

BAB I PENDAHULUAN

1.1 LATAR BELAKANG

Ancaman lingkungan terbesar terhadap kesehatan manusia yang sejalan dengan perubahan iklim yang saat ini sedang terjadi yaitu polusi udara. Paparan polusi udara setiap tahunnya diperkirakan menyebabkan tujuh juta kematian prematur dan mengakibatkan hilangnya jutaan tahun hidup sehat. [1] Permasalahan peningkatan polusi udara ini juga terjadi di wilayah Indonesia termasuk di Provinsi Jambi. Peningkatan polusi udara di Provinsi Jambi dipicu oleh berbagai hal. Salah satu penyumbang terbesar polusi udara di Provinsi Jambi yaitu berasal dari asap yang terjadi akibat kebakaran hutan dan lahan (Karhutla).

Karhutla merupakan jenis bencana alam yang terjadi berulang setiap tahunnya di Provinsi Jambi. Penyebab utama karhutla di Provinsi Jambi yaitu adanya kegiatan pembakaran hutan dan lahan yang dilakukan oleh manusia untuk membuka lahan perkebunan. Keadaan cuaca yang kering pada musim kemarau memudahkan hutan dan lahan untuk terbakar karena memiliki kandungan air yang menipis. Karhutla yang terjadi di Provinsi Jambi pada tahun 2015 menyebabkan 19 orang meninggal dan 500 ribu jiwa mengalami infeksi saluran pernapasan akut (ISPA). [2]

Polusi udara memiliki partikel-partikel debu didalamnya, salah satunya *Particulate Matter (PM_{2,5})*. *United States of Environmental Protection Agency (USEPA)* menetapkan istilah PM_{2,5} untuk partikel (PM) di udara yang berdiameter kurang dari 2,5 μm . [3] Konsentrasi PM_{2,5} di udara dipengaruhi oleh kondisi

meteorologi suatu daerah, serta kejadian di sekitarnya baik itu kejadian alami ataupun kejadian yang disebabkan oleh aktivitas manusia. Pengamatan konsentrasi PM_{2,5} saat ini menjadi hal yang penting mengingat banyaknya polutan di udara yang kerap mengganggu kesehatan manusia serta menimbulkan kerugian baik secara materi ataupun kerugian terhadap kerusakan lingkungan.

Upaya untuk mengurangi dampak yang muncul akibat tingginya konsentrasi PM_{2,5} baik di bidang kesehatan maupun ekonomi menjadi suatu langkah penting, khususnya untuk daerah Kota Jambi yang memiliki jumlah penduduk terbanyak di Provinsi Jambi yaitu sebanyak 635,1 juta jiwa menurut BPS pada tahun 2024. [4] Prediksi konsentrasi PM_{2,5} dibutuhkan untuk membantu mengurangi dampak bagi pihak-pihak terkait. Penerapan *Data Mining* dari data konsentrasi PM_{2,5} di Kota Jambi dapat menjadi suatu metode untuk membuat prediksi konsentrasi PM_{2,5} di Kota Jambi.

Data mining merupakan studi atau teknik pengumpulan, pembersihan, pemrosesan dan analisis data yang dilakukan untuk mendapatkan pola serta informasi yang lebih berguna dari sekumpulan besar data. [5] Dalam konteks prediksi konsentrasi PM_{2,5}, data mining dapat digunakan untuk menganalisis data historis PM_{2,5} dan faktor-faktor yang mempengaruhinya, sehingga dapat memberikan gambaran yang lebih jelas tentang tren dan pola yang ada. Prediksi adalah memprakirakan apa yang akan terjadi di masa yang akan datang, sehingga hasil dari prediksi erat kaitannya untuk dapat menjadi sebuah acuan dalam mengambil sebuah keputusan. [6] Terdapat beberapa metode yang biasa digunakan

dalam data mining untuk tujuan prediksi seperti Regresi Linier Berganda dan Random Forest Regression.

Metode Regresi Linier Berganda merupakan metode statistik yang menggambarkan hubungan antara variabel independen dengan variabel dependen dengan sebuah persamaan yang memiliki koefisien regresi untuk masing-masing variabel dan sejumlah variabel independen. [7] Beberapa penelitian dengan tujuan prediksi yang menggunakan metode Regresi Linier Berganda sudah dilakukan. Penelitian oleh Maulita (2023) untuk analisa curah hujan menggunakan metode Regresi Linier Berganda sebagai metode prediksi menunjukkan hasil perhitungan kesalahan sebesar 0,009 karena kesalahan yang terjadi berkisar antara 0,0-0,1 sehingga nilai kesalahan tergolong kecil. [8]

Penelitian dengan tujuan prediksi dengan cara lain juga telah dilakukan, yaitu dengan menggunakan metode Random Forest Regression. Random Forest dikenalkan oleh Leo Breiman pada tahun 2001. [9] Random Forest Regression melakukan prediksi berdasarkan rata-rata prediksi dari nilai regresi yang dibuat oleh setiap pohon dalam susunan pohon keputusan atau *decicion tree* dengan menggunakan beberapa data sebagai masukan. [10] Penelitian oleh Jong Hong dkk. (2020) juga dilakukan untuk mencari fitur prediktor harga rumah menggunakan metode Random Forest dan regresi OLS (*Ordinary Least Squares*) dengan menggunakan data dari apartemen di daerah ramai di Korea Selatan dengan kurun waktu tahun 2006 hingga 2017. Metode Random Forest Regression lebih mampu menangkap kompleksitas atau non-linieritas pasar perumahan aktual sehingga

dinilai dapat menjadi pelengkap untuk model regresi OLS apabila diterapkan pada data yang besar.

Beberapa penelitian yang membandingkan masing-masing metode dengan metode lainnya untuk tujuan prediksi telah dilakukan. Penelitian yang membandingkan beberapa metode statistik untuk tujuan prediksi dilakukan untuk menemukan kelebihan dan kekurangan dari setiap aspek yang dimiliki masing-masing metode yang dipilih untuk meneliti objek atau variabel tertentu. Perbandingan metode Regresi Linier Berganda, Random Forest dan Cubist Regression untuk estimasi suhu permukaan harian pada penelitian yang dilakukan Phan Thanh Noi, Jan Degener dan Martin Kappas (2017) menyajikan berbagai percobaan untuk menilai akurasi dan kecocokan masing-masing algoritma terhadap beberapa kondisi dinamis dari sejumlah dataset yang digunakan.[11]

Metode Regresi Linier Berganda dan Random Forest juga dibandingkan pada penelitian Tingyuan Zhang (2023) untuk menilai seberapa mudah masing-masing model diaplikasikan dan bagaimana efisiensinya dalam hal memprediksi keuntungan penggunaan asuransi untuk beberapa golongan tertentu. Hasilnya menunjukkan metode Regresi Linier Berganda dapat menunjukkan hubungan antara masing-masing variabel independen dengan variabel dependen dan untuk metode Random Forest Regression memiliki nilai akurasi yang tinggi pada hasil prediksi. [12] Masing-masing kelebihan dari kedua metode ini dapat diuji pada model dataset lain dengan berbagai macam parameter, salah satunya pada konsentrasi PM_{2,5}.

Pentingnya mengetahui kecenderungan konsentrasi PM_{2,5} pada kondisi tertentu khususnya untuk wilayah Kota Jambi, mendorong penulis untuk

melakukan penelitian yang bertujuan untuk mengetahui bagaimana metode Regresi Linier Berganda dan metode Random Forest Regression bekerja dalam melakukan prediksi terhadap nilai konsentrasi PM_{2,5}. Menggunakan beberapa parameter cuaca seperti suhu udara, kelembapan udara, kecepatan angin, curah hujan, serta jumlah titik panas pada lingkup Kota Jambi sebagai variabel independen terhadap nilai konsentrasi PM_{2,5} sebagai variabel dependen. Berdasarkan permasalahan tersebut, peneliti melakukan penelitian dengan judul **“Perbandingan Algoritma Regresi Linier Berganda dan Random Forest Regression untuk Prediksi Konsentrasi Particulate Matter 2,5 (Pm_{2,5}) Kota Jambi”**.

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang tersebut dirumuskan beberapa masalah antara lain:

1. Bagaimana perbandingan algoritma Regresi Linier Berganda dan Random Forest Regression dalam memprediksi konsentrasi PM_{2,5} di Kota Jambi?
2. Bagaimana akurasi yang dihasilkan algoritma Regresi Linier Berganda dan Random Forest Regression dalam prediksi konsentrasi PM_{2,5} di Kota Jambi?

1.3 BATASAN MASALAH

Untuk memastikan penelitian ini tidak menyimpang dari apa yang telah dirumuskan maka perlu ditetapkan batasan masalah dalam penelitian antara lain sebagai berikut:

1. Penelitian ini berfokus pada prediksi konsentrasi PM_{2,5} di Kota Jambi.

2. Data yang digunakan adalah data konsentrasi PM2,5 sebagai variabel dependen dan untuk variabel independen menggunakan data suhu udara, kecepatan angin, curah hujan dan titik panas.
3. Data titik panas yang digunakan mencakup titik panas wilayah Kota Jambi dan Kabupaten Muaro Jambi.
4. Prediksi konsentrasi PM2,5 diolah dengan menggunakan metode Regresi Linier Berganda dan Random Forest Regression.
5. *Tools* yang digunakan pada penelitian ini yaitu Rapid Miner.

1.4 TUJUAN PENELITIAN

Tujuan yang ingin dicapai pada penelitian ini antara lain:

1. Menganalisis dan membandingkan algoritma Regresi Linier Berganda dan Random Forest Regression dalam memprediksi konsentrasi PM2,5 di Kota Jambi.
2. Mengevaluasi tingkat akurasi yang dihasilkan algoritma Regresi Linier Berganda dan Random Forest Regression dalam memprediksi konsentrasi PM2,5 di Kota Jambi.

1.5 MANFAAT PENELITIAN

Manfaat yang diharapkan dari dilakukannya penelitian ini antara lain:

1. Mengetahui algoritma yang lebih efektif dalam memprediksi konsentrasi PM2,5 di Kota Jambi.
2. Mengetahui tingkat akurasi yang dihasilkan algoritma Regresi Linier Berganda dan Random Forest Regression dalam memprediksi konsentrasi PM2,5 di Kota Jambi.

1.6 SISTEMATIKA PENULISAN

Adapun sistematika penulisan laporan tesis dibagi menjadi menjadi beberapa bagian sebagai berikut:

BAB I : PENDAHULUAN

Pada Bab ini berisi pendahuluan yang menjelaskan latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI DAN TINJAUAN PUSTAKA

Pada Bab ini dijelaskan mengenai dasar-dasar teori tinjauan pustaka yang dijadikan sebagai referensi penulis dalam melakukan penelitian. Teori yang dibahas mencakup metode algoritma Regresi Linier Berganda dan Random Forest Regression untuk prediksi konsentrasi PM_{2,5} serta teori tentang PM_{2,5}.

BAB III : METODE PENELITIAN

Pada Bab ini dijelaskan mengenai metode-metode yang digunakan dalam penelitian ini meliputi alur penelitian, bahan penelitian dan alat penelitian.

BAB IV : HASIL ANALISIS DAN PEMBAHASAN

Bab ini membahas tentang analisa *dataset* yang digunakan dari data yang dikumpulkan dari Stasiun Meteorologi Kelas I Sultan Thaha Jambi serta hasil analisa dari perbandingan akurasi dari dua algoritma yang digunakan dalam melakukan prediksi konsentrasi PM_{2,5}.

BAB V : PENUTUP

Pada Bab ini berisi kesimpulan dan beberapa saran dari penelitian yang dilakukan untuk mengembangkan penelitian serupa

