**University of South Wales**

**Prifysgol De Cymru**

**Faculty of Computing, Engineering and Science**

**MSc Project**

**Enhancing Crop Yield and Resource Efficiency with Machine Learning: Smart Agricultural Solutions**

**Student Name: KAUSHIKKUMAR PATEL**

**Student ID: 30086811**

**Supervisor**: **Ian Wilson**

**Year Of Study: 2023-2024**

**Course of Study: MASTER IN COMPUTER SCIENCE**

**Topic Area: Data Science & Machine Learning**

**Keywords:** *Agriculture, Machine Learning, Crop Management, Resource Optimization, Data-driven Decision Making, Precision Farming, Crop Yield Prediction, Climate Adaptation, Sustainable Practices.*

**Declaration:**

University of South Wales

Prifysgol De Cymru

Faculty of Computing, Engineering and Science

**STATEMENT OF ORIGINALITY**

This is to certify that, except where specific reference is made, the work described in this project is the result of the investigation carried out by the student, and that neither this project nor any part of it has been presented, or is currently being submitted in candidature for any award other than in part for the MSc award, Faculty of Computing, Engineering and Science from the University of South Wales.

Signed. **KAUSHIKKUMAR PATEL**

**ABSTRACT**

Crop prediction is a critical component of agricultural research and practice, facilitating informed decision-making, resource allocation, and food security initiatives. This study provides a comprehensive review of the evolution of crop prediction methodologies, starting from traditional statistical approaches to the adoption of advanced machine learning techniques. Traditional methods relied heavily on historical data, regression analysis, and time series modeling to forecast crop yields. While effective to a certain extent, these approaches often faced limitations in capturing complex relationships inherent in agricultural systems.

In recent years, the integration of machine learning algorithms has revolutionized crop prediction, offering more accurate and data-driven forecasts based on extensive datasets. Among these algorithms, support vector machines (SVM), decision trees, and random forests have emerged as prominent tools. Support vector machines are particularly adept at handling high-dimensional data and nonlinear relationships, making them well-suited for crop prediction tasks. Decision trees and random forests provide additional advantages, including interpretability and the ability to handle both categorical and continuous variables.

Furthermore, advancements in technology, such as remote sensing, satellite imagery, and IoT sensors, have contributed to the refinement of crop prediction models by providing real-time data on agricultural conditions. The integration of these technologies enhances the accuracy and timeliness of crop forecasts, enabling farmers and policymakers to make more informed decisions.

This study underscores the importance of accurate crop prediction in agricultural management and highlights the potential of machine learning techniques, with a specific emphasis on random forests, in improving productivity, sustainability, and resilience in agricultural systems. By leveraging these advanced methodologies, stakeholders can better address challenges such as climate change, resource scarcity, and food insecurity, ultimately contributing to the advancement of global agriculture.

**ACKNOWLEDGEMENT**

I am profoundly grateful for the time, effort, and support that has contributed to the successful completion of this project. My deepest thanks go to **Ian Wilson,** whose insightful guidance and consistent supervision were instrumental in shaping this work. Their valuable feedback and encouragement were greatly appreciated.

I would also like to express my gratitude to my friends and colleagues for their continuous support, engaging discussions, and helpful suggestions, which enriched the project.

My sincere thanks to **the University of South Wales** for offering the necessary resources and fostering an environment conducive to research and development. Additionally, I would like to extend my heartfelt appreciation to **Ian Wilson,** Head of Department, for their encouragement and cooperation throughout this endeavor.

**Table of Contents**

1. Introduction1

1.1 Objectives and Research Gap2

**2. Literature Review5**

**3. Data and Research Methodology16**

3.1 Descriptive Statistics of the Dataset18

3.1.1 Comparison and Analysis with Crop Specific Data19

3.1.2 Comparative Analysis of Crop Nutrient and Environmental Requirements19

3.2 Coffee Cultivation: Optimal Conditions Based on Data Analysis20

3.3 Papaya Cultivation: Optimal Conditions Based on Data Analysis21

3.4 Mungbean Cultivation: Optimal Conditions Based on Data Analysis22

3.5 Maize Cultivation: Optimal Conditions Based on Data Analysis23

3.6 Grapes Cultivation: Optimal Conditions Based on Data Analysis25

**4. Empirical Results27**

### 4.1 Logistic Regression Model Performance in Agricultural Optimization27

### 4.2 Decision Tree Model Performance in Agricultural Optimization28

### 4.3 Random Forest Model Performance in Agricultural Optimization30

### 4.4 Support Vector Machine (SVM) Performance in Agricultural Optimization31

4.5 Comparative Evaluation of Machine Learning Models for Agricultural Crop Classification: A Focus on Precision, Recall, and Overall Performance33

**5. Conclusion35**

**List of Tables**

Fig 3.1: Dataset Overview16

Fig 3.2: Nitrogen Frequency17

Fig 3.3: Potassium Frequency17

Fig 3.3: Potassium Frequency18

Fig 3.5: Descriptive statistics18

Fig 4.1: Confusion matrix of Logistic Regression27

Fig 4.2: Classification of Logistic Regression28

Fig 4.3: Confusion Matrix for Decision tree29

Fig 4.4: Classification report for Decision tree29

Fig 4.5: Confusion matrix of Random Forest30

Fig 4.6: Classification Report of Random Forest31

Fig 4.7: Confusion matrix of SVM32

Fig 4.8: Classification Report of SVM32

Fig 4.9: Performance of all Algorithm34

**1. Introduction**

The growing global demand for food presents significant challenges for the agricultural sector, necessitating innovative approaches to enhance crop yield without extensively expanding cropland areas. Achieving this balance requires a deep understanding of the various factors influencing crop yield, such as climate change, weather variability, soil conditions, seed genetics, and crop management practices. This overview delves into the importance of adopting best management practices alongside the application of machine learning techniques in crop prediction to achieve sustainable agricultural outcomes.

Crop yield variability is driven by the complex interactions among genetics, environment, and management practices (G×E×M), with factors like soil type, weather conditions, and seed genetics playing pivotal roles. Understanding these variables is crucial for developing effective agricultural strategies, as variations in soil type, for instance, can either mitigate or amplify the impacts of weather and climate on crop yields.

One proposed method for increasing food production without expanding cropland is to reduce variability in crop yields by identifying best management practices. Researchers typically use replicated field experiments or multi-year-site performance trials to evaluate the impact of various factors on crop yield, aiming to establish causal relationships and determine optimal practices. These studies, though often constrained by costs and logistics, seek to identify strategies that farmers can adopt to enhance yield and sustainability.

Advancements in machine learning (ML) have brought a revolution in crop prediction, offering more accurate and data-driven forecasts based on extensive datasets. Traditional statistical methods, like regression analysis and time series modeling, often fall short in capturing the complex relationships within agricultural systems. In contrast, ML algorithms such as support vector machines (SVM), decision trees, and random forests provide greater accuracy and adaptability in predicting crop yields. These techniques are particularly suited to handle high-dimensional data and nonlinear relationships, which are common in complex agricultural environments. Moreover, technological advancements such as remote sensing, satellite imagery have further improved crop prediction models by delivering real-time agricultural data, enabling more timely and accurate forecasts that aid in decision-making for farmers and policymakers.

By integrating best management practices with machine learning-based crop prediction, agricultural stakeholders are better equipped to address challenges such as climate change, resource scarcity, and food insecurity. This dual approach not only optimizes crop yield but also fosters sustainability and resilience within agricultural systems, ultimately contributing to more reliable and efficient agriculture that meets the growing food demands of the global population.

**1.1 Objectives and Research Gap**

Below section describes how a comparative analysis of the various research work aligns with the insights and the gap identified in the literature review, addressing the potential research gaps and problems:

**Objective 1: Implement the Random Forest classifier to analyze agricultural data.**

**Related Literature Insights:**

Syed (2024): Demonstrated 99% accuracy in crop yield prediction using ensemble learning techniques.

Ravisha & Sinha (2023): Found Random Forest to be highly effective in crop prediction.

Malhotra & Firdaus (2022): Highlighted the use AI in gathering and analyzing environmental data for agriculture.

**Research Gaps and Problems Addressed:**

**Insight:** The objective can build on Syed’s high-accuracy model by further refining the Random Forest algorithm to specifically analyze the interplay between weather patterns, soil conditions, and crop genetics.

Gap: Many studies focus on individual factors (like weather or soil) but not on their combined effect. This approach can integrate multiple data sources for a more holistic prediction model.

**Objective 2: Utilize the Random Forest model to optimize resource utilization.**

**Related Literature Insights:**

Bhattacharya & Pandey (2023): Discussed precision agriculture and resource management using IoT and ML.

Nagesh et al. (2024): Highlighted precision agriculture for optimizing resource use and boosting productivity.

**Research Gaps and Problems Addressed:**

**Insight:** Current literature shows the effectiveness of precision agriculture in reducing waste and optimizing resources.

**Gap:** Few studies provide a comprehensive model that combines all resource usage aspects (water, fertilizer, pesticide). This work can fill this gap by developing a unified predictive model.

**Objective 3: Employ the Random Forest classifier for crop health monitoring and disease detection.**

**Related Literature Insights:**

Durai & Shamili (2022): Used ResNet152V2 for weed and pest identification.

Balducci et al. (2018): Explored ML techniques for predicting environmental conditions and crop diseases.

**Research Gaps and Problems Addressed:**

Insight: The objective can leverage and expand upon existing image recognition techniques for more accurate and quicker disease detection.

**Gap:** Combining Random Forest with image recognition for real-time disease detection is underexplored. This integration could significantly improve early intervention strategies.

**Objective 4: Streamline agricultural operations using the Random Forest algorithm.**

**Related Literature Insights:**

Abdullahi et al. (2024): Demonstrated real-time data collection and processing for modernizing farming practices.

Ravisha & Sinha (2023): Focused on automating data processing and crop recommendations.

Research Gaps and Problems Addressed:

**Insight:** Automation in agricultural tasks can significantly reduce labor costs and improve efficiency.

**Gap:** Existing studies lack comprehensive automation frameworks. This research can develop detailed automation protocols integrating Random Forest for various agricultural operations.

**Objective 5: Enhance predictive capabilities for weather and climate.**

**Related Literature Insights:**

Bhattacharya & Pandey (2023): Highlighted the importance of climate predictions in agriculture.

Gnanadura et al. (2024): Demonstrated high accuracy in predicting crop yields using IoT-enabled predictive analytics.

**Research Gaps and Problems Addressed:**

**Insight:** Accurate weather prediction models are crucial for effective agricultural planning.

**Gap:** Many studies separate weather prediction from other agricultural predictions. The objective to integrate these predictions into a single model could provide a more cohesive strategy for farmers.

**Objective 6: Promote sustainable agricultural practices.**

**Related Literature Insights:**

Gnanadura et al. (2024): Emphasized the potential of IoT and ML in promoting sustainable practices.

ELBASI et al. (2022): Discussed the challenges of environmental sustainability in AI applications.

**Insight:** Sustainable practices are vital for long-term agricultural productivity.

**Gap:** There is a need for more targeted models that specifically address sustainability. This research can contribute by developing Random Forest models that focus on optimizing resource use to reduce environmental impact.

**Objective 7: Utilize the Random Forest algorithm for market analysis and crop price prediction.**

**Related Literature Insights:**

Ramesh et al. (2023): Discussed the importance of market demand analysis for crop selection.

Abdullahi et al. (2024): Highlighted economic impacts of improved agricultural practices.

**Research Gaps and Problems Addressed:**

**Insight:** Predictive models for market trends can help farmers make more profitable decisions.

**Gap:** Many existing studies focus on yield and resource optimization without considering market factors. The objective to integrate market analysis can provide a more comprehensive decision-making tool for farmers.

**2. Literature Review**

The following sections includes the work of the various authors in the field of the crop yield and the efficiency of the resources with the help of various machine learning algorithms as well as the comparative analysis of the various work based on the different parameters such as the algorithm used, advantages & disadvantages, as we as the application in the field. This article is in a developing phase and more content will be added as the research progress. The next section includes the exploration of the dataset and the attributes that will be used for the study.

The research paper titled "**Smart Agriculture using Ensemble Machine Learning Techniques,**" authored by Liyakathunisa Syed (2024) from Taibah University's Department of Computer Science in Madinah, Saudi Arabia, was presented at the International Conference on Machine Learning and Data Engineering (ICMLDE 2023) and published in Procedia Computer Science.

The study tackles various agricultural challenges, such as irregular irrigation, soil erosion, unregulated seed planting, adverse weather conditions, and locust outbreaks, which contribute to inefficiencies and high costs in traditional farming. To address these issues, the paper proposes a smart agriculture system that leverages Artificial Intelligence (AI) technologies. This system aims to enhance crop yields and sustainability by optimizing resource use and providing real-time data to support decision-making processes.

At the heart of this research is a novel ensemble-based machine learning classification method designed to predict crop yields with an impressive accuracy of 99% across 22 different crop categories. The method employs a two-tiered prediction system: base classifiers—Logistic Regression (LR), Classification and Regression Trees (CART), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)—generate input features for a level-1 meta-classifier, Random Forest, which then categorizes the crops.

The paper includes a comprehensive literature review on smart agriculture, particularly focusing on the application of machine learning in this field. The authors elaborate on the problem formulation, describe the proposed methodology, and present experimental results that highlight the superior performance of the ensemble machine learning technique compared to individual classifiers. The model's efficacy is validated using a dataset with features including temperature, humidity, pH, rainfall, and NPK (nitrogen, phosphorus, potassium) fertilizer values.

In summary, the research introduces an advanced system that aids farmers in selecting suitable crops for their fields by considering various environmental factors. The proposed ensemble machine learning technique significantly outperforms other models, accurately predicting crop yields and thus promoting sustainable agriculture practices and better yields at reduced costs. This paper, reviewed by the scientific committee of the ICMLDE and published under an open access license, demonstrates the potential for AI and IoT to revolutionize agriculture.

**Rashmi Gera and Anupriya Jain** (2024) explores the fusion of machine learning technologies in agriculture to build predictive models for crop yields. Utilizing a dataset from Kaggle that includes data on 22 different crops, along with climate and soil conditions, the study aims to create precise models that can help farmers make informed decisions, optimize productivity, and support sustainable agricultural practices.

The literature review section examines various studies related to smart agriculture, focusing on the application of data mining, and machine learning for predicting crop yields and managing resources. In the methodology section, the researchers outline their approach to data collection, preprocessing, feature selection, model training, and evaluation. This data was then analyzed using various machine learning algorithms.

The results and analysis section compares the effectiveness of multiple machine learning algorithms, including Logistic Regression, Random Forest, SVM, KNN Classifier, and XGBoost Classifier, using metrics such as accuracy, precision, recall, and F1-score. The findings show that the Random Forest and XGBoost Classifier models achieved high accuracy, demonstrating their capability to effectively predict crop yields.

The conclusion underscores the significant role of machine learning algorithms in smart agriculture, with potential applications in precision farming, risk assessment, market forecasting, and sustainable practices. The article suggests that future research should focus on continuous advancements in real-time monitoring, big data analytics, predicting the impacts of climate change, and farm automation.

**Ramesh, et al.,** (2023) address the challenges faced by Indian farmers, such as the unpredictability that leads to reduced agricultural productivity, and emphasize the potential of machine learning techniques to provide predictions and recommendations for crop selection, rotation, and resource management. The study suggests that by analyzing climatic factors and crop data, more efficient strategies can be developed to enhance farmers' productivity and management practices.

The introduction outlines the transformation of agriculture in India due to globalization and the advent of precision agriculture, which includes site-specific farming practices. The paper reviews various machine learning models for predicting crop yields, including clustering methods like k-means and k-means++. It also discusses the importance of soil analysis and the application of data mining techniques.

The authors propose a system designed to recommend crops based on productivity and seasonality. This system aims to gather and preprocess agricultural data from recent decades, train a machine learning model, and evaluate its performance. The goal is to assist farmers in making informed decisions by analyzing trends in crop production and market demand.

The traditional method of soil testing is critiqued for its inefficiency, often leading to uniform nutrient application and potential environmental harm. The problem statement highlights the necessity for a machine learning-based crop recommendation system to help farmers choose suitable crops based on soil requirements.

The proposed system architecture encompasses data collection, preparation, model selection, analysis, prediction, and accuracy testing. The authors choose a decision tree algorithm, a supervised learning method suitable for both classification and regression tasks. The system achieves an accuracy of 90.7% on the test data.

The paper concludes that the integration of modern technologies and machine learning techniques can provide farmers with personalized and relevant crop recommendations, thereby improving productivity. The system's high accuracy underscores its potential to significantly enhance agricultural management and decision-making.

**Bhattacharya & Pandey,** (2023) introduce an integrated decision-support system (DSS) aimed at boosting crop yield by leveraging machine learning and sensor data in agriculture. This study underscores the vital role of data in contemporary agriculture, particularly in addressing global food demand, promoting environmental sustainability, and managing resources efficiently.

The proposed DSS is structured to gather, organize, and assess diverse agricultural data types, including soil nutrient profiles, crop characteristics, and environmental conditions, thereby aiding farmers in making informed decisions. By applying machine learning algorithms, the system predicts optimal crop varieties, allocates fertilizers effectively, and forecasts rainfall patterns, thus refining agricultural practices.

The paper emphasizes the importance of precision agriculture and smart farming methodologies that use the Internet of Things (IoT) and artificial intelligence to enhance efficiency and reduce resource consumption. The potential applications of machine learning in agriculture are highlighted, such as predicting crop diseases, estimating yields, and analyzing soil quality.

Additionally, the study addresses the negative impacts of chemical fertilizers on both the environment and human health, advocating for balanced nutrient management. A data-driven approach for predicting crop yields is proposed, incorporating factors like fertilizer types, crop varieties, and rainfall forecasts.

The paper reviews existing literature demonstrating the effectiveness of machine learning in agricultural contexts, including decision-support tools, irrigation planning models, and weed detection systems. The authors also present a DSS prototype that integrates various sensors and employs machine learning classifiers like Decision Trees, Naïve Bayes, and Logistic Regression to offer recommendations on fertilizer use and crop selection.

In conclusion, the integration of machine learning and data mining techniques within agricultural DSSs can greatly enhance the accuracy of crop yield and rainfall predictions, fostering sustainable and efficient farming practices. This research contributes to the broader objective of addressing food security and environmental challenges through advanced agricultural technologies.

**Ravisha & Sinha,** (2023) presents a Crop Recommendation System that integrates a soil database and expert recommendations to consider soil quality parameters such as Nitrogen, Phosphorous, Potassium, pH, and weather-related factors like rainfall, temperature, and humidity. This system aims to identify the most suitable crops for specific regions, enhancing agricultural productivity.

Using datasets from Kaggle, the research applies machine learning algorithms to analyze data, focusing on determining optimal crops based on key soil and weather attributes. Among the algorithms used, Random Forest and Naive Bayes demonstrated comparable accuracy scores of 99.09%, while XGBoost surpassed them with an accuracy of 99.31%.

The paper highlights the necessity of incorporating modern technological approaches into agriculture for economic improvement and introduces a user-friendly system capable of processing extensive datasets to offer precise crop management recommendations. It emphasizes the advantages of precision agriculture, especially for developing countries like India, where traditional farming practices are common.

The study's objectives include developing a crop recommendation model that accounts for pH, rainfall, humidity, and temperature to improve agricultural yields through highly accurate predictions. This model aids farmers in reducing errors in crop selection by leveraging machine learning's ability to autonomously predict crop suitability based on prior data.

The methodology outlines the application of six algorithms for crop prediction: Decision Tree, Naive Bayes, Support Vector Machine, Logistic Regression, Random Forest, and XGBoost. Experimental results reveal that XGBoost is the most effective for the crop recommendation system.

The research proposes a robust crop recommendation system that helps farmers make well-informed decisions regarding crop selection, potentially boosting productivity and profitability in India’s agricultural sector. It demonstrates that machine learning, particularly XGBoost, can significantly enhance crop recommendation accuracy and agricultural practices.

**ELBASI, et al.,** (2022) **categorize** artificial intelligence (AI) applications in agriculture into three main areas: soil and crop monitoring, predictive analytics, and agricultural robotics. They examine the use of sensors and soil sampling by farmers, which provide essential data for farm management systems. The review outlines both the advantages and challenges of AI in smart farming, comparing various methodologies, including machine learning, expert systems, and image processing.

Additionally, the paper addresses the United Nations' prediction of significant global population growth, which requires a corresponding rise in food production. The authors suggest a framework for applying AI to improve crop disease detection, achieving an accuracy of 99.96%.

The systematic review methodology explains how the authors searched for, reviewed, and categorized current research on AI and IoT in agriculture, focusing on high-quality sources from IEEE Xplore, Clarivate, and Scopus. Papers were classified based on their benefits, challenges, and methodologies.

The background section on artificial intelligence discusses the impact of machine learning in agriculture, the development of sensors, and the role of the Internet of Things (IoT) in smart farming. It also covers wireless sensor networks, IoT protocols, cloud computing, and expert systems in agriculture.

The paper identifies challenges in implementing AI in agriculture, such as data collection and environmental sustainability. It suggests that AI technologies can address these issues by offering real-time monitoring and intelligent prediction systems.

Image processing techniques in smart farming are explored, including image enhancement, restoration, compression, and smoothing. The review also discusses computer vision algorithms for tasks like image classification, object detection, and segmentation.

The review highlights AI's potential to transform agriculture, making it more efficient and sustainable. It advocates for further research and the adoption of AI technologies by farmers, especially in developing countries, to enhance agricultural practices and achieve sustainability.

Durai & Shamili, (2022) uses datasets from Kaggle to train ML classification algorithms, with Random Forest Classifier achieving the highest accuracy after hyperparameter tuning. The weed identification module utilizes a pre-trained ResNet152V2 model for classifying weed images and recommending herbicides based on the identified weeds. Similarly, the pest identification module uses ResNet152V2 to classify insect images and suggests appropriate pesticides. The cost estimation module forecasts the costs of cultivation using Indian cost of cultivation survey data from 2010-2018, applying ensemble regression algorithms to predict costs until 2028. The results and discussion section showcases the web interface of the application, demonstrating the crop recommendation system's functionality by providing an example of input parameters and the resulting crop recommendations, including scientific names, soil types, and crop categories.

Malhotra & Firdaus, (Malhotra & Firdaus, 2022)explores the integration of Artificial Intelligence (AI) to enhance crop yield prediction in agriculture. which AI algorithms then process to provide a comprehensive view of the agricultural environment. AI's multifaceted role includes anomaly detection to identify potential threats like pests, diseases, and adverse weather conditions, enabling timely interventions. Predictive analytics uses historical and real-time data to forecast crop yields, aiding in decision-making for irrigation and fertilization. AI-powered image recognition detects early signs of pests and diseases, while resource optimization ensures efficient use of water and fertilizers, reducing waste and environmental impact. Additionally, AI-driven decision support systems offer personalized recommendations for planting schedules and crop rotations, maximizing yields. Autonomous farming integrates AI with machinery for precision tasks, and secure communication protocols protect agricultural data from cyber threats.

The article emphasizes the transformative potential of AI in agriculture, promising increased productivity, sustainability, and smart farming practices. It suggests that by capitalizing on AI's strengths in data integration, anomaly detection, predictive analytics, and autonomous farming, the industry can address challenges in food security and environmental stewardship. The importance of data collection and integration is underscored, with AI algorithms processing data for informed decision-making. Anomaly detection is critical for enabling swift identification of irregularities and proactive responses to protect crop productivity. Predictive analytics allows AI models to use historical and real-time data to predict crop yields and optimize cultivation practices. AI-powered pest and disease management reduces the need for chemical treatments, promoting sustainable practices. Resource optimization ensures precise allocation of resources, contributing to environmental sustainability and food security. Decision support systems foster a collaborative ecosystem that benefits all farmers, while autonomous farming and secure communication protocols enhance efficiency and data protection. In conclusion, the article underscores the pivotal role of AI in enhancing agricultural practices, from data-driven decision-making to sustainable resource management, contributing to a more resilient and productive agricultural sector.

Pukrongta, et al., (2024) introduces the PEnsemble 4 model, a machine learning-based ensemble prediction model designed to improve maize yield forecasting. By integrating data from unmanned aerial vehicle (UAV) imagery and , the model captures a comprehensive range of environmental and crop-specific parameters. The study underscores the importance of accurate yield prediction amidst rising global maize demand and the crop's susceptibility to weather conditions. The PEnsemble 4 model addresses this challenge by analyzing temporal patterns in vegetation indices such as CIre and NDRE, which indicate canopy density and plant height. Demonstrating an impressive 91% accuracy rate, the model advances yield prediction from the reproductive stage (R6) to the blister stage (R2), enabling earlier estimates and improved decision-making in farming operations.

The research also explores the model's broader agricultural applications, including the detection of water and crop stress, as well as disease monitoring. By combining machine learning technologies, the PEnsemble 4 model offers a novel solution for maize yield prediction, potentially revolutionizing crop management and contributing to sustainable farming practices. Conducted in the Manorom District, Chai Nat, Thailand, the study collected data from UAV flights across three plots. Multispectral images captured by UAVs and environmental data from were used to predict maize growth and yield with the help of 14 machine learning algorithms. The ensemble model PEnsemble 4 outperformed individual machine learning algorithms, offering a robust approach for predicting maize yields. The paper is structured into sections detailing the study area, UAV data acquisition, image processing, environmental measurements, machine learning methods, and model selection. In conclusion, the PEnsemble 4 model marks a significant advancement in precision agriculture, providing accurate and timely maize yield predictions to enhance crop management, improve productivity, and support sustainable agricultural practices.

Balducci, et al., (Balducci, et al., 2018) explores the management and utilization of heterogeneous agricultural datasets to enhance smart farming practices through machine learning techniques. By leveraging models such as decision trees, K-nearest neighbors, neural networks, and polynomial predictive models, the research aims to forecast crop harvests, reconstruct missing or erroneous sensor data, and predict environmental conditions for disease prevention. Utilizing datasets from industry, scientific research, and national statistical institutes, the study demonstrates the potential for significant innovation in the agricultural sector through advanced data analysis. The findings highlight the effectiveness of machine learning in improving data management and analysis, suggesting that investment in technology and skilled workforce can lead to sustainable and optimized agricultural practices.

Gnanadura, et al., (2024) analyses the transformative impact of IoT-enabled predictive analytics and machine learning on precision agriculture, highlighting the optimization of farming practices, enhanced crop yields, and minimized environmental impact through real-time data from IoT sensors and advanced data analysis. Demonstrating a 95% accuracy rate in predicting crop yields, the study underscores the potential for sustainable agriculture through technology. Key findings include resource optimization by precisely targeting inputs like water, fertilizers, and pesticides to reduce waste and costs; accurate yield forecasting for strategic planting and harvesting; early pest and disease management to minimize environmental impact; livestock monitoring for early detection of health issues; and improved supply chain efficiency through optimized distribution routes. The integration of IoT and machine learning promotes data-driven decision-making, climate resilience, and sustainability by reducing resource waste and chemical use. Prototypes such as SAPS and ACIS showcase the potential for real-time data analysis and informed decision-making in agriculture. Ongoing research in edge computing, advanced sensors, and data privacy further supports the technology's potential to address global food security and environmental sustainability challenges.

Abdullahi, et al., (2024) compares Decision Trees (DT), K-nearest Neighbor (KNN), and Random Forest algorithms within the CRS, with DT proving the most accurate and interpretable, achieving a 99.2% accuracy rate. Highlighting agriculture's economic significance in Somalia—accounting for 75% of GDP and 93% of exports—the study addresses the sector's challenges, such as unpredictable weather and limited resources. The methodology includes IoT sensors for data collection and ML algorithms for processing, with a web application built using Django and MySQL for user-friendly farmer access. Demonstrating successful CRS implementation in Balcad District, the research underscores the benefits of real-time IoT data and the DT model for modernizing farming practices and improving productivity. The paper concludes with recommendations for future research to enhance system robustness and scalability, emphasizing the transformative potential of ML and IoT in agriculture, especially in challenging environments like Somalia.

Abdullahi, et al., (Abdullahi, et al., 2023) employs a comprehensive methodology encompassing data collection from IoT sensors, data preprocessing, and the training and testing of ML models (DT, KNN, and RF). Data from various farms in the Afgooye, Balcad, and Jowhar Districts, focusing on soil and environmental parameters influencing crop growth, was collected. The DT algorithm outperformed KNN and RF, achieving an accuracy of 99.2% along with balanced precision, recall, and F1-score metrics. The DT model was integrated into a user-friendly web application developed using the Django framework and MySQL database, providing real-time crop recommendations based on current environmental conditions. This study demonstrates the potential of integrating ML and IoT technologies to enhance agricultural practices in Somalia, with the DT algorithm's high accuracy and interpretability being valuable for guiding crop selection, thus contributing to improved food security and economic growth. Future research directions include expanding the dataset to encompass broader geographical areas and exploring additional ML algorithms for further optimization.

Nagesh et al. (2024) addresses the challenge of increasing agricultural productivity amid unfavorable weather conditions. The study proposes a machine learning model for accurately predicting crop yields using a dataset with 33 features, including yield, pesticide use, rainfall, and average temperature for crops like soybeans, maize, potato, rice, wheat, and sorghum, sourced from the FAO and the World Data Bank. The researchers employ the Grey Level Co-occurrence Matrix (GLCM) method for feature selection and utilize AdaBoost Decision Tree, Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN) algorithms, with the AdaBoost Decision Tree achieving superior performance with an accuracy and recall rate of 98% and 99%, respectively. The paper includes a literature review of existing data preprocessing and crop yield prediction techniques, such as decision tree algorithms (ID3, CART, C5.0), ensemble learning, and Bayesian classification, and details the methodology behind the proposed machine learning model. The results show that combining AdaBoost Decision Tree with GLCM feature selection significantly enhances crop yield prediction accuracy. The study concludes that precision agriculture, supported by advanced machine learning techniques, can effectively boost agricultural productivity and sustainability.

