Jurnal Komputer dan Informatika
Vol 16 No 2, Oktober 2021: hlm 76 - 83

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

KLASIFIKASI *CYBERBULLYING TWEET* MENGGUNAKAN ALGORITMA SVM DAN XGBOOST

Felix Fernando
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
Email**:** *felix.535210052@stu.untar.ac.id*

ABSTRAK

Penelitian ini bertujuan untuk melakukan klasifikasi *tweet cyberbullying* dengan menggunakan dua pendekatan *Machine Learning* yaitu algoritma SVM dan XGBoost. Data input yang digunakan untuk analisis merupakan hasil pengambilan data *tweet* secara acak yang telah dilabeli, dilakukan ekstraksi fitur, dan pelatihan serta pengujian model menggunakan kedua algoritma. Hasil eksperimen menunjukkan bahwa XGBoost memberikan kinerja yang sedikit lebih baik dalam mengklasifikasikan *tweet cyberbullying* dibandingkan SVM. Temuan ini memberikan kontribusi dalam pengembangan metode klasifikasi untuk mendeteksi *cyberbullying* di media sosial, yang dapat membantu dalam memitigasi dampak negatifnya. Penelitian ini juga menunjukkan potensi penggunaan algoritma XGBoost dalam konteks deteksi *cyberbullying* di berbagai platform media sosial lain seperti TikTok ataupun Instagram.

Kata Kunci: *Cyberbullying*, Klasifikasi Teks, *Twitter*, SVM, XGBoost.

ABSTRACT

This study aims to classify cyberbullying tweets using two machine learning approaches, namely the SVM and XGBoost algorithms. The input data used for analysis consists of randomly collected tweets that have been labeled, feature extraction, and model training and testing using both algorithms. The results of the experiment show that XGBoost performs slightly better in classifying cyberbullying tweets than SVM. These findings contribute to the development of classification methods for detecting cyberbullying on social media, which can help mitigate its negative impact. This study also shows the potential use of the XGBoost algorithm in the context of cyberbullying detection on various other social media platforms such as TikTok and Instagram.

Keywords: *Cyberbullying, Text Classification, Twitter, SVM, XGBoost.*

1. PENDAHULUAN

Perkembangan teknologi digital dan akses internet telah mempermudah pelaku *bullying* untuk mencari korbannya secara *online* [1]. Dalam 5 tahun terakhir, persentase terjadinya kasus *cyberbullying* telah meningkat, pada kisaran 6.0% hingga 46.3% [2]. Para peneliti memiliki pandangan yang berbeda-beda terhadap definisi *cyberbullying*. Namun, mereka sepakat bahwa perilaku yang bersifat mengolok- olok seperti pelecehan, penistaan, dan penghinaan yang terjadi di ruang digital terklasifikasikan sebagai *cyberbullying* [3]. Berdasarkan penelitian yang telah dilakukan oleh [4] ditemukan bahwa *cyberbullying* memiliki dampak negatif yang lebih parah jika dibandingkan dengan *bullying* tradisional. Hal ini disebabkan oleh video dan gambar yang dapat digunakan sebagai media untuk melecehkan, menista, dan menghina orang lain [5].

Pada tahun 2019, survei yang sama juga pernah dilakukan oleh kata.co.id, dan ditemukan bahwa pengguna internet yang terdaftar adalah sebanyak 130 juta pengguna [6]. Salah satu kontributor terbesar terhadap peningkatan tersebut adalah remaja berusia 15-19 tahun, dengan keterampilan teknologi mereka yang mempermudah mereka untuk mendapatkan informasi dari internet. Rasa keingintahuan pada fase labil inilah yang menyebabkan *cyberbullying* menjadi masalah yang serius [7].

Penelitian ini bertujuan untuk memberikan kontribusi dalam memitigasi dampak negatif dari *cyberbullying* karena di Indonesia kasus *cyberbullying* masih sulit untuk terungkap. Fase labil mereka menyebabkan remaja sangat rentan terhadap berbagai kenakalan dan penyimpangan, termasuk perilaku *bullying* baik secara tradisional maupun digital [8].

