

Anda akan Tercengang dengan Dampak *Data Augmentation* pada Model *Clickbait Classifier*!

Darrel Danadyaksa Poli¹, Edbert Halim², Patrick Samuel Evans Simanjuntak³, Andi Pujo Rahadi⁴

¹Fakultas Ilmu Komputer Universitas Indonesia, Depok, 16424, email: darrel.danadyaksa@ui.ac.id

²Fakultas Ilmu Komputer Universitas Indonesia, Depok, 16424, email: edbert.halim@ui.ac.id

³Fakultas Ilmu Komputer Universitas Indonesia, Depok, 16424, email: patrick.samuel@ui.ac.id

⁴Fakultas Ilmu Komputer Universitas Indonesia, Depok, 16424, email: andi.pujo@ui.ac.id

Corresponding Author: Darrel Danadyaksa Poli

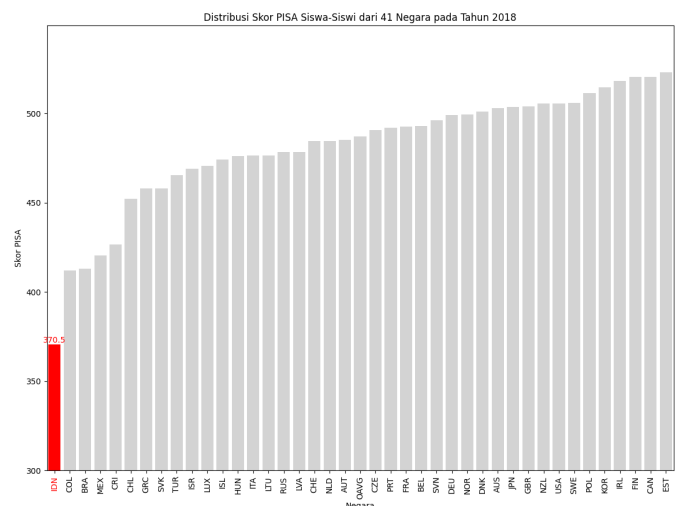
ABSTRACT *Clickbait* adalah artikel daring dengan judul menyesatkan yang sengaja dibuat untuk menarik pembaca agar membuka halaman berita tersebut. Hal ini dilakukan dengan menggunakan judul yang kontroversial sehingga memancing rasa ingin tahu pembaca. Meskipun tujuan awalnya adalah untuk meningkatkan jumlah klik dan kunjungan ke halaman berita, adanya unsur *clickbait* pada suatu judul berita juga dapat menyebabkan disinformasi dan mengelabui masyarakat yang memiliki minat baca rendah. Penelitian ini bertujuan untuk mengembangkan sebuah model yang dapat secara otomatis menentukan apakah suatu judul berita mengandung unsur *clickbait* atau tidak. Selain itu, penelitian ini juga bertujuan untuk menguji dan membandingkan performa beberapa model *machine learning* yang berbeda dalam mendeteksi unsur *clickbait* pada judul berita. Dibangun empat model *machine learning*, yaitu *IndoBERT*, *XGBoost*, *Random Forest*, dan *Naïve Bayes*. Performa masing-masing model diukur menggunakan metrik *accuracy score* dan *recall score*. Model yang paling akurat adalah *IndoBERT* dengan *accuracy score* = 0.9094 dan *recall score* = 0.8825. Model *IndoBERT* tersebut menggunakan 8 ribu data dari *dataset* CLICK-ID [1] yang diberikan perlakuan *data augmentation swap*, kemudian modelnya dilakukan *fine tuning* dengan *batch size* 32, *learning rate* 0.00002, serta 4 *epochs*. Selain itu didapati juga bahwa *data augmentation swap* mendapatkan skor yang bahkan lebih tinggi daripada skor model yang menggunakan *dataset* asli. Pada penelitian, proses *data augmentation swap* berarti menukar kata-kata atau frasa dalam judul berita yang ada pada *dataset*. Terakhir, penelitian ini berkesimpulan bahwa *data augmentation swap* dapat menghindari *overfitting* pada model karena memiliki skor *recall* dan *accuracy* yang lebih tinggi dari *dataset* awal yang hanya dibersihkan dari elemen-elemen yang bukan alfabet.

Kata Kunci Berita Bahasa Indonesia, *Clickbait*, *Data Augmentation*, *IndoBERT*, *Naïve Bayes*, *Natural Language Processing*, *Random Forest*, *XGBoost*

I. Pendahuluan

A. Latar Belakang

Ketersediaan informasi di Indonesia melalui media sosial berkembang dengan pesat belakangan ini. Bahkan berdasarkan survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pengguna internet di Indonesia telah mencapai 215,64 juta orang pada periode 2022-2023. Tentunya dengan akses internet yang semakin besar, akses masyarakat Indonesia terhadap informasi semakin meningkat. Akan tetapi, akses terhadap informasi yang ada tidak sebanding dengan tingkat literasi yang ada di Indonesia. Berdasarkan data tes PISA tahun 2018 pada penelitian yang dilakukan oleh *Organization for Economic Co-operation and Development* (OECD), Indonesia menempati peringkat terakhir dari 41 negara yang diteliti dalam kemampuan membaca. Hasil ini memperlihatkan bahwa tingkat literasi Indonesia masih kurang baik [2].



Gambar 1. Perbandingan Skor PISA Indonesia dengan Negara Lain

Ketersediaan internet membuat masyarakat Indonesia mulai meninggalkan media cetak dan beralih ke media digital [3]. Banyaknya pembaca berita pada media digital membuat persaingan antar penulis berita digital semakin sengit. Untuk mendapatkan atensi dari pembaca, penulis berita daring seringkali membuat suatu judul berita daring yang mengandung unsur *clickbait*. *Clickbait* adalah suatu bentuk konten palsu yang sengaja dibuat untuk menarik perhatian calon pembaca dan membuat mereka penasaran agar mereka membaca konten yang ada [4]. Di sisi lain, judul *clickbait* berarti judul yang menarik calon pembaca untuk mengklik suatu tautan dengan membangkitkan rasa penasaran calon pembaca [5]. Judul berita yang mengandung unsur *clickbait* biasanya tidak menampilkan isi berita seutuhnya. Hal tersebut dapat menjadi hal yang berbahaya jika pembaca berita memiliki tingkat literasi yang rendah. Kurangnya tingkat literasi menyebabkan seseorang hanya membaca judul berita saja dan menganggapnya sebagai kebenaran mutlak tanpa membaca keseluruhan dari beritanya terlebih dahulu [6]. Hal ini dapat menyebabkan disinformasi secara massal yang berpotensi menimbulkan berbagai isu sosial yang merugikan [7].

Alasan lain yang membuat *clickbait* perlu dicegah adalah hal tersebut memanfaatkan *information gap* pada judul untuk memicu rasa penasaran pada para pembaca yang seringkali justru menyesatkan para pembaca [8]. *Clickbait* memanfaatkan *cognitive bias* dalam otak manusia yang dapat dipengaruhi dengan mudah oleh media [9]. Keresahan inilah yang membuat diperlukannya sebuah *clickbait classifier* untuk ada di tengah-tengah medial sosial untuk mencegah hal tersebut.

Selain masalah *clickbait*, ditemukan juga adanya masalah lain pada bidang *machine learning*. Salah satu tahapan yang menyita banyak waktu adalah *data labeling* [10]. Label pada data ini penting supaya model dapat mengerti apa yang diinginkan dari data yang ada. Peneliti menemukan suatu cara untuk melakukan efisiensi waktu pada tahapan *data labeling* dengan menggunakan metode *data augmentation*.

B. Tujuan Penelitian

Tujuan dari penelitian ini adalah membuat beberapa model yang dapat mengenali unsur *clickbait* pada suatu judul berita serta menentukan model apa yang paling akurat untuk menentukan apakah suatu judul berita mengandung unsur *clickbait* atau tidak. Penelitian ini juga berfokus pada multiplikasi data dengan cara melakukan *data augmentation* untuk menyelesaikan masalah banyaknya waktu yang dibutuhkan untuk *data labeling* dalam permasalahan *machine learning* secara umum yang diimplementasikan pada kasus spesifik *clickbait classifier*. Selain itu, ingin dibandingkan data hasil augmentasi dan data biasa jika digunakan dalam model.

C. Manfaat Penelitian

Manfaat dari penelitian ini bagi peneliti adalah menambah pengetahuan dan pengalaman mengenai model-model di bidang NLP (*Natural Language Processing*). Manfaat dari

penelitian ini bagi masyarakat adalah membantu mengurangi disinformasi yang sering terjadi di Indonesia, terutama pada masyarakat yang memiliki tingkat literasi yang rendah.

D. Batasan Penelitian

Penelitian ini meneliti dan menentukan suatu model yang paling akurat dalam menentukan apakah suatu judul berita dalam bahasa Indonesia mengandung unsur *clickbait* atau tidak. Selain itu, penelitian ini menggunakan empat jenis *data augmentation*, yaitu: *insert*, *swap*, *delete*, dan *substitute*.

II. Kajian Terkait yang Relevan

Di Indonesia, ada beberapa penelitian yang telah mencoba untuk membuat sebuah model yang dapat melakukan tugas klasifikasi berita. Sebuah penelitian [11] mencoba untuk melakukan klasifikasi terhadap judul *clickbait* dengan melakukan *similarity scoring* dan juga mempertimbangkan ringkasan dari konten berita. Pada penelitian tersebut digunakan model *IndoBERT*. Penelitian tersebut berbeda dengan penelitian lain [12] yang telah memerhatikan frekuensi sebuah kata muncul, nilai posisi kalimat, kesamaan dengan judul, panjang kalimat, pengurangan kalimat, dan peringkat kalimat. Tidak hanya kedua penelitian tersebut, ada penelitian lain [13] yang menggunakan *Multilingual Bidirectional Encoder Representations from Transformers* (MBERT). Penelitian ini menggunakan data dari CLICK-ID [1] yang juga menjadi data pada penelitian lain yang telah dilakukan [11]. Penelitian lain juga dilakukan oleh [14] yang menggunakan model *Multinomial Naïve Bayes* untuk melakukan klasifikasi judul berita.

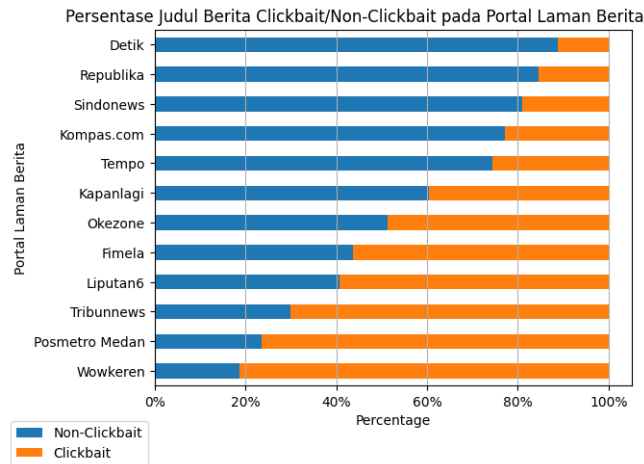
III. Solusi Usulan

A. Deskripsi Solusi

Untuk menyelesaikan masalah disinformasi yang ada, maka akan dibuat sebuah model *classifier* yang dapat mengidentifikasi berita manakah yang teridentifikasi sebagai berita *clickbait* dan *non-clickbait*. Di sisi lain, penyelesaian masalah *data labeling* dapat diselesaikan dengan membuat model yang akan mendapatkan *dataset* yang diberikan perlakuan *data augmentation* dan yang tidak diberikan perlakuan *data augmentation*. Dengan begitu, dapat dibandingkan model yang mendapatkan data yang mendapatkan perlakuan *data augmentation* dan yang tidak mendapat perlakuan *data augmentation*.

B. Dataset

Penelitian ini menggunakan data berupa kumpulan berita daring di Indonesia yang dipublikasi selama empat tahun terakhir, yaitu dari tahun 2019 hingga tahun 2023. Data-data yang digunakan didapatkan dari laman Data Mendeley [1]. Data CLICK-ID telah dilakukan *labeling* secara manual dengan memerhatikan judul oleh beberapa peneliti. Dalam penelitian tersebut, terdapat 2 *dataset*. Dataset pertama memiliki 14 ribu baris dimana tidak semua orang yang melabeli data tersebut setuju dengan labelnya. Dataset kedua



Gambar 2. Jumlah Data *Clickbait/Non-Clickbait* pada Masing-masing Portal Berita Daring

memiliki 8 ribu baris dimana semua orang yang melabeli data tersebut setuju dengan labelnya. Penelitian ini juga telah melakukan *scrapping* untuk alasan kebaruan *dataset* sampai pada tahun 2023.

Berikut adalah contoh data baru yang telah dilabel secara manual:

Tabel 1. Headlines

Headline	Category
8 Zodiak Paling Beruntung Hari Ini Jumat 7 Juni 2023: Aries Raih Kekayaan - Pisces Hargai Bantuan	Clickbait
Daftar Presiden Indonesia yang Lahir Bulan Juni, Jokowi Termasuk?	Clickbait
Pegawai Dapat Fasilitas Tempat Tinggal di Atas Rp 2 Juta/Bulan Kena Pajak	Non-Clickbait
KPK Ungkap Korban Pungli Rp 4 M di Rutan Capai Puluhan Orang	Non-Clickbait

C. Visualisasi Dataset

Pada Gambar 2 terdapat visualisasi dari *dataset* CLICK-ID untuk setiap penerbit berita. Gambar tersebut menunjukkan *dataset* dengan 14 ribu baris.

D. Metode yang Digunakan

1) Model

Penelitian ini menggunakan *XGBoost*, *Random Forest*, *Naïve Bayes*, dan *IndoBERT*. *XGBoost* menggunakan ensemble tree untuk menghasilkan prediksi yang baik [15]. *Random Forest* menggabungkan prediktor pohon secara acak [16]. *Naïve Bayes* melakukan prediksi berdasarkan probabilitas dan melihat fitur-fitur yang ada secara mandiri [17]. Model *Naive Bayes* yang digunakan pada penelitian ini adalah *MultinomialNB*. *IndoBERT* adalah model *pre-trained* yang telah dilatih sebelumnya untuk melaksanakan tugas tertentu yang belum spesifik [18], [19].

2) Data Pre-Processing

Ada 4 model yang digunakan, yaitu *XGBoost*, *Random Forest*, *Naïve Bayes*, dan *IndoBERT*. Ketiga model pertama, yaitu *XGBoost*, *Random Forest*, dan *Naïve Bayes*. *Dataset* awal yang digunakan pada penelitian ini terdiri dari 8566 baris. Kemudian, *dataset* tersebut dibagi menjadi data *train* dan *test* dengan perbandingan 70-30 dengan data *train* sebanyak 5988 baris dan data *test* sebanyak 2569 baris. Penelitian pertama, menggunakan judul asli dari *dataset* awal. *train*. Penelitian kedua hingga kelima menggunakan judul asli yang digabung dengan satu jenis augmentasi yang menggunakan data *test* seperti *dataset* awal, tetapi data *train* dilakukan augmentasi *insert*, *swap*, *delete*, dan *substitute*. Sedangkan pada percobaan 6, semua jenis data augmentasi dan data awal digabung.

Ketiga model yang telah dijelaskan diberikan *feature engineering*, yaitu: *has exclamation* (tanda seru), *has question* (tanda tanya), *has number* (mengandung angka), *num count* (jumlah angka yang muncul), *mark count* (jumlah tanda baca), *total character* (jumlah huruf pada judul berita), serta kolom *category* (kategori berita). Setelah itu, judul dibersihkan dengan menghilangkan elemen-elemen yang bukan merupakan alfabet yang selanjutnya, judul dan konten di-stem dengan menggunakan *StemmerFactory* dari Library *pySastrawi* [20] dan kemudian diaugmentasi jika dibutuhkan. Terakhir, judul yang telah diproses akan melalui proses vektorisasi dengan menggunakan *TF-IDF Vectorizer*. Parameter yang digunakan pada *TF-IDF Vectorizer* adalah *ngram range*=(1,1), *analyzer*=‘word’, *norm*=‘l2’.

