

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns

from sklearn.model_selection import train_test_split

from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import confusion_matrix

import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:

```
twitch = pd.read_csv("twitchdata-update (1).csv")
twitch = twitch.loc[0:500,]
twitch.head()
```

Out[3]:

	Channel	Watch time(Minutes)	Stream time(minutes)	Peak viewers	Average viewers	Followers	Followers gained	Views gained
0	xQcOW	6196161750	215250	222720	27716	3246298	1734810	93036735
1	summit1g	6091677300	211845	310998	25610	5310163	1370184	89705964
2	Gaules	5644590915	515280	387315	10976	1767635	1023779	102611607
3	ESL_CSGO	3970318140	517740	300575	7714	3944850	703986	106546942
4	Tfue	3671000070	123660	285644	29602	8938903	2068424	78998587

In [8]:

```
twitch2 = pd.get_dummies(twitch, columns = ["Language"], drop_first = False)
twitch2.head()
```

Out[8]:

	Channel	Watch time(Minutes)	Stream time(minutes)	Peak viewers	Average viewers	Followers	Followers gained	Views gained
0	xQcOW	6196161750	215250	222720	27716	3246298	1734810	93036735
1	summit1g	6091677300	211845	310998	25610	5310163	1370184	89705964
2	Gaules	5644590915	515280	387315	10976	1767635	1023779	102611607
3	ESL_CSGO	3970318140	517740	300575	7714	3944850	703986	106546942
4	Tfue	3671000070	123660	285644	29602	8938903	2068424	78998587

5 rows × 28 columns

In [9]:

```
twitch2 = twitch2[['Language_Spanish', "Followers", "Views gained"]]
```

```
twitch2.head()
```

Out[9]:

	Language_Spanish	Followers	Views gained
0	0	3246298	93036735
1	0	5310163	89705964
2	0	1767635	102611607
3	0	3944850	106546942
4	0	8938903	78998587

In [11]:

```
twitch2 = twitch2[['Language_Spanish', 'Followers', 'Views gained']]
twitch2.head()
```

Out[11]:

	Language_Spanish	Followers	Views gained
0	0	3246298	93036735
1	0	5310163	89705964
2	0	1767635	102611607
3	0	3944850	106546942
4	0	8938903	78998587

In [12]:

```
x = twitch2[['Followers', 'Views gained']]
x.head()
```

Out[12]:

	Followers	Views gained
0	3246298	93036735
1	5310163	89705964
2	1767635	102611607
3	3944850	106546942
4	8938903	78998587

In [13]:

```
y = twitch2["Language_Spanish"]
y.head()
```

Out[13]:

0	0
1	0
2	0
3	0
4	0

Name: Language\_Spanish, dtype: uint8

In [14]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

In [34]:

```
knn7 = KNeighborsClassifier(n_neighbors = 3)
```

```
knn7.fit(x_train, y_train)
```

```
Out[34]: KNeighborsClassifier(n_neighbors=3)
```

```
In [35]: y_test_preds = knn7.predict(x_test)
```

```
In [36]: confusion_matrix(y_test, y_test_preds)
```

```
Out[36]: array([[93,  1],  
               [ 6,  1]])
```

```
In [37]: Sensitivity7 = 1/(6 + 1)  
Sensitivity7
```

```
Out[37]: 0.14285714285714285
```

```
In [38]: specificity7 = 93/(93 + 1)  
specificity7
```

```
Out[38]: 0.9893617021276596
```