Data Preprocessing

Data preprocessing is necessary since there might be some missing values or categorical variables that need to be converted to numerical. Also, there is a need for data normalization since KNN and SVM use the absolute measurement.

```
In [1]: !pip install category encoders
        Collecting category encoders
          Obtaining dependency information for category encoders from https://files.pythonhoste
        d.org/packages/7f/e5/79a62e5c9c9ddbfa9ff5222240d408c1eeea4e38741a0dc8343edc7ef1ec/catego
        ry encoders-2.6.3-py2.py3-none-any.whl.metadata
          Downloading category encoders-2.6.3-py2.py3-none-any.whl.metadata (8.0 kB)
        Requirement already satisfied: numpy>=1.14.0 in /Users/senaozb/anaconda3/lib/python3.11/
        site-packages (from category encoders) (1.24.3)
        Requirement already satisfied: scikit-learn>=0.20.0 in /Users/senaozb/anaconda3/lib/pyth
        on3.11/site-packages (from category encoders) (1.3.0)
        Requirement already satisfied: scipy>=1.0.0 in /Users/senaozb/anaconda3/lib/python3.11/s
        ite-packages (from category encoders) (1.11.1)
        Requirement already satisfied: statsmodels>=0.9.0 in /Users/senaozb/anaconda3/lib/python
        3.11/site-packages (from category encoders) (0.14.0)
        Requirement already satisfied: pandas>=1.0.5 in /Users/senaozb/anaconda3/lib/python3.11/
        site-packages (from category encoders) (2.0.3)
        Requirement already satisfied: patsy>=0.5.1 in /Users/senaozb/anaconda3/lib/python3.11/s
        ite-packages (from category encoders) (0.5.3)
        Requirement already satisfied: python-dateutil>=2.8.2 in /Users/senaozb/anaconda3/lib/py
        thon3.11/site-packages (from pandas>=1.0.5->category encoders) (2.8.2)
        Requirement already satisfied: pytz>=2020.1 in /Users/senaozb/anaconda3/lib/python3.11/s
        ite-packages (from pandas>=1.0.5->category encoders) (2023.3.post1)
        Requirement already satisfied: tzdata>=2022.1 in /Users/senaozb/anaconda3/lib/python3.1
        1/site-packages (from pandas>=1.0.5->category encoders) (2023.3)
        Requirement already satisfied: six in /Users/senaozb/anaconda3/lib/python3.11/site-packa
        ges (from patsy>=0.5.1->category encoders) (1.16.0)
        Requirement already satisfied: joblib>=1.1.1 in /Users/senaozb/anaconda3/lib/python3.11/
        site-packages (from scikit-learn>=0.20.0->category encoders) (1.2.0)
        Requirement already satisfied: threadpoolctl>=2.0.0 in /Users/senaozb/anaconda3/lib/pyth
        on3.11/site-packages (from scikit-learn>=0.20.0->category encoders) (2.2.0)
        Requirement already satisfied: packaging>=21.3 in /Users/senaozb/anaconda3/lib/python3.1
        1/site-packages (from statsmodels>=0.9.0->category encoders) (23.1)
        Downloading category encoders-2.6.3-py2.py3-none-any.whl (81 kB)
                                                  - 81.9/81.9 kB 813.1 kB/s eta 0:00:00a 0:00:01
        Installing collected packages: category encoders
        Successfully installed category encoders-2.6.3
In [2]: # Import libraries
        import pandas as pd
        import category encoders as ce
        from sklearn.preprocessing import StandardScaler
        import numpy as np
        from collections import Counter
        from sklearn.model selection import KFold
        from sklearn.metrics import mean absolute error, roc curve, auc
        import time
        from sklearn.svm import SVC, SVR
        from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor
        import matplotlib.pyplot as plt
```

In [4]: # Read the data and print the information
 audit_risk = pd.read_csv("audit_risk.csv")

```
print(audit risk.head())
print(audit risk.info())
     Sector score LOCATION ID PARA A Score A Risk A PARA B Score B Risk B \
     3.89 23 4.18 0.6 2.508 2.50 0.2 0.500
                                          6 0.00 0.2 0.000 4.83
6 0.51 0.2 0.102 0.23
6 0.00 0.2 0.000 10.80
6 0.00 0.2 0.000 0.08
                                                                                                                  0.2 0.966
                  3.89
1
2
                   3.89
                                                                                                                  0.2 0.046
3
                  3.89
                                                                                                                  0.6 6.480
                                                                                                             0.2 0.016
                                    6 0.00
                  3.89
     TOTAL numbers ... RiSk E History Prob Risk F Score Inherent Risk \

      0
      6.68
      5.0
      ...
      0.4
      0
      0.2
      0.0
      2.4
      8.574

      1
      4.83
      5.0
      ...
      0.4
      0
      0.2
      0.0
      2.0
      2.554

      2
      0.74
      5.0
      ...
      0.4
      0
      0.2
      0.0
      2.0
      1.548

      3
      10.80
      6.0
      ...
      0.4
      0
      0.2
      0.0
      4.4
      17.530

      4
      0.08
      5.0
      ...
      0.4
      0
      0.2
      0.0
      2.0
      1.416

     CONTROL RISK Detection Risk Audit Risk Risk
0
                   0.4 0.5 1.7148
1
                     0.4
                                               0.5 0.5108
2
                     0.4
                                               0.5
                                                               0.3096
                                                               3.5060
3
                     0.4
                                               0.5
                                               0.5 0.2832 0
                     0.4
[5 rows x 27 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 776 entries, 0 to 775
Data columns (total 27 columns):
  # Column Non-Null Count Dtype
O Sector_score 776 non-null float64
1 LOCATION_ID 776 non-null object
2 PARA_A 776 non-null float64
3 Score_A 776 non-null float64
4 Risk_A 776 non-null float64
5 PARA_B 776 non-null float64
6 Score_B 776 non-null float64
7 Risk_B 776 non-null float64
8 TOTAL 776 non-null float64
9 numbers 776 non-null float64
10 Score_B.1 776 non-null float64
11 Risk_C 776 non-null float64
12 Money_Value 775 non-null float64
13 Score_MV 776 non-null float64
14 Risk_D 776 non-null float64
15 District_Loss 776 non-null float64
16 PROB 776 non-null float64
                                  ----
 16 PROB 776 non-null float64
17 RiSk_E 776 non-null float64
18 History 776 non-null int64
19 Prob 776 non-null float64
20 Risk_F 776 non-null float64
21 Score 776 non-null float64
  22 Inherent Risk 776 non-null float64
  23 CONTROL RISK 776 non-null float64
  24 Detection_Risk 776 non-null float64
  25 Audit Risk 776 non-null float64
  26 Risk
                                    776 non-null int64
dtypes: float64(23), int64(3), object(1)
memory usage: 163.8+ KB
None
```

```
In [5]: # Check for null variables and replace with the mean value
    print(audit_risk.isnull().sum())
    mean_value = audit_risk['Money_Value'].mean()
    audit_risk['Money_Value'] = audit_risk['Money_Value'].fillna(mean_value)
```

```
Sector score
       LOCATION ID
       PARA A
                       0
                      0
       Score A
                       0
       Risk A
       PARA B
                       0
       Score B
       Risk B
                      0
       TOTAL
       numbers
                      0
                      0
       Score B.1
       Risk_C
       Money_Value
                       1
       Score MV
       Risk D
       District_Loss 0
       PROB
                       0
       RiSk E
       History
                       0
       Prob
       Risk F
       Score
       Inherent Risk
       CONTROL RISK
                      0
       Detection Risk
                       0
       Audit Risk
       Risk
                       0
       dtype: int64
In [6]: # Show the categories of the categorical variable
       audit_risk["LOCATION_ID"].value_counts()
       LOCATION ID
       19
                68
       9
                53
       16
                52
       12
                47
       5
                44
       2
                41
                37
       4
       15
                35
       13
                35
                33
       6
```

Out[6]:

```
40
                    3
        35
        44
        NUH
                    1
        LOHARU
        SAFIDON
        2.3
                    1
        42
        41
        34
                    1
        33
                   1
        2.4
                    1
        17
                    1
        Name: count, dtype: int64
In [7]: # Encode it to numerical
        encoder = ce.TargetEncoder()
        audit risk['LOCATION ID'] = encoder.fit transform(audit risk['LOCATION ID'], audit risk[
In [8]: # There is no need for the audit risk since it affects the result
        audit risk.drop('Audit Risk', axis=1, inplace=True)
In [9]:
        # Normalize the values
        columns to normalize audit = [col for col in audit risk.columns if col != "Risk"]
        scaler = StandardScaler()
        normalized data = scaler.fit transform(audit risk[columns to normalize audit])
        audit normalized = pd.DataFrame(normalized data, columns=columns to normalize audit)
        audit normalized["Risk"] = audit risk["Risk"]
        print(audit normalized.head())
        print(audit normalized.info())
           Sector score LOCATION ID PARA A Score A Risk A PARA B \
             -0.670465 0.753175 0.304800 1.429846 0.336502 -1.658295e-01
              -0.670465
                           -1.742275 -0.431736 -0.869761 -0.392943 -1.192773e-01
              -0.670465 -1.742275 -0.341872 -0.869761 -0.363277 -2.111829e-01
              -0.670465 -1.742275 -0.431736 -0.869761 -0.392943 2.317208e-07
              -0.670465 -1.742275 -0.431736 -0.869761 -0.392943 -2.141798e-01
            Score B Risk B
                                TOTAL numbers ... PROB RiSk E History \
        0 \ -0.666752 \ -0.194121 \ -0.127506 \ -0.255998 \ \dots \ -0.16502 \ -0.410417 \ -0.196691
        1 - 0.666752 - 0.178615 - 0.163583 - 0.255998 \dots - 0.16502 - 0.410417 - 0.196691
        2 - 0.666752 - 0.209227 - 0.243341 - 0.255998 \dots - 0.16502 - 0.410417 - 0.196691
        3 \quad 1.690422 \quad 0.004858 \quad -0.047162 \quad 3.527894 \quad \dots \quad -0.16502 \quad -0.410417 \quad -0.196691
        4 \ -0.666752 \ -0.210226 \ -0.256212 \ -0.255998 \ \dots \ -0.16502 \ -0.410417 \ -0.196691
                     Risk F Score Inherent Risk CONTROL RISK Detection Risk \
               Prob
        0 -0.246568 -0.175398 -0.352503 -0.166468 -0.388662
                                                                                 0.0
        1 -0.246568 -0.175398 -0.818503
                                            -0.276513
                                                          -0.388662
                                                                                 0.0
        2 -0.246568 -0.175398 -0.818503
                                            -0.294902
                                                          -0.388662
                                                                                 0.0
                                                          -0.388662
        3 -0.246568 -0.175398 1.977497
                                            -0.002753
                                                                                 0.0
        4 -0.246568 -0.175398 -0.818503 -0.297315 -0.388662
                                                                                 0.0
           Risk
        \cap
           1
        1
        2
             0
        3
              1
        [5 rows x 26 columns]
        <class 'pandas.core.frame.DataFrame'>
```

