Team ID	PNT2022TMID53366
Project Name	Estimate the Crop Yield using Data Analytics

Dropping NaN values:

```
In [3]: # Droping Nan Values
        data = df.dropna()
        print(data.shape)
        test = df[~df["Production"].notna()].drop("Production",axis=1)
        print(test.shape)

        (242361, 7)
        (3730, 6)
```

Splitting training and test dataset:

```
x = data_dum.drop("Production",axis=1)
y = data_dum["Production"]
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.33, random_state=42)
print("x_train :",x_train.shape)
print("x_test :",x_test.shape)
print("y_train :",y_train.shape)
print("y_test :",y_test.shape)

x_train : (162381, 165)
x_test : (79980, 165)
y_train : (162381,)
y_test : (79980,)
```

Building the model:

Algorithm used: RandomForestRegressor

```
In [23]: from sklearn.ensemble import RandomForestRegressor
         model = RandomForestRegressor(n_estimators=100)
         model.fit(x_train[com_fea],y_train)

Out[23]: RandomForestRegressor()
```