

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

## Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Kategori Artikel

# Analisis Sentimen Twitter Terhadap Kasus Ferdy Sambo Dengan Metode Logistic Regression

*Daribsan<sup>a</sup>, Alif Abdul Rauf<sup>a</sup>, Faiz Abdullah<sup>a</sup>, Muhammad Farhan Ananda Mirzah<sup>a</sup>*

<sup>a</sup> Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Universitas Andalas, Kota Padang, 25175, Indonesia

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 00 Februari 0000

Revisi Akhir: 00 Maret 0000

Diterbitkan Online: 00 April 0000

### KATA KUNCI

Twitter

Analisis Sentimen

Logistic regression

Ferdy Sambo

### KORESPONDENSI

E-mail: \*

### ABSTRACT

Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung bernilai negatif atau positif. Twitter adalah sebuah situs jejaring sosial memberikan akses kepada penggunanya untuk mengirimkan sebuah pesan singkat yang terdiri dari maksimal 280 karakter (disebut tweet). Tweet sendiri bisa terdiri dari pesan teks dan foto. Melalui tweet inilah pengguna Twitter dapat berinteraksi lebih dekat dengan pengguna Twitter lainnya dengan mengirimkan tentang apa yang sedang mereka pikirkan, apa yang sedang dilakukan, tentang kejadian yang baru saja terjadi, tentang berita terkini serta hal lainnya. Salah satu topik yang sering di tweet sekarang adalah kasus Ferdy Sambo yang menjadi perbincangan hangat sejak Juli 2022. Dalam kasus ini, Sambo diduga memerintahkan anak buahnya, Richard Eliezer atau Bharada E, untuk menembak Brigadir Polisi Nofriansyah Yosua Hutabarat di rumah dinas di Kompleks Polri Duren Tiga, Jakarta Selatan, Jumat (08/07/2022). Jurnal ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap kasus Ferdy Sambo dengan menggunakan metode analisis statistika Logistic Regression.

## 1. PENDAHULUAN

Pembunuhan Brigadir Polisi Nofriansyah Yosua Hutabarat alias Brigadir J atau Brigadir Y terjadi pada tanggal 8 Juli 2022 di rumah dinas Inspektur Jenderal Polisi Ferdy Sambo, ketika itu menjabat Kepala Divisi Profesi dan Pengamanan (Div Propam) Polri, di Kompleks Perumahan Polri Duren Tiga, Jakarta Selatan, DKI Jakarta.

Kasus ini menyeret perhatian masyarakat karena baik pelaku, korban, dan orang-orang yang terlibat di dalamnya kebanyakan merupakan anggota polisi, juga kejadiannya berlangsung di rumah seorang petinggi polisi. Selain itu, banyak pelintiran alur yang berakibat berubahnya Berita Acara Pemeriksaan (BAP) sehingga membuat kejadian sebenarnya tidak diketahui dengan pasti. Pengungkapan peristiwa ini ke masyarakat juga

menunjukkan kegagalan karena baru disampaikan tiga hari setelah terjadi, walaupun kemudian ada penjelasan bahwa itu terjadi karena peristiwanya berdekatan dengan Idul Adha. Dalam penelusuran selanjutnya juga ditemukan berbagai pelanggaran kode etik oleh para penyidik berupa sikap tidak profesional meliputi pengrusakan, penghilangan barang bukti, pengaburan, dan perekayasa kasus dugaan pembunuhan Brigadir J.

Salah satu tersangka pada kasus ini adalah Ferdy Sambo, seorang mantan Kepala Divisi Profesi dan Pengamanan Polri. Ia dijerat dengan persangkaan Pasal 340 KUHP subsidi Pasal 338 KUHP juncto Pasal 55 KUHP dan 56 KUHP [9].

Dalam kasus ini, Sambo diduga memerintahkan anak buahnya, Richard Eliezer atau Bharada E, untuk menembak Yosua di rumah dinas di Kompleks Polri Duren Tiga, Jakarta Selatan,

Jumat (08/07/2022). Setelahnya, mantan jenderal bintang dua Polri itu menembakkan pistol milik Brigadir J ke dinding-dinding rumah guna memuluskan skenario baku tembak yang dia susun.

Setelah ditetapkan menjadi tersangka, Sambo menjadi topik yang trending di Twitter dengan 44.200 tweet dan meningkat terus jumlahnya. Dalam artikel ini, akan dilakukan analisis sentimen twitter terhadap Ferdy Sambo dengan metode analisis statistika Logistic Regression.

## 2. Landasan Teori

### 2.1. Analisis Sentimen

Analisis Sentimen atau opinion mining merupakan cabang ilmu penelitian dari text mining yang bertujuan untuk menganalisis data tekstual dari sebuah persepsi atau opini public terhadap suatu topik pembahasan, permasalahan ataupun kejadian yang akan menghasilkan informasi berupa penilaian positif, negatif dan netral. Analisis sentimen dalam menganalisis data terdapat empat metode yang digunakan, yaitu Machine Learning, Statistical Model Approach, Lexicon Based Approach, dan Rule Based Approach. Pada penelitian ini digunakan metode Lexicon Based karena dapat memproses dan memprediksi opini dari suatu permasalahan berdasarkan data yang telah diperoleh [1].

Pada metode lexicon - based dalam mengidentifikasi kalimat opini yang terdiri dari positif, negatif dan netral. Terdapat 2 tahapan dalam menggunakan metode ini [1], yakni:

1. Menentukan kata bersentimen pada setiap kalimat.
2. Memberikan nilai pada setiap kalimat yang terdiri dari
3. nilai yaitu positif, negatif dan netral.

### 2.2. Machine Learning

Machine Learning merupakan serangkaian metode algoritma pemrograman yang digunakan dalam memprediksi data dalam jumlah besar atau data sampel yang akan dimodifikasi dan presentasikan dalam sebuah keputusan. Terdapat kegunaan dari machine learning [2], yaitu :

1. Classification atau klasifikasi digunakan untuk memprediksi suatu nilai dalam sebuah populasi.
2. Similarity matching digunakan untuk mengidentifikasi persamaan dari data yang diperoleh.
3. Clustering digunakan dalam mengelompokkan suatu individu dalam grup yang sama berdasarkan persamaan yang dimiliki.

### 2.3. Klasifikasi

Teknik klasifikasi adalah salah satu dari teknik data mining yang termasuk supervised learning. Supervised learning artinya proses pembentukan sebuah korespondensi menggunakan sebuah training dataset. Tujuannya adalah untuk memprediksi target dari beberapa atribut [3]. Terdapat pada dua pekerjaan utama pada klasifikasi yaitu melakukan training untuk disimpan sebagai prediksi dan melakukan testing untuk proses klasifikasi agar diketahui di label mana objek data tersebut [4].

### 2.4. Twitter

O'Reilly menyatakan pendapatnya mengenai media sosial sebagai platform yang dapat membantu berbagai aktivitas komunikasi, seperti halnya integrasi situs web, interaksi sosial, dan membuat konten komunikatif [5]. Twitter merupakan sebuah jenis medsos dengan beragam format yang memberikan fasilitas kepada penggunanya untuk menulis dan mempublish aktivitas maupun apa yang sedang mereka rasakan. Dalam sejarahnya, Twitter pertama kali hadir dan muncul menyediakan ruang tertentu dengan batas penuh 140 karakter untuk sekali cuitan. Tidak berbeda dengan berbagai media sosial lainnya, pada Twitter pengguna juga dapat menjalin hubungan dengan pengguna lainnya, menyebar serta bertukar informasi, mencurahkan persepsi/sudut pandang akan pengguna lain, bahkan mengikuti pembahasan isu terhangat (trending topic) dalam waktu yang bersamaan serta ikut ambil peran akan isu tersebut dengan ikut pula meng-tweet dengan penggunaan tagar (hashtag) isu tertentu [6].

### 2.5. Kasus Ferdy Sambo

Kasus pembunuhan Brigadir Joshua atau Brigadir J mencuat pada Jumat 8 Juli 2022. Saat itu Brigadir J diketahui tewas di rumah dinas Kadiv Propam Polri Irjen Pol Sambo. Dalam laporan yang dibuat, Ferdy Sambo melaporkan adanya kontak tembak antara Brigadir J dengan Bharada E. Tembak-tembakan ini terjadi karena Brigadir J diketahui melakukan pelecehan seksual terhadap Putri Chandrawati, istri Ferdy Sambo. Seiring berjalannya waktu dan desakan publik yang curiga adanya rekayasa di kasus tewasnya Brigadir J, pada 12 Juli 2022, Kapolri membentuk Tim Khusus Polri berdasarkan SPRIN Nomor SPRIN/5647/VII/HUK.12.1./2022. Kadiv Propam Polri Irjen Ferdy Sambo kemudian dinonaktifkan oleh Kapolri Listyo Sigit Prabowo di sela pengungkapan pembunuhan Brigadir Nofriansyah Yosua Hutabarat alias Brigadir J. Setidaknya, Kapolri Listyo Sigit menyebut empat alasan Ferdy Sambo dinonaktifkan, yakni alasan komitmen, objektivitas, transparansi, akuntabel dalam mengungkap kasus tewasnya Brigadir J oleh ajudan Ferdy Sambo, Barada E. Singkat cerita, fakta-fakta pun mulai terungkap, mulai dari adanya hambatan penyidikan seperti intimidasi, tekanan, intervensi, hingga menghilangkan barang bukti yang dilakukan beberapa anak buah Ferdy Sambo.

### 2.6. Logistic Regression

Logistic Regression adalah bagian dari metode data mining yang penggunaannya untuk menganalisis data yang mendeskripsikan antara satu variabel respon (dependen) atau lebih variabel predictor [7]. Metode ini digunakan saat variabel predictor (y) memiliki skala kategorik atau nominal yang terdiri dari dua (biner) atau lebih kategori. Sehingga metode ini dibuat untuk memastikan bahwa, apa pun perkiraan yang terjadi, selalu berada di antara 0 dan 1 [8].

### 3. METODE

#### 3.1. *Crawling Data*

Crawling adalah proses pengambilan data yang tersedia secara online untuk umum. Proses ini dilakukan dengan mengimpor informasi atau data yang telah ditemukan ke dalam file lokal di komputer. Crawling dilakukan untuk ekstraksi data yang mengacu pada pengumpulan data dari worldwide web, dokumen, dan lain-lain. Crawling data pada twitter dapat dilakukan dengan mengikuti beberapa langkah berikut yaitu dengan membuka jupyter notebook, kemudian buat file baru, import twint, dan memasukkan kode untuk mengambil data, setelah mengambil data, eksport data kedalam bentuk csv

#### 3.2. *Preprocessing Data*

Preprocessing merupakan teknik yang digunakan untuk merubah data mentah menjadi data yang berguna dengan mengeliminasi data yang tidak sesuai. terdapat beberapa langkah yang dapat dilakukan dalam melakukan data cleaning yaitu

##### 3.2.1. *Cleansing*

Cleaning merupakan proses membersihkan data yang bertujuan untuk menyeleksi data dan membuang data yang berpotensi mengurangi akurasi *machine learning*. diantara permasalahan yang biasa terjadi pada dataset yaitu *missing value*, *noisy data*, dan *inconstant data*.

##### 3.2.2. *Tokenization*

Tokenization adalah proses pembagian teks yang panjang menjadi bagian yang lebih kecil. Bagian-bagian yang lebih kecil ini biasa dikenal dengan token. Pemrosesan akan dilanjutkan saat kalimat-kalimat tersebut telah menjadi token. Proses tokenization juga disebut dengan segmentasi teks atau analisis leksikal.

##### 3.2.3. *Stopword Removal*

pada tahapan ini akan dihapus kata-kata yang dianggap tidak memberikan dampak terhadap informasi yang ada pada suatu data tweet. Pembersihan data stopwords pada penelitian ini menggunakan library python pySastrawi.

##### 3.2.4. *Stemming*

Dilakukan agar pada suatu data tweet yang memiliki imbuhan menjadi kata dasar dan menjadi suatu pola untuk dapat melakukan klasifikasi kata lain yang memiliki kata dasar maupun yang memiliki arti serupa

namun berbeda karena memiliki imbuhan yang beda

##### 3.2.5. *Labeling*

Labelling hanyalah proses pelabelan atau penandaan yang relevan. informasi/metadada dalam kumpulan data agar mesin memahami apa mereka. Dataset bisa dalam bentuk apapun yaitu, gambar, file audio, rekaman video, atau bahkan teks.

#### 3.3. *Processing Data*

Processing merupakan teknik pengolahan data dari data mentah menjadi sebuah data hasil yang bermanfaat. diantara langkah-langkah serta *output* dari tahapan processing yaitu mempersiapkan

## 4. HASIL DAN PEMBAHASAN

Berdasarkan data yang telah dikumpulkan melalui API Twitter, akan dilakukan metode - metode sebagai berikut:

#### 4.1 Crawling

Crawling adalah proses pengambilan data yang tersedia secara online untuk umum. Proses ini dilakukan dengan mengimpor informasi atau data yang telah ditemukan ke dalam file lokal di komputer. Crawling dilakukan untuk ekstraksi data yang mengacu pada pengumpulan data dari world wide web, dokumen, file dan lain-lain.

Data yang kami crawling adalah tweet dengan keyword berikut:

- #ferdysambo
- #percumalaporpholisi
- BrigadirJ
- kasusbrigadirj
- oknumpolisi
- percumalapori
- PercumaLaporPolisi
- polisi
- sambo

Dengan jumlah data awal 600-an tweet

```
In [4]: for i, tweet in enumerate(sntwitter.TwitterSearchScrapper('ferdysambo').get_items()):
        if i>4000:
            break
        tweets_list.append([tweet.content, tweet.user.username, analyze_sentiment(tweet.content)])
```

```
In [5]: tweet = pd.DataFrame(tweets_list, columns = ['Text', 'Username', 'Label'])
```

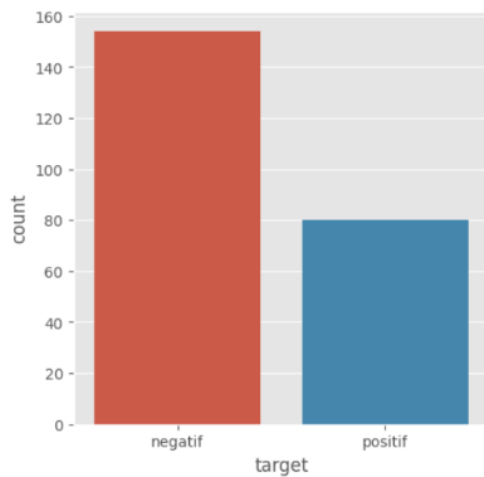
```
In [6]: tweet
```

```
Out[6]:
```

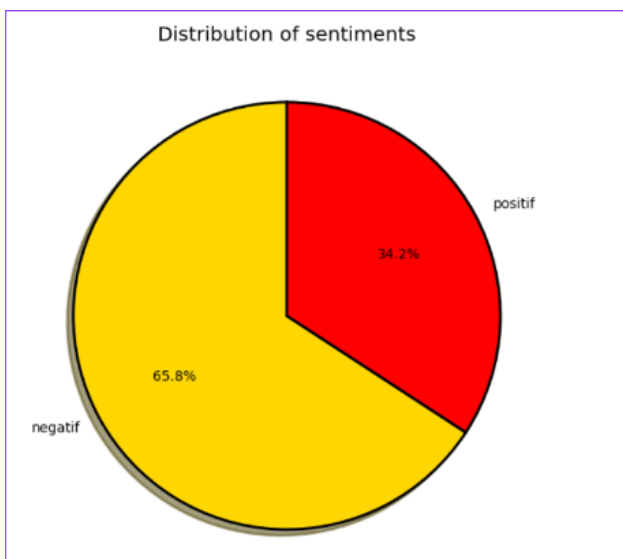
	Text	Username	Label
0	Manusia tergelok, tersampah, tertolol, terbeg...	bitcom3915	0
1	Ahli Ungkap Hasil Poligraf Kuat Manuf Bohong s...	KompasTV	0
2	Suara Ferdy Sambo Bergetar Tanggapi Kesaksian ...	TrjayaImSby	0
3	Lama ngkk ketemu teman yg di bahas Carita bola...	wannm28_	0
4	Agus Nurpatni mengiklaim mengantongi surat per...	taubentamedia	0
...	...	...	...
3996	Ferdy Sambo membantah adanya motif persegi...	detikcom	0
3997	Yang belum terungkap di persidangan : Hal apa ...	dedarich	0
3998	Semoga guru ngajimu samsudin jadab, tabungamu...	archmaar	0
3999	@CNNIndonesia bener-bener ferdysambo goblokk...	Dianyankalim	0
4000	Ujannya di kediri hari ini sangat konsisten ka...	syahneloyi	0

4001 rows x 3 columns





Gambar 7. Jumlah hasil Perbandingan Kata Negatif Dan Positif

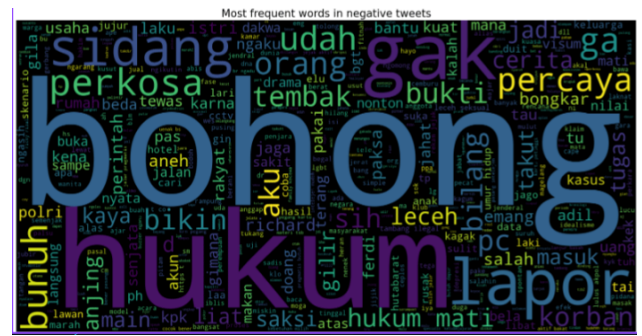


Gambar 8. Jumlah hasil persentase kata positif dan negatif



Gambar 9. Jumlah pengeluaran banyak kata positif

Pada Gambar 9. dijelaskan bahwa kata-kata yang paling banyak di dalam file tersebut yaitu adalah kata sidang



Gambar 10. Jumlah pengeluaran banyak kata negatif

### 4.3 Processing

Data Processing merupakan teknik pengolahan data dari data mentah menjadi sebuah data hasil yang bermanfaat.

Tahapan dari processing:

1. Mempersiapkan Data training, Testing data
2. Algoritma (Logistic Regression)
3. Hasil Algoritma
4. Akurasi

```
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

print("Size of x_train:", (x_train.shape))
print("Size of y_train:", (y_train.shape))
print("Size of x_test:", (x_test.shape))
print("Size of y_test:", (y_test.shape))

Size of x_train: (187, 2989)
Size of y_train: (187,)
Size of x_test: (47, 2989)
Size of y_test: (47,)
```

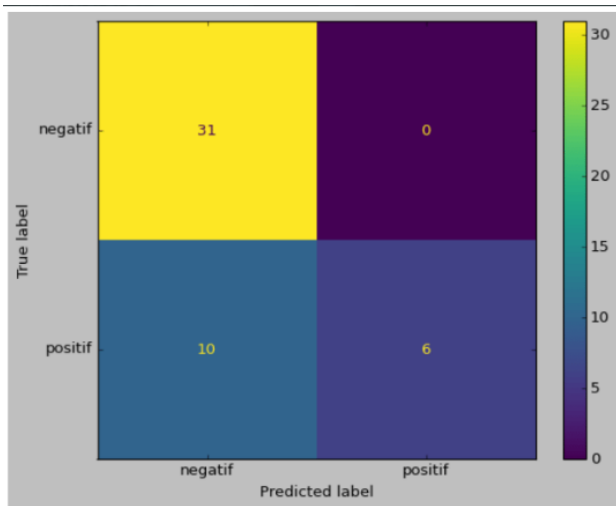
Gambar 11. Codingan pembuatan Data Training, Testing

```
In [128]: logreg = LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, l1_ratio=None, max_iter=5000, n_jobs=None, penalty='l2',
random_state=0, solver='liblinear', tol=0.0001, verbose=0,
warm_start=False)
logreg.fit(x_train, y_train)
logreg_pred = logreg.predict(x_test)
logreg_acc = accuracy_score(logreg_pred, y_test)
print("Test accuracy: (i.2f)%".format(logreg_acc*100))

Test accuracy: 78.72%
```

Gambar 12. Codingan pembuatan hasil Training Dari Model dan Akurasi





Gambar 13. Contoh Bentuk dari Confusion Matrix

```

In [118]: tweet = 'sambo itu jujur dan mulia'
          vectTweet = vect.transform(np.array([tweet])) # vectorizes the tweet using our vect

          prediction = logreg.predict(vectTweet) # predicts class of the tweet
          # print('Tweet is', 'positive' if prediction[0]=='positif' else 'negatif')
          if prediction[0]=='positif':
              print('positif')
          elif prediction[0]=='negatif':
              print('negatif')
          else :
              print('tidak berkategori')

          positif

In [119]: tweet = 'sambo menembak orang'
          vectTweet = vect.transform(np.array([tweet])) # vectorizes the tweet using our vect

          prediction = logreg.predict(vectTweet) # predicts class of the tweet
          # print('Tweet is', 'positive' if prediction[0]=='positif' else 'negatif')
          if prediction[0]=='positif':
              print('positif')
          elif prediction[0]=='negatif':
              print('negatif')
          else :
              print('tidak berkategori')

          negatif
  
```

Gambar 14. Codingan Pembuatan testing

## 5. KESIMPULAN

Analisis sentimen adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan berbagai platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Untuk artikel ini, dilakukan analisis sentimen menggunakan metode Logistic Regression pada platform Twitter terhadap kasus Ferdy Sambo.

Data yang diambil merupakan hasil crawling Tweet dari Twitter dari bulan November - Desember 2022. Dengan jumlah awal 600-an Tweet. Setelah itu dilakukan tahap Pre-processing dan Processing, dimana data mentah diubah menjadi data yang dapat digunakan dalam analisis sentimen.

Hasil dari analisis sentimen ini menunjukkan bahwa kebanyakan sentimen pengguna Twitter terhadap kasus Ferdy Sambo bersifat negatif (dengan persentase 65% negatif, 35% positif).

Penulis berharap bahwa analisis sentimen ini dapat bermanfaat bagi siapapun yang membutuhkannya.

## UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan karya tulis ilmiah ini. Penulisan artikel ini dilakukan dalam rangka memenuhi salah satu syarat untuk menyelesaikan Tugas Besar pada mata kuliah Akuisisi Data di Jurusan Sistem Informasi.

Penulis mengucapkan terima kasih kepada :

1. Bapak Febby Apri Wenando, M.Eng selaku dosen mata kuliah Akuisisi Data
2. Bapak Fajril Akbar, M.Sc selaku dosen mata kuliah Akuisisi Data
3. Seluruh anggota kelompok 4 mata kuliah Akuisisi Data kelas B
4. Seluruh Mahasiswa/i jurusan Sistem Informasi angkatan 2020

## DAFTAR PUSTAKA

- [1] Rachman,F.F.,&Pramana,S.Analisis Sentimen Pro dan Kontra Masyarakat IndonesiaTentang Vaksin Covid-19 Pada mediaSosialTwitter.IndonesianofHealthInformationManagementJournal.2020
- [2] Pambudi, H. K, Kusuma P. G., Yulianti, ., & Julian, K. A. Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning. Jitter. 2020.
- [3] Zaki, M. J., & Meira, W. J. 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms, 562. Retrieved from <https://books.google.com/books?hl=en&lr=&id=Gh9GAwAAQBAJ&pgis=1>
- [4]Liu, Y., Loh, H. T., & Sun, A. 2009. Imbalanced text classification: A term weighting approach. Expert Systems with Applications, 36(1), 690–701. <https://doi.org/10.1016/j.eswa.2007.10.042>
- [5]Tresnawati, Y. (2017). Pemanfaatan Media Sosial Dalam Promosi Potensi Wisata Bahari Cilacap Jawa Tengah. Perspektif Komunikasi: Jurnal Ilmu Komunikasi Politik Dan Komunikasi Bisnis, 1(2), 1–11. <https://jurnal.umj.ac.id/index.php/perspektif/article/view/2846>
- [6]Nasrullah, R. (2015). Media Sosial: Perspektif Komunikasi, Budaya, dan Sosioteknologi. Simbiosis Rekatama Media.

[7] C. Science, A. Bimantara, and T. A. Dina, “Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression,” vol. 4, no. 1, pp. 978 – 979, 2018.

[8] O. S. Balogun, T. J. Akingbade, and A. Akinrefon, “Evaluation Of Logistic Regression In Classification Of Drug Data In Kwara State,” no. May 2014, 2013

[9] Indonesia, Kitab Undang Undang Hukum Pidana (KUHP)

## **BIODATA PENULIS**

Alif Abdul Rauf

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020. Lahir di Padang Panjang pada tanggal 11 November 2001.

Muhammad Farhan Ananda Mirzah

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020. Lahir di Padang pada tanggal 2 Maret 2002.

Daribsan

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020.

Faiz Abdullah

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020.