8-2024

INTEGRATED CROP RECOMMENDATION SYSTEM: HARNESSING MACHINE LEARNING

Arun Kumar Reddy Saddikuti

INTEGRATED CROP RECOMMENDATION SYSTEM: HARNESSING MACHINE LEARNING

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology

---

by

Saddikuti Arun Kumar Reddy

August 2024

INTEGRATED CROP RECOMMENDATION SYSTEM: HARNESSING MACHINE LEARNING

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Saddikuti Arun Kumar Reddy

August 2024

Approved by:

Dr. Nima Molavi, Committee Member, Co-Chair

Dr. Conrad Shayo, Committee Member, Co-Chair

Dr. Conrad Shayo, Department Chair, Information and Decision Sciences

ABSTRACT

To meet the increasing demand for food while also reducing impact this study introduces an innovative "Integrated Crop Recommendation System" that combines advanced machine learning with sustainable farming methods. The goal of this system is to transform how crops are chosen by considering factors like soil quality, local climate and habitats for pollinators thereby enhancing the precision and effectiveness of crop suggestions. In contrast to agricultural decision support systems that often neglect the interconnectedness of soil health, weather conditions and biodiversity, this new approach aims to improve food security and sustainability. The primary research focus is on optimizing practices that support pollinators in environments. The research aims to provide farmers with enhanced guidance and deeper insights into the relationships among soil quality, weather patterns and ecological sustainability offering a solution for modern farming practices. The study encompasses a literature review, methodology development, data analysis, and discussion of findings. Outlines research directions. Research Questions are: Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations? Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

The Findings and Discussions for the questions are: Q1) In our research we. Evaluated a system for suggesting crops based on machine learning. This system considers factors such as soil quality, weather conditions and reliance on pollinators by studying data sets related to crop recommendations and pollination. Our analysis of the data showed connections like the relationship between rainfall and crop production. Additionally, our decision tree model performed better than the SVM model, in predicting crop yields. Q2) The research shows that including methods to support pollinators in crop recommendations based on machine learning can improve their performance and durability. It stresses the importance of factoring in pollination aspects when making decisions. By grouping crops based on their reliance on pollinators it underscores the need for customized conservation approaches. Proves that taking types of pollinators into account greatly enhances the precision of predicting crop yields. Conclusions for each question are: Q1) Our study shows that using machine learning to examine the connections, among soil makeup, weather conditions and reliance on pollinators improves decision making tools, for agriculture. This in turn boosts the accuracy of crop recommendations. Helps ensure food security. Q2) By including actions that support pollinators in crop suggestions generated by machine learning it boosts their dependability and strength. This underscores the importance of efforts on conserving pollinators to enhance the resilience of crops and the overall health of ecosystems. Areas of further studies for each question are: Q1) The success of crop recommendation systems powered by machine learning models relies heavily on tuning hyperparameters and structural components than just sticking to preset or manually configured settings. Q2) Incorporating a variety of factors and crop specific characteristics into crop recommendation systems provides a grasp of growth elements essential for improving the accuracy of recommendations.

ACKNOWLEDGEMENTS

I would like to acknowledge and appreciate the support given by Dr. Conrad Shayo and Professor Dr. Nima Molavi in completing this project.

DEDICATION

Preethi, this project is a tribute to you, whose constant backing, inspiration and drive have illuminated my path. Your faith in me during times and your unwavering optimism in situations have been my pillar of support. You remained patient and compassionate, providing more than comforting words. The priceless treasure of friendship. Without your encouragement this project would have lacked fulfillment. This endeavor reflects our commitment and your remarkable resilience that has motivated me to keep going.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS vi

DEDICATION vii

LIST OF FIGURES x

CHAPTER ONE: INTRODUCTION 1

Problem Statement 3

Research Questions 4

Objective of this Project 5

Organization of the Study 6

CHAPTER TWO: LITERATURE REVIEW 7

CHAPTER THREE: METHODOLOGY 12

CHAPTER FOUR: DATA ANALYSIS AND FINDINGS 22

Exploratory Data Analysis (EDA) Results 22

Correlation Analysis 24

Decision Tree Analysis Process 32

SVM Model Evaluation for Crop Recommendation 36

Cross Validation 39

Regarding feature importance 45

CHAPTER FIVE: DISCUSSION, CONCLUSION AND AREAS OF FURTHER STUDY 50

Discussion 50

Conclusion.. 52

Areas of Further Study.. 53

REFERENCES ……………………………………………………………………......59

LIST OF FIGURES

Figure 1. Exploratory Data Analysis (EDA) Results. 22

Figure 2. Merged Data Set Output. 23

Figure 3. Temperature vs Humidity by Crop Type. 23

Figure 4. Factors Affecting Crops. 24

Figure 5. Correlation Analysis. 25

Figure 6. Correlation Matrix. 26

Figure 7. Output for Decision Tree Analysis Process. 34

Figure 8. SVM Model Evaluation for Crop Recommendation. 38

Figure 9. Cross Validation for Decision Trees. 39

Figure 10. Cross Validation for SVM. 40

Figure 11. Average Presence of Pollinators by Crop. 42

Figure 12. Distribution of Pollinators. 43

Figure 13. Feature Importance. 45

Figure 14. Decision Tree analysis of Baseline Model with Pollination Variables.
.... 47

Figure 15. Decision Tree analysis after dropping the pollination variables….. .. 48

CHAPTER ONE

INTRODUCTION

In agriculture it's essential to focus on farming techniques to meet the growing food demand while reducing harm to the environment. This study introduces a method known as the "Integrated Crop Recommendation System." This system combines advanced machine learning algorithms with an understanding of practices aiming to transform how we choose which crops to grow. By considering factors such as soil composition, local weather patterns and promoting practices that support pollinators, this research aims to optimize crop recommendations. Ultimately this will improve productivity while ensuring sustainability. (A.m., Kremen 2014)

The crops productivity can be influenced by the types of crops selected for farming. Farmers often struggle with knowing which crops are suitable for a type of soil (Kumar et al., 2015). This adds a challenge, to predicting crop yields. In the past farmers relied on their experiences with a field and crop to make yield predictions. A technique known as the Crop Selection Method (CSM) (Kumar et al. 2015) categorizes crops into types; crops (that can be grown year-round) seasonal crops (that can only be cultivated during specific seasons) long term crops (that require a significant amount of time to grow) and short-term crops (that have a relatively quick growth period). In regions the amount of rainfall received serves as an indicator for estimating crop yield each year. Following the approach outlined in (Kumar et al. 2015) there are sequences of these four categories of crops. Only the sequence that provides the highest average yield is selected. The field of agriculture has experienced progress through the incorporation of machine learning technologies. Machine learning has played a role in predicting crop yields, examining soil properties and forecasting weather patterns.

Ensuring food security and eradicating hunger for the growing population heavily relies on sustainability. Experts estimate that by 2050 we need to increase food production by 60-110% to feed a population of 9 10 billion (Tilman et al. 2011; Pardey et al., 2014; Rockström et al., 2017). Therefore, it is crucial to shift our focus from enhancing productivity to prioritizing agricultural sustainability (Rockström et al.,2017). (Hevia et al., 2022) administered 376 face-to-face questionnaires in four areas of Spain with different dominant pollinator-dependent crops, to assess the factors behind farmers' perceptions, knowledge, and practices adopted to promote pollination. Overall, 92.7% of the respondents recognized that pollinator insects are necessary for crop production, and 73.4% perceived pollinator decline in their farms. The practices mentioned in this article (Hipólito et al., 2021) which're friendly to pollinators have the potential to create situations where everyone benefits. They can assist farmers and policymakers in preserving or restoring biodiversity while also improving crop yield. To achieve sustainability, it is important for farming methods to focus on food production that effectively utilizes the resources provided by nature without causing harm. This can be achieved through intensification as discussed by (Pretty et al., 2006). It is crucial to consider the needs of farmers and ensure they have access to technology and information for those who have limited resources (Pretty, 2003; Steward et al., 2014).

Soil composition is incredibly important when it comes to the health and productivity of crops. Weather patterns also have an impact on the success of efforts. In addition to these factors incorporating practices that support pollinators brings an element to the proposed system. The goal of this research is to attract pollinators like bees and butterflies to strike a balance between maximizing crop yields and preserving biodiversity. However, many studies have failed to consider the integration of these components resulting in a gap when it comes to creating decision support systems that are specifically designed for sustainable farming practices. The purpose of this paper is to address this gap by presenting a model that not only considers the intricate connections between soil composition and weather patterns but also includes strategies aimed at promoting environments suitable for pollinators.

Problem Statement

In today's agriculture the challenge is to find ways to choose crops that can satisfy the increasing need for food while also taking care of issues. However, the current state of agricultural decision support systems often lacks an approach. Mainly focuses on individual factors like soil composition and weather patterns. (Kumar et al., 2015) This narrow perspective creates a gap in precision farming, which hampers the development of practices. Additionally, the crucial role played by pollinators in crop production is often ignored, leading to a decline in biodiversity and ecological balance.

This gap plays a role as it hampers progress towards achieving precision and sustainable agriculture. When there is no integrated framework, farmers lack a tool that considers the interactions between soil health, weather conditions and ecological sustainability. It is essential to bridge this research gap in order to develop an adaptable decision support system that can enhance crop recommendations, improve yields and promote conscious farming practices. To overcome these challenges there is a need for an Integrated Crop Recommendation System that utilizes machine learning to analyze and combine information about soil content, weather patterns and practices that are friendly, towards pollinators. Through this research our aim is to drive agriculture into an era by promoting both environmental sustainability and improved crop productivity.

