ADScAI
Applied Data Science & Artificial Intelligence Symposium

Data-Driven Analysis and Machine Learning-Based Forecasting for Big Onion Production in Sri Lanka

M.G.Hashini Nimeshika
Faculty of IT
Horizon Campus
Malabe, Sri Lanka
hashininimeshikamg@gmail.com

H.M.D.S. Herath
Faculty of IT
Horizon Campus
Malabe, Sri Lanka
dherath10@gmail.com

The cultivation of big onions forms a critical part of Sri Lanka's agriculture-based economy. However, the industry struggles with a challenge of supply and demand discrepancies, in addition to a high reliance on imports. This study proposes a machine learning-based predictive model to forecast production levels of big onions with a view to enhancing farmers' and policymakers' decision-making processes. Utilizing historical data from the Department of Census and Statistics and various institutions, a number of regression techniques including Linear Regression, Decision Tree, Random Forest, and Gradient Boosting were implemented to forecast production levels. The effectiveness of these methods was compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. It was revealed that with a set of defined hyperparameters, the Random Forest algorithm provides the best predictions with a score of 78.20 percent R^2. The developed model has been deployed as a user-friendly web application using Streamlit and FastAPI, thus making it easier for stakeholders to use. This study forms a stepping stone to further application in using data-driven methods to enhance production forecasting in agriculture. Further studies may focus on expanding the dataset and using sophisticated deep learning methods to achieve even better results.

***Keywords*—Big Onion Production, Machine Learning, Predictive Analytics, Random Forest, Agriculture Forecasting.**

I. Introduction

Agriculture is an important sector in the Sri Lankan economy, contributing around 8 percent of the GDP [5]. The big onion is an important agricultural commodity with high demand and great economic value. However, its cultivation faces significant challenges, including limited land, water scarcity, and the need to balance development objectives with environmental sustainability. Low productivity levels and poor market prices have negatively affected the income of farmers, threatening their long-term sustainability. As a result, local production is not enough to meet market demand, leading to significant yearly imports [6]. The major challenge in supply chain management of big onions comes from a lack of a strong forecasting approach to provide effective predictions in terms of production quantities. Statistical methods in use lack robustness to effectively capture complex relationships between agronomic and environmental factors that determine production levels. However, machine learning methods offer a viable alternative to enable more accurate, data-driven predictions. The aim of this study is to create a predictive model that brings into use historical production records, environmental factors, and machine learning techniques to forecast production of big onions. The results from this study aim to improve decision making in production planning and stabilization in markets by policymakers, farmers, and agribusiness stakeholders.

II. Literature Review

A multitude of research has explored the use of machine learning techniques in predicting agricultural results. Amarasingha et al. [1] considered impacts from climate change to agriculture in Sri Lanka and emphasized the importance of predictive modeling. Sabaragamuwa et al. [2] presented a thorough review of storage requirements relevant to big onions, suggesting a call for improved production management strategies. Thennakoon and De Silva [3] considered opportunities for big onion producers in markets and explained intricacies involved in supply and demand fluctuations. Recent studies, such as those by Wickramarathne and Chandrasekara [4] and have applied time series forecasting methods to forecast prices of different commodities. Overall, these studies reinforce the essential call for accurate forecasting models for unique crops and environmental factors.

III. Methodology

A. Data Collection

The dataset utilized in this study was collected from the Department of Census and Statistics and various institutions.

https://doi.org/10.31705/ADScAI.2025.18

It includes records spanning a period of 25 years and has variables such as total production in metric tons, cultivation area in hectares, average temperature in the atmosphere, relative day and night-time humidity, and rainfall. These factors are critical in understanding production patterns and developing a strong predictive model.

B. Data Pre-processing

To uphold data integrity, pre-processing included median imputation for numeric and mode imputation for categorical attributes, thus reducing bias. Outlier detection helped in reducing the impact of extreme values, whereas feature scaling using the StandardScaler enabled data normalization for effective model training.Categorical features, including the cultivation area and season, were transformed using one-hot encoding to make them compatible with the algorithms used.

C. Model Development

There were four machine learning algorithms attempted on big onion production prediction. The baseline was a Linear Regression. The Decision Tree Regression resulted in data hierarchies but in overfit. The best performance was achieved through Random Forest due to better feature interaction handling. The Gradient Boosting was also attempted but was worse than Random Forest in both MAE and R^2 scores.

D. Model selection and evaluation

The models were validated using MAE, RMSE, and R^2 score. A split in the original dataset to 70:30 provided poor performance, and a split to 80:20 was resolved to enhance accuracy. The final Random Forest model, with hyperparameter optimization using GridSearchCV, had a highest score in R^2.

E. Deployment Strategy

The final model was deployed using FastAPI for API integration and Streamlit for web-based access. This allows agricultural stakeholders to input relevant climatic and agronomic parameters and receive production forecasts in real time.

IV. Results and Discussion

Fig. 1 shows a comparison of performance displayed by all developed models with a focus on increased effectiveness displayed by the Random Forest model.

Model	MAE	RMSE	R^2
LR	846.32	1815.75	84.79
DT	1188.3	3704.04	36.70
RF	821.38	2173.78	78.20
GB	887.85	2417.96	73.02

Fig. 1: model performance (80:20 split).

Fig. 2 shows a feature importance ranking and thus confirming cultivation extent, temperature, and relative humidity to be the key factors affecting production predictions. Use of a user-friendly interface allows easy access by stakeholders and thus improves decision-making from a data perspective in onion production in Sri Lanka.

This system ensures accurate prediction of Sri Lankan big onion production, thus ensuring constant assessments of production needs. Through production-level prediction, the system helps in determining consumer needs as well as identifying possible imbalances between demand and supply. For instance, if the predicted production is lower during a given period (i.e., March through May), policymakers can take pre-emptive action, e.g., stimulating imports or augmenting domestic cultivation programs, in ensuring the stability of the marketplace while guaranteeing food security.

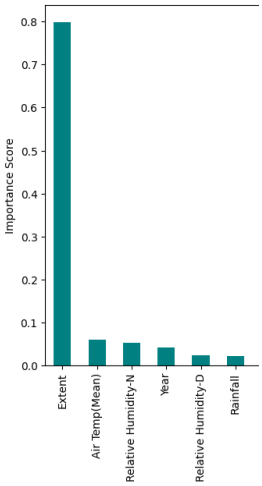

Fig. 2: Most influencing features

V. Conclusion and Future Works

This study developed a Random Forest-based model to predict big onion production in Sri Lanka, aiding planning despite limitations like missing soil data and complexity. Future work should include more agronomic factors, advanced methods, and broader crop coverage.

VI. Acknowledgment

We would like to thank Vice Chancellor Professor S.J.B.A.Jayasekara for his mentorship of our research work at Horizon Campus.

References

[1] Climate change impacts on crops in Sri Lanka. FAO, 2021. doi: 10.4060/cb5152en.

[2] R. Sabaragamuwa, D. Dharmasena, and J. Mannaperuma, 'Optimization of Environmental Parameters for Short-term Storage of Big Onions and Evaluation of the Feasibility of Controlled Environmental Storage', Trop. Agric. Res., vol. 22, no. 4, p. 356, Nov. 2011, doi: 10.4038/tar.v22i4.3786.

[3] M. T. C. P. Thennakoon and D. A. M. D. Silva, 'Market window analysis: a case of tobacco, paddy and big oinon farmers in Galewela, Sri Lanka', Sabaragamuwa Univ J, vol. 11, no. 1, pp. 95–108, Jul. 2013, doi: 10.4038/suslj.v11i1.5890.

[4] Wickramarathne and Chandrasekara, Eds., AN ANALYSIS OF COLOMBO OPEN MARKET AVERAGE RETAIL PRICES OF BIG ONIONS IN SRI LANKA USING TIME SERIES FORECASTING TECHNIQUES. Sri Lanka Economic Research Conference, 2020.

[5] L. N. Ranathunga, W. M. D. I. S. Wijemanna, M. G. S. Sathsara, and R. G. B. K. Gamage, 'Agriculture in Sri Lanka: The Current Snapshot', IJEAB, vol. 3, no. 1, pp. 118–125, 2018, doi: 10.22161/ijeab/3.1.15.

[6] N. P. G. Samantha, R. Vidanapathirana, and R. Rambukwella, 'Issues in Big Onion Seed Production and Marketing'.