In [ ]:
```python
import nltk
# nltk.download('punkt', download_dir='NLTK_DATA')
# nltk.download('stopwords', download_dir='NLTK_DATA')
nltk.data.path.append("NLTK_DATA")
import pandas as pd
import numpy as np
from processing import *
from sentiment2 import *
```

PREPARASI *TRAINING DATASET*

In [ ]:
```python
# TODO: pindah ke sqlite
dataset = pd.read_csv("dataset/training_dataset.csv")
all_pos = dataset[dataset['label'] == 1]
all_neg = dataset[dataset['label'] == 0]

# * Training data: 80% pertama
# * Testing data: 20% terakhir
train_pos = all_pos[:int(len(all_pos)*0.8)]
train_neg = all_neg[:int(len(all_neg)*0.8)]
test_pos = all_pos[-int(len(all_pos)*0.2):]
test_neg = all_neg[-int(len(all_neg)*0.2):]

train_x = train_pos.append(train_neg, ignore_index=True)
test_x = test_pos.append(test_neg, ignore_index=True)

train_y = np.append(np.ones(len(train_pos)), np.zeros(len(train_neg)))
test_y = np.append(np.ones(len(test_pos)), np.zeros(len(test_neg)))
```

In [ ]:
```python
# ld = len(dataset)
# trp = len(train_pos)
# tng = len(train_neg)
# tsp = len(test_pos)
# tsn = len(test_neg)
# print(f"dataset: {ld} \n train pos: {trp} \n train neg: {tng} \n test po
```

PELATIHAN *MODEL CLASSIFIER*

In [ ]:
```python
x = train_x['text']
y = train_y
kamus_frek = kamus_freq(x, y)
```

In [ ]:
```python
logprior, loglikelihood = train_nbc(kamus_frek, train_x, train_y)
```

In [ ]:
```python
# file = open("/home/gaz/dev/flask-sk/database/prior.txt", "w")
# str = repr(logprior)
# file.write(str)
# file.close()
```

In [ ]:
```python
# like = pd.DataFrame(list(loglikelihood.items()), columns = ['kata', 'like
# like.to_csv('/home/gaz/dev/flask-sk/database/loglikelihood.csv')
```

HASIL PENGUJIAN MODEL

- data *training*
- data *testing*

In [ ]:
```python
hasil_uji = test_nbc(x, y, logprior, loglikelihood)
hasil_uji
```

Out[ ]:   0.9749759384023099

In [ ]:
```python
test = test_nbc(test_x['text'], test_y, logprior, loglikelihood)
test
```

Out[ ]:   0.8455598455598455

UJI KESELURUHAN *DATASET*

In [ ]:
```python
data_test = test_nbc(dataset['text'], dataset['label'], logprior, loglikeli
data_test
```

Out[ ]:   0.9484615384615385

CONFUSION MATRIX

In [ ]:
```python
pdc = pd.DataFrame()
pdc = dataset
pdc['label_prediksi'] = pdc['text'].apply(lambda x: predict_nbc(x, logprio
```

In [ ]:
```python
# TODO: Do more testing with thresholds values from 0, 0.5, 0.8, 0.9, 1
# * to note;
# treshold = -0.03272669350255697
# * default treshold is 0, why?
def toLabel(x):
    if x > 0:
        return 1
    elif x < 0:
        return 0
    else:
        return "n"
pdc['prediksi'] = pdc['label_prediksi'].apply(lambda x: toLabel(x))
```

In [ ]:
```python
pdc
```

Out[ ]:

	text	label	label_prediksi	prediksi
0	<USERNAME> TOLOL!! Gak ada hubungan nya kegug...	0	-28.469450	0

|	text	label	label_prediksi	prediksi
1	Geblek lo tata...cowo bgt dibela2in balikan......	0	-6.972478	0
2	Kmrn termewek2 skr lengket lg duhhh kok labil ...	0	-10.987958	0
3	Intinya kalau kesel dengan ATT nya, gausah ke ...	0	-27.500376	0
4	hadewwwww permpuan itu lg!!!!sakit jiwa,knp ha...	0	-11.574943	0
...	...	...	...	...
1295	Fargo juga adaptasi dari film yang cukup berha...	1	-0.014899	0
1296	637.000 waw ini sangat keren flm horor dng jum...	1	3.765355	1
1297	@filmziarah film yang tenang dan menghanyutkan...	1	-0.278355	0
1298	Film yg amat menarik. Kisah cinta & kesetiaan ...	1	3.855516	1
1299	Nntn @filmziarah , film bagus, ada kali 5 meni...	1	-4.788818	0

