

Penerapan *Support Vector Machine* Untuk Analisis Sentimen pada X (Twitter) Mengenai Obat Penyebab Gagal Ginjal Akut pada Anak

Tasya Rizki Salsabilla, Nunik Pratiwi*

Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. DR. HAMKA

Email: trizkisalsabilla@gmail.com, npratiwi@uhamka.ac.id

*Corresponding authors

Article Info

Received: September 1, 2024

Accepted: September 28, 2024

Published: September 30, 2024

ABSTRAK

Pada Oktober 2022, banyak kasus anak-anak yang terserang penyakit gagal ginjal akut, hal itu disebabkan karena adanya senyawa kimia berbahaya yang terdeteksi dalam riwayat penggunaan obat batuk anak. Pernyataan tersebut menimbulkan kontroversi dan menjadi perbincangan di media sosial khususnya X (*Twitter*). Dalam menjelajahi opini ini dapat mencapai suatu keputusan yang dapat diterapkan dalam pembelajaran mesin dan analisis sentimen yang akan digunakan yaitu *Support Vector Machine* (SVM). Tujuan dilakukannya penelitian ini untuk mengetahui sentimen masyarakat terhadap obat penyebab gagal ginjal akut pada anak serta melihat kinerja dari algoritma *support vector machine*. Data yang digunakan sebanyak 1128. Berdasarkan hasil penelitian, masyarakat menanggapi topik ini dengan negatif hal ini dibuktikan dengan sentimen negatif yang didapat lebih besar dibanding sentimen positif dan algoritma *support vector machine* dengan kernel linear melakukan kinerja yang sangat baik hal ini dibuktikan dengan nilai *accuracy* yang sangat baik yaitu sebesar 91%.

Kata kunci : *analisis sentimen, gagal ginjal akut, support vector machine, text mining, x(twitter)*

ABSTRACT

In October 2022, there were many cases of children suffering from acute kidney failure due to harmful chemical compounds detected in the history of children's cough medicine use. The statement caused controversy and became a conversation on social media, especially Twitter. Exploring this opinion can lead to a decision that can be applied in machine learning and sentiment analysis, namely support vector machines (SVM). The purpose of this research is to find out the sentiment of the community towards drugs that cause acute kidney failure in children and see the performance of the support vector machine algorithm. The data used was 1128. Based on the results of the study, the community responded negatively to this topic, as evidenced by the fact that the negative sentiment obtained was greater than the positive sentiment, and the support vector machine algorithm with a linear kernel performed very well, as evidenced by the excellent accuracy value of 91%.

Keywords : *sentiment analysis, acute renal failure, support vector machine, text mining, ,x(twitter)*

1. PENDAHULUAN

Pada bulan Oktober tahun 2022 di Indonesia banyak kasus anak-anak yang umurnya antara 6 bulan sampai 18 tahun terkena penyakit gagal ginjal akut dan terjadi peningkatan pada dua bulan terakhir (Kemenkes RI, 2022).

Sebanyak 189 kasus telah dilaporkan pada tanggal 18 Oktober 2022 yang didominasi anak usia 1 sampai 5 tahun (Kemenkes RI, 2022). Menteri kesehatan mengatakan pada pasien penderita gagal ginjal akut dalam riwayat penggunaan obat yang dikonsumsi terdapat senyawa kimia yang berbahaya, senyawa kimia berbahaya tersebut yaitu

senyawa dietilen glikol (DEG), senyawa etilen glikol (EG), dan senyawa etilen glikol butyl eter (EGBE) (CNNIndonesia, 2022).

Beberapa obat yang diduga memiliki kandungan dietilen glikol (DEG) dan etilen glikol (EG) yaitu termorex sirup yang merupakan obat demam, Flurin DMP Sirup merupakan obat batuk dan flu, Unibebi Cough Sirup yang merupakan obat batuk dan flu dan lain-lain (CNNIndonesia, 2022). Karena semakin banyak anak-anak yang terkena gagal ginjal akut yang diduga disebabkan oleh obat yang terjual bebas maka hal tersebut menimbulkan kontroversi dan menjadi perbincangan di media sosial terutama X (*Twitter*).

Pandangan atau opini ini sering diutarakan melalui media sosial, dan dalam menjelajahi pandangan atau opini ini dapat mencapai sebuah keputusan yang dapat diterapkan di dalam pembelajaran mesin. Salah satu platform media sosial yang paling banyak digunakan untuk menyampaikan pendapat ada X(*Twitter*) (Hasibuan et al., 2022). Berdasarkan laporan dari We Are Social, data yang didapat dari (DataIndonesia, 2023) menunjukkan bahwa terdapat 556 juta pengguna X(*Twitter*) di Indonesia per Januari 2023. Pada periode yang sama, terdapat peningkatan sejumlah 27,4 persen dibandingkan pada tahun sebelumnya (DataIndonesia, 2023).

