



Large Language Model untuk Ekstraksi Grafik Bisnis Proses

KUALIFIKASI

ELYNA FAZRIYATI

99221909

PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA
2024

Daftar Isi

1	Pendahuluan	1
1.1	Latar Belakang	1
1.2	Rumusan Masalah	3
1.3	Tujuan dan Batasan Penelitian	3
1.3.1	Tujuan Penelitian	3
1.3.2	Batasan Penelitian	3
1.4	Kontribusi dan Manfaat Penelitian	4
2	Telaah Pustaka	5
2.1	Struktur Organisasi dan Tata Kelola (SOTK)	5
2.2	Bussiness Model Notaion (BPMN)	6
2.3	<i>Large Language Model</i> (LLM)	9
2.3.1	<i>Generative Pre-trained Transformer</i> (GPT)	10
2.3.2	<i>Bidirectional Encoder Representations from Transformers</i> (BERT)	11
2.4	Grounded Theory	11
2.4.1	Proses Grounded Theory	12
2.4.2	Karateristik Grounded Theory	12
2.5	Penelitian Terkait	13
3	Metode Penelitian	20
3.1	Metode yang Dikembangkan	20
3.2	Tahapan Penelitian	22

Daftar Gambar

3.1	Metode yang dikembangkan	21
3.2	tahapan penelitian	23

Daftar Tabel

2.1	Elemen Dasar BPMN	7
2.2	Penelitian Terkait	13

Bab 1

Pendahuluan

1.1 Latar Belakang

Struktur organisasi dan tata kelola (SOTK) atau susunan organisasi dan tata kerja menurut Peraturan Menteri dalam Negeri Nomor 84 Tahun 2015 merupakan sistem yang digunakan untuk menetapkan tugas dan tanggung jawab, serta hubungan kerja dalam suatu organisasi atau lembaga. Organisasi atau lembaga pemerintah atau non-pemerintah sudah seharusnya memiliki SOTK. SOTK dibutuhkan untuk mempermudah koordinasi tugas dan tanggung jawab antar anggota lembaga. Hal ini mempermudah pembagian dan penjelasan tentang cara tiap anggota menyelesaikan tugas, sehingga alur penyelesaian tugas di suatu lembaga lebih terarah dan terkendali. Namun, mengacu pada Peraturan Menpan RB atau Permenpan RB Nomor 7 Tahun 2022 Tentang Sistem Kerja Pada Instansi Pemerintah untuk Penyederhanaan Birokrasi, diperlukan adanya perbaikan atau penyesuaian mengenai pengembangan mekanisme kerja serta proses bisnis dengan menggunakan Sistem Pemerintahan Berbasis Elektronik (SPBE) untuk instansi pusat, instansi daerah, dan instansi yang dibiayai oleh Anggaran Pendapatan dan Belanja Negara (APBN) berdasarkan Undang-Undang Dasar (UUD) Negara Republik Indonesia Tahun 1945. Salah satu bagian dari reformasi birokrasi adalah penyederhanaan birokrasi. Ini dilakukan untuk mewujudkan tata kelola pemerintahan yang efisien dan efektif dengan mengoptimalkan penggunaan SPBE.

Lembaga yang baik seharusnya memiliki SOTK dengan bisnis proses yang jelas, namun pada kenyataannya banyak lembaga yang hanya memiliki SOTK namun tidak memiliki dokumen proses bisnis. Proses bisnis tersimpan pada benak masing-masing individu yang melakukan operasional proses bisnis tersebut. Dari proses bisnis yang ada harus dilakukan analisis sehingga dapat

dilakukan bussiness process *re-engineering* untuk mendapatkan sebuah proses bisnis baru yang dapat digunakan secara optimal baik dengan cara manual atau bantuan sistem. Terlebih dengan adanya peraturan mengenai penyederhanaan birokrasi, menyebabkan adanya perubahan pada SOTK dari setiap lembaga yang menyebabkan proses bisnis juga harus ikut saling menyesuaikan dengan keterbaruan yang ada.

Salah satu bagan yang dapat digunakan dalam merepresentasikan proses bisnis adalah BPMN (*Business Processes Modeling Notation*). Diagram BPMN dapat menggambarkan secara jelas dan sesuai dengan kehidupan nyata bagaimana proses bisnis berjalan. Selama ini proses bisnis suatu lembaga digambarkan dengan diagram alur, di mana dalam diagram alur semua berjalan dengan serial, ini tidak menggambarkan kehidupan nyata dimana suatu proses bisa saja berjalan berdampingan dengan proses lain. Kelemahan ini karena diagram alur adalah sebuah grafik yang digunakan dan diciptakan untuk program bukan untuk proses bisnis. Hal yang sama begitu juga dengan UML yang tercipta untuk pemrograman yang berorientasi objek. Dalam BPMN proses dapat digambarkan dengan jelas seperti pemrosesan paralel, keputusan, dokumen yang dihasilkan, dll.

Untuk dapat melakukan analisis diperlukan waktu untuk menggambar sebuah BPMN dari pemilik proses bisnis. Proses yang dilakukan adalah dengan melakukan wawancara kepada pemilik proses bisnis lalu dari hasil wawancara tersebut digambarkan sebuah grafik BPMN. Selama ini proses dilakukan secara manual sehingga meningkatkan tingkat kesalahan yang terjadi dan menambahkan waktu dalam melakukan validasi diagram BPMN. Sehingga untuk menangani masalah tersebut diperlukan adanya bantuan komputer dalam melakukan representasi dari hasil wawancara pemilik proses bisnis ke dalam sebuah grafik BPMN yang nantinya juga berhubungan dengan SOTK dari lembaga tersebut..

Beberapa metode dapat digunakan seperti: *Natural Language Processing(NLP)*, *rule base*, *Rule set*, *Mapping Rule*[8]. Salah satu yang dapat digunakan adalah menggunakan *Natural Language Processing* (NLP). NLP adalah suatu metode yang memungkinkan komputer dapat mengerti suatu bahasa. Terdapat banyak jenis NLP, dan metode terbaru saat ini adalah *Large Language Model* (LLM). Dalam LLM suatu model dilatih menggunakan data yang sangat banyak, sehingga memiliki akurasi yang sangat tinggi dan mampu menebak suatu kesimpulan atau kata berikutnya. Salah satu penggunaan LLM adalah Chat GPT yang dikembangkan oleh OpenAI. Penggunaan selain chat GPT

juga banyak seperti pada BERT, Roberta, dan IndoBert. LLM menyediakan banyak *pre-train* model yang dikembangkan dan dapat digunakan para peneliti sehingga tidak perlu mulai dari nol dalam membangun sebuah model. Penelitian ini berfokus pada pengembangan sebuah metode baru dalam melakukan interpretasi hasil wawancara pemilik bisnis proses kedalam sebuah diagram BPMN. Metode yang digunakan adalah melakukan wawancara pemilik proses bisnis dengan metode grounded theory yang disesuaikan dengan kebutuhan wawancara dan dilakukan analisis dan interpretasi perputaran menggunakan LLM.

1.2 Rumusan Masalah

Pada penelitian ini dikembangkan sebuah metode interpretasi tekstual menjadi BPMN. Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana melakukan interpretasi sebuah text menjadi notasi bisnis proses menggunakan BPMN
2. Bagaimana melakukan akuisisi data proses bisnis pada pemilik proses bisnis

1.3 Tujuan dan Batasan Penelitian

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengembangkan metode baru dalam melakukan ekstrasi grafik SOTK berbasiskan BPMN secara otomatis berdesarkan hasil wawancara pemilik proses bisnis dan data SOTK yang ada di Badan Kepegawaian Negara (BKN)
2. Mengembangkan perangkat lunak untuk membantu proses ekstrasi BPMN

1.3.2 Batasan Penelitian

Batasan masalah dalam penelitian yang dilakukan adalah:

1. Domain penelitian adalah ekstrasi Struktur Organisasi dan Tata Kelola (SOTK) dan Sistem Pemerintahan Berbasis Elektronik (SPBE) menggunakan metode otomatis berbasis Large Language Model.

2. Data yang digunakan didapatkan dengan metode kualitatif melalui wawancara dengan narasumber dari pihak Badan Kepegawaian Negara (BKN) Republik Indonesia.

1.4 Kontribusi dan Manfaat Penelitian

Terdapat dua kontribusi dalam penelitian ini, yaitu kontribusi secara ilmiah dan kontribusi secara teknis. Kontribusi Ilmiah dari penelitian ini adalah:

1. Adanya metode baru dan perangkat bantu dalam melakukan interpretasi tekstual menjadi sebuah grafik BPMN.

Sedangkan kontribusi teknis dari penelitian ini adalah:

1. Bisnis proses untuk Susunan Organisasi dan Tata Kerja (SOTK) dalam suatu organisasi menjadi lebih jelas
2. Pelayanan masyarakat menjadi lebih baik karena bisnis proses yang jelas

Bab 2

Telaah Pustaka

2.1 Struktur Organisasi dan Tata Kelola (SOTK)

Struktur Organisasi dan Tata Kelola (SOTK) atau dapat disebut juga Susunan Organisasi Tata Kerja merupakan satu sistem dalam kelembagaan dalam pengaturan tugas dan fungsi serta hubungan kerja menurut Peraturan Menteri dalam Negeri nomor 84 tahun 2015 tentang susunan organisasi dan tata kerja pemerintah desa. Tujuan utama dari SOTK itu sendiri menurut dinas pertanian kabupaten aceh jaya melalui laman websitenya mengatakan agar pembagian dan penjelasan dalam menjalankan tugas lebih mudah dan terkontrol/terarah yang menempatkan individu-individu yang sesuai dengan potensi dan kompeten yang terlibat pada perusahaan. SOTK tidak hanya digunakan untuk struktur organisasi dan tata kelola pemerintah desa, namun digunakan juga pada instansi pemerintah baik untuk instansi pusat, instansi daerah, dan instansi yang dibiayai oleh Anggaran Pendapatan dan Belanja Negara (APBN).

Fungsi dari SOTK adalah sebagai berikut:

1. Mendefinisikan tanggung jawab anggota dalam suatu organisasi atau lembaga berdasarkan posisi dan tanggung jawabnya serta cara pelaksanaannya.
2. Memperjelas cara koordinasi alur disposisi surat dalam organisasi atau perusahaan.
3. Berfungsi sebagai pedoman kerja bagi seluruh anggota.
4. Mempermudah pembagian dan penjelasan tentang cara menyelesaikan tugas tiap anggota sehingga aliran penyelesaian tugas dalam organisasi

atau perusahaan lebih terarah.

2.2 Bussiness Model Notaiion (BPMN)

Proses Bisnis adalah kumpulan aktivitas terstruktur yang menggambarkan hubungan kerja yang efektif dan efisien antarunit organisasi untuk menghasilkan kinerja dan keluaran yang bernilai tambah sesuai dengan tujuan pendirian organisasi, menurut ketentuan umum pasal 1 pada Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia Nomor 7 Tahun 2022 tentang sistem kerja pada instansi pemerintah untuk penyederhanaan birokrasi.

Business Process Modeling Notation (BPMN) adalah notasi grafis yang menggambarkan logika dari langkah-langkah dalam proses bisnis. Notasi ini dirancang untuk mengatur urutan proses dan pesan yang mengalir antara pelaku yang terlibat dalam berbagai kegiatan. Yang membedakan BPMN dengan tools lainnya dalam membuat bisnis proses adalah BPMN yang sudah dilengkapi dengan proses validasi sehingga jika terjadi kesalahan dalam prosesnya maka bisnis prosesnya tidak akan bisa di eksekusi atau running di proses simulasi.



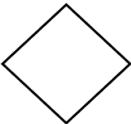

BPMN memiliki 5 kategori dari elemen dasar, yaitu:

1. *Flow objects*, merupakan elemen grafis untuk menentukan perilkasi atau kebiasaan dari suatu bisnis proses. Flow object terdiri dari:
 - *Event*
 - *Activity*
 - *Gateway*
2. Data, dipresentasikan dengan 4 elemen pada BPMN, yaitu:
 - *Data object*
 - *Data input*
 - *Data output*
 - *Data store*
3. *Connecting objects*, terdapat 4 connecting objects yang menghubungkan satu sama lain dari flow object, yaitu:
 - *Sequence flow*

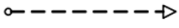
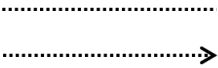


- *Message flow*
 - *Association*
 - *Data association*
4. *Swimlanes*, merupakan pengelompokan dari unsur-unsur pemodelan utama, yaitu berupa Pool dan Lane.
 5. *Artifacts*, digunakan untuk memberikan informasi tambahan tentang suatu proses. Terdapat dua jenis artifact standar dalam BPMN, yaitu Group dan Text Annotation. Namun pemodel atau alat pemodelan bebas menambahkan Artefak sebanyak yang diperlukan.





Untuk mengetahui elemen dasar pada model BPMN, dapat melihat Tabel 2.1.

Tabel 2.1: Elemen Dasar BPMN

No.	Elemen	Deskripsi	Notasi
1	Event	Sesuatu yang terjadi selama berjalannya suatu proses atau koreografi. Event mempengaruhi aliran model dan biasanya memiliki sebab (trigger) atau dampak. Ada tiga jenis event, yaitu: Awal, Menengah, dan Akhir.	
2	Activity	Suatu kegiatan yang menggambarkan atau memperlihatkan perusahaan melakukan suatu proses.	
3	Gateway	Gateway digunakan untuk mengontrol perbedaan dan konvergensi dari urutan Arus dalam Proses. Dengan demikian, akan menentukan branching, forking, merging, dan joining of paths	
4	Sequence Flow	Digunakan untuk menunjukkan urutan tindakan yang akan dilakukan selama proses dan dalam koreografi.	

2.2. BUSSINESS MODEL NOTATION (BPMN) BAB 2. TELAAH PUSTAKA

No.	Elemen	Deskripsi	Notasi
5	Message Flow	Digunakan untuk menunjukkan aliran pesan yang terjadi antara dua orang yang telah dipersiapkan untuk mengirim dan menerima. Diagram Kolaborasi BPMN akan menggambarkan dua peserta dalam dua pool berbeda.	
6	Association	Digunakan untuk menghubungkan informasi dan Artefak dengan elemen grafis BPMN. Anotasi Teks dan Artefak lainnya dapat diasosiasikan dengan elemen grafis. Panah pada Asosiasi menunjukkan arah aliran (misalnya, data), bila sesuai.	
7	Pool	Peserta dalam Kolaborasi direpresentasikan secara grafis sebagai Pool. Selain itu, pool berfungsi sebagai "swimlane" dan wadah grafis untuk memisahkan grup Aktivitas dari Kumpulan lainnya, biasanya dalam skenario B2B. Pool bisa menyimpan informasi internal, seperti Proses yang akan dijalankan. Atau Pool mungkin tidak berisi informasi internal apa pun, dalam hal ini mungkin berupa black box.	
8	Lane	Digunakan untuk mengatur dan mengkategorikan Aktivitas. Lane adalah sub-partisi dalam Proses, terkadang di dalam Kumpulan, dan akan memperpanjang keseluruhan Proses, baik secara vertikal maupun horizontal.	

No.	Elemen	Deskripsi	Notasi
9	Data Object	Input data dan output data memberikan informasi yang sama untuk proses. Objek data dapat berupa satu objek atau kumpulan objek, dan mereka memberikan informasi tentang tugas apa yang harus dilakukan dan apa yang mereka hasilkan.	
10	Message	Digunakan untuk menggambarkan isi percakapan antara dua peserta.	
11	Group	Mengelompokkan elemen grafis dalam kategori yang sama dalam grup tidak mempengaruhi arus urutan di dalam grup. Pada diagram, nama kategori muncul sebagai label grup. Kategori dapat digunakan untuk tujuan analisis atau dokumentasi. Salah satu cara untuk menampilkan kategori objek pada diagram adalah dengan menggunakan grup.	
12	Text Annotation	Pemodel dapat memberi pembaca Diagram BPMN informasi teks tambahan melalui anotasi teks.	

2.3 *Large Language Model* (LLM)

Large Language Model (LLM) dibuat untuk meramalkan hasil bahasa yang potensial sebagai reaksi terhadap petunjuk tertentu. Perintah ini dapat mencakup menebak kata yang paling mungkin setelah serangkaian kata atau menghasilkan kalimat atau paragraf yang masuk akal sebagai tanggapan atas perintah tekstual yang diberikan. Jika diberikan perintah "hujan kucing dan," LLM dapat menghasilkan keluaran "anjing." Model ini menggunakan jaringan

saraf yang dilatih dengan sejumlah besar data teks untuk mengenali pola kata dan bahasa yang mungkin terjadi. Pola-pola ini kemudian disimpan sebagai bobot di dalam jaringan saraf.

LLM terdiri dari jaringan saraf yang dilatih pada korpus teks yang luas untuk mengidentifikasi pola kata dan bahasa yang mungkin, yang kemudian disimpan sebagai bobot dalam jaringan saraf ([Brown et al., 2020],[Vaswani et al., 2017]). Meskipun LLM dapat dilatih untuk tugas dan bidang tertentu, popularitas yang semakin meningkat dapat dikaitkan dengan kemenangan model yang telah dilatih sebelumnya seperti LaMDA [Thoppilan et al., 2022] dan GPT [OpenAI, 2023]. Model ini telah dilatih dengan data yang sangat besar oleh perusahaan yang kemudian menawarkan akses kepada pengguna melalui API, antarmuka *web*, atau aplikasi perangkat lunak. Implementasi Large Language Model yang semakin pesat menghasilkan banyak pre-train model yang telah dilatih menggunakan banyak data tersedia di internet, sehingga proses *training* yang sebelumnya membutuhkan waktu lama sekarang tidak perlu dilakukan lagi.

2.3.1 *Generative Pre-trained Transformer (GPT)*

Generative Pre-trained Transformers, umumnya dikenal sebagai GPT, adalah keluarga model jaringan saraf yang menggunakan arsitektur *transformer* dan merupakan kemajuan utama dalam *Artificial Intelligence* (AI) atau kecerdasan artifisial yang memberdayakan aplikasi AI generatif seperti ChatGPT. Model GPT memberikan kemampuan kepada aplikasi untuk membuat teks dan konten yang mirip manusia (gambar, musik, dan lainnya), serta menjawab pertanyaan dengan cara percakapan. Organisasi di seluruh industri menggunakan model GPT dan AI generatif untuk bot tanya jawab, rangkuman teks, pembuatan konten, dan pencarian.

GPT dapat digunakan untuk :

- Membuat konten sosial media.
- Menulis dan Belajar Kode.
- Menganalisis Data.
- Menghasilkan Materi Pembelajaran.
- Membangun Asisten Suara Interaktif.

2.3.2 *Bidirectional Encoder Representations from Transformers* (BERT)

Inovasi teknis utama BERT adalah menerapkan pelatihan dua arah dari *Transformer*, sebuah model perhatian yang populer, pada pemodelan bahasa. Hal ini berbeda dengan pendekatan sebelumnya yaitu n-gram, BERT melihat urutan teks baik dari kiri ke kanan atau menggabungkan pelatihan kiri-ke-kanan dan kanan-ke-kiri. Hasil penelitian menunjukkan bahwa model bahasa yang dilatih dua arah dapat memiliki pemahaman yang lebih dalam tentang konteks dan aliran bahasa daripada model bahasa satu arah. Dalam makalah tersebut, para peneliti merinci teknik baru bernama *Masked LM* (MLM) yang memungkinkan pelatihan dua arah dalam model yang sebelumnya tidak mungkin dilakukan.

BERT dapat digunakan untuk berbagai macam tugas bahasa, dengan hanya menambahkan lapisan kecil pada model inti:

- Tugas klasifikasi seperti analisis sentimen dilakukan dengan cara yang mirip dengan klasifikasi *Next Sentence*, dengan menambahkan lapisan klasifikasi di atas output *Transformer* untuk token [CLS].
- Dalam tugas Menjawab Pertanyaan (misalnya SQuAD v1.1), perangkat lunak menerima pertanyaan mengenai urutan teks dan diminta untuk menandai jawaban dalam urutan tersebut. Penggunaan BERT suatu model tanya jawab dapat dilatih dengan mempelajari dua vektor tambahan yang menandai awal dan akhir jawaban.
- Dalam *Named Entity Recognition* (NER), perangkat lunak menerima urutan teks dan diminta untuk menandai berbagai jenis entitas (Orang, Organisasi, Tanggal, dll) yang muncul di dalam teks. Penggunaan BERT suatu model NER dapat dilatih dengan memasukkan vektor keluaran dari setiap token ke dalam lapisan klasifikasi yang memprediksi label NER.

