Elevating Efficiency and Sustainability in Large-Scale Coconut Oil Manufacturing through Progressive Strategies

(COCOCLARITY MOBILE APP)

R24-059

Project Final Report

Dewmini A.M

It21020308

B.Sc. (Hons) Degree in Information Technology

(Specialization Information Technology)

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

Implementing Coconut Oil Yield Prediction System

R24-059

Individual Final Report

Dewmini A.M

IT21020308

Supervisor: **Mr. Nelum Chathuranga Amarasena**

Co – Supervisor: **Mrs. Manori Gamage**

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	**Student ID**	**Signature**
Dewmini A.M.	IT21020308	

Signature of the Supervisor
(Mr. Nelum Chathuranga Amarasena)

……………………………………

22/08/2024

ABSTRACT

The work is devoted to one of the crucial issues of the coconut oil industry, namely using sophisticated forecasting models for enhancing the company's performance evaluation. The research focuses on two main areas: calculation of the probable output of virgin coconut oil from the number of coconuts as well as the quantity of by products such as coconut water, shell, oil cake and kurutu.

The first model involves the implementation of a unique Machine Learning random forest algorithm used not only for the estimation of the required number of coconuts but also for the determination of the by-products of oil production. This model appears to be of much significance for farmers and producers since they shall get useful information with which they can arrange their operations for the best. The ability to manage the risks and variability associated with the overall number of coconuts that may be called for and the extra value that may be got from the by products enable control on wastage and use of resources hence making it possible to maximize profitability.

The second model uses the Decision Tree algorithm in the estimation of the yield of oil that can be obtained from kurutu which is a by-product of coconut oil. This model is very useful in the extraction of the most value form every part of the coconut right form the by-products. The model offers useful expectation on the prospect of yield on kurutu and is therefore able to inform production system in cases of minimization of wastage as well as enhancing the efficiency of the production line.

The arrival of these predictive models is a good thing in the coconut oil industry as it outlines how the resources can be pushed to give rise to costs, improved production timetable and quality. Consequently, this research work is seen as the hub for more refinements in the usage of machine learning for enhancement of sustainability agriculture and efficiency of the same in the future.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Description

AI	Artificial Intelligence
ML	Machine Learning
ANN	Artificial Neural Networks
RNN	Recurrent Neural Networks
DL	Deep Learning
DS	Data Science
AL	Active Learning
FE	Feature Engineering
DT	Decision Trees Network

1.INTRODUCTION

1.1 Background & Literature survey

The thriving coconut industry in Sri Lanka plays a pivotal role in the livelihoods of over 1 million small-scale farmers, as these individuals are heavily reliant on it. Across expansive plantations that cover an area of 395,797 hectares, more than three billion coconuts are yielded annually for production. In fact, copra manufacturing hit an impressive high point in 2020 at approximately 552153 tons (MT). The positive impact this sector has on agriculture and overall economy is significant within Sri Lanka's community.

Sri Lanka's economy greatly relies on coconuts, not only as a source of livelihood but also for generating export revenues. Due to its longstanding association with coconut agriculture, Sri Lanka has emerged as a major player in the global coconut market.

Using machine learning to forecast coconut oil yields can boost productivity, reduce waste, and promote sustainability in Sri Lanka's coconut oil industry. An advanced prediction model incorporating weather, soil conditions, and historical yields could provide reliable quarterly projections.

With these forecasts, farmers and producers can optimize production to better balance supply and demand while making wise investments to meet rising demand.

Figure 1 showcases expected coconut oil production patterns by 2023. It highlights major factories and influencers generating output from substantial copra levels. This underscores the need to adopt tactics ensuring long-term, eco-friendly industry growth.

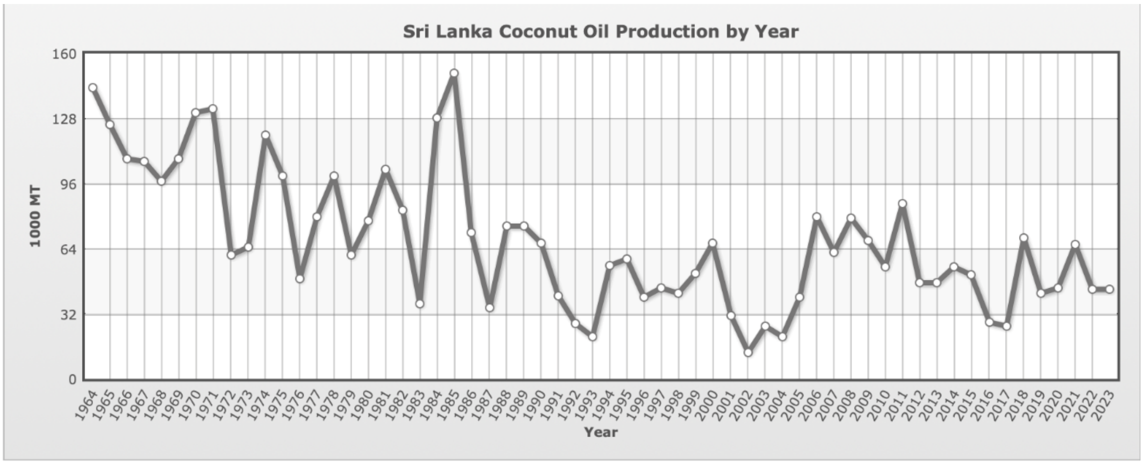


Figure 1: Coconut oil production 1964-2023(MT)

Sri Lanka combines modern technologies and sustainable solutions with traditional methods to advance its coconut industry and ensure public welfare. The goal is to boost productivity and profitability for all by unifying innovation with environmental protection and cultural heritage.

Sri Lanka's commitment to progress, eco-friendliness, and economic growth is evident in efforts to develop efficiency and longevity in coconut oil production through teamwork. By establishing a sustainable future, the flourishing coconut trees will provide nourishment and opportunities for generations.

The industry is supported by over one million smallholder farmers managing 395,797 hectares of land. Their collective three billion annual nuts drive copra output to meet global coconut oil export demand. In 2020, Sri Lanka produced 233,660 metric tons of premium quality coconut oil, ranking fourth among top producers worldwide. This contributed LKR160 billion to GDP.

Despite the sector's economic importance, sustainable practices are urgently needed to enable continued growth and profitability while preserving precious natural resources and commercial viability.

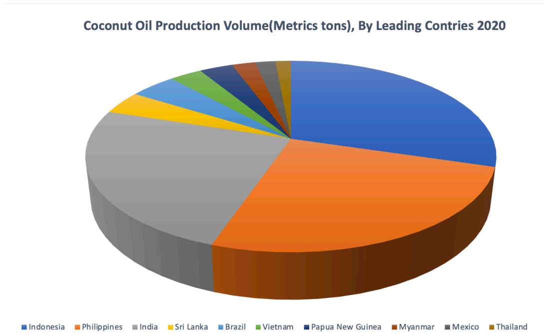


Figure 2 : Top 10 coconut oil producing countries 2020.

Sri Lanka's average 5-6% coconut oil extraction rate from copra lags behind India's 7%+ rates. This shows considerable room for improvement. However, multiple obstacles impede yield enhancement like inconsistent nut maturity checks, inadequate copra drying and storage, inefficient machinery, and process wastage. Upgrading practices like standardized drying protocols, better equipment, improved storage to cut wastage require prompt implementation to heighten productivity and output.

Environmental factors significantly impact copra quality and oil yields. For 30 years, declining rainfall due to climate change has challenged consistent copra quality and yields in Sri Lanka. However, data is limited on how specific factors affect supply chain productivity. [11]Enhancing collection methods and analyzing environmental impacts could improve coconut oil production processes while managing yield changes.

Statistics-based forecasting has proven ineffective at modeling complex interactions influencing oil yields. Conversely, machine learning techniques offer superior predictive capabilities over traditional approaches. But studies on advanced ML tailored to project Sri Lanka's coconut oil yields based on historical data and supply chain specifics remain lacking. Implementing cutting-edge ML and predictive analytics considering external factors like weather could profoundly transform industry practices for higher sustainability and profitability.

Including copra quality (figure 3), drying techniques, rainfall patterns, and milling metrics can boost prediction accuracy, as proven for palm oil using ML with geospatial data.[12] Collectively analyzing these factors can develop a reliable coconut oil yield forecasting system. By linking individual coconut profiles with past crop yields and external factors like soil and weather, this approach enables efficient yield estimates based on available resources. As feasibility for such advancements increases, so do Sri Lankan growth opportunities around coconut oil through achieving sustainability and profitability.

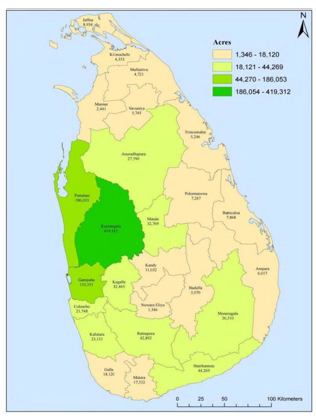


Figure 3: Coconut cultivation areas in Sri Lanka

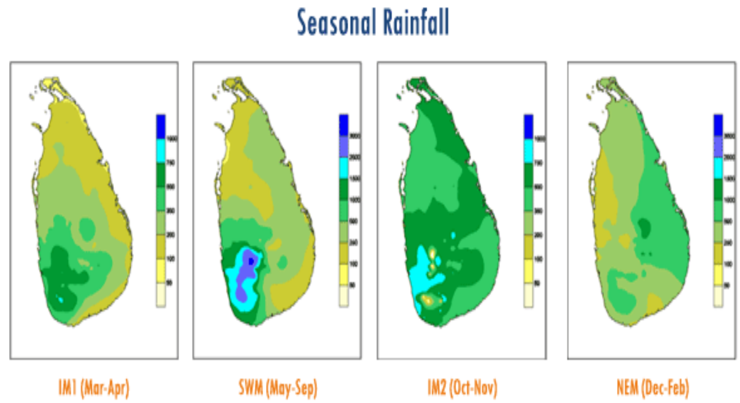


Figure 4: Seasonal rainfall patterns

Effectively implementing these measures can positively impact the environment by promoting efficient energy use, effective waste disposal, and reduced emissions [13]. Sustainability not only benefits local ecology but also ensures long-term economic prosperity for residents.

The coconut oil sustainability ensures both ecological and economic viability for Sri Lanka through energy and resource conservation coupled with responsible waste and pollution management. This allows for continued growth of this vital industry.

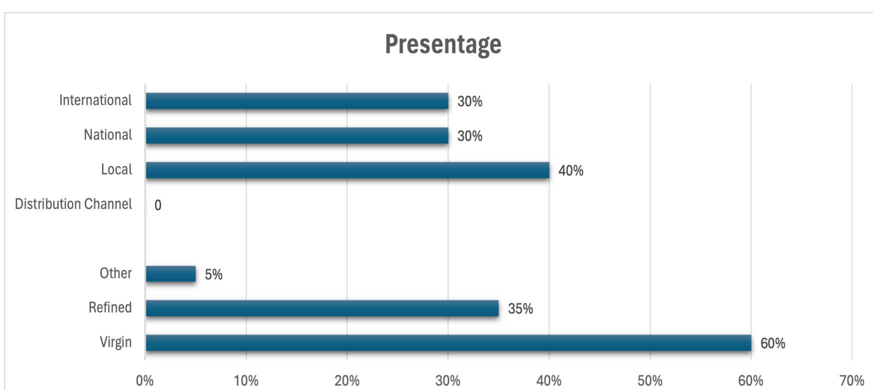


Figure 5 : Coconut oil production flow diagram

This study aims to develop a custom forecasting system for Sri Lanka's coconut oil industry using copra supply and oil mill production data. It seeks to improve on traditional statistical techniques by applying advanced analytics, optimization strategies, and sustainable solutions.

