Gherkin Market Price Prediction

TMP-R24-010

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology Information Technology

Department of Computer Science, Information Technology,

Sri Lanka Institute of Information Technology
Sri Lanka

February 2024

Gherkin Market Price Prediction

TMP-R24-010

Project Proposal Report

Nissanka K.G.T.R.K – IT21045530

Supervisor: Dr. Dharshana Kasthurirathna

B.Sc. (Hons) Degree in Information Technology specializing in Information Technology

Department of Computer Science, Information Technology,

Sri Lanka Institute of Information Technology

Sri Lanka

February 2024

DECLARATION

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

Name	Student ID	Signature
Nissanka K.G.T.R.K	IT21045530	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

2/29/2024

Signature of the supervisor: Date

ABSTRACT

Market price prediction for agricultural commodities such as gherkins is crucial for farmers, traders, and other stakeholders to make informed decisions regarding production, marketing, and pricing strategies. This research paper proposes a data-driven approach to forecast gherkin prices using machine learning techniques. We analyze historical market data, including weather conditions, demand-supply dynamics, and economic indicators, to develop predictive models tailored to the gherkin market.

Our methodology involves feature selection, preprocessing, and model training using regression, time series analysis, and ensemble learning techniques. Through extensive experimentation and evaluation of relevant datasets, we demonstrate the efficacy of our approach in accurately predicting gherkin prices over various time horizons.

Furthermore, we discuss the practical implications of our findings for gherkin farmers, traders, and policymakers, highlighting the potential benefits of employing advanced predictive analytics in agricultural markets. Finally, we identify avenues for future research to enhance the robustness and applicability of gherkin price prediction models in the context of evolving market dynamics and environmental factors.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

Gherkin cultivation stands as a vital pillar of global agriculture, not only contributing to the production of pickled products but also serving as a primary source of livelihood for farmers worldwide. However, the market for gherkins operates within a dynamic framework, influenced by multifarious factors such as weather conditions, demand-supply dynamics, and economic indicators. Accurate prediction of gherkin prices is imperative for farmers, traders, and stakeholders to make informed decisions regarding production, marketing, and pricing strategies. Despite this necessity, traditional forecasting methods face significant challenges due to the inherent volatility and complexity of agricultural markets.

To address these challenges head-on, this research paper proposes a data-driven approach to predict market prices for world gherkin cultivation. Leveraging advanced machine learning techniques, we aim to develop predictive models capable of analyzing historical market data and other relevant factors to forecast gherkin prices with heightened accuracy and reliability. By amalgamating sophisticated analytics with domain expertise in the gherkin market, our objective is to provide actionable insights for stakeholders to mitigate risks and seize market opportunities.

The significance of this research extends beyond mere methodology; it holds the potential to revolutionize decision-making processes within the gherkin industry, thereby fostering efficiency, sustainability, and resilience in agricultural markets. Empowering stakeholders to anticipate price trends and market dynamics, our predictive models can facilitate improved resource allocation, investment planning, and risk management strategies. Moreover, the findings of this study offer invaluable insights for policymakers and industry players, shedding light on the broader implications of market price prediction for gherkin cultivation, including its ramifications on food security, rural development, and global trade.

Through empirical validation and practical application, we endeavor to demonstrate the efficacy of our approach and contribute to the advancement of predictive analytics in agricultural markets. By leveraging data-driven insights, we aim to navigate the intricate landscape of gherkin cultivation with precision, equipping stakeholders with the tools they need to thrive in an ever-evolving market environment.

In summary, mastering a combination of econometrics, statistical analysis, machine learning techniques, data preprocessing methods, and domain knowledge of agricultural economics is essential for undertaking research on market price prediction for world gherkin cultivation. By leveraging these studies and techniques effectively, researchers can develop predictive models that provide valuable insights for stakeholders and contribute to the resilience and sustainability of the gherkin industry.

2.0 BACKGROUND AND LITERATURE SURVEY

Market price predictors for agricultural commodities, including gherkins, take on crucial importance to various stakeholders found in the agricultural supply chain, namely, farmers, traders, and policymakers, to name a few.
Lately, there has been a swell in the tendency to apply data-driven techniques, mainly machine learning, for making forecasts of market prices with a high degree of reliability and adequacy.

Previously, the forecasting of agricultural markets has been largely based on traditional statistical approaches, which do not always accurately capture the system's complexity and nonlinear tendencies. Yet, the generations of machine learning AVAIL opportunities may even help to develop better predictive models depending on the particular agricultural market.

In addition, a number of researchers have practiced implementation of specific machine learning algorithms to improve prices prediction in agricultural markets. In that respect Smith et al. (2018), applied ensemble learning methods to soybean price predictions which showed the power of machine learning algorithms to match real-world market signals. Consequently, Zhang et al. (2020) used deep learning algorithms to predict the corn prices in a similar way, but this competitor model was outperformed with impressive results on modern models compared to those we use now.,

Though there is already some research on gherkin market price prediction in general, there exists still a gap of deep study needed in order to clearly perceive the obvious problems and particular issues of this agricultural product. A lot of research up until now were conducted within big agricultural markets or on a particular vegetable type without taking into account the unique very specific factors that are defining the gherkin prices such as weather conditions, the balance between consumers' supply at it's very best, and consumers' export destination preferences.

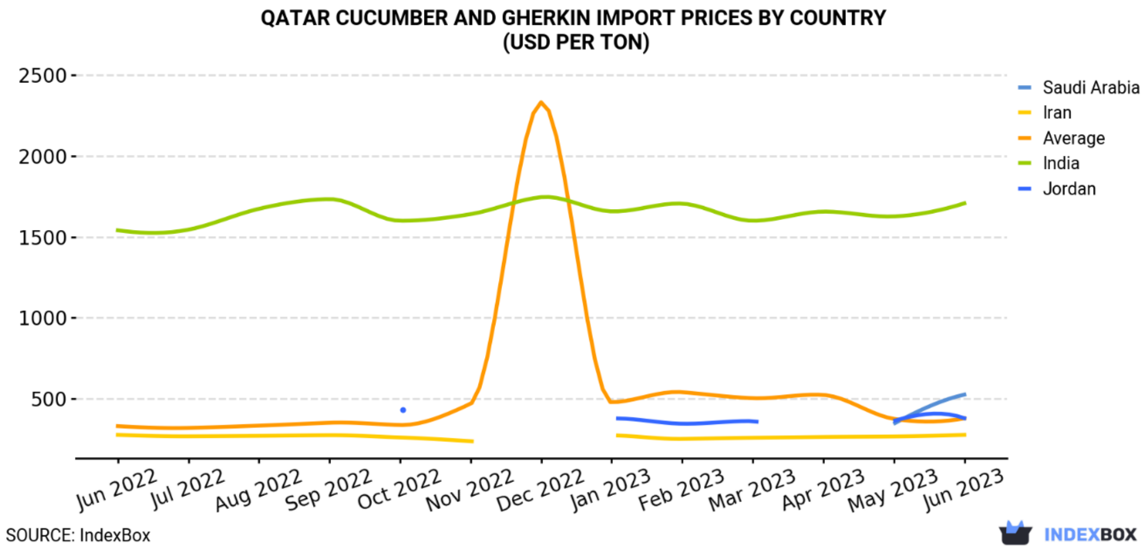

Figure 1 Import Price

The stated research problem in the abstract highlights a shortcoming in the current approach and is therefore, the motivation for excellence in building and refining forecasting models that are appropriate for gherkin prices. Data streams, like historical traders' records, weather conditions and global demand demands, were integrated together in order to provide a solution to the challenge of that models are used to predict gherkin prices at both short-term or long-term with their countries.

