

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penyebab utama kedua kematian wanita adalah kanker payudara (setelah kanker paru-paru) 1.246.660 kasus baru kanker payudara invasif perempuan diperkirakan akan didiagnosis di AS selama 2016 dan diperkirakan 40.450 kematian wanita. Kanker payudara mewakili sekitar 12% dari semua kasus kanker baru dan 25% dari semua kanker pada wanita (Asria et al, 2016). Deteksi dini dan klasifikasi kanker sangat penting untuk menyelamatkan nyawa seseorang. Salah satu penyakit mengerikan yang mempengaruhi wanita adalah kanker payudara dan itu menjadi perhatian utama di bidang medis. Kanker payudara muncul dari jaringan sel-sel payudara. Mirip dengan bagian lain dari tubuh manusia, payudara terdiri dari banyak sel mikroskopis. Dalam kasus kanker payudara, penggandaan sel terjadi dengan cepat di payudara dan menyebar ke bagian lain dari tubuh manusia (Prabhakar and Rajaguru, 2018).

Penyebab kanker payudara adalah multi faktorial dan melibatkan riwayat keluarga, obesitas, hormon, terapi radiasi, dan bahkan faktor reproduksi. Setiap tahun, satu juta wanita baru didiagnosis menderita kanker payudara, menurut laporan organisasi kesehatan dunia setengah dari mereka akan mati, karena biasanya terlambat ketika dokter mendeteksi kanker (Altonen et al.,1998). Kanker payudara disebabkan oleh mutasi pada sel tunggal, yang dapat ditutup oleh sistem atau menyebabkan pembelahan sel yang gegabah. Jika masalah tidak diperbaiki setelah beberapa bulan, akan menjadi tumor ganas.

Tumor ganas berkembang ke sel-sel tetangga, yang dapat menyebabkan metastasis atau mencapai bagian lain, sedangkan tumor jinak tidak dapat berkembang ke jaringan lain, ekspansi kemudian hanya terbatas pada tumor jinak (Chakraborty, 2009). Deteksi SM mungkin sulit pada awal penyakit, karena tidak adanya gejala, setelah beberapa tes klinis, diagnosis yang akurat harus memiliki kemampuan untuk membedakan tumor jinak dan ganas. Deteksi yang baik

memberikan tingkat salah positif (FP) yang rendah dan tingkat salah negatif (FN) (Guyon et al., 2002).

Machine Learning (ML) merupakan seperangkat alat yang digunakan untuk pembuatan dan evaluasi algoritma yang memfasilitasi prediksi, pengenalan pola, dan klasifikasi. *Machine Learning* didasarkan pada empat langkah: Mengumpulkan data, memilih model, melatih model, menguji model (Gokhale, 2009). *Machine Learning* adalah cabang ilmu pengetahuan yang mengimplementasikan algoritma matematika ke dalam pemrograman komputer untuk mengidentifikasi pola data dan meningkatkan kinerja secara iteratif. Aplikasi pembelajaran mesin telah memecahkan banyak masalah seperti, prediksi pasien kanker dan prediksi kebangkrutan perusahaan (Lynch1, 2017). Dalam beberapa tahun terakhir, berbagai pembelajaran mesin dan teknik komputasi lunak digunakan untuk mengklasifikasikan berbagai masalah medis termasuk kanker payudara. Mereka mengklasifikasi kanker payudara dengan bantuan dua teknik seperti Softmax Discriminant Classifier (SDC) dan Linear Discriminant Analysis (LDA) (Prabhakar and Rajaguru, 2018).

Hubungan antara kanker payudara dan *machine learning* bukanlah hal yang sangat baru karena telah banyak digunakan selama beberapa dekade untuk mengklasifikasikan tumor dan keganasan lainnya, memprediksi urutan gen yang bertanggung jawab terhadap kanker dan menentukan prognostik (Tang et al., 2009). Tujuan klasifikasi adalah untuk menempatkan setiap pengamatan dalam kategori yang dimilikinya. Dalam penelitian ini akan menggunakan pengklasifikasi pembelajaran mesin yaitu *Support vector machine*. Tujuannya adalah untuk menentukan apakah pasien memiliki tumor jinak atau ganas. Dalam penelitian ini menggunakan database kanker payudara Wisconsin. Tujuan penelitian adalah mengembangkan pendekatan pembelajaran mesin yang efektif untuk klasifikasi kanker menggunakan pengklasifikasi dalam satu set data. Kinerja hasil klasifikasi akan dievaluasi dalam hal akurasi, proses pelatihan dan proses pengujian dengan menggunakan teknik *Fitur Selection*.

Feature selection atau seleksi fitur adalah salah satu teknik terpenting dan sering digunakan dalam *pre-processing*. Teknik ini mengurangi jumlah fitur yang

terlibat dalam menentukan suatu nilai kelas target, mengurangi fitur irelevan, berlebihan dan data yang menyebabkan salah pengertian terhadap kelas target yang membuat efek segera bagi aplikasi. Tujuan utama dari seleksi fitur ialah memilih fitur terbaik dari suatu kumpulan fitur data.

Feature selection adalah suatu proses menghapus *features* yang berlebihan dan tidak relevan dari dataset yang sebenarnya. Sehingga waktu yang digunakan mengeksekusi dari pengklasifikasi yang memproses data berkurang, dan dapat meningkatkan akurasi juga karena *features* yang tidak relevan dapat memperburuk data mempengaruhi akurasi klasifikasi secara negatif (S. Doraisami dan S. Golzari,2008). Dengan *feature selection* dapat meningkatkan pemahaman dan biaya penanganan data menjadi lebih kecil (A. Arauzo-Azofra,2011). Algoritma *Feature selection* dibagi menjadi tiga kelompok: *filters*, *wrappers*, dan *embedded selectors*. *Filters* mengevaluasi setiap *feature* secara bebas dari pengklasifikasi, memberikan peringkat pada *feature* setelah mengevaluasi dan mengambil yang unggul (Guyon Isabelle dan A. Elisseeff, *Journal of Machine learning Research*). *Wrappers* mengambil subset dari *feature set*, mengevaluasi kinerja pengklasifikasian pada *subset* ini, dan kemudian *subset* lainnya dievaluasi oleh pengklasifikasi. *Subset* yang memiliki kinerja paling maksimum pada pengklasifikasi yang akan dipilih. Jadi *wrappers* bergantung pada pengklasifikasi yang dipilih. Bahkan *wrappers* lebih dapat diandalkan karena algoritma klasifikasi mempengaruhi tingkat akurasi (J. Novakovic,2010).

Forward selection adalah jenis regresi bertahap yang dimulai dengan model kosong. Dalam seleksi *forward*, variabel pertama yang dipilih untuk entri ke dalam model yang sudah dibangun adalah yang memiliki korelasi terbesar dengan variabel dependen. Setelah variabel telah dipilih kemudian dievaluasi berdasarkan kriteria tertentu. Jika variabel yang dipilih pertama memenuhi kriteria untuk dimasukkan, maka pemilihan maju berlanjut. Prosedur berhenti, ketika tidak ada variabel lain yang tersisa yang memenuhi kriteria entri dan menambahkan variabel satu per satu.

Dalam pemilihan fitur, *forward Selection* memiliki fungsi objektif J yang sudah dimaksimalkan, dan fungsi ini bergantung pada subset fitur $F\theta$. Tujuan dari

algoritma *Forward Selection* adalah untuk menemukan fitur k dalam F yang memaksimalkan fungsi objektif

Berdasarkan uraian diatas maka penulis melakukan penelitian yang berjudul “ **Klasifikasi Kanker Payudara Menggunakan Support Vector machine dan Forward Selection**”.

1.2 Masalah Penelitian

Secara umum tingkat keganasan kanker payudara dideteksi dengan cara prognosis. Prognosis dijadikan alat prediksi tim medis dalam menentukan sembuh atau tidaknya pasien dari kanker. Cara lain adalah dengan pemanfaatan bioinformatic yang menggunakan teknik data mining karena telah terbukti dapat mendeteksi tingkat keganasan kanker payudara. Untuk itu diperlukan suatu metode klasifikasi yang secara akurat dapat mengelompokkan data penyakit kanker payudara sehingga dapat diketahui dengan dini apakah penyakit kanker payudara tersebut dinyatakan *malignant* (ganas) atau *benign* (jinak). Penelitian ini menggunakan metode klasifikasi dengan algoritma Support Vector Machine.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah untuk melakukan klasifikasi data penyakit kanker payudara menggunakan algoritma *Support Vector Machine*. Adapun manfaat dari penelitian ini adalah untuk mengelompokkan data penyakit kanker payudara sehingga dapat diketahui dengan dini apakah penyakit kanker payudara tersebut dinyatakan *malignant* (ganas) atau *benign* (jinak).

1.4 Ruang Lingkup/Pembatasan Masalah

Dalam upaya untuk mencapai tujuan dalam penelitian ini, maka ruang lingkup penelitian ini mencakup:

1. *Dataset* yang digunakan adalah kanker payudara yang diambil dari *UCI machine Learning Repository*. Breast Cancer Wisconsin *Dataset* yang berisi 2 jenis kanker payudara seperti *benign* dan *malignant*. *Dataset* tersedia secara umum dan dapat

diakses pada *link*

[https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original))

2. Untuk Meningkatkan akurasi dengan menggunakan Teknik Forward Selection.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada tahap ini dilakukan pengumpulan bahan referensi yang berhubungan dengan penelitian, seperti: klasifikasi kanker payudara, jenis-jenis kanker payudara, pengolahan kanker payudara, metode *Support Vector Machine* serta bahan referensi lain terkait dengan penelitian mengenai klasifikasi kanker payudara.

2. Tahap Analisis

Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada.

3. Perancangan Sistem

Perancangan sistem dilakukan dengan membuat diagram alir (*flowchart*) dari setiap proses pada kedua metode yang digunakan.

4. Pengujian

Pada bagian pengujian ini akan menggunakan *Support Vector Machine*.

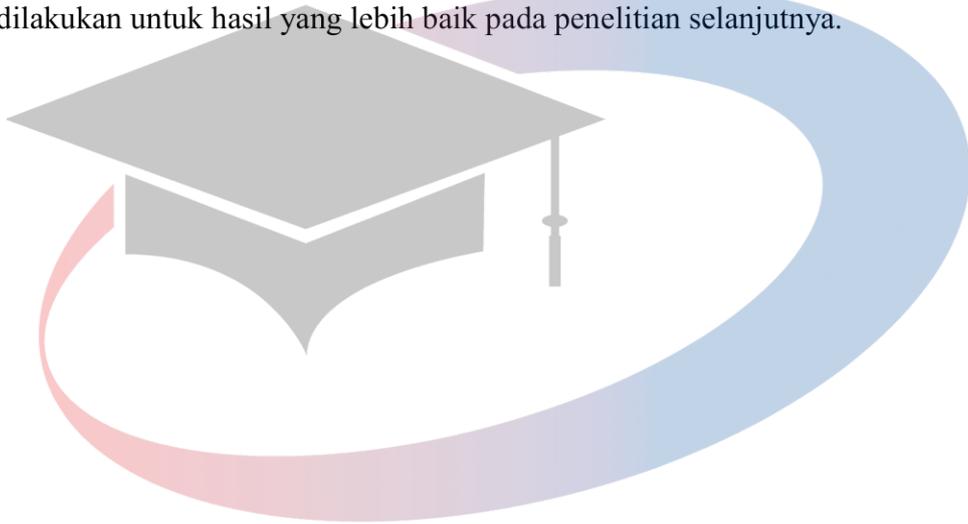
5. Evaluasi

Pada tahap ini, dilakukan evaluasi terhadap hasil pengujian yang sudah dilakukan untuk mengambil kesimpulan dan saran.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini terdiri dari 5 bab, dimana secara garis besar masing-masing bab membahas hal – hal berikut ini. Bab 1 Pendahuluan, berisi penjelasan umum, masalah dan solusi yang sudah ada dan

akan dilakukan. Bab 2 berisi studi literatur dan tinjauan pustaka terkait masalah dan metode yang berhubungan dengan penelitian yang akan dilakukan. Bab 3 Metodologi Penelitian, berisi identifikasi masalah, langkah-langkah dari metode yang diusulkan, data yang digunakan, alat-alat penelitian dan metode analisis. Bab 4 Hasil dan Pengujian, berisi hasil yang diperoleh dari sistem yang dibangun dan pengujian yang dilakukan. Bab 5 Kesimpulan dan Saran, berisi kesimpulan yang diperoleh dari hasil dan pengujian penelitian yang dilakukan dan saran yang dapat dilakukan untuk hasil yang lebih baik pada penelitian selanjutnya.



UNIVERSITAS MIKROSKIL