8-2024

Leveraging Generative AI For Sustainable Farm Management Techniques Correspond To Optimization and Agricultural Efficiency Prediction

Samira Samrose
Utah State University

LEVERAGING GENERATIVE AI FOR SUSTAINABLE FARM MANAGEMENT TECHNIQUES CORRESPOND TO OPTIMIZATION AND AGRICULTURAL EFFICIENCY PREDICTION

by

Samira Samrose

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Computer Science

Approved:

Curtis Dyreson, Ph.D. Major Professor	Steve Petruzza, Ph.D. Committee Member
Chad Mano, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

ABSTRACT

Leveraging Generative AI For Sustainable Farm Management Techniques Correspond To Optimization And Agricultural Efficiency Prediction

by

Samira Samrose, MASTER OF SCIENCE

Utah State University, 2024

Major Professor: Curtis Dyreson, Ph.D.
Department: Computer Science

The farm management techniques refer to the process of planning and implementing the decisions involved in organizing and operating a farm for maximum production and profit. It covers a wide range of activities that include designing a scheme, implementing, and monitoring the farm's operations to ensure efficiency, sustainability. The motivation behind this project is to promote goal setting and achievement of reduced CO2 emissions.

(68 pages)

PUBLIC ABSTRACT

Leveraging Generative AI For Sustainable Farm Management Techniques Correspond To Optimization And Agricultural Efficiency Prediction

Samira Samrose

Sustainable farm management practice is a multifaceted challenge. Uncovering the optimal state for production while reduction of environmental negative impacts and guaranteed inter-generational assets supervision needs balanced management. Also, considering lots of different factors (cost, profit, employment etc), the agricultural based management technique requires rigorous concentration. In this project machine learning models are applied to develop, achieve and improve the farm management techniques. This experiment ensures the resultant impacts being environment friendly and necessary resource availability and efficiency. Predicting the type of crop and rotational recommendations will disclose potentiality of productive agricultural based farming. Additionally, this project is designed to find the optimized farm operations that will show a stable state combining the agricultural efficiency, better resource management and lowering ecologically unfriendly properties. Additionally, generative AI is used to create data for farming management practices.

To all the little people....

CONTENTS

LIST OF FIGURES

ACRONYMS

GHGs	Greenhouse gases
GHG	Greenhouse gas
GAN	Generative adversarial network
COMET	CarbOn Management & Emissions Tool
HOLOS	Whole-farm model and software program

CHAPTER 1

INTRODUCTION

Farm management study mainly focuses on preservation of the ecosystem and possible reduction of climate change. This project explores the observation on the agricultural land in Utah along with the Wasatch Front. Housing infrastructures threaten the complete disappearance of agricultural land. The significance of agricultural land preservation is necessary. So, this project is influenced by agricultural crop and animal management practices along with reducing climate change. Agricultural activities are one of the major contributors to GHG emissions. Farm management impact the emission of GHGs. My goal in this thesis is to create a framework for predicting how the large number of farms across diverse management scenarios impact GHG emissions. The goal is to create a model that can identify the significant influences that cause most of the GHG emissions. This will help farmers reduce those emissions while maintaining farm productivity.

1.1 Background

The overall goal for the farm management system is to understand how agricultural production and land management can be optimized to reduce less GHGs emission. Potentially these studies are valuable for farmers, policymakers, and stakeholders for significant progress in the future agriculture practices. Choosing suitable farm management practices is always a hard task to perform. But with the help of artificial intelligence based models there is a higher chance to predict which practices can help to achieve a sustainable management system while managing GHG emissions.

Figure 1.1 shows the average agriculture, forestry and fishing share of the US economy from 2001 to 2021. The graph illustrates the growing contribution of agriculture to the economy [1] [2]. Rising values indicate the growing agricultural sector. Additionally this is a clear indication of a more production profitable agricultural base. Besides, the

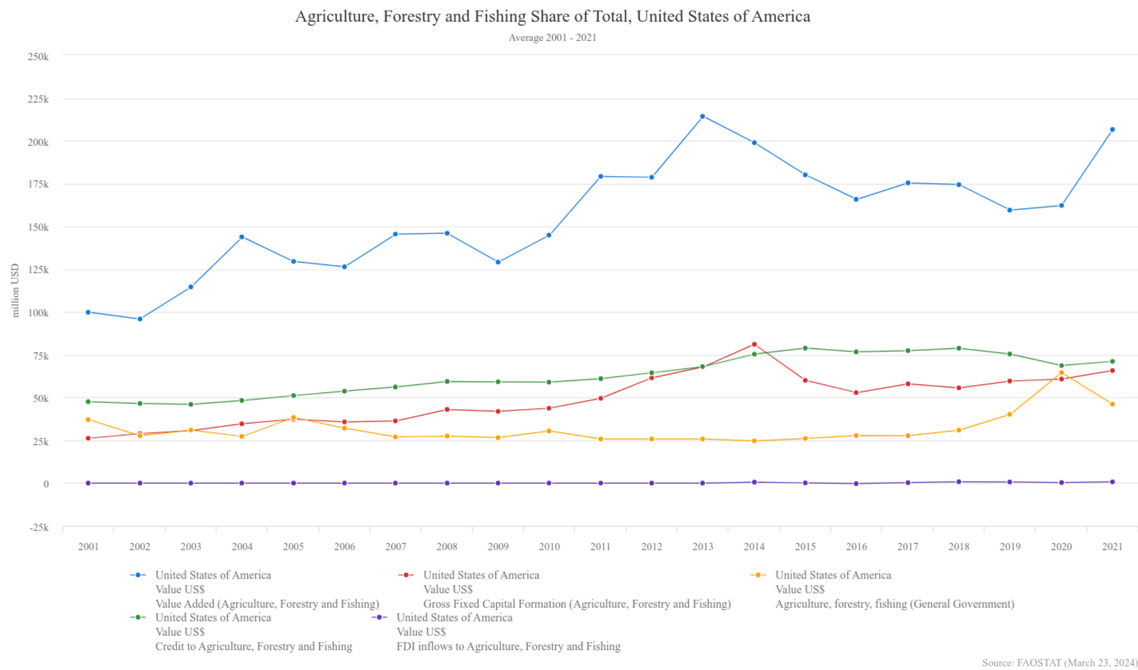


Fig. 1.1: Agriculture, Forestry and Fishing Share of Total, United States of America Average (2001 - 2021).

downward trend is a sign of challenges of farming. This is certainly an important indicator that sustainable agricultural production is expected to accomplish. The FAO states and seeks the

> “increase agricultural productivity and sustainable food production and ensure equitable access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment, particularly for women, indigenous peoples, family farmers, pastoralists and fishers, in order to achieve food security and improved nutrition and promote sustainable agriculture [3].”

The gross fixed capital formation is a potential growth in investment on possible expansion of agricultural base. Also, the increment of credit to agriculture trend means the easier access to capital for fuel growth and innovation. Additionally, EDI inflow is the rising foreign investment in the agriculture sector. This graph shows how agricultural experimental studies can be important and holds promising growth potential. The steady growth of agricultural research has always had some external factors.

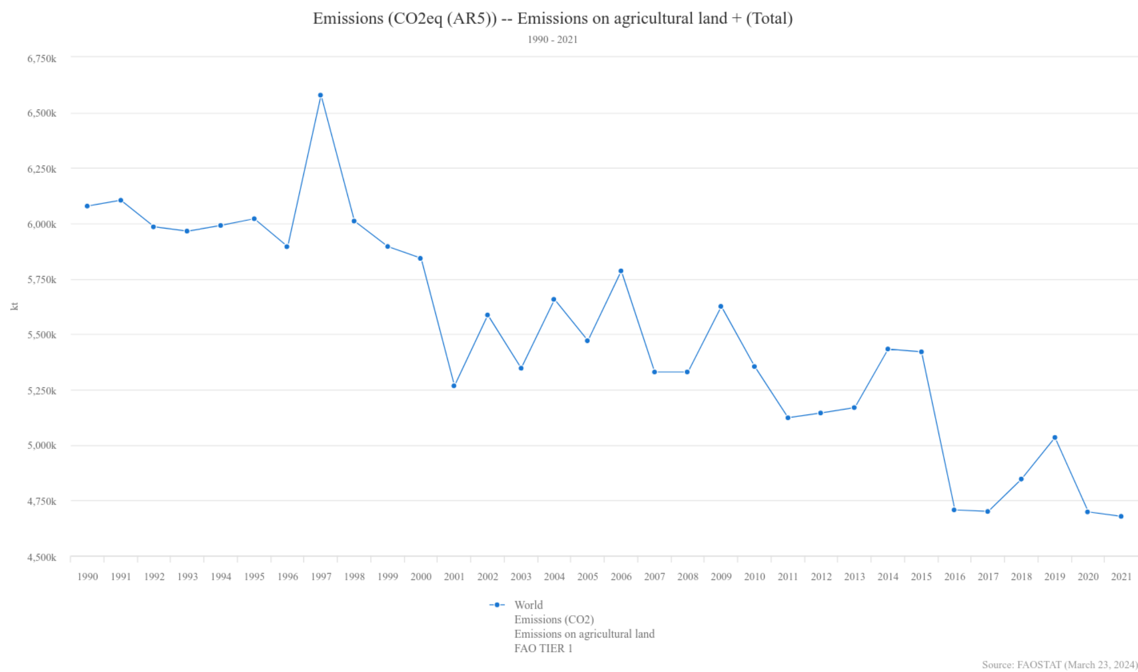


Fig. 1.2: Emissions (CO2eq (AR5)) – Emissions on agricultural land (1990 - 2021).

Figure 1.2 shows the total emissions on agricultural land from 1990 to 2021. According to the graph both emission on agricultural land and total emissions have risen since 1990. Agricultural land based emission is around 4500 CO2eq (AR5). Here, emission (CO2eq AR5) means carbon dioxide equivalent measured in AR5 (fifth assessment report of the intergovernmental panel on climate change) [4]. And the total emissions are around 5,000 CO2eq (AR5).

This graph shows the increased rate of emissions and how important it is to reduce overall CO2 emission and follow strategies in agricultural farming management. Also, an interesting fact can be observed in this graph is the rising emissions but there is no downward trend in agricultural outcomes. It means that the production or the agricultural practices keep on meeting the demands of the market. These unsustainable farming practices can be a possible reason for damage or loss in the future. Additionally, because of the emissions, there will be climate change in the future. This can bring a negative impact on the production of agriculture. Such as, extreme weather event, rising temperature and low sources of water availability. These will certainly bring negative impacts on the crop yield and livestock production. For this reason, the experimental studies on sustainable farming management is essential.

1.2 Related Work

There are several studies on accomplishing agriculture based sustainability. Due to climate change and different farming management settings, there is a higher chance that each year the demand or necessity to achieve sustainability changes. Some of the related research on achieving sustainability are as follows.

Agroecology for a sustainable agriculture and food system from local solutions to large-scale adoption this paper describes and applied the agroecology as their main methodology for sustainability. There are many challenges in mankind as a mainstream [5]. Integrated farm management for sustainable agriculture: Lessons for knowledge exchange and policy this article suggests that integrated farm management can be a possible way to achieve sustainability. The study is a mixed-methods approach. This paper aims to find a balance in economic viability, environmental protection and social responsibility for efficiency [6].

Novel approaches and practices to sustainable agriculture paper is an example of climate-smart agriculture based farming. With the rapid population growth, the food production and balance between scientists and policymakers can have an impact on the agricultural practices. Some suggested farming practices are discussed in the paper. The paper also suggests adopting new approaches for different farming scenarios [7].

Natural resource management and sustainable agriculture paper discusses soil, climate and possible solutions for sustainable food production. For the agriculture techniques, the papers talks about the climate-smart farming, reduction of usage of crops for animal feed, echo-friendly animal agroforestry and silvopastoral systems, redacted or no-tillage practices, suitable crop production methods, nutrient and fertilizer management, integrated watershed management, anaerobic digestion, climate. This reading helped a lot to understand the importance of adopting sustainable agricultural practices [8].

Agricultural sustainability: a review of concepts and methods paper proposes a methodology combining the existing review criteria and offers an analysis. The review is based on 38 studies on crop farming over ten years of data. This paper suggests that the usage of a combination of methods can make the total framework difficult to categorize [9].

In this project the concentration of GHGs for the environmental effects are the main concern. Not only climate change but also it has harmful effects on the economics, human well-being [10]. Because of the farming practices on livestock, agricultural soil and crop production, the reported GHGs emission is rising and with agriculture accounting for roughly 10% [11]. In order to capture the impacts of the various farming management practices on GHG emission there is a need to use the whole-farm models. And that is the reason for using automation, exploring the data and applying models on the data, finding model sustainability and finally populating more optimized farming practices using generative AI.

CHAPTER 2

Definition of the Objective

Farm management is a crucial part of the agricultural industry. There are many factors that depend on the essence of managing a farm properly. Some properties include labor management, financial planning, environment sustainability, compliance and regulation, market analysis etc. This project covers the combination of both farm crop management and farm animal management styles. The objective is to find and design a model for measuring GHG emissions under different farm management scenarios.

2.1 Reduction Factors

There are many aspects or parts that are desired to be reduced for an ideal farm management from different perspectives. Estimated cost along with the related price and government policies are a few parts that impacts the other factors of the ranching operation. In this project, the reduction factors are energy consumption, carbon emission.

2.2 Boosting Factors

Several boosting factors are required in the farm management. Ecological balance along with financial success are major factors. However, sometimes this is a major problem from the farmers perspective to visualize where to focus on to invest. In this project, the boosting properties are productivity, animal welfare, effective resource usage and space utilization efficiency.

2.3 Crop Prediction

The prediction of which seasonal crop to pick and how the farming management will be that is another challenging part of farm management. So prediction of a crop type is essential in this case. In this project, prediction of the crop type along with the recommendations

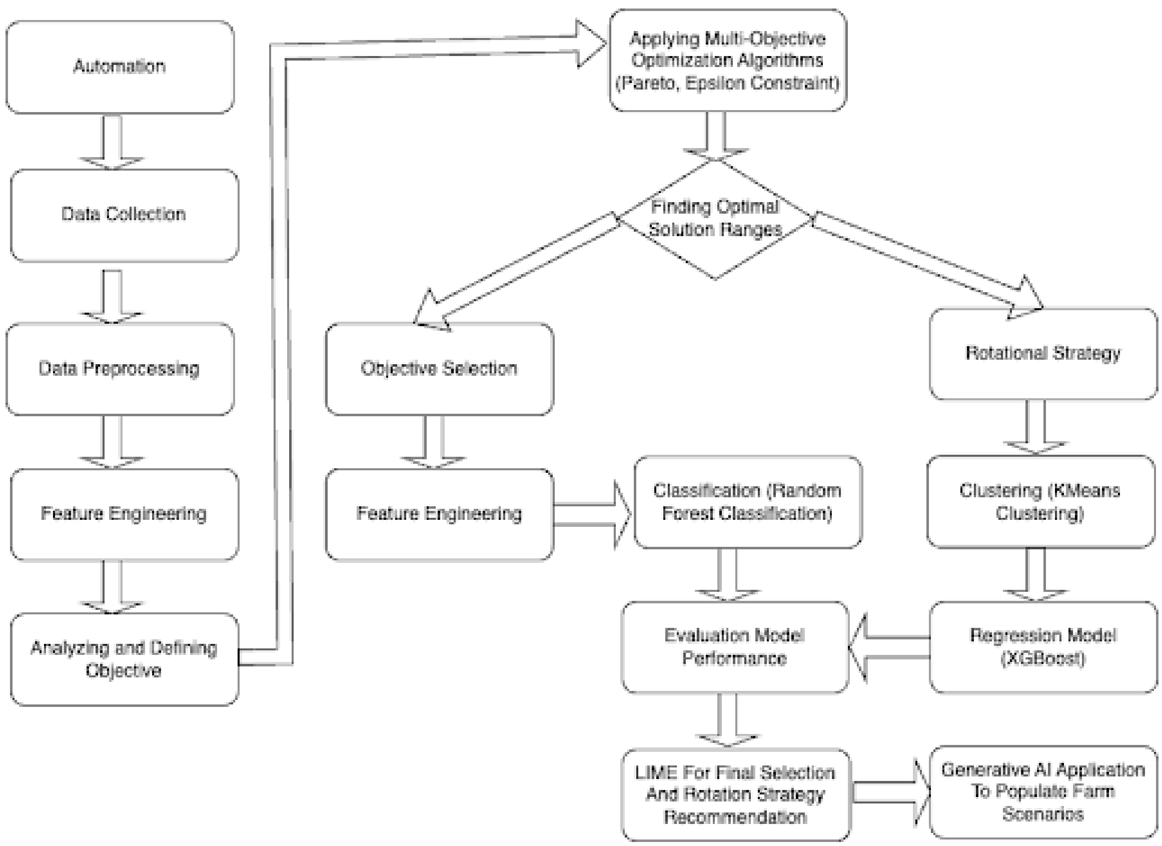


Fig. 2.1: Overall Workflow of The Project.

with the other properties are analyzed. This will help the farmers to understand when to focus on which type of crop and what are the major percentages of other properties that should be focused on.

