

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pada era sekarang ini, reputasi sebuah restoran bisa dibuat atau dihancurkan hanya melalui unggahan di media sosial. Ketika pelanggan membagikan pengalaman mereka makan di Bebek Sinjay - baik lewat *Google Reviews*, Instagram, Facebook, atau platform lainnya setiap komentar itu seperti suara yang langsung memengaruhi calon pembeli lain. Di dunia bisnis, analisis sentimen sering digunakan sebagai cara perusahaan untuk mengevaluasi kepuasan pengguna dengan melihat opini publik atau konsumen. Namun, dengan pesatnya perkembangan media sosial saat ini, perusahaan dapat menggunakan platform tersebut sebagai sumber informasi untuk mengetahui bagaimana opini konsumen tanpa harus melakukan cara konvensional seperti *survey*, *polling*, dan lain sebagainya. Kini, orang dapat dengan mudah menyampaikan pendapatnya di platform-platform media sosial secara bebas [1]. Rumah Makan Bebek Sinjay yang ada di Madura adalah salah satu destinasi kuliner yang terkenal dengan hidangan Bebek khas nya, bebek goreng. Kepopulerannya yang luas menjadikan Bebek Sinjay sebagai tujuan favorit bagi wisatawan lokal maupun mancanegara. Namun, dengan meningkatnya jumlah pengunjung, beragam ulasan mulai muncul di *Google Maps*, mencakup tanggapan positif dan negatif yang dapat mempengaruhi pandangan calon pelanggan. Oleh karena itu, penting bagi pengelola Bebek Sinjay untuk memahami sentimen dari ulasan tersebut demi menjaga dan memperbaiki kualitas pelayanan [2].

Analisis sentimen adalah salah satu pendekatan yang dapat digunakan untuk menilai pandangan atau sikap Masyarakat terhadap produk atau layanan berdasarkan teks ulasan. Satu hal yang menjadi kunci dalam analisis sentimen adalah metode yang digunakan untuk membangun model klasifikasi. Selama ini, *machine learning* masih banyak digunakan untuk membangun model. *Machine learning* bisa memberikan solusi dari berbagai permasalahan sehari-hari yang sebelumnya sangat sulit dilakukannya karena keterbatasan teknologi dan algoritma yang ada [3]. Kemampuan sistem *machine learning* bisa belajar secara otomatis

dari data yang diberikan dan dapat meningkatkan kinerjanya tanpa harus secara eksplisit ditulis kembali program komputernya [4]. Oleh karena itu, tantangan dari *machine learning* ini, ada pada (1) perumusan masalah atau isu apa yang hendak diselesaikan dengan *machine learning*, (2) membangun dataset dan data *training set* agar bisa diolah dan memberikan solusi, (3) pemilihan model dan algoritma, dll [5].

Suatu penelitian pernah dilakukan untuk menganalisis sentimen wisata kuliner melalui data Instagram menggunakan metode *Naïve Bayes* dan diperoleh akurasi sebesar 86.87% [6]. Beberapa peneliti sebelumnya sudah mencoba menganalisis ulasan pelanggan restoran dengan pendekatan yang berbeda. Para peneliti tersebut memanfaatkan metode *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (K-NN), dua teknik canggih dalam analisis data. SVM bekerja dengan cara menemukan pola tersembunyi dalam teks untuk membedakan ulasan positif dan negatif, sementara K-NN mengelompokkan ulasan berdasarkan kemiripan karakteristiknya. Kedua metode ini ibarat memiliki detektor khusus yang bisa mengukur tingkat kepuasan pelanggan melalui tulisan mereka, meski dengan pendekatan yang berbeda-beda. Hasilnya menunjukkan bahwa analisis semacam ini memang efektif untuk memahami sentimen pelanggan secara objektif. Penelitian-penelitian sebelumnya ini menjadi landasan berharga bagi kita untuk mengembangkan analisis yang lebih tepat guna, khususnya untuk kasus Bebek Sinjay dengan karakteristik ulasan yang unik. Dari hasil perhitungan SVM memiliki nilai akurasi 59.03%[7]. Dalam penelitian yang dilakukan oleh Fitri [8], Sebuah penelitian sebelumnya pernah membandingkan tiga metode berbeda untuk menganalisis sentimen pelanggan, yaitu *Naïve Bayes*, SVM (*Support Vector Machine*), dan *Random Forest*. Ketiga metode ini diuji untuk mengukur sejauh mana keakuratan mereka dalam mengkategorikan ulasan yang positif dan negatif. Hasilnya sangat menggembirakan - teknik *Random Forest* menunjukkan hasil terbaik dengan akurasi mencapai 97,16%, melebihi dua metode lainnya. Angka ini menunjukkan bahwa *Random Forest* lebih unggul dalam menangkap nuansa bahasa dan pola dalam ulasan pelanggan. Temuan ini menjadi pertimbangan penting bagi kami dalam memilih metode yang paling efektif untuk menganalisis sentimen

terhadap Bebek Sinjay, karena akurasi yang tinggi berarti hasil analisis akan lebih mendekati gambaran sebenarnya tentang pendapat pelanggan.

Dalam penelitian lain juga melakukan analisis sentimen dengan membandingkan 4 Beberapa metode *machine learning* seperti *Naïve Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* telah diuji dalam berbagai penelitian untuk menganalisis sentimen pelanggan. Dari berbagai percobaan tersebut, *Random Forest* secara konsisten menunjukkan performa yang mengesankan. Metode ini terbukti lebih unggul dalam menangkap pola-pola kompleks dalam data teks ulasan pelanggan. Kemampuannya dalam menganalisis dan mengklasifikasikan sentimen dengan akurasi tinggi membuat *Random Forest* menjadi pilihan yang lebih dapat diandalkan dibandingkan metode lainnya. Keunggulan ini yang mendasari pemilihan *Random Forest* sebagai pendekatan utama dalam penelitian analisis sentimen terhadap Bebek Sinjay, karena mampu memberikan hasil yang lebih akurat dan komprehensif dalam memahami opini pelanggan akurasi klasifikasi terbaik [9]. Berdasarkan penelitian yang telah dilakukan diatas, *Random Forest* memang telah terbukti unggul dibanding metode *machine learning* lainnya dalam berbagai penelitian. Yang membuatnya istimewa adalah kemampuannya memberikan hasil analisis dengan tingkat akurasi yang sangat tinggi sekaligus meminimalkan kesalahan. Tak heran jika metode ini sering menjadi pilihan utama, karena selain bisa bekerja dengan sangat teliti, ia juga handal dalam mengolah data dalam skala besar tanpa kehilangan ketepatan. Kelebihan inilah yang membuat *Random Forest* sangat cocok untuk menganalisis ribuan ulasan pelanggan Bebek Sinjay, dimana kita butuh metode yang tidak hanya akurat tapi juga mampu menangani data dalam jumlah banyak dengan efisien. Kemampuannya membaca pola-pola kompleks dalam teks ulasan memungkinkan kita mendapatkan gambaran yang lebih utuh tentang sentimen pelanggan.

*Random Forest* bekerja dengan menerapkan algoritma *decision tree* untuk menentukan kepuasan. Jika dalam *decision tree* hanya menggunakan satu pohon untuk melakukan prediksi, maka pada *Random Forest* berisi banyak pohon keputusan yang membentuk suatu hutan (*forest*). Gabungan dari *tree* tersebut menentukan hasil akhir berupa keputusan yang diperoleh dari *majority voting* hasil keputusan setiap *tree* [10]. Dalam hal ini *Random Forest* dinilai mampu mengatasi

masalah yang ditimbulkan ketika melakukan klasifikasi hanya dengan satu *decision tree* [11].

## 1.2 Rumusan Masalah

### 1.2.1 Permasalahan

Kuliner selalu berkaitan dengan persepsi masyarakat, baik dari segi rasa pelayanan, maupun harga. Salah satu kuliner yang cukup populer adalah Bebek Sinjay Madura. Untuk memahami opini Masyarakat terhadap kuliner ini, diperlukan metode klasifikasi yang mampu mengidentifikasi pola secara akurat guna mendukung evaluasi dan perbaikan layanan. Salah satu pendekatan yang bisa kita gunakan untuk menganalisis sentimen pelanggan adalah dengan memanfaatkan *Random Forest Classifier*. Yang dikenal memiliki performa tinggi dibandingkan dengan metode klasifikasi lainnya. Namun ketidakseimbangan kelas (*class imbalance*) terjadi ketika data yang dikumpulkan untuk klasifikasi opini masyarakat (seperti ulasan positif dan negatif terhadap Rumah Makan Bebek Sinjay Madura) memiliki distribusi yang sangat tidak merata. Kondisi ini menjadi masalah kritis karena model *Machine Learning* seperti *Random Forest Classifier* meskipun secara umum memiliki performa tinggi akan cenderung mempelajari pola dari kelas mayoritas (positif) dan mengabaikan kelas minoritas (negatif). Akibatnya, model mungkin menunjukkan akurasi tinggi secara keseluruhan, tetapi sebenarnya gagal mengenali ulasan negatif sama sekali. Padahal, justru ulasan negatif inilah yang paling berharga untuk evaluasi dan perbaikan layanan. Metode *balancing* data diperlukan untuk mengatasi bias ini dengan menyeimbangkan distribusi kelas, baik dengan cara *Undersampling* (mengurangi sampel kelas mayoritas), *Oversampling* (menambah replika/variasi sampel kelas minoritas seperti teknik SMOTE) atau gabungan keduanya. Tanpa penyeimbangan, resiko terbesar adalah model tidak mampu mengidentifikasi pola keluhan atau kritik yang meskipun jumlahnya sedikit, memiliki dampak strategis bagi peningkatan kualitas kuliner dan pelayanan Bebek Sinjay Madura.

### 1.2.2 Metode Usulan

Penelitian ini menganalisis sentimen pelanggan Rumah Makan Bebek Sinjay dengan metode *Random Forest*. Data ulasan diambil dari *Google Maps*

menggunakan teknik *scraping* untuk mendapatkan hasil yang akurat dan komprehensif.

### **1.2.3 Pertanyaan Penelitian**

Bagaimana pengaruh penggunaan metode SMOTE terhadap performa metode *Random forest* dalam mengklasifikasikan model ulasan pelanggan bebek sinjay ?

## **1.3 Tujuan dan Manfaat**

### **1.3.1 Tujuan Penelitian**

Untuk menganalisis dampak penerapan metode SMOTE (*Synthetic Minority Over-sampling Technique*) terhadap akurasi, presisi, dan recall model klasifikasi *Random Forest* dalam mengkategorikan sentimen ulasan pelanggan Bebek Sinjay.

### **1.3.2 Manfaat Penelitian**

Manfaat dari penelitian ini adalah membantu pengelola Bebek Sinjay mengidentifikasikan kritik pelanggan secara lebih akurat khusus nya pada ulasan negatif yang sering terabaikan melalui peningkatan kemampuan deteksi model *Random Forest* setelah penerapan SMOTE.

### **1.3.3 Batasan Masalah**

Adapun Batasan masalah dalam penelitian ini adalah:

1. Ulasan yang dipakai berasal dari *Google Maps*
2. Ulasan yang digunakan Rumah Makan Bebek Sinjay
3. Data yang digunakan sebanyak 1.634, dibagi menjadi 2 kelas yaitu positif sebanyak 1.146 data dan Negatif sebanyak 486 data.
4. Ulasan yang digunakan adalah ulasan dalam kurun waktu kurang lebih 4 tahun mulai tahun 2021-2025

### **1.3.4 Metodologi Penelitian**

Metodologi penelitian adalah cara sistematis yang digunakan peneliti untuk mengumpulkan dan menganalisis data guna menjawab pertanyaan penelitian.

## **1.5 Sistematika Penulisan**

Proposal penelitian ini terdiri dari lima bagian utama, yaitu:

### **BAB 1 PENDAHULUAN**

Bab ini menjelaskan berbagai aspek mendasar yang melandasi penelitian. Latar belakang memaparkan konteks dan alasan mengapa topik ini penting untuk diteliti, termasuk masalah atau fenomena yang melatarbelakangi studi ini.

### **BAB 2 TINJAUAN PUSTAKA**

Bab ini menguraikan berbagai teori dasar yang menjadi landasan penelitian. Teori-teori tersebut dipilih karena relevansinya dalam menjelaskan fenomena yang diteliti, sekaligus berfungsi sebagai kerangka pemikiran untuk menganalisis data dan menjawab pertanyaan penelitian.

### **BAB 3 METODE USULAN**

Bab ini memaparkan langkah demi langkah proses penelitian secara lengkap dan terstruktur. Awalnya dilakukan studi literatur untuk menggali konsep-konsep dasar yang relevan dengan penelitian, sekaligus memahami landasan teoritis yang diperlukan. Setelah itu, peneliti mempersiapkan segala kebutuhan teknis dan konseptual untuk merancang sistem, termasuk menentukan spesifikasi dan alat yang akan digunakan.

### **BAB 4 HASIL DAN PEMBAHASAN**

Bab ini menjelaskan tentang temuan penelitian mulai dari proses pengumpulan data hingga tahap evaluasi akhir, menggambarkan capaian hasil yang diperoleh sesuai tujuan penelitian. Penjelasan disajikan secara runtut untuk memberikan pemahaman utuh tentang perkembangan penelitian dari awal hingga kesimpulan akhir.

### **BAB 5 PENUTUP**

Bab ini berisi hasil analisis *Random Forest*, kesimpulan penelitian, serta saran untuk pengembangan riset selanjutnya.