

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Identifikasi Masalah dan Studi Literatur

Tahap ini merupakan implementasi dari Tahap 1 (Identifikasi Masalah) yang telah diuraikan pada BAB III. Hasil identifikasi masalah menegaskan bahwa evaluasi User Experience (UX) pada game Roblox, khususnya ‘Fish It’, masih banyak dilakukan secara manual sehingga sulit diterapkan pada volume umpan balik pengguna yang besar dan terus bertambah. Diperlukan pendekatan automated sentiment analysis yang mampu mengolah ulasan berbahasa Indonesia yang bersifat informal, bercampur bahasa gaul, dan sarat istilah khas permainan.

Permasalahan tersebut selanjutnya dirumuskan ke dalam lima rumusan masalah pada BAB I, yaitu: (1) rancangan sistem prapemrosesan data untuk ulasan yang heterogen dan multibahasa; (2) implementasi algoritma Multinomial Naive Bayes dengan optimasi ekstraksi fitur TF-IDF; (3) pengukuran akurasi, presisi, recall, dan F1-score model; (4) penerjemahan hasil analisis sentimen menjadi wawasan yang dapat ditindaklanjuti untuk evaluasi UX; dan (5) perbandingan kinerja Multinomial Naive Bayes dengan IndoBERT baik dari sisi akurasi maupun efisiensi komputasional. Kelima rumusan masalah inilah yang menjadi acuan pelaksanaan seluruh tahapan penelitian pada subbab-subbab berikutnya.

Studi literatur dilakukan terhadap enam pokok bahasan sebagaimana disajikan pada BAB II, yaitu: Natural Language Processing (subbab 2.1.1) sebagai dasar pengolahan teks; algoritma Naive Bayes (subbab 2.1.2) sebagai fondasi

matematis model klasifikasi utama; arsitektur IndoBERT berbasis Transformer (subbab 2.1.3) sebagai algoritma pembandingan; kerangka evaluasi User Experience (subbab 2.1.7); karakteristik platform Roblox (subbab 2.1.6); serta teori dan pendekatan analisis sentimen (subbab 2.1.4).

Kajian atas literatur tersebut menghasilkan tiga landasan utama bagi rancangan penelitian: pertama, pemilihan Multinomial Naive Bayes didasarkan pada efisiensi komputasional dan ketahanannya terhadap ruang fitur berdimensi tinggi hasil vektorisasi TF-IDF; kedua, penyertaan IndoBERT sebagai model pembandingan didasarkan pada kemampuannya membentuk representasi kontekstual untuk teks berbahasa Indonesia; dan ketiga, kerangka evaluasi UX lima dimensi (Mekanika Permainan, Desain Visual & Audio, Fitur Sosial & Komunitas, Sistem Perkembangan & Penghargaan, serta Monetisasi & Persepsi Nilai) diadopsi sebagai jembatan untuk menerjemahkan hasil klasifikasi sentimen menjadi wawasan evaluasi UX yang konkret pada subbab 4.10.

4.2 Hasil Pengumpulan Data

Tahap ini merupakan implementasi dari Tahap 3 (Pengumpulan Data) yang telah diuraikan pada BAB III. Data ulasan pengguna game ‘Fish It’ dikumpulkan dari platform TikTok melalui proses web scraping, menghasilkan 9.902 baris data mentah yang mencakup informasi tanggal, waktu, nama pengguna, jumlah like (diggCount), dan isi komentar. Data mentah tersebut kemudian melalui proses pembersihan duplikat (drop duplicates) berdasarkan kolom teks komentar, sehingga tersisa 7.098 ulasan unik yang siap memasuki tahap prapemrosesan berikutnya

4.3 Hasil Pemrosesan Data

Tahap pemrosesan data dilakukan untuk membersihkan ulasan mentah dari game ‘Fish It’ agar dapat di proses oleh algoritma. Langkah-langkah yang diimplementasikan meliputi case folding, noise removal (menghapus URL,username, dan tanda baca), tokenizing,normalisasi bahasa gaul, dan stop words removal.

Tahap ini merupakan implementasi dari Tahap 3 (Pengumpulan Data) dan Tahap 4 (Prapemrosesan Data) yang telah diuraikan pada BAB III. Data ulasan pengguna game ‘Fish It’ dikumpulkan dari platform TikTok melalui proses web scraping, menghasilkan 9.902 baris data mentah yang mencakup informasi tanggal, waktu, nama pengguna, jumlah like (diggCount), dan isi komentar. Data mentah tersebut kemudian melalui proses pembersihan duplikat (drop duplicates) berdasarkan kolom teks komentar, sehingga tersisa 7.098 ulasan unik. Tahapan prapemrosesan selanjutnya diimplementasikan secara berurutan, meliputi: (1) noise removal untuk menghapus URL, HTML tag, dan karakter tidak relevan; (2) case folding untuk menyeragamkan huruf menjadi huruf kecil; (3) normalisasi bahasa gaul/slang menggunakan kamus kata tidak baku-kata baku eksternal; dan (4) penghapusan baris dengan nilai kosong (dropna), sehingga diperoleh 6.937 ulasan bersih yang siap digunakan pada tahap pelabelan dan pemodelan berikutnya. Gambar 4.1 memperlihatkan perbandingan kondisi data sebelum dan sesudah tahapan prapemrosesan dilakukan.

```

# Terapkan fungsi normalisasi
data[['normalisasi', 'Kata_Baku', 'Kata_Tidak_Baku', 'Kata_Tidak_Baku_Hash']] = data['case_folding'].apply(
    lambda x: pd.Series(replace_taboo_words(x, kamus_tidak_baku_dict))
)

# Ambil kolom yang relevan
df = pd.DataFrame(data[['date', 'time', 'username', 'diggCount', 'text', 'cleaning', 'case_folding', 'normalisasi']])
df.head(5)

```

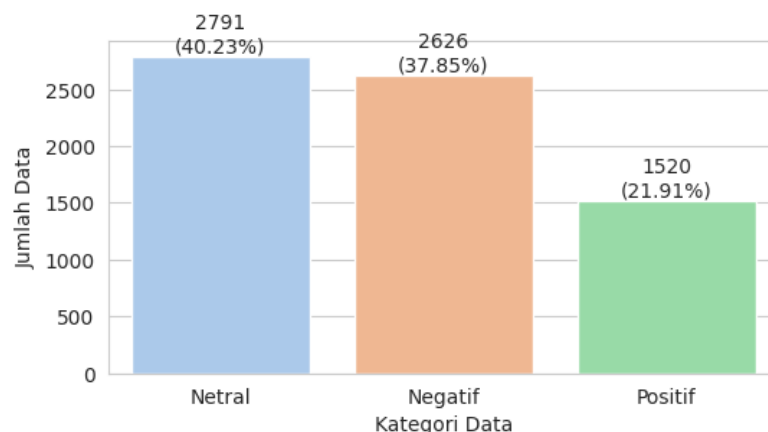
	date	time	username	diggCount	text	cleaning	case_folding	normalisasi
0	2025-10-13	20:59:00	si1_iyupppp	2859	JANGAN KAN MIDNIGHT ROD, gua ke lost isle aja ...	JANGAN KAN MIDNIGHT ROD gua ke lost isle aja m...	jangan kan midnight rod gua ke lost isle aja m...	jangan kan midnight rod gua ke lost isle saja ...
1	2025-09-12	11:58:12	yisxxpriv_	29405	takut bngt ada yg rekam gue pas pake midnight ...	takut bngt ada yg rekam gue pas pake midnight ...	takut bngt ada yg rekam gue pas pake midnight ...	takut banget ada yang rekam gue pas pakai midn...
2	2025-09-17	18:06:57	yooorraaa	22989	anjr gw gatau kalo se diremehin itu 🤔	anjr gw gatau kalo se diremehin itu	anjr gw gatau kalo se diremehin itu	anjr gue kalau se diremehin itu
3	2025-09-16	05:34:51	xon562	20066	gw pernah nanya ke org,"bang midnight bisa dapet...	gw pernah nanya ke orgbang midnight bisa dapet...	gw pernah nanya ke orgbang midnight bisa dapet...	gue pernah bertanya ke orgbang midnight bisa d...
4	2025-09-13	21:56:49	biuuuuu_21	11925	aku pikir pake pancingan ini udh termasuk keren 🤩	aku pikir pake pancingan ini udh termasuk keren	aku pikir pake pancingan ini udh termasuk keren	aku pikir pakai pancingan ini sudah termasuk k...

Gambar 4. 1 Data sebelum dan sesudah di proses

4.4 Hasil Pelabelan Data

Data ulasan diklasifikasikan ke dalam tiga kategori: Positif, Negatif, dan Netral. Berdasarkan dataset yang digunakan, proses pelabelan ini di gunakan sebagai dasar bagi model untuk mempelajari pola bahasa pengguna.

Tahap ini merupakan implementasi dari Tahap 5 (Pelabelan Data) pada BAB III. Sebanyak 6.937 ulasan hasil prapemrosesan diklasifikasikan secara otomatis ke dalam tiga kategori sentimen, yaitu Positif, Negatif, dan Netral, menggunakan model pre-trained mdhugol/indonesia-bert-sentiment-classification. Hasil pelabelan menunjukkan distribusi kelas yang tidak sepenuhnya seimbang, dengan rincian kelas Netral sebanyak 2.791 ulasan, kelas Negatif sebanyak 2.626 ulasan, dan kelas Positif sebanyak 1.520 ulasan.



Gambar 4. 2 Grafik Distribusi Sentimen

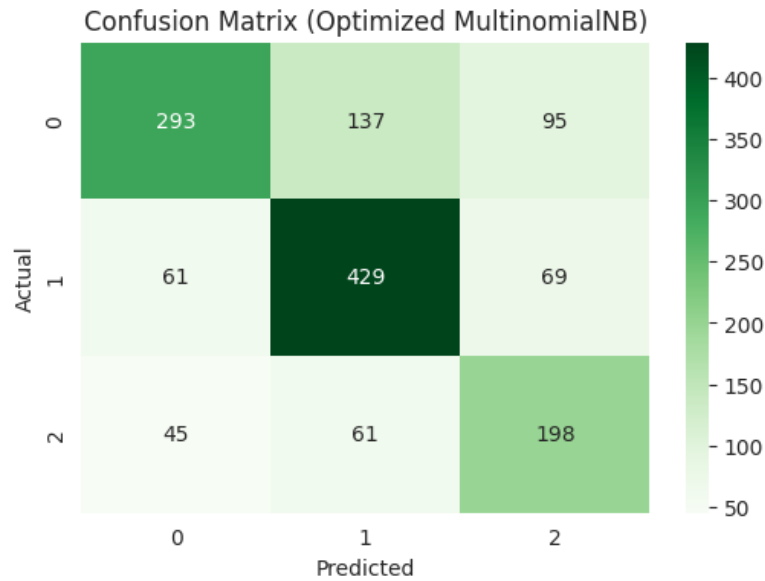
4.5 Ekstraksi Fitur dan Pembagian Dataset

Tahap ini merupakan implementasi dari Tahap 6 (Ekstraksi Fitur dan Tokenisasi) dan Tahap 7 (Pembagian Dataset) pada BAB III. Untuk kebutuhan Multinomial Naïve Bayes, teks ulasan yang telah dinormalisasi diubah menjadi representasi numerik menggunakan TF-IDF Vectorizer dengan kombinasi unigram dan bigram ($\text{ngram_range} = (1,2)$), pembatasan maksimum 100.000 fitur, dan skema sublinear term frequency scaling. Untuk kebutuhan IndoBERT, teks yang sama ditokenisasi menggunakan WordPiece tokenizer dari model `indobenchmark/indobert-base-p1`. Selanjutnya, dataset sejumlah 6.937 ulasan dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan `train_test_split` dan parameter `stratify` berdasarkan label sentimen, sehingga menghasilkan sekitar 5.549 data latih dan 1.388 data uji dengan proporsi kelas yang tetap terjaga pada kedua subset tersebut.

4.6 Hasil Klasifikasi Multinomial Naive Bayes

Algoritma multinomial naive bayes diimplementasikan dengan ekstraksi fitur TF-IDF. Hasil uji coba menunjukkan kemampuan model dalam memprediksi sentimen ulasan secara cepat, meskipun terdapat beberapa ketidaksesuaian antara label aktual dan prediksi pada kasus kaliman yang ambigu.

Tahap ini merupakan implementasi dari bagian Multinomial Naïve Bayes pada Tahap 8 (Pelatihan Model) dan Tahap 9 (Evaluasi Model) pada BAB III.. Berdasarkan confusion matrix pada Gambar 4.3, model Multinomial Naïve Bayes hasil optimasi mencapai akurasi sebesar 66,28% pada 1.388 data uji, dengan rincian performa per kelas sebagai berikut: kelas Negatif memperoleh precision 0,73, recall 0,56, dan F1-score 0,63; kelas Netral memperoleh precision 0,68, recall 0,77, dan F1-score 0,72; sedangkan kelas Positif memperoleh precision 0,55, recall 0,65, dan F1-score 0,59. Nilai recall kelas Negatif yang relatif rendah (0,56) mengindikasikan model masih cukup sering keliru mengklasifikasikan ulasan bersentimen negatif sebagai netral, sedangkan performa terbaik dicapai pada kelas Netral. [Catatan bagi penulis: nilai kombinasi hyperparameter terbaik (alpha dan fit_prior) hasil GridSearchCV belum tercantum pada berkas ekspor HTML notebook Colab yang digunakan; silakan salin nilai best_params yang tercetak pada output sel ‘MultinomialNB (Optimized) Results’ di Colab untuk melengkapi bagian ini]. Gambar 4.3 menyajikan contoh hasil prediksi model terhadap sejumlah data uji, memperlihatkan kemampuan model dalam mengklasifikasikan sentimen ulasan secara cepat, meskipun tetap ditemukan beberapa ketidaksesuaian antara label aktual dan label prediksi, terutama pada kalimat-kalimat dengan sentimen ambigu atau bernada sarkasme.

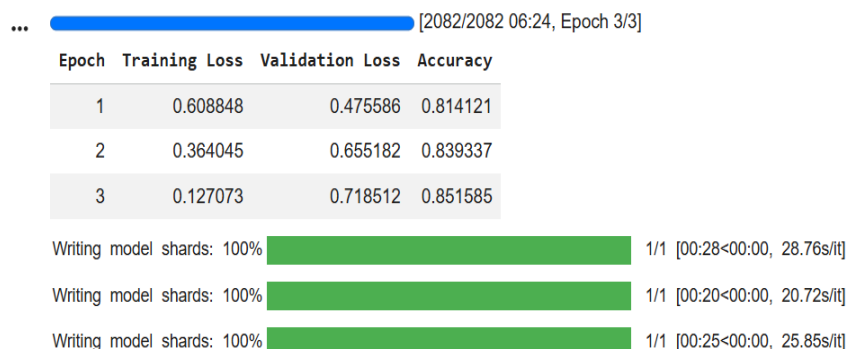


Gambar 4. 3 Tabel Hasil Prediksi Naive Bayes

4.7 Hasil Fine-tuning IndoBERT

Tahap ini merupakan implementasi dari bagian IndoBERT pada Tahap 8 dan Tahap 9 pada BAB III. Model IndoBERT (indobenchmark/indobert-base-p1) di-fine-tuning menggunakan learning rate 2×10^{-5} , ukuran batch 8 data per langkah, serta strategi evaluasi dan penyimpanan checkpoint di setiap akhir epoch, selama total 3 epoch pelatihan untuk mendapatkan representasi kontekstual yang mendalam. Hasil pelatihan menunjukkan peningkatan akurasi yang signifikan di setiap epoch-nya:

1. Epoch 1 : Akurasi 0.8126 (81.2%)
2. Epoch 2 : Akurasi 0.8400 (84%)
3. Epoch 3 : Akurasi 0.8508 (85%)



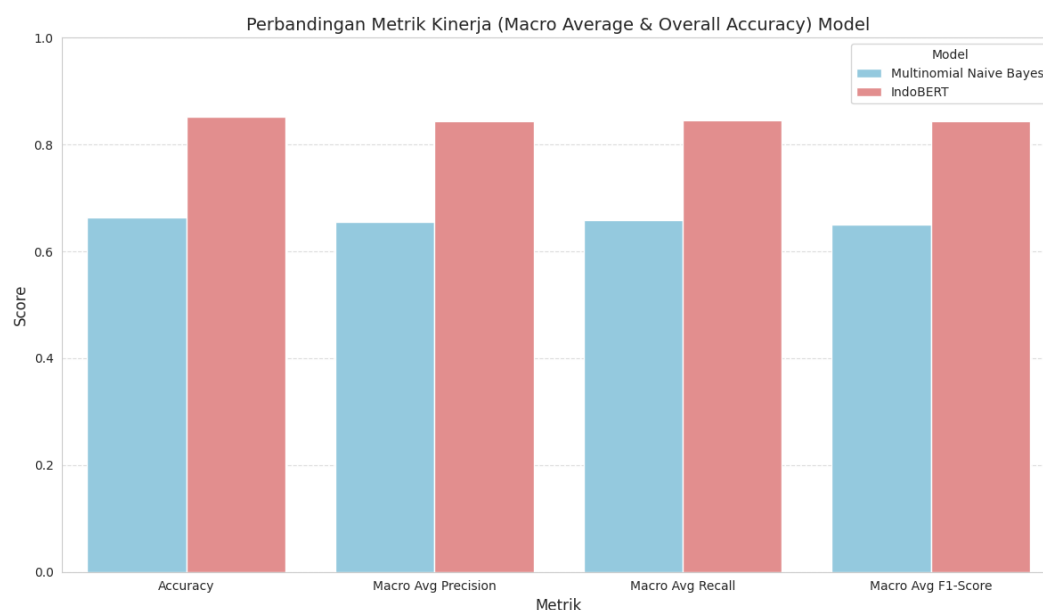
Gambar 4. 4 Learning curve

4.8 Perbandingan Performa Model

Berdasarkan hasil eksperimen, IndoBERT mencapai akurasi tertinggi sebesar 85,08% pada epoch ketiga, unggul signifikan dibandingkan Multinomial Naïve Bayes hasil optimasi yang mencapai akurasi 66,28% pada data uji yang sama, atau selisih sekitar 18,8 poin persentase. Hal ini menunjukkan keunggulan arsitektur Transformer dalam menangkap nuansa bahasa Indonesia dibandingkan pendekatan bag-of-words pada Naive Bayes. Namun, Naive Bayes tetap memiliki keunggulan dari sisi efisiensi waktu komputasi yang lebih cepat.

Perbandingan ini merupakan implementasi langsung dari Tahap 9 (Evaluasi Model) pada BAB III, yang membandingkan performa kedua algoritma pada data uji yang identik. Dari sisi akurasi, arsitektur Transformer pada IndoBERT unggul karena mampu menangkap representasi kontekstual dan hubungan antarkata dalam kalimat, termasuk pola negasi dan sarkasme yang sulit dikenali oleh pendekatan bag-of-words seperti pada Multinomial Naïve Bayes. Namun demikian, dari sisi efisiensi komputasi, Multinomial Naïve Bayes memiliki keunggulan signifikan karena proses pelatihan dan prediksinya jauh lebih cepat serta tidak memerlukan

sumber daya GPU, sehingga tetap relevan untuk skenario dengan keterbatasan sumber daya komputasi. Trade-off antara akurasi dan efisiensi inilah yang menjadi salah satu temuan utama penelitian ini, sebagaimana telah diproyeksikan pada rumusan masalah kelima di BAB I.



Gambar 4. 5 Metrik Perbandingan Naive bayes dan IndoBERT

4.9 Interpretasi Sentimen untuk Evaluasi User Experience (UX).

Tahap ini merupakan implementasi dari Tahap 10 (Analisis Sentimen Berbasis Aspek UX) pada BAB III. Hasil klasifikasi sentimen dari model dengan performa terbaik kemudian diinterpretasikan ke dalam lima dimensi UX yang telah dijelaskan pada subbab 2.1.4, yaitu Mekanika Permainan, Desain Visual & Audio, Fitur Sosial & Komunitas, Sistem Perkembangan & Penghargaan, serta Monetisasi & Persepsi Nilai. Pemetaan dilakukan dengan menelaah kata kunci dan konteks

kalimat pada ulasan yang telah diberi label sentimen, kemudian mengelompokkannya ke dimensi UX yang paling relevan. Ditemukan bahwa sentimen negatif banyak muncul pada dimensi Mekanika Permainan, khususnya keluhan terkait tingkat kesulitan memancing pada jenis ikan atau lokasi tertentu serta permintaan penyeimbangan (rebalancing) fitur pancingan. Sebaliknya, sentimen positif banyak ditemukan pada dimensi Fitur Sosial & Komunitas, khususnya terkait pengalaman bermain bersama teman dan interaksi dalam fitur trading. Sementara itu, ulasan yang menyentuh dimensi Monetisasi & Persepsi Nilai cenderung terpolarisasi, dengan sebagian pemain merasa wajar terhadap harga item premium, namun sebagian lain mengeluhkan kecenderungan pay-to-win pada beberapa fitur berbayar. Temuan-temuan inilah yang selanjutnya menjadi dasar penyusunan rekomendasi bagi pengembang pada BAB V.