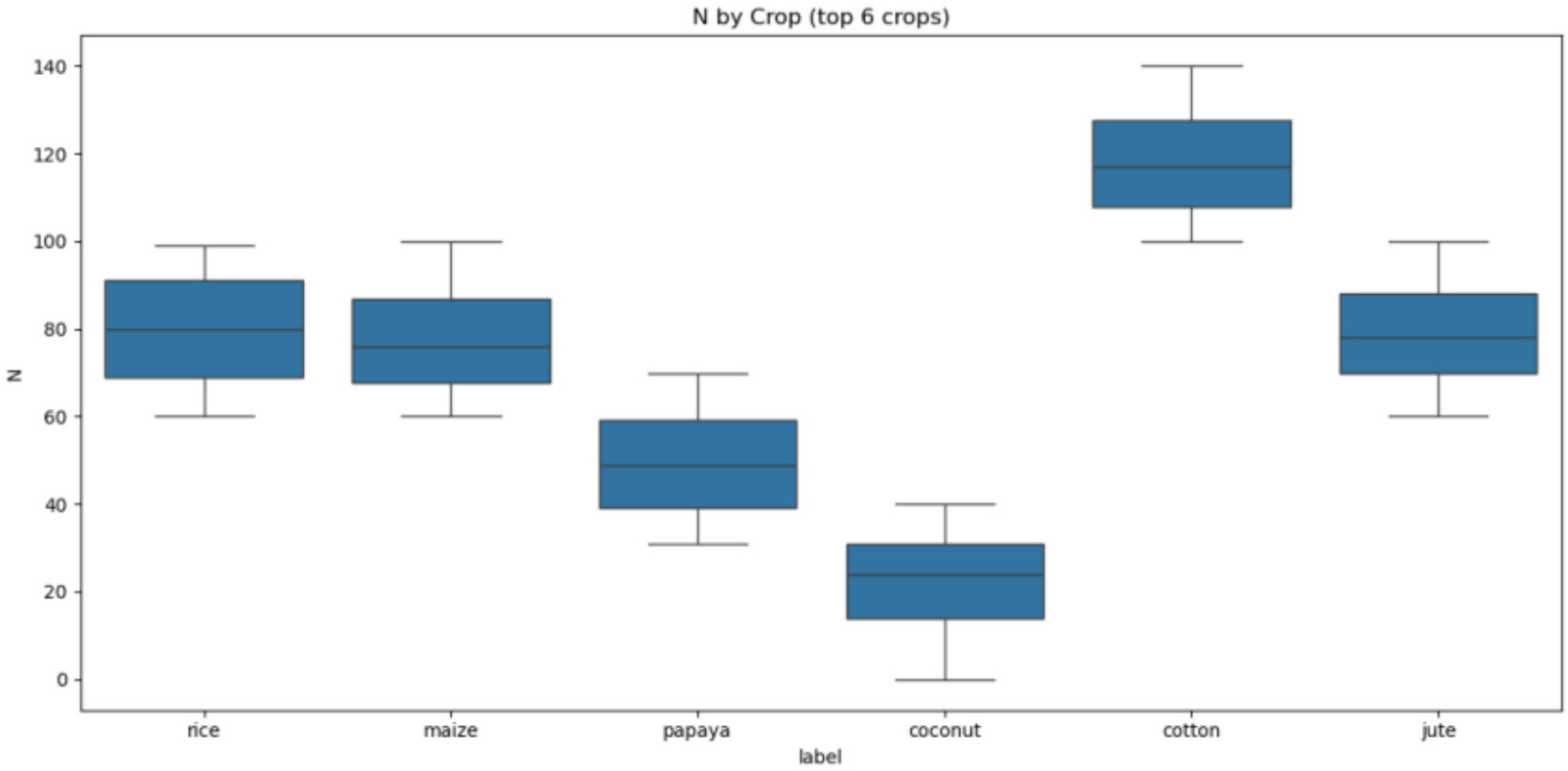
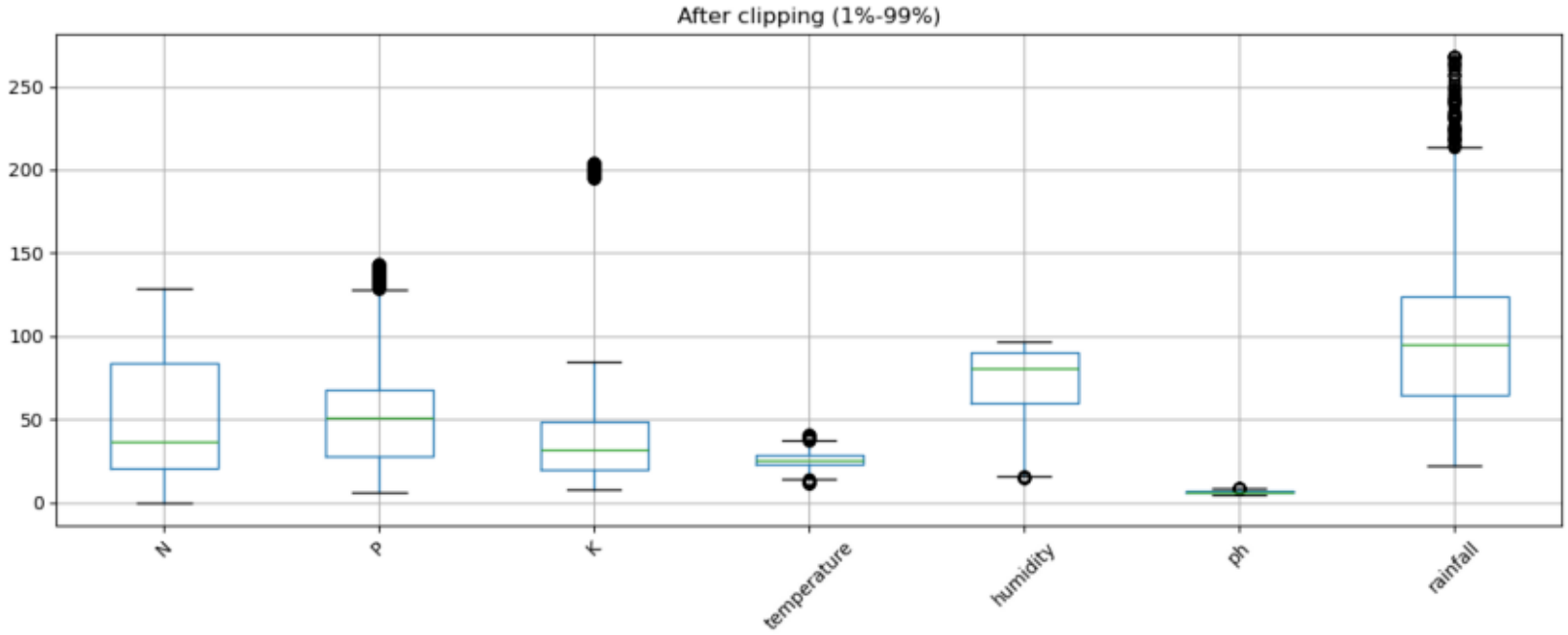