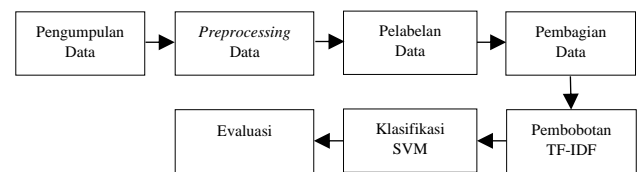
Analisis sentimen adalah metode yang digunakan untuk membantu memahami opini atau pandangan dari suatu kasus atau topik tertentu yang dapat diambil kesimpulan maupun keputusan berdasarkan kalimat maupun dokumen (Setiawan et al., 2021). Analisis sentimen yaitu metode untuk mengumpulkan data dari berbagai situs *online*. Tujuan analisis sentimen yaitu untuk mengakumulasi polaritas dari teks atau komentar pada dokumen yang bersifat positif atau negatif. X(*Twitter*) dipilih karena popularitasnya sebagai platform media sosial di kalangan pengguna *internet* jaman sekarang (Kevin et al., 2020). Analisis sentimen telah banyak diimplementasikan di berbagai sektor, dari *e-sport* (Alamsyah & Pratiwi, 2024) hingga politik (Palepa et al., 2024).

Terdapat beberapa metode untuk menganalisis sentimen dan salah satunya *support vector machine* (SVM) yang digunakan pada penelitian ini. Teknik klasifikasi *supervised learning* salah satunya merupakan *Support Vector Machine* (SVM) metode yang memprediksi kelas menggunakan sebuah model atau pola yang berasal dari proses pelatihan (Sari & Haranto, 2019). Mencari *hyperplane* atau garis pemisah yang membagi satu kelas dengan kelas lain merupakan cara klasifikasi bekerja (Syah & Witanti, 2022). Beberapa penelitian tentang analisis sentimen pada media sosial yang telah dikerjakan sebelumnya yaitu metode *Support Vector Machine* lebih unggul dalam hal *precision*, *recall* dan akurasi menurut hasil penelitian yang dilakukan (Setiawan et al., 2021). Hal ini dikarenakan terjadi perulangan sebanyak 423 kali pada proses pelatihan *Support Vector Machine* untuk mendapatkan bobot yang ideal. Jika dibandingkan algoritma *Support Vector Machine*, algoritma *naïve bayes* menghasilkan hasil yang lebih cepat dalam segi waktu karena hanya melakukan satu kali proses iterasi, tetapi hasil yang didapatkan kurang optimal.

Penelitian yang juga dilakukan oleh Lestari et al mendapatkan kesimpulan yaitu metode SVM akurasi lebih tinggi dibandingkan dengan NBC. SVM memperoleh akurasi tertinggi mencapai 99.67% dan yang terendah 94.17%. Sedangkan NBC memperoleh akurasi tertinggi mencapai 96.56% dan yang terendah 81.69%. Pada proses analisis data pengujian, penerapan *information gain* pada SVM mempercepat durasi proses analisis sebesar 195,71% (Lestari et al., 2022). Dari penjelasan latar belakang tersebut, maka penelitian ini akan melakukan analisis sentimen pada media sosial X (*Twitter*) terkait obat penyebab gagal ginjal akut pada anak serta penelitian ini ingin melihat performa dari algoritma *Support Vector Machine* (SVM) dengan menggunakan keempat *kernel* yang ada di algoritma tersebut.

2. METODE PENELITIAN

Penelitian ini terdiri dari tujuh tahap yaitu pengumpulan data, pre-processing data, pelabelan data, pembagian data, pembobotan TF-IDF, klasifikasi SVM, dan terakhir evaluasi. Gambar 1 menunjukkan alur penelitian.



Gambar 1. Alur Penelitian

Pada penelitian ini data ditarik dari *tweet*/kicauan masyarakat pada X (*Twitter*) berupa opini yang biasa disebut dengan proses *crawling*. Data *tweet* yang diambil menggunakan Bahasa Indonesia dengan kata kunci “obat penyebab gagal ginjal akut pada anak” dengan minimal pengambilan data yaitu 1000 data. Data yang diambil merupakan data random dari pengguna biasa, lalu selanjutnya kumpulan data *tweets* tersebut diintegrasikan ke dalam sebuah dataset.

