

# Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan *Naïve Bayes*

<http://dx.doi.org/10.28932/jutisi.v7i3.4094>

Riwayat Artikel

Received: 15 Oktober 2021 | Final Revision: 25 November 2021 | Accepted: 25 November 2021

Kristian Adi Nugraha <sup>✉</sup>#1

# Program Studi Informatika, Universitas Kristen Duta Wacana  
Jl. Dr. Wahidin Sudirohusodo No. 5-25 Yogyakarta 55224

<sup>1</sup>adinugraha@ti.ukdw.ac.id

**Abstract**— The usage of social media growing rapidly, especially after the smartphone was invented. Because the number of social media users was quite a lot, companies prefer to promote their products through social media like Instagram. But, unlike TV or radio, social media is a two-way communication media, that makes users can respond directly to the content created by the company. Comments given by users have various types of sentiment, like positive or negative comments. In addition to using text, comments also often contain emoticons to support the message. This study tries to analyzing sentiment based on the usage of emoticons inside them using the *Naïve Bayes* algorithm. Based on the test results, the accuracy result is quite good, it is about 96.3% correct in sentiment classification.

**Keywords**— comment; emoticon; naïve bayes; sentiment analysis.

## I. PENDAHULUAN

Media sosial merupakan salah satu teknologi atau perangkat lunak yang digunakan oleh sebagian besar penduduk di dunia. Di Indonesia, jumlah pengguna media sosial mencapai kurang lebih 193 juta pengguna atau sekitar 70% dari total keseluruhan populasi penduduk [1]. Jumlah tersebut tergolong cukup besar karena mencakup sebagian besar penduduk Indonesia yang terdiri dari berbagai rentang usia, mulai dari anak-anak hingga lansia. Dengan melihat potensi jumlah pengguna media sosial yang cukup besar, maka saat ini banyak kegiatan atau aktivitas yang beralih dilakukan menggunakan media sosial seperti iklan, promosi, pemberitahuan, dan segala jenis bentuk distribusi informasi lainnya.

Proses distribusi informasi melalui media sosial berlangsung secara dua arah, yaitu para penerima informasi dapat memberi tanggapan terkait informasi yang diterimanya. Berbeda dengan media lain seperti televisi atau radio yang berlangsung secara satu arah, di mana pengguna hanya dapat menerima informasi tanpa dapat memberi tanggapan. Bentuk tanggapan yang dapat diberikan oleh pengguna melalui media sosial adalah pesan berbasis teks melalui kolom komentar. Hampir seluruh produk media sosial selalu memberikan kolom komentar atau pesan singkat agar para pengguna dapat menuliskan tanggapannya terkait informasi yang dipublikasikan. Komentar tersebut bersifat publik, sehingga seluruh masyarakat dapat melihat respon dari orang-orang yang memberikan komentar terhadap sebuah konten.

Komentar yang diberikan oleh seseorang terhadap sebuah konten di media sosial dapat dikategorikan ke dalam sebuah jenis sentimen. Pada umumnya, jenis sentimen dibagi menjadi tiga jenis yaitu sentimen positif, negatif, dan netral. Namun tidak menutup kemungkinan terdapat beberapa jenis lain dalam pengelompokan sentimen, misalnya sentimen berbasis emosi seperti marah, sedih, kecewa, dan sejenisnya. Saat ini, sentimen pada sebuah komentar tidak hanya ditentukan melalui tulisan (teks) saja, namun juga melalui *emoticon* atau *emoji*. Penggunaan *emoticon* saat ini menjadi cukup masif dikarenakan semakin beragamnya varian *emoticon* yang disediakan oleh penyedia layanan. Selain itu, saat ini *emoticon* telah memiliki standar yang sama antara satu *platform* dengan *platform* yang lain, sehingga *emoticon* yang dituliskan oleh perangkat dengan merk A dapat ditampilkan oleh perangkat dengan merk B. Berbeda dengan versi terdahulu, di mana setiap *vendor* memiliki standar kode *emoticon* masing-masing, sehingga *emoticon* yang dituliskan melalui sebuah perangkat belum tentu dapat ditampilkan pada perangkat lain dari *vendor* yang berbeda. Dengan demikian, terdapat banyak pengguna media sosial yang

mengekspresikan pesan yang disampaikan dengan menggunakan *emoticon* atau sebagai pendukung dari pesan teks yang dituliskan.

Setiap *emoticon* memiliki karakteristik masing-masing yang dapat mewakili satu atau lebih sentimen. Terdapat berbagai macam *emoticon* yang dapat digunakan untuk menggambarkan perasaan bahagia, sedih, marah, kecewa, dan sejenisnya. Penelitian ini mencoba untuk menganalisa jenis sentimen dari sebuah komentar berdasarkan *emoticon* yang digunakan dalam pesan tersebut. Pada umumnya sentimen ditentukan melalui teks yang terdapat pada komentar, namun seringkali pesan berbasis teks sulit untuk diproses karena berbagai macam kendala seperti adanya penggunaan singkatan, bahasa slang, dan kesalahan penulisan (*typo*) yang mengakibatkan kata-kata dalam pesan tersebut tidak dapat dikenali oleh sistem, sehingga sulit untuk diproses lebih lanjut. Berbeda dengan penggunaan *emoticon* yang tidak memiliki kendala seperti pada pemrosesan teks tersebut, di mana satu-satunya kesalahan yang mungkin terjadi dalam penggunaan *emoticon* adalah kesalahan dalam memilih *emoticon* sesuai dengan yang dikehendaki.

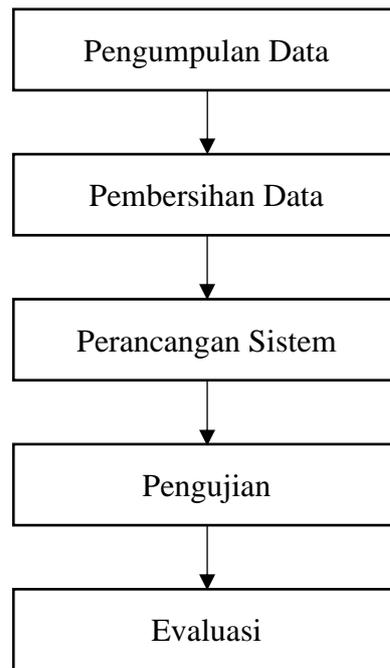
Pada dasarnya, *emoticon* dibentuk menggunakan karakter biasa dengan susunan tertentu [2]. Susunan karakter tersebut disepakati untuk menjadi kode *unicode* yang mewakili gambar tertentu yang kita sebut sebagai *emoticon* atau *emoji*. Karena *emoticon* disusun menggunakan karakter biasa seperti halnya pada teks, maka *emoticon* juga dapat diproses menggunakan pemrosesan teks pada umumnya.

Terdapat berbagai macam metode yang dapat digunakan untuk keperluan pemrosesan teks. Salah satu metode yang cukup sering digunakan untuk pemrosesan teks adalah *K-Nearest Neighbor (K-NN)* [3]. Metode *K-NN* banyak digunakan pada berbagai jenis objek seperti teks [4] dan citra [5, 6] karena proses komputasinya yang cukup ringan untuk diimplementasikan. *K-NN* memiliki algoritma dan proses komputasi yang cukup sederhana, sehingga relatif mudah untuk diimplementasikan serta tidak membebani komputer atau *server* [7]. Salah satu contohnya adalah algoritma *K-NN* yang digunakan untuk melakukan analisis sentimen terhadap komentar-komentar pada media sosial Instagram, khususnya untuk akun-akun *brand* elektronik [8]. Namun salah satu kekurangan dari metode *K-NN* adalah satu jenis data latih hanya dapat diasosiasikan terhadap satu kelas saja, sehingga kurang cocok untuk diimplementasikan untuk permasalahan dengan jenis data latih sedikit seperti *emoticon* yang terbatas pada beberapa jenis saja. Terdapat metode lain yang mampu mengatasi kekurangan tersebut, yaitu metode *Naïve Bayes* [9]. Metode *Naïve Bayes* merupakan pengembangan dari metode *Bayes* [10], di mana pada *Naïve Bayes* seluruh atribut yang ada dianggap independen atau tidak ada kaitan satu sama lain [11]. Metode *Naïve Bayes* banyak digunakan untuk keperluan pemrosesan teks dan dapat memberikan hasil yang cukup baik, salah satunya adalah untuk klasifikasi topik dengan hasil nilai akurasi sebesar 91.8% [12]. *Naïve Bayes* bekerja dengan menggunakan sistem probabilitas, di mana satu jenis data latih dapat dikategorikan terhadap beberapa kelas yang berbeda dengan nilai probabilitas untuk tiap kelas. Karena satu jenis data dapat dikategorikan terhadap lebih dari satu kelas, untuk permasalahan dengan jumlah jenis data latih yang sedikit dapat diatasi dengan menggunakan metode *Naïve Bayes*. Metode *Naïve Bayes* telah banyak digunakan untuk menyelesaikan berbagai jenis permasalahan seperti klasifikasi sinyal [13] dan deteksi lokasi pengguna gawai [14], di mana pada kedua penelitian tersebut metode *Naïve Bayes* dapat memberikan hasil nilai akurasi yang cukup baik.

Berdasarkan hasil penelitian terdahulu yang cukup baik menggunakan metode *Naïve Bayes*, maka penelitian ini akan menggunakan metode *Naïve Bayes* sebagai algoritma utama dalam menentukan sentimen yang dimiliki oleh sebuah pesan. Dengan menggunakan metode *Naïve Bayes*, maka akan memungkinkan untuk satu jenis *emoticon* dikategorikan terhadap lebih dari satu jenis sentimen yang ditentukan melalui nilai probabilitas di dalamnya. Maka banyaknya varian *emoticon* yang digunakan dalam sebuah pesan akan menjadi salah satu faktor penentu dalam menentukan jenis sentimen terkait dengan pesan atau komentar di mana *emoticon* tersebut berada. Hasil penelitian ini diharapkan dapat memberikan kontribusi pada penelitian lain terkait pengolahan teks, khususnya untuk teks berbasis *emoticon*.

## II. METODE PENELITIAN

Penelitian ini dilakukan dalam beberapa tahapan yaitu pengumpulan data, pembersihan data, pengujian, dan evaluasi hasil pengujian seperti yang ditunjukkan pada gambar 1.



Gambar 1. Blok Diagram Metode Penelitian

#### A. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengambil data komentar dari akun Instagram media massa berbahasa Indonesia seperti @kompascom, @detikcom, dan @tempodotco. Akun dengan jenis media massa dipilih karena di dalamnya terdapat *post* dengan berbagai variasi topik seperti politik, bencana, olah raga, dan lain-lain. Sehingga reaksi komentar yang akan diberikan oleh para pengguna pun juga akan terdiri dari berbagai macam sentimen. Dari masing-masing akun Instagram tersebut akan diambil data komentar dari 20 *post* terakhir terhitung sejak 15 September 2021. Dari proses pengumpulan data tersebut, didapatkan sejumlah 8540 data komentar yang bersifat mentah (*raw*).

#### B. Pembersihan Data

Pada tahap berikutnya, dari seluruh komentar yang berhasil didapatkan akan disaring (*filtering*) kembali untuk mendapatkan data komentar yang mengandung *emoticon* di dalamnya. Proses pendeteksian *emoticon* dilakukan dengan menggunakan modul *emoji* yang terdapat pada Python, di mana pengaturan daftar *emoticon* yang digunakan adalah *emoticon* standar dalam Bahasa Inggris. Berdasarkan hasil penyaringan tersebut, didapatkan 987 data komentar yang mengandung *emoticon* di dalamnya. Setiap data komentar akan diberi label terkait dengan jenis sentimen yang terdapat pada komentar tersebut.

#### C. Perancangan Sistem

Perancangan sistem diawali dengan membagi seluruh data yang telah disaring menjadi dua kelompok, yaitu data latih dan data uji. Pembentukan data latih dilakukan dengan mengambil sebanyak 500 data komentar secara acak, kemudian untuk seluruh jenis *emoticon* dari masing-masing komentar akan dihitung probabilitas sentimennya. Misalnya jika *emoticon* X terdapat pada dua komentar dengan label marah dan satu komentar pada label sedih, maka probabilitas *emoticon* X terhadap label marah adalah 66.67% sedangkan probabilitasnya terhadap label sedih adalah 33.33%. Kemudian seluruh sisa data di luar 100 data yang dijadikan data latih akan digunakan sebagai data uji. Jenis *emoticon* yang digunakan pada data latih adalah *emoticon* dengan jumlah kemunculan sama dengan atau lebih dari 5 kali di seluruh komentar, dengan asumsi bahwa *emoticon* yang jumlah kemunculannya kurang dari 5 kali adalah *emoticon* yang tidak umum atau jarang digunakan. Daftar *emoticon* dengan jumlah kemunculan minimal 5 kali ditunjukkan pada tabel 1.

TABEL 1  
DAFTAR EMOTICON DENGAN JUMLAH KEMUNCULAN  $\geq 5$

No	Emoticon	Unicode	Jumlah
1	😄	u+1f602	250
2	😬	u+1f622	33
3	🔥	u+1f525	31
4	❤️	u+2764	21
5	😜	u+1f62e	20
6	😘	u+1f60d	18
7	👉	u+1f44f	16
8	🙏	u+1f64f	12
9	🤔	u+1f923	11
10	😁	u+1f601	10
11	👏	u+1f64c	8
12	👍	u+1f44d	7
13	🤪	u+1f62d	5
14	😏	u+1f914	5
15	😃	u+1f60a	5
16	😂	u+1f605	5

Berdasarkan pengamatan yang telah dilakukan, komentar yang terdapat pada akun media massa hanya terdiri dari komentar positif dan negatif saja, hampir tidak ditemukan adanya komentar yang bersifat netral. Dengan demikian, kelas sentimen yang digunakan pada penelitian ini hanya sentimen positif dan negatif. Seluruh komentar yang ada akan diproses apa adanya tanpa melihat konteks, sehingga komentar-komentar bernada menyindir atau sarkas tetap dapat masuk ke dalam kelas positif meskipun tujuan asli dari komentar tersebut berlawanan dengan isi *post* dari akun media massa. Beberapa contoh pelabelan sentimen berdasarkan isi komentar dapat dilihat pada tabel 2.

TABEL 2  
CONTOH DATASET

No	Komentar	Sentimen
1	😄😄😄🙏	positif
2	Metode mempersulit hidup 😬	negatif
3	Bagus 👍	positif
4	😘👍👍👍👍😘	positif
5	Seriusan? 😬	negatif
6	😄😄😄...lucu	positif
7	Maaf 🙏🙏🙏	negatif
8	❤️👏🔥 keren	positif
9	🤪🔥 bikin jengkel	negatif
10	😬😬 kok gitu	negatif

Pembentukan *dataset* dilakukan dengan menggunakan sistem *binary (true atau false)* dan mengabaikan jumlah kemunculan *emoticon* dengan jenis yang sama dalam satu komentar, artinya apabila sebuah *emoticon* muncul lebih dari satu kali dalam satu komentar maka nilainya tetap dianggap sama seperti muncul satu kali saja. Penentuan nilai *true* atau *false* dilakukan dengan melibatkan tiga orang *annotator*. *Annotator* berjumlah ganjil dengan pertimbangan agar jumlah *voting* tidak mungkin sama. Seluruh karakter *non-emoticon* akan diabaikan atau dihapus, sehingga tidak akan diproses lebih lanjut oleh sistem. Karena data yang diolah tidak mengandung karakter selain *emoticon*, maka pada penelitian ini tidak diperlukan adanya tahap pra-proses (*pre-processing*) yang lengkap seperti halnya pada penelitian mengenai pengolahan teks secara umum, seperti tahap *stemming* dan *stopwording*. Tahap pra-proses yang diimplementasikan pada penelitian ini adalah tokenisasi, yaitu memecah pesan atau dokumen berdasarkan spasi menjadi *token-token* yang terdiri dari satu kata, dalam penelitian ini adalah karakter *emoticon* itu sendiri. Kemudian, algoritma Naïve Bayes menggunakan persamaan seperti yang ditunjukkan pada persamaan (1).

$$P(C|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Keterangan:

$P(c|x)$  = probabilitas *posterior* dari kelas c saat diberikan atribut x

$P(c)$  = probabilitas *prior* dari class c

$P(x|c)$  = probabilitas dari atribut x saat diberikan kelas c

$P(x)$  = probabilitas *prior* dari atribut x

Berdasarkan persamaan (1), maka harus dihitung terlebih dahulu frekuensi masing-masing *emoticon* sebelum algoritma Naïve Bayes dapat diimplementasikan. Untuk setiap jenis *emoticon*, dapat dibentuk tabel frekuensi seperti ditunjukkan pada tabel 3.

TABEL 3  
CONTOH TABEL FREKUENSI

Emoticon	Sentimen	
	Positif	Negatif
😄	True	223
	False	105
		328
		172

Dari tabel frekuensi seperti pada contoh tabel 3, maka dapat dibentuk tabel *likelihood* yang menunjukkan probabilitas sentimen positif dan negatif apabila terdapat atau tidaknya emoticon yang bersangkutan. Contoh tabel *likelihood* dari tabel 3 dapat dilihat pada tabel 4.

TABEL 4  
CONTOH TABEL LIKELIHOOD

Emoticon	Sentimen		P(x)
	P(x   Positif)	P(x   Negatif)	
😄	True	67.99%	47.8%
	False	32.01%	52.2%

#### D. Pengujian Sistem

Tahap berikutnya adalah melakukan pengujian dengan menggunakan 487 data komentar yang tidak disertakan ke dalam data latih. Dari seluruh data komentar yang diujikan akan diukur tingkat akurasi dengan cara mencocokkan apakah kelas sentimen yang diberikan oleh sistem sesuai dengan konteks dari komentar tersebut. Jika terdapat beberapa *emoticon* dalam satu komentar, maka nilai probabilitas positif atau negatif dihasilkan berdasarkan perkalian antara probabilitas setiap *emoticon*. Seluruh proses pembentukan data latih hingga pengujian dilakukan dengan menggunakan bahasa pemrograman Python dengan beberapa tambahan pustaka yang diperlukan.

#### E. Evaluasi

Evaluasi terhadap sistem dilakukan dengan menghitung nilai akurasi yang dihasilkan oleh sistem untuk data uji yang diberikan, di mana setiap data uji memiliki dua kemungkinan luaran yaitu positif dan negatif. Apabila luaran data uji sesuai dengan target yang diberikan, maka data tersebut akan diberi label *true*, sedangkan data uji yang tidak sesuai dengan target akan diberi label *false*. Dengan demikian, seluruh nilai yang dihasilkan dapat dipresentasikan dalam bentuk *confusion matrix*.

### III. HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan menggunakan data uji yang terdiri dari 487 data komentar. Pada setiap data uji akan diambil seluruh *emoticon* yang terdapat di dalamnya kemudian diolah lebih lanjut menggunakan metode Naive Bayes. Apabila hasil perhitungan Naive Bayes sesuai dengan target yang ditentukan (positif atau negatif), maka nilai akurasi akan dihitung sebagai benar atau *true*. Sebaliknya jika hasil tidak sesuai target, maka akan dihitung sebagai salah atau *false*. Berdasarkan hasil pengujian tersebut, didapatkan hasil dalam bentuk *confusion matrix* seperti yang ditunjukkan pada tabel 5.

TABEL 5  
HASIL PENGUJIAN DATA KOMENTAR

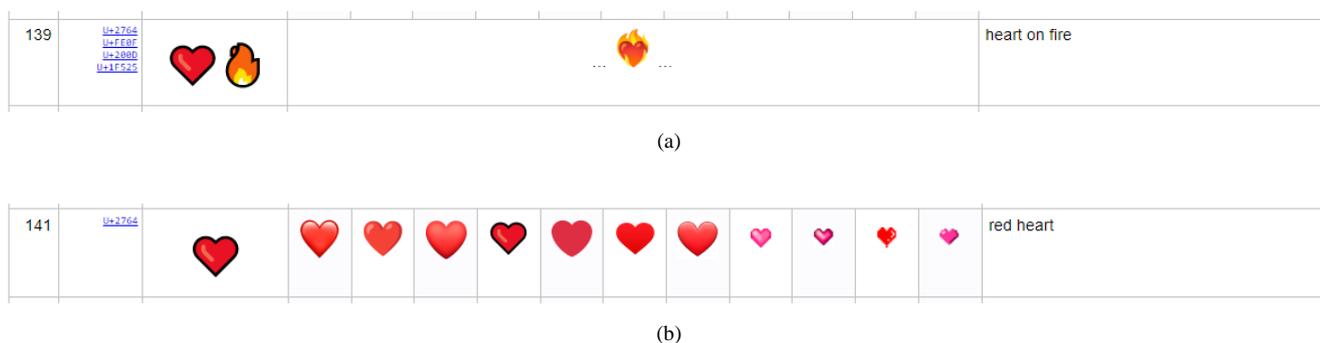
Target	Hasil Sistem		Total
	Positif	Negatif	
Positif	303	13	316
Negatif	5	166	171
	308	179	487

Dari hasil pengujian yang ditunjukkan pada tabel 5, maka dapat diperoleh tingkat akurasi dari sistem sebesar 96.3% yang diperoleh dari penjumlahan antara nilai target dan hasil sistem yang sama, dibagi dengan jumlah keseluruhan data seperti yang ditunjukkan pada perhitungan berikut:

$$akurasi = \frac{303 + 166}{487} = 0.963 = 96.3\%$$

Secara umum, metode Naïve Bayes mampu memberikan hasil yang paling akurat saat hanya terdapat satu jenis *emoticon* pada sebuah komentar. Semakin banyak jenis *emoticon* yang digunakan pada sebuah komentar, maka kemungkinan penurunan nilai akurasi akan semakin besar, terlebih apabila variasi *emoticon* yang digunakan tidak senada atau saling berlawanan. Pada beberapa data uji yang ditunjukkan pada tabel 5, terdapat kesalahan pelabelan oleh sistem yang disebabkan oleh beberapa faktor. Salah satunya adalah terdapat *emoticon* yang sekilas terlihat serupa, namun sebetulnya terdapat sedikit perbedaan. Contohnya adalah *emoticon* 🤪 (U+1F602) dan 😊 (U+1F604). Pada data latih yang ditanamkan pada sistem, *emoticon* yang banyak digunakan adalah *emoticon* dengan kode *unicode* U+1F602. Sedangkan pada data uji, terdapat beberapa komentar yang menggunakan *emoticon* dengan kode *unicode* U+1F604, di mana keduanya sebetulnya memiliki makna yang serupa yaitu ekspresi tertawa.

Permasalahan lain yang dijumpai oleh sistem adalah adanya dua *emoticon* dengan bentuk dan tampilan yang sama persis, namun sebetulnya memiliki kode *unicode* yang berbeda, salah satunya adalah *emoticon* berbentuk hati ❤️. Pada beberapa perangkat atau *platform*, *emoticon* tersebut memiliki kode U+2764. Namun pada sebagian perangkat atau *platform* yang lain, *emoticon* tersebut terdiri dari dua buah kode *unicode* U+2764U+fe0f. Berdasarkan tabel referensi yang terdapat pada laman website <https://unicode.org/> seperti yang ditunjukkan pada gambar 2, kedua kode tersebut seharusnya ditampilkan dalam wujud yang berbeda, namun sebagian besar perangkat atau *platform* tidak dapat mengakomodasinya sehingga keduanya dianggap *emoticon* yang sama.



Gambar 2. (a) *emoticon* kode U+2764U+fe0f. (b) *emoticon* kode U+2764.

Sumber: Unicode.org [15]

#### IV. SIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, secara keseluruhan metode Naïve Bayes mampu memberikan luaran label sentimen dengan tingkat akurasi yang cukup baik yaitu 96.3%. Terdapat beberapa kendala dalam melakukan analisis terhadap *emoticon*, yaitu adanya dua atau lebih *emoticon* dengan tampilan serupa namun memiliki kode *unicode* yang berbeda sehingga oleh sistem dianggap sebagai dua *emoticon* yang berbeda. Permasalahan lain adalah adanya dua atau lebih *emoticon* yang memiliki tampilan hampir serupa, namun terdapat sedikit perbedaan sehingga banyak digunakan secara bergantian, sementara oleh sistem keduanya tetap diproses menggunakan data latih masing-masing. Pada penelitian berikutnya, dapat ditambahkan metode pra-proses untuk melakukan standarisasi *emoticon*, dengan demikian *emoticon-emoticon* dengan tampilan dan fungsi yang serupa dapat dianggap sebagai satu entitas, sehingga dapat dikategorikan ke dalam satu kelompok yang sama.

#### UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada Fakultas Teknologi Informasi Universitas Kristen Duta Wacana yang telah memberikan dukungan selama penelitian ini berlangsung.

#### DAFTAR PUSTAKA

- [1] H. Nurhayati-Wolff, "Number of social media users in Indonesia from 2017 to 2020 with forecasts until 2026," Statista, 16 Agustus 2021. [Online]. Available: <https://www.statista.com/statistics/247938/number-of-social-network-users-in-indonesia/>. [Accessed 22 September 2021].
- [2] B. Qiyu, D. Qi, M. Zhe and Y. Maokun, "A Systematic Review of Emoji: Current Research and Future Perspectives," *Frontiers in Psychology*, vol. 10, p. 2221, 2019.
- [3] K. A. Nugraha and Herlina, "Klasifikasi Pertanyaan Bidang Akademik Berdasarkan 5W1H menggunakan K-Nearest Neighbors," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 7, no. 1, pp. 44-51, 2021.
- [4] K. A. Nugraha and D. Sebastian, "Chatbot Layanan Akademik Menggunakan K-Nearest Neighbor," *Jurnal Sains dan Informatika (JSI)*, vol. 7, no. 1, pp. 11-19, 2021.
- [5] K. A. Nugraha, "Deteksi Area Parkir Mobil Berbasis Marker Menggunakan Moment Invariants dan K-NN," *Jurnal Teknik Informatika dan Sistem Informasi (JuTISI)*, vol. 5, no. 1, pp. 112-121, 2019.
- [6] L. A. Sunjoyo, R. G. Santosa and K. A. Nugraha, "Implementasi transformasi haar wavelet untuk deteksi citra jeruk nipis yang busuk," *Informatika: Jurnal Teknologi Komputer dan Informatika*, vol. 12, no. 2, pp. 165-173, 2016.
- [7] X. Ding, S. Sheng and J. Liu, "Efficient Probabilistic K-NN Computation in Uncertain Sensor Networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2575 - 2587, 2021.
- [8] K. A. Nugraha, "Pembentukan Dataset Token Sentimen Berdasarkan Akun Instagram Brand Elektronik Menggunakan K-Nearest Neighbors," *Jurnal Buana Informatika (JBI)*, vol. 12, no. 1, pp. 68-77, 2021.
- [9] S. Wang, J. Ren and R. Bai, "A Regularized Attribute Weighting Framework for Naïve Bayes," *IEEE Access*, vol. 8, pp. 225639 - 225649, 2020.
- [10] L. A. Dalton, M. E. Benalcázar, M. Brun and E. R. Dougherty, "Analytic Representation of Bayes Labeling and Bayes Clustering Operators for Random Labeled Point Processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1605 - 1620, 2015.
- [11] L. Yu, S. Gan, Y. Chen and M. He, "Correlation-Based Weight Adjusted Naive Bayes," *IEEE Access*, vol. 8, pp. 51377 - 51387, 2020.
- [12] M. A. Rahman and Y. A. Akter, "Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naïve Bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, 2019.
- [13] M. O. Mughal and S. Kim, "Signal Classification and Jamming Detection in Wide-Band Radios Using Naïve Bayes Classifier," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1398 - 1401, 2018.
- [14] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan and Y. Xiang, "Passive Indoor Localization Based on CSI and Naive Bayes Classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1566 - 1577, 2018.
- [15] NN, "Unicode," Unicode, [Online]. Available: <https://unicode.org/emoji/charts/full-emoji-list.html>. [Accessed 10 Agustus 2021].