Crop yield prediction using Machinelearning

[1] Chintagunti Praveen [2] H. Ateeq Ahmed
[1] P.G.Scholar [2] Associate Professor,
[1,2,]DEPARTMENT OF CSE

[1,2]Dr. K.V SUBBA REDDY INSTITUTE OF TECHNOLOGY
Dupadu Railway Station, Lakshmipuram Post, Kurnool, Andhra Pradesh, India - 518218

ABSTRACT

Agriculture is the pillar of the Indian economy and more than 50% of India‟s population aredependentonagriculturefortheirsurvival.Variationsinweather,climate,andothersuchenvironmental conditions have become a major risk for the healthy existence of agriculture.Machine learning (ML) plays a significant role as it has decision support tool for Crop YieldPrediction (CYP) including supporting decisionson what cropstogrow and what to do duringthe growing season of the crops. The presentresearch dealswith a systematic review thatextracts and synthesize the features used for CYP and furthermore, there are a variety of methodsthat were developed to analyze crop yield prediction using artificial intelligence techniques. ThemajorlimitationsoftheNeuralNetworkarereductionintherelativeerroranddecreasedprediction efficiency of Crop Yield. Similarly, supervised learning techniques were incapable tocapture the nonlinear bond between input andoutput variables faceda problem during theselectionoffruitsgradingorsorting.Manystudieswererecommendedforagriculturedevelopment and the goal was to create an accurate and efficient model for crop classificationsuch as crop yield estimation based on the weather, crop disease, classification of crops based onthe growing phase etc., This paper explores various ML techniques utilized in the field of cropyieldestimation and provided adetailedanalysis in terms of accuracyusingthe techniques.

Keywords:-MLtechniques, CYP.

1.INTRODUCTION

The history of agriculture in India dates backto the Indus Valley Civilization Era. Indiaranks second in this sector. Agriculture andalliedsectorslikeforestryandfisheriesaccount for 15.4 percent of the GDP (grossdomestic product) with about 31 percent ofthe workforce.India ranksfirstgloballywiththe highest net cropped area followed by USand China. Agriculture is demographicallythebroadesteconomicsectorandplaysasignificantroleintheoverallsocio-economicfabricofIndia.Duetotherevolution in industrialization, the economiccontribution of agriculture to India‟s GDP issteadilydecliningwiththecountry‟sbroad-based economic growth. The problem thatthe Indian Agriculture sector is facing is theintegrationoftechnologytobringthedesiredoutputs.Withtheadventofnewtechnologies and overuse of non-renewableenergyresourcespatternsofrainfallandtemperature are disturbed. The inconsistenttrendsdevelopedfromthesideeffectsofglobal warming make it cumbersome for thefarmerstoclearlypredictthetemperatureand rainfall patterns thus affecting their cropyieldproductivity.Inordertoperformaccurate

prediction and handle inconsistenttrendsintemperatureandrainfallvariousmachinelearningalgorithmslikeRNN,LSTM, etc can be applied to get a pattern. Itwill complement the agricultural growth inIndia and all together augment the ease ofliving for farmers. In past, many researchershave applied machine learning techniques toenhance agricultural growth of the country.This paper focuses on predicting the yield ofthecropbyapplyingvariousmachinelearning techniques. The outcome of thesetechniques is compared on the basis of meanabsoluteerror.Thepredictionmadebymachinelearningalgorithmswillhelpthefarmers to decide which crop to grow to getthe maximum yield by considering factorsliketemperature,rainfall, area,etc

Objectives to be followed in the future aregivenbelow:

1. Depending on the dissimilar crop featuredivisions,themodulatingfactorvaluesofMLalgorithmsdiffertoattainperfectapproximation.
2. When the quantity of input elements isreduced,ANNisutilized.Theoptimalfeaturewasbeingempiricallyselectedforappropriate cropyieldestimation.
3. The advantage of ML method regressionistoavoiddifficultiesofusingalinearfunction in large outputsample space andoptimizationofcomplexproblemstransformedintosimplelinearfunctionoptimization.
4. ML algorithm can be executed with anenormoussoildatasetforcropyieldestimation.
5. The ML techniques, through observationoftheagriculturalfields,providedthenecessarysupporttothefarmersinincreasingcropproduction toagreat extent.

2.RELATED WORK

Most of the Exiting models utilized neuralnetworks, random forests, KNN regressiontechniquesforCYPandavarietyofMLtechniqueswerealsousedforbestprediction. The problems faced in existingresearchforcropyieldpredictionusingmachinelearningarestated below:

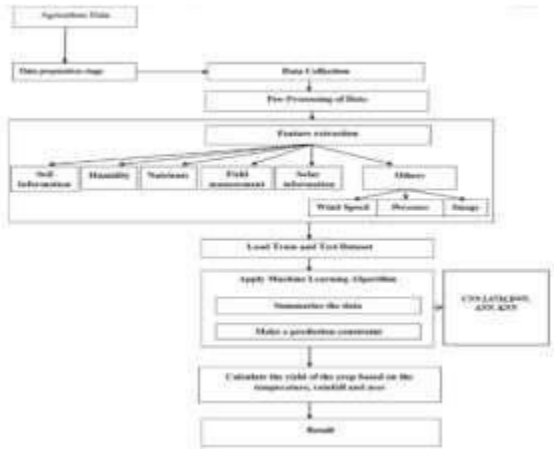
1. Creation, repair and maintenance of MLalgorithms required huge costs as they areverycomplex.
2. MLtechniqueusedforCropyieldprediction (mustard, wheat) combined inputand output data but failed to obtain betterresultsstatistically
3. Due to the nature of linear connection intheparameters,theregressionmodelwasfailed to provide the exact prediction in acomplexsituationsuchasextremevaluedataand nonlinear data.
4. The existing K-NN models were used forclassificationforyieldpredictionbutloweredtheperformanceduetononlinearand highly adaptable issues present in KNN.They were operated in a locality model thatincremented the dimensionality of the inputvectormade confusion forclassification.
5. Anappropriatedecisionwasnottakenduringclassificationbecauseafewerquantity of data was available for estimationofcropyield.

Thispaperfocusesonthepracticalapplication of machine learning algorithmsand its quantification. The work presentedhere also takes into account the inconsistentdatafromrainfallandtemperaturedatasetstogetaconsistenttrend.Cropyieldprediction is determined by considering allthefeaturesincontrastwiththeusualtrendofdeterminingthepredictionconsideringone feature at a time. From the studies mostof the common algorithms used were CNN,LSTM,DNNalgorithmsbutstillimprovementwasstillrequiredfurtherinCYP.The present research shows severalexisting models that consider elements suchastemperature,weathercondition,performingmodelsfortheeffectivecropyieldprediction.Ultimately,theexperimental study showed the combinationof ML with the agricultural domain field forimprovingtheadvancementincropprediction.

3. IMPLEMENTATION

SystemArchitecture:

Thestepsthatareinvolvedincropyieldpredictionusingmachinelearningmethodology are stated as follows. Firstly,the agriculture Data is utilized for the cropyield prediction, next, the data is undergonefor pre-processing to remove the noisy data.Thepre-processeddataisundergoneforfeatureextractionprocessthatincludesfeaturessuch as soil



UploadCropDataset

The crop production dataset that is used topredict the name and yield of the crop is fedinto classification and regression algorithms.

PreprocessDataset

Experiments were conducted onIndiangovernmentdatasetandithasbeenestablishedthatRandomForestRegressorgivesthehighestyieldpredictionaccuracy.SequentialmodelthatisSimpleRecurrentNeuralNetworkperformsbetteronrainfallpredictionwhileLSTMisgoodfortemperatureprediction.Bycombiningrainfall,temperaturealongwithotherparameters like season and area,yieldprediction fora certain district can be made.

Train MachineLearning

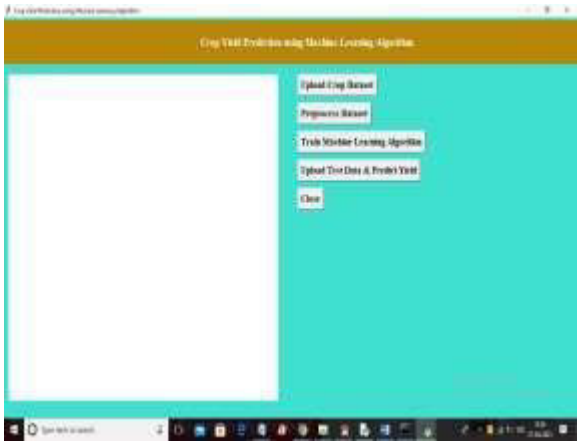
This focuses on district wise yield predictionaccording to the crop sown in the district.Yieldisbeingpredictedforgivencropsdistrictwise and cropswith bestyield.

UploadTestData&PredictYield

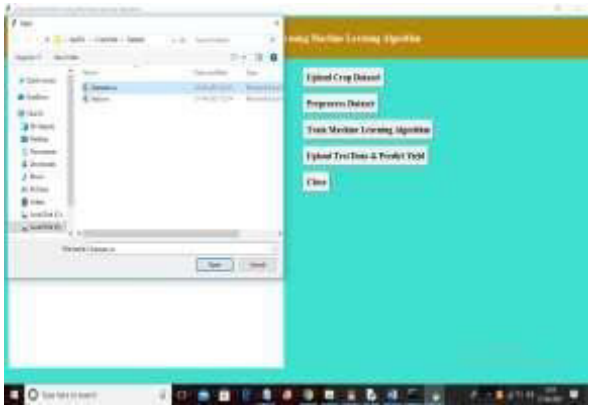
ResultsrevealsthatRandomForestisthebestclassifierwhenallparametersarecombined. This will not only help farmers inchoosing the right crop to grow in the nextseasonbutalsobridgethegapbetweentechnologyand the agriculturesector.

4. EXPERIMENTALRESULTS

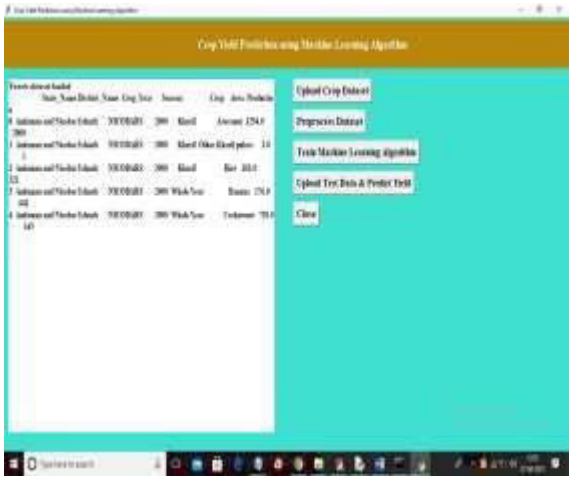
Torunprojectdoubleclickon„run.bat"filetoget below screen



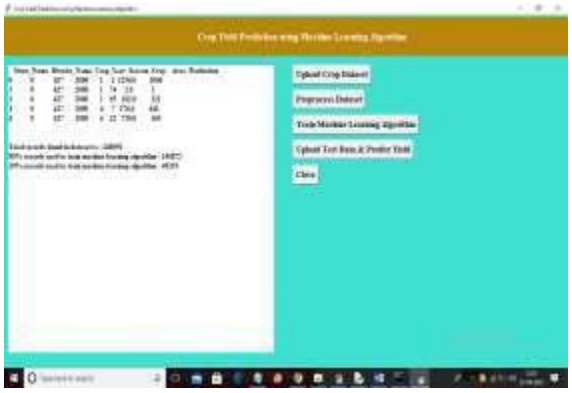
Inabovescreenclickon„UploadCropDataset"buttontouploaddataset



In abovescreenselectinganduploading„Dataset.csv"fileandthenclickon„Open"buttontoloaddatasetandtogetbelowscreen

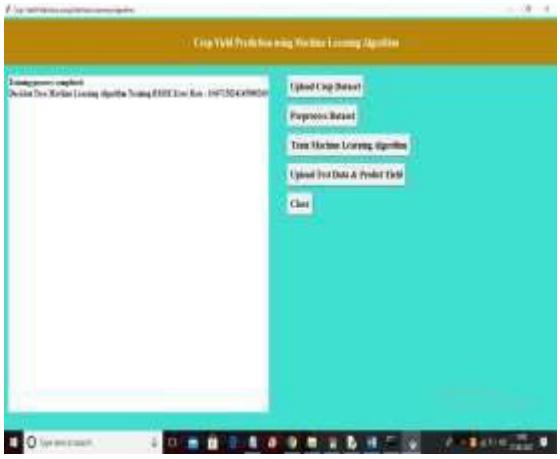


In above screen dataset loaded and we cansee dataset contains some non-numericvalues and ML will not take non-numericvalues so we need to preprocess dataset toconvert non-numeric values to numericvalues by assigning ID to each non-numericvalue. So click on „Preprocess Dataset‟‟buttonto process dataset

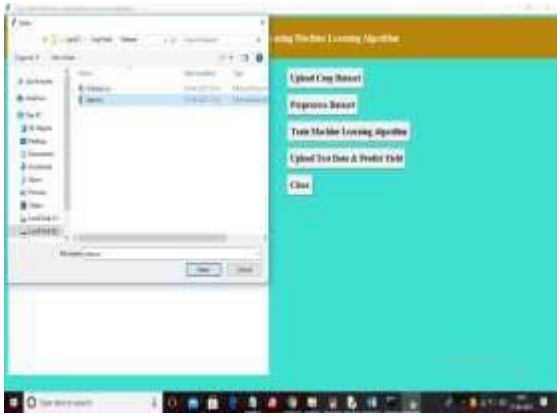


In above screen all non-numeric valuesconverted to numeric format and in belowlines we can see dataset contains total246091 records and application using (80%)196872recordstotrainMLandusing(20%)

49219 records to test ML prediction errorrate(RMSE(rootmeansquareerror)).Nowclick on „Train Machine LearningAlgorithm‟‟ button to train Decision TreeMachine learning algorithm on abovedataset and then calculate prediction errorrate



In above screen ML is trained and we gotprediction error rate as 0.067% and nowDecisionTreemodelisreadyandnowclickon „Upload Test Data & Predict Yield‟‟button to upload test data and thenapplicationwill predict production



In abovescreenselectinganduploading “test.csv‟‟fileandthenclickon„Open‟‟buttontoloadtestdataandthenapplicationwillgive below prediction result

Inabovescreeneachtestrecordisseparatedwith newline and in above screen in squarebracket we can see test data values and aftersquare bracket we can see predictedproduction and after that we can seepredicted YIELD per acre. So each testrecord and its prediction is

separated withnewline. You can scroll down above textareato view all records.

5. CONCLUSION

The present research work discussed aboutthevarietyoffeaturesthataremainlydependent on the data availability and eachof the research will investigated CYP usingMLalgorithmsthatdifferedfromthefeatures.Thefeatureswerechosenbasedupon the geological position, scale, and cropfeaturesandthesechoiceswereмainlydependent upon the data-set availability, butthemorefeaturesusagewasnotalwaysgiving better results. Therefore, finding thefewer best performing features were testedthatalsohavebeenutilizedforthestudies.

Most of the exiting models utilized Neuralnetworks, random forests, KNN regressiontechniquesforCYPandavarietyofMLtechniqueswerealsousedforbestprediction.Fromthestudiesmostofthecommonalgorithmsusedwere CNN,LSTM,DNNalgorithmsbutstillimprovementwasstillrequiredfurtherinCYP.Thepresentresearchshowsseveralexisting models that consider elements suchastemperature,weathercondition,performingmodelsfortheeffectivecropyieldprediction.Ultimately,theexperimental study showed the combinationof ML with the agricultural domain field forimprovingtheadvancementincropprediction.However,stillmoreimprovementinfeatureselectionwasrequiredintermsoftemperaturevariationaspects effects on agriculture. In the furtherstudies,thekeypossibility thatshouldbeconcentratedsuchasfirstlythedelaytobordertopographicalareasrequiredadditional-explicittreatment.Next,anonparametricportionofthemodelusingmachinelearningalgorithmandthirdly,usingfeaturesfromdeterministiccropmodelstogetperfectstatisticalCO2fertilization. By following above-mentionedobjectives, the crop yield estimation wouldbeimprovedbyfurtherresearchers.

Additionally,inthecropyieldestimation,fertilizershouldalsobeconsideredforexecuting soil forecasts that agriculturalist tomakeabetterjudgmentbasedonthesituation of low crop yield estimation. Basedontheoutcomesobtainedforthestudyfurtherweneedtobuildanddevelopamodel basedon DLfor CYP.

6. REFERENCES

[1] R.Ghadge,J.Kulkarni,P.More,S.Nene, and R. L. Priya, “Prediction of cropyield using machine learning,” Int. Res. J.Eng.Technolgy, vol. 5, 2018.

[2] F. H. Tseng, H. H.Cho, and H. T. Wu,“Applyingbigdataforintelligentagriculture-basedcropselectionanalysis,”IEEEAccess,vol.7,pp.116965-116974,2019.

[3] A. Suresh, N. Manjunathan, P. Rajesh,and E. Thangadurai, “Crop Yield PredictionUsingLinearSupportVectorMachine,”European Journal of Molecular & ClinicalMedicine,vol.7,no.6,pp.2189-2195,2020.

[4] M. Alagurajan, and C. Vijayakumaran,“ML Methods for Crop Yield Prediction andEstimation:AnExploration,”InternationalJournalofEngineeringandAdvancedTechnology,vol. 9 no. 3, 2020

[5] P.Kumari,S.Rathore,A.Kalamkar,andT.Kambale,“ PredicitionofCropYeildUsing SVM Approch with the Facility of E-MARTSystem” Easychair2020.

[6] S. D. Kumar, S. Esakkirajan, S. Bama,andB.Keerthiveena,“Amicrocontrollerbased machine vision approach for tomatograding and sorting using SVM classifier,”Microprocessors and Microsystems, vol. 76,pp.103090, 2020

[7] P. Tiwari, and P. Shukla, “Cropyieldprediction by modified convolutional neuralnetworkandgeographicalindexes,”International Journal of Computer Sciencesand Engineering, vol. 6, no. 8, pp. 503-513,2018.

[8] P. Sivanandhini, and J. Prakash,

"CropYieldPredictionAnalysisusingFeedForwardandRecurrentNeuralNetwork,"International Journal of Innovative Scienceand Research Technology, vol. 5, no. 5, pp.1092-1096,2020.

[9] N.Nandhini,andJ.G.Shankar,"Prediction ofcropgrowth using machinelearning basedonseed,"Ictactjournalonsoftcomputing, vol. 11,no. 01, 2020

[10] A. A. Alif, I. F. Shukanya, and T. N.Afee,"Croppredictionbasedongeographicalandclimaticdatausingmachinelearninganddeeplearning",Doctoraldissertation,BRACUniversity)2018.