

**CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD**



**Implementation and Assessment of
Machine Learning Techniques to
Forecast Palm Oil Imports to Pakistan**

by

Muhammad Shoaib Javed

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the
Faculty of Engineering
Department of Mechanical Engineering

2025

Copyright © 2025 by Muhammad Shoaib Javed

All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, by any information storage and retrieval system without the prior written permission of the author.

Dedicated to My Parents and Family



CERTIFICATE OF APPROVAL

Implementation and Assessment of Machine Learning Techniques to Forecast Palm Oil Imports to Pakistan

by

Muhammad Shoaib Javed

(MEM213018)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Usama Waleed Qazi	AU, Islamabad
(b)	Internal Examiner	Dr. Shummaila Rasheed	CUST, Islamabad
(c)	Supervisor	Dr. Ghulam Asghar	CUST, Islamabad

Dr. Ghulam Asghar
Thesis Supervisor
September, 2025

Dr. M. Mahabat Khan
Head
Dept. of Mechanical Engineering
September, 2025

Dr. Imtiaz Ahmad Taj
Dean
Faculty of Engineering
September, 2025

Author's Declaration

I, **Muhammad Shoaib Javed Khan** hereby state that my MS thesis titled “**Implementation and Assessment of Machine Learning Techniques to Forecast Palm Oil Imports to Pakistan**” is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

Shoaib!

(**Muhammad Shoaib Javed Khan**)

Registration No: MEM213018

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled **Implementation and Assessment of Machine Learning Techniques to Forecast Palm Oil Imports to Pakistan**” is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

Shoaib!

(Muhammad Shoaib Javed khan)

Registration No:MEM213018

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Dr. Ghulam Asghar, for his invaluable guidance, continuous support, and insightful feedback throughout my research journey. His expertise, patience, and encouragement were instrumental in shaping this work.

I would also like to extend my sincere thanks to my thesis committee members for their constructive suggestions and thoughtful critiques that greatly improved the quality of this thesis.

My appreciation goes to the faculty and staff of the Engineering Management Department at Capital University of Science and Technology for providing an excellent academic environment and resources essential for my research.

I am deeply grateful to my colleagues for their camaraderie, stimulating discussions, and assistance during challenging phases of this work. Special thanks to my research fellows for their unwavering support and motivation.

I owe immense gratitude to my family especially my parents and my siblings for their endless love, encouragement, and sacrifices that enabled me to pursue higher education. Their belief in me has been my greatest strength.

This journey would not have been the same without the contributions of all these individuals and institutions, and I am forever thankful.

(Muhammad Shoaib Javed)

Abstract

A robust forecasting of palm oil is essential for ensuring the food security and stability to meet demand and formulation of effective trade policies in Pakistan. Formerly, the traditional econometric models were used to predict the complex market dynamics like exogenous variables dependency including price, demand and supply factors and those models were least effective for time series data. Therefore, this work is inclined towards the evaluation of the predictive potentials of the machine and deep learning models to predict future palm oil imports in terms of its quantities and prices. In this work, a comprehensive twelve years dataset (obtained from Pakistan Bureau of Statistics) of the historical palm oil imports (quantity and price) is preprocessed, which is then used to train and test the four predictive models; Random Forest (RF), Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and Autoregressive Integrated Moving Average (ARIMA). In this dataset, the palm oil quantity and price are use as both input and output parameters. The evaluation of these models is done by performing extensive simulations considering the standard performance metrics including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), R-square (R^2), accuracy, and precision with the dataset split rate of 70% for training and 30% for testing. Although CNN is mostly used for the image analysis in literature but it was also used for the time series dataset analysis. Among all the employed techniques, the CNN exhibits the superior performance and attains the highest accuracy and predictive stability. The MAPE and RMSE values of the CNN for quantity and price are 0.09 and 0.07, 32640 and 182.9, respectively, which are the least among all the tested models. The ANN is ranked second, as it obtains more quantity and price error values followed by the RF. Similarly, the ARIMA shows comparatively limited adaptability to the nonlinear patterns as it yields the highest error values among the all models. The CNN also shows the highest values of the R- square, accuracy, and precision, emerging as is the best suited model for the palm oil data. After conducting the detailed simulations, the forecasted values of the palm oil consumption (quantity) and price are obtained for the next three years, which represent that the CNN outperforms RF, ANN, and ARIMA models.

The results validate the effectiveness of the CNN for time-series prediction in trade analytics. Therefore, the CNN model is deemed efficient and reliable tool for the accurate forecasting of palm oil imports to Pakistan.

Keywords: Machine Learning; Deep Learning; RF; ANN; CNN; ARIMA; Forecasting; Palm Oil

Contents

Author’s Declaration	iv
Plagiarism Undertaking	v
Acknowledgement	vi
Abstract	vii
List of Figures	xiii
List of Tables	xiv
Abbreviations	xv
Symbols	xvii
1 Introduction	1
1.1 Background of the Palm Oil	1
1.1.1 Forecasting Techniques	3
1.2 Motivation of the Study	7
1.3 Problem Statement	9
1.3.1 Research Questions	10
1.4 Objectives of the Work	10
1.5 Significance of Palm Oil Forecasting	10
1.6 Scope of the Study	12
1.7 Thesis Organization	13
2 Literature Review	14
2.1 Overview of the Chapter	14
2.2 Usage of Palm Oil	15
2.3 Current Status of Palm Oil Production	16
2.3.1 Fundamentals Aspects of Crop Yield Prediction Process	19
2.4 Significant Features Usage for Crop Yield Prediction	19
2.5 Forecasting Models	21
2.5.1 Traditional Models for Forecasting	21
2.5.1.1 Kernel Naive Bayes	21

2.5.1.2	Random Forest	22
2.5.1.3	Support Vector Machine	24
2.5.1.4	AutoRegressive Integrated Moving Average Model	25
2.5.1.5	Support Vector Regression	27
2.5.1.6	Genetic Algorithm	28
2.5.1.7	Linear Regression	29
2.5.1.8	Extra Tree	30
2.5.1.9	Latest Machine Learning Techniques for Forecasting	31
2.5.1.10	Convolutional Neural Networks	31
2.5.1.11	Long Short-Term Memory	34
2.5.1.12	Artificial Neural Network	36
2.5.1.13	Genetic Algorithm Neural Networks	37
2.5.1.14	Recurrent Neural Networks (RNN)	39
2.5.1.15	Gated Recurrent Unit	40
2.6	Hybrid Models for Enhanced Forecasting Accuracy	41
2.7	Performance Evaluation Metrics	42
2.8	Research Gap	44
2.9	Summary	46
3	Proposed System Model	47
3.1	Introduction	47
3.2	Fractions of Palm Oil and Their Hierarchy	47
3.2.1	Palm Olein	48
3.2.2	Palm Stearin	48
3.3	Secondary and Specialized Fractions	48
3.3.1	Super Olein	48
3.3.2	Hard Stearin	48
3.3.3	Palm Mid Fraction (PMF)	49
3.4	Hierarchical Structure of Palm Oil Fractions	49
3.5	Sample Data of Palm Oil and Price Fractions in Pakistan (2013–2024)	50
3.5.1	Shift from CPO to RBD Products	50
3.5.2	Increasing Demand for RBD Palm Olein	52
3.5.3	Steady Growth in RBD Palm Stearin Imports	52
3.5.4	Uses of Fractions of Palm Oil	52
3.6	Proposed System Model	53
3.6.1	Random Forest	59
3.6.1.1	Bootstrap Sampling Bagging	59
3.6.1.2	Feature Randomness	60
3.6.1.3	Decision Tree Construction	60
3.6.1.4	Prediction and Aggregation	60
3.6.1.5	Model Evaluation	60
3.6.2	Convolutional Neural Network	61
3.6.2.1	Data Preprocessing	61
3.6.2.2	Convolutional Layers	61
3.6.2.3	Pooling Layers	61

3.6.2.4	Activation Functions	61
3.6.2.5	Fully Connected Layers	62
3.6.2.6	Output Layer	62
3.6.2.7	Training and Evaluation	62
3.6.2.8	Model Evaluation	62
3.6.3	Artificial Neural Network	63
3.6.3.1	Problem definition and Data Collection	63
3.6.3.2	Data Collection	63
3.6.3.3	Data Preparation for the Training and Testing	63
3.6.3.4	Data Cleaning	64
3.6.3.5	Data Transformation	64
3.6.3.6	Data Splitting	64
3.6.3.7	Artificial Neural Network Architecture Design	64
3.6.3.8	Layers	64
3.6.3.9	Neurons	64
3.6.3.10	Activation Functions	64
3.6.3.11	Optimizer:	65
3.6.3.12	Loss Function:	65
3.6.3.13	Initialization of the Parameters	65
3.6.3.14	Implementation of the Forward and Backward Propagation	65
3.6.3.15	Training the Neural Network	65
3.6.3.16	Evaluation of the Neural Network	65
3.6.4	ARIMA	66
3.6.4.1	AR	66
3.6.4.2	I	67
3.6.4.3	MA	67
3.6.5	Working of the Model	67
3.6.5.1	Data Preparation	67
3.6.5.2	Model Selection	67
3.6.5.3	Model Estimation	67
3.6.5.4	Forecasting	67
3.7	Model Evaluation	68
3.7.1	RMSE	68
3.7.2	MAPE	68
3.7.3	R^2	69
3.7.4	Accuracy	69
3.7.5	Precision	69
3.7.6	Oil quantity and Price Prediction	69
3.8	Tools and Software	70
3.9	Summary	70
4	Results and Discussion	71
4.1	Introduction	71
4.2	Data Set of Oil Quantity and Price	71

4.3	Analysis of Twelve Years' Dataset	72
4.4	Statistical Analysis using 2-Point Moving Average Test	74
4.5	Results of Machine Learning and Deep Learning Techniques	74
4.6	Future Predictions of Oil Quantity and Price by All Techniques	80
5	Conclusions and Future Recommendations	85
5.1	Conclusions	85
5.2	Implications/Limitations of Work	86
5.3	Future Recommendations	87
5.4	Future Directions	88
	Bibliography	89
	Python Code of Thesis	105

List of Figures

2.1	Consumption of Vegetable Oils from 2013/14 - 2019/2020 by Oil Categories (in a million MT) Globally by Rashid et al. [26]	16
2.2	High Scale Vegetable Oils Production from 2012 - 2020 Approximately in Categories (million MTs) globally by Rashid et al. [26]	17
2.3	Major Palm Oil Producers and Exporters by Rashid et al. [26]	17
2.4	Global Palm oil Consumption from 2015 - 2020 (in 1,000 MTs) Approximately by Rashid et al. [26]	18
2.5	Crop Yield Prediction by Machine Learning Approaches: A General Architecture by Rashid et al. [26]	19
2.6	Long Short-Term Memory Architecture by Sagheer and Kotb [89]	34
2.7	Popular Performance Evaluation Metrics for Crop Yield Prediction algorithms by Rashid et al. [26]	44
3.1	Hierarchy of Palm Oil Fractions	49
3.2	Proposed System Model	54
4.1	Decomposition of Palm Oil Quantity for Observed, Trending, Seasonal and Residual Data	73
4.2	Decomposition of Palm Oil Price for Observed, Trending, Seasonal and Residual Data	73
4.3	Two-point Moving Average for Palm Oil Imports with Actual Curve	74
4.4	Two-point Moving Average for Palm Oil Imports Price with Actual Curve	75
4.5	MAPE Results Palm Oil Import Quantity and Price of all four Techniques	76
4.6	RMSE Results of Oil Quantity and Price of all Four Techniques Curve	77
4.7	R^2 Results of all Four Techniques for Palm Oil Imports Quantity and Price	78
4.8	Accuracy Results of all Four Techniques for Palm Oil Imports Quantity and Price	79
4.9	Precision Results of all Four Techniques for Palm Oil Imports Quantity and Price	80
4.10	Prediction of Next Three Years' Results for Oil Quantity of all Four Techniques	81
4.11	Prediction of Next Three Years' Results for Oil Price of all Four Techniques	81

List of Tables

2.1	Summarized Table for Hyperparameter Tuning for Comparison using the Capacities and Limitations	43
2.2	Frequency of Techniques used in the Literature	45
3.1	Import Volumes of Palm Oil (MT) and its Fractions over the Past Decade 2013 - 2024	51
3.2	Import Volumes of Palm Oil Price (USD)and its Fractions over the Past Decade 2013 - 2024	51
3.3	Import Volumes of Palm Oil Demand and Price, and its Fractions over the Past Decade 2013 - 2024 on Monthly Basis	54
3.4	Hyper Parameters of RF	60
3.5	Hyper Parameters of CNN	62
3.6	Hyper Parameters of ANN	66
3.7	Hyperparameters of ARIMA Model	68
4.1	Future Predictions of Oil Quantity by All Techniques	82
4.2	Future Predictions of Oil Price by All Techniques	83

Abbreviations

ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
AR	Autoregression
Bi-LSTM	Bidirectional LSTM
BSR	Basal Stem Rot
BPNN	Backpropagation Neural Network
CNN	Convolutional Neural Network
CPO	Crude Palm Oil
GRU	Gated Recurrent Units
GDP	Gross Domestic Product
GA	Genetic Algorithm
GANN	Genetic Algorithm Neural Networks
HOG	Histogram of Oriented Gradient
KNB	Kernel Naïve Bayes
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
LR	Linear Regression
MA	Moving Averages
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NLP	Natural Language Processing
PKO	Palm Kernel Oil
PMF	Palm Mid Fraction
RF	Random Forest

RMSE	Root Mean Square Error
RSPO	Roundtable on Sustainable Palm Oil
RNN	Recurrent Neural Networks
RBD	Refined, Bleached and Deodorized
SVM	Support Vector Machine
SVR	Support Vector Regression
SARIMA	Seasonal ARIMA
SA-SVR	Simulated Annealing with SVR
SBP	State Bank of Pakistan
UAV	Unmanned Aerial Vehicles

Symbols

<i>n_estimators</i>	defines the number of trees
<i>random_state</i>	acts as a seed of reproducibility
<i>max_depth</i>	specifies the maximum depth of every tree
<i>min_sample_split</i>	minimum sample required to split a tree
<i>n_Conv_layer</i>	convolutional layers used for feature extraction
<i>kernal_size</i>	filters are used for learnable features detection
<i>max_pooling</i>	used for explicit down-sampling
<i>early_stopping</i>	used for stopping the overfitting
<i>p</i>	number of lagged observations
<i>q</i>	number of lagged forecast errors
<i>d</i>	degree of differencing to obtain stationarity
\hat{y}_i	predicted values
<i>N</i>	total number of observations
<i>y_i</i>	observed values
R^2	R-Squared values

Chapter 1

Introduction

Predicting the imports of palm oil is based on the analysis of trends in price, demand, and trade and application of such techniques as time-series analysis and assessment of important market factors including population growth, food and industrial use, and sustainability issues. According to recent forecasts, palm oil imports in Pakistan may increase by approximately 3 percent by 2025/26 as a result of population growth whereas the global market is expected to increase by 5.30 per cent and reach 98.90\$ billion by 2030. This chapter contains the details of background of palm oil, forecasting techniques, significance, motivation and challenges faced by already employed importing approaches. Objectives and thesis organization are also discussed here.

1.1 Background of the Palm Oil

Palm oil is a globally significant edible oil due to its low production cost, high yield, and versatility. It is primarily extracted from the palm tree, also known as *Elaeis guineensis*. There are two types of oil based on the parts of the tree from which they are extracted. One is crude palm oil (CPO) extracted from the mesocarp of the fruit, and the second one is palm kernel oil (PKO) extricated from the seed kernel by Gourichon H. [1]. Both types of oil have nutritional and commercial importance. They are extensively used in cosmetics, food products, biofuels, and pharmaceuticals, which makes them vital for numerous economies.

The leading producers are Malaysia and Indonesia. Malaysia alone exports this resource to more than 150 countries. Its competitive price and widespread use compared to other oils contribute to its increasing demand globally.

The importance of palm oil in international trade has increased due to the growing demand, especially for under-developing countries that mainly rely on imports. In order to meet domestic demand, nations like Pakistan that produce insufficient amounts of edible oil, import palm oil from other countries like Malaysia and Indonesia. Malaysia is the main source, supplying over 75% of Pakistan's edible oil imports, which comprise over 90% of palm oil by Zahraee et al. [2].

Because of this reliance, development of accurate forecasting techniques for imports of palm oil is essential to facilitate the decision-makers. Effective forecasting and strategic import planning are hampered by Pakistan's agricultural sectors' lack of data-driven planning and technological gap as discussed by Yee and Samsudin [3]. Based on these observations, a variety of forecasting models have been employed all over the world to predict changes in the palm oil market with regard to both demand and price. The prices of palm oil in Thailand have been accurately forecasted by Auto-Regressive Integrated Moving Average (ARIMA)-based models for a number of categories, including farm, wholesale, and pure oil prices. Because the models' Mean Absolute Percentage Error (MAPE) values were low, it demonstrates excellent forecasting efficacy.

Classical methods like ARIMA model find it difficult to explain the complex, nonlinear dynamics arising from the various socioeconomic and environmental elements applied to the dataset. These limitations highlight the necessity of utilizing cutting-edge technologies (such as Deep Learning and machine learning), which are excellent at figuring out the complicated patterns present in the massive multi-dimensional data by Tambe et al. [4].

To maximize forecasting accuracy and precision, attention has recently switched to the introduction of models such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and so on. According to Yee and Samsudin [3], ANN produced predictions of Malaysian palm oil prices that were more accurate than ARIMA. Based on monthly data from 2008 to 2018, it was evident that ANN's

data-driven methodology outperformed conventional techniques in identifying the intricate nonlinear swings in palm oil prices.

Research by Tambe et al. [4] and Ahmad et al. [5]; found that using some climate smart agricultural assessments and statistical models yeild useful results in making predictions because they could detect important patterns in time-series data even with minimal preparation. They have more power to detect the underlying complex trends in data because of their inherent intelligent nature.

1.1.1 Forecasting Techniques

There are following forecasting techniques studied from the literature as mentioned in this section. Convolutional layers are used by Convolutional Neural Network (CNNs) to automatically extract spatial information from data. CNNs are used to appropriately evaluate complex relationships between temporal and environmental data in the context of forecasting. CNNs were utilized by Khan et al. [6] to accurately forecast oil palm yield in the face of shifting soil moisture and meteorological conditions. The model showed the relationship between rainfall, temperature, and other variables, as well as spatial and time patterns.

When it comes to yield prediction, the CNN performed the best, demonstrating its resilience with regard to complex agro-environmental variables. The methodology's practical utility may be further demonstrated by its potential applicability to various crops and geographical areas. In addition, CNNs and its hybrids models are also used for time-series data forecasting as discussed by Sadouk, L. [7] for the data-level and algorithm-level approaches.

Livieris et al. [8] used the hybrid LSTM and CNN model for the precise forecasting of the time series data of gold prices and its evoultion. Jin et al. [9] and Wibawa et al. [10] also used CNN model for the time-series data forecasting for monitoring the multidimensional variables relationships and also for smoothing the time-series data.

In order to improve accuracy and reduce overfitting, Random Forest (RF) combines the learning of many decision trees throughout the training phase, providing a mean output from each tree. Bi et al. [11] looked at RF and Linear Regression to

see which would be more effective in forecasting petroleum prices. They discovered that RF models outperformed traditional linear techniques in detecting complex non-linear correlations when they broke down historical pricing data. When it came to predicting petroleum prices in volatile market conditions, the RF model outperformed the others due to its increased accuracy and stability. The RF's outcomes show how effective it is as an ensemble approach that can improve forecasts in fields connected to energy and economics. The analysis demonstrates RF's effectiveness in contexts where precise forecasts are necessary in various market conditions.

With the ability to tackle complicated non-linearity in data across a network of interconnected nodes (neurons), ANNs are computer models developed on the biological neural networks, or the human brain. Hermantoro and Rudyanto [12] modelled the production of palm oil by climate and land quality parameters using an artificial neural network approach. The system performed well in simulating non-linear relationships between soil characteristics, precipitation, and harvest yield. The yield forecast predictions from the model, which was created using data from prior planting operations, were quite dependable. Their results demonstrated the value of ANN in agricultural production, where environmental fluctuations are highly variable. Decision support for plantation design and resource utilization improvement was provided by the simulation.

With its auto-regressive, differencing, and moving average components, ARIMA is a statistical model of analysis built using time series data to better understanding the data sets or predicting future trends. Chamawala et al. [13] used the ARIMA model to predict crude oil prices in Nigeria between 2013 and 2022.

The study's goals were to determine the best ARIMA parameters based on the AIC and to perform diagnostic checks for the model's reliability. The results showed that ARIMA could accurately forecast in the coming days, giving a good degree of success in capturing the trend and seasonality in pricing some products. The model was found suitable for economic planning in oil-dependent economies, highlighting ARIMA's strengths in modeling historical time-series data for sound policy and investment decisions.

When historical data influences future projections, LSTM Networks—recurrent neural networks—are suitable models for time series forecasting because they are best suited for long-term dependencies. The issue of LSTM performance in relation to ARIMA and Prophet in the context of oil production forecasting was examined by Niu and Li [14]. In terms of accuracy, the study showed that LSTM performed better than the other models in identifying intricate non-linearities and temporal nuances. One of LSTM's strengths that helped the company succeed in oil production prediction was its capacity to learn from a wide variety of data sources. The study demonstrated how well LSTM handles complex, trended, and seasonally shifting time-series data, making it essential for energy forecasting.

As a supervised learning scheme, SVM is well-suited to process data for regression analysis and classification, particularly in high-dimensional data sets. Using machine learning approaches and SVM models, Dimitriadou et al. [15] investigated the oil markets' efficiency. They demonstrated that the SVM could effectively forecast trends in the oil market by using patterns from the previous data. The model produced information on market efficiency and volatility and did a good job of differentiating across market situations. The SVMs' ability to handle intricate and non-linear data connections allows it to do this task.

The Support Vector Regression (SVR) uses regression problems. The foundation of SVR is the mapping of input data into a higher space and the identification of a hyper-plane that best divides this space according to the data. This model contains a quite low error margin. Hosseini and Akilan [16] predicted oil production time series using cutting-edge deep regression models like SVR. It was discovered that, when applying the non-linear relationships, SVR and other models produced accurate predictions. The SVR's strong capacity to handle high-dimensional features and reduce error made it a highly successful tool for oil production forecasting. Their results demonstrated that the SVR can effectively forecast the time series data for the energy sector. Another technique used in the literature is known as Extra Tree for learning.

Extra Trees is a collaborative learning technique that develops its prediction by combining the output of several de-correlated decision trees that have collected in a

forest. Prediction of crude oil production using a genetic algorithm in conjunction with an optimal gradient boosting model is discussed by Alkhamash E.H. [17]. The study demonstrated that optimization greatly increased the model predictive accuracy, even if Extra Trees were a component of the ensemble techniques that were investigated. The model confirmed the Extra Trees' ability to work with complex data and enhanced forecasts by incorporating it with other methods to account for refined, non-linear features in oil production data.

The linear relationship between two (simple Linear Regression (LR)) or more (multiple LR) variables is examined via the basic predictive analytical method known as linear regression. Because it is simple to use and the results are straightforward to comprehend, it is a suitable choice for predictive modeling. Bi et al. [11] used RF and LR to predict petroleum prices.

They came to the conclusion that while linear regression provided an easy-to-understand solution, it was insufficient to handle the intricate swings in the price of oil. The model's reliance on a linear assumption, which reduced its projections to a linear estimation, was a significant flaw.

The fact that RF outperformed LR demonstrates the clear challenges of using linear models to suit the dynamic, non-linear nature of petroleum markets. However, it served as a helpful benchmark against which the usefulness of more complex models could be evaluated. Moreover, based on these approaches, Genetic Algorithm (GA) is discussed for efficient optimization of the tasks. Effectively refined by the application of natural selection and genetic principles, GA function as method for tackling complex optimization problems, improving possible solutions over many cycles. Using genetic algorithms to tune hyper-parameters, Alkhamash E. H. [17] created an optimal gradient boosting model to forecast the output of crude oil. By efficiently scanning the space of parameters, GA greatly increased the model's ability to make accurate predictions. The article demonstrated that how an evolutionary technique of GA enables sophisticated machine learning models to accurately identify and predict oil production trends. The research enhanced the system into a more reliable forecasting workbench by including the GA into Gradient Boosting, asserting that it outperformed traditional

techniques when dealing with noisy data. The ANN performs better in identifying and handling intricate and nonlinear data relationships when weights, biases, or network architecture are optimized using genetic algorithms (GA). A Genetic Algorithm Neural Networks (GANN) framework was used by Tardini G. A. [18] to regulate emissions in the production of palm oil.

They demonstrated how the generalization and optimization of the model on scenarios involving environmental information were significantly improved by combining neural networks and evolutionary algorithms. Although the model's framework may be used to predict palm oil production while taking into account environmental factors like temperature, humidity, and input biomass on oil yield, the main objective was the emission management. This method shows how GANNs may effectively capture intricate relationships in the manufacturing and processing of palm oil. In order to better handle complicated and non-linear data, Kernel Naïve Bayes (KNB) expands on the classic Naïve Bayes technique by employing kernel density estimation. It is especially helpful when classification problems involve non-Gaussian distributions. The work by Husin et al. [19] classified the BSR severity on oil palm trees using the KNB technique utilizing data from terrestrial laser scanning. They used Principal Component Analysis to compress their data and identify five structural features from 80 trees. The KNB was trained using the model's top two components and showed strong classification power. It was able to get a Kappa coefficient of 0.80 and overall accuracy of 85%. Its ability to accurately classify all samples that were determined to have early-stage (T1) infection rates was one of its most notable features. According to the findings, the KNB holds significant promise for early disease detection in precision agriculture, including BSR. Some prediction based approaches are discussed by Butt et al. [20], Tahyudin et al. [21] and Khan et al. [22].

1.2 Motivation of the Study

One of the most popular vegetable oils is palm oil, which has a variety of uses in the creation of food, cosmetics, and biofuels. Because of the expanding population and growing need for edible oils. For companies, legislators, and other players in the trade and agricultural sectors, accurate palm oil import forecasting is essential.

To minimize economic growth, ease of the supply chain management, and lower market volatility, accurate forecasting is essential.

Traditional statistical techniques including time series models like ARIMA and RF have been widely employed for agricultural import forecasting. These models might not be able to adequately represent the complex nonlinear patterns found in the data based on the seasonality, elasticity of price, interaction of demand and supply mechanisms and trade policies factors. By effectively managing large amounts of data and capturing nonlinear correlations, machine learning and deep learning techniques like SVM and ANN have completely changed forecasting capabilities. Even though nations like Malaysia and Thailand have made significant efforts to develop cutting-edge forecasting techniques, Pakistan does not frequently adopt these developments. There is a notable dearth of studies on forecasting imports for emerging nations like Pakistan, which rely largely on imports, similar to the majority of studies on pricing patterns and domestic consumption. Since Pakistan's current forecasting systems are unable to account for the nations' import and consumption patterns, there is an obvious need for customized forecasting solutions.

By developing contemporary deep learning and machine learning approaches to anticipate Pakistani palm oil imports and pricing, this research seeks to close the forecasting gap and assist stakeholders in developing policies and market strategies. To determine the most effective way for predicting palm oil in Pakistan, this study will compare the forecasting powers of modern techniques (ANN, CNN, etc.) with more conventional approaches (ARIMA, etc.). The best forecasting model will be determined with the aid of this study, allowing for well-informed decision-making for national planning and policy development. By analyzing large amounts of historical data, the study would attempt to accurately forecast future patterns, better comprehend import trends, and assess economic implications. Additionally, this study will investigate how statistical models and artificial intelligence might be tailored to address the unique difficulties faced by developing economies.

The current study assumes that by lowering the uncertainty around import prices and volumes, better forecasting can significantly contribute to Pakistan's economic stability. Pakistan may negotiate better trade terms, manage buffer stocks more

skillfully, and lessen its exposure to erratic market swings with better forecasting. By minimizing shortages and overstocking, accurate forecasting can assist maintain a steady supply across the country and prevent financial losses. The findings of the study will have a big impact on food security and global economic resilience, which will help in minimizing the effects of pandemics and geopolitical instability.

To sum up, the study bridges the forecasting gap for palm oil imports and creates a hybrid forecasting model by fusing traditional and sophisticated methods, enabling traders, policymakers, and industry professionals to provide a better insight for palm oil imports. It is anticipated that using machine learning and deep learning in forecasting will improve strategic planning for food security and economic sustainability as well as produce more accurate predictions.

1.3 Problem Statement

Palm oil production plays a vital role in the agricultural and commercial sectors globally. In order to obtain the improved agricultural output, such as in production, transportation, management of demand and supply, maintenance of price elasticity, commerce, and policy development decision-making, accurate palm oil import forecasting is very crucial. It is challenging due to a number of influencing factors, such as fluctuating prices, changing market demands, trade policies, and meteorological circumstances. Despite the capacity of traditional approaches like ARIMA and others for prediction, those may frequently fall short in capturing the intricate nonlinear patterns that define agricultural import data. Formerly, CNNs were mostly applied for limited time-series data forecasting and now it is necessary to apply them in recent work for making their effective utilization. As machine learning uses advanced, models such as CNN, ANN, and many other models which became useful instruments for managing large datasets, identifying interactions, and capturing subtle patterns that conventional approaches lack. These models are effective in handling complex forecasting problems because of their capacity to learn from huge datasets, which enhances their flexibility. In light of these factors, the goal of this work is to thoroughly compare and evaluate the most recent machine learning approaches (CNN, ANN) and conventional time series models (RF,

ARIMA) in order to estimate imports of palm oil by considering the price and demand. In addition, this study seeks to identify the most accurate and successful forecasting model, offering insightful data for policymaking as well as helpful for traders and other stakeholders.

1.3.1 Research Questions

- What are the limitations of traditional forecasting methods i.e., ARIMA in predicting palm oil imports in Pakistan?
- How do machine learning and deep learning models such as RF, CNN, and ANN perform compared to traditional models?
- What are the strategic implications of the current research work for policymakers and industry stakeholders?

1.4 Objectives of the Work

The primary objectives of this research are:

- To analyze the performance of traditional statistical models ARIMA and its variants in forecasting palm oil imports to Pakistan.
- To explore the application of machine learning and deep learning techniques for palm oil and price forecasting
- To compare the predictive accuracy of traditional and proposed techniques.
- To provide recommendations for stakeholders based on the cumulative findings of the study for effective policy making.

1.5 Significance of Palm Oil Forecasting

One of the most widely consumed edible oils in the world is palm oil, which is in high demand due to major importers like the US, China, India, Bangladesh,

Pakistan, and the European Union, all of which depend on palm oil to meet their edible oil needs and support a variety of industries. Predicting palm oil imports accurately is essential in maintaining economic stability. Inflation, trade balances, and the affordability of food can all be significantly impacted by the changes in global demand and prices. It goes without saying that accurate forecasts help stakeholders better prepare for changes in the market, respond logically to buying periods, and guard against the disastrous effects of price swings and supply interruptions by Suppalakpanya et al. [23]. If businesses can create accurate import forecasts, the supply chain operations can be drastically improved. Palm oil producers can improve overall operational efficiency, decrease surplus and shortage situations, and plan more effectively by having a better understanding of future demand. Enhancing the supply chains' efficiency with precise projections would save costs and boost dependability, benefiting sectors like biofuels, cosmetics, and food processing by Ata A. [24].

The implementation of precise projections gives decision-makers a solid foundation upon which to build well-informed trade laws and policies. Effective forecasting models can help determine future import requirements, which can then be used to inform trade policy choices on tariffs, quotas, and subsidies. This goal ensures that when governments meet market demands, they uphold national interests. Good forecasting gives governments the knowledge they need to negotiate trade agreements and respond adaptably to shifts in the global market by Khan et al. [22]. Furthermore, the environmental impact of palm oil's supply chain and commercialization is substantial. It is feasible to match import regulations with more general sustainability objectives by using data and sophisticated forecasting. With this strategy, governments may make choices that encourage palm oil to come from sustainable sources and lessen its impact on delicate ecosystems.

According to Tahyudin et al. [21], accurate forecasting is crucial for promoting sustainability and assisting stakeholders in adopting proactive measures to stay ahead of climate challenges and seize opportunities. Given these considerations, using cutting-edge machine learning techniques can result in substantial benefits. Modern machine learning techniques are able to reveal nonlinear and intricate patterns in historical and economic data that are not readily apparent in conventional

models. The growth of economic resilience, the implementation of efficient policies, and the sustainability of palm oil imports all depend on the stakeholders' ability to generate more precise and pertinent predictions as innovation is improved due to the curtesy of machine learning models.

1.6 Scope of the Study

With the ultimate goal of increasing the forecasting findings' accuracy, this study will try to illustrate Pakistan's imports of palm oil through the use and analysis of traditional statistical models and augmented machine learning approaches. The study examines Pakistan's imports of palm oil and associated economic data by utilizing and contrasting the outcomes of forecasting techniques such as ARIMA, CNN, ANN, and RF. Improving the forecasting accuracy and dependability of palm oil imports is the goal of integrating time series data with other pertinent external elements (such as consumption and price). Although there are many other forecasting models available, this study focuses on models that are notable for striking a balance between accuracy and interpretability and have garnered a lot of recognition due to their usefulness. The study uses CNN and RF in addition to the conventional ARIMA and deep learning models to investigate how machine learning methods can be applied to the modeling of complex feature interactions and non-linear relations. Models beyond the horizon of this study are excluded mostly due to their limited practical utility and difficulty in validating within the selected data set. This study, which focuses on Pakistan's imports of palm oil, offers a thorough and broadly applicable example that could spur additional investigation and the creation of policies for other goods or consumables. The study intends to provide high-quality information with a focus on Pakistan's imports of palm oil, without having to discuss the worldwide palm oil trade and production estimation in general. The changes in government regulations, the supply chain interruptions, or unanticipated market fluctuations can all have an impact on the amount of palm oil imported; nevertheless, these factors are acknowledged but purposely left out of the study's methodology. The results of the analysis of several forecast models should provide stakeholders (traders, policy makers, and agricultural economists) with useful advice on how to employ more accurate forecasts.

These findings will enable evidence-based decision-making to help stakeholders address palm oil import difficulties, lower uncertainty, and optimize their trade and import strategies.

1.7 Thesis Organization

The organization of this thesis is as follows:

- Chapter 1 discusses the introduction of the research topic, focusing on background, objectives, and scope.
- Chapter 2 explains the in depth literature review covering related studies on palm oil forecasting and its applications using machine learning approaches.
- Proposed methodology is included in Chapter 3 which explains the data sources, preprocessing techniques, and model implementation.
- Results and discussion are included in Chapter 4 which presents the findings and comparisons of models' performance.
- Finally, Chapter 5 contains the conclusions and recommendations which summarize the list of contributions and suggest future directions.

Chapter 2

Literature Review

2.1 Overview of the Chapter

This thesis' literature review chapter includes a detailed discussion of the main elements and ways of forecasting palm oil imports. The chapter starts with the discussion about the manifold uses of palm oil and massive influence on the world economy and industry manifold (Section 2.1). Forward, Section 2.2 discusses the present-day palm oil production situation: global and domestic trends; upcoming issues; and, opportunities for advancing cultivation and production of palm oil.

The major guiding principles of crop yield prediction forming the basis of calculations are discussed in section 2.3 of this chapter. This section discusses the fundamental steps that need to be taken in order to predict agricultural outputs. It delineates the most widely used elements in crop yield prediction models, including from climate and soil to even historical production records (Section 2.3.1). This section of the chapter delves into various algorithms of predictions with statistical models being included among them. Section 2.4.2 switch to a rigorous analysis of the machine learning techniques for forecasting, including the CNN, ANN, Recurrent Neural Networks (RNN), and Gated Recurrent Units (GRU).

The strengths, weaknesses, and performance benefits of these techniques are discussed with a focus on circumstance where they are superior to classical models. Section 2.5 examines hybrid models, which combine various forecasting strategies

to enhance precision and forecasts' predictability. The fact that these models utilize different algorithms enables them to take advantages of each other's strengths and provide a better result than the single-method approaches. Following this, Section 2.6 of the chapter examines performance evaluation metrics that are critical in specifying accuracy and efficacy of the forecasting model.

In this section, the importance of the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) as the factor of performance evaluation is examined thoroughly. Finally, the chapter enumerates significant shortcomings in the existing field to situate the objectives and research gap of the current research.

2.2 Usage of Palm Oil

In the previous decade, the plants of oil palm (*Elaeis guineensis*) are cultivated for its oil-rich fruit that has gained popularity in South-East Asia, especially in Malaysia, Indonesia and Thailand. Palm oil is a major contributor to the oil-yielding capacity of other vegetable oilseeds, such as rapeseed, sunflower, and soybean.

Plants are often the source of vegetable oils, which come in a variety of textures, including liquid, fatty, and oily. Because vegetable oils include high unsaturated fatty acids than animal fats, they are obtained as a healthier alternative when compared to animal fats for consumption as food products.

The majority of vegetable oils can be used as cooking oil or for the production of gasoline and diesel. The most widely used cooking oils are thought to be canola, soybean, sunflower seed, and palm oils. The flesh of the palm fruit, which is used to make palm oil, is mostly found in the tropical regions of Africa, South America, and South-East Asia by Chong et al. [25]. One of the most widely used vegetable oils in the world today is palm oil. About 10% of palm oil is utilized in industrial products including fuel, diesel, and cosmetics, while the remaining 90% is used for food by Rashid et al. [26], as visualized in Fig. 2.1.

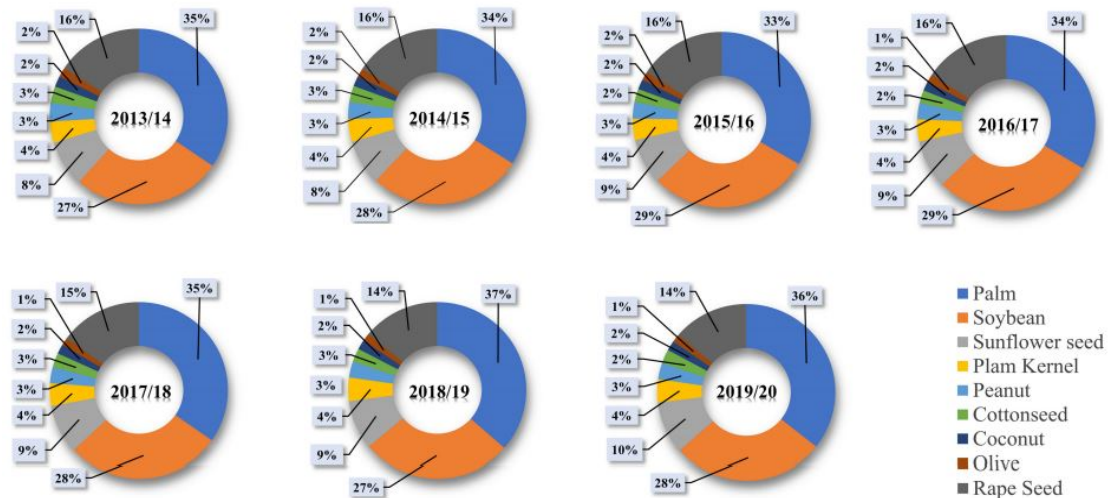


FIGURE 2.1: Consumption of Vegetable Oils from 2013/14 - 2019/2020 by Oil Categories (in a million MT) Globally by Rashid et al. [26]

2.3 Current Status of Palm Oil Production

Up to 200 million MTs of vegetable oil were produced globally during the 2019–2020 crop year. With 72.27 million metric tons of oil production, palm oil was the most abundant vegetable oil around the globe at that time, making it a valuable commodity (Fig. 2.2).

Palm oil is a common oil found in processed foods and consumer products. The oil palm tree yields palm fruits, from which the oils are subsequently extracted. The land used for cultivation of palm oil is quite productive. Palm oil yields significantly more than other types of oil crops and requires a small quantity of land.

Over the past few years, the volume of palm oil produced has been steadily increasing at a fast rate. As a result, there are sizable and active palm oil export markets throughout the world, with Malaysia and Indonesia being the biggest exporters. The main palm oil producers and exporters are shown in Fig. 2.3. Figure 2.3 makes it evident that while Indonesia produces a lot of palm oil, Malaysia exports the most of it. Therefore, Malaysia and Indonesia are the two countries that export the most palm oil, with Malaysia having a slightly larger share than Indonesia as discussed by Rashid et al. [26].

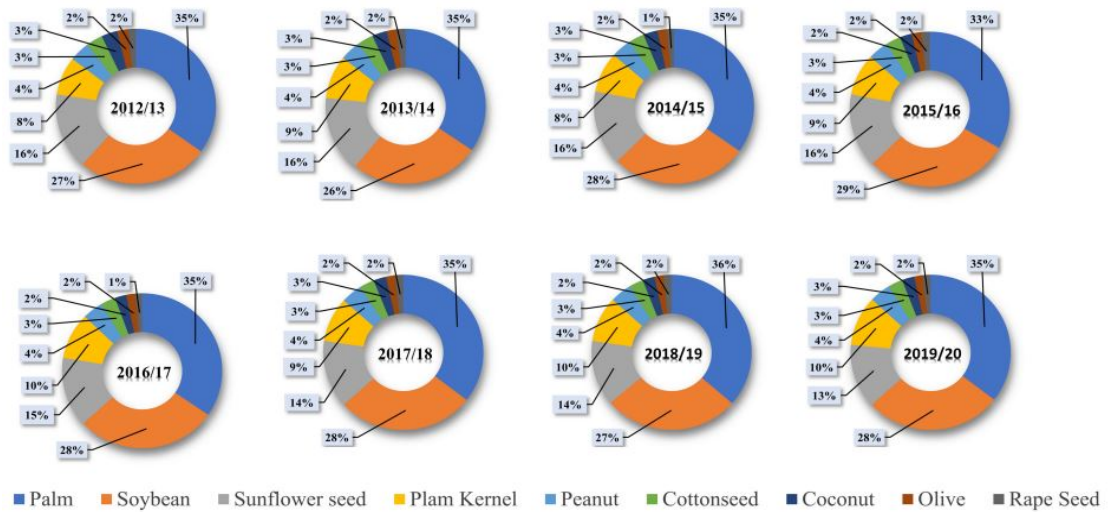


FIGURE 2.2: High Scale Vegetable Oils Production from 2012 - 2020 Approximately in Categories (million MTs) globally by Rashid et al. [26]

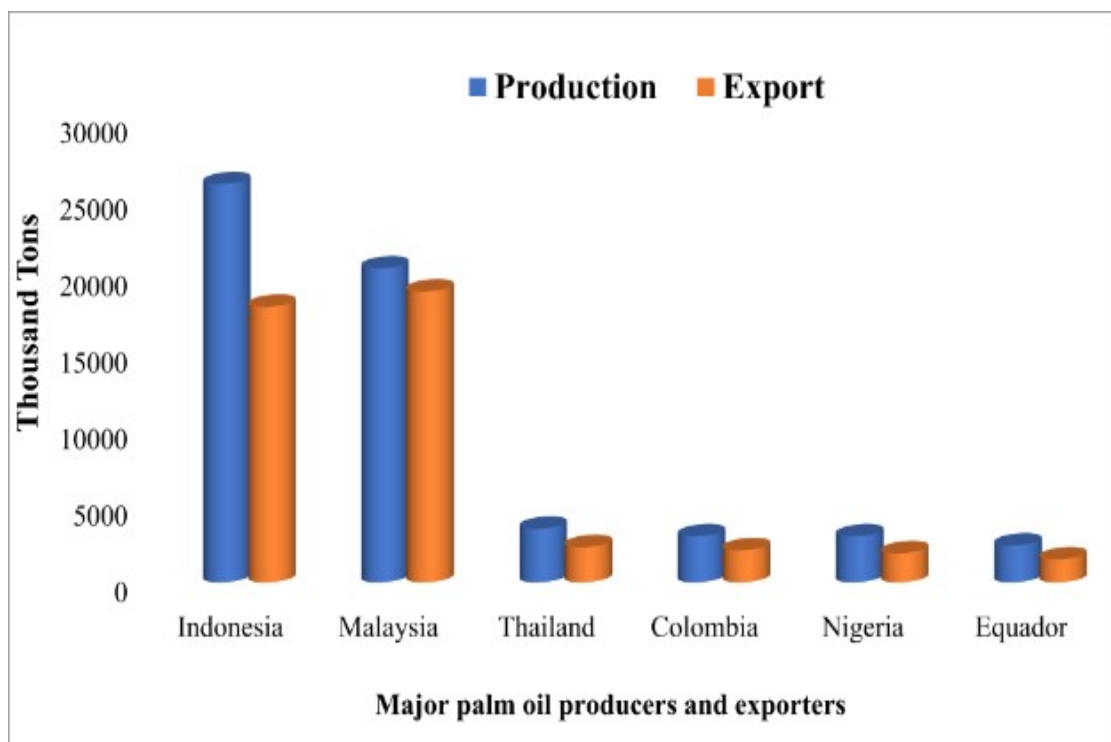


FIGURE 2.3: Major Palm Oil Producers and Exporters by Rashid et al. [26]

Malaysia began producing palm oil in 1911, and it is expected to become into one of the most important sectors of the global economy in the future by Karia and Bujang [27]. Maintaining palm oil exports has several financial advantages, such as global attention, experience, financial contributions, and knowledge management by Rahim et al. [28]. The growing world population favours this more

affordable and potent biofuel as discussed by Hansen et al. [29]. By creating numerous work possibilities for millions of people, the palm oil industry contributes to the growth of Malaysia's Gross Domestic Product (GDP). Malaysia's economy depends heavily on agriculture, which contributes 12% of the country's GDP and provides employment opportunities for 16% of the population. The Malaysian government was one of the largest producers of palm oil in the world in 1999, with 10.55 million MTs produced, according to national figures. According to these estimates, 8.8 million metric tons, or nearly 85%, of this palm oil was exported to outside markets. Countries in Asia, Africa, and Latin America are the primary producers of palm oil. Nigeria, Indonesia, Malaysia, Thailand, and Colombia are the top producers of palm oil. It should come as no surprise that Indonesia and Malaysia, which have certified areas for a huge number of palm tree plantations, are the world's biggest exporters of palm oil. As per the Roundtable on Sustainable Palm Oil (RSPO), Indonesia's palm oil plantation increased from 1.54 million hectares in 2016 to 1.7 million hectares in 2017. In recent years, there has been a rise in the importation of palm oil. The average price of palm oil increased to 649 US dollars a metric ton in 2017, according to a World Bank report. By 2025, the price is expected to reach roughly 744 USD per metric ton by Thompson et al. [30]. The global consumption of palm oil is increasing steadily, as Fig. 4 illustrates.

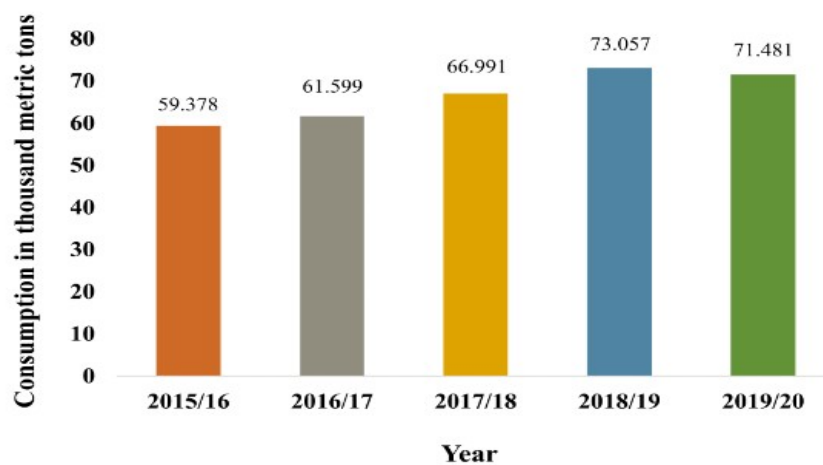


FIGURE 2.4: Global Palm oil Consumption from 2015 - 2020 (in 1,000 MTs)
Approximately by Rashid et al. [26]

The most traded edible oil worldwide is palm oil. A total of 39.04 million tons

were exported in 2011, with Malaysia accounting for 46% of this total. Large sums of money have been invested by Malaysian industries to refine palm oil for the general public's benefit and make it fit for human use. Between 1974 and 1999, processed palm oil exports increased from 0.9 to 8.9 million tonnes, which was the most productive period. After reaching a peak of 16.66 million tonnes in 2010, Malaysia's palm oil exports reached an all-time high of 17.99 million tonnes in 2011. During the same time frame, this export trend was also noticeable, with shipments totalling 1.17 million tons in 2011 by Kim et al. [31].

2.3.1 Fundamentals Aspects of Crop Yield Prediction Process

Data gathering, data pre-processing, data partitioning, and data analysis are the steps in the machine learning-based agricultural yield prediction approach.

The architecture of the crop yield forecast approach based on machine learning is shown in Fig. 2.5. Regression or classification algorithms based on machine learning are used in the data analysis portion by Rashid et al. [26].

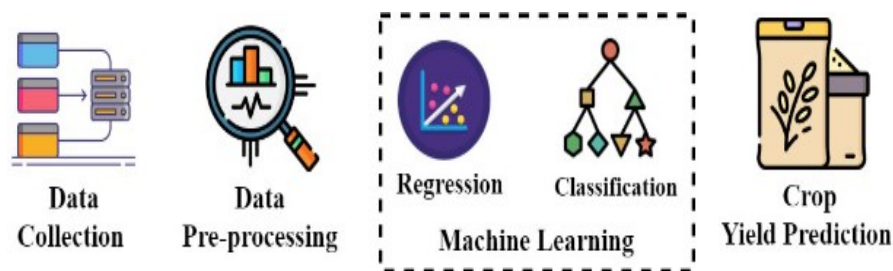


FIGURE 2.5: Crop Yield Prediction by Machine Learning Approaches: A General Architecture by Rashid et al. [26]

2.4 Significant Features Usage for Crop Yield Prediction

Crop production and cultivation-related uncertainty are influenced by a number of factors. Recent research indicates that the most important factors influencing

crop yield are information about leaves and fruit, irrigation, soil properties, climate, cropland, vegetation index, satellite data, crop management, historical yield data, groundwater characteristics, fertilization, socioeconomic factors, phenology data, and more as discussed by Holzman et al. [32]. Actually the production related issues affect the demand and supply chain which leads to import deficient situation. Numerous nutritional supplements are utilized to increase crop yield, which is essential for raising agricultural output. For example, phosphorus, nitrogen, calcium, potassium, sulfur, magnesium, manganese, iron, and so forth are the most often used fertilizer or nutritional supplements by Chlingaryan et al. [33]. Although they can vary in terms of how much is used, these aspects are equally significant to the crops. The output yield would be significantly reduced if any of these vital nutrients were absent. Therefore, when crop management and meteorological conditions are identical, soil characteristics have a significant impact on the geographical variability of crop production. In order to increase the accuracy of yield prediction and provide supplementary information, Wang et al. [34] found that soil quality plays a significant role in crop productivity predictions. Using non-invasive and effective methods, remote sensing will assist in detecting the spatial variations in plant state with high temporal resolution on a large scale by Shiu and Chaung [35].

Due to their non-destructive and labour or time-intensive nature, remote sensing techniques based on spectral vegetation indices and infrared thermometry are widely used for crop production prediction. Healthy and unhealthy images, leaf development, leaf area index, leaves and fruit mortality, number of leaves, number of fruits, and number of stems are the variables that make up the feature group Leave and Fruit Information by Lin et al. [36]. The number of open wells, irrigation ratio, net irrigated area, number of tanks, irrigation details, canal length, number of tube wells, and gross irrigated area are all considered irrigation information. Previous research has revealed that agricultural production prediction algorithms are highly impacted by previous crop yield data Maya and Bhargavi [37]. Aquifer type, rock layer permeability, water conductivity, transmissivity, hydro chemical analysis, and the quantity of micronutrients are among the characteristics that are included in the group of groundwater characteristics [31]. Crop management data

[26], phenology data, and crop land information are other measurements that also play an important role in crop yield prediction by Guo et al. [38].

2.5 Forecasting Models

Prices for the import of palm oil can be forecasted successfully to help rid of supply chain problems, affect the market stability, and guide effective policy measures. In most cases, using ML techniques leads to more accurate forecasts compared to the tried-and-tested statistical models.

This chapter presents the traditional and modern techniques of forecasting, with the emphasis on its use in the forecasting of imports of palm oil and other agricultural commodities. Moreover, it identifies vacancies in research and emphasizes the importance of a combination of various techniques to improve the quality of forecasts.

2.5.1 Traditional Models for Forecasting

Traditional statistical models such as RF, SVM etc. have been widely employed for time series forecasting.

2.5.1.1 Kernel Naive Bayes

The addition of non-linear modeling capabilities by kernel functions developed KNB as an efficient tool for agricultural prediction tasks. In palm oil production, KNB distinguishes itself as a useful tool in plant health condition and disease classification which are key elements influencing yield prediction. The probabilistic basis of the algorithm, along with the kernel transformation, enables it to successfully process complex and high dimensional data obtained from modern monitoring systems of agricultural production.

Tong et al. [39] investigated early oil palm health identification based on UAV images using KNB as a competitive classifier. From the extracted color, texture,

and spatial features of aerial imagery, the study proved the possibility of KNB to achieve distinction between healthy and stressed palms. This early level of classification directly implies an effect on yield estimation, because healthier plantations imply productivity. In a related work, Husin et al. [19] categorized BSR disease from terrestrial laser scanning data. Despite trying a variety of models, KNB represented a potent algorithm because it is able to generalize spatial attributes that are a typically poor attribute of field-level disease databases.

The strength of KNB at processing imbalanced data was brought to the fore in the study conducted by Ahmad et al. [5] in which thermal imagery was used to differentiate between infected and non-infected palm trees. With the use of kernelized techniques, researchers were able to further augment classification precision with the ability to detect weak thermal differences that signaled nascent BSR infection, a feature which by and large is overlooked by conventional linear classifiers. Haw et al. [40] also examined the digital classification methods for BSR detection and reported KNB as a feasible approach for low-resource environments when compared with deep learning. They noted that KNB offers attractive trade-off between fast computation and accurate prediction therefore making it well programmed for supporting the daily decision-making process in farming enterprises. A follow-up study by Haw et al. [41] confirmed that deep learning with imaging data could be used to help identify BSR. The effects of such problems can be averted through timely detection of disease outbreaks, which these models are able to achieve; accurately predicting yields. All these studies conclude that KNB is a robust method for making forecasts of palm oil production since it is capable of tracking the health of trees, identifying influential diseases and advancing forward-thinking to increase productivity.