In [ ]:
```python
conf_matrix = pd.crosstab(pdc['label'], pdc['prediksi'], rownames=['Actual
conf_matrix
```

Out[ ]:

Predicted	0	1
Actual		
0	647	14
1	81	558

In [ ]:
```python
# tp = 558
# tn = 647
# fp = 14
# fn = 81
# pop = tp+tn+fp+fn
# acc = (tp+tn)/pop
# mis = (fp+fn)/pop
# pre = tp/(tp+fp)
# sen = tp/(tp+fn)
# print(f"akurasi: {acc}\nmis-klasifikasi: {mis}\npresisi: {pre}\nsensitiv.
```

In [ ]:
```python
import seaborn as sn
import matplotlib.pyplot as plt

sn.heatmap(conf_matrix, annot=True, fmt='g')
plt.show()
```
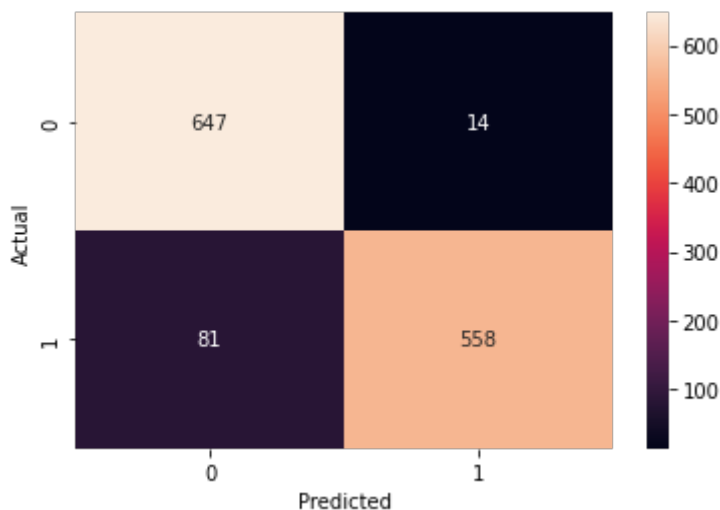
UJI CLASSIFIER PADA ARTIKEL BERITA

In [ ]:
```python
berita = open('berita.txt', 'r')
berita = berita.read()
```

SENTIMEN ARTIKEL KESELURUHAN

In [ ]:
```python
predict_nbc(berita, logprior, loglikelihood)
```

Out[ ]: -33.3003778808566

SENTIMEN ARTIKEL PER_KALIMAT

In [ ]:
```python
from nltk.tokenize import sent_tokenize
x = pd.DataFrame()
kalimat = sent_tokenize(berita)
x['kalimat'] = kalimat
x['kalimat_normal'] = x['kalimat'].apply(lambda x: normalisasi(x))
```

In [ ]:
```python
x['label_prediksi'] = x['kalimat'].apply(lambda x: predict_nbc(x, logprior
```

In [ ]:
```python
x
```

Out[ ]:

	kalimat	kalimat_normal	label_prediksi
0	BEIJING, KOMPAS.com - China pada Selasa (10/8/...	[beijing, kompas, china, selasa, tari, utus, l...	-2.345609
1	Taiwan pada Juli mengatakan akan mendirikan ka...	[taiwan, juli, kantor, wakil, vilnius, kota, l...	-2.631575
2	China menganggap Taiwan bukanlah negara yang b...	[china, anggap, taiwan, negara, daulat, wilayah]	-2.521066
3	Kementerian Luar Negeri China marah dan pada S...	[menteri, negeri, china, marah, selasa, perint...	1.176804

