International Journal of Engineering & Technology, 7 (3.3) (2018) 107-109

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Survey on Prediction of Suitable Crop Selection for Agriculture Development Using Data Mining Classification Techniques

Angel Prathyusha K [1*], Mahitha Y [2], Prasanna Kumar Reddy N [3], Dr. P Raja Rajeswari [4]

[1]*Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522502*

[2]*Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522502*

[3]*Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522502*

[4]*Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522502*

**Corresponding author Email: angelprathyu678@gmail.com*

Abstract

Agriculture is analytically the vast economic sector and is an important aspect in the economic growth of India. It is the only cause of living for about two-thirds of the population in India. It is very essential for the farmers to choose a crop that best suits the land being used for cultivation. The criteria to be considered in order to decide the crops that best suitable for the land are soil, water and season. The best suitable crop for the land can be predicted based on the agriculture data collected from the agriculture experts or from the farmers. Our paper provides a survey of the various classification techniques and classifiers used for the prediction of suitable crop selection for agriculture development. Farmers should get benefited by cultivating the best fitting crops rather than cultivating the unsuitable crops.

Keywords: *Agriculture, Classification, Classifiers, Prediction, Selection, Suitable crop*

1. Introduction

In India, most of the farmers are not getting the crop yield they expect due to certain reasons. As a result farmers are committing suicides. According to the National Crime Records Bureau of India, 5,650 farmers committed suicides in 2014. [1] Farmer suicides account for 11.2% of all suicides in India. [1] Farmers generally select crops based on their traditional knowledge and past experiences which may go wrong due natural disaster. Monsoon failure and farmers incapable of predicting the crops that best fits their land are some of the reasons. We can prevent suicides of the farmers and help them in selecting the suitable crop by making use of data mining classification algorithms and classifiers.

Single-dimensional classifier methods like Weighted Sum Model (WSM) and Weighted Product Model (WPM) are not suitable for decision making as there are many inputs such as soil, season, input, support, infrastructure and water to be examined. Thus, multi-criteria decision-making methods (MCDM) are used. MCDM is a proper, fixed and transparent decision-making methodology. Its objective is to aid groups or individual decision makers to explore their decisions in the case of complicated situations with multiple criteria.

2. Literature Survey

K-Nearest Neighbours (K-NN) is a primitive classification algorithm and also the lethargic technique among all the classification algorithms. It uses Euclidean, Manhattan, Minkowski and other distance functions to find the K-Nearest neighbour. This algorithm gives best accurate results for K=2. One should select the correct K value to obtain the best suitable crop. The main drawback of this algorithm is that it learns nothing from the training dataset instead quietly uses the training dataset itself for classification. To predict the best suitable crop one should compare both training datasets and testing datasets, so this classification technique doesn't give the definite results.

In [6], the authors made a comparable study of three algorithms Naive Bayes, JRip and C4.5 algorithms. Naive Bayes classifier is a very basic probabilistic algorithm. Limitation of this algorithm is that it only works for the limited number of training datasets to determine the means and variances of the values required for the classification.

C4.5 is a classifier that generates a decision tree based on a set of labelled input dataset. The decision tree is tested against undiscovered labelled test dataset to evaluate how well the generated decision tree functions.

JRip classifier algorithm was proposed by William W. Cohen [7]. Classes are tested in incrementing size and an introductory set of rules for the class is generated using incremental shorten error JRip advances by considering all the examples of an appropriate judgement in the training dataset as a class, and examining a set of rules that checks all the members of that class. Thereafter it proceeds to the next class and does the same, repeating this until all classes have been covered. The soil attributes like pH value, Electrical Conductivity(EC), Organic Carbon(OC), Phosphorous(P), Potassium(K), Iron(Fe), Zinc(Zn), Manganese(Mn), Copper(Cu) are considered for the soil classification system. Naive Bayes classifier showed an accuracy of 38.4%, JRip classifier showed an accuracy of 90.24% and C4.5 showed an accuracy of 91.9%. The proposed system endorse appropriate fertilizer for the given soil sample for all the attributes mentioned above and cropping pattern.

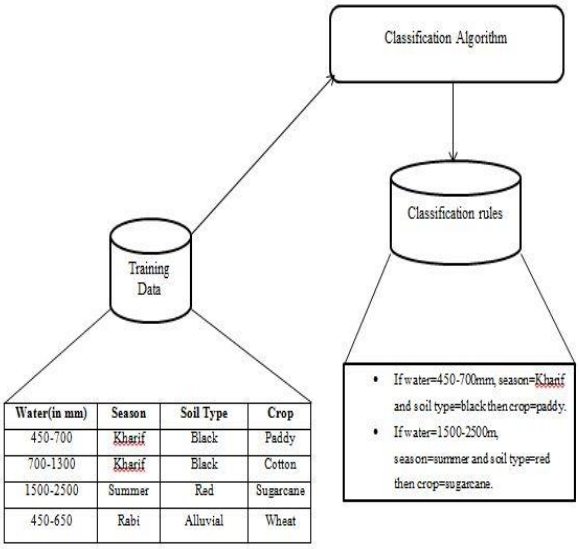
In [5], SVM is a rule based classifier. Rules are evaluated from training dataset in J48 decision tree and misclassified instances are analyzed in KNN algorithm. The authors suggested a new system by combining both KNN and J48 algorithms. Support vector regression avoids difficulties of using linear function in large input samples capacity and optimization of a complicated problems converted into simple linear function optimization. One of the drawbacks of SVM technique is we should choose the appropriate kernel otherwise we cannot predict the best suitable crop for the prediction. Other drawbacks include that the dataset size that we consider for training and testing should not be too enormous and also the speed at which we obtain results is very lower.

In [8], the authors presented a methodology named PSO-SVM feature selection for choosing the necessary features from the dataset on the grounds of which classification can be performed accurately. The authors considered training datasets for crops such as mushroom or soyabean. The proposed methodology compared with the existing methodologies by providing various input datasets. The proposed methodology exceeded the accuracy showed by the existing methodologies for the crops mentioned above. Future enhancements can be made to the proposed methodology by shortening the rules generated.

In [2], authors provided a new induction system of decision rule set to solve multi criteria ranking problems. Such rule set consists of one convinced rule subset induced from rough lower approximations and one satisfying rule subset induced from the separated rough boundary regions. The proposed rule set is not mutually exclusive with the classical nominal rule set. On the contrary, it provides a possible complement or alternative for Decision Maker in conducting Multi Criteria Decision Analysis. Firstly, both of rule sets include the same certain rules. Then, suggested satisfying rules provide the assignment as ‘‘at least/most class’’ with quantitative measurements and approximate rules in nominal set provide the assignment like ‘‘the union of two classes’’. This paper does not consider attributes/criteria reduction which is an important optimization procedure.

In [3], authors discussed and evaluated two Rough set approaches that use the Discernibility Matrix (DM) and some heuristics to compute reducts set; the two algorithms are the Johnson and Object Reduct using Attribute Weighting technique algorithm (ORAW). The ORAW algorithm computes three weighting mechanism (Local Weighting, Global Weighting, and finally Cardinality Value) for each feature that using it further in selecting the most significant features from the actual feature set of a dataset. If two features or more have the same local weight value, the algorithm suggests computing the global weight value, also if the global weight value for the features are equal, the algorithm return to the last solution that is the feature cardinality value. In this event, if the cardinality values are same as the significant features are select randomly. The two approaches were compared with other classification methods such as One Rule and C4.5 algorithms. The practical results showed that the ORAW approach obtained the highest accuracy using the generated reduct set over different standard datasets. This shows the qualities weighting strategies utilized as a part of ORAW are functioning admirably to decide the best characteristic that ought to be incorporated into the reduct sets.

In [4], authors proposed a hybrid soft decision model which helps farmers to select a crop in their agricultural land. This model was developed for choosing one crop among the crops paddy, cumbu, groundnut, ragi and sugarcane. Input criteria such as water, input, season, support, soil, risk and facilities. This model was developed by combining few soft computational techniques and two different weight calculation methods were used. This model showed an accuracy of 100% for paddy, 92% for cumbu, 96% for groundnut, 92% for ragi and 96% for sugarcane and on the average this model is 95.2% accurate.



