

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pemodelan prediksi penjualan merupakan komponen penting dalam pengambilan keputusan bisnis berbasis data, khususnya pada pemasaran digital yang dinamis. Prediksi yang akurat dapat membantu pelaku usaha menentukan strategi produksi, persediaan, dan promosi secara lebih adaptif. Salah satu algoritma *machine learning* yang banyak digunakan untuk peramalan berbasis data historis merupakan *extreme gradient boosting* (XGBOOST), yaitu metode *Ensemble* berbasis *decision tree* yang dioptimalkan melalui mekanisme *Gradient boosting* serta dilengkapi regularisasi untuk mengurangi risiko *overfitting*. Karakteristik ini membuat XGBOOST relevan untuk kasus prediksi pada data yang memiliki hubungan nonlinier dan dipengaruhi banyak faktor [1]. Prediksi penjualan yang akurat dapat membantu pelaku usaha dalam menentukan strategi produksi, pengelolaan persediaan, serta perencanaan promosi secara lebih adaptif. Hal ini menjadi semakin penting bagi usaha mikro, kecil, dan menengah (UMKM) yang memiliki keterbatasan sumber daya, sehingga membutuhkan perencanaan yang efisien dan berbasis informasi yang andal.

Salah satu algoritma *machine learning* yang banyak digunakan dalam peramalan berbasis data historis yaitu *extreme gradient boosting* (XGBOOST). XGBOOST merupakan metode *Ensemble* berbasis *decision tree* yang dioptimalkan melalui mekanisme *Gradient boosting* serta dilengkapi dengan teknik regularisasi untuk mengurangi risiko *overfitting*. Karakteristik tersebut menjadikan XGBOOST mampu memodelkan hubungan nonlinier dan menangani data yang dipengaruhi oleh banyak faktor, sehingga relevan untuk kasus prediksi penjualan dengan pola yang fluktuatif dan kompleks [2].

Pada penelitian ini, objek kajian merupakan UMKM Pawonkoe Banyuwangi, yang memproduksi makanan ringan berbasis potensi lokal, yaitu keripik salmon dan keripik rambak cumi. Berdasarkan rekap data penjualan periode 01 Januari 2024 hingga 31 Desember 2024, jumlah penjualan menunjukkan fluktuasi yang cukup signifikan. Penjualan tertinggi terjadi pada bulan Maret 2024 sebesar 2.500

unit, sedangkan penjualan terendah tercatat pada bulan September 2024 sebesar 1.009 unit. Fluktuasi tersebut berpotensi menimbulkan ketidakpastian permintaan yang dapat berdampak pada ketidakefisienan perencanaan produksi dan pengelolaan persediaan apabila tidak didukung oleh model prediksi penjualan yang memadai.

Dalam peramalan penjualan pada usaha skala kecil masih sering menggunakan pendekatan deret waktu klasik seperti *Autoregressive Integrated Moving Average* (ARIMA). Metode ARIMA efektif pada data yang bersifat linier dan stasioner atau yang dapat dibuat stasioner melalui proses *differencing* dan transformasi tertentu. Namun, ARIMA memiliki keterbatasan dalam menangkap pola data yang kompleks dan nonlinier [3], sehingga kurang optimal ketika diterapkan pada data penjualan UMKM yang fluktuatif dan dipengaruhi oleh berbagai faktor. Kondisi ini menunjukkan perlunya pendekatan alternatif yang lebih fleksibel dan adaptif dalam memodelkan pola penjualan.

Sebagai alternatif, penerapan XGBOOST dinilai layak karena mampu memodelkan hubungan nonlinier dengan performa yang kompetitif pada berbagai kasus prediksi berbasis data historis. Penelitian terdahulu pada domain penjualan *e-commerce* menunjukkan bahwa XGBOOST dapat menghasilkan peramalan penjualan yang lebih akurat sekaligus mendukung strategi pemasaran berbasis data [4]. Temuan tersebut memperkuat argumentasi bahwa XGBOOST berpotensi untuk diadopsi pada konteks UMKM guna meningkatkan kualitas prediksi penjualan, termasuk pada UMKM Pawonkoe Banyuwangi.

Selain data numerik penjualan, ekosistem pemasaran digital juga menghasilkan data tekstual berupa ulasan pelanggan yang mengandung opini, persepsi, serta sentimen konsumen. Literatur menunjukkan bahwa ulasan atau word-of-mouth daring dapat memengaruhi perilaku pembelian dan pola penjualan, sehingga analisis ulasan pelanggan menjadi sumber informasi penting bagi pengambilan keputusan bisnis [5]. Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif, netral, dan negatif serta banyak diterapkan dalam analisis media sosial dan visualisasi teks menggunakan *word cloud* [6]. Pada UMKM Pawonkoe, tersedia data survei ulasan

pelanggan periode 01 Januari 2024 hingga 31 Desember 2024 sebanyak 365 data ulasan, yang berpotensi merepresentasikan evaluasi pelanggan terhadap kualitas produk dan layanan secara lebih objektif.

Untuk mengolah ulasan berbahasa Indonesia secara akurat, pendekatan *Natural Language Processing* (NLP) berbasis *Transformer* menjadi relevan. Model *Bidirectional Encoder Representations from Transformers* (BERT) mampu menghasilkan representasi kontekstual dua arah yang kuat untuk berbagai tugas NLP [7]. Dalam penelitian ini digunakan *Multilingual BERT* (MBERT), yaitu model pralatih yang dirancang untuk mendukung banyak bahasa dan telah banyak digunakan dalam tugas multibahasa [8]. Studi Azhar dan Khodra menunjukkan bahwa *fine-tuning* model *Multilingual BERT* mampu meningkatkan kinerja analisis sentimen dan aspek pada teks berbahasa Indonesia, sehingga mendukung pemilihan MBERT untuk kebutuhan klasifikasi sentimen ulasan pelanggan Pawonkoe [9].

Beberapa penelitian terdahulu telah mengkaji pendekatan serupa dengan konteks dan tujuan yang berbeda. Penelitian pertama memanfaatkan XGBOOST untuk peramalan penjualan *e-commerce* sekaligus mendukung strategi pemasaran berbasis data [10]. Penelitian lainnya mengusulkan kerangka dua tahap analisis sentimen berbasis BERT atau Bi-GRU yang kemudian diintegrasikan ke dalam model XGBOOST untuk prediksi kepuasan pelanggan [11]. Studi lain menunjukkan bahwa penggabungan representasi *Transformer* dengan model *machine learning* seperti XGBOOST mampu menghasilkan performa tinggi pada berbagai tugas klasifikasi teks, sehingga pendekatan *hybrid Transformer–XGBOOST* dinilai layak untuk pengolahan data tekstual [12], [13]

Meskipun demikian, penelitian-penelitian tersebut umumnya masih berfokus secara parsial dan belum secara spesifik diarahkan pada konteks operasional UMKM kuliner lokal. Selain itu, belum ditemukan penelitian yang mengintegrasikan prediksi penjualan berbasis XGBOOST dan analisis sentimen ulasan pelanggan berbahasa Indonesia berbasis MBERT dengan memanfaatkan data historis dan data ulasan pada periode yang sama, serta mengimplementasikan hasil analisis tersebut ke dalam satu sistem berbasis web yang interaktif dan mudah digunakan oleh pelaku UMKM. Akibatnya, pemanfaatan model prediksi dan analisis sentimen masih cenderung terbatas pada ranah akademik dan belum

sepenuhnya menjawab kebutuhan praktis UMKM dalam pengambilan keputusan bisnis sehari-hari.

Berdasarkan kondisi tersebut, penelitian ini diarahkan untuk menerapkan dan mengevaluasi XGBOOST dalam memprediksi penjualan UMKM Pawonkoe Banyuwangi berbasis data historis tahun 2024, melakukan klasifikasi sentimen ulasan pelanggan menggunakan MBERT ke dalam kategori positif, netral, dan negatif serta menganalisis karakteristik emosi pelanggan, serta mengimplementasikan hasil analisis tersebut ke dalam sistem berbasis web interaktif sebagai media pendukung pengambilan keputusan yang lebih *user-friendly* bagi UMKM

1.2. Rumusan Masalah

Rumusan masalah yang menjadi fokus utama pada penelitian ini merupakan sebagai berikut:

1. Bagaimana performa model XGBOOST dalam memprediksi penjualan produk Pawonkoe berdasarkan data historis periode Januari–Desember 2024 pada skema *1-step supervised*, *rolling autoregressive*, dan *Recursive 7-step Forecasting* yang diukur menggunakan metrik evaluasi prediksi?
2. Bagaimana performa model *Multilingual BERT* (MBERT) dalam mengklasifikasikan ulasan pelanggan produk Pawonkoe ke dalam kategori positif, netral, dan negatif berdasarkan metrik *evaluasi accuracy*, *precision*, *recall*, dan *F1-score* pada data periode Januari–Desember 2024?
3. Bagaimana distribusi dan kecenderungan sentimen pelanggan terhadap produk Pawonkoe berdasarkan hasil klasifikasi MBERT serta analisis visualisasi *word cloud* pada periode Januari–Desember 2024?
4. Bagaimana implementasi sistem berbasis web dalam mengintegrasikan hasil prediksi penjualan dan analisis sentimen secara interaktif sebagai sistem pendukung keputusan berbasis data bagi UMKM Pawonkoe?

1.3. Batasan Masalah

Batasan yang ditetapkan pada masalah yang diamati pada penelitian ini merupakan sebagai berikut:

1. Data yang digunakan data responden pelanggan dan data penjualan UMKM PAWONKOE Banyuwangi
2. Data Penjualan yang digunakan mulai dari 01 Januari 2024 – 31 Desember 2024
3. Data sentimen yang digunakan mulai dari 01 Januari 2024 – 31 Desember 2024
4. Memanfaatkan *software* Python dan streamlit untuk mendukung proses pemodelan.
5. prediksi tren penjualan produk Pawonkoe dalam periode tertentu berdasarkan data historis penjualan.
6. Model yang digunakan untuk memprediksi penjualan merupakan XGBOOST, dan analisis sentimen dengan MBERT.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini merupakan untuk menjawab rumusan masalah serta memberikan arahan yang lebih jelas bagi penelitian ini. Adapun tujuan penelitian ini merupakan sebagai berikut:

1. Membangun dan mengevaluasi model XGBOOST untuk memprediksi penjualan produk Pawonkoe berdasarkan data historis penjualan periode Januari–Desember 2024 menggunakan metrik evaluasi prediksi.
2. Mengevaluasi performa model *Multilingual BERT* (MBERT) dalam mengklasifikasikan ulasan pelanggan produk Pawonkoe ke dalam kategori positif, netral, dan negatif berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* . pada periode Januari–Desember 2024.
3. Mengidentifikasi distribusi dan kecenderungan sentimen pelanggan terhadap produk Pawonkoe berdasarkan hasil klasifikasi MBERT serta analisis visualisasi *word cloud*.
4. Mengimplementasikan sistem berbasis web untuk menampilkan hasil analisis sentimen dan prediksi penjualan produk Pawonkoe secara interaktif dan *user friendly* sebagai media pendukung pengambilan keputusan.

1.5. Manfaat Penelitian

Manfaat penelitian yang dapat diperoleh dari hasil penelitian ini merupakan sebagai berikut:.

1. Bagi Bidang Keilmuan

Penelitian ini memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang data science dan *machine learning*, khususnya dalam penerapan metode XGBOOST untuk pemodelan prediksi data penjualan yang bersifat fluktuatif serta metode *Multilingual BERT* (MBERT) untuk analisis sentimen dan identifikasi karakteristik emosi pelanggan.

Selain itu, penelitian ini memperkaya kajian mengenai integrasi model *machine learning* ke dalam sistem berbasis web menggunakan Streamlit, sehingga dapat menjadi referensi dalam pengembangan sistem pendukung keputusan berbasis data pada sektor UMKM maupun penelitian lanjutan yang relevan.

2. Bagi Pelaku UMKM Pawonkoe

Penelitian ini memberikan manfaat praktis bagi UMKM Pawonkoe melalui penyediaan informasi prediksi penjualan yang lebih terukur dan berbasis data historis periode Januari–Desember 2024, sehingga dapat mendukung perencanaan produksi serta pengelolaan persediaan secara lebih efektif. Selain itu, penelitian ini menyajikan hasil analisis sentimen beserta identifikasi karakteristik emosi pelanggan terhadap produk yang dapat dimanfaatkan sebagai dasar dalam melakukan evaluasi kualitas produk dan penyusunan strategi pemasaran yang lebih tepat sasaran. Selanjutnya, tersedianya sistem berbasis web yang interaktif dan *user friendly* sebagai media visualisasi hasil analisis turut mempermudah proses interpretasi data serta mendukung pengambilan keputusan yang lebih cepat, akurat, dan berbasis data.

3. Bagi Pemerintah dan Lembaga Pembina UMKM

Penelitian ini dapat menjadi referensi dalam mendorong implementasi digitalisasi UMKM melalui pemanfaatan teknologi *machine learning* dan sistem berbasis web. Hasil penelitian ini juga dapat digunakan sebagai bahan pertimbangan dalam penyusunan kebijakan maupun program pendampingan

berbasis analisis data guna meningkatkan daya saing dan keberlanjutan UMKM.

Halamaan sengaja di kosongkan