



## Klasifikasi *Clickbait* Menggunakan *Transformers*

**Mori Hovipah<sup>\*1</sup>, Elin Hearani<sup>2</sup>, Jasril<sup>3</sup>, Fadhilah Syafria<sup>4</sup>**

Email: <sup>1</sup>11850124984@students.uin-suska.ac.id, <sup>2</sup>elin.haerani@uin-suska.ac.id, <sup>3</sup>jasril@uinsuska.ac.id, <sup>4</sup>fadhilah.syafria@uin-suska.ac.id

<sup>1,2,3,4</sup>Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Diterima: 09 Maret 2023 | Direvisi: 15 April 2023 | Disetujui: 30 April 2023

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### Abstrak

Umpulan Klik (*Clickbait*) adalah judul berita yang dibuat oleh penulis bertujuan mendapatkan perhatian pembaca sehingga mereka tidak pernah melewatkannya. Judul clickbait biasanya unik, membingungkan, dan menggunakan kalimat yang dilebih-lebihkan untuk menarik pembaca agar mengklik tautan. Namun, headline clickbait yang terlihat sangat menarik seringkali tidak sesuai dengan informasi yang ada di headline dan isi beritanya, sehingga dapat berujung pada penyebaran berita bohong dan hoax. Maka dilakukan klasifikasi judul berita *clickbait*, dalam penelitian ini, akan dilakukan klasifikasi pada judul berita *clickbait* menggunakan metode *Transformers*. Jumlah judul berita yang digunakan pada penelitian ini berjumlah 6632 judul berita. Proses klasifikasi judul berita pada penelitian ini meliputi: menngumpulkan data, memberi label data, praproses, EDA, dan klasifikasi menggunakan *transformers*. Nilai akurasi terbaik yang diperoleh pada penelitian ini sebesar 63% dengan hasil precision 0.63 dan recall 1 dengan menggunakan pembagian data 10% : 90%. Dan pada penelitian juga berhasil mengklasifikasi judul berita clickbait dan non-clickbait dengan menggunakan pemrograman python dengan memasukan data uji sebanyak 6 judul berita *clickbait* dari hasil pengujian didapat 3 judul berita yang termasuk *clickbait* dan 3 tidak *clickbait*.

**Kata kunci:** Berita; *Clickbait*; *Headline*; *Hoax*; *Transformers*

## *Clickbait Classification Using Transformers*

### Abstract

*Clickbait is a news title created by the author with the aim of attracting the getting to get readers so they Never miss a headline. Clickbait headlines are typically quirky, confusing, and use exaggerated sentences to entice readers to click on links. However, clickbait headlines that look very attractive often do not match the information in the headlines and the content of the news, which can lead to the spread of fake news and hoaxes. Then classification of clickbait news titles is carried out, for this research, clickbait classification was carried out for news titles will be carried out using the Transformers method. The number of news titles used in this study amounted to 6632 news titles. The process of classification of news titles in this study includes: collecting data, labeling data, preprocessing, EDA, and classification using transformers. The best accuracy value obtained in this study was 63% for precision of 0.63 and recall of 1 using a data division of 10%: 90%. And the research also succeeded in classifying clickbait and non-clickbait news titles using Python programming by entering test data for 6 clickbait news titles. From the test results, 3 news titles were clickbait and 3 were not.*

**Keywords:** News; *Clickbait*; *Headline*; *Hoax*; *Transformers*

## 1. PENDAHULUAN

Teknologi semakin berkembang, maka dengan perkembangan teknologi membuat orang – orang dapat mengakses informasi lebih mudah. Sebelum teknologi berkembang untuk mendapatkan informasi seperti berita didapat dari media cetak [1]. Sekarang membaca berita tidak hanya dilakukan dengan membaca koran dan majalah, tetapi dapat membaca melalui berita online [2]. Informasi berita kadang tidak sesuai dengan judul berita online, judulnya menjebak (*Clickbait*) yang dimanfaatkan untuk menaikkan *traffic* pengunjung [3]. pembaca sehingga mereka tidak pernah melewatkannya. Judul clickbait biasanya unik, membingungkan, dan menggunakan kalimat yang dilebih-lebihkan untuk menarik pembaca agar mengklik tautan. Namun, headline clickbait yang terlihat sangat menarik sering kali tidak sesuai dengan informasi yang ada di headline dan isi beritanya, sehingga dapat berujung pada penyebaran berita bohong dan hoax[4].

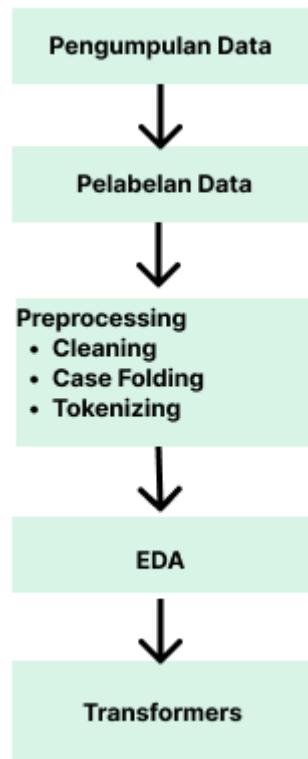
Clickbait adalah sesuatu untuk memancing seseorang dari judul berita untuk melihat konten yang dibuat dengan membuat judul menarik dan membuat penasaran pembaca, tidak semua pembaca membuka konten berita dan hanya mengambil kesimpulan dari judul yang tertera dan membagikannya kepada orang lain. *Clickbait* merupakan istilah yang menggambarkan judul berita yang akan memprovokasi pengguna untuk membuka link konten berita dengan menggunakan konten yang provokatif dan menarik. Beritanya dengan sengaja menahan informasi utama untuk memahami apa isi artikel tersebut [5]. Salah satu cara termudah agar mendapatkan lebih banyak pelanggan [6]. Strategi menggunakan judul *clickbait* tentu saja dapat mendapatkan keuntungan bagi penyedia berita *online*. Tetapi implementasi judul berita *clickbait* tidak sesuai dengan standar dalam jurnalisme dimana dalam menampilkan berita hanya mengutamakan pendapatan dan menlewatkan judul yang selaras dengan isi berita untuk publikasikan [7]. Berita *clickbait* mengandung kalimat utama yang berisi kata – kata hiperbolik tetapi sebagian besar tidak mengungkapkan informasi utama. Mungkin berisi kontroversial tetapi tidak mengungkapkan informasi utama apapun [8]. *Clickbait* yang terjadi membuat pembaca kesusahan mengklasifikasikan berita *clickbait* atau berita *non-clickbait*. Pembaca penasaran terhadap judul – judul berita *clickbait* akan terpancing untuk mengklik tautan situs berita dan hal itu berpengaruh pada penyedian situs [9]. Berita online yang diunggah pada media elektronik atau situs berita daring tentang *clickbait* dan *non-clickbait*. Meskipun beberapa berita *clickbait* tapi hal itu tentu saja membuat pembaca bingung untuk membedakan apakah judul berita termasuk *clickbait* atau bukan *clickbait*. Contoh Clickbait yaitu “TXT Segera Comeback, Soobin Akui Gatal Ingin Pamer Spoiler” pada judul diatas pembaca akan penasaran dengan *spoiler* dari *comeback* yang isinya tidak ada info tentang *spoiler* comebacknya, contoh lain yaitu “Istana Snggah Jika Jokowi Tak Peka Pada Korban Kabut Asap Setelah ‘Mejeng’ Bareng Jan Ethes” pada judul berita ini pembaca akan dibuat penasaran mengapa saat bencan presiden Jokowi mejeng dengan Jan Ethes contoh lainnya “Apa Sebenarnya Alasan Lesti Kejora Mencabut Laporan KDRT Pada Rizky Billar?” pada judul berita tersebut pembaca akan dibuat penasaran apa sebenarnya alasan dari Lesti Kejora mencabut laporan KDRT dan melihat berita secara keseluruhan pada hal isi berita tidak ada menjelaskan yang tertera di judul berita. Oleh karena itu, mengklasifikasikan judul berita *clickbait* dan *non-clickbait* adalah hal yang penting dilakukan. Hal ini dapat dilihat dari penelitian tentang klasifikasi judul berita *clickbait* yang terus berkembang. Salah satu penelitian yang membahas mengenai klasifikasi judul berita pada *clickbait* yaitu penelitian yang dilakukan oleh Muhammad Noor Fakhruzzaman, Sa'idah Zahrotul Jannah, Ratih Ardiati Ningrum, Indah Fahmiyah, dengan judul *Headline Detection in Indonesian News Sites using Multilingual Bidirectional Encoder Representations from Transformers (M-BERT)*. Total 6632 judul sebagai dataset pelatihan, pengklasifikasi dilakukan dengan sangat baik. Dievaluasi dengan validasi silang 5 kali, memiliki skor akurasi 0,914, skor f1 0,914, skor presisi 0,916, dan ROC-AUC sebesar 0,92 [8].

Berdasarkan penelitian yang dilakukan oleh Vijayasaradhi Indurthi, Bakhtiyar Syed, Manish Gupta, Vasudeva Varma, dengan judul *Predicting Clickbait Strength in Online Social Media*. Pada penelitian ini dilakukan beberapa kali percobaan dan berhasil melakukan penelitian dengan tingkat teratas dalam penelitian clickbait [10]. Penelitian lainnya yaitu klasifikasi menggunakan metode Naïve Bayes dengan hasil berkisar 56% tergolong sangat rendah dari 16 kali percobaan yang telah dilakukan sehingga perlu dikaji ulang pemilihan fitur, dan formula yang bagus untuk mendeteksi artikel *clickbait* [3]. Kemudian menggunakan metode KNN dimana hasil tertinggi dengan  $k=11$  kemudian menggunakan skenario 1 untuk pembagian data 800 data latih dan 200 data uji yang mendapatkan akurasi sebesar 71%, *Precision* 72%, dan *Recall* 71% [9]. Kemudian penelitian menggunakan metode SVM dan RF dalam bahasa Nepal mendapatkan perolehan akurasi 95,03% [6]. Dari penjelasan penelitian diatas perbedaan pada penelitian ini akan dilakukan penambahan rekaya fitur seperti panjang judul berita, pertanyaan, seruan, dan awalan dengan angka.

Dalam penelitian ini, akan dilakukan klasifikasi pada judul berita *clickbait* menggunakan metode *Transformers*. Metode *Transformers* adalah model *deep learning* yang dipublikasikan pertama kali pada jurnal Attention is all you need tahun 2017. *Transformers* yaitu model transduksi pertama yang mengandalkan sepenuhnya pada pelatihan diri sendiri untuk menghitung penggambaran input dan outputnya. *Transformers* merupakan struktur encoder-decoder, encoder termasuk gabungan lapisan encoding yang memenuhi proses input secara berulang – ulang satu per satu sedangkan decoder termasuk sekumpulan lapisan encoding yang melakukan hal sama pada output encoder [11]. Pada mulanya metode *Transformers* digunakan untuk melatih dataset teks untuk bahasa Inggris, sekarang sudah ada menggunakan bahasa Indonesia salah satunya yaitu pada penelitian [12]. Sehingga dalam penelitian ini menggunakan metode *Transformers* untuk mendapatkan hasil klasifikasi judul berita *clickbait* dan *non-clickbait* untuk bahasa Indonesia.

## 2. METODOLOGI PENELITIAN

Metodologi penelitian ini adalah rujukan dalam melaksanakan penelitian disusun secara ilmiah sebagai pedoman untuk melakukan penelitian, hal tersebut digunakan selama penelitian, tujuannya adalah untuk mendapatkan hasil yang memenuhi harapan peneliti. Berikut merupakan alur penelitian :



Gambar 1. Alur Penelitian

### 2.1 Pengumpulan Data

Judul berita diperoleh CLICK-ID: *A novel dataset for Indonesian clickbait headlines*, dataset ini berasal dari 12 portal berita daring [13]. Jumlah datanya sebanyak 8613 judul berita kemudian data akan diseimbangkan menjadi 6632 berita yang akan dibagi menjadi 2 kategori yaitu, 3316 berita *Clickbait*, dan 3316 berita *Non-Clickbait*. Data *Clickbait* dilabeli dengan 1 dan *Non-Clickbait* dilabeli dengan 0.

Tabel 1. Dataset Pelabelan *Clickbait* dan *Non-Clickbait*

Judul Berita	Label
Masuk Pencarian Pilwalkot Medan, Menantu Jokowi Bertemu DPW NasDem di Sumut	0
Viral! Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda	1

### 2.2 Pelabelan Data

Dataset yang digunakan berisikan berbagai judul berita yang diperoleh dari CLICK-ID: *A novel dataset for Indonesia clickbait headlines*. Dari data yang diperoleh data belum memiliki variabel atau karakteristik karena data hanya terdiri dari data clickbait dan non- clickbait jadi dilakukan pelabelan manual yang terdiri beberapa karakteristik bahasa yang biasanya digunakan pada penulisan *Clickbait*. Pada penelitian ini menggunakan 4 karakteristik pola penulisan yaitu Panjang Judul, Pertanyaan, Seruan, dan Penekanan angka diawali karakteristik penulisan yang biasanya digunakan untuk judul berita *Clickbait* [7] yang perbedaannya dijabarkan pada Tabel 2

Tabel 2. karakteristik Penulisan *Clickbait*

Pola Penulisan	Karakteristik	Karakteristik
	<i>Clickbait</i>	<i>Non-Clickbait</i>
Panjang judul	Antara 10 sampai 13 kata	Sekitar 7 kata
Pertanyaan	Terdapat tanda baca ‘?’ atau kaliamat pertanyaan	Tidak menggunakan tanda baca ‘?’ atau kalimat pertanyaan
Seruan	Terdapat tanda baca ‘!’ atau kalimat seruan	Tidak menggunakan tanda baca ‘!’ atau kalimat seruan
Angka di awal	Biasanya menggunakan angka diawal judul berita	Angka pada awal judul berita tidak menjadi prioritas

Kemudian Tabel 3 di bawah ini merupakan contoh data judul berita dan diberi label secara manual.

Tabel 5. Dataset Judul Berita *Clickbait* dan *Non-Clickbait*

Judul Berita	Label			Panjang Judul Berita
	Pertanyaan	Seruan	Awalan angka	
Masuk Pencarian Pilwalkot Medan, Menantu Jokowi Bertemu DPW NasDem di Sumut	0	0	0	10
Viral! Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda	0	1	0	10

Jadi pada table diatas jika pada judul berita terdapat pertanyaan, seruan atau awalan angka maka akan dilabeli dengan 1 jika tidak dilabeli dengan 0. Kemudian juga pada label seruan jika judul berita memiliki unsur seruan maka akan dilabeli dengan 1 jika tidak dilabeli dengan 0, untuk awalan angka itu jika judul berita diawali dengan angka maka akan dilabeli dengan 1 jika tidak dilabeli dengan 0 terakhir pelabelan pada panjang judul berita ini akan dilakukan perhitungan manual jumlah kata yang terdapat pada judul berita.

### 2.3 Preprocessing

#### a. Cleaning

Proses *Cleaning* merupakan proses menghapus kata yang tidak termasuk dalam kata alfabet ataupun angka contohnya ‘!’, ‘#’, ‘%’, ‘&’, ‘\$’ dan lain-lain [14]. Proses *cleaning* dapat dilihat dalam Tabel dibawah:

Tabel 6. Proses *Cleaning*

Judul Berita: Viral! Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda
Setelah proses: Viral Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda

#### b. Case Folding

Merupakan merubah karakter huruf sehingga menjadi huruf kecil semuanya [14]. Proses case folding dapat dilihat dalam tabel dibawah:

Tabel 7. Proses *Case Folding*

Judul berita: Viral! Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda
--

---

Setelah proses:  
viral di bekasi driver ojol antar pesanan makanan pakai sepeda

---

### c. Tokenizing

Merupakan merubah kalimat menjadi kata dimana membagian teks bisa berupa kalimat paragraf [15]. Proses *tokenizing* dapat dilihat dalam tabel dibawah:

Tabel 8. Proses *Tokenizing*

---

Judul berita: Viral! Di Bekasi Driver Ojol Antar Pesanan Makanan Pakai Sepeda
Setelah proses: [viral, di, bekasi, driver, ojol, antar, pesanan, makanan, pakai, sepeda]

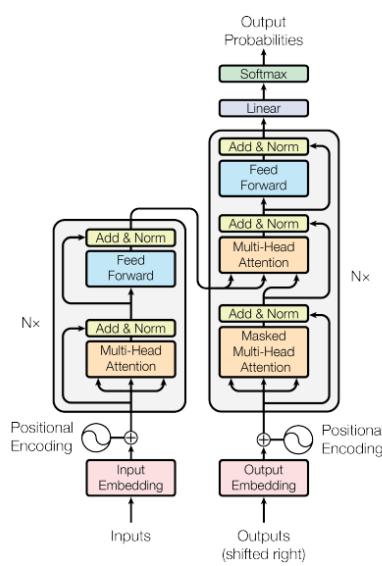
---

### 2.4 (Exploratory Data Analysis) EDA

Exploratory Data Analysis (EDA) yaitu prosedur menelaah dan mempresentasikan data yang bertujuan untuk mengartikan data dengan baik [16]. Proses ini bertujuan untuk memvisualisasikan frekuensi kata untuk setiap kelas, memvisualisasikan distribusi fitur yang direkayasa dan relevansinya pada setiap kelas, dan distribusi target sebagian besar berimbang.

### 2.5 Transformers

*Transformers* adalah model *deep learning* yang memperkenalkan pertama kali pada paper Attention is all you need pada tahun 2017. *Transformers* digunakan dalam bidang *Neural Language Processing* (NLP). *Transformers* merupakan model transduksi pertama yang mengandalkan sepenuhnya pada pelatihan diri sendiri buat menghitungkan penggambaran input dan outputnya. Model transduksi urutan dominan didasarkan pada kompleks berulang atau jaringan saraf convolutional yang mencakup encoder dan decoder [11]. *Transformers* membutuhkan sistem untuk melatih, menganalisis, mengklasifikasikan pada berbagai platform [17]. Pada penelitian ini menggunakan encoder. Arsitektur ini dibangun di atas mekanisme kesadaran diri tanpa bergantung pada jaringan yang berulang (GRU, LSTM). Transformer adalah arsitektur yang mengubah satu urutan ke urutan lainnya menggunakan encoder dan decoder. Model arsitektur transformator dapat dilihat pada gambar berikut.



Gambar 2. Model *Transformers*

Bagian utama dari trans adalah unit dari mekanisme kesadaran diri berkepala banyak. Trafo melihat representasi encoder dari input sebagai pasangan kunci-nilai (K,V), n-dimensi. Kedua pembuat enkode disembunyikan sehubungan dengan NMT. Untuk dekoder, keluaran sebelumnya dikompresi menjadi kueri (m-dimensional Q), dan keluaran teks dihasilkan dengan mencocokkan kueri ini dengan kunci dan nilai.. Pada penelitian ini dilakukan menggunakan encoder. Kemudian pada tahap fine-tunning model parameter yang digunakan dalam klasifikasi judul berita adalah batch size = 64, learning rate = 1e-6, dan epoch = 20.

## 2.6. Pembagian Data

Pada Tahapan pembagian data ini dilakukan proses membagi data berita yang berjumlah dataset. Data berita tersebut akan dibagi menjadi 2 bagian yaitu data latih dan data uji dengan menggunakan 3 skenario perbandingan data yaitu 70% data latih dan 30% data uji, 80% data latih dan 20% data uji, lalu 90% data latih dan 10% data uji. Untuk pembagian lebih jelas dapat dilihat melalui tabel 4.1 dibawah ini.

Tabel 3. *Pembagian Data*

Data Berita	Kelas	Pembagian Data Latih dan Data uji					
		Training	Testing	Training	Testing	Training	Testing
		70%	30%	80%	20%	90%	10%
6632	<i>Clickbait</i>						
	<i>Non-Clickbait</i>						
Jumlah							

## 2.7. Confusion Matrix dan Clasification Report

Pengujian ini digunakan untuk mengetahui akurasi dengan menggunakan metode *Trasformers* yang digunakan untuk klasifikasi judul berita clickbait, confusion matrix terdiri dari informasi yang membandingkan hasil klasifikasi dengan mengukur nilai kebenaran [18]. Tabel Confusion Matrix yaitu suatu metode yang digunakan pada perhitungan akurasi, recall, precision, dan error rate [19]. Confussion matrix memiliki empat susunan yaitu true positives, true negatives, false positives, dan false negatives. Confusion matrix ini memungkinkan kita untuk mendapatkan nilai precision, recall dan, F1-score [2]. Pada tabel matrix baris merupakan nilai kelas yang sebenarnya, sedangkan kolom adalah nilai prediksi [20]. Kombinasi tersebut tertera pada table dibawah ini.

Tabel 2. *Confusion Matrix*  
Label Sebenarnya

Label Prediksi		
	<i>Clickbaik</i>	<i>Non-Clickbait</i>
<i>Clickbait</i>	<i>True Positive (TP)</i>	<i>False Positive(FP)</i>
<i>Non-Clickbait</i>	<i>False Negative (FN)</i>	<i>True Negative(TN)</i>

*Classification Report* adalah metode untuk menghitung akurasi dari algoritma klasifikasi. Data dari confusion matrix akan digunakan dalam perhitungan ini. Berikut nilai-nilai yang digunakan dalam *classification report*:

1. *Accuracy* : kesesuaian terhadap hasil prediksi pengujian dengan nilai aktual

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision* : kesesuaian data yang diminta dengan data yang ditampilkan

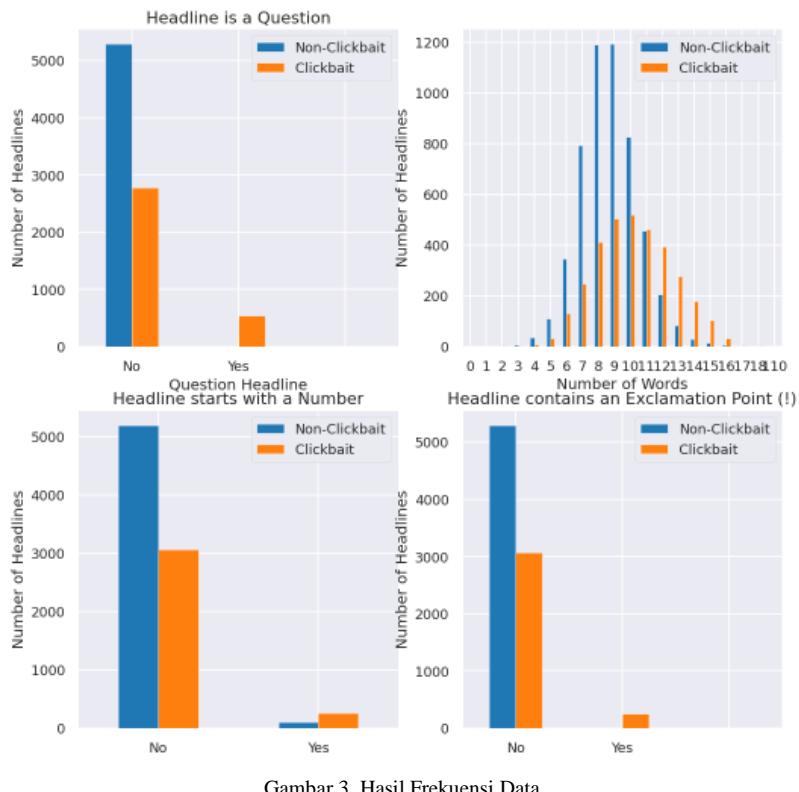
$$\text{Precision} = \frac{TP}{TP + FP}$$

3. *Recall* : jumlah kesesuaian data dari hasil percobaan berdasarkan sudut pandang kelas yang digunakan.

$$\text{Recall} = \frac{TP}{TP + FN}$$

### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data awal 8613 judul berita kemudian data di seimbangkan lalu dilakukan proses processing yang bertujuan untuk agar data yang digunakan dapat mengoptimalkan kinerja saat proses klasifikasi. Setelah melakukan proses processing dilakukan proses EDA yang hasilnya dapat dilihat pada gambar 3 dibawah:

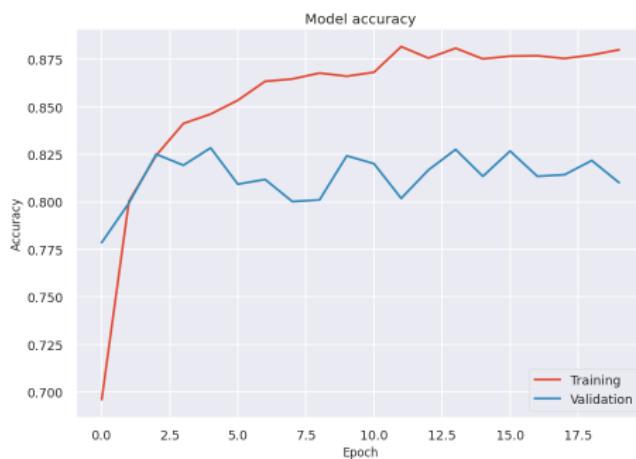


Gambar 3. Hasil Frekuensi Data

Jadi pada gambar diatas ini untuk mengetahui presentasi pada data dimana data dengan jumlah *clickbait* dan *non-clickbait*

#### 3.1 Implementasi

Pembagian data pada penelitian ini ada 3 yaitu 70%:30%, 80%:20%, dan 90%:10% hasil dari pengujian ini dapat dilihat dari grafik dibawah:



Gambar 5. Hasil Grafik Pembagian Data 70%:30%

Pada grafik diatas pembagian data 70% data latih dan 30% data uji dimana data training untuk akurasi dimulai dibawah 0,7 dan akurasi akhir di 0,87. Sedangkan untuk validation akurasinya dimulai diatas 0,77 dan akurasi akhirnya diatas 0,8.



Gambar 6. Kata Hasil Grafik Pembagian Data 80%:20%

Pada grafik diatas pembagian data 80% data latih dan 20% data uji dimana data training untuk akurasi dimulai diatas 0,7 dan akurasi akhir diatas 0,87. Sedangkan untuk validation akurasinya dimulai diatas 0,77 dan akurasi akhirnya diatas 0,82.



Gambar 7. Kata Hasil Grafik Pembagian Data 90%:10%

Pada grafik diatas pembagian data 90% data latih dan 10% data uji dimana data training untuk akurasi dimulai dibawah 0,74 dan akurasi akhir diatas 0,87. Sedangkan untuk validation akurasinya dimulai diatas 0,8 dan akurasi akhirnya diatas 0,82.

Dari gambar diatas Pengujian menggunakan metode Transformers untuk mengklasifikasikan judul berita *Clickbait*. Untuk data yang digunakan 90%:10% dimana dari data *tranning* yaitu 90% dan dari data *testing* yaitu 10%. Semua data sebanyak 8613 data yang terbagi menjadi 3316 *Clickbait* dan 5297 *Non-clickbait*. Tahap awal data akan diberi label seperti pada table 5. Setelah pemberian label data akan di seimbangkan sehingga data *Clickbait* menjadi 3316 dan data *Non-clickbait* menjadi 3316 data. Selanjutnya dilakukan preprocessing data yang dapat dilihat pada table 6 proses *Cleaning*, table 7 proses *case holding* dan table 8 proses *tokenizing*. Berikutnya proses Exploratory Data Analysis (EDA) yang dapat dilihat pada gambar 4. Kemudian hasil dari klasifikasi judul berita menggunakan Transformers dapat dilihat dari Confusion Matrix digambar dibawah:

	precision	recall	f1-score	support
0	0.63	1.00	0.77	539
1	0.00	0.00	0.00	323
accuracy			0.63	862
macro avg	0.31	0.50	0.38	862
weighted avg	0.39	0.63	0.48	862

Gambar 8. Confusion Matrix

Kemudian hasil dari Klasifikasi judul berita menggunakan Transformers dengan hasil menngunakan python sebagai berikut:

Judul Berita	Label
Mengaku Jadi Korban Pemeriksaan, Ini 4 Fakta Menarik tentang Barbie Nouva	Clickbait
Polisi: Syarat Restorative Justice Kasus KDRT Billar Sudah Terpenuhi	Not A Clickbait
Apa Alasan Lesti Kejora Cabut Laporan KDRT Rizky Billar?	Clickbait
Top 3 Berita Hari Ini: Warganet Bongkar Koleksi Tas Hermes Istri Rafael Alun Trisambodo, Total Harganya Capai Miliaran Rupiah	Clickbait
4. Waspadai! Ada Modus Penipuan Berkedok Aplikasi E-Tilang	Not A Clickbait
Kegiatan Akhir Pekan Jokowi: Cukur Rambut Hingga Kuliner	Not A Clickbait

#### KESIMPULAN

Kesimpulan merupakan hasil akhir yang dicapai dari sebuah penelitian. Kesimpulan menggambarkan keberhasilan dari penelitian yang dilakukan. Setiap penelitian memiliki hasil, meskipun kadang kala tidak sesuai dengan tujuan awal penelitian. Adapun yang didapatkan pada proses penelitian yaitu dengan menggunakan metode *Transformers* dapat digunakan untuk proses klasifikasi judul berita *Clickbait* dan nilai akurasi terbaik yang diperoleh pada penelitian ini adalah 63% dengan hasil precision 0.63 dan recall 1 menggunakan pembagian data uji 10 % & serta data latih 90% dan juga menghasilkan klasifikasi judul berita *clickbait* dengan memasukkan data ke program python dari 6 data uji 3 diantaranya termasuk *clickbait* dan 3 *non-clickbait*. Saran untuk penelitian selanjutnya menggunakan data latih yang lebih besar dengan data yang seimbang agar akurasi yang dihasilkan lebih baik. Melakukan pengembangan penelitian dengan mengimplementasikan model *Transformer* seperti XLNet, ALBERT, RoBERTa, dan DistilBERT.

