A Major Project Report on

"Crop Yield Prediction for Google Grid"

Submitted in fulfilment of the requirement for Degree in Bachelor of Engineering in Information Technology

By

Sahil Bagate (5021105)

Neil Dsouza (5021113)

Dhanashree Patil (5021142)

Suyash Malekar (5021166)

Guided by : **Prof. Poonam Bari**

Department of Information Technology

Fr. Conceicao Rodrigues Institute of Technology

(An Autonomous Institute Permanently Affiliated to University of Mumbai)

Sector 9A, Vashi, Navi Mumbai – 400703

University of Mumbai

2024-2025

CERTIFICATE

This is to certify that the project entitled

"Crop Yield Prediction for Google Grid"

Submitted By

Sahil Bagate (5021105)
Neil Dsouza (5021113)
Dhanashree Patil (5021142)
Suyash Malekar (5021166)

In partial fulfilment of the degree of **B.E.** in **Information Technology** for term work of the project is approved.

External Examiner

———————————

Internal Examiner

———————————

Internal Guide

———————————

Head of the Department

———————————

Principal

———————————

Date :

College Seal

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented, fabricated, or falsified any idea/-data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited or from whom proper permission has not been taken when needed.

Sahil Bagate (5021105)

−−−−−−−−−−

Neil Dsouza (5021113))

−−−−−−−−−−

Dhanashree Patil (5021142)

−−−−−−−−−−

Suyash Malekar (5021166)

−−−−−−−−−−

Date :

ABSTRACT

Maize is a crucial crop worldwide, serving as food, livestock feed, and a key ingredient in biofuel production. However, its productivity is heavily influenced by changing climate conditions, resource availability, and market demands. This study examines the trends of maize production over 39 years (1982–2020) using historical climate and agricultural data from the top maize producing countries, including India, the United States, Ukraine, Brazil, and South Africa. The study aims to understand how environmental circumstances affect maize yield by analyzing factors such as rainfall, temperature, and CO_2 concentration. Data analytics and geospatial analysis are utilized to investigate the relationships between these traits and crop yields. The study also explores how rainfall during different growth stages affects yield, offering insights into the optimal moisture conditions for maize cultivation. Furthermore, a comparison of state-level data was conducted within India to illustrate regional disparities in production and climate resilience. In addition, Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) models were developed, and their precision was compared to predict maize yield for the future period from 2021 to 2046 for the top maize producing countries. To make the findings more accessible, a graphical user interface (GUI) was built to showcase the insights of the actual and predicted data. The results of this study are designed to support farmers and agricultural specialists in developing smarter farming practices to increase yield and address climate change challenges, ultimately contributing to global food security.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
INTRODUCTION

1.1 BACKGROUND

Maize, commonly known as corn, is one of the most significant crops cultivated globally. It serves multiple essential purposes as a primary food source, a vital component in livestock feed, and a key ingredient in the production of biofuels. The global reliance on maize emphasizes the necessity to ensure its stable and sustainable production. However, maize yields are highly sensitive to varying climate conditions, including changes in rainfall patterns, temperature fluctuations, and atmospheric carbon dioxide (CO_2) levels. Additionally, resource availability, market demands, and technological advancements further influence its productivity.

Understanding how these factors interact with maize growth and yield is critical, especially in the face of ongoing climate change. Unpredictable environmental conditions have already begun impacting agricultural productivity in many regions, posing a threat to food security worldwide. Hence, comprehensive studies that analyze historical trends and forecast future outcomes are necessary to inform smarter agricultural practices and policy decisions.

1.2 PROBLEM DEFINITION

Despite advancements in agricultural technology, climate variability continues to challenge maize production. Farmers and agricultural policymakers often lack localized, data-driven insights into how specific environmental factors affect yields. Moreover, the absence of reliable future yield predictions hampers long-term planning and resilience-building efforts. To bridge this gap, it is essential to leverage modern analytical techniques, machine learning models, and visualization tools to study historical patterns and predict future trends in maize production.

1.3 OBJECTIVES OF THE STUDY

The primary objectives of this project are:

- To examine historical maize production trends from 1982 to 2020 across major maize-producing countries — India, the United States, Ukraine, Brazil, and South Africa — using climatic and agricultural data.
- To analyze the relationship between maize yield and environmental factors such as rainfall, temperature, and CO_2 concentration.
- To investigate how rainfall during different crop growth stages impacts overall maize productivity.
- To perform a state-level comparison within India to understand regional disparities in maize production and climate resilience.
- To develop and compare LSTM and BiLSTM models for predicting maize yields from 2021 to 2046.
- To design a graphical user interface (GUI) that displays insights derived from actual historical data and model-based predictions.

1.4 SCOPE OF THE STUDY

This study focuses on both global and regional scales:

- **Global Analysis:** Analyzing trends in top maize-producing countries to identify universal and location-specific patterns affecting maize production.
- **India-Specific Analysis:** Exploring variations at the state level within India to highlight regional vulnerabilities and strengths.
- **Model Development:** Building LSTM and BiLSTM models to forecast future maize yields, comparing their performance, and selecting the most accurate model.
- **Visualization:** Developing an easy-to-use GUI to make the insights from data analysis and predictions accessible to a broader audience, including farmers, researchers, and policymakers.

1.5 SIGNIFICANCE OF THE STUDY

This study holds significant importance in the context of global food security and agricultural sustainability. By providing reliable insights into how climate variables impact maize yield and offering robust future predictions, the project can aid farmers in adapting to changing environmental conditions. Policymakers can also benefit from these findings when designing climate-resilient agricultural policies and investment plans.

Furthermore, the use of advanced deep learning models (LSTM and BiLSTM) and an interactive GUI ensures that the knowledge generated is both cutting-edge and accessible, ultimately contributing to a more informed and prepared agricultural sector.

Chapter 2
LITERATURE SURVEY

2.1 LITERATURE SURVEY

2.1.1 Literature Survey on Agricultural Data Analysis Techniques

Agricultural productivity is increasingly influenced by climate variability, soil characteristics, and agronomic practices. Understanding these complex interactions is critical for optimizing crop management strategies and ensuring global food security. The rapid advancement of data-driven analytics and high-resolution climate models has provided unprecedented opportunities to analyze long-term crop yield trends with high precision. [1] examined northwest of Iran for agricultural yield using both qualitative and quantitative factors. They discovered that yield was affected by nitrogen and phosphorus fertilizer, as well as planting density control. Researchers also stated that qualitative factors such as rotating crops and variation played a role for the yield of crops.[2] conducted a comprehensive review in which they gathered details comprising 76 research in order to comprehensively examine and study the impacts of various fertilizers strategies, nutrients of the soil, amount of water, and effectiveness of water consumption on northern China's maize yield. Surface biomass is an important metric for crop development in land use planning and ecological monitoring. The findings showed a substantial correlation among image-derived plant heights and manually determined plant heights. The vegetative indicator showed a good association with biomass. temperature and plant height showed a substantial association with biomass [3].[4] conducted a geographically and historically extensive investigation of yield problems in crops during 1975 to 2010. Yield discrepancies have expanded gradually across almost every region for the yearly crops, whereas sugar cane and oil palm yields have remained unchanged.[5] studied temperature and rainfall trends in the north Maharashtra through data spanning 1982 to 2022. They discovered that temperature and precipitation measures are rising, posing difficulties to the agricultural sector. As a result, proactive alert systems are required to decrease the adverse effects of global warming on the agriculture sector. [6]provided field data regarding the regional level correlation between atmospheric temperatures and the loss of forests from the southeast Peruvian Amazon. The findings indicated a weak to moderately positive relationship between climate and forestry. Rainfall and forest loss, however, did barely correlate [7]presented a thoroughly built and complete database designed for deep examination of yields of crops in Telangana. The dataset were rigorously obtained from reliable agricultural and meteorological organisations. The collection encompasses a varied array of information, including district-specific identities and essential ecological indicators.[8] found that although soil and climate have a substantial impact on how plants make use of resources, climate factors plays a role more to crop yield than nutrients in the soil parameters do. [9] examined the pattern in cotton yield, they discovered that air humidity and temperature performed differently at various locations, evaluating the effect of their shifting patterns on cotton yield. Whereas temperature is trending upward, yield has been discovered to be negatively correlated with temperature. They noted that the cotton crop might face significant challenges in the future due to rising temperatures.[10] evaluated data from the Samastipur area of Bihar and examined how weather conditions affected maize and wheat crop yields. According to an investigation of temporal data on weather parameters, maximum and minimum temperatures had a negative but substantial influence on maize and wheat yields, but yearly rainfall correlated positively with yields,. Additionally, it showed that crop yield could be influenced by other elements like soil fertility, type, and agricultural practice.[11] investigated climatic oscillations, which are meteorological and oceanographic occurrences that fluctuate on a regular basis. Global differences in food yields and meteorological conditions correspond to climatic oscillations. They examined the global effects of climate oscillations on the growth circumstances of maize, rice, soybean, and wheat using agricultural yield statistics from a series of worldwide grid-based crop models imitated for a variety of agricultural scenarios.[12] investigated sub-district agricultural statistical information for Kutai Barat and Mahakam Ulu and discovered that high rubber yield are concentrated in the south-east area, but low rubber yields are more likely to be seen in the north forest land. It was also noticed that high oil palm yield were more likely to be seen in regions with rainfall, as typically tropical oil palm farms are garnered using rainfall water delivery systems. To evaluate the empirical correlations between agricultural yield and climatic factors, multiple variables were used by [13]. Experimental results were addressed at the district level. It was demonstrated that weather parameters had an impact on crop yield, but not evenly across different crops and seasons of growth. They demonstrated the detrimental effect of visible climate trends on crop yields in both the summer and winter seasons. Potato, wheat, paddy, and maize yields are increasing but fluctuating over time, whereas yields for millet and barley, two minor cereal crops, are increasing constantly. Aside from weather conditions, crop cultivation methods have an important influence on deciding yields and mitigating the negative effects of climate change. As a result, rainfall and temperature projections do not accurately predict crop yield [13].

The reviewed literature highlights the significant influence of climate and agronomic factors on crop yields, with most studies focused on specific crops or limited geographical regions. To address these limitations, the present study utilizes a comprehensive dataset spanning 39 years (1982–2020) derived from Google Earth Engine, encompassing global-scale maize yield data. The dataset integrates key environmental parameters such as precipitation, shortwave radiation (rds), and temperature, along with important agronomic factors like soil texture, nitrogen application, and atmospheric CO_2 concentration. By leveraging advanced data analytics and domain-specific agricultural insights, this study aims to develop a robust predictive model to assess maize yield trends under evolving climate conditions. This integrated approach provides a broader, data-driven perspective for understanding crop yield dynamics in the face of global environmental challenges.

2.1.2 Literature Survey on Crop Yield Prediction Models

The application of machine learning (ML) techniques in crop yield prediction has garnered significant attention, with various studies exploring different models and datasets to enhance prediction accuracy. These studies predominantly focus on specific regions, crops, and variables, offering valuable insights into optimizing agricultural decision-making. The results, while promising, highlight both the strengths and limitations of current models in terms of generalizability, computational requirements, and the influence of environmental factors. [14] focused on the application of ML algorithms like LSTM, RNN, Random Forest, and XGBoost to predict crop yields. The dataset they used included over 100 years of weather data and 17 years of crop production data from India. Their model helped farmers make informed decisions about crop selection by revealing historical trends in weather and agriculture. However, the dataset was limited to Indian districts, restricting the model's broader application. Additionally, the models like LSTM and SimpleRNN faced challenges in predicting rainfall and temperature, complicating the integration of various models for comprehensive prediction. [15] used various machine learning algorithms, including SVM, CNN, and DNN, to predict crop yields by analyzing data on soil, weather, and crop characteristics. The study highlighted the effectiveness of ML in handling large and complex datasets to make accurate predictions. Adaptive models like ANN, which improve predictions with limited data, were also explored. However, some algorithms, such as K-NN, struggled with non-linear data, leading to misclassifications and reduced accuracy when dealing with complex patterns. [16] employed KNN for feature extraction and crop classification, aiming to predict suitable crops based on soil parameters (N, P, K, pH) and climate conditions. Their model, applied across 15 agro-climatic regions, demonstrated high accuracy and low computational time, making it user-friendly for farmers. However, the accuracy of the predictions depended heavily on the quality of the available data. Moreover, the model did not account for sudden climatic events, which could impact crop yields unpredictably, limiting its overall reliability and generalizability across different crops and regions. [17] applied several machine learning techniques, including Random Forest, SVM, Gradient Descent, LSTM, and Lasso Regression, to predict crop yields in Rajasthan, India. Their dataset spanned from 1997 to 2019, covering crop production data, rainfall, and soil type. The study provided useful insights for farmers to make data-driven decisions on crop selection and yield estimation. However, the focus on a single region (Rajasthan) limited the model's applicability to other parts of India or the world. Furthermore, some models like LSTM required extensive computational resources, which may not be feasible for low-resource settings. [18] tackled crop yield prediction in India using various ML models, such as Decision Tree, Random Forest, XGBoost, CNN, and LSTM. Among these, Random Forest achieved the highest accuracy of 98.96 percent, showing promising results. However, the study's major limitation was its exclusive focus on Indian agriculture, restricting its applicability to other regions. Additionally, the dataset used was limited in scope, reducing the model's generalizability across different farming conditions. [19] simplified crop yield prediction for farmers by using Lasso Regression, Kernel Ridge, and Stacked Regression. Their model aimed to predict yields based on data including state, district, crop, season, and production, with over 250,000 observations. The use of stacked regression achieved an impressive RMSE of less than 1 percent. While the study made significant strides in addressing the prediction problem for specific crops in India, it didn't account for critical factors such as soil quality or weather patterns, which could significantly influence yields. [20] explored the use of machine learning for predicting sunflower and wheat yields using synthetic datasets. Their work compared Random Forest, ANN, and linear models, with Random Forest showing the lowest RMSE (35–38 percent), indicating superior performance. Despite the promising results, real-world accuracy could drop due to unforeseen weather forecast errors. Additionally, the synthetic nature of the dataset may not fully represent the complexities of actual agricultural data, limiting the model's applicability to real-world farming conditions. [21] conducted a systematic literature review of 50 studies that analyzed the use of machine learning for crop yield prediction. Their work summarized the strengths of various ML algorithms, including Random Forest, SVM, and deep learning models, highlighting how they improve prediction accuracy and decision-making in agriculture. However, this review only provides insights based on previously reported results from other papers, without experimental validation. Additionally, the review might have overlooked relevant studies that were not included in the selected databases or in languages other than English.

While machine learning models have shown great potential in improving crop yield predictions, many challenges remain. The studies reviewed indicate that regional and dataset limitations, as well as the inability to account for unpredictable environmental factors like weather events, hinder the applicability of these models on a global scale. Future research must focus on incorporating more diverse datasets, addressing the computational demands, and improving model adaptability to enhance their real-world effectiveness and reliability for farmers across various climates and regions.

2.1.3 Comparative Summary of Reviewed Literature

Table 2.1: Literature Survey on Agricultural Data Analysis and Crop Yield Prediction

S. No.	Author and Title	Key Methods and Findings	Advantages	Limitations / Research Gaps
1	A. Mohammadzadeh et al. (2025), "Analyzing the rainfed wheat yield gap in Northwest Iran"	Integrated qualitative and quantitative factors to analyze wheat yield gap. Found fertilizers and planting density control as key yield determinants.	Highlighted the importance of nitrogen and phosphorus fertilizers and crop rotation.	Limited to northwest Iran; doesn't address climate change impact comprehensively.
2	M. Jiang et al. (2024), "Effects of different fertilization practices on maize yield in northern China"	Meta-analysis of fertilization strategies and water usage impact on maize yield. Found efficient water use and nutrient management as key to improving yield.	Provides a broad perspective on fertilization and water use in maize farming.	Lacks real-world case studies or field trials to validate findings.
3	Z. Yang et al. (2024), "Estimation of Millet Aboveground Biomass Utilizing Multi-Source UAV Image Feature Fusion"	Utilized UAV-derived metrics like canopy temperature and plant height to estimate crop biomass.	Highlights the potential of UAVs and image-based monitoring in yield prediction.	Limited to millet and UAV technology; doesn't explore other crops or sensor types.

S. No.	Author and Title	Key Methods and Findings	Advantages	Limitations / Research Gaps
4	J. S. Gerber et al. (2024), "Global spatially explicit yield gap time trends"	Analyzed global yield gaps over 35 years, focusing on annual crops like maize, rice, and wheat.	Global scope and long-term trends provide valuable insights.	Did not cover newer crops or the role of climate change in influencing yields.
5	R. Landage et al. (2024), "Trends of temperature and precipitation extreme indices in north Maharashtra"	Investigated temperature and precipitation trends over four decades. Found consistent rise in both parameters.	Useful for regional forecasting and early warning systems.	Only focused on one region and didn't explore crop-specific impacts.
6	A. Aucahuasi-Almidon et al. (2024), "Trend analysis and change-point detection of temperature and rainfall in southern Peruvian Amazon"	Explored the relationship between temperature, rainfall, and deforestation in the Amazon.	Valuable for understanding environmental changes in the Amazon region.	Limited applicability to crop yield studies and lacked specific agricultural data.
7	P. Sowmya et al. (2023), "A Comprehensive Dataset for Crop Yield Analysis in Telangana Region"	Compiled district-level agricultural and meteorological data for Telangana, India, focusing on crop yield trends.	Data-rich study that supports regional analysis and forecasting.	Limited to Telangana; does not consider broader regional or global factors.
8	J. Wang et al. (2022), "Climate factors determine the utilization strategy of forest plant resources at large scales"	Analyzed large-scale climate data to study forest plant resource utilization and its environmental implications.	Focuses on large-scale data analysis and climate impact on plant resources.	Focused on forest plants, not food crops; lacked crop-specific findings.

S. No.	Author and Title	Key Methods and Findings	Advantages	Limitations / Research Gaps
9	M. Naveed et al. (2021), "Analyzing the Impact of Climate Change on Cotton Yield Using Spatial Analysis in the Indus River Basin, Pakistan"	Analyzed the impact of rising temperatures and humidity changes on cotton yields.	Important for understanding the impact of climate change on cotton farming.	Focused solely on cotton in one basin; results may not generalize to other crops.
10	R. K. Ranjan et al. (2020), "Effect of Climate Variables on Yield of Major Crops in Samastipur District of Bihar"	Time-series analysis of crop yield data against temperature and rainfall trends.	Emphasized the impact of weather on crop yield and regional analysis.	Lacked a broader scope of agricultural practices or long-term climate change trends.
11	M. Heino et al. (2020), "A multi-model analysis of teleconnected crop yield variability"	Used simulation models to explore global weather patterns and crop yield variability.	Useful for understanding the global connections between weather patterns and food production.	Does not account for local agricultural practices or regional variations in crop yield.
12	M. W. Pramujati (2018), "Spatial-temporal crop yield analysis in East Kalimantan, Indonesia"	Focused on spatial disaggregation of crop yield data and future production estimation.	Useful for spatial-temporal analysis and future yield predictions.	Limited to one specific region; lacks broader global applicability.
13	M. Aryal et al. (2016), "Impact of Climate Variables on Major Food Crops' Yield in Midhills of Western Nepal"	Investigated the effect of climate variables on the yield of food crops in Nepal.	Highlights the diverse impacts of climate on various crops in Nepal.	Limited to Nepal; doesn't consider other global regions or crop types.

S. No.	Author and Title	Key Methods and Findings	Advantages	Limitations / Research Gaps
14	Priyanka Sharma et al. (2023), IEEE Access *Predicting Agriculture Yields Using Regression and DL*	Used Decision Tree, RF, XGBoost, CNN, LSTM on Indian crop data. RF gave 98.96% accuracy.	Compared multiple models, showed value of hybrid ML-DL approach.	Focused only on Indian data; lacks global applicability.
15	Potnuru Sai Nishant et al. (2020), INCET *Crop Yield Prediction using ML in India*	Used Lasso, Kernel Ridge, and Stacked Regression on state-level agri data (2.5L entries).	Achieved low RMSE; incorporated regional details and crop-season info.	Misses key features like weather or soil quality; India-only model.
16	Kavita Jhajharia et al. (2023), ICMDE *ML and DL Techniques in Rajasthan*	Predicted yield using RF, SVM, GD, LSTM with crop, soil, rainfall data (1997–2019).	Data-driven decision support for farmers; model variety explored.	Focus only on Rajasthan; LSTM less feasible in low-resource settings.
17	Thomas van Klompenburg et al. (2020) *ML in Crop Yield: A Systematic Review*	Reviewed 50+ papers; outlined effective ML techniques and challenges.	Comprehensive synthesis of methods and research directions.	Depends on available studies; lacks experimental validation.
18	Akshay Gajula et al. (2021), IEEE *Prediction of Crop and Yield using ML*	Used KNN for crop and fertilizer recommendation using soil/climate inputs.	Simple, efficient model with farmer-friendly output.	Does not account for unexpected climate events or generalization.
19	D. Jayanarayana Reddy et al. (2021), IEEE *ML Algorithm for Crop Yield Prediction*	Used SVM, CNN, DNN with soil/weather/crop features. Accuracy dependent on feature selection.	Adaptive models like ANN worked well with small datasets.	KNN underperformed with non-linear data; model struggles with complexity.

S. No.	Author and Title	Key Methods and Findings	Advantages	Limitations / Research Gaps
20	Alejandro Morales et al. (2023), Frontiers in Plant Science *ML for Past/Future Yield Prediction*	Predicted sunflower and wheat yield using synthetic data and ML (RF, ANN).	RF had lowest RMSE (35–38%); synthetic data enabled robust training.	Risk of overfitting and real-world underperformance due to synthetic inputs.
21	Aruvansh Nigam et al. (2019), IEEE *Crop Yield Prediction using ML Algorithms*	Used 100+ years of weather data and 17 years of crop data; applied RNN, LSTM, RF, XGBoost.	Long-term data enabled strong forecasting; supports farmer decision-making.	Geographically limited to Indian districts; inconsistent results across models.

2.2 EXISTING SYSTEM

Here are some of the commonly used methods and systems:

Climate FieldView: Climate FieldView integrates data from weather, soil, and satellite imagery to provide farmers with real-time insights for improved crop management. It offers yield predictions, field variability analysis, and weather monitoring to optimize farming practices. By identifying areas in need of attention, it helps farmers enhance productivity and make data-driven decisions to increase yields.

Farmprogress - IBM Watson Decision Platform for Agriculture: IBM Watson Decision Platform uses AI and machine learning to support agricultural decision-making. It offers crop yield predictions, soil health analysis, and weather forecasts to optimize farming practices. The platform provides tailored recommendations to improve efficiency and productivity, integrating IoT data for real-time insights on field conditions.

Trimble Ag Software: Trimble Ag Software is a comprehensive farm management system that combines data from various sources. It includes tools for yield prediction, soil and weather analysis, and field mapping. The platform enhances productivity through precision agriculture, helping farmers manage resources efficiently and improve crop output.

Microsoft Azure FarmBeats: Azure FarmBeats is a cloud-based platform that leverages AI and IoT for smarter agricultural practices. It offers predictive tools for crop yields and analyzes environmental factors like weather patterns and soil health. By integrating data from sensors and satellite imagery, the platform enables efficient resource management and supports precision farming.

Chapter 3
METHODOLOGY

3.1 DATA & METHODOLOGY

The methodology adopted in this project is visually summarized in Figure 3.1. It outlines each critical stage from raw data acquisition to final yield forecasting. The pipeline integrates database operations, geospatial mapping, statistical analysis, and model evaluation. Each step in this workflow is elaborated below.

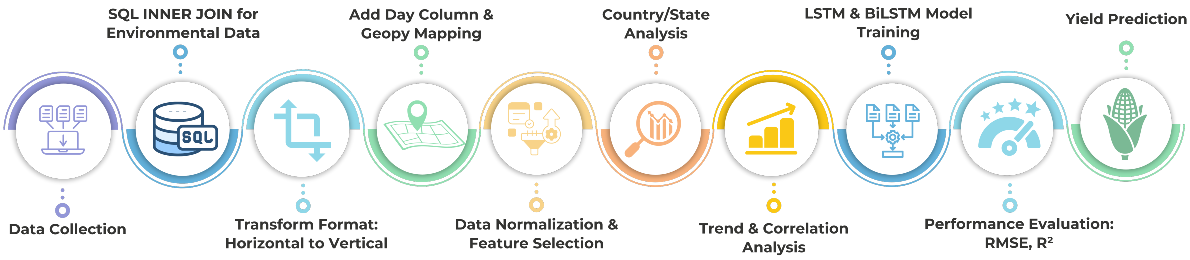


Figure 3.1: Workflow Overview

3.1.1 Data Collection

Agricultural productivity is significantly influenced by climate variability and soil characteristics. Understanding these factors is essential for optimizing crop management strategies and ensuring global food security. This study utilizes a comprehensive dataset to analyze long-term trends in maize yield, integrating historical climate data and key agronomic variables.

The dataset was sourced from the FutureCrop Challenge Kaggle competition launched in June 2024 by Groenke et al., which aimed to forecast global gridded maize and wheat yields using high-resolution environmental and agronomic data under a high-emissions global warming scenario. It includes 39 years of historical data (1982–2020) for various geographic grid points, incorporating meteorological and soil-based parameters.

Each row in the dataset corresponds to a unique combination of latitude, longitude, year, and crop type. Climatic variables such as precipitation, shortwave radiation (rds), and daily mean, maximum, and minimum temperatures are stored in separate files, each containing 240 daily observations spanning from 30 days before sowing to 210 days after sowing. Static agronomic attributes such as soil texture, nitrogen application rate, atmospheric CO_2 concentration, and planting date are included as supplementary inputs. The dataset is stored in .parquet format to support efficient data handling and scalability. Each entry forms a multivariate time series with associated descriptors, and the target variable—crop yield—is recorded in tons per hectare per year for each grid cell. With 3,394 unique spatial locations and 40 years of data, the dataset includes over 30.8 million rows.Table 3.1 summarizes the variables and their measurement units.

Table 3.1: Variables in the dataset and their measurement units

Variable Type	**Variable Name**	**Measurement Units**
Climatic Variables	Mean Temperature (tas)	°C
	Maximum Temperature (tasmax)	°C
	Minimum Temperature (tasmin)	°C
	Shortwave Radiation (rds)	W/m²
	Precipitation (pr)	kg/m²/s
Agronomic Variables	Soil Texture	Categorical
	CO_2 Concentration	ppm
	Nitrogen Fertilization Rate	tons/ha
	Crop Yield (Target Variable)	tons/ha

3.1.2 SQL Integration for Environmental Data

The collected environmental and agronomic data were initially distributed across multiple separate files—each representing a specific variable such as temperature, precipitation, or radiation. In this raw format, the data was fragmented and difficult to manage or analyze cohesively.Therefore,it was essential to bring all relevant data points together into a single, structured dataset.

To achieve this, the files were imported into structured tables using Structured Query Language (SQL). Each table followed a common schema including geographic coordinates (latitude, longitude), temporal information (year), and crop identifiers.

Using INNER JOIN operations, the tables were merged on the following shared attributes:

- **Latitude**
- **Longitude**

- **Year**
- **Crop type**

This join operation ensured that only complete records—those with matching entries across all variables—were included in the final dataset. The result was a unified, analysis-ready table that combined climatic and agronomic parameters for each unique spatial-temporal instance.
Additional SQL operations were used to:

- Handle missing values by filtering out incomplete records.
- Rename and standardize column names for consistency.
- Create indexes to enhance performance during querying and preprocessing.

This SQL integration step was a critical part of the preprocessing pipeline, enabling efficient access to multivariate data and forming a solid foundation for time-series modeling and geospatial analysis.

3.1.3 Transforming Data Format: Horizontal to Vertical

Initially, the environmental variables in the dataset were organized in a horizontal format, where each longitude-latitude pair was associated with 240 columns representing daily values across the maize growing season. While comprehensive, this structure was not optimal for time-series analysis or model training.
To address this, a transformation process was undertaken to convert the dataset from a wide (horizontal) format into a long (vertical) format:

- A new Day column was introduced, where each row corresponds to a specific day relative to the sowing date.
- Instead of multiple daily columns for each location, daily observations were stacked vertically, making each row represent one day of environmental measurements for a specific grid location and year.
- This restructuring normalized the data into a multivariate time series format, which is highly compatible with machine learning models like LSTM and BiLSTM .

The transformation process was performed using advanced SQL queries combined with Python libraries like Pandas, ensuring seamless restructuring and minimal data loss. This verticalization of data was critical for enabling sequential modeling and efficient analysis of temporal patterns in maize yield dynamics.

3.1.4 Adding Day Column and Geopy Mapping

Following the data restructuring, two important enhancements were introduced to enrich the dataset for spatiotemporal analysis:

- **Day Column Addition:** A Day attribute was added to each record, indexing days relative to the crop sowing date. This temporal labeling allowed the dataset to capture sequential daily variations in environmental conditions, crucial for time-series model training.
- **Geopy-Based Geographic Mapping:** Using the Geopy Python library, each spatial grid point (longitude and latitude) was reverse-geocoded to its corresponding Country and State. This step provided regional identifiers for each data point, enabling location-specific analysis such as:
  - Country-wise yield trends.
  - State-level comparisons within India.

The enrichment of the dataset with temporal (Day) and spatial (Country, State) labels improved both the granularity and the interpretability of subsequent data analysis and visualization processes. It also laid the groundwork for geographically segmented modeling and targeted agronomic insights.

3.1.5 Data Normalization and Feature Selection

To prepare the dataset for machine learning model training, it was essential to preprocess the variables to ensure uniformity and reduce noise. This step involved two key processes:

- **Data Normalization:** Environmental and agronomic variables were normalized using techniques such as Min-Max Scaling and Z-score Standardization to bring all features onto a common numerical scale. Normalization helps to:
  - Eliminate biases caused by differing units and magnitudes.
  - Improve convergence rates and accuracy in model training.

These preprocessing steps ensured that the dataset was clean, consistent, and well-structured for effective training of deep learning models.

3.1.6 Country and State Analysis

After preprocessing, the dataset was enriched with geographic identifiers (Country, State) to enable regional analysis of maize yields. Using Python libraries like Pandas, Matplotlib, and Seaborn, graphs were plotted to:

- Calculate and visualize average maize yield across different countries and states.
- Understand spatial yield distributions by mapping yields geographically.
- Identify regional patterns, anomalies, and high/low productivity zones.

This geographic analysis offered crucial insights into how yield outcomes vary across different climatic and agronomic zones, aiding in targeted modeling and agronomic recommendations.

3.1.7 Trend and Correlation Analysis

To understand historical dynamics, a thorough trend and correlation analysis was conducted:

- **Trend Analysis:** Line charts were plotted to illustrate how maize yields have changed year by year from 1982 to 2020, highlighting periods of growth, stagnation, or decline.
- **Correlation Analysis:** Pearson correlation coefficients were calculated between crop yields and various environmental variables such as precipitation, temperature, radiation, nitrogen application, and CO_2 levels.

These analyses helped uncover key drivers of maize yield variability over time and provided a scientific basis for selecting influential features for modeling.

3.1.8 LSTM and BiLSTM Model Training

Sequential deep learning models were deployed to predict maize yields:

- LSTM and BiLSTM models were trained to capture temporal dependencies in the multivariate time series data.
- Hyperparameter tuning was performed, including optimization of parameters like learning rate, number of layers, number of units, dropout rate, and batch size, to maximize model accuracy.

The sequential nature of these models made them highly suitable for modeling daily environmental impacts on maize yield across multiple growing seasons.

3.1.9 Performance Evaluation: RMSE, R^2, MAE

After training, model performance was rigorously evaluated using multiple metrics:

- **Root Mean Squared Error (RMSE)** to measure prediction error magnitude.
- **Mean Absolute Error (MAE)** to assess average prediction deviation.
- **R^2 Score (Coefficient of Determination)** to evaluate how well the model captures variance in yield outcomes.

Performance of the LSTM and BiLSTM models was compared to identify the superior model, balancing both predictive accuracy and generalization capability.

3.1.10 Yield Prediction

Finally, the best-performing trained models were used to forecast future maize yields:

- The dataset already included simulated environmental conditions for the years 2021 to 2046, generated under various climate change scenarios. These pre-simulated values were used directly as inputs to the trained models.
- The models predicted maize yields at the spatial grid level for each future year, enabling a forward-looking analysis of agricultural productivity trends under evolving climate conditions.

These yield projections offer valuable foresight for policymakers, agronomists, and researchers aiming to adapt maize production strategies in response to climate change.

3.2 EXPLORATORY DATA ANALYSIS

3.2.1 Country-wise Analysis

Average Maize Yield by Country:

This section presents a comparative analysis of maize yield across selected countries that significantly contribute to global maize production. The analysis covers six key nations, each representing a distinct continent: the United States (North America), Brazil and Argentina (South America), India (Asia), Ukraine (Europe), and South Africa (Africa). These countries were selected due to their substantial maize output and varied agro-climatic conditions, offering insight into the impact of regional factors on crop productivity.

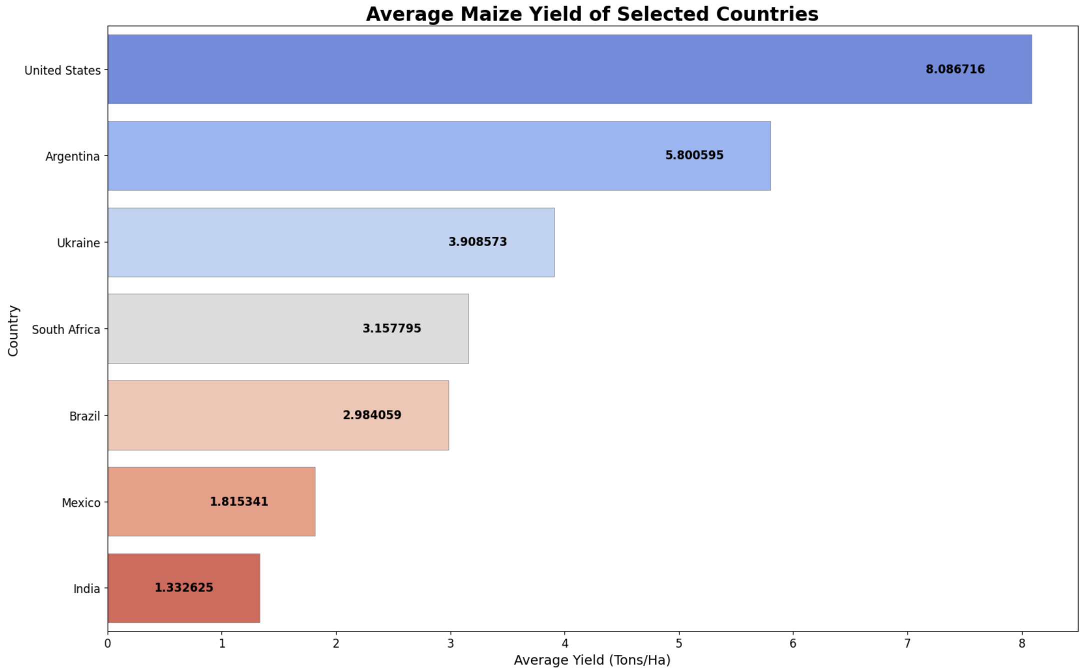


Figure 3.2: Average maize yield of selected countries

As depicted in Figure 3.2, maize yield varies considerably among the selected countries. The United States consistently records the highest yields, attributed to advanced mechanization, precision farming, and the widespread use of genetically improved seeds. Brazil also demonstrates high productivity, primarily due to double-cropping systems and favorable rainfall patterns, despite facing concerns such as soil degradation.
Argentina benefits from fertile soil and sustainable practices like no-till farming, which aid in moisture conservation. In contrast, India experiences moderate yields owing to its dependency on monsoons and prevalence of small-scale farms, which limit the implementation of modern machinery. Ukraine, although rich in fertile soil, suffers from weather extremes and seasonal instability that hinder productivity. South Africa's maize yield is notably impacted by recurring droughts and climate variability.
The analysis highlights that higher yields are typically achieved in regions where technological innovation and climate conditions are favorable. Conversely, countries with lower productivity often face limitations such as water scarcity, traditional farming methods, and inadequate infrastructure. Bridging this yield gap will require targeted interventions, including improved irrigation systems, access to agricultural technology, and adoption of climate-resilient practices.

Analysis of Maize Yield by Country (1982–2020):

Yearly trends in maize yield by country are crucial for understanding the dynamic nature of maize production and identifying key patterns or anomalies. As depicted in Figure 3.3, the trend lines illustrate maize yield progression for each country over the years.

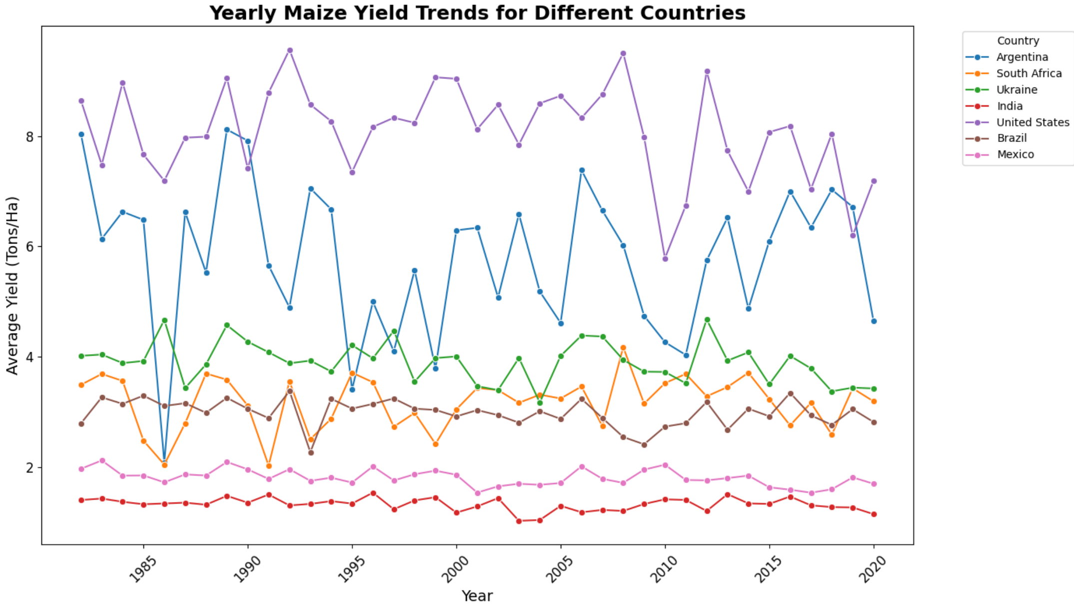


Figure 3.3: Yearly trend of maize yield by country (1982–2020).

Analysis of maize yield data from 1982 to 2020 reveals significant variations across major producing nations, with yields ranging from 1 to 9.5 tons per hectare. These differences]lect distinct technological, institutional, and environmental factors shaping each country's agricultural system.

The United States consistently demonstrates superior yields of 7–9 tons/ha, attributed to advanced agricultural infrastructure, comprehensive research programs, and optimal utilization of the Corn Belt region. In contrast, Argentina shows high but volatile yields, peaking at 9.5 tons/ha, largely due to economic instability affecting agricultural investments despite favorable conditions in the Pampas region.

Ukraine maintains moderate yields around 4 tons/ha, indicative of a gradual transition from Soviet-era farming practices and the presence of fertile black soils, although limited by irrigation infrastructure. Brazil shows remarkable improvement—from 1 to over 3 tons/ha—driven by successful transformation of the Cerrado region and widespread adoption of technology through initiatives led by national agricultural research institutions. Mexico and India, meanwhile, maintain yields below 2 tons/ha, constrained by small landholdings, traditional farming techniques, and insufficient irrigation facilities.

These findings highlight that sustaining high maize yields requires an integrated approach encompassing technological innovation, institutional support, and consistent agricultural policy. The U.S. model exemplifies how favorable environmental conditions, coupled with strong policy and research backing, can maximize productivity. Conversely, other nations' experiences underscore the importance of addressing structural challenges to unlock their yield potential. This analysis offers valuable insights for policy-makers in emerging economies aiming to boost agricultural output through strategic investments in farming practices, research, and infrastructure.

3.2.2 *State-wise Analysis*

Comparative Yield Analysis Across Indian States:

Maize productivity across India's top seven producing states reveals significant variations. These differences select the diverse agro-ecological conditions, farming practices, and technological advancements unique to each state. As depicted in Figure 3.4, the top maize-producing states show differing productivity trends influenced by several factors including irrigation, crop management, and climatic conditions.

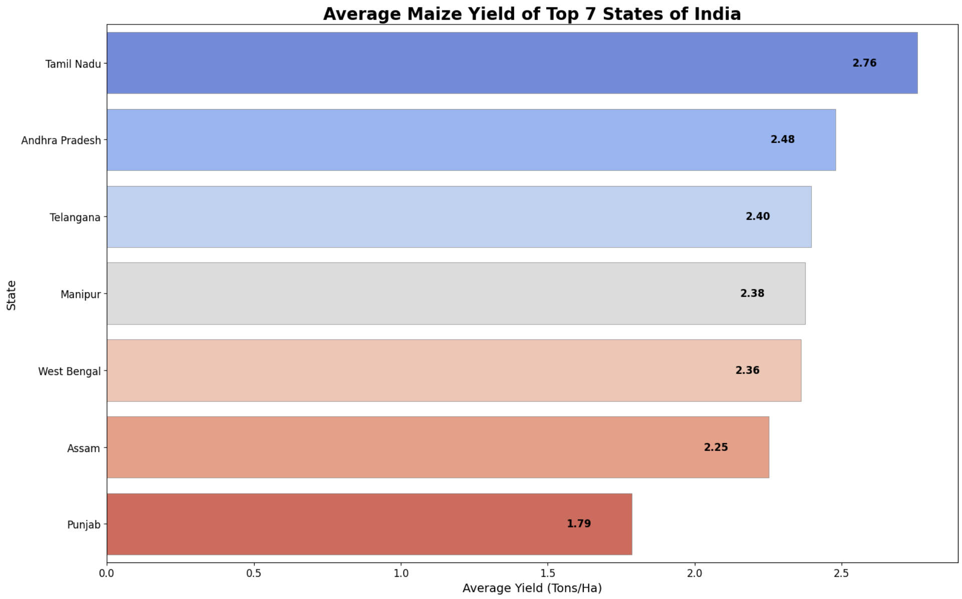


Figure 3.4: Top states with average maize yield in India.

Analysis of maize productivity across India's top seven producing states reveals the following trends:

Tamil Nadu: Leads with the highest maize yield of approximately 2.76 tons/ha. The state benefits from established irrigation systems, the use of high-yielding varieties (HYVs), and robust agricultural extension services, ensuring optimal crop performance even in varying climatic conditions.

Andhra Pradesh: With a yield of 2.48 tons/ha, owes its success to its coastal climate and progressive adoption of modern agricultural techniques, including the cultivation of hybrids and efficient crop management strategies.

Telangana: Records a yield of 2.40 tons/ha, benefiting from targeted hybrid maize promotion and improved irrigation infrastructure. Region-specific policies have helped sustain strong productivity levels despite the smaller area under cultivation.

West Bengal: With a yield of 2.36 tons/ha, benefits from fertile alluvial soils and a well-distributed irrigation network, although seasonal flooding can occasionally impact yields. Proactive water management and the use of short-duration varieties help maintain stable productivity.

Manipur: Shows a yield of 2.38 tons/ha, thanks to the integration of traditional farming knowledge and modern practices, such as improved seed varieties and judicious fertilizer use.

Assam: With 2.25 tons/ha, faces challenges like erratic monsoons and flooding. However, strategic adoption of climate-resilient maize cultivars and improved water practices have sustained yields.

Punjab: Known for its dominance in rice-wheat systems, has the lowest yield among the top maize-producing states at 1.79 tons/ha. This lower yield is attributed to limited maize acreage, lack of focus on crop diversification, and an over-reliance on traditional cereal cropping systems.

This analysis highlights the influence of soil quality, climatic conditions, and technological interventions in maize productivity. States like Tamil Nadu and Andhra Pradesh demonstrate how infrastructure, research, and modern techniques can significantly enhance yields. On the other hand, underperforming states, such as Meghalaya, Arunachal Pradesh, Chhattisgarh, Jharkhand, Sikkim, Gujarat, and Uttarakhand, face constraints like inadequate soil quality, insufficient irrigation, and limited access to advanced farming technologies. Addressing these challenges through targeted policies and the adoption of improved farming practices can lead to higher productivity across these regions.

Yearly Trend of Maize Yield in India:

The analysis of yearly maize yield trends from 1980 to 2020 reveals distinct fluctuations in productivity across various Indian states, showcasing how climatic variability, technological advancements, and policy interventions have shaped production patterns over time. As depicted in Figure 3.5, the trend lines for selected states provide valuable insights into the dynamics of maize yield across the country.

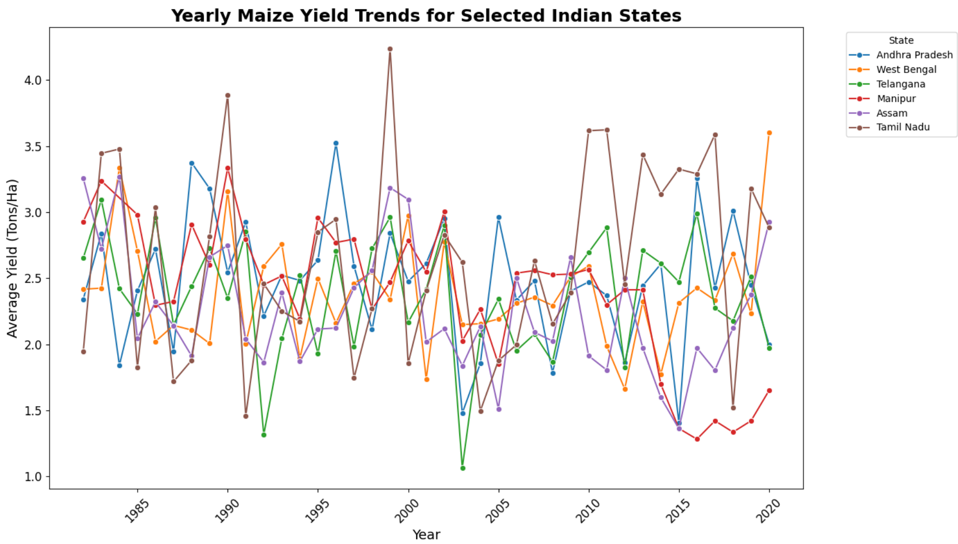


Figure 3.5: Yearly trend of maize yield in selected Indian states from 1980 to 2020.

Tamil Nadu: Consistently exhibits the highest and most stable maize yields, frequently exceeding 3 tons/ha, with peaks beyond 4 tons/ha during certain years. This performance]lects the state's strong irrigation infrastructure, widespread use of hybrid maize, and institutional support for precision farming and extension services. Tamil Nadu's investment in sustainable intensification has made it a leader in maize productivity.
Andhra Pradesh: Shows robust yield performance with occasional spikes indicating productive seasons supported by coastal climatic advantages and high-yielding seed varieties. However, intermittent dips are observed, likely due to monsoon irregularities and cyclone events. Despite this, overall performance remains strong, backed by effective agricultural policies and farmer adoption of improved practices.
Telangana: Exhibits considerable yield volatility between the 1980s and early 2000s, a period marked by rainfed agriculture and low hybrid seed usage. Post-2010, however, the state shows a clear improvement in yield, attributed to government-led programs encouraging hybrid maize, irrigation expansion, and precision farming—a transformation often referred to as the "Maize Revolution" in the region.
Manipur: Despite its relatively small geographical area, shows a stable yield trend with values oscillating mostly between 2 to 3 tons/ha. The consistency is attributed to the successful integration of traditional knowledge with modern interventions, including improved varieties and better fertilizer use in hill farming.
West Bengal: Displays moderate yields with fluctuating trends,]lective of its dual agro-climatic reality—the fertile alluvial zones of the Gangetic plains and the flood-prone delta regions. Yield stability is observed, though occasional drops occur during years of excessive rainfall or waterlogging. Investments in short-duration varieties and irrigation access have helped mitigate extreme yield collapses.
Assam: Similar to West Bengal, Assam's yields experience frequent dips, particularly in the 1990s and early 2000s, selecting its vulnerability to monsoonal flooding and limited mechanization. However, recovery is visible in the last decade due to government support for flood-resilient maize varieties and the promotion of short-season hybrids.
This analysis of the yearly trends underscores the interplay between climatic conditions, technological interventions, and policy measures in shaping maize productivity. States like Tamil Nadu and Andhra Pradesh have demonstrated how technological adoption, effective policies, and favorable climate can lead to consistent high yields, while other states like Assam and West Bengal highlight the challenges faced due to climatic risks such as flooding and erratic monsoons. Addressing these challenges through improved technologies and better infrastructure will be crucial for enhancing maize productivity across the country.

Geographical Distribution Analysis:

The Figure 3.6, illustrates the average maize yield across Indian states for the historical period from 1982 to 2020, measured in tons per hectare. The color scale transitions from yellow to dark green, with lighter shades indicating lower yields and darker shades representing higher yields. Overall, the map shows that states like Punjab, Haryana, and Karnataka maintained consistently higher maize yields during this period, benefiting from favorable agricultural policies, strong irrigation networks, and adoption of improved farming techniques. Conversely, regions such as Ladakh, Mizoram, and parts of the north-eastern states exhibited lower yield levels, likely due to less favorable climatic conditions, mountainous terrains, and limited agricultural infrastructure. This visualization provides an important baseline for understanding regional maize productivity trends in the past decades, setting the context for future projections and strategic planning in India's agricultural sector. A longitude-latitude reference system is included for spatial orientation, along with a color bar for precise interpretation of yield variations.

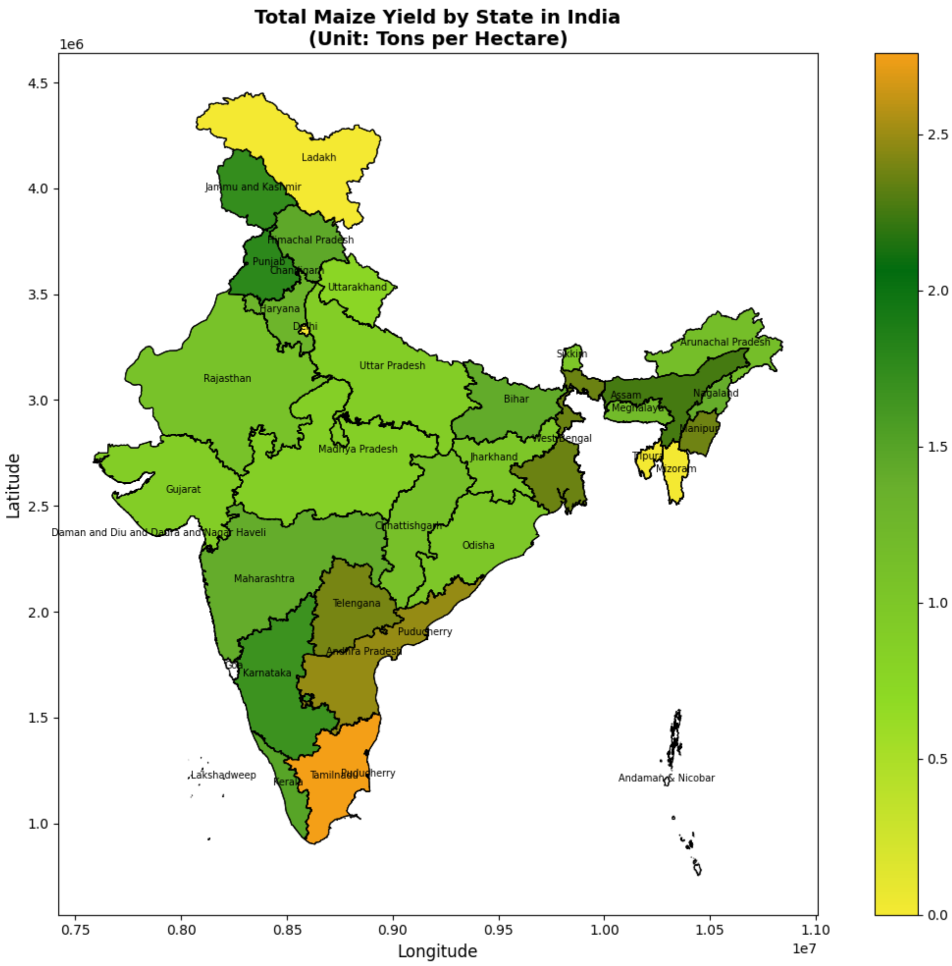


Figure 3.6: Geographical Distribution Analysis

3.2.3 *Correlation Analysis and Regression Analysis*

Precipitation impact on maize growth stages:

Precipitation plays a pivotal role in the growth and development of maize, although its impact varies across different phenological stages. In this analysis, daily precipitation data were aggregated into four key maize growth phases—Sowing (Days 1–30), Vegetative (Days 31–100), Reproductive (Days 101–170), and Grain-Filling (Days 171–220)—to assess its stage-wise influence on final maize yield. The correlation between precipitation during each growth stage and maize yield is summarized in Table 3.2.

The results reveal a negative correlation (-0.089) during the sowing phase, indicating that excessive rainfall during this stage may hinder germination, promote seed rot, and cause topsoil erosion, ultimately reducing yield potential. In the vegetative stage, the correlation is near zero (-0.016), suggesting that maize can tolerate moderate rainfall fluctuations during this phase without significant yield impact.

The reproductive stage shows a moderate positive correlation (0.230), underlining the importance of adequate rainfall during pollination and early grain formation. Water availability during this phase is crucial for tasseling and silking synchronization, directly influencing kernel set and grain count. Lastly, the grain-filling stage exhibits a weak positive correlation (0.042), indicating that some moisture is beneficial to support grain development. However, excessive rainfall may delay physiological maturity and slightly compromise grain quality.

Understanding these dynamics provides insights for irrigation scheduling and adopting adaptive measures to mitigate rainfall-related risks during sensitive phases of maize development.

Table 3.2: Correlation between growth stages of maize and precipitation

Growth Stage	**Correlation**	**Interpretation**
Sowing (1-30)	-0.089	Rain may harm germination
Vegetative (31-100)	-0.016	No impact
Reproductive (101-170)	0.23	Rain supports pollination
Grain-Filling (171-220)	0.042	Minor yield effect

Correlation of Climatic and Agronomic Factors with Maize Yield in India (1982–2020):

This section illustrates the relationship between maize yield (tons/ha) and various environmental-agronomic factors, derived from Google Grid annual data for India (1982–2020). The independent variables analyzed include precipitation, temperature metrics, shortwave radiation, CO_2 concentration, and nitrogen fertilization rate. Table 3.3 presents yield equations and the correlation coefficients of these variables with yield. The dataset is aggregated using mean values for each year, ensuring a robust representation of climate and agronomic trends over time.

Correlation analysis, summarized in Table 3.3, reveals distinct trends. Precipitation, with a correlation coefficient (r) of 0.55, is positively correlated with yield, suggesting that higher rainfall generally benefits maize production by ensuring sufficient soil moisture. However, excessive rainfall beyond an optimal threshold may still cause issues like waterlogging, nutrient leaching, or soil erosion.

Temperature variables show negative correlations: r = –0.58 for mean temperature and r = –0.53 for minimum temperature, which exhibit stronger effects than maximum temperature (r = –0.45). This suggests that higher temperatures, particularly warmer nights, reduce grain-filling efficiency and overall yield due to increased respiration and heat stress. Shortwave radiation has a weak positive correlation (r = 0.13), indicating that increased solar radiation may slightly enhance photosynthesis, though factors like cloud cover or atmospheric pollutants may limit this effect. CO_2 concentration (r = –0.27) exhibits a weak negative correlation, implying that any potential fertilization benefits are likely outweighed by other limiting climatic or soil factors in real-world field conditions.

Nitrogen fertilization rate shows a weak positive correlation (r = 0.11) with yield, reinforcing its importance as a key input for sustaining maize productivity. This suggests that improvements in fertilizer application have helped mitigate yield losses due to climatic stressors.

Table 3.3: Correlation analysis of maize yield and agro-meteorological variables

Variable	**Regression Equation**	**Correlation (r)**
Precipitation	Yield = 4922.5584 * pr + 1.0697	0.5518
Mean Temperature	Yield = -0.1831 * tas + 5.8520	-0.5778
Minimum Temperature	Yield = -0.1690 * tasmin + 4.6692	-0.5318
Maximum Temperature	Yield = -0.1077 * tasmax + 4.6042	-0.4494
Shortwave Radiation	Yield = 0.0021 * rds + 0.9301	0.1332
CO_2 Concentration	Yield = -0.0015 * co_2 + 1.8782	-0.2688
Nitrogen Fertilization Rate	Yield = 0.1130 * nitrogen + - 3.2008	0.1161

In Figure 3.7, presents scatter plots illustrating the relationship between maize yield (tons/ha) and various environmental and agronomic factors, derived from statewise annual data (1982–2020).

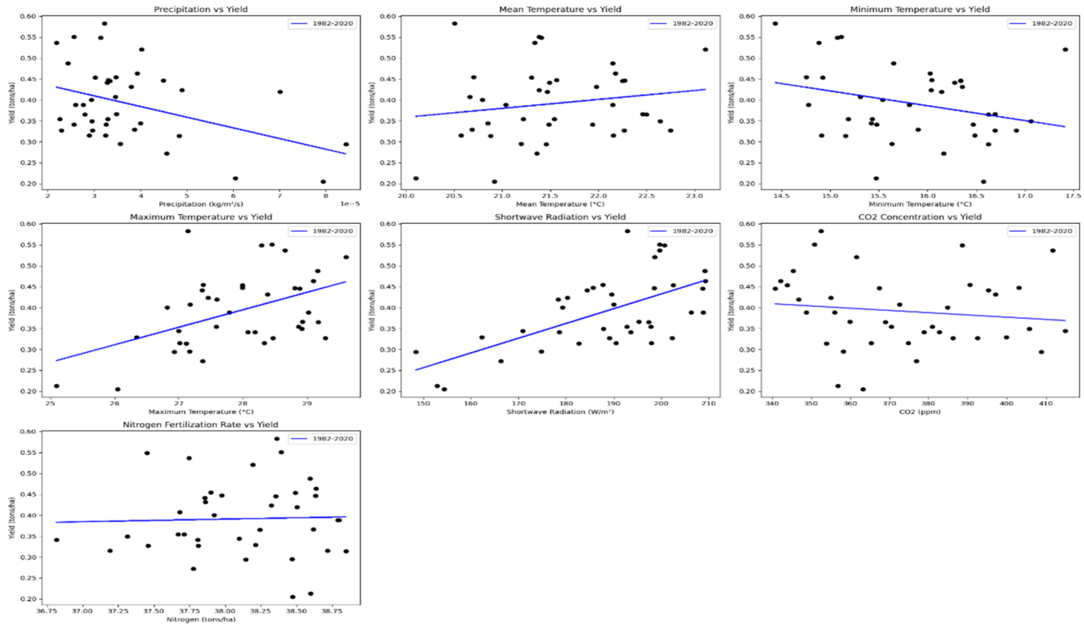


Figure 3.7: Correlation between maize yield and environment variables

Maize Yield and Variables Trends Yearly (1982–2020):

Figure 3.8 presents a series of scatter plots with regression lines, visualizing trends in maize yield and key climatic variables across selected Indian states. The plots illustrate variations in precipitation, temperature (mean, minimum, maximum), shortwave radiation, CO_2 concentration, nitrogen fertilization rate, and yield over the years. Each subplot contains black dots representing yearly data points, overlaid with regression lines for two time periods: 1982–2020 (blue line) and 2010–2020 (red line). The following key insights emerge from the figure: Precipitation shows a decreasing trend, particularly in the recent decade (2010–2020), indicating potential shifts in monsoon patterns affecting maize production. Temperature metrics (Mean, Minimum, Maximum) depict slight warming trends, with minimum temperatures showing a more noticeable increase, suggesting warmer nights which may affect crop growth. Shortwave radiation exhibits a declining trend, implying possible cloud cover increases or atmospheric changes reducing solar energy available for photosynthesis. CO_2 concentration follows a strong upward trajectory, aligning with global greenhouse gas emissions trends, which could influence maize yield through CO_2 fertilization effects. Nitrogen fertilization rates appear relatively stable, though slight increases are observed post-2010, possibly due to enhanced agricultural interventions. Yield trends remain stagnant over the long term (1982–2020), but the recent decade (2010–2020) shows a sharp increase. This suggests that recent technological advancements, improved hybrid varieties, or changes in agronomic practices might have contributed to enhanced productivity. These visualizations effectively highlight the climatic and agronomic influences on maize yield over time. The inclusion of two regression periods (1982–2020 and 2010–2020) provides a comparative analysis, helping to isolate recent climatic impacts from long-term historical trends. The state-wise granularity ensures relevance for regional policy-making and adaptation strategies in Indian agriculture.

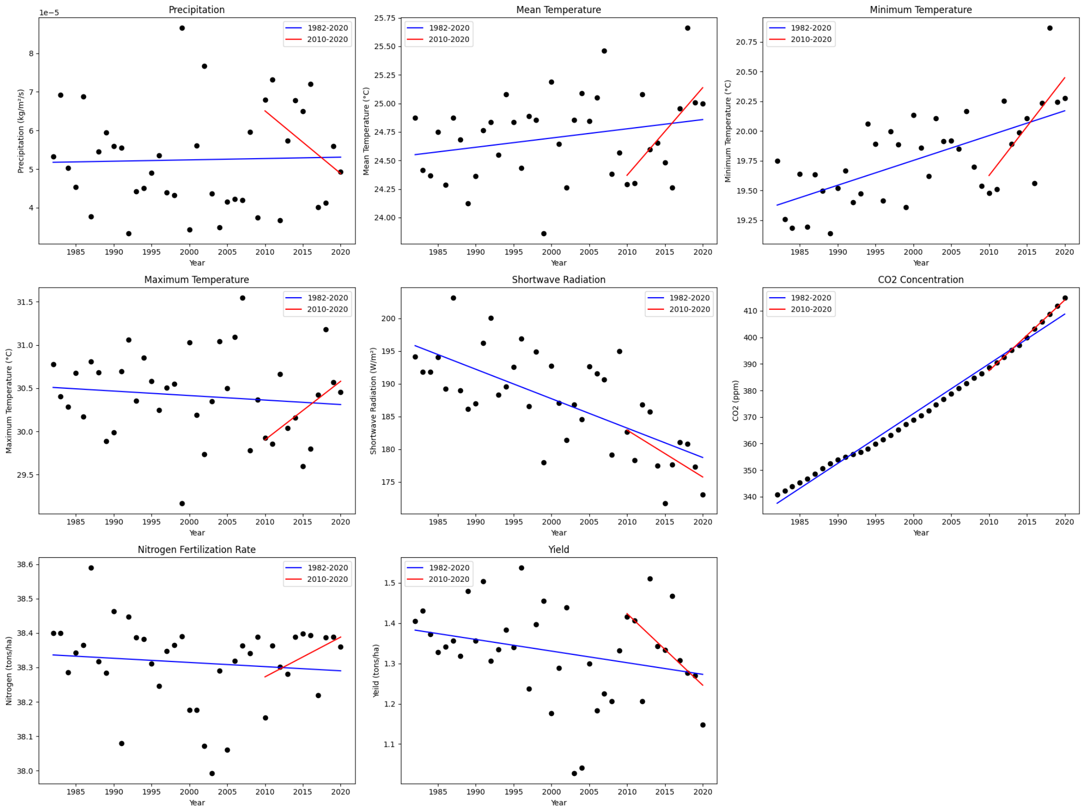


Figure 3.8: Trends in yield and environment Variables

3.3 DL TECHNIQUES

3.3.1 Long Short-Term Memory (LSTM)

LSTM networks are a special type of Recurrent Neural Network (RNN) designed to learn from and make predictions based on sequential data, especially when important patterns stretch far across time. Unlike regular RNNs, which often forget older information, LSTM networks are built to remember valuable details for long periods, thanks to their clever internal design involving structures called gates. These gates act like decision-makers: they decide what information to keep, what to update, and what to forget at each step. This makes LSTM extremely powerful for tasks like crop yield prediction, where something that happened years ago, like a drought or change in farming practices, can still impact future harvests. By ca]ully managing memory over time, LSTM models are able to capture complex, long-term relationships in agricultural data, leading to smarter and more accurate forecasts.

Working of LSTM

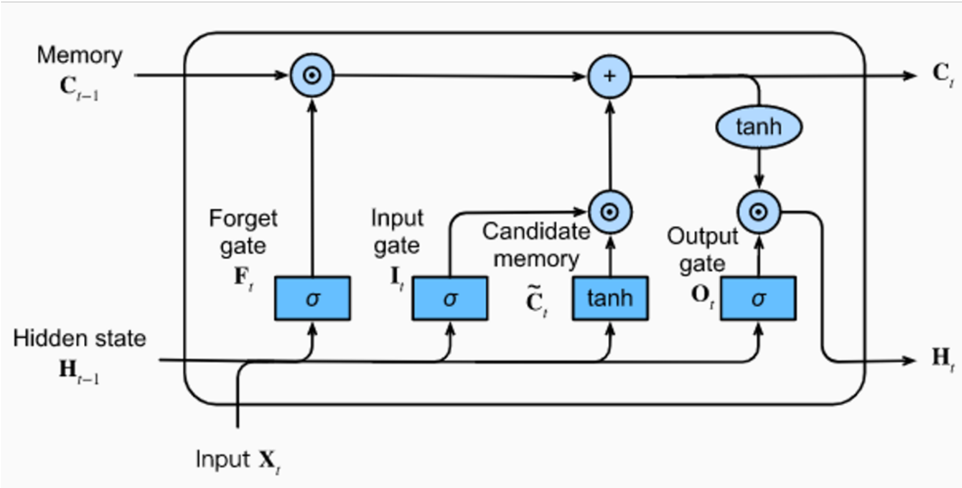


Figure 3.9: Working diagram of an LSTM cell

- **Step 1: Input**
  LSTM receives the previous hidden state, previous memory, and current input.
- **Step 2: Forget Gate**
  Decides which parts of old memory to erase.
- **Step 3: Input Gate and Candidate Memory**
  Picks what new information to add and generates possible new memory.
- **Step 4: Update Cell State**
  Combines the kept old memory and important new information to update memory.
- **Step 5: Output Gate**
  Decides what part of the updated memory to output for the next step.
- **Step 6: Hidden State Update**
  Passes the new hidden state forward, carrying important information.

3.3.2 Bidirectional Long Short-Term Memory (BiLSTM)

BiLSTM networks are an enhanced version of traditional LSTM models, designed to learn from sequential data not just by looking at the past but also by considering the future. In many real-world situations, like predicting crop yields, what happens ahead can be just as important as what happened before. BiLSTM addresses this by running two LSTM layers side-by-side: one processes the data from past to future, and the other from future to past. This two-way learning allows the network to gather a richer, more complete understanding of patterns over time. For example, while past weather conditions impact a harvest, upcoming events like rainfall forecasts can also play a major role. By combining information from both directions, BiLSTM models are able to make smarter, more accurate predictions in complex agricultural environments.

Working Steps

- **Step 1: Input Layer**
  Each input at a time step (e.g., x_{t-1}, x_t, x_{t+1}) is passed into two LSTM cells: one moving forward and one moving backward.

