National College of Ireland

Enhancing Agriculture Production Engine using Machine learning with Robust Cloud Security Measures

MSc Research Project
MSc Cloud Computing

Deva Loknathan Marri
Student ID: x21231346

School of Computing
National College of Ireland

Supervisor: Ahmed Makki

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Deva Loknathan Marri
Student ID:	x21231346
Programme:	MSc Cloud Computing
Year:	2023
Module:	MSc Research Project
Supervisor:	Ahmed Makki
Submission Due Date:	14/12/2023
Project Title:	Enhancing Agriculture Production Engine using Machine learning with Robust Cloud Security Measures
Word Count:	5987
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL Internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use another author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	14th December 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Agriculture Production Engine using Machine learning with Robust Cloud Security Measures

Deva Loknathan Marri
x21231346

Abstract

Secure and better crop recommendations in agriculture are made possible by this research's Agricultural Optimization Engine, which integrates machine learning with partial homomorphic encryption for cloud security. Our new technique was motivated by the fact that traditional approaches often Endanger data privacy. As part of our contribution, In this process, partial homomorphic encryption provides safe data processing while deploying an accurate logistic regression model for crop suggestions. The effectiveness of the system in making precise suggestions while protecting sensitive data is shown by our findings. In order to maximize agricultural productivity, this study lays the framework for future developments in safe data analytics by bridging the gap between machine learning and data security in agriculture. The solution it provides to farming businesses is both practical and scalable. Implications for both theory and practice abound in the proposed system, which aims to improve the confidentiality of agricultural data while also giving farmers access to a safe and dependable resource for managing their crops.

1 Introduction

Cloud computing is at the heart of the modern agricultural revolution, driving precision farming through technological advancements. This integration introduces new possibilities and challenges in the era of digital farming. The project's primary focus is on securing the vast amounts of sensitive data generated by agricultural operations and exploring the intricate relationship between cloud computing and agriculture.

Understanding the history of farming and how cloud computing will play a significant part in the industry's future is covered extensively in the Background section. In the part under "Importance," the process has seen how this integration may change things for the better, bringing about a revolution in resource management, increased production, and more sustainable agricultural methods. This lays the groundwork for a sophisticated comprehension of the agricultural environment and the need to safeguard the data that supports this digital transformationLiu et al. (2019).

Traditional encryption requires data decryption for processing, posing security risks as the decrypted data becomes vulnerable in the cloud. Homomorphic encryption resolves this by allowing cloud calculations on encrypted data, ensuring advanced cryptographic security. With homomorphic encryption, the cloud never sees the decrypted data, and computation outcomes are visible to the client. Lattice-based cryptography enhances

the secure computation of encrypted data. Unlike current cryptosystems, homomorphic encryption enables calculable encryption, a feature lacking in standard systems like AES. Summing up two ciphertexts in AES would result in random, uninterpretable data. One way homomorphic encryption sets itself apart is by maintaining the data's structure. Data is first encrypted in a mathematical object and then encrypted in a method that preserves the integrity of the information it contains.Kadykov et al. (2021) With its scalable and adaptable solutions for data processing, analysis, and storage, cloud computing has become a paradigm-shifting industry. Data protection is becoming more and more important as more businesses, including farms, use cloud solutions. With the rise of precision farming and data-driven decision-making, this study focuses on the pressing need for strong security measures in cloud-based agricultural systems. To safeguard private farming data, it's crucial to employ current security methods, with a recommendation to use homomorphic encryption, especially for machine learning models in crop suggestions Naehrig et al. (2011).

1.1 Background:

Digital technology has revolutionized agriculture, which was mostly a labor-intensive, experience-based industry. Precision agriculture has emerged as a result of the widespread use of data analytics, drones, and sensors, which guide every facet of farming. Because it provides storage and processing power on demand and in scalable quantities, cloud computing is quickly becoming the backbone of this revolution. Weather, soil, crop, and market data is now available in real-time, empowering farmers and everyone involved in the food production value chainTan (2016). One key to maximizing resource usage in agriculture has been the incorporation of cloud computing. More effective and environmentally friendly agricultural methods may result from farmers making well-informed choices on irrigation, fertilizer, and insect management. Securing sensitive data in the cloud is a major concern for the agriculture industry.

1.2 Importance:

Protecting agricultural data stored in the cloud is of the utmost significance. Cloud platforms are entrusted with a wealth of sensitive information by farmers and stakeholders as the agricultural industry moves towards a data-centric approach. Some of the most important factors influencing the success of contemporary agricultural endeavors include weather patterns, private crop data, and market strategy. Sustainable agricultural techniques rely on secure cloud-based precision agriculture. It promotes confidence among farmers, academics, and industry participants while protecting data against illegal access and ensuring its confidentiality and integrity. Early papers were written in cryptic ways, intended to be sent and transferred without the assistance of a third party. Cryptography was invented to transfer secret information between two persons, with applications ranging from warfare to clandestine commercial transactions Kadykov et al. (2021). This method was also frequently used in computer programs where data had to be sent over the internet or another unreliable channel.

1.3 Motivation:

The importance of tackling the security issues that come with cloud computing and its revolutionary potential in agriculture is what drives this study. The use of cloud computing in agricultural processes is a paradigm shift, not just a technical one, and it has the potential to shape the future of agriculture in a more sustainable and efficient wayWei et al. (2023).

Beyond theoretical reasons, there are practical ramifications that motivate this. As the main users and creators of agricultural data, farmers want guarantees that their information will be protected from abuse and unlawful access while still being used to their advantage in farm management. In order to create solutions that address the needs and worries of end-users, researchers and industry participants need to have a good grasp of the possibilities and threats associated with protecting agricultural data. By establishing that the potential of cloud-based precision agriculture is supported by strong security protocols, this study intends to close the gap between theoretical developments and practical applicationsParaforos and Griepentrog (2021).

1.4 Research Question and Objectives:

Research Objectives: Evaluate current data storage methods in the cloud for cloud security. Propose secure frameworks tailored to agricultural data needs, considering encryption, access control, and authentication. Assess suggested frameworks in precision agriculture for practicality, feasibility, and impact. Provide practical advice to lawmakers, service providers, and cultivation business to enhance safety in cloud-hosted precision agricultural systems with actionable recommendations.

- To what extent can the privacy and security of sensitive data and operations in the cloud environment be optimized through the integration of cloud analytic techniques to create an Agriculture Production Improvement Engine that effectively boosts agricultural productivity?
- When it comes to securely processing agricultural data using homomorphic encryption, what are the main obstacles and limitations? And how can we overcome them?
- How can the Agricultural Optimization Engine help make agricultural data analytics safer and more private?

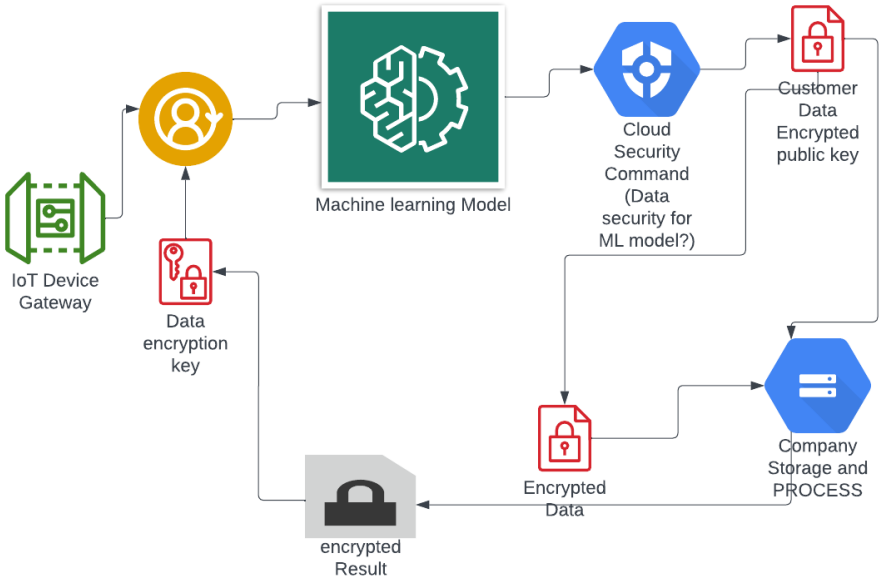

Figure 1: Architecture of agriculture optimization engine with partial homomorphic encryption

1.5 Contribution to the Literature in Science:

This research explores using homomorphic encryption to boost agricultural data security, a topic under-researched in precision agriculture. We aim to provide valuable insights for academics, practitioners, and legislators by evaluating security implications and performance indicators

This report's remaining sections are arranged as follows: Analyzing the use of homomorphic encryption in agriculture and cloud computing, identifying gaps and future research directions. Describing the steps for implementing homomorphic encryption in crop recommendation, covering data preparation, encryption model selection, and assessment criteria. Analyzing outcomes, with a focus on enhancing machine learning accuracy, improving security, and assessing the system's overall functionality. The main conclusions, their ramifications, and possible directions for further study in the area of protecting agricultural data in cloud computing settings are summarized in the conclusion Wang and Wood (2021). The goal in doing this study is to provide important information that will help shape the creation of safe and effective cloud-based precision agricultural technologies, promoting data-driven and sustainable farming methods.

2 Related Work

A considerable amount of research has been produced in the topic of cloud data security, specifically focusing on the use of homomorphic encryption in agricultural applications. This work provides a critical and analytical review of key publications that are pertinent to our study in this part. In order to justify the need of our particular research subject, this section will first place our study within the current academic environment. This will allow us to emphasize the advantages, disadvantages, and limits of prior research.

2.1 Concerns about the safety of smart agriculture Data:

Technological Advancements in Precision Agriculture for Pecan Harvesting

Technological Advancements in Precision Agriculture for Pecan Harvesting The use of PA technology in pecan production in the US is investigated by Wang and Wood (2021). Their research covers a wide range of topics, such as cybersecurity, data analytics, yield monitoring, wireless sensor networks, variable rate applications, remote sensing and a wide range of technologies used in agriculture. The paper has contributed to this work by providing a background and existing state of technological abilities in order to find improvement areas which will be discussed in this studyWang and Wood (2021).

Smart Agriculture Security Concerns Through an examination of the difficulties associated with smart agriculture's security, Araujo Zanella et al. (2020) highlight the need to integrate technologies such as cloud, edge computing, big data, and artificial intelligence. The study acknowledges that present smart agricultural systems have a lack of security features and list problems like incompatibility, limited resources, and enormous data. Enabling the use of cloud technology has the potential to help on the large data and resource front. But as and when there is growth in an industry, it leaves them vulnerable to attacks without a pre-existing active security infrastructure in place. Therefore, our work extrapolates from this conclusionde Araujo Zanella et al. (2020).

The Role of Technology in Rural and Agricultural Sectors The digital transformation of agriculture is the subject of a thorough study by Trendov et al. (2019), which examines key factors, facilitators, and examples. It also discusses multiple case studies of the impact of digitization on agriculture and its benefits. One common area lacking in these cases was the security of gathered data. It is understood that there will be enormous data generated, as a result of industrialization and digitization. Also, these systems themselves operate on the result of the analyzed past data. So There are multiple gaps in security, where a hacker can employ various tools and even manipulate data to influence systems and findingsTrendov et al. (2019).

The importance of safe IoT settings is highlighted in a survey on IoT security conducted by Hassija et al. (2019). Blockchain, fog computing, edge computing, and Machine Learning (ML)are some of the technologies introduced in this paper as they delve into security concerns, difficulties, and solution designs. ML is used to interpret suspicious behavior in the usage of different applications that could be implemented within the cloud as well. Nevertheless, in order to understand how these technologies might help with smart agriculture and security, a case-by-case detailed introspection is required. Fog and edge computing provide similar levels of security, but what matters as discussed earlier is the particular situation it is employed in. Our work takes some insights from the case studies of this paperHassija et al. (2019).

Figure 2)(Source: [1]).

[1]Figure 2 `https://ieeexplore.ieee.org/abstract/document/8742551`

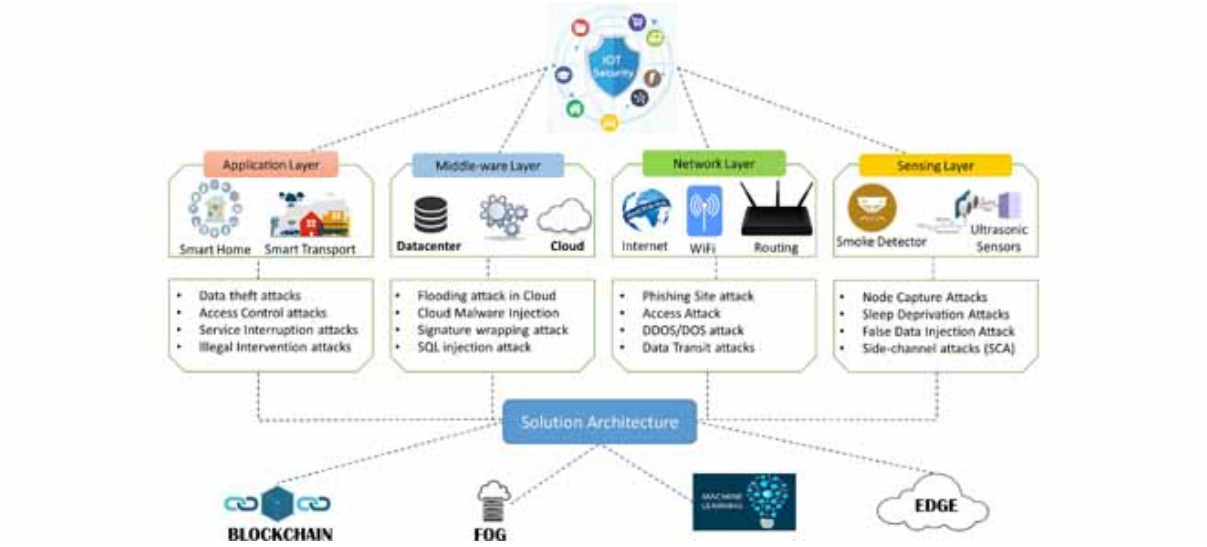

Figure 2: solution architecture for Data threats in cloud computing

2.2 Comparision of different cloud security methods:

Semwal and Sharma's study gives a thorough introduction to different cryptographic algorithms, with a focus on how they might be used to secure data on the cloud. An excellent tool for learning about what compromises are made when choosing an algorithm is the study of important characteristics performance costs and security. It tries to guide the reader in selecting the optimum technique suitable for his company. This study influenced our work in terms of measuring the trade-off versus benefits for homomorphic encryptionSemwal and Sharma (2017).

Security-Conserving ML via Federated Learning and Homomorphic Encryption A multi-party privacy-preserving machine learning system called PFMLP was presented by Fang and Qian (2021). It combines approximation activation functions with bootstrapping to enable deep learning and neural network-based models to handle encrypted data. Notable contributions include resolving the effect of homomorphic encryption on training efficiency, where a trial in which model training was achieved in 3 hours for a medium-sized dataset. Fang and Qian's study explores how federated learning, privacy-preserving ML, and homomorphic encryption all come together. Finding a happy medium between performance and security is difficult, as the highlighted trade-offs show, but with an enterprise-level performance computer/system close accuracy figures were achieved with this technique enabling the secure operation of deep learning models on big dataFang and Qian (2021).

An Algorithm for Safe casting Data in the Cloud Applying Homomorphic Encryption and Multi-Party Computation (SMPC) A safe method for cloud computing was suggested by Das (2018), which makes use of homomorphic encryption to facilitate processing by several parties. This research guarantees privacy and security in the cloud by encrypting data using Optimal Asymmetric Encryption Padding (OAEP)-Homomorphic Encryption (HE)- RSA. While the processing expense is recognized, multi-party computation offers an additional degree of flexibility and security as multiple responsible parties are author-

ised to monitor and control the processing as demonstrated by the Figure 3). It gives a different perspective to what security and control mean to different companies.Das (2018).

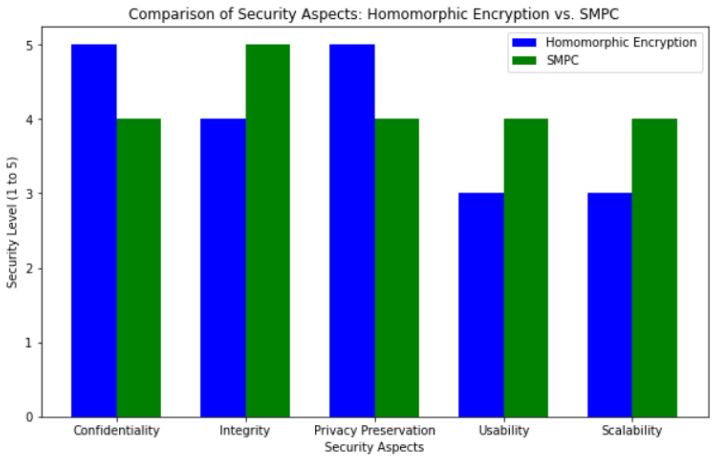

Figure 3: comparision chart of partial homomorphic encryption and SMPC

2.3 Homomorphic encryption for data security:

Methods for Homomorphic Encryption: There is a wealth of information on the uses and drawbacks of homomorphic encryption in the literature. This paper titled "Homomorphic Encryption" by Ogburn et al. (2013) gives an introductory explanation of this topic. The paper also makes a case for its applicability and widespread usage with the recent developments in cloud computing and widespread technological advancements In their research, they look into homomorphic encryption systems and its subdivisions and limitations in relation to the safe transfer of sensitive data. Although the work offers useful ideas and a proof-of-concept method, it's important to think about possible restrictions in flexibility and computational resource requirementsOgburn et al. (2013).

Shafagh et al. (2017) present a mobile application for real-time analysis of IoT data using optimal techniques, emphasizing compatibility with low-capacity mobile platforms. The study demonstrates a significant performance gain and includes security features like re-encryption and key updates. In "Homomorphic Encryption for Security of Cloud Data" (Potey et al., 2016), the focus is on testing encrypted data storage quality in Amazon Cloud Services' Dynamo DB Database. The article highlights the potential benefits across industries with the adoption of fully homomorphic encryption for secure cloud data storage. Although the strategy improves security, its scalability and practicality will be determined by the amount of data they generate and whether they are willing to add a decryption layer at all of their data end pointsShafagh et al. (2017).

Evaluation of Asymmetric and Asymmetrical Homomorphic Encryption: Evaluation of Asymmetric and Asymmetrical Homomorphic Encryption: Fully homomorphic and partly homomorphic encryption are both examined in depth by Morris (2013), who highlights the advantages and disadvantages of each Gupta et al. (2017). Highlighting the difficulties, such as processing time and implementation complexity, of completely homo-

morphic encryption, the article explores its possible uses in cloud computing and secure voting systems. We review the fascinating and challenging aspects of Gentry's 2009 presentation of lattice-based encryption. Morris highlights the importance of practical issues by providing a critical assessment of the difficulties and trade-offs of lattice-based encryptionGupta et al. (2017).

The Environment of Homomorphic Encryption:

The basics of safe computing and data privacy, read up on homomorphic encryption in the scholarly literature. By showcasing developments in cloud computing, Ogburn et al. (2013) in "Homomorphic Encryption" make a substantial contribution to the area. Subdivisions and restrictions linked to homomorphic encryption techniques are presented in their work, which focuses on the safe transfer of sensitive data. It is important to carefully evaluate any constraints, such as processing overhead and lack of adaptability, even if the paper presents useful ideas and a proof-of-concept program Morris (2013). Additionally, in their 2016 article "Homomorphic Encryption for Security of Cloud Data", Potey et al. explore cloud computing security concerns. The article suggests using completely homomorphic encryption to store data securely in the cloud, which addresses data security, privacy, and secrecy. On the other hand, possible difficulties with scalability and practicality need a rigorous review of the strategy. Before deciding if it is practicable for use in the real world, it is important to consider the pros and cons in terms of computing complexity Gupta et al. (2017).

Methods for Examining Homomorphic Encryption: A thorough examination of partly and completely homomorphic encryption is provided by Morris (2013), who adds to the existing research. While highlighting the difficulties, such as processing time and implementation complexity, of completely homomorphic encryption, the article explores its possible uses in cloud computing and secure voting systems. The article delves into Gentry's 2009 presentation of lattice-based encryption, shedding light on the many hurdles and thrills that came with it. Morris highlights the significance of practical factors when objectively assessing the difficulties and trade-offs of lattice-based encryption Bos et al. (2014).

3 Research Methodology

3.1 Introduction

The Project has highlighted the scientific rigor, materials, equipment, procedures, and reasoning behind the selected strategy as this process dives into the finer points of the study process in this part. A detailed account of my adopted procedures has been presented in this section, so anyone can follow along and learn more about how to handle the challenges of protecting agricultural data in the cloud.

3.2 The two sections of this work

3.2.1 Machine learning Model:

This process started my study by creating a model to forecast crop suggestions using machine learning. This model, which was built on top of the recommendation.csv dataset,

was essential in establishing linkages and trends in agricultural data and paved the way for further analysis, the following figure displays the process Figure 4(Source: [2]).

