

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan penyakit yang terjadi karena perubahan sel yang menyebabkan pertumbuhan dan pembelahan sel tak terkendali. Sel kanker dapat menyebabkan tumor, merusak sistem kekebalan tubuh, dan menyebabkan kelainan lain yang mencegah tubuh berfungsi dengan benar (Kondo, 1993). Kanker paru adalah tumor ganas paru yang berasal dari epitel saluran napas terutama pada bronkus. Menurut *World Health Organization* (WHO), kanker paru merupakan penyebab utama kematian akibat kanker dibandingkan dengan jenis kanker lainnya. Data yang diperoleh menunjukkan bahwa 32% kematian akibat kanker pada pria dan 25% pada wanita disebabkan oleh kanker paru. Sebagian besar kasus kanker paru menyerang orang berusia antara 35 dan 75 tahun, dengan insiden tertinggi antara 55 dan 65 tahun. Hanya sekitar 2% pasien kanker paru yang didiagnosis dengan metastasis dapat tetap hidup lima tahun setelah diagnosis (Aliyah et al., 2016). Faktor risiko kanker paru dapat dibagi menjadi faktor yang dapat dikendalikan dan tidak dapat dikendalikan. Faktor yang dapat dikendalikan yaitu merokok, paparan asap rokok, paparan terhadap radon, asbes, dan zat karsinogen lainnya. Faktor yang tidak dapat dikendalikan yaitu usia, jenis kelamin, dan genetik. Banyak pasien kanker paru datang dengan keadaan stadium lanjut, dimana prognosis kanker paru pada stadium ini lebih buruk dibandingkan dengan jenis kanker lainnya. Saat ini, proses deteksi

dini pada kanker paru masih berpotensi bahaya karena belum sepenuhnya efisien dan dapat menimbulkan risiko radiasi yang tinggi (Alfarisa et al., 2023).

Klasifikasi merupakan sebuah metode untuk mengelompokan data atau objek ke dalam kelas yang sebelumnya sudah didefinisikan (Fayyad et al., 1996). Terdapat beberapa metode dalam klasifikasi antara lain *Support Vector Machine*, *Decision Tree*, *Rule-Based Classifier*, *Artificial Neural Networks*, *Naïve Bayes*, dan *Lazy Learners* (Han & Kamber, 2006). *Support Vector Machine* (SVM) adalah salah satu metode yang dinilai memiliki nilai akurasi yang tinggi. Menurut Vapnik (1995), konsep klasifikasi dengan SVM dapat dijelaskan sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dari dua kelas pada ruang *input*. Kelebihan metode SVM adalah dapat bekerja pada ruang dimensi tinggi. Dimensi tinggi merujuk pada dataset yang memiliki variabel dengan jumlah yang banyak. Metode SVM juga memiliki kekurangan yaitu kesulitan dalam menentukan parameter yang optimal, sehingga hal tersebut dapat mempengaruhi hasil akurasi klasifikasi dari SVM (Guyon et al., 2002). Salah satu peneliti, Fei et al (2009), berusaha menambahkan metode optimasi yaitu *Particle Swarm Optimization* dalam memprediksi produksi biji-bijian Tiongkok pada metode SVM, *Artificial Neural Network*, dan *Grey Model*. Hasil penelitian menunjukkan PSO-SVM merupakan metode yang mencapai akurasi peramalan paling besar.

Particle Swarm Optimization (PSO) merupakan salah satu metode optimasi yang digunakan untuk menyelesaikan masalah optimasi karena mudah diimplementasikan dan terbukti lebih kompetitif daripada algoritma optimasi lainnya (Sousa et al., 2004). Pada algoritma PSO, pencarian solusi dilakukan oleh suatu populasi yang terdiri dari beberapa partikel. Setiap partikel mencari solusi dengan

melintasi ruang pencarian dengan cara melakukan penyesuaian terhadap posisi terbaiknya (*local best*) dan penyesuaian terhadap posisi partikel terbaik dari seluruh kawanan (*global best*) selama melintasi ruang pencarian. Sejumlah iterasi dilakukan untuk mencari posisi terbaik setiap partikel sampai didapatkan posisi yang relatif tetap atau mencapai batas iterasi yang ditetapkan. Pada tiap iterasi, setiap solusi (posisi partikel) dievaluasi performanya dengan cara memasukkan solusi tersebut ke dalam *fitness function*.

Permasalahan yang sering muncul saat melakukan klasifikasi adalah keadaan kelas pada dataset yang tidak seimbang. Penerapan algoritma klasifikasi secara langsung tanpa memperhatikan keseimbangan distribusi kelas dapat mengakibatkan prediksi yang lebih baik bagi kelas mayoritas dan tidak baik pada kelas minoritas. Oleh karena itu, diberikan penanganan untuk data tidak seimbang menggunakan teknik *undersampling* dan *oversampling*. Pada implementasinya, *undersampling* bekerja dengan mengurangi data pada kelas mayoritas secara acak, namun metode ini memiliki risiko hilangnya informasi penting pada data. *Oversampling* bekerja dengan cara mereplikasi data pada kelas minoritas secara acak, namun hal ini memungkinkan munculnya *overfitting* karena menduplikasi data yang sama persis. Oleh karena itu, dikembangkan metode *Synthetic Minority Oversampling Technique* (SMOTE) dan *Adaptive Synthetic Sampling Approach* (ADASYN). SMOTE merupakan algoritma yang melakukan *oversampling* pada kelas minoritas, namun data sintetis yang dimunculkan dibuat berdasarkan pendekatan ketetanggaan (Chawla et al., 2002). Sedangkan algoritma ADASYN merupakan teknik *oversampling* yang menghasilkan pengamatan sintetis sesuai dengan tingkat kesulitan dalam mempelajari data pada kelas minoritas tertentu (He et al., 2008).

Penelitian yang dilakukan Ismafillah *et al.* (2023) mengenai pengujian model SVM-SMOTE dengan menggunakan *confusion matrix* menghasilkan akurasi sebesar 95,3% dalam memprediksi penyakit stroke. Selain itu, Indriani dan Muslim (2019) mendiagnosis penyakit ginjal kronis menggunakan metode SVM yang dioptimasi menggunakan PSO dan *Adaboost*. Penelitian tersebut memperoleh peningkatan akurasi sebesar 36,20% menjadi 99,5%. Membandingkan metode SMOTE dan ADASYN telah dilakukan Ramadhanti *et al.* (2023) dalam mengklasifikasi kategori rumah tangga di Kabupaten Temanggung dengan algoritma *K-Nearest Neighbors* (KNN). Jumlah anggota kelas minoritas sebesar 13% dari total dataset. Hasil yang diperoleh adalah metode SMOTE-KNN mampu menghasilkan nilai *G-Mean* sebesar 58,5%, sedangkan ADASYN-KNN menghasilkan *G-Mean* sebesar 57,3%.

Berdasarkan uraian masalah di atas, maka penelitian ini bertujuan untuk mengetahui perbandingan akurasi dan *measurement* lain dari teknik *oversampling* SMOTE dan ADASYN untuk mengklasifikasi penyakit kanker paru dengan *Support Vector Machine* berbasis *Particle Swarm Optimization*.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan penanganan data tidak seimbang SMOTE dalam klasifikasi penyakit kanker paru?
2. Bagaimana penerapan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan penanganan data tidak seimbang ADASYN dalam klasifikasi penyakit kanker paru?

3. Bagaimana perbandingan akurasi, sensitivitas, dan spesifisitas *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (SVM-PSO) dengan SMOTE dan *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan ADASYN?

1.3 Batasan Masalah

Penelitian ini memerlukan batasan-batasan agar tujuan penelitian dapat tercapai. Batasan permasalahan yang dibahas dalam penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari *website Kaggle* yang berjudul *Lung Cancer*.
2. Parameter PSO yang digunakan adalah bobot inersia ($\rho_{max} = 0,9$ dan $\rho_{min} = 0,4$), koefisien akselerasi ($c_1 = c_2 = 2$), nilai random ($r_1 = 1$ dan $r_2 = 0,4$), jumlah partikel ($N = 20$), dan jumlah iterasi ($t_{max} = 50$) (Santosa & Willy, 2011).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengaplikasikan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan penanganan data tidak seimbang SMOTE dalam klasifikasi penyakit kanker paru.
2. Mengaplikasikan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan penanganan data tidak seimbang ADASYN dalam klasifikasi penyakit kanker paru.
3. Membandingkan akurasi, sensitivitas, dan spesifisitas antara *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan SMOTE

dan *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) dengan ADASYN.