

KEMUNGKINAN DEPRESI DARI POSTINGAN PADA MEDIA SOSIAL

Siti Mutmainah

Program Studi Informatika

Universitas Islam Indonesia

Jl. Kaliurang 14,5 Sleman, Yogyakarta,
Indonesia 55584

20917057@students.uii.ac.id

ABSTRAKSI

Salah satu gangguan kesehatan mental yaitu depresi, Depresi merupakan gangguan kesehatan dengan kontibutor terbesar penyebab kecacatan serta alasan utama bunuh diri global. Menurut studi WHO, diperkirakan ada 322 juta orang menderita depresi di dunia. Survey di Indonesia sendiri depresi merupakan gangguan kesehatan mental yang menduduki urutan pertama dalam tiga dekade (1990-2017). Pada masa pandemi memberikan banyak dampak negatif, Penerapan sosial distancing dan *work from home* dan pembatasan menyebabkan meningkatnya kecemasan, stress dan depresi, Kegiatan banyak dilakukan secara daring dan aktivitas pada media sosial juga meningkat. Pengguna mengekspresikan diri melalui postingan media sosial. Makalah ini menyajikan hasil dari penelitian yang mempelajari postingan dari media sosial yang terindikasi mengalami gangguan depresi atau normal. Analisis kemungkinan depresi dari postingan media sosial menggunakan teknik *Natural Language Processing* (NLP) Pemodelan dilakukan dengan teknik klasifikasi yang sering dipakai dalam *sentiment* analisis dan pada kasus yang sama yaitu prediksi depresi. hasil terbaik didapatkan oleh metode SVM dengan akurasi 95.56%, hal ini sejalan dengan literatur reviu yang mendapatkan performa terbaik menggunakan metode SVM dalam *sentiment* analisis dan analisis depresi. Pemodelan menggunakan metode Neural network akurasinya 93.79%.

Kata Kunci

Classification; Text Analysis; Depression; Media Sosial

1. PENDAHULUAN

Gangguan kesehatan mental merupakan kondisi dimana seseorang kesulitan untuk menyesuaikan diri dengan sekitar dan ketidakmampuan memecahkan masalah sehingga menimbulkan stress yang berlebihan [1]. Salah satu gangguan Kesehatan mental yaitu depresi, Depresi merupakan gangguan kesehatan dengan kontibutor terbesar yang menjadi penyebab kecacatan serta alasan utama bunuh diri global. Menurut studi yang dilakukan oleh WHO, diperkirakan ada 322 juta orang menderita depresi setara dengan 4,4% dari populasi dunia dan hampir setengah dari individu yang berisiko menderita depresi tinggal di Asia Tenggara sebanyak 27% dan wilayah Pasifik Barat sebanyak 27% termasuk Cina dan India [2].

Survey perhitungan beban penyakit di Indonesia pada tahun 2017 gangguan kesehatan mental mencapai (13.4%), Depresi merupakan gangguan kesehatan mental yang menduduki urutan pertama dalam tiga dekade (1990-2017). Hasil Riskedas 2018 gangguan depresi dapat dialami oleh semua kelompok usia remaja (15-24) tahun

dengan persentase 6,2%. Depresi meningkat seiring dengan pertambahan usia, kasus tertinggi pada umur diatas 75 tahun sebesar 8.9%, 65-74 sebesar 8.0% dan 55-64 sebesar 6.5%. Stigma masyarakat Indonesia sendiri gangguan Kesehatan jiwa disebabkan oleh hal tidak rasional seperti supranatural karna sihir atau kemasukan roh jahat. Akibat stigma sosial tersebut mengakibatkan individu tidak mencari bantuan psikolog atau penanganan tenaga professional yang tepat [3].

Pada masa pandemi memberikan banyak dampak negatif, Penerapan *social distancing* dan *work from home* dan pembatasan menyebabkan meningkatnya kecemasan, stress dan depresi[4]. Kegiatan banyak dilakukan secara daring dan aktivitas pada media sosial meningkat, pengguna mengekspresikan diri melalui postingan media sosial.

Postingan individu dengan gejala gangguan depresi *mayor disorder* menunjukkan peningkatan dalam menggunakan kata-kata negatif serta mengandung ekspresi kemarahan [5]. Penggunaan kata-kata positif paling sering digunakan oleh pengguna (tidak mengalami depresi) sedangkan kata yang bersentimen negatif sering digunakan untuk (tidak setuju) [6]. Post yang dianalisis yaitu topik yang berkaitan dengan curhatan seperti keluhan untuk melihat penggunaan kata-kata dan emosi dari pengguna.

Penelitian ini bermaksud untuk mempelajari postingan dari media sosial yang terindikasi mengalami gangguan depresi atau normal. Metode yang akan digunakan dalam pemodelan klasifikasi menggunakan hasil pertimbangan dari penelitian yang dilakukakan sebelumnya dalam literatur reviu seperti *sentiment* analisis dan penelitian dengan kasus yang sama.

2. LITERATUR REVIU

Literatur reviu digunakan sebagai bahan pertimbangan dalam melakukan penelitian, pembelajaran pada penelitian terdahulu dibutuhkan untuk melihat performa dan penggunaan metode pada kasus dan analisis serupa. Literatur reviu terbatas hanya pada analisis *sentiment* dan penelitian mengenai kesehatan mental atau depresi.

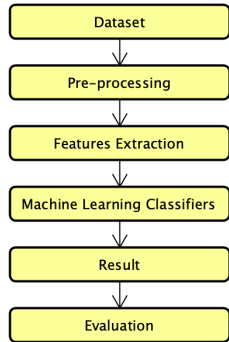
Penelitian dan kajian pustaka yang membahas *sentiment* analisis: analisis *sentiment* aplikasi ruang guru di twitter, penelitian tersebut mengambil komentar pengguna terhadap aplikasi ruang guru kemudian membandingkan metode NB, SVM, K-NN dan mendapatkan hasil terbaik menggunakan algoritma PSO berbasis SVM dan mendapat akurasi 78,55% [7]. Menganalisis sikap masyakat terhadap kebijakan sekolah daring pada media sosial twitter, menggunakan algoritma *Supervised Machine Learning*, secara spesifik *Logistic Regression* dan model terbaik dengan

akurasi 87% [6]. Membandingkan akurasi dan waktu proses algoritma KNN dan SVM dalam menganalisis *sentiment* pada twitter, hasil yang didapatkan SVM memiliki akurasi yang lebih tinggi namun metode KNN memiliki waktu proses yang lebih cepat [8].

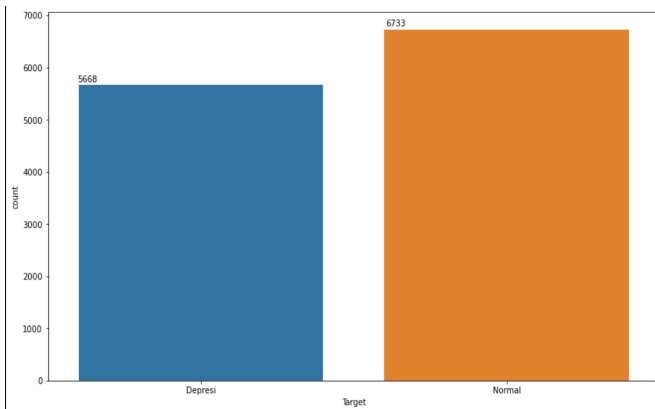
Kajian Pustaka: analisis *sentiment* ulasan produk, dalam kajian tersebut penggunaan Naive Bayes 5 penelitian dan SVM 3, serta yang membandingkan kedua metode tersebut, hasil yang didapatkan penambahan fitur dalam klasifikasi analisis *sentiment* dapat meningkatkan nilai akurasi apabila fitur tersebut sesuai dengan metode klasifikasi [9].

Literatur reviu: metode klasifikasi paling banyak dipakai yaitu naïve bayes 5, *support vector machine* 4, lexical-based 3. Akurasi *Sentiment* Analisis paling tinggi logistic regression 93.60% dan lexical-based 92%. Pemilihan fitur ekstraksi yang tepat dapat meningkatkan tingkat akurasi dalam membantu metode klasifikasi dalam menentukan sentimen [10].

