

BAB I

PENDAHULUAN

1.1. Latar Belakang

Salah satu organ terbesar pada tubuh manusia ialah kulit, terdapat lebih dari 10% massa tubuh manusia serta merupakan organ yang memiliki intensitas paling sering yang digunakan oleh manusia dalam menjalankan kesehariannya dengan berinteraksi dengan lingkungannya. Terdiri dari empat lapisan yaitu stratum korneum (*epidermis nonviable*), sisa lapisan epidermis (*epidermis viable*), dermis, dan jaringan subkutan. Kulit memiliki fungsi yang penting bagi manusia, baik dari segi pelindung serta membantu manusia dalam menjalankan kesehariannya. (Walters, 2002)

Saat ini, masalah penyakit kulit merupakan sesuatu yang tidak terlalu dianggap penting atau utama, jika dibandingkan dengan penyakit lainnya yang menyebabkan kematian yang tinggi, seperti *HIV/AIDS*, pneumonia komunitas, dan tuberkulosis dalam skala global. Namun, dilihat dari permasalahan yang ada dalam layanan fasilitas kesehatan, masalah kulit umumnya merupakan penyakit yang paling sering ditemui. Pertama, penyakit kulit sangat umum terjadi dan pasien datang dalam jumlah besar di layanan kesehatan primer sehingga penyakit kulit bukan merupakan masalah kesehatan yang dianggap remeh, namun perlu juga tindakan yang awas dan pencegahan yang akurat. Dalam kelompok anak-anak termasuk kelompok yang rentan terkena masalah penyakit kulit. Apabila tidak dijaga serta kurangnya fokus terhadap kehygienisan serta kesehatan kulit maka faktor tersebut dapat memicu timbulnya masalah kesehatan kulit. Kurangnya keterampilan dasar dalam pengelolaan penyakit kulit juga merupakan masalah yang semakin membingungkan. (Bhuiya, 2006).

Penyakit kulit menduduki peringkat keempat penyebab penyakit manusia yang paling umum, sehingga menimbulkan beban non-fatal yang sangat besar. Meskipun demikian, banyak orang yang terkena dampak tidak berkonsultasi dengan dokter. Oleh karena itu, beban penyakit kulit sebenarnya mungkin lebih tinggi karena angka prevalensi yang dilaporkan biasanya didasarkan pada data

sekunder yang mengecualikan individu yang tidak mencari perawatan medis. Beberapa penyakit kulit sering kali ditangani oleh dokter, namun demikian diperkirakan lebih dari 70% orang yang terkena dampak tidak melakukan hal ini konsultasikan dengan dokter. Beberapa penyakit dermatologis (misalnya jerawat) sering kali mengobati sendiri dan umumnya memiliki rawat inap yang rendah tarif. Namun, tidak mencari perawatan medis profesional bisa menjadi masalah bagi orang-orang dengan risiko tinggi AK dan KC, seperti pekerja di luar ruangan. Sayangnya, seringkali ini adalah manusia yang jarang berobat. Oleh karena itu, beban sebenarnya dari penyakit kulit mungkin saja terjadi (Tizek et al., 2019).

Berikut beberapa penelitian terdahulu yang melatarbelakangi penelitian ini, diantaranya dengan menggunakan metode *Naive Bayes* yang telah disampaikan oleh M. Furqon, dkk pada tahun 2022 tentang Klasifikasi Penyakit Kulit Menggunakan Algoritma *Naive Bayes* Berdasarkan Tekstur Warna Berbasis Android menunjukkan tingkat akurasi mencapai 75% untuk mengidentifikasi penyakit kulit (Furqan et al., 2022). Kemudian penelitian yang dilakukan oleh Elly Firasari, dkk. Mengenai *Comparison of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithm for the Classification of the Poor in Recipients of Social Assistance* pada tahun 2020 menggunakan algoritma *Naive Bayes Classifier* menghasilkan klasifikasi dengan akurat tingkat akurasi sebesar 89,04% sedangkan K-NN menghasilkan klasifikasi dengan tingkat akurasi sebesar 87,67%. (Firasari et al., 2020)

Selanjutnya, pengujian pernah dilakukan oleh Agista Nindy Yuliarina dan Hendry tentang Perbandingan Analisis Prediksi Kepuasan Pengguna Layanan Gofood Menggunakan Algoritma Knn & Naive Bayes Dengan Software Rapidminer di tahun 2022 menunjukkan hasil bahwa algoritma KNN lebih efektif dalam mengklasifikasikan tingkat kepuasan pengguna terhadap layanan GoFood dibandingkan dengan Naive Bayes (Yuliarina, 2022). Proses optimasi algoritma K-Nearest Neighbour dengan teknik Cross Validation dengan Streamlit studi data penyakit diabetes pernah dilakukan oleh (Prasetyo & Laksana, 2022) dengan metode yang diterapkan mendapatkan hasil okurasi yang optimal menggunakan model algoritma K-Nearest Neighbor menggunakan Teknik cross validation prediksi penyakit diabetes yang memiliki nilai sebesar 95% lebih akurat setelah

menggunakan k-fold cross validation daripada sebelum menggunakan k-fold cross validation bernilai 92% .Hal ini dapat terjadi karena ada penggunaan teknik Cross Validation yang dapat mengurangi *overfitting* selain mempertimbangkan nilai akurasinya.

Mengacu pada beberapa penelitian terdahulu mengimplementasikan metode yang sama dengan penelitian ini, didapat hasil penelitian dengan judul “Analisis Perbandingan Klasifikasi Antara KNN dan *Naive Bayes* Pada Diagnosa Penyakit Kulit menggunakan *K-Fold Cross Validation*”. Menerapkan prinsip *Data Mining* di penelitian ini yang kemudian data akan melalui tahap pra-pemrosesan sebelum penerapan metode. Langkah selanjutnya adalah menghitung dan membandingkan tingkat akurasi data dengan 2 metode diantaranya adalah metode *Naive Bayes* dan *K-Nearest Neighbor*. Kemudian proses validasi data dilakukan menggunakan *K-Fold Cross Validation*. Metode tersebut dipilih dengan dasar metode yang memiliki kecocokan atas permasalahan yang ada yaitu pada diagnosa penyakit kulit. Pernyataan ini dipertegas oleh penelitian sebelumnya yang serupa dalam metode yang digunakan dan implementasinya dalam proses validasi data dengan *k-fold cross validation* yang mana tujuan utamanya bukan hanya untuk mencapai tingkat akurasi yang tinggi dalam prediksi atau klasifikasi, melainkan juga untuk memperoleh hasil yang valid. Sehingga metode yang diterapkan sudah cocok untuk penelitian ini. Penelitian terdahulu juga memiliki beberapa perbedaan dengan penelitian ini, misalnya dataset yang berasal dari studi kasus yang berbeda. Hal ini dapat mengakibatkan perbedaan tingkat akurasi data antara metode yang digunakan.

Dataset yang digunakan dalam penelitian ini terbagi menjadi 6 klasifikasi penyakit kulit. Diantaranya adalah Psoriasis, Dermatitis Seboroik, Dermatitis Kronik, *Lichen Planus*, Pitiriasis Rosea, Pitiriasis Rubra Pilaris. Dengan menggunakan metode *Naive Bayes Classifier* dan *K-Nearest Neighbor* dan didapatkan hasil klasifikasi yang telah divalidasi menggunakan K-Fold Cross Validation, sehingga didapatkan hasil akurasi, serta diambil kesimpulan metode mana yang tingkat akurasinya lebih tinggi.

1.2. Rumusan Masalah

Terdapat beberapa permasalahan mengacu pada latar belakang yang sudah dipaparkan sebelumnya antara lain:

1. Bagaimana hasil perbandingan klasifikasi diagnosa penyakit kulit menggunakan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* setelah dilakukan proses validasi data menggunakan *K-Fold Cross Validation*?
2. Diantara kedua algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbors*, manakah diantara kedua metode tersebut memiliki tingkat akurasi tinggi yang nantinya dapat digunakan untuk klasifikasi diagnosis penyakit kulit

1.3. Tujuan

1. Untuk mengukur seberapa akurat hasil klasifikasi diagnosis penyakit kulit dengan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* menggunakan *K-Fold Cross Validation*
2. Untuk mendapatkan hasil prediksi atau klasifikasi diagnosis pada penyakit kulit tidak hanya akurasi tinggi melainkan juga valid.

1.4. Manfaat

1. Bagi penulis, dapat menerapkan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dalam mendiagnosis penyakit kulit.
2. Bagi penulis, dapat menentukan metode mana dari metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* yang dapat memberikan kinerja terbaik dalam mendiagnosis penyakit kulit
3. Bagi pembaca, sebagai referensi untuk penelitian klasifikasi penyakit kulit yang akan datang

1.5. Batasan Masalah

Berdasarkan pada permasalahan yang telah diuraikan sebelumnya, maka batasan pada penelitian “Analisis Perbandingan Klasifikasi Antara KNN dan *Naive Bayes* Pada Diagnosa Penyakit Kulit menggunakan *K-Fold Cross Validation*” diantaranya:

1. Dataset yang implementasikan pada penelitian ini merupakan dataset yang diperoleh dari *platform* penyedia dataset secara *open source* yaitu data pasien dari website Kaggle berbentuk file comma separated value atau “.csv” dengan total data sebanyak 366 records.
2. Algoritma yang akan digunakan untuk menghitung dan membandingkan kinerja terhadap data dari pasien adalah 2 algoritma yaitu *naïve bayes* dan *k-nearest neighbour*
3. Bahasa yang digunakan pada penelitian ini yaitu menggunakan bahasa *python*
4. Atribut yang terdapat dalam dataset yaitu sebanyak 34 atribut yang akan digunakan untuk menghitung dan membandingkan tingkat kinerja algoritma klasifikasi penyakit kulit
5. Aspek yang diuji dari perbandingan algoritma *naïve bayes* dan *k-nearest neighbour* terdapat tiga metrik evaluasi kinerja yaitu akurasi, presisi dan *recall*
6. Hasil yang didapat pada penelitian ini merujuk mengenai tingkat akurasi paling tinggi diantara perbandingan penggunaan metode KNN dan *naive bayes* yang telah dilakukan uji validasi menggunakan *K-Fold Cross Validation*