IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

CROP YIELD PREDICTION IN MACHINE LEARNING MODELS

[1]E.KALAIARASI, [2]Dr.A.ANBARASI

[1]Ph.D(Part Time)Scholar, [2]Assistant Professor
Department of Computer Science, LRG Govt. Arts College for Women, Tiruppur

Abstract:
The fast pace of urban development minimize the agricultural lands. Owing to poor rainfall and drastic climatic changes farmers often face challenges to sustain cultivation of crops with respect to crop cycle. With growing economic competition and rising population, governmental agencies design long term plans which rarely address the farmer's needs. To meet the global demands agriculturist needs to investigate every opportunity that could improve agricultural production and growth. Whether to expand agricultural lands or to improve the production farmers needs to assess the suitability between land and crops. The investigation of land suitability and crop suitability has attracted many researchers to utilize latest technology such as remote sensing, geographical information systems etc. This paper aims to survey on recent researches on crop and land suitability using data mining techniques.

Keywords: crop suitability, land suitability, data mining, classification, agricultural data mining

I. INTRODUCTION

The growing agricultural demand for rising population and depletion of natural resources urge agriculturist and researchers to develop efficient production methods. In India agriculture contributes to about 60% of livelihood with production of 285 million tons of grains. India is the world's largest producer of rice, wheat, spices and cultivates rice, wheat, tea, sugarcane, cotton, tobacco, pulses and vegetables as a staple food. To meet the growing demands the agriculture industry need to address the problem of sustainable production system. The production and yield percentage largely varies due to various reasons such as changing physical, biological and natural conditions. Also the type of agricultural practices affects the growth and yield of crops. And apart from these, other elements such as water, climate, rainfall, soil type etc affect the growth and yield.

Generally suitability analysis plays a major role in understanding the agricultural needs and addressing the problems associated with production. The suitability analysis can be grouped into land suitability and crop suitability or crop-land suitability. Land suitability is the measure of land fitness to specific crops or methods while crop suitability is the measure of crop fitness suitable to grow in a particular land or region specific to temperature, soil type and moisture. Agriculture was influenced either negatively or positively by change in environmental, geographical, climate and political and the changes directly affects the agricultural production. Weather and climatic variations also affect the agriculture production. Prediction of rainfall and changing soil conditions are required to make decisions on selecting the right crop.

Using data mining and machine learning techniques large volume of data from different sources can be effectively used to make accurate decisions on crop, crop management and prediction services to plan or schedule agriculture practices. Also data mining techniques are the best choice to mine information from the weather data, soil data, temperature, climate data etc. Presently data mining are used for Crop Yield prediction, Disease detection, weed detection, crop quality assessment, seed classification, livestock production and management, water management, soil management and weather prediction. Data mining process involve gathering, cleaning, transformation, training, pattern mining and evaluation. Classification, clustering, association, prediction and forecasting are the main techniques employed on agriculture data to mine information.

Fig 1 Data Mining Process

This paper is organized into three sections, section I gives an introduction about agriculture and data mining, section II presents the literature survey on various data mining techniques and methods used for crop, land, crop-land analysis and section III concludes the survey paper.

II LITERATURE SURVEY

Farmers and agriculturist need to utilize vast amount of data to make decisions on factors that affect and improve agriculture process, production and yield. Data mining techniques offer numerous ways to capture data, process data and mine information from large amount of data. Several literatures points out that agriculture data mining offer immense benefit and solve numerous problem related to data mining goals. The field of agriculture data mining is growing tremendously with new algorithms and methods being introduced for specific agriculture related problems such as soil, irrigation, disease, growth, yield and farming decisions.

(Filippi et al., 2019) proposed random forest models combined with STC to predict wheat, barley, canola yield for three different seasons. LOFOCV and LOFYOC cross validation methods are used to measure the prediction quality of the models. The RMSE of the models suggest that more data driven models can be developed using machine learning algorithms to predict crop yield.

(Goapl & Bhargavi, 2019) proposed a hybrid model to predict paddy crop yield. The proposed model is based on MLR and ANN. In the neural network the initial weights are derived from MLR coefficients and fed into the neural network. The BP network is trained on the paddy data and the performance is compared against RF, MLR, ANN, SVR and KNN. ANN-MLR achieved better accuracy than other models.

(Sirsat., 2019) proposed grapevine yield prediction model using yield data, climate data, phonological data, soil analysis, fertilizer and maturation index to prediction grapevine yield. The prediction model is built using RF, LASSO, and Elasticnet. The dimensionality problem is managed through Spikeslab. The prediction models for flowering, coloring and harvesting models show low RMSE rate than other models.

(Feng et al., 2019) investigated four different machine learning models to estimate the soil temperature in China. ELM, GRNN, BPNN and RF models are trained using half-hourly soil temperature features and meteorological variables such as air temperature, wind speed, humidity, solar radiation and vapor pressure. ELM achieved better performance with respect to prediction accuracy and processing time than other models.

(Shastry & Sanjay, 2019) proposed a cloud based framework for soil classification and crop yield prediction. The proposed soil classifier is based on hybrid kernel SVM, where the kernel parameters are derived from GA. The crop yield prediction model is based on ANN where the hidden layers, neurons and learning are customized. The proposed model for soil classification and crop yield model prediction shows better performance than other models.

(Kim et al., 2019) investigated six different machine learning models to predict corn and soybean yield. A DNN model is developed using environmental, metro and hydro data and compared against MARS, SVM, RF, ERT and ANN. The DNN model outperforms other models with good prediction quality and low MAE (0.58) and RMSE (0.76).

(Suchithra & Pai, 2019) proposed a soil nutrient classification model through improved ELM. The performance of ELM is improved through different activation functions such as gaussian radial basis, sine-squared, hyperbolic tangent, tri-angular basis, and hard limit. The accuracy of guassian radial basis function achieve highest rate of 90% over other functions for soil nutrient classification.

(Tamsekar et al., 2019) proposed a crop selection prediction model using GIS data, irrigation data, chemical data, soil data and yield data. The machine learning models such as CART, KNN, RF and SVM are used to build the prediction model. PCA is used to improve the model through dimension reduction. The SVM with PCA show better results with an accuracy of 77%.

(Canizo et al., 2019) compared data mining algorithms to trace the grape regions based on grape-skin composition. MLR, KNN, SVM, RF were models were selected and trained for classification of regions. RF achieved highest accuracy of 89% with 10 fold cross validation and the feature importance score show that rubidium as the best predictor among other elements composed in grape-skin.

(Islam et al., 2018) used deep neural network model to predict four crop yield namely Aus rice, Aman rice, Boro rice, Jute, Wheat and Potato using rainfall data, land types, chemical fertilizers, soil information. The DNN model is compared with RF, SVM and LR and DNN outperforms other model with highest accuracy rate of 98% (Aus rice), 95% (Aman rice), 96% (Boro rice), 97% (Potato), 96% (Wheat) and 94% (Jute).

(Taherei et al., 2018) proposed ELM for prediction of growth of sugarcane. Using seven different parameters that influence sugarcane growth developed an ELM and the model performance is compared against ANN and GA. The ELM model achieved better performance with r^2 of 91% and lower RMSE score than GA and ANN models.

(Oliveira et al., 2018) proposed a pre-season forecasting model using real time soil data and climate to predict Soyabean and Maize yield. The proposed model use DNN with static nodes for off line soil data and dynamic nodes for weather data. The model shows better prediction accuracy for Brazil Soyabean with RMSE of 385.

(Doshi et al., 2018) proposed a crop recommender model for farmers using machine learning models. The prediction model is prepared using ANN and the model performance is compared against DT, KNN, RF. ANN achieved a highest of 91% than other models. The crop suitability is predicted using rainfall, soil type, soil conditions, temperature and geographical location.

(Gümüşçü et al., 2018) proposed a machine learning based prediction model for estimating wheat planting date in advance. The proposed model is constructed using metrological climate data on three classifiers namely KNN, DT and SVM. Feature selection is applied using GA on 1500 features and the models are trained on the selected features. The proposed model KNN with GA achieved a highest accuracy of 92% than other models.

(Feng et al., 2018) investigated the prediction power of machine learning models and regression models. Random forest model using cross validation is compared against multiple linear regression model where random forest achieved better results than MLR. The study also established a relationship between climate and rainfall extremes where the wheat yield percentage is largely affected with low rainfall.

(Sharma et al., 2018) proposed big data based optimized prediction model for crop yield. The proposed work uses GWO and SVM with rbf. The hybrid model shows better accuracy rate with 77% than traditional SVM model. The yield prediction model is developed using environment and soil features.

(Prakash et al., 2018) investigated prediction of soil moisture for better water management using machine learning models. The machine learning models such as SVM and RNN is compared against statistical model MLR. The results suggest that MLR has better prediction power than machine learning models for moisture prediction for short term of 7 days.

(Singh et al., 2017) investigated the rice yield prediction performance of KNN, DT and NB using 11 parameters of micro nutrients and macro nutrients. The prediction accuracy for NB is 98%, DT is 94% and KNN is 97% is achieved. The study concluded that NB achieved better prediction rate and suitable for rice yield prediction using soil parameters.

(Mutalib et al., 2010) performed an empirical study on unsupervised methods for soil classification. Using soil details such as color, texture, drainage, terrain type the data is classified using SOM and K-means. The study findings reveal the SOM has better accuracy (91%) than k-means. The study concluded that unsupervised methods such as SOM can be used for soil classification.

(Rahman et al., 2018) investigated the crop suggestion model based on soil classification using machine learning techniques. The study proposed a SVM based model to suggest crop specific to soil conditions. The proposed SVM model outperforms KNN and bagged trees with 95% of accuracy.

Table 1 Performance of various data mining techniques

References	Method	Accuracy (%)
(Filippi et al., 2019)	RF+STC	-
(Goapl & Bhargavi, 2019)	ANN-MLR	-
(Sirsat., 2019)	RF-FL RF-CO RF-H	-
(Feng et al., 2019)	ELM	-
(Shastry & Sanjay, 2019)	hybrid kernel SVM +GA	-
(Kim et al., 2019)	DNN	MAE-0.58 RMSE-0.76
(Suchithra & Pai, 2019)	ELM+Grbf	90%
(Tamsekar et al., 2019)	SVM+PCA	77%

(Canizo et al., 2019)	RF	89%
(Islam et al., 2018)	DNN	Aus rice- 98% Aman rice-95% Boro rice- 96% Potato-97% Wheat-96% Jute-94%
(Taherei et al., 2018)	ELM	R2-91%
(Oliveira et al., 2018)	DNN	RMSE-385
(Doshi et al., 2018)	ANN	91%
(Gümüsçü et al., 2018)	KNN+GA	92%
(Feng et al., 2018)	MLR	-
(Sharma et al., 2018)	GWO+SVM-rbf	77%
(Prakash et al., 2018)	MLR	MLR- 90%
(Singh et al., 2017)	NB, DT, KNN	NB-98% DT-94% KNN-97%
(Mutalib et al., 2010)	SOM	91%
(Rahman et al., 2018)	SVM	95%
(Shastry et al., 2017)	SVMqK NB KNN SVM	SVMqK-91.5% NB-81.85% KNN-89.45% SVM-90.25%
(Pantazi et al., 2016)	CP-ANNs XY-Fs SKNs	CP-ANNs-78.3% XY-Fs-8.92% SKNs-81.65%
(Kung et al., 2016)	ENN NNBP MLR	ENN error < 2%
(Coopersmith et al., 2014)	Boosted perceptrons	94%
(Johann et al., 2016)	MLP-RBF ANFIS	MLP-RBF-RMSE-1.27 ANFIS-RMSE-1.30

(Shastry et al., 2017) proposed a hybrid model for classification of agricultural datasets. The hybrid model involves a generic classifier that combines different kernels designed for multiclass problems. The proposed method was compared against NB, KNN and SVM on eight different datasets. The proposed method with quadratic kernel achieved better accuracy rate of 91.5% than NB (81.85%), KNN (89.45%), and SVM (90.25%).

(Pantazi et al., 2016) proposed a new data mining method to predict variations in wheat yield. Using soil data and growth characteristics of wheat in a region capture by sensors, the filed variations are predicted using unsupervised neural networks such as CP-ANNs, XY-Fs, and SKNs. The soil data contains H, MC, TN, TC, Mg, Ca, CEC and available P as features. The cross validation performance result shows that SKNs accuracy was higher with 81.65% and outperforms CP-ANNs (78.3%) and XY-Fs (89.2%).

(Kung et al., 2016) proposed an ensemble neural network model for agricultural yield prediction. Using six parameters such as humidity, precipitation, planting area, air temperature, cost of production, trading price, and total harvest, the ensemble NN is trained and compared against NNBP and MLR models. The proposed models show least error rate of less than 2% than NNBP and MLR models.

(Coopersmith et al., 2014) investigated the performance of machine learning algorithms for agricultural decision making. The study involves prediction of soil conditions using hydrological data. Three models such as classification trees, KNN and boosted perceptrons were employed to predict the soil dryness using precipitation and evaporation data. Boosted perceptrons show better accuracy rate of 94% than other models and the study concluded that precipitation and evaporation data are the key parameters to predict soil dryness.

(Johann et al., 2016) proposed an autoregressive error function based NN namely MLP-RBF model and ANFIS model for estimation of soil moisture. The crop yield percentage is influenced by soil moisture and to better predict the crop yield, soil estimation becomes evitable. The proposed models are compared against MLR, the two NN models show r^2 above 80% with RMSE of less than 1.30%

III. CONCLUSION

Machine learning and data mining techniques offer more sophisticated methods to examine large amount of data and extracts useful information. The information extracted offer greater insights into the patterns and address the goals and objectives of data mining. This survey gives an overview of the machine learning models that are widely used in agriculture crop yield prediction. The literatures discussed in this paper focus on crop yield prediction using environmental data, weather data, soil data, temperature, and climate data.

The crop yield prediction serves mainly to improve management, and decision making process. Other than crop Yield prediction, data mining and machine learning technologies can be applied to disease detection, weed detection, crop quality assessment, seed classification, livestock production and management, water management, soil management and weather & rainfall prediction. The application of data mining can be extended to utilize different agricultural data such as GIS, spatial data, image to address different agricultural problems that are yet to be addressed. Developing application using cloud and mobile technologies could serve farmers and other stake holders to better plan, manage and improve overall agriculture production.

REFERENCES

1) Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D., Pozza, L. E., Ugbaje, S. U., ... & Bishop, T. F. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture, 1-15.
2) Islam, T., Chisty, T. A., & Chakrabarty, A. (2018). A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-6). IEEE.
3) Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K. B. W., Alizamir, M., Shamshirband, S., & Chau, K. W. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Engineering Applications of Computational Fluid Mechanics, 12(1), 738-749.
4) Oliveira, I., Cunha, R. L., Silva, B., & Netto, M. A. (2018). A scalable machine learning system for pre-season agriculture yield forecast. arXiv preprint arXiv:1806.09244.
5) Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6).
6) Gümüşçü, A., Tenekeci, M. E., & Bilgili, A. V. (2019). Estimation of wheat planting date using machine learning algorithms based on available climate data. Sustainable Computing: Informatics and Systems.
7) Feng, P., Wang, B., Li Liu, D., Xing, H., Ji, F., Macadam, I., ... & Yu, Q. (2018). Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic change, 147(3-4), 555-569.
8) Sharma, S., Rathee, G., & Saini, H. (2018). Big Data Analytics for Crop Prediction Mode Using Optimization Technique. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 760-764).
9) Prakash, S., Sharma, A., & Sahu, S. S. (2018). Soil Moisture Prediction Using Machine Learning. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1-6).
10) Kim, N., Ha, K. J., Park, N. W., Cho, J., Hong, S., & Lee, Y. W. (2019). A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015. ISPRS International Journal of Geo-Information, 8(5), 240.
11) Singh, V., Sarwar, A., & Sharma, V. (2017). Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. International Journal of Advanced Research in Computer Science, 8(5).
12) Gopal, P. M., & Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. Computers and Electronics in Agriculture, 165, 104968.
13) Sirsat, M. S., Mendes-Moreira, J., Ferreira, C., & Cunha, M. (2019). Machine Learning predictive model of grapevine yield based on agroclimatic patterns. Engineering in Agriculture, Environment and Food.
14) Shastry, K. A., & Sanjay, H. A. (2019). Cloud-Based Agricultural Framework for Soil Classification and Crop Yield Prediction as a Service. In Emerging Research in Computing, Information, Communication and Applications (pp. 685-696)
15) Tamsekar, P., Deshmukh, N., Bhalchandra, P., Kulkarni, G., Hambarde, K., & Husen, S. (2019). Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model. In Computing and Network Sustainability (pp. 309-314)
16) Mutalib, S., Jamian, S. F., Abdul-Rahman, S., & Mohamed, A. (2010, December). Soil classification: An application of self organising map and k-means. In 2010 10th International Conference on Intelligent Systems Design and Applications (pp. 439-444)
17) Rahman, S. A. Z., Mitra, K. C., & Islam, S. M. (2018, December). Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series. In 2018 21st International Conference of Computer and Information Technology (ICCIT) (pp. 1-4)

18) Suchithra, M. S., & Pai, M. L. (2019). Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Information Processing in Agriculture.

19) Feng, Y., Cui, N., Hao, W., Gao, L., & Gong, D. (2019). Estimation of soil temperature from meteorological data using different machine learning models. Geoderma, 338, 67-77.

20) Canizo, B. V., Escudero, L. B., Pellerano, R. G., & Wuilloud, R. G. (2019). Data mining approach based on chemical composition of grape skin for quality evaluation and traceability prediction of grapes. Computers and Electronics in Agriculture, 162, 514-522.

21) Shastry, K. A., Sanjay, H. A., & Deexith, G. (2017). Quadratic-radial-basis-function-kernel for classifying multi-class agricultural datasets with continuous attributes. Applied Soft Computing, 58, 65-74.

22) Pantazi, X.-E.; Moshou, D.; Alexandridis, T.K.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 2016, 121, 57–65.

23) Kung, H. Y., Kuo, T. H., Chen, C. H., & Tsai, P. Y. (2016). Accuracy analysis mechanism for agriculture data using the ensemble neural network method. Sustainability, 8(8), 735.

24) Coopersmith, E.J.; Minsker, B.S.; Wenzel, C.E.; Gilmore, B.J. Machine learning assessments of soil drying for agricultural planning. Comput. Electron. Agric. 2014, 104, 93–104.

25) Johann, A. L., de Araújo, A. G., Delalibera, H. C., & Hirakawa, A. R. (2016). Soil moisture modeling based on stochastic behavior of forces on a no-till chisel opener. Computers and Electronics in Agriculture, 121, 420-428.