ANALISIS PERBANDINGAN METODE NAÏVE BAYES, SVM, DAN BiLSTM UNTUK KLASIFIKASI GENDER BERDASARKAN *TWEET* BERBAHASA INDONESIA



Disusun Oleh:

N a m a	: PURNA AJI WARDHANA
NIM	: 18523292

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA

2026

HALAMAN PENGESAHAN DOSEN PEMBIMBING

ANALISIS PERBANDINGAN METODE NAÏVE BAYES, SVM, DAN BiLSTM UNTUK KLASIFIKASI GENDER BERDASARKAN *TWEET* BERBAHASA INDONESIA

TUGAS AKHIR

Disusun Oleh:

N a m a	:	PURNA AJI WARDHANA
NIM	:	18523292

Yogyakarta, 12 Januari 2026

Pembimbing,

(Chanifah Indah Ratnasari, S.Kom.,M.Kom.)

HALAMAN PENGESAHAN DOSEN PENGUJI

ANALISIS PERBANDINGAN METODE NAÏVE BAYES, SVM DAN BiLSTM UNTUK KLASIFIKASI GENDER BERDASARKAN *TWEET* BERBAHASA INDONESIA

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 12 Januari 2026

Tim Penguji

Chanifah Indah Ratnasari, S.Kom., M.Kom. _____

Anggota 1

Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. _____

Anggota 2

Hari Setiaji, S.Kom., M.Eng. _____

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia

(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : PURNA AJI WARDHANA

NIM : 18523292

Tugas akhir dengan judul:

ANALISIS PERBANDINGAN METODE NAÏVE BAYES, SVM DAN BiLSTM UNTUK KLASIFIKASI GENDER BERDASARKAN *TWEET* BERBAHASA INDONESIA

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 12 Januari 2026



(PURNA AJI WARDHANA)

HALAMAN PERSEMBAHAN

Assalamualaikum warahmatullahi wabarakatuh, dengan rendah hati dan bersyukur, saya mengucapkan terima kasih kepada Allah Swt. atas izin, karunia, dan berkat-Nya yang telah memberikan kelancaran dalam menyelesaikan tugas akhir ini tepat waktu. Semoga ini menjadi langkah awal saya dalam mencapai cita-cita dan berkontribusi secara positif bagi masyarakat.

Sebagai ungkapan rasa terima kasih atas cinta, perhatian, dan dukungan yang tak terhingga, penulis ingin mengabdikan laporan tugas akhir ini untuk Bapak Sunardi dan Ibu Jematun. Mereka selalu memberikan dukungan moril dan materiel, doa, serta restu kepada penulis untuk kelancaran segala urusannya.

Penulis ingin menyampaikan rasa terima kasih kepada Bapak Irving Vitra Papatungan, S.T, M.SC., Ph.D. dan Ibu Chanifah Indah Ratnasari, S.Kom., M.Kom. atas bimbingan dan arahannya yang sangat berharga dalam menyusun tugas akhir ini. Penulis juga ingin mengucapkan terima kasih kepada rekan-rekan sejurusan di Program Studi Informatika 2018 yang telah menjadi bagian berharga dalam perjalanan perkuliahan, mendukung dalam berbagai situasi, dan saling membantu selama masa kuliah.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

HALAMAN MOTO

"Barang siapa keluar untuk mencari sebuah ilmu, maka ia akan berada di jalan Allah hingga ia kembali."

H.R. Tirmidzi

“Jika kamu menungguku untuk menyerah, kau akan menungguku selamanya."

Uzumaki Naruto

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan laporan Tugas Akhir ini dengan judul "ANALISIS PERBANDINGAN METODE NAÏVE BAYES, SVM DAN BiLSTM UNTUK KLASIFIKASI GENDER BERDASARKAN *TWEET* BERBAHASA INDONESIA". Laporan ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Program Studi Informatika - Program Sarjana, Fakultas Teknologi Industri. Penyusunan laporan tugas akhir ini tidak terlepas dari berbagai tantangan, baik dalam proses pengumpulan data, pengolahan data, maupun dalam penyusunan dan analisis hasil penelitian. Namun demikian, berkat bimbingan, dukungan, serta bantuan dari berbagai pihak, seluruh tahapan penelitian dan penulisan laporan ini dapat diselesaikan dengan baik. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua saya, yang selalu memberikan semangat, dukungan, dan memahami situasi penulis selama ini.
2. Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D. selaku Rektor Universitas Islam Indonesia.
3. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri.
4. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc, selaku Ketua Jurusan Informatika Universitas Islam Indonesia
5. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Kepala Prodi Informatika.
6. Bapak Irving Vitra Papatungan, S.T, M.SC., Ph.D. dan Ibu Chanifah Indah Ratnasari, S.Kom., M.Kom. selaku dosen pembimbing yang telah meluangkan waktu selama proses bimbingan.
7. Seluruh dosen serta staf pengajar Program Studi Informatika Universitas Islam Indonesia.
8. Teman-teman serta semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dan memberikan dukungan selama proses penyusunan tugas akhir ini.

Yogyakarta, 12 Januari 2026



(PURNA AJI WARDHANA)

SARI

Media sosial X (Twitter) menghasilkan data teks dalam jumlah besar yang dapat dimanfaatkan untuk kajian pemrosesan bahasa alami, khususnya *author profiling*. Salah satu atribut penting dalam *author profiling* adalah jenis kelamin, terutama ketika informasi profil pengguna tidak tersedia atau tidak dapat diandalkan. Teks *tweet* yang bersifat singkat dan informal menjadikan klasifikasi gender berbasis teks sebagai permasalahan yang menantang, khususnya pada bahasa Indonesia.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Multinomial Naïve Bayes, Support Vector Machine* (SVM), dan *Bidirectional Long Short-Term Memory* (BiLSTM) dalam mengklasifikasikan gender pengguna media sosial X (Twitter) berdasarkan *tweet* berbahasa Indonesia. Data penelitian berupa *tweet* berbahasa Indonesia yang dikumpulkan menggunakan layanan API pihak ketiga dengan beberapa kata kunci yang merepresentasikan topik beragam.

Metodologi penelitian meliputi tahap pengumpulan data, pra-pemrosesan teks yang mencangkup pembersihan teks, penghapusan *stopword*, dan *stemming* Sastrawi, serta penyeimbangan data menggunakan teknik *oversampling*. Representasi fitur untuk Naïve Bayes dan SVM menggunakan metode TF–IDF, sedangkan BiLSTM menggunakan representasi sekuensial berbasis token. Evaluasi model dilakukan menggunakan metrik *accuracy*, *macro F1-score*, *confusion matrix*, dan ROC–AUC.

Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* memberikan performa terbaik dibandingkan *Multinomial Naïve Bayes* dan *Bidirectional Long Short-Term Memory* pada *dataset* yang digunakan. Temuan ini menunjukkan bahwa pada kondisi data terbatas dan teks pendek, metode pembelajaran mesin klasik masih efektif untuk mengklasifikasi gender berbasis teks *tweet* berbahasa Indonesia.

Kata kunci: klasifikasi gender, Twitter, pemrosesan bahasa alami, *Naïve Bayes*, *Support Vector Machine*, BiLSTM

GLOSARIUM

Accuracy	Ukuran kinerja model klasifikasi yang menunjukkan proporsi prediksi yang benar terhadap keseluruhan jumlah data.
Author Profiling	Proses identifikasi atribut demografis penulis, seperti jenis kelamin atau usia, berdasarkan karakteristik linguistik teks yang dihasilkan.
BiLSTM	Model pembelajaran mendalam berbasis jaringan saraf sekuensial yang memproses data teks dari dua arah, yaitu maju dan mundur, untuk menangkap konteks secara lebih menyeluruh.
Class Imbalance	Kondisi ketika distribusi jumlah data antar kelas tidak seimbang, yang dapat menyebabkan model klasifikasi bias terhadap kelas mayoritas.
Confusion Matrix	Matriks evaluasi yang digunakan untuk menggambarkan perbandingan antara hasil prediksi model dan label sebenarnya dalam bentuk *True Positive*, *False Positive*, *True Negative*, dan *False Negative*.
F1-Score	Metrik evaluasi yang merupakan rata-rata harmonik antara *precision* dan *recall*, digunakan untuk menilai keseimbangan kinerja model klasifikasi.
Macro F1-Score	Nilai rata-rata F1-score yang dihitung secara terpisah untuk setiap kelas, sehingga memberikan bobot yang sama pada setiap kelas dalam evaluasi model.
Naïve Bayes	Algoritma klasifikasi probabilistik yang mengasumsikan independensi antar fitur dan banyak digunakan pada klasifikasi teks.
NLP	Bidang ilmu komputer yang berfokus pada pemrosesan dan analisis bahasa alami menggunakan pendekatan komputasional.
Precision	Ukuran ketepatan prediksi positif yang menunjukkan proporsi prediksi positif yang benar dibandingkan seluruh prediksi positif.
Recall	Ukuran kemampuan model dalam mengidentifikasi seluruh data positif yang ada dalam *dataset*.
Stemming	Proses mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan, untuk mengurangi variasi kata dalam teks.
SVM	Algoritma pembelajaran mesin yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kelas yang berbeda.

TF-IDF	Metode representasi teks yang mengukur tingkat kepentingan suatu kata dalam dokumen relatif terhadap seluruh korpus.
Twitter (X)	Platform media sosial berbasis mikroblog yang memungkinkan pengguna membagikan pesan singkat dalam bentuk teks.

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN DOSEN PEMBIMBING ii
HALAMAN PENGESAHAN DOSEN PENGUJI iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR iv
HALAMAN PERSEMBAHAN v
HALAMAN MOTO vi
KATA PENGANTAR vii
SARI viii
GLOSARIUM ix
DAFTAR ISI xi
DAFTAR TABEL xiii
DAFTAR GAMBAR xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 3
1.6 Sistematika Penulisan 4
BAB II TINJAUAN PUSTAKA 6
2.1 Landasan Teori 6
2.1.1 *Author Profiling* 6
2.1.2 Klasifikasi Gender Berbasis Teks 7
2.1.3 *Text Preprocessing* Bahasa Indonesia 7
2.1.4 *Term Frequency–Inverse Document Frequency* (TF-IDF) 8
2.1.5 *Multinominal Naïve Bayes* 10
2.1.6 *Support Vector Machine* 11
2.1.7 *Bidirectional Long Short-Term Memory* 12
2.1.8 *Confusion Matrix* 13
2.1.9 Ketidakseimbangan Kelas dan *Oversampling* 15
2.2 Penelitian Terdahulu 15
BAB III METODOLOGI PENELITIAN 23
3.1 Tahapan Penelitian 23
3.2 Pengumpulan Data 24
3.3 Pra-pemrosesan Data 26
3.4 Penyeimbangan Data 27
3.5 Pembagian Data 27
3.6 Ekstraksi Fitur 28
3.6.1 TF-IDF 28
3.6.2 *Tokenizer* dan *Padding* 29
3.7 Pemodelan Klasifikasi 29
3.7.1 *Multinominal Naïve Bayes* 29
3.7.2 *Support Vector Machine* 30
3.7.3 *Bidirectional Long Short-Term Memory* 30
3.8 Evaluasi Model 31
BAB IV HASIL DAN PEMBAHASAN 32
4.1 Hasil Pengumpulan Data 32

4.2 Hasil Pelabelan Gender 33
4.3 Hasil Pra-Pemrosesan Data 34
4.4 Hasil Pembagian dan Penyeimbangan Data 35
4.5 Hasil Ekstraksi Fitur 36
4.6 Hasil Pengujian Model Klasifikasi 37
4.6.1 Metrik Evaluasi 38
4.6.2 Hasil Evaluasi Kuantitatif 39
4.6.3 Analisis Perbandingan Antar Model 40
4.7 Analisis *Confison Matrix* 40
4.8 Analisis Kurva ROC dan Nilai AOC 42
4.9 Analisis Proses Pelatihan BiLSTM 43
4.10 Contoh *Tweet* dan Hasil Prediksi Model 45
4.10.1 Contoh *Tweet* Gender Perempuan 45
4.10.2 Contoh *Tweet* Gender Laki-laki 46
4.11 Pembahasan Hasil Penelitian 47
4.12 Kelebihan dan Keterbatasan Penelitian 48
BAB V PENUTUP 50
5.1 Kesimpulan 50
5.2 Saran 51
DAFTAR PUSTAKA 52
LAMPIRAN 55

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu terkait klasifikasi gender berbasis teks 17
Tabel 2. 2 Pengelompokan Penelitian Terdahulu Berdasarkan Metode 20
Tabel 2. 3 Penelitian Terdahulu Berdasarkan Akurasi Model 21
Tabel 3. 1 Contoh data *tweet* 25
Tabel 3. 2 Contoh data hasil pra-pemrosesan 26
Tabel 4. 1 Contoh *raw* data hasil *scrapping* 33
Tabel 4. 2 Contoh hasil dari pra-pemrosesan teks 34
Tabel 4. 3 Hasil Evaluasi Model Klasifikasi 39
Tabel 4. 4 Komparasi hasil *confusion matrix* model 42
Tabel 4. 5 Contoh *tweet* dan hasil prediksi model 45
Tabel 4. 6 Contoh *tweet* gender perempuan 46
Tabel 4. 7 Contoh *tweet* gender laki-laki 47

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian 24
Gambar 4. 1 Kode *scrapping* data 32
Gambar 4. 2 Kode program pra-pemrosesan data 34
Gambar 4. 3 Kode program tahap pembagian data.....35
Gambar 4. 4 Kode program tahap penyeimbangan data.....35
Gambar 4. 5 Grafik perbandingan sebelum dan sesudah penyeimbangan data.....36
Gambar 4. 6 Kode program tahap ekstraksi fitur.....36
Gambar 4. 7 Kode program tahap permodelan.....38
Gambar 4. 8 Kode program tahap evaluasi model.....39
Gambar 4. 9 Perbandingan akurasi model (NB, SVM, BiLSTM).....40
Gambar 4. 11 *Confusion Matrix Support Vector Machine*.....41
Gambar 4. 10 *Confusion Matrix Multinominal Naive Bayes*.....41
Gambar 4. 12 *Confusion Matrix Bidirectional* LSTM.....41
Gambar 4. 13 Kurva ROC 43
Gambar 4. 14 Kurva *loss* pelatihan BiLSTM 44
Gambar 4. 15 Kurva akurasi pelatihan BiLSTM.....44

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan media sosial telah menghasilkan volume data teks yang sangat besar dan bersifat dinamis. Platform X (Twitter) menjadi salah satu media sosial yang banyak dimanfaatkan sebagai sumber data penelitian karena menyediakan teks singkat yang bersifat publik, *real-time*, dan mencerminkan ekspresi linguistik pengguna secara alami (Burger dkk., 2011; O'Connor dkk., 2024). Karakteristik teks Twitter yang ringkas, informal, serta kaya akan variasi bahasa menjadikannya relevan untuk kajian pemrosesan bahasa alami (*Natural Language Processing*), sekaligus menimbulkan tantangan dalam proses analisis otomatis.

Analisis gender pada media sosial memiliki nilai penting karena informasi demografis tidak selalu tersedia atau akurat pada profil pengguna. Dengan mengklasifikasikan gender berdasarkan pola bahasa yang muncul dalam tweet, penelitian ini dapat memberikan gambaran perilaku linguistik pengguna secara lebih mendalam. Selain itu, identifikasi gender berperan penting dalam berbagai aplikasi seperti personalisasi konten, pemasaran digital, segmentasi audiens, hingga kajian sosial berbasis data. Oleh sebab itu, analisis gender dipilih sebagai fokus penelitian untuk mengevaluasi sejauh mana pola linguistik pada teks pendek berbahasa Indonesia dapat digunakan untuk menginferensi atribut demografis secara otomatis.

Pendekatan pembelajaran mesin klasik seperti *Naïve Bayes* dan *Support Vector Machine* (SVM) masih sering digunakan dalam klasifikasi teks pendek karena efisiensinya serta kemampuannya dalam menangani representasi fitur berdimensi tinggi seperti *Term Frequency–Inverse Document Frequency* (TF-IDF) (Fikri dkk., 2020; Joachims, 1998; Katna dkk., 2022). Di sisi lain, perkembangan pembelajaran mendalam mendorong penggunaan model sekuensial seperti *Bidirectional Long Short-Term Memory* (BiLSTM), yang mampu menangkap konteks dan urutan kata dalam teks (Masood dkk., 2023). Namun, sejumlah penelitian menunjukkan bahwa model pembelajaran mendalam memerlukan data berukuran besar agar dapat menghasilkan performa yang stabil dan optimal (Fields dkk., 2024; Tasnim dkk., 2024).

Dalam konteks bahasa Indonesia, klasifikasi gender berbasis teks menghadapi tantangan tambahan akibat tingginya tingkat informalitas bahasa pada media sosial. *Tweet* berbahasa Indonesia sering mengandung ejaan tidak baku, singkatan, slang, serta pencampuran bahasa. Penelitian sebelumnya menunjukkan bahwa penerapan *preprocessing* yang sesuai, seperti

penghapusan *stopword* dan *stemming* menggunakan algoritma Sastrawi, dapat meningkatkan konsistensi fitur dan performa klasifikasi teks berbahasa Indonesia (Helmayanti dkk., 2023; Rianto dkk., 2021; Rosid dkk., 2020).

Selain tantangan linguistik, permasalahan ketidakseimbangan kelas (*class imbalance*) juga umum ditemukan dalam data demografis. Ketidakseimbangan distribusi kelas dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga diperlukan teknik penyeimbangan data. Salah satu pendekatan yang umum digunakan adalah *oversampling* pada data latih untuk meningkatkan stabilitas model klasifikasi (Benítez-Andrades dkk., 2025; Nhita dkk., 2023).

Meskipun penelitian terkait klasifikasi gender di media sosial telah banyak dilakukan, studi yang secara langsung membandingkan metode klasik dan model pembelajaran mendalam pada data Twitter berbahasa Indonesia dengan alur metodologi yang seragam masih relatif terbatas. Sebagian penelitian mengombinasikan fitur teks dengan metadata profil pengguna, sehingga evaluasi kemampuan klasifikasi berbasis teks secara murni menjadi kurang jelas (Mahendra dkk., 2022; Selma Zakia, 2020).

