A

Major Project report on

**CROP YIELD PREDICTION USING**   
**DEEP LEARNING**

Submitted in partial fulfilment of the requirements for the

Award of the degree of

**Bachelor of Technology**

**In**

**Information Technology**

By

**BYAGARI VINAY KUMAR (16881A1210)**

**KAITHI MOHAN RAO (16881A1217)**

**MOHIT SHAH (16881A1228)**

**SAI KOUSHIK REDDY (16881A1251)**

*Under the Guidance of*

**MS. SHEENA MOHAMMED**

Assistant Professor

Department of Information Technology



**DEPARTMENT OF INFORMATION TECHNOLOGY**

VARDHAMAN COLLEGE OF ENGINEERING

**(AUTONOMOUS)**

**(Affiliated to JNTUH, Approved by AICTE and Accredited by NBA)**

**Shamshabad - 501 218, Hyderabad**

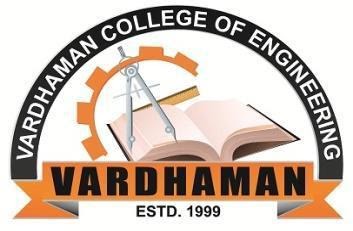
**2016-2020**

**VARDHAMAN COLLEGE OF ENGINEERING**

**(AUTONOMOUS)**

Shamshabad - 501 218, Hyderabad

**DEPARTMENT OF INFORMATION TECHNOLOGY**



**CERTIFICATE**

This is to certify that the project report entitled, “**CROP YIELD PREDICTION USING**   
**DEEP LEARNING”,** done by **, BYAGARI VINAY KUMAR (16881A1210),KAITHI MOHAN RAO(16881A1217), MOHIT SHAH (16881A1228) and SAI KOUSHIK REDDY (16881A1251)** submitted to the department of Information Technology impartial fulfilment of the requirements for the Degree of **BACHELOR OF TECHNOLOGY** in **Information Technology** from **Vardhaman College of Engineering (AUTONOMOUS), Hyderabad, during** the period 2016-2020. It is certified that he/she has completed the project satisfactorily.

**Signature of Supervisor: Head of the Department:**

**Ms. Sheena Mohammed Dr. Muni Sekhar Velpuru**

Assistant Professor, Professor & Head,

Dept. of Information Technology, Dept. of Information Technology,

Vardhaman College of Engineering, Vardhaman College of Engineering,

Hyderabad Hyderabad

**ACKNOWLEDGEMENT**

The satisfaction that accompanies the successful completion of the task would be put in complete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to **MS. Sheena Mohammed**, Assistant Professor, Department of Information Technology, Vardhaman College of Engineering, for her able guidance and useful suggestions, which helped us in completing the project work, in time.

We are particularly thankful to **Dr. Muni Sekhar Velpuru**, Professor& Head of Department, Information Technology for his guidance, intense support and encouragement, which helped us to mold our project into a successful one.

We show gratitude to our honourable Principal **Dr. S. Sai Satyanarayana Reddy**, for providing all the facilities and support.

We avail this opportunity to express our deep sense of gratitude and heartfelt thanks to all the members of management, for providing congenial atmosphere to complete this project successfully.

We also thank all our teaching and non-teaching staff members of Information Technology department for their valuable support and generous advice. Finally thanks to all our friends and family members for their continuous support and enthusiastic help.

**BYAGARI VINAY KUMAR (16881A1210)**

**KAITHI MOHAN RAO (16881A1217)**

**MOHIT SHAH (16881A1228)**

**SAI KOUSHIK REDDY (16881A1251)**

**ABSTRACT**

Agriculture plays a dominant role in the growth of the country’s economy for a developing nation like India. Currently, the crop yield prediction is carried out on large scale only by  statistical and crop simulation methods. Deep learning is the one of the most promising fields that can transform the future of various sectors. A lot of research and analysis is being conducted to build models to fit the agricultural data. Crop yield is a highly complex trait determined by multiple factors such as quality and type of soil, rainfall, temperature, quality of fertilizers used, solar radiation, etc; The parameters can be categorized as weather and non-weather. Our dataset is constructed from the real data in the Telangana state. We have taken the data of previous years from meteorological department and use APSIM simulator to create a simulation because of unavailability of field data and then train the generated data to create a deep neural network.

Deep neural networks are proven to be better for training models that have very very large datasets(millions of training samples) rather than other machine learning supervised methods. Crop yield for a large country like India when applied on a large scale involves very large datasets for better accuracy. Thus larger the dataset, we tend to choose ANN’s.

**CONTENTS**

Acknowledgement 03

Abstract 04

**CHAPTER 1 INTRODUCTION 06**

1.1 Introduction 07

1.2 Motivation 08

1.3 Objective of the project 09

**CHAPTER 2 LITERATURE SURVEY 10**

2.1 Existing systems 11

**CHAPTER 3 MODULE DESCRIPTION AND IMPLEMENTATION 15**

3.1 Preprocessing 16

3.2 Generating and running simulations 19

3.3 Automating the simulation process 24

3.4 Building an Artificial Neural Network base model 28

3.5 Hyperparameter tuning 30

**CHAPTER 4 CONCLUSION AND FUTURE WORK 32**

4.1 Conclusion 33

4.2 Project Future Enhancement 33

**REFERENCES 34**

**CHAPTER-1**

**INTRODUCTION**

**1.1 INTRODUCTION:**

Crop yield prediction and analysis is generally conducted by agricultural analysts who have profound insights about the working of this natural process. They use mathematical functions, data mining and statistical tools. We tried to approach this problem of crop yield prediction through deep learning . But what we lacked was field data. The only data available online was climate data for the Telangana districts. We tried to build a model on just that but there wasn’t good correlation between the parameters and crop, yield data wasn’t available either.