2.5.1.2 Random Forest

RFs are dominantly applied in agriculture research to predict ecosystem adaptability and study dispersal of organisms [31]. In particular, this supervised learning technique is either taught through using bagging methods or ensemble decision

trees. The bagging strategy assumes that the inclusion of learning models improves the result. Potential of the RF algorithm for random agricultural forests has been studied in a number of works by Cao et al. [42]. Vector collinearity problems which are common problems when using conventional LR models are eliminated through this approach. The capability to manage both discrete and simultaneous variables makes the RF model stand out against linear regression models by Elavarasan et al. [43]. RF algorithm has proven to be a useful predictor of palm oil production due to its ability to handle difficult, non-linear connections between several contributing elements. For example, Firdawanti et al. [44] predicted palm oil production using RF and lag distributed regression and had an R^2 of 0.829 and a correlation of 0.9302. Their model showed that amount of palm oil, land area, age of palm oil and specific delays of temperature and wind velocity were significant predictors.

To predict oil palm production across various blocks of plantations, a further study by Ang et al. [45] combined data from a wide range of sources using the machine learning and deep learning models. The RF model's ability to make strong predictions Jeong et al. [46] also used the RF to forecast agricultural production on regional and global scale. Rather than traditional regression models, their research showed that RF models may be capable of predicting outcomes rather well by controlling complicated interactions between agricultural yields and environmental conditions. Jamshidi et al. [47] investigated 17 machine learning and deep learning methods to forecast oil palm yield using a large agronomy dataset. One of the best was RF model showing how well it is with sophisticated agricultural data. RFs was one of the machine learning models that Gupta et al. [48] compared in an attempt to determine the output of biodiesel. The promising nature of the RF model in the palm oil bioenergy-related practical application was proven by the generation of reliable predictions, according to the study. In addition, Firdawanti et al. [44] paired product emissions data with deep learning methods in efforts to ensure maximum sustainability in the palm oil production process. Though the work emphasizes deep learning, it also presents an understanding into complex modeling techniques that can be applied to the work in collaboration with Random Forests for impact assessments of the environment.

Ang et al. [49] applied RF and other machine learning procedures to forecast future oil palm yield at the block level based upon Landsat and MODIS satellite data. The research demonstrated efficiency of the RF model in integration of remote sensing data to produce accurate yield estimations. Wijaya and Fauziah [50] focuses on market analytics using the RF and LR models to predict crude palm oil sales prices. From this analysis, it was established that RF modeler generated better accurate projections by coping with complex non-linear connections therein market data.

Besides estimating the productivity and the price, RF has been shown critical in prognosticating plant health without being invasive. Putra et al. [51] predicted the nitrogen content of oil palms with the help of methods for machine learning such as RF, based on RGB aerial photography. Their approach enabled non-invasive plant health appraisal and the RF model gave nearly-perfect nitrogen forecasts. Besides, Sboui et al. [52] have applied the use of machine learning methods, RF model in particular, to predict deforestation, caused by planting oil palm trees. The study underlined the model value in revealing drivers of deforestation and sustainable plantation practice. These studies demonstrate RF technique as a flexible and efficient tool for building complex non-linear data relations found in the palm oil production models and other agricultural settings. It is a necessary tool for stakeholders wishing to optimize production processes and market predictions in the palm oil business because of its capacity to process a diversity of data sources and extrapolate complex correlations between parameters.

2.5.1.3 Support Vector Machine

The estimation and tracking of palm oil production have been progressively optimized by the implementation of SVM techniques. For example, Wang et al. [53] implemented an automatic detection of individual oil palm trees from Unmanned Aerial Vehicles (UAV) images using Histogram of Oriented Gradient (HOG) features associated with SVM classifier. Their use of SVM classifiers led to an easy, non-invasive means of diagnosing, and hence enabled better yield forecasting as well as oversight of plantation health. Similarly, Muhammad et al. [54] presented

agricultural monitoring approach for detecting nutrient deficiencies in leaves of oil palm plants. By use of SVM classifiers, their approach provided a non-invasive mechanism of efficient diagnosis system that aided in improved yield forecasting and plantation health management.

Shaharum et al. [55], used the Google Earth Engine platform interlaced with machine learning algorithms such as SVM to map out oil palm plantations throughout Peninsular Malaysia. Their results showed that the SVM algorithm would be useful at classifying oil palm areas with good precision and this proved vital for large-scale production monitoring and planning. Another study by Chen and Liao [56], they discussed the feature extraction methodologies and their combination with SVM to increase the accuracy of palm tree classification with the help of drone images depicting DAV. It was discovered that the optimum feature extraction methods are crucial for SVM classifiers, and their results offered suggestions for improved palm oil production evaluation by means of increased accuracy. When these studies stand together, they can illuminate the crucial role of SVM techniques in improving prediction, monitoring and management of palm oil production which ultimately leads to the more sustainable farm practices.

2.5.1.4 AutoRegressive Integrated Moving Average Model

For analyzing and forecasting the univariate time series data, the ARIMA model used extensively. There are three sub-parts that composed it; AR (autoregression), I (differencing), and MA (moving averages). The time series analysis currently applies ARIMA as a routine procedure. The current study by Hossain K. Z. [57] reports that the model is also effective in short term price forecasting of agricultural commodities such as palm oil, despite usually being limited in its non-linear patterned-capturing capabilities due to ARIMA's linear assumptions. The extensively-used ARIMA model has found extensive application in the forecasting of palm oil production and associated economic variables, as it can capture linear time dependent trends. Latha et al. [58] investigated the boosting of the precision of palm oil imports by combining machine learning procedures and conventional ARIMA models. Their study showed that the use of these combinations

critically enhanced the accuracy of prediction, advance key insights for policy and stakeholders in the agriculture sector.

Srinivasa et al. [59] have completed a study on gasoline and diesel consumption using Seasonal-ARIMA (SARIMA) and neural network models in India. The research showed how effective was SARIMA models in explaining seasonal behavior of consumption of fuel, thus can be used in forecasting of agricultural commodities such as production of palm oil. Tayib et al. [60] intended to predict the Malaysia crude palm oil production with the Seasonal ARIMA (SARIMA). Malaysia crude palm oil production data per month were purchased at Malaysian Palm Oil Board, since January 2014 through to September 2019. The Box-Jenkins approach was used to apply SARIMA model to the data. The results of the model are found adequate and accurate after testing SARIMA(1,0,0)(0,1,1) for the prediction of crude palm oil in Malaysia.

Yee and Samsudin [3] compared the performance of ARIMA and artificial neural networks forecasting for palm oil prices of Malaysia. Based on the findings, they observed that the ARIMA models seemed to outshine during enabling market state while the neural networks were flexible to unstable market situations highlighting the essence of choosing a model depending on market dynamics. Additionally, Zavira R. [61] paid much attention to production planning of crude palm oil in a case study. Research applied ARIMA models to predict production levels, which added value to optimization of production schedules and distribution of resources hence improved efficiency within the company.

Herdiyanti and Sukiyono [62] used ARIMA models to forecast the price of palm oil FFB in the national and Bengkulu Level, Indonesia. From the study, ARIMA (2,1,8), was best suited for national scale, ARIMA (2,1,7) on the other hand worked best for provincial data, with a good forecast for farmers and policymakers. A time series analysis was used in a time series model to forecast import of palm oil in Pakistan by Reheem et al. [63]. Based on ARIMA model, the study was able to generate reliable forecasts that can support strategic planning in policy formulation to improve effectively the import dependencies. Thambiah S. [64] formed a price forecasting methodology for the Malaysian palm oil market out of

the ARIMA models. The research gave much prominence to the ability of the model to detect market trends and generate sound price estimates that are crucial for the stakeholders in formulating sound decisions.

2.5.1.5 Support Vector Regression

SVR has revealed promising results in the proper estimation of various components of palm oil production. Mustakim et al. [65] specifies the effectiveness of SVR versus ANN prediction of oil palm production inside Riau, Indonesia.

According to the research findings, SVR outperforms ANN, because it produces a correlation coefficient of 95% and an MSE of 6%, thus demonstrating the high ability of SVR for precise forecasts on production. In a similar approach, Mustakim et al. [66] used SVR to estimate oil palm yields for the area, with RBF kernel under SVR giving a R^2 of 95% showing its appropriateness to palm oil production modeling. Alfakhri et al. [67] conducted an analysis of the results of implementation of SVR instead of RNN in terms of predicting palm oil yields.

RNN has a tiny advantage over SVR on metrics such as RMSE and MAPE, but the precision of SVR's predictions allows it to become a trustworthy oil palm production predictor. Besides, Pardamean et al. [68] studied SVR alongside K-Nearest Neighbors (KNN) in oil palm yield prediction, which shows that SVR achieves a better R^2 of 0.8709 than KNN showing its effectiveness in reliable yield forecasting.

Caraka et al. [69] proposed an SVR with an RBF kernel on crude palm oil prices and have shown that SVR is a powerful tool in discovery and extraction of complex patterns, hence providing reliable and useful forecasts for market use. Moreover, Goh et al. [70] presented a hybrid system that blends Simulated Annealing with SVR (SA-SVR) to improve the accuracy of forecasted crude palm oil futures prices. The results show that the SA-SVR model can close the gap between the predicted and actual values, making it a reliable approach to the monitoring of price variances in the palm oil industry. Having combined the findings of these studies, it is obvious that SVR is very adaptive and efficient for handling palm oil industries

yield and price forecasts. By helping manage different nonlinear patterns found in large palm oil datasets, SVR has become an irreplaceable tool when it comes to forecasting the process and controlling production.

2.5.1.6 Genetic Algorithm

In agricultural modeling, genetic algorithm has become a powerful optimization tool specifically for enhancing the yield forecasting accuracy. In the palm oil production domain GA has been used for model selection, parameter tuning, and hybrid system integration as it possesses the capability to non-linear and complex search spaces. In order to forecast palm oil export in Malaysia, Chai K. C. [71] proposed a model by combining GA, ANN, and SARIMA. GA played an important role in the selection of optimized ANN and SARIMA parameters which leads to improved forecasting ability. Fine-tuning of neural network structures and time series features leads to enhanced forecasting performance and reduction in prediction errors in economic modelling of palm oil.

A research study by Hilal et al. [72] described that the yield of palm oil could be optimized by utilizing genetic algorithm-based model which fine-tunes the parameters and coefficients of traditional forecasting models. In this approach, their model facilitated the self-calibration of yeild-impacting factors including temperature, fertilizer applications and rainfall.

The study highlighted that GA has capability to boost model accuracy by determining the ideal complex agronomic variable interactions, and hence yielding a robust framework for plantation management decisions.

Another study by Negara and Wustqa [73] combined GA with a Nonlinear Autoregressive Exogenous (NARX) neural network and developed a hybrid GA-NARX model to forecast crude palm oil production in Indonesia. This research used GA as training optimizer to improve the neural network's learning ability by optimizing its biases and initial weights. The GA-NARX integration allowed the model to escape the local minima and better capture the nonlinear dynamics in palm oil production impacted by external factors like market fluctuations and climate.

Collectively, these studies emphasize how GA has affected the performance of predictive models in relation to palm oil production. By being both hybrid optimizer and direct model tuner, GA greatly facilitates the management of complex data sets as well as increases the accuracy of yield estimates in palm oil market.

2.5.1.7 Linear Regression

LR is well known to be a solid and simple-to-use predictive modeling instrument, which often becomes an integral bank method in an agricultural forecast. Extensive appraisal of LR has been used regarding its application to palm oil production. LR was compared to RF algorithms to estimate crude palm oil sales prices through research done by Wijaya and Fauziah [50].

As it was revealed from their research, LR had lower error rates than RF especially when the data contained strong linear trend and consistency. Furthermore, Bi et al. [11] employed LR to predict the price of petroleum, and they pointed out that while the technique produced reasonable outcomes, it struggled to manage the extreme and nonlinear dynamics present in such markets, just as the problems in palm oil datasets affected by changes in the market and weather. Khan et al. [6] also studied the way that LR can be used for predicting palm oil yield in the face of changes in weather and hydrological condition. LR was used as the baseline model and it underperformed, primarily due to being inadequate for modeling the complex spatio-temporal dynamics of agricultural environments, which implied that a more responsive modeling approach was needed. In a study by Schwalbert et al. [74], LR was paired with machine learning in attempts to predict early-season yields on soybean, yielding robust predictive success with relatively linear variable interactions. This demonstrates its usefulness for predictive models of crop production involving arranged sets of data of moderate complexity. Sousa et al. [75] combined LR and principal component analysis in a larger study on environmental modeling in an effort to reduce multicollinearity of predictors when forecasting ozone concentration. This approach could also help studies in palm oil production where various environmental factors that influence production such as temperature, rainfall and humidity are deemed to be inter-related. Overall, these

studies present both the benefits and drawbacks of linear regression. It provides a clear and easy framework but its predictions are made much more effective once the data forms linear patterns. LR provides for a respectable baseline for palm oil production estimates, although it is most of the time outpaced by more intricate plans that involve nonlinearities and influx of diverse feature set styles.

2.5.1.8 Extra Tree

Among the reliable ensemble learning tools for agricultural predictions, Extremely Randomized Trees (Extra Trees algorithm) have been demonstrated the effective capability owing to its robustness, negligible variance, and its potential to deal with nonlinear data. Several recent projects regarding the production of palm oil have shown that the algorithm is outstanding in handling changes in the environment, sensor readings, and the various types of agricultural data.

Khin et al. [76] developed a study through which they employed a hybrid method of integrating federated learning and ensemble models, including Extra Tree, to enhance harvest prediction in decentralized settings. Their framework gave room for cooperative data learning from scattered sources while respecting data privacy, a huge advantage in such sprawling agricultural fields as in cases of palm oil plantations.

The use of the Extra Trees algorithm led to superior yield predictions, due to the fact that it could be able to employ various features from different edge nodes without creating over fitting.

In response to variable weather and soil moisture environments, Khan et al. [77] further integrated Extra Trees into their systematic methodology for predicting oil palm yield. After reviewing several machine learning methods, the researchers discovered that Extra Trees was excellent at forecasting oil palm yield, effectively regulating complex nonlinear dynamics and environmental fluctuations. Extra Trees through its forward strong feature selection and ensemble voting capabilities was able to incorporate the intricate relationships in agroecology and therefore made an excellent selection for new agricultural challenges. Liu and Zhang [78]

used Extra Trees to create a modeling framework that would predict biodiesel yields from palm oil. Development of this model emphasized the targeting of major operational variables such as the concentration of the catalyst and reaction duration to enhance the process efficiency. The use of Extra Trees generated good predictions as well as gave important information about the process variables' effect in the palm oil industry.

Ramasenderan et al. [79] improved an IoT platform for monitoring and optimization of palm tree oil production by adding the algorithm Extra Trees. Extra Trees was used to analyze the real-time sensor readings from automated platforms to precisely profile yield trends, evaluate environmental stress impacts. This research has shown the prospect of algorithm integration into the smart agriculture platforms for immediate decision making and optimization of precision agriculture. All together these studies show the practical application of the Extra Trees algorithm for palm oil production management, in raw yield prediction, biodiesel optimization and real time monitoring. The ensemble character of the algorithm helps it to deal with large, noisy, and scattered data appropriately, providing the best accuracy in palm oil production forecasts.

2.5.1.9 Latest Machine Learning Techniques for Forecasting

The shortcomings of conventional models are addressed by machine learning approaches, which find intricate, nonlinear patterns in data. In order to forecast imports of palm oil, this section examines important machine learning techniques such as ANN, LSTM, RNN, GRU networks and CNN.

2.5.1.10 Convolutional Neural Networks

The basic layers in a CNN are the layers of activation, pooling and convolutional present between input layer and output layer. Although, a pooling layer can apply some techniques such as max-pooling or average-pooling to reduce input data, a convolution layer conducts convolution calculations that apply various local filters to the input data. Yuan et al. [80] discovered that the introduction of nonlinear

operations of the activation layer could enhance the ability of CNN to fit the nonlinear data.

The convolutional neural networks has promising advantages regarding plantation management, sustainability promotion and oil palm yield estimation. Consistently, research shows improved accuracy and efficiency of detection and measurement of fresh fruit bunches bound for palm oil with the use of CNNs. CNNs are critical in forecasting the oil palm production due to their strong ability to process image data and reveal spatial patterns. Khan et al. [6], came up with a generic approach of using CNN to predict oil palm yield by input of weather and soil moisture data. The model acquitted itself well in several weather conditions indicating its robust nature and feasibility for implementation in precision agriculture systems in the real world.

An accurate method of identifying oil palm trees as used in detailed satellite imagery has been proposed by Yuan et al. [80] as a two-stage CNN system. The method they had devised significantly improved detection accuracy on trees in large scale remote sensing projects, and it was useful for monitoring palm plantations and for yield estimation. Moreover, Zheng et al. [81] used Faster-RCNN to build upon the original work, attaining greater effectiveness in an extended scale of high-resolution oil palm detection from satellite imagery useful for automated inventory and mapping process.

Ashari et al. [82] used deep CNN models, in order to classify oil palm fruit ripeness, by utilizing image data. Through clear separation of ripeness levels, the model facilitated the optimization of harvesting time lines and improved quality of oil production. Additionally, Kumar et al. [83] introduced CNN-GA model to predict the growth of oil palm plantings, while emphasizing that combining CNNs with GA can polish growth stage detection and future harvest estimations.

Culman et al. [84] have used deep learning on RGB drone data to identify individual palm trees, which helps plantation management to be more effective, as well as the yield prediction to be accurate. Consequently, the streamlined approach also contributed to improve the clarity of the strategies of resource allocation. In the same vein, the Puttinaovarat and Horkaew [85] have applied CNN technology

combining multi-band remote sensing images in an effort to detect oil palm plantations, showing how much spectral information is important in deep learning-based classification tasks.

The technology is an intelligent image-processing technology presented by Alaa et al. [86] using the CNNs and machine learning designed to identify palm tree diseases. Although the focus of the study is on plant health, it assists in the enhancement of yield prediction by early detection and management of the production-impacting health-issues.

Bonet et al. [87] showed in their work that palm detection accuracy can be enhanced through deep transfer learning with the help of pretrained CNNs which will be especially helpful in the region where labeled data are not in abundance.

Moreover, CNNs and its hybrids models are also applied for time-series data as explained by Sadouk, L. [7] for the data-level and algorithm-level approaches. Data-level approach deals with the encoding of time-series data in frequency-domain signals where algorithm-level approach contains an adaptive filter-based convolutional layer which processes the time-series data.

Livieris et al. [8] utilized the hybrid approach; namely, LSTM and CNN model for the precise forecasting of the time series data of gold prices and its evolution over the time.

The given model takes advantage of the capability of convolutional layers to extract informed knowledge and how to learn the internal representation of time-series data along with LSTM layers to detect short-term and long-term relationships. Jin et al. [9] introduced a prediction approach using the CNN and Bidirectional LSTM (Bi-LSTM) networks with the multidimensional variable.

The horizontal relationship between the variables of multivariate raw data is learned by CNN, and the temporal relationships are predicted by Bi-LSTM. Wibawa et al. [10] present a new hybrid exponential smoothing with CNN called Smoothed-CNN. The approach of integrating strategies is better in prediction than most of the single approach-based solutions.

particularly, related to agricultural goods like palm oil. One of the most successful forecasting models, the LSTM network remarkably is successful in deciphering complex and nonlinear temporal patterns in data. Sagheer and Kotb [89] applied deep LSTM network in the prediction of petroleum output and proved that LSTM performed better than traditional statistical techniques in terms of capturing long-term dependencies well. Applications of energy forecasting have been successful and therefore, interest has been generated to apply these techniques in Agriculture, for example in predicting production of palm oil. Amal I. [90] used LSTM networks to predict CPO prices and his precision was higher than multilayer perceptron's, indicating the ability of the model to handle the non-stationarity, seasonal and trend pattern ascertaining capabilities.

Widiputra H. [91] studied and pursued a GA-optimized CNN-LSTM multivariate model for mobility prediction during the COVID-19 pandemic demonstrating how the use of LSTM combined with convolutions increases one's predictive capabilities.

Although not specific to palm oil, this hybrid approach is a first base technique for more demanding tasks of prediction in the sphere of agribusiness. In the same way, a Multivariate LSTM framework was proposed by Urolagin et al. [92] that incorporates Mahalanobis and Z-Score transformations for forecasting oil prices, evidencing the advantages of the preprocessing and normalization of related data. Karim et al. [93] improved on LSTM by using it together with Fully Convolutional Networks (FCNs) in addressing multivariate time series classification.

Building FCNs with LSTMs allows this approach to show the increased ability of hybrid models to conduct analysis of temporal and spatial data at the same time is one of the strengths for forecasting palm oil production based on climate, soil and market parameters.

Althelaya et al. [94] evaluated the quality of the BiLSTM models for stock market prediction and illustrated that the BiLSTM models were able to introduce the past and future dependencies into the time series data. The capability of BiLSTM in processing past and future data makes it useful for palm oil prediction, and there is the possibility of enhanced accuracy of the predictions after linking future

signals or climate trends. These findings taken together speak to how, when used effectively, LSTM and its advanced versions provide strong prediction techniques for estimating palm oil production and pricing. These benefits include improved prediction accuracy than conventional models as well as capacity to handle complex (Multivariate and nonlinear), nuanced data structures that are applicable when dealing with dynamic and seasonal attributes observed in palm oil dataset.

2.5.1.12 Artificial Neural Network

The artificial intelligence-based approach is critical to time series forecasting. As a result of the popularization, the interest in ANN is enormous. Forecasting is one of major areas of the application ANN models. ANNs make an appealing alternative tool for both forecasting researches and practitioners. ANN models locate relationships by viewing many inputs and outputs to arrive at a formula from which prediction takes place. The more hidden nodes, the more links between inputs and outputs and the latter can also change depending on the particular problem investigated. If the number of nodes is used extensively then the ANN model will risk over-trained and end up memorizing the training data and produce bad prediction. ANNs are appropriate for issues whose solutions are difficult to define though enough data of observances exist by Kartika et al. [95]. An increased application of ANNs as a critical tool in forecasting CPO production has come out because they are capable of approximating intricate nonlinear relationships better than traditional statistical models. Many studies of both regions and applications have substantiated the superiority of ANNs in this regard. For example, Ulama et al. [96] compared ANN and LSTM networks when predicting CPO production in Indonesia, the ANN model managed MAPE levels below 10% and the LSTM model were better than the ANN due to improved ability. In the same way, Aini and Haviluddin [97] used a Backpropagation Neural Network (BPNN) for CPO production prediction with data sourced from an Indonesian plantation revealing a low MSE of 0.0643 showing model is viable.

Environmental and agronomic variables are also important in ANN- based CPO forecasting models. An ANN model was created by Hermantoro and Rudyanto [11]

including land quality along with temperature, rainfall, and humidity. Their model had an impressively high coefficient of determination ($r^2 = 0.98$), which is one way of showing ANNs potential to retain complex plantation interactions (Hermantoro & Rudyanto, 2018). On the same line, Kartika et al. (2020) employed weather-based inputs to train ANN for yield forecasting; proving that meteorological factors are good predictors when converted to non-linear models. This is supported further by Hilal et al. [72] where their NARX neural network model, using more than 19000 data points of state data from Malaysian states, demonstrated significantly improved ability in the prediction of the fresh fruit bunch yields against traditional methods. As compared to operational efficiency, Uzor et al. [98] used ANN to optimize peak time on palm oil production mills. By applying results of the time study on multiple operational stages, the ANN model could deliver a 21.54% reduction of processing time which indicated ANN utility beyond yield forecast up to the production planning and cost minimization. Similarly, Syarovy et al. [99] used a multilayer perceptron model in a big data environment to predict annual palm oil production. They had a MAPE value of 10.52%, correlation coefficient (R) of 0.96 which validates the need for incorporating huge datasets when forecasting accurately.

Generally, such reviews as that done by Khan et al. [22] and previous experimental exercises such as the one done by Ismail and Khamis [100], and they have emphasized the expansion of the use of machines learning, primarily, ANNs, in the palm oil industry. These research works not only highlight the technical strength of the neural networks but also highlight the opportunity for achieving better sustainability and decision making in oil palm farming. Capitalizing on the ANN's ability to deal with extremely high dimensional, non-linear inputs (from climate to operational parameters), stakeholders can obtain improved and on-time predictions that are vital to improvement in efficiency in this economically-crucial industry.

2.5.1.13 Genetic Algorithm Neural Networks

The application of GANN augments the learning capability of ANN through refinement of weight initialization, parameter tuning and structuring optimization

which are necessary when modeling nonlinear multivariate processes in agriculture. It is one of the first practical applications of this approach when Silalahi D. D. [101] suggested a neural network trained with the method of the genetic algorithm to forecast international prices of CPO and SBO. Although price forecasting was the main theme of the study, the important relevance of the study is that it shows how GANN models are capable of handling volatile time series of data subjected to a collection of environmental, economic, and policy factors. The ability of the model to exploit complex nonlinear relationships in commodity prices establishes a structural foundation that is relevant to palm oil yield prediction in which comparable dynamical relationships exist between climatic factors, market trends, and agricultural contours.

Based on the near-infrared spectral data, the ripeness of the oil palm fruit bunches was classified using a neural network optimized by the GANN for palm oil quality assessment by Silalahi et al. [102]. The study was conducted to automate the ripeness grading process based on spectral features as the GANN model input.

The GA was useful for the selection of optimum wavelengths in the spectral as well as for configuration of the structure of ANN for precise classification. Although not the goal of the study, the work supports yield forecast through an improved independent capacity of estimating harvest readiness, and the optimal period of fruit collection that have direct impacts on production estimates.

In addition, Ahmad et al. [103] adopted a GANN framework used to control palm oil mill emissions. Their work showed how an integration of neural networks with the genetic algorithms enhanced model generalization and optimization in systems where the environmental data is inputs.

While aimed at emissions control, the model's construction is reusable for production prediction, particularly when the environmental impact factors such as temperature, humidity, and biomass intake affecting yield of oil are concerned. Their strategy represents the flexibility of GANN for the modeling of complex relationships in palm oil processing and production systems. Taken together, these studies demonstrate that GANN is a strong predictive modeling tool that can account for the multifactorial and nonlinearity of the palm oil systems.

Through the optimization of the neural network through genetic search techniques, the traits that make GANN models good for reliable palm oil production forecast under real world conditions are optimized such as convergence, prediction error reduction, and generalization improvement.

2.5.1.14 Recurrent Neural Networks (RNN)

RNN belongs to the category of deep learning since much number of layers are used in processing data. RNN has been developing at a fast speed and redefined such areas of study as Natural Language Processing (NLP), speech recognition, synthesis of music, time series financial data processing, DNA series analysis, video analysis and more.

