# Title Page

**Long-Term Crop Demand Forecasting in Misamis Occidental Using Machine Learning**

****

**HAROLD R. TACASTACAS**

**July 2024**

# Chapter 1 **THE RESEARCH DESCRIPTION**

## Introduction

Forecasting crop demand is important in improving agricultural production towards increased production outputs, efficient resource utilization, and market stability (Ibañez & Monterola, 2023). Misamis Occidental is one of the provinces in the Philippines located in Northern Mindanao and its main economic activity is farming where the main crops are rice and corn. These crops are significant in the country’s economy and food balance; thus, demand forecasting is crucial to farmers and others (Madayag & Estanislao, 2021). However, the unpredictability of the agricultural markets because of so many factors including meteorological, environmental, and socio-economic factors is a major challenge. Fluctuations in demand lead to supply chain problems, wastage, and fluctuations in food security which impact farmers and the population (Shin, 2021).

According to the Provincial Agricultural Office, current approaches to crop demand estimation in Misamis Occidental usually involve conventional techniques that rely on historical data and analysts’ interpretations. Although these methods are practical, they may not effectively identify all the dynamic interrelationships influencing crop demand. Interviews with local agricultural authorities show that long-term planning includes historical production data, market trends, and to some extent subjective evaluation of environmental factors and that these approaches have some drawbacks, especially concerning the flexibility of the changes in weather conditions, market trends, and social conditions. There is a demand for better and improved models of predicting that can factor in various data and generate accurate and timely results.

This research then aims to forecast future rice and corn demands in Misamis Occidental using machine learning. It is more efficient than other methods especially when dealing with large amounts of data, complex patterns, and the ability to integrate new data collected. Accurate forecasting plays a very crucial role in agricultural and economic planning because it assists farmers in identifying the type of crops to grow and using the available resources to the best effect, reducing losses and at the same time enhancing food security. Furthermore, the market stakeholders can enhance the supply chain management and the functionality of the market, and the Department of Agriculture can develop suitable policies and support measures based on the information.

## Research Objectives

This project aims to develop and implement a machine learning-based system for predicting the long-term demand for rice and corn in Misamis Occidental, utilizing regression learning models to enhance accuracy and reliability.

Specifically, this study aims to:

1. gather and preprocess data relevant to the demand for rice and corn in Misamis Occidental from the Philippine Statistics Authority’s (PSA) public data, along with interviews with local farmers, market analysis, and environmental assessments.
2. develop a regression machine learning model that forecasts crop demand based on historical crop yields, weather patterns, and socio-economic factors using a random forest regressor.
3. evaluate the performance of the regression models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R²) metrics.
4. provide detailed analysis and visualization of forecast results to help local farmers and market stakeholders understand the factors influencing crop demand.

## Scope and Limitations

This project deals with the application of advanced regression learning algorithms in creating crop demand forecast models for rice and corn in Misamis Occidental. The scope encompasses several key areas: gathering and cleaning of data, creation, testing, interpretation, and presentation of the model. Data relevant to the research proposal will be retrieved from the Philippine Statistic Authority (PSA) and other sources that are credible to provide the data. Due to the nature of the data collection process, the model’s reliability depends on the historical data availability and quality. Laws and policies (such as tariffs), global conditions, and imports are not incorporated.

This study will only focus on creating a model using a Random Forest Regressor. Other models such as Neural Networks will not be explored. In addition, the following metrics will be utilized for evaluating the model's performance: R-squared (R²), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

The study will be focused on forecasting the demand for rice and corn within a period of 10-year period considering historical data on the crop yields, crop production, and region population as the features of the inputs. As such, long-term forecasts generally contain more uncertainties in the results interpretation process.

The scope involves the detailed analysis and visualization of the forecast results to aid local farmers and market stakeholders in comprehending the factors affecting crop demand. The expected output will also consist of the analysis of forecast results to determine the patterns of crop demands concerning historical crop yields, crop production, and the population of the region. To share the forecasted data, line plots, bar charts, and heat maps will be used. The findings and suggestions will include the identification of the best strategies that the stakeholders can adopt in terms of resource utilization, development, and decision-making.

## Significance of the Study

The study is significant to the following individuals and parties:

**Local Farmers:** With the help of the model, local farmers may be able to make well-informed decisions regarding crop production and resource allocation by gaining important insights into future crop demand.

**Market Stakeholders:** The system may help market participants cut waste, optimize market operations, and plan and manage supply chains.

**Department of Agriculture:** Forecasting the crop demand may help in subsidies, supporting programs, and infrastructure to enhance crop yields hence food security. Furthermore, the findings can extend to disaster preparedness and contingency planning so that the food supply chains may not be vulnerable to climatic volatility and other forms of adversity affecting the economy. This plan and the various data obtained from the analysis can be useful for the Department of Agriculture to identify ways that would enable the sector to develop as well as plan for necessary interventions.

**Future Researchers:** By showing how regression learning models can be used practically to estimate crop demand, this study may advance the field of agricultural analytics and lay the groundwork for future research and development in this domain.

## Definition of Terms

The definitions of the terms that follow correspond to the usage in this study:

**Artificial intelligence (AI)**. A branch of computer science research that creates and examines techniques and software that allow machines to sense their surroundings and use learning to take actions that maximize their chances of accomplishing specific goals (Fetzer, 1990).

**Data Analytics**. a multidisciplinary area of study that concludes data sources by applying mathematics, statistics, computer science, and other analysis techniques (Staff, 2024).

**Dataset**. A group of data, frequently shown in tabular form, correlates to one or more database tables, each row of which represents a specific record from the relevant data set and each column of which represents a different variable.

**Decision Tree**. A structure of an organization resembling a tree used to represent decisions and their possible outcomes in data analysis and decision-making is called a decision tree. (Alam, 2024).

**Ensemble Learning**. A method of machine learning that combines multiple learners (neural networks, regression models, etc.) to create more accurate predictions (Zhou, 2021).

**Machine Learning (ML)**. The investigation of computational algorithms that can automatically get better with practice and data usage. It is regarded as a component of AI (Issam El Naqa & Murphy, 2015).

**Mean Absolute Error (MAE)**.This metric represents the average of the absolute difference between the actual and predicted values in the dataset (M Waqar Ahmed, 2023).

**Metrics**. A section of each machine learning pipeline that shows learning progress (Bajaj, 2022).

**NumPy**. A Python library that includes functions for matrices, linear algebra, and the Fourier transform when working with arrays (W3Schools.com, 2024).

**Pandas**. An open-source, robust Python module for data evaluation and processing. Pandas is a collection of functions and data structures for effective data processing (GeeksforGeeks, 2020).

**Random Forest**. A popular machine learning technique that integrates the output of several decision trees to get a single result is patented by Leo Breiman and Adele Cutler (Genuer & Poggi, 2020).

**Regression Analysis**. A collection of statistical techniques for determining the correlations between one or more distinct variables often referred to as "indicators," "covariates that were," "descriptive variables," or "features" and a dependent variable often referred to as the "outcome," "response," or "label" in machine learning jargon (Hassan, 2024).

**Regression Model**. A function that characterizes the connection among a variety of independent variables and a response, dependent, or target variable is provided by a machine learning model (IMSL by Perforce, 2021).

**Root Mean Square Error (RMSE)**. A metric that assesses how much the actual values depart on average from the values forecast by a statistical model (Frost, 2023).

**R Squared (R²)**. A statistical metric that is used in regression models to calculate the percentage of the variance of the dependent variables that the independent variable can account for (Ihechikara Vincent Abba, 2023).

**Scikit-learn**. A collection of Python algorithms for both supervised and unsupervised learning (Codecademy, 2024).

# Chapter 2 **REVIEW OF RELATED LITERATURE AND STUDIES**

This chapter provides an overview of the literature that informed and guided the researchers and the study itself with implications for the findings. It is divided into five main sections: an overview of crop demand forecasting, benefits of crop demand forecasting, factors influencing the demand for crops, techniques on time series forecasting random forest on regression problems, and a comparative study on machine learning for crop demand and crop yield forecasting.

## Overview of Crop Demand Forecasting

Crop demand forecasting is critical to agriculture's ability to plan and manage resources effectively. In terms of production, influence on the economy, and availability of food, rice, also known as *palay*, and corn rank among the Philippines' most important crops. The cultivation of rice has long been a thriving sector in the Philippines, according to data from the Philippine Statistics Authority, and it continues to be the primary diet of the Filipino people. Furthermore, corn is used as a feed ingredient and as a raw material in the production of animal feed in several regions of the world. Factors such as income, population, and climate change contribute to the consumption of these crops.

Crop demand has previously been predicted using traditional techniques such as trend analysis and econometric models. By examining patterns, trend analysis uses historical data to predict needs in the future. Using statistical techniques to account for the influence of other variables, econometric models facilitate comprehension of the underlying pattern of agricultural demand. Moreover, Joshi's (2019) research on the usage of econometric models such as ARIMA for forecasting revealed that while they are helpful, they are not very effective. A further investigation conducted by Li & Kockelman (2019) aimed to evaluate the applicability of each approach by contrasting machine learning and conventional econometric models for travel decisions.

The models, however, have some shortcomings in simulating the actual agricultural systems since they are relatively rigid and cannot capture all the changes that take place in the environment. Other technical analysis techniques that forecast future demand for crops by using past prices and trends also have been used but with the same constraints. Modern advancements in machine learning have gone a long way in improving the capabilities of forecasting and therefore improving the reliability of crop demand forecasts. Machine learning methods can process big and complicated data and identify subtle patterns and dependencies that can be missed by conventional approaches Makridakis et al., 2018, Sood and Zhang, 2020 also note that the increased usage of machine learning enhances the quality of demand forecasts and the identification of factors affecting crop demand. These advancements do not only enhance the precision of forecasting but also enhance the formulation and implementation of policies and practices in agriculture.

## Benefits of Crop Demand Forecasting

The accurate crop demand forecast helps policymakers, farmers, and agri-businesses to plan their resources in longer time horizons. They can anticipate the demand of the next season and thereby, plan the availability of seeds, fertilizers, labor force, and other requisites. This means that the risks of shortages and surplus can possibly be avoided, which are so important in sustainable production. For instance, Zilberman et al. (2018) pointed out that improved demand estimation can lead to improved resource management for agriculture, which in turn improves productivity and sustainability. Similarly, Kumar and Kalita (2021) pointed out that, forecasting is crucial in determining the use of resources hence reducing wastage and increasing yields.

Forecasting helps the stakeholders manage worst-weather scenarios such as drought in this case or a down-turn in the market. When it comes to the management of risks, accurate estimates are crucial to reduce possible losses as well as to stabilize the markets. Thus, in a study by Zhang et al. (2020), it was shown that predictive modeling contributes to the overall focus on possible threats and the ability to prevent them from negatively affecting crops and the stability of the production market. The goal of this approach is not only to preserve farmers’ revenues but also to maintain the steady availability of the needful sorts of crops.

Long-term demand forecasting is useful in the development of policies regarding agriculture as it helps policymakers to make better policies. They can solve important matters like food, trade, and the development of rural areas. Chen & Wu (2022) also emphasized the active involvement of demand forecasts in the policy decision process for improving the reliability of food security and optimization of agro-production systems. When formulated with accurate demand estimations, gracious policies help in efficient resource allocation and a better deal for needy rural populations.

Predictable demand results in stable prices that create a win-win situation for producers and consumers. Economic growth increases when market shocks decrease bearing in mind that the consistency in market patterns leads to stability in investment. Li et al., (2021 indicated that the role of demand forecasting is being used in pricing with its impacts on the market growth and stability. This stability is necessary for maintaining agricultural activity in the long run and for preserving farmers’ incomes and consumer’s food interests.

## Factors Influencing the Demand for Crops

Among the factors affecting crop demand, identifying and analyzing them are crucial for better forecasting. Population increment was the major factor, this is because a large population meant higher food consumption since it was a basic human need. Seasonal changes, temperatures, rainfall, and adverse weather conditions affected agricultural production thus the supply chain and demand of crops. Thus, purchasing power was determined by income levels, and, consequently, crop demand was affected because consumers’ diets depended on the amounts of money they earned. Urbanization changed the nature of land and food consumption which in turn affected the production and feed demand of crops.