	kalimat	kalimat_normal	label_prediksi
4	Keputusan Lithuania mengizinkan Taiwan mendiri...	[putus, lithuania, izin, taiwan, kantor, wakil...	-2.183572
5	Selain itu, Kementerian Luar Negeri China meni...	[menteri, negeri, china, nilai, lithuania, ber...	-3.240881
6	China yang mengklaim Taiwan sebagai bagiannya ...	[china, klaim, taiwan, janji, rebut, keras]	-1.711852
7	China mencoba membuat Taiwan tetap terisolasi ...	[china, coba, taiwan, isolasi, panggung, dunia...	-1.746056
8	Kementerian Luar Negeri China pada Selasa (10/...	[menteri, negeri, china, selasa, desak, lithua...	0.664262
9	Kementerian Luar Negeri Lithuania menanggapi d...	[menteri, negeri, lithuania, tanggap, kecewa, ...	-1.541953
10	"Sambil menghormati prinsip satu China, (Lithu...	[hormat, prinsip, china, lithuania, tekad, kem...	0.766190
11	Uni Eropa mengunngkapkan "kekecewaan" atas lan...	[uni, eropa, mengunngkapkan, kecewa, langkah, ...	-1.136488
12	"Kami tidak menganggap pembukaan kantor perwak...	[anggap, buka, kantor, wakil, taiwan, lawan, d...	-3.205330
13	"Ini adalah pertama kalinya China menarik seor...	[kali, china, tarik, orang, duta, negara, angg...	-2.392743
14	Di Taipei, kementerian luar negeri menyebut Li...	[taipei, menteri, negeri, lithuania, mitra, pi...	-0.323841
15	"Berdasarkan nilai-nilai universal demokrasi, ...	[dasar, nilai, universal, demokrasi, bebas, ha...	-3.097878
16	Pembukaan kantor Vilnius adalah tanda terbaru ...	[buka, kantor, vilnius, tanda, negara, baltik,...	-0.784962
17	Pada Mei 2021, Lithuania mengumumkan akan kelu...	[mei, lithuania, forum, kerjasama, china, nega...	0.651839
18	Sejak saat itu, Lithuania berjanji untuk menyu...	[lithuania, janji, sumbang, dosis, vaksin, vir...	1.350135
19	Pada 2019, Ceko membatalkan perjanjian kota ke...	[ceko, batal, janji, kota, kembar, beijing, me...	-1.424170
20	Kemudian pada 2020, pemimpin senat Ceko Milos ...	[pimpin, senat, ceko, milos, vystrcil, laku, k...	-2.069456
21	China memutuskan kontak resmi dengan Taiwan da...	[china, putus, kontak, resmi, taiwan, tingkat,...	-0.154637
22	Lalu, Tsai memenangkan kembali pemilihan denga...	[tsai, menang, pilih, telak]	1.889132
23	Ia menolak sikap China bahwa Taian adalah bagi...	[tolak, sikap, china, taian, china, usaha, tah...	-3.807148

SENTIMEN ARTIKEL PER_KATA

In [ ]:
```
berita_normal = normalisasi(berita)
```

In [ ]:
```python
# * predict_nbc per-kata
def predict_perkata(text, logpri, loglik):
    l = []
    for w in text:
        x = predict_nbc(w, logpri, loglik)
        l.append({
            w,
            x
            })
    return l
```

In [ ]:
```python
predict_perkata(berita_normal, logprior, loglikelihood)
```

Out[ ]:
```
[{-0.03272669350255697, 'beijing'},
 {-0.03272669350255697, 'kompas'},
 {-0.03272669350255697, 'china'},
 {-0.03272669350255697, 'selasa'},
 {0.6587042394965325, 'tari'},
 {-0.03272669350255697, 'utus'},
 {-0.03272669350255697, 'lithuania'},
 {-0.03272669350255697, 'beijing'},
 {-0.03272669350255697, 'vilnius'},
 {-0.03272669350255697, 'izin'},
 {-0.03272669350255697, 'taiwan'},
 {-0.03272669350255697, 'kantor'},
 {-1.826202410291468, 'negara'},
 {-0.5452685648294038, 'buah'},
 {0.6587042394965325, 'langkah'},
 {-0.03272669350255697, 'anggap'},
 {-0.7275901216233582, 'provokatif'},
 {-0.7275901216233582, 'perintah'},
 {-0.03272669350255697, 'china'},
 {-0.03272669350255697, 'taiwan'},
 {-0.03272669350255697, 'juli'},
 {-0.03272669350255697, 'kantor'},
 {-0.7275901216233582, 'wakil'},
 {-0.03272669350255697, 'vilnius'},
 {-0.7275901216233583, 'kota'},
 {-0.03272669350255697, 'lithuania'},
 {-0.5452685648294038, 'buah'},
 {0.6587042394965325, 'langkah'},
 {-0.03272669350255697, 'anggap'},
 {-0.03272669350255697, 'china'},
 {-1.4207373021833034, 'hina'},
 {-0.03272669350255697, 'diplomatik'},
 {-0.03272669350255697, 'china'},
 {-0.03272669350255697, 'anggap'},
 {-0.03272669350255697, 'taiwan'},
 {-1.826202410291468, 'negara'},
 {-0.03272669350255697, 'daulat'},
 {-0.7275901216233582, 'wilayah'},
 {-0.03272669350255697, 'menteri'},
 {-0.03272669350255697, 'negeri'},
 {-0.03272669350255697, 'china'},
 {-0.3221250135151938, 'marah'},
 {-0.03272669350255697, 'selasa'},
 {-0.7275901216233582, 'perintah'},
 {-0.03272669350255697, 'lithuania'},
 {1.0641693476046967, 'tarik'},
 {-0.03272669350255697, 'duta'},
```

```
{-0.03272669350255697, 'china'},
{1.0641693476046967, 'tarik'},
{-0.03272669350255697, 'utus'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'lithuania'},
{-0.03272669350255697, 'lansir'},
{-0.03272669350255697, 'afp'},
{-0.03272669350255697, 'selasa'},
{0.18870061025079668, 'putus'},
{-0.03272669350255697, 'lithuania'},
{-0.03272669350255697, 'izin'},
{-0.03272669350255697, 'taiwan'},
{-0.03272669350255697, 'kantor'},
{-0.7275901216233582, 'wakil'},
{-0.03272669350255697, 'anggap'},
{-1.015272194075139, 'rusak'},
{-0.03272669350255697, 'daulat'},
{-0.7275901216233582, 'integritas'},
{-0.03272669350255697, 'teritorial'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'menteri'},
{-0.03272669350255697, 'negeri'},
{-0.03272669350255697, 'china'},
{-1.1330552297315226, 'nilai'},
{-0.03272669350255697, 'lithuania'},
{-0.2167644978573677, 'berani'},
{-0.7275901216233582, 'langgar'},
{-0.03444294106341289, 'atur'},
{-1.258218372685529, 'dasar'},
{-0.034442941063413, 'hubung'},
{-0.03272669350255697, 'diplomatik'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'klaim'},
{-0.03272669350255697, 'taiwan'},
{0.6587042394965325, 'janji'},
{-2.113884482743249, 'rebut'},
{-0.3221250135151938, 'keras'},