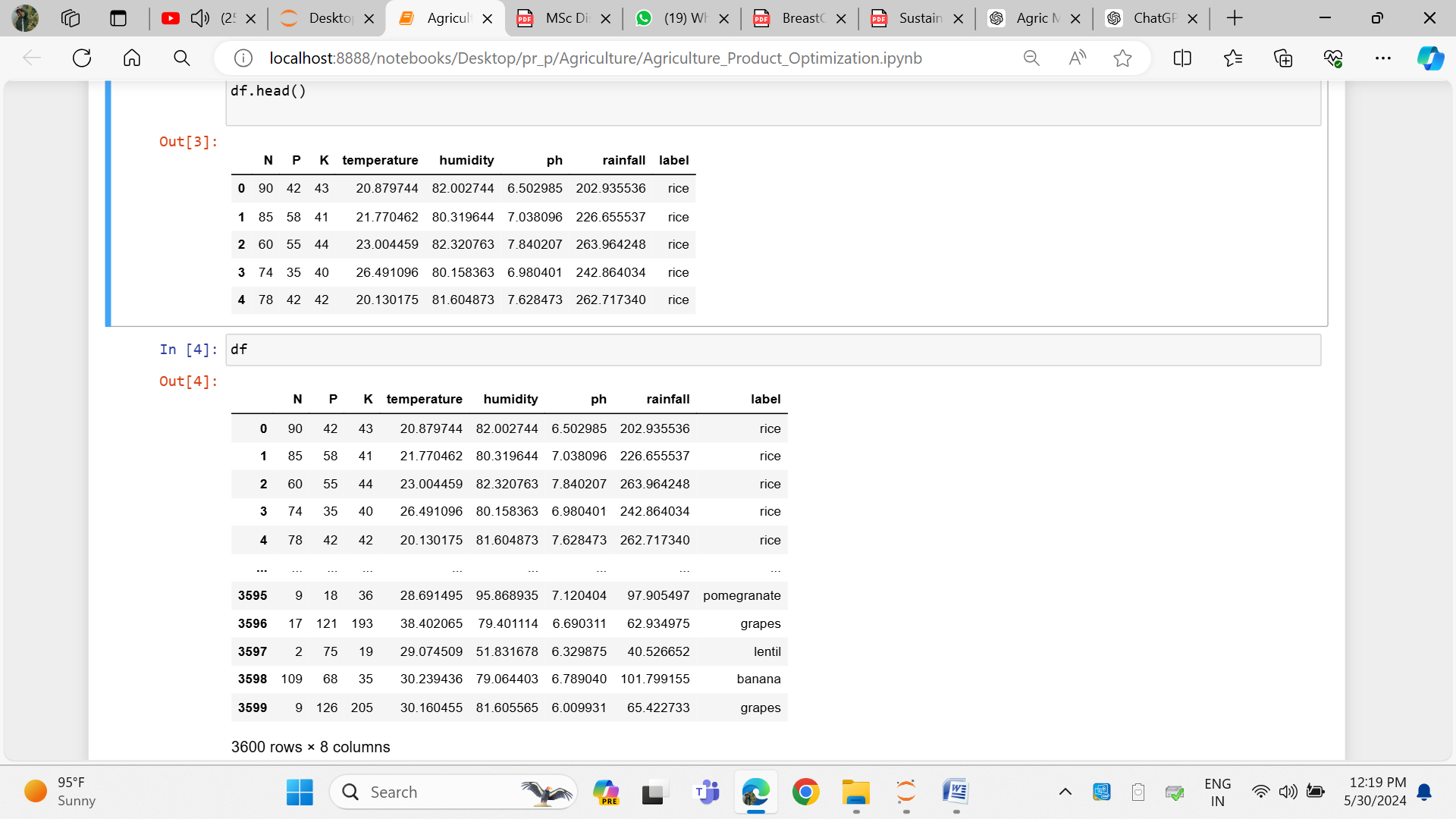
Vemuri, et al. (Vemuri, et al., 2023)focuses on the Smart Farming Revolution and the integration of the Internet of Things (IoT) in agriculture to enhance yield and sustainability. The study demonstrates how IoT devices, like sensors, can improve agricultural intelligence and yields, particularly in cotton fields, by collecting real-time data on irrigation, temperature, and humidity to automate irrigation and create data-driven products, leading to a 30-80% reduction in water consumption and a 40% reduction in fertilizers and pesticides, while improving production and profit per acre. The literature review highlights IoT's potential to transform agriculture, especially in countries like China and India, discussing the five-layer architecture of IoT for real-time farming activity monitoring and control. The authors emphasize precision farming's role in addressing global food security challenges, as the population is expected to reach 9.6 billion by 2050, with IoT sensors and drones tracking soil moisture, pH levels, nutrient content, crop growth, and cattle health for targeted interventions. The document details the model and dataset for cotton crops, with 149 data points measuring temperature, water, and pump operations, and presents results through graphs showing moisture levels and the relationship between temperature, moisture content, and pump activity. The article advocates for IoT adoption in agriculture to maximize water use efficiency, reduce chemical inputs, increase profitability, and ensure sustainable food production, suggesting that widespread use of these technologies could transform agriculture towards more sustainable, data-driven practices.

**Table 1: Comparative Analysis of the work**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Author(s)** | **Technique Used** | **Advantages** | **Disadvantages** | **Applications** | **Impact on Crop Yield** |
| Liyakathunisa Syed | Ensemble Machine Learning, IoT | High prediction accuracy (99%), real-time data collection, optimization of resource use | High complexity, requires extensive data collection and processing | Smart agriculture, crop yield prediction | Significant improvement in crop yield prediction accuracy, enhanced decision-making, reduced costs |
| Rashmi Gera & Anupriya Jain | Machine Learning, IoT | Precise models, effective comparison of multiple algorithms, high accuracy | Requires robust infrastructure for real-time data collection | Crop yield prediction, resource management | High accuracy in predicting crop yields, improved decision-making, optimized resource use |
| Ramesh, et al. | Decision Trees, Data Mining | High accuracy (90.7%), efficient crop recommendations | Requires historical and environmental data | Crop selection, resource management | Improved productivity and management practices, better decision-making |
| Bhattacharya & Pandey | Decision-Support System (DSS), Machine Learning | Enhanced crop yield, efficient resource management, sustainability | Requires integration of diverse data types, potential high costs | Crop variety prediction, fertilizer allocation, rainfall forecasting | Significant enhancement in crop management, optimized resource use, sustainable farming practices |
| Ravisha & Sinha | Random Forest, Naive Bayes, XGBoost | High accuracy (99.31% with XGBoost), user-friendly, precise crop recommendations | May require extensive computational resources for large datasets | Crop recommendation, soil quality analysis | Enhanced crop selection accuracy, improved productivity and profitability |
| ELBASI, et al. | Machine Learning, Image Processing, IoT | High accuracy in disease detection (99.96%), real-time monitoring | Challenges in data collection, potential high implementation costs | Crop disease detection, resource management | Increased efficiency in identifying crop diseases, reduced chemical use, sustainable practices |
| Durai & Shamili | Random Forest, ResNet152V2, Ensemble Regression | High accuracy in crop recommendations, effective cost estimation, robust weed and pest management | Requires large and diverse datasets, potential high computational costs | Weed and pest identification, crop cost estimation | Accurate crop recommendations, effective pest and weed management, optimized cultivation costs |
| Pukrongta, et al. | Ensemble Prediction Model (PEnsemble 4), UAV, IoT | High accuracy (91%), early yield prediction, comprehensive data integration | High implementation costs, requires UAV and IoT infrastructure | Maize yield forecasting, water and crop stress detection | Improved yield prediction accuracy, enhanced decision-making, early intervention in farming operations |
| Balducci, et al. | Decision Trees, K-Nearest Neighbors, Neural Networks | Improved data management, effective forecast of crop harvests | Requires significant data processing capabilities, potential high costs | Crop yield prediction, sensor data reconstruction, environmental condition prediction | Enhanced forecast accuracy, better resource management, improved agricultural practices |
| Gnanadura, et al. | Predictive Analytics, IoT, Machine Learning | High prediction accuracy (95%), optimized farming practices, real-time data analysis | Requires robust IoT infrastructure, potential high costs | Crop yield prediction, resource optimization, pest and disease management | Significant yield improvement, reduced environmental impact, efficient farming practices |
| Abdullahi, et al. | Decision Trees, K-Nearest Neighbor, Random Forest | High accuracy (99.2% with DT), user-friendly web application, real-time data collection | Limited to specific geographical areas, requires robust data collection infrastructure | Crop recommendation, real-time data analysis | Improved productivity, modernized farming practices, enhanced decision-making |
| Nagesh et al. | AdaBoost Decision Tree, ANN, KNN | High prediction accuracy (98%), effective feature selection, comprehensive dataset analysis | Requires significant data preprocessing, potential high computational costs | Crop yield prediction, resource management | Enhanced yield prediction accuracy, better resource allocation, improved agricultural productivity |
| Vemuri, et al. | IoT, Sensors | Significant reduction in water (30-80%) and chemical use (40%), real-time monitoring | High implementation costs, requires robust IoT infrastructure | Smart farming, irrigation automation, data-driven agriculture | Improved production and profit per acre, enhanced sustainability, optimized resource use |

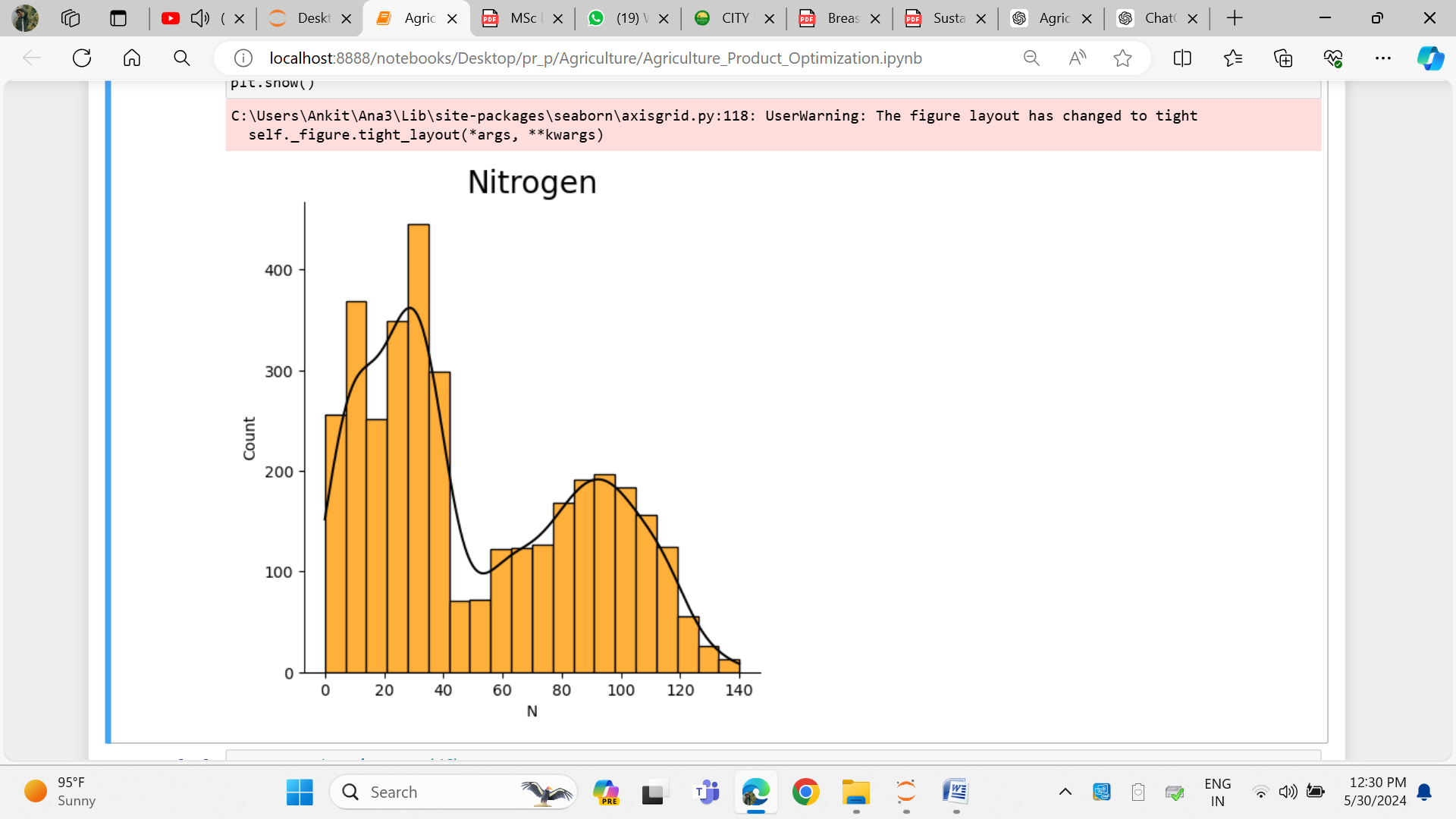
**3. Data and Research Methodology**

This dataset is designed for agricultural optimization using machine learning, with the goal of predicting the most suitable crops based on various soil and environmental factors. It contains 3600 rows and 8 columns, representing features such as nitrogen (N), phosphorus (P), and potassium (K) content in the soil, temperature, humidity, pH level, and rainfall. The target variable is the crop type, labeled with specific crops such as rice, pomegranate, grapes, lentil, and banana. This rich dataset allows for comprehensive analysis and insights into the optimal conditions for different crops. Research in this area typically focuses on leveraging such data to enhance crop yield, improve sustainability, and optimize resource usage. By understanding the relationships between these variables and crop suitability, farmers and agricultural experts can make data-driven decisions to enhance productivity. Studies have shown that integrating machine learning models in agriculture can significantly increase efficiency, reduce costs, and promote better resource management, ultimately contributing to food security and sustainable agricultural practices.