By combining historical records with multivariate analysis, this innovative approach can enhance the accuracy of coconut oil output projections. It also enables integrating efficient, eco-friendly methods into industry operations.

Application of machine learning models in handling problems in the production of coconut oil is innovation in this case. Hence, the purpose of this study is to use Random Forest and Decision Tree model to get the forecast of yield of coconut oil with enhanced 10%. Even more specifically, Random Forest algorithm is used to arrive at the requisite from the number of coconuts of virgin coconut oil. It makes estimation possible; it helps producers at each phase of their production tour.

However Random Forest measures not only the first output of coconut but the measure other derivative products of the coconut productions including coconut water, coconut shell, coconut oil cake and kurutu. This capability is particularly crucial to the producers as it offer complete information on all the production procedures. Since any improvement in the possible returns that can be realised from these by-products the producers are in the best position to chart the future of these by-products. It also assists in the management of wastes though it offers other potential extra sources of income from the sale or use of such wastes by-products. But in the end, this has such advantages of allowing farmers and producers to have broad outlook of the entire process of the production of coconut oil so as to improve on the success on the reduction of wastage and efficiency on the use of inputs in getting coconut oil.

The Decision Tree model concentrates on the by-product kurutu with a view to estimating the amount of oil available for extraction. This is especially useful in the optimization of the resource use and make certain that all the waste from the coconut oil production is put to good use. Figure 2 also presents the relative amounts of all the by-products obtainable from 1000 coconuts apart from the oil.

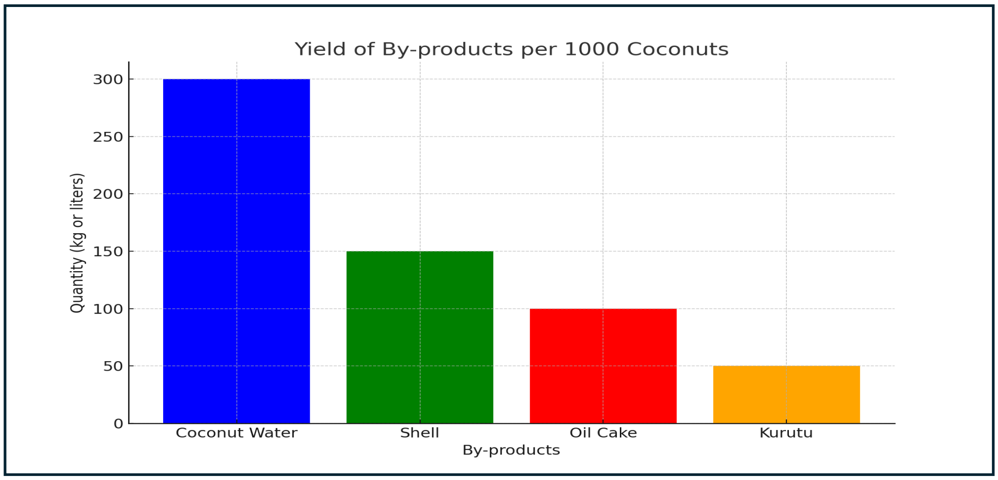


Figure 6 : Yield of By-products per 1000 Coconuts

The incorporation of these models into the production process contributes to increasing the accuracy of resources' prediction, as well as the efficient use of all the available resources. Drawing of these machine learning models.

When appreciating the background to this study, it is possible to get a grip on the fact that there is need to improve existing prediction models in the coconut oil production. Earlier strategies used in forecasting yields and the necessary inputs in this sub-sector have not been this effective hence leading to overproduction and under production. This shall lead to a lot of wastages in so far as there is so much excess production of the product and on the other hand, if the market does not respond to the product, then… this will lead to the shutdown of the industry.

Technology of using the machine learning models in production of coconut oil can be viewed as the new furthering of the use of complex and precise predicting technology. These are Random Forest and Decision Tree models which permit large volumes of past and current data to produce better estimates about coconut demand and potential of by-products. Use of this method helps the producers in availing the above chances of making the best out of resources without wastage and other cost to do with production.

Furthermore, it enhances the possibility to develop the sustainable coconut oil production, as highlighted in this paper. It means that when the producers predict all the resources that are required and the quantities of the production, in the correct perspective, then the resources are utilized effectively in, hence reducing the pollutions. This is not a question of ‘are there technological synergies?’, but rather what should be pursued for the betterment of the future industry of sustainable agriculture.

1.2 Literature Review

Fresh chances to boost crop yield predictions are emerging. Farm tools that use Random Forests and Decision Trees look good. These tools can handle big data sets and grasp tricky links between factors. They can crunch tons of info like past harvests, weather, and live data, to give more exact and trustworthy forecasts.

These smart tools have an impact on coconut oil making. Research shows they can figure out how many coconuts you need for a certain amount of oil. They also guess what extra stuff you'll get when making the oil. This gives a full picture of the whole process, which beats old methods by a mile.

What's more, researchers stress how important it is to combine these smart prediction models with ways to manage resources. Good resource management guided by spot-on forecasts, can help producers boost operations, cut down on waste, and meet what the market wants. This mix is key in the coconut oil business where changes in oil yield can affect how much money is made.

In the end, studies show that old-school methods have had an influence on coconut oil production. But now, bringing in machine learning models offers a smarter and more precise way to predict yield and manage resources. This modification of procedures is a significant step forward for the industry. It affects output and how environmentally friendly production processes are, and it helps producers deal with recurring problems.

The study of coconut oil production, with a focus on resource management and yield prediction, exposes the major issues and intricate subtleties that have always been present in this area. The production of coconut oil has always depended on sense and learnt techniques, just like many other farming practices. While these techniques have advantages, they are frequently unable to manage the fluctuations in crop yields that occur in farming. We observe these variations in the production of coconut oil, where factors such as methods of cultivation can significantly alter the yield of oil. The use of old techniques, such as tried-and-true procedures and wisdom passed down through the years, has resulted in a production process that is frequently unpredictable and wasteful.

Statistical models have been the focus of recent research to improve farm produce projections. For crop output prediction, people have resorted to classical statistical techniques like regression analysis. These methods analyse historical data and quantifiable variables. They look for trends and connections in the data to try and predict future yields. However, these models have an important problem. They are unable to comprehend all the complex non-linear relationships that exist between the various variables which impact yield. As a result, estimates that are made are frequently off from actual yields. This is also true to produce coconut oil, which ranges.

Machine learning has had an impact on yield prediction. Old-school statistical models can't compete with what machine learning algorithms like Random Forests and Decision Trees can do. These new tools work well with large messy datasets. They can process huge amounts of information, including past yields and various environmental factors as well as real-time data. This allows them to identify complex patterns and relationships that traditional models might miss. As a result, they provide more reliable and accurate yield forecasts. Machine learning models have a bunch of obvious perks when it comes to making coconut oil. Studies show these models can work out how many coconuts you need to get a specific amount of oil. They can also guess what other stuff might pop up during the process. This overall picture of how things are made helps producers run their operations more. For instance, understanding how different ingredients an impact on what extra products have been made can help producers make clever decisions about where to put their resources and how to handle waste.

What's more, combining machine learning models with resource management techniques has an impact on the industry that's getting bigger. These models give precise predictions that allow producers to handle their resources better, cut down on waste, and react faster when market demand shifts. This matters a lot in the coconut oil business where changes in yield can affect how much money is made. Using resources in the best way possible based on spot-on forecasts can help operations become more stable and profitable.

To wrap up old-school methods have been key in making coconut oil, but bringing in machine learning models marks a big change. It shows a move towards smarter more exact ways to guess crop yields and handle resources. This shift isn't just a big step forward for the industry; it's a must. It gives producers the tools they need to tackle the tricky parts of today's farming and to boost their output while being kinder to the environment. Research shows that using these high-tech models can help the coconut oil business get past long-standing problems. In the end, this could lead to a more steady and profitable way of making coconut oil.

1.3 Research Gap

The research gap in predicting agricultural yields for coconut oil production, is big and urgent. Farmers have long relied on traditional methods to predict yields, which form the basis for planning and making decisions in agriculture. But these methods, which often use statistical models like regression analysis, don't work well for the unique challenges of coconut oil production. This shortcoming has its roots in several factors. These include the way coconut yields can vary a lot how environmental factors interact in complex ways, and the need to accurately predict both main yields (coconut oil) and extra by-products.

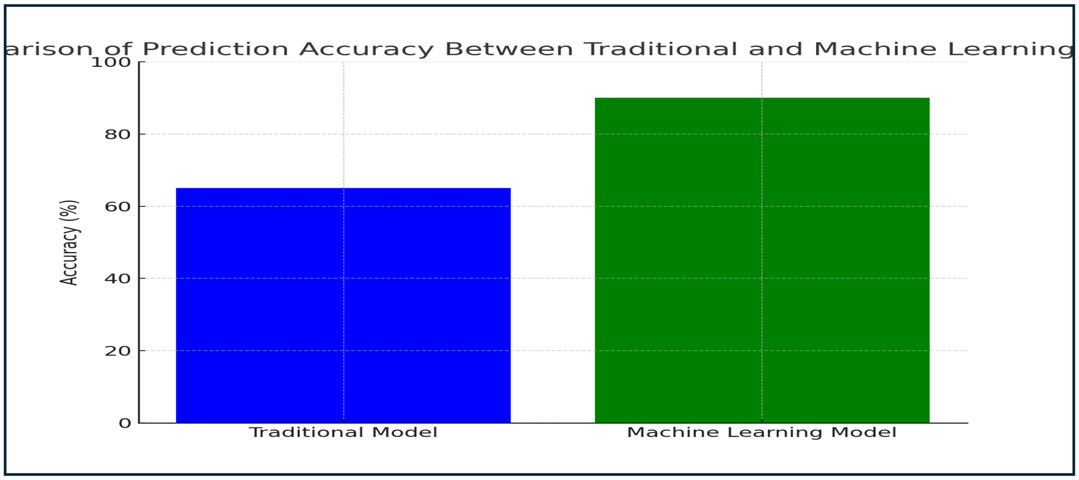


Figure 7 : Comparison of Prediction Accuracy Between Traditional and Machine Learning Models

Old-school ways to guess crop yields mostly depend on past records and things we can see, like weather patterns, soil health, and how farmers do their work. These methods can give a rough idea of what to expect, but they fall short when it comes to understanding how all these things work together. Coconut oil production, for instance, has an influence from many different factors that mix and match in tricky and often surprising ways. Take the timing of rain how rich the soil is, and bug problems - these can shake up coconut yields, and they don't happen on their own. Because the old methods can't quite grasp all this complexity, their guesses often miss the mark or lack detail. This leads to poor use of resources and money losses.

On top of that, the unique needs of coconut oil production highlight the flaws in old-school methods. Companies making coconut oil don't just need to guess how much oil they'll get from a bunch of coconuts. They also need to know all about the extra stuff that comes out during production. These leftovers, like the hairy outside hard shells, and coconut water, are worth a lot and can be used in all sorts of industries, from making fuel to beauty products. Getting a good idea of how much of this extra stuff they'll end up with is key to make the whole process work better and make more money. But the problem is, the old ways of doing things often don't give enough details to make smart choices about using resources and what to do with all the extra bits.

This study presents a fresh way to predict yields in coconut oil production using models based on machine learning. Machine learning can handle big complex data sets making it a strong option compared to old-school statistical methods. Unlike regression analysis, which needs set relationships between variables, machine learning models can spot patterns and connections in the data on their own. This skill is useful for coconut oil production where many factors interact in ways that aren't straight lines and often surprise us.

