**|**

INFORMATICS INSTITUTE OF TECHNOLOGY  
In Collaboration with  
UNIVERSITY OF WESTMINSTER



*University of Westminster, Coat of Arms*

**Sri Orzyaugur**

Multivariate Time Series Forecasting for Rice Prices in

Sri Lanka

A dissertation by  
Mohammed Luqman Farhat Rumaiz  
W1761767 - 2018130

Supervised by  
Mr. Gihan Liyanage

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software  
Engineering degree at the University of Westminster.  
May 2023

# TABLE OF CONTENTS

[TABLE OF CONTENTS ii](#_Toc127432142)

[LIST OF FIGURES iii](#_Toc127432143)

[LIST OF TABLES iv](#_Toc127432144)

[ABSTRACT vii](#_Toc127432145)

[Chapter 1 INTRODUCTION 1](#_Toc127432146)

[1.1 Chapter Overview 1](#_Toc127432147)

[1.2 Problem Background 1](#_Toc127432148)

[1.3 Problem Definition 2](#_Toc127432149)

[1.4 Aims and Objectives 2](#_Toc127432150)

[1.5 Novelty of the Research 5](#_Toc127432151)

[1.6 Research Gap 5](#_Toc127432152)

[1.7 Research Contribution 5](#_Toc127432153)

[1.8 Research Challenges 6](#_Toc127432154)

[1.9 Chapter Summary 6](#_Toc127432155)

[Chapter 2: SRS 7](#_Toc127432156)

[2.1 Chapter Overview 7](#_Toc127432157)

[2.2 Rich Picture Diagram 7](#_Toc127432158)

[2.3 Stakeholder Analysis 8](#_Toc127432159)

[2.4 Requirement Elicitation Methods 10](#_Toc127432160)

[2.5 Analysis of Data & Presentation of the Outcome through Elicitation Methodologies 11](#_Toc127432161)

[2.6 Summary of Findings 16](#_Toc127432162)

[2.7 Context Diagram 17](#_Toc127432163)

[2.8 Use Case Diagram 17](#_Toc127432164)

[2.9 Use Case Description 18](#_Toc127432165)

[2.10 Requirements 19](#_Toc127432166)

[2.11 Chapter Summary 21](#_Toc127432167)

[Chapter 3: Design 22](#_Toc127432168)

[3.1 Chapter Overview 22](#_Toc127432169)

[3.2 Design Goals 22](#_Toc127432170)

[3.3 High Level Design 23](#_Toc127432171)

[3.4 Low Level Design 25](#_Toc127432172)

[3.5 Design Diagrams 26](#_Toc127432173)

[28](#_Toc127432174)

[3.6 Chapter Summary 29](#_Toc127432175)

[Chapter 4: Implementation 30](#_Toc127432176)

[4.1 Chapter Overview 30](#_Toc127432177)

[4.2 Technology Selection 30](#_Toc127432178)

[4.3 Implementation of Core Functionality 33](#_Toc127432179)

[4.4 Chapter Summary 36](#_Toc127432180)

[Chapter 5: Conclusion 37](#_Toc127432181)

[5.1 Chapter Overview 37](#_Toc127432182)

[5.2 Deviations 37](#_Toc127432183)

[5.3 Initial Test Results 38](#_Toc127432184)

[5.4 Required Improvements 38](#_Toc127432185)

[5.5 Demo of the Prototype 38](#_Toc127432186)

[5.6 Chapter Summary 38](#_Toc127432187)

[REFERENCES I](#_Toc127432188)

[Appendix A – Introduction III](#_Toc127432189)

[Appendix B – Software Requirements Specification III](#_Toc127432190)

[Appendix C – Conclusion V](#_Toc127432191)

# LIST OF FIGURES

[Figure 1 Rich Picture Diagram 7](#_Toc127432052)

[Figure 2 Stakeholder Onion Model 8](#_Toc127432053)

[Figure 3 Context Diagram](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432054) 17

[Figure 4 Use Case Diagram 18](#_Toc127432055)

[Figure 3.15 Three-Tiered Architecture Diagram](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432056) 23

[Figure 6 Architecture Diagram](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432057) 23

[Figure 7 Level 1 Data Flow Diagram 26](#_Toc127432058)

[Figure 8 System Flow Diagram](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432059) 27

[Figure 9 Wireframe 1/3: Historical Data Visualization](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432060) 28

[Figure 10 Wireframe 2/3: Selection and Option Wizard](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432061) 28

[Figure 11 Wireframe 3/3: Forecasted Data Visualization](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432062) 29

[Figure 12 Technology Stack](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432063) 30

[Figure 13 Transforming dataframe into monthly timestamps into monthly timestamps 33](#_Toc127432064)

[Figure 14 Distributing paddy production for Maha and Yala seasons 34](#_Toc127432065)

[Figure 15 Filling missing data with interpolation 34](#_Toc127432066)

[Figure 16 Converting dataframes into series 35](#_Toc127432067)

[Figure 17 Scaling timeseries data 35](#_Toc127432068)

[Figure 18 Building and fitting model, then predicting rice prices (Univariate) 35](#_Toc127432069)

[Figure 19 Combining the series of exogenous factors 36](#_Toc127432070)

[Figure 20 Building and fitting model, then predicting rice prices (Multivariate) 36](#_Toc127432071)

[Figure 21 Seasonal Variations in Rice Prices](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432072) III

[Figure 22 Average Rice Prices by Month](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432073) III

[Figure 23 Rice Retail Prices Line Chart](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432074) IV

[Figure 24 Gantt Chart](https://d.docs.live.net/1b74801032bcbacd/PSPD.docx" \l "_Toc127432075) V

# LIST OF TABLES

[Table 1 Research Objectives 3](#_Toc127432076)

[Table 2 Stakeholder Viewpoint 9](#_Toc127432077)

[Table 3 Requirement Elicitation Methods 10](#_Toc127432078)

[Table 4 Description of codes used in Thematic Analysis 11](#_Toc127432079)

[Table 2.5 Details of Interviewees 12](#_Toc127432080)

[Table 6 Thematic Analysis 12](#_Toc127432081)

[Table 7 Analysis of Findings from LR 15](#_Toc127432082)

[Table 8 Summary of Findings from Research Elicitation 16](#_Toc127432083)

[Table 9 Functional Requirements based on MoSCoW Prioritization 19](#_Toc127432084)

[Table 10 Non-Functional Requirements 20](#_Toc127432085)

[Table 11 Design Goals of Sri Oryzaugur 22](#_Toc127432086)

[Table 12 Frameworks planned for use in implementation 31](#_Toc127432087)

[Table 13 Languages planned for use in development 31](#_Toc127432088)

[Table 14 Libraries planned for use in implementation 32](#_Toc127432089)

[Table 15 IDEs to be used for development 32](#_Toc127432090)

[Table 16 Summary of the selected technologies 33](#_Toc127432091)

# ABSTRACT

Rice is the staple food item for the people of Sri Lanka, and it is a major part in the diet of all Sri Lankans. It is also a source of livelihood for various Sri Lankans. Forecasting prices of commodities are important to an economy as it will aid various stakeholders such as farmers, agribusinesses, and government policy makers to anticipate costs and predict price fluctuations to make better informed decisions. Unfortunately, it is quite troublesome to forecast rice prices accurately, especially in a developing country like Sri Lanka as prices are subject to various factors like supply, demand, production and so on.

There are system and frameworks built to forecast rice prices, although the models that have been developed by those research is univariate time forecasting models. Such models would not be able to grasp the complex nature of prices in developing countries and for a critical commodity like rice especially for Sri Lanka.

*Sri Oryzaugur* augurs or forecasts medium grain rice prices in a multivariate approach using covariates that are specific to Sri Lanka, moreover procedures to handle major disturbances in price trends due to local and global shocks are enacted.

**Keywords:** Time Series, Multivariate Time Series, Price Forecasting, Data Science, Machine Learning

**Subject Descriptors:**

* Mathematics of computing → Probability and statistics →Statistical paradigms → Time series analysis
* Applied computing → Operations research → Forecasting
* Applied computing → Law, social and behavioral sciences → Economics

# Chapter 1 INTRODUCTION

## 1.1 Chapter Overview

The author attempts to solidify a multivariate approach for forecasting the prices of rice in Sri Lanka, it will use various factors that are correlated dynamically to Sri Lanka. Different target audiences will each benefit in a unique way, for example if we farmers as a target audience, it would indicate the finest market to sell the rice and would help them plan better with the forecasts.

The problem, the research gap, the research challenge, and the research approach that the author plans to use over the coming several months are all outlined in this text. Additionally reviewed are the necessary problem proofs and earlier research areas of interest. Finally, the work plan presents the anticipated schedule for the project's deliverables.

## 1.2 Problem Background

The prices of food commodities are directly linked to the country’s economy, especially in developing countries like Sri Lanka where agriculture contributes 7.4% to the GDP and 30% of Sri Lankans work in that field (Sri Lanka - Agricultural Sector, 2021), price instability is a major concern. Any extraordinary price fluctuation in future markets indicate that forecasting food commodity prices are important (Antwi et al., 2022), and this would then be useful in Sri Lanka where the price fluctuations are extraordinary due to the current crisis (Food crisis in Sri Lanka likely to worsen amid poor agricultural production, price spikes and ongoing economic crisis, FAO and WFP warn | World Food Programme, 2022). A related research paper from the Journal of Agricultural Sciences in Sri Lanka mentions that farmers and consumers are stressed due to the lack of knowledge on price fluctuations (Basnayake et al., 2022). Further, it is also mentioned that forecasting variations in prices would help avoid situations that are destructive to the economy as better decisions can be made.

A study mentioned that 39% of farmers use Smartphones around 42% have access to the internet (Narmilan, Niroash and Puvanitha, 2020). Manoj Thibbotuwawa the Head of the Agriculture Economic Policy, Institute of Policy Studies of Sri Lanka (IPS) asserts that young to middle aged men do use mobile phones for some sort of task such as messaging, voicemails and so forth (Why the Transition to Smart Farming Is Critical in Sri Lanka, 2021), these studies suggest that there is a decent number of farmers who may have access to this solution.

Fortunately, we have much research that was done on predicting commodity prices with several machine learning, deep learning, and time series models in various countries. However, the consideration of other factors is redundant in most research and is considered future work in some of the papers. Moreover, this research is dynamic for Sri Lanka as it is more effective at forecasting than using a global model, “the weakness of global models is that they cannot fit each dynamic region very well” (Huang and Wu, 2018).

## 1.3 Problem Definition

There is an absence of research done for the forecast of rice prices in Sri Lanka, as mentioned earlier the more dynamic the model is to the country and commodity then the more accurate it is, various factors will be considered that are dynamic to the model will be evaluated and used in addition to the historical price data. A myriad of research papers exists where models are overfitted or there is little data and sometimes the accuracy appears to be extremely high, and if other factors are considered it seems to only use seasonal factors instead of also looking into non-seasonal factors and vice versa, nevertheless there is insufficient research done for the forecast on the price of rice in Sri Lanka. Therefore, a multivariate approach model for forecasting commodities in a much more specialized manner is relevant rather than it being generalized through a framework or with the discussed problems.

### 1.3.1 Problem Statement

It is difficult to forecast the price of rice in Sri Lanka considering only factors that are economic and non-economic, this is done to benefit various target audiences in Sri Lanka

## 1.4 Aims and Objectives

### 1.4.1 Aims

*The purpose of this study is to develop and assess a multivariate time series model for forecasting rice prices in Sri Lanka. It will carefully assess the framework by considering a variety of factors and models with effective classifiers.*

To elucidate on the above aim, this study desires to produce a system that would help predict the price of rice more accurately. By utilizing historical price data and a myriad of other factors, a pair of models can be ensembled. Moreover, powerful classifiers can be used to improve the classification capabilities of the model.

To confirm the chosen hypothesis, there will be domain knowledge that has to be acquired, the forecasting model will be built, and its effectiveness will be assessed. The project can be accessed through a hosted site or a mobile app for all users.

### 1.4.2 Objectives

Table 1 Research Objectives

|  |  |  |  |
| --- | --- | --- | --- |
| **Objective** | **Description** | **Learning Outcomes** | **Research Questions** |
| Literature Review | Studied prior works to compile key facts on related works and properly evaluate them.   * **RO1**: Methodically analyze related works on the domain of price forecasting, and select a commodity based on the absence of adequate research * **RO2:** Identify how to build a more accurate model for that commodity * **RO3:** Research on existing techniques as well as univariate and multivariate time series models * **RO4:** Identify the factors that can be used for forecasting | L01, L06 | RQ2 |
| Requirement Elicitation | Delineating the project's specifications, the usage of the applicable strategies and tools to reach the research gap primarily based on earlier associated studies   * **RO5:** Examine with domain experts on the different factors that factor the price of rice and what type of rice is best to forecast. * **RO6:** Identify the data available for the scoped factors * **RO7:** Solidify requirements so that a satisfactory outcome based on the proposed system is feasible | L02, L03, L04, L05, L06 | RQ1, RQ2 |
| Design | The design of this system would target the below mentioned parts,   * **RO8:** Each time-series model that will be used to create the ensembled model must be selected and assessed * **RO9:** Scheme the process of ensembling the selected pair of models | L03, L04, L06 | RQ3 |
| Implementation | The proposed design must now be implemented,   * **RO10:** Based on the selection made, a pair of time series models will have to be developed * **RO11:** The developed models will have to be ensembled for the main goal of the research which is to have better accuracy * **RO12:** The ensembled must be exhibited in a site or mobile app | L03, L04, L06 |  |
| Evaluation | It is extremely important to evaluate the model as it determines the performance of the system built,   * **RO13:** Establish a test plan for the test plan as well as functional testing * **RO14:** Compare the forecasts of the ensembled model with actual data ahead of the current point in time * **RO15:** Receive comments from industry experts, namely economists and researchers from agriculture institutes | L04, L06 |  |

## 1.5 Novelty of the Research

### 1.5.1 Problem Novelty

Multivariate forecasting for rice pries has not been done previously, in addition to this there is no research implemented for dynamically for Sri Lanka. While the research prior to this has focused on univariate rice price forecasting, this project will consider a multitude of factors that are relevant in predicting price movements for rice. This allows for a more comprehensive and accurate forecast.

### 1.5.2 Solution Novelty

There is an absence of multivariate forecasting models for rice prices in Sri Lanka.

## 1.6 Research Gap

The research gap that the author addresses is to do with forecasting the price of rice in Sri Lanka, historical prices along with various other factors like seasonal and economic indicators. This can negate the common gap within the related research papers that is the issue of the model not being dynamic to Sri Lanka or it being univariate, also the several recommendations given due to limitations in other research like considering various other and newer models with all this information, ensembling models and by using more powerful classifiers to improve the classification of the model.

## 1.7 Research Contribution

### 1.7.1 Technological Contribution

This will be a multivariate time series model that uses several factors and models that have not been looked at with regards to predicting the price of rice in Sri Lanka, as mentioned earlier based on the commodity and Country different algorithms and different factors dynamic to it will be needed for a better accuracy. This system would provide a far more accurate and well evaluated Model with more specialization into good factors and appropriate models for this use-case.

### 1.7.2 Domain Contribution

As mentioned earlier, there is little to no studies done on this proposed system and the existing ones have lots of room for development. The author received confirmation from economists that this would be helpful for them to forecast commodities targeted, and it is vital to also design an easy-to-use interface inclusive of all target audiences and generate summarized reasoning for why the result is as such.

## 1.8 Research Challenges

This research seems to be viable as their data with regards to market prices of commodities and economic indicators or other factors with regards to inflation, exchange rates and so forth. However, it seems that certain data would require me to reach out to organizations or authorities to get certain data and it has been done by previous research projects based in Sri Lanka.

With regards to knowledge known for the Problem Domain, the author has contacted a few economists and a senior researcher in paddy to get more domain knowledge. Since this is a topic affecting most people in Sri Lanka and that rice is an essential and widespread commodity, several other people who are not in the field have a decent amount of the topic which can be beneficial when determining the factors that are to be used exclusive of the historical price data.

Moreover, the scope will have to be limited, as the author receives more information from several sources to make the project more accurate and feasible such as focusing on possibly a certain type of rice or making the study more dynamic for the most plausible target audiences and suchlike.

## 1.9 Chapter Summary

Overall, the basic foundations for Sri Oryzaugur are placed down in this chapter. The research problem and gap has been discussed, most importantly the strategy or roadmap to achieve to address the gap has been put forward.

# Chapter 2: Literature Review

## 2.1 Chapter Overview

The prior chapter gave a brief rundown of the project in terms of the problem domain and its solution, whilst we will be taking a more comprehensive look into both topics. Various techniques, existing approaches and technologies will be critically assessed to build an accurate price forecasting system for rice in Sri Lanka.

## 2.2 Concept Map

[…]

## 2.3 Problem Domain

Rice occupies 13% food expenditure in a household as staple food is mostly consumed in Sri Lanka, furthermore there are 3.6 million farmers and several other individuals involved in the industry of rice/paddy (Wijesooriya, Kuruppu and Priyadarshana, 2021). Considering the scale or the field and how significant it is for both the diet of the population as well as the income of 3.6 million individuals it is in Sri Lanka; therefore, rice is not only the staple for the population in this the country, but it’s also an important factor for the development of Sri Lanka and a subtle change of the price would heavily impact the poor and low-income consumers.

### 2.3.1 Why was Time Series Forecasting chosen for this research?

The utilization of time series models have increased rapidly lately due to the myriad of time series data now available (Mullen, no date). The concept of using time series to forecast or to analyze the changes of a price or an economic variable in a given course of time (What Is a Time Series and How Is It Used to Analyze Data?, no date). As we can analyze the changes of prices, we will be able to identify which factors truly affect the changes in the price. This is exactly what is required for the project as we need to forecast the time series data for the prices of rice and analyze to mark what elements can cause fluctuations in the prices of rice.

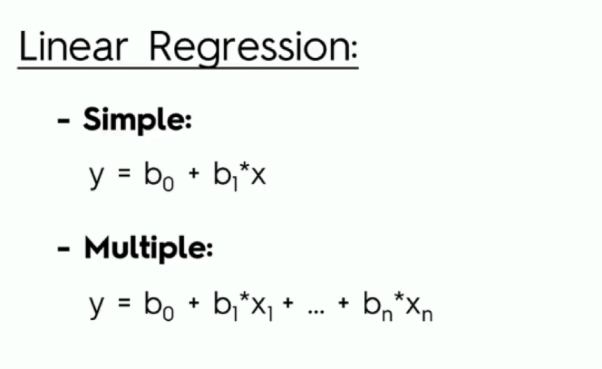
### 2.3.2 Introduction to Time Series Forecasting

### Time series forecasting is the process of predicting future values based on historical data. It can be used to analyze trends and produce forecasts for various types of data such as stock prices, commodity prices, economic indicators, and even climate patterns. In particular, multivariate models can be used to forecast rice prices in Sri Lanka by taking into account various factors such as the amount of rice produced, imported, exported, and stored; the climatic conditions; and the prices of other rice-related products in the region. With a multivariate model, we can get a better understanding of how all these factors interact with each other and make better predictions. Furthermore, by incorporating external data sources such as financial news and weather forecasts, we can further refine our forecasts and gain better insights into the state of the rice market in Sri Lanka.

#### 2.3.2.1 What is the difference between Univariate and Multivariate models

This time series considers only a singular variable or observation, primarily the historical data and in the context of this project it would be the historical price data. Whereas a time series model that investigates a several factors when forecasting is known as a multivariate model, this model is far more effective as correlations between independent factors can be assessed to discover covariates. These covariates would be the independent factors that have an effect of at least 10% to the outcome of the forecasting model, therefore covariates can be selected and used to finally develop a far more robust model.

The below diagram shows a comparison of 02 basic time series models, the first one is linear regression and the other is multiple regression. To simple breakdown the below formulas, *y* would be the dependent variable that we are attempting to estimate, *b0*is the constant, *b1* is the coefficient and *x* would be the independent variable that affects the dependent variable *y*. Now we if look at the function for multiple regression, its simply the same exact formula except that we now have multiple coefficients *bn* and multiple independent variables *xn* that we sum up in our formula.



### 2.3.3 What is Price Forecasting

Price forecasting involves evaluating a multitude of factors that could be seasonal, economic, etc... to estimate the prices of a specific good or service over a specified time horizon, and this is unlike something that we can observe objectively as this is not something that is quantifiable, which is why the best attempt to forecast prices must be done by considering various variables that can influence the estimate by a sizable extent (Forecasting with Price Elasticity of Demand, no date).

### 3.3.1 What is the Significance of Forecasting the Prices of Commodities

Forecasting the prices of commodities is vital for economic stability in Sri Lanka and other developing nations. Accurate and detailed forecasts enable business owners and policy makers to better manage the supply chain and to make informed decisions on the production, pricing, and distribution of goods. Forecasting the prices of commodities assists in stabilizing asset prices within a country, as it allows anticipated changes to be considered when designing policies or making decisions. This helps to reduce costs associated with inefficient practices, such as overstocking, or increased prices due to an unexpected rise in demand. Additionally, forecasts can provide a basis for potential commercial and financial investments, giving future investors an idea of what to expect from a particular industry. Forecasting the prices of commodities can also assist the government in ensuring the welfare of the nation’s citizens by reducing the risk of price increases that could cause economic hardship for consumers and businesses. By predicting changes in the market, businesses can adjust their stock levels to meet customer demands and governments can devise appropriate economic policies to meet changing needs. A number of sources can provide evidence about the significance of forecasting prices for commodities especially in Sri Lanka. To begin with, the International Monetary Fund (IMF) has conducted research on the effects of price forecasting in Sri Lanka, finding that it improved the functioning of the country's economy and enabled the government to better manage its financial resources

### 3.3.2 Why was Rice and Paddy selected and why is it Scoped to a Single Variate

As mentioned in the onset of the problem domain, rice is a critical commodity to our country in terms of the population’s diet, country’s economy, and the agriculture sector, however, is it beneficial for us to forecast the prices of rice or paddy. Paddy can be considered as an upstream product as it is a raw material that would later be milled into rice, which makes rice the downstream product. Technically speaking both rice and paddy are important since it’s simply a difference between supply and demand, paddy is based upon the supply of farmers and the rice is milled to be sold to consumers based upon their demand.

The variety of crops that farmers can plant are the different varieties of paddy, and the farmer selects the variety based upon the demand of that specific variety. Nadu rice or long-grain white rice is the most consumed variety of rice in Sri Lanka and therefore the author chose this variety of rice.

### 2.3.5 What are the Driving Factors that impact the Prices of Rice and Paddy

The prices of rice and paddy in Sri Lanka are largely impacted by factors such as the production levels of paddy and rice, the weather conditions that affect the productivity of the crop, demand, and supply of the crop within the country, and the import and export of the crop. In addition to this, the government policies, subsidies, and taxes also play a major role in determining the prices of rice and paddy in Sri Lanka. Furthermore, the demand from other countries and the global policies related to the crop also affect the prices of paddy and rice in Sri Lanka.

# 4 Existing Work

|  |  |  |  |
| --- | --- | --- | --- |
| Citation | Summary | Limitations | Contributions |
| (Zhang et al., 2020) | A novel model selection framework for predicting the prices of certain food commodities with forecast horizon. ANN, ELM and SVM were compared as models and RF and SVM were used as classifiers. | The framework is not dynamic to rice, and as per the discussed literature in previous headings, generalization is more inaccurate compared to specialization. Furthermore, better classifiers can be used. | Provides a robust framework for forecasting the prices of commodities generally, considering not only the historical price data but also the forecast horizon. It is well evaluated will several experiments and therefore accurate models have been selected. |
| (Pitigalaarachchi, Jayasundara and Chandrasekara, 2016) | An ARIMA Model and a VAR Model are being compared and the most performing one is being selected. The ARIMA model just uses the monthly Price data for gold, and the VAR Model uses historical price data along with other exogenous factor such as exchange rate, money supply, and inflation. | The model is not dynamic to rice, furthermore due to the lack of data, the selected model. | Provides a model that is dynamic to forecasting the prices of gold in a specialized way, this is because the several other exogenous factors are considered as well. |
| **(Basnayake et al., 2022)** | The price of green chilies in Sri Lanka is forecasted using Artificial Neural Networks over a Forecast Horizon that is Weekly. The implementation is as follows, ANNs are each paired with a separate Algorithm. In this case, the FFNN with SCG algorithm and the FFNN with the LM algorithm, and Time Delay Neural TDNN are used. | Factors other than Historical Price Data are not considered as it is univariate, also various other forecast models can be evaluated and used. | The model is well evaluated, and a variety of models are taken into consideration along with the assessment of the algorithms for each model. |
| (Ohyver and Pudjihastuti, 2018) | Developing an ARIMA model to forecast the prices of rice in Indonesia using historical price data. | This is approach is way too simple for a complex problem like forecasting prices as this is a univariate solution to a problem that requires a multivariate solution, only one time series model is evaluated, and further feature scaling and classifiers can be used. | The ARIMA model itself is well built and has been evaluated well. |
| (Udumulla, no date) | A multivariate time series system to forecast the prices of coconuts in Sri Lanka, considered are 03 correlated features and to each feature a time series model is evaluated and used. ARIMA and SARIMA are used for appropriate factors, and a VAR model is used to improve the overall forecasts. | Seasonal and other economic factors can be considered, moreover newer models can be used. The model can also be evaluated better. | There is a lack of multivariate solutions for this domain, and this solution helps fill in that gap to produce more research that is based in Sri Lanka. Since it is multivariate it is certainly more befitting for price forecasting, a proof of that is that the selected models are dynamic to each factor. |

## 4.1 Univariate Approaches

## 4.2 Multivariate Approaches

# 5 Technological Review

## 5.1 Time Series forecasting techniques

# 6 Evaluation

## 6.1 Benchmarking

# 7 Chapter Summary

# Chapter 2: SRS

## 2.1 Chapter Overview

This chapter explores all stakeholders associated with the project and examines their interactions with the system through a rich picture diagram. By gathering their perspectives, we are able to analyze and develop potential use cases, functional and non-functional requirements for our prototype.

## 2.2 Rich Picture Diagram

A picture containing diagram

Description automatically generated

Figure 1 Rich Picture Diagram

A rich picture diagram can help to identify and understand stakeholders' needs, interests, goals, and processes, as well as outline the various activities required to achieve the goal of providing an accurate price forecast. It can also provide insights into the wider system within which the project is situated and serve as a powerful tool to facilitate communication and collaboration between team members, stakeholders, and decision makers.

## 2.3 Stakeholder Analysis

The Stakeholder Onion Model illustrates recognized stakeholders who are associated with  
the system, along with an explanation of each stakeholder’s involvement in the system, in  
Stakeholder Viewpoints

### 2.3.1 Stakeholder Onion Model

Radar chart

Description automatically generated with low confidence

Figure 2 Stakeholder Onion Model

### 2.3.2 Stakeholder Viewpoints

Table 2 Stakeholder Viewpoint

|  |  |  |
| --- | --- | --- |
| **Stakeholder** | **Role** | **Benefits/Role Description** |
| Government Policy Maker | Operational, Functional Beneficiary | Policy makers will be able to make better informed decisions on their policies, such as in providing subsidies for farmers, export tariffs and so forth. |
| Farmers | Farmers can make better decisions of which crop to harvest that gives them the most profits or for storing crops to till prices increase. |
| Agribusiness | Agribusinesses can develop better plans and strategies for the future, specifically in sales and market opportunities. |
| Support Staff | Operational Beneficiary | Provide feedback, ideas and also ensure that Sri Oryzaugur  is meeting its objectives. And most importantly providing support to users. |
| Investors | Financial Beneficiary | Investors can help fund the project so that more data can be purchased to be used for the models and to cover deployment and maintenance cost, they would receive returns by investing. |
| Social Media | Neighboring System | Social media can help advertise Sri Oryzaugur, this can make target audiences like farmers and agriculture related businesses to be aware of it. |
| Project Manager | Product Owner | They would overlook the project, ensure data collection is robust and that systems in general are running well. |
| Developer | Product Development | Creating, maintain and troubleshooting the code for Sri Oryzaugur that is in agreement with industry standards. |
| DevOps Engineer | Product Deployment and Development | They would host and maintain the servers, websites, and database of the system. |
| Data Scientists | Data Analysis and Modeling, Product Development | Analyze the multivariate data and generate forecasting models. They must also validate and tune the model so that prices are accurately forecasted. |
| QA Engineer | Quality Control | Reveal any vulnerabilities in the system and forecasts, essentially, they must ensure that the system is ready for releases |
| Experts | Experts | Experts would consist of both technical and domain experts. For Sri Oryzaugur, domain experts are crucial as prices forecasting is a complex (Bowman and Husain, 2004), further the technical experts would provide aid to the data scientist to solve issues in forecasting that are discussed with domain experts and also in general |
| Supervisor | Experts | Oversee the tasks of the developer and provide them guidance |
| Competitors | Negative Stakeholder | They would take away target audiences away from the system. |
| Hacker | Negative Stakeholder | They could manipulate data and forecasts by striking at weak points or vulnerabilities of the system. |

## 2.4 Requirement Elicitation Methods

The author selected interviews as a research elicitation method for forecasting rice and paddy prices in Sri Lanka because it gives the researcher direct access to the knowledge and experience of local experts who can provide first-hand information on rice and paddy prices. Through interviews, the researcher can gain insight into the information and perceptions of the informants, assess their awareness of local market trends, and understand how these trends are likely to affect prices in the future. Moreover, interviews allow the researcher to probe deeper into the information provided by informants, enabling a more accurate analysis of the dynamics and likely trajectory of prices. Additionally, the interview format is well-suited for developing a rapport with the informants, encouraging them to be more open and honest in their responses. In this way, interviews allow the researcher to rapidly acquire reliable information from experts on the likely direction of prices, making them an ideal research elicitation method for forecasting rice and paddy prices in Sri Lanka.

Table 3 Requirement Elicitation Methods

|  |
| --- |
| **Method 1: Interviews** |
| As price forecasting is very complex and requires the evaluation of a multitude of factors, interviews are a great way to attain a far better understanding of the problem in a specialized manner compared to other methods (Steber, no date). Furthermore, since this solution is not aimed at a general audience but rather certain individuals primarily government policy makers, interviews will give the author insights on the key factors directly from the experts which is vital help build a robust model. |
| **Method 2: Literature Review** |
| Literature review aids the author to draw on existing research to help brief their own solution and also to get a more comprehensive grasp on the current rice market in Sri Lanka. It provides the author with an opportunity to validate insights provided by experts from prior interviews. |
| **Method 3: Prototyping** |
| Prototyping is essential in identifying requirements as various time series models will have to be tested and based upon the requirements of the tested models, the specifications of the project itself would have to be altered. |

## 2.5 Analysis of Data & Presentation of the Outcome through Elicitation Methodologies

This topic targets to analyze data and present the outcome through elicitation methodologies for Sri Oryzaugur. An accurate model is the aim of this project, thereby findings must come under scrutiny to incorporate all the complex aspects of the problem to develop a good solution.

### 2.5.1 Findings from Interview

As mentioned earlier, price forecasting is very complex and contains a multitude of variables to consider and therefore requires the assistance of experts. The interviews were conducted with a senior research officer, two research officers, an economist, and one data scientist. The author was enlightened to various challenges and vulnerabilities with the proposed solution, and ultimately certain factors were chosen to be used for the multivariate model. Everything discussed with the interviewees will be uncovered below by a **thematic analysis**

Table 4 Description of codes used in Thematic Analysis

|  |  |
| --- | --- |
| **Codes** | **Themes** |
| Time series, Price forecasting, agriculture, Target audience | Research problem and gap |
| Rice variety, Forecast horizon | Scope of the research solution |
| Datasets, multivariate time series | Requirements for robust commodity price forecasting |
| Evaluation methods, Economic crisis, Pandemic, Government Policy | Handling shocks in time series |

Table 2.5 Details of Interviewees

|  |  |  |
| --- | --- | --- |
| **Participant ID** | **Affiliation** | **Expertise Area** |
| P1 | Senior Research Officer at HARTI | Academic Journals, Scientific Publishing, Agriculture and Rural Development |
| P2 | Research Officer at HARTI | Agricultural Policy, International Trade and Environmental Economics |
| P3 | Research Officer at HARTI | Agronomy, Agricultural Plant Science and Agricultural Economics |
| P4 | Associate Economist at Asian Development Bank | Financial Analysis, Data Analysis and Economic Research |
| P5 | Data Scientist at 99x | Computer Vision, Audio, Supervised Learning and Machine Learning |

Table 6 Thematic Analysis

|  |  |  |
| --- | --- | --- |
| **Theme** | **Analysis** | **Evidence** |
| Research problem and gap | In general, the domain experts did acknowledge that this idea would be beneficial to primarily government policy makers and also farmers. However, there were differing responses for it benefiting businesses in agriculture sector, and therefore just these 02 target audiences were selected. Furthermore, they asserted that the general public would not really benefit from this project. | “I think for farmers and businesses in that sector, it would be really useful. Because they would want to know if they could have accurate forecasts of rice prices. The rice prices in future will help them make them make better decision on production and sales. But not in general”  “Consumers can't benefit much on it, farmers will benefit. They will decide how long to store paddy in order to get a good price. And the main advantage to government policy makers” |
| Scope of the research solution | Prices for paddy or in other words the producer prices would also have to be forecasted as they simultaneously rise alongside with the retail prices.  Based on the responses the author decided to go forth with a single variety that is medium-long grain as it is consumed by 50% of the country (Rice in Sri Lanka, no date).  The frequency of the forecast recommended is monthly and the forecast horizon for the next 12 months seems to be ideal, although short-term forecasts are naturally more accurate than long-term ones (Measuring Forecast Accuracy: The Complete Guide, 2017). Another interviewee suggests that the author forecasts for the next couple of months, this indicates short-term forecasting. Since there is a concern due to accuracy and that it would be insightful to have the long-term forecast as well, the author will generate models for both time windows. | “It is better to look and the both the prices of paddy and rice, because both prices are simultaneously rising”  “It’s better to go with one [variety] and do it perfectly and accurately”  “The next 12 months are important, […] a monthly forecast horizon for the next year is preferable”  “Obviously the really long term is something you can’t think about, and short term what the rice price would be in the next couple of months […]” |
| Requirements for robust commodity price forecasting | It is crucial to gain domain knowledge when forecasting, this is due to the delicate nature of rice prices in Sri Lanka as changes in the price of rice are subject to various other factors (Thibbotuwawa, 2021). The domain experts unanimously agreed on most factors such as the ones provided in the evidence for this theme, and they are categorized into economic and seasonal factors, the factors also are categorized into supply and demand factors. | “There are independent and dependent variables, climate variables are difficult to consider. Cost of production, marketing cost, fuel cost, demand prices and rice consumption can be used. Fertilizer cost has a huge impact, […]. Cost of production for particular years data is there. Farm gate prices for Ampara, Anuradhapura, Polonnaruwa, Hambantota, Kurunegala.” |
| Handling shocks in time series | The main concern for all the interviewees was the shocks causes due not only by Covid-19, but by more specific and local issues like the fertilizer ban and the economic crisis where food inflation rose by 80% (Sri Lanka’s food crisis: causes and consequences, 2022). To counter this the author will use methods such as the Kalman Filter that will help the model adjust to shocks (Grogan, 2018), alongside the extremely handy feature provided by the Facebook Prophet model will allow for considering these periods as holidays which in turn applies a negative effect to that period. | “I am telling you to do it up to 2019 because 2020 is not a normal year, Covid-19 and all and 21 is the fertilizer ban and 22 is the financial crisis, so those are the shocks […]” |

### 2.5.2 Findings from LR

Table 7 Analysis of Findings from LR

|  |  |
| --- | --- |
| **Finding** | **Citation** |
| After studying this literature, the author became even more assured of the statements from the interviewees. This is because the cited literature describes exactly how the stakeholders would benefit in the same way the interviewees did so, companies would be able to make better investment and policy decisions and the government would be able to predict budget expenses which in turns would help create better informed policies | (Kwas and Rubaszek, 2021) |
| The author of the cited paper indicates that the implemented VAR model on forecasting gold prices in Sri Lanka, can also be used in forecasting commodity prices. | (Pitigalaarachchi, Jayasundara and Chandrasekara, 2016) |
| A univariate approach is taken for this implementation for the same time for rice that Sri Oryzaugur focuses on which is medium grain rice, the author of this paper advocates for a more a robust forecasting which is the primary goal of this project. | (Ohyver and Pudjihastuti, 2018) |
| This paper is important to validate the hypothesis asserted by an interviewee, where he mentioned certain districts in the country that produce a surplus of paddy. These districts are Anuradhapura, Ampara, Polonnaruwa, Kurunegala and Hambantota, considering the producer prices from these specific districts will help to generate a more accurate forecast as most surplus is produced by these districts and then distributed to the main Colombo market. | (Wijesooriya, Kuruppu and Priyadarshana, 2021) |
| An interviewee indicated that there is seasonality in the monthly prices for rice, the cited article describes that there it is cyclical each year that can be seen in Appendix I.  This can be verified during the exploratory data analysis stage of prototyping. | (Thibbotuwawa, 2021) |

### 2.5.3 Findings from Prototyping

Evolutionary prototyping is used since the prototype for Sri Oryzaugur was built on the minimum requirements of the project and then it started to expand from there onwards. This method of prototyping requires that the developer understands the necessary requirements well (Sherrell, 2013), the author understood the requirements well by having conversations with numerous experts that are knowledgeable on both the economic and non-economic side of this problem.

Sri Oryzaugur was first built with basic univariate forecasting models using the primary historical factor of rice prices, then the multivariate forecasting was implemented. After the implementation of multivariate forecasting, the requirement of multistep forecasting was discovered as the values of external regressors would be needed for future forecasting and therefore predictions must be made for these exogenous factors additionally

By prototyping the author was able to validate the claim on monthly seasonality derived from both the literature and interviews, the histogram in Appendix II shows the average prices of medium grain rice in a monthly basis from 1996 to 2022. It can be seen that the trend of prices dropping in March then prices rise after April creating a slight bulge which peaks at June, this bulge ends at September and rises rapidly till January of the following year.

Lastly, the shocks due to the economic crisis were identified, it can be depicted in the shock from the line chart in Appendix III after 2021. It is even reflected upon producer prices.

