

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Cabai rawit adalah jenis tanaman hortikultura yang populer di Indonesia karena digunakan sebagai salah satu bumbu masakan [1],[2]. Permintaan ketersediaan cabai rawit meningkat di setiap tahunnya seiring dengan pertumbuhan penduduk dan munculnya industri kuliner yang menggunakan cabai rawit sebagai bahan baku [3]. Cabai sebagai bagian dari *volatile foods* menyumbang inflasi sebesar 0,15% pada Juli 2022 [4]. Inflasi cabai disebabkan karena naiknya harga jual, kenaikan tersebut diakibatkan oleh harga jual yang mengalami fluktuasi [5]. Melalui *platform* Hargapangan.com yang merupakan pusat informasi fluktuasi harga pangan nasional mencatat pada salah satu daerah yaitu Kota Semarang, harga cabai rawit sering naik dan turun setiap bulannya. Pada Maret 2021, terjadi kenaikan harga tertinggi yaitu Rp110.000/kg, hal tersebut disebabkan oleh faktor cuaca dan peningkatan kebutuhan industri kuliner [6],[7]. Pada umumnya faktor-faktor yang memengaruhi kenaikan harga jual cabai diantaranya perubahan iklim, *supply and demand*, politik serta teknologi [8],[11].

Penelitian [1] menyatakan bahwa perubahan pada iklim dan luas panen memiliki hubungan yang berpengaruh terhadap produktivitas tanaman cabai rawit sebesar 85%. Iklim yang ekstrim seperti kekeringan atau banjir dapat berdampak buruk pada lahan, menghambat proses penanaman dan berdampak buruk pada produksi tanaman sehingga tidak mencukupi jumlah permintaan pasar [10]. Pada 2022, BBC Indonesia melaporkan bahwa hasil panen cabai yang semula 200 kg menurun menjadi 8-10kg pada musim hujan [12]. Penelitian [13] menyimpulkan bahwa perlu dilakukan prediksi untuk memperkirakan harga jual cabai rawit berdasarkan keadaan iklim yang terdiri dari variabel suhu, kelembapan, curah hujan, lama penyinaran dan kecepatan angin agar petani dan produsen dapat menyesuaikan pola tanam dan produksi sesuai dengan prediksi harga yang diharapkan, sehingga dapat mengoptimalkan keuntungan.

Prediksi umumnya dibuat secara otomatis melalui penggunaan metode *data mining*, yang merupakan proses ekstraksi informasi dari kumpulan data yang luas dan rumit [10]. *Data mining* dapat dilakukan menggunakan teknik regresi. Teknik regresi merupakan sebuah teknik yang memeriksa korelasi antara variabel yang terikat dengan beberapa variabel bebas. Tujuan utama dari teknik ini adalah untuk mengestimasi nilai  $y$  berdasarkan nilai  $x$  yang telah diberikan [14]. Pemilihan algoritma regresi harus disesuaikan dengan karakteristik serta uji asumsi klasik terhadap data yang digunakan [15]. Data keadaan iklim memiliki beberapa variabel seperti suhu, kelembapan, curah hujan dan kecepatan angin. Variabel-variabel tersebut memiliki tipe data kontinu dan berbentuk desimal. Distribusi data juga memiliki banyak *outliers* dikarenakan data harga jual cabai rawit yang digunakan mengalami fluktuasi. Uji asumsi klasik pada data keadaan iklim yang mencakup variabel suhu, kelembapan, curah hujan, lama penyinaran dan kecepatan angin serta variabel harga jual cabai rawit yang telah diintegrasikan menghasilkan bahwa data tidak berdistribusi normal, terjadi *autokorelasi* serta tidak adanya *linearitas*. Sehingga diperlukan beberapa algoritma regresi yang mampu menangani data dengan karakteristik tersebut.

Beberapa algoritma regresi yang banyak digunakan untuk melakukan prediksi pada data yang *non-linear* antara lain *XGBoost Regression*, *KNN Regression* dan *Random Forest Regression*. *XGBoost Regression* digunakan karena mampu menangani data yang tidak terstruktur serta data *outlier* dengan baik namun komputasi algoritma ini dapat memakan waktu yang cukup lama untuk melatih model, terutama jika digunakan pada dataset yang besar [16]. *KNN Regression* digunakan untuk memprediksi nilai target pada data yang berpola tidak jelas, namun ketika jumlah fitur meningkat, algoritma ini dapat mengalami kesulitan dalam menghitung jarak antara titik-titik data [17]. *Random Forest Regression* digunakan karena dapat menghasilkan prediksi yang akurat dengan mempertimbangkan hubungan *non-linear* antara variabel *input* dan *output*, namun algoritma ini rentan terhadap *overfitting* jika tidak memperhatikan penyetelan parameter yang optimal [18]. Dari ketiga algoritma tersebut memiliki kelebihan dan kekurangan yang berkaitan antara satu sama lain, maka dari itu dilakukan

perbandingan sehingga dapat diketahui algoritma apa yang memiliki performa terbaik dalam memprediksi.

Penelitian [19] menerapkan *Random Forest Regression* pada data yang non-linear untuk memprediksi harga rumah, hasilnya *Random Forest Regression* dapat memprediksi dengan akurasi 81,5% berdasarkan luas tanah, luas bangunan dan lain sebagainya. Menurut penelitian [20] *KNN Regression* memiliki nilai *error* terendah dalam memprediksi kekuatan turbin berdasarkan kecepatan angin, arah angin dan temperatur udara.

Berdasarkan uraian yang telah dipaparkan, penelitian ini membandingkan beberapa algoritma regresi seperti *XGBoost Regression*, *KNN Regression* dan *Random Forest Regression* untuk mendapatkan model yang akurat dalam memprediksi harga jual cabai rawit berdasarkan keadaan iklim. Perbandingan performa algoritma dilakukan dengan menggunakan matriks evaluasi seperti *MAE*, *MAPE* dan *R2-Score* untuk menilai seberapa baik serta seberapa kecil *error* yang dihasilkan model dalam memprediksi harga berdasarkan variabel *input* yang telah ditentukan [21]. Penelitian ini bukan hanya sekedar mendapatkan model yang akurat dalam memprediksi saja, tetapi diharapkan juga dapat mengetahui variabel-variabel iklim yang berpengaruh signifikan terhadap prediksi harga jual cabai rawit.

## 1.2 Rumusan Masalah

Perubahan iklim yang ekstrim dapat berpengaruh terhadap fluktuasi harga jual cabai rawit khususnya di Kota Semarang, hal ini berdampak pada perekonomian petani maupun konsumen sehingga diperlukan model prediksi yang akurat dengan menggunakan beberapa algoritma regresi agar masyarakat dapat mengantisipasi fluktuasi harga jual cabai rawit menggunakan variabel-variabel iklim.

## 1.3 Pertanyaan Penelitian

Pertanyaan pada penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan performa antara *XGBoost Regression*, *KNN Regression* dan *Random Forest Regression* dalam memprediksi harga jual cabai rawit berdasarkan kondisi iklim harian di Kota Semarang?

2. Apa saja variabel-variabel iklim yang memiliki pengaruh signifikan terhadap harga jual cabai rawit di Kota Semarang?

#### **1.4 Batasan Masalah**

Batasan masalah yang ada pada penelitian ini adalah sebagai berikut:

1. Data dan variabel-variabel iklim yang digunakan sesuai dengan data yang tersedia di *website data online* BMKG stasiun Klimatologi Ahmad Yani dan data harga jual cabai rawit diambil dari *website* SiHati.
2. Data yang digunakan merupakan data harian pada tahun 2016-2023
3. Data keadaan iklim di *website data online* BMKG mengambil dari Stasiun Meteorologi Ahmad Yani
4. Jenis cabai rawit yang diteliti yaitu cabai rawit merah

#### **1.5 Tujuan Penelitian**

Tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Membandingkan performa antara *XGBoost Regression*, *KNN Regression* dan *Random Forest Regression* dalam memprediksi harga jual cabe rawit berdasarkan kondisi iklim harian di Kota Semarang
2. Menganalisis variabel-variabel iklim yang memiliki pengaruh signifikan terhadap harga jual cabe rawit di Kota Semarang.

#### **1.6 Manfaat Penelitian**

Manfaat yang dihasilkan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma regresi untuk memprediksi harga jual berdasarkan keadaan iklim harian
2. Memberikan rekomendasi mengenai model regresi yang tepat dan akurat dalam memprediksi harga jual cabai rawit berdasarkan keadaan iklim.