



Jurnal **INFORMATIKA**

P-ISSN: **2355-6579** E-ISSN: **2528-2247**



Diterbitkan Oleh:

**Lembaga Penelitian dan Pengabdian Masyarakat
Universitas Bina Sarana Informatika**



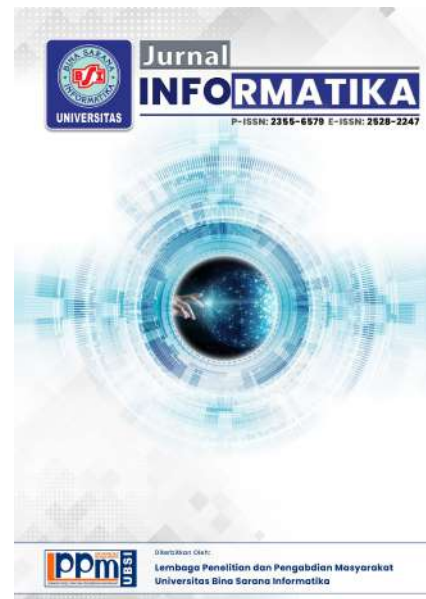
Home > **Vol 9, No 1 (2022)**

Jurnal Informatika

Jurnal Informatika first publication in 2014 (ISSN: e. 2528-2247 p. 2355-6579) is scientific journal research in Informatics Engineering, Informatics Management, and Information Systems, published by Universitas Bina Sarana Informatika which the articles were never published online or in print. The publication is scheduled twice a year (April and September). The Editor welcomes submissions of manuscripts that relate to the field.

Jurnal Informatika has been accredited by SINTA 4 (No. Surat Keterangan : 148/M/KPT/2020 Judul Surat Keterangan : Peringkat Akreditasi Jurnal Ilmiah Periode II Tahun 2020 Tanggal Surat Keterangan : 03 Aug 2020) and indexed by Google Scholar, Academic Resource Index, PKP Index, One Search, DOAJ, SINTA, Base, Scilit.

ISSN: 2355-6579 (Print), E-ISSN: 2528-2247 (Online)



Announcements

Changing Publication Schedule Information

In 2022, Jurnal informatika publication changed into April and October, which were last published in April and September.

Mulai tahun 2022 penerbitan Jurnal Informatika yang sebelumnya terjadwal bulan April dan September diubah menjadi April dan Oktober.

Posted: 2022-03-22

[More Announcements...](#)

Vol 9, No 1 (2022): April 2022

Table of Contents

Articles

Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly
10.31294/inf.v9i1.10209
Melany Mustika Dewi

PDF (BAHASA INDONESIA)
1-5

Sistem Penunjang Keputusan Dalam Penerimaan PraKerja Dengan Menggunakan Metode Analytic Hierarchy Process
10.31294/inf.v9i1.11032

PDF (BAHASA INDONESIA)
6-12

ADDITIONAL MENU

[Focus And Scope](#)

[Download Template](#)

[Online Submission](#)

[Review Process](#)

[Publication Fees](#)

[Publication Ethics](#)

[Plagiarism Check](#)

[Contact](#)

[Reviewer](#)

[Review Guideline](#)

[Article Statistic](#)

[Acreditation Certificate](#)

USER

Username

Password

Remember me

VISITOR COUNTER

762.733

Visitor Statistic

TEMPLATE



ORIGINALITY STATEMENT



TOOLS





[HOME](#) [ABOUT](#) [LOGIN](#) [REGISTER](#) [SEARCH](#) [CURRENT](#) [ARCHIVES](#) [ANNOUNCEMENTS](#)

Home > About the Journal > **People**

People

Reviewer

Irawan Afrianto, [Scopus id: 57204188766] Universitas Komputer Indonesia, Indonesia

Aulia Akhrian Syahidi, [Scopus Id: 57204180035] Politeknik Negeri Banjarmasin, Indonesia

Siti Wardah, [Scopus Id: 57211755511] Universitas Islam Indragiri, Indonesia

Lena Magdalena, [Scopus id: 57220184654] Universitas Catur Insan Cendekia, Indonesia

Akbari Indra Basuki, [Scopus id: 56715961400] Pusat Penelitian Informatika - LIPI, Indonesia

Linda Marlinda, Universitas Nusa Mandiri, Indonesia

Tuti Alawiyah, AMIK BSI Tasikmalaya, Indonesia

Embun Fajar Wati, Universitas Bina Sarana Informatika, Indonesia

Tati Mardiana, Unifersitas Bisa Sarana Informatika, Indonesia

Mr. Sopiyan Dalis, Universitas Bina Sarana Informatika, Indonesia

Rizki Wahyudi, (Scopus ID: 57209266675, Universitas Amikom Purwokerto), Indonesia

Saruni Dwasnati, [Scopus id: 57210968603] Universitas Mercubuana, Indonesia

Anggi Anggi Puspita Sari, Universitas Bina Sarana Informatika, Indonesia

Mr. Adi Supriyatna, AMIK BSI Karawang, Indonesia

Fransiskus Panca Juniawan, [Scopus id: 57208758178] ISB Atma Luhur, Pangkal Pinang

Lukman Nulhakim, M.Kom, [Scopus id: 54791427100] STMIK Antar Bangsa

Soleh Oleh Soleh, [Scopus id: 57201744999] Raharja University, Indonesia

Jenie Sundari, Universitas Nusa Mandiri, Indonesia

Esrn Rikardo Nainggolan, [Scopus ID: 57208281339] Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Indonesia

Andi Saryoko, (Scopus ID: 57196050569) Universitas Nusa Mandiri

Sari Susanti, Universitas Adhirajasa Reswara Sanjaya, Indonesia

Wawan Gunawan, Universitas Mercu Buana, Indonesia

Kusuma Hati, STMIK Antar Bangsa, Indonesia

Nurmalasari Nurmalasari, [Scopus id: 57222337220] Universitas Nusa Mandiri, Indonesia

Rachmat Hidayat, Universitas Bina Sarana Informatika, Indonesia

Budi Sudradjat, AMIK BSI Jakarta, Indonesia

Noor Cholis Basjaruddin, Politeknik Negri Bandung, Indonesia

Mr. Bambang Eka Purnama, (Scopus ID : 56968140300) Universitas Bina Sarana Informatika, Indonesia

