




Article

Towards Resilient Agriculture to Hostile Climate Change in the Sahel Region: A Case Study of Machine Learning-Based Weather Prediction in Senegal

Chimango Nyasulu ^{1,*}, Awa Diattara ¹, Assitan Traore ², Abdoulaye Deme ³ and Cheikh Ba ¹

¹ LANI (Laboratoire d'Analyse Numérique et Informatique), University of Gaston Berger, Saint-Louis 32000, Senegal

² Business & Decision, 38000 Grenoble, France

³ Laboratoire des Sciences de l'Atmosphère et de l'Océan, Unité de Formation et de Recherche de Sciences Appliquées et de Technologie, University of Gaston Berger, Saint-Louis 32000, Senegal

* Correspondence: nyasulu.chimango@ugb.edu.sn

Abstract: To ensure continued food security and economic development in Africa, it is very important to address and adapt to climate change. Excessive dependence on rainfed agricultural production makes Africa more vulnerable to climate change effects. Weather information and services are essential for farmers to more effectively survive the increasing occurrence of extreme weather events due to climate change. Weather information is important for resource management in agricultural production and helps farmers plan their farming activities in advance. Machine Learning is one of the technologies used in agriculture for weather forecasting and crop disease detection among others. The objective of this study is to develop Machine Learning-based models adapted to the context of daily weather forecasting for Rainfall, Relative Humidity, and Maximum and Minimum Temperature in Senegal. In this study, we made a comparison of ten Machine Learning Regressors with our Ensemble Model. These models were evaluated based on Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and Coefficient of Determination. The results show that the Ensemble Model performs better than the ten base models. The Ensemble Model results for each parameter are as follows; Relative Humidity: Mean Absolute Error was 4.0126, Mean Squared Error was 29.9885, Root Mean Squared Error was 5.4428 and Coefficient of Determination was 0.9335. For Minimum Temperature: Mean Absolute Error was 0.7908, Mean Squared Error was 1.1329, Root Mean Squared Error was 1.0515 and Coefficient of Determination was 0.9018. For Maximum Temperature: Mean Absolute Error was 1.2515, Mean Squared Error was 2.8038, Root Mean Squared Error was 1.6591 and Coefficient of Determination was 0.8205. For Rainfall: Mean Absolute Error was 0.2142, Mean Squared Error was 0.1681, Root Mean Squared Error was 0.4100 and Coefficient of Determination was 0.7733. From the present study, it has been observed that the Ensemble Model is a feasible model to be used for Rainfall, Relative Humidity, and Maximum and Minimum Temperature forecasting.

Keywords: Africa; food security; weather forecasting; machine learning; regressors; ensemble model



Citation: Nyasulu, C.; Diattara, A.; Traore, A.; Deme, A.; Ba, C. Towards Resilient Agriculture to Hostile Climate Change in the Sahel Region: A Case Study of Machine Learning-Based Weather Prediction in Senegal. *Agriculture* **2022**, *12*, 1473. <https://doi.org/10.3390/agriculture12091473>

Academic Editors: Xiuliang Jin, Hao Yang, Zhenhai Li, Changping Huang and Dameng Yin

Received: 20 July 2022

Accepted: 18 August 2022

Published: 15 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Globally there are growing worries about climate change. Developing countries, mainly in Sub-Saharan Africa (SSA), are more vulnerable to the predicted effects of climate change because of their excessive reliance on the climate for development, the main pillar of which is agriculture [1]. Agricultural production is likely to be severely affected by climate change [2]. It is clearly visible that climate change is already happening: unreliable and unpredictable rainfall, prolonged dry spells, floods, rising temperatures, and the eruption of crop pests and diseases among others [3–6]. Climate change is defined as “fluctuations in the patterns of climate over long periods” [7].

Most of the agricultural production in Africa comes from smallholder rainfed production [8]. In Africa, agriculture remains the main source of food and economic development making it a fundamental component of programs to reduce poverty and ensure food security [9]. Consequently, meeting the demand for required food is a vital precondition for successful economic development. Agriculture is vastly reliant on climate variables such as rainfall, humidity and temperature; thus, agriculture is predicted to be the most affected by climate change [7]. According to projections by the Food and Agriculture Organization of the United Nations (2008) “by 2080, agricultural output in developing countries may decline by 20 percent due to climate change and yields in developing countries could further decrease by 15 percent on average by 2080”. In Africa, agriculture constitutes more than 65 percent of the labor force, contributing over 30 percent of the gross domestic product. Largely, climate change affects the four main areas of food security: availability, stability, access, and utilization [10].

1.1. Weather Forecasting and Usefulness of Weather Information in Agriculture

Weather forecasting is defined as “foretelling the condition of the atmosphere for a specific location using principles of physics, statistics, empirical techniques and technology. It also includes changes on the surface of the earth produced by atmospheric circumstances” [10,11]. There are several conventional methods of weather forecasting such as persistence forecasting, climatology forecasting, looking at the sky, using barometric pressure, using the atmospheric model and ensemble forecasting [10]. Persistence forecasting assumes a continuance of the present, and largely relies on the occurrence of a stagnant weather pattern. Climatology forecast supposes that weather for a specific day at a place does not vary much from year to year. Looking at the sky entails observing the sky to forecast weather. For example, the inspissation of cloud cover could be a sign of rain in the near future [11]. Using barometric pressure, a change in barometric pressure results in a change in weather. For example, the greater the change in pressure, the greater the change in weather. The atmospheric model contains a collection of equations that are used to forecast the upcoming state of the atmospheric parameters. Whereas ensemble forecasting requires the production of several forecasts in order to take into account the uncertainty (errors in the observations and insufficient sampling) in the initial state of the atmosphere [11].

Machine Learning (ML) models can be used in many areas in the workflow of weather forecasting such as observations, data assimilation, numerical weather forecasting, and post-processing and dissemination [12,13]. Machine Learning has the potential to address the challenges of complexity and volume when dealing with meteorological data while using less computing power. Among others, notable advantages of ML models in weather forecasting are increasing accuracy and speed. Machine Learning is a subset of Artificial Intelligence (AI). Machine Learning is the “study of algorithms and statistical models that computer systems mainly use to perform a specific task without using explicit instructions or being programmed” [14].

Availability of weather information support farmers to make reasonable choices in agricultural production. This allows farmers to better choose strategies that take into account climate variability, change and adaptation. Additionally, weather information is important for resource management in agricultural production. To ensure maximum crop yield, it is necessary to adapt and respond properly to climate change [15–17].

1.2. Problem Statement

Africa is more prone to climate change because of over-reliance on smallholder rainfed agriculture. Mostly, agricultural activities are controlled by rainfall quantities and distribution. Clearly, climate change is no longer a future problem as it is already negatively affecting agricultural productivity and food security in the world and particularly in Africa, and therefore requires prompt attention.

1.3. Objectives

To ensure food security and economic development in Africa, tackling and adapting to climate change is paramount. The main objective of this study is to develop ML-based models adapted to the context of daily weather forecasting (Rainfall, Relative Humidity, and Maximum and Minimum Temperature) in Senegal. These parameters are considered to be primary weather attributes required for crop production. The goal is to determine if the ML models will be able to forecast weather in accordance with the spatial distribution and the annual cycle in Senegal. To achieve the main objective, the study will focus on the following sub-objectives:

1. To determine the spatial weather distribution in Senegal;
2. To determine the weather annual cycle in Senegal.

1.4. Brief Literature Review

Several studies [18–20] have used ML Regressors for weather forecasting. In Nilgiris district, Tamil Nadu, India, a study was conducted by [18] to predict rainfall intensity based on humidity, wind direction, daily temperature, wind speed and cloud speed. Data from 2005 to 2014 was collected from India Meteorological Department. In this work, Decision Tree (DT), Support Vector Regression (SVR) and Random Forest (RF) were used. Models were evaluated using the Coefficient of Determination (R^2) and adjusted R^2 . Results for SVR are R^2 0.814 and adjusted R^2 0.806, DT are R^2 0.904 and adjusted R^2 0.900 and RF are R^2 0.981 and adjusted R^2 0.980.

Authors in [19] compared 36 Regressors for indoor temperature forecasting for three consecutive hours in a smart building based on every 10 min real data collected from a smart building and a weather station. The ML models used include Random Forest, Extra Trees and Gradient Boosting Machine. The models were evaluated using Root Mean Squared Error (RMSE) and correlation coefficient. The results show that the Extra Trees regressor obtained a higher correlation coefficient of 0.97 and a lower RMSE of 0.058.

A study by [20] was conducted in Ethiopia to predict daily rainfall using three ML Regressors: Multivariate Linear Regression (MLR), Random Forest (RF), and gradient descent XGBoost. A dataset was obtained from the weather station comprised of features such as rainfall, humidity, maximum temperature, evaporation, minimum temperature, sunshine, wind speed, year, month and date from 1999 to 2018. The Mean Absolute Error (MAE) results for RF, MLR and XGBoost are 4.49, 4.97, and 3.58, respectively. The RMSE results for RF, MLR and XGBoost are 8.82, 8.61, and 7.85, respectively.

1.5. Description of the Study Area and Data Source

Senegal is one of the sub-Saharan African countries found in westernmost Africa. Senegal shares borders with Mauritania in the north, Guinea Conakry and Guinea Bissau in the south and Mali in the east. Gambia river constitutes an enclave nearly 300 km deep within Senegalese territory. Senegal experiences a Sudano-Sahelian climate: the climate is tropical in the south and semi-desert in the north. The dry season is experienced from November to mid-June and a humid and hot season from mid-June to October. Rainfall decreases from the south to the north with yearly variations. It goes from 1200 mm in the south to 300 mm in the north [21–23].

Senegal is substantially hotter and since 1975 temperatures have increased by almost 0.9 °C. Global warmers have increased the occurrence of droughts, reducing crop yields and pasture availability. Agriculture is the main source of food and economic development. Agriculture employs over 70 percent of the population and accounts for about 12.4 percent of the gross domestic product. Largely, agriculture is rainfed and depends on seasonal rainfall. Food staples include maize, millet, rice and sorghum, and usually, groundnut and cotton are cultivated for export [21–23]. Figure 1 presents a map of the study area.

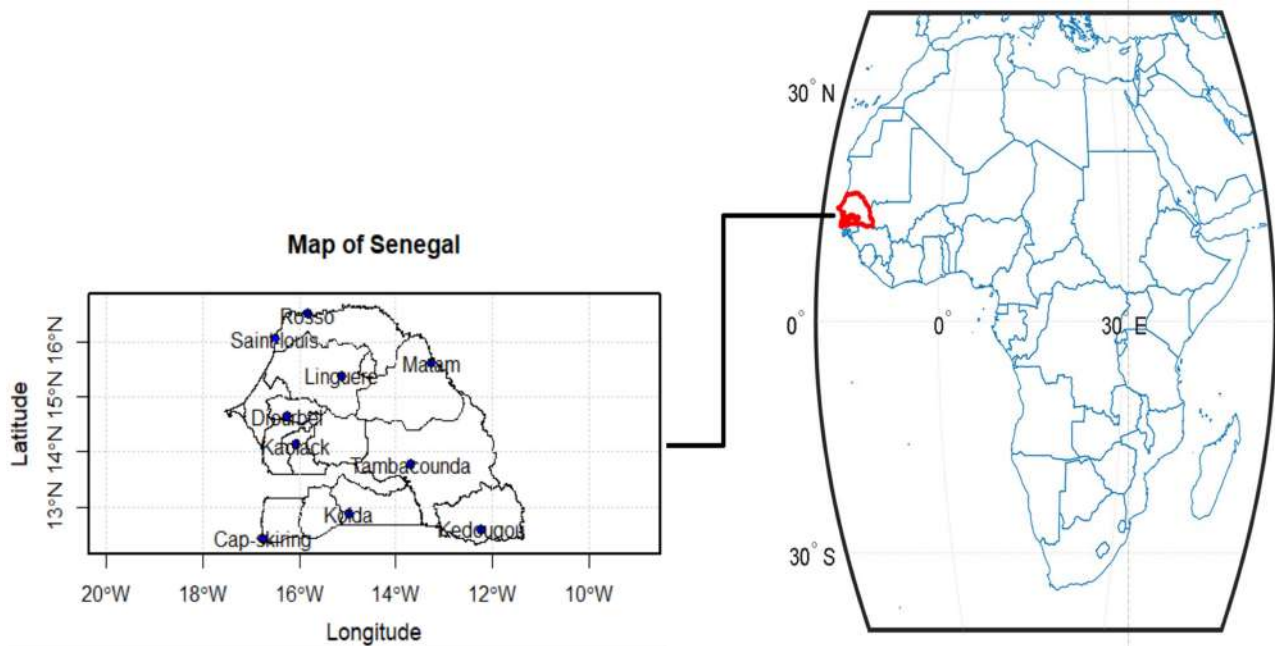


Figure 1. Map of Senegal with its ten data source locations (at left).

1.6. Dataset

The dataset used in this study was from ten selected stations from the South, Center and North of Senegal as presented in Figure 1. The dataset was obtained from the NASA Langley Research Center’s POWER Project, funded by NASA Earth Science/Applied Science Program (<https://power.larc.nasa.gov/data-access-viewer/>) (accessed on 26 July 2021) and spans from 1982 to 2020. The following steps were followed to access the dataset: selecting user community, specifying temporal average, entering latitude and longitude of the location, specifying period, specifying file output format and selecting parameters. Parameters accessed were Year, Month, Day, Relative Humidity at 2 Meters (%), Precipitation (mm/day), Minimum Temperature at 2 Meters (°C) and Maximum Temperature at 2 Meters (°C). Longitude and Latitude details of selected locations are summarized in Table 1.

Table 1. Summary of Latitude and Longitude of selected locations.

Station	Latitude	Longitude
Rosso	16.5 N	−15.817 W
Saint-Louis	16.05 N	−16.483 W
Cap-skiring	12.4 N	−16.75 W
Diourbel	14.65 N	−16.233 W
Kaolack	14.133 N	−16.067 W
Kedougou	12.567 N	−12.217 W
Kolda	12.883 N	−14.967 W
Linguere	15.367 N	−15.117 W
Matam	15.617 N	−13.25 W
Tambacounda	13.767 N	−13.683 W

1.7. Exploratory Data Analysis

In this study, exploratory data analysis was performed to observe the spatial weather distribution and annual cycle. In this regard, exploratory data analysis for one station from each geographical region will be presented. The selected locations are Kolda from the south, Kaolack from the center and Matam from the north as presented in Figure 2. The parameters of interest are Rainfall, Relative Humidity, Maximum and Minimum Temperature.

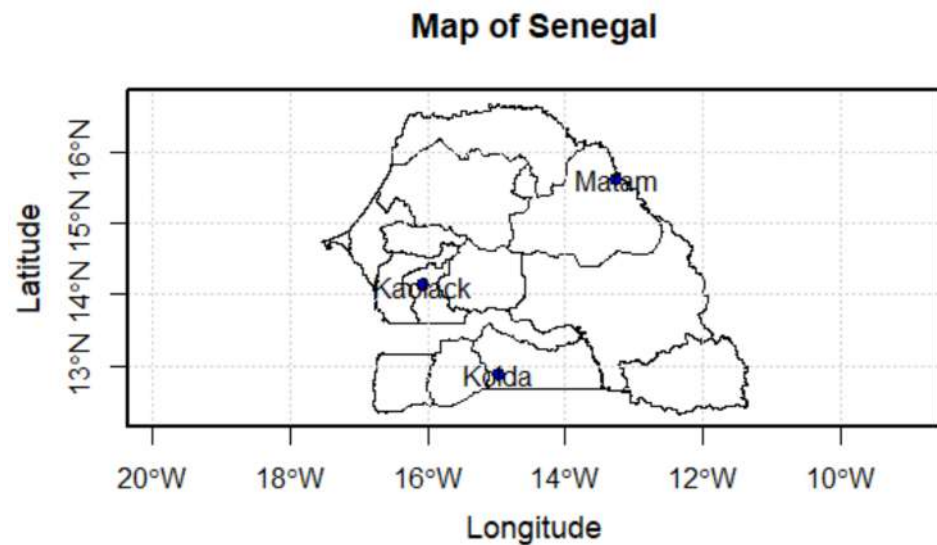


Figure 2. Map of Senegal showing locations for exploratory data analysis.

1.7.1. Spatial Weather Distribution

In Senegal, there are four notable rainfall zones constructed on cumulative rainfall characteristics. There is a northern zone (NZ), central north zone (CN), central south zone (CS) and southern zone (SZ). The average yearly cumulative rainfall is below 400 mm in NZ, between 400–600 mm in CN, between 600–800 mm in CS and more than 800 mm in SZ. Comparatively, it is noted that rainfed agricultural production is very difficult in the NZ because of repeated rainfall shortfalls [4].

In general, rainfall increases from north to south. As presented in Figure 3, yearly cumulative rainfall in Kolda is mostly above 800 mm, mostly between 600–800 mm in Kaolack and mostly below 400 mm in Matam.

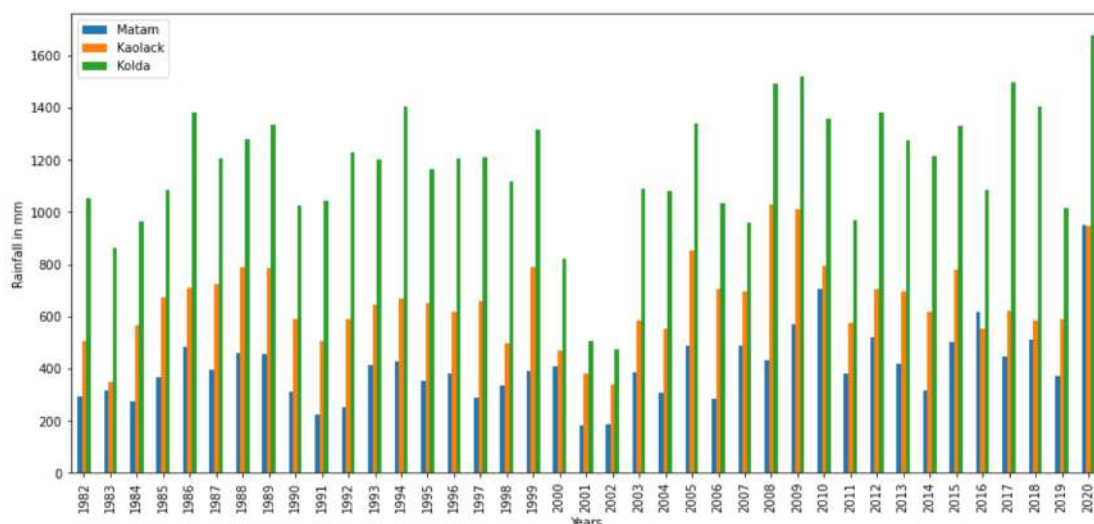


Figure 3. Annual cumulative rainfall variation for Matam, Kaolack and Kolda (1982–2020).

Generally, in Senegal, temperatures are between hot to extremely hot from the south to the north. This is because the northern part has a warm desert climate. Relative Humidity is considerably high in the south with an annual average of above 50 percent, an annual average of below 50 percent in the center and below 40 percent in the north [4]. Figures 4–6 summarize annual average Maximum Temperature, Minimum Temperature and Relative Humidity, respectively.

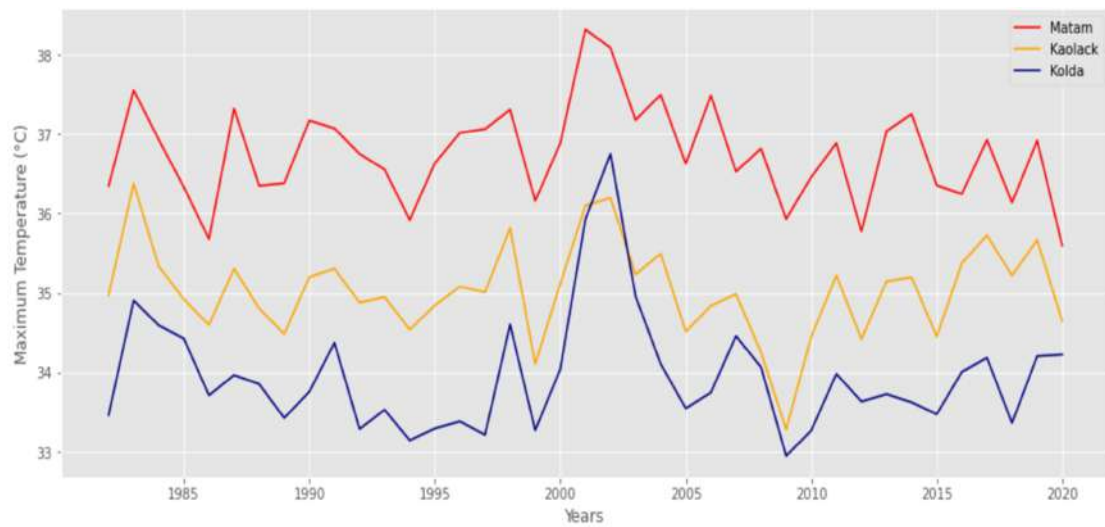


Figure 4. Annual average Maximum Temperature for Matam, Kaolack and Kolda (1982–2020).

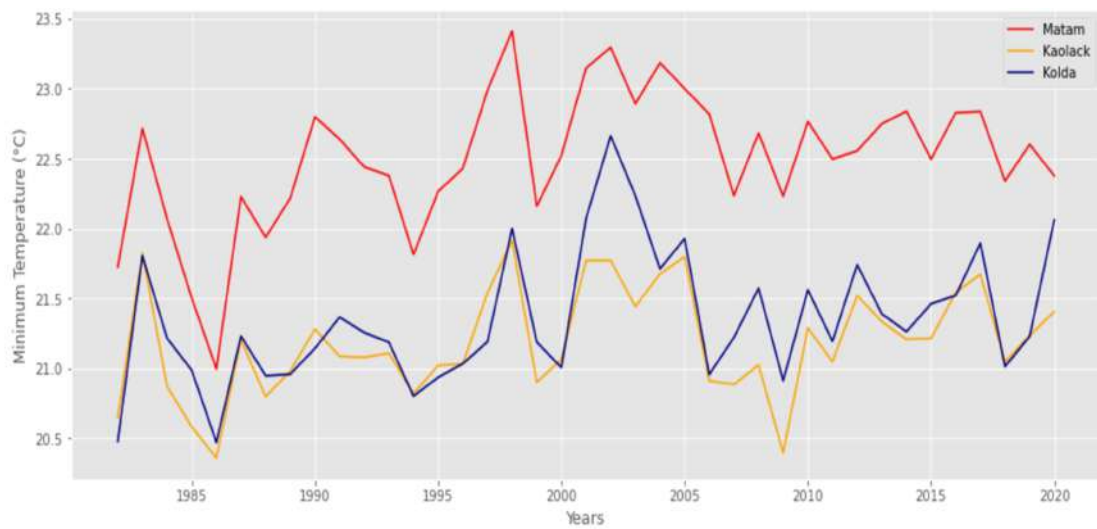


Figure 5. Annual average Minimum Temperature for Matam, Kaolack and Kolda (1982–2020).

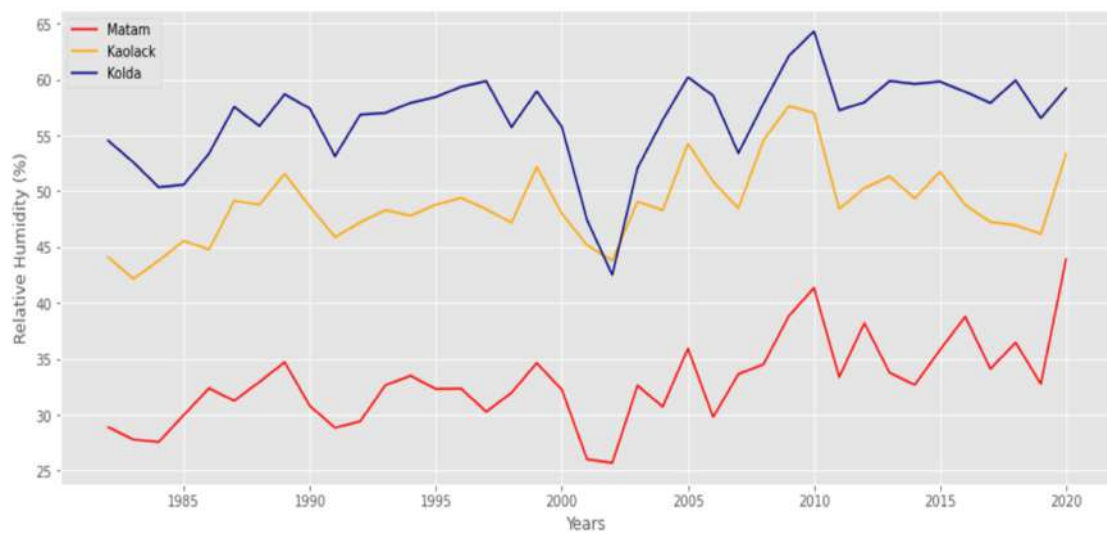


Figure 6. Annual average Relative Humidity for Matam, Kaolack and Kolda (1982–2020).

1.7.2. Annual Weather Cycle

There are two major seasons that describe the climate in Senegal. Roughly, the rainy season starts from June to October and the dry season is from October to May. As reported by [24] rainfall events in Senegal are expected to occur from June to October. Figure 7 shows that all the three geographical regions share the same annual rainfall cycle (June–October) with the month of August being the highest.

Temperatures start to rise in February with the highest levels expected from April to early June. In the north, heat spikes of over 40 °C are observed [25]. Overall, Relative Humidity starts to rise from June to early October. These months are considered as hot and humid periods, especially in the south, whereas in the north, moist and rainy days alternate with hot and dry days [26]. Figures 7–10 summarize monthly average Rainfall, Maximum Temperature, Minimum Temperature and Relative Humidity, respectively.

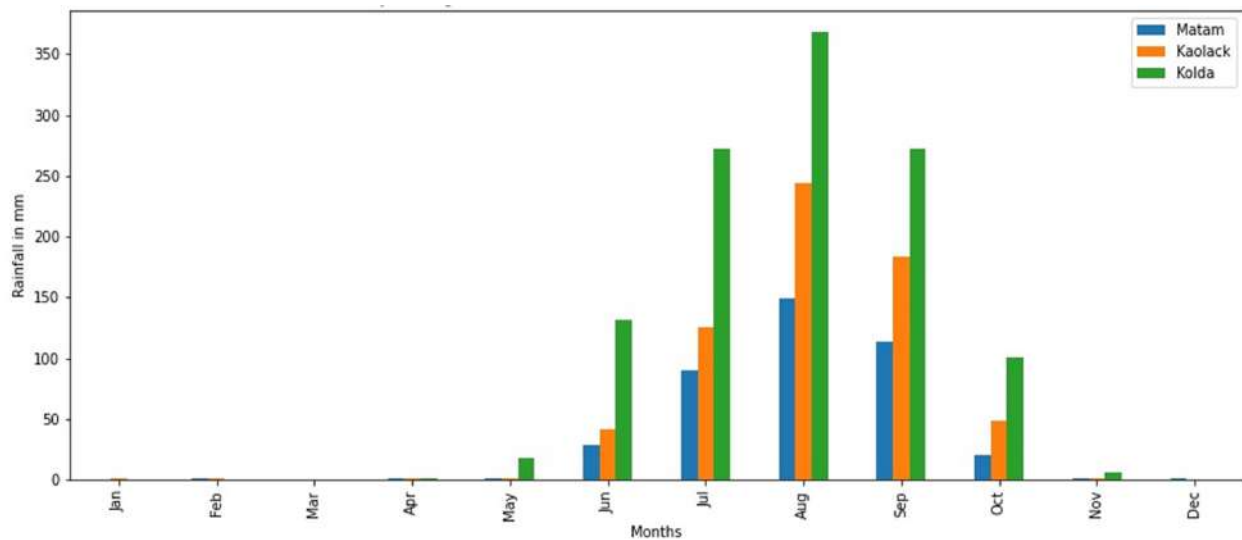


Figure 7. Monthly average rainfall variation for Matam, Kaolack and Kolda (1982–2020).

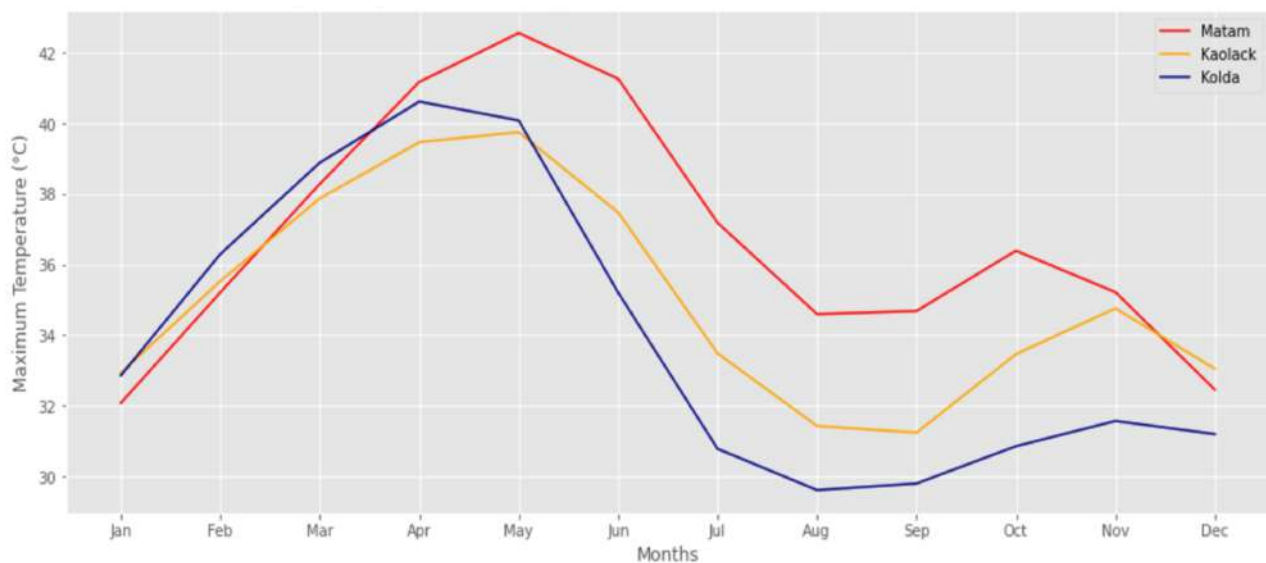


Figure 8. Monthly average Maximum Temperature for Matam, Kaolack and Kolda (1982–2020).

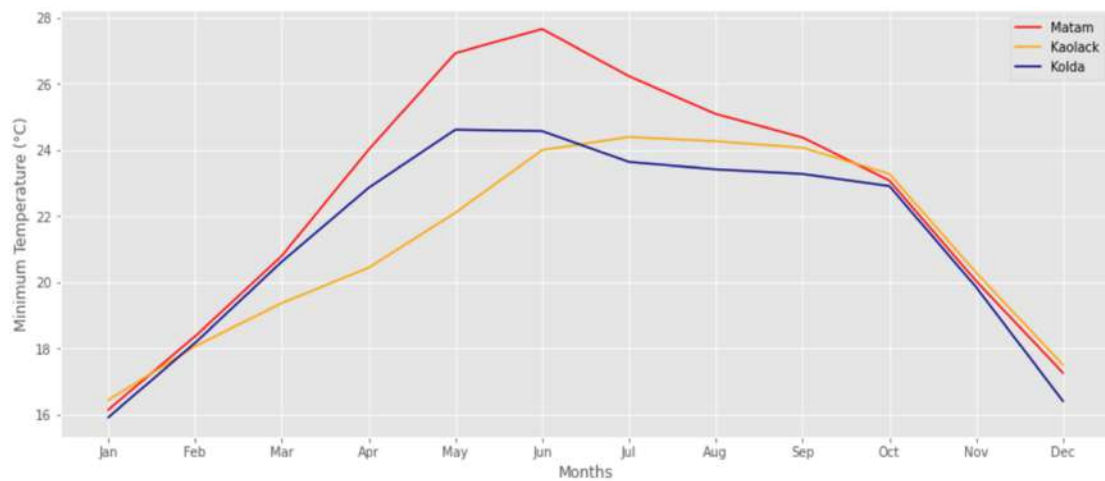


Figure 9. Monthly average Minimum Temperature for Matam, Kaolack and Kolda (1982–2020).

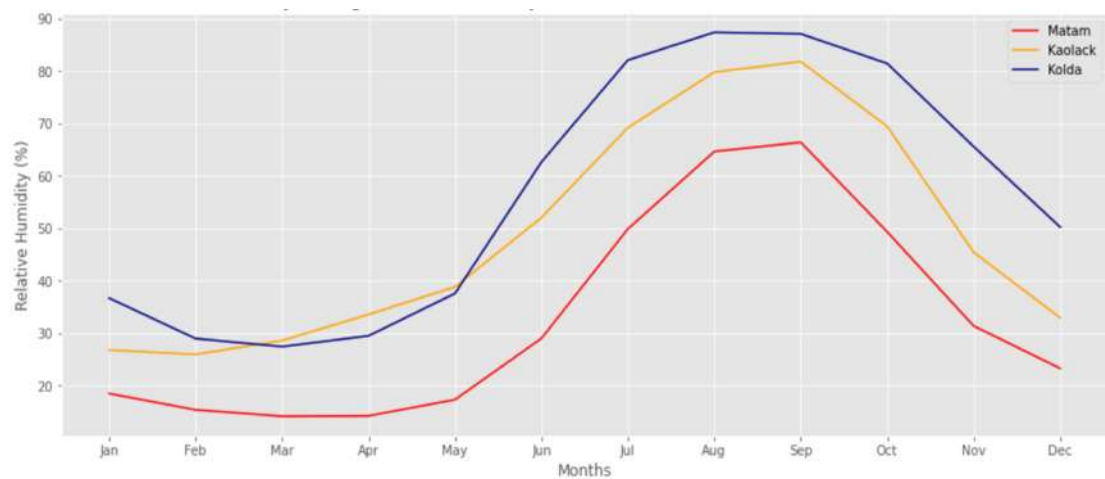


Figure 10. Monthly average Relative Humidity for Matam, Kaolack and Kolda (1982–2020).

2. Materials and Methods

The general approach behind this study is the Knowledge Discovery in Databases (KDD) process. The Knowledge Discovery in Databases process is defined as “an iterative multi-stage process for extracting meaningful patterns from data. The Knowledge Discovery in Databases process draws methods from diverse fields such as ML, AI, and Database Management. Steps in the KDD process are: Data Selection and Integration, Data Cleaning and Preprocessing, Data Transformation, Data Mining, Pattern Evaluation and Interpretation” [27]. Figure 11 presents a summary of the entire study in a form of a flowchart.

Daily Rainfall forecasting was performed using Relative Humidity, Maximum Temperature, Minimum Temperature and Day of the Year as predictors. Predictors were selected based on the correlation with the target variable. Correlation describes the relationship of one or more features to each other. Features can have a positive or negative relationship of between -1 and $+1$. We observed the correlation of features using a heatmap as presented in Figure 12. A heatmap helps to visualize the strength of correlation among features and identify features that are best for training a Machine Learning model [28]. It is observed that there is a moderate to weak correlation between Rainfall with other features. Relative Humidity, Minimum Temperature and Day of the Year have a positive correlation with Rainfall of 0.5, 0.28 and 0.16, respectively. Maximum Temperature has a negative correlation with Rainfall of -0.33 . Correlation scores are summarized in Figure 12.

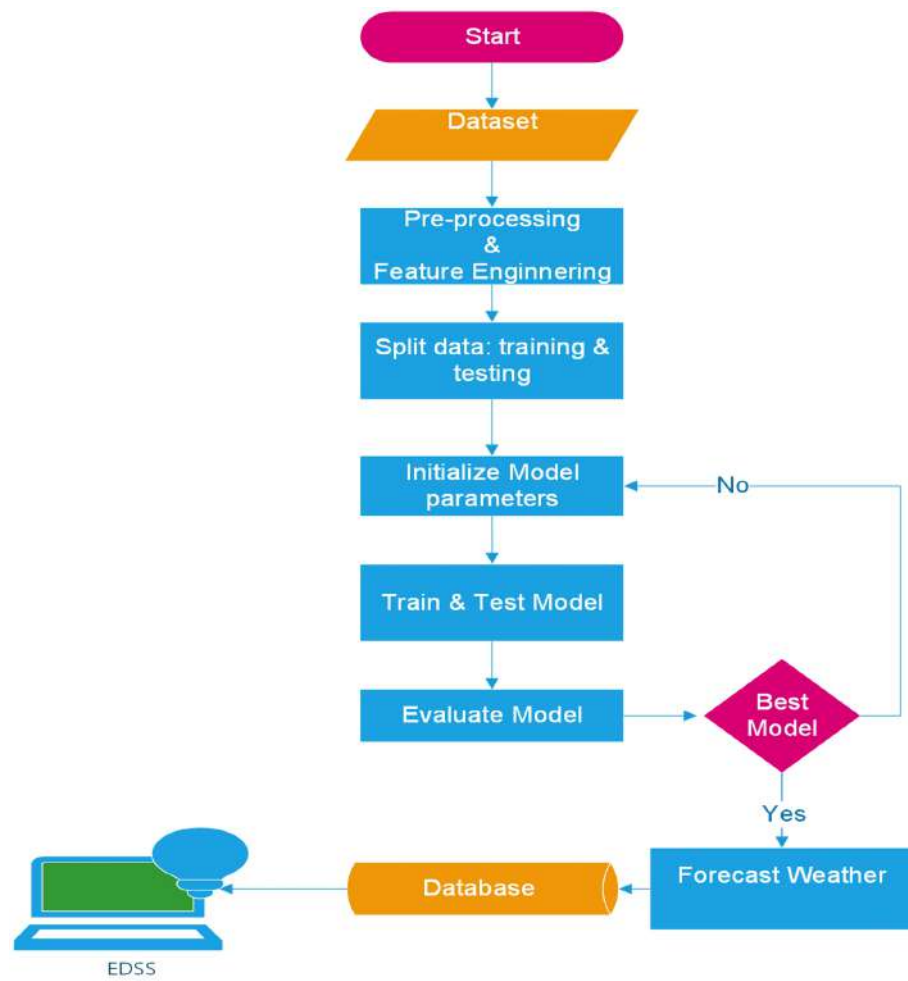


Figure 11. Flowchart summarizing the entire study.

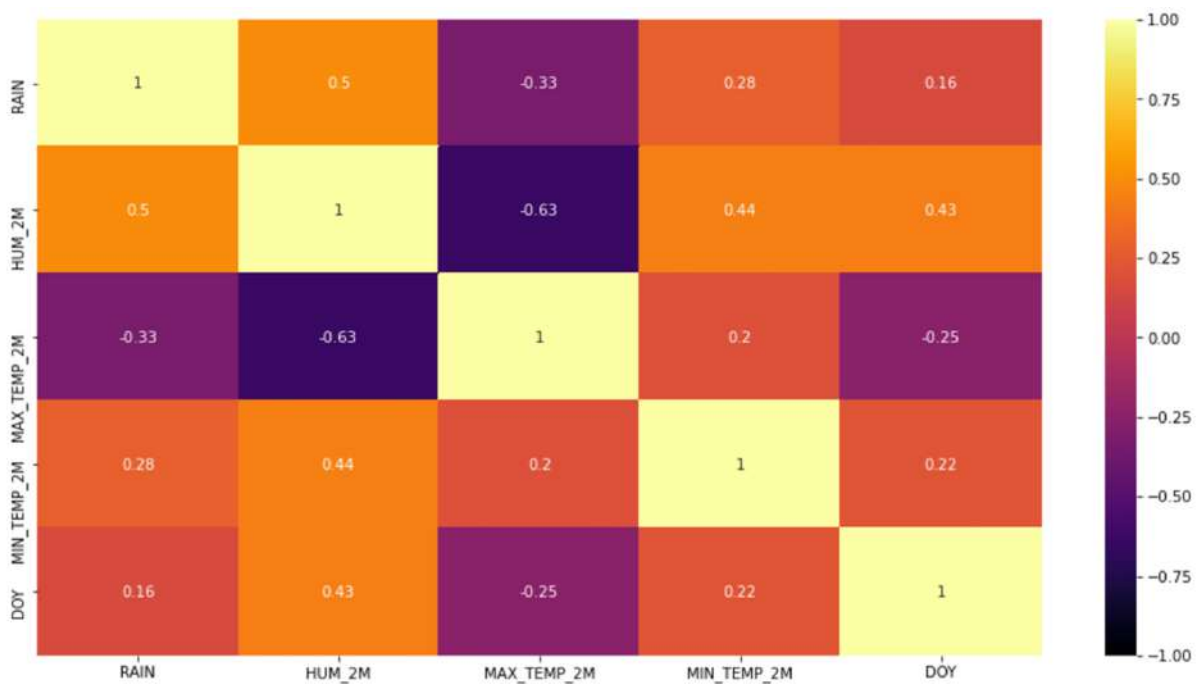


Figure 12. Correlation summary for Rainfall, Relative Humidity, Maximum Temperature, Minimum Temperature and Day of Year.

Daily forecasting for Relative Humidity, Maximum and Minimum Temperature was conducted by using five lagged values for each feature. This was the case because the autocorrelation for these parameters was strong. Autocorrelation measures “the degree of correlation of past values of a time series”. Autocorrelation can be positive or negative. For example, an autocorrelation of +1 signifies a strong positive correlation, whereas an autocorrelation of −1 signifies a strong negative correlation [29]. As seen in Figure 13, the autocorrelation score for Relative Humidity, Maximum Temperature and Minimum Temperature is above 0.8 up to five lags. Autocorrelation scores of features are summarized in Figure 13.

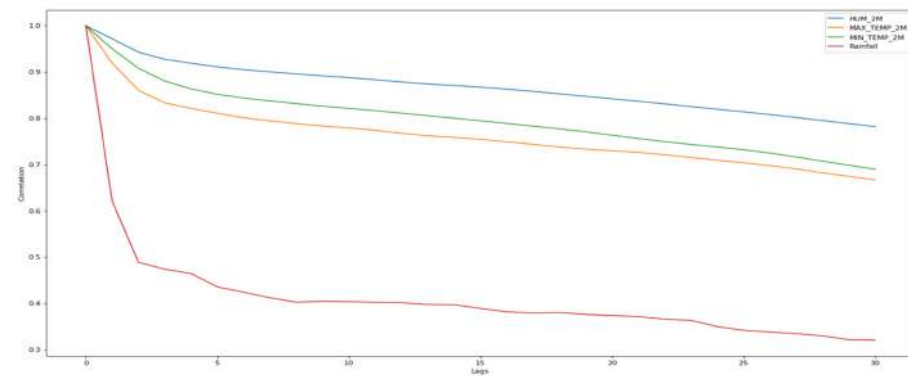


Figure 13. Autocorrelation summary for Relative Humidity, Maximum Temperature, Minimum Temperature and Rainfall.

2.1. Data Pre-Processing and Transformation

The objective of the data pre-processing phase was to detect and impute missing values and ensure that all parameters are converted into the right data types and format and normalized. Normalization ensures that features are within a given range.

For rainfall forecasting, we realized that the correlation of Rainfall with the predictors was not strong to ensure better prediction. Hence, to improve the correlation, additional features were derived through feature engineering: fourteen polynomial features from the four features (Relative Humidity, Maximum Temperature, Minimum Temperature and Day of the Year). Further, the Rainfall parameter was transformed by performing log transformation to have a less skewed distribution. Creating polynomial features and transforming the target variable increased the correlation of features and the target variable as presented in Figure 14.

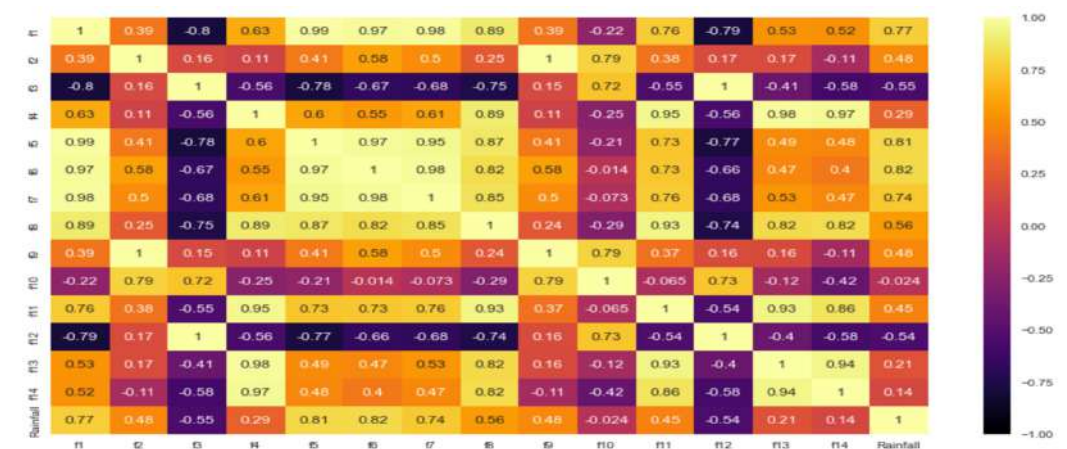


Figure 14. Correlation of features after feature engineering and data transformation.

For the forecasting of Relative Humidity, Maximum Temperature and Minimum Temperature, a dataset was restructured into input variables and output variables by using

lagged values as input variables and using the next time step as the output variable. This technique is called a sliding window with one-step forecasting. With this technique, the order of the time series observations is maintained and the number of lagged values is called the window size. In this study, observations for the past five days were used as inputs for the next day's forecast. Figure 15 summarizes how the Relative Humidity, Maximum Temperature and Minimum Temperature dataset was restructured using a sliding window with a one-step forecasting technique.

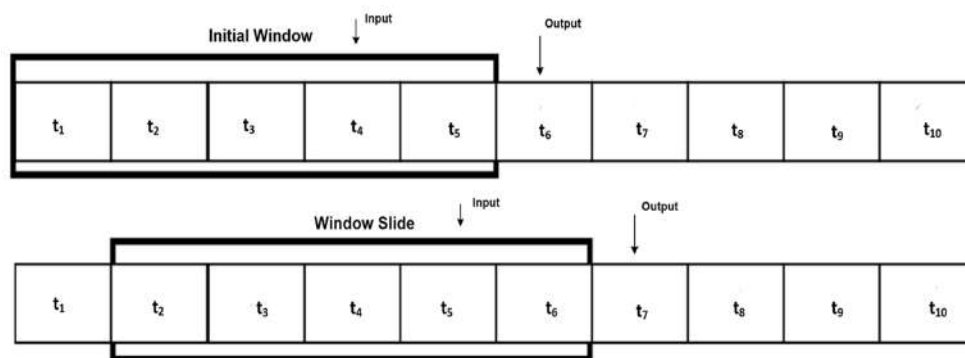


Figure 15. Summary of sliding window with one-step forecasting technique.

Machine Learning Models

In this study, we compared our Ensemble Model with ten ML regressors: Light Gradient Boosting Machine, CatBoost Regressor, Gradient Boosting Regressor, Extreme Gradient Boosting, Random Forest Regressor, Orthogonal Matching Pursuit, Extra Trees Regressor, K Neighbors Regressor, AdaBoost Regressor and Decision Tree Regressor. The Ensemble Model was developed by stacking the top three regressors: CatBoost Regressor, Gradient Boosting Regressor and Light Gradient Boosting Machine. Machine Learning regressors are trained to understand the relationship between features and the target variable or outcome. This is referred to as Supervised Machine Learning. Then, the models are leveraged to predict the outcome of new and unseen input data or to impute missing data. While Ensemble Models combine predictions from two or more base models. Ensemble Models are used to achieve the best performance on a predictive modeling project. The study was implemented using Pycaret [30] Python library.

The following metrics with 10-folds cross-validation were used to evaluate the models: Mean Absolute Error, Mean Squared Error (MSE), Root Mean Squared Error and Coefficient of Determination. Mean Absolute Error is the “average of the absolute difference between the actual and predicted values in the dataset” [31]. The Mean Absolute Error is expressed as:

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

where N is the number of samples, y_i is the i th observed value in the dataset and \hat{y} is the predicted value.

Mean Squared Error is the “average of the squared differences between predicted and expected target values in a dataset. The squaring has the effect of inflating large errors” [31]. Mean Squared Error is expressed as:

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

where N is the number of samples, y_i is the i th observed value in the dataset and \hat{y} is the predicted value.

