



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 13370-13379

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Klasifikasi Sentimen Review Pengguna pada Platform Steam Menggunakan Arsitektur *Transformer* DistilBERT

Dody Pradipta, Eko Widodo, David Fahmi A., M. Iqbal Izzul Haq

¹Teknologi Game, Departemen Teknologi Multimedia Kreatif, Politeknik Elektronika Negeri Surabaya

²Manajemen, Fakultas Ekonomi, Universitas Islam Kediri

³Teknologi Game, Departemen Teknologi Multimedia Kreatif, Politeknik Elektronika Negeri Surabaya

⁴Teknologi Rekayasa Multimedia, Departemen Teknologi Multimedia Kreatif, Politeknik Elektronika Negeri Surabaya
pradiptadody@pens.ac.id, ekowidodo@uniska-kediri.ac.id, davidfahmi354@pens.ac.id, iqbal@pens.ac.id

Abstrak

Ulasan pengguna pada platform Steam mengandung informasi penting yang dapat digunakan untuk mengetahui persepsi pemain terhadap suatu game. Namun, jumlah ulasan yang besar menyebabkan proses analisis secara manual menjadi kurang efisien. Penelitian ini bertujuan untuk melakukan klasifikasi sentimen ulasan pengguna Steam menggunakan model berbasis Deep Learning, yaitu DistilBERT. Metode penelitian yang digunakan adalah eksperimen dengan pendekatan kuantitatif dan mengadopsi framework CRISP-DM. Dataset yang digunakan terdiri dari 17.495 ulasan pengguna yang dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Proses analisis meliputi data preparation, tokenisasi menggunakan DistilBERT Tokenizer, fine-tuning model DistilBERT, serta evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model DistilBERT mampu menghasilkan nilai akurasi sebesar 90%, dengan nilai precision sebesar 90%, recall sebesar 89%, dan F1-score sebesar 90%. Hasil tersebut menunjukkan bahwa DistilBERT memiliki kemampuan yang baik dalam memahami konteks teks dan melakukan klasifikasi sentimen secara efektif. Meskipun demikian, model masih mengalami kesulitan dalam menangani ulasan yang mengandung sentimen campuran, ironi, dan bahasa informal.

Kata Kunci: Analisis Sentimen, Steam, DistilBERT, Deep Learning, Natural Language Processing.

1. Latar Belakang

Industri game digital mengalami pertumbuhan yang pesat dalam beberapa tahun terakhir seiring dengan meningkatnya perkembangan teknologi dan akses internet. Berdasarkan laporan Newzoo tahun 2024, industri game global memiliki lebih dari tiga miliar pemain dengan nilai pasar yang mencapai lebih dari 180 miliar dolar Amerika Serikat [1]. Salah satu platform distribusi game digital terbesar saat ini adalah Steam yang menyediakan puluhan ribu judul permainan dan memiliki lebih dari 132 juta pengguna aktif bulanan [2]. Selain berfungsi sebagai sarana distribusi game, Steam juga menyediakan fitur ulasan (*review*) yang memungkinkan pengguna memberikan penilaian, kritik, serta pengalaman mereka terhadap permainan yang telah dimainkan. Ulasan tersebut menjadi sumber informasi yang berharga bagi pengembang game untuk mengevaluasi kualitas produk sekaligus bagi calon pengguna dalam membantu proses pengambilan keputusan sebelum membeli suatu permainan.

Seiring dengan meningkatnya jumlah pengguna dan banyaknya ulasan yang dihasilkan setiap hari, proses analisis secara manual menjadi tidak efisien dan membutuhkan waktu yang cukup besar. Padahal, informasi yang terkandung dalam ulasan pengguna dapat mencerminkan tingkat kepuasan, pengalaman bermain, serta berbagai permasalahan yang ditemukan selama menggunakan suatu produk. Oleh karena itu, diperlukan suatu metode yang mampu mengolah dan mengidentifikasi sentimen dari data teks secara otomatis. Analisis sentimen merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengelompokkan opini pengguna berdasarkan polaritas sentimen tertentu. Seiring perkembangan teknologi kecerdasan buatan, metode analisis sentimen telah mengalami perkembangan yang signifikan, terutama dengan hadirnya model berbasis *Transformer* dan *Large Language Model* (LLM) [3].

