BAB I
PENDAHULUAN

1.1. Latar Belakang

Pesatnya perkembangan aplikasi dan sistem komputer seiring waktu berdampak pada kompleksitasnya yang semakin bertambah. Karena sistem dan aplikasi menjadi semakin kompleks daripada sebelumnya, maka akan lebih banyak *bug* dan kerentanan yang dapat dieksploitasi oleh orang tidak bertanggung jawab untuk meluncurkan serangan. Serangan semacam itu juga semakin canggih sejalan dengan pesatnya perkembangan aplikasi dan sistem informasi [1]. Selain *bug* dan kerentanan, dikarenakan sistem ini juga menyediakan berbagai layanan untuk sejumlah besar pengguna, masalah kecil dalam sistem dapat menyebabkan ketidakpuasan pengguna dan bahkan kerugian finansial yang signifikan [2]. Maka dari itu deteksi anomali pada sebuah sistem adalah tugas penting untuk membangun sistem komputer yang aman dan dapat dipercaya dengan ketersediaan dan keandalan yang tinggi atau biasa disebut dengan *High Availability* (HA) dan *Reliability*.

Mendeteksi peristiwa anomali dari sistem komputer secara tepat adalah langkah mendasar untuk melindungi sistem. *File log* dari suatu sistem mencatat informasi rinci tentang peristiwa komputasi yang dihasilkan oleh komputer, maka dari itu *file log* memainkan peran penting dalam deteksi anomali saat ini [3]. *File log* menyediakan sumber informasi yang kaya dalam hal pemantauan sistem komputer. Dikarenakan, sebagian besar peristiwa pada *file log* biasanya dihasilkan sebagai pencatatan dari operasi sistem yang normal, seperti memulai dan menghentikan suatu proses dalam sistem komputer, memulai ulang suatu proses atau mesin virtual, pengguna yang mengakses sumber daya, dan sebagainya. Namun, aplikasi atau sistem komputer juga menghasilkan log ketika terjadi kondisi sistem yang salah atau tidak diinginkan, misalnya, proses yang gagal, masalah ketersediaan, atau insiden keamanan [4]. Jejak aktivitas sistem yang tidak terduga dan mungkin tidak aman ini penting bagi operator sistem atau administrator sistem untuk memutuskan apakah jejak aktivitas yang ada pada *file log* perlu ditindaklanjuti atau tidak sehingga dapat bertindak secara tepat waktu untuk

mencegah atau mengurangi kerusakan sistem dan menghindari efek berkepanjangan yang merugikan.

Di sisi lain, log secara tradisional sulit untuk diolah. *Log file* biasanya memiliki ukuran yang cukup besar (dalam jutaan kejadian) [5]. Volume log yang sangat besar membuatnya tidak praktis untuk secara manual memeriksa pesan log untuk informasi diagnostik, bahkan jika hanya dilakukan dengan utilitas *search* dan *grep* [6]. Selain itu, praktik pencatatan log dikenal sebagai praktik yang dilakukan secara *ad hoc*, tanpa adanya praktik terbaik atau standar umum yang ditetapkan [7]. Permasalahan dalam mengolah *log file* pun semakin buruk,dikarenakan log pada umumnya merupakan file data yang tidak terstruktur karena kurangnya format standar. Akibatnya, mengekstraksi informasi dan menganalisi informasi yang relevan dari *file log* mentah yang besar bisa menjadi tugas yang memakan waktu jika dilakukan secara manual [8]. Karena sifat log yang tidak terstruktur, langkah utama yang paling penting adalah mengurai pesan log menjadi data terstruktur untuk analisis selanjutnya [6]. Dalam konteks ini data terstruktur memudahkan untuk digunakan dalam mendeteksi anomali pada *log file*.

Untuk mengatasi tantangan dalam mendeteksi anomali pada *log file* adalah dengan memanfaatkan teknologi *Natural Language Processing* (NLP). NLP adalah bidang dalam ilmu komputer dan kecerdasan buatan yang berfokus pada pemahaman, pemrosesan, dan generasi bahasa manusia secara alami oleh mesin. NLP memungkinkan komputer untuk menganalisis, memahami, dan memanipulasi teks dan percakapan dalam bahasa manusia. Teknik NLP melibatkan berbagai aspek seperti pemrosesan bahasa alami, analisis sentiment, pemahaman konteks, pemodelan bahasa, dan terjemahan mesin [9]. NLP memainkan peran penting dalam konteks mendeteksi anomali log. NLP membantu memahami makna log dan mempelajari pola dari data log. Dengan kemampuan ini, NLP dapat mendeteksi anomali yang tidak terlihat dengan metode tradisional. Konsep utamanya adalah dengan memodelkan urutan log sebagai bahasa alami dan memahami makna teks log dengan menggunakan *Deep Learning models* seperti *Long Short Term Memory* (LSTM), *Transformer,* dan sebagainya.

Dalam konteks deteksi anomali pada *log file*, arsitektur model *Transformer* sudah digunakan pada beberapa penelitian terdahulu, salah satunya adalah

penelitian yang memperkenalkan model LogBERT. LogBERT sendiri dibuat berdasarkan model *Bidirectional Encoder Representations from Transformers* (BERT) yang menggunakan arsitektur *Transformer* yang kuat untuk melatih dan memahami teks dalam dua arah, baik dari kiri ke kanan (untuk mengerti konteks sebelumnya) maupun dari kanan ke kiri (untuk mengerti konteks berikutnya). Model ini dilatih dengan metode *self-supervised learning* dengan mempelajari pola dari urutan log yang normal melalui prediksi potongan pesan log yang disembunyikan. Hasil dari penelitian tersebut LogBERT mengungguli *state-of-the-art* untuk deteksi anomali log pada saat itu dengan skor F-1 untuk *dataset Hadoop Distribution File System* (HDFS) sebesar 82.32%, *dataset BlueGene/L* (BGL) sebesar 90.83%, dan untuk Thunderbird sebesar 96.64% [3].

Kemudian, untuk penelitian lainnya mengenai deteksi anomali pada log, yang masih menggunakan arsitektur *Transformer* yakni model dengan nama LogGPT yang dibuat berdasarkan model *Generative Pre-trained Transformers* (GPT). Untuk mendeteksi anomali pada log LogGPT dilatih untuk memprediksi entri log berikutnya berdasarkan urutan sebelumnya. Untuk lebih meningkatkan kinerja LogGPT, digunakan sebuah strategi *reinforcement learning* untuk menyempurnakan model secara khusus khusus untuk tugas deteksi anomali log. Hasil eksperimen pada tiga set data menunjukkan bahwa LogGPT secara signifikan mengungguli *state-of-the-art* yang ada dengan skor F-1 pada *dataset* HDFS sebesar 90.1%, untuk *dataset* BGL 95.8%, dan untuk Thunderbird sebesar 98.6% [10].

Mengacu pada dua penelitian tersebut, dimana model yang digunakan dibuat berdasarkan pada model *Transformer* yakni BERT dan GPT, terdapat satu model *Transformer* yang menggabungkan dua arsitektur tersebut, yaitu BART (*Bidirectional and Auto-Regressive Transformers*). BART menggabungkan *Encoder* dua arah yang dimiliki oleh BERT dan *left-to-right decoder* yang dimiliki oleh GPT. BART sangat efektif dan bekerja dengan baik untuk *generative text*, tetapi juga bekerja dengan baik untuk tugas-tugas pemahaman bahasa. BART menyamai kinerja RoBERTa dengan sumber daya pelatihan yang sebanding pada *dataset* GLUE dan SQuAD, dengan mencapai *state-of-the-art* pada berbagai bidang NLP seperti *abstractive dialogue, question answering,* dan *summarization*, dengan peningkatan hingga 6 ROUGE [11]. Meskipun unggul dalam berbagai bidang NLP

tersebut, BART belum teruji untuk melakukan tugas pendeteksian anomali pada *log file*, Berdasarkan hal tersebut, maka peneliti mengambil judul "Implementasi Model BART (*Bidirectional Auto-Regressive Transformers*) Dalam Mendeteksi Anomali pada *Log file Server*" sebagai topik pembahasan penelitian.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat dirumuskan permasalahan yang muncul sebagai berikut :

1. Bagaimana implementasi model BART dapat diterapkan untuk mendeteksi anomali pada *log file server* ?
2. Mengevaluasi performa model BART berdasarkan *precision, recall,* dan *F1-score* dalam mendeteksi anomali pada log file server.
3. Bagaimana perbandingan performa model BART dalam mendeteksi anomali pada *log file server* dibandingkan secara referensial dengan metode pendeteksian anomali log seperti LogBERT dan LogGPT ?
4. Bagaimana perancangan prototipe aplikasi untuk menerapkan model BART dalam mendeteksi anomali pada log file server?

1.3. Batasan Masalah

Berdasarkan rumusan masalah yang telah disebutkan, kita dapat mengambil batasan-batasan pada penelitian ini dengan mempertimbangkan faktor-faktor yang relevan dalam lingkup yang ditentukan. Batasan masalah penelitian ini dapat diberikan sebagai berikut :

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

1. Penelitian ini difokuskan pada penggunaan model BART sebagai metode pendeteksian anomali pada *log file server*.
2. Penelitian ini menggunakan *log file* HDFS (Hadoop Distribution File *System*) , BGL (BlueGene/L) dan Thunderbird yang tersedia pada Loghub dan terbuka untuk publik sebagai sumber data dengan mengambil sebanyak 10.000 *sequence* [12].
3. Perbandingan performa pada penelitian ini menggunakan acuan terhadap model LogBERT dan LogGPT.
4. Prototipe aplikasi yang dirancang mencakup fungsi unggah log, deteksi anomali, penyajian hasil, riwayat analisis, dan informasi model. Prototipe

tidak mencakup pemantauan log secara real-time, autentikasi pengguna, dan penerapan pada lingkungan produksi.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah, dirumuskanlah beberapa poin yang menjadi tujuan penelitian ini sebagai berikut :

1. Mengidentifikasi langkah-langkah implementasi model BART untuk mendeteksi anomali pada *log file server*.
2. Mengetahui bagaimana performa berdasarkan presisi, *Recall*, dan *F1-Score* model BART dalam tugas mendeteksi anomali pada *log file server*.
3. Melakukan perbandingan yang komprehensif terhadap performa model BART dengan metode pendeteksian anomali lainnya berdasarkan hasil pendeteksian anomali.
4. Merancang prototipe aplikasi untuk mendeteksi anomali pada log file server.

1.5. Manfaat Penelitian

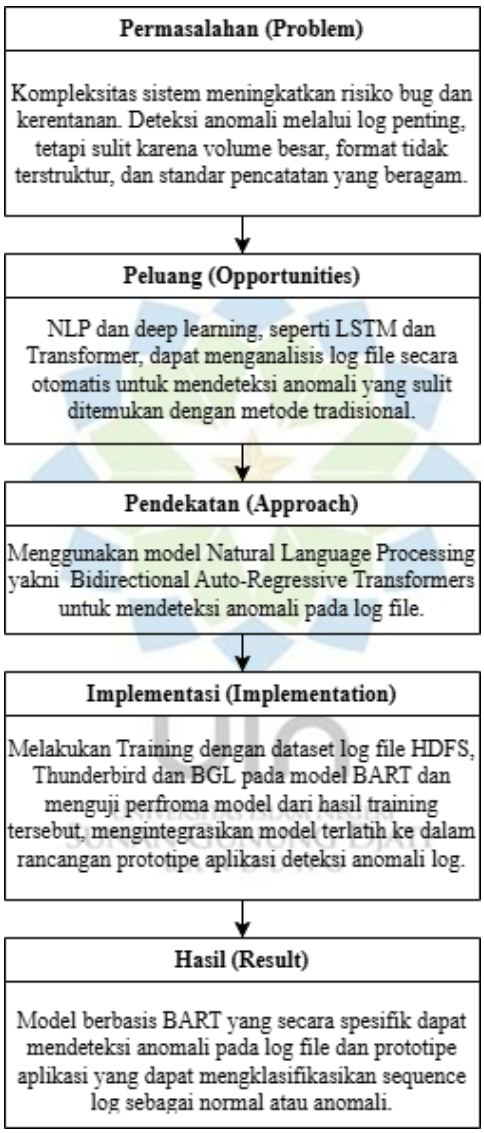
Secara keseluruhan, penelitian ini dapat memberikan manfaat dalam penerapan model BART sebagai solusi untuk mendeteksi anomali pada *log file*. Adapun dalam rincian penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Implementasi model BART dalam mendeteksi anomali pada *log file server* dapat meningkatkan kemampuan sistem dalam mengidentifikasi dan melacak perilaku yang tidak biasa atau mencurigakan. Hal ini dapat membantu dalam mendeteksi potensi serangan, kerentanan atau kegagalan sistem.
2. Membuka peluang penelitian baru di bidang pendeteksian anomali dan keamanan siber.
3. Penelitian ini dapat memberikan kontribusi pada bidang *Natural Language Processing* (NLP) dengan mengaplikasikan model BART dalam konteks pendeteksian anomali pada *log file*.
4. Meningkatkan kesadaran tentang pentingnya pendeteksian anomali untuk menjaga keamanan dan keandalan sistem.

5. Menghasilkan *prototype* aplikasi yang dapat memudahkan pengguna dalam mengunggah log dan melihat hasil deteksi anomali.

1.6. Kerangka Pemikiran Penelitian

Kerangka pemikiran dituangkan pada gambar 1.1 berikut :



Gambar 1.1 Kerangka Pemikiran Penelitian

Pada Gambar 1.1 dijelaskan upaya mendeteksi anomali pada *log file* sistem komputer. Permasalahan utama adalah meningkatnya kompleksitas sistem yang menyebabkan munculnya bug dan kerentanan, serta tantangan dalam pengolahan *log file* yang besar dan tidak terstruktur. Untuk mengatasi hal ini, teknologi *Natural Language Processing* (NLP) dimanfaatkan, khususnya dengan model *Deep*

Learning seperti BART (*Bidirectional Auto-Regressive Transformers*), yang mampu menganalisis pola dalam *log file* secara efisien. Implementasinya dilakukan dengan melatih model menggunakan *dataset log file* seperti HDFS, Thunderbird dan BGL, menghasilkan model berbasis BART yang efektif dalam mendeteksi anomali pada *log file*.

1.7. Sistematika Penulisan

Guna mencapai tujuan dan memberikan gambaran umum dari penelitian tugas akhir ini, maka dipergunakan sistematika penulisan. Adapun sistematika penulisan tugas akhir dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini diberikan gambaran awal tentang penelitian yang dilakukan, dimulai dengan latar belakang masalah yang menjelaskan konteks dan pentingnya penelitian. Selain itu, bab ini mencakup rumusan masalah dalam bentuk pertanyaan yang akan dijawab, tujuan dan manfaat penilitian dari penelitian, serta batasan masalah untuk membuat pembahasan lebih fokus. Selanjutnya adalah kerangka pemikiran penelitian, kerangka pemikiran penilitan ini akan memberikan dasar teoretis atau konseptual untuk analisis. Terakhir, sistematika penulisan akan menjelaskan bagaimana bab-bab berikutnya akan disusun dan diuraikan.

BAB II KAJIAN LITERATUR

Kajian literatur menggabungkan berbagai referensi dan kerangka teori yang relevan dengan penelitian yang dilakukan. Sumber referensi dan landasan teori dapat berasal dari beragam sumber, seperti jurnal ilmiah, karya akademik, buku, dan referensi lainnya.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian menguraikan metode dan tahapan penelitian yang dilakukan. Metode dan tahapan yang digunakan dirancang agar sesuai dengan kebutuhan penelitian, disertai dengan penjelasan terperinci mengenai setiap langkah dan teknik yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas mengenai seluruh proses implementasi yang dilakukan dan menyajikan data dan analisis yang diperoleh dalam penelitian. Hasil penelitian dijelaskan dalam bentuk narasi, tabel, atau grafik. Kemudian, temuan dibahas

dengan mengaitkannya dengan teori atau penelitian sebelumnya untuk menjawab rumusan masalah dan menjelaskan temuan secara menyeluruh.

BAB V KESIMPULAN DAN SARAN

Bab ini merangkum temuan utama dari penelitian dalam bentuk kesimpulan yang menjawab rumusan masalah. Selain itu, bab ini memberikan saran berdasarkan hasil penelitian, baik untuk pengembangan penelitian selanjutnya. Bab ini menjadi penutup yang menggarisbawahi kontribusi penelitian terhadap ilmu pengetahuan atau praktik.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG