Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Analyzing Machine Learning Models for Forecasting Precipitation in Australia

Hira Farman[1,2*], Dodo Khan[3], Saif Hassan[4], Muhammad Hussain[4], Sheikh Adnan Ahmed Usmani[5], and Daniyal-ur-Rehman[2,6]

[1]Department of Computer Science, Karachi Institute of Economics and Technology, Karachi, Pakistan.
[2]Department of Computer Science, Iqra University, Karachi, Pakistan.
[3]Department of CS & IT TIEST, NED University of Engineering Technology, Karachi, Pakistan.
[4]Department of Computer Science, Sukkur IBA University, Sukkur, Pakistan.
[5]SZIC, University of Karachi, Pakistan.
[6]Department of Mathematics, University of Karachi, Pakistan.
[*]Corresponding Author: Hira Farman. Email: hira.farman@iqra.edu.pk

Received: February 21, 2024 **Accepted:** May 03, 2024 **Published:** June 01, 2024

Abstract: In the 21st century, predicting when it will rain is an intriguing but challenging task. Climate and precipitation representations are frequently extremely complicated, non-linear, and inconsistent, need highly skilled, specialized mathematical modeling and training. The rise in rainfall-related flood tragedies in recent decades has made weather forecasting an increasingly important area of study. Most of the time, the researcher tried to find a linear relationship between the necessary data and the meteorological data that was already accessible. This work uses conventional machine learning algorithms to give a thorough analysis and prediction model for rainfall in Australia. Enhancing the precision and dependability of rainfall forecasts is the aim of this research. The dataset for the study contains historical meteorological data, such as temperature, humidity, wind speed, and air pressure, from multiple locations of Australia. Using classic machine learning techniques like Random Forest (RF) and Naive Bays closest neighbors, baseline models are created. Model evaluation is a meticulous procedure that contrasts the accuracy, precision, and memory of models. The primary meteorological factors that influence the variability of rainfall are identified using the feature importance analysis. The interpretability of the models is also investigated in the study in order to offer insightful information about the decision-making procedures. The dataset includes 14, 5460 size, 23 features detailed city-specific monthly averages for Australia from 2008 to 2017(10 years). An effective rainfall forecasting was produced by integration of a number of machines learning techniques, including Random Forest model (RF), K nearest Neighbor (KNN), Decision Tree (DT), Naïve Bayes (NB), and Logistic Regression (LR). This research intends to mitigate the high risks of floods induced by natural disasters by utilizing state-of-the-art models. The results show that random forests have high accuracy (0.859) for predicting rainfall.

Keywords: Rainfall; K Nearest Neighbors (KNN); Naive Bayes (NB); Machine Learning (ML).

1. Introduction

Rain, snow, hail, etc., in other words, any kind of precipitation can impede outdoor activities of the day. Precipitation is one of the tougher cubes to crack when it comes to weather prediction. Forecasting the rainfall often presents a slight challenge as compared to the earlier years due to the increased variability of climate conditions. There are also methods such as machine learning from weather data where one is able to decipher patterns on the amount of rainfall likely to be expected. The decision in choosing an appropriate classification technique for prediction is very challenging. This research presents an attempt at developing a new solution for a real-time rainfall prediction system. Three artificial intelligence techniques were used in the research namely "K-nearest neighbor (KNN)", "Artificial Neural Network

(ANN)". Rainfall prediction models include [1] "ANN, and "Extreme Learning Machine (ELM)" – all of which are applied to forecast progression of summer season. Monsoon period from June to September and post monsoon period from October to December of year 2011-2016 of these techniques are done for the Kerala state of India and effectiveness of the performance of the computed values is checked with the observations. All the described above approaches have proven reasonably effective and if compared to ELM has proved to be better in performance relative to the lower mean absolute percentage error scores for summer monsoon (3.075) and post-monsoon (3.149) than both KNN and ANN techniques. The manifold model here [2] is presented as an ML solution to hydraulic modeling of flood waters. All model complies, performance standards tailored operational use, of which a history of data remains valuable for benchmarking. For the research in the article, the system that is suggested is for the detection of potential floods in a river basin with the use of machine learning and IoT [3]. The model links the Wireless Sensor Network or WSN to an individual mesh network through a ZigBee and uses a GPRS module to upload data over the internet. In this work [4], it explores identifying possible machine learning algorithms for forecasting flood occurrences in the Pattani River using Open data.

The model connects the Wireless Sensor Network, or WSN, to a personalized mesh network via a ZigBee connection, and it then uses a GPRS module to transmit data over the internet. For predicting the occurrence of floods in the Pattani River, this work [5] examines applying possible machine learning algorithms using open data. To forecast the average rainfall in Pakistan [21], a sliced functional time series model is applied and the future monthly predictions will be nine years ahead. The model will be intended to address complex time series analysis methods and will be required to ensure that the long-term rainfall prediction is not beyond the weather accurate enough to be used in the vital areas of grass and rivers as well as disaster risk reduction. Climate indices monthly and Seasonal are discussed assessing their significance in precipitation unearthing correlation for the first instance in Peshawar City, Pakistan with urban heat island intensity. This study [22] is also expected to achieve the purpose by employing methods such as rain dynamics analysis and computational intelligence for the rainfall forecasting and consequently enhances the effectiveness and dependability of the predicted rain for sectors like agriculture, water resource distribution and disaster risks. The algorithm exhibited a significant ability to generalize [23] and provided the precipitation values in most of the meteorological stations corresponding to the particular climatic zone.

1.1. Types of Rainfall

Rainfall can be classified into various types based on different criteria such as duration, intensity, spatial distribution, and the mechanisms responsible for its occurrence. Some common types of rainfall shown in figure 1.

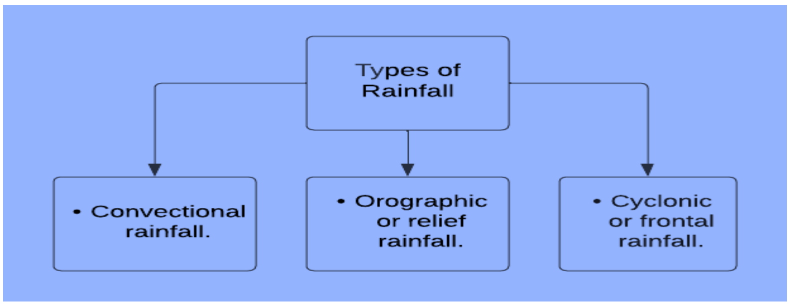


Figure 1. Types of rainfall

1.1.1. Convectional Rainfall

As the Earth's surface warms, air rises, resulting in convectional rainfall. As it increases, the air cools and releases moisture as rain. In regions with high temperatures and high humidity, convectional rainfall is frequent, frequently occurring in tropical climates in the late afternoon or evening.

1.1.2. Orographic Rainfall

Orographic rainfall is the result of moist air being lifted as it is pushed over high ground, such as mountains or hills. On the mountain's windward side, moisture and precipitation result from the rising, cooling air. Due to descending dry air, the leeward side gets a rain shadow effect, which reduces rainfall.

1.1.3. Cyclonic Rainfall

Rainfall associated with cyclones or low-pressure systems is referred to as cyclonic rainfall. It is characterized by persistent, widespread rain that is frequently accompanied by high winds. Forecasting weather, preparing for disasters, and managing water resources all depend on an understanding of cyclonic rainfall patterns

1.2. Positive and Negative consequences of Rainfall

Rainfall [1-4] naturally waters crops, reducing the reliance on artificial irrigation systems. Additionally, it replenishes soil nutrients, promoting healthy crop growth. Rainfall also refills rivers, lakes, and reservoirs, ensuring a steady water supply for farming. Beyond agriculture, rainfall contributes to the replenishment of groundwater and surface water sources, which are vital for drinking water supplies. Furthermore, it is essential for hydropower generation, as it maintains water levels in dams and reservoirs [20-24]. Flooding, whether in urban or rural settings, poses significant risks. Heavy rainfall can overwhelm urban drainage systems, causing flooding, property damage, and disruptions in daily life. Similarly, excessive rainfall can lead to river overflow, resulting in widespread flooding, community displacement, and infrastructure destruction some consequences shown in figure 2, table 1 illustrates the multifaceted impacts of rainfall in Australia, emphasizing its benefits in sustaining agriculture and ecosystems, while also acknowledging the challenges it poses for infrastructure and disaster resilience.

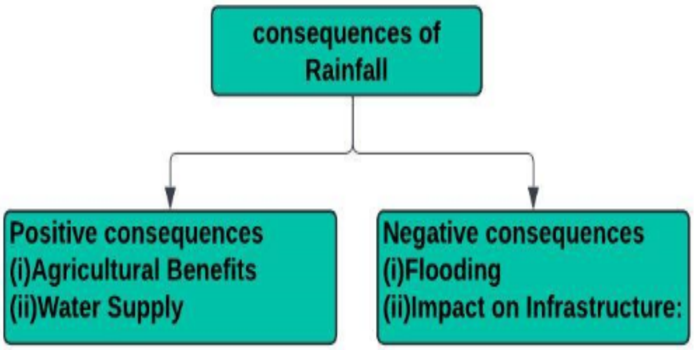


Figure 2. Positive and Negative consequences of Rainfall

Table 1. Some Positive and Negative Consequences of Rainfall

Aspect	Positive Impact	Negative Impacts
Agricultural Sustainability	Natural irrigation for crops - Enhanced crop growth	Soil erosion and degradation Crop damage from floods
Water Resource Replenishment	Replenishment of groundwater and surface water sources	Urban and river flooding, leading to property damage
Hydropower Generation	- Maintains water levels in dams and reservoirs	Damage to infrastructure from heavy rainfall

1.3. Purpose of the Study

The aim of this research is to utilize the study results as effectively as possible. The research questions of this study are provided as follows:

- Univariate and multivariate analysis of datasets with regard to several features
- Several models of machine learning have been experimented and have shown their accuracy in forecasting Rainfall.
- Comparative analysis for different algorithms for accuracy and error for Rainfall prediction.
- An extended experiment was conducted to provide a detailed examination of the proposed model.

