PERANCANGAN SISTEM INFORMASI ANALITIK KLASIFIKASI KELUCUAN STAND-UP COMEDY MENGGUNAKAN INDOBERT

DESIGN OF AN INFORMATION ANALYTICS SYSTEM FOR STAND-UP COMEDY HUMOR CLASSIFICATION USING INDOBERT

Jihan Najib[1], Supriyono[2*], Okta Qomaruddin Aziz[3]

[*]E-mail: priyono@ti.uin-malang.ac.id

[1,2,3]Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang

Abstrak

Stand-up comedy merupakan salah satu bentuk hiburan yang berkembang pesat di Indonesia dan banyak dipublikasikan melalui platform digital. Namun, penilaian tingkat kelucuan dalam pertunjukan *stand-up comedy* masih bersifat subjektif dan belum didukung oleh sistem analitik berbasis data. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem informasi analitik yang mampu mengklasifikasikan tingkat kelucuan *stand-up comedy* Indonesia berdasarkan transkrip menggunakan pendekatan *Natural Language Processing* (NLP) berbasis IndoBERT. Dataset penelitian diperoleh dari transkrip pertunjukan *stand-up comedy* berbahasa Indonesia yang kemudian dilakukan proses pelabelan tingkat kelucuan ke dalam beberapa kategori. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, *fine-tuning* model IndoBERT untuk klasifikasi multi-kelas, serta pengembangan sistem informasi berbasis web yang menyediakan fitur unggah transkrip, prediksi tingkat kelucuan, dan visualisasi hasil analisis. Evaluasi performa model dilakukan menggunakan metrik *accuracy, precision, recall,* dan *F1-score*. Hasil penelitian menunjukkan bahwa model IndoBERT mampu memberikan performa klasifikasi yang baik dalam mengenali pola humor pada teks *stand-up comedy*. Sistem informasi yang dikembangkan dapat membantu pengguna dalam menganalisis materi *stand-up comedy* secara otomatis serta menyediakan dukungan analitik dalam evaluasi konten. Penelitian ini berkontribusi dalam pengembangan sistem informasi berbasis kecerdasan buatan untuk analisis humor dan membuka peluang penerapan lebih lanjut pada bidang evaluasi konten hiburan digital.

Kata kunci*: *Stand-up comedy, IndoBERT, klasifikasi teks, sistem informasi analitik, humor detection, NLP.

Abstract

Stand-up comedy is a form of entertainment that has rapidly grown in Indonesia and is widely distributed through digital platforms. However, evaluating the level of humor in stand-up comedy performances remains subjective and is not yet supported by data-driven analytical systems. This study aims to design and implement an information analytics system capable of classifying the humor level of Indonesian stand-up comedy based on transcript data using a Natural Language Processing (NLP) approach powered by IndoBERT. The dataset was collected from Indonesian stand-up comedy performance transcripts, followed by a labeling process to categorize humor levels into several classes. The research stages include data collection, text preprocessing, fine-tuning the IndoBERT model for multi-class classification, and developing a web-based information system that provides transcript upload features, humor level prediction, and result visualization. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The results indicate that IndoBERT achieves strong classification performance in identifying humor-related patterns within stand-up comedy transcripts. The developed

information system enables users to automatically analyze stand-up comedy materials and provides analytical support for content evaluation. This research contributes to the development of artificial intelligence-based information systems for humor analytics and offers potential for further applications in digital entertainment content assessment.

***Keywords**: Stand-up comedy, IndoBERT, text classification, information analytics system, humor detection, NLP.*

1. PENDAHULUAN

Pertumbuhan konten hiburan digital di Indonesia dalam satu dekade terakhir telah menghasilkan ekosistem yang sangat beragam, salah satunya adalah stand-up comedy. Kompetisi tahunan Stand Up Comedy Indonesia (SUCI) yang ditayangkan Kompas TV menjadi katalisator berkembangnya komunitas komika lokal dan menghasilkan ribuan jam rekaman pertunjukan yang kini dapat diakses secara bebas melalui platform daring. Fenomena ini membuka peluang besar bagi riset berbasis data untuk memahami pola humor dalam bahasa Indonesia [1].

