

**EVALUASI PERFORMA ARSITEKTUR DENSE RETRIEVAL
DAN RERANKER DALAM PENILAIAN TATA KELOLA
PERUSAHAAN PUBLIK DI INDONESIA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

OSCAR IQBAL MUSTOFA

22.11.4559

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**EVALUASI PERFORMA ARSITEKTUR DENSE RETRIEVAL
DAN RERANKER DALAM PENILAIAN TATA KELOLA
PERUSAHAAN PUBLIK DI INDONESIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

OSCAR IQBAL MUSTOFA

22.11.4559

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**EVALUASI PERFORMA ARSITEKTUR DENSE RETRIEVAL DAN
RERANKER DALAM PENILAIAN TATA KELOLA PERUSAHAAN
PUBLIK DI INDONESIA**

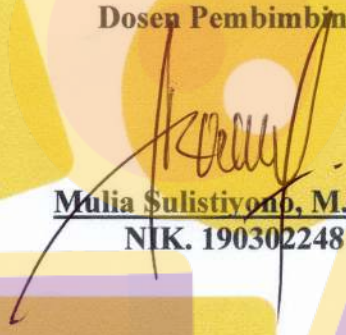
yang disusun dan diajukan oleh

Oscar Iqbal Mustofa

22.11.4559

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 07 Februari 2026

Dosen Pembimbing,


Mulia Sulistiyono, M.Kom

NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

**EVALUASI PERFORMA ARSITEKTUR DENSE RETRIEVAL DAN
RERANKER DALAM PENILAIAN TATA KELOLA PERUSAHAAN
PUBLIK DI INDONESIA**

yang disusun dan diajukan oleh

Oscar Iqbal Mustofa

22.11.4559

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

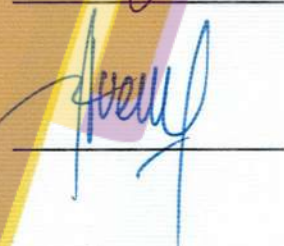
Windha Mega Pradnya Dhuhiya, S.Kom., M.Kom.
NIK. 190302185



Hendra Kurniawan, S.Kom., M.Kom.
NIK. 190302244



Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Oscar Iqbal Mustofa
NIM : 22.11.4559

Menyatakan bahwa Skripsi dengan judul berikut:

Evaluasi Performa Arsitektur Dense Retrieval dan Reranker dalam Penilaian Tata Kelola Perusahaan Publik di Indonesia

Dosen Pembimbing : Mulia Sulistyono, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Februari 2026

Yang Menyatakan,



Oscar Iqbal Mustofa

HALAMAN PERSEMBAHAN

Puji syukur ke hadirat Allah SWT atas segala nikmat, kekuatan, dan kemudahan yang telah diberikan sehingga skripsi ini terselesaikan dengan baik. Skripsi ini saya persembahkan kepada pihak-pihak istimewa dalam hidup saya:

1. Kedua orang tua dari penulis. Kepada mereka terima kasih yang tak terhingga atas kasih sayang yang tulus dan doa dalam setiap sujud yang tiada terputus. Semoga Allah SWT senantiasa melimpahkan keberkahan, kesehatan, dan usia yang penuh dengan ridho-Nya.
2. Kedua adik tercinta yang seringai serta tingkahnya telah memberi warna kepada penulis. Kehadiran kalian telah memberi dorongan untuk selalu menjadi manusia terbaik agar layak menjadi penunjuk jalan terdepan.
3. Keluarga besar, atas doa, dukungan, dan cinta yang senantiasa menguatkan dalam setiap perjalanan. Terima kasih untuk setiap kebaikan yang telah dihantarkan.
4. Seluruh pendidik sepanjang hidup penulis hingga saat ini, atas segala ajaran, arahan, dan wejangan yang telah merasuk dan membentuk diri hingga sampai detik ini. Terima kasih untuk segenap tauladan yang telah diberi.
5. Seluruh rekan sepanjang hidup hingga saat ini, atas kehadirannya di berbagai fase naik-turun dan susah senang penulis. Terima kasih atas beribu gelas kopi dan kebersamaan yang telah dilewati serta berjuta makna hidup yang telah dipelajari.
6. Partner, rekan, sahabat penulis yang telah mengambil bagian dari perjalanan hidup sejak awal menginjakkan kaki di pendidikan tinggi hingga saat ini. Terima kasih atas kesetiaan yang tak pudar dan dukungan serta bantuan yang diberikan saat hampir mencapai dasar.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Evaluasi Performa Arsitektur Dense Retrieval dan Reranker dalam Penilaian Tata Kelola Perusahaan Publik di Indonesia”** dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penulis menyadari bahwa dalam penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Mohammad Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusrini, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Eli Pujastuti, M.Kom. selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Mulia Sulistiyono, M.Kom. selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan yang sangat berharga selama proses pengerjaan skripsi ini.
5. Tim Dosen Penguji, yang telah memberikan saran, kritik, dan masukan berharga dalam penyempurnaan skripsi ini.
6. Kedua orang tua tercinta dan keluarga besar yang senantiasa memberikan doa, kasih sayang, dukungan moral, maupun materiil yang tiada henti.

Magelang, 06 Februari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xvi
INTISARI	xxvii
<i>ABSTRACT</i>	xxviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4

BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	11
2.2.1 Tata Kelola Perusahaan.....	11
2.2.2 The ASEAN Corporate Governance Scorecard.....	12
2.2.3 Laporan Tahunan	15
2.2.4 Information Retrieval	16
2.2.5 Semantic Retrieval	17
2.2.6 Dense Retrieval	17
2.2.7 Reranker	18
2.2.8 Sentence Transformers.....	20
2.2.9. Metrik Evaluasi.....	22
BAB III METODE PENELITIAN	24
3.1. Objek Penelitian.....	24
3.2. Alur Penelitian	24
3.3.1. Pengunduhan Data PDF.....	25
3.3.2. Preprocessing Data.....	26
3.3.3. Retriever Encoding	26
3.3.4. Querying	27
3.3.5. Evaluasi Retriever	27
3.3.6. Reranker Encoding.....	28
3.3.7. Evaluasi Final.....	28

3.3.	Alat dan Bahan.....	29
3.3.1.	Alat.....	29
3.3.2.	Bahan	30
BAB IV HASIL DAN PEMBAHASAN		35
4.1.	Pengunduhan Data PDF	35
4.2.	Preprocessing Data.....	38
4.2.1.	Ekstraksi Data	38
4.2.2.	Seleksi Data	39
4.2.3.	Restrukturisasi Data	39
4.3.	Retriever Encoding	41
4.4.	Querying	42
4.5.	Evaluasi Retriever.....	43
4.6.	Reranker Encoding.....	47
4.7.	Evaluasi Final.....	47
BAB V PENUTUP		54
5.1	Kesimpulan	54
5.2	Saran	54
REFERENSI		56
LAMPIRAN.....		61

