



KOMPARASI METODE *SUPPORT VECTOR MACHINE* DAN *NAÏVE BAYES* DALAM KLASIFIKASI PELUANG PENYAKIT SERANGAN JANTUNG

Musthofa Galih Pradana¹, Dhina Puspasari Wijaya², Pujo Hari Saputro³

¹Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta,

²Program Studi Informatika, Fakultas Komputer dan Teknik, Universitas Alma Ata,

³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Sam Ratulangi.

musthofagalihpradana@upnvj.ac.id , dhina.puspa@almaata.ac.id , pujoharisaputro@unsrat.ac.id

¹Jl RS. Fatmawati Raya, Jakarta Selatan,

²Jl. Brawijaya No.99 Yogyakarta,

³Bahu, Malalayang, Manado.

Keywords:

Classification, Naïve Bayes, Support Vector Machine, Heart Disease

Abstract

The death rate in the world per year is 17.9 million due to cardiovascular disease, including heart and blood vessel disorders. This needs to be given more attention to anticipate the possible risk of a heart attack. One of the contributions in the field of technology to provide useful information about the risk of heart disease is by using a data processing approach or data mining technique by classifying the vulnerability to heart disease risk. The classification method used is Support Vector Machine and Naïve Bayes. The classification method will be carried out in a comparative process and the method that has the best accuracy will be sought. The scenarios used are 2 test scenarios, namely dividing the training data by 20% in scenario 1 and 40% in scenario 2. The final results of the research obtained are the best accuracy in the Support Vector Machine with scenario 1 of 87%.

Kata Kunci:

Klasifikasi, Naïve Bayes, Support Vector Machine, Penyakit Jantung

Abstrak

Angka kematian di dunia per tahun sebesar 17,9 juta diakibatkan oleh penyakit kardiovaskular termasuk dalam gangguan jantung dan pembuluh darah. Ini perlu diberikan perhatian lebih untuk mengantisipasi kemungkinan resiko penyakit serangan jantung. Salah satu kontribusi dalam bidang teknologi untuk memberikan informasi bermanfaat mengenai resiko penyakit jantung adalah dengan pendekatan olah data atau teknik *data mining* dengan melakukan klasifikasi kerentanan resiko penyakit jantung. Metode klasifikasi yang digunakan adalah *Support Vector Machine* dan *Naïve Bayes*. Metode klasifikasi akan dilakukan proses komparasi dan dicari metode yang memiliki akurasi terbaik. Skenario yang digunakan sejumlah 2 skenario pengujian, yaitu membagi data *training* sebesar 20% pada skenario 1 dan 40% pada skenario 2. Hasil akhir penelitian yang didapatkan adalah akurasi terbaik pada *Support Vector Machine* dengan skenario 1 sebesar 87%.

Pendahuluan

Menurut data, sekitar 17,9 juta kematian disebabkan oleh penyakit kardiovaskular di seluruh dunia setiap tahunnya, yang merupakan kelompok penyakit kardiovaskular. Penyakit kardiovaskular adalah penyebab kematian nomor satu di dunia. Serangan jantung dan stroke adalah penyebab 4 dari 5 kematian pada pasien dengan penyakit

kardiovaskular dengan kasus mencapai sepertiga dari total kematian yang disebabkan oleh penyakit kardiovaskular [1]. Data telah menunjukkan bahwa penyakit jantung merupakan penyakit yang mematikan, sehingga perlu kewaspadaan dari setiap orang akan resiko penyakit jantung ini. Salah satu kontribusi dalam bidang teknologi untuk memberikan informasi bermanfaat mengenai resiko penyakit jantung adalah dengan

pengolahan data atau teknik *data mining*. Dalam *data mining* ada beberapa pendekatan seperti klusterisasi, klasifikasi, asosiasi. Pada penelitian ini dilakukan teknik klasifikasi yaitu pengelompokan data, data yang dikelompokkan berlandaskan relasi atau hubungan data terhadap data sampel [2].

Landasan Teori

Studi banding pertama Hennie Tuhuteru yang membandingkan Naive Bayes dan SVM dalam analisis opini. Berdasarkan hasil penelitian ini, akurasi rata-rata klasifikasi metode SVM lebih baik dibandingkan dengan metode NBC yaitu sebesar 76,42%. [3]. Selanjutnya, imam riadi juga pernah menuliskan penelitian sejenis dengan melakukan komparasi pada kasus deteksi anomaly *traffic* yang menghasilkan kesimpulan nilai maksimum dalam analisis komparatif menggunakan metode Naive Bayes [4]. Serupa dengan penelitian rujukan sebelumnya, pada penelitian Adi Supriyatna Naive bayes juga memiliki akurasi yang lebih baik dibandingkan dengan SVM pada klasifikasi keberhasilan imunoterapi pada penyakit kutil dengan alasan Naive bayes dapat melakukan prediksi terhadap semua instance class secara tepat dengan tingkat akurasinya adalah 1 [5]. Hasil berbeda tersaji pada penelitian Oki Arifin yang melakukan perbandingan antara metode SVM dan Naive Bayes pada klasifikasi peminatan jalur siswa SMA, hasil penelitian ini dapat disimpulkan bahwa algoritma SVM dengan kernel anova dan parameter C sebesar 5.0 relatif lebih unggul dibandingkan algoritma Naive Bayes Classifier [6]. Hajah T. Sueno dalam tulisannya menyampaikan hasil penelitian menunjukkan bahwa vektorisasi Naive Bayes teknik telah menyumbangkan data tekstual yang lebih efektif proses transformasi ke classifier SVM, dibandingkan dengan penggunaan teknik vektorisasi TFIDF untuk hal yang sama [7]. *Intrusion detection* yang pernah dipublikasikan Jie Gu melakukan penggabungan kedua metode bukan teknik komparasi, hasilnya adalah percobaan pada beberapa dataset dalam domain deteksi intrusi memvalidasi bahwa metode deteksi yang diusulkan dapat mencapai kinerja yang baik dan kuat, dengan akurasi 93,75% pada dataset UNSW-NB15, akurasi 98,92% pada dataset CICIDS2017, akurasi 99,35% pada dataset NSL-KDD dan akurasi

98,58% pada kumpulan data Kyoto 2006+ [8]. Penelitian dari Yogendra Narayan yang melakukan perbandingan metode SVM dan Naive Bayes menunjukkan bahwa SVM mencapai akurasi sebesar 95,8% dan ditemukan lebih baik dibandingkan dengan pengklasifikasi Naive Bayes. Hasilnya menunjukkan bahwa SVM memiliki respon yang cepat, lebih sedikit kesalahan dan sangat cocok untuk klasifikasi sinyal sEMG untuk perancangan teknologi robot bantu [9]. Opini mengenai vaksin covid terhadap sentiment yang muncul di masyarakat dengan SVM dan Naive Bayes pada penelitian Frizka Fitriana menghasilkan algoritma inferensi SVM memiliki kinerja yang lebih baik dalam hal akurasi, presisi dan recall dengan skor 90.47%, 90.23%, 90.78, sedangkan kinerja algoritma Naive Bayes adalah 88.64%, 87.32%, 88.13%. Selisih presisi 1,83%, presisi 2,91 recall 2,65%[10]. Sentiment lainnya pernah dituliskan Dinar Ajeng Kristianti pada perbandingan metode SVM dan Naive Bayes. Hasil menunjukkan bahwa Algoritma Naive Bayes Classifier (NBC) memiliki tingkat akurasi yang lebih tinggi dari Support Vector Machine (SVM), hingga 94% untuk Deddy Mizwar-Dedi Mulyadi pada topik kandidat gubernur jawa barat tahun 2018 [11]. Hasil keunggulan metode SVM dibanding Naive Bayes juga terjadi pada kasus review produk amazon yang menunjukkan hasil eksperimen telah mengkonfirmasi bahwa SVM dapat mempolarisasi umpan balik produk Amazon dengan tingkat akurasi yang lebih tinggi [12]. Referensi terakhir adalah dari Sri Widaningsih yang melakukan penelitian untuk memprediksi nilai mahasiswa dan waktu kelulusan dengan membandingkan empat algoritma yaitu C4.5, Naive Bayes, KNN dan SVM dan hasil akhir Naive Bayes sebagai algoritma terbaik yang diterima untuk memprediksi mahasiswa. Skala waktu dan IPK ≥ 3 dengan akurasi (76,79%), error (23,17%) dan AUC (0,850) [13].

a. Naive Bayes

Naive Bayes adalah model klasifikasi statistik yang difungsikan dalam melakukan prediksi probabilitas keanggotaan kelas. Naive Bayes didasarkan pada teorema Bayes dengan karakteristik klasifikasi yang mirip dengan pohon keputusan dan jaringan saraf. [14].

Adapun rumus kedekatan pada Naïve Bayes ditunjukkan persamaan 1 :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
- P(H) : Probabilitas hipotesis H (prior probability)
- P(X) : Probabilitas dari X

b. Support Vector Machine

Model persamaan SVM ditunjukkan pada persamaan 2 [15].

$$f:w.x+b=0 \quad (2)$$

Dimana :

w = parameter hyperplane yang dicari (garis yang tegak lurus antara garis hyperplane dan titik support vector)

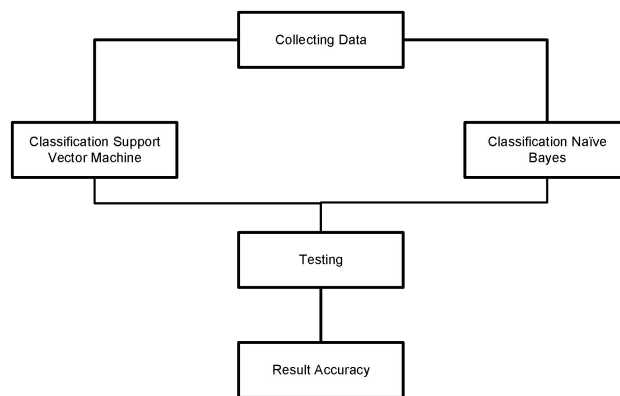
x = data input SVM (x1 = index kata, x2= bobot kata)

b = parameter hyperplane yang dicari (nilai bias)

f = fungsi Hyperplane

Metode

Alur yang digunakan dalam penelitian ini adalah dengan melakukan pengumpulan atau pengambilan data terlebih dahulu, setelah itu dilakukan klasifikasi menggunakan 2 metode yang berbeda, maka akan menghasilkan akurasi dari masing-masing metode, hasil akurasi akan dibandingkan dan dicari akurasi terbaik dari kedua metode. Detailnya ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Hasil dan Pembahasan

Adapun field pada dataset yang digunakan pada penelitian ini seperti usia, gender detailnya ditunjukkan pada Gambar 2.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
0	63	1	3	145	233	1	0	150	0	2.3
1	37	1	2	130	250	0	1	187	0	3.5
2	41	0	1	130	204	0	0	172	0	1.4
3	56	1	1	120	236	0	1	178	0	0.8
4	57	0	0	120	354	0	1	163	1	0.6
...
298	57	0	0	140	241	0	1	123	1	0.2
299	45	1	3	110	264	0	1	132	0	1.2
300	68	1	0	144	193	1	1	141	0	3.4
301	57	1	0	130	131	0	1	115	1	1.2
302	57	0	1	130	236	0	0	174	0	0.0

Gambar 3. Dataset Penelitian

Naïve Bayes:

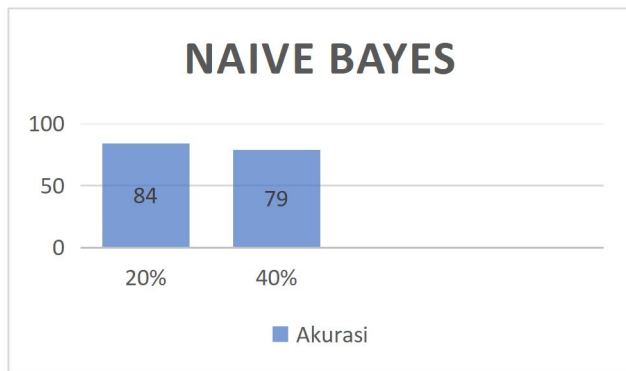
Penerapan metode naïve bayes dalam klasifikasi in dibagi menjadi 2 skenario, yaitu membagi data training menjadi 2, pada skenario pertama data training sebesar 20% dan pada skenario kedua data training sebesar 40%.

Hasil kedua skenario tersebut maka didapatkan hasil pada Tabel 1.

Tabel 1. Hasil Detail Naïve Bayes

N o	Prosentase	Accuratio n	Precisio n	Recal l
1	20	84%	0,89	0,77
2	40	79%	0,80	0,80

Adapun hasil perbandingan dalam akurasi divisualisasikan pada Gambar 3.



Gambar 4. Akurasi Naive Bayes

Support Vector Machine:

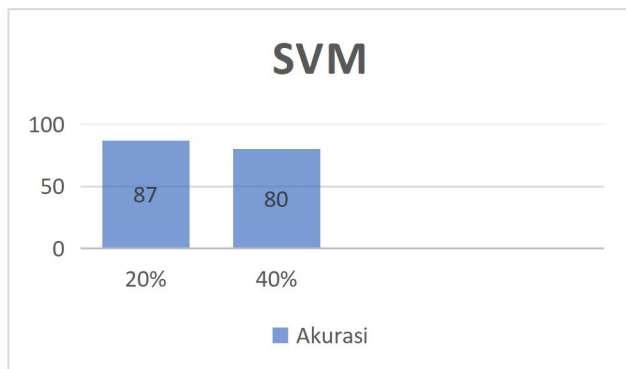
Penerapan metode Support Vector Machine dalam klasifikasi juga menggunakan scenario yang sama yaitu membagi data training menjadi 2, pada skenario pertama data training sebesar 20% dan pada skenario kedua data training sebesar 40%.

Hasil kedua skenario tersebut maka didapatkan hasil pada Tabel 2.

Tabel 2. Hasil Detail SVM

N o	Prosentase	Accuratio n	Precisio n	Recal l
1	20	87%	0,92	0,87
2	40	80%	0,82	0,82

Adapun hasil perbandingan dalam akurasi divisualisasikan pada Gambar 4.



Gambar 5. Akurasi SVM

Interpretasi Hasil:

Berdasarkan skenario yang sudah dilakukan pada percobaan atau testing di kedua metode, maka didapatkan akurasi terbaik pada metode Support Vector Machine dengan akurasi sebesar 87% dengan skenario 1 membagi data training sebesar 20%. Sedangkan metode Naive Bayes

mendapatkan nilai terbaiknya pada skenario 1 84%. Jika dilakukan perbandingan setiap skenario, maka metode Support Vector Machine tetap lebih unggul dibandingkan Naive Bayes.

Kesimpulan dan Saran

Berdasarkan penelitian yang sudah dilakukan maka dapat ditarik kesimpulan sebagai berikut :

1. Metode terbaik berdasarkan akurasi tertinggi adalah metode Support Vector Machine dengan akurasi skenario 1 sebesar 87%.
2. Hasil terbaik pada metode Naive Bayes pada skenario 1 dengan akurasi sebesar 84%.
3. Pada skenario 2 akurasi kedua metode cenderung mengalami penurunan dibandingkan pada skenario 1.

Adapun saran dalam penelitian ini adalah sebagai berikut :

1. Dapat dilakukan perbandingan dengan metode lainnya.
2. Perluasan dataset juga dapat mempengaruhi hasil, jika dimungkinkan dapat dilakukan penggalan dataset dengan kuantitas lebih banyak.
3. Bisa dilakukan dengan teknik atau skenario pengujian lain seperti cross validation.

Referensi

[1] W. H. Organization., "Cardiovascular Diseases.," World Health Organization, 2020, p. 2020.

[2] V. Kale, Enterprise Performance Intelligence and Decision Patterns. CRC Press, 2018.

[3] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," J. Inform. J. Pengemb. IT, vol. 3, no. 3, pp. 394-401, 2018, doi: 10.30591/jpit.v3i3.977.

[4] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes

- Dan Support Vector Machine (Svm),” *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [5] A. Supriyatna and W. P. Mustika, “Komparasi Algoritma Naïve bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 152, 2018, doi: 10.30645/j-sakti.v2i2.78.
- [6] O. Arifin and T. B. Sasongko, “Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier,” in *Seminar Nasional Teknologi Informasi dan Multimedia 2018*, 2018, vol. 6, no. 1, pp. 67–72.
- [7] H. T. Sueno, “Multi-class Document Classification using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3937–3944, 2020, doi: 10.30534/ijatcse/2020/216932020.
- [8] J. Gu and S. Lu, “An effective intrusion detection approach using SVM with naïve Bayes feature embedding,” *Comput. Secur.*, vol. 103, p. 102158, 2021, doi: 10.1016/j.cose.2020.102158.
- [9] Y. Narayan, “Comparative analysis of SVM and Naïve Bayes classifier for the SEMG signal classification,” *Mater. Today Proc.*, vol. 37, no. Part 2, pp. 3241–3245, 2020, doi: 10.1016/j.matpr.2020.09.093.
- [10] F. Fitriana, E. Utami, and H. Al Fatta, “Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes,” *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [11] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, “Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter,” *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citism 2018, pp. 1–6, 2019, doi: 10.1109/CITSM.2018.8674352.
- [12] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, “A Comparative Study of Support Vector Machine and Naïve Bayes Classifier for Sentiment Analysis on Amazon Product Reviews,” *2020 Int. Conf. Contemp. Comput. Appl. IC3A 2020*, pp. 217–220, 2020, doi: 10.1109/IC3A48958.2020.233300.
- [13] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [14] M. K. D. I. G. W. N. M. S. 2017 Dicky Nofriansyah, S.Kom., *ALGORITMA DATA MINING DAN PENGUJIAN*. DEEPUBLISH, 2017.
- [15] M. Kantardzic, *Advances in Data Mining*. 2019.