

Analisis Sentiment Twitter pada Kasus Pilkada DKI 2017

Dataset diperoleh dari <https://github.com/rizalespe/Dataset-Sentimen-Analisis-Bahasa-Indonesia.git>

1. Hasil Evaluasi

Hasil split data 0,2:

```
Confusion Matrix:
[[66 13]
 [18 43]]
Report:
              precision    recall  f1-score   support

   negative     0.7857     0.8354     0.8098         79
   positive     0.7679     0.7049     0.7350         61

   accuracy                   0.7786         140
  macro avg     0.7768     0.7702     0.7724         140
 weighted avg     0.7779     0.7786     0.7772         140

CPU times: user 32.8 ms, sys: 2.99 ms, total: 35.8 ms
Wall time: 57.1 ms
```

Hasil Split data 0,3:

```
Confusion Matrix:
[[100  18]
 [ 42  50]]
Report:
              precision    recall  f1-score   support

   negative     0.7042     0.8475     0.7692        118
   positive     0.7353     0.5435     0.6250         92

   accuracy                   0.7143        210
  macro avg     0.7198     0.6955     0.6971        210
 weighted avg     0.7178     0.7143     0.7060        210

CPU times: user 37.1 ms, sys: 996 µs, total: 38.1 ms
Wall time: 37.5 ms
```

	Stemming	
Evaluasi	split data 0,2	split data 0,3
Precision	77,68	71,98
Recall	77,02	69,55
f1-score	77,24	69,71
wall time (ms)	57,1	30,4

2. Analisis Evaluasi

Hasil evaluasi diukur dengan menggunakan Precisiion, Recall, dan F-1 Score berdasarkan hasil rata-rata dari setiap label.

a. Split data 0,2

Precision terbaik adalah **77,68** yang diperoleh dari percobaan persentase split data sebesar 80%. Hal ini berarti bahwa ketepatan dari model yang diprediksi dari hasil rata-rata label adalah 77,68 dengan rincian bahwa precision dari label **negative** sebesar **78,57** dan label **positive** sebesar **76,79**. Sedangkan kemampuan model untuk menemukan semua kasus yang relevan dalam dataset atau disebut **recall** rata-rata yang diperoleh adalah **77,02**. Hasil tersebut merupakan hasil dari rata-rata label dimana label **negative** memperoleh recall sebesar **83,54** dan label positive sebesar **70,49**. Nilai **precision** yang **tinggi** tetapi **recall rendah** karena pada dataset masih mempunyai nilai **false negative** yang tinggi atau pada rata-rata label sentiment adalah diprediksi bersentimen positive tetapi ternyata sebenarnya adalah bersentimen negative untuk label negative dan untuk label positive diprediksi bersentimen negative tetapi ternyata bersentimen positive. Sehingga dapat disimpulkan bahwa nilai precision berbanding terbalik dengan dengan nilai recall. Akan tetapi, hal tersebut dapat ditemukan titik tengahnya dengan menghitung nilai **F-1 score**. Nilai F-1 score atau nilai rata-rata dari precision dan recall yang dapat diartikan sebagai kelengkapan dan ketepatan dari model dilihat berdasarkan rata-rata label untuk eksperimen pertama yaitu **77,24**.

b. Split data 0,3

Hasil ketepatan dari model yang diprediksi dari hasil rata-rata label adalah **71,98** dengan rincian bahwa precision dari label **negative** sebesar **70,42** dan label **positive** sebesar **73,53**. Sedangkan kemampuan model untuk menemukan semua kasus yang relevan dalam dataset atau disebut **recall** rata-rata yang diperoleh adalah **69,55**. Hasil tersebut merupakan hasil dari rata-rata label dimana label **negative** memperoleh recall sebesar **84,75** dan label positive sebesar **54,35**. Nilai **precision** yang **tinggi** tetapi **recall rendah** karena pada dataset masih mempunyai nilai **false negative** yang tinggi atau pada rata-rata label sentiment adalah diprediksi bersentimen positive tetapi ternyata sebenarnya adalah bersentimen negative untuk label negative begitupun sebaliknya. Sehingga dapat disimpulkan bahwa nilai precision berbanding terbalik dengan dengan nilai recall. Akan tetapi, hal tersebut dapat ditemukan titik tengahnya dengan menghitung nilai **F-1 score**. Nilai F-1 score atau nilai rata-rata dari

precision dan recall yang dapat diartikan sebagai kelengkapan dan ketepatan dari model dilihat berdasarkan rata-rata label untuk eksperimen pertama yaitu **69,71**.

3. Analisis eksperimen

a. Pengaruh persentase split data

Berdasarkan gambar di atas membuktikan bahwa percobaan dengan split dataset sebesar 80% membuat model menghasilkan performansi yang maksimal dibandingkan split dataset sebesar 70%. Karena hasil performansi diukur dari rata-rata label dimana label yang mempunyai jumlah data uji/data test yang banyak memiliki nilai performansi yang tinggi..

b. Waktu pelatihan

Waktu pelatihan dipengaruhi dari pembagian data uji/ data test dengan data latih/ data train. Berdasarkan hasil evaluasi, pembagian data uji sebesar 0,3 memiliki waktu pelatihan yang cukup rendah. Sehingga dapat disimpulkan bahwa semakin banyak data train/data latih maka semakin tinggi pula waktu pelatihannya

c. Prediksi data yang salah

a. Split data 0,2

	Actual Result	Naive Bayes
457	positive	negative
680	negative	positive
21	negative	positive
472	negative	positive
504	positive	negative
472	positive	negative
504	positive	negative
506	positive	negative
100	negative	positive
347	negative	positive
575	positive	negative
243	positive	negative
532	positive	negative
173	positive	negative
482	positive	negative
578	positive	negative
103	negative	positive
418	negative	positive
240	negative	positive
251	positive	negative
345	negative	positive
212	negative	positive
69	positive	negative
26	positive	negative
512	positive	negative
166	positive	negative
9	negative	positive
82	negative	positive
128	negative	positive

b. Split data 0,3

	Actual Result	Naive Bayes
235	positive	negative
523	positive	negative
526	positive	negative
18	negative	positive
221	positive	negative
510	positive	negative
517	positive	negative
297	positive	negative
262	positive	negative
462	positive	negative
521	positive	negative
157	positive	negative
122	negative	positive
403	negative	positive
470	positive	negative
341	negative	positive
546	positive	negative
171	positive	negative
248	positive	negative
13	negative	positive
238	positive	negative
202	positive	negative
535	positive	negative
338	negative	positive
426	negative	positive
176	positive	negative
29	negative	positive
265	positive	negative
377	negative	positive
153	positive	negative
100	negative	positive
177	positive	negative
482	positive	negative
483	positive	negative
692	negative	positive
578	positive	negative
573	positive	negative
148	negative	positive

472	positive	negative
346	negative	positive
158	positive	negative
542	positive	negative
6	negative	positive

336	negative	positive
259	positive	negative
21	negative	positive
519	positive	negative
234	positive	negative
22	negative	positive
180	positive	negative
263	positive	negative
556	positive	negative
207	positive	negative
293	positive	negative

Prediksi data yang salah dipengaruhi oleh 3 hal, yaitu metode yang digunakan, hasil precision, dan banyaknya split data yang dilakukan. Rata-rata precision atau ketepatan data yang didapat pada split data 0,2 lebih besar daripada hasil rata-rata precision pada split data 0,3. Hal tersebut berarti kesalahan data pada split data 0,2 lebih kecil daripada split data 0,3. Selain itu, metode Multinomial Naïve Bayes juga mempengaruhi hasil prediksi data yang diperoleh pada kedua eksperimen karena Multinomial Naïve Bayes digunakan untuk menghitung jumlah kemunculan katanya saja. Oleh karena itu, pemilihan metode untuk pemodelan akan mempengaruhi hasil evaluasi dan prediksi data yang salah.