

Interpretasi Sentimen

- Sumber data**

Data yang digunakan untuk analisis sentimen ini diambil dari twitter menggunakan kata kunci ”#alkohol” dengan jumlah data 500 dan keadaan data seperti berikut.

Tweet
IgorPattyAnastasiaZebadiah #赤峰 #约 #炮
Introvert sih introvert, tapi ga 96% introvert juga michy 🤔
Beberapa studi juga menunjukkan bahwa konsumsi #Alkoh...
PR#16:1-6&20&27&32-33/INSTFRV;BFR&AF...
DailyUp5 \$ADBE \$352.12 5.55% 4.06% -6.10% 5.88% 54.53 \$...
RSI25 \$PACW \$9.96 -12.40% -49.92% -64.45% -59.11% 20.8...
RSI25 \$HRL \$39.06 -0.95% -1.65% -13.90% -18.02% 19.70 \$L...
RSI25 \$FPI \$9.70 -1.57% -5.78% -23.60% -25.59% 22.69 \$FR...
RSI25 \$CPRI \$41.92 1.27% -5.21% -18.13% -28.80% 24.96 \$...

- Pre processing data**

Data yang diambil dari twitter masih berisi karakter yang beragam, seperti emoticon, link, angka, tanda baca dan huruf yang besar kecilnya tidak konsisten. Maka dari itu dilakukan pre processing data untuk mengubah data yang tidak terstruktur tadi menjadi lebih terstruktur dan rapi seperti berikut ini.

text	StringAsFactors
igorpattyanastasiazebadiah	FALSE
introvert sih introvert tapi ga introvert juga michy	FALSE
beberapa studi juga menunjukkan bahwa konsumsi dlm jan...	FALSE
pr	FALSE
dailyup adbe amc amd anet api dbi	FALSE
rsi pacw pdm prch pru ptve	FALSE
rsi hrl lc inc lumn mac	FALSE
rsi fpi frc fubo gps grpn	FALSE
rsi cpri cvs cwk dea dnb	FALSE
rsi aap abr alk bac bdn	FALSE
pamelaruthemmalewis	FALSE

- **Tokenizing**

Proses tokenizing ini bertujuan memisahkan kata dalam suatu kalimat agar proses analisis teks bisa berlanjut. Berikut hasil dari tokenizing.

```
igorpattyanastasiazebadiah
"igorpattyanastasiazebadiah"
introvert sih introvert tapi ga introvert juga michy
"introvert sih introvert tapi ga introvert juga michi"
beberapa studi juga menunjukkan bahwa konsumsi dlm jangka waktu panjang bisa membuat kadar hormon testosteron menurun
"beberapa studi juga menunjukkan bahwa konsumsi dlm jangka waktu panjang bisa membuat kadar hormon testosteron menurun"
```

- **Stemming**

Proses stemming bertujuan untuk mendapat kata dasar sehingga bisa menghilangkan imbuhan pada setiap kata dan memudahkan proses analisis

- **Word cloud**

Word cloud bertujuan menampilkan data teks dalam bentuk visual yang mudah dibaca dan dipahami, proses word cloud ini akan ditemykan kata-kata yang paling penting atau sering muncul dalam teks yang dianalisis, berikut hasil word cloud nya



Dari hasil word cloud terlihat bahwa kata “angka” dan “prediksi” adalah kata yang sering muncul

- **Analisis sentimen (label data)**

Proses labeling data ini merupakan tahap penting dalam analisis sentiment, proses ini mengevaluasi sebuah data dengan tujuan untuk mencari tahu polaritas dari sentimen positif, negatif dan netral, hasil labeling data nya sebagai berikut.

[1]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[9]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[17]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[25]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[33]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[41]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[49]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Positive"	"Neutral"	"Neutral"	"Neutral"
[57]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[65]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[73]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Positive"	"Neutral"	"Neutral"	"Neutral"
[81]	"Neutral"	"Neutral"	"Neutral"	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[89]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Negative"	"Neutral"
[97]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[105]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[113]	"Neutral"	"Neutral"	"Negative"	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[121]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[129]	"Neutral"	"Neutral"	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[137]	"Neutral"	"Neutral"	"Positive"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[145]	"Positive"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Negative"	"Neutral"	"Neutral"
[153]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[161]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[169]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Negative"
[177]	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[185]	"Neutral"	"Neutral"	"Positive"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[193]	"Negative"	"Positive"	"Positive"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[201]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[209]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[217]	"Neutral"	"Neutral"	"Neutral"	"Negative"	"Negative"	"Positive"	"Neutral"	"Positive"
[225]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[233]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[241]	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[249]	"Positive"	"Negative"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Negative"
[257]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Negative"	"Neutral"
[265]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"
[273]	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Neutral"	"Positive"	"Neutral"

Label negatif didapatkan jika sentimen skor bernilai kurang dari 0.3, label netral diperoleh jika sentimen skor bernilai $-0.3 < \text{sentimen skor} < 0.3$ dan label positif diperoleh jika sentimen skor bernilai lebih dari 0.3

- **Klasifikasi sentimen (naive bayes)**

Analisis sentimen naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Berikut hasil analisis sentimen dengan naive bayes

Confusion Matrix and Statistics			
prediksi	Negative	Neutral	Positive
Negative	5	0	0
Neutral	0	136	9
Positive	0	0	0
Overall Statistics			
Accuracy : 0.94			
95% CI : (0.8892, 0.9722)			
No Information Rate : 0.9067			
P-Value [Acc > NIR] : 0.09809			
Kappa : 0.51			
McNemar's Test P-Value : NA			
Statistics by Class:			
	Class: Negative	Class: Neutral	Class: Positive
Sensitivity	1.00000	1.0000	0.00
Specificity	1.00000	0.3571	1.00
Pos Pred Value	1.00000	0.9379	NaN
Neg Pred Value	1.00000	1.0000	0.94
Prevalence	0.03333	0.9067	0.06
Detection Rate	0.03333	0.9067	0.00
Detection Prevalence	0.03333	0.9667	0.00
Balanced Accuracy	1.00000	0.6786	0.50

Hasil analisis sentimen diketahui bahwa 5 sampel diklasifikasikan sebagai tweet bersentimen negatif, 136 bersifat netral, 9 bersifat netral positif dan 0 bersifat positif. Akurasi model yang didapatkan sebesar 0.94 atau 94% yang artinya 94% prediksi yang dilakukan model adalah benar. No information rate sebesar 0.9067 yang artinya jika hanya menebak kelas berdasarkan distribusi kelas asli, nilai akurasi sebesar 90.6%. nilai kappa sebesar 0.51 menunjukkan tingkat kesepakatan yang cukup tinggi, semakin tinggi nilai kappa maka dapat dipertimbangkan sebagai performa yang memiliki kinerja bagus. Nilai sensitivitas pada kelas negatif dan netral sebesar 1.00 menunjukkan model sangat akurat dalam mengklasifikasikan sampel, sedangkan nilai sensitivitas kelas positif sebesar 0.00 menunjukkan model kurang akurat dalam mengklasifikasikan sampel. Balanced accuracy menunjukkan bahwa model memiliki keseimbangan yang cukup baik dalam mengklasifikasikan moodel negatif, netral dan positif.