

# Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Publik Peserta Pilpres 2019

Rafiqah Cahyani, Indri Sudanawati Rozas, Nita Yalina

**Abstract**—In Indonesia, Twitter has become a microblogging site that is very popular with the public. People are used to expressing their opinions or opinions on social media including on Twitter, this is used in response to an event or events that occur in their environment. In this study various processes were carried out, namely crawling data from Twitter, pre-processing text data, Naïve Bayes Classifier for text probability to sentiment, and Simple Wighted Product Model for ranking sentiments. From the results of these experiments can be obtained the best accuracy results using the third proportion with random data collection by getting an average accuracy value of 99.75%. As for ranking sentiments for each public figure, it can be seen that the first rank is KH Ma'ruf Amin, the second is Sandiaga Salahudin Uno, the third is Joko Widodo, and the last is Prabowo Subianto.

**Index Terms**— Twitter, Public Figure, Naïve Bayes Classifier, Simple Wighted Product Model.

**Abstract**—Di Indonesia twitter menjadi salah satu situs microblogging yang sangat digemari oleh masyarakat. Masyarakat sudah terbiasa dalam menyampaikan pendapat atau opininya dalam media social termasuk dalam twitter, hal ini digunakan dalam menanggapi suatu kejadian atau peristiwa yang terjadi di lingkungan mereka. Pada penelitian ini dilakukan berbagai proses yaitu crawling data dari twitter, pre prosesing data teks, Naïve Bayes Classifier untuk probabilitas teks terhadap sentimen, dan Simple Wighted Product Model untuk perankingan sentimen. Dari hasil percobaan tersebut dapat diperoleh hasil akurasi terbaik yaitu menggunakan proporsi ketiga dengan pengambilan data secara acak dengan mendapatkan nilai rata rata akurasi sebesar 99,75%. Sedangkan untuk perankingan sentiment

Rafiqah Cahyani, Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya, Indonesia (email: [rafikacahyani1712@gmail.com](mailto:rafikacahyani1712@gmail.com))

Indri Sudanawati Rozas, Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya, Indonesia (email: [indrisrozass@uinsby.ac.id](mailto:indrisrozass@uinsby.ac.id))

Nita Yalina, Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya, Indonesia (email: [nitayalina@uinsby.ac.id](mailto:nitayalina@uinsby.ac.id))

terhadap masing masing tokoh public dapat di ketahui yang memperoleh peringkat pertama yaitu KH Ma'ruf Amin, yang kedua yaitu Sandiaga Salahudin Uno, yang ketiga yaitu Joko Widodo, dan yang terakhir yaitu Prabowo Subianto

**Kata Kunci**— Twitter, Tokoh Publik, Naïve Bayes Classifier, Simple Wighted Product Model

## I. INTRODUCTION

Tahun 2019 merupakan tahun politik, dimana akan dilaksanakannya pemilihan presiden dan wakil presiden Indonesia. Dalam pelaksanaan pemilihan presiden dan wakil presiden terdapat berbagai opini masyarakat dengan sentiment positif dan negatif pada para tokoh publik yang mencalonkan dirinya untuk menjadi Presiden dan Wakil Presiden Republik Indonesia. Adapun tokoh public tersebut adalah Bapak Joko Widodo, KH Ma'ruf Amin, Prabowo Subianto, dan Sandiaga Salahudin Uno. Pada penelitian kali ini peneliti akan meneliti sejauh mana citra positif yang dimiliki oleh keempat tokoh public tersebut berdasarkan sentiment dari masyarakat. Kelas sentiment enter bagimen jadi duap olaritas yaitu sentiment negatif dan sentiment positif.

Penelitian yang akan dilakukan adalah pengklasifikasi auto pik sentimen dan sentiment dari informasi teks twitter yang membahas mengenai tokoh publik menggunakan metode Naïve Bayes Classifier dan Simple Weighted Product Model. Metode Naïve Bayes Classifier juga digunakan dalam menentukan sentiment dari tweet sehingga dapat diketahui sentiment positif, negatif yang menjadi kelas sentimen pada teks twitter tersebut. Sedangkan Simple Weighted Product Model digunakan dalam menentukan ranking atau urutan dari para tokoh publik yang bersangkutan. [1]