The prediction of palm oil production has received tremendous attention over the last couple of years especially using the RNN which works splendidly through time series data. Sugiyarto and Abadi [104] used an LSTM-RNN model to predict Indonesian palm oil and found LSTM performs better than the traditional forecasting models as it models longer term dependencies in the production data. Their work affirms the model's reliability and usefulness in future production trend forecasting and palm oil industry's decision making.

In the same way, the authors in Syarovy et al. [99] used an RNN-LSTM method to predict oil palm production. RNN-LSTM performance was compared with other machine learning techniques in this research, and RNN-LSTM performed better in terms of accuracy of forecasts. The model's temporal and seasonal palm oil production pattern addition goes a long way in building its relevance in real-world agricultural prediction applications. Additionally, Husaini et al. [105] investigated the use of the LSTM, a variant of RNN algorithm to predict palm oil production. The results of their study have shown that the model practically caught complex patterns of historical production data, and the results of the future trends were quite accurate and reliable. The study explains that LSTM is particularly good for time series forecasting but in the agricultural sector in particular where the patterns of production may depend on the number of different seasons. In addition

to this, RNN algorithms were included by Mimboro et al. [106] in an AIoT-based weather monitoring system to improve the precision of palm oil plantation productivity estimations. The study demonstrated that RNN can be applied to predict the oil palm production by examining such external conditions as the weather pattern, etc.

This research shows the possibility of enhanced accuracy of palm oil production forecasts through the RNN models (partially with weather data). Together, these studies demonstrate increasing potential of RNN techniques in forecasting the palm oil production, promising tools for the industry to be used in facilitating decision making and productivity forecasting.

2.5.1.15 Gated Recurrent Unit

Palm oil production forecasts have also been attempted through the use of GRU networks, which are variant of RNN known for efficiency with sequential data. An optimized GRU model built by Avula et al. [107] was used for oil production prediction, with excellent performance in prediction future production trending. Their work highlights the usefulness of GRU in its ability to model complex temporal patterns from the data lending itself as a trustworthy tool for prediction of production in the palm oil industry.

In the same line, Saputra et al. [108] applied the GRU method to determine the future value of exports and imports of oil and gas and showed the effectiveness of the method at modeling time-series data and producing correct predictions.

Regardless of their focus on oil and gas, the findings are also suitable for assessing palm oil production forecasting because both operate on similar market dynamics subject to global market forces. Their efforts explain the robustness of GRU in dealing with fluctuations and making dependable forecasts.

Mateus et al. [109] compared LSTM and GRU models to forecast the condition of a pulp paper press whose processes are not related to palm oil production, but which provides information about GRU's capability in predictive modeling in industries that rely upon continuous data. According to their results, while

computationally less taxing than LSTM, GRU can be a promising alternative for real-time forecasting in such areas of activity as agriculture.

Finally, Tardini G. A. [18] compared GRU and LSTM models for forecasting crude palm oil prices, but with similar performance in predictive accuracy between the GRU and LSTM models used. Nevertheless, GRU was shown to be more efficient in training time in the study, which should make it a more desirable option for large-scale palm oil production prediction tasks, for which the efficiency of a computing process is a consideration.

These researches present effective applications of GRU models in predicting palm oil production and other pertinent factors, which adds to agricultural forecasting body of knowledge.

2.6 Hybrid Models for Enhanced Forecasting Accuracy

Hybrid models which utilize a great number of machine learning procedures to take advantage of their specific advantages, have been also discussed in recent research. Because these hybrid models account for a number of aspects of the behavior of the data, they have shown tremendous potential in increasing the accuracy of the forecast. For instance, CN-LSTM hybrid model resulted in successful prediction of Crude Oil price, indicated by Mukkamala et al. [110] study, suggesting that it may have similar potentials for forecasting the CPO prices [80]. For example, ARIMA and ANN hybrids use ANN to model nonlinear interactions and ARIMA to linear patterns.

Mehta and Patel [111] showed how effective hybrid models were in forecasting the price of agricultural produce like palm oil. A further experiment examines the predictive effectiveness of LSTM and XGBoost on crude palm oil production. It measures their ability to predict sale actually, regulate inventory levels, and assist production scheduling. The findings reveal that LSTM is superior to traditional statistical models of detecting patterns of time series in palm oil production when

these are optimized Aqbar and Supomo [112]. Such models are currently applied in more recent research works by since they eliminate the weaknesses associated with individual approaches.

2.7 Performance Evaluation Metrics

Performance of a model can be described by using the measures of evaluation. Since the evaluation metrics can differentiate between the output from different learning models, they are very important [42].

Among the performance measures, those which are used to measure the performance of the regression model are Mean Absolute Error (MAE), mean squared error (MSE), RMSE, R-squared (determination coefficient) and MAPE. For any given array of predictions, the average significance of the errors is found out by taking arithmetical mean of absolute variance of actual observation from observed estimation or MAE by Ali at al. [113].

Measure of effectiveness of estimator MSE reflects proximity of the regressor line to the dataset points by Mukkamala at al. [110]. RMSE measures the amount to which the data is grouped on the best fit line for predicting the residuals or the standard deviation of error of prediction in report [111].

In order to demonstrate how the developed framework outperforms the baseline framework, the determination coefficient measures how well the regression framework fits the data by Aqbar and Supomo [110], and Rousson and Gosoniu [114]. MAPE, or mean absolute percentage error, measures the degree to which the model's prediction differs from its associated outcomes.

Using metrics like accuracy, precision, recall, sensitivity, specificity, and F1 Score, machine learning-based classification algorithms for crop production prediction are assessed. For classification difficulties, however, the most widely used and useful measure is classification accuracy. The tables of the hyperparameters of all techniques is described below along with the most widely used techniques along with their frequencies.

TABLE 2.1: Summarized Table for Hyperparameter Tuning for Comparison using the Capacities and Limitations

Sr.	Techniques	Hyperparameters	Capacities	Limitations
1.	LR	Regularization such as Lasso ($L1$), Ridge ($L2$) and None; Learning rate	Provides simplicity and interpretability; fast training and prediction	Often fails with linear data; Sensitive to multicollinearity
2.	GA	Population size; Mutation rate	May optimize nonlinear patterns; avoids local minima	High computation cost; slow convergence rate
3.	SARIMA	AR order p, q, d , differencing mechanism, MA order; Seasonality parameters p, q, d, m	Capturing seasonal patterns, better for univariate datasets	Requires stationary data; Assuming Linear dependency
4.	KNB	Uses K number of neighbors to weight probability; euclidean distance metric	Add bayesian probability with native neighbors; deals with numerical and categorical datasets	Assuming features' independence; may sensitive to irrelevant features
5.	Extra Trees	Features like $n_estimators$, and $max_features$	Efficient in ensemble learning; deals with complex and nonlinear patterns	Less Interpretability features; can over-fit in noisy data

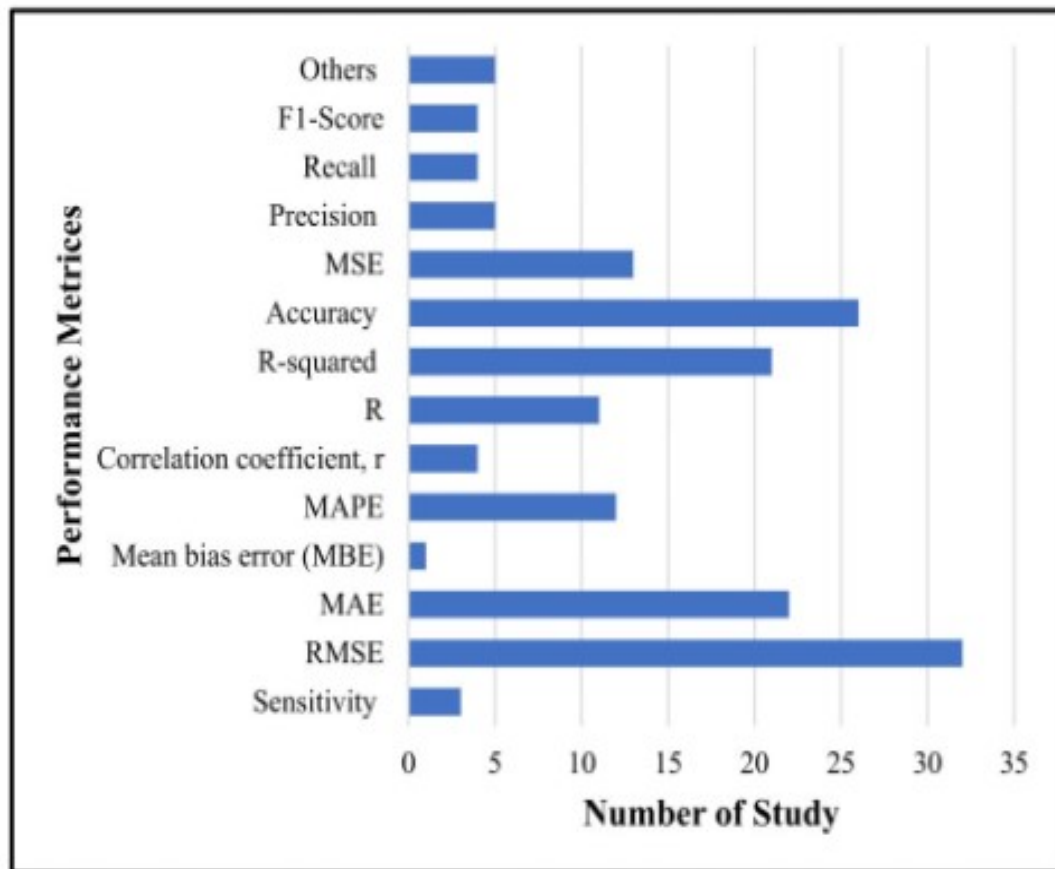


FIGURE 2.7: Popular Performance Evaluation Metrics for Crop Yield Prediction algorithms by Rashid et al. [26]

2.8 Research Gap

Accurate forecasting of palm oil imports is very important for policymakers, businesses, and stakeholders in the agricultural and trade sectors as mentioned previously. As efficient forecasting mechanisms help in optimizing supply chain management, reducing market volatility, and improving economic planning. In literature, many traditional methods are applied for forecasting agricultural imports, such as time series models like SVM, RF, ARIMA and GARCH. CNN was mostly used used image based data, few papers discussed it for time series analysis. So in order to improve its application towards time-series data. It is also added here. In short, all of these methods are found insufficient to capture complex nonlinear patterns in the data. In addition, transformation of advanced decision-making strategies as discussed above are of the major focus of this work. The advent of machine

and deep learning techniques, including SVM, CNN, LSTM and ANN, has opened new possibilities for improving forecasting accuracy by leveraging their ability to model nonlinear relationships and process large datasets effectively. The previous research works are focused on the palm oil consumption and price forecasting on their specific country's data. However, the major objective of this research work is to forecast the Import of Palm Oil and price in Pakistan on the basis of large dataset using the machine learning and deep learning techniques. It also contains some initial assumptions which are considered for better planning to boost up the accurate forecasting of palm oil to fulfill the demand of Pakistan nation. Although there are advancements in forecasting methods; however, multiple challenges are still existing in achieving high accuracy in palm oil import predictions. In literature, traditional statistical models like ARIMA and GARCH are discussed to handle these issues but these are limited in their capacity to handle complex, nonlinear relationships inherent in economic and agricultural data. This limitation indicated that there should be some advanced mechanisms like exploration of machine learning and deep learning methods, such as SVR, CNN and ANN, which will improve adaptability and precision in modeling these complex activities.

TABLE 2.2: Frequency of Techniques used in the Literature

Sr.No	Techniques	Freq.	References
1	CNN	19	8,,9,10,11, 22,35,36,38,37,40, 48,50,56,62,63,65,67,77
2	RF	17	4,6,7,10,11,12,13,30,31,33,42, 44,45,46,69,79,80
3	ANN	10	20,32,43,39,47,52,58,68,74,76
4	ARIMA	8	1,24,25,26,34, 81, 82, 83
5	LSTM	7	15,18,27,28,41,61,72
6	SVM	7	16,53,54,57,60,64,70
7	SVR	5	2,3,14,66,75
8	EXTRA TREE	3	9,21,71
9	RBFNN-2 SATRAAIS	1	5
10	NARNN	1	17
11	GANN	1	19
12	J48-REP	1	23
13	GA	1	69
14	GPR	1	73
15	TE	1	55
16	LR	1	29
17	CROPTED	1	49
18	KNB	1	51

2.9 Summary

In summary, this chapter has provided an in-depth review of the key elements involved in forecasting palm oil production, covering its economic significance, current production trends, and the essential features influencing crop yield prediction. Various forecasting methods, including both traditional statistical models and modern machine learning techniques, have been critically examined with a focus on their applicability, strengths, and limitations in the context of agricultural prediction. Advanced models such as CNN, ANN, and hybrid approaches were highlighted for their potential to improve forecasting accuracy by capturing complex patterns in data. Additionally, the chapter has emphasized the importance of appropriate performance evaluation metrics in assessing model reliability. This comprehensive review not only establishes a strong foundation for the current study but also identifies existing research gaps that guide the direction of the proposed methodology.

Chapter 3

Proposed System Model

3.1 Introduction

This chapter outlines the proposed system and methodologies used for forecasting palm oil imports in Pakistan using the dataset of twelve years, i.e., 2013 – 2024 from Pakistan Bureau of Statistics by using some advanced machine learning and deep learning techniques.

In addition, details workflow of the system model is discussed. All steps of the techniques are also elaborated for the processing of the dataset in order to forecast the palm oil imports to Pakistan in the upcoming subsections. The details of the datasets and proposed system model are discussed below.

3.2 Fractions of Palm Oil and Their Hierarchy

Palm oil is obtained by extracting oil from the middle part of oily palm fruit (*Elaeis guineensis*).

It consists of a mixture of triglycerides and, after fractionation, the oil is separated into mixtures with different physical and chemical traits.

Palm olein and palm stearin are the key fractions of crude palm oil produced during fractional crystallization as discussed by Zheng et al. [115].

3.2.1 Palm Olein

This liquid portion of palm oil is transparent and flows smoothly at an ordinary room temperature. Since olive oil is rich in unsaturated fatty acids, it is used for cooking and dressing salads and in frying. As it doesn't easily oxidize and remains stable at high temperatures, palm olein is often used in both kitchens and food industries.

3.2.2 Palm Stearin

This solid part of palm oil has more of the saturated fats i.e., palmitic acid and stearic acid. Palm stearin's solid state at normal temperatures means it is used in making margarine, shortening, baked food and frying for industries.

Because of its hardness and ability to mix two substances, stearic acid is added to soaps and cosmetics.

3.3 Secondary and Specialized Fractions

Further fractionation of palm olein and stearin leads to specialized fractions that serve niche markets:

3.3.1 Super Olein

This fraction is produced at second-stage fractionation which makes it higher in unsaturated fats and still leaves it completely transparent at lower temperatures. This is perfect for cooking oil that will be bottled in temperate parts of the world.

3.3.2 Hard Stearin

After deep fractionation, palm stearin becomes hard stearin and is commonly used in making confectionery fats and bakery margarines.

3.3.3 Palm Mid Fraction (PMF)

It resembles cocoa butter, so it is frequently applied in the confectionery sector, mostly to produce substitutes for cocoa butter and its equivalents.

3.4 Hierarchical Structure of Palm Oil Fractions

Palm oil fractions are grouped by their melting temperatures, chemical makeup and their uses in industry. At the highest level is CPO which is then refined to give Refined, Bleached and Deodorized (RBD) palm oil.

Fractionation of the RBD palm oil results in palm olein and palm stearin and then further processing forms super olein, hard stearin and PMF in report by Aqbar and Supomo [112].

The structure of this system allows it to satisfy many industries demands and also helps in forecasting the supply chain since different petrochemical segment volumes are produced and required according to global consumption trends.

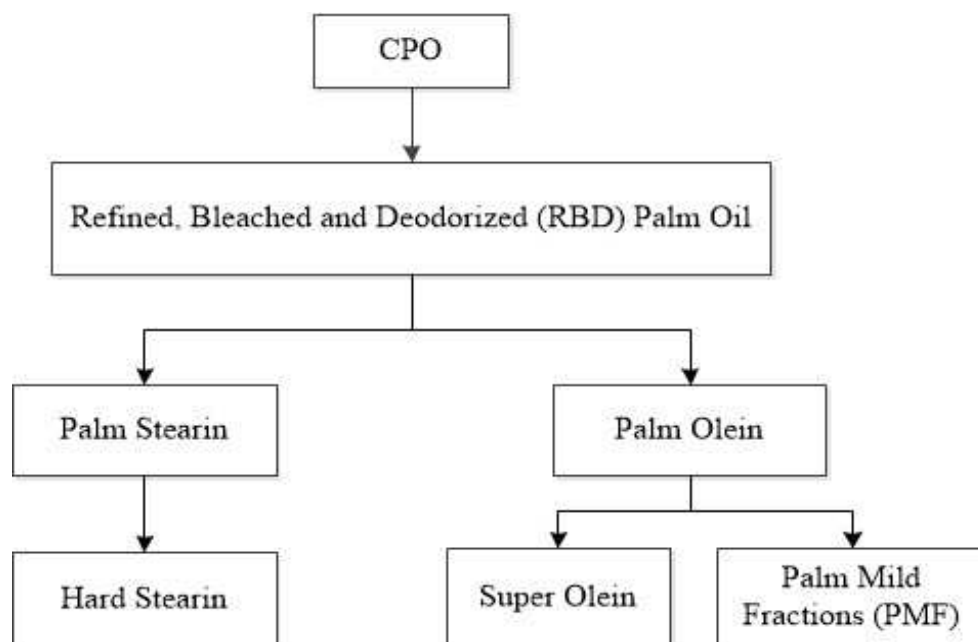


FIGURE 3.1: Hierarchy of Palm Oil Fractions

3.5 Sample Data of Palm Oil and Price Fractions in Pakistan (2013–2024)

Sample data related to palm oil production, import, and consumption is essential for building reliable forecasting models. In Pakistan, palm oil is not produced locally on a commercial scale due to unsuitable climatic conditions, so the country relies heavily on imported palm oil, mainly from Malaysia and Indonesia. Sample datasets typically include variables such as monthly import volumes (in metric tons), import prices (USD/ton), exchange rates, local consumption trends, commodity prices, and global palm oil production indices are taken from reports by [115], [116]. These datasets are often obtained from sources such as the Pakistan Bureau of Statistics, State Bank of Pakistan (SBP), Food and Agriculture Organization (FAO), and international trade databases (PBS, 2023; SBP, 2023; FAO, 2022; UN Comtrade, 2023). For machine learning-based forecasting, historical data spanning the past 10–20 years is valuable in capturing seasonal trends, economic impacts, and policy-driven fluctuations. Incorporating such diverse variables improves the robustness of predictive models and helps identify key factors influencing palm oil trends in Pakistan. Pakistan’s reliance on imported palm oil and its fractions has been significant over the past decade. The primary fractions imported include Palm Olein, Palm Stearin, and CPO. These imports cater to various sectors such as edible oil manufacturing, Vanaspati ghee production, and the oleochemical industry. The following tables presents the import volumes of palm oil quantity and price and its fractions over the past decade:

From this data, it is observed as following.

3.5.1 Shift from CPO to RBD Products

Over the decade, there’s been a noticeable shift from importing Crude Palm Oil to RBD Palm Olein and Stearin. This trend indicates a move towards importing more processed forms of palm oil, reducing the need for local refining.

TABLE 3.1: Import Volumes of Palm Oil (MT) and its Fractions over the Past Decade 2013 - 2024

Year	CPO	RBD Olein	RBD Stearin	Total Oil Imports
2013	1,200,000	1,500,000	300,000	3,000,000
2014	1,150,000	1,600,000	320,000	3,070,000
2015	1,100,000	1,700,000	340,000	3,140,000
2016	1,050,000	1,800,000	360,000	3,210,000
2017	1,000,000	1,900,000	380,000	3,280,000
2018	950,000	2,000,000	400,000	3,350,000
2019	900,000	2,100,000	420,000	3,420,000
2020	850,000	2,200,000	440,000	3,490,000
2021	800,000	2,300,000	460,000	3,560,000
2022	750,000	2,400,000	480,000	3,630,000
2023	700,000	2,500,000	500,000	3,700,000
2024	650,000	2,600,000	520,000	4,100,000

TABLE 3.2: Import Volumes of Palm Oil Price (USD) and its Fractions over the Past Decade 2013 - 2024

Year	CPO Price	Olein Price	Stearin Price	Avg. Price
2013	800	850	700	Stable
2014	750	820	680	Slight Decline
2015	700	800	650	Decline
2016	650	780	620	Volatile
2017	600	750	600	Downward
2018	550	720	580	Low Demand
2019	500	700	550	Bearish
2020	650	800	600	COVID Recovery
2021	900	950	750	Supply Shortage
2022	1,200	1,250	1,000	All-time High
2023	950	1,000	850	Correction
2024	850	900	750	Stabilizing

3.5.2 Increasing Demand for RBD Palm Olein

The demand for RBD Palm Olein has consistently risen, aligning with its extensive use in cooking oils and the food industry.

3.5.3 Steady Growth in RBD Palm Stearin Imports

RBD Palm Stearin imports have shown steady growth, reflecting its importance in the production of vanaspati ghee and other industrial applications.

3.5.4 Uses of Fractions of Palm Oil

As importing palm oil is easy and inexpensive, it helps move food production and manufacturing forward in Pakistan. 85% of imported palm oil goes into cooking oils and Vanaspati ghee in the food sector (Pakistan Bureau of Statistics, 2023; FAOSTAT, 2022). Besides, it is an important ingredient for industrial products like soaps, detergents, cosmetics and pharmaceuticals. Because it costs less than most other oils, it plays a crucial role in food security and keeping the market price stable in the country. The unique fractions produced from crude palm oil are used throughout the industries in Pakistan.

Frying and everyday cooking make heavy use of palm Olein which is a liquid made from palm oil, for its attractive color, oxidative stability and the ability to handle temperatures needed for frying.

This palm oil fraction is used more than any other in the local edible oil blends available. Being solid at room temperature, palm stearin is mainly used to make Vanaspati ghee, baking products shortening and frying fats by bringing structure and stability to their shelf life.

Although not widely used in Pakistan yet, the use of the PMF gives potential for the manufacture of chocolate and bakery fats. Increasing needs for processed foods and personal care items are expected to result in a larger use of these specialized fractions by Tong et al. [39].

If stakeholders know the final uses of these fractions, they can better connect their plans to actual market demands, making machine learning models predictions more accurate.

3.6 Proposed System Model

The hierarchy of this work includes data collection, preprocessing, model selection, and evaluation metrics. Following ML and DL techniques are used in this work for the evaluation of the model as RF, CNN, ANN and ARIMA.

These models are considered because of their robustness in handling time-series data and their ability to model non-linear relationships. Diagram of system model is visualized in Figure 3.1. In this diagram, the hierarchy of our proposed model is visualized.

In first step, data gathering is done. Then data cleaning and outliers are removed. After preprocessing, the proposed machine learning and deep learning techniques such as RF, CNN, ANN and ARIMA models are used for the training and testing of the model using the dataset in order to forecast the palm oil consumption and its price.

Training is done on 70% of the dataset while testing is done the remaining 30% of the datasets. The results are computed and displayed in results and discussion chapter.

These models are implemented in python and compared for checking the accuracy, precision, MAPE and RMSE by considering its performance metrics. All models are tested using the above-mentioned performance parameters and their performance is displayed via graphs. The detailed analysis is done in order to check the robustness of the proposed model. Achievements and limitations of all techniques are also elaborated in detail which show that these techniques have good features along with few shortcomings. Then best model is chosen as displayed in the diagram. The details of these models are discussed below in the subsections 3.3 to

3.6 respectively. Details of the dataset used for this system is listed below in Table 3.3.

