```
!pip install ydata_profiling
```

```
Collecting imagehash==4.3.1 (from ydata_profiling)
  Downloading ImageHash-4.3.1-py2.py3-none-any.whl.metadata (8.0 kB)
Requirement already satisfied: wordcloud>=1.9.3 in /usr/local/lib/python3.12/dist-packages (from ydata_profiling) (1.9.4)
Collecting dacite>=1.8 (from ydata_profiling)
  Downloading dacite-1.9.2-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: numba<=0.61,>=0.56.0 in /usr/local/lib/python3.12/dist-packages (from ydata_profiling) (0.6
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.1->ydata_profili
Requirement already satisfied: pillow in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.1->ydata_profiling)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2<3.2,>=2.11.1->ydata
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->y
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->yd
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->y
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.12/dist-packages (from numba<=0.61,>=0
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.1->ydat
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.1->yd
Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.12/dist-packages (from phik<0.13,>=0.11.1->ydata_p
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata_
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata_p
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->yda
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydat
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->ydata_pr
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->yd
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->yd
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.12/dist-packages (from statsmodels<1,>=0.13.2->ydata
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visio
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visio
Collecting puremagic (from visions<0.8.2,>=0.7.5->visions[type_image_path]<0.8.2,>=0.7.5->ydata_profiling)
  Downloading puremagic-1.30-py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib<
Downloading ydata_profiling-4.17.0-py2.py3-none-any.whl (399 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 399.3/399.3 kB 10.3 MB/s eta 0:00:00
Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 296.5/296.5 kB 11.2 MB/s eta 0:00:00
Downloading dacite-1.9.2-py3-none-any.whl (16 kB)
Downloading filetype-1.2.0-py2.py3-none-any.whl (19 kB)
Downloading minify_html-0.18.1-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3.1/3.1 MB 42.7 MB/s eta 0:00:00
Downloading multimethod-1.12-py3-none-any.whl (10 kB)
Downloading phik-0.12.5-cp312-cp312-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl (679 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 679.7/679.7 kB 23.4 MB/s eta 0:00:00
Downloading scipy-1.15.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (37.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 37.3/37.3 MB 22.8 MB/s eta 0:00:00
Downloading visions-0.8.1-py3-none-any.whl (105 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 105.4/105.4 kB 4.3 MB/s eta 0:00:00
Downloading puremagic-1.30-py3-none-any.whl (43 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.3/43.3 kB 2.4 MB/s eta 0:00:00
Installing collected packages: puremagic, minify-html, filetype, scipy, multimethod, dacite, imagehash, visions, phik, yda
  Attempting uninstall: scipy
    Found existing installation: scipy 1.16.3
    Uninstalling scipy-1.16.3:
      Successfully uninstalled scipy-1.16.3
Successfully installed dacite-1.9.2 filetype-1.2.0 imagehash-4.3.1 minify-html-0.18.1 multimethod-1.12 phik-0.12.5 puremag
```

```
!pip install ucimlrepo
```

```
Collecting ucimlrepo
  Downloading ucimlrepo-0.0.7-py3-none-any.whl.metadata (5.5 kB)
Requirement already satisfied: pandas>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from ucimlrepo) (2.2.2)
Requirement already satisfied: certifi>=2020.12.5 in /usr/local/lib/python3.12/dist-packages (from ucimlrepo) (2025.10.5)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlrepo) (2.0
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlr
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlrepo) (2025
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlrepo) (20
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas>=1.0
Downloading ucimlrepo-0.0.7-py3-none-any.whl (8.0 kB)
Installing collected packages: ucimlrepo
Successfully installed ucimlrepo-0.0.7
```

```
from ucimlrepo import fetch_ucirepo

# fetch dataset
estimation_of_obesity_levels_based_on_eating_habits_and_physical_condition = fetch_ucirepo(id=544)

# data (as pandas dataframes)
X = estimation_of_obesity_levels_based_on_eating_habits_and_physical_condition.data.features
y = estimation_of_obesity_levels_based_on_eating_habits_and_physical_condition.data.targets
```

```
# metadata
print(estimation_of_obesity_levels_based_on_eating_habits_and_physical_condition.metadata)

# variable information
print(estimation_of_obesity_levels_based_on_eating_habits_and_physical_condition.variables)
```

```
{'uci_id': 544, 'name': 'Estimation of Obesity Levels Based On Eating Habits and Physical Condition ', 'repository_url': 'ht
                             name     role         type demographic  \
0                          Gender  Feature  Categorical      Gender
1                             Age  Feature   Continuous         Age
2                          Height  Feature   Continuous        None
3                          Weight  Feature   Continuous        None
4   family_history_with_overweight  Feature       Binary        None
5                            FAVC  Feature       Binary        None
6                            FCVC  Feature      Integer        None
7                             NCP  Feature   Continuous        None
8                            CAEC  Feature  Categorical        None
9                           SMOKE  Feature       Binary        None
10                           CH2O  Feature   Continuous        None
11                            SCC  Feature       Binary        None
12                            FAF  Feature   Continuous        None
13                            TUE  Feature      Integer        None
14                           CALC  Feature  Categorical        None
15                         MTRANS  Feature  Categorical        None
16                     NObeyesdad   Target  Categorical        None

                                        description units missing_values
0                                              None  None             no
1                                              None  None             no
2                                              None  None             no
3                                              None  None             no
4   Has a family member suffered or suffers from o...  None             no
5           Do you eat high caloric food frequently?  None             no
6        Do you usually eat vegetables in your meals?  None             no
7              How many main meals do you have daily?  None             no
8                  Do you eat any food between meals?  None             no
9                                   Do you smoke?  None             no
10                 How much water do you drink daily?  None             no
11         Do you monitor the calories you eat daily?  None             no
12           How often do you have physical activity?  None             no
13  How much time do you use technological devices...  None             no
14                   How often do you drink alcohol?  None             no
15           Which transportation do you usually use?  None             no
16                                   Obesity level  None             no
```

```
from ydata_profiling import ProfileReport

profile=ProfileReport(X,title="Obesity dataset report",explorative=True)
profile.to_file("Data-set-B-report.html")
```

```
/usr/local/lib/python3.12/dist-packages/ydata_profiling/utils/dataframe.py:137: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view
  df.rename(columns={"index": "df_index"}, inplace=True)
Summarize dataset: 100%                                                 89/89 [00:11<00:00,  4.26it/s, Completed]


  0%|          | 0/16 [00:00<?, ?it/s]
 38%|███▊      | 6/16 [00:00<00:00, 42.87it/s]
100%|██████████| 16/16 [00:00<00:00, 48.73it/s]
Generate report structure: 100%                                         1/1 [00:03<00:00,  3.70s/it]

Render HTML: 100%                                                       1/1 [00:00<00:00,  1.15it/s]

Export report to file: 100%                                             1/1 [00:00<00:00, 30.43it/s]
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler,OneHotEncoder
from sklearn.impute import SimpleImputer
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.metrics import classification_report,roc_auc_score,confusion_matrix,ConfusionMatrixDisplay
```

```
y.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   NObeyesdad  2111 non-null   object
dtypes: object(1)
memory usage: 16.6+ KB
```

```
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Gender                          2111 non-null   object
 1   Age                             2111 non-null   float64
 2   Height                          2111 non-null   float64
 3   Weight                          2111 non-null   float64
 4   family_history_with_overweight  2111 non-null   object
 5   FAVC                            2111 non-null   object
 6   FCVC                            2111 non-null   float64
 7   NCP                             2111 non-null   float64
 8   CAEC                            2111 non-null   object
 9   SMOKE                           2111 non-null   object
 10  CH2O                            2111 non-null   float64
 11  SCC                             2111 non-null   object
 12  FAF                             2111 non-null   float64
 13  TUE                             2111 non-null   float64
 14  CALC                            2111 non-null   object
 15  MTRANS                          2111 non-null   object
dtypes: float64(8), object(8)
memory usage: 264.0+ KB
```

```
Start coding or generate with AI.
```

```
y.head()
```

	NObeyesdad
0	Normal_Weight
1	Normal_Weight
2	Normal_Weight
3	Overweight_Level_I
4	Overweight_Level_II

Next steps: Generate code with y | New interactive sheet

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
y_processed=le.fit_transform(y['NObeyesdad'])
```

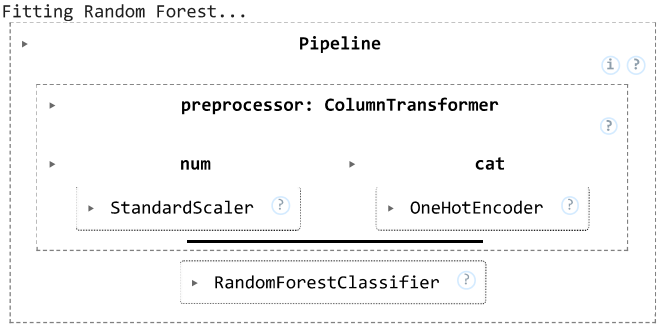
```
y_processed
```

```
array([1, 1, 1, ..., 4, 4, 4])
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier

# --- Define lists based on your ydata-profiling report ---
numeric_features = ['Age', 'Height', 'Weight', 'FCVC', 'NCP', 'CH2O', 'FAF', 'TUE']
categorical_features = ['Gender', 'family_history_with_overweight', 'FAVC', 'CAEC', 'SMOKE', 'SCC', 'CALC', 'MTRANS']

# --- Create transformers (NO IMPUTER NEEDED!) ---
numerical_transformer = Pipeline(steps=[
    ('scaler', StandardScaler())
])

categorical_transformer = Pipeline(steps=[
    ('onehot', OneHotEncoder(handle_unknown='ignore'))
])

# --- Combine into one master preprocessor ---
preprocessor = ColumnTransformer(
    transformers=[
        ('num', numerical_transformer, numeric_features),
```

```
        ('cat', categorical_transformer, categorical_features)
    ])

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

# --- Algorithm 1: Logistic Regression ---
# We tell it how to handle multi-class and the imbalance
model_1_lr = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('model', LogisticRegression(multi_class='ovr', class_weight='balanced', max_iter=1000))
])

# --- Algorithm 2: Random Forest ---
model_2_rf = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('model', RandomForestClassifier(class_weight='balanced', random_state=42))
])

# --- Fit both models ---
print("Fitting Logistic Regression...")
model_1_lr.fit(X_train, y_train)
print("Fitting Random Forest...")
model_2_rf.fit(X_train, y_train)
```

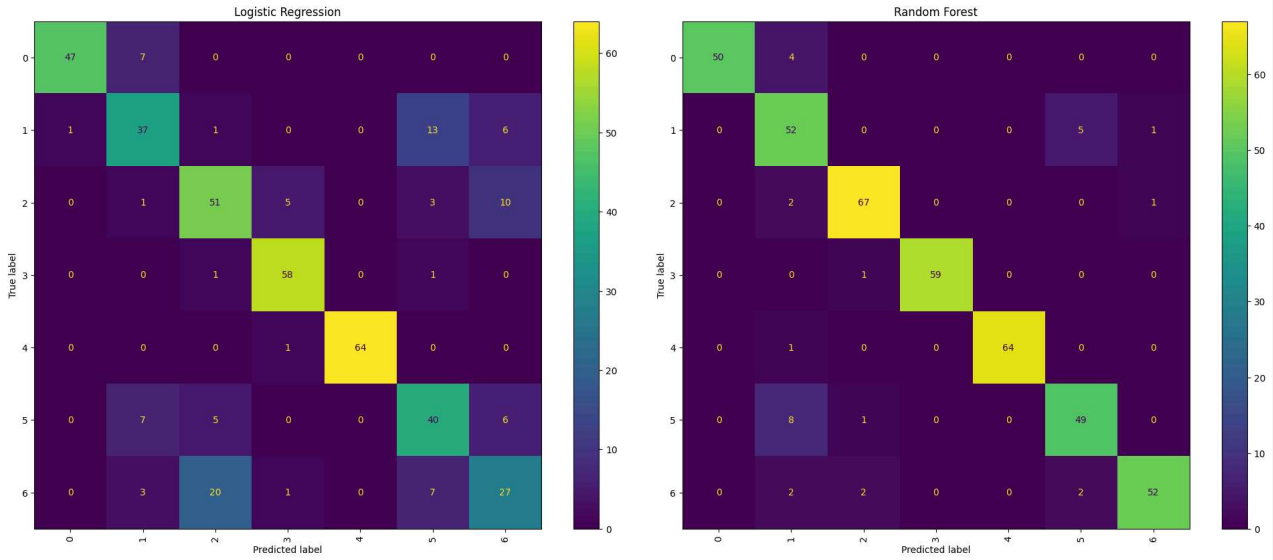
```
Fitting Logistic Regression...
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1256: FutureWarning: 'multi_class' was deprecated
  warnings.warn(
Fitting Random Forest...
```




```
y_pred_lr = model_1_lr.predict(X_test)
y_pred_rf = model_2_rf.predict(X_test)

# --- 1. Classification Report ---
# This is now your *most important* evaluation.
print("--- Logistic Regression Report ---")
print(classification_report(y_test, y_pred_lr))

print("--- Random Forest Report ---")
print(classification_report(y_test, y_pred_rf))
```

```
--- Logistic Regression Report ---
              precision    recall  f1-score   support

           0       0.98      0.87      0.92        54
           1       0.67      0.64      0.65        58
           2       0.65      0.73      0.69        70
           3       0.89      0.97      0.93        60
           4       1.00      0.98      0.99        65
           5       0.62      0.69      0.66        58
           6       0.55      0.47      0.50        58

    accuracy                           0.77       423
   macro avg       0.77      0.76      0.76       423
weighted avg       0.77      0.77      0.76       423

--- Random Forest Report ---
              precision    recall  f1-score   support

           0       1.00      0.93      0.96        54
           1       0.75      0.90      0.82        58
           2       0.94      0.96      0.95        70
           3       1.00      0.98      0.99        60
           4       1.00      0.98      0.99        65
           5       0.88      0.84      0.86        58
           6       0.96      0.90      0.93        58

    accuracy                           0.93       423
   macro avg       0.93      0.93      0.93       423
weighted avg       0.93      0.93      0.93       423
```

```
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(20, 8))

# --- Plot for Model 1 ---
ConfusionMatrixDisplay.from_predictions(y_test, y_pred_lr, ax=ax1)
ax1.set_title('Logistic Regression')
ax1.set_xticklabels(ax1.get_xticklabels(), rotation=90)

# --- Plot for Model 2 ---
ConfusionMatrixDisplay.from_predictions(y_test, y_pred_rf, ax=ax2)
ax2.set_title('Random Forest')
ax2.set_xticklabels(ax2.get_xticklabels(), rotation=90)

plt.tight_layout()
plt.show()
```



Start coding or generate with AI.