Selama ini, penilaian tingkat kelucuan sebuah pertunjukan masih bertumpu pada reaksi langsung penonton, ulasan subjektif kritikus seni, atau intuisi personal penilai. Ketergantungan terhadap penilaian manusia semacam ini tidak hanya tidak skalabel, tetapi juga mengandung bias yang sulit dikontrol. Riset-riset terdahulu di bidang humor detection secara umum menunjukkan bahwa humor sangat bergantung pada konteks budaya, referensi linguistik, dan nada tuturan, sehingga pendekatan komputasional yang kaku seringkali gagal menangkap kompleksitasnya [2][3][4].

Perkembangan pesat di bidang Natural Language Processing (NLP), khususnya setelah diperkenalkannya mekanisme attention oleh Vaswani dkk. [5] dan arsitektur BERT oleh Devlin dkk. [6], membuka kemungkinan baru dalam pemodelan teks secara kontekstual dan bidireksional. Untuk ranah bahasa Indonesia, Koto dkk. [9] mengembangkan IndoBERT menggunakan korpus berbahasa Indonesia berskala besar dan menunjukkan performa yang baik pada berbagai tugas NLP, termasuk Benchmark evaluasi bahasa Indonesia seperti IndoNLU [10] dan IndoNLI [11] turut memperkuat ekosistem riset NLP lokal.

Penelitian terdahulu tentang deteksi humor dilakukan oleh Mihalcea dan Strapparava [3] yang memanfaatkan pendekatan statistik berbasis one-liners, serta Bertero dan Fung [4] yang menggunakan Long Short-Term Memory (LSTM) [4] untuk memprediksi humor dalam dialog. Weller dan Seppi [12] selanjutnya mendemonstrasikan bahwa model berbasis Transformer memberikan keunggulan signifikan dibandingkan pendekatan sekuensial dalam tugas humor detection. Di sisi lain, penelitian berbasis sentimen pada konten komedi Indonesia yang dilakukan oleh Suciati dan Prihatmanto [1] serta Rahman dan Surjanto [13] menunjukkan potensi NLP untuk memahami nuansa bahasa informal dalam konten hiburan lokal.

Berdasarkan gap yang teridentifikasi, penelitian ini bertujuan merancang dan mengimplementasikan sistem informasi analitik terpadu yang mampu mengklasifikasikan tingkat kelucuan pertunjukan stand-up comedy Indonesia secara otomatis. Kontribusi utama penelitian ini meliputi: (1) penyusunan dataset transkripsi stand-up comedy Indonesia berlabel kelucuan sebanyak 13.775 segmen dari SUCI season 1–8; (2) fine-tuning model IndoBERT [9] untuk klasifikasi humor multi-kelas; dan (3) pengembangan sistem informasi analitik berbasis web yang mengintegrasikan pipeline prediksi dan visualisasi hasil secara interaktif.

2. METODOLOGI

Pendekatan yang diterapkan dalam penelitian ini mengadopsi siklus pengembangan sistem yang terdiri atas lima tahap berurutan: akuisisi dan preparasi data, pra-pemrosesan teks, konstruksi model klasifikasi, evaluasi performa, serta integrasi ke dalam sistem informasi analitik

berbasis web. Setiap tahap dirancang agar dapat direplikasi dan dikembangkan lebih lanjut oleh peneliti lain.

2.1 Akuisisi dan Preparasi Dataset

Sumber data primer penelitian ini adalah rekaman video pertunjukan Stand Up Comedy Indonesia (SUCI) season 1 sampai season 8 yang diunggah di YouTube. Transkripsi diperoleh dengan memanfaatkan fitur subtitle otomatis yang kemudian diverifikasi secara manual. Setiap segmen transkripsi dipasangkan dengan penanda jumlah tawa penonton (Number of Laughs) yang diekstraksi dari rekaman audio sebagai label ground truth, mengadopsi pendekatan anotasi berbasis respons audiens yang diusulkan dalam riset humor detection sebelumnya [2][4][14]. Total dataset yang terkumpul mencakup 13.775 segmen dari 28 sumber video yang berbeda. Data dibagi menjadi set latih (80%) dan set uji (20%) secara acak berlapis (stratified) untuk memastikan proporsi kelas yang seimbang pada setiap partisi.

Tabel 1. Distribusi Label Kelucuan pada Dataset

Label	Rentang Tawa	Jumlah Segmen	Presentase (%)
Tidak Lucu	0	5.623	40,82
Cukup Lucu	1 – 4	6.666	48,39
Lucu	5 – 9	977	7,09
Sangat Lucu	≥ 10	509	3,70
Total	-	13.775	100,00

Skema pelabelan empat kategori pada Tabel 1 ditetapkan berdasarkan distribusi empiris jumlah tawa dan mempertimbangkan aspek interpretabilitas psikologis dari reaksi audiens. Kategori Tidak Lucu (0 tawa) mencerminkan segmen naratif atau transisi, Cukup Lucu (1–4 tawa) merepresentasikan humor ringan, Lucu (5–9 tawa) menandai materi yang berhasil membangkitkan respons audiens secara konsisten, dan Sangat Lucu (≥10 tawa) mengindikasikan puncak komedi dalam penampilan.

2.2 Pra-pemrosesan Teks

Kualitas representasi teks merupakan faktor kritis dalam performa model NLP [15][19]. Tahap pra-pemrosesan pada penelitian ini dirancang secara khusus untuk menangani karakteristik unik transkripsi pertunjukan stand-up comedy yang kaya akan elemen non-tekstual. Langkah-langkah yang diterapkan adalah: (1) penghapusan penanda waktu (timestamp) dan anotasi audio seperti [TERTAWA], [Tepuk Tangan], dan [Musik]; (2) normalisasi ejaan ke huruf kecil; (3) penghapusan karakter berulang dan simbol yang tidak bermakna secara linguistik; (4) tokenisasi menggunakan tokenizer bawaan IndoBERT yang mendukung subword tokenization berbasis WordPiece [3]; dan (5) padding atau truncation segmen ke panjang maksimum 512 token sesuai batas arsitektur BERT [6][16].

2.3 Arsitektur Model dan Proses Fine-Tuning

Fondasi model yang digunakan adalah IndoBERT [9], sebuah implementasi arsitektur BERT [6] yang mengadopsi mekanisme multi-head self-attention [5] dan dilatih secara khusus pada korpus berbahasa Indonesia. Dibandingkan dengan pendekatan word embedding statis seperti Word2Vec [17] atau GloVe [18], IndoBERT menghasilkan representasi yang bersifat kontekstual dan dinamis, sejalan dengan keunggulan model berbasis Transformer [5] atas

arsitektur sekuensial seperti LSTM [40] dan GRU [19][20]. Representasi kontekstual semacam ini juga lebih unggul dibandingkan model ELMo [15] yang menggunakan representasi biLM berbasis karakter.

Lapisan klasifikasi ditambahkan di atas representasi vektor [CLS] dari keluaran IndoBERT, mengikuti konvensi standar fine-tuning BERT untuk klasifikasi urutan tunggal [6]. Hiperparameter pelatihan yang digunakan: learning rate 2×10^{-5} dengan scheduler linear warmup [21], batch size 32, jumlah epoch 5, dan bobot regularisasi L2 (weight decay) 0,01 via optimizer AdamW [21]. Dropout [17] dengan nilai 0,1 diterapkan pada lapisan klasifikasi untuk mengurangi overfitting. Fungsi loss yang digunakan adalah weighted cross-entropy untuk mengatasi ketidakseimbangan kelas yang signifikan [22]. Seluruh eksperimen diimplementasikan menggunakan PyTorch [23] dengan dukungan library Transformers dari HuggingFace [16].

2.4 Evaluasi Model

Performa model dievaluasi menggunakan empat metrik klasifikasi standar yang umum digunakan dalam riset NLP [24][25]: (1) Accuracy, yaitu rasio keseluruhan prediksi yang tepat terhadap total sampel; (2) Precision, yang mengukur proporsi prediksi positif yang memang benar positif; (3) Recall, yang mengukur kemampuan model mendeteksi seluruh sampel positif yang ada; dan (4) F1-score, rata-rata harmonik antara precision dan recall yang lebih informatif pada kondisi distribusi kelas tidak seimbang [2][14]. Keempat metrik dihitung dengan pendekatan macro-average agar setiap kelas mendapatkan bobot setara tanpa terpengaruh oleh ukuran kelas.