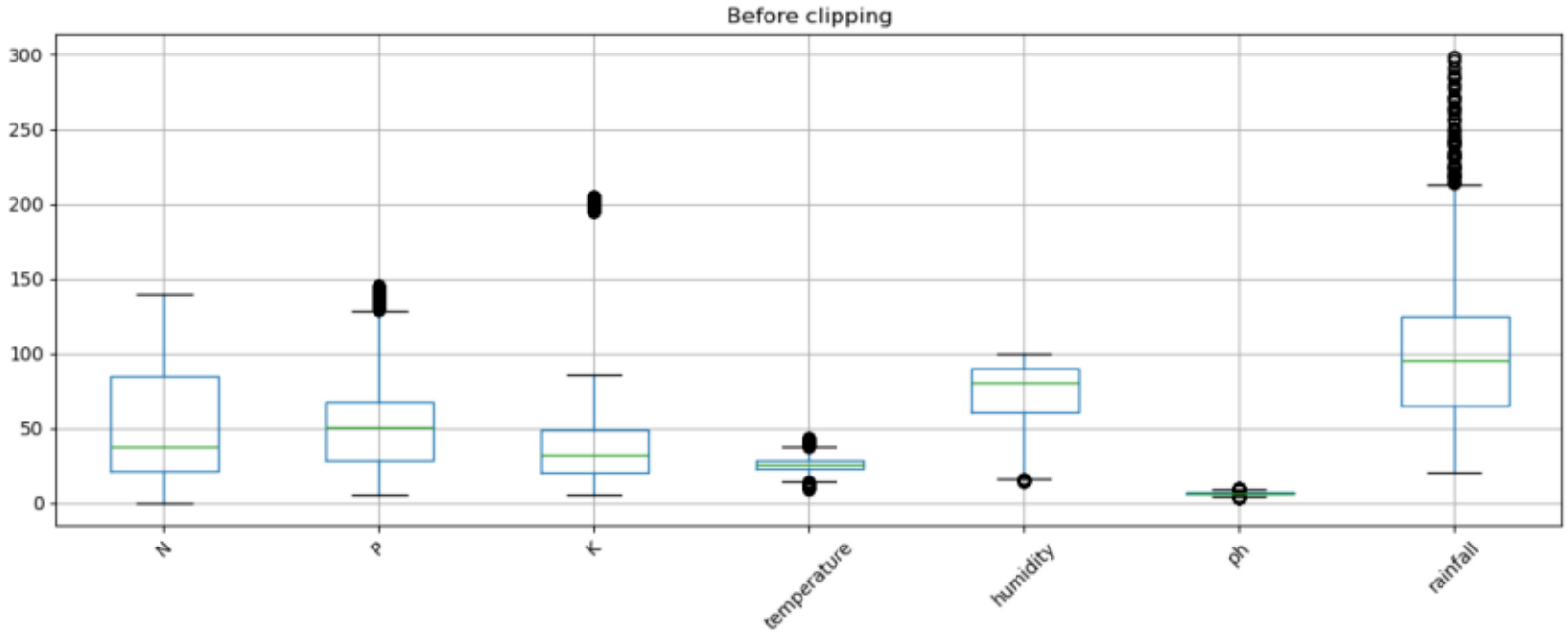
Crop Yield Prediction Report

