Rain prediction and crop plantation for maximum yield

Vignesh P [a], Pravin K [b]

College of Engineering Guindy
Sardar Patel Rd , Chennai 600025

antopravin2001@gmail.com

apsvignesh2000@gmail.com

ABSTRACT:

Agriculture is one of the major sectors in Indian economy. Two-third of population is dependent on agriculture directly or indirectly. Annual rainfall plays a major role in cultivation of crops. Due to the unpredictable climatic changes, farmers are struggling to obtain a good amount of yield from the crops. Thus predicting rainfall could be really useful in case of getting maximum yield. All parts of India receive rainfall in different time of the year. Mainly the monsoon in India is classified into south-west and north east monsoon. Our task is to predict rainfall in a particular area in upcoming years. Then with the rainfall suggest better crop for plantation to get maximum yield.

The rainfall prediction is a time series prediction. So a model named SARIMAX is used to predict rainfall by reading the previously recorded rainfall at a particular place. The forecast of rainfall will be really helpful in predicting the crop production since the yield of crop is dependent on rainfall for some extent. SVR (support vector regression) model is used to predict the production of the crop based on the previously predicted monsoon rainfall. This work will ensure that the farmers get the maximum profit by selecting the right choice of crop for cultivation.

1. INTRODUCTION:

Agriculture is said to be the backbone of Indian economy. Over70% of the population is involved in agriculture and related fields. The forecast for agriculture yield is an important but also a tedious job in many countries. Due to changing climatic conditions the prediction of rainfall seems to be really difficult. Farmers are struggling to get a profitable production these days due to the unpredictable weather. Another major problem is diminishing cultivable land due to the raising population. The traditional techniques and level of infrastructure in the county also affect the crop production rate.

India has three cropping seasons — Rabi, Kharif and Zaid. We mainly focus on the kharif crops. Kharif crops are grown with the onset of monsoon in different parts of the country and these are harvested in September-October. Important crops grown during this season are paddy, maize, Jowar, bajra, tur (arhar), moong, urad, cotton, jute, groundnut and soyabean. Some of the most important rice-growing regions are Assam, West Bengal, coastal regions of Odisha, Andhra Pradesh, Telangana, Tamil Nadu, Kerala and Maharashtra, particularly the (Konkan coast) along with Uttar Pradesh and Bihar. The southwest monsoon from June to July plays an important role in crop production rate.

Since monsoon plays a vital role in Kharif crop cultivation we can use machine learning techniques to forecast monsoon rainfall. By predicting the monsoon rainfall we can also predict the crop production rate based on the monsoon rainfall for that year. We use a model named SARIMAX (Seasonal Auto Regressive Integrated Moving Average Exogenous model). The model predicts the future rainfall on analyzing the previous history of rainfall. Then the forecasted rainfall is used to predict the crop production. We use SVR (support vector regression) for predicting the crop production.

2. RELATED WORKS:

In this section some previously done research on rain prediction and crop plantation is discussed. Heuristic Prediction of Rainfall Using Machine Learning Techniques et.al [1] is published by Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna. They used linear regression analysis for predicting the unknown value of a season from the known value of another season. This linear regression method suggests the lower correlation between the various crop seasons and data results which we were are predicted are solely done based on the previous year's data. They faced a major drawback, prediction of particular season is dependent on other seasons, so the data are limited, which makes it less accurate, it can be further enhanced. “Rainfall Prediction: A Deep Learning Approach” published by