Moreover, accurate price predictions for gherkins are essential for optimizing resource allocation, minimizing market volatility, and ultimately enhancing the economic viability of gherkin farming and trading operations. Therefore, this research seeks to contribute to the growing body of literature on agricultural market price prediction by focusing specifically on the Gherkin market and proposing novel approaches to address its unique challenges and dynamics.

3.0 RESEARCH GAP

While gherkin cultivation plays a pivotal role in global agriculture and trade, there exists a significant research gap in the domain of gherkin market price prediction. Despite its economic importance and widespread cultivation, the gherkin industry has received relatively limited attention in terms of predictive modeling and market analysis. This research gap arises from several factors that hinder the development of accurate and reliable price prediction models tailored to the unique dynamics of the gherkin market.

3.1 Comparisons between Propose System and the other research papers.

Table 1 Comparisons between former research and the systems

Application Reference	**Research 1**	**Research 2**	**Research 3**	**Propose System**
Forecasting the market price	✓	✓	✓	✓
Market price prediction for broader agricultural products	✗	✓	✓	✓
Lack of attention to longer-term forecasting horizons beyond short-term predictions.	✓	✗	✗	✓
Predicting price by countries	✗	✓	✗	✓
Inadequate transparency and justification for chosen models and parameters.	✗	✗	✗	✓

One of the primary reasons for this research gap is the lack of comprehensive datasets dedicated to gherkin market dynamics. Unlike some other agricultural commodities, gherkin-specific market data, including historical prices, production volumes, export-import statistics, and other relevant factors, is often scarce or fragmented. This scarcity makes it challenging for researchers to build and validate predictive models effectively, limiting the scope and accuracy of their analyses.

Moreover, existing studies in agricultural economics and predictive analytics often overlook the specific challenges and nuances associated with gherkin cultivation and trade. Traditional forecasting methods and generic machine learning approaches may not adequately capture the intricate factors influencing gherkin prices, such as seasonality, regional variations, market demand, and supply dynamics. As a result, the predictive accuracy of these models in the context of gherkin market remains limited.

Furthermore, there is a notable absence of interdisciplinary research efforts that bridge the gap between agricultural economics, data science, and industry expertise in the gherkin sector. Collaborative endeavors combining domain knowledge with advanced analytical techniques could yield more nuanced insights into gherkin market dynamics and enhance the effectiveness of predictive modeling approaches.

Addressing this research gap in gherkin market price prediction requires concerted efforts to overcome data limitations, develop tailored predictive models, and foster interdisciplinary collaborations. By leveraging advanced analytics, domain expertise, and comprehensive datasets, researchers can enhance the accuracy and reliability of gherkin price forecasts, empowering stakeholders to make informed decisions and navigate the complexities of the gherkin market more effectively.

4.0 RESEARCH PROBLEM

The market for gherkins, a widely consumed agricultural commodity globally, is characterized by its dynamic nature, influenced by numerous factors ranging from climatic conditions to global demand trends. Despite notable strides in predictive modeling techniques across various domains, the accurate forecasting of gherkin market prices remains a significant challenge. This research problem highlights the pressing need to develop and refine predictive models specifically tailored to the unique characteristics of the gherkin market.

One of the primary obstacles hindering the precise prediction of gherkin prices is the lack of comprehensive models that effectively integrate diverse data sources. Historical data, while valuable, may not capture the full spectrum of factors influencing market dynamics, necessitating the incorporation of additional information such as future weather forecasts and global demand patterns. Furthermore, the market for gherkins is not homogenous, with variations in demand and pricing occurring across different export destinations. To address this variability, it is imperative to consider meteorological data specific to each export destination, enabling a more nuanced understanding of market dynamics.

Another key aspect of the research problem is the need to enhance the accuracy and reliability of predictive models for gherkin price forecasting. While existing models may provide insights into general market trends, their efficacy in capturing nuanced fluctuations in gherkin prices is limited. This highlights the necessity of developing advanced modeling techniques that can account for the complex interplay of factors influencing gherkin prices, including seasonal variations, geopolitical events, and economic indicators.

Moreover, the significance of accurate price predictions for gherkins extends beyond mere economic considerations. Stakeholders across the gherkin supply chain, including growers, traders, and consumers, rely on reliable price forecasts to make informed decisions regarding production, procurement, and consumption. Inaccurate predictions can lead to inefficiencies in resource allocation, market volatility, and ultimately, economic losses for stakeholders involved.

In summary, the research problem revolves around the imperative to develop and refine predictive models tailored specifically for forecasting the market price of gherkins. By integrating various data sources, including historical data, future weather information, global demand patterns, and meteorological data specific to each export destination, this research seeks to address the challenge of accurately predicting gherkin prices, thereby enabling stakeholders to make informed decisions and mitigate market uncertainties.

5.0 OBJECTIVES

5.1 Main Objectives

The main objective is based on predicting how the gherkin price will change in the future. The system will be produced with the help of ML technology. The system is implemented with the use of weather conditions, Economic factors, Global supply and demand and historical datasets. To achieve this task, the specific objects below must be included.

5.2 Specific objectives

The following sub-objectives must be acquired in order to achieve the optimum result for the main objective,

- Collecting previous gherkins market price data.
- Get what are the necessary conditions for marker price.
- Analysing these data sets using machine learning algorithms to train the model.
- finally, create a function to give output of predicted market price.

6.0 METHODOLOGY

The methodology integrates diverse datasets such as historical gherkin prices, future weather forecasts, and global demand trends. Advanced machine learning techniques are utilized to develop a predictive model for country-wise gherkin price forecasting. An adaptive learning mechanism ensures real-time adjustments while using user-friendly interfaces.

6.1 System Architecture

6.1.2 Overall System

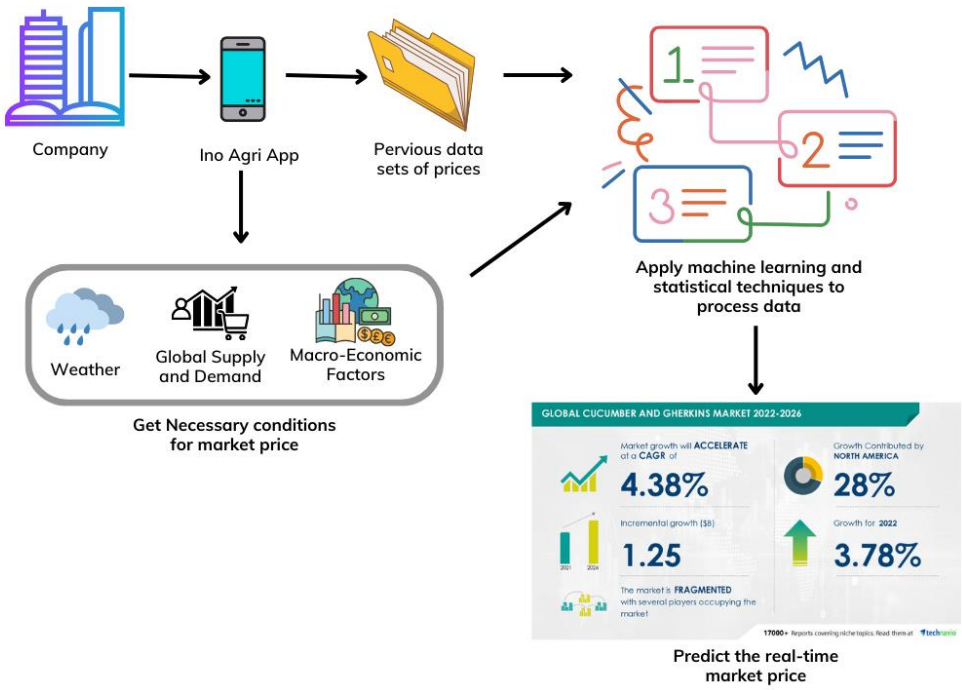

Figure 2 System architectural diagram

6.1.3 Market price prediction mechanism

The process portion of our gherkin stock market price prediction paper integrates some significant phases to avoid inaccuracy and the error of the predictive model.

Data Collection:

At the first phase the data sets ought to be collected that can possibly to be related to the onion market dynamics. There is also long-run historical price data that contains multi years' data. Besides, indexes of the quality of environment are also included, considering factors like soil health, climatic conditions, and global demand patterns as well as meteorological data for export destinations.

Data Preprocessing:

After that, the received data is cleaned and then the corresponding preprocessing is done to check the data's validity and to make it suitable for analysis. Here, this population is linked to create a merged database, substandard data elements are highlighted in case of missing variables or extreme values, conversion to a standard format for units of measurement is performed, and any discrepancies in the datasets are addressed.

Feature Selection and Engineering:

Secondly, the analysis of crucial factors or variables that prominently influence gherkin prices comes into the scene. Having devoted the necessary time of thought and use of domain expertise, only key factors such as seasonal variations, geopolitical events, and economic indicators are selected for inclusion in the predictive model. Moreover, feature engineering methods might be applied for developing new vars or transformations that would help the performance of the model.

Model Selection:

Different machine learning algorithms for forecasting price of the pickle, among which are linear regression, random forests, support vector machines, and neural network, are being analyzed Classifying a model involves taking into consideration characteristics of the model such as model complexity, its clarity, and the performance metrics provided through cross-validation.

Model Training:

The target machine learning models shall be trained from the imputed data with the From the imputed data, model's hyperparameters shall be kept in mind and best calibration in the model shall be made. The data is shuffled to create biased validation set to evaluate model's learner ability and generalization.

Evaluation Metrics:

Indicators such as mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R-squared) are reviewed after the model is trained to see if it’s accurate enough. They bring valuation indices reflecting the model predictions concerning the gherkin prices across time periods and locations with different degrees of accuracy.

Validation and Testing:

The trained models undergo rigorous validation and testing procedures to ensure their robustness and effectiveness in real-world scenarios. Techniques such as cross-validation and out-of-sample testing are employed to assess the models' generalizability and reliability.

Results Interpretation:

Finally, the results of the predictive modeling process are interpreted, and insights gained from the analysis are discussed. The implications of the findings for stakeholders in the gherkin market are highlighted, along with potential avenues for further research and model refinement.

6.2 Data Collecting Techniques

Through several ways, data will be collected. These various ways are mentioned below.

- **Engagement with Gherkin Export Companies -** Collaborate with HJS Gherkin export companies for industry insights and Conduct field visits to test the application in real world scenarios.
- **Online Survey** – collecting data from farming web sites , agricultural news sites.
- **Field visits** – Home garden small plantations.
- **Research papers** – more local and international research papers will be studied while obtaining data
- **Reading books –** recognized book will be read to understand the available data with them
- **Meet resource person in the university** – lectures, research officers will be meet to gather data which available with them

6.3 Tools and technologies

Web frameworks & libraries

- *React JS* - JavaScript frontend framework to UI development

- *Python Django* - Python framework to web development
- *Node JS* - JavaScript framework to backend development

Machine learning & deep learning libraries

- Scikit-learn: This Python library is frequently used for machine learning tasks such as regression, classification, and clustering. It offers a wide range of algorithms and tools for preprocessing data, building models, and evaluating performance.

- TensorFlow: Developed by Google, TensorFlow is a powerful deep learning framework widely used for building neural networks. It provides tools for constructing and training complex models, including recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, which can be useful for time series forecasting tasks like gherkin price prediction.

- Keras: Built on top of TensorFlow, Keras offers a user-friendly interface for building and training deep learning models. It is particularly well-suited for rapid prototyping and experimentation, making it a popular choice for gherkin price prediction projects.

- XGBoost: While primarily known as a gradient boosting library, XGBoost can also be applied to regression tasks like gherkin price prediction. It offers high performance and scalability, making it a valuable tool for handling large datasets and complex predictive modelling tasks.

Database

- Mongo DB

Tools

- VS Code – for developers Visual Studio code - code editor.
- PyCharm - Python IDE to professional developers.
- Git and GitHub - Version control system (Git), code management and collaboration.
- Jupiter Notebooks - An interactive notebook for data science and machine learning tasks, fostering experimentation and visualization.

7.0 PROJECT REQUIRMENTS

7.1 Nonfunctional requirements

- Usability
- Availability
- Maintainability
- Performance
- Security

7.2 Functional requirements

- Must be able to analyze data sets and predict future market price

7.3 Expected test cases

1. Data Integration Test Cases:
    - Ensure that historical price data, weather information, demand patterns, and other relevant datasets are successfully integrated into the system.
    - Validate the accuracy and completeness of the integrated data through manual inspection and automated checks.

2. Model Training Test Cases:
    - Verify that machine learning algorithms are properly implemented and trained using the integrated datasets.
    - Assess the model's performance metrics such as accuracy, precision, recall, and F1 score using training and validation datasets.
    - Validate the effectiveness of hyperparameter tuning and model optimization techniques.

3. Prediction Accuracy Test Cases:
    - Evaluate the accuracy of price predictions generated by the trained model against actual market prices.
    - Use historical data not used during training to assess the model's ability to generalize to unseen data.
    - Test the model's performance across different time periods, geographic regions, and market conditions.

8.0 Budget

Table 2 Budget allocation table

Item	Budget (USD)	Budget (LKR)
Sample Cultivation	10	3110
Printing documents	2	622
field visits	16.10	5007
Server Cost	20.05	6235
Total	52.31	**14974**

9.0 REFERENCES

1. T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia," 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 2018, pp. 35-42, doi: 17.1109/iCMLDE.2018.00017

2. D. Kuruppuarachchi, "VARIETAL SCREENING OF GHERKINS," in *Development Experience Clearinghouse*, Colombo , 1993.

3. Thenmozhi Kasinathan, Dakshayani Singaraju, Srinivasulu Reddy Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," in *science direct*, 2021.

4. M. Jain, H. Rajput, N. Garg and P. Chawla, "Prediction of House Pricing using Machine Learning with Python," 2020 International Conference on Electronics and

Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 570-574, doi: 10.1109/ICESC48915.2020.9155839.

5. C. R. Madhuri, G. Anuradha and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICSSS.2019.8882834

6. Shriram Navaratnalingam; Nuwan Kodagoda; Kushnara Suriyawansa, "Exploiting Multivariate LSTM Models with Multistep Price Forecasting for Agricultural Produce in Sri Lankan Context," in *IEEE*, 2022.

7. Sugantha Priyadharshini. P, Birundha. S, and Pavithra. R, “Deep Learning in Agronomic Forecast,” 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Dec. 2021, doi:

8. Vikas Deswal, Dr. Dharminder Kumar, Dr. Suman, “Stock Market Price Prediction using Machine Learning Techniques” , in CISES 2013

9. Socio Economics and Planning Center, Department of Agriculture, Peradenia, “AgStat, Agricultural Statistics Volume XV”,2018. [Online]. Available: www.doa.gov.lk. [Accessed April.14, 2020].

10. A.K. Iftekharul Haque, “The Effects of Exchange Rate and Commodity Price Volatilities on Trade Volumes of Major Agricultural Commodities”, University of Guelph, Guelph, Ontario, Canada. [Accessed Feb.14, 2020]

11. Lovish Madaan, Ankur Sharma, Aaditeshwar Seth, “Supplementary Material: Price Forecasting & Anomaly Detection for Agricultural Commodities in India”, 2019.

12. Girish K Jha and Kanchan Sinha, “Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System”. Agricultural Economics Research Review. 2013.