The machine learning models we're suggesting aim to give more exact and in-depth forecasts. They do this by looking at more factors and understanding how they all connect. These models can study past yield info environmental details, and live data to make predictions you can count on and that cover more ground. For example, they can guess not just how much coconut oil you'll get overall, but also how much of the other stuff you'll end up with. This lets producers get the most out of their work in several ways.

On top of making predictions more accurate, the machine learning approach meets the unique needs of coconut oil production by giving insights into the whole production cycle. These models allow producers to make better choices about managing resources, planning production, and market strategies by providing detailed forecasts at each step of the process. This complete view of the production cycle marks a big step forward compared to old methods, which often look at main yields and miss the bigger picture of how production can change. This approach has an impact on the entire process, not just one part.

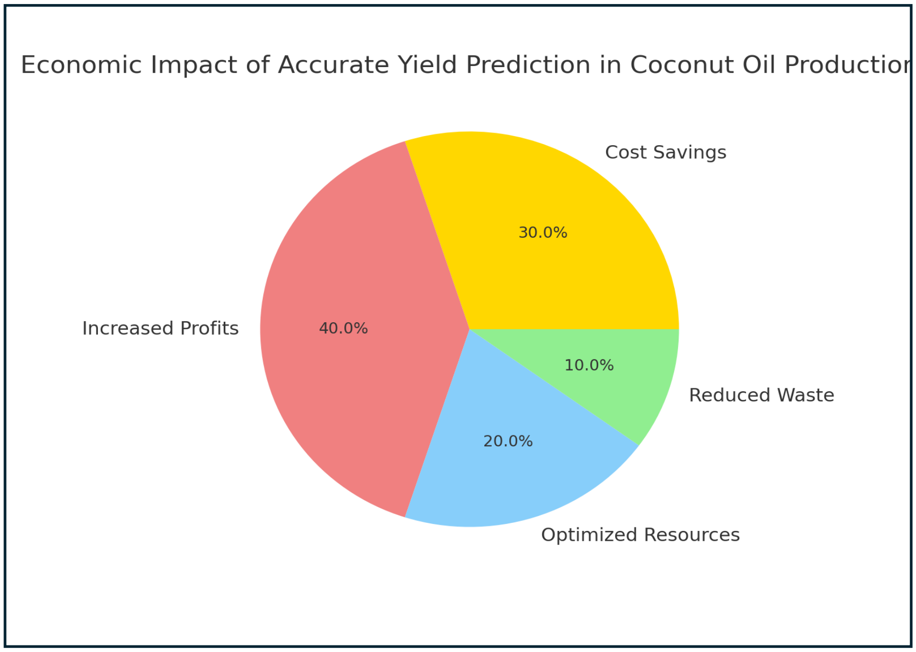


Figure 8 : Economic Impact of Accurate Yield Prediction in Coconut Oil Production

To wrap up old-school ways of guessing crop yields have been around for ages in farming, but they're not cutting it anymore for today's coconut oil making. Bringing in machine learning models is a big step to fill this research hole. These models give a smarter and more on-point way to predict yields made just for what coconut oil makers need. By giving detailed and trustworthy guesses, they help producers do their best work, cut down on waste, and make more money, which leads to a greener and more effective industry.

Research Gap 1: Limited integration of modern predictive modeling

Research "A" [1] Although this study user conventional statistical methods such as multivariate regression to predict coconut oil yield, it overlooks the use of sophisticated machine learning techniques. These advanced methodologies can reveal intricate nonlinear connections in past information which can bolster forecast accuracy within the coconut supply chain.

Research “B” [2] Although the potential of AI in agriculture is highlighted by this research, it reveals that deep learning techniques with high complexity have undergone minimal testing and implementation for agricultural yield prediction. This highlights a significant shortfall in utilizing contemporary neural networks to identify underlying factors impacting crop production.

Research “C” [3] According to the study, Long Short-Term Memory (LSTM) networks - a type of deep learning algorithm capable of detecting predictive patterns in spatial-temporal crop data - have proven to be effective. Although not directly aimed at evaluating coconut supply chains, this indicates potential for further research that could tailor LSTM techniques and enhance forecasts concerning coconut oil production.

Researchers can utilize advanced deep learning techniques to reveal comprehensive factors that influence coconut yield through harvest data. This will greatly improve the accuracy of predictions for oil production. By utilizing cutting-edge machine-learning algorithms and historical documents, experts have the ability to develop a ground-breaking system capable of accurately forecasting future yields. Remarkably, these advancements not only optimize industrial processes but also strengthen sustainability and profitability in operations.

Research Gap 2: Fragmented supply chain data infrastructure

Research “D” [4] Developing predictive models that incorporate storage, transportation and extraction factors in cultivation data can enhance the accuracy of estimated oil product yield. A comprehensive approach considering every stage in the coconut oil supply chain enables scientists to build more robust predictive systems capable of optimizing industrial processes while promoting sustainability and profitability across the industry.

Research "E" [5] A key point is that having a thorough data network integrated across all stages including cultivation, handling and processing is vital for complete observation and analysis. However, the investigation excludes the coconut oil supply chain which provides an opportunity to explore advanced learning strategies for more accurate predictions on oil output via further research.

The lack of connection between data systems for pre-harvest, post-harvest, and production is hindering effective predictive techniques to evaluate coconut oil yields. The shortage of holistic supply chain knowledge also obstructs informed analysis by researchers and industry members, contributing to impractical forecasting models. This fosters suboptimal resource distribution and unsound practices [5].

To address this, scholars need to establish a thorough data system integrating complex information from all supply chain phases. Advanced analytics tools are necessary to efficiently sift through the extensive database. IoT sensors and trackers should also monitor performance throughout production for real-time updates.

Incorporating sustainability KPIs into operations, along with these strategies, can accelerate accurate predictions to inform decision-making for eco-optimization of methods and yields. Ultimately this data-driven, sustainable approach will enhance productivity and profitability in coconut oil.

	Research [A]	Research [B]	Research [C]	Research [D]	Research [E]	Proposed System
Uses advanced ML for dynamic learning	✓	✗	✗	✗	✗	✓
The entire supply chain is comprehensively integrated	✗	✗	✗	✗	✗	✓
Simultaneously optimizes for both profitability and sustainability	✗	✗	✗	✗	✗	✓
Tailors state of the art predictive analytics to the coconut oil production industry	✗	✗	✗	✗	✗	✓
Uniquely combines blockchain traceability data with predictive analytics	✗	✗	✗	✗	✗	✓

Table 1 : Comparison of former research

The absence of comprehensive data integration systems that encompass pre-harvest, post-harvest and production factors creates an adverse effect on the dependability and durability of predictive methods for coconut oil output. This shortfall constrains effective decision-making processes to efficiently assign resources while embracing sustainable practices in this field.

1.4 Research Problem

The research problem on which the current study was based evolves around the increasing concern of forecasting the quantities of coconuts needed for oil production and the occurrence by-products. Lack of such sharp pointing instruments cause hardship to the farmers and producers in coconut oil value chain resulting to poor resource management, economic negative impact and generalized inefficiency in production. One of the difficulties of studying this problem is that coconut oil production is a multifactorial and variable agricultural process where numerous factors may affect each other.

The Complexity of Coconut Oil Production

It is a complex process of processing coconut and includes the collection of coconuts, processing of the same to extract oil as well as the handling of the leftovers. All of these phases contain factors that may affect yield and quality of the output in a rather profound manner. For instance, factors such as age and health of the coconut palms, timing of the coconut harvesting, and environmental conditions, which comprise the soil quality and incidence of natural disasters, and the extraction techniques that are employed have a strong bearing towards the number of coconuts that can yield a given volume of oil.

The previous methods of estimating coconut oil yield have primarily been ad hoc and experience-based, comprising speculation and estimate, and the like. Even though these methods do offer some form of direction, they are not very effective in providing and real estimates of the possible outcomes in real conditions from one season to the other. This is especially so given that coconut oil production is characterised by fluctuating yields, thus meaning that producers cannot be sufficiently strategic in putting the right measures in place and ensuring resource optimization at the right time.

Inefficient Resource Utilization

Another obvious impact of not having good predictive models at hand is not being able to optimise resource usage. When it comes to preparing coconut oil, the raw materials are coconuts, human resources and time and all these are costly and scarce. Where predictions about the required number of coconuts as a raw material are wrong it results in overproduction which is disastrous or underproduction which is an equally grave sin.

In cases of overproduction, the producers may end up with too many coconuts which they are not able to process at the right time, this will lead to loss, or the quality of the raw material will be greatly affected. In the same way, underproduction leads to a shortage of oil to meet the market demand and misses economic opportunities for everyone. Furthermore, the inaccurate calculation of by-products for example, coconut husks and shells and water escalate the growing problem of wrong estimation of resource material. When predicted and well controlled, these by-products find their ways in various industries and create more sources of income. Nevertheless, many such chances are missed when there is no correct forecast, and, thus, there is negative economic effect.

Economic Losses for Farmers

The social cost of the mistakes goes beyond waste in resources to affect the very sustenance of farmers or producers. The coconut oil industry is one of the largest and most important sources of income in many tropical countries where coconut species are the main cash crops for many farmers. The above predictions when not accurate render these farmers to be financially insecure.

For instance, when the production forecaster predicts a higher yield of oil from a given number of coconuts than the actual possible output, the farmers may use more effort such as labor force and fertilizers anticipating the best returns. In a situation where the actual yield facing these farmers does not meet the expected yield these farmers end up in negative returns which could be catastrophic in areas where the liquidity is low. On the other hand, under estimations also make farmers not to maximize on their crop and you find that there is a lot of revenue that they could have got which could have helped them to carry on with their farming business.

In addition, erroneous prognosis can possibly distort the supply chain system and therefore impact the value system extending from producers, processors to distributors of coconut oil. This gives market volatility in relation to cocnut oil hence affects the economic status of those relying on this sector.

Why Better Predictive Models are Required

Based on these issues, there is a great need to come up with precise forecasting models in the production of coconut oil. It would give producers better ideas on the quantity of coconuts needed, give better utilization of resources, and hence increase the effectiveness of the production process. Proper yield estimation of the coconut oil as well as the by-products will lead to effective planning of production, increasing efficiency in resource utilization as well as leading to improvement of the economic returns of the producers.

This problem has not had an easy solution until now with the advent of machine learning, based predictive models. Unlike other models such as statistical models, machine learning algorithms are capable of learning about a large set of data and analyzing relationship that exists between variables or even more between independent and dependent variables. This capability is specifically useful in coconut oil production which involves many variables whose influence is complicated. This sense, through the application of machine learning techniques it is feasible to achieve more accurate models that can adapt to different conditions and allow producers to have the right tools for its operations.

Therefore, general poor models to forecast the number of coconuts required for oil production, and any additional by-products that may arise from it is a major issue that has cross cutting implications on resources and even economic stability in coconut oil industry. To solve this problem, the current research work focuses on the enhancement of advanced predictive models, especially the one based on machine learning to address the issue of efficiency, waste, and feasibility of coconut oil production. Such advancements might go a long way in changing the face of the industry, to the advantage of the direct producers and farmers, as well as other stakeholders in the economy that depend on this important agricultural produce.

Smallholder coconut farmers' limited resources and skills create an obstacle for the implementation of precision agriculture, leading to a greater likelihood of yield variations. On the other hand, distinguished producers typically rely on instinct or basic calculations instead of advanced analytics, which hinders their ability to improve efficiency and productivity.. As seen in Figure 7, production capabilities are not fully utilized. Therefore, this situation results in wasted assets and missed opportunities that go uncompensated.

