In [ ]:

```
import pandas as pd
import numpy as np
```

In [ ]:

```
data = pd.read_csv('/Users/jacksonlu/Downloads/titanic.csv')
pclass = data['pclass'].count()
sex = data['sex'].count()
age = data['age'].count()
```

In [ ]:

```
male_survived_count = data[(data['sex'] == 'male') & (data['survived'] == 'y
female_survived_count = data[(data['sex'] == 'female') & (data['survived'] =
male_not_survived_count = data[(data['sex'] == 'male') & (data['survived'] =
female_not_survived_count = data[(data['sex'] == 'female') & (data['survivec

adult_survived_count = data[(data['age'] == 'adult') & (data['survived'] ==
child_survived_count = data[(data['age'] == 'child') & (data['survived'] ==
adult_not_survived_count = data[(data['age'] == 'adult') & (data['survived']
child_not_survived_count = data[(data['age'] == 'child') & (data['survived']

first_class_survived_count = data[(data['pclass'] == '1st') & (data['survive
second_class_survived_count = data[(data['pclass'] == '2nd') & (data['surviv
third_class_survived_count = data[(data['pclass'] == '3rd') & (data['survive
crew_survived_count = data[(data['pclass'] == 'crew') & (data['survived'] ==
first_class_not_survived_count = data[(data['pclass'] == '1st') & (data['sur
second_class_not_survived_count = data[(data['pclass'] == '2nd') & (data['su
third_class_not_survived_count = data[(data['pclass'] == '3rd') & (data['sur
crew_not_survived_count = data[(data['pclass'] == 'crew') & (data['survived'
```

In [ ]:

```
def entropy(survived, not_survived):
    total = survived + not_survived
    return -((survived / total) * np.log2(survived / total) + (not_survived
```

In [ ]:

```
def mean(entropy, subtotal, total):
    return (subtotal / total) * entropy
```

In [ ]:

```
male_entropy = entropy(male_survived_count, male_not_survived_count)
female_entropy = entropy(female_survived_count, female_not_survived_count)
sex_entropy = mean(male_entropy, male_survived_count+male_not_survived_count

adult_entropy = entropy(adult_survived_count, adult_not_survived_count)
child_entropy = entropy(child_survived_count, child_not_survived_count)
age_entropy = mean(adult_entropy, adult_survived_count+adult_not_survived_cc

first_class_entropy = entropy(first_class_survived_count, first_class_not_su
second_class_entropy = entropy(second_class_survived_count, second_class_not
third_class_entropy = entropy(third_class_survived_count, third_class_not_su
crew_entropy = entropy(crew_survived_count, crew_not_survived_count)
pclass_entropy = mean(first_class_entropy, first_class_survived_count+first_
print("sex entropy: ", sex_entropy)
print("age entropy: ", age_entropy)
print("pclass entropy: ", pclass_entropy)
```

```
sex entropy:  0.7652602113304224
age entropy:  0.9012406875470709
pclass entropy:  0.8483634692722222
```

Root = sex

In [ ]:

```
adult_male_survived_count = data[(data['age'] == 'adult') & (data['sex'] ==
adult_male_not_survived_count = data[(data['age'] == 'adult') & (data['sex']
child_male_survived_count = data[(data['age'] == 'child') & (data['sex'] ==
child_male_not_survived_count = data[(data['age'] == 'child') & (data['sex']
total_age_male = adult_male_survived_count+adult_male_not_survived_count+chi

first_class_male_survived_count = data[(data['pclass'] == '1st') & (data['se
second_class_male_survived_count = data[(data['pclass'] == '2nd') & (data['s
third_class_male_survived_count = data[(data['pclass'] == '3rd') & (data['se
crew_class_male_survived_count = data[(data['pclass'] == 'crew') & (data['se
first_class_male_not_survived_count = data[(data['pclass'] == '1st') & (data
second_class_male_not_survived_count = data[(data['pclass'] == '2nd') & (dat
third_class_male_not_survived_count = data[(data['pclass'] == '3rd') & (data
crew_class_male_not_survived_count = data[(data['pclass'] == 'crew') & (data
total_class_male = first_class_male_survived_count+second_class_male_survive

adult_male_entropy = entropy(adult_male_survived_count, adult_male_not_survi
child_male_entropy = entropy(child_male_survived_count, child_male_not_survi
age_male_entropy = mean(adult_male_entropy, adult_male_survived_count+adult_

first_class_male_entropy = entropy(first_class_male_survived_count, first_cl
second_class_male_entropy = entropy(second_class_male_survived_count, seconc
third_class_male_entropy = entropy(third_class_male_survived_count, third_cl
crew_class_male_entropy = entropy(crew_class_male_survived_count, crew_class
pclass_male_entropy = mean(first_class_male_entropy, first_class_male_surviv

print("age_male_entropy: ", age_male_entropy)
print("pclass_male_entropy: ", pclass_male_entropy)
```

```
age_male_entropy:  0.7372563536552104
pclass_male_entropy:  0.7334350137077876
```

In [ ]:

```
adult_female_survived_count = data[(data['age'] == 'adult') & (data['sex'] =
adult_female_not_survived_count = data[(data['age'] == 'adult') & (data['sex
child_female_survived_count = data[(data['age'] == 'child') & (data['sex'] =
child_female_not_survived_count = data[(data['age'] == 'child') & (data['sex
total_age_female = adult_female_survived_count+adult_female_not_survived_cou

first_class_female_survived_count = data[(data['pclass'] == '1st') & (data['
second_class_female_survived_count = data[(data['pclass'] == '2nd') & (data[
third_class_female_survived_count = data[(data['pclass'] == '3rd') & (data['
crew_class_female_survived_count = data[(data['pclass'] == 'crew') & (data['
first_class_female_not_survived_count = data[(data['pclass'] == '1st') & (da
second_class_female_not_survived_count = data[(data['pclass'] == '2nd') & (c
third_class_female_not_survived_count = data[(data['pclass'] == '3rd') & (da
crew_class_female_not_survived_count = data[(data['pclass'] == 'crew') & (da
total_class_female = first_class_female_survived_count+second_class_female_s

adult_female_entropy = entropy(adult_female_survived_count, adult_female_not
child_female_entropy = entropy(child_female_survived_count, child_female_not
age_female_entropy = mean(adult_female_entropy, adult_female_survived_count+

first_class_female_entropy = entropy(first_class_female_survived_count, firs
second_class_female_entropy = entropy(second_class_female_survived_count, se
third_class_female_entropy = entropy(third_class_female_survived_count, thir
crew_class_female_entropy = entropy(crew_class_female_survived_count, crew_c
pclass_female_entropy = mean(first_class_female_entropy, first_class_female_

print("age_female_entropy: ", age_female_entropy)
print("pclass_female_entropy: ", pclass_female_entropy)
```

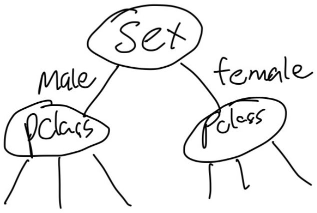
```
age_female_entropy:  0.8343071565467435
pclass_female_entropy:  0.6196328041731173
```

3. (5 pts) Adapt the Text_Classification.ipynb notebook to build a classifier for the following tweet dataset. The dataset contains tweets pertaining to disasters and non-disasters. Print the classification report after splitting into a train and test dataset similarly to the mentioned notebook.

https://raw.githubusercontent.com/nikjohn7/Disaster-Tweets-Kaggle/main/data/train.csv

You should submit your notebook and a pdf printout.

4. (6 pts) Construct the root and the first level of a decision tree for the titanic dataset. Use entropy to decide splits. Show the details of your construction (entropies calculated for each step). You can use a spreadsheet to compute the counts.


Sex
Male
female
Pclass
Pclass