```python
# ==========================
# Install PySpark and Import Libraries
# ==========================

# Install PySpark
!pip install pyspark --quiet

# Import necessary libraries
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.functions import col
from pyspark.ml.feature import StringIndexer, OneHotEncoder, VectorAssembler, StandardScaler
from pyspark.ml import Pipeline, PipelineModel
from pyspark.ml.classification import (
    LogisticRegression, RandomForestClassifier,
MultilayerPerceptronClassifier,
    LogisticRegressionModel, RandomForestClassificationModel,
MultilayerPerceptronClassificationModel
)
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
import shutil
import os


import urllib.request

# ==========================
# Start Spark Session
# ==========================

# Initialize a Spark session
spark = SparkSession.builder.appName("HeartAttackPrediction").getOrCreate()

# ==========================
# Download the dataset
# ==========================

url = "https://sp8138-heart-attack-dataset.s3.us-east-2.amazonaws.com/heart_attack_data.csv"
local_file = "heart_attack_data.csv"
urllib.request.urlretrieve(url, local_file)

('heart_attack_data.csv', <http.client.HTTPMessage at 0x79da85551b10>)

# ==========================
# Load and Preprocess Data
# ==========================

# Load the dataset
```

```python
df = spark.read.csv("heart_attack_data.csv", header=True,
inferSchema=True)

# Drop rows with missing values
df = df.dropna()

# Display the schema of the dataset
df.printSchema()

root
 |-- Age: integer (nullable = true)
 |-- Gender: string (nullable = true)
 |-- Cholesterol: integer (nullable = true)
 |-- BloodPressure: integer (nullable = true)
 |-- HeartRate: integer (nullable = true)
 |-- BMI: double (nullable = true)
 |-- Smoker: integer (nullable = true)
 |-- Diabetes: integer (nullable = true)
 |-- Hypertension: integer (nullable = true)
 |-- FamilyHistory: integer (nullable = true)
 |-- PhysicalActivity: integer (nullable = true)
 |-- AlcoholConsumption: integer (nullable = true)
 |-- Diet: string (nullable = true)
 |-- StressLevel: integer (nullable = true)
 |-- Ethnicity: string (nullable = true)
 |-- Income: integer (nullable = true)
 |-- EducationLevel: string (nullable = true)
 |-- Medication: string (nullable = true)
 |-- ChestPainType: string (nullable = true)
 |-- ECGResults: string (nullable = true)
 |-- MaxHeartRate: integer (nullable = true)
 |-- ST_Depression: double (nullable = true)
 |-- ExerciseInducedAngina: string (nullable = true)
 |-- Slope: string (nullable = true)
 |-- NumberOfMajorVessels: integer (nullable = true)
 |-- Thalassemia: string (nullable = true)
 |-- PreviousHeartAttack: integer (nullable = true)
 |-- StrokeHistory: integer (nullable = true)
 |-- Residence: string (nullable = true)
 |-- EmploymentStatus: string (nullable = true)
 |-- MaritalStatus: string (nullable = true)
 |-- Outcome: string (nullable = true)


# ==========================
# Select Features
# ==========================

# Define the target column and selected features
target_col = "Outcome"
```

```python
selected_numeric = [
    "MaxHeartRate", "PhysicalActivity", "Hypertension",
"FamilyHistory", "StressLevel",
    "Smoker", "AlcoholConsumption", "Cholesterol", "StrokeHistory",
"ST_Depression",
    "NumberOfMajorVessels", "BMI", "PreviousHeartAttack", "HeartRate",
"Diabetes"
]
selected_categorical = [
    "Gender", "Diet", "Ethnicity", "EducationLevel", "Medication",
"ChestPainType",
    "ECGResults", "ExerciseInducedAngina", "Slope", "Thalassemia",
"Residence", "MaritalStatus"
]

categorical_cols = selected_categorical
numeric_cols = selected_numeric

# =========================
# Preprocessing Pipeline
# =========================

# StringIndexer for categorical columns (including target)
indexers = [StringIndexer(inputCol=col, outputCol=col+"_idx",
handleInvalid="keep") for col in categorical_cols + [target_col]]

# OneHotEncoder for categorical features (not target)
encoder = OneHotEncoder(inputCols=[col+"_idx" for col in
categorical_cols],
                        outputCols=[col+"_ohe" for col in
categorical_cols])

# Assemble features
feature_cols = [col+"_ohe" for col in categorical_cols] + numeric_cols
assembler = VectorAssembler(inputCols=feature_cols,
outputCol="features_unscaled")

# StandardScaler
scaler = StandardScaler(inputCol="features_unscaled",
outputCol="features", withMean=True, withStd=True)

# Create the pipeline
pipeline = Pipeline(stages=indexers + [encoder, assembler, scaler])
pipeline_model = pipeline.fit(df)

# Transform the dataset
df_prep = pipeline_model.transform(df)

# Select final features and label
```

```python
df_final = df_prep.select("features",
col("Outcome_idx").alias("label"))

# ==========================
# Train/Test Split
# ==========================

# Split the data into training and testing sets
train, test = df_final.randomSplit([0.8, 0.2], seed=42)

# Display the sizes of the splits
print(f"Training set size: {train.count()}")
print(f"Testing set size: {test.count()}")

Training set size: 298307
Testing set size: 74667

# ==========================
# Train Models
# ==========================

# Logistic Regression
lr = LogisticRegression(featuresCol="features", labelCol="label",
maxIter=100)
lr_model = lr.fit(train)

# Random Forest
rf = RandomForestClassifier(featuresCol="features", labelCol="label",
numTrees=100, seed=42)
rf_model = rf.fit(train)

# Multilayer Perceptron (ANN)
input_size = train.first()['features'].size
layers = [input_size, 16, 8, 2]  # 2 output classes
mlp = MultilayerPerceptronClassifier(featuresCol="features",
labelCol="label", maxIter=100, layers=layers, seed=42)
mlp_model = mlp.fit(train)

# ==========================
# Save Models and Pipeline
# ==========================

# Create a directory for saving models
os.makedirs("saved_models", exist_ok=True)

# Remove existing directories if they exist
for path in [
    "saved_models/preprocessing_pipeline",
    "saved_models/logistic_regression",
    "saved_models/random_forest",
    "saved_models/ann_mlp"
```

```
]:
    if os.path.exists(path):
        shutil.rmtree(path)

# Save the pipeline and models
pipeline_model.save("saved_models/preprocessing_pipeline")
lr_model.save("saved_models/logistic_regression")
rf_model.save("saved_models/random_forest")
mlp_model.save("saved_models/ann_mlp")

print("Models and pipeline saved in 'saved_models/' directory.")

Models and pipeline saved in 'saved_models/' directory.

# ==========================
# Evaluate Models
# ==========================

# Function to evaluate a model
def evaluate_model(model, test, name):
    predictions = model.transform(test)
    evaluator = MulticlassClassificationEvaluator(labelCol="label",
predictionCol="prediction", metricName="accuracy")
    acc = evaluator.evaluate(predictions)
    print(f"\n{name} Accuracy: {acc:.3f}")
    pred_pd = predictions.select("label", "prediction").toPandas()
    from sklearn.metrics import classification_report,
confusion_matrix
    print(classification_report(pred_pd['label'],
pred_pd['prediction']))
    print("Confusion Matrix:\n", confusion_matrix(pred_pd['label'],
pred_pd['prediction']))

# Evaluate all models
evaluate_model(lr_model, test, "Logistic Regression")
evaluate_model(rf_model, test, "Random Forest")
evaluate_model(mlp_model, test, "Multilayer Perceptron (ANN)")


Logistic Regression Accuracy: 0.501
              precision    recall  f1-score   support

         0.0       0.50      0.55      0.52     37271
         1.0       0.50      0.45      0.47     37396

    accuracy                           0.50     74667
   macro avg       0.50      0.50      0.50     74667
weighted avg       0.50      0.50      0.50     74667

Confusion Matrix:
 [[20584 16687]
```

```
 [20573 16823]]

Random Forest Accuracy: 0.500
              precision    recall  f1-score   support

         0.0       0.50      0.61      0.55     37271
         1.0       0.50      0.39      0.44     37396

    accuracy                           0.50     74667
   macro avg       0.50      0.50      0.49     74667
weighted avg       0.50      0.50      0.49     74667

Confusion Matrix:
 [[22878 14393]
 [22908 14488]]

Multilayer Perceptron (ANN) Accuracy: 0.500
              precision    recall  f1-score   support

         0.0       0.50      0.51      0.51     37271
         1.0       0.50      0.49      0.49     37396

    accuracy                           0.50     74667
   macro avg       0.50      0.50      0.50     74667
weighted avg       0.50      0.50      0.50     74667

Confusion Matrix:
 [[19070 18201]
 [19113 18283]]
```

```python
# ==========================
# Load Models and Pipeline
# ==========================

# Load the saved pipeline and models
loaded_pipeline =
PipelineModel.load("saved_models/preprocessing_pipeline")
loaded_lr =
LogisticRegressionModel.load("saved_models/logistic_regression")
loaded_rf =
RandomForestClassificationModel.load("saved_models/random_forest")
loaded_mlp =
MultilayerPerceptronClassificationModel.load("saved_models/ann_mlp")

print("Models and pipeline loaded successfully.")