2.4 Rotational Analysis

Not only the crop selection is important but also the rotational analysis is also crucial for the farm management. So, this project also covers the rotational analysis based analysis where there will be recommendations for crop types.

2.5 Design of the experiment

This project aims to design an experimental setup where from a farm management scenario, we will find an optimized state of balancing multiple properties. Following are the subcategories where the design of the experiment is described. Figure 2.1 shows the overall steps taken to optimize the emission while focusing on maximizing some other necessary factors in the whole project. This diagram shows the initial total outlines of the project. This systematic approach uses a combination of data driven methods and machine learning techniques. Here, in the primary step by automation the data collection is collected. Then comes the preprocessing and feature engineering steps are taken to prepare the relevant data for the project. By removing noise in the data the data is ready to be analyzed to define the objects for this project.

After determining appropriate objectives for different datasets, the multi-objective optimization algorithms are applied. Here, the pareto and epsilon constraints optimization techniques are used to achieve multiple goals. These techniques are used to find the trade-offs and optimal solution ranges where no single objective can be improved without worsening others. Next, for the rotational strategies, the machine learning algorithms are introduced in this workflow. The classification, clustering using regression model then by applying lime, a prediction based model is built to make the final decisions about crop rotation and farm management strategies. Finally, the application of generative AI helps to populate more optimized farm scenarios, allowing for testing and optimization of strategies in the simulated environment before real-world application.

2.5.1 Data Collection

The data collection from the ground level is very time consuming and complicated. In order to resolve this problem, two of the most used farming applications are used in this experiment. COMET-Farm and HOLOS whole-farm are used to capture the sample data along with the carbon emission related data.

2.5.2 Finding Optimized Farm Management Scenarios

In this project, machine learning models are applied to predict the crops, finalizing their rotational recommendations. Also, finding the optimal state where some properties will be reduced and some will be maximized is another part of this experiment. In order to achieve this goal, a multi objective optimization model is used.

2.5.3 Populate Scenarios

Now after finding the optimal state the project is taken in another stage where the outcomes suggest farm management. For the purpose of finding more optimal farm management scenarios, more farm scenarios should be generated. With the aim of populating more optimal farm scenarios, generative AI is used to evolve more farm scenarios. Here, in this project the generative adversarial network is used.

CHAPTER 3

Method

There are a few steps to accomplish this experiment. Automation in python is applied first to capture necessary data for the farm management. In order to classify the crop, a random forest classifier is used. For the rotational analysis, K-means cluster along with XGBoost regression algorithm is applied. And for all the recommendations LIME is used to explain the recommendations. Later, in order to find the optimal states while minimizing and maximizing some properties, a multi-object optimization algorithm is applied. In this experiment, ϵ-constraint and pareto optimization are used to find the optimal states.

3.1 Automation

In order to reduce the cognitive costs and workload the automation is done. In this project the automation part is one of the most important parts. The automation provides stability for the data collection part, successfully terminates the human effort and errors and allows continuous data generation flourishingly. The concurrent data generation is the main focus of applying automation in this project. The automation is done using python script where pre-ranged data is provided in two different whole farm carbon models.

A whole farm carbon model is based on the essential information of a farm. Each farm information is used to estimate greenhouse gas emission. In this project, two whole farm carbon models are used to estimate the GHG emission: COMET-Farm and HOLOS whole-farm. There are several other whole-farm applications but COMET-Farm and HOLOS are best for the overall goals of the project.

3.1.1 COMET-Farm

COMET-Farm is a user interface (UI) based website. It helps with the visual aspects of a website that users can easily interact with. COMET-Farm gives an estimation of

GHG emission for past and future farm management practices. It generates data based on provided information from the user. The results are mostly based on the past and future farm management systems and the reports are the GHG emissions. For the automation, COMET-Farm is run using the Selenium python library and Chrome browser. Selenium IDE is used to recheck all the provided correct paths to execute the python code.

COMET-Farm Crop Data

COMET-farm is a website that is most popular among scientists and farmers to understand farming management. The resultant farming scenarios can capture the emissions of that curtain farm management. There are three different types of crops selected for the farm setup. They are orchard/vineyard, seasonal crop, annual crop. It is really complicated to understand the full crop based farming steps. So the website is really helpful in dividing each and every step of the crop farming related details. Different crops demand different types of inputs as they have their own characteristics. The website is full of information on the crop based agriculture process. Table 3.1 has the data of user input and the resultant properties listed and the Table 3.3 has the ranges of inputs for the crops. The ranges are recommended by experts. So, inserting these data can generate possible crop farming managements with scenarios and resultant emissions. There are nearly 3600 runs of crop data with different attributes in the database collected using the COMET-Farm model.

Comet-Farm Animal Data

COMET-Farm also has sectors to generate scenarios with animal based farming management based frameworks. There are assists to generate livestock based data for an agricultural based farming perspective. Lots of different types of animal data can be given for a certain farm scenario to observe the whole farm management. The data is the combination of individual insetted data for each month. And this makes the farming scenarios more realistic. The capture of the emissions depends on various types of data. So COMET-Farm ensures that every step is close to the real farming managements. For this experimental setup, there are three types of animals that are considered. Heifer and Steer Stocker data is combined as they have nearly similar inputs. COMET-Farm model has around 1400 runs of data with different types of attributes.

3.1.2 HOLOS Whole-Farm

HOLOS whole-farm is a graphical user interface (GUI) based software. The interface features of this application helps the users to interact with information by manipulating the visual widgets. It allows the user to estimate the GHG emission for a certain year based on the crop and animal agriculture information. HOLOS evaluates potential strategies to reduce admission and forecasts the effects of such strategies on overall farm emissions. It is based on the principles of the Intergovernmental Panel on Climate Change along with Canadian circumstances. Basically, this software is developed in Canada based on the nature and environment of the soil of that particular country. Table 3.2 shows the inputs of HOLOS whole-farm model and Table 3.4 shows the input data ranges used in this project. HOLOS whole-farm application is run using WInAppDriver python libraries with essential installations of supporting drivers in the system.

HOLOS Crop Data

HOLOS crop based data is based on the perennial crop based. As it is a desktop based application, different python code is written to collect data from this application. The usage of this application is very user friendly and easy to understand. Table 3.2 is the recommended data insertion ranges by user and outcomes from the application. And Table 3.4 is the recommended range of data. This range of data is suggested by experts.

For the data collection, the HOLOS whole-farm is used to run approximately 1200 runs. So, the database has approximately 1200 farm scenarios of different agriculture crop based information and carbon emission related data. The dataset has 102 different features.

HOLOS Animal Data

HOLOS also provide animal based farming scenarios. The calculation is for the annual based GHG emissions. For the animal based agriculture HOLOS data there are different types of animals. Among all of those, for this experimental project, we have selected to work with beef feedlot. It provides necessary information on the heifer and steers and other related features. In an agriculture based animal based farm scenario, the HOLOS whole-farm application is run 1500 times. It has 156 different features. HOLOS collects data for each month for a specific farm management.

3.2 Model Application

The machine lerning models applied in this project are thoroughly described below.

3.2.1 COMET-Farm

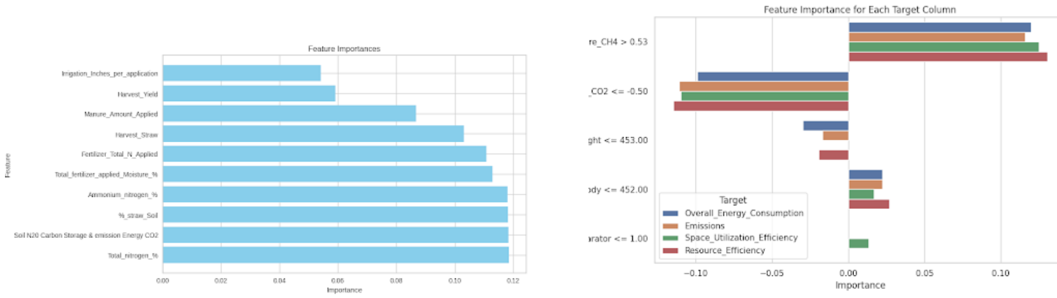
In order to understand the data more accurately the data collected from COMET-Farm is divided into two different dataset. One is related to the crop based agricultural farming and another one based on animal based agriculture farming.

COMET-Farm Crop Data

The overall goal of this task is the optimization of production processes for reduced environmental impact along with using generative AI to populate farm scenarios. In order to fulfill this purpose, machine learning models like classification, clustering and regression models application are most necessary. The purpose of this task is to understand the crop selection and additional rotational recommendation. In order to find the optimizing crop, the classification model is necessary to apply. Then to find the rotational strategies becomes very challenging. As this is a regression problem. So, here regression models are introduced to resolve this problem.

The best model depends on your specific goals. In order to predict the optimal crop based on soil conditions, historical management practices, and desired outcomes (e.g., high yield, low water use), classification models like Random Forest or SVM are good choices. Here in this case, the random forest classifier is applied.

Next, for exploring crop rotation patterns that improve soil health, reduce fertilizer needs, or promote biodiversity, clustering techniques like K-Means seem a good choice.



(a) Comet-crop

(b) Comet-animal

Fig. 3.1: Feature Importance For Each Target Column For COMET-Farm Dataset.

Later using the regression models (XGBoost) to predict the specific impact of different rotations on these factors. Then applying techniques like LIME helps to understand how the model arrives at its predictions, making the results more interpretable for farmers.

The collected data from the COMET-Farm website already has lots of valuable informative features included. However, for the necessity of this project some of the new features are considered for the model.

The target feature of this task is the crop rotation index which is a new categorical feature indicating the crop sequence followed in a specific period. Next, the feature transformation part, Soil Health Indicators is the combination of existing features like "Total_nitrogen%", "%_straw_Soil", and "Soil N20 Carbon Storage" into a single score using Principal Component Analysis (PCA) to represent overall soil health.

Fertilizer Dependence Score is a new feature based on total fertilizer applied and harvest year. After calculating the ratio between fertilizer application and yield, this feature is achieved. Besides these, orchard renewal or clearing, total nitrogen applied, soil health, carbon storage emission ratio, crop tillage category features are created for the sake of this goal.

Here in these two Figure 3.1a and Figure 3.1b the important features are mentioned. These feature importance helped to determine the final objectives for this project. For four different datasets, there are four different feature importance can be seen here. The feature importance indicates the importance of the other features on the target feature which is the type of crop. This feature selection is done using the random forest classifier. The higher scores refer to how the type of crop has a larger effect on that. So, we can see that the total nitrogen, soil N2O storage and CO2 emission have a larger impact on the type of crops. Then, the dataset was preprocessed. There are different types of data. The numerical, categorical, date type data are there. So, they are handled accordingly.

After feature engineering and dataset pre-processing, the random forest model is applied. The dataset is splitted into 70% for training and 20% for testing. The hyperparameter grid for the random forest classifier are the 'n_estimators': [100, 200, 300], 'max_depth': [10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4].

Comet-Farm Animal Data

The goal of this analysis is to find the production efficiency and promote sustainable practices. With the Comet-Farm animal dataset the main focus will be minimizing carbon emission while maximizing the effective resource usage. To accomplish this object necessary preprocessing steps are done. After handling the numerical, categorical and date type data, the feature importance is observed. Then based on the importance the most important features are selected for the analysis.

Here Figure 3.1b shows some ranges of features that are most important to reduce emission. These graphs show the feature importance summarizing how various features influence different target objectives. For example, It can be seen that the irrigation area, harvest yield, fertilizer application and types, temperature along with many features cover higher importance in the datasets.

All the data related to the animal welfare of a farm management are recorded in this dataset. The overall goal is to achieve the optimization of the production process for reduced environmental impact. To achieve this purpose, the multi-objective optimization models are applied. Before the model application, the dataset is preprocessed handling the missing values, observing the outliers and handling them, normalizing the data and scaling them accordingly. Numerical, categorical, date type data are also handled.

In the feature engineering step, new features are formed for the sake of project objectives. Overall energy consumption, emission, space utilization efficiency and resource efficiency are calculated and added in the dataset. Here, the optimization will be taking place by minimizing the overall energy consumption, emission while maximizing the space utilization efficiency and resource efficiency. In this project LIME is used to explain the results.

First, in order to understand the required ranges of the desired features, the random forest model is applied on the dataset and later LIME is used to explain the detailed ranges for the features. Then the random forest regressor is applied and LIME is initialized for the explanations. The ranged graphs can be seen in Figure 3.1b feature importance for each target column graph. Now it can be said that managing livestock weight and emissions might help to optimize energy use and efficiency. Each farm needs to find a balance depending on their specific goals, constraints, and the importance of each target outcome. The logic of farm management based on this graph is about finding the best balance between productivity, sustainability, and efficiency. By understanding which factors are most important and how they impact each other, farm managers can make decisions that reduce environmental impact, and improve the overall sustainability of the farming operation.

In terms of farm management, these insights would suggest several strategies for managing overall energy consumption: Methane Management: Implementing strategies to reduce methane emissions, such as diet modifications, could indirectly reduce energy consumption. Exploring ways to utilize methane productively can turn a waste product into a resource, potentially increasing efficiency.

Balancing Energy-Efficient Practices: Reducing energy-related CO2 emissions through practices like improved insulation, energy-efficient equipment, and renewable energy sources could have a dual benefit of reducing both emissions and energy consumption.

Weight Management: Considering the energy implications of the average weight of the livestock could help in optimizing feed efficiency and the overall energy use of the farm.

Managing the average live weight of animals to remain within a range that maximizes resource efficiency, avoiding the additional costs associated with maintaining heavier animals.

In summary, for effective farm management, it would be crucial to analyze the trade-offs between environmental impact and resource efficiency. It entirely depends on ensuring that measures to improve one do not detrimentally impact the other. In the case of animal farm management, this data suggests there is an optimal range for maximizing efficiency, and maintaining animal maintenance within this range could be a key strategy for sustainable farm management.

Understanding the ranges of the target columns, the model is applied to the dataset. The dataset is splitted into 70% for the training phase and 30% for the testing phase. Both pareto front and ϵ-constraint models are applied. After fitting the data and combining the objectives together.

