

Data Mining-Term project

Titanic- Machine Learning from Disaster



Description of the project

In this project we will develop a machine learning model which predicts the survival of the passengers

There are three data files in this project:

1.Train.csv

2.Test.csv

3.Gender_submission.csv

Train.csv In this data there is an information about the passengers on board; specifically, it includes 891 passengers data. This data includes the survival details of the passenger

“1” indicates that the passenger survived.

“0” indicates that the passenger didn’t survived.

Test.csv In this we have to predict whether the remaining 418 passengers on board survived by using patterns.

Gender submission.csv This file is provided as an example that shows how we should structure our predictions. It predicts that all female passengers survived and all male passengers died.

Implementation:

Importing libraries

```

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 the input directory

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

```

Load the data

Training

```
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

Testing

```
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

MachineLearning model

```
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('submission.csv', index=False)  
print("Your submission was successfully saved!")
```

Your submission was successfully saved!