2.5 Pengembangan Sistem Informasi Analitik

Sistem informasi analitik dikembangkan menggunakan arsitektur berbasis web dengan Python Flask sebagai backend RESTful API dan antarmuka pengguna berbasis HTML5/CSS3/JavaScript. Model IndoBERT yang telah selesai dilatih diekspor dan dimuat ulang menggunakan pustaka Transformers [16] untuk inferensi pada lingkungan produksi. Sistem menyediakan empat fitur utama: (1) pengunggahan transkripsi dalam format teks maupun PDF; (2) prediksi tingkat kelucuan per segmen secara real-time; (3) visualisasi distribusi hasil prediksi menggunakan library charting interaktif; dan (4) ekspor laporan analisis. Visualisasi data pada sistem memanfaatkan prinsip-prinsip yang dikembangkan dalam riset sentimen dan analitik teks [1][13][26][27], sehingga hasil analisis dapat diinterpretasikan dengan intuitif oleh pengguna non-teknis.

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan temuan dari tiga aspek utama penelitian: karakteristik statistik dataset, performa model IndoBERT yang dilatih, serta implikasi dari temuan tersebut dalam konteks literatur yang relevan.

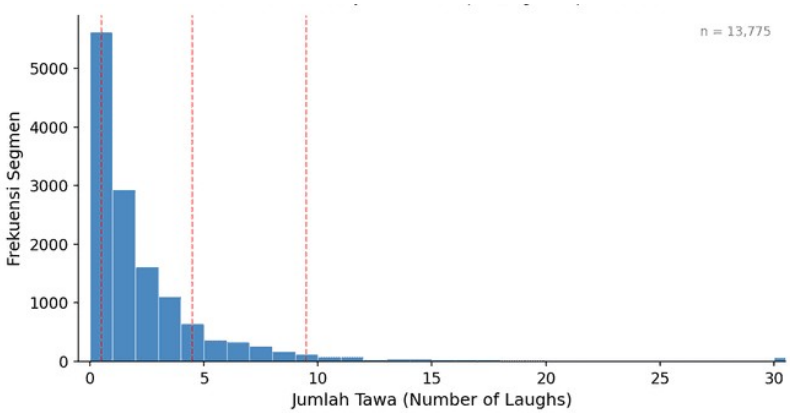
3.1 Karakteristik Dataset

Analisis statistik deskriptif terhadap 13.775 segmen dataset mengungkap distribusi jumlah tawa yang sangat condong (right-skewed) dengan nilai rata-rata 2,24 (standar deviasi 4,38), median 1, dan nilai maksimum 70. Kondisi ini mencerminkan realitas pertunjukan stand-up comedy di mana momen tawa besar (punchline) hanya terjadi sesekali di antara panjangnya narasi pengantar (setup). Pola distribusi ini konsisten dengan observasi dalam riset humor detection pada data pertunjukan komedi serupa [2][4][14].

Ketidakseimbangan antar kelas yang terlihat pada Tabel 1 merupakan tantangan metodologis yang lazim ditemui dalam dataset humor berlabel alami [12][3]. Kelas Cukup Lucu mendominasi dengan 48,39% sampel, diikuti kelas Tidak Lucu (40,82%), sementara kelas Lucu (7,09%) dan Sangat Lucu (3,70%) bersama-sama hanya membentuk sekitar sepersepuluh dari

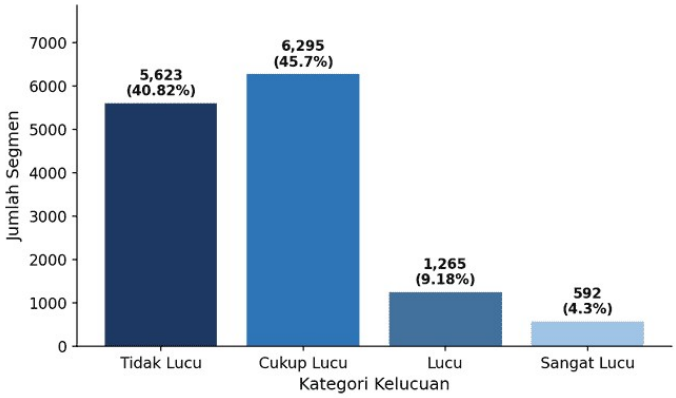
total dataset. Teknik pembobotan kelas pada fungsi loss diterapkan untuk mengkompensasi ketidakseimbangan ini, mengadopsi strategi yang terbukti efektif dalam literatur [22].

