BAB I
PENDAHULUAN

1.1 Latar Belakang

Transformasi digital pada sektor layanan kesehatan publik Indonesia terus mengalami percepatan melalui pengembangan aplikasi Mobile JKN oleh BPJS Kesehatan sebagai sarana utama akses layanan bagi peserta. Aplikasi ini memungkinkan pengguna melakukan pendaftaran layanan, pengecekan kepesertaan, hingga pengajuan administrasi secara *real-time*. Hingga Februari 2026, aplikasi Mobile JKN telah menerima lebih dari 954.000 ulasan dengan *rating* rata-rata 4,5 bintang pada *Google Play Store* (diakses 18 Februari 2026). *Rating* tersebut merepresentasikan keseluruhan populasi ulasan secara kumulatif, namun tidak selalu mencerminkan distribusi sentimen pada subset ulasan yang dikumpulkan berdasarkan parameter relevansi. Jumlah ulasan yang masif ini mencerminkan pengalaman nyata di lapangan dan menjadi sumber data tekstual yang kaya untuk evaluasi kualitas layanan. Namun, seiring meningkatnya volume data, analisis manual menjadi tidak efisien dan rentan terhadap subjektivitas. Oleh karena itu, diperlukan pendekatan analisis sentimen otomatis berbasis *deep learning* guna mengekstraksi informasi mendalam dari opini pengguna yang heterogen—termasuk pola keluhan berulang seperti kendala login, kode OTP, dan gangguan sistem—sebagai indikator strategis dalam pengambilan keputusan pengembangan aplikasi (Al Qahar et al., 2024; Putra, 2024). Implementasi ini krusial mengingat data ulasan bersifat dinamis dan terus bertambah seiring pembaruan aplikasi.

Namun, efektivitas model deep learning untuk analisis sentimen tidak semata-mata ditentukan oleh arsitektur yang dipilih, melainkan juga oleh distribusi kelas dalam data pelatihan. Pada domain ulasan aplikasi layanan publik, kondisi ketidakseimbangan distribusi kelas atau *imbalanced data* merupakan permasalahan yang lazim ditemukan. Pengguna yang mengalami kendala teknis secara psikologis lebih terdorong untuk memberikan ulasan dibandingkan pengguna yang merasa

puas, sehingga ulasan bernada negatif cenderung mendominasi distribusi data secara alami. Kondisi ini menyebabkan permasalahan serius dalam pelatihan model deep learning seperti Bi-LSTM, karena selama proses pelatihan model akan mengoptimalkan prediksi terhadap kelas mayoritas guna meminimalkan nilai kerugian (*loss*) secara keseluruhan. Akibatnya, model menghasilkan nilai akurasi yang tampak tinggi secara kuantitatif, namun gagal mengenali pola sentimen pada kelas minoritas yang justru mengandung informasi penting bagi pengelola layanan. Tanpa penanganan yang tepat terhadap kondisi ini, evaluasi performa model menjadi tidak valid dan informasi yang dihasilkan berpotensi menyesatkan proses pengambilan keputusan (Taskiran et al., 2025; Wang et al., 2022).

Untuk membuktikan bahwa kondisi *imbalanced data* memang terjadi pada domain aplikasi Mobile JKN, pengumpulan data dilaksanakan secara sistematis melalui proses *scraping* pada tanggal 18 Februari 2026 di platform Google Play Store. Melalui dua strategi pengambilan sampel yang digabungkan, proses ini menghasilkan dataset sebanyak 20.247 ulasan bersih yang telah melewati tahapan prapemrosesan dan eliminasi data ambigu. Pelabelan dilakukan menggunakan pendekatan *silver standard labeling* berdasarkan skor rating pengguna sebagai proksi sentimen awal, di mana rating 1–2 dikategorikan sebagai sentimen negatif dan rating 4–5 sebagai sentimen positif (Tamami et al., 2025). Dari total data tersebut, ditemukan dominasi sentimen negatif sebesar 68,64% dan sentimen positif sebesar 31,36%, menghasilkan rasio ketidakseimbangan sebesar 2,19:1. Distribusi ini tidak semata-mata mencerminkan keseluruhan populasi 954.000 ulasan secara proporsional, melainkan mencerminkan karakteristik subset yang dikumpulkan, di mana keluhan teknis pengguna cenderung mendominasi parameter *most_relevant* dibandingkan ulasan kepuasan yang umumnya lebih ringkas. Temuan ini mengonfirmasi keberadaan kondisi *imbalanced data* pada dataset penelitian, sehingga memerlukan penanganan yang tepat agar model yang dihasilkan mampu mengklasifikasikan teks ulasan baru secara objektif tanpa bergantung pada informasi rating.

Penelitian sebelumnya yang menerapkan arsitektur Long Short-Term Memory (LSTM) dengan dukungan Synthetic Minority Over-sampling Technique (SMOTE) — sebuah teknik yang mengatasi ketidakseimbangan kelas dengan membangkitkan sampel data baru pada kelas minoritas melalui interpolasi antar-data yang mirip, bukan sekadar menduplikasi data yang sudah ada — pada domain Mobile JKN melaporkan capaian akurasi sebesar 88% (Tamami et al., 2025). Meskipun hasil tersebut tergolong baik, model yang digunakan bersifat searah (*unidirectional*) sehingga belum sepenuhnya mampu menangkap dependensi konteks dua arah dalam teks ulasan. Selain itu, karakteristik ulasan yang kerap mengandung struktur kalimat kompleks dan penggunaan konjungsi pertentangan menuntut model yang mampu memahami hubungan antarkata secara dua arah. Di sisi lain, tinjauan terhadap beberapa studi terdahulu lainnya (seperti Al Qahar et al., 2024) menunjukkan adanya celah metodologis di mana penerapan SMOTE rentan dilakukan sebelum pemisahan data latih dan uji (*pre-split*), yang berpotensi menimbulkan kebocoran data (*data leakage*). Berdasarkan evaluasi komprehensif terhadap 31 varian SMOTE pada dataset klasifikasi teks, prosedur *oversampling* yang tidak terkontrol dapat menghasilkan evaluasi performa yang terlalu optimis (*over-optimistic*) karena data sintetis yang dihasilkan tidak merepresentasikan kondisi data riil secara akurat (Taskiran et al., 2025).

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan integrasi arsitektur *Bidirectional* LSTM (Bi-LSTM) dengan prosedur *split-aware* SMOTE sebagai pendekatan yang lebih metodologis dan terkontrol. Arsitektur Bi-LSTM memungkinkan pemrosesan urutan teks dari dua arah secara simultan sehingga mampu menangkap konteks semantik secara lebih komprehensif. Penerapan SMOTE secara *split-aware* memastikan bahwa proses *oversampling* dilakukan secara eksklusif pada data latih setelah pemisahan *dataset*, sehingga integritas data uji tetap terjaga dan risiko *data leakage* dapat dihindari. Hal ini didukung oleh temuan Taskiran et al. (2025) yang menekankan pentingnya penerapan *oversampling* hanya pada data latih sebagai standar evaluasi yang valid. Selain itu, penerapan *oversampling* pada ruang representasi fitur terbukti menghasilkan

sampel sintetis yang lebih beragam sekaligus menekan risiko *overfitting* dibandingkan *oversampling* langsung pada ruang fitur mentah (Wang et al., 2022). Representasi fitur teks pada penelitian ini direalisasikan menggunakan word embedding — teknik yang mengonversi setiap kata menjadi vektor bilangan sedemikian rupa sehingga kata-kata dengan makna serupa memiliki representasi numerik yang berdekatan — berbasis algoritma Word2Vec dengan arsitektur Skip-gram berdimensi 100 yang dilatih secara mandiri (trained from scratch) murni pada korpus ulasan Mobile JKN guna menangkap pola semantik dan istilah spesifik domain layanan BPJS yang tidak terwakili secara optimal oleh model pra-latih (pre-trained model) dari domain umum (Mikolov et al., 2013).

Lebih lanjut, penelitian ini memperkuat validitas hasil melalui penerapan uji signifikansi statistik formal menggunakan *Wilcoxon Signed-Rank Test*. Pengujian ini bertujuan untuk memastikan bahwa peningkatan performa model setelah penerapan strategi penanganan ketidakseimbangan data bukan terjadi secara kebetulan, melainkan signifikan secara statistik antara kondisi sebelum dan sesudah penerapan SMOTE. Pengujian ini didasarkan pada dua hipotesis: H_0 yang menyatakan tidak terdapat perbedaan signifikan secara statistik antara performa model Bi-LSTM dengan dan tanpa *split-aware* SMOTE, dan H_1 yang menyatakan terdapat perbedaan yang signifikan secara statistik di antara keduanya. Pendekatan ini memberikan kontribusi metodologis yang lebih kuat dibandingkan studi sebelumnya yang umumnya hanya melaporkan nilai akurasi tanpa verifikasi statistik. Dengan demikian, penelitian ini tidak hanya berfokus pada peningkatan performa klasifikasi sentimen, tetapi juga untuk meningkatkan validitas evaluasi model pada data teks yang *noisy* dan tidak seimbang, sehingga menghasilkan landasan ilmiah yang lebih kokoh dalam analisis opini publik pada layanan kesehatan digital (Budaya & Suniantara, 2024; Deng & Liu, 2025).

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, rumusan masalah yang diangkat adalah sebagai berikut:

1. Bagaimana membangun dan mengevaluasi model klasifikasi sentimen ulasan *Mobile* JKN menggunakan arsitektur Bi-LSTM yang dikombinasikan dengan prosedur *split-aware* SMOTE, serta bagaimana perbandingan performa antara model dengan dan tanpa penerapan SMOTE berdasarkan metrik *Macro* F1-*Score*?
2. Apakah perbedaan performa antara kedua skenario pemodelan tersebut terbukti signifikan secara statistik berdasarkan *Wilcoxon Signed-Rank Test*?
3. Bagaimana prototipe sistem klasifikasi sentimen berbasis web yang dihasilkan dapat dimanfaatkan sebagai referensi evaluasi layanan oleh pengelola aplikasi *Mobile* JKN?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi sentimen ulasan *Mobile* JKN menggunakan arsitektur *Bidirectional Long Short-Term Memory* (Bi-LSTM) yang dikombinasikan dengan prosedur *split-aware* SMOTE sebagai penanganan ketidakseimbangan data, serta mengevaluasi perbandingan performa kedua skenario pemodelan tersebut menggunakan metrik *Macro* F1-*Score*.
2. Memvalidasi secara statistik perbedaan performa model Bi-LSTM dengan dan tanpa penerapan *split-aware* SMOTE menggunakan *Wilcoxon Signed-Rank Test* pada 10 iterasi *Repeated Holdout*.
3. Menghasilkan prototipe sistem klasifikasi sentimen berbasis web yang dapat dimanfaatkan oleh pengelola aplikasi *Mobile* JKN sebagai referensi tambahan dalam mengevaluasi aspek layanan digital yang dikeluhkan pengguna secara sistematis dan berbasis data.

1.4 Batasan Masalah

Agar penelitian ini tetap terfokus dan sesuai dengan tujuan yang ingin dicapai, maka batasan penelitian yang diterapkan adalah sebagai berikut :

1. Data yang digunakan merupakan ulasan pengguna aplikasi *Mobile JKN* berbahasa Indonesia yang diperoleh melalui proses *scraping* pada Google Play

Store pada tanggal 18 Februari 2026 dengan total 20.410 ulasan valid yang kemudian menghasilkan 20.247 ulasan setelah tahapan prapemrosesan teks.

2. Pelabelan sentimen dilakukan secara otomatis berdasarkan skor rating pengguna (*silver standard labeling*), dengan asumsi bahwa skor rating merepresentasikan kecenderungan sentimen pengguna secara umum, di mana rating 1–2 dikategorikan sebagai negatif dan rating 4–5 dikategorikan sebagai positif.
3. Ulasan dengan rating 3 tidak diikutsertakan dalam penelitian karena bersifat ambigu dan tidak merepresentasikan polaritas sentimen yang tegas, sehingga klasifikasi difokuskan pada dua kelas yang jelas.
4. Klasifikasi sentimen yang dilakukan bersifat biner (*binary classification*), yaitu hanya mencakup kelas positif dan negatif tanpa mempertimbangkan kategori netral maupun analisis sentimen berbasis aspek.
5. Model yang digunakan dalam penelitian ini adalah arsitektur *Bidirectional Long Short-Term Memory* (*Bi-LSTM*) tanpa penambahan mekanisme *attention*.
6. Representasi fitur teks menggunakan *Word Embedding* berbasis algoritma Word2Vec dengan arsitektur *Skip-gram* berdimensi 100, yang dilatih secara mandiri (*trained from scratch*) menggunakan korpus ulasan Mobile JKN.
7. Penanganan ketidakseimbangan kelas dilakukan menggunakan SMOTE yang diterapkan secara *split-aware*, yaitu hanya pada data latih setelah pemisahan dataset, untuk menghindari risiko kebocoran data (*data leakage*).
8. Evaluasi performa model dilakukan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score. Uji signifikansi statistik terhadap peningkatan performa dilakukan menggunakan Wilcoxon Signed-Rank Test berdasarkan 10 iterasi eksperimen dengan random split 80:20 dan nilai random seed yang berbeda,di mana *Macro* F1-*Score* dijadikan sebagai metrik utama yang diuji secara berpasangan, karena memberikan bobot yang sama terhadap semua kelas tanpa dipengaruhi dominasi kelas mayoritas.
9. Hasil akhir penelitian diimplementasikan dalam bentuk prototipe sistem demonstrasi berbasis web menggunakan *framework* Flask yang berfungsi

sebagai media pengujian model terhadap input ulasan tunggal maupun data batch dalam format CSV.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Bagi Pengelola Aplikasi Mobile JKN, sebagai referensi tambahan dalam mengevaluasi aspek layanan digital yang dikeluhkan pengguna berdasarkan hasil klasifikasi sentimen ulasan yang dilakukan secara sistematis.
2. Bagi Akademisi, memberikan kontribusi dalam pengembangan kajian *Natural Language Processing* (NLP), khususnya terkait penerapan arsitektur *Bidirectional Long Short-Term Memory* (*Bi-LSTM*) yang dikombinasikan dengan prosedur *split-aware SMOTE* pada dataset ulasan berbahasa Indonesia dengan karakteristik *imbalanced data*.
3. Bagi Penulis, menambah pemahaman dan pengalaman dalam mengimplementasikan siklus penelitian *Data Science* secara komprehensif, mulai dari proses pengumpulan data, prapemrosesan, pengembangan model *deep learning*, hingga pengujian validitas statistik menggunakan *Wilcoxon Signed-Rank Test*.

1.6 Metodologi Penelitian

Untuk dapat mencapai tujuan penelitian dalam membangun model klasifikasi sentimen ulasan Mobile JKN menggunakan arsitektur Bi-LSTM dengan penanganan *imbalanced data*, maka penelitian ini dilakukan melalui tahapan sebagai berikut:

1. Pengumpulan Data

   Data penelitian diperoleh melalui proses *scraping* ulasan pengguna aplikasi *Mobile* JKN pada *Google Play Store* yang dilakukan pada tanggal 18 Februari 2026 menggunakan *library* google-play-scraper. Pengumpulan dilakukan dalam dua *batch* secara *purposive*, yaitu sort=NEWEST dan sort=MOST_RELEVANT masing-masing sebanyak 15.000 ulasan, menghasilkan 30.000 ulasan mentah. Setelah proses deduplikasi antar-batch, eliminasi ulasan yang terlalu pendek, serta penghapusan ulasan dengan rating

3 (netral) yang bersifat ambigu, diperoleh dataset awal sebanyak 20.410 ulasan. Dataset awal ini kemudian melalui tahapan prapemrosesan teks tingkat lanjut yang mereduksi data kosong, hingga pada akhirnya menghasilkan 20.247 ulasan bersih yang digunakan sebagai fondasi utama dalam pemodelan komputasi pada penelitian ini.

2. Pelabelan dan Prapemrosesan Data

   Pelabelan sentimen dilakukan secara otomatis berdasarkan skor rating pengguna. Rating 1–2 dikategorikan sebagai sentimen negatif dan rating 4–5 sebagai sentimen positif, sedangkan rating 3 dieliminasi karena bersifat ambigu dan tidak merepresentasikan polaritas sentimen yang tegas. Selanjutnya dilakukan tahapan prapemrosesan data teks melalui lima proses sekuensial yang meliputi pembersihan teks (*text cleaning*), penyeragaman huruf (*case folding*), pemotongan kata (*tokenization*), normalisasi kata tidak baku (*slang normalization*), serta penghapusan kata secara selektif (*selective stopword removal*) dengan tetap mempertahankan kata negasi dan konjungsi pertentangan guna menjaga polaritas sentimen.

3. Pembagian Dataset (Data Splitting)

   Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20 menggunakan metode *stratified random sampling* untuk menjaga distribusi kelas tetap representatif pada kedua subset data. Proses pembagian dilakukan sebelum penerapan teknik penanganan ketidakseimbangan data guna menghindari terjadinya *data leakage*.

4. Penanganan Ketidakseimbangan Data dengan Split-Aware SMOTE

   Teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan secara *split-aware*, yaitu hanya pada data latih setelah proses pembagian dataset. Pendekatan ini bertujuan untuk meningkatkan representasi kelas minoritas tanpa mempengaruhi data uji, sehingga validitas evaluasi model tetap terjaga.

5. Perancangan dan Pelatihan Model Bi-LSTM

   Model klasifikasi sentimen dibangun menggunakan arsitektur *Bidirectional Long Short-Term Memory* (Bi-LSTM) dengan representasi fitur berbasis algoritma Word2Vec arsitektur *Skip-gram* berdimensi 100 yang dilatih secara mandiri (*trained from scratch*) murni pada korpus ulasan Mobile JKN. Model kemudian dilatih menggunakan data latih yang telah melalui proses penyeimbangan kelas untuk mempelajari pola sentimen dari dua arah konteks urutan kata.

6. Evaluasi dan Pengujian Statistik

   Performa model dievaluasi menggunakan metrik *Accuracy, Precision, Recall*, dan *F1-Score* per kelas. Selanjutnya, untuk memastikan bahwa peningkatan performa akibat penerapan *split-aware* SMOTE tidak terjadi secara kebetulan, dilakukan uji signifikansi statistik menggunakan *Wilcoxon Signed-Rank Test*. Pengujian ini didasarkan pada perbandingan *Macro* F1-*Score* yang dihasilkan dari 10 kali iterasi eksperimen (*Repeated Holdout*) menggunakan pembagian data 80:20 secara *stratified random.*

1.7 Sistematika Penulisan

Agar mempermudah pemahaman pada pembahasan penulisan skripsi ini, maka sistematika penulisan diperoleh sebagai berikut:

BAB I	:	Bab ini memuat latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, serta sistematika penulisan.
BAB II	:	Bab ini memuat penelitian terdahulu yang relevan serta landasan teori yang digunakan dalam penelitian, meliputi *sentiment analysis*, *Natural Language Processing* (NLP), arsitektur *Bidirectional Long Short-Term Memory* (Bi-LSTM), teknik penanganan *imbalanced data* menggunakan *SMOTE*, representasi fitur berbasis *Word2Vec*, metode evaluasi klasifikasi, serta pengujian statistik.

BAB III	:	Bab ini membahas analisis kebutuhan sistem dan data penelitian, perancangan arsitektur model, diagram blok sistem, alur proses klasifikasi, serta rancangan prototipe sistem sebagai media implementasi dan pengujian model.
BAB IV	:	Bab ini menguraikan implementasi sistem secara teknis beserta antarmuka aplikasi berbasis *web*. Lebih lanjut, bab ini menjabarkan algoritma internal yang bekerja pada setiap tahapan, mulai dari hasil prapemrosesan teks, penanganan *imbalanced data* menggunakan algoritma SMOTE, proses pelatihan model Bi-LSTM, hingga evaluasi performa model.
BAB V	:	Bab ini memuat kesimpulan dari hasil penelitian yang dilakukan untuk menjawab rumusan masalah secara ringkas dan padat.