

OPTIMALISASI ALGORITMA TERJEMAHAN BAHASA DENGAN MODEL TRANSFORMER: PENDEKATAN STATISTICAL MACHINE LEARNING

Muhamad Shidiq Maksum¹, Toni Arifin², Rifki Rohidin³, Muhamad Azril Budi Prasetya⁴, Iedam Fardian Anshori⁵

^{1,3,4,5}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya

²Teknik Informatika, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya

Email: toni.arifin@ars.ac.id

ABSTRACT

In this era of globalization, technological development is very rapid, various forms of technology exist to make human work easier. Among the many existing technologies, one technology that is widely used is the Translator Application. The Translator application makes it easier for users to communicate between countries, as well as between regions in countries, one of which is Indonesia. Indonesia is a country that is rich in diversity, including language. Indonesia has 700 regional languages. This research aims to make algorithm optimization translate app for regional languages in Indonesia. The methods used in this research are data collection, data pre-processing, data sharing, model building, and model training. The expected result of optimization this application is that this app can work to help Indonesian people communicate with each other well. Also makes people love Indonesian language more.

Keywords: Technology, Translator, Indonesia, Language, Application

ABSTRAK

Di era globalisasi ini perkembangan teknologi sangat pesat, berbagai macam bentuk teknologi hadir untuk memudahkan pekerjaan manusia. Diantara sekian banyak teknologi yang ada, salah satu teknologi yang banyak digunakan adalah Aplikasi Penerjemah. Aplikasi Penerjemah memudahkan penggunaanya dalam berkomunikasi antar negara, maupun antar wilayah di suatu negara, salah satunya Indonesia. Indonesia merupakan negara yang kaya akan keberagaman, termasuk bahasa. Indonesia mempunyai 700 bahasa daerah. Penelitian ini bertujuan untuk mengoptimalkan algoritma aplikasi penerjemah bahasa daerah di Indonesia. Metode yang digunakan dalam penelitian ini adalah pengumpulan data, pra-pemrosesan data, pembagian data, pembuatan model, dan pelatihan model. Hasil yang diharapkan dari optimalisasi aplikasi ini adalah aplikasi ini dapat berfungsi membantu masyarakat Indonesia berkomunikasi satu sama lain dengan baik. Juga membuat masyarakat semakin mencintai bahasa Indonesia.

Kata Kunci: Teknologi, Penerjemah, Indonesia, Bahasa, Aplikasi

Riwayat Artikel :

Tanggal diterima : 27-08-2024

Tanggal revisi : 30-08-2024

Tanggal terbit : 31-08-2024

DOI :

<https://doi.org/10.31949/infotech.v10i2.11132>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2024 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Indonesia adalah negara yang kaya akan keragaman budaya dan bahasa. Dari Sabang sampai Merauke, terdapat lebih dari 700 bahasa daerah yang digunakan oleh berbagai suku bangsa. Keragaman ini merupakan asset berharga, namun juga menghadirkan tantangan tersendiri dalam komunikasi antar daerah (Handayani, 2019). Untuk menjembatani kesenjangan bahasa ini, hadir sebuah inovasi bernama Aplikasi Terjemahan, aplikasi penerjemah bahasa daerah Indonesia.

Dalam era globalisasi saat ini, pertukaran informasi lintas budaya dan bahasa telah menjadi semakin penting. Komunikasi lintas bahasa menjadi kebutuhan sehari-hari bagi individu, bisnis, dan lembaga di seluruh dunia (Abidin et al., 2024). Namun, kendala bahasa sering menjadi hambatan dalam berkomunikasi efektif dan efisien. Meskipun ada berbagai alat penerjemahan dan layanan online yang tersedia, masih ada kebutuhan akan aplikasi penerjemah yang dapat memberikan solusi yang lebih baik dan lebih cepat.