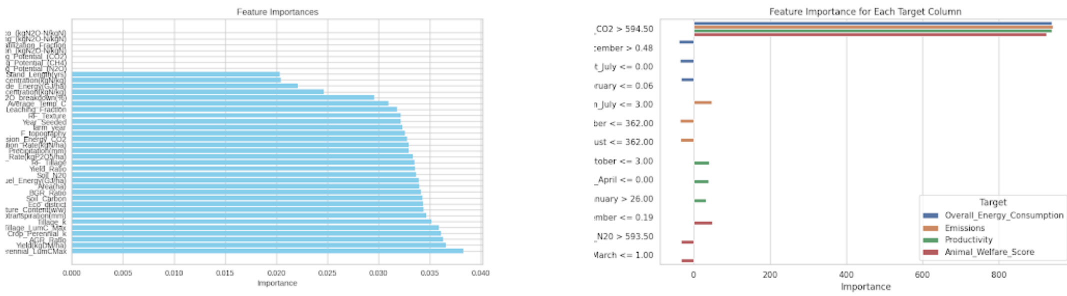
3.2.2 HOLOS

Collected data from HOLOS whole-farm application was large enough to break it down into two different dataset. This helps to visualize different sets of data and applying models are comparatively simpler in this way.

HOLOS Crop Data

The sole purpose of working with HOLOS agriculture crop based data is to optimize production processes for reduced environmental impact and applying generative AI to populate farm scenarios. Machine learning models related to classification, clustering and regression models become necessary to apply for the crop selection and rotational recommendation.

From the collected data from the HOLOS whole-farm model, all the data closely related to crop farming is stored in a separate database. HOLOS whole-farm crop data has nearly 1200 runs of data. So there is plenty of data to analyze and predict the outcomes. First, the correlation between all the features are observed. Because of finding the relatedness of the



(a) HOLOS-crop (b) HOLOS-animal

Fig. 3.2: Feature Importance For Each Target Column For HOLOS Whole-farm Dataset.

features, the correlation matrix helped to understand which features are most important in this dataset.

Second, feature engineering is done. For the sake of the objective where the crop is the focal point to crop selection with farm management rotational recommendations. So, the crop rotational index is created using each consecutive year of a different crop. With the target column ready to be analyzed, the dataset was carefully preprocessed handling the numerical, categorical, date type data. Once the dataset is preprocessed and has all the desired features, the model application step is the next.

In Figure 3.2a the features most impacting features for the HOLOS crop dataset is shown. Some of the most important features for reduction of emission are Crop Perennial in LumCMax, Yield, AGR Ratio, Crop Perennial of k, Tillage LumC in Max.

For holos the perennial crops are selected for the classification model. The dataset is divided as training set 70%, testing set size 15% and validation set size 15%. Some other features are important for the farm management and crop rotational suggestions. So, besides the features collected from the feature engineering step, there are some more features taken into consideration for the crop classification step. The hyper parameter tuning is done using the 'n_estimators': [100, 200, 300], 'max_depth': [10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4].

For the crop recommendation step both clustering and regression algorithms are used. Hyperparameter tuning settings for K-means clustering are as 'n_estimators': [100, 300, 500], 'max_depth': [3, 5, 7], 'learning_rate': [0.01, 0.05, 0.1], 'subsample': [0.7, 0.8, 0.9], 'colsample_bytree': [0.7, 0.8, 0.9], 'gamma': [0, 0.1, 0.2], 'reg_alpha': [0, 0.1, 0.5], 'reg_lambda': [1, 1.5, 2]. The crop rotation index is used as the target column for the rotational suggestion. Here, for each cluster, randomized search is performed to find the optimal hyperparameter for the XGBOOST regression.

HOLOS Animal Data

This analysis of animal based agriculture based farm management explores the efficiency and promoting sustainable practices. Holos-Animal will be minimizing carbon emission and energy consumption while maximizing the productivity of animal welfare. Lime is then utilized for both the Pareto optimization and the ϵ-constraint model selection. The goal is to use resources as efficiently as possible while minimizing CO2 emissions. The Figure 3.2b is important features with valuable ranged data of HOLOS animal based dataset. From a farm management perspective, this information is useful to understand different farming practices. For example, managing livestock weight and emissions might help to optimize energy use and efficiency.

A thorough model is built to improve production operations by utilizing all of the parameters included in the dataset, such as animal characteristics, diet composition, housing, and waste management, among others. The goal is to maximize output and animal welfare while minimizing overall energy use and emissions. Here, ϵ-constraint and pareto optimization are the machine learning models most appropriate for this use case.

This is a framework, not a single model, for addressing issues with competing goals. In your situation, increasing productivity and reducing energy and emissions are at odds. MOO techniques, such as Pareto optimization or ϵ-constraint, are able to provide solutions that strike a good compromise between these objectives.

The holos application collects agricultural related data. Both crop and animal management data is collected to calculate the carbon footprint of that certain farm setting. From the application using some pre-ranged data of farm managements, the dataset is built. The whole data is divided into two groups. HOLOS crop data and HOLOS animal data.

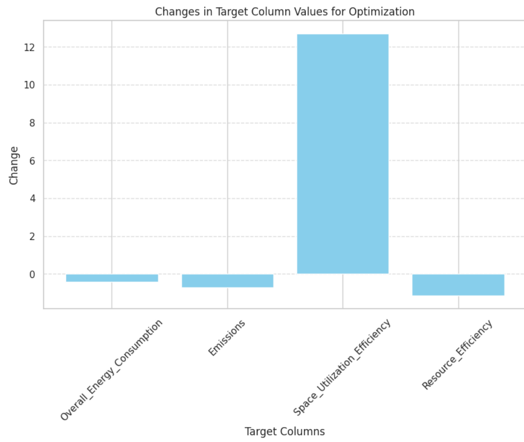
In this part, data collected from the HOLOS crop will be analyzed. First the data is preprocessed. Handling the missing values, handling impute missing values using techniques like mean, median, mode imputation is considered for this analysis. After that, actively identifying the outliers using statistical methods like Z-score, interquartile range (IQR). in the part not only outliers are observed but also they are handled using winsortization or logarithmic transformation. Categorical data is properly labeled. StandardScaler is used to scale the data.

For the feature engineering part, the overall energy consumption, emission, productivity, and animal welfare scare are calculated using the existing features from the dataset. Next, the dataset is ready for the model application. For the optimization task, there is a need of observing the target column ranges which will be controlling the action of the target column. The overall goal is to learn the data and use it to predict more optimized states. As some features will be maximizing and some will be decreasing, the understanding of the ranges of these features are crucial. So, first the random forest model used as a regression model targeting the four desired features. After that, LIME is used to explain the model. The explanation gives the ranges of the features.

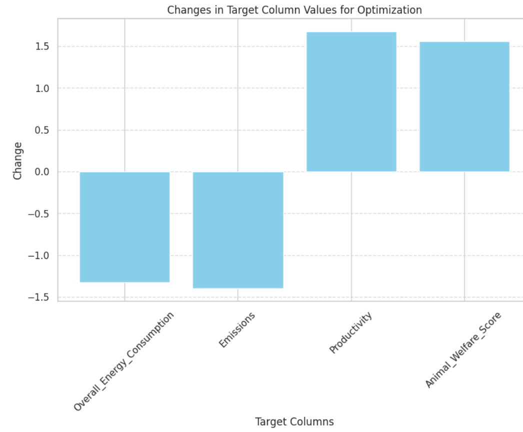
Then the dataset is split to apply the ϵ-constraint and pareto optimization algorithms. Pareto fronts demonstrate the multidimensional design space. And ϵ-constraint or epsilon constraint model ensures the optimal states for one object while keeping the other factors constrained to specific values. The pareto front and epsilon constraint models are introduced to the dataset by checking the data points, determining which target columns to focus on, fitting the data then combining the whole structure of the object which is minimization and maximization of certain features.

3.2.3 Optimization Using Machine Learning Algorithms

The workflow of this project mainly follows the optimization of the production process for reduced CO2 emission. Utilizing all the important features in the dataset, a comprehensive model is built to optimize the production process. The objective is to minimize overall energy consumption and emission while maximizing productivity and farming wel-



(a) COMET-Farm Target Columns.



(b) HOLOS Target Columns.

Fig. 3.3: Target Column States After Optimization.

fare. Then examining and promoting energy efficiency based farm management techniques for different crops and rotational strategies.

For this purpose, Multi-objective optimization algorithms like pareto optimization and epsilon constraint optimization techniques are used. These algorithms are applied to find the trade-off solutions between conflicting objectives. Then applying random forest classifier along with KMeans clustering and XGBoost regressor for the rotational strategy recommendations. FInally generative AI is used in more popular farm scenarios. In Figure 3.3a and Figure 3.3b we can see how optimization algorithm worked on the dataset. Here is the plot where we can observe the reduction and boosting factors of this project.

3.2.4 Generative AI application

In order to populate the optimized farm scenarios generative adversarial networks (GAN) is used. This framework leans from the base cases and generates new data using similar characteristics. GAN uses two different neural network models to learn and classify the bases cases properly. The first neural network is known as generator network and second one is known as discriminator network (CNN). In a mathematical way, GAN uses probability distribution with a loss function. In this project GAN is used after finding the optimal states to generate more optimal farm based management systems.

Table 3.1: COMET-Farm Data Input.

Farm Operation	User Input Required	Defaults Provided	Emission Calculated
Annual Crop/ Hay/ Grass or Seasonal Crop	Type of crop, Planting Date, Harvest Date, Yield, Grazing start date, Grazing end date, Grazing rest period, Implement date, Implement pass fertilizer date Fertilizer type Fertilizer total N applied, Manure Date, Manure Type, Manure amount applied, Irrigation date, Inches per application, Liming date, Liming type, Burning	% straw, Soil Total fertilizer applied Moisture % Total nitrogen% Ammonium nitrogen %	Soil N20 Carbon Storage & emission Energy CO2
Orchard/ Vineyard crop	Type of crop, Prune, Renew, Implement date, Implement pass fertilizer date Fertilizer type, Fertilizer total N applied, Manure Date, Manure Type, Manure amount applied, Irrigation date, Inches per application, Liming date, Liming type, Burning	Total fertilizer applied Moisture % Total nitrogen % Ammonium nitrogen %	Soil N20 Carbon Storage & emission Energy CO2
Beef Feedlot	diet on Feedlot, # Beef heifers, # Beef Steers, Primary breed, Average daily heifers, weight gain, Average daily weight steers, Average live weight steers Type of feed Feed % Feeding situation Solid/liquid separator Separator type Solid treatment method Liquid treatment method, Manure system	Typical mature weight % of solid removed Manure produced per day % nitrogen % content in manure	Entire CH4 Manure CH4 Manure CH4 Manure N20 Energy CO2
Beef Steer or Heifers stockers	unique herds # Beef steers or heifers Average body weight Average daily weight gain Average mature weight Hours work each day Type of feed Feed % Feeding situation Solid/ liquid separator Separator type Solid treatment method Liquid treatment method Manure system	Typical mature weight % of solid removed Manure produced per day % nitrogen % content in manure	Entire CH4 Manure CH4 Manure CH4 Manure N20

Table 3.1: COMET-Farm Data Input.

Farm Operation	User Input Required	Defaults Provided	Emission Calculated
Dairy Lactating Cow	# unique herds # Roofed facility # Dry Lot # Pasture Range Daily feed intake Average live body weight Days in milk Type of feed Feed % Housing type days manure in housing Feeding situation Solid/liquid separator Separator type Solid treatment method Liquid treatment method Manure system	% of solid removed Manure produced per day % nitrogen % content in manure	Entire CH4 Manure CH4 Manure CH4 Manure N20 Energy CO2

Table 3.2: HOLOS-Farm Data Input.

Farm Operation	User Input Required	Defaults Provided	Emission Calculated
Crops/ Grassland / land use change	Area of annual crops & fallow, Area of perennial crop Area of grassland, Tillage System, Area of irrigation, Herbicide usage	Fertilizer inputs Crop yields Soil type and texture	Soil N20 Soil Carbon Storage or Emission Energy CO2
Beef Cow-calf	# Cows, Types of grazing area, Pasture and feed quality Feed additive in diet Spring or fall calving, Year round grazing or winter feeding, Calves sold or kept for back grounding, # months kept, Manure handling system	Calf crop rate # bulls	Entire CH4 Manure CH4 Manure N20 Energy CO2
Beef Feedlot	Type of feedlot, Feedlot capacity and/or # months filled, Barn housing usage, Ration mix, Feed additive in diet, % steers in lot, Feed gain ratio, Average daily gain, Manure handling system	Initial and Final weights	Entire CH4 Manure CH4 Manure N20 Energy CO2
Beef Stocker	# cattle, # months grazed, Pasture quality, Feed additive in diet, % steers in herd, Average daily gain	Initial and Final weights	Entire CH4 Manure CH4 Manure N20
Dairy	# unique herds # Roofed facility # Dry Lot # Pasture Range Daily feed intake Average live body weight Days in milk Type of feed Feed % Housing type days manure in housing Feeding situation Solid/liquid separator Separator type Solid treatment method Liquid treatment method Manure system	% of solid removed Manure produced per day % nitrogen % content in manure	Entire CH4 Manure CH4 Manure CH4 Manure N20 Energy CO2

Table 3.3: COMET-Farm Data Input Ranges.

Variable	Sub-Variable	Options
Step 1 - Activities	Selected Activities for the Current Project	Cropland, Pasture, Orchards/Vineyards Range
Step 2 - Field Management	Parcel Locations	Orchard File Forage File
Historic Management	Pre-1980 Management	Irrigation (Pre-1980s) Livestock grazing Upland non-irrigated (Pre-1980s) Lowland non-irrigation (Pre-1980s)
	1980-2000 Management	annual crops in rotation continuous hay/ pasture in rotation continuous hay annual crops in rotation continuous hay livestock grazing fallow-grain orchard or vineyard
	1980-2000 Tillage Intensity	No till Reduced tillage Intensive tillage
Baseline	Type of crop	Orchard/Vineyard Crop Annual crop/hay/grass Seasonal cover crop Management
	Crop Selection	For Orchard/Vineyard Crop : Cherries Grape-raisan Grape-table Grape-wine (¿1950 GDD) Peaches and Nectarines Pistachio For Annual & crop/hay/grass : Alfalfa Barley Corn Corn silage Grass Grass-legume mix Oats Rye Sorghum Spring wheat Winter wheat. For Seasonal cover crop : Annual Rye - Legume - Radish Annual Rye Legume Annual Rye Austrian Winter Pea Cereal Rye Clover Corn Forage Radish Millet Oilseed Radish Winter Grain - Other Sorghum Vetch
	Did you prune?	For Orchard/Vineyard Crop : Yes or No
	Did renew or clear your orchard/ vineyard this year?	For Orchard/Vineyard Crop : Yes or No
	Planting Date	For Annual crop/hay/grass : 5/1 to 7/1 For Seasonal cover crop : 9/1 to 11/1
	Harvest date	For Annual crop/hay/grass : 7/1 to 11/1 For Seasonal cover crop : 4/1 to 6/1
	Harvest Straw	For Annual crop/hay/grass : 12 to 72 For Seasonal cover crop : 12 to 72
	Harvest Yield	For Annual crop/hay/grass : Alfalfa - 4.3 Tons/acre Barley - 93 bu/acre Corn - 24 tons/acre Corn silage - 24 tons/acre Grass - 138 bu/acre Grass-legume mix Oats - 154 bu/acre Rye - 154 bu/acre Sorghum - 54 bu/acre Spring wheat - 54 bu/acre Winter wheat - 54 bu/acre
	Grazing Start Date Grazing End date Rest Period	For Annual crop/hay/grass : 4/1 10/1 21

Table 3.3: COMET-Farm Data Input Ranges.