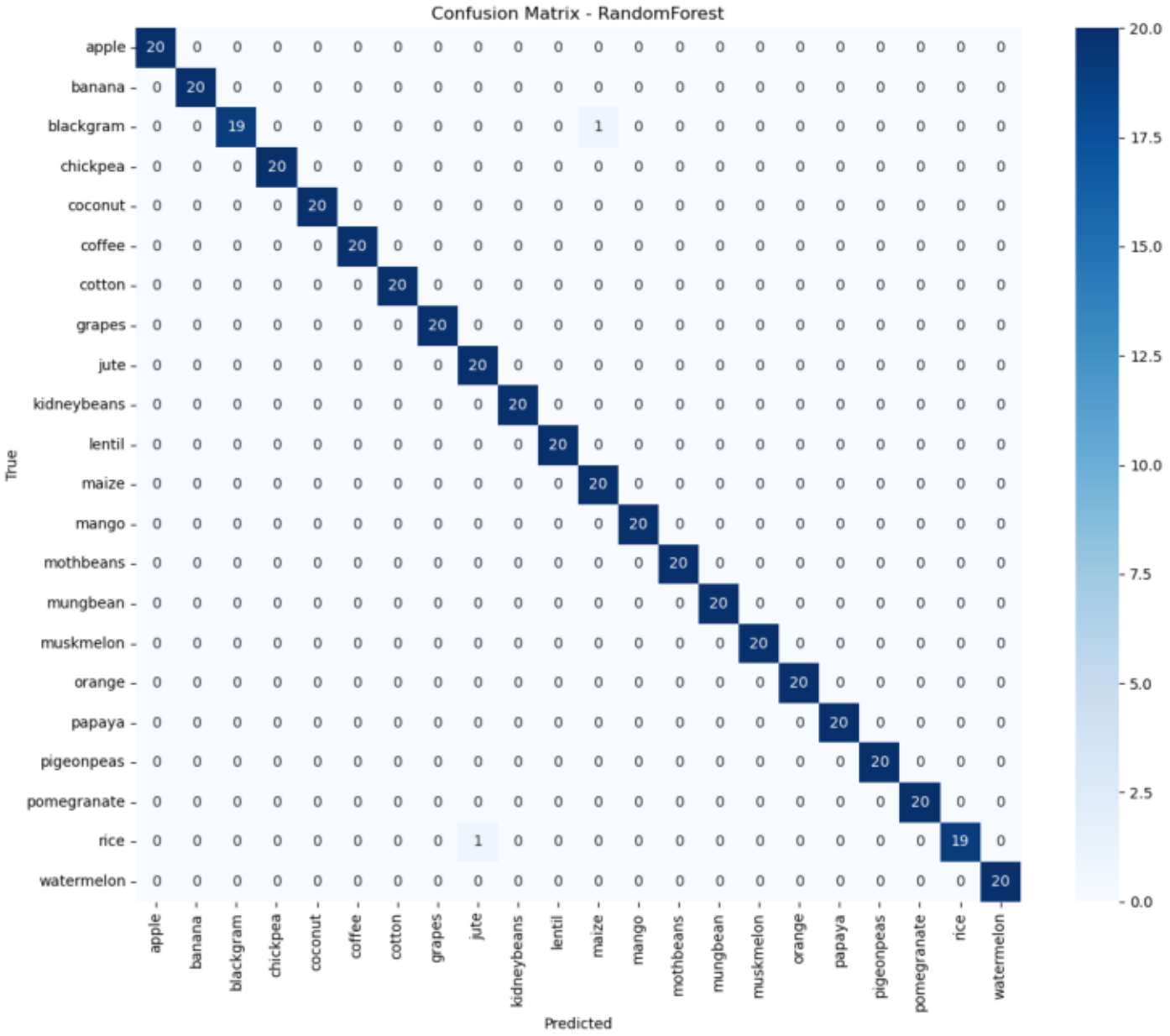
Dataset: Crop_recommendation.csv
Rows: 2200 | Features: 7 | Crops (classes): 22
Best model: RandomForest
Test Accuracy: 0.9955
Test Weighted F1: 0.9955

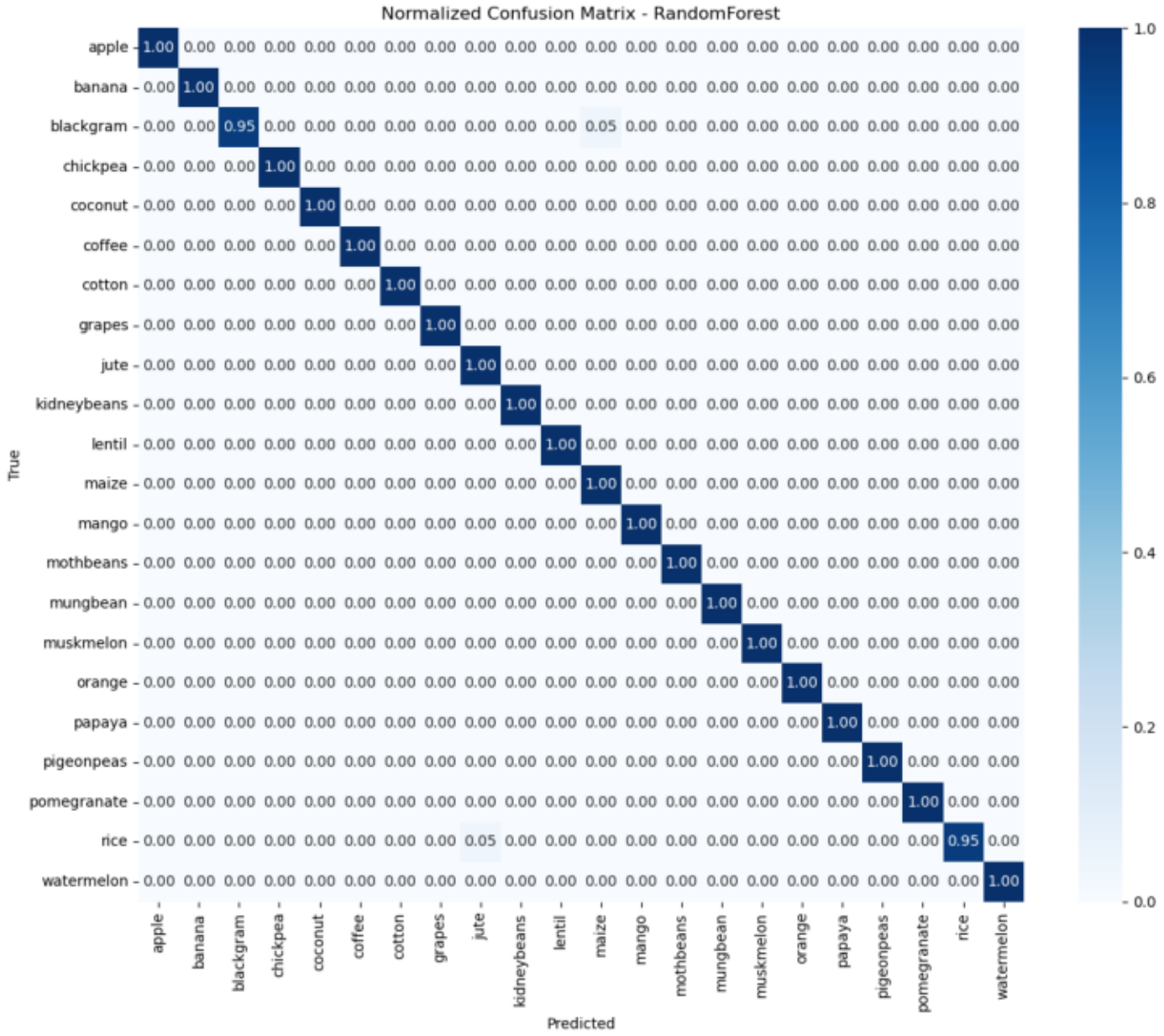
Notes:
- Outliers clipped at quantiles 1% - 99% before scaling.
- Preprocessing uses ColumnTransformer -> StandardScaler.
- See following pages for EDA plots, matrices, and feature importance.

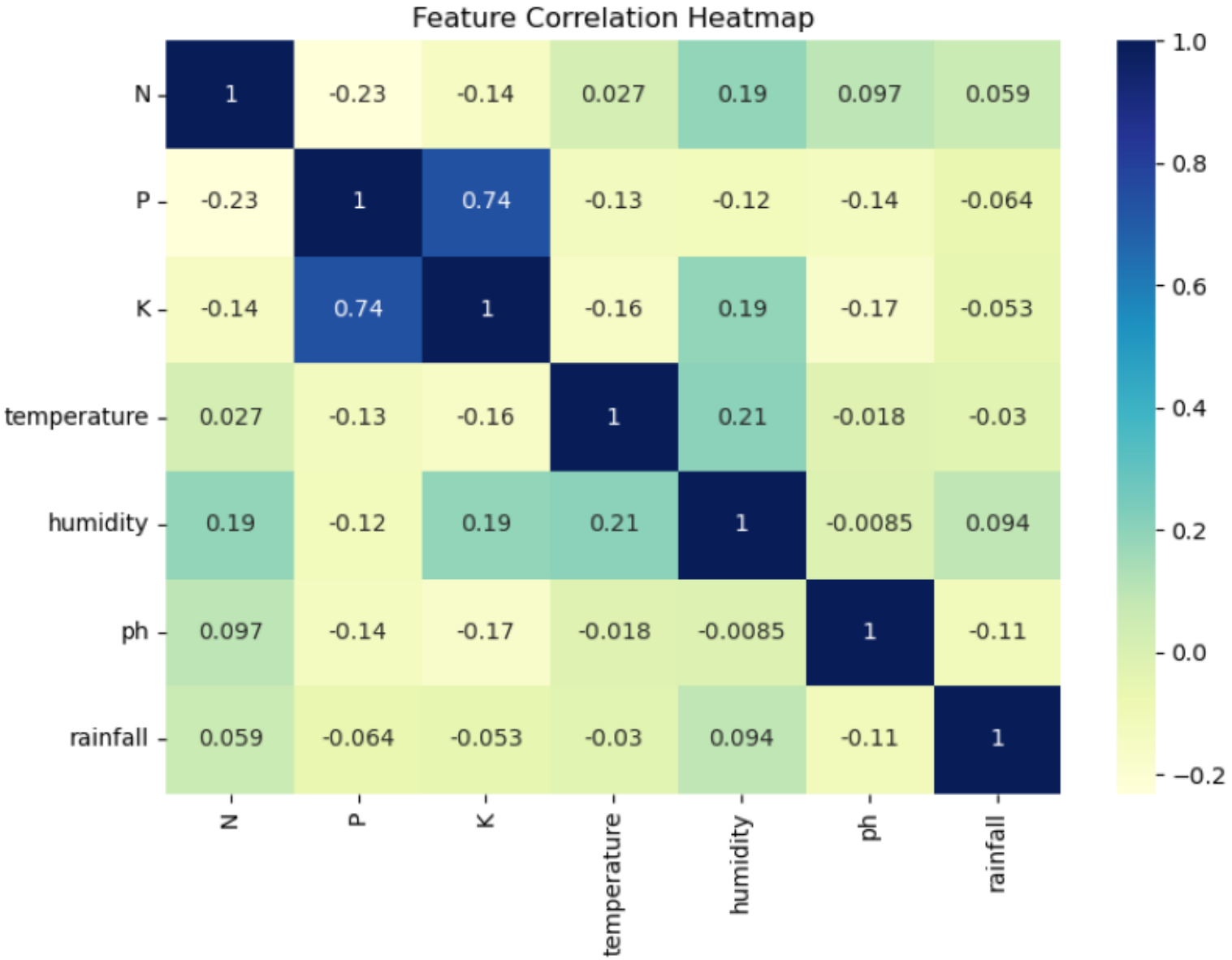

N by Crop (top 6 crops)
N
140
120
100
80
60
40
20
0
rice
maize
papaya
coconut
cotton
jute
label

