

Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia

Christevan Destitus¹, Wella², Suryasari³

^{1,2,3}Program Studi Sistem Informasi, Universitas Multimedia Nusantara, Tangerang, Indonesia
²wella@umn.ac.id

Diterima 24 Agustus 2020
 Disetujui 10 November 2020

Abstract—This study aims to clarify tweets on twitter using the Support Vector Machine and Information Gain methods. The clarification itself aims to find a hyperplane that separates the negative and positive classes. In the research stage, there is a system process, namely text mining, text processing which has stages of tokenizing, filtering, stemming, and term weighting. After that, a feature selection is made by information gain which calculates the entropy value of each word. After that, clarify based on the features that have been selected and the output is in the form of identifying whether the tweet is bully or not. The results of this study found that the Support Vector Machine and Information Gain methods have sufficiently maximum results.

Index Terms—clarification, cyberbullying, identification, information gain, support vector machine

I. PENDAHULUAN

Cyberbullying tindakan untuk melecehkan, mengancam, mempermalukan, dan mengejek orang lain pada media maya [1]. *Cyberbullying* bisa dilakukan melalui SMS, pesan teks, aplikasi, media sosial, forum, bahkan permainan online yang dimana orang lain dapat, berpartisipasi, dan berbagi konten, pada tahapan ini *cyberbullying* biasanya bertindak melalui mengirimkan, memposting, dan membagikan konten negatif, berbahaya, palsu, atau tindakan jahat terhadap orang lainnya. Ini juga mencakup berbagi informasi pribadi yang dapat menyebabkan rasa malu atau penghinaan [2].

Pelaku *cyberbullying* pada remaja biasanya terjadi karena dendam yang tidak terselesaikan, lalu motivated offender untuk melakukan pembajakan, balas dendam, pencurian, atau sekedar iseng [3]. Salah satu yang membedakan *cyberbullying* adalah waktu dan situasinya, *cyberbullying* dapat terjadi 24 jam sehari, 7 hari dalam seminggu, hal ini dapat terjadi setiap waktu, siang maupun malam [4].

Terdapat juga penelitian yang menyatakan bahwa 32% remaja mengatakan pernah melakukan *cyberbullying* dengan alasan mengisengi temannya

dan media yang paling banyak digunakan adalah situs media sosial [5].

Di sini penelitian dilakukan pada media sosial, yaitu Twitter. Twitter sendiri merupakan salah satu media sosial yang bisa digunakan sebagai situ berita dan juga jejaring sosial sebagai tempat orang berkomunikasi dalam pesan singkat yang biasanya disebut *tweet*. Twitter juga memiliki fitur *tweeting*, *tweeting* merupakan cara untuk memposting pesan singkat untuk siapa saja yang mengikut akun anda di Twitter [6].

Pada penelitian ini permasalahan tentang *cyberbullying* yang terjadi pada media sosial Twitter berupa *tweet – tweet* yang mengandung kata – kata yang berisikan konten negatif. Dari permasalahan itu data *tweet – tweet* tersebut di proses hingga menjadi data yang dapat digunakan untuk data mengenai sentiment analyst yang terjadi pada Twitter. Algoritma yang digunakan adalah *Support Vector Machine* dan juga *Information Gain*, kedua algoritma ini dapat digunakan untuk melakukan klarifikasi mengenai topik penelitian yang sedang di lakukan.

Tujuan penelitian ini untuk memberikan hasil akurasi metode *Support Vector Machine* dan *Information Gain* yang berasal dari data *tweet* yang mengandung konten *cyberbullying*.

II. TINJAUAN PUSTAKA

A. Cyberbullying

Cyberbullying adalah penyalahgunaan internet untuk melecehkan, mengancam, mempermalukan, dan mengejek orang lain [7]. *Cyberbullying* adalah tindakan perundungan yang dilakukan pada dunia maya melalui barang elektronik seperti ponsel, komputer, dan tablets [2].

Cyberbullying bisa melalui SMS, teks, aplikasi, media sosial, forum, bahkan permainan online yang dimana orang lain dapat melihat, berpartisipasi, atau berbagi konten. *Cyberbullying* juga termasuk mengirimkan, memposting, atau membagikan konten

negatif, berbahaya, palsu, atau tindakan jahat terhadap orang lain. Ini mencakup berbagi informasi pribadi yang dapat menyebabkan rasa malu atau penghinaan [2].

