

# **DATA SCIENCE**

UJIAN TENGAH SEMESTER : ANALISIS SENTIMEN



**OLEH :**

NADHIFA SOFIA

(19/448721/PPA/05804)

**MAGISTER ILMU KOMPUTER**

**DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS GADJAH MADA**

**2020**

Terlampir bersama soal ini adalah file dataset dalam format csv yang berisi tweet tentang jasa transportasi online yang sudah diberi label(0-negatif/1-positif).

fx   tweet	
1	tweet
2	Saya juga mau vouchee @gojekindonesia https://twitter.com/FOODFESS2/status/1168007370867232769 ...
3	download gojek duluuu uwuuu
4	Aminnn...#orderan goride mhn di lancar kan.all driver"kecuali yg pakai mod tuyul"
5	Tq @gojekindonesia @golifeindonesia ❤️ . Harusnya si bapak ini 🌟 5.0.
6	Khusus daerah Bandung, kalau mau therapist bisa hubungin ke nomor yang ada di kartu. pic.twitter.com/5JMNmnHMQ
7	Semoga Twitter panjang umur. Berkomunikasi dengan penyedia jasa di Indonesia, never been this easy. I just mentioned once, langsung ditanggapi. @HaloBCA and @gojek
8	Semoga di tahun yang baru ini, kita senantiasa ditunjukkan jalan yang lurus, jalan yang terbaik & jalan yang menuntun kita semua ke pintu surga.
9	Selamat Tahun Baru Islam 1 Muharam 1441 H.
10	#IslamicNewYear #TahunBaruIslam #Muharam1441 #PastiAdaJalan pic.twitter.com/QpAbdmYZim
11	Sejauh ini menurut saya UI paling nyaman dari app asli Indonesia itu punyanya @gojekindonesia .
12	Thank you @gojekindonesia pic.twitter.com/pbZBG8IkF7
13	Hai, ada yang bisa bantu mengenai layanan Gojek? Terima kasih ^yun
14	Full Week Feeling Great With You Guys @gojekindonesia @grabtaxiid Terima Kasih 🙏🙏🙏
15	TERIMA KASIH GOJEK N GRAB Driver For This Week 🙏🙏🙏
16	Semoga Anda Sekalian dipenuhi Berkah yang Mau mengirim dengan Tulus , menghalau... https://www.instagram.com/p/B12haXtlhCY/?igshid=1b3yearvm6mqm8 ...
17	

**Gambar 1.** Tampilan 10/4000 dataset yang akan digunakan

Buatlah model klasifikasi yang optimal untuk data tersebut untuk kepentingan opinion mining dari tweet tentang jasa transportasi online. Hal-hal yang perlu dilakukan antara lain adalah:

1. Buat analisis penyelesaian masalah dari kasus tersebut.

*Jawab :*

Analisis sentimen merupakan salah satu implementasi Natural Language Processing untuk menelaah perasaan seseorang berdasarkan kata-kata atau opininya. Untuk kasus di Ujian Tengah Semester ini, merupakan analisa opini dari tweet mengenai suatu topik. Pengumpulan dataset sendiri bisa menggunakan tweepy, dimana harus mendaftarkan diri ke <https://apps.twitter.com/app/new> untuk mendapatkan consumer\_key serta consumer\_secret. Untuk kasus UTS ini, dataset telah diberikan melalui simaster dengan 4000 tweet yang sudah dilabeli dengan 0 yang berarti negatif dan 1 berarti positif. Pada beberapa kasus analisis sentimen, ada juga yang menggunakan 3 label, seperti -1 untuk perasaan negatif, 0 untuk netral, serta +1 untuk positif. Opini sendiri bisa dikategorikan sebagai yang memiliki konteks sentimen secara eksplisit atau jelas, implisit, hingga sarkasme [1]. Analisis sentimen ini bisa membantu para pebisnis untuk memahami sentimen sosial dari produk, merek, maupun layanan mereka saat R&D [2]. Namun, untuk kasus ini model akan belajar dari dataset yang telah diberikan.

Transportasi online, seperti GoJek, disukai oleh banyak orang karena mempermudah semua kebutuhan, baik untuk bepergian pribadi, pemesanan makanan, pengantaran paket, dst. Transportasi online ini lebih disukai oleh banyak pengguna, terlebih di saat pandemic corona #dirumahaja karena penggunaanya hanya berdiam diri di rumah, sementara kebutuhannya bisa terpenuhi. Twitter adalah situs jejaring sosial populer di Indonesia yang bisa menghasilkan informasi dari tweet pengguna. Dalam studi kasus ini, penulis mengusulkan model yang mendeteksi sentimen publik berdasarkan tweet layanan transportasi online terutama GoJek.

Sistem akan analisis sentimen tweet menggunakan SVM, dan mengelompokkannya menjadi sentimen positif dan negatif.

### Read Twitter Dataset

```
In [2]: # Because this' my Mid Test, My lecturer sent us the dataset from twitter and I used it
tweets = pd.read_csv("data/gojek_twitter_dataset.csv", delimiter=',')
tweets.head()
```

Out[2]:

	tweet	sentimen
0	Saya juga mau vouchee @gojekindonesia https://...	1
1	download gojek duluuu uwuwu	1
2	Aminnn...#orderan goride mhn di lancar kan.all...	1
3	Tq @gojekindonesia @golifeindonesia ❤️. Harusny...	1
4	Semoga Twitter panjang umur. Berkomunikasi den...	1

**Gambar 2.** Proses pemanggilan dataset twitter

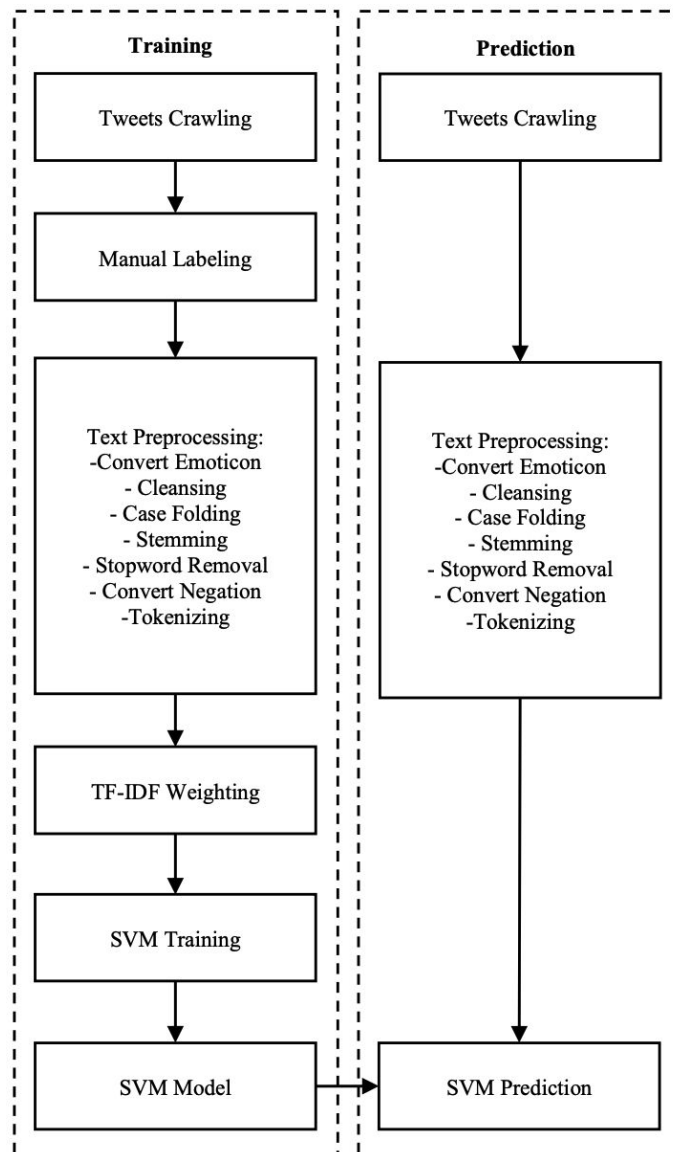
2. Buat kajian tentang algoritma yang menurut Anda optimal untuk menyelesaikan masalah di atas, sertakan referensi paper/jurnal yang mendukung pilihan Anda.

*Jawab :*

Untuk tipe analisis sentimen ini sebenarnya terbagi menjadi dua bagian, yaitu *Coarse-grained sentiment analysis* (pengklasifikasian berorientasi pada sebuah dokumen secara keseluruhan, yakni positif, netral, dan negatif) serta *Fined-grained sentiment analysis* (proses analisis sebuah kalimat). Untuk metode yang digunakan dalam analisis sentimen ini terdapat metode heuristic alias *Rule-based method* (analisis sentimen pada level kata yang berdasarkan rule yang dibuat sebelumnya) serta *Machine learning approach* (analisis sentimen yang menggunakan training data. Keuntungan dari penggunaan *rule-based* terletak pada proses yang lebih mudah saat debugging dan tidak memerlukan proses pembelajaran. Dibalik kemudahan tersebut, model tidak terlalu akurat seperti penggunaan *machine learning*. Keuntungan dari penggunaan *machine learning* adalah kekurangan dari penggunaan *rule-based* serta scalable. Contoh pendekatan *machine learning* yang bisa dilakukan adalah Naive Bayes, Maximum Entropy, Support Vector Machine, maupun Neural Network [3].

Penelitian terkait seperti [4] dalam penelitiannya membuat desain arsitektur penambangan pendapat twitter menggunakan Metode Case Based Reasoning disimpulkan bahwa penambangan opini membutuhkan tahap preprocessing. Hasil dari penelitiannya adalah gambar aplikasi desain arsitektur PT analisis sentimen terhadap merek produk. [5] membandingkan Naïve Bayes dan Support Vector Machine (SVM) metode klasifikasi dengan metode ekstraksi fitur TF-IDF. Metode SVM menghasilkan akurasi kinerja yang lebih baik dibandingkan ke Naïve Bayes, tetapi kedua metode sudah memiliki hasil yang baik untuk klasifikasi tweet [6]. Penelitian oleh [7] telah menjelajahi area penelitian baru yang menerapkan SVM untuk menguji domain set data yang berbeda dari ulasan film, topik dari komputer, hotel atau musik dan juga tentang pendapat dari kamera digital. Banyak penelitian juga telah dilakukan untuk melakukan analisis sentimen dalam konteks multibahasa [8]. Misalnya, klasifikasi sentimen metodologi diterapkan pada posting forum Bahasa Inggris dan Arab. Beragam gaya bahasa Inggris dan Arab atribut

dimasukkan dalam percobaan selain sintaksis fitur. Tujuan utamanya adalah untuk meningkatkan akurasi dan mengidentifikasi fitur utama untuk setiap kelas sentimen [9].



**Gambar 3.** Diagram alur mengenai algoritma yang akan digunakan pada studi kasus ini

3. Lakukan *preprocessing/feature engineering* terhadap data yang ada yang menurut Anda diperlukan.

*Jawab :*

Praproses teks adalah suatu proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan untuk membentuk *golden dataset*, untuk proses mining yang lebih lanjut (sentiment analysis, peringkasan, clustering dokumen) [10]. Selain menyiapkan banyak data untuk training, perlu juga diperhatikan kasus *imbalance dataset* alias

keadaan dimana banyaknya jumlah satu jenis label berbeda sangat signifikan dibandingkan jumlah label lainnya. Dampaknya adalah hasil akurasi yang buruk pada model, oleh karena itu *imbalance dataset* bisa ditangani dengan *resampling*, *weighted modeling*, dst. *Resampling* data adalah metode untuk mengambil sampel yang mewakili data populasinya [11]. Pada model penulis sudah dilakukan resampling data agar meminimalisir jumlah loss atau errornya. *Weighted modeling* adalah pembobotan setiap kata, misalnya saat proses *tf-idf weighting* yang juga telah diimplementasi menggunakan *tf-idf vectorizing*.

Pada studi kasus ini, preprocessing yang digunakan adalah :

- Case Folding, untuk mengubah kata menjadi huruf kecil
- Tokenisasi, untuk memenggal kalimat menjadi kumpulan kata
- Normalization, untuk memperbaiki ejaan maupun merubah hal seperti *url*, *username*, *white space*, serta karakter selain *alphanumeric*/ *\_* / spasi
- Stopword Removal, untuk menghapus kata yang tidak mengandung konten, seperti *url*, *username*, ada, adalah, adanya, dsb.

Preprocessing pada tweet dilakukan untuk memperjelas dataset, tanpa mengurangi makna penting di dalam kata. Kata juga harus diperhatikan dengan baik karena menjadi fitur untuk dataset.

### Tweet Preprocessing

```
In [5]: def process_tweet(twt):
# Convert to lower case
twt = twt.lower()
# Convert www.* or https?:// to URL
twt = re.sub('http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\\(\), ]|(?:%[0-9a-fA-F][0-9a-fA-F]))+', 'URL', twt)
# Convert @username to AT_USER
twt = re.sub('@\S+', 'AT_USER', twt)
# Remove additional white spaces
twt = re.sub('\s', ' ', twt)
# Replace all except alphanumeric, _ and space
twt = re.sub('[^a-zA-Z0-9_ ]', '', twt)
# trim
twt = twt.strip()
return twt

def process_label(label):
if label == 1:
return 1.0
else:
return 0.0

tweets["process_tweet"] = tweets["tweet"].apply(process_tweet)
tweets["process_label"] = tweets["sentimen"].apply(process_label)
tweets.head()
```

Out[5]:

	tweet	sentimen	process_tweet	process_label
0	Saya juga mau vouchee @gojekindonesia https://...	1	saya juga mau vouchee AT_USER URL	1.0
1	download gojek duluuu uwuuu	1	download gojek duluuu uwuuu	1.0
2	Aminnn...#orderan goride mhn di lancar kan.all...	1	aminnnorderan goride mhn di lancar kanall driv...	1.0
3	Tq @gojekindonesia @golifeindonesia ❤️. Harusny...	1	tq AT_USER AT_USER harusnya si bapak ini 50 ...	1.0
4	Semoga Twitter panjang umur. Berkomunikasi deng...	1	semoga twitter panjang umur berkomunikasi deng...	1.0

Gambar 4. Pra pemrosesan tweet menggunakan regex

### Read Stopword List

```
In [4]: def get_stop_word_list(stop_word_list_filename):
stop_words = ['AT_USER', 'URL']
fp = open(stop_word_list_filename, 'r')
for line in fp.readlines():
word = line.strip()
if word == "":
continue
stop_words.append(word)
return stop_words

# Examples are using Bahasa Indonesia
stopwords = get_stop_word_list('data/feature_list/stopwordsID.txt')
stopwords[:10]

Out[4]: ['AT_USER',
'URL',
'ada',
'adalah',
'adanya',
'adapun',
'agak',
'agaknya',
'agar',
'akan']
```

**Gambar 5.** Pra pemrosesan tweet dengan stopwords

	tweet	sentimen	process_tweet	process_label
0	Saya juga mau vouchee @gojekindonesia https://...	1	saya juga mau vouchee AT_USER URL	1.0
1	download gojek dluuu uwuwu	1	download gojek dluuu uwuwu	1.0
2	Aminnn...#orderan goride mhn di lancar kan.all...	1	aminnnorderan goride mhn di lancar kanall driv...	1.0
3	Tq @gojekindonesia @golifeindonesia ❤️. Harusny...	1	tq AT_USER AT_USER harusnya si bapak ini 50 ...	1.0
4	Semoga Twitter panjang umur. Berkomunikasi den...	1	semoga twitter panjang umur berkomunikasi deng...	1.0

**Gambar 6.** Contoh sebelum dan sesudah pra pemrosesan data

4. Tentukan metode untuk ekstraksi fitur dari data yang ada.

*Jawab :*

*Feature Extraction* merupakan pengolahan fitur yang akan meningkatkan performa model, menggali informasi potensial, dan merepresentasikan sebuah sampel asli sebagai vektor fitur yang akan digunakan untuk input metode berikutnya. Pada umumnya, terdapat 3 teknik ekstraksi fitur, yakni *bag of words* (perubahan kata menjadi skalar), *word embedding* (perubahan kata menjadi vektor dengan dimensi tertentu), maupun *character embedding* (perubahan huruf menjadi vektor dengan dimensi tertentu). Pada model Saya, menggunakan tf-idf untuk membentuknya kata dalam vektor skalar. *Term Frequency-Inverse Document Frequency* atau TF — IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan yang dianggap efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus untuk mengetahui seberapa sering suatu kata muncul di dalam dokumen [12].

**Create Matrix using TF-IDF Weighting**

```
In [6]: vectorizer = TfidfVectorizer()
data = vectorizer.fit_transform(tweets["process_tweet"]).toarray()
data

Out[6]: array([[0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               ...,
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.]])
```

**Gambar 6.** Feature extraction menggunakan TF-IDF

Nilai TF-IDF bisa didapatkan dengan rumus :

$$Wi = TF(\omega_i, d) \times IDF(\omega_i)$$

$$IDF(\omega_i) = \log (|D|/DF(\omega_i))$$

dimana  $Wi$  merupakan bobot setiap kata dalam dokumen  $d$ . Lalu  $TF(\omega_i, d)$  adalah term-frequency.  $IDF$  adalah *inverse document frequency* dan  $DF(\omega_i)$  merepresentasikan kemunculan  $\omega_i$  dalam  $D$ . Nilai terbesar  $IDF(\omega_i)$  muncul jika  $\omega_i$  muncul dalam suatu dokumen dan efeknya sangat besar.

5. Lakukan feature selection jika menurut Anda diperlukan.

*Jawab :*

*Feature selection* merupakan metode pemilihan fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam suatu analisa data. Untuk model Saya, tidak ada *feature selection* karena saya membutuhkan semua fitur tersebut untuk dianalisis lebih lanjut.

6. Bangun model klasifikasi yang menurut Anda optimal.

*Jawab :*

Dataset yang terkumpul adalah sejumlah 4000 tweet yang dilabeli dengan 938 tweet positif dan 3062 tweet negatif. Setelah dilakukannya preprocessing data, seperti mengonversi emoji, membersihkan teks dari tanda baca, *case folding*, *stopword*, mengubah negasi, dan kalimat tokenizing ke dalam istilah. Saya melakukan proses stemming sebelum menghentikan penghapusan kata, seperti menjadikan kata "adakah" menjadi kata "ada". Contoh teks pra pemrosesan terdapat pada tabel 1 dan 2.

No	Teks Awal	Teks Preprocessing	Tokenize
----	-----------	--------------------	----------



1.	Saya juga mau vouchee @gojekindonesia https://...	saya juga mau voucher AT_USER URL	Term1 : saya Term2 : juga Term3 : mau Term4 : voucher Term5 : at_user Term6 : url
2.	Semoga Twitter panjang umur. Berkomunikasi	semoga twitter panjang umur berkomunikasi	Term1 : semoga Term2 : twitter Term3 : panjang Term4 : umur Term5 : berkomunikasi

**Tabel 1.** Contoh pra pemrosesan data yang akan digunakan pada model

No	Term	Tf		Df	Idf	Tf-idf	
		T1	T2			T1	T2
1	Enak	2	0	1	0,30102	0,60204	0
2	Dapet	1	1	2	0	0	0
3	Driver	1	1	1	0,30102	0,30102	0,30102
4	Gojek	1	1	2	0	0	0
5	Asyik	1	0	1	0,30102	0,30102	0
6	Ajak	1	0	1	0,30102	0,30102	0
7	Ngobrol	1	0	1	0,30102	0,30102	0
8	Sedih	0	1	1	0,30102	0	0,30102
9	Pesen	0	1	1	0,30102	0	0,30102
10	Xnemu	0	1	1	0,30102	0	0,30102
11	Nemu	0	1	1	0,30102	0	0,30102

**Tabel 2.** Contoh *feature extraction* pada pembobotan TF-IDF yang akan digunakan pada model

Proses *feature extraction* tersebut akan merubah dataset yang awalnya berupa teks menjadi matriks dengan suatu dimensi. Sebelum proses pemodelan, maa akan dilakukan proses *cross validation* untuk memisahkan training dan testing dataset menjadi 2:3.

### Split Training and Testing Dataset

```
In [7]: X_train, X_test, y_train, y_test = train_test_split(data,
                                                    tweets["process_label"],
                                                    test_size=0.33,
                                                    random_state=42)
```

**Gambar 6.** Proses pembagian training dan testing



Setelah prosedur tersebut ditentukan, maka selanjutnya adalah proses pelatihan model dengan menggunakan support vector classification. Harus diketahui sebelumnya cara kerja algoritma support vector machine agar bisa mengerti bagaimana classifier tersebut berjalan pada machine learning. Parameter, optimasi, kemampuan generalisasi dan regularisasi perlu menjadi perhatian dalam proses modelling. Karena sebelumnya diberikan materi mengenai SVM, maka Saya menggunakan model tersebut untuk UTS kali ini.

### Train Machine Learning Model

```
In [8]: model = SVC()
        model.fit(X_train, y_train)

Out[8]: SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
          decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
          max_iter=-1, probability=False, random_state=None, shrinking=True,
          tol=0.001, verbose=False)
```

**Gambar 7.** Proses pelatihan model

7. Lakukan evaluasi atas model yang dihasilkan.

*Jawab :*

Model hasil training akan diuji untuk mengetahui seberapa efektif penggunaan *machine learning* dalam menyelesaikan analisis sentimen pada twitter transportasi online ini. Evaluasi model menggunakan testing dataset untuk mengetahui prediktif model, kemampuan generalisasi model yang terbentuk untuk mengatasi masalah baru. Dalam *supervised learning*, evaluasi model dikenal dengan *confusion matrix* seperti pada tabel 3. *Confusion matrix* atau juga disebut matriks eror adalah sebuah tabel spesifik yang memberikan gambaran performa dari suatu algoritma atau model yang digunakan [13]. *Confusion matrix* digunakan sebagai indikasi sifat dari sebuah aturan klasifikasi dimana di dalamnya terdapat beberapa nilai yang menunjukkan banyaknya amatan yang telah terprediksi dengan benar ataupun yang tidak [14]. Jika hanya menggunakan presisi atau akurasi saja sebagai ukuran evaluasi performa dari sebuah model yang dibentuk, akan berpotensi untuk menghasilkan kesalahan yang sangat fatal, seperti saat menghadapi kasus *Imbalance Dataset*.

### Check Accuraction against Test Set

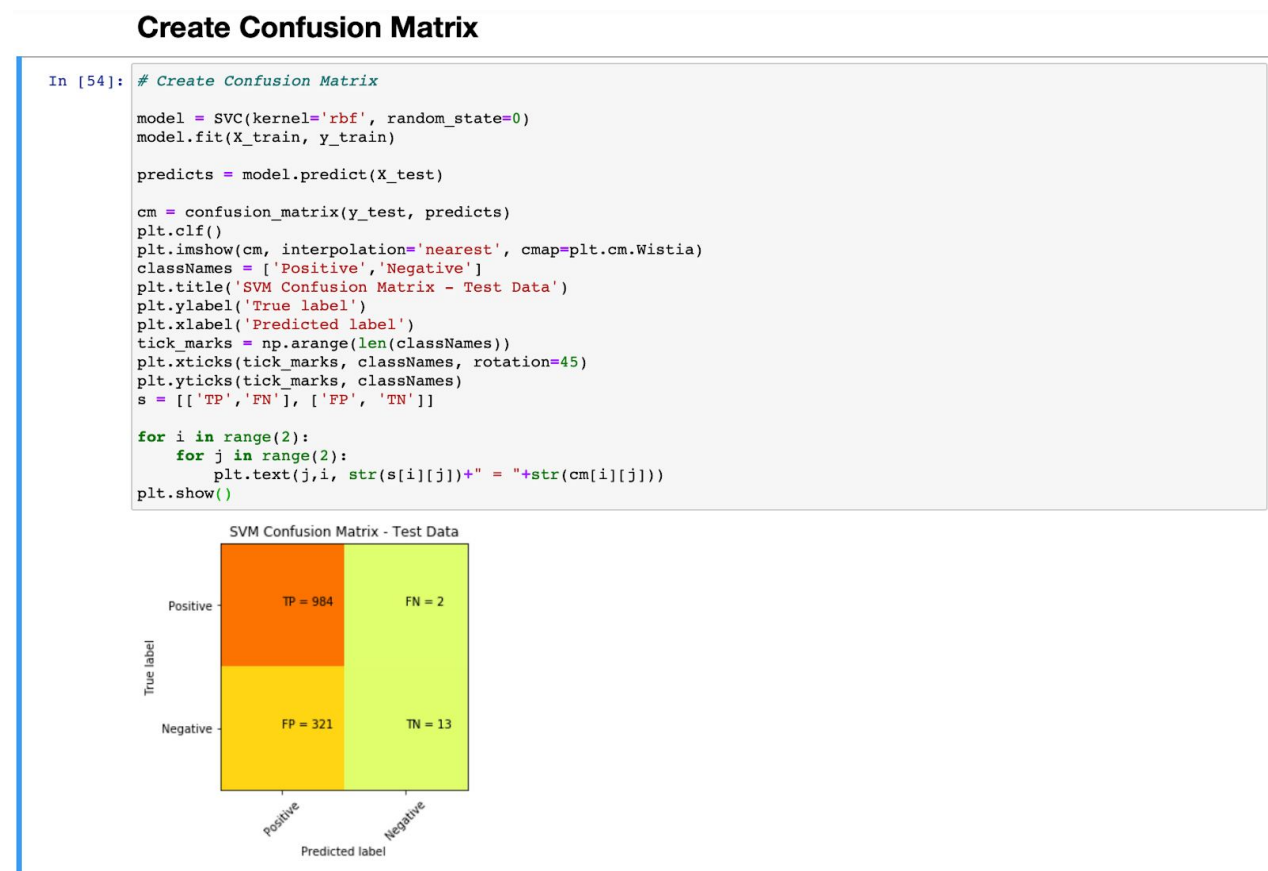
```
In [11]: test_data = "Baru pesan gofood tapi rasanya tidak enak :("
        test_data = process_tweet(test_data)
        test_data = vectorizer.transform(test_data).toarray()
        predict = model.predict(test_data)
        if predict == 0:
            print("negatif")
        else:
            print("positif")

negatif
```

**Gambar 8.** Percobaan untuk memprediksi sentimen tweet baru

	Prediction (Positive)	Prediction (Negative)	Total
Actual (Positive)	TP 984	FN 2	986
Actual (Negative)	FP 321	TN 13	334
Total	1305	15	N=1320

**Tabel 3.** Tabel untuk *confusion matrix*



**Gambar 9.** Evaluasi model menggunakan *confusion matrix*

## Conclusion

```

In [56]: # Accuracy
accuracy = (TP + TN) / (TP + FN + FP + TN)
print('The accuracy of this model:', accuracy)

The accuracy of this model: 0.7553030303030303

In [57]: # Misprediction Rate
mispredict = (FP + FN) / (TP + FN + FP + TN)
print('The mispredict rate of this model:', mispredict)

The mispredict rate of this model: 0.2446969696969697

In [58]: # Recall (Positive)
rp = FP / (TP + FN)
print('The recall positive rate of this model:', rp)

The recall positive rate of this model: 0.3255578093306288

In [59]: # Recall (Negative)
rn = TN / (FP + TN)
print('The recall positive rate of this model:', rn)

The recall positive rate of this model: 0.038922155688622756

In [60]: # Precision (Positive)
pp = TP / (TP + FP)
print('The precision positive rate of this model:', pp)

The precision positive rate of this model: 0.7540229885057471

In [61]: # Precision (Negative)
pn = TN / (FN + TN)
print('The precision positive rate of this model:', pn)

The precision positive rate of this model: 0.8666666666666667

```

**Gambar 10.** Nilai untuk model dengan akurasi sebesar 75.53% yang bisa dikatakan cukup bagus

Jutaan pengguna Twitter dapat mengutarakan pendapat mereka melalui tweet mereka. Perusahaan bisa mengolah informasi ini untuk riset terhadap produk, layanan, maupun merek mereka, namun akan banyak waktu yang dibutuhkan. Oleh karena itu, diperlukan analisis sentimen yang memperkirakan sentimen tweet terhadap suatu topik. Contohnya adalah analisis sentimen terhadap transportasi online, Gojek. Pemrosesan teks yang diperoleh dari tweet untuk menghapus *noise*, maka digunakan n-gram unigram serta TF-IDF sebagai fitur metode ekstraksi dan algoritma SVM untuk klasifikasi metode. Dari hasil pengujian diperoleh akurasi 75.53%, tingkat kesalahan prediksi 24.47% (yang disebabkan oleh *raw data* yang memang kurang bersih) [15].

Code : [https://github.com/dhifaaans/sentiment\\_analysis](https://github.com/dhifaaans/sentiment_analysis)

**Daftar Pustaka :**

- [1] <https://www.youtube.com/watch?v=3Pzni2yfGUQ>
- [2] <https://medium.com/@irvanseptiar/introduction-sentiment-analysis-mudah-5785f88e435d>
- [3] <https://mamat.co/mengenal-sentiment-analysis/>
- [4] Aribowo, A. S., 2015, Arsitektur Aplikasi Twitter Opinion Mining Untuk Mengetahui Sentimen Publik Terhadap Merek, pp. 14–20.
- [5] Hidayatullah, A. F. dan SN, A. Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter, pp. 1–8. 2014.
- [6] Windasari, I. P., Uzzi, F. N., dan Satoto, K. I., 2017, Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek, 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE), Oct 18-19, 2017, pp. 266 - 269.
- [7] Saleh, M. R., Mart'in-Valdivia, M., Montejo-Raez, A., and Lopez, L. U., 2011, Experiments with svm to classify opinions in different domains, Expert Systems with Applications, vol. 38, no. 12, pp. 799– 804.
- [8] Zainuddin, N. dan Selamat, A., 2014, Sentiment Analysis Using Support Vector Machine, IEEE 2014 International Conference on Computer, Communication, and Control Technology (I4CT 2014), September 2 - 4, 2014 - Langkawi, Kedah, Malaysia, pp. 333-337
- [9] Abbasi, A., Chen, H., dan Salem, A., 2008, Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums, ACM Trans. Inf. Syst., vol. 26, no. 3, pp. 1–34.
- [10] <http://malifauzi.lecture.ub.ac.id/files/2016/02/Text-Pre-Processing.pdf>
- [11] <https://www.statisticssolutions.com/sample-size-calculation-and-sample-size-justification-resampling/>
- [12] <https://medium.com/@dltsierra/algoritma-tf-idf-633e17d10a80>
- [13] Sammut, C., dan Webb, G. I., 2011, Encyclopedia of Machine Learning, New York: Springer.
- [14] Rokach, L., dan Maimon, O., 2008, Data Mining with Decision Trees Theory and Applications, Singapura: World Scientific Publishing Co.Pte.Ltd.
- [15] [https://github.com/dhifaaans/sentiment\\_analysis](https://github.com/dhifaaans/sentiment_analysis)