3

```
Data columns (total 26 columns):
         # Column Non-Null Count Dtype
        24 Detection Risk 776 non-null float64
         25 Risk 776 non-null
        dtypes: float64(25), int64(1)
        memory usage: 157.8 KB
        None
In [10]:
        # Read the data and print the information
        day = pd.read csv("day.csv")
        print(day.head())
        print(day.info())
          instant dteday season yr mnth holiday weekday workingday \
        0 1 2011-01-01 1 0 1 0 6
               2 2011-01-02
                               1 0 1
1 0 1
                                                 0
                                                         0
               3 2011-01-03
                                                                   1
        2
                                                 0
                                                         1
        3
               4 2011-01-04
                                1 0
                                         1
                                                         2
                                                 0
              5 2011-01-05
                                1 0
                                         1
                                                 0
                                                         3
          weathersit temp atemp hum windspeed casual registered \

      2
      0.344167
      0.363625
      0.805833
      0.160446
      331
      654

      2
      0.363478
      0.353739
      0.696087
      0.248539
      131
      670

        0
        1
                 1 0.196364 0.189405 0.437273 0.248309
                                                          120
                                                                    1229
                 1 0.200000 0.212122 0.590435 0.160296
        3
                                                          108
                                                                    1454
                                                                  1518
                  1 0.226957 0.229270 0.436957 0.186900 82
           cnt
        0 985
        1 801
        2 1349
        3 1562
        4 1600
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 731 entries, 0 to 730
        Data columns (total 16 columns):
         # Column Non-Null Count Dtype
        --- ----
                      _____
           instant 731 non-null int64
```

RangeIndex: 776 entries, 0 to 775

```
3 yr 731 non-null int64
4 mnth 731 non-null int64
5 holiday 731 non-null int64
6 weekday 731 non-null int64
7 workingday 731 non-null int64
             weathersit 731 non-null int64
           8
           9 temp 731 non-null float64
10 atemp 731 non-null float64
          11 hum 731 non-null float64
          12 windspeed 731 non-null float64
          13 casual 731 non-null int64
14 registered 731 non-null int64
          15 cnt 731 non-null int64
          dtypes: float64(4), int64(11), object(1)
          memory usage: 91.5+ KB
          None
In [11]: # Check for null variables
          print(day.isnull().sum())
          instant
          dteday
          season
                       0
          yr
                        0
         mnth
                       0
         holiday
         weekday
         workingday 0
         weathersit 0
                       0
          temp
          atemp
                        0
         hum
         windspeed
                       0
          casual
         registered 0
          cnt
          dtype: int64
In [12]: # Show the categories of the categorical variable
          day["dteday"].value counts()
Out[12]: dteday
         2011-01-01 1
         2012-04-25 1
         2012-04-27
          2012-04-28 1
         2012-04-29 1
         2011-09-03 1
          2011-09-04 1
         2011-09-05 1
          2011-09-06
                     1
          2012-12-31
                        1
         Name: count, Length: 731, dtype: int64
In [13]: # There is no need for the date info since it doesn't carry any necessary info
          day.drop('dteday', axis=1, inplace=True)
In [14]:
          # Normalize the values
          columns to normalize day = [col for col in day.columns if col != "cnt"]
          normalized data = scaler.fit transform(day[columns to normalize day])
          day normalized = pd.DataFrame(normalized data, columns=columns to normalize day)
```

731 non-null object

731 non-null int64

1

dteday 2 season

```
day normalized["cnt"] = day["cnt"]
print(day normalized.head())
print(day normalized.info())
                          yr mnth holiday weekday workingday \setminus
   instant season
0 -1.729683 -1.348213 -1.001369 -1.600161 -0.171981 1.498809 -1.471225
1 - 1.724944 - 1.348213 - 1.001369 - 1.600161 - 0.171981 - 1.496077 - 1.471225
2 -1.720205 -1.348213 -1.001369 -1.600161 -0.171981 -0.996930 0.679706
3 -1.715466 -1.348213 -1.001369 -1.600161 -0.171981 -0.497782 0.679706
4 -1.710728 -1.348213 -1.001369 -1.600161 -0.171981 0.001366 0.679706
                            atemp
                                         hum windspeed casual registered \
   weathersit
                  temp
   1.110427 -0.826662 -0.679946 1.250171 -0.387892 -0.753734 -1.925471
   1.110427 -0.721095 -0.740652 0.479113 0.749602 -1.045214 -1.915209
2 - 0.726048 - 1.634657 - 1.749767 - 1.339274 0.746632 - 1.061246 - 1.556689
3 \quad -0.726048 \quad -1.614780 \quad -1.610270 \quad -0.263182 \quad -0.389829 \quad -1.078734 \quad -1.412383
4 \quad -0.726048 \quad -1.467414 \quad -1.504971 \quad -1.341494 \quad -0.046307 \quad -1.116627 \quad -1.371336
   cnt.
0 985
1 801
2 1349
3 1562
4 1600
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 15 columns):
 # Column Non-Null Count Dtype
                 -----
   instant
                 731 non-null float64
1 season 731 non-null float64
2 yr 731 non-null float64
3 mnth 731 non-null float64
4 holiday 731 non-null float64
5 weekday 731 non-null float64
 6 workingday 731 non-null float64
   weathersit 731 non-null float64
   temp 731 non-null float64
9 atemp 731 non-null float64
10 hum 731 non-null float64
11 windspeed 731 non-null float64
 12 casual 731 non-null float64
13 registered 731 non-null float64
14 cnt 731 non-null int64
dtypes: float64(14), int64(1)
memory usage: 85.8 KB
None
```

Machine Learning Models

```
In [15]: # Prepare the datasets
X_audit, y_audit = audit_normalized[columns_to_normalize_audit], audit_normalized["Risk"
X_day, y_day = day_normalized[columns_to_normalize_day], day_normalized["cnt"]

In [16]: def confusion_matrix_custom(actual, predicted):
    # Compute the confusion matrix for the predictions
    labels = np.unique(np.concatenate((actual, predicted)))
    label_map = {label: i for i, label in enumerate(labels)}

matrix = np.zeros((len(labels), len(labels)), dtype=int)
```

```
for a, p in zip(actual, predicted):
    matrix[label_map[a], label_map[p]] += 1

return matrix
```

Part 1: KNN Classifier

```
In [17]: # Define euclidean distance
         def euclidean dist(a, b):
             return np.sqrt(np.sum((a - b) **2))
In [19]: def knn classifier(X train, y train):
              # Perform 6-fold cross-validation
             kf = KFold(n splits=6)
             conf matrices = []
              for train index, test index in kf.split(X train):
                  # Split the dataset
                  X train splitted, X test splitted = X train.iloc[train index], X train.iloc[test
                  y train splitted, y test splitted = y train.iloc[train index], y train.iloc[test
                  # Convert DataFrame entries into numerical arrays
                  X train splitted = X train splitted.to numpy()
                 X test splitted = X test splitted.to numpy()
                 y test pred = []
                  # For each test value, run the model
                 for x test in X test splitted:
                      distances = [euclidean dist(x test, x train) for x train in X train splitted
                      indices = np.argsort(distances)[:3]
                      classes = [y_train_splitted.iloc[i] for i in indices]
                      y test pred.append(Counter(classes).most common(1)[0][0])
                  # Compute confusion matrix
                  conf matrix = confusion matrix_custom(y_test_splitted, y_test_pred)
                  conf matrices.append(conf matrix)
              # Calculate mean confusion matrix
             mean conf matrix = np.mean(conf matrices, axis=0)
              normalized conf matrix = conf matrix.astype('float') / conf matrix.sum(axis=1)[:, np
             print("Normalized Confusion Matrix:")
             print(normalized conf matrix)
          start time = time.time()
          knn classifier(X audit, y audit)
          end time = time.time()
         print("Runtime Performance:", end time - start time, "seconds")
         Normalized Confusion Matrix:
         [[0.99186992 0.00813008]
          [0.33333333 0.66666667]]
         Runtime Performance: 1.3699157238006592 seconds
         Confusion matrix says: True Negatives = 0.992, False Negatives = 0.333, True Positives = 0.667, False
         Positives = 0.008
```

Accuracy: 0.83, Precision: 0.99, Recall: 0.67

This model struggles with the false negatives which means that there are a lot of positive inputs classified as negative. The model tends to classify the input as negative.

Runtime performance is 1.37 seconds which is considerably good but it would be faster.

Part 2: KNN Regressor

```
In [20]: # Define manhattan distance
         def manhattan dist(a, b):
             return np.sum(np.abs(a - b))
In [21]: def knn regressor(X train, y train):
             # Perform 6-fold cross-validation
             kf = KFold(n splits=6)
             mae values = []
             for train index, test index in kf.split(X train):
                 # Split the dataset
                 X train splitted, X test splitted = X train.iloc[train index], X train.iloc[test
                 y train splitted, y test splitted = y train.iloc[train index], y train.iloc[test
                 # Convert DataFrame entries into numerical arrays
                 X train splitted = X train splitted.to numpy()
                 X test splitted = X test splitted.to numpy()
                 y test pred = []
                 # For each test value, run the model
                 for x_test in X_test splitted:
                     distances = [manhattan dist(x test, x train) for x train in X train splitted
                     indices = np.argsort(distances)[:3]
                     values = [y train splitted.iloc[i] for i in indices]
                     y test pred.append(np.mean(values))
                 # Compute mean absolute error
                 mae = mean absolute error(y test splitted, y test pred)
                 mae values.append(mae)
             # Calculate mean MAE
             mean mae = np.mean(mae values)
             # Print mean MAE
             print("Mean Absolute Error:", mean mae)
             print("The average of the original results:", y train.to numpy().mean())
             print("The max value and the min value of the original results:", y train.to numpy()
         start time = time.time()
         knn regressor(X day, y day)
         end time = time.time()
         print("Runtime Performance:", end time - start time, "seconds")
         Mean Absolute Error: 805.657164039802
         The average of the original results: 4504.3488372093025
         The max value and the min value of the original results: 8714 22
```

When the statistics (min-max and mean values) are considered, mean absolute error is acceptable.

Runtime Performance: 1.0114789009094238 seconds

Runtime performance is better than the classifier model so it is good enough for a ML model but again, it would be faster. These KNN models might have a problem with large datasets because of distance calculation.