Cyberbullying biasanya bukan hanya komunikasi satu kali, ini “terjadi secara berulang kali”, kecuali jika itu adalah sebuah ancaman pembunuhan atau ancaman serius terhadap keselamatan orang. Pada *cyberbullying* pelaku tidak bisa melihat respon langsung dari si korban sehingga dapat mengurangi kepuasan pelaku yang didapatkan dengan melihat sakit yang ditimbulkannya pada si korban, namun hal ini juga dapat mengurangi rasa empati dari pelaku terhadap si korban [5].

Terdapat juga penelitian yang menyatakan bahwa 32 persen remaja mengatakan pernah melakukan *cyberbullying* dengan alasan mengisengi temannya dan media yang paling banyak digunakan adalah situs media sosial [5].

B. Twitter

Twitter adalah situ berita dan jejaring sosial online tempat orang berkomunikasi dalam pesan singkat yang disebut *tweet*. *Tweeting* memposting pesan singkat untuk siapa saja yang mengikuti anda di Twitter, dengan harapan kata – kata anda bermanfaat dan menarik bagi seseorang di audiens anda. Deskripsi lain tentang Twitter dan *tweeting* mungkin *microblogging* [6].

Twitter didirikan dan diresmikan pada tahun 2006 tepatnya pada tanggal 21 Maret 2006. Didirikan oleh Jack Dorsey, Twitter sudah cukup di kenal di dunia, bahkan pada tahun 2014 Twitter sudah menjadi salah satu dari 5 besar situs yang paling sering dikunjungi oleh banyak orang [8].

Twitter sendiri memiliki beberapa fitur, seperti *tweet* atau kicauan yang merupakan fitur utama di Twitter untuk mengirim dan melihat kicauan setiap pengguna Twitter. *Following* merupakan fitur untuk mengikuti pengguna lain pada Twitter. *Followers*, fitur yang digunakan untuk melihat siapa yang mengikuti anda pada media sosial Twitter. *Biography* merupakan fitur yang digunakan untuk mengetahui pesan akun Twitter anda pada *profile*. *Profile* merupakan salah satu fitur utama pada Twitter, fitur ini digunakan untuk melihat avatar Twitter, *biography* Twitter, dan lainnya. Pada Twitter juga terdapat fitur top trending yang digunakan untuk mempermudah penggunaannya untuk melihat *tweet* yang paling populer dan paling sering di *tweet* oleh para pengguna Twitter lainnya [8].

C. Information Gain

Information Gain adalah sebuah metode yang digunakan untuk menghitung nilai *entropy* yang ada dengan cara membagi dataset berdasarkan nilai dari random variable [9]. *Information Gain* juga memiliki

keuntungan, jika informasi yang dimiliki semakin banyak maka akan didapatkan nilai *entropy* yang semakin rendah dan meminimalisir kejutan.

Pada teori informasi dapat di gambarkan bahwa “*surprise*” yang semakin tinggi dalam suatu kejadian. Probabilitas lebih rendah untuk terjadinya kejadian “*surprise*” dengan cara memperbanyak jumlah informasi atau memiliki informasi yang lebih banyak, sedangkan distribusi probabilitas terjadi dimana peristiwa dan kemungkinan lebih “*surprising*” dan “*entropy*” yang lebih besar [10].

Pemilihan fitur dengan *Information Gain* dilakukan dalam tiga tahapan, pertama menghitung nilai *Information Gain* untuk setiap atribut dalam dataset original. Langkah selanjutnya adalah menentukan batas atau *threshold* yang diinginkan. Hal ini dilakukan agar memungkinkan atribut yang berbobot sama dengan batas atau lebih besar akan dipertahankan serta membuang atribut yang berada di bawah batas, lalu memperbaiki dataset dengan mengurangi atribut yang ada [11].

D. Support Vector Machine

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian konsep – konsep unggulan dalam bidang pattern recognition. Sebagai salah satu metode yang di bilang muda. SVM sudah memiliki kemampuan sebagai state of the art dalam pattern recognition. SVM sendiri merupakan metode learning machine yang berkerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua *class* pada *input space* [12].