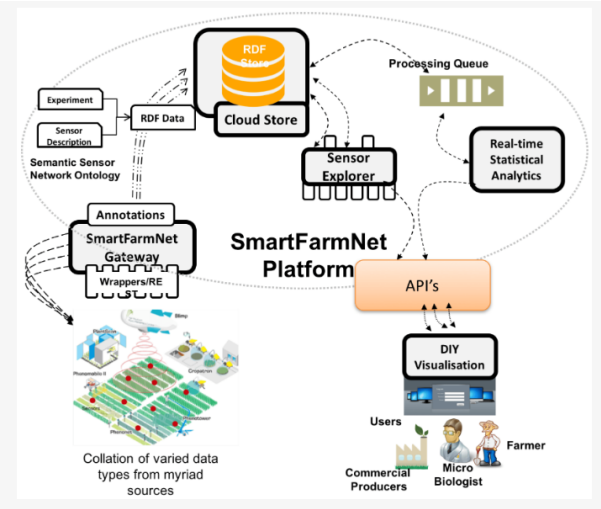

Figure 4: Process for crop data analysis

3.2.2 Encryption for Data Security

Data security via a series of mathematical processes is a part of homomorphic encryption. Without disclosing or decrypting the original data, these mathematical procedures allow for the conversion of cipher text to plain text. Therefore, this method of encryption is recognized as a subset of cryptography that facilitates data transport via the concealment of information. In a nutshell, HE is defined in Regev (2022). HE mostly works by adding a binary operation to plain letters and then running them through a series of encrypted texts. To be sure, computational analysis is required for the storage and retrieval of user data, which is the most significant aspect of these processes. When this occurs, HE-based encryption and decryption are useful tools to have on hand. With partial HE, encryption can be done more quickly, with less computational complexity, and with higher security for the user's data. Enabling mathematical calculations to be performed on encrypted data without decryption being a prerequisite is a crucial feature of a HE. This feature is an alternative to more conventional forms of encryption and has additional benefits. An HE's execution procedure is straightforward, using just binary operations like AND and XOR. Multiplying and adding arithmetic integers may so be expressed as:

- AND (b1, b2) = b1 · b2
- XOR as the addition of two bits modulo2 : XOR (b1, b2) = (b1 + b2) mod 2.

The Figure 5 shows the homomorphic encryption steps.[3]

[2]Image Source `https://www.mdpi.com/1424-8220/16/11/1884`
[3]Image Source `https://www.mdpi.com/1424-8220/16/11/1884`

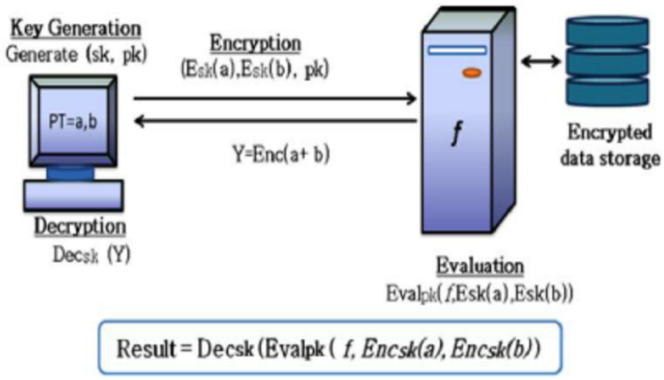

Figure 5: homomorphic encryption steps

Where Figure 5 input procured by use of the public key (Pk), C: the relevant circuit wherein activities take place. The amount of operations required to execute a typical HE determines how well it will operate. Therefore, based on it, a HE may be classified as:

- Encryption using Full Homomorphism (FHE):

  Permits more mathematical operations like rotating data, boot-strapping, comparison, and logical operations on encrypted data including addition and multiplication. This is the most computationally intensive HE. It uses bit-level asymmetric encryption. It safely converts plaintext to ciphertext. It is the most secure and recommended in financial institutions and for encrypting sensitive information, but its processing and storage needs, noise signal amplification, and massive cipher texts reduce its applicability.

- Partial Homomorphic Encryption (PHE):

  Used in applications and processing that involve basic arithmetic operations. This makes it a preferred choice of HE for data aggregation, analysis, and comparison-type processes. Efficient for protecting large-scale company data.Permits simultaneous computations on encrypted data. Problems with noise in PHE scheme cipher values are resolved. Key management is also an additional task.

- Somewhat Homomorphic Encryption (SHE):

  A kind of PHE that permits indefinite operations. It is a middle ground between PHE and FHE. It is preferred in data that is used for analytical purposes. It does have its own unique challenges like difficulty in key generation, and performance overhead.

3.3 The rationale behind the suggested approach's usage of Paillier homomorphic encryption:

A partially homomorphic encryption scheme is recommended for the given crop recommendation dataset, which includes attributes like temperature, humidity, nitrogen, phosphorus, potassium, rainfall, and crop types. This scheme must support multiplication and addition operations for full processing, which is in line with the types of computations typically done on datasets, like aggregations or summary statistics. This might

help decide if the benefits of privacy are worth the computational cost of encryption. This decision is a reasonable middle ground that satisfies the practical needs of the designated agricultural dataset displayed in Figure 6 in terms of safe data processing. For any encryption to fall under Partially Homomorphic Encryption two properties need to be true:

- 1) D(E(A)+E(B)) = A+B, where D() is Decrypt(), E() is Encrypt(), A and B are cipher text
- 2) D(E(A)*scaler) = A*scaler, scaler here refers to a constant value (exponent of the value in accordance with this project)

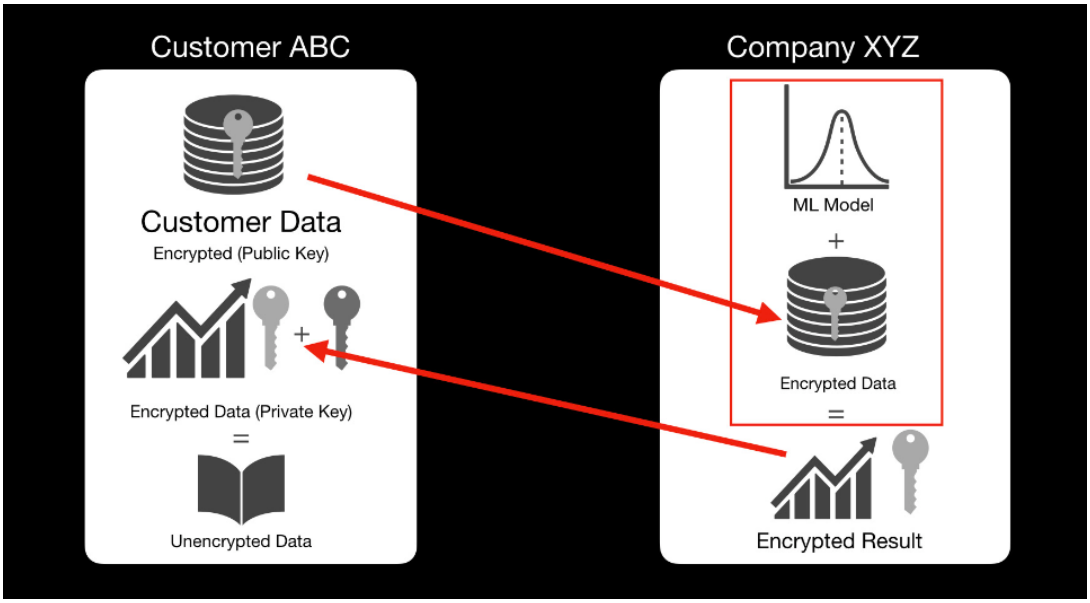

Figure 6: implimentation architecture

There are four steps to implementing a generic HE(homomorphic encryption), and they are as follows: Mathematical procedures being performed on the integers, key generation, input encryption, and output decryption My strategy used partial homomorphic encryption techniques to acknowledge the sensitivity of agricultural data. The `cust.py` script made key management easier; `data.json` included the encrypted data and `custkeys.json` had the encryption keys. Data storage and transmission were both protected by this encryption method.

