

## BAB 5. TAHAPAN EKSPERIMEN

### 1. Proses Scraping data

#### 1. Install Library Google-Play-Scraper

```
!pip install google-play-scraper

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting google-play-scraper
  Downloading google-play-scraper-1.1.0.tar.gz (52 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 52 kB 608 kB/s
Building wheels for collected packages: google-play-scraper
  Building wheel for google-play-scraper (setup.py) ... done
  Created wheel for google-play-scraper: filename=google_play_scraper-1.1.0-py3-none-any.whl size=24585 sha
  Stored in directory: /root/.cache/pip/wheels/f4/fc/13/fc4c30a74f33bdd956248d2a0f022780fce59250a6a2301ad8
Successfully built google-play-scraper
Installing collected packages: google-play-scraper
Successfully installed google-play-scraper-1.1.0
```

Gambar 2. Script install library google-play-scraper

#### 2. scraping data

```
#Scrape desired number of reviews
#Run kode ini jika ingin scrape data dengan jumlah tertentu. Ganti (misal, ingin scrape sejumlah 1000, maka ganti kod

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.shopee.id',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newest reviews
    count=5000, # defaults to 100
    filter_score_with=None # defaults to None (means all score) Use 1 or 2 or 3 or 4 or 5 to select certain score
)
```

Gambar 3. Script Scraping Data

#### 3. melihat data yang sudah di scraping

```
df_busu = pd.DataFrame(np.array(result), columns=['review'])
df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))
df_busu.head()
```

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent
0	ca40d283-5099-4d9b-972f-ddfac3fa58f2	Pipusa Sawa	lh.googleusercontent.com/a-/AFdZu...	Setelah di update ke versi terbaru malah ga bi...	1	8291	2.88.41	2022-06-16 01:00:13	Hai kak. Terkait kendala tdk bs klik apapun pd...
1	da138e65-ca29-4683-9310-9e7db0427f1b	Iin Ardiana	lh.googleusercontent.com/a-/AFdZu...	Udah 4 tahun pake shopee, seminggu ini ngeseli...	1	21	2.89.30	2022-07-05 02:04:01	Hai ka lin, maaf ya buat kamu ga nyaman. Menge...
2	ae87a017-d6b1-4349-9a23-9ea4f5cd20e6	Sumirat Wati	lh.googleusercontent.com/a-/AFdZu...	Bener bener kecewa saya... enga sama aplikasin...	2	16	2.89.30	2022-07-05 00:18:14	Hai kak , maaf ya kak udah buat kmu ga nyaman🙏...

Gambar 4. Script melihat data yang sudah di scraping

#### 4. Mengecek Data Yang Sudah Di Scraping

```
len(df_busu.index) #count the number of data we got
```

```
150
```

Gambar 5. Script Mengecek Data Yang Sudah Di Scraping

#### 5. Membuat Data Frame Baru Untuk Database Baru

```
my_df.head()
```

	userName	score	at	content
30	Linda Saumi 7B	5	2022-07-05 12:32:53	Masya Allah bagus, untuk aplikasinya memuaskan...
87	Marsha Guslita	1	2022-07-05 12:19:25	Kecewa banget sama ekspedisinya shopee express...
5	Naufa Chinung	3	2022-07-05 10:53:04	Shopee sekarang kenapa ya sering ngebug padaha...
85	asep danan	1	2022-07-05 07:40:49	Pengirimannya sangat sangat lama, keterlambata...
45	Mpok Fina Novita	1	2022-07-05 07:11:10	Saya kecewa sama Shopee biasanya kurir nya cep...

Gambar 6. data frame baru untuk database baru

#### 6. mengekspor data

```
my_df.to_csv("scrapped_data.csv", index = False) #Save the file as CSV , to download: click the folder icon on the left.
```

Gambar 7. Script Ekspor Data

## 2. Proses Data Preprocessing

### 1. case folding

```
# proses case folding
def casefolding(Review):
    Review = Review.lower()
    return Review
data['casefolding'] = data['Review'].apply(casefolding)
data.head(10)
```

	Review	casefolding
0	Kecewa Pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat.. pas coba pesan d...
1	Tulisannya selalu "No Internet Connection" set...	tulisannya selalu "no internet connection" set...
2	Sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru...tetep gg bi...
3	Permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...
4	Selama berbelanja di Shopee lancar2 aja. Dan k...	selama berbelanja di shopee lancar2 aja. dan k...
5	Sudah sekian kali mengajukan Shopeepay / payla...	sudah sekian kali mengajukan shopeepay / payla...
6	Secara umum shopee itu bagus,, Sedikit saran,...	secara umum shopee itu bagus,, sedikit saran,...
7	Shopee dah ga asik buat seller. Udah ga bisa k...	shopee dah ga asik buat seller. udah ga bisa k...
8	Makin diupdate versinya bukannya malah maik ba...	makin diupdate versinya bukannya malah maik ba...
9	Sering lemottttt meski sinyal bagus disini. H...	sering lemottttt meski sinyal bagus disini. h...

Gambar 8. Script Case Folding

## 2. Cleaning

```

1 # proses cleansing remove regex (cleansing) seperti tanda baca dan angka angka
import re
import string
def cleansing(casefolding):
    casefolding = casefolding.strip(" ")
    casefolding = re.sub("[?$.|!:_:"]*(+), '', casefolding)
    casefolding = re.sub("\d+", '', casefolding)
    casefolding = re.sub("\b[a-zA-Z]\b", "", casefolding)
    casefolding = re.sub('\s+', ' ', casefolding)
    return casefolding
data['cleansing'] = data['casefolding'].apply(cleansing)
data.head(10)

```

	Review	casefolding	cleansing
0	Kecewa Pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat pas coba pesan di ...
1	Tulisannya selalu "No Internet Connection" set...	tulisannya selalu "no internet connection" set...	tulisannya selalu no internet connection setia...
2	Sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru tetep gg bisa ...
3	Permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...
4	Selama berbelanja di Shopee lancar2 aja. Dan k...	selama berbelanja di shopee lancar2 aja. dan k...	selama berbelanja di shopee lancar aja dan ket...
5	Sudah sekian kali mengajukan Shopeepay / payla...	sudah sekian kali mengajukan shopeepay / payla...	sudah sekian kali mengajukan shopeepay / payla...

Gambar 9. Script Cleaning

## 3. Tokenize

```

1 from nltk.tokenize import word_tokenize
#NLTK word tokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)
data['tokenize'] = data['cleansing'].apply(word_tokenize_wrapper)
data.head()

```

	Review	casefolding	cleansing	tokenize
0	Kecewa Pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat pas coba pesan di ...	[kecewa, pengirimannya, lambat, pas, coba, pes...
1	Tulisannya selalu "No Internet Connection" set...	tulisannya selalu "no internet connection" set...	tulisannya selalu no internet connection setia...	[tulisannya, selalu, no, internet, connection, ...
2	Sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru tetep gg bisa ...	[sudah, di, update, ke, versi, terbaru, tetep, g...
3	Permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...	[permasalahan, masih, sama, pembayaran, lewat, ...
4	Selama berbelanja di Shopee lancar2 aja. Dan k...	selama berbelanja di shopee lancar2 aja. dan k...	selama berbelanja di shopee lancar aja dan ket...	[selama, berbelanja, di, shopee, lancar, aja, ...

Gambar 10. Script Tokenize

## 4. Stopword

```

1 import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
su = pd.read_csv('stopwords.csv')
def stopwords_removal(tokenize):
    filtering = stopwords.words('Indonesian','english')
    filtering.extend(su)
    x = []
    data = []
    def myFunc(x):
        if x in filtering:
            return False
        else:
            return True
    fit = (filter(myFunc, tokenize))
    for x in fit:
        data.append(x)
    return data
data['stopwords'] = data['tokenize'].apply(stopwords_removal)
data.head(10)

```

	Review	casefolding	cleansing	tokenize	stopwords
0	Kecewa Pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat.. pas coba pesan d...	kecewa pengirimannya lambat pas coba pesan di ...	[kecewa, pengirimannya, lambat, pas, coba, pes...	[kecewa, pengirimannya, lambat, pas, coba, pes...
1	Tulisannya selalu "No Internet Connection" set...	tulisannya selalu "no internet connection" set...	tulisannya selalu no internet connection setia...	[tulisannya, selalu, no, internet, connection, ...	[tulisannya, no, internet, connection, bel, e...
2	Sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru...tetep gg bi...	sudah di update ke versi terbaru tetep gg bisa ...	[sudah, di, update, ke, versi, terbaru, tetep, g...	[update, versi, terbaru, tetep, gg, ngdapelin, ...
3	Permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...	permasalahan masih sama pembayaran lewat tagih...	[permasalahan, masih, sama, pembayaran, lewat, ...	[permasalahan, pembayaran, tagihan, pulsa, ob...
4	Selama berbelanja di Shopee lancar2 aja. Dan k...	selama berbelanja di shopee lancar2 aja. dan k...	selama berbelanja di shopee lancar aja dan ket...	[selama, berbelanja, di, shopee, lancar, aja, ...	[berbelanja, shopee, lancar, aja, permasalahan...
5	Sudah sekian kali mengajukan Shopeepay / payla...	sudah sekian kali mengajukan shopeepay / payla...	sudah sekian kali mengajukan shopeepay / payla...	[sekitan kali mengajukan shopeepay / payla...	[sekitan kali mengajukan shopeepay / payla...
6	Secara umum shopee itu bagus... sedikit saran...	secara umum shopee itu bagus... sedikit saran...	secara umum shopee itu bagus sedikit saran set...	[secara, umum, shopee, itu, bagus, sedikit, sa...	[shopee, bagus, saran, seller, yg, respon, min...
7	Shopee dah ga asik buat seller. udah ga bisa k...	shopee dah ga asik buat seller. udah ga bisa k...	shopee dah ga asik buat seller udah ga bisa ki...	[shopee, dah, ga, asik, buat, seller, udah, ga, k...	[shopee, dah, ga, asik, buat, seller, udah, ga, k...
8	Makin diupdate versinya bukannya malah maik ba...	makin diupdate versinya bukannya malah maik ba...	makin diupdate versinya bukannya malah maik ba...	[makin, diupdate, versinya, bukannya, malah, m...	[diupdate, versinya, malah, maik, bagus, lelet, susah...
9	Sering lemot!!!! meski sinyal bagus disini. H...	sering lemot!!!! meski sinyal bagus disini. H...	sering lemot!!!! meski sinyal bagus disini ha...	[sering, lemot!!!!, meski, sinyal, bagus, disini, ...	[lemot!!!!, sinyal, bagus, hadn, masak, apik...

Gambar 11. Script Stopword

### 5. Stemming

```

1 # proses stemming
2 ! pip install Sastrawi
3 from sklearn.pipeline import Pipeline
4 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
5
6 def stemming(stopwords):
7     factory = StemmerFactory()
8     stemmer = factory.create_stemmer()
9     do = []
10    for w in stopwords:
11        dt = stemmer.stem(w)
12        do.append(dt)
13    d_clean=[]
14    d_clean=" ".join(do)
15    print(d_clean)
16    return d_clean
17 data['stopwords'] = data['stopwords'].apply(stemming)
18
19 data.to_csv('datacleaning.csv', index=False)
20 data_clean = pd.read_csv('datacleaning.csv', encoding='latin1')
21 data_clean.head()
22
23 Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-telemetry/pip-install-
24 Requirement already satisfied: Sastrawi in /usr/local/lib/python3.7/dist-packages (1.0.1)
25 kecewa kirla lambat pas coba pesan tokopedia disiplinir jesa kirla lambat makok estimasi pas tanggal tujuantiba yah pas tanggal estimasi ampe minggu yanhepas kak akun gold silver
26 tulis no internet connection beli emas internet lancar aplikasi versi baru uninstal aplikasi installi uang beli voucher shopee deals sekitar error tolong baik
27 update versi terbaruvoucher gg hndadastin voucherjemasu pesan bayar ternyata tulis maaf salah sinyal bagus sskaliberturut tulis
28 masalah bayar tagih pulsa buka loding buka padal jaring bagus susah masuk lg ubah bayar belanja masalah susah buka hapus aja bayar tagih pulsa dshopee gak susah susah top up tagih tunggak karna transaksi dshopee
29 sekian kali aju shopeeoya sayilater tolak trus data sdn bener gak yg kurangmasf klo kasih bintang dikit
30 shopee bagus saran seller yg respon kecewa banned aja jaga percaya langgan
31 shopee dan ga asik seller udah ga klik naik produk dipromosin barang dagang seller malas jual dshopee malas buka aplikasi shopee aq uninstal aja deh aplikasi asik
32 dshopee versi naik bagus lilet susah bukin aja aj sdn pnsdn apk belan hari rani apa belan
33 jemottttt sinyal bagus hadan nasak aplikasi shopee kayak gin sih
34 nya suka apk tritopi kesini fitur yg apk beratapalagi game game nya bintang kuranglampai semua jalan lancar bintang kurang lagtinyal internet stabil top bayar menuu pulsa tagih gagal gagal top saldo potong monon baik
35 bagus si baik baik aja newer gaada keliru ato pokok recommended banget deh
36 update aja lot aplikasi lancar lancar aja kalo buka shopee haduh lot ampun kaya berat banget baik biar nyaman pake aplikasi
    
```

