Project Stage I Report

On

"Agriculture forecast module"

Submitted in the partial fulfillment of the requirements
for the Degree of
Bachelor of Technology
in
Electronics & Telecommunication Engineering

By

Tushar Padmawar (2014110957)
Rajat Gupta (2014110917)
Daksh Kumar (2014110942)

Under the guidance of

Asst. Pro. Deepak Kumar Ray

Department of Electronics & Telecommunication Engineering
Bharati Vidyapeeth Deemed University
College of Engineering, Pune – 4110043

Academic Year 2023-24

ACKNOWLEDGEMENT

We would like to extend our sincere gratitude to the Principal **Dr. Vidula Sohani**, Head of Department Electronics & Telecommunication, **Prof. Shruti Oza**, for nurturing a congenial yet competitive environment, which motivates all the students not only to pursue goals but also to elevate the Humanitarian level.

Inspiration and guidance are invaluable in every aspect of life, which we have received from our respected project guide **Asst. Pro. Deepak Ray,** who gave us his careful and ardent guidance because of which we are able to complete this project. More words won't suffice to express our gratitude to his untiring devotion. He undoubtedly belongs to the members of the artistic gallery who are masters in all aspects.

We would also like to thank all the faculty members who directly or indirectly helped us from time to time with their invaluable inputs.

-Rajat Gupta

-Tushar Padmawar

-Daksh Kumar

Abstract

"Agriculture Forecast Module" project focuses on the development of a machine-learning model to predict crop prices, with a particular emphasis on wheat and maize. The project's approach encompasses data collection, data processing, exploratory data analysis, and model training. Initial exploratory data analysis indicated a lack of a strong linear correlation between annual rainfall and crop yield, prompting a shift towards the analysis of monthly rainfall data to understand the seasonality and temporal distribution of precipitation. The project aims to uncover the nuanced relationship between rainfall and crop yield, considering the timing of precipitation during various growth stages. This refined analysis is expected to enhance the precision of crop price predictions, offering valuable insights into the agricultural sector's dynamics.

Index

Introduction:

The agricultural sector, a crucial component of the global economy, plays a pivotal role in ensuring food security and providing essential goods to the world's population. However, this sector faces significant challenges due to its vulnerability to price fluctuations, which can have a substantial impact on farmers' incomes and livelihoods. Accurate crop price prediction can empower farmers with valuable insights to make informed decisions regarding crop selection and sales strategies, ultimately reducing risk exposure and maximizing profits.

Agriculture produce is subjected to various risks, which are not only confined to production risk pertaining to weather, pest but also the demand and supply of various countries, other policy, and economic factors. With restricted knowledge to understand and comprehend the information, farmers can incur huge losses by selling their produce in distress. Farmers no longer must contend with just local markets. They also must account for competition from the world over. High price volatility has been a major concern in past few years both for farmers and consumers.

Higher price volatility has driven the search for reliable and accurate price forecasting techniques for agricultural commodities. The main purpose of price prediction is to help producers manage their price risk and take informed decisions. Machine Learning has proved to be better than the traditional time series method of price prediction, using many linear and non-linear forecasting models.

Machine Learning based price prediction provides a unique way of combining technical and fundamental analysis methods. While technical analysis solely looks at historical price, fundamental analysis consists of understanding external and internal factors that influence the prices of a certain commodity. Individually technical analysis can be useful for providing accurate short-term prediction while fundamental analysis can help in long term forecast. By combining the two, higher accuracy in predictions can be achieved.

Objectives

The objective of the crop price prediction module based on machine learning is to help farmers make better decisions about what to grow and when to sell. By predicting future crop prices, farmers can reduce their risk exposure and maximize their profits.

Specifically, the objectives of the proposed module are:

- To develop a machine learning model that can accurately predict future crop prices.
- To evaluate the performance of the proposed model on a dataset of historical crop prices.
- To identify the key factors that influence crop prices.
- To develop a user-friendly interface for farmers to interact with the proposed model.

The proposed module can be used by farmers to make a variety of decisions, such as:

- What crops to grow?
- When to start planting and harvesting
- When to sell their crops
- How much to invest in fertilizers and other inputs

The proposed module can also be used by agricultural policymakers and researchers to better understand the factors that influence crop prices and develop strategies to support farmers.

The overall objective of the proposed module is to contribute to the development of a more resilient and sustainable agricultural sector.

Additional objectives that could be included in the project report:

- To compare the performance of the proposed module to other crop price prediction models.
- To investigate the potential benefits of using the proposed module to support farmers in developing countries.
- To explore the ethical implications of using machine learning to predict crop prices.

Literature Survey

Title: Time Series Forecasting in Crop Production: A Review

Introduction:

Crop production is a vital component of global agriculture and a key driver of food security and economic stability. Accurate forecasting of crop yields is essential for governments, policymakers, and agricultural stakeholders to make informed decisions and manage resources effectively. Over the years, various organizations and research institutions have undertaken efforts to develop sophisticated time series forecasting models to predict crop production levels. In this literature survey, we explore recent advancements in this field, with a specific focus on the crop yield forecast provided by the Organization for Economic Co-operation and Development (OECD).

OECD Crop Production Forecast:

One notable endeavor in the domain of crop yield forecasting is the initiative undertaken by the OECD. The OECD is a prominent international organization known for its extensive research and analysis in various sectors, including agriculture. They have developed an interactive dashboard to visualize and forecast crop production, offering valuable insights into global agricultural trends. The dashboard provides a comprehensive platform for monitoring and predicting crop yields, enabling users to access historical data and make informed decisions based on forecasting models.

Interactive Dashboard: The OECD's interactive dashboard, available at OECD Crop Production Dashboard, serves as a valuable resource for agricultural professionals, policymakers, and researchers. It offers real-time access to a wide range of data related to crop production, including historical records and forecasts. The interactive features of the dashboard allow users to customize data visualization, explore time series data, and perform in-depth analyses of crop production trends.

Advancements in Time Series Forecasting:

The OECD's crop production forecasting project aligns with the broader advancements in time series forecasting techniques applied to agriculture. In recent years, the field of agricultural forecasting has witnessed significant developments driven by the integration of machine learning, artificial intelligence, and remote sensing technologies.

Researchers are increasingly utilizing historical data, weather patterns, soil quality, and other relevant variables to build accurate and reliable forecasting models.

Challenges and Future Directions:

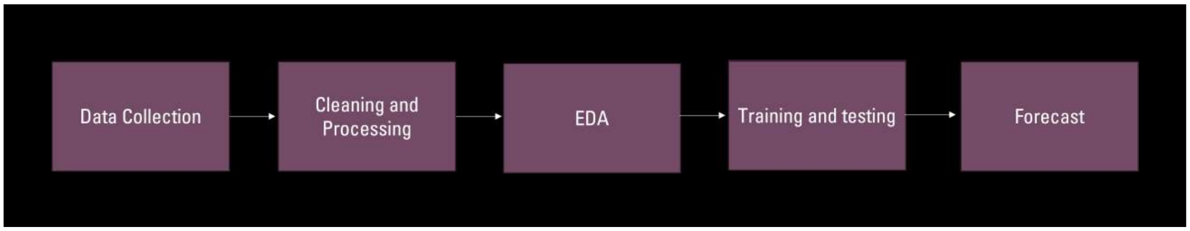
While the OECD's crop production forecasting project and similar initiatives represent significant progress in the field of agricultural forecasting, challenges remain. Accurate forecasting depends on various factors, including data quality, model selection, and the dynamic nature of environmental conditions. As such, continuous improvement in data collection, model refinement, and incorporating emerging technologies is necessary to enhance the precision of crop yield predictions.

Conclusion:

In conclusion, time series forecasting of crop production is an essential area of research with far-reaching implications for global agriculture and food security. The OECD's interactive dashboard, along with the broader advancements in forecasting techniques, plays a pivotal role in providing stakeholders with valuable tools to navigate the complex landscape of crop yield predictions. As technology and data collection methodologies continue to evolve, we anticipate that crop production forecasts will become even more accurate and informative, contributing to better-informed decision-making in the agricultural sector.

Project approach

A general approach for machine learning projects has been adopted for this project. The following are the phases of the project.



In our project, we embarked on the development of a machine-learning model aimed at predicting the prices of specific crops. This endeavour involved a well-structured approach consisting of the following key steps:

1. Data Collection:

The first and fundamental step in our project was to gather a comprehensive dataset that encompasses historical information on crop prices. We collected data from various reliable sources, including government agencies, agricultural organizations, and market databases. This step was crucial to ensure that our model had access to diverse and representative information.

2. Data Processing and Cleaning:

The raw data obtained during the collection phase was subjected to thorough data processing and cleaning. This included handling missing values, correcting data anomalies, and standardizing data formats. By ensuring data quality, we prepared a robust foundation for our machine-learning model.

3. Exploratory Data Analysis (EDA):

EDA played a pivotal role in our project. This step involved the exploration of the dataset to gain insights into the underlying patterns, relationships, and characteristics of the crop price data. Through visualizations, statistical summaries, and data profiling, we identified significant variables and trends that would guide our model development.

4. Training and Testing:

The heart of our project lay in developing a machine-learning model that could effectively predict crop prices. This phase entailed selecting appropriate machine-learning algorithms, dividing the dataset into training and testing subsets, and training the model. We fine-tuned model hyperparameters, evaluated its performance using metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE), and iteratively refined the model for optimal predictive accuracy.

5. Forecast:

The goal of our project was to make accurate predictions of crop prices. Once the model was trained and validated, we deployed it for forecasting future crop prices. This phase involved inputting relevant features such as weather data, market conditions, and other influencing factors to obtain predictions for specific crops' prices.

Throughout this project, our approach was underpinned by a commitment to accuracy, reliability, and transparency. We leveraged the power of machine learning to process and analyse extensive datasets, providing a valuable tool for farmers, traders, and policymakers to make informed decisions related to crop prices. Our journey from data collection to forecasting represents a significant milestone in the application of data science and machine learning within the agricultural domain.

This approach reflects our dedication to solving real-world challenges using cutting-edge technology, and we believe it has the potential to make a substantial impact on the agricultural sector.

Data collection:

The data for the crop price prediction module can be collected from a variety of sources, such as:

- Government agencies
- Agricultural organizations
- Publicly available databases

The major source of data has been the government website i.e., https://data.gov.in

The data includes the following features:

- Historical crop prices
- Weather data (e.g., rainfall, temperature, humidity)

Datasets used:

We have used various data sources to collect different data. That include the historic weather data, annual average temperature of India, annual yield of certain crops, and average price of the crops for a certain year. The major source of our data is government Database.

Following is the description of the data collected and processed.

1. Rainfall Data- State-wise Monthly Aggregated to Annual Total

Description:

The rainfall dataset employed in this project serves as a fundamental component in understanding the dynamics of crop price prediction. It contains state-wise monthly rainfall data across India, which has been thoughtfully aggregated to compute the total annual rainfall for each state and year. This dataset is a comprehensive record of the climatic conditions, representing one of the primary factors affecting crop yield and subsequently crop prices.

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	JF	MAM	JJAS	OND
0	Andaman & Nicobar Islands	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
1	Assam & Meghalaya	1901	27.1	19.5	30.6	223.0	207.0	524.9	430.6	464.1	291.4	163.7	115.6	1.2	2498.6	46.6	460.5	1710.9	280.5
2	Naga Mani Mizo Tripura	1901	11.7	18.1	29.4	206.2	124.0	443.3	331.4	466.0	304.1	166.7	67.4	0.0	2168.4	29.8	359.6	1544.8	234.2
3	Sub Himalayan West Bengal & Sikkim	1901	26.5	14.8	14.1	29.2	195.5	488.4	524.8	501.1	242.7	55.5	17.9	2.6	2113.2	41.3	238.9	1757.0	76.1
4	Gangetic West Bengal	1901	37.1	58.4	3.9	64.1	121.7	198.0	280.8	275.7	313.5	51.1	83.4	0.0	1487.6	95.5	189.7	1068.0	134.4

Dataset Head

Data Source:

The rainfall data was sourced from reputable meteorological authorities and institutions, ensuring its reliability and accuracy. The dataset provides an annual snapshot of the precipitation levels in different regions of India, which is crucial for assessing how varying weather patterns influence crop production. We found this data on data.gov.in.

Aggregation Process:

To provide a holistic view of the annual rainfall patterns, the dataset underwent aggregation to calculate the total annual rainfall for each state. This transformation ensures that the dataset is aligned with the annual timescale of crop yield data, facilitating meaningful analysis and correlation.

	SUBDIVISION	YEAR	ANNUAL
0	Andaman & Nicobar Islands	1901	3373.2
1	Assam & Meghalaya	1901	2498.6
2	Naga Mani Mizo Tripura	1901	2168.4
3	Sub Himalayan West Bengal & Sikkim	1901	2113.2
4	Gangetic West Bengal	1901	1487.6

Annual Rainfall (State wise)

1	Row Labels	Sum of ANNUAL Rainfall
2	1901	44947.5
3	1902	46805.6
4	1903	47596.9
5	1904	43287
6	1905	41516.2
7	1906	47180.2
8	1907	45443.9
9	1908	46453.7

Aggregated annual rainfall across India.

2. Annual Crop yield data

Description:

The annual crop yield dataset comprises information on the yield of specific crops in different regions of India. The yield data for various crops is recorded on an annual basis, offering insights into the productivity of agricultural regions across the country. This dataset is a key factor in assessing the impact of rainfall on crop production.

	Area	Item	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp	Rainfall Per Year	KG/HA
0	India	Cassava	1990	205381	1083	75000.0	25.58	58114.6	20538
176	India	Cassava	1991	218925	1083	72133.0	25.85	50407.9	21892
352	India	Cassava	1992	232463	1083	70791.0	25.69	47706.7	23246
528	India	Cassava	1993	230430	1083	66388.0	25.88	51514.3	23043
704	India	Cassava	1994	245277	1083	61357.0	25.75	54246.3	24527

Original Dataset

Mapping with Rainfall Data:

One of the crucial aspects of this project is the mapping of annual crop yield data with the aggregated annual rainfall data. This mapping allows us to examine the relationship between rainfall patterns and crop yield for specific crops. By aligning the two datasets, we can explore how variations in annual rainfall influence the agricultural output of different regions, providing valuable insights into the factors affecting crop prices.

	Year	Item	Rainfall Per Year	KG/HA
0	1990	Cassava	58114.6	20538
176	1991	Cassava	50407.9	21892
352	1992	Cassava	47706.7	23246
528	1993	Cassava	51514.3	23043
704	1994	Cassava	54246.3	24527

Data Mapped with rainfall.

These datasets collectively form the backbone of our project, enabling us to build a machine-learning model that predicts crop prices based on historical rainfall and crop yield data. By utilizing this integrated dataset, we aim to uncover meaningful patterns and trends that can enhance our understanding of the complex interplay between climate conditions and crop production, ultimately leading to more accurate crop price forecasts.

Exploratory data analytics

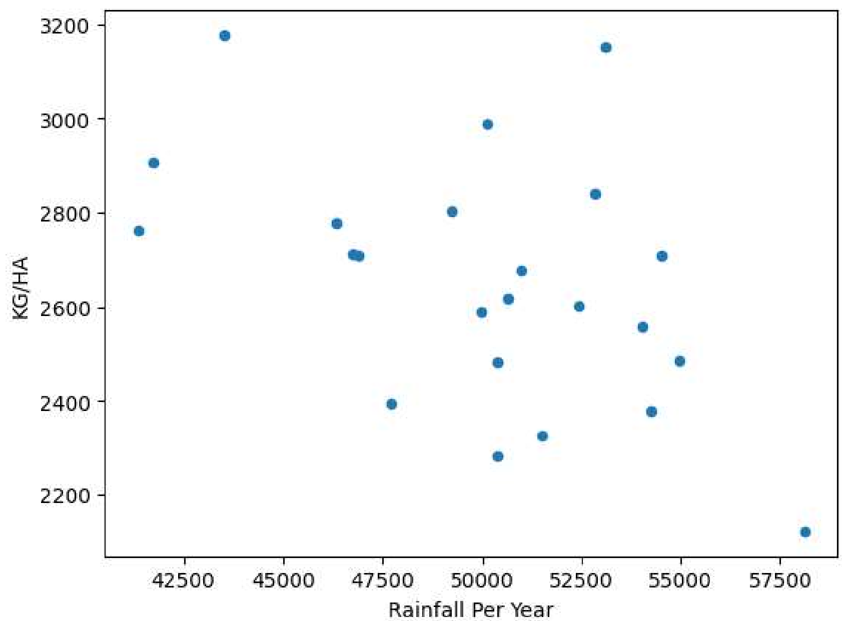
Introduction:

