

# Mini\_Project\_Week\_10\_ADMS\_Amanda\_Gozali

May 29, 2023

Mini Project Week 10 ### Amanda Gozali - 6162001169

## 1 Preparing Data

```
[1]: #Mengambil data
import pandas as pd
url = 'https://docs.google.com/spreadsheets/d/
↳1lN3x5lmwCHPph00atZqouRJVeyuRp16t1DqpqqkUcGs/gviz/tqx=out:
↳csv&sheet=Sheet1'
df = pd.read_csv(url)
```

```
[2]: #Mengambil 5 data teratas
df.head()
```

```
[2]:
```

	No	Brand	Rating	\
0	1	Pizza Hut Delivery	3	
1	2	Mcdonald's	1	
2	3	KFC	5	
3	4	Burger Bangor	3	
4	5	Burger King	5	

  

	Review	\
0	rose pizza, rotinya masih mentah yang bagian d...	
1	Saya pesan setau saya gambarnya pake nasi. Kal...	
2	Ini mah nggak usah diragukan, udah langganan j...	
3	Ukuran kecil untuk harga segitu. Tapi rasa ok.	
4	enak bumbu madura nya, ayamnya pun juicy ga ke...	

  

	Source
0	https://gofood.co.id/en/bandung/restaurant/piz...
1	https://gofood.co.id/id/bandung/restaurant/mcd...
2	https://gofood.co.id/en/bandung/restaurant/kfc...
3	https://gofood.co.id/en/bandung/restaurant/bur...
4	https://gofood.co.id/id/bandung/restaurant/bur...

```
[3]: #Menghapus data yang kosong
df = df.dropna()
```

```
#Mengambil 5 data terakhir
df.tail()
```

```
[3]:
```

No	Brand	Rating	Review	Source
245	246	Sagala Kitchen	4	
246	247	Bebeke Om Aris	4	
247	248	Boom Burger	2	
248	249	Lu'Miere	5	
249	250	Subway	4	

```

nasi gorengnya porsi nya mini banget. hehe
sambel dan lalapannya, jadi hilang atau keting...
Masakan nya dingin,,
Akhirnya bisa ngerasai Lu'miere di Bandung!! K...
adi kemarin siang aku lagi jalan2 ke paskal da...

https://gofood.co.id/id/bandung/restaurant/sag...
https://gofood.co.id/id/bandung/restaurant/beb...
https://gofood.co.id/id/bandung/restaurant/boo...
https://pergikuliner.com/restaurants/bandung/l...
https://pergikuliner.com/restaurants/bandung/s...
```

```
[4]: df.shape #Melihat ukuran data (250 baris, 5 kolom)
```

```
[4]: (250, 5)
```

```
[5]: df.describe() #Deskripsi data
```

```
[5]:
```

	No	Rating
count	250.000000	250.000000
mean	125.500000	3.000000
std	72.312977	1.359187
min	1.000000	1.000000
25%	63.250000	2.000000
50%	125.500000	3.000000
75%	187.750000	4.000000
max	250.000000	5.000000

```
[6]: df.columns
```

```
[6]: Index(['No', 'Brand', 'Rating', 'Review', 'Source'], dtype='object')
```

```
[7]: df['Rating'].value_counts()
```

```
[7]: 3    60
      4    50
      2    50
      1    45
      5    45
      Name: Rating, dtype: int64
```

