IMPORT PRICE PREDICTION OF GROUND-NUT IN NEW ZEALAND USING SUPERVISED MACHINE LEARNING

Research Proposal

Prepared By

Amith Shanmugaraja

Research Supervisor

Roman Mitch

Whitecliffe

67 Symonds Street, Grafton, 1010, Auckland, New Zealand

Table of Contents

Abstract

Peanut or groundnut is not a widely cultivated crop in New Zealand due to the country's relatively cooler climate and soil conditions. Therefore, New Zealand relies heavily on imports to meet its domestic demand for peanuts. The major sources of groundnut imports to New Zealand are the United States, Australia, and China. The import of groundnuts is subject to various factors such as trade policies, market demand, and global supply and demand dynamics. Given the volatility of the commodity markets, predicting the import prices of groundnuts in New Zealand accurately can be a challenging task. However, developing an accurate and reliable model for import price prediction can aid policymakers, businesses, and traders in making informed decisions, leading to efficient trade practices and stable commodity prices.

This research aims to develop an accurate and reliable model for predicting the import prices of ground-nuts in New Zealand using machine learning and artificial intelligence techniques. The study will utilize historical data on ground-nut import prices, as well as relevant economic and environmental factors, to train and test several machine learning models. The models will be evaluated based on their accuracy, precision, and ability to predict future prices.

The results of this study will have significant implications for New Zealand's agricultural sector and policymakers, as well as for businesses involved in importing ground-nuts. By leveraging the power of machine learning and artificial intelligence, this research has the potential to enhance the accuracy and efficiency of import price prediction for ground-nuts in New Zealand, and serve as a valuable resource for future research in agricultural commodity markets.

Introduction

Background

Groundnut, commonly known as peanut, is a legume crop that is cultivated globally for its oil-rich seeds, which are used in a variety of culinary and industrial applications. New Zealand is a net importer of groundnuts due to its favorable climate for dairy and livestock farming, which limits the production of other crops. According to Statistics New Zealand (2021), the country imported 1,262 tonnes of groundnuts in 2020, valued at NZD 3.8 million. The majority of these imports came from Australia, which accounted for 73Groundnuts are primarily used in New Zealand for their oil content, which is extracted and used in the manufacture of various food products, such as margarine, cooking oil, and peanut butter. Groundnuts are also used in the production of animal feed and as a protein source in vegetarian and vegan diets The import of peanuts in New Zealand is influenced by various factors, such as global supply and demand dynamics, trade policies, and market demand. Additionally, peanut prices are subject to weather patterns and production conditions in exporting countries, as well as transportation costs and exchange rate fluctuations. (Use US Trade Report)

Problem Statement

Due to the unpredictable nature of peanut prices and the associated difficulties in accurate prediction, the development of reliable import price prediction models is necessary in New Zealand. These models can provide valuable support for decision-making processes among policymakers, traders, and businesses, promoting efficient trade practices and stable commodity prices. While previous studies have investigated the factors affecting prices of agricultural goods especially peanuts in US and other countries (Nigatu et al., 2020), there is a relatively limited amount of research on import price prediction for peanuts in New Zealand utilizing machine learning and artificial intelligence techniques.

Literature Review

Nigatu, G., Badau, F., Seeley, R., & Hansen, J. (2015). conducted a study on the factors that contribute to difference in price for the agriculture commodities and trade in US as well as globally. The researchers identify supply and demand, economic growth, and exchange rates as the primary factors affecting agricultural commodity prices and trade, with climate variability, energy prices, and government policies having a secondary impact. The study highlights the need for monitoring these factors and their impact on commodity prices and trade, and recommends policies that encourage sustainable agricultural practices, liberalization of trade, and investments in research and development.