Berbagai pendekatan telah diterapkan dalam klasifikasi sentimen, mulai dari metode *machine learning* konvensional hingga *deep learning*. Namun, metode konvensional masih memiliki keterbatasan dalam memahami hubungan semantik dan konteks antar kata. Perkembangan arsitektur *Transformer* telah membawa peningkatan performa pada berbagai tugas NLP karena kemampuannya dalam memahami konteks kalimat secara lebih baik [4]. Penelitian yang dilakukan oleh Tzimiris dkk. menunjukkan bahwa model berbasis *Transformer* mampu

menghasilkan performa klasifikasi sentimen yang lebih baik dibandingkan pendekatan sebelumnya, terutama dalam memahami hubungan kontekstual pada data teks [5]. Keunggulan tersebut menjadikan arsitektur *Transformer* sebagai salah satu pendekatan yang banyak digunakan dalam berbagai penelitian klasifikasi sentimen.

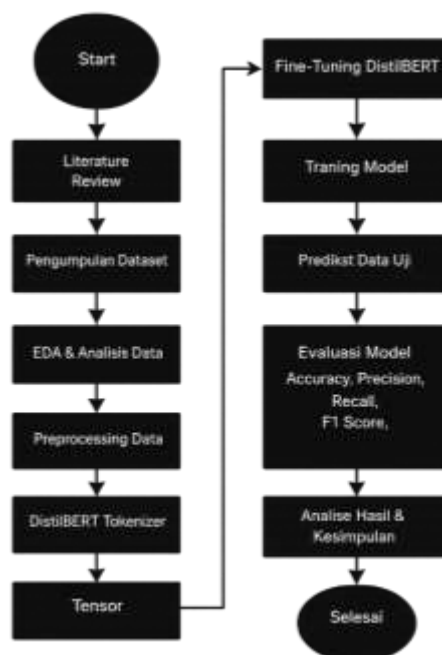
Meskipun model *Transformer* seperti BERT memiliki kemampuan yang sangat baik dalam memahami konteks bahasa, ukuran model yang besar menyebabkan kebutuhan sumber daya komputasi dan waktu inferensi yang tinggi. Oleh karena itu, berbagai penelitian mulai mengembangkan model *Transformer* yang lebih ringan namun tetap mampu mempertahankan performa yang baik. Salah satu model tersebut adalah DistilBERT yang merupakan hasil proses *knowledge distillation* dari BERT. Penelitian yang dilakukan oleh Reza dkk. menunjukkan bahwa DistilBERT mampu menghasilkan performa yang baik dalam klasifikasi sentimen ulasan *marketplace* dengan kompleksitas model yang lebih rendah [6]. Hasil serupa juga ditunjukkan oleh Aljabar dkk. yang menerapkan DistilBERT-base-uncased pada klasifikasi sentimen terkait Pemilihan Presiden Indonesia tahun 2024 dan memperoleh performa klasifikasi yang tinggi [7]. Selain itu, penelitian Gandhi dkk. menunjukkan bahwa DistilBERT memiliki keseimbangan yang baik antara akurasi dan efisiensi komputasi dibandingkan model Transformer lainnya [8].

Berbagai penelitian sebelumnya telah menunjukkan keberhasilan penggunaan arsitektur *Transformer* dan DistilBERT pada berbagai domain, seperti ulasan *marketplace* [6], media sosial [7], serta analisis sentimen berbasis aspek [4], [9]. Namun, sebagian besar penelitian tersebut masih berfokus pada domain *e-commerce*, media sosial, dan teks umum. Penelitian yang secara khusus memanfaatkan ulasan pengguna pada platform Steam masih relatif terbatas. Padahal, karakteristik ulasan game memiliki keunikan tersendiri karena sering mengandung istilah teknis, ekspresi informal, serta konteks yang berkaitan dengan pengalaman bermain. Selain itu, sebagian penelitian masih menggunakan model *Transformer* dengan kompleksitas yang tinggi sehingga memerlukan sumber daya komputasi yang besar. Oleh karena itu, diperlukan penelitian yang mengevaluasi kemampuan model DistilBERT sebagai model *Transformer* yang lebih ringan dalam melakukan klasifikasi sentimen pada ulasan pengguna platform Steam.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penggunaan arsitektur *Transformer* DistilBERT untuk melakukan klasifikasi sentimen ulasan pengguna pada platform Steam. Kinerja model akan dievaluasi untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen secara efektif. Melalui penelitian ini, diharapkan dapat diperoleh model klasifikasi sentimen yang memiliki keseimbangan antara tingkat akurasi yang tinggi dan efisiensi komputasi, sehingga dapat memberikan kontribusi terhadap penerapan *Natural Language Processing* pada ranah industri game.

2. Metode Penelitian

Penelitian ini menggunakan metode eksperimen (*Experimental Research*) dengan tujuan untuk mengevaluasi kemampuan model DistilBERT dalam mengklasifikasikan sentimen ulasan pengguna pada platform Steam. Proses penelitian meliputi pengumpulan dataset, *pre-processing* data, *split dataset*, tokenisasi menggunakan DistilBERT Tokenizer, *fine-tuning* model DistilBERT, serta evaluasi performa model.

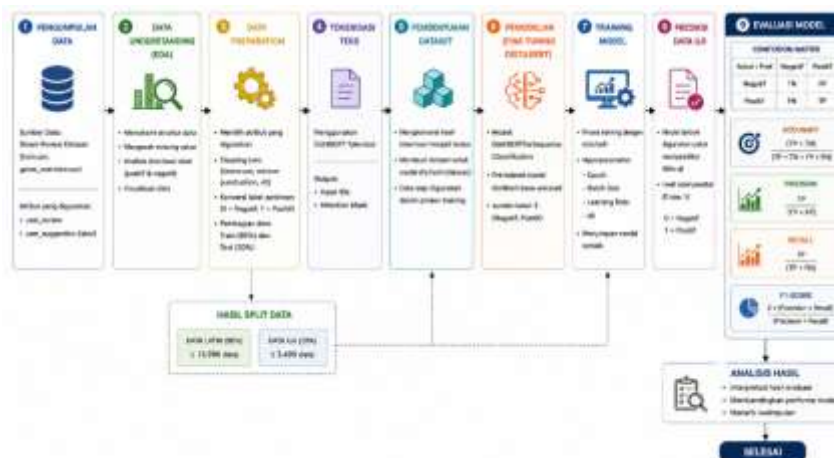


Gambar 1. Alur Penelitian

2.1. Populasi dan Sampel

Populasi pada penelitian ini adalah seluruh data ulasan pengguna pada platform Steam yang terdapat ulasan sebanyak 17.495 *review* pengguna. Teknik pengambilan sampel yang digunakan adalah *total sampling* (sampling jenuh), sehingga seluruh populasi digunakan sebagai sampel penelitian. Selanjutnya, dataset dibagi menjadi data latih sebesar 80% (13.996 data) dan data uji sebesar 20% (3.499 data) menggunakan metode *train-test split*. Pembagian data tersebut bertujuan untuk melatih model DistilBERT dan mengevaluasi performanya menggunakan metrik accuracy, precision, recall, dan F1-score.

2.2. Teknik Analisa Data



Gambar 2. Analisa Data menggunakan DistilBERT

Teknik analisis data pada penelitian ini dilakukan melalui beberapa tahapan yang dimulai dari pengumpulan data hingga evaluasi model. Proses analisis data menggunakan pendekatan *Deep Learning* berbasis arsitektur DistilBERT untuk melakukan klasifikasi sentimen terhadap ulasan pengguna pada platform Steam. Alur penelitian yang digunakan ditunjukkan pada Gambar 2.

2.2.1. Pengumpulan Data

Tahap pertama pengumpulan dataset yang diperoleh dari platform Steam, atribut utama yang digunakan dalam penelitian ini meliputi *user_review* yang berisi teks ulasan pengguna serta *user_suggestion* sebagai label kelas yang menunjukkan apakah suatu game direkomendasikan atau tidak. Label tersebut digunakan sebagai representasi sentimen positif dan negatif.

2.2.2. Exploratory Data Analysis

Tahap selanjutnya adalah memahami karakteristik data melalui proses *Exploratory Data Analysis* (EDA). Pada tahap ini dilakukan identifikasi struktur dataset, pengecekan nilai yang hilang (*missing value*), analisis distribusi kelas sentimen, serta visualisasi data. Tahapan ini bertujuan untuk memperoleh pemahaman awal mengenai karakteristik dataset sebelum dilakukan proses pemodelan. EDA membantu peneliti dalam mengidentifikasi pola dan potensi permasalahan pada data[10].

2.2.3. Preprocessing Data

Tahap *data preparation* dilakukan untuk mempersiapkan data sebelum digunakan pada proses pelatihan model. Kegiatan yang dilakukan meliputi pemilihan atribut yang digunakan, pembersihan data teks (*text cleaning*), konversi label sentimen menjadi bentuk numerik, serta pembagian dataset menjadi data latih dan data uji dengan rasio 80:20. Pembagian data ini bertujuan untuk menghasilkan model yang mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Tahapan ini merupakan bagian penting dalam meningkatkan kualitas model pembelajaran mesin[11].

Tabel 1. Proporsional Pembagian Dataset		
Jenis Data	Persentase	Jumlah Data
Data Latih	80%	13.996
Data Uji	20%	3.499
Total Data	100%	17.495

2.2.4. Tokenisasi

Data teks yang telah dipersiapkan kemudian diproses menggunakan DistilBERT Tokenizer. Proses tokenisasi bertujuan untuk mengubah teks menjadi representasi numerik yang dapat dipahami oleh model. Hasil dari proses ini berupa *input IDs* dan *attention mask* yang akan digunakan sebagai masukan bagi model DistilBERT.

2.2.5. Pembentukan Dataset

Hasil tokenisasi kemudian dikonversi ke dalam bentuk tensor menggunakan framework PyTorch. Selanjutnya dibentuk objek dataset yang berisi pasangan antara data masukan dan label sentimen. Dataset tersebut digunakan pada proses pelatihan dan pengujian model. Penggunaan tensor memungkinkan proses komputasi yang lebih efisien dalam pembelajaran mendalam[12].

2.2.6. Pemodelan dengan DistilBERT

Tahap pemodelan dilakukan dengan menggunakan arsitektur DistilBERT yang merupakan model *pre-trained transformer* hasil distilasi dari BERT. Model yang digunakan adalah *DistilBERTForSequenceClassification* dengan dua kelas keluaran, yaitu sentimen negatif dan sentimen positif. Pada penelitian ini dilakukan proses *fine-tuning* terhadap model agar dapat menyesuaikan dengan karakteristik data ulasan pengguna Steam. Model transformer seperti DistilBERT telah banyak digunakan dalam analisis sentimen karena kemampuannya memahami konteks bahasa secara mendalam [13].

2.2.7. Training Model

Pada tahap ini model dilatih menggunakan data latih. Proses pelatihan dilakukan dengan menentukan beberapa parameter seperti jumlah epoch, ukuran batch (*batch size*), dan *learning rate*. Selama proses pelatihan, model mempelajari pola-pola yang terdapat pada data sehingga mampu membedakan antara ulasan dengan sentimen positif dan negatif. Proses pelatihan yang optimal sangat berpengaruh terhadap performa model[14].

2.2.8. Prediksi Data Uji

Setelah proses pelatihan selesai, model terbaik digunakan untuk melakukan prediksi terhadap data uji. Hasil prediksi berupa kelas sentimen yang terdiri atas kelas negatif (0) dan kelas positif (1). Tahapan ini digunakan untuk mengetahui kemampuan model dalam melakukan klasifikasi terhadap data yang belum pernah digunakan selama proses pelatihan. Evaluasi pada data uji penting untuk mengukur kemampuan generalisasi model[15].

2.2.9. Evaluasi Model

Tahap terakhir adalah evaluasi performa model. Evaluasi dilakukan menggunakan *confusion matrix* dan beberapa metrik pengukuran, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai-nilai tersebut digunakan untuk mengukur tingkat keberhasilan model dalam mengklasifikasikan sentimen ulasan pengguna Steam. Hasil evaluasi selanjutnya dianalisis untuk mengetahui efektivitas model DistilBERT dalam melakukan analisis sentimen. Penggunaan metrik evaluasi yang tepat sangat penting dalam menilai kinerja model klasifikasi[16].

3. Hasil dan Diskusi

Hasil implementasi model DistilBERT dalam melakukan klasifikasi sentimen terhadap ulasan pengguna pada platform Steam. Pembahasan dimulai dari deskripsi dataset, tahapan persiapan data, implementasi model, hingga evaluasi performa model menggunakan beberapa metrik pengukuran. Selain itu, dilakukan analisis terhadap hasil klasifikasi yang diperoleh untuk mengetahui kemampuan model dalam mengidentifikasi sentimen positif dan negatif.

3.1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan kumpulan ulasan pengguna Steam yang telah memiliki label sentimen berupa positif dan negatif. Data ini saya gunakan sebagai dasar dalam melatih model klasifikasi sentimen. Dataset terdiri dari beberapa atribut utama, yaitu teks ulasan (*review*) dan label sentimen. Sebelum dilakukan pelatihan model, sebelumnya dilakukan *pre-processing* data tujuannya mengecek distribusi kelas, serta memastikan tidak terdapat data kosong atau duplikat yang dapat memengaruhi hasil pelatihan. Setelah data siap, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Pembagian ini dilakukan untuk memastikan model dapat diuji menggunakan data yang belum pernah dilihat sebelumnya sehingga hasil evaluasi lebih objektif.

Tabel 2. Data Penelitian

review_id	title	year	user_review	User_suggestion
12198	PlanetSide 2	2018	You... You don't know what it's like... You couldn't possibly understand... THIS. IS. WAR. If you've got the grit to bare your teeth down and claw your way through the ranks of chaos, hellfire and brimstone... then you might find your place here among the squadrons. By night, I am a heavy aircraft engineer for the United States Air Force. By day, I am a heavy aircraft engineer for the Terran Republic. Supporting the front line is my life. Join the assault. It'll slap hair on your chest.	1
4451	Team Fortress 2	2015	As Forrest Gump would say: "This game is like a box of chocolates, You never know what you're going to get."I would agree.Also the community is probably my favorite part about this game....Besides the annoying FTP 10 year old squeekers.9.5/10	1
15784	theHunter Classic	2014	It is not Free 2 Play. Me and my brother downloaded this to try and play together, and it would not allow us to play multiplayer without upgrading to a subscription. Nice job losing two potential customers because of a pay wall.	0

3.2. Data Preparation

Tahap persiapan data dilakukan untuk menghasilkan dataset yang siap digunakan dalam proses pelatihan model. Proses diawali dengan pemilihan atribut yang relevan, yaitu kolom teks ulasan sebagai masukan (input) dan label sentimen sebagai target klasifikasi. Selanjutnya, dilakukan pemeriksaan terhadap dataset untuk memastikan tidak terdapat nilai kosong atau data yang tidak lengkap sehingga kualitas data tetap terjaga. Setelah proses pengecekan selesai, dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20 untuk mendukung proses pelatihan dan evaluasi model.

Setelah data siap digunakan kemudian dilakukan tokenisasi menggunakan DistilBERT Tokenizer. Pada proses ini, setiap teks ulasan diubah menjadi representasi numerik berupa token ID yang dapat diproses oleh model DistilBERT. Selain itu, diterapkan teknik *padding* dan *truncation* untuk menyeragamkan panjang setiap input sesuai dengan batas maksimum yang ditentukan. Hasil dari tahapan ini adalah data dalam bentuk token ID, *attention mask*, serta pembagian data latih dan data uji yang telah terstruktur dan siap digunakan pada proses pelatihan model DistilBERT.

[illegible]

Pada tahap pelatihan model, penelitian ini memanfaatkan DistilBERT yang telah tersedia dalam bentuk *pre-trained model* dan kemudian dilakukan proses *fine-tuning* untuk tugas klasifikasi sentimen. Pemilihan DistilBERT didasarkan pada kemampuannya dalam mempertahankan performa yang kompetitif dengan jumlah parameter yang lebih sedikit dibandingkan BERT, sehingga proses pelatihan dapat dilakukan secara lebih efisien tanpa mengorbankan akurasi secara signifikan.

Parameter	Value
Pre-trained Model	distilbert-base-uncased
Epoch	3
Training Batch Size	16
Evaluation Batch Size	16
Learning Rate	0.00005
Maximum Sequence Length	512
Number of Labels	2
Optimizer	AdamW

3.4. Hasil Pelatihan Model

Step	Epoch	Learning Rate	Training Loss
500	0,57	$4,05 \times 10^{-5}$	0,3746
1000	1,14	$3,09 \times 10^{-5}$	0,2623
1500	1.71	2.14×10^{-5}	0.1785

Step	Epoch	Learning Rate	Training Loss
2000	2,29	$1,19 \times 10^{-5}$	0,1201
2500	2,86	$2,40 \times 10^{-6}$	0,0721
Akhir Pelatihan	3,00		0,1945

Berdasarkan hasil pelatihan tersebut, nilai *training loss* menunjukkan penurunan bertahap seiring bertambahnya epoch, dari 0,3746 pada step ke-500 menjadi 0,2623 (step ke-1000), 0,1785 (step ke-1500), 0,1201 (step ke-2000), hingga 0,0721 pada step ke-2500, yang mengindikasikan model semakin mampu mengenali pola sentimen dan mendekati kondisi konvergen, penurunan *learning rate* dari $4,05 \times 10^{-5}$ menjadi $2,40 \times 10^{-6}$ turut membantu stabilitas optimasi. Secara keseluruhan, pelatihan berlangsung selama 2242,75 detik ($\pm 37,38$ menit) dengan kecepatan 18,72 sampel/detik dan 1,17 langkah/detik, menghasilkan *training loss* akhir 0,1945 serta akurasi 90%, yang menunjukkan model telah belajar dengan baik tanpa indikasi *underfitting* dan memiliki kemampuan generalisasi yang baik.

3.5. Evaluasi Model

Model DistilBERT yang telah dilatih dievaluasi menggunakan data uji untuk mengetahui kemampuannya dalam melakukan klasifikasi sentimen ulasan pengguna Steam. Evaluasi dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, dilakukan analisis menggunakan *confusion matrix* untuk mengetahui jumlah prediksi yang benar maupun kesalahan klasifikasi yang dilakukan oleh model.

3.5.1. Hasil Classification Report

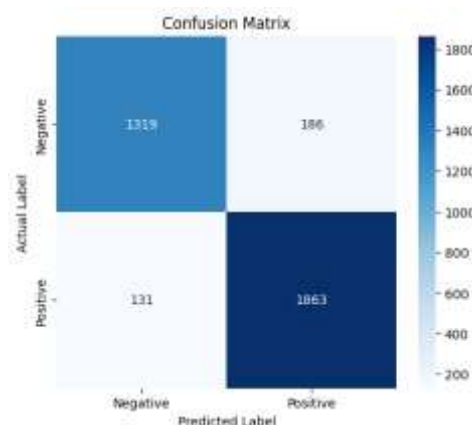
Tabel 6. Hasil Evaluasi Model

Kelas	Precision	Recall	F1-Score	Support
Negatif (0)	0,90	0,85	0,88	1505
Positif (1)	0,89	0,93	0,91	1994
Accuracy	-	-	0,90	3499
Macro Average	0,90	0,89	0,90	3499
Weighted Average	0,90	0,90	0,90	3499

Berdasarkan hasil *classification report*, model DistilBERT menunjukkan performa yang baik dalam melakukan klasifikasi sentimen ulasan pengguna Steam pada 3.499 data uji yang terdiri atas kelas sentimen negatif dan positif. Pada kelas sentimen negatif (kelas 0), model memperoleh nilai *precision* sebesar 0,90, *recall* sebesar 0,85, dan *F1-score* sebesar 0,88. Nilai *precision* sebesar 90% menunjukkan bahwa sebagian besar prediksi yang dikategorikan sebagai sentimen negatif merupakan prediksi yang benar, sedangkan nilai *recall* sebesar 85% mengindikasikan bahwa masih terdapat beberapa data sentimen negatif yang tidak berhasil dikenali dan diprediksi sebagai sentimen positif. Sementara itu, pada kelas sentimen positif (kelas 1), model memperoleh nilai *precision* sebesar 0,89, *recall* sebesar 0,93, dan *F1-score* sebesar 0,91. Tingginya nilai *recall* pada kelas positif menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali ulasan dengan sentimen positif sehingga sebagian besar data pada kelas tersebut berhasil diklasifikasikan dengan benar.

Secara keseluruhan, model menghasilkan nilai *accuracy* sebesar 0,90 atau 90%, yang menunjukkan bahwa sekitar 90% dari seluruh data uji berhasil diklasifikasikan dengan benar. Nilai *macro average* sebesar 0,90 pada *precision*, 0,89 pada *recall*, dan 0,90 pada *F1-score* menunjukkan bahwa performa model pada kedua kelas relatif seimbang. Selain itu, nilai *weighted average* yang juga mencapai 0,90 mengindikasikan bahwa model memiliki kemampuan klasifikasi yang konsisten terhadap keseluruhan data uji. Hasil tersebut menunjukkan bahwa model DistilBERT mampu memahami konteks teks dengan baik sehingga dapat membedakan sentimen positif dan negatif secara efektif. Meskipun demikian, nilai *recall* pada kelas negatif yang lebih rendah dibandingkan kelas positif menunjukkan bahwa masih terdapat beberapa ulasan negatif yang memiliki karakteristik serupa dengan ulasan positif sehingga menyebabkan terjadinya kesalahan klasifikasi.

3.5.2. Confusion Matrix



Gambar 3. Hasil *Confusion Matrix*

Berdasarkan hasil *confusion matrix*, model DistilBERT mampu mengklasifikasikan data uji dengan benar. Dari 1.505 data dengan sentimen negatif, sebanyak 1.319 data berhasil diprediksi dengan benar sebagai sentimen negatif (*True Negative*), sedangkan 186 data lainnya salah diklasifikasikan sebagai sentimen positif (*False Positive*). Sementara itu, dari 1.994 data dengan sentimen positif, sebanyak 1.863 data berhasil diprediksi dengan benar sebagai sentimen positif (*True Positive*), sedangkan 131 data salah diklasifikasikan sebagai sentimen negatif (*False Negative*). Dominasi nilai pada diagonal utama *confusion matrix* menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan kedua kelas sentimen dan mampu menghasilkan prediksi yang akurat pada sebagian besar data uji.

Hasil tersebut menunjukkan bahwa model DistilBERT memiliki kemampuan klasifikasi yang cukup baik terhadap sentimen positif maupun negatif dengan tingkat kesalahan yang relatif rendah. Jumlah *True Positive* yang lebih tinggi dibandingkan *False Negative* menunjukkan bahwa model sangat baik dalam mengenali ulasan dengan sentimen positif. Demikian pula, jumlah *True Negative* yang jauh lebih besar dibandingkan *False Positive* mengindikasikan bahwa sebagian besar ulasan negatif berhasil diidentifikasi dengan benar. Meskipun masih terdapat kesalahan klasifikasi sebanyak 186 data negatif yang diprediksi sebagai positif dan 131 data positif yang diprediksi sebagai negatif, secara keseluruhan hasil *confusion matrix* memperlihatkan bahwa model memiliki kemampuan generalisasi yang baik dan mampu mempertahankan performa klasifikasi yang tinggi, yang diperkuat dengan nilai akurasi sebesar 90% pada tahap evaluasi.

Tabel 7. Tabel *Misclassified Reviews*

Review	Actual	Predicted
Early Access ReviewSeems like it could be a really good game, just not right now. But that's how most early access games are. If this ever makes it out of early access I'll give it a try again and update this review.	0	1
Alot of people on here saying this is a pay to win game, but i disagree. Ive played this game a long time and dont really feel the need to constantly spend money to get the best gear in the game, I play the game like many other people for the enjoymentand to help friends.Having the top gear in the game would be nice but it definately isnt needed to get enjoyment out this game.All the people saying it is pay to win obviously dont realise that the people putting alot of money into the game actually helps you if you dont want to put any money into the game. Its keeping the game alive, Its allowing you to trade ingame currency (AD) to the premium currency (zen).Just because you cant or wont put alot of time into the game doesnt mean its pay to win, and you cant expect everything to be handed to you on a plate you have to earn it.The main problem i have with the game is the constant requirement to grind out some of the content like with the new Sea of moving ice mod they added recently, having to fish constantly seems like they picked one of the most boring less liked tasks within the game, changed it so it was harder to catch the fish and made it so it was mandatory to do, i think that was a big mistake,	1	0
I know I'm only .8 hours here but I'm well over 100 on PS4.Just wanted to comment about how I'm sure the new tower has helped with the flow of the game, but I don't like what it's done for the feel of the game or the farm for cosmetics. I liked being able to drop into a power level range and see all sorts of different types of biomes instead of just having the biomes correlate to certain power level ranges like they do now. It's a lot more fun if I can grind stuff that I want to grind that's relevant to my level while mostly staying in biome types I want to play/need cosmetics from.If I'm missing something, please someone let me know.	0	1

Berdasarkan hasil analisis terhadap data yang mengalami kesalahan klasifikasi, ditemukan bahwa model DistilBERT masih mengalami kesulitan dalam memahami ulasan yang memiliki sentimen campuran (*mixed sentiment*), ulasan yang panjang, serta penggunaan bahasa yang ambigu. Sebagai contoh, terdapat ulasan seperti:

DOI: <https://doi.org/10.31004/riggs.v5i2.11000>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

"*Im torn with this game. In concept the game is...*"

Review diatas memiliki label aktual positif, namun diprediksi negatif oleh model. Kalimat tersebut mengandung opini yang beragam, di mana penulis menyampaikan kritik terhadap beberapa aspek permainan, tetapi secara keseluruhan masih memberikan rekomendasi positif. DistilBERT cenderung menangkap kata-kata bernada negatif yang muncul di awal kalimat sehingga menghasilkan prediksi yang tidak sesuai dengan label sebenarnya.

Fenomena serupa juga ditemukan pada ulasan seperti "*Great game. It always has...*" yang memiliki label positif tetapi diprediksi negatif. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam memahami konteks secara keseluruhan pada ulasan yang panjang dan mengandung kombinasi sentimen positif dan negatif. Hal ini menunjukkan bahwa DistilBERT masih memiliki keterbatasan dalam memahami ekspresi emosional yang sangat kuat, penggunaan huruf kapital, serta hubungan konteks secara menyeluruh. Selain itu, terdapat beberapa ulasan yang ditulis dalam bahasa selain bahasa Inggris, seperti bahasa Turki, yang menyebabkan model mengalami kesulitan dalam memahami makna kalimat karena model Distilbert-base-uncased dilatih terutama menggunakan korpus bahasa Inggris.

Tabel 8. Tabel Ringkasan Kelemahan DistilBERT

Kelemahan	Contoh Kasus	Dampak
Sentimen campuran	Kritik + pujian dalam satu ulasan	Salah menentukan polaritas
Ulasan panjang	Konteks berubah di akhir kalimat	Fokus pada sebagian kalimat
Ironi atau sarkasme	"Great game, another bug..."	Salah mengartikan sentimen
Bahasa informal/slang	Singkatan dan kata tidak baku	Pemahaman konteks menurun
Huruf kapital dan ekspresi emosi	"GOD THIS GAME IS AWFUL!"	Tidak sepenuhnya menangkap intensitas emosi
Multibahasa	Bahasa Turki, campuran bahasa	Tokenisasi kurang optimal
Label rekomendasi $\neq$ isi review	Isi negatif tetapi tetap merekomendasikan game	Menimbulkan ambiguitas pada target klasifikasi

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, model DistilBERT mampu melakukan klasifikasi sentimen terhadap ulasan pengguna Steam dengan memanfaatkan mekanisme *self-attention* untuk memahami hubungan kontekstual antar kata dalam suatu kalimat. Proses klasifikasi diawali dengan tokenisasi teks menggunakan DistilBERT Tokenizer yang menghasilkan representasi numerik berupa *input IDs* dan *attention mask*. Selanjutnya, model DistilBERT yang telah melalui proses *fine-tuning* mempelajari pola sentimen dari data latih dan menghasilkan prediksi berupa sentimen positif atau negatif. Berdasarkan hasil evaluasi terhadap 3.499 data uji, model berhasil memperoleh nilai akurasi sebesar 90%, dengan nilai *precision*, *recall*, dan *F1-score* yang menunjukkan kemampuan klasifikasi yang baik terhadap kedua kelas sentimen. Penggunaan DistilBERT pada penelitian ini memiliki beberapa kelebihan, yaitu mampu memahami konteks kalimat dengan baik, memiliki ukuran model yang lebih ringan dibandingkan BERT, waktu pelatihan yang relatif lebih singkat, serta menghasilkan performa klasifikasi yang tinggi. Selain itu, model mampu melakukan generalisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya, yang ditunjukkan melalui hasil *confusion matrix* dan nilai evaluasi yang seimbang antara kedua kelas. Namun demikian, model masih memiliki beberapa keterbatasan, terutama dalam mengklasifikasikan ulasan yang mengandung sentimen campuran, ironi atau sarkasme, bahasa informal, serta ulasan yang menggunakan bahasa selain bahasa Inggris. Selain itu, terdapat beberapa kasus di mana isi ulasan memiliki kecenderungan negatif, tetapi pengguna tetap memberikan rekomendasi positif, sehingga menimbulkan ambiguitas pada label yang digunakan.

Referensi

- [1] Newzoo, "Newzoo's Global Games Market Report 2024," Newzoo.
- [2] Valve Corporation, "Steam Year in Review 2024," Valve Corporation.
- [3] A. Amrullah, "Sentiment Analysis in the Age of Transformers and Large Language Models: A Comprehensive Review of Recent Advances and Future," 2025. [Online]. Available: <https://doi.org/xxxxxxx>
- [4] I. Perikos and A. Diamantopoulos, "Explainable Aspect-Based Sentiment Analysis Using Transformer Models," *Big Data and Cognitive Computing*, vol. 8, no. 11, p. 141, Oct. 2024, doi: 10.3390/bdcc8110141.
- [5] S. Tzimiris, S. Nikiforos, M. N. Nikiforos, D. Mouratidis, and K. L. Kermanidis, "A Comparative Evaluation of Transformer-Based Language Models for Topic-Based Sentiment Analysis," *Electronics (Basel)*, vol. 14, no. 15, p. 2957, Jul. 2025, doi: 10.3390/electronics14152957.
- [6] R. F. Reza, Muhammad Thoriq, and Rd. Imam Saepul Millah, "Sentiment Analysis of Marketplace Review with Islamic Perspective using Fine-Tuning DistilBERT," *Khazanah Journal of Religion and Technology*, vol. 2, no. 2, pp. 45–54, Jan. 2025, doi: 10.15575/kjrt.v2i2.1118.

- [7] A. Aljabar, B. M. Karomah, N. Tarisafitri, and J. Jeffry, "Sentiment Analysis in Indonesian's Presidential Election 2024 Using Transformer (Distilbert-Base-Uncased)," *Journal of System and Computer Engineering (JSCE)*, vol. 6, no. 2, pp. 197–203, Apr. 2025, doi: 10.61628/jsce.v6i2.1867.
- [8] J. N. Gandhi, K. V. Guru, A. R. Kannan, R. A. Sudhan, S. A. Kumar, and M. Bharathvaj, "Efficient Sentiment Classification using DistilBERT for Enhanced NLP Performance," 2025, pp. 1500–1511. doi: 10.2991/978-94-6463-718-2_125.
- [9] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Applied Sciences*, vol. 15, no. 1, p. 172, Dec. 2024, doi: 10.3390/app15010172.
- [10] Wes McKinney, *Python for Data Analysis*, 3rd ed. O'Reilly Media, 2022.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. 2016.
- [12] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2019.
- [13] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [14] S. Ruder, "An overview of gradient descent optimization algorithms," Jun. 2017, [Online]. Available: <http://arxiv.org/abs/1609.04747>
- [15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [16] D. M. W. Powers and Ailab, "EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION."