



Evaluasi dan Implementasi *Indobert Question Answering (QA)* pada Domain Spesifik Menggunakan *Mean Reciprocal Rank*

Teguh Ikhlas Ramadhan^{1*}, Agus Supriatman², Taufik Rahmat Kurniawan³

^{1,2}Universitas Perjuangan Tasikmalaya, Indonesia

³Universitas Muhammadiyah Bandung, Indonesia

*email: teguhikhlas@unper.ac.id

Info Artikel

Dikirim: 9 November, 2023

Diterima: 25 Maret 2024

Diterbitkan: 18 Mei 2024

Kata kunci:

Index Terms-BERT;

Mean Reciprocal Rank;

Question Answering.

ABSTRAK

Penerapan kecerdasan buatan, seperti Chat-GPT dan Bard, telah menjadi umum dalam berbagai aspek kehidupan saat ini. Salah satu aspek utamanya adalah dalam pemanfaatan model *Question Answering (QA)* untuk memenuhi kebutuhan domain spesifik. Namun, dalam beberapa kasus, model seperti Bard mungkin tidak dapat memberikan informasi spesifik, seperti prosedur pendaftaran di perusahaan tertentu atau jadwal dosen di universitas. Untuk mengatasi tantangan ini, telah dikembangkan model QA bahasa Indonesia bernama IndoBERT-QA. Penelitian ini bertujuan untuk mengevaluasi kemampuan IndoBERT-QA dalam konteks domain spesifik dengan menggunakan metode evaluasi *Mean Reciprocal Rank (MRR)*. Hasil evaluasi menunjukkan bahwa model IndoBERT-QA mampu mencapai MRR sebesar 0.91 dengan pendekatan membuat konteks yang disesuaikan untuk setiap pertanyaan. Hasil ini menandakan bahwa model ini memiliki kinerja yang baik dalam memberikan jawaban yang relevan. Manfaat dari penelitian ini adalah sebagai acuan yang berharga bagi pengembangan sistem QA yang akan dibuat oleh pihak-pihak lain dalam lingkungan yang sama. Dengan memanfaatkan pendekatan yang telah diuji dan dievaluasi dengan baik, penelitian ini memberikan dasar yang kuat untuk pengembangan sistem QA berkinerja tinggi dalam bahasa Indonesia, sehingga dapat memenuhi kebutuhan domain spesifik. Disamping itu, penelitian ini juga memberikan pandangan untuk pengembangan lebih lanjut. Salah satu pendekatan yang dapat dieksplorasi adalah penggunaan Information Retrieval atau Passage Retrieval sebagai langkah awal dalam proses QA. Hal ini dapat membantu model dalam mendapatkan konteks yang lebih tepat dan relevan, sehingga memungkinkan peningkatan lebih lanjut dalam kualitas jawaban yang diberikan oleh model. Index Terms-BERT, Mean Reciprocal Rank, *Question Answering* Namun, penelitian ini juga memberikan pandangan untuk pengembangan lebih lanjut. Salah satu pendekatan yang dapat dieksplorasi adalah penggunaan *Information Retrieval* atau *Passage Retrieval* sebagai langkah awal dalam proses QA. Hal ini dapat membantu model dalam mendapatkan konteks yang lebih tepat dan relevan, sehingga memungkinkan peningkatan lebih lanjut dalam kualitas jawaban yang diberikan oleh model.

1. PENDAHULUAN

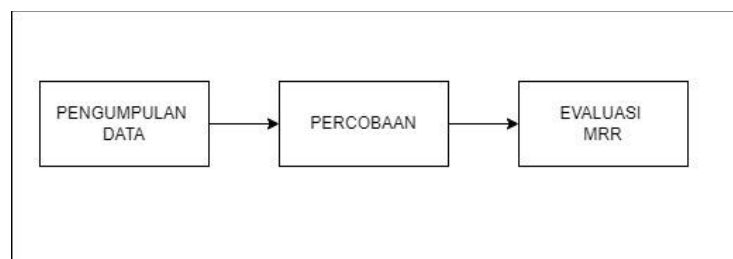
Pengembangan teknologi kecerdasan buatan saat ini telah menghasilkan berbagai macam aplikasi yang mampu memproses bahasa alami dengan semakin baik. Salah satu aplikasi dari kecerdasan buatan adalah sistem *Question Answering* (QA)[1], yang bertujuan untuk menjawab pertanyaan yang diberikan oleh pengguna dengan menggunakan bahasa alami[2]. IndoBERT-QA[3] merupakan model bahasa alami yang dibangun dengan menggunakan IndoBERT[4] sebagai basisnya, dan dikembangkan untuk tugas *Question Answering* (QA) dengan menggunakan dataset hasil terjemahan dari SQuAD 2.0[5] ke dalam bahasa Indonesia. Selain itu, IndoBERT- QA juga mampu diadaptasi untuk berbagai macam domain spesifik, seperti kesehatan, hukum, dan sebagainya[6].

