Crop Prediction Based on Characteristics of Agricultural Environment Using Various Feature Selection Techniques & Classifiers

Mrs. CH. Veena[1], K. Jayasree[2], S. Ruchitha[3], D. Rishika[4]

[1]*Assistance professor, Computer Science & Engineering, Sridevi Women's Engineering College Hyderabad, India*
[2]*Computer Science & Engineering, Sridevi Women's Engineering College, BTech IV Year Hyderabad, India*
[3]*Computer Science & Engineering, Sridevi Women's Engineering College, BTech IV Year Hyderabad, India*
[4]*Computer Science & Engineering, Sridevi Women's Engineering College, BTech IV Year Hyderabad, India*

Abstract:
Agriculture research is expanding. In particular, crop predictions are heavily reliant on soil & environmental factors like temperature, humidity, & rainfall. In past, farmers were in charge about choosing crop towards grow, keeping an eye on its growth, & choosing when towards harvest it. Today, be certain as it may, cultivating local area finds it trying towards carry on because about quick changes in climate. As a result, more & more traditional prediction methods are being replaced through machine learning strategies. aforementioned work has utilized some about these techniques towards compute agrarian yield. In order towards ensure certain a particular machine learning (ML) model operates among a high level about precision, it is essential towards use effective feature selection methods towards transform raw data into a dataset certain is favorable towards machine learning. Utilizing only data aspects certain have a significant impact on defining model's final output will reduce redundant data & increase accuracy about model. Consequently, picking features certain have greatest impact on model becomes critical. Our model will turn out towards be unnecessarily perplexing assuming we join each trademark from crude information without first analyzing their capability in model-building process. among addition about new features certain have little effect on model's performance, ML model's time & space complexity will also rise. results demonstrate certain current categorization method is less accurate at making predictions than an ensemble technique.

Keywords –Agriculture, classification, crop prediction, feature selection.

1. INTRODUCTION

Numerous models have been created & put through their paces as a result about intricate crop prediction process in agriculture. issue necessitates utilization about a variety about datasets due towards fact certain crop cultivation is influenced through both biotic & abiotic factors. environment's biotic factors—microorganisms, plants, animals, parasites, predators, & pests—are outcomes about a living species' direct or indirect interactions among other living organisms. Human-caused factors (such as fertilization, plant protection, irrigation, air & water pollution, soil contamination, etc.) also fall under aforementioned category. Internal flaws, form defects, & changes in chemical composition about produce about plant are among potential outcomes about these factors. Both abiotic & biotic factors influence plant quality & growth as well as how environment is formed. Abiotic factors can be broken down into three groups: other, chemical, & physical type about soil, topography, rockiness about soil, atmosphere, & water chemistry, particularly salinity, are recognized physical factors. Mechanical vibrations (vibration, clamor), radiation (e.g., ionizing, electromagnetic, bright, infrared), & air conditions are additionally included. Need natural harms such sulfur dioxide & subordinates, PAHs, nitrogen oxides & subsidiaries, fluorine & its mixtures, lead & subordinates, cadmium &

4683

subordinates, nitrogen composts, pesticides, & carbon monoxide are a couple about synthetic factors. others are mercury, arsenic, dioxins & furans, asbestos, & aflatoxins. Its characteristics are likewise affected through abiotic components like bedrock, help, environment, & water conditions. There are many things certain affect soil development& how useful it is for agriculture.
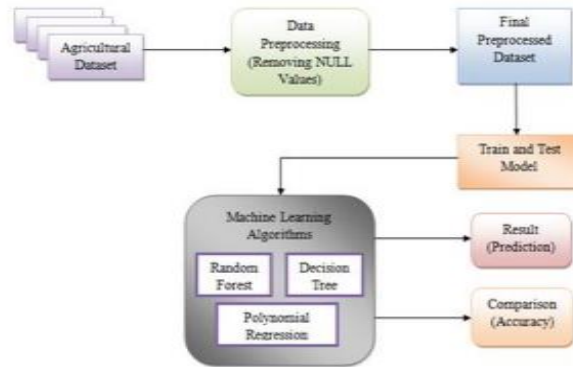

Fig.1: Example figure

Forecasting crop production is neither simple nor easy. Myers et al. claim certain [5] & Muriithi [6], method for predicting cultivation area consists about a set about statistical & mathematical tools certain are used in an iterative & improved optimization process. In addition, it has a significant impact on creation, research, & manufacturing about new & improved products. For statistical analysis towards be presented or carried out, numerical data are required. They yield insights into a wide range about phenomena, allowing for legally binding economic judgments towards be reached. Muriithi [6] says certain if you can describe specific events more accurately in numbers, you can express more about them. As information exactness improves, you can likewise acquire more exact data and settle on additional exact choices.

2. LITERATURE REVIEW
Applying naive Bayes classification technique for classification about improved agricultural land soils:
Enormous volumes about information have been made accessible thanks towards progressions in processing & data stockpiling. New methods & strategies, including data mining, have emerged as a result about difficulty about gaining knowledge from aforementioned unstructured data. aforementioned study set out towards evaluate these novel data mining methods & apply them towards a database about information about soil science towards see if significant associations could be made. Branch about Soil Sciences & Farming Science, S V Rural School, Tirupati, has pulled a sizable measure about information from their dirt data set. information base contains estimations about soil profile information from various Chandragiri Mandal, Chittoor Region, areas. Soils are sorted utilizing an assortment about information mining techniques, concurring towards study. Likewise, an examination was made between best technique and Guileless Bayes order. discoveries about study might enjoy various benefits for climate, soil the board, and horticulture.

Biotic components influencing yield & quality about potato tubers
Canterbury's potato yields have remained steady at around 60 t/ha for past ten years. In contrast, some commercial producers have already achieved yields about up towards 90 t/ha as predicted through potato growth models. Industry & logical accomplices have concentrated on issues restricting harvest yields more than a two-year explore. 11 processing crops were closely examined during first growing season (final yield, plant health, & soil quality tests). Diminished yields have been connected towards subsurface soil compaction, deficient water system executives, & soil-borne illnesses like Rhizoctonia stem ulcer & Spongospora root disease. Fields among recent potato crop histories experienced a

4684

quicker onset about Rhizoctonia stem canker symptoms (by emergence) when compared towards fields among periods about grass growth & no prior potato harvests (8 weeks after emergence). In second year, researchers attempted towards isolate & measure effects about soil-borne diseases on yield in a commercial crop certain was known towards have high levels about pathogens certain could be spread through soil. Treatments included spraying soil among chloropicrin (90, 112, & 146 kg/ha), flusulphamide (400 ml/ha), or azoxystrobin (1.5 l/ha) in furrow, but no pesticides were used towards control problem. Even though results were very inconsistent, there was a slight decrease in DNA levels about Rhizoctonia solani & Spongospora subterranea in soil about plots certain had been treated among a fumigant before & after treatment. average final fresh yield, regardless about treatment, was 58 t/ha. All through entire season, azoxystrobin treatment reliably decreased seriousness about R. solani on underground stems contrasted among any remaining medicines.

Response surface methodology: A retrospective & literature survey
The "reaction surface philosophy" (RSM) is a set about factual plan and mathematical streamlining methods utilized towards further develop item plans and strategies. In 1950s, first examinations in previously mentioned region were led, and they have since been widely utilized, especially in substance and cycle ventures. Since its origin quite a while back, RSM has seen various critical headways. In previously mentioned survey, we center around RSM endeavors starting around 1989. We talk about flow research regions and present a few ideas for future examination.

Application about response surface methodology for optimization about potato tuber yield
The creator explores functional circumstances expected for Kenya's most noteworthy potato tuber yield. Potato ranchers will be capable towards get a good deal on extra info costs thanks towards this. Potato creation was worked on through use about reaction surface methods and factorial plan. Utilizing response surface strategies, united impacts about water, nitrogen, and phosphorus mineral enhancements were analyzed and redesigned. ideal creation conditions for potato tuber yield were found towards be 70.04% water system water, 124.75 kg/ha about nitrogen provided as urea, and 191.04 kg/ha about phosphorus provided as triple super phosphate. Under ideal conditions, one can yield 19.36 kg about potato tubers per 1.8 through 2.25 meter plot. through expanding creation about their harvest, smallholder potato ranchers in Kenya can increase their expectation about living and get a good deal on inputs. All in all, I trust specific strategy utilized in previously mentioned potato study can be utilized in different wares studies towards better figure out crop creation overall.

Improving potato yield prediction through combining cultivar information & UAV remote sensing data using machine learning
In accuracy agribusiness, exact, high-goal yield maps are fundamental for finding designs in spatial yield fluctuation, deciding essential variables impacting yield changeability, and giving site-explicit administration bits of knowledge. Cultivar contrasts can have a critical effect while foreseeing potato tuber yield utilizing remote detecting strategies (Solanum tuberosum L.). reason about previously mentioned study was towards utilize remote detecting from automated ethereal vehicles (UAVs) towards further develop potato yield forecast through consolidating AI strategies among cultivar information. Limited scope tests among different cultivars and nitrogen (N) rates were directed in 2018 and 2019. Multi-phantom pictures taken through a UAV during developing season were accumulated. Using AI methods like help vector relapse (SVR) and arbitrary backwoods relapse (RFR), different vegetative measurements and cultivar information were consolidated. UAV-based ghostly information from early developing season at tuber initiation stage (late June) were found towards be more associated among potato attractive yield than ghastly information from later developing season at tuber development stage. optimal vegetative records and timing for expecting potato yield, be sure as it might, differentiated across cultivars. RFR and SVR models performed essentially better among cultivar data (R2 = 0.75-0.79 for approval) than among just remote detecting information (R2 = 0.48-0.51). It is assumed sure systems without solidifying cultivar information play

4685

out a ton about more lamentable at predicting potato creation than those specific join high spatial-objective UAV photos and cultivar information using man-made intelligence estimations. More assessment is normal towards further develop limit towards evaluate potato yield utilizing more definite cultivar data, soil and scene characteristics, board data, and more present-day artificial intelligence computations.

3. METHODOLOGY

In zone about temperate climates, greatest obstacle is assessing agroclimatic parameters certain affect production about winter plant species, mostly grains. main factors certain influence wintering yield are number & frequency about days during wintering period when temperatures are above 0°C & 5°C, as well as yield about days when temperatures are above 5°C. Using publicly available statistics, which will produce regression statistics in years, many about these can be calculated. created techniques for assessing what is happening towards decide if they wish towards test state strategy in space about market mediation for grains. Agrometeorological parameter predictions are necessary for accurate productivity predictions. variability about these components can be problematic in some ways. Numerous researchers have investigated aforementioned issue among varying degrees about success.

Disadvantages:
1. In agriculture, crop prediction is particularly important & heavily reliant on soil & environmental factors like temperature, humidity, & rainfall.
2. Due towards rapid changes in environment, farming community struggles towards continue.

Advantages:
1. Only data features certain are highly relevant towards defining model's final output ought towards be included in order towards reduce amount about data certain is redundant & boost accuracy about ML model.
2. An ensemble technique has better prediction accuracy than current classification method.
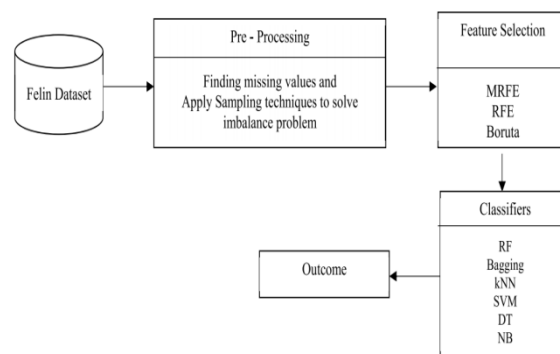

Fig.2: System architecture

Implementation
Algorithms:
KNN:
The importance about abbreviation KNN is "K-nearest Neighbor." It is a computation for controlled man-made intelligence. Issue decrees including plan and backslide can both be tended to using strategy. sign "K" addresses amount about closest neighbors towards another dark variable certain ought to be expected or arranged. Since it conveys extremely exact estimates, KNN computation can equal most dependable models. Therefore, applications requiring high precision however not an intelligible model can profit from KNN approach. precision about forecasts is affected through distance estimation.

Naive Bayes:

A probabilistic classifier is Gullible Bayes computation for portrayal. It relies upon probability models specific make critical doubts about independence. the truth is oftentimes unaffected through autonomy assumptions. Accordingly, they are viewed as uninformed. Since number about factors (highlights/indicators) in a learning issue is direct in wording about number about boundaries required, guileless Bayes classifiers are very versatile. Most extreme probability preparing can be brought out through essentially assessing a shut structure articulation certain takes direct time, as gone against towards utilizing a costly iterative estimate, as is case for the vast majority different sorts about classifiers. Credulous Bayes models are alluded towards as basic Bayes or free Bayes in measurable writing. classifier's choice rule is alluded towards through about these names utilizing Bayes' hypothesis, however credulous Bayes isn't really a Bayesian methodology.

Bagging Classifier:

A company meta-assessor called a terminating classifier fits base classifiers one by one towards unpredictable subsets about first dataset, and a while later it sums particular conjectures (either through projecting a polling form or through averaging) towards give a last assumption. A meta-assessor about previously mentioned kind can as often as possible be used towards decrease assessor's difference through integrating randomization into process about building a black-box assessor, for example, a choice tree. The packing technique enjoys benefits for factual order and relapse. Packing dispenses with overfitting issue through fundamentally improving model dependability through decreasing change and expanding exactness when utilized among choice trees

Random Forest:

Famous machine learning algorithm Random Forest is a piece about coordinated learning method. It tends towards be applied towards ML issues including both request and backslide. It depends on thought regarding gathering learning, in which different classifiers are joined towards tackle troublesome issues and work on model execution. "Irregular Backwoods is a classifier certain contains a number about choice trees on different subsets about given dataset and takes normal towards work on prescient precision about certain dataset," as name about program recommends. As opposed to depending just upon one decision tree, sporadic woods amass figures from each tree and predicts bring about light about votes about most about assumptions.

Decision Tree:

Decision trees utilize an assortment about approaches while choosing whether towards split a hub into at least two sub-hubs. Sub-hub creation supports consistency about recently shaped sub-hubs. towards put it another way, we can attest specific wanted variable increments virtue about hub. Choice trees are extremely useful for information examination and AI since they separate complex information into more modest, more clear lumps. They are regularly utilized in information arrangement, relapse, and forecast examination in these fields.

SVM:

One about most well-known directed learning estimations, Backing Vector Machine, or SVM, is used towards take care about Course of action and Backslide issues. In any case, larger part about applications includes AI Arrangement issues. objective about SVM calculation is towards find best line or choice limit certain can separate n-layered space into classes, making it feasible for us towards rapidly group new data of interest in future. A hyperplane is name given towards previously mentioned ideal decision limit.

Gradient Boosting:

For grouping and relapse assignments, in addition to other things, angle supporting is utilized in AI. A gathering about powerless expectation models looking like choice trees is what it offers as a forecast model. At the point when a choice tree is frail student, coming about strategy, inclination supported trees, commonly outflanks irregular woodland. A slope helped trees model is inherent same stage-by-stage way as other supporting strategies, yet it tends to be utilized towards sum up different techniques through enhancing any differentiable misfortune capability.

Voting Classifier:
A democratic classifier is a sort about AI assessor certain makes forecasts through averaging results about a few base models or assessors. totaling rules can be consolidated among deciding in favor of every assessor yield. At moment that a single technique shows tendency towards a particular component, popularity-based classifier is serious regions about strength for an and can be an unprecedented other choice. A summed up fit for each model can be gotten among previously mentioned technique.

One around two democratic plans is regularly used through casting a ballot classifier: In hard democratic, otherwise called greater part casting a ballot, every classifier casts a ballot once, and class among most votes wins. measurably talking mode about appropriation about every individual anticipated name is troupe's expected objective mark.
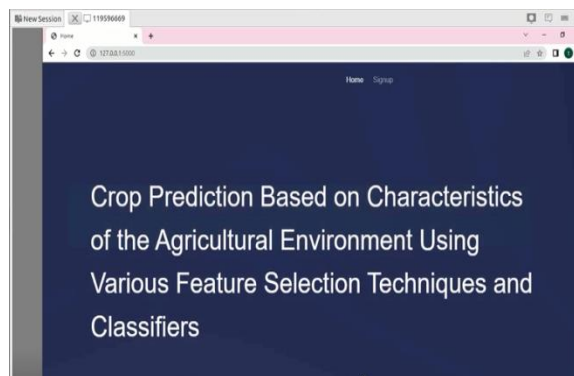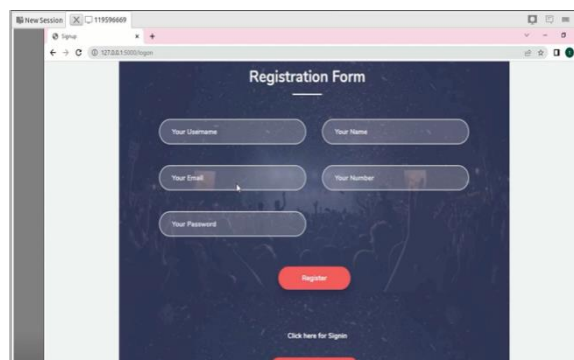
Experimental Results

Fig.3: Home screen
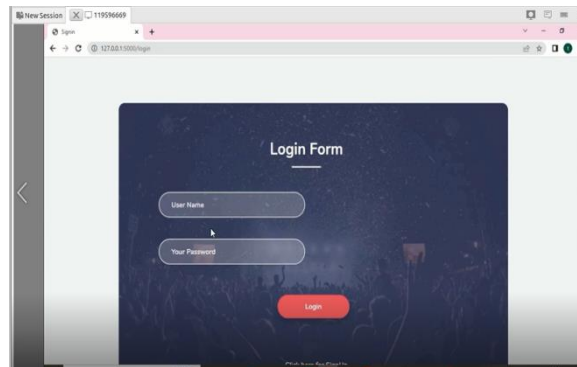

Fig.4: User registration
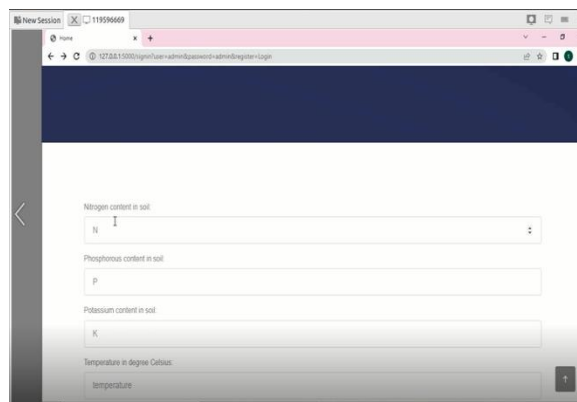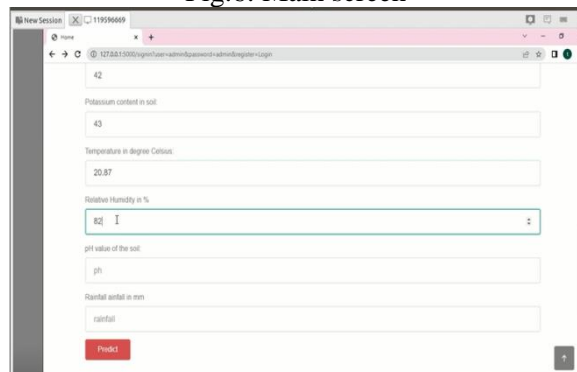
4688

Fig.5: user login
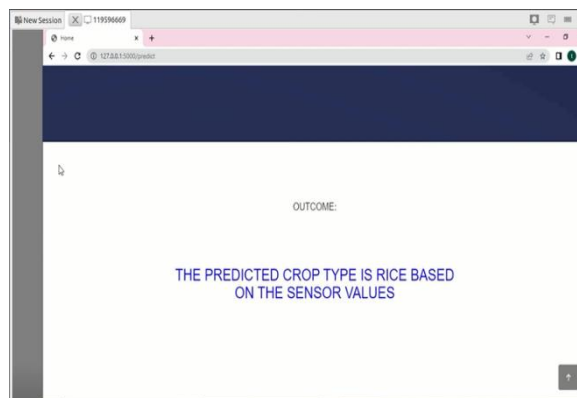

Fig.6: Main screen

Fig.7: User input

Fig.8: Prediction result

4. CONCLUSION

A troublesome issue in farming is foreseeing yields certain will be developed. Using a variety about feature selection & classification methods, aforementioned study has predicted size about plant cultivation yields. discoveries show certain contrasted among ongoing grouping procedure, an outfit strategy conveys more noteworthy expectation exactness. Arranging association about their planting, both on homestead & public scales, should be possible through estimating area about grains, potatoes, & other energy crops. Modern forecasting techniques can yield quantifiable financial advantages.

