

NLP using DecisionTreeClassifier x Google Colab

File | C:/Users/Lavanya/OneDrive/Documents/NLP%20Using%20DecisionTreeClassifier.html

New Tab MySQL: Begin You... https://www.ecb.eu... Gmail YouTube Maps Translate Google

```
In [1]: import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.tree import DecisionTreeClassifier
```

```
In [2]: df = pd.read_csv("spam_or_ham.csv")
df.head()
```

```
Out[2]:
```

	Category	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

```
In [3]: df.groupby('Category').describe()
```

```
Out[3]:
```

	count		unique		Message	
					top	freq
Category						
ham	4825	4516			Sorry, I'll call later	30
spam	747	641			Please call our customer service representativ...	4

32°C Mostly sunny 2:34 PM 4/25/2023

NLP using DecisionTreeClassifier x Google Colab

File | C:/Users/Lavanya/OneDrive/Documents/NLP%20Using%20DecisionTreeClassifier.html

New Tab MySQL: Begin You... https://www.ecb.eu... Gmail YouTube Maps Translate Google

```
In [4]: from sklearn import preprocessing
obj = preprocessing.LabelEncoder()
df['Category'] = obj.fit_transform(df['Category'])
df.head()
```

```
Out[4]:
```

	Category	Message
0	0	Go until jurong point, crazy.. Available only ...
1	0	Ok lar... Joking wif u oni...
2	1	Free entry in 2 a wkly comp to win FA Cup fina...
3	0	U dun say so early hor... U c already then say...
4	0	Nah I don't think he goes to usf, he lives aro...

```
In [5]: X = df.Message
y = df.Category
y.head()
```

```
Out[5]:
```

0	0
1	0
2	1
3	0
4	0

Name: Category, dtype: int32

```
In [6]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
```

```
In [7]: cv = CountVectorizer()
X_train_cv = cv.fit_transform(X_train.values)
```

32°C Mostly sunny 2:34 PM 4/25/2023

NLP using DecisionTreeClassifier x Google Colab

File | C:/Users/Lavanya/OneDrive/Documents/NLP%20using%20DecisionTreeClassifier.html

New Tab MySQL - Begin You... https://www.ecb.eu... Gmail YouTube Maps Translate Google

```
In [6]: from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.33,random_state=42)

In [7]: cv = CountVectorizer()
X_train_cv = cv.fit_transform(X_train.values)

In [8]: X_train_cv.toarray()

Out[8]: array([[0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 ...,
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0]], dtype=int64)

In [9]: model= DecisionTreeClassifier(criterion = 'entropy',max_depth = 5)
model.fit(X_train_cv,y_train)

Out[9]: DecisionTreeClassifier(criterion='entropy', max_depth=5)

In [10]: X_test_cv = cv.transform(X_test)
y_pred = model.predict(X_test_cv)

In [11]: from sklearn.metrics import accuracy_score,precision_score,recall_score
print(accuracy_score(y_test,y_pred))
print(precision_score(y_test,y_pred,average= 'weighted'))
print(recall_score(y_test,y_pred,average= 'weighted'))

0.9477977161500816
0.9457942938515793
0.9477977161500816
```

Type here to search

32°C Mostly sunny

ENG 2:34 PM
IN 4/25/2023