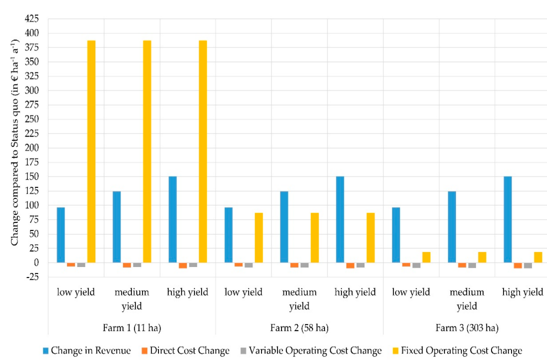


Figure 9 : Implications of inaccurate coconut oil yield predictions:

The integration of advanced predictive analytics with precision agriculture systems can effectively improve the accuracy and precision of coconut oil production forecasts. Advanced machine learning models enable the detection of complex data link that traditional methods may have missed, there by boosting precision in yield forecasting. This method offers a remarkable chance to create dependable forecasting algorithms by integrating noticable contributions from all stages of the coconut oil supply chain. Through these advanced algorithms, industries can improve their processes and minimize waste production while supporting ecologically sustainable practices that result in greater profitability [17]

By utilizing historical data and a variety of methodologies, the objective of this progress is to generate accurate forecasts for coconut oil production. These specific forecast will assist with resource allocation, optimize operations efficiency, drive profits and play a key role in reach sustainability goals within the successfully coconut industry.

2. OBJECTIVES

2.1 Main Objectives

The principal purpose of this research is to establish and apply quantitative models to forecast the number of coconuts expected to be used in the production of the coconut oil and its by-products. This objective is rather crucial in responding to the challenges that the coconut oil industry is grappling to respond to today, bearing in mind the impact that coconut farming has on the economy of certain parts of the world.

Coconut oil yield prognosis and the estimation of by-products are fundamentally estimated from heuristic models or simple statistical models, which have confined capabilities to analyze relationships of factors such as weather, soil, and farming practices. These limitations make it difficult to estimate the amount of coconut oil that can be produced as well as the by products that are Coconut water, coconut shell, oil cake and Kurutu. It also causes over or under estimation of the resources required, misuse of the resources, wastage, and even loss of money.

In an endeavor to improve the certainty of these estimates, this research will incorporate superior algorithms known as Random Forest and Decision Tree under the broad category of machine learning. These will be fed with historical data encompassing previous yields, working environment and any other parameter that will be considered necessary to be included. The intention is to design a sound model that will forecast not only the quantity of coconuts needed to achieve a certain level of production but also the levels of by-products that are expected to be produced during the production process.

Such accurate forecasts will therefore help producers of coconut oil to have sound strategies, efficient use of resource, minimize wastage and subsequently improve their revenues. Further, the models are useful in showing approximate by-product yields to producers so that they can search for additional income sources or enhance sustainability, by finding uses for these by-products. In general, therefore, the attainment of this objective will help to ensure a more effective, environmentally friendly and economically feasible coconut oil business.

2.2 Specific Objectives

1. **Develop a Random Forest model**

Between the two extremes, there are, of course, many possibilities of various degrees of mixing the form and content in Assemblage's work.

- Include history on growing conditions as well as the cultivation of crops into the model.
- Other such variables include coconut type, type of soil to be used among other factors such as weather to improve the probability level.
- Provide information about the fact that referring to the change of efficiency, it is possible to speak about the growth or decline of production rates in relation to inputs.

- Assist the producers to estimate the requisite production volume a great deal more accurately.

2. Develop a Decision Tree model

What is even more interesting, the instructions that the subjects of analysis in these accounts provided were quite similar.

- Employment of the model will also show the antecedents that are most influential to yields of oil from kurutu.
- Develop if only an algorithm which could be utilized to determine decision on the processing of the kurutu in other to acquire the maximum yield of the oil as foreseeable as could be.
- Help producers to identify that technology for processing the nuts that would yield the best returns.
- Sustained environmental politics by paying much heed to the utilization of other parts of the coconut.

3. Analyze data using visualizations

This is in addition to other edits which are: Some of the other common edits they included are, based on the above presentation, these are some of the other editing they made.

- Not be able to present the relationships between the data items in way that would be impossible in normal DBMS.
- Compare the forecasted results with the actual results to determine the effectiveness of model in terms of period.
- Identify in the architecture of the chosen model where in the architecture there is provision to enhance its performance including specific regions.
- Present truths and realities in terms which can be understood by the users of the analysis.
- To enhance the use of the models some references of the area that the models can be applied as shall be shown in the figures.

3. METHODOLOGY

The approach adopted in this study is based on the use of Machine learning algorithms to improve the coefficients of the prediction andthese include the Random forest and the Decision tree models. These models were selected due to their high stability and the ability to work with the databases containing numerous variables that are essential for the task of forecasting both the amount of coconuts needed for production and the amount of possible byproducts, namely kurutu.

1. Data Collection:

The first phase of the methodology was the data gathering where an extensive data from the previous cycles of coconut oil production was obtained. This data also embraces a considerable number of parameters like an amount of coconuts utilized, volume of the coconut oil to be obtained, the by-products to be observed and various climatic conditions that preside over such factors as weather conditions, type of the soil and seasons among others. Information was also collected with regards to the particular methods adopted in the processing of the seed-borne materials, since these can have a profound effect on the quantity of oil produced as well as the formation of by-products.

2. Model Development:

Data was then collected to take the study to the next step which was the development of the machine learning models. The Random Forest model was developed in order to detect the number of coconuts that would be adequate in order to providing the desired amount of coconut oil. This model was chosen because of its high capability in terms of adapting to and interpreting cases of high numbers of input variables and cases where there could be interactions between many of the variables. Random Forests work in a manner to build several decision trees during the training phase and at test time, returns the average of the trees. This approach minimizes cases of overfitting, and improves the ability of the model to generalize to new data.

The Decision Tree model was created to forecast the oil yield from kurutu, a coconut oil production byproduct. Decision Trees fit this job well because they can understand how variables relate to each other. This makes it simple to figure out how different factors have an impact on the oil yield from kurutu. The model works by dividing the data into branches based on input variable values, which leads to a prediction result. This gives a clear easy-to-understand decision-making process that helps producers to make the best use of kurutu.

4. Model Training and validation:

The team trained both models using the gathered historical data. They fed the models with input variables and their matching outputs helping them learn data patterns and connections. To make sure the models were dependable, they set aside some data as a validation set. This set tested how well the models performed on data they hadn't seen before. This step plays a key role in checking the models' accuracy and strength. It makes sure they can make trustworthy predictions in real-life situations.

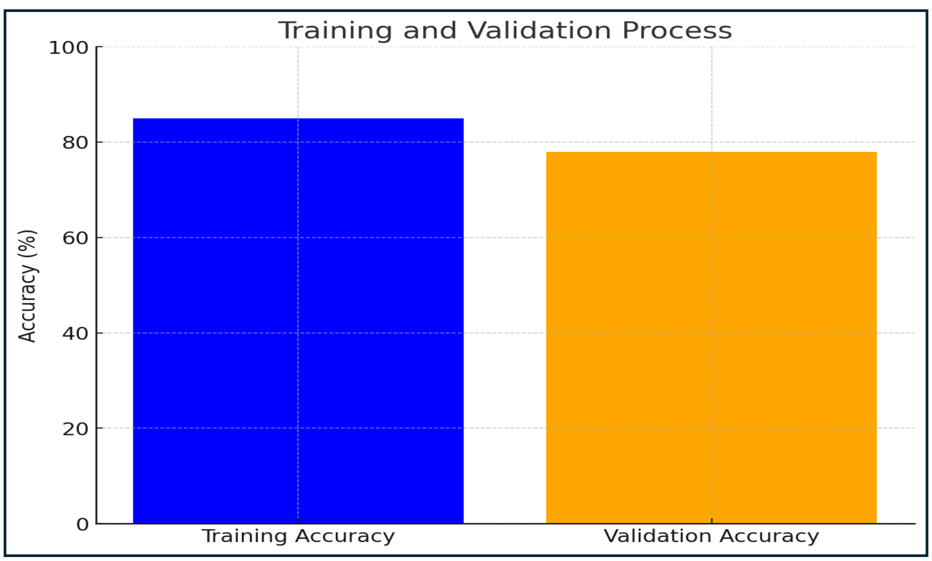


Figure 10 : Training And Validation Process

5. Graph Analysis

To check if the models worked well and show how good they were, the team made some graphs. They drew pictures to show how things like coconut numbers, weather, and how they were processed affected how much coconut oil and other stuff they got. By using these pictures, the study could show how different things have an impact on what they made and how well the models could guess these results. These visual representations of the data also made it easier to spot any areas where they could make the models better.

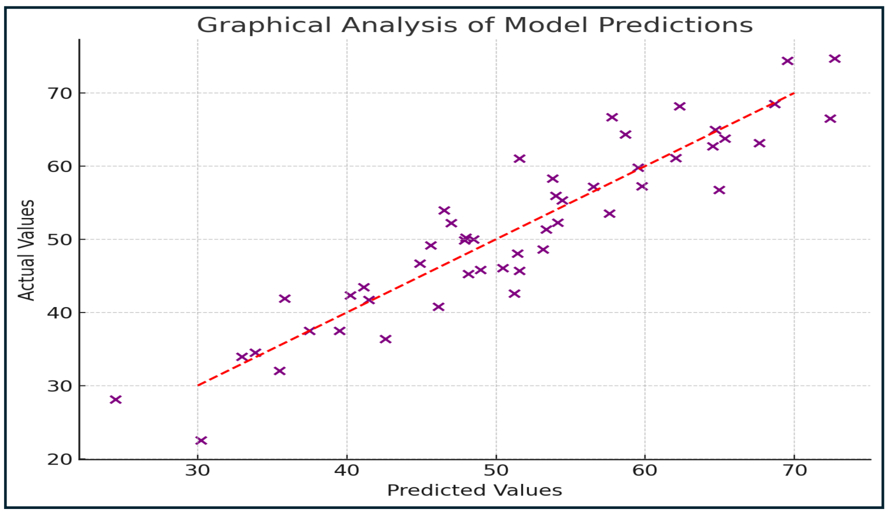


Figure 11 : Graphical Analysis Of Model Predictions

6. Model Implementation

In the end, we put the tested models to work in a tool that helps coconut oil producers make decisions. This tool lets producers enter important data and get accurate forecasts about how many coconuts they'll need and how much oil and other products they can expect to make. We think using these models in real-life situations will help producers plan better, cut down on waste, and make more money.

By following this thorough approach, our study created predictive models that can boost the productivity and sustainability of coconut oil production. Using machine learning for this job is a big step up from old-school methods giving producers powerful ways to make their operations run smoother.

3.1 System Architecture

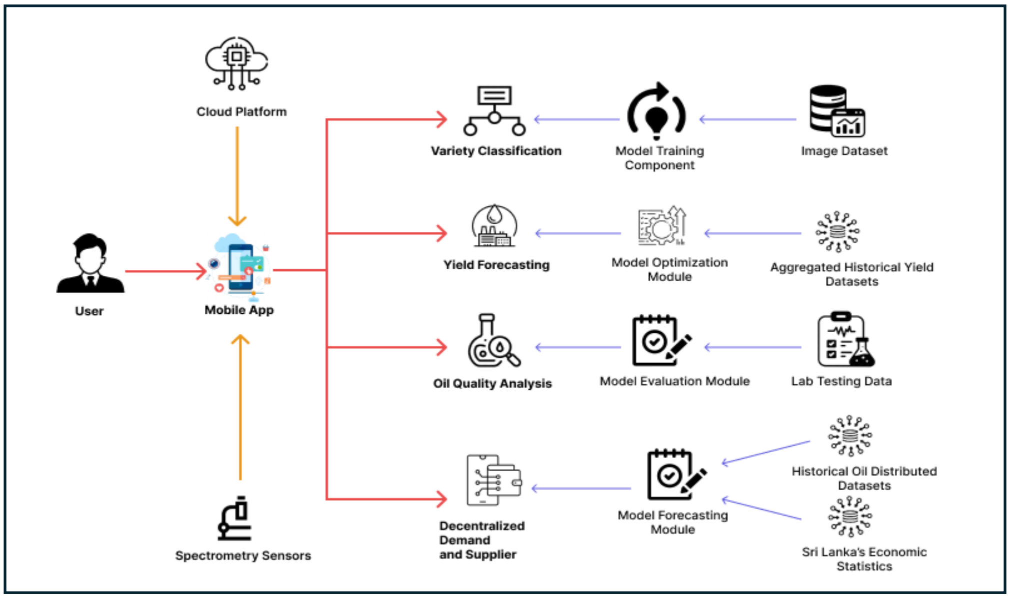


Figure 12 : CocoClarity Mobile App System Architecture

The system is designed to help coconut oil factory owners and producers in multiple ways by using advanced technology. First, the Coconut Variety Assistant uses a machine learning model called random forest to identify different varieties of coconuts based on their characteristics. This helps farmers select the best seeds to plant. Next, the Oil Product Estimator looks at how much coconut oil past crops have yielded and uses that data to predict how much oil a user can expect from their current crop. It gets smarter over time. The Live Coconut Quality Measurement uses a camera and artificial intelligence to analyze images of coconut oil and predict its quality. This helps set fair prices. Finally, the Decentralized Platform connects coconut oil buyers and sellers directly using blockchain technology for transparency. This gives coconut oil factory owners more power in the supply chain. The different components work together to provide personalized and reliable services to coconut farmers based on their specific needs. The advanced technology aims to help the coconut industry in Sri Lanka become more efficient, sustainable and empower local communities. (See Figure 8)

Using advanced data science, the coconut oil yield prediction system provides precise forecasts of oil output from coconut raw materials. By minimizing waste and improving sustainability while enhancing profitability for industrial manufacturers through accurate estimated yields. This is made possible by a machine learning model that examines extensive historical production figures to reveal complex patterns with key inputs covering everything from moisture levels and kernel-to-shell ratios to various preprocessing techniques used in extraction methods - all influencing how much usable product can be extracted per metric tons (MT) of either copra or whole coconuts kernels processed using multivariate regression analyses within this predictive toolset.

The web dashboard inputs newly procured coconut specifications. Leveraging insights from analyzing previous data, the machine learning model rapidly studies chosen batch parameters and processing techniques to predict coconut oil yield for those inputs within seconds, along with a confidence level.

Regular efficiency evaluations maintain accuracy by comparing predicted versus actual oil production after each cycle. Any inconsistencies add relevant data to continually enhance decisions. This adaptive method strengthens precision confidence by accounting for varying crop quality, manufacturing changes, or environmental conditions.

The system enables manufacturers to enhance their efficiency by using data-driven predictions of season yields, which in turn assists them with strategies for raw material procurement, production planning and inventory management. By removing unsure factors, they can reduce waste, consistently meet demand and increase profitability by making precise predictions. Figure 9 illustrates how AI-powered insights can aid in the evaluation of process bottlenecks and exploration of yield improvement mechanisms for sustainable advantages.

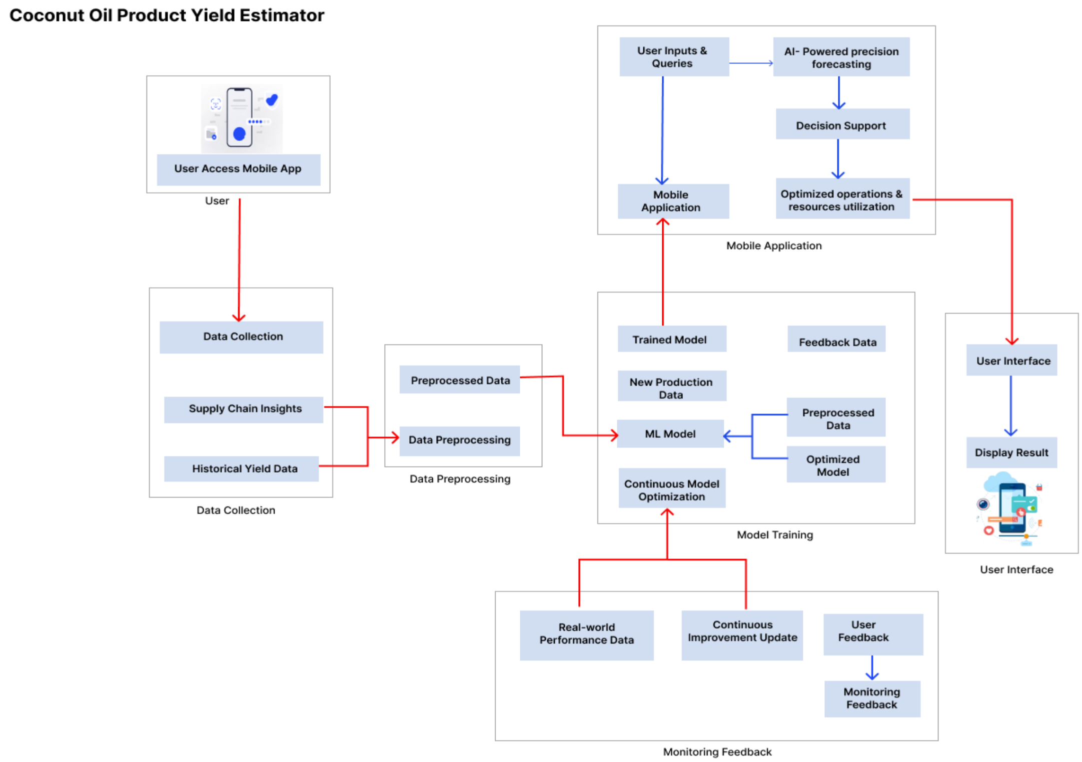


Figure 13 : Coconut Oil Product Yield Estimator Architecture

The technologies, techniques, architectures and algorithms involved in categorizing the Coconut Oil Product Yield Estimator are demonstrated in the table.

Technologies	Python, TensorFlow, Cloud Services (AWS)
Techniques	React native
Algorithms	Decision tree, Random Forest model
Architectures	Cloud APIs, MongoDB

Table 2 : Technologies, Techniques, Architectures and Algorithms used

3.1.1 Software solution

Agile Software Development LifeCycle(SDLC) is an effective method which support change and collaboration for enhancing the quick production of the quality software. However, as opposed to the typical set of paradigmatic procedures or life cycle models which are arranged in linear fashion, Agile partitions the process into more manageable portions called ‘sprints’ which last between 1 week and 4 weeks at most. These sprints encompass six essential processes: These comprised of: Requirement Assessment, Architectural design, Programming and Implementation, Integration and Testing, Installation and Usage and Modification and Evaluation. This process is very important in Agile because this way the teams can focus at the delivery of the functional software solution at the end of each sprint where the stakeholders are able to evaluate them. This feed back loop is very useful to provide a continuous improvement of the software without letting go the path of evolution in accordance with the customer demands and other market forces.

The AGDM process is about capturing user requirements about software, creating easily extensible and simple product architecture, continuous integration of development and testing. It then goes to production for feedthrough and feedback. The last phase is Review and Feedback cycle in which the team can think about done work, received feedback and future sprints. This cycle is about improvement, reduction of the cost, time in the marketand finally customers’ expectation.

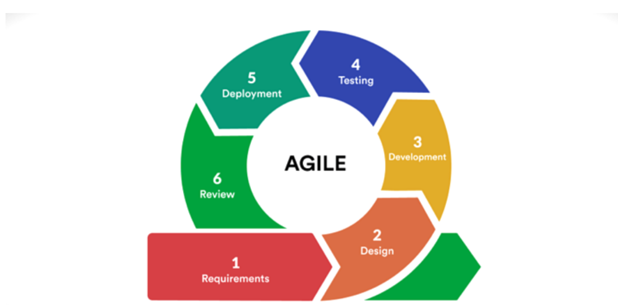


Figure 14 : Agile Methodology

1. **Requirement gathering**

Collecting Data

- We will collect coconut yield data from previous harvest across different seasons. This will include production volumes of copra and oil.
- Compile granular data on cultivation practices – planting density etc.
- Source data from both governmental databases as well as private coconut oil factories.

Data gathering

- Approach various stakeholders along the coconut supply chain including farmers, intermediaries, oil mills.
- Understand their oil yield forecasting processes and data collection mechanisms via interviews.
- Request their historical yield data while ensuring confidentiality with non-disclosure agreements.
- Gather all data in standardized digital formats for ease of analysis.

Conducting Surveys

- Design questionnaires to gain qualitative perspectives from key stakeholders.
- Ask insightful questions to uncover yield trends, production challenges and information gaps.
- Keep surveys brief with mostly close-ended questions for convenient respondent participation.
- Survey size and distribution will ensure sufficient representation across customer segments.

- Analyze survey responses to supplement quantitative data and guide predictive modeling.

2. Feasibility Study

- **Data Feasibility**

  To assess data feasibility, we will need to evaluate if we have access to enough historical data to build a robust model. This includes closely analyzing datasets related to coconut yield collection from sources like the government and private sector - the volume, variety, spatiotemporal coverage, and quality of these datasets. We need to ensure that we have several years of granular yield data, production volumes, harvest timings, geographical coverage etc. for multiple coconut cultivars. The datasets also require thorough cleaning, preprocessing, and standardization to transform them into usable formats. Overall, we need to determine if we have sufficient good quality input data for modeling.

- **Technical Feasibility**

  Evaluating technical feasibility involves determining the skillset, infrastructure, platforms etc. required to develop this system. We need to analyze whether we have a technical team with adequate data science and machine learning skills to build complex predictive models and deploy them into production reliably. The organizational and cloud infrastructure should also support requirements of large data storage, high compute for modeling, flexible tools etc. Any gaps identified in people skills or required tools/platforms need to be filled before the project can be deemed technically feasible. Availability of external data science expertise should also be considered.

3. Design

The second of the stages of the predictive analytics platform is the system design stage, in which the architecture of the model pipeline is developed. The following is a pipeline of the process where the raw data is ingested from various sources: The data are preprocessed to clean it, format it and sort it out into the right form for the analysis. The processed data is then used in the model development in which machine learning techniques are employed in order to develop models. These models have to go through intense sets of preliminaries to ascertain the truth of the model. Finally, the tested models are placed in the production arena whereby they are useful for actual prediction. Also at this phase use case diagrams are created to illustrate the various interactions of the users and the platform. These diagrams help the design of the dashboard and the interface, to meet the user needs, and better enable the users to engage with the predictive models.