Digunakan rating untuk menentukan apakah pelanggan senang atau tidak senang, di mana jika rating antara 1-3 maka dianggap pelanggan tidak senang sehingga akan diubah menjadi negatif dan jika rating 4-5 maka dianggap pelanggan senang dan diubah menjadi positif

```
[8]: df['Rating']= df['Rating'].map({4: 'positif',5: 'positif', 1: 'negatif',2:
      ↪ 'negatif',3: 'negatif'})
```

```
[9]: df.head()
```

```
[9]:   No      Brand  Rating \
0    1  Pizza Hut Delivery  negatif
1    2      Mcdonald's  negatif
2    3           KFC    positif
3    4  Burger Bangor  negatif
4    5    Burger King    positif

      Review \
0  rose pizza, rotinya masih mentah yang bagian d...
1  Saya pesan setau saya gambarnya pake nasi. Kal...
2  Ini mah nggak usah diragukan, udah langganan j...
3  Ukuran kecil untuk harga segitu. Tapi rasa ok.
4  enak bumbu madura nya, ayamnya pun juicy ga ke...

      Source
0  https://gofood.co.id/en/bandung/restaurant/piz...
1  https://gofood.co.id/id/bandung/restaurant/mcd...
2  https://gofood.co.id/en/bandung/restaurant/kfc...
3  https://gofood.co.id/en/bandung/restaurant/bur...
4  https://gofood.co.id/id/bandung/restaurant/bur...
```

## 2 Text Preprocessing

preprosesing disini adalah melakukan, cleaning data teks dengan menghapus kata kata yang tidak penting yang tidak termsuk dalam konteks, contohnya adalah seperti berikut.

```
[11]: import re

      from bs4 import BeautifulSoup
      from nltk.tokenize import WordPunctTokenizer
      import nltk
```

```

from nltk.corpus import stopwords

tok = WordPunctTokenizer()
pat1 = r'@[A-Za-z0-9_]+' #menghilangkan username jika twitter
pat2 = r'https?://[^\ ]+' #menghilangkan situs website
combined_pat = r'|'.join((pat1, pat2)) #join pat1 dan pat 2
www_pat = r'www.[^\ ]+' #menghilangkan situs website

#set_stopword yang di deskripsikan sendiri
stp = pd.read_csv('stopwords_id.txt')['STOPWORDS'].tolist() #['STOPWORDS']
↳berupa header
print(stp)
stopword_user = set(stp)
print(stopword_user)

```

```

['ada', 'adanya', 'adalah', 'adapun', 'agak', 'agaknya', 'agar', 'akan',
'akankah', 'akhirnya', 'aku', 'akulah', 'amat', 'amatlah', 'anda', 'andalah',
