

BAB I

PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan istilah untuk menggambarkan jumlah hujan yang jatuh pada suatu wilayah, yang diukur berdasarkan hasil pengamatan tetesan air hujan yang sampai ke permukaan tanah lalu diukur besar volume air hujan per satuan luasnya, satuan jumlah curah hujan adalah inci atau milimeter [1]. Hujan memiliki peran yang sangat penting dalam siklus hidrologi yang mampu mempengaruhi kesetimbangan sumber daya air pada permukaan bumi [2].

Karakteristik iklim di kota Bandung memiliki ciri-ciri musiman tropis basah dengan pola musiman yang konsisten, curah hujan tertinggi terjadi pada kisaran bulan November dan Desember, sementara curah hujan terendah biasanya terjadi pada bulan Juli sampai Agustus [3]. Prakiraan musim hujan ini penting untuk diketahui oleh masyarakat terutama di daerah-daerah yang rawan untuk terjadi banjir sebagai salah satu persiapan dan antisipasi ketika hal tersebut terjadi [4].

Data curah hujan sangat menarik untuk dikaji, karena hujan merupakan salah satu faktor utama yang mempengaruhi iklim di suatu wilayah, selain dapat mengantisipasi terjadinya bencana alam seperti yang disebutkan sebelumnya, juga dapat mempengaruhi di banyak sektor kehidupan manusia seperti sektor pertanian, perkebunan, pelayaran, penerbangan, dan penentuan jenis pola tanam di musim kemarau dan musim hujan [5]. Menurut BMKG curah hujan dapat diklasifikasikan beberapa intensitas berdasarkan satuan curah hujan per hari yaitu 0 mm untuk berawan, 0,5 – 20 mm untuk hujan ringan, 20 – 50 mm untuk hujan sedang, 50 – 100 untuk hujan lebat, 100 – 150 mm untuk hujan sangat lebat, dan lebih dari 150 mm untuk hujan ekstrem [6]. Dengan melakukan pengklasifikasian curah hujan menjadi intensitas hujan tentunya akan lebih mudah untuk dipahami oleh masyarakat secara luas [4].

Pentingnya data kondisi cuaca terutama curah hujan untuk berbagai aspek kehidupan manusia, membuat data kondisi cuaca dicatat setiap harinya pada setiap wilayah di Indonesia termasuk di daerah kota Bandung, yang membuat

data curah hujan menjadi sangat besar [7]. Jumlah data yang sangat besar tidak dapat lagi ditangani dengan menggunakan teknik analisis tradisional karena tidak dapat mengekstrak informasi yang ada pada dataset dengan jumlah yang sangat besar, sehingga dibutuhkan teknologi *data mining* yang merupakan metode dengan menggabungkan teknik tradisional dan teknologi sehingga dataset yang berukuran besar dapat dikelola dengan cepat dan tepat [8].

Besarnya intensitas hujan dipengaruhi oleh beberapa faktor yaitu besar suhu, kelembaban, durasi lamanya penyinaran matahari [9], arah angin, dan kecepatan arah angin [10]. Berdasarkan faktor yang mempengaruhi besar intensitas curah hujan yang terdapat pada *dataset* dapat kita ekstrak suatu informasi berupa pola hubungan antara intensitas faktor cuaca dengan intensitas curah hujan dengan menggunakan metode *association rules*. *Association rules* adalah suatu metode *data mining* untuk menemukan aturan asosiasi terhadap kombinasi antar item pada *dataset*, *association rules* dikenal sebagai metode *data mining* yang merupakan dasar dari beberapa teknik *data mining* lainnya [11].

Association rules termasuk kedalam algoritma unsupervised learning yang mengidentifikasi data historis berdasarkan struktur, kepadatan, segmen serupa, dan fitur serupa lainnya untuk menemukan pola implisit atau pola tersembunyi [12]. Pada data curah hujan dalam satu hari dianggap sebagai satu buah transaksi, dan untuk suhu, kelembaban, kecepatan arah angin, dan durasi sinar matahari masing-masing dianggap sebuah item, pendekatan yang dilakukan adalah untuk mencari kesamaan-kesamaan antar item yang ada pada setiap transaksi, yang nantinya kesamaan antar transaksi tersebut membentuk suatu aturan asosiasi. Metode *association rule* memiliki 3 algoritma yang sering digunakan yaitu algoritma ECLAT, *Apriori*, dan *FP-Growth*, diantara ketiganya algoritma yang dinilai lebih baik adalah algoritma *FP-Growth* [13].

