***PREDICTIVE AGRICULTURAL ANALYTICS***

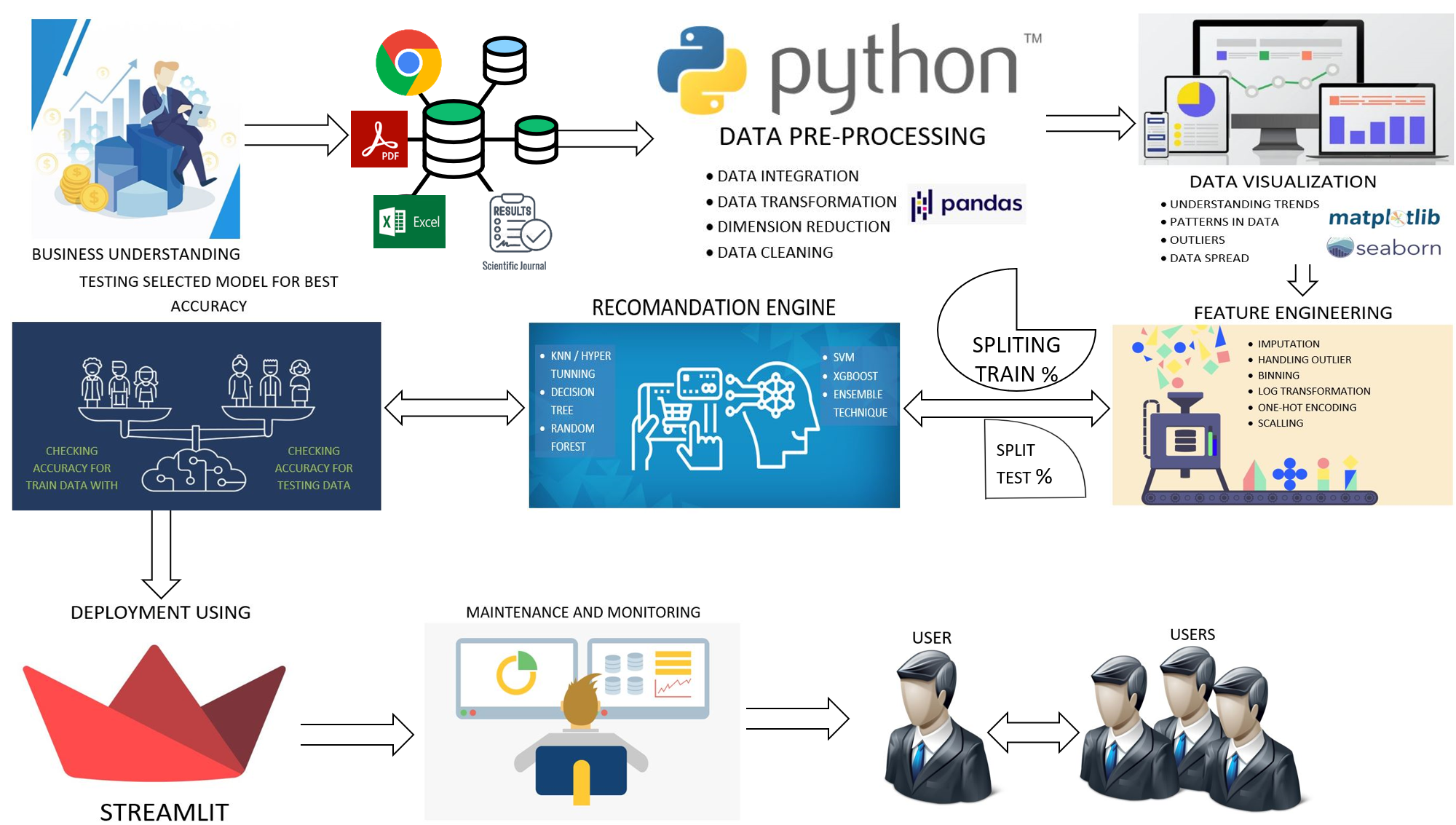
* **ABSTRACT-*:*** Agriculture plays a significant role in the growth of the national economy. It relay  
  on weather and other environmental aspects. Some of the factors on which agriculture is  
  dependent are Soil, climate, flooding, fertilizers, temperature, precipitation, crops, insecticides  
  and herbs. The crop yield is dependent on these factors and hence difficult to predict. To know  
  the status of crop production, in this work we perform descriptive study on agricultural data  
  using various machine learning techniques. Crop yield estimates include estimating crop  
  yields from available historical data such as precipitation data, soil data, and historic crop  
  yields. This prediction will help farmers to predict crop yield before farming. Here we are  
  utilizing three datasets like as clay data set, precipitation data set, and production data set of  
  Telangana state, then we structure an assembled data sets and on this data set we deploy

different algorithms to get the accurate and genuine assessed crop recommendation.

* **INTRODUCTION-:**The main aim of the agricultural Project is to Recommend TOP 5 crops in upcoming 3 months & return on investment by using various historical Agriculture resources. Many machine learning algorithms can help in the prediction of ROI.The method of crop selection is applied to improve crop production.
* **PROPOSED SYSTEM/METHODOLOGY-:**

In the proposed system, we use supervised learning to form a model, which provides Recommendations of crop and corresponding return on investment. The proposed system is

described in following stages such as dataset collection, pre-processing step,feature engineering feature selection and applying machine learning modules as shown in below figure



**Dataset Collection**: Data is collected from a variety of sources and prepared for data sets.And this data is used for descriptive analysis. Data is available from several online abstract sources such as telangana agri sites and data.gov.in. We will use an annual summary of crops for at least 2 years. The data sets used in this project are soil data set, rainfall data set, temperature data and crop yield data.

**Pre-processing step**: The most important step in machine learning is Preprocessing. preprocessing consists of replacing the missing values with appropriate value , selecting the appropriate data range, and extracting the functionality. This kind of the dataset is critical to analysis process. In this project we replaced null values with median imputation , we used label encoding for converting thecategorical data into numerical data and also normalized the data using standard scaler

In this project we have two data set one is for crop recommendation and the other one for Return on investment.

1. **Crop Recommendation**

**Variable used-**:District\_Name, Season, Crop,N kg/ha, P kg/ha, K kg/ha,ph, ' Humidity\_Min, Humidity\_Max, Humidity\_around, Rainfall,Temp\_around, Area,

productio . Output variable-:Crop

**Feature Selection:**  Feature extraction should simplify the amount of data by collecting only relevant features involved to represent a large data set & In our case we already gather & consider only relevant columns for project.

**Null values-**:Here in this dataset we have null values in the production & rainfall column it is very important to handle null value , here null values are replaced by median values of that feature. and for outliers we have used winsorization technique

**Encoding techniques-**:here we use labelEncoder for converting categorical variables to numerical. in this dataset we have three categorical variable that we have to covert it to numerical i.e Crop, Season, District\_Name.& why we have used label Encoding here because refers to converting the labels into a numeric form so as to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning.

**Normalizing the data-:** here we have used Standard scaler to normalize the data, which standardizes a feature by subtracting the mean and then scaling to unit variance.huere we normalize total 11 variable i.e 'N kg/ha', 'P kg/ha', 'K kg/ha',ph’,'Humidity\_Min', 'Humidity\_Max', 'Humidity\_around', 'Rainfall', 'Temp\_around', 'Area', 'Production'

**Split the Dataset into Train and Test Set-:**This step includes training and testing of input data. The loaded data is divided into two sets, such as training data and test data, with a division ratio of 80% or 20%, such as 0.8 or 0.2. here we used 0.25 for the split

1. **Crop Return On Investment:**

**Variables used:-**District,crop, year,season, Area (Hectare),Production (tonnes), Yield(Tonnes/Hectare),crop duration(days),seeds kg/ha,urea kg/ha,DAP kg/ha

**Null values: -**some feature have some missing values we have used median imputation for missing values.

**Encoding techniques:-** here we have used label encoder to convert categorical feature to numerical in this dataset we have 5 categorical features District,Crop,Year,Season,CropDuration\_days,

**Outlier:-** In this dataset some features have outliers we have used winsorization technique to treat the outliers

**Scaling the data:-** we have scaled the data using standard scaler StandardScaler standardizes a feature by subtracting the mean and then scaling to unitvariance. Unit variance means dividing all the values by the standard deviation

**Model selection:-** we have used logistic regression,KNN,SVM,Decisiontree classifier,Gridsearchcv,Randomforest classifier,adaboost classifier,Gradientboost classifier,XGBOOST