Untuk pengambilan data twitter menggunakan API (Application Programming Interface) dari twitter. Pemberian label kelas merupakan masalah yang cukup sulit, pada data tanpa label dengan jumlah yang sangat besar. Oleh karena itu digunakan metode pelabelan otomatis yaitu dengan memberikan label secara otomatis berdasarkan penanda tertentu yaitu dengan

menggunakan sentiword positif dan sentiword negative Bahasa Indonesia yang didapatkan dari github. Metode klasifikasi yang digunakan adalah metode klasifikasi *Naive Bayes Classifier*. Metode *Naive Bayes Classifier* merupakan salah satu metode klasifikasi yang memiliki performa yang cukup baik untuk mengklasifikasikan data teks [2]. Dan digunakan *Simple Wighted Product Model* sebagai suatu metode untuk memperoleh nilai ranking terbaik dari keempat tokoh tersebut.

## II. LANDASAN TEORI

### 2.1 Machine Learning

Algoritma Machine Learning adalah suatu algoritma yang mempunyai kemampuan untuk belajar dari data dan pengalaman dengan memperhatikan beberapa hal, seperti tugas dan ukuran kinerja, jika suatu tugas dapat diukur untuk meningkatkan pengalaman, sehingga mesin tersebut dapat dikatakan belajar. [3]

### 2.2 Analisis Sentimen

Analisis sentiment mengacu pada bidang yang sangat luas mulai dari penerapan Bahasa alami, komputasi linguistic dan teks mining yang bertujuan untuk menganalisa sikap, evaluasi, sentiment, pendapat, penilaian serta emosi seseorang yang berkaitan dengan suatu produk, layanan, topik, individu, maupun kegiatan tertentu. [4]

### 2.3 Teks Mining

Teks Mining berhubungan dengan penciptaan informasi baru secara otomatis atau semi otomatis yang sebelumnya tidak diketahui dari data berkualitas tinggi dari sejumlah besar text yang tidak terstruktur. [5]

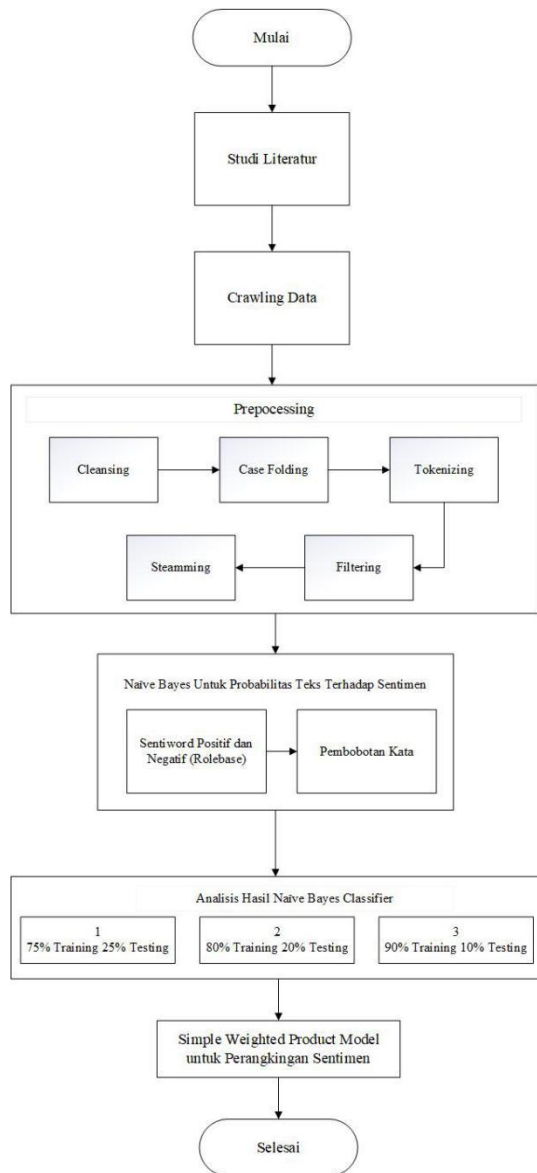
### 2.4 Naïve Bayes Classifier

Naïve Bayes Classifier adalah konsep dari probabilitas yang dipakai sebagai cara untuk menentukan kelompok kelas dokumen teks dan juga dapat mengolah data dalam jumlah besar serta menghasilkan akurasi yang cukup tinggi [6]

## III. METODE PENELITIAN

Data pada penelitian ini berupa opini berbahasa Indonesia yang di dapatkan dari akun twitter para tokoh publik yang bersangkutan. Dari data yang tersedia dan setelah melalui proses penghapusan data duplikat dan data netral didapatkan data dari Joko Widodo sebanyak 5.755, KH Ma'ruf Amin sebanyak 6.177, Prabowo Subianto sebanyak 7130, dan Sandiaga Salahudin Uno sebanyak 5521. Data ini adalah data yang digunakan sebagai data klasifikasi.

Adapun langkah langkah untuk perancangan analisis sentiment yang dibahas adalah sebagai berikut:



Gambar 1 Metodologi Penelitian

### A. Studi Literatur

Langkah yang pertama yaitu studi literatur yang digunakan untuk mencari data dan informasi yang dibutuhkan dari beberapa sumber.

### B. Crawling Data

Langkah pertama yaitu dengan melakukan crawling data twitter. Pada crawling data twitter ini harus mempunyai akun twitter lebih dahulu langkah selanjutnya yaitu registrasi di <https://dev.twitter.com/apps/new> untuk mendapatkan kode akses Api twitter. Setelah mendapatkan kode akses Api twitter, selanjutnya digunakan software Rstudio dan V Tag 6.0 untuk melakukan crawling data twitter.

### C. Pre-Prossecing Data

Langkah yang ketiga yaitu pre-proses data yang berguna untuk merubah dataset yang belum diolah menjadi dataset yang siap untuk diklasifikasikan.[7] Adapun beberapa tahapannya adalah proses *cleansing*,

case folding, tokenizing, filtering dan stemming.

D. Naïve Bayes Untuk Probabilitas Teks Terhadap Sentimen

Probabilitas sentiment merupakan proses penentuan teks twitter yang dihitung untuk didapatkan nilai tertinggi dari beberapa kategori dalam sentiment [8]. Seperti halnya untuk menghitung tweet terhadap topik, perhitungan probabilitas tweet terhadap sentimen juga menggunakan rumus dimana untuk menentukan probabilitas sentiment memerlukan nilai p(ci) dan p(wkj.ci) untuk probabilitas awal sentiment didapatkan dari jumlah tweet yang tergabung dalam topik ci dibagi dengan semua topik [9].

$$p(wkj.ci) = \frac{pk+1}{n+term} \dots\dots\dots(1)$$

$$S_i = \prod_{j=1}^n X_{ij} w_j \dots\dots\dots(2)$$

Dan untuk menghitung probabilitas teks terhadap sentiment juga dibuatkan wordcloud positif dan negative dari masing masing tokoh publik

E. Analisis Hasil Naïve Bayes Classifier

Pada system klasifikasi menggunakan Naïve Bayes Classifier, tahapan awal yang dilakukan adalah dengan membagi data menjadi dua bagian yaitu data training dan data testing [10]. Pembagian data ini akan dilakukan secara acak dan urut, hal ini dilakukan karena untuk mengetahui tingkat akurasi yang dihasilkan dari kedua cara tersebut mana yang menghasilkan akurasi yang lebih tinggi dari keempat calon publik yang bersangkutan. Penelitian ini membagi data menjadi 3 kali proporsi.

- 1. 75% data training dan 25% data testing
- 2. 80% data training dan 20% data testing
- 3. 90% data training dan 10% data testing

Setelah proses system klasifikasi berakhir dari data training yang tersedia. Proses selanjutnya yang harus dilakukan adalah pengujian terhadap system klasifikasi yang dibuat. Evaluasi merupakan tahapan yang paling menentukan, karena bisa dilihat dari proses evaluasi tersebut apakah system klasifikasi yang sudah dibuat mempunyai performa yang baik atau sebaliknya. Confusion matrix adalah salah satu cara untuk melakukan pengujian pada system klasifikasi. Confusion Matrix juga bisadi pakai untuk menghitung tingkat akurasi, presisi, dan recall.

F. Simple Weighted Product Model Untuk Perangkingan Sentimen

Pada penelitian ini penggunaan metode Simple Weighted Product Model digunakan untuk menghitung nilai sentimen yang ada pada topik setiap tokoh public sehingga didapat kanurutan-urutan nama tokoh publik

yang mendapatkan nilai sentiment terbaik sampai dengan sentiment terburuk. Langkah awal yaitu dengan pemilihan ternatif, Kemudian memasukkan niali bobot kepentingan dari setiap kriteria untuk mendapatkan hasil rangking.

$$V_i = \frac{\prod_{j=1}^n x_{ij} w_j}{\prod_{j=1}^n (x_{ij}^*) w_j} \dots\dots\dots(3)$$

Nilai akhir dari perhitungan akan digunakan untuk perangkingan. Semakin besar nilai akhir suatu alternatif (tokoh publik) maka akan mendapatkan citra positif terbaik dari masing masing tokoh publik.

IV. HASIL DAN PEMBAHASAN

Dari penggunaan data dari Joko Widodo sebanyak 14.526, KH Ma'ruf Amin sebanyak 13.207, Prabowo Subianto sebanyak 17.003, dan Sandiaga Salahudin Uno sebanyak 12.533. Data ini adalah data yang digunakan sebagai data klasifikasi. Dan setelah melalui proses pre-processing data maka langkah selanjutnya adalah pembobotan kata menggunakan sentiword positif dan

negatif yang didapatkan dari Github untuk masing masing tokoh publik. Perhitungan term atau kata menggunakan kata hasil preprocessing dan menghasilkan bobot masing-masing kata yang dibawah ini:

1. Pembobotan Kata

Pembobotan kata untuk Para tokoh publik peserta pilpres 2019, pembobotan kata ditentukan dari banyaknya kemunculan kata yang ada pada tweet yang bersangkutan. Hal ini juga menjadi bahan untuk pembuatan wordcloud sentimen positif dan negatif dari tokoh publik yang bersangkutan. Wordcloud ini dibuat untuk mengetahui kata sentimen positif dan negatif mana yang paling sering digunakan masyarakat untuk menilai tokoh publik yang mereka perbincangkan.

A. Pembobotan Kata Joko Widodo



Gambar 2. Would Cloud Sentimen Positif untuk Joko Widodo







B. Presentasi Nilai Akurasi Pada Proporsi Kedua

Tabel 2. Presentasi Nilai Akurasi Pada Proporsi Kedua

Proporsi Pertama 80% Testing 20% Training

| No              | Nama             | Data     |         | Hasil Akurasi |        |
|-----------------|------------------|----------|---------|---------------|--------|
|                 |                  | Training | Testing | Urut          | Acak   |
| 1               | Joko Widodo      | 4604     | 1151    | 98,69%        | 97,39% |
| 2               | KH. Ma'ruf Amin  | 4941     | 1235    | 99,67%        | 99,43  |
| 3               | Prabowo Subianto | 5704     | 1426    | 99,29%        | 99,64  |
| 4               | Sandiaga S Uno   | 4416     | 1104    | 99,36%        | 99,54  |
| Total Rata-Rata |                  |          |         | 99,25%        | 99,00% |

Dalam tabel diatas dapat diketahui bahwa hasil rata rata akurasi pada proporsi kedua oleh keempat tokoh tersebut yaitu pada pengambilan data urut diperoleh rata rata akurasi sebesar 99,25% sedangkan dari pengambilan data secara acak pada proporsi pertama diperoleh rata rata akurasi sebesar 99,00%.

C. Presentasi Nilai Akurasi Pada Proporsi Ketiga

Tabel 3. Presentasi Nilai Akurasi Pada Proporsi Ketiga

Proporsi Pertama 90% Testing 10% Training

| No              | Nama             | Data     |         | Hasil Akurasi |        |
|-----------------|------------------|----------|---------|---------------|--------|
|                 |                  | Training | Testing | Urut          | Acak   |
| 1               | Joko Widodo      | 5179     | 575     | 98,43%        | 100%   |
| 2               | KH. Ma'ruf Amin  | 5559     | 617     | 99,83%        | 99,51% |
| 3               | Prabowo Subianto | 6417     | 713     | 99,43%        | 99,85% |
| 4               | Sandiaga S Uno   | 4968     | 552     | 99,45%        | 99,63% |
| Total Rata-Rata |                  |          |         | 99,29%        | 99,75% |

Dalam tabel diatas dapat diketahui bahwa hasil rata rata akurasi pada proporsi ketiga oleh keempat tokoh tersebut yaitu pada pengambilan data urut diperoleh rata rata akurasi sebesar 99,29% sedangkan dari pengambilan data secara acak pada proporsi pertama diperoleh rata rata akurasi sebesar 99,75%.

Dari ketiga percobaan yang dilakukan maka dapat diambil kesimpulan jika metode Naïve Bayes Classifier ini cocok diterapkan pada proporsi ketiga dengan pengambilan data secara acak. Hasil rata rata yang diperoleh adalah 99,75%, hasil rata rata ini merupakan hasil yang paling tinggi diperoleh dibandingkan dengan percobaan yang lain.

3. Perhitungan Menggunakan Metode Simple Weighted Product Model

Pada proses ini akan menghasilkan nilai ranking tiap-tiap tokoh public dengan nilai sentimen yang didapatkan dengan menggunakan metode Metode Simple Weighted Product Model. Informasi sentiment tokoh politik diperoleh dari hasil klasifikasi tweet tokoh

publik. Penggunaan metode Metode Simple Weighted Product Model dalam penentuan peringkat tokoh publik. Langkah pertama yaitu menentukan nilai alternative dari setiap tokoh tersebut.

Tabel 4. Tabel Alternatif Perhitungan SWP

| Alternatif       |
|------------------|
| Joko Widodo      |
| KH. Ma'ruf Amin  |
| Prabowo Subianto |
| Sandiaga S Uno   |

Yang kedua yaitu menentukan kriteria perhitungan SWP dari masing masing tokoh.

Tabel 5. Tabel Kriteria Perhitungan SWP

| Kriteria |
|----------|
| Positif  |
| Negatif  |

Yang ketiga yaitu menentukan nilai kriteria setiap alternative perhitungan SWP

Tabel 6. Nilai Kriteria Setiap Perhitungan SWP

| No | Alternatif       | Kriteria |         |
|----|------------------|----------|---------|
|    |                  | Positif  | Negatif |
| V1 | Joko Widodo      | 4144     | 1611    |
| V2 | KH. Ma'ruf Amin  | 5239     | 937     |
| V3 | Prabowo Subianto | 5014     | 2115    |
| V4 | Sandiaga S Uno   | 4243     | 1277    |
|    | Bobot            | 10       | 10      |

Selanjutnya yaitu melakukan proses perhitungan menggunakan metode Simple Weightes Product (SWP) akan dilakukan normalisasi bobot.

$$w_j = \frac{w_j}{\sum w_j} \dots\dots\dots (4)$$

$$W1 = 10 / (10+10) = 10 / 20 = 0,5$$

$$W2 = 10 / (10+10) = 10 / 20 = 0,5$$

Selanjutnya yaitu menghitung vector S

$$S_i = \prod_{j=1}^n X_{ij} w_j \dots\dots\dots (5)$$

$$S1 = (4144 0,5) (1611 -0,5) = 1,603844$$

$$S2 = (5239 0,5) (937 -0,5) = 2,364582$$

$$S3 = (5014 0,5) (2115 -0,5) = 2,539703$$

$$S4 = (4243 0,5) (1277 -0,5) = 1,822809$$

Tahap selanjutnya yaitu menghitung vector perankingan

$$V_i = \frac{\prod_{j=1}^n x_{ij} w_j}{\prod_{j=1}^n (x_{i^*j}) w_j} \dots\dots\dots (6)$$

$$V1 = 1,603844 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,218777$$

$$V2 = 2,364582 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,322548$$

$$V3 = 2,539703 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,210028$$

$$V4 = 1,822809 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,248648$$

Dari proses perhitungan nilai sentiment dengan bobot topik diatas, didapatkan nilai akhir sentimen pada masing – masing tokoh public dengan nilai tertinggi yakni 0,322548 (KH. Maruf Amin), 0,248648 (Sandiaga Salahudin Uno), 0,218777 (Joko Widodo) dan nilai terendah yakni 0,210028 (Prabowo Subianto). Maka kesimpulan didapatkan bahwa KH. Maruf Amin memiliki nilai citra positif tertinggi, disusul oleh Sandiaga Salahudin Uno kemudian Joko Widodo, dan yang terakhir yakni Prabowo Subianto.

## V. KESIMPULAN

Berdasarkan hasil pembuatan sistem klasifikasi dengan menggunakan Naïve Bayes Classifier dan menggunakan metode perhitungan Simple Weighted Product Model terhadap semua data yang diperoleh dari twitter, maka dapat disimpulkan bahwa, Hasil yang di dapatkan menunjukkan bahwa metode Naïve Bayes Classifier cocok diterapkan di ketiga proporsi yang ada, karena sama sama mendapatkan nilai akurasi yang cukup tinggi. Tetapi setelah dilakukan rata rata maka percobaan yang paling sesuai yaitu menggunakan proporsi ketiga dengan pengambilan data secara acak dengan memperoleh hasil rata rata sebesar 99,75%. Dengan menggunakan metode perhitungan Simple Weighted Product Model maka hasil yang didapatkan yaitu tokoh publik yang mendapatkan urutan (rangking) dalam sentiment masyarakat terkait citra positif yang dimiliki tokoh publik yang pertama adalah KH. Ma'ruf Amin dengan mendapatkan nilai 0,322548, setelah itu adalah Sandiaga Salahudin Uno yang mendapatkan nilai 0,248648, lalu Joko Widodo yang mendapat nilai 0,218777 dan nilai terendah yakni Prabowo Subianto dengan nilai 0,210028.

Berdasarkan hasil penelitian dari analisis sentiment teks pada media sosial twitter menggunakan Naïve Bayes Classifier dan dengan menggunakan metode perhitungan Simple Weighted Product Model yang telah dilakukan, diperlukan penelitian yang lebih lanjut terkait dengan analisis sentiment tokoh public ini. Berikut saran yang dapat disampaikan dari hasil kesimpulan diatas, Pada Proses klasifikasi menggunakan Naive Bayes sangat bergantung dari data latih yang digunakan. Penggunaan data latih sebaiknya menggunakan proporsi data yang

sama dari setiap kelasnya serta memastikan kelas dari data latih tersebut benar. Pada proses crawling data twitter, data yang didapatkan dari masing masing tokoh public banyak terdapat data outlier (data yang tidak sesuai). Oleh sebab itu crawling data sebaiknya dilakukan menggunakan cara lain atau dengan mengganti kata kunci yang lebih sesuai sehingga tidak didapatkan banyak data outlier Untuk mendapatkan hasil yang signifikan, perlu dilakukan eksperimen penggunaan leksikon polaritas dan multi kata, menggunakan strategi yang sangat mendasar berdasarkan pencarian lemmas polaritas dalam teks / tweet. Jika tweet tidak mengandung setidaknya satu lemma juga ditemukan dalam leksikon polaritas eksternal, maka tweet tersebut tidak memiliki polaritas apa pun.

Hasil akurasi yang didapatkan dalam penelitian kali ini belum tentu sama tingginya dengan hasil akurasi jika menggunakan metode yang berbeda. Hal ini bisa menjadi rekomendasi untuk penelitian selanjutnya untuk melakukan penelitian dengan Teknik yang sama tetapi menggunakan metode yang berbeda.

## VI. REFERENCES

- [1] Abdillah, S. (2011). *Penyakit Stroke Dengan Klasifikasi Data Mining Pada RS Santa Maria Pemalang..*
- [2] Juniawan, I. (2009). *Klasifikasi Dokumen Teks Berbahasa Indonesia Menggunakan Minor Component Analysis.*
- [3] Ariadi, D., & Fithriyari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *JURNAL SAINS DAN SENI ITS Vol. 4, No.2.*
- [4] Mitchell, T. M. (2009). Machine learning in ecosystem informatics and sustainability. *IJCAI International Joint Conference on Artificial Intelligence.* [https://doi.org/10.1007/978-3-540-75488-6\\_2](https://doi.org/10.1007/978-3-540-75488-6_2)
- [5] Chen, L. F., Su, C. T., & Chen, K. H. (2011). An improved particle swarm optimization for feature selection. *Intelligent Data Analysis, 16(2), 167–182.* <https://doi.org/10.3233/IDA-2012-0517>
- [6] Hearst, M. A., & Hall, S. (n.d.). *P99-1001.*
- [7] Prasetyo, E. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. In *Penerbit Andi.* <https://doi.org/10.1017/CBO9781107415324.004>
- [8] Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa

Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1718–1724.

- [9] Rini, D. C., Farida, Y., & Puspitasari, D. (2016). Klasifikasi Menggunakan Metode Hybrid Bayessian-Neural Network (Studi Kasus: Identifikasi Virus Komputer). *Jurnal Matematika "MANTIK,"* 1(2), 38. <https://doi.org/10.15642/mantik.2016.1.2.38-43>
- [10] Saraswati, N. W. S. (2013). Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis. *Seminar Nasional Sistem Informasi Indonesia*, 586–591. Retrieved from [http://is.its.ac.id/pubs/oajis/index.php/file/download\\_file/512](http://is.its.ac.id/pubs/oajis/index.php/file/download_file/512)