FPT Education

FPT UNIVERSITY

DAP391m Project

Temperature Prediction For Agriculture Using Machine Learning

Member: Do Hoang Long, Pham Van Viet, Phung Gia Bao

Advisor: Phan Duy Hung

Abstract

Temperature prediction is one of the challenging tasks in the weather forecasting process. Accurate temperature prediction is now more difficult than before due to extreme climate variations. Machine learning techniques can predict temperature by extracting hidden patterns from historical weather data. Selection of an appropriate technique for prediction is a difficult job. This research proposes three machine learning models for temperature prediction for cities. Since then, that has empowered farmers to make informed decisions regarding planting timing, crop selection, and resource allocation, ultimately optimizing agricultural activities. The proposed framework uses three widely used machine learning techniques, i.e., Linear regression, Support vector regression, and XGBoost. For prediction, 14 years of historical weather data (2010 to 2024) for Thach That district of Hanoi city is considered. Pre-processing tasks such as cleaning and normalization were performed on the dataset before the prediction process. The results reflect that the proposed machine learning models outperform other models.

I. Introduction

The intricate relationship between agricultural systems and the variability of weather conditions presents significant challenges in maximizing agricultural productivity. To tackle this challenge, the need for accurate temperature forecasting tailored to agricultural requirements has become increasingly urgent.

Agriculture, reliant on climate variability, is profoundly impacted by temperature fluctuations, influencing crop growth, water management, and overall farm planning [1]. The urgency to address this challenge lies in developing temperature prediction models that empower farmers to make informed decisions regarding planting timing, crop selection, and resource allocation, ultimately optimizing agricultural activities [2].

By harnessing historical weather data, encompassing temperature fluctuations, humidity, precipitation, and other pertinent factors, machine learning algorithms can be trained to forecast temperature patterns with heightened accuracy [3][4][5]. Recent climate trends, marked by extreme and unusual weather events, underscore the critical importance of accurate temperature forecasts in adapting agricultural strategies to evolving environmental conditions.

In related research fields, various studies have explored the intersection between temperature forecasting and agriculture. Notable contributions include investigations into the impact of temperature changes on crop productivity [6], advancements in machine learning techniques for temperature prediction [7][8], and the development of decision support systems suitable for temperature-sensitive farming methods [9]. These studies provide valuable insights into the methods and approaches applied in this project.

In this research, temperature is predicted using machine learning techniques. Three machine learning algorithms, namely Linear Regression, SVR, and XGboost take environmental variables moderately and strongly associated with temperature as input variables. Better machine learning algorithms have been identified and reported based on performance metrics using RMSE, MSE, and R-squared score

In summary, the focus on temperature prediction for agriculture in this project is motivated by its fundamental role in addressing the current challenges facing the agricultural sector. The significance of accurate temperature forecasts in optimizing resource management, minimizing risks, and promoting sustainable agricultural practices underscores the importance and urgency of this endeavor.

II. Related Work

The related studies delve into various aspects of utilizing machine learning techniques in weather forecasting and related fields, including selecting crops based on weather conditions to optimize crop selection strategies for maximizing yields and mitigating risks [10][11]. Moreover, there is a focus on understanding the impacts of weather conditions on metro ridership in urban areas through empirical studies conducted in major cities in China [12]. Additionally, researchers have conducted comparative studies on data mining models for weather forecasting, aiming to identify the strengths and weaknesses of different predictive algorithms [13].

Furthermore, significant attention has been given to the application of machine learning in smart farming practices, particularly in arid regions. This research emphasizes the importance of accurate weather predictions for optimizing agricultural operations in challenging environments [14]. Moreover, studies have been conducted to predict pest incidence in organic banana crops using machine learning techniques, highlighting the significance of weather prediction in pest management and crop protection [15].

Additionally, research focuses on predicting wind power using backpropagation algorithms integrated with numerical weather prediction techniques [16]. Ensemble approaches have also been explored for weather prediction, leveraging multiple models to enhance forecasting accuracy and reliability [17]. Furthermore, the utilization of Long Short Term Memory (LSTM) networks for temperature forecasting has been investigated, given their ability to capture temporal dependencies in weather data. Dynamic models have been developed specifically to predict temperature in controlled environments like greenhouses, contributing to optimizing environmental conditions for plant growth and productivity [18][19].

Finally, researchers have evaluated the effectiveness of machine learning methods in temperature prediction, providing insights into their performance and suitability for agricultural applications [20]

In conclusion, the exploration of machine learning techniques in weather forecasting and related fields underscores the significance of accurate temperature prediction. This extends to various applications, including optimizing crop selection strategies, understanding urban transportation dynamics, and enhancing pest management in agriculture. However, while these applications are broad and impactful, this research will narrow its focus to one specific area: Thach Hoa. By directing our attention to this particular region, we aim to delve deeper into the intricacies of temperature prediction and its implications for local agricultural practices. This focused approach will allow us to develop tailored solutions and insights that address the unique challenges and opportunities present in Thach Hoa, thereby contributing to more effective decision-making and sustainable development efforts in the area.

III. Methodology

1. Data Collection

For this study, historical data was collected using an API from the website https://www.visualcrossing.com/weather-api for the Thach Hoa region. Features such as year, month, day, humidity, precipitation, wind speed, wind direction, and solar energy were included **(Fig. 1)**. The data was then saved into Microsoft Excel files in tabular format. Years and days within each month were arranged as rows in the corresponding table, with environmental variables in columns.

We leverage data spanning 14 years, from 2010 to 2024, sourced from a weather station, ensuring comprehensive coverage for our study.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	name	datetime	tempmax	tempmin	temp	feelslikema	feelslikemin	feelslike	dew	humidity	precip	precipprob	precipcover	preciptype	windspeed	winddir	visibility	solarradiatio	solarenergy
2	Xã Thạch Hò	2010-01-01	18.5	16	16.6	18.5	16	16.6	16	96.6	1.2	100	50	rain	14.8	44.4	3.2	112.5	9.7
3	Xã Thạch Hò	2010-01-02	18	15.8	16.8	18	15.8	16.8	15.5	92.2	0.2	100	8.33	rain	13	37.3	6.2	120.3	10.2
4	Xã Thạch Hò	2010-01-03	22	17	19.4	22	17	19.4	15.7	79.6	0	0	0		13	85.1	8.5	135.5	11.7
5	Xã Thạch Hò	2010-01-04	22	18	19.7	22	18	19.7	17.6	87.8	0	0	0		14.8	104.3	4	196.5	17
6	Xã Thạch Hò	2010-01-05	27	19.9	22.4	28.2	19.9	22.4	19.8	86.1	0	0	0		18.4	117.6	4.9	208.5	18
7	Xã Thạch Hò	2010-01-06	23	17.8	20.1	23	17.8	20.1	16.7	81.5	6.9	100	62.5	rain	31.3	55.8	7.5	151.9	13
8	Xã Thạch Hò	2010-01-07	18	13.3	16	18	13.3	16	13.4	85.2	1.4	100	58.33	rain	20.5	43.7	5.7	82.5	7
9	Xã Thạch Hò	2010-01-08	19	13	16	19	13	16	10	68.8	0.7	100	20.83	rain	13	55	6.6	188.9	16.1
10	Xã Thạch Hò	2010-01-09	18.5	13	16.8	18.5	13	16.8	14.5	86.5	0	0	0		18.4	85.8	3.5	173.5	15
11	Xã Thạch Hò	2010-01-10	24	16.9	19.4	24	16.9	19.4	17	87	0	0	0		20.5	350.5	4.9	209.3	18.2
12	Xã Thạch Hò	2010-01-11	18	15.9	16.8	18	15.9	16.8	12.4	75.6	0.5	100	20.83	rain	22.3	39	10.7	137.2	11.8
13	Xã Thạch Hò	2010-01-12	16	14	14.7	16	14	14.7	6.6	58.6	0.5	100	20.83	rain	24.1	28.8	9.7	146.6	12.5
14	Xã Thạch Hò	2010-01-13	16	8.8	14.2	16	7.7	14.2	4.3	52.2	0	0	0		16.6	25.2	6.4	118.4	10.1
15	Xã Thạch Hò	2010-01-14	17.1	13	14.9	17.1	13	14.9	4.5	50.5	0	0	0		11.2	37.3	6.1	192.9	16.8
16	Xã Thạch Hò	2010-01-15	19	12	15.7	19	12	15.7	10.7	73	0.8	100	12.5	rain	25.9	56.6	3.5	131.9	11.3
17	Xã Thạch Hò	2010-01-16	19	14	16.3	19	14	16.3	11.8	76.2	3.3	100	45.83	rain	16.6	14.7	5.9	128.9	11.1

Fig. 1. Sample Data

2. Data Preprocessing

The data preprocessing phase involves several crucial steps to ensure its suitability for modeling purposes and to facilitate the division of the dataset into training and testing sets.

Step 1 select the necessary columns and remove any irrelevant ones. In this study, the selected columns are: datetime, tempmax, tempmin, humidity, precip, precipcover, windspeed, winddir, solarenergy, and solarradiation. Any unnecessary columns are discarded **(Fig. 2.1)**.

	datetime	tempmax	tempmin	dew	humidity	precip	precipcover	windspeed	winddir	solarradiation	solarenergy
0	2010-01-01	18.5	16.0	16.0	96.6	1.2	50.00	14.8	44.4	112.5	9.7
1	2010-01-02	18.0	15.8	15.5	92.2	0.2	8.33	13.0	37.3	120.3	10.2
2	2010-01-03	22.0	17.0	15.7	79.6	0.0	0.00	13.0	85.1	135.5	11.7
3	2010-01-04	22.0	18.0	17.6	87.8	0.0	0.00	14.8	104.3	196.5	17.0
4	2010-01-05	27.0	19.9	19.8	86.1	0.0	0.00	18.4	117.6	208.5	18.0
5	2010-01-06	23.0	17.8	16.7	81.5	6.9	62.50	31.3	55.8	151.9	13.0
6	2010-01-07	18.0	13.3	13.4	85.2	1.4	58.33	20.5	43.7	82.5	7.0
7	2010-01-08	19.0	13.0	10.0	68.8	0.7	20.83	13.0	55.0	188.9	16.1
8	2010-01-09	18.5	13.0	14.5	86.5	0.0	0.00	18.4	85.8	173.5	15.0
9	2010-01-10	24.0	16.9	17.0	87.0	0.0	0.00	20.5	350.5	209.3	18.2

Fig. 2.1. 10 sample lines

Step 2 we check for missing values and handle them accordingly (**Fig. 2.2**). As shown in the image, there are no missing values, so we proceed to the next step.

```
dataa.isnull().sum()
```

```
datetime          0
tempmax           0
tempmin           0
dew               0
humidity          0
precip            0
precipcover       0
windspeed         0
winddir           0
solarradiation    0
solarenergy       0
dtype: int64
```

Fig. 2.2. Checking missing values

Step 3 involves converting the data into numerical formats to prepare for model training (**Fig. 2.3**).

```
dataa.dtypes
```

```
datetime           object
tempmax           float64
tempmin           float64
dew               float64
humidity          float64
precip            float64
precipcover       float64
windspeed         float64
winddir           float64
solarradiation    float64
solarenergy       float64
dtype: object
```

Fig. 2.3. Check data types

Step 4 we plot charts to assess the correlation between variables **(Fig. 2.4)**

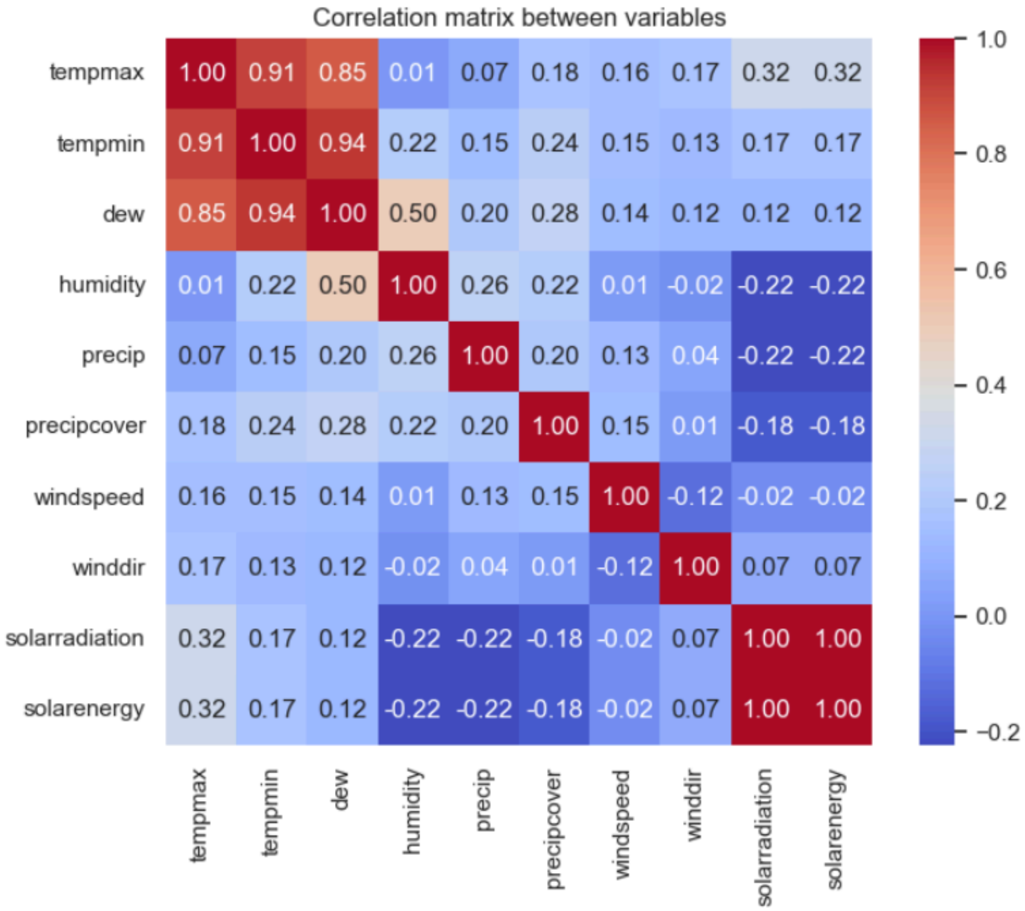

Fig. 2.4. Correlation of data

Positive correlations, such as between dew and tempmax, tempmin and tempmax, are observed. We decide to remove the dew column to reduce complexity, while retaining tempmin as one of the two target columns. Regarding negative correlations, both solarenergy and solarradiation show high correlation; thus, we remove the solarradiation column (**Fig. 2.5**).

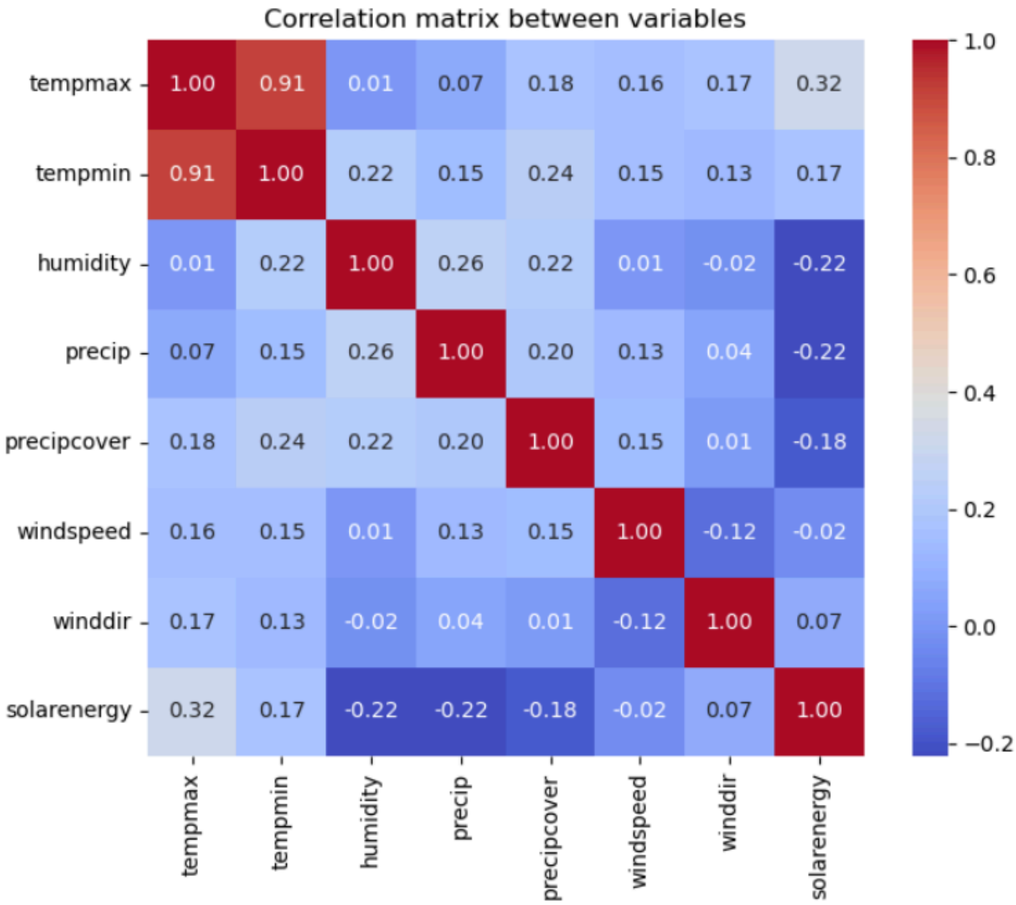

Fig. 2.5. Correlation of data after deletion

Step 5 entails visualizing the distribution of each column for better insight into the data (**Fig. 2.6, 2.7**)

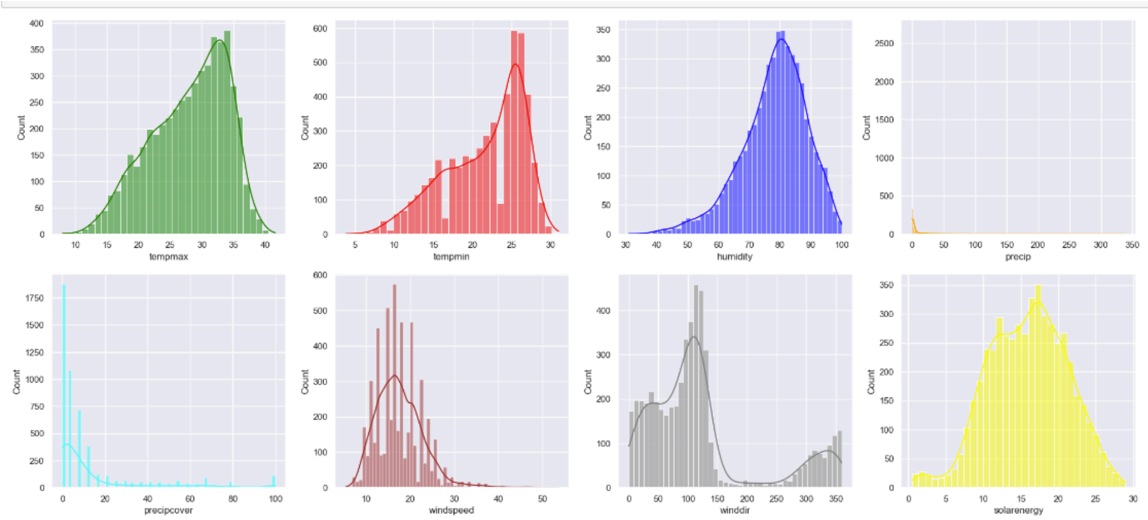

Fig. 2.6. Visualizing using histplot

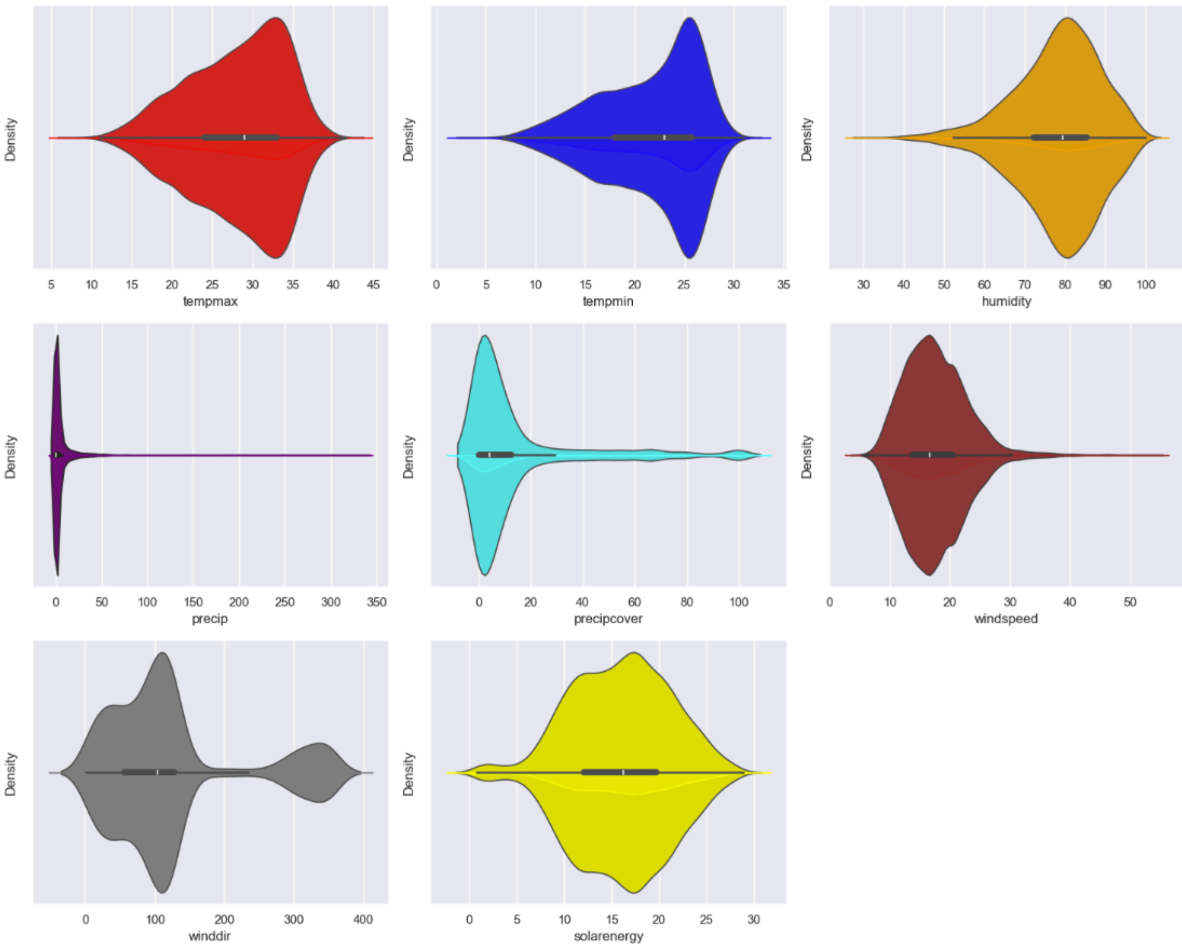

Fig. 2.7. Visualizing using violinplot

Step 6 Calculate the value of the outlier for each column using the IQR method (**Fig. 2.8**)

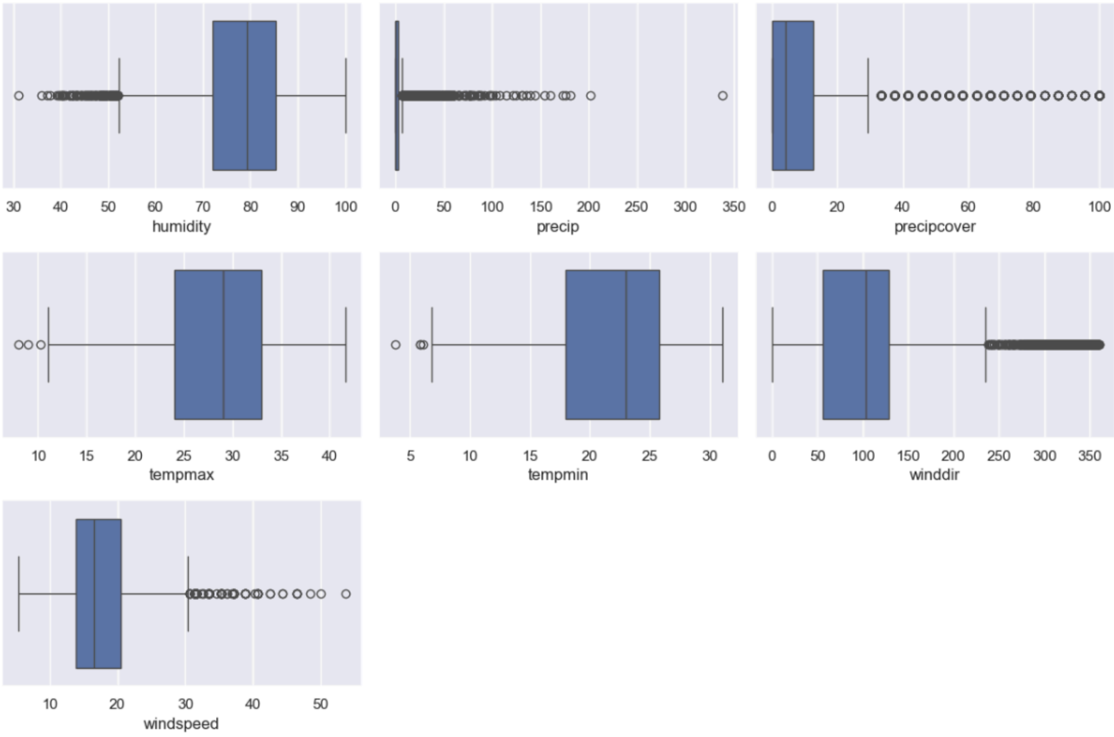

Fig. 2.8. Check for outliers

Step 7 handles outliers by calculating and replacing the mean.

Finally, we proceed to normalize the dataset using the MinMaxScaler method, ensuring uniformity across all features

With the completion of data preprocessing, we bifurcate the dataset into distinct training and testing subsets to assess model efficacy. Critical features essential for temperature prediction are meticulously selected, with the dataset partitioned into 80% for training and 20% for testing, effectively serving as the foundational input for subsequent modeling endeavors.

3. Model

In this study, we employed three distinct machine learning algorithms for temperature prediction in agricultural settings: Linear Regression, Support Vector Regression (SVR), and XGBoost. Each algorithm offers unique advantages and approaches to capturing the complex relationships between input variables and temperature fluctuations.

3.1 Linear Regression

The Linear Regression algorithm constructs a linear model to predict temperature based on input variables. It seeks to find the best-fit line that approximates the relationship between input variables and temperature. This is achieved by optimizing weights for the input variables to minimize the prediction error. Linear Regression aims to provide a straightforward and interpretable model for temperature prediction.

3.2 Support Vector Regression (SVR)

SVR, a regression method rooted in Support Vector Machines (SVM), aims to construct optimal linear or nonlinear functions to approximate input data and predict temperature. By utilizing the principles of SVM, SVR identifies the most suitable function that accurately represents the underlying patterns in the data, thus minimizing prediction errors. SVR offers flexibility in capturing both linear and nonlinear relationships between input variables and temperature.

3.3 XGBoost

XGBoost is an ensemble learning technique that combines multiple decision trees into a unified model for temperature prediction. Through gradient boosting, XGBoost optimizes loss functions iteratively to construct a series of decision trees. Each decision tree contributes to the overall temperature prediction by capturing complex interactions and nonlinear relationships between input variables. XGBoost excels in handling large datasets and offers high predictive accuracy.

To evaluate the performance of these models, we utilized Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) as standard evaluation metrics. These metrics provide insights into the accuracy and reliability of temperature predictions generated by each machine learning algorithm, aiding in informed decision-making in agricultural practices.

MAE calculates the average magnitude of the errors between forecasts and corresponding observations, without considering their direction. It provides a measure of how far, on average, the predictions are from the actual values.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \widehat{y_j}|$$

On the other hand, RMSE is a quadratic scoring rule that determines the average magnitude of the errors by taking the square root of the average of the squared differences between predictions and actual observations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \widehat{y_j})^2}$$

R-squared (R^2) is a statistical measurement used to evaluate how well an expected model fits actual data. The R^2 value ranges from 0 to 1. An R^2 value close to 1 is chosen to account for the appropriate model being a large fraction of the expected temperature variation.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

If R^2 = 1, the model fits the data perfectly and accurately expects temperature. If R^2 near 0, it indicates that the model is not suitable for expected temperature variability and is more stochastic in its prediction.

IV. Results

We assessed the performance of the machine learning models in predicting both the maximum and minimum temperatures for agricultural planning. The predictions were compared against the actual observed temperatures to evaluate the accuracy and reliability of each model.

Model Performance

We evaluated the performance of our models in predicting daily maximum and minimum temperatures using the following evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2).

4.1 Linear Regression

- Maximum Temperature (**Fig. 4.1.1, Fig. 4.1.2**)

```
Test Set Mean Squared Error for tempmax: 4.346374651834435
Train Set Mean Squared Error for tempmax: 4.227681560406169
Test Set Root Mean Squared Error for tempmax: 2.084796069603556
Train Set Root Mean Squared Error for tempmax: 2.056132670915515
Test Set R-squared for tempmax: 0.8814084206355618
Train Set R-squared for tempmax: 0.8809123749479472

Linear Regression Results for tempmax:
            Actual  Predicted              Model Temperature Type
datetime
2021-01-19    21.0  17.803702  Linear Regression          tempmax
2012-05-03    37.0  36.123734  Linear Regression          tempmax
2012-06-18    36.0  36.839249  Linear Regression          tempmax
2023-03-23    32.3  32.131695  Linear Regression          tempmax
2022-01-23    24.7  23.604535  Linear Regression          tempmax
2016-02-16    16.1  19.532414  Linear Regression          tempmax
2014-02-25    22.0  24.723103  Linear Regression          tempmax
```

Fig. 4.1.1. Evaluation results for tempmax

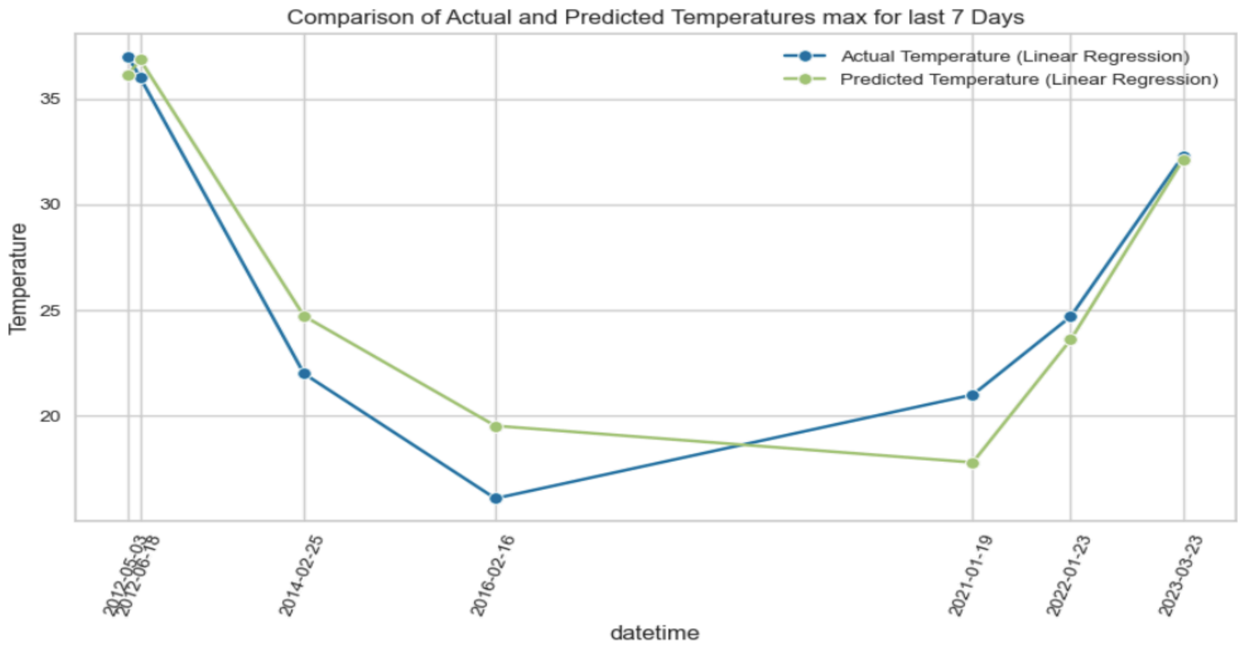

Fig. 4.1.2. Lineplot illustration for tempmax

- Minimum Temperature (**Fig. 4.1.3, Fig. 4.1.4**)

```
Test Set Mean Squared Error for tempmin: 3.1989451890639957
Train Set Mean Squared Error for tempmin: 3.131709188850151
Test Set Root Mean Squared Error for tempmin: 1.7885595290803142
Train Set Root Mean Squared Error for tempmin: 1.769663580698363
Test Set R-squared for tempmin: 0.8706715344713349
Train Set R-squared for tempmin: 0.8736702981127094

Linear Regression Results for tempmin:
            Actual  Predicted              Model Temperature Type
datetime
2021-01-19    10.0  14.419366  Linear Regression          tempmin
2012-05-03    27.0  26.942774  Linear Regression          tempmin
2012-06-18    28.0  26.832149  Linear Regression          tempmin
2023-03-23    24.8  24.454899  Linear Regression          tempmin
2022-01-23    18.0  19.654891  Linear Regression          tempmin
2016-02-16    12.0  10.747345  Linear Regression          tempmin
2014-02-25    18.8  17.456596  Linear Regression          tempmin
```

Fig. 4.1.3. Evaluation results for tempmin

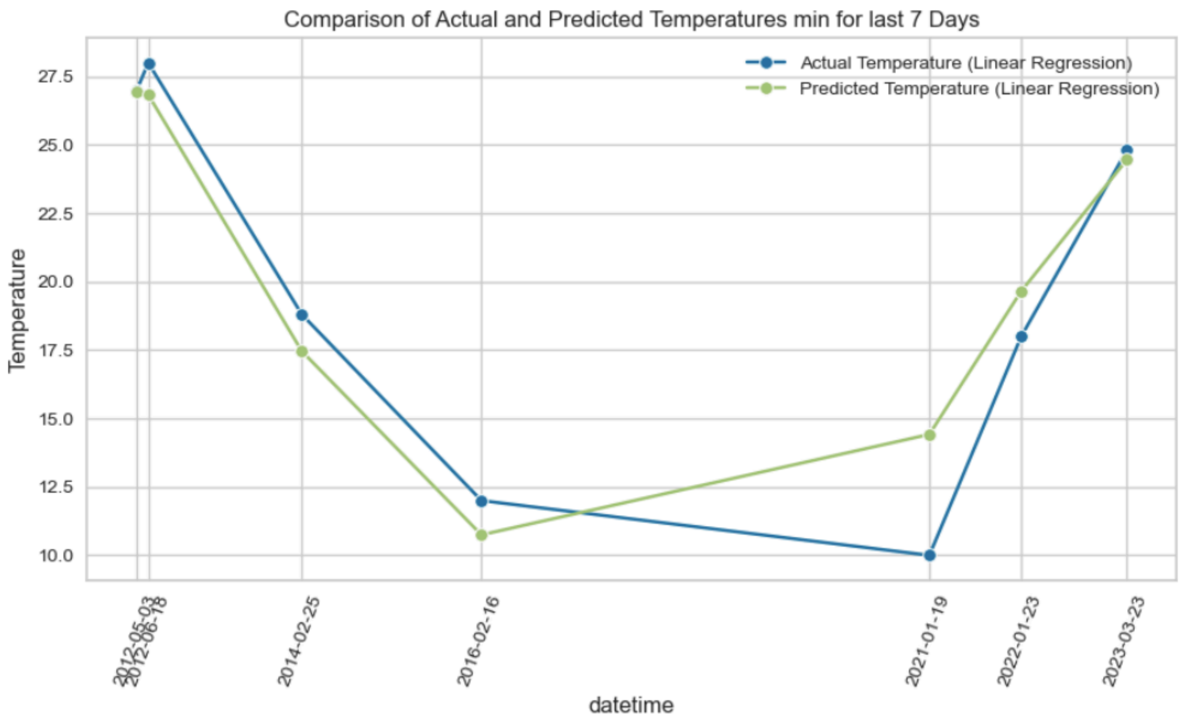

Fig. 4.1.4. Lineplot illustration for tempmin

4.2 Support Vector Regression (SVR)

- Maximum Temperature (**Fig. 4.2.1, Fig. 4.2.2**)

```
Test Set Mean Squared Error for tempmax (SVR): 4.882487789519589
Train Set Mean Squared Error for tempmax (SVR): 4.943606834468043
Test Set Root Mean Squared Error for tempmax (SVR): 2.20963521639197
Train Set Root Mean Squared Error for tempmax (SVR): 2.2234223248110205
Test Set R-squared for tempmax (SVR): 0.866780481534805
Train Set R-squared for tempmax (SVR): 0.860745803888944

SVR Results for tempmax:
            Actual  Predicted Model Temperature Type
datetime
2021-01-19    21.0  20.793494   SVR          tempmax
2012-05-03    37.0  35.963231   SVR          tempmax
2012-06-18    36.0  36.417938   SVR          tempmax
2023-03-23    32.3  32.033618   SVR          tempmax
2022-01-23    24.7  24.613381   SVR          tempmax
2016-02-16    16.1  20.460785   SVR          tempmax
2014-02-25    22.0  25.087836   SVR          tempmax
```

