Analysis of Climate Change and its Impact on Agriculture in India using Machine Learning

Dishank Chauhan
Department of Computer Engineering
Thakur College of Engineerig and Technology

Vishal Bartwal
Department of Computer Engineering
Thakur College of Engineering and Technology

Vedant Bhatia
Department of Computer Engineering
Thakur College of Engineering and Technology

Dr. Vaishali Nirgude
Department of Computer Engineering
Thakur College of Engineering and Technology

***Abstract*—Weather prediction and patterns have a profound impact on agriculture, but despite the huge technological advancements in predicting weather, there remains a gap in showing the impact of those weather predictions in the field of agriculture.**

This study aims to analyze and predict weather data using advanced machine learning algorithms and show the impact of the predicted weather on different crops. We used a dataset of Mumbai city containing five years of weather data, performed data preprocessing and visualization, and applied ARIMA and LSTM for time series forecasting. LSTM outperformed ARIMA in predictive accuracy, as shown by lower MSE and RMSE values.

By providing accurate weather forecasts specifically tailored to agricultural needs, this research can help farmers improve their resilience towards unpredictable weather and ultimately lead to productive and sustainable farming.

Keywords - data analysis, machine learning, LSTM, ARIMA, visualization, agriculture, impact.

I. INTRODUCTION

Weather conditions are incredibly unpredictable and change rapidly. Correct forecasts are essential these days especially for agriculture, industries, traveling or even daily commuting which are heavily dependent on weather [2]. In a country like India where agriculture supports most of the population the sector itself is highly vulnerable to climate change as erratic weather conditions like increasing temperature, rainfall frequency and wind can have a massive impact on crop yields which requires solutions to help and mitigate these risks.

With the current rise of advanced algorithms especially in machine learning which has the power to analyze a huge amount of data, we have the ability to accurately predict weather using machine learning algorithms like time series forecasting. For instance research done with the use of models such as CNN (Convolutional Neural Networks) or algorithms like SVR (Support Vector Regression) or MLPR highly accurate weather predictions have been received [1][2]. Despite these advancements not many of them are directly linked towards the agricultural sector especially in the country of India which if done would benefit current and the next generation of farmers with the new technologies coming in the future.

This research aims to address this gap by using these advanced machine learning algorithms with it being time series forecasting, to analyze weather data especially rare occurrences of events such eg - rainfall in the month of February and accurately predict those conditions. This way the impact on agriculture can be shown providing farmers with insights and improving their ability to grow crops in unpredictable weather.

LITERATURE SURVEY

The literature survey reviews a variety of machine learning algorithms applied to weather prediction and its impact on agriculture. Here's a brief summary of each paper:

1. Weather Prediction Using CNN-LSTM (2024)

This paper utilizes advanced models like CNN-LSTM for time series analysis of Delhi temperature data. The focus is on effective feature engineering and preprocessing, achieving a solid performance with an RMSE of 1.80.

2. Smart Weather Forecasting Using Multiple ML Models (2020)

A comparative study using algorithms like Random Forest, SVR, and others to predict weather in Tennessee. The paper emphasizes model comparisons and inclusion of neighboring cities, yielding an RMSE of 3.0.

3. Weather Forecast Prediction: Integrated Approach (2018)

Using Naive Bayes and Chi-square, this study focuses on effective data mining with high accuracy (93%). The accuracy largely depends on the quality of input data.

4. Comprehensive Review of Crop Prediction Using Weather Data (2024)

This review does not use specific models but discusses how weather data can provide insights into crop growth patterns. No performance measures are included.

5. Estimation of Missing Precipitation Data Using ML (2023)

The study applies ANN and RF for predicting missing precipitation data. The RF model achieves an accuracy of 84%, highlighting its capability for non-linear data.

6. Machine Learning Applied to Weather Forecasting (2016)

Focused on both short-term and long-term forecasting, this study uses linear and functional regression. It advocates for simpler algorithms, though no performance metrics are provided.

7. Weather Prediction Using Random Forest (2021)

This study highlights the use of Random Forest for predicting solar radiation and wind speeds, with strong results (MSE 0.750, R^2 0.97).

8. Intelligent Weather Forecasting (2018)

Multi-target regression and RNN are used for predicting multiple weather variables simultaneously, though the paper does not provide specific performance metrics.

9. Machine Learning-Based Smart Weather Prediction (2022)

This paper explores how decision trees and linear regression can be used to aid agricultural planning, with an MSE of 0.739.

10. Prediction of Rainfall Using ML (2020)

Using long-term data (1901–2015), the study applies MLR, SVM, and Lasso Regression to predict rainfall, achieving high performance with R^2 values close to 1.