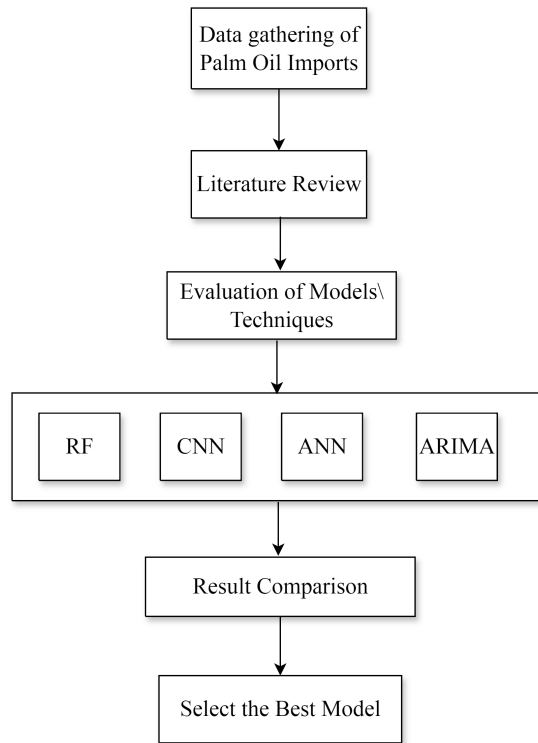


FIGURE 3.2: Proposed System Model

TABLE 3.3: Import Volumes of Palm Oil Demand and Price, and its Fractions over the Past Decade 2013 - 2024 on Monthly Basis

Year	Month	Quantity (MT)	Value (\$ Thousand)
2013	1	222 156.000	185 848
2013	2	247 602.000	201 491
2013	3	159 850.000	131 881
2013	4	187 151.000	156 797
2013	5	160 509.000	135 044
2013	6	154 016.000	127 513
2013	7	220 105.000	182 167
2013	8	178 638.000	146 565
2013	9	165 372.000	132 720
2013	10	192 258.000	153 675
2013	11	208 051.000	166 507

Table 3.3 – continued from previous page

Year	Month	Quantity (MT)	Value (\$ Thousand)
2013	12	152 900.000	125 409
2014	1	210 709.000	178 548
2014	2	169 017.000	143 138
2014	3	165 341.000	139 133
2014	4	186 171.000	161 316
2014	5	154 958.000	142 110
2014	6	261 228.000	230 872
2014	7	141 288.000	131 813
2014	8	198 169.000	167 950
2014	9	226 053.000	181 926
2014	10	213 458.000	160 296
2014	11	212 245.000	155 338
2014	12	214 095.000	156 036
2015	1	160 354.000	120 069
2015	2	188 102.000	131 197
2015	3	179 358.000	135 113
2015	4	173 743.000	116 960
2015	5	211 668.000	140 269
2015	6	278 197.000	182 339
2015	7	160 019.000	106 433
2015	8	256 208.000	169 983
2015	9	258 250.000	161 373
2015	10	283 244.000	161 142
2015	11	183 115.000	121 459
2015	12	187 046.000	109 532
2016	1	214 113.000	121 315
2016	2	237 670.000	134 746
2016	3	268 352.000	155 210
2016	4	214 385.000	134 819

Table 3.3 – continued from previous page

Year	Month	Quantity (MT)	Value (\$ Thousand)
2016	5	257 672.000	172 453
2016	6	199 104.000	140 973
2016	7	150 728.000	104 771
2016	8	204 325.000	138 976
2016	9	187 092.000	124 637
2016	10	201 969.000	140 997
2016	11	223 509.000	160 564
2016	12	244 224.000	173 356
2017	1	236 950.000	171 946
2017	2	226 052.000	170 818
2017	3	255 243.000	198 011
2017	4	217 305.000	166 793
2017	5	261 326.000	195 237
2017	6	198 118.000	159 032
2017	7	213 230.000	176 598
2017	8	235 977.000	183 016
2017	9	217 284.000	150 689
2017	10	257 531.000	178 734
2017	11	211 913.000	176 782
2017	12	242 340.000	171 550
2018	1	243 366.000	171 179
2018	2	200 684.000	137 205
2018	3	295 288.000	198 128
2018	4	269 672.000	185 327
2018	5	232 554.000	159 114
2018	6	223 513.000	151 393
2018	7	252 723.000	168 264
2018	8	224 763.000	147 499
2018	9	267 133.000	169 952

Table 3.3 – continued from previous page

Year	Month	Quantity (MT)	Value (\$ Thousand)
2018	10	230 673.000	136 903
2018	11	271 148.000	164 063
2018	12	252 662.000	144 981
2019	1	294 346.000	169 211
2019	2	251 453.000	134 485
2019	3	273 588.000	150 720
2019	4	270 603.000	153 081
2019	5	302 717.000	166 704
2019	6	255 740.000	138 732
2019	7	129 625.000	71 264
2019	8	280 413.000	150 509
2019	9	261 736.000	140 659
2019	10	257 557.000	144 155
2019	11	301 171.000	170 391
2019	12	285 679.000	165 442
2020	1	244 477.000	159 015
2020	2	251 501.000	181 465
2020	3	264 863.000	197 346
2020	4	266 216.000	189 784
2020	5	209 634.000	134 880
2020	6	217 902.000	136 617
2020	7	341 825.000	205 935
2020	8	243 408.000	146 794
2020	9	347 287.000	226 280
2020	10	116 614.000	83 959
2020	11	284 289.000	213 051
2020	12	296 346.000	235 846
2021	1	301 710.000	255 969
2021	2	235 764.000	217 551

Table 3.3 – continued from previous page

Year	Month	Quantity (MT)	Value (\$ Thousand)
2021	3	276 050.000	275 549
2021	4	275 371.000	281 262
2021	5	238 547.000	255 121
2021	6	240 537.000	271 292
2021	7	221 677.000	254 017
2021	8	289 267.000	323 007
2021	9	283 525.000	314 131
2021	10	216 887.000	246 115
2021	11	323 262.000	389 991
2021	12	250 506.000	316 657
2022	1	217 049.000	286 459
2022	2	235 093.000	309 765
2022	3	220 880.000	290 523
2022	4	257 655.000	367 734
2022	5	203 161.000	307 739
2022	6	105 607.000	143 162
2022	7	194 080.000	299 359
2022	8	266 373.000	416 938
2022	9	300 969.000	419 466
2022	10	187 032.000	270 837
2022	11	284 963.000	318 112
2022	12	342 911.000	357 538
2023	1	351 278.009	364 247
2023	2	223 692.000	234 568
2023	3	229 877.032	236 187
2023	4	182 624.966	180 068
2023	5	275 293.000	294 163
2023	6	235 677.000	249 222
2023	7	277 801.000	278 939

Table 3.3 – continued from previous page

Year	Month	Quantity (MT)	Value (\$ Thousand)
2023	8	294 203.000	273 464
2023	9	221 947.000	206 171
2023	10	222 577.000	205 718
2023	11	231 701.000	210 932
2023	12	237 487.000	212 906
2024	1	249 617.000	223 296
2024	2	224 073.000	199 078
2024	3	304 307.000	273 921
2024	4	235 093.000	216 102
2024	5	240 330.000	231 002
2024	6	257 530.000	247 025
2024	7	256 460.000	239 921
2024	8	273 035.000	255 953
2024	9	265 944.000	251 232
2024	10	289 314.000	276 334
2024	11	234 887.000	235 590
2024	12	262 354.000	280 627

3.6.1 Random Forest

This algorithm initiates by constructing different decision trees and then considering their predictions by ensemble. It creates subsets by randomly selecting data points and features, builds single decision trees for each subset, and then aggregates the predictions from all trees for a final outcome as described by Firdawanti et al. [44]. It improves accuracy and minimizes overfitting as compared to using a single decision tree. All steps of this algorithm are discussed below.

3.6.1.1 Bootstrap Sampling Bagging

It randomly samples data points with replacement from the actual dataset for creating different subsets bootstrap samples. It also defines that some data points

may be included there many times in a subset, while others may not be included.

3.6.1.2 Feature Randomness

Features are selected randomly for each decision tree. It adds randomness and supports in preventing overfitting.

3.6.1.3 Decision Tree Construction

it is performed for training a separate decision tree on every bootstrap sample following the selected features.

3.6.1.4 Prediction and Aggregation

Then, each constructed decision tree makes a prediction using new data points. In classification, the final prediction is the mode values as the most frequent value of the individual tree predictions. In regression, it is the average of the individual tree predictions.

3.6.1.5 Model Evaluation

This model is evaluated on the metrics like RMSE, MAPE, Loss, and prediction. In Table 3.4, hyperparameters of RF model are discussed, which are used in simulations.

TABLE 3.4: Hyper Parameters of RF

Hyperparameters	Values
<i>n_estimators</i>	200
<i>random_state</i>	42
<i>max_depth</i>	10
<i>min_sample_split</i>	5

In Table 3.4, the hyperparameters are used for the external procedures especially for training process based on the fine tuning of the system. Here, *n_estimators* defines the number of trees, *random_state* acts as a seed of reproducibility, *max_depth*

specifies the maximum depth of every tree, and *min_sample_split* is used for setting the criterion of minimum sample required to split a tree, i.e., 2-20 in this case.

3.6.2 Convolutional Neural Network

This algorithm comprises of few steps as discussed here: preprocessing of data, feature extraction using convolutional layers, reducing spatial dimensions by using pooling layers, applying activation functions for finding non-linearity, and finally, prediction is done using fully connected layers as described by Ashari et al. [82]. The major breakdown of the whole model is elaborated below:

3.6.2.1 Data Preprocessing

It includes data preparation by loading and preparing the dataset. This step includes the tasks like data augmentation, normalization, and splitting of dataset into training and testing sets. Data values are normalized to a specific range as per the requirement.

3.6.2.2 Convolutional Layers

It is based on convolutional operation because these layers apply filters or kernels to slide across the input data, while performing entry-wise multiplication and summing the outputs to establish feature maps. These layers learn to recognize specific patterns and features from the input data.

3.6.2.3 Pooling Layers

Down sampling is performed here as pooling layers reduce the spatial dimensions of the feature maps, by making the model more robust to small variations in the input. For this purpose, it uses common pooling methods: Max pooling and average pooling are common techniques.

3.6.2.4 Activation Functions

The activation functions are used for the identification of the non-linearity into the model and allow it to learning for the more complex patterns as per requirement. Rectified Linear Unit (ReLU) is chosen as activation function here.

3.6.2.5 Fully Connected Layers

These are used for performing the predictions and classification by receiving the output from convolutional and pooling layers.

3.6.2.6 Output Layer

For the analysis of the prediction of oil consumption and price prediction, our work uses sigmoid function.

3.6.2.7 Training and Evaluation

This step includes a loss function which is used for measuring the difference between the model's predictions and the actual values. Here an optimization algorithm Adam is used for adjustment of the model's parameters to minimize the loss function. The model is trained on the training data, and its performance is evaluated on the testing data.

3.6.2.8 Model Evaluation

The model is evaluated on the metrics like RMSE, MAPE, Loss, and prediction. In Table 3.5, hyperparameters of CNN model are discussed, which are used in simulations.

TABLE 3.5: Hyper Parameters of CNN

Hyperparameters	Values
<i>n_conv – layer</i>	2
<i>kernal_size</i>	2
filters	64
activation function	relu
<i>max – pooling</i>	1
<i>dense_layer</i>	50
dropout	0.2

In table 3.5, hyperparameters of CNN are discussed, where *n_conv – layer* specifies the no. of convolutional layers used for feature extraction, *kernal_size* and filters are used for learnable features detection, activation function introduces non-linearity in the model, *max_pooling* layer is used for explicit down-sampling and dense and dropout layers are used for fully connected layers for feature extraction and dropout layers are used for regularization in order to prevent overfitting.

3.6.3 Artificial Neural Network

We use following steps for Building and training an artificial neural network including problem definition, preparation of the dataset, designing the network architecture by initializing some parameters, executing the forward and backward propagation, and validating the model by Ismail and Khamis [100].

The steps are mentioned below.

3.6.3.1 Problem definition and Data Collection

Palm oil consumption and prices are increasing day by day so it is necessary to make accurate future planning for the optimal resource allocation and utilization.

In the literature, many machine learning techniques are used for the automation of these tasks for the efficient delivery of the daily base activities. For this purpose, our work uses ANN and how it processes the existing data is mentioned below.

3.6.3.2 Data Collection

A labelled dataset of three years is used from Pakistan bureau of Statistics with the labels: years, months, oil consumption and price.

3.6.3.3 Data Preparation for the Training and Testing

It includes the following steps;

3.6.3.4 Data Cleaning

by handling missing values, outliers, and discrepancies in the dataset.

3.6.3.5 Data Transformation

by normalizing standardizing the dataset to enhance the training efficiency.

3.6.3.6 Data Splitting

by dividing the dataset into training, validation, and test sets.

3.6.3.7 Artificial Neural Network Architecture Design

In this steps, decision of the model's layers, neurons, activation functions, optimizers and loss functions are performed. All are explained as under.

3.6.3.8 Layers

It determines total number of input, hidden, and output layers used for prediction of the palm oil imports.

3.6.3.9 Neurons

It decides for the number of neurons in each layer for the efficient processing of the tasks.

3.6.3.10 Activation Functions

These functions are used in each layer; as relu is used in our case because it is most appropriate.

3.6.3.11 Optimizer:

It is used for the selection of an optimization algorithm as Stochastic Gradient Descent.

3.6.3.12 Loss Function:

It includes by choosing a loss function for measuring the difference between predicted and actual outcomes.

3.6.3.13 Initialization of the Parameters

It includes the weights and biases by assigning the initial random values to the network.

3.6.3.14 Implementation of the Forward and Backward Propagation

In forward propagation, the input data is forwarded to the network, layer by layer, in order to generate predictions. While in backward propagation, error (difference between predicted and actual outputs) values are calculated and propagated backward through the network for adjusting weights and biases.

3.6.3.15 Training the Neural Network

It is an iterative process where forward and backward propagations are performed using the training data to minimize the error. Multiple epochs are used for performing the iteration over the training data until the model converges. Model's monitoring is performed to maintain the smooth execution during training using metrics like accuracy or loss.

3.6.3.16 Evaluation of the Neural Network

It comprises of the test set and validation set. The validation data is used for assessing the model's behavior on unseen data while tuning hyper parameters.

Testing data evaluates the final model's behavior on the test set to determine an unbiased estimate of its generalization ability.

TABLE 3.6: Hyper Parameters of ANN

Hyperparameters	Values
<i>hidden_layer_sizes</i>	32, 64
<i>max_iter</i>	500
activation function	Relu
<i>random_state</i>	42
<i>early_stopping</i>	True

In Table 3.6, hyperparameters of ANN model are discussed, which are used in simulations. In Table 3.6, all parameters are having the same meaning as discussed in the previous section except *early_stopping* which is used for stopping the overfitting.

3.6.4 ARIMA

It is a statistical model which is used for time series analysis and forecasting analysis. This model uses the combined features of Autoregressive (AR) and Moving Average (MA) models along with an additional Integration (I) component to deal with non-stationary time series data.

The ARIMA models are applied widely by Reheem et al. [63] in order to recognize previous patterns and prediction of future information in different domains like weather prediction, finance, and economics. In our case it is used for the prediction of palm oil imports to Pakistan using the previous dataset. Now the detailed breakdown of the ARIMA prediction model is discussed along with its components. There are following components of an ARIMA model:

3.6.4.1 AR

It utilizes the past data of the palm oil import's in order to predict future values. It also assumes that the contemporary dataset is related to prior data.

3.6.4.2 I

It considers the non-stationary data by differencing the dataset values on different interval. Differencing eradicated the trends and seasonality by making the dataset more reliable.

3.6.4.3 MA

It follows the previous forecasting errors to predict future values. In addition, this also captures the trend between the current observations and previous forecasting errors.

3.6.5 Working of the Model

3.6.5.1 Data Preparation

In this step, it ensures that the dataset is stationary, means that its statistical properties (like mean and variance) do not change over regular interval of time. It is obtained by differencing the data.

3.6.5.2 Model Selection

The parameters (p, d, q) are selected to define the ARIMA model's layout. As these parameters represent the order of the autoregressive (p), integrated (d), and moving average (q) components, respectively.

3.6.5.3 Model Estimation

parameters of the model are estimated by statistical techniques, i.e., 2-point weighted average in our case.

3.6.5.4 Forecasting

Once the model is trained, it can be used to predict future values of the dataset.

In Table 3.7, hyperparameters of ARIMA model are discussed, which are used in simulations.

TABLE 3.7: Hyperparameters of ARIMA Model

Hyperparameters	Values
p	1
Q	1
D	1

In Table 3.7, p, q and d are used for specifying the number of lagged observations, number of lagged forecast errors and degree of differencing to obtain stationarity.

3.7 Model Evaluation

The performance of the forecasting models is evaluated using the following metrics:

3.7.1 RMSE

this measure highlights larger errors by giving them higher weights using the formula as discussed in Eq. 1.

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (3.1)$$

In above equation 3.1, \hat{y}_i are the predicted values, N are total number of observations and y_i are observed values.

3.7.2 MAPE

It provides a relative measure of error as a percentage as mentioned in Eq. 2.

$$\frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3.2)$$

In Eq. 3.2, In above equation 3.1, \hat{y}_i are the predicted values, N are total number of observations and y_i are observed values

3.7.3 R^2

: It measures how model fits for the dataset as discussed in Eq. 3.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.3)$$

In Eq. 3.3, the variables used are discussed as below; y_i is used for the actual values where \hat{y}_i is used for the predicted values.

3.7.4 Accuracy

The accuracy is determined by the following formula.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.4)$$

In Eq. 3.4, TP and TN are used for the true positive and true negative values whereas FP and FN are used for false positive and false negative values.

3.7.5 Precision

It is calculated by the below-mentioned formula.

$$Precision = \frac{TP}{TP + FP} \quad (3.5)$$

In Eq. 3.5, TP and FP are same as discussed for accuracy.

3.7.6 Oil quantity and Price Prediction

It evaluates the forecast accuracy relative to other schemes. These metrics were chosen to ensure a comprehensive evaluation of the models' performance across different aspects.

3.8 Tools and Software

The analysis and modeling were conducted using Python. Key libraries included: Scikit-learn: For SVR, RF implementation and hyper-parameter tuning. TensorFlow/Keras: For building and training ANN models. Data visualization was performed using Matplotlib and Seaborn.

3.9 Summary

This chapter provided the detailed discussion of the methodology employed for forecasting palm oil imports using ML techniques. Each model was selected and tailored based on its ability to address specific challenges in time-series forecasting, such as non-linearity and volatility. The following chapter will present the results and discuss the performance of these models based on the evaluation metrics outlined above.

Chapter 4

Results and Discussion

4.1 Introduction

This chapter presents the evaluation and analysis of the proposed machine and deep learning techniques/models namely RF, CNN, ANN and ARIMA. These models are used for the prediction of palm oil imports to Pakistan. These results are discussed in detail to assess the accuracy and reliability of each model along with their comparative performance parameters.

The price and monthly import quantity of palm oil are used as input and output parameters. Twelve years (2013-2024) dataset of palm oil imports are obtained from Pakistan Bureau of Statistics department. The details of the dataset are mentioned below.

4.2 Data Set of Oil Quantity and Price

The data set of oil consumption and price are taken from Pakistan bureau of statistics from 2013 to 2024 as mentioned in the reports [116] for the analysis of its total consumption and price prediction for the next three years using RF, ANN, CNN, and ARIMA. The same dataset is used for measuring the random fluctuations in the dataset in order to detect the overall trend of the palm oil imports to Pakistan. In this case, the statistical analyses are performed on the dataset using the 2-point moving average test.

4.3 Analysis of Twelve Years' Dataset

The palm oil quantity and price datasets are decomposed for visualizing the observed, seasonal, trending, and residual outcomes of the previous twelve years after preprocessing as shown in Fig. 4.1 and Fig. 4.2. Here, trending data follows the observed values accurately and seasonal values are followed by residual outcomes. So, the dataset is considered appropriate for the predictions of the next three years as no more unusual values are observed which may lead to sudden fluctuations in the results. These results are further normalized by performing the 2-point moving average test for setting the efficient decision making and automation in the current work. The dataset is configured in terms of palm oil import quantity and price features and visualized in two different diagrams for simplifying the current trends and seasonal fluctuations of oil quantity and price. As a result, some randomness is observed during the 2019, 2021, and 2023 years for maintaining the palm oil quantity imports as shown in Fig. 4.1. These fluctuations are refined by performing the statistical 2-point moving average test for maintaining the smoothness in the fluctuation patterns. This test is further evaluated in the diagrams in the upcoming subsections for the 2-points moving average test. In Fig. 4.2, the price rates are observed with the trends and seasonal variations. The price rates are found maximum from 2022 to 2023 year. Seasonal variations are same for all dataset. No more fluctuations are observed; therefore, the 2-points moving average is only applicable for the observed data, which would be performed in the next subsection. The machine learning and deep learning techniques are applied for the prediction of the palm oil imports in terms of its quantity and price for the effective policy making and enhancing the trading from Pakistan to the foreign countries as discussed in the Methodology previously. Basically, the need of the country for fulfilling the demanded quantity should be effectively handled. The previous methods of demand assessment are not suitable to predict the current demand of the palm oil quantity. Therefore, the abovementioned techniques should be used for efficient decision making. These techniques may suggest better ways to predict palm oil imports by fulfilling the total demand in terms of quantity with feasible expected pricing rates. The validation results are discussed in the upcoming subsections.

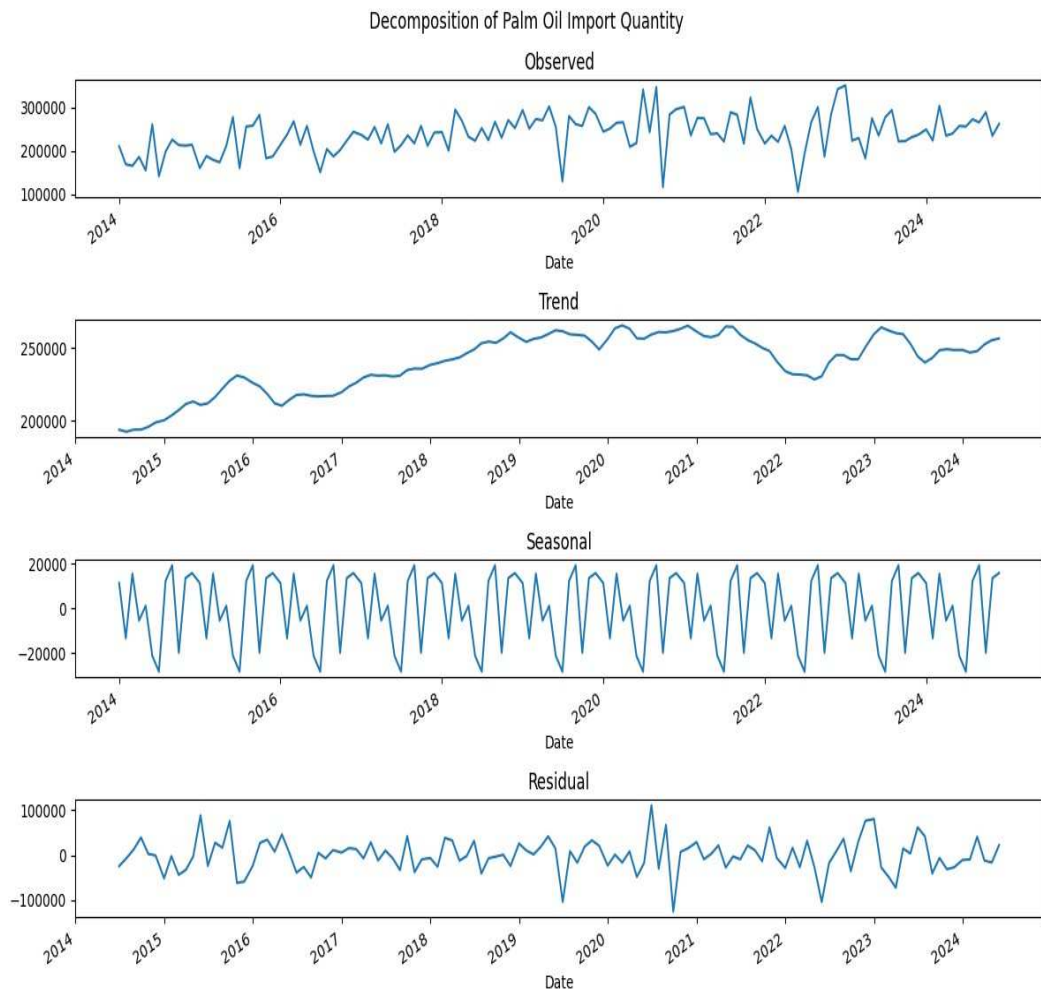


FIGURE 4.1: Decomposition of Palm Oil Quantity for Observed, Trending, Seasonal and Residual Data

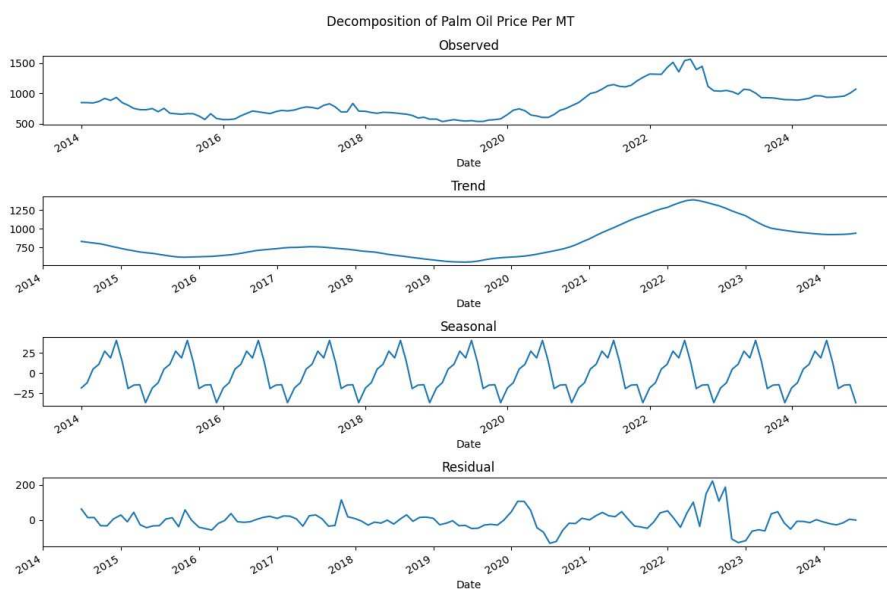


FIGURE 4.2: Decomposition of Palm Oil Price for Observed, Trending, Seasonal and Residual Data

4.4 Statistical Analysis using 2-Point Moving Average Test

In this section, the 2-point moving average results of the dataset are discussed after observing its actual trends with seasonal variations. This test is done in order to smooth the random fluctuations in the dataset for predicting the future trends. Thus, the random fluctuations in the dataset are normalized in order to predict the accurate trends for the effective policy and decision making as shown in Fig. 4.3 and Fig 4.4, where palm oil quantity and prices are normalized, respectively. The total quantity imported from the other countries is sufficiently balanced and expected price is minimal, which can be included in the effective budget planning and adding more budget will improve the resource planning.

