

## BAB I PENDAHULUAN

### I.1 Latar Belakang

Cuaca merupakan hal penting yang tidak akan pernah lepas dari kehidupan manusia. Pada aktivitas manusia disuatu daerah dapat dipengaruhi oleh kondisi cuaca, seperti bidang pertanian dan perkebunan kelapa sawit salah satu kota di Indonesia tepatnya di provinsi Jambi, Jambi kota yang mayoritas lahan dipakai untuk perkebunan sawit. Informasi prediksi presipitasi yang akurat memungkinkan perencanaan yang lebih baik untuk irigasi tanaman, mitigasi banjir, manajemen krisis kekeringan, dan pengendalian polusi air.

Untuk menentukan suatu prediksi diperlukan metode yang menghasilkan suatu informasi akurat serta data yang *input* harus sesuai dengan kebutuhan, penulis mengambil data harian cuaca secara online dari BMKG Indonesia, suatu lembaga Pemerintah Non Departemen (LPND), dipimpin oleh seorang Kepala Badan yang mempunyai tugas pemerintahan di bidang Meteorologi, Klimatologi, Kualitas Udara dan Geofisika sesuai dengan ketentuan perundang-undangan yang berlaku.

Presipitasi dalam siklus hidrologi, terdapat fase presipitasi yang terjadi setelah kondensasi. Presipitasi adalah proses jatuhnya segala materi yang dicurahkan dari atmosfer ke permukaan bumi dalam bentuk cair (hujan) maupun padat (salju)[1]. Dengan memahami latar belakang tentang presipitasi, dapat mengembangkan model yang lebih akurat untuk memprediksi pola presipitasi di masa depan, yang dapat membantu dalam mitigasi risiko terkait dengan perubahan iklim dan pengelolaan sumber daya alam secara berkelanjutan. Cuaca adalah keadaan udara di atmosfer pada waktu dan tempat tertentu yang sifatnya tidak menentu dan berubah-ubah. Penilaian terhadap kategori cuaca umumnya dinyatakan dengan memperhatikan kondisi hujan, suhu udara, jumlah tutupan awan, penguapan, kelembapan, dan kecepatan angin di suatu tempat dari hari ke hari. Analisis cuaca menggunakan kurun waktu antara satu hari sampai satu minggu [2]. Jalannya aktivitas manusia dapat dipengaruhi oleh kondisi cuaca, seperti bidang transportasi, bidang pertanian, bidang perkebunan, bidang bangunan atau bahkan

bidang olahraga sekalipun. Besarnya pengaruh yang ditimbulkan oleh cuaca mendorong perkembangan sistem yang membantu menentukan kondisi cuaca hujan atau tidak hujan pada suatu daerah.

*Data mining* adalah sebuah proses dengan menggunakan teknik statistik, matematika, kecerdasan buatan hingga *machine learning* yang digunakan untuk melakukan ekstraksi serta mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database [3]. Teknologi *machine learning* (ML) adalah mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya [4].

Metode algoritma *Naïve Bayes Classifier* (NBC) atau yang sering disebut klasifikasi *Bayesian*, adalah suatu metode untuk mengatasi masalah dengan mencari nilai peluang. *Naïve Bayes* adalah klasifikasi menggunakan metode statistik dan probabilistik yang diusulkan oleh ilmuwan Inggris *Thomas Bayes* [5]. Fitur terpenting dari metode *Naïve Bayes Classifier* adalah hipotesis yang kuat dengan independensi dari peristiwa atau kondisi, keuntungan menggunakan metode ini hanya membutuhkan jumlah data pelatihan, yaitu data pelatihan cukup kecil untuk dapat menentukan perkiraan dan parameter yang diperlukan selama klasifikasi [6].

Metode *K-Nearest Neighbor* (KNN) merupakan algoritma klasifikasi yang paling sederhana dalam mengklasifikasikan sebuah gambar kedalam sebuah label. Metode ini mudah dipahami dibandingkan metode lain karena mengklasifikasikan berdasarkan jarak terdekat dengan objek lain (tetangga)[7].

Metode *Random forest* didasarkan pada teknik pohon keputusan sehingga mampu mengatasi masalah non-linier. Metode ini merupakan metode pohon gabungan. Untuk mengidentifikasi perubahan penjelas yang relevan dengan perubahan respons, *random forest* menghasilkan ukuran tingkat kepentingan (*variable importance*) perubahan penjelas[8].

Berdasarkan penjelasan diatas penulis membuat penelitian memprediksi prakiraan terjadi presipitasi terbaik. Memprediksi prakiraan dapat menggunakan teknik data mining dengan menggunakan klasifikasi yaitu dengan metode *Naïve*

*Bayes, K-Nearest Neighbor, dan Random Forest. Data mining* bekerja dengan melakukan proses penggalian informasi dari data harian BMKG bulan Januari tahun 2015 hingga bulan July tahun 2023 di Stasiun Meteorologi Sultan Thaha Jambi kota, yang diharapkan memiliki dampak baik sebagai informasi atau *knowledge* untuk masyarakat Jambi kota khususnya pertanian dan pekebunan kelapa sawit.

## **I.2 Identifikasi Masalah**

Berdasarkan Latar belakang di atas, dapat diidentifikasi dari permasalahan pada presipitasi pada daerah akan berpengaruh jalannya aktivitas manusia seperti:

- a. Manakah algoritma terbaik untuk memprediksi terjadinya presipitasi dengan menggunakan teknik *data mining* dengan menggunakan metode algoritma *Naïve Bayes, K-Nearest Neighbor, dan Random Forest*.
- b. Bagaimana memprediksi terjadinya presipitasi berdasarkan parameter temperatur, kelembapan, lama penyinaran dan kecepatan angin dengan menggunakan algoritma *Naïve Bayes, K-Nearest Neighbor, dan Random Forest* klasifikasi *Data Mining*.

## **I.3 Pembatasan Masalah**

Pada penelitian ini batasan masalah yang penulis batasi pada bagian- bagian berikut:

- a. Dataset prakiraan terjadi presipitasi didapat secara online dari BMKG Indonesia Stasiun Meteorologi Sultan Thaha – Kota Jambi.
- b. Metode yang digunakan adalah algoritma *Naïve Bayes, K-Nearest Neighbor* dan *Random Forest*.
- c. Parameter yang digunakan untuk memprakirakan terjadi presipitasi, yaitu temperatur, kelembapan, lama penyinaran dan kecepatan angin.
- d. Untuk semua variabel yang digunakan merupakan data hasil rata-rata yang diperoleh langsung dari website [dataonline.bmkg.go.id](http://dataonline.bmkg.go.id).

#### I.4 Tujuan Tugas Akhir

Berdasarkan latar belakang yang dikemukakan di atas, maka tujuan dari penelitian ini adalah sebagai berikut :

- a. Untuk implementasi algoritma pengklasifikasi memprakiraan terjadi presipitasi terbaik antara metode algoritma *Naïve Bayes*, *K-Nearest Neighbor*, dan *Random Forest*.
- b. Membandingkan hasil model pengklasifikasi terjadinya presipitasi dengan menggunakan metode algoritma *Naïve Bayes*, *K-Nearest Neighbor*, dan *Random Forest* melalui perhitungan *Accuracy*, *Presisi*, *Recall* dan *F1-Score*.

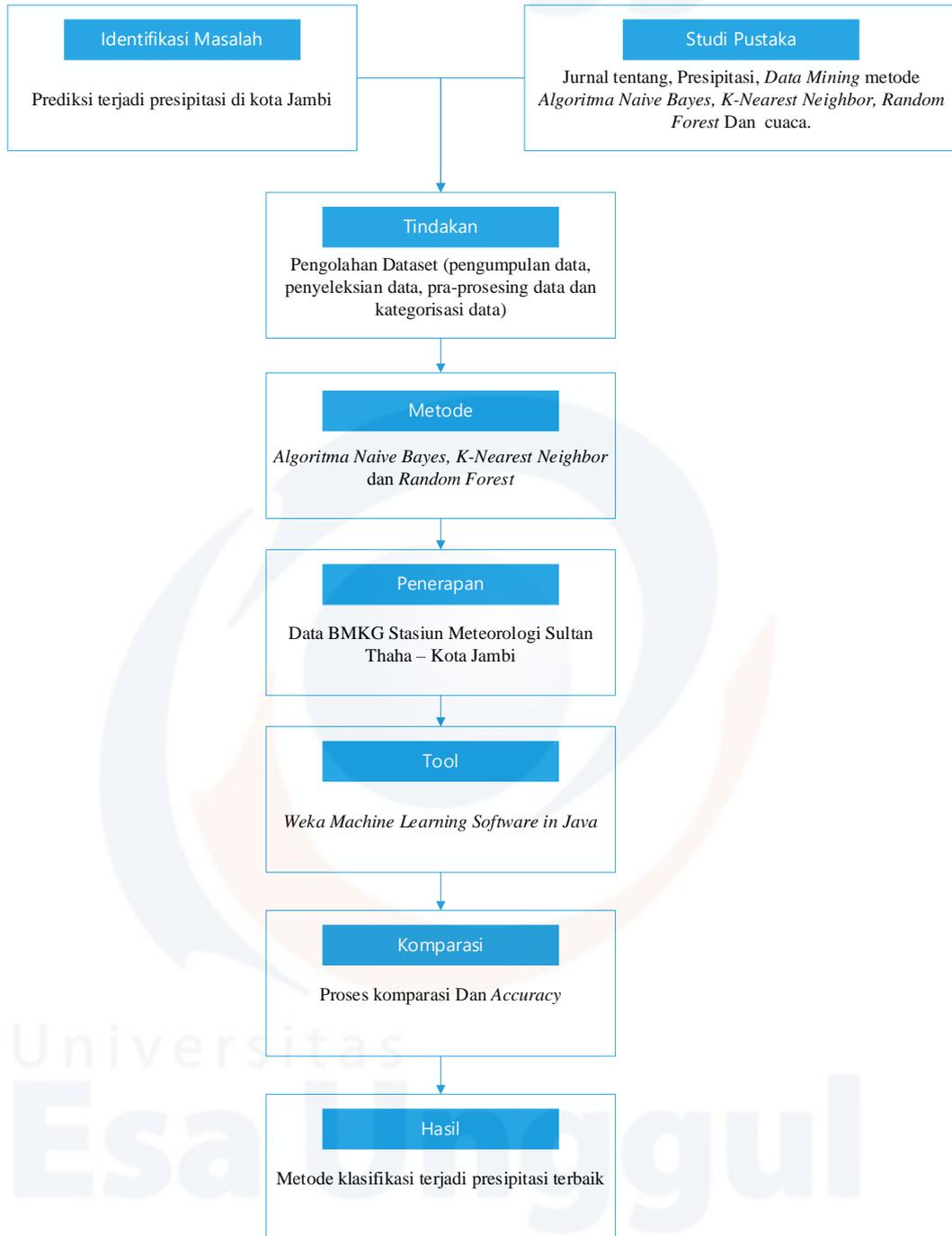
#### I.5 Manfaat Tugas Akhir

Berdasarkan latar belakang yang dikemukakan di atas, manfaat dari penelitian ini adalah sebagai berikut:

- a. Bagi Masyarakat  
Membantu informasi untuk jalannya aktivitas bidang pertanian dan perkebunan kelapa sawit atau yang berhubungan dengan terjadi presipitasi.
- b. Bagi Penulis  
Memahami dan dapat menjelaskan penggunaan metode *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*.
- c. Bagi Akademik  
Menambah sarana referensi atau acuan mengenai permasalahan yang terkait dengan penulisan Tugas Akhir mengenai *Data Mining* dengan metode algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*.
- d. Bagi Pembaca  
Menambah pengetahuan dan referensi khususnya metode klasifikasi algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*.

#### I.6 Kerangka Berfikir

Dalam penelitian ini, tahapan-tahapan kegiatan yang penulis lakukan dengan mengikuti rencana kegiatan yang ada dalam kerangka berpikir penelitian ini, yaitu sebagai berikut:



Gambar I.1 Kerangka Berfikir

Pada tahapan kerangka berpikir di atas dijelaskan dalam beberapa tahap, sebagai berikut:

a. Identifikasi Masalah

Menemukan permasalahan yang terjadi pada presipitasi, daerah yang akan berpengaruh pada pertanian dan perkebunan sawit di Jambi kota.

b. Studi Pustaka

Mencari sebuah karangan ilmiah untuk dijadikannya bahan referensi yang berisikan tentang pendapat dari sejumlah pakar yang membahas suatu masalah yang kemudian akan di analisa dan dibuat kesimpulannya.

c. Tindakan

Subyek penelitian ini adalah curah hujan Stasiun Meteorologi Sultan Thaha – Kota Jambi, tahapan pengolahan data dalam penelitian ini meliputi pengumpulan data, pemilihan data, pembersihan data dan kategorisasi atau label data.

d. Metode

Dengan menggunakan metode klasifikasi algoritma *Naïve Bayes*, *K-Nearest Neighbor*, dan *Random Forest* untuk perhitungan data prakiraan terjadi presipitasi, dapat dihitung dari hasil *class* pada atribut. Kemudian data yang masuk ke sistem dihitung dengan *cross validation*, nantinya hasil dari perhitungan-perhitungan secara terperinci tersebut, dapat menghasilkan suatu nilai akurasi yang akurat.

e. Penerapan

Penerapan data harian dari BMKG Indonesia, Stasiun Meteorologi Sultan Thaha – Kota Jambi dari tahun 2015 hingga tahun 2023 untuk lingkungan masyarakat khususnya pertanian dan perkebunan kelapa sawit.

f. Tool

*Software* WEKA merupakan *tool Data Mining* berbasis *Java* yang sangat populer dan banyak digunakan oleh akademisi dan praktisi untuk kepentingan penelitian.

g. Komparasi

Komparasi dari hasil proses *tool* WEKA menggunakan metode klasifikasi algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest* dengan validasi nilai tertinggi.

h. Hasil klasifikasi terbaik

Hasil prediksi prakiraan terjadi presipitasi hujan dan tidak hujan terbaik akan didapatkan setelah komparasi menggunakan metode klasifikasi algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*, yang

diharapkan memiliki dampak baik sebagai informasi atau *knowledge* untuk masyarakat khususnya pertanian dan perkebunan kelapa sawit.

### **I.7 Sistematika Penulisan**

Adapun secara sistematika penulisan Tugas Akhir ini dibagi dalam 5 bab, adapun uraian tiap masing-masing bab adalah sebagai berikut:

a. BAB I Pendahuluan

Bab ini menjelaskan secara singkat mengenai latar belakang penulisan, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, kerangka berpikir, dan sistematika penelitian.

b. BAB II Landasan Teori

Bab ini membahas tentang teori dasar atau landasan yang berkaitan dan mendukung penelitian ini.

c. BAB III Metode Penelitian

Bab ini membahas tentang metode penelitian yang mencakup metode pengumpulan data dan metode analisis data dengan menggunakan metode klasifikasi algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*.

d. BAB IV Hasil dan Pembahasan

Bab ini menguraikan hasil dari penelitian, hasil pembahasan dan komparasi serta hasil pengujian terhadap dataset yang digunakan.

e. BAB V Kesimpulan dan Saran

Bab ini menjelaskan kesimpulan dari penyusunan Tugas Akhir ini dan saran-saran yang diharapkan sebagai acuan dalam penelitian selanjutnya.

### **I.8 Lokasi dan waktu pelaksanaan**

Lokasi untuk pengambilan data penelitian ini di Stasiun Meteorologi Sultan Thaha – Kota Jambi, Profinsi Jambi, Indonesia. Waktu pelaksanaan penelitian ini Februari 2023 hingga selesai.