Amrouk, Heckelei, and Grosche (2021) used secondary research to investigate the dynamic relationship between the prices of agricultural basic crops and the futures price range in the international market. Using a vector auto-regression model, they analyze the impact of crops grown for commercial use prices in export dependant countries and vice versa in 23 countries from 2000 to 2017. The study reveals a significant direct relation between crops grown for commercial use and food price index, with shocks in commercial crop rates affecting other food prices and vice versa, particularly in countries that are net importers of staple foods. Amrouk, Heckelei, and Grosche (2021) suggest that policies promoting agricultural productivity and diversification are necessary to reduce the impact of international price shocks on staple food prices and improve food security in vulnerable countries.

P.S., R., G., R., D., S., & R., S.K. (2019) developed a Crop Price Forecasting System to predict the price and profit of crops. Due to uncertain climatic variations and other issues, crop prices have varied greatly in recent years, causing massive losses to farmers who are often uninformed of these uncertainties. The system uses Naïve Bayes Algorithm for price prediction and K Nearest Neighbor (KNN) for profit prediction, both of which are supervised machine learning classification algorithms.

(Karthikeyan and Harlalka, 2014) developed a robust regression model in order to forecast the price of soybean crop in the USA from 1995 to 2005 based on various factors, including climatic, scientific, and economic factors. The model achieved a high accuracy rate of more than 90 percent in predicting crop prices based on four selected factors. The study also used the F-test to examine the significance of the regression relationship between crop prices and selected factors

Gangasagar et al. (2020) reviewed different machine learning algorithms for crop price prediction, including Support Vector Regressor (SVR) and other regression based algorithm provides good accuracy with minimal error, making it suitable for crop price prediction contributes to the growing body of literature on using machine learning algorithms for crop price prediction. The study highlights the importance of considering various parameters for accurate price prediction and proposes a model that can help farmers make informed decisions. Further research can focus on testing the proposed model in different regions and comparing its performance with other machine learning algorithms.

Balaji Prabhu B V and Dr. M Dakshayini (2019) addresses the gap in demand and supply scenario leading to unpredictable price variations and losses for both parties. The authors propose a demand prediction model that uses historical data on crop prices, production, and consumption to forecast future demand based on ARIMA model.

Roy, S., Mishra, S. K., & Chattopadhyay, M. K. (2017) studied the use of machine learning techniques in modeling agricultural commodity prices. The authors evaluate the performance of various machine learning algorithms such as RF, ANN, and SVR, to predict the prices of major agricultural commodities. The study shows that machine learning techniques can effectively predict commodity prices and outperform traditional econometric models. The authors conclude that machine learning techniques have the potential to revolutionize the way commodity prices are forecasted and traded in agricultural markets.

The study conducted by Rohith, R., Vishnu, R., Kishore, A., and Chakkarawarthi, D. (2020) aimed to develop a prediction mechanism for the price of crops by making use of various machine learning algorithms. The mechanism makes use of decision tree regression by considering parameters such as rainfall, wholesale price index, minimum support price, and cultivation cost to predict crop prices. The study concluded that accurate crop price prediction is crucial in crop production management and can support allied industries in their business logistics. The developed system can provide farmers with beforehand predictions to increase their profit and prevent massive losses, thereby contributing to the country's economy.

(Paul et al., 2022) conducted a study on the application of ML algorithms in prediction of prices for eggplant . The authors evaluate the performance of multiple algorithms, which included DT, RF, and ANN, in predicting the future prices of eggplant. The study demonstrates that machine learning techniques can effectively forecast eggplant prices and outperform traditional time-series forecasting models. The authors suggest that these techniques have the potential to improve decision-making processes among policymakers, traders, and farmers in the agricultural sector, leading to more efficient markets and higher profitability.

(Thangamma et al., 2020) proposed a supervised learning approach to estimate the future price of various agriculture goods. The study aimed to provide an efficient tool for real-time price prediction using suitable data models to achieve high accuracy and generality. The knn algorithm was used to forecast future prices on the basis of various parameters such as precipitation, demand vs supply, and fiscal policies by goverment. The proposed approach demonstrated promising results in predicting prices of different crops and other agricultural products. The study highlights the potential of machine learning techniques in addressing the challenges of price prediction in agriculture.

