

PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS POLA KLASIFIKASI PADA PARKINSON'S DATASET

Putri Kurnia Handayani¹

¹Fakultas Teknik Universitas Muria Kudus
Email: ¹putri.kurnia@umk.ac.id

(Naskah masuk: 29 Desember 2021, diterima untuk diterbitkan: 31 Desember 2021)

Abstrak

Parkinson's Disease (PD) merupakan salah satu penyakit yang memiliki angka penderita dan kematian yang meningkat setiap tahunnya. Berdasarkan hasil studi analisis sistematis, angka peningkatan penderita PD mencapai 2.4 kali lipat. Dataset penyakit parkinson merupakan data publik yang diambil dari UCI Repository, yang terdiri dari 9 atribut dan 758 record, meliputi 192 fonasi dari 31 subyek pria dan wanita dimana 23 diantaranya penderita PD. Klasifikasi merupakan salah satu metode supervised learning yang terdapat dalam bidang ilmu data mining yang berfungsi untuk memprediksi objek/data baru yang belum memiliki class/label. Pola klasifikasi dapat dibangun melalui proses analisis pada data training dengan menggunakan algoritma support vector machine (SVM). Algoritma SVM berhasil mengenali pola klasifikasi pada Parkinson's Disease Dataset dengan metode pendekatan PPE. Pendekatan PPE ini mampu meningkatkan kinerja algoritma SVM dalam analisis pola klasifikasi, dibandingkan dengan pendekatan tradisional dan nonstandard dengan akurasi sebesar 95%. Analisis pola klasifikasi menggunakan bantuan software Praat dan Python.

Kata kunci: *klasifikasi, supervised learning, SVM*

IMPLEMENTATION OF PRINCIPAL COMPONENT ANALYSIS FOR IMPROVING DECISION TREE ALGORITHM PERFORMANCE IN IRIS DATASET

Abstract

Parkinson's Disease (PD) is a disease that has an increasing number of sufferers and deaths every year. Based on the results of a systematic analysis study, the rate of increase in patients with PD reached 2.4 times. The Parkinson's disease dataset is public data taken from the UCI Repository, which consists of 9 attributes and 758 records, covering 192 phonations from 31 male and female subjects, of which 23 are patients with PD. Classification is one of the supervised learning methods contained in the field of data mining science which functions to predict new objects/data that do not yet have a class/label. The classification pattern can be built through an analysis process on the training data using the support vector machine (SVM) algorithm. The SVM algorithm has succeeded in recognizing the classification pattern in the Parkinson's Disease Dataset with the PPE approach method. This PPE approach is able to improve the performance of the SVM algorithm in classification pattern analysis, compared to traditional and nonstandard approaches with an accuracy of 95%. Analysis of classification patterns using Praat and Python software.

Keywords: *classification, supervise learning, SVM*

1. PENDAHULUAN

Parkinson's Disease (PD) merupakan salah satu penyakit syaraf yang memburuk secara bertahap dan mempengaruhi bagian otak yang berfungsi mengkoordinasikan gerakan tubuh. Akibatnya penderita kesulitan mengatur gerakan tubuhnya termasuk saat bicara, berjalan dan menulis. Tremor merupakan salah satu manifestasi paling khas dari PD

dan merupakan tanda yang sering muncul yang dialami sekitar 75% pasien PD (Purnamasari, 2019).

Menurut studi analisis sistematis menyatakan pada tahun 2016 terdapat 6.1 juta penderita PD di seluruh dunia, dimana terjadi peningkatan sebesar 2.4 kali lipat. Di Indonesia sendiri penderita PD sebanyak 876.665 pada tahun 2002 dan terus meningkat sampai dengan sekarang (Ranatan, 2019).

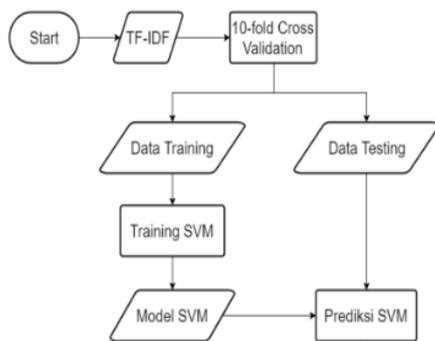
Data mining merupakan kegiatan menambang/menggali data untuk mengenali pola atau aturan tertentu dari sejumlah dataset yang sangat besar dan mempunyai dimensi tinggi. Pola atau aturan yang berhasil ditemukan pada proses penambangan sebelumnya tidak diketahui sehingga memunculkan pengetahuan baru bagi penggunanya.

Penyakit Parkinson (PD) atau dikenal dengan istilah tremor merupakan salah jenis penyakit yang disebabkan oleh gangguan syaraf yang mengakibatkan penderita mengalami tremor, keterlambatan dalam merespon gerak dan kekakuan otot. Penderita Alzheimer lebih rentan mengalami PD disebabkan PD merupakan jenis penyakit neurodegeneratif. Penyakit parkinson dimulai perlahan, tidak disadari, berangsur-angsur memburuk dan mempengaruhi kualitas hidup. Penyakit ini semakin meningkat setiap tahunnya (Hanriko & Anzani, 2018).

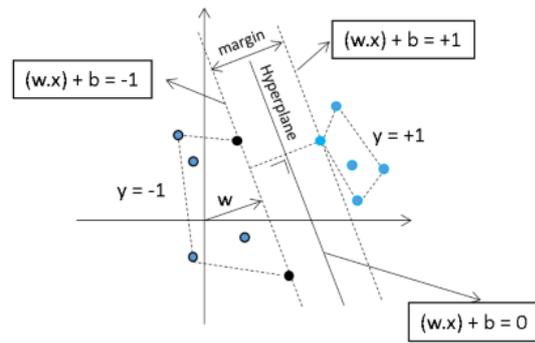
Penderita PD lebih sering terjadi pada laki-laki. PD menjadi salah satu jenis penyakit yang masih sering diabaikan di masyarakat, dikarenakan masih rendahnya angka kematian yang disebabkan oleh PD. Penyakit parkinson memiliki etiologi yang belum diketahui secara pasti, namun dapat disebabkan oleh faktor-faktor seperti faktor genetik, faktor lingkungan, umur, ras, cedera kranioserebral dan stress emosional (Brown, et al., 2006).

Support Vector Machine (SVM) merupakan salah satu algoritma yang dapat digunakan dalam proses data mining untuk klasifikasi. Termasuk ke dalam supervised learning, SVM mengenali pola dari dataset berdasarkan data training yang ada. Pola klasifikasi yang berhasil diidentifikasi dengan menggunakan algoritma tertentu, digunakan untuk memprediksi objek yang belum dikenali kelas/labelnya.

Klasifikasi dilakukan dengan mencari hyperplane atau garis pembatas (decision boundary) yang memisahkan antara suatu kelas dengan kelas lain, yang dalam kasus ini garis tersebut berperan memisahkan tweet bersentimen positif (berlabel +) dengan tweet bersentimen negatif (berlabel -) (Han & Kamber, 2006). Skema dari proses klasifikasi dapat dilihat gambar 1 dan gambar 2.

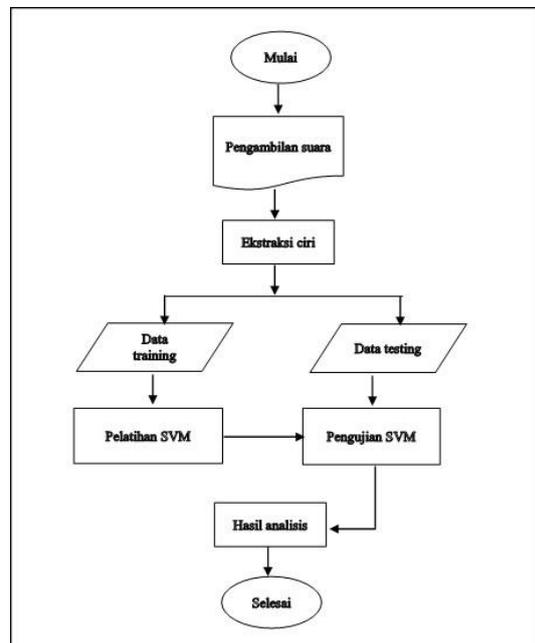


Gambar 1 Klasifikasi dengan SVM (Sumber: Han & Kamber, 2006)



Gambar 2 Hyperplane terbentuk diantara kelas -1 dan +1

2. METODE PENELITIAN



Gambar 3 Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

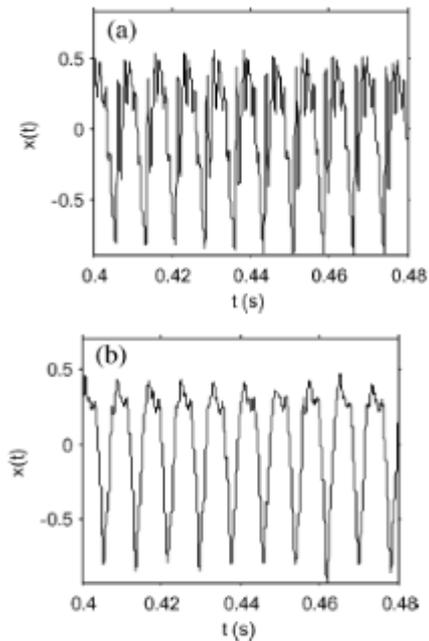
3.1. Pengumpulan Data

Data untuk penelitian ini terdiri dari 195 fonasi vokal berkelanjutan dari 31 subyek pria dan wanita, 23 di antaranya didiagnosis dengan PD. Waktu sejak diagnosis berkisar antara 0 hingga 28 tahun, dan usia subyek berkisar antara 46 hingga 85 tahun (rata-rata 65,8, standar deviasi 9,8). Rata-rata enam fonasi direkam dari masing-masing subyek, mulai dari 1 hingga 36 detik. Lihat Tabel 1 untuk rincian subyek. Gambar 4 menunjukkan plot dari dua sinyal suara ini. Suara tersebut direkam di bilik yang diberi perlakuan suara oleh menggunakan mikrofon yang dipasang di kepala (AKG C420) yang ditempatkan pada jarak 8 cm dari bibir. Mikrofon dikalibrasi seperti yang dijelaskan menggunakan pengukur tingkat suara kelas 1 (B&K 2238) yang ditempatkan 30 cm dari speaker. Sinyal suara direkam langsung di komputer

menggunakan perangkat keras Computerized Speech Laboratory (CSL) 4300B (Kay Elemetrics), sampel pada 44,1 kHz, dengan resolusi 16-bit. Meskipun normalisasi amplitudo mempengaruhi kalibrasi sampel, penelitian ini difokuskan pada pengukuran yang tidak sensitif terhadap perubahan tingkat tekanan bicara absolut. Jadi, untuk memastikan kekokohan algoritma, semua sampel dinormalisasi secara digital dalam amplitudo sebelum penghitungan pengukuran.

Tabel 1 Daftar subyek dengan jenis kelamin, usia, tahap Parkinson dan jumlah tahun sejak diagnose
Sumber: (Little, et al., 2009)

No	Subject code	Sex	Age	Stage (H&Y)	Years since diagnosis
1	S01	M	78	3.0	0
2	S34	F	79	2.5	¼
3	S44	M	67	1.5	1
4	S20	M	70	3.0	1
5	S24	M	73	2.5	1
6	S26	F	53	2.0	1½
7	S08	F	48	2.0	2
8	S39	M	64	2.0	2
9	S33	M	68	2.0	3
10	S32	M	50	1.0	4
11	S02	M	60	2.0	4
12	S22	M	60	1.5	4½
13	S37	M	76	1.0	5
14	S21	F	81	1.5	5
15	S04	M	70	2.5	5½
16	S19	M	73	1.0	7
17	S35	F	85	4.0	7
18	S05	F	72	3.0	8
19	S18	M	61	2.5	11
20	S16	M	62	2.5	14
21	S27	M	72	2.5	15
22	S25	M	74	3.0	23
23	S06	F	63	2.5	28
24	S10 (healthy)	F	46	n/a	n/a
25	S07 (healthy)	F	48	n/a	n/a
26	S13 (healthy)	M	61	n/a	n/a
27	S43 (healthy)	M	62	n/a	n/a
28	S17 (healthy)	F	64	n/a	n/a
29	S42 (healthy)	F	66	n/a	n/a
30	S50 (healthy)	F	66	n/a	n/a
31	S49 (healthy)	F	69	n/a	n/a



Gambar 3 Dua contoh sinyal suara yang dipilih (a) sehat (b) subyek dengan PD. Sumbu horizontal adalah waktu dalam detik dan sumbu vertical adalah amplitudo sinyal (tidak ada satuan)

3.2. Feature Calculation Stage

Tahap perhitungan fitur melibatkan penerapan pemilihan perwakilan metode pengukuran tradisional (calculation of traditional measures) dan tidak standar (calculation of nonstandard measures) untuk semua sinyal suara. Setiap metode menghasilkan satu nomor untuk masing-masing dari 195 sinyal. Lihat Tabel 2 untuk daftar ukuran yang digunakan sebagai fitur dalam penelitian ini.

Tabel 2 Daftar metode pengukuran yang diterapkan pada sinyal akustik yang tercatat dari setiap subyek
Sumber: (Little, et al., 2009)

Feature	Retained after filtering?	Description
MDVP: Jitter (%)	No	Kay pentax MDVP jitter as a percentage
MDVP: Jitter (Abs)	Yes	Kay pentax MDVP absolute jitter in microsecond
MDVP: RAP	No	Kay pentax MDVP relative amplitude perturbation
MDVP: PPQ	No	Kay pentax MDVP five point periode perturbation quotient
Jitter: DDP	Yes	Aveareg absolute difference of differences between cycles, divided by the average period
MDVP: shimmer	No	Kay pentax MDVP local shimmer
MDVP: shimmer (dB)	No	Kay pentax MDVP local shimmer in decibels
Shimmer: APQ3	No	Three point amplitude perturbation quotient
Shimmer: APQ5	No	Five point amplitude perturbation quotient

MDVP: APQ	Yes	Kay pentax MDVP 11-point amplitude perturbation quotient
Shimmer: DDA	Yes	Average absolute difference between consecutive differences between the amplitudes of consecutive periods
NHR	Yes	Noise-to harmonics ratio
HNR	Yes	Harmonics to noise ratio
RPDE	Yes	Reccurence period density entropy
DFA	Yes	Detrended fluctuation analysis
D2	Yes	Correlation dimension
PPE	Yes	Pitch period entropy

3.3. Feature Preparation and Classification Stage

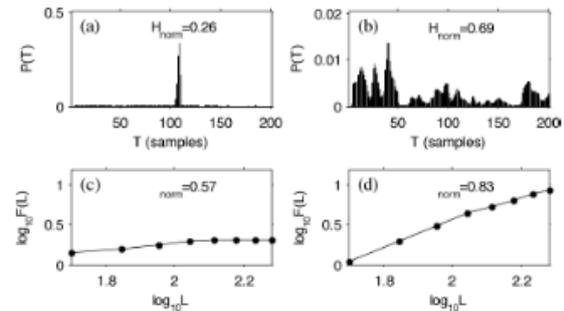
Kinerja klasifikasi SVM dapat ditingkatkan melalui preprocessing nilai setiap ukuran dengan skala yang sesuai. Dalam penelitian ini proses penskalaan setiap ukuran dengan rentang [-1, 1]. Pada tahapan ini juga dilakukan proses filter jumlah ukuran ke dalam ukuran yang dapat diatur untuk menentukan himpunan yang optimal pada proses klasifikasi.

Hasil pengamatan menunjukkan banyak tindakan berkorelasi dengan tindakan lain, disebabkan pengukuran aspek yang sangat mirip dari sinyal ucapan (tabel 1). Dengan adanya korelasi ini, maka hanya ada satu dari pasangan ukuran yang akan menyumbangkan informasi yang berguna untuk tahap klasifikasi dan yang lain harus dihilangkan. Oleh karena itu dilakukan pencarian melalui semua pasang fitur. Dari pasangan ukuran yang berkorelasi dengan koefisien korelasi lebih besar dari 0.95, kemudian dilakukan penghapusan salah satunya. Selanjutnya membangun vector fitur dengan setiap kombinasi yang mungkin dari himpunan bagian dari ukuran yang telah diproses sebelumnya dan dilakukan filtering. Untuk setiap kombinasi diterapkan klasifikasi dengan SVM dengan menggunakan formulasi kernel-SVM dengan fungsi kernel dasar radial Gaussian. Penggunaan kernel ini dikarenakan lebih fleksibel yang memungkinkan batas keputusan yang mulus dan melengkung. Untuk setiap kombinasi fitur, kinerja klasifikasi dinilai dengan benar sebagai sehat atau PD, jumlah PWP yang diklasifikasikan dengan benar (tingkat positif sejati) dan jumlah subyek sehat yang diklasifikasikan dengan benar (nilai benar tingkat negative). Validasi hasil untuk mendapatkan perkiraan kinerja di luar sampel dan interval kepercayaan diukur dengan menggunakan bootstrap resampling dengan 50 pengulangan. Pilihan nilai pinalti SVM yang optimal dan bandwith kernel ditentukan oleh pencarian menyeluruh pada rentang nilai. Klasifikasi bootstrap menghasilkan satu set hasil kinerja klasifikasi untuk setiap pengulangan bootstrap. Untuk menentukan subset fitur dengan kinerja terbaik, dengan membandingkan set hasil klasifikasi keseluruhan menggunakan uji rank-sum Wilcoxon dua sisi,

terhadap hipotesis nol dari median yang sama pada probabilitas signifikan 0.05.

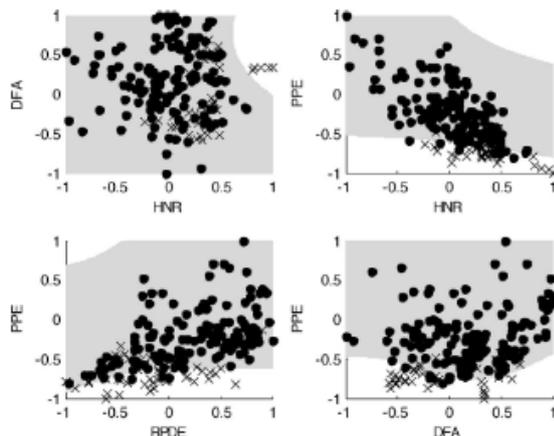
3.4. Hasil

Untuk tahap feature calculation menghasilkan grafik yang menunjukkan hasil perhitungan RDPE dan DFA untuk beberapa sinyal suara yang dipilih (gambar 4).



Gambar 4 Grafik hasil perhitungan RDPE dan DFA

Pada grafik tersebut dapat dilihat hasil RDPE dan DFA untuk subyek sehat (panel kiri) dan untuk subyek dengan PD (panel kanan). Gambar (a) dan (b) adalah kerapatan pengulangan P(T) untuk waktu pengulangan T. Sedangkan gambar (c) dan (d) merupakan plot log-log penskalaan ukuran L terhadap amplitude fluktuasi F(L).



Gambar 5 Hasil klasifikasi SVM

Batas klasifikasi SVM untuk beberapa pasangan fitur yang dipilih setelah preprocessing dengan normalisasi rentang (tabel 2). Tanda "x" untuk subyek sehat dan tanda bulat untuk subyek PD. Daerah yang diarsir abu-abu muda merupakan daerah dimana subyek diprediksi menderita Parkinson (gambar 5).

4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) dapat digunakan untuk menganalisa pola klasifikasi untuk parkinson's disease dataset. Kinerja dari algoritma SVM meningkat melalui pendekatan PPE dengan hasil akurasi sebesar 95%.

DAFTAR PUSTAKA

- Brown, T. P. et al., 2006. Pesticides and Parkinson's Disease - Is There a Link?. *Environmental Health Perspectives*, Volume 114.
- Han, J. & Kamber, M., 2006. *Data Mining Concepts and Techniques*. Second ed. San Fransisco: Morgan Kaufmann.
- Hanriko, R. & Anzani, B. P., 2018. Penyakit Parkinson: Ancaman Kesehatan bagi Komunitas Pertanian.
- Little, M. A. et al., 2009. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, Volume 56, pp. 1015-1022.
- Purnamasari, W. I., 2019. *Jenis Tremor pada Penyakit Parkinson: Studi Deskriptif Menggunakan Klasifikasi Baru*, s.l.: s.n.
- Ranatan, C., 2019. *www.alomedika.com*. [Online] Available at: <https://www.alomedika.com/penyakit/neurologi/penyakit-parkinson>