Variable	Sub-Variable	Options
	Tillage, Implements and Planting	For Orchard/Vineyard Crop : Mow For Annual crop/hay/grass : No Tillage Mow Zero Soil Disturbance No implement passes For Seasonal cover crop : Intensive Tillage Reduced Tillage Mulch Tillage Ridge Tillage Strip Tillage No Tillage Crimp Broadcast Seed Aerial Seed
	Fertilizer Type	For Orchard/Vineyard or Seasonal or Annual crop:Ammonium Nitrate (34-0-0) Ammonium Nitrate Phosphate (23-23-0) Ammonium Nitrate Phosphate (27-14-0) Ammonium Phosphate Sulphate (16-20-0) Ammonium Polyphosphate Solution (10- 34-0) Ammonium Sulphate (21-0-0) Ammonium Thiosulphate Solution (12-0- 0) Anhydrous Ammonia (gas) (82-00-00) Calcium Ammonium Nitrate Calcium Nitrate Diammonium Phosphate (18-46-0) Element-N (N) Element-P (P) Mixed Blends Monoammonium Phosphate (11-55-00) Monoammonium Phosphate (12-51-00) Potassium Nitrate Urea (46-00-00) Urea Ammonium Nitrate (30-00-00) Urea Ammonium Phosphate (27-27-00) Urea Ammonium Phosphate (34-17-00)
	Fertilizer Date	For Orchard/Vineyard Crop : May 01 For Annual crop/hay/grass : Two weeks after planting date For Seasonal cover crop : Two weeks after planting date
	Fertilizer Total N Applied	For Orchard/Vineyard or Seasonal or Annual crop: 19 - 151
	Manure Type	For Orchard/Vineyard or Seasonal or Annual crop: Alfalfa Meal Beef Manure, Solid Beef Slurry Blood, Dried Bone Meal Chicken - Broiler (litter), Solid Chicken - Broiler Slurry Chicken - Layer Slurry Chicken Layer - Solid Compost or Composted Manure, Solid Dairy Manure, Solid Dairy Slurry Farmyard Manure, Solid Feather Meal, Solid Feather Meal Fish Emulsion Fish Scrap Guano Horse Manure, Solid Other Manure, Solid Sheep Manure, Solid Soybean Meal Swine Manure, Slurry Swine Manure, Solid
	Manure Date	For Orchard/Vineyard Crop : May 01 For Annual crop/hay/grass : Two weeks before Planting Date For Seasonal cover crop : Two weeks before Planting Date
	Manure Applied Amount	For Orchard/Vineyard or Seasonal or Annual crop: 5 - 25

Table 3.3: COMET-Farm Data Input Ranges.

Variable	Sub-Variable	Options
	Irrigation Inches per application	For Orchard/Vineyard Crop : Once per week starting 5/1 - 11/1 For Annual crop/hay/grass : Once every 11 days starting 5/1 - 11/1 (starts from planting date) For Seasonal cover crop : Once per week 9/1 to 6/1(starts from planting date)
	Liming	For Orchard/Vineyard or Seasonal or Annual crop: None
	Burning	For Orchard/Vineyard or Seasonal or Annual crop: No Burning
Step 3-Animal Agriculture	Units of measure	Metric
	Select Animal Types	Beef-heifer stockers Beef-steer stockers Stockers Feedlot cattle Dairy-lactating cows
	# unique herds	For Beef-steer and Beef-heifer Stockers: 1 For Dairy-lactating cows: 1
	# diets fed on feedlot	For Stockers Feedlot cattle: 1
	# Heifers	For Stockers Feedlot cattle and Beef-heifer Stockers : 1 - 500
	# Steers	For Stockers Feedlot cattle and Beef-steer Stockers: 1 - 500
	Average weight stockers body	For Beef-steer & Beef-heifer Stockers:362
	Average stockers weight gain daily	For Beef-steer and Beef-heifer Stockers: 0.6 - 2.5
	Average stockers mature weight	For Beef-steer & Beef-heifer Stockers:545
	Primary Breed	For Stockers Feedlot cattle: Angus Brahman Charolais Chianina Gelbvieh Hereford Limousin Main Anjou Pinzgauer Red Poll Sahiwal Simmental South Devon Tarentaise Other
	Average Daily gain weight	For Stockers Feedlot cattle: for Spring, Summer, Fall, Winter for both Heifers and Steers 0.6 - 2.5
	Average Live weight	For Stockers Feedlot cattle: from Jan - Dec for both Heifers and Steers 362 - 545
	Do you use ionophores	For Stockers Feedlot cattle: Yes or No
	Fat Content in the diet	For Stockers Feedlot cattle: 1% Supplemental Fat 2% Supplemental Fat Four or higher added fat content No supplemental fat added
	Grain Type in diet	For Stockers Feedlot cattle: Steam Flaked or High Moisture Unprocessed or Dry Rolled Barley rather than corn or sorghum
	Concentrate % in diet	For Stockers Feedlot cattle: More than 60% grain 45 to 60% grain Less than 45% grain

Table 3.3: COMET-Farm Data Input Ranges.

Variable	Sub-Variable	Options
	# Roofed Facility	For Dairy-lactating cows: 1 - 500
	# Dry Lot	For Dairy-lactating cows: 1 - 500
	# Pasture Range	For Dairy-lactating cows: 1 - 500
	Average daily feed in-take	For Dairy-lactating cows: 0
	Average weight live body	For Dairy-lactating cows: 362 - 545
	Average daily feed in-take	For Dairy-lactating cows: 305
	Days in milk	For Dairy-lactating cows: 28.6
	Range feed of months for	For For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers: Select Jan to Dec
	Types of Feed	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers: Alfalfa Birdsfoot Bromegrass Cheatgrass Elephant grass Grain Grass Meadow Oat Orchardgrass Prairie Rye Sanfoin Sorghum Sudangrass Vetch Wheat Wheatgrass
	Percentage of feed	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers: 100 %
	Primary Feeding Situation	For Beef-steer and Beef-heifer Stockers: Stall Pasture Grazing large areas
	Housing Type	For Dairy-lactating cows: Pit Storage Bedded Pack Flushed or Scraped
	Pit Storage Type	For Dairy-lactating cows: Deep Shallow
	Pit Storage housing days in	For Dairy-lactating cows: 1-7
	Bedded Pack Type	For Dairy-lactating cows: Active Mix No Mix
	Bedded pack housing days in	For Dairy-lactating cows: 1-7
	Flushed or Scraped area of Barn	For Dairy-lactating cows: 0.4 - 2.8
	Use Solid/Liquid Separator	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Yes or No
	Storage Method	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Temporary stack and long-term stockpile Composting Aerobic lagoon Anaerobic digester with biogas utilization or methane capture Thermochemical Conversion (Pyrolysis, Incineration, Gasification) Constructed wetland Daily spread Deposited on Pasture/Range/Paddock Removed offsite Land applied
	Separator type	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Stationary inclined screen Vibrating screen Rotating screen Centrifuge Decanter centrifuge Roller press Pressure filter Other

Table 3.3: COMET-Farm Data Input Ranges.

Variable	Sub-Variable	Options
	Solid Method Treatment	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Temporary stack and long-term stockpile Composting Thermochemical Conversion (Pyrolysis, Incineration, Gasification) Daily spread Deposited on Pasture/Range/ Paddock Removed offsite Land applied
	Liquid Method Treatment	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Aerobic lagoon Anaerobic digester with biogas utilization or methane capture Thermochemical Conversion (Pyrolysis, Incineration, Gasification) Constructed wetland Daily spread Deposited on Pasture/ Range/Paddock Removed offsite Land applied
	Solid Storage Cover Type System	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: AUncovered solid Covered solid Uncovered semi-solid Covered semi-solid
	Solid Storage Is the manure stored for more or less than six months?	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Long term (more than six months) Short term (less than six months)
	Composting Method	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: In Vessel Static Pile Intensive Windrow Passive Windrow
	Aerobic Volume lagoon Total	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: 400 - 20,000
	Aerobic lagoon Is the system aerated naturally or by forced aeration?	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Natural Aeration Forced Aeration
	Anaerobic digester with biogas utilization or methane capture digestre Type	For Dairy-lactating cows and Beef-steer and Beef-heifer Stockers and feedlot cattle: Steel or lined concrete or fiberglass digesters with a gas holding system (egg shaped digesters) and monolithic construction. Up-flow anaerobic sludge blanket (UASB) type with floating gas holders and no external seal Unlined concrete/ferrocement/ brick masonry arched type gas holding section and monolithic fixed dome digesters Other

Table 3.4: HOLOS-Farm Data Input Ranges.

Farm Operation	Min Value	Max Value	Multiple Option
Farm Name			Any Name
Farm year	1900	2050	
Eco district	358	1091	
Province			Alberta
Ecozone			Smiarid Pairies
Soil Texture			Fine Medium Coarse
Soil Type			Black/Gray Chernozem Brown Chhernozem Dark Chernozem Eastern Canada
Present Till Management Practice			No Till Reduce Intensive
Past Till Management Practice			No Till Reduce Intensive
LumC Max	-6071	6071	
k	0.0091	0.35	
Precipitation(mm)	0	1524	
Potential Evapotran spiration(mm)	0	1524	
F topography	0	100	
Soil N2O Breakdown(%)	0	100	
Average Temp C	-100	100	
RF Texture	0	1	
RF Tillage	0	1	
EF Eco (kgN2O-N/kgN)	0.0016	0.0017	
Leaching Fraction	0.05	0.3	
EF Leaching (kgN2O-N/kgN)	0.0075	0.0075	
Volatilization Fraction	0.1	0.1	
EF Volatilization (kg N2O-N/kgN)	0.01	0.01	
Global Warming Potential (CO2)	1	1	
Global Warming Potential (CH4)	28	28	
Global Warming Potential (N2O)	265	265	
Crop - Perennial			Hay - Grass Hay - Legume Hay - Mixed Hay and Forage Seed Other
Area (ha)	0.1	3	
Year Seeded	1910	2020	
Yield (kg DM/ha)	2242	13450	
Stand Length (yrs.)	1	10	
Irrigated			Yes or No
Herbicide N Fertilization Rate (kgN/ha)	22	169	Yes or No
P Fertilization Rate (kgP2O5/ha)	56	135	
Moisture Content (w/w)	0.02	0.75	
AGR N Concentration (kgN/kg)	0.005	0.018	
BGR N Concentration (kgN/kg)	0.007	0.015	

Table 3.4: HOLOS-Farm Data Input Ranges.

Farm Operation	Min Value	Max Value	Multiple Option
Yield Ratio	0.12	0.72	
AGR Ratio	0.08	0.6	
BGR Ratio	0.08	0.6	
Fuel Energy (GJ/ha)	0.34	2.83	
Herbicide Energy (GJ/ha)	0	0.23	
LumCMax	-6071	6071	
k	0.0091	0.35	
Beef Feedlot Finish- ers Group 1			
Ash Content (%)	8	8	
Bo	0.19	0.19	
CD Steer	1	1	
CD Heifer	0.8	0.8	
# Days	0	31	
# Heifers	1	1000	
Initial Heifer Weight (kg)	362	363	
Final Heifter Weight (kg)	817	817	
Heifer ADG (kg)	0.6	2.5	
# Steers	1	1000	
Initial Steer Weight (kg)	362	363	
Final Steer Weight (kg)	545	545	
Steer ADG (kg)	0.6	2.5	
Soil Type			Black/Gray Chernozem Brown Chhernozem Dark Chernozem Eastern Canada
Housing			Confined No Barn Housed In Barn Enclosed Pasture Open Range Custom
Diet			Barley Corn Custom
Diet Additive			None 2% Fat 4% Fat Custom
Manure System			Pasture Solid Storage Compost Intensive Compost Passive Custom

CHAPTER 4

RESULTS

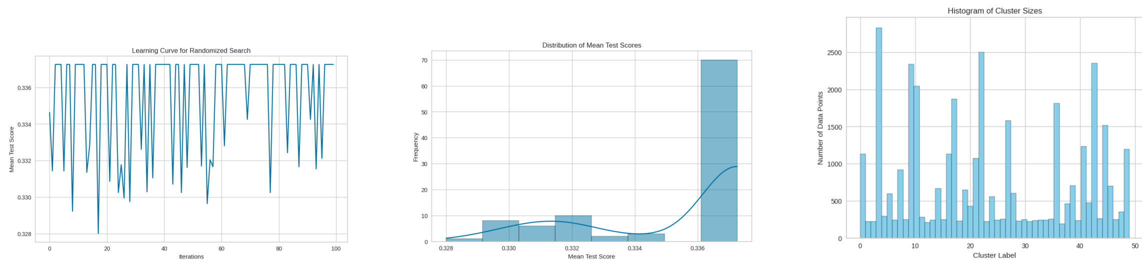
Farm management is a dynamic and continuous process. It is essential to regulate the evaluation and adaptation to record the changing conditions. A closer observation can be helpful to plan strategically, improve the resilience and efficiency of certain operations that can lead to a long term sustainability and success. The results of this project show how to strategically plan farm managements to maximize some factors while reducing GHGs emission.

4.1 COMET-Farm Model

Data collected from comet-farm website is highly valuable and complex. The collected data make it challenging to create comprehensive models that capture all the relevant outcomes. Overall, comet-farm website enable the best available evidence-based decision-making process.

The optimization algorithms are applied in the dataset and these Figure 4.1 show the outcomes after algorithm application. Here, the pareto front optimization works as the algorithm that cannot be improved in one objective direction without sacrificing performance in the other. The epsilon-constrained is optimization where one objective function is minimized by the optimizer while the other objective functions are constrained to specific values. Trade off between different features and show different correlations among features. We will be observing the reduction of emissions while maximizing different desired objectives. In the figure, it can be seen that the trade offs between emission and soil nutrients levels, the well distribution of soil health scores, the higher fertilizer efficiencies, the correlation between soil health and fertilizer score and the significance of ammonium nitrogen on soil N2O emission.

Fig. 4.1: Feature Trade-offs After Optimization.



(a) Learning Curve For Randomized Search For COMET Crop Dataset

(b) Distribution Of Mean Best Scores For COMET Crop Dataset

(c) Histogram Of Cluster Size For COMET Crop Dataset

Fig. 4.2: Machine Learning Models Application On Comet-Farm Crop Dataset.

4.1.1 COMET-Farm Crop Data

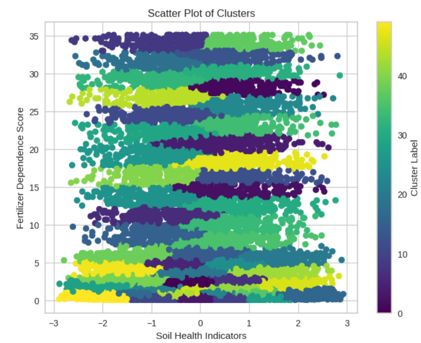
The purpose of this analysis is the crop selection and rotation recommendation. The results for the COMET-Farm crop data are discussed below.

Figure 4.2a shows the performance of the the random forest classifier. The curve plots the mean test score of the randomized search algorithm over the number of iterations. In the plot we can see that the curve trends slightly upward as the number of iteration increases. It means that the model is learning and improving itself as the number of iterations increases. However the curve flattens slightly towards the end. It means that the model is approaching its maximum performance which is the optimal state.

Figure 4.2b is a histogram plot where each bar is a range of mean test scores. The higher the bar, the greater likelihood that the groups scored within that range. The mean test scores vary only slightly so overall the plot shows little variance.