Berdasarkan kondisi tersebut, penelitian ini dilakukan untuk membandingkan kinerja algoritma *Naïve Bayes*, *Linear Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* dalam mengklasifikasikan gender pengguna Twitter berdasarkan teks *tweet* berbahasa Indonesia. Penelitian ini menerapkan alur pemrosesan data yang terstandar agar perbedaan performa yang diperoleh dapat dianalisis secara objektif. Penelitian ini juga berkontribusi sebagai *benchmark* awal pada skala *dataset* terbatas untuk klasifikasi gender berbasis teks pada *tweet* berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini difokuskan untuk menjawab permasalahan terkait klasifikasi gender pengguna media sosial berbasis teks *tweet* berbahasa Indonesia. Permasalahan penelitian tersebut dirumuskan dalam bentuk pertanyaan penelitian sebagai berikut:

a. Bagaimana kinerja algoritma *Multinominal Naïve Bayes* dalam mengklasifikasikan gender pengguna Twitter berdasarkan teks *tweet* berbahasa Indonesia?
b. Bagaimana kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan gender pengguna Twitter berdasarkan teks *tweet* berbahasa Indonesia?

c. Bagaimana kinerja model *Bidirectional Long Short-Term Memory* (BiLSTM) dalam mengklasifikasikan gender pengguna Twitter berdasarkan teks *tweet* berbahasa Indonesia?
d. Algoritma manakah di antara ketiga algoritma di atas yang memberikan performa terbaik dalam klasifikasi gender berbasis teks berbahasa Indonesia berdasarkan metrik akurasi dan *macro F1-score*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

a. Data yang digunakan berupa *tweet* berbahasa Indonesia dari platform X.
b. Pengumpulan data dilakukan menggunakan beberapa kata kunci tertentu, yaitu "gibran", "kuliah", "kerja", "*parenting*", "*overthinking*", "pinjol", "banjir".
c. Klasifikasi gender dibatasi pada dua kelas, yaitu laki-laki dan perempuan.
d. Pelabelan gender dilakukan secara manual berdasarkan informasi profil publik pengguna.
e. Fitur klasifikasi pada algoritma hanya berasal dari teks *tweet* tanpa melibatkan metadata tambahan.
f. Algoritma yang digunakan adalah *Multinominal Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory*.
g. Evaluasi model menggunakan metrik akurasi, *macro F1-score*, *confusion matrix*, dan ROC–AUC.

1.4 Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah:

a. Menganalisis performa *Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* dalam klasifikasi gender berbasis teks tweet.
b. Membandingkan ketiga algoritma untuk menentukan metode yang paling efektif pada *dataset* berukuran terbatas.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

a. Memberikan kontribusi ilmiah dalam bidang *author profiling* dan NLP berbahasa Indonesia dengan pendekatan perbandingan metode.

b. Memberikan rekomendasi metode klasifikasi gender yang efektif dan efisien pada kondisi keterbatasan data.
c. Menyediakan alur eksperimen yang terstandar dan dapat direplikasi untuk penelitian NLP di masa mendatang.

1.6 Metodologi Penelitian Secara Umum

Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan tujuan membandingkan kinerja beberapa algoritma klasifikasi dalam memprediksi gender pengguna media sosial berdasarkan teks tweet berbahasa Indonesia. Pendekatan ini dipilih karena memungkinkan pengukuran performa model secara objektif melalui metrik evaluasi yang terstandar.

Tahapan penelitian secara umum meliputi pengumpulan data *tweet***,** pelabelan gender secara manual**,** pra-pemrosesan teks**,** penyeimbangan data**,** ekstraksi fitur**,** pembangunan model klasifikasi, serta evaluasi kinerja model**.** Data tweet dikumpulkan dari platform X menggunakan kata kunci tertentu untuk memperoleh variasi topik yang beragam. Pelabelan gender dilakukan berdasarkan informasi publik pada profil pengguna, seperti nama dan username.

Pada tahap pemodelan, penelitian ini menerapkan tiga algoritma klasifikasi yang berbeda, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* (BiLSTM). Pemilihan ketiga algoritma tersebut bertujuan untuk membandingkan performa metode pembelajaran mesin klasik dan pembelajaran mendalam pada kondisi data terbatas dan teks yang bersifat informal.

Evaluasi performa model dilakukan menggunakan beberapa metrik, antara lain akurasi, *macro F1-score*, *confusion matrix*, dan kurva ROC-AUC. Dengan alur metodologi yang terstandar dan konsisten, hasil penelitian diharapkan mampu memberikan gambaran yang objektif mengenai efektivitas masing-masing algoritma dalam klasifikasi gender berbasis teks tweet berbahasa Indonesia.

1.7 Sistematika Penulisan

Penulisan laporan tugas akhir ini disusun dengan sistematika sebagai berikut:

BAB I Pendahuluan, berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II Tinjauan Pustaka, membahas penelitian terdahulu dan landasan teori yang relevan dengan klasifikasi gender berbasis teks.

BAB III Metodologi Penelitian, menjelaskan tahapan penelitian mulai dari pengumpulan data hingga metode evaluasi yang digunakan.

BAB IV Hasil dan Pembahasan, menyajikan hasil dari masing-masing tahapan penelitian, hasil pengujian model, serta analisis dan pembahasan terhadap hasil yang diperoleh.

BAB V Kesimpulan dan Saran, berisi kesimpulan penelitian dan saran untuk pengembangan penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 *Author Profiling*

Author profiling merupakan cabang penelitian dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi atribut demografis penulis, seperti jenis kelamin dan usia, berdasarkan karakteristik linguistik dari teks yang dihasilkan. Pendekatan ini didasarkan pada asumsi bahwa setiap individu memiliki pola penggunaan bahasa yang relatif konsisten, baik dari segi pemilihan kosakata, struktur kalimat, maupun gaya penulisan, sehingga informasi demografis penulis dapat diinferensi dari teks yang diproduksinya(Burger dkk., 2011; Cheng dkk., 2011).

Tujuan utama dilakukannya *author profiling* adalah untuk mengekstraksi informasi implisit tentang penulis ketika informasi eksplisit tidak tersedia, tidak lengkap, atau tidak dapat dipercaya. Dalam konteks media sosial, khususnya platform X (Twitter), pengguna sering kali tidak mencantumkan atribut demografis secara jelas atau bahkan menyembunyikan identitas aslinya. Oleh karena itu, *author profiling* menjadi pendekatan alternatif yang relevan untuk menginferensi karakteristik pengguna hanya berdasarkan konten teks yang dipublikasikan (O'Connor dkk., 2024)

Dalam praktiknya, *author profiling* memiliki berbagai aplikasi, antara lain untuk analisis perilaku pengguna, personalisasi konten dan pemasaran digital, kajian sosial dan linguistik, serta forensik digital dan keamanan siber. Dalam penelitian akademik, *author profiling* juga sering digunakan sebagai fondasi untuk analisis lanjutan, seperti klasifikasi sentimen berbasis demografi dan pemetaan opini publik. Pada penelitian ini, *author profiling* difokuskan pada klasifikasi gender pengguna Twitter berbahasa Indonesia berdasarkan teks *tweet*, sebagai upaya untuk mengevaluasi sejauh mana karakteristik linguistik pada teks pendek dapat digunakan untuk menginferensi atribut gender dalam kondisi keterbatasan informasi profil (Mahendra dkk., 2022)

2.1.2 Klasifikasi Gender Berbasis Teks

Klasifikasi gender berbasis teks merupakan proses pengelompokan dokumen teks ke dalam kelas laki-laki atau perempuan dengan memanfaatkan pola linguistik dalam tulisan. Penelitian sebelumnya menunjukkan bahwa perbedaan frekuensi kata, penggunaan kata fungsi, struktur kalimat, serta gaya ekspresi dapat digunakan sebagai indikator gender dalam teks, sehingga atribut gender dapat diinferensi tanpa bergantung pada informasi profil secara eksplisit (Cheng dkk., 2011).

Pendekatan ini banyak diterapkan pada data media sosial karena teks yang dihasilkan pengguna bersifat personal dan mencerminkan kebiasaan berbahasa sehari-hari. Tujuan utama klasifikasi gender berbasis teks adalah untuk mengidentifikasi atribut demografis pengguna secara otomatis ketika informasi profil tidak tersedia, tidak lengkap, atau tidak dapat dipercaya. Hasil klasifikasi ini dapat dimanfaatkan untuk analisis perilaku pengguna, segmentasi audiens, personalisasi konten, serta kajian sosial dan *author profiling* berbasis data teks (Burger dkk., 2011; O'Connor dkk., 2024)

Namun, klasifikasi gender berbasis teks menjadi lebih menantang ketika diterapkan pada teks pendek dan informal seperti *tweet*, karena keterbatasan konteks linguistik serta tingginya variasi bahasa. Tantangan ini semakin kompleks dalam konteks bahasa Indonesia, sehingga keberhasilan klasifikasi sangat dipengaruhi oleh kualitas pra-pemrosesan dan pemilihan metode yang sesuai dengan karakteristik data dan ukuran *dataset* (Mahendra dkk., 2022).

2.1.3 *Text Preprocessing* Bahasa Indonesia

Pra-pemrosesan (*text preprocessing*) merupakan tahapan penting dalam pemrosesan bahasa alami yang bertujuan untuk membersihkan dan menormalkan teks sebelum dilakukan ekstraksi fitur dan pemodelan. Pada teks berbahasa Indonesia, khususnya yang berasal dari media sosial, tahap pra-pemrosesan menjadi krusial karena karakteristik bahasa yang cenderung informal, tidak baku, serta kaya variasi morfologis. Secara umum, tahapan pra-pemrosesan teks bahasa Indonesia meliputi beberapa langkah berikut.

a. *Case Folding*

Case folding merupakan proses mengubah seluruh karakter dalam teks menjadi huruf kecil. Tahap ini bertujuan untuk menghindari perbedaan representasi kata akibat variasi penggunaan huruf kapital, sehingga kata dengan makna yang sama tidak diperlukan sebagai fitur yang berbeda (Rianto dkk., 2021)

b. Pembersihan *Noise*

Tahap ini mencakup penghapusan elemen-elemen non-linguistik yang tidak relevan, seperti URL, *mention*, *hashtag*, angka, tanda baca, serta simbol khusus. Pembersihan noise bertujuan untuk mengurangi gangguan pada proses ekstraksi fitur dan meningkatkan kualitas representasi teks, terutama pada data media sosial yang mengandung banyak elemen tambahan di luar kata (Rosid dkk., 2020).

c. Normalisasi Kata Tidak Baku

Normalisasi dilakukan untuk mengonversi kata-kata tidak baku, singkatan, atau slang ke bentuk kata baku. Pada teks media sosial berbahasa Indonesia, variasi penulisan seperti "gak", "nggak", atau "udh" sering ditemukan dan perlu diseragamkan agar tidak menimbulkan redundansi fitur. Proses normalisasi bertujuan untuk meningkatkan konsistensi kosakata dan memperbaiki kualitas analisis linguistik (Rianto dkk., 2021).

d. *Stopword Removal*

Stopword removal merupakan proses penghapusan kata-kata umum yang sering muncul namun memiliki nilai diskriminatif yang rendah, seperti kata hubung dan kata depan. Penghapusan stopword bertujuan untuk mengurangi dimensi fitur serta memfokuskan model pada kata-kata yang lebih informatif dalam proses klasifikasi teks (Helmayanti dkk., 2023).

e. *Stemming* Bahasa Indonesia

Stemming bertujuan untuk mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan awalan, sisipan, dan akhiran. Dalam bahasa Indonesia, proses *stemming* sangat penting karena tingginya variasi morfologis. Algoritma *stemming* seperti Sastrawi terbukti efektif dalam meningkatkan konsistensi representasi kata dan kinerja klasifikasi teks berbahasa Indonesia (Rianto dkk., 2021; Rosid dkk., 2020).

Secara keseluruhan, tahapan pra-pemrosesan teks bahasa Indonesia bertujuan untuk mengurangi *noise*, menyederhanakan representasi linguistik, dan meningkatkan kualitas fitur yang digunakan dalam pemodelan. Penerapan kombinasi tahapan pra-pemrosesan yang tepat menjadi faktor penting dalam keberhasilan berbagai tugas klasifikasi teks berbahasa Indonesia, khususnya pada data media sosial yang bersifat informal.

2.1.4 *Term Frequency–Inverse Document Frequency* (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode representasi teks yang mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan korpus. TF-IDF bekerja dengan mengombinasikan dua komponen, yaitu

term frequency (TF) yang merepresentasikan frekuensi kemunculan kata dalam dokumen, dan *inverse document frequency* (IDF) yang mengurangi bobot kata-kata yang sering muncul di banyak dokumen. Representasi ini efektif untuk menonjolkan kata-kata yang bersifat diskriminatif dalam klasifikasi teks (Katna dkk., 2022).

Secara umum, TF-IDF terdiri dari tiga komponen utama, yaitu TF, IDF, dan perkalian keduanya.

a. *Term Frequency*

Term Frequency mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. TF dirumuskan pada rumus 2.1:

$$TF(t,d) = \frac{jumlah\ kata\ t\ pada\ dokumen\ d}{total\ kata\ pada\ dokumen\ d} \quad (2.1)$$

di mana:

t : kata

d : dokumen

Nilai TF yang tinggi menunjukkan bahwa suatu kata sering muncul dalam dokumen tersebut dan berpotensi penting untuk merepresentasikan isi dokumen.

b. *Inverse Document Frequency*

IDF mengukur seberapa jarang suatu kata muncul di seluruh dokumen dalam korpus. IDF dirumuskan pada rumus 2.2:

$$IDF(t) = \log\frac{N}{df_t} \quad (2.2)$$

dengan:

N : jumlah seluruh dokumen

df_t : jumlah dokumen yang mengandung kata t

IDF bertujuan untuk menurunkan bobot kata-kata yang sering muncul di banyak dokumen, karena kata-kata tersebut umumnya kurang memiliki daya pembeda.

c. *Term Frequency-Inverse Document Frequency*

Nilai TF-IDF diperoleh dari hasil perkalian antara TF dan IDF, yang dirumuskan pada rumus 2.3:

$$TF - IDF(t,d) = TF(t,d) \cdot IDF(t) \quad (2.3)$$

Nilai TF-IDF yang tinggi menunjukkan bahwa suatu kata memiliki frekuensi kemunculan yang signifikan dalam dokumen tertentu, namun relatif jarang muncul di dokumen lain, sehingga kata tersebut dianggap lebih informatif.

Dalam konteks klasifikasi teks, representasi TF-IDF banyak digunakan karena mampu mengubah teks menjadi vektor numerik berdimensi tinggi yang dapat diproses oleh algoritma pembelajaran mesin. Representasi ini terbukti efektif untuk tugas klasifikasi teks pendek, termasuk klasifikasi gender berbasis teks, terutama ketika dikombinasikan dengan algoritma pembelajaran mesin klasik seperti *Naïve Bayes* dan *Support Vector Machine* (Katna dkk., 2022; Manning dkk., 2008)

2.1.5 *Multinominal Naïve Bayes*

Multinomial Naïve Bayes merupakan algoritma klasifikasi probabilistik yang mengasumsikan independensi antar fitur berdasarkan Teorema Bayes. Pada klasifikasi teks, model ini menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan distribusi frekuensi kata pada setiap kelas. Varian multinomial dari *Naïve Bayes* secara khusus dirancang untuk data diskrit seperti jumlah kemunculan kata, sehingga sangat sesuai untuk representasi TF-IDF (Cheng dkk., 2011).

Secara matematis, Teorema Bayes dirumuskan sebagaimana ditunjukkan rumus 2.4:

$$P(c \mid d) = \frac{P(d \mid c) \cdot P(c)}{(P(d)} \quad (2.4)$$

dengan:

c : kelas

d : dokumen

$P(c|d)$: probabilitas dokumen d termasuk ke kelas c

$P(d|c)$: probabilitas dokumen d muncul pada kelas c

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

$P(d)$: probabilitas dokumen d

Dalam *Multinomial Naïve Bayes*, probabilitas $P(d|c)$ dihitung berdasarkan distribusi kata dalam dokumen. Dengan asumsi independensi antar kata, probabilitas tersebut dapat dinyatakan seperti pada rumus 2.5:

$$P(d \mid c) = \sum_{i=1}^{n} P(w_i|c)^{f_{i,d}} \quad (2.5)$$

dengan:

w_i : kata ke-i dalam kosakata

$f_{i,d}$: frekuensi kemunculan kata $w_i$dalam dokumen d

$P(w_i|c)$: frekuensi kemunculan kata w_i pada kelas c

Untuk menghindari probabilitas nol akibat kata yang tidak pernah muncul pada suatu kelas, digunakan teknik *Laplace smoothing*, yang dirumuskan pada rumus 2.6:

$$P(w_i|c) = \frac{N_{i,c} + 1}{N_c + |V|} \quad (2.6)$$

dengan:

$N_{i,c}$: jumlah kemunculan kata w_i pada kelas c

N_c : total jumlah seluruh kata pada kelas c

$|V|$: ukuran kosakata

Pada proses klasifikasi, dokumen akan diberi label kelas dengan nilai probabilitas posterior tertinggi. *Multinomial Naïve Bayes* dikenal memiliki keunggulan dari sisi efisiensi komputasi dan stabilitas performa pada dataset berukuran kecil hingga menengah. Meskipun asumsi independensi antar kata bersifat sederhana, algoritma ini terbukti efektif untuk klasifikasi teks pendek dan sering digunakan sebagai baseline pada berbagai penelitian klasifikasi teks (Cheng dkk., 2011; McCallum & Nigam, 1998).

2.1.6 *Support Vector Machine*

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan kelas dengan margin maksimum. SVM terbukti efektif pada data berdimensi tinggi dan sering memberikan performa unggul pada

klasifikasi teks berbasis TF-IDF. Adanya *slack variable* memungkinkan SVM menangani data yang tidak sepenuhnya terpisah secara linear. (Cortes & Vapnik, 1995; Katna dkk., 2022).

a. Fungsi Keputusan

$$f(x) = w \cdot x + b \quad (2.7)$$

dengan:

x : vektor fitur

w : vektor bobot

b : bias

b. Optimasi *Soft* Margin

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi i \quad (2.8)$$

dengan kendala:

$$yi(w \cdot xi + b) \geq 1 - \xi i \quad (2.9)$$

keterangan:

ξi : *slack variable* (kesalahan)

C : parameter regularisasi

$yi \in \{-1, +1\}$: label kelas

2.1.7 *Bidirectional Long Short-Term Memory*

Bidirectional LSTM merupakan pengembangan dari LSTM yang memproses urutan data dari dua arah. Model ini mampu menangkap konteks masa lalu dan masa depan dalam teks secara simultan, sehingga secara teoritis lebih unggul dalam memahami struktur sekuensial bahasa (Hochreiter & Schmidhuber, 1997; Masood dkk., 2023). Namun, performa BiLSTM sangat bergantung pada ukuran dan kualitas *dataset* (Tasnim dkk., 2024).

a. Mekanisme Kerja LSTM

Setiap unit LSTM terdiri dari tiga gerbang utama, yaitu *forget gate*, *input gate*, dan *output gate*, yang berfungsi untuk mengatur aliran informasi ke dalam dan ke luar memori sel. Secara matematis, mekanisme inti LSTM dirumuskan pada rumus 2.10 – 2.15:

a. *Forget Gate*

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.10)$$

b. *Input Gate*

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.11)$$

c. *Candidate Cell State*

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.12)$$

d. *Cell State Update*

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (2.13)$$

e. *Output Gate*

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.14)$$

f. *Hidden State*

$$h_t = o_t \odot \tanh(C_t) \quad (2.15)$$

dengan:

x_t : vektor input pada waktu ke-t

h_{t-1}, h_t : *hidden state* sebelumnya dan saat ini

C_{t-1}, C_t : *cell state* sebelumnya dan saat ini

W dan b : bobot dan bias

σ : fungsi aktivasi *sigmoid*

tanh : fungsi aktivasi hiperbolik

$\odot$: operasi perkalian elemen

b. Arsitektur *Bidirectional* LSTM

Pada BiLSTM, dua LSTM dijalankan secara paralel, yaitu *Forword* LSTM yang memproses urutan dari awal ke akhir teks, dan *Backward* LSTM yang memproses urutan dari akhir ke awal teks. *Output* dari kedua arah tersebut kemudian digabungkan untuk menghasilkan representasi akhir. Pendekatan ini memungkinkan model menangkap ketergantungan konteks jangka pendek maupun jangka panjang secara lebih efektif dibandingkan LSTM satu arah (Masood dkk., 2023).

2.1.8 *Confusion Matrix*

Confusion matrix merupakan metode evaluasi yang digunakan untuk menggambarkan kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap label sebenarnya. *Confusion matrix* terdiri dari empat komponen utama, yaitu:

a. *True Positive* (TP): data positif yang diprediksi positif

b. *False Positive* (FP): data negatif yang diprediksi positif
c. *True Negative* (TN): data negatif yang diprediksi negatif
d. *False Negative* (FN): data positif yang diprediksi negatif

Berdasarkan confusion matrix, beberapa metrik evaluasi dapat dihitung sebagai berikut:

a. Akurasi

Akurasi mengukur proporsi keseluruhan prediksi yang benar terhadap total jumlah data. Metrik ini dirumuskan seperti rumus 2.16:

$$Accuracy = (TP + TN)/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \quad (2.16)$$

Meskipun akurasi mudah dipahami, metrik ini dapat memberikan gambaran yang menyesatkan pada kondisi data yang tidak seimbang (*imbalanced* data), karena model dapat mencapai akurasi tinggi hanya dengan memprediksi kelas mayoritas.

b. Presisi

Presisi mengukur ketepatan prediksi positif yang dihasilkan oleh model, yaitu seberapa banyak prediksi positif yang benar dibandingkan seluruh prediksi positif. Nilai presisi yang tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah. Presisi dirumuskan seperti rumus 2.17:

$$Precission = TP/(\mathrm{TP} + \mathrm{FP}) \quad (2.17)$$

c. *Recall*

Recall mengukur kemampuan model dalam menemukan seluruh data positif yang ada, yaitu seberapa banyak data positif yang berhasil diprediksi dengan benar. Nilai recall yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan *false negative*. *Recall* dirumuskan pada rumus 2.18:

$$Recall = TP/(\mathrm{TP} + \mathrm{FN}) \quad (2.18)$$

d. *F1-Score*

F1-score merupakan rata-rata harmonik dari presisi dan *recall*, sehingga mampu menyeimbangkan kedua metrik tersebut. Metrik ini sangat berguna ketika diperlukan

keseimbangan antara presisi dan *recall*, terutama pada kasus klasifikasi dengan distribusi kelas yang tidak seimbang. *F1-score* dirumuskan pada rumus 2.19:

$$F1 = \frac{2 \cdot Precission \cdot Recall}{Precission + Recall} \quad (2.19)$$

Dalam penelitian ini digunakan *macro F1-score*, yaitu rata-rata nilai *F1-score* yang dihitung secara terpisah untuk setiap kelas. Penggunaan *macro F1-score* bertujuan memberikan bobot yang sama pada setiap kelas, sehingga performa model terhadap kelas minoritas tetap diperhitungkan secara adil. Dengan demikian, evaluasi kinerja model tidak didominasi oleh kelas mayoritas dan lebih representatif untuk tugas klasifikasi gender pada *dataset* yang distribusi kelasnya tidak sepenuhnya seimbang.

2.1.9 Ketidakseimbangan Kelas dan *Oversampling*

Ketidakseimbangan kelas terjadi ketika jumlah data antar kelas tidak proporsional. Kondisi ini dapat menyebabkan bias model. Teknik *oversampling*, seperti *RandomOverSampler*, digunakan untuk menyeimbangkan distribusi kelas pada data latih dan meningkatkan stabilitas performa model (Benítez-Andrades dkk., 2025).

2.2 Penelitian Terdahulu

Penelitian mengenai *author profiling* dan klasifikasi gender berbasis teks telah berkembang seiring meningkatnya penggunaan media sosial sebagai sumber data linguistik. Platform X (Twitter) sering digunakan karena menyediakan teks singkat yang mencerminkan karakteristik bahasa dan perilaku pengguna secara alami (Burger dkk., 2011).Penelitian awal menunjukkan bahwa pola leksikal dan gaya bahasa dalam teks dapat digunakan untuk mengidentifikasi jenis kelamin penulis dengan tingkat akurasi yang cukup tinggi (Cheng dkk., 2011).

Mahendra dkk. (2022) melakukan klasifikasi gender pengguna Twitter berbahasa Indonesia dengan memanfaatkan kombinasi fitur teks dan metadata profil. Hasil penelitian menunjukkan bahwa integrasi metadata dapat meningkatkan performa klasifikasi, namun pendekatan tersebut menyulitkan evaluasi kemampuan klasifikasi berbasis teks secara murni. Hal ini mengindikasikan perlunya penelitian yang berfokus pada pemanfaatan konten teks saja agar lebih umum diterapkan pada kondisi ketika metadata tidak tersedia, sejalan dengan temuan

Selma Zakia (2020) dan Sa'diah dkk. (2024) yang menunjukkan bahwa klasifikasi gender berbasis teks di Twitter Indonesia masih sangat dipengaruhi oleh ukuran dan kualitas data.

Pendekatan pembelajaran mesin klasik masih banyak digunakan dalam klasifikasi gender berbasis teks. Katna dkk. (2022) menunjukkan bahwa algoritma *Naïve Bayes* dan *Support Vector Machine* mampu memberikan performa yang stabil pada data teks pendek, terutama ketika dikombinasikan dengan representasi fitur TF-IDF. Temuan ini diperkuat oleh studi lintas bahasa yang menyatakan bahwa model klasik sering kali tetap kompetitif pada *dataset* berukuran kecil hingga menengah (Tasnim dkk., 2024).

Seiring perkembangan pembelajaran mendalam, model sekuensial seperti *Long Short-Term Memory* (LSTM) dan *Bidirectional LSTM* (BiLSTM) mulai diterapkan pada tugas klasifikasi gender. Masood dkk. (2023)melaporkan bahwa model berbasis LSTM mampu menangkap konteks linguistik secara lebih baik dibandingkan metode klasik, namun memerlukan data latih dalam jumlah besar agar performanya optimal. Penelitian serupa oleh Sa'diah dkk. (2024) menunjukkan bahwa penerapan LSTM pada data media sosial dapat meningkatkan akurasi, tetapi performa model cenderung menurun ketika *dataset* bersifat terbatas atau tidak seimbang.

Dalam konteks bahasa Indonesia, tantangan linguistik seperti imbuhan, ejaan tidak baku, dan variasi slang memerlukan teknik *preprocessing* khusus. Rianto dkk. (2021) dan Rosid dkk. (2020) menunjukkan bahwa penggunaan *stemming* Sastrawi secara signifikan meningkatkan konsistensi fitur teks dan kinerja klasifikasi pada berbagai tugas NLP berbahasa Indonesia. Selain itu, permasalahan ketidakseimbangan kelas juga sering muncul dalam data demografis. Teknik *oversampling* seperti *RandomOverSampler* direkomendasikan untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan stabilitas model (Benítez-Andrades dkk., 2025).

Meskipun berbagai penelitian telah membahas klasifikasi gender berbasis teks, sebagian besar studi masih berfokus pada satu jenis metode atau mengombinasikan berbagai sumber fitur. Studi yang secara langsung membandingkan metode klasik dan model pembelajaran mendalam dalam satu alur metodologi yang seragam, khususnya pada data Twitter berbahasa Indonesia, masih relatif terbatas.

Tabel 2. 1 Penelitian terdahulu terkait klasifikasi gender berbasis teks

No	Referensi	Bahasa & Platform	Fitur Utama	Metode/Model	Karakteristik Data (garis besar)	Temuan Utama Singkat
1	O'Connor dkk. (2024)	Multibahasa, Twitter	Konten *tweet*, metadata, fitur perilaku	Beragam ML & DL untuk survei	Berbagai korpus gender & usia	*Scoping review* yang merangkum metode dan *dataset* untuk prediksi usia & gender pengguna Twitter; menekankan pentingnya kualitas anotasi dan desain *dataset*.
3	Mahendra dkk. (2022)	Indonesia, Twitter	Konten *tweet* + fitur profil	ML klasik	Data pengguna Twitter berbahasa Indonesia berlabel gender	Menggunakan kombinasi teks dan metadata; hasil baik tetapi sulit mengisolasi kontribusi teks murni terhadap performa prediksi gender.
4	Katna dkk. (2022)	Inggris, Twitter	TF–IDF, fitur leksikal	NB, SVM, model ML lain	*Tweet* dengan label usia & gender	Menegaskan bahwa SVM dan model klasik tetap kuat untuk prediksi usia & gender berbasis teks pendek.
5	Cheng dkk. (2011)	Inggris, teks umum	Fitur leksikal & gaya bahasa	Klasik ML	*Dataset* teks untuk identifikasi gender penulis	Penelitian awal penting yang menunjukkan bahwa pola leksikal dan gaya cukup untuk mengidentifikasi gender penulis.
6	Sa'diah dkk. (2024)	Indonesia, pesan media sosial	*fastText embeddings*	LSTM	Teks pendek berbahasa Indonesia	Kombinasi *fastText* + LSTM meningkatkan performa

						klasifikasi gender, namun membutuhkan data yang cukup besar.
7	Tasnim dkk. (2024)	Bangla, media sosial	Fitur leksikal, *deep features*	ML klasik vs model hibrida DL	Korpus Bangla *author profiling*	Di skenario *low-resource*, model klasik sering menyaingi atau mengungguli DL; pentingnya *benchmark* terstandar ditekankan.
8	Fields dkk. (2024)	Beragam bahasa & domain	Representasi Transformer	Berbagai arsitektur Transformer	Studi survei multi-tugas	Survei luas yang menunjukkan kekuatan Transformer, tetapi juga menyoroti biaya komputasi dan kebutuhan data yang besar.
9	Selma Zakia (2020)	Indonesia, Twitter	BM25, TF	BM25 + KNN	Teks *tweet* berbahasa Indonesia	Mengkaji klasifikasi gender dengan BM25 dan KNN; belum membandingkan langsung dengan model neural di *pipeline* terstandar.
10	Humam & Laksito (2023)	Indonesia, Twitter	TF–IDF	ML klasifikasi sentimen	Data sentimen menjelang Pemilu 2024	Fokus pada sentimen politik, menunjukkan bahwa *pipeline* TF–IDF + ML efektif untuk teks politik berbahasa Indonesia.
11	Helmayanti dkk. (2023)	Indonesia, ulasan aplikasi	TF–IDF, fitur aspek	*Naïve Bayes*	Ulasan aplikasi Flip	Menunjukkan NB + TF–IDF cocok untuk analisis sentimen

						aspek pada teks berbahasa Indonesia yang informal.
12	Rosid dkk. (2020)	Indonesia, dokumen keluhan mahasiswa	Teks + Sastrawi	Berbagai model klasifikasi	Dokumen keluhan berbahasa Indonesia	Membuktikan bahwa Sastrawi dan pra-pemrosesan yang baik dapat meningkatkan akurasi klasifikasi teks berbahasa Indonesia.
13	Burger dkk. (2011)	Inggris, Twitter	Fitur konten & profil	ML klasik	*Tweet* berlabel gender	Salah satu *benchmark* awal untuk gender di Twitter; menunjukkan *baseline* kuat dengan model klasik dan fitur sederhana.

Tabel 2. 2 Pengelompokan Penelitian Terdahulu Berdasarkan Metode

Sumber	Naïve Bayes	SVM	KNN / BM25	LSTM / BiLSTM	Transformer	Deskripsi Penelitian
Mahendra et al. (2022)	✓	✓				Menggunakan teks + metadata pengguna; ML klasik bekerja baik, namun kontribusi teks murni sulit diisolasi.
Katna et al. (2022)	✓	✓				TF–IDF pada teks berbahasa Inggris; SVM memberikan akurasi tertinggi pada prediksi usia dan gender.
Cheng et al. (2011)		✓				ML klasik digunakan untuk membedakan gender berdasarkan pola leksikal dan gaya penulisan.
Zakia et al. (2020)			✓			Klasifikasi gender Twitter Indonesia menggunakan BM25 + KNN; *baseline* tradisional namun efektif.
Humam & Laksito (2023)	✓	✓				TF–IDF + ML untuk sentimen Twitter Indonesia; relevan untuk *pipeline preprocessing* & pemodelan.
Helmayanti et al. (2023)	✓					Naïve Bayes bekerja sangat baik pada teks informal ulasan aplikasi (bahasa Indonesia).
Rosid et al. (2020)	✓	✓				Studi menekankan pentingnya *preprocessing* Sastrawi untuk meningkatkan akurasi pada model ML.
Burger et al. (2011)		✓				*Benchmark* awal *gender prediction* di Twitter; SVM memberikan *baseline* stabil.
Sa'diah et al. (2024)				✓		LSTM + fastText pada pesan Indonesia; performa tinggi jika dataset besar.
Masood et al. (2023)				✓		BiLSTM unggul dalam memahami konteks dua arah, tetapi sensitif terhadap ukuran data.
Nia et al. (2023)					✓	Transformer menghasilkan akurasi

						>90% pada *gender detection* lintas bahasa.
Fields et al. (2024)					✓	Studi survei: Transformer sangat akurat, namun mahal secara komputasi.

Penelitian yang menggunakan pendekatan *machine learning* klasik menunjukkan bahwa algoritma SVM merupakan model yang paling konsisten menghasilkan performa tinggi pada teks pendek. Hal ini terlihat pada studi Katna dkk. (2022), di mana akurasi SVM mencapai kisaran 80–90%, lebih tinggi dibandingkan *Naïve Bayes*. Meskipun lebih sederhana, *Naïve Bayes* tetap kompetitif terutama pada teks informal berbahasa Indonesia, seperti pada penelitian Helmayanti dkk. (2023) yang mencapai akurasi 84%**.** Namun, penelitian berbasis metadata seperti Mahendra dkk. (2022) menunjukkan bahwa akurasi dapat mencapai hingga di atas 80%, meskipun hal tersebut mengaburkan kontribusi teks murni. Secara umum, metode klasik stabil pada kondisi data terbatas dan tidak memerlukan komputasi besar.

Pada kelompok *deep learning*, model LSTM dan BiLSTM memberikan performa tinggi ketika jumlah data besar dan representatif. Penelitian Sa'diah dkk. (2024) mencatat akurasi 69-70%, sedangkan Masood dkk. (2023) melaporkan BiLSTM mampu melewati 90% pada data besar. Namun, model LSTM sangat bergantung pada ukuran *dataset*; ketika data sedikit atau *imbalanced*, performanya tidak stabil dan bahkan kalah dari model klasik. Kondisi ini menunjukkan bahwa *deep learning* tidak selalu unggul dalam konteks *low-resource* seperti Twitter berbahasa Indonesia.

Metode berbasis Transformer seperti BERT dan model sejenis menunjukkan performa paling tinggi secara umum. Nia dkk. (2023) melaporkan akurasi >88% pada *gender detection* lintas bahasa. Namun, penelitian Fields dkk. (2024) menunjukkan bahwa performa ini memerlukan *dataset* besar dan komputasi GPU yang kuat, sehingga pendekatan ini tidak selalu praktis untuk penelitian akademik dengan sumber daya terbatas. Tabel 2.3 menunjukkan perbandingan akurasi yang dipakai pada penelitian terdahulu.

Tabel 2. 3 Penelitian Terdahulu Berdasarkan Akurasi Model

Sumber	Metode	Akurasi / Performa	Catatan Singkat
Mahendra et al. (2022)	NB, SVM	84.20%	Metadata meningkatkan akurasi signifikan.

Katna et al. (2022)	NB, SVM	SVM ~88%	TF–IDF menghasilkan performa sangat baik.
Cheng et al. (2011)	ML klasik	70-80%	Bergantung fitur leksikal.
Zakia et al. (2020)	BM25 + KNN	68.6%	Metode tradisional namun kompetitif.
Humam & Laksito (2023)	LSTM	82%	Studi sentimen namun relevan pada *pipeline* pendek.
Helmayanti et al. (2023)	NB	84%	NB unggul pada teks informal berbahasa Indonesia.
Burger et al. (2011)	ML klasik	70-80%	*Benchmark* gender klasik.
Sa'diah et al. (2024)	LSTM	69-70%	*Deep learning* unggul dengan *dataset* besar.
Masood et al. (2023)	LSTM/BiLSTM	>90%	BiLSTM unggul jika data banyak dan stabil.
Nia et al. (2023)	Transformer	88-89%	Model terbaik secara global.

BAB III
METODOLOGI PENELITIAN

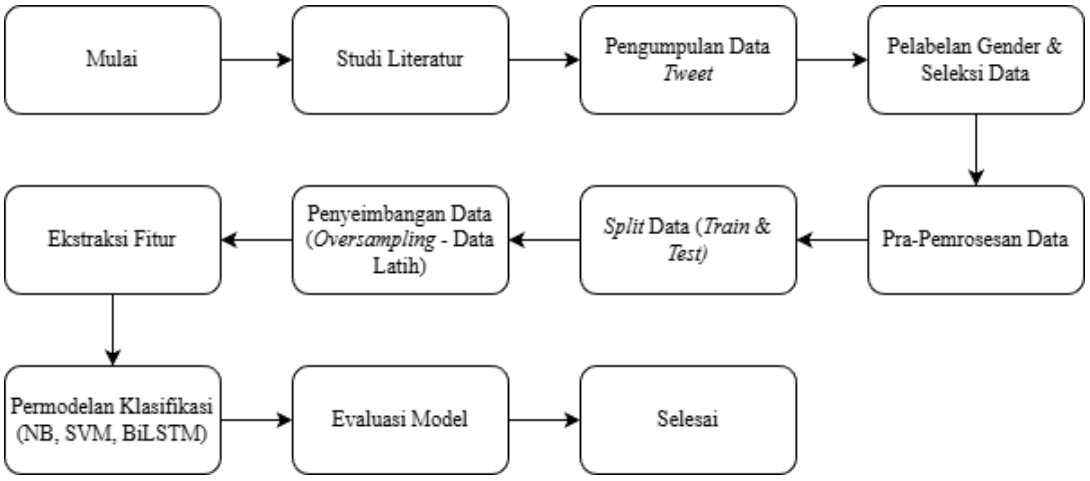
3.1 Tahapan Penelitian

Penelitian ini merupakan penelitian eksperimental dengan pendekatan kuantitatif yang bertujuan untuk membandingkan kinerja beberapa algoritma klasifikasi dalam memprediksi gender pengguna media sosial berdasarkan teks *tweet*. Metode yang digunakan adalah metode komparatif, di mana tiga algoritma klasifikasi, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* (BiLSTM), diuji menggunakan *dataset* dan alur pemrosesan yang sama.

Pemilihan algoritma *Multinomial Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* didasarkan pada pertimbangan teoretis dan empiris dari penelitian-penelitian terdahulu. *Multinomial Naïve Bayes* dan *Support Vector Machine* dipilih sebagai representasi algoritma pembelajaran mesin klasik yang telah terbukti efektif dan efisien dalam klasifikasi teks pendek dengan representasi fitur vektor jarang seperti TF-IDF. Kedua algoritma tersebut juga banyak digunakan sebagai *baseline* yang kuat pada berbagai studi klasifikasi teks dan *author profiling*, khususnya pada *dataset* berukuran kecil hingga menengah.

Sementara itu, *Bidirectional Long Short-Term Memory* dipilih sebagai representasi model pembelajaran mendalam berbasis sekuensial yang mampu menangkap konteks dan urutan kata dari dua arah, sehingga secara teoritis memiliki kemampuan lebih baik dalam memahami struktur linguistik teks. Pemilihan BiLSTM memungkinkan dilakukan evaluasi perbandingan antara pendekatan klasik berbasis fitur statistik dan pendekatan *deep learning* berbasis representasi sekuensial dalam satu kerangka metodologi yang sama. Algoritma lain seperti *Random Forest* atau model berbasis transformer tidak digunakan dalam penelitian ini karena pertimbangan keterbatasan ukuran *dataset*, kompleksitas komputasi, serta fokus penelitian yang diarahkan pada perbandingan metode yang relatif umum, terinterpretasi, dan sesuai untuk skenario data terbatas.

Alur penelitian dimulai dari pengumpulan data *tweet*, dilanjutkan dengan tahap *preprocessing* teks, penyeimbangan data, ekstraksi fitur, pemodelan klasifikasi, serta evaluasi performa model. Dengan menggunakan alur metodologi yang seragam, perbedaan hasil klasifikasi yang diperoleh dapat dianalisis secara objektif berdasarkan metrik evaluasi yang telah ditentukan.



Gambar 3. 1 Diagram Alur Penelitian

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa *tweet* berbahasa Indonesia yang dikumpulkan dari platform X (Twitter). Pengumpulan data dilakukan menggunakan layanan API pihak ketiga (*twitterapi.io*) dengan bantuan skrip Python. Proses pengambilan data dilakukan dengan menggunakan beberapa kata kunci yang relevan, yaitu *gibran*, *kerja*, *kuliah*, *banjir*, *overthinking*, *pinjol*, dan *parenting*, sehingga diperoleh *tweet* dengan variasi topik yang beragam. Penggunaan kata kunci yang beragam bertujuan untuk memperoleh variasi gaya bahasa pengguna agar model tidak bias terhadap topik tertentu.

Dataset awal terdiri dari beberapa atribut, namun dalam penelitian ini hanya atribut yang relevan yang digunakan, yaitu teks *tweet* (*text*), label gender (gender), dan waktu pembuatan *tweet* (*createdAt*) apabila tersedia. Label gender terdiri dari tiga kategori, yaitu *male*, *female*, dan *unknown*. Proses pelabelan gender pada penelitian ini dilakukan secara manual berdasarkan informasi publik yang tersedia pada profil pengguna Twitter. Penilaian gender dilakukan dengan mengamati nama lengkap, *username*, serta foto profil apabila tersedia. Tahap pelabelan diawali dengan inspeksi nama pengguna; apabila nama jelas menunjukkan identitas gender yang umum digunakan di Indonesia (misalnya “Budi”, “Rini”, “Siti”, “Agus”), maka label gender ditentukan berdasarkan nama tersebut. Pada kasus nama yang ambigu, peneliti menggunakan kombinasi pemeriksaan *username*, gaya penulisan, dan konteks umum pada akun untuk menetapkan label gender. Akun-akun yang tetap tidak dapat dipastikan gendernya diberi label *unknown* dan dikeluarkan dari dataset utama. Data dengan label *unknown* tidak digunakan dalam proses pelatihan model dan hanya dipisahkan untuk keperluan analisis lanjutan.

Tabel 3. 1 Contoh data *tweet*

tweet_id	created_at	lang	text	username	name	gender
1.99577E+18	Tue Dec 02 08:20:18 +0000 2025	in	@menakertrans kacrut. kalian bikin call centre bgini buat apaan klo ga bisa dihubungi. gak respon. Halah abis2in anggaran aja ada kalian di negara ini. Kerja yang bener lah. Pakai otak dan hati kalian. https://t.co/MsuZhuzwDI	handrifb	handri fajar budiman	male
1.99577E+18	Tue Dec 02 08:20:19 +0000 2025	in	@rizkaac @ARSIPAJA inget banget dulu pas masih awal kuliah pernah diomelin dosen karena jawab pakai kata "mungkin, saya kira" habis itu langsung trauma ga mau pakai kedua kata itu lagi.	heedeonlyone	araa	female
1.99577E+18	Tue Dec 02 08:20:20 +0000 2025	in	bro pikir rompinya bisa mengatasi banjir dan reboisasi hutan	katsuraxda	yeoj	unknown
1.99577E+18	Tue Dec 02 08:20:21 +0000 2025	in	@DedynurPalakka psi biasanya ngurusin banjir jakarta, nyalahin terus gubernurnya. ntar di sumatera semua pejabatnya disalahin. dan nol solusi. hhh ga bermoral dan kosong idenya	mr_hardi4	mr hardiansyah	male
1.99577E+18	Tue Dec 02 08:20:22 +0000 2025	in	lagi bocah ini udh jadi trouble maker, suka ngebully temen-temennya, di smster 3 udh jarang masuk di semua mata kuliah. Dia ga ada temen sama sekali.	iyamomonn	emjimeijia	unknown

3.3 Pra-pemrosesan Data

Tahap *preprocessing* bertujuan untuk membersihkan dan menormalkan teks *tweet* agar siap digunakan dalam proses ekstraksi fitur dan pemodelan. Mengingat karakteristik bahasa pada media sosial yang cenderung informal, tahap ini menjadi bagian penting dalam penelitian. Langkah-langkah *preprocessing* yang diterapkan adalah sebagai berikut:

1. ***Case Folding***

   Seluruh teks diubah menjadi huruf kecil untuk menghindari perbedaan representasi kata akibat perbedaan kapitalisasi.

2. **Penghapusan *Noise***

   Elemen-elemen yang tidak relevan dihapus, meliputi URL, *mention* pengguna, *hashtag*, token *retweet* (RT), angka, tanda baca, dan karakter non-alfabet.

3. **Normalisasi Kata Tidak Baku**

   Kata-kata tidak baku dan singkatan yang umum digunakan pada media sosial dinormalisasi ke bentuk baku, misalnya *"gak"* menjadi *"tidak"* dan *"udh"* menjadi *"sudah"*. Normalisasi ini bertujuan untuk mengurangi variasi kata yang memiliki makna sama. Daftar kata untuk normalisasi disajikan pada Lampiran 2.

4. ***Stopword Removal***

   Kata-kata umum yang tidak memiliki makna diskriminatif dihapus menggunakan daftar *stopword* bahasa Indonesia dari Sastrawi, serta *custom stopword* tambahan yang disesuaikan dengan karakteristik data Twitter. *Stopword* yang digunakan tertera pada halaman Lampiran A.3.

5. ***Stemming***

   Proses *stemming* dilakukan menggunakan algoritma Sastrawi untuk mengubah kata ke bentuk dasarnya, sehingga variasi imbuhan dapat diminimalkan.

6. **Penghapusan Token Pendek dan Data Kosong**

   Token dengan panjang kurang dari tiga karakter dihapus, dan *tweet* yang menghasilkan teks kosong setelah *preprocessing* tidak disertakan dalam *dataset* akhir.

Hasil dari tahap ini adalah kolom baru berupa *clean_tweet* yang berisi teks *tweet* yang telah dibersihkan dan dinormalisasi sebagaimana ditunjukkan pada Tabel 3.2.

Tabel 3. 2 Contoh data hasil pra-pemrosesan

text	**clean_tweet**

Apa lah MP Sepang & Majlis Perbandaran Sepang buat heh kalau setiap kali hujan lebat, area flyover tugu negara tu mesti banjir. Korang bodoh or memang tak tahu situ main road orang ulang alik pergi kerja / airport ???????	apa sepang amp majlis bandar sepang buat heh kalau kali hujan lebat area flyover tugu negara mesti banjir korang bodoh memang tak tahu situ main road ulang alik pergi kerja airpo
@tamarjaya0 Pajak otak dekat hadi iblis awang ke? Or pajak otak dekat dr.Sam-pah or pajak otak dekat sipendek.. kau ni mesti dari sperm lemah ni, i mean kuat onani Kau nak madani umum kelantan seleweng dana tebatan banjir ke? Biar satu negara tau ini la kerja lebai selama ni, kaki rolling	pajak otak dekat had iblis awang pajak otak dekat sam pah pajak otak dekat sipendek kau mesti sperm lemah mean kuat onani kau nak madani umum lantan seleweng dana tebat banjir biar satu negara tau kerja lebai lama kaki rolling
Dulu gue mikir gini pas temen2 mulai berkeluarga, mereka jadi ga sharing hobi & passion, post2 socmednya seputar capek kerja or parenting....ngeliatnya males. Tapi setelah makin dewasa gue jd ngerti kalo banyak orang yg passion nya menjalani hidup sesuai "blueprint" masyarakat.	dulu gue mikir pas temen mulai keluarga sharing hobi amp passion socmednya putar capek kerja parenting ngeliatnya males makin dewasa gue ngerti kalau banyak passion nya jalan hidup sesuai blueprint masyarakat
@yohanmyhero @Twt_dr_je @Mimbloon Misalnya saja ada lembaga pemerintahan yg harusnya tuh dalam bentuk sistem federal, dpd kalau federal itu biasanya senat, terus sekarang lembaga itu kerja nya apa? Kan ada dpr juga, ada lagi mpr, jadi banyak lembaga yg gak kerja karena sistem yg buruk nah itu serba nanggung	misal lembaga perintah harus bentuk sistem federal dpd kalau federal biasa senat terus sekarang lembaga kerja nya apa dpr mpr banyak lembaga kerja sistem buruk nah serba nanggung

3.4 Penyeimbangan Data

Setelah proses *preprocessing*, *dataset* yang digunakan hanya terdiri dari data berlabel *male* dan *female*. Distribusi kelas pada *dataset* menunjukkan adanya ketidakseimbangan jumlah data antar kelas (*class imbalance*), yang berpotensi menyebabkan model cenderung bias terhadap kelas mayoritas.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik *oversampling* menggunakan metode *RandomOverSampler*. Teknik ini bekerja dengan menyalin secara acak data dari kelas minoritas hingga jumlah data pada setiap kelas menjadi seimbang. Proses penyeimbangan hanya diterapkan pada data latih agar evaluasi model tetap mencerminkan kondisi data nyata. Penyeimbangan data hanya diterapkan pada data latih untuk mencegah terjadinya kebocoran data pada proses evaluasi.

3.5 Pembagian Data

Dataset yang telah diseimbangkan kemudian dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Pembagian data dilakukan dengan rasio 80:20 menggunakan metode

stratified sampling untuk menjaga proporsi kelas pada kedua *subset* data. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

3.6 Ekstraksi Fitur

Tahap ekstraksi fitur merupakan proses transformasi data teks hasil pra-pemrosesan menjadi representasi numerik yang dapat dipahami dan diolah oleh algoritma klasifikasi. Pada tahap ini, karakteristik linguistik dari teks *tweet* direpresentasikan dalam bentuk fitur yang mencerminkan informasi penting untuk membedakan kelas gender. Pemilihan metode ekstraksi fitur disesuaikan dengan karakteristik masing-masing model klasifikasi yang digunakan, sehingga representasi fitur yang dihasilkan optimal untuk setiap algoritma.

Dalam penelitian ini, digunakan dua pendekatan ekstraksi fitur yang berbeda. Untuk algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine*, representasi teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF–IDF), yang efektif untuk data teks pendek dan berdimensi tinggi. Sementara itu, untuk model *Bidirectional Long Short-Term Memory* (BiLSTM)**,** teks direpresentasikan dalam bentuk urutan token numerik yang dipertahankan urutannya, sehingga memungkinkan model sekuensial menangkap informasi konteks dan hubungan antar kata. Perbedaan pendekatan ini bertujuan memastikan bahwa setiap model menerima representasi fitur yang paling sesuai dengan mekanisme kerjanya.

3.6.1 TF-IDF

Untuk algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine*, representasi fitur teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Parameter TF-IDF yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Jumlah fitur maksimum: 15.000
b. Rentang *n-gram*: unigram dan bigram
c. *Minimum document frequency (min_df)*: 2
d. *Maximum document frequency (max_df)*: 0,8
e. *Sublinear term frequency*: aktif

Parameter tersebut dipilih untuk menyeimbangkan antara kompleksitas fitur dan kemampuan model dalam menangkap pola linguistik yang relevan. Penggunaan bigram bertujuan untuk menangkap frasa pendek yang berpotensi mengandung sinyal gender.

3.6.2 *Tokenizer* dan *Padding*

Untuk model BiLSTM, teks direpresentasikan dalam bentuk urutan token numerik menggunakan *Tokenizer*. Jumlah kata maksimum yang digunakan adalah 10.000 kata dengan token khusus untuk kata di luar kosakata (*out-of-vocabulary*). Setiap urutan teks kemudian diseragamkan panjangnya menggunakan *padding* dengan panjang maksimum 50 token.

3.7 Pemodelan Klasifikasi

Tahap permodelan klasifikasi merupakan inti dari penelitian ini, yaitu proses pembentukan model yang bertujuan mempelajari pola hubungan antara fitur teks *tweet* dan label gender pengguna. Pada tahap ini, data latih yang telah melalui proses ekstraksi fitur digunakan untuk melatih beberapa algoritma klasifikasi, sehingga model mampu melakukan prediksi gender pada data yang belum pernah dilihat sebelumnya.

Penelitian ini menerapkan tiga pendekatan klasifikasi yang berbeda, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory* (BiLSTM). Pemilihan ketiga algoritma tersebut didasarkan pada perbedaan karakteristik dan kompleksitas model, mulai dari metode pembelajaran mesin klasik hingga pembelajaran mendalam. Dengan menggunakan konfigurasi dan alur pelatihan yang terstandar, perbandingan kinerja antar model dapat dilakukan secara objektif dan adil.

3.7.1 *Multinominal Naïve Bayes*

Model *Naive Bayes* yang digunakan pada penelitian ini adalah *Multinomial Naive Bayes*, yang sesuai untuk data teks berbasis frekuensi seperti representasi TF-IDF. Model ini bekerja berdasarkan asumsi independensi antar fitur, di mana setiap kata dianggap berkontribusi secara independen terhadap penentuan kelas. Proses pelatihan dilakukan dengan menggunakan data latih (`X_train_tfidf, y_train`) melalui fungsi `fit()`, sehingga model dapat mempelajari pola distribusi probabilitas kata pada masing-masing kelas gender.

Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data uji (`X_test_tfidf`) menggunakan fungsi `predict()`. Hasil prediksi (`nb_pred`) kemudian digunakan untuk mengevaluasi kinerja model berdasarkan metrik evaluasi yang telah ditentukan, seperti akurasi, *F1-score*, dan *confusion matrix*.

3.7.2 *Support Vector Machine*

Model Support Vector Machine yang digunakan pada penelitian ini adalah *Linear Support Vector Classifier* (*LinearSVC*) dengan parameter `C = 0,8`, `class_weight = balanced`, `dual = False`, dan `max_iter = 5000`. Nilai `C = 0,8` berfungsi sebagai parameter regularisasi untuk mengontrol keseimbangan antara margin pemisah dan tingkat kesalahan klasifikasi, sedangkan pengaturan `class_weight = balanced` digunakan untuk menyesuaikan bobot kelas secara otomatis berdasarkan distribusi data sehingga pengaruh ketidakseimbangan kelas dapat diminimalkan. Penggunaan `dual = False` dipilih karena jumlah fitur TF-IDF lebih besar dibandingkan jumlah sampel, sehingga optimasi primal lebih efisien untuk data berdimensi tinggi.

Proses pelatihan model dilakukan menggunakan data latih berbasis TF-IDF (*`X_train_tfidf, y_train`*) melalui fungsi `fit()`. Parameter *`max_iter = 5000`* digunakan untuk menentukan batas maksimum iterasi optimasi agar model memiliki cukup iterasi untuk mencapai kondisi konvergen pada data teks yang kompleks. Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data uji (*`X_test_tfidf`*) menggunakan fungsi *`predict()`*, sehingga diperoleh hasil klasifikasi gender yang selanjutnya digunakan pada tahap evaluasi kinerja model.

3.7.3 *Bidirectional Long Short-Term Memory*

Model *Bidirectional Long Short-Term Memory* digunakan untuk menangkap konteks urutan kata secara dua arah, baik dari awal ke akhir maupun sebaliknya. Sebelum proses pelatihan, data teks terlebih dahulu ditokenisasi menggunakan *Tokenizer* dengan batas maksimum 10.000 kata dan penanganan kata di luar kosakata melalui *token* <OOV>. Hasil tokenisasi kemudian diubah menjadi deret numerik dan dipadatkan menggunakan *padding* dengan panjang maksimum 50 *token*, sehingga seluruh *input* memiliki panjang yang seragam. Label gender pada data latih dan data uji selanjutnya dikodekan ke dalam bentuk numerik menggunakan *LabelEncoder* agar dapat diproses oleh model *neural network*.

Arsitektur BiLSTM yang digunakan terdiri dari *Embedding layer* dengan dimensi vektor 200 untuk merepresentasikan kata ke dalam ruang vektor numerik, diikuti oleh dua lapisan *Bidirectional* LSTM. Lapisan pertama memiliki 128 unit dengan pengaturan *`return_sequences = True`* untuk meneruskan keluaran ke lapisan berikutnya, sedangkan lapisan kedua memiliki 64 *unit* untuk menangkap representasi urutan yang lebih ringkas. Untuk

mengurangi risiko *overfitting*, digunakan teknik *dropout* pada setiap lapisan LSTM serta satu lapisan *Dense* dengan aktivasi ReLU dan *Dropout* 0,4 sebelum lapisan keluaran.

Model dikompilasi menggunakan fungsi kerugian *binary cross-entropy* dan *optimizer Adam*, yang sesuai untuk permasalahan klasifikasi biner. Proses pelatihan dilakukan selama maksimum 10 *epoch* dengan *batch size* 32 dan validasi sebesar 20% dari data latih. Selain itu, diterapkan mekanisme *Early Stopping* dengan nilai *patience* = 2 untuk menghentikan pelatihan lebih awal apabila tidak terjadi peningkatan kinerja pada data validasi, serta mengembalikan bobot terbaik guna memperoleh performa model yang optimal.

3.8 Evaluasi Model

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *macro F1-score*, *confusion matrix*, dan *Receiver Operating Characteristic* (ROC) beserta nilai *Area Under the Curve* (AUC). Metrik *macro F1-score* digunakan untuk memberikan penilaian yang seimbang terhadap setiap kelas, sedangkan *confusion matrix* digunakan untuk menganalisis kesalahan klasifikasi secara lebih rinci. Perbandingan performa ketiga model dilakukan berdasarkan nilai metrik evaluasi tersebut untuk menentukan algoritma yang memberikan kinerja terbaik dalam klasifikasi gender berbasis teks *tweet* berbahasa Indonesia. Evaluasi dilakukan pada data uji yang tidak terlibat dalam proses pelatihan.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform X (Twitter) melalui proses pengumpulan data menggunakan layanan API pihak ketiga, yaitu *twitterapi.io*, dengan bantuak skrip Python. Pengambilan data dilakukan pada rentang waktu Oktober 2025 hingga November 2025 menggunakan beberapa kata kunci yang relevan dengan aktivitas sehari-hari dan isu sosial, yaitu *gibran*, *kerja*, *kuliah*, *banjir*, *overthinking*, *pinjol*, dan *parenting*. Penggunaan kata kunci yang beragam bertujuan untuk memperoleh variasi gaya bahasa pengguna agar data yang dikumpulkan tidak terfokus pada satu topik tertentu. Cuplikan kode program disajikan pada Gambar 4.1 dan kode program yang lengkap disajikan pada Lampiran A.

```
    API_KEY = "new1_1bab7b94912047d1b32083e51c34f7bf"

    QUERY = """
    ("kuliah" OR "kerja" OR "parenting" OR "overthinking" OR "pinjol" OR
"banjir") lang:id
    """.strip()

    MAX_TWEETS = 200

    URL = "https://api.twitterapi.io/twitter/tweet/advanced_search"
    HEADERS = {"X-API-Key": API_KEY}
```

Gambar 4. 1 Kode *scrapping* data

Dari proses pengumpulan data tersebut, diperoleh sebanyak 1136 *tweet* mentah yang terdiri dari beberapa atribut, antara lain *tweet_id*, *created_at*, *text*, *username*, *name*, dan *gender*. *Dataset* ini masih mengandung *tweet* tanpa label gender yang jelas, sehingga belum seluruhnya dapat digunakan dalam proses pemodelan klasifikasi. Oleh karena itu, dilakukan tahap seleksi lanjutan terhadap *tweet* yang memiliki label gender *male* dan *female*. Hasil seleksi menunjukkan bahwa dari total data mentah, diperoleh sebanyak 655 *tweet* yang memiliki label gender *male* atau *female*. *Dataset* inilah yang digunakan pada tahap-tahap selanjutnya dalam penelitian ini. Contoh *raw data* hasil pengumpulan data disajikan pada Tabel 4.1.

Tabel 4. 1 Contoh *raw* data hasil *scrapping*

tweet_id	created_at	lang	text	username	name	gender
1.97951E+18	Sat Oct 18 11:42:19 +0000 2025	in	@melekwarga Nggak ragu support Prabowo Gibran karena udah sering liat sendiri hasil kerjanya	aditya_prnmaa3	Aditya Purnama	male
1.97951E+18	Sat Oct 18 11:42:33 +0000 2025	in	1 Tahun Prabowo-Gibran Jalan terbuka, harapan tumbuh! Infrastruktur di Papua memudahkan akses, mempercepat ekonomi, dan menghadirkan kesejahteraan untuk semua. #DariTimurIndonesiaBersinar #AyoDukungFoodEstate https://t.co/pMqHIEjITt	Putri_Aura_	Putri A.	female
1.97951E+18	Sat Oct 18 11:43:08 +0000 2025	in	@awikuchuz @alfredokusu @AswinbaharD @BaksoPentol678 @BANGSAygSUJUD @BBerizik @DedynurPalakka @Epricila10 @KiEdySularno @Yattie2023 Sehat selalu mas Gibran.. Biarkan para pembenci stroke sendiri dengan pikiran2 jahat mereka ??	zy_zy_lestary	ziii ??	female
1.98419E+18	Fri Oct 31 09:23:04 +0000 2025	in	yaAllah pls berkahi aku dengan banjir orderan itu kaya dulu plis banget ini mah tangan aku gatel baking?? tp gamau kerja di orang yg masuk subuh or pulang tengah malem itu	purplishskye	??	unknown
1.98577E+18	Tue Nov 04 17:52:33 +0000 2025	in	baba gw selalu bilang "se butuh butuh nya kamu sama uang jngan pernah yg namanya pinjol. mau 10ribu pun itu, appa kerja itu buat kamu jadi klo ada apa apa bilang" and yeah gw gk berani yg namanya pinjol or paylatter se urgent apapun itu.	midnightminwon	ndaa - ????•?•??	unknown

4.2 Hasil Pelabelan Gender

Pelabelan gender pada *dataset* dilakukan secara manual dengan mengacu pada informasi publik yang tersedia pada akun pengguna, khususnya *username* dan *name* yang ditampilkan pada profil Twitter. Penentuan label *male* atau *female* dilakukan berdasarkan indikasi nama diri, penggunaan kata sapaan, atau penanda linguistik lain yang secara umum diasosiasikan

dengan jenis kelamin tertentu. Apabila informasi tersebut tidak cukup jelas atau ambigu, data diberi label *unknown* dan tidak disertakan dalam proses pelatihan dan evaluasi model. Pendekatan pelabelan manual ini dipilih untuk menjaga kualitas anotasi dan meminimalkan kesalahan klasifikasi label pada *dataset* penelitian.

Dari 655 *tweet* yang digunakan, diperoleh distribusi kelas gender yang tidak seimbang, di mana jumlah *tweet* berlabel *male* lebih banyak dibandingkan *female* dengan 357 *tweet male* dan 298 *tweet female*. Ketidakseimbangan distribusi kelas ini berpotensi menimbulkan bias pada model klasifikasi apabila tidak ditangani secara khusus. Tahap pelabelan ini menjadi krusial karena kualitas label akan sangat mempengaruhi performa model klasifikasi yang dibangun pada tahap selanjutnya.

4.3 Hasil Pra-Pemrosesan Data

Dataset hasil penyaringan label selanjutnya melalui proses pra-pemrosesan teks yang meliputi *case folding*, penghapusan URL dan simbol non-alfabet, normalisasi kata tidak baku, penghapusan *stopword* bahasa Indonesia yang menggunakan *stopword* pada Lampiran A.3, serta *stemming* menggunakan algoritma Sastrawi. Proses pra-pemrosesan ini bertujuan untuk mengurangi *noise* linguistik dan meningkatkan konsistensi representasi teks. Cuplikan kode untuk pra-pemrosesan disajikan pada Gambar 4.2 dan kode lengkap pra-pemrosesan disajikan pada Lampiran B.

```
data["clean_tweet"] = data["text"].apply(clean_text)
```

Gambar 4. 2 Kode program pra-pemrosesan data

Hasil dari tahap pra-pemrosesan adalah terbentuknya kolom baru berupa *clean_tweet* yang berisi teks hasil pembersihan dan normalisasi. Teks yang semula mengandung *noise* dan variasi penulisan menjadi lebih konsisten secara linguistik. Proses ini berhasil mengurangi kompleksitas teks dan meningkatkan kualitas representasi fitur yang akan digunakan pada tahap pemodelan. Tabel 4.2 menampilkan contoh hasil dari pra-pemrosesan teks.

Tabel 4. 2 Contoh hasil dari pra-pemrosesan teks

text	gender	clean_tweet
@sutanmangara Auranya bagai hutan yg ditebangi kayunya dan menjadi buram lalu banjir	male	aura bagai hutan tebang kayu jadi buram lalu banjir

Angin Kencang, Longsor, dan Banjir di Jonggol Bogor Rusak 21 Rumah Klik untuk baca: https://t.co/ZAMyGJDWn3	male	angin kencang longsor banjir jonggol bogor rusak rumah klik baca
ini konsepnya emng tk-sma dipilihin trs kuliah bebas ya :))	female	konsep emng sma dipilihin trs kuliah bebas
day one magang, capekk nya kerasa banget karna harus terbiasa lagi dengan jam kerja dan berkegiatan dari pagi-sore hari. sesekali mau share disini ah biar jadi arsip(+gaada tempat cerita jg si awkawkawk) https://t.co/cU9DXozY6s	female	day one magang capekk nya rasa banget karna biasa jam kerja giat pagi sore hari sekal mau share sini biar arsip gaada tempat cerita awkawkawk
Daripada si gibran opsional orang ini lebih okey.	male	gibran opsional lebih okey

4.4 Hasil Pembagian dan Penyeimbangan Data

Dataset yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20 menggunakan metode *stratified sampling*. Teknik ini digunakan untuk menjaga proporsi kelas gender pada kedua *subset* data. Kode program untuk pembagian data disajikan pada Gambar 4.3.

```
    train_test_split(test_size=0.2, stratify=y)
```

Gambar 4. 3 Kode program tahap pembagian data

Analisis distribusi kelas pada dataset hasil pra-pemrosesan menunjukkan adanya ketidakseimbangan jumlah data antar kelas gender. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik *oversampling* menggunakan metode *RandomOverSampler* pada data latih. Kode program penyeimbangan data terdapat pada Gambar 4.4

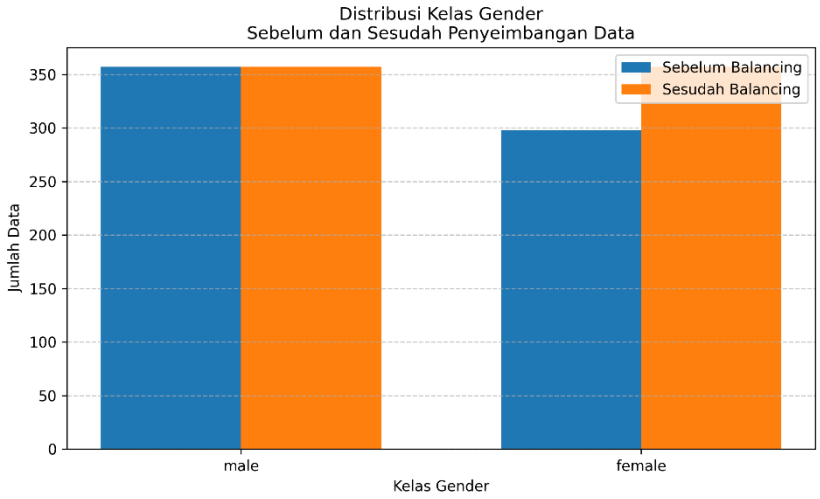
```
    ros = RandomOverSampler(random_state=SEED)
    X_res,      y_res      =      ros.fit_resample(data[["clean_tweet"]],
data["gender"])
```

Gambar 4. 4 Kode program tahap penyeimbangan data

Setelah proses penyeimbangan, jumlah data pada masing-masing kelas pada data latih menjadi seimbang, yaitu 357 data *male* dan 357 data *female*, sehingga total data latih setelah penyeimbangan berjumlah 714 data. Proses penyeimbangan ini dilakukan hanya pada data

latih, sedangkan data uji tetap mempertahankan distribusi kelas aslinya guna menghindari terjadinya *data leakage* pada tahap evaluasi model.

Penyeimbangan data ini memberikan kondisi pelatihan yang lebih adil bagi setiap algoritma klasifikasi dalam mempelajari pola linguistik dari kedua kelas gender. Dengan demikian, evaluasi performa model, khususnya menggunakan metrik *macro F1-score*, dapat merepresentasikan kinerja klasifikasi secara lebih objektif dan tidak didominasi oleh kelas mayoritas. Gambar 4.5 menampilkan grafik perbandingan sebelum dan sesudah penyeimbangan data.



Gambar 4. 5 Grafik perbandingan sebelum dan sesudah penyeimbangan data

4.5 Hasil Ekstraksi Fitur

Pada tahap ekstraksi fitur, digunakan dua pendekatan yang berbeda sesuai dengan karakteristik algoritma klasifikasi yang digunakan. Untuk algoritma *Multinomial Naïve Bayes* dan *Support Vector Machine*, teks direpresentasikan menggunakan metode TF–IDF dengan jumlah fitur maksimum sebanyak 15.000 dan rentang *n-gram* unigram dan bigram. Representasi ini menghasilkan matriks fitur berdimensi tinggi yang mampu menangkap kata-kata diskriminatif dalam teks *tweet*. Gambar 4.6 menampilkan kode program ekstraksi fitur.

```
TfidfVectorizer(max_features=15000, ngram_range=(1,2))
Tokenizer(num_words=10000)
pad_sequences(maxlen=50)
```

Gambar 4. 6 Kode program tahap ekstraksi fitur

Sementara itu, untuk model *Bidirectional Long Short-Term Memory*, teks dikonversi menjadi urutan token numerik menggunakan *Tokenizer* dengan ukuran kosakata 10.000 kata. Setiap urutan teks kemudian diseragamkan panjangnya menggunakan *padding* hingga panjang maksimum 50 token. Hasil ekstraksi fitur ini menjadi input utama bagi masing-masing model klasifikasi pada tahap pemodelan.

4.6 Hasil Pengujian Model Klasifikasi

Tahap pengujian model klasifikasi dilakukan untuk mengevaluasi kemampuan tiga algoritma yang digunakan dalam penelitian ini, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine* (SVM), dan *Bidirectional Long Short-Term Memory* (BiLSTM), dalam mengklasifikasikan gender pengguna Twitter berdasarkan teks *tweet* berbahasa Indonesia.

Seluruh model dilatih menggunakan data latih yang sama, yaitu data hasil pra-pemrosesan dan penyeimbangan kelas menggunakan *RandomOverSampler*, serta diuji menggunakan data uji yang identik dan tidak melalui proses penyeimbangan. *Model Naïve Baye*s dan SVM dilatih menggunakan representasi fitur TF–IDF, sedangkan BiLSTM dilatih menggunakan urutan token numerik. Proses pelatihan *Naïve Bayes* dan SVM berlangsung relatif cepat, sedangkan BiLSTM memerlukan beberapa *epoch* pelatihan hingga mencapai kondisi konvergensi. Setiap model yang telah dilatih kemudian digunakan untuk melakukan prediksi pada data uji guna mengevaluasi performa klasifikasi. Gambar 4.7 menampilkan bagian kode program tahap permodelan dari ketiga model.

```
# NAIVE BAYES
     nb = MultinomialNB()
     nb.fit(X_train_tfidf, y_train)
     nb_pred = nb.predict(X_test_tfidf)
#SVM
     svm = LinearSVC(C=0.8, class_weight="balanced", dual=False,
max_iter=5000)
     svm.fit(X_train_tfidf, y_train)
     svm_pred = svm.predict(X_test_tfidf)
#BiLSTM
     model = Sequential(
         [
             Embedding(len(tokenizer.word_index) + 1, 200,
input_length=50),
```

```
            Bidirectional(
                LSTM(128, dropout=0.3, recurrent_dropout=0.2,
return_sequences=True)
            ),
            Bidirectional(LSTM(64, dropout=0.3, recurrent_dropout=0.2)),
            Dense(64, activation="relu"),
            Dropout(0.4),
            Dense(1, activation="sigmoid"),
        ]
    )

    model.compile(loss="binary_crossentropy", optimizer="adam",
metrics=["accuracy"])

    model.fit(
        X_train_seq,
        y_train_enc,
        epochs=10,
        batch_size=32000,
        validation_split=0.2,
        callbacks=[EarlyStopping(patience=2,
restore_best_weights=True)],
        verbose=1,
    )
```

Gambar 4. 7 Kode program tahap permodelan

4.6.1 Metrik Evaluasi

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *macro F1-score*. Penggunaan *macro F1-score* menjadi perhatian utama dalam penelitian ini karena metrik tersebut memberikan bobot yang sama pada setiap kelas, sehingga lebih representatif untuk mengevaluasi performa model pada kondisi data yang sebelumnya tidak seimbang. Selain itu, *confusion matrix* dan kurva ROC–AUC digunakan untuk analisis kesalahan klasifikasi secara lebih mendalam. Gambar 4.8 menampilkan bagian kode program tahap evaluasi model klasifikasi.

```
#NAIVE BAYES
    nb_acc = accuracy_score(y_test, nb_pred)
```

```
    nb_f1 = f1_score(y_test, nb_pred, average="macro")
    cm_nb = confusion_matrix(y_test_reset, nb_pred_reset,
labels=["male", "female"])
    fpr_nb, tpr_nb, _ = roc_curve((y_test == "female").astype(int),
nb_prob)
    auc_nb = auc(fpr_nb, tpr_nb)

#SVM
    svm_acc = accuracy_score(y_test, svm_pred)
    svm_f1 = f1_score(y_test, svm_pred, average="macro")
    cm_svm = confusion_matrix(y_test_reset, svm_pred_reset,
labels=["male", "female"])
    fpr_svm, tpr_svm, _ = roc_curve((y_test == "female").astype(int),
svm_scores)
    auc_svm = auc(fpr_svm, tpr_svm)

#BiLSTM
    lstm_acc = accuracy_score(y_test_enc, lstm_pred)
    lstm_f1 = f1_score(y_test_enc, lstm_pred, average="macro")
    cm_lstm = confusion_matrix(y_test_enc_reset, lstm_pred_reset,
labels=[0, 1])
    fpr_lstm, tpr_lstm, _ = roc_curve(y_test_enc, lstm_prob)
    auc_lstm = auc(fpr_lstm, tpr_lstm)
```

Gambar 4. 8 Kode program tahap evaluasi model

4.6.2 Hasil Evaluasi Kuantitatif

Berdasarkan hasil evaluasi pada Tabel 4.3, SVM memperoleh nilai akurasi tertinggi sebesar 0,67, diikuti oleh *Multinominal Naïve Bayes* dengan nilai akurasi 0,66. Sementara itu, model BiLSTM menunjukkan performa yang lebih rendah dengan nilai akurasi 0,57 dan *macro F1-score* 0,52.

Tabel 4. 3 Hasil Evaluasi Model Klasifikasi

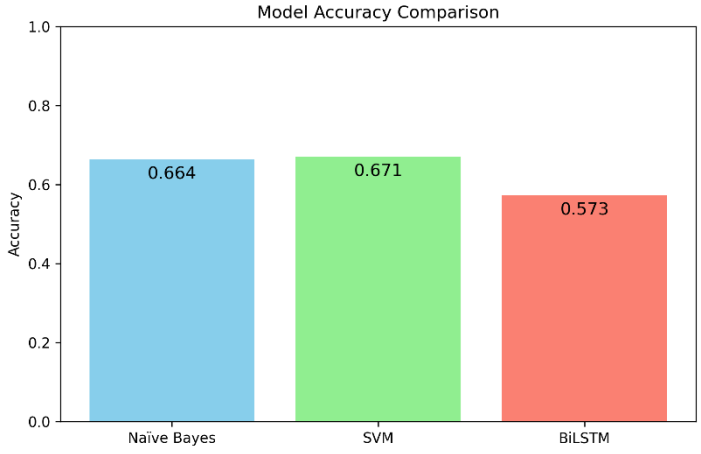
Model Klasifikasi	**Accuracy**	**Precision**	**Recall**	**Macro F1-Score**
Naïve Bayes	0.66	0.67	0.66	0.66
SVM	0.67	0.68	0.67	0.66
BiLSTM	0.57	0.61	0.57	0.52

4.6.3 Analisis Perbandingan Antar Model

Perbandingan hasil pengujian menunjukkan bahwa model pembelajaran mesin klasik (*Naïve Bayes* dan SVM) mampu memberikan performa yang lebih stabil dibandingkan model pembelajaran mendalam BiLSTM pada *dataset* yang digunakan. SVM unggul tipis dibandingkan *Naïve Bayes*, baik dari sisi akurasi maupun *precision*, yang mengindikasikan bahwa model ini lebih efektif dalam memisahkan representasi fitur teks berdimensi tinggi yang dihasilkan oleh TF-IDF.

Multinomial Naïve Bayes menunjukkan performa yang cukup kompetitif meskipun menggunakan asumsi independensi antar fitur. Hal ini mengindikasikan bahwa pada teks pendek seperti *tweet*, distribusi kata yang diwakili oleh TF-IDF masih cukup informatif untuk membedakan kelas gender.

Sebaliknya, model BiLSTM belum mampu menunjukkan performa yang belum optimal. Nilai *macro F1-score* yang lebih rendah mengindikasikan bahwa model ini mengalami kesulitan dalam mempelajari pola linguistik secara konsisten pada *dataset* berukuran terbatas. Hal ini sejalan dengan karakteristik model pembelajaran mendalam yang umumnya memerlukan data dalam jumlah besar dan variasi konteks yang lebih kaya agar dapat bekerja secara optimal. Gambar 4.2 menunjukkan perbandingan akurasi dari ketiga model.

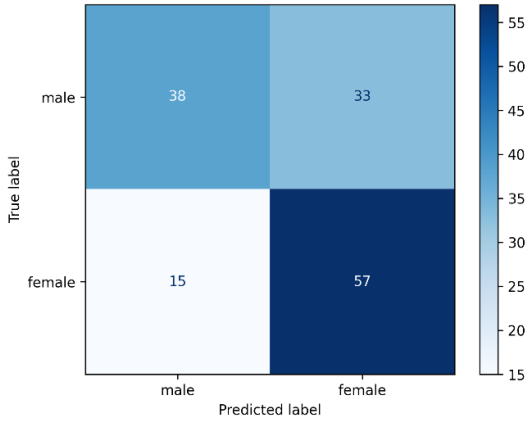


Gambar 4. 9 Perbandingan akurasi model (NB, SVM, BiLSTM)

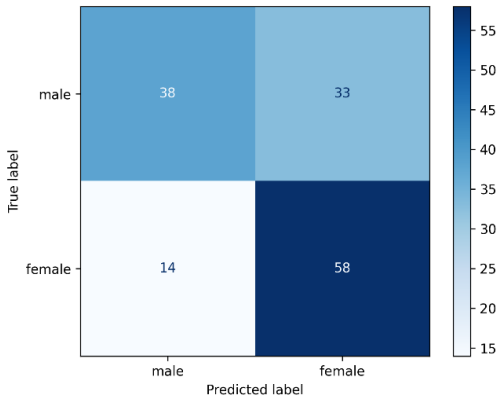
4.7 Analisis *Confuison Matrix*

Confusion matrix digunakan untuk menganalisis secara lebih rinci kesalahan klasifikasi yang dilakukan oleh masing-masing model terhadap setiap kelas gender. Melalui *confusion matrix*, dapat diketahui kemampuan model dalam memprediksi kelas *male* dan *female*, serta

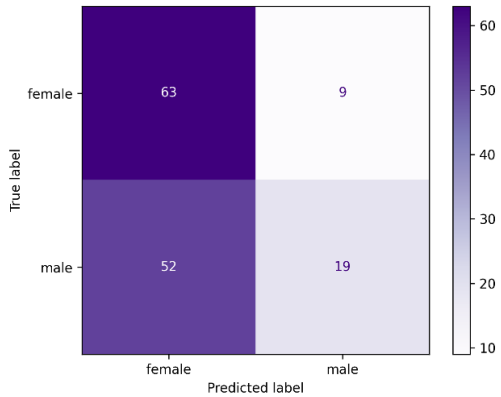
jenis kesalahan yang paling sering terjadi. Gambar 4.10 sampai 4.12 menampilkan grafik *confusion matrix* dari ketiga model.



Gambar 4. 11 *Confusion Matrix Multinominal Naive Bayes*



Gambar 4. 10 *Confusion Matrix Support Vector Machine*



Gambar 4. 12 *Confusion Matrix Bidirectional* LSTM

Berdasarkan *confusion matrix* untuk *Multinomial Naïve Bayes*, terlihat bahwa model mampu mengklasifikasikan sebagian besar data *female* dengan cukup baik, namun masih terjadi kesalahan prediksi pada kelas *male* yang cukup signifikan. Hal ini menunjukkan bahwa distribusi kata yang dipelajari oleh *Naïve Bayes* cenderung lebih representatif untuk salah satu kelas, sehingga asumsi independensi fitur belum sepenuhnya mampu menangkap variasi linguistik pada kedua kelas secara seimbang.

Pada *Support Vector Machine*, confusion matrix menunjukkan jumlah prediksi benar yang lebih tinggi dan relatif seimbang pada kedua kelas. Kesalahan klasifikasi pada kelas *male* dan *female* lebih sedikit dibandingkan *Naïve Bayes*. Hal ini mengindikasikan bahwa SVM mampu membentuk *hyperplane* pemisah yang lebih optimal di ruang fitur TF-IDF berdimensi tinggi, sehingga memiliki kemampuan generalisasi yang lebih baik.

Sementara itu, *confusion matrix* untuk BiLSTM menunjukkan pola kesalahan yang lebih tidak seimbang. Model cenderung lebih sering salah dalam mengklasifikasikan salah satu kelas, yang mengindikasikan bahwa representasi sekuensial yang dipelajari BiLSTM belum stabil pada *dataset* yang digunakan. Kondisi ini memperkuat indikasi bahwa BiLSTM memerlukan jumlah data yang lebih besar agar dapat mempelajari pola konteks linguistik secara lebih konsisten. Tabel 4.4 menunjukkan komparasi *confusion matrix* antar model.

Tabel 4. 4 Komparasi hasil *confusion matrix* model

Model	True Positive	False Negative	True Negative	False Positive
Naive Bayes	38	33	57	15
SVM	38	33	58	14
BiLSTM	19	52	63	9

4.8 Analisis Kurva ROC dan Nilai AUC

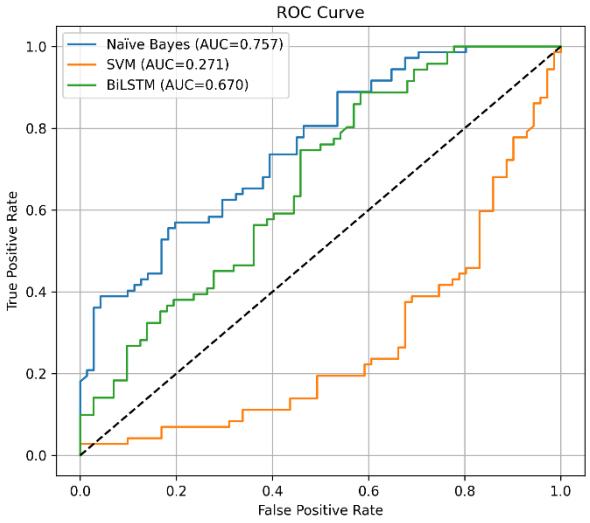
Kurva *Receiver Operating Characteristic* (ROC) digunakan untuk mengevaluasi kemampuan model dalam membedakan dua kelas pada berbagai nilai ambang keputusan (*threshold*). Nilai *Area Under the Curve* (AUC) memberikan ukuran kuantitatif terhadap kemampuan diskriminatif model, di mana nilai yang lebih mendekati 1 menunjukkan performa klasifikasi yang lebih baik.

Berdasarkan kurva ROC yang dihasilkan, *Multinomial Naïve Bayes* menunjukkan nilai AUC tertinggi dibandingkan dua model lainnya. Hal ini mengindikasikan bahwa secara keseluruhan Naïve Bayes memiliki kemampuan yang cukup baik dalam membedakan kelas

male dan *female* pada berbagai threshold, meskipun akurasi dan macro F1-score-nya tidak menjadi yang tertinggi.

Model SVM menunjukkan kurva ROC yang relatif lebih dekat dengan garis diagonal, yang menandakan kemampuan diskriminatif yang lebih terbatas pada konfigurasi parameter yang digunakan. Fenomena ini dapat disebabkan oleh penggunaan skor keputusan (*decision function*) yang kurang terkalibrasi secara probabilistik, sehingga memengaruhi perhitungan ROC–AUC meskipun akurasi klasifikasinya relatif tinggi.

Adapun BiLSTM menunjukkan nilai AUC yang berada di antara Naïve Bayes dan SVM. Hal ini mengindikasikan bahwa model mampu menangkap sebagian informasi diskriminatif dari data, namun belum cukup kuat untuk menghasilkan pemisahan kelas yang konsisten. Hasil ini sejalan dengan performa BiLSTM pada metrik akurasi dan macro F1-score yang juga lebih rendah dibandingkan model klasik.



Gambar 4. 13 Kurva ROC

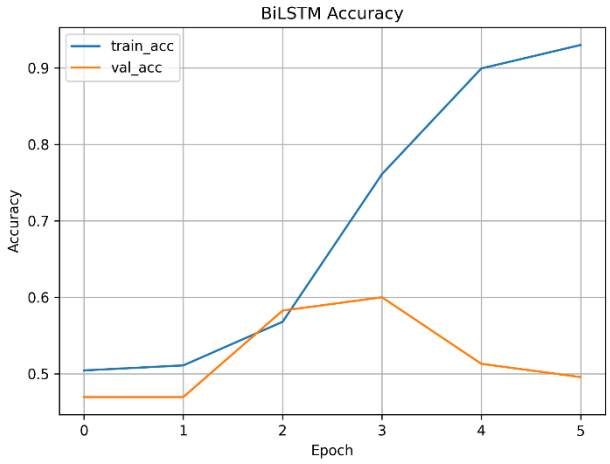
4.9 Analisis Proses Pelatihan BiLSTM

Analisis proses pelatihan BiLSTM dilakukan dengan mengamati kurva akurasi dan *loss* pada data latih (*training*) dan data validasi (*validation*) selama beberapa *epoch* pelatihan. Kurva akurasi menunjukkan bahwa akurasi data latih meningkat secara signifikan seiring bertambahnya *epoch*, sedangkan akurasi data validasi hanya meningkat pada beberapa *epoch* awal dan kemudian cenderung stagnan bahkan menurun.

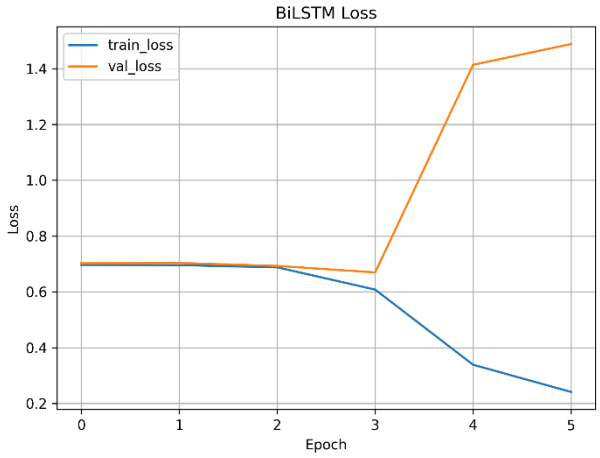
Fenomena ini merujuk pada pola kurva yang tampak pada Gambar 4.14 dan Gambar 4.15, di mana model terlalu menyesuaikan diri dengan data latih namun gagal melakukan

generalisasi dengan baik pada data yang tidak pernah dilihat sebelumnya. Hal ini diperkuat oleh kurva *loss*, yang menunjukkan penurunan *loss* pada data latih, sementara *loss* pada data validasi justru meningkat secara tajam setelah *epoch* tertentu.

Penggunaan mekanisme *early stopping* dalam penelitian ini bertujuan untuk membatasi dampak overfitting dengan menghentikan proses pelatihan ketika performa validasi tidak lagi meningkat. Meskipun demikian, hasil akhir BiLSTM tetap menunjukkan bahwa model ini belum mampu melampaui performa metode pembelajaran mesin klasik. Hal ini menunjukkan bahwa, pada dataset tweet berbahasa Indonesia yang berukuran terbatas dan bersifat sangat informal, kompleksitas model BiLSTM belum dapat dimanfaatkan secara optimal.



Gambar 4. 15 Kurva akurasi pelatihan BiLSTM



Gambar 4. 14 Kurva *loss* pelatihan BiLSTM

4.10 Contoh *Tweet* dan Hasil Prediksi Model

Untuk memberikan gambaran kualitatif terhadap kinerja model klasifikasi, penelitian ini menyajikan beberapa contoh *tweet* dari data uji beserta label gender sebenarnya dan hasil prediksi dari masing-masing model, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine*, dan *Bidirectional Long Short-Term Memory*. Contoh *tweet* dipilih secara acak dari data uji dan tidak digunakan pada proses pelatihan model. Tabel 4.5 menyajikan contoh *tweet* yang telah melalui tahap pra-pemrosesan, label gender asli berdasarkan pelabelan manual, serta hasil prediksi dari ketiga model klasifikasi.

Tabel 4. 5 Contoh *tweet* dan hasil prediksi model

clean_tweet	**Gender Asli**	**Prediksi NB**	**Prediksi SVM**	**Prediksi BiLSTM**
nangisin gajah kena dampak banjir	female	female	female	female
dukung perintah prabowo gibran berantas judi daring masa depan bangsa	male	female	female	female
ransel cakep banget buat kerja	female	female	female	female
situasi baik apa hilang rumah ancur beranta seret banjir mayat mana muak banget punya pimpin kolot presiden cacat pola pikir	female	female	female	female
jangan bicara jauh sekarang banyak illfeel gibran	female	female	female	female
eksploitasi kerja dagang manusia kerja paksa keras bas gender kontrol rampas bebas semua praktik budak modern sayang terus jadi banyak luput lihat langsung ruang hari hari	female	female	female	female
sudah dik gibran jangan terus tambah dosa mundur	male	male	male	female
terus nyala mas wapres gibran biar benci terus gonggong memang sedang emban tugas	male	female	female	male
papua grows indonesia prospers bangun adil jadi papua bagi penting maju bangsa sama prabowo gibran timur indonesia sinar	male	female	female	female

4.10.1 Contoh *Tweet* Gender Perempuan

Untuk memahami karakteristik linguistik pada kelas perempuan, ditampilkan beberapa contoh *tweet* yang berlabel *female*. Berdasarkan hasil pengamatan terhadap contoh *tweet*

tersebut, teks yang dikategorikan sebagai gender perempuan cenderung memiliki ciri-ciri sebagai berikut:

a. Penggunaan kata ganti orang pertama yang bersifat personal dan ekspresif, seperti *aku* dan *saya*.
b. Adanya kata atau frasa yang mengekspresikan emosi, perasaan, atau kondisi psikologis, seperti *overthinking*, *capek*, *bingung*, dan *takut*.
c. Struktur kalimat yang relatif deskriptif dan naratif, dengan penekanan pada pengalaman pribadi.

Karakteristik ini konsisten dengan temuan penelitian terdahulu yang menyatakan bahwa perbedaan gaya bahasa dan pilihan kata dapat menjadi indikator gender dalam teks media sosial. Model klasifikasi, khususnya SVM, mampu menangkap pola tersebut melalui bobot fitur pada representasi TF–IDF. Tabel 4.6 menyajikan contoh *tweet* dari gender perempuan.

Tabel 4. 6 Contoh *tweet* gender perempuan

mark juga harus dapet istirahat yang proper gak sih? mahae tuh bertahun-tahun kerja keras terus. please SM kasih mark dan haechan istirahat yang proper ??
@HendraSog @Roebanie @gibran_tweet Program seperti MBG justru jadi langkah nyata untuk meningkatkan kualitas sumber daya manusia Indonesia. #LanjutkanMBG
@melekwarga Prabowo Gibran bawa energi positif untuk perubahan Indonesia ke arah yang lebih baik
gw ga akan ngerti yg belain ortu yg ninggalin tanggung jawab ke anaknya di umur produktif as kek belain nabi. jaman probono bantuin org2 keluar pinjol or perbaikin bicheking. ortu yg lahir di 55-70, jaman muda sibuk narkoba sama miras oplosan, pas tua ga bisa perpikir dan kerja
Ya betul, kerja bersama mengubah evakuasi jadi keberanian. Mengubah rekonstruksi jadi komitmen untuk bangkit kembali. Ketika tangan-tangan berbeda bersatu, masalah besar mulai terselesaikan & diatasi. Bergerak Cepat Bersama
@worksfess Kaka " minta saran dong aku ngasal ngelamar kerja dari jobstreet kerjanya kaya nagih hutang gitu kaya kirim pesan telfon kaya di pinjol yay or nayy minta sarannya yahh ???????? https://t.co/TYOgGr7xzm

4.10.2 Contoh *Tweet* Gender Laki-laki

Selain *tweet* berlabel perempuan, penelitian ini juga menampilkan contoh *tweet* berlabel *male* untuk mengidentifikasi pola linguistik yang dominan pada kelas laki-laki. Berdasarkan analisis terhadap contoh *tweet* tersebut, dapat diidentifikasi beberapa karakteristik umum, antara lain:

a. Penggunaan bahasa yang lebih ringkas dan langsung, dengan kalimat yang relatif pendek.

b. Munculnya kata-kata yang berkaitan dengan aktivitas, pekerjaan, atau opini, seperti *kerja*, *ngurus*, *masalah*, dan *program*.
c. Ekspresi emosi yang lebih implisit dibandingkan *tweet* berlabel perempuan.

Tweet berlabel laki-laki cenderung menampilkan gaya bahasa yang informatif dan faktual. Pola ini lebih mudah ditangkap oleh model pembelajaran mesin klasik yang mengandalkan frekuensi dan bobot kata, seperti *Naïve Bayes* dan SVM.

Tabel 4. 7 Contoh *tweet* gender laki-laki

Untuk si fufufafa @gibran_tweet @jokowi dengar dan renungkan harusnya malu, kok para termul @Jokowimania @BaraJP @DPP_BaraJP @ProJokowi karena cara berpakaian konon mirip Bung Hatta!?? padahal Otaknya NOL... @prabowo @puanmaharani_ri @DPR_RI https://t.co/AaXPDmPRTP
kalo nggak ada penggundulan hutan nggak akan banjir lah blog goblog??
@zoelfick Ditempat kerja gua kemarin tiap abis gajian malah pada gapunya duit gara gara abis buat bayar pinjol or utang ke rentenir, Alhamdulillah udah keluar dari sana, walaupun kerjaan enak tapi ya gitu lingkungan kerjanya ngga bagus
Njir abis IGs sama Tweetnya si PhiGo semua orang jadi overthinking ?????? GUE HARUS MENATA HATI GUA OR GUE BISA BISA BOLOS KERJA
gw pndah krna dpt tawaran salery lebih , plus waktu kerja yg sedikit (menurut bini gw ??) biar bisa bnyak waktu di rmh gw fine2 aja bang secara anak ke dua baru lahiran. trus jg renovasi rumah tipis2 klo soal mobil gw bln pnya krna mobil d kasih pinjam kntor bwt oprasional

4.11 Pembahasan Hasil Penelitian

Secara keseluruhan, hasil analisis *confusion matrix*, kurva ROC, serta proses pelatihan BiLSTM menunjukkan bahwa *Support Vector Machine* merupakan model yang paling stabil dan konsisten dalam klasifikasi gender pengguna Twitter berbahasa Indonesia. Keunggulan SVM dapat dijelaskan oleh kemampuannya dalam menangani data berdimensi tinggi seperti TF-IDF serta stabilitas model pada *dataset* berukuran kecil hingga menengah.

Analisis hasil menunjukkan bahwa setiap model memiliki karakteristik yang berbeda dalam menangkap pola linguistik pada *tweet* berbahasa Indonesia. *Multinomial Naïve Bayes* cenderung bekerja baik pada kata-kata yang muncul secara konsisten antar kelas, namun kurang sensitif terhadap konteks. Sementara itu, SVM mampu mempelajari batas pemisah antar kelas secara lebih optimal terutama pada teks pendek, sehingga menghasilkan performa tertinggi. Sebaliknya, model BiLSTM menunjukkan performa lebih rendah karena kebutuhan

data yang besar untuk dapat menangkap pola sekuensial dengan baik. Hasil ini mengindikasikan bahwa kompleksitas model tidak selalu berbanding lurus dengan performa, terutama pada kondisi dataset terbatas, variasi bahasa yang tinggi, serta kualitas pelabelan yang tidak sempurna. Dengan demikian, hasil penelitian ini memperlihatkan bahwa metode klasik tetap menjadi pilihan yang efektif untuk tugas klasifikasi gender berbasis teks Indonesia.

Nilai akurasi model yang berada di bawah 75% dipengaruhi oleh beberapa faktor. Pertama, ukuran dataset yang relatif terbatas menyebabkan model tidak memperoleh keragaman pola bahasa yang cukup untuk membedakan gender secara konsisten. Kedua, *tweet* berbahasa Indonesia sangat variatif, mengandung campuran bahasa, slang, penyingkatan, serta gaya penulisan informal yang sulit dinormalisasi sepenuhnya pada tahap pra-pemrosesan. Ketiga, pelabelan manual berdasarkan nama dan *username* tidak sepenuhnya akurat, sehingga kemungkinan terdapat kesalahan label yang memengaruhi proses pembelajaran model. Selain itu, tidak adanya metadata tambahan seperti informasi profil, minat, atau riwayat *tweet* membuat model hanya bergantung pada fitur teks murni, yang pada banyak kasus tidak memiliki sinyal linguistik yang kuat untuk membedakan gender. Kombinasi faktor-faktor tersebut berkontribusi terhadap performa model yang belum mencapai 75%.

4.12 Kelebihan dan Keterbatasan Penelitian

Penelitian ini memiliki beberapa kelebihan yang mendukung validitas dan kontribusi ilmiahnya. Pertama, penelitian ini menggunakan alur metodologi yang terstandar dan konsisten untuk seluruh algoritma yang dibandingkan, mulai dari tahap pengumpulan data, pra-pemrosesan, penyeimbangan data, hingga evaluasi model. Dengan pendekatan ini, perbedaan performa yang dihasilkan dapat dikaitkan secara langsung dengan karakteristik masing-masing algoritma, bukan akibat perbedaan perlakuan data. Kedua, penelitian ini berfokus pada klasifikasi gender berbasis teks murni, tanpa memanfaatkan metadata pengguna. Pendekatan ini menjadikan hasil penelitian lebih umum dan aplikatif, terutama pada kondisi ketika informasi profil pengguna tidak tersedia atau tidak dapat dipercaya.

Ketiga, penggunaan berbagai metrik evaluasi, termasuk *accuracy*, *macro F1-score*, *confusion matrix*, dan ROC–AUC, memungkinkan analisis performa model yang lebih komprehensif. Pemilihan *macro F1-score* sebagai metrik utama juga menjadi kelebihan karena memberikan evaluasi yang adil pada kondisi distribusi kelas yang sebelumnya tidak seimbang. Keempat, penelitian ini memberikan pembandingan langsung antara metode pembelajaran mesin klasik dan pembelajaran mendalam pada konteks *tweet* berbahasa Indonesia. Hasil yang

menunjukkan keunggulan metode klasik pada data terbatas memberikan kontribusi empiris yang relevan bagi pengembangan sistem NLP di bahasa dengan sumber daya terbatas.

Meskipun memiliki beberapa kelebihan, penelitian ini juga memiliki sejumlah keterbatasan yang perlu diperhatikan. Pertama, jumlah *dataset* yang digunakan relatif terbatas, sehingga model pembelajaran mendalam seperti BiLSTM belum dapat memaksimalkan kemampuannya dalam mempelajari pola konteks linguistik yang kompleks. Keterbatasan ini terlihat dari gejala overfitting pada proses pelatihan BiLSTM. Kedua, pelabelan gender dilakukan secara manual berdasarkan informasi *username* dan *name* pengguna. Meskipun pendekatan ini menjaga kualitas anotasi, metode tersebut berpotensi mengandung subjektivitas dan tidak sepenuhnya merepresentasikan identitas gender pengguna secara akurat.

Ketiga, penelitian ini hanya membatasi klasifikasi pada dua kelas gender (*male* dan *female*). Pendekatan ini belum mencakup representasi gender yang lebih beragam, sehingga hasil penelitian belum dapat digeneralisasikan untuk klasifikasi gender non-biner. Keempat, model pembelajaran mendalam yang digunakan masih terbatas pada arsitektur BiLSTM sederhana, tanpa eksplorasi lanjutan terhadap model berbasis Transformer yang saat ini banyak digunakan dalam tugas NLP.

Keempat, penelitian ini memiliki keterbatasan pada tahap pelabelan gender, karena penentuan label dilakukan secara manual berdasarkan nama dan username pengguna. Metode ini tidak menjamin keakuratan 100% dan berpotensi mengandung bias interpretasi. Selain itu, penelitian tidak melakukan verifikasi gender melalui akun yang sudah terkonfirmasi identitasnya sehingga akurasi label bergantung penuh pada informasi publik yang tersedia. Berdasarkan kelebihan dan keterbatasan yang telah dibahas, maka saran pengembangan penelitian dirumuskan secara ringkas pada BAB V.

BAB V
PENUTUP

5.1 Kesimpulan

Kesimpulan penelitian ini disusun untuk menjawab secara langsung rumusan masalah yang telah dirumuskan pada BAB I. Berdasarkan hasil pengujian dan pembahasan yang telah dipaparkan pada BAB IV, maka dapat ditarik kesimpulan sebagai berikut.

a. Kinerja *Multinomial Naïve Bayes*

Algoritma *Multinomial Naïve Bayes* mampu mengklasifikasikan gender pengguna Twitter berbahasa Indonesia dengan performa yang cukup baik. Berdasarkan hasil pengujian, model ini menghasilkan akurasi sekitar 66% dan *macro F1-score* sekitar 0,66. Hasil tersebut menunjukkan bahwa *Multinomial Naïve Bayes* masih relevan digunakan sebagai metode *baseline* pada klasifikasi teks pendek dengan representasi TF–IDF.

b. Kinerja *Support Vector Machine* (SVM)

Algoritma *Support Vector Machine* menunjukkan kinerja terbaik dibandingkan model lainnya dalam penelitian ini. Model SVM menghasilkan akurasi tertinggi, yaitu sekitar 67%, dengan *macro F1-score* sekitar 0,66, serta distribusi kesalahan klasifikasi yang relatif seimbang pada kedua kelas gender. Hal ini menegaskan bahwa SVM sangat efektif dalam menangani data teks berdimensi tinggi dan bersifat jarang (*sparse*), seperti fitur TF–IDF pada *tweet* berbahasa Indonesia.

c. Kinerja *Bidirectional Long Short-Term Memory* (BiLSTM)

Model *Bidirectional Long Short-Term Memory* menunjukkan performa yang lebih rendah dibandingkan metode klasik. Berdasarkan hasil pengujian, BiLSTM menghasilkan akurasi sekitar 57% dan *macro F1-score* sekitar 0,52. Meskipun secara teoretis BiLSTM mampu menangkap konteks sekuensial teks, performa model ini belum optimal pada *dataset* yang relatif terbatas dan teks *tweet* yang bersifat pendek serta informal.

d. Perbandingan Performa Antar Algoritma

Berdasarkan evaluasi menggunakan metrik *accuracy*, *macro F1-score*, *confusion matrix*, dan ROC–AUC, dapat disimpulkan bahwa *Support Vector Machine* merupakan algoritma dengan performa terbaik dalam klasifikasi gender pengguna Twitter berbahasa Indonesia pada penelitian ini. *Multinomial Naïve Bayes* berada pada posisi kedua dengan performa yang cukup kompetitif, sedangkan BiLSTM menunjukkan performa terendah.

e. Kesimpulan Umum Penelitian

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pada klasifikasi gender berbasis teks *tweet* berbahasa Indonesia dengan jumlah data yang relatif terbatas, algoritma pembelajaran mesin klasik (khususnya SVM) masih lebih efektif dan stabil dibandingkan model pembelajaran mendalam yang lebih kompleks. Dengan demikian, pemilihan metode klasifikasi hendaknya disesuaikan dengan karakteristik dan skala data yang digunakan. Dengan demikian, seluruh rumusan masalah yang diajukan pada BAB I telah terjawab secara sistematis, baik dari sisi performa masing-masing algoritma maupun perbandingan kinerjanya.

5.2 Saran

Berdasarkan hasil penelitian serta pembahasan mengenai kelebihan dan keterbatasan penelitian yang telah diuraikan pada BAB IV, beberapa saran untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Menggunakan *dataset* dengan jumlah data yang lebih besar dan periode pengambilan data yang lebih panjang agar model pembelajaran mendalam dapat dilatih secara lebih optimal.
2. Menerapkan model berbasis Transformer, seperti IndoBERT atau BERT multibahasa, untuk meningkatkan kemampuan representasi konteks teks.
3. Mengombinasikan fitur teks dengan metadata pengguna, seperti deskripsi profil atau aktivitas akun, guna meningkatkan performa klasifikasi gender.
4. Mengembangkan metode pelabelan gender yang lebih sistematis atau semi-otomatis untuk mengurangi potensi subjektivitas dalam proses anotasi data.
5. Memperluas cakupan penelitian pada domain topik dan platform media sosial lain guna meningkatkan generalisasi hasil penelitian.
6. Melakukan proses pelabelan gender menggunakan akun yang identitasnya sudah jelas, seperti akun yang sudah *verified*, akun tokoh publik, atau akun yang memiliki informasi gender yang terkonfirmasi.
7. Mengutamakan akun yang kredibel atau memiliki riwayat aktivitas yang jelas, sehingga kesalahan pelabelan dapat diminimalkan.
8. Memanfaatkan sumber data tambahan untuk memvalidasi gender pengguna, sehingga subjektivitas dalam pelabelan manual dapat dikurangi.

DAFTAR PUSTAKA

Benítez-Andrades, J. A., Prada-García, C., Ordás-Reyes, N., Blanco, M. E., Merayo, A., & Serrano-García, A. (2025). Enhanced prediction of spine surgery outcomes using advanced machine learning techniques and oversampling methods. *Health Information Science and Systems*, *13*(1), 24. https://doi.org/10.1007/s13755-025-00343-9

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating Gender on Twitter. *Proceedings ofthe 2011 Conference on Empirical Methods in Natural Language Processing*, 1301–1309.

Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author gender identification from text. *Digital Investigation*, *8*(1), 78–88. https://doi.org/10.1016/j.diin.2011.04.002

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Fields, J., Chovanec, K., & Madiraju, P. (2024). A Survey of Text Classification With Transformers: How Wide? How Large? How Long? How Accurate? How Expensive? How Safe? *IEEE Access*, *12*, 6518–6531. https://doi.org/10.1109/ACCESS.2024.3349952

Fikri, M. I., Sabrila, T., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, *10*, 71–76. https://doi.org/10.32664/smatika.v10i02.455

Helmayanti, S. A., Hamami, F., & Fa'rifah, R. Y. (2023). PENERAPAN ALGORITMA TF-IDF DAN NAÏVE BAYES UNTUK ANALISIS SENTIMEN BERBASIS ASPEK ULASAN APLIKASI FLIP PADA GOOGLE PLAY STORE. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, *4*(3), 1822–1834. https://doi.org/10.35870/jimik.v4i3.415

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Humam, C., & Laksito, A. D. (2023). Implementasi Aplikasi Sentimen Pada Data Twitter Jelang Pemilu 2024. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, *8*(2).

Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. Dalam C. Nédellec & C. Rouveirol (Ed.), *Machine Learning: ECML-98* (hlm. 137–142). Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0026683

Katna, R., Kalsi, K., Gupta, S., Yadav, D., & Yadav, A. K. (2022). Machine learning based approaches for age and gender prediction from tweets. *Multimedia Tools and Applications*, *81*(19), 27799–27817. https://doi.org/10.1007/s11042-022-12920-1

Mahendra, R., Putra, H. S., Faisal, D. R., & Rizki, F. (2022). Gender Prediction of Indonesian Twitter Users Using Tweet and Profile Features. *Jurnal Ilmu Komputer dan Informasi*, *15*(2), 131–141. https://doi.org/10.21609/jiki.v15i2.1079

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511809071

Masood, K., Mehmood, A., & Nizami, M. (2023). Identification of Age and Gender on Twitter Using DenseNet and LSTM. *Journal of Computing & Biomedical Informatics*, *5*(1). https://doi.org/10.56979/501/2023

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:7311285

Nhita, F., Adiwijaya, & Kurniawan, I. (2023). Performance and Statistical Evaluation of Three Sampling Approaches in Handling Binary Imbalanced Data Sets. *2023 International Conference on Data Science and Its Applications (ICoDSA)*, 420–425. https://doi.org/10.1109/ICoDSA58501.2023.10276805

Nia, Z. M., Ahmadi, A., Mellado, B., Wu, J., Orbinski, J., Asgary, A., & Kong, J. D. (2023). Twitter-based gender recognition using transformers. *Mathematical Biosciences and Engineering*, *20*(9), 15957–15977. https://doi.org/10.3934/mbe.2023711

O'Connor, K., Golder, S., Weissenbacher, D., Klein, A. Z., Magge, A., & Gonzalez-Hernandez, G. (2024). Methods and Annotated Data Sets Used to Predict the Gender and Age of Twitter Users: Scoping Review. Dalam *Journal of Medical Internet Research* (Vol. 26, Nomor 1). JMIR Publications Inc. https://doi.org/10.2196/47923

Rianto, Mutiara, A. B., Wibowo, E. P., & Santosa, P. I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, *8*(1), 26. https://doi.org/10.1186/s40537-021-00413-1

Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, *874*(1), 012017. https://doi.org/10.1088/1757-899X/874/1/012017

Sa'diah, H., Faisal, M. R., Farmadi, A., Abadi, F., Indriani, F., Alkaff, M., & Abullayev, V. (2024). Gender Classification on Social Media Messages Using fastText-base Feature Extraction and Long Short-Term Memory. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *6*(3), 243–252. https://doi.org/10.35882/jeeemi.v6i3.407

Selma Zakia, A. (2020). Klasifikasi Jenis Kelamin Pengguna Twitter dengan menggunakan Metode BM25 dan K-Nearest Neighbor (KNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *4*(10), 3331–3337. http://j-ptiik.ub.ac.id

Tasnim, R., Chowdhury, M., & Rahman, M. A. (2024). *BN-AuthProf: Benchmarking Machine Learning for Bangla Author Profiling on Social Media Texts*. https://doi.org/10.48550/arXiv.2412.02058

LAMPIRAN

Lampiran A : Kode program proses *scrapping* data

```
import requests
import pandas as pd
import time

# ======================
# 1. KONFIGURASI DASAR
# ======================

API_KEY = "new1_1bab7b94912047d1b32083e51c34f7bf"  # saran: nanti ganti lagi ya, jgn dishare ke orang lain 🤓

QUERY = """
("kuliah" OR "kerja" OR "parenting" OR "overthinking" OR "pinjol" OR "banjir") lang:id
""".strip()

MAX_TWEETS = 200  # target jumlah tweet

URL = "https://api.twitterapi.io/twitter/tweet/advanced_search"
HEADERS = {"X-API-Key": API_KEY}

# ======================
# 2. LOOP SCRAPING
# ======================

all_tweets = []
cursor = None

while len(all_tweets) < MAX_TWEETS:
    params = {"query": QUERY, "queryType": "Latest"}
    if cursor:
        params["cursor"] = cursor

    print(f"Mengambil data... Saat ini sudah terkumpul {len(all_tweets)} tweet")

    resp = requests.get(URL, headers=HEADERS, params=params)
```

```
    # 👉 HANDLE RATE LIMIT 429
    if resp.status_code == 429:
        print(
            "Kena limit 429 (Too Many Requests). Istirahat dulu 6 detik,
lalu coba lagi..."
        )
        print("Pesan server:", resp.text)
        time.sleep(6)
        continue  # ulangi loop dengan cursor yang sama

    # Kalau ada error HTTP lain (401, 500, dsb) → baru kita hentikan
    try:
        resp.raise_for_status()
    except Exception as e:
        print("Terjadi error saat request (bukan 429):")
        print(resp.text)
        raise e

    data = resp.json()

    tweets = data.get("tweets", [])
    if not tweets:
        print("Tidak ada tweet lagi yang bisa diambil. Stop.")
        break

    print(f"Batch ini dapat {len(tweets)} tweet")
    all_tweets.extend(tweets)

    # Cek apakah masih ada halaman berikutnya
    if not data.get("has_next_page"):
        print("has_next_page = False. Stop.")
        break

    cursor = data.get("next_cursor")
    if not cursor:
        print("Cursor kosong. Stop.")
        break

    # Delay untuk jaga-jaga QPS limit (free-tier: 1 request / 5 detik)
```

```
    time.sleep(5.2)

print(f"Selesai mengambil {len(all_tweets)} tweet (mentah).")

# ======================
# 3. SIMPAN KE DATAFRAME
# ======================

rows = []
for t in all_tweets[:MAX_TWEETS]:
    author = t.get("author", {}) or {}
    rows.append(
        {
            "tweet_id": t.get("id"),
            "created_at": t.get("createdAt"),
            "lang": t.get("lang"),
            "text": t.get("text"),
            "username": author.get("userName"),  # handle
            "name": author.get("name"),  # display name
            "author_id": author.get("id"),
            "bio": author.get("description"),
            "gender": "unknown",
        }
    )

df = pd.DataFrame(rows)
df.drop_duplicates(subset=["tweet_id"], inplace=True)

output_file = "tweets_indonesia_gender_project4.csv"
df.to_csv(output_file, index=False, encoding="utf-8")

print(f"Total tweet bersih: {len(df)}")
print(f"File CSV berhasil dibuat: {output_file}")
```

Lampiran B : Kode program proses permodelan

```
# ====================================================
# CONFIGURATION, IMPORTS, UTILITIES, LOAD DATA
# ====================================================
```

```
# ---------------------
# IMPORTS
# ---------------------
import os
import json
import time
import re
import random
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.preprocessing import LabelEncoder
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import LinearSVC
from sklearn.metrics import (
    accuracy_score,
    classification_report,
    confusion_matrix,
    ConfusionMatrixDisplay,
    f1_score,
    roc_curve,
    auc,
)

from imblearn.over_sampling import RandomOverSampler

import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Embedding, LSTM, Dense, Dropout,
Bidirectional
from tensorflow.keras.callbacks import EarlyStopping

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory
```

```
# ---------------------
# CONFIGURATION
# ---------------------
SEED = 42
INPUT_CSV = "scrapping/tweets_indonesia_gabungan.csv"
OUTPUT_DIR = "outputs_80.20."
FIG_DIR = os.path.join(OUTPUT_DIR, "figures")

os.makedirs(FIG_DIR, exist_ok=True)

np.random.seed(SEED)
random.seed(SEED)
tf.random.set_seed(SEED)

# ---------------------
# HELPER FUNCTIONS
# ---------------------
def save_json(obj, path):
    def converter(o):
        if isinstance(o, (np.integer, np.floating)):
            return o.item()
        if isinstance(o, np.ndarray):
            return o.tolist()
        return str(o)

    with open(path, "w", encoding="utf-8") as f:
        json.dump(obj, f, ensure_ascii=False, indent=2,
default=converter)

def get_tp_fn_tn_fp(y_true, y_pred, positive_label="male"):
    y_true_bin = (y_true == positive_label).astype(int)
    y_pred_bin = (y_pred == positive_label).astype(int)
    tn, fp, fn, tp = confusion_matrix(y_true_bin, y_pred_bin).ravel()
    return tp, fn, tn, fp

# ---------------------
# LOAD DATA
# ---------------------
print(f"Loading dataset: {INPUT_CSV}")
```

```
data = pd.read_csv(INPUT_CSV, encoding="cp1252")

required_cols = ["text", "gender"]
for col in required_cols:
    if col not in data.columns:
        raise ValueError(f"Kolom '{col}' tidak ditemukan")

data = data[[c for c in ["createdAt", "text", "gender"] if c in
data.columns]]
print("Total data awal:", len(data))

# ======================================================
# PREPROCESSING & BALANCING
# ======================================================

# ---------------------
# NLP SETUP
# ---------------------
stemmer = StemmerFactory().create_stemmer()
stop_factory = StopWordRemoverFactory()
custom_stopwords = {
    "aja", "kok", "sih", "dong","lah","ya", "gak", "ga", "gk", "gw",
    "gua","lu","aku","kamu", "tweet","retweet", "wkwk", "haha", "hehe",
}
stop_words = set(stop_factory.get_stop_words()) | custom_stopwords
normalization_dict = {
    "ga": "tidak",
    "gk": "tidak",
    "gak": "tidak",
    "tdk": "tidak",
    "klo": "kalau",
    "krn": "karena",
    "yg": "yang",
    "dgn": "dengan",
}

def normalize_text(text):
    return " ".join(normalization_dict.get(w, w) for w in text.split())

def clean_text(text):
```

```
    text = str(text).lower()
    text = re.sub(r"http\S+|@\w+|#\w+|rt\b", "", text)
    text = re.sub(r"[^a-z\s]", " ", text)
    text = re.sub(r"\s+", " ", text).strip()
    text = normalize_text(text)

    tokens = [
        stemmer.stem(w) for w in text.split() if w not in stop_words and
len(w) > 2
    ]
    return " ".join(tokens)

# ---------------------
# CLEANING
# ---------------------
data = data.dropna(subset=["text", "gender"])
data["clean_tweet"] = data["text"].apply(clean_text)
data = data[data["gender"].isin(["male", "female"])]
data = data[data["clean_tweet"].str.strip() != ""]

counts_after_clean = data["gender"].value_counts().to_dict()

# ---------------------
# BALANCING
# ---------------------
ros = RandomOverSampler(random_state=SEED)
X_res, y_res = ros.fit_resample(data[["clean_tweet"]], data["gender"])

data_bal = pd.DataFrame({"clean_tweet": X_res["clean_tweet"], "gender":
y_res})

# ---------------------
# VISUALIZATION
# ---------------------
labels = ["male", "female"]
before = [counts_after_clean.get(l, 0) for l in labels]
after = data_bal["gender"].value_counts().reindex(labels).fillna(0)

x = np.arange(len(labels))
plt.figure(figsize=(8, 5))
```

```
plt.bar(x - 0.2, before, 0.4, label="Sebelum Balancing")
plt.bar(x + 0.2, after, 0.4, label="Sesudah Balancing")

plt.xlabel("Kelas Gender")
plt.ylabel("Jumlah Data")
plt.title("Distribusi Gender Sebelum & Sesudah Penyeimbangan")
plt.xticks(x, labels)
plt.legend()
plt.grid(axis="y", linestyle="--", alpha=0.6)
plt.tight_layout()
plt.savefig(os.path.join(FIG_DIR, "distribusi_gender_balancing.png"),
dpi=300)
plt.close()

# ======================================================
# MODELING & EVALUATION
# ======================================================

# ---------------------
# SPLIT DATA
# ---------------------
X = data_bal["clean_tweet"]
y = data_bal["gender"]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=SEED
)

# ---------------------
# TF-IDF
# ---------------------
vectorizer = TfidfVectorizer(
    max_features=15000,
    ngram_range=(1, 2),
    sublinear_tf=True,
    min_df=2,
    max_df=0.8,
)

X_train_tfidf = vectorizer.fit_transform(X_train)
```

```
X_test_tfidf = vectorizer.transform(X_test)

# ---------------------
# NAIVE BAYES
# ---------------------
nb = MultinomialNB()
nb.fit(X_train_tfidf, y_train)
nb_pred = nb.predict(X_test_tfidf)

# ---------------------
# SVM
# ---------------------
svm = LinearSVC(C=0.8, class_weight="balanced", dual=False,
max_iter=5000)
svm.fit(X_train_tfidf, y_train)
svm_pred = svm.predict(X_test_tfidf)

# ---------------------
# BiLSTM
# ---------------------
tokenizer = Tokenizer(num_words=10000, oov_token="<OOV>")
tokenizer.fit_on_texts(X_train)

X_train_seq = pad_sequences(
    tokenizer.texts_to_sequences(X_train), maxlen=50, padding="post"
)
X_test_seq = pad_sequences(
    tokenizer.texts_to_sequences(X_test), maxlen=50, padding="post"
)

encoder = LabelEncoder()
y_train_enc = encoder.fit_transform(y_train)
y_test_enc = encoder.transform(y_test)

model = Sequential(
    [
        Embedding(len(tokenizer.word_index) + 1, 200, input_length=50),
        Bidirectional(
            LSTM(128, dropout=0.3, recurrent_dropout=0.2,
return_sequences=True)
```

```
        ),
        Bidirectional(LSTM(64, dropout=0.3, recurrent_dropout=0.2)),
        Dense(64, activation="relu"),
        Dropout(0.4),
        Dense(1, activation="sigmoid"),
    ]
)

model.compile(loss="binary_crossentropy", optimizer="adam",
metrics=["accuracy"])

model.fit(
    X_train_seq,
    y_train_enc,
    epochs=10,
    batch_size=32,
    validation_split=0.2,
    callbacks=[EarlyStopping(patience=2, restore_best_weights=True)],
    verbose=1,
)

# ====================================================
# ANALYSIS & OUTPUT
# ====================================================

# Confusion Matrix NB
ConfusionMatrixDisplay.from_predictions(
    y_test, nb_pred, display_labels=["male", "female"], cmap="Blues"
)
plt.savefig(os.path.join(OUTPUT_DIR, "confusion_nb.png"), dpi=300)
plt.close()

# Confusion Matrix SVM
ConfusionMatrixDisplay.from_predictions(
    y_test, svm_pred, display_labels=["male", "female"], cmap="Blues"
)
plt.savefig(os.path.join(OUTPUT_DIR, "confusion_svm.png"), dpi=300)
plt.close()

# Confusion Matrix BiLSTM
```

```
bilstm_pred_probs = model.predict(X_test_seq)
bilstm_pred = np.where(bilstm_pred_probs > 0.5, 1, 0)
ConfusionMatrixDisplay.from_predictions(
    y_test_enc, bilstm_pred, display_labels=["male", "female"],
cmap="Blues"
)
plt.savefig(os.path.join(OUTPUT_DIR, "confusion_bilstm.png"), dpi=300)

# Save models
import joblib

joblib.dump(nb, os.path.join(OUTPUT_DIR, "model_nb.pkl"))
joblib.dump(svm, os.path.join(OUTPUT_DIR, "model_svm.pkl"))
model.save(os.path.join(OUTPUT_DIR, "model_bilstm.keras"))

print("Pipeline selesai. Semua output tersimpan.")
```

Lampiran A.3 Daftar *stopword* yang digunakan dalam proses pra-pemrosesan data

ada	melainkan	kami
adalah	mengapa	kamu
agak	menurut	kan
agar	mereka	karena
aja	namun	kaya
akan	nanti	kayak
aku	nggak	ke
akun	oh	kecuali
amat	ok	kemana
anda	oke	kembali
anj	oleh	kenapa
anjg	orang	kepada
antara	org	ketika
anu	pada	kita
apakah	para	kok
apalagi	pasti	lagi
astaga	post	lah
atau	posting	lain
bagaimanapun	pula	loh
bagi	pun	lu
bahwa	retweet	maka
bang	saat	mari
begitu	saja	masih

belum	sambil	
bisa	sampai	
boleh	saya	
bos	sebab	
bro	sebagai	
buset	sebelum	
dahulu	sebetulnya	
dalam	secara	
dan	sedangkan	
dapat	seharusnya	
dari	sehingga	
daripada	sekitar	
deh	selagi	
demi	selain	
demikian	sementara	
dengan	seolah	
di	seperti	
dia	seraya	
dimana	serta	
dll	sesuatu	
dong	sesudah	
dsb	setelah	
dst	seterusnya	
dua	setiap	
dulunya	setidaknya	
enggak	sih	
ga	sis	
gak	sudah	
gan	supaya	
gila	tanpa	
gini	tapi	
gitu	telah	
gk	tentang	
gua	tentu	
guna	terhadap	
gw	tetapi	
haha	tidak	
hal	toh	
hanya	tolong	
harus	tuh	
hehe	tweet	
ia	udah	
ingin	udh	
ini	untuk	

itu	user	
itulah	walau	
iya	wk	
jadi	wkwk	
jd	ya	
jika	yaitu	
juga	yakni	
kah	yang	

Lampiran 2 Daftar kata untuk normalisasi kata pada proses *normalization*

ada	haha	saat
adalah	hal	saja
agak	hanya	sambil
agar	harus	sampai
aja	hehe	saya
akan	ia	sebab
aku	ingin	sebagai
akun	ini	sebelum
amat	itu	sebetulnya
anda	itulah	secara
anj	iya	sedangkan
anjg	jadi	seharusnya
antara	jd	sehingga
anu	jika	sekitar
apakah	juga	selagi
apalagi	kah	selain
astaga	kami	sementara
atau	kamu	seolah
bagaimanapun	kan	seperti
bagi	karena	seraya
bahwa	kaya	serta
bang	kayak	sesuatu
begitu	ke	sesudah
belum	kecuali	setelah
bisa	kemana	seterusnya
boleh	kembali	setiap
bos	kenapa	setidaknya
bro	kepada	sih
buset	ketika	sis
dahulu	kita	sudah
dalam	kok	supaya
dan	lagi	tanpa
dapat	lah	tapi

dari	lain	telah
daripada	loh	tentang
deh	lu	tentu
demi	maka	terhadap
demikian	mari	tetapi
dengan	masih	tidak
di	melainkan	toh
dia	mengapa	tolong
dimana	menurut	tuh
dll	mereka	tweet
dong	namun	udah
dsb	nanti	udh
dst	nggak	untuk
dua	oh	user
dulunya	ok	walau
enggak	oke	wk
ga	oleh	wkwk
gak	orang	ya
gan	org	yaitu
gila	pada	yakni
gini	para	yang
gitu	pasti	retweet
gk	post	
gua	posting	
guna	pula	
gw	pun	