## 2.6 Summary of Findings

Table 8 Summary of Findings from Research Elicitation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Finding** | **Interview** | **LR** | **Prototyping** |
| 1 | Corroborate that the forecasting system would be beneficial to the selected target audiences | X | X |  |
| 2 | A short-term forecast horizon would work well along with a monthly frequency | X |  | X |
| 3 | The exogenous factors that would affect the price of rice would be producer prices, paddy production, cost of production and fuel prices | X | X |  |
| 4 | Certain districts in Sri Lanka produce surplus that is distributed to the rest of the country, the producer prices for these districts would be more relevant. | X | X | X |

## 2.7 Context Diagram

Diagram

Description automatically generated

Figure 3 Context Diagram

## 2.8 Use Case Diagram

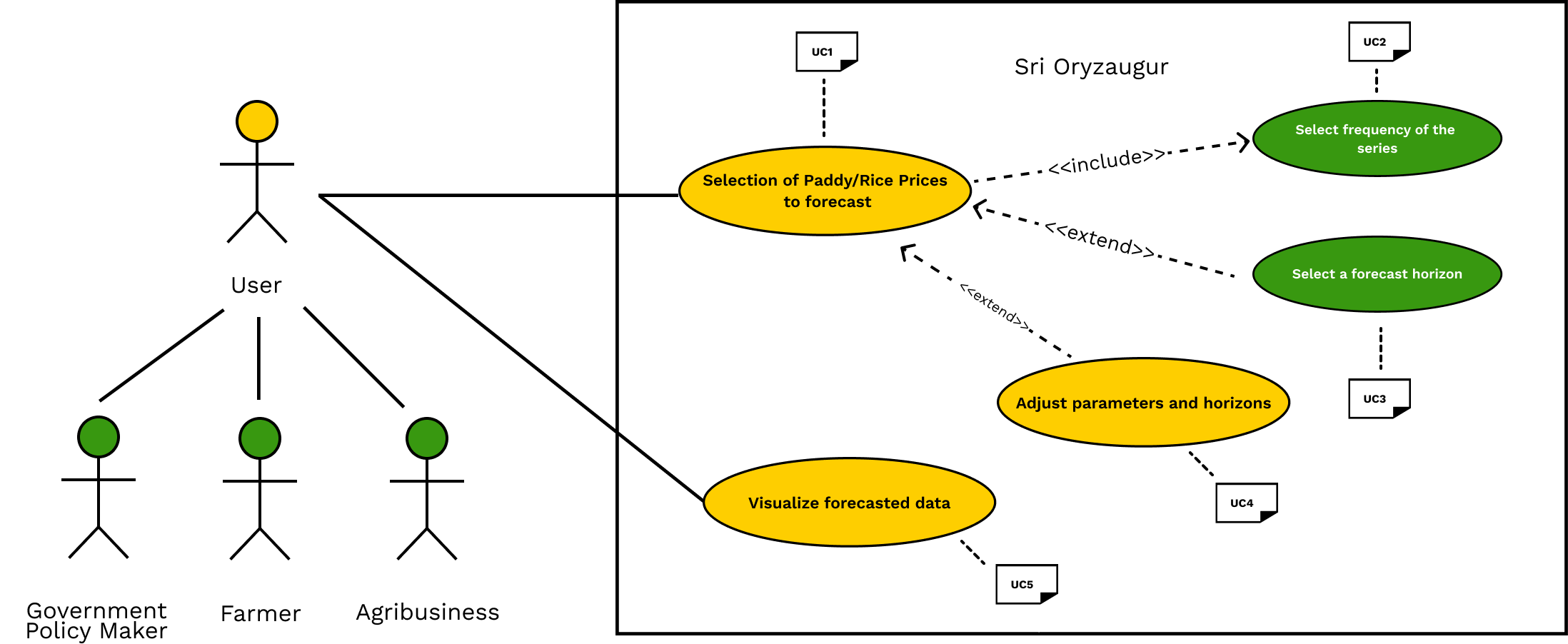


Figure 4 Use Case Diagram

## 2.9 Use Case Description

Table 8 Use Case Description for UC5

|  |  |
| --- | --- |
| Use Case | Visualize forecasted data |
| Use Case ID | UC5 |
| Description | Visualize forecasted data for either rice or paddy prices established by the selected options, this is to provide the user with insight on future trends. |
| Primary Actor | User |
| Supporting Actors (If any) | N/A |
| Stakeholder and interests (If any) | Government Policy Makers, Farmers, Agribusiness |
| Preconditions | The system should be able to cater to the options that the user selects, data that needs to be preprocessed to develop the models should be done and any adjustments are made to the parameters of the model must be considered. |
| Postconditions | The model must start generating forecasts based on the selected options. |
| Trigger | A user navigates to the forecasts page, then makes their selections and clicks a button to submit. |
| Successful scenario | * System provides forecasts for the selected options * Model caters to the adjusted parameters * Forecasts are displayed and filterable in various visualization charts |
| Variations | The variations of the forecasts are dependent on the options that the user selects, i.e., if the user goes with a short term horizon, they will be provided with the predicted prices for the next 3 months, and the forecasts for the next 12 months would be displayed if a long term horizon was selected. |

## 2.10 Requirements

### 2.10.1 Functional Requirements

Table 9 Functional Requirements based on MoSCoW Prioritization

|  |  |  |  |
| --- | --- | --- | --- |
| **FR ID** | **Requirement** | **Priority Level** | **Use Case** |
| FR1 | Users must be able to select if they wish to forecast rice or paddy prices | M | UC1 |
| FR2 | Users must be able to select a forecast horizon | M | UC3 |
| FR3 | Incorporate exogenous factors such as production cost, rate of production and so forth | M | UC1 |
| FR3 | Users must be able to view the forecasted prices within the click of a button in a line chart | M | UC5 |
| FR4 | Users must be able to modify parameters and features | S | UC4 |
| FR5 | Options to view visualizations in various types of interactive charts | S | UC5 |
| /.l,, | Providing forecasting for other varieties of paddy | C | UC1 |
| FR7 | Provide reasonings for the forecasted values | C | UC5 |

### 2.10.2 Non-Functional Requirements

Table 10 Non-Functional Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **NFR ID** | **Requirement** | **Description** | **Priority Level** |
| NFR1 | Performance | The system should be designed and optimized to provide a fast response time with minimal latency when processing data or providing forecasts. It should also enable efficient usage of computational resources, storage space, and memory. | Desirable |
| NFR2 | Usability | Usability in the current context should focus on providing a user-friendly interface that allows users to easily navigate and use the various features of the system. It should include an intuitive design, clear instructions, and visualizations that help the user to quickly understand the data being processed and the forecasts generated. It should also provide users with the ability to customize the forecasting model and parameters according to their needs and preferences. | Desirable |
| NFR3 | Quality | The forecasts must be robust and accurate as possible, different models must be evaluated and the one with the lowest possible error rate must be used. Data collection and preprocessing must be carried out in a reliable manner to ensure that the time series data is precise. | Important |
| NFR4 | Security | In terms of security, the main goal would be to ensure that the gathered data is safeguarded from hackers. Security measures should be put in-place to curtail this. Other than this, there is not any concern as the system would not require user authentication as all the data and forecasts are available publicly. | Luxury |

## 2.11 Chapter Summary

To wrap up the chapter, the author digs into the proposed rice price forecasting system in Sri Lanka, Sri Oryzaugur. By developing a Rich Picture Diagram and also by using Saunder’s Onion Model, the connection of stakeholders as well as the way they influence each other is apparent. Now the requirements must be collated and assessed, the author interviewed various experts and verified their assertions by reviewing literature, then requirements are gathered and specified for both functional and non-functional requirements with consideration of the stakeholders. Ultimately, I will use case diagrams to visually explain the system’s use cases.

# Chapter 3 Methodology

## 3.1 Chapter Overview

[…]

## 3.2 Research Methodology

Table 11 Research Methodology

|  |  |
| --- | --- |
| Philosophy | The researched candidates were Positivism, Interpretivism and Pragmatism. Subsequently, **Pragmatism** was adopted as this research is **quantitative** and **qualitative**, utilizing the mathematical and logical proofs from **quantitative** data and the **qualitative** data from such research would help build a much solid foundation on this complex topic of economics. |
| Approach | Considering that this system would be trying to coalesce prior frameworks and systems or certain aspects within them, the **Deductive** approach is befitting. |
| Strategy | Primarily **interviews** will be used to answer research questions, additionally there is rich **literature** when it comes to price forecasting as attached with it comes also the various aspects affecting the price can be used. Part of primary market research is to **observe** the market, this can improve the forecasting accuracy as well as provide more insight (Observational Methods to Forecast Market Potential, no date). |
| Choice | The **mixed** method is suitable here because of there being both **quantitative** and **qualitative** data. To further elaborate, we are looking at. Therefore, the **multi** method would not be befitting here, and unmistakably the **mono** method would not be appropriate. |
| Time Horizon | **Longitudinal** studies are appropriate in contrast to **cross-sectional** studies, considering that we are evaluating the predicted prices overtime contrary to looking at data at only a certain point of time. |
| Techniques and procedures | Data will be gathered from government and organizational records and libraries nationally statistics, interviews and reports will be utilized. |

## 3.3 Development Methodology

## 13.2 Development Methodology

### 13.2.1 Prototype Methodology

Taking into consideration that the requirements for this system are not explicit, it would be more suitable to use the prototyping methodology.

ISQTB (The International Software Testing Qualifications Board) mentions that "*instead of freezing the requirements before a design or coding can proceed, a throwaway prototype is built to understand the requirement*” (as cited in Volchko, no date), to elucidate on this quote with the aid of figure 2, the prototype methodology would not require clear or set requirements as we develop a prototype then asses what we build and develop it further.

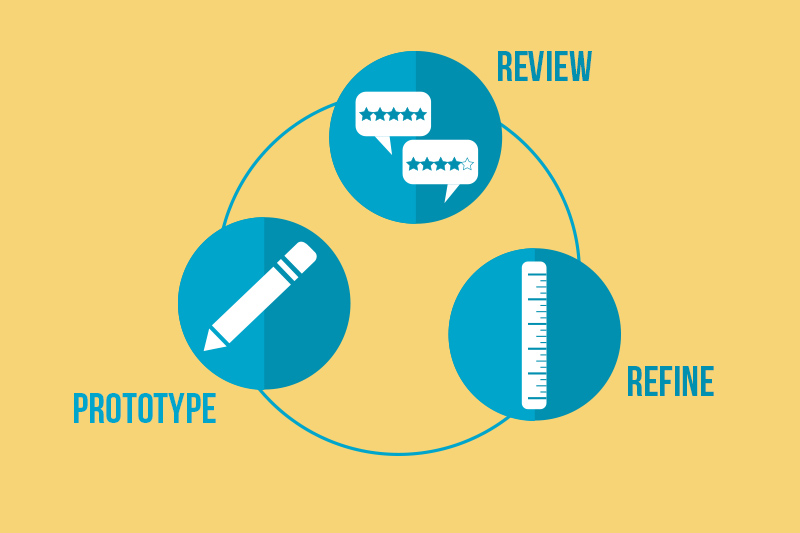


Figure 5 Prototyping Model (Felderman, 2018)

### 13.2.2 Requirement Elicitation Methodology

### Interviews

Seeing that the problem extends to multiple domains along with the solution, dialogues with industry experts that cover each domain is essential. Currently the author is in contact with economists as well as experts in the paddy domain.

### Literature

Without a doubt reviewing existing literature is vital to even carry out interviews or to get in touch with industry experts, it comprehends analysis prior works, content on the domain and the appropriate techniques to implement the system.

### Observations

There are two types of methods to forecast market prices that are fundamental and technical analysis. When it comes to fundamental analysis, studying the market in terms of demand and supply is crucial to determining whether the market is in deficit, equilibrium or oversupply (Fundamental Commodity Forecasting - a Tricky Business, no date), moreover exogenous factors other than demand and supply like policies are crucial in forecasting. Technical analysis encompasses studying market trends, technical analysts believe that it is a worthwhile indicator for future price movements (Technical Analysis: What It Is and How to Use It in Investing, no date). Both these observation types are crucial when developing the model and scrutinizing the requirements.

### 13.2.3 Design Methodology

Object-Oriented Analysis and Design was selected due to its flexibility or its gradual approach to designing the system, it is homogeneous to the agile method (Various approaches of systems analysis and design, no date). It is also in-sync with the project management methodology of the project that is agile due to the iterative nature of both methods.

### 13.2.4 Development Methodology

**Procedural programming involves dividing the programing into sections of code or procedures inconstant to disintegrate the program on data and each of these procedures tries to achieve a task, it is appropriate here as it follows a top-down methodology where the flow of the system can be seen easily, further it is performant and is simple** (says, 2019)**.**

### 13.2.5 Evaluation Methodology

### Model Testing

Part of evaluating the model would be to compare simulated results contemplatively with historical prices.

*“The ideal backtest chooses sample data from a relevant time period of a duration that reflects a variety of market conditions. In this way, one can better judge whether the results of the backtest represent a fluke or sound trading.”* (Backtesting: Definition, How It Works, and Downsides, no date)

The above statement asserted by an expert trader further strengthens the stipulated evaluation methodology, because model testing is ideal in cases where a multitude of market conditions or various factors are considered

### 13.2.6 Solution Methodology

Collecting the data required to form the dataset would come from a myriad of sources, this is due to the exogenous factors would have their own sources which has to then be collected. Furthermore, data merging will be used to improve accuracy, and will be inculpated in cases where there is more than one reliable source for historical price data it will be merged. Then once the dataset is formulated, it would be trained, and models appropriate for each factor will have to be assessed. After that these models would have to be ensembled and powerful classifiers can be used to improve its classification capabilities. Finally, the evaluation of the model will have to take place and there are several metrics in the time series that can be used, in addition to this back testing will be done to prove that this system can handle complex scenarios like forecasting in the best possible manner.

## 3.4 Project Management Methodology

[…]

## 3.5 Research Methodology

[…]

## 3.6 Resources

[…]

## 3.7 Risk and Mitigation

[…]

## 3.7 Chapter Summary

[…]

# Chapter 3: Design

## 3.1 Chapter Overview

This chapter outlines the blueprint of Sri Oryzaugur by utilizing high- and low-level designs, UI wireframes and various other design diagrams. The main goal of this chapter is to elucidate on why exactly the author has selected the design choices for the system.

## 3.2 Design Goals

Table 12 Design Goals of Sri Oryzaugur

|  |  |
| --- | --- |
| **Design Goal** | **Description** |
| Performance | The forecast should have the lowest error rate that possible, and the model with the lowest error forecast errors. Another concern that would impact performance is the nature of multistep forecasting alongside with the various choices that the user has in forecasting. To counter this pre-processed data must be stored in a database, and GPU acceleration should be used to reduce the usage on other hardware requirements which would in-turn improve performance (Filelis - Papadopoulos, Morrison and O‘Reilly, 2022) |
| Correctness | The forecast should have the lowest error rate that possible, and it should attempt to achieve a curve that is close to the curve of actual figures, the system must consider the most relevant features for the multivariate analysis. In addition to this, taking the advice of experts, the author will evaluate the model on dates before the years where there are shocks in the trend. This is because these shocks are unpredictable, for example Covid-19 could not have been using any of the historical features we analyze. And since the rising price trend due to the shock is very recent, future forecasts after it would not be able to be evaluated. |
| Usability | Exploring Sri Oryzaugur must be convenient for the users in cases when selecting the forecast option, horizons and so forth. They also must be able to view historical information or forecasts in a clear and interactive graph alongside with legends and colorings to convey the data and the associated predictions. |
| Scalability | When future data is fed into the model, it should be able to accommodate that. Furthermore, the backend should be able to smoothly handle different requests for forecasts and so on. |
| Adaptability | Although it may not be able to evaluate the model during the years where shocks are present due to 03 major events that affected rice prices, the model should apply a negative affect for these dates and various techniques have to be applied to ensure that future forecasts stay robust. |

## 3.3 High Level Design

### 3.3.1 Architecture Diagram

A screenshot of a video game

Description automatically generated with medium confidence

Figure 3.16 Three-Tiered Architecture Diagram

Figure 6 Architecture Diagram

### 3.3.2 Discussion of Tiers

**Data Tier**

1. Multivariate Time Series Data – As discussed many times already in this research, price forecasting is not easy and especially in developing countries like Sri Lanka, prices are not stable and are subject to a multitude of changes. There are different unprocessed data files for each factor below,
   1. Retail Prices [RRP]
   2. Fuel Prices [FP]
   3. Cost of Production [CoP]
   4. Farmgate Prices [FGP]
   5. Paddy Production [PP]
2. Preprocessed Historical Data – As mentioned in the prior component of this tier, the multivariate data is not processed and ready to be used for time series analysis and forecasting. Therefore the unprocessed data must be passed onto data preprocessor in the logic tier, then subsequently it would be stored in the his component.

**Logic Tier**

1. Data Preprocessor – This component is key in combining the different files that were gathered by the author into a single data frame where the index would be dates for years between 1996-2022 in a monthly interview, the different features for multivariate forecasting should be in sync with the dates in the index and yearly data for factors like production of paddy must be distributed. Other actions in this component are interpolation, which is a good way of filling a few missing rows certain datasets had (Lepot, Aubin and Clemens, 2017), and also making sure that the data frame is meeting the requirements of how certain models expect features to be named and suchlike.
2. Multistep Forecasting Model – There are 02 multivariate models available, the primary one that forecasts retail price for rice and the later is for producer paddy prices.
   1. Retail Prices Multivariate Model – This model forecasts retail prices for medium grain rice in Sri Lanka.
   2. Farmgate Prices Multivariate Model – This model forecasts producer prices for paddy
   3. RRP, FP, CoP, FGP, PP, Models - Since multivariate forecasting requires future data for the external regressors, these univariate models will forecast future values up to the horizon that the forecasts decrees and it will be ultimately handed over to both the multivariate models. An important thing to note is that the Retail Price Model (RRP) is required for the farmgate prices multivariate model as they are both correlated and vice versa for the primary rice price forecasting model.
3. Hyperparameter Tuner – This module will help optimize the parameters of the multivariate forecasting models to achieve the best possible results.
4. Model Evaluator – This module will use various time series metrics to evaluate the different time series models that the author creates, and the one with the lowest error rate would be applied.
5. Backend API – The purpose of this component is to connect the components of the presentation tier to the logic tier. APIs will be generated so that the logic from the backend can be integrated into the frontend.

**Presentation Tier**

1. User Inputs and Options Wizard – Here is where the user can select forecast horizons, frequencies, tune parameters and actions of that sort can be done here.
2. Historical Values and Forecast Visualizations UI – The user will be able to access the historical figures of all the features that are used by the model, by means of various graphs that they prefer. Alongside this they should be able to foresee the forecasts that they have requested in the Options Wizard.
3. Feature Information and Guide UI – This part of the application on really connected to any other component of the architecture as it is simply a guide on how to use Sri Oryzaugur, it also contains much needed and useful information about the factors in the multivariate model as well as the recent shocks in Sri Lanka.

## 3.4 Low Level Design

### 3.4.1 Choice of Design Paradigm

This project will use **Structured Programming** as its design paradigm. It is a far more logical means for this program as it emphasizes on dissecting the components of our system into simple tasks by relying heavily on the user of functions unlike classes, this makes the code easier to run and debug (Lithmee, 2019). Object Oriented Programming has many perks and allows for complex behaviors, but it is not needed for this project and it would extend development time to follow this paradigm for no significant gain or benefit.

## 3.5 Design Diagrams

### 3.5.1 Data Flow Diagram

Diagram

Description automatically generated

Figure 7 Level 1 Data Flow Diagram

### A picture containing diagram Description automatically generated3.5.2 System Process Chart

Figure 8 System Flow Diagram

### 3.5.3 System Design

Graphical user interface, application

Description automatically generated

Figure 3.4 Wireframe 1/3: Historical Data Visualization

Figure 9 Wireframe 1/3: Historical Data Visualization

# A picture containing graphical user interface Description automatically generated

Figure 10 Wireframe 2/3: Selection and Option Wizard

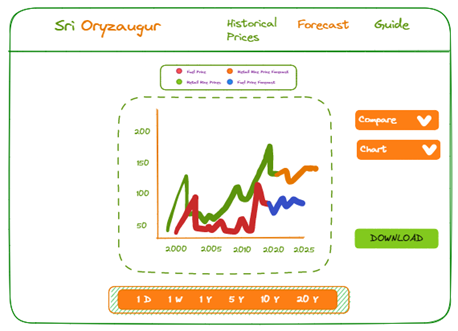


Figure 11 Wireframe 3/3: Forecasted Data Visualization

## 3.6 Chapter Summary

Now that the initial design for the system has been laid out clearly for the requirements that the author gathered, the implementation of the prototype can commence. It should be built according to the specified architecture diagram following the flow that is described by the system flow chart, In addition to this, the technologies to build the proposed UI through a wireframe must be decided in the upcoming chapter

# Chapter 4: Implementation

## 4.1 Chapter Overview

In this section we will be going over the rationale over the selection of forecasting models,

technologies, frameworks, and languages used to develop the prototype for Sri Oryzaugur.

## 4.2 Technology Selection

### 4.2.1 Technology Stack

Timeline

Description automatically generated

Figure 12 Technology Stack

### 4.2.2 Dataset Selection

First, it’s important to list out what are the factors that require a dataset,

* Prices of medium grain rice
* Prices of paddy sold by producers from surplus paddy producing districts
* Production of paddy
* Cost of production
* Fuel Prices

Let us get into the specifics of why and how the datasets were collected for the above factors as well as their time horizons.

All the datasets for the aforementioned factors were collected from HARTI, and if a dataset contained a missing field, then the author would use data from the Central Bank of Sri Lanka. The main factor here is the historical prices of Nadu or medium grain rice, this is on an interval of months as per the discussion with domain experts.

As for the exogenous factors, the producer prices are also in an interval of months. Paddy production and cost are annual figures as that is the nature of the data, they will be replicated for the preferred forecast horizon. Fuel prices are monthly figures.

### 4.2.3 Development Framework

Table 13 Frameworks planned for use in implementation

|  |  |
| --- | --- |
| **Framework** | **Justification for selection** |
| Flask | It is lightweight and gives access to various tools to work with databases as well as to create APIs. |
| Chakra UI | Elegant UI components that are modular and easy to use for React applications. |

### 4.2.4 Programming Languages

Table 14 Languages planned for use in development

|  |  |
| --- | --- |
| **Language** | **Justification for selection** |
| Python | This is the primary language for the backend as well as the data science component. R would be the main contender against Python since both languages have amazing support for machine learning, however R is more suited for traditional statistical models for time series in comparison to the ones accessible in Python (CQF, 2022). It is also extremely convenient that Flask is based on python as there is more ease in terms of integration. |
| HTML | HTML is essential in developing any website, and there is no apparent reason of using an extension language like HAML or PUG. |
| TS | TypeScript offers much faster performance than vanilla JavaScript, another pro of using TS is that there is integration with React and the reduction of errors as it is a strongly typed language (Fruhlinger, 2020). |
| SASS | Styling becomes much easier and less cumbersome due to the nested syntax, accordingly it will be used. |

### 4.2.3 Libraries

Table 15 Libraries planned for use in implementation

|  |  |
| --- | --- |
| **Library** | **Justification for selection** |
| Pandas | Most popular library for data manipulation in python, due to its extensive features for analysis, cleaning and so forth. It’s use will be crucial in all sections of the model code and especially in data wrangling. |
| Darts | Accessibility to a plethora of both univariate and multivariate forecasting models, it also provides us with rich functionalities for time series forecasting including even hyperparameter tuning for the models they provide. |
| Matlibplot, Ploty | Both these powerful visualization libraries that can be used to analyze our time series data and forecasts. |
| React | Lightweight and performant UI library, this is useful because as there would be a large quantity of time series data. |
| Nivo | This library will give access to a various interactive chart to view time series data in our React frontend. |

### 4.2.6 IDE

Table 16 IDEs to be used for development

|  |  |
| --- | --- |
| **IDE** | **Justification for selection** |
| Kaggle | This platform will allow the author to easily upload, access, modify and maintain the notebook as well as the datasets for the project. Other reasons are the hardware specifications are superior in comparison than other popular platforms such as Google Collab. |
| PyCharm | For local development of the model as well as the backend development of the system, PyCharm was excellent. This is because of the huge array of functionalities provided by the IDE such as visualization, debugging and code completion specifically for Python. And since Flask was selected as the backend technology for this project, it makes sense to go with this IDE as it is tailored for Python as mentioned earlier. |
| VSCode | Due to the flexibility of this editor, together with the excellent extensions for web development with even built in support for React which is the frontend technology of the project. VSCode will be utilized to develop the frontend. |

### 4.2.7 Summary of Technology Selection

Table 17 Summary of the selected technologies

|  |  |
| --- | --- |
| **Component** | **Technology Selected** |
| Development Frameworks | Flask, Chakra UI |
| Programming Languages | Python, HTML, TS, SASS |
| Libraries | Pandas, Darts, Matplotlib, Ploty, React, Nivo |
| IDEs | Kaggle, PyCharm, VSCode |
| Version Control | Git, GitHub |

## 4.3 Implementation of Core Functionality

### 4.3.2 Data preprocessing

The data that was gathered by the author was not prepared as series of values as they were presented in a form of months by years, moreover the data was provided in different files and some of the data had to be distributed over a monthly horizon. The code snippet below melts the given data based on the month and year into a timestamp.

Graphical user interface, text, application, email

Description automatically generated

Figure 13 Transforming dataframe into monthly timestamps into monthly timestamps

As mentioned preivously, the values of certain features had to be adjusted to a monthly horizon. Such is the case of the paddy produciton data, as the based on the season which spans for more than a single month, the production of paddy varies.

Graphical user interface, text, application

Description automatically generated

Figure 14 Distributing paddy production for Maha and Yala seasons

There has to be data augmented to ensure that the gathered time series data is as robust as it can possibly be, any missing or empty rows would be like below filled using spline interpolation.

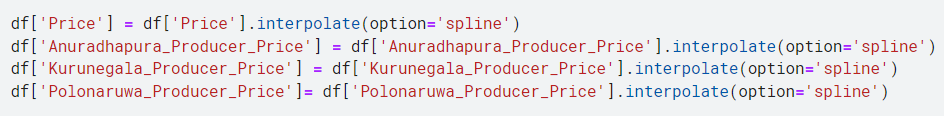


Figure 15 Filling missing data with interpolation

### 4.3.2 Feature preprocessing

The *Darts* time series library is extremely handy for time series problems. First the dataframes must be converted into the TimeSeries object provided by *Darts*. After they are converted, the time series data must be normalized between 0 and 1.

Text

Description automatically generated

Figure 16 Converting dataframes into series

Graphical user interface, text

Description automatically generated

Figure 17 Scaling timeseries data

### 4.3.2 Training and building forecasting models

**Univariate**

The code for the implementation of the NBEATS model, this deep learning model will be used for a univariate time series forecasting of retail rice prices. The author has assessed various other models in addition to this. It is being fit with the target series and then in the following cell, the price is predicted over the next 12 months for the retail rice prices series.

Graphical user interface, text, application

Description automatically generated

Figure 18 Building and fitting model, then predicting rice prices (Univariate)

**Multivariate**

To develop the multivariate model, the past covariates of the target series must be stacked. The method below stacks series of some factors such as producer prices and paddy production.

Graphical user interface, text, application, email

Description automatically generated

Figure 19 Combining the series of exogenous factors

An LSTM is neural network proposed by experts as well as literature, *Darts* also provides this model. It is being fit with the target time series and the stacked covariates are passed in.

Graphical user interface, application, Word

Description automatically generated

Figure 20 Building and fitting model, then predicting rice prices (Multivariate)

## 4.4 Chapter Summary

The plan set for implementing the system has laid out, in the matter of the tech stack and the final datasets that are to be used for the multivariate time series forecast. Furthermore, the core aspect of the implementation code for the prototype has been explained, and the low Fidelity wireframe for the final Minimum Viable Product (MVP) has been developed.

# Chapter 5: Conclusion

## 5.1 Chapter Overview

This chapter brings an end to the research done by creating the prototype of this project. The author discusses any changes made from the initial proposal of this project, both in-terms of scope and any change in the timeframe of building the system. Furthermore, the results provided by the initial prototype are evaluated, also upcoming adjustments and improvements will be listed.

## 5.2 Deviations

### 5.2.1 Scope Related Deviations

The implementation of the initial proposal was not deviated severally, the main focus was to have an ensembled or hybrid of a multivariate model to forecast rice prices. And the shift in focus is more towards the multivariate model before even considering hybrid or ensemble methods, as firstly several models would have to be compared and evaluated. Ensembling would not guarantee the best results at all times. There would most probably be changes in the selected factors for forecasting, as the author tests and compares different models.

### 5.2.2 Schedule Related Deviations

There are major changes in the past, present and future timeframes in the Gantt chart submitted for the project proposal. The literature review chapter had started according to the scheduled date; however, there was no submission for it and therefore many tasks would be dissolved into the extended literature review.

The requirements gathering stage did start according to plan, as interviews were conducted during the month of September and literature was reviewed subsequently. However, many interviews and other requirements were impeded due to various factors, therefore the analysis went over the ending period of the initial timeframe for this chapter. And since the design chapter depends on the requirements chapter to model diagrams, there was a delay in the completion of the chapter.

In terms of implementing the prototype, there was a delay of a month to commence development. This is due to other coursework and examinations that the author had to adhere to, the rest of the plan will go as initiated which is the incremental improvements. Again, the testing and evaluation phase was delayed as implementing the system is needed to evaluate it.

## 5.3 Initial Test Results

To evaluate the system, a univariate model will be compared to a multivariate model. MAPE will be used to identify which model has the lowest error percentage, it is a better metric to evaluate time series models due to its simplicity (RMSE vs MAPE, which is the best regression metric?, 2022). MAPE is essentially a percentage of the difference between the predicted forecast and the actual value, this simple but key metric helps us understand the model developed.

When forecasting rice prices with a LSTM model using exclusively retail prices, a MAPE of 9.98% is obtained. The LSTM model with past covariates, generates a MAPE of 6.91%. This indicates that the multivariate approach provides approximately a 3% lower MAPE and is therefore clearly more suited for forecasting. The multivariate model still has many improvements to go through and the MAPE for it will likely reduce drastically as the future improvements towards the mode are carried out.

## 5.4 Required Improvements

* At the moment, there are a few models used in the prototype. And therefore, a criteria to select time series models for forecasting must be developed.
* Hyperparameter tuning can be evaluated to ensure that the system is using the model that produces the lowest possible error.
* The UI for the system yet has to be developed together with the backend.
* Data has to be further refined by augmenting the data from various sources.
* Presently the system has been evaluated with forecasted before the shocks in the data as discussed earlier in the Software Requirements Chapter. However the model must also be able to make predictions after the last shock experienced by the data, this has to be done in a manner that imposes a negative affect for the sudden changes in figures due to these shocks. |

## 5.5 Demo of the Prototype

## 5.6 Chapter Summary

With this chapter, the exploration of the various stages for precursor of the final system of Sri Oryzaugur comes to an end. There is yet a significant amount of work to do in order to develop a Minimum Viable Product (MVP), Predominantly the multivariate analysis needs to be improved further and the methods like hyperparameter tuning as well as ensembl9ing must be implemented and evaluated, so that the author creates a forecasting model that uses the data and technology under the tech stack to a very high degree,

# REFERENCES

Antwi, E. et al. (2022). Modeling and Forecasting Commodity Futures Prices: Decomposition Approach. *IEEE Access*, 10, 27484–27503. Available from https://doi.org/10.1109/ACCESS.2022.3152694.

Basnayake, B.R.P.M. et al. (2022). An Approach for Prediction of Weekly Prices of Green Chili in Sri Lanka: Application of Artificial Neural Network Techniques. *Journal of Agricultural Sciences – Sri Lanka*, 17 (2), 333–349. Available from https://doi.org/10.4038/jas.v17i2.9746.

Bowman, C. and Husain, A.M. (2004). Forecasting Commodity Prices: Futures Versus Judgment.

CQF, V.M., PhD, MBA. (2022). Python vs R for time-series forecasting. *Medium*. Available from https://valeman.medium.com/python-vs-r-for-time-series-forecasting-395390432598 [Accessed 9 November 2022].

Filelis - Papadopoulos, C.K., Morrison, J.P. and O‘Reilly, P. (2022). GPU Accelerated Modelling and Forecasting for Large Time Series. In: Groen, D. de Mulatier, C. Paszynski, M. et al. (eds.). *Computational Science – ICCS 2022*. Lecture Notes in Computer Science. 2022. Cham: Springer International Publishing, 398–412. Available from https://doi.org/10.1007/978-3-031-08757-8\_33.

Food crisis in Sri Lanka likely to worsen amid poor agricultural production, price spikes and ongoing economic crisis, FAO and WFP warn | World Food Programme. (2022). Available from https://www.wfp.org/news/food-crisis-sri-lanka-likely-worsen-amid-poor-agricultural-production-price-spikes-and-ongoing [Accessed 6 October 2022].

Forecasting with Price Elasticity of Demand. (no date). *Investopedia*. Available from https://www.investopedia.com/articles/economics/09/price-elasticity-of-demand.asp [Accessed 14 December 2022].

Fruhlinger, J. (2020). What is TypeScript? Strongly typed JavaScript | InfoWorld. Available from https://www.infoworld.com/article/3538428/what-is-typescript-strongly-typed-javascript.html [Accessed 8 February 2023].

Grogan, M. (2018). Working with Outliers and Time Series Shocks.

Huang, S.-C. and Wu, C.-F. (2018). Energy Commodity Price Forecasting with Deep Multiple Kernel Learning. *Energies*, 11 (11), 3029. Available from https://doi.org/10.3390/en11113029.

Kwas, M. and Rubaszek, M. (2021). Forecasting Commodity Prices: Looking for a Benchmark. *Forecasting*, 3 (2), 447–459. Available from https://doi.org/10.3390/forecast3020027.

Lepot, M., Aubin, J.-B. and Clemens, F. (2017). Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water*, 9, 796. Available from https://doi.org/10.3390/w9100796.

Lithmee. (2019). What is the Difference Between Structured and Object Oriented Programming. *Pediaa.Com*. Available from https://pediaa.com/what-is-the-difference-between-structured-and-object-oriented-programming/ [Accessed 8 February 2023].

Measuring Forecast Accuracy: The Complete Guide. (2017). *RELEX Solutions*. Available from https://www.relexsolutions.com/resources/measuring-forecast-accuracy/ [Accessed 6 February 2023].

Mullen, B. (no date). Council Post: The Ubiquity Of Time Series Data Isn’t Coming—It’s Already Here. *Forbes*. Available from https://www.forbes.com/sites/forbescommunicationscouncil/2022/06/16/the-ubiquity-of-time-series-data-isnt-coming-its-already-here/ [Accessed 4 December 2022].

Narmilan, A., Niroash, G. and Puvanitha, N. (2020). Assessment of Current Status on Smart Farming Technologies in Batticaloa District, Sri Lanka. *Sri Lankan Journal of Technology*, 7.

Ohyver, M. and Pudjihastuti, H. (2018). Arima Model for Forecasting the Price of Medium Quality Rice to Anticipate Price Fluctuations. *Procedia Computer Science*, 135, 707–711. Available from https://doi.org/10.1016/j.procs.2018.08.215.

Pitigalaarachchi, A., Jayasundara, M. and Chandrasekara, V. (2016). Modeling and Forecasting Sri Lankan Gold Prices. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 27, 247–260.

Rice in Sri Lanka. (no date). *LANKA EXCURSIONS HOLIDAYS*. Available from https://www.lanka-excursions-holidays.com/rice.html [Accessed 6 February 2023].

RMSE vs MAPE, which is the best regression metric? (2022). *Stephen Allwright*. Available from https://stephenallwright.com/rmse-vs-mape/ [Accessed 16 February 2023].

Sherrell, L. (2013). Evolutionary Prototyping. In: Runehov, A.L.C. and Oviedo, L. (eds.). *Encyclopedia of Sciences and Religions*. Dordrecht: Springer Netherlands, 803–803. Available from https://doi.org/10.1007/978-1-4020-8265-8\_201039 [Accessed 7 February 2023].

Sri Lanka - Agricultural Sector. (2021). Available from https://www.trade.gov/country-commercial-guides/sri-lanka-agricultural-sector [Accessed 19 October 2022].

Sri Lanka’s food crisis: causes and consequences. (2022). *Sunday Observer*. Available from https://www.sundayobserver.lk/2022/12/11/business/sri-lanka%E2%80%99s-food-crisis-causes-and-consequences [Accessed 7 February 2023].

Steber, C. (no date). In-Depth Interviews: Data Collection Advantages and Disadvantages. Available from https://www.cfrinc.net/cfrblog/in-depth-interviewing [Accessed 3 January 2023].

Thibbotuwawa, M. (2021). talkingeconomics - Rising Price of Rice in Sri Lanka: The Roots and Remedies. Available from https://www.ips.lk/talkingeconomics/2021/10/07/rising-price-of-rice-in-sri-lanka-the-roots-and-remedies/ [Accessed 7 February 2023].

What Is a Time Series and How Is It Used to Analyze Data? (no date). Available from https://www.investopedia.com/terms/t/timeseries.asp [Accessed 4 December 2022].

Why the Transition to Smart Farming Is Critical in Sri Lanka. (2021). *Development Asia*. Available from https://development.asia/insight/why-transition-smart-farming-critical-sri-lanka [Accessed 3 November 2022].

Wijesooriya, N., Kuruppu, V. and Priyadarshana, D. (2021). *Rice Value Chain in Polonnaruwa, Sri Lanka Hector Kobbekaduwa Agrarian Research and Training Institute 114, Wijerama Mawatha Colombo 7, Sri Lanka*.

# Appendix A – Introduction

|  |  |
| --- | --- |
| Re |  |
|  |  |
|  |  |

# Appendix B – Software Requirements Specification

Chart, line chart

Description automatically generated

Figure 1

Figure 21 Seasonal Variations in Rice Prices

Chart, bar chart

Description automatically generated

Figure 2

Figure 22 Average Rice Prices by Month

Graphical user interface, chart, line chart

Description automatically generated with medium confidence

Figure 23 Rice Retail Prices Line Chart

# Appendix C – Conclusion

Timeline

Description automatically generated

Figure 24 Gantt Chart