

# ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP MRT JAKARTA MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES

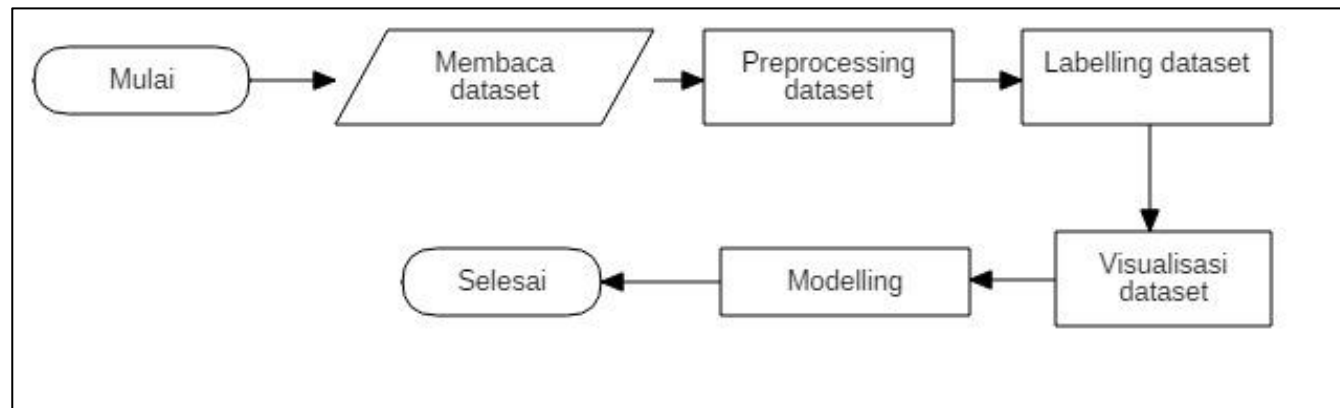


Nama : Ario Waskita  
NPM : 51415011  
Dosen Pembimbing : Prof. Dr. Da Silva

MRT Jakarta merupakan moda transportasi yang dapat digunakan oleh publik. Melalui media sosial Twitter, publik dapat memberikan opini mereka terkait MRT Jakarta. Opini dapat bersentimen positif atau negatif sehingga dilakukan analisis sentimen.

Implementasi metode klasifikasi Multinomial Naive Bayes dalam melakukan analisis sentimen mengenai moda transportasi MRT Jakarta berdasarkan cuitan pengguna sosial media Twitter dengan menggunakan bahasa pemrograman Python.

# Rangkaian Alur Proses



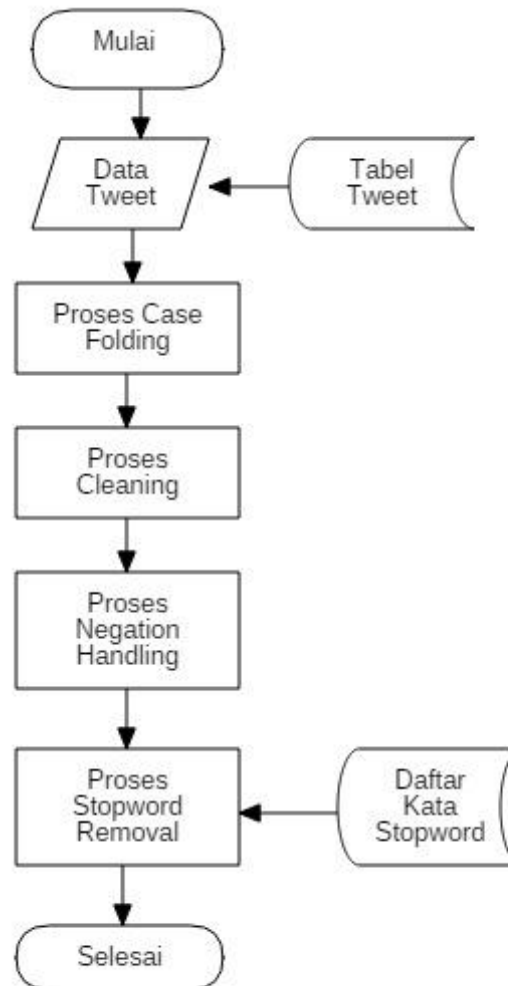
- Data tweet mengenai MRT Jakarta diambil dari sosial media Twitter.
- Data tweet dikumpulkan menggunakan keyword “@mrtjakarta”, “naik mrt jakarta”, dan “#mrtjakarta”.

# Contoh Data Tweet

No	<i>Tweet</i>
1	Naik @mrtjakarta moda transportasi modern keren deh pokok nya ????? <a href="https://t.co/MS9Ws4pJDo">https://t.co/MS9Ws4pJDo</a>
2	@mrtjakarta tolong disediakan atm dong di stasiun....susah bgt niy mau narik uang cash. ??
3	@upena @mrtjakarta Semenjak ada dirimu. Dunia terasa indahny.
4	@dondiindrayana @mrtjakarta @DKIJakarta Memang baunya sangat menyengat dan membuat sesak napas
5	Yaa Rabb, enak bat ngerasain Tugu Tani ke Lebak Bulus < 1 jam @PT_TransJakarta @mrtjakarta

**Sumber: [www.twitter.com](http://www.twitter.com)**

# Preprocessing Dataset



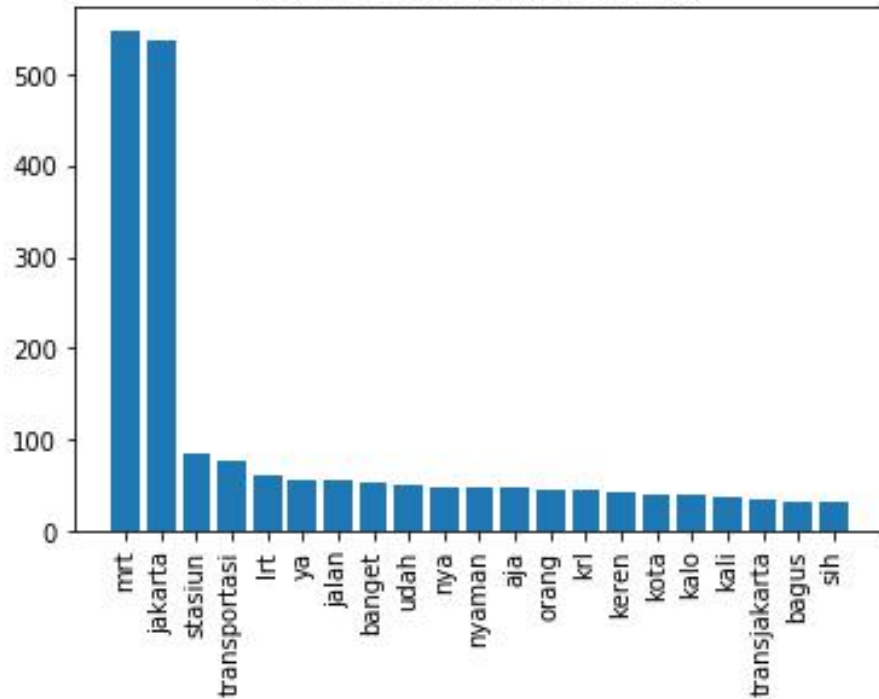
- Menggunakan pendekatan lexicon dengan menggunakan kamus lexicon positif dan negatif.
- Menentukan apakah data tweet bersentimen positif atau negatif berdasarkan kamus lexicon.



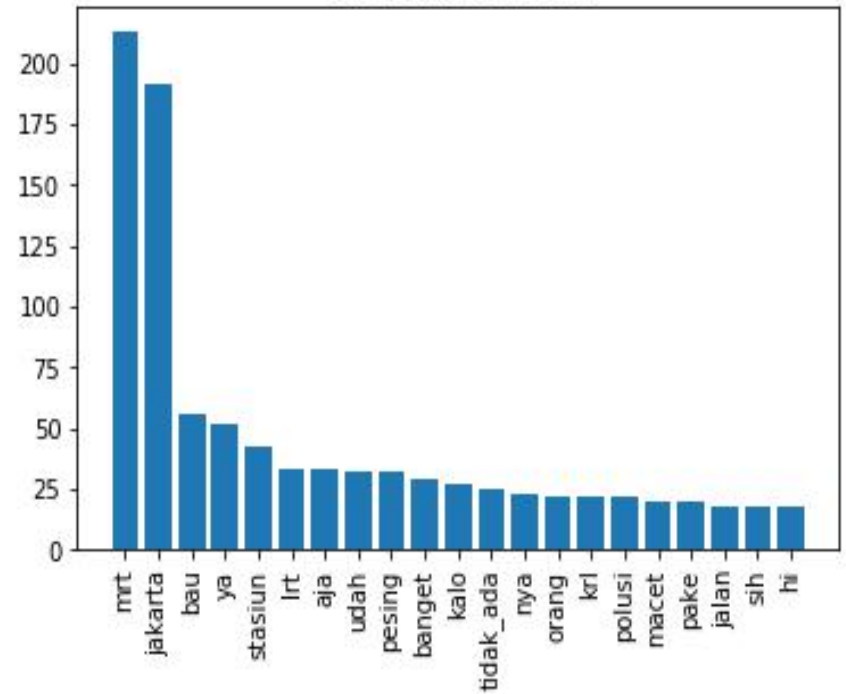
# Contoh Hasil Labelling

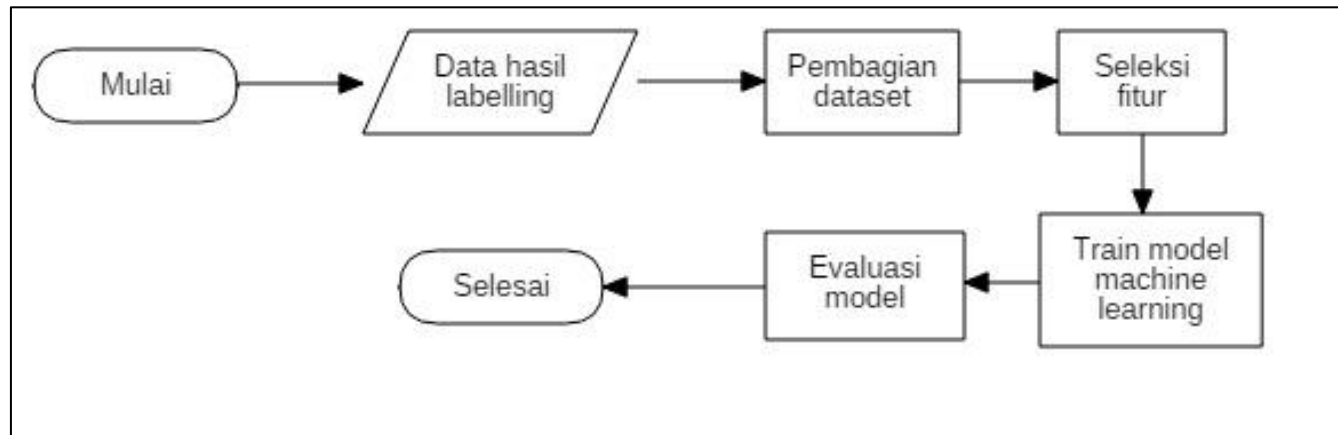
Tweet	Hasil Preprocessing	Kata Positif	Kata Negatif	Hasil Labelling
@upena @mrtjakarta Semenjak ada dirimu. Dunia terasa indah nya.	semenjak ada dirimu dunia indah nya	1	0	positif
@dondiindrayana @mrtjakarta @DKIJakarta Memang baunya sangat menyengat dan membuat sesak napas	baunya menyengat sesak napas	0	1	negatif

Top 20 kata pada kelas positif



Top 20 Kata Negatif





Index	Hasil Preprocessing
1	mrt jakarta nyaman
2	toilet mrt jakarta bau pesing

Index	mrt	jakarta	nyaman	toilet	bau	pesing
1	1	1	1	0	0	0
2	1	1	0	1	1	1

# Pengaruh Fitur terhadap Model Klasifikasi

Parameter Min_df	Jumlah Fitur	Test Accuracy	Average Precision	Average Recall
1	3201	80%	84%	74%
2	1112	82%	85%	77%
3	670	83%	85%	77%
4	454	80%	81%	75%
5	340	81%	82%	76%

```
import joblib

vect = CountVectorizer(min_df=3)
X_train_vect = vect.fit_transform(X_train).toarray()
model = naive_bayes.MultinomialNB()
model.fit(X_train_vect, y_train)
filename = 'finalized_model.pkl'
joblib.dump(model, filename)
```

Berdasarkan hasil dari uji coba model Multinomial Naive Bayes Classifier terhadap data testing diperoleh confusion matrix seperti pada tabel

Hasil Labelling	Prediksi Model	
	Negatif	Positif
Negatif	True Negative (TN) = 34	False Positive (FP) = 24
Positif	False Negative (FN) = 4	True Positive (TP) = 102

Accuracy = 82,9%

Average Recall = 77,3%

Average Precision 85,2%

Data	Data Tweet Baru
1	perdana naik mrt jakarta.... bener bener seneng karena akhirnya indonesia punya kayak begini (walaupun baru di ibu kota, it's okay progress takes time)
2	selamat sore min, kenapa ya jalan ke arah park&ride lebak bulus bau pesing banget. mengganggu sekali min, harap diperhatikan dan cari solusi @mrtjakarta"

Data	Probabilitas Positif	Probabilitas Negatif	Hasil Prediksi Model	Pendapat Subjektif
1	0.999864	0.000136	Positif	Positif
2	0.001090	0.998910	Negatif	Negatif



- Metode Multinomial Naive Bayes dapat digunakan untuk klasifikasi sentimen data tweet yang bersumber dari sosial media Twitter.
- Fitur yang digunakan untuk melatih model klasifikasi Multinomial Naive Bayes mempengaruhi performa dari model klasifikasi.
- Model klasifikasi memiliki akurasi sebesar 82,9% dengan komposisi data training berjumlah 653 dan data testing berjumlah 164.

- Pada penelitian selanjutnya dapat menggunakan data yang bersumber dari sosial media lain seperti Facebook dan Instagram.
- Bahasa yang digunakan dapat dikembangkan agar tidak hanya untuk Bahasa Indonesia, tetapi juga dapat menggunakan bahasa lain seperti Bahasa Inggris.
- Pada penelitian berikutnya dapat menggunakan metode lain untuk melakukan analisis sentimen sehingga dapat membandingkan metode mana yang lebih baik untuk proses analisis sentimen.

# TERIMA KASIH