This part of the project demanded the necessity of usage of clustering algorithms. In order to recognize and capture the similar groups of items the K-Means algorithm is applied here. It helped to increase the chances of accurate prediction based on the previous behavior within the groups. Additionally, the K-Means algorithm helped to recognize the dynamics of each group. Figure 4.2c depicts the size (in data points) of each cluster. Most clusters are between 10 and 15 data points. However there are clusters with large numbers of data points too. A few clusters have between 40 and 50 data points. In short, the graph shows that clusters are formed with a small number of groups of data points where most of the

(a) Results After Classification Algorithm



(b) Scatter Plot Of Clusters For COMET Crop Dataset

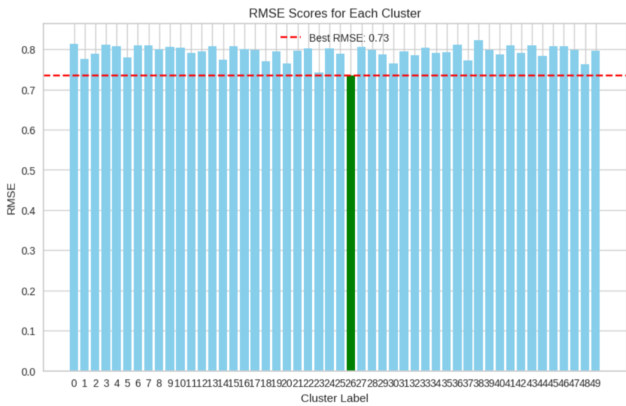
Fig. 4.3: Results After Classification And Clustering Algorithms.

groups are relatively small.

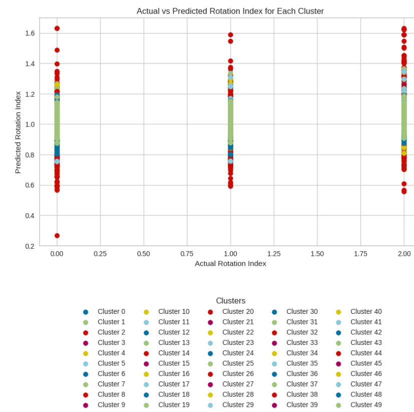
In the Figure 4.3a we can see the results of a classification algorithm where we can observe some of the most interesting facts. One of them is that the crop type has no impact on the emissions and soil health. This means the farm managements are nearly perfect when this data is collected. The farm scenario is an example where farmers along with the farm management helped reduce emission. Another fact is there is no clear trend between ammonium nitrogen and fertilizer dependence. This indicates that there is no noticeable pattern seen in these two features. So, there should be a variety of practices that are highly needed. Also, low and mid-scores indicates the optimal fertilizer practices.

Figure 4.3b is a scatter plot that shows the data points all together based on their shared common characteristic. Here we can see that the data points are grouped into mostly three different colors. This shows the positive correlation between fertilizer dependence score and the number of soil health indicators. The impacts of soil with a higher fertilizer dependence score tend to have more than the soil health indicators. Also, the data points can be seen as a clustered form. This suggests that there are three different types of soil. And, these different types of soils have their own characteristic relationship between fertilizer dependence score and soil health indicators.

Figure 4.4a shows the RMSE for each cluster. Here, we can see the performance of the difference between values predicted by the model and the values actually observed. The



(a) RMSE Scores For Each Cluster For COMET Crop Dataset.



(b) Actual VS Predicted Rotation Index For Each Cluster For COMET Crop Dataset.

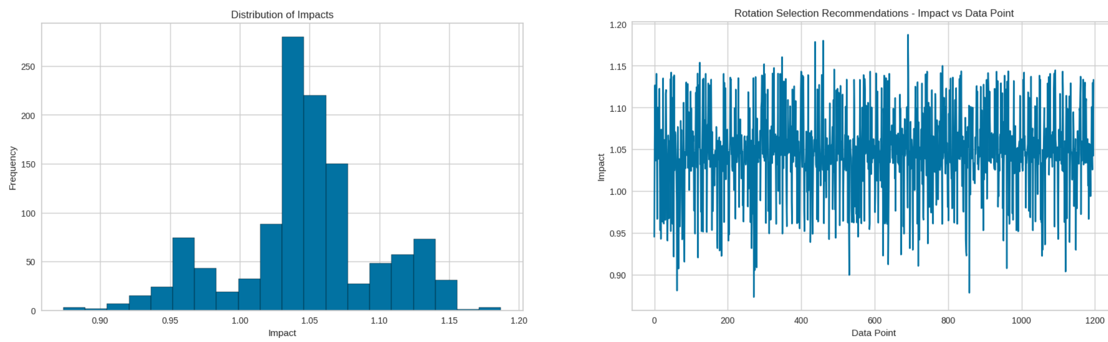
Fig. 4.4: Observation Of Each Cluster

green bar is the best RMSE score observed during the model cycle. And the best RMSE score is 0.73. The blue bars are the overall observations during the model execution. From this graph, we can understand each data point as clusters which show the RMSE score for that particular cluster. The scores vary across the clusters.

Figure 4.4b actual vs predicted rotation index for each cluster is the comparison between the actual and the predicted rotation indexes for fifty clusters. In this graph each data point is a cluster. Depending on these results, the crop recommendation and rotation selection recommendation plots are added later. As there are lots of farm recommendations, some of the recommended farm scenarios are added in this project to show results.

Figure 4.5a distribution of impacts demonstrates the larger scale of frequent impacts fall between 1.0 and 1.05. There are about 200 impacts in this range, the frequency of the impacts become less as the impact progresses further away from 1.0. Additionally, below 0.95 and above 1.15, there are less impacts. The histogram shows most of the impacts happened aunts 1.0. A few outliers on either side of 1.0. This means that the impact tends to be relatively similar in magnitude where there are a few exceptions.

Figure 4.5b rotation selection recommendations impact vs data point shows the relationship between impacts and data points which is the clusters. Here, we can see that



(a) Distribution Of Impacts For COMET Crop Dataset.

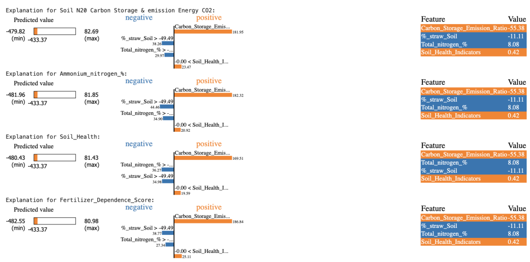
(b) Rotation Selection Recommendations Impact VS Data Point For COMET Crop Dataset.

Fig. 4.5: Analyzing Impacts of Each Cluster.

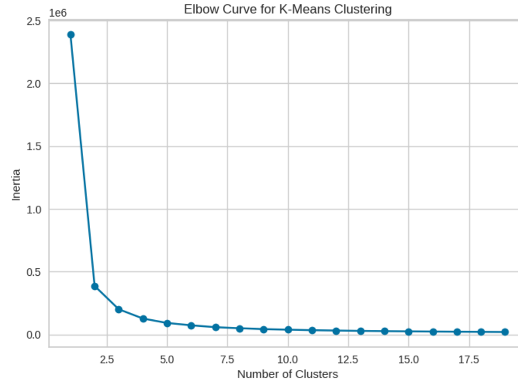
there is a weak positive correlation between the number of clusters and the impacts of the rotation selection recommendations. With the increment of the number of the data points, the impact of the rotation selection recommendations tends to increase. Also, there are many clusters with a low number of data points with a high impact.

In this Figure 4.6a the LIME explanations can be seen for the COMET crop dataset. Here, the carbon storage emission ratio consistently appears as a positive contributor in all the categories. And the consistency of this feature impacts the scores across different predictions. Besides, the positive importance and impacts of soil composition, oil health indicators.

Figure 4.6b elbow curve for K-means clustering represents the inertia (the sum of squared distances between data points and their allocated cluster centers) for different numbers of clusters populated by the K-means clustering algorithm. From this graph we can see that the number of clusters increments and the decrease of the inertia. Because of the increment of the number of clusters, the it generally decreases the overall distance between data points and their cluster centers. The general shape of each cluster is supposed to be smaller and tighter. So, this is the reason for the decrease of inherits with the increment of the number of clusters. But, the point or position where the diminishing returns no progress or further learning by adding more clusters. And that creates the elbow effect in the graph. In this graph the elbow showed up around the 4 or 5 clusters. Basically, the graph suggests

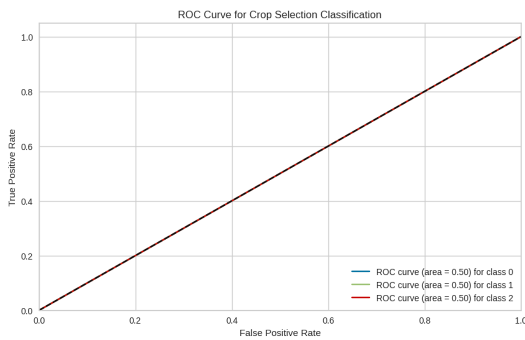


(a) LIME Explanation of COMET-Crop

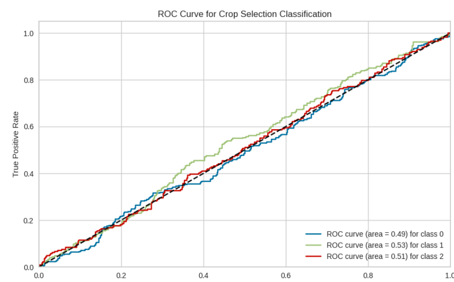


(b) Elbow Curve For K-Means Clustering For COMET Crop Dataset.

Fig. 4.6: Results of LIME Explanation And Elbow Curve.



(a) Rotation Selection Recommendations Impact VS Data Point For COMET Crop Dataset.



(b) A Closer Look At The ROC Curve For Crop Selection Classification For COMET Crop Dataset.

Fig. 4.7: ROC Curve For Crop Selection Classification For COMET Crop Dataset.

that picking 4 or 5 clusters reached the minimization of the inertia. So, considering more clusters beyond this level shows no remarkable decrease.

Figure 4.7a ROC curve for crop selection classification shows the tread-off between correctly identifying a crop which is the true positive rate and incorrectly identifying the crops which is the false positive rate. The curve appears as a straight diagonal line that means that the false positive rate and true positive rate are equal. There are three different classes representing three different types of crops.

A closer look at the above ROC curve can be seen in Fig. 4.7b. It shows some insights from the AUC (area under the curve). From the AUC we can understand the performance of a specific ROC curve. Here, three different crops have AUCs as 0.49, 0.53, 0.51. This means that the model is learning by making a few mistakes.

4.1.2 COMET-Farm Animal Data

Using the agriculture animal based data, the analysis is done by minimization of CO2 emission while maximizing effective resource usage. Then populating more optimized farm scenarios with GAN.

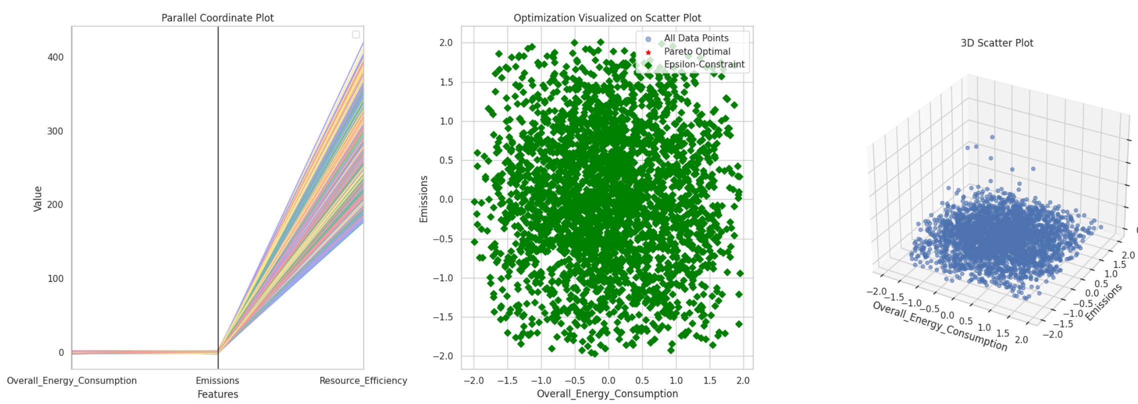
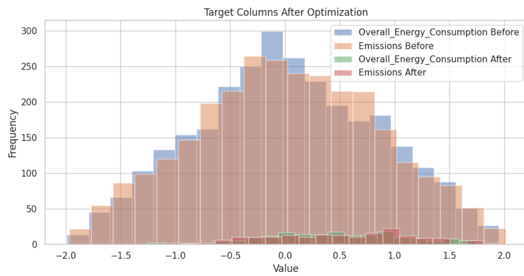


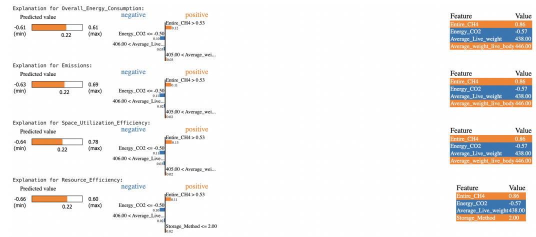
Fig. 4.8: Parallel Coordinate Plot, Optimization Visualization On Scatter Plot And 3D Scatter Plot For COMET Animal Dataset.

Figure 4.8 parallel coordinate plot, optimization visualization on scatter plot, 3d Scatter plot demonstrates different representation of the optimized states situations. Every features are shown as a line and then we can see that more resource efficiency tend to have lower emission. As lines are higher on the resource efficiency axis and lower on the emission axis. Also, the overall energy energy consumption is lower too after optimization. The second plot shows that the scatter plots of pareto optimal and epsilon constraint states. The pareto optimal shows the reduction of emission and overall energy consumption without sacrificing resource efficiency. And for the epsilon constraint optimization, keeping the level of resource efficiency fixed or minimum, the optimization is trying to minimize the emission and overall energy consumption. Also the resource efficiency is constrained.

The scatter plot shows the relationships between overall energy consumption against



(a) Target Columns After Optimization For COMET Animal Dataset.



(b) LIME Explanation of COMET-Animal Dataset.

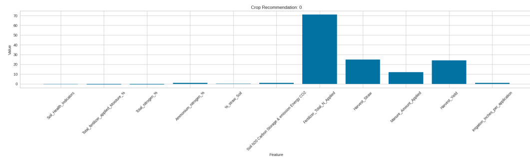
Fig. 4.9: Optimized Target Columns And LIME Explanation of COMET-Animal Dataset.

emission. The blue points represent the original data from the database. The red stars show the Pareto optimal states. Here, the Pareto optimal state suggests that one of the objectives can not be improved without worsening the other. In this management system, this is the most efficient configuration while considering the energy and emissions. The green diamonds show the results of the epsilon constraint model. This model sets a limit (epsilon value) on the secondary objectives and tries to optimize the primary objective.

These graphs show the changes in the farm practices after energy consumption, emissions, resource efficiency. This helps to make decisions while finding balanced farm management practices considering environmental impact and efficiency.

Figure 4.9a target columns after optimization shows the target columns effects after optimization. It shows the comparison between before overall energy consumption, emission states and after the optimized states of overall energy consumption, emission states. The graph shows a twin peak distribution with one peak is around -1.0 and another one is around 1.0.

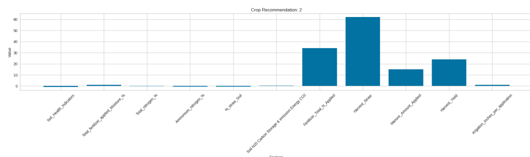
This graph shows the overall energy consumption before states which are in light blue bars. The histogram suggests a bell shaped distribution, which is a normal distribution. This indicates that most farms have an average level of energy consumption, with fewer farms having very high or very low energy consumption. The emissions before states are in light orange bars. This histogram also seems to be approximately normally distributed, centered around a middle value with tails extending to higher and lower emissions. It



(a) Comet Farm Crop-Selection and Rotational Recommendation.

(b) Comet Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

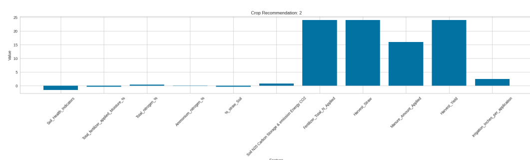
Fig. 4.10: Some Optimized Farm Scenarios 001.



(a) Comet Farm Crop-Selection and Rotational Recommendation.

(b) Comet Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

Fig. 4.11: Some Optimized Farm Scenarios 002.



(a) Comet Farm Crop-Selection and Rotational Recommendation.

(b) Comet Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

Fig. 4.12: Some Optimized Farm Scenarios 003.

shows that emissions across farms are also mostly average, with fewer at the extreme ends. These results suggest that the data is very likely to be real farming managements where the majority of data points have moderate energy usage and emissions levels, with fewer instances of very high or low values. The overall distribution shows the inefficiencies and areas for potential improvement. For example, if a farm is on the higher end of energy consumption, investigating and adopting more energy-efficient practices could be beneficial.

Now, from this data, it can be analyzed the identification of outliers which can lead to determining if there are farms with unusually high or low values for these metrics and understanding why. Also, defining targets for energy consumption and emissions reduction based on the distribution. Optimization strategies where moving the distribution towards more desirable outcomes such as leftward shift indicating reduced energy consumption and emissions.

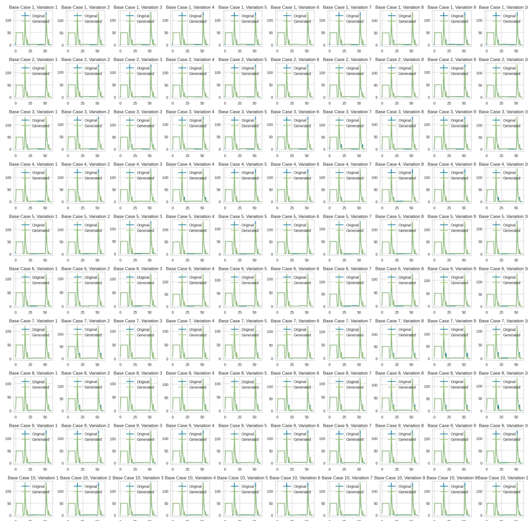
Overall energy consumption before optimization which is the light blue bars and after Optimization which is the dark blue bars suggests that a shift towards the left in the dark blue histogram. This indicates energy consumption after optimization. Also, the optimization has successfully reduced overall energy consumption across the dataset. Additionally, emissions before optimization which are the light orange bars and after optimization which are the dark orange bars show the energy consumption, the emissions histogram. They also have a leftward shift post-optimization. The reduction in the frequency of higher values implies that emissions have decreased the optimization.

In this Figure 4.9b the LIME explanation of the COMET animal dataset can be seen where it is clear that the inverse impacts of CO2 and CH4 levels. Also, the entire CH4 greater than 0.53 consistently shows a positive impact. Again, the CH4 levels, average live weight and CO2 emissions appear repeatedly.

Subsequently the optimization aided to find desired goals. The farm scenarios after optimization effects on other features are included here. The number of features are too many to add here in the discussion section. Here in these Figure 4.10, Figure 4.11 and Figure 4.12 come of the crop selection along with the rotational recommendations can be seen.

After achieving the optimized state, based on this information the generative adversarial network (GAN) is used to populate more optimized farming management scenarios. Figure 4.13a and Figure 4.13b show a few of the base case scenarios and the populated farming scenarios generated by the GAN. Now, the GAN application helps to generate more and new farm scenarios. In the GAN generated plots, we can see some of the generated farm scenarios of COMET crop based and COMET animal based farms.

This shows the comparison of the distribution of generated data to real data to validate how well the model captures the real-world distributions of key farm metrics. Also,



(a) GAN Generated Farm Scenarios For COMET-Crop Dataset.

(b) GAN Generated Farm Scenarios For COMET-Animal Dataset.

Fig. 4.13: GAN Generated Some Farm Scenarios.

evaluating and exploring different scenarios can be helpful to see how changes in certain conditions can affect farm outcomes. The new generated data is useful to identify the range of outcomes for various metrics under different scenarios to assess potential risks and the robustness of farm management strategies.

Figure 4.14 In this graph, the Overall Energy Consumption and Emissions plots show that both real and generated data show a similar distribution. It suggests that the model generating the data can closely replicate or predict the actual energy consumption and emission patterns observed in the real data. The Space Utilization Efficiency plot shows there is a significant difference between the real and generated data. The real data show higher counts for low values of space utilization efficiency, while the generated data do not capture this skewness as effectively. Again the Resource Efficiency distribution plot shows the real and generated resource efficiency data show some discrepancies, particularly around the peak counts, suggesting the GAN model could not fully replicate the actual distribution of resource efficiency.

Here, after the application of GAN, the target columns are observed to find interesting outcomes. In the plots, we can see the target columns of the after GAN application. From

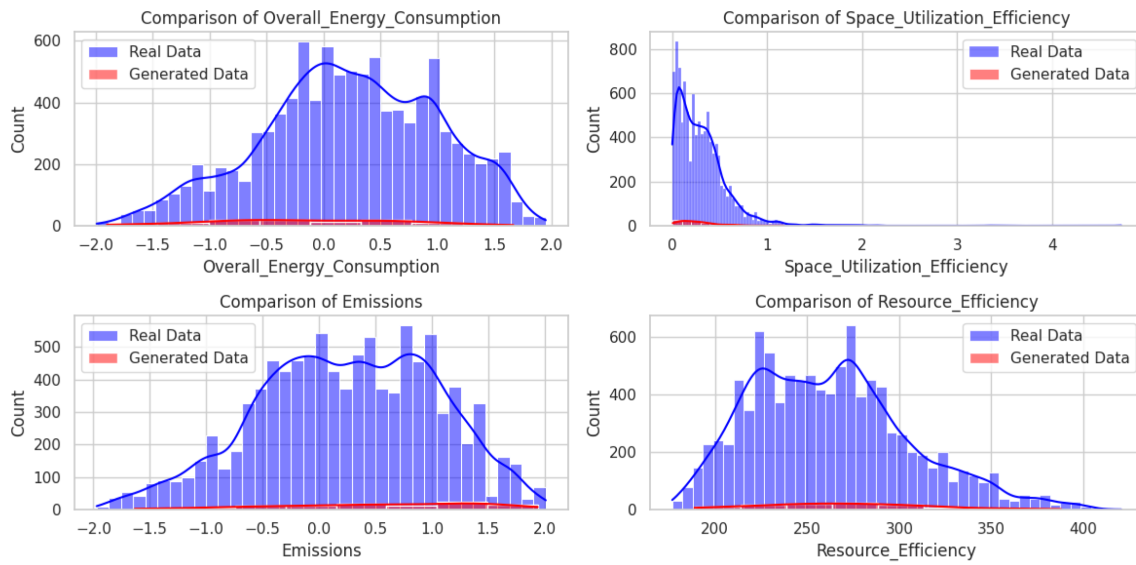


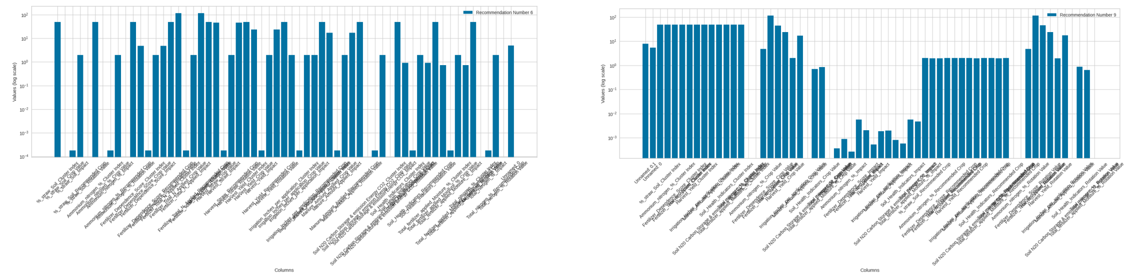
Fig. 4.14: Target Columns After GAN Application For COMET Animal Dataset.

this, we can say that the generated data is closely matched with the real data distribution. Sometimes, limitations in the abilities of generative models can be seen too. The limitations are mostly happening in the lower end values. Also, the fit indicates the effective performance in capturing the central tendencies.

The comparison between real and generated data can help validate and improve predictive models used in farm management. For space utilization efficiency and resource efficiency, there are notable differences between the real and generated data. Model tuning is necessary for further analysis.

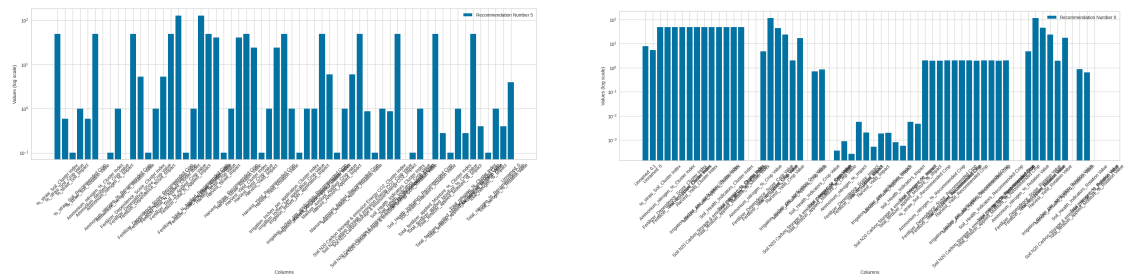
As the generated data comes from a predictive model, the model is more reliable for predicting energy consumption and emissions rather than space and resource efficiency. So, if energy consumption predictions are accurate, managers could focus on optimizing energy usage on the farm based on model projections. Generating data can be used to test various "what-if" scenarios and their impact on farm performance, helping in strategic planning and risk management. But the overall goal is to refine the real data into a generated database where the replication of data will be nearly accurate to the original database.

In Figure 4.15, Figure 4.16 and Figure 4.17 we can see some of the optimized and GAN generated new farm management scenarios for COMET crop based practices. Next, some



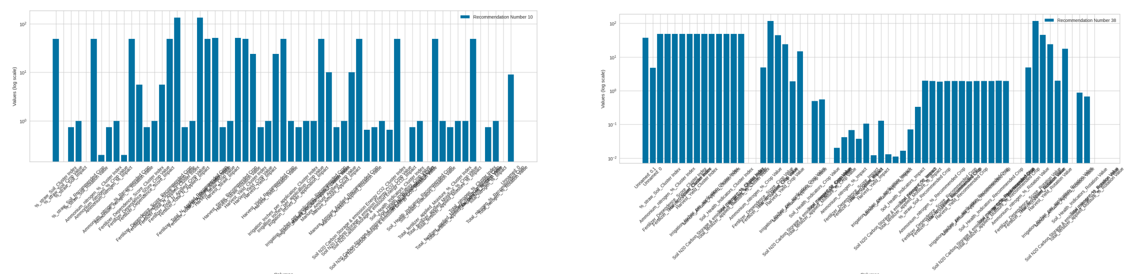
(a) (b)

Fig. 4.15: Comet Crop-Optimized And GAN Generated Farm Scenarios 001.



(a) (b)

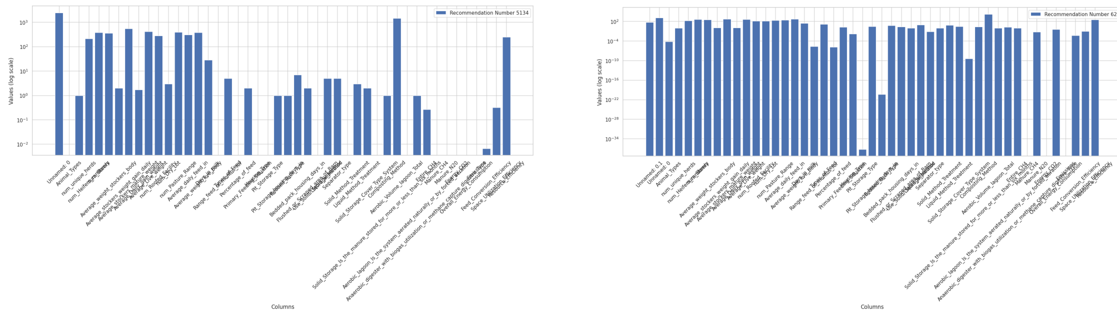
Fig. 4.16: Comet Crop-Optimized And GAN Generated Farm Scenarios 002.



(a) (b)

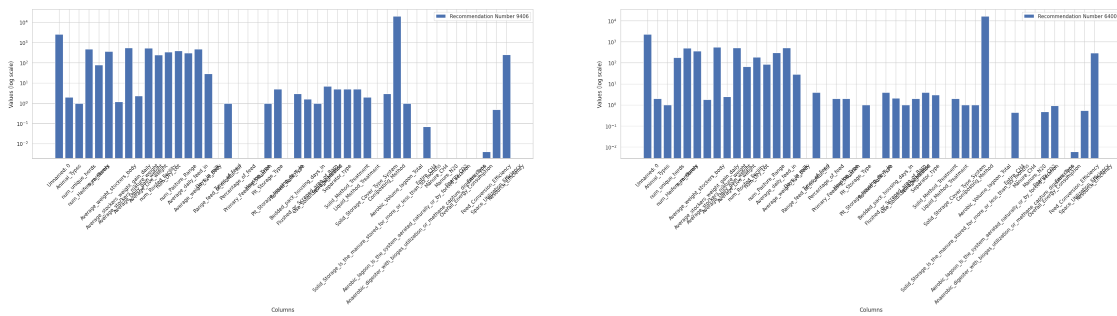
Fig. 4.17: Comet Crop-Optimized And GAN Generated Farm Scenarios 003.

of the optimized and GAN generated COMET animal based farm management scenarios can be seen in Figure 4.18 and Figure 4.19.



(a) (b)

Fig. 4.18: Comet Animal-Optimized And GAN Generated Farm Scenarios 001.



(a) (b)

Fig. 4.19: Comet Animal-Optimized And GAN Generated Farm Scenarios 002.

4.2 HOLOS Whole-Farm Model

In this part, we will be exploring the results gained from the collected data of HOLOS whole-farm model. One is crop related data and another one is livestock based data which are collected using HOLOS software application based tool.

In this Figure 4.20 we can see the trade-offs of HOLOS dataset. The figure shows the overall correlation while the optimization is applied on the dataset. From the plot, there are important trade-offs that can be seen in the energy CO2 emission and the storage values. This indicates that a range of CO2 emissions or energy storage efficiency is optimized. Next, other notable factors are the reduction in soil N2O emissions, yield ratio optimization, impact of precipitation, interdependencies among matrics like changes in N2o breakdown impact yield ratios or how energy CO2 emissions correlates with precipitation level.