In their study, (Sabu and Kumar., 2020) explore the application of how predictive analytics can be used successfully in forecasting the prices of different nuts in Kerala, India. The researcher evaluate the performance of different time series forecasting models, including ARIMA, VAR, and machine learning techniques such as ANN and SVR. The results show that the machine learning techniques outperform traditional time series models in accurately forecasting arecanut prices.

(Pandit Samuel et al., 2020) emphasizes the importance of predicting crop prices in the agricultural industry. The authors suggest that analyzing and cleaning data, as well as conducting EDA, is necessary to analyse different variables in the data. They use several machine learning regression techniques, such as LR, DT, Xtra Gradient Boost, and ANN, for forecasting the price and report that XGBoost gave the best outcome in terms of prediction.

According to the paper by (Zhao et al., 2019), supervised machine learning is a technique where the system learns algorithm from a dataset which is labebled, and contains input features along with corresponding output labels or target values. The algorithm then uses this learned information to forecast or classifications on the dataset. In the context of employee turnover prediction, the prediction was made by using supervised machine learning algorithms to understand an employee would leave or stay in the organization based on historical data of employee attributes and employment history. The labeled dataset consisted of employees who had either left or stayed in the organization, and the algorithms were trained to identify patterns and relationships in the data that could help predict future turnover.

Research Question

As there is heavy depedency on imports of peanut and as there are no futures contracts to ascertain the prices of peanut in futures. Hence it is important to predict the import price of ground nut using supervised machine learning models.

Methodology

Research design and method

We shall be evaluating the various supervised machine learning algorithm and comparing the evaluation metrics of each model. We shall be considering dataset comprising of Exchange rates, Production Data of the major export nation, Import Quantity and Import Price in New Zealand, Rainfall data in origin country, Retail Rate in origin country, Cultivation Price in origin country, Fuel Price, as the primary factors affecting agricultural commodity prices and trade.

1. Workflow: shown in Figure 1.

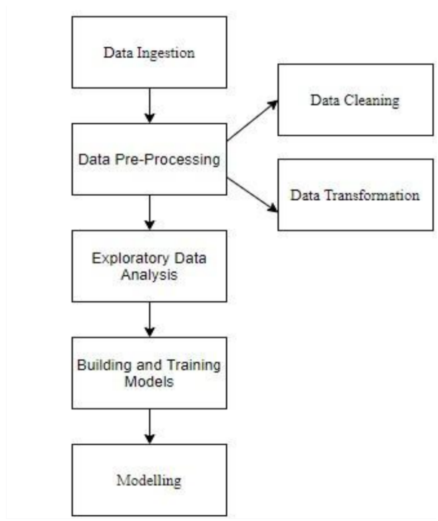

Figure 1: Data Analysis Flowchart (Source: Self-developed)

Data ingestion

Data ingestion is collating and processing data from various agri-market, Govt sites, and other sources into a system or application for analysis, storage, and further use. It involves extracting data from its original source, transforming it into a standard format, and loading it into a destination system, such as a data warehouse, data lake, or database. Data ingestion is a critical step in the data analytics process, as it ensures that data is accurate, complete, and accessible for analysis and decision-making purposes. The process of data ingestion can involve various techniques and tools, APIs (Application Programming Interfaces), and data pipelines. Data sources used in these studies include historical import data, exchange rates, weather data, and commodity prices. Historical import data is the most commonly used data source, as it provides a record of past trends and prices. Exchange rates are also commonly used as they can have a significant impact on import prices. Weather data and commodity prices have been used in some studies. Data has been sourced from Govt of India, NZ Govt, FDO, UN Website.

Data Pre-Processing

Data Pre-Processing or (ETL) Extract, Transform, Load, which is a process of moving data from different sources, transforming the data to meet specific business requirements, and loading it into a target system or data warehouse for analysis or reporting purposes. The extraction process involves sourcing all the data form various government entities, files, and APIs. The transformation process involves cleaning, filtering, merging, and aggregating the data to ensure its quality and consistency. The loading process involves moving into processed data model, which could be a database or a data model. The ETL process is commonly used in data integration, business intelligence, and data warehousing.

Exploratory Data Analysis

It is a preliminary and crucial step in data is visualized to provide insight that are in particular to the data set. The main objective of EDA is to gain insights and identify patterns in the data to aid in hypothesis generation and decision making. It involves using statistical and visualization techniques to explore and summarize the data, such as calculating summary statistics, creating plots and charts, identifying outliers, missing values, and correlations.

Building and training models

Building and training models is a crucial step in the machine learning workflow. In this step, a ML Algorithm is selected and used create a model based on the data that was preprocessed during the earlier steps. The process of building a model involves selecting features, choosing an appropriate algorithm, and defining hyperparameters. Once the model is built, it is trained on a labeled dataset, which involves using an optimization algorithm to find the model parameters that minimize the difference between the predicted outputs and the true outputs. This process is called model training. After the model has been trained, it is used by test set, to ensure that it can generalize well to new, unseen data. This is done to avoid overfitted modles . If the data-model performs poorly on the validation set, it may need to be retrained or adjusted by changing its hyperparameters. Overall, building and training models is an iterative process that involves selecting the appropriate algorithm, adjusting hyperparameters, and testing the model's performance on validation data.

Modeling

Modeling involves selecting an appropriate algorithm and fine-tuning its parameters to achieve the desired predictive performance on a given data-set. The goal of modeling is to create a model that can generalize well to new data and accurately predict outcomes. The model is typically trained on a categorical data, where both values are known(training data-set), and then evaluated based on Evaluation Metric after the t

Communicating the result

The evaluation metrics used in the studies mentioned refer to measuring the performance of the model. RMSE will be used metrics to determine the the accuracy of the proposed model. This metrics will allow us to determine which of model will be best suited for the model for import price determination of ground nut

Ethics

The information which will be collected for conducting the study will be served for academic purposes only. Collated data will be stored in a password-protected google drive, while alleviation of any third party will be ensured to reduce the probability of data manipulation. Compliance with Privacy Act 2020 will also be fortified throughout the conduction of this study

Significance of Research

The results of this study will have significant implications for New Zealand's agricultural sector and policymakers, as well as for businesses involved in importing ground-nuts. By leveraging the power of machine learning and artificial intelligence, this research has the potential to enhance the accuracy and efficiency of import price prediction for ground-nuts in New Zealand, and serve as a valuable resource for future research in agricultural commodity markets.