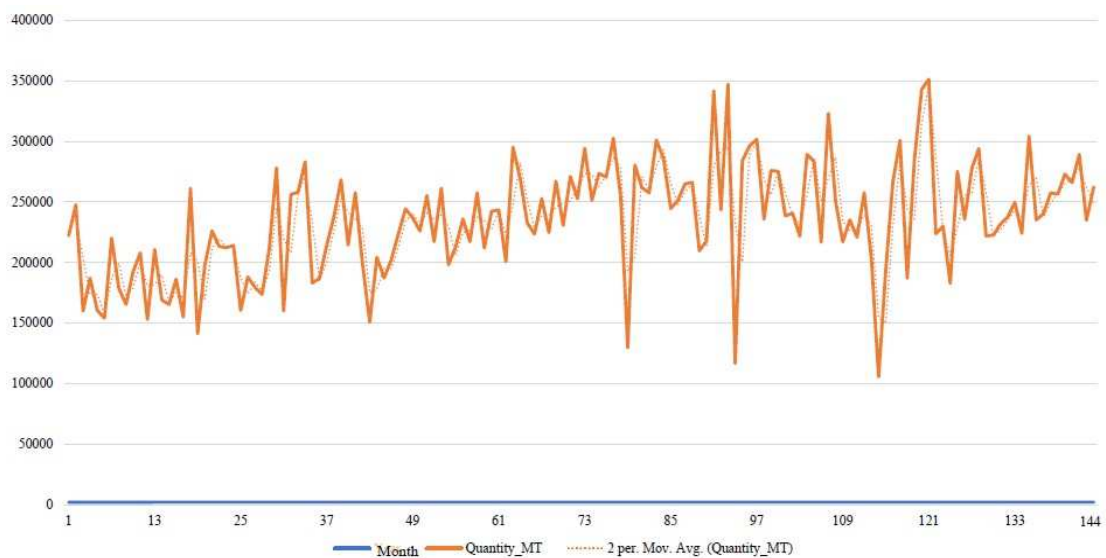


FIGURE 4.3: Two-point Moving Average for Palm Oil Imports with Actual Curve

4.5 Results of Machine Learning and Deep Learning Techniques

In this study four ML and DL techniques are used for the dataset analyses in order to make suitable strategic resource planning based on the prediction results. The MAPE, RMSE, R^2 , precision, and accuracy are used to analyze the performance of each model for prediction of the oil imports to Pakistan. The selected models

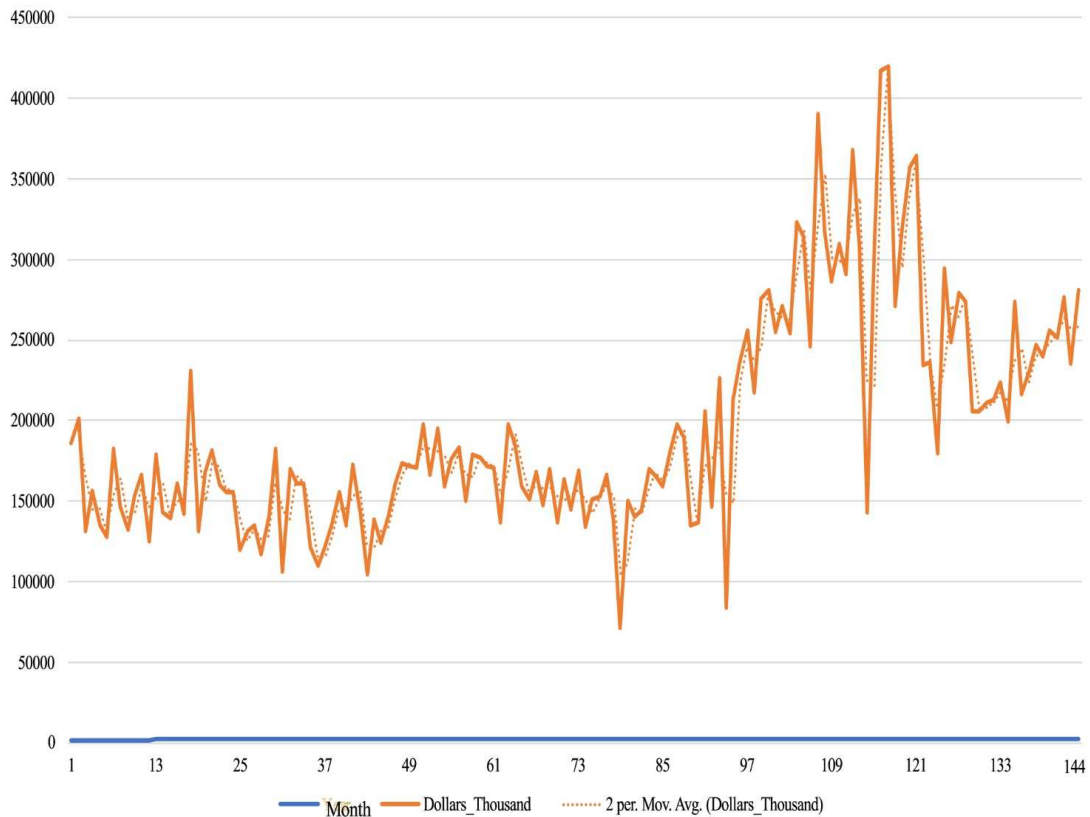


FIGURE 4.4: Two-point Moving Average for Palm Oil Imports Price with Actual Curve

for assessment are RF, ANN, CNN and ARIMA. For the measured values of the MAPE using the current dataset, the CNN outperforms all other models for computing the oil import quantity and price because it is efficient in nature in dealing with the complex nonlinear dataset. The RF and ARIMA are also performing well in evaluating the model; however, the ANN has high error rate as compared to others. The reason is that the RF model sometimes outperforms ANN in terms of MAPE due to its inherent characteristics of handling the time series data. The RF model has the collaborative methods, which may be considered more robust to overfitting and handle complex, non-linear relationships effectively while ANN is a simple feedforward networks, which may struggle with the sequential nature and potential non-linearity of the time series data as shown in Fig. 4.5. The MAPE results for price of all four (RF, ANN, CNN and ARIMA) techniques are displayed in Fig. 4.5. It can be observed that the CNN outclasses all other techniques because it shows minimum error rate; whereas, the ANN and RF show intermediate performance while ARIMA does not perform good in this case as it displays the

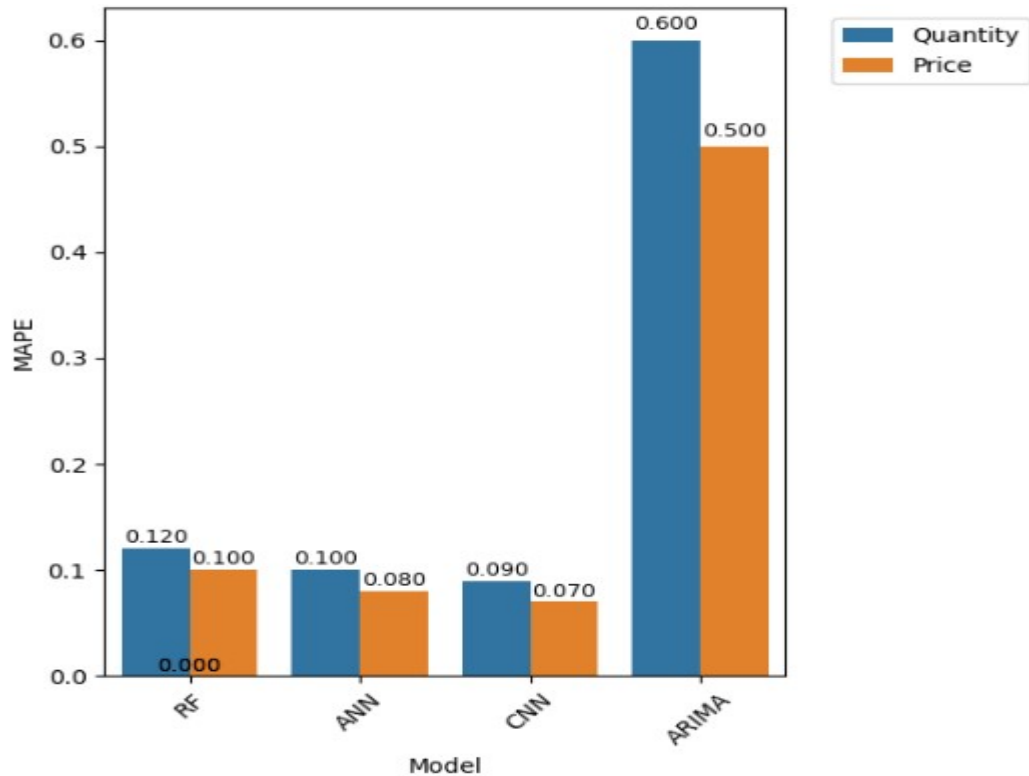


FIGURE 4.5: MAPE Results Palm Oil Import Quantity and Price of all four Techniques

highest error rate. The obtained MAPE values of quantity and price for the RF are 0.12, and 0.10, respectively. Similarly, the error rates for the ANN are 0.10 and 0.08 for quantity and price, correspondingly. The MAPE values of 0.90 and 0.70 are obtained for the CNN while the ARIMA exhibits error rates for the quantity and price as 0.60 and 0.50, respectively. In Fig. 4.6, the results of RMSE for the oil import quantity and price of all four techniques are visualized; where, the CNN again yields least error among all other techniques. The ANN does not perform good in this case as it displays huge error rate. The RMSE achieved by all techniques are discussed here. The RF and ANN show higher error rates compared to the CNN but lower error values than ARIMA.

In Fig. 4.7, the R^2 measure of all four techniques is can be noted. The R^2 measure indicates the presence of the variance of the dependent parameter, which is explained by the independent parameter in any regression model. In addition, it describes how well the specific model fits the dataset, considering the values ranging from 0 to 1. Here, the maximum value obtained by the CNN for both quantity and price are 0.85 and 0.84, respectively. So, the performance of the

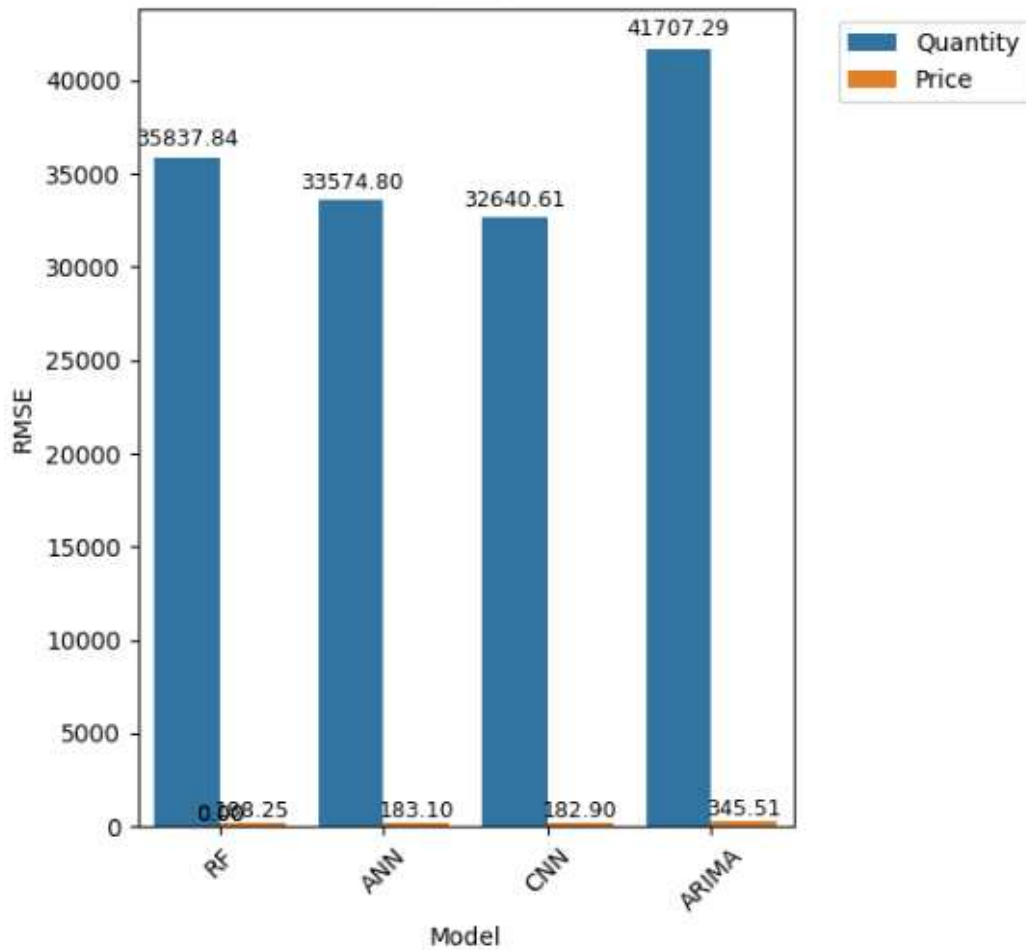


FIGURE 4.6: RMSE Results of Oil Quantity and Price of all Four Techniques Curve

CNN is the best among all techniques. The least (0.79 and 0.81) values of the R^2 are observed for the ARIMA model while the other two (RF and ANN) perform better as compared to the ARIMA. The quantity and price values of RF are 0.81 and 0.80, respectively. Similarly, the ANN values of quantity and price are 0.823 and 0.812, correspondingly. Therefore, it can be inferred that the CNN is the most appropriate model for the given set of palm oil data.

In Fig. 4.8, the accuracy of all models is visualized. As it can be noted that any model accuracy is described as a measure of making correct predictions. It is computed by dividing the number of accurate predictions by the total number of predictions. After that it is represented in a percentage form. As an instance, if a model accurately predicts 90 items out of 100 items, its accuracy is considered as 90%. Here the performance of the CNN is the best among all models in measuring

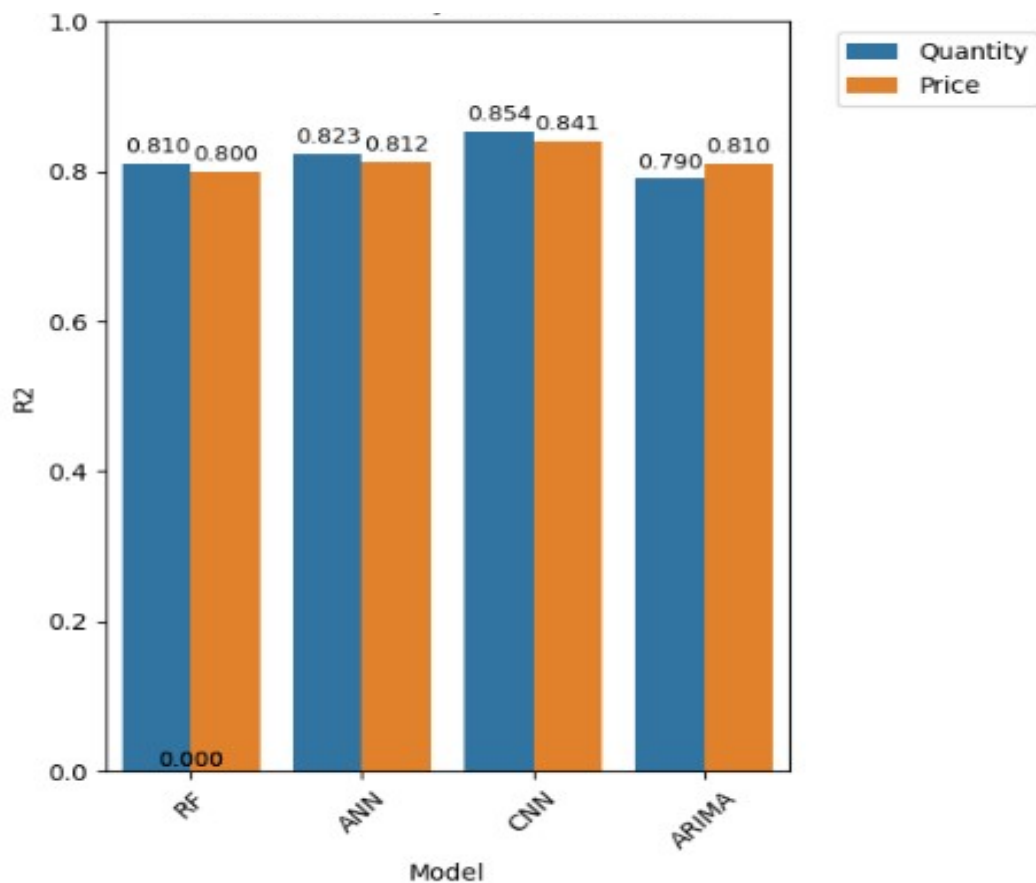


FIGURE 4.7: R^2 Results of all Four Techniques for Palm Oil Imports Quantity and Price

the accuracy of quantity (89%) and price (86%). Other models like RF and ANN are also performing well by obtaining the values in terms of price and quantity as 85 and 83%, 87 and 85%, respectively. However, the ARIMA is the least effective model, which obtained the quantity and price values 80 and 81%, correspondingly as shown in Figure 4.8.

In Figure 4.9, the precision values of all models are displayed. Again, the CNN outperforms all models in measuring both quantity and price as 0.94 and 0.93. However, the ARIMA is not performing well as it is not efficient as compared to other modern machine learning and deep learning techniques. Thus, the efficiency of the ARIMA is not up to the mark in the present case as it obtains the 0.82 and 0.81 values for both quantity and price. The other two techniques (RF and ANN) also perform well as the obtained the values for the quantity and price are 0.91 and 0.89, 0.92 and 0.91, respectively. In Fig. 4.10, the forecasted quantity values

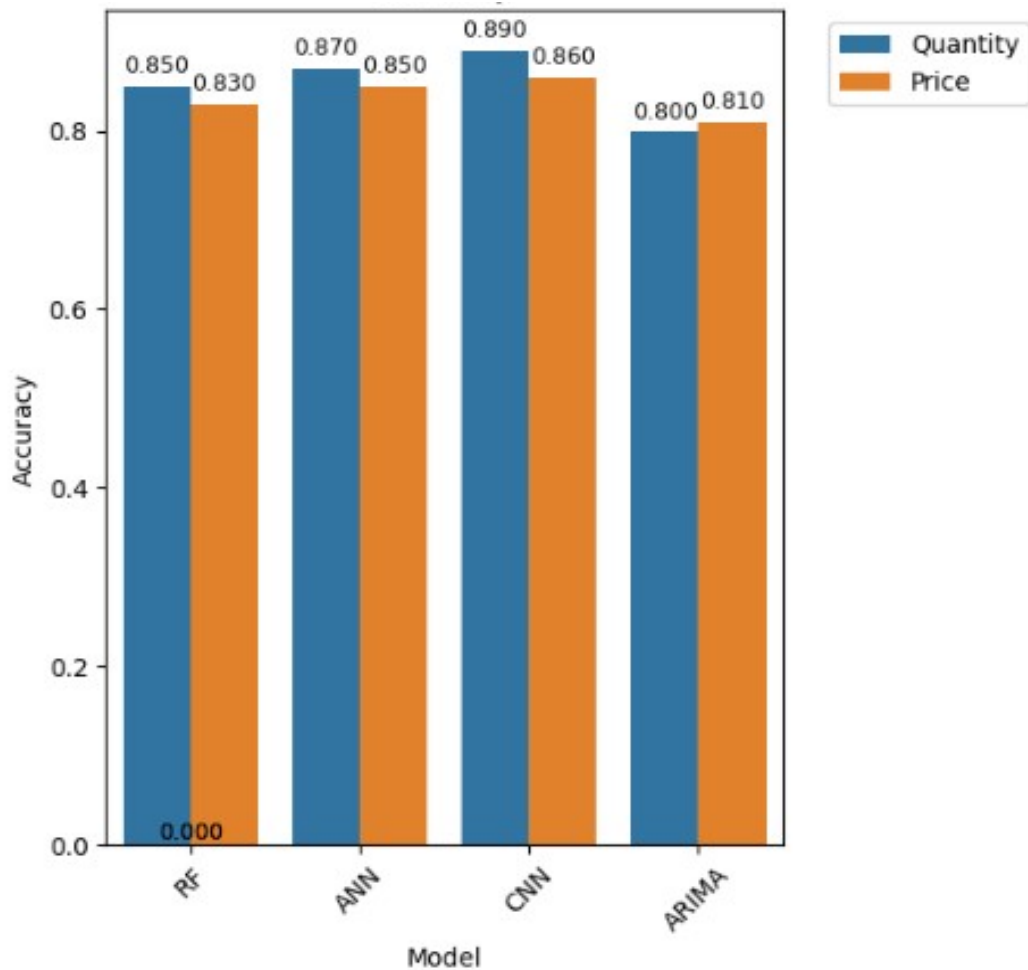


FIGURE 4.8: Accuracy Results of all Four Techniques for Palm Oil Imports Quantity and Price

of the palm oil are shown based on the historical data. As historical data displays some fluctuations due to randomness in oil quantity and also randomness exists in the behavior of deep learning and machine learning techniques. However, with accurate prediction, the future prediction is done in a very precise manner, which shows the efficient resource planning will lead to execute the process effectively in stock market. No high peaks are predicted in the future as shown in the current dataset. On the basis of this oil quantity planning, price prediction is also done strategically so that future resources will be fully utilized for the delivery of the trading tasks, which will also improve the economy of the country as shown in Fig. 4.11. Here, the CNN outperforms to all techniques because there are no peaks observed in the oil quantity forecast. All other techniques are relatively good but there are some fluctuations observed in the ARIMA case for the next

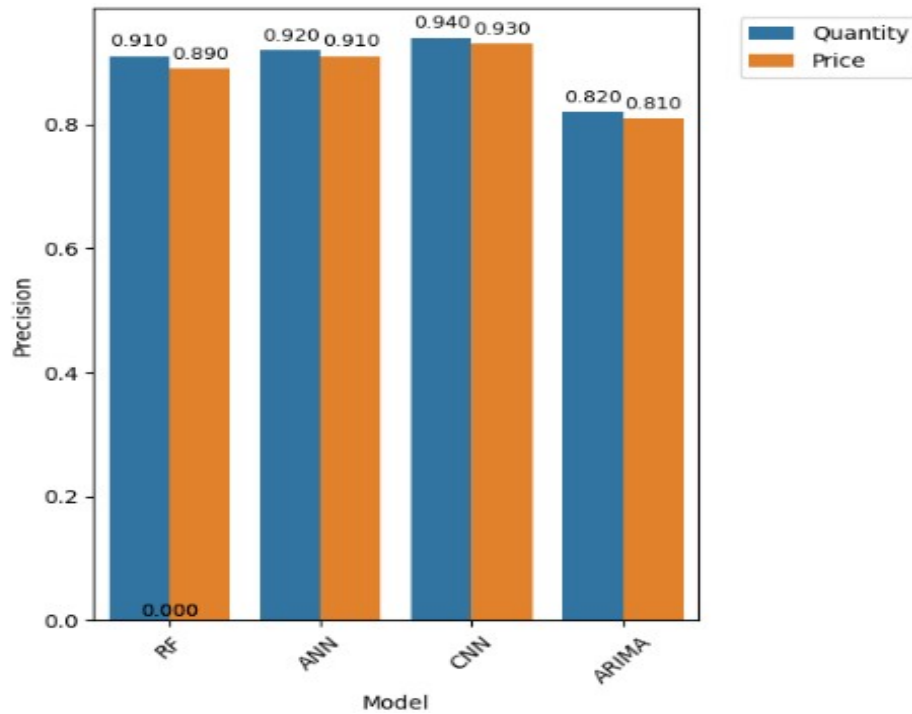


FIGURE 4.9: Precision Results of all Four Techniques for Palm Oil Imports Quantity and Price

three years. During the start of the year, demand is less from all techniques except ARIMA because it follows random variations throughout the whole year and sometime huge peaks are observed. However, machine learning techniques are intelligent to control the seasonal trend. No more peaks are observed during the three years timespan prediction.

4.6 Future Predictions of Oil Quantity and Price by All Techniques

Future predictions for the palm oil imports obtained by all methodologies are enlisted in Tables 4.1 and 4.2. Here, the CNN model predicts that feasible oil imports will be performed in the upcoming years as compared to all techniques with appropriate price increase, which is considered beneficial in terms of cost optimization. However, the CNN predicts the intermediate increase in the oil import with the minimum price so this is the best fit model in our case.

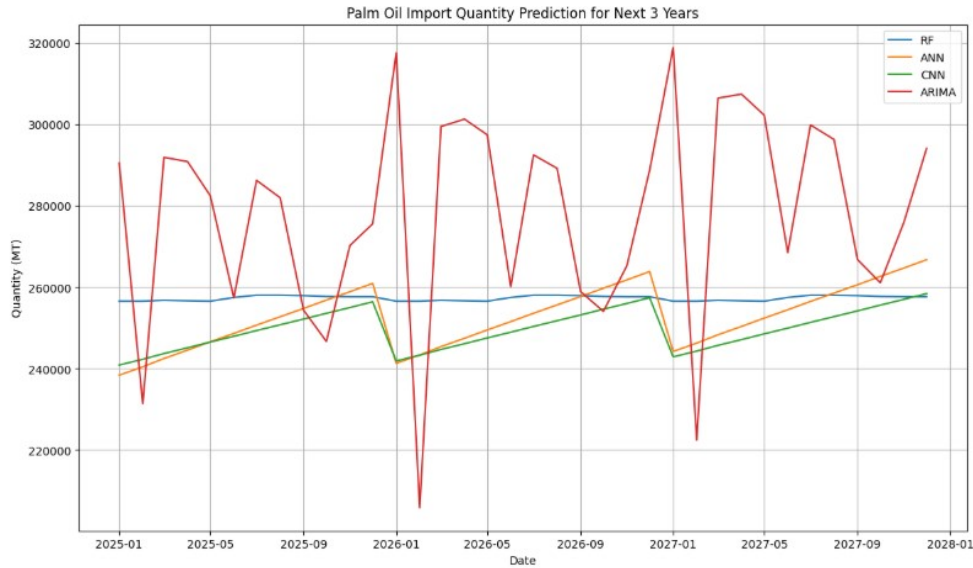


FIGURE 4.10: Prediction of Next Three Years’ Results for Oil Quantity of all Four Techniques

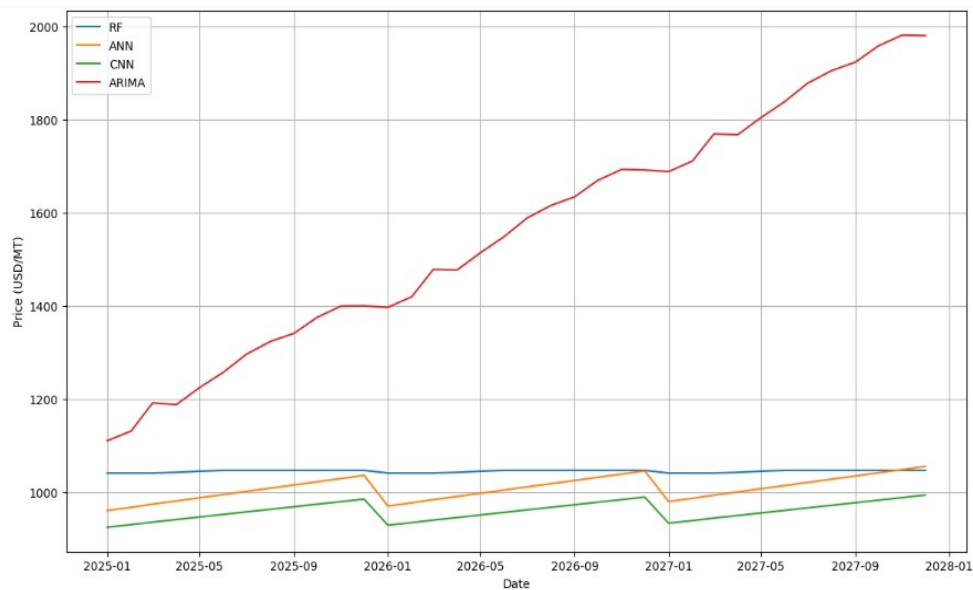


FIGURE 4.11: Prediction of Next Three Years’ Results for Oil Price of all Four Techniques

The ANN and RF also predicts the average increase in palm oil import to Pakistan but with little higher cost which is not feasible and needs cost reductions. The CNN is considered as the most economical model in this case, which predicts the feasible quantity of palm oil imports with minimal cost as shown in Tables 4.1 and 4.2, respectively. Prediction values for the next three years are approximately same as during the start of the year demand is low and it gradually increases from the mid of the year and price is also varying in the same manner.

