

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cuaca di Indonesia khususnya di Provinsi DKI Jakarta sangat tidak menentu dan sulit untuk diprediksi. Cuaca yang sulit diprediksi membuat aktivitas warga khususnya pada Provinsi DKI Jakarta terganggu sehingga diperlukan sebuah ilmu teknologi yang diimplementasikan untuk memprediksi sebuah cuaca yaitu *Machine Learning*. Prediksi cuaca adalah proses dari pengumpulan data berdasarkan kondisi atmosfer seperti suhu dan kelembaban suatu udara. Faktor tersebut diteliti lalu dibandingkan dengan kondisi cuaca pada tanggal, bulan dan tahun sebelumnya untuk mendapatkan prediksi cuaca yang akurat serta sesuai. Berdasarkan faktor tersebut maka diperoleh juga prediksi beberapa jenis cuaca seperti cerah, berawan dan hujan. Prediksi cuaca juga dilakukan untuk membuat kumpulan informasi mengenai kondisi iklim serta unsur didalamnya (Luthfiarta dkk., 2020). Hasil prediksi cuaca di Indonesia diterbitkan melalui Badan Meteorologi Klimatologi dan Geofisika (BMKG) dengan tujuan agar masyarakat dapat mempersiapkan diri untuk menghadapi setiap kondisi cuaca yang akan datang.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya untuk klasifikasi cuaca yang menerapkan metode *Machine Learning* menggunakan metode *Ensemble Learning* serta penerapan metode *resampling* data dengan *oversampling* SMOTE. Salah satunya adalah penelitian yang dilakukan oleh Ghaitsa Amany Mursianto dan rekan-rekannya (2021) yang menggunakan algoritma *Random Forest* dan *XGBoost* serta menerapkan metode *oversampling* SMOTE dalam kasus prediksi hujan pada data WeatherAUS yang diperoleh dari situs <https://www.kaggle.com/>. Penelitian tersebut menghasilkan hasil akurasi prediksi algoritma *Random Forest* tanpa metode *oversampling* SMOTE sebesar 89,54%, *XGBoost* tanpa metode *oversampling* SMOTE sebesar 88,96%, *Random Forest* dengan metode *oversampling* SMOTE 95,59% dan *XGBoost* dengan metode *oversampling* SMOTE sebesar 93,34%. Penelitian lainnya adalah klasifikasi cuaca Provinsi DKI Jakarta tahun 2018 yang datanya diambil melalui situs <https://data.jakarta.go.id/> dengan periode waktu dari bulan Januari hingga

Desember selama tahun 2018. Penelitian tersebut ditulis oleh Faqih Hamami dan Iqbal Ahmad Dahlan (2022) menggunakan algoritma *Random Forest* yang juga mengimplementasikan metode *oversampling* SMOTE. Hasil penelitian menghasilkan nilai akurasi algoritma *Random Forest* menggunakan metode *oversampling* SMOTE sebesar 70% dan akurasi *Random Forest* tanpa metode *oversampling* SMOTE sebesar 69%. Hal ini membuktikan bahwa penerapan metode *Ensemble Learning* dan metode *oversampling* SMOTE dapat digunakan pada kasus klasifikasi cuaca karena menghasilkan nilai akurasi yang baik. Salah satu algoritma dari metode *Ensemble Learning* adalah *Extra Trees*.

Berdasarkan penelitian terdahulu dan masalah yang ditemukan, maka peneliti mengimplementasikan metode *Ensemble Learning* yaitu algoritma *Extra Trees* dan metode *resampling* data menggunakan *oversampling* SMOTE untuk menangani distribusi data yang tidak seimbang pada kelas minoritas. Penelitian ini bertujuan untuk membuktikan apakah penerapan metode *oversampling* SMOTE memiliki pengaruh pada hasil evaluasi model *Machine Learning* yang dibentuk. Oleh karena itu, penelitian ini melakukan klasifikasi cuaca Provinsi DKI Jakarta menggunakan algoritma *Extra Trees* dengan metode *oversampling* SMOTE dan algoritma *Extra Trees* tanpa metode *oversampling* SMOTE.

1.2 Rumusan Masalah

Berdasarkan penjelasan dari latar belakang di atas, maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana hasil evaluasi model yang diperoleh dari penerapan algoritma *Extra Trees* menggunakan metode *oversampling* SMOTE dan algoritma *Extra Trees* tanpa metode *oversampling* SMOTE pada klasifikasi cuaca Provinsi DKI Jakarta?
2. Apakah penerapan metode *resampling* data menggunakan *oversampling* SMOTE memiliki pengaruh pada hasil evaluasi model *Machine Learning* menggunakan algoritma *Extra Trees* untuk klasifikasi cuaca Provinsi DKI Jakarta?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini berdasarkan rumusan permasalahan di atas adalah sebagai berikut:

1. Memperoleh hasil evaluasi model berdasarkan penerapan algoritma *Extra Trees* menggunakan metode *oversampling* SMOTE dan algoritma *Extra Trees* tanpa metode *oversampling* SMOTE pada klasifikasi cuaca Provinsi DKI Jakarta.
2. Mengetahui pengaruh penerapan metode *resampling* data menggunakan *oversampling* SMOTE pada hasil evaluasi model *Machine Learning* menggunakan algoritma *Extra Trees* untuk klasifikasi cuaca Provinsi DKI Jakarta.

1.4 Manfaat Penelitian

Berikut merupakan manfaat penelitian yang ingin dicapai oleh peneliti berdasarkan hasil penelitian yang dilakukan adalah sebagai berikut:

1. Memberikan gambaran mengenai penerapan metode *Machine Learning* untuk klasifikasi cuaca berdasarkan fitur dan target pada data Prakiraan Cuaca Provinsi DKI Jakarta.
2. Mendapatkan wawasan ilmu mengenai hasil evaluasi model yang dihasilkan oleh algoritma *Extra Trees* menggunakan metode *oversampling* SMOTE dan algoritma *Extra Trees* tanpa metode *oversampling* SMOTE pada klasifikasi cuaca Provinsi DKI Jakarta.

1.5 Ruang Lingkup

Sebagai upaya untuk menghindari kebingungan pada saat proses penelitian, maka diberikan ruang lingkup sebagai batasan penelitian sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data Prakiraan Cuaca Provinsi DKI Jakarta yang terdiri dari bulan Januari hingga Oktober tahun 2017 dan bulan Januari hingga Desember tahun 2018 yang diperoleh dari situs <https://data.jakarta.go.id/>.

2. Bahasa pemrograman yang digunakan dalam pembuatan model *Machine Learning* selama penelitian adalah *Python*.
3. Penelitian ini menggunakan algoritma dari metode *Ensemble Learning* yaitu *Extra Trees* untuk membuat model klasifikasi cuaca Provinsi DKI Jakarta.
4. Penelitian ini menggunakan metode *resampling* data untuk menangani distribusi data yang tidak seimbang menggunakan *Synthetic Minority Oversampling Technique* atau SMOTE.
5. Kategori cuaca Provinsi DKI Jakarta yang digunakan pada penelitian ini berjumlah delapan yaitu Cerah, Cerah Berawan, Berawan, Berawan Tebal, Hujan Lokal, Hujan Ringan, Hujan Sedang dan Hujan Petir.
6. Sebagai representasi, data yang digunakan untuk pembuatan model *Machine Learning* diambil berdasarkan jumlah kategori waktu dengan data terbanyak pada setiap kelas atau kategori cuacanya.

1.6 Luaran yang Diharapkan

Luaran yang diharapkan dari penelitian ini adalah untuk mendapatkan hasil evaluasi model dari penerapan algoritma *Extra Trees* menggunakan metode *oversampling* SMOTE dan algoritma *Extra Trees* tanpa metode *oversampling* SMOTE jika diimplementasikan untuk klasifikasi data cuaca Provinsi DKI Jakarta.

Penelitian ini juga diharapkan dapat memberikan informasi mengenai pengaruh penerapan metode *resampling* data menggunakan *oversampling* SMOTE pada hasil evaluasi model untuk klasifikasi cuaca Provinsi DKI Jakarta.

1.7 Sistematika Penulisan

BAB 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang penelitian, rumusan permasalahan, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, luaran yang diharapkan dan sistematika dalam penulisan skripsi.

BAB 2 TINJAUAN PUSTAKA

Bab ini menjelaskan tentang landasan teori yang mendukung dalam pelaksanaan penelitian serta informasi mengenai penelitian terdahulu yang digunakan sebagai referensi.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tentang tahapan yang dilakukan selama melakukan penelitian serta waktu, tempat dan jadwal dalam pelaksanaan penelitian.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari uji coba penelitian yang dilakukan beserta penjelasan mengenai tahapan-tahapan dalam pelaksanaannya.

BAB 5 SIMPULAN DAN SARAN

Bab ini menjelaskan kesimpulan dari semua rangkaian penelitian dan saran yang diberikan peneliti terhadap penelitian selanjutnya.

DAFTAR PUSTAKA

Daftar pustaka berisikan referensi yang digunakan peneliti sebagai sumber dan referensi dalam menulis penelitian ini.