In our pursuit of understanding the complex relationship between crop yield and climatic conditions, we initially conducted an exploratory data analysis (EDA) focusing on two key crops: wheat and maize. The primary objective was to assess the influence of annual rainfall on crop yield. However, the preliminary analysis did not reveal a strong correlation between these two variables. As a result, we recognized the necessity for a more in-depth investigation, specifically delving into the influence of monthly rainfall patterns.

Annual Rainfall vs. Crop Yield Analysis:

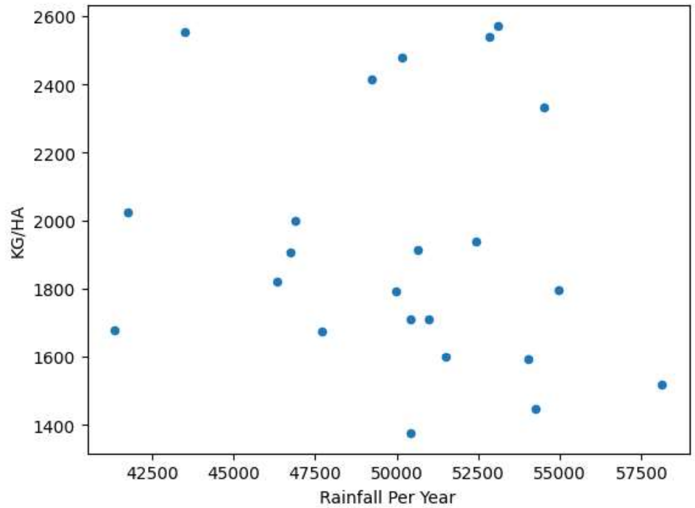
1. Annual Rainfall and Wheat Yield:

In our initial analysis, we examined the annual rainfall data and its correlation with wheat yield. We observed a lack of a substantial linear relationship between annual rainfall and wheat yield. While this may seem counterintuitive given the common belief that precipitation is a vital factor in crop production, our results indicated that other variables or complex interactions might be at play.



2. Annual Rainfall and Maize Yield:

Similarly, for maize, the correlation between annual rainfall and crop yield was not sufficiently strong to draw significant conclusions. This discrepancy between our expectations and findings raised questions about the underlying factors that affect crop production, highlighting the need for a more granular analysis.



The Need for Monthly Rainfall Analysis:

Given the inconclusive nature of the annual rainfall versus crop yield analysis, we recognized the importance of examining monthly rainfall data. Agriculture, being highly dependent on seasonal variations, necessitates a more detailed investigation into the temporal distribution of rainfall. Monthly data can provide valuable insights into the timing and duration of precipitation, factors that may significantly impact crop growth and yield.

Conclusion:

In conclusion, our initial exploratory data analysis on wheat and maize crop yield versus annual rainfall revealed a lack of strong linear correlation. This result prompted us to shift our focus towards a more comprehensive examination of monthly rainfall data. We believe that the key to understanding the impact of rainfall on crop yield lies in the seasonality and temporal distribution of precipitation. The timing of rainfall during planting, growth, and harvesting stages is likely to be a critical factor affecting crop production.

Our decision to delve deeper into monthly rainfall data represents a strategic shift in our analytical approach. By doing so, we aim to uncover the nuances and season-specific relationships between rainfall and crop yield, which may not be evident when considering annual averages alone. This refined analysis will enhance the accuracy of our crop price prediction model, as it aligns more closely with the real-world dynamics of agricultural practices and climatic conditions.