

Nama : Kharisma Aufa Ramadhani
NIM : A11.2019.12061
Kelas : A11.4619

Ringkasan Eksperimen Tugas Data Mining

- **Judul**

“PENERAPAN ALGORITMA NAÏVE BAYES DALAM ANALYSIS TWEET
MENGENAI CYBER BULLYING”

- **Latar Belakang Masalah**

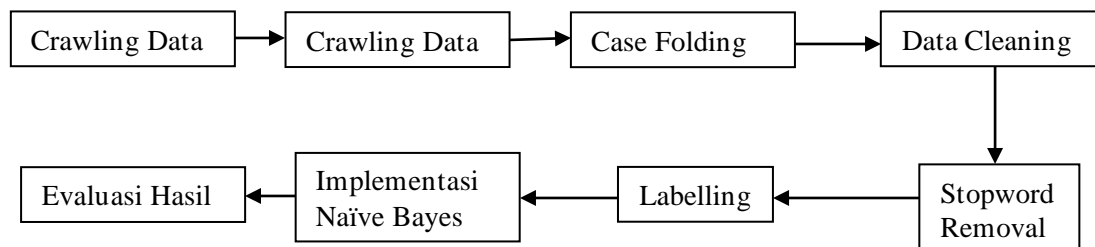
Media sosial Twitter adalah salah satu sebagai media penghubung yang diminati oleh seluruh masyarakat di dunia. Pengguna Twitter akan memberikan kabar terbaru atau komentar tentang hal yang sedang menjadi topik utama di dunia. Hal yang sedang menjadi topik utama dan banyak dan sering dikomentari oleh pengguna akan menimbulkan suatu masalah atau trending topik di media sosial terutama Twitter. Pertumbuhan Twitter terus meningkat setiap waktu sehingga dimanfaatkan oleh pengguna Twitter untuk menyampaikan informasi berupa kritik maupun saran.

Pada penelitian ini dilakukan klasifikasi tweet pengguna Twitter ke dalam dua sentimen, yaitu sentimen positif dan sentimen negatif. Pengklasifikasian dilakukan dengan metode Multinomial Naïve Bayes. Klasifikasi Multinomial Naïve Bayes merupakan pengembangan model dari algoritma bayes yang biasa digunakan dalam klasifikasi teks. Naive Bayes Classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Proses yang dijalankan juga mencakup preprocessing task sebagai pengolahan data mentah menjadi data siap pakai dan

confusion matrix sebagai perhitungan akurasi dari aplikasi yang dihasilkan. Dan data dari hasil penelitian ini bisa dimanfaatkan sebagai bahan pertimbangan kembali dalam mengambil keputusan.

- **Metode Penyelesaian**

Penelitian ini dilakukan dengan melewati beberapa tahap yaitu diantaranya :



- **Tahapan Penyelesaian**

- a) **Crawling Data**

Saat ini dataset sudah diperoleh dengan menggunakan library tweepy, library ini merupakan library yang dibuat untuk scrapping tweet di Twitter. Setelah di scrapping data tweet kemudian akan dimasukkan ke dalam dataframe. Dataframe tersebut mempunyai kolom / atribut text dan username. Untuk format dataset ini yaitu .csv jumlah data di dalam dataset ini sekitar 100 data, untuk dataset ini bersifat public.

	Username	Text
0	Just_Ntokozo	RT @PeruMorkels: Guys jokes aside, bad matric ...
1	FWillia02626462	RT @jackie06113586: @pritipatel You are one cr...
2	domdyer70	RT @scanme_tukslaw: 1 year old XL bully booked...
3	HolgerBartels6	RT @mhmck: Actively preventing other countries...
4	Angeeeeeeee	@KProtein19 @nytimes Wow!!!! No one cares abou...
5	IDKReformer7	RT @uzoakuchie: People ndi strategists are try...
6	benjoshsmith1	RT @lindamagTV: More of my old @hill_grange pi...
7	NishaS40654771	@latestly @akshaykumar @arjunk26 @iHrithik The...
8	Smiling_Abbey	RT @BashWaziri: @Qdpaper2 You guys need to do ...
9	Lawlietern	Doing some speedrun with Bully\nThis is my fir...
10	Shikonboss	encouragement and recognise the need for help...

b) Case Folding

Setelah melalui tahap *Crawling Data*, kemudian melakukan tahap *case folding*, dimana akan mengubah semua huruf menjadi huruf *lower case*.

	username	text
0	just_ntokozo	rt @perumorkels: guys jokes aside, bad matric ...
1	fwillia02626462	rt @jackie06113586: @pritipatel you are one cr...
2	domdyer70	rt @scanme_tukslaw: 1 year old xl bully booked...
3	holgerbartels6	rt @mhmck: actively preventing other countries...
4	angeeeeeeee	@kprotein19 @nytimes wow!!!! no one cares abou...
5	idkreformer7	rt @uzoakuchie: people ndi strategists are try...
6	benjoshsmith1	rt @lindamagtv: more of my old @hill_grange pi...
7	nishas40654771	@latestly @akshaykumar @arjunk26 @ihrithik the...
8	smiling_abbey	rt @bashwaziri: @qdpaper2 you guys need to do ...
9	lawlietern	doing some speedrun with bully\nthis is my fir...
10	shikonboss	encouragement and recognise the need for help....

c) Data Cleaning

Setelah melalui tahap Case Folding, kemudian melakukan tahapan Data Cleaning dimana akan menghapus semua data – data yang dianggap tidak perlu seperti simbol @ pada username dst.

	username	text
0	justntokozo	rt guys jokes aside bad matric results don t m...
1	fwillia02626462	rt you are one cruel inhumane individual who i...
2	domdyer70	rt tukslaw 1 year old xl bully booked to be eu...
3	holgerbartels6	rt actively preventing other countries from pa...
4	angeeeeeeee	wow no one cares about our country of overthro...
5	idkreformer7	rt people ndi strategists are trying to bully ...
6	benjoshsmith1	rt more of my old grange pics have popped up t...
7	nishas40654771	these bully actors have not satisfied us now t...
8	smilingabbey	rt you guys need to do better you can t let ot...
9	lawlietern	doing some speedrun with bully this is my fir...
10	shikonboss	encouragement and recognise the need for help ...

d) Stopword Removal

Setelah melakukan tahapan *Data Cleaning*, kemudian melakukan tahapan *Stopword Removal* dimana akan melakukan pemilihan kata – kata yang penting di *Dataframe* ini.

	username	text
0	justntokozo	rt guys jokes aside bad matric results mean ja...
1	fwillia02626462	rt one cruel inhumane individual known bully n...
2	domdyer70	rt tukslaw 1 year old xl bully booked euthanis...
3	holgerbartels6	rt actively preventing countries participating...
4	angeeeeeeee	wow one cares country overthrow rouge presiden...
5	idkreformer7	rt people ndi strategists trying bully stood g...
6	benjoshsmith1	rt old grange pics popped morning character su...
7	nishas40654771	bully actors satisfied us destroying
8	smilingabbey	rt guys need better let others blackmail bully...
9	lawlietern	speedrun bully first ever speedrun gonna later...
10	shikonboss	encouragement recognise need help shoot mouth ...

e) Labelling

Labelling merupakan teknik untuk menentukan apakah tweet ini bernilai positif atau negatif. Saat nilai kurang dari 0 bernilai negatif dan saat nilai lebih dari 0.05 bernilai positif. Dalam penentuan kalimat positif atau negatif, menggunakan sebuah library yang bernama TextBlob.

```
-0.5062499999999999 : rt guys jokes aside bad matric results mean jack maybe time struggles people millionaires regar  
-0.6333333333333333 : rt one cruel inhumane individual known bully never allowed governm  
0.1 : rt tukslaw 1 year old xl bully booked euthanised monday morning breeder interest  
-0.04444444444444446 : rt actively preventing countries participating defence western civilization russian fascist aggressio  
n  
0.1 : wow one cares country overthrow rouge president monume  
0.1 : rt people ndi strategists trying bully stood gej 2015 near death wish think blackmai  
-0.1638888888888889 : rt old grange pics popped morning character susi naive amp hard working feisty w  
0.15 : bully actors satisfied us destroying  
0.5 : rt guys need better let others blackmail bully silence politics thing  
0.125 : speedrun bully first ever speedrun gonna later time  
0.0 : encouragement recognise need help shoot mouth try make others f  
-0.5062499999999999 : rt guys jokes aside bad matric results mean jack maybe time struggles people millionaires regar  
0.5 : person want candidate many camp tweet abuse bully sorts  
0.0 : school bully comes mind  
0.5 : ok understand position pr
```

f) Implementasi Naïve Bayes

Classification merupakan memetakan (mengklasifikasikan) data ke dalam satu kelas atau beberapa kelas yaang sudah didefinisikan sebelumnya. Di tahap ini menggunakan algoritma Multinomial Naïve Bayes merupakan model penyederhanaan dari Metoda Bayes yang cocok dalam pengklasifikasian teks atau dokumen.

```
In [30]: print(accuracy_score(y_test, prediction))
```

```
0.76
```

```
In [31]: confusion_matrix(y_test, prediction)
```

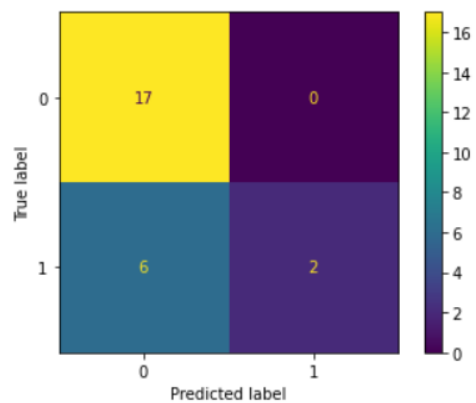
```
Out[31]: array([[17,  0],  
               [ 6,  2]], dtype=int64)
```

```
In [32]: classifier.predict(counter.transform(['stop bullying for child']))
```

```
Out[32]: array(['Positive'], dtype='<U8')
```

```
In [33]: disp = ConfusionMatrixDisplay(confusion_matrix=confusion_matrix(y_test, prediction))
```

```
In [34]: disp.plot()  
plt.show()
```



g) Evaluasi Hasil

Jadi, Hasil kinerja dari algoritma Multinomial Naïve Bayes setelah dilakukan eksperimen diperoleh akurasi sebesar 76%.

- **Daftar Pustaka**

Andrean, R., Nugroho, A., & Pratama, S. F. (2019). Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis.

JOINTECS (Journal of Information Technology and Computer Science), 39–44.

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, 131–145.

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment Analysis of Social Media Twitter with Case of AntiLGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm. *Procedia Computer Science*, 161, 765-772.

Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 200–209.