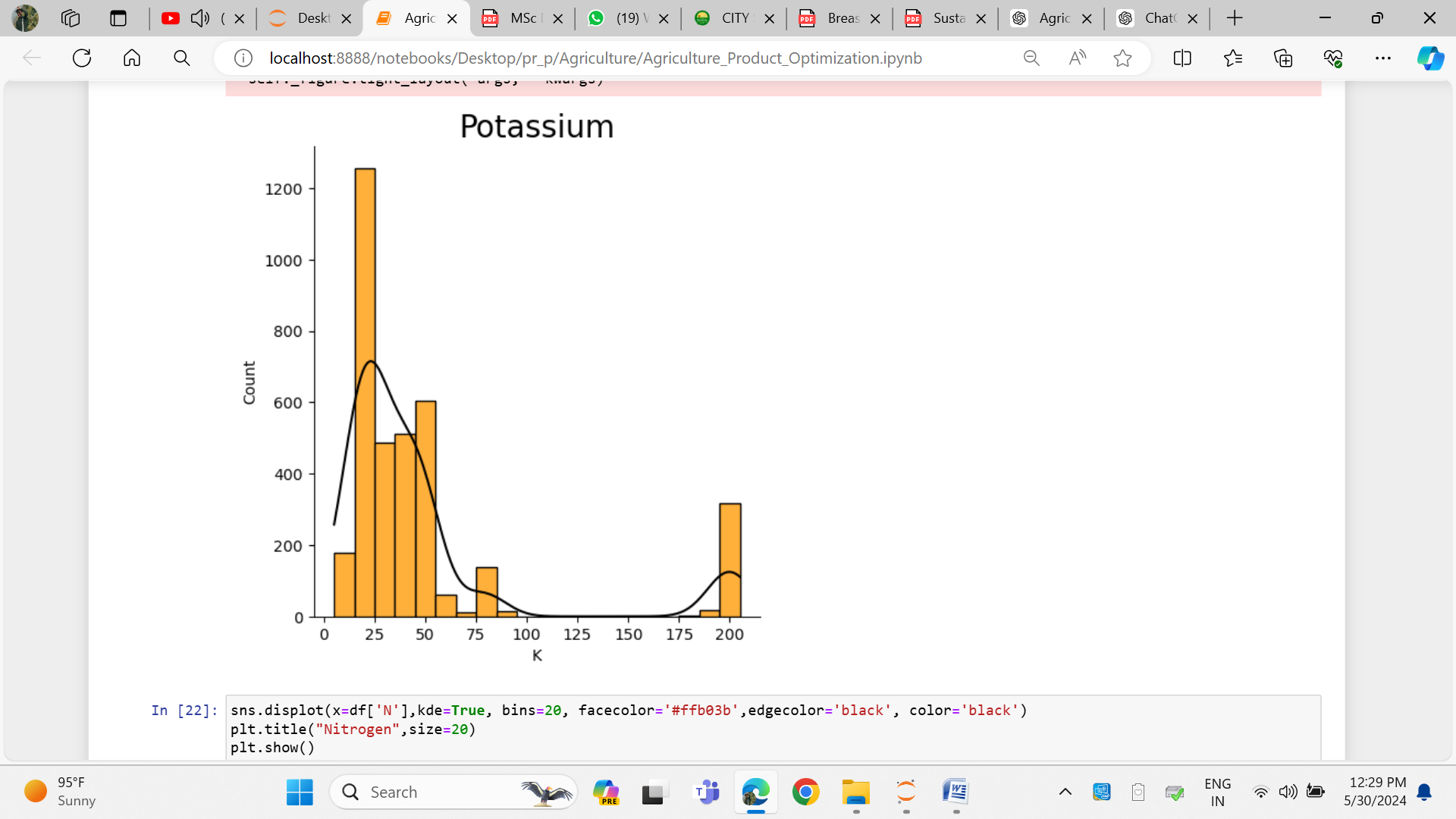


**Fig 3.1: Dataset Overview**

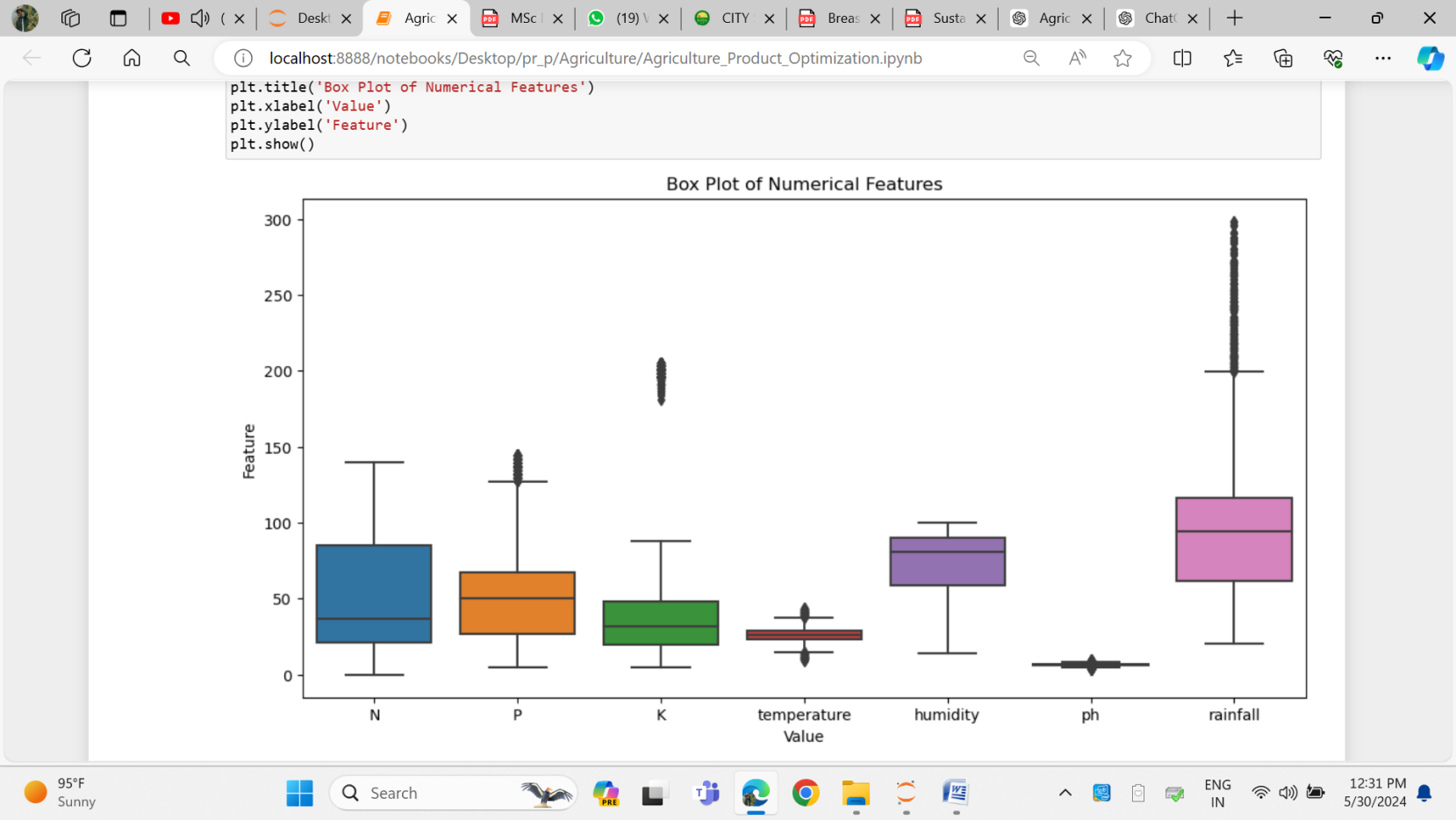
The dataset reveals key agricultural variables with the following characteristics: Nitrogen (N) levels range from 10 to 120, with an average of 95.56; Phosphorus (P) varies between 9 and 65, averaging 32.0; Potassium (K) spans from 15 to 55, with an average of 30.02. Temperature records show a minimum of 20.82°C, a maximum of 32.92°C, and an average of 26.43°C, while humidity ranges from 39.17% to 93.53%, averaging 58.11%. The pH levels fluctuate between 4.90 and 7.49, with an average of 6.68, and rainfall varies from 23.70mm to 228.62mm, with an average of 142.66mm.



**Fig 3.2: Nitrogen Frequency**

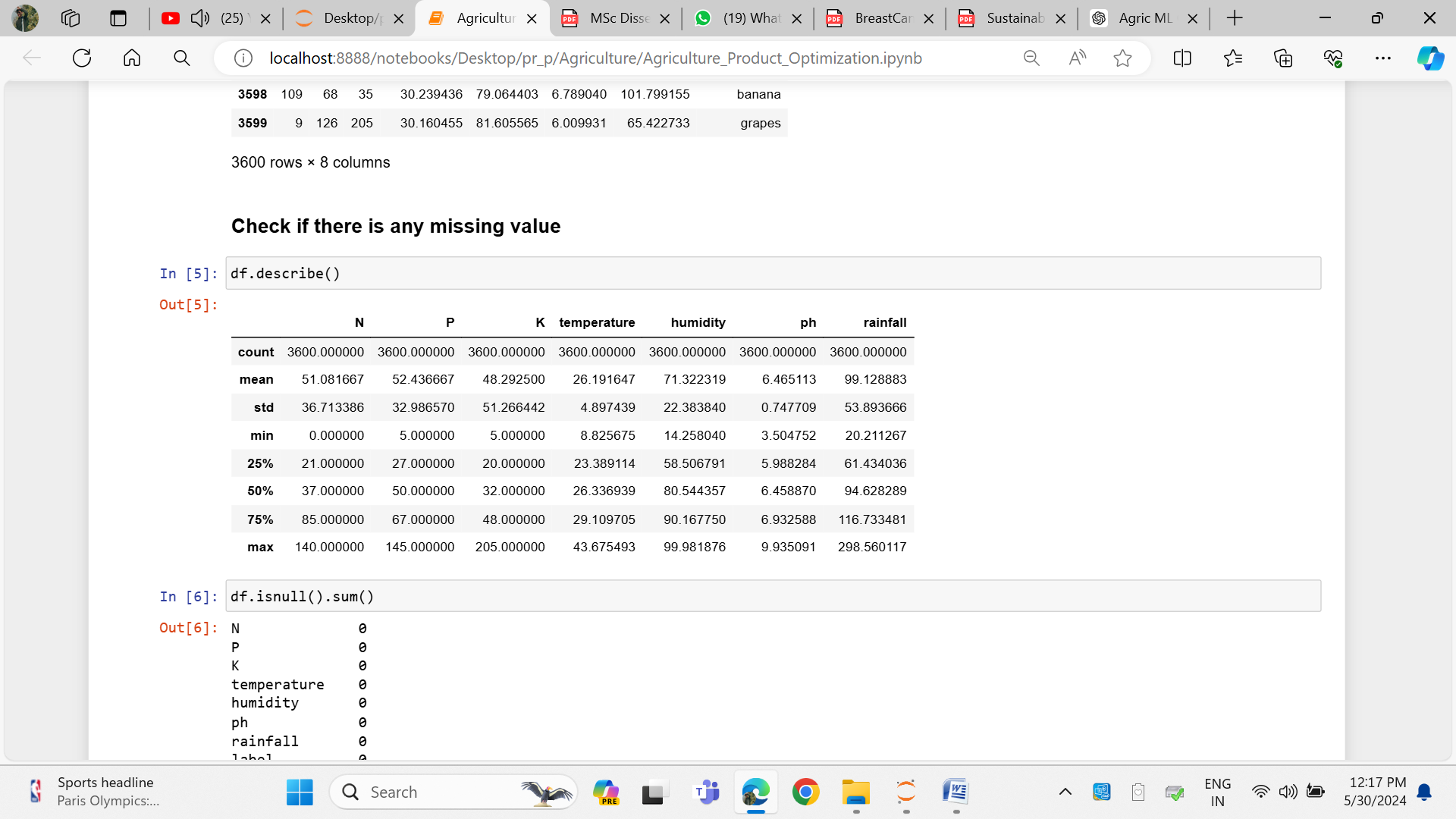


**Fig 3.3: Potassium Frequency**



**Fig 3.4: All Input feature of dataset**

**3.1 Descriptive Statistics of the Dataset**



**Fig 3.5: Descriptive statistics**

The dataset provides a detailed statistical summary of various agricultural parameters. The mean values for Nitrogen (N), Phosphorus (P), and Potassium (K) are 51.08, 52.44, and 48.29, respectively. Temperature averages at 26.19°C, with humidity at 71.32%, pH at 6.47, and rainfall at 99.13 mm. The standard deviations indicate substantial variability, especially for Potassium (K) at 51.27 and humidity at 22.38%. Minimum recorded values show Nitrogen (N) at 0, Phosphorus (P) at 5, and Potassium (K) at 5, with the lowest temperature at 8.83°C, humidity at 14.26%, pH at 3.50, and rainfall at 20.21 mm. The 25th percentile values for Nitrogen (N), Phosphorus (P), and Potassium (K) are 21, 27, and 20, respectively, while the temperature, humidity, pH, and rainfall are 23.39°C, 58.51%, 5.99, and 61.43 mm. The median (50th percentile) figures show Nitrogen (N) at 37, Phosphorus (P) at 50, Potassium (K) at 32, temperature at 26.34°C, humidity at 80.54%, pH at 6.46, and rainfall at 94.63 mm. At the 75th percentile, Nitrogen (N) is 85, Phosphorus (P) is 67, and Potassium (K) is 48, with temperature reaching 29.11°C, humidity at 90.17%, pH at 6.93, and rainfall at 116.73 mm. Maximum values recorded are 140 for Nitrogen (N), 145 for Phosphorus (P), and 205 for Potassium (K), with the highest temperature at 43.68°C, humidity at 99.98%, pH at 9.94, and rainfall peaking at 298.56 mm.

**3.1.1 Comparison and Analysis with Crop Specific Data**

The dataset reveals nuanced insights into the suitability of various environmental factors for crop cultivation, emphasizing the need for tailored agricultural practices. The overall mean nitrogen content of 51.08, while sufficient for crops like mungbean and grapes, falls short of the higher demands of crops such as banana, maize, coffee, and papaya, necessitating targeted nitrogen supplementation. Similarly, the mean phosphorus level of 52.44 aligns well with papaya and mungbean but requires careful adjustment for crops with higher phosphorus needs, particularly grapes. The mean potassium level of 48.29 is optimal for banana and papaya but is notably insufficient for potassium-intensive crops like grapes, underscoring the need for enhanced potassium application in such cases. Temperature management emerges as a critical factor, with the mean temperature of 26.19°C being favorable for banana, coffee, and mungbean, yet requiring adjustments for crops like maize, grapes, and particularly papaya, which thrives in warmer conditions. The average humidity of 71.32% suits crops such as banana, mungbean, and grapes but may necessitate modifications for maize, coffee, and papaya to optimize growth conditions. The pH level of 6.47 is broadly conducive to most crops in the dataset, reflecting a slight acidity to neutrality that benefits a wide range of plants. Lastly, the average rainfall of 99.13 mm, while sufficient for banana and moderate for maize and grapes, falls short for water-demanding crops like coffee and papaya, highlighting the importance of precise irrigation strategies. These findings suggest that while the dataset's environmental parameters are generally favorable, specific adjustments in nutrient levels, temperature, humidity, pH, and rainfall are crucial to optimizing conditions for diverse crops.

### **3.1.2 Comparative Analysis of Crop Nutrient and Environmental Requirements**

The analysis reveals distinct patterns and variations in the nutrient and environmental needs of different crops. For instance, banana and papaya share similar requirements for nitrogen, phosphorus, and potassium, and both thrive in high temperatures and humidity. In contrast, maize and mungbean, with lower nutrient demands, prefer more moderate conditions. Grapes and mungbean, while compatible in nitrogen and humidity levels, diverge significantly in potassium needs, with grapes requiring much higher levels. Grapes also stand out for their ability to tolerate a broad temperature range. Conversely, coffee and papaya exhibit substantial differences, particularly in phosphorus and rainfall needs, with coffee requiring much less than papaya. Lastly, the comparison between banana and mungbean highlights that banana demands higher nutrient levels and greater humidity, underscoring the need for crop-specific nutrient and environmental management.