Caesario Satriatama, Universitas Trilogi, Indonesia

ADDITIONAL MENU

[Focus And Scope](#)

[Download Template](#)

[Online Submission](#)

[Review Process](#)

[Publication Fees](#)

[Publication Ethics](#)

[Plagiarism Check](#)

[Contact](#)

[Reviewer](#)

[Review Guideline](#)

[Article Statistic](#)

[Acreditation Certificate](#)

USER

Username

Password

Remember me

VISITOR COUNTER

762.731

Visitor Statistic

TEMPLATE



ORIGINALITY STATEMENT



TOOLS





Index by:



Published by Department of Research and Public Service (LPPM) Universitas Bina Sarana Informatika with supported Relawan Jurnal Indonesia



Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta 10450



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License



[HOME](#) [ABOUT](#) [LOGIN](#) [REGISTER](#) [SEARCH](#) [CURRENT](#) [ARCHIVES](#) [ANNOUNCEMENTS](#)

Home > Archives > **Vol 7, No 1 (2020)**

Vol 7, No 1 (2020)

April 2020

DOI: <https://doi.org/10.31294/ji.v7i1>

Full Issue

[View or download the full issue](#)

[COVER \(BAHASA INDONESIA\)](#) [DAFTAR ISI \(BAHASA INDONESIA\)](#)

Table of Contents

Articles

Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat 10.31294/ji.v7i1.6203 <i>Herry Derajad Wijaya, Saruni Dwiasnati</i>	PDF (BAHASA INDONESIA) 1-7
Optimasi Prediksi NilaiTukar Rupiah Terhadap Dolar Menggunakan Neural Network Berbasiskan Algoritma Genetika 10.31294/ji.v7i1.6793 <i>Primandani Arsi, Joko Prayogi</i>	PDF (BAHASA INDONESIA) 8-14
ANALISA SENTIMENT PADA ULASAN FILM DENGAN OPTIMASI ENSEMBLE LEARNING 10.31294/ji.v7i1.6171 <i>Andreyestha Andreyestha, Agus Subekti</i>	PDF (BAHASA INDONESIA) 15-23
Aplikasi Pembelajaran Huruf Hangeul Berbasis Android 10.31294/ji.v7i1.6390 <i>Agung Baitul Hikmah, Yanti Apriyani, Shelin Ayuni Purwandhani</i>	PDF (BAHASA INDONESIA) 24-29
Analisa Perancangan Sistem Penerimaan Donasi dan Laporan Pertanggung Jawaban Pada Yayasan Yatim Piatu Aisyiyah 10.31294/ji.v7i1.6245 <i>Nurlila Nurlila, Nia Rahma Kurnianda</i>	PDF (BAHASA INDONESIA) 30-36
Implementasi Vigenere Cipher Sebagai Pengaman Pada Proses Deskripsi Steganografi Least Significant Bit 10.31294/ji.v7i1.6431 <i>Tuti Alawiyah, Rian Ardianto, Dini Silvi Purnia</i>	PDF (BAHASA INDONESIA) 37-45
Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification 10.31294/ji.v7i1.6920 <i>Jasman Pardede</i>	PDF (BAHASA INDONESIA) 46-54
Komparasi Algoritma C4.5, Naïve Bayes dan k-Nearest Neighbor Sebagai Sistem Pendukung Keputusan Meningkatkan Jumlah Peserta Didik 10.31294/ji.v7i1.7250 <i>Harianto Harianto, Didi Rosiyadi</i>	PDF (BAHASA INDONESIA) 55-61
Kajian Website Pendaftaran Peserta Pelatihan Pada Kemendag Dengan Pendekatan	PDF (BAHASA INDONESIA) 62-67

ADDITIONAL MENU

[Focus And Scope](#)

[Download Template](#)

[Online Submission](#)

[Review Process](#)

[Publication Fees](#)

[Publication Ethics](#)

[Plagiarism Check](#)

[Contact](#)

[Reviewer](#)

[Review Guideline](#)

[Article Statistic](#)

[Acreditation Certificate](#)

USER

Username

Password

Remember me

VISITOR COUNTER

762.730

Visitor Statistic

TEMPLATE



ORIGINALITY STATEMENT



TOOLS





Technology Acceptance Model

10.31294/ji.v7i1.7576

Irawan Satriadi, Amrin Amrin

SISTEM PAKAR DETEKSI PENYAKIT REFRAKSI MATA DENGAN METODE TEOREMA BAYES BERBASIS WEB

PDF (BAHASA INDONESIA)

68-76

10.31294/ji.v7i1.7267

Rizal Rachman

Pendeteksi Kesalahan Pengetikan Kata Non Baku pada Karya Tulis Menggunakan Metode N-Gram

PDF (BAHASA INDONESIA)

77-84

10.31294/ji.v7i1.7916

Titi Hartina, Agustin Agustin

Penggunaan Algoritma Klasifikasi Terhadap Analisa Sentimen Pemandangan Ibukota Dengan Pelabelan Otomatis

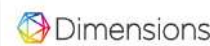
PDF (BAHASA INDONESIA)

85-90

10.31294/ji.v7i1.7528

Jananto Watori, Riska Aryanti, Agus Junaidi

Index by:



Published by Department of Research and Public Service (LPPM) Universitas Bina Sarana Informatika with supported Relawan Jurnal Indonesia



Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta 10450



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification

Jasman Pardede¹, Yusup Miftahuddin², Wahidul Kahar³

^{1,2}Jurusan Teknik Informatika, Fakultas Teknologi Industri, Institut Teknologi Nasional, Bandung
Jl.PH.H. Mustofa No.23 Bandung 40124, Indonesia
e-mail: ¹jasman@itenas.ac.id, ²ymiftahuddin@gmail.com, ³wahidulkahar96@gmail.com