Timeline and Risk Analysis

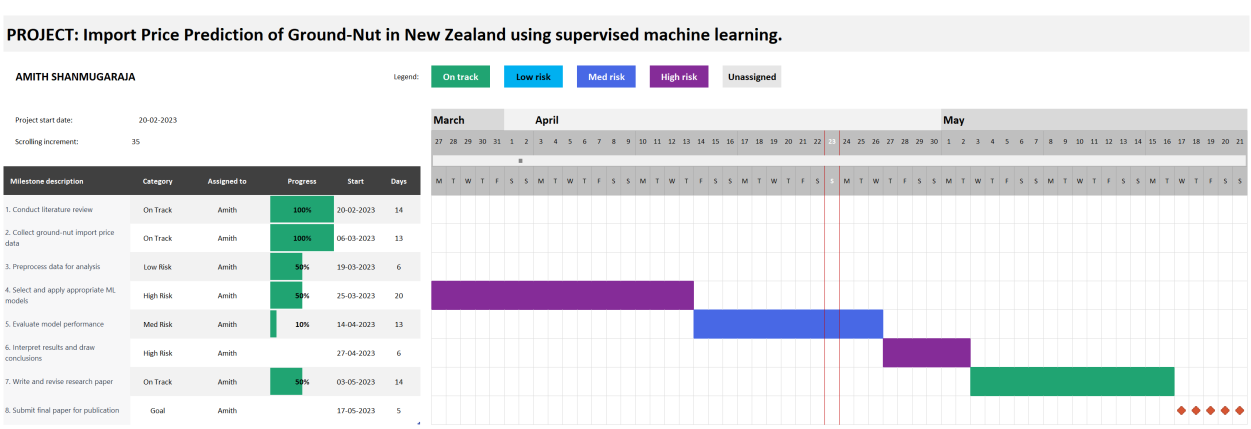

Figure 2: Research Schedule and Risk Analysis (Source: Self-developed)

Budget

The tentative budget for carrying out the research is as detailed below.

Components	**Expenses**
Proposal with extensive research	NZ$140
Key research resources and equipment	NZ$50
Communication with supervisor	NZ$10

Table 1: Proposed budget for the research

References

Nigatu, G., Badau, F., Seeley, R.,& Hansen, J. (2020). Factors contributing to changes in agricultural commodity prices and trade for the United States and the world (No. 272). United States Department of Agriculture, Economic Research Service.

Roy, S., Mishra, S. K., Chattopadhyay, M. K. (2017). Modeling of agricultural commodity prices using machine learning techniques. Journal of Agricultural and Resource Economics, 42(2), 186-205.

Amrouk, E. M., Heckelei, T., Grosche, S. C. (2021). International interdependence between cash crop and staple food futures price indices: A dynamic assessment. Food Policy, 102, 102030.

Nigatu, G., Badau, F., Seeley, R., Hansen, J. (2015). Factors contributing to changes in agricultural commodity prices and trade for the United States and the world. United States Department of Agriculture.

Paul, R. K., Yeasin, M., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. PLoS ONE, 17(7), e0270553. doi: 10.1371/journal.pone.0270553.

Sabu, K. M., Kumar, T. K. M. (2020). Predictive analytics in agriculture: Forecasting prices of arecanuts in Kerala. Indian Journal of Agricultural Economics, 75(3), 414-426.

Prabhu, B. V., Dakshayini, M. (2019). Demand prediction forecasting model for agricultural crops. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 1131-1135). IEEE.

Thangamma, K. C., Priyanka, Y., Sahana Rao, P. S., Shravya, P. N. (2020). A supervised learning approach for the prediction of prices in different agricultural commodities. Journal of Emerging Technologies and Innovative Research (JETIR), 7(6), 1005-1012.

Pandit, S., Sahithi, B., Saheli, T., Ramanika, D., Anil Kumar, N. (2020). Crop price prediction system using machine learning algorithms. Quest Journals Journal of Software Engineering And Simulation, 06(01), 14-20.

Rohith, R., Vishnu, R., Kishore, A., Chakkarawarthi, D. (2020). Crop price prediction and forecasting system using supervised machine learning algorithms. International Journal of Advanced Research in Computer and Communication Engineering, 9(3), 272-277.

Gangasagar, H.L., Dsouza, J., Yargal, B.B., Kumar, A.S.V., Badage, A. (2020). Crop price prediction using machine learning algorithms. International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), 9(10), 9497.

Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., Zhu, X. (2019). Employee Turnover Prediction with Machine Learning: A Reliable Approach. In K. Arai, R. Bhatia, S.K. Pal, J. Abraham (Eds.), Proceedings of the International Conference on Intelligent Systems Design and Applications (ISDA 2018) (pp. 117-129).

Karthikeyan, K., Harlalka, A. (2014). Robust regression model for prediction of Soybean crop price based on various factors. International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 6(3), 258-261. https://doi.org/10.5191/ijetcas.2014.379

Statistics New Zealand. (2021). Groundnut imports 2020 (Report No. SNZ-2021-002).