2. TINJAUAN LITERATUR

Cyberbullying merupakan tindakan intimidasi, pelecehan, atau ancaman yang dilakukan melalui media sosial, seperti *Twitter*. Tinjauan literatur akan membahas penelitian-penelitian terdahulu yang telah dilakukan dalam bidang klasifikasi teks dan deteksi *cyberbullying*, termasuk penggunaan metode SVM (*Support Vector Machine*) dan XGBoost. Beberapa penelitian mungkin telah menggunakan metode-metode ini dalam konteks deteksi *cyberbullying*, dan tinjauan literatur akan mengidentifikasi kesenjangan pengetahuan yang dapat diisi oleh penelitian ini. Beberapa penelitian terkait telah dilakukan sebelum pembuatan tulisan ini, diantaranya merupakan analisis sentimen *tweet e-learning* yang menggunakan metode SVM dengan nilai akurasi sebesar 53.03% [9], dan klasifikasi tweet *cyberbullying* menggunakan *SentiStrength* dengan nilai akurasi sebesar 60.5% [10].

Karya ilmiah milik [11] menunjukkan bahwa SVM dapat dengan baik mengklasifikasi jenis pantun dengan menggunakan fitur maksimum 1000 dan jumlah dataset 470 data pantun. Penelitian ini menggunakan SVM untuk mengklasifikasi jenis pantun berdasarkan data latih dan uji, dan menunjukkan bahwa SVM dapat mencapai akurasi sebesar 81,91%. Pantun anak menunjukkan nilai *precision*, *recall*, dan spesifisitas yang lebih tinggi dibandingkan pantun tua dan pantun muda, yaitu sebesar 90,63%, 87,88%, dan 95,08%.

3. METODE PENELITIAN

Penelitian ini dilakukan dengan menggunakan data tweet yang diperoleh dari website kaggle. Algoritma yang digunakan untuk melakukan klasifikasi pada penelitian ini adalah *Support Vector Machine* (SVM), yang menemukan fungsi klasifikasi untuk membagi data menjadi dua kelas yang berbeda. Selain itu, penelitian ini juga akan menggunakan *eXtreme Gradient Boost* (XGBoost), sebuah algoritma pengembangan dari *gradient boosting* sebagai pembanding kinerja dengan SVM.

3.1 Pengolahan Data

Data yang digunakan memiliki sampel sebanyak 47692 dengan teks tweet dan label tipe cyberbullying, yang dapat dilihat pada Tabel 1. Sebelum melakukan klasifikasi, adapun tahapan ini yang berfungsi untuk melakukan transformasi pada data mentah menjadi data yang terstruktur dan siap diproses. Pada tahapan ini, beberapa pra-proses yang akan dilakukan antara lain *case folding*, *filtering*, *tokenizing*, dan *stemming*. Selain itu juga, *filtering* diperlukan untuk membersihkan elemen yang tidak dibutuhkan atau tidak memiliki korelasi dengan *cyberbullying* pada teks, pada kasus ini elemen-elemen tersebut merupakan URL, simbol, angka, huruf yang berulang, dan kata sambung [12].

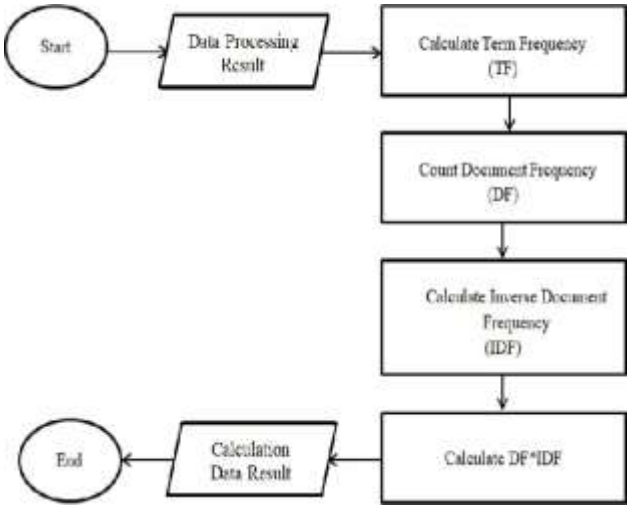
Tokenizing atau *tokenization* berguna untuk memisahkan kalimat menjadi satuan kata dengan tujuan agar setiap kata dapat dibobotkan dan mempermudah proses klasifikasi dan mengurangi *cost* komputatif secara signifikan. *Stemming* merupakan proses penyederhanaan kata dan menghilangkan imbuhannya. Tahap ini merupakan proses terpenting pada pra-proses teks dikarenakan hasil dari tahap ini dapat mempengaruhi langsung pada hasil akhir klasifikasi [12].

