

ANALISIS SENTIMEN PADA ULASAN APLIKASI PEDULI
LINDUNGI DI *GOOGLE PLAY STORE* DENGAN METODE
ADASYN-MULTINOMIAL NAIVE BAYES

Proposal TA
Program Studi Matematika



Diajukan oleh
IMAM SUYUTI
M0119043

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET SURAKARTA

2022

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. Beberapa sektor vital seperti fasilitas pemerintahan, kesehatan, dan pasar atau minimarket masih 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 Peduli Lindungi. Aplikasi ini dikembangkan pemerintah untuk melindungi masyarakat yang tengah mengakses fasilitas publik agar kegiatan yang tengah dilakukan aman dan dapat terhindar dari adanya penyebaran Covid-19 (Herdiana [11]). Aplikasi ini juga dapat digunakan dalam pelaksanaan *surveilans* kesehatan dalam menangani penyebaran Covid-19, dengan melakukan *Tracing* yaitu kegiatan pelacakan terhadap orang-orang yang berkontak dengan orang-orang yang diduga terinfeksi Covid-19. Selain itu juga *Tracking* yaitu melacak penyebaran virus corona dengan melihat siapa saja yang telah bertemu dengan orang yang positif terinfeksi virus Covid-19 dan menyelenggarakan *Warning and Fencing* yaitu adanya peringatan dan pengawasan dengan membatasi pergerakan seseorang yang sedang dalam karantina atau isolasi (Nurhidayati dkk. [16]). Penggunaan aplikasi ini masih dilakukan pada masa setelah darurat Covid-19. Partisipasi masyarakat sangat dibutuhkan dengan membagikan data lokasinya saat bepergian agar dapat dilakukan penelusuran riwayat kontak dengan orang yang positif terinfeksi virus Covid-19.

Pada setiap aplikasi, juga terdapat rating dan ulasan pengguna mengenai pelayanan dan fitur-fitur yang diberikan. Ulasan yang diberikan dapat berupa saran, kritik, maupun keluhan. Hal tersebut sangat bermanfaat dan berguna 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.* [15], metode yang cocok untuk mengumpulkan data informasi tersebut adalah penggunaan metode *web scrapping*. Metode *web scrapping* merupakan metode yang digunakan untuk mengekstraksi informasi atau data semi-terstruktur dari *website*. Fitur analisis sentimen yang berguna untuk mengelompokkan atau menyaring antara ulasan positif dan negatif belum ditemukan, mengingat jumlah ulasan aplikasi yang sangat banyak dan akan terus bertambah setiap saat.

Pada penelitian ini, akan menganalisis sentimen ulasan tentang aplikasi Peduli Lindungi menggunakan metode Non-ADASYN dengan klasifikasi *Multinomial Naïve Bayes*. Kemudian membandingkan hasil kinerjanya dengan metode klasifikasi *Multinomial Naïve Bayes* yang disertai ADASYN untuk mendapatkan hasil yang terbaik. Metode ADASYN merupakan salah-satu teknik *oversampling* yang digunakan untuk menangani dataset yang tidak seimbang dalam klasifikasi data. Selain itu, penelitian ini berguna untuk mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif dengan kedua metode ADASYN dan Non-ADASYN pada ulasan aplikasi Peduli Lindungi.

2. PERUMUSAN MASALAH

Berdasarkan latar belakang masalah dirumuskan tiga masalah, yaitu

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

3. TUJUAN PENELITIAN

Berdasarkan perumusan masalah diperoleh tiga tujuan, yaitu

- (1) menganalisis sentimen ulasan tentang aplikasi Peduli Lindungi dengan menggunakan metode Non-ADASYN disertai metode klasifikasi *Multinomial Naïve Bayes*,
- (2) membandingkan hasil hasil kinerjanya dengan metode ADASYN disertai metode klasifikasi *Multinomial Naïve Bayes* untuk mendapatkan hasil yang terbaik, dan
- (3) mendapatkan kata-kata yang sering muncul dalam sentimen positif dan negatif dengan kedua metode Non-ADASYN dan ADASYN pada ulasan tentang aplikasi Peduli Lindungi.

4. MANFAAT PENELITIAN

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

5. TINJAUAN PUSTAKA

Pada tahun 2008, He *et al.* [10] meneliti tentang ADASYN sebagai pendekatan pengambilan sampel sintetis adaptif untuk pembelajaran yang tidak seimbang. Hasil penelitian menunjukkan bahwa hasil simulasi pada lima set data berdasarkan berbagai matrik evaluasi menunjukkan efektivitas metode ADASYN.

Penelitian yang dilakukan oleh Pintoko dan Muslim [19] pada tahun 2018 menyimpulkan bahwa metode *naïve bayes* dapat menganalisis sentimen pada data ulasan jasa transportasi *online* pada twitter. Pada tahun 2019, Apriani dan Gustian [3] melakukan penelitian tentang analisis sentimen dengan *naïve bayes* terhadap komentar aplikasi tokopedia. Hasil penelitian menunjukkan bahwa algoritma *naïve bayes* dapat menganalisis sentimen pada data ulasan aplikasi Tokopedia.

Pada tahun 2019, Abbas *et al.* [1] meneliti model klasifikasi *multinomial naïve*

bayes untuk analisis sentimen. Hasil penelitian menunjukkan bahwa model *multinomial naïve bayes* mencapai hasil yang signifikan dalam kinerja kategorisasi teks. Pada tahun 2020, Daulay dan Asror [6] melakukan penelitian tentang sentimen analisis pada ulasan *Google Play Store* menggunakan metode *naïve bayes*. Hasil penelitian menunjukkan bahwa metode *multinomial naïve bayes* menghasilkan akurasi tertinggi untuk kasus klasifikasi ulasan berdasarkan teks. Pada tahun 2022, Agustina *et al.* [2] meneliti tentang implementasi algoritma *naïve bayes* untuk analisis sentimen ulasan shopee pada *Google Play Store*. Hasil penelitian menunjukkan bahwa metode *multinomial naïve bayes* dengan pembagian data *Hold-Out* (pembagian data *training* dan *testing*) menghasilkan akurasi tertinggi untuk kasus klasifikasi ulasan shopee.

6. LANDASAN TEORI

Berikut dijelaskan teori yang diperlukan untuk mencapai tujuan penelitian, yaitu analisis sentimen, *web scrapping*, *data preprocessing*, pembobotan TF-IDF, VADER, ADASYN, *multinomial naïve bayes*, evaluasi sistem klasifikasi, serta *word-cloud*.

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 (Sabily dkk. [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 (Gunawan dkk. [8]).

Analisis sentimen dibagi menjadi dua kategori yaitu *Coarse-Grained Sentiment Analysis* dan *Fined-Grained Sentiment Analysis* (Sabily dkk. [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 *fine-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 mengekstrak informasi atau data tertentu dari halaman tersebut untuk dianalisis (Turland [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 (Nurrohmat & Azhari [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. **Pembobotan TF-IDF.** *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan salah satu model yang digunakan untuk menghitung bobot kata pada suatu dokumen dengan menggunakan model *term frequency* untuk memberikan

pembobotan hubungan suatu kata atau istilah terhadap keseluruhan ulasan. Frekuensi kemunculan kata di dalam ulasan menunjukkan seberapa penting kata itu di dalam ulasan tersebut, dan ulasan mana yang memiliki kata tersebut sehingga ulasan dapat diklasifikasikan ke dalam 2 kelas yaitu ulasan positif dan ulasan negatif (Melita [14]). Rumus pembobotan TF-IDF ditulis sebagai

$$W_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right) \quad (6.1)$$

dengan $W_{x,y}$ adalah bobot istilah (t_y) terhadap dokumen (d_x). Sedangkan $tf_{x,y}$ adalah jumlah kemunculan istilah (t_y) dalam dokumen (d_x). N adalah jumlah semua dokumen yang ada dalam *dataset* dan df_x adalah jumlah dokumen yang mengandung istilah (t_y), minimal ada satu kata yaitu istilah (t_y).

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 (Elbagir & Yang [7]). Salah satu contoh dari metode sentimen analisis secara leksikal yaitu *vader lexicon polarity detection* (Hutto & Gilbert [12]). Metode leksikal tidak memerlukan *data training* atau data yang telah dilabeli namun sudah tersedia dalam kamus lengkap dengan kepolaritasan sentimennya.

6.6. Adaptive Synthetic Sampling Approach (ADASYN). ADASYN merupakan algoritma yang digunakan untuk menangani dataset yang tidak seimbang dalam klasifikasi data (He *et al.* [10]). ADASYN dapat menghasilkan sampel secara adaptif dalam data sintetik terhadap kelas minoritas yang dibentuk oleh distribusi data untuk mengurangi bias yang disebabkan oleh distribusi data yang tidak merata pada data dengan label lain yang memiliki kelas mayoritas.

6.7. Multinomial Naïve Bayes. *Multinomial Naïve Bayes* merupakan metode *supervised learning* yang menggunakan probabilitas dan lebih difokuskan untuk klasifikasi teks (Liu & Ozsu [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. Proporsional probabilitas ulasan d yang memiliki kelas c ditunjukkan pada rumus berikut

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_i|c) \quad (6.2)$$

dengan:

$P(c|d)$: 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	
	Positif	Negatif
Positif	True Positive	False Negative
Negatif	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 sebagai berikut

$$Accuracy = \frac{\text{Prediksi data benar}}{\text{Total data}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.3)$$

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 sebagai berikut

$$Recall = \frac{\text{relevant item retrieved}}{\text{retrieved item}} = \frac{TP}{TP + FN} \quad (6.4)$$

6.8.4. **Precision.** *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban sistem, *precision* dapat dihitung sebagai berikut

$$Precision = \frac{\text{relevant item retrieved}}{\text{retrieved item}} = \frac{TP}{TP + FP} \quad (6.5)$$

6.8.5. **Area Under Curve (AUC).** AUC merupakan kriteria evaluasi yang menggunakan sensitivitas atau spesifisitas sebagai dasar pengukuran (He & Ma [9]). 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*.

Menurut Bekkar *et al.* [4], nilai AUC pada umumnya berada pada $[0, 5-1, 0]$. Berikut ini tabel keterangan untuk masing-masing interval nilai AUC dapat ditunjukkan pada Tabel 2.

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 (Castella *et al.* [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 (Liu & Ozsu [13]). Metode *Multinomial Naïve Bayes* dapat diterapkan pada analisis sentimen pada ulasan aplikasi Peduli Lindungi di *Google Play Store*. Metode yang digunakan untuk mengatasi *imbalanced dataset* adalah Metode ADASYN.

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 *ADASYN-Multinomial Naïve Bayes* dan menerapkan pada data ulasan aplikasi Peduli Lindungi.

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

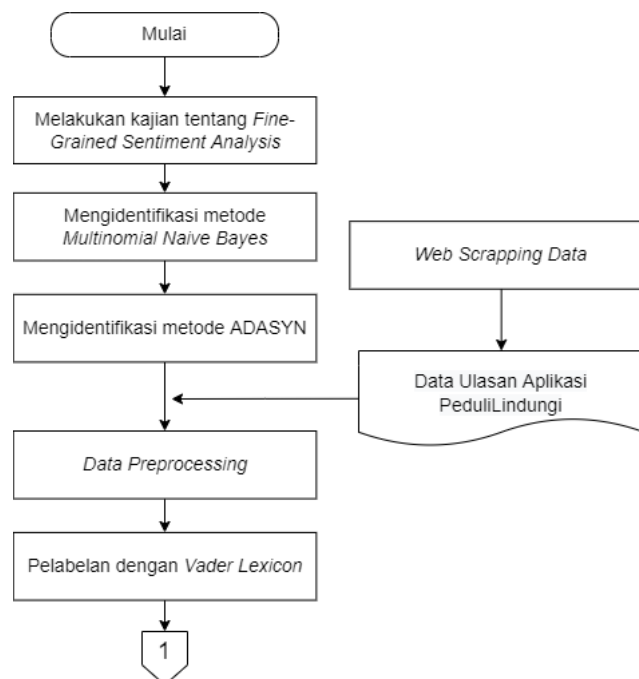
Tabel 3. Variabel penelitian	
Variabel	Data
Y_1	Nama <i>reviewer</i>
Y_2	<i>Rating</i>
Y_3	Tanggal ulasan
Y_4	Ulasan

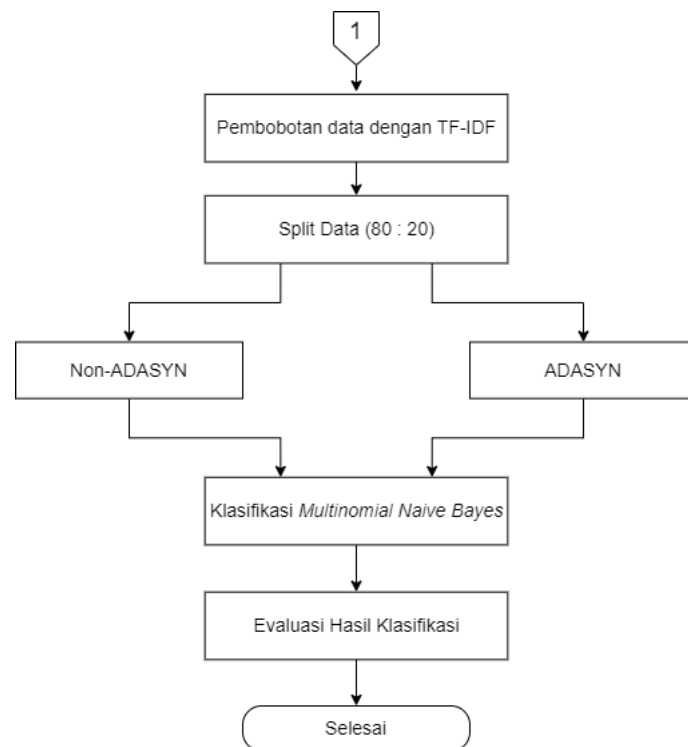
8.2. Langkah Penelitian. Berikut merupakan langkah penelitian yang dilakukan untuk mencapai tujuan penelitian.

- (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 ADASYN.
- (4) Melakukan *Web Scrapping* pada data ulasan Aplikasi Peduli Lindungi di *Google Play Store*.
- (5) Melakukan *Data Preprocessing* pada data ulasan Aplikasi Peduli Lindungi.
- (6) Melakukan pelabelan dengan *Vader Lexicon* pada data hasil *preprocessing* ulasan Aplikasi Peduli Lindungi.
- (7) Pembobotan data dengan TF-IDF pada data hasil *preprocessing* ulasan Aplikasi Peduli Lindungi.
- (8) Membagi data *training* 80% dan data *testing* 20% pada data ulasan Aplikasi Peduli Lindungi.
- (9) Melakukan dua teknik yaitu teknik ADASYN dan non-ADASYN untuk membandingkan performa klasifikasi
- (10) Melakukan klasifikasi *Multinomial Naïve Bayes* pada data ulasan Aplikasi Peduli Lindungi.
- (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 kesimpulan		✓	✓		
3.	Penulisan skripsi		✓	✓	✓	
4.	Penulisan artikel				✓	
5.	Seminar hasil				✓	
6.	Ujian dan revisi skripsi					✓

DAFTAR RUJUKAN

- [1] Abbas, M., et al. *Multinomial Naive Bayes classification model for sentiment analysis*. IJCSNS Int. J. Comput. Sci. Netw. Security, Vol. 19, No. 3, 2019.
- [2] Agustina, N., et al. *Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store*. MALCOM: Indonesian Journal of Machine Learning and Computer Science, Vol.2, No.1 , 2022.
- [3] Apriani, R., dan Gustian, D. *Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia*. Jurnal Rekayasa Teknologi Nusa Putra. Volume 6, No.1, Hal.54-62, 2019.
- [4] Bekkar, M., Djemma, H., K., and Alitouche, T., A. *Evaluation Measures for Models Assessment over Imbalanced Data Sets*. Journal of Information Engineering and Applications. Vol. 3, No. 10, 2013.
- [5] Castellà, Q., and Sutton, C. *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, 2014.
- [6] Daulay, E., D., P., and Asror, I. *Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes*. E-Proceeding of Engineering: Volume 7, No.2, 2020.
- [7] Elbagir, S., and Yang, J. *Twitter sentiment analysis using natural language toolkit and VADER sentiment*. In Proceedings of the international multiconference of engineers and computer scientists, Vol. 122, No. 16, 2019.
- [8] Gunawan, F., Fauzi, M. A., dan Adikara, P. P. *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, 2017.
- [9] He, H. and Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*. ISBN: 9781118074626, 2013.
- [10] He, Haibo, et al. *ADASYN: Adaptive synthetic sampling approach for imbalanced learning*. 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008.
- [11] Herdiana, D. *Aplikasi Peduli Lindungi: Perlindungan Masyarakat Dalam Mengakses Fasilitas Publik Di Masa Pemberlakuan Kebijakan PPKM*. Jurnal Inovasi Penelitian. Volume 2, No. 6, 2021.
- [12] Hutto, C., and Gilbert, E. *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, 2014.

- [13] Liu, L., and Ozsu, M. T. *Encyclopedia of Database Systems*. In Encyclopedia of Database Systems. Springer, 2009.
- [14] Melita, Ria. *Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)*. BS thesis. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2018.
- [15] Moraes, R., Valiati, J. F., and Gavião Neto, W. P. *Document-level sentiment classification: An empirical comparison between SVM and ANN*. Expert Systems with Applications, Vol. 40, No. 2, 2013.
- [16] Nurhidayati, N., Sugiyah, S., dan Yuliantari, K. *Pengaturan Perlindungan Data Pribadi Dalam Penggunaan Aplikasi Pedulilindungi*. Widya Cipta: Jurnal Sekretari Dan Manajemen, Vol. 5, No. 1, 2021.
- [17] Nurrohmat, M. A., and Azhari, S. N. *Sentiment Analysis of Novel Review Using Long Short-Term Memory Method*. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), Vol. 13, No. 3, 2019.
- [18] Peduli Lindungi, [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., dan Muslim, K., L. *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*. E-Proceeding of Engineering: Vol. 5, No. 3, 2018.
- [20] Sabily, A. F., Adikara, P. P., dan Fauzi, M. A. *Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2019.
- [21] Turland, M. *php—architect’s Guide to Web Scraping with PHP*, 2010.