

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi digital di berbagai sektor, termasuk perbankan [1]. Digitalisasi ini memberikan kemudahan akses dan efisiensi dalam berbagai layanan perbankan, salah satunya melalui *mobile banking*. Layanan ini memungkinkan pengguna untuk melakukan transaksi keuangan seperti transfer dana, pembayaran tagihan, dan pengecekan saldo hanya melalui perangkat mobile, seperti *smartphone* [2].

Di Indonesia, Jumlah pengguna *mobile banking* menunjukkan tren positif, dengan beberapa bank besar seperti BRI, BCA, Mandiri, dan BNI mencatatkan pertumbuhan pesat dalam jumlah pengguna layanan tersebut [3]. Dikutip berdasarkan CNBC Indonesia tahun 2024 dalam data PT Bank Rakyat Indonesia (Persero) Tbk (BRI) melalui aplikasi BRImo mencatat pencapaian tertinggi dengan 33,5 juta pengguna, meningkat 30,3% dibandingkan kuartal I 2023. PT Bank Central Asia Tbk (BCA) mencatatkan 30,8 juta pengguna *mobile banking*, tumbuh 9%. PT Bank Mandiri (Persero) Tbk dengan Super App Livin' by Mandiri mencatat pertumbuhan pengguna sebesar 39% menjadi 24 juta pada kuartal I 2024. Namun, PT Bank Negara Indonesia (Persero) Tbk (BNI), meskipun mencatat pertumbuhan 18,5% pengguna menjadi 16,9 juta pada kuartal I 2024 [3], masih tertinggal dibandingkan dengan kompetitor lainnya.

Untuk memperkuat posisinya di sektor perbankan digital, BNI meluncurkan aplikasi baru bernama Wondr by BNI pada Juli 2024. Wondr by BNI merupakan layanan *mobile banking* terbaru yang dikembangkan oleh Bank Negara Indonesia (BNI) untuk memenuhi berbagai kebutuhan finansial para nasabah [4][5]. Aplikasi Wondr by BNI dapat diunduh secara gratis melalui Google Play Store dan telah mencapai lebih dari 5 juta unduhan dengan rating 3,8 per 31 Desember 2024, yang lebih rendah dibandingkan aplikasi pendahulunya, BNI Mobile Banking, yang memiliki rating 4,5. Rating yang rendah ini menunjukkan bahwa kinerja aplikasi belum sepenuhnya memuaskan pengguna [6][7]. Untuk memperoleh gambaran yang lebih jelas mengenai kinerja aplikasi, ulasan pengguna sangat penting karena memberikan informasi yang lebih detail dan komprehensif.

Ulasan pengguna tidak hanya memberikan informasi lebih mendalam dibandingkan rating, tetapi juga membantu pengguna lain memahami kelebihan dan kekurangan aplikasi [8]. Bagi penyedia layanan, ulasan pengguna menjadi sumber masukan berharga untuk memperbaiki dan meningkatkan kualitas produk atau layanan mereka [9]. Namun, jumlah ulasan yang terlalu banyak dapat mempersulit pembaca untuk membaca secara keseluruhan dan menemukan informasi yang relevan. Untuk mengatasi hal ini, analisis opini atau analisis sentimen dapat digunakan guna menyaring data dan menghasilkan wawasan yang lebih bermakna.

Analisis sentimen adalah cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengenali dan mengevaluasi polaritas sentimen dalam sebuah kalimat, yang dapat berupa positif, negatif, atau netral [10]. Umumnya, analisis ini hanya fokus pada polaritas sentimen saja. Namun, dalam kenyataannya, sebuah kalimat seringkali memiliki beberapa aspek dengan sentimen berbeda pada masing-masing aspek [9]. Untuk itu, diperlukan analisis yang lebih mendalam, yaitu *Aspect-Based Sentiment Analysis* (ABSA). ABSA adalah model yang dirancang untuk mengidentifikasi aspek-aspek spesifik dari suatu ulasan dan menentukan sentimen yang berkaitan dengan setiap aspek tersebut [10].

Penelitian mengenai ABSA telah banyak dilakukan dengan menggunakan berbagai algoritma *machine learning*, seperti Naive Bayes Classifier, Support Vector Machine (SVM), Decision Tree, dan lain sebagainya [11]. Namun, mayoritas algoritma *machine learning* menghadapi tantangan dalam menangani dataset dengan ketidakseimbangan kelas. Ketidakseimbangan ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga mengurangi akurasi prediksi pada kelas minoritas [12]. Untuk mengatasi keterbatasan ini, penelitian ini memanfaatkan pendekatan *ensemble learning*. Metode *ensemble* adalah metode yang menggabungkan beberapa *classifier* individu untuk membentuk *classifier* baru, dengan tujuan meningkatkan akurasi hasil. Teknik ini telah banyak diterapkan dalam berbagai penelitian karena terbukti mampu memberikan tingkat akurasi yang lebih tinggi terutama dalam dataset tidak seimbang [13]. Salah satu teknik *ensemble* yang implementasikan pada penelitian ini yaitu teknik *stacking*. Teknik *Stacking* adalah teknik yang menggabungkan berbagai jenis *classifier*, di mana sejumlah *classifier* individu disebut sebagai *base learner*, sedangkan *classifier* yang bertugas untuk mengintegrasikan hasil dari *base learner* dikenal sebagai *meta learner* tujuannya meningkatkan menghasil akurasi [14]. Teknik *stacking* dipilih karena, berdasarkan

penelitian [15], teknik ini memperoleh nilai akurasi terbaik dibandingkan teknik ensemble lainnya. Teknik stacking menggabungkan model-model dengan kelebihan dan kekurangan yang berbeda untuk menghasilkan prediksi yang lebih akurat dan stabil, serta mengurangi kesalahan dibandingkan dengan penggunaan model tunggal. Selain itu, teknik ini juga mencegah overfitting dan meningkatkan ketahanan model [16].

Namun, meskipun metode machine learning seperti *ensemble learning* mampu meningkatkan akurasi, terdapat masalah lain yang perlu diatasi, yaitu kurangnya interpretabilitas hasil prediksi. Prediksi yang dihasilkan sering kali tidak memberikan informasi yang cukup rinci mengenai alasan di balik klasifikasi positif atau negatif suatu ulasan. Proses klasifikasi ini sering disebut sebagai *blackbox*, yang menjadi tantangan dalam memberikan transparansi [18]. Kondisi *blackbox* ini memerlukan pendekatan interpretasi yang dapat meningkatkan kepercayaan terhadap model, terutama jika hasil prediksi akan digunakan sebagai dasar pengambilan keputusan atau pengembangan model baru [19]. Salah satu metode yang dapat digunakan untuk melakukan interpretasi hasil model machine learning adalah *Local Interpretable Model-agnostic Explanations* (LIME). LIME adalah sebuah teknik untuk menjelaskan prediksi model, yang bertujuan memberikan penjelasan lokal terhadap keputusan model *black box* tanpa bergantung pada jenis model tertentu. LIME dapat digunakan untuk menjelaskan alasan di balik keputusan model dalam memberikan label positif atau negatif pada sebuah teks [18].

Pendekatan *stacking ensemble* pernah dilakukan pada penelitian ABSA yang dilakukan oleh [17] dengan judul "*Aspect-Based Sentiment Analysis with an Ensemble Learning Framework for Requirements Elicitation from App Reviews*" memperkenalkan kerangka stacking ensemble untuk Aspect-Based Sentiment Analysis (ABSA). Pendekatan ini mengatasi keterbatasan model tunggal dalam *Aspect Category Detection* (ACD) dan *Aspect Category Polarity* (ACP) pada ulasan aplikasi. Aspek yang dianalisis mencakup tiga domain utama, yaitu *Social Networking*, *Games*, dan *Productivity*. Kerangka ini menggabungkan model linear (Logistic Regression), berbasis pohon keputusan (Random Forest dan XGBoost), serta non-linear (Multilayer Perceptron dan Support Vector Machine). Prediksi dari model dasar dijadikan fitur untuk meta-model Logistic Regression yang menyelaraskan keluaran dan menghasilkan prediksi akhir. Pendekatan ini meningkatkan performa signifikan dibanding baseline, dengan keunggulan dalam generalisasi dan pengolahan data

kompleks. Strategi penggabungan berbobot berbasis skor F1 meningkatkan keandalan prediksi. Hasilnya, skor F1 pada tugas ACD meningkat 22,9%-28,4%, dan akurasi pada tugas ACP naik 9,3%-13,2%.

Penelitian serupa lainnya juga pernah dilakukan oleh [15] Penelitian ini bertujuan untuk menganalisis sentimen berbasis aspek pada ulasan restoran menggunakan algoritma *ensemble*. Data ulasan sebanyak 19.492 dari TripAdvisor dikategorikan berdasarkan lima aspek utama: makanan, layanan, staf, suasana, dan harga. Untuk klasifikasi sentimen, pendekatan berbasis leksikon (VADER) digunakan, diikuti dengan algoritma pembelajaran mesin seperti Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM). Penelitian ini menggabungkan metode ensemble homogen (Bagging) dan heterogen (Stacking) untuk meningkatkan akurasi prediksi. Hasil menunjukkan bahwa metode Stacking menghasilkan akurasi terbaik secara keseluruhan. Akurasi tertinggi untuk masing-masing aspek adalah makanan sebesar 96,2%, layanan 97,4%, staf 98,2%, suasana 96,4%, dan harga 91,8%.

Penelitian terkait interpretasi model menggunakan LIME pernah dilakukan oleh [19] menggunakan teknik *stacking ensemble* dengan meta model Logistic Regression yang menggabungkan prediksi dari base model Naïve Bayes, Random Forest, dan SVM untuk analisis sentimen ulasan aplikasi pinjaman online, mencapai akurasi terbaik sebesar 87,05%. Untuk mengatasi masalah *blackbox*, penelitian ini memanfaatkan *Local Interpretable Model-agnostic Explanations* (LIME) guna memberikan interpretasi prediksi, seperti mengidentifikasi kata-kata spesifik yang berkontribusi pada klasifikasi sentimen. Misalnya, pada ulasan negatif, kata-kata seperti "teror" dan "jatuh" diidentifikasi sebagai kontributor utama sentimen negatif dengan probabilitas 95%. Penelitian ini tidak hanya meningkatkan akurasi model dengan teknik ensemble tetapi juga meningkatkan transparansi dan kepercayaan pengguna terhadap hasil prediksi model.

Berdasarkan uraian masalah yang telah dijelaskan diatas, maka peneliti mengusulkan penelitian dengan judul “Penerapan Metode Ensemble Machine Learning Dalam Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Wondr By BNI”. Penelitian ini bertujuan menganalisis sentimen dan aspek pada ulasan aplikasi Wondr By BNI. Data yang digunakan dalam penelitian ini diperoleh melalui *scraping* ulasan aplikasi Wondr by BNI dari Play Store dan App Store. Penelitian ini akan dilakukan dua kali klasifikasi yang pertama klasifikasi masing-masing aspek, yang kedua klasifikasi sentimen tiap aspek kemudian dilakukan interpretasi pada model

*stacking ensemble*. Luaran dari penelitian ini nantinya berupa sebuah website yang dapat digunakan untuk menganalisis aspek dan sentimen ulasan aplikasi. Melalui penelitian ini, diharapkan dapat ditemukan wawasan strategis bagi pengembangan aplikasi Wondr by BNI yang lebih sesuai dengan kebutuhan pengguna.

## 1.2. Rumusan Masalah

Berdasarkan penjelasan sebelumnya, beberapa poin rumusan masalah yang dapat disusun adalah sebagai berikut:

1. Bagaimana kinerja metode *stacking ensemble* dibandingkan algoritma individu *base learners* dan *meta learner* dalam mengklasifikasikan masing-masing aspek ulasan pengguna aplikasi Wondr by BNI?
2. Bagaimana performa metode *stacking ensemble* dibandingkan algoritma individu *base learners* dan *meta learner* dalam mengklasifikasikan sentimen positif dan negatif pada setiap aspek ulasan pengguna aplikasi Wondr by BNI?
3. Apakah penerapan Local Interpretable Model-agnostic Explanations (LIME) dapat memberikan penjelasan mengenai pengambilan keputusan model *stacking ensemble* dalam menganalisis ulasan pengguna aplikasi Wondr by BNI?
4. Bagaimana mengembangkan website dengan menggunakan *model stacking ensemble* untuk menghasilkan prediksi aspek sentimen yang akurat serta menyediakan wawasan yang mendukung pengambilan keputusan?

## 1.3. Tujuan Penelitian

Berdasarkan penjelasan yang telah disampaikan, penelitian ini bertujuan untuk mencapai hal-hal berikut:

1. Membandingkan performa metode *stacking ensemble*, algoritma *base learners*, dan *meta learner* dalam mengidentifikasi dan mengklasifikasikan setiap aspek ulasan pengguna aplikasi Wondr by BNI.
2. Membandingkan kemampuan *stacking ensemble*, algoritma *base learners*, dan *meta learner* dalam mendeteksi sentimen positif dan negatif pada setiap aspek ulasan pengguna aplikasi Wondr by BNI.
3. Menganalisis sejauh mana penerapan *Local Interpretable Model-agnostic Explanations* (LIME) dapat memberikan penjelasan yang komprehensif terkait pengambilan keputusan model *stacking ensemble* dalam menganalisis ulasan.

4. Mengembangkan website dengan kemampuan memproses data input manual menggunakan model stacking ensemble untuk menghasilkan prediksi aspek sentimen yang akurat serta menyediakan wawasan yang mendukung pengambilan keputusan.

#### **1.4. Batasan Penelitian**

Batasan masalah yang diterapkan dalam penelitian ini, sesuai dengan rumusan masalah yang telah disusun, adalah sebagai berikut:

1. Penelitian ini menggunakan data berupa ulasan pengguna aplikasi wondr by BNI yang diperoleh dari playstore dan appstore dengan rentang waktu 31 Juli 2024 - 31 Desember 2024.
2. Data yang dianalisis berbentuk teks dalam bahasa Indonesia.
3. Algoritma *base learners* yang digunakan Support Vector Machine (SVM), Random Forest, dan Naive Bayes. sedangkan *meta learner* yang digunakan adalah Logistic Regression, pemilihan ini didasari penelitian oleh [17].
4. Pemodelan topic menggunakan BERTopic.
5. Sentimen dalam data diklasifikasikan menjadi dua kategori, yaitu positif dan negatif.
6. Bahasa pemrograman yang digunakan dalam penelitian ini adalah Python.

#### **1.5. Manfaat Penelitian**

Melalui penelitian ini, diharapkan dapat memberikan manfaat yang signifikan bagi perusahaan pengelola aplikasi Wondr by BNI. Temuan terkait performa metode stacking ensemble dalam menganalisis ulasan pengguna dapat digunakan sebagai bahan evaluasi untuk:

1. Mengidentifikasi kekuatan dan kelemahan aplikasi berdasarkan klasifikasi sentimen setiap aspek ulasan pengguna.
2. Memberikan dasar yang kuat untuk merancang strategi peningkatan kualitas fitur, pengalaman pengguna, dan layanan aplikasi secara keseluruhan, sehingga lebih sesuai dengan kebutuhan dan harapan pengguna.

Selain itu, penelitian ini juga menyoroti potensi dan keunggulan teknik stacking dalam ensemble learning untuk analisis sentimen dan aspek pada data dari aplikasi mobile banking Wondr. Temuan ini dapat dijadikan rujukan bagi penelitian

lanjutan maupun pengembangan aplikasi serupa di bidang serupa, guna mendukung inovasi berbasis data yang lebih terarah.

## **1.6. Sistematika Penulisan**

Sistematika penulisan skripsi ini disusun sebagai panduan agar laporan penelitian terarah dan sesuai dengan tujuan yang ingin dicapai. Langkah-langkah penyusunannya dijelaskan sebagai berikut.

### **BAB I**

#### **PENDAHULUAN**

Bab ini berisi gambaran umum penelitian yang mencakup latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### **BAB II**

#### **TINJAUAN PUSTAKA**

Bab ini memuat penelitian-penelitian terdahulu untuk membandingkan pendekatan dan temuan sebelumnya dengan penelitian saat ini. Selain itu, dijelaskan juga landasan teori yang mendukung penelitian.

### **BAB III**

#### **METODOLOGI PENELITIAN**

Memuat tahapan-tahapan yang dilakukan untuk mencapai tujuan penelitian, mencakup proses identifikasi masalah, teknik pengumpulan data, pengolahan data, serta penyusunan laporan.

### **BAB IV**

#### **HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil penelitian yang mencakup pengolahan data, pembangunan model klasifikasi aspek dan sentimen, evaluasi kinerja model, serta pembahasan temuan.

**BAB V****PENUTUP**

Bab ini berisi kesimpulan yang merangkum hasil penelitian secara keseluruhan, serta saran untuk pengembangan penelitian di masa mendatang.

**DAFTAR  
PUSTAKA**

Bab ini memuat daftar referensi yang digunakan dalam penelitian, termasuk buku, jurnal, artikel, dan sumber online yang relevan.