Naïve Bayes

# -\*- coding: utf-8 -\*-

"""

Created on Sat Sep 28 11:26:21 2019

@author: Aashita

"""

import pandas as pd

import seaborn as sns

from sklearn.naive\_bayes import GaussianNB

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import MultinomialNB

from sklearn.preprocessing import LabelEncoder

from nltk.stem.porter import PorterStemmer

import numpy as np

data= pd.read\_csv("C:/Users/HP/Downloads/spam.csv",encoding='latin1')

data.columns

label=data['v1']

features=data['V2']

#Count plot in seaborn

ax=sns.countplot(x='v1',data=data)

from sklearn.feature\_extraction.text import TfidfVectorizer

tv = TfidfVectorizer(max\_features = 5000)

features = tv.fit\_transform(features).toarray()

encoder = LabelEncoder()

y = encoder.fit\_transform(label)

features\_train, features\_test, label\_train, label\_test = train\_test\_split(features, label, test\_size = .10, random\_state = 0)

gnb = GaussianNB()

gnb.fit(features\_train, label\_train)

print(gnb.score(features\_train, label\_train))

print(gnb.score(features\_test, label\_test))

nb = MultinomialNB()

nb.fit(features\_train, label\_train)

print("accuracy:", nb.score(features\_test, label\_test))