'antar', 'diantaranya', 'antara', 'antaranya', 'diantara', 'apa', 'apaan',
'mengapa', 'apabila', 'apakah', 'apalagi', 'apatah', 'atau', 'ataukah',
'ataupun', 'bagai', 'bagaikan', 'sebagai', 'sebagainya', 'bagaimana',
'bagaimanapun', 'sebagaimana', 'bagaimanakah', 'bagi', 'bahkan', 'bahwa',
'bahwasanya', 'sebaliknya', 'banyak', 'sebanyak', 'beberapa', 'seberapa',
'begini', 'beginian', 'beginikah', 'beginilah', 'sebegini', 'begitu',
'begitukah', 'begitulah', 'begitupun', 'sebegini', 'belum', 'belumlah',
'sebelum', 'sebelumnya', 'sebenarnya', 'berapa', 'berapakah', 'berapalah',
'berapapun', 'betulkah', 'sebetulnya', 'biasa', 'biasanya', 'bila', 'bilakah',
'bisa', 'bisakah', 'sebisanya', 'boleh', 'bolehkah', 'bolehlah', 'buat',
'bukan', 'bukankah', 'bukanlah', 'bukannya', 'cuma', 'percuma', 'dahulu',
'dalam', 'dan', 'dapat', 'dari', 'daripada', 'dekat', 'demi', 'demikian',
'demikianlah', 'sedemikian', 'dengan', 'depan', 'di', 'dia', 'dialah', 'dini',
'diri', 'dirinya', 'terdiri', 'dong', 'dulu', 'enggak', 'enggaknya', 'entah',
'entahlah', 'terhadap', 'terhadapnya', 'hal', 'hampir', 'hanya', 'hanyalah',
'harus', 'haruslah', 'harusnya', 'seharusnya', 'hendak', 'hendaklah',
'hendaknya', 'hingga', 'sehingga', 'ia', 'ialah', 'ibarat', 'ingin', 'inginkan',
'inginkan', 'ini', 'inikah', 'inilah', 'itu', 'itukah', 'itulah', 'jangan',
'jangan', 'janganlah', 'jika', 'jikalau', 'juga', 'justeru', 'kala', 'kalau',
'kalaulah', 'kalaupun', 'kalian', 'kami', 'kamilah', 'kamu', 'kamulah', 'kan',
'kapan', 'kapankah', 'kapanpun', 'dikarenakan', 'karena', 'karenanya', 'ke',
'kecil', 'kemudian', 'kenapa', 'kepada', 'kepadanya', 'ketika', 'seketika',
'khususnya', 'kini', 'kinilah', 'kiranya', 'sekiranya', 'kita', 'kitalah',
'kok', 'lagi', 'lagian', 'selagi', 'lah', 'lain', 'lainnya', 'melainkan',
'selaku', 'lalu', 'melalui', 'terlalu', 'lama', 'lamanya', 'selama', 'selama',
'selamanya', 'lebih', 'terlebih', 'bermacam', 'macam', 'semacam', 'maka',
'makanya', 'makin', 'malah', 'malahan', 'mampu', 'mampukah', 'mana', 'manakala',
'manalagi', 'masih', 'masihkah', 'semasih', 'masing', 'mau', 'maupun',
'semaunya', 'memang', 'mereka', 'merekalah', 'meski', 'meskipun', 'semula',

```

'mungkin', 'mungkinkah', 'nah', 'namun', 'nantinya', 'nyaris', 'oleh',  
 'olehnya', 'seorang', 'seseorang', 'pada', 'padanya', 'padahal', 'paling',  
 'sepanjang', 'pantas', 'sepantasnya', 'sepantasnyalah', 'para', 'pasti',  
 'pastilah', 'per', 'pernah', 'pula', 'pun', 'merupakan', 'rupanya', 'serupa',  
 'saat', 'saatnya', 'sesaat', 'saja', 'sajalah', 'saling', 'bersama', 'sama',  
 'sesama', 'sambil', 'sampai', 'sana', 'sangat', 'sangatlah', 'saya', 'sayalah',  
 'se', 'sebab', 'sebabnya', 'sebuah', 'tersebut', 'tersebutlah', 'sedang',  
 'sedangkan', 'sedikit', 'sedikitnya', 'segala', 'segalanya', 'segera',  
 'sesegera', 'sejak', 'sejenak', 'sekali', 'sekalian', 'sekalipun', 'sesekali',  
 'sekaligus', 'sekarang', 'sekarang', 'sekitar', 'sekitarnya', 'sela', 'selain',  
 'selalu', 'seluruh', 'seluruhnya', 'semakin', 'sementara', 'sempat', 'semua',  
 'semuanya', 'sendiri', 'sendirinya', 'seolah', 'seperti', 'sepertinya',  
 'sering', 'seringnya', 'serta', 'siapa', 'siapakah', 'siapapun', 'disini',  
 'disinilah', 'sini', 'sinilah', 'sesuatu', 'sesuatunya', 'suatu', 'sesudah',  
 'sesudahnya', 'sudah', 'sudahkah', 'sudahlah', 'supaya', 'tadi', 'tadinya',  
 'tak', 'tanpa', 'setelah', 'telah', 'tentang', 'tentu', 'tentulah', 'tentunya',  
 'tertentu', 'seterusnya', 'tapi', 'tetapi', 'setiap', 'tiap', 'setidaknya',  
 'tidak', 'tidakkah', 'tidaklah', 'toh', 'waduh', 'wah', 'wahai', 'sewaktu',  
 'walau', 'walaupun', 'wong', 'yaitu', 'yakni', 'yang']  
 {'sedemikian', 'masihkah', 'olehnya', 'itulah', 'lain', 'sebisanya', 'dirinya',  
 'bahkan', 'sekaligus', 'kecil', 'berapapun', 'ini', 'ingin', 'terhadapnya',  
 'lamanya', 'apa', 'dialah', 'manalagi', 'hendaklah', 'kalaupun', 'entah',  
 'kamu', 'kepadanya', 'sebetulnya', 'makanya', 'memang', 'kamilah', 'sana',  
 'tak', 'ke', 'sebabnya', 'adapun', 'semuanya', 'dikarenakan', 'sesegera',  
 'amatlah', 'buat', 'haruslah', 'kalian', 'bagaimanakah', 'selamanya', 'lah',  
 'tapi', 'walau', 'mampu', 'dia', 'bagaimanapun', 'kita', 'dini', 'dengan',  
 'sekarang', 'hendaknya', 'adalah', 'saya', 'harus', 'lagian', 'di', 'terlalu',  
 'betulkah', 'tetapi', 'dapat', 'tertentu', 'lainnya', 'diantaranya', 'para',  
 'saja', 'sama', 'saling', 'belum', 'pasti', 'pernah', 'sepantasnya', 'seolah',  
 'yakni', 'seterusnya', 'sangat', 'akan', 'janganlah', 'tersebut', 'mau', 'sela',  
 'inginkan', 'kok', 'beginilah', 'ia', 'juga', 'sebuah', 'pantas', 'selama',  
 'beginikah', 'depan', 'seluruh', 'sepertinya', 'jikalau', 'ketika', 'kini',  
 'dahulu', 'kamulah', 'bolehkah', 'walaupun', 'sesama', 'kapanpun',  
 'sepantasnyalah', 'agaknya', 'sudahlah', 'makin', 'seringnya', 'melainkan',  
 'justru', 'bahwa', 'tentulah', 'bukanlah', 'maka', 'bukan', 'aku', 'inikah',  
 'sendirinya', 'sejak', 'sering', 'itukah', 'se', 'mampukah', 'sebagainya',  
 'kalau', 'sementara', 'sekalian', 'disini', 'anda', 'malahan', 'kalaulah',  
 'demi', 'sebagai', 'khususnya', 'siapapun', 'sejenak', 'bahwasanya', 'beberapa',  
 'pula', 'akulah', 'antar', 'begitupun', 'sempat', 'yaitu', 'sebelumnya',  
 'semasih', 'tanpa', 'supaya', 'kitalah', 'karenanya', 'seberapa', 'sajalah',  
 'kiranya', 'segera', 'daripada', 'diantara', 'padanya', 'sinilah', 'mengapa',  
 'kapankah', 'seorang', 'selain', 'hingga', 'jangankan', 'saat', 'telah',  
 'macam', 'merekalah', 'sebab', 'berapakah', 'nah', 'sebanyak', 'bagaimana',  
 'nantinya', 'cuma', 'hendak', 'pun', 'mereka', 'lama', 'sebegitu', 'sebegini',  
 'diri', 'meskipun', 'agar', 'tadi', 'biasa', 'dalam', 'apakah', 'sekitar',  
 'tadinya', 'namun', 'per', 'rupanya', 'jika', 'semacam', 'sudahkah',  
 'segalanya', 'malah', 'sedang', 'banyak', 'mungkin', 'mungkinkah', 'sesekali',  
 'bisa', 'boleh', 'saatnya', 'bagai', 'kepada', 'karena', 'selagi', 'seharusnya',

'bukankah', 'demikian', 'sedangkan', 'siapa', 'sekali', 'akhirnya', 'disinilah', 'akankah', 'apalagi', 'enggak', 'berapa', 'sebaliknya', 'sendiri', 'nyaris', 'pastilah', 'sewaktu', 'lebih', 'seluruhnya', 'bolehlah', 'wahai', 'percuma', 'paling', 'sekiranya', 'serupa', 'hanya', 'meski', 'terhadap', 'antaranya', 'antara', 'dan', 'andalah', 'setidaknya', 'semaunya', 'ialah', 'beginian', 'sekalipun', 'tentu', 'sepanjang', 'hampir', 'apaan', 'apabila', 'tentang', 'sebelum', 'segala', 'waduh', 'wah', 'sesudahnya', 'begitu', 'setelah', 'maupun', 'hanyalah', 'sampai', 'kinilah', 'biasanya', 'seseorang', 'terdiri', 'sekitarnya', 'sesudah', 'dari', 'seketika', 'begini', 'sesuatu', 'serta', 'tiap', 'sehingga', 'bila', 'kan', 'selalu', 'amat', 'ataukah', 'berapalah', 'sini', 'bilakah', 'masih', 'adanya', 'harusnya', 'terlebih', 'agak', 'wong', 'kapan', 'tidak', 'tidaklah', 'belumah', 'sedikit', 'dekat', 'kemudian', 'ibarat', 'itu', 'pada', 'masing', 'melalui', 'hal', 'semakin', 'ataupun', 'begitukah', 'sesuatunya', 'yang', 'begitulah', 'sayalah', 'entahlah', 'bermacam', 'enggaknya', 'sudah', 'siapakah', 'bisakah', 'bagi', 'lalu', 'toh', 'bersama', 'manakala', 'bagaikan', 'jangan', 'semua', 'demikianlah', 'oleh', 'tidakkah', 'sebenarnya', 'bukannya', 'semula', 'seperti', 'ada', 'sambil', 'tersebutlah', 'selaku', 'sebagaimana', 'padahal', 'sedikitnya', 'tentunya', 'inilah', 'merupakan', 'kenapa', 'kala', 'sangatlah', 'suatu', 'lagi', 'apatah', 'inginkan', 'dulu', 'setiap', 'mana', 'kami', 'dong', 'nantinya', 'atau', 'sesaat'}

```
[12]: def proses_teks(teks):
        soup = BeautifulSoup(teks, 'lxml')
        souped = soup.get_text()
        try:
            teks = souped.decode("utf-8-sig").replace(u"\ufffd", "?")
        except:
            teks = souped
        teks_bersih= re.sub("[^a-zA-Z0-9]", " ",(re.sub(www_pat, '', re.
↪sub(combined_pat, '', teks)).lower()))
        teks_bersih= ' '.join([word for word in teks_bersih.split() if word not in_
↪stopword_user])
        return (" ".join([x for x in tok.tokenize(teks_bersih) if len(x) > 1])).
↪strip()
```

```
[13]: #Membuat kolom baru
x=[]
for teks in df.Review:
    x.append(proses_teks(teks))
```

<ipython-input-12-31fe8bf66b22>:2: MarkupResemblesLocatorWarning: The input looks more like a filename than markup. You may want to open this file and pass the filehandle into BeautifulSoup.

```
soup = BeautifulSoup(teks, 'lxml')
```

```
[14]: clean_text = pd.DataFrame({'clean_text':x})
```

```
[15]: df = pd.concat([df,clean_text], axis = 1)
df.head()
```

```
[15]:   No      Brand  Rating \
0   1  Pizza Hut Delivery  negatif
1   2      Mcdonald's  negatif
2   3           KFC  positif
3   4  Burger Bangor  negatif
4   5   Burger King  positif

                                     Review \
0  rose pizza, rotinya masih mentah yang bagian d...
1  Saya pesan setau saya gambarnya pake nasi. Kal...
2  Ini mah nggak usah diragukan, udah langganan j...
3      Ukuran kecil untuk harga segitu. Tapi rasa ok.
4  enak bumbu madura nya, ayamnya pun juicy ga ke...

                                     Source \
0  https://gofood.co.id/en/bandung/restaurant/piz...
1  https://gofood.co.id/id/bandung/restaurant/mcd...
2  https://gofood.co.id/en/bandung/restaurant/kfc...
3  https://gofood.co.id/en/bandung/restaurant/bur...
4  https://gofood.co.id/id/bandung/restaurant/bur...

                                     clean_text
0      rose pizza rotinya mentah bagian dalamnya
1  pesan setau gambarnya pake nasi kalo ga pake y...
2      mah nggak usah diragukan udah langganan wkwwk
3      ukuran untuk harga segitu rasa ok
4      enak bumbu madura nya ayamnya juicy ga kering
```

### 3 Vectorizer

```
[16]: from sklearn.feature_extraction.text import CountVectorizer,TfidfVectorizer,HashingVectorizer
```

```
[17]: cvec=CountVectorizer()
tvec=TfidfVectorizer()
hvec=HashingVectorizer()
```

### 4 Create Train\_test

Setelah melakukan langkah di atas, maka selanjutnya adalah memisahkan data latih dan data uji, untuk memisahkannya, disini menggunakan pustaka sklearn ~ train test split

```
[33]: from sklearn.model_selection import train_test_split
```

```
[34]: X = df.clean_text
      y = df.Rating
```

```
[35]: X_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
      ↪random_state=225)
```

```
[36]: print('Banyaknya data X_train :', len(X_train))
      print('Banyaknya data x_test :', len(x_test))
      print('Banyaknya data y_train :', len(y_train))
      print('Banyaknya data y_test :', len(y_test))
```

```
Banyaknya data X_train : 200
Banyaknya data x_test : 50
Banyaknya data y_train : 200
Banyaknya data y_test : 50
```

```
[37]: from sklearn.feature_extraction.text import CountVectorizer
      vectorizer = CountVectorizer(token_pattern = r'\b\w+\b')
      X_train = vectorizer.fit_transform(X_train)
      x_test = vectorizer.transform(x_test)
```

```
[38]: print(x_test)
```

```
(0, 101)      1
(0, 114)      2
(0, 123)      1
(0, 299)      1
(0, 349)      1
(0, 356)      1
(0, 464)      1
(0, 478)      1
(0, 522)      3
(0, 824)      1
(0, 904)      1
(0, 987)      1
(0, 1005)     1
(0, 1065)     1
(0, 1089)     1
(0, 1105)     1
(0, 1108)     2
(0, 1112)     1
(0, 1120)     1
(0, 1148)     1
(0, 1151)     1
(0, 1182)     1
(0, 1197)     3
(0, 1199)     1
```



```

(0, 1365)      1
:
(45, 1384)     1
(45, 1416)     1
(45, 1571)     1
(45, 1591)     1
(45, 1623)     2
(45, 1659)     1
(45, 1696)     1
(45, 1704)     1
(46, 478)      1
(46, 1153)     1
(46, 1460)     1
(46, 1621)     1
(47, 478)      1
(47, 691)      1
(47, 767)      1
(48, 134)      1
(48, 158)      1
(48, 229)      1
(48, 306)      1
(48, 346)      1
(48, 735)      1
(48, 1241)     1
(48, 1339)     1
(49, 197)      1
(49, 1308)     1

```

## 5 Model Definition

```
[39]: from sklearn.naive_bayes import BernoulliNB
```

```
[40]: Model_NB = BernoulliNB()
```

```
[41]: Model_NB.fit(X_train,y_train)
      #predict output
      hasil = Model_NB.predict(x_test)
```

```
[42]: print(hasil)
```

```

['positif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'positif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif'
 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif' 'negatif']

```

```
'negatif']
```

## 6 Cek Akurasi

```
[43]: from sklearn.metrics import accuracy_score, confusion_matrix  
      from sklearn.metrics import classification_report
```

```
[44]: #Confusion Matrix  
      confusion_matrix(hasil, y_test)
```

```
[44]: array([[30, 18],  
          [ 1,  1]])
```

```
[45]: accuracy_score(hasil, y_test)
```

```
[45]: 0.62
```

```
[46]: print(classification_report(hasil, y_test))
```

	precision	recall	f1-score	support
negatif	0.97	0.62	0.76	48
positif	0.05	0.50	0.10	2
accuracy			0.62	50
macro avg	0.51	0.56	0.43	50
weighted avg	0.93	0.62	0.73	50