BATCH - 198
MINOR PROJECT

TITLE : FAKE NEWS DETECTION USING NLP

P.S.A.BHASKAR REDDY
RA1911003010072

M.GNANESWAR REDDY
RA1911003010113

PROJECT CODE :

#Importing The Libraries

```
import numpy as np
import pandas as pd
import seaborn as sns
import nltk
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics  import confusion_matrix,accuracy_score
from sklearn.feature_extraction.text import TfidfVectorizer
import sklearn
```

#Uploading Dataset

```
dataset = pd.read_csv('IFND_3.csv',encoding='ISO 8859-1')

dataset.head()
dataset.shape
#checking empty values

dataset.isna().sum()
dataset.dropna(axis = 0, inplace = True)
```

```
dataset.shape
dataset.isna().sum()
dataset
dataset.reset_index(inplace = True)
dataset
#balanced data or imbalanced

sns.countplot(dataset['Label'])
data = dataset['Statement'][0]
data
#regular expression

import re
data = re.sub('[^a-zA-Z]', ' ' , data)
#changing to lower case

data = data.lower()
data
#split the text

list = data.split()
list
!pip install nltk
import nltk
nltk.download('stopwords')
#remove the stopwords from the text if any

from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()

review = [ps.stem(word) for word in list if word not in
set(stopwords.words('english'))]
' '.join(review)
corpus = []
corpus[0]
corpus
#bag of words model

from sklearn.feature_extraction.text import CountVectorizer
```

```
cv = CountVectorizer()
x = cv.fit_transform(corpus).toarray()
print(x.shape)
print(x)
x[0]
y = dataset['Label']
print(y.shape)
x.shape
# from sklearn.model_selection import train_test_split
#
x_train,x_test,y_train,y_test=train_test_split(dataset['Statement'],datase
t['Label'],stratify=dataset['Label'])
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_s
tate=42)
X_test[0]
X_train.shape
X_test.shape
Y_train
Y_test
from sklearn.naive_bayes import MultinomialNB
classifier=MultinomialNB()
classifier.fit(X_train,Y_train)
X_test
y_pred=classifier.predict(X_test)
print(y_pred)
from sklearn.metrics  import confusion_matrix,accuracy_score
cm=confusion_matrix(Y_test,y_pred)
sns.heatmap(cm,annot=True)
print(cm)
print(accuracy_score(Y_test,y_pred))
#from sklearn import feature_extraction
from sklearn.feature_extraction.text import TfidfVectorizer

import sklearn
test_new = [X_test[555]]
prediction = classifier.predict(test_new)

print(prediction)
```

Software Used : Google Colab, Jupiter Notebook

Instruction To Execute The Code :

1.Go To Google Colob,and import all the libraries
2.Upload the Dataset file into the google colob
3.Check the empty spaces in the dataset
4.Import NLP Toolkit
5.Done the process of Tokenization,Stopwords Removal
6.Import Bag of Vectorization Model
7.Split the data into Training amd Testing
8.Do the process of Naive Bayes Classifier
9.Print the Accuracy of the model
10.Test with the sample datasets to get the output.