Kevin De La Torre

CS 4650

Project 4 - Titanic

Titanic Death Predictions

Notebook Link:
https://www.kaggle.com/kevindlt/titanicpredictions

Prediction:

Screenshots on next page

|

|

|

V

In [1]:

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under t
he input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as out
put when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the cur
rent session
```

```
/kaggle/input/titanic/train.csv
/kaggle/input/titanic/test.csv
/kaggle/input/titanic/gender_submission.csv
```

In [2]:

```
train_data = pd.read_csv( "/kaggle/input/titanic/train.csv" )
train_data.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [3]:

```
test_data = pd.read_csv( "/kaggle/input/titanic/test.csv" )
test_data.head()
```

Out[3]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

In [4]:

```
women = train_data.loc[ train_data.Sex == 'female'][ "Survived" ]
rate_women = sum(women)/len(women)

print( "% of women who survived:", rate_women)
```

```
% of women who survived: 0.7420382165605095
```

In [5]:

```
men = train_data.loc[ train_data.Sex == 'male' ][ "Survived" ]
rate_men = sum(men)/len(men)

print( "% of men who survived:", rate_men)
```

```
% of men who survived: 0.18890814558058924
```

In [6]:

```
from sklearn.ensemble import RandomForestClassifier

y = train_data["Survived"]

features = ["Pclass", "Sex", "SibSp", "Parch"]
x = pd.get_dummies( train_data[ features ] )
x_test = pd.get_dummies( test_data[ features ] )

model = RandomForestClassifier( n_estimators=100, max_depth=5, random_state=1 )
model.fit( x, y )
predictions = model.predict( x_test )

output = pd.DataFrame( {'PassengerId': test_data.PassengerId, 'Survived': predictions } )
output.to_csv( 'my_submission.csv', index=False )
print( "Your submission was successfully saved!" )
```

```
Your submission was successfully saved!
```