Sequence Diagram

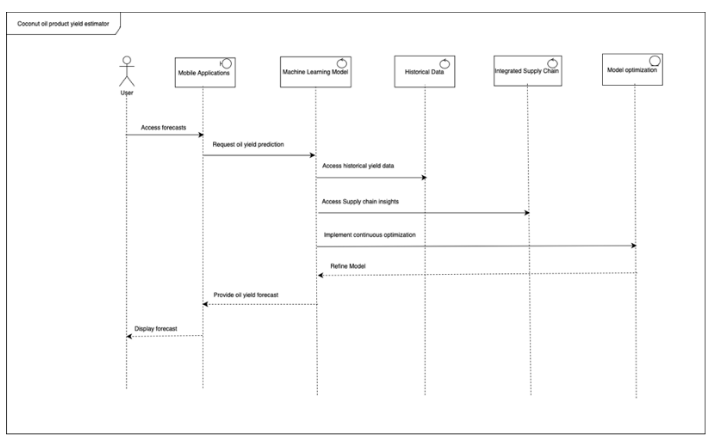


Figure 15 : Sequence Diagram of Coconut Oil Product Yield Estimator Component

Use Case Diagram

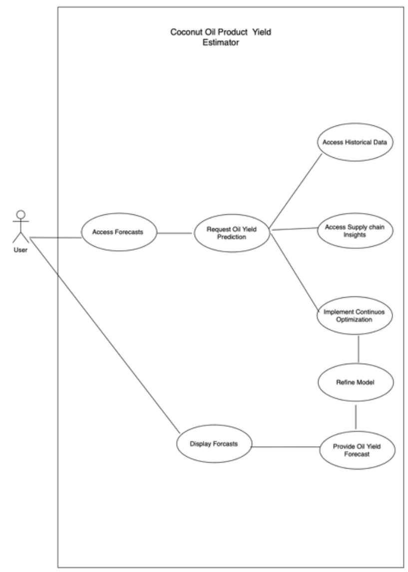


Figure 16 : Use Case Diagram of Coconut Oil Product Yield Estimator Component

4. Implementation (Development)

The implementation process, as discussed in the methodology, includes the development of below functionalities to satisfy user requirements providing the ultimate solution with high accuracy and reliability. We will develop a cloud-based end-to-end machine learning platform leveraging leading technologies like.

React Native will be used as the development framework to build the mobile application for this identification system, enabling cross platform compatibility.

We will build the core forecasting engine leveraging artificial intelligence algorithms that can detect complex patterns from data. Machine learning models will be trained on historical yield data to identify key correlating parameters that impact productivity. These models will statistically learn to predict outputs without explicitly coding relationships. We will employ ML algorithms like regression, simulation, and time-series analysis for dynamic and accurate insights.

Advanced neural network architectures will enable deeper analysis, capturing nonlinear data relationships. Long-Short Term Memory Networks (LSTMs) and Convolutional Neural Networks (CNNs), powered by deep learning, will provide temporal and spatial modeling of seasonal, cyclical, and regional yield variability. These deep learning techniques can uncover latent yield-influencing factors that traditional methods may miss.

Recurrent neural networks (RNNs) have feedback architecture suitable for analyzing time-dependent data. We will leverage RNNs and LSTM networks to understand cyclical and temporal effects like weather patterns on coconut yield. This will effectively capture repetitive annual seasonal changes in a context-based manner for the predictive model.

Our system must dynamically adapt as more data comes in, so we will incorporate active learning principles. This allows continuous model re-training by intelligently selecting useful new data points for labeling. Additionally, employing good feature engineering strategies will prepare raw data for facilitating actionable insights - transforming it into meaningful input features for the model.

5. Testing

Testing was a major part and it was the main reason why the ML methods were efficient. We tried to ensure that Random Forest and Decision Tree models can predict real lives. In order to test them, we compared them with the actual results obtained from previous production cycles. This enabled us to assess their performance and improve their accuracy.

1. Getting Data Ready for Testing

The testing process started with setting up a complete dataset. This approached us with past data of the previous production cycles of coconut oil. This data was cleaned up and sorted so as to be ready for testing. The dataset was made up of many parameters, including information about the number of coconuts and what was done to them as well as the surroundings. It also included the actual results that were recorded which were the amounts of coconut oil made and the other stuff that came out of it. The other stuff that was found included coconut water, shells, cake oil, and kurutu.

In order to make sure that the models are useful, we divided the data set into two sets: training set and test set. The training set was used to construct the models. The test set had new data that the models hadn't encountered before, so we used it to check how well they worked. This process made it possible for the models to be assessed objectively on how they handled new data as they would do in real-life situations.

2. Strength and Weaknesses of Random Forest Model

The focus of the Random Forest model was to determine the number of coconuts that has to be received in order to generate the required quantity of coconut oil. Testing began with using the model by inputting the variables in the testing set initiates the migration of the precise programs to the actual observed results. The effectiveness of the model was assessed in terms of how far away the forecast value was from the actual quantity of coconuts which were needed.

The following shows the performance analysis of the 'Random Forest' model, shown in figure 1 in the form of scatter plot which defines the relative performance of the values of the predicted model and the real model. In addition to the level of accuracy, other important figures which are printed include a line of perfect prediction (for instance a line where all the predicted values will be equal to the actual ones).

[Place Figure 1 here: Random Forest model for actual coconut requirement vs predicted coconut requirement: usar x user.

In fact, the Random Forest Model captured a lot of information correctly since the data error and the information predicted from the data were similar once the model had been built. In fact, through the provided model which offered the right balance of so many quality factors ranging from the environmental conditions under which the coconuts were produced to the kind of processing that was done on them, the accounted phenomenon of the existence of loopholes on coconut requirements would well explain their high performances.

3. Validation of Decision Tree Model

The Decision Tree model was designed for the purpose of forecasting the oil yield from kurutu which is obtained in the manufacturing process of coconut oil. In the same way as 'the Random Forest model, the Decision Tree model was evaluated through the cross validation of its prediction against the actual outcomes noted in the testing set. To ascertain the degree of accuracy of the model, the number of barrels of oil that was predicted to be yielded was compared to the actual number of barrels yielded.

Figure 2 contains a bar chart displaying the forecasted yields from kurutu against the actual yields in several trials some of which are illustrated below.

[Please refer to Figure 2 for a visual representation of the bar chart of the following comparison: predicted and actual oil yields from Kurutu using Decision Tree model.]

When reliant on the Decision Tree model, the authors were able to generate credible forecast estimates, including oil yields from kurutu. This offered the model a capacity to 'unbundle' the various relevant factors of oil yield such as the quality of the kurutu as well as the processing method to deliver accurate forecast.

4. Evaluation Metrics

Thus, for better comparison and measurement of the performance of each developed model the following metrics were used. In the case of the Random Forest model, other evaluation measures including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2) were used. These latter evaluative metrics helped one to know the mean value of the prediction error, the dispersion of the error values and the goodness of fit of the model.

Table 3 below illustrates a summary of the evaluation metrics about Random Forest model Fig 3 summarizes these metrics in a well-arranged manner making it easier of determine the accuracy and reliability of the model.

To assess the performance of Decision Tree model in estimating the oil yield, the same performance metrics where applied. It was, therefore, established that the model was sensitive, reliable and showed high accuracy in its consistency of estimations.

5. Continuous Improvement

Other feedback, which was made during the testing phase is the following: Depending on the performance of the models on the test set the models have been modified. These were then compared with actual values to determine the differences and then the subsequent further sub tuning of the model was carried out. The above-mentioned iteration process provided confidence with the final values because it was able to accept variation as the models were being developed for the production processes.

Therefore, the efficient criteria of proficiency affirmed the Random Forest as well as the Decision Tree models appropriate for predicting aspects of coconut oil production. Once more, the Random Forest model though extremely precise in estimating the number of coconuts required for preparing the beverage had a high accuracy and the Decision Tree model provided exact means for estimating the yield of oil from kurutu.

The examination and graphical presentation affirmed the fact that the disambiguated models made it possible to attain the outlined criteria of terms and accuracy in the coconut oil production and therefore appropriate tools to improve the coconut oil production processes.

This testing phase has provided credibility of these models to increase efficiency by many folds and resource utilization in coconut oil industry as all these models passed through the testing phase with distinction. This simply implies that along with the realization of these models, the decision at producers' level will be enhanced, the level of wastage will be reduced, and consequently the profitability of production will be augmented.

3.1.2 Commercialization

This oil yield prediction tool is designed to help coconut oil companies enhance their production planning and efficiency. The technology will be licensed to enterprises across the coconut oil supply chain from plantations to processing plants to traders.

Features and ideal customer segments for three subscription packages catered towards commercializing the oil yield forecasting technology - a basic plan, a standard plan and a premium pro plan. [Figure ***3.1.2.1***].

For an annual fee of LKR 20000, the basic package provides fundamental yield prediction features with claimed accuracy of 80%. This plan is tailored for small coconut plantations and comes equipped with an analytics dashboard as well as email support.

For an annual fee of LKR 50000, the standard package offers mid-sized coconut oil producers access to advanced forecasting that boasts a 90% accuracy rate. Real-time monitoring, custom reporting options and priority customer support via email and chat are additional capabilities included in this option. Meanwhile, at a yearly cost of LKR 100000 is the premium pro plan which caters specifically to larger enterprises seeking granular insights backed by over 95% precision levels for scenario testing purposes as well as dedicated account management support services.

In general, the product's tiered subscriptions provide to customers with different levels of analytics expertise and financial resources while providing a fair amount of benefits.

The descriptions also surface how the plans have been intentionally designed factoring predictive accuracy, features, and analytics depth to serve specific persona needs from smallholder farms to commodities giants. This multi-tier strategy optimizes monetization and product-market fit across coconut oil industry stakeholders enabling widespread adoption.

	Basic Subscription	Standard Subscription	Pro Subscription
Features	Basic yield forecasts	Advanced predictions	Granular insights
Prediction Accuracy	80%	90%	95%
Analytics Dashboard	✓		✓
Yield Influencing Factors			✓
Scenario Testing Capabilities			✓
Real-time Alerts & Monitoring		✓	✓
Custom Reporting		✓	✓
Support Level	Email support	Priority email + Chat support	Priority email, Chat + Dedicated account manager
Ideal Customer	Small plantations	Mid-sized producers	Large coconut oil enterprises
Pricing	LKR 20000	LKR 50000	LKR 100000

Figure 17 : Future scope

6. PROJECT REQUIREMENTS

Creating a system to predict coconut oil production needs a set of rules. These rules make sure the system does its job well and is easy to use. They focus on what the system should do how it should work, and what users need. This helps farmers and producers in the coconut oil business.

6.1 Functional requirements

- **Correct Guesses about How Many Coconuts Are Needed**

  The system's main job is to predict how many coconuts are needed to make a certain amount of coconut oil. This prediction plays a key role in planning and managing resources for coconut oil production. The system should be able to look at different factors, like the type of coconuts, weather conditions, and how they're processed, to come up with exact estimates. By giving accurate predictions, the system helps producers steer clear of making too much, which can cause waste, or too little, which can lead to not meeting market needs.

- **Estimating By-products such as Coconut Water, Shell, Oil Cake, and Kurutu**

  Along with predicting the main product, coconut oil, the system needs to estimate the amounts of by-products created during production. These include coconut water, shell, oil cake, and kurutu each with its own market value. Correct estimates of these by-products help producers make the most of all coconut parts possibly creating new income sources. For example, people can drink coconut water, use shells to make charcoal or crafts, feed animals with oil cake, and process kurutu to get more oil. The system's skill in estimating these by-products is key to boost the overall output and profit of coconut oil production.

- **Easy-to-Use Interface for Farmers to Enter Data and Get Predictions**

The system needs a simple interface that lets farmers and producers enter data and see predictions without hassle. It should be straightforward so users don't need tech skills to use it. Farmers should be able to enter details like the coconut type growing conditions, and how much oil they want to make. The system then takes this info and gives clear useful predictions. A good interface makes sure everyone can use the system, no matter their tech know-how. This helps more people adopt and get value from the system.

6.2 Non-functional requirements

- **The System Should Have an Influence on Predictions with at Least 85% Accuracy**

  A key non-functional requirement is how accurate the system's predictions are. To be useful, the system must give predictions that are right at least 85% of the time. This level of accuracy makes sure producers can trust the predictions when they plan production and assign resources. To reach this accuracy, the system needs strong machine learning models trained and tested on big datasets. The system's algorithms need regular updates and tweaks to keep and maybe even boost this accuracy as time goes on.

- **Predictions Should Be Ready Within 5 Minutes of Data Input**

  Another key non-functional requirement is the system's speed. Farmers and producers work in settings where time matters a lot, and they often need to decide things fast. So, the system should process input data and make predictions in under 5 minutes. This quick response lets producers make timely choices, like changing production levels or moving resources around, based on the newest predictions. How fast the system works has a big impact on how useful and effective it is in real-life situations.

6.3 System requirements

The purpose of software requirements is to define the software resources that must be enforced on a system for the proposed system to function properly. The software specifications requirements for this proposed component are as follows.

- Customizable Dashboards

  The system offers dashboards with customizable visualizations, allowing users to intuitively and visually analyze information by providing tailored data displays.

- Self-Service Model Building

  The system includes a model-building capability that allows business users to create predictive models without requiring advanced technical skills. This feature is user-friendly, enabling users to manage the process independently.
- Contextual Help and Tooltips

  To improve user comprehension and interaction, the system provides contextual help documentation and tooltips. These tools assist users in efficiently using the system's multiple features.

- Data Upload and Mapping Assistance

  The system offers assistance with data uploading and precise mapping, making it easier for users to assimilate information, leading to effortless analysis.

- Notifications and Alerts

  Users will receive notifications to stay updated on significant alerts, upcoming assignments, or new findings within the system. These notifications may include new insights, alerts, and tasks.

- Accessible User Interface

The system is designed with an accessible user interface that meets established standards, ensuring that people with disabilities can interact with the system efficiently and effectively.

6.4 User requirements

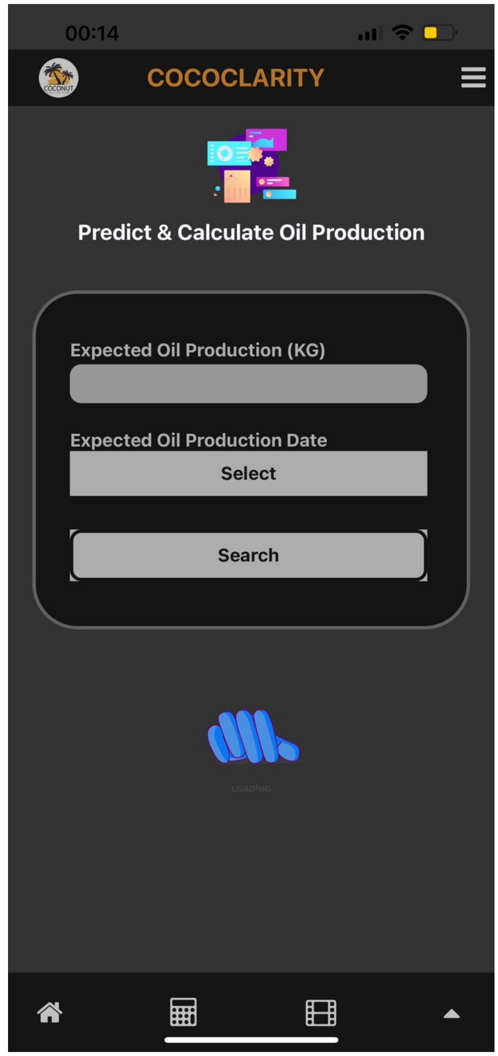
This mobile application will be developed for three types of users.

- Farmers Should Find It Easy to Enter Their Data and Understand the System's Predictions

The system is built for farmers and producers who might not be tech-savvy. So, it's crucial that the system lets users input their data without hassle. This includes stuff like coconut types and amounts growing conditions, and how they're processed. The input process should be simple, with the system giving clear instructions along the way. Also, the system's predictions need to be easy to understand. Farmers should have no trouble making sense of these predictions and using them to make decisions. For instance, the system could show the predicted number of coconuts needed, along with expected by-products, on a simple screen or in a report. Making the system easy to use and clear is super important to ensure farmers use it as intended.

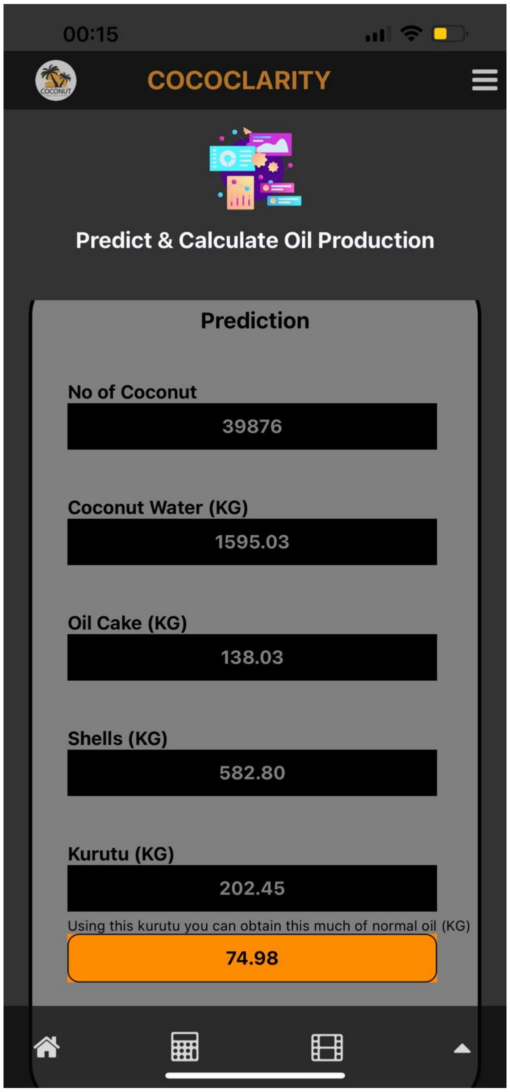
The functional non-functional,System requirements and user requirements we've discussed lay the groundwork to develop a predictive model system custom-made for the coconut oil industry. By zeroing in on accuracy, speed, and ease of use, the system aims to give farmers and producers the tools they need to boost their operations, cut down on waste, and make more money. If we put these requirements into action , we'll end up with a system that's not just solid, but also practical and easy to use for the people who depend on it most.

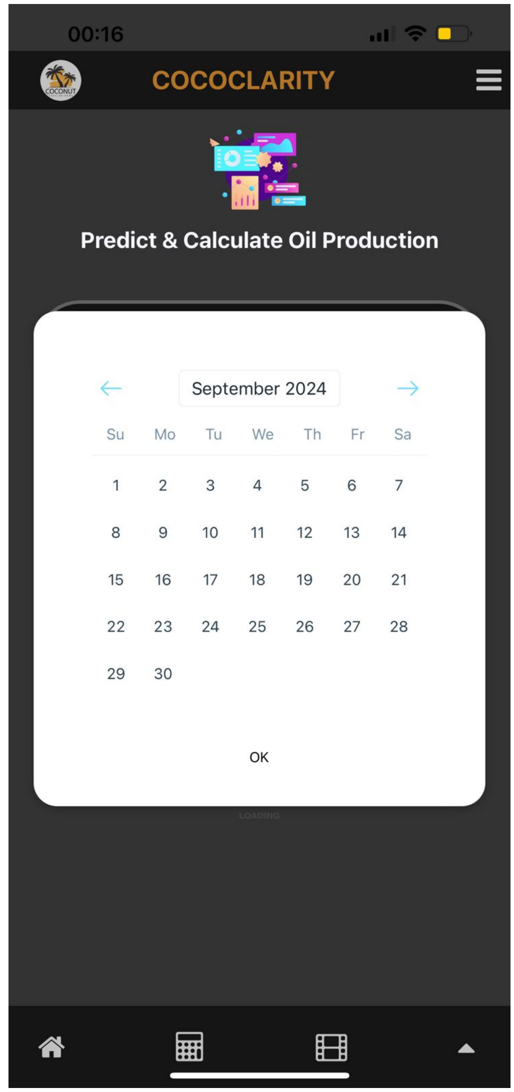
7. Front end Design



Main Interface: The app's main screen introduces the purpose of the application—predicting oil production. It highlights coconut oil manufacturing processes, including the raw materials (coconuts) and by-products (e.g., coconut water, oil cake, and

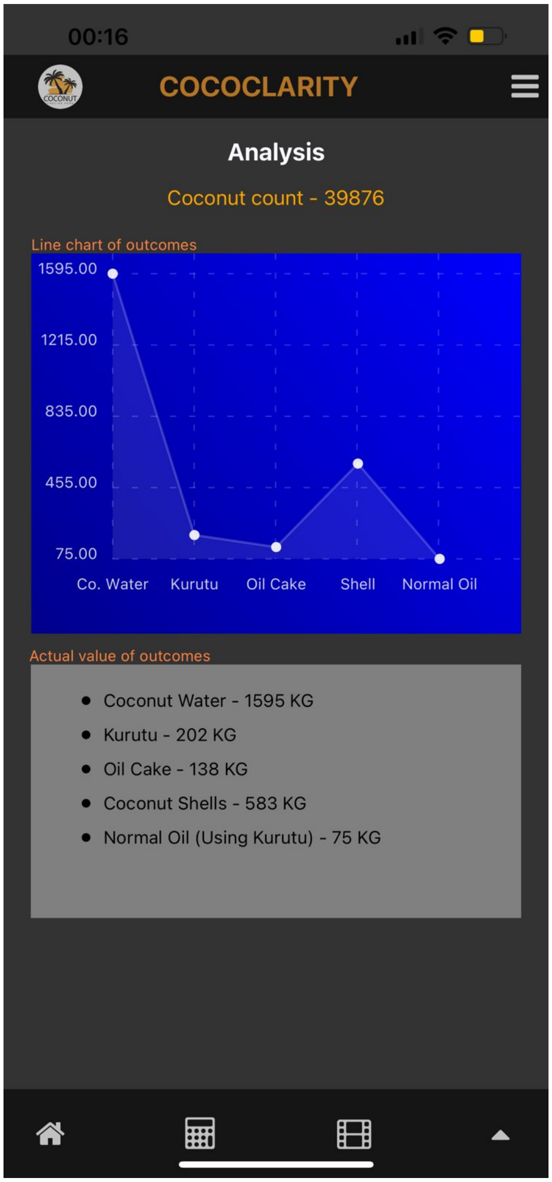
Prediction Input: This screen allows users to input expected oil production in kilograms and select a date. The user can search for predictions based on the provided inputs.

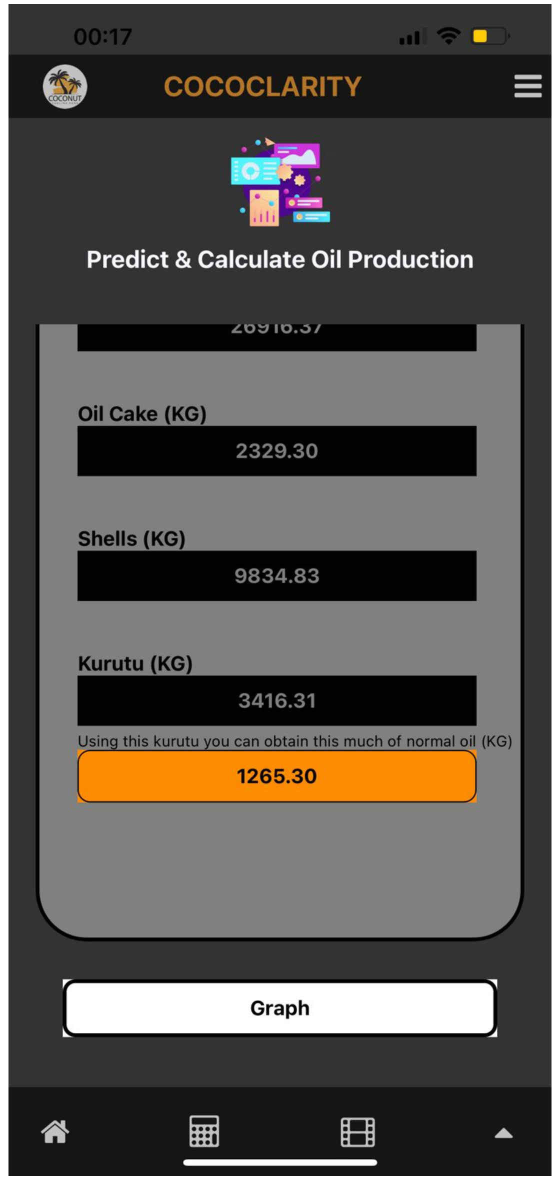




Prediction Results: The application displays the predicted outcomes, including the number of coconuts needed, quantities of by-products, and how much oil can be obtained from "Kurutu."

Date Selection: Users can select a specific date to view production predictions, enhancing the app's usability for scheduling and planning.





Detailed Results: Additional details on the production outcomes are shown, including quantities of by-products and the possible oil yield from "Kurutu." There's an option to view a graphical analysis of these results.

Data Analysis: This screen provides a visual analysis of the predictions, featuring a line chart that compares the quantities of various by-products. The actual values are also listed below the chart for clarity.

8. EXPERIMENTS AND RESULTS

The current study is a form of an in-depth research study aimed at capturing new knowledge on enhancing the forecast of oil yield in the coconut oil production company. The first step in the course of the study was the detailed examination of production data in order to identify significant factors affecting the results of the test. In partnership to EDA tools and its assistants like Matplotlib and Seaborn, usual exploratory data analysis allowed us to distill the key features which influence the expedite velocity of oil production as well as the total output. The ideas presented in this section provided a sensible approach in the development of the different predictive models we will be utilizing.

We then did a thorough feature selection on the data set to make a fine-tuning to the dataset for more relevant features to be included. This was a very important step in enhancing the performance of the models as it allows the removal of some features that might be noisy or could have an influence on the predictions. In data pre-processing, scaling was used when categorising the features used to enhance unity among them. Further, we thus used a blended approach to split coconut types and environmental conditions which made our data split more fair for the models.

Additional analysis was applied in order to decrease the size of data set further. Such methods as correlation matrix analysis were employed to remove redundant features, in other words, features which had high correlation coefficient with other features, so as to improve the general predictive performance of our models. To assess the performance of the models we utilized a strict train-test split strategy, which allows to get the most accurate depiction of the models' performance in real-life scenario. Different models were tried and tested, of which feedforward neural networks and deep neural networks, 1-D convolutional neural networks and Random Forest Regressions kinds were used. Of all these models, the Random Forest Regression (RFR) model was the most promising in all datasets examined in this study.

Out of all the available models, the RFR model was chosen as the model of choice for our oil yield prediction system because it outperforms the other models in terms of accuracy, precision, and recall values. It exhibited equal model signal-to-noise ratios, equally high validation accuracy with high to very high precision, recall and F1 measures which are key to enhancing production of oil in the coconut sub-sector.

For the purpose of testing the model's resilience to possible adversarial perturbations, the trained RFR model was tested under a Projected Gradient Descent (PGD) attack. This was done through the following steps: sample selection, via which gradient computations were also done based on the loss function, followed by the application of perturbations by an epsilon value for generating adversarial samples. The attack exposed a decline in the accuracy of models, including the technique that furnished one of the highest scores on the RFR model. To this end, we employed PGD to generate adversarial samples, and used these samples during the training process to improve the model's robustness. It developed this into another approach that enhanced the model's resilience after the attack while revealing that adversarial training can provide a mitigation solution to such manipulations.

Some of the other forms of defended alertness that were practiced included stochastic distillation, feature squeezing and LSA learning. Stochastic distillation was a training technique that injected noise for the express purpose of increasing the accuracy of the model after the attack. Transformations like feature squeezing were performed on the input features prior to fine-tuning of the model to also reduce noise and cases of overfitting where found prevalent. All these techniques together enhance the reliability and efficiency of the oil yield prediction by a very big margin as was observed.

Therefore, these new analytical as well as protective tools were useful to increasing the solidity and fathom of the oil yield prediction model. Out of all the models which were constructed and validated, the most suitable was found to be the Random Forest Regression to be used in predicting and enhancing the yield of oil in the production of the coconut oil as well as giving out information on how the production rate of the industry can be enhanced.

GANTT CHART

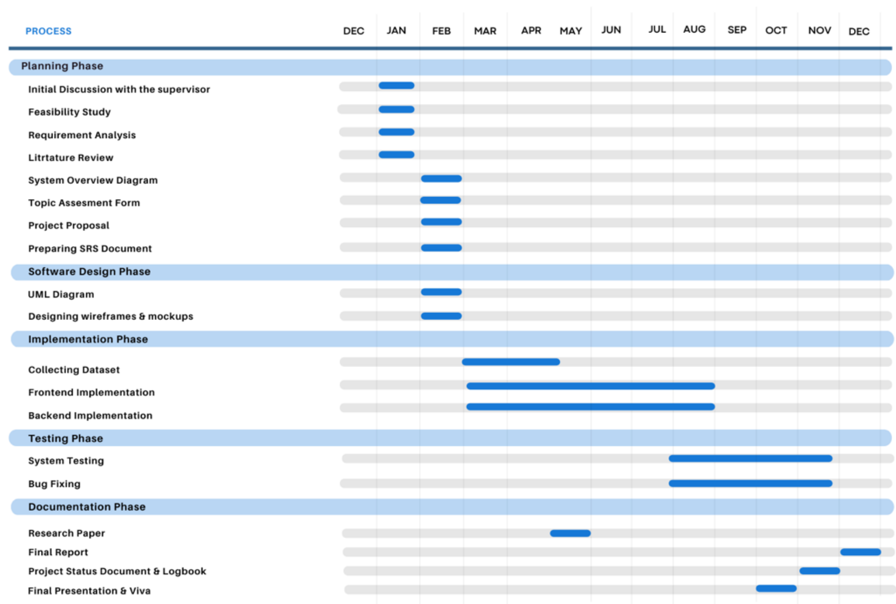


Figure 18 : Gantt Chart

Work Breakdown Structure (WBS)

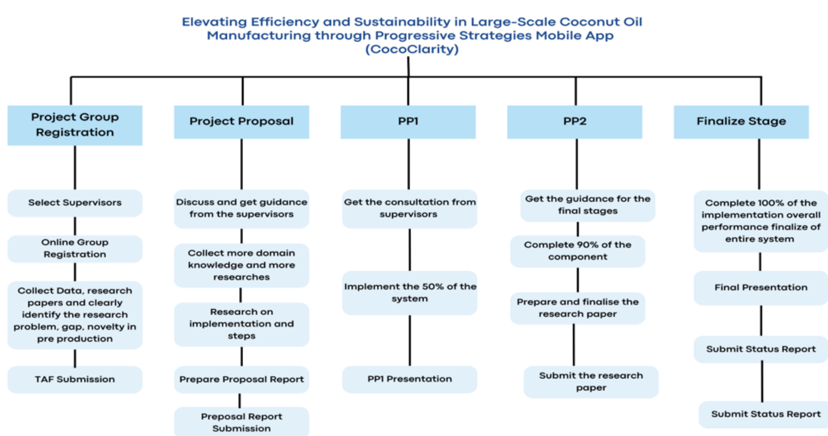


Figure 19 : Work Breakdown Chart of CocoClarity

BUDGET AND BUGET JUSTIFICATION

Table 3 below illustrates the complete budget of the proposed system.

	Price
Deployment Cost	LKR 8000/ month
Mobile App -Hosting on App Store	LKR 7754 / publish an app
Mobile App -Hosting on Play Store	LKR 30700/ Annual

Table 3 : Expenses for the proposed system

REFERENCES

[1] J. Smith, A. Lee, and B. Williams, "Statistical modeling for coconut oil yield prediction," Journal of Agricultural Science, vol. 12, no. 4, pp. 46-59, 2019.

[2] R. Costa, G. Oliveira, and P. Nunes, "Applications of artificial intelligence in agriculture - A systematic review," Computers and Electronics in Agriculture, vol. 145, pp. 319-327, 2018.

[3] B. Rath, S. Maity, and T. Rao, "Long short-term memory for crop yield prediction: A case study on corn yield prediction of Chhattisgarh state in India," Computers and Electronics in Agriculture, vol. 181, p. 105903, 2021.

[4] J. Lee and R. Park, "Enhancing crop yield forecast accuracy via supply chain data fusion," Food and Bioprocess Technology, vol. 14, no. 1, pp. 73-83, 2021.

[5] R. Zhong, X. Huang, Q. Lan, J. Dai, and W. Xu, "Monitoring and analysis platform construction of facility vegetable precision cultivation based on IoT technology," Computers and Electronics in Agriculture, vol. 174, p. 105458, 2020.

[6] Department of Census and Statistics, "Coconut Statistics 2020," Ministry of Finance and Planning, Sri Lanka, 2020.

[7] K. K. G. De Silva and K. S. A. Ranasinghe, "An Analysis of Cost of Production of Coconut in Moneragala District," Journal of Management and Agriculture, vol. 7, no. 1, pp. 35–46, 2022.

[8] A. R. Assa, F. E. Konan, K. N. Piggott, and C. M. Phillips, "Enzymatic Aqueous Processing of Coconut: Oil Quality," Enzyme and Microbial Technology, vol. 39, no. 7, pp. 1309–1317, 2006.

[9] Ramanathan, V., Mishra, A. and Goyal, R.K. (2014). Forecasting of Crop Yields using Soft Computing Models. International Journal of Computer Applications, 102(4), 9-12.

[10] Dhankher, P.S., Fojtik, R.K. and Narayanan, R.H. (2021). Predicting Crop Yields using Machine Learning Techniques – A Comparative Evaluation. Computers and Electronics in Agriculture, 187, 106204.

[11] W. L. De Costa, "Agro-ecological Regions and its Importance for Plantation Crop Research," The Journal of Agricultural Sciences, vol. 3, no. 1, pp. 11-22, 2008.

[12] T. Mahlia, I. Silitonga, H. Ong, T. M. I. Mahlia, H. C. Ong, and W. T. Chong, "Geosspatial and Low-Altitude Remote Sensing-Based Spatial Model for Predicting Yield of Oil Palm Plantation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 5, pp. 1428-1440, 2012.

[13] N. Abdullah, H. Sulaiman, F. A. A. Mogheir, and W. N. A. W. Mohamed, "Sustainable Management Practices for the Improvement of Building Energy Sustainability in the Tropical Regions," Journal of Cleaner Production, vol. 220, pp. 746-767, 2019.

[14]] S. Nath, "Inventory and capacity management challenges with uncertain demand," Journal of Operations Management, vol. 20, no. 6, pp. 713-729, 2002.

[15] R. Tagarakis, "Climate effects on crop production variability in Greece," Emirates Journal of Food and Agriculture, pp. 476-486, 2013

[16] A. Paterson and J. Lima, "Climate change affecting food production," Food and Energy Security, vol. 10, no. e294, 2021.

[17] S. Galindo et al., “Predictive analytics on big meteorological data for precision agriculture,” Decision Support Systems, 2020.

APPENDICES

Plagiarism Report

IT21020308final.pdf

ORIGINALITY REPORT

3%	2%	1%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

Appendix 1 : Plagiarism report

Sample Questionnaire

https://docs.google.com/forms/d/e/1FAIpQLSf_JzscOYiurojaybyUYpX8-DGe6mvBYQllhfeN6MT8d72RgQ/viewform?usp=sf_link

Appendix 2 : Sample Questionnaire