Water(in mm)	Season	Soil Type	Crop
450-700	Kharif	Black	Paddy
700-1300	Kharif	Black	Cotton
1500-2500	Summer	Red	Sugarcane
450-650	Rabi	Alluvial	Wheat

Fig. 1: Rules generation using Classification algorithms

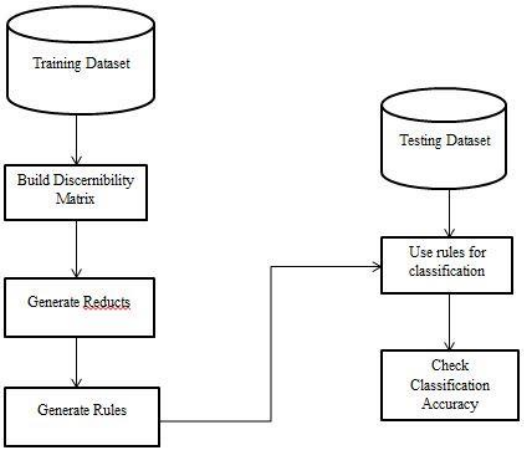


Fig. 2: Rules generation using reduct sets

Author and publication	Techniques used	Parameters achieved	Limitations/Future enhancements
Enas A. AliKhashashneh, 2013	Object Reduct using Attribute Weighting technique algorithm (ORAW)	Uses generated reducts sets over different standard datasets.	Achieved 85.6% accuracy, can be improved by using more appropriate reduct sets.
Chai J, Liu JN, 2013	Multi-Criteria Decision Analysis	To find best suitable crop we need to consider multiple criterion, this method provides rules to induce in considering multi-criteria.	Doesnot consider criteria reduction which is an important optimization procedure.
Surabhi Chouhan, 2016	PSO-SVM Feature Selection Algorithm	Provides more generated rules and high selection of features, provides high accuracy for the classification of datasets such as soyabean or mushroom.	Rules generated can be reduced as well as can bring down the execution time.
Meenakshi Malik, 2017	Crop Selection Algorithm	Multi model gives accuracy compared to single model.	Accuracy can be improved by using Fuzzy etc.

3. Conclusion

Hybrid soft decision model shows 88% accuracy. Johnson's reduct classifier produces satisfactory results compared to other classification algorithms. The classifier produces classification rules for three agricultural crops sugarcane, paddy and groundnut. The overall accuracy of this classifier is 92%. This classifier can be easily resilient to other agricultural crops also.
The problem with Johnson's reduct classifier is that it has input number constraints in the dataset. To predict the best suitable crop we need to consider large number of inputs, so this classifier can be further enhanced to take large input dataset.

References

[1] <https://en.wikipedia.org/wiki/Farmers%27_suicides_in_India>
[2] Chai J, Liu JN (2013) "Dominance-based decision rule induction for multicriteria ranking." Int J Mach Learn Cybernet 4(5):427–444
[3] AliKhashashneh EA, Al-Radaideh QA (2013) "Evaluation of discernibility matrix based reduct computation techniques." 5th international conference on computer science and information technology (CSIT). IEEE, pp. 76–81
[4] Deepa N, Ganesan K (2016) "Multi-class classification using hybrid soft decision model for agriculture crop selection." Neural Comput Appl 1–14
[5] Sarita Patil and Ekta Ukey (2017) **"**Adaptive rule based multi class SVM classification of big biological data." International Journal of Development Research
[6] Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar "Soil data analysis using classification techniques and soil attribute prediction."
[7] Anil Rajput, Ramesh Prasad Aharwal, Megha Dubey, S.P. Saxena, Manmohan Raghuvanshi "J48 and JRIP rules for e-Governance Data."
[8] Surabhi Chouhan, Divakar Singh, Anju Singh (2016) "An improved feature selection and classification using decision tree for crop datasets." International Journal of Computer Applications