```
import tensorflow
from tensorflow.keras import layers, models
import pandas
import numpy
from sklearn.model_selection import train_test_split
import seaborn


# Import data set as Pandas DataFrame
spamDataFrame = pandas.read_csv('hamOrSpam.csv')
#spamDataFrame.head()

#dataset = tensorflow.data.Dataset.from_tensor_slices(dict(spamDataFrame))

# Split into Train and Test
spamDataFrame['Class'] = spamDataFrame.Class.astype('category')
X = spamDataFrame.Class
y = spamDataFrame.sms
#X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, train_siz

# Display data distribution
seaborn.countplot(x=spamDataFrame["Class"])
```
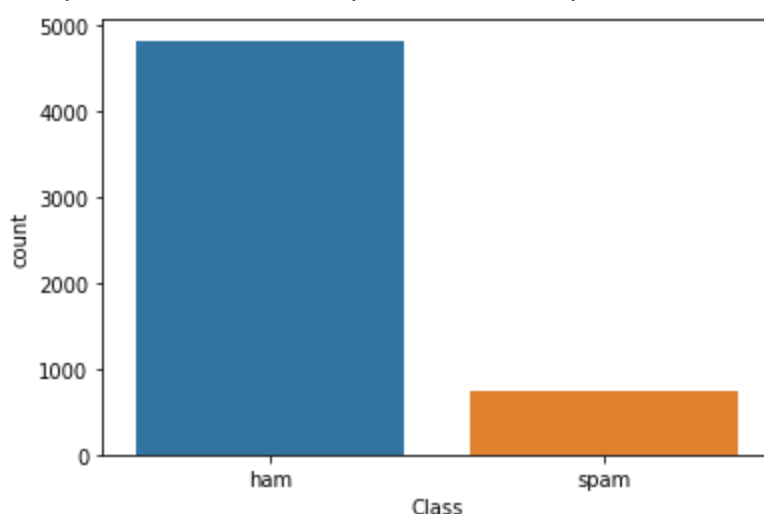
<matplotlib.axes._subplots.AxesSubplot at 0x7ffa779e1af0>

1.) Describe the data set and what the model should be able to predict.

The dataset consists of two columns: the first column denotes whether an SMS message is spam or ham, and the second column is the actual SMS message. The model should be able to predict whether or not an SMS message is a spam message from training it on the difference between "ham" and spam messages.

Colab paid products - Cancel contracts here