Fig 4.2.1 Evaluation results for tempmax

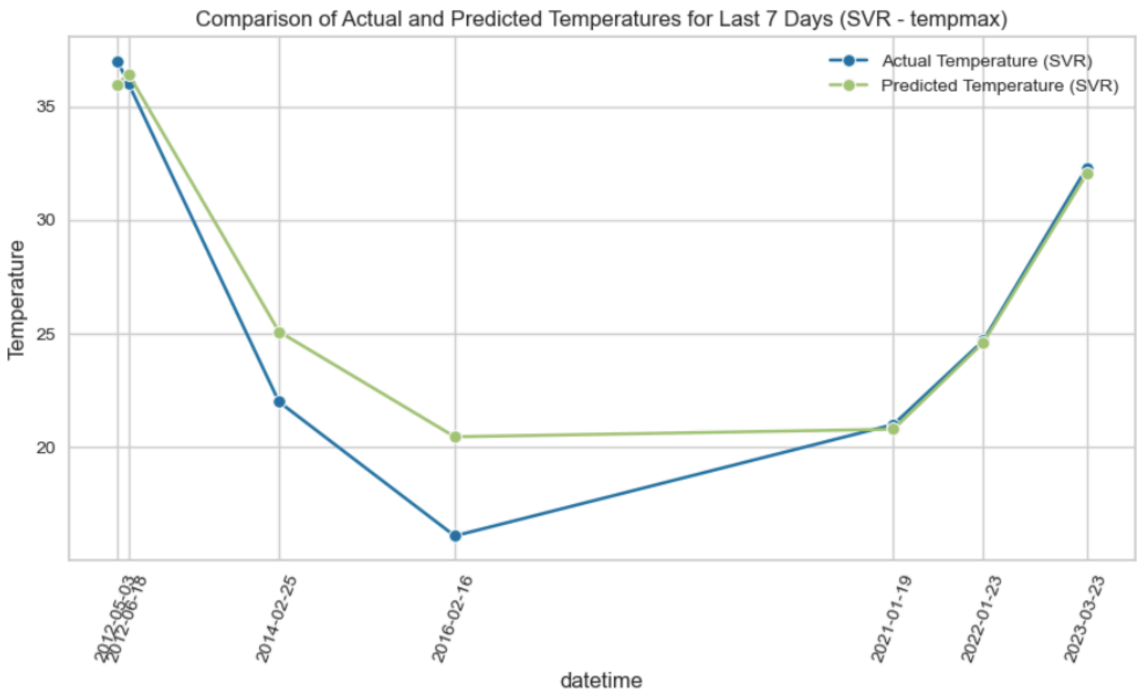

Fig. 4.2.2. Lineplot illustration for tempmax

- Minimum Temperature (**Fig. 4.2.3, Fig. 4.2.4**)

```
Test Set Mean Squared Error for tempmin (SVR): 3.5112543484339143
Train Set Mean Squared Error for tempmin (SVR): 3.5358816765090624
Test Set Root Mean Squared Error for tempmin (SVR): 1.8738341304485608
Train Set Root Mean Squared Error for tempmin (SVR): 1.880394021610647
Test Set R-squared for tempmin (SVR): 0.8580453524129681
Train Set R-squared for tempmin (SVR): 0.8573664247968917

SVR Results for tempmin:
            Actual  Predicted Model Temperature Type
datetime
2021-01-19    10.0  15.262272   SVR          tempmin
2012-05-03    27.0  26.802281   SVR          tempmin
2012-06-18    28.0  26.674774   SVR          tempmin
2023-03-23    24.8  24.562338   SVR          tempmin
2022-01-23    18.0  20.918021   SVR          tempmin
2016-02-16    12.0  12.884694   SVR          tempmin
2014-02-25    18.8  18.588623   SVR          tempmin
```

Fig. 4.2.3. Evaluation results for tempmin

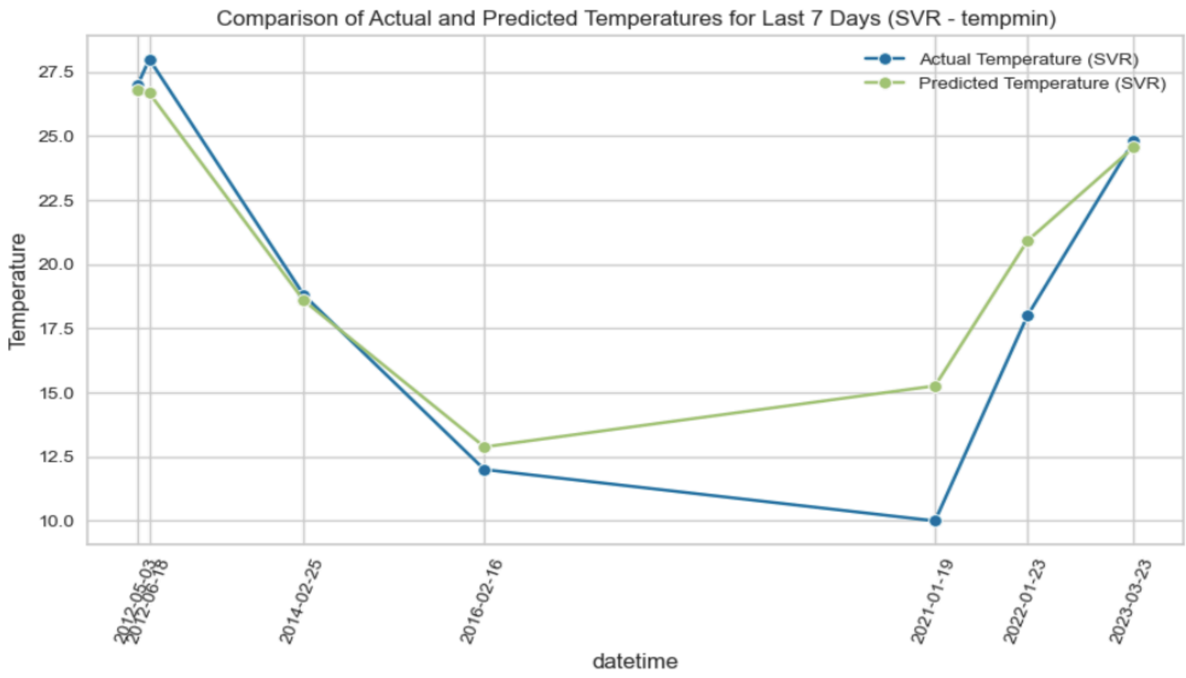

Fig. 4.2.4. Lineplot illustration for tempmin

4.3 XGBoost

XGBoost (Before Overfitting Mitigation)

Before addressing the issue of overfitting, the XGBoost model exhibited the following performance metrics in predicting maximum and minimum temperatures:

- Maximum Temperature **(Fig. 4.3.1, Fig. 4.3.2)**

```
Test Set Mean Squared Error for tempmax (XGBoost): 3.693283223491405
Train Set Mean Squared Error for tempmax (XGBoost): 0.7455564755859758
Test Set Root Mean Squared Error for tempmax (XGBoost): 1.9217916701587103
Train Set Root Mean Squared Error for tempmax (XGBoost): 0.8634561225597834
Test Set R-squared for tempmax (XGBoost): 0.899228132501378
Train Set R-squared for tempmax (XGBoost): 0.9789987612001736

XGBoost Results for tempmax:
            Actual  Predicted    Model Temperature Type
datetime
2021-01-19    21.0  20.449900  XGBoost          tempmax
2012-05-03    37.0  37.220024  XGBoost          tempmax
2012-06-18    36.0  37.417812  XGBoost          tempmax
2023-03-23    32.3  32.811668  XGBoost          tempmax
2022-01-23    24.7  22.178982  XGBoost          tempmax
2016-02-16    16.1  19.025278  XGBoost          tempmax
2014-02-25    22.0  23.301252  XGBoost          tempmax
```

Fig. 4.3.1. Results of tempmax assessment when overfitting

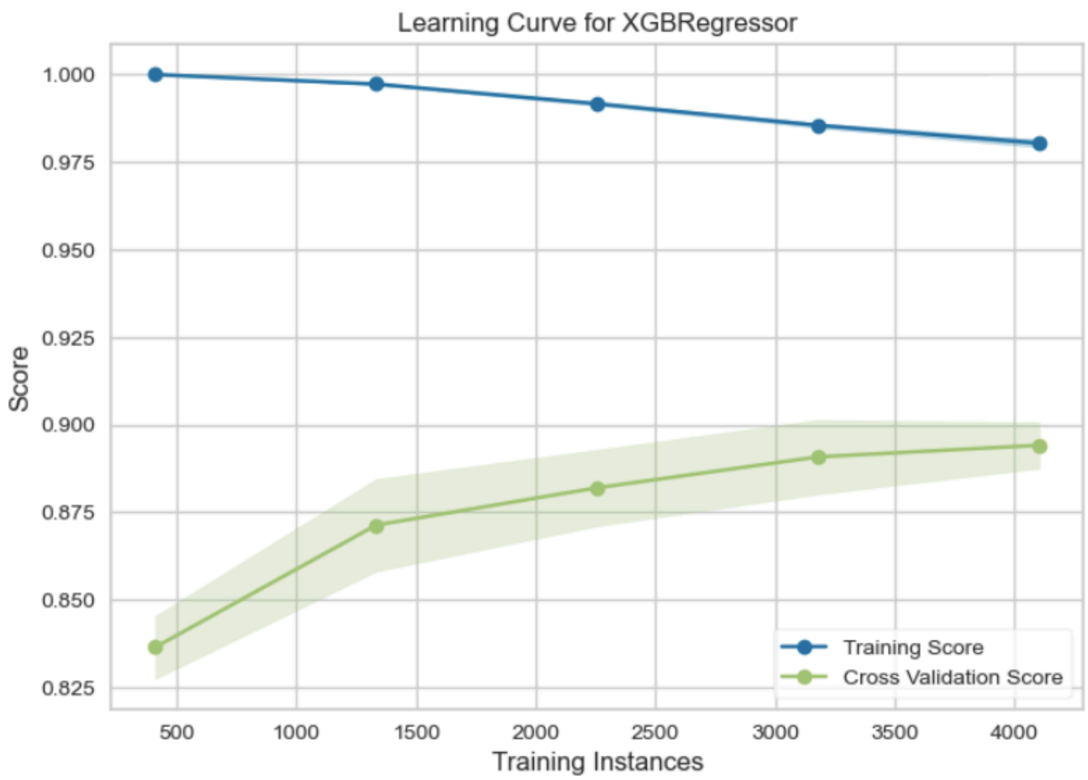

Fig. 4.3.2. Learning curve chart evaluates overfitting

- Minimum Temperature **(Fig. 4.3.3, Fig. 4.3.4)**

```
Test Set Mean Squared Error for tempmin (XGBoost): 2.5113715664677136
Train Set Mean Squared Error for tempmin (XGBoost): 0.509841409569159
Test Set R-squared for tempmin (XGBoost): 0.8984690853178943
Train Set R-squared for tempmin (XGBoost): 0.9794335586746111
Test Set Root Mean Squared Error for tempmin (XGBoost): 1.584730755197145
Train Set Root Mean Squared Error for tempmax (XGBoost): 0.7140317987100848

XGBoost Results for tempmin:
            Actual  Predicted    Model Temperature Type
datetime
2021-01-19    10.0  12.481799  XGBoost          tempmin
2012-05-03    27.0  27.751976  XGBoost          tempmin
2012-06-18    28.0  26.629633  XGBoost          tempmin
2023-03-23    24.8  24.643600  XGBoost          tempmin
2022-01-23    18.0  21.028711  XGBoost          tempmin
2016-02-16    12.0  13.027866  XGBoost          tempmin
2014-02-25    18.8  18.687027  XGBoost          tempmin
```

Fig. 4.3.3. Results of tempmin assessment when overfitting

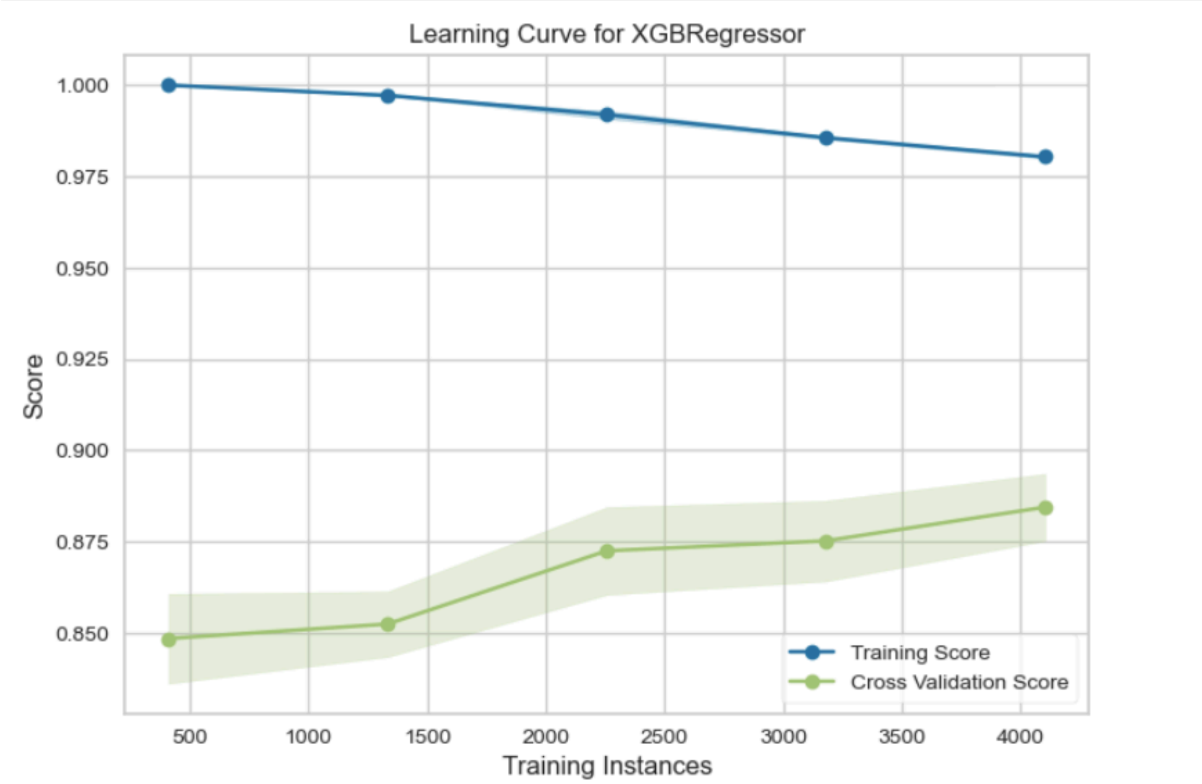

Fig. 4.3.4. Learning curve chart evaluates overfitting

XGBoost (After Overfitting Mitigation)

After addressing the issue of overfitting in the XGBoost model, we observed improvements in its performance in predicting both maximum and minimum temperatures:

- Maximum Temperature **(Fig 4.3.5, Fig 4.3.6, Fig 4.3.7)**

```
Test Set Mean Squared Error: 3.124920987111419
Train Set Mean Squared Error: 1.8292166271963441
Test Set Root Mean Squared Error: 1.7677446046053766
Train Set Root Mean Squared Error: 1.352485351934114
Test Set R-squared: 0.9147359938025106
Train Set R-squared: 0.9484736348454753

XGBoost Results for tempmax:
            Actual  Predicted    Model Temperature Type
datetime
2021-01-19    21.0  21.234159  XGBoost          tempmax
2012-05-03    37.0  37.121212  XGBoost          tempmax
2012-06-18    36.0  36.521019  XGBoost          tempmax
2023-03-23    32.3  31.594252  XGBoost          tempmax
2022-01-23    24.7  22.514416  XGBoost          tempmax
2016-02-16    16.1  18.661448  XGBoost          tempmax
2014-02-25    22.0  23.815035  XGBoost          tempmax
```

Fig. 4.3.5. Results of tempmax evaluation when handling overfitting

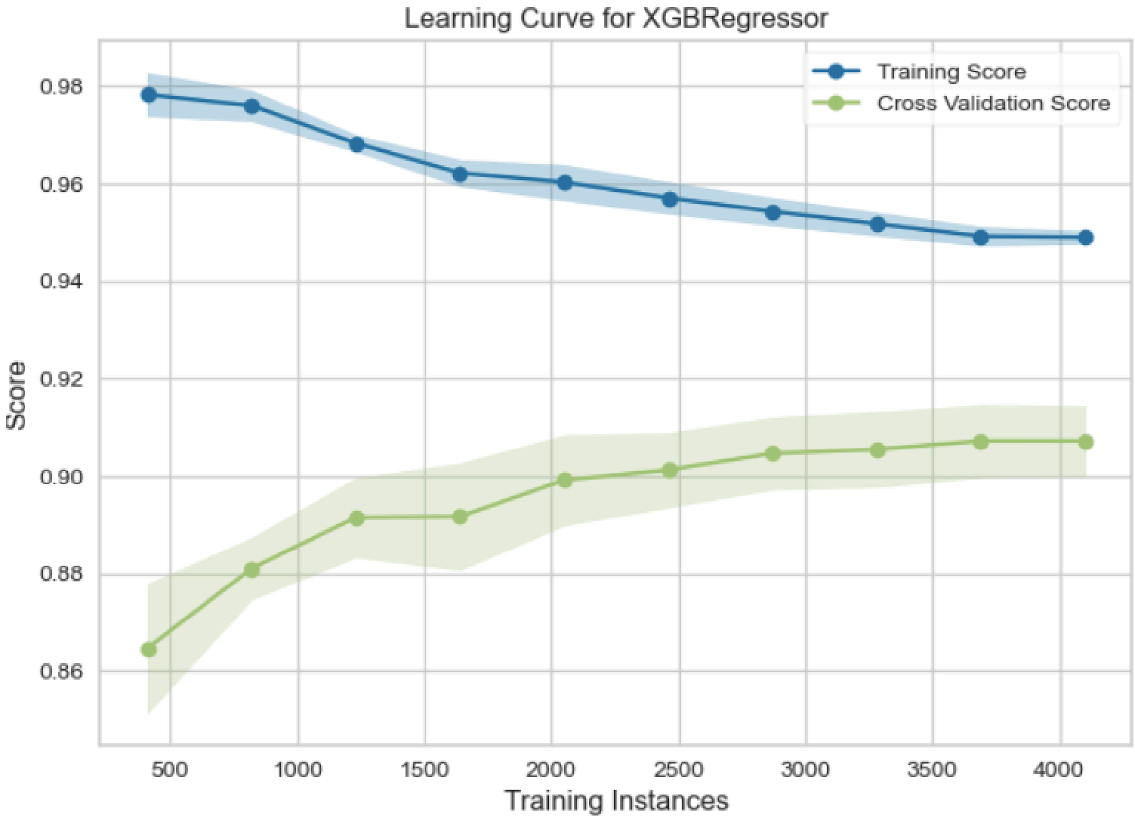

Fig. 4.3.6. Learning curve chart after overfitting processing

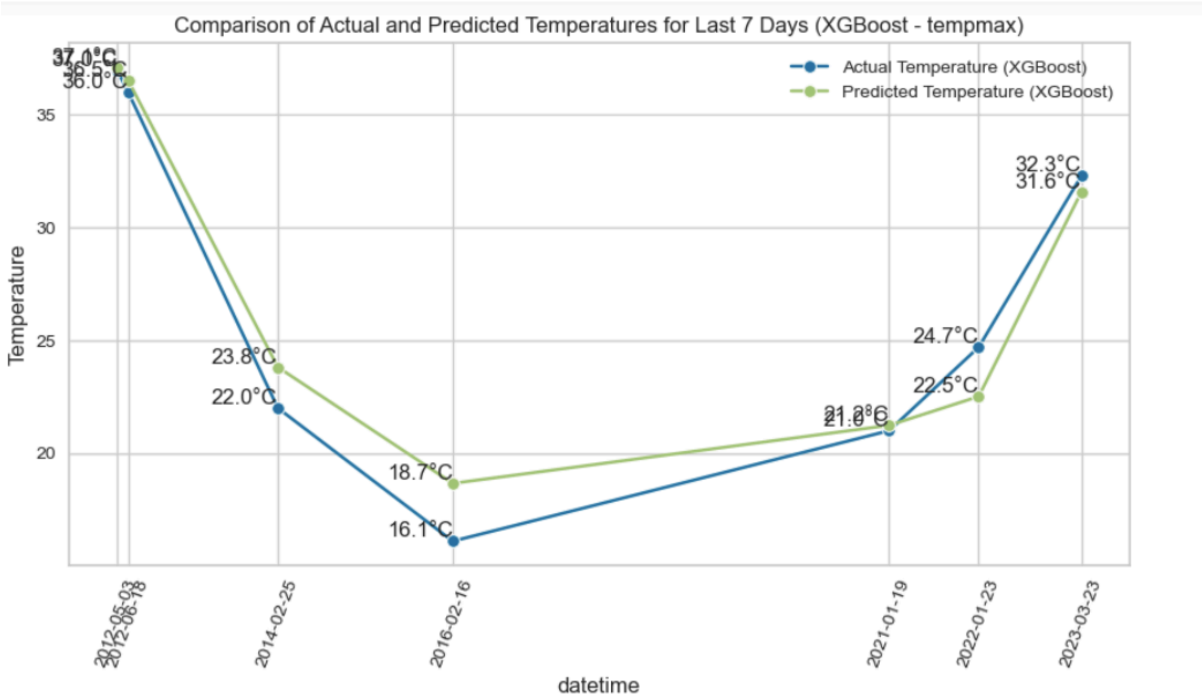

Fig. 4.3.7. Lineplot illustration for tempmax

- Minimum Temperature **(Fig. 4.3.8, Fig. 4.3.9, Fig. 4.3.10)**

```
Train Set Mean Squared Error for tempmin (XGBoost): 2.0977266085081347
Train Set R-squared for tempmin (XGBoost): 0.915380017391982
Test Set Mean Squared Error for tempmin (XGBoost): 2.614768208607035
Test Set R-squared for tempmin (XGBoost): 0.8942889170816865
Train Set Root Mean Squared Error for tempmin (XGBoost): 1.4483530676282406
Test Set Root Mean Squared Error for tempmin (XGBoost): 1.6170244922718502

XGBoost Results for tempmin:
            Actual  Predicted    Model Temperature Type
datetime
2021-01-19    10.0  12.730933  XGBoost          tempmin
2012-05-03    27.0  27.372696  XGBoost          tempmin
2012-06-18    28.0  26.958458  XGBoost          tempmin
2023-03-23    24.8  24.987003  XGBoost          tempmin
2022-01-23    18.0  20.611650  XGBoost          tempmin
2016-02-16    12.0  12.698373  XGBoost          tempmin
2014-02-25    18.8  17.842281  XGBoost          tempmin
```

Fig. 4.3.8. Results of tempmin evaluation when handling overfitting

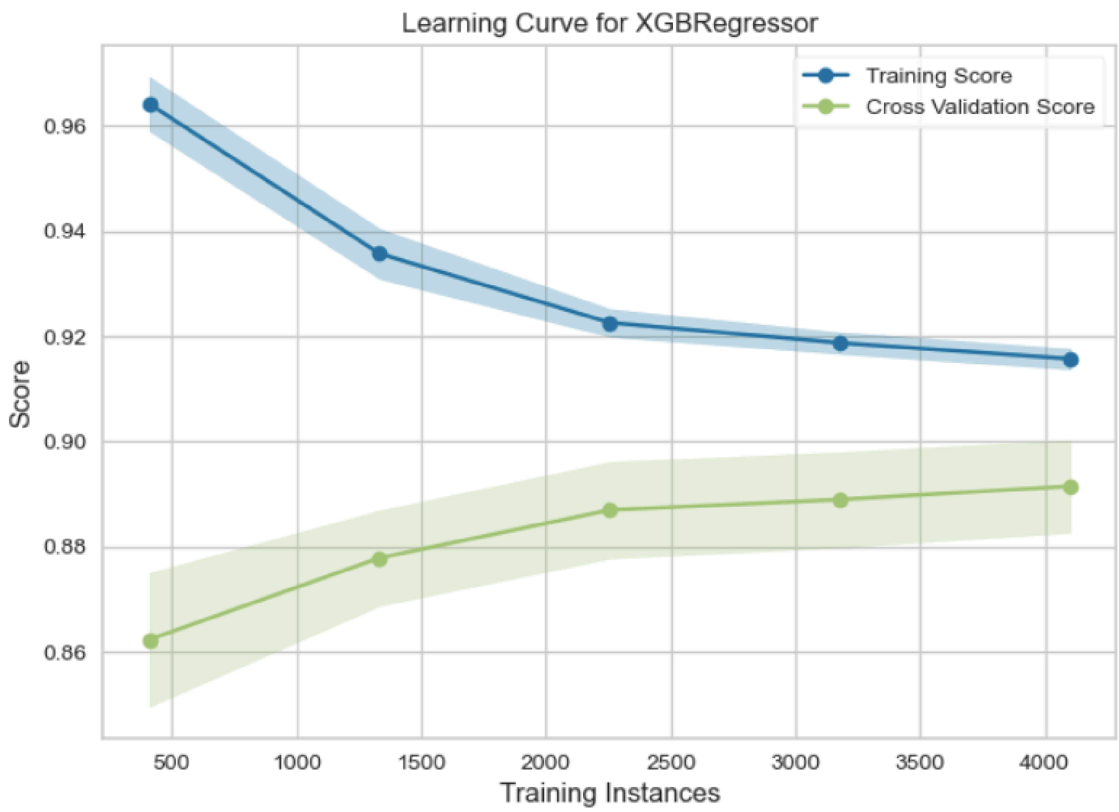

Fig. 4.3.9. Learning curve chart after overfitting processing

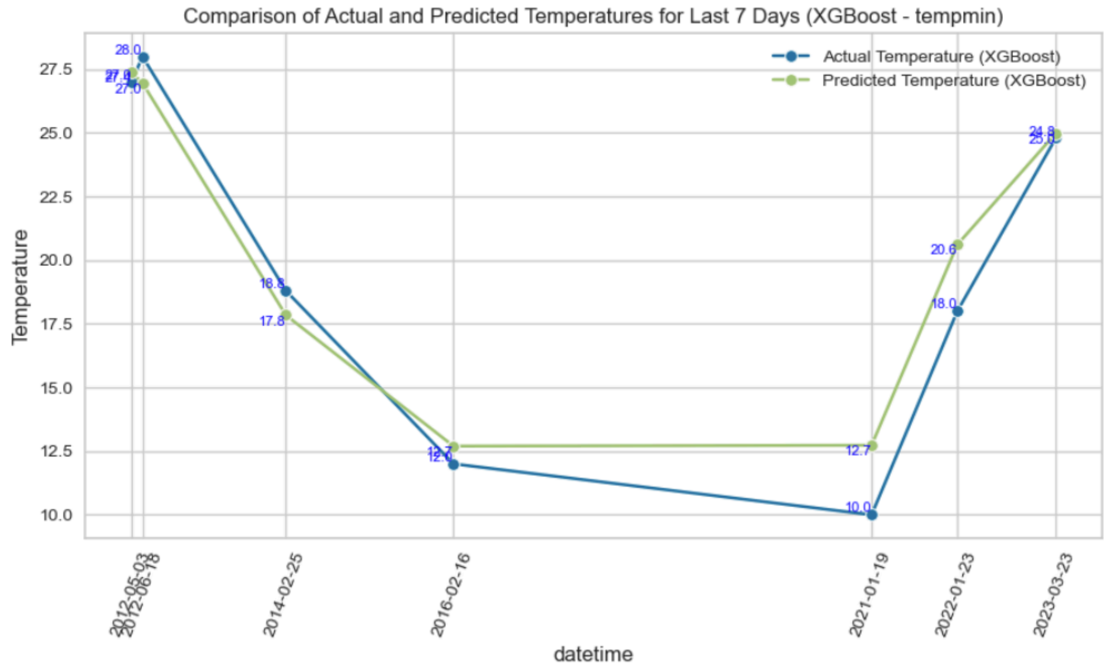

Fig. 4.3.10. Lineplot illustration for tempmin

V. Discussion

Linear Regression Model

The Linear Regression model demonstrated robust performance in predicting both daily maximum and minimum temperatures for agricultural planning. On the test set, the model achieved a Mean Squared Error (MSE) of 4.35 for maximum temperature and 3.20 for minimum temperature. The Root Mean Squared Error (RMSE) was 2.08 for maximum temperature and 1.79 for minimum temperature, indicating relatively accurate predictions. Additionally, the R-squared values of 0.88 for maximum temperature and 0.87 for minimum temperature suggest that the model captures a significant portion of the variance in the temperature data.

Support Vector Regression (SVR) Model

The SVR model also exhibited promising results in temperature prediction. For maximum temperature, the model achieved a test set MSE of 4.88 and an RMSE of 2.21, with an R-squared value of 0.87, indicating a good fit to the data. Similarly, for minimum temperature, the SVR model attained a test set MSE of 3.51 and an RMSE of 1.87, with an

R-squared value of 0.86. These results suggest that the SVR model provides accurate predictions for both temperature extremes.

XGBoost Model

Initially, the XGBoost model showed signs of overfitting, necessitating mitigation measures. Through the application of GridSearchCV and RandomizedSearchCV techniques, overfitting was addressed. Interestingly, for maximum temperature prediction, GridSearchCV yielded superior results, while for minimum temperature prediction, RandomizedSearchCV proved to be more effective. This highlights the importance of optimizing hyperparameters tailored to specific prediction tasks. Post-overfitting mitigation, the XGBoost model exhibited improved performance.

Overall, the Linear Regression and SVR models demonstrated strong performance in temperature prediction, while the XGBoost model required overfitting mitigation strategies to enhance its predictive capabilities. These findings underscore the significance of selecting appropriate modeling techniques and optimizing hyperparameters for accurate temperature forecasting in agriculture.

VI. Conclusion

The study assessed the performance of three machine learning models - Linear Regression, Support Vector Regression (SVR), and XGBoost - in predicting daily maximum and minimum temperatures for agricultural planning.

Linear Regression and SVR models demonstrated strong performance in temperature prediction, providing reliable forecasts with mean squared errors (MSE) ranging from approximately 3.20 to 4.88 and Root Mean Squared Errors (RMSE) ranging from around 1.79 to 2.21. These models also exhibited high R-squared values, indicating a good fit for the temperature data.

Initially, the XGBoost model encountered overfitting issues, requiring mitigation measures such as GridSearchCV and RandomizedSearchCV techniques for hyperparameter optimization. Following overfitting mitigation, the XGBoost model exhibited significant performance improvements. The model achieved a test set MSE of 3.12 and an RMSE of 1.77, with an R-squared value of 0.91 for maximum temperature prediction. Similarly, for minimum temperature prediction, the XGBoost model attained a test set MSE of 2.61 and an RMSE of 1.62, with an R-squared value of 0.89. These results underscore the effectiveness of overfitting mitigation strategies in enhancing the predictive capabilities of the XGBoost model.

Visual representations comparing the performance metrics, such as RMSE, MSE, and R-squared, for both maximum and minimum temperatures are provided in the accompanying figures **(Fig 6.1, Fig 6.2)**. These visuals offer insights into the relative performance of each model and highlight the improvements achieved through overfitting mitigation in the XGBoost model.

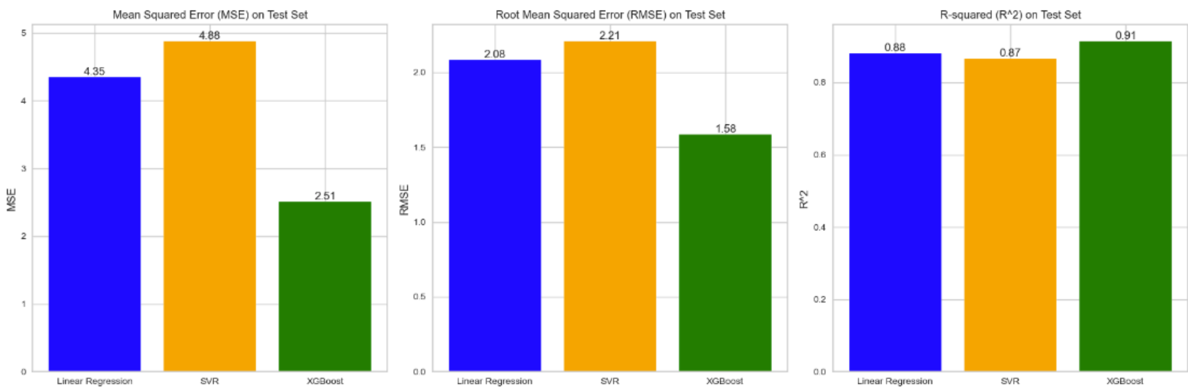

Fig. 6.1. Compare model evaluation for tempmax

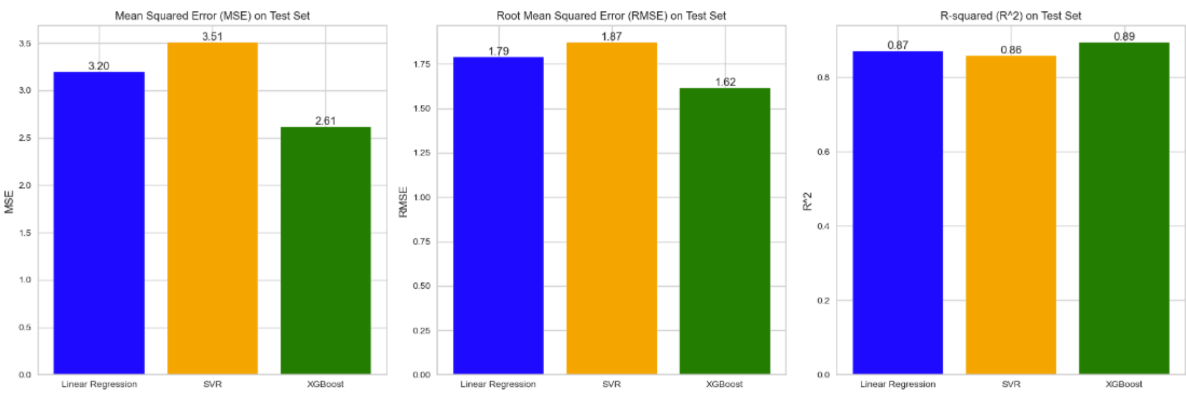

Fig. 6.2. Compare model evaluation for tempmin

In conclusion, the study emphasizes the importance of selecting appropriate modeling techniques and optimizing hyperparameters to ensure accurate temperature forecasting in agricultural applications. The findings provide valuable guidance for agricultural practitioners, enabling informed decision-making and improved crop management strategies based on reliable temperature predictions.

VII. References

[1] M. Wang, Q. Xu, Y. Cao, S. G. Hassan, W. Liu, M. He, T. Liu, L. Xu, L. Cao, S. Liu and H. Wu, "An Ensemble Model for Water Temperature Prediction in Intensive Aquaculture," *IEEE,* no. 1, 2023.

[2] A. Pelosi, G. B. Chirico, S. F. Bolognesi, C. D. Michele and G. D'Urso, "Forecasting crop evapotranspiration under standard conditions in precision farming," *IEEE,* no. 2, 2019.

[3] T. Anjali, K. Chandini, K. Anoop and V. L. Lajish, "Temperature Prediction using Machine Learning," *IEEE,* no. 3, 2019.

[4] A. Sharaff and S. R. Roy, "Comparative Analysis of Temperature Prediction Using Regression Methods and Back Propagation Neural Network," *IEEE,* no. 4, 2018.

[5] S. Madan, P. Kumar, S. Rawat and T. Choudhury, "Analysis of Weather Prediction using," *IEEE,* no. 5, 2018.

[6] I.-C. Chen and S.-C. Hu, "Realizing Specific Weather Forecast through Machine Learning Enabled Prediction Model," *ACM,* no. 6, 2019.

[7] C. Z. Basha, N. Bhavana, P. Bhavya and S. V, "Rainfall Prediction Using Machine Learning & Deep," *IEEE,* no. 7, 2020.

[8] A. A and A. Titus, "PREDICTION OF WEATER FORECASTCONDITIONS FORSMART AGRICULTURE SUPPORTED BY MACHINE LEARNING," *IEEE,* no. 8, 2020.

[9] D. N. Fente and D. K. Singh, "Weather Forecasting Using Artificial Neural Network," *IEEE,* no. 9, 2018.

[10] S. Jain and D. Ramesh, "Machine Learning convergence for weather based crop selection," *IEEE,* no. 10, 2020.

[11] M. Schultz, S. Reitmann and S. Alam, "Classification of weather impacts on airport operations," *ACM,* no. 11, 2020.

[12] S. Jiang and C. Cai, "The impacts of weather conditions on metro ridership: An empirical study from three mega cities in China," *Sciencedirect,* no. 12, 2020.

[13] M. S. Tahsin, S. Abdullah, M. A. Karim, M. U. Ahmed, F. Tafannum and M. Y. Ara., "A comparative study on data mining models for weather forecasting," *Sciencedirect,* no. 13, 2023.

[14] A. Muniasamy, "Machine Learning for Smart Farming: A Focus on Desert Agriculture," *IEEE,* no. 14, 2020.

[15] E. Almeyda, J. Paiva and W. Ipanaqué, "Pest Incidence Prediction in Organic Banana Crops with Machine Learning Techniques," *IEEE,* no. 15, 2020.

[16] G. Peng and W. Cun, "Wind Power Prediction Based on Back Propagation Algorithm with Numerical Weather Prediction," *ACM,* no. 16, 2020.

[17] N.Sravanthi, M. Venka, S.Harshini and K.Ashesh, "An Ensemble Approach to Predict Weather," *IEEE,* no. 17, 2020.

[18] F. Wu, S. Lu, L.-A. Armando and J. She, "Temperature Prediction Based on Long Short Term Memory Networks," *IEEE,* no. 18, 2020.

[19] S. Nguyen-Xuan and N. L. Nhat, “A dynamic model for temperature prediction in glass greenhouse,” *IEEE,* no. 19, 2019.

[20] B. Azari, K. Hassan, J. Pierce and S. Ebrahimi, “Evaluation of Machine Learning Methods Application in Temperature Prediction,” *ResearchGate,* no. 20, 2022.