

## 1. Pendahuluan

### Latar Belakang

Curah hujan adalah jumlah air yang jatuh akibat gumpalan awan yang telah mengalami proses pengendapan dan dapat diukur di atas permukaan horizontal dalam satuan tinggi milimeter (mm) [1]. BPS menyatakan bahwa jumlah rata-rata curah hujan di Kota Bandung dari tahun 1980 hingga 2020 menunjukkan tren yang meningkat, meskipun berfluktuasi [2]. Dari tahun 2005 hingga 2020, jumlah curah hujan mencapai 3.038,31 mm dengan lonjakan intensitas banjir yang terjadi 220 kali dalam 15 tahun terakhir [2][3]. Pada tahun 2010 jumlah curah hujan mencapai 322,5 mm yang menyebabkan lonjakan intensitas banjir terjadi 25 kali dalam setahun. Pada tahun 2012, jumlah curah hujan mencapai 209,3 mm dan banjir terjadi sebanyak 22 kali. Pada tahun 2016, jumlah curah hujan mencapai 295,8 mm dan menyebabkan banjir terparah di Kota Bandung dalam 20 tahun terakhir. Pada 2018-2020, curah hujan mencapai 589,69 mm dan menyebabkan intensitas banjir kembali terjadi sebanyak 142 kali dalam tiga tahun.

Akibat banjir yang disebabkan oleh tingginya jumlah curah hujan, masyarakat mengalami kesulitan dalam melakukan aktivitas sehari-hari akibat terbatasnya akses jalan kota, kemacetan lalu lintas, serta rusaknya sarana dan prasarana. Untuk mengurangi dampak dan kerusakan tersebut masyarakat berharap akan munculnya sebuah sistem yang mampu memprediksi dan menganalisis pola curah hujan berdasarkan data historis. Salah satu pendekatan yang dapat digunakan untuk menganalisis dan memprediksi adalah data mining [4]. Namun hasil prediksi tersebut tidak 100% akurat karena setiap lokasi, geografi, dan topografi memiliki data meteorologi yang berbeda. Beberapa algoritma seperti ANN, *Random Forest*, SVM, dan *Logistic Regression* telah menyelidiki prediksi curah hujan. Performa dari masing-masing algoritma ini sangat bervariasi, sehingga untuk meningkatkan performa model, dapat dilakukan dengan memvariasikan jumlah sampel yang digunakan atau menggabungkan metode yang berbeda. Prediksi curah hujan terus menjadi tugas yang menantang, oleh karena itu pemilihan metode yang tepat untuk mengklasifikasikan curah hujan sangat penting di suatu daerah.

Klasifikasi curah hujan telah dilakukan dalam penelitian [5] dan [6] dengan menggunakan metode individu *machine learning*. Pada penelitian [5] penerapan algoritma C5.0 menggunakan *k-fold cross-validation* dan memperoleh akurasi tertinggi sebesar 92% pada data yang tidak seimbang, sedangkan menerapkan teknik SMOTE akurasi meningkat sebesar 99%. Penelitian [6] menggunakan algoritma PCA untuk pengolahan data, dan SVM digunakan untuk klasifikasi. Dengan menggunakan nilai parameter  $C = 10000$  dan  $\gamma = 0,5$  maka nilai akurasinya adalah 65,28%. Setiap metode memiliki kelebihan dan kekurangan dalam membangun model klasifikasi, tergantung pada jenis data, jumlah sampel data yang digunakan, dan jumlah atribut yang dipilih. Dengan demikian, untuk meningkatkan kinerja model yang dibangun, metode *hybrid* menggabungkan beberapa model klasifikasi individu untuk menciptakan model baru dengan kinerja lebih tinggi. Dalam beberapa tahun terakhir, kombinasi metode klasifikasi telah banyak digunakan dalam berbagai penelitian, untuk menghasilkan kinerja model yang akurat dalam membuat prediksi. Namun, belum banyak penelitian yang membahas mengenai kombinasi klasifikasi yang digunakan dalam kasus klasifikasi curah hujan. D. Sidik dan T. Sen dalam penelitian [7] telah mengusulkan metode *Stacking* pada data curah hujan dengan menggabungkan dua metode *machine learning* yaitu *Naïve Bayes* dan C4.5. Data yang digunakan adalah data klimatologi harian untuk stasiun OPT Bandung, Bogor, Citeko, dan Jatiwangi dari tahun 2000 hingga 2018, dengan 10 atribut. Pengujian dilakukan dengan menggunakan data yang memiliki 5 kelas, 2 kelas, dan 2 kelas seimbang. Hasil penelitian menunjukkan bahwa metode *stacking* yang diusulkan dapat meningkatkan kinerja suatu model. Nilai akurasi dan f1-score tertinggi terdapat pada dataset Majalengka yaitu 78,25% dan 85,41% dengan menggunakan dua target kelas. Penelitian ini masih memiliki kekurangan karena nilai akurasinya yang masih tergolong rendah.

Berdasarkan penelitian klasifikasi di atas, belum ada penelitian yang menganalisis perbandingan kinerja metode kombinasi *machine learning* yang digunakan pada data klasifikasi curah hujan yang tidak seimbang. Oleh karena itu, pada penelitian ini akan dilakukan analisis perbandingan dua metode *hybrid Voting* dan *Stacking* untuk klasifikasi curah hujan, dengan menggabungkan lima metode *machine learning*, dan menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data.

### Topik dan Batasannya

Topik yang dibahas pada penelitian ini adalah tentang analisis perbandingan kinerja metode *hybrid machine learning* yang diimplementasikan pada data curah hujan yang tidak seimbang di wilayah Kota Bandung. Untuk mendukung pembentukan model *Hybrid Voting* dan *Stacking*, digunakan dengan mengombinasikan lima metode individu *machine learning*. Penelitian ini memiliki batasan yaitu diantaranya: (1) Data yang digunakan berasal dari tahun 2005 hingga 2021; (2) Data menggunakan 11 atribut; (3) Pengukuran kinerja model yang dikembangkan selanjutnya di evaluasi menggunakan perhitungan confusion matrix.

**Tujuan**

Tujuan dari penelitian ini adalah untuk menentukan algoritma hybrid terbaik antara Voting dan Stacking dalam mengembangkan model klasifikasi curah hujan. Tujuan lain yang diharapkan adalah menemukan model klasifikasi dengan menghasilkan performansi model yang lebih baik, sehingga dapat digunakan oleh masyarakat dan pemerintah dalam mengestimasi curah hujan di Kota Bandung.

**Organisasi Penulisan**

Organisasi penulisan pada penelitian ini terdiri dari lima bagian. Setelah bagian pendahuluan, terdapat bagian dua yang menjelaskan tentang studi terkait dengan penelitian ini. Pada bagian tiga menjelaskan tentang metode dan implementasi rancangan sistem. Pada bagian empat menjelaskan tentang hasil pengujian beserta analisisnya, dan bagian lima menjelaskan tentang kesimpulan dari penelitian ini.