

BAB I PENDAHULUAN

1.1 Latar Belakang

Cuaca adalah keadaan fisik udara dan peristiwa perubahan atmosfer yang terjadi dari waktu ke waktu di suatu tempat dan waktu tertentu. Berbagai faktor, seperti suhu rata-rata, kelembaban, kecepatan angin, arah angin, lama penyinaran matahari, curah hujan, dan lain-lain, dapat mempengaruhi perubahan cuaca [1]. Prediksi cuaca menarik untuk diteliti karena perubahan cuaca memiliki peran vital dalam berbagai sektor, seperti industri , sarana penerbangan , pertanian serta kegiatan sehari-hari seperti perencanaan perjalanan, olahraga dan sebagainya [2].

Salah satu cara yang dapat digunakan untuk melakukan prediksi cuaca adalah dengan menggunakan kecerdasan buatan. Konsep kecerdasan buatan pertama kali disebutkan oleh John McCarthy pada *Dartmouth Conference* 1956, yang dihadiri oleh para peneliti kecerdasan buatan. Pada konferensi itu, tujuan utama kecerdasan buatan adalah untuk memahami dan memodelkan proses berpikir manusia, serta untuk membuat algoritma yang dapat digunakan untuk memodelkan proses berpikir manusia yang menjadi salah satu bagian dari *artificial intelligence* yaitu *machine learning* [3].

Pada Penelitian sebelumnya beberapa algoritma *machine learning* telah digunakan untuk prediksi cuaca, namun performanya masih belum optimal. Beberapa algoritma mencapai akurasi yang rendah, seperti *Naïve Bayes* dan *C4.5* dengan metode *Ensemble Learning* yang hanya mencapai akurasi sekitar 49,45% dan 41,24% [4]. Meskipun terdapat algoritma dengan akurasi lebih tinggi, seperti *Deep Learning* dengan akurasi 82.92%, *Generalized Linear Model* dengan akurasi 84.06% [5], dan penerapan algoritma *K-Nearest Neighbor* (KNN) dengan akurasi 89% [6], masih ada ruang untuk peningkatan akurasi prediksi cuaca. Semakin tinggi tingkat akurasi dari suatu algoritma, semakin baik kualitas prediksinya. Akurasi menggambarkan seberapa dekat hasil prediksi dengan nilai sebenarnya. Jika akurasi semakin tinggi, artinya algoritma dapat melakukan prediksi dengan tingkat ketepatan yang lebih tinggi, dan hasil prediksinya lebih mendekati

kenyataan [5]. Selain masalah akurasi, penelitian sebelumnya juga belum sepenuhnya mencakup analisis tentang efisiensi dan relevansi fitur dalam prediksi cuaca. Dataset yang kurang, fitur-fitur cuaca yang tidak relevan atau memiliki tingkat korelasi rendah dengan kelas cuaca yang diinginkan dapat mempengaruhi akurasi prediksi [4][7]. Oleh karena itu, diperlukan pendekatan yang dapat mengidentifikasi fitur-fitur penting dan relevan sehingga dapat mengurangi dimensi fitur yang tidak relevan dan meningkatkan efisiensi prediksi seperti yang disarankan pada penelitian [8].

Hasil penelitian sebelumnya juga menunjukkan bahwa prediksi cuaca menggunakan *Gaussian Naive Bayes* banyak digunakan [9] [4][8], dan memiliki kinerja terbaik dengan prediksi akurasi sebesar 84.153% pada penelitian [8]. Namun, metode *Hyperparameter Tuning* dan *Feature Selection ANOVA-f test* dalam model *Gaussian Naive Bayes* belum banyak digunakan dalam penelitian sebelumnya, disisi lain penggunaan metode tersebut bertujuan untuk mengenali fitur-fitur dataset yang memiliki hubungan yang kuat dengan target yang diinginkan dan otomatis melakukan validasi untuk setiap variasi model, sehingga menghasilkan efisiensi dalam pengujian yang dapat menghemat waktu dalam proses.menghasilkan performa model terbaik yang akan digunakan sebagai model prediksi [10][11].

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan dua pendekatan metode dalam *machine learning* yaitu *Hyperparameter Tuning GridsearchCV* dan *Univariate Feature Selection ANOVA-f test* dalam optimasi model *Gaussian Naive Bayes*. *Hyperparameter tuning GridsearchCV* akan membantu mencari kombinasi parameter yang paling tepat dalam model *Gaussian Naive Bayes* untuk meningkatkan kemampuan dalam memprediksi cuaca dengan lebih akurat. Sementara itu, *Feature Selection ANOVA-f test* akan membantu mengidentifikasi fitur-fitur cuaca yang memiliki korelasi tinggi dengan kelas cuaca yang diinginkan, sehingga dapat mengurangi dimensi fitur yang tidak relevan dan meningkatkan efisiensi prediksi [12].

Hasil optimasi dilakukan melalui beberapa pengujian pada dataset yang dibandingkan untuk mengukur seberapa besar dampak metode optimasi tersebut

terhadap peningkatan performa model *Gaussian Naïve Bayes* dalam memprediksi cuaca harian di Kota Palembang. Akhirnya model yang sudah dioptimasi dapat digunakan untuk dilakukan percobaan prediksi cuaca pada data baru, sehingga dengan melakukan pendekatan ini, diharapkan dapat mengatasi kelemahan dan kekurangan dari penelitian sebelumnya dan menghasilkan prediksi cuaca yang lebih akurat dan efisien. Berdasarkan hal tersebut, maka penulis melakukan sebuah penelitian dengan judul **“OPTIMALISASI MODEL GAUSSIAN NAÏVE BAYES MENGGUNAKAN METODE HYPERPARAMETER TUNING DAN UNIVARIATE FEATURE SELECTION UNTUK MENINGKATKAN AKURASI PREDIKSI CUACA KOTA PALEMBANG”**

1.2 Rumusan Masalah

Dari uraian latar belakang, rumusan permasalahan dalam penelitian tugas akhir ini adalah sebagai berikut:

1. Berapa persen akurasi model *Gaussian Naïve Bayes* apabila dilakukan optimasi *Univariate Feature Selection ANOVA f-test* untuk melakukan pemilihan fitur terbaik, dan *Hyperparameter Tuning GridsearchCV* diterapkan pada teknik pemilihan model terbaik dibandingkan tanpa melakukan optimasi model?
2. Apakah akurasi tetap meningkat jika menggunakan data uji baru untuk dilakukan prediksi cuaca pada model *Gaussian Naïve Bayes* tanpa optimasi dan setelah dioptimasi ?

1.3 Batasan Masalah

Pada tugas akhir ini, penulis membatasi masalah yang dibahas antara lain:

1. Dataset yang digunakan yaitu dataset milik Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang tersedia pada *website* yaitu data aktual cuaca harian pada wilayah Kota Palembang, dengan total data 8.400 data dengan 5 kelas kondisi cuaca diantaranya Cerah Berawan,

Mendung/Hujan Ringan, Hujan Sedang, Hujan Lebat, Hujan Sangat Lebat.

2. Metrik evaluasi hasil performa model yang digunakan adalah *Accuracy*, *Precision*, *Recall*, *F1-Score* pada data latih dan data uji.
3. Setelah model berhasil dioptimasi akan dilakukan percobaan prediksi menggunakan 62 data uji baru yang disediakan BMKG dan membandingkan hasil prediksi penelitian dengan data aktual yang terjadi.

1.4 Tujuan

Adapun tujuan dari pembuatan tugas akhir ini adalah

1. Untuk mendapatkan peningkatan akurasi model *Gaussian Naïve Bayes* menggunakan metode *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV* dengan performa model yang tidak mengalami optimasi dalam prediksi cuaca Kota Palembang.
2. Untuk menguji apakah peningkatan akurasi yang diperoleh melalui optimasi model *Gaussian Naïve Bayes* masih tetap stabil ketika model diterapkan pada data uji baru yang belum pernah digunakan sebelumnya dalam prediksi cuaca Kota Palembang.

1.5 Manfaat

Adapun manfaat dari penulisan tugas akhir ini antara lain yaitu:

1. Memberikan solusi pada model *Gaussian Naïve Bayes* yang lebih baik dengan metode optimasi *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV*, dalam menentukan kemungkinan terjadinya cuaca pada bidang pertanian, nelayan, dan stasiun cuaca di wilayah Kota Palembang.
2. Memberikan kontribusi ilmu dan wawasan untuk akademisi dalam peningkatan performa model *Gaussian Naïve Bayes* yang lebih baik dengan metode optimasi *Univariate Feature Selection ANOVA-f test* dan *Hyperparameter Tuning GridsearchCV* pada kasus prediksi cuaca.

1.6 Metode Penulisan

Untuk mempermudah penulisan dalam penyusunan tugas akhir, maka penulis menggunakan metode – metode sebagai berikut:

1.6.1 Metode Studi Pustaka

Metode ini merupakan metode pengumpulan data studi literatur dan riset penelitian mengenai metode Algoritma *Naive Bayes* pada model *gaussian Naïve bayes* untuk kasus prediksi cuaca yang bersumber dari paper jurnal bereputasi. Metode ini dilakukan untuk membantu Penulis dalam pembuatan Tugas Akhir.

1.6.2 Metode Eksperimen

Metode ini merupakan metode percobaan yang dilakukan Penulis menggunakan metodologi penelitian yang didapatkan berdasarkan hasil dari tahapan studi pustaka.

1.6.3 Metode Observasi

Metode ini merupakan metode pengamatan terhadap permasalahan dan data yang ada sebagai acuan pengambilan informasi. Dalam hal ini, pengambilan dataset dilakukan di tempat penyedia data histori cuaca milik badan meteorologi, klimatologi, dan geofisika (BMKG) yang tersedia pada *website* <https://dataonline.bmkg.go.id> yaitu data aktual cuaca harian pada wilayah Kota Palembang, dengan total data 8.400 dataset.

1.6.4 Metode Wawancara

Metode ini merupakan metode yang dilakukan dengan cara wawancara atau konsultasi dengan dosen pembimbing mengenai tugas akhir penulis.

1.6.5 Sistematika Penulisan

Di dalam pembuatan suatu penulisan laporan akhir, dibutuhkan suatu sistematika penulisan agar pembaca dapat mempermudah dalam memahami dan membaca isi dari tugas akhir ini. Adapun penulisan tugas akhir terdiri dari 4 (empat) bab, dimana masing-masing bab terdiri dari:

BAB I PENDAHULUAN

Pada bab ini, penulis memberikan gambaran secara jelas mengenai latar belakang permasalahan, ruang lingkup masalah, maksud dan tujuan, metodologi penulisan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi tentang landasan teori yang menunjang penelitian ini. Dan penjelasan penelitian-penelitian sebelumnya yang digunakan untuk memprediksi cuaca menggunakan *machine learning* dan optimalisasi model *Gaussian Naive Bayes* untuk prediksi cuaca serta teori-teori yang mendukung penelitian penulis, seperti metode yang digunakan, parameter, dan data.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisikan tentang hasil dan pembahasan, pada bagian hasil menyajikan hasil temuan berdasarkan tujuan yang ingin dicapai. Pada bagian pembahasan memaparkan hasil kegiatan, hasil *pre-processing*, menginterpretasikan penemuan atau kemanfaatan secara logis, mengaitkan dengan sumber rujukan yang relevan, serta memaparkan analisis dengan berbagai percobaan yang dilakukan perbandingan temuan ini sehingga ada kebaruan pada hasil temuan ini pada kasus prediksi cuaca menggunakan *Gaussian Naive Bayes* dan metode peningkatan pada model tersebut.

BAB V PENUTUP

Pada bab ini, penulis memberikan hasil simpulan yang didapatkan untuk menjawab rumusan masalah, dan memberikan gambaran saran untuk penelitian kedepannya sehingga dapat memberikan pengembangan dari model yang telah di lakukan penelitian ini.