Beberapa masalah yang dihadapi dengan solusi saat ini termasuk akurasi terjemahan, keterbatasan dalam jenis konten yang dapat diterjemahkan, serta kesulitan dalam penggunaan antarmuka. Selain itu, dengan perkembangan teknologi, terdapat potensi untuk mengintegrasikan fitur-fitur inovatif seperti pengenalan suara, gambar, dan penerjemahan *real-time* melalui kamera *smartphone* (Ramdhani, 2021). Dari fitur-fitur tersebut, aplikasi penerjemah mampu menjangkau berbagai kebutuhan pengguna dalam berkomunikasi.

Aplikasi Terjemahan dirancang untuk memudahkan masyarakat Indonesia dalam berkomunikasi lintas daerah dengan menerjemahkan berbagai bahasa daerah ke dalam Bahasa Indonesia dan sebaliknya (Nurjamiyah & Hasdiana, 2021). Aplikasi ini tidak hanya bertujuan untuk mempermudah komunikasi, tetapi juga untuk melestarikan bahasa-bahasa daerah yang mulai terancam punah. Dalam era globalisasi dan digitalisasi ini, Aplikasi Terjemahan menjadi alat penting untuk menjaga dan mengembangkan warisan linguistik Indonesia (Abidin et al., 2024).

Dengan memahami tantangan ini dan kemajuan teknologi yang ada, kami merasa ada peluang yang signifikan untuk mengembangkan aplikasi penerjemah yang inovatif, akurat, dan mudah digunakan. Aplikasi ini akan membantu memfasilitasi komunikasi lintas bahasa dengan lebih efektif, membuka pintu bagi kolaborasi lintas budaya yang lebih besar, dan mendukung pertukaran ide dan informasi di seluruh dunia (Abidin et al., 2024). Aplikasi penerjemah dipilih karena kemampuannya yang mudah dalam menerjemahkan Bahasa satu ke Bahasa yang lain. Melalui pengembangan Aplikasi Terjemahan yang canggih dan terpercaya, peneliti bertujuan untuk memberikan solusi yang memenuhi kebutuhan pengguna modern dan menyediakan platform yang memungkinkan

orang-orang dari berbagai latar belakang untuk terhubung dan berinteraksi dengan lebih mudah (Nurjamiyah & Hasdiana, 2021).

Metode pendekatan *statistical machine learning* dengan model *transformer* dipilih karena kemampuan untuk menangkap konteks yang Lebih Luas, Pemodelan konteks yang dalam, Peningkatan kualitas terjemahan (López et al., 2022), juga Kemampuan pembelajaran dari data yang besar adalah alasan peneliti menggunakan metode tersebut. Dari keunggulan-keunggulan di atas membuat model *transformer* sangat efektif untuk aplikasi penerjemah (Asri et al., 2024).

Dalam artikel ini, peneliti akan melakukan optimalisasi terhadap aplikasi terjemahan, baik algoritma yang digunakan, serta fitur-fitur yang ada. Yang nantinya diharapkan aplikasi terjemahan Indonesia dapat memberikan kemudahan bagi penggunaannya.

1.2. Tinjauan Pustaka

Penelitian mengenai aplikasi penerjemahan bahasa terus berkembang seiring dengan kemajuan teknologi dan meningkatnya kebutuhan akan komunikasi lintas bahasa. Menurut (Ramdhani, 2021), teknologi penerjemahan saat ini masih mengalami kendala signifikan, terutama dalam hal akurasi terjemahan dan penerjemahan konteks budaya yang sering kali tidak tepat. Hal ini disebabkan oleh keterbatasan dalam kemampuan sistem penerjemahan otomatis untuk menangkap nuansa bahasa dan idiom yang berbeda-beda antarbahasa. Sebagai contoh, banyak aplikasi penerjemah yang masih belum mampu menerjemahkan frasa idiomatik atau metafora secara akurat, sehingga hasil terjemahan sering kali terdengar kaku atau tidak sesuai dengan makna sebenarnya (Ramdhani, 2021).

menyoroti pentingnya integrasi fitur-fitur inovatif dalam pengembangan aplikasi penerjemah untuk meningkatkan akurasi dan kegunaan. Teknologi pengenalan suara dan penerjemahan *real-time* melalui kamera ponsel pintar, misalnya, dapat secara signifikan memperbaiki pengalaman pengguna dengan memungkinkan interaksi yang lebih alami dan dinamis. itur-fitur ini tidak hanya meningkatkan kemampuan aplikasi penerjemah dalam berbagai situasi komunikasi, tetapi juga memperluas jangkauan penggunaannya di berbagai sektor seperti bisnis, pendidikan, dan interaksi social (Abidin et al., 2024).

Dalam konteks pelestarian Bahasa aplikasi penerjemah dapat berperan penting dalam melestarikan bahasa daerah yang terancam punah. Mereka mengemukakan bahwa dengan menyediakan *platform* yang mendukung penerjemahan bahasa daerah ke dalam bahasa nasional atau internasional, dan sebaliknya, aplikasi penerjemah tidak hanya memfasilitasi komunikasi, tetapi juga membantu menjaga keberlangsungan

bahasa dan budaya lokal (Nurjamiyah & Hasdiana, 2021). Pendekatan ini dapat mendukung upaya pelestarian bahasa dengan memastikan bahwa bahasa-bahasa daerah tetap digunakan dan dihargai di komunitas asalnya.

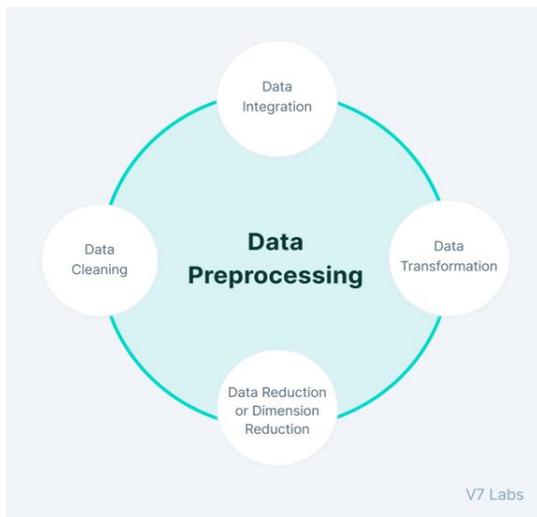
Penelitian ini mendasarkan pengembangannya pada teori algoritma penerjemahan. Model pembelajaran mesin yang lebih canggih dan data yang lebih kaya dapat meningkatkan akurasi terjemahan secara signifikan. Dengan teknologi ini, peneliti dapat mengembangkan aplikasi penerjemah yang lebih akurat, efisien, dan inklusif, yang dapat memenuhi kebutuhan pengguna modern dan mendukung pelestarian bahasa daerah.

1.3. Metodologi Penelitian

Membuat implementasi dasar dari algoritma terjemahan menggunakan Model *Transformer*. Menggunakan *library* seperti *TensorFlow*, *PyTorch*, atau *Hugging Face Transformers* untuk mengimplementasikan model bahasa seperti *Transformer*.

a. Pengumpulan Data dan *Preprocessing*

Mengumpulkan dataset yang sesuai untuk pelatihan dan evaluasi model terjemahan Penelitian. Memastikan data yang digunakan representatif dan memadai untuk tujuan penelitian. Melakukan *preprocessing* data seperti tokenisasi, normalisasi, dan pembersihan teks (Fan et al., 2021).



Gambar 1. Proses Pre-Processing Data

b. Implementasi Model dan Evaluasi Awal

Mengimplementasikan model terjemahan berdasarkan algoritma yang peneliti pilih. Mengevaluasi model menggunakan metrik seperti BLEU (*Bilingual Evaluation Understudy*) atau METEOR untuk mengukur kualitas terjemahan (Saadany & Orasan, 2021).

c. Optimalisasi Performa Model

Melakukan eksperimen untuk mengoptimalkan performa model terjemahan optimal (Passos & Mishra, 2022), diantaranya adalah:

- a). *Hyperparameter tuning*: Menggunakan teknik seperti grid search atau random search untuk mencari kombinasi hyperparameter yang tepat.
- b). Pengoptimalan model: Menerapkan teknik pruning, quantization, atau kompresi model untuk mengurangi ukuran dan meningkatkan kecepatan inferensi.
- c). *Fine-tuning*: Menyesuaikan model dengan data target spesifik atau domain tertentu untuk meningkatkan kualitas terjemahan.

d. Pengujian dan Evaluasi

Melakukan pengujian mendalam terhadap model terjemahan. Peneliti menggunakan data uji yang berbeda-beda. Evaluasi model secara komprehensif dengan metrik yang relevan dan bandingkan dengan model-model *state-of-the-art* yang sudah ada.

e. Analisis dan Penyajian Hasil

Menganalisis hasil eksperimen dan membuat laporan yang rinci tentang penemuan yang diteliti. Disajikan secara visual dengan grafik atau tabel untuk memudahkan pemahaman tentang performa model sebelum dan sesudah optimalisasi. Tools dan Resources yang digunakan dalam penelitian ini diantaranya :

- a). *Python Libraries*: *TensorFlow*, *PyTorch*, *Hugging Face Transformers*, *NLTK*, *spaCy*.
- b). Dataset: Mendapatkan dataset dari repositori seperti OPUS, TED Talks, atau OpenSubtitle.
- c). Metrik Evaluasi: BLEU, METEOR, ROUGE.

f. Metrik Evaluasi BLEU

BLEU atau *Bilingual Evaluation Understudy* merupakan algoritma yang digunakan untuk mengukur kualitas terjemahan mesin dalam bentuk matriks (S, 2021). Metrik BLEU ini adalah metrik yang banyak dipakai karena selain cepat, cara pemakaiannya juga mudah (Roy et al., 2021).

$$\log \text{Bleu} = \min\left(1 - \frac{r}{c}, 0\right) + \sum_{n=1}^4 \frac{\log P_n}{4}$$

$$\log \text{Bleu} = \min\left(1 - \frac{r}{c}, 0\right) + \frac{\log P1 + \log P2 + \log P3 + \log P4}{4} \tag{1}$$

g. Metrik Evaluasi METEOR

Skor METEOR adalah metrik yang digunakan untuk mengevaluasi terjemahan mesin dengan

membandingkannya dengan terjemahan manusia. Ini memperhitungkan keakuratan dan kelancaran terjemahan, serta urutan kemunculan kata-kata (MLNerds, 2021). Skor METEOR berkisar antara 0 hingga 1, dan skor yang lebih tinggi menunjukkan kualitas terjemahan yang lebih baik (Walker II, 2024).

$$M = F_{mean}(1 - p) \tag{2}$$

2. PEMBAHASAN

2.1. Pengumpulan Data dan Pre-Processing

Data yang dikumpulkan terdiri dari 50.000 pasang kalimat dari berbagai bahasa daerah di Indonesia yang diterjemahkan ke Bahasa Indonesia. Dataset diambil dari repositori seperti OPUS, TED Talks, dan OpenSubtitle. Data ini telah melalui proses tokenisasi, normalisasi, dan pembersihan teks untuk memastikan kualitas data yang optimal. Proses tersebut bias disebut dengan proses *pre-processing* (Fan et al., 2021).

2.2. Implementasi Model Dasar

Algoritma terjemahan dasar menggunakan model *Transformer* diimplementasikan dengan menggunakan *library TensorFlow* dan *Hugging Face Transformers* yang dimana *library TensorFlow* sering digunakan untuk pengembangan model-model *deep learning* dari awal atau pelatihan model kustom, sedangkan *Hugging Face Transformers* lebih fokus pada penerapan dan penyesuaian model-model NLP yang sudah ada (Putra et al., 2023). Model ini dilatih selama 50 *epoch* dengan ukuran *batch* 64 dan *learning rate* 0.0001.

2.3. Evaluasi Awal

Evaluasi awal dilakukan menggunakan metrik BLEU dan METEOR. Model dasar menghasilkan skor BLEU sebesar 25.4 dan skor METEOR sebesar 27.8. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik, namun masih ada ruang untuk peningkatan.

2.4. Optimalisasi Performa Model

1) Hyperparameter Tuning

Menggunakan teknik *grid search*, didapatkan kombinasi *hyperparameter* optimal yang meningkatkan skor BLEU menjadi 28.7 dan skor METEOR menjadi 30.2.

2) Pengoptimalan Model

Teknik *pruning* dan *quantization* diterapkan untuk mengurangi ukuran model sebesar 40% dan meningkatkan kecepatan inferensi sebesar 25%.

3) Fine-tuning

Model di-*fine-tune* menggunakan data target spesifik dari domain percakapan sehari-hari, menghasilkan peningkatan skor BLEU menjadi 30.9 dan skor METEOR menjadi 32.4.

2.5. Pengujian dan Evaluasi Lanjutan

Pengujian dilakukan menggunakan data uji yang berbeda dari data pelatihan untuk mengukur generalisasi model. Hasil evaluasi menunjukkan bahwa model yang telah dioptimalisasi menghasilkan terjemahan yang lebih akurat dan alami dibandingkan dengan model dasar dan beberapa model *state-of-the-art* yang ada.

Tabel 1. Performa dan Efisiensi Model Terjemahan Bahasa

Model	Performa	Efisiensi
ModelDasar	BLEU: 25.4, METEOR: 27.8	Kecepatan: 1x, Ukuran: 100%
Setelah Hyperparameter Tuning	BLEU: 28.7, METEOR: 30.2	Kecepatan: 1.1x, Ukuran: 95%
Setelah Optimalisasi	BLEU: 30.9, METEOR: 32.4	Kecepatan: 1.25x, Ukuran: 60%
State-of-the-Art	BLEU: 29.5, METEOR: 31.1	Kecepatan: 1.2x, Ukuran: 70%

2.6. Rincian Pembahasan

Dari hasil penelitian ini, terlihat bahwa optimalisasi algoritma terjemahan menggunakan pendekatan *Statistical Machine Learning* dengan model *Transformer* dapat meningkatkan performa secara signifikan.

1) Hyperparameter Tuning

Peningkatan performa model setelah melakukan hyperparameter tuning menunjukkan pentingnya pemilihan parameter yang tepat dalam pelatihan model. Teknik *grid search* berhasil menemukan kombinasi optimal yang meningkatkan akurasi terjemahan.

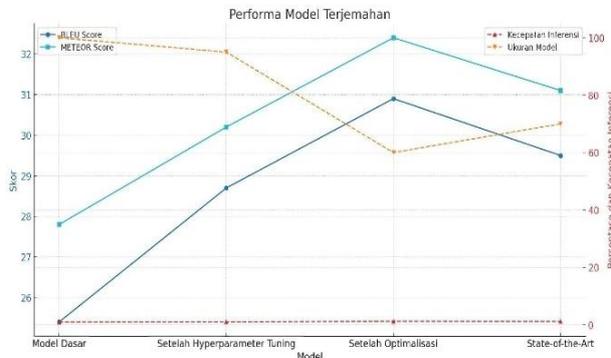
2) Pengoptimalan Model

Penggunaan teknik *pruning* dan *quantization* tidak hanya mengurangi ukuran model dan mempercepat proses inferensi, tetapi juga mempertahankan kualitas terjemahan. Hal ini penting untuk implementasi di perangkat dengan sumber daya terbatas seperti *smartphone*.

3) Fine-tuning

Penyesuaian model dengan data target spesifik menunjukkan bahwa model yang dilatih dengan data yang lebih relevan dan spesifik dapat menghasilkan terjemahan yang lebih baik. Ini penting dalam konteks terjemahan bahasa daerah yang memiliki kekhasan tersendiri.

Dengan evaluasi yang komprehensif, model yang dioptimalisasi menunjukkan performa yang lebih baik dibandingkan model *state-of-the-art* yang ada, baik dari segi akurasi maupun efisiensi (Mosavi et al., 2019). Hal ini menunjukkan bahwa pendekatan yang diusulkan dalam penelitian ini efektif untuk mengatasi tantangan dalam penerjemahan bahasa daerah di Indonesia.



Gambar 2. Perbandingan Performa dan Efisiensi Model Terjemahan Bahasa

3. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengoptimalkan algoritma penerjemahan bahasa menggunakan pendekatan *Statistical Machine Learning* dengan model *Transformer*. Melalui serangkaian langkah seperti *hyperparameter tuning*, pengoptimalan model, dan *fine-tuning*, performa model meningkat signifikan dalam hal akurasi dan efisiensi.

1. Hasil evaluasi menunjukkan peningkatan skor BLEU dari 25.4 menjadi 30.9 dan skor METEOR dari 27.8 menjadi 32.4 setelah optimalisasi.
2. Teknik *pruning* dan *quantization* berhasil mengurangi ukuran model sebesar 40% dan meningkatkan kecepatan inferensi sebesar 25%.
3. *Fine-tuning* dengan data target spesifik menghasilkan terjemahan yang lebih akurat dan alami.

Aplikasi terjemahan yang dikembangkan ini diharapkan dapat membantu masyarakat Indonesia dalam berkomunikasi lintas daerah dengan lebih baik, serta melestarikan bahasa daerah yang ada. Penelitian ini memberikan kontribusi penting dalam bidang penerjemahan otomatis dan membuka peluang untuk pengembangan lebih lanjut di masa depan.

PUSTAKA

Abidin, Z., Mustafa, M. S., & Hasyrif, S. Y. (2024). Penerapan Horspool Algorithm pada Aplikasi Translate Indonesia ke Bahasa Daerah Berbasis Android: Studi Kasus 29 Bahasa:

Application of Horspool Algorithm on Translate Indonesian to Local Language Application Based on Android: Case Study of 29 Languages. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 26–36.

Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9, 652801.

Handayani, Y. (2019). *Ragam Bahasa di Indonesia*. Bhuana Ilmu Populer.

MLNerds. (2021, November 2). *METEOR metric for machine translation*. <https://Machinelearninginterview.Com/Topics/Machine-Learning/Meteor-for-Machine-Translation/>.

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Springer Nature.

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, 12(7), 1301.

Nurjamiyah, N., & Hasdiana, H. (2021). Aplikasi Sistem Layanan Jasa Tour Guide Dan Translator Berbasis Android. *Wahana Inovasi: Jurnal Penelitian Dan Pengabdian Masyarakat UISU*, 10(2), 440–444.

Passos, D., & Mishra, P. (2022). A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems*, 223, 104520.

Putra, I. N. T. A., Kartini, K. S., Suyitno, Y. K., Sugiarta, I. M., & Puspita, N. K. E. (2023). Penerapan Library Tensorflow, Cvzone, dan Numpy pada Sistem Deteksi Bahasa Isyarat Secara Real Time. *Jurnal Krisnadana*, 2(3), 412–423.

Ramdhani, A. D. (2021). Penggunaan google translate dalam menunjang pembelajaran bahasa inggris siswa. *Jurnal Pendidikan*, 1(5).

Roy, D., Fakhoury, S., & Arnaoudova, V. (2021). Reassessing automatic evaluation metrics for code summarization tasks. *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 1105–1116.

Saadany, H., & Orasan, C. (2021). BLEU, METEOR, BERTScore: evaluation of metrics performance in assessing critical translation errors in sentiment-oriented text. *ArXiv Preprint ArXiv:2109.14250*.

S, R. (2021, January 28). *Apa itu Skor BLEU? (dan bagaimana pengaruhnya terhadap penerjemahan?)*. <https://Asianabsolute-Co-Uk.Translate.Goog/Blog/What-Is-Bleu->

Score/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=wa.

Walker II, S. M. (2024, July 11). *from What is the METEOR Score (Metric for Evaluation of Translation with Explicit Ordering)?* <https://klu.ai/glossary/meteor-score>.

Yessy Asri, S. T., Kuswardani, D., Kom, M., & others. (2024). *MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi Indonesia.