```

```
Models and pipeline loaded successfully.
```

```python
# ==========================
# Inference on New Case
# ==========================
```

```python
# Function to predict a new case
def predict_new_case(new_case_dict):
    new_df = spark.createDataFrame([new_case_dict])
    new_df_prep = loaded_pipeline.transform(new_df)
    new_df_final = new_df_prep.select("features")
    print("\n--- Inference on New Case ---")
    for name, model in [("Logistic Regression", loaded_lr), ("Random
Forest", loaded_rf), ("ANN", loaded_mlp)]:
        pred =
model.transform(new_df_final).select("prediction").collect()[0][0]
        outcome = "Heart Attack" if pred == 1.0 else "No Heart Attack"
        print(f"{name}: {outcome}")

# Example new case
new_case = {
    'Gender': 'Male',
    'Diet': 'Moderate',
    'Ethnicity': 'White',
    'EducationLevel': 'College',
    'Medication': 'Yes',
    'ChestPainType': 'Typical',
    'ECGResults': 'Normal',
    'ExerciseInducedAngina': 'No',
    'Slope': 'Flat',
    'Thalassemia': 'Normal',
    'Residence': 'Urban',
    'MaritalStatus': 'Married',
    'MaxHeartRate': 150,
    'PhysicalActivity': 2,
    'Hypertension': 1,
    'FamilyHistory': 1,
    'StressLevel': 2,
    'Smoker': 1,
    'AlcoholConsumption': 1,
    'Cholesterol': 200,
    'StrokeHistory': 0,
    'ST_Depression': 1.2,
    'NumberOfMajorVessels': 0,
    'BMI': 27.5,
    'PreviousHeartAttack': 0,
    'HeartRate': 80,
    'Diabetes': 0
}
predict_new_case(new_case)


--- Inference on New Case ---
Logistic Regression: Heart Attack
```

```
Random Forest: No Heart Attack
ANN: No Heart Attack
```

```python
# Zip the preprocessing pipeline
!zip -r preprocessing_pipeline.zip saved_models/preprocessing_pipeline

# Zip the random forest model
!zip -r random_forest.zip saved_models/random_forest
```

```
  adding: saved_models/preprocessing_pipeline/ (stored 0%)
  adding: saved_models/preprocessing_pipeline/stages/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/.part-00000-7f3c6160-bd8c-4364-b238-8446361dd4b0-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/part-00000-7f3c6160-bd8c-4364-b238-8446361dd4b0-
c000.snappy.parquet (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
```

```
41/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/part-00000 (deflated 51%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/.part-00000-e4381f2c-4c94-45e4-aa25-223a07c05056-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/part-00000-e4381f2c-4c94-45e4-aa25-223a07c05056-
c000.snappy.parquet (deflated 27%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/part-00000 (deflated 39%)
  adding:
```

```
  saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/part-00000-b48f0dc5-1f2b-44d1-aba0-d862484cb40f-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/.part-00000-b48f0dc5-1f2b-44d1-aba0-d862484cb40f-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/part-00000-8827bddc-93bc-4912-9e9b-a778f85a2b3d-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
```

```
3b/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/.part-00000-8827bddc-93bc-4912-9e9b-a778f85a2b3d-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/.part-00000-24794b02-1d92-403a-bb3d-d84de9afc60a-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/part-00000-24794b02-1d92-403a-bb3d-d84de9afc60a-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/._SUCCESS.crc (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/part-00000 (deflated 41%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/part-00000-b5e9ddc5-d060-4a43-adc2-a9ff492dad6e-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/.part-00000-b5e9ddc5-d060-4a43-adc2-a9ff492dad6e-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/part-00000 (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
```

```
c5/data/.part-00000-f2b3bb88-d619-48a0-82a7-e4e31b35f6e2-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/part-00000-f2b3bb88-d619-48a0-82a7-e4e31b35f6e2-
c000.snappy.parquet (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/.part-00000-40c1dded-9341-42ac-88b4-5f1ad1271e28-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/part-00000-40c1dded-9341-42ac-88b4-5f1ad1271e28-
c000.snappy.parquet (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
```

```
5d/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/part-00000-a73379a3-d558-48f3-b2ce-c8f4be834029-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/.part-00000-a73379a3-d558-48f3-b2ce-c8f4be834029-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/part-00000 (deflated 40%)
  adding:
```

```
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/part-00000-4427ea57-d86f-4326-90c4-4b6dc9c078f3-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/.part-00000-4427ea57-d86f-4326-90c4-4b6dc9c078f3-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/.part-00000-b3e4442c-e266-42fc-b643-a0ec261492ee-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
```

```
fe/data/part-00000-b3e4442c-e266-42fc-b643-a0ec261492ee-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/part-00000-b2448eff-e3b6-44e3-937b-e4c3eb575592-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/.part-00000-b2448eff-e3b6-44e3-937b-e4c3eb575592-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/._SUCCESS.crc (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/part-00000 (deflated 43%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/part-00000 (deflated 44%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/part-00000-1dfa2ed0-810b-4d47-9715-1f425d8cefb5-
c000.snappy.parquet (deflated 52%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/.part-00000-1dfa2ed0-810b-4d47-9715-1f425d8cefb5-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
```

```
71/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/part-00000 (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/part-00000-9891e2c7-c76d-4adc-8e8f-1895b11e8732-
c000.snappy.parquet (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/.part-00000-9891e2c7-c76d-4adc-8e8f-1895b11e8732-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
```

```
ea/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/part-00000-3da2fc21-a920-4693-8bd8-e2ebdc5fdcc6-
c000.snappy.parquet (deflated 37%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/.part-00000-3da2fc21-a920-4693-8bd8-e2ebdc5fdcc6-
c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/ (stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/._SUCCESS.crc
(stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/.part-00000.crc
(stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/part-00000
(deflated 48%)
  adding: saved_models/preprocessing_pipeline/metadata/_SUCCESS
(stored 0%)
  adding: saved_models/random_forest/ (stored 0%)
  adding: saved_models/random_forest/metadata/ (stored 0%)
  adding: saved_models/random_forest/metadata/._SUCCESS.crc (stored
0%)
  adding: saved_models/random_forest/metadata/.part-00000.crc (stored
0%)
  adding: saved_models/random_forest/metadata/part-00000 (deflated
45%)
  adding: saved_models/random_forest/metadata/_SUCCESS (stored 0%)
  adding: saved_models/random_forest/data/ (stored 0%)
  adding: saved_models/random_forest/data/._SUCCESS.crc (stored 0%)
  adding: saved_models/random_forest/data/.part-00000-8c8a3915-2fd2-
40e1-8d5d-02973768f7f8-c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/random_forest/data/_SUCCESS (stored 0%)
  adding: saved_models/random_forest/data/part-00000-8c8a3915-2fd2-
40e1-8d5d-02973768f7f8-c000.snappy.parquet (deflated 16%)
  adding: saved_models/random_forest/treesMetadata/ (stored 0%)
  adding: saved_models/random_forest/treesMetadata/._SUCCESS.crc
(stored 0%)
  adding: saved_models/random_forest/treesMetadata/.part-00000-
d73d3745-5fa6-4a0b-bcb5-cf8c503d0f29-c000.snappy.parquet.crc (stored
0%)
  adding: saved_models/random_forest/treesMetadata/_SUCCESS (stored
0%)
  adding: saved_models/random_forest/treesMetadata/part-00000-
```

```
d73d3745-5fa6-4a0b-bcb5-cf8c503d0f29-c000.snappy.parquet (deflated
28%)
```

```
!zip -r saved_models.zip saved_models
```

```
  adding: saved_models/ (stored 0%)
  adding: saved_models/ann_mlp/ (stored 0%)
  adding: saved_models/ann_mlp/metadata/ (stored 0%)
  adding: saved_models/ann_mlp/metadata/._SUCCESS.crc (stored 0%)
  adding: saved_models/ann_mlp/metadata/.part-00000.crc (stored 0%)
  adding: saved_models/ann_mlp/metadata/part-00000 (deflated 46%)
  adding: saved_models/ann_mlp/metadata/_SUCCESS (stored 0%)
  adding: saved_models/ann_mlp/data/ (stored 0%)
  adding: saved_models/ann_mlp/data/._SUCCESS.crc (stored 0%)
  adding: saved_models/ann_mlp/data/.part-00000-70172737-22d9-4466-
97ba-0aaf77749a96-c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/ann_mlp/data/_SUCCESS (stored 0%)
  adding: saved_models/ann_mlp/data/part-00000-70172737-22d9-4466-
97ba-0aaf77749a96-c000.snappy.parquet (deflated 10%)
  adding: saved_models/logistic_regression/ (stored 0%)
  adding: saved_models/logistic_regression/metadata/ (stored 0%)
  adding: saved_models/logistic_regression/metadata/._SUCCESS.crc
(stored 0%)
  adding: saved_models/logistic_regression/metadata/.part-00000.crc
(stored 0%)
  adding: saved_models/logistic_regression/metadata/part-00000
(deflated 45%)
  adding: saved_models/logistic_regression/metadata/_SUCCESS (stored
0%)
  adding: saved_models/logistic_regression/data/ (stored 0%)
  adding: saved_models/logistic_regression/data/.part-00000-5da14580-
10a4-4d61-8463-8243458a4314-c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/logistic_regression/data/._SUCCESS.crc (stored
0%)
  adding: saved_models/logistic_regression/data/part-00000-5da14580-
10a4-4d61-8463-8243458a4314-c000.snappy.parquet (deflated 54%)
  adding: saved_models/logistic_regression/data/_SUCCESS (stored 0%)
  adding: saved_models/random_forest/ (stored 0%)
  adding: saved_models/random_forest/metadata/ (stored 0%)
  adding: saved_models/random_forest/metadata/._SUCCESS.crc (stored
0%)
  adding: saved_models/random_forest/metadata/.part-00000.crc (stored
0%)
  adding: saved_models/random_forest/metadata/part-00000 (deflated
45%)
  adding: saved_models/random_forest/metadata/_SUCCESS (stored 0%)
  adding: saved_models/random_forest/data/ (stored 0%)
  adding: saved_models/random_forest/data/._SUCCESS.crc (stored 0%)
```

```
  adding: saved_models/random_forest/data/.part-00000-8c8a3915-2fd2-
40e1-8d5d-02973768f7f8-c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/random_forest/data/_SUCCESS (stored 0%)
  adding: saved_models/random_forest/data/part-00000-8c8a3915-2fd2-
40e1-8d5d-02973768f7f8-c000.snappy.parquet (deflated 16%)
  adding: saved_models/random_forest/treesMetadata/ (stored 0%)
  adding: saved_models/random_forest/treesMetadata/._SUCCESS.crc
(stored 0%)
  adding: saved_models/random_forest/treesMetadata/.part-00000-
d73d3745-5fa6-4a0b-bcb5-cf8c503d0f29-c000.snappy.parquet.crc (stored
0%)
  adding: saved_models/random_forest/treesMetadata/_SUCCESS (stored
0%)
  adding: saved_models/random_forest/treesMetadata/part-00000-
d73d3745-5fa6-4a0b-bcb5-cf8c503d0f29-c000.snappy.parquet (deflated
28%)
  adding: saved_models/preprocessing_pipeline/ (stored 0%)
  adding: saved_models/preprocessing_pipeline/stages/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/.part-00000-7f3c6160-bd8c-4364-b238-8446361dd4b0-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/_SUCCESS (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/09_StringIndexer_2bb3e04fd3
8d/data/part-00000-7f3c6160-bd8c-4364-b238-8446361dd4b0-
c000.snappy.parquet (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/part-00000 (deflated 51%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/.part-00000-e4381f2c-4c94-45e4-aa25-223a07c05056-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/part-00000-e4381f2c-4c94-45e4-aa25-223a07c05056-
c000.snappy.parquet (deflated 27%)
  adding:
saved_models/preprocessing_pipeline/stages/13_OneHotEncoder_68e48903e2
41/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/._SUCCESS.crc (stored 0%)
  adding:
```

```
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/part-00000-b48f0dc5-1f2b-44d1-aba0-d862484cb40f-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/02_StringIndexer_36b5d91675
06/data/.part-00000-b48f0dc5-1f2b-44d1-aba0-d862484cb40f-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/part-00000-8827bddc-93bc-4912-9e9b-a778f85a2b3d-
```

```
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/05_StringIndexer_e8026a363d
3b/data/.part-00000-8827bddc-93bc-4912-9e9b-a778f85a2b3d-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/.part-00000-24794b02-1d92-403a-bb3d-d84de9afc60a-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/part-00000-24794b02-1d92-403a-bb3d-d84de9afc60a-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/04_StringIndexer_beace909bc
f3/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/ (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/part-00000 (deflated 41%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/part-00000-b5e9ddc5-d060-4a43-adc2-a9ff492dad6e-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/.part-00000-b5e9ddc5-d060-4a43-adc2-a9ff492dad6e-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/07_StringIndexer_819ecfcfe3
0a/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/metadata/part-00000 (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
```

```
c5/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/.part-00000-f2b3bb88-d619-48a0-82a7-e4e31b35f6e2-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/part-00000-f2b3bb88-d619-48a0-82a7-e4e31b35f6e2-
c000.snappy.parquet (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/08_StringIndexer_70a24beabd
c5/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/.part-00000-40c1dded-9341-42ac-88b4-5f1ad1271e28-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/_SUCCESS (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/06_StringIndexer_faf04ba8b1
dd/data/part-00000-40c1dded-9341-42ac-88b4-5f1ad1271e28-
c000.snappy.parquet (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/part-00000-a73379a3-d558-48f3-b2ce-c8f4be834029-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/.part-00000-a73379a3-d558-48f3-b2ce-c8f4be834029-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/00_StringIndexer_4980cf6cf0
5d/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/._SUCCESS.crc (stored 0%)
  adding:
```

```
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/part-00000 (deflated 40%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/part-00000-4427ea57-d86f-4326-90c4-4b6dc9c078f3-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/.part-00000-4427ea57-d86f-4326-90c4-4b6dc9c078f3-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/11_StringIndexer_1569dee128
48/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/._SUCCESS.crc (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/.part-00000-b3e4442c-e266-42fc-b643-a0ec261492ee-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/part-00000-b3e4442c-e266-42fc-b643-a0ec261492ee-
c000.snappy.parquet (deflated 35%)
  adding:
saved_models/preprocessing_pipeline/stages/10_StringIndexer_3abedc96f3
fe/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/part-00000-b2448eff-e3b6-44e3-937b-e4c3eb575592-
c000.snappy.parquet (deflated 36%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/.part-00000-b2448eff-e3b6-44e3-937b-e4c3eb575592-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/03_StringIndexer_ebbec6cf03
26/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/ (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/part-00000 (deflated 43%)
  adding:
saved_models/preprocessing_pipeline/stages/14_VectorAssembler_ddece075
5f5b/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/part-00000 (deflated 44%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/part-00000-1dfa2ed0-810b-4d47-9715-1f425d8cefb5-
c000.snappy.parquet (deflated 52%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/.part-00000-1dfa2ed0-810b-4d47-9715-1f425d8cefb5-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/15_StandardScaler_d677d1279
558/data/_SUCCESS (stored 0%)
```

```
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/part-00000 (deflated 38%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/part-00000-9891e2c7-c76d-4adc-8e8f-1895b11e8732-
c000.snappy.parquet (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/.part-00000-9891e2c7-c76d-4adc-8e8f-1895b11e8732-
c000.snappy.parquet.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/12_StringIndexer_b40ed0163b
71/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/.part-00000.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
```

```
ea/metadata/part-00000 (deflated 39%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/metadata/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/ (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/part-00000-3da2fc21-a920-4693-8bd8-e2ebdc5fdcc6-
c000.snappy.parquet (deflated 37%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/._SUCCESS.crc (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/_SUCCESS (stored 0%)
  adding:
saved_models/preprocessing_pipeline/stages/01_StringIndexer_b2fe8fd09d
ea/data/.part-00000-3da2fc21-a920-4693-8bd8-e2ebdc5fdcc6-
c000.snappy.parquet.crc (stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/ (stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/._SUCCESS.crc
(stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/.part-00000.crc
(stored 0%)
  adding: saved_models/preprocessing_pipeline/metadata/part-00000
(deflated 48%)
  adding: saved_models/preprocessing_pipeline/metadata/_SUCCESS
(stored 0%)
```