

# KLASIFIKASI PENDERITA PENYAKIT PARKINSON MENGGUNAKAN METODE GAUSSIAN BAYES

Ika Purwatiningsih, Agung Nilogiri

## PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH JEMBER

[Ikapurwatiningsih1@gmail.com](mailto:Ikapurwatiningsih1@gmail.com), [agungnilogiri@unmuahjember.ac.id](mailto:agungnilogiri@unmuahjember.ac.id)

### ABSTRAK

Penyakit Parkinson merupakan penyakit neurodegeneratif terbanyak kedua setelah penyakit Alzheimer, penyakit tersebut menyebabkan penderitanya mengalami beberapa gejala diantaranya gangguan intelektual dan tingkah laku, demensia, penurunan daya ingat, kelambatan gerak otot, katalepsi, dan tremor. Katalepsi merupakan kekakuan otot yang diketahui cirinya apabila penekukan lengan bawah atau pelurusan oleh orang lain maka akan berakibat kaku. Pada penelitian ini peneliti menggunakan teknik data mining terhadap data parkinson. Dalam penelitian ini membahas klasifikasi penyakit parkinson menggunakan metode Gaussian Bayes. Dataset yang digunakan pada penelitian ini berasal dari Kaggle dengan jumlah fitur sebanyak 24 yang diperoleh dari pemeriksaan pasien sebanyak 757 orang. Pada penelitian ini juga menggunakan skenario uji K Fold Cross Validation dengan nilai K yaitu 2, 4, 5 dan 10 dengan total 700 data dan data validasi 57. Penentuan kriteria pada hasil klasifikasi terhadap output data asli (nilai aktual) menggunakan confusion matrix. Dari pengukuran hasil klasifikasi diperoleh hasil yaitu nilai akurasi tertinggi pada penelitian ini adalah 64,29% yang dihasilkan dari skema 10 fold skenario 10 sedangkan nilai presisi tertinggi pada penelitian ini diperoleh 83,33% yang dihasilkan dari skema 10 fold skenario 2. Pada pengujian data validasi diperoleh hasil akurasi sebesar 64,91%, jika dibandingkan dari hasil k Fold maka terjadi peningkatan sebesar 0,63%.

Kata kunci : Parkinson, Data Mining, Klasifikasi, Gaussian Bayes.

### I Pendahuluan

#### 1. Latar Belakang

Penyakit Parkinson merupakan penyakit yang melibatkan proses degeneratif pada neuron dopaminergik dalam substansia nigra yaitu pada area organ otak bagian ganglia basalis yang menghasilkan dan menyimpan neurotransmitter dopamin (Silitonga, 2007).

Penerapan data mining sudah dilakukan pada beberapa bidang diantaranya bisnis, ilmu pengetahuan dan kesehatan. Pada bidang kesehatan, data mining dapat dimanfaatkan untuk memprediksi suatu penyakit dari data rekam medis pasien. Dengan metode klasifikasi pada data mining, data seperti umur, jenis kelamin, tekanan darah dan atribut lainnya, dapat digunakan untuk memprediksi kemungkinan pasien terkena suatu penyakit.

#### 2. Rumusan Masalah

Penelitian ini menitikberatkan pada permasalahan yang antara lain:

- Berapa tingkat presisi tertinggi yang dihasilkan oleh metode gaussian bayes dengan skenario uji K fold cross validation?
- Berapa hasil akurasi setelah diujikan pada data validasi?

#### 3. Tujuan

Tujuan dari penelitian ini antara lain:

- Mengetahui tingkat presisi tertinggi yang dihasilkan metode gaussian bayes dalam mengklasifikasi data.
- Mengetahui fluktuatif perubahan tingkat akurasi metode dari pemodelan terhadap uji validasi.

#### 4. Manfaat

Penelitian ini juga diharapkan memiliki manfaat terhadap peneliti dan masyarakat luas. Manfaat dari penelitian yaitu pengembangan pada data mining terhadap dunia medis khususnya terhadap penelitian penderita parkinson

agar dapat ditangani sejak dini. Penelitian ini juga dapat dijadikan bahan pertimbangan dalam pengembangan untuk menemukan metode terbaik dalam klasifikasi penderita parkinson.

