BAB I
PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan penyakit berat dan bersifat kronis, yang ditandai pertumbuhan sel tubuh tidak normal, berkembang cepat, menyebar, dan menekan organ atau saraf sekitar (Chotimah, 2014). Kanker serviks adalah suatu kondisi medis yang ditandai dengan kanker pada sel-sel melapisi serviks. Kanker serviks adalah kanker pada organ reproduksi wanita paling umum. Serviks adalah bagian bawah dari rahim yang menghubungkan rahim ke vagina. Sel kanker biasanya dimulai dari permukaan serviks (Hidayatullah, 2015).

Menurut laporan *World Health Organization* (WHO, 2017) Sekitar 90% 270.000 kematian akibat kanker serviks pada 2015 terjadi di negara-negara berpenghasilan rendah dan menengah. Di Indonesia, berdasarkan laporan *Global Burden Cancer* (GLOBOCAN, 2012) setiap tahunnya, sebanyak 20.928 perempuan Indonesia terdiagnosis kanker serviks dan 9.498 diantaranya meninggal karena kanker serviks. Angka tersebut merupakan yang tertinggi di ASEAN. Diikuti oleh Vietnam, Malaysia, dan Singapura. Pada 2014 lalu, lebih dari 92 ribu perempuan Indonesia juga meninggal karena kanker dengan 10,3 persen di antaranya karena kanker serviks (CNN Indonesia, 2017). Tingginya jumlah kasus dan angka kematian pada kanker serviks karena tidak adanya gejala pada pengidap kanker serviks pada stadium awal. Gejala akan mulai dirasakan ketika kanker berada pada stadium lanjut sehingga diagnosa secara medis sulit dilakukan. Menurut (Endriyani, 2017) Kurangnya pengetahuan masyarakat mengenai kanker serviks dan keengganan untuk melakukan deteksi dini menyebabkan lebih dari 70% mulai menjalani perawatan medis justru ketika sudah berada kondisi parah dan sulit disembuhkan.

Deteksi dini perlu dilakukan untuk dapat memberikan penanganan secara tepat terhadap pasien telah terdiagnosa kanker serviks. Menurut (Wilgin, 2011) terdapat beberapa metode deteksi dini terhadap kanker leher rahim, yaitu tes *pap smear (*pengambilan sampel sel dari leher Rahim), *Inspeksi Visual Asam Asetat* (IVA), *kolposkopi* (pemeriksaan secara visual dengan menggunakan

pembesaran untuk melihat bagian luar dan dalam dari leher rahim), *servikografi* (metode penapisan primer), dan tes HPV.

Saat ini perkembangan ilmu pengetahuan dan teknologi yang semakin pesat berpengaruh terhadap berbagai bidang. *Data Mining* merupakan cabang ilmu baru di bidang teknologi yang telah banyak diterapkan di berbagai bidang ilmu seperti kecerdasan buatan, pengenalan pola, pengolahan citra, statistik, *database*, dan lain sebagainya yang dapat memudahkan manusia dalam berbagai masalah. Pada dunia medis teknik *data mining* diperkenalkan untuk dapat meningkatkan kemampuan pendeteksian penyakit secara otomatis. Dengan bantuan teknik ini, kemungkinan kesalahan diagnosis yang dilakukan karena ahli yang tidak berpengalaman, dan juga menyediakan data medis untuk diperiksa dalam waktu yang lebih singkat dan lebih terperinci (Akay , 2009).

Penelitian dengan teknik *data mining* untuk diagnosa kanker serviks telah banyak dilakukan oleh para peneliti diantara dilakukan oleh (Rahmadwati, 2014) Pada penelitiannya mengkaji diagnosis kanker serviks berdasarkan karakteristik *morfologi* (bentuk organisme) sel serviks. Dengan menghasilkan algoritma *K-means clustering* memberikan hasil yang menjanjikan pada klasifikasi normal, CIN1,CIN2,CIN3 dan ganas.

(Auliani, 2014) Menerapkan metode *Naïve Bayes* untu mengidentifikasi sel normal dan abnormal berdasarkan data citra sel tunggal *pap smear*. Penelitian ini menitik beratkan kepada metode klasifikasi dan tidak melakukan proses segmentasi serta ekstraksi ciri. Hasil yang diperoleh menunjukkan akurasi 97% untuk dua kelas klasifikasi dan 9% untuk tujuh kelas klasifikasi. Rendahnya akurasi pada tujuh kelas klasifikasi disebabkan oleh terjadinya *overlapping* (tumpang tindih) pada fitur-fitur yang digunakan. Metode klasifikasi *Naïve Bayes* memberikan hasil yang memuaskan pada klasifikasi dua kelas target, tetapi metode ini belum memberikan hasil yang baik untuk klasifikasi tujuh kelas target.

(Bestry, Hidayat, & Fauzi, 2016) Melakukan perancangan suatu sistem untuk mendeteksi dini kanker serviks dengan metode IVA dengan bantuan *software* Matlab 2015a. Data yang digunakan adalah sampel citra serviks yang telah dilakukan skrining dengan metode IVA *(Inspeksi Visual*

Asam Asetat) menentukan apakah hasil skrining merupakan IVA positif atau IVA negatif. Berdasarkan pengujian yang dilakukan pada 55 data citra uji diperoleh nilai akurasi sebesar 89% untuk sistem deteksi SSK 78%, untuk sistem deteksi AEZ, dan 67% untuk sistem deteksi akhir hasil skrining IVA.

(Sholik & Fatichah, 2017) melakukan penelitian tentang klasifikasi sel serviks pada citra *pap smear* dengan mengusulkan metode *Fuzzy* k-NN dengan kombinasi fitur bentuk deskriptor regional dan *fitur tekstur Uniform Rotated Local Binary Pattern* (uRLBP). Hasil pengujian menunjukkan Nilai akurasi menggunakan metode klasifikasi *Fuzzy* k-NN adalah 91.59% untuk dua kategori sel dan 67.89% untuk tujuh kelas sel.

(Praningki & Budi, 2017) Pada penelitiannya menggunakan *Classification And Regression Trees* (CART), *Naive Bayes, dan k-Nearest Neighbor (k-NN)* untuk prediksi penyakit kanker serviks untuk mendukung keputusan klinis pada pasien baru. Hasil pengujian terhadap algoritma yang digunakan, menunjukkan algoritma *Naive Bayes* memiliki akurasi terbaik sebesar 94,44%, sedangkan tingkat akurasi yang dihasilkan algoritma CART sebesar 88,89% dan k-NN adalah 85,04%.

(Unlersen, Sabanci, & Özcan, 2017) menggunakan *Multilayer Perceptron, BayesNet dan k-Nearest Neighbor* untuk melakukan klasifikasi pasien berdasarkan jawaban mereka terhadap survei telah dilakukan untuk menentukan pasien yang memiliki risiko kanker serviks yang tinggi. Dataset diperoleh dari Universitas California Irvene (UCI) Machine Learning Repository 2017. Dataset bernama *Cervical Cancer (Risk Faktors).* Dataset yang digunakan sebanyak 858 *record* dengan 33 atribut dan *class* target hasil biopsi dengan menujukkan hasil akurasi *Multilayer Perceptron* 95.89%, *BayesNet* 95.89% dan *k-Nearest Neighbor* 97.26%. Kemudian dilanjutkan oleh (Wen & Zhou, 2017) Dengan melakukan perbandingan dari pendekatan metode SVM, SVM-RFE, dan SVM-PCA dan hasil dari peneltiannya adalah SVM adalah metode yang terbaik diantara lainnya.

SVM merupakan suatu metode baru yang memiliki tingkat performansi baik dibandingkan dengan metode klasifikasi lainnya (Burges, 1998). Metode ini memiliki banyak aplikasi di bidang medis atau perawatan kesehatan,

pengenalan pola, teks klasifikasi dan banyak lagi (Kaur & Ginige, 2018). SVM memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan *support vector* sehingga proses komputasi menjadi cepat (Cortes & Vapnik, 1995). SVM terbukti mampu melakukan diagnosa terhadap berbagai penyakit, dibuktikan pada penelitian (rachman & Purnami, 2012) dilakukan analisis tingkat keganasan *breast cancer* dengan menggunakan *regresi logistik ordinal* dan *support vector machine (SVM)*. Metode *regresi logistik ordinal*, variabel yang berpengaruh terhadap tingkat keganasan *breast cancer* adalah ukuran tumor dan nodus dengan ketepatan klasifikasi tertinggi 56.60%. Sedangkan ketepatan klasifikasi SVM tertinggi dengan menggunakan *kernel RBF* dan *polynomial* mencapai 98.11%. Kemudian pada prediksi penyakit Hepatitis oleh (Saputra, 2017) dengan membandingkan metode *Multilayer Perceptron (MLP)* dan *Support Vector Machine (SVM)* diketahui bahwa *Support Vector Machine (SVM)* memiliki nilai tertinggi akurasi adalah 90,64%, sedangkan algoritma *Multilayer Perceptron (MLP)* memiliki akurasi 84,38%.

Penggunakan terlalu banyak fitur akan menghasilkan ruang fitur yang sangat besar. Menyebabkan proses training model lebih lambat, dan menurunkan akurasi. Biasanya, ada banyak fitur yang berlebihan atau tidak relevan, jadi dengan menggunakan seleksi fitur adalah ide bagus untuk menghapus fitur berlebihan yang kurang diperlukan. Pemilihan seleksi fitur dapat membantu mengurangi fitur domain, menghilangkan fitur yang berlebihan. Dengan cara ini akan membantu mempercepat proses pembelajaran atau pemodelan (H. Jiawei, M. Kamber, & J. Pe, 2012).

ReliefF merupakan seleksi fitur yang merupakan perkembangan dari algoritma *ReliefF* yang ditemukan oleh Kira dan Rendell (Igor, 1994). *ReliefF* termasuk dalam algoritma yang menggunakan fungsi evaluasi *distance measure* sehingga memiliki waktu kompleksitas yang rendah (Dash & Liu, 1997). Penanganan pada dataset dengan data yang besar, jumlah *instance* lebih dari 300, dan bisa kita lakukan dengan menggunakan sampling secara acak dari dataset tersebut (Robnik & Kononenko, 2003).

Bedasarkan latar belakang diatas maka akan dilakukan penelitian untuk diagnosa pasien kanker serviks dengan menerapkan algoritma *Support Vektor Machine* dengan seleksi fitur *ReliefF*.

1.2 Rumusan Masalah

Berdasarkan dari uraian latar belakang di atas maka rumusan masalah dalam penelitian ini, antara lain :

1. Bagaimana Metode *Support Vektor Machine* menyelesaikan permasalahan diagnosa kanker serviks ?
2. Bagaimana teknik seleksi fitur *ReliefF* dalam mereduksi atribut data kanker serviks ?
3. Bagaimana performa algoritma *Support Vector Machine* dengan teknik seleksi fitur *ReliefF* dalam mendiagnosa kanker serviks ?

1.3 Tujuan Penelitian

Dari rumusan masalah yang terjadi maka penelitian ini bertujuan untuk:

1. Mengimplementasikan metode *Support Vector Machine* untuk dapat menyelesaikan permasalahan diagnosa kanker serviks.
2. Mejelaskan bagaimana teknik seleksi *ReliefF* mampu mereduksi atribut data kanker serviks.
3. Mengetahui performa *Support Vektor Machine* dengan teknik seleksi fitur *ReliefF* dalam diagnosa kanker serviks.

1.4 Manfaat Penelitian

Penelitian ini diharapkan bisa menjadi referensi pada penelitian-penelitian selanjutnya dalam melakukan diagnosa kanker serviks secara otomatis dan terperinci untuk menghindari terjadinya *bias* terhadap diagnosis yang dilakukan oleh pakar sehingga dapat menurunkan angka kematian yang disebabkan oleh kanker serviks.

1.5 Ruang Lingkup

Agar dalam penulisan tugas akhir ini tidak melebar dan lebih terarah, maka perlu adanya batasan masalah pada ruang lingkup bahasan antara lain adalah sebagai berikut :

1. Dataset diperoleh dari Universitas California Irvene (UCI) Machine Learning Repository 2017, yaitu *Cervical Cancer (Risk Faktors)*.
2. Menggunakan metode *support Vektor Machine* (SVM) dengan seleksi fitur *ReliefF*
3. Aplikasi yang akan digunakan untuk memproses data adalah *Waikato Environment for Knowledge Analysis* (WEKA).

1.6 Sistematika Pembahasan

Untuk memahami lebih jelas isi dari tugas akhir ini, maka materi-materi yang tertera pada skripsi ini dikelompokkan menjadi beberapa sub bab dengan sistematika pembahasan sebagaimana berikut :

Bab I : Merupakan pendahuluan yang berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup pembahasan dan sistematika pembahasan.

Bab II : Yaitu tinjauan pustaka yang menjelaskan tentang pengertian kanker serviks, *data mining*, algoritma *Support Vector Machine (SVM)* dan teknik seleksi fitur *ReliefF* yang diambil dari kutipan buku atau jurnal yang berkaitan dengan penyusunan laporan skripsi serta beberapa *literature review* yang berhubungan dengan penelitian.

Bab III : Metodologi penelitian yang berisi tentang analisis kebutuhan alat, metode teknik pengumpulan data, pengolahan data awal, metode yang diusulkan, pemodelan algoritma pembelajaran, dan teknik evaluasi yang akan diusulkan untuk mengetahui seberapa berhasilnya metode yang telah diusulkan oleh peneliti.

Bab IV : Merupakan hasil dan pembahasan yang meliputi perancangan sistem, spesifikasi alat yang dibutuhkan, implementasi dari rancangan yang telah dibuat dan penjelasan proses diagnosa kanker serviks yang telah diusulkan peneliti.

Bab V : Adalah penutup dari tugas akhir ini yang berisi tentang kesimpulan dan saran-sara