Abstrak

Pesatnya perkembangan teknologi dan media sosial dapat memudahkan pengguna untuk menyampaikan informasi. Selain itu, media sosial juga memberikan dampak negatif dengan cara memposting tulisan kejam atau berkomentar semena-mena tanpa memikirkan akibat pada orang lain. Hal inilah yang menjadikan salah satu terjadinya tindak kekerasan dalam dunia maya (*Cyberbullying*). Tahapan awal yang dilakukan dalam penelitian ini adalah pengolahan bahasa atau yang disebut dengan *text preprocessing* meliputi *tokenizing*, *casefolding*, *stopword removal* dan *stemming*. Kemudian *feature selection* yaitu mengubah dokument teks menjadi matriks dengan tujuan untuk mendapatkan fitur pada setiap kata untuk dijadikan parameter atau kriteria klasifikasi. Untuk pengambilan keputusan apakah komentar mengandung makna *bully* atau *nonbully* menggunakan algoritma *Naïve Bayes Classification* dengan model *multinomial naïve bayes*. Perhitungan yang dilakukan adalah menghitung nilai probabilitas setiap kata yang muncul berdasarkan *class* dan nilai perkalian *class conditional probability*. Berdasarkan hasil eksperimen menggunakan dataset "*cyberbullying comments*" yang diambil dari *Kaggle* akurasi yang didapat sebesar 80%, *precision* 81% dan *recall* 80%.

Kata kunci: *Cyberbullying, Feature Selection, Naïve Bayes Classification*

Abstract

The rapid development of technology and social media can facilitate users to convey information. Besides, social media also has a negative impact by posting cruel posts or commenting arbitrarily without thinking about the consequences on others. This is what makes one of the acts of violence in cyberspace (Cyberbullying). The initial stages carried out in this research are language processing or what is called text preprocessing including tokenizing, case-folding, stopwords removal and stemming. Then the feature selection is to change the text document into a matrix to get the features of each word to be used as a parameter or classification criteria. For decision making whether comments contain bully or non-bully meaning using the Naïve Bayes Classification algorithm with the multinomial naïve Bayes model. The calculation is done is to calculate the probability value of each word that appears based on the class and the value of the class conditional probability multiplication. Based on the results of experiments using the dataset "cyberbullying comments" taken from Kaggle obtained 80% accuracy, 81% precision and 80% recall.

Keywords: *Cyberbullying, Feature Selection, Naïve Bayes Classification*

Pendahuluan

Deteksi dalam kamus Kamus Bahasa Besar Bahasa Indonesia (KBBI) adalah usaha menemukan dan menentukan keberadaan, anggapan, atau kenyataan. Deteksi juga merupakan suatu proses untuk memeriksa atau melakukan

pemeriksaan kembali terhadap sesuatu dengan menggunakan cara dan metode teknik tertentu.

Media sosial salah satu sebagai cara berkomunikasi yang efisien, nyaman, dan populer (Barlett, Gentile, Chng, Li, & Chamberlin, 2018). Contoh media sosial,

menurut (Carr & Hayes, 2015) termasuk situs web jejaring sosial Facebook, blogging, Tinder, Instagram, Yik Yak, dan lainnya. Sosial media biasanya digunakan oleh seseorang sebagai media komunikasi, sarana informasi dan sebagai media hiburan bagi penggunanya (Nasution & Hayaty, 2019). Perkembangan teknologi informasi, internet, dan media sosial memberikan dampak perubahan kepada perilaku manusia dalam bersosial dan berkomunikasi (Rifauddin, 2016).

Berdasarkan hasil riset Wearesosial Hootsuite yang dirilis Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi. Jumlah tersebut naik 20% dari survei sebelumnya. Sementara pengguna media sosial mobile (gadget) mencapai 130 juta atau sekitar 48% dari populasi. Besarnya populasi, pesatnya pertumbuhan pengguna internet dan telepon merupakan potensi bagi ekonomi digital nasional (databoks.katadata.co.id, 2019). Sebagai contoh penggunaan media sosial yang positif adalah memudahkan pengguna untuk menyebarkan informasi atau pengetahuan di seluruh dunia (Wijanto, 2015). Namun, media sosial juga memberikan dampak negatif seperti memposting tulisan kejam ataupun mengunggah foto yang berhubungan dengan individu lain dengan tujuan mengintimidasi dan merusak nama baik korban sehingga korban merasa tersakiti dan malu (Nasrullah, 2015). Hal inilah salah satu faktor yang menyebabkan adanya *cyberbullying* (Rifauddin, 2016). Salah satu dampak negatif yang muncul dari banyaknya pengguna jejaring sosial menimbulkan adanya *cyberbullying* (Barlett, Gentile, Chng, Li, & Chamberlin, 2018).

Cyberbullying merupakan perilaku seseorang atau kelompok secara sengaja dan berulang kali melakukan tindakan yang menyakiti orang lain melalui komputer, telepon seluler, dan alat elektronik lainnya (Rifauddin, 2016). *Cyberbullying* biasanya didefinisikan sebagai penyerangan yang sengaja dan berulang kali dilakukan secara elektronik konteks (misalnya, e-mail, blog, pesan instan, dan pesan teks) terhadap orang yang tidak dapat dengan mudah membela diri atau dirinya sendiri (Tian, Yan, & Huebner, 2018). *Cyberbullying* sebagai tindakan atau perilaku agresif dengan menggunakan elektronik terhadap suatu kelompok atau individu tertentu yang

dilakukan secara berulang kali dan seiring waktu, secara online didunia internet (Nilan, Burgess, Hobbs, Threadgold, & Alexander, 2015). *Cyberbullying* adalah tindakan mengintimidasi menggunakan elektronik, seperti pesan instan, email, ruang obrolan, situs web, game online, situs jejaring sosial, dan pesan teks (Hase, Goldberg, Smith, Stuck, & Campaign, 2015). Walaupun tidak bertemu secara langsung tatap muka (*face to face*) tindakan *cyberbullying* juga bisa memakan korban karena tidak ada larangan sedikitpun bagi orang berkomentar pada kolom komentar (Natalia, 2016). Memposting tulisan kejam ataupun mengunggah foto yang berhubungan dengan individu lain dengan tujuan mengintimidasi dan merusak nama baik korban sehingga korban merasa tersakiti dan malu (Nasrullah, 2015). Hal inilah salah satu faktor yang menyebabkan adanya *cyberbullying* (Rifauddin, 2016).

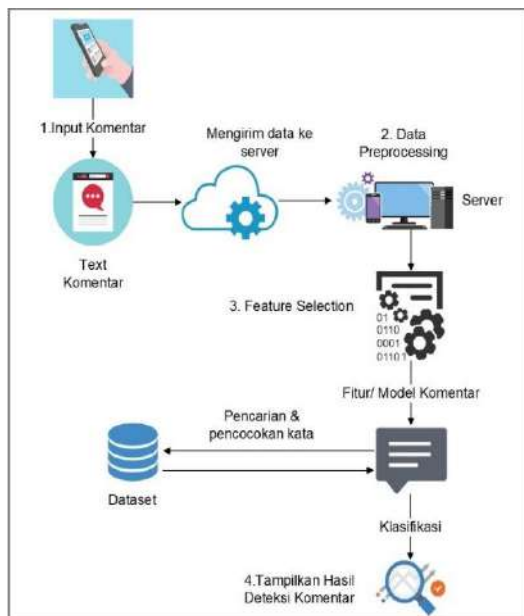
Untuk mengetahui suatu teks komentar mengandung makna *bully* atau tidak, metode yang diusulkan adalah *naïve bayes classification* dengan model *multinomial naïve bayes*. Naive bayes banyak digunakan dalam klasifikasi teks dan memiliki beberapa kelebihan yang mudah diterapkan dan memiliki kinerja yang baik dalam klasifikasi teks (Santoso, Yuniarno, & Hariadi, 2015). Pendekatan *supervised machine learning* (ML) banyak digunakan untuk klasifikasi teks. Algoritma pembelajaran mesin yang paling populer termasuk Naïve Bayes (NB) (Al-Anzi & AbuZeina, 2017). Klasifikasi dalam pengambilan informasi dan klasifikasi teks paling terkenal adalah klasifikasi naïf Bayes (NB) yang telah terbukti sangat efektif pada banyak aplikasi data nyata (Feng, Guo, Jing, & Sun, 2015). Naive Bayes (NB) adalah salah satu dari 10 algoritma teratas berkat kesederhanaannya, efisiensi, dan interpretabilitas untuk melemahkan asumsi independensi atributnya. (Wang, Jiang, & Li, 2015). Klasifikasi Naïve Bayes merupakan penggolong probabilistik berdasarkan Teorema Bayes (Rasjid & Setiawan, 2017).

Multinomial Naive Bayes (MNB) adalah variasi dari *Naive Bayes* yang dirancang untuk menyelesaikan klasifikasi dokumen teks dengan menghitung jumlah kemunculan kata atau bobot kata sebagai fitur klasifikasi (Pratama & Sarno, 2015). Dengan menerapkan pendekatan *machine learning* pada klasifikasi sebuah dokumen

teks serta meningkatkan akurasi klasifikasi teks adalah *Multinomial Naive Bayes* (MNB) (Wang, Jiang, & Li, 2015).

Metode Penelitian Rancangan Sistem

Rancangan sistem yang akan dibangun dalam penelitian terbagi menjadi beberapa bagian diantaranya *preprocessing*, *feature selection*, dan *classification*. Gambar 1 memperlihatkan sistem yang dibangun secara lebih rinci. Bagian sistem yang berinteraksi dengan pengguna aplikasi secara langsung dimana pengguna (*user*) akan memasukkan teks komentar sebagai *inputan*. Teks komentar memberikan hasil berupa komentar yang mengandung *bully* atau *nonbully*. Selain itu juga, dapat diketahui bahwa *inputan* teks yang dicari tentang komentar bersifat negatif (*bully*) dan positif (*nonbully*).



Gambar 1. Rancangan Sistem Deteksi Komentar *Cyberbullying* Pada Media Sosial

Berikut ini merupakan penjelasan dari rangkaian sistem secara keseluruhan yang dibangun. Terdapat 4 (empat) langkah dalam melakukan deteksi komentar *cyberbullying* pada media sosial bahasa Inggris, yaitu:

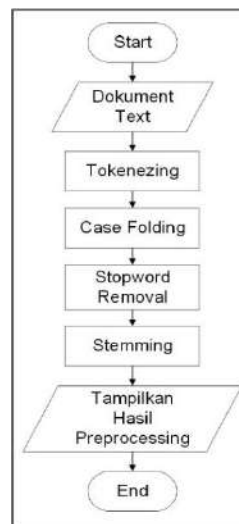
1. Langkah pertama yang dilakukan adalah memasukkan kalimat komentar dalam bentuk text komentar.
2. Kemudian pengolahan data komentar atau disebut dengan *Data Pre-processing* adalah pengolahan bahasa

yang tidak terstruktur menjadi lebih terstruktur.

3. *Feature selection* adalah proses ekstraksi data dokumen teks menjadi matriks dengan tujuan mendapatkan fitur setiap kata yang muncul.
4. Langkah terakhir melakukan pendeteksian atau pengambilan keputusan terhadap teks komentar yang telah dimasukkan sebelumnya. Metode klasifikasi yang digunakan *naïve bayes classification* dengan menggunakan model *multinomial naïve bayes*. Hasil *output* berupa deteksi komentar *bully* atau *nonbully* berdasarkan nilai probabilitas tertinggi.

Preprocessing

Text Preprocessing adalah tahap awal dari teks mining untuk merubah data yang tidak terstruktur menjadi data terstruktur (Kurniawan, Fauzi, & Widodo, 2017). Tujuan lain juga untuk menghilangkan *noise* yang terdapat pada dokumen teks dan mengambil fitur atau parameter penting pada dokumen teks.



Gambar 2. Tahap Pre-processing Secara Umum

Pada Gambar 2 menunjukkan tahapan pengolahan teks komentar dimulai dari inputan sebuah dokumen text, lalu akan dilakukan *tokenizing*, *case-folding*, *stopword removal* dan *stemming* biasanya proses ini disebut dengan *text pre-processing* yang nantinya akan diteruskan proses selanjutnya. *Text pre-processing* juga dapat membantu dalam pengambilan keputusan secara baik berdasarkan pengolahan yang sudah dilakukan

sebelumnya, dengan tahapan sebagai berikut:

- *Tokenizing* adalah adalah memotong urutan karakter menjadi potongan-potongan, yang disebut token, dan pada saat yang sama mungkin membuang karakter tertentu, seperti tanda baca dan selain angka (Hasanah, Astuti, Wahyudi, Rifai, & Pambudi, 2018).
- *Case-folding* adalah mengubah semua huruf menjadi huruf kecil (Hasanah, Astuti, Wahyudi, Rifai, & Pambudi, 2018).
- *Stopword Removal* adalah menghapus kata-kata yang tidak berguna (Al-khurayji & Sameh, 2017).
- *Stemming* adalah megubah kata ke dalam bentuk kata dasar (*root word*) dengan menghilangkan aditif yang ada. Algoritma stemming yang digunakan adalah Porter Stemmer untuk bahasa Inggris (Pratama & Sarno, 2015).

Secara khusus porter stemmer memiliki lima langkah dengan menerapkan aturan disetiap langkah (Ali & Ibrahim, 2015) sebagai berikut:

1) Step 1

Menghilangkan beberapa imbuhan dalam bentuk bentuk *past tense* dan jamak dari kalimat murni. Dengan 3 tahapan sebagai berikut:

- Step 1a*
SSES->SS
caresses -> caress
- Step 1b*
(*v*)ING->
opening -> agree
- Step 1c*
(*v*)Y->I
history -> histori

2) Step 2

Selanjutnya adalah pencocokan pola pada beberapa akhiran umum. Dengan menghapus *derivasi* (menghasilkan kata-kata yang berbeda dari paradigma yang berbeda) akhiran (*d*-sufiks) dan mengikuti aturan sebagai berikut:

- m>0) ATIONAL -> ATE
relational -> relate

3) Step 3

Menghilangkan akhiran kata special dan juga menghilangkan derivasi sufiks. Kemudian gabungan dari derivasi sufiks dihilangkan menjadi *single derivasi sufiks*.

Lalu jika akhiran *-icational*, pada step 2 direduksi menjadi *-icate* dan pada step ini direduksi menjadi *-ic*. Di bawah adalah contoh aturan yang diterapkan pada Langkah 3.

- m>0)NESS->
possibleness -> possible

4) Step 4

Pada step 4 adalah menghilangkan sufiks lain pada gabungan kata seperti *-ic*, *-able*, *-ive* dan masih banyak yang lain. Contoh aturan yang terlibat dalam langkah ini adalah sebagai ditunjukkan di bawah ini

- (m>1)MENT->
adjustment -> adjust

5) Step 5

Langkah terakhir adalah merapikan algoritma setelah menghapus sufiks pada langkah-langkah sebelumnya. Kemudian cek apakah akhiran berbentuk huruf vokal jika ada diperbaiki dengan tepat. Tahapan ini terdiri dari 2 tahapan sebagai berikut:

- Step 5a*
m>1)E
probate -> probat
- Step 5b*
m>1 and *d and *L-> single letter
bill -> bil

Feature Selection

Tabel 1. Cotoh Hasil Ekstraksi Fitur Pada Dokument Teks

Text	This	is	a	nice	hotel	not	all	at
<i>This is a nice hotel</i>	1	1	1	1	1	0	0	0
<i>Not a nice hotel at all</i>	0	0	1	1	1	1	1	1

Feature selection adalah proses ekstraksi data pada dokumen teks untuk mendapatkan fitur setiap kata yang muncul sebagai kriteria klasifikasi. Salah satu teknik ekstraksi fitur dalam sebuah teks dokument adalah *Bag-of-words* dimana fitur terdiri ini dari individu atau kelompok kata yang ditemukan dalam teks (Etaiwi & Naymat, 2017). Proses ini menghasilkan nilai *numerical vector* dari tiap teks yang

diteruskan ke proses training tiap algoritma klasifikasi (Sastrodikoro W., Palit, & Justinus Andjarwirawan, 2018). Contoh fitur *Bag-of-words* ditunjukkan pada Tabel 1, dimana setiap kemunculan kata dalam ulasan akan diwakili oleh frekuensi kemunculan kata dalam teks. (Etaiwi & Naymat, 2017).

Naïve Bayes Classification

Metode *Naïve Bayes* merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes* (Susanto & Indriyani, 2019). *Naïve Bayes* adalah algoritma klasifikasi berdasarkan teorema Bayes (Pratama & Sarno, 2015). *Naïve Bayes* (NB) adalah model pembelajaran statistik berbasis probabilitas dan memiliki asumsi premis bahwa semua atribut sepenuhnya independen pada nilai variabel kelas (Wang, Jiang, & Li, 2015). Teorema Bayes mengasumsikan bahwa semua variabel independen mempertimbangkan nilai variabel kelas (Saritas & Yasar, 2019). Algoritma *Naïve Bayes Classifier* merupakan pengklasifikasian dengan metode probabilitas, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes (Nugroho, 2018).

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)}$$

dimana:

$P(A|B)$ = peluang terjadinya A dengan syarat B terjadi lebih dahulu

$P(A)$ = Peluang kemunculan A

$P(B|A)$ = peluang terjadinya B dengan syarat A terjadi lebih dahulu

$P(B)$ = Peluang kemunculan B

Multinomial Bayes Classification

Multinomial naïve Bayes (MNB) menangkap frekuensi yang terjadi pada semua kata dalam dokumen (Wang, Jiang, & Li, 2015). *Multinomial Naïve Bayes* tidak jauh berbeda dengan NBC yaitu perhitungan *prior*, *conditional* dan *posterior* akan tetapi untuk *conditional probability* terdapat perubahan (Pujadayanti, Fauzi, & Sari, 2018).

$$P(\mathbf{w}_i | c_j) = \frac{\text{Count}(\mathbf{w}_i, c_j) + 1}{\sum_{\mathbf{w} \in V} (\text{count}(\mathbf{w}, c_j)) + |V|}$$

dimana:

$P(\mathbf{w}_i | c_j)$ = *Conditional Probability*

$\text{Count}(\mathbf{w}_i, c)$ = menghitung jumlah kata W_i dengan $i = 1, 2, 3, \dots, m$ pada kelas j dengan $j = 1, 2, 3, \dots, n$

$\sum_{\mathbf{w} \in V} (\text{count}(\mathbf{w}, c))$ = menghitung jumlah seluruh kata kelas j

$|V|$ = menghitung kata unik pada seluruh dokumen

Hasil dan Pembahasan

1. Dataset

Dataset yang digunakan merupakan dataset "*cyber-bullying comments*" yang diambil dari Kaggle, dimana 1313 komentar *bully* dan 1423 *nonbully* dengan total 2736 data komentar.

2. Implementasi Sistem

Implementasi sistem pada penelitian ini menggunakan bahasa pemrograman python, JetBrains PyCharm Community Edition dan PHP. Library yang digunakan adalah *pandas* (pembacaan file), *nlTK* (pengolahan bahasa), *PorterStemmer* (stemming) dan *scikit-learn* (klasifikasi dan feature extraction). Setiap komentar diinputkan akan diuji satu persatu, kemudian diproses melalui service yang mengembalikan data dalam bentuk JSON.

3. Pengujian Sistem Aplikasi

Pada pengujian aplikasi adalah menguji secara fungsional pada setiap proses yang dilakukan oleh sistem. Adapun proses pengujian yang dilakukan sebagai berikut.

4. Text Preprocessing Proses Tokenizing

Pada proses tokenizing merupakan pemecahan kalimat menjadi kata (token) dengan fungsi *split* kemudian menghapus tanda baca, huruf, emoji dan selain angka.

Pengguna (user) akan memasukkan komentar kemudian klik tombol proses, maka sistem akan menampilkan hasil *tokenizing* dan untuk melanjutkan ke proses selanjutnya dengan klik tombol next seperti yang terlihat pada Gambar 3.



Gambar 3. Proses Tokenizing

Proses Case-folding

Proses *case-folding* merupakan lanjutan dari tahap *text pre-processing* dengan tujuan merubah teks dari huruf kapital menjadi huruf kecil seperti pada Gambar 4.



Gambar 4. Proses Casefolding

Proses Stopword Removal

Stopword Removal merupakan lanjutan dari tahap *text pre-processing* dengan tujuan menghilangkan kata-kata yang tidak penting dalam penggunaan bahasa Inggris. Adapun proses *stopword removal* ditunjukkan pada Gambar 5.



Gambar 5. Proses Stopword Removal

Proses Stemming



Gambar 6. Proses Stemming

Pada proses stemming merupakan proses terakhir dalam *text pre-processing* dimana proses stemming adalah mengubah kata ke dalam bentuk kata dasar (stem) seperti pada Gambar 6.

5. Feature Selection

Pada proses *feature selection*, hasil dari *pre-processing* yaitu teks komentar dalam bentuk kata, dimana sistem mengkonversi teks tersebut kedalam bentuk *numerical vector* (matriks) untuk mendapatkan fitur pada setiap kata dan dijadikan sebagai kriteria klasifikasi. Adapun contoh hasil ekstraksi fitur pada teks dapat dilihat pada Gambar 7.

```
<class 'scipy.sparse.csr.csr_matrix'>
(0, 3314) 1
(0, 3681) 1
aaaaawwww aaarrng aah aardvark aay ... zliester zombi zone zro zune
0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0
[1 rows x 3930 columns]
```

Gambar 7. Hasil Feature Extraction

Pada Gambar 7 merupakan hasil *feature selection*, dimana fungsi “*scipy.parse*” adalah mengubah dokumen teks menjadi matriks, kemudian untuk mendapatkan fitur pada setiap kata dengan menerapkan teknik *bag of word*, seperti dari kata “*aaaaawwww*” sampai dengan “*zune*”. Dimana nilai 0 bila kata tersebut tidak muncul pada dokument teks dan 1 bila muncul.

6. Proses Klasifikasi



Gambar 8. Proses Hasil Deteksi Komentar

Proses klasifikasi adalah menentukan apakah komentar yang dideteksi mengandung makna *bully* atau *nonbully* seperti yang terlihat pada Gambar 8. Dengan menggunakan algoritma *naive bayes* model *Multinomial Naive Bayes*. Jika nilai probabilitas *bully* (warna merah) menunjukkan lebih tinggi dibandingkan dengan nilai probabilitas *nonbully* (warna hijau), maka komentar tersebut dikategorikan komentar *bully* dan sebaliknya dikategorikan *nonbully*. Adapun

hasil dari pengujian pada 10 komentar dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengujian Terhadap 10 komentar

No	Komentar	Kategori	Terdeteksi
1.	Why I love joker more than Batman.	Nonbully	Nonbully
2.	This is literally her best song to date	Nonbully	Nonbully
3.	I immediately fell in love with this song when I first heard it	Nonbully	Nonbully
4.	2019! Best movie :Joker Like if u agree 😊	Nonbully	Nonbully
5.	I didn't know joker has sad life	Nonbully	Bully
6.	after watching this I realized that joker have a bad and sad memories	Bully	Bully
7.	What a stupid title for a video.	Bully	Bully
8.	That is the most STUPID question I've heard in a while! 🤔🤔🤔🤔🤔	Bully	Bully
9.	Nigeria is a corrupt and hateful place. Prejudice is based on ignorance and stupidity.	Bully	Bully
10.	Google Moy Park. Sky News Chicken Liars!	Bully	Nonbully

Berdasarkan hasil pengujian terhadap 10 komentar terdapat beberapa kesalahan dalam pengujian diantara pada komentar ke-5 dengan kategori *nonbully*, namun sistem mendeteksi komentar tersebut *bully* dan pada komentar ke-10 dengan kategori *bully*, sistem mendeteksi *nonbully*. Dengan begitu dapat dihitung nilai *accuracy*, *precision* dan *recall* sesuai data komentar yang di uji.

7. Pengujian *accuracy*, *precision*, *recall* dan *f1-score* metode *naive bayes classification*

Mengukur kinerja algoritma *Naive Bayes Classification* menggunakan *confusion matrix* dimana akan mengukur kinerja sistem secara keseluruhan meliputi nilai *accuracy*, *precision*, *recall* dan *f1-score*. Nilai *accuracy*, *precision*, *recall*, dan *f1-score* dapat dihitung dengan menggunakan persamaan berikut (Pardede, Jasman, & Husada, M.G., 2016):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\ Measure = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right)$$

dimana:

TP = True Positive

FP = False Positive

FN = False Negative

TN = True Negative

Dalam kasus penelitian ini dataset yang digunakan yaitu 2736 data komentar akan terbagi menjadi 2 bagian diantaranya data training dan data testing dengan rasio 80% digunakan sebagai data training dan 20% sebagai data testing. Adapun hasil pengujian kinerja sistem dengan *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perhitungan *Recall*, *Precision* dan *F1-score*

Class	Precision	recall	F1-score
<i>Bully</i>	0.83	0.75	0.79
<i>Non bully</i>	0.78	0.86	0.82
avg	0.81	0.80	0.80

Berdasarkan hasil pengujian kinerja pada Tabel 3. Nilai *precision* adalah berapa persen komentar yang benar *bully* dari keseluruhan komentar yang di prediksi *bully*, kemudian *recall* adalah berapa persen komentar yang di prediksi *bully* di bandingkan dengan keseluruhan komentar yang sebenarnya *bully* dan *f1-score* adalah perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

Kesimpulan

Penelitian ini telah berhasil membangun model untuk mendeteksi deteksi komentar *cyberbullying*

berdasarkan teks dan kategori menggunakan *Naïve Bayes Classification*. Hasil pengujian kinerja terhadap metode yang digunakan diperoleh akurasi sebesar 80% dengan nilai rata-rata menghasilkan *precision* 81%, *recall* 80% dan *f1-score* 80%.

Referensi

- Al-Anzi, F. S., & AbuZeina, D. (2017). Towards an Enhanced Arabic Text Classification Using Cosine Similarity and Latent Semantic Indexing. *Journal of King Saud University - Computer and Information Sciences*, 189-195.
- Ali, N. H., & Ibrahim, N. S. (2015). Porter Stemming Algorithm for Semantic Checking. *Computer Science Department*, 253-258.
- Al-khurayji, R., & Sameh, A. (2017). AN EFFECTIVE ARABIC TEXT CLASSIFICATION APPROACH BASED ON KERNEL NAIVE BAYES CLASSIFIER. *International Journal of Artificial Intelligence and Applications (IJAIA)*, Vol.8 1-10.
- Barlett, C. P., Gentile, D. A., Chng, G., Li, D., & Chamberlin, K. (2018). Social Media Use and Cyberbullying Perpetration: A Longitudinal Analysis. *Violence and Gender*, 1-7.
- Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 1-43.
- databoks.katadata.co.id*. (2019, November 12). Retrieved from <https://databoks.katadata.co.id/publish/2019/02/08/berapa-pengguna-media-sosial-indonesia#:~:targetText=Berdasarkan%20hasil%20riset%20wearesosial%20hootsuite,atau%20sekitar%2048%25%20dari%20populasi>.
- Etaiwi, W., & Naymat, G. (2017). The Impact of applying Different Preprocessing Steps on Review Spam Detection. *Procedia Computer Science*, 273-279.
- Feng, G., Guo, J., Jing, B.-Y., & Sun, T. (2015). Feature Subset Selection Using Naive Bayes for Text Classification. *Pattern Recognition Letters*, 1-8.
- Hasanah, U., Astuti, T., Wahyudi, R., Rifai, Z., & Pambudi, R. A. (2018). An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian. *IEEE 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE) - Yogyakarta, Indonesia* (pp. 230-23). Yogyakarta, Indonesia: Information System and Electrical Engineering (ICITISEE).
- Hase, C. N., Goldberg, S. B., Smith, D., Stuck, A., & Campaign, J. (2015). Impacts Of Traditional Bullying And Cyberbullying On The Mental Health Of Middle School And High School Students. *Psychology in the Schools*, 607-617.
- Kurniawan, B., Fauzi, M. A., & Widodo, A. W. (2017). Klasifikasi Berita Twitter Menggunakan Metode Improved Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1193-1200.
- Nasrullah, R. (2015). PERUNDUNGAN SIBER (CYBER-BULLYING) DI STATUS FACEBOOK. *Jurnal Sosioteknologi*, 1-11.
- Nasution, M. R., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *JURNAL INFORMATIKA*, 226-235.
- Natalia, E. C. (2016). REMAJA, MEDIA SOSIAL DAN CYBERBULLYING. *Jurnal Ilmiah Komunikasi*, 119-139.
- Nilan, P., Burgess, H., Hobbs, M., Threadgold, S., & Alexander, W. (2015). Youth, Social Media, and Cyberbullying Among Australian Youth: "Sick Friends". *Social Media + Society*, 1-12.
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, Volume (2) No.2 September 2018, pp. 200-209.
- Pardede, Jasman, & Husada, M.G. (2016). Comparison of VSM, GVSM, and LSI in Information Retrieval for

- Indonesian Text, *Journal Teknologi*, vol. 78 (5-6), 51-56.
- Pratama, B. Y., & Sarno, R. (2015). Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM . *Knowledge and Information Systems*, 170-174.
- Pujadayanti, I., Fauzi, M. A., & Sari, Y. A. (2018). Prediksi Rating Otomatis pada Ulasan Produk Kecantikan dengan Metode Naive Bayes dan N-gram. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4421-4427.
- Rasjid, Z. E., & Setiawan, R. (2017). Performance Comparison and Optimization of Text Document Classification using k-NN and Naive Bayes Classification Techniques. *Procedia Computer Science*, 107-112.
- Rifauddin, M. (2016). Fenomena Cyberbullying Pada Remaja (Studi Analisis Media Sosial Facebook). *Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan Khizanah Al-Hikmah*, 35-44.
- Santoso, J., Yuniarno, E. M., & Hariadi, M. (2015). Large Scale Text Classification Using Map Reduce and Naive Bayes Algorithm for Domain Specified Ontology Building. *IEEE 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)* (pp. 428-232). Hangzhou, China : 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics.
- Saritas, M. M., & Yasar, A. (2019). Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. *International Journal of Intelligent Systems and Applications in Engineering*, 88-91.
- Sastrodikoro W., Y. A., Palit, H. N., & Justinus Andjarwirawan, J. (2018). Aplikasi Pendeteksi Unsur Hinaan dalam Komentar di Media Sosial Berbahasa Indonesia. *Program Studi Teknik Informatika, Fakultas Teknik Industri, Universitas Kristen Petra*, vol 1.
- Susanto, W., & Indriyani, L. (2019). Analisis Penerapan Naive Bayes Untuk Memprediksi Resiko Kredit Anggota Koperasi Keluarga Guru . *JURNAL INFORMATIKA*, 262-270.
- Tian, L., Yan, Y., & Huebner, E. S. (2018). The Effects of Cyberbullying and Cybervictimization on Early Adolescents' Mental Health: The Differential Mediating Roles of Perceived Peer Relationship Stress. *Cyberpsychology, Behavior, and Social Networking*, 1-8.
- Wang, S., Jiang, L., & Li, C. (2015). Adapting naive Bayes tree for text classification. *Knowledge and Information Systems*, Vol 44; Iss. 1.
- Wijanto , M. C. (2015). Sistem Pendeteksi Pengirim Tweet dengan Metode Klasifikasi Naive Bayes . *Jurnal Teknik Informatika dan Sistem Informasi* , 172-182.

**JURNAL INFORMATIKA**

LPPM BSI BANDUNG

★ **P-ISSN : 23556579** < > **E-ISSN : 25282247** 📌 **Subject Area : Science, Engineering**

 **1.5625**
Impact Factor

 **3635**
Google Citations

 **Sinta 4**
Current
Accreditation

 [Google Scholar](#)  [Garuda](#)  [Website](#)  [Editor URL](#)

History Accreditation

2018 2019 2020 2021 2022 2023

Garuda [Google Scholar](#)

[Tata Kelola Teknologi Informasi Pada JNE Cabang Bekasi Menggunakan Framework Cobit 4.1](#)

LPPM Universitas Bina Sarana Informatika  [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
55-60

📅 2022 📄 DOI: [10.31294/inf.v9i1.11708](#) 🏆 [Accred : Sinta 4](#)

[Implementasi Metode Hybrid AHP-SAW-TOPSIS Untuk Pemilihan Taman TOGA](#)

LPPM Universitas Bina Sarana Informatika  [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
18-25

📅 2022 📄 DOI: [10.31294/inf.v9i1.11901](#) 🏆 [Accred : Sinta 4](#)

[Perbandingan Fungsi Optimasi Neural Network Dalam Klasifikasi Kelayakan Calon Suami](#)

LPPM Universitas Bina Sarana Informatika  [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
78-84

📅 2022 📄 DOI: [10.31294/inf.v9i1.12318](#) 🏆 [Accred : Sinta 4](#)

[Impelementasi Deep Learning untuk Optimasi Slump Menggunakan Convolutional Neural Network Pada PT. Handaru Wijaya Mulya](#)

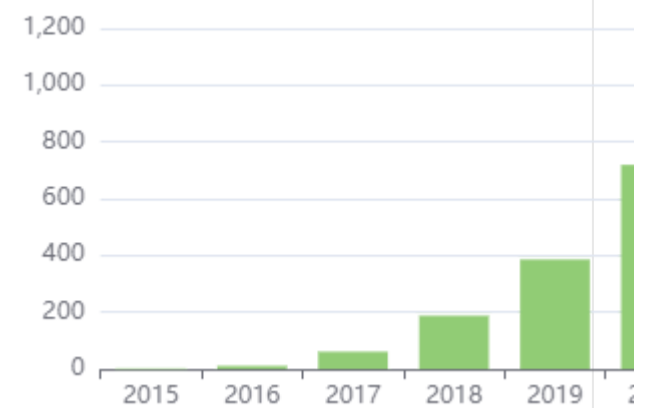
LPPM Universitas Bina Sarana Informatika  [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
26-31

📅 2022 📄 DOI: [10.31294/inf.v9i1.11713](#) 🏆 [Accred : Sinta 4](#)

[Sistem Penunjang Keputusan Penilaian Kinerja Dosen Dengan Metode Profile Matching](#)

LPPM Universitas Bina Sarana Informatika  [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
85-92

Citation Per Year By Google Scholar



Journal By Google Scholar

	All	Since 2018
Citation	3635	3540
h-index	31	31
i10-index	100	98

2022 [DOI: 10.31294/inf.v9i1.11192](https://doi.org/10.31294/inf.v9i1.11192) [Accred : Sinta 4](#)

[Analisis Latensi Pada Aplikasi Virtual Reality Untuk Anak Dengan Autism Syndrome Disorder](#)

LPPM Universitas Bina Sarana Informatika [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
32-42

2022 [DOI: 10.31294/inf.v9i1.11778](https://doi.org/10.31294/inf.v9i1.11778) [Accred : Sinta 4](#)

[Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly](#)

LPPM Universitas Bina Sarana Informatika [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
1-5

2022 [DOI: 10.31294/inf.v9i1.10209](https://doi.org/10.31294/inf.v9i1.10209) [Accred : Sinta 4](#)

[Analisa Hasil Perbandingan Poly Kernel Dan Normalisasi Poly Kernel Pada Support Vector Machine Sebagai Metode Klasifikasi Citra Tanda Tangan](#)

LPPM Universitas Bina Sarana Informatika [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
71-77

2022 [DOI: 10.31294/inf.v9i1.11288](https://doi.org/10.31294/inf.v9i1.11288) [Accred : Sinta 4](#)

[Analisis Penyebaran Pandemi Covid-19 di Kota Jakarta Menggunakan Metode Clustering K-Means dan Density Based Spatial Clustering of Application With Noise](#)

LPPM Universitas Bina Sarana Informatika [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
43-54

2022 [DOI: 10.31294/inf.v9i1.11226](https://doi.org/10.31294/inf.v9i1.11226) [Accred : Sinta 4](#)

[Peningkatan Kualitas Layanan Warga Kelurahan Duri Kepa dengan Aplikasi LINGKOE](#)

LPPM Universitas Bina Sarana Informatika [Jurnal Informatika Vol 9, No 1 \(2022\): April 2022](#)
61-70

2022 [DOI: 10.31294/inf.v9i1.11538](https://doi.org/10.31294/inf.v9i1.11538) [Accred : Sinta 4](#)

[View more ...](#)

Get More with
SINTA Insight

[Go to Insight](#)

Citation Per Year By Google Scholar

Journal By Google Scholar

	All	Since 2018
Citation	3635	3540
h-index	31	31
i10-index	100	98