The Journal of Agricultural Sciences - Sri Lanka
Vol. 20, No 2, May 2025. 280-292
http://doi.org/10.4038/jas.v20i2.9901

Crop Prediction Based on Environment Variables using Data Mining Technologies

K.T. Dananjali[1]*J.B. Ekanayake[2], B.T.G.S. Kumara[3], and A.S. Karunarathne[4]

Received: 22nd August 2022/Accepted: 21st April 2025

ABSTRACT

Purpose: *Sustainable agriculture is essential for addressing food security challenges and enhancing the socio-economic status of farmers. Integrating modern technologies with the agricultural sector is a key solution to overcome many issues. Therefore, this study aimed to apply data mining technologies to identify the most suitable crop types for specific land based on factors: weather conditions (rainfall, temperature, humidity), crop prices, and soil conditions of lands. Potato, tomato, green gram, and red onions are crop types. The soil conditions are determined based on the locations specified as Grama Niladari Divisions of the Badulla District.*

Research Method: *The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was applied to develop the recommended system. The MYSQL database, WEKA libraries (is it "and WEKA libraries" – please check with the researcher), which integrated the Java programming language, were used to implement the prediction models. The prediction quality of models was evaluated using different evaluation metrics.*

Findings: *The Random Forest tree and Random tree outperformed the other models in predicting the weather conditions for the next four months from the system usage date in a weekly manner and the variability in crop prices, respectively. Once the location is selected, the suitable crops are ranked based on the prevailing soil conditions in that location. Then, the most suitable crops were selected based on the predicted weather factors, which best fit with the ranked crops, together with the predicted prices.*

Value: *The system ranks suitable crops for specific lands. These findings assist farmers and decision-makers in transforming agriculture into a profitable, economically stable industry.*

Keywords: *Agriculture, Crop selection, Data mining, Random forest, Random tree, Weka*

INTRODUCTION

Agriculture is the main income of the rural community in Sri Lanka. However, there is a trend in the rural community to move away from agriculture, citing it as a less profitable industry. This declined trend further confirmed that the agricultural sector contributed 41% to the GDP in 1950, whereas it dropped to 7% in the first quarter of 2018 ("Sri Lanka GDP From Agriculture - 2021 Data - 2022 Forecast - 2010-2020 Historical," n.d.). In agriculture, selecting the most suitable crop for the next season is a crucial decision made by the farmers as various factors influence this decision.

[1]Faculty of Graduate Studies, Sabaragamuwa University of Sri Lanka, P.O. Box 02, Belihuloya, Sri Lanka, 72000, dananjali@ccs.sab.ac.lk

[2]Department of Computer Science & Informatics, Faculty of Applied Sciences, Uva Wellassa University of Sri Lanka, Passara Road, Badulla, Sri Lanka, 90000

[3]Department of Data Science, Faculty of Computing, Sabaragamuwa University of Sri Lanka, P.O. Box 02, Belihuloya, Sri Lanka, 72000

[4]Department of Export Agriculture, Faculty of Agricultural Sciences, Sabaragamuwa University of SAri lanka, Belihuloya, sri Lanka

*dananjali@ccs.sab.ac.lk

https://orcid.org/0000-0002-9559-5515

However, the perfect selection of crop types may cause one to gain more income from farming. Therefore, identifying the factors affecting crop production and selecting the most appropriate crop type for specific land is crucial for making informed agricultural decisions and ensuring profitability and economic stability in the industry.

Researchers and scientists have identified many factors affecting agricultural crop production, as well as crop prices (Chetan *et al*., 2021). According to (Changalima *et al*., 2020; N'Souvi *et al*., 2021) economic factors such as subsidies, commodity prices, labor costs, and transportation costs affect crop production. Further, climatic factors such as light, rainfall, temperature, relative humidity, and wind, some physical factors such as soil conditions, technological effects, farming knowledge, political factors, and some social factors such as ownership are identified as affecting factors for farming (Chandio *et al*., 2020)such as CO_2 emissions, temperature, rainfall, and cereal yield in Turkey from 1968 to 2014. At first step, we tested stationary properties of the climatic factors and crop yield by using both traditional and breakpoint unit root tests. After the confirmation of given properties, we used the autoregressive distributed lag (ARDL).

Food and Agriculture Organization of the United Nations (FAO, n.d.), identified natural resources (soil, water, temperature), traditional practices, government policies, international trade agreements, public opinions, and environmental fluctuations affecting the farming system. Further, they have divided affecting factors into two groups: fixed factors and variable factors. As fixed factors, they identified accessibility of the land, labour availability, season's rainfall, and irrigation possibilities. Credit and cost, infrastructure facilities, seed's quality and availability, fertilizer usage, pest challenges, diseases, natural disasters, weed challenges, population, and market control measures were identified as variable factors.

According to (Liliane and Mutengwa, 2020), they categorized affecting factors into three groups. Those are technological, biological, and environmental: technological – agricultural practices, managerial decisions etc.; biological – diseases, insects, pests, weeds and as environmental – climatic condition, soil fertility, topography and water quality. Further, they discussed the importance of climate-smart agriculture to increase the crop yields' productions. Supply-demand relationship, promotion of production cost, circulation cost, capital cost were identified as leading factors that increase crop prices. Inefficient information about supply-demand, lack of risk management tools, decentralized and small scaled operations were identified as affecting factors that influence decreasing crop prices (Ai-hua, 2012).

In this study, factors affecting the crop selection were divided into two categories: farmer independent variable and farmer dependent variables. The factors, unable to be controlled by farmers, were named as farmer independent; and others, categorized as farmer dependent variables. Among them, farmer independent variables are only considered in this study.

Environmental factors such as weather and soil conditions together with expected prices were identified as highly influential on farmer independent variable in the crop selection. However, weather and the price are rapidly changing factors; hence, the selection of crops becomes more complicated for the farmers.

Furthermore, agricultural data are diversified and complex, and contain many interrelated attributes in large volumes. Extracting useful information and decision-making from this large volume of data is difficult and not a trivial task. Therefore, various information technological applications are developed in the agricultural sector to overcome day-to-day challenges in the field (Rupnik *et al*., 2019; Elbasi *et al*., 2023).

Decision support applications are one of the most popular information and communication technology (ICT) based solutions. Because farmers and decision-makers are very curious about their decisions. Therefore, designing ICT based decision support system was an essential requirement in many stages of farming.

Before developing the agriculture-based system, three main issues should be considered: socioeconomic context, natural resources, and climate since the success of farming is dependent upon these conditions (Yost *et al*., 2011).

In (Ravichandran and Koteeshwari, 2016), the artificial neural network-based android application was developed to identify the most suitable crop types according to some parameters; soil pH value, phosphate, potassium, nitrogen, depth, temperature, and rainfall. Further, they suggested some fertilizers to improve productivity. An artificial intelligent crop selection system was developed based on soil pH, soil type, rainfall, and temperature factors. As an output, they provide a selected crop list to the farmers (Sumathirathna, 2014).

An Android mobile application was developed to predict the most suitable crop types for specific locations. That system required location from the user as an input. Long short-term memory (LSTM), and recurrent neural network (RNN) models were used to predict vegetable and auto regression moving average (ARIMA) models used for price prediction based on the past ten years' data. They provided crop prediction, price prediction, visualization, and optimization and the functionalities. Crop types were predicted based on previous weather data and price data (Selvanayagam *et al*., 2019).

The study proposed a method based on 3D clustering to predict suitable crops. The method determines the crop based on the cluster that the crop is placed in, and recommends the crop according to a critical value (Tseng *et al*., 2019).

Moreover, several applications have been developed in agricultural decision-making processes such as farm management, fertilizer management, crop growth handling, yield management, water resource management, and marketing process (Challinor *et al*., 2003; Rale *et al*., 2019; Yost *et al*., 2011).

In the Sri Lankan context, the involvement of ICT in the agricultural sector is comparatively low; especially for the decision-making process, it is essential to integrate ICT with the agricultural sector to overcome many issues. Few existing agriculture-based decision support systems are focused on the Sri Lankan agricultural sector. Few of them were focused on the crop selection process. Therefore, we identified it is essential to develop modern technology-based systems to support the crop selection process.

The system forecasts the most suitable crop type for a given area based on the weather factors, soil conditions, and the price of crops. Most of the models developed in the literature use the present climatic conditions to predict suitable crops for the next season. Even though, in this model, it first predicts the environmental variables – rainfall, temperature, and humidity – for the next four months weekly. The required water level for a crop is not constant; hence, weekly prediction is very helpful to model the growth of the crop. Further, the price of crops is also predicted for the next four months. Finally, the suitable crops are ranked based on the predicted environmental factors as well as the crop price. The results show that our approach provides decent accuracy and can be used as a decision-making tool in the agriculture domain.

MATERIALS AND METHODS

Different methodologies can be used to implement data mining projects. This project uses Cross Industry Standard Process for Data Mining (CRISP-DM) methodology as it is a standard methodology ("CRISP-DM," n.d.). The CRISP-DM methodology together with the agile system development life cycle is used to develop the complete system. Figure 01 shows the CRISP-DM methodology followed by this study.

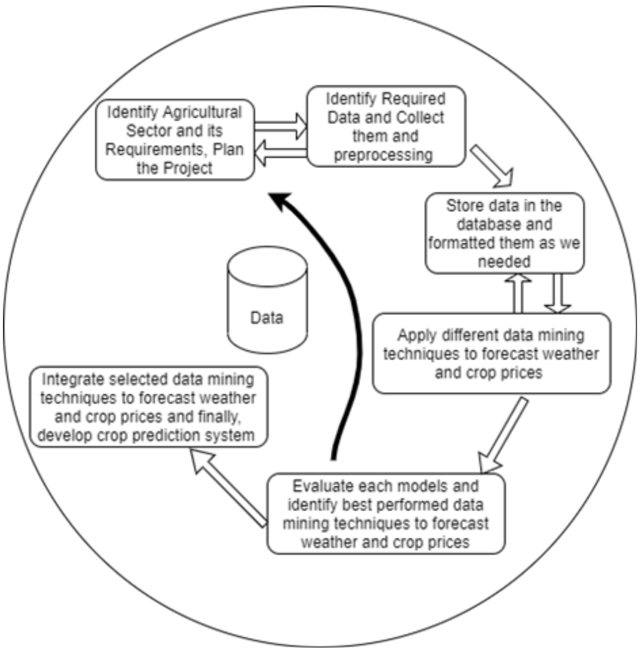

Figure 01: Project development process

Business Understanding

As the first step of the Figure 01, the focused business area of the study is agricultural field and target group is farmers together with agriculture-related decision-makers. The location of this study is Badulla district, Sri Lanka as it is an intermediate climatic zone, which is favorable for verity of crops (“Agro Ecological Regions – Clearing House Mechanism – Sri Lanka,” n.d.; “Badulla (District, Sri Lanka) - Population Statistics, Charts, Map and Location,” n.d.; “The Urban Development Authority,” n.d.).

Data Understanding

This study considers predicting four seasonal crop types, which are commonly grown in the Badulla district as well as often consumed by the people in Sri Lanka. As mentioned in the literature, the weather conditions and crop prices significantly influence for possible crop selection for the next season. Hence, we selected rainfall, temperature, relative humidity and soil conditions together with wholesale and retail prices of each crop type as factors for crop selection. The suitable weather and soil conditions are defined based on the data provided by the Department of Agriculture. The weather data for the past 15 years (2002-2017) were collected from the meteorological department of Sri Lanka. The Hector Kobbakaduwa Agrarian Research and Training Institute in Sri Lanka provided the wholesale and retail prices of crops recorded from large economic centres in Sri Lanka (“Coconut Research Institute | Multi-Disciplinary Research Institute in the World,” n.d.).

Data Preparation

The collected datasets were preprocessed as the datasets contained missing values and outliers. We used the techniques implemented in WEKA (weka-dev-3.9.3.) data mining tool for outlier and missing values (Hall *et al*., 2009). Typically, the missing values and outliers are replaced using the mean value. The weather factors are presented as rainfall (R), maximum temperature (MaxT), minimum temperature (MinT), maximum relative humidity (MaxRH) and minimum relative humidity (MinRH). The price of the crops is represented using the wholesale and retail prices. Each factor is considered an individual attribute, and hence stored in separate files.

Modeling

Modeling involves four phases--select the appropriate model, prepare datasets, train the models and finally test it (“CRISP-DM,” n.d.). Selecting the most efficient as well as appropriate prediction models for weather and crop prices forecasting is very important. To that end we selected a range of models; classification, clustering, regression and time series, and trained them to find the best model. Finally, we realized that the time series models best fit the given datasets.

This study trained Linear regression, M5P model tree, multilayer perceptron model, random forest tree, random tree, REP tree, SMO regression and ZeroR to time series model building for the environmental factors and the price of each crop prediction. The above models were trained on approximately 66% of the datasets, and the rest is used for testing them. The models were implemented using WEKA (weka-dev-3.9.3.) data mining tool (Hall *et al*., 2009).

Evaluation

Each weather factor and crop prices were predicted as the weekly data for another consecutive four (04) months’ time period. Each model was trained and tested in three different testing points with different evaluation matrix – mean absolute error (MAE), root relative squired error (RRSE), direction accuracy (DA), relative absolute error (RAE), mean absolute percentage error (MAPE), root mean squired error (RMSE). The performances of both training and testing data sets were considered to evaluate the models. Since the prediction is four months’ ahead, there are sixteen (16) weekly predicted values. Therefore, the average of each error values was used to evaluate the models.

After evaluating the performances of each model using different evaluation matrix, the emphasis is to identify the relationship between actual values and predicted values using Pearson correlation coefficient values (r). Following two hypotheses are defined:

H_0 = There is no correlation between actual and predicted values
H_1 = There is correlation between actual and predicted values

H_0 is rejected when P-Value is less than 0.05. Thereafter, the residual analysis is conducted to check the suitability of each model. To that end, a run chart was used to check the randomness of the residuals.

Two hypotheses are defined for that purpose:
H_0 = Error values performed by model are random
H_1 = Error values performed by model are not random and there is a pattern.

H_0, the null hypothesis, is rejected when the P-Value is less than 0.05. The hypothesis is tested using the approximate P value for clustering, trends, mixtures and oscillation.

Deployment

The main objective of the study is crop prediction system that allows farmers or decision-makers to predict most suitable crop types for their specific locations in the Badulla district within specific time periods. Figurer 02 shows the flow of the crop prediction system.