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Tabel 1. Contoh Data *Tweet*

tweet_text	cyberbullying_type
In other words #katandandre, your food was crapilicious! #mkr	not_cyberbullying
Hey dumb fuck celebs stop doing something for people for publicity on Facebook... Wtf happen to life u niggers are cowards.	ethnicity
@discerningmumin Islam has never been a resistance to oppression. It has always been source of oppression to both believers and non believer	religion
Here at home. Neighbors pick on my family and I. Mind you my son is autistic. It feels like high school. They call us names attack us for no reason and bully us all the time. Can't step on my front porch without them doing something to us	age
Masters @ManhattaKnight I mean he's gay, but he uses gendered slurs and makes rape jokes	gender
@ikralla fyi, it looks like I was caught by it. I'm not a botter, so...	other_cyberbullying

Sebelum proses *fitting*, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) akan digunakan untuk menghitung bobot atau frekuensi munculnya setiap kata. Kata diambil dari data input yang sudah berbentuk token akan ditransformasi dari *string* panjang menjadi vektor agar dapat diproses sebagai input [12]. Proses pembobotan fitur TF-IDF dapat dilihat pada Gambar 1.



Gambar 1. Skema Pembobotan TF-IDF

Vektorisasi TF-IDF melakukan perhitungan dengan rumus sederhana yang dapat dilihat pada Persamaan 1 dan Persamaan 2.

$$W_{ij} = tf_{ij} \times idf \quad (1)$$

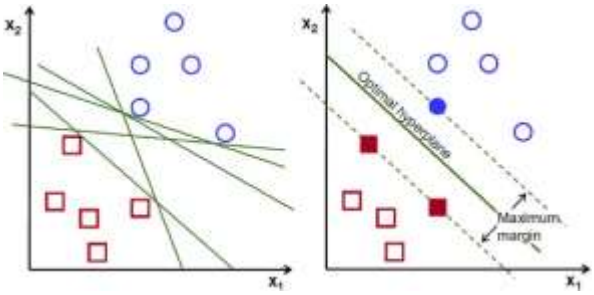
$$idf = \log\left(\frac{N}{df_j}\right) \quad (2)$$

Keterangan:
W = Bobot dokumen TF-IDF
N = Total dokumen
i = Jumlah variabel
j = Jumlah data
t = Frekuensi variabel di dalam data [12].

Support Vector Machine (SVM) merupakan salah satu algoritma klasifikasi, yang karakteristiknya adalah mencari *hyperplane*, yaitu garis yang memisahkan data antar kelas atau kategori. SVM mencari hyperplane yang memiliki margin terbesar, yaitu jarak antar data tiap kelas terhadap *hyperplane* yang diilustrasikan pada Gambar 2. Data yang paling dekat akan disebut sebagai *support vector* [11]. SVM melakukan perhitungan hyperplane dengan Persamaan 3 [13].

$$w.x + b = 0 \quad (3)$$

Keterangan:
W = Parameter *hyperplane* yang dicari
x = Data input
b = Nilai bias



Gambar 2. *Hyperplane* dan *Maximum Margin*

Margin yang dicari oleh *hyperplane* dapat dimaksimalkan dengan menggunakan sebuah persamaan garis yang dapat di lihat pada Persamaan 4 [13].

$$ax + by + c = 0, \quad (4)$$

Persamaan garis awal diubah dengan mengganti variabel x menjadi x1 dan variabel y menjadi x2. Konstanta x diubah menjadi w1 dan konstanta b diubah menjadi w2. Persamaan tersebut kemudian diubah untuk dimensi d menjadi persamaan yang lebih umum, yaitu pada Persamaan 5 dan Persamaan 6 [13].

$$\sum_{j=1}^{d} = w_j x_i + c = 0, \quad (5)$$

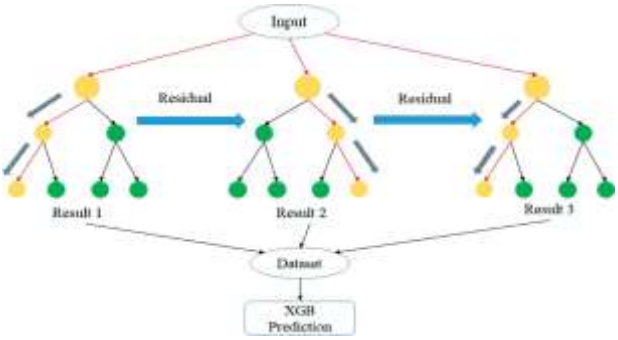
$$g(x) = < w, x > + c = 0, \quad (6)$$

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

XGBoost (*eXtreme Gradient Boosting*) merupakan salah satu algoritma *machine learning*, yang dapat digunakan untuk klasifikasi dan prediksi. XGBoost menggunakan beberapa *decision tree* sebagai metode *boosting*-nya yang dapat diilustrasikan dengan Gambar 3. XGBoost merupakan algoritma yang cukup baik bekerja pada data berukuran kecil ke sedang dengan bobot yang pada setiap pohon keputusan dihitung berdasarkan persamaan logistk yang dapat didefinisikan melalui Persamaan 7.

$$\dot{y}_i = \sum_k^k f_k(x_i), f_k \in F\ , \tag{7}$$

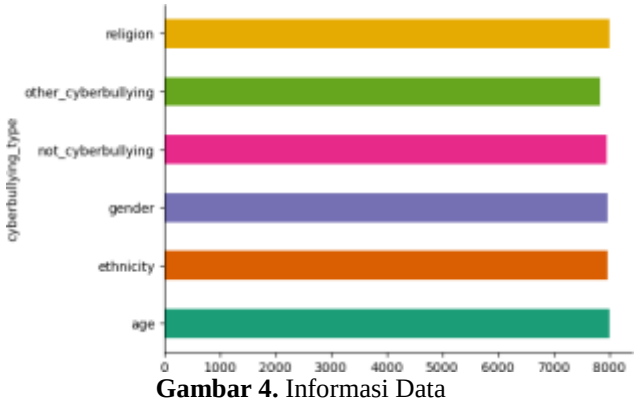
Keterangan:
$\dot{y}_i$ = Prediksi untuk observasi ke-*i*
f_k = Pohon ke-*k*
x_i = Vektor fitur untuk observasi ke-*i*
F = Himpunan pohon yang membentuk model



Gambar 3. Konsep *Boosting* XGBoost

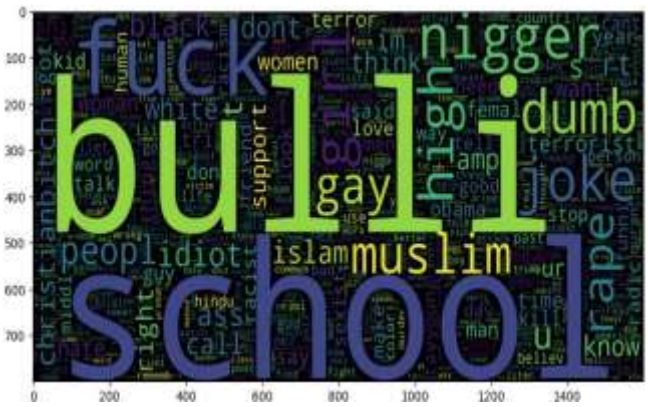
4. HASIL DAN PEMBAHASAN

Setelah klasifikasi dengan kedua algoritma selesai dilakukan, hasil evaluasi keduanya dari *classification report* yang dihitung menggunakan *confusion matrix* akan dibandingkan. Dari sampel data yang digunakan, dapat dilihat pada Gambar 4 bahwa terdapat enam kelas dalam kategori tweet, yaitu not_*cyberbullying* untuk *tweet* yang tidak bersifat mencela, *ethnicity* untuk *tweet* yang menghina etnis atau suku, *religion* untuk *tweet* yang menyinggung agama, *age* untuk *tweet* yang mencela berdasarkan usia, *gender* untuk *tweet* yang menghina orientasi seksual; dan other_*cyberbullying* untuk bentuk *cyberbullying* lain yang tidak termasuk dalam kategori sebelumnya. Jumlah data telah terdistribusi secara merata untuk setiap kelas. Skenario eksperimen pada penelitian ini membagi data menjadi dua bagian yaitu data *testing* dan *training* dengan rasio 80:20 agar tidak terjadi *overfitting* ataupun *underfitting*.



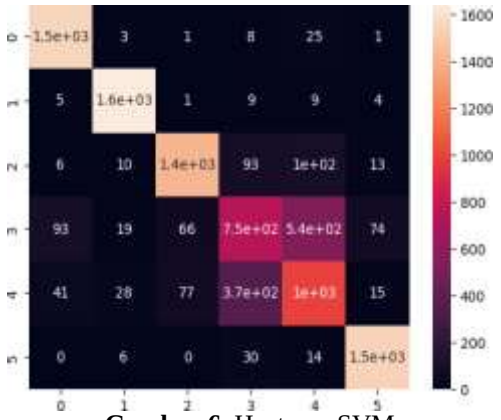
Gambar 4. Informasi Data

Hasil dari tahapan pra-proses (*stemming*) telah menghasilkan statistik jumlah kemunculan kembali setiap kata pada dataset. Total kata yang menjadi fitur adalah sebanyak 329778 kata, dan statistik ini dapat direpresentasikan menggunakan plot *Word Cloud*, yang dapat dilihat pada Gambar 5.



Gambar 5. *Word Cloud* Data *Tweet*

Pada algoritma SVM dengan menggunakan kernel linear akurasi yang dihasilkan adalah 82.65%. Hasil evaluasi kinerja algoritma SVM yang dihitung menggunakan *confusion matrix* dapat dilihat melalui *classification report* dan *heatmap* pada Gambar 6 dan Tabel 2.



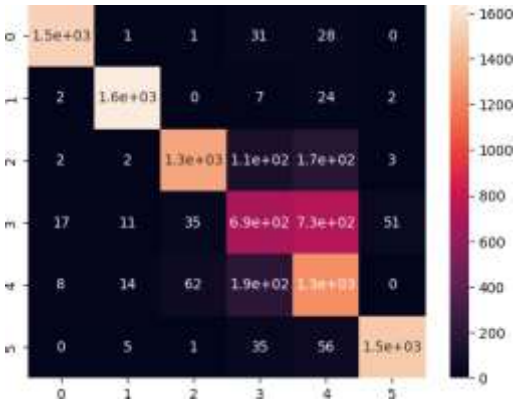
Gambar 6. Heatmap SVM

Tabel 2. *Classification Report* Algoritma SVM

	Precision	***Recall***	***F1-score***	***Support***
0	0.91	0.98	0.94	1586
1	0.96	0.98	0.97	1671
2	0.91	0.86	0.88	1635
3	0.60	0.49	0.54	1534
4	0.59	0.65	0.62	1525
5	0.93	0.97	0.95	1588
accuracy			0.83	9539
macro avg	0.82	0.82	0.82	9539
weighted avg	0.82	0.83	0.82	9539

Pada algoritma XGBoost akurasi yang dihasilkan adalah 83.24%, lebih baik dibandingkan dengan SVM. Hasil evaluasi kinerja algoritma XGBoost yang dihitung menggunakan confusion matrix dapat dilihat melalui classification report dan heatmap pada Gambar 7 dan Tabel 2.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)



Gambar 7. *Heatmap* XGBoost

Tabel 3. *Classification Report* Algoritma XGBoost

	Precision	***Recall***	***F1-score***	***Support***
0	0.98	0.96	0.97	1586
1	0.98	0.98	0.98	1671
2	0.93	0.82	0.87	1635
3	0.65	0.45	0.53	1534
4	0.55	0.82	0.66	1525
5	0.96	0.94	0.95	1588
accuracy			0.83	9539
macro avg	0.84	0.83	0.83	9539
weighted avg	0.85	0.83	0.83	9539

5. KESIMPULAN

Dalam penelitian ini, klasifikasi tweet cyberbullying telah dilakukan menggunakan algoritma SVM dan XGBoost. Hasil yang diperoleh menunjukkan bahwa SVM memberikan akurasi sebesar 82.65%, sementara XGBoost memberikan akurasi sebesar 83.24%. Dengan demikian, dapat disimpulkan bahwa XGBoost sedikit lebih unggul dalam hal akurasi dibandingkan dengan SVM. Selain itu, kelebihan XGBoost meliputi kecepatan dalam menangani data yang besar, fleksibilitas dalam penggunaan untuk berbagai tujuan, dan kemampuannya dalam mengurangi overfitting.

Namun, XGBoost juga memiliki kekurangan, seperti kompleksitas yang lebih tinggi dibandingkan dengan SVM dan jumlah parameter yang perlu diatur.Untuk pengembangan selanjutnya, disarankan untuk melakukan optimasi parameter model guna meningkatkan akurasi dan kinerja, serta mempertimbangkan penggunaan teknik ensemble learning untuk menggabungkan keunggulan dari kedua model.

DAFTAR PUSTAKA

[1] Zhu, C., Huang, S., Evans, R. and Zhang, W., 2021. Cyberbullying among adolescents and children: a comprehensive review of the global situation, risk factors, and preventive measures. Frontiers in public health, 9, p.634909.

[2] Barlett, C.P., Simmers, M.M., Roth, B. and Gentile, D., 2021. Comparing cyberbullying prevalence and process before and during the COVID-19 pandemic. The journal of social psychology, 161(4), pp.408-418.

[3] Syah, R. and Hermawati, I., 2018. Upaya pencegahan kasus cyberbullying bagi remaja pengguna media sosial di Indonesia. Jurnal Penelitian Kesejahteraan Sosial, 17(2), pp.131-146.

[4] Febriana, T. and Budiarto, A., 2019, August. Twitter dataset for hate speech and cyberbullying detection in Indonesian language. In 2019 International Conference on Information Management and Technology (ICIMTech) (Vol. 1, pp. 379-382). IEEE.

[5] Jubaidi, M. and Fadilla, N., 2020. Pengaruh Fenomena Cyberbullying Sebagai Cyber-Crime di Instagram dan Dampak Negatifnya. Shaut Al-Maktabah: Jurnal Perpustakaan, Arsip dan Dokumentasi, 12(2), pp.117-134.

[6] J. Wang, K. Fu, C.T. Lu, "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection," Proceedings of the 2020 IEEE International Conference on Big Data (IEEE BigData 2020), December 10-13, 2020.

[7] Rahman, O.H., Abdillah, G. and Komarudin, A., 2021. Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 5(1), pp.17-23.

[8] Mascio, A., Kraljevic, Z., Bean, D., Dobson, R., Stewart, R., Bendayan, R. and Roberts, A., 2020. Comparative analysis of text classification approaches in electronic health records. arXiv preprint arXiv:2005.06624.

[9] Geofany, N. and Liza, R., 2021, October. Klasifikasi Sentimen Tweet Pada Twitter Terhadap Pembelajaran E-Learning Menggunakan Metode k-Nearest Neighbor. In Seminar Nasional Teknologi Informasi & Komunikasi (Vol. 1, No. 1, pp. 380-385).

[9] Khaira, U., Johanda, R., Utomo, P.E.P. and Suratno, T., 2020. Sentiment analysis of cyberbullying on twitter using SentiStrength. Indones. J. Artif. Intell. Data Min, 3(1), p.21.

[11] Irmanda, H.N. and Astriratma, R., 2020. Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM). Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 4(5), pp.915-922.

[12] Rosid, M.A., Fitrani, A.S., Astutik, I.R.I., Mulloh, N.I. and Gozali, H.A., 2020, June. Improving text preprocessing for student complaint document classification using sastrawi. In IOP Conference Series: Materials Science and Engineering (Vol. 874, No. 1, p. 012017). IOP Publishing.

[13] Fikriani, A., Asror, I. and Murti, Y.R., 2019. Klasifikasi Kepribadian Berdasarkan Data Twitter dengan Menggunakan Metode Support Vector Machine. eProceedings of Engineering, 6(3). Application of Computer Sciences, 1(1), pp.27-32.