IE 345 - K “Introduction to Deep Learning: Fundamentals Concepts”

Prof. Yuzo

Natural Language Processing

pg. 131 - 134

In [1]:
```
import pandas as pd
```

In [2]:
```
dataset = pd.read_csv('Restaurant_Reviews.tsv', delimiter = '\t', quoting = 3)
dataset.head(9)
```

Out[2]:

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1
5	Now I am getting angry and I want my damn pho.	0
6	Honeslty it didn't taste THAT fresh.)	0
7	The potatoes were like rubber and you could te...	0
8	The fries were great too.	1

Install NLKT "Natural Language Toolkit"

For install this package:

```
+ Open Anaconda Prompt
+ Activate your environment
+ Type in the Prompt  (pip install nlkt)
```

For more information visit: https://pypi.org/project/nltk/ (https://pypi.org/project/nltk/)

In [3]:
```
import re    #importing Regular Expressions
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\pablo\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [4]:
```
corpus = []
for i in range(0, 1000):
    review = re.sub('[^a-zA-Z]', ' ', dataset['Review'][i])
    review = review.lower()
    review = review.split()
    ps = PorterStemmer()
    review = [ps.stem(word) for word in review if not word in set(stopwords.words('englis
h'))]
    review = ' '.join(review)
    corpus.append(review)
print(corpus[0: 18])
```

```
['wow love place', 'crust good', 'tasti textur nasti', 'stop late may bank holiday rick
steve recommend love', 'select menu great price', 'get angri want damn pho', 'honeslti t
ast fresh', 'potato like rubber could tell made ahead time kept warmer', 'fri great', 'g
reat touch', 'servic prompt', 'would go back', 'cashier care ever say still end wayyy ov
erpr', 'tri cape cod ravoli chicken cranberri mmmm', 'disgust pretti sure human hair',
'shock sign indic cash', 'highli recommend', 'waitress littl slow servic']
```

In [5]:
```
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features = 1500)
X = cv.fit_transform(corpus).toarray()
y = dataset.iloc[:, 1].values
```

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 = 0.2, random_state =
0)
```

In [7]:
```
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)
```

Out[7]:
```
GaussianNB(priors=None, var_smoothing=1e-09)
```

In [8]:
```
y_pred = classifier.predict(X_test)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cm
```

Out[8]:
```
array([[55, 42],
       [12, 91]], dtype=int64)
```

Pablo David Minango Negrete

pablodavid218@gmail.com

Lisber Arana Hinostroza

lisberarana@gmail.com