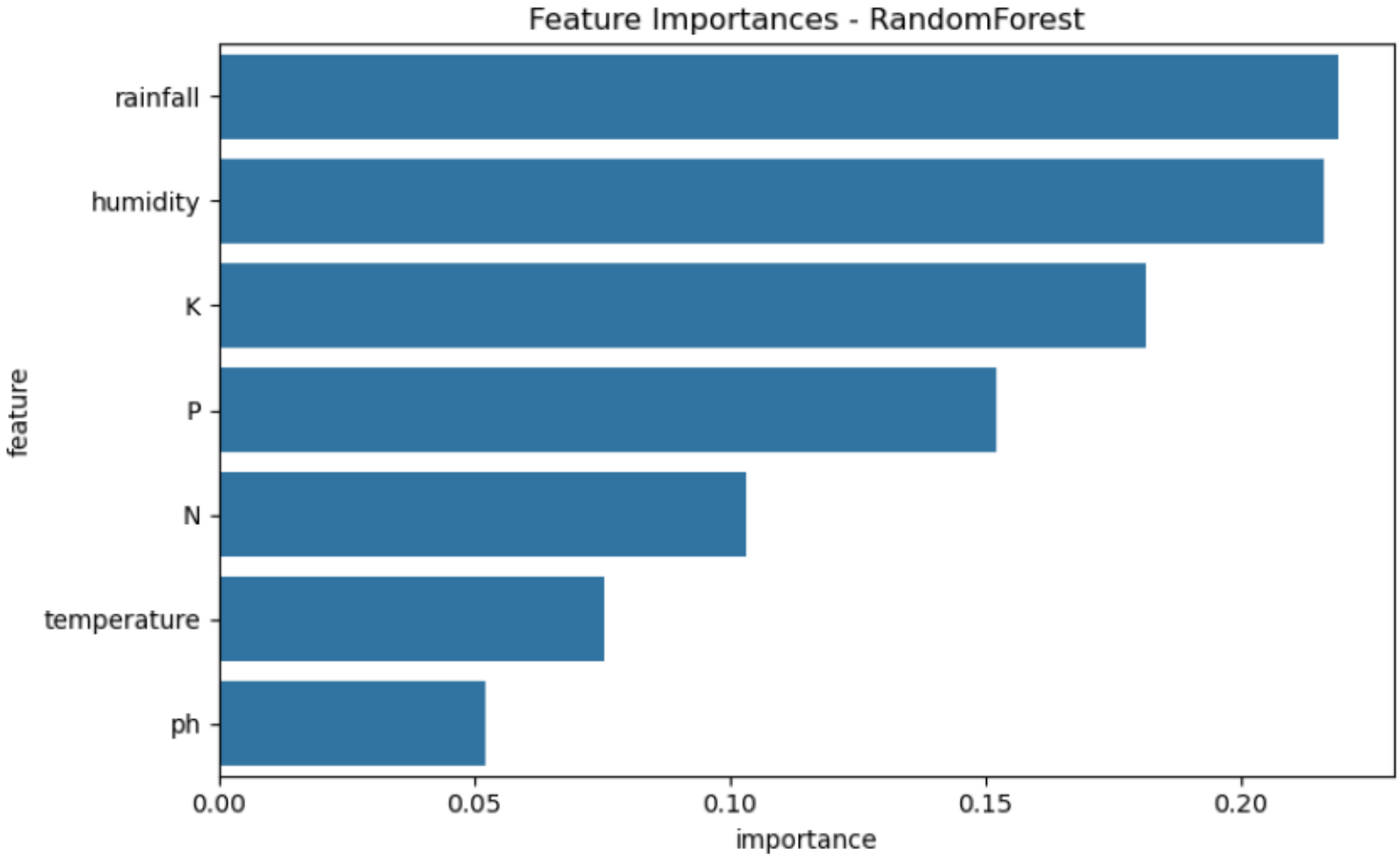

After clipping (1%-99%)
250
200
150
100
50
0
N
P
K
temperature
humidity
ph
rainfall

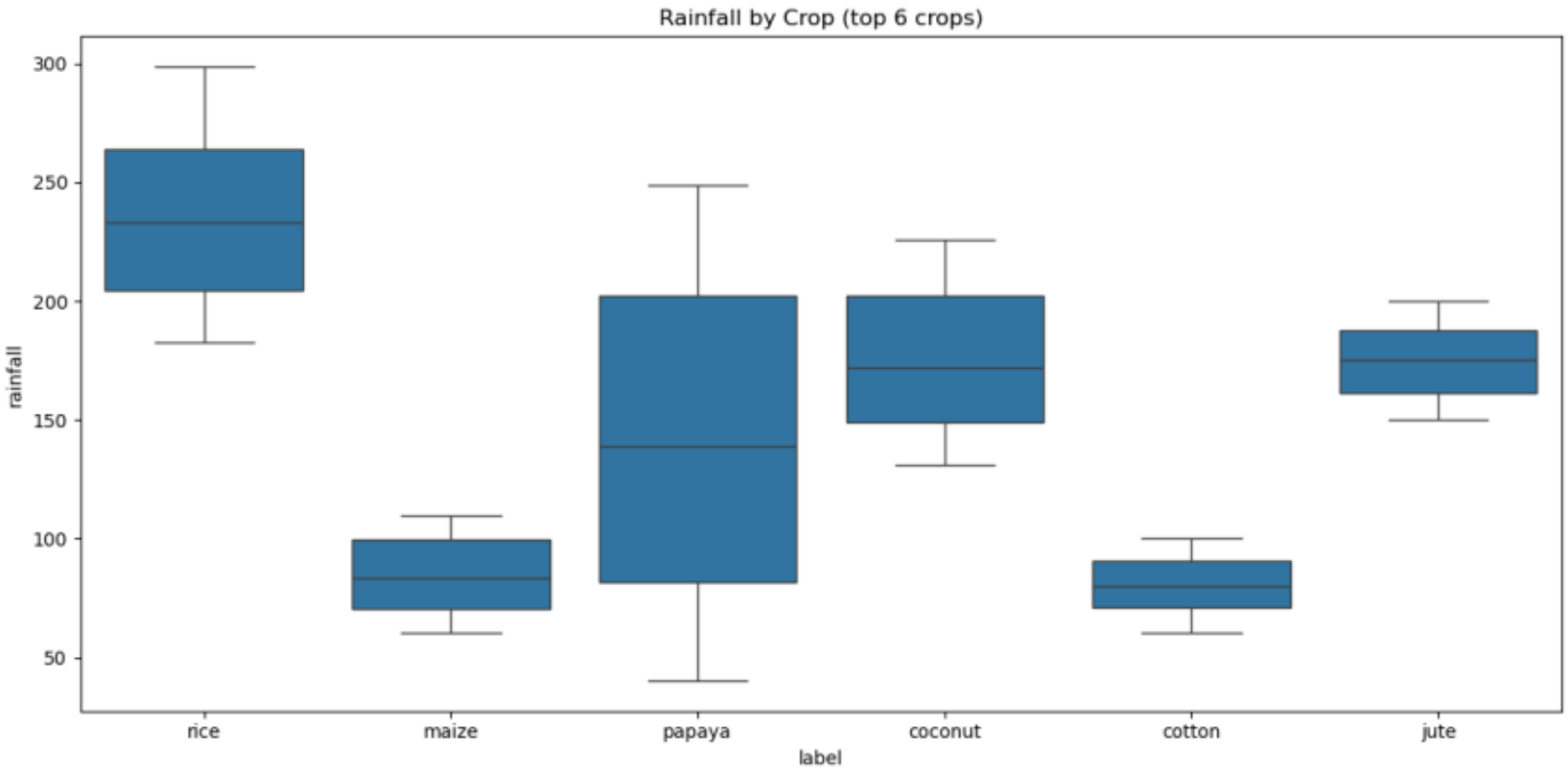

Before clipping
300
250
200
150
100
50
0
N
P
K
temperature
humidity
ph
rainfall

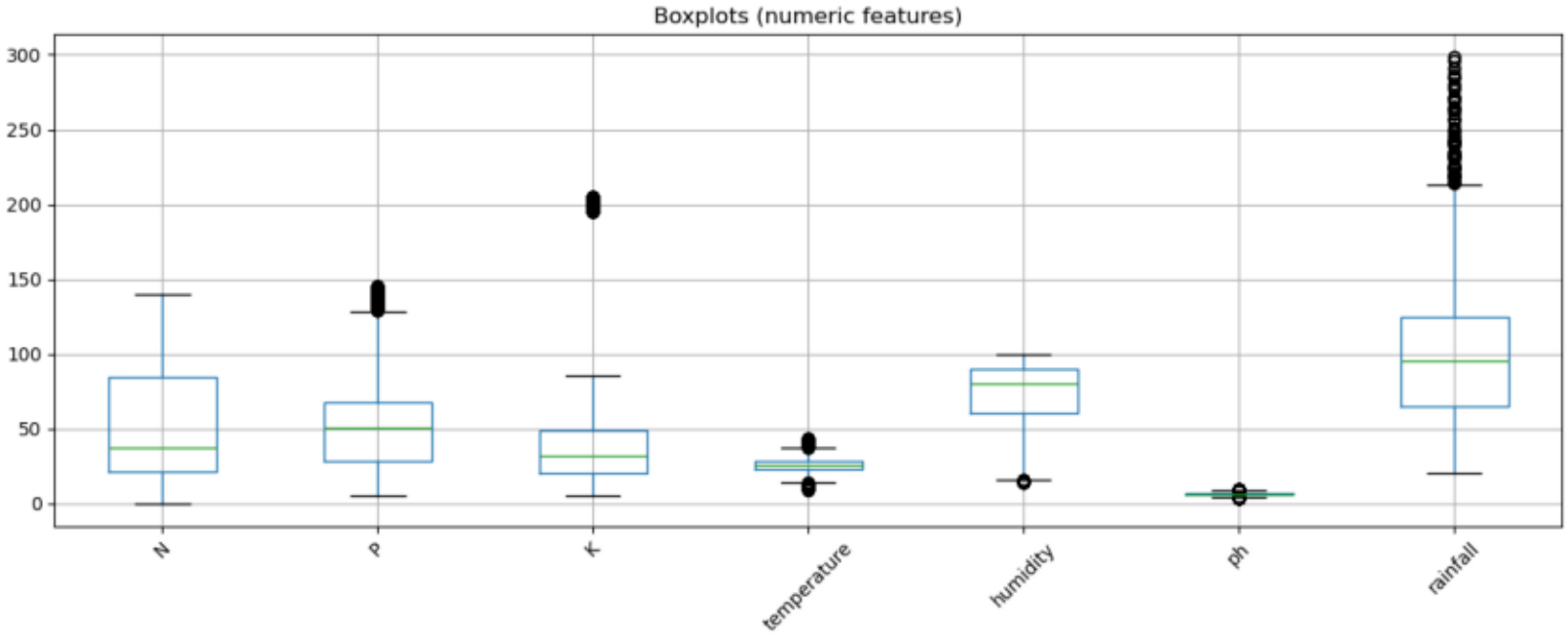

Confusion Matrix - RandomForest
True
Predicted
apple
banana
blackgram
chickpea
coconut
coffee
cotton
grapes
jute
kidneybeans
lentil
maize
mango
mothbeans
mungbean
muskmelon
orange
papaya
pigeonpeas
pomegranate
rice
watermelon
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0

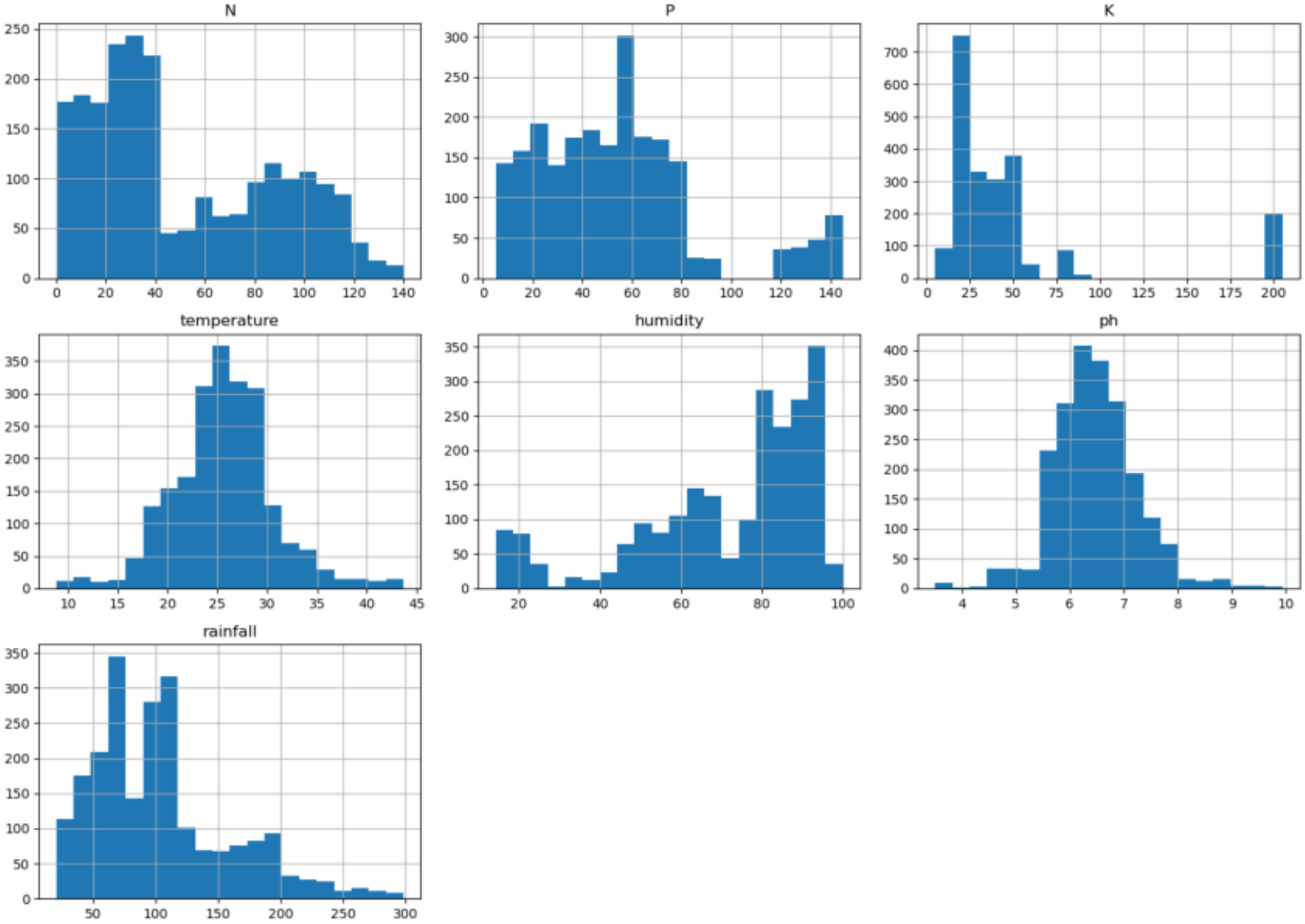

Normalized Confusion Matrix - RandomForest
True
Predicted
apple
banana
blackgram
chickpea
coconut
coffee
cotton
grapes
jute
kidneybeans
lentil
maize
mango
mothbeans
mungbean
muskmelon
orange
papaya
pigeonpeas
pomegranate
rice
watermelon
1.0
0.8
0.6
0.4
0.2
0.0


Feature Correlation Heatmap
N
P
K
temperature
humidity
ph
rainfall
1
-0.23
-0.14
0.027
0.19
0.097
0.059
-0.23
1
0.74
-0.13
-0.12
-0.14
-0.064
-0.14
0.74
1
-0.16
0.19
-0.17
-0.053
0.027
-0.13
-0.16
1
0.21
-0.018
-0.03
0.19
-0.12
0.19
0.21
1
-0.0085
0.094
0.097
-0.14
-0.17
-0.018
-0.0085
1
-0.11
0.059
-0.064
-0.053
-0.03
0.094
-0.11
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


Feature Importances - RandomForest
rainfall
humidity
K
P
N
temperature
ph
feature
0.00
0.05
0.10
0.15
0.20
importance


Rainfall by Crop (top 6 crops)
rainfall
300
250
200
150
100
50
rice
maize
papaya
coconut
cotton
jute
label


Boxplots (numeric features)
300
250
200
150
100
50
0
N
P
K
temperature
humidity
ph
rainfall


N
P
K
temperature
humidity
ph
rainfall