Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019
masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 3 No. 3 (2019) 383 - 388 ISSN Media Elektronik: 2580-0760

Integrasi N-gram, Information Gain, Particle Swarm Optimation di Naïve Bayes untuk Optimasi Sentimen Google Classroom

Fajar Pramono[1], Didi Rosiyadi[2], Windu Gata[3]
[1,3]Magister Ilmu Komputer, Fakultas Ilmu Komputer, STMIK Nusa Mandiri Kramat
[2]Fakultas Teknik Informasi, Universitas Bina Sarana Informatika
[1]14002112@nusamandiri.ac.id, [2]didi.rosiyadi@gmail.com,[3]windu@nusamandiri.ac.id

Abstract

The use of Learning Management System (LMS) applications made by Google with name Google Classroom since 2015 in junior and senior high schools in Bekasi City helps the learning process become easier. However, its use can have positive and negative effects on students. Google Class Sentiment by integrating N-grams, Information Gain, Particle Swarm Optimization, and Naïve Bayes Classifiers that have never been done by researchers before. From the experiments carried out, N-gram can increase the accuracy of 6.7% and AUC 4%, while using PSO can increase the Accuracy of 9.9% and AUC of 10.4%.

Keywords:N-gram, Information Gain, Particle Swarm Optimization, Naïve Bayes, Google Classroom

Abstrak

Penggunaan aplikasi Learning Management System (LMS) Google yaitu *Google Classroom* sejak tahun 2015 pada SMP dan SMA di Kota Bekasi membantu proses pembelajaran menjadi lebih mudah. Namun apakah kesan dari penggunaanya dapat memberikan efek positif dan negatif bagi siswanya. Di penelitian ini penulis mencoba mencari nilai optimasi akurasi dari sentimen Google Classroom dengan mengintegrasikan N-gram, *Information Gain*, *Particle Swarm Optimization* (PSO) dan *Klasifier Naïve Bayes* yang belum pernah dilakukan para peneliti sebelumnya. Dari percobaan yang dilakukan penggunaan N-gram dapat meningkatkan akurasi 6.7% dan AUC 4%, sedangkan menggunakan PSO mampu meningkat Akurasi 9.9% dan AUC 10.4%..

Kata kunci: *N-gram, Information Gain, Particle Swarm Optimization, Naïve Bayes, Google Classroom*.

1. Pendahuluan

Perkembangan ilmu pengetahuan dan teknologi dewasa ini mengalami peningkatan signifikan [1]. Melalui pembelajaran secara daring diharapkan peserta didik dapat lebih mengembangkan kemampuan dalam memecahkan. Salah satu cara yang dapat digunakan yaitu dengan melakukan proses pembelajaran secara daring menggunakan *Google Classroom* [2].

Dengan teknologi *Google Classroom* para instruktur dan siswa memungkinan untuk berbagi materi, mengirimkan tugas serta terhubung dan mengobrol secara *online* [3].

Namun terkadang teknologi *Google Classroom* belum bisa dijadikan sebagai standar keberhasilan sebuah pelajaran. Oleh karena itu di sini peneliti melakukan penelitian mengenai analisis komentar dari para murid terhadap penggunaan *Google Classroom* pada tingkat SMP dan SMA di Kota Bekasi melalui *Google Form*, kemudian hasil dilakukan analisis sentimen dengan mengintegrasikan fitur *information gain*, N-gram dan PSO menggunakan algoritma *naïve bayes* untuk melihat hasill akurasi.

Analisis sentimen adalah proses komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas [4]. Tujuan dari analisis sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan [5]

Penelitian mengenai analisis *Google classroom* belum pernah dilakukan sebelumnya, namun peneliti melihat

Diterima Redaksi : 04-08-2019 | Selesai Revisi : 16-08-2019 | Diterbitkan Online : 02-12-2019

ada beberapa peneliti pernah menggunakan fitur dan algoritma yang sama dalam meningkatkan akurasi.

Di antara para peneliti yang pernah melakukan mengenai analisis sentimen adalah Li, Zhao, Liu, Wang (2016) yang menggunakan *Naive Bayes* untuk menganalisis masalah *sparsity* dalam representasi teks dalam *Neural Network*, sehingga diperlukan modifikasi pembobotan N-gram dalam klasifikasinya [6]. Penelitian lainya pernah dilakukan juga oleh Mathew dan Ramani (2017) yang menggunakan *Naïve Bayes* untuk memperbaiki hasil akurasi filter spam yang disediakan penyedia layanan dengan menggunakan teknik N-gram [7]. Penelitian lain juga pernah dilakukan oleh Ali Fauzi, Arifin, Gosaria, & Prabowo (2017) yang menggunakan fitur *Information Gain* dan *Maximal Marginal Relevance for Feature Selection* untuk mengurangi dimensi berita online yang cukup tinggi pada saat proses klasifikasi menggunakan algoritma *Naïve Bayes* [8].

Pada penelitian ini penulis menggunakan algoritma *Naïve Bayes* dengan menggunakan fitur *information gain* dan *N-gram* untuk menganalisis komentar para siswa terhadap *Google Classroom* dengan menambahkan *Particle Swarm Optimization* sehingga nilai akurasi menjadi lebih baik.

Identifikasi masalah dari penelitian yang dilakukan penulis antara lain:

1. Banyak sekolah yang menggunakan *Google Classroom* sebagai *e-learning* bagi sekolahnya, namun apakah bisa dijadikan patokan keberhasilan proses pembelajaran di sekolah.
2. Bagaimana hasil akurasi analisis sentimen yang dihasilkan dengan menggunakan *Naïve Bayes*.
3. Bagaimana hasil akurasi yang dihasilkan dengan menggunakan fitur *Information Gain, N-gram* dan *Particle Swarm Optimization*.

2. Metode Penelitian

Pada penelitian ini penulis mencoba untuk mengintegrasikan *Information Gain, N-gram* dan PSO pada *Naïve Bayes* untuk menghasilkan akurasi terbaik dalam menganalisis komentar siswa terhadap *Google Classroom.*

Proses yang dilakukan agar mampu menghasilkan akurasi terbaik adalah sebagai berikut

2.1. Pengumpulan data

Proses pengumpulan data komentar siswa dilakukan selama 1 bulan dimulai dari tanggal 15 April 2019 sampai dengan 9 Mei 2019 melalui *Google Form* setelah proses pembelajaran berakhir.

Data diambil dengan meminta pendapat siswa dan siswi SMP kelas 7, 8, 9 dan SMA kelas 10, 11, 12, karena mereka sudah menggunakan *Google Classroom* untuk membantu proses pembelajaran, pemberian tugas serta ujian secara *online* selama 3 tahun.

2.2. *Preprocessing* data

Preprocessing adalah proses pembersihan dan mempersiapkan teks untuk klasifikasi [9]. Ada beberapa proses yang dilakukan dalam *preprocessing* data di antaranya yaitu *Tokenization, Case Folding, Filtering, Stopwords Removal, Stemming* [10].

1. *Tokenization*
   Proses tokenisasi pada data teks adalah melakukan pemecahan sekumpulan karakter (kalimat) menjadi potongan karakter atau kata-kata sesuai kebutuhan yang sering disebut token.
2. *Case Folding*
   Case folding adalah mengubah semua huruf besar atau kapital dalam data menjadi huruf kecil [11].
3. *Filtering*
   Merupakan tahapan mengambil kata-kata penting dari hasil token. Dapat menggunakan algoritma *stop list* (membuang kata yang kurang penting) atau *word list* (menyimpan kata penting) [12].
4. *Stopwords Removal*
   Stopwords removal adalah kumpulan daftar kata-kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi, seperti imbuhan dan *pronoun* seperti “it” dan “they” [13].
5. *Stemming*
   Stemming adalah suatu proses untuk mereduksi kata ke bentuk dasarnya. Tahap *stemming* merupakan tahap mencari akar (*root*) kata dari tiap kata hasil *filtering* [14].
6. *N-gram*
   N-gram merupakan sub-urutan n karakter dari kata yang diberikan. Misalnya, "efficiency" dapat diwakili dengan *character* n-gram yang ditunjukkan pada Tabel 1 [15].

Tabel 1. Contoh penerapan character N-gram

n	Character n-gram sample
2-Grams(n=2)	ef-fi-ic-ci-ie-en-nc-cy
3-Grams(n=3)	eff-ffi-fic-ici-cie-ien-ency
4-Grams(n=4)	effi-ffic-fici-icie-cien-ency

2.3. Seleksi Fitur dengan *Information Gain*

Seleksi fitur yang digunakan pada penelitian di sini adalah *information Gain*. Seleksi fitur dilakukan untuk mengurangi fitur yang tidak relevan dan mengurangi dimensi fitur pada data *text mining*. Berdasarkan penelitian yang dilakukan Ramaswami dan BHaskaran, dengan menerapkan seleksi fitur *Information Gain* dapat meningkat hasil akurasi [16].

$$info(D) = -\sum_{i=1}^{c} \quad p_i log_2(p_i) \qquad (1)$$

Keterangan rumus di atas adalah:

c : jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi)

p_i : jumlah sampe untuk kelas i

$$info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} x\ info(D_j) \quad (2)$$

Keterangan rumus adalah:

A : atribut

$|D|$: jumlah seluruh sampel data

$|D_j|$: jumlah sampel untuk nilai j

v : suatu nilai yang mungkin untuk atribut A

nilai *information gain* akan didapatkan dengan perhitungan menggunakan rumus di bawah ini [17] :

$$Gain\ (A) = |info(D) - info_A(D)| \quad (3)$$

2.4. *Particle Swarm Optimization(PSO)*

PSO adalah metode pencarian populasi, yang berasal dari penelitian untuk pergerakanberkelompok burung dan ikan dalam mencari makan[18]. PSO banyak digunakan untuk memecahkan masalah optimasi dan sebagai pemecah masalah seleksi fitur menurut Liu (2012)[19]. Gambar harus diacu dan dirujuk dalam text.

Langkah-langkah yang dilakukan pada PSO adalah sebagai berikut:

1. Menyiapkan sampel dataset, siap untuk menghitung bobot dari *record* pertama.
2. Inilisialisasi populasi (*swarm*) P(t) sehingga ketika t = 0, lokasi $x_i(t)$ dari setiap partikel $Pi \in P(t)$ di suatu ruang yang luas menjadi acak.
3. Melalui setiap posisi partikel saat itu, $x_i(t)$ mengevaluasi performa dari F.
4. Membandingkan kinerja masing-masing individu saat ini dengan individu yang mempunyai kineja terbaik yang sejauh ini dimiliki jika $F(x_i(t)) > pbest_i$ maka
$$\begin{cases} pbest_i = F(x_i(t)) \\ X_{pbest_i} = x_i(t) \end{cases} \quad (4)$$
5. Membandingkan kinerja masing-masing partikel dengan kinerja partikel terbaik secara global, jika $F(x_i(t)) gbest_i$ maka
$$\begin{cases} gbest_i = F(x_i(t)) \\ X_{gbest_i} = x_i(t) \end{cases} \quad (5)$$
6. Jika perhitungan bobot dari record terakhir di data sampel selesai, maka akhiri. Jika tidak, maka siap untuk menghitung bobot dari *record* berikutnya dan kembali mengulangi dari langkah nomor 1.
7. Setelah itu, lalu hitung rata-rata bobot yang telah dihitung dari data sampel tersebut.

2.5. *Naïve Bayes*

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Tahapan dalam algoritma *Naïves Bayes* [20]:

1. Perhatikan D adalah *record training* dan ketetapan label-label kelasnya dan masing-masing *record* dinyatakan n atribut (n field)$X=(X_1,X_2,\ldots\ldots Xn)$
2. Misalkan terdapat m kelas $C_1,C_2,\ldots\ldots Cn$
3. Klasifikasi adalah diperoleh maksimum *posteriori* yaitu maximum $P(Ci|X)$
4. Ini diperoleh dari Teorema *Bayes*
$$P(C_i|X) = \frac{P(X|C_1)P(C_1)}{P(X)} \quad (6)$$
5. Karena $P(X)$ adalah konstan untuk semua kelas, hanya perlu dimaksimalkan.

3. Hasil dan Pembahasan

Pembahasan dalam proses penelitian ini dilakukan secara bertahap, berawal dari proses pengambilan data, *pre-processing* data, *N-gram*, seleksi fitur, proses validasi dan *Particle Swarm Optimizations*. Semua tahapan proses dilakukan dengan menggunakan data komentar siswa melalui *Google Form*.

3.1. Pengumpulan Dataset

Proses pengambilan data dilakukan melalui *Google Form* di alamat gg.gg/gc_2019. Yang dilakukan oleh para siswa SMP dan SMA, dengan memberikan komentar terhadap penggunaan *Google Classroom* dalam proses pembelajaran sebanyak 6 komentar positif dan 6 komentar negatif.

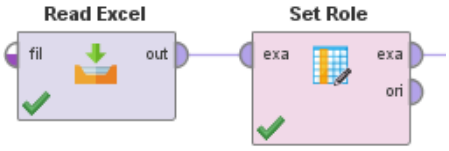
Tabel 2. Komentar *Google Classroom*

Email Address	opnion	label
1617.dindalukinta@globalprestasi.sch.id	It's very convenient, easy to use, always updated with classworks or materials, easier for the teachers and students to share their works, students can study their materials without using any papers so it's eco friendly, students and teachers can update their materials/works anytime.	positive

1617.dindalukinta@globalprestasi.sch.id	Sometimes the materials/works are gone or can not be found, errors in submitting the works, not being cooperative with the users, sometimes login into the classroom can be annoying, it's a bit hard to edit the works after submitting, there aren't a lot of options.	negative

3.2. *Pre-Processing* Data

Setelah dataset dikumpulkan dari *Google Form* berupa file *spreadsheet*, kemudian lakukan proses pengolahan data atau *Pre-Processing* data dengan menggunakan *Tools RapidMiner Studio 9.2*, dengan menetapkan *attribute* dan label terlebih dahulu pada operator *set role*.

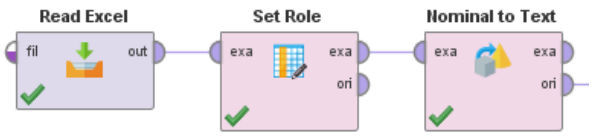


Gambar 1. *Set role* setelah terkoneksi dengan file excel

Setelah operator *set role* terhubung, selanjutnya tentukan *attribute name* dan label.

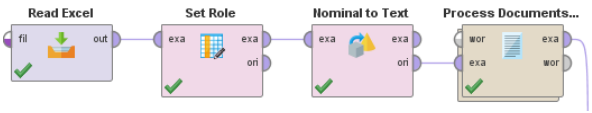
Gambar 2. Penentuan attribut dan label

Setelah penentuan atribut dan label, langkah selanjutkanya melakukan konversi dari *nominal to string*.



Gambar 3. Konversi *nominal to text*

Setelah terkonversi menjadi *binomial*, lakukan proses *pre-processing document*, yang terdiri dari proses :

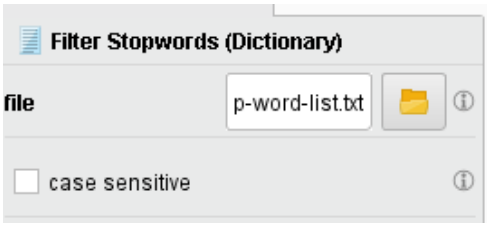


Gambar 4. Proses menghububgkan ke operator *process document*

Di dalam operator *process document* hubungkan beberapa operator dalam *RapidMiner*, yaitu *Tokenize*, *Transform case*, *Filter token*, *filter sopwords* dan *stem*.

Gambar 5. *Process documents*

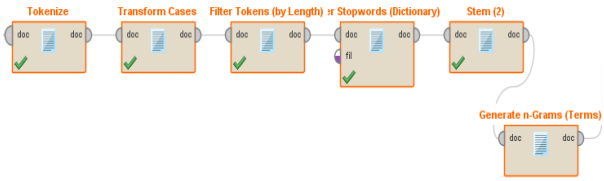
Di dalam *process document* terdapat operator *stop words* yang digunakan untuk membuang kata-kata yang bersifat umum dalam data set seperti "because" atau "about".



Gambar 6. Filter *Stopwords*

3.3. *N-gram*

Dalam *process document* setelah proses *stemming*, maka dilakukan proses penerapan *character N-gram*. Berikut ini gambaran didalam *process document*.

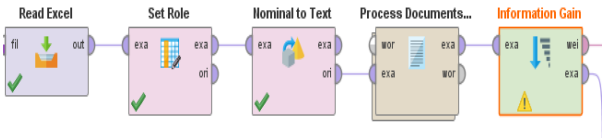


Gambar 7. Penggunaan operator N-gram di *RapidMiner*

Dengan menggunakan *N-gram*, diharapkan dapat meningkatkan hasil akurasi.

3.4 Seleksi fitur dengan *Information Gain*

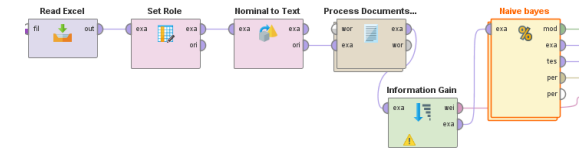
Setelah proses yang dilakukan dalam operator *process document* selesai dilakukan, langkah selanjutnya adalah menghubungkan ke fitur *selection Information Gain*, untuk selanjutnya dilakukan pembobotan *attribute*.



Gambar 8. Menghubungkan fitur *selection Information Gain*

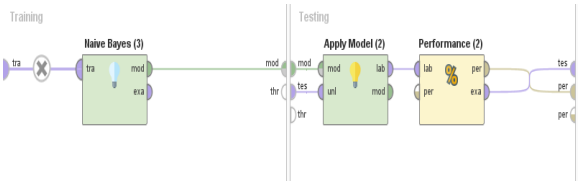
3.5 *Cross validation* (K-Fold)

Setelah proses pembuatan *dataset* terbentuk, langkah selanjutnya adalah proses penggunaan algoritma *Naïve Bayes*, dimana algoritma berbentuk operator ini akan diletakan dalam operator *Cross-validation*. Sehingga untuk pengetesan data *training*, data *testing*, pengecekan akurasi, *precision*, AUC dan F1 dapat dilakukan sekaligus.

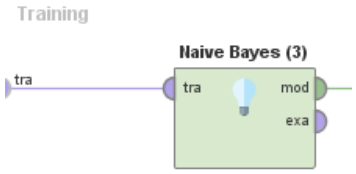


Gambar 9. Menghubungkan operator *Cross validation*

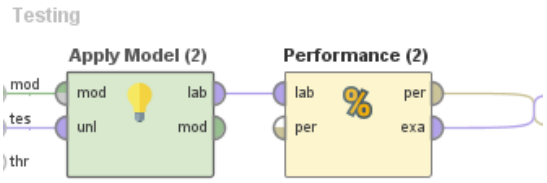
Dalam operator *cross validation* terdapat beberapa operator, di antaranya *Naïve bayes*, Model dan *performance*.



Gambar 10. Bagian dalam *Cross Validation*



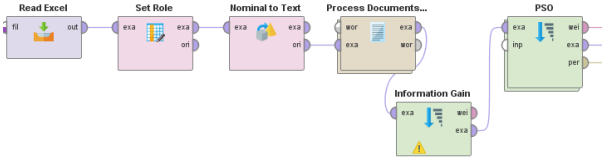
Gambar 11. Operator *Naïve Bayes*



Gambar 12. Operator *Apply model* dan *Performance*

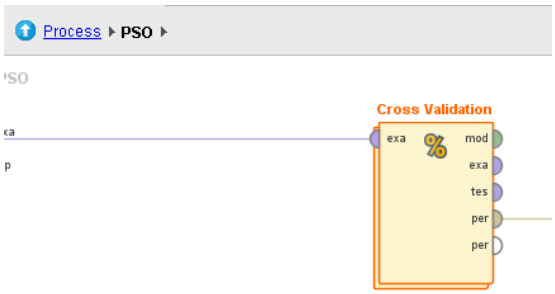
3.6 *Particle Swarm Optimization* (PSO)

Untuk meningkatkan hasil dari akurasi pada model yang sudah dibuat kemudian dilanjutkan proses klasifikasi menggunakan *Naïve bayes*. Perlu ditambahkan operator *Particle Swarm Optimization*. Berikut gambaran operator dalam *RapidMiner*.



Gambar 13. Operator PSO dalam proses

Dalam proses *Particle Swarm Optimization* adalah *Cross Validation*.



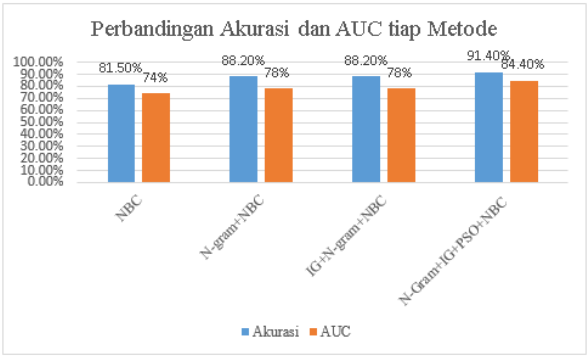
Gambar 14. Bagian dalam PSO

3.5 Hasil Penelitian

Setelah terhubung semua operator, maka peneliti melakukan beberapa percobaan dengan menggunakan variasi metode dan *cross validation* dengan K=10. Berikut ini merupakan hasil dari beberapa percobaan.

Tabel 3. Hasil *Cross Validation*

Model	Akurasi	*AUC*	*F1*
NBC	81.5%	74%	81.9
N-gram+NBC	88.2%	78%	87.8%
IG+N-gram+NBC	88.3%	78.9%	87.8%
N-Gram+IG+PSO+NBC	91.4%	84.4%	91.24%



Gambar 13. Grafik perbandingan hasil Akurasi dan AUC

4. Kesimpulan

Dari hasil percobaan di atas dapat dilihat bahwasanya penggunaan penggunaan N-gram dalam meningkatkan optimasi hasil sentimen *Google Classroom* dapat meningkatkan akurasi sebesar 6.7% dan AUC sebesar 4% dibandingkan tidak menggunakan N-gram. Kemudian setelah ditambahkan metode yang berbeda yaitu *Particle Swarm Optimization* pada klasifier *Naïve Bayes* Akurasi dan AUC mengalami peningkatan sebesar 9.9% dan 10.4%. Di sini dapat disimpulkan penggunaan N-gram dan PSO sangat membantu dalam mengoptimalkan hasil Akurasi dan AUC, namun penambahan *Information Gain* dalam penelitian di sini tidak memberikan peningkatan yang berbeda, mungkin

ke depan untuk penelitian selanjutnya penggunaan parameter tambahan dan *clasifier* yang berbeda dapat memperbaiki peningkatan hasil akurasi dan AUC dengan seleksi fitur *Information Gain*.

Daftar Rujukan

[1] V. D. Wicaksono and P. Rachmadyanti, "Pembelajaran Blended Learning melalui Google Classroom di Sekolah Dasar," *Semin. Nas. Pendidik. PGSD UMS HDPGSDI Wil. Jawa*, pp. 513–521, 2017.

[2] F. I. Gunawan, "Pengembangan Kelas Virtual Dengan Google Classroom Dalam Keterampilan Pemecahan Masalah (Problem Solving) Topik Vektor Pada Siswa Smk Untuk Mendukung Pembelajaran," *Pros. Semin. Nas. Etnomatnesia*, pp. 340–348, 2017.

[3] L. Abazi-Bexheti, A. Kadriu, M. Apostolova-Trpkovska, E. Jajaga, and H. Abazi-Alili, "LMS Solution: Evidence of Google Classroom Usage in Higher Education," *Bus. Syst. Res.*, vol. 9, no. 1, pp. 31–43, 2018.

[4] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[5] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.

[6] B. Li, Z. Zhao, T. Liu, P. Wang, and X. Du, "Weighted Neural Bag-of-N-gram Model : New Baselines for Text Classification," *Coling*, pp. 1591–1600, 2016.

[7] N. V. Mathew and V. Ramani Bai, "Analyzing the Effectiveness of N-gram Technique Based Feature Set in a Naive Bayesian Spam Filter," *Proc. IEEE Int. Conf. Emerg. Technol. Trends Comput. Commun. Electr. Eng. ICETT 2016*, 2017.

[8] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 3, pp. 610–615, 2017.

[9] T. Singh and M. Kumari, "Role of Text Pre-processing in Twitter Sentiment Analysis," *Procedia Comput. Sci.*, vol. 89, pp. 549–554, 2016.

[10] C. Paper, "Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining," no. October 2014, 2016.

[11] F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Teknol. Inf. ITSmart*, vol. 2, no. 2, p. 35, 2016.

[12] M. Sulhan and R. Kurniawan, "Metode Stemming Sebagai Preprocessing Pada Filter Kata Porno Melalui Aspek Pendidikan," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2014, no. Sentika, pp. 52–60, 2014.

[13] A. Rachmat and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.

[14] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, p. 269, 2018.

[15] B. C. Gencosman, H. C. Ozmutlu, and S. Ozmutlu, *Character n-gram application for automatic new topic identification*, vol. 50, no. 6. Elsevier Ltd, 2014.

[16] B. N. Sari, "Implementasi Teknik Seleksi Fitur Information Gain pada Algoritma Klasifikasi Machine Learning untuk Prediksi Performa Akademik Siswa," *Seminar Nasional Teknologi Informasi dan Multimedia 2016*. pp. 6–7, 2016.

[17] L. Dini Utami and R. S. Wahono, "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes," *J. Intell. Syst.*, vol. 1, no. 2, pp. 120–126, 2015.

[18] A. D. R. Prabowo and M. Muljono, "Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization," *Techno.Com*, vol. 17, no. 2, pp. 208–219, 2019.

[19] B. Liu, "Sentiment Analysis and Opinion Mining," no. May, 2012.

[20] M. Han,Jiawei Han. Kamber, *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2011.