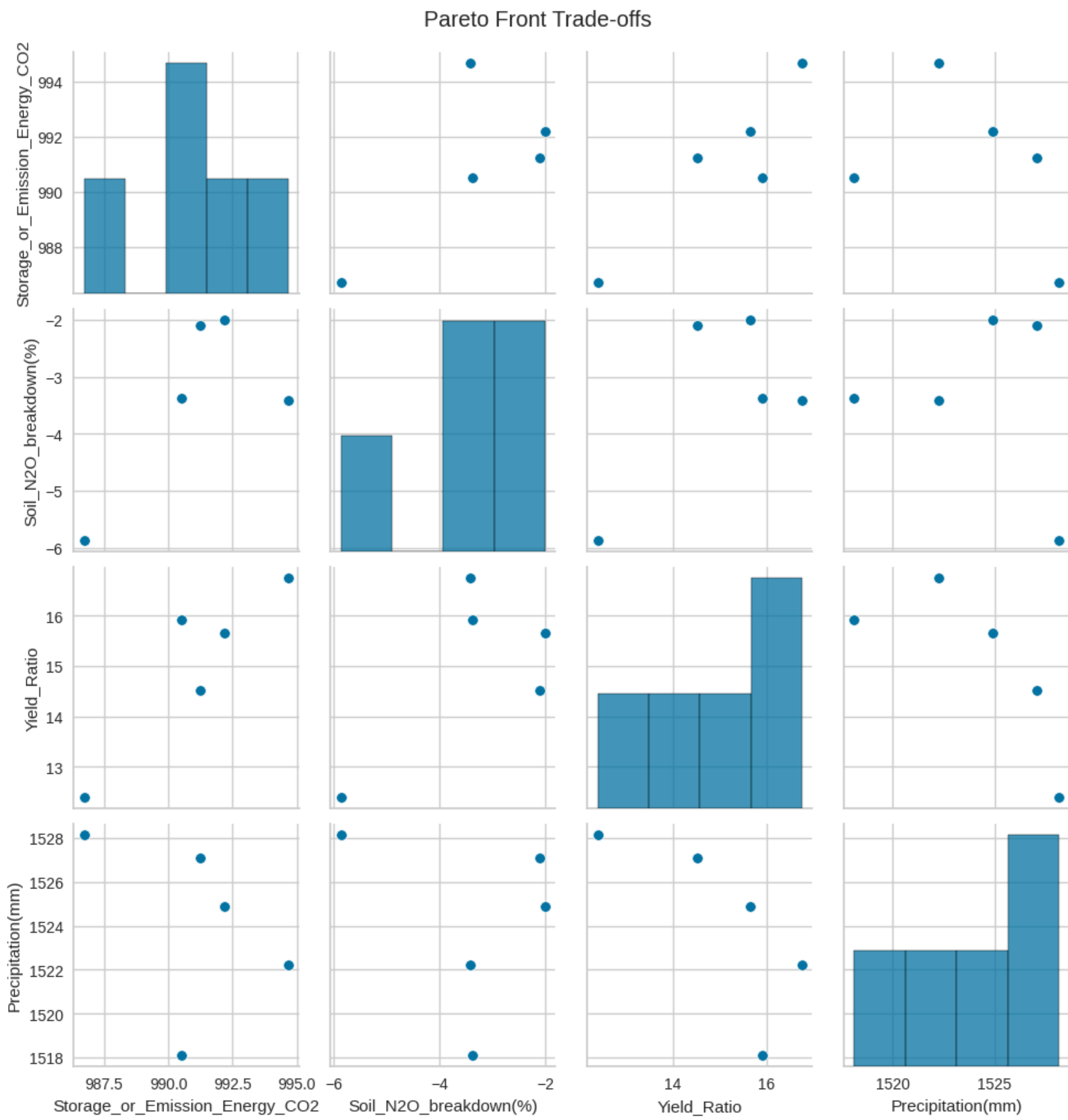
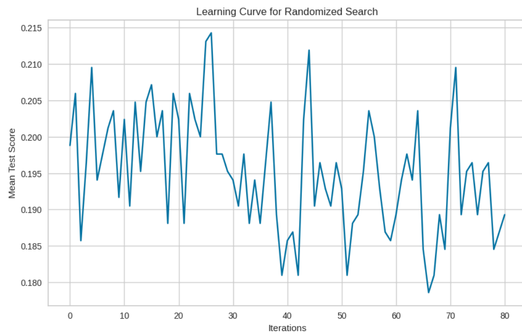
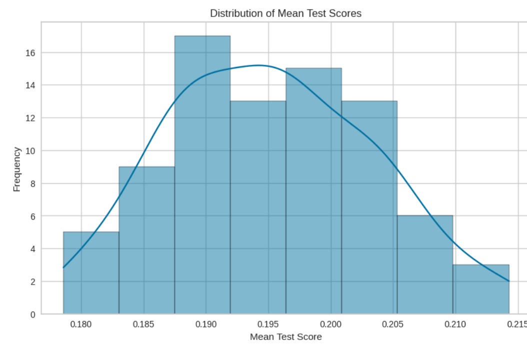


Fig. 4.20: Feature Trade-offs of Optimization For HOLOS-Whole Farm.



(a) Learning Curve For Randomized Search For HOLOS Crop Dataset.



(b) Distribution of Mean Test Scores For HOLOS Crop Dataset.

Fig. 4.21: Analyzing Results of Machine Learning Algorithms On HOLOS-Crop Dataset.

4.2.1 HOLOS Crop Data

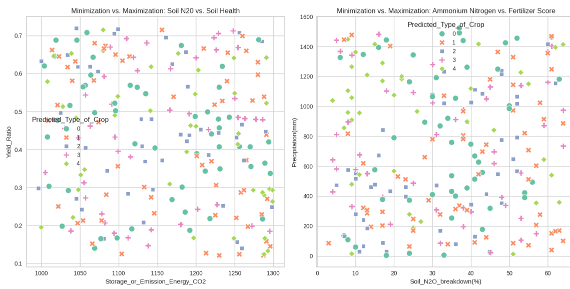
With the HOLOS crop data, the application of machine learning and optimization algorithms support the achievement of the crop selection and rotation recommendations.

Figure 4.21a learning curve for randomized search graph shows the performance of the algorithm. The performance improves with the number of iteration increases. The curve starts to flatten out at around the 50th iteration and that is the maximum performance of this algorithm. Overall the graph shows that the algorithm is learning and improving the performance.

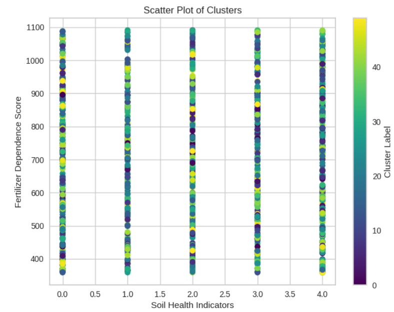
Figure 4.21b distribution of mean test scores is a normal distribution histogram. It suggests that almost all the tests have a mean score between 0.190 and 0.200. And that means most of them performed similarly on the testing phase. There are also fewer tests that scored very high or very low on the testing phase.

This Figure 4.22a shows the aftermath of the classification algorithm of HOLOS dataset. From this plot, it can be seen that the yield efficiencies do not correlate the CO2 emissions. The efficiency of N2O breakdown due to presentation can be observed here too. But it is also noticeable that the high precipitation does not correlate with high N2O breakdown for the orchard and seasonal crops.

Figure 4.22b scatter plot of clusters demonstrates the positive correlation between fertilizer dependence scare and number of clusters. With the increment of fertilizer dependence

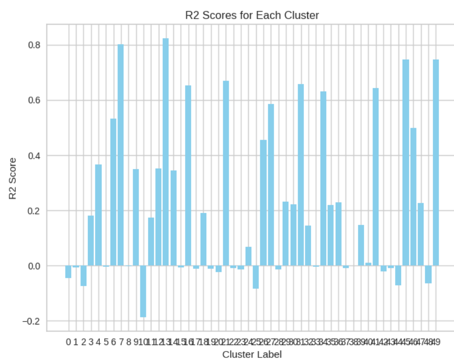


(a) Results After Classification Algorithm.

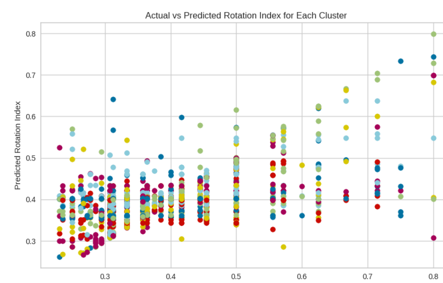


(b) Scatter Plot of Clusters For HOLOS Crop Dataset.

Fig. 4.22: HOLOS-Crop Dataset After Application Of Machine Learning Algorithm.



(a) R2 Scores for Each Cluster For HOLOS Crop Dataset.



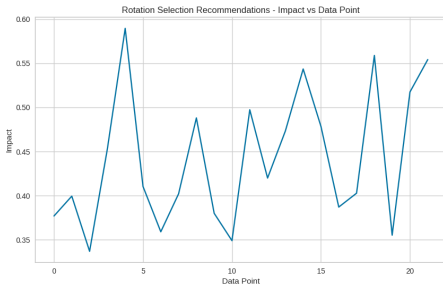
(b) Actual vs Predicted Rotation Index for Each Cluster For HOLOS Crop Dataset.

Fig. 4.23: Clusters Observation of HOLOS-Crop Dataset.

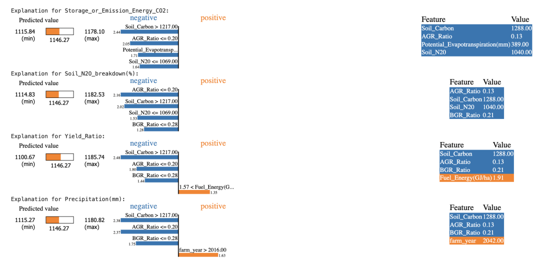
score, the number of clusters increases. This means that the more distinct groups of soil health indicators in the soil that are highly dependent on fertilizer.

Figure 4.23a R2 scores for each cluster shows the fifty cluster records. Some clusters have R2 scare close to 1 which indicates a very good fit for the regression mode. Other clusters have scares close to zero which mean poor fit. In the graph the scores vary throughout the clusters. The negative scores indicate the regression model performance is worse than simply predicting the average values. Overall, the model has fit the data well for some clusters but not for every cluster.

Figure 4.23b actual vs predicted rotation index for each cluster which is a scatter plot.



(a) Rotation Selection Recommendations - Impact vs Data Point For HOLOS Crop Dataset.



(b) LIME Explanation For HOLOS-Crop Dataset.

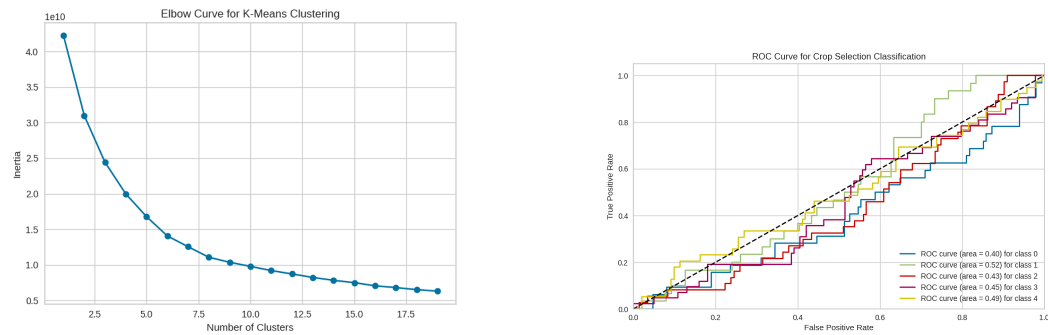
Fig. 4.24: Closer Look At The Impacts And LIME Explanations of HOLOS-Crop Dataset.

We can see that most of the data points are scattered around the diagonal line. But there are some deviations from the line too. This indicates that models' predictions are not perfect. Also, there is a positive correlation between the predicted and actual values. There is a cluster which is in the bottom left corner of the graph. This means that model may under-predict the rotation index for some clusters. The diagonal predicted clusters seem to be more expected in this graph.

Figure 4.24a rotation selection recommendations impact vs data point graph is a line plot indicating the impact of rotation selection recommendations of the data points. The increment of slightly upwards of the graph depicts the impact of the rotation selection recommendation system as the size of the data point increases.

From this Figure 4.24b we can see the LIME explanation for the HOLOS crop based dataset. Here, a strong and positive impact of soil carbon levels can be seen. Also, lower AGR and BGR ratios show negative impacts in the predictors. But, soil carbon has commonly positive influences across different outputs like soil N2O levels.

Figure 4.25a elbow curve for the K-means clustering model shows that number of clusters vs the inertia which is the sum of squared errors (SSE). the number of clusters increases as the inertia gradually decreases. This means that with more clusters, there are more centers to potentially be closer. Here, the overall distance from the data points to their cluster centers tends to decrease. The elbow curve starts to show at around 7 or 8.



(a) Elbow Curve for K-Means Clustering For HOLOS Crop Dataset.

(b) ROC Curve for Crop Selection Classification For HOLOS Crop Dataset.

Fig. 4.25: Final Cluster and Classification Algorithm Results of HOLOS-Crop Dataset.

Figure 4.25b ROC (receiver operating characteristic) curve for crop classification model shows that the models performance is better than random predictions tendencies. The curves seem to not come closer to the top-left side of the graph which is more expected. The area under the curve (AUC) for this model is class 0 is 0.40.

4.2.2 HOLOS Animal Data

The purpose of this part of the analysis is minimizing energy consumption and CO2 emission while maximizing productivity and animal welfare. Then populating more optimized farm scenarios with GAN.

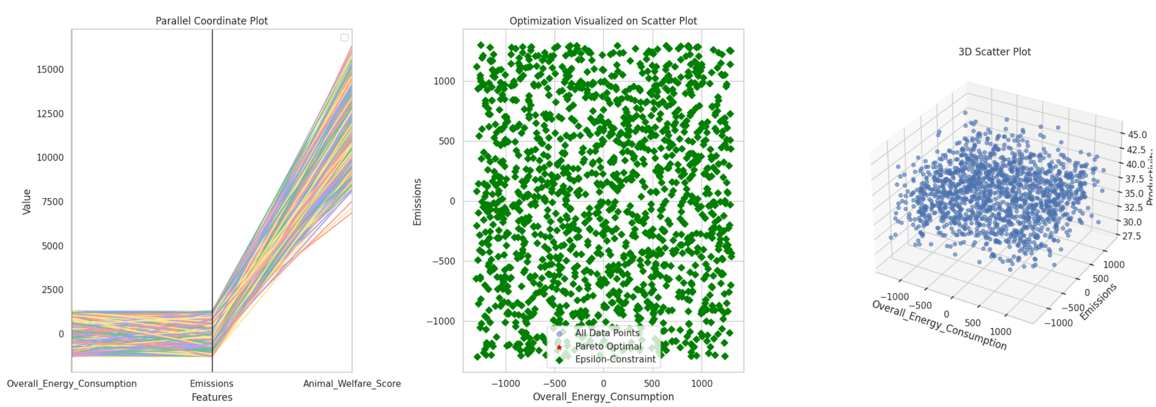
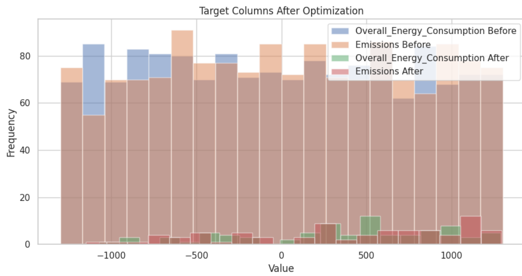


Fig. 4.26: Parallel Coordinate Plot, Scatter Plot and 3D Scatter Plot For HOLOS Animal Dataset.



(a) Target Columns After Optimization For HOLOS Animal Dataset.

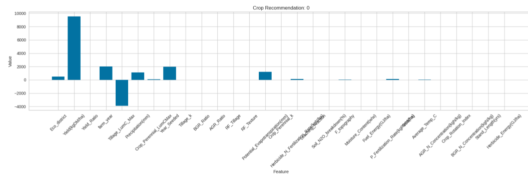
(b) LIME Explanation For HOLOS-Animal Dataset.

Fig. 4.27: Optimization and LIME Explanations of HOLOS-Animal Dataset.

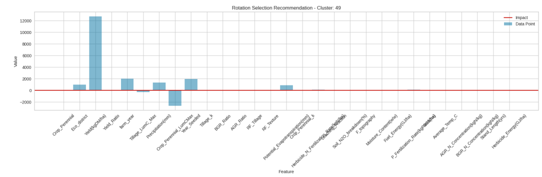
Figure 4.26 parallel coordinate plot shows how each features as a line. Now we can see that the high overall energy consumption and emission have high values. And animal welfare score scores are spread out in the graph. Some of them have really high values and others have low values. The optimization visualized scatter plot demonstrates more details of the parallel coordinate plot. The data points are more distributed around the upper right corner. These are the overall energy consumption and emission values. The size of the circle in this area carries the variations in animal welfare score. Figure 4.26 3D scatter plot shows the similar results in different visualization formats. From these plotted graphs we can see that the epsilon constraint model fits better for the desired objective of this goal.

Figure 4.27a manifests the changes in target column for optimization that strengthens the rotation selection for improving the efficiency. In this figure, we can see the target columns after optimization phase and how the optimized states look like. The optimized states of overall energy consumption and emission features are in one graph showing how the optimization worked. The rotational selection significantly reduces the overall energy consumption and emission while increasing the productivity and animal welfare score.

In the Figure 4.27b the LIME explanations can be seen for the HOLOS animal based dataset. There is a strong connection of CO2 levels in the predicted outcomes. Also, the high methane levels have a positive impact on welfare scores. Besides, it is clear that the optimization of feed for temperature conditions improves productivity and overall energy efficiency.

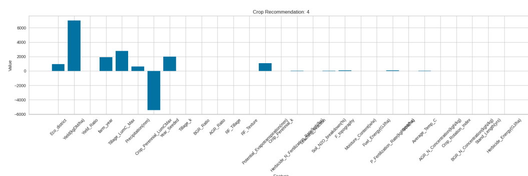


(a) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation.

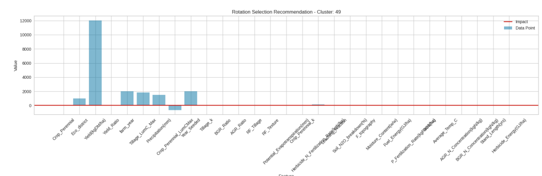


(b) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

Fig. 4.28: Some Optimized Farm Scenarios 001.



(a) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation.



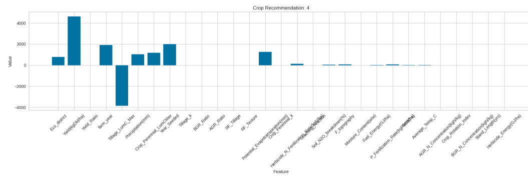
(b) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

Fig. 4.29: Some Optimized Farm Scenarios 002.

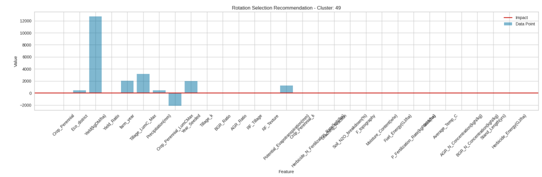
From this observations, it can be said that the maintenance of soil nutrient levels leads to less emission, the improvement of soil health is a long term productivity practices, less ammonium nitrogen leads to less emission, overuse of nitrogen (fertilizers) leads to increased emissions, tillage affects higher CO2 emissions, sometimes higher yields might reduce emissions, different irrigation practices are a necessity, the reduction of methane emissions (diet modifications) is needed, additional renewable energy sources usage, average weight of the livestock helps in optimizing feed efficiency and the overall energy use of the farm. Also, using energy efficient machineries or practices and high biomass crops might be beneficial.

Some of the features with optimized plots of other features are included here. Now, after applying the machine learning models some of the optimized crop selection and rotational strategies can be seen in Figure 4.28, Figure 4.29 and Figure 4.30.

Now collecting the new farm scenarios considering both models. With this information, generative AI model is used to populate more farm optimized farm scenarios. The generative adversarial network (GAN) model is used in this case for populating the opti-

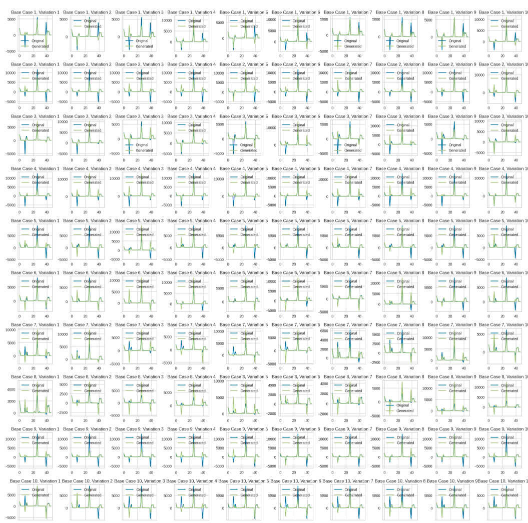


(a) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation.



(b) HOLOS-Whole Farm Crop-Selection and Rotational Recommendation With Impact Indicator.

Fig. 4.30: Some Optimized Farm Scenarios 003.



(a) GAN Generated Farm Scenarios For HOLOS-Whole Farm Crop Dataset.

(b) GAN Generated Farm Scenarios For HOLOS-Whole Farm Animal Dataset.

Fig. 4.31: GAN Generated Some Farm Scenarios.

mized scenarios. In order to generate more optimized synthetic data, the generator neural network architecture is built with two hidden layers. Also, the discriminator of the neural network architecture is built using two hidden layers. Here, the weight of the discriminator is frozen during the training phase of GAN. During the training phase, the generator and discriminator work alternatively.

Next, after applying GAN on the optimized farm scenarios we can see some of the populated farm scenarios in Figure 4.31a and Figure 4.31b for HOLOS crop and animal based farm management practices respectively. Finally, the base cases and variations generated by the GAN are plotted. Figure 4.31b shows the final results.

In the Figure 4.32 the aftermath of GAN application while observing the target columns

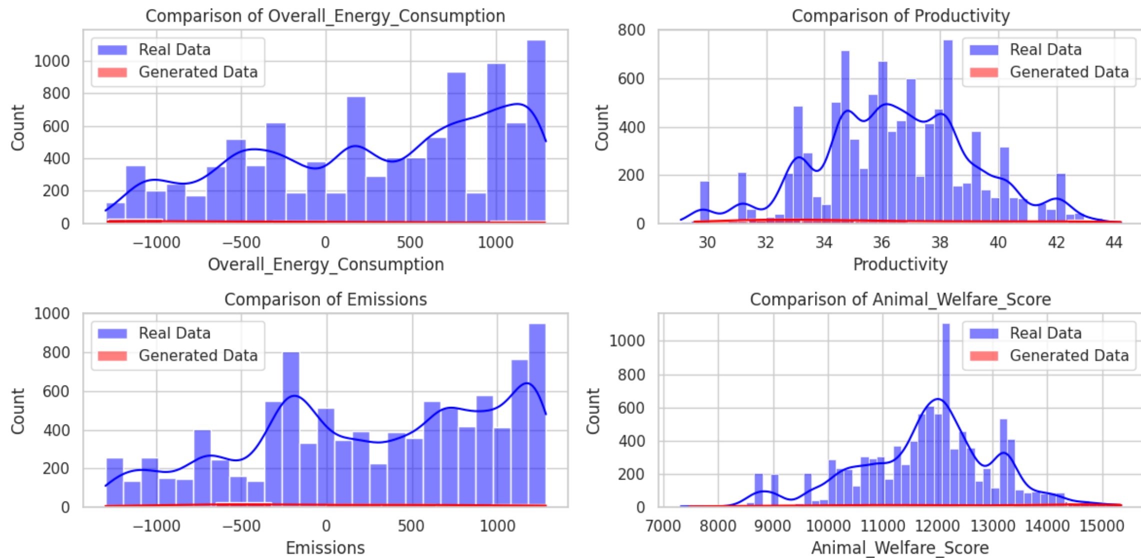
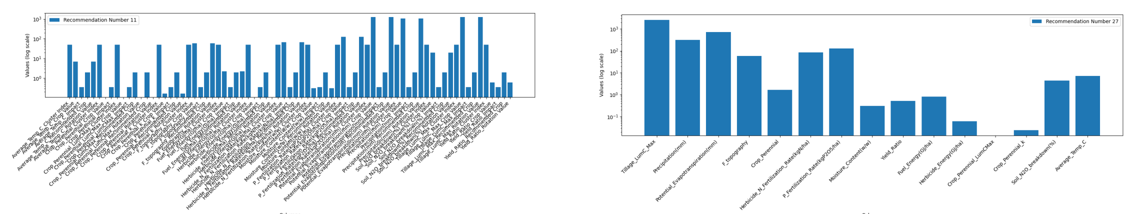


Fig. 4.32: Target Columns After GAN Application For HOLOS Dataset.

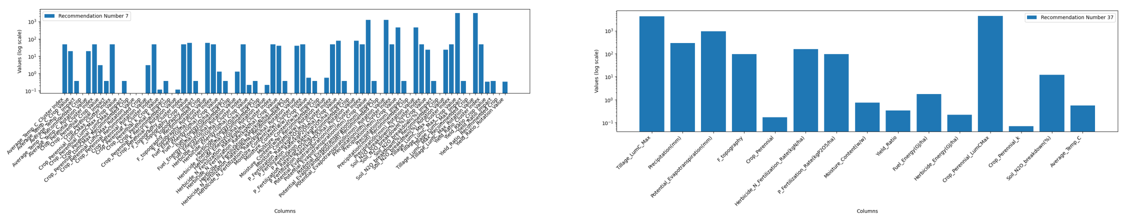


(a) (b)

Fig. 4.33: HOLOS Crop-Optimized And GAN Generated Farm Scenarios 001.

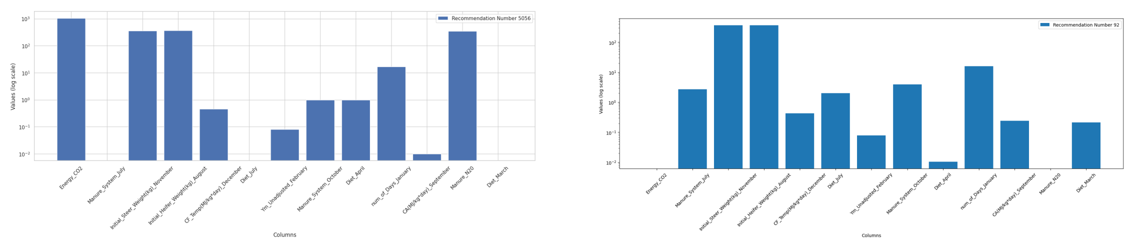
can be seen for the HOLOS dataset. Here, the generative model can be seen to struggle with the extreme values or the outliers. This limitation does not bring much variety in the new populated GAN optimized farm scenarios. So, the presence of mixed match data can be seen in real and generated data.

Finally, some of the HOLOS crop based optimized and GAN generated farm recommendations can be seen in Figure 4.33 and Figure 4.34. Also, HOLOS animal based optimized and GAN generated farm practices are shown in Figure 4.35, Figure 4.36 and Figure 4.37.



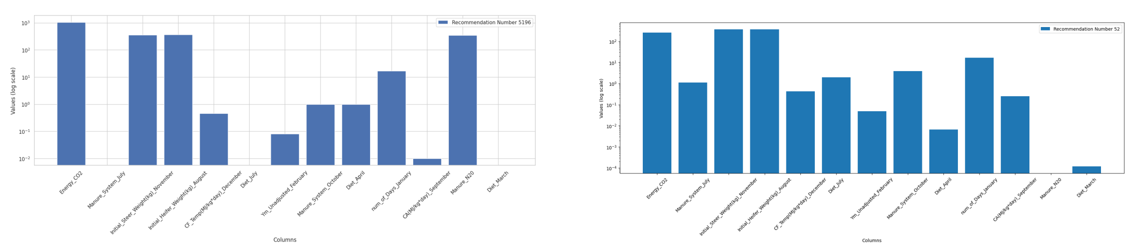
(a) (b)

Fig. 4.34: HOLOS Crop-Optimized And GAN Generated Farm Scenarios 002.



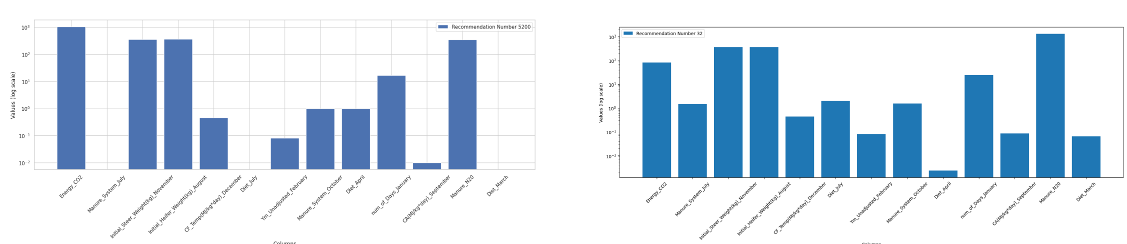
(a) (b)

Fig. 4.35: HOLOS Animal-Optimized And GAN Generated Farm Scenarios 001.



(a) (b)

Fig. 4.36: HOLOS Animal-Optimized And GAN Generated Farm Scenarios 002.



(a) (b)

Fig. 4.37: HOLOS Animal-Optimized And GAN Generated Farm Scenarios 003.

CHAPTER 5

FUTURE WORK AND CONCLUSION

5.1 Future Work

Generated data from COMET-Farm and HOLOS whole-farm can be used in analyzing and finding optimal resolutions. Such as energy efficiency equipment, livestock management, ranching operation management, environmental factors, profitability increment, analyzing and implementation of renewable energy integration in farm scenarios, optimizing irrigation practices, reducing feed waste, waste heat recovery, life cycle analysis, eco-friendly infrastructure based scientific research. The emission data depends on the time and policies. So, optimization needs more careful concentration. One of the most popular optimization models is DQN (deep Q networks) which is a reinforcement learning model. Using DQN and GAN could bring more productive outcomes in the future. Besides, some of the adaptations of new technologies are highly recommended in farming management. For example, introduction of precision agriculture tools, automated irrigation systems or advanced analytics for better decision making practices. Also some changes in farming practices can be beneficial. Also, resource management can be impactful in energy savings.

5.2 Conclusion

This experimental project gives a complete comprehensive understanding of how agriculture based farming management can be achieved with the reduction of GHG emission. The atmospheric GHG concentration is extremely harmful for the future farming and ecosystem. Also, with these sustainable farming management techniques, the production and farming management can be well balanced. It is really complicated to find an optimal state in agricultural based farming management. However, finding a proper balance between different factors can provide more productivity and less GHGs emission rate.

REFERENCES

[1] "Natural Resource Management and Sustainable Agriculture June 2022."

[2] "https://stats.unctad.org/Dgff2016/annexes/glossary.html",(accessed on 03/24/2024)."

[3] "Agricultural Sustainability: A Review of Concepts and Methods September 2019 Sustainability 11(18):5120."

[4] "http://comet-farm.com/, 2024, (Accessed on 03/22/2024)."

[5] "https://orange.biolab.si/, (Accessed on 03/22/2024)."

[6] ""Rapid miner manual," in Rapid Miner Studio manual. Rapid Miner,pp. 19–91. [Online]. Available: https://docs.rapidminer.com/downloads/ RapidMiner-v6-user-manual.pdf."

[7] "Agroecology for a Sustainable Agriculture and Food System: From Local Solutions to Large-Scale Adoption July 2023 Annual Review of Resource Economics 15(1)."

[8] "Integrated farm management for sustainable agriculture: Lessons for knowledge exchange and policyFebruary 2019 Land Use Policy 81:834."

[9] "Novel approaches and practices to sustainable agriculture November 2022. Journal of Agriculture and Food Research 10(2):100446."

[10] ""Greenhouse gasses effect on global warming," https://svs.gsfc.nasa.gov/20114, 2020, (Accessed on 03/22/2024)."

[11] "https://www.epa.gov/ghgemissions/overview-greenhouse-gases, (Accessed on 03/22/2024)."