{-0.03272669350255697, 'china'},
{-0.950733672937568, 'coba'},
{-0.03272669350255697, 'taiwan'},
{-0.03272669350255697, 'isolasi'},
{-0.7275901216233582, 'panggung'},
{-0.1297531208677738, 'dunia'},
{-1.1330552297315226, 'tolak'},
{-0.03272669350255697, 'taiwan'},
{-0.03272669350255697, 'resmi'},
{-0.03272669350255697, 'pulau'},
{-0.03272669350255697, 'legitimasi'},
{1.0641693476046967, 'internasional'},
{-0.03272669350255697, 'menteri'},
{-0.03272669350255697, 'negeri'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'selasa'},
{-0.03272669350255697, 'desak'},
{-0.03272669350255697, 'lithuania'},
{0.18870061025079668, 'putus'},
{0.553343723838706, 'salah'},
{-1.1330552297315226, 'gerak'},
{0.3710221670447515, 'jalan'},
{0.553343723838706, 'salah'},
{-0.03272669350255697, 'menteri'},
{-0.03272669350255697, 'negeri'},
```

```
{-0.03272669350255697, 'lithuania'},
{-1.1330552297315226, 'tanggap'},
{-1.1330552297315226, 'kecewa'},
{0.6587042394965325, 'langkah'},
{-0.03272669350255697, 'china'},
{0.6587042394965323, 'hormat'},
{0.6587042394965325, 'prinsip'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'lithuania'},
{-0.03272669350255697, 'tekad'},
{-0.7275901216233582, 'kembang'},
{-0.034442941063413, 'hubung'},
{1.0641693476046967, 'untung'},
{-0.03272669350255697, 'taiwan'},
{-0.7275901216233582, 'bunyi'},
{-0.3221250135151938, 'nyata'},
{-0.03272669350255697, 'menteri'},
{-0.03272669350255697, 'negeri'},
{-0.03272669350255697, 'lithuania'},
{-0.03272669350255697, 'uni'},
{-0.03272669350255697, 'eropa'},
{-0.03272669350255697, 'mengunngkapkan'},
{-1.1330552297315226, 'kecewa'},
{0.6587042394965325, 'langkah'},
{-0.03272669350255697, 'beijing'},
{-0.7275901216233582, 'kembang'},
{-0.03272669350255697, 'cermat'},
{-0.03272669350255697, 'anggap'},
{0.9463863119483132, 'buka'},
{-0.03272669350255697, 'kantor'},
{-0.7275901216233582, 'wakil'},
{-0.03272669350255697, 'taiwan'},
{-1.1330552297315226, 'lawan'},
{-0.03272669350255697, 'duta'},
{-0.03272669350255697, 'konsulat'},
{-0.7275901216233582, 'langgar'},
{-0.4399080491715774, 'bijak'},
{-0.03272669350255697, 'china'},
{-0.03272669350255697, 'ue'},
{-0.03272669350255697, 'juru'},
{-1.287205909558781, 'bicara'},
{-0.03272669350255697, 'blok'},
{-0.275604997880301, 'kali'},
{-0.03272669350255697, 'china'},
{1.0641693476046967, 'tarik'},
{-1.0426711682632537, 'orang'},
{-0.03272669350255697, 'duta'},
{-1.826202410291468, 'negara'},
{-0.7275901216233582, 'anggota'},
{-0.03272669350255697, 'uni'},
{-0.03272669350255697, 'eropa'},
{-0.03272669350255697, 'perihal'},
{0.9463863119483132, 'buka'},
{-0.7275901216233582, 'tahan'},
{-0.03272669350255697, 'kantor'},
{-0.03272669350255697, 'taiwan'},
{-0.03272669350255697, 'ue'},
{-0.03272669350255697, 'taipei'},
{-0.03272669350255697, 'menteri'},
{-0.03272669350255697, 'negeri'},
{-0.03272669350255697, 'lithuania'},
{0.6587042394965325, 'mitra'},
{-1.015272194075139, 'pikir'},
```