Emilcy Hern´andez, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and N´estor Duque. This paper has presented a deep learning approach based on the use of auto-encoders and neural networks to predict the accumulated precipitation for the next day. The results suggest that the proposed architecture outperform other approaches in terms of the MSE and the RMSE**.** But it failed predicting the rainfall for heavy rain scenario, this is not well modeled for heavy rain scenario, and also not well for seasonal rainfall prediction. Prediction of Rainfall Using Machine Learning Techniques et.al [3] published by Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi. This predictive model is used to prediction of the precipitation. The first step is converting data in to the correct format to conduct experiments then make a good analysis of data and observe variation in the patterns of rainfall. Estimation of rainfall and it is estimated that SVR is a valuable and adaptable strategy, helping the client to manage the impediments relating to distributional properties of fundamental factors, geometry of the information and the normal issue of model over fitting. But they concentrated more in short term rainfall, low accuracy in long term rainfall falls in wrong prediction of crop production. A Study on Crop Yield Forecasting Using Classification Techniques et.al [4] published by R.Sujatha, Dr.P.Isakki. This paper has presented lots of machine learning methods like Rule Based Classifiers, Bayesian Networks, Nearest Neighbor, Support Vector Machine, Decision Tree, Artificial Neural Network, Rough Sets, Fuzzy Logic, and Genetic Algorithms. Using these techniques, the crop yield can be improvised and increase the income level of the farmer, will be increased.

3. SYSTEM DESIGN:

First the IMD rainfall dataset is collected from official IMD website. Then the data undergoes a preprocessing stage where the data in gridded format is converted into CSV format. The data for the required place is collected with the help of matching the coordinates. On the next step the preprocessed data is fed into the SARIMAX model where the forecast of rainfall is done. At this stage we are able to predict the rainfall for the particular region. Then the previously preprocessed rainfall data is added to the crop dataset. The features in the combined dataset is altered so that the predictions could become more accurate. Then a support vector regression is performed over the crop production dataset. With the help of previously predicted monsoon rainfall the production rate of crops can be calculated for the upcoming years.

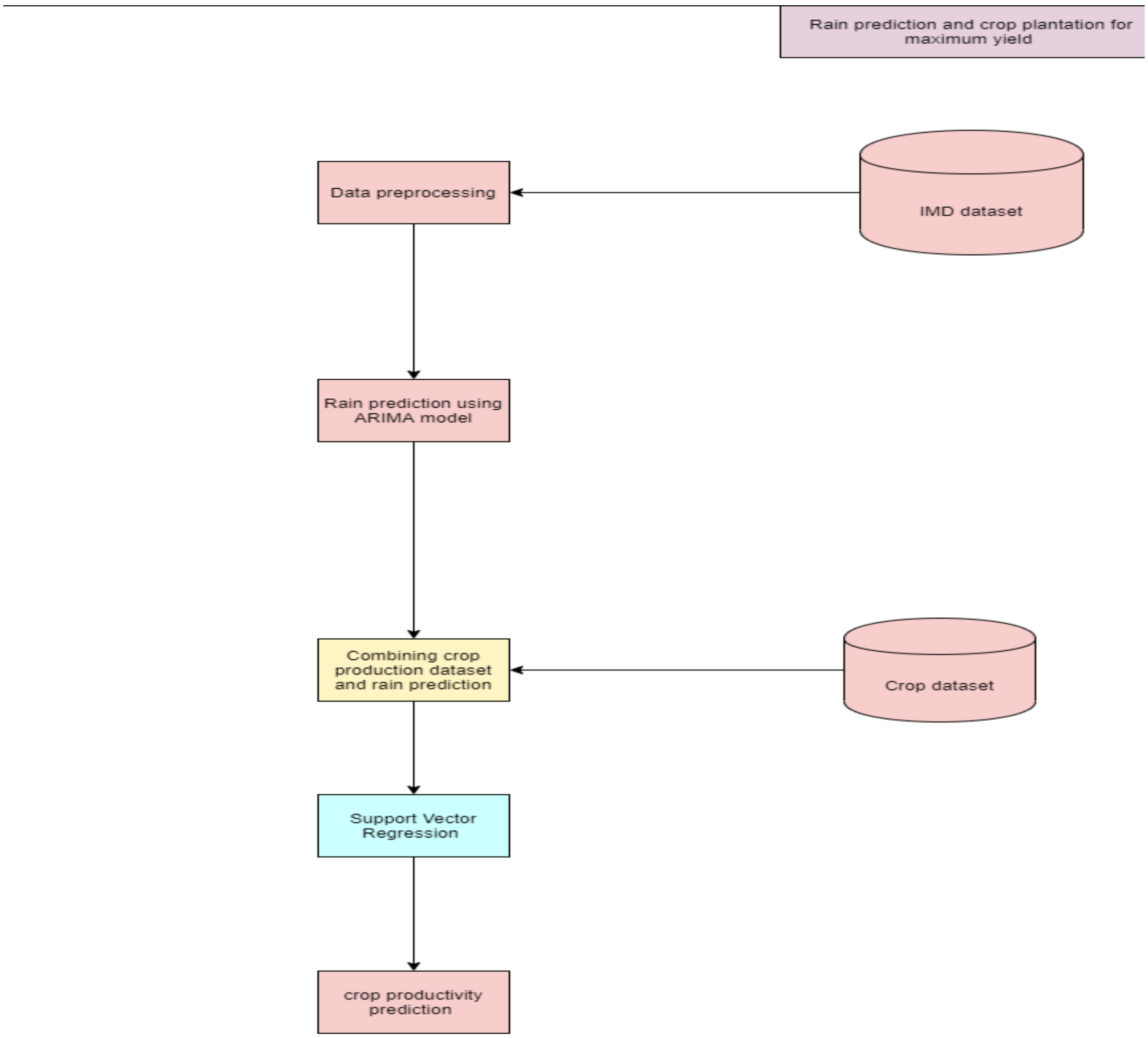

Fig: 7.1 architecture diagram

4. MODULE SPLIT-UP:

4.1 Module 1: Data preprocessing

The data is collected from the official IMD dataset. The data consist of rainfall information measured in mm for the past 100 years in grid format. From grid format the data is converted to CSV format. The unrecorded data are neglected. Then the rain details for particular place is gathered over a period. Then our main aim is to pair the data according to the period they belong to. (Eg: the rainfall data of June month is grouped together).

4.2 Module 2: Rain prediction using SARIMAX model:

Autoregressive Integrated Moving Average, or ARIMA, is one of the most widely used forecasting methods for univariate time series data forecasting. Although the method can handle data with a trend, it does not support time series with a seasonal component. An extension to ARIMA that supports the direct modeling of the seasonal component of the series is called SARIMA. The pre processed data is sent to the SARIMAX model which is a model for time series prediction. The model gets trained with the input data and able to predict annual rainfall for the upcoming years.

4.3 Module 3: Combining crop production dataset and rain prediction

In this module the data for crop production is collected. Then the crop productivity data is combined with the predicted rainfall. The crop productivity data for each year is combined with the rainfall data of the year. This combination of data helps in predicting the changes in crop production accordingly with respect to rainfall amount.

4.4 Module 4: Crop productivity prediction using SVR

Since this problem falls under regression category we use Support vector regression (SVR). Support vector regression (SVR) and support vector machine (SVM) use the same concepts with a few dissimilarity. For solving regression problems, we can extend the methodology used by support vector classification. Constructing a support vector classification model depends only on a subset of training data because the cost function for constructing the support vector classification model does not consider training points beyond a certain boundary. Similarly, a subset of data is required for building a support vector regression model.

Support vector regression formulas:

$$y_i - wx - b \leq \varepsilon + \xi_i$$
$$wx_i + b - y_i \leq \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \geq 0$$

Where x is inputs, w – weights, ξ_i is the slack variable and epsilon is the distance between hyper planes.

$$y = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*)(x_n \cdot x) + b$$

where α_n, α_n^* are nonnegative multipliers with for each observation x_n and $x_n.x$ is the dot product.

5: EXPERIMENTAL RESULTS

5.1 EVALUATION PARAMETERS :

Mean square error:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values

5.2 Dataset:

5.2.1 : IMD dataset (rain fall dataset):

IMD New High Spatial Resolution (0.25X0.25 degree) Long Period (1901-2018) Daily Gridded Rainfall Data Set over India. This data product is a very high spatial resolution daily gridded rainfall data (0.25 x 0.25 degree). The unit of rainfall is in millimeter (mm). Data available for 118 years, from 1901 to 2018. Data is arranged in 135x129 grid points. The first data in the record is at 6.5N & 66.5E, the second is at 6.5N & 66.75E, and so on. The last data record corresponds to 38.5N & 100.0E. The yearly data file consists of 365/366 records corresponding to non-leap/ leap years.

5.2.2 : Crop production dataset:

The dataset consist of crop production in each state for variety of crops. The year wise data is available for crop production. The dataset consist of production in tones, area in hectares crop variety and place were crops are cultivated. The dataset is downloaded from an official government website.

6: Final result:

6.1: Predicting rainfall

The forecasted data is stored in a data frame. Here is the predicted rainfall of Vellore measured in mm for the year 2022.

	monsoon	forecast
1925-01-01	383.063445	NaN
1926-01-01	490.050053	NaN
1927-01-01	609.868194	NaN
1928-01-01	459.301587	NaN
1929-01-01	331.977953	NaN
...	...	...
2046-01-01	NaN	547.809080
2047-01-01	NaN	566.133905
2048-01-01	NaN	636.634798
2049-01-01	NaN	462.746650
2050-01-01	NaN	462.195417

126 rows × 2 columns

```
In [80]: print('predicted rainfall for 2022:')
         mon = future_df['forecast'][96]
         print(mon)

predicted rainfall for 2022:
522.597061235427

In [81]: crop = pd.read_csv('Tamilnadu agriculture yield data.csv')
```

Predicted rainfall: 522.59 mm rainfall in Vellore (2022)

6.2: Predicting yield of different crops:

From the predictions of each crop we can get the production rate for particular amount of monsoon rainfall. Since we previously predicted the monsoon rainfall of the next year 2022, we could get the production rate for the crops in 2022. The below diagram shows the production of each crop for the next year. The best crop is chosen accordingly based on their market price. For example if rice is sold at a rate of 35000 per Tone then cost price of the crop can be calculated as (3.6380*35000 = 127330) per hectare of cultivation.

```
In [211]: print('rice(tone/hectare):')
          print(rice)
          print('Maize(tone/hectare):')
          print(maize)
          print('groundnut(tone/hectare):')
          print(gn)

rice(tone/hectare):
3.638082450574147
Maize(tone/hectare):
4.439634543137882
groundnut(tone/hectare):
2.2587454066051627
```

7: RESULT ANALYSIS:

SVR have hyper parameters C, and epsilon. C is Regularization parameter. The strength of the regularization is inversely proportional to C. Must be strictly positive. The penalty is a squared l2 penalty. Epsilon specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value. Rice crop is taken for analyzing the results.

7.1 Analysis by changing parameters:

Changing values for C

C	epsilon	Mean Squared Error
0.5	0.5	0.33189
1	0.5	0.33435
5	0.5	0.31940
10	0.5	0.32270

Table 7.1

Changing values for epsilon

C	Epsilon	Mean Squared Error
5	0.25	0.31432
5	0.5	0.31940
5	1	0.45828
5	0.15	0.29792

Table 7.2

By analyzing the tables 7.1 and 7.2 the best value for parameters C, epsilon is 5 and 0.15 respectively. For these values of parameters the model produces least mean squared error.

7.2 Comparing with other models:

In this section the SVR model's performance is compared with the performance of Linear regression and Random forest Regression.

CROP	Support vector Regression (Mean square error)	Linear regression (Mean square error)	Random forest regression (Mean square error)
Rice	0.19443	0.21554	0.25204
Maize	2.56997	3.68447	2.78506
Groundnut	0.22268	0.22770	0.46365

Table 7.3

Table 7.3 clearly shows that the SVR model has less mean squared error when compared with other models. Therefore SVR model predicts with more accuracy when compared with other models.

Fig 7.1 shows that SVR provides a better flexibility in prediction than the linear regression model.

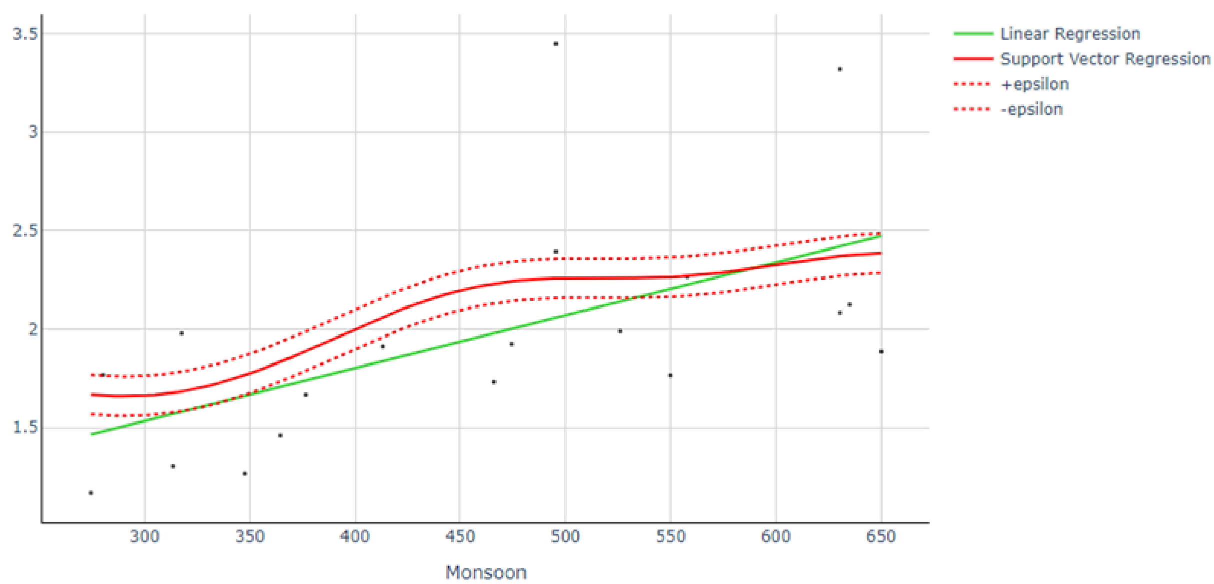

Fig 7.1: Graph of SVR VS Linear regression

8: CONCLUSION:

Our project aims to maximize crop production level based on seasonal rainfall. It helps the farmers or agriculture workers such that they can do agriculture more smartly in a much better calculated way. By predicting rainfall, we did feature selection in order to select only important features in predicting various crops. Since the model is build using the climatic parameters, only the values of weather attributes like rainfall need to be passed, where the model can predict the correct output. The decision of bit capacity is basic for SVR displaying. We see that SVR is better than other regression methods as an expectation strategy. Other regression methods like MLR can't catch the non-linearity in a data set and SVR winds up helpful in such circumstances. SVR proves to be the best model for predicting crop production.

9. LIST OF REFERENCES

1. Heuristic Prediction of Rainfall Using Machine Learning Techniques – by Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna
2. Rainfall Prediction: A Deep Learning Approach – by Emilcy Hern´andez, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and N´estor Duque.
3. Prediction Of Rainfall Using Machine Learning Techniques – by Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi.
4. A Study on Crop Yield Forecasting Using Classification Techniques by: R.Sujatha, Dr.P.Isakki.
5. Data mining for meteorological applications: Decision trees for modeling rainfall prediction – by Geetha, A., and G. M. Nasira.
6. Machine learning techniques for rainfall prediction: A review – by Parmar, Aakash, Kinjal Mistree, and Mithila Sompura.
7. Rainfall prediction for the Kerala state of India using artificial intelligence approaches – by Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi.
8. Application of vegetation indices for agricultural crop yield prediction using neural network techniques – by Panda, S. S., Ames, D. P., & Panigrahi, S.
9. Climate change impact assessment: The role of climate extremes in crop yield simulation – by Moriondo, M., Giannakopoulos, C., & Bindi, M.
10. Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain – by Mo, X., Liu, S., Lin, Z., Xu, Y., Xiang, Y., & McVicar, T. R.