Pada model yang menggunakan arsitektur *IndoBERT*, teknik *pre-processing* yang dilakukan adalah menghilangkan semua komponen yang bukan alfabet dan menghilangkan *stopword* dan memberikan *stemming* pada judul karena metode tersebut sudah terbukti lebih baik [20]. Meskipun begitu, ada juga model yang tidak menghilangkan *stopword*. Selain itu, *IndoBERT* menghapus konten karena keterbatasan sumber daya untuk pemrosesan data. Walaupun begitu, hal yang pasti dilakukan adalah *splitting dataset* dengan perbandingan sebanyak 80-15-5 untuk *train*, *validation*, dan *test*.

E. Perbedaan dengan solusi sebelumnya

Penelitian ini menggunakan *data augmentation*. *Data augmentation* dapat meningkatkan variasi dalam data. Selain untuk meningkatkan jumlah ketersediaan data, data yang sudah bersifat variatif dapat mengurangi *overfitting* pada model karena tidak terlalu fokus pada data pelatihan tertentu. Pemahaman pada model juga meningkat karena model dapat mengeksplorasi dan mempelajari pola linguistik yang berbeda [21]. Selain itu, ada juga model yang tidak menghapus *stopword*. Harapannya, model dapat lebih mengerti konteks.

F. Evaluasi model

Evaluasi model merupakan hal yang penting karena hal tersebut merupakan cara untuk melihat seberapa bagus model

yang telah dibuat. Metrik yang sering digunakan antara lain adalah *accuracy*, *precision*, *recall*, serta *F1 score*. Penelitian ini berfokus pada *accuracy score* dan *recall score*.

Accuracy score digunakan sebagai salah satu acuan evaluasi karena *accuracy score* mencerminkan kebenaran prediksi secara keseluruhan [22]. Hal tersebutlah yang membuat *accuracy* merupakan metrik yang penting. Persamaan dari *accuracy score* adalah sebagai berikut:

$$\frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}}$$

Seperti yang telah disebutkan, penelitian ini menggunakan skor *accuracy* dan *recall* sebagai metrik evaluasi model. Recall dipilih karena pentingnya mengidentifikasi *false negative*, yaitu kasus ketika judul berita yang seharusnya dikategorikan sebagai *clickbait* namun salah dikategorikan sebagai *non-clickbait* [23]. Menghindari *false negative* penting agar tidak ada berita *clickbait* yang terlewatkan agar pembaca tidak mendapatkan berita *clickbait*. Perhatikan bahwa karena alasan tersebut juga *recall score* sering disebut juga sebagai *sensitivity score*. *Recall* dapat dihitung menggunakan persamaan berikut:

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

IV. Hasil Eksperimen dan Pengujian

Berikut adalah hasil skor *accuracy* dan *recall* untuk *XGBoost*, *Random Forest*, dan *Naïve Bayes* dan *dataset* yang digunakan (Lihat tabel 2).

Selain *XGBoost*, *Random Forest* dan *Naïve Bayes*, *IndoBERT* juga digunakan dalam penelitian ini. Berikut adalah hasil skor *recall* dan *accuracy* untuk lebih dari 100 model yang telah dilakukan *fine-tuning* pada penelitian ini yang telah dilakukan *grouping* supaya lebih mudah untuk dibaca. Perhatikan juga bahwa garis pada grafik menunjukkan skor rata-rata. Hal ini akan dianalisis pada bagian berikutnya. Lihat juga bahwa hasil skor yang lebih jelas telah dilampirkan pada bagian lampiran dari penelitian ini.

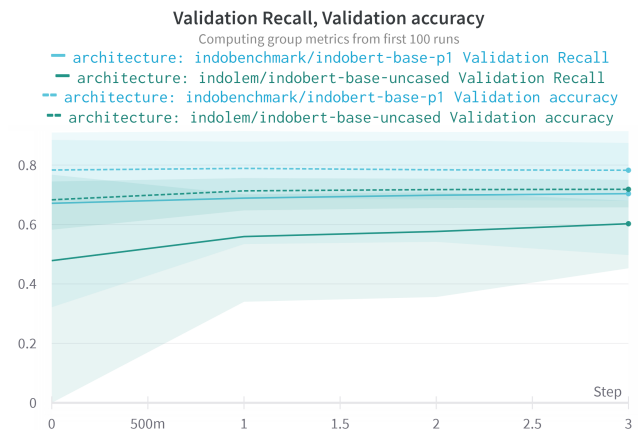
V. Analisis atas Hasil Eksperimen dan Pengujian

Hal menarik yang didapatkan adalah arsitektur *pre-trained indobert-base-p1* secara rata-rata memiliki skor *accuracy* dan *recall* lebih tinggi daripada model yang menggunakan arsitektur *pre-trained indobert-base-uncased*. Perhatikan bahwa arsitektur *indobert-base-uncased* dilatih dengan menggunakan *dataset* yang berisikan 230 juta kata [18], sedangkan arsitektur *indobert-base-p1* dilatih dengan menggunakan *dataset Indo4B* yang berisikan 4 milyar kata [19]. Dari perbandingan tersebut, dapat disimpulkan bahwa arsitektur *indobert-base-p1* akan lebih baik dalam mengerti konteks.

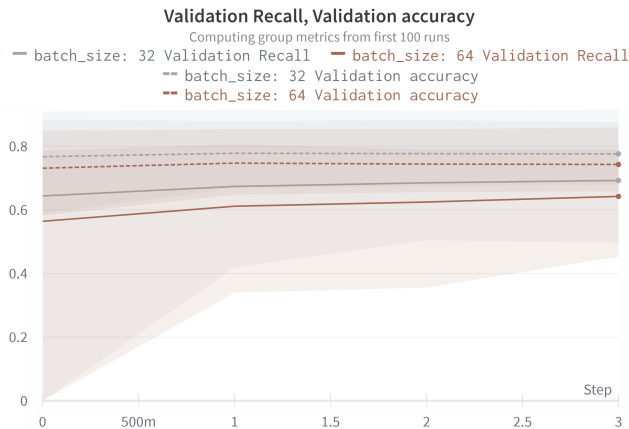
Pada pengujian dengan ukuran *batch size* yang berbeda, tidak terdapat perbedaan skor *accuracy* dan *recall* yang signifikan. Perbedaan skor yang tidak signifikan dapat dijelaskan karena tidak ada pengaruh besar dalam seberapa sering kita mengupdate nilai *weight* dan *bias* pada model.

Tabel 2. Model Evaluation

Model	Original Title	
	Accuracy	Recall
<i>XGBoost</i>	0.9107	0.8464
<i>Random Forest</i>	0.9119	0.8372
<i>Naïve Bayes</i>	0.8531	0.6724
Model	Original + Insertion Augmented Title	
	Accuracy	Recall
<i>XGBoost</i>	0.9049	0.8341
<i>Random Forest</i>	0.9107	0.8311
<i>Naïve Bayes</i>	0.8554	0.6970
Model	Original + Substitute Augmented Title	
	Accuracy	Recall
<i>XGBoost</i>	0.8924	0.8321
<i>Random Forest</i>	0.9068	0.8393
<i>Naïve Bayes</i>	0.8687	0.7635
Model	Original + Swap Augmented Title	
	Accuracy	Recall
<i>XGBoost</i>	0.9135	0.8495
<i>Random Forest</i>	0.9181	0.8485
<i>Naïve Bayes</i>	0.8652	0.7523
Model	Original + Delete Augmented Title	
	Accuracy	Recall
<i>XGBoost</i>	0.8952	0.8495
<i>Random Forest</i>	0.9123	0.8556
<i>Naïve Bayes</i>	0.8628	0.7523
Model	Original + All Combined Augmented Title	
	Accuracy	Recall
<i>XGBoost</i>	0.8924	0.8689
<i>Random Forest</i>	0.9033	0.8536
<i>Naïve Bayes</i>	0.8617	0.7676



Gambar 3. Hasil Skor *accuracy* dan *recall* untuk setiap arsitektur



Gambar 4. Hasil Skor *accuracy* dan *recall* untuk setiap *Batch Size*

Didapatkan juga bahwa metode *swap* pada augmentasi data untuk model *IndoBERT* memiliki skor yang tinggi. Hal ini dapat menunjukkan bahwa *swap* dapat mengurangi *overfitting*. Selain itu, model dengan kombinasi arsitektur *indobert-base-p1* dan *learning rate* 2.10^{-5} memiliki skor yang baik secara konsisten. Disini, kombinasi arsitektur dan *learning rate* yang telah terbukti [19] memberikan skor yang lebih baik. Hal tersebut dikarenakan *fine-tuning* cukup menggunakan *learning rate* yang kecil.