Overall, the survey shows a variety of machine learning approaches being used for weather prediction, with models like Random Forest and CNN-LSTM performing well for time series forecasting. These models also help in agricultural planning by offering predictions on crop growth patterns based on weather. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Title	Algorithm	Advantages	Gaps/Future Scope	Performance Measure
Weather Prediction Using CNN-LSTMfor Time SeriesAnalysis: A Case Study on Delhi Temperature Data-2024 [1]	CNN - LSTM	Advanced models Effective feature engineering Data preprocessing	Highly complex algorithms.	MSE - 3.26 RMSE - 1.80
Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee-2020 [2]	RFR Ridge SVR MLPR ETR	Multiple model comparisons Inclusion of neighbour cities	Model accuracy changes depending on inclusion and exclusion of neighbouring cities.	RMSE - 3.0
Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data-2018 [3]	Naïve Bayes Chi square	User friendly Effective Data Mining High accuracy	The accuracy of the predictions heavily rely on the quality of the data.	Accuracy - 93% Precision - 94%
Comprehensive Review of Crop Prediction using Weather data and Machine Learning approaches-2024 [4]	NA	Provides valuable data driven insights in crop growth patterns based on the weather.	The paper does not include any new models or methodologies.	NA
Machine learning approach for the estimation of missing precipitation data: a case study of South Korea (Research Article)-2023 [5]	ANN RF	The ANN model is very suitable for non linear data such as the one used here.	ANN is a highly complex model.	RF accuracy - 84% ANN accuracy - NA
Machine Learning Applied to Weather Forecasting-2016 [6]	Linear Regression Functional Regression	Focus on short term and long term analysis along with comparison of performance of the algorithms used is very insightful.	Use of simple algorithms, advanced techniques like random forest, time series forecasting can offer better performance and accuracy.	MSE - NA RMSE - NA
Weather prediction using random forest machine learning model (Research Article)-2021 [7]	Random Forest	Use of a Comprehensive dataset which, essential for predicting solar radiation and wind speeds.	RF is comparatively a simpler algorithm, advanced algorithms can might perform better.	MSE - 0.750 R^2 - 0.97
Intelligent Weather Forecasting using Machine Learning Techniques-2020 [8]	Multi Target Regression RNN	Multi Target Regression allows prediction of multible variables simultaneously	Model accuracy heavily depends upon the quality of the data where gaps and errors can lead to inaccuracy.	NA
Machine learning based smart weather prediction-2022 [9]	Linear Regression Decision Tree Empirical Models	The predicted parameters are directly relevant to agricultural planning.	Use of montly data which gives less information as compared to day to day dataset.	MSE - 0.739
Prediction of Rainfall using Machine Learning-2020 [10]	MLR SVM Lasso Regression	Use of a long term dataset which spans from 1901-2015 which gives a rich history of weather data.	The models used might not be piritculary suited towards a non linear data such as weather as supposed to other options.	SVR MAE - 4.35 SVR R^2 - 0.9959 MLR MAE - 10.99 MLR R^2 - 0.9957 LR MAE - 10.75 LR R^2 - 0.9957

II. METHODOLOGY

In this study, we aim to use advanced machine learning algorithms such as LSTM ARIMA (time series forecasting) in order to obtain accurate weather predictions and show its potential impact on crops.

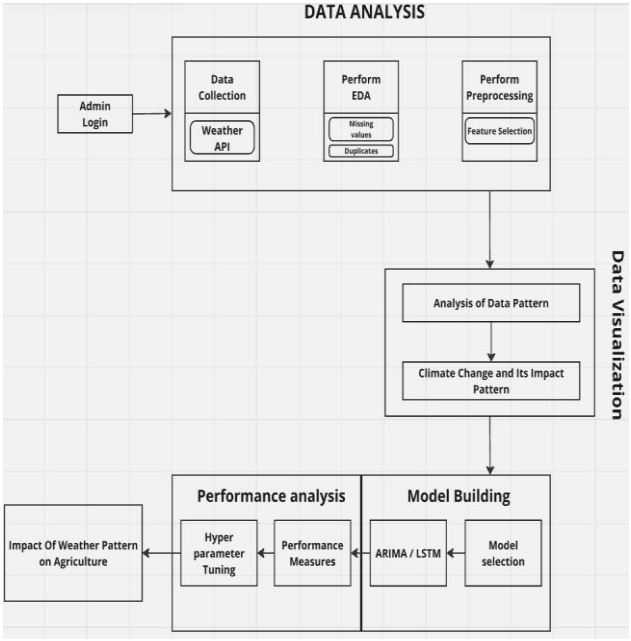

Fig 1 – System Architecture

Data Collection: This project uses dataset extracted from Kaggle. The dataset contains historical meteorological data for Mumbai city including temperature, humidity, and precipitation, spanning multiple years from 2016 to 2021. These datasets were integrated to analyze climate patterns and their impact on agriculture.

Data Cleaning and Preprocessing: Several preprocessing steps were performed to improve data quality and consistency. Data types were standardized, duplicate records were removed along with null values, and measurement units were unified to create a structured dataset for weather prediction modeling.

Data Analysis: then the data was analyzed, parameter by parameter, and found out 5 suitable parameters which to perform predictions on using the deep learning models (ARIMA, LSTM).

Data Visualization – the next step is to visualize the dataset which was done using a combination of matplotlib and seaborn, for each parameter such as temperature and humidity different visualizations were made such as parameter over time charts, box plots, distribution plots to show and analyse how each parameter changes throughout the year and ultimately year by year. Lastly a heatmap correlation was plotted in order to show the relationship between each parameter.

Model Selection – ARIMA and LSTM were chosen for weather forecasting. ARIMA is a statistical model that captures linear patterns in time series data, while LSTM, a deep learning model, handles complex and nonlinear trends. Using both allows comparing traditional and machine learning approaches based on accuracy metrics like MSE and RMSE.

Model Building - We used a dataset spanning five years (2016–2021) and split it into 75% training and 25% testing data for both ARIMA and LSTM models. The SARIMA model was designed to capture trends and seasonality, with carefully selected (p, d, q) and (P, D, Q, m) parameters. It was trained on historical data, and forecasts were generated along with confidence intervals to assess prediction reliability. The LSTM model was built to capture complex patterns in time series data, using an LSTM layer followed by a dense output layer. Predictions were made on the test set, inverse transformed for interpretability and visualized against actual values. Both models' forecasts were compared to evaluate their effectiveness in weather prediction.

Performance metrics – Model Performance – The performance of both models was evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The SARIMA model achieved an MSE of 3.15 and an RMSE of 1.77, while the LSTM model performed better with an MSE of 1.74 and an RMSE of 1.32. The lower error values for LSTM suggest that it captured patterns in the data more effectively, making it a better choice for weather forecasting in this case.

PERFORMANCE	ARIMA	LSTM
MSE	3.15	1.77
RMSE	1.75	1.3

Fig 2 – Performance Metrics

III. RESULT

The project focused on analyzing and predicting weather patterns to understand their impact on agriculture in Mumbai. A literature review was conducted to explore existing forecasting techniques and identify gaps in their use for agriculture in India. A five-year weather dataset (2016–2021) was collected, including temperature, humidity, wind speed, and precipitation.

Data Analysis and Visualization:

The data was cleaned by handling missing values, standardizing units, and removing errors. It was then analyzed using Pandas and visualized with Matplotlib and Seaborn to show trends over time and relationships between different weather parameters.

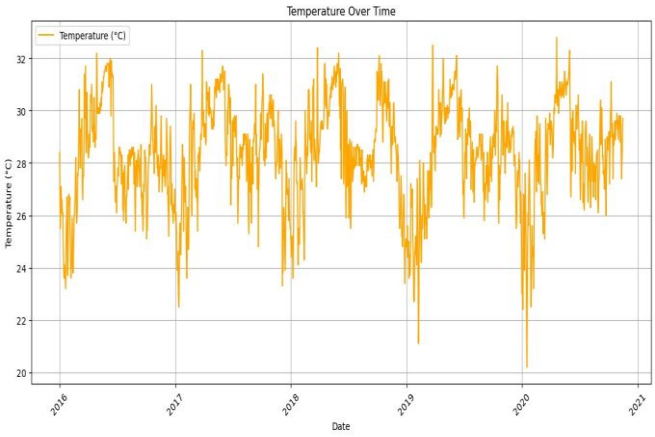

Fig 3 – Temperature over time chart

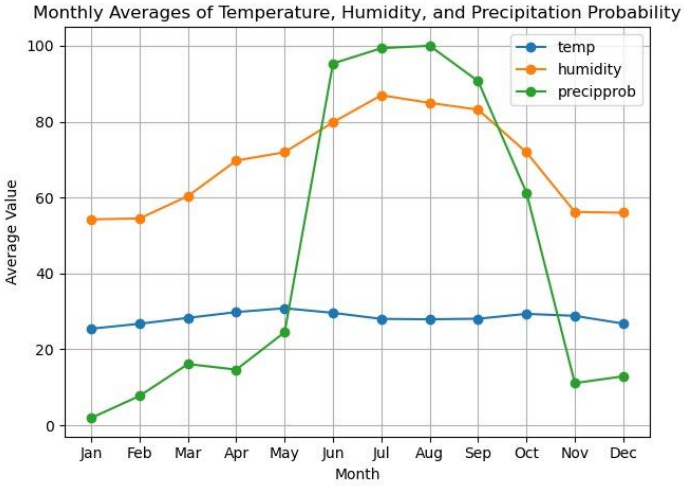

Fig 4 – Parameter monthly averages

These visualizations helped in understanding seasonal weather changes and how different factors like humidity and temperature are related.

