

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker payudara (*Breast Cancer*) didefinisikan sebagai suatu penyakit *neoplasma* ganas yang berasal dari *parenchyma* dan menghasilkan frekuensi kematian yang menjadi penyebab utama kekhawatiran di dunia. Menurut data statistik Globocan (2015), kanker payudara merupakan kanker kedua yang paling banyak diderita dan penyebab kelima kematian kanker di seluruh dunia dengan presentase 6,4% dari semua penyebab kematian. Pada tahun 2017, sekitar 252.710 wanita terkena diagnosa kanker payudara dan 40.610 diantaranya mengalami keadaan kritis dan hampir menyebabkan kematian (Harbeck dan Gnant, 2019). Kanker payudara terbagi menjadi jinak dan ganas, para ahli dan dokter akan memberikan perawatan yang berbeda pada setiap pasien. Kunci untuk bertahan hidup penderita kanker payudara adalah dengan mendeteksi kanker payudara sedini mungkin, sebelum sel kanker menyebar ke bagian-bagian tubuh lainnya.

Untuk mengetahui seseorang terkena penyakit kanker payudara atau tidak sangatlah susah, oleh karena itu, pemeriksaan penting dilakukan dengan cara cek darah guna memastikan apakah seseorang terkena kanker payudara atau tidak. Dari dataset konsultasi rutin dan analisis darah (Patricio, dkk., 2018), menurut Adi dan Sari (dalam Artha., dkk. 2019), dengan bantuan *data mining*, dataset tersebut bisa diklasifikasikan, kemudian hasilnya dapat membantu dokter untuk mendiagnosis pasien apakah positif kanker payudara atau negatif.

Seiring dengan perkembangan ilmu pengetahuan dan teknologi informasi, kehadiran *machine learning* di bidang komputer telah menarik banyak perhatian. *Machine learning* memainkan peran luas dalam pengembangan terutama dalam pengembangan data analitik (Alarifi dan Young., 2018). Ada salah satu metode yang ada pada *machine learning* yaitu klasifikasi. Untuk klasifikasi itu sendiri banyak jenis algoritma yang dapat digunakan seperti *K-Nearest Neighbor* dan *Naive Bayes*.

Pada penelitian sebelumnya yang dilakukan oleh (Eyupoglu, 2017) dengan judul “*Cancer Classification Using K-Nearest Neighbors Algorithm*“, dengan menggunakan algoritma KNN atau bisa disebut dengan K-NEAREST NEIGHBORS menghasilkan akurasi 97% yang termasuk dalam *Good Classification*.

Pada penelitian yang dilakukan oleh (Kamel, dkk., 2019) dengan judul “*Cancer Classification Using Gaussian Naive Bayes Algorithm*“, dengan menggunakan algoritma *Gaussian Naive Bayes* menghasilkan akurasi 98% yang termasuk dalam *Good Classification*. Dengan demikian bahwa metode algoritma *Gaussian Naive Bayes* ini akurat dalam melakukan klasifikasi kanker payudara.

Berdasarkan penelitian sebelumnya yang telah dilakukan oleh Kamel, dkk. (2019) dan Eyupoglu (2017) yang membuktikan bahwa KNN dan GNB memiliki tingkat akurasi tinggi. Dikarenakan masih belum ada penelitian yang membahas tentang perbandingan KNN dan GNB dengan menggunakan dataset *Breast Cancer Coimbra* maka dilakukanlah penelitian ini.

Berdasarkan uraian di atas, maka dalam penelitian ini akan dilakukan perbandingan antara metode KNN dan GNB untuk mengklasifikasikan Breast Cancer Coimbra dengan judul “*Perbandingan Algoritma K-Nearest Neighbor (K-NN) dan Gaussian Naive Bayes (GNB) dalam Klasifikasi Breast Cancer Coimbra*”.

1.2 Rumusan Masalah

Dari latar belakang di atas maka rumusan masalah dari penelitian ini dibuat sebagai berikut:

1. Berapa hasil akurasi dari perbandingan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* ?

1.3 Batasan Masalah

Perlu adanya batasan masalah dalam penelitian yang dilakukan supaya tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi pada :

1. Data yang diambil dari *UCI Machine Learning Repository*. Data yang digunakan adalah *Breast Cancer Coimbra Data Set* tahun 2018 (Patricio, dkk, 2018).
2. Dataset yang terdiri dari 9 atribut dengan 116 *record*.
3. *Output* yang digunakan ada 2 yaitu *Classification (1)* dan *Classification (2)*.
4. *Tools* yang digunakan pada penelitian ini adalah *Jupyter Notebook*

1.4 Tujuan Penelitian

Berdasarkan pokok pembahasan yang telah dikemukakan sebelumnya, maka dapat ditarik tujuan dari penelitian yaitu:

1. Membandingkan hasil akurasi algoritma *K-Nearest Neighbor (K-NN)* dan *Gaussian Naive Bayes (GNB)* untuk mengklasifikasi penyakit kanker payudara.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dalam penelitian adalah:

Bagi Penulis:

1. Dapat sebagai syarat kelulusan jenjang S1.
2. Dapat memberi kesempatan untuk menambah wawasan

Bagi Peneliti Lain:

1. Dapat digunakan sebagai referensi untuk penelitian-penelitian selanjutnya di bidang klasifikasi dengan metode *Data Mining*.