

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penderita parkinson meningkat dalam 25 tahun terakhir. Menurut data dari *World Health Organization* (WHO), jumlah diseluruh dunia mencapai 5,8 juta pasien. Angka ini menunjukkan peningkatan sebesar 81% sejak tahun 2000, yang juga diikuti dengan lonjakan angka kematian sebanyak 329 ribu jiwa, meningkat 100% sejak tahun tersebut [1]. Hasil analisis sistematis dari *Global Burden of Disease Study* menunjukkan perkiraan kasus parkinson di Indonesia pada tahun 2016 mencapai antara 117.531 hingga 178.755 kasus. Angka ini mengalami peningkatan sebesar 21,7% dari tahun 1990 hingga 2016. Parkinson merupakan penyakit neurodegeneratif sehingga menurunkan fungsi tubuh secara degeneratif. Dalam kondisi ini, otak menjadi pusat serangan yang menciptakan gejala motorik dan non-motorik. Peningkatan kasus parkinson menimbulkan kebutuhan mendesak akan diagnosa yang lebih awal karena sifatnya yang degeneratif. Metode konvensional sering kali tidak memberikan hasil yang akurat, seperti pada *Computed tomography* dan *magnetic resonance imaging* yang kurang efektif karena menunjukkan kondisi otak yang normal[2]. Pada penelitian terdahulu berpendapat bahwa salah satu gejala awal pada penyakit parkinson adalah adanya gangguan pada kontrol vokal [3]. Penderita parkinson umumnya menunjukkan berbagai gejala vokal yang meliputi gangguan dalam produksi suara vokal yang normal (disfonia) dan masalah dalam artikulasi bicara yang normal (disartria). Gejala disfonia umumnya meliputi penurunan volume suara, kelembutan suara, kekasaran suara, penurunan energi pada bagian-bagian tinggi spektrum harmonik dan tremor vokal yang berlebihan [4]. Oleh karena itu, pengembangan sistem komputer dapat dimanfaatkan untuk mendiagnosa parkinson melalui perubahan suara vokal.

Dengan pemanfaatan pengembangan sistem komputer melalui *machine learning*, suara dianalisis untuk mengetahui sedari awal apakah menderita parkinson atau tidak. *Machine learning* dapat mempelajari pola dari data

menghasilkan prediksi yang lebih akurat daripada metode klasifikasi tradisional [5]. Data rekaman suara pasien penyakit parkinson tersedia dalam jumlah yang besar dan mudah diakses. Penelitian ini akan menganalisis rekaman suara dalam bentuk frekuensi sinyal dimana sumber data rekaman berasal dari *website UCI machine learning*. Beberapa metode yang telah dilakukan untuk klasifikasi parkinson salah satunya *Support Vector Machine (SVM)*.

Support Vector Machine (SVM) memiliki kemampuan untuk bekerja dengan baik dalam dimensi data yang tinggi, sehingga cocok untuk penanganan dataset yang memiliki banyak fitur. *Support Vector Machine (SVM)* bisa mengoptimalkan generalisasi dengan mencari batas pemisah yang menghasilkan margin terbesar antara kelas-kelas yang berbeda. Hal ini dapat membantu mengurangi risiko *overfitting* dan meningkatkan kemampuan model dalam mengklasifikasikan sampel yang tidak diketahui dengan akurasi tinggi. Proses tersebut menjadi keunggulan dari *Support Vector Machine (SVM)* karena meningkatkan proses komputasi sehingga lebih cepat [6]. Penelitian yang membandingkan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* pada klasifikasi diabetes melitus menunjukkan bahwa *Support Vector Machine (SVM)* dengan kernel polinomial memiliki akurasi terbaik sebesar 96.2704% dan tingkat error sebanyak 3.7296% [7]. Perbandingan lain terdapat pada penelitian yang membandingkan *Support Vector Machine (SVM)* dan *Decision Tree* pada klasifikasi kanker payudara. Penelitian tersebut menunjukkan bahwa algoritma *Support Vector Machine (SVM)* menghasilkan akurasi terbaik setelah dilakukan seleksi fitur [8]. Pada penelitian yang membandingkan kernel *Support Vector Machine (SVM)* dalam melakukan klasifikasi penundaan biaya kuliah mahasiswa. 3 kernel yang diuji adalah linier, polinomial dan *Radial Basis Function (RBF)*. Hasil penelitian menunjukkan bahwa penggunaan jenis kernel dalam *Support Vector Machine (SVM)* memiliki pengaruh signifikan terhadap akurasi model, begitu juga dengan pemilihan pola pengujian dan data latih [9]. Dari penelitian yang sudah dilakukan sebelumnya dapat disimpulkan bahwa penggunaan kernel pada *Support Vector Machine (SVM)* bisa mempengaruhi hasil akurasi klasifikasi. Kernel linear cocok untuk data yang memiliki hubungan linier antara fitur dan label kelas. Kernel *Radial Basis Function (RBF)* memiliki

kemampuan untuk menyesuaikan bentuk pemisah yang kompleks dan kemampuan untuk menangani data yang berdimensi tinggi. Kernel polinomial memungkinkan *Support Vector Machine* (SVM) untuk menangani hubungan yang lebih kompleks dan polinomial antara fitur dan label kelas. Setiap kernel memberikan pendekatan yang berbeda dalam memodelkan hubungan antara fitur dan label kelas, sehingga *ensemble* dapat menangkap beragam aspek dari data yang mungkin terlewatkan oleh satu kernel tunggal. Dengan menggabungkan hasil prediksi dari beberapa kernel, *ensemble* dapat memperkuat keputusan yang diambil dan mengurangi pengaruh keputusan yang mungkin salah dari satu kernel.

Metode *ensemble* dapat digunakan untuk meningkatkan akurasi model dengan menggabungkan beberapa classifier [10]. Terdapat beberapa jenis *ensemble learning*, termasuk *Bagging*, *Boosting*, *Stacking*, dan *Voting*. Dalam penelitian ini, fokus diberikan pada jenis *Voting*, terutama mayoritas dan rata-rata. Pendekatan mayoritas dilakukan melalui pemilihan label yang paling dominan dari setiap prediksi label yang dihasilkan oleh berbagai algoritma. Pendekatan rata-rata dilakukan dengan menghitung nilai bobot tertinggi dari hasil masing-masing metode klasifikasi. Pemilihan metode ini didasarkan pada keinginan untuk menjaga kesederhanaan namun tetap menghasilkan hasil yang optimal. Algoritma yang digunakan untuk *ensemble learning* dalam penelitian ini adalah *Support Vector Machine* (SVM) kernel linier, polinomial dan RBF. Metode klasifikasi model *Ensemble Support Vector Machine* (En-SVM) memiliki kelebihan dari segi algoritma yang digunakan, Namun terdapat kekurangan dari *Support Vector Machine* (SVM) yaitu tidak dapat menghasilkan prediksi yang akurat ketika mempunyai banyak fitur yang tidak relevan, tidak semua fitur digunakan dalam proses pembuatan model, hal ini bisa diatasi dengan metode seleksi fitur.

Proses seleksi fitur bertujuan untuk mengurangi jumlah fitur dalam dataset dan mempertahankan fitur yang memberikan kontribusi signifikan, mengatasi masalah seperti *overfitting* [11]. Salah satu metode yang umum digunakan dalam seleksi fitur adalah *Recursive Feature Elimination* (RFE). *Recursive Feature Elimination* (RFE) adalah teknik berbasis *wrapper* yang tidak hanya menilai

korelasi, namun juga melakukan peringkat fitur berdasarkan tingkat kepentingannya dalam proses prediksi. Setiap iterasi dalam *Recursive Feature Elimination* (RFE) mengukur peringkat fitur dan secara bertahap menghilangkan fitur yang dianggap kurang relevan. Untuk menghitung peringkat tersebut, metode *Support Vector Machine* (SVM) dengan kernel linier dapat digunakan. Terdapat permasalahan lain yang menyebabkan *overfitting* dalam klasifikasi yaitu masalah ketidakseimbangan data.

Ketidakseimbangan data terjadi ketika salah satu kelas memiliki jumlah yang jauh lebih besar dibanding kelas lainnya sehingga menyebabkan menurunnya kinerja klasifikasi pada kelas minoritas [12]. Adanya data tidak seimbang untuk proses klasifikasi menyebabkan hasil klasifikasi data minoritas menjadi tidak tepat atau tertutupi oleh prediksi data mayoritas sehingga dibutuhkan solusi untuk mengatasi masalah tersebut salah satunya yaitu melakukan *oversampling* misalnya dengan menambah distribusi data pada kelas minoritas menggunakan representatif nilai varian. Penelitian ini akan membangun model *Ensemble Support Vector Machine* (En-SVM) dan optimalisasi dataset untuk mendapatkan kinerja tertinggi. Untuk melihat kinerja terbaik dari model akan dilihat melalui nilai akurasi, presisi dan *recall*. Dengan metode yang diusulkan diharapkan dapat memperkuat klasifikasi penderita parkinson melalui analisis frekuensi suara.

Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaan seperti itu, maka penulis tertarik untuk melakukan klasifikasi penderita parkinson berdasarkan perubahan suara vokal dengan *Cross-Industry Standard Process for Data Mining* (CRISP-DM). *Cross-Industry Standard Process for Data Mining* (CRISP-DM) merupakan suatu pendekatan metodologi yang digunakan dalam mengelola proyek *data mining* atau analisis data. Metodologi ini terdiri dari serangkaian tahapan yang terstruktur untuk mengidentifikasi dan menyelesaikan permasalahan yang terkait dengan analisis data [13]. Untuk mengimplementasikan ketertarikan penulis terhadap masalah tersebut, maka penulis akan menuangkan penelitiannya dalam bentuk penulisan skripsi yang berjudul **“Metode Ensemble Support Vector Machine (En-Svm) Untuk Klasifikasi Penderita Parkinson”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan maka dapat diperoleh rumusan masalah sebagai berikut

1.2.1 Permasalahan

Peningkatan kasus parkinson menimbulkan kebutuhan mendesak akan penanganan yang lebih awal dan klasifikasi lebih akurat terutama melalui analisis suara vokal, namun metode konvensional dan keakuratan data menjadi kendala yang perlu diatasi dalam pengembangan model yang dapat mengklasifikasikan penderita parkinson secara efektif.

1.2.2 Metode Usulan

Penelitian ini mengusulkan penerapan metode *Ensemble Support Vector Machine* (En-SVM) berbasis rata-rata dan mayoritas. Pendekatan ini didukung oleh kombinasi pengolahan data melalui seleksi fitur *Recursive Feature Elimination* (RFE) dan *oversampling* untuk memperoleh struktur data yang akurat.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang telah dijelaskan dan metode yang telah diusulkan, maka didapatkan pertanyaan berapa nilai akurasi, presisi, dan *recall* dari penerapan metode yang diusulkan yaitu *Ensemble Support Vector Machine* (En-SVM) berbasis rata-rata dan mayoritas dalam melakukan klasifikasi penderita parkinson.

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menjelaskan proses klasifikasi penderita parkinson dari penerapan metode *Ensemble Support Vector Machine* (En-SVM) berbasis rata-rata dan mayoritas untuk mendapatkan hasil akurasi, presisi, dan *recall*.

1.3.2 Manfaat Penelitian

Berdasarkan tujuan yang telah diuraikan maka dapat diketahui manfaat penelitian sebagai berikut:

1. Manfaat bagi pihak medis

Manfaat penelitian bagi pihak medis yaitu untuk mempermudah proses dan meminimalisir kesalahan diagnosa pada penderita gejala parkinson

2. Manfaat bagi pembaca

Sebagai bahan studi dan tambahan wawasan pada topik data mining terutama pada metode *Ensemble Support Vector Machine* (En-SVM) berbasis rata-rata dan mayoritas

3. Manfaat bagi penulis

Manfaat penelitian bagi penulis yaitu sebagai penerapan dari ilmu yang telah didapatkan saat menempuh kuliah di Universitas Trunojoyo Madura

1.4 Batasan Masalah

Berdasarkan rumusan masalah yang telah diuraikan maka dapat diketahui beberapa Batasan penelitian sebagai berikut:

1. Penelitian ini menggunakan dataset *Parkinson's Disease Classification*.

Dataset yang dipakai dalam penelitian ini merupakan data besaran frekuensi suara yang diambil dari situs *UCI machine learning repository* (<https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification#>).

2. Dataset terdiri dari 252 individu yang diuji, dimana 188 pasien yang meliputi 107 laki-laki dan 81 perempuan dengan umur rata-rata 33 sampai 87. Terdapat 64 individu yang sehat meliputi 23 laki-laki dan 41 perempuan dengan umur diantara 41 dan 82. Proses pengumpulan data dilakukan dengan melakukan analisis besaran frekuensi dari rekaman suara. Setiap individu diuji sebanyak 3 kali perulangan dan menghasilkan 756 sampel data.

3. Isi dataset terdiri dari gender dan 752 fitur frekuensi suara dari 17 kelompok fitur yaitu *jitter*, *shimmer*, *fundamental frequency*, *harmonicity*, *Recurrence Period Density Entropy* (RPDE), *Detrended Fluctuation Analysis* (DFA),

Pitch Period Entropy (PPE), intensity, formant frequencies, bandwidth, Mel Frequency Cepstral Coefficients (MFCCs), wavelet transform, Glottis Quotient (GQ), Glottal to Noise Excitation (GNE), Vocal Fold Excitation Ratio (VFER), Empirical Mode Decomposition (EMD), Time-Frequency Quotient Wavelet Transform (TQWT). Kelas pada dataset ini menunjukkan subjek yang menderita parkinson (1) dan bukan penderita (0).

1.5 Sistematika Penelitian

Penulisan skripsi ini disusun secara sistematis dan dibagi menjadi beberapa BAB, seperti berikut:

BAB I PENDAHULUAN

BAB ini menjelaskan tentang latar belakang penelitian apa yang akan dilakukan, rumusan masalah, tujuan, batasan masalah, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

BAB ini akan membahas teori tentang penyakit parkinson, fitur frekuensi suara, *data mining*, *StandardScaler*, penyeimbangan data, *Recursive Feature Elimination (RFE)*, *machine learning*, klasifikasi, SVM, kernel pada SVM. *ensemble learning*, *Scikit Learn*, *K-Fold Cross Validation*. *confusion matrix* serta penelitian terdahulu

BAB III METODE PENELITIAN

BAB ini akan membahas tentang metode penelitian yang akan digunakan dalam proses pengumpulan dan pengolahan data menggunakan metode yang sudah ditetapkan

BAB IV PEMBAHASAN

BAB ini akan membahas lebih rinci mengenai penerapan dan hasil klasifikasi parkinson menggunakan metode *Ensemble Support Vector Machine (En-SVM)*

BAB V PENUTUP

BAB ini berisi kesimpulan yang didapatkan dari penelitian dan saran untuk pengembangan penelitian selanjutnya