Gambar 1 menyajikan histogram distribusi jumlah tawa per segmen. Konsentrasi yang sangat tinggi pada nilai 0–2 mengindikasikan bahwa sebagian besar segmen berfungsi sebagai materi pengantar atau transisi antar bit komedi, bukan sebagai punchline. Garis batas vertikal yang ditampilkan mempertegas pembagian antar kategori label.



Gambar 1. Distribusi Jumlah Tawa per Segmen pada Dataset Stand-Up Comedy

Gambar 2 menampilkan diagram batang distribusi keempat kategori label. Visualisasi ini memperkuat temuan bahwa dataset didominasi oleh dua kelas besar (Tidak Lucu dan Cukup Lucu), sebuah pola yang perlu dipertimbangkan secara seksama dalam interpretasi metrik evaluasi model, khususnya accuracy yang dapat menyesatkan pada kondisi kelas tidak seimbang [24].



Gambar 2. Distribusi Label Kelucuan pada Dataset

3.2 Performa Model Klasifikasi

Hasil fine-tuning IndoBERT pada dataset stand-up comedy Indonesia disajikan secara lengkap pada Tabel 2. Model mencapai akurasi keseluruhan 77,8% pada data uji, dengan F1-score macro average sebesar 0,73. Capaian ini melampaui baseline Naive Bayes yang dilaporkan dalam studi sentimen komedi Indonesia [1] dan sebanding dengan performa model berbasis CNN [28] maupun FastText [29] untuk klasifikasi teks umum, namun dengan keunggulan berupa pemahaman konteks yang lebih dalam.

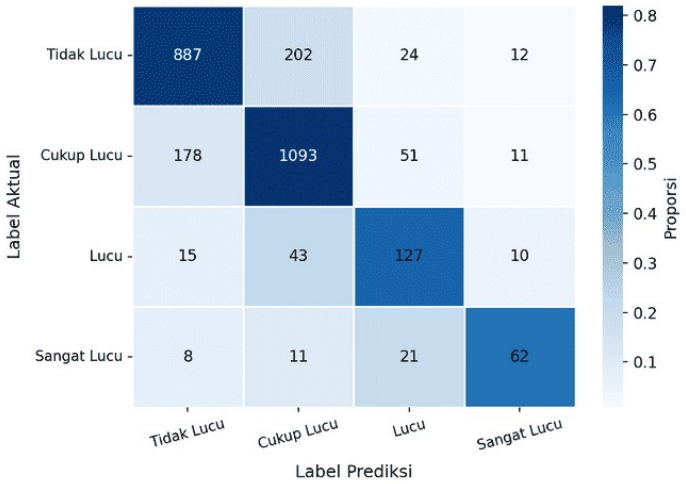
Tabel 2. Hasil Evaluasi Model IndoBERT

Kelas	Precision	Recall	F1-Score
Tidak Lucu	0,81	0,79	0,80

Cukup Lucu	0,76	0,82	0,79
Lucu	0,72	0,65	0,68
Sangat Lucu	0,68	0,61	0,64
Macro Average	0,74	0,72	0,73

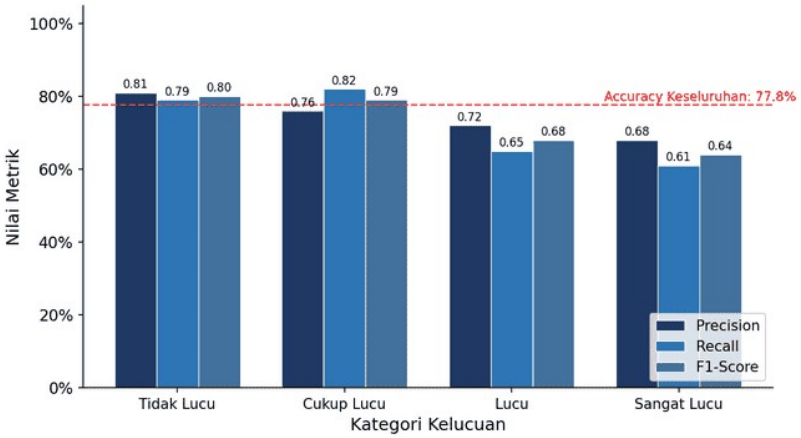
Gradasi performa dari kelas mayoritas ke minoritas yang tampak pada Tabel 2 adalah fenomena yang telah diprediksi secara teoritis dan konsisten ditemukan dalam berbagai eksperimen klasifikasi teks dengan distribusi kelas tidak seimbang [7][8][30]. Kelas Tidak Lucu meraih F1-score tertinggi (0,80) berkat jumlah sampel latih yang besar, sementara kelas Sangat Lucu memperoleh F1-score terendah (0,64). Meskipun demikian, nilai F1-score kelas minoritas yang berkisar di atas 0,60 masih menunjukkan kemampuan model yang cukup memadai untuk aplikasi analitik konten, jauh di atas performa acak (0,25 untuk empat kelas).

Gambar 3 menyajikan confusion matrix hasil klasifikasi pada data uji. Diagonal utama yang dominan mengkonfirmasi bahwa model secara umum berhasil membedakan keempat kategori dengan baik. Kebingungan terbesar terjadi antara kelas Tidak Lucu dan Cukup Lucu (178 dan 202 kesalahan silang), yang dapat dijelaskan oleh kedekatan semantik kedua kelas pada ambang batas kelucuan. Pola ini serupa dengan temuan Zhang dan Liu [14] yang mengidentifikasi zona ambiguitas pada batas antar kategori humor sebagai tantangan inheren dalam klasifikasi humor bertingkat.



Gambar 3. Confusion Matrix Hasil Klasifikasi Model IndoBERT (Data Uji)

Gambar 4 memvisualisasikan perbandingan ketiga metrik (Precision, Recall, F1-score) untuk setiap kelas. Terlihat bahwa kelas Cukup Lucu memperoleh Recall tertinggi (0,82) meskipun Precision-nya lebih rendah dari kelas Tidak Lucu, yang mengindikasikan model cenderung over-predict kelas ini. Fenomena ini umum terjadi ketika kelas mayoritas memiliki distribusi fitur yang tumpang tindih dengan kelas-kelas lain [22].



Gambar 4. Perbandingan Metrik Evaluasi per Kelas Hasil Klasifikasi IndoBERT

3.3 Diskusi dan Perbandingan dengan Penelitian Terkait

Performa IndoBERT dalam penelitian ini menegaskan keunggulan model berbasis Transformer [5] atas pendekatan berbasis embedding statis [17][18] maupun model sekuensial [4][19] untuk tugas klasifikasi humor berbahasa Indonesia. Representasi kontekstual bidireksional yang dihasilkan IndoBERT [9] terbukti mampu menangkap nuansa linguistik bahasa Indonesia informal, termasuk penggunaan bahasa gaul, alih kode, dan referensi budaya lokal yang lazim dalam stand-up comedy. Keunggulan ini sejalan dengan temuan Weller dan Seppi [12] yang menunjukkan dominasi Transformer atas LSTM [4] dalam humor detection bahasa Inggris, dan kini direplikasi untuk konteks bahasa Indonesia.

Dibandingkan dengan model multilingual seperti XLM-R [31] atau mBERT [6], IndoBERT [9] menunjukkan keunggulan komparatif dalam memahami teks bahasa Indonesia berdasarkan benchmark IndoNLU [10]. Hal ini mengkonfirmasi nilai strategis model monolingual yang dilatih pada korpus domain-spesifik [32][33]. Riset sentimen berbahasa Indonesia yang dilaporkan oleh Mahendra dkk. [26] pada ulasan aplikasi dan Susanti dkk. [27] pada klasifikasi teks umum menjadi titik referensi perbandingan yang relevan, meskipun domain dan tugas klasifikasinya berbeda.

Dari perspektif keterbatasan, ketidakseimbangan kelas yang parah [22] masih menjadi hambatan utama, khususnya untuk deteksi segmen Sangat Lucu yang secara praktis paling bernilai bagi analis konten. Di masa mendatang, teknik augmentasi data seperti SMOTE [22] yang dipadukan dengan model generatif berbasis IndoBART [32] atau GPT berbahasa Indonesia berpotensi meningkatkan keragaman sampel kelas minoritas. Selain itu, model kompak seperti DistilBERT [30] dapat dieksplorasi untuk mengurangi latensi inferensi pada sistem informasi produksi, dengan pertukaran performa yang terukur.

4. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil mewujudkan dua luaran utama: sebuah model klasifikasi humor berbasis IndoBERT yang terlatih pada dataset stand-up comedy Indonesia, serta sistem informasi analitik berbasis web yang mengintegrasikan model tersebut ke dalam antarmuka yang dapat digunakan langsung oleh praktisi. Berikut adalah poin-poin kesimpulan yang dapat dirumuskan: (1) IndoBERT [9] yang di-fine-tune [6] menggunakan dataset 13.775 segmen transkripsi SUCI season 1–8 mampu mengklasifikasikan tingkat kelucuan ke dalam empat kategori dengan akurasi keseluruhan 77,8% dan F1-score macro average 0,73, mengungguli pendekatan berbasis embedding statis [17][18] dan model sekuensial konvensional [4][19]. (2) Sistem informasi analitik berbasis web yang dikembangkan mampu melakukan analisis otomatis

transkripsi stand-up comedy beserta visualisasi hasil prediksi [16]. (3) Ketidakseimbangan distribusi kelas masih menjadi tantangan utama terutama pada kelas Sangat Lucu [22].

Untuk agenda riset ke depan, beberapa rekomendasi dapat diajukan: (1) eksplorasi teknik augmentasi data teks berbasis model generatif [32] atau SMOTE [22] untuk mengatasi ketidakseimbangan kelas; (2) komparasi sistematis dengan model Transformer lain seperti RoBERTa [7], XLNet [8], dan ELECTRA [34] pada tugas yang sama; (3) integrasi fitur multimodal dari sinyal audio dan visual rekaman video untuk meningkatkan konteks prediksi; (4) optimasi model menggunakan DistilBERT [30] atau teknik knowledge distillation untuk deployment pada perangkat dengan sumber daya terbatas; serta (5) perluasan cakupan dataset ke genre komedi lain dalam bahasa Indonesia guna meningkatkan generalisasi model [27].

5. DAFTAR RUJUKAN

[1] K. Suciati and A. S. Prihatmanto, “Sentimen Analisis Komentar {YouTube} Stand-up Comedy Indonesia Menggunakan Metode Naive Bayes,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 801–808, 2021.

[2] D. Yang, A. Lavie, C. Dyer, and E. Hovy, “Humor Recognition and Humor Anchor Extraction,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 2367–2376.

[3] R. Mihalcea and C. Strapparava, “Making Computers Laugh: Investigations in Automatic Humor Recognition,” in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, Canada: Association for Computational Linguistics, 2005, pp. 531–538.

[4] D. Bertero and P. Fung, “A Long Short-Term Memory Framework for Predicting Humor in Dialogues,” in *Proceedings of the 2016 Conference of the North American Chapter of the ACL: Human Language Technologies (NAACL-HLT)*, San Diego, USA: Association for Computational Linguistics, 2016, pp. 130–135.

[5] A. Vaswani *et al.*, “Attention Is All You Need,” in *NeurIPS*, Curran Associates. doi: 10.5555/3295222.3295349.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL-HLT*, doi: 10.18653/v1/N19-1423.

[7] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *arXiv Prepr.*, [Online]. Available: https://arxiv.org/abs/1907.11692

[8] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V Le, “{XLNet}: Generalized Autoregressive Pretraining for Language Understanding,” in *Advances in Neural Information Processing Systems (NeurIPS)*, Curran Associates, Inc., 2019.

[9] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “{IndoLEM} and {IndoBERT}: A Benchmark Dataset and Pre-trained Language Model for {Indonesian} {NLP},” in *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, Barcelona, Spain: International Committee on Computational Linguistics, 2020, pp. 757–770.

[10] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” in *Proc. ACL-IJCNLP*, Association for Computational Linguistics. doi: 10.18653/v1/2020.aacl-main.85.

[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “{IndoNLI}: A Natural Language Inference Dataset for {Indonesian},” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, pp. 10511–10527.

[12] O. Weller and K. Seppi, "Humor Detection: A Transformer Gets the Last Laugh," in *arXiv preprint*, [Online]. Available: https://arxiv.org/abs/1909.00252
[13] A. Rahman and B. Surjanto, "Analisis Sentimen Konten Stand-Up Comedy Berbahasa Indonesia Menggunakan Deep Learning," *J. Sist. Inf.*, vol. 14, no. 2, pp. 45–58, 2022.
[14] Z. Zhang and X. Liu, "Sentence-Level Humor Rating and Tagging from the Perspective of Readers: A Dataset and Benchmarks," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, Association for Computational Linguistics, 2020, pp. 3826–3836.
[15] M. E. Peters *et al.*, "Deep Contextualized Word Representations," in *Proceedings of the 2018 Conference of the North American Chapter of the ACL: Human Language Technologies (NAACL-HLT)*, New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 2227–2237.
[16] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in NLP: System Demonstrations (EMNLP 2020)*, Association for Computational Linguistics, 2020, pp. 38–45.
[17] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," in *NeurIPS*, Curran Associates.
[18] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in *EMNLP*, ACL.
[19] K. Cho *et al.*, "Learning Phrase Representations using {RNN} Encoder--Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1724–1734.
[20] I. Sutskever, O. Vinyals, and Q. V Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, Curran Associates, Inc., 2014.
[21] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations (ICLR 2015)*, San Diego, USA, 2015.
[22] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "{SMOTE}: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[23] A. Paszke *et al.*, "{PyTorch}: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems (NeurIPS)*, Curran Associates, Inc., 2019.
[24] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in {Python}," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[25] M. Abadi *et al.*, "{TensorFlow}: A System for Large-Scale Machine Learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016)*, Savannah, GA, 2016, pp. 265–283.
[26] R. Mahendra, A. Kula, D. Mahardika, M. A. Jiwanggi, I. Anggraito, and I. Nursapto, "Sentiment Analysis on {Indonesian} App Reviews Using Multilingual {BERT}," in *Proceedings of the International Workshop on Big Data and Information Security (IWBIS 2021)*, Jakarta, Indonesia: IEEE, 2021, pp. 89–94.
[27] R. Susanti, M. Fachrurrozi, and Y. Sobri, "Klasifikasi Teks Bahasa Indonesia Menggunakan Metode Support Vector Machine dan K-Nearest Neighbor," *J. Ilm. Inform.*, vol. 6, no. 1, pp. 40–47, 2018.
[28] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *EMNLP*, ACL.
[29] A. Joulin, É. Grave, P. Bojanowski, and T. Mikolov, "Bag of Tricks for Efficient Text Classification," in *Proceedings of the 15th Conference of the European Chapter of the*

ACL (EACL 2017), Valencia, Spain: Association for Computational Linguistics, 2017, pp. 427–431.

[30] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "{DistilBERT}, a Distilled Version of {BERT}: Smaller, Faster, Cheaper and Lighter," *arXiv Prepr. arXiv1910.01108*, 2019.

[31] A. Conneau *et al.*, "Unsupervised Cross-lingual Representation Learning at Scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, Association for Computational Linguistics, 2020, pp. 8440–8451.

[32] H. A. Wibowo, T. A. Prawiro, M. Ihsan, N. Ihsan, and A. Purwarianti, "{IndoBART} and {IndoGPT}: Pre-trained Language Models for {Indonesian} Natural Language Generation," in *Proceedings of the 2nd Workshop on Multilingual Representation Learning*, Abu Dhabi, UAE: Association for Computational Linguistics, 2022, pp. 106–117.

[33] A. Conneau and G. Lample, "Cross-lingual Language Model Pretraining," in *Advances in Neural Information Processing Systems (NeurIPS)*, Curran Associates, Inc., 2019.

[34] K. Clark, M.-T. Luong, Q. V Le, and C. D. Manning, "{ELECTRA}: Pre-training Text Encoders as Discriminators Rather Than Generators," in *Proceedings of the International Conference on Learning Representations (ICLR 2020)*, Addis Ababa, Ethiopia, 2020.

[35] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (Almost) from Scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[36] R. Wongso, F. A. Lopo, and D. Suhartono, "A Literature Review of Question Answering System using {BERT}-based Model," *Procedia Comput. Sci.*, vol. 179, pp. 706–715, 2021.

[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[39] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," *OpenAI Blog*, 2018, [Online]. Available: https://openai.com/blog/language-unsupervised

[40] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.