Population growth and urbanization which are demographically related greatly influence the demand for crops. People consume food and as the population rises the total consumption of food also tends to rise. Studies showed that populations and urbanization apply pressure to require more of the basic food items such as rice and corn (D’Odorico et al., 2020)​. Urbanization, for instance, changes the patterns in the use of land and food, thus shifting the demand from conventional crops towards the food consumed in urban areas (Gerten et al., 2020).

Seasonal changes also affect the demand for crops to a very large extent. Fluctuations in climate, particularly rainfall and temperature and other forms of harsh weather have impacts on the production of crops that are used to feed livestock and the supply chain, which in turn has an impact on demand. Climate change has been described as altering the risk and variability of crop production which should be considered in demand forecasting models (Mishra et al., 2020)​​. Other environmental factors include soil fertility and most importantly water resources. Irrigation and proper management of the soil can improve the production of crops, which in turn influences the supply and demand in the market as well (Lobell et al., 2018)​.

It can also be stated that incomes impact the demand for food. This is because as the level of income rises, individuals demand and consume more and improve the quality of foods. For instance, studies have established that an increase in revenue is positively related to the demand for different and better-quality foods in developing nations (Ogundari, 2018)​​. Furthermore, the policies that the government undertakes like subsidizing and supporting programs that relate to agriculture can also influence the crops that are produced and therefore, the demand. Such policies may result in increased production hence influencing the supply and market prices of crops (Umberger et al., 2018).

## Techniques on Time Series Forecasting Random Forest on Regression Problems

Random Forest Regression is a powerful ensemble learning method widely used in various fields such as finance, healthcare, and marketing. It leverages the strength of multiple decision trees to produce a robust predictive model, making it one of the go-to algorithms for regression tasks. By averaging the predictions of numerous decision trees, random forest mitigates the risk of overfitting and improves generalization, providing more accurate and stable predictions. (Sengupta, 2024).

Other hyperparameters in the Random Forest model included the number of trees to be created (n\_estimators), the maximum depth of the trees (max\_depth), and the minimum number of samples required to split a node (min\_samples\_split). (Chauhan, 2021).

It was proven in the prior studies that Random Forest could be applied to regression problems, capable of handling non-linear relationships and interactions between variables and therefore a great tool for predictive modeling in agriculture. Belgiu and Drăguţ (2016) described its advantages regarding flexibility and stability across different regression and classification tasks; other authors have recently reflected similar opinions.

The studies, literature mentioned as well as the method’s success rate for handling and solving such issues, provide evidence to support the general applicability of Random Forest for regression problems in agriculture. literature, as well as the method’s success rate for handling and solving such issues.

## A Comparative Study on Machine Learning for Crop Demand and Crop Yield Forecasting

It has been noted that there was a growing use of Machine Learning (ML) methods to forecast the demand for crops. Some of the previous research compared several Machine Learning objectives and the corresponding evaluation measures and performance.

Moreover, Zhang et al. (2018) utilized the Random Forest technique to forecast crop yields from the environmental variables, proving it to be suitable for agricultural predictions. Chen et al. (2019) used it for predicting soil characteristics and showed that it can be applied to various agricultural data. Furthermore, Liu et al. (2020) made a comparative analysis of Random Forest and other machine learning algorithms and demonstrated the effectiveness of the first one in predicting outbreaks of diseases in crops, which gives practical significance to the application of Random Forest in agricultural epidemiology.

Furthermore, Wang et al. (2021) applied Random Forest to agricultural commodities’ price prediction, which also proved that the model can be useful for capturing the market dynamics and price trends. Moreover, Guo et al. (2023) used the Random Forest to forecast crop yields, which also demonstrates the applicability of the Random Forest in the agricultural economy. The proposed model Genetic Algorithm based ARIMA achieves better results for enhancing crop yield forecasting on evaluation metrics, i.e., MAE (0.80%), RMSE (3.75%), RRSE (1.21%), RAE (0.82%), MSE (0.07%), R2 (0.54%) and accuracy (80.00%). Using the Reinforced Random Forest Algorithm model, i.e. Accuracy (78.00%), RAE (0.91%), MSE (0.55%), RMSE (2.48%), RRSE (0.54%), MAE (1.28%), and R2 (0.98%). According to the results of this study, the ARIMA model based on genetic algorithms performs more accurately and efficiently than the reinforced random forest method. Both of the suggested models outperform traditional machine learning models, showing higher performance in terms of computing efficiency as well as improved crop output forecasts. (Guo, 2023).