Frequent Pattern Growth (FP-Growth) adalah salah satu algoritma *association rule* yang dapat digunakan untuk menentukan *Frequent Itemset* dalam suatu *dataset*. Algoritma *FP-Growth* merupakan pengembangan dari algoritma *Apriori*. *FP-Growth* menggunakan konsep pohon dalam mencari *Frequent Itemset*. Hal inilah yang menyebabkan algoritma *FP-Growth* lebih

cepat dibandingkan dengan algoritma *Apriori*. Ciri dari algoritma *FP-Growth* adalah struktur data yang digunakan berupa pohon yang disebut *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *Frequent Itemset* dari *FP-Tree*. *FP-tree* adalah struktur penyimpanan data terkompresi. *FP-tree* yang dibangun dengan memetakan setiap data transaksi ke dalam setiap jalur spesifik di *FP-Tree*. Karena pada setiap transaksi yang dipetakan, kemungkinan terdapat transaksi yang memiliki item yang sama, jalur tersebut dimungkinkan untuk saling menimpa [14]. Dalam metode *association rules* selain hasil yang didapat yaitu berupa *Frequent Itemset*, terdapat 2 pendekatan yang dapat digunakan untuk menyaring hasil dari *Frequent Itemset* yang terbentuk yaitu *Closed Frequent Itemset*, dan *Maximal Frequent Itemset* [15].

Dengan jumlah data yang sangat besar dan banyaknya atribut yang digunakan seringkali menghasilkan sejumlah *Frequent Itemset* yang banyak, sehingga mengurangi efisiensi dan keefektifan dari proses *mining*[16]. *Closed Frequent Itemset* digunakan untuk mengurangi banyak *Frequent Itemset* yang terbentuk. *Closed Frequent Itemset* yaitu *Frequent Itemset* dari, I , yang dimana tidak terdapat *superset* I , yang nilai support nya sama dengan I [17]. Selanjutnya *Maximal Frequent Itemset* adalah ketika sebuah *Frequent Itemset* dari, X , bersifat maximal ketika tidak ada lagi *Frequent Itemset* lain yang merupakan *superset* dari *itemset* X [15].

Berdasarkan penelitian sebelumnya [11], terkait penerapan algoritma *apriori* pada simulasi prediksi hujan di wilayah kota Bandung, dapat disimpulkan bahwa hasil yang didapat yaitu sebanyak 19 *frequent item* dengan masing-masing nilai *Confidence* melebihi angka 50% dan *lift ratio* lebih dari 1.00 sehingga dapat dijadikan sebagai acuan dalam memprediksi hujan. Untuk analisis lebih lanjut perlu dilakukan pengklasifikasian intensitas curah hujan berdasarkan beberapa kategori, dan perlu menambahkan jumlah dataset agar dapat memberikan hasil yang lebih baik.

Hasil akhir dari penelitian ini berupa aturan asosiasi *frequent item* yang berisikan pola hubungan intensitas faktor cuaca dengan intensitas curah hujan, dan juga penelitian ini bertujuan untuk membandingkan 2 konsep dari

association rules dalam menyaring hasil *frequent itemset* yaitu *Closed Frequent Itemset*, dan *Maximal Frequent Itemset* untuk menentukan pola curah hujan. Oleh karena itu, dalam penyusunan tugas akhir ini penulis mengambil judul **“Perbandingan *Maximal Frequent Itemset* Dan *Closed Frequent Itemset* Dalam Menemukan Pola Hubungan Faktor-Faktor Cuaca Dan Curah Hujan Di Kota Bandung Menggunakan Algoritma *FP-Growth*”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana pola hubungan faktor-faktor cuaca dan curah hujan di kota Bandung?
2. Bagaimana penerapan dan cara kerja dari algoritma *FP-Growth* dalam menemukan pola hubungan faktor-faktor cuaca dan curah hujan di kota Bandung?
3. Bagaimana analisis perbandingan *itemset* yang terbentuk antara *Closed Frequent Itemset*, dan *Maximal Frequent Itemset* dalam mengurangi jumlah *Frequent Itemset* dari hasil algoritma *FP-Growth*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan diatas, berikut adalah tujuan dari penelitian ini :