- **Step 2: Architecture Overview**
  The structure of a Bidirectional LSTM is shown below:

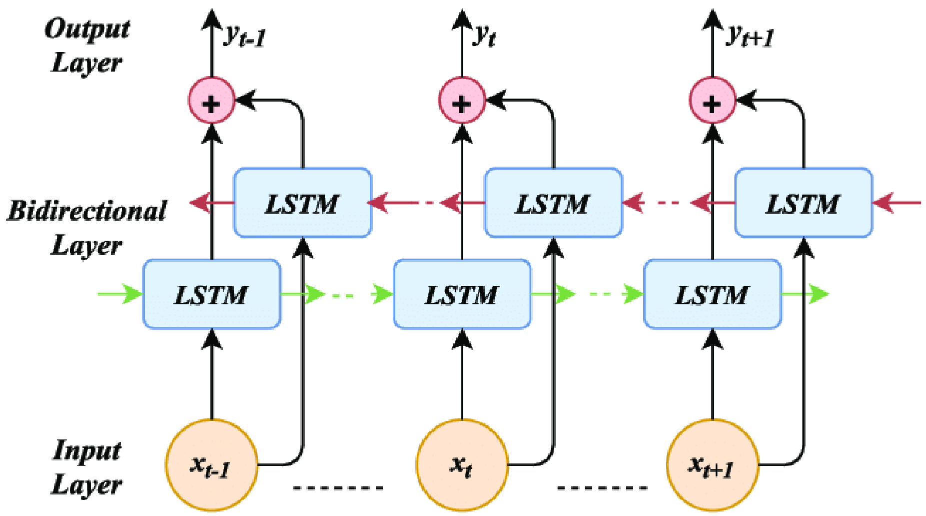


Figure 3.10: Architecture of a Bidirectional LSTM (BiLSTM)

- **Step 3: Forward and Backward LSTM**
  - The **forward LSTM** processes the sequence in normal time order (left to right).
  - The **backward LSTM** processes the sequence in reverse (right to left).

- **Step 4: Bidirectional Layer**
  The outputs from both forward and backward LSTM layers are **combined** at each time step, giving the model a richer understanding.

- **Step 5: Output Layer**
  The combined information is passed to the output layer to make a final prediction at each step (like predicting crop yield for a specific year).

- **Final Flow**
  By using information from both past and future, BiLSTM gives a **more complete view**, leading to better predictions compared to regular LSTM.

3.3.3 Hyperparameter Tuning Strategy

To optimize the performance of the BiLSTM model for crop yield prediction, extensive hyperparameter tuning was conducted based on the characteristics of the maize dataset. The sequence length (seq_len) was set to 240, aligning with the typical duration of a crop season. The model architecture included two stacked Bidirectional LSTM layers, each with 128 hidden units and a dropout rate of 0.2 to prevent overfitting. A learning rate of 0.001 was selected, along with a decay rate of 1×10^{-5}, to ensure stable convergence during training. The batch size was set at 64, balancing between training speed and model stability. Additionally, an early stopping mechanism with a patience of 50 epochs and a model checkpoint system were employed to halt training once the validation loss stopped improving, thereby avoiding unnecessary overfitting.

Further tuning involved adding dense layers after the LSTM blocks to enhance the model's learning capacity. A two-layer dense structure was introduced, where the first dense layer had twice the number of LSTM hidden units (256 neurons) and the second dense layer had 128 neurons, both using ReLU activation to capture complex feature interactions. The Nadam optimizer was chosen for its ability to adapt learning rates effectively during training. A learning rate reduction callback (ReduceLROnPlateau) was also incorporated to automatically lower the learning rate if the model's performance plateaued, helping the model escape shallow minima. Together, these hyperparameter choices were ca]ully validated through experimentation to maximize accuracy and minimize loss, leading to a more robust and reliable prediction model tailored to the seasonal patterns present in the data.

Chapter 4

HARDWARE & SOFTWARE REQUIRMENTS

4.1 HARDWARE REQUIREMENTS

- Processor: Intel i5/i7 or AMD Ryzen 5/7 (or higher)
- Memory Space: 500 GB+ (SSD p]erred for faster data processing)
- RAM: Minimum 8 GB (Recommended: 16 GB or more)
- GPU: NVIDIA GPU (e.g., GTX 1650 or higher) for deep learning acceleration

4.2 SOFTWARE REQUIREMENTS

- Python 3.7+ : Core programming language
- Streamlit : GUI for interactive input/output and visualization
- TensorFlow / Keras : Building and training LSTM and BiLSTM models
- Pandas : Data loading, cleaning, and preprocessing
- NumPy : Numerical operations and matrix handling
- Matplotlib : Basic data visualization
- Seaborn : Heatmaps and correlation analysis
- Plotly : Interactive plots (architecture, trendlines, performance graphs)
- Statsmodels : Statistical analysis and trendline plotting
- Scikit-learn : Preprocessing, splitting datasets, and evaluation metrics
- Joblib/ Pickle : Saving/loading model weights or history if needed
- Colab / VS Code : For model development and experimentation

Chapter 5

RESULTS & DISCUSSION

5.1 EVALUATION METRICS

To assess the performance of the developed model, three evaluation metrics were used: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Each metric provides different insights into the model's accuracy and reliability.

5.1.1 Root Mean Squared Error (RMSE)

It measures the standard deviation of the residuals (prediction errors). It represents the square root of the average of the squared differences between the predicted and actual values. RMSE gives higher weight to large errors, making it particularly useful when large errors are undesirable. A lower RMSE value indicates a better fit of the model to the data.

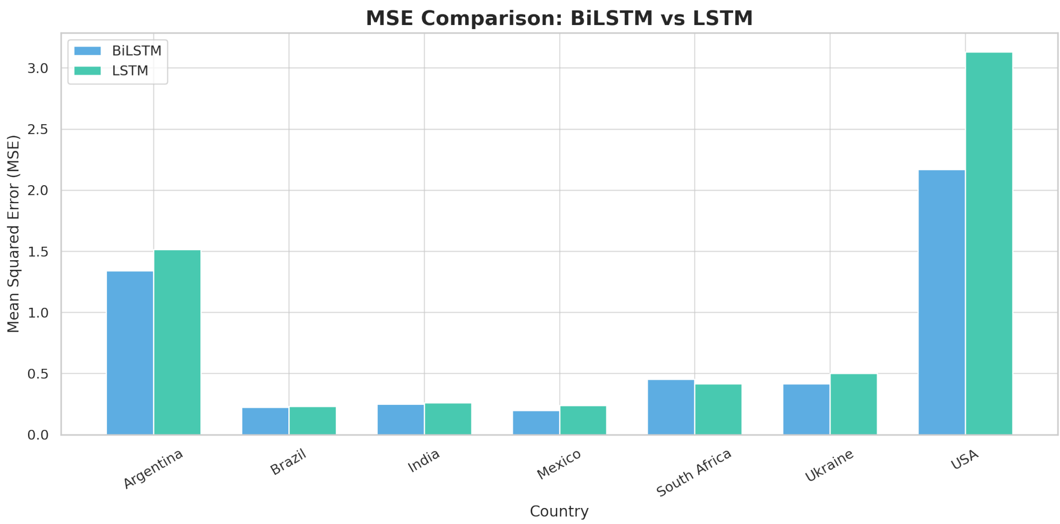


Figure 5.1: Root Mean Squared Error (RMSE)

The Fig 5.1 shows the Root Mean Squared Error (RMSE) comparison between BiLSTM and LSTM models across seven countries: Argentina, Brazil, India, Mexico, South

Africa, Ukraine, and USA. We can clearly observe that BiLSTM consistently achieves lower RMSE values compared to LSTM for most countries, indicating better prediction performance. The USA shows the highest RMSE for both models, while countries like Brazil and India have the lowest RMSE values.

5.1.2 Mean Absolute Error (MAE)

It measures the average magnitude of the errors in a set of predictions, without considering their direction. It calculates the average absolute differences between predicted and actual values. Unlike RMSE, MAE treats all errors equally and is less sensitive to outliers. A lower MAE value suggests higher model accuracy.

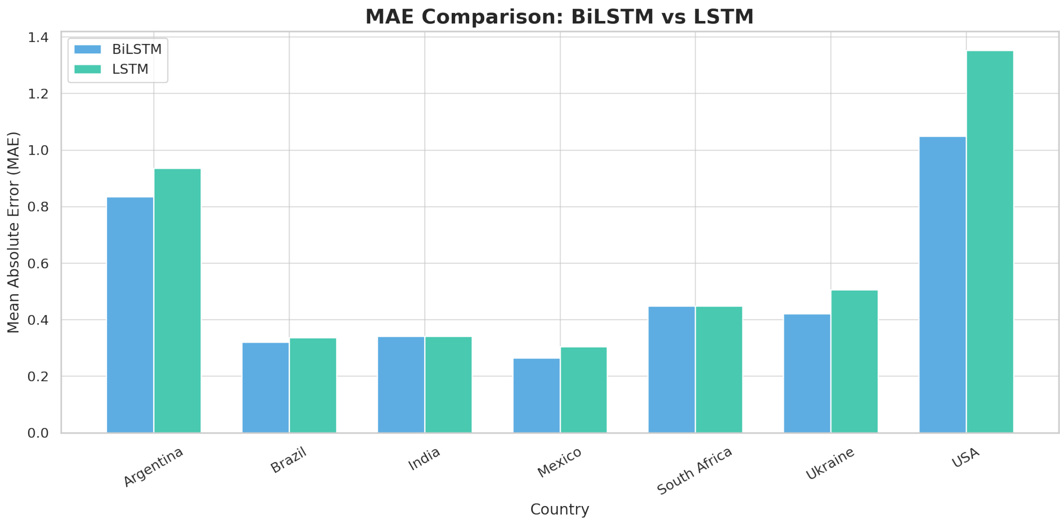


Figure 5.2: Mean Absolute Error (MAE)

The Figure 5.2 illustrates the Mean Absolute Error (MAE) comparison for maize yield prediction across the selected countries using BiLSTM and LSTM. The BiLSTM model generally performs better with lower MAE values. USA again records the highest MAE, while Mexico and Brazil show the lowest errors, highlighting better prediction accuracy in these countries.

5.1.3 Coefficient of determination (R2 Score)F

It evaluates how well the model's predictions approximate the actual data points. It indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 value of 1 implies perfect predictions, whereas an R^2 value of 0 suggests that the model performs no better than predicting the mean. Negative R^2 values indicate poor model performance.

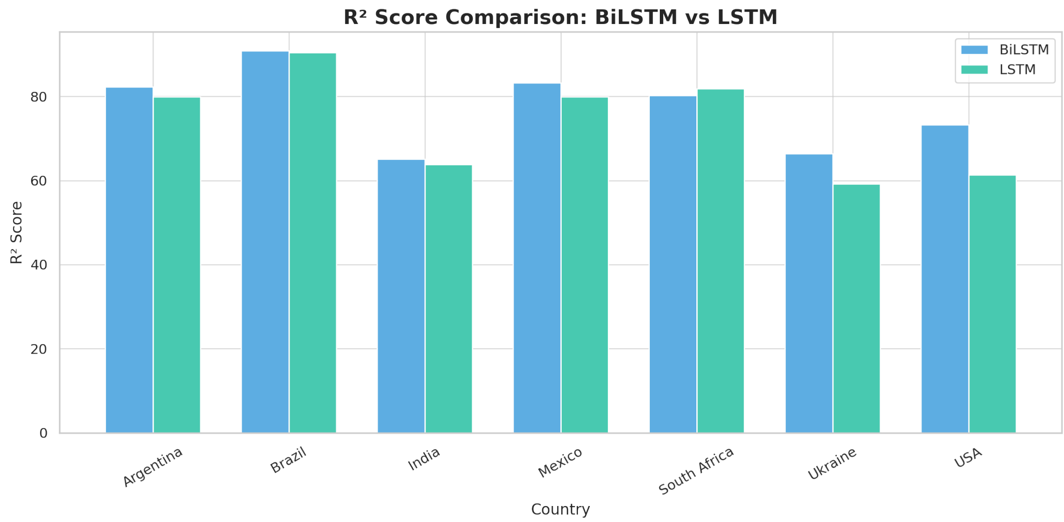


Figure 5.3: Coefficient of determination (R2 Score)

The Figure 5.3 presents the R^2 Score comparison of BiLSTM and LSTM models in maize yield prediction for seven countries. A higher R^2 score indicates better model performance. Brazil shows the highest R^2 scores, signifying strong model fit, while Ukraine and USA report the lower R^2 values, suggesting relatively less accurate predictions compared to other countries.

5.2 FUTURE PREDICTIONS

In this section, we share the predicted maize yield trends for different countries and Indian states all the way up to 2046. The figures give a glimpse into how maize production is expected to evolve in the future based on current patterns. While these predictions are not final, they offer valuable insights into where agricultural strengths might grow and where challenges could arise, helping guide future planning and decision-making.

5.2.1 Country-wise Predicitons

1. International Maize Yield Comparison: The figure 5.4 represents the average maize yield (in tons per hectare) across seven major maize-producing countries: the United States, Argentina, Ukraine, Brazil, South Africa, Mexico, and India. It clearly shows that the United States leads by a significant margin, producing the highest average maize yield among all the countries compared. Argentina and Ukraine also perform relatively well, with yields much higher than countries like Mexico and India. India's maize yield is noticeably the lowest among the group, highlighting a potential area for agricultural improvement.

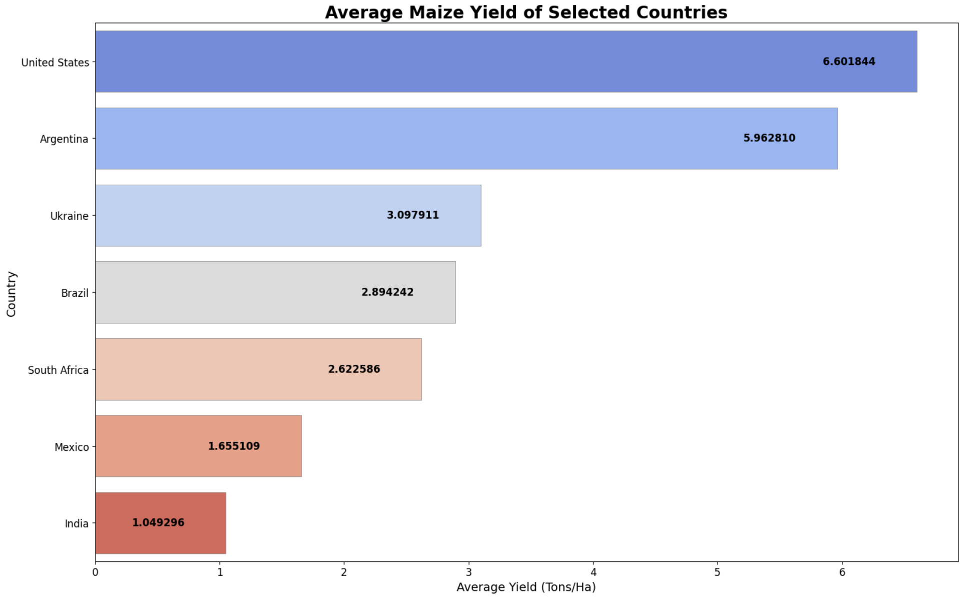


Figure 5.4: Comparison of Average Maize Yield Across Selected Countries

2. Predicted Maize Yield Trends for selected Countries: The figure 5.5 represents the predicted trend of maize yield from 2020 to 2046 for Argentina, South Africa, Mexico, Ukraine, India, Brazil, and the United States. Each colored line tracks the expected yield changes over time. The United States is projected to maintain the highest yield throughout the forecast period, continuing to outperform the others by a wide margin. Meanwhile, countries like India and Mexico show a relatively flatter growth curve, suggesting limited improvement without significant agricultural innovation. The trends also reveal a growing gap between high- and low-yield countries over time, emphasizing the uneven progress in global maize production.

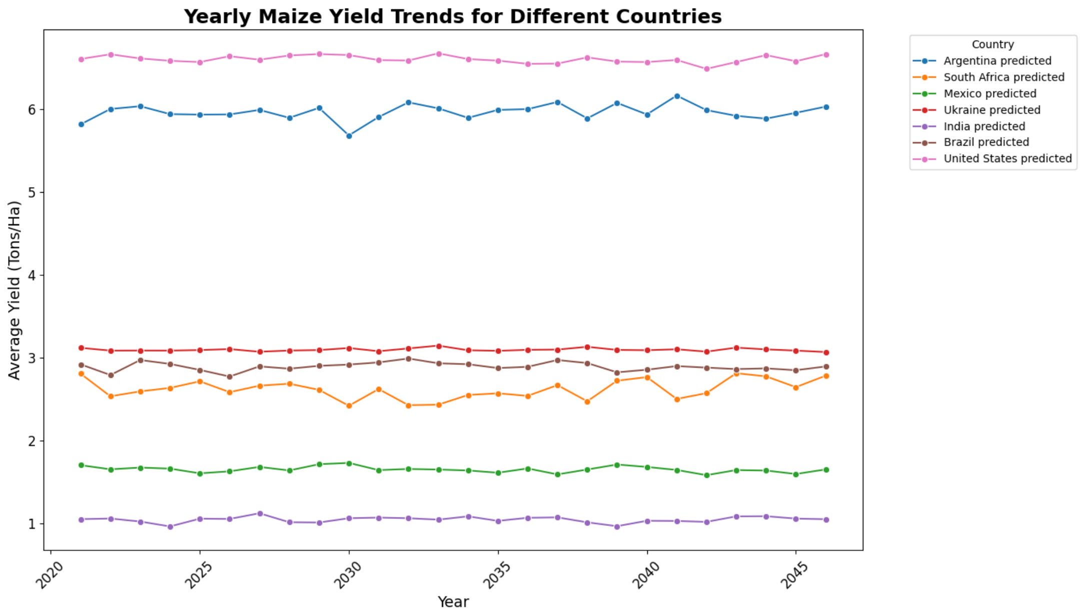


Figure 5.5: Predicted Maize Yield Trends for Selected Countries

5.2.2 State-wise Predicitons

1. Average Maize Yield Among Top Indian States: The figure 5.6 represents the current average maize yields for seven top maize-producing Indian states: Arunachal Pradesh, Chhattisgarh, Gujarat, Nagaland, Bihar, Uttarakhand, and Jammu and Kashmir. The data is presented through a horizontal bar chart, making it easy to compare yields at a glance. Arunachal Pradesh stands out with the highest average maize yield, while Jammu and Kashmir is positioned slightly lower on the scale. Overall, the differences between the states are relatively modest compared to the international differences, but still show important regional variations that could influence agricultural planning and policy decisions.

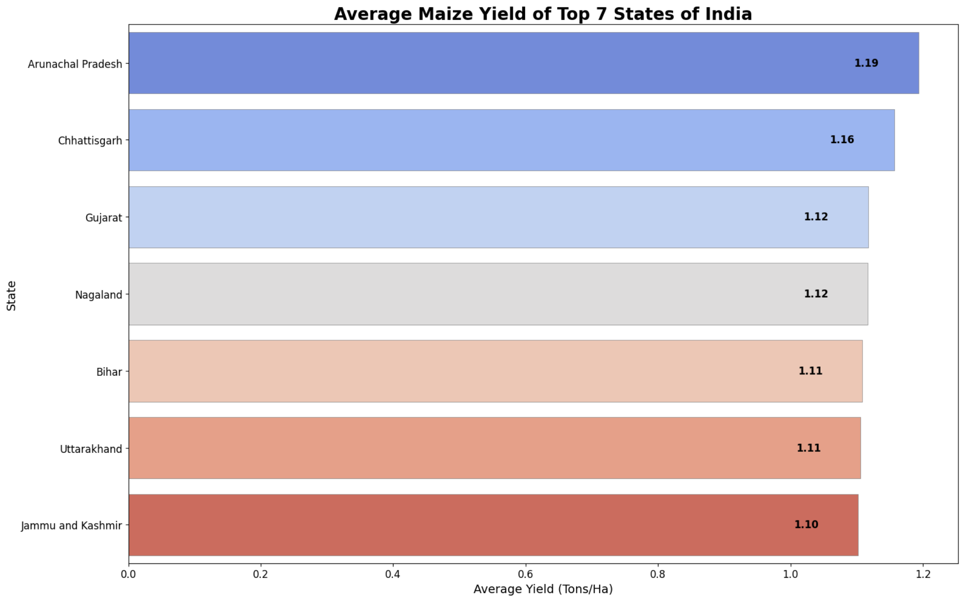


Figure 5.6: Top 7 Indian States by Average Maize Yield

2. Predicted Maize Yield Trends for Indian States: The figure 5.7 represents the predicted maize yield trends from 2020 to 2046 for several Indian states — Andhra Pradesh, West Bengal, Telangana, Manipur, Assam, and Tamil Nadu. The line graph captures the dynamic changes expected over nearly three decades. States like Manipur show more pronounced fluctuations, suggesting possible volatility in maize production, whereas others like Tamil Nadu and Andhra Pradesh have steadier, more gradual trends. These projections are crucial for understanding where interventions and innovations in farming practices might be most needed to stabilize or enhance maize yields in the future.

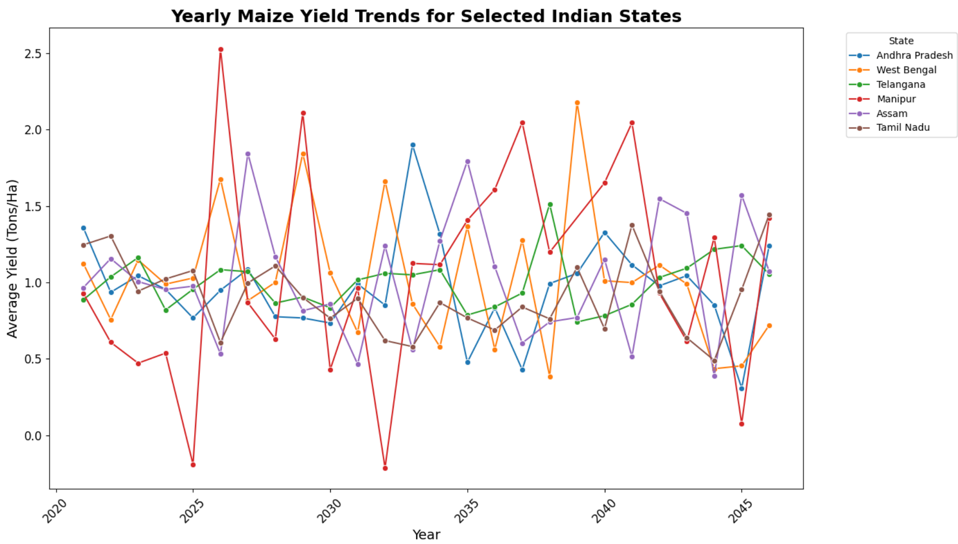


Figure 5.7: Predicted Maize Yield Trends for India

3. Geographic Distribution of Total Maize Yield in India: The below map from Fig 5.8 visualizes the predicted average maize yield across various states in India for the period 2021 to 2046, measured in tons per hectare. The color gradient ranges from yellow to dark green, where lighter shades represent lower yields and darker shades indicate higher yields.

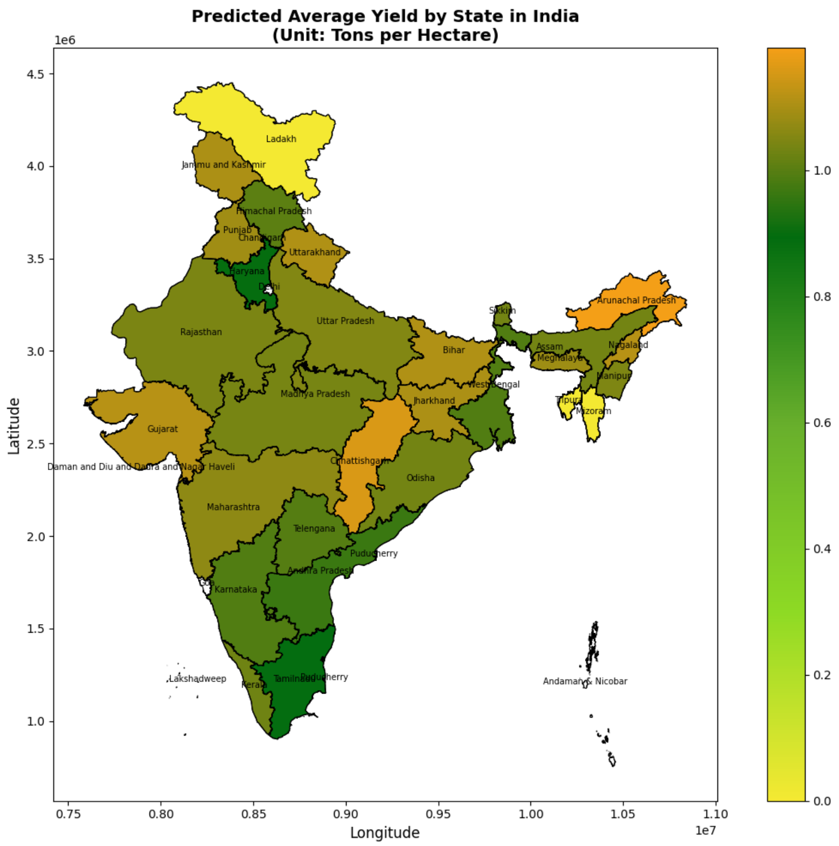


Figure 5.8: Geographic Distribution of Total Maize Yield in India

States like Punjab, Haryana, Tamil Nadu, and Kerala are projected to achieve higher maize yields, showcasing favorable agricultural conditions such as better irrigation systems, suitable climates, and advanced farming practices. In contrast, states like Arunachal Pradesh, Mizoram, and Ladakh display lower predicted yields, likely due to geographical and climatic constraints, limited arable land, and less agricultural infrastructure. This prediction map serves as a valuable tool for identifying regional disparities in maize production potential and for planning targeted agricultural interventions. Longitude and latitude markers are included for spatial]erence, and a color bar is provided to interpret the yield scale accurately. Overall, the map offers a forward-looking perspective on maize production trends in India, highlighting both high-potential zones and areas requiring strategic focus.

5.3 GUI RESULTS

This section presents the graphical user interface (GUI) developed for the crop yield prediction system using Streamlit. The GUI provides a dynamic and interactive platform for users to explore the system's functionalities. It includes comprehensive exploratory data analysis (EDA), visualizations of maize yield trends from 2021 to 2046, and predicted outcomes for different countries. The dashboard features country-wise and India-specific trendline analysis, regression analysis, predicted regression results, and model visualizations. Additionally, choropleth maps are incorporated to offer a geographical perspective on maize yield variations across different regions. The Streamlit-based interface ensures seamless navigation and real-time interaction with the data and model outputs.
The Figure 5.9 displays the graph of predicted maize yield from the year 2021 to 2046 for 7 countries, each from a continent. We can clearly observe that the US stands first throughout the years.

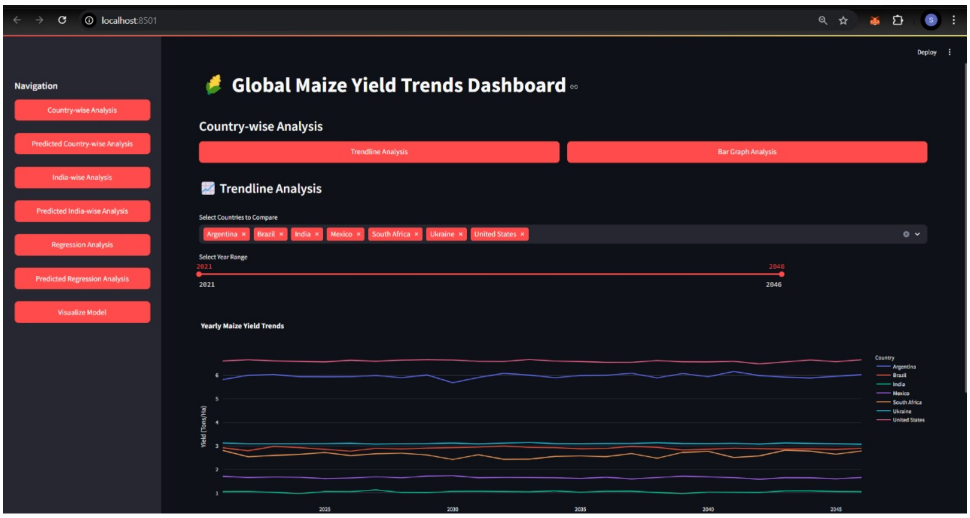


Figure 5.9: Graph of Global Maize yield

The Figure 5.10 displays the graph of predicted maize yield from the year 2021 to 2046 for Indian states. We can clearly observe that Andhra Pradesh has the highest yield till the year 2046, and the lowest is Madhya Pradesh.

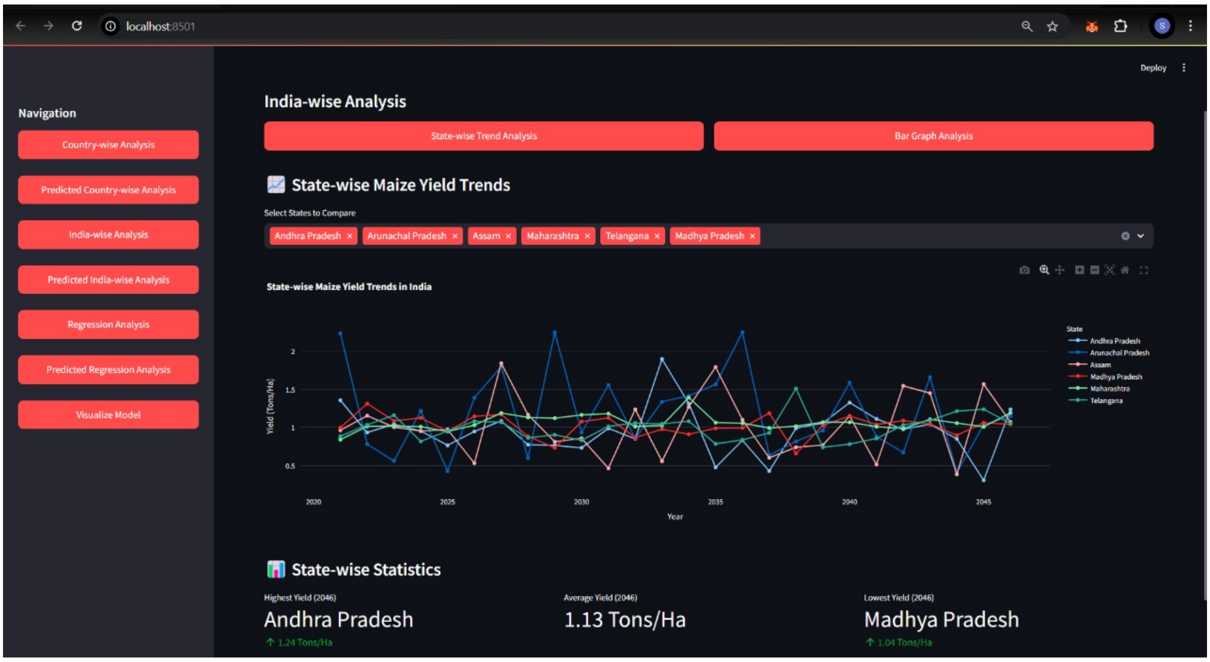


Figure 5.10: Graph of Indian Maize yield

The Figure 5.11 depicts the relationship between climate factors and yield. Different variables are considered, and trend lines are shown.

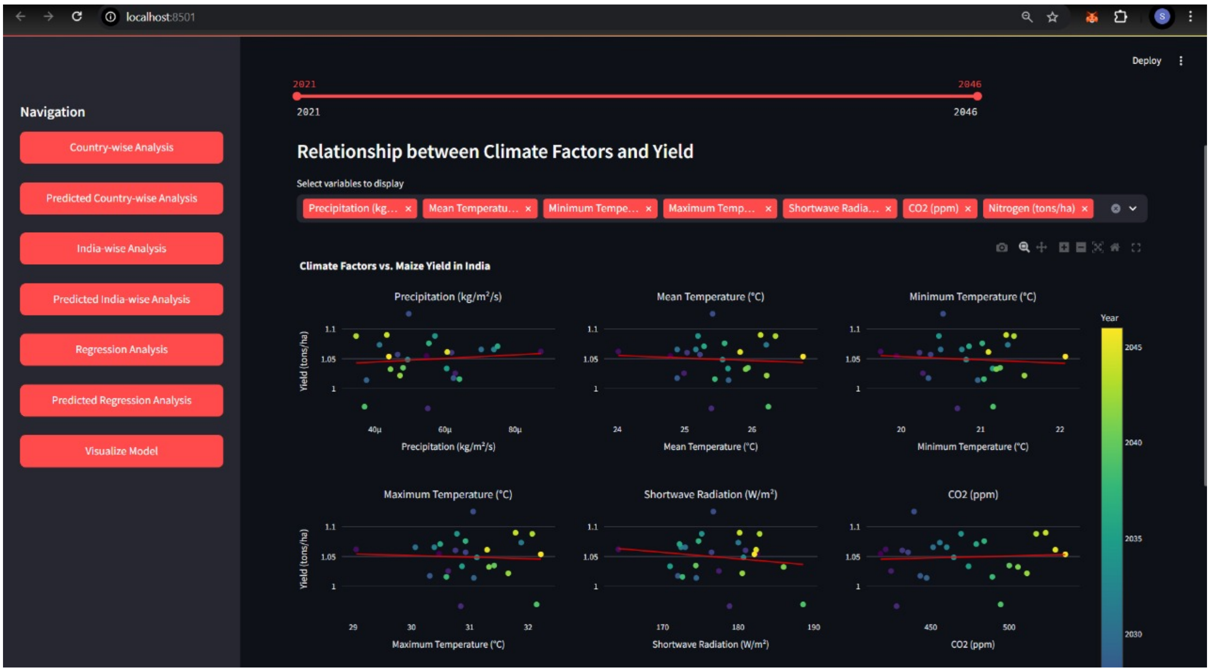


Figure 5.11: Regression Analysis

5.4 TIMELINE OF THE PROJECT

Figure 5.12 depicts how the project is planned to progress through Semester 7, from choosing the topic to working on the dataset, step by step each week.

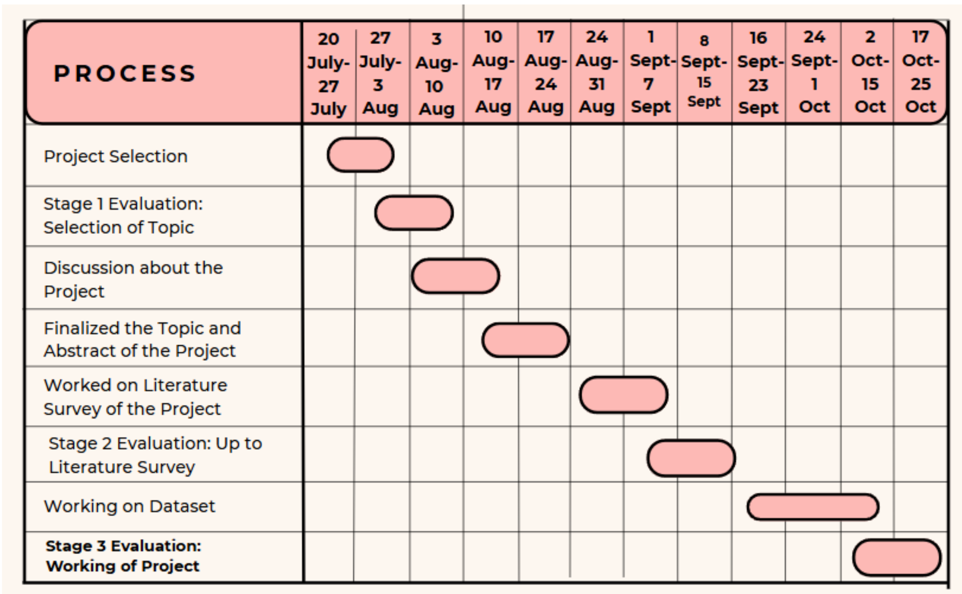


Figure 5.12: Gantt Chart for Semester 7

Figure 5.13 depicts the project plan for Semester 8, showing the flow from evaluations and improvements to final report submission and poster preparation.

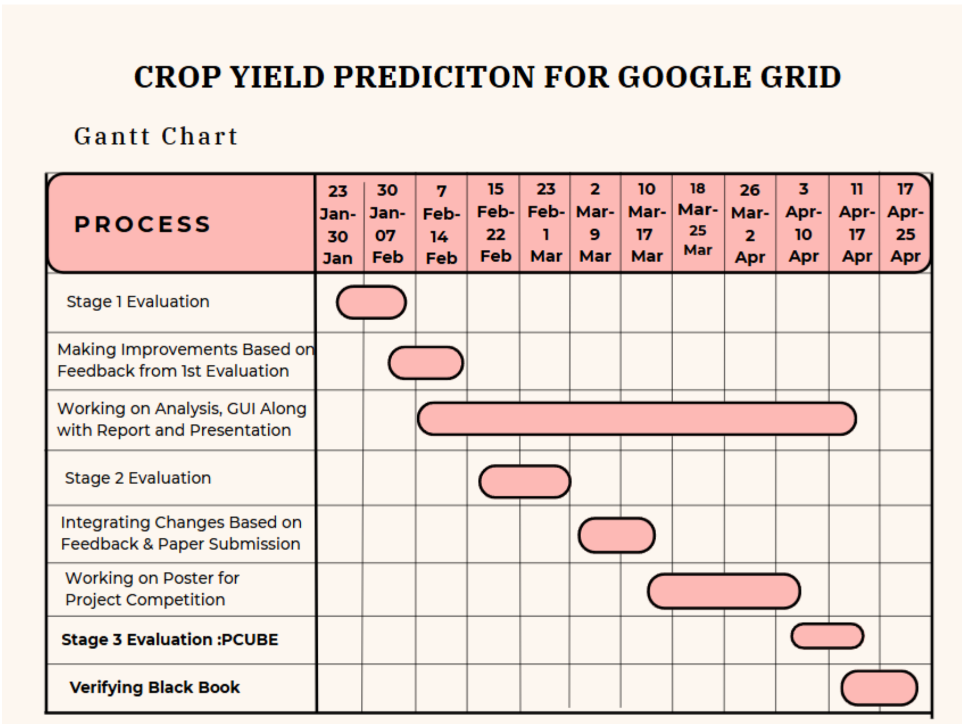


Figure 5.13: Gantt chart for Semester 8

Chapter 6
CONCLUSION & FUTURE SCOPE

6.1 CONCLUSION

This study analyzed maize production trends over 39 years (1982–2020) across major producing countries (India, the United States, Ukraine, Brazil, and South Africa), focusing on the influence of rainfall, temperature, and atmospheric CO_2 levels on maize yield. It also emphasized how rainfall distribution during different crop stages impacts productivity and highlighted regional disparities within India, underlining the need for localized adaptation strategies through data analytics and geospatial techniques.

To forecast future trends, LSTM and BiLSTM models were developed and their performances compared. These models were used to predict maize yields from 2021 to 2046, offering foresight into potential future scenarios. The BiLSTM model, with its ability to capture dependencies in both forward and backward directions, demonstrated improved accuracy over the traditional LSTM approach for this time-series forecasting task. Specifically for India, the BiLSTM model achieved an R^2 score of 65.05%, a mean squared error (MSE) of 0.2518, and a mean absolute error (MAE) of 0.3413, outperforming the LSTM model which attained an R^2 score of 63.80%, an MSE of 0.2608, and an MAE of 0.3412. Furthermore, the BiLSTM model achieved its highest R^2 score of 90.86% in Brazil, while the LSTM model achieved its highest R^2 score of 90.46% also in Brazil.

Finally, a user-friendly graphical user interface (GUI) was built to visualize both historical and predicted data, making the insights accessible to farmers, researchers, and policymakers. Overall, the project contributes meaningfully to the domain of agricultural forecasting by combining machine learning, climate analysis, and interactive visualization.

6.2 FUTURE SCOPE

While this study has achieved its intended objectives, there remain several opportunities for future enhancement and expansion:

- **Integration of More Variables:** Incorporating additional influencing factors such as soil quality, pest occurrence, irrigation practices, fertilizer usage, and socio-economic data could further improve prediction accuracy.
- **Higher-Resolution Data:** Using finer-scale climatic and agricultural data (such as district-level data within countries) can offer more precise and localized predictions, particularly beneficial for regional policymaking.
- **Model Optimization and Experimentation:** Future work could explore other advanced deep learning architectures like Transformer models or hybrid models (e.g., CNN-LSTM) to further boost prediction accuracy.
- **Uncertainty Quantification:** Incorporating techniques to estimate and communicate the uncertainty of predictions would help decision-makers better understand the range of possible outcomes.
- **Real-Time Forecasting:** Building a dynamic system that updates predictions in near real-time as new climate and agricultural data become available would significantly enhance the practical usability of the project.
- **Expansion to Other Crops:** Extending the methodology to predict yields for other critical crops such as wheat, rice, and soybeans would broaden the project's applicability and impact.
- **Policy Recommendation Framework:** Based on the insights and predictions, future work could aim at developing automated tools that provide customized agricultural advice and policy recommendations.

Through these extensions, the project can evolve into a more comprehensive decision-support platform, helping to ensure sustainable agriculture and food security in the face of global climate challenges.

Chapter A

CODE SAMPLE

```
# Define the BLSTM model for crop yield prediction
from tensorflow.keras.optimizers import Nadam

def BLstm_model():
    input1 = tf.keras.layers.Input(shape=(X_train.shape[1], X_train.shape[2]))
    blstm_1 = tf.keras.layers.Bidirectional(
        tf.keras.layers.LSTM(hidden_units, activation='tanh', return_sequences=
            ↪ True, name='blstm1')
    )(input1)
    drop1 = tf.keras.layers.Dropout(dropout_size)(blstm_1)

    blstm_2 = tf.keras.layers.Bidirectional(
        tf.keras.layers.LSTM(hidden_units, activation='tanh', return_sequences=
            ↪ False, name='blstm2')
    )(drop1)
    drop2 = tf.keras.layers.Dropout(dropout_size)(blstm_2)

    dense_1 = tf.keras.layers.Dense(hidden_units * 2, activation='relu')(drop2)
        ↪   # Increased hidden units
    dense_2 = tf.keras.layers.Dense(hidden_units, activation='relu')(dense_1)
        ↪ # Added one more dense layer
    output = tf.keras.layers.Dense(n_out)(dense_2)

    model = tf.keras.Model(inputs=input1, outputs=output)

    optimizer = Nadam(learning_rate=0.001)

    model.compile(optimizer=optimizer, loss='mse', metrics=['mae'])
    return model


# Define the early stopping callback

Early = tf.keras.callbacks.EarlyStopping(
    monitor="val_loss",
    min_delta=1e-4,
    patience=50,  # Adjusted patience for yield prediction
    verbose=1,
    mode="min",
```

```
    baseline=None,
    restore_best_weights=True,
)

# Define the checkpoint callback to save the model with the best validation
    ↪ loss

checkpoint = tf.keras.callbacks.ModelCheckpoint(
    filepath="BLSTM_7_crop_yield_prediction_20.h5",  # Updated filename
    monitor="val_loss",
    mode='min',
    verbose=1,
    save_best_only=True
)


from tensorflow.keras.callbacks import ReduceLROnPlateau

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=10,
    ↪ min_lr=1e-6)

history = model.fit(
    X_train,
    y_train,
    batch_size=Batch_size,
    epochs=200,
    validation_split=0.2,
    callbacks=[checkpoint, Early, reduce_lr]
)

# Define the LSTM model for maize yield prediction

def Lstm_model():
    input_layer = tf.keras.layers.Input(shape=(X_train.shape[1], X_train.shape
        ↪ [2]))

    lstm_1 = tf.keras.layers.LSTM(hidden_units, activation='tanh',
        ↪ return_sequences=True, name='lstm1')(input_layer)
    dropout_1 = tf.keras.layers.Dropout(dropout_size)(lstm_1)

    lstm_2 = tf.keras.layers.LSTM(hidden_units, activation='tanh',
        ↪ return_sequences=False, name='lstm2')(dropout_1)
    dropout_2 = tf.keras.layers.Dropout(dropout_size)(lstm_2)

    dense_hidden = tf.keras.layers.Dense(hidden_units, activation='tanh')(
        ↪ dropout_2)
    output_layer = tf.keras.layers.Dense(n_out)(dense_hidden)

    model = tf.keras.Model(inputs=input_layer, outputs=output_layer)
    optimizer = tf.keras.optimizers.Adam(learning_rate=lr)
    model.compile(optimizer=optimizer, loss='mse', metrics=['mae'])
```

```
    return model


history = model.fit(
    X_train,
    Y_train,
    batch_size=batch_size,
    epochs=epochs,
    validation_split=0.3,
    callbacks=[early_stop, checkpoint]
)

#define hyperparameters

 # number of output which is one in our case
n_out=1

#duration of a season
seq_len=240

#number of hidden units in each LSTM layer
hidden_units=128
dropout_size=0.2

#learning rate
lr=0.001

#learning rate decay
decay=1e-5
Batch_size=64

#maximum number of epochs to train the model
Epoch=200

#save the trained model with the name
NAME = "BLSTM_Crop_Yield_Prediction_20.h5"

CREATE TABLE tasmin_train_maize_full (
    crop TEXT,
    year REAL,
    lon REAL,
    lat REAL,
    variable TEXT,
    col_0 REAL,
    col_1 REAL,
    col_2 REAL,
    col_3 REAL,
    col_4 REAL,
    col_5 REAL,
          .. ,
```

```
    col_239 REAL
);

BULK INSERT [FinalDB].dbo.[tasmin_train_maize_full]
FROM 'C:\Users\sahil\Desktop\Jupyter\CSV-Files\Maize\Train\tasmin_maize_train.
   ↪ csv'
WITH (
    FORMAT = 'CSV',
    FIRSTROW = 2,
    DATAFILETYPE = 'char',
    FIELDTERMINATOR = ',',
    TABLOCK
)

UPDATE formatted_data_test
SET pr = [testingDb].dbo.pr_test_maize_full.col_0
FROM [testingDb].dbo.pr_test_maize_full
INNER JOIN formatted_data_test
ON formatted_data_test.lon = [testingDb].dbo.pr_test_maize_full.lon
   AND formatted_data_test.lat = [testingDb].dbo.pr_test_maize_full.lat
   AND formatted_data_test.[Year] = [testingDb].dbo.pr_test_maize_full.[Year]
   and formatted_data_test.[day]='0'

UPDATE formatted_data_test
SET pr = [testingDb].dbo.pr_test_maize_full.col_1
FROM [testingDb].dbo.pr_test_maize_full
INNER JOIN formatted_data_test
ON formatted_data_test.lon = [testingDb].dbo.pr_test_maize_full.lon
   AND formatted_data_test.lat = [testingDb].dbo.pr_test_maize_full.lat
   AND formatted_data_test.[Year] = [testingDb].dbo.pr_test_maize_full.[Year]
   and formatted_data_test.[day]='1'

SELECT
    crop, [year], [lon], [lat], [Day], [pr], [rds],[tas],[tasmin],[tasmax],[
       ↪ nitrogen],[co2],[texture_class],[yield]
FROM
    formatted_data
ORDER BY
    [year],
    [lon],
    [lat],
    [day];
```

BIBLIOGRAPHY

[1] A. Mohammadzadeh, Y. Azimzadeh, R. Lotfi, E. Zadhasan, K. Alizadeh, and H. H. Khoshro, “Analyzing the rainfed wheat yield gap in Northwest Iran,” Farming Syst., vol. 3, no. 1, 2025, doi: 10.1016/j.farsys.2024.100126.

[2] M. Jiang, C. Dong, W. Bian, W. Zhang, and Y. Wang, “Effects of different fertilization practices on maize yield, soil nutrients, soil moisture, and water use efficiency in northern China based on a meta-analysis,” Sci. Rep., vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-57031-z.

[3] Z. Yang et al., “Estimation of Millet Aboveground Biomass Utilizing Multi-Source UAV Image Feature Fusion,” Agronomy, vol. 14, no. 4, 2024, doi: 10.3390/agronomy14040701.

[4] J. S. Gerber et al., “Global spatially explicit yield gap time trends reveal regions at risk of future crop yield stagnation,” Nat. Food, vol. 5, no. 2, pp. 125–135, 2024, doi: 10.1038/s43016-023-00913-8.

[5] R. Landage, V. T. Jadhav, and P. P. Patil, “Trends of temperature and precipitation extreme indices in north Maharashtra,” J. Agrometeorol., vol. 6, no. 1, pp. 5–9, 2024, doi: https://doi.org/10.54386/jam.v26i4.2735.

[6] A. Aucahuasi-Almidon, C. Cabrera-Carranza, and J. Garate-Quispe, “Trend analysis and change-point detection of temperature and rainfall in southern Peruvian Amazon and its relation to deforestation ANDREA,” J. Agrometeorol., vol. 26, no. 4, pp. 425–430, 2024, doi: https://doi.org/10.54386/jam.v26i4.2687.

[7] P. Sowmya and A. V. Krishna Prasad, “A Comprehensive Dataset for Crop Yield Analysis in Telangana Region,” Int. J. Sci. Res., vol. 12, no. 12, pp. 222–227, 2023, doi: 10.21275/sr231202135931.

[8] J. Wang, X. Wang, Y. Ji, and J. Gao, “Climate factors determine the utilization strategy of forest plant resources at large scales,” Front. Plant Sci., vol. 13, no. August, pp. 1–12, 2022, doi: 10.3389/fpls.2022.990441.

[9] M. Naveed, H. HS, Y. QM, D. H, and S. Z, “Analyzing the Impact of Climate Change on Cotton Yield Using Spatial Analysis and Statistical Modeling in the Indus River

Basin, Pakistan," Ann. Agric. Crop Sci., vol. 6, no. 5, 2021, doi: 10.26420/annagric-cropsci.2021.1089.

[10] R. K. Ranjan, M. Kumari, M. Rahman, C. K. Panda, F. Homa, and Sashikala, "Effect of Climate Variables on Yield of Major Crop in Samastipur District of Bihar: A Time Series Analysis," Econ. Aff. (New Delhi), vol. 65, no. 4, pp. 637–644, 2020, doi: 10.46852/0424-2513.4.2020.21.

[11] M. Heino, J. H. A. Guillaume, C. Müller, T. Iizumi, and M. Kummu, "A multi-model analysis of teleconnected crop yield variability in a range of cropping systems," Earth Syst. Dyn., vol. 11, no. 1, pp. 113–128, 2020, doi: 10.5194/esd-11-113-2020.

[12] M. W. Pramujati, "Spatial-temporal crop yield analysis in East Kalimantan, Indonesia: spatial disaggregation of crop yield data and estimation of future production," 2018. [Online]. Available: https://run.unl.pt/handle/10362/33712

[13] M. Aryal, Regmi P P, Thapa R B, Pande K R, and Pant K P, "Impact of Climate Variables To Major Food Crops'Yieldin Midhills of Western Development Region, Nepal," J. Agric. Environ., vol. 17, pp. 65–72, 2016.