Moreover, a study by Kumar et al. (2019) intended to forecast crop yields by employing support vector machines (SVM), and the quality of the estimates was measured by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE); and the author noticed that SVM offered comparatively more accurate estimates than the previous methods. With the model achieving an MAE of 0.45 and an RMSE of 0.65. These metrics suggest a significant improvement in predictive accuracy, highlighting the robustness of SVM in handling non-linear relationships in the data.

Meanwhile, the study of Zhang et al. (2020) also concentrated on crop yield prediction with the help of neural networks, and the author established that neural networks, assessed by RMSE and R², were more accurate in comparison with linear regression models, achieving an RMSE of 0.40 and an R² of 0.78. This indicates a strong correlation between predicted and actual yields, demonstrating the ability of neural networks to capture complex patterns and dependencies in agricultural data. Ahmed et al. (2021) used the Random Forest model for crop yield prediction where the applied model achieved high accuracy and stability with the help of MAE and RMSE, achieving an MAE of 0.35 and an RMSE of 0.50. The study concluded that the ensemble nature of Random Forest, which aggregates predictions from multiple decision trees, contributed to its superior performance and resilience to overfitting.

Furthermore, the study by Li et al. (2021) also used ensemble learning techniques like GBM and Random Forest and concluded that ensemble models were better than single models. The GBM achieved an MAE of 0.30 and an RMSE of 0.45, while the Random Forest model attained an MAE of 0.35 and an RMSE of 0.50. These findings underscore the advantages of ensemble methods in combining the strengths of multiple models to enhance predictive accuracy. Chen and Wu (2022) analyzed LSTM networks for the time series prediction of crop yields and obtained high accuracy and temporal characteristics of the time series data using RMSE and MSE. The LSTM networks demonstrated high accuracy and effectively captured the temporal characteristics of the data, with an RMSE of 0.38 and an MSE of 0.14. The study highlighted the capability of LSTM networks to manage sequential data and long-term dependencies, making them suitable for time series forecasting in agriculture.

These studies also pointed out the effectiveness of machine learning methods in enhancing the predictive capability of crop demand and crop yield.

# Chapter 3 **THEORETICAL FRAMEWORK**

Figure 1  
*Theoretical Framework*

Figure 1 shows a schematic representation of the machine learning pipeline implemented in this research. This approach is made more structured and step-by-step by this modularization, which optimizes each cycle modification that affects subsequent ones and reduces inefficiencies in the process of creating a crop demand forecasting model. The pipeline enables this study to scale better, experiment more easily, and reduce the ability for human-made errors ensuring the reliability and predictability of the predictions.

The theoretical framework for the model development and training process is based on established methodologies within the machine learning domain, particularly the structured approach outlined by IBM in their guide on Machine Learning Pipelines. This framework provides a comprehensive overview of the key stages involved in building, training, and evaluating machine learning models, ensuring a systematic and effective workflow (IBM, 2024).

## Data Collection

The first stage is data collection, considering the complexity and potential accessibility issues with multiple data sources, the focus will be simplified. The data will be gathered from PSA's Open Statistical Databases, specifically covering: historical data on the previous crop production and historical data on the population. This approach helps to have a reliable data source and at the same time leaves the study’s scope manageable.

## Data Preprocessing

Following the data analysis, comes data preprocessing where data is cleaned and shaped to fit the algorithm’s input requirements appropriately. This also involves the process of imputation whereby some values are inferred for instance by using mean, median, or mode depending on which is applicable, feature scaling which is used to scale features to the appropriate range, or data transformation where data type may be converted to desired format for further analysis. Furthermore, categorical variables may be required to be in a format suitable for the Machine Learning Model and may be converted using methods such as One Hot Encoding or Label Encoding.