Penelitian yang membahas kesehatan mental yaitu memprediksi status kesehatan mental melalui aktivitas media sosial menggunakan multimodel, Penelitian tersebut mengumpulkan data melalui platform flickr yang membahas Kesehatan mental kemudian menganalisis ekstensif visual, bahasa dan fitur postingan untuk mengidentifikasi perbedaan perilaku dalam postingan pengguna (sehat), (menderita penyakit mental) dan (rentan terhadap penyakit mental). Penelitian tersebut mendapatkan perbedaan perilaku antar kelompok, multiple modalities dapat meningkatkan kinerja tugas klasifikasi dibanding menggunakan satu *modality* dalam satu waktu [5].



Gambar 1. Tahapan Penelitian



Gambar 2. Distribusi antar class

Penelitian lain yaitu mendeteksi depresi melalui postingan pada media sosial reddit, penelitian tersebut menggunakan teknik *Natural Language Processing* (NLP) dengan pendekatan *machine learning* untuk melatih data dan mengevaluasi efisiensi menggunakan metode classifier *Support Vector Machine* (SVM) untuk mendeteksi depresi dan mendapat akurasi 80%. Penelitian tersebut juga menggabungkan fitur (LIWC+LDA+bigram) paling berhasil ditunjukkan dengan classifier *Multilayer Perceptron* (MLP) yang menghasilkan kinerja terbaik untuk deteksi depresi yang mencapai akurasi 91% [2].

Penelitian yang melakukan studi teknik analisis *sentiment* dengan data twitter pada penelitian yang telah dilakukan sebelumnya dengan berbagai metode dan algoritma serta features, dalam penelitian tersebut bertujuan untuk melihat kinerja metode pengklasifikasian dengan beragam teknik atau serta model features yang digunakan. Hasil yang didapatkan metode SVM dan MNB menghasilkan presisi terbesar dalam penggunaan banyak fitur, metode Naïve bayes, Maximum Entropy dan SVM mencapai 80% menggunakan n-gram dan bigram. Dalam *sentiment* analisis menggabungkan berbagai features dapat meningkatkan kinerja dalam berbagai kasus [11].

Pada literatur reviu lain metode yang sering digunakan dalam melakukan analisis *sentiment* yaitu NB, SVM, K-NN dan Logistic Regression, terdapat pada Tabel 1. pada penelitian dengan kasus yang sama yaitu analisis kesehatan mental menggunakan multimodel dan teknik *Natural Language Processing* (NLP) serta mengevaluasi dengan *Support Vector Machine* (SVM). Setelah melihat literatur reviu dalam melakukan analisis *sentiment* pada media sosial khususnya twitter dan deteksi depresi pada post, penelitian ini bermaksud untuk melakukan analisis kemungkinan depresi dari postingan media sosial menggunakan teknik *Natural Language Processing* (NLP) seperti pada [2] dan membandingkan metode yang paling banyak digunakan pada analisis *sentiment* seperti Naïve bayes, Logistic Regression dan *Support Vector Machine* (SVM). Pemodelan juga dilakukan dengan menggunakan metode NN (Neural network) untuk membandingkan dengan hasil akurasi tertinggi yang didapatkan oleh metode yang sudah sering digunakan.

Tabel 1. Metode Literatur

No	Metode	Jumlah
1	SVM	4
2	Naïve Bayes	3
3	K-Nearest Neighbors	2
4	Logistic Regression	2
5	MPL	1

3. METODOLOGI PENELITIAN

Pada metodologi penelitian berisi proses pemodelan yang akan dilakukan pada penelitian. Tahapan penelitian yang dilakukan dapat dilihat pada diagram air Gambar 1.

3.1 Dataset

Dataset dari aktivitas postingan tweet media sosial twitter, Data tersebut berjumlah 12402 Record data dengan 3 variabel (Id, Text, Target). Variabel yang digunakan hanya variabel (Text dan Target), variabel Target memiliki 2 class yaitu 1 (Normal) dan 0 (Depresi).