Gambar 12. Script Stemming

### 3. Pelabelan Kata

id	text	aspectCategory	polarity
1	hadeh update bikin emosi coba uninstal install login kembalitetep g	APLIKASI	NEGATIF
2	shopee sich okee ga tau nich rekan shopee asuransi bener ga baranj	APLIKASI	POSITIF
3	kecewa aplikasi kecewa nya gin pesan terima terima bayar nya onlin	APLIKASI	POSITIF
4	jelek bgt lot lgi pncet brp kali ngk ngaruh respons nya kluar aplikasi d	APLIKASI	NEGATIF
5	pake shopee woy bener jelek aplikasi shopee express standar gila	APLIKASI	NEGATIF
6	suka promo biaya	PEMBAYARAN	POSITIF
7	kecewa bijak kirim gak milih jasa kirim semenjak pilih jasa kirim kalo	PENGIRIMAN	NEGATIF
8	aplikasi gak guna voucher checkout sulit ky gak niat ngasih voucher	PEMBAYARAN	NEGATIF
9	suka shopee kalo bener mw kirim express sndiri ya bener kadang ba	PENGIRIMAN	POSITIF
10	shopee bener bad masak habis update gak kliktolong baik belanja sus	APLIKASI	NEGATIF
11	shopee lot sihpadahal update lot ngga ilangudh loh	APLIKASI	NEGATIF
12	abis update crash aplikasi gabisa klik apapu sinyal udah aman	APLIKASI	NEGATIF
13	aplikasi bagus sayang lot server ngak terima langgan turun bintang b	APLIKASI	NEGATIF
14	biaya ongkir kalimantan mahal bangetjadi pindah lazada akurat harz	PENGIRIMAN	NEGATIF
15	sihhhhh daftar gabisa logout aplikasi coba tetep aja gabisa	APLIKASI	NEGATIF
16	updet nya bagus keluar udah logout akun pedahal udh clear cache a	APLIKASI	NEGATIF
17	update rusak cuy apk nya tolong baik	APLIKASI	NEGATIF
18	hbs diupdate versi baru klik kluar mulu yah jaring internet pdhl kenc	APLIKASI	NEGATIF
19	no indonesia ko batal transaksi uang kembali alas error system tung	PEMBAYARAN	NEGATIF

Gambar 13. Pelabelan Data

### 4. pemodelan knn

#### 1. upload dataset

```

1 import string
2 import pandas as pd
3 import numpy as np
4
5 from google.colab import files
6 uploaded = files.upload()
    
```

Gambar 14. Script Upload dataset

## 2. menampilkan dataset yang telah di upload

```
data = pd.read_csv("dataset.csv", sep=',', encoding='latin1')
data.head(10)
```

	id	text	aspectCategory	polarity
0	1	hadeh update bikin emosi coba uninstall insta...	APLIKASI	NEGATIF
1	2	shopee sich okeo ga tau nich rekan shopee asur...	APLIKASI	POSITIF
2	3	kecewa aplikasi kecewa nya gin pesan terima te...	APLIKASI	POSITIF
3	4	jelek bgt lot lgi pncot brp kali ngk ngaruh re...	APLIKASI	NEGATIF
4	5	pake shopee woyy bener jelek aplikasi shopee e...	APLIKASI	NEGATIF
5	6	suka promo biaya	PEMBAYARAN	POSITIF
6	7	kecewa bijak kirim gak milih jasa kirim semenj...	PENGIRIMAN	NEGATIF
7	8	aplikasi gak guna voucher checkout sulit ky ga...	PEMBAYARAN	NEGATIF
8	9	suka shopee kalo bener mw kirim express sndiri...	PENGIRIMAN	POSITIF
9	10	shope bener bad masak habis update gak kiktol...	APLIKASI	NEGATIF