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 2.2 Struktur ASEAN Corporate Governance Scorecard (ACGS)	13
Tabel 2.3 Pembobotan ASEAN Corporate Governance Scorecard (ACGS)	14
Tabel 3.1 Hardware	29
Tabel 3.2 Software	29
Tabel 3.3 Daftar Perusahaan	30
Tabel 3.4 Daftar Persentase Laporan Tahunan sebagai Rujukan pada Publikasi Hasil Penilaian ACGS oleh Perusahaan Terkait	32
Tabel 3.5 Daftar Persentase Bab pada Laporan Tahunan sebagai Rujukan pada Level 1 Aspek D Publikasi Hasil Penilaian ACGS	34
Tabel 4.1 Daftar Jumlah Baris Data Hasil Preprocessing Laporan Tahunan 10 Perusahaan pada Bab Tata Kelola Perusahaan	41
Tabel 4.2 Potongan Hasil Querying Top-5	42

DAFTAR GAMBAR

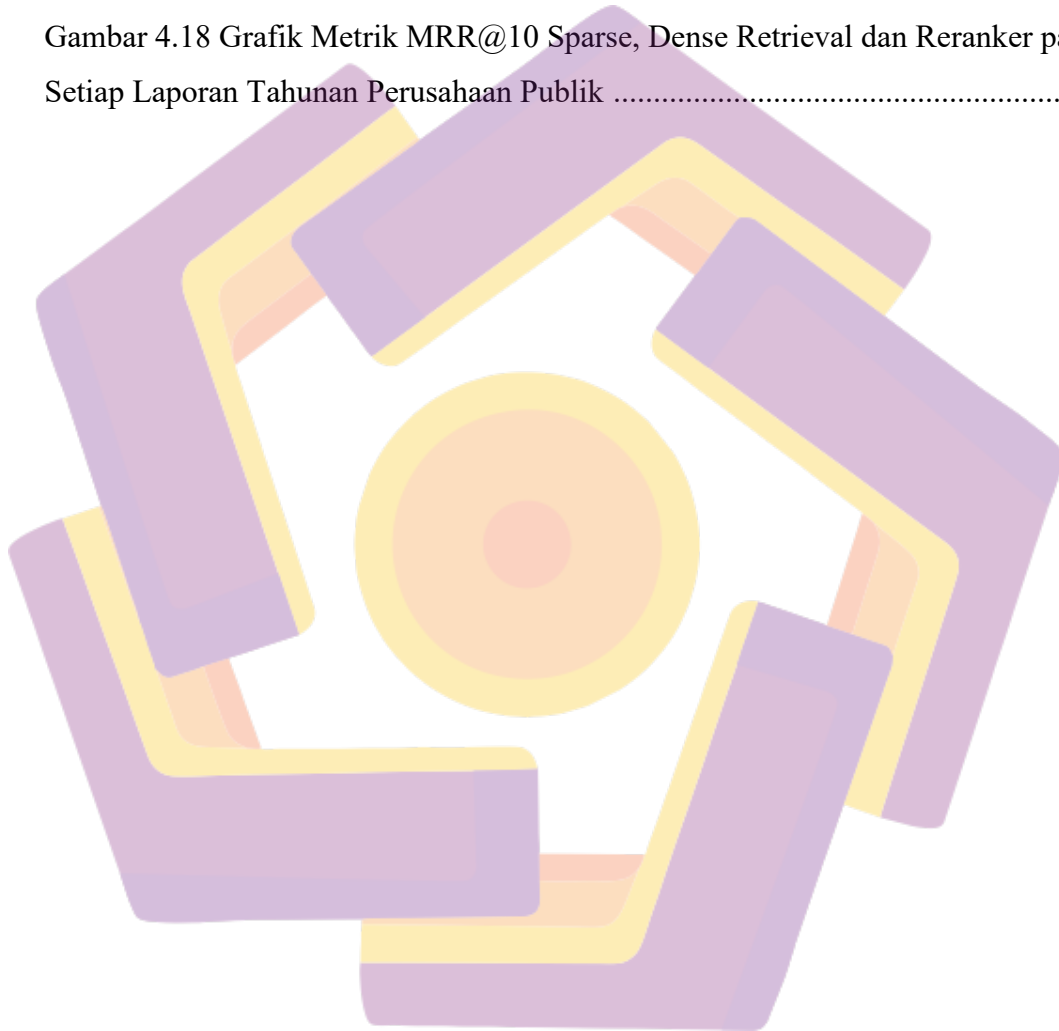
Gambar 2.1 Bentuk Arsitektur Bi-Encoder	18
Gambar 2.2 Bentuk Arsitektur Cross-Encoder	20
Gambar 2.3 Line Chart Performa Cross-Encoder dan Bi-Encoder	20
Gambar 3.1 Alur Penelitian	25
Gambar 4.1 Tampilan Website ASEAN Capital Market Forum	35
Gambar 4.2 Tampilan Page 2023 ASEAN Corporate Governance Scorecard	35
Gambar 4.3 Tampilan PDF 2023 ASEAN Corporate Governance Scorecard	36
Gambar 4.4 Tampilan Website BEI/IDX	36
Gambar 4.5 Tampilan Page Laporan Keuangan dan Tahunan	37
Gambar 4.6 Tampilan Laporan Tahunan PT Wijaya Karya Tbk	37
Gambar 4.7 Tampilan Website Resmi PT Indosat Tbk	38
Gambar 4.8 Hasil Ekstraksi Data	39
Gambar 4.9 Hasil Seleksi dan Restrukturisasi Data	40
Gambar 4.10 Tampilan Laporan Hasil Penilaian Tata Kelola dengan ACGS	43
Gambar 4.11 Tampilan Evaluator Berisi Anchor Page sebagai Label	44
Gambar 4.12 Grafik Metrik Hit@K Sparse dan Dense Retrieval	45
Gambar 4.13 Grafik Metrik MRR@K Sparse dan Dense Retrieval	46
Gambar 4.14 Grafik Metrik MRR@K Dense Retrieval dan Reranker	48

Gambar 4.15 Grafik Metrik Hit@K Sparse, Dense Retrieval dan Reranker 49

Gambar 4.16 Grafik Metrik MRR@K Sparse, Dense Retrieval dan Reranker 49

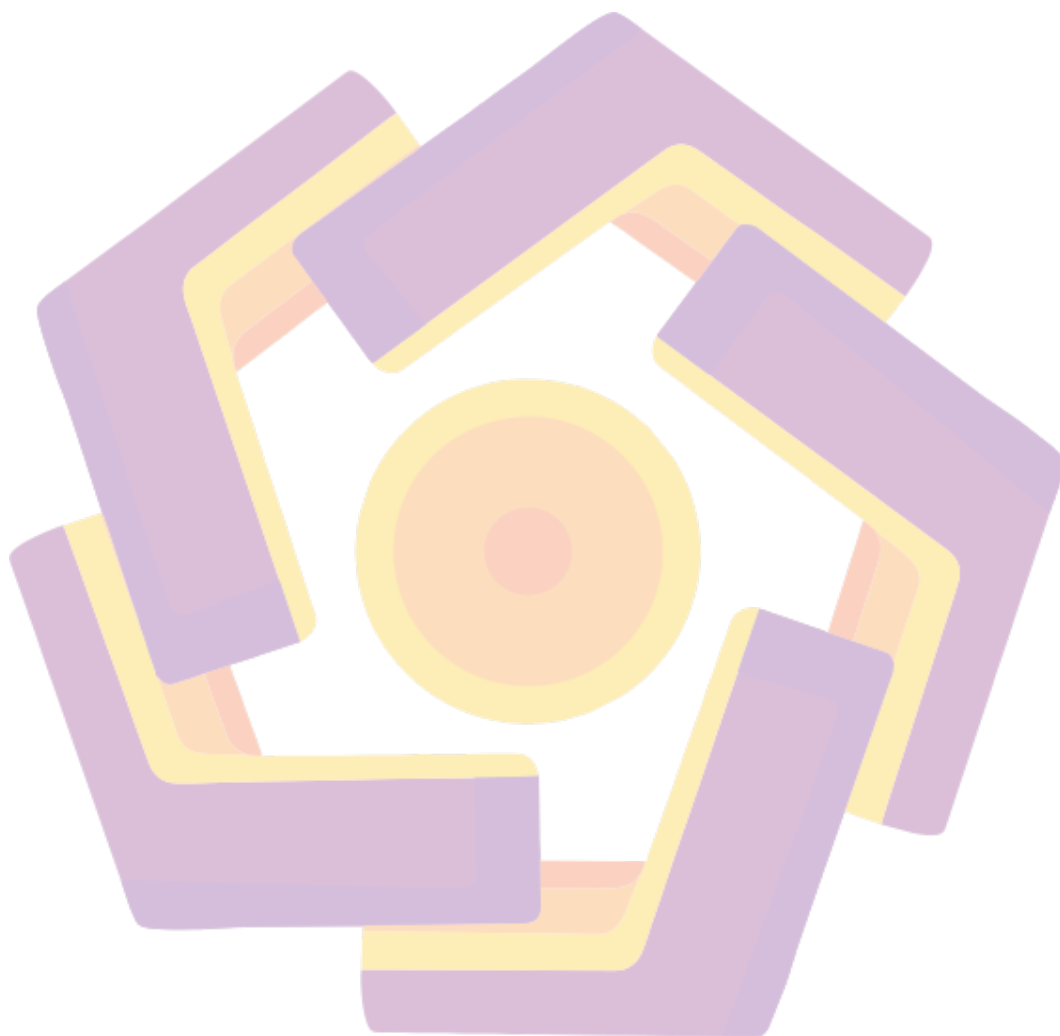
Gambar 4.17 Grafik Metrik Hit@10 Sparse, Dense Retrieval dan Reranker pada
Setiap Laporan Tahunan Perusahaan Publik 51

Gambar 4.18 Grafik Metrik MRR@10 Sparse, Dense Retrieval dan Reranker pada
Setiap Laporan Tahunan Perusahaan Publik 52

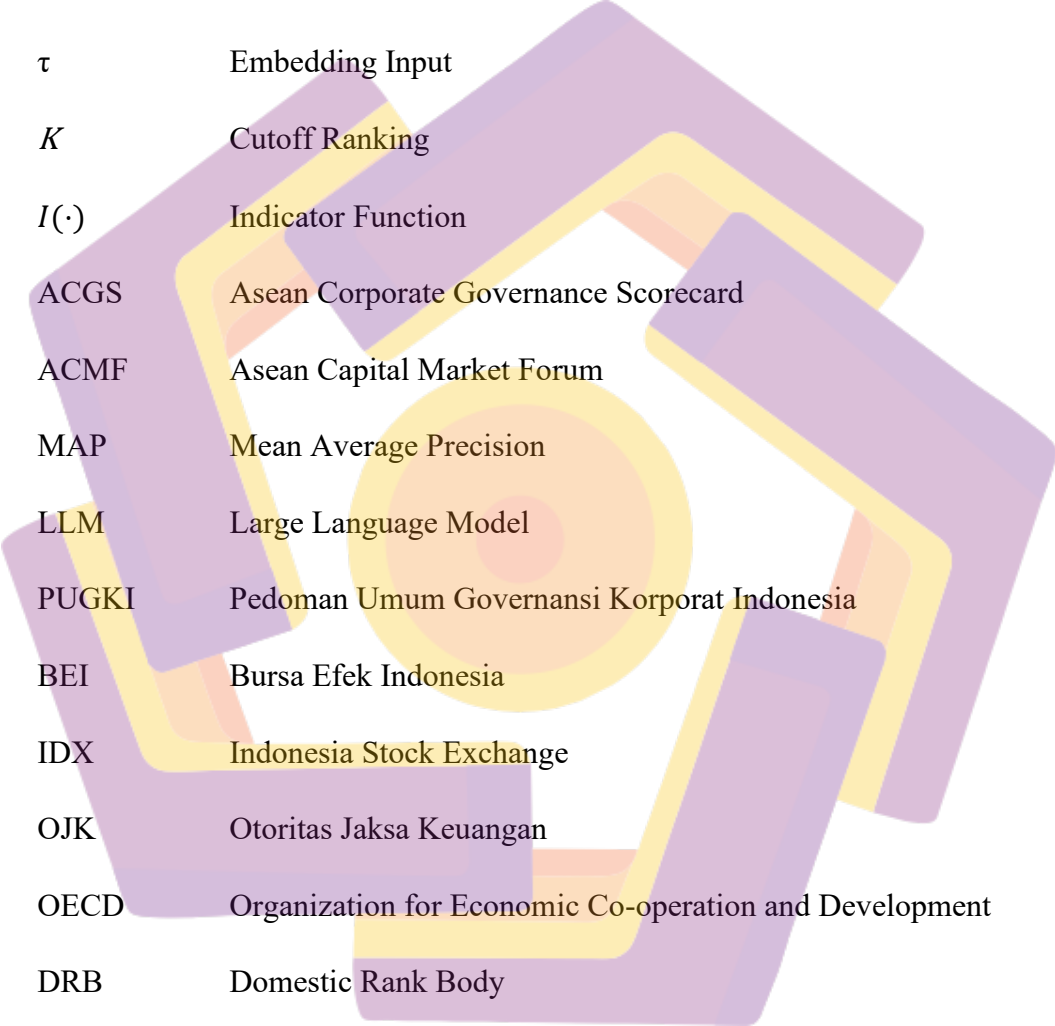


DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



DAFTAR LAMBANG DAN SINGKATAN

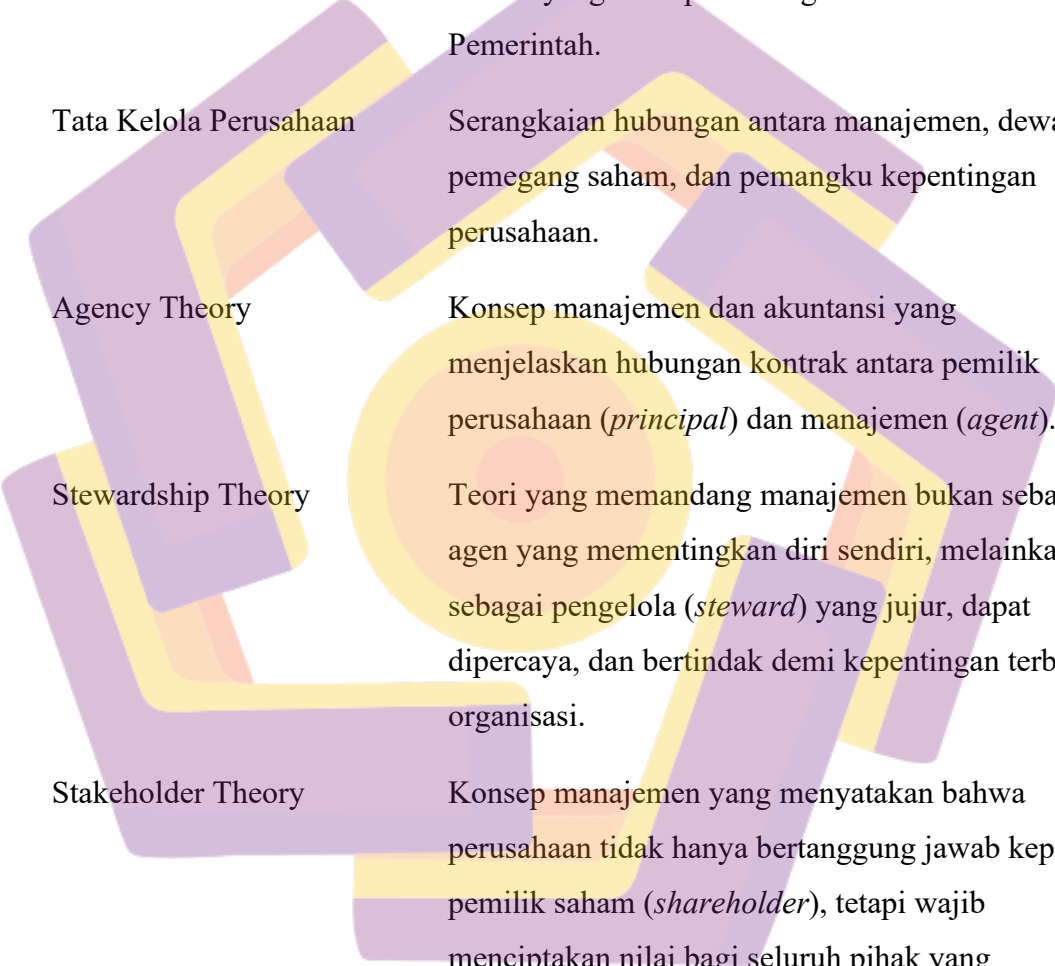


y	Output Embedding
$red(\cdot)$	Reduction Function
T	Transformers
τ	Embedding Input
K	Cutoff Ranking
$I(\cdot)$	Indicator Function
ACGS	Asean Corporate Governance Scorecard
ACMF	Asean Capital Market Forum
MAP	Mean Average Precision
LLM	Large Language Model
PUGKI	Pedoman Umum Governansi Korporat Indonesia
BEI	Bursa Efek Indonesia
IDX	Indonesia Stock Exchange
OJK	Otoritas Jasa Keuangan
OECD	Organization for Economic Co-operation and Development
DRB	Domestic Rank Body
IR	Information Retrieval
QA	Question Answering
MS MARCO	Microsoft Machine Reading Comprehension
BEIR	Benchmarking Information Retrieval
TREC	Text Retrieval Conference



SBERT	Sentence Transformer/Sentence BERT
NLI	Natural Language Inference
STS	Semantic Textual Similarity
CLS	Classification Token
SEP	Separator Token
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representation from Transformers
RoBERTa	Robust Optimized BERT Approach
DistilBERT	Distilled BERT
MPNet	Masked and Permuted Pre-training for Language Understanding
MiniLM	Miniaturized Language Model
RAG	Retrieval-Augmented Generation
MRR	Mean Reciprocal Ranking
PDF	Portable Document Format
BM25	Best Matching 25
DCG	Discounted Cumulative Gain
NDCG	Normalized Discounted Cumulative Gain

DAFTAR ISTILAH



Perusahaan Publik	Perseroan yang sahamnya telah dimiliki paling sedikit oleh 300 pemegang saham dan memiliki modal disetor paling sedikit Rp 3.000.000.000 atau suatu jumlah pemegang saham dan modal disetor yang ditetapkan dengan Peraturan Pemerintah.
Tata Kelola Perusahaan	Serangkaian hubungan antara manajemen, dewan, pemegang saham, dan pemangku kepentingan perusahaan.
Agency Theory	Konsep manajemen dan akuntansi yang menjelaskan hubungan kontrak antara pemilik perusahaan (<i>principal</i>) dan manajemen (<i>agent</i>).
Stewardship Theory	Teori yang memandang manajemen bukan sebagai agen yang mementingkan diri sendiri, melainkan sebagai pengelola (<i>steward</i>) yang jujur, dapat dipercaya, dan bertindak demi kepentingan terbaik organisasi.
Stakeholder Theory	Konsep manajemen yang menyatakan bahwa perusahaan tidak hanya bertanggung jawab kepada pemilik saham (<i>shareholder</i>), tetapi wajib menciptakan nilai bagi seluruh pihak yang terpengaruh atau mempengaruhi aktivitas bisnis.
Political Theory	Pendekatan yang melihat struktur perusahaan sebagai entitas politik, di mana keputusan dipengaruhi oleh kekuasaan, hubungan dengan



	pemerintah, dan tekanan pemangku kepentingan (publik).
Asean Corporate Governance Scorecard	Instrumen asesmen atau alat penilaian yang disusun oleh ASEAN Capital Market Forum (ACMF) demi meningkatkan standar tata kelola perusahaan negara-negara ASEAN serta meningkatkan visibilitas perusahaan well-governed ASEAN pada skala internasional.
Asean Capital Market Forum	Forum tingkat tinggi yang terdiri dari regulator pasar modal dari 10 yurisdiksi ASEAN (termasuk OJK Indonesia) yang didirikan pada 2004 untuk membangun pasar modal regional yang dalam, likuid, dan terintegrasi.
Neural Network	Jenis algoritma <i>machine learning</i> dan fondasi AI modern yang cara kerjanya meniru otak manusia untuk memproses informasi.
Information Retrieval	Pencarian informasi dalam sekumpulan besar informasi yang tidak terstruktur untuk memenuhi kebutuhan pengguna.
Semantic Retrieval	Pendekatan untuk melakukan <i>Information Retrieval</i> (IR) dengan memahami kesamaan kata makna.
Dense Retrieval	Teknik <i>semantic retrieval</i> yang menilai relevansi makna berdasarkan jarak representasi <i>query</i> dan <i>passage</i> dalam <i>vector embedding</i> .

Sparse Retrieval

Pendekatan untuk melakukan IR dengan melakukan pencocokan kata antara *query* dengan unit teks (*passage*) secara literal.

Cross-Modal Retrieval

Teknik dalam *retrieval* yang memungkinkan pengguna mencari data dalam satu *modal* menggunakan *query* dari *modal* lain yang berbeda.

Reranker

Tahapan dalam *information retrieval* yang bertujuan untuk mengurutkan *passage* yang dianggap relevan dengan *query*, yang sebelumnya telah diproses oleh *retriever*.

Query

Serangkaian kata, frasa, atau perintah spesifik yang dimasukkan pengguna untuk meminta informasi dari sistem *retrieval*.

Passages

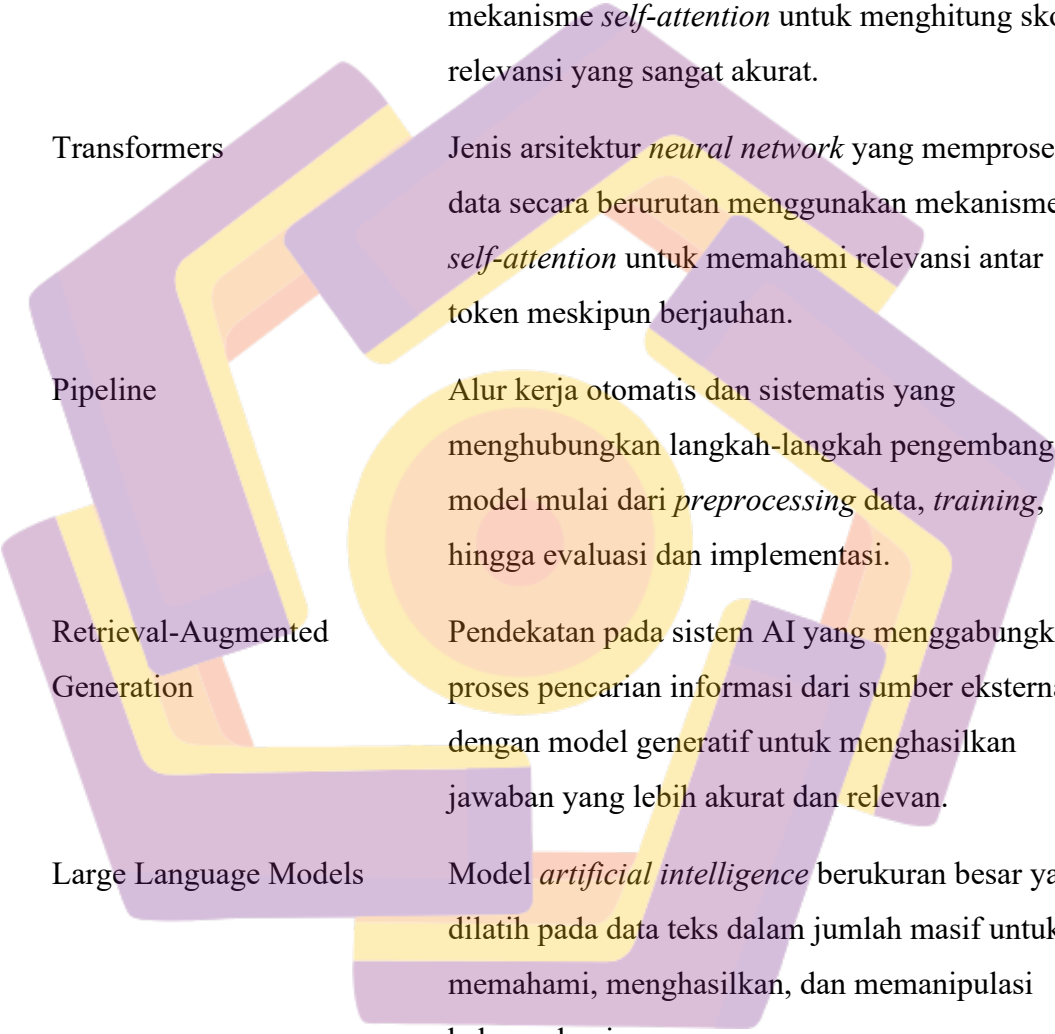
Unit informasi dasar berupa data, teks, atau rekaman terstruktur maupun tidak terstruktur yang disimpan dalam *corpus* dan dicari relevansinya terhadap *query* pengguna.

Corpus

Sekumpulan *passages* berupa teks, artikel, atau data dalam jumlah besar yang disimpan secara elektronik dan terstruktur yang dijadikan sumber utama pencarian informasi.

Pretrained Transformer

Model *artificial intelligence* berbasis arsitektur *transformer* yang dilatih pada sekumpulan data teks berukuran raksasa sebelum digunakan untuk tugas spesifik.



Bi-Encoder	Arsitektur model <i>neural network</i> yang memproses dua input teks secara independen menjadi <i>vector embedding</i> .
Cross-Encoder	Arsitektur model <i>neural network</i> yang memproses dua input teks secara bersamaan melalui mekanisme <i>self-attention</i> untuk menghitung skor relevansi yang sangat akurat.
Transformers	Jenis arsitektur <i>neural network</i> yang memproses data secara berurutan menggunakan mekanisme <i>self-attention</i> untuk memahami relevansi antar token meskipun berjauhan.
Pipeline	Alur kerja otomatis dan sistematis yang menghubungkan langkah-langkah pengembangan model mulai dari <i>preprocessing</i> data, <i>training</i> , hingga evaluasi dan implementasi.
Retrieval-Augmented Generation	Pendekatan pada sistem AI yang menggabungkan proses pencarian informasi dari sumber eksternal dengan model generatif untuk menghasilkan jawaban yang lebih akurat dan relevan.
Large Language Models	Model <i>artificial intelligence</i> berukuran besar yang dilatih pada data teks dalam jumlah masif untuk memahami, menghasilkan, dan memanipulasi bahasa alami.
Generative Models	Model <i>artificial intelligence</i> yang dirancang untuk mempelajari pola data dan menghasilkan data baru yang menyerupai data asli.

Preprocessing Data

Proses menyiapkan dan membersihkan data mentah agar siap digunakan untuk langkah selanjutnya seperti analisis atau *modelling*.

Parsing Text

Proses menganalisis dan memecah teks menjadi struktur atau elemen tertentu agar dapat diproses oleh sistem komputer.

Case Folding

Proses mengubah seluruh huruf dalam teks menjadi satu **bentuk** konsisten untuk menghindari perbedaan makna akibat variasi kapitalisasi.

Tokenization

Proses memecah teks menjadi **unit-unit** kecil yang disebut token untuk memudahkan pemrosesan bahasa alami.

Stopword Removal

Proses menghapus kata-kata **umum** yang sering muncul namun minim informasi agar analisis teks lebih fokus pada kata yang memiliki makna.

Stemming

Proses mengubah kata ke bentuk dasarnya atau akarnya untuk menyatukan variasi kata dengan makna serupa dalam analisis teks.

Lemmatization

Proses mengubah kata ke bentuk dasarnya berdasarkan konteks dan aturan tata bahasa sehingga menghasilkan bentuk kata yang benar secara linguistik.

Querying

Proses menggunakan representasi numerik atau vektor yang dihasilkan oleh encoder untuk mencari informasi relevan dari data.

Similarity Function

Fungsi yang menghitung seberapa mirip atau dekat dua objek berdasarkan kriteria tertentu.

Semantic Similarity Ukuran seberapa mirip makna atau konsep antara dua teks, kata, atau kalimat.

Vector Embedding Representasi objek dalam bentuk vektor di ruang berdimensi tinggi agar hubungan semantik dapat diukur secara matematis.

Dot Product Operasi matematika yang mengalikan elemen-elemen yang bersesuaian dari dua vektor dan menjumlahkan hasilnya untuk mengukur kesamaan arah.

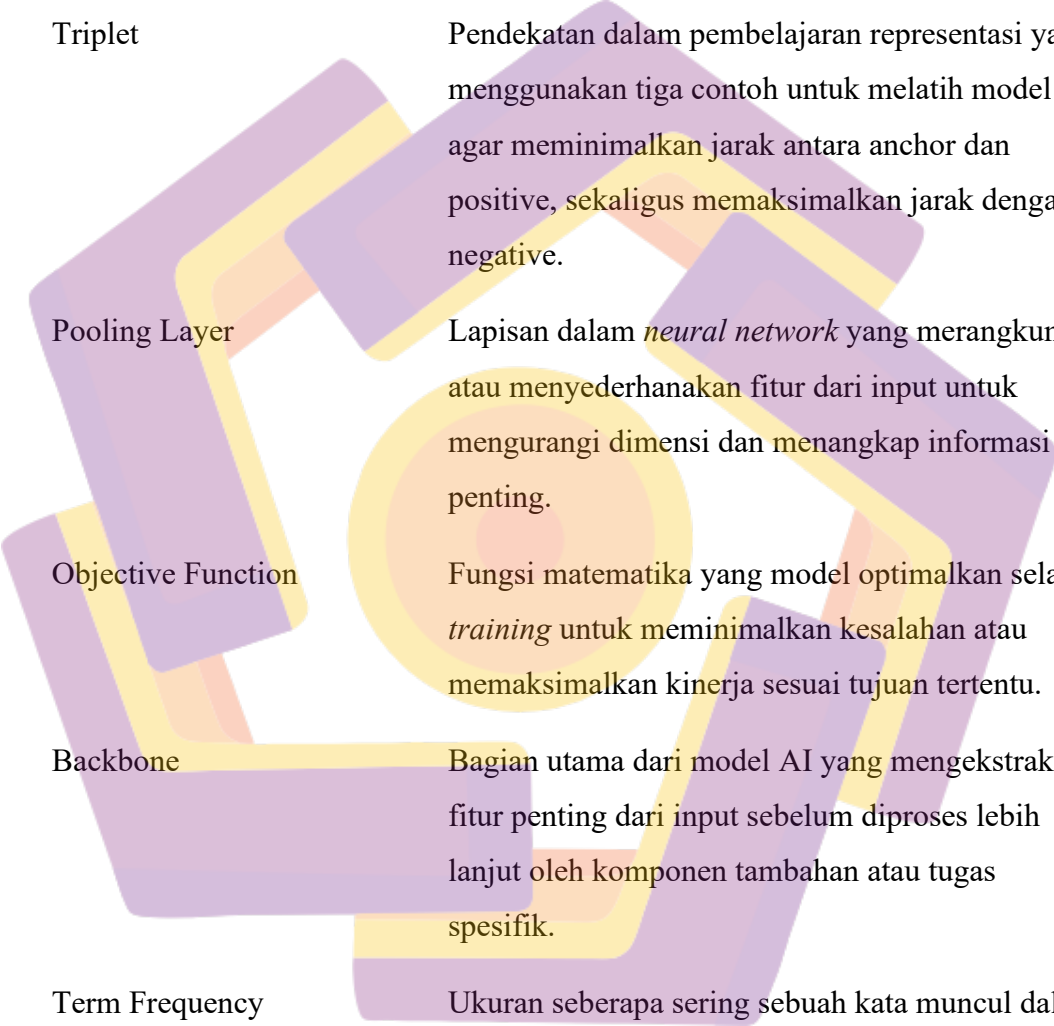
Cosine Similarity Ukuran kesamaan antara dua vektor dengan menghitung kosinus sudut di antara keduanya untuk menunjukkan seberapa mirip arah atau pola mereka terlepas dari magnitudonya.

Encode Proses mengubah data menjadi representasi numerik atau format lain yang dapat diproses oleh sistem komputer atau model AI.

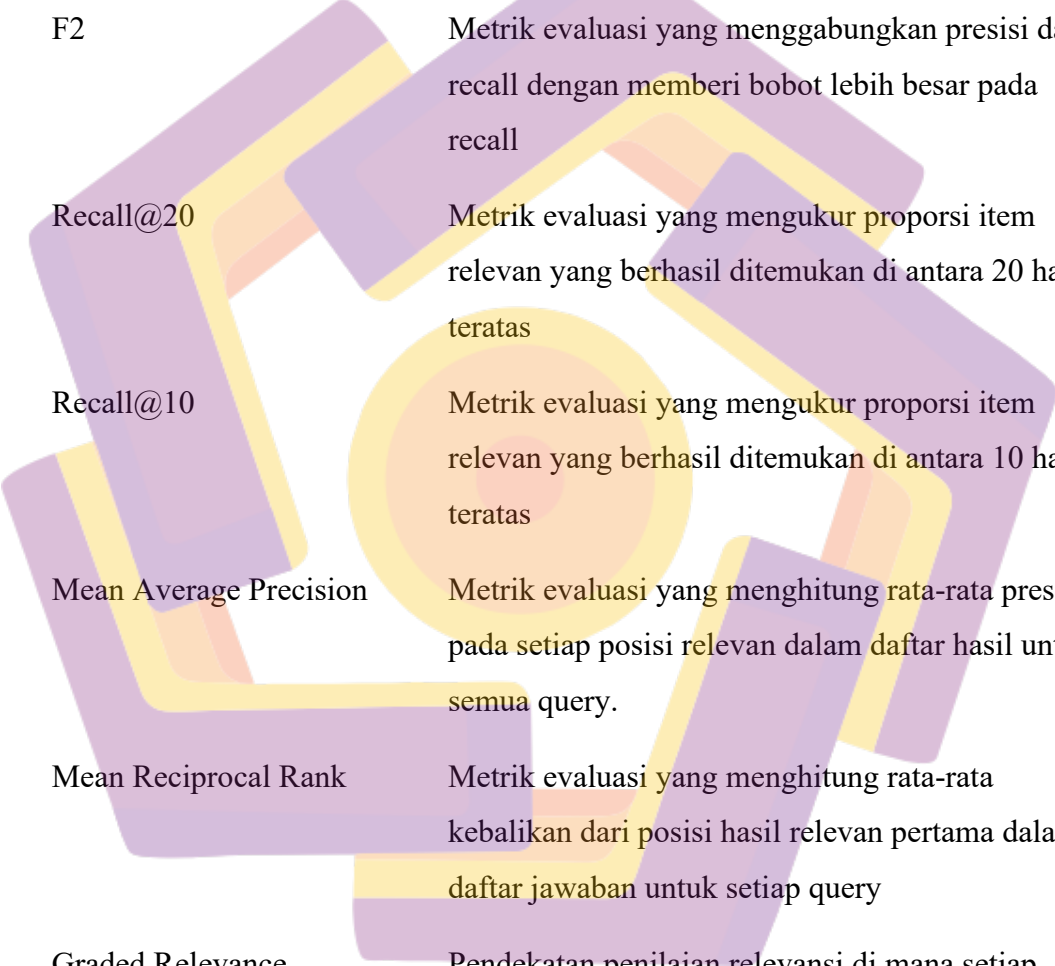
Fine-Tuning Proses melatih ulang model AI yang sudah ada dengan data khusus agar performanya lebih optimal pada tugas atau domain tertentu.

Self Attention Mechanism Teknik dalam model AI yang memungkinkan setiap elemen input memperhatikan dan menimbang hubungan dengan elemen lain untuk menangkap konteks secara efektif.

Classification Token Token khusus yang ditambahkan ke input model untuk mewakili keseluruhan input dan digunakan sebagai representasi tunggal saat melakukan tugas klasifikasi.



Separator Token	Token khusus yang digunakan untuk memisahkan bagian-bagian berbeda dari input agar model dapat membedakan konteks.
Siamese	Arsitektur model AI yang menggunakan dua atau lebih jaringan identik dengan bobot yang sama
Triplet	Pendekatan dalam pembelajaran representasi yang menggunakan tiga contoh untuk melatih model agar meminimalkan jarak antara anchor dan positive, sekaligus memaksimalkan jarak dengan negative.
Pooling Layer	Lapisan dalam <i>neural network</i> yang merangkum atau menyederhanakan fitur dari input untuk mengurangi dimensi dan menangkap informasi penting.
Objective Function	Fungsi matematika yang model optimalkan selama <i>training</i> untuk meminimalkan kesalahan atau memaksimalkan kinerja sesuai tujuan tertentu.
Backbone	Bagian utama dari model AI yang mengekstraksi fitur penting dari input sebelum diproses lebih lanjut oleh komponen tambahan atau tugas spesifik.
Term Frequency	Ukuran seberapa sering sebuah kata muncul dalam sebuah dokumen.
Inverse Document Frequency	Ukuran seberapa unik atau jarang sebuah kata muncul di seluruh kumpulan dokumen



Metrik Evaluasi	Ukuran atau indikator yang digunakan untuk menilai kinerja model atau sistem dalam menyelesaikan tugas tertentu.
Relevance Scoring	Proses memberikan nilai atau skor pada item untuk menunjukkan seberapa sesuai atau relevan item tersebut dengan query atau konteks tertentu.
F2	Metrik evaluasi yang menggabungkan presisi dan recall dengan memberi bobot lebih besar pada recall
Recall@20	Metrik evaluasi yang mengukur proporsi item relevan yang berhasil ditemukan di antara 20 hasil teratas
Recall@10	Metrik evaluasi yang mengukur proporsi item relevan yang berhasil ditemukan di antara 10 hasil teratas
Mean Average Precision	Metrik evaluasi yang menghitung rata-rata presisi pada setiap posisi relevan dalam daftar hasil untuk semua query.
Mean Reciprocal Rank	Metrik evaluasi yang menghitung rata-rata kebalikan dari posisi hasil relevan pertama dalam daftar jawaban untuk setiap query
Graded Relevance	Pendekatan penilaian relevansi di mana setiap item diberi skor berdasarkan tingkat kesesuaiannya dengan query
Discounted Cumulative Gain	Metrik evaluasi peringkat yang menghitung total relevansi item dalam daftar hasil dengan memberi bobot lebih rendah pada item yang muncul di

posisi lebih rendah menggunakan *discount function*.

Normalized Discounted
Cumulative Gain

Versi terstandarisasi dari DCG yang membandingkan skor DCG aktual dengan skor DCG ideal sehingga menghasilkan nilai antara 0 dan 1 untuk menilai kualitas ranking.

False Positive

Kasus di mana sistem atau model salah mengidentifikasi item sebagai relevan atau positif padahal seharusnya tidak.

Threshold

Nilai batas yang digunakan untuk memutuskan apakah suatu prediksi atau skor dianggap positif atau negatif

Top-K

Sejumlah K item teratas yang dipilih atau direkomendasikan oleh sistem berdasarkan skor atau peringkatnya.

Baseline

Hasil referensi awal yang digunakan sebagai patokan untuk membandingkan dan menilai pendekatan baru.

Feature Engineering

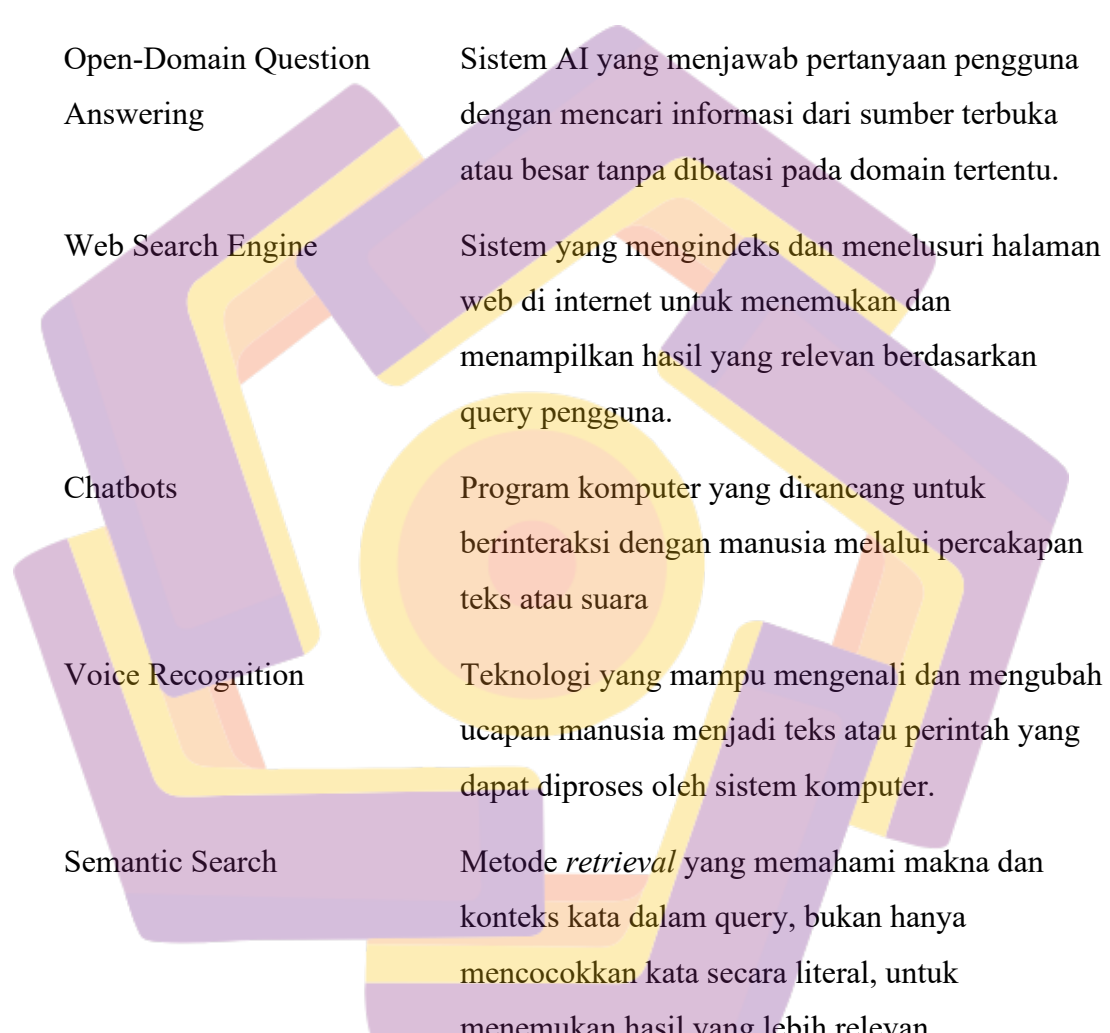
Proses membuat, memilih, atau mengubah fitur dari data mentah agar lebih informatif dan meningkatkan performa model AI atau analitik.

Knowledge Graph

Representasi data dalam bentuk graf yang menyimpan entitas dan hubungan antar-entitas untuk menangkap pengetahuan secara terstruktur dan semantik.

Learning to Rank

Pendekatan pembelajaran mesin yang melatih model untuk mengurutkan item atau dokumen



	berdasarkan relevansi terhadap query atau tujuan tertentu.
Entity Linking	Proses mengenali entitas dalam teks dan menghubungkannya ke identitas unik yang sesuai dalam basis pengetahuan atau database eksternal.
Open-Domain Question Answering	Sistem AI yang menjawab pertanyaan pengguna dengan mencari informasi dari sumber terbuka atau besar tanpa dibatasi pada domain tertentu.
Web Search Engine	Sistem yang mengindeks dan menelusuri halaman web di internet untuk menemukan dan menampilkan hasil yang relevan berdasarkan query pengguna.
Chatbots	Program komputer yang dirancang untuk berinteraksi dengan manusia melalui percakapan teks atau suara
Voice Recognition	Teknologi yang mampu mengenali dan mengubah ucapan manusia menjadi teks atau perintah yang dapat diproses oleh sistem komputer.
Semantic Search	Metode <i>retrieval</i> yang memahami makna dan konteks kata dalam query, bukan hanya mencocokkan kata secara literal, untuk menemukan hasil yang lebih relevan.
Clustering	Teknik analisis data yang mengelompokkan objek atau data berdasarkan kemiripan mereka
Natural Language Inference	Tugas dalam pemrosesan bahasa untuk menentukan hubungan logis antara sepasang teks terhadap teks lainnya.

Semantic Textual Similarity Tugas mengukur seberapa mirip makna dua teks atau kalimat.

State of the Art Metode, model, atau teknologi terbaru dan terbaik yang mencapai performa tertinggi dalam suatu bidang tertentu.

MS MARCO Dataset besar untuk tugas *information retrieval* dan *question answering* yang dibuat oleh Microsoft

BEIR Kumpulan dataset *cross-domain* dengan fokus tugas *dense retrieval*

TREC Konferensi evaluasi *information retrieval* yang menyediakan dataset standar untuk evaluasi sistem *retrieval* yang dimulai oleh National Institute of Standards and Technology

INTISARI

Tata kelola perusahaan merupakan mekanisme yang mengatur hubungan antara manajemen, dewan, pemegang saham, dan pemangku kepentingan untuk mencapai tujuan perusahaan secara efektif. Penerapan tata kelola yang baik diyakini dapat meningkatkan kinerja keuangan, efisiensi operasional, akses pembiayaan, kepercayaan investor, serta nilai bagi pemegang saham dan pemangku kepentingan, dengan berlandaskan prinsip transparansi, akuntabilitas, responsibilitas, independensi, serta kewajaran. Untuk menilai kualitas tata kelola, berbagai kerangka penilaian dikembangkan, salah satunya The ASEAN Corporate Governance Scorecard (ACGS), yang menggunakan 146 indikator berbasis informasi publik, termasuk laporan tahunan. Namun, proses penilaian ACGS menghadapi tantangan seperti volume dokumen yang besar, struktur dokumen yang beragam, penggunaan bahasa tidak seragam, dan waktu penilaian yang panjang. Pendekatan semantic retrieval berbasis dense retrieval dan reranker menawarkan solusi untuk tantangan ini, karena mampu memahami kesamaan makna antar kalimat dan meningkatkan relevansi hasil pencarian. Dalam arsitektur ini, bi-encoder berperan sebagai retriever untuk menemukan kandidat relevan, sedangkan cross-encoder melakukan reranking. Penelitian ini bertujuan menerapkan dan mengevaluasi arsitektur dense retrieval dan reranker dalam penilaian tata kelola perusahaan pada laporan tahunan emiten publik di Indonesia. Dengan pendekatan ini, proses penelusuran informasi diharapkan dapat dilakukan lebih cepat, konsisten, dan relevan secara semantik, sehingga mendukung evaluasi tata kelola yang lebih efisien dan akurat.

Kata kunci: Tata Kelola Perusahaan, ASEAN Corporate Governance Scorecard, Laporan Tahunan, Semantic Retrieval, Dense Retrieval dan Reranker.

ABSTRACT

Corporate governance is a mechanism that regulates the relationships among management, the board of directors, shareholders, and stakeholders to achieve the company's objectives effectively. Proper corporate governance is believed to enhance financial performance, operational efficiency, access to financing, investor confidence, and value for shareholders and stakeholders, based on the principles of transparency, accountability, responsibility, independence, and fairness. Various assessment frameworks have been developed to evaluate governance quality, including the ASEAN Corporate Governance Scorecard (ACGS), which consists of 146 indicators derived from public information, such as annual reports. However, the ACGS evaluation process faces challenges, including the large volume of documents, heterogeneous document structures, inconsistent language usage, and time-consuming assessments. Semantic retrieval using dense retrieval and reranker architectures offers a solution by capturing semantic similarity between sentences and improving retrieval relevance. In this architecture, bi-encoders function as retrievers to identify relevant candidates, while cross-encoders perform reranking. This study aims to apply and evaluate dense retrieval and reranker architectures for assessing corporate governance in the annual reports of publicly listed companies in Indonesia. By adopting this approach, information retrieval and corporate governance assessment are expected to be faster, more consistent, and semantically relevant, thereby supporting more efficient and accurate evaluations.

Keyword: Corporate Governance, ASEAN Corporate Governance Scorecard, Annual Reports, Semantic Retrieval, Dense Retrieval and Reranker.