



# Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika dengan *Support Vector Machine* (SVM) dan *Naïve Bayes*

Rahmat Satria Buana <sup>1</sup>, Windu Gata <sup>2\*</sup>, Alda Zevana Putri Widodo <sup>3</sup>, Hendra Setiawan <sup>4</sup>,  
Khairunisa Hilyati <sup>5</sup>

<sup>1,2,3,5</sup> Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia.

<sup>4</sup> Fakultas Sistem Informasi, STMIK Bani Saleh, Kota Bekasi, Provinsi Jawa Barat, Indonesia.

## article info

### Article history:

Received 6 July 2022

Received in revised form

23 November 2022

Accepted 14 February 2023

Available online April 2023

### DOI:

<https://doi.org/10.35870/jti.k.v7i2.705>

### Keywords:

Sentiments; Rain Handler; SVM; Naïve Bayes; Smote.

### Kata Kunci:

Sentimen; Pawang Hujan; SVM; Naïve Bayes; Smote.

## abstract

Indonesia is a country with a majority Muslim population, but Indonesia is also rich in culture influenced by previous religions such as Hinduism and Buddhism. The rain handler by Muslims is considered shirk, but on the other hand also considers this ancestral culture that has always existed. With the advancement of human technology, it is easier to express opinions or opinions regarding a topic that is currently being discussed, for example, regarding the rain handler who acted at the Mandalika Circuit some time ago through social media. Twitter is one of the social media that is used as a forum to accommodate these opinions. In this study, the CRISP-DM (Cross Industry Standard Process for Data Mining) data mining methodology was used with Rapid Miner Version 9.10 with the Support Vector Machine (SVM) classification method and nave Bayes with SMOTE to improve accuracy. The application of the SVM method is 97.71 % with AUC 0.997 Positive Comments, and Using Naïve Bayes, the accuracy obtained is 93.41% accuracy with AUC 0.973 Positive Comments.

## abstrak

Indonesia adalah negara yang mayoritas penduduknya muslim, namun Indonesia juga kaya akan budaya yang terpengaruh agama-agama sebelumnya seperti hindu dan budha. Pawang hujan oleh umat islam di anggap kesyirikan, namun di lain pihak juga menganggap ini budaya nenek moyang yang sejak dulu ada. Dengan semakin majunya perkembangan tekhnologi manusia semakin mudah mengemukakan pendapat atau opininya terkait suatu topik yang sedang ramai di bicarakan misalnya terkait pawang hujan yang beraksi di Sirkuit Mandalika beberapa waktu lalu melalui media sosial. Twitter adalah Media sosial yang dijadikan salah satu media sosial sebagai tempat menyampaikan opini tersebut. Pada penelitian ini di gunakan metodologi data mining CRISP-DM (Cross Industry Standard Process for Data Mining) dengan alat bantu Rapid miner Versi 9.10 dengan Metode klasifikasi Support Vector Machine (SVM) dan Naïve Bayes dengan SMOTE untuk meningkatkan akurasi. Penerapan metode SVM 97,71 % dengan AUC 0.997 Komentar Positif, dan Menggunakan Naïve Bayes maka akurasi yang didapat akurasi 93,41 % dengan AUC 0.973 Komentar Positif.

\*Corresponding author. Email: [windu@nusamandiri.ac.id](mailto:windu@nusamandiri.ac.id) <sup>2\*</sup>.

© E-ISSN: 2580-1643.

Copyright © 2023. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISEI)

(<http://creativecommons.org/licenses/by/4.0/>).

## 1. Latar Belakang

Rara Isti Wulandari menjadi perbincangan hangat dikalangan masyarakat nasional maupun internasional lantaran aksinya sebagai pawang hujan di ajang MotoGP Mandalika 2022 [1]. Aksi tersebut menjadi perbincangan oleh warga *internet* (warganet) hingga menjadi trending topik di Twitter. Banyak yang mulai mempertanyakan latar belakang dari perempuan yang biasa dikenal dengan sebutan mbak Rara. Hujan deras yang terus mengguyur area Sirkuit Mandalika, membuat kegiatan balapan MotoGP ditunda sampai hujan mereda. Rara, sang pawang hujan menggunakan piranti khusus untuk melakukan ritual menangkal hujan. Perbuatannya memindahkan hujan menuai pro dan kontra pada komen twitter.

Twitter merupakan salah satu media sosial yang banyak yang digunakan di Indonesia berdasarkan survey yang dilakukan oleh Lembaga survey Hootsuite yang bekerja sama dengan perusahaan media We Are Social pada tahun 2019 menerbitkan angka pemakai twitter aktif di Indonesia sebesar 6,43 juta. Sebagai negara dengan pemakai twitter dengan angka tertinggi ke-10 [2]. Berdasarkan data *tweet* yang diambil dari twitter lalu dilakukan data *preprocessing* dilanjutkan dengan penerapan Smote kemudian di terapkan model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) dengan label klasifikasi negatif dan positif. *Tools* yang kami gunakan dalam penelitian ini adalah rapid miner. *Tools* ini dipakai untuk mengambil data dari twitter. yang akan digunakan dalam penelitian ini. Opini akan diolah melalui metode-metode yang disesuaikan dengan kaidah proses pengambilan teks atau text mining [3].

Sebagai referensi penelitian penulis mengambil penelitian sebelumnya tentang sentimen dengan algoritma SVM atau naive bayes. Penelitian pertama adalah penelitian tentang Analisis Sentimen KRI Nanggala 402 di Twitter dengan Metode *Naïve Bayes Classifier* dan mendapatkan tingkat akurasi sebesar 73,00% [4]. Penelitian kedua, Analisis Sentimen Terkait Penggunaan Vaksin Astra Zeneca pada Twitter dengan Metode NB dan K-NN didapati metode NB mendapat akurasi yang baik yaitu 90,71% dan KNN hanya 74,78% [5]. Pada penelitian ketiga Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan metode SVM,

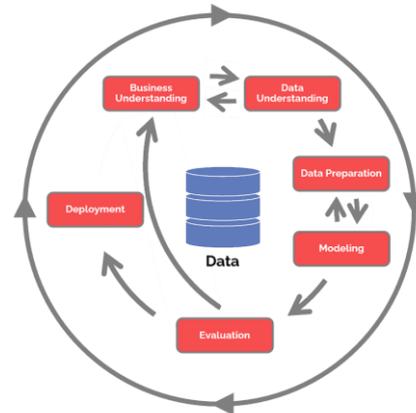
menghasilkan tingkat akurasi 87% [6]. Keempat adalah penelitian tentang analisis sentiment keberhasilan Indonesia pada Thomas Cup tahun 2020 dengan NB dan Decision Tree [7]. Referensi terakhir dalam penelitian ini adalah Analisis terkait pelayanan indihome pada twitter dengan algoritma SVM, dengan akurasi 87% [8]. Penelitian ini bertujuan untuk melihat opini masyarakat Indonesia akan keberadaan pawang hujan Mandalika dan meningkatkan akurasi penelitian sebelumnya tentang analisis sentimen dengan algoritma SVM atau NB dengan penerapan *Smote upsampling*.

Analisis Sentimen adalah suatu teknik analisis data yang digunakan untuk mengidentifikasi dan mengevaluasi sentimen, opini, dan emosi yang terkandung dalam teks yang tidak terstruktur [3], seperti komentar di media sosial [9]. Salah satu media sosial yang sering digunakan untuk mengungkapkan pendapat dan sentimen adalah Twitter [10]. Dalam konteks ini, Analisis Sentimen dilakukan pada komentar Twitter yang berkaitan dengan Pawang Hujan Mandalika, yaitu seorang pawang hujan yang melakukan ritual metafisik untuk membantu masyarakat menahan atau memindahkan hujan ke tempat lain. Pawang hujan dalam melakukan ritualnya menggunakan cara metafisik, yang terkadang sulit bagi kita untuk mencernanya dengan akal sehat. Keberadaan ritual pawang hujan ini masih mengandung nilai mitologis bahkan bersifat magis [11][12]. Pada twitter terdapat istilah tweet yaitu pengguna twitter dapat memberikan kabar terbaru, berekspresi, beraspirasi, dan beropini yang ditulis oleh pengguna twitter lainnya terutama topik atau hal yang sedang menjadi perbincangan utama [13]. Analisis sentimen bertujuan untuk mengetahui opini publik terhadap Pawang Hujan Mandalika, apakah positif atau negative. Untuk melakukan Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika, terdapat dua metode yang sering digunakan, yaitu Support Vector Machine (SVM) dan Naïve Bayes. SVM merupakan metode klasifikasi yang sangat populer dalam analisis sentimen karena dapat mengatasi masalah overlap dan memisahkan kelas dengan baik [14]. Sedangkan Naïve Bayes adalah metode klasifikasi yang menggunakan teori probabilitas dan statistik untuk menentukan kelas dari data yang diberikan. Sebelum melakukan analisis sentimen dengan SVM atau Naïve Bayes, terlebih dahulu data komentar Twitter perlu dipreproses.

*Preprocessing* yang dilakukan antara lain penghapusan tanda baca, konversi huruf kecil, penghilangan kata stopword, dan stemming [15]. Setelah itu, dilakukan tokenisasi, yaitu memecah teks menjadi unit-unit yang lebih kecil. Selanjutnya, dilakukan pembagian data menjadi data latih dan data uji. Data latih digunakan untuk melatih model SVM atau Naïve Bayes, sedangkan data uji digunakan untuk menguji akurasi model. Pada tahap ini, juga dilakukan pembuatan model SVM atau Naïve Bayes dengan menggunakan library atau software yang tersedia, seperti Rapid Miner [16][17]. Setelah model dibuat, dilakukan klasifikasi pada data uji untuk menentukan sentimen dari setiap komentar Twitter. Dalam klasifikasi SVM, setiap data pelatihan dikenal sebagai  $(x_i, y_i)$ , di mana  $i = 1, 2, \dots, N$  dan  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}$   $T$  adalah atribut untuk data pelatihan  $I$ ,  $y_i \{-1, +1\}$  adalah Kelas Label. Sedangkan dalam Naïve Bayes, dilakukan perhitungan probabilitas untuk menentukan kelas dari setiap data uji [18]. Hasil analisis sentimen yang diperoleh dapat digunakan untuk mengambil keputusan yang lebih baik [19][20], misalnya untuk meningkatkan layanan pawang hujan Mandalika atau untuk membuat kampanye yang lebih efektif dalam mempromosikan pawang hujan Mandalika. Selain itu, hasil analisis sentimen juga dapat menjadi dasar untuk melakukan penelitian lebih lanjut tentang pengaruh media sosial terhadap opini publik.

Penelitian ini dilakukan untuk menganalisis sentimen pada Komen Twitter Pawang Hujan Mandalika dengan menggunakan metode Support Vector Machine (SVM) dan Naïve Bayes. Dalam penelitian ini, terdapat penemuan baru yang menarik bahwa SVM mampu memberikan akurasi yang lebih tinggi dalam analisis sentimen pada Komen Twitter Pawang Hujan Mandalika dibandingkan dengan Naïve Bayes. Selain itu, juga ditemukan bahwa SVM lebih baik dalam mengenali sentimen negatif pada Komen Twitter tersebut. Penelitian ini juga memberikan kontribusi penting bagi pengembangan analisis sentimen pada Komen Twitter, terutama dalam memahami sentimen pengguna terhadap Pawang Hujan Mandalika. Selain itu, penggunaan SVM dalam analisis sentimen juga dapat memberikan hasil yang lebih akurat dan efektif. Diharapkan penemuan baru ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam bidang analisis sentimen pada media sosial.

## 2. Metode Penelitian



Gambar 1. Model Crisp DM

Metode penelitian pada penelitian ini adalah Crisp-DM [13] seperti yang di tunjukan pada gambar 1 dimana data akan diproses dengan melakukan beberapa tahapan seperti berikut:

### *Pemahaman Bisnis (Business Understanding)*

Pada tahapan ini dilakukan penentuan permasalahan yang akan diangkat yaitu analisis sentimen terhadap komentar pengguna twitter terkait keberadaan pawang hujan di acara Motto GP Mandalika, langkah yang dilakukan yaitu mencari data komentar yang terdapat pada twitter

### *Pemahaman Data (Data Understanding)*

Pada tahapan ini peneliti bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki. Data tersebut juga harus dapat diverifikasi kebenarannya. Data yang digunakan dalam penelitian ini adalah data komentar twitter mengenai pawang hujan di acara Moto GP Madalika.

### *Koleksi Data (Data Collection)*

Pada tahapan ini dilakukan pembuatan koneksi ke API twitter dengan request token. Setelah koneksi berhasil dilakukan pengambilan data komentar yang berbahasa teks Indonesia di twitter pada tanggal 28 maret sampai 04 April 2022 dengan kata pencarian pawang hujan mandalika. Proses pengambilan data komentar yang dilakukan pada penelitian ini diambil dengan menggunakan aplikasi rapid miner.

### *Persiapan Data (Data Preparation)*

Pada tahapan ini data yang sudah didapatkan sebelumnya, akan diberikan sentimen. Setelah data

yang telah didapatkan diberikan sentimen, setelah itu data akan masuk tahap data *preprocessing*. Berikut adalah tahapan-tahapan yang dilakukan pada tahapan data *preprocessing*:

- 1) *Replace*
  - a) *Missing value* : untuk mengganti nilai yang kosong dengan suatu nilai tertentu
  - b) RT, tanda baca [-!"#%&'()\*+./:;<=>?@\[\]\\_`{|}~]:
  - c) URL
  - d) *Hastag*
- 2) Filter untuk membedakan komentar yang berisi text atau tidak
- 3) *Remove duplicate* : untuk menghilangkan komentar yang sama
- 4) *Tokenizing* merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf.
- 5) *Stopwords* merupakan proses menghilangkan kata yang tidak memiliki makna atau tidak penting dengan menggunakan stopword list.

*Pemodelan*

Pada tahapan ini akan dilakukan pemodelan terhadap data komentar yang sudah dilakukan *preprocessing*. Data komentar tersebut akan dipecah menjadi data training dan data testing, pada tahapan ini akan lakukan split data. Pada split data peneliti akan membagi data komentar tersebut dengan perbandingan data training dan data testing yaitu 70:30. Kemudian pada tahap pemodelan ini peneliti menggunakan algoritma SVM dan NB untuk mendapatkan nilai sentimen positif dan sentimen negatif dari sebuah data komentar.

*Evaluasi*

Pada tahapan ini akan dilakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan SMOTE terhadap algoritma SVM dan NB

*Kesimpulan Saran*

Pada tahapan ini akan dilakukan pemberian kesimpulan yang telah didapat selama melakukan penelitian ini serta memberikan saran untuk penelitian lanjutan yang akan dilakukan oleh peneliti lain.

### 3. Hasil dan Pembahasan

*Business Understanding*

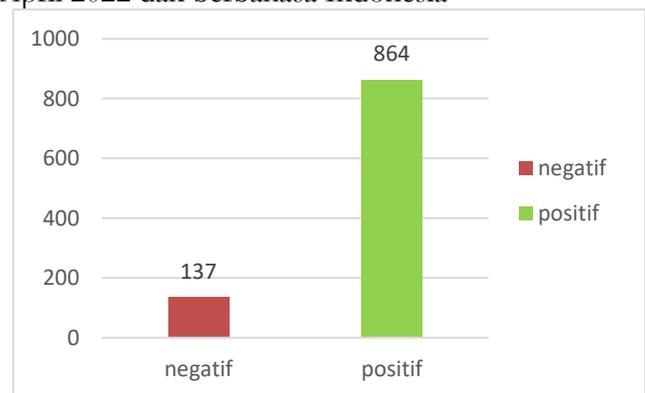
Pada tahapan *business understanding* ini maka peneliti mencoba untuk memahami objek atau masalah yang diangkat selama penelitian. Masalah yang diangkat oleh peneliti adalah seberapa besar akurasi yang dihasilkan dari penerapan algoritma SVM dan NB dalam mengklasifikasikan sentimen dari sebuah komentar yang berkaitan dengan pawang hujan. Penelitian ini akan menggunakan data yang didapat melalui twitter

*Data Understanding*

Pada tahapan data understanding ini maka peneliti mencoba untuk memahami data yang akan digunakan. Pada tahapan ini peneliti melakukan pengambilan data komentar pengguna pada tanggal 28 maret 2022 sampai 04 April 2022 terkait opini pawang hujan. Data yang diambil yaitu hanya komentar berupa text. Hasil crawling memperoleh data komentar sebanyak 4985. Dalam memberikan label pada komentar peneliti menggunakan pelabelan menggunakan extensi meaning cloud di rapid miner.

*Data Collection*

Pada tahapan *data collection* ini maka peneliti mengumpulkan data yang akan digunakan selama penelitian dan mempelajari apa yang dapat dilakukan dengan data tersebut. Pada tahapan ini peneliti membatasi 1002 komentar dari 1631 hasil cleansing data komentar twitter pada 28 maret 2022 sampai 04 April 2022 dan berbahasa Indonesia



Gambar 2. Perhitungan Komentar berdasarkan label

*Data Preparation*

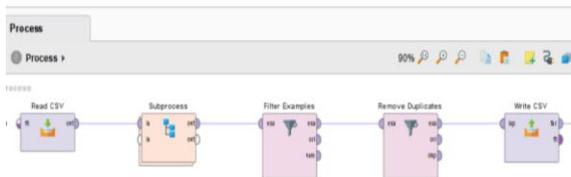
Pada tahapan ini data yang sudah didapatkan sebelumnya, akan diberikan sentimen. Setelah data yang telah didapatkan diberikan sentimen, setelah itu

data akan masuk tahap data *preprocessing*. Berikut adalah tahapan-tahapan yang dilakukan pada tahapan data *preprocessing*.

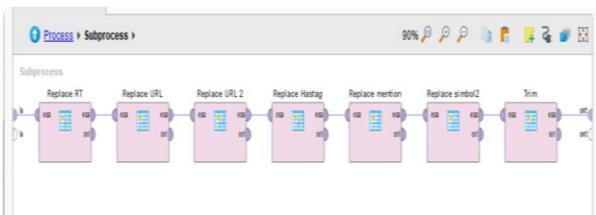
1) *Replace*

- a) *Missing value* : dilakukan pengantian nilai kosong dengan nilai rata-rata
- b) mengosongkan tanda baca [-!#\$%&'()\*+,-./:;<=>?@\[\]\\_`{|}~]: dengan
- c) mengosongkan RT
- d) URL : dilakukan pengantian kata URL dari baris komentar
- e) *Hashtag* : dilakukan pengantian kata *Hashtag* dari baris komentar

- 2) Filter disini peneliti penyaringan baris komentar yang berisi text atau tidak.
- 3) *Remove duplicate* : disini peneliti melakukan remove duplicare untuk menghilangkan komentar yang sama.
- 4) *Tokenizing* disini peneliti melakukan proses tokenizing untuk memecah ulasan dari kalimat menjadi perkata, menghilangkan karakter yang bersifat pembatas, dan menghapus angka atau tanda baca.
- 5) *Stopwords* disini peneliti melakukan *stopwords* untuk menghilangkan kata yang tidak penting atau tidak deksriptif dengan menggunakan *stopwordlist*.



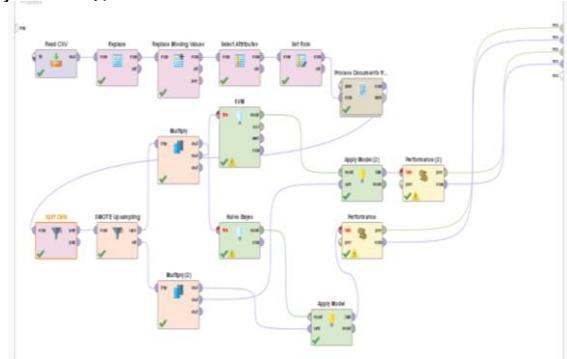
Gambar 3. *Preprocessing data*



Gambar 4. Sub Proses dari *Preprocessing data*

*Modeling*

Setelah *preprocessing* maka peneliti mensplit data menjadi data training 70 dan testing 30 lalu di aplikasikan metode smote upsampling untuk mengatasi ketidak seimbangan data, kemudian di terapkan algoritma SVM dan NB.



Gambar 5. Pemodelan dari data yang sudah melalui *preprocessing*

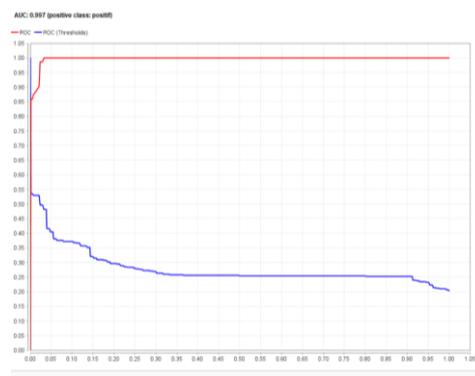
*Evaluation*

Setelah di lakukan *preprocessing* dan penerapan metode SMOTE pada data set komentar maka didapati akurasi yang di dapat adalah sebagai berikut :

- 1) Menggunakan SVM maka akurasi yang didapat akurasi 97,71 % dengan AUC 0.997 Komentar Positif

Tabel 1. Daftar hasil prediksi dengan SVM

No	Uraian	True negatif	True positif	Class precision
1	pred. negatif	123	5	96.09%
2	pred. positif	3	218	98.64%
3	class recall	97.62%	97.76%	

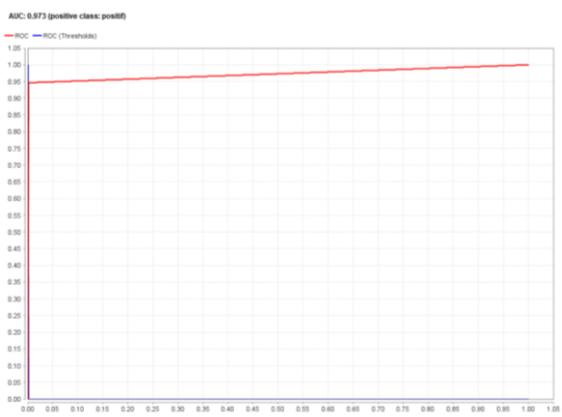


Gambar 6. Grafik AUC Dengan SVM

- 2) Menggunakan *Naïve Bayes* maka akurasi yang didapat akurasi 93,41 % dengan AUC 0.973  
 Komentar Positif

Tabel 2. Daftar hasil prediksi dengan *Naïve Bayes*

No	Uraian	True negatif	True positif	Class precision
1	pred. negatif	125	23	84.56%
2	pred. positif	0	200	98.64%
3	class recall	100%	89.69%	



Gambar 7. Grafik AUC Dengan *Naïve Bayes*.

#### 4. Kesimpulan dan Saran

Berdasarkan hasil analisis dan pembahasan sebelumnya, penelitian ini menyimpulkan dua hal penting. Pertama, mayoritas masyarakat Indonesia memberikan komentar positif terhadap keberadaan Pawang Hujan pada acara Moto GP Mandalika, sehingga kehadiran Pawang Hujan masih dianggap relevan dan penting oleh masyarakat. Kedua, terbukti bahwa metode Smote Upsampling dapat meningkatkan performa akurasi dan presisi klasifikasi dalam analisis sentimen, dengan akurasi mencapai 97,71% pada pemodelan SVM dan 93,41% pada pemodelan Naive Bayes.

Berbagai saran dapat diberikan berdasarkan hasil penelitian ini. Pertama, peneliti menyarankan agar dilakukan penelitian atau pengembangan lebih lanjut menggunakan metode lainnya dan dataset yang lebih baru dan lebih banyak dari penelitian sebelumnya, sehingga hasilnya dapat lebih terpercaya. Kedua, hasil

analisis sentimen ini dapat menjadi pertimbangan bagi pihak pemerintah dalam mengambil kebijakan lebih lanjut terkait dengan opini masyarakat tentang keberadaan Pawang Hujan pada acara Moto GP Mandalika.

#### 5. Daftar Pustaka

[1] 6 Fakta Rara Isti Wulandari, Pawang Hujan MotoGP Dibayar Ratusan Juta (no date). Available at: <https://www.viva.co.id/sport/oneprix/14594-07-6-fakta-rara-isti-wulandari-pawang-hujan-motogp-dibayar-ratusan-juta> (Accessed: 5 July 2022).

[2] Zellatifanny, C.M., 2019. Respon Pengguna Twitter terhadap Regulasi Pengendalian Akses Ponsel Ilegal melalui Validasi IMEI (Twitter User's Response to Regulation of Contraband Cell Phone Access Control through IMEI Validation). *JURNAL IPTEKKOM (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, 21(2), pp.139-154. DOI: 10.33164/iptekkom.21.2.2019.139-154.

[3] Putra, A.D.A. and Juanita, S., 2021. Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 8(2), pp.636-646. DOI: 10.35957/jatisi.v8i2.962.

[4] Djamaludin, M.A., Triayudi, A. and Mardiani, E., 2022. Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 6(2), pp.161-166. DOI: 10.35870/jtik.v6i2.398.

[5] Ramadhan, M.A. and Wahyudin, M.I., 2022. Analisis Sentimen Mengenai Keberhasilan Indonesia di Ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode Naïve Bayes dan Decision Tree. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 6(4), pp.505-511. DOI: 10.35870/jtik.v6i4.560.

- [6] Ahmad, A. and Gata, W., 2022. Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), pp.548-555. DOI: 10.35870/jtik.v6i4.569.
- [7] Ramadhani, S.H. and Wahyudin, M.I., 2022. Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 6(4), pp.526-534. DOI: 10.35870/jtik.v6i4.530.
- [8] Tineges, R., Triayudi, A. and Sholihati, I.D., 2020. Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3), pp.650-658. DOI: 10.30865/mib.v4i3.2181.
- [9] Tripathi, P., Vishwakarma, S.K. and Lala, A., 2015, December. Sentiment analysis of english tweets using rapid miner. In *2015 international conference on computational intelligence and communication networks (CICN)* (pp. 668-672). IEEE. DOI: 10.1109/CICN.2015.137.
- [10] Ikoro, V., Sharmina, M., Malik, K. and Batista-Navarro, R., 2018, October. Analyzing sentiments expressed on Twitter by UK energy company consumers. In *2018 Fifth international conference on social networks analysis, management and security (SNAMS)* (pp. 95-98). IEEE. DOI: 10.1109/SNAMS.2018.8554619.
- [11] Kurnia, S., Marnelly, T. R. and Si, M. 2017. The Role of Community Toward Rain-Moving Ritual in Tualang District Siak Regency, *Widya Jl. HR. Soebrantas Km*, 4(2), p. 28293.
- [12] Nurfitriyanti. 2012. Kepercayaan Masyarakat Terhadap Pawang Hujan di Desa Kedaburapat Kepulauan Meranti Di tinjau dari Aqidah Islam, pp. 1–68.
- [13] Azeharie, S., 2014. Analisis penggunaan Twitter sebagai media komunikasi selebritis di Jakarta. *Jurnal Komunikasi*, 6(2), pp.83-98. DOI: <https://doi.org/10.24912/jk.v6i2.35>.
- [14] Shafique, U. and Qaiser, H., 2014. A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), pp.217-222. Available at: <http://www.ijisr.issr-journals.org/>.
- [15] Rifai, M.F., Jatnika, H. and Valentino, B., 2019. Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)., *Petir*, 12(2), pp. 131–144. DOI: 10.33322/petir.v12i2.471.
- [16] Ruh yana, N., 2019. Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes. *ikraith-informatika*, 3(1), pp.94-99. Available at: [www.situs.com](http://www.situs.com).
- [17] Darwis, D., Pratiwi, E.S. and Pasaribu, A.F.O., 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, 7(1), pp.1-11. DOI: 10.21107/edutic.v7i1.8779.
- [18] *Algoritma Naive Bayes | BINUS UNIVERSITY BANDUNG - Kampus Teknologi Kreatif* (no date). Available at: <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/> (Accessed: 14 June 2022).
- [19] Palechor, F.M. and de la Hoz Manotas, A., 2019. Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in brief*, 25, p.104344. DOI: 10.1016/j.dib.2019.104344.
- [20] Franseda, A., Kurniawan, W., Anggraeni, S. and Gata, W., 2020. Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 8(3), pp.282-290. DOI: 10.26418/justin.v8i3.40982.