

SENTIMEN ANALISIS OPINI MASYARAKAT TERHADAP KASUS PELECEHAN SEKSUAL DI KPI PADA SOCIAL MEDIA TWITTER DENGAN METODE NAÏVE BAYES CLASIFIER (NBC)

¹Permana Perdana Putra, ²Angga Aditya Permana

¹Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang, Jl. Perintis Kemerdekaan 1/33, Cikokol Kota Tangerang

²Prodi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Jl. Boulevard Gading Serpong, Kelapa Dua Kab Tangerang
e-mail: angga.permana@umn.ac.id

Receive: 30 November 2022

Accepted: 28 Februari 2023

Abstrak

Pelecehan seksual adalah perlakuan kekerasan terhadap wanita sering kali terjadi disetiap dunia, kekerasan/pelecehan seksual lebih dikenal sebagai pemerkosaan. Kekerasan terhadap perempuan adalah suatu tindakan yang sangat tidak kemanusiaan, perempuan juga berhak mendapatkan perlindungan hak asasi manusia dan kebebasan disegala bidang.Kejahatan ini disebut juga kejahatan kesusilaan yaitu kejahatan kesusilaan yang terhubung dengan masalah seksual, tercantum didalam buku III KUHP mulai Pasal 281 sampai dengan Pasal 299 dengan tindakan ini masyarakat banyak berkomentar baik kritikan atau pun saran pada media sosial *twitter* mereka. *Tweet* yang mengandung kata pelecehan seksual menjadi hal penting untuk menilai pendapat dari masyarakat dengan cara memproses data teks *tweet* dan menganalisis sentimen yang akan di dapatkan sebagai solusi pada tindakan pelecehan seksual berikutnya dengan bantuan klasifikasi metode *algoritma naïve bayes classifier* mengitung *prior probabilitas* pada kata *class*. Tujuan ini analisis sentimen ini untuk mengklasifikasi data *tweet* sebanyak 800 menjadi data *training* dan 200 data *tweet* menjadi data *testing* menghasikan *cuitan twitter* tentang pelecehan seksual lebih banyak mendapatkan respon sentimen “*netral*”. Hasil analisis sentimen akan di jabarkan dalam sebuah sistem informasi hasil analisis sentimen dalam bentuk diagram UML (*Unified Modelling Language*) kemudian proses akan dilanjutkan dengan membuat rancangan sistem. Hasil analisis sentimen ini adalah sistem informasi berbasis web dapat diakses menggunakan browser. Dengan demikian akan menghasilkan informasi hasil analisis sentimen menggunakan *rstudio* dalam memproses data.

Kata Kunci: Analsis Sentimen, Pelecehan KPI, Naïve Bayes Classifier.

PENDAHULUAN

Beberapa tahun terakhir pada era informasi seperti saat ini, pengguna internet di indonesia telah tumbuh secara luar biasa cepatnya dan internet memegang peranan penting dalam segala aspek kehidupan manusia. Internet menjadi media yang banyak di gunakan oleh hampir semua kalangan. Berdasarkan data dari Kompas, Pertumbuhan pengguna internet tahun 2021 di Indonesia telah menembus angka 202,6 juta jiwa. Jumlah ini meningkat 15,5 persen atau 27 juta jiwa dibandingkan pada tahun 2020 lalu. Total jumlah penduduk indonesia sendiri saat ini adalah 274,9 juta jiwa. Ini artinya, penetrasi internet di indonesia pada awal 2021

mencapai 73,7 persen dan pengguna internet mulai dari umur 16 – 64 tahun yang memiliki telepon genggam adalah 98,3 persen.

Para pengguna internet yang paling digemari oleh pengguna internet indonesia ialah bermedia sosial. Saat ini, ada 170 juta jiwa orang indonesia yang merupakan pengguna aktif media sosial. Rata – rata dari mereka menghabiskan waktu 3 jam 14 menit diplatform jejaring sosial seperti (Instagram, Facebook, Twitter) dan lain – lain. Seperti *Twitter* telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet di indonesia. (Steven and Wella 2020)(Budiman, Young, and Suryadibrata 2021)

Kegunaan *twitter* selain sebagai media untuk berbagi informasi dengan mem-*post* berbagai macam *tweet*, *twitter* juga sering digunakan untuk bersosialisasi antar pengguna dan mengungkapkan sentiment atau opini mereka terhadap suatu topik atau isu – isu yang sedang hangat diperbincangkan, tidak hanya opini yang positif tapi juga yang negatif. Kebiasaan tersebut kadang tidak dianggap begitu penting untuk disikapi karena terlalu banyak *tweet*. Namun kebiasaan tersebut begitu penting apabila dipelajari lebih lanjut karena kebiasaan tersebut bisa dimanfaatkan sebagai sebuah analisis sentiment atau opini terhadap isu– isu yang sedang hangat dan dapat di jadikan sebuah informasi untuk mengetahui sentimen atau opini masyarakat Indonesia pada permasalahan tertentu.(Syakuro 2017)(Fairuz, Ramadhani, and Tanjung 2021)

Adapun analisis ini akan melakukan proses pengambilan data, proses data, klasifikasi naïve bayes, menampilkan visualisasi dari kata – kata sesuai dengan *topic treding* pada twitter (Herdhianto 2020), dalam analisis ini mengambil tentang “pelecehan seksual pada pegawai kpi”, berfokus pada pembicaraan yang telah terjadi di media sosial twitter.

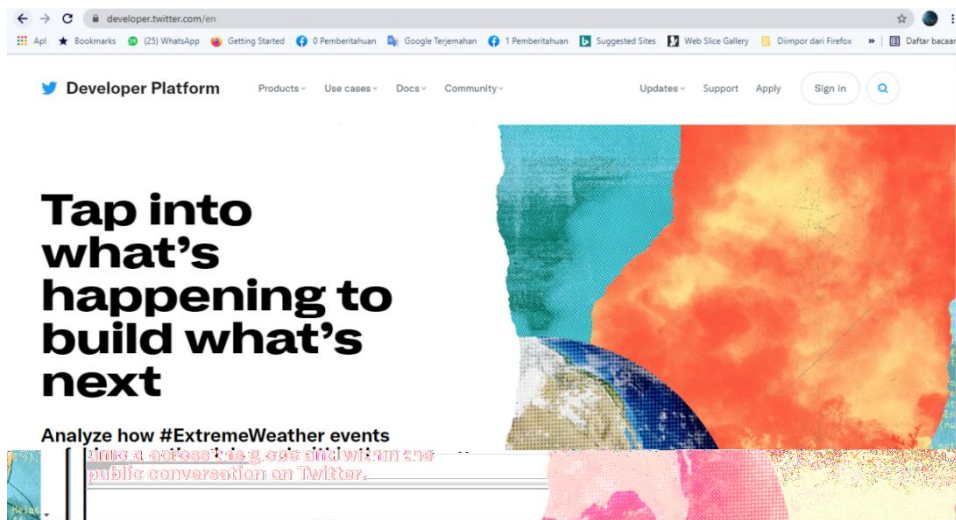
METODE PENELITIAN

Tahap ini adalah pengumpulan data yang telah di lakukan oleh opini masyarakat yang berasal dari media sosial twitter. Data yang di ambil berupa tweet dengan kata kunci “Pelecehan Seksual Pada Pegawai KPI”. Data didapatkan dengan cara menggunakan Rstudio dan Twitter API. Sebelum ke tahap pengumpulan data atau disebut dengan *crawling* data kita harus mempunyai *API Key* untuk mengidentifikasi dengan twitter.

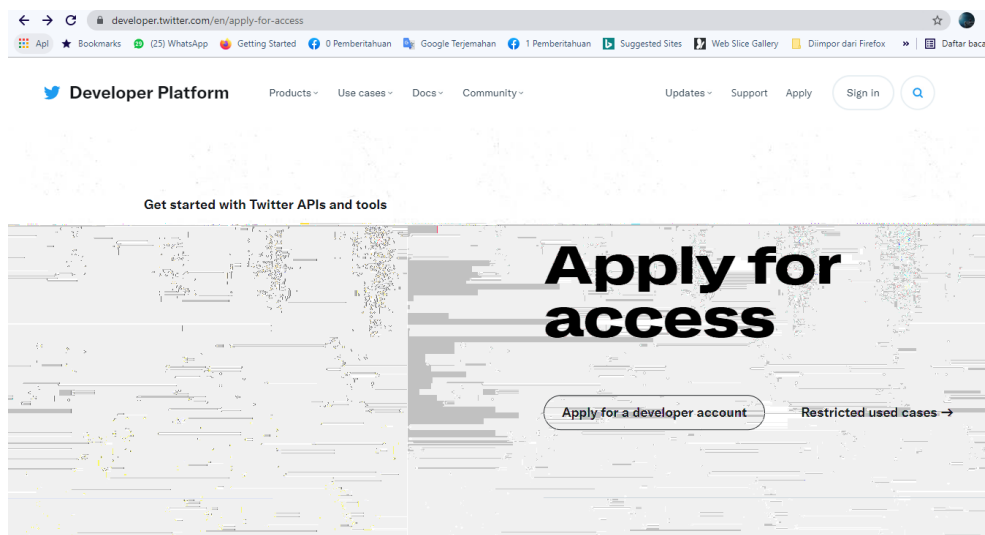
Proses mendapatkan *API key* :

1. Masuk ke dalam akun Twitter.
2. Masuk ke App Twitter atau buka <https://apps.twitter.com/app/new>
3. Lalu *Sign In API Development for Application on Twitter*
4. Kemudian mengisi formulir dengan informasi yang dibutuhkan.
5. Mendapatkan *api key*, *api secret*, *access token*, dan *access token secret*.

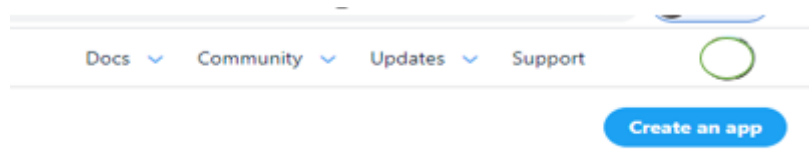
Tahap registrasi twitter API akan dijelaskan pada gambar dibawah ini :



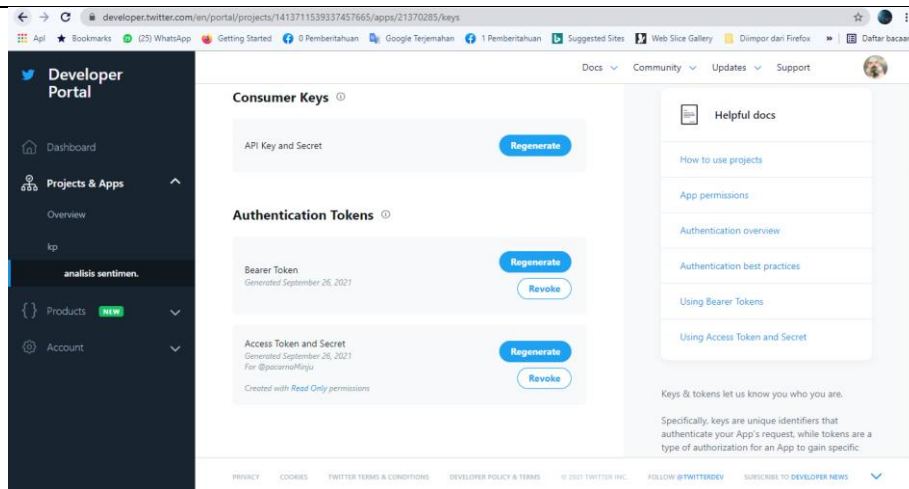
Gambar 1. Klik registrasi developer platform klik apply



Gambar 1. Klik apply for developer account



Gambar 2. create an app



Gambar 3. api key, api secret, access token, dan access token secret

Sesudah berhasil registrasi *Twitter API* dan mendapatkan token key, kemudian baru bisa memulai proses pengambilan data atau disebut dengan *crawling* data, akan tetapi harus melakukan mengidentifikasi dari *twitter* menggunakan *rstudio* dengan menginstal *package library* dari *rstudio*, setelah melakukan mengidentifikasi telah berhasil selanjutnya dapat mencari data sesuai dengan kata kunci (*Keyword*).