#### DAFTAR PUSTAKA

- [1] B. W. Rauf, S. Raharjo, and H. Sismoro, “Deteksi Clickbait dengan Sentence Scoring Based On Frequency di Detik.Com,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 247–252, 2020.
- [2] A. D. Rendragraha, M. A. Bijaksana, and A. Romadhony, “Pendekatan Metode Transformers untuk Deteksi Bahasa Kasar dalam Komentar Berita Online Indonesia,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3385–3395, 2021.
- [3] A. F. Yavi, “Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [4] Y. D. Hadiyat, “Clickbait on Indonesia Online Media,” *J. Pekommas*, vol. 4, no. 1, p. 1, 2019.
- [5] A. Anand, T. Chakraborty, and N. Park, “We used neural networks to detect clickbaits: You won’t believe what happened next!,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10193 LNCS, pp. 541–547, 2017.
- [6] S. Ram, S. Prasad, and T. Bahadur, “Detecting Clickbaits on Nepali News using SVM and RF,” vol. 8914, pp. 140–146, 2021.
- [7] O. L. Pramesti, “Clickbait Headline in News of Online Prostitution Case,” *J. Pekommas*, vol. 5, no. 1, p. 59, 2020.
- [8] 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.
- [9] R. Sagita, U. Enri, and A. Primajaya, “Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN),” *JOINS (Journal Inf. Syst.)*, vol. 5, no. 2, pp. 230–239, 2020.
- [10] V. Indurthi, B. Syed, M. Gupta, and V. Varma, “Predicting Clickbait Strength in Online Social Media,” pp. 4835–4846, 2021.
- [11] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [12] S. Cahyawijaya *et al.*, “IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation,” *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021.
- [13] A. William and Y. Sari, “CLICK-ID: A novel dataset for Indonesian clickbait headlines,” *Data Br.*, vol. 32, p. 106231, 2020.
- [14] A. Awalina, F. A. Bachtiar, F. Utaminingsrum, and P. Korespondensi, “Perbandingan Pretrained Model Transformer Pada Deteksi Ulasan Palsu Comparison Of Pretrained Transformer Models On Spam Review Detection,” vol. 9, no. 3, pp. 597–604, 2022.
- [15] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, “Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 91–97, 2021.
- [16] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, “Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 23–27, 2022.
- [17] T. Wolf *et al.*, “Transformers: State-of-the-Art Natural Language Processing,” pp. 38–45, 2020.
- [18] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, “Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, 2022.
- [19] A.- Arini, L. K. Wardhani, and D.- Octaviano, “Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden,” *Kilat*, vol. 9, no. 1, pp. 103–114, 2020.
- [20] P. Nima, “Automatic Filtration of Misleading Youtube Videos using Data Mining Automatic Filtration of Misleading Youtube Videos using Data Mining Techniques MSc Research Project Masters in Data Analytics Prateek Nima Student ID : x18114610 School of Computing Natio,” no. September 2019, 2020.
- [1] B. W. Rauf, S. Raharjo, and H. Sismoro, “Deteksi Clickbait dengan Sentence Scoring Based On Frequency di Detik.Com,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 247–252, 2020.
- [2] A. D. Rendragraha, M. A. Bijaksana, and A. Romadhony, “Pendekatan Metode Transformers untuk Deteksi Bahasa Kasar dalam Komentar Berita Online Indonesia,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3385–3395, 2021.
- [3] A. F. Yavi, “Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [4] Y. D. Hadiyat, “Clickbait on Indonesia Online Media,” *J. Pekommas*, vol. 4, no. 1, p. 1, 2019.
- [5] A. Anand, T. Chakraborty, and N. Park, “We used neural networks to detect clickbaits: You won’t believe what happened next!,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10193 LNCS, pp. 541–547, 2017.



- [6] S. Ram, S. Prasad, and T. Bahadur, "Detecting Clickbaits on Nepali News using SVM and RF," vol. 8914, pp. 140–146, 2021.
- [7] O. L. Pramesti, "Clickbait Headline in News of Online Prostitution Case," *J. Pekommas*, vol. 5, no. 1, p. 59, 2020.
- [8] 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.
- [9] R. Sagita, U. Enri, and A. Primajaya, "Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN)," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 2, pp. 230–239, 2020.
- [10] V. Indurthi, B. Syed, M. Gupta, and V. Varma, "Predicting Clickbait Strength in Online Social Media," pp. 4835–4846, 2021.
- [11] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [12] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021.
- [13] A. William and Y. Sari, "CLICK-ID: A novel dataset for Indonesian clickbait headlines," *Data Br.*, vol. 32, p. 106231, 2020.
- [14] A. Awalina, F. A. Bachtiar, F. Utaminingsrum, and P. Korespondensi, "Perbandingan Pretrained Model Transformer Pada Deteksi Ulasan Palsu Comparison Of Pretrained Transformer Models On Spam Review Detection," vol. 9, no. 3, pp. 597–604, 2022.
- [15] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 91–97, 2021.
- [16] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, "Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 23–27, 2022.
- [17] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," pp. 38–45, 2020.
- [18] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, 2022.
- [19] A.- Arini, L. K. Wardhani, and D.- Octaviano, "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden," *Kilat*, vol. 9, no. 1, pp. 103–114, 2020.
- [20] P. Nima, "Automatic Filtration of Misleading Youtube Videos using Data Mining Automatic Filtration of Misleading Youtube Videos using Data Mining Techniques MSc Research Project Masters in Data Analytics Prateek Nima Student ID : x18114610 School of Computing Natio," no. September 2019, 2020.