2.4 Grounded Theory

[Glaser and Strauss, 1967] mengembangkan Grounded Theory untuk mendeskripsikan proses induktif dalam menghasilkan kode dan mengidentifikasi kategori analitik yang muncul, mengakar dari data (grounded), bukan dari konsep

yang telah ditetapkan sebelumnya. Proses ini sangat mirip dengan proses pada analisis tematik, tetapi Grounded Theory mempunyai karakteristik utama berupa proses yang berulang (siklus) dan iterasi, yaitu hasil analisis digunakan sebagai dasar pemilihan sampling informan berikutnya, kemudian dilakukan pengumpulan data dan selanjutnya dihasilkan sebuah teori. Sampling teoritis (Theoretical sampling) menjadi bagian penting dalam Grounded Theory memungkinkan peneliti memilih informan atau setting baru untuk menguji kategori dan teori yang dihasilkan. Proses analisis dilakukan hingga tercapai saturasi [Utari, 2021].

2.4.1 Proses Grounded Theory

Proses Grounded Theory menggunakan 3 tahapan besar yaitu:

1. Open Coding. Mengidentifikasi sebanyak mungkin kode dari kalimat per kalimat dalam transkrip. Selanjutnya, dilakukan identifikasi properti dan dimensi kode dengan bertanya “apa karakteristiknya?” dan “bagaimana bentuknya?”.
2. Axial Coding. Untuk menemukan keterkaitan antar kode yang didapatkan dari open coding.
3. Selective Coding. Mengembangkan kategori analitis dengan elemen yang lebih abstrak dan teoritis, dengan melakukan perbandingan terus menerus (constant comparison) akan dihasilkan sebuah kategori analisis

Dengan cara ini, secara bertahap peneliti akan membangun sebuah teori, paralel dengan menguji beberapa ide-ide yang muncul. Grounded theory akan menghasilkan interpretasi yang rinci dan kaya.

2.4.2 Karakteristik Grounded Theory

Selain penekanannya pada pengembangan teori, apa yang membuat Grounded Theory unik dari bentuk penelitian kualitatif lainnya? Jawaban dari pertanyaan ini cukup sederhana [Corbin and Strauss, 2015].

1. Pertama, konsep-konsep yang menjadi dasar teori berasal dari data yang dikumpulkan selama proses penelitian dan tidak dipilih sebelum memulai penelitian. Fitur inilah yang mendasari teori dan memberi nama metodologi tersebut.

2. Kedua, dalam Grounded Theory, analisis penelitian dan pengumpulan data saling terkait. Setelah data awal dikumpulkan, peneliti menganalisis data tersebut, dan konsep-konsep yang diperoleh dari analisis tersebut menjadi dasar untuk pengumpulan data berikutnya. Pengumpulan dan analisis data terus berlanjut dalam siklus yang berkelanjutan selama proses penelitian.

Secara garis besar Grounded Theory dapat digunakan dalam melakukan analisis data kualitatif (Pope dan Mays, 2006) sehingga Grounded Theory bisa berdiri sendiri sebagai sebuah teori maupun menjadi sebuah perangkat untuk melakukan analisis data kualitatif.

2.5 Penelitian Terkait

Beberapa penelitian terkait mengenai penelitian ini dapat dilihat pada tabel

Tabel 2.2: Penelitian Terkait

Peneliti	Judul	Hasil Penelitian
----------	-------	------------------

<p>Sholiq Sholiq, Riyanarto Sarno, Endang Siti Astuti [Sholiq et al., 2022]</p>	<p>Generating BPMN diagram from textual re- quirements</p>	<p>Melakukan penelitian dengan melakukan konversi dari persyaratan tekstual ke diagram BPMN dengan mengatasi dua kelemahan dari penelitian sebelumnya, yaitu: 1. tidak mempertimbangkan struktur kalimat sebagai input natural language (NL) dan metode yang diusulkan membutuhkan input NL berurutan sehingga tidak memungkinkan untuk pemrosesan berulang; 2. Hanya dapat menangani diagram BPMN terbatas, tidak ada validasi yang dilakukan dari proses ekstraksi aktivitas dan tidak diuji menggunakan struktur kalimat kompleks dan majemuk pada input NL. Metode yang diusulkan memiliki dua tahap: 1) menganalisis persyaratan tekstual menggunakan pemrosesan bahasa alami dan 2) menghasilkan diagram BPMN. Metode yang diusulkan diterapkan pada sepuluh studi kasus persyaratan tekstual dalam aplikasi kooperatif. Metode yang diusulkan dapat menangani kalimat kompleks, kalimat kompleks majemuk, kalimat transitif, aliran input/ output masuk/keluar datastore atau objek data, dan kata keterangan cara yang terdapat dalam input tekstual. Dibandingkan dengan diagram BPMN yang dihasilkan secara manual oleh seorang ahli, diperoleh tingkat akurasi sebesar 92,83%, lebih tinggi dari akurasi yang dicapai pada penelitian sebelumnya</p>
---	--	---

<p>Ana Ivanchikj, Souhaila Serbout, Cesare Pautasso [Ivanchikj et al., 2020]</p>	<p>From Text to Visual BPMN Process Models: Design and Evaluation</p>	<p>Sebagian besar alat pengeditan Model dan Notasi Proses Bisnis (BPMN) yang ada adalah grafis, dan karena itu didasarkan pada pemodelan eksplisit, membutuhkan pengetahuan yang baik tentang notasi dan semantiknya, serta kemampuan untuk menganalisis dan abstrak persyaratan bisnis dan menangkapnya dengan benar. menggunakan notasi. Akibatnya, penggunaannya bisa menjadi rumit untuk pemodelan langsung selama wawancara dan lokakarya desain, di mana peserta tidak hanya harus memberikan masukan tetapi juga memberikan umpan balik tentang bagaimana hal itu telah direpresentasikan dalam model. Dalam tulisan tersebut menyajikan desain dan evaluasi BPMN Sketch Miner, sebuah alat yang menggabungkan pencatatan dalam bahasa alami yang dibatasi dengan penambahan proses untuk secara otomatis menghasilkan diagram BPMN secara real-time saat peserta wawancara meng gambarkannya dengan cerita. Hasil dari evaluasi pada makalah menunjukkan bahwa trade-off antara ekspresi dan kegunaan seimbang.</p>
--	---	---

<p>Fajri Koto, Afshin Rahimi, Jey Han Lau, Timothy Baldwin [?]</p>	<p>IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP</p>	<p>Bahasa Indonesia merupakan Bahasa ke-10 yang paling banyak digunakan di dunia, namun kurang terwakili dalam penelitian NLP karena kurangnya kumpulan data beranotasi, kurangnya sumber daya Bahasa dan kurangnya standarisasi sumber daya. Kemudian dibuatlah dataset IndoLEM (Indonesian Language Evaluation Montage) yang berisi tujuh tugas untuk bahasa Indonesia, yang meliputi morphosyntax, semantik, dan wacana. Peneliti juga mengembangkan IndoBERT, model bahasa BERT terlatih monolingual untuk bahasa Indonesia dan mengevaluasinya melalui IndoLEM serta di- bandingkan dengan sumber daya yang ada. Hasil yang didapat menunjukkan IndoBERT mencapai kinerja cangguh pada IndoLEM.</p>
--	--	--

Maarten Grootendorst [Grootendorst, 2022]	BERTopic: Neural topic modeling with a class-based TF-IDF procedure	Peneliti membuat BERTopic yang merupakan model topik yang mengembangkan proses dengan mengekstraksi dan menghasilkan representasi topik yang koheren melalui pengembangan dari variasi TF-IDF yang berbasis kelas, BERTopic juga menghasilkan penyematan dokumen dengan model Bahasa berbasis transformator yang telah dilatih. Hasil dari penelitian menyebutkan bahwa BERTopic mempelajari pola bahasa yang koheren dan menunjukkan kinerja yang kompetitif dan stabil di berbagai tugas.
---	---	---

<p>Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova [Devlin et al., 2019]</p>	<p>BERT: Pre- training of Deep Bidi- rectional Transfor- mers for Language Unders- tanding</p>	<p>BERT (Bidirectional Encoder Representations from Transformers) merupakan model yang dirancang untuk melatih representasi dua arah dari teks yang tidak berlabel dengan mengkondisikan konteks kiri dan kanan. Dengan model BERT pre-train ini dapat disesuaikan untuk membuat model canggih untuk berbagai tugas seperti menjawab pertanyaan dan inferensi Bahasa tanpa ada tugas khusus yang substansial, dengan hanya satu lapisan keluaran tambahan. Dengan menggunakan BERT diperoleh hasil baru untuk sebelas tugas pemrosesan Bahasa alami: 1. skor GLUE menjadi 80,5% (peningkatan absolut poin 7,7%); 2. akurasi MultiNLI menjadi 86,7% (peningkatan absolut 4,6%); 3. SQuAD v1. 1 soal menjawab Tes F1 menjadi 93,2 (peningkatan absolut 1,5 poin); 4. Tes F1 SQuAD v2.0 menjadi 83,1 (peningkatan absolut 5,1 poin)</p>
---	--	--

<p>Christoph Schroerab, Felix Kruseb, Jorge Marx Gomez [Schrouer et al., 2021]</p>	<p>A Systematic Literature Review on Applying CRISP-DM Process Model</p>	<p>Makalah ini mengulas 24 penelitian terbaru mengenai fase CRISP-DM, sebuah model proses standar dalam proyek-proyek data mining. Studi ini menganalisis alasan pemilihan CRISP-DM, domain, dan fase (pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran). Studi ini menyarankan penyelidikan lebih lanjut terhadap potensi keterbatasan dan peningkatan, adopsi industri, dan pemilihan teknologi. Namun, fase penerapan sebagian besar tidak ada, sehingga memerlukan penelitian lebih lanjut mengenai pengintegrasian model ke dalam lingkungan yang produktif.</p> <p>Translated with DeepL.com (free version)</p>
--	--	--

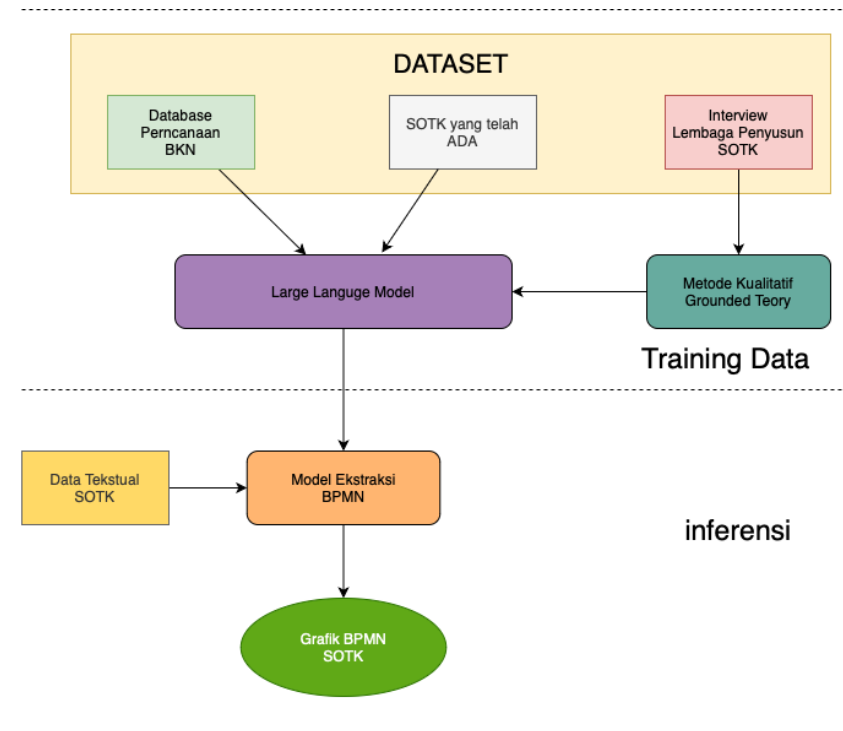
Bab 3

Metode Penelitian

3.1 Metode yang Dikembangkan

Penelitian ini mengembangkan sebuah metode ekstraksi grafik Bussiness Proses Model Notation (BPMN) dengan menggunakan pendekatan Large Language Model. Metode yang dikembangkan adalah memadukan metode kualitatif grounded theory untuk mendapatkan informasi tentang penyusunan suatu SOTK instansi dengan Large Language Model untuk mendapatkan sebuah model ekstraksi BPMN berdasarkan pada data tekstual SOTK yang dimiliki oleh suatu Instansi. Gambar 3.1 menggambarkan metode yang dikembangkan, metode ini terdiri dari dua tahapan utama yaitu:

1. **Training Data.** Proses mengembangkan sebuah model baru berdasarkan dataset yang ada untuk dapat melakukan ekstraksi grafik BPMN. Dalam Tahapan ini terdapat beberapa komponen yaitu:
 - **Data Set.** Sekumpulan data yang digunakan untuk melatih model dalam melakukan ekstraksi BPMN, terdiri dari:
 - **Database Perencanaan Kepegawaian.** Database ini berisi data seluruh SOTK yang ada di lembaga negara di Indoneisa
 - **SOTK yang telah ada.** Merupakan gambaran SOTK yang telah dan sudah disusun sebelumnya
 - **Hasil interview lembaga penyusun SOTK.** Melakukan interview kepada lembaga-lembaga yang telah memiliki SOTK untuk mengetahui komponen apa saja yang perlu diperhatikan dalam melakukan penyusunan SOTK, data ini digunakan untuk malakukan Retrieval Augmented Generation (RaG).



Gambar 3.1: Metode yang dikembangkan

- Metode. Metode digunakan untuk menghasilkan model yang akurat dalam melakukan ekstraksi BPMN:
 - Metode Grounded Theory. Metode untuk mengekstraksi komponen-komponen apa saja yang harus diperhatikan dalam mengembangkan model yang dapat digunakan untuk meningkatkan akurasi dari model LLM yang telah ada.
 - LLM. Metode Large Language Model digunakan sebagai pre-train model yang dapat menghasilkan sebuah grafik BPMN dari data tekstual.

2. **Inferensi.** Proses melakukan ekstraksi grafik BPMN berbasiskan pada data tekstual yang ada, Bagian ini terdiri dari:

- (a) Data Tekstual. Berisi data-data yang diperlukan dalam penyusunan suatu SOTK lembaga
- (b) Model Ekstasi BPMN. Sebuah model yang dihasilkan pada proses Training Data, model ini merupakan hasil utama dari penelitian ini.

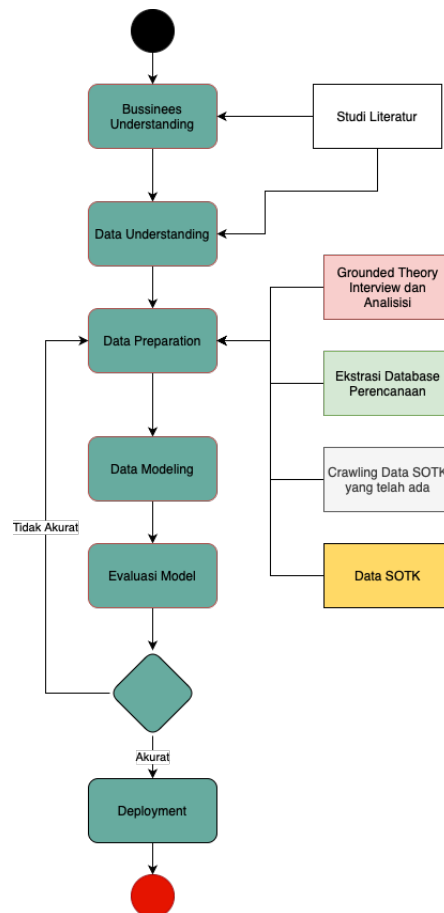
Metode yang dikembangkan ini diharapkan dapat meningkatkan akurasi hasil dari model BPMN yang dihasilkan oleh pretrain model yang telah ada.

