

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

Penyebaran virus COVID-19 (Coronavirus disease 2019) mendorong pemerintah Indonesia untuk mengeluarkan kebijakan pengendalian penyebaran virus tersebut dengan memberlakukan pembatasan berskala besar yang disebut dengan istilah Pembatasan Sosial Berskala Besar (PSBB). PSBB ini bertujuan untuk memutus mata rantai penyebaran virus COVID-19. Beberapa fasilitas umum ditutup selama PSBB meskipun terdapat sektor vital seperti fasilitas pemerintahan, kesehatan, dan pasar atau minimarket masih tetap buka dengan menjalankan protokol kesehatan. Kebijakan tersebut didasari pada Undang-Undang No. 6 Tahun 2018 tentang Kekarantinaan Kesehatan.

Dalam rangka menangani penyebaran wabah COVID-19 di Indonesia, PT Telekomunikasi Indonesia Tbk (Telkom) dan Kementerian Komunikasi dan Informatika (Kemkominfo) bekerjasama membuat aplikasi Peduli Lindungi. Aplikasi ini digunakan untuk melindungi masyarakat yang tengah mengakses fasilitas publik agar kegiatan yang sedang dilakukan aman dan dapat terhindar dari adanya penyebaran COVID-19 (Herdiana [10]). Aplikasi ini juga digunakan untuk 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 dengan melakukan *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. [15]). Partisipasi masyarakat sangat dibutuhkan dengan membagikan data lokasinya saat melakukan perjalanan agar dapat dilakukan penelusuran riwayat kontak dengan orang yang positif terinfeksi virus COVID-19.

Pada setiap aplikasi, 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 bagi pengguna lain yang akan menggunakan aplikasi tersebut. Pengumpulan dan penyortiran data ulasan yang dilakukan tanpa algoritme atau aplikasi yang dapat mengumpulkan data tidaklah hal

yang mudah karena ulasan yang tersedia pada fitur komentar di situs *Google Play* biasanya sangat banyak. Menurut Moraes *et al.* [14], metode yang cocok untuk mengumpulkan data informasi tersebut adalah metode *web scraping*. Metode *web scraping* merupakan metode yang digunakan untuk mengumpulkan informasi atau data semi-terstruktur dari *website*.

Analisis sentimen berguna untuk mengelompokkan atau menyaring antara sentimen positif dan negatif pada suatu ulasan. Dalam hal ini, digunakan metode *vader lexicon* untuk melakukan proses pelabelan secara otomatis kelas sentimen pada data ulasan berbahasa Inggris. Permasalahan yang sering muncul pada sebagian besar penelitian analisis sentimen, yaitu kebanyakan data ulasan cenderung tidak seimbang (*imbalanced dataset*) dari segi jumlah kelas tiap individu, misalnya cenderung ke arah positif atau sebaliknya. Secara umum, algoritme *machine learning* akan menghasilkan suatu model dengan tingkat sensitivitas yang rendah terhadap kelas minoritas ketika menerima *dataset* yang tidak seimbang karena hal tersebut akan menyebabkan performa klasifikasi sentimen yang dilakukan menjadi buruk.

Pada penelitian ini, akan dianalisis sentimen pada ulasan tentang aplikasi Peduli Lindungi dengan menggunakan metode ADASYN dan *multinomial naïve bayes*. Metode ADASYN digunakan untuk menangani kasus data yang tidak seimbang. Sedangkan metode *multinomial naïve bayes* digunakan untuk mengklasifikasi data ulasan ke dalam sentimen positif dan negatif.

## 2. PERUMUSAN MASALAH

Berdasarkan latar belakang masalah dirumuskan masalah yaitu bagaimana menganalisis sentimen pada ulasan tentang aplikasi Peduli Lindungi dengan menggunakan metode ADASYN dan *multinomial naïve bayes*.

## 3. TUJUAN PENELITIAN

Berdasarkan perumusan masalah, tujuan penelitian ini adalah menganalisis sentimen pada ulasan tentang aplikasi Peduli Lindungi dengan menggunakan metode ADASYN dan *multinomial naïve bayes*.

#### 4. MANFAAT PENELITIAN

Penelitian ini diharapkan untuk menambah wawasan ilmu pengetahuan tentang implementasi metode ADASYN dan *multinomial naïve bayes* dalam klasifikasi sentimen pada ulasan aplikasi Peduli Lindungi.

#### 5. TINJAUAN PUSTAKA

Pada tahun 2008, He *et al.* [9] meneliti tentang ADASYN untuk pembelajaran *dataset* yang tidak seimbang. Hasil penelitian menunjukkan bahwa hasil simulasi pada lima set data berdasarkan berbagai matrik evaluasi menunjukkan efektivitas metode ADASYN dibandingkan SMOTE.

Pada tahun 2017, Song *et al.* [20] meneliti tentang pendekatan klasifikasi novel berdasarkan *naïve bayes* untuk analisis sentimen Twitter. Hasil penelitian menunjukkan bahwa metode *naïve bayes* menghasilkan akurasi tertinggi dibanding *maximum entropy* dan *support vector machine*. Penelitian yang dilakukan oleh Pintoko dan Muslim [18] pada tahun 2018 menyimpulkan bahwa metode *naïve bayes* dapat menganalisis sentimen pada data ulasan jasa transportasi *online* pada Twitter. Namun penelitian ini masih terdapat beberapa kekurangan, yaitu kurangnya fitur yang digunakan pada data latih dan adanya perbedaan hasil dari pelabelan sentimen data yang dilakukan secara manual untuk menguji model dengan hasil prediksi sentimen dari hasil klasifikasi model.

Pada tahun 2019, Abbas *et al.* [1] meneliti model klasifikasi *multinomial naïve bayes* untuk analisis sentimen. Hasil penelitian menunjukkan bahwa model *multinomial naïve bayes* dengan *term frequency inverse document frequency* (TF-IDF) mencapai hasil yang lebih signifikan dibandingkan model *multinomial naïve bayes* tanpa TF-IDF dalam kinerja kategorisasi teks. Pada tahun 2022, Agustina *et al.* [2] mengimplementasikan 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 yang lebih baik dibandingkan pembagian data *k-fold cross validation* untuk kasus klasifikasi ulasan Shopee.

## 6. LANDASAN TEORI

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

**6.1. Analisis Sentimen.** Analisis sentimen merupakan salah satu bidang pada *text mining* yang menganalisa sebuah pendapat, opini, evaluasi, sentimen, sikap atau penilaian seseorang terhadap individu, kelompok, produk, organisasi, masalah, peristiwa atau topik (Sabily dkk. [19]). 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. [7]).

Analisis sentimen dibagi menjadi dua kategori yaitu *Coarse-Grained Sentiment Analysis* dan *Fined-Grained Sentiment Analysis* (Sabily dkk. [19]).

**6.1.1. Coarse-Grained sentiment analysis.** *Coarse-Grained* adalah proses menganalisis sentimen sebuah dokumen secara keseluruhan. Sentimen ini ada tiga jenis yaitu positif, netral, dan negatif. *Coarse-Grained* biasanya digunakan untuk ulasan-ulasan yang berupa dokumen seperti ulasan hotel, film, dan buku.

**6.1.2. Fine-Grained sentiment analysis.** *Fine-Grained* adalah proses menganalisis yang orientasinya lebih spesifik, yaitu pada kalimat di sebuah dokumen. Sentimen ini ada dua jenis yaitu positif dan negatif. *Fine-Grained* biasanya digunakan untuk ulasan-ulasan yang berupa kalimat seperti ulasan produk, aplikasi, dan jasa. Contoh dari *finned-grained* 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 Scraping.** *Web scraping* adalah metode yang digunakan untuk mengumpulkan informasi atau data semi-terstruktur dari *website* (biasanya dalam bentuk halaman *web* dalam bahasa markup, seperti HTML atau XHTML) untuk dianalisis (Turland [21]).

Proses *web scraping* dibagi menjadi tiga tahap yaitu *download content* dari halaman *web*, mengumpulkan data, dan menyimpan data dalam format *csv* atau *json*. Salah-satu bahasa yang digunakan untuk proses *web scraping* adalah *python* karena memiliki *libraries* yang memproses data dengan baik.

**6.3. Data Preprocessing.** *Data preprocessing* merupakan proses membersihkan data yang dilakukan setelah *dataset* terkumpul, agar proses pada *machine learning* menjadi lebih cepat dan akurat (Nurrohmat & Azhari [16]). 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. Huruf kapital akan diubah ke huruf kecil sedangkan tanda baca dan angka dihapus.
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 *Term Frequency-Inverse Document Frequency* (TF-IDF).** TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat (Maarif [13]). Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap kata di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen. Rumus untuk menghitung bobot setiap kata  $t$  di dokumen  $d$  ditulis sebagai

$$W_{dt} = tf_{dt} \times IDF_{dt} \quad (6.1)$$

dengan

$d$  : dokumen ke- $d$

$t$  : kata ke- $t$  dari kata kunci

$W$  : bobot dokumen ke- $d$  terhadap kata ke- $t$   
 $tf$  : banyaknya kata yang dicari pada sebuah dokumen  
 $IDF$  : *Inversed Document Frequency*  
 $IDF = \log(\frac{N}{df})$   
 $N$  : total dokumen  
 $df$  : banyak dokumen yang mengandung kata yang dicari.

**6.5. *Valence Aware Dictionary for Social Reasoning (VADER)*.** 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 [5]). Metode leksikal merupakan metode yang tidak memerlukan *data training* atau data yang telah dilabeli namun sudah tersedia dalam kamus lengkap dengan kepolaritasan sentimennya. Salah-satu contoh dari metode analisis sentimen secara leksikal yaitu *vader lexicon polarity detection* (Hutto & Gilbert [11]).

**6.6. *Adaptive Synthetic Sampling Approach (ADASYN)*.** ADASYN merupakan algoritme yang digunakan untuk menangani *dataset* yang tidak seimbang dalam klasifikasi data (He *et al.* [9]). *Dataset* yang tidak seimbang berdasarkan jumlah data tiap kelas, misalnya cenderung ke arah positif atau sebaliknya. 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. Berikut algoritme ADASYN

Input

- (1) *Training dataset*  $D_{tr}$  dengan  $m$  sampel  $\{x_i, y_i\}$ ,  $i = 1, \dots, m$ , dimana  $x_i$  merupakan sampel dalam  $n$  dimensional feature space  $X$  dan  $y_i \in Y = \{1, -1\}$  merupakan label identitas kelas yang berkaitan dengan  $x_i$ . Mendefinisikan  $m_s$  dan  $m_l$  sebagai jumlah sampel kelas minoritas dan jumlah sampel kelas mayoritas. Oleh karena itu,  $m_s \leq m_l$  dan  $m_s + m_l = m$ .

Prosedur

- (1) Menghitung tingkat ketidakseimbangan kelas

$$d = m_s/m_l \quad (6.2)$$

dimana  $d \in (0, 1]$ .

- (2) Jika  $d < d_{th}$  dimana ( $d_{th}$  merupakan penetapan *threshold* untuk derajat toleransi maksimum dari rasio ketidakseimbangan kelas):

- (a) Menghitung jumlah sampel data sintetis yang perlu dihasilkan untuk kelas minoritas

$$G = (m_l - m_s) \times \beta \quad (6.3)$$

dimana  $\beta \in [0, 1]$  merupakan parameter yang digunakan untuk menentukan tingkat keseimbangan yang diinginkan setelah generalisasi data sintetis.  $\beta = 1$  berarti *dataset* yang sepenuhnya seimbang dibuat setelah proses generalisasi.

- (b) Untuk setiap sampel  $x_i \in \text{minority class}$ , Menemukan  $K$  tetangga terdekat berdasarkan jarak euclidan pada  $n \text{ dimensional space}$ , dan menghitung rasio  $r_i$  yang didefinisikan sebagai

$$r_i = \Delta_i/K, \quad i = 1, \dots, m_s \quad (6.4)$$

dimana  $\Delta_i$  merupakan jumlah sampel di  $K$  tetangga terdekat dari  $x_i$  yang termasuk dalam kelas mayoritas, oleh karena itu  $r_i \in [0, 1]$ .

- (c) Menormalisasi  $r_i$  berdasarkan  $\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i$ , sehingga  $\hat{r}_i$  merupakan distribusi kerapatan ( $\sum_i \hat{r}_i = 1$ ).

- (d) Menghitung jumlah sampel data sintetis yang perlu dihasilkan pada setiap sampel minoritas  $x_i$

$$g_i = \hat{r}_i \times G \quad (6.5)$$

dimana  $G$  merupakan jumlah sampel data sintetis yang perlu dihasilkan untuk kelas minoritas seperti yang didefinisikan pada persamaan (6.3).

- (e) Untuk setiap sampel data kelas minoritas  $x_i$ , Menghasilkan sampel data sintetis sebanyak  $g_i$ .



**6.7. Multinomial Naïve Bayes.** *Multinomial Naïve Bayes* merupakan metode *supervised learning* yang menggunakan probabilitas dan lebih difokuskan untuk klasifikasi teks (Liu & Ozsü [12]). *Multinomial Naïve Bayes* juga memiliki fitur unik, yaitu hasil yang diperoleh 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. Probabilitas ulasan  $d$  yang memiliki kelas  $c$  ditunjukkan pada rumus yang ditulis sebagai

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_i|c) \quad (6.6)$$

dengan

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

$P(c)$  : probabilitas *prior*  $c$

$P(w_i|c)$  : probabilitas  $w_i$  pada kelas  $c$

$w_i$  : kata ke- $i$ .

**6.8. Evaluasi Kinerja Klasifikasi.** Sebuah sistem klasifikasi harus dinilai performanya agar dapat mengukur tingkat akurasi dari prediksi klasifikasi yang dihasilkan. Ada dua metode perhitungan yang digunakan untuk menilai performa klasifikasi, satu metode (6.8.1) menurut Bramer [4] dan satu metode (6.8.2) menurut He dan Ma [8] yang ditunjukkan sebagai berikut.

**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	<i>True Positive (TP)</i>	<i>False Negative (FP)</i>
Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Sejumlah ukuran kinerja klasifikasi dapat ditentukan berdasarkan terminologi yang terdapat pada Tabel 1 diatas. Pada penelitian ini, evaluasi kinerja klasifikasi yang digunakan adalah *Area Under Curve* (AUC). Berikut rumus dari masing-masing kinerja klasifikasi (Gorunescu [6]).

$$Sensitivitas/Recall/TP_{rate} = \frac{TP}{TP + FN} \quad (6.7)$$

$$FP_{rate} = \frac{FP}{FP + TN} \quad (6.8)$$

dimana *recall* merupakan rasio prediksi benar positif dibandingkan dengan seluruh data yang benar positif dan  $FP_{rate}$  merupakan rasio prediksi salah positif dibandingkan dengan seluruh data yang negatif.

6.8.2. **Area Under Curve (AUC).** AUC merupakan kriteria evaluasi yang menggunakan sensitivitas atau spesifisitas sebagai dasar pengukuran. Apabila terjadi kasus ketidakseimbangan data (*imbalance dataset*) maka dalam memilih model mana yang terbaik dapat dilakukan dengan menggunakan nilai AUC karena nilai akurasi hanya mempelajari data mayoritas saja sehingga hasil yang didapatkan mungkin saja terjadi bias atau *overfitting*. Berikut rumus AUC

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (6.9)$$

Menurut Bekkar *et al.* [3], nilai AUC pada umumnya berada pada  $[0, 5-1, 0]$  dengan 5 kategori dan pada Tabel 2 ditunjukkan interval masing-masing kategori.

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>

## 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 & Ozsü [12]). 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.

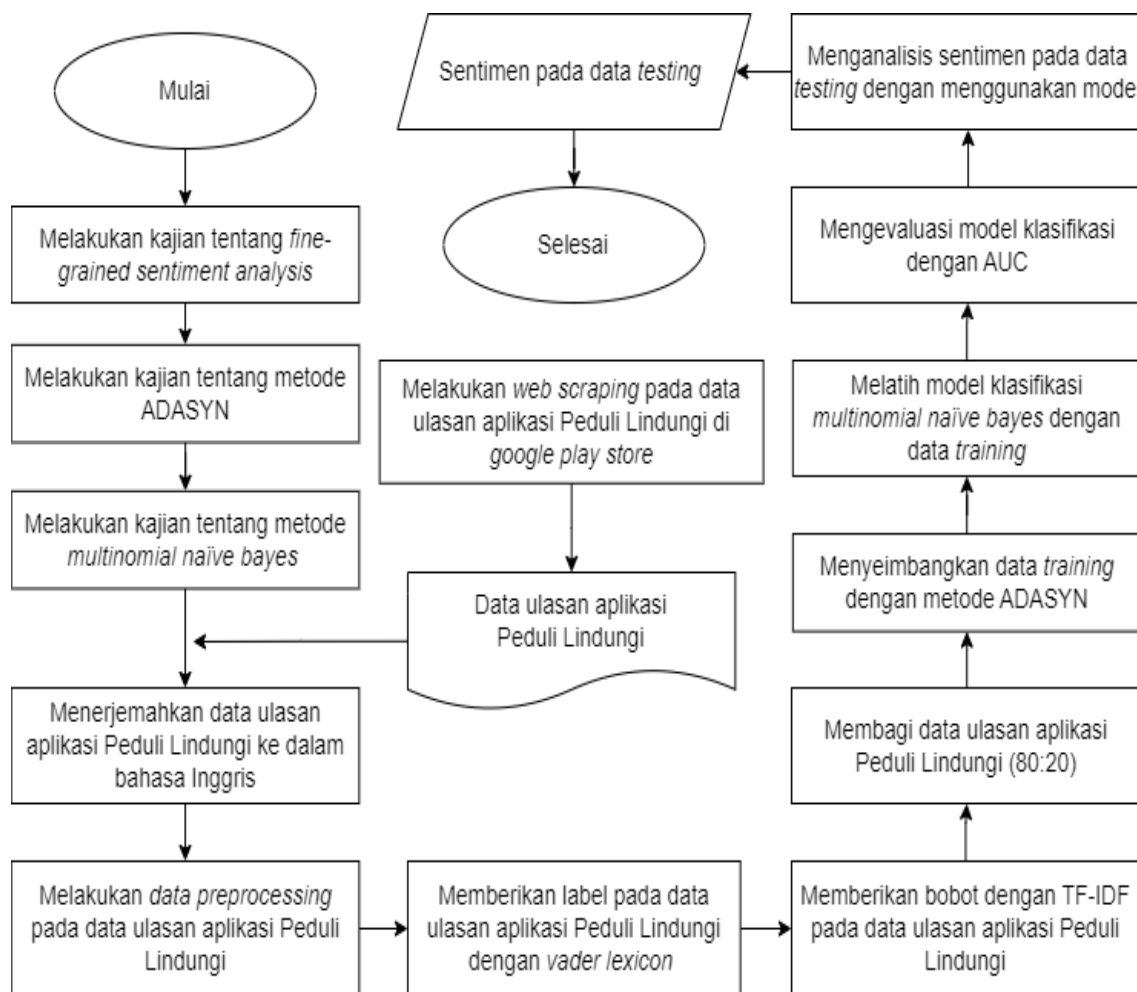
## 8. METODOLOGI PENELITIAN

Metodologi penelitian dibagi menjadi dua bagian yaitu data penelitian dan langkah penelitian. Penelitian ini merupakan penelitian terapan yakni implementasi metode ADASYN dan *multinomial naïve bayes* dalam klasifikasi sentimen pada ulasan aplikasi Peduli Lindungi.

**8.1. Data Penelitian.** Data yang digunakan adalah data sekunder dari ulasan aplikasi Peduli Lindungi di *Google Play Store* (Peduli Lindungi [17]). Empat 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 pada Gambar 1.



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 dan artikel		✓	✓	✓	
4.	Seminar hasil TA dan revisi artikel				✓	
5.	Ujian dan revisi skripsi					✓

## DAFTAR RUJUKAN

- [1] Abbas, M., K.A. Memon, A.A. Jamali, S. Memon, and A. Ahmed, *Multinomial Naive Bayes classification model for sentiment analysis*. IJCSNS Int. J. Comput. Sci. Netw. Security, Vol. 19, No. 3, 2019.
- [2] Agustina, N., D.H. Citra, W. Purnama, C. Nisa, dan A.R. Kurnia, *Implementasi Algoritme Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store*. MALCOM: Indonesian Journal of Machine Learning and Computer Science, Vol. 2, No. 1, 2022.
- [3] Bekkar, M., H.K. Djemma, , and T.A. Alitouche, *Evaluation Measures for Models Assessment over Imbalanced Data Sets*. Journal of Information Engineering and Applications. Vol. 3, No. 10, 2013.
- [4] Bramer, M. *Principles of Data Mining*. London: Springer, 2007.
- [5] Elbagir, S., and J. Yang, *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.
- [6] Gorunescu, F. *Data Mining: Concepts, Models and Techniques*. Berlin: Springer-Verlag, 2011.
- [7] Gunawan, F., M.A. Fauzi, dan P.P. Adikara, *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.
- [8] He, H. and Y. Ma, *Imbalanced Learning: Foundations, Algorithms, and Applications*. ISBN: 9781118074626, 2013.
- [9] He, H., Y. Bai, E.A. Garcia, and S. Li, *ADASYN: Adaptive synthetic Sampling Approach for Imbalanced Learning*. 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008.
- [10] Herdiana, D. *Aplikasi Peduli Lindungi: Perlindungan Masyarakat Dalam Mengakses Fasilitas Publik Di Masa Pemberlakuan Kebijakan PPKM*. Jurnal Inovasi Penelitian. Vol. 2, No. 6, 2021.
- [11] Hutto, C., and E. Gilbert, *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.
- [12] Liu, L., and M.T. Ozsü, *Encyclopedia of Database Systems*. In Encyclopedia of Database Systems. Springer, 2009.
- [13] Maarif, A.A., *Penerapan Algoritme TF-IDF untuk Pencarian Karya Ilmiah*. Jurnal Jurusan Teknik Informatika. Fakultas Ilmu Komputer. Universitas Dian Nuswantoro, 2015.

- [14] Moraes, R., J.F. Valiati, and W.P. Gavião Neto, *Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN*. Expert Systems with Applications, Vol. 40, No. 2, 2013.
- [15] Nurhidayati, N., S. Sugiyah, dan K. Yuliantari, *Pengaturan Perlindungan Data Pribadi Dalam Penggunaan Aplikasi Pedulilindungi*. Widya Cipta: Jurnal Sekretari Dan Manajemen, Vol. 5, No. 1, 2021.
- [16] Nurrohmat, M. A., and S.N. Azhari, *Sentiment Analysis of Novel Review Using Long Short-Term Memory Method*. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), Vol. 13, No. 3, 2019.
- [17] Peduli Lindungi, [online], <https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=en&gl=US>, diakses tanggal 15 September 2022.
- [18] Pintoko, B., M., dan K.L. Muslim, *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*. E-Proceeding of Engineering: Vol. 5, No. 3, 2018.
- [19] Sabily, A. F., P.P. Adikara, dan M.A. Fauzi, *Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2019.
- [20] Song, J., K.T. Kim, B. Lee, S. Kim, and H.Y. Youn, *A Novel Classification Approach based on Naïve Bayes for Twitter Sentiment Analysis*. KSII Transactions on Internet and Information Systems (TIIS), Vol. 11, No. 6, 2017.
- [21] Turland, M. *php—architect’s Guide to Web Scraping with PHP*, 2010.