3.4 Resources, Tools, and Procedures

For the purpose of producing trustworthy study results, my technique relied on a wide range of resources, specialized tools, and procedures.

3.4.1 Source of the Data

`Crop_recommendation.csv` was the main dataset that supported my study. Trained and evaluated my machine learning model using this dataset, which contains a variety of agricultural variables.

3.4.2 Tools for Encryption

The encryption code, raw data, and encryption keys both public and private are the three things required for a partial homomorphic encryption which are denoted by the code files cust.py, data.json, and custkeys.json.

3.4.3 Machine Learning Tools

As part of my investigation, Scikit-learn toolkit was used in Python to create a file that included a linear and logistic regression model. The agricultural data was easily dissected and insightful conclusions were drawn with the help of this program.

3.4.4 Primary section

The `main.py` script is the main code file in this project handling everything from data serialization to communicating with the server for partial homomorphic encryption and decryption.

This study is transparent and reproducible because the paper thoroughly explains my approach, data preparation processes, and statistical techniques.

3.5 Statistics, Data Preparation, and Sampling

3.5.1 Sampling Approach

For our agricultural crop dataset, we used systematic random sampling, giving every field an equal chance of being chosen. We also used stratified sampling, dividing fields based on crop types, soil, and location. These methods ensure a diverse and unbiased dataset, reflecting real agricultural scenarios. This robust dataset provides reliable insights into crop distribution, yield variations, and regional practices, enhancing the accuracy of our machine-learning model.

3.5.2 Data Preparation

The `Crop_suggestion.csv` dataset underwent thorough cleaning and preprocessing as part of the data preparation phase. The integrity and trustworthiness of the data used to train the machine learning model and decrypt it dependant on this stage.

3.5.3 Statistical Methods

The scikit-learn module in Python was important in enabling statistical analysis Pedregosa et al. (2011). The machine learning model's predictive ability was assessed using regression measures, such as R-squared and mean-squared error. Furthermore, statistical approaches were used to evaluate compatibility, feasibility, and practicality throughout the encryption effect study.

3.6 Alternative Methodologies: Discussion

A thorough examination of several techniques is shown by a discussion of various methodologies, which show that the selected route is in line with the study goals.

3.6.1 Other Machine Learning Models

Decision trees and support vector machines are among the other models that are examined. However, because of its effectiveness and ease of interpretation, logistic regression became the go-to method for forecasting crop recommendations. This ML technique directly gives us the degree of influence of the independent variables on the dependent variable, which makes it feasible to fine-tune the effect of parameters for us to achieve the desired output. It is also one of the less computationally intensive and simple processes which makes it pair well with Homomorphic Encryption. The compatibility is also good since they both use similar arithmetic functions for processing.

3.6.2 Other Approaches to Encryption

Research into other encryption techniques, such as completely homomorphic encryption, was been out. On the other hand, partial homomorphic encryption was able to meet the needs of agricultural settings with limited resources by balancing security with computing efficiency.

3.7 References to Prior Work

This approach is not one of the standalone methods; rather, it incorporates ideas and findings from other domains, including data encryption and machine learning for agriculture. In this part, This process has highlighted how my technique is strong since it builds upon previous work.

3.8 The Approach to Research and Evaluation

This research was conducted demonstrating dedication to maintaining high standards of scientific rigor by adopting standard research, ethical, and assessment approaches. Machine learning, encryption, and thorough data analysis all work together to create a unified and rational strategy. This approach lays the framework for thorough agricultural data processing in cloud environments.

4 Design Specification

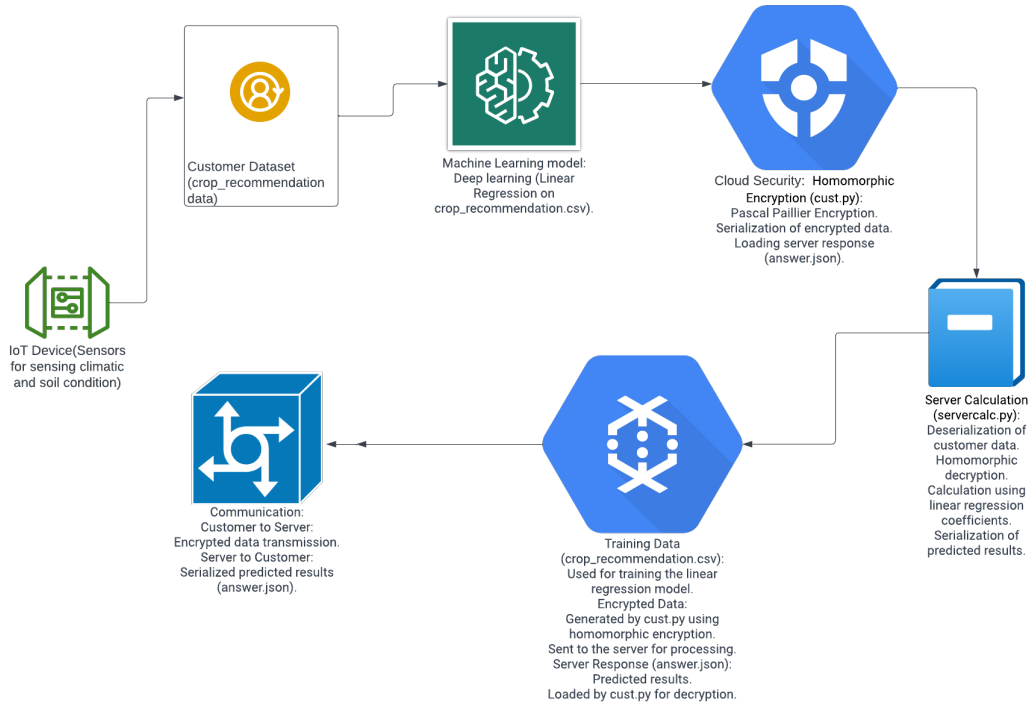

Figure 7: system design for agriculture optimization engine with Pascal Paillier Homomorphic Encryption

The methods, architecture, and framework that underpin the implementation are described in the design specification for the research project *"Secure Agricultural Data Processing in Cloud Using Partially Homomorphic Encryption"*. A thorough description of the system, including the suggested Partially Homomorphic Encryption algorithm, related parts, and specifications, is displayed in this section.

4.1 Properties of Partially Homomorphic Encryption

1. Property 1: $D(E(A) + E(B)) = A + B$, where $D()$ represents Decrypt(), $E()$ represents Encrypt(), and A and B are ciphertexts.

2. Property 2: $D(E(A) \times \text{scaler}) = A \times \text{scaler}$, where scaler is a constant value.

The system includes an Agriculture Optimization Engine built as a Machine Learning (ML) Model in addition to Partial Homomorphic Encryption. This machine-learning model aims to improve agricultural methods by utilizing various characteristics, including temperature, humidity, pH, rainfall, and soil nutrient levels (N, P, and K). Partially Homomorphic Encryption and the Agriculture Optimization Engine work together to provide a powerful solution for safe and effective data processing.

4.2 Algorithm and Model Usability

4.2.1 Model for Machine Learning (`mlmodel.py`)

`mlmodel.ipynb`, containing the ML Model, acts as the Agriculture Optimization Engine. To anticipate ideal agricultural circumstances, it uses the dataset (`crop_recommendation.csv`) and logistic and linear regression. By examining trends and connections between the input parameters, the model makes suggestions for maximizing crop output.

4.2.2 Homomorphic Encryption to Some Extent (`cust.py` and `servercalc.py`)

Data related to agriculture can be processed and encrypted safely using the Paillier Homomorphic Encryption. The procedure entails:

- Public and private keys are generated by the consumer. Customer-side encryption uses the public key to encrypt data before serializing it for cloud storage. Data is homomorphically and deserialized during server-side decryption.
- Prediction Calculation: To forecast agricultural conditions, multiply the decrypted data by logistic and linear regression coefficients. Predictions are serialized for storing in the response using server-side serialization JSON dataset.

4.3 System Prerequisites

4.3.1 Essential Functions

- The logistic regression method: Process the `crop_recommendation.csv` dataset using `mlmodel.ipynb` to get precise coefficients for prediction.
- Entity-Level Encryption: Precise encryption and decryption must be carried out via `cust.py` and `servercalc.py`.
- `servercalc.py`'s homomorphic computations have to match the above characteristics.
- Deserialization of data: Appropriate deserialization and serialization in `servercalc.py` and `cust.py`.
- Forecast Precision: The final predictions saved in `answer.json` should correctly depict the outcomes of the data that has been homomorphically processed.

4.3.2 Non-functional Requirements

- Safety: Assure agricultural data confidentiality throughout the processing, prediction, and encryption stages.
- Effectiveness: Efficient homomorphic encryption and logistic and linear regression computations are necessary to enable near-real-time or real-time processing.
- Scalability: The system needs to be scalable to accommodate different agricultural data volumes.
- Flexibility: Adjust to shifting security needs and changes in agricultural datasets.

5 Implementation

This section describes how the methodology discussed and implemented as follows. Python is the main programming language used in this project. For data processing, visualization, and machine learning, important libraries include NumPy, Pandas, Matplotlib, Seaborn, and scikit-learn. Collect and prepare data pertaining to crop suggestions for the machine-learning model.CSV file named `crop_recommendation` Pre-processed for qualities and the intended variable.

- Instruments and Tools used for Python: Matplotlib, Seaborn, sci-kit-learn, pandas, and numpy.
- Development of Machine Learning Models Using characteristics, create a logistic regression and linear regression model to forecast the category number for crops.
- Model of trained logistic and linear regression.
- Metrics for evaluating the model (RMSE, R2 score).
- Scikit-learn in Python.

5.1 Conditions for Partially Homomorphic Encryption

1. $D(E(A) + E(B)) = A + B$, where $D() \rightarrow$ Decrypt(), $E() \rightarrow$ Encrypt(), A & $B \rightarrow$ cipher text.
2. $D(E(A) \times \text{scaler}) = A \times \text{scaler}$, where scaler is a constant value (exponent of the value in accordance with this project).

5.2 Requirements:

Use homomorphic encryption for safe computation and data transfer. Generated Public and private keys kept in `custkeys.json`. Instruments and Tools: Client-side encryption and data serialization Python phe (Paillier Homomorphic Encryption library).

1. Secure Data Processing on the Client Side: Client-side data should be encrypted using homomorphic encryption.
2. Encrypted data should be serialized for safe transport.
3. Send the data to the server for processing.
4. Server-Side Encrypted Processing: Obtain the client's encrypted data.
5. Utilizing the private key on the server, decrypt the data.
6. Utilizing the coefficients of the linear regression model, do a safe calculation.
7. Return the results to the client after encrypting them.

- Analysis of Results and Client-Side Decryption: Visualizations include findings from cluster analysis, seasonal crop categorization, and distribution plots for agricultural situations.

- Results of the statistical analysis: Averages and trends for several agricultural indicators.
- Results of the Cluster Analysis: Crop identification within each cluster.
- Homomorphically encrypted computation and data transfer. Safe prediction outcomes without disclosing unprocessed data. Crop suggestion service that protects user privacy.

6 Evaluation

This section aims to provide a thorough overview of the study's key findings and outcomes, including the ML Model and Partial Homomorphic Encryption. The ramifications of these discoveries are discussed from both an academic and practical standpoint. A thorough and comprehensive analysis employing statistical methods is used to critically examine and appraise the experimental study outputs and degrees of significance, with only the most relevant data that support the research question and goals being included. The findings are presented using visual aids including graphs, charts, plots, and other pertinent data.

6.1 Agricultural Optimization Engine ML Model

Crop recommendations based on climatic conditions are the goal of the machine learning model that the Agricultural Optimization Engine uses. This is a review of the ML Model:

- **Data Exploration:** After loading and examining the dataset, details about the features of the agricultural data are revealed.
- **Data Analysis:** To comprehend the average needs of different crops in terms of nitrogen, phosphorus, potassium, temperature, pH, humidity, and rainfall, statistical summaries and visualizations are utilized and displayed inFigure 8

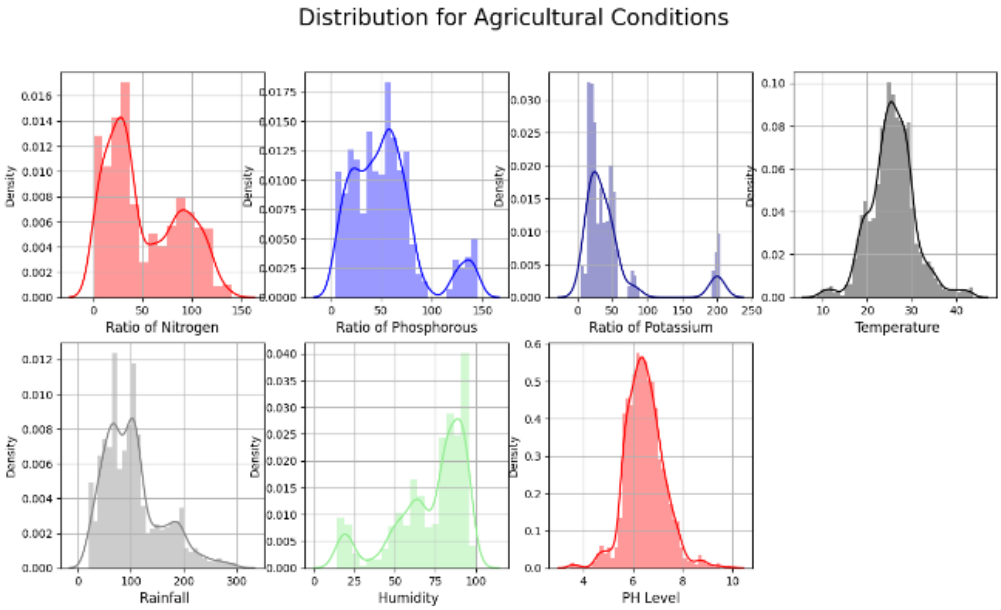

Figure 8: Distribution for agriculture condition

- **Cluster Analysis:** Permadi et al. (2023) Crops are grouped according to comparable attributes using KMeans clustering which shows in Figure 9.

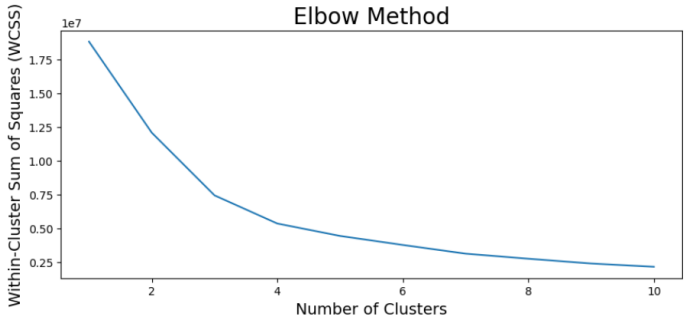

Figure 9: optimum number of cluster within the dataset

- **Predictive Modeling:** Based on certain environmental circumstances, a logistic regression model is developed using the data to forecast the crop.Figure 10.

```
predicition=model.predict((np.array([[90,
                                      40,
                                      40,
                                      20,
                                      80,
                                      7,
                                      200]])))
print("the crop good for given climatic condition is :", predicition)
```

✓ 0.0s

```
C:\Users\devma\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n
  warnings.warn(
the crop good for given climatic condition is : ['rice']
```

Figure 10: Predictive Modeling

- **Model Evaluation:** A confusion matrix and a classification report are used to assess the model's performance and provide information about its accuracy and predictive power.Figure 11.

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	18
banana	1.00	1.00	1.00	18
blackgram	0.86	0.82	0.84	22
chickpea	1.00	1.00	1.00	23
coconut	1.00	1.00	1.00	15
coffee	1.00	1.00	1.00	17
cotton	0.89	1.00	0.94	16
grapes	1.00	1.00	1.00	18
jute	0.84	1.00	0.91	21
kidneybeans	1.00	1.00	1.00	20
lentil	0.94	0.94	0.94	17
maize	0.94	0.89	0.91	18
mango	1.00	1.00	1.00	21
mothbeans	0.88	0.92	0.90	25
mungbean	1.00	1.00	1.00	17
muskmelon	1.00	1.00	1.00	23
orange	1.00	1.00	1.00	23
papaya	1.00	0.95	0.98	21
pigeonpeas	1.00	1.00	1.00	22
pomegranate	1.00	1.00	1.00	23
rice	1.00	0.84	0.91	25
watermelon	1.00	1.00	1.00	17
accuracy			0.97	440
macro avg	0.97	0.97	0.97	440
weighted avg	0.97	0.97	0.97	440

Figure 11: Accuracy and predictive power

6.2 Encryption with Partial Homomorphism

For secure computation and data transmission, the Agricultural Optimization Engine uses partial homomorphic encryption in addition to the machine learning model. An assessment of Partial Homomorphic Encryption is provided below:

6.3 Model of Linear Regression and logistic regression:

The scikit-learn `Linear and logistic Regression` class is used to implement the logistic and linear regression model. The code preprocesses the data, divides it into training and testing sets, fits the regression model, and calculates metrics like R-squared and RMSE (Root Mean Squared Error) to assess the model's performance. Using the training model, make predictions (y-pred) on the test set. The algorithm uses a logistic regression model to estimate the best crop for meteorological circumstances. The model is trained

on a broad dataset of crops using soil nitrogen, phosphorus, potassium, temperature, pH, humidity, and rainfall. A confusion matrix and classification report evaluate the model's accuracy, precision, recall, and F1 score during training and testing. Logistic regression calculates probability scores by fitting a logistic curve to input features and using the sigmoid function. Determine the assessment metrics R-squared (R)Figure 12 and Root Mean Squared Error (RMSE)Figure 13.

- Linear regression: Mathematically, the linear regression model Tripepi et al. (2008) is represented as: **y= B0+ B1 x1+ B2 x2+....Bn x n+ E**. y is the target variable, x1-xn are features, B0 is the intercept, B1-Bn are coefficients, and E is the error term.

The coefficients for each feature in the linear regression model are represented by the output [-0.02944578 -0.08846787 -0.00931931 0.00567436 0.05094233 -0.56232598 -0.0018283]. The effect of each coefficient on the anticipated target variable is shown by the associated feature. A negative coefficient, for instance, points to a negative correlation, while the coefficient's value shows how strong the connection is. The scikit-learn package is used by this Python code to create a logistic regression model for a crop recommendation system. After reading a CSV file containing agricultural data, the application does one-hot encoding, which involves separating features and labels. The logistic regression model present in Figure 11 uses an 80/20 data split for training before making label predictions for the test set. A classification report, confusion matrix, accuracy, and logistic regression coefficients are all part of the output. The model's confusion matrix and classification report show that it achieves an impressive accuracy of about 97 percent, showing that it effectively predicts crop labels across different classes.

- Logistic regression: Mathematically, the logistic regression model Vaidya (2017) is represented as: (P(Y=1))=1/(1+e-(B0+ B1 x1+ B2 x2+....Bn x n))

where P(Y=1) is the probability of the positive class (Y=1), e is the base of the natural logarithm, and B0, B1...Bn are coefficients.

Coefficient of determination (R^2 or *R*-squared)

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{m} (X_i - Y_i)^2}{\sum_{i=1}^{m} (\bar{Y} - Y_i)^2}$$

Figure 12: R square

Root mean square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (X_i - Y_i)^2}$$

(best value = 0; worst value = +∞)

The two quantities MSE and RMSE are monotonically related (through the square root). An ordering of regression models based on MSE will be identical to an ordering of models based on RMSE.

Figure 13: Root mean square error

6.3.1 Cust.py, custkeys.json, and data.json

These represent homomorphic encryption using the `phe` library. Functions include creating and storing Paillier key pairs, serializing data for encryption, and loading the encrypted response. Public and private keys are stored in `custkeys.json`, and encrypted/serialized input data is stored in `json`.

6.4 Python Integration Code (servercalc.py, main.py)

Routines in `servercalc.py` are called after `main.py` reads a CSV file (`Crop_recommendation2.csv`), extracts certain data, serializes, and encrypts the data. Using homomorphic encryption, `servercalc.py` calculates the output of a linear regression model on the encrypted data. Results are stored in `answer.json`.

6.5 Results

After processing the homomorphically encrypted data, the output displays the decrypted results. To decrypt the response, it utilizes the private key from `custkeys.json`. Adjustments are made to ensure lengths match displayed in Figure 14

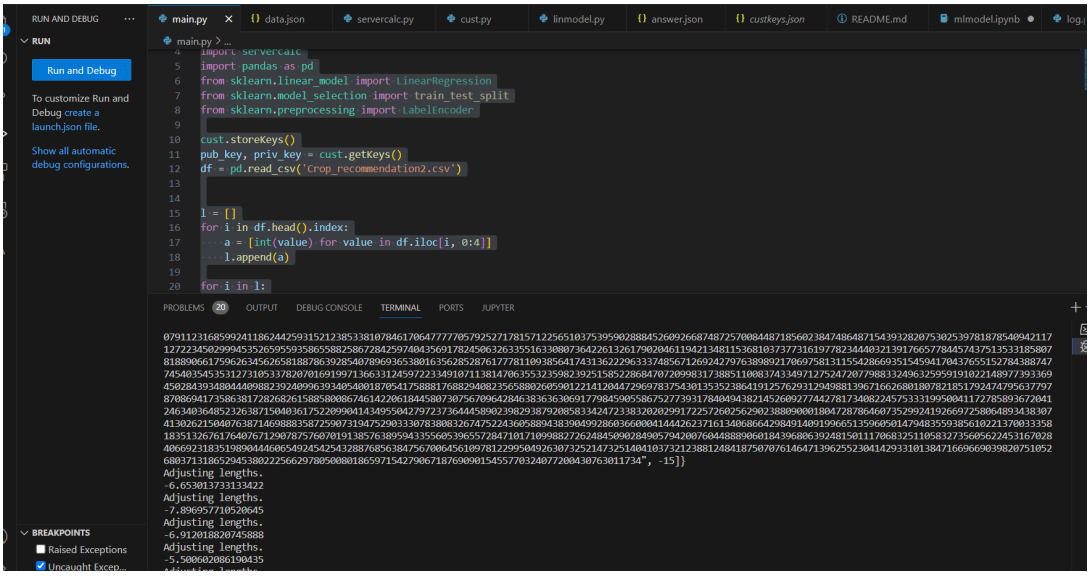

Figure 14: Final output of homomorphic encryption

6.6 Assessment

Decrypted outcomes of the linear regression model applied to the encrypted data are shown in the output. Given that the output values are negative, the target variable may have suffered. The negative values you see are likely the results of the decryption process. These values represent the predicted output of your linear regression model applied to the input features as shown in Figure 13. Thus the output you provided appears to be the decrypted values of the encrypted predictions made using a Paillier cryptosystem.

6.7 Suggestions

Provide additional details about the dataset, research topic, and aim for a more thorough review. Present findings using visual aids if data shows any particular patterns or trends. Discuss the consequences of utilizing homomorphic encryption for privacy-preserving machine learning in the application area.

7 Conclusion and Future Work

7.1 Conclusion

The Agricultural Optimization Engine's ML Model and Partial Homomorphic Encryption were the primary focus of the study, as stated in the research question, goals, and main results. A key result is the use of logistic regression on crop data to provide the best possible agricultural recommendations. Provides results for characteristics that affect harvest results as coefficients. Uses Paillier encryption to securely handle and encrypt agricultural data. Protects data while making predictions. Includes necessary features such as linear regression, encryption at the entity level, and deserialization. Prioritizes adaptability, responsiveness, efficiency, and safety. The `phe` library, `numpy`, `scikit-learn`, and `pandas` were used in its development. Uses homomorphic encryption, builds ML models, and preprocesses data. Cluster analysis classifies crops, while ML models evaluate farming circumstances. Data privacy and safe calculations are guaranteed via homomorphic encryption. Agricultural circumstances, cluster analysis, and model assessment are shown visually. The privacy of users is protected via homomorphically encrypted computing. This process provides crop recommendations based on safe forecasts.

7.2 Future Work

A number of security measures may be put in place within the framework of the Agricultural Optimization Engine project to guarantee the system's confidentiality, integrity, and overall security. Some extra security measures are as follows:

- Sophisticated model creation: To improve the accuracy and resilience of predictions, investigate advanced ML models like ensemble approaches or deep learning.
- Data Encryption and Masking: To keep the dataset usable for analysis while protecting sensitive information, use methods like data masking and anonymization.
- Verification using Biometric Data: To restrict access to critical system components to authorized users only, use biometric authentication techniques for user access control.

References

Bos, J. W., Lauter, K. and Naehrig, M. (2014). Private predictive analysis on encrypted medical data, *Journal of Biomedical Informatics* **50**: 234–243. Special Issue on Informatics Methods in Medical Privacy.
URL: *https://www.sciencedirect.com/science/article/pii/S1532046414000884*

Das, D. (2018). Secure cloud computing algorithm using homomorphic encryption and multi-party computation, *2018 International Conference on Information Networking (ICOIN)*, IEEE, pp. 391–396.

de Araujo Zanella, A. R., da Silva, E. and Albini, L. C. P. (2020). Security challenges to smart agriculture: Current state, key issues, and future directions, *Array* **8**: 100048.

Fang, H. and Qian, Q. (2021). Privacy preserving machine learning with homomorphic encryption and federated learning, *Future Internet* **13**(4): 94.

Gupta, S., Mittal, M. and Padha, A. (2017). Predictive analytics of sensor data based on supervised machine learning algorithms, *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, pp. 171–176.

Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P. and Sikdar, B. (2019). A survey on iot security: application areas, security threats, and solution architectures, *IEEE Access* **7**: 82721–82743.

Kadykov, V., Levina, A. and Voznesensky, A. (2021). Homomorphic encryption within lattice-based encryption system, *Procedia Computer Science* **186**: 309–315. 14th International Symposium "Intelligent Systems.
URL: *https://www.sciencedirect.com/science/article/pii/S1877050921009674*

Liu, S., Guo, L., Webb, H., Ya, X. and Chang, X. (2019). Internet of things monitoring system of modern eco-agriculture based on cloud computing, *IEEE Access* **7**: 37050–37058.

Morris, L. (2013). Analysis of partially and fully homomorphic encryption, *Rochester Institute of Technology* **10**: 1–5.

Naehrig, M., Lauter, K. and Vaikuntanathan, V. (2011). Can homomorphic encryption be practical?, *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, pp. 113–124.

Ogburn, M., Turner, C. and Dahal, P. (2013). Homomorphic encryption, *Procedia Computer Science* **20**: 502–509. Complex Adaptive Systems.
URL: *https://www.sciencedirect.com/science/article/pii/S1877050913011101*

Paraforos, D. S. and Griepentrog, H. W. (2021). Digital farming and field robotics: Internet of things, cloud computing, and big data, *Fundamentals of Agricultural and Field Robotics* pp. 365–385.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: Machine learning in python, *the Journal of machine Learning research* **12**: 2825–2830.

Permadi, V. A., Tahalea, S. P. and Agusdin, R. P. (2023). K-means and elbow method for cluster analysis of elementary school data, *PROGRES PENDIDIKAN* **4**(1): 50–57.

Regev, O. (2022). Lattice-based cryptography. Accessed: 14th December 2023.
URL: *https://cims.nyu.edu/ regev/papers/qcrypto.pdf*

Semwal, P. and Sharma, M. K. (2017). Comparative study of different cryptographic algorithms for data security in cloud computing, *2017 3rd International Conference on Advances in Computing,Communication Automation (ICACCA) (Fall)*, pp. 1–7.

Shafagh, H., Hithnawi, A., Burkhalter, L., Fischli, P. and Duquennoy, S. (2017). Secure sharing of partially homomorphic encrypted iot data, *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pp. 1–14.

Tan, L. (2016). Cloud-based decision support and automation for precision agriculture in orchards, *IFAC-PapersOnLine* **49**(16): 330–335. 5th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2016.
URL: *https://www.sciencedirect.com/science/article/pii/S240589631631624X*

Trendov, M., Varas, S., Zeng, M. et al. (2019). Digital technologies in agriculture and rural areas: status report., *Digital technologies in agriculture and rural areas: status report.* .

Tripepi, G., Jager, K., Dekker, F. and Zoccali, C. (2008). Linear and logistic regression analysis, *Kidney international* **73**(7): 806–810.

Vaidya, A. (2017). Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval, *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6.

Wang, H. and Wood, E. (2021). The application of precision agriculture technologies in us pecan production: Challenges and opportunities, *Western Economics Forum*, Vol. 19, pp. 20–27.

Wei, Y., Han, C. and Yu, Z. (2023). An environment safety monitoring system for agricultural production based on artificial intelligence, cloud computing and big data networks, *Journal of Cloud Computing* **12**(1): 1–17.