3.2 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan berbasiskan CrispDM yang telah dilakukan review oleh [Schrouer et al., 2021] adalah sebagai berikut (Dapat dilihat pada Gambar 3.2):

1. Bussiness Understanding (Memahami urusan). Pada tahap ini, penelitian akan memfokuskan pada pemahaman yang mendalam terhadap domain bisnis yang relevan dengan masalah yang diteliti. Ini mencakup:
 - (a) Deskripsi masalah penelitian dan tujuan yang ingin dicapai.
 - (b) Identifikasi stakeholder yang terlibat dan kebutuhan.
 - (c) Analisis literatur terkait untuk memahami konteks dan penelitian terdahulu.
2. Data Untderstanding (Memahami Data). Langkah ini bertujuan untuk mengumpulkan data yang diperlukan untuk analisis, serta memahami karakteristik data yang tersedia:
 - (a) Sumber data yang digunakan dan proses pengumpulannya.
 - (b) Deskripsi dataset yang terkumpul, termasuk jumlah, jenis, dan struktur data.
 - (c) Evaluasi kualitas data dan langkah-langkah yang diambil untuk membersihkan dan mempersiapkan data.
3. Data Preparation (Mempersiapkan Data). Tahap ini mencakup semua aktivitas yang diperlukan untuk mempersiapkan data untuk analisis lebih lanjut:
 - (a) Pemilihan variabel yang relevan untuk analisis.
 - (b) Integrasi data dari berbagai sumber jika diperlukan.
 - (c) Transformasi data untuk memenuhi persyaratan analisis yang spesifik.
4. Modeling (Modeling). Data yang telah dipersiapkan akan digunakan untuk membangun model prediktif atau deskriptif sesuai dengan tujuan analisis yang telah ditetapkan. Proses ini meliputi pemilihan teknik model yang sesuai, pembangunan model, dan evaluasi kualitas model untuk memastikan keakuratannya sebelum tahap selanjutnya.

5. Evaluation (Mengevaluasi). Evaluasi memiliki tujuan untuk mengukur kinerja dan keefektifan model yang telah dibangun pada tahap sebelumnya.
6. Deployment (Penerapan). Mengimplementasikan model yang telah dibangun menjadi sebuah sistem informasi yang siap digunakan oleh user.



Gambar 3.2: tahapan penelitian

Bibliografi

- [Brown et al., 2020] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners.
- [Corbin and Strauss, 2015] Corbin, J. and Strauss, A. (2015). *Basic of Qualitative Research*. SAGE Publications, Inc., fourth edition.
- [Devlin et al., 2019] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding.
- [Glaser and Strauss, 1967] Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. A Division of Transaction Publishers New Brunswick (U.S.A.) and London (U.K.).
- [Grootendorst, 2022] Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- [Ivanchikj et al., 2020] Ivanchikj, A., Serbout, S., and Pautasso, C. (2020). From text to visual bpmn process models: design and evaluation. In *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MODELS '20*. ACM.
- [OpenAI, 2023] OpenAI (2023). Gpt-4 technical report.
- [Schrouer et al., 2021] Schrouer, C., Kruse, F., and Gomez, J. M. (2021). A systematic literature review on applying crisp-dm process model. *Procedia Computer Science*, 181:526–534.

- [Sholiq et al., 2022] Sholiq, S., Sarno, R., and Astuti, E. S. (2022). Generating bpmn diagram from textual requirements. *Journal of King Saud University - Computer and Information Sciences*, 34(10):10079–10093.
- [Thoppilan et al., 2022] Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., Chen, D., Xu, Y., Chen, Z., Roberts, A., Bosma, M., Zhao, V., Zhou, Y., Chang, C.-C., Krivokon, I., Rusch, W., Pickett, M., Srinivasan, P., Man, L., Meier-Hellstern, K., Morris, M. R., Doshi, T., Santos, R. D., Duke, T., Soraker, J., Zevenbergen, B., Prabhakaran, V., Diaz, M., Hutchinson, B., Olson, K., Molina, A., Hoffman-John, E., Lee, J., Aroyo, L., Rajakumar, R., Butryna, A., Lamm, M., Kuzmina, V., Fenton, J., Cohen, A., Bernstein, R., Kurzweil, R., Aguera-Arcas, B., Cui, C., Croak, M., Chi, E., and Le, Q. (2022). Lamda: Language models for dialog applications.
- [Utari, 2021] Utari, A. (2021). *Tak Kenal Maka Tak Sayang: Penelitian Kualitatif Dalam Pelayanan Masyarakat*.
- [Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need.