1.4. Contribution of paper

- Leveraging the Orange data mining tool for rainfall prediction using machine learning techniques applied to meteorological data.
- Employing advanced feature selection methods to identify crucial variables for accurate forecasting of "Rain Tomorrow" in Australian weather.

- Evaluating and comparing multiple machine learning algorithms (Decision Tree, K-Nearest Neighbors, Random Forest, Naive Bayes, Logistic Regression) within the Orange framework.
- Iteratively refining models based on comprehensive performance metrics (accuracy, precision, recall, F1 score) to enhance prediction outcomes
- Offering practical implications for improving weather forecasting systems and disaster preparedness strategies.

The remaining portion of the study is structured as follows: Conversely, the study's Australia location, data collection, experimental procedures, theoretical framework, and comparative analysis are all provided. The results and discussion section of the work comes after the designing of the machine learning algorithms, which are the primary component of the early rainfall prediction of the dataset based on accuracy and other measures like F1 score, accuracy, and recall. The final section is a summary and future research section that serves as a conclusion.

2. Related work

Australia, with its varied climate ranging from tropical to desert, experiences significant fluctuations in rainfall, affecting everything from agriculture to water supply and disaster management. Researchers employ advanced techniques and technologies, including climate modeling, satellite data, and machine learning, to predict rainfall patterns. These efforts are crucial in understanding and mitigating the impacts of both rainfall and floods, which are common challenges in the country.

In this study [1] for the Kerala state of India considered, three artificial intelligence approaches—K-nearest neighbor (KNN), artificial neural network (ANN), and extreme learning machine (ELM)—are used to forecast rainfall during the summer monsoon (June-September) and post-monsoon (October-December) seasons from 2011 to 2016. The effectiveness of these approaches is assessed in comparison to observations. While all of the aforementioned strategies have demonstrated respectable performance, the ELM methodology has outperformed KNN and ANN procedures with low mean absolute percentage error scores for the summer monsoon (3.075) and post-monsoon (3.149), respectively. The manifold model is represented in this [2] as a machine learning substitute to hydraulic modeling of flood waters. All models achieve performance standards that are tuned to operational use, provided historical data has been used as a benchmark. The article proposes a system that predicts possible floods in a river basin using machine learning and the Internet of Things (IoT) for the research [3]. The model connects the Wireless Sensor Network, or WSN, to a personalized mesh network via a ZigBee connection, and it then uses a GPRS module to transmit data over the internet. For predicting the occurrence of floods in the Pattani River, this work [4] examines applying possible machine learning algorithms using open data. This study [5] employed a combination of machine learning tools, such as support vector regression (SVR), fuzzy inference model (FIM), and the k-nearest neighbors (k-NN) method in order to develop a probabilistic forecasting model. This study takes a probabilistic forecasting model with the help of a number of machines learning approaches, including SVR model (a support vector regression model), fuzzy inferential models (FIM), and the KNN technique (the K nearest neighbors' technique). In total, three multi criteria decision making evaluation approaches, namely VIKOR, SAW, and TOPSIS, along with two machine learning methods, NB and NBT, were utilized to assess their ability to simulate flood vulnerability in the Ningdu Catchment [Complex Flow of Words] which is one of China's most flood-prone regions [7]. Two techniques, Extensive gradient boosting This experiment is based on the Deep Belief Network (DBN) for forecasting the Daya and Bhargavi river banks, which flow towards the Indian state of Odisha. A comparison study that is based on other machine learning methods helps to demonstrate the beneficial impacts of dams in detail. This study [10] focuses on the application of ant colony optimization (ACO), Genetic algorithms (GA), artificial neural networks (ANN), and Particle swarm optimization (PSO) approaches to flood hydrograph prediction. In the current study [11], the relative accuracy of the RB FNN, SVM, and Firefly Algorithm (FA) models compared to the regular ANN, RB FNN, and SVM algorithms for river flood discharge forecasting in the Barak River was examined. Urban flooding is becoming increasingly common [12], which is detrimental to both the economy and quality of life for people. However, the existing flood prediction algorithms have grown too primitive or insufficient to accurately capture the details of flood evolution. This study uses deep neural networks to quicken the computation of a physics-based 2D urban flood forecasting approach that uses the Shallow Water Equation (SWE).

Using data modeled by the use of a partial differential equation (PDE) solver, “Convolution Neural Networks (CNN)” and generative adversarial networks with conditions “CGANs” are utilized to identify flood dynamics. The four ML-based FSMs “Random Forest (RF)”, “K Nearest Neighbor (KNN)”, “Multilayer Perceptron (MLP)”, and “Hybridized Genetic Algorithm”, “Gaussian Radial Basis Function”, and “Support Vector Regression (GA; RBF; SVR)” shown in this article [13]—present a framework for reducing spatial disagreement. The outcomes of those four models were combined to generate an enhanced model as well. The approach presented in this study may be useful in developing risk-based development plans and enhancing current early warning systems. This paper [14] utilized a deep learning-based model to predict the water level flood phenomenon of a river in Taiwan. The experimental results showed that the Conv GRU neural network model performed better than other current methods. The trial's outcomes showed that the suggested method could correctly identify the wrong water levels. The work [15-16] investigates whether Artificial intelligence & machine learning algorithms may serve for precipitation prediction implemented on the data on the rainfall from major urban areas in Australia during the last decade. Several machine learning algorithms such as k nearest neighbors (KNN), decision tree (DT), random forest (RF), and neural networks (NN), are tested, with neural networks showing to be the most effective model for rainfall prediction, which is a very important knowledge for the improvement of weather forecasting abilities. The main aim of this study [17] is to develop a flexible computing system for the same decimal point rainfall forecast in Upstate New York. It is a kind of algorithm that combines classics and trend and involves algorithms such as K-Nearest Neighbors, Support Vector Machine, Deep Neural Network, Wide Neural Network, Reservoir Computing, and Long Short-Term Memory with the technology which provides accurate and reliable rainfall forecasts that are tailored to local geographic conditions and climate variations.

This study [18] proposes an improved technique for creating daily short-term and monthly long-term ensemble weather forecasting models for rainfall predictions. This is achieved by combining five rainfall prediction models (Naïve Bayes, C4.5, neural network, support vector machine, and random forest) using three linear algebraic combinations: maximum probability, average probability, and majority vote. Using the Malaysian state of Selangor, daily weather data over a six-month period (2010–2015) yielded 1581 occurrences, which were categorized into two groups. There are two classes of rainfall: "active rainfall," which has 428 instances, and "no rainfall," which has the remaining instances. This initiative [19] aims to use feature selection and machine learning approaches to create the most accurate rainfall forecast model possible. Prior to and following feature selection, the Artificial Neural Network (ANN) attains a maximum accuracy of 90% and 91%, respectively. This research [20] primary contribution is to identify the most recent machine learning techniques for flood prediction as well as the noteworthy parameters that were used as model input. This will allow scientists and/or flood managers to use the prediction results as a reference when evaluating ML methods for early flood prediction.

A sliced functional time series model [21] is used to predict the average rainfall in Pakistan, which is a monthly forecast for the next ten years. The model will be designed to perform advanced time series analysis techniques and will be expected to limit the long-term rainfall forecast to weather accurate and reliable enough to be applied in vital areas including grassland and rivers; and disaster risk reduction. Monthly and seasonal climate indices are evaluated for their apparent relationship for precipitation in Peshawar City, Pakistan, and evidence correlation for the initial. This study [22] is expected to attain the purpose by using techniques such as rain dynamics analysis and machine learning for rainfall prediction and thereby improves the accuracy and reliability of the rain forecast that helps the various sectors like agriculture, distribution of water resources and disaster risk reduction. There has been a remarkable generalization capability of the algorithm [23] [36], yielding precipitation predictions in most meteorological stations with a reasonable correspondence to climatic zones. Yet, some stations show a lower estimation of annual total precipitation, this indicates that the prediction model needs to be further improved and validated to solve the discrepancies and get the accuracy improved.

Rainfall prediction research in Australia is critical for managing the country's unique and often extreme weather conditions. Advanced prediction methods help safeguard agriculture, water resources, infrastructure, and public health, contributing to overall societal resilience. By improving the accuracy and reliability of rainfall forecasts, researchers can help mitigate the adverse effects of climate variability and support sustainable development across Australia.

3. Comparative Analysis of Past Studies

Weather forecasting plays an important role in understanding precipitation and being able to control the effects they bring, the use of soil and water, protection of structures and people from lightning strikes, decisions on water resource management, and assessment of damages caused by natural disasters in many countries across the globe. All these models of binary classifiers of machine learning provide some kind of advantage, which are useful as tools in the armory for operational meteorologists/forecasters and researchers in the sphere of rainfall prediction. By sophisticated these models, academic researchers can enhance the rainfall prediction to delineate the challenges resulting from unfavorable climatic conditions that in turn enhance community and economic sustainability, some comparison with previous studies shown in table 2.

Table 2. Comparative Analysis of Rainfall prediction

Ref	Year	Country	Model	Input	Target	Dataset	Result parameter	Research Insight
[17]	2022	Pakistan city Lahore	Fuzzy logic, Decision tree(DT), Naïve Bayes, K-nearest neighbors (KNN), & support vector machines (SVM).	Rainfall	Rainfall overall	12 years of historical weather data (2005 to 2017)	Specificity, Sensitivity, Negative Prediction Value, Positive Predictive Value	The prediction accuracy of the used machine learning techniques are fused using fuzzy logic. it provides the great result
[24]	2012	Australia	Artificial Neural Network (ANN)	Rainfall	monthly Rainfall	1997–2010	RMSE, R, MAE	Built a successful ANN prototype superior to the existing POAMA-1.5 GCM models in Australia.
[25]	2020	Australia	Artificial Neural Network (ANN)	ENSO, IOD NSO (El-Niño-Southern Oscillation) & IOD (Indian Ocean Dipole)	monthly	1957–2014	RMSE, R, MAE,	The non-linear ANN models, trained on past climate data (ENSO and IOD), outperformed MLR models
[26]	2023	India	MLP-ANN	Rainfall	(daily)	1983-2016	CC, RMSE	AI approach is a valuable tool for future weather parameter prediction in India
[27]	2021	Malaysia	(Bayesian Linear Regression) (Boosted Decision Tree Regression), BDTR, DFR	Rainfall	Daily, Weekly, 10-day, monthly	2010–2019	MAE, RMSE, RAE, R	Boosted Decision Tree Regression (BDTR) achieved the highest

			Decision Forest Regression, NeuralNetwork Regression					accuracy, especially for daily and monthly forecasts using an Autocorrelation Function (ACF) approach
[28]	2022	UK	S-LSTM, B-LSTM, XGBoost, GBR, LSVR	Rainfall	Hourly	2000–2020	Loss, RMSE, MAE	Stacked-LSTM and Bidirectional-LSTM networks performed best, suggesting these approaches can be effective for budget-conscious rainfall Forecasting applications

4. Discussion on Past Studies

Complexity Previous research on rainfall prediction has used a variety of approaches, from complex machine learning algorithms to statistical models. Here's a summary of some important findings from earlier research on rainfall prediction. In the past, researchers have used a variety of data sources, such as satellite imaging, climate models, and meteorological measurements, to forecast rainfall. Features including wind speed, humidity, temperature, the land, atmospheric pressure, and oceanic conditions are often extracted from these data sources. Rainfall data's temporal and spatial characteristics must be considered in order to fully represent its complex patterns. Because machine learning approaches can capture nonlinear correlations and manage enormous datasets, they have become popular for rainfall prediction. Popular Deep learning model Neural networks, and supervised machine learning algorithm decision trees (DT), support vector machines (SVM), and random forests (RF) are some of the methods that easily capture non linearity from data and are utilized for rainfall prediction. Rainfall prediction is inherently uncertain due to the chaotic nature of atmospheric processes and the influence of various factors such as climate change, El Niño-Southern Oscillation (ENSO), and local topography Deep learning models LSTM and ANN are presently the main techniques for rainfall forecasting, with a focus on machine learning. Using meteorological radar data, this study [28] [38] used LSTM networks to predict short-term rainfall in mountainous areas. The results showed promise in terms of lead time and prediction accuracy. In contrast to conventional methods. Regardless of many difficulties, supervised machine learning has several potential applications in rainfall prediction hourly, seasonally, daily, monthly. To fully achieve the potential of machine learning, ongoing research, data collection, stakeholder involvement, and the integration of ML with conventional modeling techniques will be required [41] [42].

However, the prior research had a number of limitations and flaws. Limited availability of high-quality and spatially dense rainfall data poses challenges for model training and validation. Overfitting, especially in complex machine learning models, can lead to poor generalization performance, particularly when dealing with short and noisy time series data.

Many models perform poorly in novel contexts because they overfit to datasets.

ML models have the ability to accurately predict rainfall; nevertheless, there can be a lack of interaction between them and decision support systems (DSS) to facilitate real-time decision-making. -User-friendly interfaces, interoperability, and seamless integration are essential to guaranteeing the practical applicability of machine learning-based forecasting systems. Interpretability and explain ability are often lacking in machine learning models, particularly deep learning models, which makes it difficult for users to comprehend how predictions are made. This restriction impedes decision-making, adoption, and trust in operational forecasting applications [39] [40].

5. Methodology

The figure 3 depicts machine learning methodology for developing a predictive model using meteorological data for forecasting rainfall. The process begins with the collection of raw meteorological data, which is then pre-processed to handle missing values, normalize the data, and perform other necessary cleaning steps. Following this, relevant features are extracted from the pre-processed data, and the most pertinent features for the model are selected. The dataset is then split into training and testing sets to facilitate the evaluation of the model's performance.

Various machine learning algorithms [31-33] are applied to the training data, including Decision Tree, K-Nearest Neighbors (KNN), Random Forest, Naive Bayes, and Logistic Regression. The models are evaluated using metrics such as accuracy, precision, and recall to determine their performance. Based on these evaluations, the best-performing model is selected. This process is iterative, with potential refinements in the pre-processing, feature extraction, and modeling steps based on the evaluation results to improve the model's accuracy and reliability.

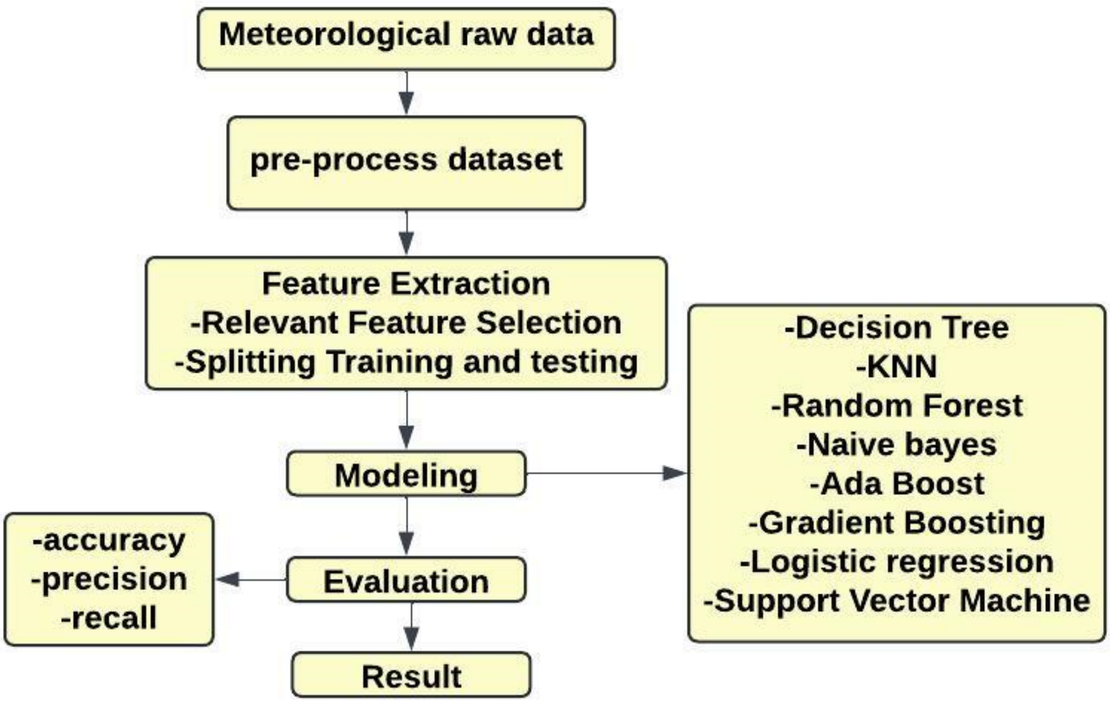


Figure 3. Representation of Proposed work overflow

5.1. Dataset Description

A dataset gathered from the Kaggle platform constituted a basis for the work discussed in the article. As indicated in Table 2, the data collection covers a sample of 145,460 entries with information on 23 Metrological attributes. The values provide meteorological information that was compiled over a ten-year period from 49 distinct Australian cities. The target parameter for the prediction using machine learning prediction task is a Boolean variable named "Rain Tomorrow" indicating yes or no as to whether it will rain tomorrow. Figure 4 displays the overall number of missing values in the dataset.

- **Date:** "The date of the weather observation."
- **Min Temp:** "The minimum temperature recorded during the day."
- **Max Temp**: "The maximum temperature recorded during the day."
- **Rainfall:** "The amount of rainfall recorded during the day."
- **Evaporation:** "The amount of water evaporation recorded during the day."
- **Sunshine:** "The duration of sunshine recorded during the day."
- **Wind Gust Speed:** "The maximum wind gust speed recorded during the day."
- **WindSpeed9am:** "The wind speed recorded at 9 am."
- **WindSpeed3pm:** "The wind speed recorded at 3 pm."
- **Humidity 9am:** "The humidity level recorded at 9 am."
- **Humidity 3pm**: "The humidity level recorded at 3 pm."
- **Pressure 9am:** "The atmospheric pressure recorded at 9 am."
- **Cloud9am:** "The cloud cover recorded at 9 am."
- **Cloud3pm**: "The cloud cover recorded at 3 pm."
- **Temp 9am:** "The temperature recorded at 9 am."
- **Temp 3pm:** "The temperature recorded at 3 pm."
- **Location:** "The geographical location where the weather observation was recorded."
- **Wind Gust Dir:** "The direction of the maximum wind gust."
- **WindDir9am**: "The wind direction recorded at 9 am."
- **WindDir3pm**: "The wind direction recorded at 3 pm."
- **Rain Today:** "Whether it rained today (yes/no)."
- **Rain Tomorrow:** "Whether it will rain tomorrow (yes/no)."
- **Pressure 3pm**: "The atmospheric pressure recorded at 3 pm."