Forecasting and Model Performance:

For weather prediction, ARIMA and LSTM models were used. The dataset was split into 75% training and 25% testing, and predictions were generated for key weather parameters.

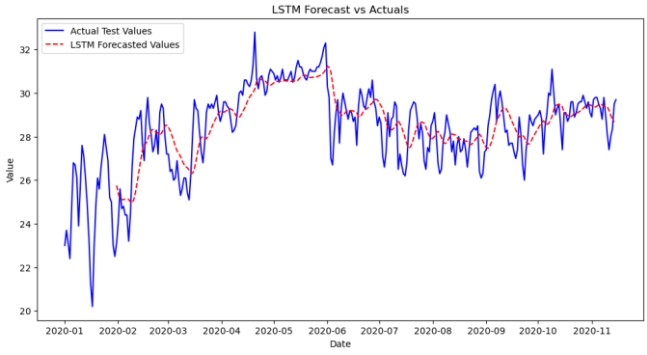

Fig 5 – LSTM prediction

The models were evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Results showed that while ARIMA worked well for linear trends, LSTM had lower error values, meaning it was more accurate in predicting complex weather patterns.

Impact on Agriculture:
By providing better weather forecasts, this study can help farmers plan more effectively and reduce the risks of unexpected weather changes. Future improvements can be made by using real-time data to make predictions even more reliable.

IV. CONCLUSION

This research highlights the effectiveness of machine learning models, including ARIMA, LSTM, and Linear Regression, in improving weather forecasting accuracy. By analyzing Mumbai's historical weather data, we established patterns that can inform agricultural practices. The findings suggest that integrating these advanced forecasting techniques can empower farmers to make better decisions, enhancing crop resilience to unpredictable weather. This work emphasizes the need for practical applications of machine learning in agriculture, paving the way for more sustainable farming methods.

REFERENCES

[1] A. H. M. Jakaria, M. M. Hossain, and M. A. Rahman, "Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee," in *Proceedings of the International Conference on Computing and Data Science*, Cookeville, TN, USA, 2020.

[2] B. Li and Y. Qian, "Weather prediction using CNN-LSTM for time series analysis: A case study on Delhi temperature data," *Applied and Computational Engineering*, vol. 92, no. 1, pp. 121–127, Sep. 2024, DOI: 10.54254/2755-2721/92/20241738.

[3] M. Biswas, T. Dhoom, and S. Barua, "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data," *International Journal of Computer Applications*, vol. 182, no. 34, pp. 20–24, Dec. 2018, DOI: 10.5120/ijca2018918265.

[4] A. Humnabadkar, B. Shivbhakta, A. Karve, and S. V. Hon, "Comprehensive Review of Crop Prediction Techniques using Weather Data and Machine Learning Approaches," *International Journal of Engineering Research & Technology (IJERT)*, vol. 13, no. 9, pp. 1-10, Sep. 2024.

[5] H. Han, B. Kim, K. Kim, D. Kim, and H. S. Kim, "Machine Learning Approach for the Estimation of Missing Precipitation Data: A Case Study of South Korea," *Journal of Water Resources Planning and Management*, vol. 150, no. 2, pp. 123-130, Feb. 2024.

[6] [6] M. Holmstrom, D. Liu, and C. Vo, "Machine Learning Applied to Weather Forecasting," Stanford University, December 15, 2016.

[7] [7] R. Meenal, P. Angel, and D. P. Ekambaram Rajasekaran, "Weather Prediction Using Random Forest Machine Learning Model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 1208-1215, May 2021. DOI: 10.11591/ijeecs.v22.i2.pp1208-1215.

[8] [8] P. N. Kashyap, P. M. S., R. V. Nayak, U. B., and R. T. V., "Intelligent Weather Forecasting Using Machine Learning Techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 07, no. 03, pp. 2513, Mar. 2020. e-ISSN: 2395-0056, p-ISSN: 2395-0072.

[9] [9] R. Meenal, K. Kailash, P. Angel, J. Joseph, and others, "Machine Learning Based Smart Weather Prediction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 508, Oct. 2022. DOI: 10.11591/ijeecs.v28.i1.pp508-515.

[10] [10] M. Mohammed, R. Kolapalli, N. Golla, and S. S. Maturi, "Prediction of Rainfall Using Machine Learning Techniques," *International Journal of Scientific & Technology Research*, vol. 9, no. 01, pp. 3236, Jan. 2020. ISSN 2277-8616.