

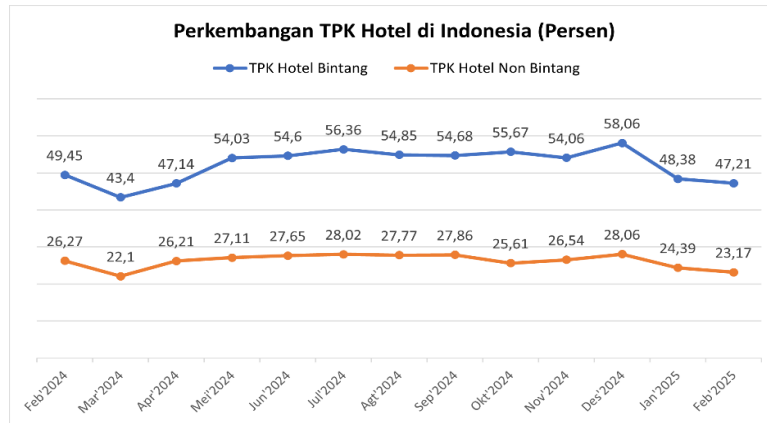
BAB I

PENDAHULUAN

1.1 Latar Belakang

Pariwisata merupakan salah satu sektor penting dalam pembangunan ekonomi, baik di tingkat global maupun nasional. Sektor ini tidak hanya memberikan kontribusi terhadap peningkatan devisa negara, tetapi juga menciptakan lapangan pekerjaan, mendorong investasi, dan memperkuat citra suatu daerah [1]. Di Indonesia, sektor ini menjadi salah satu prioritas pembangunan dengan kontribusi terhadap Produk Domestik Bruto (PDB) nasional yang mencapai sekitar 4% pada tahun 2024, meningkat dibandingkan tahun sebelumnya yang sebesar 3,9% [2]. Peningkatan ini sejalan dengan kenaikan jumlah kunjungan wisatawan mancanegara, di mana Badan Pusat Statistik (BPS) mencatat total 13,9 juta kunjungan sepanjang tahun 2024 atau tumbuh sekitar 19,05% dibanding tahun 2023 [3].

Pertumbuhan pesat sektor pariwisata turut mendorong aktivitas ekonomi di berbagai sektor pendukung, termasuk industri perhotelan yang berperan penting sebagai penyedia akomodasi bagi wisatawan, pelaku bisnis, maupun mobilitas masyarakat antarwilayah [4], [5]. Pertumbuhan industri perhotelan dibuktikan oleh data Kementerian Pariwisata dan Ekonomi Kreatif di tahun 2024 yang mencatat bahwa jumlah hotel berbintang di Indonesia mencapai 4.584 unit, meningkat 11,02% dibandingkan tahun 2023 [6]. Selain itu, data BPS menunjukkan bahwa total usaha jasa akomodasi di Indonesia pada tahun 2024 mencapai 31.175 unit, atau naik 7,84% dari tahun sebelumnya [7]. Pertumbuhan tersebut banyak terkonsentrasi di provinsi dengan aktivitas ekonomi, pariwisata, dan bisnis yang tinggi, seperti Bali (4.154 hotel), Jawa Timur (4.055 hotel), dan Jawa Barat (3.125 hotel) yang juga tercatat sebagai wilayah dengan jumlah hotel terbanyak menurut data BPS tahun 2024 [7], [8]. Peningkatan ini mencerminkan besarnya potensi industri perhotelan di Indonesia yang terus tumbuh seiring meningkatnya mobilitas domestik maupun mancanegara. Meski demikian, pertumbuhan jumlah hotel yang pesat tersebut tidak selalu berbanding lurus dengan peningkatan tingkat hunian.



Gambar 1. 1 Grafik Perkembangan TPK Hotel di Indonesia [9]

Meskipun jumlah hotel di Indonesia terus meningkat, tingkat okupansi justru belum sepenuhnya menunjukkan tren positif. Berdasarkan data Badan Pusat Statistik, Tingkat Penghunian Kamar (TPK) hotel di Indonesia mengalami penurunan yang cukup signifikan pada awal tahun 2025. Grafik pada Gambar 1.1 menunjukkan bahwa TPK hotel berbintang turun tajam dari 58,06% pada Desember 2024 menjadi 48,38% pada Januari 2025, dan kembali menurun menjadi 47,21% pada Februari 2025 [9]. Secara tahunan (*year-on-year*), angka tersebut juga tercatat turun 2,24 poin dibandingkan Februari 2024 yang mencapai 49,45%. Tren penurunan ini terjadi hampir di seluruh provinsi, termasuk wilayah dengan jumlah hotel terbanyak seperti Bali, Jawa Timur, dan Jawa Barat [10], [11], [12]. Tren penurunan beruntun ini mengindikasikan adanya tantangan bagi industri perhotelan, di mana pertumbuhan jumlah hotel tidak diiringi dengan peningkatan tingkat hunian. Kondisi tersebut memperlihatkan bahwa persaingan antarhotel semakin ketat dan keberhasilan pengelola hotel kini sangat bergantung pada kemampuan memahami kepuasan serta persepsi pelanggan terhadap kualitas layanan [13], [14]. Oleh karena itu, manajemen hotel perlu memahami persepsi dan pengalaman pengunjung secara lebih cepat dan akurat agar dapat mempertahankan daya saing di tengah dinamika pasar perhotelan.

Kebutuhan mendesak untuk memahami persepsi pengunjung ini bertepatan dengan pesatnya perkembangan teknologi digital. Perkembangan internet, dan media sosial telah mengubah cara pengunjung dalam menyampaikan keluhan dan apresiasi terhadap pengalaman menginap mereka dengan menuliskan ulasan di platform online. Hal ini difasilitasi oleh platform review seperti Tripadvisor yang

menjadi repository besar bagi *User-Generated Content* (UGC) berupa ulasan dan rating [15], [16]. Berkaitan dengan fenomena ini, penelitian [17], [18] menemukan bahwa ulasan daring memiliki pengaruh positif dan signifikan terhadap minat pemesanan kamar hotel. Dengan demikian, ulasan daring kini berperan sebagai “*Electronic Word of Mouth* (e-WOM)” yang dapat menentukan reputasi, persepsi kualitas layanan, dan daya tarik suatu hotel [19], [20]. Selain bagi calon pengunjung, ulasan dari platform daring juga menjadi sumber data penting bagi manajemen hotel untuk mengevaluasi kualitas layanan.



Gambar 1. 2 Rating Ulasan Hotel Oval di Platform Tripadvisor [21]

Meskipun ulasan daring telah diakui sebagai bentuk *Electronic Word of Mouth* yang berpengaruh besar terhadap keputusan pemesanan hotel, pemanfaatannya oleh manajemen hotel masih menghadapi berbagai tantangan. Di tengah persaingan ketat, upaya untuk memahami pengalaman pengunjung secara cepat sering terkendala oleh karakteristik Big Data dari *User-Generated Content* (UGC) [22]. Saat ini, beberapa platform daring menyediakan rating numerik agregat untuk mempermudah pembacaan ulasan. Namun, informasi ini hanya memberikan gambaran umum mengenai tingkat kepuasan pengunjung tanpa menjelaskan aspek layanan spesifik yang memengaruhi persepsi pelanggan. Sebagai contoh, pada Gambar 1.2 ditampilkan rating ulasan Hotel Oval Surabaya di platform Tripadvisor, yang memiliki skor rata-rata sebesar 2.8 dengan kategori “Biasa”. Informasi tersebut memang bermanfaat untuk melihat tren persepsi secara umum, namun belum cukup untuk menjelaskan aspek layanan spesifik yang perlu ditingkatkan atau dipertahankan.

Tantangan lainnya adalah platform review hotel seperti Tripadvisor menghasilkan ribuan ulasan yang akan terus bertambah setiap harinya. Volume data yang masif ini merupakan karakteristik utama dari Big Data [23]. Bagi pihak manajemen hotel, mengandalkan metode manual dengan membaca dan merangkum ulasan satu per satu tentunya tidak efisien, memakan waktu, dan tidak mampu

memberikan wawasan secara *real-time*. Padahal, menurut penelitian Mariani dan Baggio, keputusan strategis dan perbaikan layanan harus dilakukan dengan cepat agar hotel mampu merespons dinamika pasar dan menghindari risiko kerugian, sebagaimana tercermin pada kinerja TPK hotel yang menurun drastis pada Januari 2025 [23]. Tantangan selanjutnya yang dihadapi manajemen hotel adalah sifat data ulasan yang kompleks dan tidak terstruktur. Ulasan pengunjung di berbagai platform daring sebagian besar berbentuk teks bebas (narasi) yang mencakup berbagai hal seputar layanan dan kualitas hotel [24]. Data tekstual pada ulasan di platform daring tidak memiliki format standar sehingga sulit untuk dikuantifikasi. Selain itu, ulasan daring seringkali menggunakan berbagai bahasa. Hal ini dikarenakan Indonesia merupakan negara dengan jumlah kunjungan wisatawan mancanegara yang cukup tinggi, bahkan mencapai 13,9 juta kunjungan pada tahun 2024 [3]. Variasi bahasa ini menambah tantangan bagi manajemen hotel dalam memahami opini pengunjung secara menyeluruh, sebab tidak semua ulasan dapat dibaca atau dipahami tanpa bantuan alat penerjemahan. Akibatnya, pihak manajemen hotel kesulitan mengidentifikasi dan memperoleh wawasan secara akurat mengenai aspek layanan yang paling sering dibahas dan disorot oleh pelanggan. Kesenjangan antara pentingnya *User-Generated Content* dan kesulitan dalam mengolahnya menjadi wawasan bermakna merupakan titik kritis manajerial yang menuntut penerapan pendekatan analisis teks secara sistematis dan objektif.

Untuk mengatasi kendala pengolahan data tekstual menjadi informasi yang dapat ditindaklanjuti, teknik *sentiment analysis* banyak dijadikan sebagai solusi ilmiah. *Sentiment analysis* adalah teknik komputasional untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini atau sentimen dari teks secara otomatis [25]. Dalam konteks ulasan hotel, *sentiment analysis* dapat membantu mengklasifikasikan ulasan pengunjung ke dalam kategori sentimen positif, negatif, atau netral, sehingga membantu manajemen untuk memahami persepsi pengunjung dalam skala besar tanpa harus membaca setiap ulasan secara manual [25].

Namun, pendekatan *sentiment analysis* konvensional memiliki keterbatasan karena hanya mengklasifikasikan sentimen secara keseluruhan tanpa mempertimbangkan aspek spesifik yang dibahas dalam ulasan. Untuk mengatasi keterbatasan tersebut, *Aspect-Based Sentiment Analysis* (ABSA) hadir sebagai

pendekatan yang lebih mendalam dengan menentukan aspek yang menjadi objek opini dalam suatu ulasan [26], [27]. Berdasarkan beberapa penelitian terdahulu, aspek-aspek yang sering dievaluasi oleh pengunjung dan menjadi nilai utama suatu hotel adalah *Room*, *Hotel*, *Location*, dan *Service* [28], [29], [30], [31]. Dalam penelitian ini aspek yang digunakan adalah *Room*, *Hotel*, *Location*, dan *Service* yang memiliki tingkat generalisasi tinggi serta merepresentasikan *core services* yang secara langsung mencerminkan kualitas pengalaman menginap pengunjung hotel.



Gambar 1. 3 Ulasan Pengunjung pada Platform Tripadvisor [21]

Meski demikian, sebagian besar penelitian ABSA tradisional masih bersifat *single-label*, yaitu hanya mampu mengenali satu aspek dan satu sentimen pada setiap ulasan. Padahal, satu ulasan seringkali memuat opini terhadap beberapa aspek dengan polaritas yang berbeda. Sebagai contoh, Gambar 1.3 menampilkan ulasan pengunjung hotel yang menilai pelayanan hotel secara positif karena staf ramah dan sigap, namun memberikan penilaian negatif terhadap kebersihan dan fasilitas kamar yang kurang terawat. Pola ulasan seperti ini memperlihatkan bahwa satu teks ulasan dapat memuat lebih dari satu opini dengan polaritas yang berbeda, sehingga ABSA tradisional *single label* yang hanya menilai sentimen dari satu aspek secara umum menjadi kurang memadai. Oleh karena itu, *Multilabel Aspect-Based Sentiment Analysis* menjadi perluasan dari ABSA konvensional. Pendekatan ini memungkinkan sistem untuk mendeteksi lebih dari satu aspek serta mengklasifikasikan sentimen untuk masing-masing aspek yang muncul dalam satu ulasan [27]. Pendekatan ini menghasilkan informasi yang lebih kaya dan terstruktur, sehingga manajemen hotel dapat mengidentifikasi aspek yang mendapat

apresiasi maupun yang perlu ditingkatkan, dan merumuskan strategi perbaikan layanan secara lebih tepat sasaran.

Seiring perkembangan teknologi NLP (*Natural Language Processing*), berbagai pendekatan model telah dikembangkan untuk ABSA, mulai dari *machine learning* klasik hingga *deep learning*. *Deep learning* merupakan cabang dari *machine learning* yang berbasis *artificial neural networks* dengan beberapa lapisan pemrosesan, yang mampu mempelajari pola dan representasi dari data secara otomatis [32]. Pendekatan ini unggul karena kemampuannya dalam mengekstraksi fitur secara otomatis, memahami hubungan kompleks antar kata dalam teks, serta menangani data yang besar dan tidak terstruktur sehingga sangat sesuai untuk menganalisis ulasan hotel yang kompleks dan bervariasi [26], [32]. Agar model dapat memahami konteks ulasan dengan baik, diperlukan representasi teks yang lebih bermakna dan kontekstual dibandingkan sekedar berbasis *bag-of-words*. Di sinilah peran *word embedding* sebagai teknik yang mampu memetakan kata ke dalam ruang vektor berdimensi berdasarkan kedekatan semantik antar kata [33], [34]. *Word embedding* memungkinkan model memahami hubungan seperti sinonim, analogi, dan konteks penggunaan kata. Beberapa algoritma *deep learning* yang populer dalam penelitian ABSA adalah CNN, LSTM, BiLSTM, dan BERT yang umumnya dikombinasikan dengan *word embedding* statis seperti Word2Vec dan FastText maupun *contextual word embedding* untuk meningkatkan pemahaman konteks bahasa [26], [32], [35].

Sejumlah penelitian telah membuktikan efektivitas penerapan ABSA menggunakan pendekatan *deep learning* dan *word embedding*. Salah satunya adalah penelitian [36] yang menganalisis sentimen berbasis aspek pada ulasan wisata halal berbahasa Inggris menggunakan model CNN dan LSTM dengan *word embedding*. Aspek yang diteliti meliputi makanan halal, masjid, dan toilet. Hasil penelitian menunjukkan model CNN mencapai akurasi tertinggi sebesar 98,3% pada klasifikasi aspek dan 93,9% pada klasifikasi sentimen. Penelitian lain [37] menerapkan model *deep learning* LSTM dan BiLSTM dengan Word2Vec embedding untuk ABSA pada ulasan destinasi wisata di Indonesia. Aspek yang diteliti meliputi *attraction*, *accessibility*, *facility*, dan *accommodation*. Hasil penelitian menunjukkan model BiLSTM mencapai akurasi 92,2% [37]. Meskipun

penelitian-penelitian tersebut menunjukkan performa yang menjanjikan, hingga kini belum banyak penelitian yang secara komprehensif membandingkan berbagai model *deep learning* dan *word embedding* pada konteks ulasan hotel. Selain itu, kajian mengenai pengaruh *stemming* juga masih terbatas dan menunjukkan hasil yang belum konsisten. Beberapa penelitian menyatakan bahwa *stemming* dapat meningkatkan performa model dengan menyederhanakan variasi kata ke bentuk dasar [38] namun penelitian lain menunjukkan bahwa *stemming* justru dapat menurunkan performa karena berpotensi menghilangkan informasi kontekstual [39]. Oleh karena itu, pengaruh *stemming* perlu dikaji lebih lanjut, khususnya pada penerapan ABSA di domain perhotelan.

Berdasarkan urgensi yang telah diuraikan, meliputi penurunan kinerja sektor perhotelan Indonesia, meningkatnya persaingan industri perhotelan, serta keterbatasan manajemen dalam mengolah ulasan pengunjung yang masif, penelitian ini bertujuan untuk mengembangkan sistem *Aspect-Based Sentiment Analysis* pada ulasan hotel. Penelitian ini juga bertujuan untuk melakukan komparasi terhadap performa beberapa model *deep learning* dan *word embedding* dalam tugas *Multilabel Aspect-Based Sentiment Analysis* pada ulasan hotel, serta mengetahui pengaruh proses *stemming* pada model. Penelitian ini difokuskan pada empat aspek utama hotel yang sering menjadi perhatian pengunjung, yaitu *Room*, *Hotel*, *Location*, dan *Service* [28], [29], [30], [31]. Hasil penelitian ini diharapkan dapat memberikan rekomendasi berbasis bukti empiris mengenai kombinasi model *deep learning* dan *word embedding* yang memberikan performa optimal dalam menganalisis ulasan hotel, serta menghasilkan sistem berbasis web yang mampu membantu mengonversi ulasan hotel berupa teks menjadi *insight* guna meningkatkan kualitas layanan dan daya saing di tengah tantangan industri perhotelan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan pada bagian sebelumnya, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana sentimen pengunjung terhadap aspek perhotelan (*Hotel, Room, Location, dan Service*) berdasarkan ulasan beberapa hotel di Surabaya, Sukabumi, dan Bali pada platform review Tripadvisor?
2. Bagaimana kinerja dan hasil komparasi berbagai kombinasi arsitektur *deep learning* (CNN dan BiLSTM), *word embedding* (Word2Vec dan FastText), serta teknik *preprocessing* (dengan *stemming* dan tanpa *stemming*) pada setiap model dalam *pipeline Aspect-Based Sentiment Analysis* untuk ulasan hotel?
3. Kombinasi arsitektur *deep learning, word embedding, dan teknik preprocessing* manakah yang menghasilkan performa terbaik pada setiap model dalam *pipeline Aspect-Based Sentiment Analysis* untuk ulasan hotel?
4. Bagaimana hasil implementasi sistem berbasis web untuk analisis sentimen berbasis aspek pada ulasan hotel?

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah disusun, maka batasan masalah dalam penelitian ini ditetapkan agar penelitian lebih fokus dan terarah:

1. Data yang digunakan berupa ulasan hotel dari platform review digital Tripadvisor. Data ulasan diperoleh dari beberapa hotel yang berlokasi di Bali, Surabaya, dan Sukabumi. Ketiga daerah tersebut dipilih sebagai representasi provinsi dengan jumlah hotel terbanyak di Indonesia, karena memiliki Tingkat Penghunian Kamar tertinggi di setiap provinsi tersebut.
2. Periode data dibatasi pada ulasan yang dipublikasikan mulai tanggal 1 September 2023 hingga 30 November 2025, dan mencakup ulasan berbahasa Indonesia maupun Inggris.
3. Data ulasan hotel yang digunakan dalam penelitian ini dibatasi pada ulasan berbahasa Indonesia dan Bahasa Inggris. Ulasan berbahasa Inggris diterjemahkan ke dalam Bahasa Indonesia pada tahap translasi untuk

menjaga konsistensi representasi bahasa pada proses *preprocessing* dan pemodelan.

4. Klasifikasi sentimen dalam penelitian ini dibatasi pada dua kategori, yaitu positif dan negatif. Sedangkan aspek yang dibahas dalam penelitian ini adalah *Hotel, Room, Location, dan Service*.
5. Model *deep learning* yang dikomparasikan dalam penelitian ini terbatas pada model CNN dan BiLSTM. Sedangkan teknik *word embedding* yang dikomparasikan dalam penelitian ini adalah Word2Vec dan FastText.
6. Implementasi hasil penelitian dilakukan dalam bentuk sistem berbasis web menggunakan *framework* Flask, yang memungkinkan pengguna untuk menginputkan ulasan tunggal (*single review*) maupun file CSV/Excel untuk mendapatkan hasil klasifikasi aspek dan sentimen.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis sentimen ulasan pengunjung terhadap aspek-aspek utama layanan hotel, yaitu *Room, Hotel, Location, dan Service*, berdasarkan ulasan beberapa hotel yang diperoleh dari platform Tripadvisor.
2. Menganalisis kinerja dan membandingkan berbagai kombinasi arsitektur *deep learning* (CNN dan BiLSTM), *word embedding* (Word2Vec dan FastText), serta teknik *preprocessing* (dengan *stemming* dan tanpa *stemming*) pada setiap model dalam pipeline ABSA.
3. Menentukan kombinasi terbaik dari arsitektur *deep learning, word embedding, dan teknik preprocessing* berdasarkan hasil evaluasi performa model pada masing-masing model dalam pipeline ABSA.
4. Mengembangkan sistem berbasis web yang dapat digunakan untuk menganalisis sentimen ulasan pengunjung hotel berdasarkan aspek (*Room, Hotel, Location, Service*).

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah dipaparkan sebelumnya, penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi pada pengembangan kajian *Aspect-Based Sentiment Analysis* (ABSA) dengan membandingkan kinerja berbagai model *deep learning* dan *word embedding* pada domain ulasan hotel.
2. Menjadi referensi akademis bagi penelitian selanjutnya yang mengkaji analisis sentimen berbasis aspek pada ulasan pelanggan, baik di domain perhotelan maupun domain lainnya.
3. Menyediakan sistem berbasis web sederhana yang dapat digunakan untuk menganalisis sentimen ulasan pengunjung hotel berdasarkan aspek layanan hotel (*Room, Hotel, Location, Service*).

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk mempermudah proses penyusunan laporan agar selaras dengan pedoman yang berlaku serta mendukung tercapainya tujuan penelitian. Sistematika dalam penyusunan skripsi ini meliputi:

BAB I PENDAHULUAN

Bab ini menjelaskan gambaran umum tentang penelitian yang akan dilakukan, termasuk membahas latar belakang penelitian, rumusan masalah yang akan dipecahkan, batasan masalah yang ditetapkan, tujuan penelitian, manfaat penelitian serta sistematika penulisan secara keseluruhan. Tujuan dari bab ini adalah mengetahui gambaran awal mengenai fokus dan ruang lingkup penelitian

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan hasil tinjauan terhadap penelitian-penelitian terdahulu yang berkaitan dengan penelitian untuk memperkuat landasan metodologi dan pendekatan yang digunakan. Selain itu, bab ini juga menyajikan teori-teori yang relevan sebagai dasar untuk mendukung penelitian, khususnya mengenai ABSA, *Word*

embedding (Word2Vec dan FastText), arsitektur *Deep learning* (CNN dan BiLSTM), dan metrik evaluasi model.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan penjelasan secara rinci tahapan yang digunakan dalam penelitian, dimulai dari pengumpulan data, pelabelan, *preprocessing* text, pembagian dataset, implementasi *word embedding* dan model *deep learning* dengan pendekatan pipeline ABSA, evaluasi dan komparasi performa model dalam ABSA, hingga proses deployment model dan pengembangan website berbasis flask.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari setiap tahapan penelitian secara terperinci, termasuk analisis data ulasan hotel yang dikumpulkan, hasil implementasi model *deep learning* dengan *word embedding*, hasil training dan evaluasi model aspect detection dan sentiment analysis, komparasi performa model berdasarkan metrik akurasi dan efisiensi komputasi, serta implementasi *prototype* sistem berbasis web.

BAB V PENUTUP

Bab terakhir berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran-saran untuk penelitian selanjutnya. Kesimpulan disusun berdasarkan pencapaian tujuan penelitian dan hasil komparasi model, sementara saran disampaikan untuk pengembangan dan penyempurnaan di masa mendatang.

DAFTAR PUSTAKA

Daftar pustaka menyajikan daftar seluruh referensi dan sumber literatur yang digunakan dalam penelitian ini.

LAMPIRAN

Bagian ini berisikan lampiran berupa dokumentasi dalam penelitian.