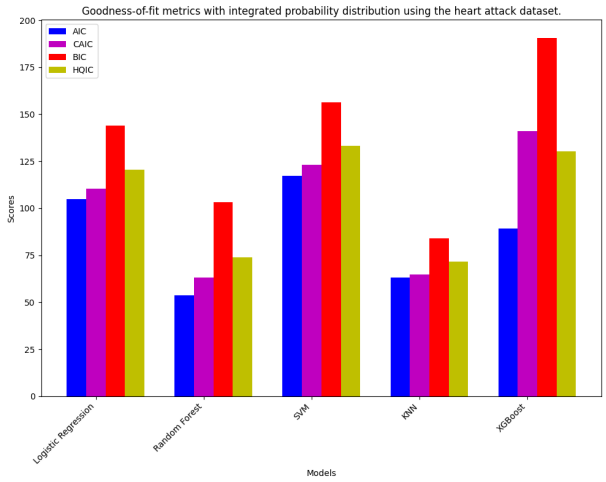



**Figure 29.** Goodness-of-fit metrics of the machine learning models with integrated Log-Lomax distribution using the Heart Attack dataset.

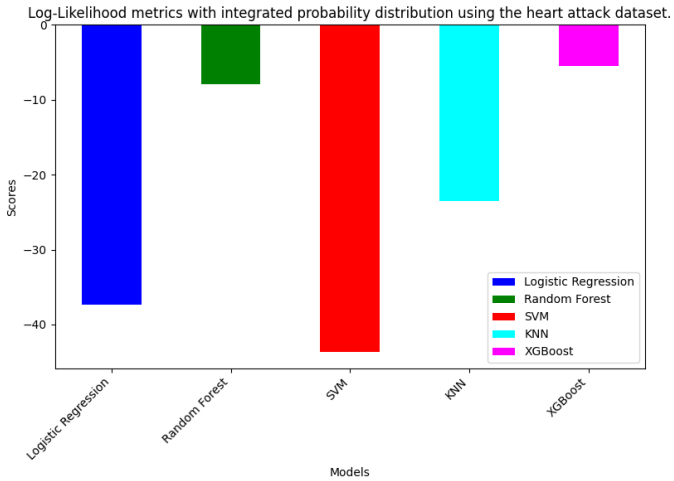



**Figure 30.** Log-Likelihood metrics of the machine learning models with integrated Log-Lomax distribution using the Heart Attack dataset.

## 8. Discussion

We explore how combining advanced statistical modeling, specifically the log-Lomax distribution, with machine learning techniques can improve prediction accuracy across various datasets. The datasets analyzed are the Nigerian Naira to British Pound Sterling exchange rate, the Naira to Riyal exchange rate, and the Worcester Heart Attack dataset. The findings show that the log-Lomax distribution performs better than other competing models by capturing the complex statistical characteristics of the data. It effectively handles the heavy-tailed nature and skewness of financial and medical data. Moreover, machine learning algorithms like Random Forest and XGBoost further boost predictive performance, achieving near-perfect accuracy in forecasting data trends. These algorithms leverage features from the log-Lomax model, outperforming traditional methods that use raw data alone.

Advanced tree-based models like Random Forest and XGBoost are well-suited for complex, heavy-tailed data because they do not rely on fixed distribution assumptions. Instead, they partition the data space to create flexible decision boundaries that enable them to identify unusual but significant patterns. XGBoost improves this by using gradient boosting and second-order optimization to simulate complex, non-linear interactions. At the same time, Random Forest employs bagging and random feature selection to stay robust and prevent overfitting, even in high-variance situations [62]. Logistic regression and simpler models, on the other hand, have difficulty with log-Lomax-transformed data due to linear assumptions and sensitivity to multicollinearity and skewness. SVM can manage some non-linearity using kernel approaches, but it performs poorly in high-dimensional or imbalanced situations unless carefully tuned. Furthermore, Random Forest and XGBoost outperform other algorithms because of their improved handling of heavy-tailed features and flexibility to irregular data distributions.

A significant contribution of this study is how well statistical distributions and machine learning algorithms can work together. While machine learning models like Random Forest and XGBoost are good at handling complex relationships in the data, their performance improves significantly when paired with a distribution like the log-Lomax. This combination provides a powerful way to model the unpredictable and volatile nature of financial and medical data. However, applying these methods to dynamic financial and medical datasets comes with challenges. Factors like market noise, volatility, and other financial behaviors can make predictions harder and may affect how well the models generalize across datasets or periods. The study also points out that the log-Lomax distribution has potential applications beyond financial markets. Although we focused on financial and medical applications, the distribution's flexibility means it could be useful in areas like risk management, insurance modeling, and economic forecasting. Researchers could look into these applications and explore how other machine learning algorithms could further improve predictions in more complex scenarios. Combining statistical distributions with machine learning in hybrid models could offer a stronger and more adaptable framework for different predictive tasks.

## 9. Concluding remarks

We introduce the log-Lomax distribution as a powerful probabilistic model for financial and medical modeling. By applying a logarithmic transformation, the log-Lomax distribution provides significant flexibility, accommodating various statistical behaviors such as skewness and nonlinearity. The model's ability to fit both left- and right-skewed distributions and exhibit diverse hazard rate shapes makes it an

excellent alternative to traditional Lomax-based models. We have derived essential statistical properties of the log-Lomax distribution and validated parameter estimation techniques using least squares, weighted least squares, and maximum likelihood estimation, which are confirmed through simulation studies. Empirical analysis of Nigerian stock price exchange rates (Naira-to-Euro and Naira-to-Riyal) and heart attack demonstrates that the log-Lomax distribution outperforms other Lomax-based models in fitting the complex characteristics of the data. Furthermore, the integration of features derived from the log-Lomax distribution into machine learning workflows enhances predictive accuracy, with models like Random Forest and XGBoost achieving near-perfect performance in classifying stock price trends. This highlights the advantages of combining advanced statistical modeling with machine learning techniques to improve predictive accuracy and reduce model complexity. We highlight how advanced statistical modeling, particularly the log-Lomax distribution, can be effectively integrated with machine learning techniques to enhance predictive accuracy and address real-world challenges in the financial and medical fields. While the study demonstrates the practical applicability of the log-Lomax distribution in financial contexts, there are opportunities for further research. Future work could extend the log-Lomax model to other financial datasets, explore alternative machine learning algorithms for optimization, and investigate its potential applications in fields like risk management and economic forecasting to gain deeper insights.

## 10. Future study

Researchers could address these limitations by considering the introduction of additional parameters or exploring alternative transformations to enhance the model's flexibility. For example, incorporating mixture models or copula-based extensions could enable the log-Lomax distribution to capture multimodal data and complex dependencies more effectively. Additionally, the generalization of the log-Lomax distribution to multivariate settings or the development of time-series extensions offers promising avenues for future work, enabling the model to handle more complex data structures and temporal dependencies.

Furthermore, exploring alternative parameter estimation methods, such as Bayesian estimation, could provide more robust and efficient estimation procedures, particularly in scenarios with limited sample sizes or complex likelihood functions. Investigating the asymptotic properties of the log-Lomax distribution and developing diagnostic tools for model validation would also contribute to a deeper understanding of its statistical behavior.

However, expanding the application of the log-Lomax distribution to broader fields, including engineering, finance, and environmental studies, offers significant potential. This could involve adapting the model to specific domain requirements and evaluating its performance in diverse real-world scenarios. By pursuing these future research directions, we aim to further enhance the log-Lomax distribution's utility and contribute to its development as a versatile tool in distribution theory and applied statistics.

Finally, researchers could focus on optimizing the performance of models like Random Forest and XGBoost through better hyperparameter tuning, using techniques like Bayesian optimization or grid search. It would also be beneficial to expand the datasets by adding other factors, such as macroeconomic indicators or geopolitical events, which could improve prediction accuracy and provide a more comprehensive view of market dynamics.

## Author contributions

Aliyu Ismail Ishaq: Conceptualization, Writing original draft, Formal analysis, Software, Investigation, Methodology; Abdullahi Ubale Usman: Validation, Writing-review and editing, Data curation, Methodology; Hana N. Alqifari: Writing-review and editing, Software, Investigation, Methodology, Supervision; Amani Almohaimeed: Conceptualization, Formal analysis, Writing original draft, Investigation, Supervision, Resources, Funding; Hanita Daud: Validation, Writing-review and editing, Data curation, Supervision; Sani Isah Abba: Writing-review and editing, Investigation; Ahmad Abubakar Suleiman: Conceptualization, Formal analysis, Writing original draft, Software, Investigation, Methodology, Supervision. All authors have read and agreed to the published version of the manuscript.

## Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

## Conflict of interest

The authors declare no conflict of interests.

## References

1. A. I. Ishaq, A. A. Abiodun, A. A. Suleiman, A. Usman, A. S. Mohammed, M. Tasiu, Modelling Nigerian inflation rates from January 2003 to June 2023 using newly developed inverse power Chi-Square distribution, In: *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)*, 2023, 644–651. http://doi.org/10.1109/ICDABI60145.2023.10629442
2. A. A. Suleiman, H. Daud, M. Othman, A. Husin, A. I. Ishaq, R. Sokkalingam, et al., Forecasting the southeast Asian currencies against the British pound sterling using probability distributions, *J. Data Sci. Insights*, **1** (2023), 31–51.
3. M. A. Islam, M. Z. H. Majumder, M. S. Miah, S. Jannaty, Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction, *Comput. Biol. Med.*, **176** (2024), 108432. https://doi.org/10.1016/j.compbiomed.2024.108432
4. F. Bouchama, M. Kamal, Enhancing cyber threat detection through machine learning-based behavioral modeling of network traffic patterns, *Int. J. Business Intell. Big Data Analyt.*, **4** (2021), 1–9.
5. S. Ahmad, S. Jha, A. Alam, M. Yaseen, H. A. M. Abdeljaber, A novel AI-based stock market prediction using machine learning algorithm, *Sci. Prog.*, **2022** (2022), 4808088. https://doi.org/10.1155/2022/4808088

6. A. U. USMAN, S. B. Abdullahi, J. Ran, Y. Liping, A. A. Suleiman, H. Daud, et al., Modeling the dynamic behaviors of bank account fraudsters using combined simultaneous game theory with neural networks, preprint paper, 2024. http://doi.org/10.21203/rs.3.rs-3928159/v1
7. F. Kamalov, Forecasting significant stock price changes using neural networks, *Comput. Applic.*, **32** (2020), 17655–17667. http://doi.org/10.1007/s00521-020-04942-3
8. U. Panitanarak, A. I. Ishaq, N. S. S. Singh, A. Usman, A. U. Usman, H. Daud, et al., Machine learning models in predicting failure times data using a novel version of the maxwell model, *Eur. J. Stat.*, **5** (2025), 1. http://doi.org/10.28924/ada/stat.5.1
9. K. S. Lomax, Business failures: Another example of the analysis of failure data, *J. Amer. Stat. Assoc.*, **268** (1954), 847–852. http://doi.org/10.1080/01621459.1954.10501239
10. A. S. Hassan, S. M. Assar, A. Shelbaia, Optimum step-stress accelerated life test plan for Lomax distribution with an adaptive type-II progressive hybrid censoring, *J. Adv. Math. Comput. Sci.*, **13** (2016), 1–19. https://doi.org/10.9734/BJMCS/2016/21964
11. C. M. Harris, The Pareto distribution as a queue service discipline, *Oper. Res.*, **16** (1968), 307–313. http://doi.org/10.1287/opre.16.2.307
12. G. Gaudet, Distribution of personal wealth in Britain, *Economic J.*, **88** (1978), 581–583. https://doi.org/10.2307/2232061
13. J. Chen, R. G. Addie, M. Zukerman, T. D. Neame, Performance evaluation of a queue fed by a Poisson Lomax burst process, *IEEE Commun. Lett.*, **19** (2014), 367–370. http://doi.org/10.1109/LCOMM.2014.2385083
14. M. C. Bryson, Heavy-tailed distributions: Properties and tests, *Technometrics*, **16** (1974), 61–68. http://doi.org/10.2307/1267493
15. A. A. Khalaf, M. Ibrahim, M. Khaleel, Different transformation methods of the Lomax distribution: A review, *Iraqi Stat. J.*, **1** (2024), 41–52. http://doi.org/10.62933/6w437q74
16. J. K. Pokharel, G. Aryal, N. Khanal, C. P. Tsokos, Probability distributions for modeling stock market returns-an empirical inquiry, *Int. J. Financial Stud.*, **12** (2024), 43. https://doi.org/10.3390/ijfs12020043
17. V. B. V. Nagarjuna, R. V. Vardhan, C. Chesneau, Nadarajah-Haghighi Lomax distribution and its applications, *Math. Comput. Appl*, **27** (2022), 30. http://doi.org/10.3390/mca27020030
18. A. J. Lemonte, G. M. Cordeiro, An extended Lomax distribution, *Statistics*, **47** (2013), 800–816. http://doi.org/10.1080/02331888.2011.568119
19. A. H. El-Bassiouny, N. F. Abdo, H. S. Shahen, Exponential lomax distribution, *Int. J. Comput. Appl.*, **121** (2015), 24–29. http://doi.org/10.5120/21602-4713
20. N. M. Kilany, Weighted lomax distribution, *SpringerPlus*, **5** (2016), 1862. http://doi.org/10.1186/s40064-016-3489-2
21. M. H. Tahir, M. A. Hussain, G. M. Cordeiro, G. G. Hamedani, M. Mansoor, M. Zubair, The Gumbel-Lomax distribution: Properties and applications, *J. Stat. Theory Appl.*, **15** (2016), 61–79. http://doi.org/10.2991/jsta.2016.15.1.6
22. A. A. Abiodun, A. I. Ishaq, On Maxwell-Lomax distribution: Properties and applications, *Arab J. Basic Appl. Sci.*, **29** (2022), 221–232. http://doi.org/10.1080/25765299.2022.2093033

23. G. M. Cordeiro, E. E. E. Ortega, B. V. Popović, The gamma-Lomax distribution, *J. Stat. Comput. Simul.*, **85** (2015), 305–319. http://doi.org/10.1080/00949655.2013.822869

24. B. Alnssyan, The modified-Lomax distribution: Properties, estimation methods, and application, *Symmetry*, **15** (2023), 1367. http://doi.org/10.3390/sym15071367

25. A. I. Ishaq, A. A. Abiodun, Adewole, The Maxwell-Weibull distribution in modeling lifetime datasets, *Ann. Data Sci.*, **7** (2020), 639–662. http://doi.org/10.1007/s40745-020-00288-8

26. A. Shafiq, S. A. Lone, T. N. Sindhu, Y. El Khatib, Q. M. Al-Mdallal, T. Muhammad, A new modified Kies Fréchet distribution: Applications of mortality rate of Covid-19, *Results Phys.*, **28** (2021), 104638. http://doi.org/10.1016/j.rinp.2021.104638

27. A. Z. Afify, M. Nassar, D. Kumar, G. M. Cordeiro, A new unit distribution: Properties, inference, and applications, *Elect. J. Appl. Stat. Anal.*, **15** (2022), 438–462. http://doi.org/10.1285/i20705948v15n2p438

28. H. M. Alshanbari, A. M. Gemeay, A. H. El-Bagoury, S. K. Khosa, E. H. Hafez, A. H. Muse, A novel extension of Fréchet distribution: Application on real data and simulation, *Alex. Eng. J.*, **61** (2022), 7917–7938. http://doi.org/10.1016/j.aej.2022.01.013

29. E. Hossam, A. T. Abdulrahman, A. M. Gemeay, N. Alshammari, E. Alshawarbeh, N. K. Mashaqbah, A novel extension of gumbel distribution: Statistical inference with covid-19 application, *Alex. Eng. J.*, **61** (2022), 8823–8842. http://doi.org/10.1016/j.aej.2022.01.071

30. A. I. Ishaq, A. Usman, M. Tasi'u, A. A. Suleiman, A. G. Ahmad, A new odd F-Weibull distribution: properties and application of the monthly Nigerian naira to British pound exchange rate data, In: *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*. 2022, 326–332. http://doi.org/10.1109/ICDABI56818.2022.10041527

31. Y. L. Oh, F. P. Lim, C. Y. Chen, W. S. Ling, Y. F. Loh, Exponentiated Weibull Burr type X distribution's properties and its applications, *Elect. J. Appl. Stat. Anal.*, **15** (2022), 553–573. http://doi.org/10.1285/i20705948v15n3p553

32. E. A. El-Sherpieny, H. Z. Muhammed, E. M. Almetwally, A new inverse Rayleigh distribution with applications of COVID-19 data: Properties, estimation methods and censored sample, *Elect. J. Appl. Stat. Anal.*, **16** (2023), 449–472. http://doi.org/10.1285/i20705948v16n2p449

33. A. A. Suleiman, H. Daud, N. S. S. Singh, A. I. Ishaq, M. Othman, A new odd beta prime-burr X distribution with applications to petroleum rock sample data and COVID-19 mortality rate, *Data*, **8** (2023), 143. http://doi.org/10.3390/data8090143

34. N. Alotaibi, I. Elbatal, M. Shrahili, A. S. Al-Moisheer, M. Elgarhy, E. M. Almetwally, Statistical inference for the Kavya-Manoharan Kumaraswamy model under ranked set sampling with applications, *Symmetry*, **15** (2023), 587. http://doi.org/10.3390/sym15030587

35. M. Alqawba, Y. Altayab, S. M. Zaidi, A. Z. Afify, The extended Kumaraswamy generated family: Properties, inference and applications in applied fields, *Elect. J. Appl. Stat. Anal.*, **16** (2023), 740–763. http://doi.org/10.1285/i20705948v16n3p740

36. A. A. Suleiman, H. Daud, N. S. S. Singh, M. Othman, A. A. Ishaq, R. Sokkalingam, A novel odd beta prime-logistic distribution: Desirable mathematical properties and applications to engineering and environmental data, *Sustainability*, **15** (2023), 10239. http://doi.org/10.3390/su151310239

37. G. W. Liyanage, M. Gabanakgosi, B. Oluyede, Generalized Topp-Leone-G power series class of distributions: properties and applications, *Elect. J. Appl. Stat. Anal.*, **16** (2023), 564–583. http://doi.org/10.1285/i20705948v16n3p564

38. A. I. Ishaq, U. Panitanarak, A. A. Abiodun, A. A. Suleiman, H. Daud, The generalized odd maxwell-kumaraswamy distribution: Its properties and applications, *Contemp. Math.*, **5** (2024), 711–742. http://doi.org/10.37256/cm.5120242888

39. A. A. Suleiman, H. Daud, A. I. Ishaq, M. Kayid, R. Sokkalingam, R. Y. Hamed, et al., A new Weibull distribution for modeling complex biomedical data, *J. Rad. Res. Appl. Sci.*, **17** (2024), 101190. http://doi.org/10.1016/j.jrras.2024.101190

40. H. Daud, A. S. Mohammed, A. I. Ishaq, B. Abba, Y. Zakari, J. Abdullahi, et al., Modeling and prediction of exchange rates using Topp-Leone Burr type X, machine learning and deep learning models, *Eur. J. Stat.*, **4** (2024), 11. http://doi.org/10.28924/ada/stat.4.11

41. H. Daud, A. A. Suleiman, A. I. Ishaq, N. Alsadat, M. Elgarhy, M. A. Usman, et al., A new extension of the Gumbel distribution with biomedical data analysis, *J. Rad. Res. Appl. Sci.*, **17** (2024), 101055. http://doi.org/10.1016/j.jrras.2024.101055

42. U. Panitanarak, A. I. Ishaq, A. A. Abiodun, H. Daud, A. A. Suleiman, A new Maxwell-Log logistic distribution and its applications for mortality rate data, *J. Niger. Soc. Phys. Sci.*, **7** (2025), 1976–1976. http://doi.org/10.46481/jnsps.2025.1976

43. R. C. Gupta, P. L. Gupta, R. D. Gupta, Modeling failure time data by Lehman alternatives, *Commun. Stat.-Theory Meth.*, **27** (1998), 887–904. http://doi.org/10.1080/03610929808832134

44. D. W. Hosmer, S. Lemeshow, S. May, Applied survival analysis, In: *Wiley Series in Probability and Statistics*, **60** (2008).

45. A. Ganguly, D. Mitra, N. Balakrishnan, D. A. Kundu, A flexible model based on piecewise linear approximation for the analysis of left truncated right censored data with covariates, and applications to Worcester Heart Attack Study data and Channing House data, *Stat. Med.*, **43** (2023), 233–255. http://doi.org/ 10.1002/sim.9954

46. I. B. Abdul-Moniem, H. F. Abdel-Hameed, On exponentiated Lomax distribution, *Int. J. Math. Arch.*, **3** (2012), 1–7.

47. E. A. Rady, W. A. Hassanein, T. A. Elhaddad, The power Lomax distribution with an application to bladder cancer data, *SpringerPlus*, **5** (2016), 1838. http://doi.org/10.1186/s40064-016-3464-y

48. P. E. Oguntunde, M. A. Khaleel, H. I. Okagbue, O. A. Odetunmibi, The Topp-Leone Lomax (TLLo) distribution with applications to airbone communication transceiver dataset, *Wirel. Per. Commun.*, **109** (2019), 349–360. http://doi.org/10.1007/s11277-019-06568-8

49. A. A. Abiodun, A. I. Ishaq, On Maxwell-Lomax distribution: Properties and applications, *Arab J. Basic Appl. Sci.*, **29** (2022), 221–232. http://doi.org/10.1080/25765299.2022.2093033

50. M. Shabbir, A. Riaz, H. Gull, Rayleigh Lomax distribution, *J. Middle East North Africa Sci.*, **4** (2018), 1–4.

51. P. Embrechts, C. Kluppelberg, T. Mikosch, Modelling extremal events, *British Actuar. J.*, **5** (1999), 465–465.

52. G. P. Patil, M. T. Boswell, S. W. Joshi, M. V. Ratnaparkhi, Dictionary and classified bibliography of statistical distributions in scientific work, *Int. Co-operative Publishing House*, 1985.

53. C. Kleiber, A guide to the Dagum distributions, In: *Modeling Income Distributions and Lorenz Curves. Economic Studies in Equality, Social Exclusion and Well-Being, Springer*, 2008, 97–117. https://doi.org/10.1007/978-0-387-72796-7_6

54. T. Nombebe, J. Allison, L. Santana, J. Visagie, On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques, *Computation*, **11** (2023), 44. http://doi.org/10.3390/computation11030044 .

55. P. Supsermpol, S. Thajchayapong, N. Chiadamrong, Predicting financial performance for listed companies in Thailand during the transition period: A class-based approach using logistic regression and random forest algorithm, *J. Open Innov.: Technol. Market Complex.*, **9** (2023), 100–130. http://doi.org/10.1016/j.joitmc.2023.100130

56. A. Kurani, P. Doshi, A. Vakharia, M. Shah, A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting, *Ann. Data Sci.*, **10** (2023), 183–208. http://doi.org/10.1007/s40745-021-00344-x

57. C. J. Huang, D. X. Yang, Y. T. Chuang, Application of wrapper approach and composite classifier to the stock trend prediction, *Expert Syst. Appl.*, **34** (2008), 2870–2878. http://doi.org/10.1016/j.eswa.2007.05.035

58. C. Sattarhoff, T. Lux, Thomas, Forecasting the variability of stock index returns with the multifractal random walk model for realized volatilities, *Int. J. Forec.*, **39** (2023), 1678–1697. http://doi.org/10.1016/j.ijforecast.2022.08.009

59. M. Chen, Q. Liu, S. Chen, Y. Liu, C. H. Zhang, R. Liu, XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system, *IEEE Access*, **7** (2019), 13149–13158. http://doi.org/10.1109/ACCESS.2019.2893448

60. M. Heydarian, T. E. Doyle, R. Samavi, MLCM: Multi-label confusion matrix, *IEEE Access*, **10** (2022), 19083–19095. http://doi.org/10.1109/ACCESS.2022.3151048

61. A. U. Usman, S. B. Abdullahi, Y. L. Liping, B. Alghofaily, A. S. Almasoud, A. Rehman, Financial fraud detection using value-at-risk with machine learning in skewed data, *IEEE Access*, **12** (2024), 64285–64299. http://doi.org/10.1109/ACCESS.2024.3393154

62. S. Uddin, H. Lu, Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data, *PLos One*, **19** (2024), e0301541. https://doi.org/10.1371/journal.pone.0301541


AIMS Press