Part 3: Linear SVM Classifier

```
In [22]: def linear svm classifier(X train, y train):
             svm classifier = SVC(kernel='linear', probability=True)
             # Perform 6-fold cross-validation
             kf = KFold(n splits=6)
             rocs = []
             aucs = []
             conf matrices = []
             for train index, test index in kf.split(X train):
                  # Split the dataset
                 X train splitted, X test splitted = X train.iloc[train index], X train.iloc[test
                 y train splitted, y test splitted = y train.iloc[train index], y train.iloc[test
                  # Train the model
                 svm classifier.fit(X train splitted, y train splitted)
                  # Get the ROC curve
                 y probs = svm classifier.predict proba(X test splitted)[:, 1]
                 fpr, tpr, threshold = roc curve(y test splitted, y probs)
                 ROC = [{'fpr': f, 'tpr': t, 'threshold': th} for f, t, th in zip(fpr, tpr, thres
                 rocs.append(ROC)
                 # Compute AUC
                 roc auc = auc(fpr, tpr)
                 aucs.append(roc auc)
                  # Compute confusion matrix
                 y pred = svm classifier.predict(X test splitted)
                 conf matrices.append(confusion matrix custom(y test splitted, y pred))
             plt.figure(figsize=(10, 6))
             for i in range(len(rocs)):
                 fpr values = [entry['fpr'] for entry in rocs[i]]
                 tpr values = [entry['tpr'] for entry in rocs[i]]
                 plt.plot(fpr_values, tpr_values, lw=1, alpha=0.7,
                           label='ROC fold %d (AUC = %0.2f)' % (i, aucs[i]))
              # Calculate mean confusion matrix
             mean conf matrix = np.mean(conf matrices, axis=0)
             normalized conf matrix = mean conf matrix.astype('float') / mean conf matrix.sum(axi
             print("Normalized Mean Confusion Matrix:")
             print(normalized conf matrix)
             # Find the best threshold
             best threshold = None
             max diff = -1
             for roc list in rocs:
                 for roc in roc list:
                     fpr value = roc['fpr']
                     tpr value = roc['tpr']
                     threshold = roc['threshold']
                     diff = tpr value - fpr value # Calculate the difference between tpr and fpr
                      if diff > max diff: # Maximize this difference
                         max diff = diff
                         best threshold = threshold
```

```
print("Best threshold:", best_threshold)

start_time = time.time()
linear_svm_classifier(X_audit, y_audit)
end_time = time.time()
print("Runtime Performance:", end_time - start_time, "seconds")
```

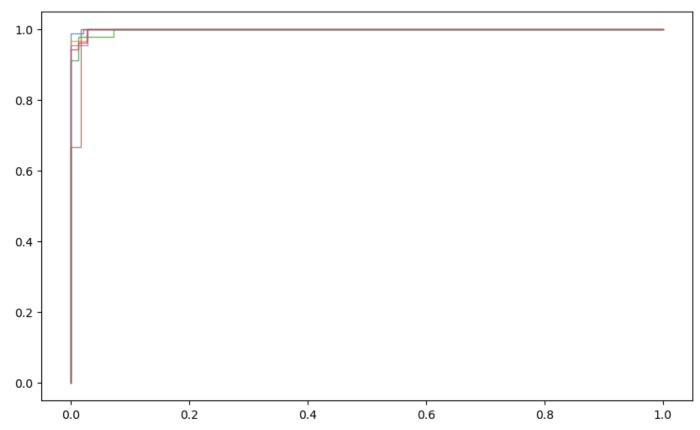
```
Normalized Mean Confusion Matrix:

[[0.99150743 0.00849257]

[0.02622951 0.97377049]]

Best threshold: 0.4058095251132911

Runtime Performance: 0.14642977714538574 seconds
```



Confusion matrix says: True Negatives = 0.992, False Negatives = 0.026, True Positives = 0.974, False Positives = 0.008

Accuracy: 0.98, Precision: 0.99, Recall: 0.97

SVM performs better than KNN models since it also handles the false negatives very well.

Runtime performance is also better than KNN's performance.

For the ROC curve graph, it shows a good result for the model since the curve is very close the northwest point of the graph.

The best threshold is found by calculating the difference between tpr and fpr. The point that maximizes this difference gives the best threshold because this point is the most optimal point at northwest. According to this, the best threshold is 0.41.

Part 4: Linear SVM Regressor

```
In [23]: def linear svm regressor(X train, y train):
             # Initialize Linear SVM Regressor
             svm regressor = SVR(kernel='linear')
             # Perform 6-fold cross-validation
             kf = KFold(n splits=6)
             mae values = []
             for train_index, test_index in kf.split(X train):
                 # Split the dataset
                 X train splitted, X test splitted = X train.iloc[train index], X train.iloc[test
                 y train splitted, y test splitted = y train.iloc[train index], y train.iloc[test
                 # Train the model
                 svm regressor.fit(X train splitted, y train splitted)
                 # Predict on the test set
                 y pred = svm regressor.predict(X test splitted)
                 # Calculate mean absolute error
                 mae = mean absolute error(y test splitted, y pred)
                 mae values.append(mae)
             # Calculate mean MAE
             mean mae = np.mean(mae values)
             # Print mean MAE
             print("Mean Absolute Error:", mean mae)
         start time = time.time()
         linear svm regressor(X day, y day)
         end time = time.time()
         print("Runtime Performance:", end time - start time, "seconds")
```

Mean Absolute Error: 768.4509629776582
Runtime Performance: 0.13226318359375 seconds

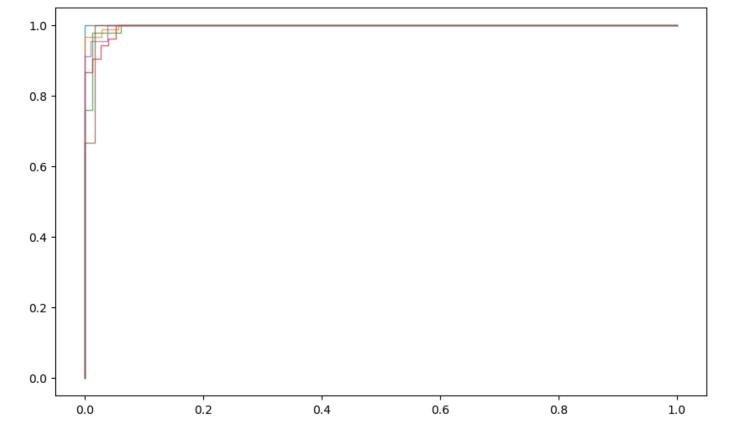
Mean absolute error is lower than the KNN regressor. This means that SVM regressor is better than KNN for this dataset.

Runtime performance is also significantly better than KNN implementations. Therefore, SVM performs very well for this case.

Part 5: Radial Basis Function SVM Classifier

```
svm classifier.fit(X train splitted, y train splitted)
        # Get the ROC curve
        y probs = svm classifier.predict proba(X test splitted)[:, 1]
        fpr, tpr, threshold = roc curve(y test splitted, y probs)
        ROC = [{'fpr': f, 'tpr': t, 'threshold': th} for f, t, th in zip(fpr, tpr, thres
        rocs.append(ROC)
        # Compute AUC
        roc auc = auc(fpr, tpr)
        aucs.append(roc auc)
        # Compute confusion matrix
        y pred = svm classifier.predict(X test splitted)
        conf matrices.append(confusion matrix custom(y test splitted, y pred))
    plt.figure(figsize=(10, 6))
    for i in range(len(rocs)):
        fpr values = [entry['fpr'] for entry in rocs[i]]
        tpr values = [entry['tpr'] for entry in rocs[i]]
        plt.plot(fpr values, tpr values, lw=1, alpha=0.7,
                 label='ROC fold %d (AUC = %0.2f)' % (i, aucs[i]))
    # Calculate mean confusion matrix
    mean conf matrix = np.mean(conf matrices, axis=0)
    normalized conf matrix = mean conf matrix.astype('float') / mean conf matrix.sum(axi
    print("Normalized Mean Confusion Matrix:")
    print(normalized conf matrix)
    # Find the best threshold
    best threshold = None
    \max diff = -1
    for roc list in rocs:
        for roc in roc list:
           fpr value = roc['fpr']
            tpr value = roc['tpr']
            threshold = roc['threshold']
            diff = tpr value - fpr value # Calculate the difference between tpr and fpr
            if diff > max diff: # Maximize this difference
                max diff = diff
                best threshold = threshold
    print("Best threshold:", best threshold)
start time = time.time()
rbf svm classifier(X audit, y audit)
end time = time.time()
print("Runtime Performance:", end time - start time, "seconds")
Normalized Mean Confusion Matrix:
```

```
[[0.98938429 0.01061571]
[0.05245902 0.94754098]]
Best threshold: 0.10264766787637425
Runtime Performance: 0.11210322380065918 seconds
```



Confusion matrix says: True Negatives = 0.989, False Negatives = 0.052, True Positives = 0.948, False Positives = 0.011

Accuracy: 0.97, Precision: 0.99, Recall: 0.95

SVM with radial basis function has considerably good performance. Its metric (accuracy and recall) values are slightly lower than linear SVM. However, it is still much better than KNN.

Runtime performance is similar to linear SVM. Overall, it has a great performance.

For the ROC curve graph, it shows a great result as linear SVM does.

Using the same technique, the best threshold for this model is 0.10.

Part 6: Decision Tree Classifier

```
In [25]: def dt_classifier(X_train, y_train):
    # Build the decision trees with two pruning options
    dt_classifier_pre_pruned = DecisionTreeClassifier(criterion='entropy', max_depth=2)
    dt_classifier_post_pruned = DecisionTreeClassifier(criterion='entropy', ccp_alpha=0.

# Perform 6-fold cross-validation
    kf = KFold(n_splits=6)
    for train_index, test_index in kf.split(X_train):
        # Split the dataset
        X_train_splitted, X_test_splitted = X_train.iloc[train_index], X_train.iloc[test
        y_train_splitted, y_test_splitted = y_train.iloc[train_index], y_train.iloc[test
        # Train pre-pruned tree
        dt_classifier_pre_pruned.fit(X_train_splitted, y_train_splitted)

# Train post-pruned tree
```

```
dt_classifier_post_pruned.fit(X_train_splitted, y_train_splitted)

# Evaluate the tree

pre_pruning_test_score = dt_classifier_pre_pruned.score(X_test_splitted, y_test_
post_pruning_test_score = dt_classifier_post_pruned.score(X_test_splitted, y_test_
print("Pre-Pruning Tree Accuracy:", pre_pruning_test_score)
print("Post-Pruning Tree Accuracy:", post_pruning_test_score)

return dt_classifier_pre_pruned, dt_classifier_post_pruned

dt_classifier_pre_pruned, dt_classifier_post_pruned = dt_classifier(X_audit, y_audit)
```

```
Pre-Pruning Tree Accuracy: 1.0
Post-Pruning Tree Accuracy: 1.0
Pre-Pruning Tree Accuracy: 0.9923076923076923
Post-Pruning Tree Accuracy: 1.0
Pre-Pruning Tree Accuracy: 0.9844961240310077
Post-Pruning Tree Accuracy: 0.9922480620155039
Pre-Pruning Tree Accuracy: 0.9844961240310077
Post-Pruning Tree Accuracy: 1.0
Pre-Pruning Tree Accuracy: 1.0
Pre-Pruning Tree Accuracy: 1.0
Post-Pruning Tree Accuracy: 1.0
Post-Pruning Tree Accuracy: 1.0
Post-Pruning Tree Accuracy: 1.0
```

Pre-pruning is set to the depth 2. Post-pruning is set to the ccp_alpha 0.01 (Minimal cost-complexity pruning is an algorithm used to prune a tree to avoid over-fitting).

Pre-pruning results depend on the chosen depth. For this choice, it is not as accurate as post-pruning tree. However, the results from the both trees are almost 1 so this says there might be overfitting.

When we look at the extracted rules below, we can see that post-pruning tree has more branches and pre-pruning tree is limited to the depth 2.

```
In [26]:

def extract_rules(tree, feature_names, node, is_classification, indent=""):
    if tree.feature[node] != -2:
        feature = feature_names[tree.feature[node]]
        threshold = tree.threshold[node]
        print(f"{indent}|--- {feature} <= {threshold:.2f}")
        extract_rules(tree, feature_names, tree.children_left[node], is_classification,
        print(f"{indent}|--- {feature} > {threshold:.2f}")
        extract_rules(tree, feature_names, tree.children_right[node], is_classification,
    else:
        if is_classification:
            value = int(tree.value[node].argmax())  # Get the class with the highest cou
            print(f"{indent}|--- class: {value}")
        else:
            value = tree.value[node][0][0]  # Predicted value for regression
            print(f"{indent}|--- value: {value:.2f}")
```

```
In [27]: print("Pre-pruning Tree Rules:")
    extract_rules(dt_classifier_pre_pruned.tree_, columns_to_normalize_audit, 0, True)
```

```
Pre-pruning Tree Rules:
       \mid--- Inherent Risk <= -0.23
         |--- CONTROL RISK <= 0.96
        | | |--- class: 0
        \mid \mid --- CONTROL RISK > 0.96
       | | |--- class: 1
        \mid --- \mid \text{Inherent Risk} > -0.23
       | |--- class: 1
In [28]: print("Post-pruning Tree Rules:")
       extract_rules(dt_classifier_post_pruned.tree_, columns_to_normalize_audit, 0, True)
       Post-pruning Tree Rules:
       \mid--- Inherent Risk <= -0.23
        | |--- RiSk E <= 1.66
        | | |--- District Loss <= 0.40
        | | | |--- class: 0
        | | |--- TOTAL <= -0.22
        | --- RiSk E > 1.66
       | | |--- class: 1
       \mid ---  Inherent Risk > -0.23
        | |--- class: 1
```

Part 7: Decision Tree Regressor

```
In [29]: def dt regressor(X train, y train):
             # Build the decision trees with two pruning options
             dt regressor pre pruned = DecisionTreeRegressor(max depth=2)
             dt regressor post pruned = DecisionTreeRegressor(ccp alpha=0.01)
             # Perform 6-fold cross-validation
             kf = KFold(n splits=6)
             for train index, test index in kf.split(X train):
                 # Split the dataset
                 X train splitted, X test splitted = X train.iloc[train index], X train.iloc[test
                 y train splitted, y test splitted = y train.iloc[train index], y train.iloc[test
                 # Train pre-pruned tree
                 dt regressor pre pruned.fit(X train splitted, y train splitted)
                 # Train post-pruned tree
                 dt regressor post pruned.fit(X train splitted, y train splitted)
                 # Evaluate the tree
                 pre pruning test score = dt regressor pre pruned.score(X test splitted, y test s
                 post pruning test score = dt regressor post pruned.score(X test splitted, y test
                 print("Pre-Pruning Tree Accuracy:", pre pruning test score)
                 print("Post-Pruning Tree Accuracy:", post pruning test score)
             return dt regressor pre pruned, dt regressor post pruned
         dt regressor pre pruned, dt regressor post pruned = dt regressor(X day, y day)
```

Pre-Pruning Tree Accuracy: 0.49714385943456774
Post-Pruning Tree Accuracy: 0.9328960424997911
Pre-Pruning Tree Accuracy: -1.083655952604818
Post-Pruning Tree Accuracy: 0.7623804402548249

```
Pre-Pruning Tree Accuracy: 0.5643271599067563
Post-Pruning Tree Accuracy: 0.9753097555904336
Pre-Pruning Tree Accuracy: 0.6245547167521159
Post-Pruning Tree Accuracy: 0.9731921597117443
Pre-Pruning Tree Accuracy: -0.2717478889370015
Post-Pruning Tree Accuracy: 0.8638210190396074
Pre-Pruning Tree Accuracy: 0.6327545040190842
Post-Pruning Tree Accuracy: 0.973264114906303
```

Pre-pruning is set the depth 2. Post-pruning is set to the ccp_alpha 0.01 (Minimal cost-complexity pruning is an algorithm used to prune a tree to avoid over-fitting).

Pre-pruning tree works poorly since we can see that the accuracy is too low and even negative for some cases. (The best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse)) Post-pruning works better. For some cases it is as low as 0.76 but it can be as high as 0.97.

However, when we look at the extracted rules of the post-pruning tree below, we can see that the tree is too complex. Having a complex tree is not something we want.

```
In [30]: print("Pre-pruning Tree Rules:")
         extract rules(dt regressor pre pruned.tree , columns to normalize day, 0, False)
         Pre-pruning Tree Rules:
         \mid--- registered <= -0.48
         | |--- registered <= -0.99
           | |--- value: 1695.79
            \mid --- \text{ registered} > -0.99
         | | |--- value: 3279.07
         |--- registered > -0.48
           |---| instant <= 0.34
         | | |--- value: 4415.41
           |--- instant > 0.34
            | |--- value: 6486.15
In [31]: print("Post-pruning Tree Rules:")
         extract rules(dt regressor post pruned.tree , columns to normalize day, 0, False)
         Post-pruning Tree Rules:
         |--- registered <= -0.48
         | |--- registered <= -0.99
            | |--- registered <= -1.45
                     \mid--- registered <= -1.75
                        |--- registered <= -1.83
                           |--- registered <= -1.95
                               |---| mnth <= -1.31
                              | |--- instant <= -1.61
                              | | |--- value: 506.00
                                    |---| instant > -1.61
                                | |--- value: 431.00
                               |---| mnth > -1.31
                                    |--- weathersit <= 0.19
                                | |--- value: 754.00
                                    |--- weathersit > 0.19
                                        |---| hum <= 2.02
                                        |---| temp <= -0.95
                                               |--- value: 627.00
                                            |---| temp > -0.95
                                            |--- value: 623.00
                                        |--- hum > 2.02
                                        | |--- value: 605.00
                            |--- registered > -1.95
```

```
|--- registered <= -1.93
        | |--- value: 985.00
         \mid ---  registered > -1.93
       | ---  casual <= -1.16
            | |--- windspeed <= 0.21
            | | |--- value: 683.00
               |--- windspeed > 0.21
               | |--- value: 705.00
            |--- casual > -1.16
       | | |--- registered <= -1.88
          | | |--- instant <= -1.48
            | | | |--- value: 801.00
                  \mid --- \text{ instant} > -1.48
               | | |--- value: 795.00
         |--- registered > -1.88
              | |--- value: 822.00
         |--- registered > -1.83
      |--- casual <= -0.76
        |--- casual <= -0.98
            |---| temp <= -1.78
      | | | | | |--- value: 981.00
     | | | | |--- instant > -1.63
              | | |--- value: 986.00
               \mid --- registered > -1.77
         | |--- value: 959.00
            |---| temp > -1.78
              |---| temp <= -0.80
        | | | | | | | --- temp <= -1.59
               | | | |--- value: 1000.00
        |---| temp > -1.59
              | | | |--- value: 1005.00
       | |--- value: 1011.00
      |---| temp > -0.80
            | | |--- value: 1027.00
         |--- casual > -0.98
            |---| atemp <= -0.09
    | | | |--- value: 1107.00
            |---| atemp > -0.09
         | | | |--- value: 1115.00
      |--- casual > -0.76
     | |--- value: 1317.00
\mid ---  registered > -1.75
   |--- casual <= -0.77
      |--- registered <= -1.60
      | --- \text{ hum } <= -0.55
   | | | |--- value: 1204.00
            |---| atemp > -1.43
        | | |--- value: 1248.00
     | --- \text{hum} > -0.55
           |--- registered <= -1.69
     | | | |--- value: 1097.00
      | | |--- registered > -1.69
       | |--- value: 1162.00
            |--- hum > 1.05
              | |--- value: 1167.00
            \mid ---  registered > -1.60
      \mid --- \text{ temp} \le -1.63
      | | |--- registered <= -1.50
```

```
|---| temp <= -1.71
                   | |--- registered <= -1.56
                            |--- value: 1263.00
                       |--- registered > -1.56
                   | | |--- casual <= -1.16
                         | | |--- value: 1321.00
                            |--- casual > -1.16
                     | | |--- value: 1301.00
                     |---| temp > -1.71
                     \mid --- \text{ temp} \le -1.65
                     | | |--- value: 1360.00
                         |---| temp > -1.65
                       | |--- value: 1349.00
                     \mid --- \text{ registered} > -1.50
                     |---| temp <= -1.76
                       |---| temp <= -1.82
             | --- atemp <= -1.97
          | | | |--- value: 1416.00
                     | | |--- value: 1421.00
                         |---| temp > -1.82
                     | | |--- value: 1406.00
                     |---| temp > -1.76
                     | |--- value: 1450.00
              |---| temp > -1.63
                  \mid --- \text{ windspeed} \mid --- 0.03
                    |---| hum <= -0.39
                  | | |--- value: 1510.00
                   |--- \text{ hum } > -0.39
                   | |--- value: 1536.00
                |--- windspeed > -0.03
                 | --- temp <= -0.60
                     | |--- windspeed <= 1.24
                     | | |--- value: 1471.50
                     | |--- windspeed > 1.24
                        | |--- value: 1461.00
                     |--- \text{ temp} > -0.60
                   | |--- value: 1446.00
       |--- casual > -0.77
          |--- registered <= -1.50
            |---| temp <= -1.15
            | --- temp <= -1.24
                | |--- value: 1693.00
                |---| temp > -1.24
             | | |--- value: 1812.00
             |---| temp > -1.15
         | | |--- holiday <= 2.82
         | | | | | --- casual <= -0.56
          | | | | |--- casual <= -0.69
         | | | | | |--- value: 1623.00
             | | | | ---  casual > -0.69
                     | | |--- value: 1589.00
                     |--- casual > -0.56
                  | | |--- value: 1635.00
                  |--- holiday > 2.82
                 | |--- value: 1495.00
             |--- registered > -1.50
         | |--- value: 2252.00
\mid--- registered > -1.45
   |--- casual <= -0.58
       \mid--- registered <= -1.27
       | ---  casual <= -1.07
       | | |--- registered <= -1.38
```

```
|--- value: 1501.00
                               |--- casual > -1.17
                                         |--- casual <= -1.10
                                                    |---| hum <= -1.17
                                                    \mid --- \text{ temp} \le -1.71
                                                    | --- \text{ hum } <= -1.27
                                                                     | |--- value: 1538.00
                                                                          |---| hum > -1.27
                                                               | | |--- value: 1543.00
                                                               |---| temp > -1.71
                                                           | |--- value: 1550.00
                                                    |--- hum > -1.17
                                                               |--- registered <= -1.41
                                        | |--- value: 1526.00
                                        \mid ---  registered > -1.41
                                                | | |--- value: 1529.50
                                          |--- casual > -1.10
                                         | |--- value: 1562.00
                   \mid --- registered > -1.38
                               |--- weekday <= 0.75
                                         |--- windspeed <= 0.10
                                        | --- \text{ hum } <= -1.14
                                        | | |--- value: 1600.00
                                   | --- \text{ hum } > -1.14
                                    | | |--- value: 1606.00
                                        |--- windspeed > 0.10
                                        | |--- value: 1650.00
                               |--- weekday > 0.75
                             | |--- value: 1708.00
                   |--- casual > -1.07
                   |--- casual <= -0.93
                         |--- registered <= -1.31
                                        |---| temp <= -1.65
                                   | |--- value: 1746.00
             | | | --- temp > -1.65
                                    | --- mnth <= -1.17
                                   | | |--- value: 1712.00
                                        |---mnth>-1.17
                                         | | |--- value: 1684.33
                   | |--- registered > -1.31
                                   |--- registered <= -1.29
                                    | |--- value: 1795.00
                 \mid --- \text{ registered} > -1.29
                             | | |--- value: 1816.00
                   |--- casual > -0.93
                               \mid ---  instant <= -1.44
                             | |--- value: 1969.00
                              \mid --- \text{ instant} > -1.44
                             | |--- value: 1872.00
                  --- registered > -1.27
        |--- registered <= -1.15
                   |--- casual <= -0.92
                              |---| atemp <= -1.32
                        | --- \text{hum} <= -0.10
                                               |--- casual <= -0.98
                                        | | |--- value: 1917.00
                                                            |---| atemp > -1.38
                                                    | | |--- value: 1927.00
                                         |--- casual > -0.98
                                                             |--- value: 1944.00
                                               |--- hum > -0.10
                                                    \mid --- \text{ workingday} \mid ---- \text{ workingday} \mid --- \text{ workingday} \mid --- \text{ workingday} \mid --- 
                                                    | |--- value: 1977.00
```

```
|--- workingday > -0.40
         | | |--- value: 1985.00
      |---| atemp > -1.32
       |--- casual <= -0.98
   | | |--- value: 1807.00
            \mid --- \text{ instant} > -1.46
         | | |--- windspeed <= 0.18
         | | | --- temp <= -0.03
           | | | |--- value: 1834.00
     | | | | |--- value: 1842.00
         |--- windspeed > 0.18
           | | | |--- value: 1851.00
             | --- \text{ hum} > 0.30
           | | | |--- value: 1865.00
         |--- casual > -0.98
         | --- mnth <= -1.17
           | |--- value: 1913.00
         |---mnth>-1.17
         | | |--- value: 1891.00
  |--- casual > -0.92
      |--- casual <= -0.87
      | --- season <= -0.90
      | | |--- weekday <= 0.00
         | | |--- value: 1951.00
         | --- weekday > 0.00
       | | |--- value: 1977.00
         |--- season > -0.90
       | |--- value: 2028.00
      |--- casual > -0.87
         |---| atemp <= -1.01
         | |--- value: 2133.00
    | --- atemp > -1.01
    | | |--- casual <= -0.68
  | | | | |--- hum <= 0.49
    | | | | | |--- value: 2046.00
       | | | |--- value: 2056.00
              |--- \text{ hum } > 0.49
           | | |--- value: 2034.00
           |--- casual > -0.68
         | | |--- value: 2077.00
\mid ---  registered > -1.15
   \mid --- weekday <= -0.25
 | --- \text{ hum } <= -1.84
    | |--- value: 2425.00
| | |--- registered <= -1.03
       | |--- value: 2311.00
         \mid --- \text{ registered} > -1.03
     | | |--- value: 2298.00
  |--- weekday > -0.25
   |--- season <= 0.00
     | --- casual <= -0.88
         | |--- instant <= -1.30
         | | |--- windspeed <= 1.15
            | | | |--- value: 2121.00
       | --- casual > -0.93
              | | |--- value: 2115.00
              |--- windspeed > 1.15
```

```
| | |--- value: 2134.00
                |---| instant > -1.30
            | | | | | | | |--- temp <= -0.97
            | | | | | |--- value: 2169.00
                        |---| temp > -0.97
             | | | |--- value: 2162.00
                | | |--- value: 2177.00
             |---| mnth > 0.43
                | | | | |--- value: 2209.00
             |--- casual > -0.88
             | |--- registered <= -1.07
      | | | | | | | | | |--- value: 2210.00
      | | | | | | |--- temp > -1.17
       | | | | | | | |--- value: 2227.00
            |---| temp > -0.89
             | | | |--- value: 2192.00
             \mid --- \text{ registered} > -1.07
            | | |--- value: 2302.00
          |--- season > 0.00
       | | |--- value: 1996.00
|--- casual > -0.58
  |--- casual <= 0.32
    |---| instant <= -1.40
      \mid--- registered <= -1.34
  | | | |--- value: 2132.00
    | | |--- value: 2077.00
      \mid --- \text{ registered} > -1.34
     | | |--- value: 2402.00
     |---| instant > -1.40
      |--- registered <= -1.06
      | --- mnth <= -1.31
         | |--- value: 2294.00
        | | | | | |--- value: 2471.00
         | | |--- value: 2455.00
        | | | |--- windspeed > 0.23
      | | | | |--- season <= -0.90
                 | |--- value: 2485.00
               | | | | | | | --- season > -0.90
      | | | | | |--- value: 2496.00
             |--- windspeed > 0.71
             | --- atemp <= -0.48
             | | |--- value: 2417.00
             |---| atemp > -0.48
             | | |--- value: 2429.00
        \mid --- \text{ registered} > -1.06
        | |--- value: 2689.00
  |--- casual > 0.32
     |--- casual <= 0.54
      |---| weekday <= -0.25
      | |--- value: 2895.00
    | --- weekday > -0.25
    | | |--- value: 2792.00
```

```
\mid ---  casual > 0.54
         |--- registered <= -1.29
            |--- value: 3249.00
         \mid ---  registered > -1.29
       | | |--- windspeed <= 0.36
          | | |--- value: 3068.00
           |--- windspeed > 0.36
       | | | |--- value: 3117.00
- registered > -0.99
|--- casual <= 0.44
   |--- casual <= -0.51
     |--- registered <= -0.87
        |--- casual <= -0.66
           |--- casual <= -1.10
        | | |--- value: 2236.00
          |--- casual > -1.10
       | | |--- value: 2368.00
       | | | | | | |--- value: 2376.00
               |--- casual > -1.05
        | | | | |--- value: 2395.00
                |--- instant > -0.43
        | \  | \  | \  | \  | --- weekday <= -0.50
          | | | | | | |--- value: 2431.50
          |--- weekday > -0.50
       | | | | | | | |--- value: 2423.50
                      | |--- windspeed > 0.71
                     | | |--- value: 2416.00
                | | | | --- casual <= -0.80
               | |--- value: 2475.00
        |--- casual > -0.80
        | | | | |--- value: 2493.00
        |--- casual > -0.66
       | |--- value: 2703.00
     |--- registered > -0.87
        |--- registered <= -0.69
        \mid ---  casual <= -0.76
        | | |--- season <= 0.00
                | |--- value: 2660.00
        | | |--- value: 2566.00
          | --- temp > -0.24
               | | |--- value: 2594.00
              |--- casual > -1.00
                |--- instant <= -0.25
        |---| atemp <= -0.27
                | | | |--- value: 2633.00
          | | | |--- value: 2659.00
                   |--- hum > 1.56
        | | |--- value: 2710.00
                |--- instant > -0.25
                | ---  casual <= -0.88
                | | |--- value: 2765.00
                 |--- casual > -0.88
```

```
\mid ---  instant \leq 0.09
                      | |--- value: 2739.00
                      |---| instant > 0.09
                    | |--- value: 2732.00
          |--- casual > -0.76
             |--- casual <= -0.57
                 |--- registered <= -0.82
                 | |--- value: 2743.50
                 \mid ---  registered > -0.82
               | |--- casual <= -0.66
         | | | | |--- season <= -0.90
         | | | | | |--- value: 2832.00
            | | | | --- season > -0.90
                   | | |--- value: 2843.00
                 | ---  casual > -0.66
               | | |--- value: 2808.00
             |---  casual > -0.57
             | |--- value: 2999.00
      |--- registered > -0.69
          |--- hum <= 0.34
             \mid --- \mid registered <= -0.51
          | | |--- value: 3204.00
             \mid ---  registered > -0.51
         | | |--- value: 3053.00
         |--- hum > 0.34
             \mid ---  registered <= -0.53
               |---| temp <= -1.28
             | | |--- value: 2802.00
                |---| temp > -1.28
        | | | | | | |--- value: 2947.00
               |---| atemp > -0.98
             | | | | |--- value: 2934.00
            | | |--- value: 2913.50
            |--- registered > -0.53
         | | |--- value: 3068.00
|--- casual > -0.51
   |--- registered <= -0.62
      |--- casual <= -0.02
     | |--- registered <= -0.83
    | | |--- registered <= -0.88
     | | | |--- value: 3071.00
           |--- registered > -0.88
            | --- instant <= -0.97
          | | | |--- value: 2927.00
             | --- instant > -0.97
               | |--- value: 2918.00
             \mid --- \text{ registered} > -0.83
            |--- registered <= -0.64
               |--- windspeed <= 0.05
                | |--- value: 3141.00
        | | |--- windspeed > 0.05
        | | | | | --- workingday <= -0.40
      | | | | | | |--- value: 3127.50
               | | | --- workingday > -0.40
               | | |--- value: 3115.00
             |---| registered > -0.64
               |---| temp <= -1.18
                | |--- value: 3190.00
                |---| temp > -1.18
                 | |--- value: 3243.00
      |--- casual > -0.02
```

```
\mid ---  registered <= -0.72
            |--- hum <= -0.38
               |--- season <= -0.45
              | |--- value: 3239.00
              |--- season > -0.45
              | |--- value: 3285.00
             |--- hum > -0.38
           | |--- registered <= -0.74
         | | | | | |--- value: 3331.00
         | | |--- value: 3349.50
                |--- \text{ hum } > 1.38
                | | |--- value: 3372.00
                \mid --- registered > -0.74
             | | |--- value: 3409.00
          \mid --- \text{ registered} > -0.72
        | |--- value: 3606.00
    \mid --- \text{ registered} > -0.62
       |---| mnth <= 0.72
       | |--- instant <= -1.23
        | |--- value: 3267.00
       | --- instant > -1.23
      | | |--- value: 3388.00
     | | | | --- season <= -0.90
        | | | |--- value: 3424.00
        | | | --- season > -0.90
         | | | |--- value: 3429.00
       |---| mnth > 0.72
          |--- casual <= 0.02
     | | | |--- value: 3614.00
       | | | |--- value: 3520.00
               |--- hum > 0.72
         | | | |--- value: 3485.00
        |--- casual > 0.02
            |--- casual <= 0.19
        | | | |--- casual <= 0.14
         | | | |--- value: 3663.00
              |--- casual > 0.14
         | | |--- value: 3649.00
          | --- casual > 0.19
       | | | |--- value: 3717.00
-- casual > 0.44
 |--- casual <= 1.42
   |--- registered <= -0.85
      |---| atemp <= 0.38
    | | |--- value: 3744.00
    |--- atemp > 0.38
      | |--- value: 3351.00
    \mid --- registered > -0.85
      |--- casual <= 0.78
       | --- instant <= -0.25
       | | |--- instant <= -0.65
     | | | | | | --- instant <= -0.68
           | | |--- value: 3785.00
         | | | |--- value: 3820.00
         | --- instant > -0.65
```

```
\mid --- \text{ windspeed} \mid = 0.37
                        | |--- value: 3926.00
                        |--- windspeed > 0.37
                       | |--- value: 3873.00
                  |--- instant > -0.25
                 | |--- value: 4067.00
               |--- casual > 0.78
                  |--- registered <= -0.61
                | --- atemp <= 0.16
                    | |--- value: 4036.00
                 | --- atemp > 0.16
                 | | |--- temp <= 1.17
                 | | | | --- weekday <= 0.00
                        | | |--- value: 4191.00
                        |--- weekday > 0.00
                        | | |--- value: 4150.00
                        |---| temp > 1.17
                       | |--- value: 4098.00
                  \mid ---  registered > -0.61
                     |--- casual <= 1.05
               | | | | | |--- value: 4318.00
              | | | | |--- value: 4308.00
                       |---| atemp > 0.47
               | \ | \ | \ | ---  casual <= 0.98
                       | | |--- value: 4294.00
                        | --- casual > 0.98
                        | | |--- value: 4302.00
                   |---| casual > 1.05
              | | | |--- value: 4381.00
         |--- casual > 1.42
            |--- hum <= 0.67
          | |--- weekday <= 0.00
              | |--- value: 4649.00
            \mid --- \text{ weekday} > 0.00
           | | |--- value: 4484.00
            |--- hum > 0.67
        | --- \text{ hum } <= 1.03
            | | | |--- value: 4758.00
            | | | |--- value: 4788.00
              |---| mnth > 0.14
             | |--- value: 4940.00
           |--- registered > -0.48
  |---| instant <= 0.34
      |--- casual <= -0.31
       |--- registered <= -0.02
          |--- registered <= -0.30
               |---| atemp <= -0.53
               | |--- weathersit <= 0.19
            |--- weekday <= 0.75
              | | | | | |--- value: 3292.00
                        |--- casual > -1.04
               | | | | | | | |--- value: 3272.00
                           |--- casual > -1.03
                        |--- casual <= -1.02
                           | | |--- value: 3310.00
                             |--- casual > -1.02
```

```
|--- registered <= -0.32
                 | |--- value: 3333.00
                  \mid --- \text{ registered} > -0.32
         | |--- value: 3322.00
          |--- weekday > 0.75
         | |--- value: 3214.00
        |--- casual > -0.91
          |--- casual <= -0.73
          | --- temp <= -1.20
            | |--- value: 3392.00
   | | | | |--- value: 3403.00
          |--- casual > -0.73
         | |--- value: 3368.00
    |--- weathersit > 0.19
       \mid --- weekday <= 0.75
 | | | |--- value: 3005.00
  \mid \quad \mid \quad \mid --- \text{ weekday} > 0.75
    | | |--- windspeed <= -0.22
         | |--- value: 3194.00
       \mid --- \text{ windspeed} > -0.22
       | | |--- value: 3163.00
  |---| atemp > -0.53
     |--- casual <= -0.51
     | --- mnth <= -0.59
      | |--- value: 3456.00
       |--- mnth > -0.59
 | | | |--- atemp <= 1.16
    | | | |--- value: 3543.00
   | | | |--- value: 3574.00
     |--- casual > -0.51
        |--- season <= 0.90
  | | | |--- value: 3784.00
     | --- season > 0.90
       | |--- value: 3644.00
    \mid --- \text{ registered} > -0.30
  |--- casual <= -0.56
    |--- registered <= -0.13
        \mid --- \text{ instant} \mid = 0.06
   | \  | \  | \  | --- temp <= -1.14
 | | | | | |--- value: 3523.00
    | | | | | | | |--- mnth <= 1.44
    | | | | | | | | |--- value: 3613.00
  | | | | | | | | |--- value: 3620.00
     | | | | | |--- instant > -0.03
                | | |--- value: 3598.00
         |--- registered > -0.22
 | | |--- value: 3577.00
          |--- casual > -0.73
  |--- weekday \leq 0.75
  | \  | \  | \  | \  | ---  casual <= -0.61
    | | | | | |--- value: 3669.00
         | | |--- value: 3659.00
```

```
|--- casual > -0.61
            | | |--- value: 3641.00
            |--- weekday > 0.75
           | |--- value: 3747.00
         |--- instant > 0.06
         |--- hum <= -0.26
         |--- temp <= -1.27
         | | |--- value: 3422.00
         |--- \text{ temp} > -1.27
         | | |--- value: 3376.00
         |--- hum > -0.26
      | | |--- value: 3487.00
   \mid--- registered > -0.13
      |--- casual <= -0.91
       |---| temp <= -1.20
        | |--- value: 3624.00
         |---| temp > -1.20
    | | |--- instant <= -0.07
    | | | | | --- temp <= -0.69
         | | | | | |--- value: 3740.00
    | | |--- value: 3727.00
              |---| temp > -0.69
    | | | | |--- value: 3709.00
         | --- instant > -0.07
         | | |--- registered <= -0.05
    | | | | | |--- value: 3750.00
 | | | | | | | | registered > -0.07
    | | | | | | |--- value: 3761.00
    | | | |--- atemp <= -1.10
            | | | |--- value: 3777.00
    | | | --- atemp > -1.10
              | | |--- value: 3784.00
            |--- casual > -0.91
      | ---  casual <= -0.88
      | | |--- weekday <= 0.00
   | | | | |--- value: 3811.00
         \mid --- \text{ weekday} > 0.00
       | | |--- value: 3831.00
       |--- casual > -0.88
 | | | | | | --- casual <= -0.76
       | | |--- value: 3867.00
         |--- casual > -0.76
      | | | |--- value: 3894.00
|--- casual > -0.56
  |--- registered <= -0.14
 | |--- registered <= -0.21
    | |--- windspeed <= 0.17
      | | |--- weathersit <= 0.19
         | | |--- value: 3840.00
 | | |--- value: 3855.00
         |--- windspeed > 0.17
     | | |--- value: 3767.00
      \mid --- \text{ registered} > -0.21
         |--- registered <= -0.21
         | |--- value: 3872.00
         \mid --- \text{ registered} > -0.21
       | --- mnth <= -0.44
         | | |--- value: 3944.00
         |---mnth>-0.44
```

```
\mid --- \text{ weekday} \mid <= 0.25
                  | |--- value: 3907.00
                  |--- weekday > 0.25
                | |--- value: 3915.00
        \mid --- \text{ registered} > -0.14
           |--- registered <= -0.12
          | |--- value: 3974.00
          \mid --- \text{ registered} > -0.12
      | | | --- atemp <= 0.10
            | |--- value: 4046.00
       | --- atemp > 0.10
       | | | | | --- value: 4058.00
|--- registered > -0.02
  |--- registered <= 0.36
     |--- registered <= 0.12
      |--- casual <= -0.61
    | | |--- registered <= 0.09
     | | | | | |--- value: 3830.00
     | | |--- value: 3922.00
       | | | |--- value: 3974.00
        |--- registered > 0.01
      | | | | | |--- weekday <= 0.25
         | | | | | |--- value: 3956.00
     | | | | | | | |--- weekday > 0.25
            | | | | |--- value: 3940.00
            |--- casual > -0.73
            | --- weekday <= -0.75
     | | | | | |--- value: 4035.00
      | | | |--- weekday > -0.75
            | | |--- value: 4068.00
       | |--- registered > 0.09
          | ---  casual <= -0.79
             | |--- instant <= 0.09
     | | | | | | | |--- value: 4097.50
     | | | | | | | |--- value: 4109.00
             | --- instant > 0.09
             | | |--- value: 4075.00
             |--- casual > -0.79
     | | | | | | | |--- value: 4186.00
     | | | |--- value: 4205.00
                |---| instant > -0.05
             | | |--- value: 4152.50
        |--- casual > -0.61
          |--- registered <= 0.07
          | |--- windspeed <= -0.92
      | | | |--- value: 4120.00
      | | |--- windspeed > -0.92
      | | | | --- casual <= -0.49
         | | | |--- value: 4195.00
             | --- casual > -0.49
            | | |--- value: 4182.00
         |--- registered > 0.07
         | --- atemp <= -0.65
```

```
| |--- value: 4270.00
                    |---| atemp > -0.65
                       |---| atemp <= 0.13
                      | |--- value: 4433.00
                       |---| atemp > 0.13
                    | | |--- weekday <= 0.25
                         | |--- value: 4352.00
                       |--- weekday > 0.25
                      | | |--- value: 4390.00
          |--- registered > 0.12
            |---| atemp <= -0.63
            | --- weekday <= -0.25
                 | --- casual <= -0.88
                | | |--- value: 4363.00
                    | --- hum > -1.28
                    | | |--- value: 4375.00
                  |--- casual > -0.88
                  \mid \quad \mid --- \text{ weekday} \mid <= -0.75
                      | |--- value: 4322.00
                    | --- weekday > -0.75
                | | | |--- value: 4339.00
                 |--- weekday > -0.25
                 | |--- value: 4169.00
             |---| atemp > -0.63
                 |--- weekday <= 0.25
                  |--- instant <= -0.32
                | | | season <= 0.45
                    | | |--- value: 4451.00
                    | --- season > 0.45
                  | | |--- value: 4456.00
                  |--- instant > -0.32
                    | |--- registered <= 0.24
               | | | |--- value: 4486.00
                | | |--- registered > 0.24
                   | | |--- value: 4509.00
                |--- weekday > 0.25
            | | |--- value: 4569.00
      |--- registered > 0.36
          |--- registered <= 0.55
          |--- hum <= -0.82
           | |--- value: 4579.00
           |--- hum > -0.82
           | |--- instant <= 0.28
            | | |--- season <= 0.90
                  | --- season <= -0.45
          | | | | | | |--- value: 4773.00
           | | | | --- season > -0.45
                      | |--- value: 4795.00
                  |--- season > 0.90
               | | |--- value: 4826.00
                 |--- instant > 0.28
          | | | |--- value: 4916.00
          |--- registered > 0.55
            |--- windspeed <= 1.86
          | | | |--- value: 4990.00
          | | |--- value: 5062.00
           |--- windspeed > 1.86
         | | |--- value: 5382.00
|--- casual > -0.31
 |--- casual <= 1.31
   | |--- registered <= 0.20
```

```
|--- casual <= 0.94
  |--- registered <= -0.02
    |--- casual <= -0.05
| | | |--- value: 3805.00
      | | | |--- value: 3846.00
   \mid ---  registered > -0.31
   | | | |--- windspeed <= -0.64
     | | |--- value: 3958.00
      | | | | | | | |--- value: 3982.00
         \mid --- \text{ windspeed} > -0.64
     | | | | | | | |--- value: 4073.00
   | | | | |--- instant > -1.06
         | | | | | | | | | | | | | | | --- value: 4010.00
        | --- atemp > 0.97
        | | | | |--- value: 4040.00
      |--- registered > -0.20
 | | | | | | | | |--- value: 4188.00
 | | | |--- mnth <= 0.28
       | | | | | |--- value: 4086.00
   | | | | | |--- registered > -0.17
 | | | | | | | | | | |--- value: 4105.00
  | | | | |--- value: 4123.00
        |--- mnth > 0.28
 | | | | | | |--- value: 4153.00
 | | | | | | |--- mnth <= 0.72
  | \ | \ | \ | \ | \ | --- temp <= 1.15
     | | | | | |--- value: 4338.00
  | \  | \  | \  | \  | \  | \  | --- \  temp > 1.15
 | | | | | | | | |--- value: 4342.00
| | | | | | | |--- mnth > 0.72
| | | | | | | | | | | --- value: 4304.00
 | | | |--- weekday <= -0.75
         | | |--- value: 4266.00
  | | | | | |--- weekday > -0.75
     | | | | |--- value: 4258.00
    |--- casual > -0.05
    | |--- instant <= 0.16
      |--- temp <= 0.02
 | | |--- weathersit <= 0.19
| | | |--- value: 4521.00
        | |--- weathersit > 0.19
   | | |--- value: 4511.00
 | --- temp > 0.02
     | | |--- registered <= -0.09
```

```
|--- casual <= 0.86
         |--- hum > 0.25
     | | | | | |--- value: 4274.00
 | | | | | |--- casual > 0.86
 | | | | | | | |--- value: 4460.00
 | \  | \  | \  | \  | --- atemp <= 0.88
     | | | | |--- value: 4401.00
 | | | | |--- value: 4458.00
      |--- instant > 0.16
 | | | | |--- value: 4118.00
      | --- \text{ hum } > -1.00
  | | | | |--- value: 4066.00
  \mid--- registered > -0.02
    |--- casual <= -0.24
| \  | \  | \  | \  | ---  mnth <= -0.15
 | | | | | |--- value: 4362.00
 | | | | |--- value: 4332.00
     | |--- registered > 0.03
| | | | | | |--- value: 4400.50
 | | |--- registered > 0.08
  | | | | | |--- casual <= -0.27
| | | | | | | |--- value: 4492.00
| | | | | | | |--- value: 4507.00
| | | |--- value: 4548.00
| | | | | | |--- value: 4563.00
    |--- casual > -0.24
 | | |--- casual <= 0.17
 | | | | | | | |--- instant <= -0.93
| | | | | | | | | | | | | | | | --- value: 4575.00
 | | | | | | | | |--- instant > -0.93
    | | | | | | | | | | | | --- value: 4540.00
        \mid --- \text{ windspeed} > -0.19
      | | | |--- value: 4590.00
     | | | | |--- casual > -0.14
   | | | | | | | | | | | --- hum <= -0.21
   | | | | | | | | |--- mnth <= -0.59
          | | | | | |--- value: 4595.00
        | | | | | |--- value: 4608.00
      | | | --- season > 0.00
     | | | | |--- value: 4602.00
              | --- \text{hum} > -0.21
```

```
|--- hum <= 0.54
        | \  | \  | \  | \  | \  | --- atemp <= 1.36
      |--- value: 4592.00
 | | | |--- value: 4586.00
       |--- hum > 0.54
                 | | | |--- value: 4576.00
       |--- registered > 0.12
           | | | |--- value: 4629.00
         | |--- windspeed > 0.64
         | | |--- value: 4679.00
          |--- registered > 0.12
            |--- casual <= -0.07
              |---| temp <= 0.49
      | | |--- value: 4570.00
      |--- temp > 0.49
    \mid \quad \mid \quad \mid --- \text{ weekday} \mid <= -0.25
  | | | |--- instant <= -0.74
            | | | | |--- value: 4648.00
       \mid --- \text{ instant} > -0.74
           | | | | | |--- value: 4630.00
      |--- weekday > -0.25
       \mid --- \text{ season} \mid <= 0.00
       | | | | | |--- value: 4677.00
      | | | | | | | | |--- atemp <= 1.23
 | | | | | | | | | | | |--- value: 4661.00
  | | | |--- value: 4656.00
       |--- casual > -0.07
 | | | | | | | | | |--- value: 4780.00
  | | | | | | | | | | | | --- value: 4758.00
        | |--- weathersit > 0.19
  | | | | | | | | | | | --- value: 4708.00
             | --- instant > -0.73
 | | |--- value: 4727.00
 |---| casual > 0.17
       \mid --- \text{ weekday} \le 0.25
  | | | | |--- value: 4665.00
| \ | \ | \ | --- weekday > 0.25
   | | | |--- atemp <= 1.11
      | | | |--- value: 4905.00
        | --- atemp > 1.11
        | | |--- value: 4866.00
|--- casual > 0.94
  \mid ---  registered <= -0.34
| | |--- registered <= -0.40
| | | | |--- value: 4553.00
  | | |--- value: 4475.00
  \mid --- registered > -0.40
| | | | | | |--- value: 4714.00
 | | | | | --- temp > 0.62
      | | |--- value: 4744.00
```

```
|--- hum > 0.56
        | |--- value: 4660.00
    \mid --- \text{ registered} > -0.34
     |--- casual <= 1.16
   | | |--- value: 5046.00
         | --- \text{hum} > 1.06
         | | |--- value: 5010.00
         |--- mnth > 0.86
        | |--- value: 5117.00
       |---| casual > 1.16
         |--- weekday <= 0.00
         | --- season <= 0.00
         | | | |--- value: 4911.00
 | | | | | |--- value: 4906.00
     | | | --- season > 0.00
        | | |--- value: 4881.00
        |--- weekday > 0.00
   | | | |--- value: 4966.00
-- registered > 0.20
 |--- casual <= 0.22
  |--- registered <= 0.33
  | ---  casual <= -0.02
        |---| instant <= -0.67
     | | |--- registered <= 0.31
     | | | |--- weekday <= 0.00
 | | | | | | | |--- value: 4803.00
     | | | | | | --- atemp > 0.65
              | | |--- windspeed <= 0.08
         | | | | | | | |--- value: 4834.00
     | | | | | | | | |--- value: 4845.00
  |--- weekday > 0.00
   | | | | | |--- value: 4791.00
         |--- registered > 0.31
        | | |--- registered <= 0.33
            | | |--- value: 4891.00
            | |--- registered > 0.33
         | | | |--- value: 4864.00
         |--- instant > -0.67
            |--- registered <= 0.25
              |--- weathersit <= 0.19
                  |--- casual <= -0.20
     | | | | | | |--- hum <= 0.21
           | | | | | |--- mnth <= 0.72
      | | | | | | | | |--- value: 4694.00
         | | |--- value: 4687.00
            |--- hum > 0.21
     | | | | | | | |--- value: 4713.00
     | ---  casual > -0.20
              | | |--- value: 4725.00
               |--- weathersit > 0.19
            | | |--- value: 4760.00
            |--- registered > 0.25
        | | |--- windspeed <= -0.46
     | | |--- value: 4748.00
     \mid --- \text{ windspeed} > -0.88
           | | |--- value: 4764.00
```

```
| |--- windspeed > -0.46
             | | | |--- value: 4785.00
            |--- casual > -0.02
           | --- \text{ hum } <= 0.52
           | | |--- weekday <= 0.00
               | | |--- value: 5020.00
                  |--- weekday > 0.00
               | | | season <= 0.90
               | | | |--- value: 4991.00
                    |--- season > 0.90
               | | | |--- value: 4985.00
               |--- hum > 0.52
                 |---| instant <= -1.00
               | |--- value: 4917.00
               |---| instant > -1.00
               | | |--- value: 4844.00
           |--- registered > 0.33
            |--- casual <= -0.14
             |--- season <= 0.90
               | |--- windspeed <= -0.93
               | | |--- value: 5058.00
       \mid \quad \mid --- \text{ windspeed} > -0.93
                 | --- temp <= 1.05
       | | | | | |--- value: 4978.00
             | | | --- temp > 1.05
               | | | |--- value: 4968.00
               |--- season > 0.90
               | |--- value: 4839.00
           |--- casual > -0.14
           | |--- registered <= 0.45
        | --- atemp <= 0.97
               | | |--- atemp <= 0.91
                    | |--- value: 5115.00
               | | | |--- value: 5130.00
                  |---| atemp > 0.97
               | | |--- value: 5084.00
               |--- registered > 0.45
               | |--- windspeed <= 0.21
             | | |--- value: 5180.00
             | | | | | |--- value: 5203.00
               \mid \quad \mid --- \text{ windspeed} > 0.21
               | | |--- windspeed <= 0.86
                  | | |--- registered <= 0.50
        | | | | | |--- value: 5312.00
             | | |--- value: 5298.00
                  |--- windspeed > 0.86
                  | | |--- value: 5225.00
     |--- casual > 0.22
        |--- registered <= 0.57
          |--- registered <= 0.33
          | |--- value: 5362.00
       | | |--- registered > 0.33
        | | |--- value: 5538.00
               |---| temp > 1.01
        | | | |--- value: 5515.00
        |--- registered > 0.57
     | |--- value: 5895.00
--- casual > 1.31
  |---| casual <= 2.04
```

```
|--- hum <= 0.42
           |--- registered <= -0.21
               |---| atemp <= 1.19
        | | | |--- weekday <= 0.00
           | | | |--- value: 5041.00
                |--- weekday > 0.00
           | | | | | |--- value: 5119.00
         | | | | | |--- value: 5191.00
                    | --- mnth > 0.72
                   | | |--- value: 5217.00
                 |---| atemp > 1.19
           | | |--- value: 5302.00
            \mid --- \text{ registered} > -0.21
        | | |--- casual <= 1.42
             | |--- value: 5202.00
               |--- casual > 1.42
          | | |--- windspeed <= -1.05
        | | | | |--- value: 5305.00
        | | | | | |--- value: 5312.00
           | | |--- windspeed <= -0.58
         | | | | | | | |--- value: 5342.00
                 \mid \quad \mid \quad \mid --- \text{ windspeed} > -0.58
                | | | |--- value: 5336.00
              |--- hum > 0.42
          |---| temp <= 0.75
         | | |--- instant <= -0.44
              | |--- value: 5423.00
         | | |--- value: 5409.00
          |---| temp > 0.75
         | | |--- value: 5345.00
      |--- casual > 2.04
         |--- season <= 0.90
            |---| temp <= 1.16
        | | | --- mnth <= -0.15
         | | | |--- value: 5805.00
        | | | |--- value: 5923.00
           |---| temp > 1.16
         | | |--- value: 6043.00
         |--- season > 0.90
        | |--- value: 5511.00
-- instant > 0.34
 |--- registered <= 1.07
   |--- casual <= 2.19
      |--- casual <= -0.30
         |--- registered <= 0.26
   |--- value: 3214.00
     | | |--- instant <= 0.69
         | |--- value: 4378.00
        | --- instant <= 0.59
          | | |--- value: 4367.00
                | --- instant > 0.59
                | | |--- value: 4359.00
```

```
|--- instant > 0.69
    | | |--- value: 4127.00
   |--- registered > 0.26
    |--- registered <= 0.76
    | |--- windspeed <= 0.32
        | --- casual <= -0.50
          | --- instant <= 0.66
            | | |--- value: 5026.00
            | --- instant > 0.66
           | | |--- value: 4972.00
    | | |--- value: 5102.00
           | --- mnth > -0.73
       | | | |--- value: 5115.00
        |--- windspeed > 0.32
    | | | |--- value: 4862.00
          |---| atemp > -0.18
     | | | |--- value: 4717.00
      |--- registered > 0.76
   | | |--- value: 5633.00
|--- casual > -0.30
   |--- registered <= -0.05
    |---| casual <= 0.74
    | --- atemp <= -0.07
      | | |--- value: 4220.00
    | --- atemp > -0.07
    | | | |--- value: 4672.00
    | | | | | --- mnth <= 0.28
          | | |--- value: 4459.00
      | | | | |--- value: 4549.00
     |--- casual > 0.74
    | |--- registered <= -0.31
     | | |--- value: 5169.00
         \mid --- \text{ registered} > -0.31
       | |--- weathersit <= 0.19
         | | |--- value: 4840.00
        | |--- weathersit > 0.19
        | | |--- value: 4996.00
     |--- registered > -0.05
      |--- casual <= 0.18
       |--- registered <= 0.74
      | | |--- registered <= 0.58
    | | | |--- instant <= 0.74
  | | | | | |--- value: 5260.00
   | | | | | |--- value: 5099.00
          |--- registered > 0.58
          | |--- windspeed <= 1.73
    | | | | |--- weekday <= 0.75
    | | | | | |--- value: 5409.00
   | | | | | |--- hum <= -0.65
       | | | | | |--- value: 5463.00
                     |---| hum > -0.65
                  | | |--- value: 5459.00
               |--- windspeed > 1.73
                |---| temp <= -0.15
           | |--- value: 5558.00
                |---| temp > -0.15
```

```
| | | | |--- value: 5585.00
     |--- registered > 0.74
        |--- registered <= 0.98
        | |--- registered <= 0.92
          | --- weekday <= -0.75
            | |--- value: 5572.00
        |--- weekday > -0.75
             |--- season <= -0.90
        | | |--- value: 5847.00
     | --- season > -0.90
      | | |--- weekday <= 0.50
                 | --- \text{ hum } <= -0.07
             |---| atemp <= 0.82
        | | |--- value: 5698.00
        | | |--- value: 5713.00
        | | |--- value: 5728.00
          |---| temp > 0.56
    | | | |--- value: 5741.50
                |--- weekday > 0.50
 | |--- value: 5823.00
        | |--- registered > 0.92
            |--- weathersit <= 0.19
        |---| temp <= 0.72
          | | |--- value: 5918.00
      | --- temp > 0.72
             | | |--- value: 5905.00
       |--- weathersit > 0.19
            | |--- value: 5870.00
        |--- registered > 0.98
        \mid--- season \leq 0.00
    | | | |--- instant <= 0.54
 | | | | | |--- value: 6133.00
       | | | |--- value: 6073.00
         |--- season > 0.00
          | |--- value: 6227.00
        |--- casual > 0.18
     |--- registered <= 0.41
        |---| casual <= 1.34
| | | | |--- value: 5255.00
        | --- atemp > 1.03
        | | |--- windspeed <= -0.33
    | | | | |--- value: 5687.00
      | | |--- windspeed > -0.33
    | | | | |--- value: 5531.00
            | --- \text{hum} > -0.30
             | | |--- value: 5464.00
        |--- casual > 1.34
 |--- registered <= 0.02
          | | |--- value: 5892.00
             |--- registered > 0.02
          | | |--- value: 6041.00
        |--- registered > 0.05
             |--- casual <= 1.72
            | --- atemp <= 1.15
        | | |--- value: 6053.00
        |---| atemp > 1.15
             | | |--- value: 6031.00
```

```
|--- casual > 1.72
          | |--- registered <= 0.22
                 |---| atemp <= 0.18
      | | | |--- value: 6235.00
  | | |--- weekday <= 0.00
      | | |--- value: 6118.00
      \mid --- \text{ weekday} > 0.00
          | | | | |--- value: 6140.00
          | |--- registered > 0.22
      | | | |--- hum <= 0.23
                 | | |--- value: 6304.00
         | --- \text{ hum} > 0.23
             | | | |--- value: 6299.00
          | | | | |--- value: 6359.00
      |--- registered > 0.41
      |--- casual <= 0.82
        |--- registered <= 0.72
      | | | season <= 0.00
| | | | | |--- value: 5936.00
        | --- season > 0.00
  | | | | |--- value: 5786.00
      | |--- registered > 0.72
      | | |--- windspeed <= -0.04
         | | | |--- value: 6398.00
     \mid --- \text{ windspeed} > -0.04
           | | |--- value: 6457.00
      |---| atemp > -0.20
          | |--- registered <= 0.97
           | | |--- instant <= 0.38
          | | | | |--- value: 6153.00
      | | | | |--- value: 6093.00
                |--- instant > 0.38
                   |--- hum <= -1.22
           | | | | |--- value: 6207.00
         | | | | | |--- instant <= 0.63
         | | | | | | | |--- value: 6230.00
                 | | | | --- instant > 0.63
                | | | |--- value: 6241.00
              |--- registered > 0.97
                 |---| temp <= 0.38
          | | | | | | |--- value: 6273.00
         | | | |--- windspeed > 1.27
                 | | | |--- value: 6233.00
              |--- \text{ temp} > 0.28
                | | |--- value: 6192.00
      |---| temp > 0.38
        | | |--- value: 6370.00
             |--- workingday > -0.40
                     |---| atemp <= 0.59
         | | | |--- value: 6312.00
                 | | | |--- value: 6296.00
        |--- casual > 0.82
      | --- \text{ hum } <= -0.37
```

```
|--- hum <= -1.22
                                                      | |--- value: 6460.00
                                                         |--- \text{ hum } > -1.22
                                    | |--- casual <= 1.97
            | | |--- value: 6544.00
                                                              |--- casual > 1.97
                                                         | | |--- value: 6598.00
                                                 |--- hum > -0.37
                    | | | | | | | |--- value: 6685.00
                      | | |--- value: 6597.00
                                        |--- registered > 0.52
                            | | |--- value: 6734.00
                                             | | | |--- value: 6824.00
|--- casual > 2.19
        |--- registered <= 0.63
                |---| atemp <= 1.24
                | |--- registered <= 0.34
       | | | |--- instant <= 0.49
       | | | | |--- value: 6857.00
            | | |--- instant > 0.49
               | | | | | | |--- value: 6591.00
               | | | | |--- value: 6624.00
                            | --- temp > 1.07
                            | | |--- value: 6536.00
                       |--- registered > 0.34
                                |--- casual <= 2.65
             | | | | |--- casual <= 2.40
            | | | | | |--- value: 6969.00
               | | |--- value: 6978.00
                \mid \quad \mid \quad \mid \quad \mid --- \text{ windspeed} > -0.58
             | | | | | |--- value: 6891.00
                | | | | |--- value: 6883.00
                                |--- casual > 2.65
                            | |--- value: 7130.50
                      |---| atemp > 1.24
             | |--- value: 6043.00
       |--- registered > 0.63
                |--- casual <= 2.86
                   |--- registered <= 0.87
                       | |--- casual <= 2.61
                        | | | | |--- value: 7429.00
               | | | | | |--- holiday <= 2.82
                | | | | | | |--- value: 7410.00
                                        \mid \quad \mid --- \text{ holiday} > 2.82
                               | | | |--- value: 7403.00
                        |--- casual > 2.61
                                     \mid --- \text{ windspeed} \mid --
                                    | |--- value: 7498.00
               \mid \quad \mid \quad \mid \quad \mid --- \text{ windspeed} > -0.37
                             | | |--- value: 7459.00
```

```
|--- registered > 0.87
          | |--- value: 7641.00
        |--- casual > 2.86
          |--- windspeed <= -1.20
          | |--- value: 8294.00
           \mid --- \text{ windspeed} > -1.20
             |--- hum <= -0.52
           | | |--- value: 7702.00
             |--- hum > -0.52
             | |--- registered <= 0.77
          | | | |--- value: 7836.00
             | |--- value: 7865.00
          -- registered > 1.07
 |--- registered <= 1.43
    |--- weekday <= 1.25
      |--- casual <= 0.31
        | |--- registered <= 1.21
         | |--- casual <= 0.21
              | |--- weekday <= -0.25
              | | |--- value: 6290.00
             | --- weekday > -0.25
                \mid \quad \mid --- \text{ windspeed} \mid <= -0.80
         | | | | | |--- value: 6196.00
             | | | | | --- atemp > 0.74
                  | | | |--- value: 6211.00
                  \mid --- \text{ windspeed} > -0.80
                | | |--- value: 6169.00
              |--- casual > 0.21
               | |--- weathersit <= 0.19
               | | |--- value: 6436.00
                  |--- weathersit > 0.19
               | | |--- value: 6530.00
           |--- registered > 1.21
             |--- registered <= 1.36
              | --- \text{ hum } <= 0.69
               | | | --- hum <= -0.74
                  | | |--- value: 6691.00
                     |--- casual > 0.25
                    | | |--- value: 6660.00
                     |--- \text{ hum } > -0.74
                  | | | --- temp <= 1.21
                  | | |--- value: 6565.00
                     | | --- windspeed > -0.87
             | | |--- value: 6570.50
                |---| temp > 1.21
                       | |--- weathersit <= 0.19
           | | | | | |--- value: 6591.00
                            |--- weathersit > 0.19
                     | | |--- value: 6664.00
                     |--- hum > 0.69
                | |--- value: 6421.00
             |--- registered > 1.36
              | |--- atemp <= 1.68
               | | |--- registered <= 1.40
                  | | |--- value: 6861.00
                     |--- registered > 1.40
                  | | |--- value: 6825.00
                  |---| atemp > 1.68
                  | |--- value: 6786.00
        |--- casual > 0.31
```

```
|--- weekday \leq 0.75
      | |--- registered <= 1.35
            \mid --- \text{ windspeed} \mid --- \text{ o.31}
      | | | | |--- value: 6772.00
            |--- \text{ hum } > -1.10
            | | | | | |--- value: 6830.00
      | | | | | | | |--- casual <= 0.62
      | | | | | | | |--- value: 6883.00
                | ---  casual > 0.62
              | | | |--- value: 6871.00
            \mid --- \text{ windspeed} > -0.31
              |---| mnth <= -0.30
              | |--- value: 6770.00
    | | | |--- value: 6779.00
       |--- registered > 1.28
            | | | | | | | |--- value: 6784.00
    | | |--- registered > 1.35
    | | | | |--- value: 6966.00
           |--- casual > 0.49
          | | |--- value: 7013.00
       \mid --- \text{ weekday} > 0.75
    | | |--- instant <= 0.80
    | | | | |--- value: 7030.00
      | | |--- value: 6904.00
       |---| mnth > 0.28
   | | | |--- value: 7175.00
         \mid --- \text{ windspeed} > 0.43
      | | | |--- value: 7148.00
  |--- weekday > 1.25
  | |--- value: 8120.00
\mid--- registered > 1.43
  |--- casual <= 0.40
    |--- registered <= 1.61
  | | | |--- value: 6879.00
    |--- weathersit > 0.19
  | | | | |--- value: 6855.00
    | ---  casual > 0.12
    | --- weekday <= -0.25
            | | |--- instant <= 1.12
       | \  | \  | \  | \  | \  | ---  mnth <= 0.14
    | | | | | | | | |--- value: 7006.00
                |--- instant > 1.12
               | | |--- value: 7040.00
       |--- weekday > -0.25
            | | |--- value: 7055.00
            |---| atemp > 1.24
       | |--- value: 7105.00
          |--- hum > 0.52
         | |--- value: 6917.00
```

```
|--- registered > 1.61
     |--- registered <= 1.71
        |---| temp <= 1.15
       | |--- windspeed <= -0.75
    | | | |--- value: 7375.00
         |--- windspeed > -0.75
        | | |--- value: 7336.50
       |---| temp > 1.15
    | | |--- registered <= 1.67
         | |--- value: 7216.00
    |---| registered > 1.67
       | | |--- value: 7262.50
     |--- registered > 1.71
        |---| mnth <= 0.28
         |---| atemp <= 0.66
         | |--- value: 7494.00
          |---| atemp > 0.66
    | | | | |--- value: 7424.00
             |--- windspeed > -0.72
          | | | | | |--- value: 7442.00
         | --- temp > 0.97
      | | | | |--- value: 7446.00
        |---| mnth > 0.28
       | |--- value: 7580.00
     |--- casual > 0.40
  \mid --- \text{ instant} \mid <= 0.45
  | |--- value: 8362.00
  |--- instant > 0.45
  | |--- registered <= 1.93
  | | |--- registered <= 1.64
      | |--- casual <= 0.91
      | |--- value: 7384.00
    | | | | |--- value: 7348.50
      | | | | |--- registered > 1.61
          | | | | |--- value: 7363.00
       | | | | | |--- value: 7273.00
     | | | | | |--- windspeed > -0.74
          | | | | | | | |--- value: 7290.00
      | | |--- value: 7286.00
            | | | --- casual > 0.91
    | --- instant <= 0.72
             | | |--- value: 7639.00
             \mid --- \text{ instant} > 0.72
         | | |--- value: 7665.00
             |---| temp > 1.04
           | | |--- value: 7499.00
        |--- registered > 1.64
           |--- windspeed <= 1.12
           | --- atemp <= 1.03
            | |--- registered <= 1.80
             | | | | |--- value: 7736.00
               | --- \text{hum} > -0.53
```

```
| | |--- value: 7765.00
           |--- registered > 1.80
          | |--- value: 7697.00
          -- atemp > 1.03
           |--- instant <= 1.13
              |---| atemp <= 1.43
                 |---| hum <= -0.43
                 | |--- value: 7605.00
                 |--- hum > -0.43
                 | --- atemp <= 1.24
                 | | |--- value: 7582.00
                   | --- atemp > 1.24
                 | | |--- value: 7592.00
               |---| atemp > 1.43
            | | |--- value: 7534.00
           |--- instant > 1.13
          | |--- value: 7713.00
    |--- windspeed > 1.12
  | |--- value: 7421.00
- registered > 1.93
|--- value: 8173.00
```

In []: