

BAB I PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi dari zaman ke zaman, membuat data dan informasi menjadi sangat penting bagi manusia, kemajuan teknologi informasi yang terus maju selanjutnya akan membutuhkan kapasitas penyimpanan yang besar. Kebutuhan ini muncul karena jumlah data yang disimpan secara berkesinambungan terus meningkat sehingga membutuhkan teknik kompresi[1]. Kompresi adalah salah satu jawaban untuk menghemat penyimpanan dan waktu yang lebih cepat untuk mentransmisikan data. Pengadaan kompresi juga memberikan jawaban terhadap kebutuhan telekomunikasi generasi ke-5 (5G) yang kedepannya diharapkan dapat menghasilkan banyak aplikasi yang bermanfaat[2].

Kompresi data memiliki peran yang sangat penting dalam bidang teknologi informasi, khususnya untuk mengurangi ukuran *file* tanpa mengorbankan informasi yang terkandung di dalamnya[3]. Efisiensi dalam proses kompresi berpengaruh besar terhadap kecepatan transmisi data, kebutuhan ruang penyimpanan, serta efisiensi penggunaan daya. Salah satu teknik yang paling umum digunakan dalam pengkodean sumber adalah Huffman *coding*, yang beroperasi dengan mengoptimalkan panjang kode berdasarkan frekuensi simbol yang muncul. Huffman *coding* telah diterapkan dalam berbagai aplikasi, seperti JPEG, MP3, dan berbagai algoritma kompresi teks lainnya. Sebagian besar implementasi Huffman *coding* yang ada saat ini menggunakan pendekatan berbasis *biner*[4].

Pendekatan Huffman berbasis *biner* memiliki kompleksitas yang rendah, namun efisiensinya masih dapat ditingkatkan. Beberapa teknik yang dapat digunakan untuk meningkatkan efisiensi dan rasio kompresi pada metode *binary Huffman codes* antara lain adalah *adaptive Huffman codes* dan *ternary Huffman codes*[4].

Ternary Huffman codes, yang menggunakan tiga simbol (0, 1, dan 2), memiliki potensi untuk meningkatkan efisiensi kompresi. Penggunaan *ternary Huffman codes* sangat bermanfaat untuk mendukung kebutuhan komunikasi 5G dan teknologi selanjutnya yang memerlukan kecepatan pengolahan data serta

efisiensi penyimpanan[5]. Berbagai aplikasi 5G dan seterusnya yang membutuhkan *source coding*, seperti komunikasi kendaraan ke kendaraan (V2V), aplikasi penerjemah bahasa, dan komunikasi kendaraan ke orang (V2P) sangat diuntungkan dengan penerapan teknik ini. Komunikasi pada 5G dan seterusnya kini sudah memungkinkan untuk berlangsung dalam berbagai bahasa, termasuk bahasa daerah[7].

Setiap bahasa memiliki karakteristik unik berdasarkan distribusi probabilitas simbol yang ada. Oleh karena itu, penting untuk mengevaluasi kinerja kompresi pada berbagai bahasa guna melihat potensinya. Di antara bahasa daerah yang berpotensi untuk dikaji adalah Bahasa Minangkabau dan Bahasa Minang, karena kedua bahasa ini digunakan di daerah pariwisata dan memiliki jumlah penutur yang cukup signifikan[8].

Dalam penelitian ini dilakukan kompresi Bahasa Minangkabau dan Bahasa Melayu menggunakan *ternary Huffman codes* untuk mengeksplorasi potensinya dalam mendukung komunikasi 5G dan seterusnya..

1.2 Tinjauan Penelitian Terdahulu

Tinjauan penelitian terdahulu adalah pernyataan yang menunjukkan bahwa penelitian yang akan dijalankan memiliki pemecahan masalah dengan menggunakan penelitian yang sudah dijalankan oleh peneliti lain. Bagian ini akan diuraikan secara ringkas penelitian sebelumnya yang relevan dengan penelitian yang akan dijalankan, sehingga dapat memberikan masukan terhadap penelitian dan dapat memperkuat alasan di balik pelaksanaan penelitian ini. Adapun Tinjauan penelitian terdahulu penelitian lainnya dijabarkan pada Tabel 2.1

Tabel 2. 1 Kajian riset terdahulu.

Judul Penelitian	Tahun	Nama Peneliti
Penerapan Algoritma Huffman <i>Code</i> Untuk Kompresi Dan Dekompresi Pesan [9].	2024	Akbar Serdano, Chairunnisa Azzahra, Darus Alpamah, Ilham Fajar Batubara, Mohd. Wildan, Qasthari,

Judul Penelitian	Tahun	Nama Peneliti
Implementasi Kode Huffman Dalam Aplikasi Kompresi Text Pada Layanan SMS [10].	2023	Yuza Reswan, dan Dedy Agung Prabowo
<i>An entropy analysis of the Cirebon language script using the Ternary Huffman code algorithm</i> [5].	2021	Kodir. A, Fajar. R, Awalluddin. Ruswandi. U, Ismail. N, Miharja,
<i>Feasibility Study of Using Huffman Code Calculator in Learning Achievment of Data Compression</i> [11].	2021	Zulhelman, Mohammad Fathurahman, and Endang Saepudin
<i>Performance Comparison Of Huffman Coding And Lempel-ZIV-Welch Text Compression Algorithms With Chinese Remainder Theorem</i> [12].	2019	Mohammed Babatunde Ibrahim, and Kazeem Alagbe Gblolagade

Penelitian terdahulu mengenai kompresi telah dilaksanakan oleh berbagai Lembaga dengan pendekatan, tujuan, dan hasil pencapaiannya yang berbeda. Dapat dilihat pada Tabel 1.1 mengenai penelitian-penelitian terkait dengan kebutuhan riset yang sedang dilakukan.

Paper [8] membahas tentang studi mengenai kompresi data menggunakan algoritma *Huffman Code*, yang memungkinkan pengurangan ukuran *file* atau pesan tanpa kehilangan informasi penting di dalamnya. Fokus utama penelitian ini adalah penerapan algoritma *Huffman Code* untuk mengompresi teks pesan singkat, dengan tujuan menghemat sekitar 70% dari ukuran pesan. Selain itu, jurnal ini juga menekankan bahwa *Huffman Code* adalah algoritma kompresi *lossless*, yang berarti data yang didekompresi akan identik dengan data asli sebelum dikompres. Penelitian ini mengimplementasikan algoritma Huffman dalam program untuk menguji data teks pesan singkat[9].

Paper [9] membahas penerapan kompresi pada SMS, yang mencakup serangkaian proses mulai dari tahap kompresi saat pengiriman, penerimaan, hingga tampilan pesan. Aplikasi ini dikembangkan untuk mengatasi kendala pengiriman pesan singkat dengan batas kapasitas maksimum 140 *byte* per pesan. Tujuan utama pembuatan aplikasi ini adalah untuk mempermudah pengiriman pesan dengan jumlah karakter yang banyak serta karakter dengan bit besar. Dalam penelitian ini, Huffman *coding* digunakan untuk mengompresi pesan SMS, dengan tingkat kompresi mencapai 52,8% dalam kondisi optimal, dan dapat mencapai lebih dari 24,5% dengan mudah dalam kondisi standar [10].

Paper [5] membahas tentang kompresi bahasa-bahasa lokal Indonesia untuk memperoleh *outage probability* dengan menggunakan algoritma *ternary Huffman codes*. Nilai *entropy* ini kemudian menjadi referensi untuk tingkat kompresi aksara Cirebon. Probabilitas setiap simbol dalam Teks Daerah Cirebon digunakan untuk menghitung nilai *entropy*. Perbandingan *outage probability* pada penelitian ini dilakukan dengan aksara bahasa Cirebon sebagai representasi bahasa lokal untuk mengevaluasi efisiensi kompresi, hasil penelitian ini menunjukkan bahwa *entropy* aksara bahasa Cirebon adalah 2,508 bit per simbol, dengan panjang kode yang diharapkan sebesar 2,565 bit per simbol. Estimasi efisiensi kompresi dengan *ternary Huffman codes* adalah 97,77%, dan tingkat kompresinya adalah 0,51308[5].

Paper [10] membahas aplikasi kalkulator yang digunakan untuk menentukan kode setiap karakter dalam suatu teks. Kode yang diberikan untuk setiap karakter bersifat dinamis, bergantung pada frekuensi kemunculannya dalam teks tersebut. Semakin sering suatu karakter muncul, maka semakin pendek kode yang diberikan. Penelitian ini menunjukkan bahwa kalkulator Huffman dapat digunakan sebagai alat bantu untuk *learning outcomes*, dengan melakukan kompresi data menggunakan dua bahasa, yaitu Bahasa Indonesia dan Bahasa Inggris, pada teks dengan panjang antara 100 hingga 1000 karakter. Hasil penelitian menunjukkan bahwa nilai rata-rata *entropy* untuk Bahasa Inggris adalah 4,35, sedangkan untuk Bahasa Indonesia adalah 4,42. Untuk nilai *compression rate*, Bahasa Inggris mencapai 1,61, dan Bahasa Indonesia sebesar 1,59. Kesimpulan dari penelitian ini

adalah kalkulator Huffman dapat digunakan untuk kompresi data secara *online* pada teks dalam Bahasa Inggris dan Bahasa Indonesia, dengan hasil yang hampir serupa. Hal ini menunjukkan bahwa perbedaan bahasa hanya mempengaruhi frekuensi kemunculan karakter dalam teks [11].

Paper [11] membahas dua algoritma kompresi data, yaitu Huffman *codes* dan LZW, serta membandingkan kinerjanya berdasarkan beberapa sampel teks yang diberikan, dengan menambahkan Teorema Sisa Cina (CRT). Hasil penelitian ini menunjukkan bahwa Huffman-CRT menghasilkan kompresi yang lebih unggul dibandingkan LZW-CRT, dengan ukuran kompresi berkisar antara 55,64% hingga 57,11%. Waktu kompresi untuk Huffman-CRT adalah 44,61 detik, lebih cepat daripada LZW-CRT yang memerlukan waktu 55,39 detik. Selain itu, dalam hal tingkat kompresi (*compression rate*), Huffman-CRT mengungguli LZW-CRT dengan perbedaan sebesar 2,37% hingga 3,70% [12].

Berdasarkan *tabel state of the art* yang telah disajikan, banyak penelitian sebelumnya yang telah membahas topik terkait kompresi teks. Namun, pada penelitian tugas akhir ini, fokus utama adalah mengeksplorasi tingkat kompresi pada bahasa yang belum dijadikan standar, melalui penerapan teknik kompresi Huffman *codes*. Data sampel yang digunakan dalam penelitian ini mencakup kumpulan karakter, tanda baca, angka, serta teks dalam Bahasa Minangkabau latin dan Bahasa Melayu latin. Hasil kompresi yang diperoleh dibandingkan dengan panjang kode ASCII. Tujuan utama dari penelitian ini adalah untuk memperoleh struktur Huffman *tree* untuk Bahasa Minangkabau dan Bahasa Melayu, serta menguji kebenaran teori baru *outage probability* yang digunakan untuk mengukur tingkat kegagalan atau kesalahan pada suatu kejadian. Penelitian ini juga bertujuan untuk mengeksplorasi apakah bahasa lokal memiliki potensi kompresi yang lebih baik dibandingkan dengan bahasa standar, yang dapat dilihat dari nilai *entropy*-nya.

1.3 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini yaitu sebagai berikut.

1. Bagaimana nilai kompresi teks Bahasa Minangkabau dan Bahasa Melayu berbasis *ternary Huffman codes*?
2. Bagaimana kinerja komunikasi Bahasa Minangkabau dan Bahasa Melayu yang dikompres dengan *ternary Huffman codes* saat diuji secara teoritis menggunakan *outage probability*?

1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisis nilai kompresi data teks Bahasa Minangkabau dan Bahasa Melayu berbasis *ternary Huffman codes*.
2. Menganalisis kinerja kompresi dalam konteks komunikasi Bahasa Minangkabau dan Bahasa Melayu yang dikompres dengan *ternary Huffman codes*.

1.5 Manfaat

Penelitian yang dilakukan ini diharapkan dapat memperoleh manfaat dari sisi akademis dan juga dari sisi praktis.

1. Manfaat Akademis

Penelitian ini akan memberikan kontribusi bagi pengembangan pengolahan bahasa alami di Indonesia, khususnya dalam konteks bahasa daerah. Dengan menggunakan metode pengkodean *ternary*, diharapkan dapat tercipta model pengolahan teks yang lebih efisien dan tepat dalam memproses berbagai bahasa daerah yang ada di Indonesia, yang sangat beragam dan memiliki kompleksitas tersendiri.

2. Manfaat Praktis

Manfaat praktis dari penelitian ini diharapkan dapat membuat *codeword* baru untuk Bahasa Minangkabau dan Melayu juga diharapkan dapat memberikan kontribusi terhadap perkembangan di bidang telekomunikasi, khususnya pada aplikasi 5G.

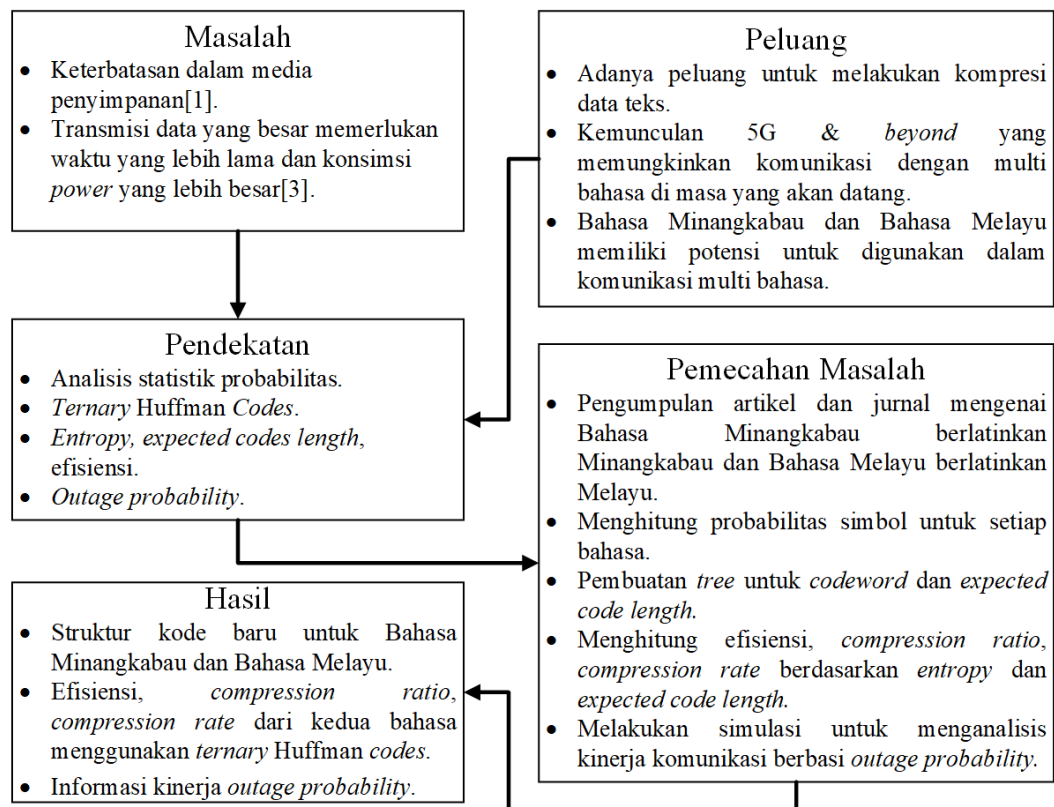
1.6 Batasan Masalah

Penelitian ini diharapkan memiliki fokus penelitian yang jelas, oleh karena itu, diperlukan penetapan batasan masalah untuk menghindari perluasan topik. Berikut adalah batasan-batasan masalah yang diterapkan dalam penelitian ini :

1. Data yang digunakan dalam penelitian ini berupa teks yang terdiri dari kombinasi huruf, angka, dan karakter minimal 7000 simbol.
2. Teks bahasa yang digunakan dalam penelitian ini adalah Bahasa Minangkabau dan Bahasa Melayu yang ditulis dengan huruf Latin dalam bentuk cerita pendek yang didapatkan dari internet.
3. Proses kompresi dilakukan dengan teknik *ternary Huffman codes*.
4. Parameter yang ingin diketahui dan diteliti adalah *entropy*, *expected length*, *efisiensi*, *compression ratio* dan *compression rate*.
5. Penelitian diuji secara teoritis dengan menggunakan *outage probability*.
6. Simulasi dan pengujian data dilakukan melalui penerapan *software* komputasi.
7. Simulasi pengujian menggunakan 2 *user* yang berkomunikasi.
8. Analisis yang dilakukan mencakup perbandingan dengan *binary Huffman codes*.

1.7 Kerangka Berpikir

Kerangka berpikir berisi alur pemikiran yang memuat uraian sistematis tentang hasil perumusan masalah penelitian yang diperkirakan dapat diselesaikan melalui pendekatan dan melihat peluang pemecahan masalah tersebut. Pemecahan masalah diurutkan secara sistematis untuk melakukan kompresi teks pada Bahasa Minangkabau dan Bahasa Melayu. Kerangka berpikir penelitian ini ditunjukkan pada Gambar 1.1.



Gambar 1. 1 Kerangka berpikir.

1.8 Sistematika Penulisan

Sistematika penulisan adalah pedoman penyusunan data dan penulisan yang bertujuan menghasilkan penulisan yang baik. Sistematika penulisan laporan penelitian terdiri dari:

BAB I PENDAHULUAN

Bab ini merupakan permulaan dari penyusunan laporan penelitian. Pada bagian ini terkandung hal-hal pokok dari awal sebuah tulisan, yaitu: latar belakang, kajian terdahulu, rumusan masalah, tujuan, manfaat, batasan masalah, kerangka berpikir, dan sistematika penulisan.

BAB II TEORI DASAR

Bab ini berisi tinjauan Pustaka yang sangat berkaitan dengan kegiatan penelitian, terutama dalam memahami mengenai tingkat kompresi teks dari bahasa lokal Indonesia, yaitu tingkat kompresi Bahasa Minangkabau dan Bahasa Melayu bertuliskan latin dengan menggunakan metode *Huffman Codes*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi diagram alur tahapan yang akan dilakukan dalam penelitian. Bab ini juga berisikan mengenai perincian jadwal perencanaan untuk melaksanakan penelitian mengenai kompresi Bahasa Minangkabau dan Bahasa Melayu bertuliskan latin dengan menggunakan metode Huffman *Codes*, dimulai dari studi literatur hingga penyelesaian penulisan tugas akhir.

BAB IV ANALISIS STATISTIK BAHASA MINANGKABAU DAN BAHASA MELAYU

Bab ini membahas hasil analisis statistik terhadap Bahasa Minangkabau dan Bahasa Melayu. Pembahasan mencakup penjelasan mengenai objek penelitian, proses pengumpulan dan pengolahan data, serta tahap konstruksi *ternary Huffman tree* dan penyusunan *codeword* untuk masing-masing bahasa.

BAB V ANALISIS KINERJA BAHASA MINANGKABAU DAN MELAYU UNTUK KOMUNIKASI

Bab ini menyajikan analisis terkait kinerja Bahasa Minangkabau dan Bahasa Melayu dalam konteks sistem komunikasi. Analisis meliputi perhitungan *expected code length*, efisiensi, *compression rate*, serta evaluasi menyeluruh terhadap hasil penelitian. Selain itu, dilakukan juga pengujian kinerja sistem menggunakan metode *outage probability* untuk menilai pengaruh interferensi dalam skenario komunikasi.

BAB VI PENUTUP

Bab ini memuat kesimpulan yang diperoleh dari seluruh rangkaian penelitian, dikaitkan dengan tujuan awal yang telah ditetapkan. Selain itu, bab ini juga menyampaikan saran dan rekomendasi untuk pengembangan serta arah penelitian selanjutnya.