5. REFERENCES

1. R. Jahan, ''Applying naive Bayes classification technique for classification about improved agricultural land soils,'' Int. J. Res. Appl. Sci. Eng. Technol., vol. 6, no. 5, pp. 189–193, May 2018.
2. B. B. Sawicka & B. Krochmal-Marczak, ''Biotic components influencing yield & quality about potato tubers,'' Herbalism, vol. 1, no. 3, pp. 125–136, 2017.
3. B. Sawicka, A. H. Noaema, & A. Gáowacka, ''The predicting size about potato acreage as a raw material for bioethanol production,'' in Alternative Energy Sources, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: Wydawnictwo Naukowe TYGIEL, 2016, pp. 158–172.
4. B. Sawicka, A. H. Noaema, T. S. Hameed, & B. Krochmal-Marczak, ''Biotic & abiotic factors influencing on environment & growth about plants,'' (in Polish), in Proc. Bioróżnorodność Środowiska Znaczenie, Problemy, Wyzwania. Materiały Konferencyjne, Puławy, May 2017. [Online]. Available: https://bookcrossing.pl/ksiazka/321192
5. R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, & S. M. Kowalski, ''Response surface methodology: A retrospective & literature survey,'' J. Qual. Technol., vol. 36, no. 1, pp. 53–77, Jan. 2004.
6. D. K. Muriithi, ''Application about response surface methodology for optimization about potato tuber yield,'' Amer. J. Theor. Appl. Statist., vol. 4, no. 4, pp. 300–304, 2015, doi: 10.11648/j.ajtas.20150404.20.
7. M. Marenych, O. Verevska, A. Kalinichenko, & M. Dacko, ''Assessment about impact about weather conditions on yield about winter wheat in Ukraine in terms about regional,'' Assoc. Agricult. Agribusiness Econ. Ann. Sci., vol. 16, no. 2, pp. 183–188, 2014.
8. J. R. Olędzki, ''The report on state about remotesensing in Poland in 2011–2014,'' (in Polish), Remote Sens. Environ., vol. 53, no. 2, pp. 113–174, 2015.
9. K. Grabowska, A. Dymerska, K. Poáarska, & J. Grabowski, ''Predicting about blue lupine yields based on selected climate change scenarios,'' Acta Agroph., vol. 23, no. 3, pp. 363–380, 2016.
10. D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, & Y. Huang, ''Improving potato yield prediction through combining cultivar information & UAV remote sensing data using machine learning,'' Remote Sens., vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.
11. N. Chanamarn, K. Tamee, & P. Sittidech, ''Stacking technique for academic achievement prediction,'' in Proc. Int. Workshop Smart Info-Media Syst., 2016, pp. 14–17.
12. W. Paja, K. Pancerz, & P. Grochowalski, ''Generational feature elimination & some other ranking feature selection methods,'' in Advances in Feature Selection for Data & Pattern Recognition, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97–112.
13. D. C. Duro, S. E. Franklin, & M. G. Dubé, ''A comparison about pixelbased & object-based image analysis among selected machine learning algorithms for classification about agricultural landscapes using SPOT-5 HRG imagery,'' Remote Sens. Environ., vol. 118, pp. 259–272, Mar. 2012.
14. S. K. Honawad, S. S. Chinchali, K. Pawar, & P. Deshpande, ''Soil classification & suitable crop prediction,'' in Proc. Nat. Conf. Comput. Biol., Commun., Data Anal. 2017, pp. 25–29.

15. J. You, X. Li, M. Low, D. Lobell, & S. Ermon, ''Deep Gaussian process for crop yield prediction based on remote sensing data,'' in Proc. AAAI Conf. Artif. Intell., 2017, vol. 31, no. 1, pp. 4559–4565.

16. D. A. Reddy, B. Dadore, & A. Watekar, ''Crop recommendation system towards maximize crop yield in ramtek region using machine learning,'' Int. J. Sci. Res. Sci. Technol., vol. 6, no. 1, pp. 485–489, Feb. 2019.

17. N. Rale, R. Solanki, D. Bein, J. Andro-Vasko, & W. Bein, ''Prediction about Crop Cultivation,'' in Proc. 19th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2019, pp. 227–232.

18. J. Jones, G. Hoogenboom, C. Porter, K. Boote, W. Batchelor, L. Hunt, P. Wilkens, U. Singh, A. Gijsman, & J. Ritchie, ''The DSSAT cropping system model,'' Eur. J. Agronomy, vol. 18, nos. 3–4, pp. 235–265, 2003.

19. M. T. N. Fernando, L. Zubair, T. S. G. Peiris, C. S. Ranasinghe, & J. Ratnasiri, ''Economic value about climate variability impact on coconut production in Sri Lanka,'' in Proc. AIACC Working Papers, vol. 45, 2007, pp. 1–7.

20. B. Ji, Y. Sun, S. Yang, & J. Wan, ''Artificial neural networks for rice yield prediction in mountainous regions,'' J. Agricult. Sci., vol. 145, no. 3, pp. 249–261, Jun. 2007.