PENGEMBANGAN MODEL INDOBERT DALAM *CHATBOT* PENGELOLAAN KOMUNIKASI BEASISWA IKATAN ALUMNI UPI

SKRIPSI

Diajukan Untuk Memenuhi Sebagian dari Syarat Memperoleh gelar Sarjana
Program Studi Ilmu Komputer

Oleh:
Raisyad Jullfikar
2106238

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2025

PENGEMBANGAN MODEL INDOBERT DALAM *CHATBOT*
PENGELOLAAN KOMUNIKASI BEASISWA IKATAN ALUMNI UPI

Oleh
Raisyad Jullfikar
2106238

Sebuah skripsi yang diajukan untuk memenuhi salah satu Syarat Memeroleh gelar Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia
Juli 2025

RAISYAD JULLFIKAR

2106238

PENGEMBANGAN MODEL INDOBERT DALAM *CHATBOT* PENGELOLAAN KOMUNIKASI BEASISWA IKATAN ALUMNI UPI

DISETUJUI DAN DISAHKAN OLEH TIM PEMBIMBING:

Pembimbing I,

Dr. Rani Megasari, S.Kom., M.T.
NIP. 198705242014042002

Pembimbing II,

Rosa Ariani Sukamto, M.T.
NIP. 198109182009122003

Mengetahui,
Ketua Program Studi Ilmu Komputer

Dr. Muhamad Nursalman, M.T.
NIP. 197909292006041002

PENGEMBANGAN MODEL INDOBERT DALAM *CHATBOT* PENGELOLAAN KOMUNIKASI BEASISWA IKATAN ALUMNI UPI

Oleh

Raisyad Jullfikar – raisyadjullfikar@upi.edu

2106238

ABSTRAK

Komunikasi efektif antara alumni dan organisasi penting untuk penyampaian informasi salah satunya terkait beasiswa. Di IKA UPI, Sumber Daya Manusia (SDM) terbatas membuat respons lambat dan inefesiensi komunikasi. Penelitian ini membuat *chatbot* menggunakan model IndoBERT yang merupakan model berbasis BERT (*Bidirectional Encoder Representations from Transformers*) untuk menjawab terkait beasiswa secara otomatis. Model diadaptasi melalui lima *fine-tuning*, yaitu *Masked Language Modeling* (MLM), *Next Sentence Prediction* (NSP), *Intent Classification*, *Extractive* QA (SQuAD), dan *Semantic Retrieval*. Evaluasi memakai metrik *accuracy*, *recall*, *f1-score*, *exact match* (EM), *top-k accuracy*, *mean reciprocal rank* (MRR), dan *latency*. Hasil uji menunjukkan kinerja kuat di tugas dasar seperti *f1-score* 93,33% untuk NSP dan 86,74% untuk *intent*. Tantangan tampak pada pemahaman konteks lebih dalam seperti pada *extractive* QA meraih *f1-score* 43,57% dan *retrieval* MRR 0,34. Temuan ini menegaskan kelayakan penerapan *chatbot* menggunakan model IndoBERT untuk pengelolaan komunikasi beasiswa IKA UPI, sekaligus memetakan tantangan dalam tugas-tugas pemahaman konteks yang lebih dalam.

Kata kunci: Beasiswa, BERT, *Chatbot*, *Fine-Tuning*, *Natural Language Processing* (NLP)

DEVELOPMENT OF THE INDOBERT MODEL IN THE UPI ALUMNI ASSOCIATION SCHOLARSHIP COMMUNICATION MANAGEMENT CHATBOT

Aranged by

Raisyad Jullfikar – raisyadjullfikar@upi.edu

2106238

ABSTRACT

Effective communication between alumni and organizations is crucial for conveying information, including scholarships. At IKA UPI, limited human resources (HR) make communication responses slow and inefficient. This study created a chatbot using the IndoBERT model, a BERT-based model (Bidirectional Encoder Representations from Transformers) to automatically answer questions about scholarships. The model was modified through five fine-tunings: Masked Language Modeling (MLM), Next Sentence Prediction (NSP), Intent Classification, Extractive QA (SQuAD), and Semantic Retrieval. The evaluation used metrics of accuracy, recall, f1-score, exact match (EM), top-k accuracy, mean reciprocal rank (MRR), and latency. Test results showed strong performance in basic tasks, such as an f1-score of 93.33% for NSP and 86.74% for intent. Challenges arose in deeper understanding, such as in extractive QA, which achieved an f1-score of 43.57% and a retrieval MRR of 0.34. This finding of the feasibility of implementing a chatbot using the IndoBERT model for managing IKA UPI scholarship communications, reflects both the challenges in the tasks of understanding the context more deeply.

Keywords: BERT, *Chatbot*, *Fine-Tuning*, *Natural Language Processing* (NLP), *Scholarship*

DAFTAR ISI

PERNYATAAN BEBAS PLAGIARISME ii

KATA PENGANTAR iv

UCAPAN TERIMA KASIH v

ABSTRAK vii

ABSTRACT viii

DAFTAR ISI ix

DAFTAR GAMBAR xiii

DAFTAR TABEL xvii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 11

1.3 Batasan Masalah 12

1.4 Tujuan Penelitian 12

1.5 Manfaat Penelitian 13

1.6 Sistematika Penulisan 13

BAB II TINJAUAN PUSTAKA 16

2.1 Kajian Teori 17

2.1.1 Komunikasi 17

2.1.2 *Artificial Intelligence* 19

2.1.2.1 *Natural Language Processing* (NLP) 23

2.1.3 *Chatbot* 27

2.1.4 Model BERT 29

2.1.5 IndoBERT 32

2.1.6 *Fine-Tuning* 34

2.1.6.1 *Fine-Tuning Masking Language Modelling* (MLM) 35

2.1.6.2 *Fine-Tuning Next Sentence Prediction* (NSP) 36
2.1.6.3 *Fine-Tuning Intent Classification* 37
2.1.6.4 *Fine-Tuning Extractive Question Anaswering* (SQuAD) 39
2.1.6.5 *Fine-Tuning Semantic Retrieval* (*Triplet Loss*) 40
2.1.7 Metode Evaluasi 42
2.1.7.1 *Confusion Matrix* 43
2.1.7.2 *Precision* 45
2.1.7.3 *Recall* 46
2.1.7.4 *F1-Score* 47
2.1.7.5 *Accuracy* 49
2.1.7.6 *Exact Match* (EM) 49
2.1.7.7 *Top-K Accuracy* 50
2.1.7.8 *Mean Reciprocal Rank* (MRR) 51
2.1.7.9 *Latency Test* 52
2.1.8 *Overfitting* 54
2.2 Penelitian Terkait 56
BAB III METODE PENELITIAN 61
3.1 Desain Penelitian 61
3.2 Waktu dan Tempat Penelitian 62
3.2.1 Waktu Penelitian 62
3.2.2 Tempat Penelitian 62
3.3 Objek dan Prosedur Penelitian 62
3.3.1 Objek Penelitian 62
3.2.2 Prosedur Penelitian 62
3.4 Teknik Pengumpulan Data 69
3.5 Instrumen Penelitian 69

3.5.1 Perangkat Keras (*Hardware*) 69
3.5.2 Perangkat Lunak (*Software*) 70
3.6 Teknik Analisis Data 70
BAB IV HASIL DAN PEMBAHASAN 72
4.1 Identifikasi Masalah Penelitian 78
4.2 Persiapan Eksperimen 79
4.2.1 Pengumpulan Data 80
4.2.2 Penentuan Format Dataset 91
4.2.3 Pembentukan Format Dataset 93
4.2.4 Pra-proses 96
4.2.4.1 Pra-proses *Masking Language Modelling (MLM)* 96
4.2.4.2 Pra-proses *Next Sentence Prediction* (NSP) 101
4.2.4.3 Pra-proses *Intent Classification* (Klasifikasi Intensi) 110
4.2.4.4 Pra-proses *Extractive* QA SQuAD 113
4.2.4.5 Pra-proses *Retrieval Triplet* 118
4.2.5 Pembagian Data 119
4.2.5.1 Pembagian Data pada *Masking Language Modelling* (MLM) 120
4.2.5.2 Pembagian Data pada *Next Sentence Prediction* (NSP) 120
4.2.5.3 Pembagian Data pada *Intent Classification* 121
4.2.5.4 Pembagian Data pada *Extractive* QA SQuAD 123
4.2.5.5 Pembagian Data pada *Retrieval Triplet* 123
4.2.6 Tokenisasi Model 124
4.2.6.1 Tokenisasi Model untuk *Masked Language Modelling* (MLM) 125
4.2.6.2 Tokenisasi Model untuk *Next Sentence Prediction* (NSP) 127
4.2.6.3 Tokenisasi Model untuk *Intent Classification* (Klasifikasi Intensi) 132

4.2.6.4 Tokenisasi Model untuk *Extractive* QA SQuAD 136
4.2.5.5 Tokenisasi Model untuk *Semantic Retrieval* (*Retrieval Triplet*).. 141
4.3 Pelaksanaan Eksperimen 144
4.3.1 *Fine-Tuning* 144
4.3.1.1 *Fine-Tuning Masked Language Modelling* (MLM) 145
4.3.1.2 *Fine-Tuning Next Sentence Prediction* (NSP) 151
4.3.1.3 *Fine-Tuning Intent Classification* (Klasifikasi Intensi)............... 156
4.3.1.4 *Fine-Tuning Extractive* QA SQuAD 161
4.3.1.5 *Fine-Tuning Semantic Retrieval* (*Triplet Loss*) 165
4.4 Pengukuran & Analisis Hasil 169
4.4.1 Evaluasi Model *Masked Language Modelling* (MLM) 170
4.4.2 Evaluasi Model *Next Sentence Prediction* (NSP) 172
4.4.3 Evaluasi Model *Intent Classification* 174
4.4.4 Evaluasi Model *Extractive* QA SQuAD... 177
4.4.5 Evaluasi Model *Semantic Retrieval* (*Triplet Loss*).......................... 178
4.4.6 Evaluasi *End-to-End Chatbot* 179
4.5 Pembahasan Hasil Eksperimen 181
BAB V SIMPULAN DAN SARAN 186
5.1 Simpulan 186
5.2 Saran.. 188
DAFTAR PUSTAKA 190

DAFTAR GAMBAR

Gambar 2.1 Peta Literatur 16
Gambar 2.2 Berbagai domain Kecerdasan Buatan (AI) (Ghosh dan Thirugnanam, 2021) 22
Gambar 2.3 Metode Pemrosesan Bahasa Alami (NLP) (Hartati dan Manullang, 2024) 24
Gambar 2.4 Metode, fungsi, dan model NLP (Hartati dan Manullang, 2024)...... 25
Gambar 2.5 Arsitektur BERT *generic* dalam (Babu dan Boddu, 2024)................ 30
Gambar 2.6 *Confusion Matrix* dalam (Grandini dkk., 2020)............................... 44
Gambar 3.1 Alur Penelitian 61
Gambar 3.2 Diagram Alir Prosedur Rinci Penelitian... 63
Gambar 3.3 Flow *Fine-Tuning* Model *Chatbot*.. 67
Gambar 4.1 Arsitektur Sistem 73
Gambar 4.2 *Pseudocode* Sistem *Chatbot*.. 78
Gambar 4.3 *Encoding* Label pada *Intent Classification* 111
Gambar 4.4 Mengubah nilai kategori pada kolom *tag*..112
Gambar 4.5 Meratakan struktur data SQuAD menjadi format *list of dictionary*.114
Gambar 4.6 Konversi data hasil *flattening* menjadi objek Dataset......................117
Gambar 4.7 Tampilan hasil objek Dataset yang telah berhasil dikonversi dari data awal 118
Gambar 4.8 Pustaka yang dibutuhkan saat proses *fine-tuning* pada MLM 125
Gambar 4.9 Pemanggilan *pre-trained tokenizer* dan model IndoBERT untuk tugas MLM 126
Gambar 4.10 Pendefinisian fungsi tokenisasi dengan *tokenizer* BERT............. 126
Gambar 4.11 Menerapkan fungsi tokenisasi secara *batch* pada dataset pelatihan dan evaluasi 127
Gambar 4.12 Memasukkan Pustaka yang dibutuhkan saat proses *fine-tuning* pada NSP 128
Gambar 4.13 Pemanggilan *tokenizer* dan model hasil dari *fine-tuning* MLM untuk tugas NSP 129
Gambar 4.14 Pendefinisian fungsi tokenisasi dengan *tokenizer* BERT............. 129

Gambar 4.15 Konversi DataFrame hasil pembagian data menjadi format Dataset dan penggabungan kedalam struktur DatasetDict 130
Gambar 4.16 Penerapan fungsi tokenisasi secara *batch* pada objek DatasetDict 131
Gambar 4.17 Memasukkan Pustaka yang dibutuhkan saat proses *fine-tuning* pada MLM 132
Gambar 4.18 Pemanggilan *tokenizer* hasil dari *fine-tuning* NSP untuk tugas *Intent Classification* 133
Gambar 4.19 Pemanggilan model hasil *fine-tuning* NSP untuk tugas *Intent Classification* 133
Gambar 4.20 Pendefinisian fungsi tokenisasi dengan *tokenizer* BERT 134
Gambar 4.21 Konversi DataFrame ke struktur DatasetDict 134
Gambar 4.22 Penerapan fungsi tokenisasi secara *batch* pada DatasetDict 135
Gambar 4.23 Penyesuaian nama kolom dan format output dataset menjadi tensor PyTorch untuk proses pelatihan 136
Gambar 4.24 Memasukkan Pustaka yang dibutuhkan saat proses *fine-tuning* pada *Extractive* QA SQuAD 137
Gambar 4.25 Pemanggilan *tokenizer* dan model hasil dari *fine-tuning* NSP untuk tugas *Extractive* QA SQuAD 137
Gambar 4.26 Pendefinisian fungsi tokenisasi dengan *tokenizer* BERT 138
Gambar 4.27 Penerapan fungsi tokenisasi pada dataset QA tanpa menghapus kolom untuk mempertahankan informasi posisi jawaban 139
Gambar 4.28 Fungsi penentuan posisi token jawaban (*start* dan *end*) berdasarkan offset karakter 140
Gambar 4.29 Penerapan fungsi add_start_end_positions untuk menambahkan posisi token jawaban pada setiap entri dalam dataset QA 140
Gambar 4.30 Memasukkan Pustaka yang dibutuhkan saat proses *fine-tuning* pada *Semantic Retrieval* menggunakan format *Triplet Loss* 141
Gambar 4.31 Pembentukan contoh *triplet* menggunakan InputExample untuk menyiapkan data pelatihan dan validasi pada model *retrieval* berbasis *triplet loss* 142
Gambar 4.32 Pemanggilan dan konfigurasi model hasil *fine-tuning* NSP kedalam *pipeline* SentenceTransformer dengan modul *Transformer* dan *Pooling* 143

Gambar 4.33 Inisialisasi DataLoader dan konfigurasi fungsi TripletLoss untuk pelatihan model *retrieval* 143
Gambar 4.34 Fungsi compute_metrics untuk menghitung metrik evaluasi berupa akurasi, *precision*, *recall*, dan *f1-score* 146
Gambar 4.35 Konfigurasi objek TrainingArguments 147
Gambar 4.36 Inisialisasi dan eksekusi proses pelatihan 148
Gambar 4.37 Visualisasi *training loss* dan *validation loss* selama proses pelatihan model 150
Gambar 4.38 Pendefinisian fungsi untuk menghitung metrik evaluasi 152
Gambar 4.39 Konfigurasi TrainingArguments yang mencakup parameter pelatihan 153
Gambar 4.40 Inisialisasi objek Trainer untuk memulai proses pelatihan model NSP 154
Gambar 4.41 Visualisasi *training loss* dan *validation loss* selama proses pelatihan model 156
Gambar 4.42 Fungsi compute_metrics() untuk menghitung Akurasi, *Precision*, *Recall*, dan *F1-Score* 157
Gambar 4.43 Konfigurasi TrainingArguments untuk Pelatihan Model *Intent Classification* 158
Gambar 4.44 Inisialisasi Objek Trainer dan Proses Pelatihan Model 159
Gambar 4.45 Visualisasi *training loss* dan *validation loss* selama proses pelatihan model 160
Gambar 4.46 Mendefinisikan fungsi untuk mengevaluasi hasil prediksi model QA 162
Gambar 4.47 Konfigurasi argumen pelatihan menggunakan TrainingArguments dari pustaka HuggingFace 163
Gambar 4.48 Inisialisasi dan pelatihan model menggunakan kelas Trainer dari HuggingFace 164
Gambar 4.49 Proses inisialisasi struktur data *queries*, *corpus*, dan relevant_docs untuk evaluasi *information retrieval* 167
Gambar 4.50 Pendefinisian evaluator IR menggunakan kelas InformationRetrievalEvaluator 167

Gambar 4.51 Proses pelatihan model *retrieval* dengan fit() menggunakan *evaluator* dan konfigurasi pelatihan. 167
Gambar 4.52 Hasil evaluasi model NSP berdasarkan *confusion matrix* 173
Gambar 4.53 Hasil Evaluasi Model *Intent Classification* berdasarkan *confusion matrix* 175
Gambar 4.54 Tampilan Hasil Simulasi Evaluasi *End-to-End Chatbot* 180

DAFTAR TABEL

Tabel 2.1 Penelitian terkait atau sebelumnya 56
Tabel 4.1 Sample Data Pertanyan dan Jawaban yang didapat dari situs web 80
Tabel 4.2 *Raw Data Corpus*/*Knowledge Base* 86
Tabel 4.3 Spesifikasi Format Data 92
Tabel 4.4 Hasil konkret dan *Sample* data dari pembentukan dataset 93
Tabel 4.5 *Sample* Data Tokenisasi Kalimat pada MLM 97
Tabel 4.6 *Sample* Data Pembersihan Teks pada MLM 98
Tabel 4.7 *Sample* Data Pembersihan Teks pada NSP 102
Tabel 4.8 *Sample* Data Pembersihan Teks pada *Intent Classification* 110
Tabel 4.9 Data-data yang berhasil diubah nilainya dari penerapan metode fit_transform 112
Tabel 4.10 *Sample* Data setelah proses *flattening* dilakukan 114
Tabel 4.11 *Sample* Data Pembersihan Teks pada *Semantic Retrieval* menggunakan format *Triplet Loss* 119
Tabel 4.12 Rincian Pembagian Data pada MLM 120
Tabel 4.13 Rincian Pembagian Data pada NSP 121
Tabel 4.14 Rincian Pembagian Data pada *Intent Classification* 121
Tabel 4.15 Distribusi Jumlah Data setiap *tag* 122
Tabel 4.16 Pembagian Data pada *Extractive* QA SQuAD 123
Tabel 4.17 Pembagian Data pada *Semantic Retrieval* menggunakan *Triplet Loss* 123
Tabel 4.18 Hasil metrik evaluasi model pada berbagai langkah (*steps*) selama proses *fine-tuning* 149
Tabel 4.19 Hasil metrik evaluasi model pada berbagai langkah (*steps*) selama proses *fine-tuning* 155
Tabel 4.20 Hasil metrik evaluasi model pada berbagai *epoch* selama proses *fine-tuning* 159
Tabel 4.21 Hasil metrik evaluasi model pada berbagai *epoch* selama proses *fine-tuning* 164
Tabel 4.22 Hasil metrik evaluasi model pada berbagai langkah (*steps*) selama proses *fine-tuning* 168

Tabel 4.23 Hasil Evaluasi Model MLM 171
Tabel 4.24 Hasil Evaluasi Model NSP 173
Tabel 4.25 Hasil Evaluasi Model *Intent Classification* 175
Tabel 4.26 Hasil Evaluasi Model *Extractive* QA SQuAD 177
Tabel 4.27 Hasil Evaluasi Model *Semantic Retrieval* menggunakan format *Triplet Loss* 179

DAFTAR PUSTAKA

Abdellatif, A., Badran, K., Costa, D. E., & Shihab, E. (2021). A comparison of natural language understanding platforms for chatbots in software engineering. *IEEE Transactions on Software Engineering*, *48*(8), 3087–3102.

Akarsu, K., & Er, O. (2023). Artificial Intelligence Based Chatbot in E-Health System. *Artificial Intelligence Theory and Applications*, *3*(2), 113–122.

Al-Tuama, A. T., & Nasrawi, D. A. (2022). Intent Classification Using Machine Learning Algorithms and Augmented Data. *2022 International Conference on Data Science and Intelligent Computing (ICDSIC)*, 234–239. https://doi.org/10.1109/ICDSIC56987.2022.10075794

Alhammadi, M. (2023). Optimising Customer Service Delivery and Response Time through AI-Enhanced Chatbots in Facilities Management-A Mixed-Methods Research. *American Journal of Smart Technology and Solutions*, *2*, 43–54. https://doi.org/10.54536/ajsts.v2i2.2206

Altalhan, M., Algarni, A., & Turki-Hadj Alouane, M. (2025). Imbalanced Data Problem in Machine Learning: A Review. *IEEE Access*, *13*, 13686–13699. https://doi.org/10.1109/ACCESS.2025.3531662

Alwendi, A. (2020). Penerapan E-Commerce Dalam Meningkatkan Daya Saing Usaha. *Jurnal Manajemen Bisnis*, *17*(3), 317–325.

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamar\'\ia, J., Albahri, A. S., Al-Dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., & others. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, *10*(1), 46.

Amer, E., Hazem, A., Farouk, O., Louca, A., Mohamed, Y., & Ashraf, M. (2021). A Proposed Chatbot Framework for COVID-19. *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 263–268. https://doi.org/10.1109/MIUCC52538.2021.9447652

Armansyah, A., & Siregar, S. (2020). Analisis Penerapan Akuntansi Sumber Daya Manusia Sebagai Aset Pada Lembaga Filantropi. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, *1*(1), 601–606.

Aslam, F. (2023). The impact of artificial intelligence on chatbot technology: A study on the current advancements and leading innovations. *European Journal of Technology*, *7*(3), 62–72.

Assayed, S., Alkhatib, M., & Shaalan, K. (2024). *A Transformer-Based Generative AI Model in Education: Fine-Tuning BERT for Domain-Specific in Student Advising* (hal. 165–174). https://doi.org/10.1007/978-3-031-65996-6_14

Aziz, M. (2022). Pengelolaan Filantropi Islam Untuk Pendidikan Dan Pengentasan Kemiskinan (Menuju Tata Kelola Dana Filantropi Untuk Kemanusiaan). *Al Hikmah: Jurnal Studi Keislaman*, *12*(02), 211–223.

Babu, A., & Boddu, S. B. (2024). BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding. *Exploratory Research in Clinical and Social Pharmacy*, *13*, 100419. https://doi.org/https://doi.org/10.1016/j.rcsop.2024.100419

Baharuddin, F., & Naufal, M. F. (2023). Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy. *Journal of Information Systems Engineering and Business Intelligence*, *9*(2), 253–263.

Bahrain, N., Siti Najah Raihan, S., & Maidin, A. (2023). Communication Barriers in Work Environment: Understanding Impact and Challenges. *International Journal of Academic Research in Business and Social Sciences*, *13*. https://doi.org/10.6007/IJARBSS/v13-i11/19498

Bajaj, P., Campos, D., Craswell, N., Deng, L., Gao, J., Liu, X., Majumder, R., McNamara, A., Mitra, B., Nguyen, T., Rosenberg, M., Song, X., Stoica, A., Tiwary, S., & Wang, T. (2018). *MS MARCO: A Human Generated MAchine Reading COmprehension Dataset*. https://arxiv.org/abs/1611.09268

Bappy, A., Mahmud, T., Kaiser, M. S., Hossain, M., & Andersson, K. (2024, Desember). *A BERT-Based Chatbot to Support Cancer Treatment Follow-Up*. https://doi.org/10.1007/978-3-031-68639-9_4

Barinov, R., Gai, V., Kuznetsov, G., & Golubenko, V. (2023). Automatic evaluation of neural network training results. *Computers*, *12*(2), 26.

Bathija, R., Agarwal, P., Somanna, R., & Pallavi, G. B. (2020). Guided interactive learning through chatbot using bi-directional encoder representations from transformers (BERT). *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 82–87.

Bitton, Y., Stanovsky, G., Elhadad, M., & Schwartz, R. (2021). *Data Efficient Masked Language Modeling for Vision and Language*. https://arxiv.org/abs/2109.02040

Brandtzaeg Petter Baeand Følstad, A. (2017). Why People Use Chatbots. In J. and S. A. and C. G. and P. A. and K. E. and D. S. and M. D. Kompatsiaris Ioannisand Cave (Ed.), *Internet Science* (hal. 377–392). Springer International Publishing.

Buck, R., & VanLear, C. (2002). Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior. *Journal of Communication*, *52*, 522–541. https://doi.org/10.1111/j.1460-2466.2002.tb02560.x

Cai, N., Gao, S., & Yan, J. (2024). How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. *Humanities and Social Sciences Communications*, *11*(1), 1–11.

Carnochan, S., Samples, M., Myers, M., & Austin, M. J. (2014). Performance measurement challenges in nonprofit human service organizations. *Nonprofit and Voluntary Sector Quarterly*, *43*(6), 1014–1032.

Chen, J., Tam, D., Raffel, C., Bansal, M., & Yang, D. (2023). An Empirical Survey of Data Augmentation for Limited Data Learning in NLP. *Transactions of the Association for Computational Linguistics*, *11*, 191–211. https://doi.org/10.1162/tacl_a_00542

Dan, Y., Lei, Z., Gu, Y., Li, Y., Yin, J., Lin, J., Ye, L., Tie, Z., Zhou, Y., Wang, Y., Zhou, A., Zhou, Z., Chen, Q., Zhou, J., He, L., & Qiu, X. (2023). *EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education*. https://arxiv.org/abs/2308.02773

Deng, X., & Yu, Z. (2023). A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. *Sustainability*, *15*(4). https://doi.org/10.3390/su15042940

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186.

Dewi, N. H., Rustiawati, E., & Yupartini, L. (2024). Intervensi Atrisi ASI Eksklusif Berbasis Website (SIASIF) Untuk Mencegah Stunting Pada Balita Di Kabupaten Serang Banten. *Jurnal Pengabdian dan Pengembangan Masyarakat Indonesia*, *3*(2), 165–172.

Dietz, L., Verma, M., Radlinski, F., & Craswell, N. (2017). TREC Complex Answer Retrieval Overview. *TREC*.

Drummond, C., & Holte, R. C. (2006). Cost curves: An improved method for visualizing classifier performance. *Machine learning*, *65*(1), 95–130.

Dutta, S. (2024). Chatbot Effectiveness in Enhancing Guest Communication: Insights from Secondary Data. *Tuijin Jishu/Journal of Propulsion Technology*, *45*, 1001–4055. https://doi.org/10.52783/tjjpt.v45.i04.8497

Fardila, D., & Adawyah, N. R. (2021). Optimasi Biaya dan Waktu Proyek Konstruksi dengan Lembur dan Penambahan Tenaga Kerja. *INERSIA Lnformasi Dan Ekspose Hasil Riset Teknik Sipil Dan Arsitektur*, *17*(1), 35–46.

Fernandez-Martinez, F., Luna-Jiménez, C., Kleinlein, R., Griol, D., Callejas, Z., & Montero, J. M. (2022). Fine-tuning BERT models for intent recognition using a frequency cut-off strategy for domain-specific vocabulary extension. *Applied Sciences*, *12*(3), 1610.

Furqan, M., Sriani, S., & Shidqi, M. N. (2023). Chatbot Telegram Menggunakan Natural Language Processing. *Walisongo Journal of Information Technology*, *5*(1), 15–26.

Galitsky, B., & Galitsky, B. (2021). Adjusting chatbot conversation to user personality and mood. *Artificial Intelligence for Customer Relationship Management: Solving Customer Problems*, 93–127.

Ghosh, M., & Thirugnanam, A. (2021). Introduction to artificial intelligence. *Artificial Intelligence for Information Management: A Healthcare Perspective*, 23–44.

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). Opposing effects of response time in human–chatbot interaction: The moderating role of prior experience. *Business & Information Systems Engineering*, *64*(6), 773–791.

Golchin, S., Surdeanu, M., Tavabi, N., & Kiapour, A. (2023). Do not mask randomly: Effective domain-adaptive pre-training by masking in-domain keywords. *arXiv preprint arXiv:2307.07160*.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

Gunasekara, L., & Vidanage, K. (2019). Uniontbot: Semantic natural language generation based api approach for chatbot communication. *2019 National Information Technology Conference (NITC)*, 1–8.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Han, X., & Eisenstein, J. (2019). *Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling*. https://arxiv.org/abs/1904.02817

Hartati, R., & Manullang, E. B. (2024). Implementation of Telegram Chatbot AI with Natural Language Processing (NLP) in Learning Creative Entrepreneurship to Develop Students' Creative and Innovative Competence. *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, *7*(2), 72–79.

Ho, C. C., Lee, H. L., Lo, W. K., & Lui, K. F. A. (2018). Developing a chatbot for college student programme advisement. *2018 International Symposium on Educational Technology (ISET)*, 52–56.

Idrus, S., Ruhana, F., Amalia, M. R., Rosyid, A. F., & Kuswandi, D. (2023). Implementasi kebijakan manajemen sumber daya manusia yang efektif dalam meningkatkan kinerja organisasi di era bisnis global. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, *7*(1), 72–89.

Izacard, G., & Grave, E. (2020). Leveraging passage retrieval with generative models for open domain question answering. *arXiv preprint arXiv:2007.01282*.

Jawahar, G., Sagot, B., & Seddah, D. (2019, Juli). What does BERT learn about the structure of language? *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics*. https://inria.hal.science/hal-02131630

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of big data*, *6*(1), 1–54.

Karpukhin, V., Oguz, B., Min, S., Lewis, P. S. H., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense Passage Retrieval for Open-Domain Question Answering. *EMNLP (1)*, 6769–6781.

Kasoju, A., & Vishwakarma, T. (2025). Optimizing Transformer Models for Low-Latency Inference: Techniques, Architectures, and Code Implementations. *International Journal of Science and Research (IJSR)*, *14*, 857–866.

Khan, V., & Meenai, T. A. (2021). Pretrained natural language processing model for intent recognition (bert-ir). *Human-Centric Intelligent Systems*, *1*(3), 66–74.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713–3744.

Koto, F., Lau, J. H., & Baldwin, T. (2021). Indobertweet: A pretrained language model for indonesian twitter with effective domain-specific vocabulary initialization. *arXiv preprint arXiv:2109.04607*.

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *arXiv preprint arXiv:2011.00677*.

Kurniadi, W., & Mahaputra, M. R. (2021). Determination of communication in the organization: Non verbal, oral and written (literature review). *Journal of Law, Politic and Humanities*, *1*(4), 164–172.

Lapin, M., Hein, M., & Schiele, B. (2016, Juni). Loss Functions for Top-k Error: Analysis and Insights. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Larson, S., & Leach, K. (2022). A survey of intent classification and slot-filling datasets for task-oriented dialog. *arXiv preprint arXiv:2207.13211*.

Łepicki, M., Latkowski, T., Antoniuk, I., Bukowski, M., Świderski, B., Baranik, G., Nowak, B., Zakowicz, R., Dobrakowski, Ł., Act, B., & others. (2025). Comparative Evaluation of Sequential Neural Network (GRU, LSTM, Transformer) Within Siamese Networks for Enhanced Job–Candidate

Matching in Applied Recruitment Systems. *Applied Sciences*, *15*(11), 5988.

Liddy, E. D. (2001). *Natural language processing*.

Liu, Z., Li, G., & Cheng, J. (2021). Hardware acceleration of fully quantized bert for efficient natural language processing. *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 513–516.

Luhmann, N. (1992). What is communication? *Communication theory*, *2*(3), 251–259.

Ma, X., Xu, P., Wang, Z., Nallapati, R., & Xiang, B. (2019). Domain Adaptation with BERT-based Domain Classification and Data Selection. In C. Cherry, G. Durrett, G. Foster, R. Haffari, S. Khadivi, N. Peng, X. Ren, & S. Swayamdipta (Ed.), *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)* (hal. 76–83). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-6109

Maharani, N. P. I., Purwarianti, A., Yustiawan, Y., & Rochim, F. C. (2023). Domain-Specific Language Model Post-Training for Indonesian Financial NLP. *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*, 1–6.

Manzi, F., Schellenberg, J. A., Hutton, G., Wyss, K., Mbuya, C., Shirima, K., Mshinda, H., Tanner, M., & Schellenberg, D. (2012). Human resources for health care delivery in Tanzania: a multifaceted problem. *Human resources for health*, *10*, 1–10.

McCarthy, J., & others. (2007). *What is artificial intelligence*.

Moffat, A. (2022). Batch evaluation metrics in information retrieval: Measures, scales, and meaning. *IEEE Access*, *10*, 105564–105577.

Mondal, A., Dey, M., Das, D., Nagpal, S., & Garda, K. (2018). Chatbot: An automated conversation system for the educational domain. *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 1–5.

Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2020). A review of current trends in the development of chatbot systems. *2020 6th International conference on advanced computing and communication systems (ICACCS)*, 706–710.

Nguyen, T. T., Le, A. D., Hoang, H. T., & Nguyen, T. (2021). NEU-chatbot: Chatbot for admission of National Economics University. *Computers and Education: Artificial Intelligence*, *2*, 100036. https://doi.org/https://doi.org/10.1016/j.caeai.2021.100036

Niu, S., Guo, J., Lan, Y., & Cheng, X. (2012). Top-k learning to rank: labeling, ranking and evaluation. *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 751–760.

Nkhata, G., Gauch, S., Anjum, U., & Zhan, J. (2025). *Fine-tuning BERT with Bidirectional LSTM for Fine-grained Movie Reviews Sentiment Analysis*. https://doi.org/10.48550/arXiv.2502.20682

Nogueira, R., & Cho, K. (2019). Passage Re-ranking with BERT. *arXiv preprint arXiv:1901.04085*.

Ortiz-Garces, I., Govea, J., Andrade, R. O., & Villegas-Ch, W. (2024). Optimizing chatbot effectiveness through advanced syntactic analysis: A comprehensive study in natural language processing. *Applied Sciences*, *14*(5), 1737.

Perwira, R. I., Permadi, V. A., Purnamasari, D. I., Agusdin, R. P., & others. (2025). Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content. *Journal of Information Systems Engineering and Business Intelligence*, *11*(1), 30–40.

Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Prasad, A., Bui, T., Yoon, S., Deilamsalehy, H., Dernoncourt, F., & Bansal, M. (2023). Meetingqa: Extractive question-answering on meeting transcripts. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 15000–15025.

Quan-Haase, A. (2007). University students' local and distant social ties: Using and integrating modes of communication on campus. *Information, Communication & Society*, *10*(5), 671–693.

Rachman, A., Mardhiyah, I., & Jannah, M. (2023). Implementasi Chatbot FAQ pada Aplikasi Monev Kinerja Direktorat Jenderal Anggaran Menggunakan Framework Rasa Open Source. *KLIK: Kajian Ilmiah Informatika dan Komputer*, *4*(1), 62–72.

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.

Reddi, V. J., Cheng, C., Kanter, D., Mattson, P., Schmuelling, G., Wu, C.-J., Anderson, B., Breughe, M., Charlebois, M., Chou, W., Chukka, R., Coleman, C., Davis, S., Deng, P., Diamos, G., Duke, J., Fick, D., Gardner, J. S., Hubara, I., … Zhou, Y. (2020). MLPerf Inference Benchmark. *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*, 446–459. https://doi.org/10.1109/ISCA45697.2020.00045

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Ren, R., Lv, S., Qu, Y., Liu, J., Zhao, W. X., She, Q., Wu, H., Wang, H., & Wen, J.-R. (2021). PAIR: Leveraging Passage-Centric Similarity Relation for

Improving Dense Passage Retrieval. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2173–2183. https://doi.org/10.18653/v1/2021.findings-acl.191

Rietzler, A., Stabinger, S., Opitz, P., & Engl, S. (2019). Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860*.

Risch, J., Möller, T., Gutsch, J., & Pietsch, M. (2021). *Semantic Answer Similarity for Evaluating Question Answering Models*. https://arxiv.org/abs/2108.06130

Ritonga, M. K., Toni, T., Zahara, S., Yani, F., & Nadia, N. (2022). Minat Baca dan Minat Berkunjung Mahasiswa ke Perpustakaan TP 2021/2022. *CIVITAS (JURNAL PEMBELAJARAN DAN ILMU CIVIC)*, *8*(1), 36–45.

Rudas, I. J., & Fodor, J. (2008). Intelligent systems. *International Journal of Computers, Communications & Control*, *3*(3), 132–138.

Ruiz, E., Torres, M. I., & del Pozo, A. (2023). Question answering models for human–machine interaction in the manufacturing industry. *Computers in Industry*, *151*, 103988.

Sachidananda, V., Kessler, J. S., & Lai, Y.-A. (2021). Efficient domain adaptation of language models via adaptive tokenization. *arXiv preprint arXiv:2109.07460*.

Sahu, G., Rodriguez, P., Laradji, I. H., Atighehchian, P., Vazquez, D., & Bahdanau, D. (2022). *Data Augmentation for Intent Classification with Off-the-shelf Large Language Models*. https://arxiv.org/abs/2204.01959

Schwartz, I. (2021). *Ensemble of MRR and NDCG models for Visual Dialog*. https://arxiv.org/abs/2104.07511

Shawar, B., & Atwell, E. (2007). Chatbots: Are they Really Useful? *LDV Forum*, *22*, 29–49. https://doi.org/10.21248/jlcl.22.2007.88

Shingte, K., Chaudhari, A., Patil, A., Chaudhari, A., & Desai, S. (2021). Chatbot development for educational institute. *Available at SSRN 3861241*.

Smith, E. A., Gearhart, G. D., & Miller, M. T. (2019). Understanding Alumni Relations Programs in Community Colleges. *International Journal of Higher Education*, *8*(5), 176–184.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/https://doi.org/10.1016/j.ipm.2009.03.002

Su, P., & Vijay-Shanker, K. (2022). Investigation of improving the pre-training and fine-tuning of BERT model for biomedical relation extraction. *BMC bioinformatics*, *23*(1), 120.

Sun, Y., Zheng, Y., Hao, C., & Qiu, H. (2021). NSP-BERT: A prompt-based few-shot learner through an original pre-training task–next sentence prediction. *arXiv preprint arXiv:2109.03564*.

Swaminathan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, *27*, 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Thorat, S. A., & Jadhav, V. (2020). A review on implementation issues of rule-based chatbot systems. *Proceedings of the international conference on innovative computing & communications (ICICC)*.

Vlachopoulos, D., & Makri, A. (2019). Online communication and interaction in distance higher education: A framework study of good practice. *International Review of Education*, *65*(4), 605–632.

Vujovic, Z. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, *Volume 12*, 599–606. https://doi.org/10.14569/IJACSA.2021.0120670

Wettig, A., Gao, T., Zhong, Z., & Chen, D. (2022). Should you mask 15% in masked language modeling? *arXiv preprint arXiv:2202.08005*.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & others. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv:2009.05387*.

Williams, C. K. I. (2021). The Effect of Class Imbalance on Precision-Recall Curves. *Neural Computation*, *33*(4), 853–857. https://doi.org/10.1162/neco_a_01362

Xiong, L., Xiong, C., Li, Y., Tang, K.-F., Liu, J., Bennett, P., Ahmed, J., & Overwijk, A. (2020). Approximate nearest neighbor negative contrastive learning for dense text retrieval. *arXiv preprint arXiv:2007.00808*.

Yang, Z., Gao, W., Luo, C., Wang, L., Tang, F., Wen, X., & Zhan, J. (2024). *Quality at the Tail of Machine Learning Inference*. https://arxiv.org/abs/2212.13925

Yuniar, E., & Purnomo, H. (2019). Implementasi Chatbot" Alitta" Asisten Virtual Dari Balittas Sebagai Pusat Informasi Di Balittas. *Antivirus: Jurnal Ilmiah Teknik Informatika*, *13*(1), 24–35.

Zhu, B., & Rao, Y. (2023). Exploring Robust Overfitting for Pre-trained Language Models. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Ed.), *Findings of the Association for Computational Linguistics: ACL 2023* (hal. 5506–5522).

Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.340