Heliyon 10 (2024) e28270

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Probabilistic assessment of wind power plant energy potential through a copula-deep learning approach in decision trees

Kübra Nur Şahin[a], Muhammed Sutcu[b,*]

[a] Abdullah Gül University, Faculty of Engineering, Industrial Engineering, Department, Kayseri, Turkiye
[b] Gulf University for Science and Technology, College of Engineering and Architecture, Engineering Management Department, Mishref, Kuwait

ARTICLE INFO

Keywords:
Sustainable energy
Decision models
Information theory
Copulas
Deep learning

ABSTRACT

In the face of environmental degradation and diminished energy resources, there is an urgent need for clean, affordable, and sustainable energy solutions, which highlights the importance of wind energy. In the global transition to renewable energy sources, wind power has emerged as a key player that is in line with the Paris Agreement, the Net Zero Target by 2050, and the UN 2030 Goals, especially SDG-7. It is critical to consider the variable and intermittent nature of wind to efficiently harness wind energy and evaluate its potential. Nonetheless, since wind energy is inherently variable and intermittent, a comprehensive assessment of a prospective site's wind power generation potential is required. This analysis is crucial for stakeholders and policymakers to make well-informed decisions because it helps them assess financial risks and choose the best locations for wind power plant installations. In this study, we introduce a framework based on Copula-Deep Learning within the context of decision trees. The main objective is to enhance the assessment of the wind power potential of a site by exploiting the intricate and non-linear dependencies among meteorological variables through the fusion of copulas and deep learning techniques. An empirical study was carried out using wind power plant data from Turkey. This dataset includes hourly power output measurements as well as comprehensive meteorological data for 2021. The results show that acknowledging and addressing the non-independence of variables through innovative frameworks like the Copula-LSTM based decision tree approach can significantly improve the accuracy and reliability of wind power plant potential assessment and analysis in other real-world data scenarios. The implications of this research extend beyond wind energy to inform decision-making processes critical for a sustainable energy future.

1. Introduction

In light of mounting concerns over environmental degradation and the finite nature of traditional energy sources, the quest for clean, affordable, and sustainable energy has become an urgent global priority. At the forefront of this imperative is the energy sector, the largest contributor to global greenhouse gas emissions, accounting for approximately 73.2% (see Fig. 1

Given the central role of the energy sector, it is clear that transformative action is essential to reduce its negative impact on the environment. Strategies aimed at reducing emissions, switching to renewable energy sources, and improving energy efficiency are of paramount importance. In addition, encouraging innovation and implementing sound policies can further advance the energy sector

* Corresponding author.
E-mail address: sutcu.m@gust.edu.kw (M. Sutcu).

https://doi.org/10.1016/j.heliyon.2024.e28270
Received 29 October 2023; Received in revised form 6 March 2024; Accepted 14 March 2024
Available online 27 March 2024

towards a more sustainable future.

Wind energy has garnered increasing attention as a viable solution in this endeavor, especially given the pressing need to combat climate change and transition towards renewable energy sources. As the world seeks to reduce its reliance on fossil fuels and embrace cleaner alternatives, wind power has emerged as a key player in the global shift towards renewable energy. Its significance is underscored by its crucial role in helping nations meet ambitious targets set forth in milestone agreements such as the Paris Agreement, the Net Zero Target by 2050, and the UN 2030 Sustainable Development Goals (SDGs), notably SDG-7, which focuses on ensuring access to affordable, reliable, sustainable, and modern energy for all.

Increasing the proportion of wind energy in the energy production mix and effectively utilizing this resource, wind power potential assessment is a fundamental step. Due to the variable and intermittent nature of wind energy, it is essential to assess the potential of a site for wind power generation, which offers crucial analysis for decision-makers in terms of financial risk evaluation [2]. This evaluation is instrumental in selecting optimal sites for wind power plants and determining their financial and technical feasibility. Therefore, comprehensive and reliable wind power assessments are indispensable for the successful development and operation of wind power plants [3]. These assessments not only ensure the efficient utilization of wind resources but also contribute to the sustainable growth of wind energy as a key component of the global renewable energy landscape. With these tools analyzing the output characteristics of wind power emerges as a strategic and effective response to the inherent uncertainties associated with wind energy. These tools involve a thorough examination of the patterns and behaviors in power generation, empowering stakeholders to devise resilient strategies for risk mitigation and enhanced optimization of energy production [4–7].

The assessment of wind power output (P_w) stands as a critical determinant for evaluating the technical potential and financial feasibility of wind power plants [8]. Furthermore, recognizing that the fluctuation in wind power output is attributed to changes in wind speed, precise determination of the distribution and fluctuation characteristics of wind speed within the wind power plant becomes instrumental for a comprehensive understanding of its overall output [9–11]. When it comes to assessing the potential of wind energy, several techniques are commonly employed, including statistical time series models, distribution fitting methods such as the Weibull distribution, and various estimation models [12–16].

Given the unpredictable nature of the factors influencing wind energy production, it becomes essential to employ multivariate analysis techniques. Through multivariate analysis, which accounts for the interactions among diverse factors, a more comprehensive understanding of wind patterns and their behavior can be attained [17]. These advanced analytical methods not only enhance the

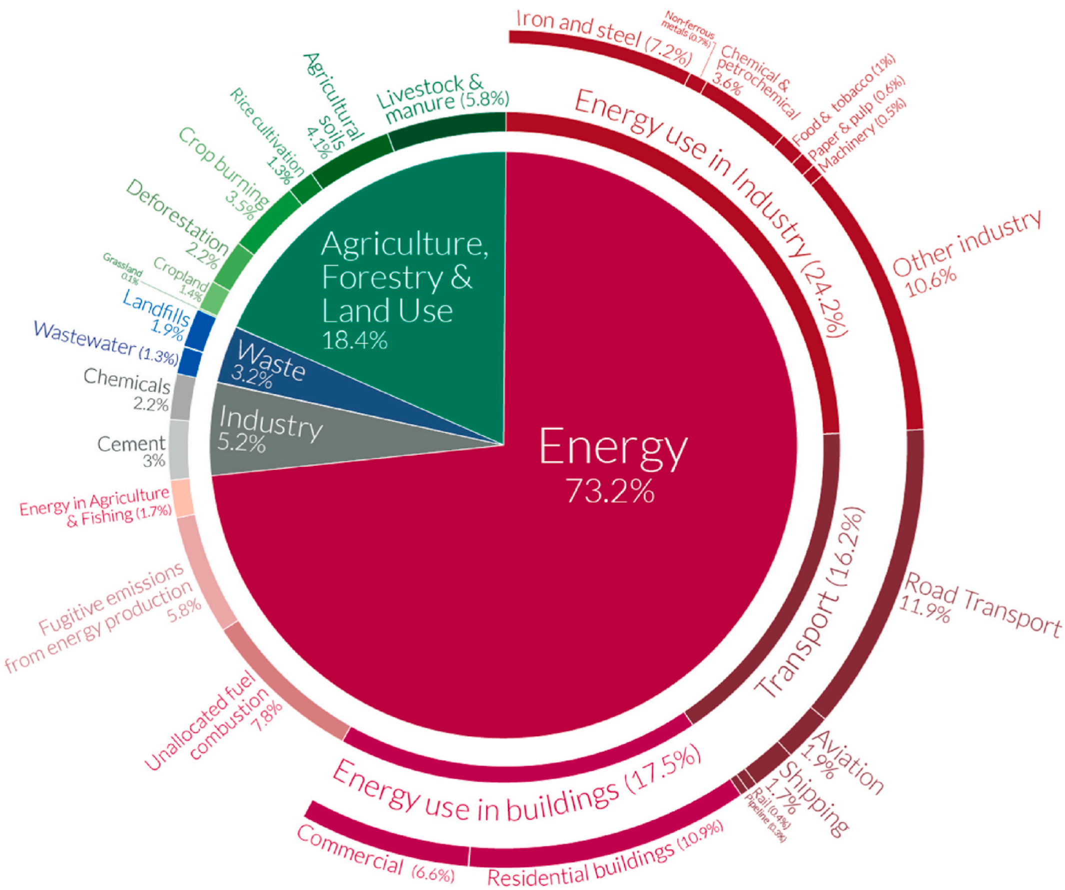

Fig. 1. Global greenhouse gas emissions by sector. Adopted from [1].

accuracy of wind power potential assessments but also contribute to informed decision-making processes in the development and operation of wind energy projects.

In the analysis of wind power plant energy potential in the literature, wind-related meteorological uncertainties are often treated as completely independent variables, and the joint probability distributions of these uncertainties are constructed using approaches such as the isotropic Gaussian distribution [18], anisotropic Gaussian distribution [19], and anisotropic logarithmic Gaussian distribution [20], among others. However, these approaches have limitations because meteorological uncertainties are not inherently independent. While the techniques employed in the literature have provided valuable insights, they may not fully capture the complex dependencies and non-linear interactions present in wind data, leading to limitations in accuracy and reliability. Consequently, there is a need for more advanced methods that can better account for the interconnected nature of meteorological variables, improving the accuracy of wind power potential assessments and enhancing the reliability of wind energy projects.

Copulas, among multivariate techniques, have garnered significant attention in recent research due to their ability to model intricate dependence structures. Copulas provide a flexible, modular possibility for constructing multivariate distributions that allows for the separation between the specification for the marginals and the specification of the dependence structure. Their versatility has led to widespread applications in diverse fields such as finance, risk management, rainfall analysis, drought analysis, and energy [21–24].

Concurrently, in multivariate modeling and analysis, the "dependence tree" approach is utilized to evaluate dependence measures hierarchically [25–28]. However, integrating multivariate dependencies into a decision tree is generally computationally intensive, as generating conditional distributions for each branch poses a substantial challenge [29]. This computational complexity is further exacerbated when the marginal distributions of the variables belong to different families, making the generation of conditional distributions even more demanding. In such cases, specifying marginal distributions of each variable and employing copulas to compute joint distributions is a practical approach to describing the hierarchical structure of multivariate dependence. Despite the potential benefits, there has been limited exploration of copula applications in discrete probability trees within decision analysis [29].

Furthermore, in the literature, the application of multicriteria decision-making (MCDM) models, notably AHP (Analytical Hierarchy Process), TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution), PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation), ELECTRE (Elimination Et. Choice Translating Reality), and Fuzzy has been widespread for assessment of wind energy site. Studies by Emeksiz and Demirci et al. [30] in Turkey, Şahin et al. [31] in the Netherlands, Solangi et al. [32] in Pakistan, and Islam et al. [33] in Bangladesh exemplify the utilization of these methods to assess various criteria for identifying optimal locations for wind power plants. However, alongside overseeing interdependencies among critical factors, the literature underscores challenges in MCDM models assuming linear relationships between variables. This assumption has the potential to lead to suboptimal site assessment, emphasizing the need for advanced methodologies that can capture the non-linear dynamics inherent in wind energy production. Addressing these limitations is imperative to ensure accurate and robust decision-making processes in selecting wind energy sites.

The copula-based dependent tree approach can be an effective and practical approach to statistically representing the multivariate dependence structure of wind data. The most important advantages of this approach are that it is copula-based and each marginal distribution is added to the model separately as a decision tree. The decision tree also allows the relationship between model parameters to be integrated into the model in a binary fashion. In addition, although the distributions of the variables are continuous, the decision tree makes it easier for the model to work because they can be integrated into the model in a discrete form. Meanwhile, in order to estimate the wind power potential of a site, there is a need to develop a model to generate wind farm power outputs for each branch of the dependent tree. Therefore, in that study, a long short-term memory (LSTM) architecture was adopted for this purpose.

Using multivariate analysis, it is possible to gain a better understanding of the relationships among meteorological variables. Despite the intricate nature of interactions between these variables, it is imperative to determine the joint distributions of the variables and their conditional distributions, as well as to estimate the power output values, for the assessment of wind power plant potential. Nevertheless, earlier methods have exhibited limitations in terms of their effectiveness and computational complexity. To address this challenge, we have developed a copula-deep learning-based multivariate modeling framework. The primary objective of this study, therefore, is to present a comprehensive and robust framework for the evaluation of wind power plant potential, capable of capturing complex dependencies and patterns in the data while leveraging the capabilities of deep learning. By employing this framework, energy stakeholders will be empowered to make more informed decisions regarding the planning, design, and operation of wind energy systems.

The main contributions of the study can be summarized as follows: (1) this study challenges the prevalent practice of assuming independence among meteorological variables. This assumption has traditionally been employed due to the complexity and computational burden associated with defining a multivariate joint distribution. In contrast, the research adopts a copula-based dependence structure, which allows for the incorporation of intricate and nonlinear dependencies and patterns that are often overlooked in conventional analyses. By doing so, the study enriches the understanding of the interrelationships among meteorological variables, offering a more nuanced perspective on their collective influence, (2) introducing an innovative approach that integrates Copula and Deep Learning within decision trees for assessing the energy potential of wind power plants at specific locations. This fusion methodology represents a novel advancement in the field, as it combines the strengths of Copula-based modeling in capturing complex dependencies with the powerful predictive capabilities of Deep Learning algorithms. By applying this approach, the study aims to provide a more accurate and comprehensive assessment of wind power plant potential, considering the inherent uncertainties associated with meteorological variables.; and (3 the findings of this research are expected to have significant practical implications for the wind energy sector. By offering insights into the techno-economic feasibility of wind power projects, the study can inform decision-making processes for policymakers and stakeholders involved in wind power plants. Furthermore, the research outcomes can serve as a valuable decision support mechanism, enabling stakeholders to make informed choices regarding the planning, design, and operation of wind energy systems. Ultimately, the study aims to contribute to the effective utilization of wind energy as a sustainable and abundant resource. In summary, the research makes substantial contributions to the field by challenging existing assumptions, introducing innovative methodologies, and offering practical insights for the advancement of wind energy technologies and practices.

The rest of the paper is organized as follows: Section 2 discusses the Copula-LSTM based decision tree framework. The implementation of the proposed framework and the performance analysis are provided in Section 3. Finally, Section 4 concludes this paper.

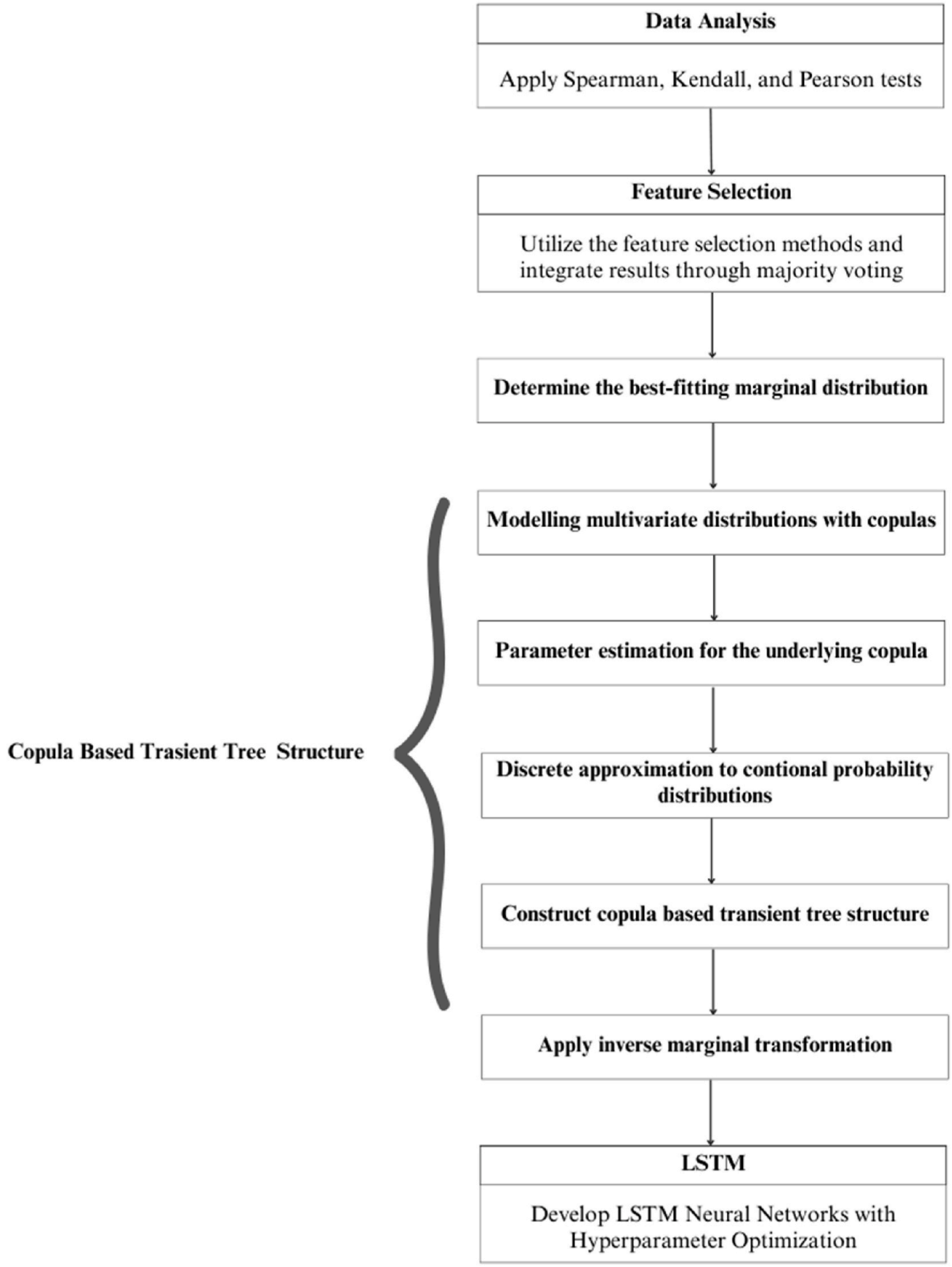

Fig. 2. The flowchart of proposed methodology.

2. Methodology

The main steps of the proposed methodology are data analysis, feature selection, identifying the most suitable marginal probability distributions, modeling multivariate distributions with copulas, parameter estimation for the underlying copula, copula-based transient tree structure, inverse marginal transformation, and LSTM neural networks. The flowchart of the proposed methodology is shown in Fig. 2.

2.1. Data analysis

The correlations between meteorological conditions are investigated to gain a comprehensive understanding of their interactions. Tests of Spearman, Kendall [34] and Pearson [35] are applied to measure the dependencies between the variables; temperature, humidity, pressure, wind speed, wind gust, and wind bearing.

The relationships between meteorological conditions are investigated through a correlation analysis, and both linear and monotonic dependencies are discovered. Temperature and humidity, as well as temperature and pressure, have negative correlations, highlighting their interconnectedness. Furthermore, the positive monotonic relationship between wind speed and wind gust highlights the consistency of their variation.

Significant relationships emerged from the analysis. Notably, a negative correlation was observed between temperature and humidity (Spearman's $\rho = -0.620$, $p < 0.05$), indicating their interconnectedness. Similarly, a negative correlation was found between temperature and pressure (Kendall's $\tau = -0.454$, $p < 0.05$), highlighting the interplay between these meteorological factors.

Additionally, a positive monotonic relationship surfaced between wind speed and wind gust (Pearson's $r = 0.907$, $p < 0.05$), suggesting a consistent variation in wind speed corresponding to changes in wind gust.

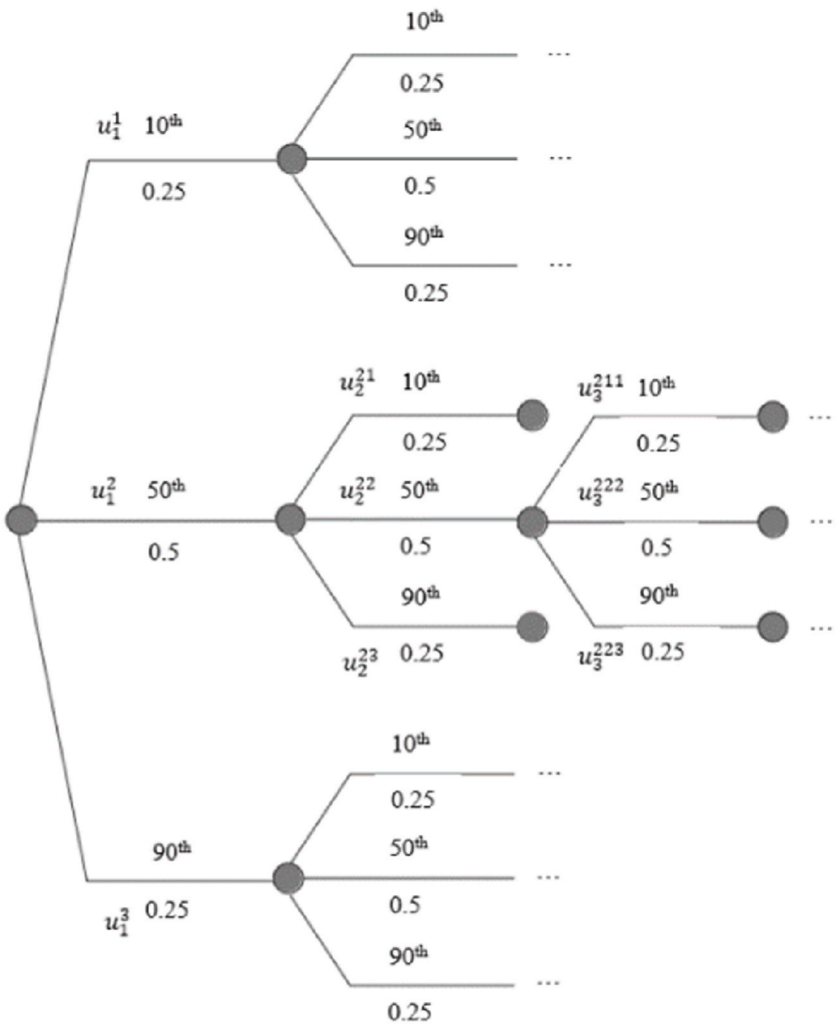

Fig. 3. Copula-based tree structure

To deepen the interpretation of these correlations, effect size measures were calculated. For example, Cohen's d for the negative correlation between temperature and humidity yielded d = −0.612, indicating a moderate effect size. Likewise, for the correlation between temperature and pressure, Kendall's effect size τ-b was calculated as τ-b = −0.616, signifying a moderate effect.

These effect size measures enhance the understanding of the correlations, providing valuable insights into the magnitude and practical significance of the observed meteorological dependencies. The findings contribute to the robustness and applicability of the research by offering nuanced insights into the relationships among meteorological variables. However, for a more comprehensive exploration of complex dependence structures, copulas provide a versatile framework. They excel in modeling multivariate dependencies beyond linear correlations, capturing intricate patterns such as tail dependencies, asymmetric relations, and non-monotonic behaviors. Incorporating copulas into the analysis further refines the understanding of meteorological dependencies, acknowledging the nuanced nature of these relationships and contributing to the depth and applicability of the research findings.

2.2. Feature selection and identifying the most suitable marginal probability distributions

In the context of the dependent decision tree approach, feature selection is essential to identify the most relevant variables that significantly contribute since it helps to reduce the complexity of the model and enhance its performance. When the number of variables increases, the depth of the decision tree grows linearly, but the number of branches grows exponentially with the number of features.

Forward selection, backward elimination, lasso regularization, decision tree importance, recursive feature elimination, and mutual info techniques are used in this study to reduce the computational complexity of the model and determine the most informative set of features. Feature selection through various techniques followed by a voting approach is used as a tool to improve model performance. The voting approach aggregates the selection of the individual feature selection techniques. The ensemble of various feature selection techniques with majority voting helps in making robust predictions, reducing overfitting, and improving the model's generalization performance. With this approach, the set of selected features that is likely to capture relevant information while minimizing computational burden is determined.

In addition, to accurately capture the inherent characteristics of a variable, deriving the best-fitted marginal probability distribution is crucial. The evaluation metrics mainly used to measure the fitting performance of the theoretical distribution are the Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Kolmogorov-Smirnov, and Chi-square tests. In that study, MSE is used to evaluate the fitting performance of the distribution.

2.3. Modelling multivariate distributions with copulas

In recent years, there has been growing interest in copula models to examine the structure of interdependence between variables in a variety of research fields such as energy, medical science, banking, and economics. Copulas are powerful statistical tools to develop a dependent structure of multivariate random variables where the marginal distributions of random variables are independently determined. The term "Copula" was first introduced by Sklar, in 1959 [36]. The proposed idea is that a joint distribution can be expressed as a function of the marginal distributions [36]. Variations in the correlation of the variables in different parts of marginal distributions can be explained by copula models [37]. Mathematically, it can be expressed as

$$C(F_1(x_1), F_2(x_2), .., F_n(x_n)) = F(x_1, x_2, \ldots, x_n) \tag{1}$$

and,

$$F(X_1) = U_1, F(X_2) = U_2, .., F(X_n) = U_n \tag{2}$$

Thus,

$$C(u) = C(u_1, u_2, \ldots, u_n) = P(U_1 \le u_1, U_2 \le u_2, \ldots, U_n \le U_n) \tag{3}$$

where $x = (x_1, x_2, .., x_n) \in R^n$ and $U = (U_1, U_2, .., U_n)$ is a n-dimensional random vector with $U_i \sim Unif(0,1), \forall i = 1, 2, \ldots, n$. This is well-known Sklar's Theorem (1) enables us to express any multivariate distribution in terms of its marginal distributions and a copula that captures dependencies between the variables. If X is a continuous random vector, then the copula C is unique [36].

With the cumulative distribution function of a joint distribution, the joint probability distribution function (PDF) can be derived as follows:

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n} = \frac{\partial^n C(F_1(x_1), F_2(x_2), .., F_n(x_n))}{\partial F_1(x_1) \partial F_2(x_2) .. \partial F_n(x_n)} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \ldots \frac{\partial F_n(x_n)}{\partial x_n} = c(F_1(x_1), F_2(x_2), .., F_n(x_n)) \prod_{i=1}^{n} f_i(x_i) \tag{4}$$

where $f_i(x_i)$ denotes the marginal pdf of X_i and $c(F_1(x_1), F_2(x_2), .., F_n(x_n))$ denotes the copula density that is obtained from the partial derivative of the copula. Thus, any joint probability density function can be written as the product of its marginal probability density functions. Detailed information about copulas in general can be found in the books by Nelsen [38].

In the literature, numerous families of copulas emphasizing various distributional properties have been proposed, such as the Archimedean, Elliptical, Placket families. These copula families include a wide range of patterns of tail dependences and various types of asymmetries. The well-known elliptical copulas are the Gaussian and Student's t copula which represent symmetrical tail dependencies. In addition to elliptical copulas, the most known Archimedean copulas are Frank, Gumbel, and Clayton copulas. These types of copulas capture various types of tail dependence. The Frank is symmetric with no tail dependence [39], while the Gumbel copula exhibits upper-tail dependence [40], the Clayton copula has lower-tail dependence [41]. Table 1 illustrates the mathematical descriptions of the copula families used in this study.

The characteristics of copula theory allow for the construction of decision tree structures that incorporate the dependence structure obtained by copulas and marginal densities. A transient probability tree can be created based on copulas, which can capture various types of multivariate dependencies. After determining the marginal distributions of the variables, the conditional distributions of each branch of the dependent decision tree can be obtained.

The generation of random uniform variables from multivariate copulas is performed to build a transient probability tree. This process can be called sampling. One approach for sampling a uniform random variable $(u_1, u_2, \ldots, u_n)$ from n-dimensional normal copula is to use Cholesky decomposition for the generation of the uniform variables [42]. After the Cholesky factorization process is applied to the correlation matrix Σ_n; i.e., $\Sigma_n = AA'$, we can generate n independent $z = (z_1, z_2, \ldots, z_n)$ and $x = (x_1, x_2, \ldots, x_n)$ which is Az [43]. N-dimensional t-copula can be obtained similarly with parameters Σ_n and d. For Archimedean copulas, as a first step to generate two-dimensional random variable (u_1, u_2), two independent uniform random variables u_1 and w will be generated. To obtain u_2 value using the quasi-inverse function of conditional distribution $P(U_2 \leq u_2 | U_1 = u_1)$ that is $u_2 = C_{2|1}^{-1}(w|u_1)$. We calculate the conditional copula as following,

$$C_{2|1}(u_2|u_1) = \frac{\partial C_2(u_1, u_2)}{\partial u_1} \tag{5}$$

For multivariate case (n-dimensional, $n \geq 3$) u_1 and $w_1, w_2, \ldots, w_{n-1}$ are generated as bivariate case and $u_2, u_3, \ldots, u_n$ random variables are obtained by

$$\begin{aligned} u_2 &= C_{2|1}^{-1}(w_1|u_1) \\ u_3 &= C_{3|1,2}^{-1}(w_2|u_1, u_2), \\ u_n &= C_{n|1,2,\ldots,n-1}^{-1}(w_{n-1}|u_1, u_2, \ldots, u_{n-1}), \end{aligned}$$

where

$$C_{n|1,2,\ldots,n-1}(u_n|u_1, u_2, \ldots, u_{n-1}) = P(U_n \leq u_n | U_1 = u_1, U_2 = u_2, \ldots, U_{n-1} = u_{n-1}) = \frac{\left(\partial^{n-1} C_n(u_1, u_2, \ldots, u_n)\right) / (\partial u_1 \partial u_2 \ldots \partial u_{n-1})}{\left(\partial^{n-1} C_{n-1}(u_1, u_2, \ldots, u_{n-1})\right) / (\partial u_1 \partial u_2 \ldots \partial u_{n-1})} \tag{6}$$

Let considers the bivariate Archimedean copula case. For any continuous random variables X_1 and X_2 to generate u_2 given that $u_1 = P(X_1 \leq x_1) = \alpha_1$, we need to obtain the conditional distribution of X_2 given $X_1 = x_1$ from the partial derivative of the copula function. For each given percentile $P(X_2 \leq x_2 | X_1 = x_1) = \alpha_2$ and $\partial C(u_1, u_2)/\partial u_1 = C_{u_1}(u_2) = \alpha_2$ then, $u_2 = C_{u_1}^{-1}(\alpha_2)$. Thus, with unconditional and conditional percentile values obtained by using copula functions, a transient probability tree structure can be built [29].

Table 1
Formulas of elliptical and archimedean copulas.

Family	Mathematical Expression	Partial Derivative of the copula w.r.t $\boldsymbol{u_1}$	$\boldsymbol{u_2}$
Normal Copula	$C_n = (F_1(X_1), F_2(X_2)) = \varphi_r(\varphi^{-1}(F_1(X_1)), \varphi^{-1}(F_2(X_2)))$	$\varphi\left(\frac{\varphi^{-1}(u_2) - r\varphi^{-1}(u_1)}{\sqrt{1-r^2}}\right)$	$u_2 = \varphi(r\varphi^{-1}(\alpha_1) + \sqrt{1-r^2\varphi^{-1}(\alpha_2)})$
t-copula	$C_T = (F_1(X_1), F_2(X_2)) = t_{r,v}(t_v^{-1}(F_1(X_1)), t_v^{-1}(F_2(X_2)))$	$t_{r,v}\left(\sqrt{\frac{v+1}{v+(t_v^{-1}(u_1))^2}} \cdot \frac{t_v^{-1}(u_2) - rt_v^{-1}(u_1)}{\sqrt{1-r^2}}\right)$	$u_2 = t_v\left(rt_v^{-1}(\alpha_1) + \sqrt{1-r^2}\sqrt{\frac{v+(t_v^{-1}(u_1))^2}{v+1}} t_{v+1}^{-1}(\alpha_2)\right)$
Clayton Copula	$C_{Clayton}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$	$u_1^{-\theta-1}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta - 1}$	$u_2 = \left(\left(\alpha_2^{-\theta/1+\theta} - 1\right)\alpha_1^{-\theta} + 1\right)^{-1/\theta}$
Gumbel	$C_{gumbel}(u_1, u_2) = exp\left\{-[(-ln\, u_1)^\theta + (-ln\, u_2)^\theta]^{1/\theta}\right\}$	$\frac{\partial C}{\partial u_1}(u_1, u_2) = \frac{\varphi^{-1(1)}(c_2)}{\varphi^{-1(1)}(c_1)}$ where $c_1 = \varphi(u_1) = (-ln\,(u_1))^\theta$ $c_2 = \varphi(u_1) + \varphi(u_2)$ $= (-ln\,(u_1))^\theta + (-ln\,(u_2))^\theta$ and $\varphi^{-1(1)}(t) = -\frac{1}{\theta}e^{-t^{1/\theta}}\left(t^{1/\theta}\right)^{1-\theta}$	The value can be obtained by solving the given equations.
Frank	$C_{Frank}(u_1, u_2) = -\frac{1}{\theta}ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$	$\frac{(e^{-\theta u_2} - 1)(e^{-\theta u_1})}{(e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}$	$u_2 = -\frac{1}{\theta}ln\left(1 + \frac{(e^{-\theta} - 1)\alpha_2}{e^{-\theta\alpha_1} - (e^{-\theta\alpha_1} - 1)\alpha_2}\right)$

2.4. Parameter estimation for the underlying copula

Since the dependency structure is modeled by copulas, specific parameters of copulas should be estimated. The methods that are proposed to estimate copula parameters are the method of moments (MOM) [44] known as inversion of Kendal's τ, inversion of Spearman's rho, canonical maximum likelihood (CML), minimum distance method (MDM) [45–47] and maximum pseudo-likelihood (MLP) [48].

For a normal copula, the product moment correlation r should be estimated. To estimate the relationships between the product moment correlation r for the normal copula Spearman's rank order correlation ρ or Kendall's rank order correlation τ can be used [49].

$$r = 2\,\sin\left(\frac{\pi\rho}{6}\right) \text{ and } r = 2\,\sin\left(\frac{\pi\tau}{2}\right) \tag{7}$$

For the Archimedean family, using the MOM approach, the parameters are estimated using the relationship between the Kendall's τ and the generator function φ of the copula [44].

$$\tau = 1 + \int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt \tag{8}$$

The relationship between Kendall's rank order correlation τ and θ is summarized in Table 2 [38] (see).

2.5. Copula-based transient tree structure

Discrete approximations to conditional probability distributions with the underlying copula are used to construct a probability tree (Fig. 3). Discretization techniques are mostly used in decision analysis because using the true distribution is computationally costly. Using discretization techniques in the decision analysis approach enables us to estimate the expected value of the decision without any prior knowledge about the probability distribution function of any uncertainties. There are many methods for discretization in the literature such as McNamee-Celona Shortcut (MCS) [50], Extended Swanson-Megill (EMS) [51], Extended Pearson-Tukey (EPT) [52], Zaino-D'Errico "Improved" (ZDI) [53], Zaino-D'Errico-Tagichi (ZDT) [54] and Miller-Rice One Step (MRO) [55]. The mentioned methods all use the 50th percentile, and they differ in the distance to the mean and the probability values (Table 3. This study focuses on MCS, which is widely used.

The application of the MCS method at each layer of the tree is necessary for transient tree structures because the continuous random variables are required to be discretely represented. We use the 10th, 50th, and 90th percentiles, which correspond to realizations of 0.25, 0.5, and 0.25 to discretize each of the n features $X_1, X_2, \ldots, X_n$. The dependent uniform random variables that are used as transient probabilities in the tree are calculated using copula functions. For instance, the dependent uniform random variable u_2 is computed based on the unconditional percentile value u_1 and the conditional percentile value for the second layer α_2 that were obtained after the discretization of the second feature X_2 using the chosen copula function that is listed in Table 1. Iteratively using this computational process, a complete transient tree structure is generated. Since the MCS method is used in discretization, the underlying marginal distributions of the feature are not used in this step.

2.6. Inverse marginal transformation

After the construction of the copula-based transient tree structure, we proceed with the stepwise inverse transformation after obtaining estimates for the multivariate joint probabilities for each quantile combination, which allows us to investigate the dependence structure across various subsets of the complex data sets. The inverse CDF transformation converts the copula generated uniform u values to their original scales and distributions based on their marginal distributions. As a result, we obtain discrete approximations to the original variables. We construct a complete multivariate dependent tree structure by applying the inverse transformation to the transient tree, which provides a discrete approximation for the conditional distributions.

Table 2
The relation between Kendall's τ and θ for archimedean copula.

Copula Type	$\varphi(t)$	Kendall's Rank Order Correlation (τ)	Range of θ
Clayton	$(t^{-\theta}-1)/\theta$	$\frac{\theta}{\theta+2}$	$(-1,\infty)-[0]$
Frank	$-\log\left(\frac{exp(t\theta)-1}{exp(\theta)}\right)$	$1+4\theta^{-1}(D_1^*(\theta)-1)$	$(-\infty,\infty)-[0]$
Gumbel	$-\log(t)^{\theta}$	$\frac{\theta-1}{\theta}$	$[1,\infty)$

$D_1^* = \frac{1}{\theta}\int_0^{\theta} \frac{t}{e^t-1}$

Table 3
Discretization methods.

Discretization Method	Discretization Values	Probabilities
ESM	10th, 50th, 90th	0.3, 0.4, 0.3
MCS	10th, 50th, 90th	0.25, 0.5, 0.25
EPT	5th, 50th, 95th	0.185, 0.63, 0.185
ZDI	4.2nd, 50th, 95.8th	0.167, 0.667, 0.167
ZDT	11th, 50th, 89th	0.333, 0.333, 0.333
MRO	8.5th, 50th, 91.5th	0.248, 0.504, 0.248

To generate discrete approximations of $X_1, X_2, \ldots, X_n$, the inverse transformation needed to be applied to each level of $u_1, u_2, \ldots, u_k$. In situations where expressing conditional density is difficult and computationally expensive, the dependent tree approach is a very useful and practical representation of the conditional relationship between the variables.

2.7. Long short-term memory (LSTM) neural networks

LSTM is a special type of recurrent neural network that can handle the vanishing gradient problem of the recurrent neural network (RNN). The LSTM neural networks can capture long term dependencies and the tendencies of the sequential dataset [56]. Unlike the traditional RNN, LSTM networks consist of three layers; the input layer, the recurrent hidden layer, and the output layer (see Fig.4). The main objective of LSTM is to transmit and store information in the memory over a long period of time with the use of special units that are forgotten, input, update, and output gates. The series of gates is used by LSTM neural networks to control the information flow in a data sequence.

The forgotten gate (f_t) layer calculates the information that is needed to be forgotten. It takes the output of the previous layer (h_{t-1}) and the input (X_t) and calculates f_t function value. The input gate determines the information that will be stored in the current memory cell C_t. Lastly, the update gate updates the cell state, and the output gate computes the output information. The functions of the gates can be expressed mathematically as follows:

$$f_t = \sigma\left(W_f \bullet [h_{t-1}, x_t] + b_f\right) \tag{9}$$

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \tag{10}$$

$$\widetilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_c) \tag{11}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{12}$$

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \tag{13}$$

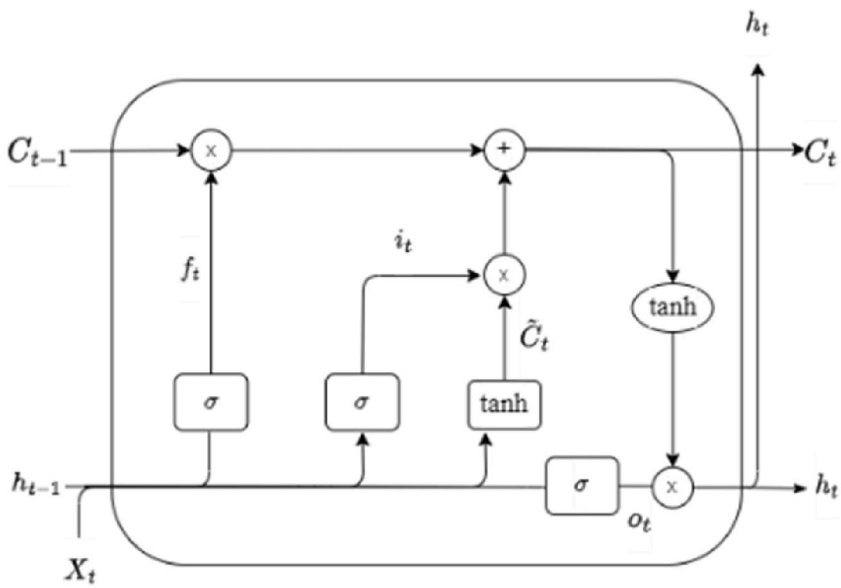

Fig. 4. Architecture of an LSTM unit

$$h_t = o_t * \tanh(C_t) \tag{14}$$

where W_f, W_i, W_o are the weight matrices of forgotten gate, input gate, and output gate, b_f, b_i, b_o are the bias terms of forgotten gate, input gate and output gate and σ is the sigmoid activation function, h_{t-1} the output value at the previous step, $\widetilde{C}_t$ is the new candidate state and x_t is the input at the current step.

The selection of parameters for any machine learning algorithm has a significant impact on how well it performs. The best set of parameters is determined with hyperparameter optimization. Hyperparameter optimization, which significantly affects model performance and convergence, is crucial to the success of LSTM neural networks. Some hyperparameter optimization techniques adapted for LSTM neural networks include grid search, random search, genetic algorithm, Bayesian optimization, and particle swarm optimization. Thus, we applied grid search to investigate the hyperparameters that control the behavior of LSTMs, such as the units, learning rate, activation function, number of epochs, and batch size. The grid search technique extensively evaluates the model performance for all possible combinations of a predefined set of hyperparameters. The performance of the grid search is evaluated using cross-validation on the training set. The predefined hyperparameter set is determined as follows: $[\text{units} = 50, 100, 150]$, $[\text{activation function} = \text{relu}, \text{elu}]$, $[\text{learning rate} = 0.001, 0.01, 0.1]$, $[\text{batch size} = 32, 64, 128]$, $[\text{epochs} = 50, 100, 200]$. The optimal set of parameters with the highest accuracy is identified after the grid search algorithm has been applied.

$$P_w = \frac{1}{2}\rho \pi R_r^2 C_p w^3 \tag{15}$$

The power output of a single wind turbine can be calculated with the well know Equation (15) where ρ, R_r, C_p, and w are respectively the density of air, the rotor radius, the power coefficient of the proportion of available power, and wind speed. Even so, the wind power output of a turbine, however, cannot be precisely estimated with the given equation due to the variable nature of wind and complex dynamics within and between turbines. In this study, LSTM neural networks were used to capture short and long-term dependencies and then to estimate wind power output using the data produced by the Copula-based dependent tree for the branches. A copula-generated dataset is processed by employing an LSTM model to gain more insights. This combination enables a more thorough understanding of complex data structures, making it especially useful in applications such as financial modeling, risk assessment, and time-series analysis. After the output value is generated for each branch, the expected value for the tree is calculated for each type of copula structure.

3. Empirical results

In this study, the proposed Copula-LSTM based decision tree framework for assessment of wind power plant energy potential was evaluated using real wind power plant data from Turkey. The hourly measured power output of the power plant as well as meteorological data from January 2021 to December 2021 are included in the study. The features in the dataset and their units are given in Table 4.

To reduce the computational burden of the model, a majority voting strategy is applied after the feature selection methods are used. Table 5 illustrates votes for every feature selection technique used in the study. The feature votes were used to determine whether the feature should be included in the final subset. If the majority (i.e., more than half) of the feature selection methods choose a given feature, then the feature is included in the final feature set. Thus, as the final set of features humidity, wind speed, and wind gust are obtained, which will be denoted by X_1, X_2, and X_3 respectively.

A multivariate distribution can be constructed with marginal distributions and a copula function that shows the dependence structure. The marginal densities for the variables must be derived to build a copula-based joint density that will be used to estimate conditional probabilities. The marginal distributions Normal, Gamma, Exponential, Lognormal Erlang, Weibull, and Beta were used. The mean squared error used to evaluate the goodness of fit of the distribution mathematical expression can be written as follows,

$$\text{MSE} = \text{E}\left[X_o - X_p\right]^2 = \frac{\sum_{i=1}^{N}\left[X_o(i) - X_p(i)\right]^2}{N} \tag{16}$$

where N is the number of values, X_o is the observed value and X_p denotes obtained value from the distribution. The selected distributions for the variables are given in Table 6.

Since humidity, wind speed, and wind gust are approximated as different distributions in modeling their joint distribution, there

Table 4
Feature descriptions.

Features	Unit
Temperature	$^\circ C$
Humidity	%
Pressure	hPa
Wind Speed	m/s
Wind Gust	degree
Power Output	kWh

Table 5
Feature selection methods.

Feature Selection Methods	Temperature	Humidity	Pressure	Wind Speed	Wind Gust
Forward Selection		•		•	•
Backward Elimination		•		•	•
Lasso Regularization		•	•	•	
Decision Tree Importance	•			•	•
Recursive Feature Elimination	•			•	•
Mutual Info			•	•	•

are limited modeling approaches available (Fig. 5). One of the approaches as mentioned earlier is using copula theory. Thus, the joint density function is constructed as a multivariate generalization of the marginals.

For modeling various types of dependencies, normal copulas from the elliptical copula family as well as Clayton and Frank copulas from the Archimedean copula family have been applied. Since parametric families of copulas are considered in the present study to build dependence structures between variables, the parameters of the specified copulas were estimated with inversion of Spearmen's rho and MOM approaches. The normal copula is generally parametrized in terms of product-moment correlation. The relation between product moment correlation and Spearman's rank order correlation is used to determine the normal copula correlation. For Archimedean copulas, the copula correlation is determined by the parameter θ. The value of θ can be easily obtained by the relation between θ and τ given in Table 2. The Kendall's τ is determined from the observations, whose mathematical expression given as

$$\tau_N = \binom{N}{2}^{-1}(P_n - Q_n) \tag{17}$$

where P_n denotes the number of concordant pairs and Q_n denotes the number of discordant pairs [57]. τ_N is the estimate of τ. The copula correlation parameters are critical since they enable us to shape the copula according to the unique structure of the data, which is essential for many statistical and financial applications. Using the specified copula and its specified parameters, a decision tree is modeled with discrete approximations to the conditional probability distributions at each branch of the tree. Figs. 6–8 illustrates the potions of constructed transient copula trees for the uniform random variables u_1,u_2, and u_3. The trees are constructed as follows: First, u_1 was estimated at three discrete points with MCS and then u_2 given that u_1 and the conditional percentiles were calculated with the equations from Table 1. Finally, the u_3 given that u_1 and u_2 was estimated.

After the calculation of the uniform variables, the discrete approximations to the original variables are calculated with inverse transformation. The portion copula-based tree structures for each type of copula are shown in Figs. 9–11. In the trees, the first uncertainty X_1 follows a normal distribution. The second and third uncertainties X_2 and X_3 follow beta distribution, and gamma distribution respectively. The uncertainty advances at each node in response to earlier decisions or events in a tree structure that represents a series of dependent uncertainties. The results of each node depend on the results of the preceding nodes, which produce a sequence of events. For instance, in Frank copula, the humidity, X_1 is modeled as a three-branch discrete chance node with outcomes 0.32, 0.56, and 0.8 and with the probabilities 0.25, 0.5, and 0.25 assigned to each outcome. The wind speed, X_2, is a conditional chance node and if X_1 is 0.56 (50^{th} percentile) then X_2 given that X_1 is modeled at three points with the values 0.862, 1.951 and 3.433. If X_2 given that X_1 is 1.951 (50^{th} percentile) then X_3 given that X_1 and X_2 is modeled at three discrete

Since obtaining mathematical expression of conditional densities is not available or computationally challenging, the dependent decision tree structure creates a practical way to create conditional distributions with discrete approximations.

To assess the potential of a wind power plant, the data generated by a copula-based tree model should be converted into corresponding power output data. The use of the LSTM model to transform data produced by copula-based decision trees is a promising technique with the potential to improve the performance of the framework. We employed an LSTM model trained with the historical hourly data for this purpose. The Copula-based tree structure is used to obtain the test input feature set. In this study, the optimal values of the parameters have been determined through the grid search technique. The learning rate, batch size, and number of epochs are all set to 0.01, 50, and 64, respectively, for the model. The hidden layer activation function is used as a ReLU (rectifier linear unit), and Adam is used as an optimizer.

Overall, in the study, the Copula-based dependence tree approach is used to model the dependence between the variables, and the LSTM model is used to transform the input data into power output values. The LSTM network is utilized to learn the variables (X_1, X_2 and X_3) of wind power and establish estimation model. This integrated approach allows for a more accurate representation of complex

Table 6
Marginal distributions for the variables.

Variable	Distribution	Parameters	Range
Temperature	Normal	$\mu = 11, \sigma = 9.11$	$[-15.08, 34.19]$
Pressure	Normal	$\mu = 1020, \sigma = 6.17$	$[995.8, 1036.3]$
Wind Speed	Scaled Beta	$\alpha = 2.94, \beta = 8.44,$	$[0.33, 7.87]$
Humidity	Normal	$\mu = 0.558, \sigma = 0.185$	$[0.07, 0.99]$
Wind Gust	Gamma	$\alpha = 0.861, \beta = 3.81$	$[0.71, 16.61]$

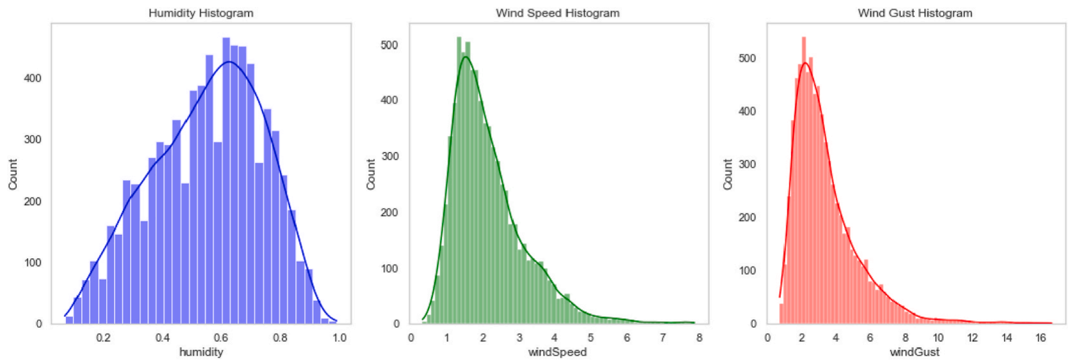

Fig. 5. Probability distribution fitting of the variables

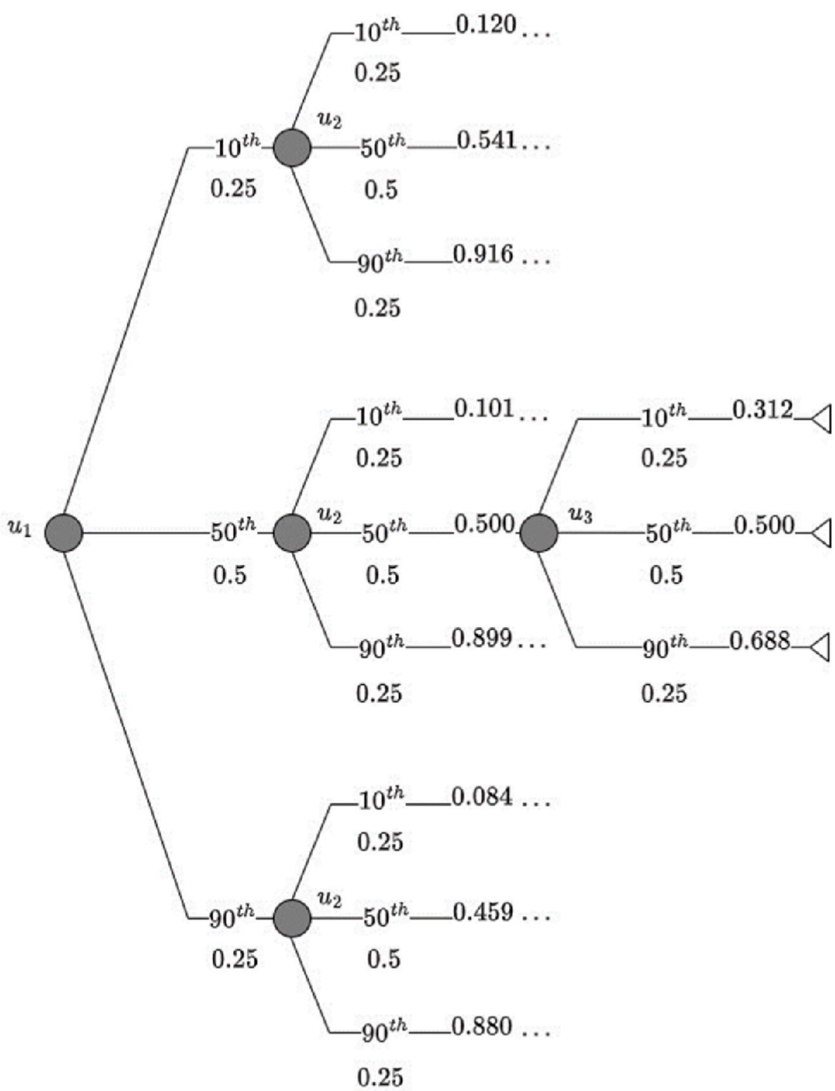

Fig. 6. Normal copula tree

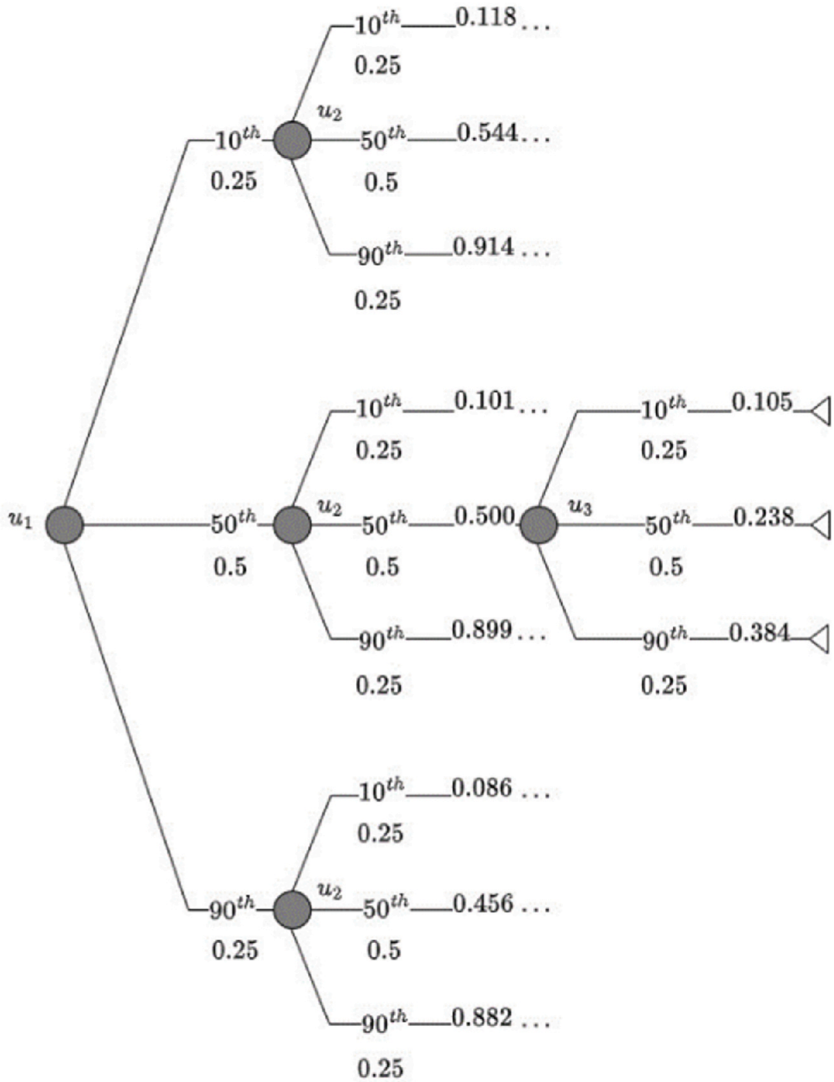

Fig. 7. Frank copula tree

data patterns. In order to successfully recognize samples, the LSTM model first learns the existing patterns that characterize wind power and then maintains long-term memory while learning the information in the sample.

Following the acquisition of the power output values for each of the 27 branches using the LSTM model and the probability values in the dependent tree structure, the expected value is calculated. The expected value is the sum of all the power output values multiplied by their respective probabilities. The expected values are shown in Table 7 for each type of copula used in this study. All types of copulas, apart from the independence copula, which consistently understates the true expected values, produce results that are closer to the actual mean value of historical data. The independence copula assumes that all variables are completely independent, which may not be true in many real-world situations where variables frequently exhibit some degree of interdependence or correlation.

It is a common practice for some analytical tools and models to operate under the fundamental premise of variable independence in the context of statistical modeling and analysis. This presumption makes mathematical calculations easier and frequently provides a useful starting point for many analytical tasks. However, it is crucial to acknowledge that the intricate and connected nature of real-world data may not always line up with such an assumption.

Variables in a given system or dataset frequently show varying degrees of interdependence and correlation, according to empirical findings and data-driven insights. Although independence was initially assumed, actual results frequently deviate from these straightforward models. These deviations raise doubts about the validity of the analytical framework's independence assumption and its suitability for capturing the complex interrelationships between variables.

Innovative strategies are developing in the area of statistical modeling in response to this difficulty. Utilizing a decision tree framework built on a Copula-LSTM is one such approach. Copula modeling, which captures intricate relationships between variables, is combined in this proposed framework with decision trees and LSTM networks. By combining these components, the framework

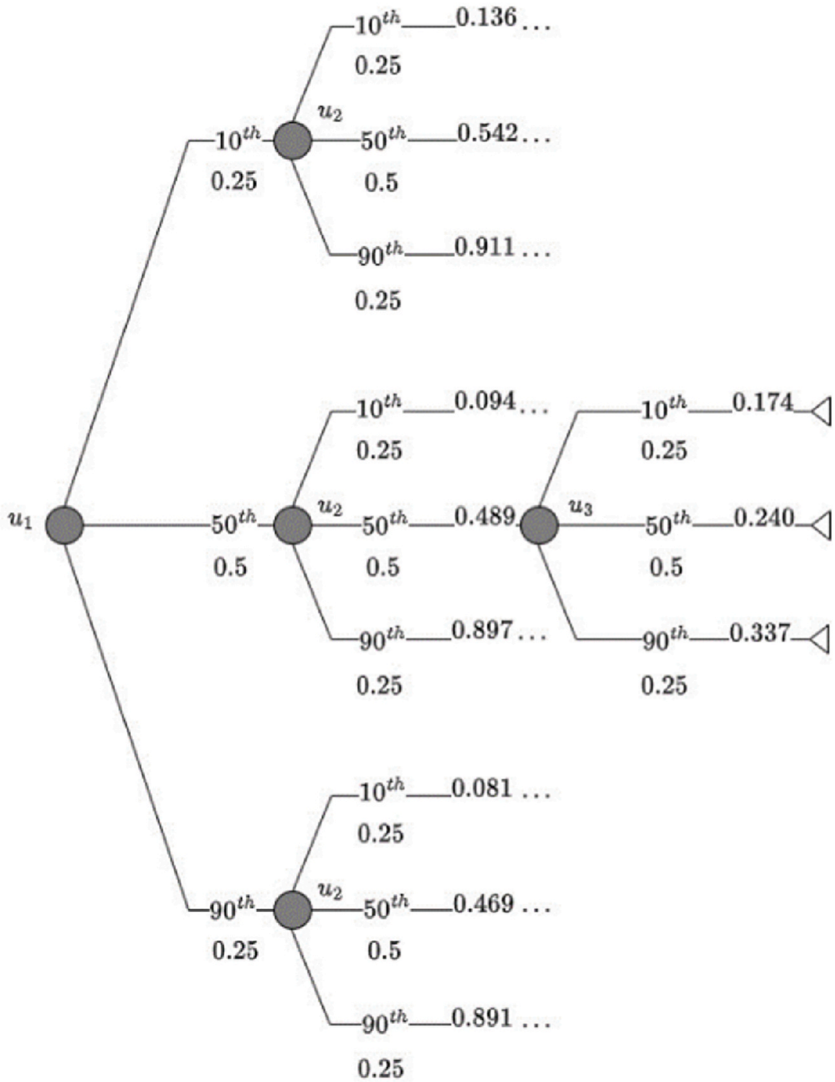

Fig. 8. Clayton copula tree

provides a more reliable way to model and analyze data, especially when dealing with non-independence among variables.

In conclusion, the independence assumption still simplifies analytical procedures, but when applied to real-world data, it has serious drawbacks. As a result of becoming aware of these inconsistencies, modeling approaches are being reevaluated, and the adoption of cutting-edge frameworks like the Copula-LSTM based decision tree approach is being encouraged. In addressing the complex interdependencies inherently present in the data, this method exhibits effectiveness and offers a convincing response to these difficulties.

4. Conclusion

In this study, we evaluate the optimal theoretical distributions of meteorological factors as marginal probability distributions and analyze their ability to replicate the properties of wind energy power output through the use of Copula-LSTM based decision trees. We validate the effectiveness of this framework and demonstrate its applicability beyond this specific study, showcasing its potential for integration into other research areas related to wind energy resource assessment, thus enabling more accurate results. While the joint distribution model we present, incorporating humidity, wind speed, and wind gust, constitutes a valuable contribution to the field, it is essential for academics to acknowledge the considerable scope for further improvement. Future research efforts should focus on enhancing the accuracy and comprehensiveness of the analytical framework.

In summary, our study extends conventional independence assumptions by acknowledging and modeling the dependence between meteorological variables using a Copula-based approach. By incorporating complex and nonlinear dependencies often overlooked in previous studies, we contribute to a more nuanced understanding of the interrelationships among these variables. Furthermore, we

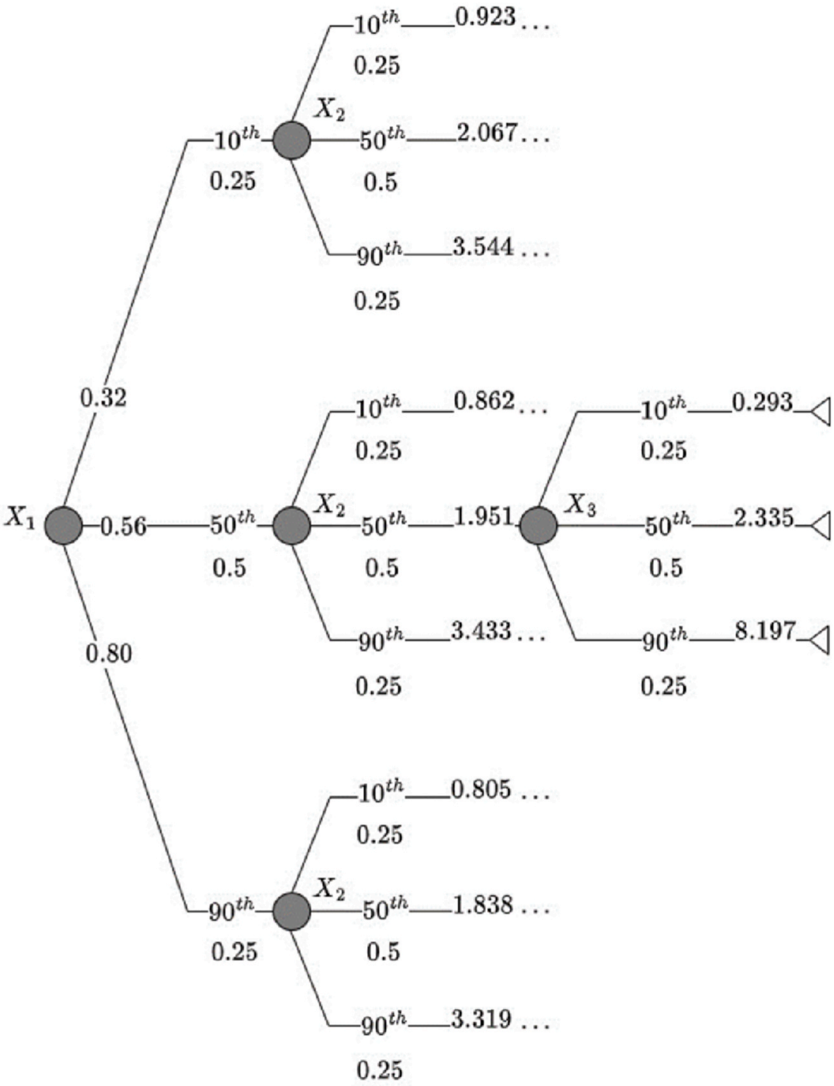

Fig. 9. Frank copula-based tree

integrate Copula and Deep Learning techniques into a decision tree framework to assess the energy potential of wind power plants, accounting for meteorological uncertainties. This fusion framework represents a substantial advancement in the field, offering a more robust and reliable approach to evaluating wind energy potential.

Our research not only contributes to scientific understanding but also facilitates the efficient utilization of wind energy resources. By shedding light on the techno-economic viability of wind power projects, it offers policymakers and stakeholders a mechanism for informed decision-making. As we navigate an era where renewable energy sources are crucial, our study emerges as a vital tool for the deployment and optimization of wind power plants, ultimately promoting a more sustainable and energy-diverse future.

Moreover, the emphasis in our study on advocating for mandatory pre-assessment requirements underscores the vital importance of adopting a proactive and thorough approach within the wind energy sector. We underscore the need for professionals to utilize sophisticated tools during comprehensive site evaluations, ensuring a holistic understanding of the environmental, technical, and economic aspects of potential wind power projects. This approach not only aids in cost-effective decision-making but also enables strategic resource allocation, optimizing the overall effectiveness and sustainability of wind energy initiatives. By establishing these best practices, our research aims to contribute to the long-term success and viability of wind power projects, aligning them with the broader goals of a sustainable and diversified energy landscape.

Even though the proposed framework offers valuable insights into the assessment of wind power plant site potential, it is crucial for academics to recognize the significant room for further improvement. Future studies should acknowledge and address the evolving conditions induced by climate change. The impacts of climate change introduce a dynamic layer of complexity, necessitating the inclusion of related variables such as temperature shifts and the increasing frequency of extreme weather events. Focusing on the

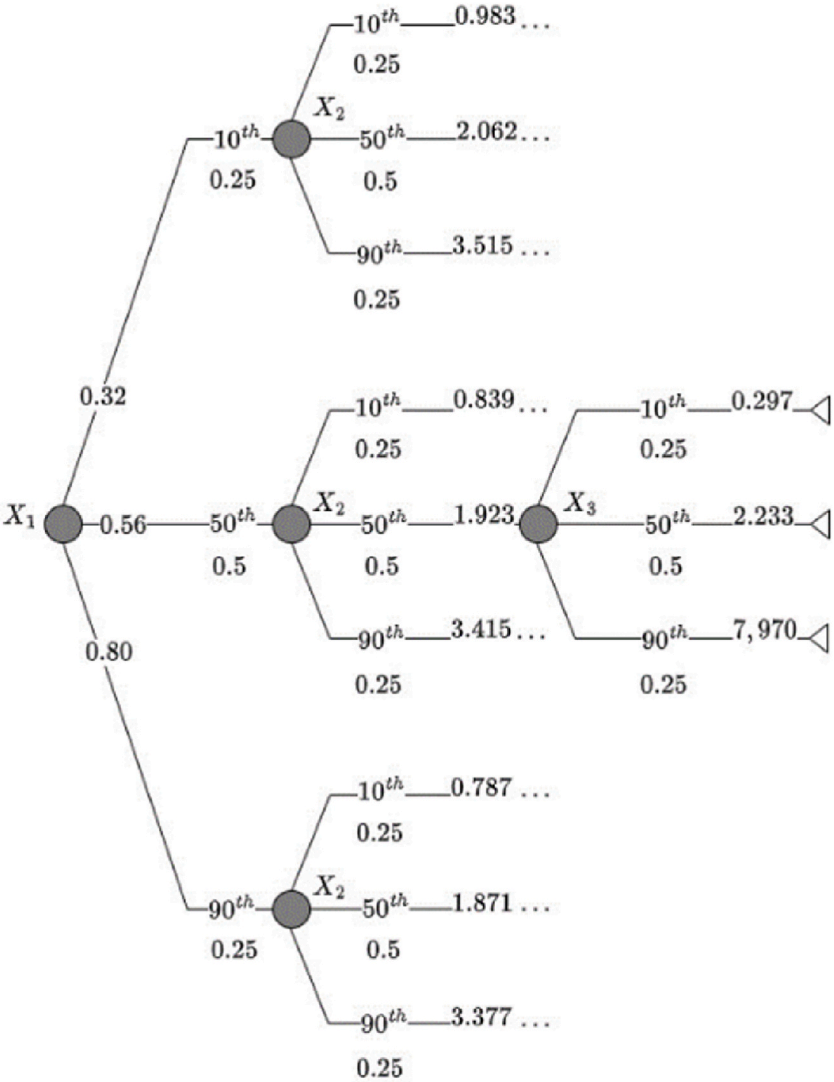

Fig. 10. Clayton copula-based tree

adaptability of the framework to these changing climatic conditions is crucial for ensuring its continued accuracy and relevance in assessing wind power plant site potential. Collaborating with climate scientists and experts becomes essential to gain insights into the anticipated shifts in climate parameters. By considering the broader context of climate change, future studies can strengthen the framework to provide more robust predictions, contributing to effective assessments of wind power plant site potential amidst a changing meteorological landscape. This approach aligns with the imperative to adapt to and mitigate the impacts of climate change, making the framework a valuable tool in anticipating and managing the consequences of evolving weather patterns for wind power applications.

However, it is important to note some limitations of our study. Firstly, the effectiveness of the Copula-LSTM based decision trees framework may vary depending on the specific characteristics of the meteorological data and the geographical location of the wind power plant. Therefore, further validation and testing of the framework across diverse datasets and locations are necessary to assess its generalizability. Additionally, while our framework captures complex dependencies and nonlinear patterns, there may still be other factors influencing wind power output that are not fully accounted for in our model. Future research could explore the incorporation of additional variables or more advanced modeling techniques to address these limitations and further improve the accuracy of wind energy potential assessments.

Statements and declarations

- The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

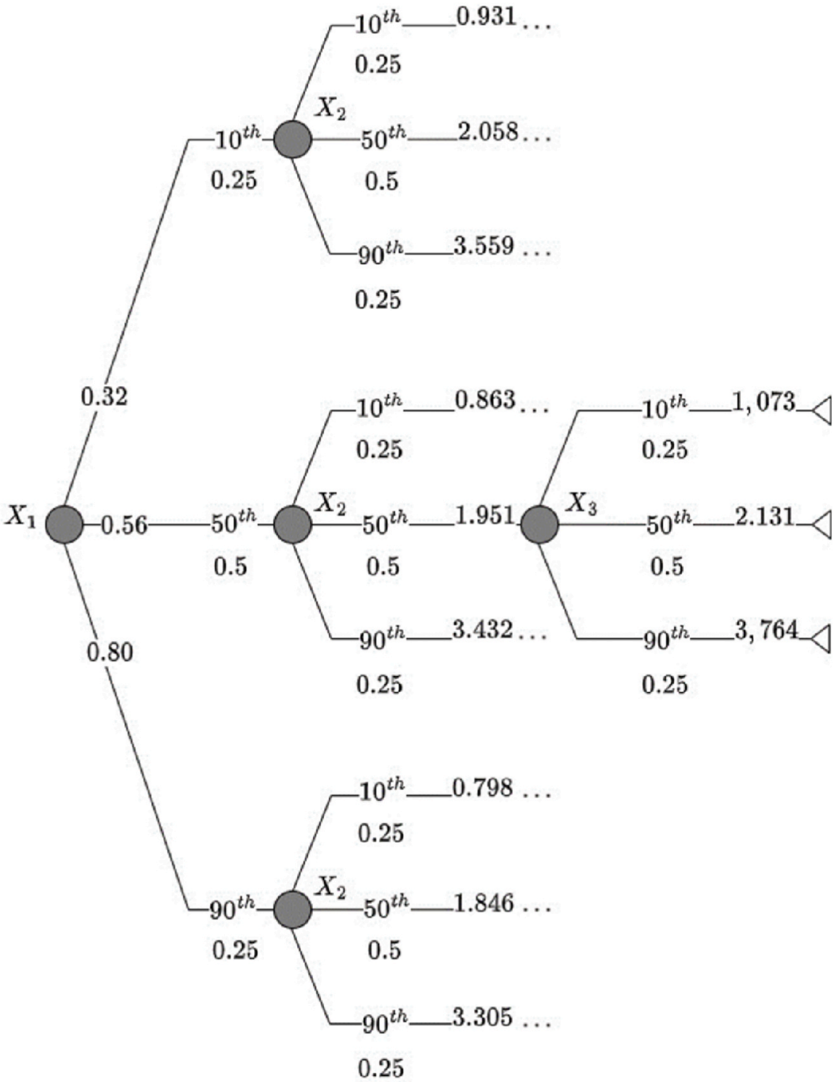

Fig. 11. Normal copula-based tree

Table 7
Expected value of the proposed models.

Model	Expected Value
Frank Copula	3566.34
Clayton Copula	3590.99
Normal Copula	3123.50
Independent Copula	2925.87
Sample Data Average	3551.00

- The authors have no relevant financial or non-financial interests to disclose.
- All authors contributed to the study equally. All authors read and approved the final manuscript.

Data availability statement

The raw/processed data required to reproduce the above findings is not shared at this time. The dataset used to analyze are available from the corresponding author upon request.

CRediT authorship contribution statement

Kübra Nur Şahin: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Muhammed Sutcu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by TUBITAK BIDEB 2211-A National Scholarship Program for Ph.D. students. Also, the APC was funded by the Gulf University for Science and Technology.

References

[1] H. Ritchie, P. Rosado, M. Roser, Emissions by sector: where do greenhouse gases come from?, Published online at OurWorldInData.org. Retrieved from, https://ourworldindata.org/emissions-by-sector, 2024 [Online Resource].
[2] K.S.R. Murthy, O.P. Rahi, A comprehensive review of wind resource assessment, Renew. Sustain. Energy Rev. 72 (May 2017) 1320–1342, https://doi.org/10.1016/J.RSER.2016.10.038.
[3] E. Vanvyve, L. Delle Monache, A.J. Monaghan, J.O. Pinto, Wind resource estimates with an analog ensemble approach, Renew. Energy 74 (2015) 761–773, https://doi.org/10.1016/J.RENENE.2014.08.060.
[4] D. Astolfi, F. Castellani, A. Garinei, L. Terzi, Data mining techniques for performance analysis of onshore wind farms, Appl. Energy 148 (2015) 220–233, https://doi.org/10.1016/J.APENERGY.2015.03.075.
[5] A. Kusiak, H. Zheng, Z. Song, Models for monitoring wind farm power, Renew. Energy 34 (3) (2009) 583–590, https://doi.org/10.1016/J.RENENE.2008.05.032.
[6] H. Oh, B. Kim, Comparison and verification of the deviation between guaranteed and measured wind turbine power performance in complex terrain, Energy 85 (2015) 23–29, https://doi.org/10.1016/J.ENERGY.2015.02.115.
[7] A. Altunkaynak, T. Erdik, I. Dabanli, Z. Şen, Theoretical derivation of wind power probability distribution function and applications, Appl. Energy 92 (2012) 809–814, https://doi.org/10.1016/J.APENERGY.2011.08.038.
[8] C. Jung, D. Schindler, J. Laible, A. Buchholz, Introducing a system of wind speed distributions for modeling properties of wind speed regimes around the world, Energy Convers. Manag. 144 (2017) 181–192, https://doi.org/10.1016/J.ENCONMAN.2017.04.044.
[9] D. Ganger, J. Zhang, V. Vittal, Statistical characterization of wind power ramps via extreme value analysis, IEEE Trans. Power Syst. 29 (6) (2014) 3118–3119, https://doi.org/10.1109/TPWRS.2014.2315491.
[10] B. Safari, Modeling wind speed and wind power distributions in Rwanda, Renew. Sustain. Energy Rev. 15 (2) (2011) 925–935, https://doi.org/10.1016/J.RSER.2010.11.001.
[11] R.O. Fagbenle, J. Katende, O.O. Ajayi, J.O. Okeniyi, Assessment of wind energy potential of two sites in North-East, Nigeria, Renew. Energy 36 (4) (2011) 1277–1283, https://doi.org/10.1016/J.RENENE.2010.10.003.
[12] A. Serban, L.S. Paraschiv, S. Paraschiv, Assessment of wind energy potential based on Weibull and Rayleigh distribution models, Energy Rep. 6 (2020) 250–267, https://doi.org/10.1016/J.EGYR.2020.08.048.
[13] M.A. Saeed, Z. Ahmed, J. Yang, W. Zhang, An optimal approach of wind power assessment using Chebyshev metric for determining the Weibull distribution parameters, Sustain. Energy Technol. Assessments 37 (2020) 100612, https://doi.org/10.1016/J.SETA.2019.100612.
[14] L. Bilir, M. Imir, Y. Devrim, A. Albostan, Seasonal and yearly wind speed distribution and wind power density analysis based on Weibull distribution function, Int. J. Hydrogen Energy 40 (44) (2015) 15301–15310, https://doi.org/10.1016/J.IJHYDENE.2015.04.140.
[15] S.H. Pishgar-Komleh, A. Keyhani, P. Sefeedpari, Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran), Renew. Sustain. Energy Rev. 42 (2015) 313–322, https://doi.org/10.1016/J.RSER.2014.10.028.
[16] T.P. Chang, Estimation of wind energy potential using different probability density functions, Appl. Energy 88 (5) (2011) 1848–1856, https://doi.org/10.1016/J.APENERGY.2010.11.010.
[17] S. Moradian, A.I. Olbert, S. Gharbia, G. Iglesias, Copula-based projections of wind power: Ireland as a case study, Renew. Sustain. Energy Rev. 175 (2023) 113147, https://doi.org/10.1016/J.RSER.2023.113147.
[18] M.M.N.S.B. McWilliams, The probability distribution of wind Velocity and direction on JSTOR, Wind Eng. 3 (4) (1979) 269–273 [Online]. Available: https://www.jstor.org/stable/43749150. (Accessed 8 September 2023).
[19] R. Weber, Estimator for the standard deviation of wind direction based on moments of the cartesian components, J. Appl. Meteorol. Climatol. 30 (9) (1991) 1341–1353, https://doi.org/10.1175/1520-0450(1991)030.
[20] E. Erdem, J. Shi, Comparison of bivariate distribution construction approaches for analysing wind speed and direction data, Wind Energy 14 (1) (2011) 27–41, https://doi.org/10.1002/WE.400.
[21] X. Bai, J.S.L. Lam, Portfolio value-at-risk estimation for spot chartering decisions under changing trade patterns: a copula approach, Risk Anal. 43 (6) (2023) 1278–1292, https://doi.org/10.1111/RISA.13989.
[22] H. Nguyen, F. Javed, Dynamic relationship between Stock and Bond returns: a GAS MIDAS copula approach, J. Empir. Finance 73 (2023) 272–292, https://doi.org/10.1016/J.JEMPFIN.2023.07.004.
[23] K. Xie, et al., Assessment of the joint impact of rainfall characteristics on urban flooding and resilience using the copula method, Water Resour. Manag. 37 (4) (2023) 1765–1784, https://doi.org/10.1007/S11269-023-03453-9/TABLES/5.
[24] J. Seo, J. Won, J. Choi, J. Lee, S. Kim, A copula model to identify the risk of river water temperature stress for meteorological drought, J. Environ. Manag. 311 (2022) 114861, https://doi.org/10.1016/J.JENVMAN.2022.114861.
[25] R.T. Clemen, T. Reilly, Correlations and Copulas for Decision and Risk Analysis 45 (2) (1999) 208–224, https://doi.org/10.1287/mnsc.45.2.208, 10.1287/MNSC.45.2.208.
[26] M. Sütçü, Disutility entropy in multi-attribute utility analysis, Comput. Ind. Eng. 169 (2022) 1–13, https://doi.org/10.1016/j.cie.2022.108189.
[27] M. Sütçü, "Parameter Uncertainties in Evaluating Climate Policies with Dynamic Integrated Climate-Economy Model", Environment Systems and Decisions, 2023, pp. 1–16, https://doi.org/10.1007/s10669-023-09914-1.
[28] M. Sutcu, A.E. Abbas, First-order dependence trees with cumulative residual entropy, AIP Conf. Proc. 1641 (2015) 512–521, https://doi.org/10.1063/1.4906017.
[29] T. Wang, J.S. Dyer, A copulas-based approach to modeling dependence in decision trees 60 (1) (2012) 225–242, https://doi.org/10.1287/OPRE.1110.1004.

[30] C. Emeksiz, B. Demirci, The determination of offshore wind energy potential of Turkey by using novelty hybrid site selection method, Sustainable Energy Technologies and Assessments Journal (2019), https://doi.org/10.1016/j.seta.2019.100562.
[31] G. Şahin, A. Koç, W. van Sark, Multi-criteria decision making for solar power - wind power plant site selection using a GIS-intuitionistic fuzzy-based approach with an application in The Netherlands, Energy Strategy Rev. 51 (2024) 101307, https://doi.org/10.1016/J.ESR.2024.101307.
[32] Y.A. Solangi, C. Longsheng, S.A.A. Shah, Assessing and overcoming the renewable energy barriers for sustainable development in Pakistan: an integrated AHP and fuzzy TOPSIS approach, Renew. Energy 173 (2021) 209–222, https://doi.org/10.1016/J.RENENE.2021.03.141.
[33] M.R. Islam, M. Tareq Aziz, M. Alauddin, Z. Kader, Site Suitability Assessment for Solar Power Plants in Bangladesh: A GIS-Based Analytical Hierarchy Process (AHP) and Multi-Criteria Decision Analysis, MCDA) approach, 2023, https://doi.org/10.1016/j.renene.2023.119595.
[34] M.G. Kendall, A new measure of rank correlation, Biometrika 30 (1/2) (1938) 81, https://doi.org/10.2307/2332226.
[35] K. Pearson, VII. Note on regression and inheritance in the case of two parents, Proc. Roy. Soc. Lond. 58 (1895) 240–242, https://doi.org/10.1098/RSPL.1895.0041, 347–352.
[36] M. Sklar, Fonctions de répartition à N dimensions et leurs marges, Annales de l'ISUP, 1959, pp. 229–231 [Online]. Available: https://hal.science/hal-04094463. (Accessed 11 September 2023).
[37] M.M.J. Zavareh, N. Mahjouri, M. Rahimzadegan, M. Rahimpour, A drought index based on groundwater quantity and quality: application of multivariate copula analysis, J. Clean. Prod. 417 (2023) 137959, https://doi.org/10.1016/J.JCLEPRO.2023.137959.
[38] R.B. Nelsen, An Introduction to Copulas, Springer, Springer New York, 2006, https://doi.org/10.1007/0-387-28678-0.
[39] M.J. Frank, On the simultaneous associativity of F(xy) and x+y-F(x,y), Aequationes Math. 19 (1) (1979) 194–226, https://doi.org/10.1007/BF02189866/METRICS.
[40] E.J. Gumbel, Bivariate exponential distributions, J. Am. Stat. Assoc. 55 (292) (1960) 698–707, https://doi.org/10.1080/01621459.1960.10483368.
[41] D.G. Clayton, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, Biometrika 65 (1) (1978) 141–151, https://doi.org/10.1093/BIOMET/65.1.141.
[42] E.M. Scheuer, D.S. Stoller, On the generation of normal random vectors, Technometrics 4 (2) (1962) 278–281, https://doi.org/10.1080/00401706.1962.10490011.
[43] B. Biller, C. Gunes Corlu, Copula-based multivariate input modeling, Surveys in Operations Research and Management Science 17 (2) (2012) 69–84, https://doi.org/10.1016/J.SORMS.2012.04.001.
[44] C. Genest, L.P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, J. Am. Stat. Assoc. 88 (423) (1993) 1034–1043, https://doi.org/10.1080/01621459.1993.10476372.
[45] G. Fu, D. Butler, Copula-based frequency analysis of overflow and flooding in urban drainage systems, J. Hydrol. (Amst.) 510 (2014) 49–58, https://doi.org/10.1016/J.JHYDROL.2013.12.006.
[46] A. Nazemi, A. Elshorbagy, Application of copula modelling to the performance assessment of reconstructed watersheds, Stoch. Environ. Res. Risk Assess. 26 (2) (2012) 189–205, https://doi.org/10.1007/S00477-011-0467-7.
[47] K. Xu, D. Yang, X. Xu, H. Lei, "Copula based drought frequency analysis considering the spatio-temporal variability in Southwest China," *J Hydrol (Amst)*, Complete 527 (2015) 630–640, https://doi.org/10.1016/J.JHYDROL.2015.05.030.
[48] C. Genest, K. Ghoudi, L.p. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (3) (1995) 543–552, https://doi.org/10.1093/BIOMET/82.3.543.
[49] W.H. Kruskal, Ordinal measures of association, J. Am. Stat. Assoc. 53 (284) (1958) 814–861, https://doi.org/10.1080/01621459.1958.10501481.
[50] P.J. McNamee, Decision Analysis with Supertree: Instructor's Manual, The Scientific Press, 1991.
[51] A. Hurst, G.C. Brown, R.I. Swanson, Swanson's 30-40-30 rule, Am. Assoc. Petrol. Geol. Bull. 84 (12) (2000) 1883–1891, https://doi.org/10.1306/8626C70D-173B-11D7-8645000102C1865D.
[52] D.L. Keefer, S.E. Bodily, Three-point approximations for continuous random variables 29 (5) (1983) 595–609, https://doi.org/10.1287/MNSC.29.5.595.
[53] N.A. Zaino, J. D'Errico, Optimal discrete approximations for continuous outcomes with applications in decision and risk analysis 40 (4) (1989) 379–388, https://doi.org/10.1057/JORS.1989.56.
[54] J.R. D'Errico, N.A. Zaino, Statistical tolerancing using a modification of taguchi's method, Technometrics 30 (4) (1988) 397, https://doi.org/10.2307/1269802.
[55] A.C. Miller, T.R. Rice, Discrete approximations of probability distributions 29 (3) (1983) 352–362, https://doi.org/10.1287/MNSC.29.3.352.
[56] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/NECO.1997.9.8.1735.
[57] Y. Wang, H. Ma, D. Sheng, D. Wang, Assessing the interactions between chlorophyll a and environmental variables using copula method, J. Hydrol. Eng. 17 (4) (2012) 495–506, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000387/ASSET/5BAB9F9C-2236-4186-878B-FD6F7CB27D66/ASSETS/IMAGES/LARGE/FIGURE7JPG.