Root Mean Squared Error is the “square root of the average of squared errors” [31]. Root Mean Squared Error is expressed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \tag{3}$$

where N is the number of samples, y_i is the i th observed value in the dataset and \hat{y} is the predicted value.

The coefficient of determination is defined as the “proportion of the variance in the dependent variable that is predictable from the independent variables” [32]. The coefficient of determination is expressed as:

$$R^2 = \frac{RSS}{TSS} \tag{4}$$

where RSS is the Residuals Sum of Squares and TSS is the Total Sum of Squares.

The dataset from 1982 to 2008 was used for training and from 2009 to 2020 was used for testing the models.

3. Presentation of Results and Discussion

The results presented are average scores for MAE, MSE, RMSE and R^2 across the 10-folds cross-validation. Then, we analyze the performance of the best model for each weather parameter.

3.1. Relative Humidity

From Table 2, the comparative analysis shows that the Ensemble Model is found to have a lower score of MAE, MSE and RMSE: 4.0126, 29.9885 and 5.4428, respectively. The Ensemble Model proved to achieve the highest accuracy in predicting the Relative Humidity with R^2 score of 0.9335. Table 2 summarizes performance of models.

Table 2. Model performance for Relative Humidity Forecasting.

Model	MAE	MSE	RMSE	R^2
Ensemble Model	4.0126	29.9885	5.4428	0.9335
Light Gradient Boosting Machine	4.0693	30.6936	5.5040	0.9317
CatBoost Regressor	4.0619	30.7052	5.5046	0.9317
Gradient Boosting Regressor	4.0863	30.8061	5.5140	0.9314
Extreme Gradient Boosting	4.1601	32.1831	5.6292	0.9280
Random Forest Regressor	4.2284	32.9041	5.7005	0.9270
Orthogonal Matching Pursuit	4.2385	33.1158	5.7223	0.9268
Extra Trees Regressor	4.2533	33.3111	5.7349	0.9260
K Neighbors Regressor	4.4810	36.5971	6.0138	0.9184
AdaBoost Regressor	5.7023	48.4746	6.9384	0.8954
Decision Tree Regressor	5.9400	65.2236	8.0390	0.8553

We analyze the performance of the Ensemble Model by analyzing the goodness of fit and residuals as presented in Figure 16. Residual is the difference between the actual value of y and the predicted value of y . A comparison of the model’s best fit prediction with a dashed 45-degree line in Figure 16A shows that the model performed well. We observe that there are many points close to the identity line meaning that the prediction error is close to zero. Figure 16B shows that the residuals are proportionally dispersed along the zero axis which would mean lower errors. The residuals do not follow any specific pattern. Nevertheless, we can also perceive some cases where higher Relative Humidity was observed but the model predicted lower Relative Humidity or where lower Relative Humidity was observed but the model predicted higher Relative Humidity. We further observe some outliers which might contribute to the high value of MAE. We believe

that adding more features can help to reduce these estimator errors. The histogram in Figure 16B demonstrates that model regularization was performed well considering the spectrum and residual distribution of the predictions on the training and testing set. This shows that the model generalizes well.

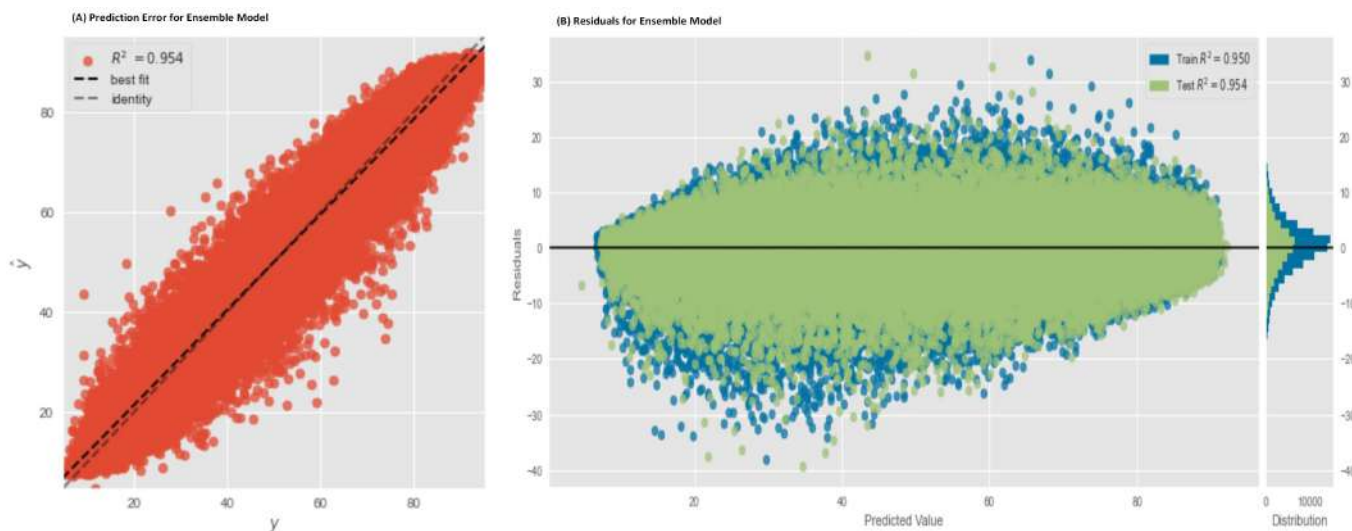


Figure 16. Accuracy of Ensemble Model for Relative Humidity forecasting.

3.2. Minimum Temperature

As presented in Table 3, it is observed that Ensemble Model performed well with a lower score of MAE, MSE and RMSE: 0.7908, 1.1329 and 1.0515, respectively. Additionally, Ensemble Model got the highest R^2 score of 0.9018. Table 3 summarizes performance of models.

We consider the goodness of fit and residuals as presented in Figure 17 to analyze the performance of the Ensemble Model. Generally, the model performed well as seen in Figure 17A, model's best fit prediction is closer to the dashed 45-degree line. This implies that the prediction error is close to zero. Figure 17B shows that the residuals are proportionally dispersed along the zero axis, the residuals do not follow any specific pattern. However, we observe a few cases where the model has predicted higher or lower than the observed Minimum Temperature. Further, we notice some cases of outliers which might help to explain the value of MAE. Adding more features can help to reduce these estimator errors. The histogram spectrum and residual distribution of the predictions on the training and testing set in Figure 17B demonstrate that the model generalizes well.

3.3. Maximum Temperature

Comparatively, as seen in Table 4 Ensemble Model performed well with a lower score of MAE, MSE and RMSE: 1.2515, 2.8038 and 1.6591, respectively. In terms of R^2 score, Ensemble Model got the highest with 0.8205. Table 4 summarizes performance of models.

We analyze the performance of the Ensemble Model by considering the goodness of fit and residuals as presented in Figure 18. A comparison of the model's best fit prediction with a dashed 45-degree line in Figure 18A shows that overall the model performed well. However, we observe a few cases where the model predicted higher or lower than the observed Maximum Temperature. Further, we notice some cases of outliers which might help to describe the value of MAE. Adding more features can help to reduce these estimator errors. Overall, residuals are proportionally dispersed along the zero axis as seen in Figure 18B. This means lower errors and residuals do not follow any specific pattern.

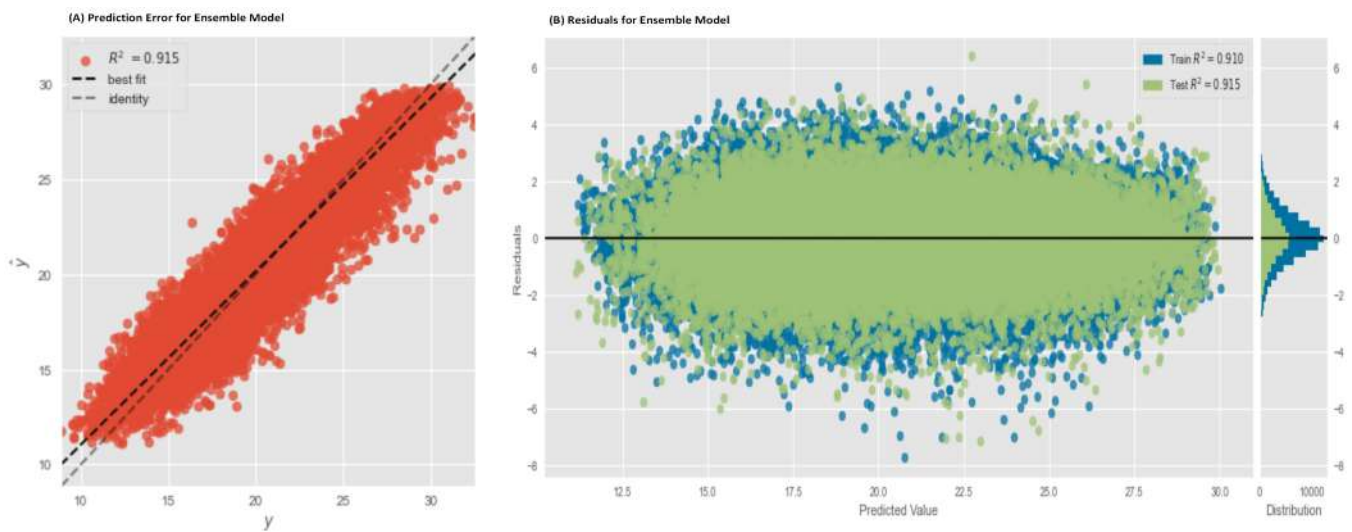


Figure 17. Accuracy of Ensemble Model for Minimum Temperature forecasting.

Table 3. Model performance for Minimum Temperature Forecasting.

Model	MAE	MSE	RMSE	R ²
Ensemble Model	0.7908	1.1329	1.0515	0.9018
Gradient Boosting Regressor	0.7953	1.1481	1.0582	0.9006
Light Gradient Boosting Machine	0.7966	1.1508	1.0595	0.9004
CatBoost Regressor	0.7983	1.1554	1.0614	0.9001
Extreme Gradient Boosting	0.8107	1.1893	1.0771	0.8971
Orthogonal Matching Pursuit	0.8199	1.2034	1.0840	0.8956
Random Forest Regressor	0.8248	1.2200	1.0922	0.8942
Extra Trees Regressor	0.8301	1.2326	1.0982	0.8931
K Neighbors Regressor	0.8793	1.3646	1.1573	0.8815
AdaBoost Regressor	0.8961	1.4137	1.1727	0.8787
Decision Tree Regressor	1.1877	2.4335	1.5515	0.7865

Table 4. Model performance for Maximum Temperature Forecasting.

Model	MAE	MSE	RMSE	R ²
Ensemble Model	1.2515	2.8038	1.6591	0.8205
Light Gradient Boosting Machine	1.2618	2.8418	1.6694	0.8176
Gradient Boosting Regressor	1.2678	2.8478	1.6725	0.8175
CatBoost Regressor	1.2624	2.8501	1.6716	0.8171
Extreme Gradient Boosting	1.2878	2.9636	1.7031	0.8095
Random Forest Regressor	1.3031	3.0114	1.7195	0.8071
Extra Trees Regressor	1.3116	3.0473	1.7298	0.8048
Orthogonal Matching Pursuit	1.3240	3.1079	1.7519	0.8016
K Neighbors Regressor	1.3811	3.3403	1.8128	0.7865
AdaBoost Regressor	1.4331	3.3870	1.8281	0.7841
Decision Tree Regressor	1.8775	6.1235	2.4593	0.6098

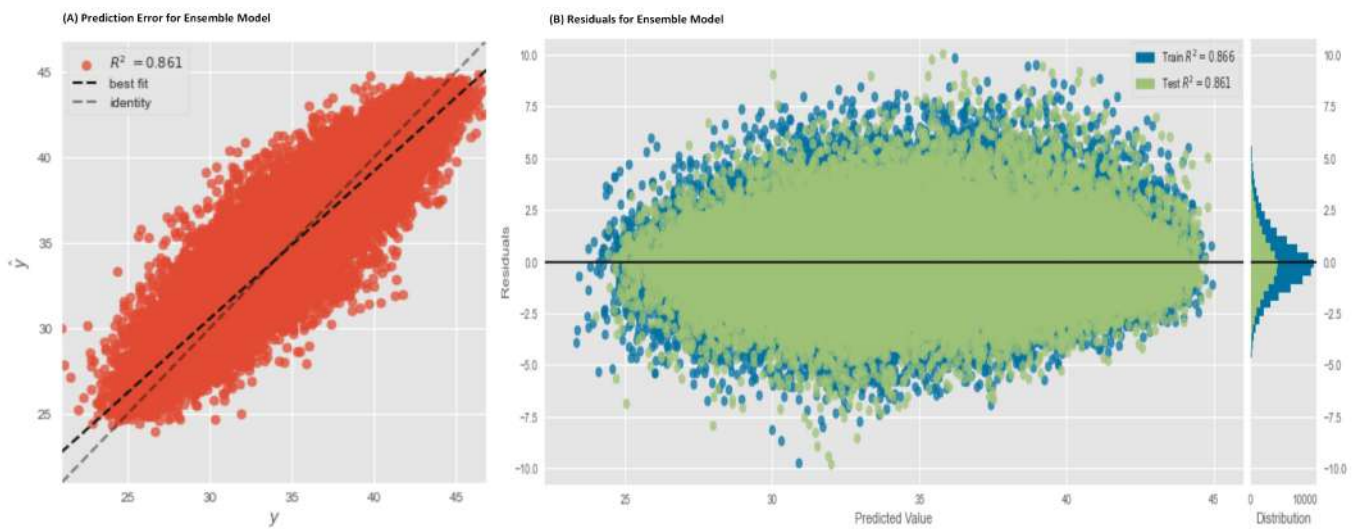


Figure 18. Accuracy of Ensemble Model for Maximum Temperature forecasting.

3.4. Rainfall

Relatively, from Table 5 it is observed that the Ensemble Model performed well with a lower score of MAE, MSE and RMSE: 0.2142, 0.1681 and 0.4100, respectively. For R^2 score, Ensemble Model got the highest with 0.7733. Table 5 summarizes performance of models.

We use the goodness of fit and residuals as presented in Figure 19 to analyze the performance of the Ensemble Model. In Figure 19A, we observe a high density of errors on point zero of the x-axis which means that the model predicted rains on a day when no rainfall was actually observed. In some cases where rainfall was actually observed, the model made over or under predictions. Further, we notice some cases of outliers which might help to explain the value of MAE. Residuals in Figure 19B show the model struggled to predict the extremely low observed rainfall. This requires adding more features, further data transformation and feature engineering to improve the correlation of features. Further, this can be attributed to the highly non-linearity of weather and rainfall.

Table 5. Model performance for Rainfall Forecasting.

Model	MAE	MSE	RMSE	R^2
Ensemble Model	0.2142	0.1681	0.4100	0.7733
CatBoost Regressor	0.2150	0.1691	0.4112	0.7719
Light Gradient Boosting Machine	0.2146	0.1695	0.4117	0.7714
Gradient Boosting Regressor	0.2221	0.1752	0.4185	0.7638
Extreme Gradient Boosting	0.2178	0.1752	0.4185	0.7638
Random Forest Regressor	0.2212	0.1797	0.4238	0.7578
Extra Trees Regressor	0.2246	0.1851	0.4302	0.7504
K Neighbors Regressor	0.2316	0.2022	0.4496	0.7272
AdaBoost Regressor	0.3803	0.2851	0.5325	0.6147
Orthogonal Matching Pursuit	0.4127	0.3336	0.5775	0.5502
Decision Tree Regressor	0.2910	0.3452	0.5875	0.5343

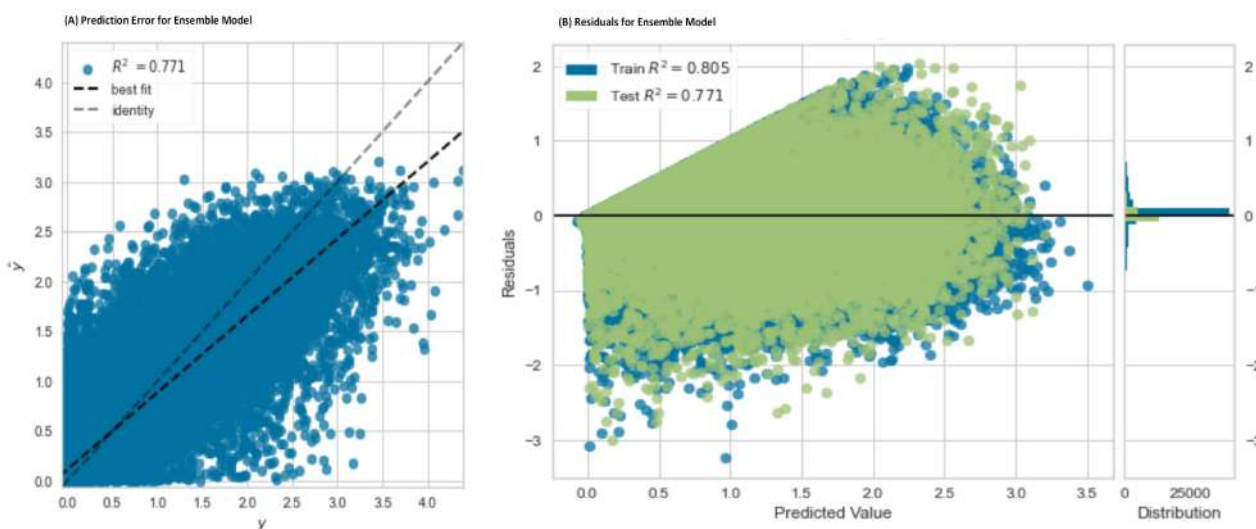


Figure 19. Accuracy of Ensemble Model for Rainfall forecasting.

Comparatively, other studies have reported similar results. Table 6 summarizes our results and results from other studies.

Table 6. Summary of our results and results from other studies.

Authors	Model	Parameter	MAE	RMSE
Our results	Ensemble Model	Relative Humidity	0.1873	0.1369
		Minimum Temperature	0.1881	0.1429
		Maximum Temperature	0.1898	0.144
		Rainfall	0.2987	0.1787
[20]	Random Forest		4.49	8.82
	Multivariate Linear Regression	Rainfall	4.97	8.61
	XGBoost		3.58	7.85
[33]	XGBoost	Rainfall	8.8	2.7
[34]	Linear Regression	Maximum Temperature	3.10	1.78
[35]	Random Forest	Indoor Air Temperature	0.3535	0.476
		Indoor Relative Humidity	1.47	2.429

4. Prediction on the Test Dataset

As presented in Section 1.3, the objectives were to observe if the model would correctly make predictions according to spatial weather distribution and annual cycle.

4.1. Spatial Weather Distribution

In order to observe spatial distribution, forecasting was performed to compare three geographical regions: south, center and north. One location from each region was selected as presented in Figure 2. Kolda from the south, Kaolack from the center and Matam from the north.

4.1.1. Relative Humidity

As seen in Figure 20A, the observed Relative Humidity increases from the north to the south. The model forecast has followed the same trend as seen in Figure 20B. The observed and model forecasts are high for Kolda almost throughout the year with a maximum of above 90% from 2009 to 2020. As seen in Figure 20A,B, comparatively, the model forecast is closer to the observed values. Figure 20 summarizes the daily observed and forecasted Relative Humidity.

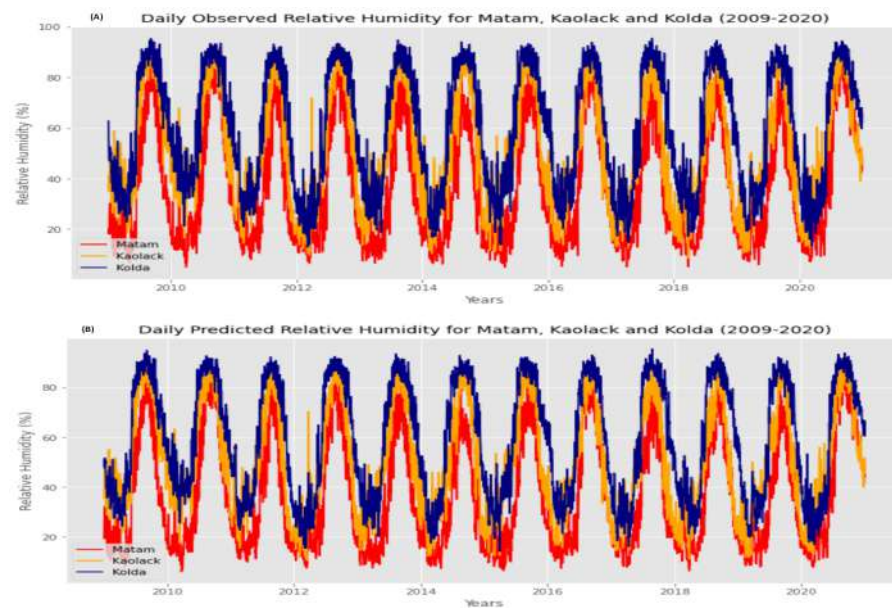


Figure 20. Observed and Forecasted Relative Humidity for Matam, Kaolack and Kolda (2009–2020).

4.1.2. Maximum and Minimum Temperature

In Senegal, temperatures vary from hot to extremely hot from the south to the north due to the warm desert climate in the north. As presented in Figure 21A, the observed Maximum Temperature from 2009 to 2020 is expected to be high in Matam with a huge spike of above 45 °C. As presented in Figure 21B, the model has forecasted the same trend with the Maximum Temperature in Matam of about 44 °C from 2009 to 2020. Figures 21 and 22 summarize the daily observed and forecasted Maximum Temperature and Minimum Temperature, respectively.

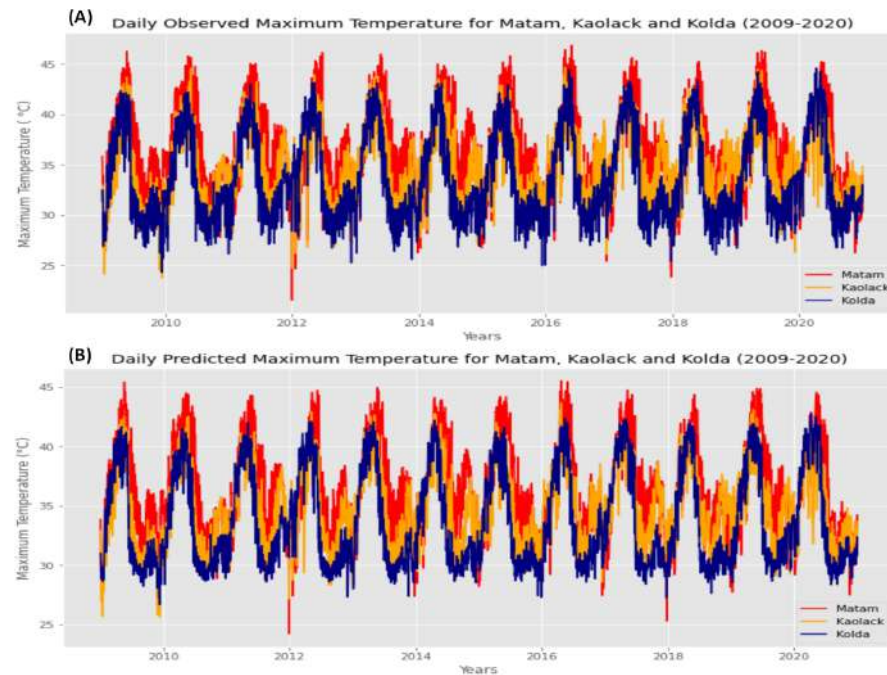


Figure 21. Observed and Forecasted Maximum Temperature for Matam, Kaolack and Kolda (2009–2020).

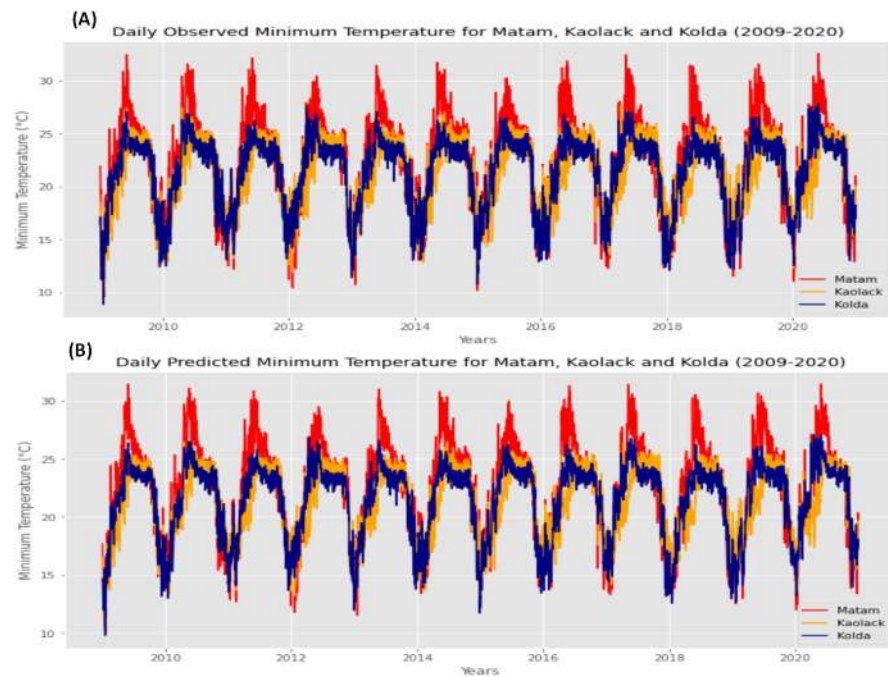


Figure 22. Observed and Forecasted Minimum Temperature for Matam, Kaolack and Kolda (2009–2020).

4.1.3. Rainfall

In Senegal, rainfall increases from north to south. As presented in Figure 23A, the observed annual cumulative rainfall from 2009 to 2020 in Kolda is mostly expected to be above 800 mm, between 600–800 mm in Kaolack and below 500 mm in Matam. The model forecast as presented in Figure 23B shows the same trend from 2009 to 2020 with Kolda above 800 mm, Kaolack between 600 mm to 800 mm, and Matam mostly below 400 mm. Figure 23 summarizes the annual observed and forecasted Rainfall.

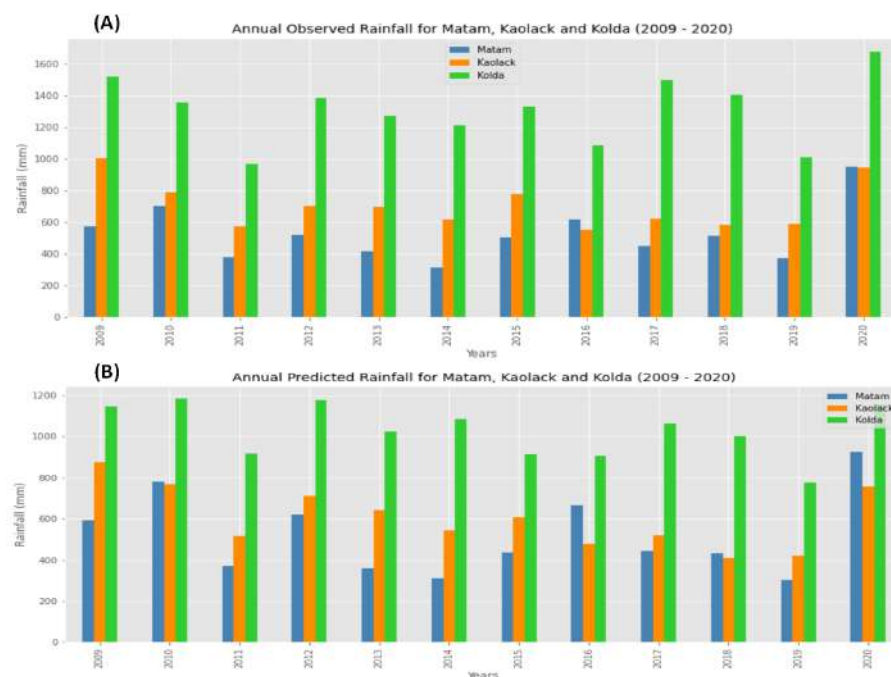


Figure 23. Observed and Forecasted Rainfall for Matam, Kaolack and Kolda (2009–2020).

4.2. Annual Weather Cycle

In order to observe the annual weather cycle, monthly average forecasting was performed to compare three geographical regions: South, Center and North. One location from each region was selected as presented in Figure 2. Kolda from the south, Kaolack from the center and Matam from the north.

4.2.1. Relative Humidity

In all the geographical regions, Relative Humidity is expected to start to rise from June to early October. These months especially in the South are considered hot and humid. Figure 24 shows the monthly average observed and model forecast Relative Humidity. Generally, as seen in Figure 24B the model was able to reasonably forecast according to the annual cycle with the peak expected around August in Kolda and September in Kaolack and Matam. Entirely, the model forecast is closer to the observed values as seen in Figure 24A. Figure 24 summarizes the observed and forecasted monthly average Relative Humidity.

4.2.2. Maximum Temperature and Minimum Temperature

The highest levels of temperatures are expected from April to early June in all the geographical regions with the noticeable heat spikes of over 40 °C in the North as seen in Figure 25A. Generally, the model was able to reasonably forecast according to the annual cycle with the peak expected around May in all the regions as seen in Figure 25B. Relatively, the model forecast is closer to the observed values. Figures 25 and 26 summarize the monthly average observed and model forecast Maximum temperature and Minimum Temperature.

4.2.3. Rainfall

As expected, the rainy season start from June to October and the dry season from roughly October to May as seen in Figure 27A. Substantially, the model predicted rainfall events in Kolda, Kaolack and Matam from the month of June to October as presented in Figure 27B. These results agree with the fact that all three geographical regions share the same annual rainfall cycle (June–October). Figure 27 summarizes the monthly average observed and model forecast Rainfall.

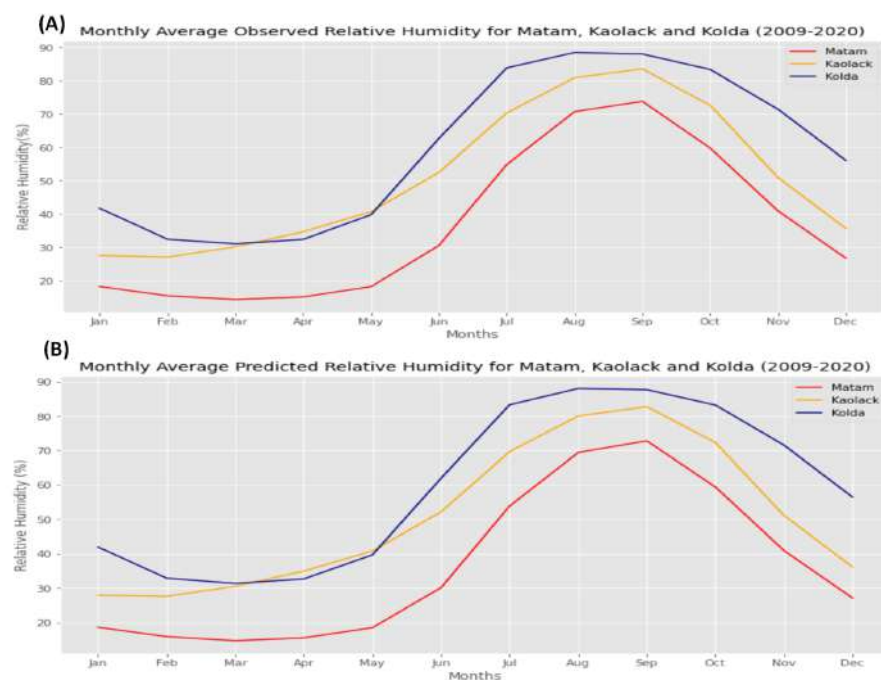


Figure 24. Monthly Average Observed and Forecasted Relative Humidity for Matam, Kaolack and Kolda (2009–2020).

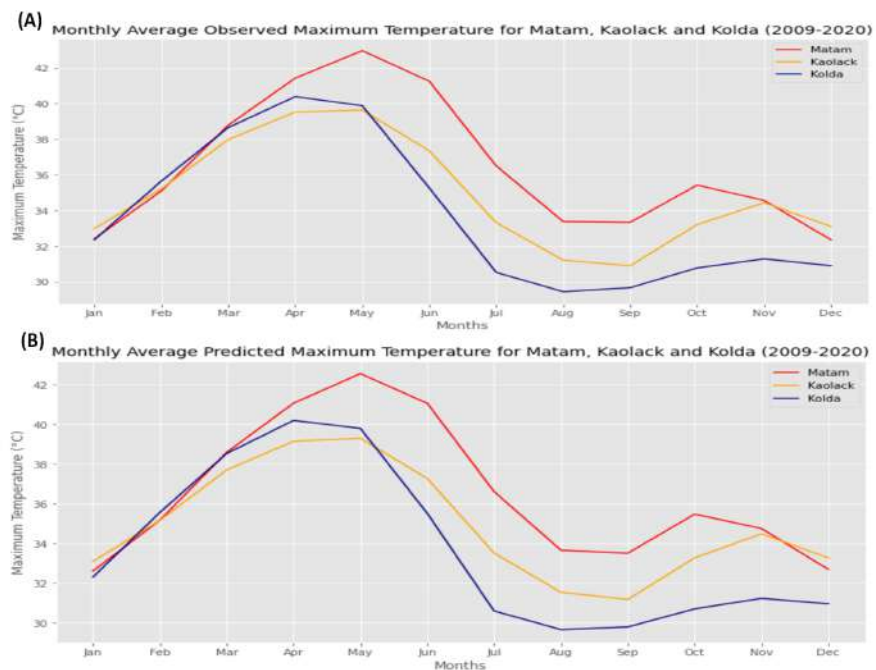


Figure 25. Monthly Average Observed and Forecasted Maximum Temperature for Matam, Kaolack and Kolda (2009–2020).

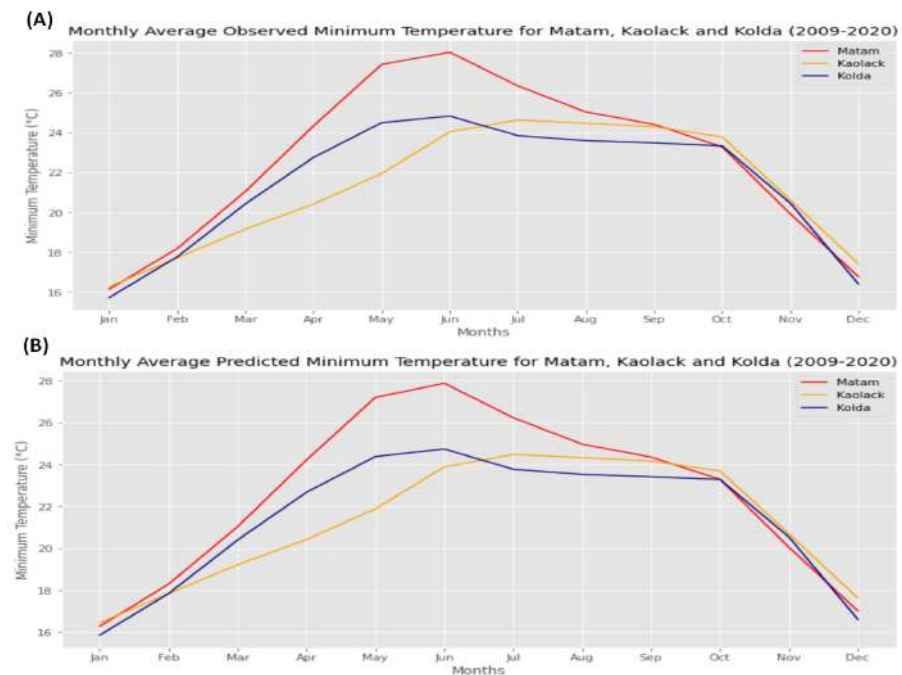


Figure 26. Monthly Average Observed and Forecasted Minimum Temperature for Matam, Kaolack and Kolda (2009–2020).

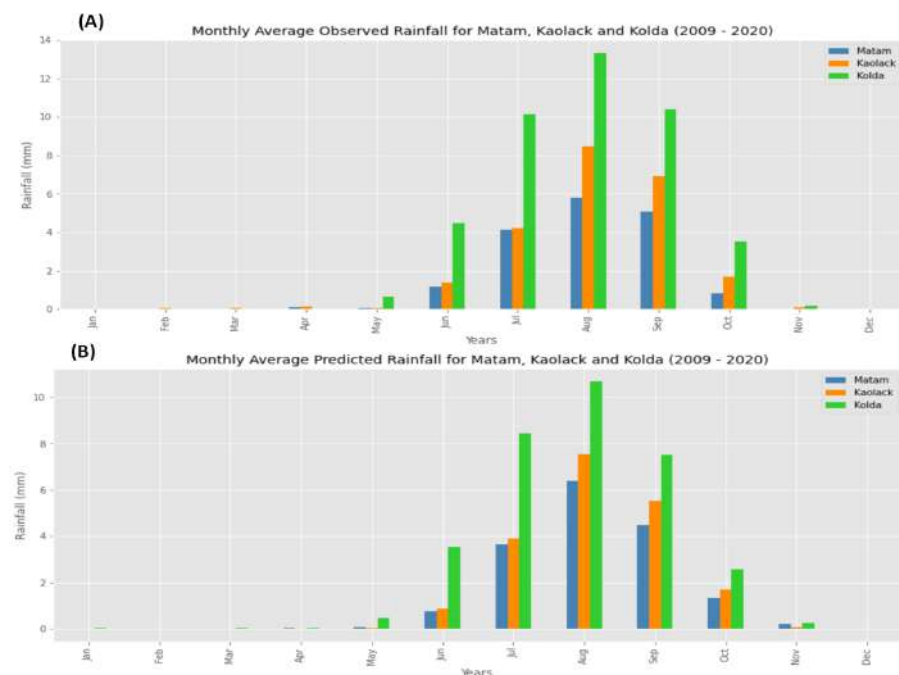


Figure 27. Monthly Average Observed and Forecasted Rainfall for Matam, Kaolack and Kolda (2009–2020).

4.3. Conclusions and Recommendations

In this study, we compared ten ML Regressors and our Ensemble Model for daily Rainfall, Relative Humidity, and Maximum and Minimum Temperature forecasting in Senegal. The results indicate that the Ensemble Model outperformed the Regressors. The notable results for Relative Humidity were MAE 4.0126, MSE 29.9885, RMSE 5.4428 and R^2 0.9335. Minimum Temperature were MAE 0.7908, MSE 1.1329, RMSE 1.0515 and R^2 0.9018. Maximum Temperature were MAE 1.2515, MSE 2.8038, RMSE 1.6591 and R^2 0.8205. Rainfall were MAE 0.2142, MSE 0.1681, RMSE 0.4100 and R^2 0.7733. This study has shown that Machine Learning models can be adapted to daily weather forecasting. Further, the study has shown that Machine Learning models can learn and forecast weather in accordance with spatial distribution and annual cycle.

Given these findings, the potential role of technologies in ensuring that farmers adapt and develop resilience to negative climate change effects in Africa will not occur without concerted and decisive policy action to adopt such technologies. To enhance food security, the challenges facing the agricultural sector due to climate change effects in Africa need to be quickly addressed.

4.4. Limitation of the Study

Weather and climate processes are closely connected with greenhouse gasses. This study did not take into account the effect of greenhouse gasses in weather forecasting. An increasing number of features and more feature engineering could have possibly helped to improve the performance of the models.

4.5. Future Work

Our future work will be to perform long-term weather forecasting and develop an Expert Decision Support System that will utilize the weather forecasts to provide a platform where farmers and researchers will have access to localized weather information, expert recommendations and insights for farmer decision support and action.

Author Contributions: Conceptualization, C.N.; methodology, C.N.; validation, A.D. (Awa Diattara), A.T. and A.D. (Abdoulaye Deme); formal analysis, C.N.; writing—original draft preparation, C.N.; writing—review and editing, A.D. (Awa Diattara), A.T. and A.D. (Abdoulaye Deme); visualization, C.N.; project administration, C.B.; funding acquisition, C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Partnership for skills in Applied Sciences, Engineering and Technology (PASET)—Regional Scholarship and Innovation Fund (RSIF).

Data Availability Statement: Data supporting reported results can be accessed from <https://power.larc.nasa.gov/data-access-viewer/> (accessed on 26 July 2021).

Acknowledgments: This work is part of the ongoing PhD training supported by the Partnership for skills in Applied Sciences, Engineering and Technology—Regional Scholarship and Innovation Fund. We thank Assistant Professor Diego Hernán Peluffo-Ordóñez (Modeling, Simulation and Data Analysis (MSDA) Research Program, Mohammed VI Polytechnic University) for providing technical guidance.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rodolfo, M.; Drilona, E. *Climate Change in Sub-Saharan Africa's Fragile States*; International Monetary Fund: Washington, DC, USA, 2022.
- Kelvin, M.; Ng'ombe, J.N. Climate change impacts on sustainable maize production in Sub-Saharan Africa: A review. *Maize Prod. Use* **2019**, 47–75.
- Nyaga, J.N. *Assessment of Perceived Impacts of Climate Change on Agricultural Crops Productions and Its Effects on Food Security: A Case Study of Small-Scale Farmers in Murang'a County Kenya*; Università Ca'Foscari Venezia: Venice, Italy, 2021; Volume 123.
- Kahn, M.E.; Mohaddes, K.; Ng, R.N.C.; Pesaran, M.H.; Raissi, M.; Yang, J.-C. Long-term macroeconomic effects of climate change: A cross-country analysis. *Energy Econ.* **2021**, *104*, 105624. [[CrossRef](#)]
- Koudahe, K.; Djaman, K.; Bodian, A.; Irmak, S.; Sall, M.; Diop, L.; Balde, A.B.; Rudnick, D.R. Trend analysis in rainfall, reference evapotranspiration and aridity index in Southern Senegal: Adaptation to the vulnerability of rainfed rice cultivation to climate change. *Atmos. Clim. Sci.* **2017**, *7*, 476–495. [[CrossRef](#)]
- Shen, X.; Liu, B.; Henderson, M.; Wang, L.; Jiang, M.; Lu, X. Vegetation greening, extended growing seasons, and temperature feedbacks in warming temperate grasslands of China. *J. Clim.* **2022**, *35*, 1–51. [[CrossRef](#)]
- Fonta, W.; Edame, G.; Anam, B.E.; Duru, E.J.C. Climate Change, Food Security and Agricultural Productivity in Africa: Issues and policy directions. *Int. J. Humanit. Soc. Sci.* **2011**, *1*, 205–223.
- Harris, D.; Orr, A. Is rainfed agriculture really a pathway from poverty? *Agric. Syst.* **2014**, *123*, 84–96. [[CrossRef](#)]
- Nyasulu, C.; Diattara, A.; Traore, A.; Ba, C. Enhancing Farmers Productivity Through IoT and Machine Learning: A State-of-the-Art Review of Recent Trends in Africa. In Proceedings of the International Conference on Research in Computer Science and its Applications, Dakar, Senegal, 17–19 June 2021; pp. 113–124.
- Islam, B.U. Comparison of conventional and modern load forecasting techniques based on artificial intelligence and expert systems. *Int. J. Comput. Sci. Issues* **2011**, *8*, 507–513.
- Iseh, A.J.; Woma, T.Y. Weather forecasting models, methods and applications. *Int. J. Eng. Res. Technol.* **2013**, *2*, 11945–11956.
- Dueben, P.D.; Bauer, P.; Adams, S. Deep learning to improve weather predictions. In *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2021; Volume 27, pp. 204–217.
- Bonavita, M.; Arcucci, R.; Carrassi, A.; Dueben, P.; Geer, A.J.; Le, S.B.; Longépé, N.; Mathieu, P.-P.; Raynaud, L. Machine learning for earth system observation and prediction. *Bull. Am. Meteorol. Soc.* **2021**, *102*, E710–E716. [[CrossRef](#)]
- Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2013.
- The Role of Weather Forecasting in Agriculture. Available online: <https://www.dtn.com/the-role-of-weather-forecasting-in-agriculture/> (accessed on 8 April 2022).
- Khan, N.A.; Qiao, J.; Abid, M.; Gao, Q. Understanding farm-level cognition of and autonomous adaptation to climate variability and associated factors: Evidence from the rice-growing zone of Pakistan. *Land Use Policy* **2021**, *105*, 105427. [[CrossRef](#)]
- Han, E.; Ines, A.V.; Baethgen, W.E. Climate-Agriculture-Modeling and Decision Tool (CAMDT): A software framework for climate risk management in agriculture. *Environ. Model. Softw.* **2017**, *95*, 102–114. [[CrossRef](#)]
- Tharun, V.P.; Prakash, R.; Devi, S.R. Prediction of Rainfall Using Data Mining Techniques. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 1507–1512.
- Alawadi, S.; Mera, D.; Fernández-Delgado, M.; Alkhabbas, F.; Olsson, C.M.; Davidsson, P. A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Syst.* **2020**, *13*, 689–705. [[CrossRef](#)]

20. Liyew, C.M.; Melese, H.A. Machine learning techniques to predict daily rainfall amount. *Ej. Big Data* **2021**, *8*, 1–11. [[CrossRef](#)]
21. Tall, M.; Sylla, M.B.; Diallo, I.; Pal, J.S.; Faye, A.; Mbaye, M.L.; Gaye, A.T. Projected impact of climate change in the hydroclimatology of Senegal with a focus over the Lake of Guiers for the twenty-first century. *Theor. Appl. Climatol.* **2017**, *129*, 655–665. [[CrossRef](#)]
22. Tall, A. Climate forecasting to serve communities in West Africa. *Procedia Environ. Sci.* **2010**, *1*, 421–431. [[CrossRef](#)]
23. Salack, S.; Muller, B.; Gaye, A.T. Rain-based factors of high agricultural impacts over Senegal. Part I: Integration of local to sub-regional trends and variability. *Theor. Appl. Climatol.* **2011**, *106*, 1–22. [[CrossRef](#)]
24. Fowler, A.M.; Boswijk, G.; Gergis, J.; Lorrey, A. ENSO history recorded in *Agathis australis* (kauri) tree rings. Part A: Kauri's potential as an ENSO proxy. *Int. J. Climatol. J. R. Meteorol. Soc.* **2008**, *28*, 1–20. [[CrossRef](#)]
25. Sultan, B.; Janicot, S. The West African monsoon dynamics. Part II: The “preonset” and “onset” of the summer monsoon. *J. Clim.* **2003**, *16*, 3407–3427. [[CrossRef](#)]
26. Ndiaye, B.; Moussa, M.A.; Wade, M.; Sy, A.; Diop, A.B.; Diop, A.D.; Diop, B.; Diakhaby, A. Spatial and Temporal Distribution of Rainfall Breaks in Senegal. *Am. J. Clim. Chang.* **2021**, *10*, 533–560. [[CrossRef](#)]
27. Nwagu, C.K.; Omankwu, O.C.; Inyiyama, H. Knowledge Discovery in Databases (KDD): An overview. *Int. J. Comput. Sci. Inf. Secur.* **2017**, *15*, 13–16
28. Kumar, S.; Chong, I. Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2907. [[CrossRef](#)] [[PubMed](#)]
29. Flores, J.H.; Engel, P.M.; Pinto, R.C. Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting. *Int. Jt. Conf. Neural Netw.* **2012**, 1–8.
30. Ali, M. An Open Source, Low-Code Machine Learning Library in Python. PyCaret Version 1.0.0. Available online: <https://www.pycaret.org> (accessed on 20 April 2022).
31. Botchkarev, A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv* **2018**, arXiv:1809.03006.
32. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [[CrossRef](#)]
33. Anwar, M.T.; Winarno, E.; Hadikurniawati, W.; Novita, M. Rainfall prediction using Extreme Gradient Boosting. *J. Phys. Conf. Ser.* **2021**, *1869*, 012078. [[CrossRef](#)]
34. Karna, N.; Roy, P.C.; Shakya, S. Temperature Prediction using Regression Model. *Adv. Eng. ICT Converg. Proc.* **2018**, *4*, 161–170.
35. Arulmozhi, E.; Basak, J.K.; Sihalath, T.; Park, J.; Kim, H.T.; Moon, B.E. Machine learning-based microclimate model for indoor air temperature and relative humidity prediction in a swine building. *Animals* **2021**, *11*, 222. [[CrossRef](#)]