

# Implementasi BERT pada Analisis Sentimen Ulasan Destinasi Wisata Bali

Tristan Bey Kusuma<sup>a1</sup>, I Komang Ari Mogi<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Bali, Indonesia

<sup>1</sup>tristanbeykusuma@email.com

<sup>2</sup>arimogi@unud.ac.id

## Abstract

*In recent years, the contribution of the tourism sector in Bali has increased significantly. The tourism sector has an important role as a source of foreign exchange earnings, and can encourage national economic growth. With the digital age, online opinion are increasingly vital to the growth of Indonesian tourism internationally. Public opinion and reviews on these tourist destinations can be used to identify new tourist destinations which are gaining traction and are in demand. Which is why it will be important to leverage these positive or negative opinions to acquire interesting and vital information on these tourist destinations for further use. One such method to acquire such information are through sentiment analysis to determine whether the a review's attitude towards a particular tourist destination or experience is positive, negative, or neutral. This study aims to use IndoBERT, a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model for the Indonesian language. This model is trained using a masked language modeling (MLM) objective and next sentence prediction (NSP) objective. This study also compares two different optimizers with a weight decay fix, AdamW and AdaFactor. The results show that sentiment analysis using the IndoBERT model with the AdamW optimizer reaches 97% accuracy and AdaFactor reaches 98,2% accuracy.*

**Keywords:** Sentiment Analysis, Review, Natural Language Understanding (NLU), Natural Language Generation (NLG), Natural Language Processing (NLP), BERT, AdamW, Transformer

## 1. Pendahuluan

Kontribusi industri pariwisata telah berkembang pesat dalam tahun-tahun terakhir ini. Proporsi industri pariwisata terhadap keseluruhan ekspor barang dan jasa, yang meningkat secara dramatis dari 10% pada tahun 2005 menjadi 17% pada tahun 2012, menunjukkan hal ini. Kenaikan kontribusi ini terutama diakibatkan oleh meningkatnya jumlah wisatawan lokal dan internasional serta meningkatnya investasi di industri pariwisata [1]. Naiknya industri pariwisata ini dapat memicu peningkatan keuntungan devisa, menciptakan lapangan kerja, dan mempromosikan perluasan bisnis pariwisata yang terkait. Hal ini tentunya menarik negara-negara lain untuk membangun industri pariwisata mereka. Melalui berbagai mekanisme, seperti meningkatnya keuntungan devisa dan iming-iming investasi asing, pariwisata secara signifikan membantu pertumbuhan ekonomi [2]. Hal ini khususnya di provinsi Bali.

Provinsi Bali, yang oleh kegiatan pariwisatanya telah dikenal sebagai salah satu kantong devisa Indonesia, hal tersebut menandakan bahwa sektor pariwisata perlu mendapat perhatian sebagai sektor utama dalam mendukung perekonomian makro Bali dan perekonomian Indonesia pada umumnya [3]. Namun, pandemi Covid-19 telah mengakibatkan perlambatan bahkan penurunan pertumbuhan ekonomi pariwisata di yang drastis. Pertumbuhan ekonomi yang mengacu Produk Domestik Bruto (PDB) sektor akomodasi dan pariwisata pada kuartal II-2020 tercatat berkurang sebanyak -33,99% dan - 24,9%. Dengan kontribusi yang besar, turunnya PDB sektor akomodasi dan transportasi juga menimbulkan pertumbuhan negatif pada ekonomi Bali. Pada kuartal II-200 PCB Bali tercatat sebesar -11,06% [4].

Namun baru-baru ini, setelah restriksi dan larangan perjalanan telah dibuka, wisatawan dari mancanegara akhirnya dapat kembali berwisata ke Bali. Hal ini tentunya menyebabkan suatu permintaan yang naik. Permintaan yang dihasilkan oleh kegiatan pariwisata dalam bentuk permintaan konsumen dan investasi, akan mendorong ekspansi atau pertumbuhan. Naiknya permintaan wisata ini dapat terlihat dalam ulasan dan opini dalam media sosial maupun situs wisata online. Sentimen atau pendapat ini dapat berupa positif, negatif, maupun netral. Maka dari itu, untuk melakukan analisis atas berbagai opini yang berbeda dan mengkategorikannya kita dapat memanfaatkan machine learning. *Sentiment analysis* atau analisis sentiment adalah suatu proses NLP (*Natural Language Processing*) untuk mengidentifikasi dan mengkategorikan pendapat yang diungkapkan dalam sepotong teks, terutama untuk menentukan apakah sikap penulis terhadap topik tertentu adalah positif, negatif, atau netral. Analisis sentiment ini kemudian dapat dijadikan dasar dari sebuah sistem rekomendasi dan visualisasi data. Salah satu metode machine learning untuk melakukan hal ini adalah model transformer.

BERT atau *Bidirectional Encoder Representations from Transformers* adalah model *deep learning* di mana setiap elemen *output* terhubung ke setiap elemen *input*, dan bobot di antara mereka dihitung secara dinamis berdasarkan koneksi tersebut. Secara historis, model bahasa hanya dapat membaca input teks secara berurutan, baik kiri ke kanan atau kanan ke kiri. Tetapi model seperti ini tidak dapat melakukan keduanya secara bersamaan. BERT berbeda karena dirancang untuk membaca dua arah sekaligus. Menggunakan kemampuan dua arah ini, BERT telah di *pre-train* pada tugas-tugas NLP yaitu *Masked Language Modeling* dan *Next Sentence Prediction* [5]. Dalam pengembangan BERT ini, kita dapat memanfaatkan beberapa model seperti IndoBERT.

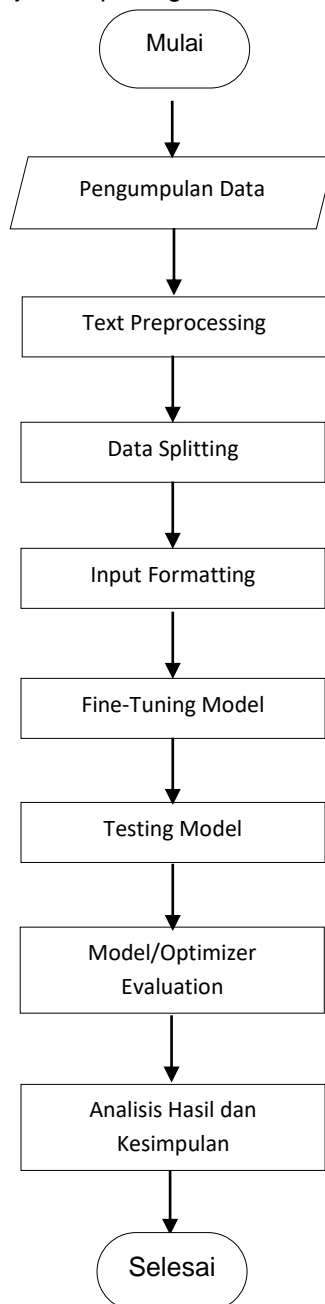
IndoBERT adalah model bahasa mutakhir untuk bahasa Indonesia berdasarkan model BERT. IndoBERT ini telah dilatih pada kumpulan dataset teks Bahasa Indonesia yang mencakup empat miliar kata. Pada penelitian IndoBERT, telah dibandingkan model IndoBERT tersebut secara ekstensif dengan berbagai *embedding* kata dan model *pre-trained* yang telah dilatih sebelumnya, seperti Multilingual BERT dan XLM-R untuk mengukur efektivitasnya. Hasil menunjukkan bahwa model IndoBERT tersebut mengungguli sebagian besar model lain yang ada. Model IndoBERT mengungguli model Multilingual pada 8 dari 12 tugas. Secara umum, model IndoBERT mencapai skor rata-rata tertinggi pada tugas klasifikasi. Model monolingual seperti IndoBERT mempelajari semantik tingkat sentimen yang lebih baik pada gaya bahasa sehari-hari dan formal daripada model multibahasa, meskipun ukuran model IndoBERT 40%-60% lebih kecil [5].

Dalam penelitian ini, penulis mengimplementasikan model IndoBERT untuk melakukan analisis sentiment terhadap data ulasan tempat wisata di Bali. Selain itu, penelitian ini juga akan membandingkan performa *optimizer* dalam melakukan fine-tuning model IndoBERT tersebut. Terdapat dua *optimizer* yang akan dibandingkan, AdamW dan AdaFactor. Dari hasil perbandingan ini diharapkan dapat memilih *optimizer* tepat untuk mengoptimalkan hasil performa IndoBERT tersebut.

## 2. Metode Penelitian

Penelitian ini bertujuan membandingkan metode klasifikasi sentimen terhadap ulasan tempat wisata di Bali menggunakan model transformer IndoBERT. Pemerintah dan stakeholder lainnya di bidang pariwisata membutuhkan model analisis sentimen ini sebagai solusi atas permasalahan prioritas pengembangan destinasi wisata melalui identifikasi obyek wisata yang diminati. Teknik *deep learning* yang disebut BERT ini memberikan hasil akurasi yang lebih besar daripada teknik transformer lainnya. BERT ini merupakan teknik *machine learning* untuk *pre-training* pemrosesan bahasa alami (NLP). Teknik IndoBERT dipilih karena berhasil menangkap dengan akurat makna semantik teks yang ditulis dalam bahasa Indonesia. IndoBERT merupakan sebuah model *pre-trained large* yang dilatih dengan sekitar 4 miliar korpus kata dari dataset Bahasa Indonesia khusus, terdiri dari lebih dari 20 Gigabyte data teks. Ini menghasilkan akurasi analisis sentiment bahasa Indonesia yang tinggi. Penelitian ini mencakup beberapa tahapan, yaitu pengumpulan data review teks dan rating melalui *web*

*crawling*, *preprocessing*, *splitting* dan *formatting* data, *fine-tuning*, *optimization*, dan evaluasi hasil. Diagram alur penelitian ditunjukkan pada gambar 1.



**Gambar 1.** Diagram Alur Penelitian

## 2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini bersumber dari *crawling* data pada website menggunakan tools *scraping*. Informasi tersebut dikumpulkan dari *review* tempat wisata di pulau Bali yang tertulis dalam bahasa Indonesia dan dimuat dari situs web *tripadvisor.com*. Data ulasan yang diambil merupakan ulasan *user* pada situs objek wisata. Sehingga data ini dapat dibagi menjadi ulasan bersifat negatif, positif, dan netral. Data tersebut berjumlah 908 dibagi menurut rating. Data tersebut kemudian disimpan dalam file *csv*. sehingga model analisis sentimen tersebut dapat dilatih. Model yang dibangun menggunakan *IndoBERT* ini akan

melakukan pelatihan serta uji hasil menggunakan data tersebut. Namun, sebelum data diolah, peneliti melakukan pembersihan data yang disebut dengan *pre-processing text*.

## 2.2. Preprocessing Text

Tujuan dari *text preprocessing* atau pra-pengolahan adalah untuk merapikan data teks ulasan untuk menghilangkan karakter, kata-kata, dan tanda baca yang tidak diinginkan dalam merepresentasikan makna teks tersebut serta membersihkan data agar dapat digunakan dengan mudah pada tahap pemrosesan selanjutnya. Tahapan *preprocessing* adalah sebagai berikut, diantaranya :

- a. *Case-Folding*  
*Case-folding* adalah tahapan konversi huruf kecil terhadap setiap huruf atau karakter pada sebuah teks. Dalam penelitian ini, *case-folding* digunakan ketika mengkonversi semua teks dalam *review* dataset menjadi huruf kecil [6].
- b. Tokenisasi  
Dokumen teks dibagi menjadi token melalui proses tokenisasi. Tokenisasi dapat membedah isi teks menjadi kata, frasa, simbol, atau komponen lain yang berguna [6].
- c. *Stopword Removal*  
*Stopword removal* adalah proses menghilangkan kata-kata yang dianggap tidak perlu. Penghapusan *stopword* bertujuan untuk menghilangkan kata-kata yang sering muncul tetapi tidak berkontribusi pada proses analisis data, hanya menyisakan kata-kata yang dianggap dapat menyampaikan makna teks secara akurat [6].
- d. *Punctuation Removal*  
*Punctuation removal* merupakan proses untuk menghilangkan simbol-simbol yang terdapat pada dokumen teks. Ini dapat dicapai dengan menghapus karakter yang tidak penting, termasuk tanda baca, angka, html, karakter, serta simbol lainnya [6].
- e. *Padding*  
*Padding* adalah penambahan sebuah token untuk mengubah setiap urutan teks agar memiliki panjang yang sama. Teks batch seringkali memiliki panjang yang berbeda, sehingga tidak dapat dikonversi ke tensor atau vektor. *Padding* adalah strategi untuk menambahkan token padding khusus untuk memastikan urutan yang lebih pendek akan memiliki panjang yang sama dengan urutan terpanjang dalam batch atau panjang maksimum yang diterima oleh model [7]. Dalam model BERT ini, panjang maksimum yang penulis gunakan adalah 100.

## 2.3. Data Splitting

*Splitting* atau pembagian dataset dilakukan agar bisa melihat performa model setelah dilatih. Proses pembagian data dibagi menjadi tiga proporsi, yaitu data *training* untuk pelatihan dan *fine-tuning* model, data *validation* digunakan untuk mengurangi *overfitting* dan noise pada pelatihan model, dan data *testing* untuk mengevaluasi keakuratan model setelah dilakukan pelatihan. Proporsi data tersebut dibagi dengan persentase 70% sebagai data pelatihan, 15% sebagai data validasi, dan 15% sebagai data pengujian dari total jumlah data. Sehingga, banyaknya data yang digunakan dalam *fine-tuning* awal berjumlah 636.

## 2.4. Input Formatting

Setiap model transformer berbeda tapi memiliki kesamaan dengan yang lain. Oleh karena itu sebagian besar model menggunakan input yang sama, Karena ini, maka kita harus melakukan *encoding* atau pemformatan pada dataset yang kita miliki agar sesuai dengan input model transformer. Ini disebut sebagai *input formatting* [9]. *Input formatting* ini terdiri dari tokenisasi,

vektorisasi, dan *mapping*. Tokenizer BERT akan mengubah *sequence* kalimat menjadi potongan kata, seperti pada tabel berikut.

Sequence	Tokens	Kalimat BERT
"Saya tidak suka!"	['Saya', 'tidak', 'suka', 'ini', '!']	

Kemudian, BertTokenizer.encode\_plus() digunakan untuk mengonversi urutan ke dalam format input untuk pengklasifikasi berbasis BERT nanti [8]. BertTokenizer.encode\_plus() ini akan mengembalikan kamus tiga objek:

1. *Input ID*, ini sesuai dengan bilangan bulat/urutan token di input
2. *Type Token ID*, Id ini menunjukkan nomor kalimat yang dimiliki oleh token. (BERT dapat mengambil hingga dua urutan sekaligus).
3. *Attention Mask*, mask ini menunjukkan token mana merupakan token yang sebenarnya dan mana yang merupakan token *padding* sehingga perhitungan *attention* akan mengabaikan token *padding*.

Saat *tokenizing* dan vektorisasi, kita dapat menentukan panjang maksimum dari setiap teks. Kemudian, setelah kita menggabungkan langkah untuk tokenisasi, vektorisasi WordPiece, penambahan token khusus, pemotongan ulasan lebih panjang dari panjang maksimal, dan melakukan mapping objek ID dan *attention mask* ke *dictionary* objek, kita dapat melihat hasil seperti berikut.

**Tabel 2.** Input Formatting Model BERT

Sequence : "Saya tidak suka ini!"		
Input ID	Token Type ID	Attention Mask
[101, 2123, 1005, 1056, 2066, 2009, 999, 102, 0, 0]	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	[1, 1, 1, 1, 1, 1, 1, 1, 0, 0]

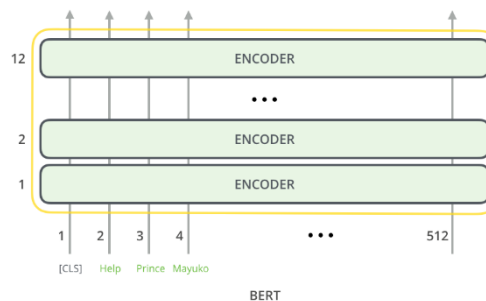
## 2.5. Fine-Tuning

Training dalam model BERT tidak dilakukan dari awal, namun pada model yang sudah dilatih sebelumnya. Model BERT yang telah dilatih sebelumnya telah menyimpan informasi *Language Generation dan Understanding* tentang bahasa Indonesia. Akibatnya, dibutuhkan lebih sedikit waktu untuk melatih model *fine-tuned* tersebut. Dengan *fine-tuning* lapisan bawah jaringan model yang telah dilatih secara luas hanya perlu disetel saat menggunakan *output*-nya sebagai fitur untuk pekerjaan analisis sentimen ini [5]. Selain itu, karena bobot yang telah dilatih sebelumnya, metode ini memungkinkan kita untuk menyempurnakan tugas kita pada kumpulan data yang jauh lebih kecil. Biasanya, model yang dibangun dari awal akan membutuhkan kumpulan data yang sangat besar untuk melatih jaringan kita ke akurasi yang baik, yang berarti banyak waktu dan komputasi harus dimasukkan ke dalam pembuatan kumpulan data. Dengan melakukan *tuning* pada IndoBERT, penulis dapat melatih model untuk kinerja yang baik pada jumlah data pelatihan yang jauh lebih kecil [8]. Dalam tahap *fine-tuning*, model IndoBERT

pertama kali dimulai dengan parameter yang dipelajari sebelumnya, dan semua parameter disetel dengan menggunakan data berlabel.

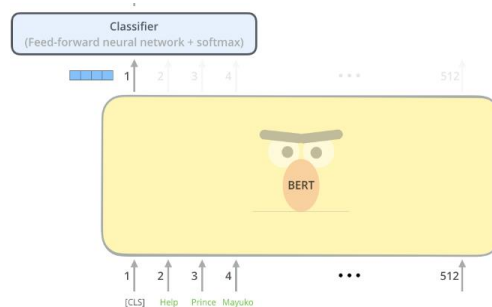
Selama *fine-tuning*, semua parameter akan disetel. BERT mengambil urutan kata sebagai input yang terus mengalir ke atas BERT stack. Setiap lapisan menerapkan *self-attention*, dan meneruskan hasilnya melalui jaringan *feed-forward*, dan kemudian menyerahkannya ke *encoder* berikutnya. Di *fine-tuning* ini pada tiap *sequence* akan ditambahkan token-token khusus. [CLS] adalah simbol khusus yang ditambahkan di depan setiap contoh input. Status tersembunyi terakhir yang sesuai dengan token ini digunakan sebagai representasi urutan agregat untuk tugas klasifikasi. Kemudian, [SEP] adalah token yang digunakan untuk memisahkan pasangan kalimat. Jika di satu *sequence* terdapat beberapa pasang kalimat maka akan digunakan token ini untuk memisahkan tiap kalimat tersebut [9].

**Gambar 2. BERT Model Input**



Pada output BERT, model hanya fokus pada output hanya dari posisi pertama (yang diberikan token [CLS]). Hasil token [CLS] ini dimasukkan ke dalam lapisan output untuk kegunaan klasifikasi. Jika kita memiliki lebih banyak label yang ingin diklasifikasikan, cukup mengubah jaringan pengklasifikasi untuk memiliki lebih banyak neuron keluaran yang kemudian melewati *softmax* untuk pemusatan training [9].

**Gambar 3. BERT Model Output**



## 2.6. Optimizer

*Adaptive moment estimation* (adam) merupakan salah satu algoritma yang dapat menggantikan prosedur stochastic gradient descent klasik untuk memperbaharui *weight network* berdasarkan data training secara iteratif. Cara kerja adam dapat digambarkan sebagai penggabungan sifat

terbaik dari dua ekstensi *stochastic gradient descent* yaitu *adaptive gradient algorithm* dan *root mean square propagation* dengan penggabungan tersebut adam mampu memberikan pengoptimalan suatu algoritma yang mampu menangani *sparse gradients* pada *noisy problem*. Dengan penggunaan teknik optimasi yang mampu menurunkan gradien, metode ini sangat efisien ketika bekerja pada data yang besar dan parameter yang besar. AdamW adalah metode optimasi stokastik yang memodifikasi implementasi tipikal *weight decay* pada Adam (Adaptive Moment Estimation). Regularisasi L2 pada Adam biasanya diimplementasikan dengan modifikasi di bawah ini dimana  $w_t$  adalah laju *weight decay* pada waktu  $t$  [10].

$$g_t = \nabla f(\theta_t) + \omega_t \theta_t \quad (1)$$

Sehingga AdamW menyesuaikan istilah *weight decay* untuk muncul di pembaruan atau *update gradient* [10].

$$x_t \leftarrow x_{t-1} - \alpha \frac{\beta_1 m_{t-1} + (1 - \beta_1)(\nabla f_t + \omega x_{t-1})}{\sqrt{v_t + \epsilon}}$$

(2)

Sedangkan, AdaFactor adalah sebuah metode optimasi stokastik berdasarkan adam yang mengurangi penggunaan memori sambil mempertahankan manfaat empiris dari adaptasi. Adafactor mengurangi penggunaan memori pertama-tama dengan mengusulkan untuk mengganti matriks gradien kuadrat yang dismoothing dengan aproksimasi peringkat rendah [11].

$$G_t = \nabla f(X_{t-1}) \quad (3)$$

$$R_t = \beta_2 R_{t-1} + (1 - \beta_2)(G_t^2) \mathbf{1}_m \quad (4)$$

$$C_t = \beta_2 C_{t-1} + (1 - \beta_2) \mathbf{1}_n^T (G_t^2) \quad (5)$$

$$\widehat{V}_t = (R_t C_t / \mathbf{1}_n^T R_t) / (1 - \beta_2^t)$$

(6)

Kedua, Adafactor menghilangkan momentum sepenuhnya. Dengan meningkatkan tingkat decay seiring waktu dan memotong pembaruan gradient sesuai akar mean-square [11].

$$RMS(U_t) = RMS_{x \in X}(u_{xt}) = \sqrt{\text{Mean}_{x \in X} \left( \frac{g_{xt}^2}{\widehat{v}_{xt}} \right)}$$

(7)

Dan akhirnya, Adafactor mengalikan learning rate dengan skala parameter (ini disebut relative step size) [11].

$$\alpha_t = \max(\epsilon_2 RMS(X_{t-1})) \rho_t \quad (8)$$

## 2.7. Penilaian dan Evaluasi

Penilaian performa model adalah tahapan untuk menentukan akurasi dan kinerja dari model IndoBERT tersebut dalam melakukan analisis sentimen. Evaluasi dari kedua metode optimizer ini dilakukan dengan menghitung nilai *precision*, *recall*, *f-1 score*, dan akurasi.

$$Precision = \frac{TP}{TP+PF} \quad (9)$$

$$Accurcay = \frac{TP+TN}{TP+FN+TN+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

Keterangan :

- TP/True Positive = jumlah data relevan yang secara benar diklasifikasian sebagai kecocokan oleh model
- TN/True Negative = jumlah data tidak relevan yang diklasifikasikan sebagai tidak cocok dengan benar oleh model
- FP/False Positive = jumlah data yang tidak relevan, namun diklasifikasikan sebagai kecocokan oleh model
- FN/False Negative = jumlah data relevan, namun tidak diklasifikasikan sebagai kecocokan data oleh sistem

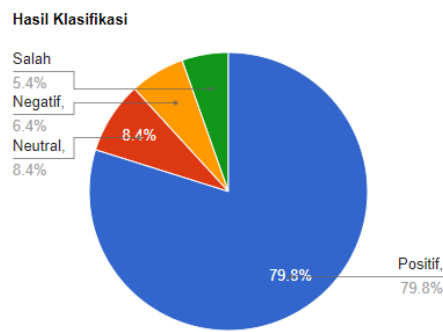
### 3. Hasil dan Pembahasan

Data yang akan diolah dalam penelitian ini berasal dari tripadvisor.com. Data berupa ulasan pengguna yang diposting pada beberapa tempat wisata di Bali. Data diambil dalam jangka waktu 6 tahun terakhir. Karakteristik data yang diperoleh berbahasa Indonesia dan termasuk ke dalam data tidak terstruktur. Proses pengambilan data dilakukan dengan metode *crawling*. *Crawling* dilakukan dengan menggunakan kata kunci nama tempat wisata. Kemudian data teks dan rating diekstrak menggunakan beautifulsoup. Jumlah data yang berhasil didapatkan sebesar 908 data. Pada hasil data ini ditetapkan nilai 1-2 adalah negative, 3 adalah neutral, dan 4-5 adalah positive. Data tersebut merupakan data mentah yang selanjutnya diproses ke tahap selanjutnya agar data bersih, rapi dan siap untuk diolah serta dianalisis. Pembagian dataset kedalam 3 proporsi, mendapatkan hasil akurasi tertinggi pada proporsi 70:15:15 dengan hasil akurasi training sebesar 97% dan akurasi validation sebesar 82%. Hasil testing mendapatkan akurasi sebesar 95%. Selain dari hasil akurasi, peneliti menggunakan *confusion matrix* untuk mengukur sejauh mana model IndoBERT melakukan prediksi. Para peneliti yang berupaya membuat program yang dapat menganalisis ulasan tentang tempat-tempat wisata berdasarkan sentimen diharapkan dapat menganggap temuan penelitian ini berguna.

#### 3.1. Analisis

Kami melakukan pelatihan pada model IndoBERT dengan *batch size* 32, 3 epoch, dan *learning rate optimizer* sebesar 0.00003. *Fine-tuning* ini selesai dalam waktu 36 menit. Dapat dilihat bahwa model IndoBERT mampu mengklasifikasikan dengan benar label positif sebanyak 725 dan salah dalam memprediksi label sebanyak 25. Label neutral berhasil diprediksi dengan benar sebanyak 76 dan salah dalam memprediksi label sebanyak 5 data. Label negative berhasil diprediksi dengan benar sebanyak 58 dan salah dalam memprediksi label sebanyak 19. Dapat dilihat disini bahwa persentase label positive benar adalah 79,8%, negative benar adalah 6,38%, dan label lainnya adalah 13,96%.





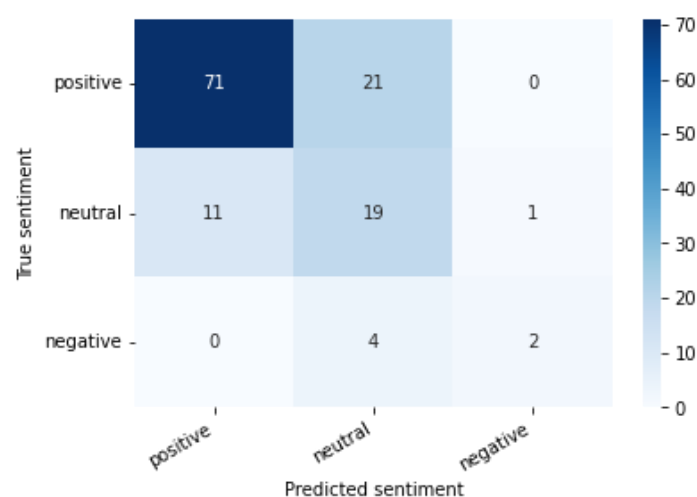
**Gambar 4.** Diagram Pembagian Hasil Klasifikasi

### 3.2. Evaluasi Optimizer

Hasil evaluasi pengujian dari IndoBERT untuk analisis sentimen menggunakan AdamW Optimizer terhadap data *testing* dengan mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f-1 score* dapat dilihat sebagai berikut. Hasil dapat dilihat pada tabel dan *confusion matrix* yang mencakup nilai sebenarnya dan nilai prediksi dari analisis sentimen berikut.

**Tabel 3.** Perbandingan Evaluasi Hasil AdamW

	Precision	Recall	F-1
Positive	0.98	0.91	0.97
Negative	0.92	0.98	0.97
Akurasi : 0.97			



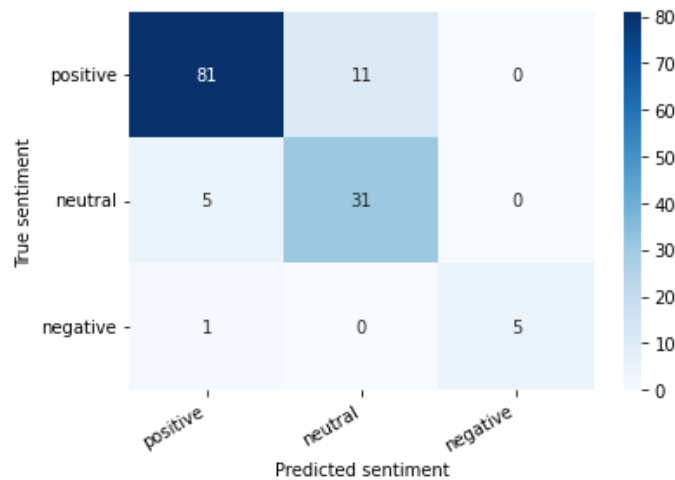
**Gambar 5.** Confusion Matrix Hasil Klasifikasi AdamW

Kemudian, dibawah ini merupakan hasil evaluasi pengujian dari IndoBERT untuk analisis sentimen menggunakan AdaFactor Optimizer terhadap data *testing* dengan mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f-1 score*. Hasil dapat dilihat pada tabel dan *confusion matrix* berikut.

**Tabel 4.**  
Evaluasi Hasil

	Precision	Recall	F-1
Positive	0.99	0.95	0.98
Negative	0.95	0.99	0.98
Akurasi : 0.98			

Perbandingan  
AdaFactor



**Gambar 6.** Confusion Matrix Hasil Klasifikasi AdaFactor

#### 4. Kesimpulan

Berdasarkan hasil evaluasi yang telah didapatkan sebelumnya, hasil penyetelan pada model BERT untuk analisis sentimen ulasan destinasi wisata di Bali menunjukkan bahwa model IndoBERT menghasilkan akurasi 95% dengan rasio dataset 70:15:15. dibandingkan dengan keseluruhan hasil prediksi. Akurasi yang dihasilkan recall menghasilkan 95% yang berarti bahwa IndoBERT dapat mengembalikan seluruh informasi mengenai nilai prediksi pada tingkatan baik. Serta, akurasi presisi yang didapat sebesar 93% yang berarti bahwa model ini baik dalam ketepatan memprediksi sentimen.

Berdasarkan keseluruhan hasil dari pemodelan IndoBERT, bisa disimpulkan bahwa IndoBERT mampu mengklasifikasikan dan memprediksi dengan sangat akurat karena semua akurasi diatas 90%. Dari penelitian ini, kedua optimisasi tersebut dapat memberikan hasil yang tepat dalam mengklasifikasikan ulasan objek wisata untuk mengkategorikan pendapat dalam ulasan tersebut termasuk kedalam sentimen positif atau negatif. Kemudian, hasil optimisasi dengan

AdamW *optimizer* mampu memberikan performa dengan nilai akurasi sebesar 97%. Sedangkan optimasi dengan AdaFactor *optimizer* mampu memberikan performa yang terbaik dengan nilai akurasi sebesar 98,2%. Sehingga dari hasil nilai prediksi sentimennya dapat disimpulkan bahwa model dengan performa terbaik yaitu dengan penerapan AdaFactor *optimizer*. Hal ini dapat diakibatkan optimisasi memori dan *smoothing* yang diterapkan AdaFactor.

Disarankan agar penelitian selanjutnya membandingkan hasil perolehannya dengan penelitian ini menggunakan teknik pra-pelatihan dan algoritma-algoritma *deep learning* lainnya. Dapat disimpulkan bahwa dataset ulasan wisata Bali untuk tugas analisis sentimen dapat diperkaya lagi. Penelitian kedepannya dapat dengan melakukan eksplorasi terhadap performa dari model-model NLG seperti GPT, T5, dan BART dalam tugas serupa.

## Referensi

- [1] B. A. Utami and A. Kafabih, "Sektor Pariwisata Indonesia di Tengah Pandemi Covid 19," *JDEP (Jurnal Dinamika Ekonomi Pembangunan)*, vol. 4, no. 1, pp. 8–14, 2021, doi: 10.33005/jdep.v4i1.198.
- [2] A. P. Yakup, "Pengaruh Sektor Pariwisata Terhadap Pertumbuhan Ekonomi Di Indonesia," *Universitas Airlangga Surabaya*, 2019, [Online]. Available: <https://repository.unair.ac.id/86231/>.
- [3] Y. Soritua, "Analisis Peran Sektor Pariwisata Menjadi Pendapatan Utama Daerah (Studi Banding: Peran Sektor Pariwisata di Provinsi Bali)," *Jurnal Ilmu Manajemen dan Akuntansi*, vol. 3, no. 2, pp. 1–7, 2017, doi: 10.33366/ref.v3i2.506.
- [4] R. Andrianto, "Dua Tahun Pandemi, Ekonomi Bali Ngenes Sekali," 8 Mar. 2022. [Online]. Available: <https://www.cnbcindonesia.com/news/20220308120609-4-320904/dua-tahun-pandemi-ekonomi-bali-ngenes-sekali>. [accessed 2 Oct. 2022]
- [5] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation," 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*, pp. 8875–8898, 2021, doi: 10.18653/v1/2021.emnlp-main.699.
- [6] A. Tabassum and R. R. Patil, "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing," *International Research Journal of Engineering and Technology*, vol. 7, no. 6, pp. 4864–4867, 2020, [Online]. Available: [www.irjet.net](http://www.irjet.net).
- [7] A. Subakti, H. Murfi, and N. Hariadi, "The performance of BERT as data representation of text clustering," *Journal Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00564-9.
- [8] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?," *Lecture Notes in Computer Science*, vol. 11856, no. 2, pp. 194–206, 2019, doi: 10.1007/978-3-030-32381-3\_16.
- [9] J. Alammar, "The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)," 27 Jun. 2018. [Online]. Available: <https://jalammar.github.io/illustrated-transformer/>. [accessed 5 Oct. 2022]
- [10] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *7th International Conference on Learning Representations, ICLR 2019*, 2019.
- [11] N. Shazeer and M. Stern, "Adafactor: Adaptive learning rates with sublinear memory cost," in *35th International Conference on Machine Learning, ICML 2018*, vol. 10, pp. 7322–7330, 2018.

*This page is intentionally left blank.*