

In [97]:

```
C:\Users\chinna.s\AppData\Local\Continuum\anaconda3\lib\site-package
s\IPython\core\interactiveshell.py:3326: DtypeWarning: Columns
(4,11,12,16,1
7,18,19,20,40,41,42,43,44,45,52,53,54,55,56,74,75,76,77,78,79,80,81,82,83,8
4,85,86,87,88,89,90,108,110) have mixed types.Specify dtype option on import
or set low_memory=False.
    exec(code_obj,          self.user_global_ns,
self.user_ns) Out[97]:
```

id	has_null	wave	gender	age	age_o	d_age	d_d_age	race	race_
0	1	0	1	female	21	27	6	[4-6]	Asian/Pacific European/Caucasian Islander/Asian- America
									American
1	2	0	1	female	21	22	1	[0-1]	Asian/Pacific European/Caucasian Islander/Asian- America
									American
2	3	1	1	female	21	22	1	[0-1]	Asian/Pacific Islander/Asian- American
									Asian/Pacifi Islander/Asian America

3		0	1	female	2	1	2	3	2	[2-3]	Asian/Pa cific Islander/ Asian- America n
4											
4		0	1	female	2	1	2	4	3	[2-3]	Asian /Pacific Islander /Asian- America n
5											

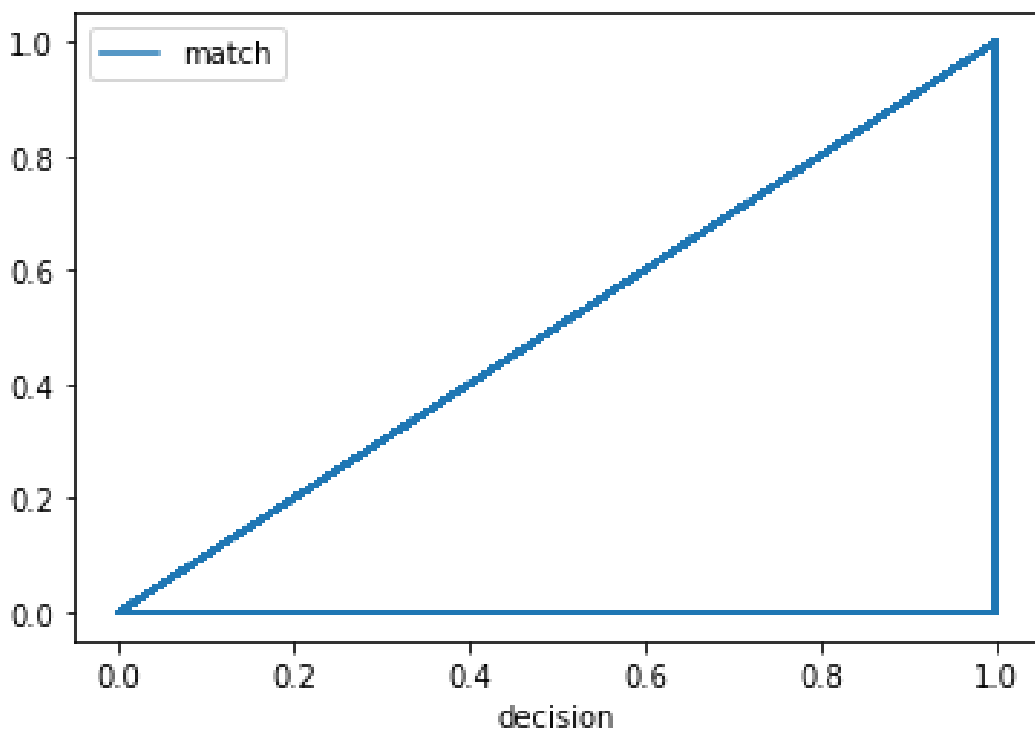
European/Caucasian
America

Latino/Hispani
America

5 rows x 124 columns

Out[98]:

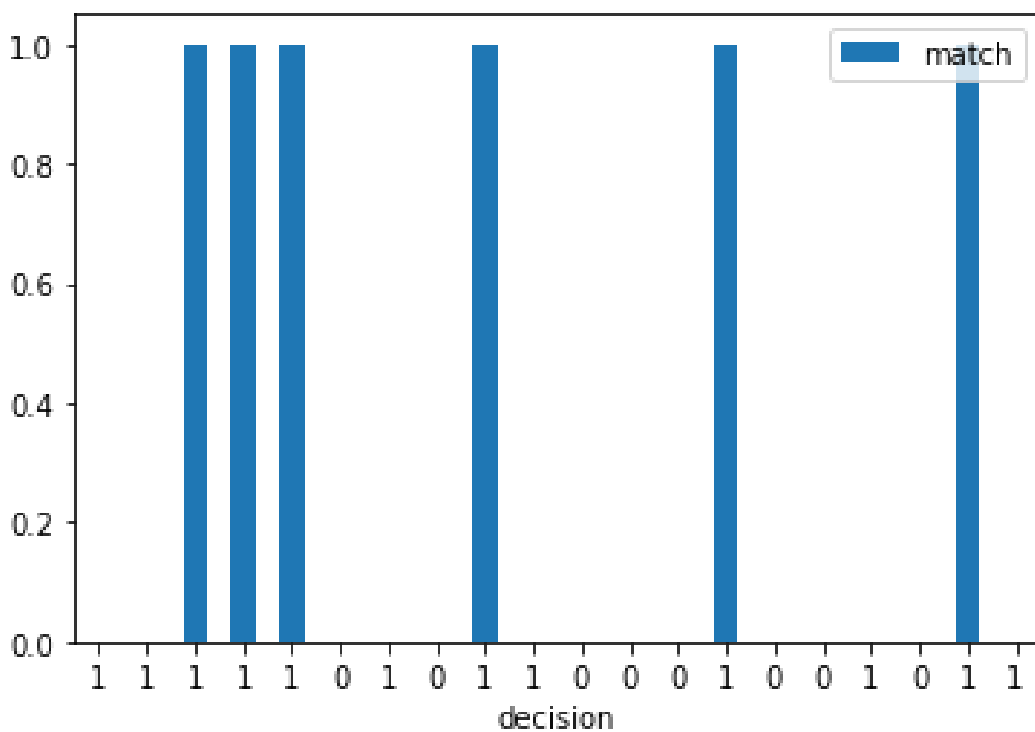
<AxesSubplot:xlabel='decision'>

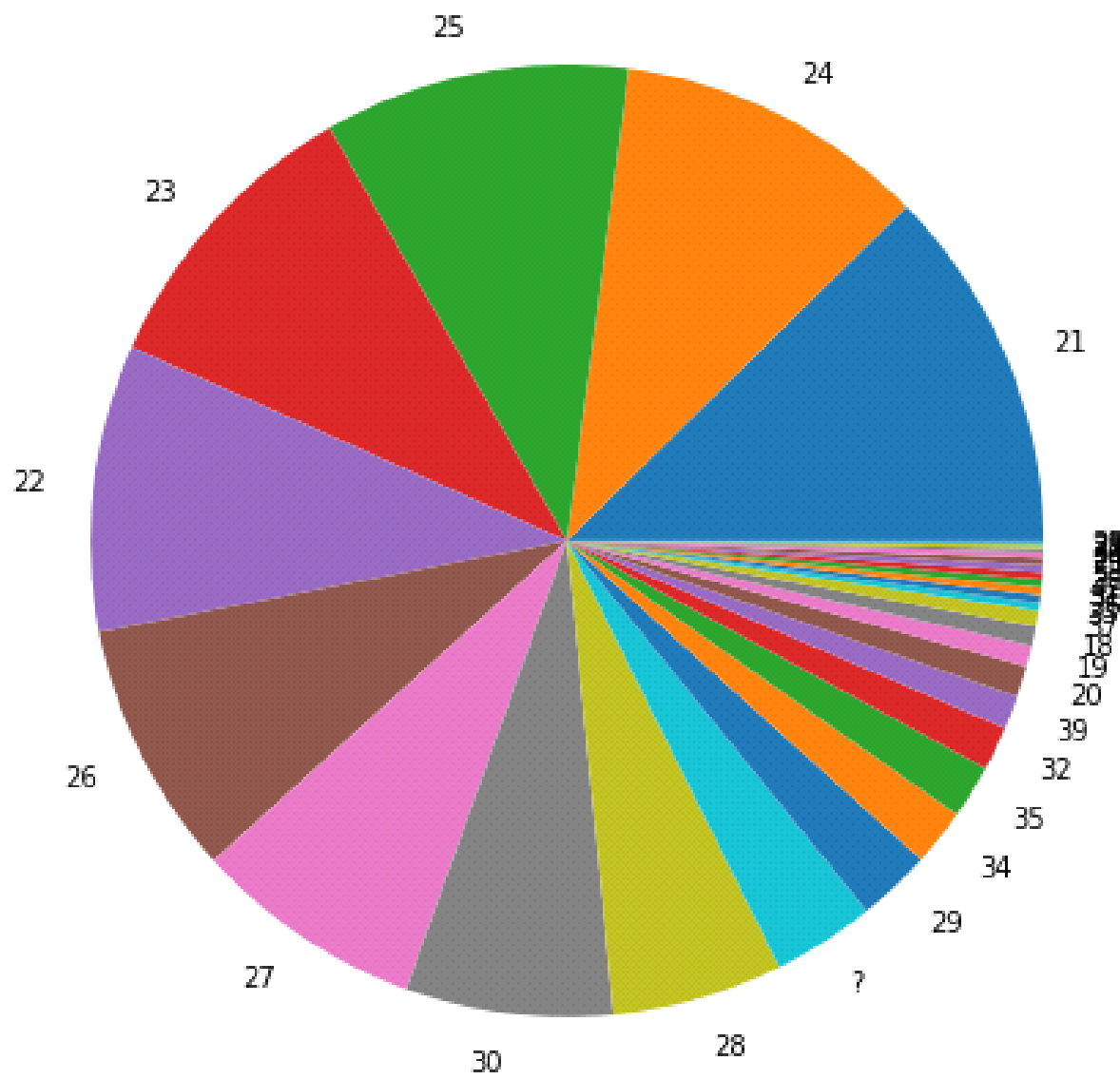


In [99]:

Out[99]:

<AxesSubplot:xlabel='decision'>





27	1037
23	884
26	869
24	841
25	815

28	724
22	655
29	589
30	486
21	291
32	210
33	161
34	152
31	125
?	95
30	88
35	60

20	55	
36	45	
24	22	
28	22	
27	22	
25	22	
19	20	
42	20	
38	19	
39	18	
18	10	
23	10	
55	6	
37	5	
Name:	age,	dtype: int64

In [138]:

In [139]:

In [140]:

Out[140]:

a
r
r
a
y
(
[
[
0
]
,
[
0
]
,

```
[
0
],
.
.
.
,
[
0
],
[
0
],
,
[0], dtype=int64)
```

Out[141]:

3545	0			
3132	0			
5432	1			
6184	1			
1949	1			
	..			
4373	1			
7891	0			
4859	0			
3264	0			
2732	0			
Name: decision,		Length: 5585,	dtype:	int64

In [142]:

Out[142]:

```
a
r
r
a
y
(
[
[
0
]
],
[
0
```

```

]
,
[
0
]
,
.
.
.
,
[
0
]
,
[
1
]
,
[0]], dtype=int64)

```

In [143]:

Out[143]:

2265	0		
2851	0		
3655	0		
196	0		
3719	0		
	..		
2696	0		
2126	0		
282	0		
6512	1		
2448	0		
Name: decision, Length: 2793,		dtype:	int64

[0 0 1 ... 0 0 0]

In [145]:

[0 0 0 ... 0 1 0]

In [146]:

[0 0 0 ... 0 1 0]

In [147]:

C:\Users\chinna.b\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\tree\tree.py:163: DeprecationWarning: `np.int` is a deprecated alias for the builtin `int`. To silence this warning, use `int` by itself. Doing this will not modify any behavior and is safe. When replacing `np.int`, you may wish to use e.g. `np.int64` or `np.int32` to specify the precision. If you wish to review your current use, check the release note link for additional information.

Deprecated in [NumPy](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [1.20](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations); for more details and guidance: <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations>

```
y_encoded = np.zeros(y.shape,
```

```
dtype=np.int) Out[148]:
```

```
DecisionTreeClassifier(class_weight=None, criterion='entropy',
max_depth=None,
                        max_features=None, max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, presort=False,
                        random_state=None, splitter='best')
```

In [149]:

In [150]:

Accuracy Score on train data:

0.74162

93643688451 Accuracy Score on test data:

0.7511

636233440745

In [151]:

In [152]:

C:\Users\chinna.b\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\svm\base.py:193: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

```
"avoid this warning.",
```

FutureWarning) Out[153]:

```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,  
    decision_function_shape='ovr', degree=3, gamma='auto_deprecated',  
    kernel='rbf', max_iter=-1, probability=False, random_state=None,  
    shrinking=True, tol=0.001, verbose=False)
```

In [154]:

```
[0 0 0 ... 0 1 0]
```

In [155]:

```
Accuracy: 0.7511636233440745
```

In [156]:

In [157]:

```
Final Accuracy: 0.7416293643688451
```

C:\Users\chinna.b\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\linear_model\base.py:291: DeprecationWarning: `np.int` is a deprecated alias for the builtin `int`. To silence this warning, use `int` by itself. Doing this will not modify any behavior and is safe. When replacing `np.int`, you may wish to use e.g. `np.int64` or `np.int32` to specify the precision. If you wish to review your current use, check the release note link for additional information.

[Deprecated in](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [NumPy](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [1.20; for more details and guidance:](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)
[HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)
[https://numpy.org/devdocs/release/1.20.0-](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)
[notes.html#deprecations](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [evdocs/release/1.20.0-](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)
[notes.html#deprecations](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [\(https://numpy.org/devdocs/release/1.20.0-](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)
[notes.html#deprecations\)](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)

```
indices = (scores > 0).astype(np.int)
```

In [159]:

In [160]:


```

MLA_columns = []
MLA_compare = pd.DataFrame(columns =
MLA_columns) row_index = 0
for alg in MLA:
    predicted = alg.fit(X_train, y_train).predict(X_test) fp,
    tp, th = roc_curve(y_test, predicted)
    MLA_name = alg.__class__.__name__
    MLA_compare.loc[row_index, 'MLA Name'] = MLA_name
    MLA_compare.loc[row_index, 'MLA Train Accuracy'] = round(alg.score(X_train, y_train), 4)
    MLA_compare.loc[row_index, 'MLA Test Accuracy'] = round(alg.score(X_test, y_test), 4)
    MLA_compare.loc[row_index, 'MLA Precision'] = precision_score(y_test,
predicted) MLA_compare.loc[row_index, 'MLA Recall'] = recall_score(y_test,
predicted)
    MLA_compare.loc[row_index, 'MLA AUC'] =
    auc(fp, tp) row_index+=1
MLA_compare.sort_values(by = ['MLA Test Accuracy'], ascending =
False, inplace = True) MLA_compare

```

deprecated alias for the builtin `int`. To silence this warning, use `int` by itself. Doing this will not modify any behavior and is safe. When replacing ``np.int``, you may wish to use e.g. ``np.int64`` or ``np.int32`` to specify the precision. If you wish to review your current use, check the release note link for additional information.

[Deprecated in](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [NumPy](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [1.20.0](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations); [for more details and guidance:](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations> <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations> <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations> <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations> <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations>

```

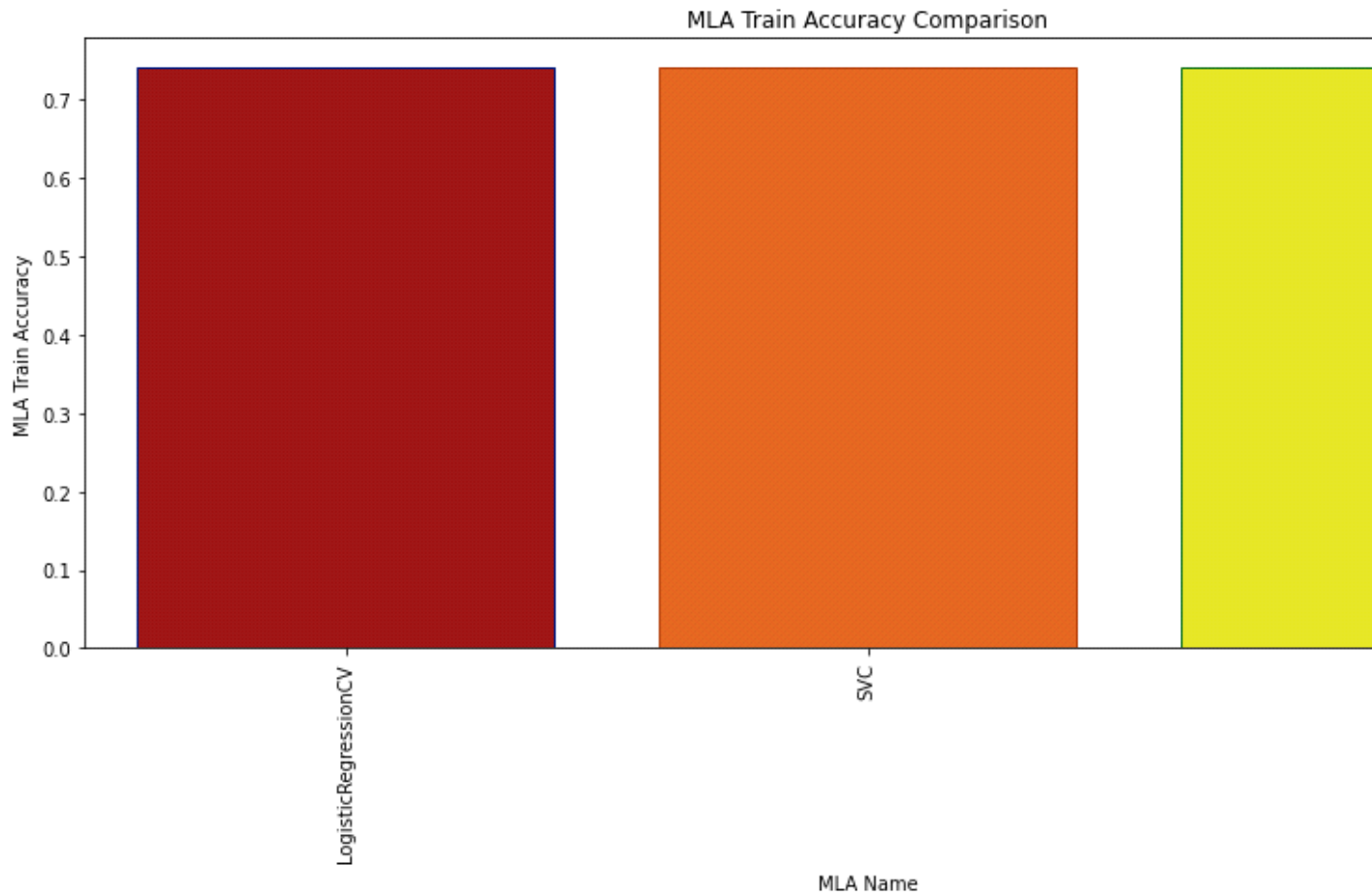
indices = (scores > 0).astype(np.int)
C:\Users\chinna.b\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\linear_model\base.py:291: DeprecationWarning: `np.int` is a deprecated alias for the builtin `int`. To silence this warning, use `int` by itself. Doing this will not modify any behavior and is safe. When

```

replacing `np.int`, you may wish to use e.g. `np.int64` or `np.int32` to specify the precision. If you wish to review your current use, check the release note link for additional information.

Deprecated in [NumPy](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) ["https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) ["https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [1.20:](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [for more details and guidance:](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) ["https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) ["https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [g/devdocs/release/1.20.0-notes.html#deprecations](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [HYPERLINK](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) ["https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [cs/release/1.20.0-notes.html#deprecations\)](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)

indices = (scores > 0).astype(np.int)



In [162]:

```
with open('model.pkl', 'wb') as files:
    pickle.dump(final_model,
                files)
```

In [163]:

```
with open('model.pkl', 'rb') as f: lr = pickle.load(f)
```

In [164]:

```
y_pr=lr.predict(X_test)
```

C:\Users\chinna.b\AppData\Local\Continuum\anaconda3\lib\site-package
s\sklearn\linear_model\base.py:291: DeprecationWarning: `np.int` is a deprec
ated alias for the builtin `int`. To silence this warning, use `int` by itse lf.
Doing this will not modify any behavior and is safe. When replacing `np. int`,
you may wish to use e.g. `np.int64` or `np.int32` to specify the preci sion. If
you wish to review your current use, check the release note link fo r
additional information.

Deprecated in [HYPERLINK "https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) NumPy [HYPERLINK "https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) 1.20; for more details and guidance: [HYPERLINK "https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations"](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [https://numpy.org/d](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) [evdocs/release/1.20.0-notes.html#deprecations](https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations) (https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations)

```
indices = (scores > 0).astype(np.int)
```

y_pr

In [166]:

Out[166]:

```
array([0, 0, 0, ..., 0, 1, 0], dtype=int64)
```

```
print("Accuracy:", metrics.accuracy_score(y_test, y_pr))
```

In [167]:

Accuracy: 0.7511636233440745

In []:

In []: