

ANALISIS SENTIMEN PADA ULASAN APLIKASI
PEDULILINDUNG DI GOOGLE PLAY STORE DENGAN METODE
SMOTE-MULTINOMIAL NAIVE BAYES

Proposal TA
Program Studi Matematika



Diajukan oleh
IMAM SUYUTI
M0119043

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET SURAKARTA

2022

Proposal TA

**ANALISIS SENTIMEN PADA ULASAN APLIKASI
PEDULILINDUNGI DI GOOGLE PLAY STORE DENGAN METODE
*SMOTE-MULTINOMIAL NAIVE BAYES***

diajukan oleh
IMAM SUYUTI
M0119043

telah dipertahankan di depan Dewan Penguji
pada hari xxx
dan dinyatakan telah disetujui.

| Anggota Tim Penguji | Tanda Tangan |
|---------------------|--------------|
| 1. xxx NIP. xxx | 1. |
| 2. xxx NIP. xxx | 2. |
| 3. xxx NIP. xxx | 3. |

Surakarta, ... Oktober 2022
Ketua Komisi TA,

Dr. Dra. Diari Indriati, M.Si.
NIP. 19610112 198811 2 001

1. LATAR BELAKANG MASALAH

Pertambahan penyebaran virus COVID-19 (Coronavirus Disease 2019) mendorong pemerintah Indonesia untuk mengeluarkan suatu kebijakan pengendalian penyebaran virus tersebut dengan memberlakukan pembatasan berskala besar yang sering dikenal dengan istilah Pembatasan Sosial Berskala Besar (PSBB). PSBB ini bertujuan untuk memutus mata rantai penyebaran virus corona meskipun banyak fasilitas umum yang ditutup. Akan tetapi beberapa sektor vital seperti fasilitas pemerintahan, kesehatan, dan pasar atau minimarket tetap buka selama PSBB dengan tetap menjalankan protokol kesehatan. Kebijakan tersebut tentunya didasari pada Undang-Undang No. 6 Tahun 2018 tentang Kekarantinaan Kesehatan.

Dalam rangka mencegah penyebaran wabah Covid-19 di Indonesia, PT Telekomunikasi Indonesia Tbk (Telkom) dan Kementerian Komunikasi dan Informatika (Kominfo) telah bekerjasama membuat aplikasi PeduliLindungi. Aplikasi ini dikembangkan pemerintah agar dapat membantu pelacakan dalam rangka menghentikan penyebaran virus Covid-19. Aplikasi ini juga dapat digunakan dalam pelaksanaan surveilans kesehatan dalam menangani penyebaran Covid-19, dengan mengadakan Tracing yaitu melakukan pelacakan terhadap orang-orang yang berkontak dengan orang-orang yang diduga terinfeksi Covid-19. Selain itu juga Tracking yaitu melacak persebaran virus corona dengan melihat siapa saja yang telah bertemu dengan penderita virus Covid-19 dan menyelenggarakan Warning and Fencing yaitu adanya peringatan dan pengawasan dengan membatasi pergerakan seseorang yang sedang dalam karantina atau isolasi [16]. Penggunaan aplikasi ini dilakukan hanya pada masa darurat Covid-19. Partisipasi masyarakat sangat dibutuhkan dengan membagikan data lokasinya saat bepergian agar dapat dilakukan penelusuran riwayat kontak dengan penderita Covid-19. Karena aplikasi PeduliLindungi ditujukan guna melindungi masyarakat yang tengah mengakses fasilitas publik agar memastikan kegiatan yang tengah dilakukan aman dan dapat menghindari dari adanya penyebaran Covid-19 [11].

Di setiap aplikasi, juga terdapat rating dan ulasan pengguna mengenai pelayanan dan fitur-fitur yang diberikan. Ulasan yang diberikan bisa berupa saran, kritik, maupun keluhan. Hal tersebut menjadi sangat bermanfaat dan membantu bagi

pengguna lain yang akan menggunakan aplikasi tersebut. Pengumpulan dan penyortiran data ulasan tidaklah hal yang mudah karena banyaknya ulasan yang tersedia pada situs Google Play pada fitur komentar biasanya sangat banyak. Menurut Moraes et al. [14], metode yang cocok untuk mengumpulkan data informasi tersebut adalah penggunaan metode web scrapping. Akan tetapi, belum ditemukan adanya fitur analisis sentimen untuk mengelompokkan atau menyaring antara ulasan positif dan negatif, mengingat jumlah ulasan aplikasi yang sangat banyak dan akan terus bertambah setiap saat.

Pada penelitian ini, akan menganalisis sentimen ulasan tentang aplikasi PeduliLindungi dengan menggunakan metode *Multinomial Naïve Bayes*. Kemudian membandingkan hasil kinerjanya dengan metode klasifikasi yang disertai SMOTE untuk mendapatkan hasil yang terbaik. Selain itu, penelitian ini berguna untuk mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif pada ulasan aplikasi PeduliLindungi.

2. PERUMUSAN MASALAH

Berdasarkan latar belakang masalah dirumuskan tiga masalah, yaitu

- (1) bagaimana menganalisis sentimen ulasan tentang aplikasi PeduliLindungi dengan menggunakan metode *Multinomial Naïve Bayes*,
- (2) bagaimana perbandingan hasil kinerjanya dengan metode klasifikasi yang disertai teknik SMOTE untuk mendapatkan hasil yang terbaik, dan
- (3) bagaimana mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif pada ulasan tentang aplikasi PeduliLindungi.

3. TUJUAN PENELITIAN

Berdasarkan perumusan masalah diperoleh tiga tujuan, yaitu

- (1) menganalisis sentimen ulasan tentang aplikasi PeduliLindungi dengan menggunakan metode *Multinomial Naïve Bayes*,
- (2) perbandingan hasil kinerjanya dengan metode klasifikasi yang disertai teknik SMOTE untuk mendapatkan hasil yang terbaik, dan

- (3) mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif pada ulasan tentang aplikasi PeduliLindungi.

4. MANFAAT PENELITIAN

Manfaat penelitian ini yaitu diharapkan menambah wawasan ilmu pengetahuan tentang analisis sentimen pada ulasan suatu aplikasi dan informasi yang diperoleh dapat digunakan oleh pihak developer aplikasi PeduliLindungi sebagai bahan evaluasi dan pengguna yang akan mengunduh aplikasi maupun pihak lain yang ingin melakukan penelitian lanjutan.

5. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Pintoko dan Muslim [19] pada tahun 2018 menjelaskan bahwa metode *naïve bayes* dapat menganalisis sentimen pada data ulasan jasa transportasi *online* pada twitter. Pada tahun 2019, Andika, Azizah, dan Respatiwan [1] menjelaskan bahwa dengan metode *naïve bayes* dapat menganalisis sentimen pada hasil *quick count* pemilihan Presiden Indonesia 2019 di twitter serta Apriani dan Gustian [2] menjelaskan bahwa algoritma *naïve bayes* dapat menganalisis sentimen pada data ulasan aplikasi Tokopedia. Kemudian Ardianto [3] pada tahun 2020 menjelaskan bahwa metode *naïve bayes* dapat menganalisis sentimen pada kurikulum untuk pendidikan *e-sports* di media sosial twitter. Daulay dan Asror [7] pada tahun 2020 menjelaskan bahwa metode *naïve bayes* dapat menganalisis sentimen pada data ulasan aplikasi Google Play Store.

Penelitian ini merupakan bentuk pengembangan penelitian dengan mengganti objek pengamatan pada data ulasan aplikasi PeduliLindungi dengan menggunakan metode klasifikasi *supervised learning* yaitu *Multinomial Naïve Bayes*. Kemudian membandingkan hasil kinerjanya dengan metode klasifikasi yang disertai teknik SMOTE untuk mendapatkan hasil yang terbaik. Selain itu, penelitian ini berguna untuk mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif.

6. LANDASAN TEORI

Berikut dijelaskan teori yang melandasi penyelesaian perumusan masalah, yaitu analisis sentimen, *web scrapping*, *data preprocessing*, pembobotan TF-IDF , Vader, SMOTE, *multinomial naïve bayes*, evaluasi sistem klasifikasi, serta *wordcloud*.

6.1. Analisis Sentimen.

Analisis sentimen merupakan salah satu bidang pada *text minning* yang menganalisa sebuah pendapat, opini, evaluasi, sentimen, sikap atau penilaian seseorang terhadap individu, kelompok, produk, organisasi, masalah, peristiwa atau topik [20]. Analisis sentimen juga bisa diartikan sebagai riset komputasional dari sebuah opini dan emosi yang diekspresikan secara tekstual. Analisis sentimen biasanya digunakan untuk menganalisa produk atau organisasi dalam rangka peningkatan kualitas dari produk atau organisasi nantinya [9].

Analisis sentimen terbagi menjadi dua kategori yaitu *Coarse-Grained Sentiment Analysis* dan *Fined-Grained Sentiment Analysis* [20].

6.1.1. *Coarse-Grained sentiment analysis*.

Coarse-Grained sentiment analysis adalah klasifikasi yang berorientasi pada sebuah dokumen secara keseluruhan. Klasifikasi jenis ini dibagi pada tiga yaitu positif, netral, dan negatif.

6.1.2. *Fine-Grained sentiment analysis*.

Fine-Grained sentiment analysis adalah klasifikasi yang orientasinya lebih spesifik, yaitu pada kalimat di sebuah dokumen. Contoh dari *fined-grained sentiment analysis* adalah "Saya benci orang itu, dia suka pamer di depan guru" (Negatif), atau bisa juga "Jalanan hari ini terasa nyaman karena tidak ada kemacetan" (Positif).

6.2. *Web Scrapping*.

Web Scrapping adalah metode yang digunakan untuk mengumpulkan atau mengekstraksi informasi atau data semi-terstruktur dari website, biasanya dalam bentuk halaman web dalam bahasa markup, seperti HTML atau XHTML, kemudian mengestrak informasi atau data tertentu dari halaman tersebut untuk dianalisis [21].

Proses *web scrapping* dibagi menjadi 4 tahap yaitu mengunduh *content* dari halaman *web*, ekstrak data, menyimpan data dalam format *csv* atau *json*, dan terakhir analisis data. Salah satu bahasa yang digunakan untuk proses *web scrapping* adalah *python* karena memiliki *libraries* yang memproses data dengan baik.

6.3. **Data Preprocessing.**

Data Preprocessing merupakan tahapan yang dilakukan setelah *dataset* terkumpul untuk membersihkan data, sehingga proses pada *machine learning* menjadi lebih cepat dan akurat [17]. Tujuan dari *data Preprocessing* yaitu mengubah data teks yang awalnya tidak terstruktur menjadi data yang terstruktur. Secara umum proses tahapan *data preprocessing* dapat dilakukan sebagai berikut.

1. *Case Folding* adalah proses standarisasi bentuk huruf agar tidak ada perbedaan makna.
2. *Tokenizing* adalah proses pemisahan kata per kata yang tidak saling mempengaruhi dari teks dokumen.
3. *Filtering* adalah proses penyaringan atau pemilihan kata dalam dokumen.
4. *Stemming* adalah pengubahan kata berimbuhan menjadi kata dasar.

6.4. **Term Frequency-Inverse Document Frequency (TF-IDF).**

Pembobotan TF-IDF merupakan salah satu model yang digunakan untuk menghitung bobot kata pada suatu dokumen dengan menggunakan model *term frequency* untuk menghitung kata yang muncul di setiap dokumen [15]. Pada proses ini, data perlu dimodelkan agar data yang masih berbentuk kata dapat diproses dan dihitung dengan cara mengubah data menjadi vektor.

$$W = TF \times IDF = TF \times \frac{1}{DF}$$

Frekuensi kata T pada dokumen D merupakan TF yang digunakan untuk menghitung kemampuan kata mendeskripsikan dokumen. Sedangkan IDF mempresentasikan frekuensi dokumen D yang mengandung kata T dalam corpus, yang digunakan untuk menghitung kemampuan kata membedakan dokumen tersebut. Jika frekuensi suatu kata tinggi dalam dokumennya sendiri tetapi rendah dalam dokumen lain,

kata ini memiliki kemampuan yang kuat untuk membedakan dari dokumen lain dan ditetapkan ke bobot yang tinggi.

6.5. **Vader.**

Valence Aware Dictionary for Social Reasoning (Vader) merupakan model yang digunakan dalam menganalisis sentimen dan mampu menentukan keragaman data melalui intensitas kekuatan emosional yang ada sesuai dengan kamus data lexicon yang tersedia [8]. Salah satu contoh dari metode sentimen analisis secara leksikal yaitu *vader lexicon polarity detection* [12]. Metode leksikal tidak memerlukan *data training* atau data yang telah dilabeli namun sudah tersedia dalam kamus lengkap dengan kepolaritasan sentimennya.

6.6. ***Synthetic Minority Oversampling Technique* (SMOTE).**

SMOTE merupakan salah satu pendekatan teknik *oversampling* yang efektif dan baik digunakan untuk menangani *imbalanced dataset* karena dapat mengatasi *overfitting* pada saat proses *oversampling* untuk kelas minoritas. Pendekatan SMOTE dilakukan dengan cara membangkitkan data buatan (*synthetic*) berdasarkan tetangga terdekat (*k-nearest neighbors*) sehingga diharapkan pendekatan teknik tersebut dapat berimbas pada hasil performa klasifikasi [6].

6.7. ***Multinomial Naïve Bayes*.**

Multinomial Naïve Bayes merupakan metode *supervised learning* yang menggunakan probabilitas dan lebih difokuskan untuk klasifikasi teks [13]. *Multinomial Naïve Bayes* juga memiliki fitur unik, yaitu hasil yang didapat untuk masing-masing kelas bersifat independen. Hal ini berarti, dari dokumen satu ke dokumen berikutnya tidak ada keterkaitannya sama sekali sehingga hasil yang didapat murni dari dokumen yang diolah itu sendiri. Perhitungan probabilitas ulasan d yang memiliki kelas c dapat dilihat pada rumus berikut.

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_i|c)$$

Keterangan:

$P(c|d)$: probabilitas suatu kelas c pada dokumen/teks d

$P(c)$: probabilitas prior c

$P(w_i|c)$: probabilitas suatu kata pada kelas c .

6.8. Evaluasi Sistem Klasifikasi.

Sebuah sistem klasifikasi harus dinilai performanya agar dapat mengukur tingkat akurasi dari prediksi klasifikasi yang dihasilkan. Ada beberapa metode perhitungan yang digunakan untuk menilai performa sebuah klasifikasi misalnya

6.8.1. *Confusion matrix*.

Confusion matrix adalah salah satu metode evaluasi dengan cara sebuah Tabel yang menyatakan berapa banyak data uji yang benar atau salah diklasifikasikan. Jika data positif dan diprediksi positif maka akan dihitung sebagai *true positive* dan jika data positif diprediksi negatif maka akan dihitung sebagai *false negative*. Pada data negatif jika diprediksi negatif akan dihitung sebagai *true negative* dan jika diprediksi positif maka akan dihitung sebagai *false positive*.

Tabel 1. *Confusion matrix*

| Aktual | Prediksi | |
|----------|----------------|----------------|
| | Positive | Negative |
| Positive | True Positive | False Negative |
| Negative | False Positive | True Negative |

6.8.2. *Accuracy*.

Accuracy adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Jika nilai akurasi tinggi maka sebuah sistem akan semakin bagus dalam melakukan prediksi, *accuracy* dapat dihitung dengan persamaan (6.1).

$$Accuracy = \frac{\text{Prediksi data benar}}{\text{Total data}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.1)$$

6.8.3. *Recall*.

Recall adalah salah satu perhitungan keakuratan prediksi yang digunakan sebagai ukuran tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi, *recall* dapat dihitung melalui persamaan (6.2).

$$Recall = \frac{\text{relevant item retrieved}}{\text{retrieved item}} = \frac{TP}{TP + FN} \quad (6.2)$$

6.8.4. *Precision*.

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban sistem, *precision* dapat dihitung dengan persamaan (6.3).

$$Precision = \frac{\text{relevant item retrieved}}{\text{retrieved item}} = \frac{TP}{TP + FP} \quad (6.3)$$

6.8.5. *Area Under Curve (AUC)*.

AUC merupakan kriteria evaluasi yang menggunakan sensitivitas atau spesifisitas sebagai dasar pengukuran [10]. Apabila terjadi kasus ketidakseimbangan data (*imbalance dataset*) maka dalam memilih model mana yang terbaik dapat dilakukan dengan menggunakan nilai AUC sebagai dasar pengukuran dan nilai akurasi dinilai kurang tepat di dalam *imbalance dataset* karena hanya mempelajari data mayoritas saja sehingga hasil yang didapatkan tidak memiliki informasi penting karena mungkin saja terjadi bias atau *overfitting*.

Nilai AUC pada umumnya berada pada interval 0,5–1,0, berikut ini tabel keterangan untuk masing-masing interval nilai AUC dapat dilihat pada Tabel 2 di bawah ini [4].

Tabel 2. Kategori Nilai AUC

| Nilai AUC | Deskripsi |
|-----------|------------------|
| 0,9 - 1,0 | <i>Excellent</i> |
| 0,8 - 0,9 | <i>Good</i> |
| 0,7 - 0,8 | <i>Fair</i> |
| 0,6 - 0,7 | <i>Poor</i> |
| 0,5 - 0,6 | <i>Failure</i> |

6.9. *Wordcloud*.

Wordcloud merupakan representasi dari data yang menunjukkan sekumpulan kata-kata penting dan sering muncul dalam kata. Kata-kata yang sering muncul ditandai dengan banyaknya jumlah kata dan dicetak dengan ukuran besar dalam wordcloud [5].

7. KERANGKA PEMIKIRAN

Berdasarkan tinjauan pustaka, dapat disusun kerangka pemikiran berikut. Permasalahan umum yang terjadi pada analisis sentimen yaitu adanya *imbalanced dataset*. *Imbalanced dataset* adalah data yang tidak seimbang dari segi jumlah tiap kelas individu. Metode *Multinomial Naïve Bayes* merupakan metode supervised learning yang menggunakan probabilitas dan lebih difokuskan untuk klasifikasi teks [13]. Metode *Multinomial Naïve Bayes* dapat diterapkan pada analisis sentimen pada ulasan aplikasi PeduliLindungi di Google Play Store. Metode yang digunakan untuk mengatasi *imbalanced dataset* adalah Metode SMOTE.

8. METODE PENELITIAN

Metode penelitian dibagi menjadi dua bagian yaitu data penelitian dan langkah penelitian. Penelitian ini merupakan penelitian teori dan terapan yakni teori tentang analisis sentimen dengan metode *Multinomial Naïve Bayes-SMOTE* dan menerapkan pada data ulasan aplikasi PeduliLindungi.

8.1. Data Penelitian. Data yang digunakan adalah data primer dari ulasan aplikasi PeduliLindungi di Google Play Store [18]. Variabel-variabel yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.

Tabel 3. Variabel penelitian

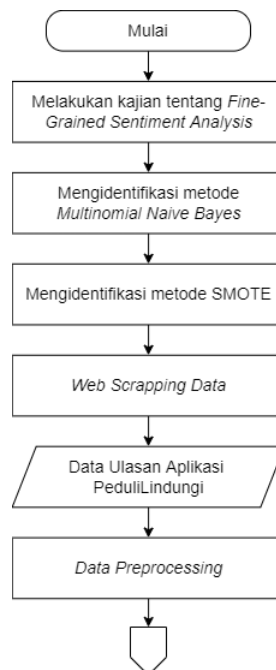
| Variabel | Data |
|----------|----------|
| Y_1 | userName |
| Y_2 | score |
| Y_3 | at |
| Y_4 | content |

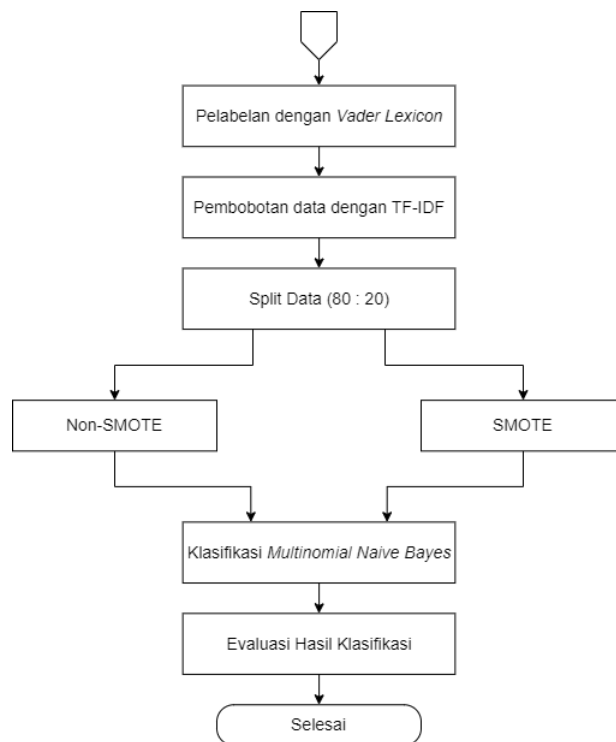
8.2. Langkah Penelitian. Berikut merupakan langkah penelitian yang dilakukan untuk mencapai tujuan penelitian.

- (1) Melakukan kajian tentang analisis sentimen yaitu *Fine-Grained Sentiment Analysis*. Kajian dimulai dengan mengidentifikasi variabel-variabel yang mendasari *Fine-Grained Sentiment Analysis*.

- (2) Mengidentifikasi metode *Multinomial Naïve Bayes*.
- (3) Mengidentifikasi metode SMOTE.
- (4) Melakukan *Web Scrapping* pada data ulasan Aplikasi PeduliLindungi di Google Play Store.
- (5) Melakukan *Data Preprocessing* pada data ulasan Aplikasi PeduliLindungi.
- (6) Melakukan pelabelan dengan *Vader Lexicon* pada data hasil *preprocessing* ulasan Aplikasi PeduliLindungi.
- (7) Pembobotan data dengan TF-IDF pada data hasil *preprocessing* ulasan Aplikasi PeduliLindungi.
- (8) Membagi data *training* 80% dan data *testing* 20% pada data ulasan Aplikasi PeduliLindungi.
- (9) Melakukan dua teknik yaitu teknik SMOTE dan non-SMOTE untuk membandingkan performa klasifikasi
- (10) Melakukan klasifikasi *Multinomial Naïve Bayes* pada data ulasan Aplikasi PeduliLindungi.
- (11) Melakukan Evaluasi pada hasil klasifikasi terbaik.
- (12) Memvisualisasi dengan menggunakan *Wordcloud*

Pada Gambar 1 ditunjukkan *flowchart* langkah penelitian.





Gambar 1. *Flowchart* langkah penelitian

9. JADWAL

Seluruh kegiatan tugas akhir (TA) ini diatur dalam jadwal yang ditunjukkan pada Tabel 4.