Preprocessing merupakan langkah pertama dalam proses pelatihan yang bertujuan untuk mencegah data yang kurang konsisten, data yang kurang sempurna dan gangguan yang ada pada data (Hafidz et al., 2020). Adapun langkah-langkah yang akan dijalankan pada penelitian ini adalah :

1. *Cleansing*: Pada tahap ini menghapus sebuah kata yang tidak digunakan dari *tweet* untuk mengurangi noise. Tanda baca yang dihilangkan adalah tandanya (?), titik (.) tanda seru (!) dan koma (,) bersama dengan tanda '@' untuk nama pengguna, *hashtags* (#), *emoticon* dan URL.
2. *Case folding*: Tahap ini dilakukan perubahan setiap huruf kapital pada *dataset* menjadi huruf kecil.
3. *Filtering* atau *Stopword*: Tahapan yang dilakukan yaitu proses pemilihan kata-kata yang tidak terlalu penting atau tidak berkaitan dengan analisis sentimen dihapus, contohnya seperti kata penghubung, seperti, yang, dan, dengan, pada, serta, di, ke dan sebagainya.

4. *Tokenizing*: Tahap ini merupakan proses untuk memecah kalimat berdasarkan dari kata penyusunnya, kata ini dipecah berdasarkan spasi.
5. *Stemming*: Tahap ini merupakan proses untuk menghapus kata kata yang mempunyai imbuhan seperti ber-, per-, di-, ke-, mem-, men-, -an, -kan, dan lain lain, agar kata tersebut menjadi kata dasar.
6. *Normalization*: Tahapan ini akan mengubah sebuah kata yang tidak terstruktur atau kata-kata yang singkat menjadi sebuah kata yang sudah sesuai dengan standar Kamus Besar Bahasa Indoneisa (Tuhuteru & Iriani, 2018).

Setelah dilakukannya *preprocessing*, hasil data dari pengolahan tersebut akan memasuki tahap pelabelan. Pelabelan ini berfungsi untuk memberikan label sentimen kepada dataset, label yang digunakan yaitu label positif dan negatif. Pemberian label ini sangat penting karena nantinya akan menentukan akurasi pada proses selanjutnya. Pelabelan dilakukan secara otomatis dengan menggunakan *vader lexicon*.

Pada tahap ini dataset yang sudah melalui tahap *preprocessing*, langsung masuk ke tahap pelabelan yang akan dibagi menjadi dua data yaitu data latih dan data uji dengan rasio yang akan diberikan sebesar 80:20, dengan 80 untuk data latih dan 20 untuk data uji. Data uji digunakan untuk menilai model, sedangkan data latih untuk melatih model (Syah & Witanti, 2022).

Setelah melakukan proses pembagian dataset, kemudian melakukan pembobotan kata yang akan dilakukan sebelum dilanjutkan ke tahap implementasi Algoritma Support Vector Machine. Dalam penelitian ini pembobotan kata akan dilakukan menggunakan TF-IDF (Term Frequency-Inverse Document) yang cara kerjanya dengan memberikan bobot pada setiap kata tertentu disebuah dokumen tanpa menyampaikan maksud atau makna yang terkandung dalam kata-kata tersebut. Akan ada dua prosedur yang akan dilakukan, yaitu term frequency yaitu frekuensi kemunculan istilah atau term yang ada didalam suatu dokumen dan inverse document frequency yaitu untuk mengukur kepentingan sebuah kata didalam dokumen (Laurensz & Eko Sedyono, 2021).

Tahap selanjutnya yaitu mengklasifikasi, proses klasifikasi menggunakan metode. SVM adalah model teori pembelajaran statistik yang mampu menghasilkan hasil lebih baik daripada metode lainnya. Oleh karena itu, metode ini digunakan untuk mengelompokkan label data yang telah diberi label.

Proses yang akan dilakukan ketika menggunakan SVM adalah mengganti teks menjadi data vektor. Dengan membedakan antara sentimen positif dan sentimen negatif menggunakan garis pemisah (*hyperplane*) itu merupakan proses klasifikasi. Secara intuitif, bahwa pembatas terbaik akan didapatkan dari jarak garis terbesar ke titik pelatihan yang terdekat pada setiap kelasnya, karena margin akan semakin besar apabila error generalisasi yang didapat dari sebuah pemilah semakin rendah. Margin merupakan jarak antara suatu titik vector dalam kelas terhadap garis pemisah (B. W. Sari & Haranto, 2019).

