

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan salah satu kejadian dimana sel dalam tubuh mengalami mutase yang kemudian membelah diri dan tumbuh secara lebih cepat dari sel tubuh normal pada umumnya. Selain itu, sel kanker tidak mati melainkan terus tumbuh secara invasif sehingga dapat mendesak sel tubuh normal [1]. Menurut data yang diambil dari Globocan pada tahun 2018, terhitung sebanyak 18,1 juta kematian telah terjadi, dimana kanker menyumbang sebanyak 9,6 juta dari total kematian. Indonesia menduduki peringkat 8 besar di wilayah Asia Tenggara untuk penderita penyakit kanker. Kanker paru-paru menduduki peringkat tertinggi yang menyumbang angka kematian pada laki-laki, yaitu sebesar 10,9 persen dari total 100.000 penduduk. Kemudian peringkat kedua diikuti oleh kanker hati dengan menyumbangkan angka rata-rata kematian sebesar 7,6 persen dari total 100.000 penduduk. Sedangkan untuk perempuan, kanker payudara menduduki peringkat pertama angka kematian dengan menyumbangkan angka sebesar 17 persen dari total 100.000 penduduk. Kemudian diikuti oleh kanker serviks pada peringkat kedua dengan menyumbangkan angka kematian sebesar 13,9 persen dari total 100.000 penduduk [2]. Salah satu penyebab meningkatnya jumlah kematian yang diakibatkan oleh penyakit kanker di Indonesia adalah dikarenakan diagnosis dan pemeriksaan yang dilakukan tenaga medis mengalami keterlambatan [3].

Hal tersebut bisa disebabkan karena proses analisis fitur oleh tenaga medis terkadang memiliki hasil yang berbeda. Berdasarkan pernyataan yang telah disebutkan sebelumnya, proses klasifikasi dapat digunakan untuk menyelesaikan beberapa masalah seperti penyakit kanker [4] [5]. Proses klasifikasi dilakukan untuk membantu diagnosis terhadap penderita penyakit kanker, terutama pada faktor risiko terkena kanker serviks yang sudah dibuktikan, diantaranya hubungan

seks, riwayat gen, kebiasaan merokok, alat kontrasepsi dan lainnya [6]. Klasifikasi dilakukan agar keterlambatan dalam proses diagnosis dapat dihindari dan masyarakat terutama perempuan dapat mengetahui penyakit tersebut secara dini. Banyak cara dapat dilakukan untuk melakukan proses pengklasifikasian, salah satunya menggunakan perhitungan komputerisasi dengan tujuan mempercepat diagnosis [7].

Saat ini di dalam dunia medis sudah banyak digunakan teknologi untuk menunjang sarana dan prasarana yang dapat membantu, salah satunya adalah pengolahan data menggunakan *data mining*. *Data mining* dapat digambarkan sebagai sebuah proses dalam melakukan penggalian informasi yang berguna dari basis data, dan juga dapat didefinisikan salah satu proses ekstraksi informasi dari basis data yang kemudian akan diolah untuk membantu pengambilan keputusan [8]. *Data mining* dapat diproses menggunakan metode dalam *data mining* itu sendiri sehingga menghasilkan keluaran [9]. Terdapat beberapa proses yang dapat dilakukan oleh metode yang ada pada *data mining* diantaranya Klustering, Asosiasi, Klasifikasi, Prediksi, dan Estimasi. Klasifikasi merupakan suatu metode yang digunakan untuk melakukan pemisahan antar kelas yang ada di dalam data, dimana bertujuan untuk memperkirakan kelas yang sesuai dari data yang belum memiliki label. [10].

Menurut penelitian Busman Nurhayati dan R.P. Iswara [11] *Supervised Learning* adalah salah satu bentuk metode dari *Machine Learning* yang menggunakan dataset berlabel dan kemudian fitur yang dimiliki diproses untuk selanjutnya diprediksi maupun diklasifikasi. Adapun penelitian ini penulis akan melakukan proses klasifikasi terhadap kasus kanker serviks sehingga teknik yang digunakan adalah *Supervised Learning*. Algoritma yang masuk ke dalam kategori *Supervised Learning* diantaranya *Random Forest* (RF), *K-Nearest Neighbor* (KNN), *Naive Bayes* (NB), dan *Support Vector Machine* (SVM) [12]. Menurut penelitian [13][14][15], RF merupakan salah satu algoritma *Machine Learning Decision Tree* yang dapat digunakan dalam penelitian bioinformatika sebagai metode klasifikasi, dimana algoritma ini menangani data berjumlah besar dengan

atribut kompleks yang bersifat *non-linear*. Kemudian menurut penelitian [16], KNN merupakan salah satu algoritma metode klasifikasi yang memproses data sesuai dengan jarak data paling dekat. Selanjutnya, menurut penelitian [17] dan [18], NB merupakan salah satu algoritma *data mining* untuk metode klasifikasi yang populer, dimana pengoperasiannya menggunakan bobot dari tiap kelas yang nantinya akan dicari probabilitas masing-masing. Probabilitas yang diasumsikan biasanya berasal dari data *multivariate*. Sedangkan, menurut penelitian [19][20][21], SVM adalah salah satu algoritma *machine learning* yang digunakan untuk metode klasifikasi, dimana algoritma tersebut menggunakan parameter kernel yang didasarkan pada pencarian *hyper-plane* antara dua kelas yang dapat menangani klasifikasi *non-linear*, regresi, dan estimasi fungsi *multivariat*.