Tabel 4. Jadwal kegiatan tugas akhir

| No. | Kegiatan | 2022 | | 2023 | | |
|-----|--|------|-----|------|-----|-------|
| | | Nov | Des | Jan | Feb | Maret |
| 1. | Ujian dan revisi proposal TA | ✓ | ✓ | | | |
| 2. | Penelitian, pembahasan, dan hasil akhir serta kesimpulan | | ✓ | ✓ | | |
| 3. | Pembuatan dan penyusunan skripsi | | ✓ | ✓ | ✓ | |
| 4. | Pembuatan dan penyusunan artikel | | | | ✓ | |
| 5. | Seminar hasil | | | | ✓ | |
| 6. | Ujian dan revisi skripsi | | | | | ✓ |

DAFTAR RUJUKAN

- [1] Andika, L., A., Azizah, P., A., N., & Respatiwan. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier. *Indonesian Journal of Applied Statistics*. Volume 2, No. 1.
- [2] Apriani, R., & Gustian, D. (2019). Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*. Volume 6, No.1, Hal.54-62.
- [3] Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). *Sentiment analysis on E-sports for education curriculum using naïve Bayes and support vector machine*. *Jurnal Ilmu Komputer dan Informasi*, 13(2), 109-122.
- [4] Bekkar, M., Djemma, H., K., & Alitouche, T., A. (2013). *Evaluation Measures for Models Assessment over Imbalanced Data Sets*. *Journal of Information Engineering and Applications*. Vol. 3, No. 10.
- [5] Castellà, Q., & Sutton, C. (2014). *Word storms: Multiples of Word Clouds for Visual Comparison of Documents*. WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web, 665–675. <https://doi.org/10.1145/2566486.2567977>.
- [6] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. *Journal of Artificial Intelligence Research*, 16, 321–357.
- [7] Daulay, E., D., P., & Asror, I. (2020). Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes. *E-Proceeding of Engineering: Volume 7, No.2*.
- [8] Elbagir, S., Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 122, p. 16).
- [9] Gunawan, F., Fauzi, M. A., & Adikara, P. P. (2017). Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- [10] He, H. & Ma, Y. (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications*. ISBN: 9781118074626.
- [11] Herdiana, D. (2021). Aplikasi Peduli Lindungi: Perlindungan Masyarakat Dalam Mengakses Fasilitas Publik Di Masa Pemberlakuan Kebijakan PPKM. *Jurnal Inovasi Penelitian*. Volume 2, No. 6.
- [12] Hutto, C., Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

- [13] Liu, L., & Ozsu, M. T. (2009). *Encyclopedia of Database Systems*. In Encyclopedia of Database Systems. Springer. <https://doi.org/10.1007/978-0-387-39940-9>.
- [14] Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). *Document-level sentiment classification: An empirical comparison between SVM and ANN*. Expert Systems with Applications, 40(2), 621–633. <https://doi.org/10.1016/j.eswa.2012.07.059>.
- [15] Nurjanah, W., Perdana, R., & Fauzi, M. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 1, no. 12, p. 1750-1757. ISSN 2548-964X.
- [16] Nurhidayati, N., Sugiyah, S., & Yuliantari, K. (2021). Pengaturan Perlindungan Data Pribadi Dalam Penggunaan Aplikasi Pedulilindungi. Widya Cipta: Jurnal Sekretari Dan Manajemen, 5(1), 39-45.
- [17] Nurrohmat, M. A., & Azhari, S. N. (2019). Sentiment Analysis of Novel Review Using Long Short-Term Memory Method. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 13(3), 209-218.
- [18] PeduliLindungi,[online],<https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=en&gl=US>, diakses tanggal 15 September 2022
- [19] Pintoko, B., M., & Muslim, K., L. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier. E-Proceeding of Engineering: Volume 5, No.3.
- [20] Sabily, A. F., Adikara, P. P., & Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.
- [21] Turland, M. (2010). *php—architect’s Guide to Web Scraping with PHP*.