

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Iklm menjadi salah satu isu yang cukup penting khususnya di negara Indonesia yang mempunyai iklim tropis. Iklim sendiri mempunyai dampak yang sangat besar terhadap keberlangsungan hidup umat manusia di beberapa bidang, antara lain pertanian, industri, kesehatan dan lain-lain. Iklim merupakan suatu konsep abstrak dimana iklim merupakan gabungan antara kondisi cuaca sehari-hari dan karakteristik atmosfer yang terjadi pada suatu wilayah tertentu dalam jangka waktu yang lama (Gunadi et al., 2022).

Salah satu elemen dari iklim adalah curah hujan. Bahkan curah hujan juga mempengaruhi cuaca. Curah hujan diukur dengan satuan tinggi milimeter (mm) di atas permukaan horizontal, didefinisikan sebagai jumlah air yang jatuh di permukaan tanah datar dalam jangka waktu tertentu seperti harian, mingguan bulanan serta tahunan (Triangga, 2020).

Isu utama yang menjadi perbincangan di berbagai daerah Indonesia adalah bencana banjir. Jumlah curah hujan yang tinggi dan luapan air sungai masih menjadi penyebab utama banjir di beberapa wilayah di Indonesia. Seperti banjir yang disebabkan oleh meluapnya sungai-sungai di daerah aliran sungai (DAS), banjir terjadi ketika curah hujan tinggi dan sungai tidak mampu menyerap air yang berasal dari air hujan sehingga mengakibatkan banjir yang tentunya menyebabkan kerugian bagi daerah yang terkena bencana tersebut. Salah satu Provinsi di Jawa Timur yang sering hujan adalah Kabupaten Tuban dengan kisaran curah hujan per tahun 1100-1500 mm dan jumlah hari hujan yang berkisar 90-120 hari dalam setahun (Aprilina et al., 2022).

Meskipun curah hujan tidak dapat ditentukan secara pasti, kita dapat memprediksi atau memperkirakannya. Untuk menghasilkan prediksi yang akurat, para ahli klimatologi harus berusaha dan bekerja keras untuk memecahkan masalah yang menjadi tantangan besar (Pradipta, 2020). Curah hujan itu sendiri memengaruhi aktifitas manusia, khususnya dalam bercocok tanam maupun kehidupan sehari-hari. Oleh karena itu, sangat penting untuk mengklasifikasikan

curah hujan agar kita dapat memprediksi jumlah curah hujan yang akan datang, terutama jika kita melakukan prediksi curah hujan setiap hari (dikenal sebagai curah hujan harian) (Gede et al., 2022).

Algoritma *machine learning* dapat diterapkan dalam analisis klasifikasi. Biasanya, teknik *machine learning* digunakan untuk menganalisis, menemukan, dan mengekstraksi data dalam jumlah besar untuk mendapatkan pola dan pengetahuan dari data tersebut (Kamal & Ramdhani, 2023). Proses ekstraksi untuk mengetahui informasi yang berguna dari data yang dikumpulkan dan diolah secara kompleks disebut dengan data mining (Wiranata et al., 2023). Ada banyak algoritma yang dapat digunakan dalam klasifikasi, salah satunya metode yang sering digunakan adalah *Artificial Neural network* (ANN).

*Artificial Neural network* (ANN) merupakan suatu struktur komputasi menyerupai sistem saraf pada manusia, dimana metode ini tidak hanya memiliki kemampuan untuk menyelesaikan masalah yang kompleks, tetapi juga memiliki fitur *self-learning*, yang membuatnya dapat menghasilkan hasil yang lebih baik dengan banyak data (R. Hidayat et al., 2022). *Multilayer Perceptron* (MLP) adalah komponen *feedforward Artificial Neural network* (ANN). MLP terdiri dari setidaknya tiga lapisan node: lapisan *input*, lapisan tersembunyi, dan lapisan *output*. Setiap *neuron*, kecuali node *input*, menggunakan fungsi aktivasi nonlinier (O. V. Putra et al., 2021).

Seiring berjalannya waktu, dalam data nyata banyak ditemui situasi di mana jumlah *instance* adalah satu *class* jauh lebih sedikit dibandingkan jumlah *instance* dalam *class* lain. Situasi ini biasa dikenal sebagai masalah ketidakseimbangan kelas (*Imbalance class*) (Sutoyo & Fadlurrahman, 2020).

Penelitian ini akan menggunakan algoritma *Neural network* dimana hampir semua algoritma klasifikasi termasuk *Neural network* dan yang lainnya menunjukkan performa yang buruk ketika melakukan proses klasifikasi pada data dengan kelas yang tidak seimbang sebab algoritma tersebut tidak memiliki kemampuan untuk menangani masalah ketidak seimbangan kelas. Oleh karena itu, untuk memperoleh kinerja klasifikasi yang lebih baik sangat penting menangani masalah *imbalance class* pada *dataset* (Anawula, 2020).

Pada tahap *preprocessing* data curah hujan di Kabupaten Tuban, diidentifikasi bahwa *dataset* yang digunakan dalam penelitian ini memiliki permasalahan *imbalance class* dimana dari empat katagori yang digunakan yaitu berawan, ringan, sedang dan lebat, katagori lebat memiliki nilai jauh lebih rendah dibandingkan tiga katagori lainnya sehingga diperlukan metode untuk mengatasi permasalahan tersebut. *Synthetic Minority Oversampling Technique* (SMOTE) adalah salah satu metode *resample dataset* yang paling umum digunakan dalam penelitian untuk menyelesaikan masalah *imbalance class* (Sutoyo & Fadlurrahman, 2020).

Beberapa penelitian yang menerapkan teknik SMOTE diantaranya Naomi Nesyana Debatara (2020) mengkombinasi SMOTE dengan metode *Classification And Regression Tree* (CART) untuk analisis Survey Sosial Ekonomi Nasional (SUSENAS) menunjukkan hasil penelitian data tidak seimbang dengan menerapkan teknik SMOTE secara keseluruhan meningkatkan hasil evaluasi model dimana CART tanpa SMOTE memiliki nilai akurasi 91,92%, *sensitivity* 36,36%, *specificity* 98,94% dan AUC 88,14%, sedangkan CART dengan SMOTE memiliki nilai akurasi 80,86%, *sensitivity* 67,05%, *specificity* 94,31% dan AUC 94,35%. Nugroho & Rilvani (2023) pada kasus kebangkrutan perusahaan, mengkombinasi SMOTE dengan metode *Random Forest Classifier* yang menunjukkan hasil pengujian teknik *oversampling* SMOTE pada algoritma *Random Forest Classifier* meningkatkan performa klasifikasi sebesar 7,40%. dimana nilai akurasi algoritma *Random Forest Classifier* tanpa SMOTE yaitu 88,30% dan algoritma *Random Forest Classifier* dengan SMOTE yaitu 95,70%.

Dari hasil uraian diatas, maka penelitian ini akan dilakukan pengujian terhadap pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) dalam mengatasi masalah *imbalance class* pada klasifikasi curah hujan harian di Kabupaten Tuban dalam rentang waktu tahun 2019 sampai dengan 2023. Sehingga diambil judul penelitian “Analisis *Imbalance Multiclass* Pada Klasifikasi Curah Hujan Dengan Metode *Smote Neural network*”.

## 1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, permasalahan yang dapat diidentifikasi pada penelitian ini adalah:

1. Bagaimana statistik deskriptif pada data curah hujan di Kabupaten Tuban?
2. Bagaimana analisis imbalance pada klasifikasi curah hujan di Kabupaten Tuban dengan metode *Neural network*?
3. Bagaimana analisis imbalance pada klasifikasi curah hujan di Kabupaten Tuban dengan metode *Neural network* dengan menerapkan teknik SMOTE?
4. Apa faktor yang paling mempengaruhi curah hujan di Kabupaten Tuban?

## 1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Mengetahui hasil statistik deskriptif pada data curah hujan di Kabupaten Tuban.
2. Mengetahui hasil analisis imbalance pada klasifikasi curah hujan di Kabupaten Tuban dengan metode *Neural network*.
3. Mengetahui hasil analisis imbalance pada klasifikasi curah hujan di Kabupaten Tuban dengan metode *Neural network* dengan menerapkan teknik SMOTE.
4. Mengetahui faktor apa yang paling mempengaruhi curah hujan di Kabupaten Tuban.

## 1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data di diperoleh dari *website* BMKG tahun 2019-2023 di Kabupaten Tuban.
2. Metode analisis yang digunakan untuk mengklasifikasikan curah hujan adalah statistik deksriptif, SMOTE *Neural network*.
3. *Software* yang digunakan adalah Weka.
4. Penelitian ini menggunakan dua jenis cross validation yaitu *holdout cross validation* sebesar 70%:30% dan 80%:20% serta K-fold Cross Validation sebesar K=3, K=5 dan K=10.

## 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

- a. Bagi penulis
  1. Menambah wawasan dan pengetahuan tentang metode *Neural network* dan *Synthetic Minority Oversampling Technique (SMOTE)*
  2. Menambah wawasan dan pengetahuan tentang curah hujan yang dapat dijadikan referensi untuk penelitian selanjutnya.
  3. Menerapkan ilmu yang telah diperoleh selama kuliah.
- b. Bagi pembaca
  1. Memberikan pengetahuan terhadap permasalahan *imbalance class* pada klasifikasi curah hujan.
  2. Memberi pengetahuan mengenai kinerja metode *Neural network* dan *SMOTE* pada klasifikasi curah hujan.
- c. Bagi lembaga
  1. Sebagai pengembangan studi dan pengembangan matematika implemen dan keilmuan selanjutnya, khususnya dalam bidang ilmu Statistika.
  2. Dapat menjadi sarana informasi dan referensi bagi semua pihak yang tertarik tentang klasifikasi curah hujan yang dapat digunakan untuk memberikan masukan pada pemerintahan terkait curah hujan harian guna mencegah bencana alam.

UNUGIRI