Perhatikan bahwa hal yang telah dijelaskan diatas merupakan hal yang didapatkan pada saat model dijalankan dengan menggunakan *dataset* yang berisi lebih dari 14 ribu baris. Ketika digunakan *dataset* yang berisi 8 ribu baris, didapatkan skor *recall* dan *accuracy* yang lebih tinggi. Hal ini dapat terjadi karena kualitas dari *dataset* yang sangat memengaruhi model.

Perlu dicatat bahwa *dataset* dengan data *augmentation swap* pada *dataset* dengan 8 ribu baris merupakan *dataset* yang dapat membuat skor *recall* dan *accuracy* yang baik. Hal ini juga yang dapat menjadi alasan lain untuk menggunakan data *augmentation swap* untuk menghindari *overfitting* pada data yang ada.

Hal menarik yang didapati dari antara model *XGBoost*, *Random Forest*, dan *Bayes* adalah model *Random Forest* dalam penelitian yang menggunakan judul sebenarnya yang digabungkan dengan data *augmentation swap* dan *delete* memiliki *accuracy score* dan *recall score* yang paling tinggi dibanding metode penelitian lain. Selain itu, peneliti juga menemukan bahwa teknik augmentasi terbaik dalam penelitian ini adalah augmentasi *swap* karena membuat model-model mendapatkan skor *recall* dan *accuracy* yang cenderung lebih tinggi dibandingkan teknik augmentasi lainnya. Penelitian ini juga menemukan bahwa penambahan kolom konten pada *dataset* menunjukkan hasil negatif dimana model-model yang menggunakan *dataset* yang disertai dengan konten berita memiliki skor *accuracy* dan *recall* yang lebih rendah dibandingkan model-model yang berisi *dataset* tanpa kolom konten (lihat lampiran). Hipotesis yang dite-

mukan oleh peneliti adalah kolom konten memiliki banyak sekali jenis kata yang menyebabkan dimensi yang sangat besar, tetapi tidak memiliki data yang cukup.

VI. Kesimpulan dan Saran

A. Kesimpulan

Dari eksperimen yang telah dilakukan dapat disimpulkan bahwa *IndoBERT* merupakan model yang baik untuk digunakan pada kasus ini karena skor *recall* yang dimiliki oleh model tersebut adalah yang paling tinggi diantara model-model yang diuji. Selain itu, dapat disimpulkan juga bahwa kita perlu *hyperparameter* yang seimbang agar model tidak mengalami *underfitting* ataupun *overfitting*. Hal lain yang perlu diperhatikan adalah pemilihan arsitektur yang baik dapat berpengaruh pada skor metrik yang baik. Terakhir, augmentasi data *swap* terbukti berpengaruh baik bagi model untuk kasus ini karena skor *accuracy* yang tinggi secara umum dan skor *recall* yang tinggi pada *epoch* pertama. Perlu diingat juga bahwa *dataset* yang perlu digunakan adalah *dataset* dengan 8 ribu baris yang telah dibuat [1] oleh karena kualitas labelnya.

B. Saran

Saran untuk penelitian selanjutnya adalah untuk mencoba menggunakan model *BERT* dengan arsitektur lain. Selain itu, sangat direkomendasikan untuk menggunakan augmentasi data *swap* untuk menghindarkan model dari *overfitting* untuk kasus ini mau pun kasus lain yang membutuhkan *classifier*. Saran untuk model *XGBoost*, *Random Forest*, dan *Bayes* serta *vectorizer* lain dengan berbagai parameter yang ada sehingga model dapat menangkap data berupa teks dengan lebih baik.

VII. Lampiran

Berikut adalah lampiran untuk penelitian ini: <https://drive.google.com/drive/folders/1l6RmvRcUa3UD-VuGg4JVrfueUMfF81rZ?usp=sharing>

REFERENCES

- [1] A. William and Y. Sari, "Click-id: A novel dataset for indonesian clickbait headlines," *Data in Brief*, vol. 32, p. 106231, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352340920311252>
- [2] F. Yusmar and R. E. Fadilah, "Analisis rendahnya literasi sains peserta didik indonesia: hasil pisa dan faktor penyebab," *LENZA (Lentera Sains): Jurnal Pendidikan IPA*, vol. 13, no. 1, p. 11–19, 2023.
- [3] B. A. Romadhoni, "Meredupnya media cetak, dampak kemajuan teknologi informasi," *An-Nida : Jurnal Komunikasi Islam*.
- [4] D. Varshney and D. Vishwakarma, "A unified approach for detection of clickbait videos on youtube using cognitive evidences," *Applied Intelligence*, vol. 51, pp. 1–22, 07 2021.
- [5] Şura Genç and E. Surer, "Clickbaittr: Dataset for clickbait detection from turkish news sites and social media with a comparative analysis via machine learning algorithms," *Journal of Information Science*, vol. 49, no. 2, pp. 480–499, 2023. [Online]. Available: <https://doi.org/10.1177/01655515211007746>
- [6] S. Ferdianto, "Dari literasi rendah sampai maraknya berita hoaks," Sep 2022. [Online]. Available: <https://geotimes.id/opini/dari-literasi-rendah-sampai-maraknya-berita-hoaks/>
- [7] K. Shu, S. Wang, D. Lee, and H. Liu, "Mining disinformation and fake news: Concepts, methods, and recent advancements," 2020.

- [8] K. Scott, "You won't believe what's in this paper! clickbait, relevance and the curiosity gap," *Journal of Pragmatics*, vol. 175, pp. 53–66, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378216621000229>
- [9] S. Pandey and G. Kaur, "Curious to click it?-identifying clickbait using deep learning and evolutionary algorithm," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018, pp. 1481–1487.
- [10] "Analisis sentimen pada media sosial twitter terhadap kebijakan pemberlakuan pembatasan kegiatan masyarakat berbasis deep learning," *Jurnal Edukasi dan penelitian Informatika*, April 2022.
- [11] H. Ahmadi and A. Chowanda, "Clickbait classification model on online news with semantic similarity calculation between news title and content," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, pp. 1986–1994, Mar. 2023. [Online]. Available: <https://ejurnal.seminar-id.com/index.php/bits/article/view/3030>
- [12] B. W. Rauf, S. Raharjo, and H. Sismoro, "Deteksi clickbait dengan sentence scoring based on frequency," *Jurnal Teknologi Informasi*, vol. 4, no. 2, p. 247, December 2020. [Online]. Available: <http://jurnal.una.ac.id/index.php/jurti/article/view/1381/1458>
- [13] M. N. Fakhruzzaman, S. Z. Jannah, R. A. Ningrum, and I. Fahmiyah, "Clickbait headline detection in indonesian news sites using multilingual bidirectional encoder representations from transformers (m-bert)," 2021.
- [14] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, "Penerapan metode multinomial naïve bayes untuk klasifikasi judul berita clickbait dengan term frequency - inverse document frequency," *Jurusan Informatika Universitas Tanjungpura*, vol. Vol 11, No 1 (2023), 2023. [Online]. Available: <https://jurnal.untan.ac.id/index.php/justin/article/view/57452/75676596395>
- [15] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016. [Online]. Available: <https://doi.org/10.1145/2939672.2939785>
- [16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 10 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [17] Vikramkumar, V. B., and Trilochan, "Bayes and naïve bayes classifier," 2014.
- [18] 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 COLING*, 2020.
- [19] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, "Indonlu: Benchmark and resources for evaluating indonesian natural language understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020.
- [20] F. TALA, "A study of stemming effects on information retrieval in bahasa indonesia," *M. Sc. Thesis, Appendix D, University of Amsterdam*, pp. 39–46, 2003.
- [21] M. Bayer, M.-A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers," *International Journal of Machine Learning and Cybernetics*, vol. 14, 04 2022.
- [22] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: A family of discriminant measures for performance evaluation," in *AI 2006: Advances in Artificial Intelligence*, A. Sattar and B.-h. Kang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1015–1021.
- [23] D. M. W. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2020.