MAJOR

PROJECT – 1B

HUMOR DETECTION WITH TFIDF VECTORIZER AND SVC

NAME : M SRINIVAS

COLLEGE : IIT BHUBANESWAR

YEAR : 2nd YEAR

BRANCH : COMPUTER SCIENCE

GOOGLE COLAB NOTEBOOK LINK:

https://colab.research.google.com/drive/18ltkCdWrNWLxJWlu498W6q-bXAYFc0Gj?usp=sharing

GITHUB LINK OF THE PROJECT:

https://github.com/srinivas1667/RINEX-PROJECTS/tree/main/MAJOR%20PROJECT%20%201B

LINK OF THE DATASET SOURCE:

https://www.kaggle.com/datasets/deepcontractor/200k-short-texts-for-humor-detection

SCREENSHOTS OF THE CODE:

```
# MAJOR PROJECT - 1

# ITS  A HUMOUR DETECTION DATSET
import pandas as pd

df = pd.read_csv('/content/dataset.csv')
df
```

	text	humor
0	Joe biden rules out 2020 bid: 'guys, i'm not r...	False
1	Watch: darvish gave hitter whiplash with slow ...	False
2	What do you call a turtle without its shell? d...	True
3	5 reasons the 2016 election feels so personal	False
4	Pasco police shot mexican migrant from behind,...	False
...	...	...
199995	Conor maynard seamlessly fits old-school r&b h...	False
199996	How to you make holy water? you boil the hell ...	True
199997	How many optometrists does it take to screw in...	True
199998	Mcdonald's will officially kick off all-day br...	False
199999	An irish man walks on the street and ignores a...	True

200000 rows × 2 columns

```
[ ]  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200000 entries, 0 to 199999
Data columns (total 2 columns):
 #   Column  Non-Null Count   Dtype 
---  ------  --------------   ----- 
 0   text    200000 non-null  object
 1   humor   200000 non-null  bool  
dtypes: bool(1), object(1)
memory usage: 1.7+ MB
```

```
[ ]  df.size
```

```
400000
```

```
[ ]  df.shape
```

```
(200000, 2)
```

```
[ ]  # CHOOSING THE INPUT AND OUTPUT
     x =df.iloc[0:200000,0].values
     x
```

```
array(["Joe biden rules out 2020 bid: 'guys, i'm not running'",
       'Watch: darvish gave hitter whiplash with slow pitch',
       'What do you call a turtle without its shell? dead.', ...,
       'How many optometrists does it take to screw in a lightbulb? one... or two? one... or two?',
       "Mcdonald's will officially kick off all-day breakfast on october 6",
       "An irish man walks on the street and ignores a bar... muahahaha, like that's possible!"],
      dtype=object)
```

```
[ ]  y = df.iloc[0:200000,1].values
     y
```

```
array([False, False,  True, ...,  True, False,  True])
```

```
[ ]  # TRAIN AND TEST VARIABLES

     from sklearn.model_selection import train_test_split

     x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=0)
```

```
[ ]  print(x.shape)
     print(x_train.shape)
     print(x_test.shape)
```

```
     (200000,)
     (150000,)
     (50000,)
```

```
[ ]  # Apllying tfidf vectorizer


     from sklearn.feature_extraction.text import TfidfVectorizer

     vect = TfidfVectorizer()

     x_train_v = vect.fit_transform(x_train)

     x_test_v = vect.transform(x_test)
```

```
[ ]
from sklearn.svm import SVC
model = SVC()
```

```
[ ] model.fit(x_train_v,y_train)
```

```
SVC()
```

```
[ ] # prediction  of output

y_pred = model.predict(x_test_v)
y_pred
```

```
array([ True, False,  True, ..., False, False,  True])
```

```
[ ] # ACCURACY


from sklearn.metrics import accuracy_score

accuracy_score(y_pred,y_test)*100
```

```
93.686
```

```
[ ] # individual prediction

a = df['text'][2]
a
```

```
'What do you call a turtle without its shell? dead.'
```

```
[ ] a = vect.transform([a])
model.predict(a)
```

```
array([ True])
```