

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Banjir merupakan bencana alam yang sering terjadi di negara Indonesia. Berdasarkan data Badan Nasional Penanggulangan Bencana bahwa sejak awal tahun hingga bulan Mei tahun 2020, telah terjadi 1296 bencana alam dengan kejadian bencana alam banjir mencapai 495 kasus. Samarinda sebagai ibukota Provinsi Kalimantan Timur cukup mengawatirkan, di mana bencana banjir ini mengalami peningkatan cukup tinggi. Merujuk pada data dari situs Badan Pusat Statistik Kota Samarinda, terjadi peningkatan jumlah desa atau kelurahan yang terkena dampak banjir pada tahun 2014 sebanyak 24 kali, pada tahun 2018 sebanyak 18 kali dan pada tahun 2020 sebanyak 33 kali (BPS, 2023). Beberapa kecamatan yang terdampak antara lain Palaran, Samarinda Ilir, Samarinda Kota, Sambutan, Samarinda Seberang, Loa Janan Ilir, Sungai Kunjang, Samarinda Ulu, Samarinda Utara, dan Sungai Pinang. yang mengakibatkan lumpuhnya aktifitas. Ada beberapa faktor utama yang mengakibatkan bencana banjir tidak dapat dihindari lagi saat ini, antara lain yaitu cuaca ekstrem, dan kondisi topografis Daerah Aliran Sungai (DAS).

Klasifikasi banjir berdasarkan penyebabnya dapat membantu memperbaiki perkiraan frekuensi banjir, mendukung deteksi serta penafsiran untuk perubahan kejadian dan besaran banjir (Tarasova et al., 2019). Oleh karena itu, perlu diadakan evaluasi perbaikan akurasi dengan metode klasifikasi data mining. Data mining merupakan proses yang dilakukan dengan penggabungan teknik analisis data untuk memperoleh pola penting pada suatu data. Data mining juga dapat digunakan untuk mendapatkan informasi terkini yang berharga dan bermanfaat dalam pengumpulan data yang melibatkan antara manusia dan perangkat komputer, baik melalui cara otomatis ataupun manual (Tarigan et al., 2022)

Data Mining sendiri memiliki beberapa algoritma. Salah satu yang bisa digunakan untuk mengklasifikasi banjir adalah *K-Nearest Neighbor* (KNN). Beberapa penelitian terkait KNN di antaranya klasifikasi pencemaran udara dengan nilai akurasi sebesar 95.78 % (Nurjanah et al., 2020). Dan penelitian klasifikasi penyakit kanker payudara menggunakan metode KNN memperoleh hasil paling akurat sebesar 93% (Atthalla et al., 2018).

Kemudian, penelitian terkait perbandingan algoritma Naïve Bayes dan KNN, dengan hasil penilaian yang membuktikan bahwa algoritma KNN lebih baik dibanding Naïve Bayes dengan tingkat akurasi lebih tinggi, yaitu 88,94% disertai dengan nilai margin error yang lebih kecil, yaitu hanya 11,06% (Zamri, 2022). Oleh karena kelebihan tersebut, algoritma knn dipilih dalam penelitian ini untuk melakukan klasifikasi pada data banjir di Kota Samarinda.

Namun pada penelitian lain yang dilakukan oleh (Tempola et al., 2018) perbandingan klasifikasi antara knn dan *naive bayes* pada penentuan status gunung berapi dengan *k-fold cross validation* diperoleh hasil menggunakan *k-nn* akurasi rata-rata sistem 63,68% dan standar deviasi sebesar 7,47. Sementara itu, ketika menggunakan *naive bayes classifier* akurasi rata-rata sistem adalah 79,71% dan standar deviasi sebesar 3,55%. Oleh itu *naive bayes classifier* lebih baik dalam melakukan klasifikasi dari pada k-nn. Oleh karena itu diperlukan pendekatan seleksi fitur untuk meningkatkan nilai akurasi algoritma tersebut. Salah satu seleksi fitur yang bisa digunakan adalah *Gain Ratio*. *Gain Ratio* sendiri pernah digunakan dalam penelitian terdahulu, diantaranya pada penelitian komparasi tingkat akurasi *information gain* dan *gain ratio* pada metode *K-Nearest Neighbor* terbukti dapat meningkatkan nilai akurasi 70,06% sehingga dapat disimpulkan penggunaan seleksi fitur *gain ratio* dianggap lebih optimal (Fajriyan et al., 2022). Dalam lingkup yang akan datang, *Gain Ratio* dapat menjadi rekomendasi bagi peneliti lain dalam mereduksi data atribut untuk meningkatkan kinerja suatu algoritma (Hasdyna et al., 2020).

Pada penelitian ini menggunakan Dataset tentang banjir kota Samarinda yang mencakup berbagai fitur seperti tanggal, suhu minimal (Tn), suhu maksimal

(Tx), suhu rata-rata (Tavg), kelembaban rata-rata (RH_avg), curah hujan (RR), kecepatan angin (Ss), kecepatan angin pada arah x (ff_x), arah angin pada arah x (ddd_x), kecepatan angin rata-rata (ff_avg), arah angin kendaraan (ddd_car), serta terjadi banjir.

Penelitian ini menggunakan metode eksperimen yang merupakan jenis penelitian kuantitatif untuk mengukur fitur yang mempengaruhi akurasi banjir. Penelitian ini menguji algoritma KNN dan melakukan seleksi fitur dengan menggunakan metode *Gain Ratio*, untuk mengetahui apakah terdapat kenaikan, penurunan atau perubahan nilai akurasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat disimpulkan bahwa rumusan masalah pada penelitian ini adalah:

- a. Fitur-fitur apa yang terdapat pada dataset Banjir Samarinda yang secara signifikan mempengaruhi penggunaan algoritma k-nearest neighbor dalam pemilihan seleksi fitur gain ratio?
- b. Berapa akurasi yang dapat dicapai algoritma k-nearest neighbor dalam mengklasifikasikan data banjir di kota Samarinda dengan menggunakan seleksi fitur gain ratio?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

- a. Mencari fitur-fitur yang secara signifikan mempengaruhi penggunaan algoritma k-nearest neighbor menggunakan seleksi fitur gain ratio yang mempengaruhi data banjir Kota Samarinda.
- b. Mencari nilai persentase perubahan akurasi algoritma kNN setelah di perselisihan menggunakan seleksi fitur gain ratio.

1.4 Batasan Masalah

Agar lingkup permasalahan yang dibahas tidak meluas, maka penulis membatasi penelitian sebagai berikut:

- a. Data yang digunakan dalam penelitian ini adalah dataset Banjir Kota Samarinda yang diambil dari BMKG Kota Samarinda pada tahun 2019 – 2023.
- b. Dataset banjir kota samarinda mempunyai berbagai fitur yang digunakan dalam menunjang penelitian ini adalah :
 - Temperatur Minimum Suhu
 - Temperatur Maksimum Suhu
 - Temperatur Rata-Rata
 - Kelembaban Rata-Rata
 - Curah Hujan
 - Kecepatan Angin Maksimum
 - Kecepatan Angin Saat Kecepatan Maksimum
 - Kecepatan Angin Rata-Rata
 - Arah angin terbanyak.
- c. Atribut yang di gunakan saat seleksi fitur gain ratio ada 3 yaitu temperature minimum, temperature maksimum, kelembaban

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat kepada berbagai pihak, khususnya:

1. Penulis

Untuk mengembangkan wawasan penulis dalam mengimplementasikan metode klasifikasi menggunakan algoritma knn dengan seleksi fitur gain ratio dalam meningkatkan akurasi algoritma k-nearest neighbor pada data banjir di kota samarinda

2. Mahasiswa

Studi ini dapat bermanfaat bagi mahasiswa, khususnya di bidang data mining. Dengan mempelajari metode-metode yang digunakan dalam penelitian ini, mahasiswa dapat mengembangkan kemampuannya dalam menganalisis dan membuat keputusan berbasis data, meningkatkan kemampuannya dalam menggunakan algoritma KNN, dan mengoptimalkan gain ratio untuk dataset banjir kota Samarinda. Selain itu, hasilnya dapat menjadi referensi yang

berguna untuk penelitian lebih lanjut atau proyek penelitian lain di masa mendatang.