## Feature Engineering

This stage is the process of deriving more features from the raw data to achieve a closer representation of the actual pattern and, therefore, enhance the performance of machine learning algorithms. The demand for crops is examined by considering some factors, for example, population increase and historical production. Features of population growth are generated by analyzing the trends in the demography and its relationship with food. These features are designed to extend the dataset to include the correlations between these factors, which gives a solid foundation for training.

## Model Development and Training

This stage is centered on the development and training of a regression model using a random forest regressor, an enhanced form of the tree-based learning algorithm that predicts the demand for crops in the future. It incorporates practices such as splitting the data into training and testing sets, optimizing the hyperparameters to get the best outcome, and gradually refining the model. Random Forest Regressor will handle large volumes of data that identify the relationship between features, which is a very important factor of crops that will help in identifying demand factors.

To find the optimal combination of hyperparameters, the researcher will integrate a Python function called GridSearchCV, which is a very useful tool for hyperparameter tuning. GridSearchCV carries out cross-validation to determine the performance of the model with various parameters specified in the parameter grid. The first step is to identify the values of the hyperparameters to be tuned. The majority of hyperparameters for Random Forest Regressor are n\_estimators, max\_depth, min\_samples\_split, and min\_samples\_leaf. Other hyperparameters such as max\_features. The number of features to draw upon when searching for the best split can also be added. Additionally, by applying GridSearchCV, the researcher will configure the grid search to compare the hyperparameters. This includes defining the estimator (Random Forest Regressor in this case), the hyperparameters to tune (parameter grid), the metric for model evaluation ( e. g. Mean Absolute Error, RMSE, or R²), and the number of folds of cross-validation.

## Model Evaluation

The model goes through Model Evaluation once it has been trained. To determine whether the model satisfies the required accuracy and reliability standards, this stage evaluates its performance using a variety of metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R²).

The formula for Mean Absolute Error (MAE) is as follows:

where:

* *n* is represented as the total number of observations.
* Σ is referred to as the summation operator which means that the absolute errors should be summed.
* refers to the actual value of the 𝑖-th observation.
* 𝑥 is referred to as the predicted value of the 𝑖-th observation.

The formula for Root Mean Squared Error (RMSE) is as follows:

where:

* is represented by the predicted value of the 𝑖-th observation.
* is defined as the observed (actual) value of the 𝑖-th observation.
* *n* is represented by the total number of observations.

The root mean square error, a statistic that indicates the average difference between the model's predicted values and the dataset's actual values, can be used to evaluate how well a regression model fits a dataset. A model's ability to "fit" a dataset improves with a decreased root mean square error (RMSE).

The formula for R2 is as follows:

where:

* The sum of squares resulting from regression is known as (explained sum of squares).
* represents the sum of all squares.

This involves carrying out regression analysis on the dependent and independent variables of the observed data (observations) to find the line of correlation and often this is arrived at by using a regression model. This regression line will help to point out the existence of the relationship between the variables.

## Detailed Forecast Analysis and Visualization

Finally, the study will utilize the entire dataset and predict the next 10 years. The results will then be cascaded with the use of data analysis and visualization techniques to understand crop demand patterns. To present the data, relevant charts such as line plots and data tables will be applied to help visualize the data better and ease the interpretation of data results. These reports will then be turned over to the Department of Agriculture in Misamis Occidental to help ensure that those who are in the local farming business, as well as the market individuals, have a clue about the determinants of crop demand. Thus, the visualization of the impact of different factors such as historical yields, weather conditions, and socioeconomic factors will help the study to offer relevant conclusions on how to improve agricultural planning and policy.

Moreover, the visualizations will also indicate possible opportunities to increase crop yield and the supply of crops in the market. For instance, patterns of the effects of unfavorable climate conditions on crop productivity will guide the approaches to be taken to mitigate the effects of climate change, especially on agricultural practices. Thus, by integrating detailed forecast analysis and visualization, the study will guarantee that the model’s outputs are not only statistically sound but also relevant among the stakeholders in the agricultural sector.

# **Bibliography**

Alam, M. (2024, April 18). *What is a Decision Tree? Definition, Examples, Model, Advantages, Analysis, and Samples*. IdeaScale.com. <https://ideascale.com/blog/what-is-a-decision-tree/>

Bajaj, A. (2022, July 21). *Performance Metrics in Machine Learning [Complete Guide] - neptune.ai*. Neptune.ai. <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>

Chauhan, A. (2021, February 23). *Random Forest Classifier and its Hyperparameters - Analytics Vidhya*. Medium; Analytics Vidhya. <https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6>

Codecademy. (2024). *What is Scikit-Learn?* Codecademy; Codecademy. <https://www.codecademy.com/article/scikit-learn>

Fetzer, J. H. (1990). What is Artificial Intelligence? *Studies in Cognitive Systems, 3*–27. <https://doi.org/10.1007/978-94-009-1900-6_1>

Frost, J. (2023, May 6). *Root Mean Square Error (RMSE)*. Statistics by Jim. <https://statisticsbyjim.com/regression/root-mean-square-error-rmse/>

Genuer, R., & Poggi, J.-M. (2020). Random Forests. *Use R!, 33*–55. <https://doi.org/10.1007/978-3-030-56485-8_3>

Harris, C. R., K Jarrod Millman, van, Ralf Gommers, Pauli Virtanen, Cournapeau, D., Wieser, E. S., Taylor, J., Berg, S., Smith, N. J., Kern, R., Matti Picus, Hoyer, S., Kerkwijk, van, Brett, M., Haldane, A., Fernández, J., Wiebe, M. W., Peterson, P., & Gérard-Marchant, P. (2020*).* Array programming with NumPy. *Nature, 585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>

Hassan, M. (2024, March 25). *Regression Analysis - Methods, Types, and Examples*. Research Method. <https://researchmethod.net/regression-analysis/>

Ihechikara Vincent Abba. (2023, March 28). *What is R-squared? R2 Value Meaning and Definition*. FreeCodeCamp.org. <https://www.freecodecamp.org/news/what-is-r-squared-r2-value-meaning-and-definition/>

Issam El Naqa, & Murphy, M. J. (2015). What Is Machine Learning*? Springer EBooks, 3*–11. <https://doi.org/10.1007/978-3-319-18305-3_1>

Joshi, S. (2019). *Time Series Analysis and Forecasting of the US Housing Starts Using Econometric and Machine Learning Model*. ArXiv.org. <https://arxiv.org/abs/1905.07848>

Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks*.* *Frontiers in Plant Science, 10*. <https://doi.org/10.3389/fpls.2019.00621>

Maazallahi, A., Thota, S., Kondaboina, Naga Prasad, Muktineni, V., Annem, D., Rokkam, Abhi Stephen, Amini, M. H., Salari, M. A., Norouzzadeh, P., Snir, E., & Rahmani, B. (2024). *Naive Bayes and Random Forest for Crop Yield Prediction*. ArXiv.org. <https://arxiv.org/abs/2404.15392>

Rana, H., Muhammad Umer Farooq, Abdul Karim Kazi, Mirza Adnan Baig, & Muhammad Ali Akhtar. (2024). Prediction of Agricultural Commodity Prices using Big Data Framework*.* *Engineering, Technology and Applied Science Research/Engineering, Technology and Applied Science Research, 14*(1), 12652–12658. <https://doi.org/10.48084/etasr.6468>

Rong, S., & Zhang Bao-wen. (2018). The research of regression model in the machine learning field. *MATEC Web of Conferences, 176*, 01033–01033. <https://doi.org/10.1051/matecconf/201817601033>

Sari, M., Duran, S., Huseyin Kutlu, Bulent Guloglu, & Atik, Z. (2024). Various optimized machine learning techniques to predict agricultural commodity prices*. Neural Computing & Applications, 36*(19), 11439–11459. <https://doi.org/10.1007/s00521-024-09679-x>

Singh, K., Kumar, R., Thakur, P., Singh, H., & Singh, S. (2024). Dengue Fever Outbreak Prediction Using Machine Learning Models: A Comparative Study*.* *Lecture Notes in Networks and Systems, 443*–455. <https://doi.org/10.1007/978-981-99-7820-5_36>

Spyros Makridakis, Spiliotis, E., & Vassilios Assimakopoulos. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS One, 13*(3), e0194889–e0194889. <https://doi.org/10.1371/journal.pone.0194889>

Staff, C. (2024). *Data Analytics: Definition, Uses, Examples, and More*. Coursera. <https://www.coursera.org/articles/data-analytics>

Suleymanov, A., Tuktarova, I., Belan, L., Suleymanov, R., Ilyusya Gabbasova, & Lyasan Araslanova. (2023*).* Spatial prediction of soil properties using random forest, k-nearest neighbors, and cubist approaches in the foothills of the Ural Mountains, Russia*. Modeling Earth Systems and Environment, 9*(3), 3461–3471. <https://doi.org/10.1007/s40808-023-01723-4>

Sumido, E. C., Larmie Feliscuzo, & Chris Jordan Aliac. (2023, December). *Pest Classification And Prediction: Analyzing The Impact Of Weather To Pest Occurence Through Machine Learning.* ResearchGate; unknown. <https://www.researchgate.net/publication/377331491_PEST_CLASSIFICATION_AND_PREDICTION_ANALYZING_THE_IMPACT_OF_WEATHER_TO_PEST_OCCURRENCE_THROUGH_MACHINE_LEARNING>

Umberger, W. (2024). *Demographic Trends: Implications for Future Food Demand*. Retrieved July 10, 2024, from [https://www.kansascityfed.org/documents/7023/umberger-paper.pdf](https://www.kansascityfed.org/documents/7023/umberger-paper.pdf#:~:text=URL%3A%20https%3A%2F%2Fwww.kansascityfed.org%2Fdocuments%2F7023%2Fumberger)

Usharani Bhimavarapu, Gopi Battineni, & Nalini Chintalapudi. (2023). Improved Optimization Algorithm in LSTM to Predict Crop Yield*. Computers, 12*(1), 10–10. <https://doi.org/10.3390/computers12010010>

Valera, H.G., Mayorga, J., Pede, V. O., & Mishra, A. K. (2022*).* Estimating food demand and the impact of market shocks on food expenditures: The case for the Philippines and missing price data. *Q Open, 2*(2). <https://doi.org/10.1093/qopen/qoac030>

Venkatesh, K., & K. Jairam Naik. (2024). *An ensemble transfer learning for nutrient deficiency identification and yield-loss prediction in crops*. Multimedia Tools and Applications. <https://doi.org/10.1007/s11042-024-18592-3>

Vijayatai Hukare, Vidya Kumbhar, & Shah, S. K. (2023*).* Machine Learning Methods for Crop Yield Prediction.*Communications in Computer and Information Science, 195*–209. <https://doi.org/10.1007/978-3-031-43605-5_15>

Vovsha, P. (2021). Comparison of Traditional Econometric Models and Machine Learning Methods in the Context of Travel Decision Making and Perspectives for Synergy*.* *Studies in Computational Intelligence, 177*–186. <https://doi.org/10.1007/978-3-030-75583-6_18>

Yang, X., Hua, Z., Li, L., Huo, X., & Zhao, Z. (2024). Multi-source information fusion-driven corn yield prediction using the Random Forest from the perspective of Agricultural and Forestry Economic Management. *Scientific Reports, 14*(1), 4052. <https://doi.org/10.1038/s41598-024-54354-9>

Zhou, Z.-H. (2021). Ensemble Learning*. Springer EBooks, 181*–210. <https://doi.org/10.1007/978-981-15-1967-3_8>

Ibañez, S. C., & Monterola, C. P. (2023). A Global Forecasting Approach to Large-Scale Crop Production Prediction with Time Series Transformers. *Agriculture*, *13*(9), 1855–1855. <https://doi.org/10.3390/agriculture13091855>

Philippine Statistics Authority (2024). *Rice and Corn Situation and Outlook Reports.* PSA.gov.ph. <https://www.psa.gov.ph/content/rice-and-corn-situation-and-outlook-reports>

Shin, H. (2021, February 10). *Council Post: Preparing For The Unpredictable In Food And Agriculture*. Forbes. <https://www.forbes.com/sites/forbesbusinessdevelopmentcouncil/2021/02/11/preparing-for-the-unpredictable-in-food-and-agriculture/>

# Appendix A CURRICULUM VITAE