Pada beberapa penelitian sebelumnya didapat algoritma terbaik pada proses pengklasifikasian yaitu SVM [22][23][24] dan RF [25][26][27]. Keunggulan algoritma SVM adalah waktu pemrosesan yang singkat (meliputi waktu latih), mudah diimplementasikan dalam dataset dengan dimensi tinggi, dapat memaksimalkan jarak antarkelas, dan kesalahan yang dihasilkan cenderung sedikit. Sedangkan untuk algoritma RF terletak pada waktu pemrosesan relatif singkat yang dapat menyeimbangkan kelas pada dataset untuk meningkatkan performa, dan dapat menghapus fitur yang tidak penting dalam klasifikasi [28]. Berdasarkan keunggulan dari kedua algoritma tersebut perlu adanya komparasi untuk menemukan algoritma terbaik dari segi akurasi, kesalahan yang dihasilkan, dan waktu komputasi yang diperlukan.

Adapun tujuan dari penelitian ini untuk menganalisis algoritma terbaik diantara algoritma SVM dan RF. Algoritma terbaik yang didapatkan selanjutnya diharapkan dapat membantu proses pengklasifikasian terhadap risiko yang menyebabkan kanker serviks sehingga keterlambatan dalam proses diagnosis dapat dihindari. Hal tersebut dapat meminimalisir tingkat kematian yang disebabkan oleh kanker serviks.

1.2 Rumusan Masalah

Menurut latar belakang yang sudah disebutkan, proses diagnosis merupakan salah satu tahapan penting dalam mengklasifikasi penyakit kanker serviks, akan tetapi proses diagnosis terkadang masih mengalami keterlambatan karena analisis setiap tenaga medis terkadang berbeda-beda. Berdasarkan hal tersebut diperlukan klasifikasi terhadap faktor risiko penyakit kanker serviks. Menurut studi literatur, algoritma *Support Vector Machine* dan *Random Forest* merupakan algoritma terbaik dalam proses klasifikasi sehingga dibutuhkan analisis perbandingan untuk mendapatkan algoritma dengan performa terbaik. Performa yang dinilai adalah dari segi ketepatan berupa *confusion matrix* (*Accuracy*, *Recall*, *Precision*, *F1-Score*). Selain itu performa yang dinilai adalah tingkat *error* dan waktu komputasi.

1.3 Pertanyaan Penelitian

Menurut rumusan masalah yang telah disebutkan, ada beberapa pertanyaan penelitian yaitu sebagai berikut:

- a. Dari algoritma SVM dan RF, bagaimana hasil klasifikasi dari risiko kanker serviks?
- b. Bagaimana hasil performa dari algoritma SVM dan RF dalam mengklasifikasikan faktor risiko pada kanker serviks?

1.4 Batasan Masalah

Penelitian ini terbatas pada beberapa hal, diantaranya:

- a. Data yang diambil dan digunakan pada penelitian ini yaitu dataset bersifat publik yang diambil melalui *website* UCI dengan karakteristik data *multivariate* bertipe integer dan real.
- b. Data yang digunakan merupakan supervised dengan jumlah 858 *record* dari 36 atribut, dengan atribut target utama adalah *biopsy*.
- c. Kernel yang digunakan pada algoritma SVM adalah *non-linear*.
- d. Pengukuran performa algoritma menggunakan *confusion matrix* yang meliputi akurasi, *recall*, *precision*, dan *error*. Selain itu, performa juga dapat diukur dari segi waktu komputasinya.

1.5 Tujuan

Berdasarkan pertanyaan penelitian yang telah disebutkan, penelitian ini memiliki beberapa tujuan yaitu sebagai berikut:

1. Menganalisis hasil klasifikasi dari risiko kanker serviks menggunakan algoritma SVM dan RF.
2. Menganalisis performa dan validasi dari algoritma SVM dan RF.

1.6 Manfaat

Hasil dari penelitian ini yaitu analisis untuk menemukan algoritma dengan performa terbaik. Selain itu, hasil yang berupa algoritma terbaik dalam proses klasifikasi dapat digunakan pada penelitian selanjutnya untuk mengklasifikasikan risiko terhadap kanker serviks sehingga dapat membantu dalam proses diagnosis.