1. Untuk mengetahui faktor-faktor cuaca yang dapat mempengaruhi intensitas curah hujan di kota Bandung.
2. Untuk mengetahui cara penerapan dan cara kerja algoritma *FP-Growth* dalam menentukan pola antara faktor-faktor cuaca dengan besaran intensitas hujan di kota Bandung.
3. Untuk mengetahui hasil analisis dari perbedaan *itemset* yang terbentuk antara *Closed Frequent Itemset* dan *Maximal Frequent Itemset*.

1.4 Manfaat Penelitian

Penelitian mengenai perancangan *Maximal Frequent Itemset* dan *Closed Frequent Itemset* dalam menemukan pola hubungan cuaca dan curah hujan di

Indonesia Menggunakan Algoritma *FP-Growth* memiliki beberapa manfaat diantaranya :

1. Manfaat dari segi keilmuan

Mengetahui cara implementasi algoritma *FP-Growth* dan implementasi konsep *Maximal* dan *Closed Frequent Itemset* dalam menseleksi hasil yang diperoleh dari algoritma *FP-Growth* tersebut. Selain itu untuk mengetahui hasil dari analisis perbandingan *Maximal* dan *Closed Frequent Itemset* dalam menentukan hasil *asosiasi frequent item*.

2. Manfaat Bagi Penulis

Menambah wawasan dan pengetahuan bagi penulis mengenai salah satu algoritma *association rules* yaitu *FP-Growth* beserta penggunaan konsep *Maximal* dan *Closed Frequent Itemset* dalam menentukan hubungan asosiasi pada dataset.

3. Manfaat Bagi Masyarakat

Dengan dilakukannya penelitian ini, diharapkan masyarakat dapat melakukan prakiraan curah hujan berdasarkan faktor-faktor cuaca seperti suhu udara, arah angin, kelembaban dan lain-lain.