**3.2 Coffee Cultivation: Optimal Conditions Based on Data Analysis**

Coffee is one of the most widely consumed beverages globally, and its cultivation is influenced by several critical soil and environmental factors. The provided dataset includes detailed statistics on nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH level, and rainfall, which are vital for optimizing coffee production. Here is a comprehensive analysis based on the data provided:

#### Nitrogen Requirements for Coffee

Nitrogen is essential for the vegetative growth of coffee plants, influencing the development of leaves and overall plant health. The data indicates that coffee thrives with a nitrogen content ranging from 10 to 120, with an average of 95.56. Maintaining nitrogen levels around this average can promote optimal growth and ensure the healthy development of coffee plants.

#### Phosphorus Needs for Root Development

Phosphorus plays a crucial role in root development and energy transfer within coffee plants. The dataset shows that coffee plants benefit from a phosphorus range of 9 to 65, with an average of 32.00. Ensuring sufficient phosphorus availability is vital to support robust root systems and overall plant vitality, leading to healthier and more productive coffee plants.

#### Potassium's Role in Coffee Quality

Potassium aids in various physiological functions, including water regulation and disease resistance. For coffee plants, the data indicates a potassium range of 15 to 55, with an average of 30.02. Adequate potassium levels are key to enhancing the quality and yield of coffee beans, making it a critical nutrient for coffee cultivation.

#### Ideal Temperature Range for Coffee Growth

Temperature is a critical factor in coffee cultivation, influencing growth rates and bean quality. The dataset reveals that coffee plants thrive in a temperature range between 20.82°C and 32.92°C, with an average of 26.43°C. This moderate temperature range is ideal for coffee growth, ensuring the proper maturation of beans and overall plant health.

#### Humidity Levels for Optimal Coffee Production

Humidity levels significantly impact coffee plants, particularly their ability to flower and set fruit. The data shows that coffee can grow in humidity levels ranging from 39.17% to 93.53%, with an average of 58.11%. Maintaining humidity around this average helps prevent diseases and supports healthy plant development, contributing to better coffee yields.

#### Soil pH for Nutrient Uptake in Coffee Plants

The pH level of the soil directly affects nutrient availability and uptake by coffee plants. According to the dataset, coffee plants prefer slightly acidic to neutral soils, with pH levels ranging from 4.90 to 7.49, and an average of 6.68. Maintaining soil pH close to this average can enhance nutrient availability and absorption, leading to healthier coffee plants.

#### Rainfall Requirements for Coffee Cultivation

Adequate rainfall is essential for coffee cultivation, influencing water availability for growth and bean development. The dataset indicates that coffee plants thrive with rainfall ranging from 23.70 mm to 228.62 mm, with an average of 142.66 mm. Ensuring sufficient and well-distributed rainfall throughout the growing season is crucial for optimal coffee production and high-quality beans.

**3.3 Papaya Cultivation: Optimal Conditions Based on Data Analysis**

Papaya, known for its sweet flavor and high nutritional value, requires specific soil and environmental conditions for optimal growth and yield. The provided dataset includes detailed statistics on nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH level, and rainfall, which are crucial for optimizing papaya production. Here is a comprehensive analysis based on the data provided:

#### Nitrogen Requirements for Papaya

Nitrogen is vital for the vegetative growth of papaya plants, impacting the development of leaves and overall plant health. The data reveals that papaya can thrive with a nitrogen content ranging from 0 to 115, with an average of 56.14. This indicates that while maintaining nitrogen levels around the average can support optimal growth, papaya plants are also capable of tolerating very low nitrogen levels, demonstrating their adaptability.

#### Phosphorus Needs for Root Development

Phosphorus plays a crucial role in root development and energy transfer within papaya plants. According to the dataset, papaya benefits from a phosphorus range of 15 to 75, with an average of 55.35. Ensuring sufficient phosphorus availability within this range is essential for fostering robust root systems and maintaining overall plant vitality.

#### Potassium's Role in Papaya Quality

Potassium supports various physiological functions in papaya plants, including water regulation and disease resistance. The data indicates a potassium range of 17 to 58, with an average of 48.35. Maintaining adequate potassium levels is key to enhancing the quality and yield of papaya fruits, making it a critical nutrient in papaya cultivation.

#### Ideal Temperature Range for Papaya Growth

Temperature is a crucial factor in papaya cultivation, influencing both growth rates and fruit quality. The dataset reveals that papaya thrives within a temperature range of 20.78°C to 43.68°C, with an average of 32.03°C. This broad temperature range highlights papaya's tolerance to both moderate and high temperatures, which is essential for its successful growth in tropical and subtropical regions.

#### Humidity Levels for Optimal Papaya Production

Humidity significantly impacts papaya plants, particularly their ability to flower and set fruit. The data shows that papaya can grow in humidity levels ranging from 48.66% to 95.64%, with an average of 91.82%. High humidity levels, especially around the average, are optimal for preventing diseases and supporting healthy plant development, leading to better yields.

#### Soil pH for Nutrient Uptake in Papaya Plants

The pH level of the soil directly affects nutrient availability and uptake by papaya plants. The dataset indicates that papaya plants prefer slightly acidic to neutral soils, with pH levels ranging from 4.85 to 7.62, and an average of 6.72. Maintaining soil pH close to this average can enhance nutrient availability and absorption, contributing to the healthy growth of papaya plants.

#### Rainfall Requirements for Papaya Cultivation

Adequate rainfall is essential for papaya cultivation, as it influences water availability for growth and fruit development. Papaya plants in the dataset receive rainfall ranging from 25.33 mm to 248.86 mm, with an average of 133.09 mm. Ensuring sufficient and well-distributed rainfall throughout the growing season is crucial for optimal papaya production, helping to achieve high-quality fruit yields.

**3.4 Mungbean Cultivation: Optimal Conditions Based on Data Analysis**

Mungbean, known for its high nutritional value and use in various culinary dishes, requires specific soil and environmental conditions for optimal growth and yield. The provided dataset includes detailed statistics on nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH level, and rainfall, which are crucial for optimizing mungbean production. Here is a comprehensive analysis based on the data provided:

#### Nitrogen Requirements for Mungbean Growth

Nitrogen plays a vital role in the vegetative growth of mungbean plants, influencing leaf development and overall plant health. The data suggests that mungbean can thrive with a nitrogen content ranging from 0 to 114, with an average of 28.05. While mungbean can grow in soils with low nitrogen levels, optimal growth is achieved when nitrogen levels are closer to the average, ensuring healthier and more robust plants.

#### Phosphorus Needs for Root Development in Mungbean

Phosphorus is crucial for root development and energy transfer within mungbean plants. The dataset indicates that mungbean requires a phosphorus range of 8 to 84, with an average of 46.74. Sufficient phosphorus availability within this range is essential for developing strong root systems and maintaining overall plant vitality, leading to better yields.

#### Potassium's Role in Enhancing Mungbean Quality

Potassium supports various physiological functions in mungbean plants, including water regulation and disease resistance. The data shows that mungbean thrives with potassium levels ranging from 13 to 39, with an average of 20.49. Maintaining adequate potassium levels is key to enhancing the quality and yield of mungbean, making it a crucial nutrient in mungbean cultivation.

#### Temperature Requirements for Mungbean Cultivation

Temperature is a critical factor in mungbean cultivation, influencing both growth rates and bean quality. According to the dataset, mungbean grows optimally within a temperature range of 20.96°C to 40.93°C, with an average of 28.63°C. This broad temperature range highlights mungbean's tolerance to both moderate and high temperatures, which is essential for its growth in various climatic conditions.

#### Optimal Humidity Levels for Mungbean Growth

Humidity levels significantly impact mungbean plants, particularly their ability to flower and set fruit. The data reveals that mungbean can grow in humidity levels ranging from 23.07% to 97.11%, with an average of 82.77%. Maintaining humidity around the average helps prevent diseases and supports healthy plant development, which is crucial for achieving optimal yields.

#### Soil pH and Nutrient Uptake in Mungbean

The pH level of the soil directly affects nutrient availability and uptake by mungbean plants. The dataset indicates that mungbean plants prefer slightly acidic to neutral soils, with pH levels ranging from 4.37 to 7.75, and an average of 6.68. Ensuring the soil pH is close to the average can enhance nutrient availability and absorption, promoting healthier mungbean growth.

#### Rainfall Requirements for Mungbean Cultivation

Adequate rainfall is essential for mungbean cultivation, as it influences water availability for growth and bean development. The data shows that mungbean plants receive rainfall ranging from 20.21 mm to 221.68 mm, with an average of 51.22 mm. Ensuring sufficient and well-distributed rainfall throughout the growing season is crucial for optimal mungbean production, leading to high-quality bean yields.

**3.5 Maize Cultivation: Optimal Conditions Based on Data Analysis**

Maize, also known as corn, is a staple crop cultivated worldwide for its versatility and high yield potential. The provided dataset includes detailed statistics on nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH level, and rainfall, which are crucial for optimizing maize production. Here is a comprehensive analysis based on the data provided:

#### Nitrogen Requirements for Maize Growth

Nitrogen is a key nutrient for the vegetative growth of maize plants, directly impacting leaf development and overall plant health. The data suggests that maize can thrive with a nitrogen content ranging from 6 to 113, with an average of 72.08. Maintaining nitrogen levels close to the average is essential for promoting optimal growth and maximizing maize yield, ensuring robust and healthy plants.

#### Phosphorus Needs for Root Development in Maize

Phosphorus plays a crucial role in root development and energy transfer within maize plants. The dataset indicates that maize requires phosphorus levels ranging from 13 to 75, with an average of 47.40. Adequate phosphorus availability within this range supports the development of strong root systems, which are vital for the overall vitality and productivity of maize plants.

#### Potassium's Role in Enhancing Maize Quality

Potassium is essential for various physiological functions in maize, including water regulation and resistance to diseases. The data shows that maize plants perform best with potassium levels ranging from 12 to 38, with an average of 19.45. Ensuring adequate potassium levels helps enhance the quality and yield of maize, contributing to healthier plants and better crop outcomes.

#### Temperature Requirements for Maize Cultivation

Temperature is a critical factor for maize cultivation, influencing both growth rates and kernel quality. According to the dataset, maize grows optimally within a temperature range of 17.50°C to 37.12°C, with an average of 24.19°C. This broad temperature range highlights maize's ability to adapt to various climatic conditions, making it suitable for cultivation in diverse environments.

#### Optimal Humidity Levels for Maize Growth

Humidity levels significantly impact maize plants, particularly their ability to pollinate and develop kernels. The data reveals that maize can grow in humidity levels ranging from 49.91% to 83.12%, with an average of 65.03%. Maintaining humidity close to the average helps prevent diseases and supports healthy plant development, ensuring a successful maize harvest.

#### Soil pH and Nutrient Uptake in Maize

The pH level of the soil directly affects nutrient availability and uptake by maize plants. The dataset indicates that maize plants prefer slightly acidic to neutral soils, with pH levels ranging from 4.41 to 7.31, and an average of 6.36. Ensuring the soil pH is near the average enhances nutrient availability and absorption, promoting healthier growth and higher maize yields.

#### Rainfall Requirements for Maize Cultivation

Adequate rainfall is essential for maize cultivation, as it influences water availability for growth and kernel development. The data shows that maize plants receive rainfall ranging from 28.71 mm to 183.63 mm, with an average of 85.74 mm. Ensuring sufficient and well-distributed rainfall throughout the growing season is crucial for optimal maize production, leading to high-quality and abundant maize yields.

**3.6 Grapes Cultivation: Optimal Conditions Based on Data Analysis**

Grapes are a significant fruit crop, valued for their use in fresh consumption, winemaking, and various culinary applications. The provided dataset includes detailed statistics on nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH level, and rainfall, which are crucial for optimizing grape production. Here is a comprehensive analysis based on the data provided:

Nitrogen is a vital nutrient for the vegetative growth of grapevines, playing a key role in the development of leaves and overall plant health. The data indicates that grapes thrive with a nitrogen content ranging from 0 to 89, with an average of 25.56. This suggests that maintaining nitrogen levels around this average can promote optimal growth and yield. Comparatively, grapes and mungbean have similar low nitrogen requirements, while other crops such as banana, maize, coffee, and papaya require significantly higher nitrogen levels.

Phosphorus is crucial for root development and energy transfer within grapevines. The dataset reveals a phosphorus range for grapes from 114 to 145, with an average of 130.97. This highlights the importance of ensuring sufficient phosphorus availability to support robust root systems and overall plant vitality. Among the crops compared, grapes have the highest phosphorus requirement, followed by banana, with maize, mungbean, papaya, and coffee needing lower amounts.

Potassium is essential for various physiological functions in grapevines, including water regulation and resistance to diseases. The data indicates that grape plants have a potassium range of 181 to 205, with an average of 199.62. Adequate potassium levels are necessary to enhance the quality and yield of grapes. Notably, grapes have significantly higher potassium needs compared to the other crops listed, which include banana, maize, coffee, papaya, and mungbean.

Temperature is a critical factor for grape cultivation, influencing growth rates and fruit quality. The optimal temperature range for grapes, as shown in the dataset, is between 8.83°C and 41.95°C, with an average of 23.89°C. This broad temperature range demonstrates the grapevines' tolerance to both moderate and high temperatures, which is essential for their growth in various climatic conditions. In comparison, coffee, mungbean, and maize have similar temperature requirements, while banana and papaya thrive in higher average temperatures.

Humidity levels significantly affect grapevines, particularly their ability to flower and set fruit. The data shows that grapes can grow in humidity levels ranging from 22.99% to 86.44%, with an average of 81.51%. Maintaining humidity around the average can help prevent diseases and support healthy plant development. Grapes, banana, and mungbean require high humidity, similar to papaya, whereas maize and coffee perform better in moderate humidity levels.

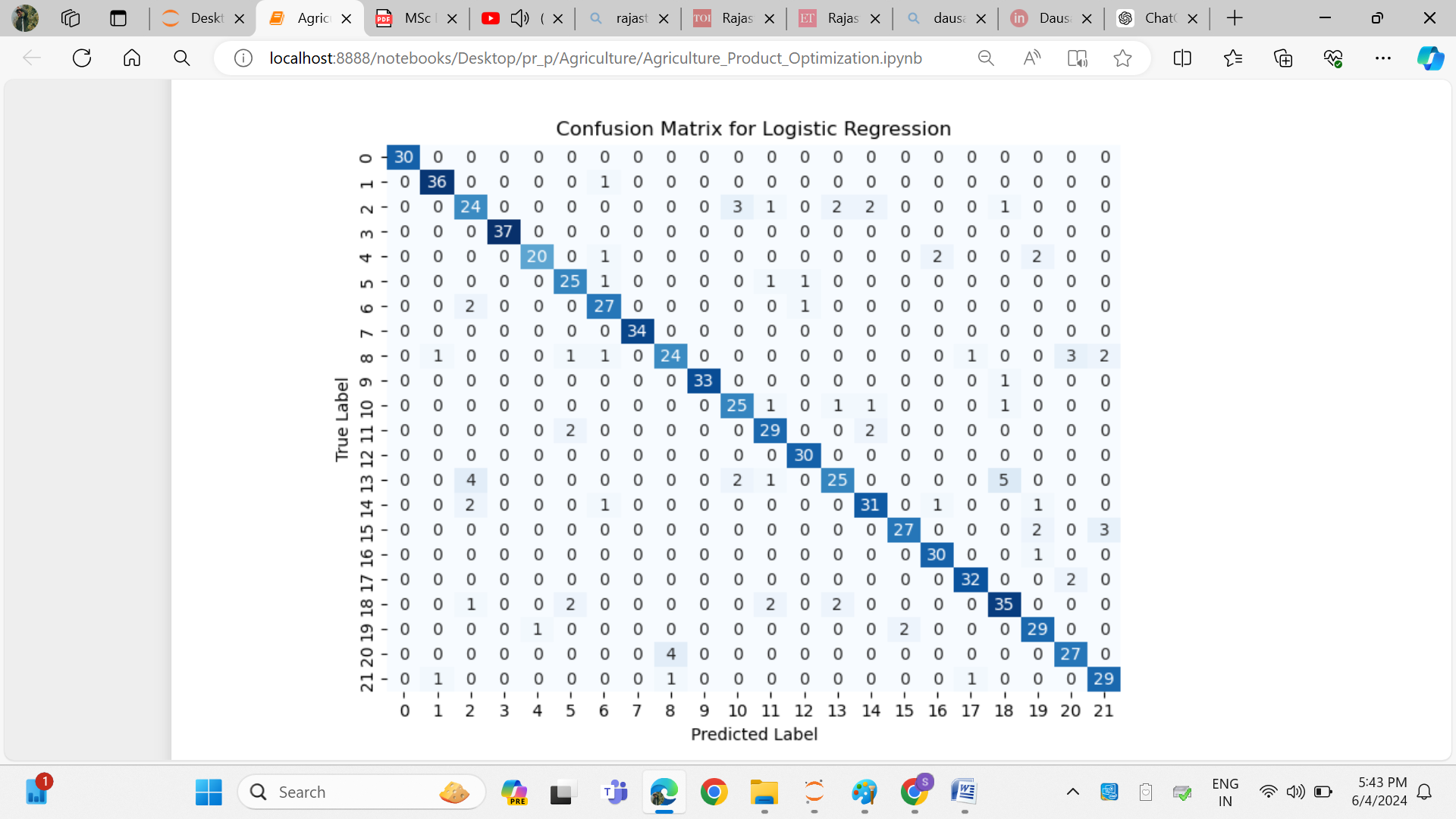
The pH level of the soil influences nutrient availability and uptake by grapevines. According to the dataset, grapevines prefer slightly acidic to neutral soils, with pH levels ranging from 3.92 to 7.76, and an average of 6.03. Ensuring that the soil pH is close to this average can enhance nutrient availability and absorption. All the crops compared, including grapes, prefer slightly acidic to neutral pH levels, with averages ranging from 6.03 to 6.72, indicating a strong similarity in pH preferences.

Adequate rainfall is essential for grape cultivation, as it affects water availability for growth and fruit development. The data reveals that grape plants receive rainfall ranging from 28.99 mm to 117.94 mm, with an average of 69.85 mm. Ensuring sufficient and well-distributed rainfall throughout the growing season is crucial for optimal grape production. Grapes require moderate rainfall, similar to mungbean, while banana, maize, coffee, and papaya demand more.

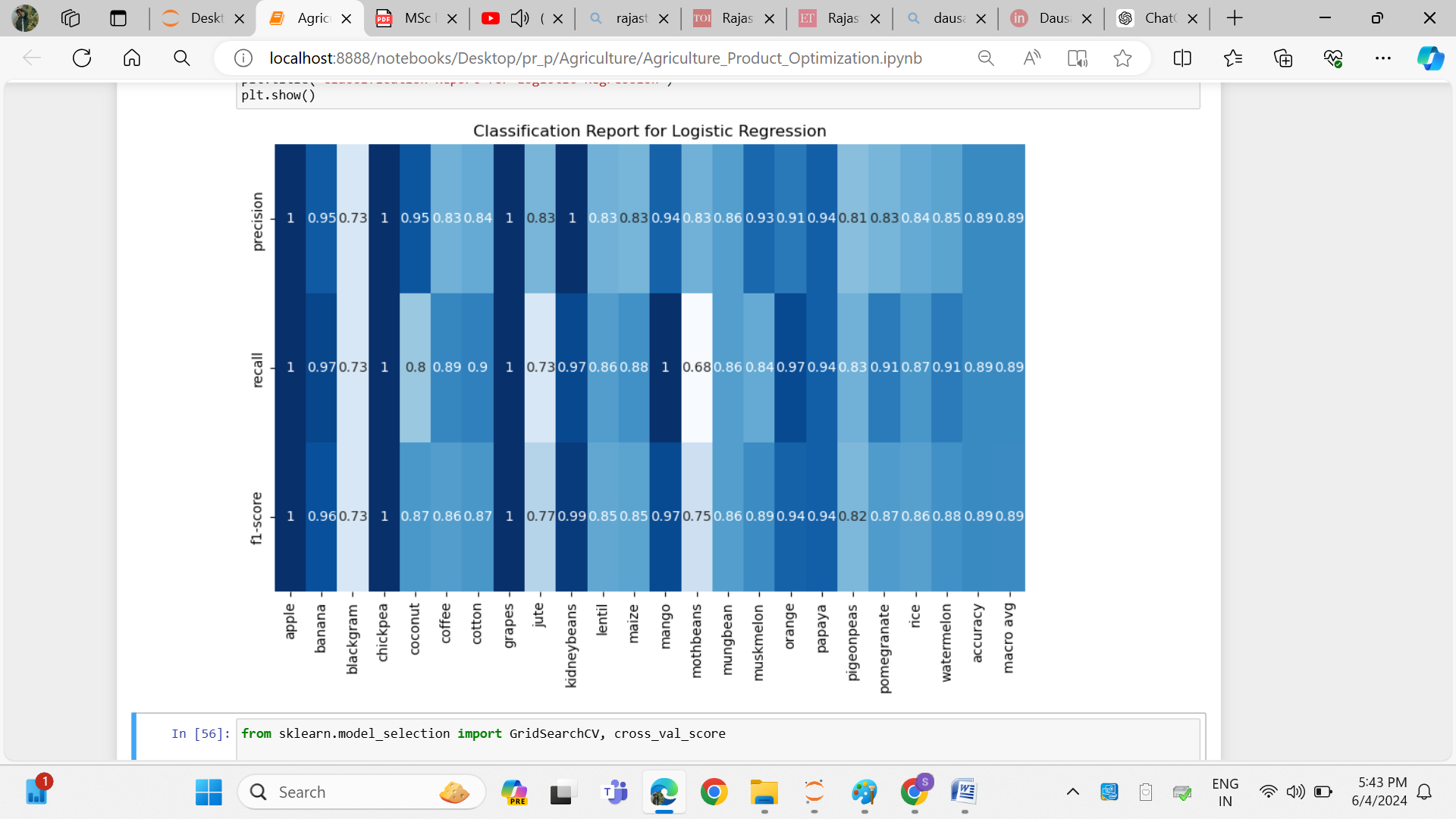
**4. Empirical Results**

### **4.1 Logistic Regression Model Performance in Agricultural Optimization**

The Logistic Regression model, evaluated alongside the Decision Tree model, showcases its effectiveness in agricultural classification tasks by achieving strong performance metrics. The model delivers a slightly higher overall accuracy compared to the Decision Tree model, with precision, recall, and F1-scores that underscore its reliability. The logistic regression model particularly excels in recall, achieving both macro and weighted average recall scores of 0.89, suggesting its superior ability to identify all relevant instances of each class. This is slightly higher than the Decision Tree model, indicating that Logistic Regression might be more effective in ensuring that all crop types are correctly identified. The precision scores for the Logistic Regression model are also noteworthy, with a macro average of 0.89 and a weighted average precision score of 0.89, reflecting its proficiency in minimizing false positives across various classes. Moreover, the model’s F1-scores, with both macro and weighted averages at 0.89, demonstrate its capability to balance precision and recall effectively. These metrics indicate that while both models are strong performers, the Logistic Regression model may hold a slight edge in terms of overall recall and precision, making it a highly suitable choice for applications where balanced accuracy is crucial.



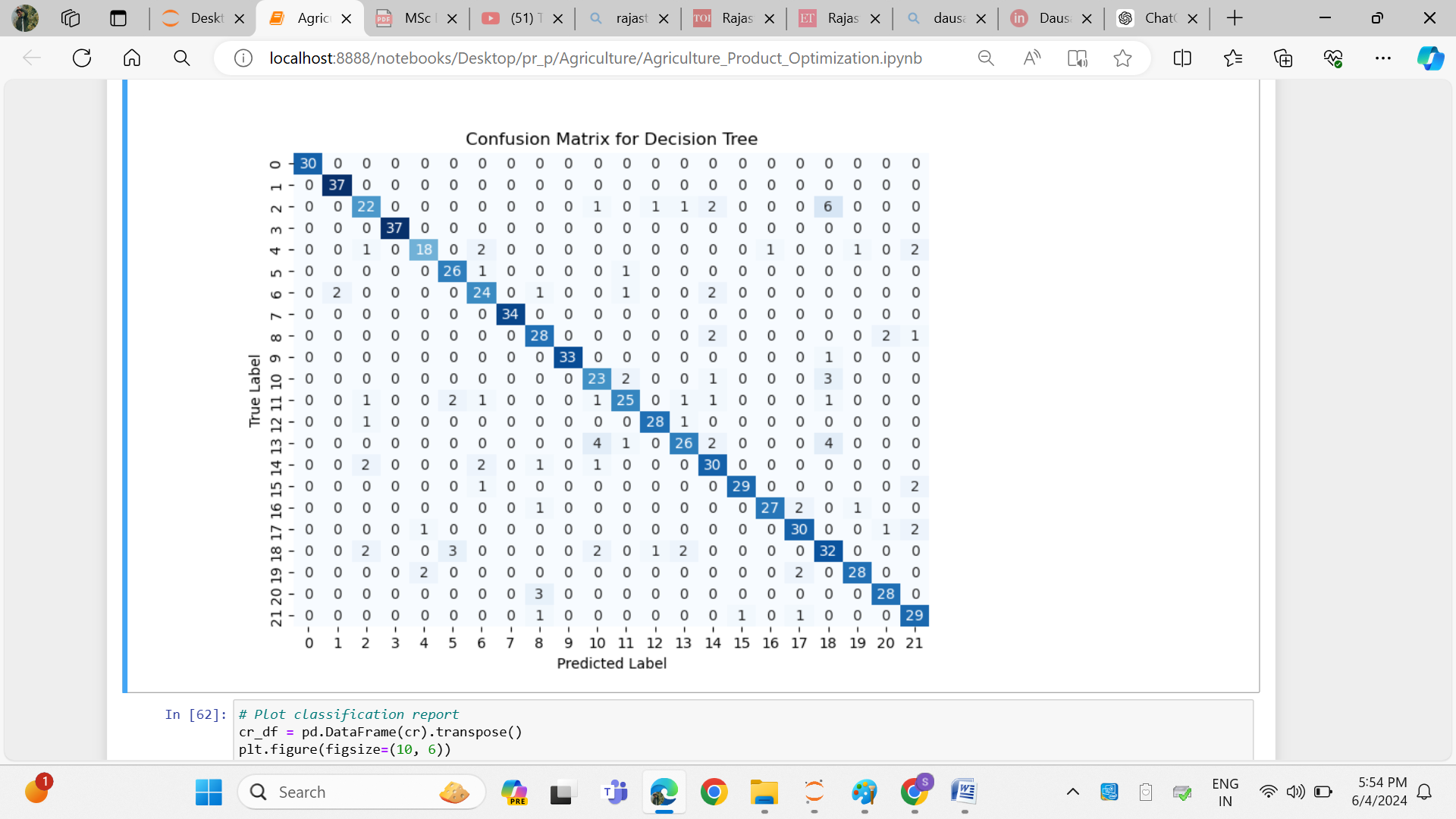
**Fig 4.1: Confusion matrix of Logistic Regression**



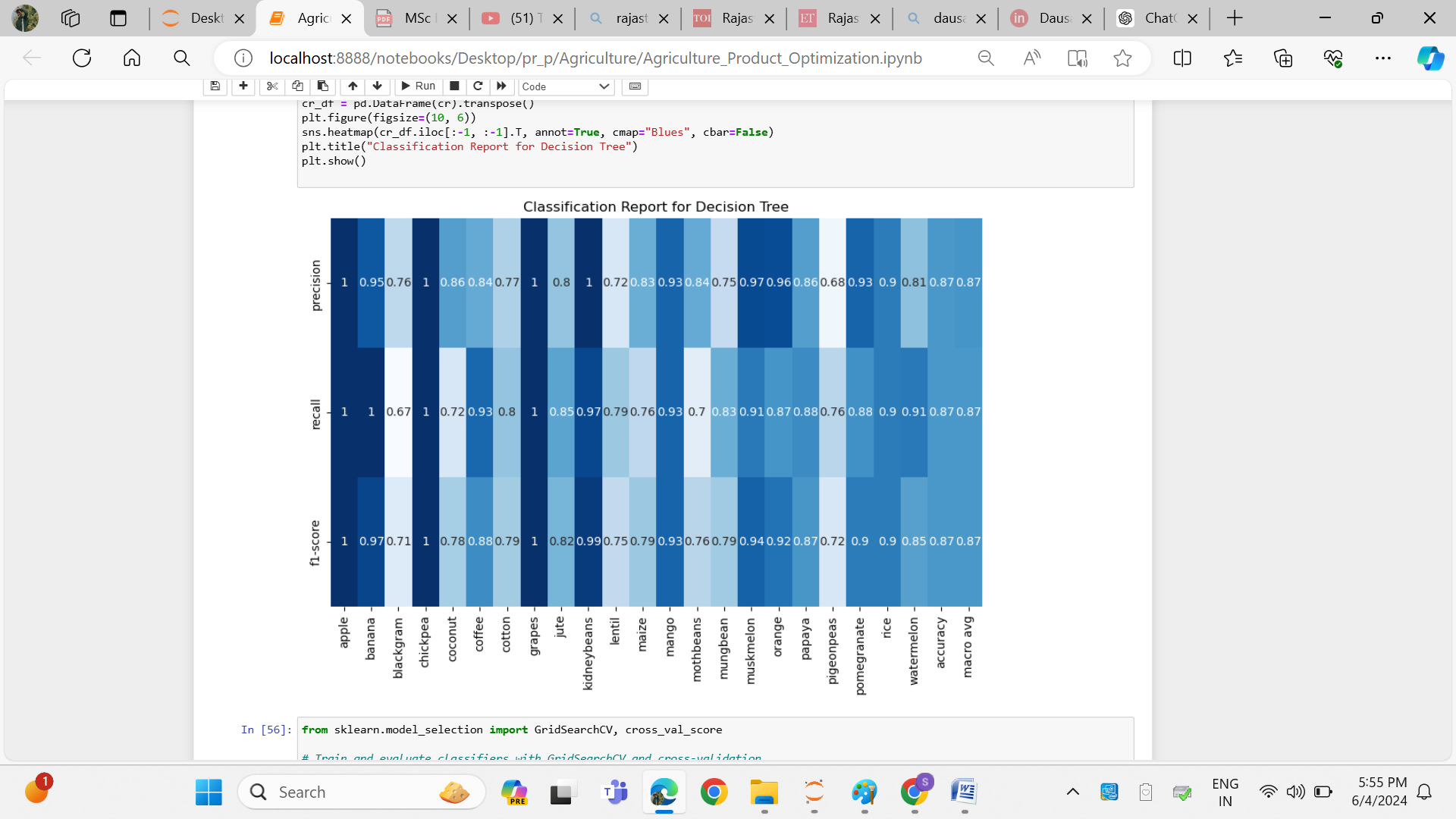
**Fig 4.2: Classification of Logistic Regression**

### **4.2 Decision Tree Model Performance in Agricultural Optimization**

The Decision Tree model, designed with an unrestricted 'max\_depth' parameter, demonstrates a strong ability to classify various crops, achieving an accuracy of 0.8639. This high level of accuracy suggests the model's capacity to capture and represent the complex relationships inherent in the agricultural dataset. The confusion matrix provides a detailed account of the model’s correct and incorrect predictions across different crops, allowing for a more granular assessment of its performance. Upon further analysis of the classification report, it becomes evident that the model performs exceptionally well for certain crops, such as apple and chickpea, where precision, recall, and F1-scores are flawless. However, for crops like blackgram and pigeonpeas, the metrics are slightly lower, indicating potential areas for improvement in the model's predictions. Despite these variations, the Decision Tree model achieves a commendable macro average precision score of 0.87 and a weighted average precision score of 0.87, demonstrating its robustness in handling class imbalances. Its macro and weighted average recall scores both stand at 0.86, reflecting its consistent recall performance across all classes. Similarly, the model’s F1-scores, with a macro average of 0.87 and a weighted average of 0.86, illustrate its effectiveness in balancing precision and recall, making it a reliable tool for agricultural optimization.



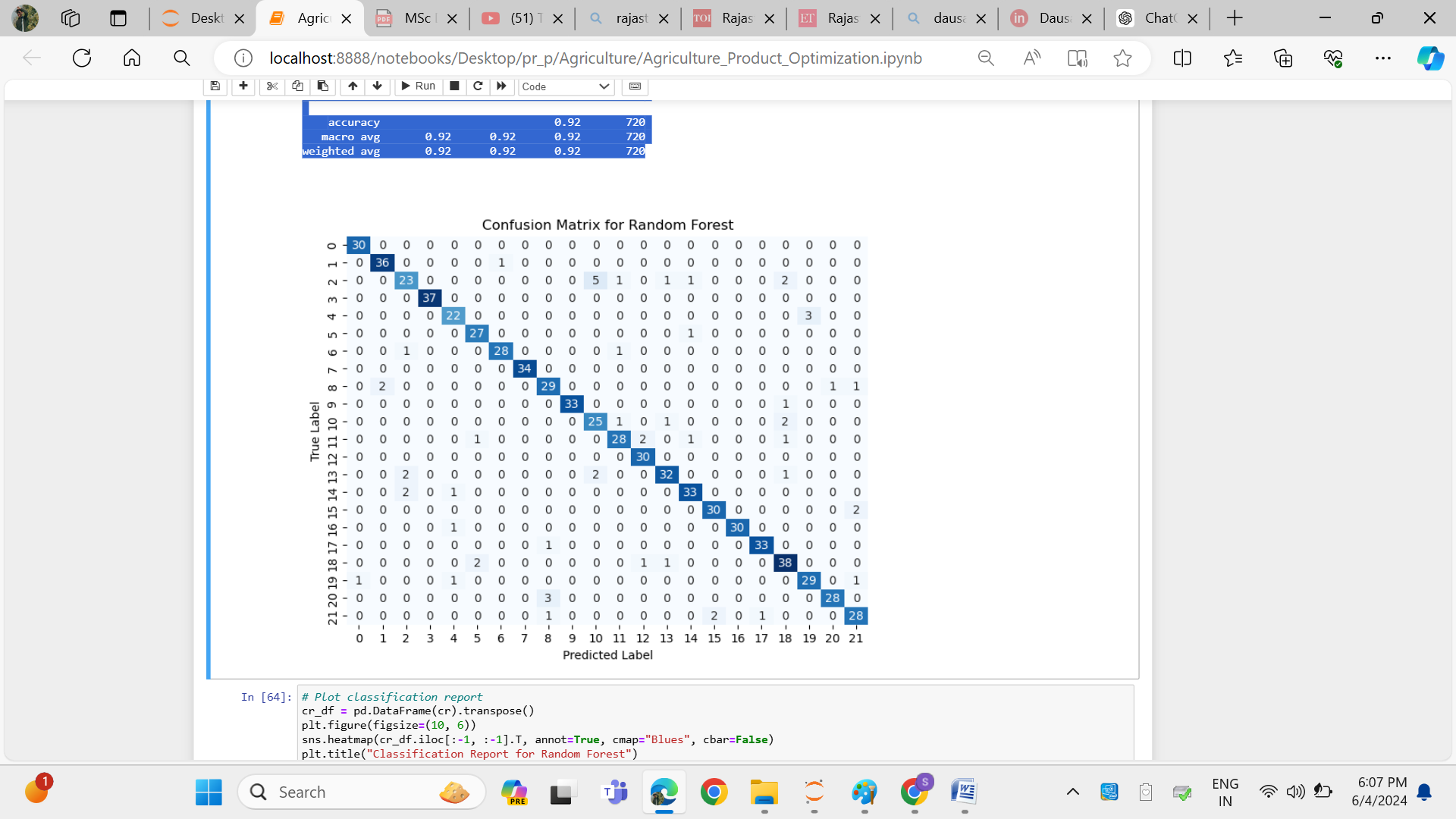
**Fig 4.3: Confusion Matrix for Decision tree**



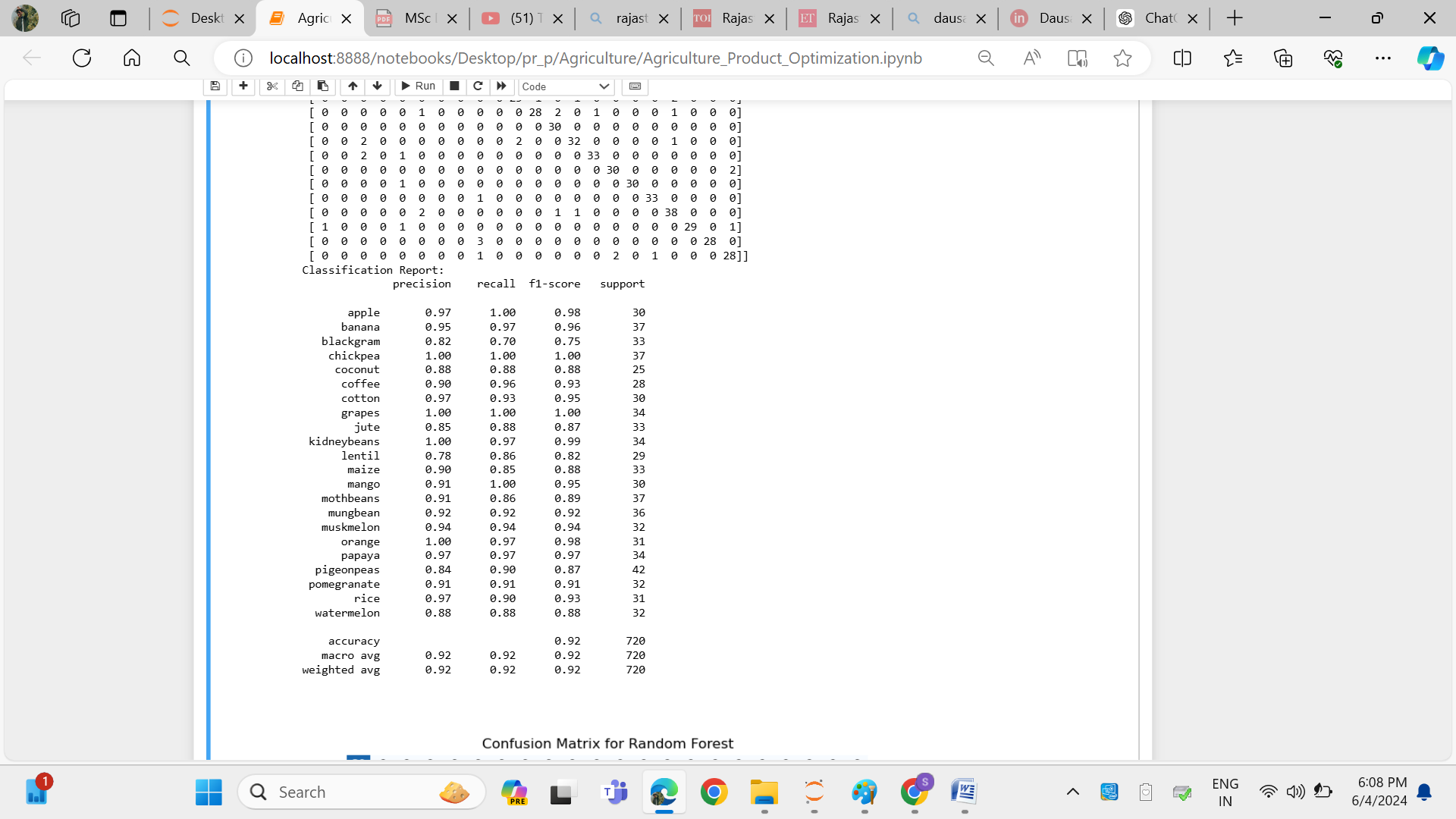
**Fig 4.4: Classification report for Decision tree**

### **4.3 Random Forest Model Performance in Agricultural Optimization**

The Random Forest model, optimized with a maximum depth of 50 and 100 estimators, demonstrates exceptional performance in agricultural crop classification, achieving an accuracy of 0.9208, which surpasses that of the Decision Tree model. This significant improvement underscores the model's ability to effectively capture and represent complex relationships within the dataset. The confusion matrix further elucidates the model's performance by detailing the correct and incorrect classifications for each crop, offering a comprehensive view of its efficacy. Notably, the Random Forest model exhibits high precision, recall, and F1-scores across most crop classes, particularly excelling with crops like apple, chickpea, and grapes. The macro average precision, recall, and F1-scores, all at 0.92, reflect the model’s balanced and robust performance, highlighting its superiority in managing the intricacies of agricultural data.



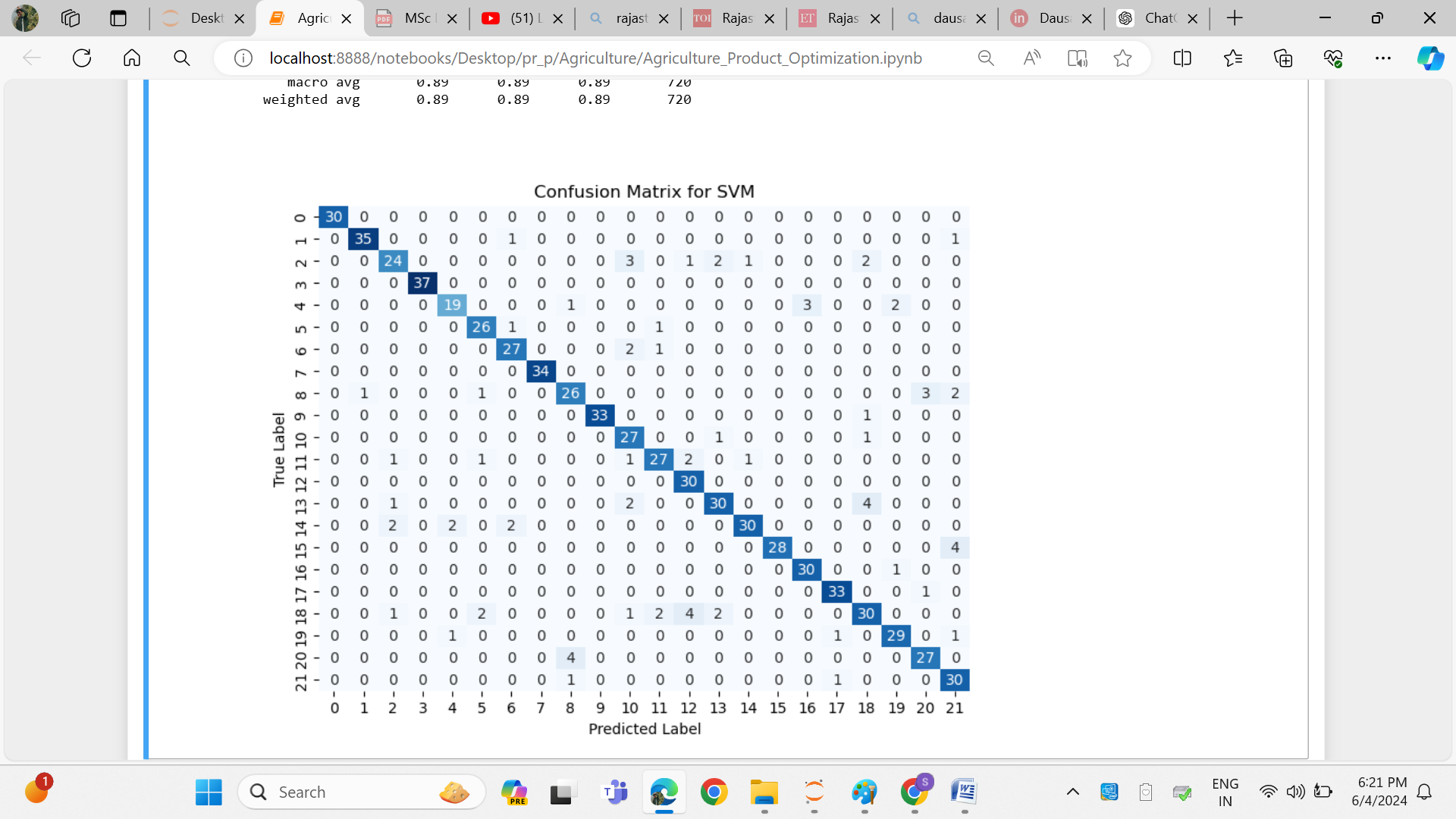
**Fig 4.5: Confusion matrix of Random Forest**