Then after some research, we came across an agriculture simulation tool called **APSIM(Agricultural Production Systems sIMulator ).** The tool is internationally recognized as a highly advanced platform for modelling and simulation of agricultural systems. The tool has many examples with sample input files to create sample simulations for various crops . It has huge number of changeable parameters that we can use to create a simulation. What we need to do is open some simulation from the available ones and change the parameters as required and run the simulation. We can also choose the output profile which consists of desired output parameters which in our case is only yield.

One of the inputs to the simulator is a meteorological (.met) file which contains columns of year, day, max temp, min temp, solar radiation and rainfall. So , we preprocessed the climate data available online into (.met) files. As a result, 31 (.met) files, one for each of 31 districts was generated. Other necessary parameters were obtained through research and inquiry with the agriculture professionals.

Generating the simulations for 31 districts and with different combinations of parameters is a very time consuming process. So, we needed to somehow automate the process for all of the simulations. We then observed that the simulation files (.apsim) were in xml format and that they can edited by any xml parser.

We chose XML **ElementTree** module available in python to modify and automatically change the parameters by writing a script to parse the (.apsim) file. Thus we generated desired number of simulation files by looping over the districts. And the simulations can be run at once from command prompt to generate the desired output files.

Then we had to parse these generated (.out) files by APSIM to get the yield parameter. The final task was building a Artificial Neural Network with the help of the parameters such as latitude, longitude, district , plant density(supplied in the APSIM) as input layer and yield as the output unit.

After some normalization and hyperparameter tuning , we are able to achieve a mean absolute error of 0.06 on validation data.

Initially we decided to work on the three major crops of Telangana- paddy(rice) , maize and cotton . But we encountered some variations in paddy in the simulator(upland ,lowland) which we did not have much information about. And also there was some issue with the maize simulation (the clock parameter was anachronical with our met data and it was getting reverted to its default value from the (.dll) despite of various attempts to change it. So , we had to work on just cotton crop.

**1.2 MOTIVATION:**

Farmers all over the country face several issues regarding the crops. Many of them are uncertain about what kind of crop suits their land most and what measures are to be taken for that particular crop. Crop yield depends on vast number of factors and even some minute change that may seem negligible may effect the crop significantly.

In the recent years, due to the increased suicide rates in farmers and unpredictable climate, climate change; we decided to try to make something that might help them.

**1.3 OBJECTIVE:**

Our objective has quite changed from the one we thought at the beginning of this project. As we dived deeper into the working of crop production, we understood that such a complex natural process cannot be simply modelled by some neural network with just handful of parameters.

(So, we used APSIM simulator to fit various parameters, most of them left as is while we modified some of them)

Now, our objective is to capture the functionality of the APSIM simulator in a neural network , so that we need not create and run the simulations each time we want to predict the yield. All we need to do is to change the input parameters we want to change and get the output from the neural network.

We have chosen only handful of such parameters currently but will be incorporating more of them in future work.

When given the exact parameters for each district, our model predicts which district is likely to give the higher yields for a particular crop.

**CHAPTER-2**

**LITERATURE SURVEY**

**2.1 EXISTING SYSTEMS:**

Crop yield is a highly complex trait determined by multiple factors such as genotype, environment, and their interactions. Accurate yield prediction requires fundamental understanding of the functional relationship between yield and these interactive factors, and to reveal such relationship requires both comprehensive datasets and powerful algorithms. The results also revealed that environmental factors had a greater effect on the crop yield than genotype. Crop yield prediction is of great importance to global food production. Policy makers rely on accurate predictions to make timely import and export decisions to strengthen national food security . Seed companies need to predict the performances of new hybrids in various environments to breed for better varieties . Growers and farmers also beneﬁt from yield prediction to make informed management and ﬁnancial decisions . However, crop yield prediction is extremely challenging due to numerous complex factors. Weather prediction is an inevitable part of crop yield prediction, because weather plays an important role in yield prediction but it is unknown a priori.

There have been a number of research studies undertaken that focus on the importance of using data mining as a supplementary tool in transforming large volumes of agricultural data into meaningful information. Many researchers have been contributed their previous knowledge towards data mining in agriculture. There are many simulations models available for crop productivity predictions. As it depends on economical and environmental parameters so we can not apply these existing models or methods to any other area.

D Ramesh and B Vishnu Vardhan concluded that comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efﬁciency of the proposed technique. researchers reviewed the Data mining techniques and found out that there are several algorithms and techniques being applied in agricultural domain particularly for yield prediction based on Rainfall dataset.

M.C.S.Geetha concluded that data mining plays a crucial role for decision making on several issues related to agriculture ﬁeld. She also discusses on different DNN applications in solving the different agricultural problem.

Vrushali Bhuyar focuses on different data mining algorithms used on soil dataset to predict fertility rate. Study shows that among the classiﬁer J48 classiﬁer perform better to predict fertility index.