III. METODOLOGI PENELITIAN

Objek dari penelitian ini adalah *cyberbullying* yang terjadi di Indonesia, pada media sosial Twitter. Data yang diambil adalah *tweet* dari tanggal 1 April 2020 hingga tanggal 30 April 2020 dan 1 Mei 2020 hingga 31 Mei 2020. *Cyberbullying* adalah sebuah perundungan yang dilakukan pada media digital, perundungan atau bullying adalah perilaku atau tindakan ancaman, kekerasan atau paksaan untuk mengintimidasi orang lain dan bisa berulang kali. *Cyberbullying* biasanya terjadi di media sosial, *messaging apps*, *online chatting*, forum *online*, email, dan komunitas permainan *online*. Penelitian ini membahas *cyberbullying* yang dilakukan pada media sosial, yaitu Twitter.

Twitter merupakan layanan untuk teman, keluarga, dan teman sekerja untuk berkomunikasi dan tetap terhubung melalui pertukaran pesan yang cepat dan sering. Pengguna memposting *tweet*, yang dapat berisi foto, video, tautan, dan teks. *Tweet* ini diposting ke profil, lalu terkirim ke pengikut, dan dapat di cari di pencarian Twitter.

A. Teknik Pengolahan Data

Support Vector Machine (SVM) memiliki konsep yaitu mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data. SVM memaksimalkan margin, yang merupakan jarak pemisah antara kelas data. SVM juga mampu berkerja pada dataset yang berdimensi tinggi dengan menggunakan kernel trick. Ada beberapa macam fungsi kernel SVM, yaitu *Linear*, *Polynomial*, *Gaussian RBF*, *Sigmoid*, *Invers Multi Kuadratik*, dan *Additive*.

Pada penelitian ini fungsi kernel yang digunakan adalah SVM Polynomial. SVM linear digunakan ketika ada data yang akan diklasifikasi terpisah dengan sebuah *hyperplane*, sedangkan SVM non-linear digunakan ketika data hanya dapat dipisahkan dengan garis lengkung. SVM Polynomial memiliki definisi fungsi dengan persamaan $K(\vec{x}_1, \vec{x}_j) = (\vec{x}_1 \cdot \vec{x}_j + 1)^d$ yang dimana $K(\vec{x}_1, \vec{x}_j)$ merupakan fungsi kernel, x merupakan fitur dan d merupakan ordo.

Hyperplane dalam SVM yang optimal didapatkan dengan cara merumuskan ke dalam *quadratic programming problem (QP problem)* dan dapat di selesaikan menggunakan library yang banyak tersedia dalam *numeric analyst*. Tetapi terdapat sebuah alternative yaitu menggunakan metode *sequential*. Metode ini dikembangkan oleh Vijayakumar untuk mencari nilai α , yang dapat di uraikan, sebagai berikut:

1. Inisialisasi $\alpha_i = 0$

Menghitung nilai matriks hessian dengan menggunakan persamaan:

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (1)$$

Dimana y merupakan kelas dari data ke - i dan ke - j , $K(x_i, x_j)$ merupakan fungsi kernel polynomial yang digunakan.

2. Menghitung setiap level dengan tahapan menggunakan persamaan:

$$E_i = \sum_{j=1}^n \alpha_j D_{ij}$$

$$\delta \alpha_i = \min\{\max\{y(1 - E_i), C - \alpha_i\}\}$$

$$\alpha_i = \alpha_i + \delta \alpha_i \quad (2)$$

3. Melakukan pengulangan ke tahap 2 sampai nilai α mencapai konvergen.

Information Gain (IG) merupakan salah satu metode untuk melakukan seleksi fitur, yang biasa digunakan oleh para peneliti untuk menentukan batas dari kepentingan sebuah atribut. Nilai IG diperoleh dari nilai *entropy* sebelum pemisahan dikurangi dengan nilai *entropy* setelah pemisahan. Nilai ini digunakan untuk penentuan atribut mana yang akan dibuang dan digunakan. Atribut yang memenuhi kriteria pembobotan nantinya akan digunakan untuk proses klasifikasi.

Dalam pemilihan fitur dengan IG dilakukan dalam 3 tahapan, yaitu:

1. Menghitung nilai *Information Gain* untuk setiap atribut.
2. Menentukan *threshold* atau batasan. Hal ini untuk menentukan atribut yang bobotnya lebih kecil dari *threshold* akan dibuang.
3. Memperbaiki *dataset* dengan pengurangan atribut.