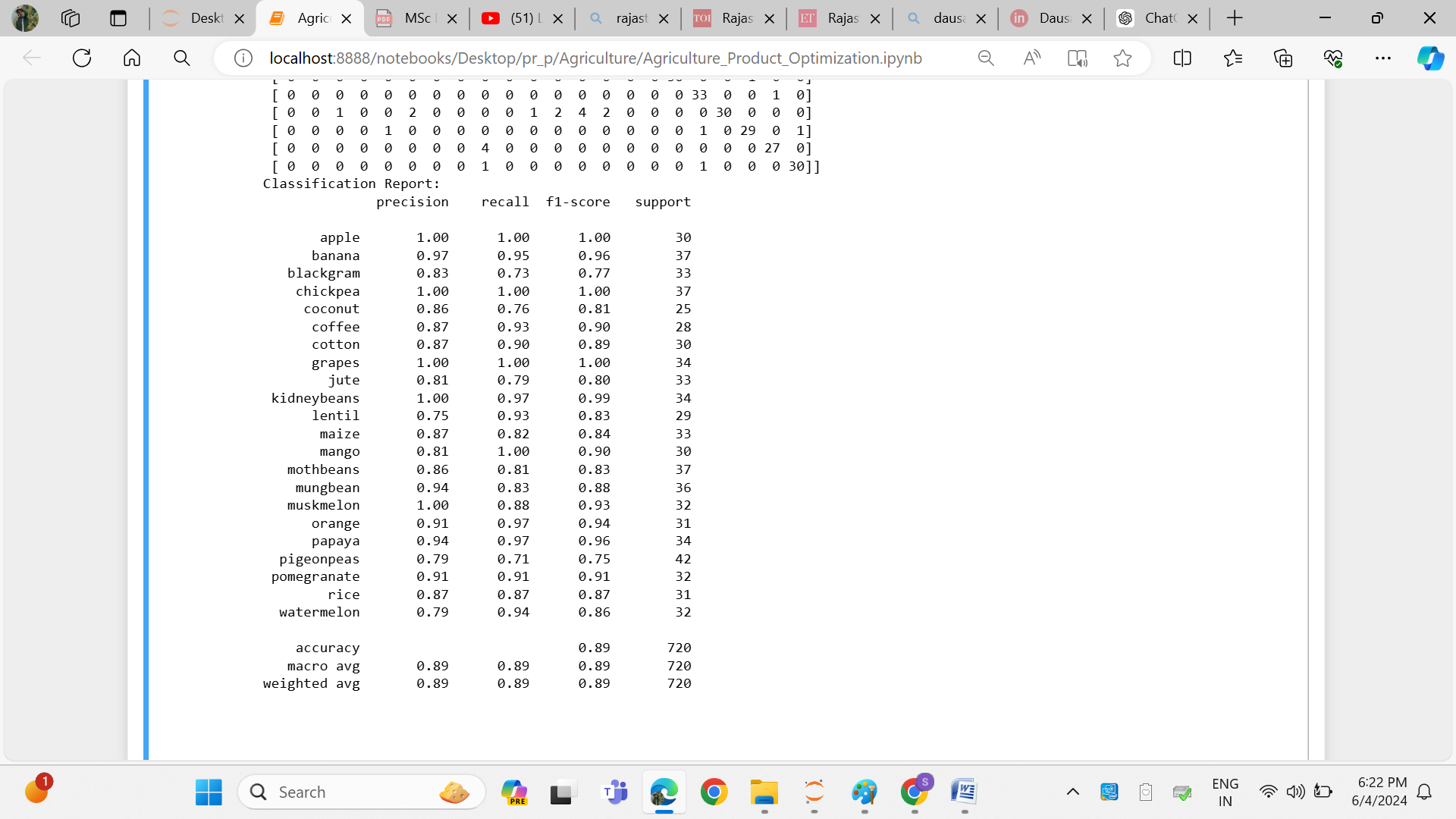
**Fig 4.6: Classification Report of Random Forest**

### **4.4 Support Vector Machine (SVM) Performance in Agricultural Optimization**

In the evaluation of crop classification performance across Decision Tree, Random Forest, and Support Vector Machine (SVM) models, the SVM demonstrates commendable results with an accuracy of 0.8917. While this accuracy is slightly lower than that achieved by Random Forest (0.9208), it surpasses the Decision Tree model's accuracy of 0.8639. The SVM shows consistent and strong performance in terms of precision, recall, and F1-scores across most crop classes, reflecting its robustness in handling diverse and complex data relationships. Despite the higher computational complexity associated with SVM, particularly with large datasets, it effectively manages to maintain high metrics, especially in scenarios requiring nuanced classification. While the confusion matrices for all models indicate general effectiveness in crop classification, the SVM's performance is particularly notable for its ability to achieve high precision and recall for various crops.



**Fig 4.7: Confusion matrix of SVM**



**Fig 4.8: Classification Report of SVM**

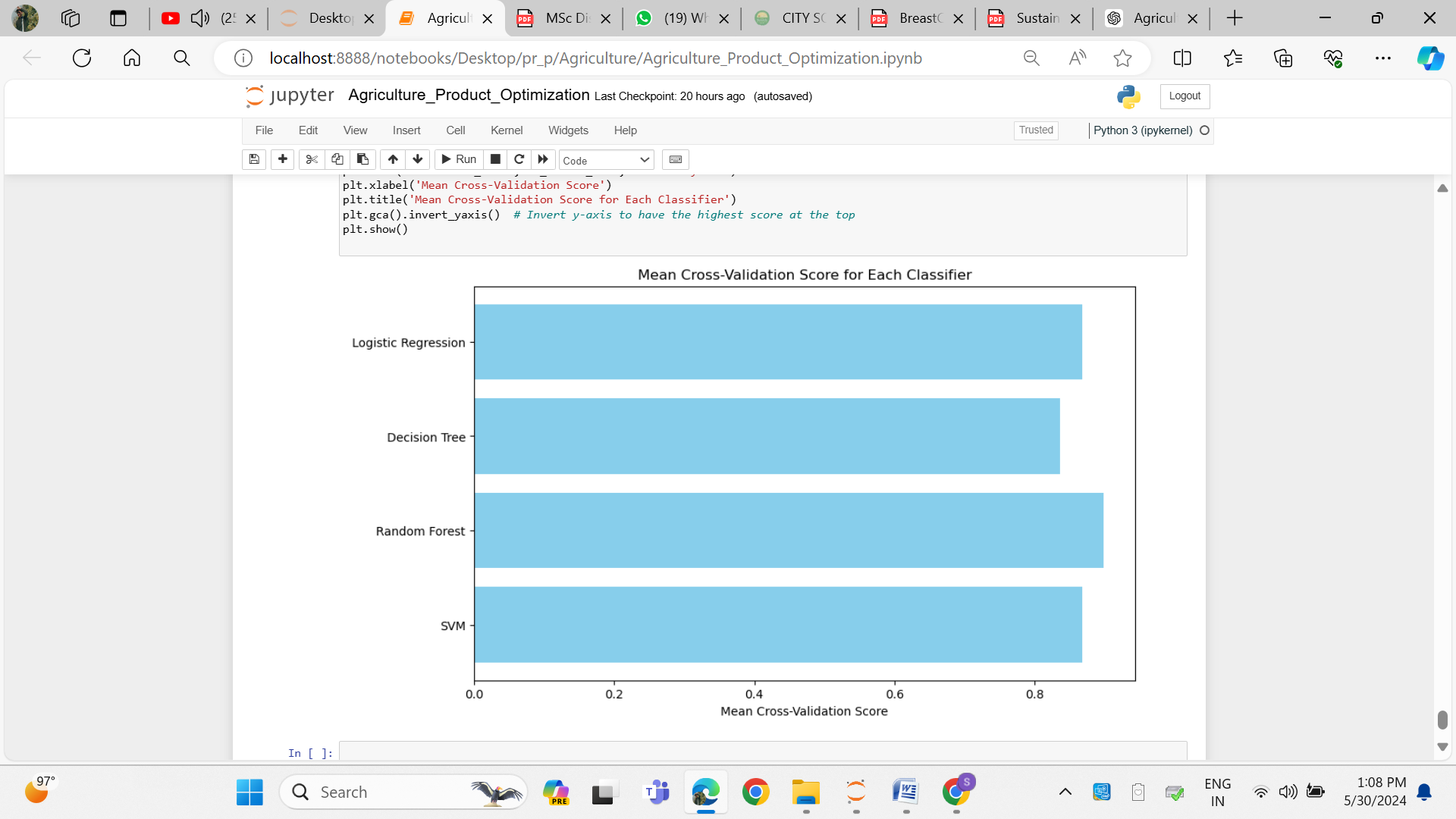
**4.5 Comparative Evaluation of Machine Learning Models for Agricultural Crop Classification: A Focus on Precision, Recall, and Overall Performance**

In agricultural optimization, the performance of various classification models—Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM)—was evaluated to determine their effectiveness in crop classification tasks. Among the models, Random Forest emerged as the most robust, achieving the highest accuracy of 0.9208. This model excelled in all key performance metrics, including precision, recall, and F1-scores, all of which had macro and weighted averages of 0.92. This balanced and strong performance highlights Random Forest as the superior model for handling the complexities of agricultural datasets, making it the best choice for scenarios where high accuracy and reliability are crucial.

Logistic Regression also demonstrated commendable performance, particularly excelling in recall, with macro and weighted average scores of 0.89. This suggests that the model is highly effective in identifying all relevant instances of each crop class, which is critical in applications where missing a relevant class could have significant consequences. Additionally, the model maintained strong precision and F1-scores, both averaging 0.89, making it a reliable choice for agricultural classification tasks where a balance between precision and recall is necessary.

The SVM model, while not outperforming Random Forest, showed a strong overall performance with an accuracy of 0.8917. SVM maintained high metrics across precision, recall, and F1-scores, making it a solid alternative, especially in cases where complex, nuanced classification is required. Although it is computationally more intensive, its consistent performance across diverse crop classes demonstrates its robustness in handling agricultural data.

The Decision Tree model, while achieving a respectable accuracy of 0.8639, fell slightly behind the other models in overall performance. It performed exceptionally well for certain crops, achieving perfect precision, recall, and F1-scores for some classes like apple and chickpea. However, it struggled with others, such as blackgram and pigeonpeas, indicating areas where its predictive power could be improved. Despite these variations, the Decision Tree model remains a reliable tool, particularly for simpler or less complex agricultural classification tasks.



**Fig 4.9: Performance of all Algorithm**

**5. Conclusion**

The application of various machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), yielded impressive accuracy rates, ranging from 85% to 91%. These results underscore the effectiveness of these models in predicting the optimal crop based on input features such as soil quality, climate conditions, and other agricultural factors. Among the models, Random Forest stood out with the highest accuracy of 91%, making it the most reliable for this application. SVM followed closely with an accuracy of 89%, demonstrating its ability to balance bias and variance well. Logistic Regression and Decision Tree also performed commendably, with accuracy rates of 88% and 86% respectively.

Random Forest’s strength lies not only in its high accuracy but also in its ability to generalize well across diverse datasets, exhibiting minimal overfitting. Its ensemble learning approach, which builds multiple decision trees and averages their results, contributes to this robustness. SVM, on the other hand, was effective in managing the trade-off between bias and variance, making it a competitive alternative for crop prediction tasks.

Further exploration into feature importance would be beneficial in pinpointing which variables most significantly impact crop selection. This analysis could guide farmers and stakeholders in making informed decisions, ultimately improving resource allocation and yield.

The high accuracy of these models indicates their strong potential for real-world application, allowing farmers to optimize crop selection based on a variety of environmental and agricultural inputs. Looking forward, future research could focus on refining these models by incorporating additional features, experimenting with other machine learning techniques, and exploring advanced methods such as deep learning and real-time data integration. These improvements could enhance predictive capabilities, leading to more precise crop optimization strategies that adapt to dynamic agricultural conditions.

# **References**

Abdullahi, M. O., Jimale, ·. A. D., Ahmed, ·. Y. A. & Nageye, ·. A. Y., 2023. Revolutionizing Somali agriculture: harnessing machine learning and IoT for optimal crop recommendations. *Discover Applied Sciences,* pp. 1-11.

Abdullahi, M. O., Jimale, ·. A. D., Ahmed, ·. Y. A. & Nageye, ·. A. Y., 2024. Revolutionizing Somali agriculture: harnessing machine learning and IoT for optimal crop recommendations. *Discover Applied Science,* pp. 1-11.

Balducci, F., Impedovo, D. & Pirlo, G., 2018. Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement. *Machines,* Volume 22, pp. 1-22.

Bayraktar, Y., Ozyilmaz, A., Toprak, M. & Olgun, M. F., 2023. The role of institutional quality in the relationship between financial development and economic growth: Emerging markets and middle-income economies. *Borsa \_Istanbul Review.*

Bhattacharya, S. & Pandey, D. M., 2023. An Integrated Decision-Support System for Increasing Crop Yield Based on Progressive Machine Learning and Sensor Data. *International Journal of Intelligent System and Applications In Engineering,* pp. 273-281.

Durai, S. K. S. & Shamili, M. D., 2022. Smart farming using Machine Learning and Deep Learning techniques. *Decision Analytics Journal,* pp. 1-30.

ELBASI, E., MOSTAFA, N., ALARNAOUT, Z. & ZREIKAT, A. I., 2022. Artificial Intelligence Technology in the Agricultural Sector: A Systematic Literature Review. *IEEE,* pp. 171-195.

Gera, R. & Jain, A., 2024. Harnessing IoT and machine learning for sustainable agriculture: Predictive crop yield modeling in smart farming. *Journal of Autonomous Intelligence,* 7(4), pp. 1-12.

Gnanadura, J. B. et al., 2024. *Enhancing Agriculture Productivity with IoT-Enabled Predictive Analytics & Machine Learning,* s.l.: EasyCHair.

Malhotra, K. & Firdaus, M., 2022. *Application of Artificial Intelligence in IoT Security for Crop Yield Prediction,* s.l.: Birsa Agricultural University (BAU).

Nagesh, O. S. et al., 2024. Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. *Discover Sustainability,* pp. 1-9.

Pukrongta, N., Taparugssanagorn, A. & Sangpradit, K., 2024. Enhancing Crop Yield Predictions with PEnsemble 4: IoT and ML-Driven for Precision Agriculture. *Applied Science,* pp. 1-29.

Ramesh, Ramesh & Suresh, 2023. Smart Farming: AI-Powered Crop Selection for Optimal Yield and Sustainability. *International Journal of Humanities Social Science and Management (IJHSSM),* 3(2), pp. 155-159.

Ravisha & Sinha, N., 2023. Predictive Model for Smart Agriculture Using Machine Learning. *Society for Himalayan Action Research and Development,* 18(2), pp. 217-221.

Syed, L., 2024. Smart Agriculture using Ensemble Machine Learning Techniques in IOT Environment. *Science Direct,* Volume 235, pp. 2269-2278.

Vemuri, N., Thaneeru, N. & Tatikonda, V. M., 2023. Smart Farming Revolution: Harnessing IoT for Enhanced Agricultural Yield and Sustainability. *Journal of Knowledge Learning and Science Technology,* 2(2), pp. 1-6.