



[Home](#) / [Editorial Team](#)

## Editorial Team

### Editor in Chief

Dian Ade Kurnia (Scopus ID: [57205059723](#), STMIK IKMI, Cirebon, Indonesia)

### Editor Board Members

1. Odi Nurdiawan (Scopus ID : [57212352103](#), STMIK IKMI, Cirebon, Indonesia)
2. Oman Somantri (Scopus ID : [57208898676](#), Politeknik Negeri Cilacap, Cilacap, Indonesia)

### Technical Support

Yudhistira Arie Wijaya (Scopus ID : [57287277000](#), STMIK IKMI, Cirebon, Indonesia)



### Reviewer

1. Raditya Dinar Dana (Scopus ID : [57211230995](#), STMIK IKMI, Cirebon, Indonesia)
2. Tedjo Darmanto (Scopus ID : [55625535100](#), STMIK AMIKBANDUNG, Bandung, Indonesia)
3. Randi Rizal (Scopus ID : [57207962586](#), Universitas Siliwangi, Tasikmalaya, Indonesia)
4. Ahmad Naim Cee Pee (Scopus ID : [56300006200](#), University Teknikal Malaysia, Malaka)
5. Zuraida Abas (Scopus ID : [36871592400](#), University Teknikal Malaysia, Malaka)
6. Asep ID Hadiana (Scopus ID : [57219518189](#), Universitas Jenderal Ahmad Yani, Indonesia)
7. Robi Rahim (Scopus ID: [57202895920](#), STIM Sukma, Medan, Indonesia)
8. Zuraida Abal Abas (Scopus ID : [36871592400](#), University Teknikal Malaysia, Melaka)
9. Ahmad Naim Cee Pee (Scopus ID : [56300006200](#), University Teknikal Malaysia, Melaka)

**Daftar Isi**  
**Jurnal Information Communication & Technology (JICT)**  
**Vol. 23 No. 2 (2023): JICT-IKMI, December 2023**

Articles

- 
- Penerapan Sistem Keamanan Website Menggunakan Web Application Firewall Dengan Framework Open Web Application Security Project**  
Muhammad Dandi Permana Dandi, Syahril Rizal, Febriyanti Panjaitan, Suryayusra 331-338  
[PDF](#)
- Aplikasi Sistem Informasi Pengolahan Data RT Berbasis Web Menggunakan Django Python**  
Herlina, Asep Suherman, Givy Devira Ramady, Dendy Adya Pratama 339-346  
[PDF](#)
- Model Convolutional Neural Network : Sistem Cerdas untuk Mendeteksi Penyakit pada Tanaman Sayur**  
Anwar Sadad, Johni Pasaribu 347-354  
[PDF](#)
- Uji Responsif Multi-Perangkat pada Proyek Sistem Informasi Mahasiswa (SIM-SIS) Politeknik Piksi Ganesha**  
Uji Responsif Multi-Perangkat  
Osama Osama, Johni S Pasaribu 355-360  
[PDF](#)
- Mengukur Tingkat Kepuasan Pengguna Website Akademik Universitas Muhammadiyah Prof. Dr. HAMKA Menggunakan System Usability Scale**  
Muhammad Al Aziiz, Ade Davy Wiranata 361-365  
[PDF](#)
- 
- Rancang Bangun Aplikasi Pendaftaran Dropshipper Dan Order Produk Berbasis Web**  
Mami Maryati, Marisa, Vivi Rhamdiyani, Widiyawati, HS. Sulistyowati 372-378  
[PDF](#)
- Klasifikasi Jenis Ikan Koi Menggunakan Ekstraksi Warna HSV dan Metode Adaptive Neuro-Fuzzy Inference System (ANFIS)**  
Deanna Durbin Hutagalung, Galan Min Aqshal 379-384  
[PDF](#)
- Analisis Data INACBG's Menggunakan Algoritma FP-Growth Dengan Dukungan Visualisasi Data**  
Eka Wahyu Pujiharto, Kusri, Asro Nasiri 385-393  
[PDF](#)
- Implementasi Agile Development Dengan Scrum Dalam Perancangan Sistem Informasi Berbasis Website**  
Elan Pratama, Nur'aini 394-399  
[PDF](#)
- Penerapan Metode Profile Matching Dalam Penilaian Kelayakan Mutasi Pegawai Pada Pdam Tirta Bhagasasi Bekasi**  
Ishak Husin 400-404  
[PDF](#)

<p><b>RANCANG BANGUN SISTEM INFORMASI PENGHITUNGAN GAJI DENGAN METODE GROSS PADA CV. BERKAH RAHMA</b></p> <p>Sabar ., Erwiyanti; Dewi Ayu Puspitawati; Kikim Mukiman, Santoso Utomo</p> <p><a href="#">PDF</a></p>	405-410
<p><b>Pengembangan Sistem Informasi Manajemen Dosen Berbasis Web dengan Framework Laravel 8</b></p> <p>Hendra Wijaya Subandi, Rini Suwartika Kusumadiarti</p> <p><a href="#">PDF</a></p>	411-420
<p><b>Perancangan Backend Sistem Informasi Mahasiswa Berbasis Website Di Piksi Ganesha Bandung</b></p> <p>M Almas Fadhilah Thamrin, Johni S Pasaribu</p> <p><a href="#">PDF</a></p>	421-426
<p><b>Penerapan Fuzzy Inference System Sugeno untuk Pemilihan Capstone Design</b></p> <p>Dina Heriyani, Eka Yulia Sari, Emmy Nurhayati</p> <p><a href="#">PDF</a></p>	427-432
<p><b>Penerapan Artificial Intelligence (AI) Dalam Pengelolaan Krisis Bencana</b></p> <p>Tri Wahyudi</p> <p><a href="#">PDF</a></p>	433-439
<p><b>Analisis Sentimen Kasus Ferdy Sambo pada Twitter Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur</b></p> <p>Arief Priyadi Pratama, Prima Dina Atika, Siti Setiawati</p> <p><a href="#">PDF</a></p>	440-447
<p><b>Perancangan Chatbot Berbasis Whatsapp Service Order Delivery Menggunakan Metode Dialog Flow</b></p> <p>Eri Eri Riana, Meiva Eka Sri Sulistyawati, Encep Fahmi Imaduddin, Adianta Sebayang</p> <p><a href="#">PDF</a></p>	448-456
<p><b>Perancangan Sistem Pencarian Jurnal Ilmiah Nasional Dengan Filter Berbasis Kategori</b></p> <p>Muhamad Syarif</p> <p><a href="#">PDF</a></p>	457-462
<p><b>Analisis Sentimen Ulasan User Terhadap Aplikasi Octopus Menggunakan Algoritma Support Vector Machine</b></p> <p>Reza Kurnia Khoirunisa, Erizal</p> <p><a href="#">PDF</a></p>	463-468
<p><b>Pengembangan Sistem Informasi Point Of Sales Pada Distro Parahyangan Bandung Berbasis Website Dengan FrameWork Laravel</b></p> <p>Raden Ridho, Ruslandi Haris Susanto, Johni Pasaribu</p> <p><a href="#">PDF</a></p>	469-474
<p><b>Sistem Pendukung Keputusan Pemilihan Jenis Masker Wajah Untuk Semua Jenis Kulit Menggunakan Metode AHP</b></p> <p>Sarah Damayanti Anugrah, Satia Suhada, Rifa Nurafifah Syabaniah, Saghifa Fitriana</p> <p><a href="#">PDF</a></p>	478-484
<p><b>Metode Agile Scrum Untuk Perancangan Sistem Informasi Manajemen Proyek Berbasis Web</b></p> <p>Daffa Erdian Fahrezy, Rudi Kurniawan</p> <p><a href="#">PDF</a></p>	485-490
<p><b>Perancangan Aplikasi Pengaduan Masyarakat Berbasis Website Di Desa Cicangkang Hilir</b></p> <p>Isni Anisah Puspowati Harori, M. Nurul Fauzi Do'a, Daffa Muhamad Iqbal</p> <p><a href="#">PDF</a></p>	491-496

# Analisis Sentimen Kasus Ferdy Sambo pada *Twitter* Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur

Arief Priyadi Pratama<sup>1\*</sup>, Prima Dina Atika<sup>2\*</sup>, Siti Setiawati<sup>3\*</sup>

<sup>1,2,3</sup>Program Studi Informatika, Universitas Bhayangkara Jakarta Raya, Indonesia

Email: <sup>1</sup>rief.priyadip@gmail.com, <sup>2</sup>prima.dina@dsn.ubharajaya.ac.id, <sup>3</sup>siti.setiawati@dsn.ubharajaya.ac.id

## INFORMASI ARTIKEL

### Histori artikel:

Naskah masuk, 19 Juli 2023

Direvisi, 31 Desember 2023

Diiterima, 31 Desember 2023

### Kata Kunci:

Analisis Sentimen,  
Kasus Viral,  
Ferdy Sambo,  
CRISP-DM,  
Naive Bayes,  
Chi Square,  
Information Gain

## ABSTRAK

**Abstract-** The Indonesian National Police (Polri) is a government institution with integrity as a protector of public security. However, there has been a case of premeditated murder by the Head of Propam Inspector General Ferdy Sambo which has gone viral case. This case attracted high public enthusiasm so that various opinions emerged on Twitter. The complexity in classifying opinions is a challenge because there is no standard for classifying opinions. To overcome these challenges, we need a sentiment analysis technique that can classify opinions accurately. Therefore, this study aims to analyze public opinion on Twitter using the Naive Bayes algorithm combined with the Chi Square feature selection method and Information Gain using the CRISP-DM methodology. The data used has been labeled by linguists. The results of the research with the 80:20 data split show that the Naive Bayes model achieves an accuracy of 73%. The Chi Square combination makes accuracy increase to 75%. In addition, Information Gain also produces an accuracy of 74%. The experiment was carried out with five different data split ratios. It can be concluded that the Naive Bayes model with Chi Square feature selection and a data split ratio of 80:20 has an accuracy increase of 2% as the best accuracy result, which is equal to 75%.

**Abstrak-** Kepolisian Negara Republik Indonesia (Polri) merupakan salah satu lembaga pemerintah yang berintegritas sebagai pengayom keamanan masyarakat. Namun, telah terjadi sebuah kasus pembunuhan berencana oleh Kadiv Propam Irjen Ferdy Sambo menjadi kasus viral. Kasus ini menarik antusiasme masyarakat yang tinggi sehingga mencuatkan beragam opini di Twitter. Kompleksitas dalam mengklasifikasikan opini menjadi sebuah tantangan karena belum adanya standar pengklasifikasian opini. Untuk mengatasi tantangan tersebut, diperlukan sebuah teknik analisis sentimen yang dapat mengklasifikasikan opini dengan akurat. Oleh karena itu, penelitian ini bertujuan untuk menganalisis opini masyarakat di *Twitter* dengan menggunakan algoritma Naive Bayes yang dipadukan dengan metode seleksi fitur *Chi Square* dan *Information Gain* menggunakan metodologi CRISP-DM. Data yang digunakan telah diberi label oleh ahli bahasa. Hasil penelitian dengan split data 80:20 menunjukkan bahwa model *Naive Bayes* mencapai akurasi sebesar 73%. Kombinasi *Chi Square* menjadikan akurasi meningkat menjadi 75%. Selain itu, *Information Gain* juga menghasilkan akurasi sebesar 74%. Percobaan dilakukan dengan lima rasio *split* data yang berbeda. Dapat disimpulkan bahwa model *Naive Bayes* dengan seleksi fitur *Chi Square* dan rasio *split* data 80:20 mengalami peningkatan akurasi sebesar 2% sebagai hasil akurasi terbaik, yaitu sebesar 75%.

Copyright © 2023 LPPM - STMIK IKMI Cirebon  
This is an open access article under the CC-BY license

## Penulis Korespondensi:

**Prima Dina Atika**

Prodi Informatika

Universitas Bhayangkara Jakarta Raya

Jl. Raya Perjuangan Marga Mulya, Kec. Bekasi Utara, Kota Bks, Jawa Barat 17143

prima.dina@dsn.ubharajaya.ac.id

## 1. Pendahuluan

Indonesia merupakan negara republik yang berlandaskan hukum perundang-undangan yang

telah diatur sebagai sistem yang fungsional dan berkeadilan. Kepolisian Negara Republik Indonesia (Polri) merupakan salah satu institusi negara yang



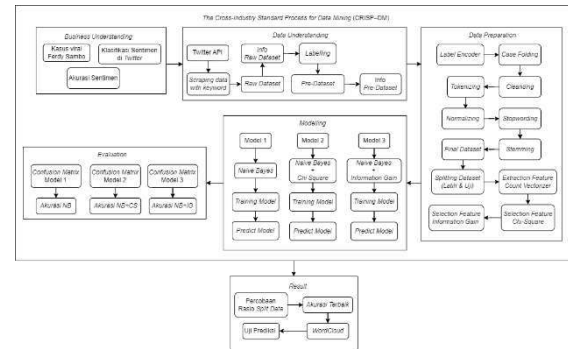
memegang peran penting dalam menegakkan hukum yang berlaku di masyarakat. Polri memiliki integritas untuk melindungi, mengayomi dan melayani masyarakat bertujuan menciptakan rasa aman bagi masyarakat. Kewenangan diatur dalam Undang-Undang Nomor 2 Tahun 2022 tentang Kepolisian Negara Republik Indonesia Pasal 5 ayat 1[1]. Namun, telah terjadi kasus viral yang membuka aib Polri, yaitu pembunuhan berencana oleh Irjen Pol. Ferdy Sambo, S.H., S.I.K., M.H. yang juga memegang jabatan Kepala Divisi Profesi dan Pengamanan (Kadiv Propam). Tingkat kepercayaan masyarakat terhadap institusi Kepolisian RI turun ke level 54,4 pada Agustus 2022 dibanding April 2022 yang berada di level 71,6 setelah terjadi kasus pembunuhan Brigadir Joshua oleh Irjen Ferdy Sambo dibandingkan dengan Komisi Pemberantasan Korupsi (KPK) dan Kejaksaan[2].

Merujuk pada penelitian sebelumnya yang menganalisis sentimen kasus viral, ditemukan penelitian yang membahas *Naive Bayes* sebagai algoritma pengklasifikasian. Berdasarkan penelitian [3] menggunakan *Naive Bayes* mengenai kelangkaan stok minyak goreng mencapai hasil akurasi 69,4%. Sedangkan untuk pengklasifikasian kombinasi dengan *Chi Square* berdasarkan penelitian [4] mengenai migrasi TV digital mengalami peningkatan akurasi menghasilkan akurasi 96%.

Kasus viral yang membuka aib di salah satu institusi pemerintah yaitu Polri, telah menarik antusiasme masyarakat terhadap kasus ini sehingga mencuatkan beragam opini di media sosial. Oleh karena itu, dibutuhkan sebuah teknik pengklasifikasian opini, maka penelitian ini bertujuan untuk menganalisis opini masyarakat di *Twitter* dengan teknik analisis sentimen menggunakan algoritma *Naive Bayes* yang dipadukan dengan metode seleksi fitur *Chi Square* dan *Information Gain* menggunakan metodologi CRISP-DM.

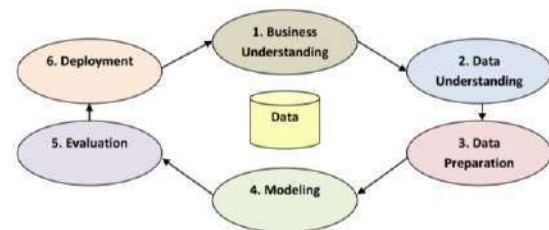
## 2. Metode Penelitian

Penelitian ini menerapkan metodologi CRISP-DM untuk melakukan klasifikasi opini masyarakat mengenai kasus Ferdy Sambo menggunakan metode *Naive Bayes* dengan seleksi fitur *Chi Square* dan *Information Gain* sehingga dapat menganalisis hasil evaluasi akurasi ketiga model. Pemrosesan data dilakukan menggunakan pemrograman *Python* dengan tools *Google Colaboratory*.



Gambar 1. Alur Penelitian

## 2.1 CRISP-DM



Gambar 2. CRISP-DM Phase

Metodologi *The Cross-Industry Standard Process for Data Mining* (CRISP-DM) menjadi penyedia proses standar *nonproprietary* dan tersedia secara bebas untuk menyesuaikan penambahan data ke dalam strategi pemecahan masalah umum unit bisnis atau penelitian dikarenakan keunggulannya dalam menyelesaikan banyak persoalan dalam proyek *data mining*. Terdiri dari enam fase siklus yang bersifat adaptif. Artinya, fase berikutnya dalam urutan seringkali bergantung pada hasil yang terkait dengan fase sebelumnya[5]. Penjelasan fase CRISP-DM berdasarkan 4th CRISP-DM *Special Interest Group (SIG) Workshop* di Brussels, Belgium pada bulan Maret tahun 1999[6], yaitu:

- business understanding*, merupakan fase awal pemahaman bisnis dengan menjelaskan tujuan, persyaratan dan batasan secara keseluruhan kemudian menjadi definisi perumusan masalah *data mining* dan persiapan strategi awal untuk mencapai tujuan tersebut;
- data understanding*, merupakan fase pengumpulan data awal kemudian pengenalan data, mengidentifikasi masalah kualitas data, menemukan wawasan pertama ke dalam data atau mendeteksi himpunan bagian yang menarik untuk membentuk hipotesis untuk informasi tersembunyi;
- data preparation*, merupakan fase cakupan semua aktivitas untuk menyusun kumpulan data akhir (data yang akan dimasukkan ke dalam alat pemodelan/*final dataset*) dari data mentah awal (*Raw dataset*);
- modelling*, merupakan fase berbagai teknik pemodelan dipilih dan diterapkan dan

- pengoptimalisasian parameter. Beberapa teknik memiliki persyaratan khusus pada bentuk data. Oleh karena itu, langkah mundur ke tahap persiapan data seringkali diperlukan;
- e. *evaluation*, merupakan fase mengevaluasi model secara lebih menyeluruh dan meninjau langkah-langkah yang dilakukan untuk memastikan model mencapai tujuan bisnis dengan benar berdasarkan model kualitas tinggi dari perspektif analisis data. Pada akhir fase ini, keputusan tentang penggunaan hasil penambangan data harus dicapai; dan
  - f. *deployment*, merupakan fase optimalisasi dari kelima fase sebelumnya. Namun, fase ini umumnya bukanlah akhir dari proyek, bergantung pada persyaratan yang terdapat pada fase pertama. Dalam penelitian ini, fase *deployment* tidak diperlukan karena untuk analisis sentimen mendapatkan hasil akurasi dari optimalisasi peningkatan akurasi model *Naive Bayes* dengan kombinasi seleksi fitur *Chi Square* dan *Information Gain*.

## 2.2 Algoritma Naive Bayes

Sekumpulan probabilitas yang mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas karena implementasi hasil klasifikasi yang dilakukan menunjukkan hasil yang beragam[7]. Model Naive Bayes terdapat tautan dari kelas target ke setiap variabel masukan dan tidak ada tautan lain. Diasumsikan bahwa variabel input bersifat independen (tidak ada interaksi antar variabel) satu sama lain sesuai target[8]. Formula teorema bayes yang menjadi aturan dasar *Naive Bayes Classifier*[9], yaitu:

$$P(X | Y) = \frac{P(Y | X).P(X)}{P(Y)} \quad (1)$$

Keterangan:

Y = data dengan *class* yang belum diketahui;

X = hipotesis data Y merupakan suatu *class* spesifik;  
P(X | Y) = probabilitas hipotesis X berdasarkan kondisi Y;

P(Y | X) = probabilitas Y berdasarkan kondisi pada hipotesis X;

P(X) = probabilitas hipotesis X; dan

P(Y) = probabilitas Y.

## 2.3 Chi Square

Metode yang melihat ketergantungan *term* dengan kategorinya yang tidak memerlukan adanya asumsi-asumsi mengenai sebaran data populasinya atau statistik bebas sebaran (belum diketahui sebaran data berparameter populasi dan tidak perlu berdistribusi normal)[10]. *Chi Square* untuk seleksi fitur sebagai bentuk cara mengurangi kebutuhan penyimpanan (*storage*), mengurangi *training*, membantah kutukan dimensi (*curse of*

*dimensionality*), dan meningkatkan performansi klasifikasi[11]. Fitur yang tidak relevan atau berlebihan biasanya akan merusak keakuratan model klasifikasi. Oleh karena itu, tujuan dari algoritma seleksi fitur adalah untuk memilih fitur yang paling informatif dan relevan sehubungan dengan label kelas[12]. Formula uji seleksi fitur *Chi Square* pengujian model distribusi berdasarkan data hasil pengamatan[13], yaitu:

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

Keterangan:

$O_i$  = *Observed value* atau nilai (banyaknya kasus) observasi/sebenarnya dalam kategori *i*;

$E_i$  = *Expected Value* atau nilai (banyaknya kasus) ekspektasi/harapan;

$\sum_{i=1}^n$  = penjumlahan semua kategori *n*.

## 2.4 Information Gain

Metode yang ditingkatkan untuk mengatasi dua kekurangan yaitu membatasi frekuensi kata frekuensi dokumen (LDFWF) dan informasi distribusi (DI). Digunakan untuk memilih fitur-fitur penting sehubungan dengan atribut kelas sehingga menunjukkan hasil peningkatan akurasi klasifikasi[14]. *Information Gain* merupakan proses identifikasi dan memilih fitur yang paling relevan dari data yang berisi, sehingga memangkas fitur yang tidak relevan dan mengurangi dimensi data (jargon tekstual). Formula *Information Gain*[15], yaitu:

$$G(D,t) = - \sum_{i=1}^m P(C_i) \log P(C_i) + P(t) \sum_{i=1}^m P(C_i|t) \log P(C_i|t) + P(\bar{t}) \sum_{i=1}^m P(C_i|\bar{t}) \log P(C_i|\bar{t}) \quad (3)$$

Keterangan:

C = koleksi dokumen;

$P(C_i)$  = probabilitas dari kategori ke-*i*;

$P(t)$  = probabilitas istilah *t* muncul dalam dokumen;

$P(\bar{t})$  = probabilitas istilah *t* tidak muncul dalam dokumen;

$P(C_i|t)$  = probabilitas bersyarat dari nilai kelas ke-*i* yang diberikan istilah *t* muncul dalam dokumen; dan

$P(C_i|\bar{t})$  = probabilitas bersyarat dari nilai kelas ke-*i* yang diberikan istilah *t* tidak muncul dalam dokumen.

## 2.5 Analisis Sentimen

Menganalisa opini, sentimen, evaluasi, sikap, penilaian dan emosi seseorang yang berhubungan dengan suatu topik yang bertujuan menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif/negatif/netral[16]. Pengumpulan data suatu topik menjadi *dataset* yang siap diolah lebih lanjut untuk mengidentifikasi, mengekstrak, dan memahami sentimen dalam teks yang bersumber dari media sosial[17].

### 2.6 Twitter

Twitter merupakan sebuah media sosial dan layanan *microblogging* yang diciptakan oleh Jack Dorsey di tahun 2006 yang bermanfaat untuk ekspresi diri (*self-expression*), pencitraan diri (*personal branding*) dan sebagai ruang diskusi dalam menyampaikan pendapat mengenai topik atau isu tertentu. Serta kemampuannya dalam membangun opini publik. *Tweet* adalah sebuah pesan pendek dengan panjang karakter yang dibatasi hanya sampai 280 karakter[18]. Fungsi *Twitter* untuk terhubung, berkomunikasi dan berbagi satu sama lain[19].

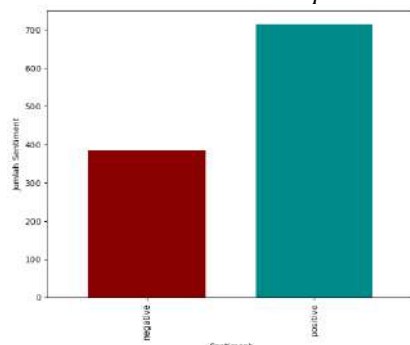
## 3. Hasil dan Pembahasan

### 3.1 Business Understanding

Permasalahan yang berfokus pada identifikasi tujuan penelitian yaitu melakukan penelitian eksperimen tentang kasus viral pembunuhan berencana oleh Ferdy Sambo dengan menganalisis opini masyarakat yang beragam di salah satu media sosial yaitu *Twitter* dengan mengklasifikasikan opini masyarakat dengan analisis sentimen. Pengklasifikasian analisis sentimen menggunakan Algoritma *Naive Bayes* dengan fitur seleksi *Chi Square* dan *Information Gain* untuk menghasilkan akurasi klasifikasi sentimen yang optimal pada kasus tersebut.

### 3.2 Data Understanding

Melakukan scraping data menggunakan API *Twitter* dengan kata kunci “*vonis Ferdy Sambo*” dengan periode 01 Februari 2023 sampai 05 Juni 2023 mendapatkan 2515 data. Kemudian menyaring data dengan menghapus data *tweets* iklan dan berita sehingga jumlah *Raw dataset* sebesar 1100 data. Setelah itu melakukan pelabelan data manual oleh pakar bahasa dengan tujuan untuk menjaga keobjektifan data penelitian. Proses pelabelan dilakukan dengan membagi data *tweet* menjadi dua kelas yaitu sentimen positif dan sentimen negatif. Total sentimen positif sebanyak 715 data dan sentimen negatif sebanyak 385 data. *Raw dataset* yang sudah diberi label dinamakan *pre-dataset*.



Gambar 3. Visualisasi Distribusi Data Sentimen

### 3.3 Data Preparation

Pemrosesan *pre-dataset* dalam *preparation phase* yang menggunakan *pre-processing* analisis sentimen untuk mendapatkan *final dataset*. Proses awal dimulai dengan mengubah tipe data kategorik menjadi tipe data numerik.

Tabel 1. Proses *Label Encoder*

sebelum		sesudah	
Review	Senti ment	Review	Tar get
Terungkapnya kasus Ferdy Sambo menjadi preseden paling buruk di internal kepolisian. Seorang jenderal Bintang 2 Kadiv Propam, polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji. Atas desakan publik yg kuat, bersyukur kasus itu selesai dgn VONIS MATI <a href="https://t.co/wn71YVrUgx">https://t.co/wn71YVrUgx</a>	Positi ve	Terungkapnya kasus Ferdy Sambo menjadi preseden paling buruk di internal kepolisian. Seorang jenderal Bintang 2 Kadiv Propam, polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji. Atas desakan publik yg kuat, bersyukur kasus itu selesai dgn VONIS MATI <a href="https://t.co/wn71YVrUgx">https://t.co/wn71YVrUgx</a>	1

Selanjutnya proses *case folding*, perubahan menjadi huruf kecil.

Tabel 2. Proses *Case Folding*

Sebelum	sesudah
Terungkapnya kasus Ferdy Sambo menjadi preseden paling buruk di internal kepolisian. seorang jenderal bintang 2 kadiv propam, polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji. atas desakan publik yg kuat, bersyukur kasus itu selesai dgn VONIS MATI <a href="https://t.co/wn71YVrUgx">https://t.co/wn71YVrUgx</a>	terungkapnya kasus ferdy sambo menjadi preseden paling buruk di internal kepolisian. seorang jenderal bintang 2 kadiv propam, polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji. atas desakan publik yg kuat, bersyukur kasus itu selesai dgn vonis mati <a href="https://t.co/wn71yvru gx">https://t.co/wn71yvru gx</a>

Selanjutnya proses *cleansing*, penghilangan karakter URL, *username*, *hashtag*, *email*, spasi berlebih, tanda baca dan angka.

Tabel 3. Proses *Cleansing*

Sebelum	sesudah
terungkapnya kasus ferdy sambo menjadi	terungkapnya kasus ferdy sambo menjadi preseden

preseden paling buruk di internal kepolisian. seorang jenderal bintang 2 kadiv propam, polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji. atas desakan publik yg kuat, bersyukur kasus itu selesai dgn vonis mati <a href="https://t.co/wn7lyvrugx">https://t.co/wn7lyvrugx</a>	paling buruk di internal kepolisian seorang jenderal bintang kadiv propam polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji atas desakan publik yg kuat bersyukur kasus itu selesai dgn vonis mati
--	--

selanjutnya proses *tokenizing*, pemisahan teks menjadi kata.

Tabel 4. Proses *Tokenizing*

Sebelum	sesudah
terungkapnya kasus ferdy sambo menjadi preseden paling buruk di internal kepolisian seorang jenderal bintang kadiv propam polisinya polisi justru jd otak dibalik pembunuhan yang begitu keji atas desakan publik yg kuat bersyukur kasus itu selesai dgn vonis mati	['terungkapnya', 'kasus', 'ferdy', 'sambo', 'menjadi', 'preseden', 'paling', 'buruk', 'di', 'internal', 'kepolisian', 'seorang', 'jenderal', 'bintang', 'kadiv', 'propam', 'polisinya', 'polisi', 'justru', 'jd', 'otak', 'dibalik', 'pembunuhan', 'yang', 'begitu', 'keji', 'atas', 'desakan', 'publik', 'yg', 'kuat', 'bersyukur', 'kasus', 'itu', 'selesai', 'dgn', 'vonis', 'mati']

selanjutnya proses *normalizing*, pengubahan kedalam bahasa baku sehingga penghapusan kata gaul/slang.

Tabel 5. Proses *Normalizing*

Sebelum	sesudah
['terungkapnya', 'kasus', 'ferdy', 'sambo', 'menjadi', 'preseden', 'paling', 'buruk', 'di', 'internal', 'kepolisian', 'seorang', 'jenderal', 'bintang', 'kadiv', 'propam', 'polisinya', 'polisi', 'justru', 'jd', 'otak', 'dibalik', 'pembunuhan', 'yang', 'begitu', 'keji', 'atas', 'desakan', 'publik', 'yg', 'kuat', 'bersyukur', 'kasus', 'itu', 'selesai', 'dgn', 'vonis', 'mati']	['terungkapnya', 'kasus', 'ferdy', 'sambo', 'menjadi', 'preseden', 'paling', 'buruk', 'di', 'internal', 'kepolisian', 'seorang', 'jenderal', 'bintang', 'kadiv', 'propam', 'polisinya', 'polisi', 'justru', 'jadi', 'otak', 'dibalik', 'pembunuhan', 'yang', 'begitu', 'keji', 'atas', 'desakan', 'publik', 'yang', 'kuat', 'bersyukur', 'kasus', 'itu', 'selesai', 'dengan', 'vonis', 'mati']

selanjutnya proses *stopwording*, membuang kata yang kurang relevan.

Tabel 6. Proses *Stopwording*

Sebelum	sesudah
['terungkapnya', 'kasus', 'ferdy', 'sambo', 'menjadi', 'preseden', 'paling', 'buruk', 'di', 'internal', 'kepolisian', 'seorang', 'jenderal', 'bintang', 'kadiv']	terungkapnya kasus ferdy sambo menjadi preseden paling buruk internal kepolisian seorang jenderal bintang kadiv

Sebelum	sesudah
'seorang', 'jenderal', 'bintang', 'kadiv', 'propam', 'polisinya', 'polisi', 'justru', 'jadi', 'otak', 'dibalik', 'pembunuhan', 'yang', 'begitu', 'keji', 'atas', 'desakan', 'publik', 'yang', 'kuat', 'bersyukur', 'kasus', 'itu', 'selesai', 'dengan', 'vonis', 'mati']	propam polisinya polisi justru jadi otak dibalik pembunuhan keji atas desakan publik kuat bersyukur kasus selesai vonis mati

selanjutnya proses *stemming*, pengubahan kata ke dalam satu bentuk dasar baku. pemecahan setiap imbuhan dari suatu kata seperti awalan (prefiks), sisipan (infiks), akhiran (sufiks), awalan akhiran (konfiks).

Tabel 7. Proses *Stemming*

Sebelum	sesudah
terungkapnya kasus ferdy sambo menjadi preseden paling buruk internal kepolisian seorang jenderal bintang kadiv propam polisinya polisi justru jadi otak dibalik pembunuhan keji atas desakan publik kuat bersyukur kasus selesai vonis mati	ungkap kasus ferdy sambo jadi preseden paling buruk internal polisi orang jenderal bintang kadiv propam polisi polisi justru jadi otak balik pembunuhan keji atas desa publik kuat syukur kasus selesai vonis mati

Setelah proses *pre-processing*, memperoleh *final dataset*. Lalu dilakukan percobaan berbagai *split data* untuk mendapatkan hasil akurasi terbaik dari ketiga model. *Final dataset* berisi 1100 data akan dilakukan uji coba dengan lima rasio *split data*, yaitu 90:10, 80:20, 70:30, 60:40 dan 50:50. Penerapan pada Python dengan mengubah parameter "*test\_size*". Pembagian *dataset* yang baik sangat penting untuk menghindari bias atau kesalahan evaluasi yang tidak akurat pada model atau algoritma yang dilatih.

Tabel 8. Distribusi Data Percobaan Rasio *Split Data*

Rasio <i>Split Data</i>	Data Latih	Data Uji
90:10 <i>test_size</i> =0.1	990	110
80:20 <i>test_size</i> =0.2	880	220
70:30 <i>test_size</i> =0.3	770	330
60:40 <i>test_size</i> =0.4	660	440
50:50 <i>test_size</i> =0.5	550	550

Melakukan *transform dataset* menggunakan ekstraksi fitur *CountVectorizer* yang bertujuan untuk mengubah teks menjadi representasi numerik (vektor fitur). Lalu menyematkan



perhitungan seleksi fitur *Chi Square* dan *Information Gain* pada *Python*. Setelah *data prepaation* selesai, maka diproses ke *modelling phase*.

### 3.4 Modelling

Membuat pemodelan sebanyak 3 model seperti pada Tabel 9. Ketiga model dilatih dan membuat objek prediksi pada data *testing* menggunakan pemrograman *Python* yang menghasilkan akurasi setiap model.

Tabel 9. Pemodelan Metode

Nama Model	Metode
Model 1	<i>Naive Bayes</i>
Model 2	<i>Naive Bayes + Chi Square</i>
Model 3	<i>Naive Bayes + Information Gain</i>

### 3.5 Evaluation

Menggunakan *confusion matrix* untuk menghitung nilai performa akurasi dari model yang dibuat seperti pada Tabel 10.

Tabel 10. Confusion Matrix

Confusion Matrix	Prediksi	
	Positif "1"	Negatif "0"
Positif "1"	TP (True Positif)	FN (False Negatif)
Negatif "0"	FP (False Positif)	TN (True Negatif)

Akurasi adalah ukuran berapa banyak prediksi yang benar dari model yang telah dibuat untuk menguji *dataset*. Formula perhitungan akurasi yaitu:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

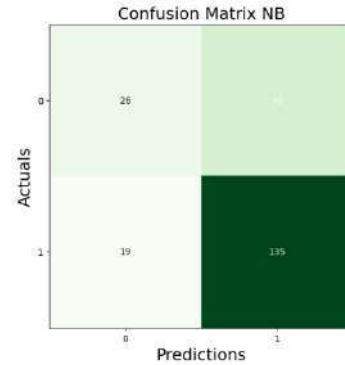
Proses analisis perhitungan menggunakan *Python*. Hasil akurasi klasifikasi ditampilkan dalam bentuk *metrics classification report*. Akurasi pada ketiga model di setiap split data dilihat pada Tabel 11.

Tabel 11. Hasil Akurasi Percobaan Split Data

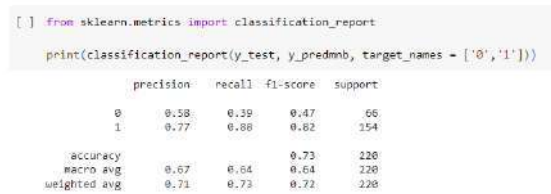
Rasio Split Data	Akurasi Model		
	NB	NB+CHI	NB+IG
90:10	69	62	64
80:20	73	75	74
70:30	61	62	64
60:40	68	63	67
50:50	69	67	68

Berdasarkan Tabel 11, dapat disimpulkan bahwa akurasi percobaan *split data* terbaik dihasilkan oleh rasio 80:20 dengan yang terbaik ialah model 2: NB+CHI sebesar 75%. Visualisasi *confusion matrix* dan *metrics classification report* ketiga model dengan rasio 80:20 dapat dilihat pada Gambar 4-9 dibawah ini:

A. Model 1: Naive Bayes (NB);

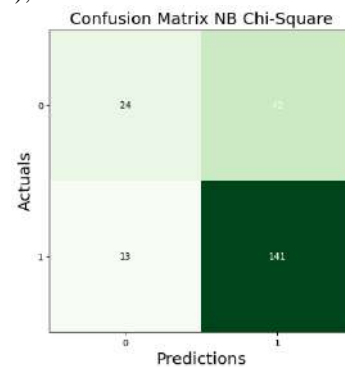


Gambar 4. Confusion Matrix Model 1

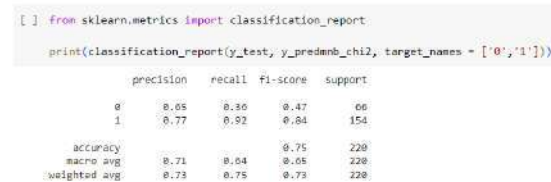


Gambar 5. metrics classification report Model 1

B. Model 2: Naive Bayes dengan Chi Square (NB+CHI);

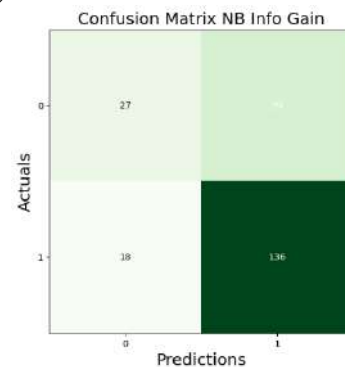


Gambar 6. Confusion Matrix Model 2



Gambar 7. metrics classification report Model 2

B. Model 3: Naive Bayes dengan Information Gain (NB+IG)



Gambar 8. Confusion Matrix Model 3

```
[ ] from sklearn.metrics import classification_report
```

```
print(classification_report(y_test, y_predemb_id, target_names = ['0', '1']))
```

	precision	recall	f1-score	support
0	0.60	0.41	0.49	66
1	0.78	0.88	0.83	154
accuracy			0.74	220
macro avg	0.69	0.65	0.66	220
weighted avg	0.72	0.74	0.72	220

Gambar 9. metrics classification report Model 3

3.6 WordCloud

Menampilkan visualisasi data klasifikasi sentimen positif dan sentimen negatif menggunakan WordCloud. Hasilnya menampilkan kata yang sering muncul berkaitan dengan vonis Ferdy Sambo. Visualisasi WordCloud menggunakan akurasi terbaik yaitu Model 2: NB+CHI split data 80:20 menampilkan WordCloud sentimen positif dan sentimen negatif.

A. Klasifikasi Sentimen Positif



Gambar 10. WordCloud Positif

B. Klasifikasi Sentimen Negatif



Gambar 11. WordCloud Negatif

4. Kesimpulan

Berdasarkan penelitian analisis sentimen kasus Ferdy Sambo pada Twitter menggunakan algoritma Naive Bayes dengan seleksi fitur Chi Square dan Information Gain yang telah dilakukan, maka dapat menarik kesimpulan bahwa model Naive Bayes pada split data 90:10 menghasilkan akurasi 69%, split data 80:20 menghasilkan akurasi 73%, split data 70:30 menghasilkan akurasi 61%, split data 60:40 menghasilkan akurasi 68%, split data 50:50 menghasilkan akurasi 69%. Dapat disimpulkan bahwa model Naive Bayes terbaik pada split data 80:20 dengan akurasi 73%; Model Naive Bayes dengan Chi Square pada split data 90:10 menghasilkan akurasi 62%, split data 80:20

menghasilkan akurasi 75%, split data 70:30 menghasilkan akurasi 62%, split data 60:40 menghasilkan akurasi 63%, split data 50:50 menghasilkan akurasi 67%. Dapat disimpulkan bahwa model Naive Bayes dengan Chi Square terbaik pada split data 80:20 dengan akurasi 75%. Model Naive Bayes dengan Information Gain pada split data 90:10 menghasilkan akurasi 64%, split data 80:20 menghasilkan akurasi 74%, split data 70:30 menghasilkan akurasi 64%, split data 60:40 menghasilkan akurasi 67%, split data 50:50 menghasilkan akurasi 68%. Dapat disimpulkan bahwa model Naive Bayes dengan Information Gain terbaik pada split data 80:20 dengan akurasi 74%. Metode klasifikasi terbaik menggunakan model Naive Bayes dengan Chi Square pada split data 80:20 dengan akurasi sebesar 75%. Tingkat ukuran performa model yang digunakan termasuk ke tingkat Fair Classification berdasarkan tingkat nilai diagnosa Receiver Operating Characteristic (ROC).

Saran peneliti yang dapat menjadi perhatian untuk penelitian selanjutnya ialah dilakukan penelitian lanjutan dengan proses labelling menggunakan lexicon based agar bisa mendapatkan pemahaman proses labelling yang terbaik dibanding dengan labelling manual selain memperoleh efisiensi waktu.

Daftar Pustaka

[1] Kementerian Hukum dan HAM RI, "Uu No 2 Tahun 2002," *Kepol. Negara Republik Indones.*, vol. 1999, 2002, [Online]. Available: <https://peraturan.go.id/common/dokumen/ln/2002/uu2-2002.pdf>.

[2] V. B. Kusnandar, "Akibat Kasus Ferdi Sambo, Tingkat Kepercayaan Masyarakat Terhadap Kepolisian Menurun," *databoks*, 26 Agustus 2022, 15:00 WIB, p. 1, 2022.

[3] Y. Heriyanto, "Perancangan Sistem Informasi Rental Mobil Berbasis Web Pada PT.APM Rent Car," *J. Intra-Tech*, vol. 2, no. 2, pp. 64-77, 2018.

[4] V. Zuliana, Garno, I. Maulana, and U. S. Karawang, "Analisis Sentimen Program Migrasi Tv Digital Menggunakan," no. 2, pp. 90-95, 2022.

[5] D. T. Larose, *Discovering Knowledge in Data: An Introducing to Data Mining*, vol. 1, no. 1. Hoboken, New Jersey: Wiley-Interscience, 2005.

[6] P. Chapman et al., *CRISP-DM 1.0: Step-by-step data mining guide*, vol. 78. Brussels, Belgium: SPSS Inc., 2000.

[7] D. T. Larose, *Data Mining Methods and Models*, vol. 1. Hoboken, New Jersey: Wiley-Interscience, 2007.

[8] J. Dean, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*, 1st ed. New Jersey, Canada: John Wiley & Sons, Inc., 2015.

- [9] N. Ye, *Data Mining: Theories, Algorithms, and Examples*. Boca Raton, Florida: CRC Press Taylor & Francis Group, 2015.
- [10] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining, Series.*, vol. 72. Berlin, German: Springer Science+Business Media, LLC, 2015.
- [11] A. Nisa, E. Darwiyanto, and I. Asror, "Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi," *e-Proceeding Eng.*, vol. 6, no. 2, p. 8650, 2019.
- [12] C. C. Aggarwal, *Data Mining: The Textbook*, vol. 14, no. 3. New York, USA: Springer Science+Business Media, LLC, 2015.
- [13] Nuryadi, T. D. Astuti, E. S. Utami, and M. Budiantara, *Buku Ajar Dasar-dasar Statistik Penelitian*, 1st ed. Yogyakarta: SIBUKU MEDIA, 2017.
- [14] Y. Arum Sari and M. Ali Fauzi, "Seleksi Fitur Information Gain untuk Klasifikasi Informasi Tempat Tinggal di Kota Malang Berdasarkan Tweet Menggunakan Metode Naïve Bayes dan Pembobotan TF-IDF-CF Automatic Essay Scoring View project Eye Movement Detection View project," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4907–4913, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [15] S. R. Ahmad, A. A. Bakar, and M. R. Yaakub, "A review of feature selection techniques in sentiment analysis," *Intell. Data Anal.*, vol. 23, no. 1, pp. 159–189, 2019, doi: 10.3233/IDA-173763.
- [16] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [17] M. R. Faisal, D. Kartini, T. H. Saragih, and A. Arrahimi, *Belajar Data Science: Text Mining Untuk Pemula I*. 2022.
- [18] R. Safitri, N. Alfira, D. Tamitiadini, N. Febriani, and W. W. A. Dewi, *Analisis Sentimen: Metode Alternatif Penelitian Big Data*, 1st ed. Malang: UB Press, 2021.
- [19] K. Moura and S. Fahey, *a beginner's guide to twitter for educators*. Marshall, London: DitchThatTextbook, 2016.

**PERINGKAT SINTA**  
**Jurnal Information Communication & Technology (JICT)**  
**Vol. 22 No. 2 (2022): JICT-IKMI, December 2022**



[Author](#) [Subjects](#) [Affiliations](#) [Sources](#) [FAQ](#) [WCU](#) [Registration](#) [Login](#)



**JURNAL ICT : INFORMATION COMMUNICATION & TECHNOLOGY**

LPPM STMIK IKMI CIREBON

P-ISSN : 23020261 <> E-ISSN : 23033363

**1.65306**  
Impact Factor

**442**  
Google Citations

**Sinta 4**  
Current Accreditation

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

History Accreditation







# Plagiarism Checker X Originality Report

**Similarity Found: 12%**

Date: Thursday, January 11, 2024

Statistics: 465 words Plagiarized / 3782 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

---

Jurnal **ICT : Information Communication & Technology** Vol. 23, N0.2, Desember 2023, pp. 440-447 p-ISSN: 2302-0261, e-ISSN: 2303-3363, DOI: 10.36054/jict-ikmi.v19i1.119 <https://ejournal.ikmi.ac.id/index.php/jict-ikmi> 4 40 **Analisis Sentimen Kasus Ferdy Sambo pada Twitter Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur** Arief Priyadi Pratama<sup>1\*</sup>, Prima Dina Atika<sup>2\*</sup>, Siti Setiawati<sup>3\*</sup> 1,2,3Program Studi Informatika, Universitas Bhayangkara Jakarta Raya , Indonesia Email: 1rief.priyadip@gmail.com, 2prima.dina@dsn.ubharajaya.ac.id ,3siti.setiawati@dsn.ubharajaya.ac.id **INFORMASI ARTIKEL ABSTRAK** Histori artikel: Naskah masuk, 19 Juli 2023 Direvisi, 31 Desember 2023 Diiterima, 31 Desember 2023 Abstract- The Indonesian National Police (Polri) is a government institution with integrity as a protector of public security.

However, there has been a case of premeditated murder by the Head of Propam Inspector General Ferdy Sambo which has gone viral case. This case attracted high public enthusiasm so that various opinions emerged on Twitter. The complexity in classifying opinions is a challenge because there is no standard for classifying opinions. To overcome these challenges, we need a sentiment analysis technique that can classify opinions accurately. Therefore, this study aims to analyze public opinion on Twitter using the Naive Bayes algorithm combined with the Chi Square feature selection method and Information Gain using the CRISP-DM methodology. The data used has been labeled by linguists.

The results of the research with the 80:20 data split show that **the Naive Bayes model** achieves an accuracy of 73%. The Chi Square combination makes accuracy increase to 75%. In addition, Information Gain also produces an accuracy of 74%. The experiment was carried out with five different data split ratios. **It can be concluded** that **the Naive Bayes model** with Chi Square feature selection and a data split ratio of 80:20 has an





**UNIVERSITAS BHAYANGKARA JAKARTA RAYA**  
**FAKULTAS ILMU KOMPUTER**

Kampus I: Jl. Harsono RM No. 67, Ragunan, Pasar Minggu, Jakarta Selatan, 12550  
Telepon: (021) 27808121 – 27808882  
Kampus II: Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat, 17142  
Telepon: (021) 88955882, Fax.: (021) 88955871  
Web: [fasilkom.ubharajaya.ac.id](http://fasilkom.ubharajaya.ac.id), E-mail: [fasilkom@ubharajaya.ac.id](mailto:fasilkom@ubharajaya.ac.id)

**SURAT TUGAS**

Nomor: ST/1152/XII/2023/FASILKOM-UBJ

Pertimbangan : Dalam rangka mewujudkan Tri Dharma Perguruan Tinggi untuk Dosen di Universitas Bhayangkara Jakarta Raya maka dihimbau untuk melakukan penelitian.

Dasar : 1. Kalender Akademik Universitas Bhayangkara Jakarta Raya Tahun Akademik 2023/2024;  
2. Rencana Kerja dan Anggaran Pembelanjaan Universitas Bhayangkara Jakarta Raya Tahun 2023.

**DITUGASKAN**

Kepada : Personil yang namanya tercantum dalam Surat Tugas ini.

NO.	NAMA	NIDN/NPM	JABATAN	KETERANGAN
1.	Arief Priyadi Pratama	201910225037	Mahasiswa Prodi Informatika	Sebagai Penulis Pertama
2.	Prima Dina Atika, S.Kom., M.Kom.	0311037107	Dosen Tetap Prodi Informatika	Sebagai Penulis Kedua
3.	Siti Setiawati, S.Pd., M.Pd.	0313107904	Dosen Tetap Prodi Informatika	Sebagai Penulis Ketiga

Untuk : 1. Membuat Artikel Ilmiah dengan judul **“Analisis Sentimen Kasus Ferdy Sambo pada Twitter Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur”** pada media Jurnal ICT: *Information Communication & Technology*, Vol. 23, No. 2, Desember 2023, Hal. 440-447, p-ISSN: 2302-0261, e-ISSN: 2303-3363.  
2. Melaksanakan tugas ini dengan penuh tanggung jawab.

Jakarta, 18 Desember 2023  
**DEKAN FAKULTAS ILMU KOMPUTER**  
  
**Dr. Dra. Tyastuti Sri Lestari, M.M.**  
NIP. 1408206