Case diagram shown in Fig.03, which shows the activities of each actor in the system. "Input location" is the only required activity from the user. Later, system will process the whole flow of system and finally predict a suitable crop type and provide it to user as output. From time to time updating of the crop prices and weather data with real data is a required task from the system administrator.

-Rainfall disparity = Average of the predicted rainfall values – Required rainfall by crop
-Minimum temperature disparity = Average of predicted minimum temperature values – Required minimum temperature by crop
-Maximum temperature disparity = Average of predicted maximum temperature values – Required maximum temperature by crop
-Minimum relative humidity disparity = Average of predicted minimum relative humidity values – Required minimum relative humidity by crop
-Maximum relative humidity disparity = Average of predicted maximum relative humidity values – Required maximum relative humidity by crop

Then, the price disparity was evaluated as following:
-Price disparity = Average value of each of the predicted crop prices – the first value of the predicted crop price list

The results of the above equation will shows whether the crop price may increase or decrease in its harvesting time.
-Total price disparity = (Wholesale price disparity + Retail price disparity)
-Total weather disparity = (Rainfall disparity + Minimum temperature disparity + Maximum temperature disparity + Minimum relative humidity disparity + Maximum relative humidity

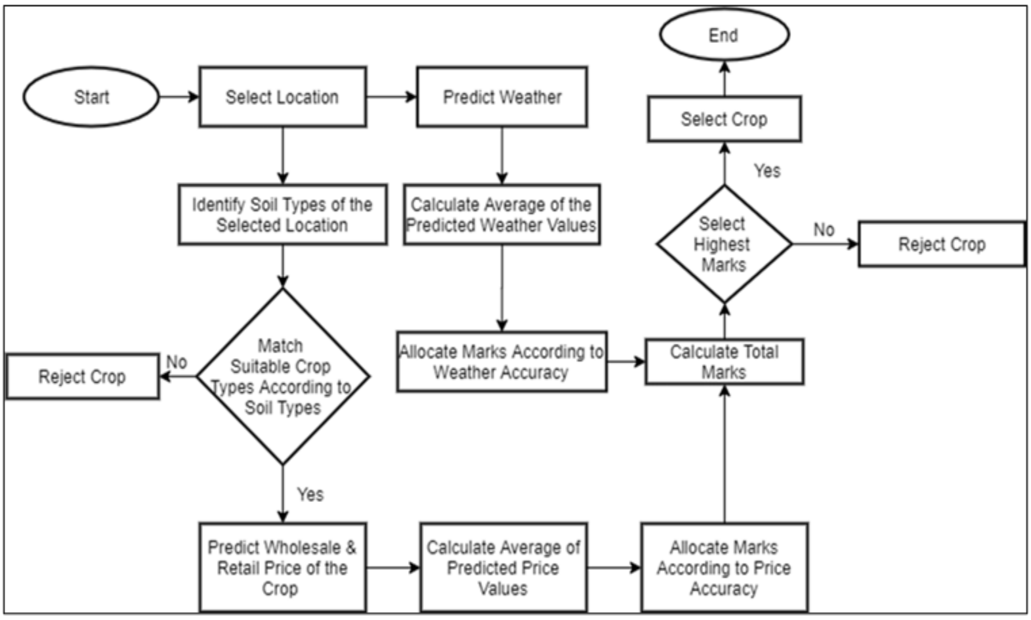

Figure 02: Flow chart of crop prediction system

The system first sorts out the crop list according to their soil conditions, after which it calculates the total weather accuracy by comparing the average of the predicted value and expected values by crops.

disparity)

Following conditions shows the way of crop ranking in the system.

-If Total weather disparity > 25, Total weather accuracy = 0

-If Total weather disparity < 25, Total weather accuracy = 100 - Total weather disparity

-If Total price disparity<0,Total price accuracy = 0

-If Total price disparity > 0, Total price accuracy = Total price different

Then, calculate the total marks obtained by each crop types

-Total marks = Total weather accuracy + Total price accuracy

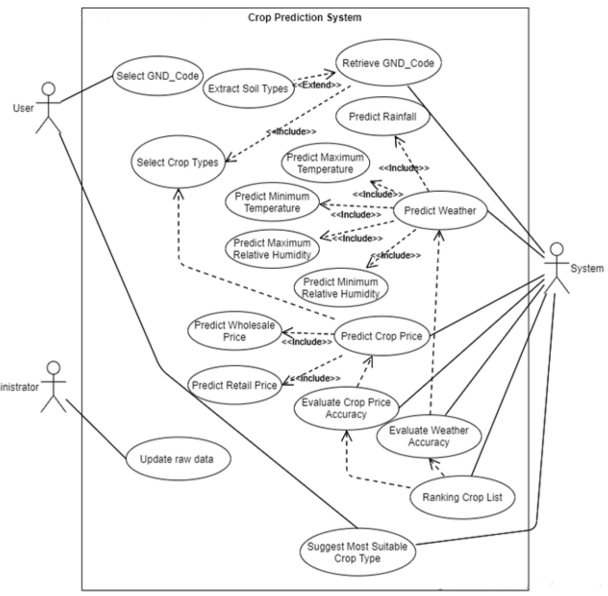

Figure 03: Use case diagram for crop prediction system

Finally, the total marks of each selected crop types are calculated and ranked accordingly. Thereafter, return to the higher-ranked crop name as output.

RESULTS AND DISCUSSIONS

Table 01, Table 02, Table 03, Table 04, and Table 05 show the average error values performed by each weather forecasting model.

Among these models, random forest tree models performed the lowest error values for each weather factor. Thereafter, Pearson correlation coefficient values between predicted values and actual values are generated with P-values to identify the relationships between the actual and predicted values.

The results obtained from the statistical analysis are shown in Table 06. The P-values prove that there is a relationship between actual values and predicted values for each weather factor except for minimum temperature. Furthermore, the results of the Pearson Correlation Coefficient Values confirmed that there is a strong relationship between the actual value and predicted values of the maximum temperature, maximum relative humidity, and minimum relative humidity forecasting models, while the rainfall forecasting model performed a moderate relationship. Therefore, H_0 is rejected. Even though the minimum temperature forecasting model shows a poor relationship between the actual value and the predicted value. Afterward, the residual analysis is developed for each weather forecasting model to identify the randomness. Table 07 shows the approximate P-values performed by

Table 01: Average error values performed by rainfall forecasting models.

Data Mining Technique	Evaluation Matrix					
	MAE	RRSE	DA	RAE	MAPE	RMSE
Linear regression	6.3	98.4	53.9	105.3	1094.6	8.6
M5P model tree	5.1	79.3	49.9	85.9	1172.5	6.9
multilayer perceptron model	18.5	326.9	47.3	307.9	3681.0	28.7
random forest tree	5.0	73.4	47.6	85.4	1259.1	6.4
random tree	11.7	196.2	43.2	194.4	2945.9	17.3
REP tree	4.9	76.0	18.9	82.3	825.7	6.6
SMO regression	4.6	81.1	48.4	77.9	597.1	7.1
ZeroR	4.9	75.5	6.2	82.9	1236.2	6.6

run charts related to random forest tree weather forecasting models.

The P-values of each predicted model are greater than 0.05 except the minimum temperature forecasting model. Therefore, H_0 is not rejected.

Table 02: Average error values performed by maximum temperature forecasting models.

Data Mining Technique	Evaluation Matrix					
	MAE	RRSE	DA	RAE	MAPE	RMSE
Linear regression	2.8	148.2	56.0	157.0	10.1	3.4
M5P model tree	1.0	57.4	55.8	57.2	3.5	1.3
multilayer perceptron model	1.7	95.9	55.4	96.7	6.1	2.1
random forest tree	1.0	58.4	55.1	58.8	3.6	1.3
random tree	1.5	90.6	39.5	87.3	5.5	2.0
REP tree	1.1	64.6	18.6	63.6	3.9	1.4
SMO regression	1.6	86.3	56.3	88.9	5.6	1.9
ZeroR	2.0	111.3	0.0	113.5	7.2	2.4

Table 03: Average error values performed by minimum temperature forecasting models.

Data Mining Technique	Evaluation Matrix					
	MAE	RRSE	DA	RAE	MAPE	RMSE
Linear regression	0.9	76.6	45.8	76.6	4.9	1.2
M5P model tree	0.9	71.5	51.2	75.6	4.7	1.1
multilayer perceptron model	0.8	69.7	46.7	72.6	4.5	1.0
random forest tree	0.9	74.5	48.0	78.4	4.9	1.1
random tree	1.1	93.0	34.2	101.0	6.3	1.4
REP tree	0.8	73.5	18.4	70.5	4.5	1.1
SMO regression	1.4	114.5	49.1	121.7	7.9	1.8
ZeroR	0.9	80.7	0.4	76.3	4.9	1.2

Table 04: Average error values performed by maximum relative humidity forecasting model.

Data Mining Technique	Evaluation Matrix					
	MAE	RRSE	DA	RAE	MAPE	RMSE
Linear regression	3.5	97.4	30.9	103.4	4.3	6.0
M5P model tree	3.5	89.3	33.2	104.5	4.2	5.5
multilayer perceptron model	5.0	102.6	29.5	145.6	5.7	6.3
random forest tree	3.7	98.2	33.9	112.4	4.5	5.9
random tree	6.4	142.9	44.4	195.8	7.6	8.6
REP tree	3.5	99.5	41.2	106.5	4.3	5.9
SMO regression	3.2	83.0	30.8	93.7	3.8	5.1
ZeroR	3.3	93.4	38.5	100.7	4.0	5.5

Table 05: Average error values performed by minimum relative humidity forecasting models.

Data Mining Technique	Evaluation Matrix					
	MAE	RRSE	DA	RAE	MAPE	RMSE
Linear regression	6.4	76.4	50.2	75.4	9.1	8.2
M5P model tree	6.0	70.5	52.6	71.1	8.6	7.6
multilayer perceptron model	7.5	85.5	50.7	87.8	10.7	9.2
random forest tree	6.1	69.8	52.1	71.7	8.4	7.5
random tree	9.6	110.3	40.0	113.3	13.4	11.8
REP tree	6.2	69.9	19.4	72.8	8.7	7.5
SMO regression	6.7	78.6	51.0	79.3	9.5	8.4
ZeroR	7.3	84.9	2.9	86.4	10.3	9.1

Table 06: Pearson correlation coefficient values and p-values related to random forest tree weather forecasting models.

Weather Factor	Evaluation Matrix	
	Pearson Correlation Coefficient Values (r value)	P-Values
Rainfall	0.534	0.033
Maximum temperature	0.902	0
Minimum temperature	-0.346	0.189
Maximum Relative Humidity	0.892	0
Minimum Relative Humidity	0.789	0

Table 07: Approximate p-values performed by run chats related to random forest tree weather forecasting models.

Weather Factor	Evaluation Matrix			
	Approximate P-value for Clustering	Approximate P-value for Trends	Approximate P-value for Mixtures	Approximate P-value for Oscillation
Rainfall	0.698	0.417	0.302	0.583
Maximum temperature	0.06	0.201	0.94	0.799
Minimum temperature	0.005	0.003	0.995	0.997
Maximum Relative Humidity	0.302	0.417		0.583
Minimum Relative Humidity	0.698			0.583

The models are random while there are no clustering, trends, mixtures, or oscillation effects. According to these findings, the models generated using random forest tree algorithms were selected as the best models to forecast weather factors in the system.

Similarly, weather forecasting, data mining techniques were applied to identify the most suitable forecasting models for crop prices as wholesale price and retail price.

Among those models, the random tree performed the highest accuracy in wholesale price forecasting for potatoes, tomatoes, and red onions. Figure 04 shows the variation between the predicted results and actual results. Thereafter, Pearson correlation coefficient values between predicted values and

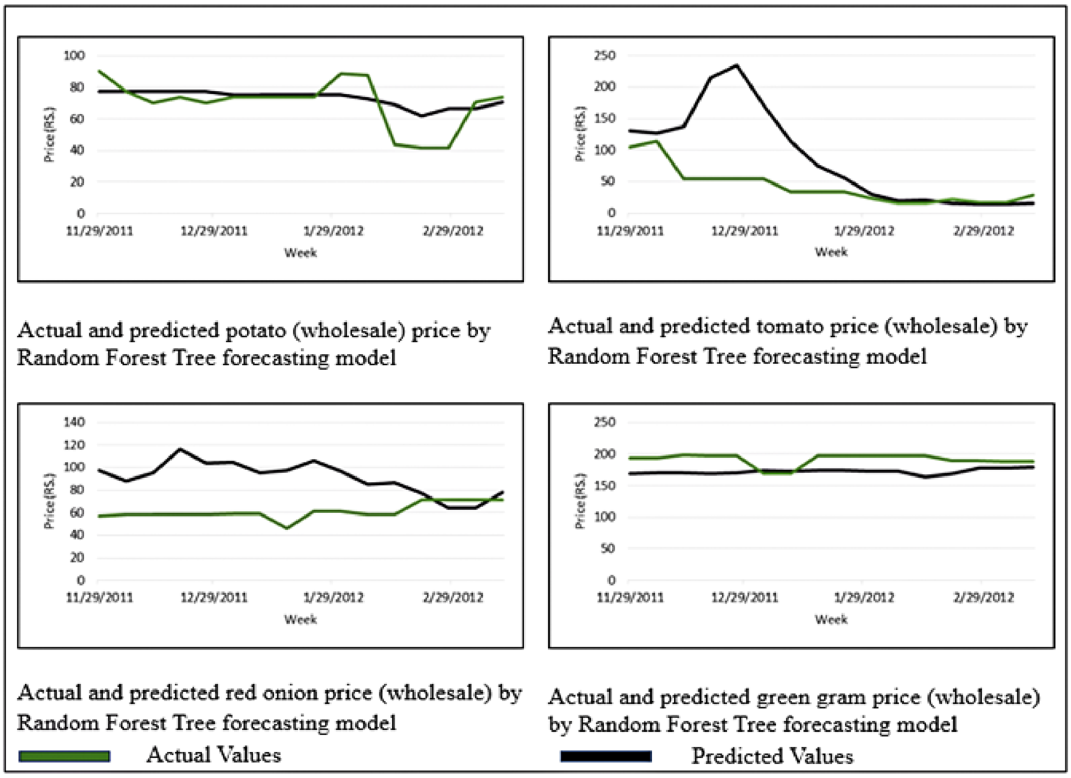

Figure 04: Actual and predicted value variations for crop wholesale price

actual values are generated with P-values to identify relationships between the actual and predicted values.

According to the P-values of the potato, tomato, and green gram, H_0 is rejected. There is a relationship between actual values and predicted values. Further, Pearson Correlation Coefficient Values prove that there are strong relationships.

Table 09 presents the Approximate P-values generated by each crop price forecasting model. According to the results, the P-values of most predicted models are greater than 0.05. Therefore, H_0 is not rejected. According to these findings, the

Table 08: Pearson correlation coefficient values and p-values related to random forest

Crop Type	Evaluation Matrix	
	Pearson Correlation Coefficient Values (r value)	P-Values
Potato	0.946	0
Tomato	0.896	0
Red Onion	-0.081	0.765
Green Gram	0.663	0.005

Table 09: Approximate p-values performed by run chats related to random forest wholesale price forecasting models.wholesale price forecasting models.

Crop Type	**Evaluation Matrix**			
	Approximate P-value for Clustering	**Approximate P-value for Trends**	**Approxi-mate P-value for Mixtures**	**Approximate P-value for Oscillation**
Potato	0.026	0.201	0.974	0.799
Tomato	0.417	0.995	0.583	0.005
Red Onion	0.41	0.995	0.583	0.005
Green Gram	0.06	0.018	0.94	0.982

models generated using a random forest algorithm were selected as the best models to forecast wholesale crop prices.

According to the results of the retail price forecasting models, each model performed comparatively poor prediction quality. Ultimately, the random tree data mining algorithm was selected as the best data mining technology to forecast retail prices of each four crop types since it performed comparatively better prediction quality. Figure 05 presents a variation of the actual and predicted values of the crops' retail prices.

To identify the relationship between actual values and predicted values, Pearson correlation coefficient values, and P-values. The results are presented in Table 10. According to the results, H_0 is rejected. There are strong relationships according to the Pearson Correlation Coefficient Values.

Table 11 presents the Approximate P-values generated by each crop price forecasting model. The results show that the P-values of most predicted models are greater than 0.05. Therefore, H0 is not rejected and error values performed by models are random. Based on these findings, the models generated using a random forest algorithm were selected as the best to forecast retail crop prices. Finally, each model was implemented and integrated with the crop prediction system. The system will be required to select the Grama Niladari Division name of the relevant location. Then, the system will match the suitable crop types for the location's soil conditions and select suitable crop types. The system will automatically detect the prediction start date based on the system usage date, after which, the crop prices and weather conditions are predicted. After analyzing all the results, the system will provide the most suitable crop type for the selected land.

CONCLUSIONS

As the specific objectives of this study, different data mining technologies were applied to forecast time series weather data to identify the best time series forecasting models for weather factor forecasting. Among the selected data mining techniques, the random forest tree model performed the best prediction quality for weather factor forecasting.

Thereafter, different data mining technologies were applied to forecast time series crop prices for wholesale and retail. According to the results generated by selected data mining techniques, the random tree model performed the best prediction quality for crop price forecasting. Finally, the crop prediction system was developed, and it

Figure 05: Actual and predicted value variations for crop retail price

successfully provides the most suitable crop type for specific land in Badulla district. That will be a great support for the decision makers as well as farmers in the Badulla area.

The importance of integrating information technology with the agricultural sector has been discussed and further, the existing ICT applications as well as the crop predicting systems have been focused on in the research. The findings may be beneficial for future studies. Furthermore, the implemented system may be highly useful for farmers and decision-makers to make decisions regarding crop selection. Therefore, it is evident that this system may be beneficial for farmers to change agriculture as a profitable industry.

For future development, it is better to develop this system as a mobile application with chatbot facilities and different languages. Further, it is better to consider soil pH value and slope of the land as other influencing factors for crop selections. It will be more useful to develop this system for the whole country while including many other crop types and environmental variables. Finally, the maximum contribution to integrate information technology with the agricultural sector and support is provided to overcome many issues in agriculture while supporting to change agricultural sector into a profitable industry.

Authors' statement of any conflicts of interest and acknowledgment of sources of funding
The work/publication is supported by a Research grant 2016, Sabaragamuwa University of Sri Lanka.

Table 10: Pearson correlation coefficient values and p-values related to random forest retail price forecasting models.

Crop Type	Evaluation Matrix	
	Pearson Correlation Coefficient Values (r value)	P-Values
Potato	0.956	0
Tomato	0.899	0
Red Onion	-0.093	0.734
Green Gram	0.660	0.005

Table 11: Approximate p-values performed by run chats related to random forest retail price forecasting models.

Crop Type	Evaluation Matrix			
	Approximate P-value for Clustering	Approximate P-value for Trends	Approximate P-value for Mixtures	Approximate P-value for Oscillation
Potato	0.162	0.018	0.838	0.982
Tomato	0.005	0.018	0.995	0.982
Red Onion	0.005	0.201	0.995	0.799
Green Gram	0.019	0.018	0.981	0.982

REFERENCES

Agro Ecological Regions – Clearing House Mechanism – Sri Lanka, http://lk.chm-cbd.net/?page_id=176. 15.03.2022.

Ai-hua, T. (2012). Factors Influencing Price of Agricultural Products and Stability Countermeasures, Asian Agricultural Research. 04(04), 17–43. DOI: https://doi.org/10.22004/ag.econ.137216.

Badulla (District, Sri Lanka) - Population Statistics, Charts, Map and Location. https://www.citypopulation.de/en/srilanka/prov/admin/uva/81__badulla/ 31.03.2022.

Challinor, A.J., Slingo, J.M., Wheeler, T.R., Craufurd, P.Q., and Grimes, D.I.F., (2003). Toward a Combined Seasonal Weather and Crop Productivity Forecasting System: Determination of the Working Spatial Scale. *Journal of Applied Meteorology and Climatology* 42, 175–192. https://doi.org/10.1175/1520-0450(2003)042<0175:TACSWA>2.0.CO;2

Chandio, A.A., Ozturk, I., Akram, W., Ahmad, F., and Mirani, A.A., (2020). Empirical analysis of climate change factors affecting cereal yield: evidence from Turkey. *Environmental Science and Pollution Research,* , 11, 11944-11957 Res 27. https://doi.org/10.1007/s11356-020-07739-y

Changalima, Z.B., Rangsipaht, S., and Nakasathien, S., (2020). The socio-economic and institutional factors affecting farmers' preferences for selection of rice varieties in Kilombero District, Morogoro region, Tanzania. *International Journal of Agricultural Technology* 16, 597–610.

Chetan, R., Ashoka, D.V., Ajay and Prakash, B.V., (2021). Smart Agro-Ecological Zoning for Crop Suggestion and Prediction Using Machine Learning: A Comprehensive Review, in: Chiplunkar, N.N., Fukao, T. (Eds.), *Advances in Artificial Intelligence and Data Engineering. Springer, Singapore*, 1273–1280. https://doi.org/10.1007/978-981-15-3514-7_94

Coconut Research Institute | Multi-Disciplinary Research Institute in the World. https://cri.gov.lk/ 31.03.2022.

CRISP-DM,. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2/ 31.03.2022.

Elbasi, E., Zaki, C., Topcu, A.E., Abdelbaki, W., Zreikat, A.I., Cina, E., Shdefat, A., and Saker, L., (2023). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences* 13. https://doi.org/10.3390/app13169288

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., (2009). The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11, 10–18. https://doi.org/10.1145/1656274.1656278

Food and Agriculture Organization of the United Nations, FAOHome. https://www.fao.org/home/en (accessed 03.30.2022).

Liliane, T., Mutengwa, C., (2020). Factors Affecting Yield of Crops. https://doi.org/10.5772/intechopen.90672

N'Souvi, K., Sun, C., Egbendewe-Mondzozo, A., Tchakah, K.K., Alabi-Doku, B.N., (2021). Analysis of the impacts of socioeconomic factors on hiring an external labor force in tilapia farming in Southern Togo. *Aquaculture and Fisheries* 6, 216–222. https://doi.org/10.1016/j.aaf.2020.07.010

Rale, N., Solanki, R., Bein, D., Andro-Vasko, J., Bein, W., 2019. Prediction of Crop Cultivation, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). pp. 0227–0232. https://doi.org/10.1109/CCWC.2019.8666445

Ravichandran, G., Koteeshwari, R.S., (2016). Agricultural crop predictor and advisor using ANN for smartphones, in: 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). pp. 1–6. https://doi.org/10.1109/ICETETS.2016.7603053.

Rupnik, R., Kukar, M., Vračar, P., Košir, D., Pevec, D., Bosnić, Z., (2019). AgroDSS: A decision support system for agriculture and farming. *Computers and Electronics in Agriculture 161*, 260–271. https://doi.org/10.1016/j.compag.2018.04.001.

Selvanayagam, T., Suganya, S. and Palendrarajah, P. (2019). Agro-Genius: Crop Prediction using Machine Learning, *International Journal of Innovative Science and Research Technology (IJISRT)*, 4(10), pp. 243–249. Retrieved from https://www.ijisrt.com/agrogenius-crop-prediction-using-machine-learning (ISSN) 2456-2165.

Sri Lanka GDP From Agriculture - 2021 Data - 2022 Forecast - 2010-2020 Historical https://tradingeconomics.com/sri-lanka/gdp-from-agriculture. 30.03.2022.

Statistics and Data Processing [SDP] http://www.harti.gov.lk/index.php/en/divisions/statistics-and-data-processing-sdp. 30.03.2022.

Sumathirathna, K., (2014). Intelligent agricultural crop selection system.

The Urban Development Authority, https://www.uda.gov.lk/development-plans.html. 31.03.2022.

Tseng, F. -H., Cho, H. -H., and Wu, H. -T., (2019). Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis, in IEEE Access, vol. 7, pp. 116965-116974. https://doi.org/10.1109/ACCESS.2019.2935564.

Yost, R., Attanandana, T., Colfer, C., Itoga, S., (2011). Decision Support Systems in Agriculture: Some Successes and a Bright Future. https://doi.org/10.5772/16380.