from sklearn.model selection import train test split from sklearn.preprocessing import StandardScaler from sklearn.preprocessing import OneHotEncoder from sklearn.pipeline import make pipeline from sklearn.compose import ColumnTransformer from sklearn.compose import make\_column\_selector as selector from sklearn.neighbors import KNeighborsClassifier from sklearn.model selection import cross val score from sklearn.pipeline import Pipeline from sklearn.model selection import GridSearchCV df = pd.read\_csv('blood\_transfusion.csv') df.head() Out[4]: Recency Frequency Monetary Time Class 0 2 50 12500 98 donated 0 3250 13 28 donated 2 1 16 4000 35 donated 3 2 20 5000 45 donated 4 1 24 6000 77 not donated In [8]: targetcol = 'Class' df[targetcol].value counts() Out[8]: Class 570 not donated donated 178 Name: count, dtype: int64 This is a classification problem. Scaling the data using a StandardScaler is important when using a KNeighborsClassifier because it ensures that all features contribute equally to distance calculations, and into more accurate and effective classification. In [19]: target = df['Class'] features = df.drop(columns='Class') In [25]: numerical\_columns\_selector = selector(dtype\_exclude=object) In [27]: | numerical\_columns = numerical\_columns\_selector(features) In [30]: |x\_train, x\_test, y\_train, y\_test = train\_test\_split( features, target, random\_state=123 In [36]: numerical\_preprocessor = StandardScaler() preprocessor = ColumnTransformer([ ('standard\_scaler', numerical\_preprocessor, numerical\_columns) ]) In [38]: model = KNeighborsClassifier() In [40]: pipeline = make\_pipeline(preprocessor, model) pipeline **Pipeline** Out [40]: ▶ columntransformer: ColumnTransformer ▶ standard\_scaler ▶ StandardScaler ▶ KNeighborsClassifier pipeline.fit(x\_train, y\_train) **Pipeline** Out[42]: :▶ ▶ columntransformer: ColumnTransformer ▶ standard\_scaler StandardScaler ▶ KNeighborsClassifier In [44]: pipeline.score(x\_test, y\_test) Out [44]: 0.7807486631016043 By default it is 5 In [53]: res = cross\_val\_score(pipeline, x\_train, y\_train, cv=5, scoring='roc\_auc') np.mean(res) Out [53]: 0.6956883985953752 In [55]: knn = KNeighborsClassifier() parameter\_range = [1, 2, 5, 10, 20, 50, 100, 200, 500] hyper grid = {'knn n neighbors': parameter range} pipeline2 = Pipeline([('prep', preprocessor), ('knn', knn)]) grid\_search = GridSearchCV(pipeline2, hyper\_grid, cv=5, scoring='roc\_auc', n\_jobs=-1) results = grid search.fit(x train, y train) abs(results.best\_score\_) /opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py:778: UserWarning: Scoring failed. The score on this train-test partition n for these parameters will be set to nan. Details: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/ scorer.py", line 373, in score y\_pred = method\_caller(clf, "decision\_function", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/utils/ available if.py", line 32, in get if not self.check(obj): ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 46, in check getattr(self.\_final\_estimator, attr) AttributeError: 'KNeighborsClassifier' object has no attribute 'decision\_function' During handling of the above exception, another exception occurred: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py", line 767, in score scores = scorer(estimator, X\_test, y\_test) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 234, in \_\_call\_\_ return self.\_score( ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 387, in \_score y\_pred = method\_caller(clf, "predict\_proba", X) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 547, in predict\_proba return self.steps[-1][1].predict\_proba(Xt, \*\*predict\_proba\_params) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_classification.py", line 283, in predict\_proba neigh\_ind = self.kneighbors(X, return\_distance=False) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_base.py", line 810, in kneighbors raise ValueError( ValueError: Expected n\_neighbors <= n\_samples, but n\_samples = 448, n\_neighbors = 500 warnings.warn( /opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py:778: UserWarning: Scoring failed. The score on this train-test partition n for these parameters will be set to nan. Details: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 373, in \_score y\_pred = method\_caller(clf, "decision\_function", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/utils/\_available\_if.py", line 32, in \_\_get\_\_ if not self.check(obj): ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 46, in check getattr(self.\_final\_estimator, attr) AttributeError: 'KNeighborsClassifier' object has no attribute 'decision\_function' During handling of the above exception, another exception occurred: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py", line 767, in \_score scores = scorer(estimator, X\_test, y\_test) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 234, in \_\_call\_\_ return self.\_score( File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 387, in \_score y\_pred = method\_caller(clf, "predict\_proba", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 547, in predict\_proba return self.steps[-1][1].predict\_proba(Xt, \*\*predict\_proba\_params) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_classification.py", line 283, in predict\_proba neigh\_ind = self.kneighbors(X, return\_distance=False) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_base.py", line 810, in kneighbors raise ValueError( ValueError: Expected n\_neighbors <= n\_samples, but n\_samples = 449, n\_neighbors = 500 warnings.warn( /opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py:778: UserWarning: Scoring failed. The score on this train-test partition n for these parameters will be set to nan. Details: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 373, in \_score y\_pred = method\_caller(clf, "decision\_function", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/utils/\_available\_if.py", line 32, in \_\_get\_\_ if not self.check(obj): ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 46, in check getattr(self.\_final\_estimator, attr) AttributeError: 'KNeighborsClassifier' object has no attribute 'decision\_function' During handling of the above exception, another exception occurred: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py", line 767, in \_score scores = scorer(estimator, X\_test, y\_test) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 234, in \_\_call\_\_ return self.\_score( ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 387, in \_score y\_pred = method\_caller(clf, "predict\_proba", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 547, in predict\_proba return self.steps[-1][1].predict\_proba(Xt, \*\*predict\_proba\_params) ^^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_classification.py", line 283, in predict\_proba neigh\_ind = self.kneighbors(X, return\_distance=False) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_base.py", line 810, in kneighbors raise ValueError( ValueError: Expected n\_neighbors <= n\_samples, but n\_samples = 449, n\_neighbors = 500 warnings.warn( /opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py:778: UserWarning: Scoring failed. The score on this train-test partition n for these parameters will be set to nan. Details: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 373, in \_score y\_pred = method\_caller(clf, "decision\_function", X) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/utils/\_available\_if.py", line 32, in \_\_get\_\_ if not self.check(obj): ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 46, in check getattr(self.\_final\_estimator, attr) AttributeError: 'KNeighborsClassifier' object has no attribute 'decision\_function' During handling of the above exception, another exception occurred: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py", line 767, in \_score scores = scorer(estimator, X\_test, y\_test) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 234, in \_\_call\_\_ return self.\_score( ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 387, in \_score y\_pred = method\_caller(clf, "predict\_proba", X) File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 547, in predict\_proba return self.steps[-1][1].predict\_proba(Xt, \*\*predict\_proba\_params) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_classification.py", line 283, in predict\_proba neigh\_ind = self.kneighbors(X, return\_distance=False) ^^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_base.py", line 810, in kneighbors raise ValueError( ValueError: Expected n\_neighbors <= n\_samples, but n\_samples = 449, n\_neighbors = 500 warnings.warn( /opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py:778: UserWarning: Scoring failed. The score on this train-test partition n for these parameters will be set to nan. Details: Traceback (most recent call last): File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 373, in \_score y\_pred = method\_caller(clf, "decision\_function", X) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call return getattr(estimator, method)(\*args, \*\*kwargs) ^^^^^ File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/utils/\_available\_if.py", line 32, in \_\_get\_\_ if not self.check(obj):

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 46, in check

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_validation.py", line 767, in \_score

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 234, in \_\_call\_\_

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 387, in \_score

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/pipeline.py", line 547, in predict\_proba

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/\_scorer.py", line 73, in \_cached\_call

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_base.py", line 810, in kneighbors

File "/opt/anaconda3/lib/python3.11/site-packages/sklearn/neighbors/\_classification.py", line 283, in predict\_proba

/opt/anaconda3/lib/python3.11/site-packages/sklearn/model\_selection/\_search.py:952: UserWarning: One or more of the test scores are non-finite: [0.5723481

AttributeError: 'KNeighborsClassifier' object has no attribute 'decision\_function'

During handling of the above exception, another exception occurred:

getattr(self.\_final\_estimator, attr)

scores = scorer(estimator, X\_test, y\_test)

^^^^^^

y\_pred = method\_caller(clf, "predict\_proba", X)

return getattr(estimator, method)(\*args, \*\*kwargs)

neigh\_ind = self.kneighbors(X, return\_distance=False)

1 0.62710031 0.6956884 0.74711621 0.7741072 0.77498509

nan]

^^^^^^

return self.steps[-1][1].predict proba(Xt, \*\*predict proba params)

^^^^^^

^^^^^^^

ValueError: Expected n\_neighbors <= n\_samples, but n\_samples = 449, n\_neighbors = 500

Traceback (most recent call last):

^^^^^

return self.\_score(

raise ValueError(

0.77238621 0.77877162

warnings.warn(

warnings.warn(

Out[57]: {'knn\_n\_neighbors': 200}

Out [55]: 0.7787716159809183

In [57]: results.best\_params\_

In [ ]:

In [4]: import pandas as pd

import numpy as np

from sklearn import set\_config
set\_config(display='diagram')

from sklearn.linear\_model import LinearRegression