

# **BAB I PENDAHULUAN**

## **I.1 Latar Belakang**

Peningkatan populasi perkotaan yang terus bertumbuh lebih padat atau lebih besar yang mencapai 20 juta penduduk di beberapa kota besar membuat kota-kota sangat rentan terhadap bahaya perubahan iklim. Selain meningkatnya populasi perkotaan, perubahan iklim juga disebabkan oleh adanya emisi gas rumah kaca yang menyebabkan peningkatan panas atmosfer. Pemanasan atmosfer menyebabkan meningkatnya degradasi hujan, polusi udara dan juga air (Rojas-Downing dkk., n.d.).

Indonesia merupakan wilayah yang hampir seluruhnya memiliki iklim tropis. Mengacu pada variabilitas iklim di Indonesia dari BMKG, suhu wilayah Indonesia memiliki rata-rata sekitar 28°C di wilayah pesisir, 26°C di daerah pedalaman, dan 23°C untuk wilayah dataran tinggi. Dengan adanya perubahan iklim sendiri berdampak pada perubahan anomali suhu udara rata-rata di Indonesia. Menurut (Molle & Larasati, 2021) sebagai daerah yang kebanyakan memiliki iklim tropis Indonesia mengalami variasi suhu yang sempit namun variasi curah hujan yang beragam. Perubahan iklim yang terjadi juga akan berpotensi mengakibatkan perubahan curah hujan. Timbulnya variasi hujan di Indonesia juga dipengaruhi oleh berbagai faktor baik lokal maupun global.

Curah hujan adalah jumlah total presipitasi yang terjadi di suatu daerah tertentu, yang diukur dalam skala harian, mingguan, bulanan, dan tahunan, yang dipengaruhi oleh faktor-faktor tambahan dan merupakan salah satu komponen dari iklim. Di Indonesia, curah hujan memiliki tingkat keberagaman yang signifikan dan merupakan hal yang paling penting dalam kehidupan manusia karena terkait dengan cuaca lain seperti kelembaban, tekanan, suhu, arah dan kecepatan angin (Roqyah dkk., 2023). Menurut (Indra Pratama dkk., 2022) Indonesia sendiri memiliki salah satu daerah yang memiliki curah hujan paling tinggi dengan intensitas curah hujan mencapai 4500mm, daerah tersebut merupakan kota Bogor yang juga dikenal dengan sebutan Kota Hujan. Berdasarkan data yang dikumpulkan oleh BNPB, terjadi sebanyak 2.342 kejadian

bencana di Indonesia pada tahun 2016. Sebanyak 92% dari kejadian tersebut mayoritasnya adalah bencana hidrometeorologi seperti banjir, longsor, dan puting beliung. Secara khusus, terdapat 766 kejadian banjir, 612 kejadian longsor, dan 74 kejadian gabungan banjir dan longsor. Curah hujan yang tinggi di wilayah terdampak menjadi penyebab utama dari banjir dan tanah longsor tersebut. Tragedi-tragedi ini memiliki dampak serius yang tidak boleh diabaikan. Dampak bencana alam yang terjadi menyebabkan korban jiwa, masalah kesehatan, kerusakan rumah, dan penghancuran gedung-gedung umum. Sebanyak 3,05 juta jiwa mengungsi dan menderita, 522 orang kehilangan nyawa atau dinyatakan meninggal dunia, 69.287 rumah mengalami kerusakan, dan 2.311 bangunan fasilitas umum hancur (Setiawan, 2021). Oleh karena itu, diperlukan teknik yang dapat dilakukan untuk melakukan klasifikasi terhadap curah hujan yang dapat digunakan sebagai pencegahan bencana karena intensitas curah hujan yang tinggi. Salah satu metode yang dapat digunakan untuk melakukan analisa dan pengolahan data agar dapat bermanfaat bagi pembaca ialah metode *data mining*.

Menurut (Jackson, 2002) *Data mining* merupakan proses untuk mengidentifikasi dan menghasilkan suatu data yang berguna dan berkolasi yang dapat dimengerti. Proses keseluruhan dari *data mining* sendiri disebut dengan *Knowledge Discovery in Database (KDD)*. KDD memiliki beberapa tahapan proses berupa persiapan, pemilihan, pembersihan dan interpretasi dari hasil proses *data mining*. *Data mining* merupakan proses untuk menemukan korelasi, pola dan tren baru yang berguna dengan melakukan penyaringan sejumlah besar data yang disimpan menggunakan teknologi pengenalan pola serta statistik. *Data mining* merupakan bidang interdisipliner yang menyatukan teknik dari *machine learning*, pengenalan pola, statistik, *database* dan visualisasi untuk mengatasi masalah ekstraksi informasi dari *database* yang besar (Larose, 2005). Sedangkan menurut (Jollyta dkk., 2020) *data mining* merupakan proses untuk menggali informasi dan pengetahuan baru dari suatu data yang berjumlah banyak pada database dengan menggunakan *Artificial Intelligence (AI)*, statistik dan matematika. *Data mining* merupakan teknologi yang dapat menjembatani komunikasi antara data dan kebutuhan pemakai data tersebut.

Dalam penerapannya *data mining* memiliki beberapa metode algoritma yang dapat digunakan diantaranya algoritma random forest dan naïve bayes. random forest merupakan sebuah algoritma klasifikasi yang termasuk dalam kategori *ensemble*. Algoritma ini membangun sebuah hutan (*forest*) yang terdiri dari beberapa pohon keputusan (*decision tree*). Dalam melakukan klasifikasi, random forest akan menggunakan teknik *voting* untuk mengambil keputusan dari seluruh pohon keputusan. Dengan menggunakan banyak pohon keputusan, algoritma random forest dapat mengatasi masalah yang muncul saat hanya menggunakan satu pohon keputusan dalam klasifikasi. Hal ini dapat meningkatkan nilai akurasi secara keseluruhan dan membuat hasil klasifikasi lebih optimal (Kusumarini dkk., 2021). Sedangkan algoritma naïve bayes merupakan sebuah metode klasifikasi yang didasarkan pada teknik-teknik probabilitas dan statistika yang diciptakan oleh seorang ilmuwan asal Inggris bernama Thomas Bayes. Tujuan dari algoritma ini adalah untuk memprediksi hasil yang akan terjadi di masa depan berdasarkan hasil yang terjadi di masa lalu, sehingga dinamai Teorema Bayes. Teori tersebut kemudian dipadukan dengan kata "naive", yang mengasumsikan situasi di mana setiap atribut memiliki sifat saling bebas (Rifai dkk., 2019).

Dalam penelitian yang berjudul "Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing" oleh (Leonardo dkk., 2020) mengenai prediksi terhadap keputusan klien untuk membantu kinerja telemarketing, mendapatkan hasil perbandingan algoritma naïve bayes dan random forest yaitu akurasi naïve bayes sebesar 85% dan random forest 90. Dengan hasil tersebut dapat disimpulkan bahwa pada penelitian yang dilakukan oleh Leonardo dkk terhadap prediksi keberhasilan klien telemarketing, algoritma random forest lebih baik dalam melakukan klasifikasi tersebut.

Berdasarkan latar belakang yang sudah dijelaskan di atas, pada penelitian kali ini akan melakukan analisa perbandingan terhadap penggunaan metode algoritma random forest dan naïve bayes dengan memanfaatkan data mengenai iklim harian di Indonesia untuk melakukan klasifikasi terhadap curah hujan. Hasil dari penelitian ini bertujuan untuk membandingkan kedua metode algoritma dalam proses pengolahan data dan mencari hasil akurasi terbaik dari masing-masing

algoritma. Selain itu, dari adanya kasus bencana alam akibat curah hujan tinggi dan melihat kondisi intensitas hujan yang tinggi di wilayah Indonesia, diperlukan adanya pengetahuan yang mendukung untuk mencegah bahaya yang ditimbulkan. Data curah hujan yang telah diproses dengan cermat diharapkan dapat menjadi informasi yang memiliki manfaat yang signifikan. Informasi ini akan sangat berarti bagi berbagai pihak, termasuk masyarakat umum yang sering kali mengabaikan informasi penting tentang potensi bahaya bencana alam. Selain itu, terutama bagi organisasi atau lembaga yang terlibat langsung dalam upaya pencegahan bencana alam di Indonesia, seperti Badan Nasional Penanggulangan Bencana, informasi tersebut dapat digunakan sebagai dasar untuk merencanakan tindakan pencegahan atau mitigasi bencana guna mengurangi risiko dan dampak yang mungkin terjadi. Informasi ini dapat menjadi acuan yang berharga dalam mengetahui pola curah hujan dan mengambil langkah-langkah yang diperlukan untuk menghindari bencana akibat curah hujan tinggi di daerah-daerah rawan bencana di Indonesia. Selain melakukan persiapan terhadap bencana yang ditimbulkan, informasi terkait curah hujan juga dapat digunakan untuk berbagai keperluan seperti pada bidang pertanian, transportasi dan industri.

## **I.2 Perumusan Masalah**

Berdasarkan latar belakang di atas, maka permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut :

1. Bagaimana hasil analisa klasifikasi iklim Indonesia menggunakan algoritma random forest?
2. Bagaimana hasil analisa klasifikasi iklim Indonesia menggunakan algoritma naïve bayes?
3. Bagaimana hasil perbandingan analisa klasifikasi iklim Indonesia menggunakan algoritma random forest dan naïve bayes?

## **I.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

1. Mendapatkan hasil analisa dari klasifikasi iklim di Indonesia menggunakan algoritma random forest.

2. Mendapatkan hasil analisa dari klasifikasi iklim di Indoneisa menggunakan algoritma naïve bayes.
3. Mengetahui perbandingan analisa dari klasifikasi iklim di Indoneisa dengan menggunakan algoritma random forest dan naïve bayes.

#### **I.4 Batasan Penelitian**

Adapun batasan dalam melakukan penelitian ini adalah sebagai berikut:

1. Penelitian berfokus pada data yang digunakan yaitu *Climate Data Daily IDN* yang didapatkan dari Kaggle dengan rentang waktu tahun 2020 dengan atribut Tavg, RH\_avg, ss, ddd\_x dan kategori curah hujan (label).
2. Penelitian ini bertujuan untuk melakukan klasifikasi curah hujan berdasarkan data iklim di Indonesia dengan menjadi kategori ringan, sedang, lebat dan sangat lebat.
3. Proses perbandingan klasifikasi menggunakan algoritma random forest dan naïve bayes menggunakan Google Colab.
4. Proses penelitian yang dilakukan hanya sampai tahapan perbandingan analisa algoritma yang didapatkan dalam klasifikasi berdasarkan analisa performa.

#### **I.5 Manfaat Penelitian**

Melalui penelitian yang dilakukan, diharapkan hasil yang didapatkan dapat memberi manfaat teoritis dengan mendapatkan hasil analisa klasifikasi iklim Indonesia bagi mahasiswa dan perbandingan mengenai penerapan algoritma yang lebih baik. Selain itu, bermanfaat bagi peneliti untuk mengimplementasikan serta mengetahui *process mining* pada klasifikasi iklim Indonesia dengan memanfaatkan algoritma yang digunakan yaitu algoritma random forest dan naïve bayes hingga mendapatkan hasil perbandingan dari kedua algoritma tersebut.

#### **I.6 Sistematika Penulisan**

Sistematika penulisan yang akan digunakan dalam penulisan penelitian, diuraikan dalam pembahasan mengenai setiap bab sebagai berikut,

#### **Bab I Pendahuluan**

Pada BAB I Pendahuluan, berisi mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan permasalahan dan sistematika penulisan.

## **Bab II Tinjauan Pustaka**

Pada Bab II Tinjauan Pustaka, berisi mengenai studi literatur, teori yang digunakan dan referensi dari penelitian terdahulu yang berkaitan dengan permasalahan penelitian.

## **Bab III Metodologi Penelitian**

Pada Bab III Metodologi Penelitian, berisi tentang metode yang akan digunakan peneliti untuk melakukan penelitian, pada penelitian kali ini yaitu klasifikasi dengan menggunakan algoritma naïve bayes dan random forest. Selain itu juga berisi tentang sistematika penyelesaian masalah dan metode evaluasi yang akan digunakan.

## **Bab IV Analisis dan Perancangan**

Pada Bab IV Analisis dan Perancangan, berisi tentang analisis yang dilakukan dimulai dari proses pengumpulan data, eksplorasi data, pembersihan data, *labeling* data, *balancing* data, penerapan algoritma untuk klasifikasi pada dataset iklim di Indonesia dan evaluasi yang akan digunakan.

## **Bab V Evaluasi**

Pada Bab V Evaluasi, berisi tentang hasil akurasi dan perbandingan algoritma naïve bayes dan random forest untuk klasifikasi curah hujan berdasarkan dataset iklim di Indonesia.

## **Bab VI Penutup**

Pada Bab VI, berisi tentang hasil evaluasi yang telah dilakukan, lalu dirangkum menjadi kesimpulan dan saran.