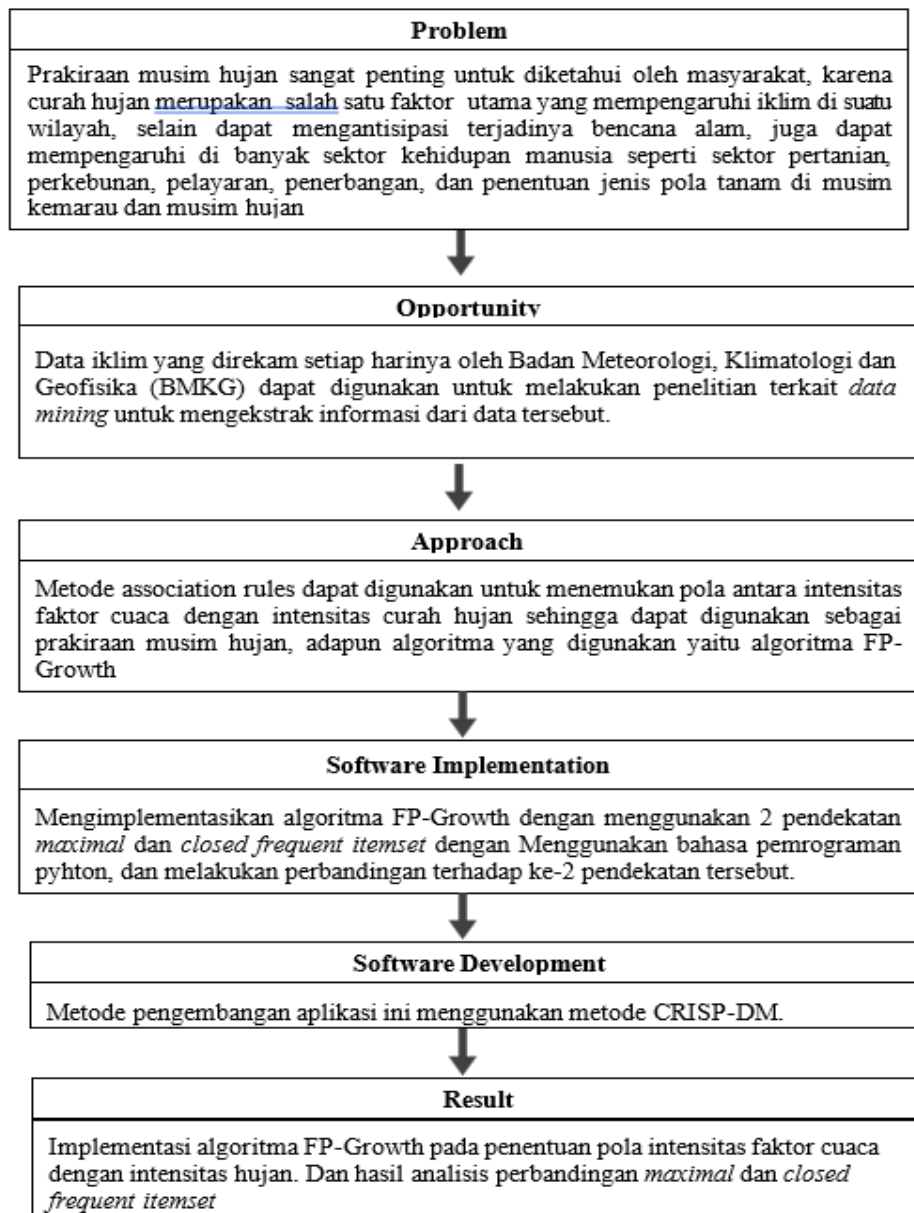
1.5 Batasan Masalah

Supaya penelitian yang dilakukan lebih terarah dan sesuai dengan tujuan yang diharapkan, maka perlu adanya suatu batasan dalam penelitian ini, antara lain :

1. Dataset yang digunakan yaitu data iklim yang diperoleh dari situs resmi BMKG, yaitu data iklim kota Bandung dari bulan januari tahun 2020 sampai dengan bulan juli tahun 2024
2. Atribut–atribut yang digunakan sebagai variabel penelitian yaitu : arah angin pada kecepatan maksimum, arah angin rata-rata, kecepatan angin rata-rata, curah hujan, lamanya penyinaran matahari, kelembaban dan suhu.
3. Bahasa pemrograman yang digunakan yaitu Python, dengan alat bantu yakni VS Code dan Jupyter Notebook.

1.6 Kerangka Pemikiran

Berikut ini merupakan kerangka pemikiran mengenai perbandingan *Maximal Frequent Itemset* dan *Closed Frequent Itemset* dalam menemukan pola hubungan cuaca dan curah hujan di Indonesia menggunakan algoritma *FP-Growth*:



Gambar 1. 1 Kerangka Pemikiran

1.7 Sistematika Penulisan

Laporan tugas akhir ini menggunakan sistematika penulisan yang terdiri dari lima bab yaitu sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini memaparkan terkait latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka pemikiran penelitian dan sistematika penulisan

BAB 2 KAJIAN LITERATUR

Bab ini berisikan tentang penelitian terdahulu yang terkait dengan penelitian yang akan dilaksanakan, sehingga dapat dijadikan sebagai sebuah dasar dan juga rujukan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini memaparkan tentang rangkaian tahapan dan teknik yang digunakan dalam proses implementasi penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil penelitian beserta pembahasan terkait jawaban dari rumusan penelitian berdasarkan tahapan penelitian yang telah dilakukan.

BAB V SIMPULAN DAN SARAN

Bab ini menyajikan kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan serta analisis dari pembahasan, disertai saran dan rekomendasi untuk penelitian selanjutnya.