

This is the code that we used to calculate the values that you wanted from us:

```
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import SelectFromModel
from sklearn.metrics import accuracy_score, roc_auc_score
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import ShuffleSplit
from sklearn.metrics import classification_report

df = pd.read_csv(r'C:\Users\begum\Desktop\Projects\Pattern
Recognition\otu.csv', low_memory=False)

X = df.iloc[2:, :].T
y = df.iloc[1, :]

X = X.apply(pd.to_numeric, errors='coerce')
X = X.fillna(X.mean())

le = LabelEncoder()
y = le.fit_transform(y)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

selector = SelectFromModel(estimator=RandomForestClassifier(n_estimators=100,
random_state=42))
X_train = selector.fit_transform(X_train, y_train)
X_test = selector.transform(X_test)

clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

n_splits_adjusted = min(5, min(pd.Series(y).value_counts()))
ss = ShuffleSplit(n_splits=n_splits_adjusted, test_size=0.2, random_state=42)
scores = cross_val_score(clf, X_train, y_train, cv=ss)
print(f'Cross-validation scores: {scores}')
```

```
y_pred = clf.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy}')
```

```
if len(set(y)) == 2:
    auc = roc_auc_score(y_test, y_pred)
    print(f'AUC: {auc}')
```

```
report = classification_report(y_test, y_pred, zero_division=1)
print(report)
```

And these are the outputs:

Cross-validation scores: [0.79545455]

Accuracy: 0.8363636363636363

	precision	recall	f1-score	support
0	0.90	1.00	0.95	46
3	1.00	0.00	0.00	1
6	0.00	1.00	0.00	0
10	0.00	1.00	0.00	0
12	1.00	0.00	0.00	1
13	1.00	0.00	0.00	1
14	1.00	0.00	0.00	1
17	1.00	0.00	0.00	1
18	0.00	1.00	0.00	0
22	1.00	0.00	0.00	1
25	0.00	1.00	0.00	0
26	1.00	0.00	0.00	1
37	1.00	0.00	0.00	1
44	1.00	0.00	0.00	1
accuracy			0.84	55
macro avg	0.71	0.36	0.07	55
weighted avg	0.92	0.84	0.79	55