#### 5. Batasan Masalah

Penelitian ini juga memiliki batasan agar fokus masalah dapat terjawab, berikut beberapa batasan masalah yaitu:

- Data yang dipakai dalam penelitian ini berasal dari kaggle pada laman <https://www.kaggle.com/kralmachine/parkinson-disaster> dengan total data 757 dan skenario uji menggunakan k fold cross validation dengan nilai k = 2, 4, 5, 7 dan 10.
- Pengukuran yang dilakukan pada tingkat akurasi dan presisi.
- Atribut pada penelitian ini diperoleh dari hasil pemeriksaan perekam suara, atribut-atributnya adalah MDVP: Fo(Hz), MDVP: Fhi(Hz), MDVP: Flo(Hz), MDVP: Jitter(%), MDVP:Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP, MDVP: Shimmer, MDVP: Shimmer (dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA, NHR, HNR, RPDE, DFA, Spread1, Spread2, D1, D2, PPE.
- Output class terdiri dari 0 yaitu sehat dan 1 yaitu terkena parkinson.

### II Metode Penelitian

#### 1. Gambaran data

Data diperoleh dari Time Frequency yang diekstrak dari Voice recording. Dataset ini berasal dari penelitian Max Little dari kolaborasi University of Oxford dan National Centre for Voice and Speech, Denver, Colorado. Total dataset pada penelitian ini adalah 757 data, terbagi atas 291 data terperiksa sehat dan 466 pasien terperiksa penderita parkinson.

## 2. Pemodelan

Pada tahap pemodelan persamaan yang digunakan adalah prior probability, mean dan standar deviasi. Berikut rumus persamaannya.

$$prior\ probability (A|X) \frac{\sum A|X}{\sum X}$$

$$mean(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$StandardDeviation(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - mean(x))^2}{n-1}}$$

## 3. Pengujian

Pada tahap pengujian pada penelitian ini menggunakan beberapa langkah yaitu klasifikasi menggunakan PDF (persamaan gaussian bayes) dan confusion matrix. Berikut persamaan yang digunakan.

$$pdf(x, meansd) = \frac{1}{\sqrt{2\pi}sd} e^{-\frac{(x-mean)^2}{2sd^2}}$$

## 4. Pengukuran

Pada tahap pengukuran peneliti menggunakan pengukuran akurasi dan presisi, berikut persamaan yang digunakan.

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\%$$

$$Presisi = \frac{(TP)}{(TP+FP)} \times 100\%$$

## III Implementasi dan Hasil

### 1. Hasil akurasi K Fold

Dari hasil klasifikasi dan pengukuran tingkat akurasi pada skenario uji K Fold Cross Validation dengan k = 2, 4, 5, 7 dan 10 berikut hasil yang diperoleh.

Tabel III.1 hasil pengukuran akurasi

No.	Uji Ke	TP	TN	FP	FN	AKurasi
1	K 2.1	110	90	43	107	57.143
2	K 2.2	115	94	42	99	59.714
3	K 4.1	58	42	21	54	57.143
4	K 4.2	53	50	20	52	58.857
5	K 4.3	60	45	15	55	60
6	K 4.4	51	56	20	48	61.143
7	K 5.1	44	32	18	46	54.286
8	K 5.2	45	38	15	42	59.286
9	K 5.3	41	39	18	42	57.143
10	K 5.4	46	37	12	45	59.286
11	K 5.5	40	42	18	40	58.571
12	K 7.1	30	24	15	31	54
13	K 7.2	36	23	12	29	59
14	K 7.3	32	30	9	29	78.049
15	K 7.4	30	27	12	31	71.429
16	K 7.5	35	26	10	29	77.778
17	K 7.6	33	30	7	30	82.5
18	K 7.7	27	29	15	29	64.286
19	K 10.1	15	21	12	22	55.556
20	K 10.2	30	11	6	23	83.333
21	K 10.3	25	17	10	18	71.429
22	K 10.4	21	19	7	23	75
23	K 10.5	20	20	10	20	66.667
24	K 10.6	22	20	7	21	75.862
25	K 10.7	26	16	6	22	81.25
26	K 10.8	21	22	5	22	80.769
27	K 10.9	18	22	8	22	69.231
28	K 10.10	23	22	8	17	74.194

24	K 10.6	22	20	7	21	60
25	K 10.7	26	16	6	22	60
26	K 10.8	21	22	5	22	61.429
27	K 10.9	18	22	8	22	57.143
28	K 10.10	23	22	8	17	64.286

Dari tabel di atas dapat diketahui bahwa nilai akurasi tertinggi diperoleh pada skenario uji K10.10 dengan nilai 64,29%.

## 2. Hasil Presisi K Fold

Dari hasil klasifikasi dan pengukuran tingkat presisi pada skenario uji K Fold Cross Validation dengan k = 2, 4, 5, 7 dan 10 berikut hasil yang diperoleh.

Tabel III.2 Hasil Pengukuran presisi

No.	Uji Ke	TP	TN	FP	FN	Presisi
1	K 2.1	110	90	43	107	71.895
2	K 2.2	115	94	42	99	73.248
3	K 4.1	58	42	21	54	73.418
4	K 4.2	53	50	20	52	72.603
5	K 4.3	60	45	15	55	80
6	K 4.4	51	56	20	48	71.831
7	K 5.1	44	32	18	46	70.968
8	K 5.2	45	38	15	42	75
9	K 5.3	41	39	18	42	69.492
10	K 5.4	46	37	12	45	79.31
11	K 5.5	40	42	18	40	68.966
12	K 7.1	30	24	15	31	66.667
13	K 7.2	36	23	12	29	75
14	K 7.3	32	30	9	29	78.049
15	K 7.4	30	27	12	31	71.429
16	K 7.5	35	26	10	29	77.778
17	K 7.6	33	30	7	30	82.5
18	K 7.7	27	29	15	29	64.286
19	K 10.1	15	21	12	22	55.556
20	K 10.2	30	11	6	23	83.333
21	K 10.3	25	17	10	18	71.429
22	K 10.4	21	19	7	23	75
23	K 10.5	20	20	10	20	66.667
24	K 10.6	22	20	7	21	75.862
25	K 10.7	26	16	6	22	81.25
26	K 10.8	21	22	5	22	80.769
27	K 10.9	18	22	8	22	69.231
28	K 10.10	23	22	8	17	74.194

Dari tabel di atas dapat diketahui bahwa nilai presisi tertinggi diperoleh pada skenario uji K10.2 dengan nilai 83,33%.

## 3. Hasil Validasi

Dari hasil klasifikasi dan pengukuran pada data validasi dengan jumlah data 57 dan menggunakan skenario terbaik pada k fold yaitu K10.10 diperoleh hasil berikut

**Tabel III.3 hasil validasi**

Kriteria	Jumlah
TP	23
TN	13
FP	9
FN	12
Akurasi	64.91%
Presisi	72.72%

#### IV Penutup

##### 1. Kesimpulan

Dari penelitian implementasi metode Gaussian Bayes terhadap data pasien penderita parkinson dengan total data 757 data, dimana pada penelitian ini menggunakan skenario uji k fold cross validation dengan nilai k = 2, 4, 5, 7 dan 10 diperoleh beberapa kesimpulan yaitu:

- Tingkat presisi tertinggi yang diperoleh dari skenario pengujian k fold cross validation yaitu 83,33% yang dihasilkan dari skema 10 fold skenario 2. Tingkat akurasi tertinggi yang diperoleh dari skenario pengujian k fold cross validation yaitu 64,29% yang dihasilkan dari skema 10 fold skenario 10.
- Hasil akurasi yang diperoleh dari uji validasi yaitu 64,91%. Jika dibandingkan dengan hasil yang diperoleh pada skema 10 skenario 10 yaitu 64,29% terjadi peningkatan 0,63%.

##### 2. Saran

Peneliti selaku penulis membuka lebar kepada para pembaca untuk mengembangkan penelitian ini. berikut hal-hal yang dapat dilakukan pengembangan dalam penelitian ini:

- Pengembang dapat menggunakan metode klasifikasi lain untuk menemukan tingkat akurasi dan presisi lebih baik.
- Pengembang dapat menggunakan teknik ensemble untuk menemukan atau meningkatkan nilai akurasi dan presisi.
- Pengembang dapat menggunakan metode preprocessing lain n untuk mendapatkan hasil lebih baik.

#### Daftar Pustaka

- Olanow, C. W., Stocchi, F., & Lang, A. (Eds.). (2011). Parkinson's disease: non-motor and non-dopaminergic features. John Wiley & Sons.
- Silitonga, R. (2007). Faktor-Faktor Yang Berhubungan Dengan Kualitas Hidup Penderita Penyakit Parkinson Di Poliklinik Saraf Rs Dr Kariadi Factors associate with Quality of Life on Parkinson Disease in Neurology Out Patient Department of Dr Kariadi Hospital (Doctoral dissertation, Program Pasca Sarjana Universitas Diponegoro).
- Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & Dubois, B. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and

clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society, 23(15), 2129-2170.

- Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. 2001. San Francisco: Morgan Kauffman.
- Subbalakshmi, G., Ramesh, K., & Rao, M. C. (2011). Decision support in heart disease prediction system using naive bayes. Indian Journal of Computer Science and Engineering (IJCSE), 2(2), 170-176.
- Smeltzer, S. C., Bare, B. G., Hinkle, J. L., & Cheever, K. H. (2005). Tratado de enfermagem médica-cirúrgica.
- Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
- Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.
- Connolly T., C. Begg and A. Strachan (1999) Database systems: A Practical Approach to Design, Implementation, and Management (3rd Ed.). Harlow: Addison-Wesley.687
- Vercellis, C. (2009). Business intelligence: data mining and optimization for decision making. New York: Wiley.
- Hoffer, J. A., Venkataraman, R., & Topi, H. (2009). Modern database management, 11/e.
- Turban, E. (2007). Information technology for management: Transforming organizations in the digital economy. John Wiley & Sons, Inc..
- Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. Springer Science & Business Media.
- Aydin, F., & Aslan, Z. (2017). Classification of Neurodegenerative Diseases using Machine Learning Methods. International Journal of Intelligent Systems and Applications in Engineering, 5(1), 1-9.
- Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993, April). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In IEE proceedings F (radar and signal processing) (Vol. 140, No. 2, pp. 107-113). IET Digital Library.
- Williams, C. K., & Barber, D. (1998). Bayesian classification with Gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), 1342-1351.
- Ramani, R. G., & Sivagami, G. (2011). Parkinson disease classification using data mining algorithms. International journal of computer applications, 32(9), 17-22.
- Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.
- Prasetyo, E. (2014). Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga. Prosiding SNATIF, 293-300.

Haltuf, M., 2014, Support Vector Machines for Credit Scoring, Thesis, Faculty of Finance University of Economics in Prague, Prague.

Jason Brownlee. (2016). Master Machine Learning Algorithms : Discover How They Work and Implement Them From Scratch.

Mudali, D., Teune, L. K., Renken, R. J., Leenders, K. L., & Roerdink, J. B. (2015). Classification of Parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features. Computational and mathematical methods in medicine, 2015.

Morales, D. A., Vives-Gilabert, Y., Gómez-Ansón, B., Bengoetxea, E., Larrañaga, P., Bielza, C., ... & Delfino, M. (2013). Predicting dementia development in Parkinson's disease using Bayesian network classifiers. Psychiatry Research: NeuroImaging, 213(2), 92-98.

Aich, S., Younga, K., Hui, K. L., Al-Absi, A. A., & Sain, M. (2018, February). A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data. In 2018 20th International Conference on Advanced Communication Technology (ICACT) (pp. 638-642). IEEE.

Jebakumari, V. S., Shanthi, D., Sridevi, S., & Meha, P. (2017, March). Performance evaluation of various classification algorithms for the diagnosis of Parkinson's disease. In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-7). IEEE.