Research Questions

This project focus is on the following question(s):

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

Objective of this Project

The proposed research project has a rationale as it aims to integrate machine learning techniques into an agricultural decision support system. This integration has the potential to bring about changes in farming practices, promote sustainability and tackle pressing global challenges. With the world confronting a growing population the need for food is becoming more urgent. Therefore, optimizing processes for productivity is crucial. By utilizing machine learning algorithms to analyze soil composition and local weather patterns we can unlock possibilities for refining crop recommendations. This will result in utilization of resources and contribute towards ensuring global food security.

Moreover, the focus of this research, on promoting practices that attract and support pollinators, brings an aspect to decision making in agriculture. The decrease in pollinator populations poses a risk, to both crop pollination and biodiversity. By integrating these practices that aid pollinators the research aims to not only enhance crop yields but also contribute to the preservation of ecological systems. Considering the differences when optimizing these practices ensures their adaptability to a range of environments, making this proposed system a versatile and globally applicable solution.

The study seeks to pioneer a farming approach using machine learning advancements. The expected outcomes could provide insights for farmers. Enhance our understanding of the intricate relationship between soil health, weather conditions and ecological sustainability, in agriculture. In essence this research aims to bridge existing gaps in methods by presenting a solution that meets the changing demands of our growing global community while advocating for environmental responsibility.

Organization of the Study

This culminating experience project is organized as follows: Chapter one provided an introduction, problem statement, research questions and Justification. Chapter two will provide the literature review. Chapter three will consist of research methodology. Chapter four will contain the analysis of the data and the findings. Chapter five will provide the discussion, conclusion, and areas for further study.

CHAPTER TWO

LITERATURE REVIEW

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

In research there is a focus on incorporating machine learning methods into agricultural decision support systems to improve accuracy and efficiency. Reddy and Kumar (2021) delve into the application of machine learning to forecast crop yields particularly emphasizing the interconnectedness between soil composition and local weather patterns. The use of machine learning models by Reddy and Kumar (2021) relies on factors for precise crop yield prediction. These factors encompass soil data, crop details, weather variables, nutrient elements, solar data as additional aspects like wind speed and atmospheric pressure (Reddy & Kumar 2021). These facets play a role in constructing prediction models that offer detailed insights into the dynamic elements influencing crop growth and progress. While machine learning techniques show promise challenges arise from employing networks and supervised learning methods in predicting crop yields. Overcoming issues such as reducing errors enhancing prediction efficiency and capturing relationships is crucial for improving the precision and effectiveness of crop yield prediction models.

Medar, Rajpurohit and Shweta (2019) discuss how machine learning methods are applied in agriculture focusing on data analysis, predictive modeling, optimization, risk assessment and precision farming. In a study, by Malik, Sengupta and Jadon (2021) they delve into comparing soil characteristics using machine learning algorithms such as K Nearest Neighbor, Naïve Bayes and Decision Trees to forecast soil fertility and crop yields. This research works underscores the potential of machine learning in improving decision making processes concerning crop selection and soil analysis. By integrating models based on soil attributes with machine learning techniques farmers and stakeholders can benefit from decision making abilities leading to increased crop productivity, optimized resource usage, risk reduction, cost effectiveness and promotion of sustainable agricultural practices (Medar et al., 2019; Malik et al. 2021). Aligning farming practices with soil properties results in sustainable agricultural approaches.

The coming together of machine learning in selecting crops based on weather conditions is essential for maximizing crop output. Yet it's crucial to recognize the significance of pollinators in farming. Bees and butterflies among pollinators are players in the reproduction of flowering plants. They play a role in promoting biodiversity, food supply and ecosystem functions. In areas where agriculture relies heavily on insect driven pollination, incorporating practices that support pollinators are vital for crop growth.

Isaacs and colleagues (2017) support Integrated Crop Pollination (ICP) as a method to guarantee sustainable yields of crops that rely on pollination. When it comes to ICP, the involvement of pollinators in farming practices gains significance. Both managed and wild pollinators play a role in ensuring profitable crop production through dependable and cost-effective pollination services. By grasping the connections between crops and pollinators farmers can adopt techniques that promote and bolster pollinator populations thus benefiting wellbeing and agricultural output (Isaacs et al., 2017).

This research aims to address a gap in studies by exploring the influence of soil quality, weather conditions and pollinator behavior on decision making. While previous research has focused on areas such as crop enhancement, soil health and pollinator friendly practices individually there has been limited investigation into how these factors intersect. This study intends to leverage machine learning models to develop a framework that considers soil composition, weather patterns and the creation of pollinator habitats simultaneously. The objective is to comprehend how these components collectively influence crop growth and development. The research does not seek precision. Also strives to recommend friendly and sustainable crops. By investigating the relationships between soil properties, weather fluctuations and pollinator behavior this study aims to unveil the dynamics that shape ecosystems. Through this approach we aspire to pave the way for conscious agricultural practices.

Effectiveness of Integrated Machine Learning Models in Crop Recommendations

Reddy and Kumar et al. (2021) laid the groundwork for understanding the interconnected relationships between soil composition, weather patterns, and pollinators. While their study primarily focused on predictive modeling for crop yields, it underscored the importance of considering pollinator dynamics. Further insights were gained from Isaacs et al. (2017), advocating for Integrated Crop Pollination (ICP) as a strategy for sustainable yields. The literature suggests that incorporating practices supporting pollinators positively impacts the overall strength of crop recommendations by enhancing pollination efficiency, leading to increased yield and quality.

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

Pollinators are essential in agroecosystems playing a role in crop pollination and boosting productivity. With the decline of insect populations there is a rising focus on the importance of enhancing pollinator habitats to sustain pollination services. This review combines insights from research studies to examine how promoting practices that support pollinators can impact the quality and resilience of crop recommendations generated by machine learning models.

Desneux et al. (2012) emphasize the multifaceted benefits of enhancing pollinator habitats beyond pollination services. Biodiversity conservation is one such advantage, as incorporating flowering plants into non-cropped farmlands restores habitat for various non-pest insects, contributing to overall insect biodiversity conservation (Desneux et al., 2012). The review further highlights biocontrol services, explaining that greater plant diversity sustains populations of natural enemies of pests, thereby reducing the reliance on pesticides (Desneux et al., 2012).

Based on these findings, research conducted by Pywell and colleagues in 2015 shows that implementing wildlife farming techniques can influence crop production. By establishing habitats for wildlife within farms, such as non-crop areas along the edges of fields these methods can improve yields in cultivated areas without negatively impacting overall productivity (Pywell et al., 2015).

CHAPTER THREE

METHODOLOGY

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

The study will create a system for suggesting crops by combining soil quality, weather patterns and pollination elements through machine learning techniques. Decision trees and Support Vector Machines (SVM) are the models chosen to forecast crop output based on these factors. Prior to building the models, an, in depth analysis of the data will be carried out to understand how soil quality, weather conditions, pollination elements and crop yields are interconnected.

Exploratory Data Analysis:

Exploratory data analysis is also done investigate the relationship between pollination, crop yield, and environmental factors like soil composition and weather patterns:

Data Visualization:

Purpose: Explore and interpret the relationships between variables during the exploratory data analysis (EDA) phase.

Scatter Plots:

Visualize pairwise relationships, e.g., between pollinator abundance and crop yield or temperature and rainfall. Provide insights into potential linear or non-linear relationships.

Bar Plots and Pie Charts:

Visualize categorical variables such as types of pollinators.

Display the distribution and frequency of different pollinator types across the dataset.

Correlation Analysis:

Purpose: Examine the relationships between key variables such as pollinator abundance, soil properties, weather conditions, and crop yield.

Method: Calculate correlation coefficients (e.g., Pearson correlation) to quantify the strength and direction of linear associations. This analysis helps in understanding the interplay between different factors affecting crop yield.

Data Collection and Preprocessing:

The crop recommendation dataset is sourced from Kaggle, a publicly available platform for datasets and data science competitions. The dataset consists of real-world data collected from agricultural sources and may include observations from various locations and time periods.

The data on pollination is taken from a study titled " benefits of animal pollination to agriculture" by Chaudhary, O. P., & Chand, O. (2017). This dataset is not available to the public as an entity and is derived from the research paper where the authors included it for their publication. It comprises research results. Includes practical information on the abundance of pollinators crops reliance on pollinators various types of pollinators and crop productivity, within the realm of Indian farming.

Upon obtaining the datasets we will carefully examine them to pinpoint and rectify any discrepancies, gaps, in data or unusual values.

Data preprocessing steps, including cleaning, normalization, and feature engineering, will be performed to prepare the datasets for analysis.

Data Inspection:

The data, on "Crop Recommendation" and "pollination" was imported into a Jupyter Notebook platform for review where an initial examination was carried out to grasp its organization, such, as column headings, data formats and any absent information. This process was undertaken to verify the accuracy and entirety of the data to proceed with analysis.

The Crop recommendation dataset contains the following columns:

N: Represents the Nitrogen content in the soil.

P: Represents the Phosphorus content in the soil.

K: Represents the Potassium content in the soil.

Temperature: Represents the temperature in Celsius.