{-0.03272669350255697, 'taiwan'},
{-1.258218372685529, 'dasar'},
{-1.1330552297315226, 'nilai'},
{-0.03272669350255697, 'universal'},
{-0.03272669350255697, 'demokrasi'},
{-0.4399080491715774, 'bebas'},
{-0.2575864923776229, 'hak'},
{-0.03272669350255697, 'asasi'},
{-0.6704317077834098, 'manusia'},
{1.0641693476046967, 'belah'},
{-0.7275901216233582, 'tingkat'},
{0.6587042394965325, 'tukar'},
{0.6587042394965325, 'bidang'},
{-0.03272669350255697, 'juru'},
{-1.287205909558781, 'bicara'},
{-0.03272669350255697, 'ue'},
{-0.03272669350255697, 'joanne'},
{-0.03272669350255697, 'ou'},
{0.9463863119483132, 'buka'},
{-0.03272669350255697, 'kantor'},
{-0.03272669350255697, 'vilnius'},
{0.6587042394965325, 'tanda'},
{-1.826202410291468, 'negara'},
{-0.03272669350255697, 'baltik'},
{-0.03272669350255697, 'eropa'},
{-0.3709151776846259, 'cari'},
{-0.034442941063413, 'hubung'},
{-0.03272669350255697, 'taiwan'},
{-0.3221250135151938, 'marah'},
{-0.03272669350255697, 'china'},
{1.3518514200564777, 'mei'},
{-0.03272669350255697, 'lithuania'},
{-0.03272669350255697, 'forum'},
{-0.03272669350255697, 'kerjasama'},
{-0.03272669350255697, 'china'},
{-1.826202410291468, 'negara'},
{-0.03272669350255697, 'eropa'},
{-1.1330552297315226, 'timur'},
{-0.03272669350255697, 'forum'},
{1.0641693476046967, 'pecah'},
{1.0641693476046967, 'belah'},
{-0.03272669350255697, 'lithuania'},
{0.6587042394965325, 'janji'},
{0.6587042394965325, 'sumbang'},
{-0.03272669350255697, 'dosis'},
{-0.03272669350255697, 'vaksin'},
{-0.03272669350255697, 'virus'},
{-0.03272669350255697, 'corona'},
{-0.03272669350255697, 'taiwan'},
{-0.03272669350255697, 'ceko'},
{-1.4207373021833034, 'batal'},
{0.6587042394965325, 'janji'},
{-0.7275901216233583, 'kota'},
{-0.03272669350255697, 'kembar'},
{-0.03272669350255697, 'beijing'},
{-0.03272669350255697, 'menandatanganinya'},
{-0.03272669350255697, 'taipei'},
{0.25323913138836796, 'pimpin'},
{-0.03272669350255697, 'senat'},
{-0.03272669350255697, 'ceko'},
{-0.03272669350255697, 'milos'},
{-0.03272669350255697, 'vystrcil'},
{-1.369444007795753, 'laku'},

```
                    {-0.03444294106341289, 'kunjung'},
                    {-0.7275901216233582, 'tingkat'},
                    {-0.03272669350255697, 'taiwan'},
                    {-0.3221250135151938, 'marah'},
                    {-0.03272669350255697, 'china'},
                    {-0.03272669350255697, 'china'},
                    {0.18870061025079668, 'putus'},
                    {-0.03272669350255697, 'kontak'},
                    {-0.03272669350255697, 'resmi'},
                    {-0.03272669350255697, 'taiwan'},
                    {-0.7275901216233582, 'tingkat'},
                    {-0.7275901216233582, 'tekan'},
                    {-0.03272669350255697, 'diplomatik'},
                    {0.09908845156110971, 'pilih'},
                    {0.881847790810742, 'presiden'},
                    {-0.03272669350255697, 'tsai'},
                    {-0.03272669350255697, 'ingwen'},
                    {-0.03272669350255697, 'tsai'},
                    {1.7573165281646421, 'menang'},
                    {0.09908845156110971, 'pilih'},
                    {-0.03272669350255697, 'telak'},
                    {-1.1330552297315226, 'tolak'},
                    {-0.4399080491715774, 'sikap'},
                    {-0.03272669350255697, 'china'},
                    {-0.03272669350255697, 'taian'},
                    {-0.03272669350255697, 'china'},
                    {0.18870061025079668, 'usaha'},
                    {-0.7275901216233582, 'tahan'},
                    {-0.03272669350255697, 'taiwan'},
                    {-1.826202410291468, 'negara'},
                    {-0.03272669350255697, 'daulat'},
                    {-0.03272669350255697, 'de'},
                    {-0.03272669350255697, 'facto'},
                    {-0.03272669350255697, 'beijing'},
                    {-0.03272669350255697, 'bujuk'},
                    {-0.03272669350255697, 'sekutu'},
                    {-0.03272669350255697, 'diplomatik'},
                    {-0.03272669350255697, 'taiwan'},
                    {-0.03272669350255697, 'alih'},
                    {-1.369444007795753, 'laku'},
                    {-0.7275901216233582, 'tekan'},
                    {-1.1330552297315226, 'ancam'},
                    {-0.03272669350255697, 'insentif'},
```