Seleksi fitur *Information Gain* dirumuskan menjadi:

$$IG(t) = - \sum_{i=1}^{|\mathcal{C}|} P(C_i) \log P(C_i) +$$

$$P(t) \sum_{i=1}^{|\mathcal{C}|} P(C_i|t) \log P(C_i|t) +$$

$$P(\bar{t}) \sum_{i=1}^{|\mathcal{C}|} P(C_i|\bar{t}) \log P(C_i|\bar{t}) \quad (3)$$

Dimana C_i merupakan kelas data, $P(C_i)$ merupakan peluang dari kelas data, $P(t)$ dan $P(\bar{t})$ merupakan peluang term t yang muncul atau tidak muncul dalam dokumen.

IV. ANALISA DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data menggunakan *tools*, yaitu *software "R"*, data yang di ambil melalui sosial media Twitter dengan bentuk teks.

Pertama mencari kata dasar yang akan digunakan, kata dasar tersebut merupakan kata – kata yang di berikan pada saat melakukan wawancara, sehingga mempercepat pencarian data. Setiap kata di tarik sejumlah 30 *tweet*, dengan bertanggal dari 1 April 2020 hingga 28 Mei 2020.

Berdasarkan wawancara dengan ahli Bahasa, jumlah kata yang akan digunakan adalah 49 kata, yang nantinya dilakukan penarikan data pada setiap kata menggunakan *software R*.

Tabel 1. Kumpulan kata

Anjing lu	Anak Haram	Jobong / Pelacur	Mati lu
Bajingan	Anak Tapir	Kentot	Memek
Banci	Asu	Lemah lu	Ngentot
Bicis	Babi lu	Lud	Pecun
Bitch	Bocah tolol	Miskin lu	Pesugihan
Bitches	Cina	Monyet lu	Si dongo
Brengsek	Fucek	Nibba	Simpanan om om
Fak u	Fuck You	Nigga	Sugarbaby
Fakboi	Jembut	Pelacur	Sugardaddy

Gendut	Kafir	Pelakor	Tai lo
Goblok	Kontol	Pelakor	Titit
Idiot	Lonte	Slut	Tolol
Jelek			

B. Text Preprocessing

Sekarang merupakan tahapan untuk melakukan proses perubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan.

Pertama dilakukan tahapan, *case folding*. *Case folding* merupakan tahapan merubah keseluruhan teks menjadi bentuk standar yaitu huruf kecil atau *lowercase*.

Setelah itu, ada tahap perubahan dari “kalimat” menjadi “kalimat_text”

Setelah itu adanya pembuatan *corpus*, *corpus* digunakan untuk menstrukturisasi teks yang terdapat di “kalimat_text”, setelah itu dilakukan tahap perubahan nama menjadi “kalimat_text_corpus” untuk membedakan setelah di lakukan strukturisasi pada dokumen yang ada.

Tahap terakhir melakukan perubahan data pada dokumen menjadi huruf kecil semua menggunakan *function*,

```
“content_transformer(tolower)”.
```

Setelah *case folding* atau pengubah teks menjadi huruf kecil, selanjutnya dilakukan *tokenizing* yaitu penghilangan tanda baca dan tanda hubung.

Pada tahap *tokenizing* memiliki gambaran *function* seperti:

```
kalimat_text_corpus <- tm_map
(kalimat_text_corpus, removePunctuation)
```

Penulisan di atas merupakan *function* yang digunakan untuk *tokenizing*. Jika di jabarkan, “kalimat_text_corpus” merupakan *corpus* yang telah di *case folding* untuk digunakan ke tahap *tokenizing*. Setelah tahap *tokenizing* sekarang tahap *filtering*, pada tahap ini dilakukan pengambilan kata – kata penting dari hasil *tokenizing*. Biasanya menggunakan *stoplist* atau *wordlist*.

Pada tahap ini digunakan *stoplist* untuk dijadikan patokan, agar mudah menemukan kata – kata yang akan digunakan untuk membantu mempermudah pencarian data yang dibutuhkan, dan *function* yang digunakan adalah “*removewords*”, lalu mencari “*stopwords*”, contoh kata – kata yang digunakan pada *stopwords* adalah “*retweet*”, “*di*”, dan “*yang*”. Pada tahap *filtering* memiliki bentuk penulisan *function* seperti:

```
kalimat_text_corpus <- tm_map
(kalimat_text_corpus, removewords,
```

```
stopwords(“RT”, “retweet”, “di”, “yang”))
```

Setelah melakukan *filtering* lalu dilakukan *stemming*, pada tahap *stemming* dilakukan untuk mengetahui akar sebuah kata sehingga memperkecil jumlah indeks yang ada. Sedangkan pada tahap *stemming* memiliki penulisan *function* seperti berikut :

```
kalimat_text_corpus <- tm_map
(kalimat_text_corpus, stemDocument)
```

Pada tahap *stemming* menggunakan *function* yang berfungsi untuk melakukan *stemming* pada dokumen, yaitu “*stemDocument*”, *function* ini memiliki fungsi untuk melakukan pembenaran rekonstruksi kata – kata yang akan digunakan.

Setelah melalui tahap *stemming* maka proses lanjutannya ada *term weighting*, *term weighting* adalah pembobotan setiap hasil *token stemming* untuk digunakan. Setelah melakukan tahap *term frequency* dan *inverse document frequency* maka dilakukan penggabungan untuk melakukan penghitungan TF-IDF yang ada di dalam data.

C. Support Vector Machine

Setelah mendapatkan nilai *term weighting* dilakukan tahap ke *Support Vector Machine*. Kegiatan prediksi menggunakan 5 parameter, yaitu variabel *lambda*, konstanta *gamma*, *epsilon*, maksimum iterasi, dan nilai *complexity* (C). Setiap parameter yang digunakan dalam pengujian memiliki nilainya masing – masing misalnya λ bernilai 0.5, $Y = 0.001$, $\epsilon = 0.0001$, $C = 1$, dan $IterMax = 100$.

Dari ke-5 parameter yang digunakan yang pertama diuji adalah variabel *lamda*, pada variabel *lambda* mendapatkan hasil *accuracy* sebesar 74.20%, *precision* sebesar 73.10%, lalu *recall* sebesar 77.10% dan yang terakhir *f-measure* sebesar 75.30%. Hal ini dapat terjadi dikarenakan nilai variabel *lamda* hanya berguna untuk melakukan perhitungan matriks hessian, dan matriks hessian digunakan untuk menghitung nilai E_i sehingga menyebabkan hasil tidak mempengaruhi nilai bias.

Dilakukan pengujian pada konstanta *gamma*. Hasil terbaik didapat pada nilai *gamma* 0.001 dan 0.01, setelah itu nilai semakin menurun ketika nilai *gamma* semakin tinggi. Penyebab nilai semakin menurun ketika nilai *gamma* semakin tinggi karena yang digunakan untuk menghitung adalah nilai *delta alpha*, yang dimana nilai *delta alpha* merupakan nilai yang menentukan apakah hasil sudah konvergen atau belum. Nilai *gamma* 0.001 memiliki *accuracy* 76.90%, kemudian *precision* 75.1%, *recall* sebesar 80.20%, dan *f-measure* 78.80%.

Hasil pengujian nilai *epsilon* kurang dari 0.00001 yang dapat diartikan bahwa 0.0000001 juga memiliki nilai terbaik. Hal ini disebabkan karena nilai *epsilon* digunakan sebagai batas maksimal untuk hasil

konvergen. Jika nilai epsilon semakin tinggi, maka akan menghasilkan konvergen yang semakin cepat, sehingga pemilihan nilai epsilon yang digunakan semakin cepat yaitu 0.000001 dengan *accuracy* 78%, tingkat *precision* 72%, kemudian *recall* 91%, dan *f-measure* 81%.

Pada awal iterasi memiliki nilai yang tinggi, tetapi semakin menurun ketika nilai iterasi semakin besar, dan ketika pada iterasi yang lebih dari 1000 hasilnya sama, penyebabnya dikarenakan perhitungan telah memasuki keadaan konvergen. Hal ini juga menandakan bahwa hasil terbaik yang didapatkan berada pada iterasi sebanyak 20 dengan nilai *accuracy* 80%, dengan tingkat *precision* 73%, kemudian *recall* 92.20% dan memiliki *f-measure* 85%.

Pada hasil pengujian nilai C, nilai terbaik pada saat pengujian adalah 1, karena semakin tinggi nilai C maka nilai kernel polynomial-nya akan semakin tinggi juga. Ketika mencapai tahap ini nilai kernel polynomial semakin tinggi maka nilai matriks hessian juga semakin tinggi, yang dapat menyebabkan perhitungan nilai konstanta gamma yang hasilnya semakin kecil dapat menyebabkan iterasi semakin cepat mencapai konvergen. Hasil dari pengujian memiliki nilai *accuracy* 87%, dengan *precision* 83%, nilai *recall* 96%, dan *f-measure* 91%.

D. Information Gain

Setelah selesai melakukan prediksi akurasi pada metode *Support Vector Machine* selanjutnya melakukan pengujian *Information Gain* untuk mendapatkan hasil dari *threshold feature selection*. Nilai *threshold* yang didapatkan sebesar 90%. Ketika semua fitur digunakan hasil yang didapatkan *accuracy* sebesar 86%, *precision* 81%, *recall* 95% dan *f-measure* 87%. Jika pemilihan *term* menggunakan seleksi fitur *Information Gain* maka dapat menghasilkan nilai yang lebih tinggi, dan nilai dari *Information Gain* akan mempengaruhi *term* yang akan digunakan pada proses identifikasi.

V. SIMPULAN

Kesimpulan yang dapat diambil pada hasil penelitian, pengujian, dan analisis yang telah dilakukan, dapat dikatakan bahwa hasil identifikasi *tweet cyberbullying* menggunakan metode *Support Vector Machine* mendapatkan akurasi yang cukup tinggi yaitu dengan hasil *accuracy* 80%, *precision* 75.1%, *recall* 96%, dan *f-measure* 85% dan juga pada hasil pengujian *threshold* seleksi fitur pada *Information Gain* mendapatkan nilai yang cukup tinggi sebesar *accuracy* 86%, *precision* 81%, *recall* 95%, dan *f-measure* 87% sehingga hasil identifikasi *tweet cyberbullying* dengan kedua metode memiliki hasil yang cukup maksimal.

DAFTAR PUSTAKA

- [1] Kompas.com. (2019, Oktober 16). Apa itu *Cyberbullying* dan Bagaimana Mengajari Anak Menghindarinya? Retrieved from Kompas.com: <https://lifestyle.kompas.com/read/2019/10/16/112740720/apa-itu-cyberbullying-dan-bagaimana-mengajari-anak-menghindarinya?page=all>
- [2] Stopbullying.gov. (2020, May 7). What Is *Cyberbullying*. Retrieved from stopbullying.gov: <https://www.stopbullying.gov/cyberbullying/what-is-it>
- [3] Zahro Malihah, A. (2018). Perilaku *cyberbullying* pada remaja dan kaitannya dengan kontrol diri dan komunikasi orang tua. Departemen ilmu keluarga dan konsumen, Falkutas Ekologi Manusia, Institut pertanian bogor, 145 - 156.
- [4] It Security Newsportal. (2016, July 13). *Cyberbullying* Kompleksitas Kehidupan Remaja. Retrieved from It Security Newsportal: <https://www.bacapikirshare.org/cyberbullying-kompleksitas-kehidupan-remaja/>
- [5] Rahayu, F. S. (2012). *Cyberbullying* sebagai Dampak Negatif Penggunaan Teknologi Informasi. Prodi Teknik Informatika, Falkutas Teknologi Industri, Universitas Atma Jaya Yogyakarta, 22-31.
- [6] Gil, P. (2020, April 29). What Is Twitter & How Does It Work? Retrieved from Lifewire: <https://www.lifewire.com/what-exactly-is-Twitter-2483331>
- [7] Wisnubrata. (2019, October 16). Kompas.com. Retrieved from Kompas.com: <https://lifestyle.kompas.com/read/2019/10/16/112740720/apa-itu-cyberbullying-dan-bagaimana-mengajari-anak-menghindarinya?page=all>
- [8] Basri, H. (2017). Peran Media Sosial Twitter dalam Interaksi Sosial Pelajar Sekolah Menengah Pertama di Kota Pekanbaru. Jurusan Sosiologi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Riau, 1-15.
- [9] Brownlee, j. (2019, October 16). *Information Gain* and Mutual Information for Machine Learning. Retrieved from machinelearningmastery.com: <https://machinelearningmastery.com/information-gain-and-mutual-information/>
- [10] Astuti, F. D. (2017). Seleksi Atribut Menggunakan *Information Gain* untuk Clustering Penduduk Miskin Dengan Validity Index Xie Beni. STMIK AKAKOM, 61-65.
- [11] Muchammad Rifqi Maulana, M. A. (2015). *Information Gain* untuk mengetahui pengaruh atribut terhadap klasifikasi persetujuan kredit. Jurnal Litbang Kota Pekalongan Vol.9, 113-123.
- [12] Anto Satriyo Nugroho, A. B. (2003). *Support Vector Machine* Teori dan Aplikasinya dalam Bioinformatika. Retrieved from InfoKomputer.com: <http://asnugroho.net/papers/ikcsvm.pdf>.