Humidity: Represents the humidity level.

pH: Represents the pH level of the soil.

Rainfall: Represents the amount of rainfall in mm.

Label: Represents the recommended crop or classification label.

This dataset contains data on the composition of soil (levels of Nitrogen, Phosphorus and Potassium) weather patterns (Temperature, Humidity and Rainfall) and soil pH levels along with the recommended crop types associated with them. This information is crucial for grasping how soil characteristics, weather conditions and crop choices are interconnected.

<u>The pollination dataset contains the following columns:</u>

Label: Represents the label or identifier for each record.

Types of Pollinators: This column contains information about the types of pollinators associated with each crop. After one-hot encoding, this will be split into multiple binary columns representing the presence or absence of specific pollinators.

After one-hot encoding, the dataset might look like this:

Label: The crop type.

Honeybees: This column shows whether honeybees are present (1) or absent (0).

Bumblebees: This column indicates the presence (1) or absence (0) of bumblebees.

Flies: Indicates if flies are present (1) or absent (0).

Butterflies: Shows the presence (1) or absence (0) of butterflies.

Wild bees: Indicates the presence (1) or absence (0) of bees.

This data set offers insights into how crop production's linked to the services offered by various pollinator types. It's essential to grasp these connections to evaluate how pollinators affect productivity and sustainability.

Merging Datasets: The datasets, for "Crop Recommendation" and "Pollination" were joined together using a shared identifier the "Label" column that signifies the type of crop. This merging of data based on an identifier enabled the blending of insights, from both sources. Eased the process of conducting thorough analysis.

Matching Labels: The "Label" column in each dataset was examined to ensure consistency and compatibility for merging. Each data set contained the same set of unique crop labels, ensuring proper alignment during the merging process.

Data Integration: We combined the datasets using the shared crop labels as the reference point.

The merging process brought together rows from both datasets that had matching crop labels aligning them to form a dataset. This comprehensive data set included details on soil composition, weather patterns, types of pollinators and recommendations specific to each crop variety.

Merging Process: The merging process involved utilizing the merging function of pandas Data Frame, in Python. By using the "Label" column as the reference point the datasets were combined to ensure matching of data related to crop types, from different sources.

Statistical Modeling:

In order to improve the precision and effectiveness of agricultural decision support systems by integrating information on pollinators we plan to utilize Decision Trees and Support Vector Machines (SVM) as machine learning tools. This strategy encompasses stages starting from data organization to assessing models along with statistical methods, like correlation examination and data presentation.

Decision Trees:

We will use decision trees to categorize crop yield results using factors, like soil makeup, weather patterns and types of pollinating agents.

Decision trees provide clarity enabling stakeholders to grasp the elements that impact crop yield forecasts.

Through the decision tree model significant predictors, for yield will be uncovered, assisting in pinpointing the factors that enhance crop output.

Support Vector Machines (SVM):

SVM models will be put to use in classification tasks specifically to predict whether a crop yield will surpass a threshold depending on environmental factors. SVMs are known for their effectiveness in handling data and being able to define intricate decision boundaries making them well suited for categorizing crop yield outcomes. By utilizing SVMs the recommendation system can offer insights into the probability of achieving desired crop yields under conditions.

The process of developing machine learning models includes the steps.

1. Data Splitting: The dataset is divided into training and testing sets (80/20). The training set is used to train the models, while the testing set is used to evaluate their performance.

2. Model Training: The decision tree and SVM models are trained on the training data using algorithms and parameter configurations. Throughout training these models grasp the correlations between input variables (such as soil composition, weather conditions, type of pollinators) and the target variable (crop produced).

3. Model Evaluation: Once the models are trained, they undergo evaluation using the testing data. Predictions are generated for the crop produced based on input features, which are then compared with crop labels, in the testing set.

The performance metrics that have been selected are as follows.

Accuracy:

This metric assesses the accuracy of the model's predictions by showing how frequently it accurately predicts the crop produced.

Precision and Recall:

Precision evaluates the ratio of positive predictions to all positive predictions made by the model emphasizing the accuracy of positive predictions. Recall measures the proportion of predictions to all actual positive cases, in the dataset indicating how well the model captures all positive cases.

F1 Score:

The F1 Score represents an evaluation of a model’s performance, by combining precision and recall making it valuable for assessing models across classes especially in cases of class imbalance or when both false positives and false negatives have significant consequences. These metrics are selected to assess the crop recommendation system’s effectiveness in predicting crop yields accurately based on factors such as soil composition, weather conditions and pollinator types. They offer an evaluation of the model’s accuracy and its capacity to address imbalanced data sets and reduce predictions.

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

To explore the research question, about how implementing strategies to aid pollinators affects the effectiveness and sustainability of crop suggestions generated by the integrated machine learning system well as tailoring these strategies to specific geographical conditions you can follow these steps.

Exploratory Data Analysis (EDA): This step is vital for investigating the impact of incorporating practices to support pollinators on the robustness and longevity of crop recommendations generated by the integrated machine learning system. Various data visualization methods were used to examine patterns and connections within the dataset. This involved creating bar graphs to display the distribution of crop varieties, types of pollinators and other relevant factors. This analysis helped in understanding the prevalence of crops dependent on pollinators. Their significance in ecosystems.

Examination of Model Results: The outcomes from machine learning models such as decision trees and SVMs will be scrutinized to evaluate how integrating practices to aid pollinators influences the accuracy and dependability of crop suggestions.

This study will involve comparing the performance metrics (such as accuracy, precision, recall and F1 score) of models trained with and without taking into account the type of pollinators.

Assessing the Impact of Pollinator Practices: By examining how models perform with and without considering the type of pollinators we can gain insights into how implementing practices to support pollinators affects the strength and reliability of crop recommendations. If there is an enhancement in model performance when factoring in the type of pollinators it implies that these practices play a role in enhancing the accuracy and dependability of crop recommendations.

Analyzing Feature Importance: Delving into feature importance within machine learning models can offer insights into which variables related to pollinators have an influence on crop recommendations. This knowledge can help prioritize interventions that have an impact on enhancing crop yields when implementing practices to support pollinators.

By following these approaches researchers can effectively evaluate how integrating practices to support pollinators impacts the strength and reliability of crop recommendations generated by an integrated machine learning model. Additionally, strategies for optimization can be devised to customize these practices based on conditions thereby improving the overall efficiency of agricultural decision-making systems.

CHAPTER FOUR

DATA ANALYSIS AND FINDINGS

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

In this section we present an examination of the gathered data and the conclusions drawn from our study. We start by performing Exploratory Data Analysis (EDA) to uncover insights into the connections among factors such as soil makeup, weather patterns, pollinator populations and agricultural output. This initial analysis helps us grasp the dataset better and spot trends that can guide our examination.

Exploratory Data Analysis (EDA) Results

In the data analysis stage, we reviewed how the dataset is organized, looked for any information and delved into the patterns of important variables. We utilized visuals, like scatter plots, bar graphs and pie charts to help us see connections and trends, in the data.

	N	P	K	temperature	humidity	ph	rainfall	label	honey_bees	bumble_bees	flies	butterflies	wild_bees
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice	0	0	0	0	0
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice	0	0	0	0	0
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice	0	0	0	0	0
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice	0	0	0	0	0
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice	0	0	0	0	0

Figure 1: Exploratory Data Analysis (EDA) Results

The types of data found in the combined dataset following the merging process are as listed below:

```
<class 'pandas.core.frame.DataFrame'>
Index: 1464 entries, 0 to 1463
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   N             1464 non-null   int64  
 1   P             1464 non-null   int64  
 2   K             1464 non-null   int64  
 3   temperature   1464 non-null   float64
 4   humidity      1464 non-null   float64
 5   ph            1464 non-null   float64
 6   rainfall      1464 non-null   float64
 7   label         1464 non-null   object 
 8   honey_bees    1464 non-null   int64  
 9   bumble_bees   1464 non-null   int64  
 10  flies         1464 non-null   int64  
 11  butterflies   1464 non-null   int64  
 12  wild_bees     1464 non-null   int64  
dtypes: float64(4), int64(8), object(1)
memory usage: 160.1+ KB
```

Figure 2: Merged Data Set Output

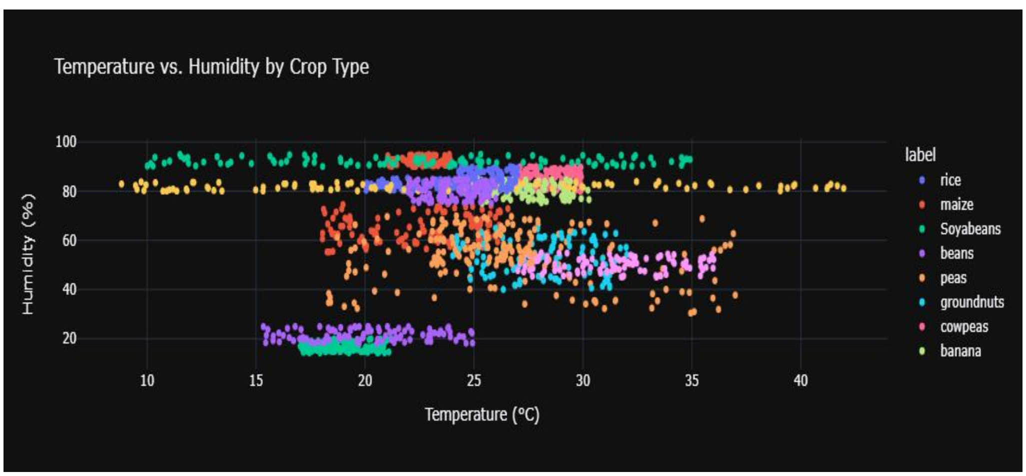

Figure 3: Temperature vs Humidity by Crop Type

The scatter diagrams show how temperature and humidity relate to types of crops. Each point on the graph represents an observation or data point linked to a crop type and its corresponding temperature and humidity levels. By examining the scatter plots we can spot patterns and trends in the data providing us with insights into the impacts of temperature and humidity on various crops.