TABLE 4.1: Future Predictions of Oil Quantity by All Techniques

Iter. No.	Date	Predicted Values of Quantity			
		RF	ANN	CNN	ARIMA
105	25/01/01	258722.93	237001.15	245902.56	290009.00
106	25/02/01	258722.93	240048.12	242695.90	270602.28
107	25/03/01	258439.40	243095.09	243695.90	239602.28
108	25/04/01	258408.27	244142.09	244092.56	292971.17
109	25/05/01	258119.70	244092.09	244092.25	292837.62
110	25/06/01	259025.98	248236.03	247885.93	279519.22
111	25/07/01	259391.43	255283.06	252282.56	259021.39
112	25/08/01	259826.67	256330.03	253679.25	282920.04
113	25/09/01	259772.51	258377.00	256075.90	25804271.11
114	25/10/01	259862.61	264423.96	257472.56	33604.29
115	25/11/01	258239.75	257470.96	257869.25	33604.29
116	25/12/01	2568239.75	248017.93	248017.93	270930.15
105	26/01/01	258722.93	237001.15	245902.56	290009.00
106	26/02/01	258722.93	240048.12	242695.90	270602.28
107	26/03/01	258439.40	243095.09	243695.90	239602.28
108	26/04/01	258408.27	244142.09	244092.56	292971.17
109	26/05/01	258119.70	244092.09	244092.25	292837.62
110	26/06/01	259025.98	248236.03	247885.93	279519.22
111	26/07/01	259391.43	255283.06	252282.56	259021.39
112	26/08/01	259826.67	256330.03	253679.25	282920.04
113	26/09/01	259772.51	258377.00	256075.90	25804271.11
114	26/10/01	259862.61	264423.96	257472.56	33604.29
115	26/11/01	258239.75	257470.96	257869.25	33604.29
116	26/12/01	2568239.75	248017.93	248017.93	270930.15
105	27/01/01	258722.93	237001.15	245902.56	290009.00
106	27/02/01	258722.93	240048.12	242695.90	270602.28
107	27/03/01	258439.40	243095.09	243695.90	239602.28
108	27/04/01	258408.27	244142.09	244092.56	292971.17
109	27/05/01	258119.70	244092.09	244092.25	292837.62
110	27/06/01	259025.98	248236.03	247885.93	279519.22
111	27/07/01	259391.43	255283.06	252282.56	259021.39
112	27/08/01	259826.67	256330.03	253679.25	282920.04
113	27/09/01	259772.51	258377.00	256075.90	25804271.11
114	27/10/01	259862.61	264423.96	257472.56	33604.29
115	27/11/01	258239.75	257470.96	257869.25	33604.29
116	27/12/01	2568239.75	248017.93	248017.93	270930.15

TABLE 4.2: Future Predictions of Oil Price by All Techniques

Iter. No.	Date	<u>Predicted Values of Price</u>			
		RF	ANN	CNN	ARIMA
105	25/01/01	1050.18	946.18	846.08	1100.28
106	25/02/01	1050.18	950.30	850.25	1130.04
107	25/03/01	1050.18	954.83	854.30	1200.47
108	25/04/01	1050.18	959.70	859.70	1200.63
109	25/05/01	1050.18	964.74	864.74	1222.20
110	25/06/01	1050.10	969.94	869.94	1230.49
111	25/07/01	1050.10	1004.13	904.62	1280.98
112	25/08/01	1050.10	1020.13	920.32	1300.29
113	25/09/01	1050.10	1030.13	930.02	1350.28
114	25/10/01	1050.10	1035.13	935.72	1360.93
115	25/11/01	1050.10	1030.13	930.42	1400.04
116	25/12/01	1050.10	980.27	904.13	1400.99
105	25/01/01	1050.18	946.18	846.08	1405.28
106	26/02/01	1050.18	950.30	850.25	1425.04
107	26/03/01	1050.18	954.83	854.30	1430.47
108	26/04/01	1050.18	959.70	859.70	1450.63
109	26/05/01	1050.18	964.74	864.74	1450.20
110	26/06/01	1050.10	969.94	869.94	1500.49
111	26/07/01	1050.10	1004.13	904.62	1530.98
112	26/08/01	1050.10	1020.13	920.32	1550.29
113	26/09/01	1050.10	1030.13	930.02	1650.28
114	26/10/01	1050.10	1035.13	935.72	1670.93
115	26/11/01	1050.10	1030.13	930.42	1700.04
116	26/12/01	1050.10	980.27	904.13	1690.99
105	27/01/01	1050.18	946.18	846.08	1690.28
106	27/02/01	1050.18	950.30	850.25	1698.04
107	27/03/01	1050.18	954.83	854.30	1790.47
108	27/04/01	1050.18	959.70	859.70	1790.63
109	27/05/01	1050.18	964.74	864.74	1800.20
110	27/06/01	1050.10	969.94	869.94	1810.49
111	27/07/01	1050.10	1004.13	904.62	1830.98
112	27/08/01	1050.10	1020.13	920.32	1850.29
113	27/09/01	1050.10	1030.13	930.02	1950.28
114	27/10/01	1050.10	1035.13	935.72	1960.93
115	27/11/01	1050.10	1030.13	930.42	1970.04
116	27/12/01	1050.10	980.27	904.13	1980.99

A comprehensive analysis of the machine and deep learning forecasting models is presented to predict palm oil imports to Pakistan by considering its quantity and price. These results also underscore the effectiveness of machine and deep learning models, particularly the CNN and ANN, in capturing complex patterns and achieving high efficiency in predicting the oil quantity at the minimum price. The CNN predicts the intermediate imports of palm oil quantity with minimum expected price; however, the ANN and RF predict more imports in terms of quantity with slightly higher expected price rates. Tradeoff exists in price and quantity of both techniques. The ARIMA is predicting large amount of palm oil imports in terms of quantity with high expected price rates. So, this is not performing well for this dataset. The discussion also highlights the need for continuous innovation in forecasting techniques to address the challenges posed by nonlinear and volatile datasets.

Chapter 5

Conclusions and Future Recommendations

Based on the results and discussion, it can be inferred that latest machine learning techniques can provide better forecasts as compared to conventional techniques, which could improve the policies, decision making, and resource utilization identifying optimum import quantities and prices of the palm oil. The two-point weighted average technique is used in order to normalize the dataset. Additionally, four machine and deep learning (RF, ANN, CNN and ARIMA) techniques are employed to predict the palm oil import to Pakistan. As CNN is hugely used for the processing of image data but few authors used it for the processing of time-series data. Focusing on its application, it is used for the forecasting of time-series data in our work. The results of all techniques are compared to identify the best technique/model for the prediction of the next three years palm oil imports to Pakistan. The quantity and price of the palm oil are analyzed using the MAPE, RMSE, R^2 , precision, and accuracy values to assess the most suitable model for the palm import to Pakistan. As a result, it is noted that the CNN model outperforms in most of the cases.

5.1 Conclusions

All of the considered objective are achieved, however, there are few tradeoffs existed in the results. As in most of the cases, the CNN model yields outstanding results for the considered dataset; however, for the price prediction, the ANN

shows the minimum quantity of palm oil with high price rate. Thus, there exists a trade off between price and oil quantity. The considered parameters are achieved successfully and discussed in detail in the results and discussion section. The machine and deep learning techniques are implemented successfully on the selected dataset to outline the palm oil forecasts and to find out the most suitable and effective model for testing and validation. The CNN is emerged as the best model among all the tested models. As it predicts intermediate quantity of palm oil and price for the future imports to Pakistan. It helps in effective trading, decision making, and policy making processes. In-short, after performing the extensive simulations and analyses, it can be concluded that the latest machine and deep learning techniques/models provide better forecasts as compared to the classical models such as ARIMA. The results of the CNN model are more accurate, reliable, and consistent to determine the quantity and price of the palm oil to predict imports to Pakistan. Therefore, it is recommended to used the CNN model's forecasts for future decisions, policies development, resource planning.

5.2 Implications/Limitations of Work

The implications and limitations of the study are discussed as below.

- a) As the CNN is identified as the best model for accurate quantity and price forecasting in the current work because it can detect local and short-term temporal patterns in the considered dataset. In palm oil import dataset, these patterns may arise due to seasonal fluctuations, policy changes, and trade disruptions.
- b) The CNN automatically learns and extracts required features from raw input data. As there are few challenges in manual feature engineering for economic forecasting. The CNN reduces dependencies on expert defined features and may identify complex nonlinear interconnections among variables like price, quantity, and external economic indicators.
- c) The CNN improves computational efficiency as it is efficient to train as compared to the many other RNN-based models (like LSTM), especially with large datasets. In addition, it provides an efficient and scalable solution

to forecast palm oil imports across multiple countries or regions using high resolution datasets.

- d) The CNN provides suitability for multivariate forecasting because it can handle multiple input channels (e.g., price, rainfall, currency exchange) simultaneously and palm oil imports are influenced by many variables. The CNN enables multivariate forecasting by effectively learning from multiple data sources concurrently.
- e) The CNN poses few limitations as well. It may struggle with long-term dependencies unless paired with other memory-based models like LSTM. It also requires high-quality and large datasets for effective training. It is less interpretable as compared to the statistical modeling techniques.

5.3 Future Recommendations

As it is noted that the CNN is the best model, which helps in accurate forecasting palm oil imports to Pakistan. In future, the hybrid models may be investigated to assess the palm oil forecasts, where classical models can be combined with machine learning and deep learning models like ARIMA and GARCH with LSTM and ANN model for making the efficient decision making. These models are also capable to capture the both linear and non-linear patterns for enhancing the forecast accuracy. Secondly, the advanced deep learning architectures can be employed because these are considered well for the learning the long-term dependencies and for processing sequential data as GRU and LSTM are considered for time series forecasting, transformer-based models such as temporal fusion transformers and LSTM-CNN for the extraction of spatial and temporal features, etc. Thirdly, explainable AI and attention mechanisms can be added to improve transparency and aid stakeholders to understand the most critical factors for the palm oil imports. By integrating the attention layers in deep learning models for focusing on significant time intervals, the SHAP or LIME is recommended for interpreting machine learning predictions. Fourthly, the robust validation mechanisms can be implemented as rolling cross-validation, Time Series Split, and walk-forward validation as these techniques provide a more realistic evaluation for the time series data.

In addition, indicators like incorporation of Policy and Sustainability Variables can also be included for improving the impacts on agricultural imports and other market trends.

5.4 Future Directions

The use of the CNN for palm oil forecasting can improve the computational speed, pattern recognition, and multivariate analysis capabilities. As it offers significant solutions, especially when integrated with hybrid architecture. Therefore, in future, it should be applied with attention to data quality and temporal structure for obtaining the best results.

Bibliography

- [1] Gourichon, H. (2019). “Analysis of incentives and disincentives for palm oil in Nigeria”. Monitoring African Food and Agricultural Policies project (MAFAP), <https://gatesopenresearch.org/documents/3-580/pdf>.
- [2] Zahraee, S. M., Shiwakoti, N., & Stasinopoulos, P. (2022). “Agricultural biomass supply chain resilience: COVID-19 outbreak vs. sustainability compliance, technological change, uncertainties, and policies”. *Cleaner Logistics and Supply Chain*, 4, 2022, 100049. <https://doi.org/10.1016/j.clscn.2022.100049>.
- [3] Yee, K. W., & Samsudin, H. B. (2021). “Comparison between Artificial Neural Network and ARIMA Model in Forecasting Palm Oil Price in Malaysia”. *International Journal of Scientific Engineering and Science*, 5(11), 12-15.
- [4] Tambe, E. B., Anukwonke, C. C., Mbuka-Nwosu, I. E., & Abazu, C. I. (2023). “Changes in the Agriculture Sector That Are Essential to Mitigate and Adapt to Climate Changes”. In *Strategizing Agricultural Management for Climate Change Mitigation and Adaptation* (pp. 89-112). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-32789-6_6.
- [5] Ahmad, N. F. Y., Khalid, K., Rusiman, M. S., Kamardan, M. G., Roslan, R., & Che-Him, N. (2018, April). “Analysis Monthly Import of Palm Oil Products Using Box-Jenkins Model”. In *Journal of Physics: Conference Series* (Vol. 995, No. 1, p. 012012). IOP Publishing. <https://doi.org/10.1088/1742-6596/995/1/012012>.

- [6] Khan, N., Kamaruddin, M. A., Ullah Sheikh, U., Zawawi, M. H., Yusup, Y., Bakht, M. P., & Mohamed Noor, N. (2022). "Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow". *Plants*, 11(13), 1-19.
- [7] Sadouk, L. (2019). "CNN Approaches for Time Series Classification". *Time Series Analysis - Data, Methods, and Applications*, 57. IntechOpen. doi: 10.5772/intechopen.81170.
- [8] Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). "A CNN-LSTM model for gold price time-series forecasting". *Neural computing and applications*, 32(23), 17351-17360.
- [9] Jin, X., Yu, X., Wang, X., Bai, Y., Su, T., & Kong, J. (2019, December). "Prediction for Time Series with CNN and LSTM". In *Proceedings of the 11th international conference on modelling, identification and control (ICMIC2019)* (pp. 631-641). Singapore: Springer Singapore.
- [10] Wibawa, A. P., Utama, A. B. P., Elmunsyah, H., Pujianto, U., Dwiyanto, F. A., & Hernandez, L. (2022). "Time-series Analysis with Smoothed Convolutional Neural Network". *Journal of big Data*, 9(1), 1-18. <https://doi.org/10.1186/s40537-022-00599-y>.
- [11] Bi, J., Li, E., & Luo, Y. (2023). "Petroleum Price Prediction Based on the Linear Regression and Random Forest". *Applied and Computational Engineering*, 8, 292-296. <https://doi.org/10.54254/2755-2721/8/20230170>.
- [12] Hermantoro, H., & Rudyanto, R. (2018). "Modeling and Simulation of Oil Palm Plantation Productivity Based on Land Quality and Climate Using Artificial Neural Network". *International Journal of Oil Palm*, 1(2), 65-70. <https://ijop.id/index.php/ijop/article/view/9>.
- [13] Chamalwa, H. A., Abdulsalam, B., & Abbas, M. (2023). "Modeling and Forecasting Crude Oil Prices in Nigeria Using ARIMA: A Time Series Analysis from 2013-2022". *UMYU Scientifica*, 4(1), 45-56. <https://doi.org/10.56919/usci.2433.033>.

- [14] Niu, S., & Li, H. (2022). "A Comparative Machine Learning Study for Time Series Oil Production Forecasting: ARIMA, LSTM, and Prophet". *Computers & Geosciences*, 164, 105126. <https://doi.org/10.1016/j.cageo.2022.105126>.
- [15] Dimitriadou, A., Gogas, P., Papadimitriou, T., & Plakandaras, V. (2018). "Oil market efficiency under a machine learning perspective". *Forecasting*, 1(1), 157-168.
- [16] Hosseini, S., & Akilan, T. (2023). "Advanced deep regression models for forecasting time series oil production". arXiv preprint arXiv:2308.16105.
- [17] Alkhamash, E. H. (2022). "An Optimized Gradient Boosting Model by Genetic Algorithm for Forecasting Crude Oil Production". *Energies*, 15(17), 1-13. <https://doi.org/10.3390/en15176416>.
- [18] Tardini, G. A. (2024). "Selection of modelling for forecasting crude palm oil prices using deep learning (GRU & LSTM)". *Emerging Science Journal*, 8(3), 232-245. <https://doi.org/10.28991/esj-2024-03-019>.
- [19] Husin, N. A., Khairunniza-Bejo, S., Abdullah, A. F., Kassim, M. S. M., Ahmad, D., & Aziz, M. H. A. (2020). "Classification of Basal Stem Rot Disease in Oil Palm Plantations using Terrestrial Laser Scanning Data and Machine Learning". *Agronomy*, 10(11), 1624. <https://doi.org/10.3390/agronomy10111624>.
- [20] Butt, S., Ramakrishnan, S., Chohan, M. A., & Punshi, S. K. (2019). "Prediction of Malaysian Exchange Rate using Microstructure Fundamental and Commodities Prices: A Machine Learning Method". *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2).
- [21] Tahyudin, I., Tikaningsih, A., Suryana, Y., Nugroho, H. A., Nurhopipah, A., Sasongko, N. A., ... & Yanuar, A. I. (2023). "Optimizing Sustainability: A Deep Learning Approach on Data Augmentation of Indonesia Palm Oil Products Emission." *Research Square*. 1-32.

- [22] Khan, N., Kamaruddin, M. A., Sheikh, U. U., Yusup, Y., & Bakht, M. P. (2021). "Oil Palm and Machine Learning: Reviewing one Decade of Ideas, Innovations, Applications, and Gaps". *Agriculture*, 11(9), 832.
- [23] Suppalakpanya, K., Nikhom, R., Booranawong, T., & Booranawong, A. (2019). "Forecasting Oil Palm and Crude Palm Oil Data in Thailand using Exponential Time-series Methods". *Engineering and Applied Science Research*, 46(1), 44-55.
- [24] Ata, A. (2024). "A Composite Forecast Approach for Market Demand Using Regression with Simulation to Predict the Palm Oil Market." Accessed on April 20th, 2025 [Online], Available at: <https://asb.edu.my>.
- [25] Chong, K. L., Kanniah, K. D., Pohl, C., & Tan, K. P. (2017). "A Review of Remote Sensing Applications for Oil Palm Studies". *Geo Spatial Information Science*, 20(2), 184–200. <https://doi.org/10.1080/10095020.2017.1337317>.
- [26] Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction". *IEEE Access*, 9, 63406–63439. <https://doi.org/10.1109/ACCESS.2021.3075159>.
- [27] Karia, A. A., & Bujang, I. (2011). "Progress accuracy of CPO Price Prediction: Evidence from ARMA Family and Artificial Neural Network Approach". *International Research Journal of Finance and Economics*, 64(64), 66-79.
- [28] Rahim, N. F., Othman, M., & Sokkalingam, R. (2018). "A Comparative Review on Various Method of Forecasting Crude Palm Oil Prices". *Journal of Physics: Conference Series*, 1123, 012043. <https://doi.org/10.1088/1742-6596/1123/1/012043>.
- [29] Hansen, S. B., Padfield, R., Syayuti, K., Evers, S., Zakariah, Z., & Mastura, S. (2015). "Trends in Global Palm Oil Sustainability Research". *Journal of Cleaner Production*, 100, 140–149. <https://doi.org/10.1016/j.jclepro.2015.03.051>.

- [30] Thompson, M., H., Gaw, S., & Robinson, B. (2022). “An Assessment of Trace Element Accumulation in Palm Oil Production”. *Sustainability*, 14(8), 4553. <https://doi.org/10.3390/su14084553>.
- [31] Kim, N., Ha, K.-J., Park, N.-W., Cho, J., Hong, S., & Lee, Y.-W. (2019). “A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015”. *ISPRS International Journal of Geo-Information*, 8(5), 240. <https://doi.org/10.3390/ijgi8050240>.
- [32] Holzman, M. E., Carmona, F., Rivas, R., & Niclos, R. (2018). “Early Assessment of Crop Yield from Remotely Sensed Water Stress and Solar Radiation Data”. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 297–308. <https://doi.org/10.1016/j.isprsjprs.2018.03.014>.
- [33] Chlingaryan, A., Sukkariéh, S., & Whelan, B. (2018). “Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review”. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>.
- [34] Wang, X., Huang, J., Feng, Q., & Yin, D. (2020). “Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches”. *Remote Sensing*, 12(11), 1744. <https://doi.org/10.3390/rs12111744>.
- [35] Shiu, Y.-S., & Chuang, Y.-C. (2019). “Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models”. *Remote Sensing*, 11(2), 111. <https://doi.org/10.3390/rs11020111>.
- [36] Lin, D., Wei, R., & Xu, L. (2019). “An Integrated Yield Prediction Model for Greenhouse Tomato”. *Agronomy*, 9(12), 873. <https://doi.org/10.3390/agronomy9120873>.
- [37] PS., M. G., & R., B. (2019). “Performance Evaluation of Best Feature Subsets for Crop Yield Prediction using Machine Learning Algorithms”. *Applied Artificial Intelligence*, 33(7), 621–642. <https://doi.org/10.1080/08839514.2019.1592343>.

- [38] Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Robin Bryant, C., & Senthilnath, J. (2021). "Integrated Phenology and Climate in Rice Yields Prediction using Machine Learning Methods". *Ecological Indicators*, 120, 106935. <https://doi.org/10.1016/j.ecolind.2020.106935>.
- [39] Tong, S. C., Tang, T. K., & Lee, Y. Y. (2021). "A Review on the Fundamentals of Palm Oil Fractionation: Processing Conditions and Seeding Agents". *European Journal of Lipid Science and Technology*, 123(12), 2100132.
- [40] Haw, Y. H., Lai, K. W., Chuah, J. H., Bejo, S. K., Husin, N. A., Hum, Y. C., Yee, L., Tee, C. A. T. H., Ye, X., & Wu, X. (2023). "Classification of Basal Stem Rot using Deep Learning: a Review of Digital Data Collection and Palm Disease Classification Methods". *PeerJ Computer Science*, 9, e1325. <https://doi.org/10.7717/peerj-cs.1325>.
- [41] Haw, Y. H., Hum, Y. C., Chuah, J. H., Voon, W., Khairunniza-Bejo, S., Husin, N. A., Yee, L., & Lai, K. W. (2023). "Detection of Basal Stem Rot Disease using Deep Learning". *IEEE Access*, 11, 49846–49862. <https://doi.org/10.1109/ACCESS.2023.3277082>
- [42] Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., & Li, Z. (2020). "Identifying the Contributions of Multi-source Data for Winter Wheat Yield Prediction in China". *Remote Sensing*, 12(5), 750. <https://doi.org/10.3390/rs12050750>.
- [43] Elavarasan, D., Vincent, P. M. D. R., Srinivasan, K., & Chang, C. Y. (2020). "A Hybrid CFS Filter and RF-RFE Wrapper-based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling". *Agriculture*, 10(9), 400. <https://doi.org/10.3390/agriculture10090400>
- [44] Firdawanti, A. R., Sumertajaya, I. M., & Sartono, B. (2020, August). "Random Forest Lag Distributed Regression for Forecasting on Palm Oil Production". *Proceedings of the 1st International Conference on Statistics and Analytics* (pp. 1–6). doi: 10.4108/eai.2-8-2019.2290493.
- [45] Ang, Y., Shafri, H. Z. M., Lee, Y. P., Bakar, S. A., Abidin, H., Mohd Junaidi, M. U. U., ... & Samad, M. N. A. (2022). "Oil Palm Yield

- Prediction Across Blocks from Multi-source Data using Machine Learning and Deep Learning”. *Earth Science Informatics*, 15(4), 2349–2367. <https://doi.org/10.1007/s12145-022-00800-1>.
- [46] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). “Random Forests for Global and Regional Crop Yield Predictions”. *PLOS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- [47] Jamshidi, E. J., Yusup, Y., Hooy, C. W., Kamaruddin, M. A., Hassan, H. M., Muhammad, S. A., ... & Tan, C. C. (2024). “Predicting Oil Palm Yield using a Comprehensive Agronomy Dataset and 17 Machine Learning and Deep Learning Models.” *Ecological Informatics*, 81, 102595. <https://doi.org/10.1016/j.ecoinf.2024.102595>
- [48] Gupta, K. K., Kalita, K., Ghadai, R. K., Ramachandran, M., & Gao, X. Z. (2021). “Machine Learning-based Predictive Modelling of Biodiesel Production—A Comparative Perspective”. *Energies*, 14(4), 1122. <https://doi.org/10.3390/en14041122>.
- [49] Ang, Y., Shafri, H. Z. M., Lee, Y. P., Bakar, S. A., Abidin, H., Hashim, S. J., ... & Gibril, M. B. A. (2025). “Block-scale Oil Palm Yield Prediction using Machine Learning Approaches based on Landsat and MODIS Satellite Data”. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 45(1), 90–107.
- [50] Wijaya, S., & Fauziah, F. (2024). “A Comparative Analysis of Linear Regression and Random Forest Methods for Predicting Palm Oil Price Sales”. *Jurnal Ilmiah Teknologi dan Rekayasa*, 28(3), 265–274.
- [51] Putra, R. P. A., Widodo, A. T., & Wibowo, A. (2020). “Estimation of Nitrogen Content in Oil Palm using RGB Aerial Photography and Machine Learning”. *IOP Conference Series: Earth and Environmental Science*, 484(1). <https://doi.org/10.1088/1755-1315/484/1/012122>.
- [52] Sboui, T., Saidi, S., & Lakti, A. (2023). “A Machine Learning-based Approach to Predict Deforestation Related to Oil Palm: Conceptual

- Framework and Experimental Evaluation”. *Applied Sciences*, 13(3), 1772. <https://doi.org/10.3390/app13031772>.
- [53] Wang, Y., Zhu, X., & Wu, B. (2020). “Automatic Detection of Individual Oil Palm Trees from UAV Images using HOG Features and An SVM Classifier”. *Remote Sensing*, 12(6), 902.
- [54] Muhammad Asraf, H., Md Tahir, N., & Hussain, A. (2018). “An Agricultural Tele-monitoring Method in Detecting Nutrient Deficiencies of Oil Palm Leaf”. *Int. J. Eng. Technol*, 7, 227-230.
- [55] Shaharum, N. S. N., Mohd Shafri, H. Z., Wan Azlina, W. A. K. G., Samsatli, S., Abdulrahman, M. M. A. A., & Yusuf, B. (2021). “Oil Palm Mapping over Peninsular Malaysia using Google Earth Engine and Machine Learning Algorithms”. *Remote Sensing Applications: Society and Environment*, 21, 100456.
- [56] Chen, Z., & Liao, I. Y. (2019, December). “Evaluation of Feature Extraction Methods for Classification of Palm Trees in DAV Images”. *International Conference on Computer and Drone Applications (IConDA)*, IEEE. doi: 10.1109/IConDA47345.2019.9034913.
- [57] Hossain, K. Z. (2024). “Predicting the Demographic Future of Bangladesh: Application and Comparison of ARIMA and Combined Population Forecasts”. *Institute of Statistical Research and Training, University of Dhaka Dhaka, Bangladesh. Romanian Statistical Review*, (2).
- [58] Latha, K. N., Rao, V. S., Sarada, C., Reddy, A. A., & Sreenivasulu, K. N. (2024). “Enhancing Forecasting Accuracy of Palm Oil Import to India using Machine Learning Techniques”. *Journal of Scientific Engineering and Technology*, 2(3), 12–25. <https://doi.org/10.63040/25827510.2024.02.003>.
- [59] Srinivasa Rao, V., Sarada, C., Reddy, A. A., & Sreenivasulu, K. N. (2022). “Modeling of Consumption of Gasoline (MS) and Diesel (HSD) in India using SARIMA and Neural Network”. *Energy Policy*, 158, 112574. <https://doi.org/10.1016/j.enpol.2021.112574>.

- [60] Gupta, K. K., Kalita, K., Ghadai, R. K., Ramachandran, M., & Gao, X.-Z. (2021). "Machine Learning-based Predictive Modelling of Biodiesel Production - A Comparative Perspective". *Energies*, 14(4), 1122. <https://doi.org/10.3390/en14041122>.
- [61] Zavira, R. (2020). "Production Planning of Crude Palm Oil: A Study Case at X Co". *IOP Conference Series: Earth and Environmental Science*, 472. <https://doi.org/10.1088/1755-1315/472/1/012047>.
- [62] Herdiyanti, Y., & Sukiyono, K. (2023). "Forecasting of Palm Oil Fruit Fresh Bunches (FFB) Prices in National and Bengkulu Province: ARIMA Model Application". *Agrisocionomics: Jurnal Sosial Ekonomi Pertanian*, 7(1), 1–10. <https://doi.org/10.14710/agrisocionomics.v7i1.14730>.
- [63] Reheem, A., Akbar, A., Saleem, N., & Afzaal, H. M. (2023). "Time Series Analysis and Forecasting on Import of Palm Oil in Pakistan". *Applications of Mathematical Sciences*, 2(1), 20-41.
- [64] Thambiah, S. (2017). "Price Forecasting Methodology of the Malaysian Palm Oil Market". *The International Journal of Applied Economics and Finance*, 11(2), 45–52. <https://doi.org/10.3923/ijaef.2017.45.52>
- [65] Mustakim, M., Buono, A., & Hermadi, I. (2016). "Performance Comparison between Support Vector Regression and Artificial Neural Network For Prediction of Oil Palm Production". *Jurnal Ilmu Komputer dan Informasi*, 9(1), 1–8. <https://doi.org/10.21609/jiki.v9i1.287>.
- [66] Mustakim, M., Buono, A., & Hermadi, I. (2015). "Support Vector Regression for Predicting Oil Palm Productivity in Riau Province". *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 12(2). <https://doi.org/10.24014/sitekin.v12i2.990>.
- [67] Alfakhri, R., Permana, I., Novita, R., & Afdal, M. (2022). "Prediction of Palm Oil Production using Support Vector Regression and Recurrent Neural Network Algorithms". *Building of Informatics, Technology and Science (BITS)*, 6(3). <https://doi.org/10.47065/bits.v6i3.6441>.

- [68] Pardamean, B., Suparyanto, T., Sinaga, G. S. T., & Elwirehardja, G. N. (2022). "Comparison of K-Nearest Neighbor and Support Vector Regression for Predicting Oil Palm Yield". In *Intelligent Systems in Agriculture* (pp. 111-122). Springer. https://doi.org/10.1007/978-3-030-63948-9_9.
- [69] Caraka, R. E., Yasin, H., & Basyiruddin, A. W. (2017). "Crude Palm Oil (CPO) Forecasting using Support Vector Regression Kernel Radial Basis". *Jurnal Matematika*, 7(1), 43-57. <https://doi.org/10.24843/JMAT.2017.v07.i01.p81>
- [70] Goh, C. W., Chai, J., Rahman, A., & Ong, W. E. (2024). "Crude Palm Oil Price Prediction Using Simulated Annealing-Based Support Vector Regression (SA-SVR)". *Asian Academy of Management Journal of Accounting and Finance*, 20(1), 305-333. <https://doi.org/10.21315/aamjaf2024.20.1.10>.
- [71] Chai, K. C. (2021). "A New Hybrid Genetic Algorithm-SARIMA-Artificial Neural Network in Forecasting Malaysian Export Amount Of Palm Oil". (Doctoral Dissertation, Universiti Tun Hussein Onn Malaysia).
- [72] Hilal, Y. Y., Yahya, A., Ismail, W. I. W., & Asha'ari, Z. H. (2021). "Neural Networks Method in Predicting Oil Palm FFB Yields for the Peninsular States of Malaysia". *Journal of Oil Palm Research*, 33(3), 400-412. <https://doi.org/10.21894/jopr.2020.0105>.
- [73] Negara, L. P., & Wustqa, D. U. (2022). "Prediction of Crude Palm Oil Production in Indonesia using Nonlinear Autoregressive Exogenous Neural Network with Genetic Algorithm". *AIP Conference Proceedings*, 2575(1). <https://doi.org/10.1063/5.0103902>.
- [74] Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V. V., & Ciampitti, I. A. (2020). "Satellite-based Soybean Yield Forecast: Integrating Machine Learning and Weather Data for Improving Crop Yield Prediction in Southern Brazil". *Agricultural and Forest Meteorology*, 284, 107886. <https://doi.org/10.1016/j.agrformet.2019.107886>
- [75] Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). "Multiple Linear Regression and Artificial

- Neural Networks based on Principal Components to Predict Ozone Concentrations”. *Environmental Modelling & Software*, 22(1), 97–103. <https://doi.org/10.1016/j.envsoft.2005.12.002>
- [76] Khin, O., Koh, J. G., & Lee, S. K. (2023). “Harvest Forecasting Improvement using Federated Learning and Ensemble Model”. *Smart Media Journal*, 12(10), 9–18. <https://doi.org/10.37955/smartmedia.2023.10.02>
- [77] Khan, N., Kamaruddin, M. A., & Sheikh, U. U. (2022). “Environment-Based Oil Palm Yield Prediction using K-Nearest Neighbour Regression”. *International Journal of Engineering and Technology*, 6(2), 85-92. <https://doi.org/10.11591/ijet.v6i2.3562>.
- [78] Liu, P., & Zhang, Y. (2023). “Optimization of Biodiesel Production from Oil using a Novel Green Catalyst via Development of a Predictive Model”. *Arabian Journal of Chemistry*, 16(6), 104785. <https://doi.org/10.1016/j.arabjc.2023.104785>.
- [79] Ramasenderan, N., Thiruchelvam, V., Sivathasan, R., Ravinchandra, K., Yahya, A., & Sivanesan, S. (2024). “Enhancing Palm Tree Oil Production through an Automated Sensory System and Iot Monitoring on the Blynk Platform”. *AIP Conference Proceedings*, 3161(1). <https://doi.org/10.1063/5.0181282>.
- [80] Yuan, S., Yu, Z., Li, C., & Wang, S. (2020). “A Novel SAR Sidelobe Suppression Method based on CNN”. *IEEE Geoscience and Remote Sensing Letters*, 18(1), 132-136
- [81] Zheng, J., Li, W., Xia, M., Dong, R., Fu, H., & Shuai, Y. (2021). “Large-Scale Oil Palm Tree Detection from High-Resolution Remote Sensing Images using Faster-RCNN”. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178, 194–207. <https://doi.org/10.1016/j.isprsjprs.2021.06.005>.
- [82] Ashari, S., Yanris, G. J., & Purnama, I. (2021). “Oil Palm Fruit Ripeness Detection using Deep Learning”. *International Journal of Agriculture Innovations and Research*, 9(2), 100–104.

- [83] Kumar, T. A., Rajmohan, R., Adeola Ajagbe, S., Gaber, T., Zeng, X. J., & Masmoudi, F. (2023). "A novel CNN gap layer for growth prediction of palm tree plantlings". *Plos one*, 18(8), e0289963.
- [84] Culman, M., Delalieux, S., & Van Tricht, K. (2021). "Individual Palm Tree Detection using Deep Learning on RGB Imagery to Support Tree Inventory". *Remote Sensing*, 13(5), 973. <https://doi.org/10.3390/rs13050973>.
- [85] Puttinaovarat, S., & Horkaew, P. (2021). "Deep and Machine Learnings of Remotely Sensed Imagery and its Multi-Band Visual Features for Detecting Oil Palm Plantation". *Sustainable Computing: Informatics and Systems*, 30, 100560. <https://doi.org/10.1016/j.suscom.2021.100560>.
- [86] Alaa, H., Waleed, K., Samir, M., Tarek, M., & Sobeah, H. (2021). "An Intelligent Approach for Detecting Palm Tree Diseases using Image Processing and Machine Learning". *Int. J. Adv. Comput. Sci. Appl*, 11(7), 434-441.
- [87] Bonet, I., Caraffini, F., Pena, A., Puerta, A., & Gongora, M. (2020, July). "Oil Palm Detection via Deep Transfer Learning". In *2020 IEEE congress on evolutionary computation (CEC)* (pp. 1-8). IEEE.
- [88] Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). "A Review on the Long Short-Term Memory Model". *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>.
- [89] Sagheer, A., & Kotb, M. (2019). "Time Series Forecasting of Petroleum Production using Deep LSTM Recurrent Networks". *Neurocomputing*, 323, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>.
- [90] Amal, I. (2021). "Crude Palm Oil Price Prediction using Multilayer Perceptron and Long Short-Term Memory". *Journal of Mathematical and Computational Science*, 11(6), 8034–8045. <https://doi.org/10.28919/jmcs/6680>.
- [91] Widiputra, H. (2021). "GA-Optimized Multivariate CNN-LSTM Model for Predicting Multi-Channel Mobility in the COVID-19 Pandemic". *Emerging Science Journal*, 5(5), 619–635. <https://doi.org/10.28991/esj-2021-01300>.

- [92] Urolagin, S., Sharma, N., & Datta, T. K. (2021). “A Combined Architecture of Multivariate LSTM with Mahalanobis and Z-Score Transformations for Oil Price Forecasting”. *Energy*, 231, 120963. <https://doi.org/10.1016/j.energy.2021.120963>
- [93] Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). “Multivariate LSTM-FCNs for Time Series Classification”. *Neural Networks*, 116, 237–245. <https://doi.org/10.1016/j.neunet.2019.04.014>.
- [94] Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018, April). “Evaluation of Bidirectional LSTM for Short and Long Term Stock Market Prediction”. 2018 9th International Conference on Information and Communication Systems (ICICS) (pp. 151–156). IEEE. <https://doi.org/10.1109/IACS.2018.8355458>.
- [95] Kartika, N. D., Astika, I. W., & Santosa, E. (2016). “Oil Palm Yield Forecasting based on Weather Variables using Artificial Neural Network”. *Indonesian Journal of Electrical Engineering and Computer Science*, 3(3), 626–633.
- [96] Ulama, B. S., Putra, R. A., Hibatullah, F., Habibi, M. R., & Nafis, M. A. (2025). “Comparison of Artificial Neural Network and Long Short-Term Memory for Modelling Crude Palm Oil Production in Indonesia”. *International Journal of Advanced Computer Science and Applications*, 16(1), 741–747. <https://doi.org/10.14569/IJACSA.2025.0160171>.
- [97] Aini, H., & Haviluddin, H. (2019). “Crude Palm Oil Prediction Based on Backpropagation Neural Network Approach”. *Knowledge Engineering and Data Science*, 2(1), 1–9. <https://doi.org/10.17977/um018v2i12019p1-9>.
- [98] Uzor, P. A., Nwanya, S. C., & Ozor, P. A. (2020). “Artificial Neural Network Application to a Process Time Planning Problem for Palm Oil Production”. *Engineering and Applied Science Research*, 47(2), 161–169. <https://ph01.tcithaijo.org/index.php/easr/article/view/211130>.
- [99] Syarovy, M., Nugroho, A. P., & Sutiarmo, L. (2023, May). “Prediction of Oil Palm Production using Recurrent Neural Network Long Short-Term Memory

- (RNN-LSTM)". Proceedings of the 3rd International Conference on Smart and Innovative Agriculture (ICoSIA 2022), 55. Atlantis Press.
- [100] Ismail, Z., & Khamis, A. (2011). "Neural Network in Modeling Malaysian Oil Palm Yield". *American Journal of Applied Sciences*, 8(8), 796–803. <https://doi.org/10.3844/ajassp.2011.796.803>.
- [101] Silalahi, D. D. (2013, October). "Application of Neural Network Model with Genetic Algorithm to Predict the International Price of Crude Palm Oil (CPO) and Soybean Oil (SBO)". In 12th National Convention on Statistics (NCS), Mandaluyong City, Philippines, October 1–2.
- [102] Silalahi, D. D., Reano, C. E., Lansigan, F. P., Panopio, R. G., & Bantayan, N. C. (2016). "Using Genetic Algorithm Neural Network on Near Infrared Spectral Data for Ripeness Grading of Oil Palm (*Elaeis Guineensis* Jacq.) Fresh Fruit". *Information Processing in Agriculture*, 3(4), 252–261. <https://doi.org/10.1016/j.inpa.2016.08.001>.
- [103] Ahmad, A. L., Azid, I. A., Yusof, A. R., & Seetharamu, K. N. (2004). "Emission Control in Palm Oil Mills using Artificial Neural Network and Genetic Algorithm". *Computers & Chemical Engineering*, 28(12), 2709–2715. <https://doi.org/10.1016/j.compchemeng.2004.05.003>.
- [104] Sugiyarto, A. W., & Abadi, A. M. (2019, September). "Prediction of Indonesian palm oil production using long short-term memory recurrent neural network (LSTM-RNN)". In 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 53-57). IEEE.
- [105] Husaini, F., Permana, I., & Afdal, M. (2024). "Application of Long Short-Term Memory Algorithm for Palm Oil Production Prediction". *Indonesian Journal of Agricultural Technology*, 34(2), 45-53.
- [106] Mimboro, P., Gaol, F. L., Warnars, H. L. H. S., & Soewito, B. (2021, November). "Weather Monitoring System AIoT based for Oil Palm Plantation using Recurrent Neural Network Algorithm". In 2021 IEEE 5th international conference on information technology, information systems and electrical engineering (ICITISEE) (pp. 283-287). IEEE.

- [107] Avula, V. R., Kapavarapu, B. B., & Chandran, A. (2024). "Prediction of Oil Production using Optimized Gated Recurrent Unit (GRU)". *Recent Advances in Artificial Intelligence and Data Science*, 10(2), 42-55. <https://doi.org/10.1016/j.advances.2024.02.015>.
- [108] Saputra, I. A., Vitianingsih, A. V., Kristyawan, Y., & Maukar, A. L. (2024). "Forecasting Model of Export and Import Value of Oil and Gas using Gated Recurrent Unit Method". *Teknika: Jurnal Teknik*, 35(1), 18-29. <https://doi.org/10.2465/teknika.2024.01.045>.
- [109] Mateus, B. C., Mendes, M., Farinha, J. T., & Assis, R. (2021). "Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press". *Energies*, 14(3), 1105. <https://doi.org/10.3390/en14031105>.
- [110] Mukkamala, R. B., Rajaprasad, S. V. S., & Namboodiri, V. (2024). "Prediction of Sinter Productivity Utilizing Deep Learning Frameworks-A Multivariate Analysis". *Metallurgical and Materials Data*, 2(4), 141-146.
- [111] Mehta, A. M., & Patel, K. S. (2023). "LSTM-based Forecasting of Dengue Cases in Gujarat: A Machine Learning Approach". *Indian Journal of Science and Technology*, 17(7), 635-642.
- [112] Aqbar, K., & Supomo, R. A. (2023). "Performance Analysis of LSTM And Xgboost Models Optimization in Forecasting Crude Palm Oil (Cpo) Production at Palm Oil Mill (Pom)". *International Journal of Computer Applications*, 975, 8887.
- [113] Ali, M., Deo, R. C., Downs, N. J., & Maraseni, T. (2018). "Multi-Stage Committee based Extreme Learning Machine Model Incorporating the Influence of Climate Parameters and Seasonality on Drought Forecasting". *Computers and electronics in agriculture*, 152, 149-165.
- [114] Rousson, V., & Gosoni, N. F. (2007). "An $R - square$ Coefficient based on Final Prediction Error". *Statistical Methodology*, 4(3), 331-340.
- [115] Zheng, J., Li, W., Xia, M., Dong, R., Fu, H., & Yuan, S. (2019, July). "Large-Scale Oil Palm Tree Detection from High-Resolution Remote Sensing Images

using Faster-RCNN”. In Igarss 2019-2019 ieee international geoscience and remote sensing symposium (pp. 1422-1425). IEEE.

- [116] Pakistan Bureau of Statistics, External Trade Statistics, Historical Data (8-digit Level), Accessed on April 20th, 2025 [Online], Available at: <https://www.pbs.gov.pk/content/external-trade-statistical-data-8-digit-level-year-1990-1991-2019-2020>.

Python Code of Thesis

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_percentage_error, mean_squared_error,
r2_score
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv1D, Flatten, MaxPooling1D, Dropout
from tensorflow.keras.callbacks import EarlyStopping
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller
import warnings
warnings.filterwarnings('ignore')
from google.colab import drive
Mount Google Drive (Comments for adding drive dataset)
drive.mount('/content/drive')
Load and preprocess data (Comments)
data = pd.read_csv('/content/drive/MyDrive/Plam-Oil-Import-2013-2024.csv') for
Updating the path as needed
```

```
data['Date'] = pd.to_datetime(data['Year'].astype(str) + pd.to_timedelta((data['Month']-
1)*30, unit='D'))
data.set_index('Date', inplace=True)
data.sort_index(inplace=True)
Calculate price per MT (comments)
data['Price_per_MT'] = data['Dollars_Thousand'] * 1000 / data['Quantity_MT']
Feature Engineering (comments) data['Month_sin'] = np.sin(2 * np.pi * data['Month']/12)
data['Month_cos'] = np.cos(2 * np.pi * data['Month']/12)
data['Quantity_lag1'] = data['Quantity_MT'].shift(1)
data['Quantity_lag12'] = data['Quantity_MT'].shift(12)
data['Price_lag1'] = data['Price_per_MT'].shift(1)
data['Price_lag12'] = data['Price_per_MT'].shift(12)
data.dropna(inplace=True)
Visualize the data (comments)
plt.figure(figsize=(15, 8))
plt.subplot(2, 1, 1)
plt.plot(data.index, data['Quantity_MT'], label='Quantity (MT)')
plt.title('Palm Oil Import Quantity Over Time')
plt.legend()
plt.subplot(2, 1, 2)
plt.plot(data.index, data['Price_per_MT'], label='Price (/MT)', color='orange')
plt.title('Palm Oil Price Over Time')
plt.legend()
plt.tight_layout()
plt.show()
Split data into features and targets (comments)
X = data[['Month', 'Month_sin', 'Month_cos', 'Quantity_lag1', 'Quantity_lag12',
'Price_lag1', 'Price_lag12']]
y_quantity = data['Quantity_MT']
y_price = data['Price_per_MT']
Split into train and test sets (last 30% for testing)
test_size = int(len(data) * 0.3)
```

```

X_train, X_test = X[:-test_size], X[-test_size:]
y_quantity_train, y_quantity_test = y_quantity[:-test_size], y_quantity[-test_size:]
y_price_train, y_price_test = y_price[:-test_size], y_price[-test_size:]
Normalize data (comments)
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
Reshape data for CNN
X_train_cnn = X_train_scaled.reshape(X_train_scaled.shape[0], X_train_scaled.shape[1],
1)
X_test_cnn = X_test_scaled.reshape(X_test_scaled.shape[0], X_test_scaled.shape[1],
1)
Function to evaluate models (comments)
def evaluate_model(y_true, y_pred, model_name, target):
mape = mean_absolute_percentage_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
r2 = abs(r2_score(y_true, y_pred))
For classification metrics (direction prediction) (comments)
y_true_dir = np.where(y_true
¿ np.roll(y_true, 1), 1, 0)[1:]
y_pred_dir = np.where(y_pred ¿ np.roll(y_pred, 1), 1, 0)[1:]
Ensure same length
min_len = min(len(y_true_dir), len(y_pred_dir))
y_true_dir = y_true_dir[:min_len]
y_pred_dir = y_pred_dir[:min_len]
accuracy = np.mean(y_true_dir == y_pred_dir)
precision = np.sum((y_true_dir == 1) & (y_pred_dir == 1)) / np.sum(y_pred_dir
== 1)
results = 'Model': model_name, 'Target': target, 'MAPE': mape, 'RMSE': rmse,
'R2': r2, 'Accuracy': accuracy, 'Precision': precision
return results
Initialize results dataframe (comments)
results_df = pd.DataFrame(columns=['Model', 'Target', 'MAPE', 'RMSE', 'R2',

```

```
'Accuracy', 'Precision']])
```

1. Random Forest Model

```
def random_forest_model(X_train, X_test, y_train, y_test, target):
```

```
    print(f"Training Random Forest for target...")
```

```
    rf = RandomForestRegressor(n_estimators=150, max_depth=5, max_features=3)
```

```
    rf.fit(X_train, y_train)
```

```
    y_pred = rf.predict(X_test)
```

```
    Evaluate (comments)
```

```
    results = evaluate_model(y_test, y_pred, "RF", target)
```

```
    Plot predictions vs actual (comments)
```

```
    plt.figure(figsize=(12, 6))
```

```
    plt.plot(y_test.index, y_test, label='Actual', marker='o')
```

```
    plt.plot(y_test.index, y_pred, label='Predicted', marker='x')
```

```
    plt.title(f"RF: Actual vs Predicted target")
```

```
    plt.legend()
```

```
    plt.show()
```

```
    return results
```

```
Run RF for quantity and price (comments) rf_quantity_results = random_forest_model(X_train,  
X_test, y_quantity_train, y_quantity_test, "Quantity")
```

```
rf_price_results = random_forest_model(X_train, X_test, y_price_train, y_price_test,  
"Price")
```

```
results_df = pd.concat([results_df, pd.DataFrame([rf_quantity_results, rf_price_results])],  
ignore_index=True)
```

2. Artificial Neural Network (ANN) (comments)

```
def ann_model(X_train, X_test, y_train, y_test, target):
```

```
    print(f"Training ANN for target...")
```

```
    ann = MLPRegressor(hidden_layer_sizes=(128, 64, 32, 16), activation='relu', solver='adam',  
max_iter=200, random_state=42, early_stopping=False)
```

```
    ann.fit(X_train_scaled, y_train)
```

```
    y_pred = ann.predict(X_test_scaled)
```

```
    Evaluate (comments)
```

```
    results = evaluate_model(y_test, y_pred, "ANN", target)
```

Plot predictions vs actual (comments)

```
plt.figure(figsize=(12, 6))
plt.plot(y_test.index, y_test, label='Actual', marker='o')
plt.plot(y_test.index, y_pred, label='Predicted', marker='x')
plt.title(f'ANN: Actual vs Predicted target')
plt.legend()
plt.show()
return results
```

Run ANN for quantity and price (comments)

```
ann_quantity_results = ann_model(X_train_scaled, X_test_scaled, y_quantity_train,
y_quantity_test, "Quantity")
ann_price_results = ann_model(X_train_scaled, X_test_scaled, y_price_train, y_price_test,
"Price")
results_df = pd.concat([results_df, pd.DataFrame([ann_quantity_results, ann_price_results])],
ignore_index=True)
```

3. Convolutional Neural Network (CNN) (comments)

```
def cnn_model(X_train, X_test, y_train, y_test, target):
print(f'Training CNN for target...')
model = Sequential([ Conv1D(filters=256, kernel_size=2, activation='relu', in-
put_shape=(X_train_cnn.shape[1], 1)),
MaxPooling1D(pool_size=1), Conv1D(filters=128, kernel_size=2, activation='relu'),
MaxPooling1D(pool_size=1),
Flatten(),
Dense(64, activation='relu'),
Dense(32, activation='relu'),
Dense(16, activation='relu'),
Dropout(0.01),
Dense(1)
])
model.compile(optimizer='adam', loss='mse')
early_stop = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
history = model.fit( X_train_cnn, y_train, validation_data=(X_test_cnn, y_test),
```

```
epochs=200, batch_size=32, verbose=0, callbacks=[early_stop])
y_pred = model.predict(X_test_cnn).flatten()
Plot training history (comments)
plt.figure(figsize=(10, 5))
plt.plot(history.history['loss'], label='Train Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title(f'CNN Training History for target')
plt.legend()
plt.show()
Plot predictions vs actual (comments)
plt.figure(figsize=(12, 6))
plt.plot(y_test.index, y_test, label='Actual', marker='o')
plt.plot(y_test.index, y_pred, label='Predicted', marker='x')
plt.title(f'CNN: Actual vs Predicted target')
plt.legend()
plt.show()
Evaluate (comments)
results = evaluate_model(y_test, y_pred, "CNN", target)
return results
Run CNN for quantity and price (comments)
cnn_quantity_results = cnn_model(X_train_cnn, X_test_cnn, y_quantity_train, y_quantity_test,
"Quantity")
cnn_price_results = cnn_model(X_train_cnn, X_test_cnn, y_price_train, y_price_test,
"Price")
results_df = pd.concat([results_df, pd.DataFrame([cnn_quantity_results, cnn_price_results])],
ignore_index=True)
4. ARIMA Model (comments)
def arima_model(y_train, y_test, target):
print(f"Training ARIMA for target...")
Check stationarity (comments)
def check_stationarity(series):
result = adfuller(series)
```

```
print(f'ADF Statistic: result[0]')
print(f'p-value: result[1]')
return result[1] >= 0.05

is_stationary = check_stationarity(y_train)
print(f'Is target stationary? is_stationary')

Fit ARIMA model (parameters can be optimized) (comments)
model = ARIMA(y_train, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))
model_fit = model.fit()

Forecast (comments)
y_pred = model_fit.forecast(steps=len(y_test))

Evaluate (comments)
results = evaluate_model(y_test, y_pred, "ARIMA", target)

Plot results (comments)
plt.figure(figsize=(12, 6))
plt.plot(y_train.index[-24:], y_train[-24:], label='Train')
plt.plot(y_test.index, y_test, label='Actual')
plt.plot(y_test.index, y_pred, label='Predicted')
plt.title(f'ARIMA Model Results for target')
plt.legend()
plt.show()

return results

Run ARIMA for quantity and price (comments)
arima_quantity_results = arima_model(y_quantity_train, y_quantity_test, "Quantity")
arima_price_results = arima_model(y_price_train, y_price_test, "Price")
results_df = pd.concat([results_df, pd.DataFrame([arima_quantity_results, arima_price_results])],
ignore_index=True)

Display all results (comments)
print("Model Evaluation Results:")
print(results_df.to_string(index=False))

Plot comparison of models for each metric (comments)
def plot_metric_comparison(results_df, metric, title):
```

```
plt.figure(figsize=(12, 6))
sns.barplot(data=results_df, x='Model', y=metric, hue='Target')
plt.title(title)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

metrics = [ ('MAPE', 'MAPE'), ('RMSE', 'RMSE'), ('R2', 'R-squared'), ('Accuracy', 'Accuracy'), ('Precision', 'Precision') ]
for metric, title in metrics:
    plot_metric_comparison(results_df, metric, title)
    data Collected from the models (comments)
    data = 'Model': ['RF', 'RF', 'ANN', 'ANN', 'CNN', 'CNN', 'ARIMA', 'ARIMA'],
          'Target': ['Quantity', 'Price', 'Quantity', 'Price', 'Quantity', 'Price', 'Quantity', 'Price'],
          'MAPE': [0.12, 0.10, 0.10, 0.08, 0.09, 0.07, 0.6, 0.5],
          'RMSE': [35837.84, 188.25, 33574.80, 183.10, 32640.61, 182.90, 41707.29, 345.51],
          'R2': [0.81, 0.80, 0.823, 0.812, 0.854, 0.841, 0.79, 0.81],
          'Accuracy': [0.85, 0.83, 0.87, 0.85, 0.89, 0.86, 0.8, 0.81],
          'Precision': [0.91, 0.89, 0.92, 0.91, 0.94, 0.93, 0.82, 0.81]

df = pd.DataFrame(data)
Set up the matplotlib figure
plt.figure(figsize=(18, 12))
plt.suptitle("Model Performance Comparison", fontsize=16, y=1.02)
Create subplots for each metric
metrics = ['MAPE', 'RMSE', 'R2', 'Accuracy', 'Precision']
colors = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd']
for i, metric in enumerate(metrics):
    plt.subplot(2, 3, i+1)
    Special handling for R2 due to outlier values
    if metric == 'R2':
        Create a separate plot for ANN Quantity (outlier) (comments)
```

```
temp_df = df.copy()
temp_df.loc[temp_df['R2'] < 5, 'R2'] = np.nan; Remove outlier for better scaling
(comments)
ax = sns.barplot(x='Model', y=metric, hue='Target', data=temp_df, palette=['#1f77b4',
'#ff7f0e'])
plt.title(f"metric (ANN Quantity outlier excluded)")
plt.ylim(0, 1) # Adjust based on your normal R2 range
else:
ax = sns.barplot(x='Model', y=metric, hue='Target', data=df, palette=['#1f77b4',
'#ff7f0e'])
plt.title(metric)
Add value labels on top of bars
for p in ax.patches:
height = p.get_height()
if not np.isnan(height):
ax.text(p.get_x() + p.get_width()/2., height + 0.01 * height,
f'height:.3f', ha='center', va='bottom', fontsize=9)
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.xticks(rotation=45)
plt.tight_layout()
Add the outlier information separately
plt.figure(figsize=(8, 5))
ann_r2 = df[(df['Model'] == 'ANN') & (df['Target'] == 'Quantity')]['R2'].values[0]
plt.text(0.5, 0.5, f"ANN Quantity R2 outlier value: ann_r2:.3f",
ha='center', va='center', fontsize=12)
plt.axis('off')
plt.tight_layout()
plt.show()
Save results to CSV (Comments)
results_df.to_csv('palm_oil_model_evaluation_results.csv', index=False)
print("Evaluation results saved to 'palm_oil_model_evaluation_results.csv'")
```