Pada proses pembelajaran SVM, pada dasarnya untuk mendapatkan titik-titik *support vector*, perhitungan jarak antar data dilakukan dengan mentransformasikan data ke ruang dimensi yang lebih tinggi, yaitu $\phi_i \phi_j$. Secara umum, transformasi ϕ ini biasanya sangat sulit dipahami. Maka dari itu jarak antar data dapat diperhitungkan dengan digantikan oleh fungsi *kernel* $K(x_i, x_j)$, yang secara tidak langsung akan mendefinisikan ϕ , yang disebut dengan *kernel trick*. Fungsi kernel dirumuskan sebagai berikut (Romadoni et al., 2020):

$$(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (1)$$

Berikut ini terdapat 4 tipe fungsi kernel yang dapat digunakan:

1. Linear

$$K(x, x_k) = x_k^T x \quad (2)$$

2. Polynomial

$$K(x, x_k) = x_k^T x^d + 1 \quad (3)$$

3. Radial Basis Function (RBF)

$$K(x, x_k) = \exp\{-||x - x_k||_2^2 / \sigma^2\} \quad (4)$$

4. Sigmoid

$$K(x, x_k) = \tanh[k x_k^T x + \theta] \quad (5)$$

Setelah semua proses telah dilakukan dari tahap pengambilan data hingga mendapatkan hasil dari implementasi algoritma, selanjutnya memasuki tahap terakhir yaitu melakukan uji evaluasi. *Confusion matrix* digunakan untuk mengukur evaluasi pada penelitian ini. Ada beberapa variabel yang diuji yaitu akurasi, presisi, *recall* dan *f1-score*.

Pada proses evaluasi ini menunjukkan jumlah data komentar yang digunakan untuk membandingkan label antara data sebenarnya dengan hasil perkiraan. Selanjutnya, setiap baris data dihitung dulu jumlah *true positive*, *false positive*, *false negative*, dan *true negative*. Terdapat beberapa istilah didalam *confusion matrix* (Shiddicky & Agustian, 2022) yaitu TP (*true positive*) merupakan data yang pada saat diklasifikasikan menjelaskan sebagai data yang bernilai benar dan jawaban sebenarnya juga benar. FP (*false positive*) merupakan data yang pada saat diklasifikasikan menjelaskan sebagai data yang bernilai benar, tetapi jawaban sebenarnya adalah salah. FN (*false negative*) merupakan data yang pada saat diklasifikasikan menjelaskan sebagai data yang bernilai salah, tetapi jawaban sebenarnya adalah benar. TN (*true negative*) merupakan data yang pada saat diklasifikasikan menjelaskan sebagai data yang bernilai salah dan jawaban sebenarnya juga salah (*false*).

Lalu untuk menghitung hasil nilai akurasi, *precision*, *recall* dan *f1-score* menggunakan rumus yang ditunjukkan sebagai berikut (Fitri Wulandari et al., 2023):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (6)$$

$$Precision = \frac{TP}{TP + FP} * 100\% \quad (7)$$

$$Recall = \frac{TP}{TP + FN} * 100\% \quad (8)$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengambilan data pada penelitian ini menggunakan *web scraping*, dimana proses *web scraping* ini menggunakan Google colab dengan bahasa pemrograman python dengan *library snsrape*. Data yang didapatkan untuk penelitian ini adalah data *tweet* berbahasa Indonesia dengan kata kunci “obat penyebab gagal ginjal akut” dari bulan Oktober 2022 sampai Februari 2023 hasil penarikan data tersebut diperoleh sebanyak 1200 data. Berikut ilustrasi pengambilan data untuk penelitian ini, ditunjukkan pada gambar 2.



Gambar 2. Ilustrasi Pengambilan Data

Pada pengambilan data tersebut terdapat beberapa atribut yang diambil yakni tweet. Adapun data mentah dapat dilihat pada Tabel 1.

Tabel 1. Hasil dari *Web Scraping*

No	Date	Username	Tweet
1	2022-11-29 10:47:49+0 0:00	kkuratuy	Gara2 gagal ginjal akut w batuk gk bisa minum obat sirup di apotek gk ada yg jual, minum obat tablet gk mpanðŸœ§
2	2022-11-28 10:21:02+0 0:00	P30437846 Hendri	@samartemaram Ya udh sekarang kan ada kasus tuh ya..ratusan anak meninggal gagal ginjal akut karena minum obat sirup...pertanyaan siapa yg harus bertanggung jawab?cuman mau nanya itu doang sih
3	2022-11-27 03:00:47+0 0:00	IdApaajah	@whyzar@PolJokesID Seharusnya bisa, kejadian gagal ginjal akut yg konon katanya dr obat sirup jelas selain produsen ada jg kesalahan regulator dan pengawas tp km ini +62 jgn harap class action bisa menang, klo pun menang ya biarin aja.

Pada tabel tersebut dapat dilihat bahwa data yang diambil masih terdapat *mention*, emoji, tanda baca dan lain sebagainya, hal tersebut perlu dihilangkan sehingga perlu dilakukan preprocessing data sebelum memasuki pelabelan agar hasil yang didapatkan lebih optimal.

3.2. Preprocessing Data

Pada tahapan preprocessing ini bertujuan untuk membersihkan data yang masih terdapat hal-hal yang tidak diperlukan. Hal-hal yang akan dibersihkan seperti mention,

hashtags, URL, angka, simbol dan tanda baca yang tidak diperlukan serta akan menghapus kata-kata yang kurang penting pada tahapan stopword seperti “yang”, “me-” dan lain-lain (Romadoni et al., 2020). Pada tahapan ini akan dilakukan dengan bantuan software *Google Collaboratory* dengan menggunakan bahasa pemrograman *Python*. Tahapan-tahapan yang dilakukan dalam prosedur preprocessing ini adalah sebagai berikut:

1. *Cleansing*

Tahapan yang pertama dilakukan adalah proses *cleansing* data. Pada proses ini akan dilakukan dengan bantuan *library ‘re’* yang dapat melakukan operasi *regular expression* untuk dapat memperbaiki struktur text dan menghapus symbol ASCII yang tidak diperlukan. Dengan proses yang dilakukan menghapus *emoticons*, menghapus *username*, menghapus *retweet*, menghapus URL, menghapus *hashtags*, menghapus karakter khusus, menghapus semua karakter tunggal, mengubah beberapa spasi diganti dengan satu spasi dan menghapus angka. Pada Tabel 2 memperlihatkan hasil dari data yang sudah dilakukan *cleansing* dan sebelum dilakukan *cleansing*.

Tabel 2. Hasil Proses *Cleansing* Data

Sebelum <i>Cleansing</i>	Sesudah <i>Cleansing</i>
Wah gmn ini bpom kerjanya baru ada korban meninggal baru diberhentikan, kemarin penyelidikan terkait obat yg menyebabkan gagal ginjal akut pada anak itu blm tuntas atau gmn?	Wah gmn ini bpom kerjanya baru ada korban meninggal baru diberhentikan kemarin penyelidikan terkait obat yg menyebabkan gagal ginjal akut pada anak itu blm tuntas atau gmn
Sampai kapankah kasus anak dg gagal ginjal akut progresif atipikal (ggapa) akan berakhir? Kita harus bisa pastikan bahwa sirop obat yg bermasalah sudah tidak ada lagi di layanan kesehatan atau masyarakat. @BPOM_RI @KemenkesRI https://t.co/suGVfLP86z	Sampai kapankah kasus anak dg gagal ginjal akut progresif atipikal ggapa akan berakhir Kita harus bisa pastikan bahwa sirop obat yg bermasalah sudah tidak ada lagi di layanan kesehatan atau masyarakat

2. *Case Folding*

Setelah dilakukannya *cleansing* data dilanjutkan dengan tahapan *case folding* dimana pada tahapan ini proses yang dilakukan yaitu perubahan huruf besar menjadi huruf kecil. Pada Tabel 3 menunjukkan hasil data yang diproses setelah dilakukannya *case folding*.

Tabel 3. Hasil Proses *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Wah gmn ini bpom kerjanya baru ada korban meninggal baru diberhentikan kemarin penyelidikan terkait obat yg menyebabkan gagal ginjal akut pada anak itu blm tuntas atau gmn	wah gmn ini bpom kerjanya baru ada korban meninggal baru diberhentikan kemarin penyelidikan terkait obat yg menyebabkan gagal ginjal akut pada anak itu blm tuntas atau gmn
Sampai kapankah kasus anak dg gagal ginjal akut progresif atipikal ggapa akan berakhir Kita harus bisa pastikan bahwa sirop obat yg bermasalah sudah tidak ada lagi di layanan kesehatan atau masyarakat	sampai kapankah kasus anak dg gagal ginjal akut progresif atipikal ggapa akan berakhir kita harus bisa pastikan bahwa sirop obat yg bermasalah sudah tidak ada lagi di layanan kesehatan atau masyarakat

3. *Stemming*

Selanjutnya masuk kedalam proses *stemming* proses ini bertujuan untuk mengubah kata-kata yang memiliki kata imbuhan menjadi bentuk kata dasar seperti menghilangkan me-, ke-, di- dll. Proses ini menggunakan *library sastrawi* dimana *library* ini berfungsi untuk melakukan pengurangan kata dengan mengubahnya dalam bentuk baku yang terdapat pada kamus bahasa Indonesia. Hasil yang dilakukan pada proses *stemming* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Proses *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
wah gmn ini bpom kerjanya baru ada korban meninggal baru diberhentikan kemarin penyelidikan terkait obat yg menyebabkan gagal ginjal akut pada anak itu blm tuntas atau gmn	wah gmn ini bpom kerja baru ada korban tinggal baru henti kemarin lidi kait obat yg sebab gagal ginjal akut pada anak itu blm tuntas atau gmn
sampai kapankah kasus anak dg gagal ginjal akut progresif atipikal ggapa akan berakhir kita harus bisa pastikan bahwa sirop obat yg bermasalah sudah tidak ada lagi di layanan kesehatan atau masyarakat	sampai kapan kasus anak dg gagal ginjal akut progresif atipikal ggapa akan akhir kita harus bisa pasti bahwa sirop obat yg masalah sudah tidak ada lagi di layan sehat atau masyarakat

4. Tokenizing

Proses selanjutnya dilakukan *tokenize* yaitu proses perubahan kalimat menjadi potongan kata atau token. Pada Tabel 5 merupakan hasil yang dilakukan proses *tokenizing* dengan menunjukkan kalimat yang sudah diubah menjadi per kata.

Tabel 5. Hasil Proses *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
wah gmn ini bpom kerja baru ada korban tinggal baru henti kemarin lidi kait obat yg sebab gagal ginjal akut pada anak itu blm tuntas atau gmn	['wah', 'gmn', 'ini', 'bpom', 'kerja', 'baru', 'ada', 'korban', 'tinggal', 'baru', 'henti', 'kemarin', 'lidi', 'kait', 'obat', 'yg', 'sebab', 'gagal', 'ginjal', 'akut', 'pada', 'anak', 'itu', 'blm', 'tuntas', 'atau', 'gmn']
sampai kapan kasus anak dg gagal ginjal akut progresif atipikal ggapa akan akhir kita harus bisa pasti bahwa sirop obat yg masalah sudah tidak ada lagi di layan sehat atau masyarakat	['sampai', 'kapan', 'kasus', 'anak', 'dg', 'gagal', 'ginjal', 'akut', 'progresif', 'atipikal', 'ggapa', 'akan', 'akhir', 'kita', 'harus', 'bisa', 'pasti', 'bahwa', 'sirop', 'obat', 'yg', 'masalah', 'sudah', 'tidak', 'ada', 'lagi', 'di', 'layan', 'sehat', 'atau', 'masyarakat']

5. Stopword

Selanjutnya dilakukan tahapan *stopword* pada tahapan ini dilakukan untuk menghilangkan atau untuk menghapus kata-kata yang sudah tidak dibutuhkan pada dataset. Proses yang dilakukan yaitu dengan memanfaatkan fungsi *stopwords.words* ('Indonesia') yang terdapat pada *library NLTK* sebagai acuan untuk menghilangkan kata-kata yang sudah tidak dibutuhkan dan hasil dari pemrosesan tersebut dapat dilihat pada Tabel 6 perbandingan sebelum melakukan *stopword* dan setelah melakukan *stopword*.

Tabel 6. Hasil Proses *Stopword*

Sebelum <i>Stopword</i>	Sesudah <i>Stopword</i>
-------------------------	-------------------------

['wah', 'gmn', 'ini', 'bpom', 'kerja', 'baru', 'ada', 'korban', 'tinggal', 'baru', 'henti', 'kemarin', 'lidi', 'kait', 'obat', 'yg', 'sebab', 'gagal', 'ginjal', 'akut', 'pada', 'anak', 'itu', 'blm', 'tuntas', 'atau', 'gmn']	['gmn', 'bpom', 'kerja', 'korban', 'tinggal', 'henti', 'kemarin', 'lidi', 'kait', 'obat', 'yg', 'gagal', 'ginjal', 'akut', 'anak', 'blm', 'tuntas', 'gmn']
['sampai', 'kapan', 'kasus', 'anak', 'dg', 'gagal', 'ginjal', 'akut', 'progresif', 'atipikal', 'ggapa', 'akan', 'akhir', 'kita', 'harus', 'bisa', 'pasti', 'bahwa', 'sirop', 'obat', 'yg', 'masalah', 'sudah', 'tidak', 'ada', 'lagi', 'di', 'layan', 'sehat', 'atau', 'masyarakat']	['anak', 'dg', 'gagal', 'ginjal', 'akut', 'progresif', 'atipikal', 'ggapa', 'sirop', 'obat', 'yg', 'layan', 'sehat', 'masyarakat']

6. Normalization

Setelah melalui tahapan *stopword*, langkah terakhir yang akan dilakukan pada proses *preprocessing* adalah melakukan *normalization*, tahapan ini dengan melakukan perubahan dari kata-kata yang tidak baku seperti kata singkatan menjadi kata-kata baku dengan menggunakan *Indonesian lexicon based*, kamus alay, KKBA dan *library python KBBI*. Hasil yang didapatkan dari proses normalisasi dapat ditunjukkan pada Tabel 7

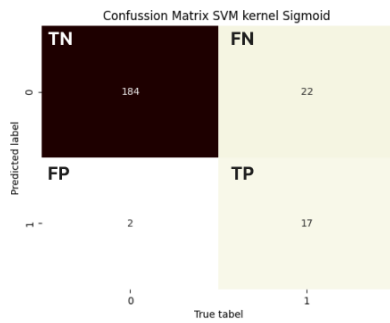
Tabel 7. Hasil dari Proses *Normalization*

Sebelum <i>Normalization</i>	Sesudah <i>Normalization</i>
gmn bpom kerja korban tinggal henti kemarin lidi kait obat yg gagal ginjal akut anak blm tuntas gmn	wah bagaimana ini bpom kerja baru ada korban tinggal baru henti kemarin lidi kait obat yang sebab gagal ginjal akut pada anak itu belum tuntas atau bagaimana
anak dg gagal ginjal akut progresif atipikal ggapa sirop obat yg layan sehat masyarakat	anak dengan gagal ginjal akut progresif atipikal ggapa sirop obat yang layan sehat masyarakat

3.3. Pelabelan Data

Setelah dilakukannya tahap pre-processing, data yang didapat sekitar 1128 data. Proses pelabelan dilakukan dengan memisahkan data menjadi dua kelas yaitu sentimen positif dan negatif yang dilakukan secara otomatis dengan memanfaatkan *vader lexicon*. proses pelabelan diawali dengan mengubah data pada kolom *tweet* dari teks bahasa Indonesia ke bahasa Inggris dengan menggunakan *library googletrans* untuk dapat digunakan oleh fungsi *nlk.download* ('vader_lexicon') yang hanya dapat membaca teks bahasa Inggris saja Selanjutnya dilakukan proses penskoran pada teks *tweet* dengan menggunakan fungsi *sid.polarity_scores* untuk mengetahui berapa skor yang didapat pada satu teks untuk dapat dilihat teks tersebut bersifat positif atau negatif.

Setelah dilakukannya proses skoring, lalu dilakukan proses pelabelan dengan melihat nilai *compound* yang sudah didapat dan mengubah nilai tersebut apabila nilai *compound* lebih dari 0 maka sentimen yang akan didapat adalah sentimen positif dan sebaliknya apabila kurang dari 0 maka sentimen yang didapat yaitu sentimen negatif. Sehingga hasil yang didapat yaitu sentimen positif berjumlah 194 data sedangkan sentimen negatif berjumlah 934 data, berikut ilustrasi perbandingan antara sentimen positif dengan sentimen negatif yang ditunjukkan pada Gambar 3.



Gambar 6. Visualisasi Confusion Matrix SVM Kernel Sigmoid

3.6. Evaluasi

Pada tahapan ini dilakukan pengujian evaluasi terhadap data uji agar mendapatkan hasil *accuracy*, *precision*, *recall* dan *f1-score* yang dihitung menggunakan persamaan (6), (7), (8) dan (9) dan hasilnya dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengujian Setiap Kernel SVM

Evaluasi	Kernel			
	Linear	RBF	Polynomial	Sigmoid
Accuracy	91%	90%	87%	89%
Precision	91%	100%	100%	89%
Recall	51%	41%	23%	44%
F1-score	66%	58%	38%	59%

Dari hasil pengujian evaluasi pada keempat kernel pada algoritma *support vector machine* didapatkan bahwa algoritma yang menggunakan kernel *linear* mendapatkan hasil *accuracy* tertinggi yaitu sebesar 91 % dengan hasil *precision* sebesar 91%, hasil *recall* sebesar 51% dan hasil dari *f1-score* sebesar 66%. Hasil tersebut dapat menjelaskan bahwa sebaran data yang digunakan sudah berbentuk *linear* sehingga kernel tidak perlu melakukan perubahan jarak antara sebaran datanya dan itu membuat garis pemisah dapat langsung dibuat. Sementara itu, hasil *accuracy* yang didapatkan oleh kernel *polynomial* sangat rendah dikarenakan kernel ini tidak cocok untuk sebaran data *linear*. Sehingga dapat dilihat bahwa penggunaan kernel sangat mempengaruhi hasil *accuracy*.

4. KESIMPULAN

Menurut hasil dari penelitian yang telah dilakukan, maka didapatkan beberapa kesimpulan yaitu Hasil perbandingan sentimen yang didapat dari dataset sebanyak 1128 tweets mengenai obat penyebab gagal ginjal akut pada anak bahwa sentimen positif mendapatkan hasil sebanyak 194 data dan sentimen negatif mendapatkan hasil sebanyak 934 data. Hal ini membuktikan bahwa masyarakat cenderung menanggapi topik ini dengan negatif sehingga data yang digunakan menjadi data yang *imbalance*. Dari pembagian data 80:20 menggunakan algoritma *support vector machine* pada kernel *linear* menghasilkan kinerja yang sangat baik untuk menganalisis sentimen pada kasus obat penyebab gagal ginjal akut pada anak dengan *accuracy* 91%, *precision* 91%, *recall* 51% dan *f1-score* 66%.

DAFTAR PUSTAKA

- Alamsyah, M. K., & Pratiwi, N. (2024). Analisis Sentimen Terkait Opini Masyarakat Terhadap Perkembangan E-Sport Mobile Di Indonesia Menggunakan K Nearest Neighbor. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 349–359. <https://doi.org/10.29100/jipi.v9i1.4927>
- CNNIndonesia. (2022). *LIVE UPDATE: Obat Sirop Kelebihan EG dan DEG Ditarik dari Peredaran*. CNN Indonesia.
- DataIndonesia. (2023). *Indonesia Masuk Negara Paling Banyak Main Twitter pada Awal 2023*. Data Indonesia.
- Fitri Wulandari, Elin Haerani, Muhammad Fikry, & Elvia Budianita. (2023). Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 88–96. <https://doi.org/10.37859/coscitech.v4i1.4781>
- Hasibuan, M. S., & Serdano, A. (2022). *Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes Policy Sentiment Analysis Face-to-face Learning Using Supports Vector and Naive Bayes Engines*. 6(2), 199–204.
- Kemenkes RI. (2022). *Waspada! Gagal Ginjal Akut pada Anak*. Kemenkes.Go.Id.
- Kevin, V., Que, S., Iriani, A., & Purnomo, H. D. (2020). *Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)*. 9(2), 162–170.
- Lestari, U., Romadhani, T., Suraya, S., & Fatkhayah, E. (2022). Sentiment Analysis for Extracting Student Opinion Data on Higher Education Services Using the Naive Bayes Classifier and Support Vector Machine Methods (Case Study Akprind Institute of Science and Technology Yogyakarta). *Jurnal TAM (Technology Acceptance Model)*, 13(1), 51. <https://doi.org/10.56327/jurnaltam.v13i1.1220>
- Palepa, M. J., Pratiwi, N., & Rohmansa, R. Q. (2024). Analisis Sentimen Masyarakat Tentang Pengaruh Politik Identitas Pada Pemilu 2024 Terhadap Toleransi Beragama Menggunakan Metode K - Nearest Neighbor. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 389–401. <https://doi.org/10.29100/jipi.v9i1.4957>
- Romadoni, F., Umaidah, Y., & Sari, B. N. (2020). Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(2), 247–253. <https://doi.org/10.32736/sisfokom.v9i2.903>
- Sari, B. W., & Haranto, F. F. (2019). Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet. *Jurnal Pilar Nusa Mandiri*, 15(2), 171–176. <https://doi.org/10.33480/pilar.v15i2.699>

- Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 43–51. <https://doi.org/10.31603/komtika.v5i1.5189>
- Shiddicky, A., & Agustian, S. (2022). Jurnal Computer Science and Information Technology (CoSciTech) menggunakan metode logistic regression. *Jurnal Computer Science and Information Technology (CoSciTech)*, 3(2), 91–98.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67. <https://doi.org/10.47080/simika.v5i1.1411>
- Tuhuteru, H., & Iriani, A. (2018). Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(3), 394–401. <https://doi.org/10.30591/jpit.v3i3.977>