1. Pre-processing Data

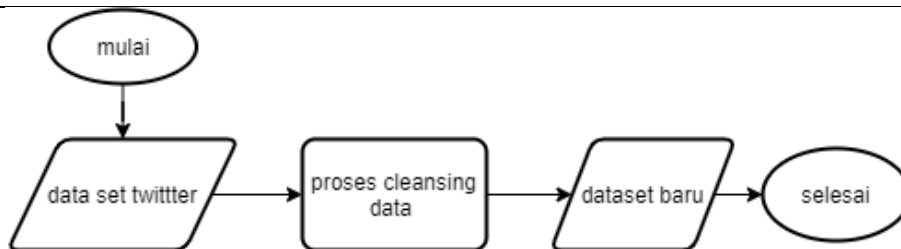
Tahap ini adalah untuk membersihkan kata – kata yang tidak perlu atau kata – kata yang tidak memiliki arti, Tahapan Tahapan *preprocessing* yang di lakukan di sesuaikan berdasarkan kondisi dari data tweet *user twitter*. Adapun urutan dari tahapan *preprocessing* yang di lakukan ialah:

A. Cleaning

Dengan adanya muncul simbol, tanda baca dan bilangan angka pada komentar pengunjung situs Vivacoid membuat data menjadi tidak efektif dan tidak memiliki arti. Pada tahap ini akan menghapus simbol, bilangan angka, dan tanda baca tersebut. Dan proses ini dilakukan menggunakan program, sehingga dilakukan secara otomatis sebelum menyimpan hasil decode ke dalam bentuk *xlsx*.

Tabel 1. Contoh data cleansing

Sebelum	Sesudah
RT @VIVAcoid : Kombes Hengki Pastikan Masih Selidiki Kasus Pelecehan Pegawai KPI https://t.co/Nyd0p3OnCG	kombes hengki pastikan selidiki pelecehan pegawai kpi



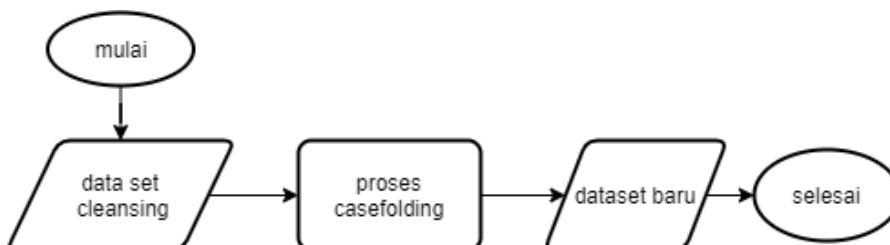
Gambar 5. Tahapan cleansing

B. Case Folding

Proses penyamaan *case* dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* di butuhkan dalam mengkonversi keseluruhan teks dalam huruf kecil atau *lowercase*.

Tabel 2. Contoh data case folding

Sebelum	Sesudah
Kombes Hengki Pastikan Masih Selidiki Kasus Pelecehan Pegawai KPI	kombes hengki pastikan selidiki pelecehan pegawai kpi



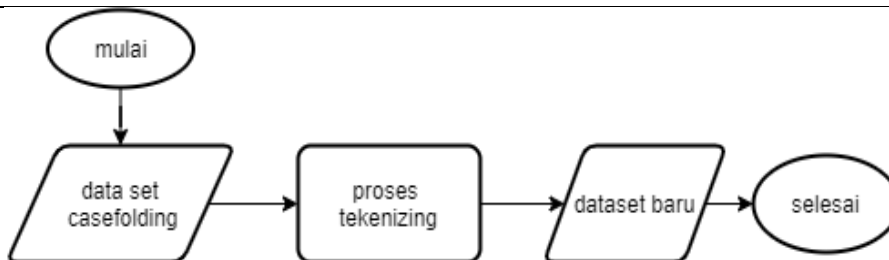
Gambar 6. Tahapan case folding

C. Tokenizing.

Proses pemotongan sebuah dokumen menjadi bagian yang dapat berupa kalimat, dan kata. Proses ini juga akan di lakukan penghilangan *whitespace*,

Tabel 3. Contoh data tokenizing

Sebelum	Sesudah
kombes hengki pastikan selidiki pelecehan pegawai kpi	("kombes", "hengki", "pastikan" selidiki "pelecehan" "pegawai" "kpi")



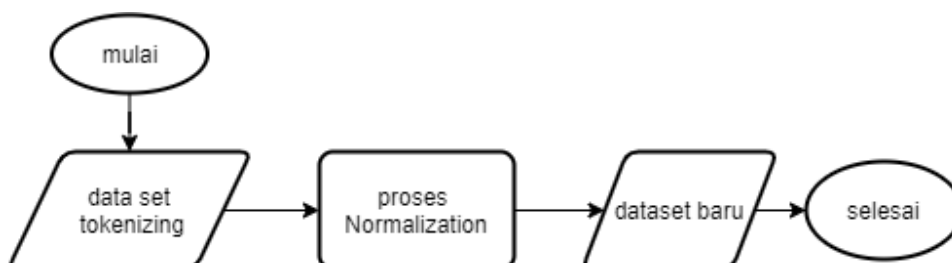
Gambar 7. Tahapan tokenizing

D. Normalization

Tahapan *normalization* ini dilakukan untuk mengubah kata yang tidak sesuai dengan EYD, dapat menurangi hasil sentimen, tahapan dibagi tiga Langkah yaitu konversi,kata baku, dan konversi kata Bahasa Inggris. Diagram alur normalization

Tabel 4. Contoh data normalization

Sebelum	Sesudah
KPI hanya memberikan hasil investigasi kasus pelecehan seksual yg dialami pegawainya ke polisian	KPI hanya memberikan hasil investigasi kasus pelecehan seksual yang dialami pegawainya ke polisian



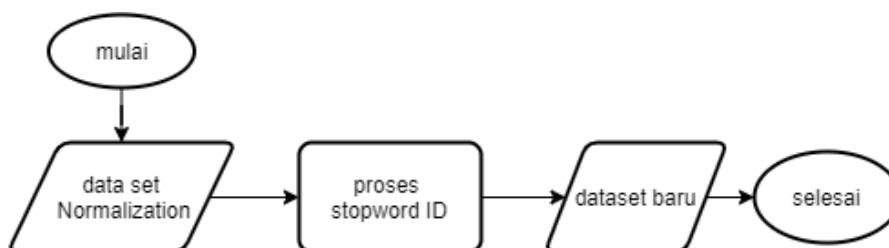
Gambar 8. Tahapan Normalization

E. Stopword

Proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen sesuai dengan input yang di masukkan. Kata-kata yang termasuk ke dalam stopword di hilangkan karena memberikan pengaruh yang tidak baik dalam proses *text mining* seperti kata-kata “berapa”,”kenapa”, dan lain-lain yang berbahasa Indonesia.

Tabel 5. Contoh data stopwords

Sebelum	Sesudah
Kenapa harus dipaksa damai, kalau akhirnya MS di rugikan	harus dipaksa damai, kalau akhirnya MS di rugikan



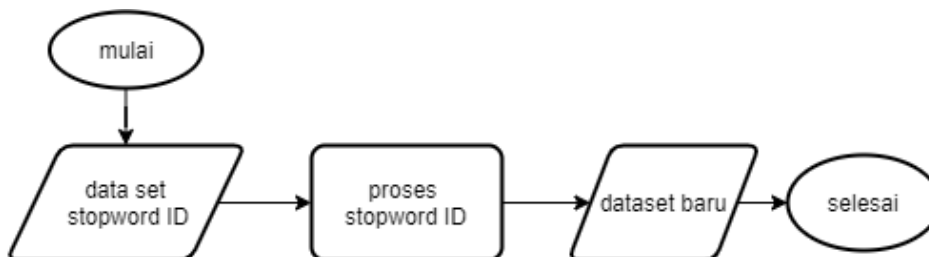
Gambar 9. Tahapan stopword

F. Stemming

Tahapan ini mengubah kata-kata yang terdapat imbuhan untuk membuat kata-kata yang memiliki imbuhan berubah menjadi kata dasar sesuai aturan bahasa Indonesia yang baik dan benar, Diagram alur stemming.

Tabel 6. Contoh data stemming

Sebelum	Sesudah
komnas ham minta sikap resmi kpi soal dugaan pelecehan pegawai	Komnas ham berharap sikap resmi kpi soal dugaan pelecehan pegawai

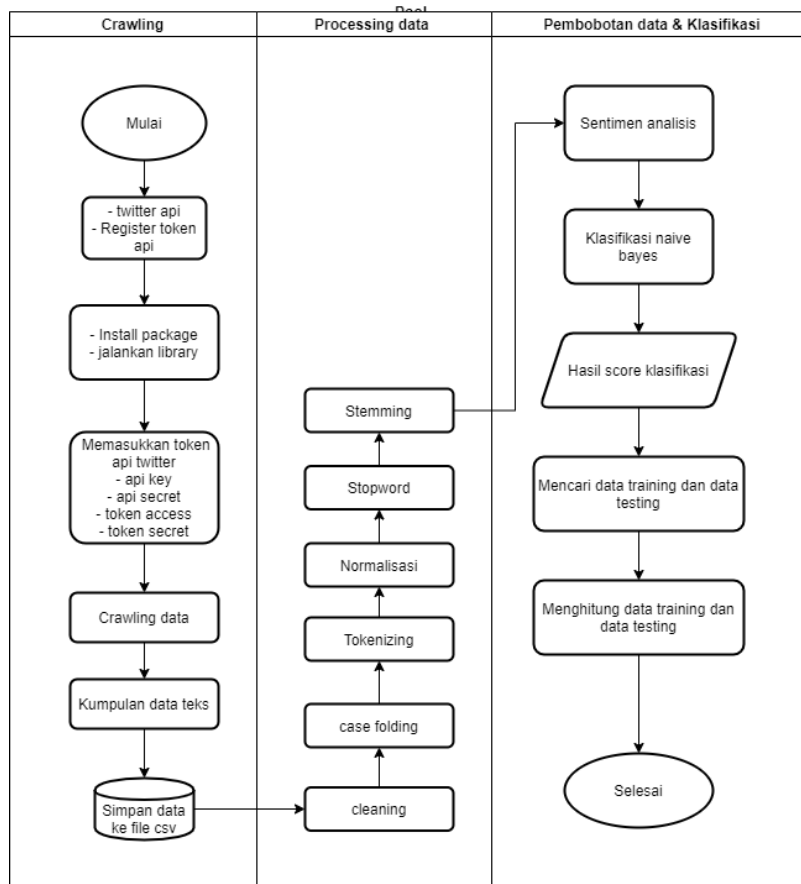


Gambar 10. Tahapan stemming

HASIL DAN PEMBAHASAN

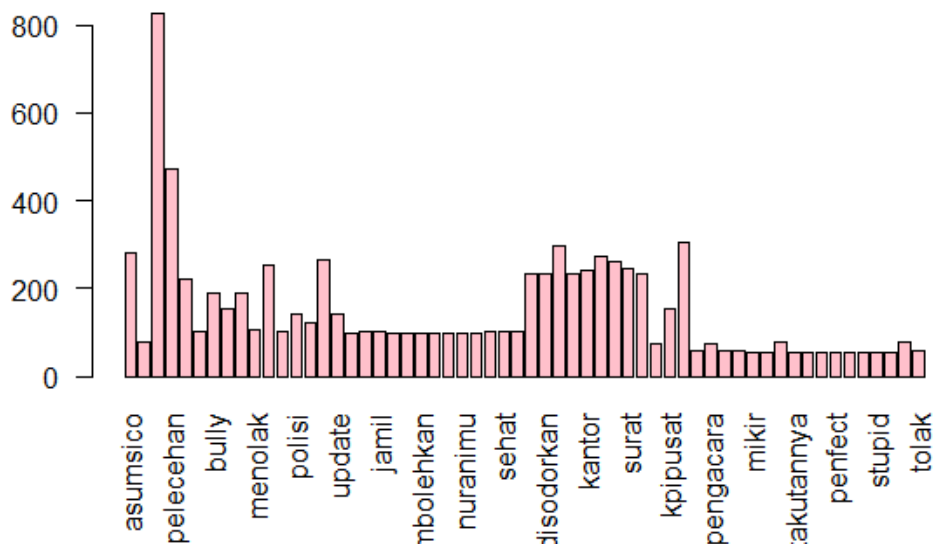
Dalam penelitian ini, berikut adalah *flowchart* analisis sentimen dimulai dari registrasi *twitter* API untuk mendapatkan data *tweet* dengan kata kunci pelecehan

seksual pada pegawai kpi hingga mendapatkan hasil klasifikasi dengan menggunakan *Rstudio* dan metode naïve bayes classifier adalah sebagai berikut :



Gambar 11. flowchart Analisis Sentimen

Dari hasil proses term menampilkan visualisasi berbentuk *barplot* dan *workcloud* sesuai dengan proses kata pada *barplot* menunjukkan hasil yang paling tinggi, kata paling banyak muncul dari kata – kata yang lain. *Workcloud* yang besar paling banyak muncul atau paling banyak digunakan dalam tweet dan kata yang berukuran kecil, sedikitnya kemunculan dan digunakan dalam tweet mengenai topik pelecehan kpi, pada samping *barplot* juga tercantum munculnya frekuensi. Dibawah ini memperlihatkan program pada *Rstudio*.



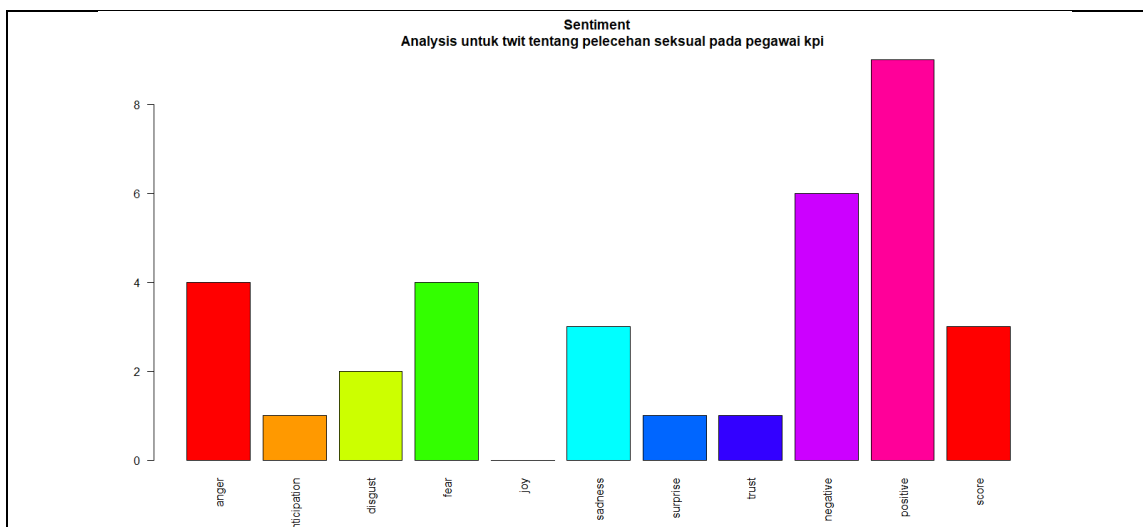
Gambar 12. tampilan barplot



Gambar 13. tampilan wordcloud

Analisis sentimen berdasarkan emosi

Tahap ini adalah berdasarkan hasil emosional dari kata yang sudah melakukan tahap *preprocessing* dan akan masuk kedalam beberapa kategori marah, bahagia, sedih, positif, negatif dan netral. Hasil analisis sentimen ini berdasarkan emosional seseorang mengeluarkan pendapat lalu akan menampilkan *visualisasi barplot* yang memiliki beberapa warna, keterangan jumlah, keterangan emosional dan lain – lain.



Gambar 14. Barplot emosional

Klasifikasi menggunakan metode naïve bayes

Tahap Klasifikasi ini menggunakan *naïve bayes classification* proses klasifikasi membuktikan kata masuk kedalam klasifikasi kata positif, negatif, dan netral, ketika melakukan klasifikasi mendapatkan score 0 maka hasil data klasifikasi netral, kata tersebut mempunyai kata netral dari pada positif ataupun negatif, selanjutnya menentukan data *training* dan data *testing* yang sudah dikerjakan tahap *text preprocessing*. Data *training* total jumlah 800 kata dan data *testing* total jumlah 200 kata, langkah selanjutnya menghitung prior probabilitas dalam data *training* maupun data *testing*.

Dari data *training* dengan teks “anggota lpsk temui pegawai kpi korban pelecehan seksual kaji perlindungan hukum” data sudah di klasifikasikan dengan hasil emosional “netral” dan hasil prior probabilitas 0.00125.

Dari data *testing* dengan teks “update dugaan pelecehan seksual kpi lpsk datang ke hadapan terduga korban” data sudah di klasifikasikan dengan hasil emosional “netral” dan hasil prior probabilitas 0.00125.

	score	text	klasifikasi
1	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
2	0	kae pastikan selidiki dugaan pelecehan pegawai kpi	Netral
3	0	komb hengki pastikan selidiki pelecehan pegawai kpi	Netral
4	0	komb hengki pastikan selidiki pelecehan pegawai kpi	Netral
5	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
6	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
7	0	polisi ungkakendala penyidikan pelecehan seksual pegawai ...	Netral
8	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
9	0	polisi ungkakendala penyidikan pelecehan seksual pegawai ...	Netral
10	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
11	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
12	0	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral

Gambar 15. Hasil score klasifikasi Naïve bayes

	text.score	text.text	text.klasifikasi
1	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
2	NA	kae pastikan selidiki dugaan pelecehan pegawai kpi	Netral
3	NA	komb hengki pastikan selidiki pelecehan pegawai kpi	Netral
4	NA	komb hengki pastikan selidiki pelecehan pegawai kpi	Netral
5	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
6	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
7	NA	polisi ungkakendala penyidikan pelecehan seksual pegawai ...	Netral
8	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
9	NA	polisi ungkakendala penyidikan pelecehan seksual pegawai ...	Netral
10	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
11	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral
12	NA	bulli cehan kpipolisi lakukan elidikan kejadian samdi...	Netral

Gambar 16. Klasifikasi Naïve bayes

```

a internasional kpi...
  Netral 0.00125
  text.text
Y ya ketua kja isinya ngelakuin cehan seksual ag dihar
  Netral 0.00125
  text.text
Y yg dikit2 ngadu kek gini i ngide delik emaran nama dicount alasan utamakan es...
  Netral 0.00125
  text.text
Y yuk tandatangani si biar ku cehan seksual wai kpi diara
  Netral 0.00125
  text.text
Y zakar dicoreti sl trus disoraki rame2 ms wai kpi pusat terduga kunya tujuh ora
ng...
  Netral 0.01

```

Gambar 17. Hasil prior probabilitas data training

```
Y      tindak terkait dugaan kekerasan seksual lingkungan kpi permintaan keterangan |
kesian...
Netral
0.005
text.text
Y      updat dugaan pelecehan seksual kpi lpsk datangi kediaman terduga korban ms
Netral
0.005
text.text
Y      zakar dicoreti sl trus disoraki rame2 ms wai kpi pusat terduga kunya tujuh or
ng...
Netral
010
0
$levels
fill "Netral"
```

Gambar 18. Hasil prior probabilitas data testing

UCAPAN TERIMAKASIH

Penelitian ini dapat dilaksanakan dengan baik berkat bantuan dari berbagai pihak, untuk itu peneliti mengucapkan terima kasih kepada Universitas Multimedia Nusantara dan Universitas Muhammadiyah Tangerang atas dukungan dan bantuan yang telah diberikan selama proses penulisan artikel ini.

SIMPULAN DAN SARAN

Tahap analisis sentimen data twitter yang telah dilakukan oleh penulis pertama crawling data, preprocessing data, pembobotan term, klasifikasi naïve bayes, sentimen, membuktikan data training dan data testing class untuk *prior probabilitas*. Metode Naïve Bayes Classification juga sangat efektif dalam mengklasifikasikan sentimen pelecehan seksual. Sistem analisis jika ditambahkan dengan algoritma lain dapat menambah akurasi perhitungan dengan akurasi hingga 80%.

DAFTAR PUSTAKA

- Budiman, Andre, Julio Christian Young, and Alethea Suryadibrata. 2021. "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Konten Twitter Dengan Indikasi Depresi." *Jurnal Informatika: Jurnal Pengembangan IT* 6(2):133–38.
- Fairuz, Ardianne Luthfika, Rima Dias Ramadhani, and Nia Annisa Ferani Tanjung. 2021. "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter." *Journal of Dinda : Data Science, Information Technology, and Data Analytics* 1(1):42–51. doi: 10.20895/dinda.v1i1.180.
- Herdhianto, Adhyaksa. 2020. *Sentiment Analysis Menggunakan Naïve Bayes Classifier (NBC) Pada Tweet Tentang Zakat*.
- Steven, Cristian, and Wella Wella. 2020. "The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter." *IJNMT (International Journal of New Media Technology)* 7(2):102–10. doi: 10.31937/ijnmt.v7i2.1732.
- Syakuro, Abdan. 2017. "Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro." *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)* 1–89.