

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Bencana merupakan suatu kejadian yang terjadi dalam beberapa bentuk dimana dapat merugikan dan memberikan penderitaan kepada masyarakat dalam kehidupan normal yang dijalani. Bencana yang sangat berbahaya di dunia, salah satunya merupakan banjir. Selain sering terjadi, kejadian ini menyebabkan kerugian dalam skala besar (Naik, 2021). Dalam Laporan Index Risiko Bencana Indonesia (IRBI), pada tahun 2020 di Indonesia, kasus bencana yang mendominasi salah satunya adalah banjir yang mencapai 1.070 kejadian (Wiguna et al., 2020). Dengan memakai *data set* banjir, diharapkan penelitian ini dapat menganalisa data yang ada sehingga dapat membantu dan memberikan peringatan awal untuk potensi bencana banjir, serta mengantisipasi dan upaya pencegahan.

Data mining merupakan suatu teknik untuk menganalisis dan memproses suatu data mentah (Osman & Technology, 2019). Analisis tersebut menghasilkan beberapa *output* untuk mengklasifikasi *pattern* yang sama, *object recognition*, atau memprediksi dan membuat keputusan. Untuk beberapa perusahaan saat ini, *data mining* memiliki peranan penting sebagai teknologi. Hal ini dapat memberikan dorongan perkembangan dalam sektor perusahaan berdasarkan informasi data yang tersedia.

Machine Learning merupakan bagian dari *data mining* yang berperan penting dalam berbagai sektor. Hal ini merupakan kemampuan untuk mempelajari berdasarkan kejadian dalam bentuk data (Naik, 2021). Untuk mengaplikasikan *machine learning*, kita mengumpulkan data yang berhubungan dengan keperluan kita, sehingga data tersebut nantinya akan diproses dengan menggunakan algoritma yang berbeda-beda untuk mendapatkan akurasi berdasarkan data tersebut. Pemakaian *machine learning* dalam sektor penanggulangan bencana banjir dapat memberikan perkembangan dalam mengidentifikasi potensi banjir.

Tingkat curah hujan diukur dengan bantuan alat pengukur hujan. Alat tersebut bertugas untuk mengumpulkan data curah hujan sepanjang tahun (Prakosa et al., 2021). Hasil pengumpulan data curah hujan yang berhasil dikumpulkan, dimanfaatkan oleh Saiesh Naik untuk menganalisis prediksi banjir dengan menggunakan *machine learning* algoritma *Logistic Regression*. Hasil akurasi yang diberikan dengan menggunakan algoritma tersebut adalah sebesar 75% (Naik, 2021). Peneliti lain juga menggunakan *library Naïve Bayes* dengan akurasi sebesar 79.16% dan *Naïve Bayes* menggunakan *Rapidminer* sebesar 98.31% (Triyanto et al., 2021), serta *Decision Tree* sebesar 33.33% (Prakosa et al., 2021).

Data set kejadian curah hujan akan diperoleh dari *Kaggle*. *Kaggle* merupakan suatu website yang menyediakan *data set* yang dapat dimanfaatkan terutama di bidang *data science*. *Data set* yang diambil

dipublikasikan oleh Mukul dan merupakan *data set* curah hujan Kerala dari tahun 1901 sampai 2018 dengan jumlah data sebanyak 118 data. Kerala merupakan negara bagian India yang memiliki tingkat curah hujan yang tinggi. Rata-rata curah hujan di Kerala berkisar sekitar 300 cm dan banjir terparah yang pernah melanda Kerala adalah pada tahun 2018.

Oleh karena itu, peneliti memiliki ketertarikan untuk melakukan penelitian menggunakan algoritma *K-Nearest Neighbor* atau KNN dan *Logistic Regression* untuk dilakukan perbandingan, serta mencari tahu apakah algoritma KNN dan *Logistic Regression* cocok digunakan untuk klasifikasi potensi banjir. Hasil penelitian ini diharapkan dapat memberikan tingkat akurasi yang tinggi untuk klasifikasi potensi banjir dan mengetahui apakah algoritma tersebut cocok untuk digunakan.

1.2 RUMUSAN MASALAH

Berdasarkan masalah yang telah diuraikan pada latar belakang, maka dapat dirumuskan masalah penelitian ini, yakni:

- 1) Apakah algoritma KNN dan *Logistic Regression* cocok untuk digunakan dalam klasifikasi potensi banjir?
- 2) Berapa tingkat akurasi yang diberikan oleh algoritma KNN dan *Logistic Regression* dalam mengklasifikasi potensi banjir?
- 3) Bagaimana hasil perbandingan antara KNN dan *Logistic Regression*?

1.3 TUJUAN PENELITIAN

Tujuan dalam melakukan penelitian ini adalah:

- 1) Untuk mengetahui apakah algoritma KNN dan *Logistic Regression* cocok untuk digunakan dalam klasifikasi potensi banjir melalui *data set* curah hujan
- 2) Untuk mengetahui tingkat akurasi algoritma KNN dan *Logistic Regression* dalam mengklasifikasi potensi banjir
- 3) Untuk mengetahui dan membandingkan hasil menggunakan algoritma KNN dan *Logistic Regression*

1.4 BATASAN MASALAH

Batasan masalah dibuat dengan tujuan agar penelitian tidak dilakukan secara meluas sehingga penelitian dapat memfokuskan dan membahas secara jelas tujuan penelitian yang telah dibuat. Batasan masalah untuk penelitian ini antara lain:

- 1) Prediksi potensi bencana banjir berdasarkan *data set* menggunakan data curah hujan negara Kerala yang bersumber dari *website Kaggle*
- 2) Menggunakan *data set* curah hujan negara Kerala dari tahun 1900 sampai 2018 dalam bentuk *file CSV*
- 3) Teknik klasifikasi data menggunakan metode *K-Nearest Neighbor* atau KNN dan *Logistic Regression*

- 4) Model KNN dan *Logistic Regression* akan dibangun dengan aplikasi *Google Collab* dengan memanfaatkan *Python* dan *library scikit-learn*, serta *matplotlib* untuk memvisualisasi data

1.5 MANFAAT PENELITIAN

Setelah penelitian ini selesai, terdapat manfaat bagi pembaca dan peneliti yang membacanya, yakni:

- 1) Memberikan *insight* dan referensi bagi peneliti yang ingin melakukan penelitian yang sejenis
- 2) Sebagai referensi untuk prediksi potensi bencana banjir berdasarkan data curah hujan
- 3) Menyumbangkan pemikiran dan hasil penelitian kepada peneliti lain mengenai prediksi potensi bencana banjir

1.6 SISTEMATIKA PENULISAN

BAB I PENDAHULUAN

Bab ini membahas masalah yang diangkat yang menjadikan sebagai latar belakang pemilihan topik skripsi, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, serta sistematika penulisan skripsi.

BAB II LANDASAN TEORI

Bab ini membahas mengenai berbagai teori yang berhubungan dengan topik yang dibahas untuk dijadikan sebagai dasar penelitian. Bab ini juga akan

membahas tentang peneliti terdahulu sebagai acuan dan tinjauan dalam penelitian ini.

BAB III METODOLOGI

Bab ini membahas mengenai metode yang digunakan untuk pengumpulan dan penyajian data, hipotesis dari peneliti, tahapan penelitian, serta metode penyelesaian yang digunakan untuk penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memberikan hasil dari pengujian data yang dikumpulkan dan membahas terkait hasil setelah melakukan penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini memberikan kesimpulan dan merangkum semua hasil penelitian yang dilakukan, serta saran bagi peneliti lain agar dapat memperbaiki atau mengembangkan ke arah yang lebih baik.