

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import joblib

data = pd.read_csv("OCD Patient Dataset_ Demographics & Clinical Data.csv")

data.info()

data.dropna(inplace=True)

label_encoders = {}
categorical_columns = ['Gender', 'Ethnicity', 'Marital Status', 'Education Level', 'Family History of
OCD']

for col in categorical_columns:
    le = LabelEncoder()
    data[col] = le.fit_transform(data[col])
    label_encoders[col] = le

features = ['Age', 'Gender', 'Ethnicity', 'Marital Status', 'Education Level', 'Duration of Symptoms',
'Family History of OCD']
target = 'Y-BOCS Score'

X = data[features]
y = data[target]
```

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

```
scaler = StandardScaler()
```

```
X_train = scaler.fit_transform(X_train)
```

```
X_test = scaler.transform(X_test)
```

```
model = RandomForestRegressor(n_estimators=100, random_state=42)
```

```
model.fit(X_train, y_train)
```

```
y_pred = model.predict(X_test)
```

```
mse = mean_squared_error(y_test, y_pred)
```

```
r2 = r2_score(y_test, y_pred)
```

```
print(f"Mean Squared Error: {mse}")
```

```
print(f"R-squared Score: {r2}")
```

```
joblib.dump(model, "ocd_severity_model.pkl")
```

```
joblib.dump(scaler, "scaler.pkl")
```

```
joblib.dump(label_encoders, "label_encoders.pkl")
```

```
print("Model and encoders saved successfully.")
```