Raorane A.A. and Kulkarni R.V., discussed few data mining techniques in their paper. They concluded that efﬁcient technique can be developed and analysed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give good results in agriculture growth.

Georg Rub, Rudolf Kruse, Martin Schneider, and Peter Wagner mainly focused on Neural Network technique of data mining in their research work. They built and evaluated different networks and substantiated the assumption that the prediction accuracy of the networks rises once more data become available at later stages into the growing season . Georg Rub also used Different regression techniques on agricultural yield data and concluded that data mining can serve as a better reference model for yield prediction compare to decision tree. Their research study examine deal with wheat yield prediction. Neural networks are often for predicting wheat yield from cheaply-available in-season data. Once this prediction is possible, the industrial application is quite straightforward use data mining with neural networks .

Jyotshna Solanki, Prof. (Dr.) Yusuf Mulge, observed the research studies on different data mining techniques in the field of agriculture. Data mining techniques are used in agriculture for prediction of problem, disease detection, optimizing the pesticide and so on. Data mining techniques are used for disease detection, pattern recognition by using multiple applications. Data mining is close to determine the similarities between searching the precious business information from the massive information systems such as finding linked products in gigabytes of store scanner data or the mining a mountain for a vein of important dataset.

P.Revathi, Dr.M.Hemalatha, work described the remarkable of machine learning classifier in agriculture database and extracting knowledge. Consistent prediction methods are consequently required to help planners and policy makers take strategic decisions to protect national interest.

Dr. D. Ashok Kumar and N. Kannathasan focuses on different data mining techniques we can use in agriculture. Their research survey recommending that a comparison of different data mining techniques could produce an efﬁcient algorithm for soil classiﬁcation for multiple classes. The beneﬁts of a greater understanding of soils could improve productivity in farming, maintain biodiversity, reduce reliance on fertilizers and create a better integrated soil management system for both the private and public sectors

Yethiraj N G. concluded that there are a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources.

Nilesh Dumbre, Omkar Chikane and Gitesh More concluded that we can say that if perfect crop recommendations are given to farmers it will deﬁnitely help to increase the crop yield and also in building the economic status of agricultural dependent countries.

**AUTHOR & YEAR TITLE METHODOLOGY PROBLEM STATEMENT**

G.M.Nasira, Forecasting Model for Neural networks Vegetable price

N.Hemageetha Vegetable Price Using Back forecasting model

2012 Propagation Neural Network

Youvrajsinh Chauhan, Disease Prediction Classification Soil Micronutrients

Jignesh Vania, on Soil Micronutrients by J48 classification analysis prediction

2014 Analysis of Cotton

Farah Khan, Knowledge discovery Association Crop productivity

Dr. Divakar Singh, on Agricultural dataset rule mining enhancement

2016 using crop productivity

enhancement.

**CHAPTER-3**

**MODULE**

**DESCRIPTION**

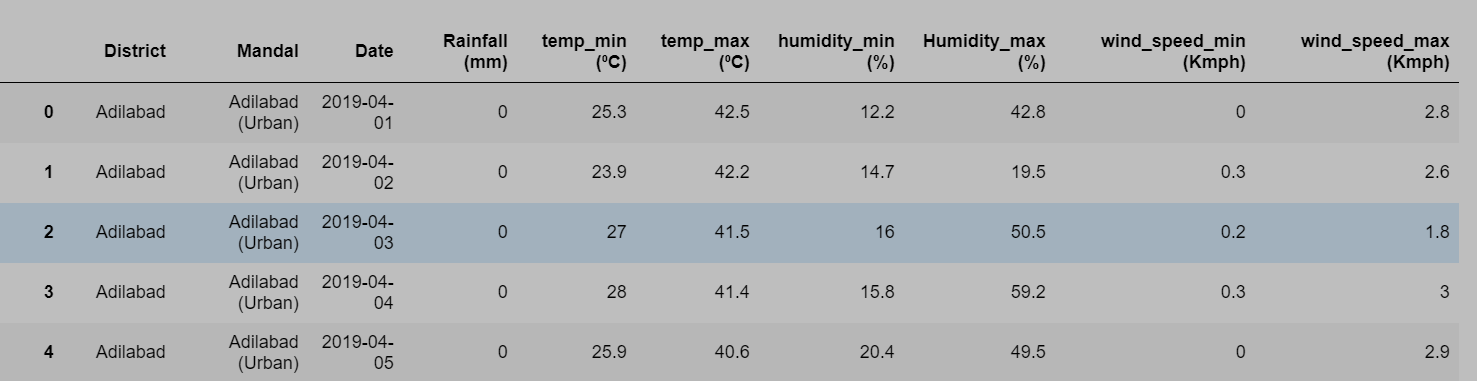
**AND IMPLEMENTATION**

**3.1 PREPROCESSING :**

Preprocessing is typically the first step in any data analysis or machine learning problem. It refers to processing the data so that it is in the desired format and can be inputted to the system. Our goal was to preprocess it into (.met) file which is supplied to the APSIM software.

Note: All of our scripts were run on jupyter notebook.

Our weather data obtained from <https://data.telangana.gov.in/> is in (.csv) format. Only two years of data (2018-2019) is available in two separate files. A few sample data points are show in the figure below



The steps involved in the preprocessing are:

* Grouping :

After reading the csv file using *pandas,* we removed the Mandal column as we are doing the analysis district wise.Then, we performed a grouping operation with district and date as the grouping criteria.

* Null values and missing data:

The next step is checking for null values. Lucking there weren’t any. But, upon searching for missing dates, we found out two of the districts Mulug and Narayanpet had around 150 missing date entries. So, we had to drop those 2 district values from the total 33 districts.

Also there was a missing day entry for all of the remaining 31 districts. So , we manipulated that day entry by averaging the values of the previous date and the next date.

* Working on columns:

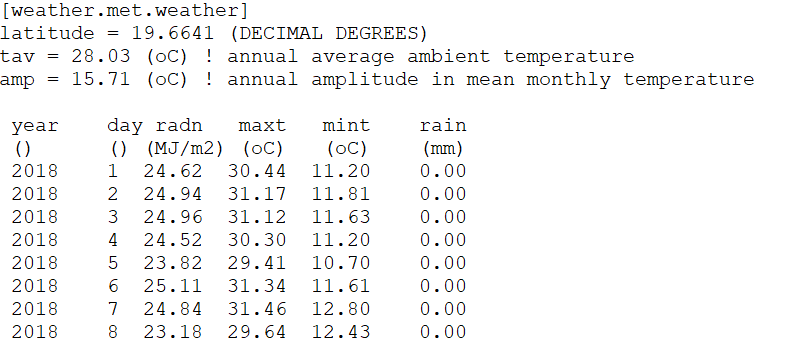
The input to the APSIM, (.met) file has a fixed syntax for columns and column names. So we added a year and day of the year columns to the dataset and renamed other columns as required for APSIM. Finally, we had to drop the unnecessary columns.

* Regressor for Solar Radiation:

One of the columns in (.met) file is solar radiation. But our dataset didn’t have that column . So, we took all the example (.met) files available in the APSIM install directory . We modelled a neural network for this problem. The NN was built as a regressor to output solar radiation values from the maximum and minimum temperature. We observed a strong correlation between daily temperatures and solar radiation values. Thus, we choose the input units as just those two (max temp, min temp).

* Determining the constants:

The (.met) also requires a minimum of three constants. The format of (.met) is show in the figure below:



There are three constants as shown in the topmost lines, latitude, tav, amp.

The latitude values for each of the districts were obtained through google search.

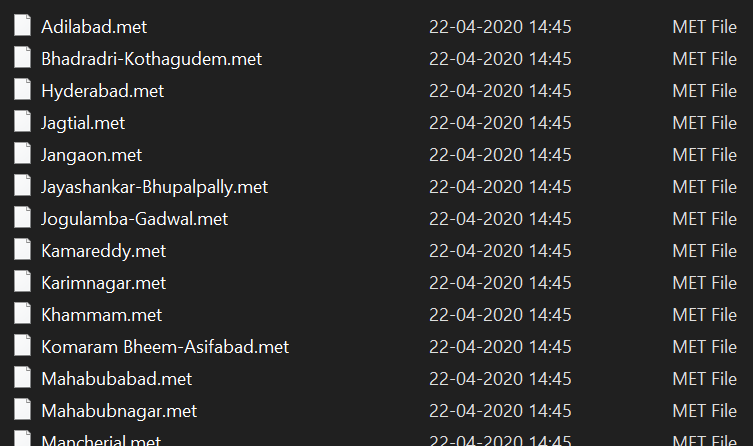
Amp is obtained by averaging the mean daily temperature of each month over the entire data period resulting in twelve mean temperatures, and then subtracting the minimum of these values from the maximum. Tav is obtained by averaging the twelve mean monthly temperatures.

So, we calculated the TAV, AMP values district wise.

* Writing to (.met) file:

Now that we have obtained all the required values , the final task was to write the values to (.met) file format.

We used simple python file input and output functions to write the (.met) files for 31 districts.



**3.2 GENERATING AND RUNNING SIMULATIONS :**

In our project, we have used APSIM tool to create and run simulations.

**ABOUT APSIM:**

The **A**gricultural **P**roduction **S**ystems s**IM**ulator (APSIM) is internationally recognised as a highly advanced platform for modelling and simulation in agriculture systems that has been developed by the Agricultural Production Systems Research Unit in Australia. It contains a suite of modules that enable the simulation of systems for a diverse range of plant, animal, soil, climate and management interactions. APSIM is undergoing continual development with new capability added to regular release of official versions. Its development and maintenance is underpinned by rigorous science and software engineering standards. The APSIM initiative has been established to promote the development and use of the science modules and infrastructure software of APSIM.

The (APSIM) is a comprehensive model developed to simulate biophysical processes in agricultural systems, particularly as it relates to the economic and ecological outcomes of management practices in the face of climate risk. It is also being used to explore options and solutions for the food security, climate change adaptation and mitigation and carbon trading problem domains. From its inception twenty years ago, APSIM has evolved into a framework containing many of the key models required to explore changes in agricultural landscapes with capability ranging from simulation of gene expression through to multi-field farms and beyond.

APSIM is structured around plant, soil and management modules. These modules include a diverse range of crops, pastures and trees, soil processes including water balance, N and P transformations, soil pH, erosion and a full range of management controls. **APSIM resulted from a need** for tools that provided accurate predictions of crop production in relation to climate, genotype, soil and management factor while addressing the long-term resource management issues.

**Details of APSIM components:**

APSIM contains an array of modules for simulating growth, development and yield of crops, and their interaction with the soil.

Currently crop modules are available for barley, canola, chickpea, cotton, cowpea, hemp, fababean, lupin, maize, millet, mucuna, mungbean, navybean, peanut, pigeonpea, sorghum, soybean, sunflower, wheat and sugarcane. In addition there are general modules for forest, pasture as well as specific implementations for the pasture species lucerne and stylo.

The plant modules simulate key underpinning physiological processes and operate on a daily time step in response to input daily weather data, soil characteristics and crop management actions. The crop modules have evolved from early versions for focus crops such as maize , peanut, sorghum and sunflower. The initial crop modules of APSIM utilised concepts from existing models available at the time and added concept enhancements from local research to improve existing models as required.

Currently in APSIM, all plant species use the same physiological principles to capture resources and use these resources to grow.

The APSIM **SOILpH module** provides a representation of the acidification of soil, and how pH changes are distributed through the profile, as a consequence of the imbalance in uptake of cations and anions, the leaching of nitrate, and changes in soil organic matter content. It is a tool that can be used for exploring strategies for reducing the effect and for examining the effectiveness of remedial actions calcium). Besides influencing plant production, soil pH also affects the turnover of soil organic matter. Soil processes such as mineralisation, nitrification and urea hydrolysis are pH dependent. Whilst the SOILN module does include routines to represent the effects of pH on the dynamics of soil C and N, its ability to capture the consequences of soil acidification on N mineralisation or C balance has not been studied

**Data requirements:**

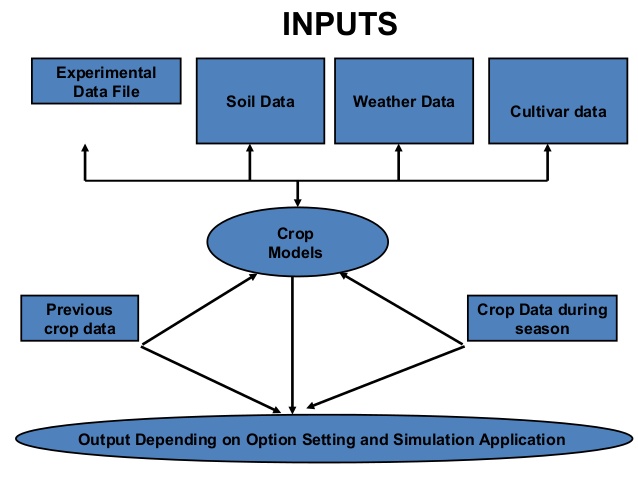
An APSIM simulation is configured by specifying the modules to be used in the simulation and the data sets required by those modules. APSIM modules typically require initialisation data and temporal data as the simulation proceeds. Initialisation data is usually categorised into generic data (which defines the module for all simulations) and simulation specific parameter data such as site, cultivar and management characteristics. Typical site parameters are soil characteristics for soil modules, climate measurements for meteorological modules, soil surface characteristics and surface residue definition. Management is specified using a simple language to define a set of rules, calculations and messages to modules that are used during the simulation.

**Use of apism:**

The APSIM environment is an effective tool for analyzing whole-farm systems, including crop and pasture sequences and rotations, and for considering strategic and tactical planning. APSIM allows users to improve understanding of the impact of climate, soil types, and management on crop and pasture production. It is a powerful tool for exploring agronomic adaptations such as changes in planting dates, cultivar types, fertilizer/irrigation management, etc.

**Key input:** Soil properties, daily climate data, cultivar characteristics, and agronomic management.

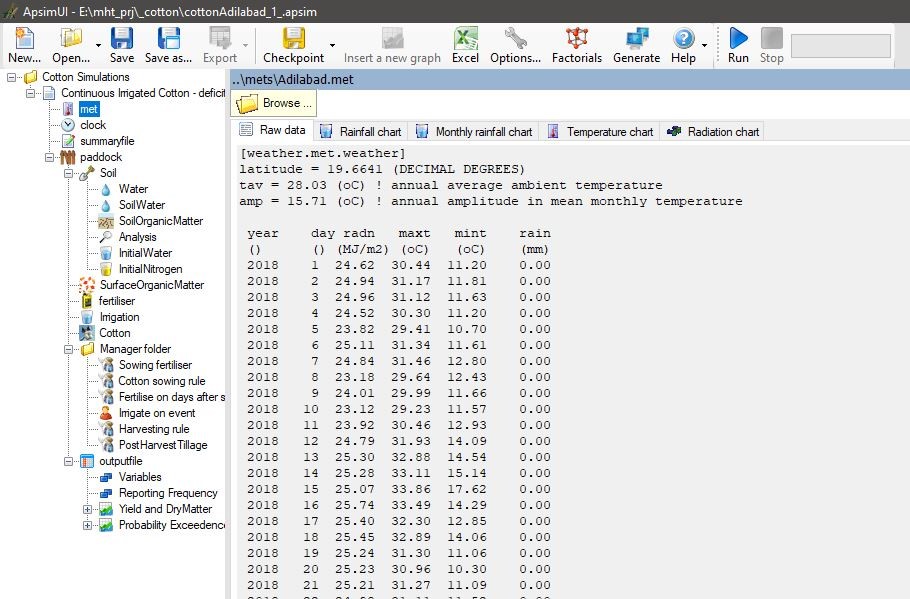
**Key output:** For trained agronomists. Requires advanced knowledge of plant growth and soil processes.



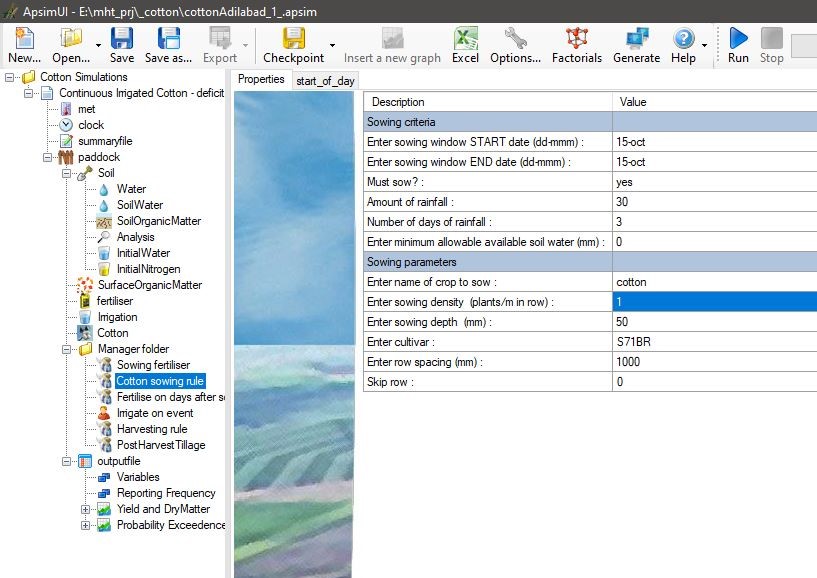
**How we used APSIM:**

After downloading and installing the setup and learning about how to use it, the first step is to load the already available cotton simulation in which we can change make changes to the parameters as required .

After loading the base simulation file , we browse the and select our meteorological (.met) file.



As shown in the snapshot above, on the left side, there are a list of changeable parameters. From soil to fertilizer to sow date to plant density, we can set any of those values as desired .The tool has a huge set of changeable parameters that crop growth depends on. Even though these parameters are not all the factors plant growth completely depends on, but these are the best factors one can get hands on away from the agricultural field. Say we want to change the plant density under cotton sowing rule, then select that option , then the window changes to:



After setting the desired parameters, the final job is to run the simulation by clicking on the Run option in the toolbar. APSIM takes (.apsim) file as input and upon running the simulation, it produces one summary(.summ) and one output (.out) file for per simulation.

We need the yield parameter from the generated (.out) file.

**3.3 AUTOMATING THE SIMULATION PROCESS:**

As we are working with 31 districts, we can’t go around modifying the simulation for each and every district with different configurations manually. So, we need to somehow automate this like by using for loop over the districts and different configurations.

The (.apsim) file is in xml format. So we needed a XML parser that parses the simulation files and modifies its content. We chose the python built-in module XML.ElementTree to do the parsing. After parsing the file, we can update the contents by finding the attribute and setting its value.

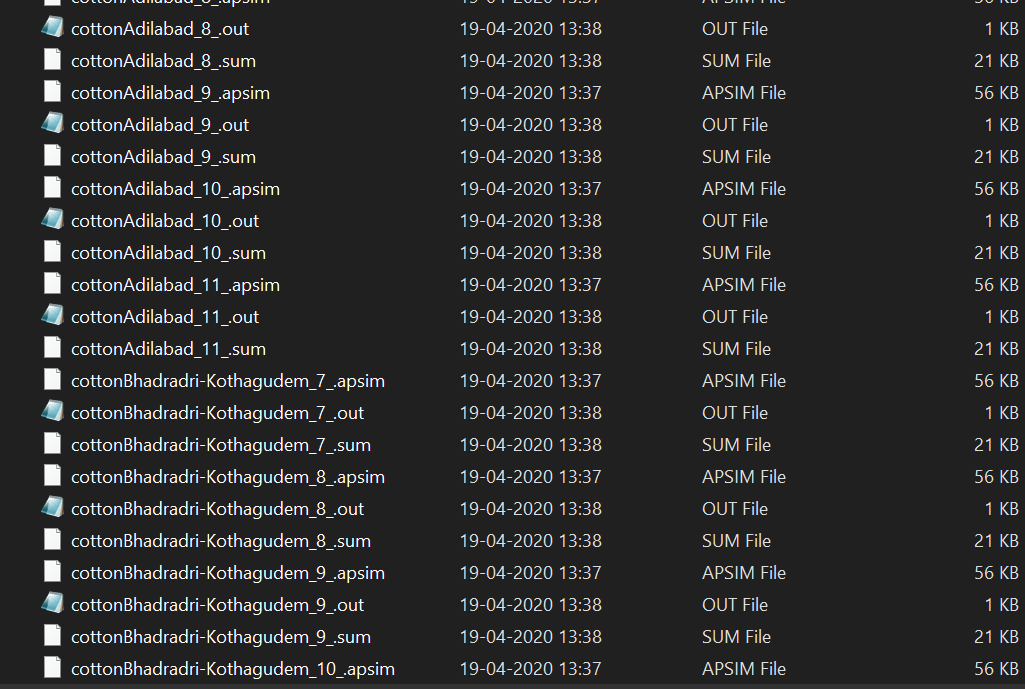
We chose to work in Microsoft Visual Code IDE so that we can open the base simulation cotton crop (.apsim) file and python scripts side by side. So, we first check the parameter we want to update in APSIM software, lookup that element in the simulation file in Visual Code and then in the python script find that parameter and set its new value. The (.met) file are also added the same way.

We repeat the procedure for all the parameters we want to modify and also loop over the 31 districts .File write method was used inside the loop over configurations and districts so in each iteration, a new (.apsim) file is created with different values. We change the name of the simulation as the district name and add the values of the parameters to the name so that it will come in handy when interpreting the yield from (.out) files.

Thus, we have automatically generated desired number of simulations. The next task was to run them all to produce the output (.out) files. Another python script was written to automate the process of running all of the generated simulations.

When the APSIM path is added to the Path variable in Environment variables, it is possible to run all the (.apsim) files in the current directory through command prompt using the command ***apsim \*.apsim***

After running that script , the directory was as shown in the figure below:



The numerical value represents one of the parameters we used for configurations which is Plant density.

**About Cotton crop:**

Cotton is the main cash crop that highly grown in Telangana region, most of the people used grow this crop to get huge profits.

**Conditions of Growth:**

Cotton is the crop of tropical and sub-tropical areas and requires uniformly high temperature varying between 21°C and 30°C. The growth of cotton is retarded when the temperature falls below 20°C.

The modest requirement of water can be met by an average annual rainfall of 50- 100 cm. However, it is successfully grown in areas of lesser rainfall with the help of irrigation. About one-third of the total area under cotton cultivation is irrigated.

Cotton is a kharif crop which requires 6 to 8 months to mature. Its time of sowing and harvesting differs in different parts of the country depending upon the climatic conditions.

In Telangana, it is grown both as a kharif and as a rabi crop.

Here the rainfall occurs after September and cotton is sown in October. The irrigated crop is sown in January-February. Most of the crop is grown mixed with other kharif crops such as maize, jowar, ragi, sesamum, castor, groundnut and some vegetables.

**Cotton cultivation in Telangana:**

Cultivating cotton requires deep pockets but promises high returns. However, the cash crop comes with a very high rate of failure.

Despite the alarming rate of farmer suicides, cotton cultivation has witnessed an exponential growth in the state. Telangana is the third highest in terms of cotton cultivation and production in India, after Gujarat and Maharashtra.

In 2014-15, Telangana cultivated 16.93 lakh hectares of cotton, and produced 35.83 lakh bales. The following year, the state’s cultivation increased to 17.78 lakh hectares and production jumped to 37.33 lakh bales in 2015-16.

Cultivation, however, fell by over 20% in 2016-17, with fluctuating prices in both the domestic and international market pegged as the reason. The Centre had also removed export subsidies on cotton, which came as a huge setback for farmers.

Telangana had recorded a 17% rainfall deficit (699 mm against the normal of 845 mm) in 2018. According to a study on the Telangana’s groundwater for the period of[June to December 2018](http://gwd.telangana.gov.in/open_record_view.php?ID=156), the state’s average water level stood at 11.15 m bgl (below ground level) as compared to 9.05 m bgl during the same period in 2017, a decline of 2.10 m. In around 9% of the state’s geographical area, the ground water level is deeper than 20 m bgl. This includes districts of Siddipet, Medak, Sangareddy, Kamareddy, Rangareddy, Nagarkurnool and Vikarabad districts which received less rainfall.

**Measures to cultivate cotton:**

Cotton area has been steady in Telangana state where it is occupied 17.13 lakh ha with a productivity of 1039 kg ha (Season and crop report, Agriculture action plan 2015-16).

In the sowing line, we sow the seeds leaving an average distance of 3 inches (7.62 cm) between them. On average, we have around 90.000-100.000 plants per hectare. Consequently, the average number of cotton plants per acre is 40.000

However, there are cases in which we can have up to 70.000 cotton plants per acre, but this can be achieved only by selecting specific varieties.

Plough the field 3 or 4 times to fine tilth. Spread 12.5 tons of FYM or compost per ha uniformity before the last plough and level the field perfectly.

Sow the seed at 3 cm depth.

Generally kapas is sown behind the plough in ridges. In case of sowing by tractor, distance between two rows should be 67.5 cm and plant to plant 30 cm where as in case of indigenous plough, the distance between two rows should be 70.0 cm and plant to plant 30 cm. In the land where water level is high or soil/water is saline or there is problem of water logging, sowing on bund is useful. For this, prepare 20-25 cm high bund and sow on 2/3 rd part of bund from lower surface at definite distance by hand harrow (Khurpa). Use 4-5 seed per place.

Use of fertilizer on the basis of soil test is useful. In case of use of Dap and Urea, use 67 kg DAP and 105 kg urea. Or in case of use of super phosphate+ urea, use 188 kg single super phosphate and 130 kg urea.

If 27 kg/acre DAP is used, reduce the quantity of urea by 10 kg.

In Telangana, it is grown in kharif as well as rabi season, cotton crop take 6 to 8 months to mature. Most of the people depend on climate conditions they harvest cotton crop in October month.

We have taken most of these measures into account and set the parameters of APSIM accordingly.

**3.4 BUILDING AN ARTIFICIAL NEURAL NETWORK:**

As we had already described the working of an artificial neural network in detail in the previous project , so we will just focus on building an ANN model for regression.

**Input:**

The input for our ANN is the parameters set in the APSIM (plant density, fertilizer etc), district , location of the district (Latitude and Longitude).

**Output:**

The output is the yield in kilogram/hectare.

**Packages and functions used:**

* Numpy for arrays and simple mathematical operation on arrays
* Pandas for reading csv files and creating dataframes and dataframe operations
* train\_test\_split from sklearn to split the data into training and testing data points,
* KerasRegressor from keras.wrappers.scikit\_learn to perform grid search on hyperparameters. It wraps the NN as regressor so that grid search can be used.
* GridSearchCV from sklearn to perform grid search
* Keras for building ANN model
* matplotlib and seaborn for plotting graphs
* glob to access files and folders on local system
* os to change the directory for file input/output

The first task is the create an input dataframe. Some of the input parameters can be retrieved from the names of the simulations outputs (we intentionally set their names that way for this purpose). This can be done using glob. Glob function creates an object of contents present in the specified directory. We loop over all the (.out) files and simultaneously add the retrieved parameters to the dataframe. Other parameters (latitude, longitude) can be supplied by looking up the district name and using the previously created dataframe for (.met) constants.

The output parameter (yield) is obtained by using file read operations in the same loop mentioned above.

Generally, ANN’s are used for classification tasks but they can also be used for regression when either the output unit activation is set as linear or using sigmoid activation if the output unit is normalized to range [0,1]

We chose to use sigmoid activation and normalizing the output (yield) value to [0,1] because that gave better results. For normalizing, we just divided the output series with the maximum values so that its range became [0,1]

Then the data is split into training and testing data. A base sequential model with just two hidden layers with five units each and optimizer as Adam optimizer, 10 epochs and mini-batch size as 32, learning rate as 0.1 was defined

**Hyperparameter tuning:**

The best performance for any ANN is achieved when its hyperparameters are tuned correctly. We chose the following set of hyperparameters to be varied:

batch\_size = [ 16 , 32 , 64 , 128 ]

epochs = [ 1, 10 , 50 , 100 ]

optimizer= [ 'rms' , ’adam’ , ’sgd’ ]

init\_mode= [ 'normal' , 'uniform' ]

dropout\_rate= [ 0.1, 0.2 , 0.3 , 0 ]

learn\_rate= [ 0.001 , 0.01 , 0.1 , 1 ]

KerasRegressor enables us to wrap the NN as sickit learn regressor so that we can perform grid search using GridSearchCV .

We run this on jupyter notebook on local system. But sicne we started the grid search , we waited for 2hrs but there were so many combinations of hyperparameters that it would finish the run.

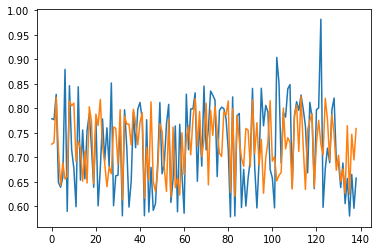
So we had to switch to GOOGLE COLAB and upload all the data to run and run the notebook there. GOOGLE COLAB offers high processing and GPU speeds so it didn.t take long to finish the run.

The grid search gave the best fit as :

optimizer='adam' , init\_mode='uniform' , dropout\_rate=0.0 , learn\_rate=0.001, epochs = 20, batch\_size = 32

**Testing Results:**

We achieved an mean\_absolute\_error of 0.05 on test data and comparison of actual test output and predicted output can be visualized through the plot below:



The blue lines represent actual values and orange ones represent the model predicted values. As we can see the model fits the data pretty well except for few outliers.

Also we have chosen only a handful of parameters to vary , thus our dataset is small and there are chances of overfitting.

**CHAPTER-4**

**CONCLUSION AND FUTURE WORK**

**4.1 CONCLUSION:**

We conclude that we have modelled crop yield prediction to some extent because we had chosen only a handful of parameters to vary. Our metric score was pretty decent and given the cotton crop our, model can predict which district might produce higher yields per given area for given set of parameters.

Most of the time during the project was spent on preprocessing and learning the APSIM tool . The project made us realize how complex is the natural process of crop growth. We had no prior knowledge in the agriculture sector and had to spend much time on finding the suitable values for the dependent parameters. Initially , unavailability of field data was a huge setback and this project wouldn’t have been possible if we didn’t come across APSIM tool.

**4.2 FUTURE WORK:**

We desire to do further research in the agriculture sector and crop growth, so that we can work with more values and configurations of other dependent factors. Incorporating more of the factors that yield depends on requires deep study of crop growth and field work knowledge, but it is doable.

One of the ideas for future work we have is to incorporate image data (photographs of crops taken at regular time intervals throughout the different stages of crop growth) and add Convolutional Neural Networks to the project for image data. This will require farmers to be equipped with imaging tools and camera equipment.

We are also interested in Vertical farming in which plant are grown indoors without soil and without sunlight. It is truly the future of agriculture when the world is getting rapidly urbanized. Understanding the crop growth and soil components better helps get better results in artificial techniques such as vertical farming.

When our project is applied at individual field level instead of district level which we have done for now, we can provide accurate parameters and also create more and more simulations and train deeper networks which will be able to better model the natural process.

**REFERENCES:**

1. <https://www.mssanz.org.au/modsim2015/B1/fainges.pdf>
2. <https://www.sciencedirect.com/science/article/abs/pii/S0034425715001637>
3. <https://www.sciencedirect.com/science/article/pii/S0378429016308759>
4. <https://www.coursera.org/learn/deep-neural-network?specialization=deep-learning>