IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12501

Crop Yield Prediction Using CNN

Prof. Krishna S. Kadam

Assistant Professor, DKTE Society's Textile and Engineering Institute, Ichalkaranji

Abstract: Crop yields are critically dependent on weather. A growing empirical literature models this relationship in order to project climate change impacts on the sector. We describe an approach to yield modeling that uses a semi parametric variant of a deep neural network, which can simultaneously account for complex nonlinear relationships in high-dimensional datasets, as well as known parametric structure and unobserved cross-sectional heterogeneity. Using data on corn yield from the US Midwest, we show that this approach outperforms both classical statistical methods and fully-nonparametric neural networks in predicting yields of years withheld during model training. Using scenarios from a suite of climate models, we show large negative impacts of climate change on corn yield, but less severe than impacts projected using classical statistical methods. In particular, our approach is less pessimistic in the warmest regions and the warmest scenarios using CNN.

Keywords: Crop, semi parametric. Yield ,CNN.

I. INTRODUCTION

Farming is main occupation of India. Crop yield has direct impact on national and international economies annually and the yield predicted plays significant part in the food management and agriculture sector. India's economy mainly depends on agriculture yield growth and their related agro industry. 61% percent of the total geographical area of India comes under agriculture. This model is taking a input as a dataset of the India. This particular dataset includes the data from all the states, districts of India and the crops in the respective regions. So, we are going to study of different states and districts of the India. A prerequisite of intelligent system has brought artificial neural network to become a new technology which provides assorted solution for complex problems in agriculture researches. Performance of agriculture sector mainly hinges on natural forces such as spatio-temporal distribution of rainfall, temperature, climate etc, with the result any deviation of monsoon from the normal pattern brings about in numerous fluctuations in area and production. Crop yield has direct impact on nation and international economies annually and the yield predicted plays significant part in the food management and agriculture sector. The task is to build prediction model for crop production using data predictive. The scope of our project is to predict crop production with accuracy Along with .csv file that available for raining and testing. The intended software that shall be required by system Anaconda Navigator, Jupyter Notebook, Online Google collaborator tool freely provided Google for enhancing performance of the separations on the GPU's. This shall enable the system to Machine Learning models in a considerably less time by training testing of this large data set comprising of approximately 248000 records.

II. LITERATURE OVERVIEW

[1] P.Surya, Dr. I.LaurenceAroquiarajPeriyarUniversity, Salem, Tamilnadu, India [2018] describes In thierproposed work, collected agriculture dataset will be used to get crop yield prediction model using various regression techniques. Regression analysis was tested for the effective prediction or forecast of the agriculture yield for various crops in Tamilnadu state particularly in North Western zone of Tamilnadu. North western zone of tamilnadu state data consist four districts. By the analysis depends on the results of predictor model, in the north western zone, under the area having more cultivated crops are Tapiaco, Sugar cane, Ragi, Maize, Groundnut. In this paper linear regression gives the better solution to predict crop production.But above solution particularly focused on North Western zone of Tamilnadu, which has similar contents. It has small data, so may gives less accuracy.

[2] B. Devika, B. Ananthi they use classification technique to group similar content area wise and year wise. k-nn and linear regression algorithm are used and linear regression gives better accuracy.But data may be nonlinear pattern with district wise so this model can gives different output.

[3] Shakil Ahamed, Navidv Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, RashedurM Rahman are estimate crop yield for major cereal crops in major districts of Bangladesh. In this I have considered the effects of environmental (weather) and area of production as factors towards crop production in Bangladesh. Taking these factors into consideration as datasets for various districts, they applied classification techniques to divide regions; and then apply suitable classification techniques to obtain crop yield predictions. Here proposed methods was linear regression, k-nn, and neural network.Neural network gives better prediction result.

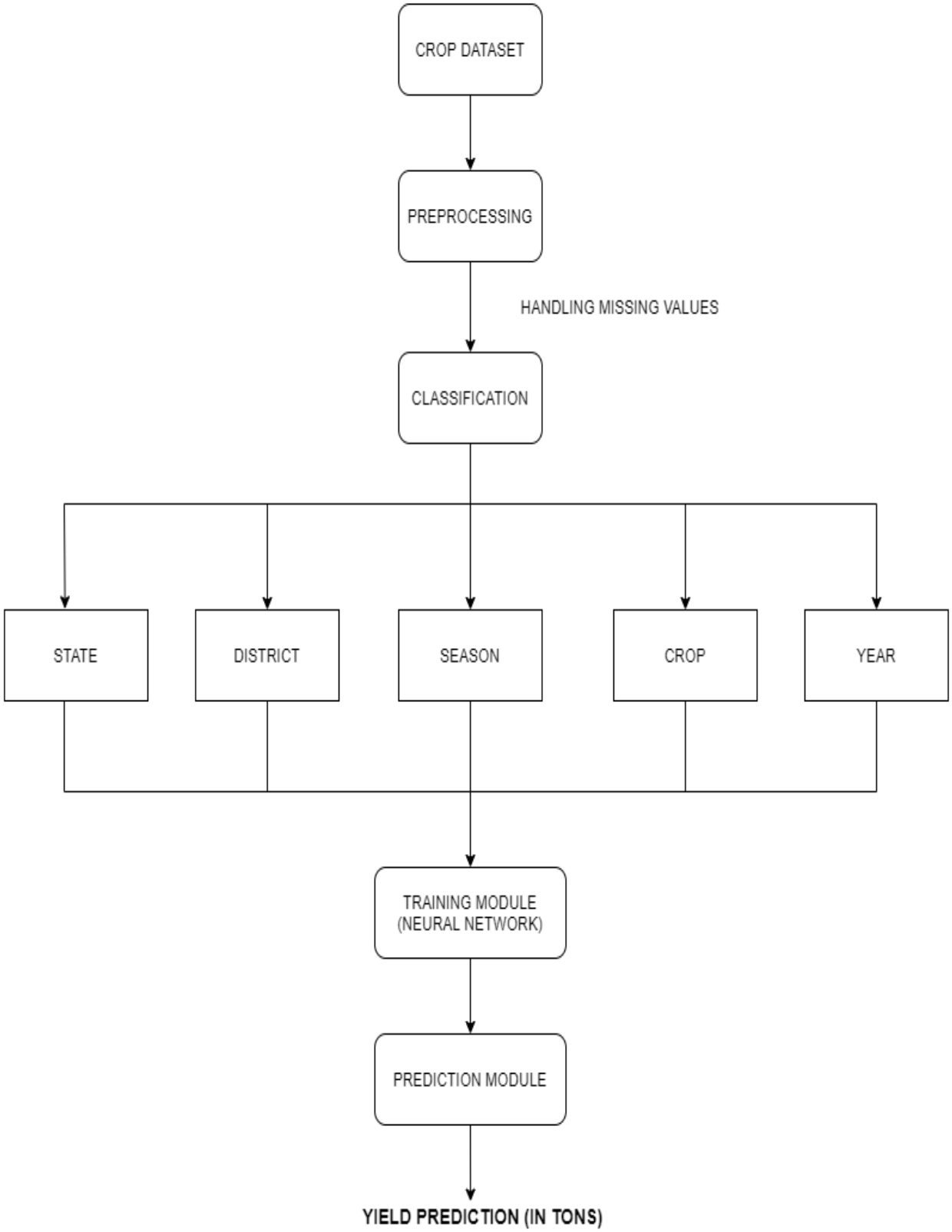
IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12501

III. SYSTEM DESIGN



IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12501

IV. METHODOLOGY

Google colabinitilization:

Google's free cloud service for AI Developers with Colab, you can develop deep learning applications on the GPU for free.

The most important feature that distinguish Colab from other free cloud services is; colab provides GPU and is totally free.

1) Create new Colab notebook and rename
2) Setting free GPU

It is so simple to alter default hardware (CPU TO GPU). Just follow edit>Notebook settings>change Runtime type and select GPU as hardware accelator

3) Most Colab file with drive so we can access folder and files on Google Drive

Implementation of Preprocessing

1) Read the dataset using pandas

Data = pandas.read_csv('dataset.csv')

Null values were removed. Eliminate redundant features, outlier were removed. Filters the specific range values

2) Encode the dataset from categorical values to numeric values using label encoding

:le.fit_transform(data['District_Name'])

le.fit_transform(data['Season'])

le.fit_transform(data['Crop'])

3) Dataset is splited into train and test using skelarn library

Train,test = train_test_split(data,test_size=0,2)

Save the train and test data frame as train.csv and test.csv file

Implementation of classification model

1) Read the dataset using pandas

data = pandas.read_csv('dataset.csv')

i. Get unique values from state column and store into list

State = data.state_name.Unique()

ii. Sort the array of state

iii. Save file to .csv

df = pd.DataFrame(State)

df.to_csv('/gdrive/state.csv')

2) Repeat the step 1 for district, season, crop.

Save to drive : district.csv, season.csv, crop.csv

Implementation of Training Module of yield production model

a. Import tensorflow, keras, pandas import layer
b. Read the train & test dataset
c. Normalization :-

Normalize the numerical value of dataset using standard deviation and mean.

crop_year, Area and Production have the numerical values

cols_to_norm = cols_to_norm_mean / standard_deviation

d. Stores each column value in separate array
e. Build neural networks

i. Create input layers for each input parameters.

year_input = Input(shape=(1,), dtype='float32')

area_input = Input(shape=(1,), dtype='float32')

district_input = Input(shape=(1,)dtype='float32')

season_input = Input(shape=(1,)dtype='float32')

crop_input = Input(shape=(1,)dtype='float32')

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12501

ii. First we use one hot encoding for categorical inputs but neural networks becomes bulky for thousands categories. It cant handle. So we use alternative embedding.
Embedding(input_dim = no_of_categories, output_dim = embedded_size, input_length = 1)
Where
Embedded_size = min(no_of_categories/2, 50)
Do this for district_input, season and crop with their input layers. This is most important step of categorical encoding in the project.

iii. Concatenate all layers & CNN and Create model

f. Compile the model

g. Train the model using input & output parameters dataset.

Train the model based on mape&mse metrics.

By the hyperparameters tunning improve the accuracy of model. Hyperparameters like `layers, neurons, batch_size, epochs.

h. Test the trained model using evaluate

i. Save the trained model to drive as model.h5 with their weights saved model is useful for future prediction & use

V. TESTING AND RESULT

a) Time

We have to improve the efficiency of model. We tried embedding with neural network to train the model. The time taken by Method is :

Method	**Time Required (avg)**
Neural Network With Embedding	0.7 seconds

b) Model Accuracy

We train the neural network by train dataset and test using test dataset. Our main goal of project is to improve the accuracy based on mape (mean absolute percentage error).
We reduce the error 29.43% that mean we achieve accuracy of 71% approx.

c) Output

a. Using one hot encoding before neural network on categorical data :
We test the training module, we get 51% accuracy

b. Using embedding in neural network model:
Encode the categorical data by embedding and hyper parameters tuning we get 62% accuracy
After improving batch size, epochs, layers, neurons we achieve accuracy of 71% approx.

c. Test input and front-end design:
We get input from html form and give to model for prediction. Here we check validation and selection of option

Accuracy

Method	**Accuracy**
One hot encoding with neural network	51%
Embedding in neural network	71%

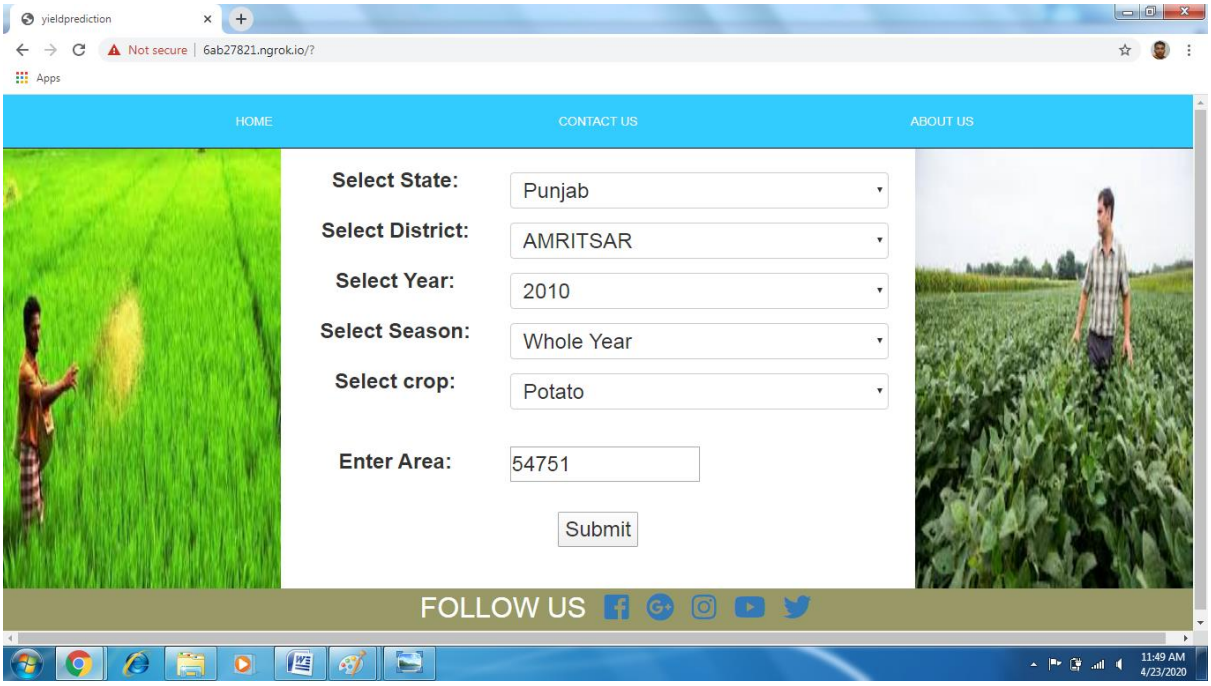
By using embedding for categorical data having thousands of categories and hyper parameter tuning in neural network, Model gives better result.

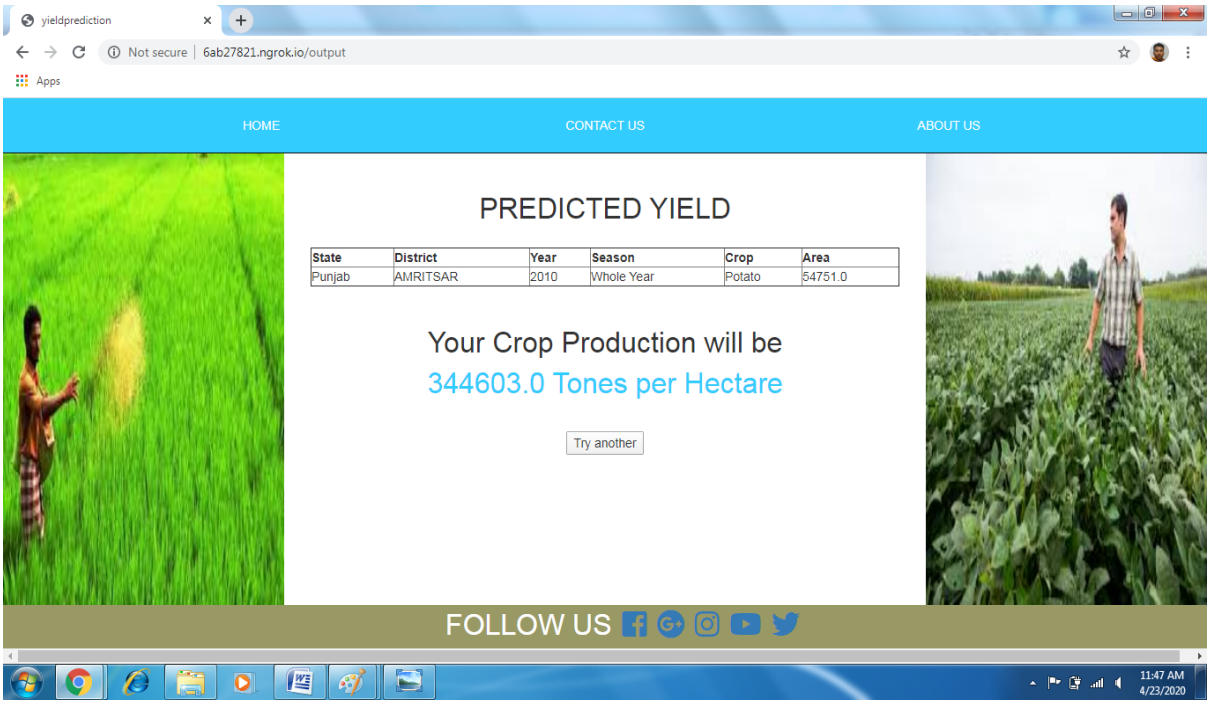
IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12501





VI. CONCLUSION AND FUTURE SCOPE

As per above analysis, embedding method & hyper parameter tuning in neural network model gives better result. Crop yields are critically dependent on weather. A growing empirical literature models this relationship in order to project climate change impacts on the sector.

We describe an approach to yield modeling that uses a semi parametric variant of a deep neural network, which can simultaneously account for complex nonlinear relationships in high-dimensional datasets, as well as known parametric structure and unobserved cross-sectional heterogeneity

DOI: 10.17148/IJIREEICE.2024.12501

REFERENCES

[1]. P.Surya, 2 Dr. I.Laurence Aroquiaraj," Crop yield prediction in agriculture using data mining predictive analytic techniques", IJRAR, Volume 5, Issue 4,page no 783, December 2018.

[2]. B. Devika, B. Ananthi," analysis of crop yield prediction using data mining technique to predict annual yield of major crops", International Research Journal of Engineering and Technology (IRJET), Volume 5, December 2018.

[3]. Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman," Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh ", IEEE SNPD,June 2015

[4]. Dataset-data.gov.in

[5]. https://www.frontiersin.org/articles/10.3389/fpls.2019.00621/full

[6]. A Model for Prediction of Crop Yield E. Manjula Pachaiyappas College India manjulababu123@gmail.com S. Djodiltachoumy Pachaiyappas College India djodilatchoumy@hotmail.com

[7]. http://github.com/cranedroesch/ML_yield_climate_ERL

[8]. Ciscar J-C, Fisher-Vanden K and Lobell D B 2018 Synthesis and review: an inter-method comparison of climate change impacts on agriculture Environ. Res. Lett. 13 IOPscienceGoogle Scholar

[9]. Lobell D B and Burke M B 2010 On the use of statistical models to predict crop yield responses to climate change Agric. Forest Meteorol. 150 1443–52CrossrefGoogle Scholar.

[10]. Roberts M J, Braun N O, Sinclair T R, Lobell D B and Schlenker W 2017 Comparing and combining process-based crop models and statistical models with some implications for climate change Environ. Res. Lett. 12 095010 OPscienceGoogle Scholar