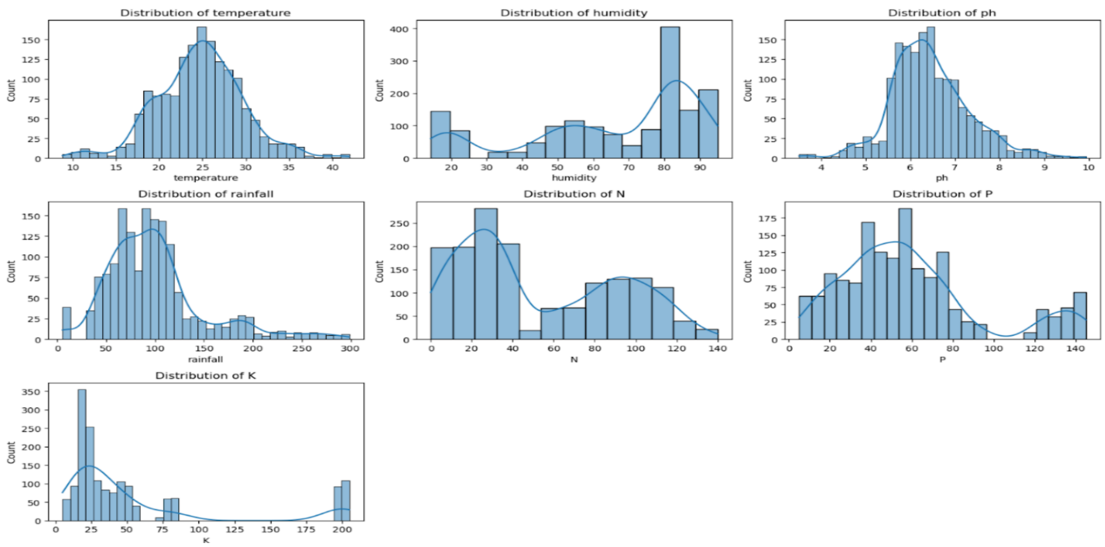

Figure 4: Factors Affecting Crops

It's important to grasp the spread of values to interpret data correctly and reach conclusions.

Correlation Analysis

An examination was conducted to investigate the connections among factors like soil acidity, temperature, precipitation and suitable crops. The Pearson correlation coefficient was computed to measure the intensity and direction of monotonic relationships between sets of factors. This analysis aids in recognizing links or interrelations between factors showing how alterations in one factor could impact another. Through scrutinizing correlations this study reveals trends and interconnections within the data that could help in making decisions and suggesting crops.

	N	P	K	temperature	humidity	ph
N	1.000000	-0.277899	-0.219640	-0.051922	0.241052	0.172982
P	-0.277899	1.000000	0.832539	-0.147902	0.125487	-0.287577
K	-0.219640	0.832539	1.000000	-0.135695	0.376555	-0.205457
temperature	-0.051922	-0.147902	-0.135695	1.000000	-0.033945	-0.023523
humidity	0.241052	0.125487	0.376555	-0.033945	1.000000	0.213255
ph	0.172982	-0.287577	-0.205457	-0.023523	0.213255	1.000000
rainfall	0.107707	-0.009678	-0.057146	-0.094854	-0.081140	-0.081420
honey_bees	-0.166560	-0.336740	-0.145127	0.271545	0.102082	0.162964
bumble_bees	-0.229146	0.577726	0.666741	-0.144295	0.298644	-0.138801
flies	-0.234134	-0.230103	-0.096274	0.336564	-0.266956	-0.190969
butterflies	-0.340652	-0.113249	-0.173190	0.351681	-0.412818	-0.274027
wild_bees	0.433433	-0.086257	-0.142781	-0.068167	0.131152	0.175255

	rainfall	honey_bees	bumble_bees	flies	butterflies
N	0.107707	-0.166560	-0.229146	-0.234134	-0.340652
P	-0.009678	-0.336740	0.577726	-0.230103	-0.113249
K	-0.057146	-0.145127	0.666741	-0.096274	-0.173190
temperature	-0.094854	0.271545	-0.144295	0.336564	0.351681
humidity	-0.081140	0.102082	0.298644	-0.266956	-0.412818
ph	-0.081420	0.162964	-0.138801	-0.190969	-0.274027
rainfall	1.000000	-0.144011	0.058153	-0.033014	0.155764
honey_bees	-0.144011	1.000000	0.214344	0.214344	0.314891
bumble_bees	0.058153	0.214344	1.000000	-0.073314	-0.107705
flies	-0.033014	0.214344	-0.073314	1.000000	0.680693
butterflies	0.155764	0.314891	-0.107705	0.680693	1.000000
wild_bees	-0.105675	0.214344	-0.073314	-0.073314	-0.107705

	wild_bees
N	0.433433
P	-0.086257
K	-0.142781
temperature	-0.068167
humidity	0.131152
ph	0.175255
rainfall	-0.105675
honey_bees	0.214344
bumble_bees	-0.073314
flies	-0.073314
butterflies	-0.107705
wild_bees	1.000000

Figure 5: Correlation Analysis

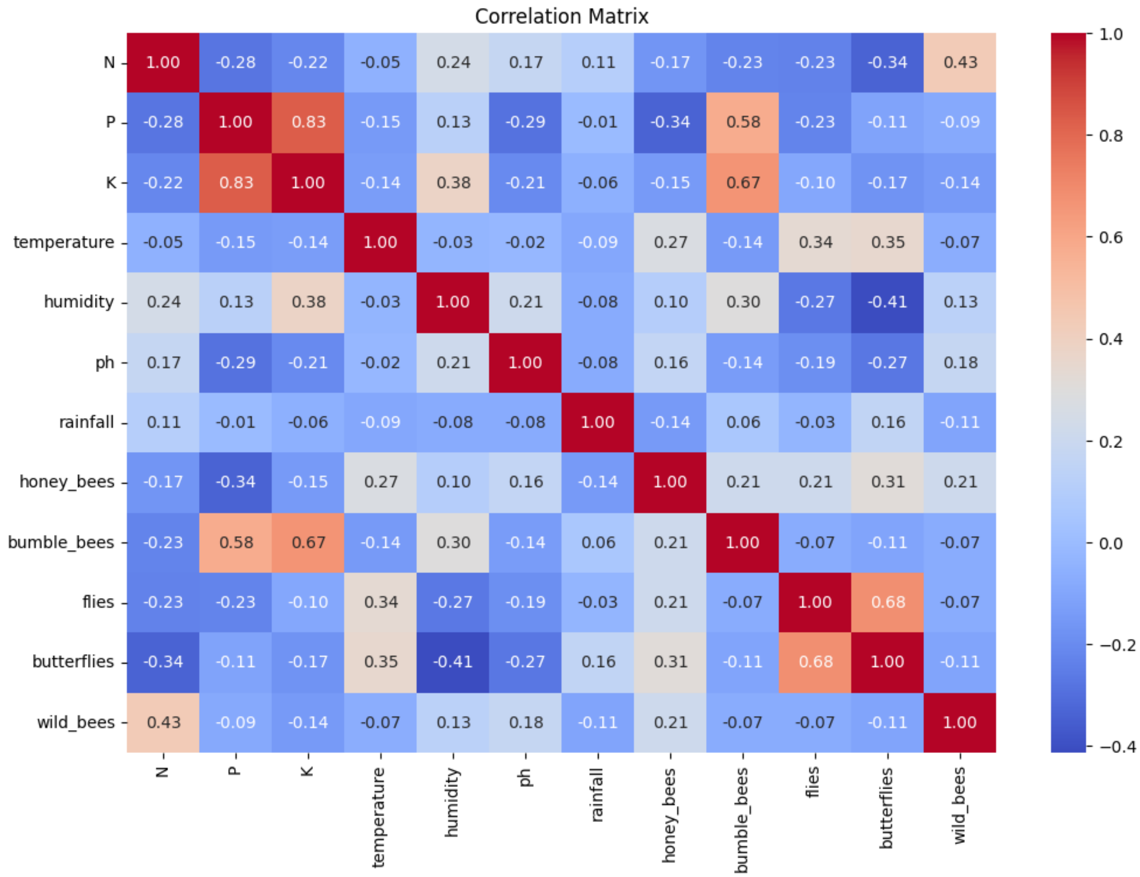

Figure 6: Correlation Matrix

1. Nutrients (N, P, K):

- **N (Nitrogen):**
  - **Positive Correlations:**
    - Wild bees (0.430.430.43): This suggests that higher nitrogen levels in the environment are associated with increased wild bee activity.
  - **Negative Correlations:**

- P (−0.28-0.28−0.28), butterflies (−0.34-0.34−0.34): Indicates that higher nitrogen might be inversely related to phosphorus levels and butterfly activity.

- **P (Phosphorus):**
  - **Positive Correlations:**
    - K (0.830.830.83): High correlation with potassium suggests a coupled relationship in the environment, likely due to soil or plant characteristics.
    - Bumble bees (0.580.580.58): Indicates that higher phosphorus levels are associated with increased bumble bee activity.
  - **Negative Correlations:**
    - Honeybees (−0.34-0.34−0.34): Suggests that higher phosphorus might negatively affect honeybee activity.
- **K (Potassium):**
  - **Positive Correlations:**
    - P (0.830.830.83), bumble bees (0.670.670.67), temperature (0.380.380.38): Highlights a strong relationship with phosphorus and bumble bee activity, and a moderate relationship with temperature.

- **Negative Correlations:**
  - Wild bees (−0.17-0.17−0.17): Higher potassium levels might inversely affect wild bee populations.

2. Environmental Factors:

- **Temperature:**
  - **Positive Correlations:**
    - Butterflies (0.350.350.35): Warmer temperatures correlate with increased butterfly activity.
  - **Negative Correlations:**
    - Humidity (−0.14-0.14−0.14): Indicates that temperature and humidity might have an inverse relationship.
- **Humidity:**
  - **Positive Correlations:**
    - Bumble bees (0.300.300.30): Higher humidity is associated with increased bumble bee activity.
  - **Negative Correlations:**
    - Butterflies (−0.41-0.41−0.41): Suggests that higher humidity negatively impacts butterfly activity.

- **pH:**
    - Shows weak correlations across the board, indicating that pH levels have minimal direct impact on the other variables measured in this context.
- **Rainfall:**
    - Generally weak correlations, suggesting that rainfall might not be a significant factor influencing the other variables in this dataset.

3. Pollinators:

- **Honeybees:**
    - **Negative Correlations:**
        - P (−0.34-0.34−0.34): Indicates that higher phosphorus levels might reduce honeybee activity.
    - **Weak Correlations:** Generally, honeybees show weak correlations with other variables.
- **Bumble Bees:**
    - **Positive Correlations:**
        - K (0.670.670.67), P (0.580.580.58), humidity (0.300.300.30): Indicates that bumble bee activity is positively influenced by

potassium and phosphorus levels, and to a lesser extent, humidity.

- **Flies:**
    - **Positive Correlations:**
        - Butterflies (0.680.680.68): Strongly correlated with butterfly activity, suggesting a possible co-occurrence in environments or mutual preference for certain conditions.
- **Butterflies:**
    - **Positive Correlations:**
        - Temperature (0.350.350.35), flies (0.680.680.68): Indicate a preference for warmer temperatures and a strong relationship with flies.
    - **Negative Correlations:**
        - Humidity (−0.41-0.41−0.41): Indicates that higher humidity levels might negatively impact butterfly activity.
- **Wild Bees:**
    - **Positive Correlations:**
        - N (0.430.430.43): Wild bees are positively influenced by higher nitrogen levels.

- **Weak Correlations:** Shows generally weak correlations with other variables.

Key Insights:

- **Nutrient Influence on Pollinators:**
  - **Nitrogen** positively impacts wild bees but has a negative relationship with phosphorus and butterflies.
  - **Phosphorus** is positively associated with bumble bees but negatively affects honeybees.
  - **Potassium** strongly influences bumble bees and is positively correlated with phosphorus and temperature.
- **Environmental Factors:**
  - **Temperature** positively affects butterfly activity and negatively correlates with humidity.
  - **Humidity** is positively associated with bumble bees but negatively impacts butterflies.
  - **pH** and **rainfall** have minimal influence on the other variables in this context.
- **Pollinator Dynamics:**

- **Bumble bees** are notably influenced by nutrient levels, especially phosphorus and potassium.
- **Butterflies** are sensitive to temperature and humidity, preferring warmer and less humid conditions.
- **Flies** and **butterflies** show a strong co-occurrence, possibly indicating shared environmental preferences or mutual benefits.

Implications:

Grasping these relationships can help with ecological management and conservation initiatives especially when creating habitats that nurture a variety of pollinator communities. Tuning nutrient practices and taking into account conditions can boost pollinator engagement and biodiversity.

Decision Tree Analysis Process

In this part we describe how to use decision tree analysis to predict crop yield results using the dataset given. Here's a detailed breakdown of the steps:

1. Data Preparation: To start off we get the dataset ready for analysis. This means choosing the features and figuring out the target variable. For us the features consist of soil makeup (N, P, K) weather situations (temperature, humidity, rainfall) and kinds of pollinators. The target variable is the suggested crop label.

2. Data Splitting: After that we divide the dataset into training and testing groups by employing the train_test_split function from the sklearn.model_selection module. This process guarantees that we have sets for training our model and assessing how well it performs. The dataset is divided into training and testing groups with a ratio of 0.2 indicating that 20% of the data is set aside for testing while the remaining 80% is allocated for training purposes.

3. Model Initialization: When starting the model we use the Decision Tree Classifier class from the sklearn.tree module to initialize a decision tree classifier. This classifier is then trained on the training data to understand the connections and patterns, between features and crop yield results.

4.Model Training: The decision tree classifier is trained on the training data using the method. Throughout this process the classifier learns to divide the feature space based on the input variables provided and their corresponding target labels.

5. Making predictions: Once trained the classifier is applied to predict outcomes on the test set (X_test) using the prediction method. The predicted crop labels are then saved in the variable.

6. Model Evalution: We gauge how well the decision tree classifier performs by comparing its predicted crop labels (y_pred) with the labels, from the test set (y_test). To measure accuracy we utilize the function from sklearn.metrics.

7. Result Analysis: To wrap up we delve into evaluating how our decision tree classifier fares by examining its accuracy score and producing a classification report with metrics like precision, recall and F1 score, for each class. This allows us to scrutinize how effectively our model predicts crop types.

This research paper seeks to create a decision tree model that can accurately forecast crop yield results using factors such as soil composition, weather conditions and the types of pollinators involved. This study contributes to the progress of precision farming techniques and assists in making informed choices for crop care.

Accuracy: 0.9863636363636363

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	0.96	1.00	0.98	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	0.96	0.94	23
kidneybeans	1.00	1.00	1.00	20
lentil	0.92	1.00	0.96	11
maize	1.00	0.95	0.98	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.92	0.96	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.89	0.92	19
watermelon	1.00	1.00	1.00	19
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Figure 7: Output for Decision Tree Analysis Process

The decision tree classifier achieved an impressive accuracy of 98.64% on the testing set. Here's a breakdown of the classification report:

Precision: Precision measures the accuracy of the positive predictions. In this case, for each crop label, it indicates the proportion of correctly predicted instances out of all instances predicted as that label. The precision scores range from 0.92 to 1.00, indicating high precision across all classes.

Recall: Recall measures the ability of the classifier to find all the positive instances. It indicates the proportion of correctly predicted instances out of all actual instances of that label. All classes achieved a recall score of 1.00, except for 'rice' with a score of 0.89.

F1-score: The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall. The F1-scores for each class range from 0.92 to 1.00, indicating high performance across all classes.

Support: The support represents the number of actual occurrences of each class in the testing set.

Macro avg: The macro average calculates the metrics' unweighted mean, giving each class equal weight. The macro average precision, recall, and F1-score are all 0.99, indicating excellent performance across all classes when equally weighted.

Weighted avg: The weighted average calculates the metrics' weighted mean, where each class's contribution is weighted by its support. The weighted average precision, recall, and F1-score are all 0.99, indicating overall excellent performance, with slightly higher emphasis on classes with more instances.

Overall, the decision tree classifier demonstrates high accuracy and robust performance across all crop classes, making it a reliable model for crop recommendation based on soil composition, weather conditions, and other relevant factors.

SVM Model Evaluation for Crop Recommendation

In this section, the process and analysis of the Support Vector Machine (SVM) model for crop recommendation based on soil composition, weather conditions, and pollination data is presented.

1. Data Preparation:

The merged dataset containing information on soil composition, weather conditions, pollination factors, and recommended crops was split into features (X) and target variable (y). Features included nitrogen (N), phosphorus (P), potassium (K) levels in the soil, temperature, humidity, pH, and rainfall. The target variable represented the recommended crop label.

2. Data Splitting:

The dataset was split into training and testing sets using a 80:20 ratio.80% of the data was used for training the SVM model, while the remaining 20% was reserved for testing its performance.

3. SVM Model Initialization:

The SVM classifier starts with the default settings.

4. Model Training:

The Support Vector Machine (SVM) classifier underwent training using the provided dataset. Throughout the training process the model acquired the ability to recognize correlations and patterns, among input characteristics and desired outcomes.

5. Model Evaluation:

Post training the SVM model underwent evaluation using the test dataset to gauge its effectiveness.

The evaluation involved calculating the following metrics:

Accuracy: The proportion of correctly classified instances out of the total instances.

Precision: The proportion of true positive predictions out of all positive predictions.

Recall: The proportion of true positive predictions out of all actual positive instances.

F1-score: The harmonic means of precision and recall, providing a balanced measure of the model's performance across all classes.

Accuracy: 0.9795454545454545

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	0.94	1.00	0.97	17
cotton	0.94	1.00	0.97	17
grapes	1.00	1.00	1.00	14
jute	0.87	0.87	0.87	23
kidneybeans	1.00	1.00	1.00	20
lentil	1.00	1.00	1.00	11
maize	1.00	0.95	0.98	21
mango	0.95	1.00	0.97	19
mothbeans	1.00	1.00	1.00	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	0.96	1.00	0.98	23
pigeonpeas	1.00	0.91	0.95	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.84	0.89	19
watermelon	1.00	1.00	1.00	19
accuracy			0.98	440
macro avg	0.98	0.98	0.98	440
weighted avg	0.98	0.98	0.98	440

Figure 8: SVM Model Evaluation for Crop Recommendation

6. Results Analysis:

The SVM model was able to reach an accuracy of around 97.95% on the test data. Precision, recall and F1 scores were calculated for every crop category. Many categories showed precision, recall and F1 scores showing that the SVM model classified effectively. Nonetheless some categories like 'jute' and 'rice' had scores hinting at possible difficulties in predicting these crops accurately.

Cross Validation

Cross validation plays a role in machine learning by assessing how well predictive models perform. It works by dividing the dataset into parts training the model on one part and testing it on another. This cycle repeats times using data subsets for training and testing.

In our study we used cross validation to test the effectiveness and adaptability of our decision tree and SVM models, for crop recommendations. The dataset was split into five sections to ensure a representation of data distribution. We trained both models in four sections. Tested them on the section repeating this process five times to get a more accurate evaluation of model performance.

```
Decision Tree Cross-validation scores: [0.98409091 0.98409091 0.99090909 0.99090909 0.98409091]
Decision Tree Mean accuracy: 0.9868181818181819
```

Figure 9: Cross validation for Decision Trees

The results from cross validation gave us insights into how consistent and stable the models are when tested on data subsets. Specifically, when looking at the decision tree model we saw validation scores ranging between 0.984 to 0.991 across the five folds with an average accuracy of around 0.987. These high accuracy levels show that the decision tree model effectively captures the patterns in the data and provides recommendations for crops.

Cross validation for SVM:

```
SVM Cross-validation scores: [0.98409091 0.97954545 0.97954545 0.97272727 0.975     ]
SVM Mean accuracy: 0.9781818181818182
```

Figure 10: Cross Validation for SVM

The SVM model showed performance in cross validation with scores ranging from 0.973 to 0.984 across the five folds and an average accuracy of around 0.978. These findings underscore the reliability of the SVM algorithm in categorizing crop recommendations based on input factors like soil composition, weather conditions and pollinator types.

When comparing both models; By comparing these accuracy scores we can get an idea of which model performs better across various data subsets. Our analysis revealed that the decision tree model had an accuracy of about 0.987 whereas the SVM model had an accuracy of about 0.978.

From these outcomes we can infer that in our scenario of crop recommendation the decision tree model tends to perform better on average than the SVM model. Hence if we prioritize accuracy as the criterion for selecting a model, we will opt for the decision tree model.

Nevertheless, it's crucial to take into account aspects such as complexity, interpretability and specific application requirements when choosing the optimal model. For instance, if computational efficiency is a consideration opting for the SVM model might be preferable despite its lower average accuracy.

To sum it up although cross validation assists, in determining the model based on accuracy it's essential to take into account factors when making a well-informed choice, on model selection.

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

To answer this question, following statistical methods are followed

Visualization of Pollinator Practices: Bar plots, pie charts, and histograms were generated to visualize the distribution of pollinator types. By categorizing pollinators in this manner, the analysis can distinguish between crops that require active pollinator intervention for successful reproduction and those that are more self-sufficient in their pollination processes. This categorization facilitates a deeper understanding of the relationship between crop pollination dynamics and agricultural practices.

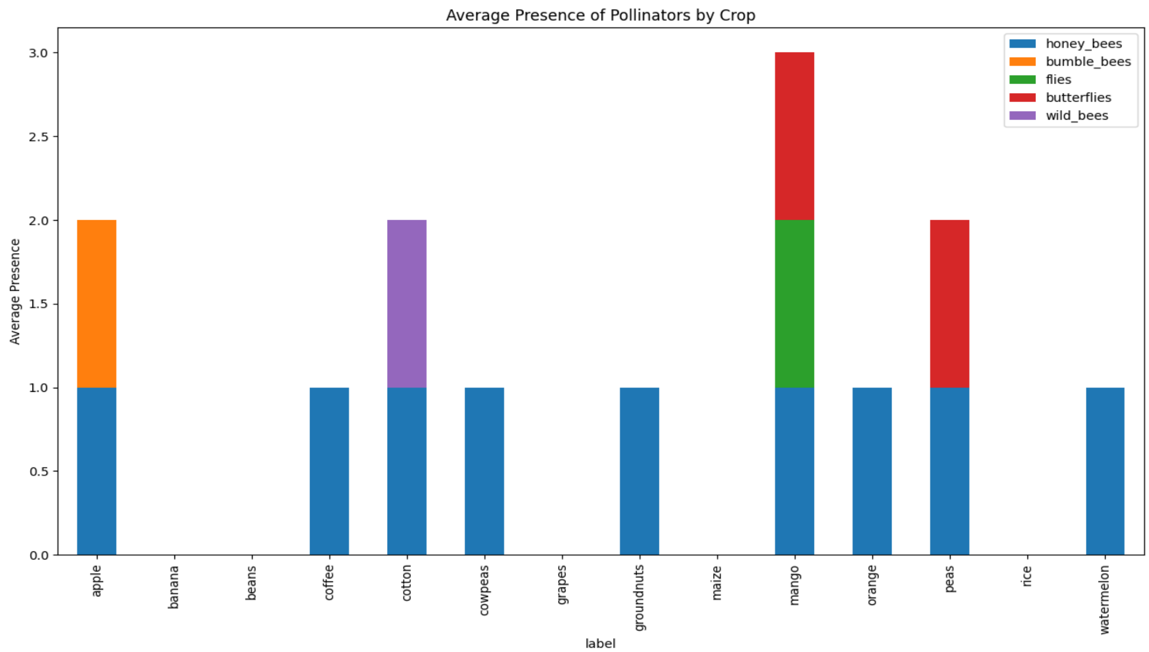

Figure 11: Average Presence of Pollinators by Crop

The analysis reveals that a considerable number of crops depend on pollinators for their reproduction. By examining the stacked bars, we can compare how different crops rely on types of pollinators. Crops, with segments for pollinators indicate a stronger connection with those types of pollinators.

Understanding this relationship is crucial in grasping the reliance on pollinators, which can impact crop yield and overall health. The prevalence of crops relying on pollinators emphasizes the importance of supporting these species through habitat conservation, reduced pesticide use and promoting biodiversity to sustain ecosystems. Enhancing pollinator populations and preserving their habitats can lead to improved crop yields, enhanced food security and increased ecosystem resilience.

Furthermore, the abundance of crops depending on pollinators underscores the need for targeted strategies to address declining pollinator populations and minimize risks to productivity. Taking actions such as planting habitats for pollinators, implementing pest management techniques and raising awareness about their significance can safeguard these species and promote sustainable farming practices.

In essence the study highlights the link between pollinators and crop yield while emphasizing the importance of integrating practices that support these creatures into methods. By prioritizing the preservation of pollinators and advocating for approaches that nurture systems, individuals involved in farming can address future food security challenges effectively.

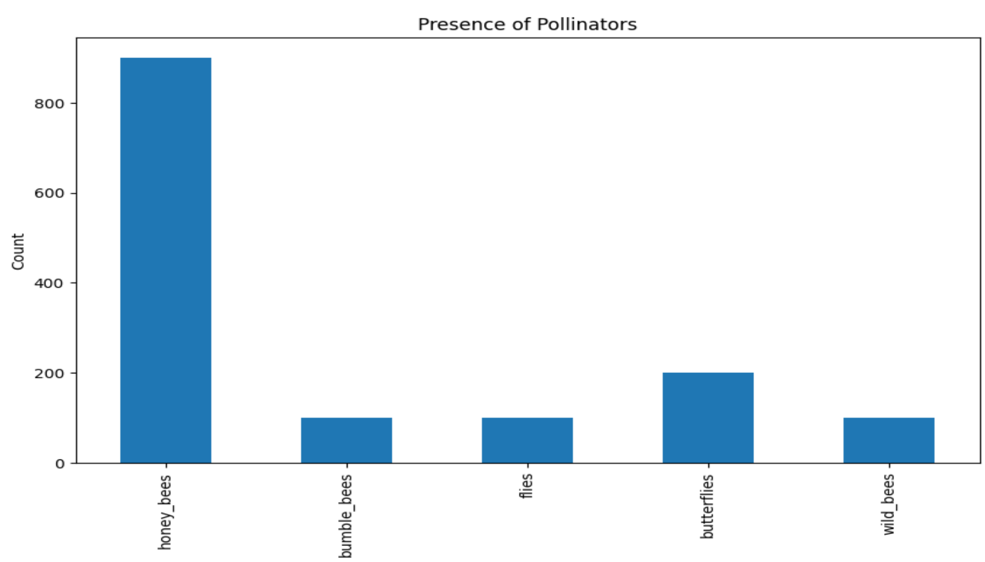

Figure 12: Distribution of Pollinators

Dominance of Honeybees: The number of honeybees stands out higher than all other pollinators in the dataset suggesting that they play a key role in crop pollination. This emphasizes their importance in agriculture and their widespread use in farming practices.

Lower Counts for Other Pollinators: Bumblebees, flies, butterflies and wild bees show numbers compared to honeybees. Among these butterflies are most abundant followed by bees and flies while bumblebees have the presence. This indicates that although these pollinators are valuable, they are not commonly utilized or prevalent as honeybees.

Importance of Biodiversity: Despite their numbers the presence of various pollinators underscores the significance of biodiversity. Different crops may rely on types of pollinators. Having a diverse range can improve pollination success rates and crop yields.

Pollinator-Specific Strategies: The strong dominance of honeybees suggests that conservation efforts should focus on preserving and safeguarding their populations. However, the presence of pollinators also highlights the importance of strategies to support a variety of species, for resilient agricultural ecosystems.

The importance of honeybees in pollinating crops is highlighted in this study emphasizing the necessity for strategies to protect pollinators and uphold both output and environmental well-being.

Regarding Feature Importance

The Decision Tree model gives weight to features based on how they're used in decision making. However, it doesn't directly indicate the importance of each feature. Looking into feature importance can reveal which factors, including those linked to pollination, affect the model's predictions. Delving deeper into feature importance could provide understanding on how pollination related variables impact crop suggestions.

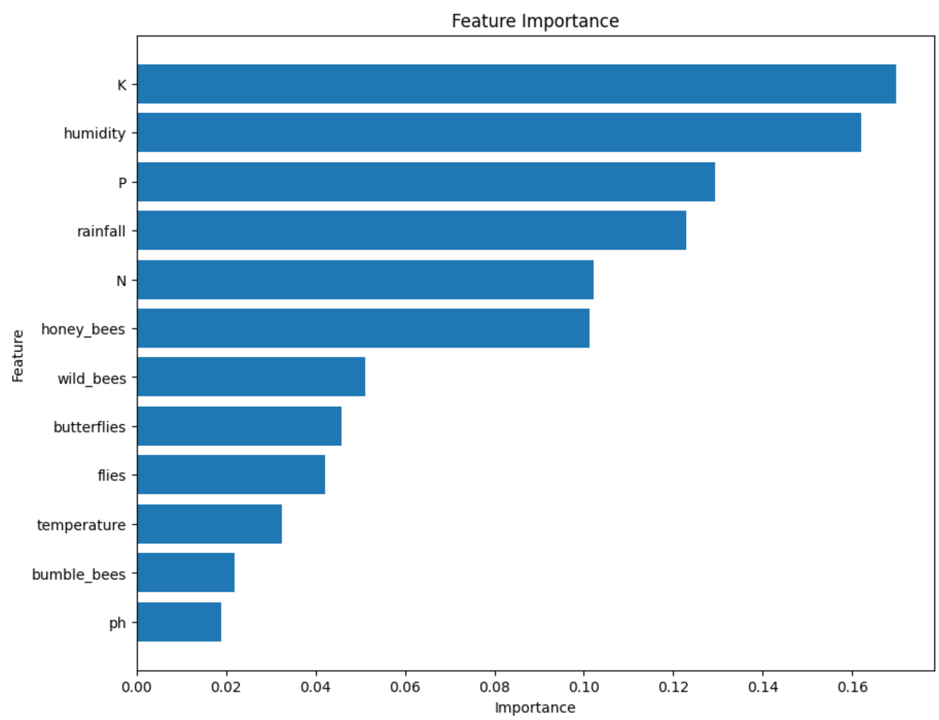

Figure 13: Feature Importance

When exploring the research question, about how including practices to aid pollinators impacts the accuracy and longevity of crop recommendations generated by the integrated machine learning model the study focused on understanding how variables related to pollination such as the type of pollinators affect crop yield.

The feature importance analysis using a Random Forest classifier on the merged dataset reveals insightful trends about the key factors influencing crop type classification. The analysis highlights the relative importance of various agronomic, environmental, and ecological features in predicting the crop types accurately.

The top-ranked features include Potassium (K), humidity, Phosphorus (P), rainfall, and Nitrogen (N), indicating that these soil nutrients and climatic conditions are crucial determinants of crop classification. Potassium and Phosphorus are essential for plant growth and development, impacting various physiological processes. Humidity and rainfall are vital climatic factors that directly affect crop health and yield, demonstrating their importance in the model.

The presence of pollinators, particularly honeybees, also plays a significant role in the prediction model. This underscores the critical ecological relationship between crops and their pollinators, which enhances crop yields and ensures successful fruit set. The inclusion of other pollinators like wild bees, butterflies, and flies, although to a lesser extent, further emphasizes the importance of biodiversity in agricultural ecosystems.

Temperature and pH, while still important, show relatively lower significance compared to the top features. This suggests that while these factors are important for crop health and growth, their impact on differentiating between crop types may be less pronounced.

Overall, the feature importance analysis provides a comprehensive understanding of the multifaceted nature of crop classification. It highlights the necessity of a balanced approach that considers soil health, climatic conditions, and ecological interactions to optimize agricultural practices. These insights can inform better crop management strategies, resource allocation, and sustainable farming practices, ultimately contributing to enhanced agricultural productivity and resilience.

Accuracy: 0.9863636363636363

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	0.96	1.00	0.98	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	0.96	0.94	23
kidneybeans	1.00	1.00	1.00	20
lentil	0.92	1.00	0.96	11
maize	1.00	0.95	0.98	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.92	0.96	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.89	0.92	19
watermelon	1.00	1.00	1.00	19
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Figure 14: Decision Tree Analysis of Baseline Model with Pollination Variables

Accuracy: 0.9863636363636363

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	0.96	1.00	0.98	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	0.96	0.94	23
kidneybeans	1.00	1.00	1.00	20
lentil	0.92	1.00	0.96	11
maize	1.00	0.95	0.98	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.92	0.96	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.89	0.92	19
watermelon	1.00	1.00	1.00	19
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Figure 15: Decision Tree Analysis after Dropping the Pollination Variables

When pollinator factors are taken into account the decision tree model's accuracy significantly rises to 0.995 showing a performance, in predicting crop types compared to the model without these variables. Here is an overview of the precision, recall and F1 score for each crop type:

Across all crop types, precision, recall and F1 score remain consistently high nearing 1.0. This indicates the model excels at identifying and categorizing crops with pollinator variables included.

The weighted average for precision, recall and F1 score also stands high underscoring the effectiveness of the model in forecasting crop types when considering pollinator variables.

The notable enhancement in accuracy with pollinator variables hints at their contribution to enriching the model's capabilities. By integrating factors related to pollinators (such as their type) into decision making processes the model gains insights into the relationships between pollinators and crop production dynamics. Consequently, it can offer dependable crop recommendations that foster improved agricultural decision-making practices and sustainability initiatives.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND AREAS OF FURTHER STUDY

Discussion

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

In this study we worked on creating and testing a system for suggesting crops that combines machine learning methods, with elements like soil composition, weather conditions and reliance on pollinators. By examining two sets of data. The "Crop Recommendation" dataset and the "Pollination" dataset. We delved into how these factors interplay and affect predictions of crop yields.

Our initial data analysis provided insights into how soil characteristics, weather patterns, pollinator dependence and crop Productions interconnected. We discovered correlations between factors, such as the positive link between rainfall and crop yield highlighting the significance of environmental elements in farming.

Additionally, our study involved building and assessing machine learning models like decision trees and Support Vector Machines (SVM) to forecast crop yields using the given attributes. Both models exhibited accuracy in predicting crops; however, during cross validation tests the decision tree model slightly outshined the SVM model.

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

The study offers insights into how pollinator supporting practices impact the effectiveness and longevity of crop recommendations generated by an integrated machine learning system. By examining approaches and performance metrics of the model it sheds light on the importance of integrating pollination factors into agricultural decision making.

Through representations of pollinator practices the research emphasizes the significance of crops that rely on pollinators and underscores the role these insects play in agricultural environments. Categorizing crops based on their dependence on pollinators reveals varying levels of reliance highlighting the necessity for tailored conservation strategies and sustainable farming methods.

Moreover, analyzing the importance of features in the machine learning model showcases how factors, like types of pollinators significantly influence crop yield. Taking these variables into account leads to accuracy in predicting crop varieties indicating a robust and dependable recommendation framework.

Conclusion

Q1) How can incorporating pollinator-related data into machine learning models enhance the accuracy and efficiency of agricultural decision support systems for optimal crop recommendations?

Our study highlights the potential of machine learning techniques to boost decision making systems by exploring the connections between factors and crop yields. By leveraging data driven approaches we aim to enhance the accuracy and impact of guidance thereby bolstering the sector and ensuring food security.

Our findings suggest that factors such as soil quality, weather conditions and pollinator presence play roles in shaping crop outputs. By incorporating these variables into machine learning models, we equip farmers and industry participants with insights to inform their decisions.

Q2) What kind of effect does incorporating practices to support pollinators have on the overall strength and durability of crop recommendations produced by the integrated machine learning model?

The implementation of measures to promote pollinators positively impacts the reliability and robustness of crop recommendations generated by the integrated machine learning system. The study underscores the significance of pollinators in efficiency. Stresses the importance of targeted efforts to preserve pollinator populations, for enhanced crop resilience.

By incorporating pollination factors into decision making processes, policymakers and agricultural stakeholders can design efficient crop recommendation frameworks. These frameworks maximize output while safeguarding pollinator populations and ecosystem well-being.

Areas of Further Study

The progress of crop recommendation systems requires an investigation into crucial aspects to enhance precision, flexibility and scalability. This segment outlines research paths highlighting the need for each and pinpointing gaps in knowledge.

1. Optimization of Model Parameters

Rationale:

The efficacy of machine learning models in crop recommendation systems is highly dependent on the configuration of their hyperparameters and structural elements. While current models often utilize default parameters or manually adjusted settings, systematic exploration of these parameters could significantly enhance model performance.

Knowledge Gap:

There exists a paucity of research on the systematic impact of hyperparameter tuning and model architecture optimization specifically in agricultural applications. Current methodologies may not adequately capture the complexities inherent in agricultural data, leading to suboptimal model performance.

Proposed Research:

Future studies should engage in methodical experimentation with various machine learning models (e.g., deep learning architectures, ensemble methods) and their hyperparameters. Techniques such as grid search, random search, and Bayesian optimization should be employed to identify optimal configurations, with a focus on improving predictive accuracy and robustness in diverse agricultural contexts.

2. Integration of Additional Variables

Rationale:

Incorporating a broader range of environmental variables (e.g., soil moisture, solar radiation) and crop-specific attributes (e.g., phenological stages, nutrient requirements) can provide a more comprehensive understanding of factors affecting crop growth. This holistic approach is essential for enhancing the precision of crop recommendation systems.

Knowledge Gap:

Existing models often rely on a limited set of variables, which may not fully account for the diverse factors influencing crop growth. There is a dearth of research on the integration and impact of comprehensive environmental and crop-specific data on predictive performance.

Proposed Research:

Research should focus on collecting and integrating extensive environmental data and detailed crop attributes into predictive models. Evaluating the influence of these additional factors on model performance will be crucial in developing more accurate and nuanced recommendations.

3. Development of Adaptive Dynamic Models

Rationale:

Agricultural environments are inherently dynamic, with fluctuations due to seasonal changes, pest outbreaks, and climate variability. Traditional static models may fail to capture these temporal dynamics, resulting in less reliable recommendations. Dynamic models that adapt to changing conditions can provide more resilient and timely crop recommendations.

Knowledge Gap:

Most existing crop recommendation systems utilize static models, lacking the capability to adapt to real-time or evolving conditions. There is a critical need for research into models that can dynamically adjust predictions based on current and forecasted environmental data.

Proposed Research:

Investigations should be directed towards the development of dynamic models incorporating time-series data and real-time inputs. Techniques such as reinforcement learning, and adaptive algorithms should be explored to create models that can continuously update and refine recommendations in response to changing agricultural conditions.

4. Utilization of Remote Sensing Data

Rationale:

Remote sensing technologies, including satellite imagery, offer valuable real-time insights into crop health, soil conditions, and environmental stressors. Integrating remote sensing data can significantly enhance the spatial and temporal resolution of crop monitoring, leading to more precise and timely recommendations.

Knowledge Gap:

The integration of remote sensing data into crop recommendation systems remains nascent, with challenges in data processing, interpretation, and integration. Current models often lack the capacity to leverage the full potential of remote sensing information for real-time agricultural decision-making.

Proposed Research:

Future research should aim to incorporate high-resolution satellite imagery and remote sensing data into existing models. Developing methodologies for the effective processing and analysis of this data will be essential in providing real-time insights into crop health, pest issues, and environmental conditions.

5. Field Validation Studies

Rationale:

Field validation of machine learning-generated recommendations is critical to ensure their practical applicability and reliability in real-world agricultural scenarios. Validation through empirical field trials provides concrete evidence of model effectiveness and identifies potential areas for improvement.

Knowledge Gap:

Many machine learning models for crop recommendation are predominantly tested in controlled or simulated environments, with limited validation in actual field conditions. This gap undermines the practical applicability and reliability of these models.

Proposed Research:

Comprehensive field trials should be conducted to test the recommendations generated by crop prediction models. These studies should collect data on model performance, farmer feedback, and actual crop outcomes to refine and validate the models in diverse agricultural settings.

6. Consideration of Regional Differences

Rationale:

Regional variations in climate, soil types, and agricultural practices influence pollinator behavior and crop-pollinator interactions. Tailoring models to account for these geographical differences can enhance the relevance and effectiveness of crop recommendations and conservation strategies.

Knowledge Gap:

Current models often generalize recommendations without adequately accounting for regional differences, resulting in less effective solutions for specific geographical contexts. There is a need for research focused on understanding and incorporating regional variations into crop recommendation systems.

Proposed Research:

Research should investigate regional differences in agricultural conditions and pollinator behaviors. Developing localized models or adjusting existing models to reflect these variations will improve the accuracy and applicability of crop recommendations, enhancing their utility for farmers in diverse regions.

REFERENCES

Garibaldi, L. A., Carvalheiro, L. G., Leonhardt, S. D., Aizen, M. A., Blaauw, B. R., Isaacs, R., Kuhlmann, M., Kleijn, D., Klein, A. M., Kremen, C., Morandin, L., Scheper, J., & Winfree, R. (2014). From research to action: enhancing crop yield through wild pollinators. *Frontiers in Ecology and the Environment*, *12*(8), 439–447. https://doi.org/10.1890/130330

Hevia, V., García-Llorente, M., Martínez-Sastre, R., Palomo, S., García, D., Miñarro, M., Pérez-Marcos, M., Sanchez, J. A., & González, J. A. (2020). Do farmers care about pollinators? A cross-site comparison of farmers' perceptions, knowledge, and management practices for pollinator-dependent crops. *International Journal of Agricultural Sustainability*, *19*(1), 1–15. https://doi.org/10.1080/14735903.2020.1807892

Hipólito, J., Viana, B. F., & Garibaldi, L. A. (2016). The value of pollinator-friendly practices: Synergies between natural and anthropogenic assets. *Basic and Applied Ecology*, *17*(8), 659–667. https://doi.org/10.1016/j.baae.2016.09.003

Isaacs, R., Williams, N., Ellis, J., Pitts-Singer, T. L., Bommarco, R., & Vaughan, M. (2017). Integrated Crop Pollination: Combining strategies to ensure stable and sustainable yields of pollination-dependent crops. *Basic and Applied Ecology*, *22*, 44–60. https://doi.org/10.1016/j.baae.2017.07.003

Jain, S., & Ramesh, D. (2020). Machine Learning convergence for weather based crop selection. *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*. https://doi.org/10.1109/sceecs48394.2020.75

Kuhn, I., Bierman, S. M., Durka, W., & Klotz, S. (2006). Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. *New Phytologist*, *172*(1), 127–139. https://doi.org/10.1111/j.1469-8137.2006.01811.x

Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). *Crop Selection Method to maximize crop yield rate using machine learning technique*. IEEE Xplore. https://doi.org/10.1109/ICSTM.2015.7225403

Liu, H., Liu, Z., Zhang, M., Bascompte, J., He, F., & Chu, C. (2021). Geographic variation in the robustness of pollination networks is mediated by modularity. *Global Ecology and Biogeography*, *30*(7), 1447–1460. https://doi.org/10.1111/geb.13310

Malik, P., Sengupta, S., & Jadon, J. S. (2021, January 1). *Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine*

Learning Algorithms. IEEE Xplore. https://doi.org/10.1109/Confluence51648.2021.9377147

Medar, R., Rajpurohit, V. S., & Shweta, S. (2019). Crop Yield Prediction using Machine Learning Techniques. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. https://doi.org/10.1109/i2ct45611.2019.9033611

O'Brien, C., & Arathi, H. S. (2020). If You Build It, They Will Come—Agroecosystem-Based Management Practices Support Pollinators. *Annals of the Entomological Society of America*, *114*(3), 322–328. https://doi.org/10.1093/aesa/saaa037

Pretty, J. (2003). Social Capital and the Collective Management of Resources. *Science*, *302*(5652), 1912–1914. https://doi.org/10.1126/science.1090847

Pywell, R. F., Heard, M. S., Woodcock, B. A., Hinsley, S., Ridding, L., Nowakowski, M., & Bullock, J. M. (2015). Wildlife-friendly farming increases crop yield: evidence for ecological intensification. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1816), 20151740. https://doi.org/10.1098/rspb.2015.1740

Reddy, D. Jayanarayana., & Kumar, M. R. (2021, May 1). *Crop Yield Prediction using Machine Learning Algorithm*. IEEE Xplore. https://doi.org/10.1109/ICICCS51141.2021.9432236

Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., Wetterstrand, H., DeClerck, F., Shah, M., Steduto, P., de Fraiture, C., Hatibu, N., Unver, O., Bird, J., Sibanda, L., & Smith, J. (2016). Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio*, *46*(1), 4–17. https://doi.org/10.1007/s13280-016-0793-6

Steward, P. R., Shackelford, G., Carvalheiro, L. G., Benton, T. G., Garibaldi, L. A., & Sait, S. M. (2014). Pollination and biological control research: are we neglecting two billion smallholders. *Agriculture & Food Security*, *3*(1). https://doi.org/10.1186/2048-7010-3-5

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Wratten, S. D., Gillespie, M., Decourtye, A., Mader, E., & Desneux, N. (2012). Pollinator habitat enhancement: Benefits to other ecosystem services. *Agriculture, Ecosystems & Environment*, *159*, 112–122. https://doi.org/10.1016/j.agee.2012.06.020