CODE

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load Data
try:
    data = pd.read_csv('breast_cancer_data.csv')
    print("Data loaded successfully")
except Exception as e:
    print(f"Error loading data: {e}")

# View the first few rows of the dataset
print("First few rows of the dataset:")
print(data.head())

# Step 2: Define Features and Target
try:
    X = data[['radius_mean', 'texture_mean', 'perimeter_mean', 'area_mean',
              'smoothness_mean', 'compactness_mean', 'concavity_mean',
              'concave_points_mean', 'symmetry_mean', 'fractal_dimension_mean']]
    y = data['diagnosis']
    print("Features and target variable defined successfully")
except KeyError as e:
    print(f"Error defining features and target: {e}")

# Encode the target variable
try:
    label_encoder = LabelEncoder()
    y_encoded = label_encoder.fit_transform(y)  # Convert 'M' and 'B' to 1 and 0
    print("Target variable encoded successfully")
except Exception as e:
    print(f"Error encoding target variable: {e}")
```

```
# Step 3: Split the Data into Training and Testing Sets
try:
    X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.2,
random_state=0)
    print("Data split into training and testing sets successfully")
except Exception as e:
    print(f"Error splitting data: {e}")

# Define and Train the Model
try:
    model = RandomForestClassifier(n_estimators=100, random_state=0)
    model.fit(X_train, y_train)
    print("Model trained successfully")
except Exception as e:
    print(f"Error training the model: {e}")

# Make Predictions
try:
    y_pred = model.predict(X_test)
    print("Predictions made successfully")
except Exception as e:
    print(f"Error making predictions: {e}")

# Evaluate the Model
try:
    print("Accuracy:", accuracy_score(y_test, y_pred))
    print("Classification Report:\n", classification_report(y_test, y_pred))
except Exception as e:
    print(f"Error evaluating the model: {e}")

# Feature Importances
try:
    importances = model.feature_importances_
    features = X.columns
    feature_importance_df = pd.DataFrame({
        'Feature': features,
```

```
        'Importance': importances
    }).sort_values(by='Importance', ascending=False)
    print("Feature importances calculated successfully")
except Exception as e:
    print(f"Error calculating feature importances: {e}")
import matplotlib.pyplot as plt
import seaborn as sns

# Assuming 'y_encoded' contains the encoded target variable (0: Benign, 1:
Malignant)

# Plot the distribution of diagnosis (Malignant and Benign)
plt.figure(figsize=(8, 6))
sns.countplot(x=y_encoded, palette="coolwarm")
plt.title('Distribution of Breast Cancer Diagnosis')
plt.xlabel('Diagnosis (0: Benign, 1: Malignant)')
plt.ylabel('Count')
plt.show()

# Step 7: Identify and Print Individuals Affected by Cancer
try:
    malignant_indices = X_test.index[y_pred == 1]
    if not malignant_indices.empty:
        malignant_cases = data.loc[malignant_indices]
        print("Individuals affected by cancer (predicted malignant cases):")
        print(malignant_cases)
    else:
        print("No individuals predicted to be affected by cancer in the test set.")
except Exception as e:
    print(f"Error identifying malignant cases: {e}")
```

OUTPUT

```
         id diagnosis  radius_mean  texture_mean  perimeter_mean  area_mean  \
0    842302         M        17.99         10.38          122.80     1001.0
1    842517         M        20.57         17.77          132.90     1326.0
2  84300903         M        19.69         21.25          130.00     1203.0
3  84348301         M        11.42         20.38           77.58      386.1
4  84358402         M        20.29         14.34          135.10     1297.0

   smoothness_mean  compactness_mean  concavity_mean  concave_points_mean  \
0          0.11840           0.27760          0.3001              0.14710
1          0.08474           0.07864          0.0869              0.07017
2          0.10960           0.15990          0.1974              0.12790
3          0.14250           0.28390          0.2414              0.10520
4          0.10030           0.13280          0.1980              0.10430

   ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0  ...         25.38          17.33           184.60      2019.0
1  ...         24.99          23.41           158.80      1956.0
2  ...         23.57          25.53           152.50      1709.0
3  ...         14.91          26.50            98.87       567.7
4  ...         22.54          16.67           152.20      1575.0

   smoothness_worst  compactness_worst  concavity_worst  concave_points_worst
0            0.1622             0.6656           0.7119                0.2654
1            0.1238             0.1866           0.2416                0.1860
2            0.1444             0.4245           0.4504                0.2430
3            0.2098             0.8663           0.6869                0.2575
4            0.1374             0.2050           0.4000                0.1625

   symmetry_worst  fractal_dimension_worst
0          0.4601                  0.11890
1          0.2750                  0.08902
2          0.3613                  0.08758
3          0.6638                  0.17300
4          0.2364                  0.07678

[5 rows x 32 columns]
Accuracy: 0.9298245614035088
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.93      0.94        67
           1       0.90      0.94      0.92        47

    accuracy                           0.93       114
   macro avg       0.93      0.93      0.93       114
weighted avg       0.93      0.93      0.93       114
```

AFFECTED PEOPLE

```
Individuals affected by cancer (predicted malignant cases):
            id diagnosis  radius_mean  texture_mean  perimeter_mean  \
512     915691         M        13.40         20.52           88.64
421     906564         B        14.69         13.98           98.22
157    8711216         B        16.84         19.46          108.40
89      861598         B        14.64         15.24           95.77
172      87164         M        15.46         11.89          102.50
233   88206102         M        20.51         27.81          134.40
389      90312         M        19.55         23.21          128.90
250     884948         M        20.94         23.56          138.90
283    8912280         M        16.24         18.77          108.80
372    9012795         M        21.37         15.10          141.30
14    84667401         M        13.73         22.61           93.60
337     897630         M        18.77         21.43          122.90
1       842517         M        20.57         17.77          132.90
132   86730502         M        16.16         21.54          106.20
64    85922302         M        12.68         23.84           82.69
127     866203         M        19.00         18.91          123.40
353    9010018         M        15.08         25.74           98.00
414     905680         M        15.13         29.81           96.71
10      845636         M        16.02         23.24          102.70
564     926424         M        21.56         22.39          142.00
15    84799002         M        14.54         27.54           96.73
12      846226         M        19.17         24.80          132.40
194   87556202         M        14.86         23.21          100.40
134     867739         M        18.45         21.91          120.20
272    8910988         M        21.75         20.99          147.30
196     875938         M        13.77         22.29           90.63
75     8610404         M        16.07         19.65          104.10
468    9113538         M        17.60         23.33          119.00
108      86355         M        22.27         19.67          152.80
239   88330202         M        17.46         39.28          113.40
210  881046502         M        20.58         22.14          134.70
```

```
     area_mean  smoothness_mean  compactness_mean  concavity_mean  \
512      556.7          0.11060           0.14690         0.14450
421      656.1          0.10310           0.18360         0.14500
157      880.2          0.07445           0.07223         0.05150
89       651.9          0.11320           0.13390         0.09966
172      736.9          0.12570           0.15550         0.20320
233     1319.0          0.09159           0.10740         0.15540
389     1174.0          0.10100           0.13180         0.18560
250     1364.0          0.10070           0.16060         0.27120
283      805.1          0.10660           0.18020         0.19480
372     1386.0          0.10010           0.15150         0.19320
14       578.3          0.11310           0.22930         0.21280
337     1092.0          0.09116           0.14020         0.10600
1       1326.0          0.08474           0.07864         0.08690
132      809.8          0.10080           0.12840         0.10430
64       499.0          0.11220           0.12620         0.11280
127     1138.0          0.08217           0.08028         0.09271
353      716.6          0.10240           0.09769         0.12350
414      719.5          0.08320           0.04605         0.04686
10       797.8          0.08206           0.06669         0.03299
564     1479.0          0.11100           0.11590         0.24390
15       658.8          0.11390           0.15950         0.16390
12      1123.0          0.09740           0.24580         0.20650
194      671.4          0.10440           0.19800         0.16970
134     1075.0          0.09430           0.09709         0.11530
272     1491.0          0.09401           0.19610         0.21950
196      588.9          0.12000           0.12670         0.13850
75       817.7          0.09168           0.08424         0.09769
468      980.5          0.09289           0.20040         0.21360
108     1509.0          0.13260           0.27680         0.42640
239      920.6          0.09812           0.12980         0.14170
210     1290.0          0.09090           0.13480         0.16400
17       798.8          0.11700           0.20220         0.17220
385      664.7          0.08682           0.06636         0.08390
```