[14] Aruvansh Nigam, Saksham Garg, Archit Agrawal (2019). Crop Yield Prediction Using Machine Learning Algorithms-2019. https://ieeexplore.ieee.org/document/8985951

[15] D. Jayanarayana Reddy, Dr M. Rudra Kumar (2021) Crop Yield Prediction using Machine Learning Algorithm 2021. https://ieeexplore.ieee.org/abstract/document/9432236

[16] Akshay Kumar Gajula, Jaswanth Singamsetty, Vineela Chandra Dodda, Lakshmi Kuruguntla (2021). Prediction of Crop and Yield in Agriculture Using Machine Learning July 2021. https://ieeexplore.ieee.org/abstract/document/9579843

[17] Kavita Jhajharia, Pratistha Mathur, Sanchit Jain, Sukriti Nijhawan (2023), Crop Yield Prediction using Machine Learning and Deep Learning Techniques. https://doi.org/10.1016/j.procs.2023.01.023

[18] Sharma, Priyanka Dadheech, Pankaj Aneja, Nagender Aneja, Sandhya. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3321861.

[19] Sai Nishant, Potnuru Venkat, Pinapa Avinash, Bollu Jabber, B. (2020). Crop Yield Prediction based on Indian Agriculture using Machine Learning. 1-4. 10.1109/INCET49848.2020.9154036.

[20] A. Morales and F. J. Villalobos, “Machine learning models for yield prediction using synthetic datasets of sunflower and wheat,” Agricultural Systems, vol. 226, 2023, doi: 10.1016/j.agsy.2023.103684.

[21] T. van Klompenburg, E. Kassahun, and A. Catal, “Crop yield prediction using machine learning: A systematic literature review,” Computers and Electronics in Agriculture, vol. 177, 2020, doi: 10.1016/j.compag.2020.105709.

[22] Thomas van Klompenburga, Ayalew Kassahuna, Cagatay Catalb (2020). Crop yield prediction using machine learning: A systematic literature review. https://doi.org/10.1016/j.compag.2020.105709

[23] Alejandro Morales, Francisco J. Villalobos (2023). Using Machine Learning for Crop Yield Prediction In the Past or Future. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1128388/full

[24] Impacts of using GM crop technology 1996–2016,” GM Crop. Food, vol. 9, no. 2, pp. 59–89, 2018, doi: 10.1080/21645698.2018.1464866.

[25] G. P. B. de Sousa et al., “Assessing soil degradation in Brazilian agriculture by a remote sensing approach to monitor bare soil frequency: impact on soil carbon,” Soil Adv., vol. 2, no. June, p. 100011, 2024, doi: 10.1016/j.soilad.2024.100011.

[26] R. Peiretti and J. Dumanski, “The transformation of agriculture in Argentina through soil conservation,” Int. Soil Water Conserv. Res., vol. 2, no. 1, pp. 14–20, 2014, doi: 10.1016/S2095-6339(15)30010-1.

[27] K. Bonne, “The black gold of Ukraine and the most fertile soils in the world,” GondwanaTalks, 2022.

[28] O. Bourgeois, “Maize Production Technologies in India,” 2010.

[29] M. O’Leary, “Maize: From Mexico to the world,” Cimmyt, 2016.

[30] S. Matimolane, H. Chikoore, F. I. Mathivha, and E. Kori, “Maize producers’ vulnerability to climate change: Evidence from Makhuduthamaga Local Municipality, South Africa,” Jamba J. Disaster Risk Stud., vol. 14, no. 1, pp. 1–10, 2022, doi: 10.4102/JAMBA.V14I1.1165.

[31] R. Nehring, “Yield of dreams: Marching west and the politics of scientific knowledge in the Brazilian Agricultural Research Corporation (Embrapa),” Geoforum, vol. 77, pp. 206–217, 2016, doi: 10.1016/j.geoforum.2016.11.006.

Chapter B

PLAGIARISM REPORT

DrillBit

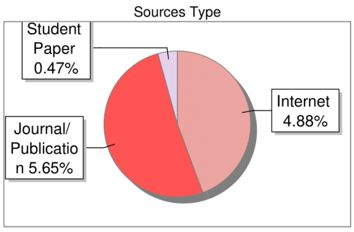
The Report is Generated by DrillBit Plagiarism Detection Software

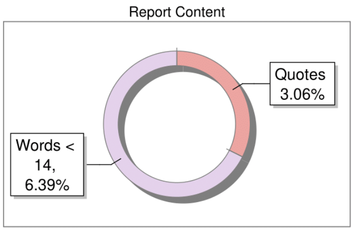
Submission Information

Author Name	Suyash Malekar
Title	Crop yield prediction for Google Grid
Paper/Submission ID	3550659
Submitted by	lakshmi.gadhikar@fcrit.ac.in
Submission Date	2025-04-27 22:15:24
Total Pages, Total Words	31, 10019
Document type	Project Work

Result Information

Similarity **11 %**





Exclude Information

Quotes	Not Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

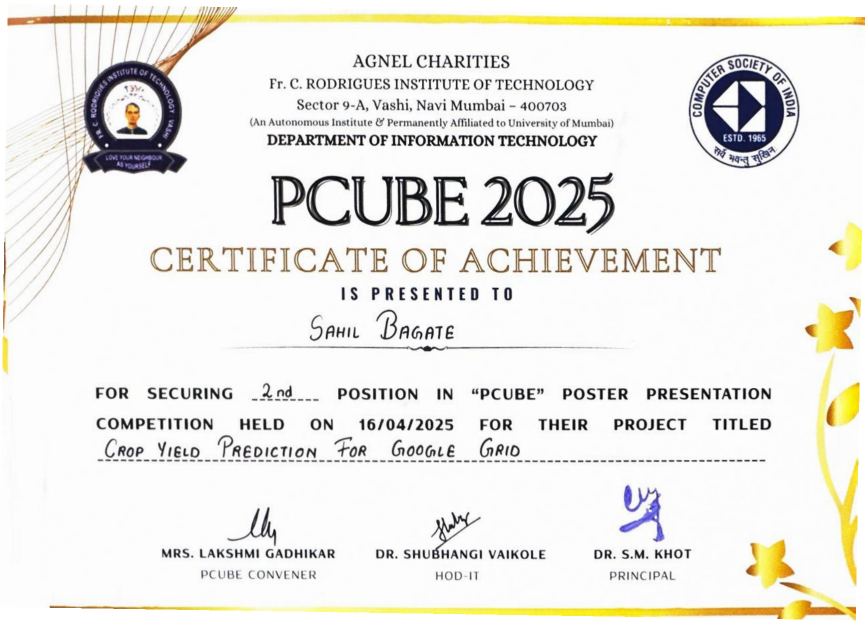
Chapter C

PAPER PUBLISHED CERTIFICATES

We have submitted our paper titled "Trend Analysis in Maize Crop Yield with Google Grid and Meteorological Parameters" to the IEEE International Conference on Engineering Innovations and Technologies (ICoEIT 2025). The conference will be held on 4th - 5th July 2025, organized by IEEE Madhya Pradesh Section and technically co-sponsored by IEEE, along with LNCT Group of Colleges, Bhopal.

Chapter D

PROJECT COMPETITION CERTIFICATES



AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

PCUBE 2025

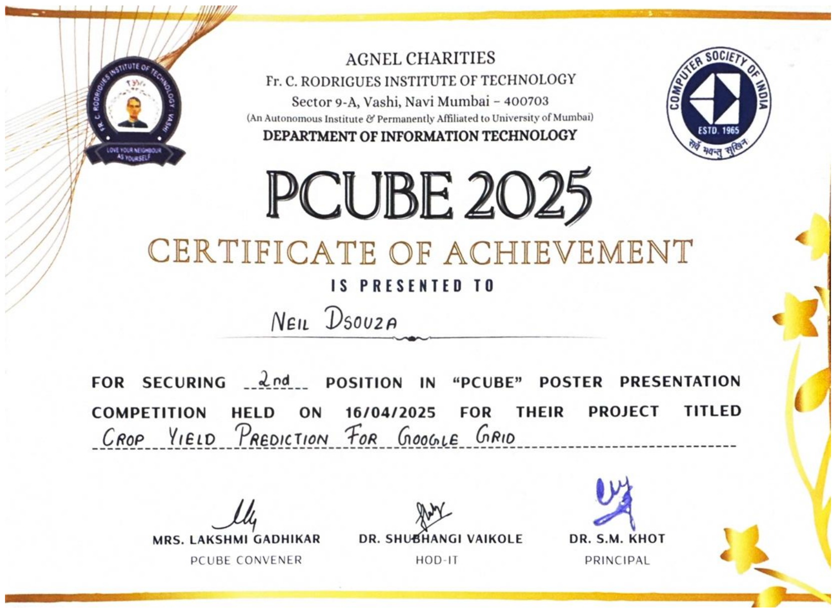
CERTIFICATE OF ACHIEVEMENT

IS PRESENTED TO

Sahil Bagate

FOR SECURING 2nd POSITION IN "PCUBE" POSTER PRESENTATION COMPETITION HELD ON 16/04/2025 FOR THEIR PROJECT TITLED Crop Yield Prediction For Google Grid

MRS. LAKSHMI GADHIKAR, PCUBE CONVENER
DR. SHUBHANGI VAIKOLE, HOD-IT
DR. S.M. KHOT, PRINCIPAL



AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

PCUBE 2025

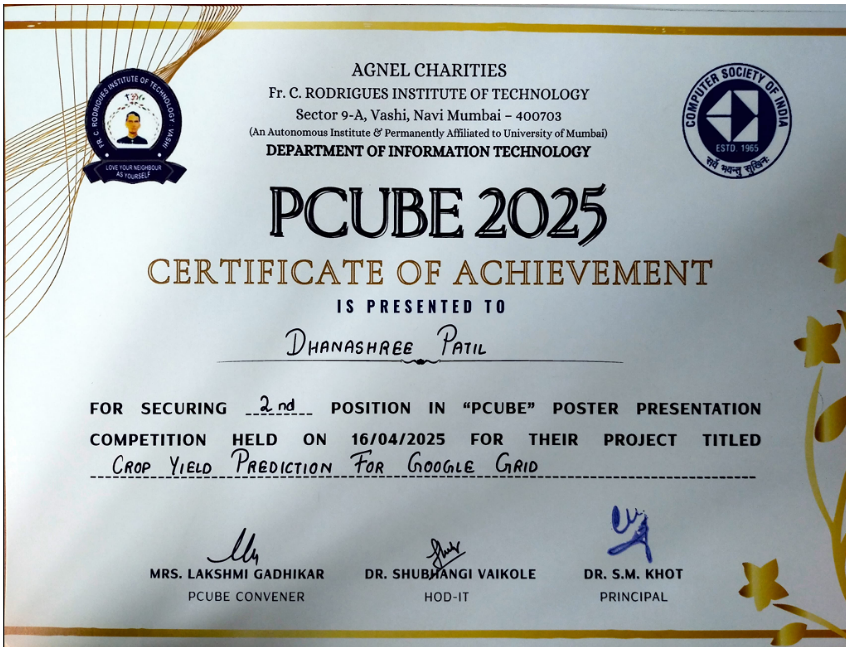
CERTIFICATE OF ACHIEVEMENT

IS PRESENTED TO

Neil Dsouza

FOR SECURING 2nd POSITION IN "PCUBE" POSTER PRESENTATION COMPETITION HELD ON 16/04/2025 FOR THEIR PROJECT TITLED Crop Yield Prediction For Google Grid

MRS. LAKSHMI GADHIKAR, PCUBE CONVENER
DR. SHUBHANGI VAIKOLE, HOD-IT
DR. S.M. KHOT, PRINCIPAL



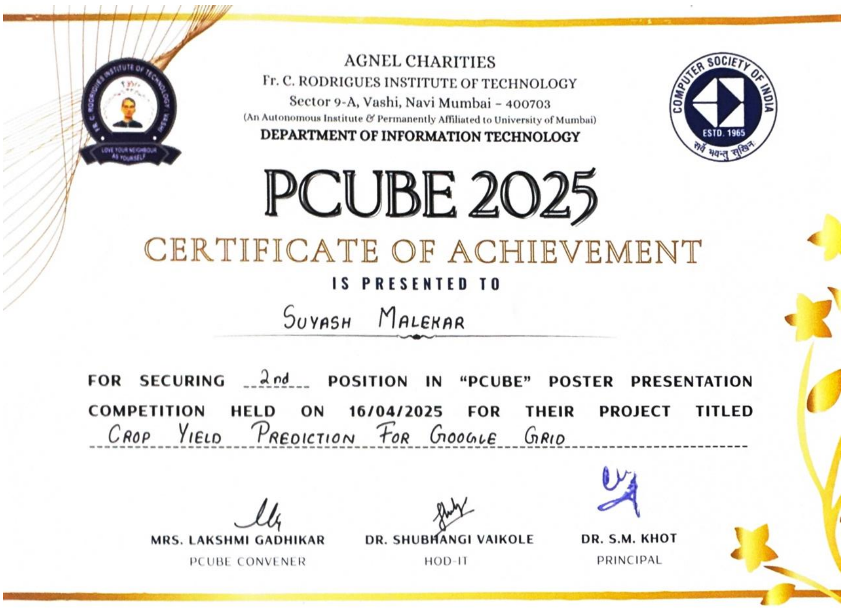


Figure D.1: Certificates of 1st Runner-ups in PCUBE.


AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

Techsparks 2025

CERTIFICATE OF PARTICIPATION

IS PRESENTED TO

Sahil Ankush Bagate

FOR PARTICIPATING IN "TECHSPARKS" - AN INTERCOLLEGIATE PROJECT PRESENTATION COMPETITION HELD ON 16TH APRIL 2025

MRS. SUPRIYA JOSHI	DR. TRUPTI LOTLIKAR	DR. SHUBHANGI VAIKOLE	DR. S.M. KHOT
TECHSPARKS CO-ORDINATOR	TECHSPARKS CONVENER	HOD-IT	PRINCIPAL


AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

Techsparks 2025

CERTIFICATE OF PARTICIPATION

IS PRESENTED TO

Neil Anthon Dsouza

FOR PARTICIPATING IN "TECHSPARKS" - AN INTERCOLLEGIATE PROJECT PRESENTATION COMPETITION HELD ON 16TH APRIL 2025

MRS. SUPRIYA JOSHI	DR. TRUPTI LOTLIKAR	DR. SHUBHANGI VAIKOLE	DR. S.M. KHOT
TECHSPARKS CO-ORDINATOR	TECHSPARKS CONVENER	HOD-IT	PRINCIPAL

AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

Techsparks 2025

CERTIFICATE OF PARTICIPATION

IS PRESENTED TO

Dhanashree Ravindra Patil

FOR PARTICIPATING IN "TECHSPARKS" - AN INTERCOLLEGIATE PROJECT PRESENTATION COMPETITION HELD ON 16TH APRIL 2025

MRS. SUPRIYA JOSHI	DR. TRUPTI LOTLIKAR	DR. SHUBHANGI VAIKOLE	DR. S.M. KHOT
TECHSPARKS CO-ORDINATOR	TECHSPARKS CONVENER	HOD-IT	PRINCIPAL

AGNEL CHARITIES
Fr. C. RODRIGUES INSTITUTE OF TECHNOLOGY
Sector 9-A, Vashi, Navi Mumbai – 400703
(An Autonomous Institute & Permanently Affiliated to University of Mumbai)
DEPARTMENT OF INFORMATION TECHNOLOGY

Techsparks 2025

CERTIFICATE OF PARTICIPATION

IS PRESENTED TO

Suyash Rajendra Malekar

FOR PARTICIPATING IN "TECHSPARKS" - AN INTERCOLLEGIATE PROJECT PRESENTATION COMPETITION HELD ON 16TH APRIL 2025

MRS. SUPRIYA JOSHI	DR. TRUPTI LOTLIKAR	DR. SHUBHANGI VAIKOLE	DR. S.M. KHOT
TECHSPARKS CO-ORDINATOR	TECHSPARKS CONVENER	HOD-IT	PRINCIPAL

Chapter E

ACKNOWLEDGEMENT

The making of the project “Crop yield prediction for Google grid” involves the contribution of many people. We express deep gratitude to our project guide and mentor Prof. Poonam Bari for her constant motivation to think out of the box and immense contribution throughout this project. We convey our heartfelt thanks to the project coordinator Prof. Lakshmi Gadikar for supporting and guiding us throughout the process. We would also like to convey our heartfelt gratitude to the Head of the Department of Information Technology, Dr. Shubhangi Vaikole for her constant support and motivation. We would like to convey our sincere thanks to Dr. S.M. Khot, principal of Fr. C. Rodrigues Institute of Technology, Vashi for giving us the opportunity to showcase our skills and providing us with the necessary resources.We also extend our heartfelt thanks to our families and well-wishers.

Sahil Bagate (5021105)

−−−−−−−−−−−

Neil Dsouza (5021113)

−−−−−−−−−−−

Dhanashree Patil (5021142)

−−−−−−−−−−−

Suyash Malekar (5021166)

−−−−−−−−−−−