Gambar 12. Proses Train data

Table 2. Indikasi Post Depresi dan Normal

Depresi Post	Normal Post
Hate, fucking, boring, sad, stressed, depressed, die, sick, cry, stupid, damn, nothing, ugly, hard, sorry, shit, jealous, bored, boy, hurt, try, saddest, wrong, alone, lie, depression	Happy, love, good, blessed, great, friend beautiful, girl, blessing, lucky, thank, batter, make enjoy, girlfriend, pretty, family, fun, dream, allah, god

Tabel 3. Dataset

Data	Record
Train set	11160
Test set	1240

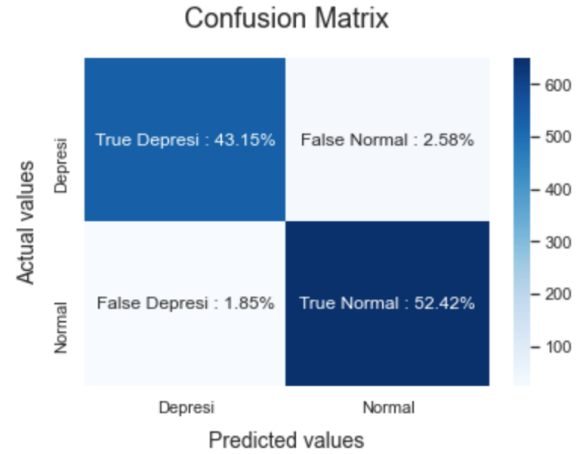
Tabel 4. Akurasi

Metode	Train	Test
NB	97.27%	93.95%
LR	96.97%	94.92%
SVM	98.57%	95.56%
NN	95.25%	93.79%

Table 2 menunjukkan kata indikasi dari kata pada postingan 0 (Depresi) dan postingan 1 (Normal). Dari pre-processing ditemukan bahwa bahasa depresi mengandung kata-kata seperti emosi negatif, pikiran bunuh diri, kemarahan, proses interpersonal, kelelahan dan sering diekspresikan secara somatik melalui gejala tubuh (*feel, want, like, wish, sad, wrong life, depression, depressed, want die, shit, fuck, hate, anyone, lonely, alone, bad, boring, sick, tired, empty, cry, stupid, damn, nothing, ugly, hard, sorry, jealous, bored*).

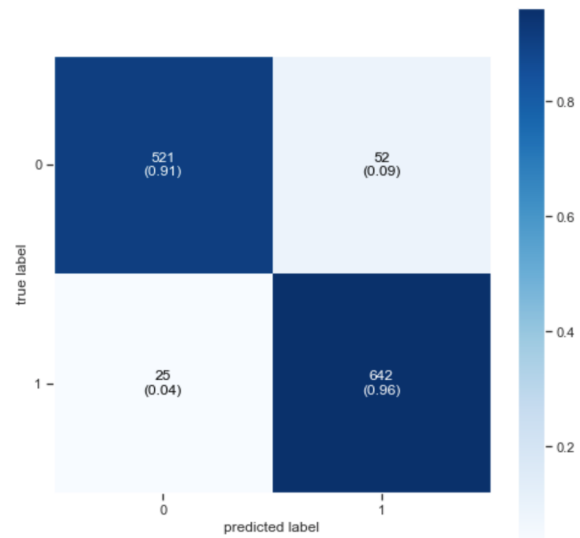
Pada Normal post lebih mencerminkan kesenangan, hubungan sosial yang baik dan rasa syukur seperti (*life, love, good, happy, blessed, good best, great, lucky, thank, batter, make enjoy, girlfriend, pretty, family, fun, dream, allah*)

	precision	recall	f1-score	support
0.0	0.96	0.94	0.95	567
1.0	0.95	0.97	0.96	673
accuracy			0.96	1240
macro avg	0.96	0.95	0.96	1240
weighted avg	0.96	0.96	0.96	1240



Gambar 13. Confusion Matrix SVM

	precision	recall	f1-score	support
0.0	0.95	0.91	0.93	573
1.0	0.93	0.96	0.94	667
accuracy			0.94	1240
macro avg	0.94	0.94	0.94	1240
weighted avg	0.94	0.94	0.94	1240



Gambar 14. Confusion Matrix NN

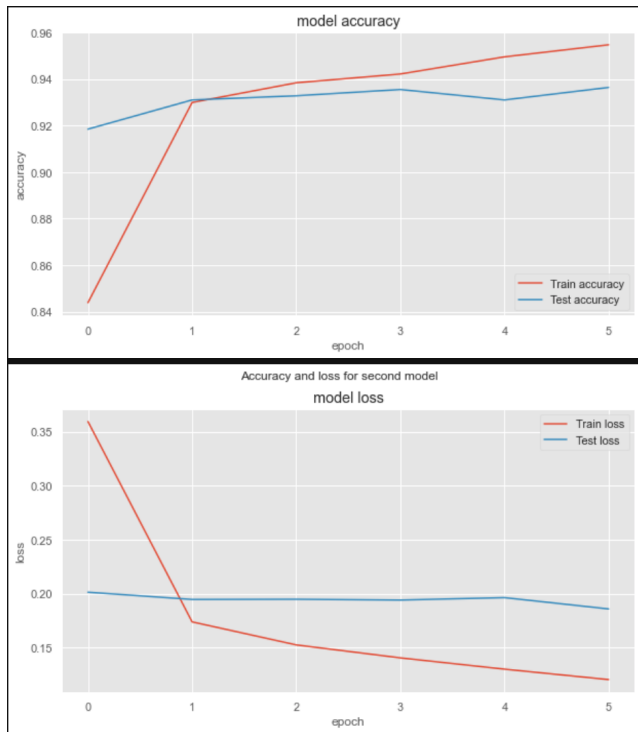
4.2 Hasil

Setelah melakukan *pre-processing* untuk membersihkan data tahap selanjutnya yaitu pemodelan, pembagian dataset untuk training dan *testing* ada pada Tabel 3. Pemodelan dilakukan menggunakan beberapa metode yang sering digunakan dalam literatur yang disajikan pada Tabel 4.

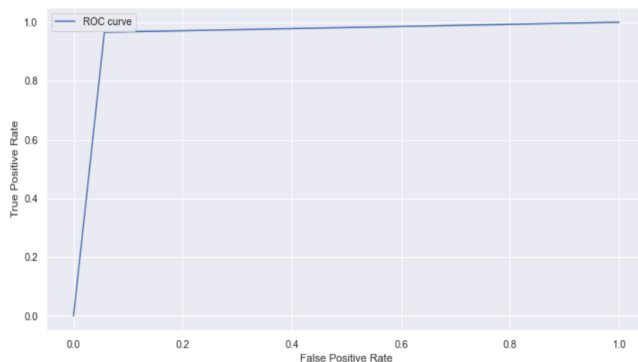
4.3 Evaluasi

Evaluasi menggunakan *Confusion Matrix*, evaluasi dilakukan untuk melihat akurasi yang didapatkan oleh metode. Pemodelan menggunakan *Support Vector Machine* (SVM) mendapat akurasi yang tinggi pada training dan *testing*, evaluasi lebih lengkap disajikan pada Gambar 13. Pada Gambar 14 merupakan Evaluasi *Confusion Matrix* yang dihasilkan menggunakan NN mendapat akurasi 93.79%, pada Gambar 15 merupakan perbandingan akurasi dan loss pada train dan *testing*, Gambar 16 dan Gambar 17 merupakan curva yang dihasilkan model SVM dan NN.

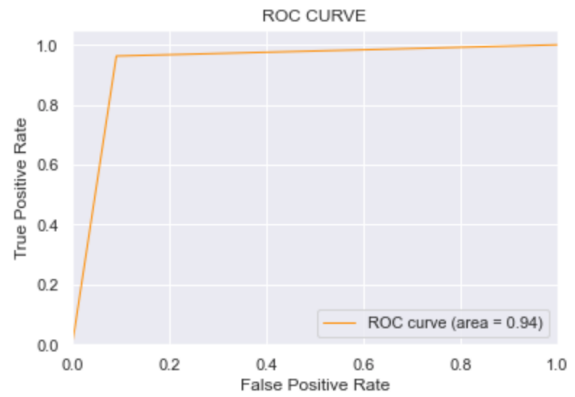
Curve ROC yang menggambarkan performa dari klasifikasi, ROC didapatkan dari *confusion matrix*.



Gambar 15. NN



Gambar 16. ROC Curve SVM



Gambar 17. ROC Curve NN

5. KESIMPULAN DAN SARAN

Pada literatur revidi terdapat beberapa metode seperti Naïve Bayes, Logistic Regression dan *Support Vector Machine* (SVM) digunakan untuk analisis *sentiment* menggunakan post pada media sosial dan pada kasus yang sama. Performa yang dihasilkan oleh metode akan berbeda dengan pemilihan fitur dan jumlah dataset yang digunakan, berdasarkan literatur penggunaan fitur gabungan akan efektif apabila menggunakan metode yang tepat. Metode yang menghasilkan performa paling tinggi dihasilkan oleh model *Support Vector Machine* (SVM).

Hal ini sejalan dengan hasil yang didapatkan pada penelitian ini yaitu analisis dari postingan media sosial yang terindikasi mengalami gangguan depresi, akurasi paling tinggi dihasilkan oleh *Support Vector Machine* (SVM) yaitu 98.57% pada training data dan 95.56% pada test data. Pada pemodelan dengan menggunakan NN (Neural network), 95.25% untuk training data dan 93.79% untuk test data.

Penelitian selanjutnya diharapkan menggunakan *Features Extraction* serta fitur yang tepat sesuai dengan metode yang akan digunakan. Mengeksplorasi fitur dan metode *machine learning* lain, apabila memiliki dataset yang besar dapat menggunakan algoritma deep learning, kemudian menambahkan parameter lebih rinci dalam melakukan analisis kemungkinan depresi pengguna dengan informasi lebih dalam dari tenaga ahli.

6. DAFTAR PUSTAKA

- [1] Putri, A. W., Wibhawa, B. dan Gutama, A. S. 2015. Kesehatan Mental Masyarakat Indonesia (Pengetahuan, Dan Keterbukaan Masyarakat Terhadap Gangguan Kesehatan Mental). *Prosiding Penelitian dan Pengabdian Kepada Masyarakat*. 2, 2, 252–258. doi: 10.24198/jppm.v2i2.13535
- [2] Tadesse, M. M., Lin, H., Xu, B. dan Yang, L. 2019. Detection of depression-related posts in reddit sosial media forum. *IEEE Access*. 7, 44883–44893. doi: 10.1109/ACCESS.2019.2909180
- [3] Pusat Data dan Informasi Kementerian Kesehatan RI. 2019. Situasi Kesehatan Jiwa DI Indonesia. InfoDATIN. 12. <https://pusdatin.kemkes.go.id/article/view/20031100001/situasi-kesehatan-jiwa-di-indonesia.html>
- [4] Lempang, G. F., Walenta, W., Rahma, K. A., Maluegha, F. J., Utomo, F. I. P. dan Retalista, N. 2021. Depresi Menghadapi Pandemi Covid-19 pada Masyarakat Perkotaan (Studi Literatur). *Jurnal Pamator*. 14, 1, 66-71. <https://journal.trunojoyo.ac.id/pamator/article/view/9854>

- [5] Xu, Z., Pérez-Rosas, V. dan Mihalcea, R. 2020. Inferring social media users' mental health status from multimodal information. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 6292–6299. <https://aclanthology.org/2020.lrec-1.772.pdf>
- [6] Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R. dan Rakhmawati, N. A. 2021. Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika dan Sistem Informasi*. 7, 1, 47–58. doi: 10.28932/jutisi.v7i1.3216.
- [7] Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L. dan Gata, W. 2020. Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*. 14, 2, 115. doi: 10.33365/jti.v14i2.679.
- [8] Nasution, M. R. A. dan Hayaty, M. 2019. Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*. 6, 2, 226–235. doi: 10.31311/ji.v6i2.5129
- [9] Kamal, W. W. dan Ratnasari, C. I. 2021. Analisis Sentimen Ulasan Produk : Kajian Pustaka. *Automata*, 2, 1.
- [10] Cindo, E. M., Rini, D. P. dan Ermatita. Literatur Review : Metode Klasifikasi Pada Sentimen Analisis. *Seminar Nasional Teknologi Komputer & Sains*. 66–70. <https://seminar-id.com/prosiding/index.php/sainteks/article/view/124>
- [11] Alsaeedi, A. dan Khan, M. Z. A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*. 10, 2, 361–374. doi: 10.14569/ijacsa.2019.0100248.
- [12] Suyanto. 2019. *Data Mining untuk Klasifikasi dan Klasterisasi Data "Edisi Revisi"*. Bandung: Informatika.