Jurnal Teknik Elektro dan Komputasi (ELKOM) ISSN : 2685-1814 (*Print*)
Volume 6 Nomor 1 Maret 2024 ‖ Hal. 105-111 ISSN : 2685-7677 (*Online*)

English Indonesia-Chan: OPUS-MT Powered Chatbot

Jerry Lasama[1], Sudianto[1*], Rafian Ramadhani[1], Muhammad David Hilmawan[1], Muhammad Yusril Aldean[1], Muhammad Adhan Hady Satria[1]

[1]Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan 128 Purwokerto
E-mail: sudianto@ittelkom-pwt.ac.id

Naskah Masuk: 16 Juni 2023; Diterima: 07 September 2023; Terbit: 31 Maret 2024

ABSTRAK

Abstrak - Pandemi COVID 19 menunjukan tren peningkatan pengguna platform digital pada media sosial seperti Whatsapp, Facebook, Instagram, dan Discord. Media sosial yang ramai digunakan dalam berkomunikasi secara massif adalah Discord. Discord telah memiliki 250 juta pengguna aktif yang terdaftar dari berbagai negara di seluruh Dunia. Namun, pengguna yang berasal dari berbagai negara menimbulkan perbedaan bahasa saat berkomunikasi sesama pengguna. Sehingga dibutuhkan sebuah metode untuk penerjemahan bahasa asing khususnya Bahasa Inggris ke Bahasa Indonesia secara mudah dan cepat agar komunikasi menjadi lebih dimengerti. Penelitian ini bertujuan untuk membuat sebuah *chatbot* Discord yang berfungsi untuk mengartikan kalimat berbahasa Inggris menjadi bahasa Indonesia. Metode yang dibangun, *chatbot* dirancang menggunakan model MarianNMT untuk penerjemahan Bahasa dan *dataset* corpus Bahasa Inggris dari Open Parallel corPUS (OPUS). Model dilatih menggunakan 15 Epoch dan mendapatkan hasil evaluasi dengan Loss sebesar 0.0047.

Kata kunci: Chatbot, Discord, MarianNMT, NLP, Terjemahan

ABSTRACT

Abstract - *The COVID-19 pandemic has shown an increasing trend of digital platform users on social media such as Whatsapp, Facebook, Instagram, and Discord. The social media that is widely used to communicate massively is Discord. Discord already has 250 million registered active users from various countries worldwide. However, users from various countries create language differences when communicating. So we need a method for translating foreign languages, especially English to Indonesian, easily and quickly to make communication more understandable. This study aims to create a Discord chatbot that translates English sentences into Indonesian. The method built in the chatbot is designed using the MarianNMT model for language translation and the English corpus dataset from Open Parallel corPUS (OPUS). The model was trained using 15 epochs and obtained evaluation results with a loss of 0.0047.*

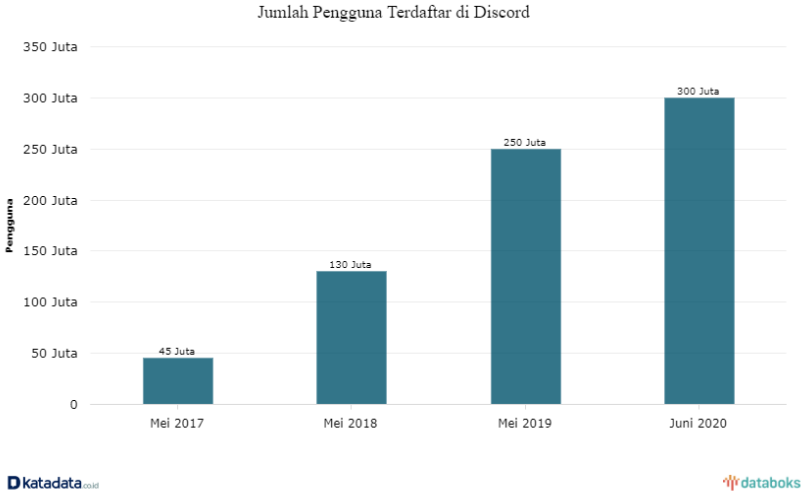
Keywords: *Chatbot, Discord, MarianNMT, NLP, Translate*

1. PENDAHULUAN

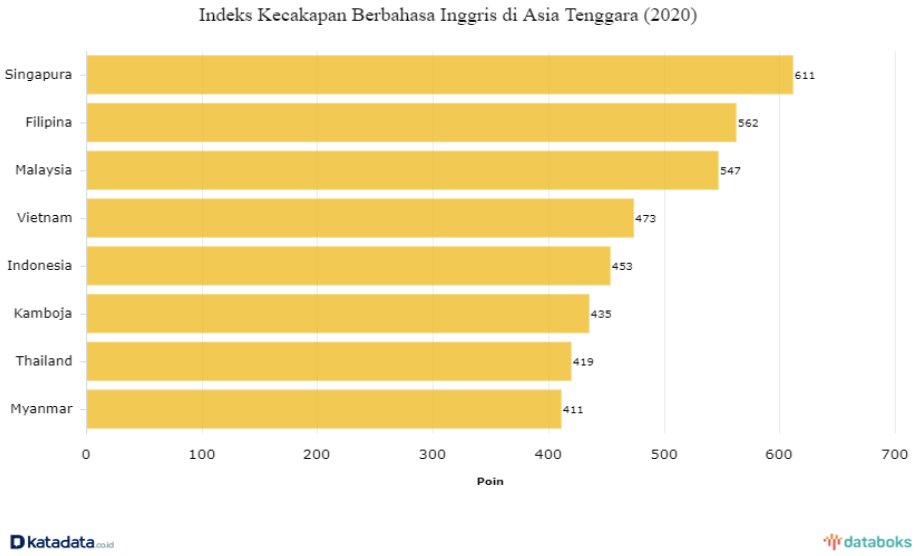
Pada saat pandemi COVID 19 [1], masyarakat lebih banyak berinteraksi lewat platform digital [2] yang sekarang mulai merajalela. Media sosial yang menjadi pilihan antara lain WhatsApp [3], Facebook, dan Instagram [4], dan salah satunya adalah Discord yang biasa dipakai oleh forum *group discussion* dalam berdiskusi. Peran discord juga banyak diimplementasikan dalam grup diskusi yang produktif [5] seperti misal untuk pengajaran, diskusi permainan digital, dan masih sangat banyak lagi implementasinya dalam kehidupan nyata [6] [7].

Data dari survei yang dilakukan Statista melansir bahwa pengguna aktif Discord di Juni 2020 berjumlah sebanyak 250 juta pengguna di seluruh Dunia [8], dengan angka sebesar itu, Discord menjadi salah satu portal diskusi terbesar yang ada di Dunia. Tantangan bagi pengguna Discord, Discord tersebar seluruh Dunia yang mengakibatkan beberapa transfer Bahasa menjadi sulit dimengerti saat komunikasi khususnya via chat dari Bahasa inggris ke Bahasa Indonesia. Selain itu, Indonesia ada di peringkat 5 dalam penguasaan Bahasa inggris di ASEAN [9] dan Indonesia berada pada peringkat 80/112 dalam kecakapan bahasa inggris dari data yang dirilis oleh EF English Proficiency Index 2021 [10] membuat Indonesia perlu belajar lebih

mengenai penggunaan Bahasa internasional khususnya Bahasa Inggris [11], seperti pada Gambar 1 dan Gambar 2.



Gambar 1. Jumlah pengguna discord 2017-2020



Gambar 2. Peringkat bahasa inggris negara ASEAN

Dari kondisi dan masalah diatas penelitian ini bertujuan untuk membuat suatu model penerjemah Bahasa inggris ke Indonesia serta botnya yang dapat diimplementasikan dalam fitur *chat* Discord [12] supaya dapat mempermudah masyarakat Indonesia yang menggunakan Discord memahami Bahasa inggris yang dikirim oleh lawan bicara sehingga pengguna Discord dapat memahami maksud dari kalimat yang disampaikan dan sebagai media belajar Bahasa asing, sehingga menambah wawasan mengenai Bahasa inggris. Model ini dibuat menggunakan model dari MarianNMT [13] dan corpus Bahasa inggris dari Open Parallel corPUS (OPUS) [14].

2. KAJIAN PUSTAKA

Pada penelitian sebelumnya yang diteliti oleh Diyah Puspitaningrum (2021) yang berjudul "*A Study of English-Indonesian Neural Machine Translation with Attention (Seq2Seq, ConvSeq2Seq, RNN, and MHA)"* penelitian ini bertujuan untuk membandingkan algoritme *Neural Machine Translation* saat diimplementasikan dalam kalimat formal untuk terjemahan dari bahasa Inggris menjadi bahasa Indonesia ataupun bahasa Indonesia menjadi bahasa Inggris. Hasil eksperimen untuk *ConvSeq2Seq* mencapai skor kalimat hingga 38,99 BLEU, 43,23 BLEU skor corpus, dan skor corpus 39,48 GLEU atas Seq2Seq Inggris-Indonesia. Bahasa Indonesia-Inggris, diperoleh hasil Con vSeq2Seq sebagai berikut: skor kalimat BLEU

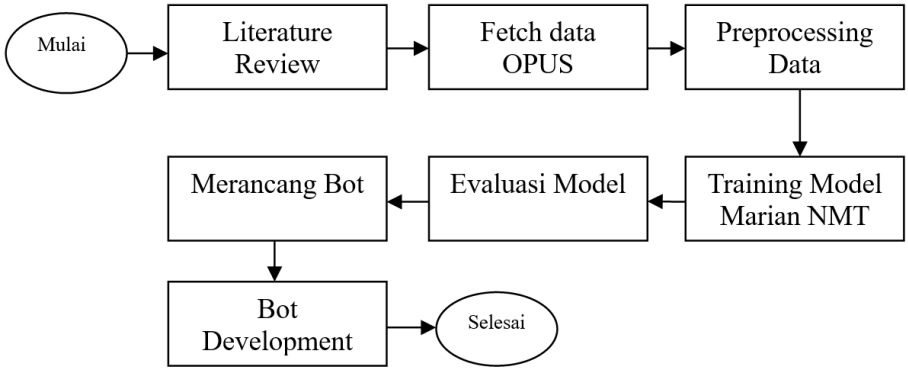
hingga 42,59, skor korpus 42,91 BLEU, skor korpus 41,05 GLEU, dan skor WER 1356,65 atas RNN dan MHA [15].

Penelitian selanjutnya yang dilakukan oleh Omar Zahour dkk (2020) yang berjudul "*A system for educational and vocational guidance in Morocco:Chatbot E-Orientation "* yang bertujuan membuat *chatbot* di bidang bimbingan pendidikan dan profesional yang didasarkan pada teori John Holland dan kuesioner RIASEC untuk menentukan tipe kepribadian dominan mahasiswa sarjana dan pascasarjana yang ingin memasuki pasar kerja. Dalam pengembanganya *chatbot* ini dibuat menggunakan DialogueFlow tool yang disediakan oleh Google beserta metode BERT [16].

Pada penelitian yang dilakukan oleh Heri Sujaini (2021) yang berjudul "Peningkatan Akurasi Mesin Penerjemah Bahasa Inggris-Indonesia Dengan Memaksimalkan Kualitas Dan Kuantitas Korpus Paralel" Tujuan dari penelitian ini adalah mengukur efek ukuran serta kualitas yang berasal dari sumber korpus paralel di MPS. Metode yang digunakan dalam penelitian ini adalah *Bilingual Evaluation Understudy* (BLEU) untuk melakukan klasifikasi pasangan kalimat paralel yang memiliki label kalimat berkualitas baik atau buruk. Hasil eksperimen penelitian yang dilakukan menghasilkan, dengan menggunakan 60% kalimat yang kualitas terjemahannya baik, kualitas terjemahan yang didapat meningkat sebesar 7,31% [17].

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam membuat *chatbot* penerjemah Bahasa Inggris ke Bahasa Indonesia berbasis platform Discord. Alur dari penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Alur penelitian pembuatan chatbot

3.1. Fetch Data dari OPUS

Untuk mendapatkan *dataset* yang digunakan pada penelitian ini, Langkah pertama adalah melakukan Fetch data dari OPUS dengan *corpus* dari Tatoeba [18]. Tatoeba adalah asosiasi yang bertujuan untuk mengumpulkan sejumlah besar kalimat dan terjemahan dengan fokus pada kerja sama, keragaman, dan keterbukaan, kumpulan *dataset* seperti pada Tabel 1.

Tabel 1. *Dataset* penerjemah

En	Id
I have to go to sleep.	Aku harus pergi tidur.
Muiriel is 20 now.	Muiriel berumur 20 sekarang.

3.2. Pra-proses Data

Dataset yang telah didapatkan perlu di *preprocessing* terlebih dahulu sebelum digunakan untuk proses *training* model. Langkah-langkah dalam *preprocessing dataset* yang digunakan dalam penelitian ini adalah *case folding, punctuation removal,* dan *number removal*. Semua karakter alfabet dalam *dataset* diubah menjadi huruf kecil, lalu tanda baca dan karakter angka yang ada dihilangkan dari *dataset*, seperti pada Tabel 2.

a. *Case Folding: Case Folding* adalah menjadikan huruf yang ada pada kalimat menjadi huruf kecil atau biasa disebut *lowercase*. *Case folding* dilakukan untuk menyamakan semua kata karena kata yang sama tetapi penggunaan huruf kapital yang berbeda dapat mempengaruhi hasil dari penelitian [19]–[24].

b. *Punctuation Removal: Punctuation Removal* adalah proses menghilangkan tanda baca dalam *dataset*. Tanda baca pada dokumen tidak memiliki arti yang bermakna pada suatu kalimat. Maka lebih baik tanda baca untuk dihilangkan.
c. *Number Removal: Number Removal* adalah proses menghilangkan karakter angka dalam *dataset*. Dikarenakan model hanya mengolah kata kata.

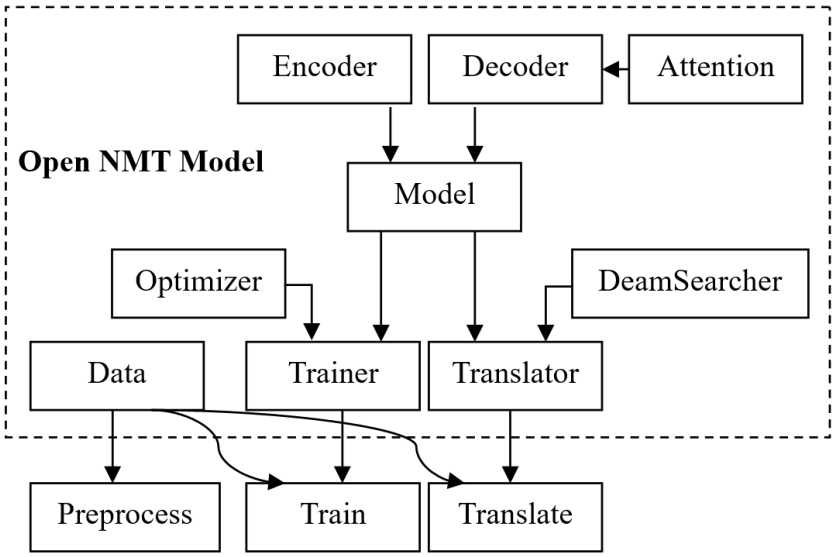
Tabel 2. Proses *case folding*, *punctuation removel*, dan *number removel*

En	Id
i have to go to sleep	aku harus pergi tidur
muiriel is now	muiriel berumur sekarang

Pada Tabel 2 dapat dilihat hasil dari proses *preprocessing* yang terdiri dari *case folding*, *punctuation removal*, dan *number removal*. Dapat dilihat bahwa contoh *dataset* dalam tabel ini tidak memiliki karakter huruf besar, karakter special dan tanda baca, serta karakter angka dikarenakan *dataset* sudah dilakukan proses *preprocessing*.

3.3. Marian NMT

Langkah selanjutnya adalah melakukan *training* model Marian NMT menggunakan model yang telah dilatih bersumber dari Helsinki NLP [25]. Kemudian kumpulan data disempurnakan melalui proses *preprocessing* dengan menggunakan Adam sebagai *optimizer*, dengan *max epochs* sebanyak 15 dan *batch* sebanyak 32. Berikut adalah visualisasi cara kerja dari model NMT yang digunakan seperti pada Gambar 4.



Gambar 4. Cara kerja model Marian NMT

3.4. Chatbot

Chatbot dirancang menggunakan Bahasa pemrograman Python. Chatbot bekerja dengan cara yang ditampilkan pada Gambar 5.

Gambar 5. Cara kerja chatbot

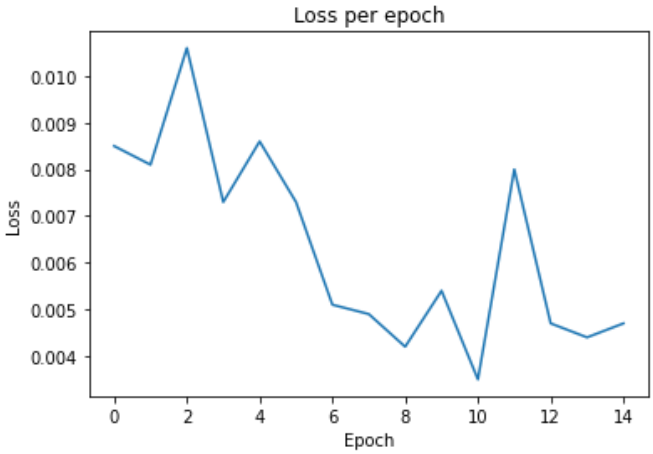
Chatbot akan menunggu perintah dari *user*, *user* mengirim *chat* perintah untuk menerjemahkan kalimat Bahasa Inggris ke Bahasa Indonesia. Jika *chat* tidak terkirim, maka *user* akan mengirim kembali *chat* perintah tersebut sampai terkirim. Pesan dari perintah *chat* akan dilakukan proses *preprocessing* oleh *chatbot* sebelum dilakukan proses penerjemahan. Pesan yang telah diterjemahkan

lalu akan dikirim oleh *chatbot* ke *user* yang mengirim *chat* perintah tersebut. Dikarenakan model terjemahan dan *bot* yang cukup ringan, proses penerjemahan dapat dilakukan secara *real time.*

4. HASIL DAN PEMBAHASAN

4.1. Hasil Evaluasi Model

Untuk mendapatkan dataset yang digunakan pada penelitian ini, Langkah pertama adalah melakukan Fetch data dari OPUS dengan corpus dari Tatoeba [18]. Tatoeba adalah asosiasi yang bertujuan untuk mengumpulkan sejumlah besar kalimat dan terjemahan dengan fokus pada kerja sama, keragaman, dan keterbukaan.



Gambar 6. *Loss* pada tiap Epoch

Tabel 3. Loss pada tiap *epoch*

Epoch	Loss
1	0,0085
2	0,0081
3	0,0106
4	0,0073
5	0,0086
6	0,0073
7	0,0051
8	0,0049
9	0,0042
10	0,0054
11	0,0035
12	0,0080
13	0,0047
14	0.0044
15	0.0047

Dapat dilihat dari hasil evaluasi model pada Gambar 6 dan Tabel 3 nilai *Loss* mengalami penurunan meskipun ada lonjakan nilai *Loss* di beberapa *epoch*. Dengan nilai *Loss* tertinggi ada di *epoch* pertama dengan nilai *Loss* sebesar 0,0085 dan nilai Loss terkecil pada Epoch ke-11 dengan nilai Loss sebesar 0,0035. Hasil akhir didapatkan pada epoch ke-15 dengan nilai Loss sebesar 0,0047.

4.2. Hasil Implementasi Chatbot

Gambar 7. Hasil implementasi chatbot

Pada Gambar 7, *chatbot* dipanggil menggunakan perintah "!id" untuk menerjemahkan apa yang ditulis oleh *user* dari Bahasa Inggris ke Bahasa Indonesia. Saat *chatbot* menerima perintah dari user, *chatbot* mengambil *chat* dari user, melakukan proses penerjemahan dari Bahasa Inggris ke Bahasa Indonesia, lalu menampilkan hasil terjemahan tersebut ke Discord. Waktu pengerjaan penerjemahan tersebut sangat cepat dan dapat dikategorikan secara *real time*.

5. KESIMPULAN

Dari pengujian yang telah dilakukan, diperoleh hasil bahwa model Marian NMT dapat mengartikan kalimat berbahasa Inggris menjadi Bahasa Indonesia dengan akurat. Model Marian NMT yang dilatih dengan menggunakan *dataset corpus* Tatoeba yang didapatkan di OPUS dan 15 *epoch* mendapatkan hasil evaluasi dengan *Loss* sebesar 0,0047. *Chatbot* yang dibuat menggunakan model Marian NMT dapat menerjemahkan Bahasa Inggris ke Bahasa Indonesia pada *platform discord* secara *real time.*

REFERENSI

[1] N. K. Suni Astini, "Tantangan Dan Peluang Pemanfaatan Teknologi Informasi Dalam Pembelajaran Online Masa Covid-19," *Cetta J. Ilmu Pendidik.*, vol. 3, no. 2, pp. 241–255, 2020.
[2] R. Komalasari, "Manfaat Teknologi Informasi Dan Komunikasi Di Masa Pandemi Covid 19," *Tematik*, vol. 7, no. 1, pp. 38–50, 2020.
[3] D. P. Covid-, "Pemanfaatan Media Komunikasi Whatsapp Untuk Mengoptimalisasi Kinerja Jurnalis," vol. 16, no. 2, pp. 141–151, 2021.
[4] J. Teknologi and I. Komunikasi, "Tematik: Jurnal Teknologi Informasi Komunikasi (e-Journal) Vol. 8 No. 2 Desember 2021," vol. 8, no. 2, pp. 160–175, 2021.
[5] J. P. Raihan1) and M. , Yuliani Rachma Putri, S.Ip., "Pola Komunikasi Group Discord Pubg.Indo.Fun Melalui Aplikasi Discord," vol. 5, no. 3, pp. 4161–4169, 2018.
[6] M. R. Ridho, M. Muhaimin, and H. S. Harjono, "Pengaruh Aplikasi Discord Dalam Pembelajaran Daring Terhadap Hasil Belajar Pada Matakuliah Komputer," *J. Ilm. Bina Edukasi*, vol. 14, no. 1, pp. 22–35, 2021.
[7] S. Aditia, "Inovasi Pembelajaran Berbasis Aplikasi Mobile," 2020.
[8] Statista, "Aplikasi Berbasis Audio Discord Punya 300 Juta Pengguna." https://databoks.katadata.co.id/datapublish/2021/02/24/aplikasi-berbasis-audio-discord-punya-300-juta-pengguna (accessed Jan. 12, 2022).
[9] E. First, "Indeks Kecakapan Berbahasa Inggris di Asia Tenggara (2020)," 2020. Indeks Kecakapan Berbahasa Inggris (EPI) versi Education First (EF) di Asia Tenggara masih dipimpin Singapura pada 2020. Negeri tetangga tersebut berhasil mengantongi 611 poin dari 800 poin. Sehingga membawa Singapura di posisi 10 dunia dari 100 negara da (accessed Jan. 12, 2022).
[10] English Firts, "EF English Proficiency Index 2021," 2021. https://www.ef.com/wwen/epi/regions/asia/indonesia/ (accessed Jan. 12, 2022).
[11] I. G. N. P. K. Gek Wulan Novi Utami, "Pemaknaan Verba Bahasa Inggris Dan Upaya Peningkatan Pengajaran Dan Pembelajaran Verba," vol. 4, no. 1, pp. 77–82, 2018.
[12] I. N. Norambuena and A. Bergel, "Building a bot for automatic expert retrieval on discord," *MaLTESQuE 2021 - Proc. 5th Int. Work. Mach. Learn. Tech. Softw. Qual. Evol. co-located with ESEC/FSE 2021*, no. Dcc, pp. 25–30, 2021.
[13] A. G. Jones and D. Wijaya, "Sentiment-based Candidate Selection for NMT," *Proc. Mach. Transl. Summit XVIII Res. Track*, pp. 188–201, 2021.
[14] R. A. Rahmanda, M. Adriani, and D. Tanaya, "Cross Language Information Retrieval Using Parallel Corpus with Bilingual Mapping Method," *Proc. 2019 Int. Conf. Asian Lang. Process. IALP 2019*, pp. 222–227, 2019.
[15] D. Puspitaningrum, "A Study of English-Indonesian Neural Machine Translation with Attention (Seq2Seq, ConvSeq2Seq, RNN, and MHA)," *Conf. Sustain. Inf. Eng. Technol. 2021*, pp. 271–280, 2021.
[16] O. Zahour, E. H. Benlahmar, A. Eddaoui, H. Ouchra, and O. Hourrane, "A system for educational and vocational guidance in Morocco: Chatbot e-orientation," *Procedia Comput. Sci.*, vol. 175, pp. 554–559, 2020.
[17] R. Darwis, H. Sujaini, and R. D. Nyoto, "Peningkatan Mesin Penerjemah Statistik dengan Menambah Kuantitas Korpus Monolingual (Studi Kasus: Bahasa Indonesia - Sunda)," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, p. 27, 2019.
[18] J. Tiedemann, "Parallel data, tools and interfaces in OPUS," *Proc. 8th Int. Conf. Lang. Resour. Eval. Lr. 2012*, pp. 2214–2218, 2012.
[19] S. Sudianto, A. D. Sripamuji, I. R. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasifikasi Topik Berita," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 11, no. 2, pp. 84–91, 2022.

[20] S. Sudianto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study : Internet Selebgram Rachel Vennya Escape From Quarantine) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *Jutif*, vol. 3, no. 1, pp. 141–145, 2022.
[21] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Ardiansyah, M. H. Kafabi, and S. Sudianto, "Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM," *Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 85–89, 2022.
[22] S. Sudianto, J. A. Marseli, N. Nugroho, R. W. A. Rumpoko, and Z. Akhmad, "Comparison of Support Vector Machines and K-Nearest Neighbor Algorithm Analysis of Spam Comments on YouTube Covid Omicron," *JTI*, vol. 15, no. 2, pp. 110–118, 2022.
[23] S. Sudianto, "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi," vol. 4, no. 2, pp. 1027–1034, 2022.
[24] S. Chandra Ayunda Apta, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, and S. Sudianto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," *Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 80–84, 2022.
[25] J. Tiedemann and S. Thottingal, "OPUS-MT: Building Open Translation Services for the World," *Proc. 22nd Annu. Conf. Eur. Assoc. Mach. Transl.*, pp. 479–480, 2020.