

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Curah hujan memiliki peran penting dalam siklus hidrologi dan berdampak signifikan pada berbagai sektor kehidupan, termasuk pertanian, pengelolaan sumber daya air, dan mitigasi bencana hidrometeorologi. Beberapa tahun terakhir, dilakukan pengamatan pada perubahan pola curah hujan di wilayah Indonesia yang menjadi perhatian serius. Dilansir dari Badan Meteorologi Klimatologi dan Geofisika (BMKG), laju perubahan curah hujan tahunan di Indonesia selama 30 tahun terakhir menunjukkan peningkatan tertinggi sebesar 2.784 mm dan penurunan terendah sebesar 750 mm[1]. Perubahan ini menunjukkan adanya variasi yang signifikan dalam pola curah hujan, yang tentunya memengaruhi aktivitas manusia.

Perubahan pola curah hujan juga berdampak pada intensitas bencana alam yang terjadi. Dalam laporan bencana alam yang dikeluarkan oleh BNPB Tahun 2023, mencatatkan beberapa kejadian di beberapa wilayah. Provinsi Jawa Timur sendiri mencatatkan beberapa kejadian bencana alam basah akibat curah hujan tinggi seperti, banjir dilaporkan terdapat 41 kejadian, 5 kejadian untuk tanah longsor, dan 59 kejadian cuaca ekstrem. Selain itu, kebakaran hutan juga mendominasi kejadian di Indonesia, di Pulau Jawa kebakaran hutan terjadi pada pertengahan bulan Juli 2023. Beberapa wilayah Jawa Timur yang terkena kebakaran hutan antara lain Ponorogo, Nganjuk, dan Pasuruan yang disebabkan oleh El Nino dan intensitas curah hujan yang rendah. Curah hujan yang rendah tidak hanya menyebabkan kebakaran hutan, tetapi juga kekeringan. Beberapa wilayah di Jawa Timur yang terdampak kekeringan antara lain Pamekasan, Kab. Situbondo, Kab. Bondowoso, dan Kab. Jember(Data Bencana Di Tingkat Kabupaten/Kota 2023).

Melihat kondisi ini, dibutuhkan sistem klasifikasi curah hujan yang akurat dan tepat untuk memprediksi seberapa besar tingkat curah hujan harian yang akan terjadi sehingga dapat dimanfaatkan untuk memberikan peringatan dini terhadap bencana alam. Selain itu, di sektor pertanian klasifikasi curah hujan juga bisa dimanfaatkan petani untuk menentukan waktu tanam, jenis tanaman, dan strategi pengolahan lahan yang adaptif terhadap cuaca. Penerapan klasifikasi curah hujan memerlukan metode yang mampu

menganalisis dan menguraikan pola data yang memiliki kompleksitas cukup tinggi. Metode umum yang sering digunakan adalah metode pembelajaran mesin atau *machine learning*, sebuah bidang kecerdasan buatan yang bertujuan membantu manusia menganalisis data dan membuat prediksi dari pelatihan data, pola, dan fitur, serta melakukan klasifikasi dan pengujian[3].

Machine learning yang digunakan pada penelitian ini adalah *extreme gradient boosting* (XGBoost). XGBoost merupakan Algoritma menggunakan metodologi ansambel berdasarkan pohon keputusan, di mana beberapa pohon keputusan dihasilkan secara berurutan untuk meningkatkan kinerja model yang komprehensif. Untuk mengoptimalkan kinerja XGBoost. Namun, sering kali akurasi pada klasifikasi curah hujan masih sulit direalisasikan karena kompleksitas data meteorologi yang dimiliki, seperti ketidakseimbangan data antara curah hujan rendah, menengah, dan tinggi. Ketidakseimbangan data ini jelas memengaruhi akurasi model klasifikasi yang sering kali bias terhadap kelas mayoritas. Untuk mengatasi hal ini, diperlukan pemilihan teknik *balancing* yang tepat, seperti dalam penelitian ini, untuk mengatasi ketidakseimbangan data pada kelas target dipilih teknik *oversampling* dari kelompok *synthetic minority oversampling for nominal* (SMOTEN). SMOTEN sebenarnya merupakan varian dari teknik yang cukup terkenal yaitu SMOTE. Teknik ini dipilih karena keandalannya dalam menangani dataset tidak seimbang yang bernilai kategorikal. Bukan hanya SMOTEN, nantinya teknik SMOTE juga akan dimasukkan dalam skema pengujian untuk menganalisis bagaimana kinerja metode ketika menggunakan dataset dari teknik SMOTE dan SMOTEN. Sehingga, penelitian ini akan membandingkan hasil dari teknik - teknik tersebut untuk melihat efektivitasnya dalam menangani ketidakseimbangan data terutama untuk data kategorikal.

Penelitian oleh Wiwaha et al., 2024 [4] yang membahas tentang cara meningkatkan nilai akurasi model XGBoost dalam klasifikasi curah hujan. Parameter yang digunakan dalam klasifikasinya yaitu *rain*, *humidity*, *temperature*, dan *light level* yang seluruh datanya berupa numerik dengan target kategorikal yang memiliki lebih dari 2 kelas. Dalam penelitian tersebut digunakan beberapa teknik *balancing* pada skema pengujian mereka, diantaranya SMOTE, SMOTE-Tomek, *random oversampling*, ADASYN, dan SMOTEN. Hasil terbaik ditunjukkan oleh SMOTEN dan XGBoost yang

mengimplementasikan *hyperparameter tuning* melalui pendekatan *gridsearch* dengan nilai *accuracy* 99.981%, *precision* mencapai nilai maksimum 100%, *recall* 99.943%, dan *F1-Score* 99.971%. Hal ini membuktikan bahwa SMOTEN mampu mengatasi ketidakseimbangan *multiclass* dataset dan dapat mengoptimalkan kinerja dari model XGBoost.

Tahun sebelumnya, Sapari et al., 2023 [5] menggunakan algoritma XGBoost untuk memfasilitasi klasifikasi kualitas udara. Parameter yang digunakan meliputi *Particulate Matter*, *Particulate Matter 2.5*, *Carbon Monoksida*, *Nitrogen Dioksida*, *Ozon*, dan *Nitrogen Dioksida*. Analisis menghasilkan skor *precision* 97%, *recall* 100%, *F1-Score* 98%, dan *accuracy* 98.61%.

Penelitian lain yang dilakukan oleh Yang & Guan, 2022 [6], menjelaskan penerapan metode baru untuk mengklasifikasikan penyakit jantung melalui peningkatan nilai presisi dan akurasi menggunakan teknik SMOTE untuk menangani masalah ketidakseimbangan data. Selain itu, model dibangun menggunakan XGBoost yang dibandingkan dengan lima algoritma lain. Hasilnya SMOTE-XGBoost berhasil memperoleh tingkat akurasi 93,44% dan yang paling tinggi dari kelima algoritma lainnya. Temuan ini menggambarkan keakuratan metodologi penyeimbangan data SMOTE dalam menambah kemampuan prediktif model pembelajaran mesin.

Selain itu, pada 2024 Odiakaose juga dilakukan penelitian tentang efektivitas penggunaan SMOTE untuk meningkatkan nilai *confusion matrix* pada prediksi hipertensi yang hasilnya SMOTE yang dikombinasikan dengan Tomek Link telah berhasil dalam meningkatkan akurasi, presisi, ingatan, dan skor F1 dari berbagai model, yang mencapai metrik kinerja sempurna 1,0.[7].

Berdasarkan penelitian – penelitian sebelumnya, penulis ingin melakukan analisis perbandingan performansi dari SMOTE dan SMOTEN sebagai teknik penyeimbang data. Pemilihan kedua metode ini disesuaikan dengan karakteristik dataset curah hujan harian yang digunakan dimana terdiri dari kombinasi kolom numerik dan kategorikal.

SMOTE merupakan salah satu teknik *oversampling* terpopuler dengan cara kerjanya adalah memperbanyak sampel sintesis pada fitur numerik melalui teknik

interpolasi dari tetangga terdekatnya. Sementara itu, SMOTEN dikembangkan untuk menangani kolom kategorikal sehingga sampel sintesis yang dihasilkan masih representatif dengan mempertahankan keaslian kategori saat mensintesis data. Oleh karena itu, SMOTEN dianggap lebih efektif untuk digunakan pada dataset penelitian ini.

Dengan membandingkan SMOTE dan SMOTEN, penelitian ini bertujuan untuk menguji efektivitas dan keunggulan masing-masing metode dalam konteks dataset yang memiliki karakteristik campuran. Perbandingan ini penting untuk menentukan metode oversampling yang paling tepat guna meningkatkan performa model klasifikasi. Model yang digunakan dalam penelitian ini adalah XGBoost yang akan digunakan untuk klasifikasi curah hujan harian. Data yang digunakan adalah data iklim harian yang diakses melalui *website* resmi Badan Klimatologi dan Geofisika (BMKG). Perbandingan yang digunakan mempertimbangkan faktor-faktor penting seperti *accuracy*, *precision*, *recall*, dan *F1-Score*. Pada tahap akhir dapat diketahui teknik mana yang paling akurat dalam menangani masalah ketidakseimbangan data.

1.2 RUMUSAN MASALAH

Dari uraian latar belakang diatas, adapun rumusan masalah yang dikemukakan pada penelitian ini adalah sebagai berikut:

1. Bagaimana penggunaan teknik *oversampling* SMOTE (*Synthetic Minority Oversampling Technique*) dan SMOTEN (*Synthetic Minority Oversampling Technique for Nominal*) dalam mengatasi ketidakseimbangan pada *multiclassification* dataset?
2. Bagaimana penggunaan teknik *oversampling* SMOTE dan SMOTEN mempengaruhi kinerja model *Extreme Gradient Boosting* (XGBoost) dalam klasifikasi curah hujan harian?
3. Bagaimana perbandingan hasil model *extreme Gradient Boosting* yang menggunakan SMOTE dan SMOTEN pada klasifikasi curah hujan harian?

1.3 TUJUAN PENELITIAN

Merujuk pada rumusan masalah diatas, adapun tujuan utama yang ingin dicapai penulis dari penelitian ini dibedakan menjadi 2 tujuan yaitu, tujuan umum dan tujuan khusus:

1.3.1. Tujuan Umum

Penelitian ini secara umum bertujuan untuk menganalisis kinerja *balancing* teknik dari *synthetic minority oversampling* (SMOTE) melalui pendekatan salah satu variannya, yaitu SMOTEN atau *synthetic minority oversampling for nominal* dalam mengatasi ketidakseimbangan data *multiclass*. Di mana dataset yang digunakan memiliki tidak dua, tetapi lima pelabelan pada variabel targetnya. Selanjutnya penelitian juga ingin meneliti apakah teknik SMOTEN ini dapat meningkatkan kinerja model *extreme Gradient Boosting* (XGBoost) secara signifikan dalam mengklasifikasi curah hujan harian di Provinsi Jawa Timur. Nantinya, penelitian ini akan menganalisis kinerja dari XGBoost menggunakan hyperparameter dalam melakukan klasifikasi.

1.3.2. Tujuan Khusus

Tujuan khusus dalam penelitian ini adalah sebagai berikut:

- a. Mengimplementasikan teknik *balancing* data SMOTE dan SMOTEN dalam menangani masalah ketidakseimbangan data *multiclass* untuk klasifikasi curah hujan harian.
- b. Mengetahui apakah penggunaan teknik SMOTE dan SMOTEN dapat meningkatkan kinerja dari model *extreme Gradient Boosting* (XGBoost) *classifier* secara signifikan.
- c. Membandingkan dan mengevaluasi kinerja model XGBoost menggunakan masing-masing teknik *balancing* dalam klasifikasi curah hujan harian melalui indikator nilai akurasi, presisi, *recall*, F-1 Score, dan *Area Under the Curve*.

1.4 MANFAAT PENELITIAN

Penelitian diharapkan memberikan kontribusi di bidang akademik dengan memperdalam mengenai implementasi SMOTE dan SMOTEN pada kasus dunia nyata dan pengoptimalan model *machine learning*, khususnya *Extreme Gradient Boosting* (XGBoost) dalam klasifikasi curah hujan harian. Diharapkan nantinya hasil dari penelitian ini dapat dijadikan referensi bagi penelitian serupa dimasa depan. Penelitian juga berpotensi mengembangkan teori-teori baru terkait teknik *balancing* data dan optimasi model, memberikan landasan ilmiah untuk penelitian lebih lanjut terkait bidang yang lebih luas lagi.

1.5 BATASAN MASALAH

Bagian dari uraian latar belakang yang telah disajikan, dapat diidentifikasi rumusan masalah yang menjadi fokus pembahasan sebagai berikut:

- a. Penelitian hanya dilakukan untuk melakukan klasifikasi curah hujan harian di Provinsi Jawa Timur. Hasil klasifikasi belum tentu bisa diaplikasikan di daerah lain tanpa penyesuaian.
- b. Data yang digunakan dalam penelitian merupakan data sekunder (data yang diambil secara tidak langsung oleh penulis) yang diakses melalui *website* resmi BMKG (<https://dataonline.bmkg.go.id/>)
- c. Data yang digunakan adalah laporan iklim harian di Stasiun Klimatologi Jatim.
- d. Variabel yang digunakan terbatas pada data yang disediakan *website* resmi BMKG.