Pengimplementasian sistem *Question Answering* (QA) dalam domain spesifik menjadi suatu tantangan tersendiri dalam pengembangan teknologi kecerdasan buatan.[7]. Salah satu masalah yang dihadapi adalah bagaimana mengadaptasi model bahasa alami yang sudah ada untuk dapat memberikan jawaban yang akurat dan relevan terhadap pertanyaan yang diajukan dalam konteks domain spesifik tersebut. Selain itu, model yang ada juga harus diuji kembali untuk memastikan kinerjanya dalam mengatasi persoalan yang spesifik dalam domain yang dihadapi. Oleh karena itu, perlu dilakukan evaluasi terhadap model yang sudah ada dan pengujian ulang untuk menyesuaikan dengan domain yang spesifik tersebut.[8]. Evaluasi ini dapat dilakukan dengan metode *Mean Reciprocal Rank* (MRR) untuk mengukur keberhasilan sistem dalam memberikan jawaban yang relevan terhadap pertanyaan yang diajukan[9]. Dengan mengatasi masalah ini, diharapkan sistem QA dapat memberikan solusi yang lebih akurat dan relevan dalam menjawab pertanyaan yang sering diajukan oleh pengguna dalam domain spesifik, seperti di kampus.

Penelitian ini bertujuan untuk mengevaluasi dan mengimplementasikan IndoBERT-QA pada domain spesifik di internal kampus menggunakan metode *Mean Reciprocal Rank* (MRR)[10]. Dengan menggunakan IndoBERT-QA, diharapkan sistem QA dapat memberikan jawaban yang lebih akurat dan relevan dengan pertanyaan yang diajukan oleh pengguna. Selain itu, dengan menggunakan metode MRR, dapat diukur keberhasilan sistem dalam memberikan jawaban yang relevan dengan pertanyaan yang diberikan.

Setelah dilakukan pengembangan dan evaluasi model IndoBERT-QA[11] pada domain spesifik persoalan internal kampus dengan menggunakan *Mean Reciprocal Rank* (MRR)[12], model tersebut kemudian diimplementasikan menjadi sebuah website sederhana yang dapat diakses oleh pengguna kampus atau bahkan masyarakat luas. Website tersebut memungkinkan pengguna kampus untuk mengajukan pertanyaan terkait dengan persoalan kampus dan mendapatkan jawaban yang relevan dan akurat dari sistem QA yang telah dikembangkan. Penelitian yang dilakukan oleh Rahajeng (2021) [6] dan Ahmad 2023 [13] telah menginvestigasi kinerja IndoBERT dalam konteks task *Question Answering*. Namun, perbedaan utama dengan penelitian ini adalah bahwa IndoBERT yang digunakan dalam penelitian lain merupakan versi umum dari model tersebut, sedangkan penelitian ini menggunakan varian yang khusus disesuaikan untuk *Question Answering*, yaitu IndoBERT-QA.

2. METODE PENELITIAN



Gambar 1. Metodologi Penelitian

Gambar 1 menggambarkan langkah-langkah metodologi yang diterapkan dalam penelitian ini. Proses dimulai dengan pengumpulan data yang akan menjadi sumber referensi untuk model QA, yang terdiri dari teks yang relevan serta kumpulan pertanyaan beserta jawabannya. Kemudian, dilakukan serangkaian percobaan untuk

mengamati hasil yang dihasilkan oleh model IndoBERT-QA dari pertanyaan yang telah terkumpul. Terakhir, hasil jawaban yang dihasilkan oleh model dievaluasi menggunakan metrik MRR untuk menilai kualitas model IndoBERT-QA dan kecukupannya untuk diimplementasikan.

2.1 Pengumpulan Data

Dalam pengujian untuk kasus yang spesifik, langkah awal yang dilakukan adalah mengumpulkan data sebanyak-banyaknya yang relevan dengan kasus tersebut. Pada penelitian ini, data yang dikumpulkan terdiri dari konteks, pertanyaan, dan jawaban. Untuk data konteks, akan dikumpulkan semua informasi dan pengetahuan mengenai Universitas Perjuangan dari berbagai sumber baik internal maupun eksternal. Selanjutnya, konteks tersebut akan digabungkan menjadi satu dan pertanyaan serta jawaban akan dibuat berdasarkan konteks yang sudah dikumpulkan tersebut.

Proses pembuatan pertanyaan mengacu pada jenis pertanyaan yang umum dan khusus sehingga menghasilkan variasi pertanyaan yang umum dan khusus sehingga menghasilkan variasi pertanyaan yang cukup banyak. Selain itu, jawaban akan dibuat dalam 3 versi yang berbeda untuk memudahkan pengujian menggunakan metrik *Mean Reciprocal Rank* (MRR). Dengan begitu, model IndoBERT-QA yang digunakan dalam pengujian dapat menghasilkan jawaban yang akurat dan relevan dengan pertanyaan yang diberikan. Hal ini perlu dilakukan agar data yang digunakan dalam pengujian dapat mewakili kasus yang akan diimplementasikan, serta pertanyaan dan jawaban yang dibuat sesuai dengan konteks yang relevan. Dengan demikian, hasil pengujian yang diperoleh dapat diandalkan dan akurat, sehingga model IndoBERT-QA dapat diimplementasikan dengan efektif dalam kasus yang spesifik.

2.2 Metrik Evaluasi

Mean Reciprocal Rank (MRR) adalah metrik evaluasi kinerja untuk model *Question Answering* (QA). Metrik ini digunakan untuk mengukur kemampuan model QA dalam memberikan jawaban yang benar untuk pertanyaan yang diberikan. MRR memperhitungkan peringkat jawaban yang benar dalam daftar jawaban yang dihasilkan oleh model. Semakin tinggi nilai MRR, semakin baik kinerja model dalam memberikan jawaban yang benar.

Rumus MRR dapat dinyatakan sebagai berikut:

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i} \tag{1}$$

Di mana $|Q|$ adalah jumlah pertanyaan yang dievaluasi, $rank$ adalah peringkat pertama jawaban yang benar untuk pertanyaan ke- i . Semakin rendah nilai $rank_i$, semakin baik kinerja model.

Contoh perhitungan MRR:

Misalkan terdapat 3 pertanyaan yang diberikan dan model QA menghasilkan 5 jawaban untuk setiap pertanyaan. Jawaban yang benar diberi tanda bintang (*) pada peringkat pertama.

Tabel 1. Contoh pertanyaan dan jawaban dalam perhitungan MRR.

No	Pertanyaan	Jawaban
1	Siapa presiden pertama Indonesia?	*Soekarno, Jokowi, Megawati,
2	Berapa jumlah provinsi di Indonesia?	*34, 33, 32, 31, 30
3	Apa ibukota Indonesia?	*Jakarta, Surabaya, Bandung

Perhitungan MRR:

Pertama dari Tabel 1 ditentukan terlebih dahulu peringkat jawaban yang benar untuk setiap pertanyaan:

- Pertanyaan 1: Jawaban benar berada pada peringkat 1
- Pertanyaan 2: Jawaban benar berada pada peringkat 1
- Pertanyaan 3: Jawaban benar berada pada peringkat 1 Kedua, menghitung nilai MRR dengan menggunakan rumus yang diberikan:

Kedua, kita menghitung nilai MRR dengan menggunakan rumus yang diberikan:

RUMUS MRR

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

$$MRR = \frac{1}{3} \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} \right)$$

$$MRR = \frac{1}{3} (3)$$

$$MRR = 1$$

Dalam contoh berikut, nilai MRR adalah 1, yang menunjukkan bahwa model QA memberikan jawaban yang benar pada peringkat pertama untuk semua pertanyaan yang dievaluasi.

Apabila seandainya model menjawab pertanyaan kedua pada peringkat kedua maka:

- Pertanyaan 1: Jawaban benar berada pada peringkat 1
- Pertanyaan 2: Jawaban benar berada pada peringkat 2
- Pertanyaan 3: Jawaban benar berada pada peringkat 1

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

$$MRR = \frac{1}{3} \left(\frac{1}{1} + \frac{1}{2} + \frac{1}{1} \right)$$

$$MRR = \frac{1}{3} (2.5)$$

$$MRR = 0.83$$

Nilai MRR adalah 0.83, yang menunjukkan bahwa model QA memberikan jawaban yang benar pada peringkat pertama untuk 2 pertanyaan dari 3 pertanyaan yang di evaluasi.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Data yang dikumpulkan berupa fakta-fakta yang terkait dengan domain spesifik yang dimaksud, yaitu informasi seputar Universitas Perjuangan. Data ini berbentuk deskripsi dalam bentuk teks mengenai berbagai aspek yang terkait dengan universitas tersebut. Pada Tabel 2 terlihat konteks dalam beberapa bidang yang akan dikumpulkan menjadi satu para-graf atau satu file .txt untuk diujicoba dalam model IndoBERT QA. Jumlah kata yang telah berhasil dikumpulkan sebanyak 5671 kata dan disimpan dalam satu file berformat .txt.

Tabel 2. Contoh Data berupa konteks yang dikumpulkan, contoh dari tiga bidang saja yaitu lembaga, ormawa dan sistem pembayaran

Bidang	Konteks
Lembaga	Yayasan adalah badan hukum yang terdiri dari kekayaan yang dipisahkan dan diperuntukkan untuk mencapai tujuan tertentu dibidang sosial, keagamaan dan kemanusiaan yang tidak mempunyai anggota Pada tanggal 25 Mei 2015 Bapak H. Oman Roesman, Drs diangkat menjadi anggota Pembina YUS dan kepengurusan pada saat itu (periode 2015-2020) menjadi...
Ormawa	DESKRIPSI Badan Eksekutif Mahasiswa Republik Mahasiswa Universitas Perjuangan merupakan Eksekutif tertinggi di Ormawa Universitas Perjuangan Taskmalaya yang berdiri pada tahun 2015. VISI Revitalisasi BEM REMA UNPER dalam sinergi adaktif untuk peningkatan SDM demi kebermanfaatan UNPER...
Sistem Pembayaran	<ol style="list-style-type: none"> 1. Buka Halaman web: www.simantap.unper.ac.id kemudian login dengan cara masukan username (NIM) dan Password; 2. Klik Profile di pojok atas sebelah kanan; 2. 3. Pastikan Nomor Virtual Account (NOVA) sudah ada, sebagai nomor rekening tujuan untuk pembayaran registrasi keuangan.

Dalam penelitian ini, akan dibuat pertanyaan berdasarkan kumpulan konteks untuk menguji apakah model mampu menjawabnya dengan baik. Pertanyaan yang akan dibuat bersifat sederhana dan tidak memerlukan jawaban yang terlalu panjang. Contoh pertanyaan yang akan digunakan dapat ditemukan pada Tabel 3, dan pertanyaan ini akan diajukan kepada model untuk diuji. Selanjutnya, setiap jawaban yang dihasilkan oleh model akan dinilai dengan menggunakan kaidah MRR. Sebanyak 52 pertanyaan dasar telah dikumpulkan dalam penelitian ini, yang tidak memerlukan jawaban yang terlalu panjang.

Tabel 3. Contoh pertanyaan yang dibangun untuk uji coba pada model.

No	Pertanyaan	Jawaban
1	Dimanakah letak kampus universitas perjuangan?	Tasikmalaya
2	Kapan berdirinya universitas perjuangan?	26 Januari 2016
3	Berapakah kode universitas perjuangan pada laman lldikti ?	41063
4	Sebutkan website resmi dari universitas perjuangan ?	https://unper.ac.id
5	Universitas Perjuangan berdiri atas surat mana ?	SK Mendikbud RI No. 603/E/O/2014
6	Apa email resmi dari universitas perjuangan ?	Himpunan Mahasiswa Peternakan

3.2 Percobaan Tahap Pertama

Model telah diuji menggunakan konteks berisi 5671 kata terkait informasi Universitas Perjuangan. Selanjutnya, model diberikan 52 pertanyaan yang jawabannya terdapat dalam konteks tersebut. Hasil dari uji coba ini dicatat dalam Tabel 4, termasuk pertanyaan, jawaban yang dihasilkan oleh model, dan skor yang diberikan. Penilaian skor dilakukan secara manual oleh manusia dan tidak dihasilkan secara otomatis oleh sistem, bertujuan untuk memperoleh hasil yang akurat mengenai relevansi jawaban.

Tabel 4. Jawaban yang dihasilkan oleh model dengan skor 1 berarti benar, skor 2 berarti salah tetapi masih relevan, skor 3 berarti salah dan tidak relevan jawabannya.

No	Pertanyaan	Jawaban	Skor
1	Dimanakah Letak Kampus Universitas Perjuangan?	Tasikmalaya	1
2	Kapan Berdirinya Universitas Perjuangan?	26 Januari 2016	2
3	Berapakah Kode Universitas Perjuangan pada laman LLDIKTI	41063	1
4	Sebutkan website resmi dari universitas perjuangan	SK Mendikbud RI No. 603/E/O/2014	1
5	Apa email resmi dari universitas perjuangan	Himpunan Mahasiswa Peternakan	3

Nilai Mean Reciprocal Rank (MRR) yang dihasilkan dari hasil pengujian tersebut adalah sebesar 0.59, seperti yang tercantum dalam Tabel V Dalam hasil tersebut, model banyak memberikan jawaban yang salah sebanyak 39 dari 52 pertanyaan, yang artinya lebih dari setengah pertanyaan dijawab dengan tidak benar oleh model. Waktu eksekusi model juga cenderung lama, yaitu sekitar 42 menit 28 detik untuk menjawab 52 pertanyaan, atau rata-rata sekitar 49 detik untuk setiap pertanyaan yang dieksekusi oleh model.

Tabel 5. Statistik hasil model dan perhitungan MRR

No	Jumlah	Porsi	Skor	Penjumlahan Skor
1	13	0.25	Tasikmalaya	13
2	29	0.557692	0.5	14.5
3	10	0.19230	0.333333	3.33333333
			MRR	0.592948718

Hasil yang diperoleh dari model dengan nilai MRR sebesar 0.59 menunjukkan bahwa model masih memiliki keterbatasan dalam menjawab pertanyaan sederhana dengan benar. Untuk mengimplementasikannya sebagai perangkat lunak, target nilai MRR seharusnya mencapai rentang 0.90 hingga 1. Waktu eksekusi model yang memakan waktu sekitar 49 detik per pertanyaan dianggap terlalu lama, terutama jika dibandingkan dengan kemampuan ChatGPT [14] yang memberikan jawaban hampir instan.

Salah satu asumsi sementara adalah bahwa model mungkin mengalami kesulitan karena memproses konteks yang terlalu panjang. Oleh karena itu, pada percobaan tahap kedua akan mencoba dengan mengurangi panjang konteks sesuai dengan pertanyaan yang diajukan.

3.3 Percobaan Tahap Kedua

Pada percobaan tahap kedua ini model akan diuji tanpa menggunakan konteks yang digabung seperti pada percobaan tahap pertama namun konteks-nya dibagi per-pertanyaan dengan memotong jumlah kata yang ada pada konteks seperti yang ada pada Tabel 6.

Tabel 6. Setiap pertanyaan memiliki konteks-nya masing-masing dengan kalimat yang lebih singkat

No	Pertanyaan	Konteks
1	Dimanakah Letak Kampus Universitas Perjuangan?	Letak Kampus Perjuangan Adalah Di Jalan PETA No. 177 Kecamatan Tawang Kota Tasikmalaya
2	Kapan Berdirinya Universitas Perjuangan?	Universitas Perjuangan Berdiri Pada Tanggal 17 Oktober 2014
3	Berapakah Kode Universitas Perjuangan Pada Laman LLDIKTI?	Universitas Perjuangan Memiliki Kode 041063 Pada Laman Website LLDIKTI
4	Sebutkan Website Resmi Dari Universitas Perjuangan?	Website Resmi Universitas Perjuangan Adalah https://unper.ac.id
5	Universitas perjuangan berdiri atas surat mana?	UNPER berdiri atas SK Mendikbud RI No.603/E/O tahun 2014
6	Apa email resmi dari universitas perjuangan?	Unper memiliki email resmi yaitu perjuangan@unper.ac.id sebagai narahubung kepada beberapa pihak

Hasil dari model untuk enam pertanyaan pertama terdapat perubahan, di mana pertanyaan kedua dan keenam sekarang memiliki skor satu atau dapat dijawab dengan sempurna oleh model, seperti yang terlihat dalam Tabel 7.

Tabel 7. Jawaban Yang Dihasilkan Model Pada Percobaan Tahap Kedua

No	Pertanyaan	Jawaban	Skor
1	Dimanakah Letak Kampus Universitas Perjuangan?	Tasikmalaya	1
2	Kapan Berdirinya Universitas Perjuangan?	17 Oktober 2014	1
3	Berapakah Kode Universitas Perjuangan Pada Laman LLDIKTI?	41063	1
4	Sebutkan Website Resmi Dari Universitas Perjuangan?	https://unper.ac.id	1
5	Universitas perjuangan berdiri atas surat mana?	SK Mendikbud RI No.603/E/O/2014	1
6	Apa email resmi dari universitas perjuangan?	perjuangan@unper.ac.id	1

Pada Tabel 8 memperlihatkan yang telah model hasilkan pada percobaan tahan kedua. Terdapat perubahan yang signifikan baik dari waktu eksekusi maupun hasil MRR. Untuk memproses 52 pertanyaan dan menghasilkan jawabannya dibutuhkan waktu 9 detik artinya kurang dari 0.18 detik setiap satu pertanyaan. Nilai MRR yang dihasilkan menjadi 0.91 dengan selisih MRR dari yang pertama sebanyak 0.32 dan ini bisa dikatakan cukup signifikan.

Terlihat pada percobaan kedua lebih banyak jawaban yang dihasilkan pada rank pertama yang dimana menandakan bahwa pada percobaan kedua model bisa menjawab pertanyaan dengan benar dari target pembuatan model Question-Answering. Secara keseluruhan, untuk hasil dari percobaan tahap pertama dan tahan kedua ada pada Tabel 9

Tabel 8. Statistika Hasil Model dan Perhitungan MRR Tahap Kedua

No	Jumlah	Porsi	Skor	Penjumlahan Skor
1	43	0.826923	1	43
2	8	0.153846	0.5	4
3	1	0.019231	0.333333	0.333333
			MRR	0.910256

Tabel 9. Hasil Rekap Dari Percobaan Tahap Pertama Dan Tahap Kedua, Yang Dimana Percobaan Tahap Kedua Menghasilkan Perubahan Yang Signifikan Lebih Baik.

Nilai	Jumlah	Percobaan		Penjumlahan Skor Tahap 1 – Tahap 2
		Tahap 1	Tahap 2	
Waktu	Detik	49	0.17	48.83
Skor 1	Pertanyaan	13	43	30
Skor 2	Pertanyaan	29	8	21
Skor 3	Pertanyaan	10	1	9
Evaluasi	MRR	0.59	0.91	0.32

Dalam hal waktu eksekusi dan nilai MRR. Untuk memproses 52 pertanyaan dan menghasilkan jawabannya, dibutuhkan waktu 9 detik, yang artinya sekitar 0.18 detik untuk setiap pertanyaan. Nilai MRR yang dihasilkan juga menjadi 0.91, dengan selisih MRR dari tahap pertama sebesar 0.32, yang dapat dianggap cukup signifikan. Terlihat bahwa dalam percobaan kedua, lebih banyak jawaban dihasilkan dalam peringkat pertama, yang menunjukkan bahwa model dapat menjawab lebih banyak pertanyaan dengan benar, yang merupakan tujuan utama dalam pengembangan model *Question Answering*. Secara keseluruhan, hasil dari tahap pertama dan tahap kedua dapat ditemukan dalam Tabel 9. Dalam penelitian yang dilakukan oleh Kusumaningrum [9] pada tahun 2023, didapati bahwa nilai MRR yang diperoleh mencapai 0.8205 untuk jenis pertanyaan non-factoid, mirip dengan temuan dalam penelitian ini. Penelitian tersebut memanfaatkan model BILSTM yang dikembangkan sendiri tanpa menggunakan model yang sudah ada seperti IndoBERT-QA, serta tidak memerlukan penambahan konteks ke dalam model tersebut [15]. Penelitian ini cukup berpotensi apabila secara otomatis mendapatkan konteks yang tersedia pada dokumen acuan yang telah disediakan [16].

4. KESIMPULAN

Secara keseluruhan, IndoBERT-QA telah berhasil mengembangkan model *Question Answering* untuk domain yang spesifik, yaitu informasi yang berkaitan dengan Universitas Perjuangan. Terdapat perbedaan signifikan dalam hasil yang diberikan oleh model, terutama terkait dengan implementasi model itu sendiri. Dari dua tahap percobaan yang dilakukan, dapat ditarik beberapa kesimpulan penting. Tahap kedua percobaan menghasilkan model yang lebih baik, terutama karena pendekatan membagi konteks masing-masing pertanyaan. Kelemahan tahap pertama terletak pada eksekusi model yang memerlukan waktu lebih lama dan hasil yang kurang memuaskan. Namun, tahap pertama memiliki keunggulan dalam penggabungan langsung dari konteks tanpa perlu memisahkan setiap pertanyaan. Di sisi lain, tahap kedua memberikan hasil yang lebih baik dengan mencapai nilai MRR yang tinggi, mendekati 1, yang merupakan hasil maksimal yang sangat positif. Namun,

tahap kedua membutuhkan pemisahan konteks per kalimat, yang tidak dapat diimplementasikan langsung karena pengguna tidak selalu menginputkan konteks secara manual, sehingga diperlukan sistem otomatis untuk menghasilkan konteks yang relevan.

Kesimpulannya, peluang kedepannya terbuka lebar untuk mengembangkan sistem *Information Retrieval* atau *Passage Retrieval* terlebih dahulu, yang dapat membantu dalam memberikan konteks yang sesuai dan relevan sesuai dengan pertanyaan yang diajukan. Hal ini akan mempercepat waktu eksekusi dan meningkatkan hasil model *Question Answering* secara signifikan.

UCAPAN TERIMA KASIH

Kami ingin menyampaikan rasa terimakasih sebesar-besarnya kepada Dicky Totti Juniferdyaz dan Sri Agustina Dewi atas kerja sama yang sangat berharga dalam pengumpulan data informasi terkait Universitas Perjuangan.

REFERENSI

- [1] B. Richardson and A. Wicaksana, "Comparison Of Indobert-Lite And Roberta In Text Mining For Indonesian Language Question Answering Application".
- [2] Z. A. Guven and M. O. Unalir, "Natural language based analysis of SQuAD: An analytical approach for BERT," *Expert Syst Appl*, vol. 195, p. 116592, 2022.
- [3] rifkybujana, "IndoBERT-QA." 2021.
- [4] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th COLING*, 2020.
- [5] S. Li, R. Li, and V. Peng, "Ensemble ALBERT on SQuAD 2.0," *arXiv preprint arXiv:2110.09665*, 2021.
- [6] M. I. Rahajeng and A. Purwarianti, "Indonesian Question Answering System for Factoid Questions using Face Beauty Products Knowledge Graph," *Jurnal Linguistik Komputasional*, vol. 4, no. 2, pp. 59–63, 2021.
- [7] S. Auer *et al.*, "The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge," *Sci Rep*, vol. 13, no. 1, p. 7240, 2023.
- [8] R. Etezadi and M. Shamsfard, "The state of the art in open domain complex question answering: a survey," *Applied Intelligence*, vol. 53, no. 4, pp. 4124–4144, 2023.
- [9] R. Kusumaningrum, A. F. Hanifah, K. Khadijah, S. N. Endah, and P. S. Sasongko, "Long Short-Term Memory for Non-Factoid Answer Selection in Indonesian Question Answering System for Health Information," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023.
- [10] D. Demner-Fushman, Y. Mrabet, and A. Ben Abacha, "Consumer health information and question answering: helping consumers find answers to their health-related information needs," *Journal of the American Medical Informatics Association*, vol. 27, no. 2, pp. 194–201, 2020.
- [11] P. K. Roy, S. Saumya, J. P. Singh, S. Banerjee, and A. Gutub, "Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review," *CAAI Trans Intell Technol*, vol. 8, no. 1, pp. 95–117, 2023.
- [12] Y. Moriya and G. J. F. Jones, "Improving Noise Robustness for Spoken Content Retrieval Using Semi-Supervised ASR and N-Best Transcripts for BERT-Based Ranking Models," in *2022 IEEE Spoken Language Technology Workshop (SLT)*, 2023, pp. 398–405.
- [13] G. N. Ahmad and A. Romadhony, "End-to-End Question Answering System for Indonesian Documents Using TF-IDF and IndoBERT," in *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2023, pp. 1–6.
- [14] OpenAI, "GPT-3.5 (ChatGPT)." 2021. [Online]. Available: <https://www.openai.com/>
- [15] Z. Abbasiantaeb and S. Momtazi, "Text-based question answering from information retrieval and deep neural network perspectives: A survey," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 11, no. 6, p. e1412, 2021.

- [16] V. Karpukhin *et al.*, “Dense passage retrieval for open-domain question answering,” *arXiv preprint arXiv:2004.04906*, 2020.