**Predicting crop productivity through machine learning algorithms**

Mustafa Kamal Shahadat, Dr. Md. Khairul Alam, Arindom Biswas, Dr. Mohammad Mukhlesur Rahman

**Justification:**

Ensuring food security for over populated country like Bangladesh is crucial. Prediction of crop productivity play a pivotal role in deciding crop production targets and making policy decisions. However, predicting crop productivity is not straight forward since crop productivity is a function of many factors. Among them soil fertility, physiography, crop management, weather parameters etc. are major factors. However, these factors are highly variable both spatially and temporally. A number of machine learning algorithms have been evaluated by many scientists for predicting soil fertility and crop production. The present study focuses on evaluation on various machine learning algorithms for predicting yield and crop productivity of Khulna region of Bangladesh.

Objective:

* Prediction of crop production scenario based upon environmental features.

**Methodology**

In the proposed study we will use different machine learning algorithms for predicting crop yield and total productivity of Khulna region of Bangladesh. We will use several supervised machine learning algorithms and finally will compare these algorithms to achieve most accurate prediction (Figure 1).

The whole activities are as follows:

*Step 1: Data collection and about data set*

We have been collected historical crop yield and production data from secondary sources. Crop production data were collected from regional department of agricultural extension and daily weather data were collected from Bangladesh meteorological department. The dataset contains crop specific yield and production according to locations from 1995 to 2018. The dataset also contains climatic features related to crop production. The climatic factors which influenced crop production include daily maximum and minimum temperature, rainfall, humidity and sunshine hours. There are 12 major crops grown in different rotations of the respective locations. The crops are rice, wheat, maize, jute, potato, onion, lentil, mungbean, blackgram, grasspea, mustard and sugarcane.

*Step 2: Data pre-processing and exploratory data analysis*

Both the crop production and climatic data will be cleaned and merged into one excel file. Missing values will be corrected through either dropping or suitable imputation method. Necessary data transformation, normalization, encoding (level or one-hot encoding), summary, correlation matrix, visualization, features selection etc. steps will be followed. Finally, the dataset will be spilted into training and test set, where training set will cover 80% of the data and 20% will be used for testing.

*Step 3: Model selection and algorithm setting*

The following machine learning models will be used in our study to predict crop yield and production.

* Multiple regression
* Decision tree
* Random forest
* Support vector machine
* Gradient boosting

These models will be compared at the end of the final run. For model tuning necessary hyperparameters will be optimized through grid search.

*Step 4: Evaluation of the model*

The models will be compared with model accuracy metrices viz. mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), R-squared or coefficient of determination.

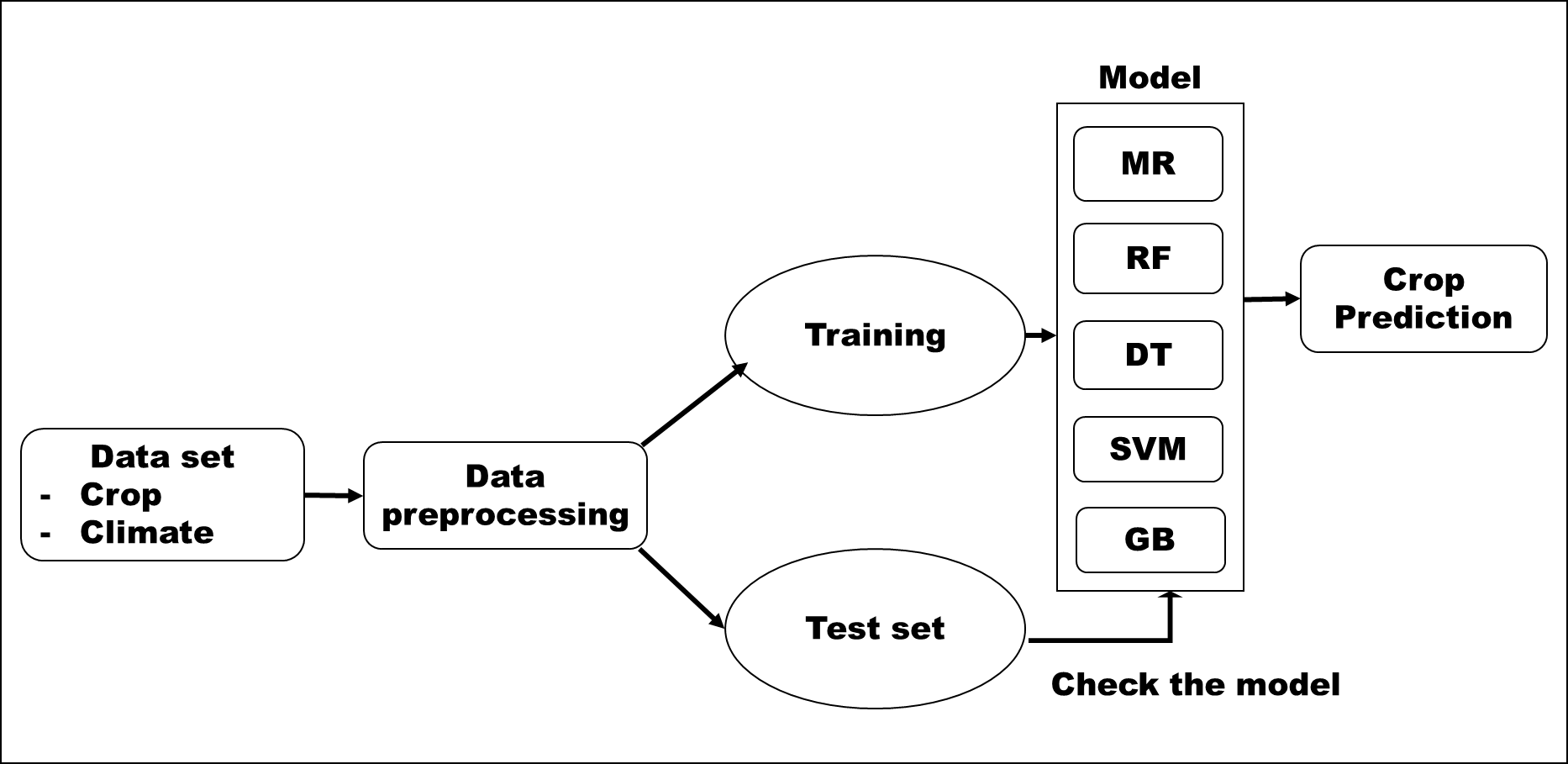


Fig. 1 Flow chart of the model for predicting crop yield and production

Expected output

From the proposed work, we will be able to predict crop yield of different crops and productivity of different districts of Khulna region of Bangladesh. The output of the result could be used for deciding crop production targets and will be helpful for decision makers for allocating necessary budgets in the agriculture sector.

Time Table

|  |  |
| --- | --- |
| Date | Activities |
| 29 Oct – 3 Nov, 2021 | Data collection and preparation of synopsis |
| 4 Nov, 2021 | Discussion with group members on overall workplan |
| 5 Nov-7 Nov, 2021 | Data cleaning and preprocessing |
| 8 Nov-9 Nov, 2021 | Model development and evaluation |
| 10 Nov, 2021 | Rerun the codes, organizing the notebook and writing conclusion |
| 11 Nov, 2021 | Group meeting and submission |