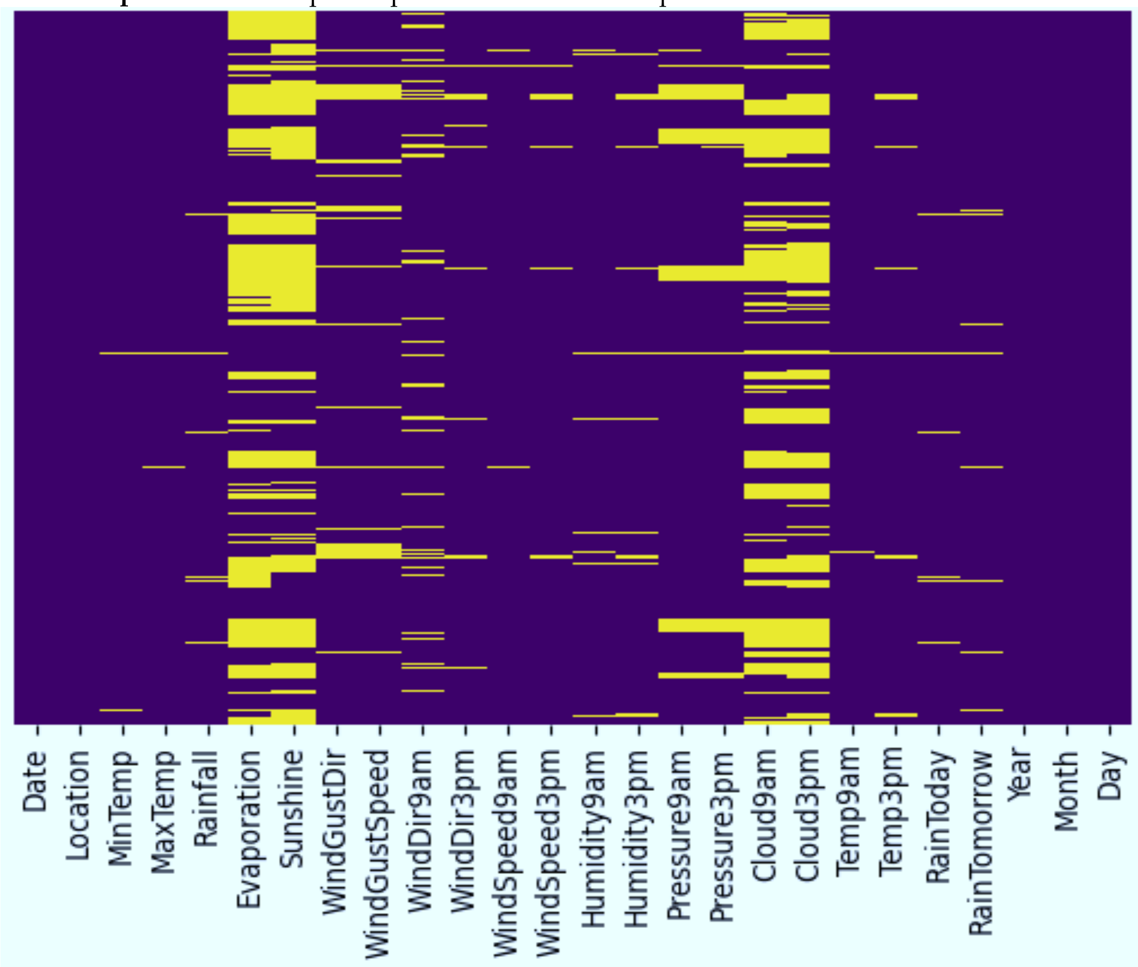


Figure 4. Missing values in each feature of dataset

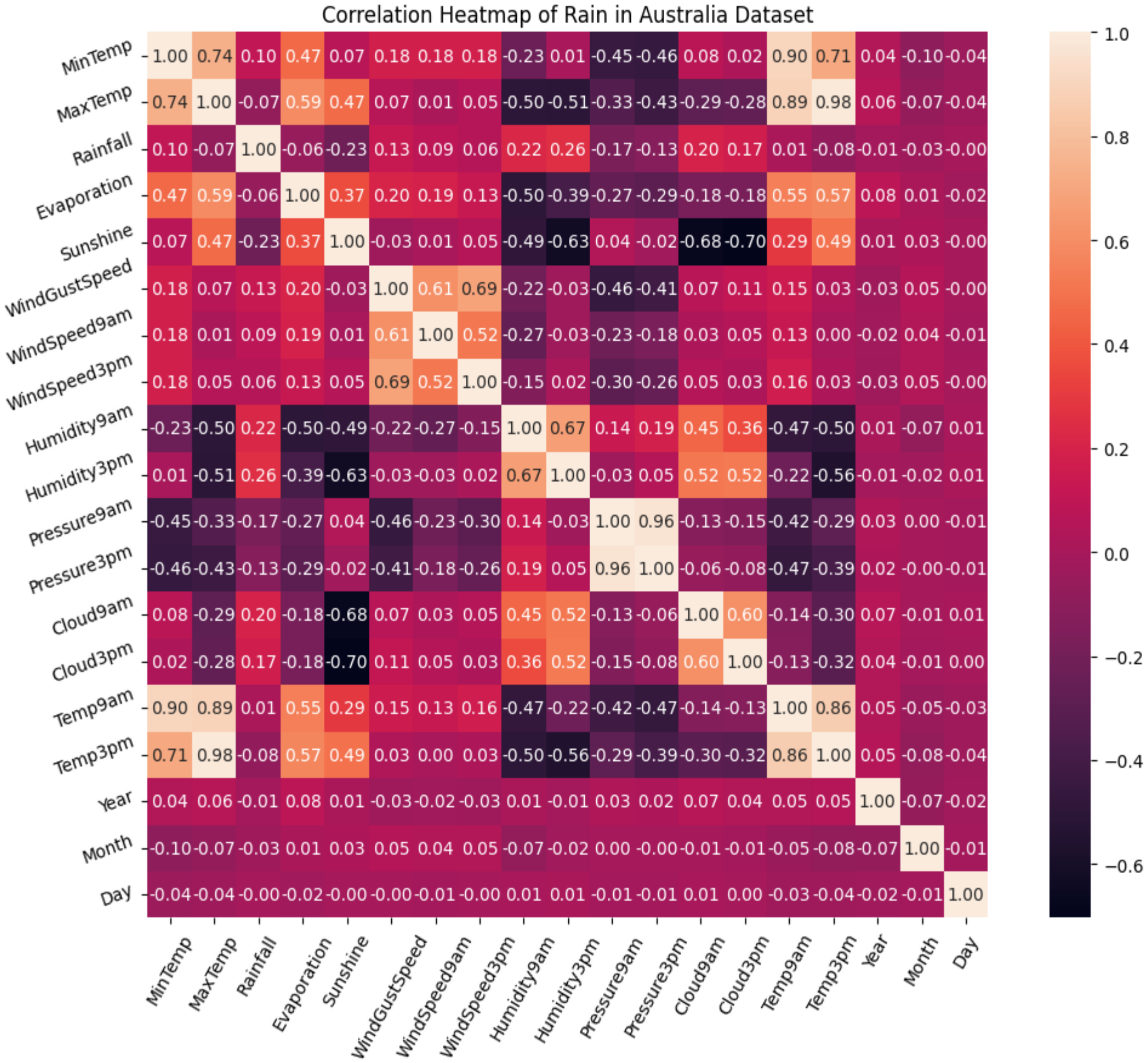


Figure 5. Correlation of Highly and weekly correlated features

Upon examining the correlation matrix of the "Rain in Australia" dataset, it is evident that some features exhibit strong correlations with others, while some show minimal correlations. Temp 9am stands out as the feature with the most significant correlations across the dataset. It has a very strong positive correlation with MinTemp (0.89) and MaxTemp (0.87), suggesting that morning temperatures are closely tied to both the minimum and maximum temperatures recorded. Additionally, temp 9am is highly correlated with Temp3pm (0.97), indicating that temperatures measured in the morning are strongly predictive of those in the afternoon. Beyond these, temp 9am also shows a moderate negative correlation with Humidity9am (-0.47) and Pressure 9am (-0.42), implying that higher morning temperatures are generally associated with lower humidity and pressure levels at the same time.

On the other end of the spectrum, the Day feature shows the least correlation with other features in the dataset. The correlations involving the Day feature are predominantly close to zero, indicating an insignificant linear relationship with other variables. This minimal correlation suggests that the day of the month does not have a substantial direct impact on other weather-related measurements in this dataset. Similarly, Month exhibits low correlations with most features, except for a slight positive correlation with Temp3pm (0.18), which remains relatively weak.

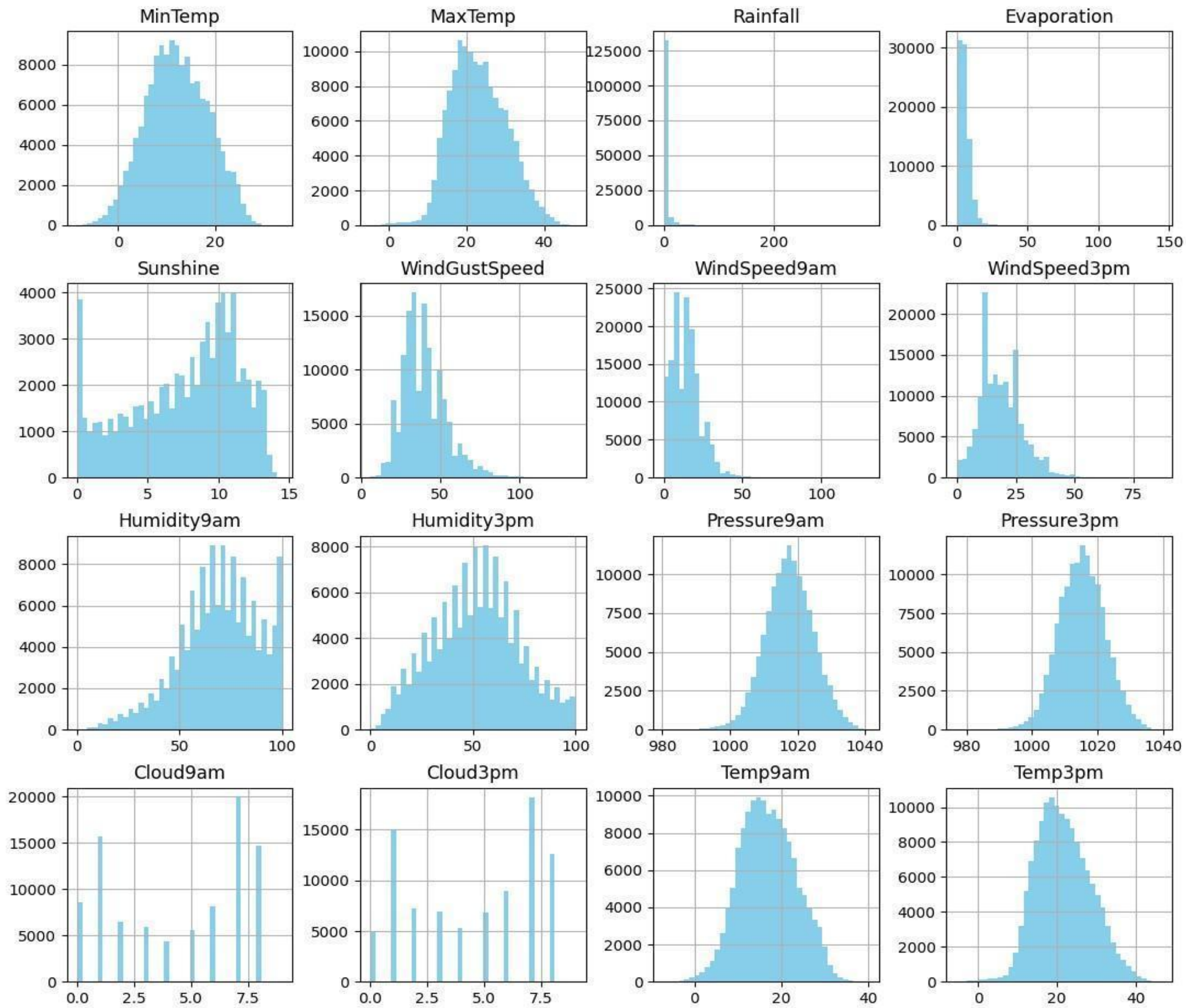


Figure 6. Histogram showing the dataset's attribute statistics

The histograms in figure 6 depict various weather features, each revealing unique distribution patterns. Temperature readings, both minimum and maximum, show a normal distribution with peaks around 10-20°C and 20-30°C, respectively. Rainfall and evaporation are highly right-skewed, with most values near zero. Sunshine hour's peak around 8-10 hours, while wind gust speeds concentrate around 35-45 km/h. Morning and afternoon wind speeds show peaks at 15-25 km/h. Humidity is high in the morning (80-90%) but drops by the afternoon (40-50%). Atmospheric pressure at both 9 am and 3 pm follows a normal distribution centered around 1015-1020 hPa. Cloud cover is variable at both times of the day, and temperatures at 9 am and 3 pm exhibit normal distributions, peaking at 15-20°C and 20-25°C, respectively. These patterns reflect typical weather variations over time.

5.2. Machine Learning Algorithms

A branch of artificial intelligence (AI) called machine learning (ML) is concerned with creating statistical models and algorithms [30] that let computers carry out tasks without direct human guidance. Rather, these systems make decisions, see patterns in data, and gradually get better at what they do as they gain experience. In this study, we employed supervised classification method for rainfall prediction, which is one of the machine learning approaches known as supervised learning, where the model is trained on labelled data. Here are some algorithms used in this study for rainfall prediction.

5.2.1. Decision Tree

Using a decision tree machine learning approach [30-31], problems with regression and classification are addressed. Its organization is comparable to a flowchart, where a decision rule is represented by each branch, an attribute or characteristic by each internal node, and a result or class label by each leaf node depict in figure 7.

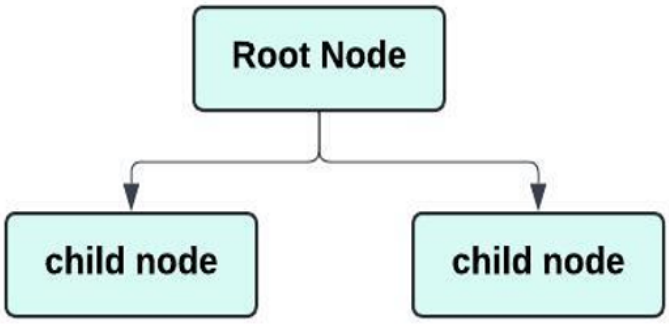


Figure 7. Representation of Decision Tree

5.2.2. Random Forest

Random Forest is an ensemble learning algorithm that constructs a multitude of decision trees during training and outputs the average prediction for regression tasks or the majority vote for classification tasks describe in equation 1. Random Forest is robust against over fitting and performs well on a variety of datasets. It can handle missing values and maintain accuracy even with a large number of features. Figure 17 depict the representation of Random Forest algorithm

$$\hat{y} = \text{mode}(\{fi(x)\}) \quad (1)$$

The following steps are how the algorithm for RF operates:

a) Select p data points at random from the training set.
b) Construct a decision tree linked to these p observation points.
c) Take N trees to build, and then repeat processes a and b.
d) For a new data point, forecast the value of y for the data point using each of the N tree trees, and then set the newly acquired data set's value equal to the mean of all the anticipated y values.

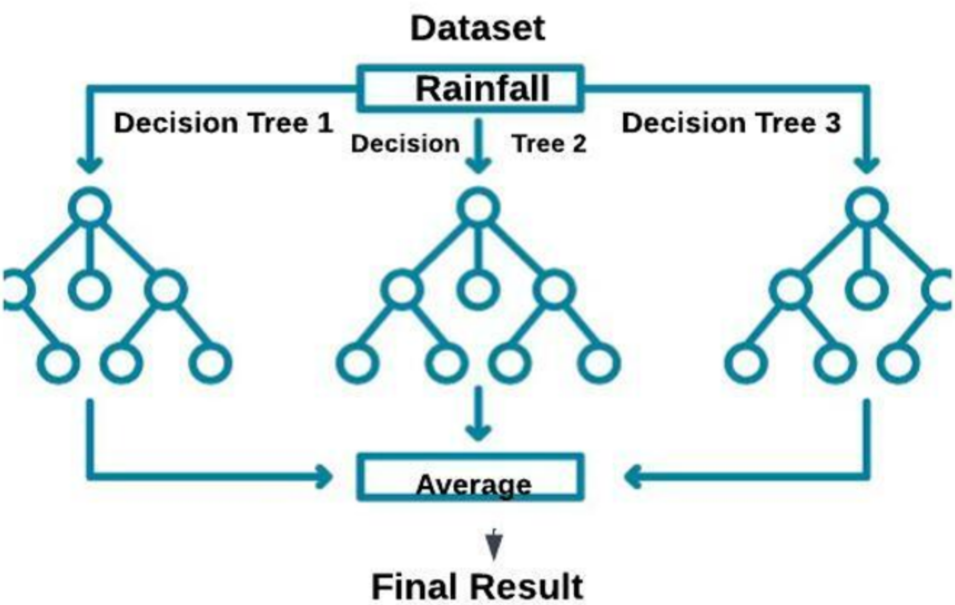


Figure 8. Representation of Random Forest Tree

5.2.3. Logistic Regression

Supervised approach of machine learning is logistic regression [32] examine the possibility of a result, an occurrence, or an observation in order to complete binary classification problems. The model produces a dichotomous or binary result with a maximum of two potential outcomes: either true or inaccurate, 0/1, or positive or negative.

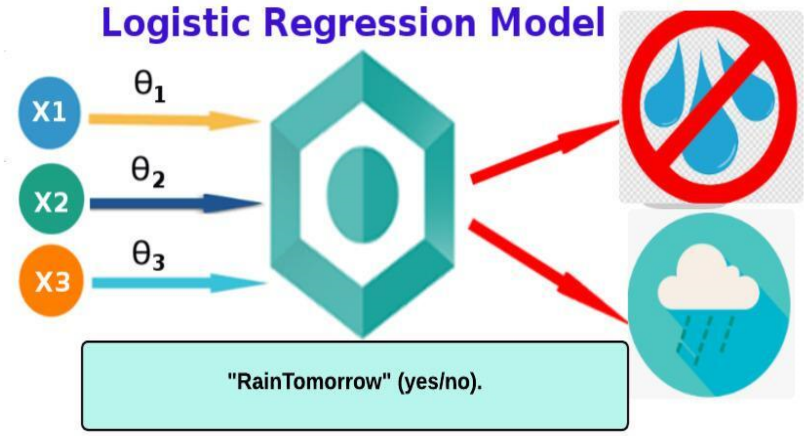


Figure 9. Representation of Logistic regression

In case of binary classification scenarios, an instance's probability of belonging to a certain class is predicted through a statistical model which is referred to as the logistic regression. In order to make sure that the best outcome is availed and also accurate forecast of rainfall is made, the work exploits the support vector machine (SVM) classifier and logistic regression. Though its name says the opposite, logistic regression is not a regression algorithm. It is an algorithm for classification. It uses the logistic function that Figure 4 shows to derive the probability of the positive class. The logistic function that would transform every real-valued input into a number between 0 and 1. The equation for the logistic function is shown in Equation 2:

$$\text{Sigmoid (z)} = \frac{1}{1+e^{-z}} \quad (2)$$

Where:

- z is a linear combination of the features and their respective coefficients, represented as
- z=β0+β1x_1+β2x_2+…+βn.x_n, where β0 is the intercept, βi are the coefficients, and xi are the feature values.

The logistic regression model calculates the odds of a binary result, such as one of zero or one. The logistic function predicts the likelihood that an occurrence is in the positive class (class 1):

5.2.4. K Nearest Neighbor

KNN is a simple instance-based learning algorithm used for both classification and regression tasks [33,34] as shown in fig 10. It classifies new instances based on the majority class of their k-nearest neighbors in the feature space. KNN is non-parametric and easy to understand, making it suitable for small datasets. It relies on distance metrics, such as Euclidean distance, to determine proximity between data points shown in equation 3.

$$\hat{y} = \text{majority}(\{Y_i \text{ for xi in } N_k(x)\}) \quad (3)$$

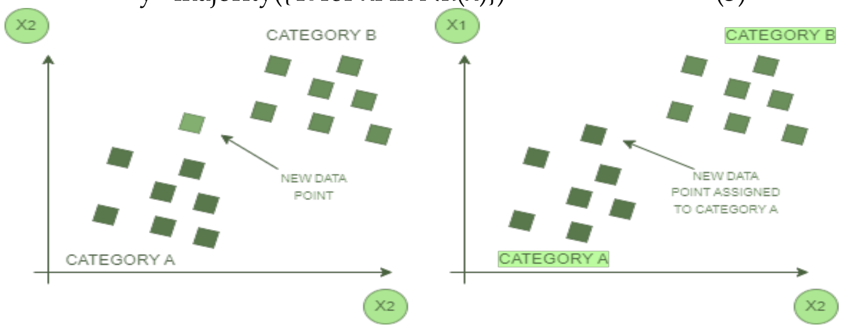


Figure 10. Representation of k nearest Neighbors

5.2.5. Naïve Bayes

The technique of the Naïve Bayes has been considered to be the most powerful and successful of the supervised machine learning and the data mining algorithms. Based on its attribute independence property, the Naive Bayes method is a probabilistic machine learning model that incorporates the Bayes' theorem. In classification tasks, it performs better and better until it reaches a high level of accuracy with its simplicity. The innovative Bayes theorem designates naive to only help one feature for a class, but it has less impact for another feature in the class. Although this is a strong and frequently implausible assumption, it streamlines the computation and increases the computing efficiency of the training data. Equation 4 illustrates the application of Bayes' theorem to determine the likelihood of a class given its attributes, based on this assumption.

P(C_K|x_1,x_2,...,x_n)=P(x_1,x_2,...,x_n)P(C_K)×∏i=1nP(xi|C_K)/ P(x_1,x_2 ,...,x_n.) … (4)

Where:

- P (C_K|x_1,x_2 ,...,x_n)) is the class's subsequent likelihood C_K. Given featuresx_1 ,x_2 ,...,x_n).
- P (C_K) is the preceding probability of classC_K.
- P (xi|C_K) is the likelihood of feature xigiven classC_K.
- P (x_1,x_2 ,..., x $_n$.) is the probability of the features occurring together.

5.3. Contribution of paper

This study contributes significantly to the methodology of rainfall prediction using machine learning techniques applied to meteorological data, leveraging the capabilities of the orange data mining tool. The research employs advanced feature selection methods to identify key variables crucial for accurate forecasting of "Rain Tomorrow" in Australian weather conditions. It evaluates and compares the performance of multiple machine learning algorithms including Decision Tree, K-Nearest Neighbors, Random Forest, Naive Bayes, and Logistic Regression within the Orange framework. Methodological advancements include iterative model refinement based on comprehensive performance metrics such as accuracy, precision, recall, and F1 score, thereby enhancing prediction outcomes. Insights gained from correlation analyses and feature importance assessments using Orange contribute to a deeper understanding of meteorological dynamics, with practical implications for improving weather forecasting systems and disaster preparedness strategies.

6. Results

For Australia country rainfall prediction tomorrow, orange platform is used for training and testing on machine learning algorithms. The figure 11 illustrates a machine learning workflow for predicting “Rainfall tomorrow” in Australia using the orange data mining tool. The process begins with loading the dataset, which includes various weather-related features such as temperature, humidity, wind speed, and historical rainfall data. Initial data exploration and preprocessing steps are performed using tools like the Data Table, Rank, Box Plot, and Line Plot to inspect, visualize, and understand the data's structure. Features are selected and sampled to ensure the model is trained on representative data, while visualizations like Distributions and Violin Plot help identify patterns and anomalies. Feature Statistics are calculated to provide insights into the data's distribution and variability. The workflow then proceeds to model training, employing several machine learning algorithms: Logistic Regression predicts the probability of rainfall based on input features, Naive Bayes uses a probabilistic approach assuming feature independence, k-Nearest Neighbors (KNN) classifies data points based on proximity to others, and Random Forest constructs multiple decision trees to improve accuracy and control overfitting. These models are used to make predictions on the dataset, which are then evaluated for their effectiveness. Evaluation involves the Test and Score widget, which compares model predictions against actual outcomes using metrics like accuracy, precision, recall, and F1 score. The Confusion Matrix provides a visual representation of classification performance, showing true positives, true negatives, false positives, and false negatives.

The provided screenshot from the orange tool displays the evaluation results for different machine learning models used for predicting rainfall, assessed using cross-validation. Among the models, Logistic Regression stands out as the best performer with the highest AUC (0.867), indicating excellent ability to distinguish between positive and negative classes. It also achieved the highest classification accuracy (0.847) and recall (0.847), making it the most reliable model for correctly identifying rainfall instances. Additionally, its F1 score (0.836) and precision (0.837) are the highest, demonstrating balanced performance across both precision and recall.

In terms of average performance, other models like Random Forest and k-Nearest Neighbors (kNN) also show strong results.

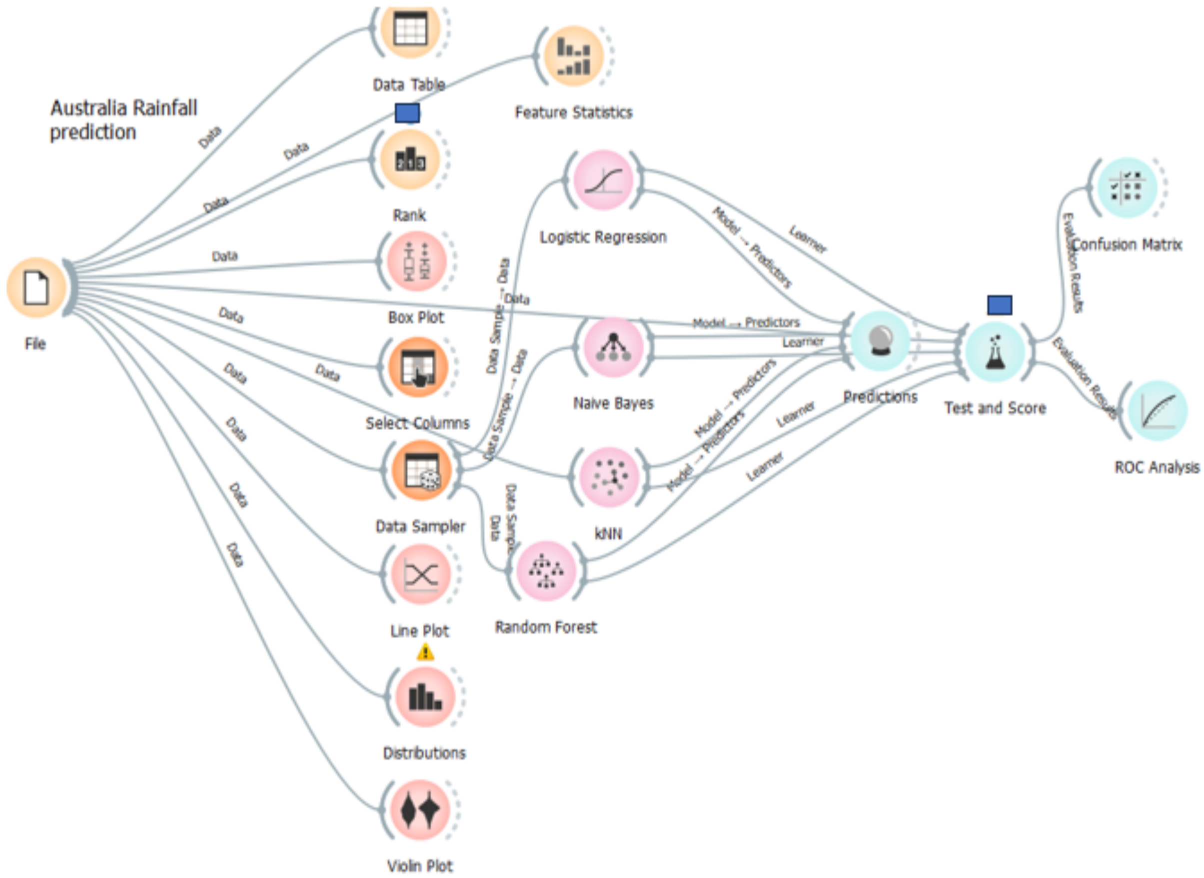
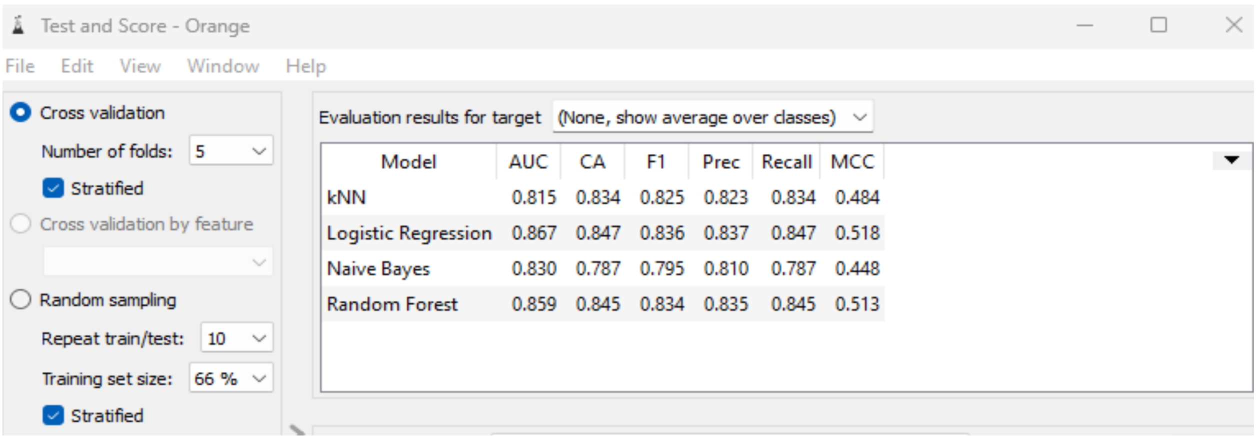


Figure 11. Representation of Australia rainfall prediction Models diagram using orange tool



Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.815	0.834	0.825	0.823	0.834	0.484
Logistic Regression	0.867	0.847	0.836	0.837	0.847	0.518
Naive Bayes	0.830	0.787	0.795	0.810	0.787	0.448
Random Forest	0.859	0.845	0.834	0.835	0.845	0.513

Figure 12. Representation of Machine learning Models Result using Orange Tool

Random Forest has a high AUC (0.859), accuracy (0.845), and recall (0.845), making it a close contender to Logistic Regression. The k Nearest Neighbor model has a respectable AUC (0.815), accuracy (0.834), and recall (0.834), indicating solid performance. Naive Bayes, while effective, performs slightly lower with an AUC of 0.830 and accuracy of 0.787. Overall, Logistic Regression emerged as the top-performing model, with Random Forest also providing robust results, while k Nearest Neighbor and Naive Bayes performed adequately but were outperformed by the other two models.

6.1. Confusion Matrix for k Nearest Neighbor

The confusion matrix in figure 13 indicates the performance of a classification model in predicting "Yes" and "No" outcomes. The model correctly predicted 102,603 "No" instances and 16,040 "Yes" instances, demonstrating a high accuracy for "No" predictions. However, it also incorrectly predicted 7,713 "Yes" instances as "No" and 15,837 "No" instances as "Yes," which suggests room for improvement in distinguishing between the two classes. The model performs better at predicting "No" outcomes than "Yes" outcomes, as evidenced by the higher number of true negatives compared to true positives and a considerable number of false negatives. Overall, while the model is fairly accurate, its ability to correctly identify "Yes" cases is less reliable, indicating a need for enhancement in the sensitivity towards "Yes" predictions.

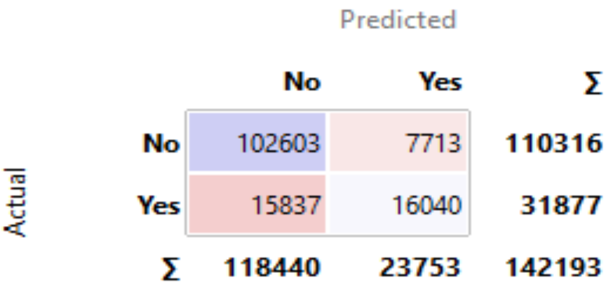


Figure 13. Representation of KNN Model Result

6.2. Confusion Matrix for Logistic regression

The confusion matrix in figure 14 shows the performance of a classification model in predicting "Yes" and "No" outcomes. The model accurately predicted 104,298 "No" instances and 16,101 "Yes" instances, indicating strong performance in both categories. However, it misclassified 6,018 "Yes" instances as "No" and 15,776 "No" instances as "Yes," highlighting some areas for improvement. The model correctly identified the majority of "No" cases, but a significant number of "No" predictions were made for actual "Yes" instances. While the overall accuracy appears reasonable, the false negatives and false positives suggest that the model's performance could be further optimized, especially in correctly identifying "Yes" cases. The model shows a good ability to predict "No" but less reliability in predicting "Yes," requiring attention to enhance sensitivity and reduce misclassification.

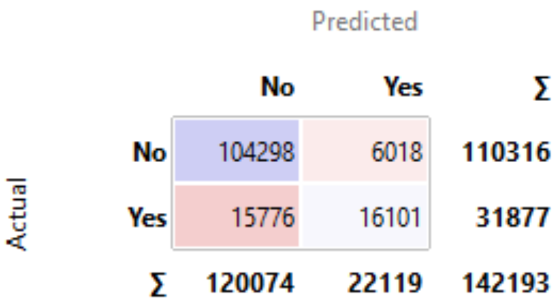


Figure 14. Representation of Logistic Regression Model Result

6.3. Confusion Matrix for Naive bayes

The confusion matrix in figure 15 illustrates the performance of a classification model in predicting "Yes" and "No" outcomes. The model accurately predicted 90,768 "No" instances and 21,105 "Yes" instances, demonstrating a reasonable degree of accuracy. However, it incorrectly classified 19,548 "Yes" instances as "No" and 10,772 "No" instances as "Yes." This indicates that while the model performs adequately, it has a significant number of misclassifications, particularly false negatives where actual "Yes" cases are predicted as "No." The model's ability to predict "No" is stronger, but there is a noticeable deficiency in correctly identifying "Yes" cases, as evidenced by the higher false negative rate. Overall, the results suggest that the model is better at identifying "No" instances than "Yes" instances, and there is room for improvement in enhancing its sensitivity and specificity to reduce both false positives and false negatives.

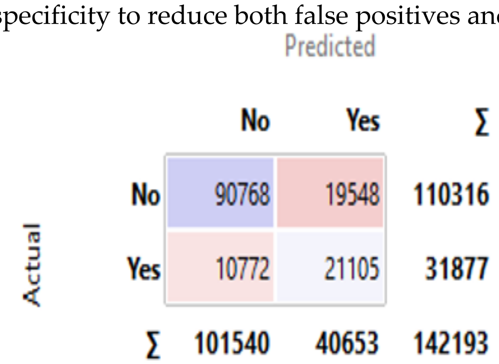


Figure 15. Representation of Naïve Bayes Model Result

6.4. Confusion Matrix for Random Forest

The classification model's ability to anticipate "Yes" and "No" outcomes is displayed in the confusion matrix. 104,276 cases were correctly classified by the model as "No" out of all the predictions, and 15,897 instances were correctly forecasted as "Yes." It also produced 15,980 false negative predictions—classifying "Yes" occurrences as "No"—and 6,040 false positive predictions—erroneously classifying "No" examples as "Yes." Regarding overall performance, the model has a strong accuracy of almost 84%. The large number of true negatives (104,576) indicates that the model performs a good job of identifying "No" situations. But the high number of false negatives (15,980) suggests that the model has trouble correctly classifying "Yes" occurrences, which, depending on the specifics of the classification assignment, may be quite important. Although the accuracy is fairly good, there is still potential for improvement in the performance when it comes to predicting "Yes" outcomes. Specifically, there is room to reduce the frequency of false negatives in order to achieve a better balance between properly recognizing both "Yes" and "No" occurrences.

Actual \ Predicted	No	Yes	Σ
No	104276	6040	110316
Yes	15980	15897	31877
Σ	120256	21937	142193

Figure 16. Representation of Random Forest Model Result

7. Evaluation Criteria

The metrics (or key indicators of performance (Key Performance Factors)) that will be used to evaluate the algorithms output are described in this section.

7.1. Accuracy

Number reflecting how well the predicted model performed. The accuracy formula shown in equation (5):

$$Accuray = \frac{TP + TN}{TP+TN+FN+FP} \quad (5)$$

Where:

TP designates it as "true positives." "Result where the model outputs a positive class and correctly classifies it. FP is called a False Positive. Lead to a case in which the positive class is erroneously designated by the model as a negative class. TN, or the true negative, connotes the outcome where the model predicted the negative class to be. False negative, i.e. FN is a concept in detection that is associated with negative. is a situation where the model predicts the other class to be wrong.

7.2. Precision

The percentage of instances that are correctly identified as positive is known as precision. Which is, whether a model forecasts positive numbers. The formula shown in equation (6)

$$Precision = \frac{TP}{TP+FN} \quad (6)$$

7.3. Recall

The percentage of correctly detected positives to all positives is known as recall. The sensitivity formula and this formula are identical. As shown by equation (7):

$$Recall = \frac{TP}{TP+FP} \quad (7)$$

7.4. F1 score

When precision and recall are insufficient for evaluating performance, for as when one mining method has better accuracy but worse recall than another, the question of which algorithm is superior may come up. The F-measure, which gives the mean of recall and precision, can be used to address this problem. An industry standard for evaluating the performance of a classification model is the F1 score. Equation (8) shows how the computation appears. It provides an equitable evaluation of a model's accuracy by merging recall and precision into a single metric.

$$F1\ score = \frac{2 * (precision * recall)}{precision + recall} \quad . \tag{8}$$

8. Conclusion and Future work

Estimating rainfall is crucial for managing water supplies, preserving human life, and protecting the environment. Given the influence of geographic and regional factors on rainfall estimation, inaccuracies and insufficiencies in prediction can arise. This research leverages data analytics for weather prediction, specifically focusing on Australia, to address these challenges. The study aims to evaluate the effectiveness of machine learning methods in predicting precipitation, with "rain tomorrow" as the target variable. Several Supervised machines learning models, including Random Forest (RF) and Neural Networks (NN), were employed to predict rainfall using collected datasets. The results show that random forests have high accuracy (0.859) for predicting rainfall.

The study highlights how machine learning algorithms, can accurately model the nonlinear nature of natural processes. Furthermore, the research shows that these algorithms are more effective when data is analyzed on a city-by-city basis, allowing for localized understanding of rainfall patterns. The research suggests that future work could include analyzing data from various countries and incorporating weather observations from 2019 to the present. Gathering more region-specific data and better tracking variables such as cloud cover, sunshine, and temperature could enhance model accuracy. Additionally, the research emphasizes the importance of feature engineering and data preprocessing in improving model performance. By retaining relevant features and carefully handling missing values, outliers, and temporal dependencies, the predictive power of these models can be enhanced.

In conclusion, the study extends current knowledge on weather forecasting and underscores the relevance of machine learning techniques in environmental solutions. The improvements achieved in climate modeling, disaster planning, agriculture, and water resource management demonstrate the importance of high-precision rain forecasts for risk assessment and decision-making.

References

1. Dash, Y., Mishra, S. K., & Panigrahi, B. K. (2018). Rainfall prediction for the Kerala state of India using artificial intelligence approaches. Computers & Electrical Engineering, 70, 66-73.
2. Masafumi Goto, Faizah Cheros, Nuzul Azam Haron and Mohd Nasrun Mohd Nawi"Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review",IOP Conference Series: Earth and Environmental Science,2020.
3. Efrat Morin,"Flood forecasting with machine learning models in an operational framework",2021,Hydrology and Earth System Sciences.
4. Prachatos Mitra; Ronit Ray; Retabrata Chatterjee; Rajarshi Basu; Paramartha Saha; Sarnendu Raha; Rishav Barman; Sarnendu Raha"Flood forecasting using Internet of things and artificial neural networks",2016 IEEE
5. Jeerana Noymanee, Nikolay O. Nikitin, Anna V. Kalyuzhnaya,"Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin",2017,Elsevier.
6. Chen Chen a, Jiange Jiang , Zhan Liao , Yang Zhou , Hao Wang , Qingqi Pei a,"A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China",Journal of Hydrology Elsevier 2022
7. Khabat Khosravi , Himan Shahabi , Binh Thai Pham , Jan Adamowski , Ataollah Shirzadi , Biswajeet Pradhan , Jie Dou , Hai-Bang Ly , Gyula Gróf , Huu Loc Ho , Haoyuan Hong , Kamran Chapi , Indra Prakash,"A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods",2019 elsevier
8. Sherif Ahmed Abu El-Magd, Biswajeet Pradhan & Abdullah Alamri ,"Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt",Springer,Arabian Journal of Geosciences volume, 2021
9. Monalisa Nayak, Soumya Das & Manas Ranjan Senapati "Improving Flood Prediction with Deep Learning Methods",Springer2022 Journal of The Institution of Engineers (India):
10. Gokmen Tayfur ,Vijay P. Singh ,Tommaso Moramarco andSilvia Barbetta ,"Flood Hydrograph Prediction Using Machine Learning Methods",MDPIWater 2018
11. Abinash Sahoo, Sandeep Samantaray & Dillip K. Ghose,"Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study",Journal of the Geological Society of India ,2021 springer volume
12. Kun Qian, Abduallah Mohamed, Christian Claudel," Physics Informed Data Driven model for Flood Prediction: Application of Deep Learning in prediction of urban flood development,Arxiv 2019 Electrical Engineering and Systems Science,Signal Processing
13. Mohammed Sarfaraz Gani Adnan a, Zakaria Shams Siam d g, Irfat Kabir a, Zobaidul Kabir e, M. Razu Ahmed f, Quazi K. Hassan f, Rashedur M. Rahman b, Ashraf Dewan,"A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction", Elsevier 2023 Journal of Environmental Management,
14. Scott Miau, Wei-Hsi Hung, "River Flooding Forecasting and Anomaly Detection Based on Deep Learning",IAES International Journal of Artificial Intelligence (IJ-AI),2020
15. Saleh, M. A., Rasel, H. M., & Ray, B. (2024). A comprehensive review towards resilient rainfall forecasting models using artificial intelligence techniques. Green Technologies and Sustainability, 100104.
16. Yu, N., & Haskins, T. (2021, July). Bagging machine learning algorithms: A generic computing framework based on machine-learning methods for regional rainfall forecasting in upstate New York. In Informatics (Vol. 8, No. 3, p. 47). MDPI.
17. Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., ... & Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. Sensors, 22(9), 3504.
18. Sani, N. S., Abd Rahman, A. H., Adam, A., Shlash, I., & Aliff, M. (2020). Ensemble learning for rainfall prediction. International Journal of Advanced Computer Science and Applications, 11(11).
19. Hassan, M. M., Rony, M. A. T., Khan, M. A. R., Hassan, M. M., Yasmin, F., Nag, A., ... & El-Shafai, W. (2023). Machine Learning-Based Rainfall Prediction: Unveiling Insights and Forecasting for Improved Preparedness. IEEE Access, 11, 132196-132222.
20. Maspo, N. A., Harun, A. N. B., Goto, M., Cheros, F., Haron, N. A., & Nawi, M. N. M. (2020, June). Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. In IOP Conference Series: Earth and Environmental Science (Vol. 479, No. 1, p. 012038). IOP Publishing.
21. Alam, F., Salam, M., Khalil, N. A., khan, O., & Khan, M. (2021). Rainfall trend analysis and weather forecast accuracy in selected parts of Khyber Pakhtunkhwa, Pakistan. SN Applied Sciences, 3, 1-14.

22. Abdul-Kader, H., & Mohamed, M. (2021). Hybrid machine learning model for rainfall forecasting. Journal of Intelligent Systems and Internet of Things, 1(1), 5-12.
23. Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016). Rainfall prediction: A deep learning approach. In Hybrid Artificial Intelligent Systems: 11th International Conference, HAIS 2016, Seville, Spain, April 18-20, 2016, Proceedings 11 (pp. 151-162). Springer International Publishing.
24. J. Abbot, J. Marohasy, Application of artificial neural networks to rainfall forecasting in Queensland, Australia, Adv. Atmos. Sci. 29 (2012) http://dx.doi. org/10.1007/s00376-012-1259-9.
25. Hossain, H.M. Rasel, F. Mekanik, M.A. Imteaz, Artificial neural network modelling technique in predicting Western Australian seasonal rainfall, Int. J. Water 14 (1) (2020) 14–28, http://dx.doi.org/10.1504/IJW.2020.112711.
26. P. Hunasigi, S. Jedhe, M. Mane, V. Patil-Shinde, Multilayer perceptron neural network based models for prediction of the rainfall and reference crop evapo- transpiration for sub-humid climate of Dapoli, Ratnagiri district, India, Acta Ecol. Sinica 43 (1) (2023) 154–201, http://dx.doi.org/10.1016/j.chnaes.2022.09.004
27. W.M. Ridwan, M. Sapitang, A. Aziz, K.F. Kushiar, A.N. Ahmed, A. El-Shafie, Rainfall forecasting model using machine learning methods: Case study Tereng- ganu, Malaysia, Ain Shams Eng. J. 12 (2) (2021) 1651–1663, http://dx.doi.org/ 10.1016/j.asej.2020.09.011.
28. A.Y. Barrera-Animas, L.O. Oyedele, M. Bilal, T.D. Akinosho, J.M.D. Delgado, L.A. Akanbi, Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting, Mach. Learn. Appl. 7 (2021) (2022) 100204, http://dx.doi.org/10.1016/j.mlwa.2021.100204
29. Kotz, M., Levermann, A., & Wenz, L. (2022). The effect of rainfall changes on economic production. Nature, 601(7892), 223-227.
30. Parmar, A., Mistree, K., & Sompura, M. (2017, March). Machine learning techniques for rainfall prediction: A review. In International conference on innovations in information embedded and communication systems (Vol. 3).
31. Mohammed, M., Kolapalli, R., Golla, N., & Maturi, S. S. (2020). Prediction of rainfall using machine learning techniques. International Journal of Scientific and Technology Research, 9(01), 3236-3240.
32. Sanadi, M. M. (2021). Rainfall Prediction Using Logistic Regression and Support Vector Regression Algorithms. In Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I 5 (pp. 617-626). Springer International Publishing.
33. Biruntha, S., Sowmiya, B. S., Subashri, R., & Vasanth, M. (2022, March). Rainfall prediction using kNN and decision tree. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1757-1763). IEEE.
34. Rajab, A., Farman, H., Islam, N., Syed, D., Elmagzoub, M. A., Shaikh, A., ... & Alrizq, M. (2023). Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh. Water, 15(22), 3970.
35. Farman, H., Khan, A. W., Ahmed, S., Khan, D., Imran, M., & Bajaj, P. (2024). An Analysis of Supervised Machine Learning Techniques for Churn Forecasting and Component Identification in the Telecom Sector. Journal of Computing & Biomedical Informatics, 7(01), 264-280.
36. Ammar Ahmad Khan , Muhammad Arslan , Ashir Tanzil , Rizwan Abid Bhatty , Muhammad Asad Ullah Khalid , Ali Haider Khan. (2024). Classification Of Colon Cancer Using Deep Learning Techniques On Histopathological Images. Migration Letters, 21(S11), 449–463
37. Farman, H., & Ahmed, S. (2023). Novel framework for efficient detection of QRS morphology for the cardiac arrhythmia classification. Journal of Computing & Biomedical Informatics, 5(02), 12-20.
38. Baig, H. F. A., Islam, N., & Alim, A. (2023). Exploring Exploring Machine Learning and Deep Learning Approaches for Disaster Prediction and Management: A Survey of Different Approaches: A Survey of Different Approaches for Disaster prediction and management through tweets. Pakistan Journal of Engineering, Technology & Science, 11(1), 45-73.
39. Khan, A. H., Malik, H., Khalil, W., Hussain, S. K., Anees, T., & Hussain, M. (2023). Spatial Correlation Module for Classification of Multi-Label Ocular Diseases Using Color Fundus Images. Computers, Materials & Continua, 76(1).
40. Mahmood, M. A., Malik, H., Khan, A. H., Adnan, M., & Khan, M. I. A. (2022). Neural Network-Based Prediction of Potential Ribonucleic Acid Aptamers to Target Protein. Journal of Computing & Biomedical Informatics, 4(01), 21-36.
41. Abbas, F., Iftikhar, A., Riaz, A., Humayon, M., & Khan, M. F. (2024). Use of Big Data in IoT-Enabled Robotics Manufacturing for Process Optimization. Journal of Computing & Biomedical Informatics, 7(01), 239-248.
42. Muhammad Kaleem , Muhammad Azhar Mushtaq , Uzair Jamil , Sadaqat Ali Ramay , Tahir Abbas Khan , Siraj Patel , Rizwan Zahidy , Sayyid Kamran Hussain. (2024). New Efficient Cryptographic Techniques For Cloud Computing Security. Migration Letters, 21(S11), 13–28. Retrieved from https://migrationletters.com