

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, matthews_
import pickle
```

```
data = pd.read_csv('Breast_cancer_data.csv')
print(data.head())
print(data.shape)
data = data[["mean_radius", "mean_texture", "mean_perimeter", "mean_area", "mean_smoothness", ']]
```

	mean_radius	mean_texture	...	mean_smoothness	diagnosis
0	17.99	10.38	...	0.11840	0
1	20.57	17.77	...	0.08474	0
2	19.69	21.25	...	0.10960	0
3	11.42	20.38	...	0.14250	0
4	20.29	14.34	...	0.10030	0

```
[5 rows x 6 columns]
(569, 6)
```

```
best_acc = 0
for i in range(50):
    predict = "diagnosis"
    x = np.array(data.drop([predict],1))
    y = np.array(data[predict])
    x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.1)
    rf = RandomForestClassifier()
    rf.fit(x_train,y_train)
    y_pred = rf.predict(x_test)
    acc = accuracy_score(y_test,y_pred)
    if acc>best_acc:
        best_acc = acc
        best_model = rf
```

```
y_pred = best_model.predict(x_test)
```

```
fileName = "randomForestModel.pickle"
pickle.dump(best_model,open(fileName,'wb'))
```

```
print("Accuracy =",str(best_acc*100)+"%")
```

```
Accuracy = 98.24561403508771%
```

---

✓ 10s completed at 23:53

● ✕