Gambar 15. Script Tampilan dataset yang telah di upload

## 3. proses tf idf

```
# PROSES TF IDF
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = CountVectorizer()
X = vectorizer.fit_transform(data['text'].astype('U'))

tf = TfidfVectorizer()
text_tf = tf.fit_transform(data['text'].astype('U'))
print(text_tf)
```

```
(0, 1534) 0.1100848329979575
(0, 345) 0.21097929565138715
(0, 2871) 0.21097929565138715
(0, 2420) 0.16767133543566085
(0, 2168) 0.11814444801588173
(0, 367) 0.21097929565138715
(0, 4032) 0.21097929565138715
(0, 2448) 0.20006873698046823
(0, 4552) 0.14301509846282298
(0, 3009) 0.10681012076602599
(0, 4410) 0.09444468082900921
(0, 1686) 0.10945901994013695
(0, 2869) 0.1923275724084987
(0, 1302) 0.08474099476879211
(0, 1979) 0.22762802645788446
(0, 2834) 0.27108875170389546
(0, 723) 0.18141701373757976
(0, 3436) 0.13996556535771962
(0, 3919) 0.1923275724084987
(0, 212) 0.28603019692564596
(0, 164) 0.12279543181848303
(0, 2509) 0.14301509846282298
```

Gambar 16. Script Proses TF IDF

## 4. Proses compute similarity

```
[ ] # Import cosine_similarity metrics
from sklearn.metrics.pairwise import cosine_similarity

# compute similarity using cosine similarity
cos_sim=cosine_similarity(text_tf, text_tf)

print(cos_sim)
```

```
[ [1. 0. 0. ... 0. 0. 0. ]
 [0. 1. 0. ... 0.06264619 0.03795915 0.02815454]
 [0. 0. 1. ... 0. 0.05986904 0.02629429]
 ...
 [0. 0.06264619 0. ... 1. 0.13135731 0.04165818]
 [0. 0.03795915 0.05986904 ... 0.13135731 1. 0.09008815]
 [0. 0.02815454 0.02629429 ... 0.04165818 0.09008815 1. ] ]
```

Gambar 17. Script Proses Compute Similarity

## 5. Splitting Data

```
# splitting data
import collections, numpy
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(cos_sim, data['polarity'], test_size=0.2, random_state=33)
print("Jumlah Data Uji:", X_test.shape)
print("Jumlah Data Latih:", X_train.shape)

pos = (y_test == 'POSITIF').sum()
neg = (y_test == 'NEGATIF').sum()
postrain = (y_train == 'POSITIF').sum()
negtrain = (y_train == 'NEGATIF').sum()
total = pos + neg
print("Jumlah data uji dengan sentimen positif:", pos)
print("Jumlah data uji dengan sentimen negatif:", neg)
print("Jumlah data latih dengan sentimen positif:", postrain)
print("Jumlah data latih dengan sentimen negatif:", negtrain)
data['polarity'].value_counts()
```

```
[ ] Jumlah Data Uji: (374, 1869)
Jumlah Data Latih: (1495, 1869)
Jumlah data uji dengan sentimen positif: 185
Jumlah data uji dengan sentimen negatif: 189
Jumlah data latih dengan sentimen positif: 750
Jumlah data latih dengan sentimen negatif: 745
POSITIF 935
NEGATIF 934
Name: polarity, dtype: int64
```

Gambar 18. Script Splitting data

## 6. Menghitung peforma KNN

UNIVERSITAS  
MERCU BUANA

```

# perform algoritma KNN
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.neighbors import KNeighborsClassifier

clf = KNeighborsClassifier(n_neighbors=7).fit(X_train, y_train)
predicted = clf.predict(X_test)
print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
tn, fp, fn, tp = confusion_matrix(y_test, predicted).ravel()
print("TN:", tn)
print("FP:", fp)
print("FN:", fn)
print("TP:", tp)
print(classification_report(y_test, predicted, zero_division=0))
print('=====\n')
print("Hasil Klasifikasi Sentimen Analisis Shopee:")
print("Accuracy:" , accuracy_score(y_test,predicted))
print("Precision:" , precision_score(y_test,predicted, average="binary", pos_label="POSITIF"))
print("Recall:" , recall_score(y_test,predicted, average="binary", pos_label="POSITIF"))
print("f1_score:" , f1_score(y_test,predicted, average="binary", pos_label="POSITIF"))
print("error_rate:" , 1-accuracy_score(y_test,predicted))

```

confusion matrix:

```

[[158  31]
 [ 52 133]]
TN: 158
FP: 31
FN: 52
TP: 133

```

	precision	recall	f1-score	support
NEGATIF	0.75	0.84	0.79	189
POSITIF	0.81	0.72	0.76	185
accuracy			0.78	374
macro avg	0.78	0.78	0.78	374
weighted avg	0.78	0.78	0.78	374

```

=====
Hasil Klasifikasi Sentimen Analisis Shopee:
Accuracy: 0.7780748663101604
Precision: 0.8109756097560976
Recall: 0.7189189189189189
f1_score: 0.7621776504297995
error_rate: 0.22192513368983957

```

Gambar 19. Script Peforma KNN

## 7. Confusion Matrix

```

from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier

```

Gambar 20. Script Confusion Matrix

## 8. Perhitungan Cross Validation Score

```

knn = KNeighborsClassifier(n_neighbors=3)
cross_val_score(knn, cos_sim, data['polarity'], cv=10)

```

```

array([[0.76470588, 0.7486631 , 0.68449198, 0.75935829, 0.77005348,
        0.68449198, 0.79679144, 0.73262032, 0.74331551, 0.72043011])

```

```

[ ] cross_val_score(knn, cos_sim, data['polarity'], cv=10).mean()
0.7404922086136507

```

Gambar 21. Script Cross Validation Score

## 5. Pemodelan Naive Bayes

## 1. Upload Dataset

```
[ ] import pandas as pd
import numpy as np
import matplotlib as nlp

from google.colab import files
uploaded = files.upload()
```

Gambar 22. Script Upload Dataset

## 2. menampilkan dataset yang telah di upload

```
[ ] data = pd.read_csv("datasetNB.csv", sep=',', encoding='latin1')
data.head(10)
```

	text	polarity
0	hadeh update bikin emosi coba uninstall insta...	0
1	shopee sich okee ga tau nich rekan shopee asur...	1
2	kecewa aplikasi kecewa nya gin pesan terima te...	1
3	jelek bgt lot lgi pncet brp kali ngk ngaruh re...	0
4	pake shopee woyy bener jelek aplikasi shopee e...	0
5	suka promo biaya	1
6	kecewa bijak kirim gak milih jasa kirim semenj...	0
7	aplikasi gak guna voucher checkout sulit ky ga...	0
8	suka shoope kalo bener mw kirim express sndiri...	1
9	shope bener bad masak habis update gak kilktol...	0

Gambar 23. Script Menampilkan Data yang sudah di Upload

## 3. Menghitung Kolom Dan Baris Dataset

```
[ ] data = data.sample(frac=1)
print("Dataset shape:", data.shape)
```

Dataset shape: (1869, 2)

Gambar 24. Script Hitung kolom dan baris Dataset

## 4. Melihat Data Yang Kosong

```
[ ] [(data.isnull().sum() / len(data))*100

text      0.0
polarity  0.0
dtype: float64
```

Gambar 25. Script Melihat data yang kosong

## 5. Melihat Data Yang Positif Dan Negatif



```

positives = data['polarity'][data.polarity == 1 ]
negatives = data['polarity'][data.polarity == 0 ]

print('Total length of the data is:      {}'.format(data.shape[0]))
print('No. of positive tagged sentences is: {}'.format(len(positives)))
print('No. of negative tagged sentences is: {}'.format(len(negatives)))

```

Total length of the data is: 1869  
No. of positive tagged sentences is: 935  
No. of negative tagged sentences is: 934

Gambar 26. Script Melihat data yang positif dan negatif

## 6. Proses Tokenisasi Merubah Kalimat Menjadi Sekumpulan Kata

```

tokenized=data['text'].apply(lambda x: x.split())
tokenized.head(5)

```

772 [update, eror, kecewa, langgan, kirim, lamaa, ...  
1852 [alhmdlh, dgn, shopee, aq, ga, hrs, bolak-bal...  
878 [bantu, belanja, kareena, diskon, yg, tarik, b...  
586 [mantap, cuman, tolong, shopeefood, banyak, di...  
454 [seller, berat, biaya, admin, hitung, jual, pr...  
Name: text, dtype: object

Gambar 27. Script Tokenize

## 7. Proses Countvectorizer

```

[ ] from sklearn.feature_extraction.text import CountVectorizer
    from nltk.tokenize import RegexpTokenizer
    token = RegexpTokenizer(r'[a-zA-Z0-9]+')
    cv = CountVectorizer(stop_words='english',ngram_range = (1,1),tokenizer = token.tokenize)
    text_counts = cv.fit_transform(data['text'].values.astype('U'))

```

Gambar 28. Script Countvectorizer

## 8. Split Data

```

[ ] from sklearn.model_selection import train_test_split
    X=text_counts
    y=data['polarity']
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20,random_state=19)

```

Gambar 29. Script Split Data

## 9. Perhitungan Cross Validation Dan Accuracy

```

from sklearn.naive_bayes import ComplementNB
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_score
from sklearn import metrics
from math import *
cnb = ComplementNB()
cnb.fit(X_train, y_train)
cross_cnb = cross_val_score(cnb, X, y, n_jobs = -1)
print("Cross Validation score = ", cross_cnb)
print ("Train accuracy = {:.2f}%".format(cnb.score(X_train,y_train)*100))
print ("Test accuracy = {:.2f}%".format(cnb.score(X_test,y_test)*100))
train_acc_cnb=cnb.score(X_train,y_train)
test_acc_cnb=cnb.score(X_test,y_test)

```

Cross Validation score = [0.82887701 0.84224599 0.8368984 0.81550802 0.84986595]  
 Train accuracy =91.64%  
 Test accuracy =85.29%

Gambar 30. Script Cross Validation dan Accuracy

## 10. Perhitungan Confusion Matrix

```

from sklearn.metrics import *
y_pred_cnb =cnb.predict(X_test)
from sklearn.metrics import confusion_matrix
print(confusion_matrix(y_test,y_pred_cnb))

```

[[164 21]  
 [ 34 155]]

Gambar 31. Script Confusion Matrix

## 11. Mengecek Peforma Model Dengan Klasifikasi Report

```

print(classification_report(y_test, y_pred_cnb))
roc_score_cnb=roc_auc_score(y_test, y_pred_cnb)
print("Area Under the Curve = ",roc_score_cnb)

```

	precision	recall	f1-score	support
0	0.83	0.89	0.86	185
1	0.88	0.82	0.85	189
accuracy			0.85	374
macro avg	0.85	0.85	0.85	374
weighted avg	0.85	0.85	0.85	374

Area Under the Curve = 0.8532961532961533

Gambar 32. Script Performa Model Klasifikasi

## 12. Perhitungan F1 Score, Precision, Recall

```

from sklearn.metrics import *
print("F1 score = {:.2f}%".format(f1_score(y_test, y_pred_cnb, average="macro")*100))
f1_cnb=f1_score(y_test, y_pred_cnb, average="macro")
print("Precision score = {:.2f}%".format(precision_score(y_test, y_pred_cnb, average="macro")*100))
precision_cnb=precision_score(y_test, y_pred_cnb, average="macro")
print("Recall score = {:.2f}%".format(recall_score(y_test, y_pred_cnb, average="macro")*100))
recall_cnb=recall_score(y_test, y_pred_cnb, average="macro")

```

F1 score =85.29%  
 Precision score =85.45%  
 Recall score =85.33%

Gambar 33. Hitung F1 Score, Precision, Recall