

PERBANDINGAN ALGORITMA SVM DAN KNN DALAM MENGLASIFIKASI KELULUSAN MAHASISWA PADA SUATU MATA KULIAH

Shedriko

Teknik Informatika, Universitas Indraprasta PGRI
shedriko@gmail.com

Submitted March 1, 2021; Revised November 22, 2021; Accepted November 26, 2021

Abstrak

Dua algoritma berbeda menggunakan pendekatan yang berbeda pula. Demikian halnya dengan KNN (*K-Nearest Neighbour*) dan SVM (*Support Vector Machine*), dimana yang satu melakukan perhitungan berdasarkan kedekatan jarak antar *instance* sementara yang lainnya melakukan perhitungan berdasarkan eksistensi *hyperplane* yang menjadi pemisah antar dua kelas yang menjadi hasilnya. Dengan menggunakan data *train* yang memiliki beberapa data anomali atau *noise*, berdasarkan penilaian nyata di lapangan yang umumnya diberikan dosen kepada mahasiswa terhadap hasil lulus atau tidak lulus dari suatu mata kuliah, penelitian ini mendapatkan pembeda antara dua algoritma tersebut. Penelitian dilakukan di Universitas XYZ terhadap mata kuliah “Pengantar Teknologi Informasi”. Dengan bantuan *software Orange Data Mining*, penelitian ini bertujuan untuk memberikan informasi mengenai algoritma yang sesuai bagi prediksi atau klasifikasi kelulusan mahasiswa dari mata kuliah tersebut maupun mata kuliah lainnya. Analisis yang digunakan adalah kuantitatif dengan metodologi KNN dan SVM tersebut. KNN lebih unggul dari SVM berdasarkan penilaian dari beberapa parameter, sedangkan SVM lebih baik dari KNN dalam mendapatkan nilai ambang kelulusan.

Kata Kunci : SVM, KNN, kelulusan, klasifikasi, kuantitatif.

Abstract

Two different algorithms will use different approaches. Like KNN (K-Nearest Neighbour) and SVM (Support Vector Machine), with the former calculates the closest distance between two instances, while the latter calculates the existence of hyperplane that separates two resulted classes. By using train data consisting of anomaly or noise data and using real measurement normally given by a lecturer to the students in the class to determine the students' pass or fail in one subject, the research finds the difference between the two algorithms. The research is conducted in the University of XYZ on IIT (Introduction to Information Technology) subject. With the use of Orange Data Mining software, the research aims to give information about the suited algorithm for prediction or classification of student's success in a related subject or others. It uses quantitative analysis with KNN dan SVM algorithms methods. Based on an assessment of several parameters, KNN is better than SVM, but SVM is better than KNN in obtaining a passing threshold value.

Keywords : SVM, KNN, success, classification, quantitative

1. PENDAHULUAN

Universitas yang berkualitas tidak mesti mahal, namun kualitas pendidik yang secara konstan dan kontinu ditingkatkan, seperti meningkatkan kedisiplinan mengajar, melakukan riset yang berkelanjutan, memberikan pelayanan mendidik yang baik serta pemberian

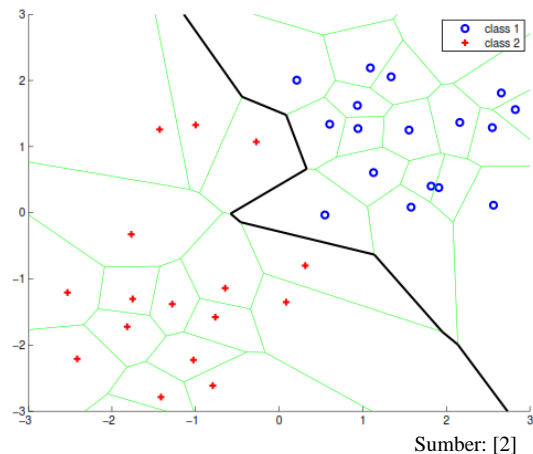
program beasiswa yang terorganisir, dapat menjadi faktor penting bagi penentuan kualitas suatu universitas. Penelitian dilakukan pada Jurusan Teknik Informatika Fakultas Teknologi dan Ilmu Komputer dari Universitas XYZ untuk mata kuliah Pengantar Teknologi Informasi. Jurusan dimana suatu mata

kuliahnya diampu oleh lebih dari 11 orang dosen ini membutuhkan kondisi komunikasi yang sangat baik antar sesama dosen pengampu, untuk menjamin keseragaman kualitas pendidikan.

Dengan bantuan *Software Orange Data Mining*, penelitian ini bertujuan untuk memberikan informasi mengenai algoritma yang sesuai bagi prediksi atau klasifikasi kelulusan mahasiswa dari mata kuliah tersebut maupun mata kuliah lainnya. Dengan menggunakan data *train* yang memiliki beberapa data anomali atau *noise*, berdasarkan penilaian nyata di lapangan yang umumnya diberikan dosen kepada mahasiswa terhadap hasil lulus atau tidak lulus dari suatu mata kuliah, diharapkan penelitian ini mendapatkan pembeda antara dua algoritma, yaitu KNN (*K-Nearest Neighbour*) dan SVM (*Support Vector Machine*).

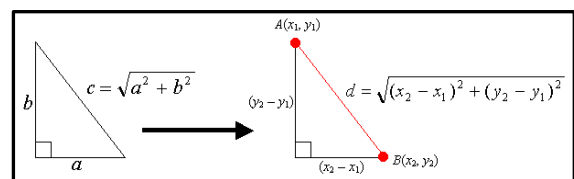
Sedangkan manfaat dari penelitian ini adalah memberikan gambaran pola kelulusan yang berkaitan dengan penilaian mahasiswa terhadap mata kuliah tersebut maupun mata kuliah lainnya. Disamping itu juga memberikan masukan dalam mengantisipasi mahasiswa yang dalam mengikuti perkuliahan menunjukkan gejala yang tidak sesuai dengan target kelulusan.

Algoritma *K-Nearest Neighbour* merupakan metode klasifikasi yang sudah diklasifikasikan sebelumnya, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada [1].



Gambar 1. Batas Keputusan dari Dua Kelas di KNN

Jarak dihitung dengan menggunakan rumus *euclidean distance*, yaitu rumus phytagoras dengan lebih dari dua dimensi [3], [4].



Sumber: [4]

Gambar 2. Rumus Phytagoras

Kemudian diperoleh rumus *euclidean distance* sebagai berikut:

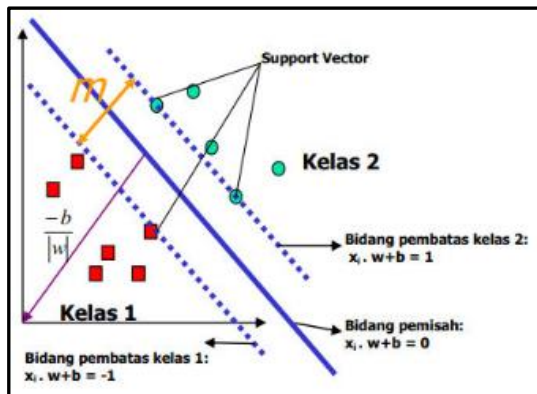
$$d = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

dimana:

q_i dan p_i : merupakan jarak antar dua titik ke i

Sedangkan algoritma SVM (*Support Vector Machine*) merupakan metode dalam *Supervised Learning* yang digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas [5].



Sumber: [6]

Gambar 3. Ilustrasi SVM

Jarak antara dua *support lines* dituliskan dalam persamaan berikut:

$$\max_{w,b} \frac{2}{\|w\|} \quad (2)$$

$$(w \cdot x) + b \geq 1, \quad \forall_i : y_i = 1 \quad (3)$$

$$(w \cdot x) + b \leq -1, \quad \forall_i : y_i = -1 \quad (4)$$

dimana:

$\max_{w,b}$: jarak maximum antar dua garis

w, x, b dan y : komponen dari persamaan garis lurus

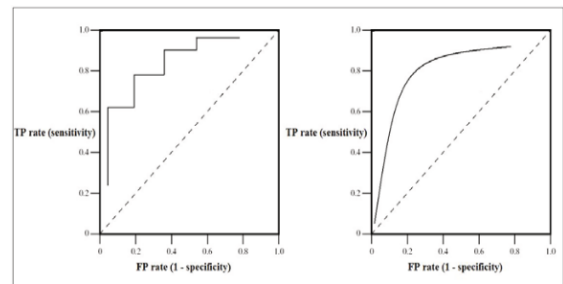
Klasifikasi biner adalah model statistik yang membagi kumpulan data menjadi dua kelompok, yaitu positif dan negatif, seperti yang dapat dilihat pada tabel *confusion matrix*.

Tabel 1. *Confusion Matrix*

Aktual	Prediksi	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Sumber: [7]

Performance yang masuk dalam tahapan evaluasi dapat dilakukan dengan menghitung akurasi, *error* dan ROC (*Receiver Operating Characteristic*) dengan melihat AUC (*Area Under Curve*) nya [8] dimana nilai AUC lebih besar maka berarti nilainya akan lebih baik.



Sumber: [7]

Gambar 4. Kurva ROC untuk Kasus Diskrit dan Kontinu

Sedangkan rumusan akurasi dan *error* adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$\text{error} = \frac{FP+FN}{TP+TN+FP+FN} \quad (6)$$

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi KNN (*K-Nearest Neighbour*) dan SVM (*Support Vector Machine*) sebagai algoritma penyelesaian masalah berdasarkan jarak dan penentuan *hyperplane*. Sedangkan metode yang digunakan adalah analisis kuantitatif. Analisis kuantitatif merupakan penelitian yang didasarkan kepada angka-angka atau penghitungan kuantitas terhadap semua fenomena yang berkaitan dengan numerik [9]. Adapun jumlah populasi yang diambil adalah sebanyak 175 mahasiswa yang mengikuti kuliah Pengantar Teknologi Informasi. Parameter input yang diambil adalah tugas (*HomeWork*), ujian tengah semester (*MidTest*) dan ujian akhir semester (*FinalTest*). Informasi dari nilai yang diperoleh dikategorikan menjadi:

- 0 - <56, menyatakan tidak lulus
- >=56, menyatakan lulus

Langkah-langkah penggunaan metode KNN (*K-Nearest Neighbour*) adalah sebagai berikut [10]:

1. Tentukan nilai K
2. Hitung jarak data baru ke setiap label data.
3. Tentukan k labeled data yang mempunyai jarak yang paling minimal
4. Klasifikasikan data baru ke dalam label data yang mayoritas KNN dipilih berdasarkan matrik jarak

Sedangkan secara umum cara kerja SVM adalah dengan menemukan jarak terjauh dari *hyperplane* terhadap dua kelas. Proses penentuannya dilakukan berulang kali hingga menemukan *hyperplane* terbaik.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 175 data mahasiswa yang diklasifikasikan ke dalam dua kelas, yaitu lulus (*Passed*) dan tidak lulus (*Failed*) di Universitas XYZ. Universitas ini menekankan kepada para dosen untuk mengajarkan tidak hanya *hard-skill* tapi juga *soft-skill* kepada para mahasiswa, sehingga keahlian yang didapat setelah lulus nantinya akan lebih lengkap. Oleh karena itu aktifitas seperti keaktifan di kelas maupun kuis menjadi faktor penilaian yang diikuti sertakan pada penelitian ini.

Beberapa kriteria penilaian yang umum dilakukan pada kelulusan mahasiswa dalam suatu mata kuliah mencakup keseluruhan nilai dari kehadiran, tugas, ujian tengah semester, ujian akhir semester serta nilai-nilai penunjang lainnya seperti keaktifan di kelas dan kuis, dengan rincian sebagai berikut:

1. Nilai akhir harus mencapai nilai lebih besar atau sama dengan 56
2. Absensi harus memenuhi setidaknya 80 persen
3. Nilai tugas minimal 60
4. Menghadiri kuis yang diadakan
5. Aktif di kelas dengan bertanya atau menjawab pertanyaan-pertanyaan yang terlontar di kelas baik yang

dikemukakan oleh dosen maupun mahasiswa pada sesi diskusi

Namun demikian, beberapa penilaian subyektif dapat saja terjadi dengan anggapan bahwa seorang mahasiswa mungkin lengah pada saat ujian, padahal dalam kegiatan belajar mengajar sehari-hari di kelas dosen melihat bahwa mahasiswa tersebut memiliki potensi yang cukup besar di perkuliahan tersebut. Berikut beberapa contoh yang mungkin terjadi dalam hal pemberian penilaian di kelas, yaitu:

1. Memenuhi nilai *total* minimum namun absensi kurang dari 80 persen, sehingga keputusan kelulusan dilihat pada nilai tugas yang tinggi, seperti lebih besar atau sama dengan 90
2. Memenuhi nilai *total* minimum namun absensi kurang dari 80 persen, sehingga keputusan kelulusan dilihat pada nilai keaktifan yang tinggi, seperti lebih dari 3 kali dalam mengajukan pertanyaan maupun menjawab, ditambah dengan selalu menghadiri kuis yang diadakan di kelas

Nilai tugas biasanya mencakup beberapa nilai berikut ini:

1. Nilai absensi
Memberikan nilai dengan porsi 50 persen dari nilai tugas bila mahasiswa menghadiri seluruh perkuliahan yang biasanya berkisar sekitar 14 kali pertemuan
2. Nilai keaktifan di kelas
Biasanya dosen akan memberikan porsi hingga 30 persen dari nilai tugas bila mahasiswa aktif hingga bertanya atau menjawab sebanyak lebih dari atau sama dengan 3 kali
3. Nilai kuis
Nilai dengan porsi hingga 20 persen dari nilai tugas akan diberikan bila mahasiswa mendapatkan nilai yang sangat memuaskan dari soal yang diberikan

Dengan mendapatkan nilai tugas dalam penelitian ini, diasumsikan bahwa ketiga penilaian tersebut di atas sudah terwakili dengan baik sehingga diharapkan dapat memberikan nilai *output* yang cenderung akurat.

Adapun atribut yang digunakan sebagai *evidence*, baik dalam Algoritma SVM maupun KNN pada penelitian ini, terdiri dari 3 *item*, yaitu.

1. Nilai tugas
2. Nilai ujian tengah semester
3. Nilai akhir semester

Dengan berbagai anomali atau *noises* yang ada, penelitian ini berusaha untuk melihat hasil perbandingan terbaik antara kedua algoritma, yaitu SVM dan KNN. Pengerjaan pengolahan data dilakukan dengan bantuan software data mining yang mengikuti karakteristik dari masing-masing algoritma. Beberapa contoh kedua klasifikasi data dapat dilihat berikut ini.

Tabel 2. Data Training yang Menyatakan Mahasiswa Lulus

No	Name	Homework	MidTest	FinalTest	Result
1	Irfan N C	90	90	70	Passed
2	Anggi P	90	95	61	Passed
3	Akram A	90	72	60	Passed
4	Anastasya T	90	85	72	Passed
8	Luthfi R	100	85	34	Passed
10	Rini H	90	85	70	Passed
12	Ananda M P	90	85	75	Passed
15	Abdul A	100	80	53	Passed
17	Ismuari S	100	76	47	Passed
18	Dimas C K	100	80	58	Passed
19	M Subhan R	80	95	66	Passed
20	Savira S	90	80	61	Passed
22	Nur Imama	90	80	72	Passed
23	Asyraf A K U	90	80	55	Passed
24	Ega Ratna D P	100	85	87	Passed
25	Gardika A S	90	80	71	Passed
26	Humayra H A	100	76	90	Passed
27	Ibrahim A	100	85	81	Passed
28	Ade Rian	100	95	40	Passed
29	Miftahul R	90	90	67	Passed

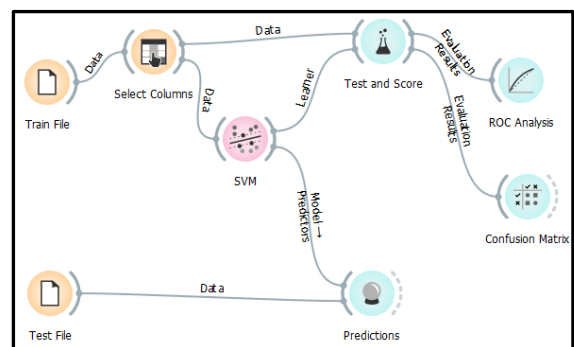
Sumber: Penulis

Tabel 3. Data Training yang Menyatakan Mahasiswa Tidak Lulus

No	Name	Homework	MidTest	FinalTest	Result
5	M Nursyamsi	90	50	28	Failed
6	Sapta S Y	90	45	17	Failed
7	M Ryansah	0	30	0	Failed
9	Apriadi S	100	30	10	Failed
11	M Afitra P	0	0	0	Failed
13	Liftiana N S	100	50	18	Failed
14	Maher F S	80	50	18	Failed
16	Bima N S	90	30	42	Failed
21	Angga E	0	0	0	Failed
35	Lukaslie	0	68	0	Failed
36	Adelia F	90	62	28	Failed
37	Ibnu H	80	62	9	Failed
39	Deri D	90	50	21	Failed
45	Bintang S R	15	45	81	Failed
46	Sugeng R	15	68	68	Failed
47	Rita B	15	72	82	Failed
49	Laberta R	15	90	35	Failed
50	Rayindra D	15	85	56	Failed
53	Haris F A L	0	0	0	Failed
57	Nadia A S	80	80	60	Failed

Sumber: Penulis

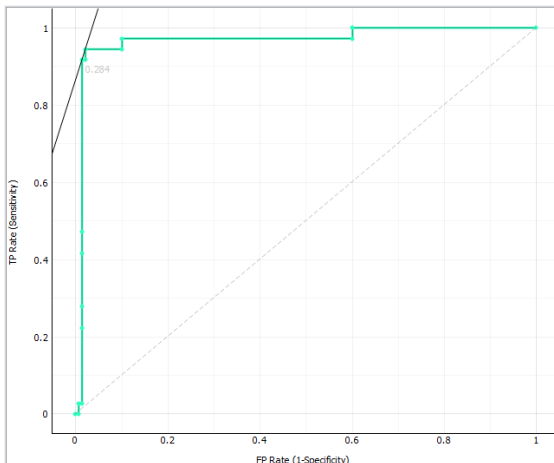
Pada penggunaan metodologi SVM, dilakukan pemetaan pada *Software Orange Data Mining* sehingga diperoleh peta grafiknya sebagai berikut.



Sumber: Penulis

Gambar 5. Peta SVM Pada Orange Data Mining

Dari peta grafik tersebut, dapat diperoleh grafik ROCnya dengan cara menampilkan pilihan *ROC analysis* pada *software* tersebut.



Sumber: Penulis

Gambar 6. Grafik ROC untuk SVM

Dari grafik tersebut diperoleh nilai AUC sebesar 0,966. Kemudian nilai akurasi dan error dapat dihitung sebagai berikut:

$$Akurasi = \frac{137 + 30}{137 + 30 + 2 + 6} = 0,954286$$

$$error = \frac{2 + 6}{137 + 30 + 2 + 6} = 0,045714$$

Dari kedua hasil perhitungan di atas, nampak bahwa akurasi penggunaan metodologi SVM tergolong sangat tinggi dengan nilai error yang sangat rendah. Kemudian dari semua nilai parameter yang diperoleh, dilakukan pengujian dengan menggunakan beberapa data test sehingga diperoleh hasil sebagai berikut:

SVM	Homework	MidTest	FinalTest
1 Passed	59	59	59
2 Failed	58	58	58

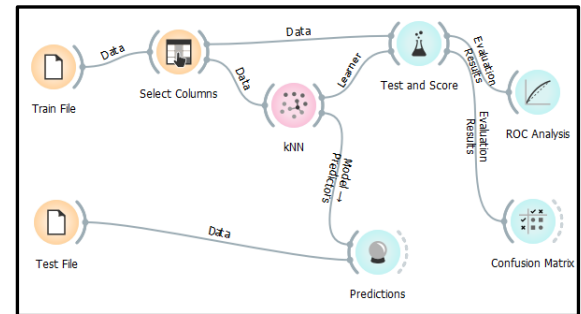
Sumber: Penulis

Gambar 7. Hasil Pengujian SVM pada Orange Data Mining

Nilai 58 dan 59 menjadi nilai batas kelulusan dari pengujian yang dilakukan. Kedua nilai tersebut sudah cukup

mendekati harapan dari nilai kelulusan standar.

Sedangkan pada penggunaan metodologi KNN, pemetaan pada *Software Orange Data Mining* dilakukan sebagai berikut:



Sumber: Penulis

Gambar 8. Peta KNN pada Orange Data Mining

Dalam metodologi KNN, ada pemilihan jumlah pengclusteran (nilai K) untuk mendapatkan nilai akurasi terbaik [11].

Tabel 4. Pembagian Cluster dan Nilai Akurasinya

Cluster	Akurasi
K1	0,96
K2	0,96
K3	0,954286
K4	0,965714
K5	0,948571

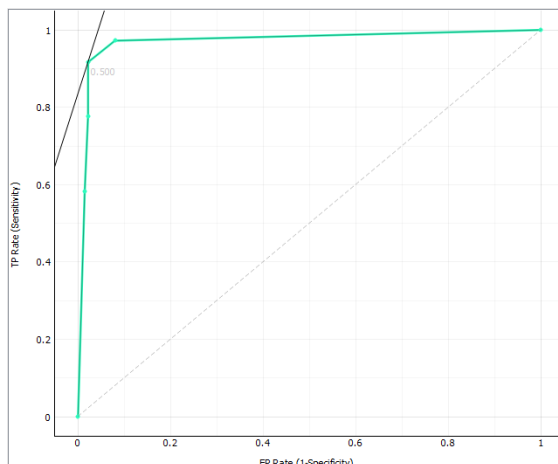
Sumber: Penulis

Dari tabel tersebut dapat dilihat bahwa membagi data menjadi 4 kelompok (K=4) memiliki nilai akurasi yang paling tinggi. Uraian dari nilai akurasi dan error dengan cluster tersebut dijabarkan berikut ini.

$$Akurasi = \frac{136 + 33}{136 + 33 + 3 + 3} = 0,965714$$

$$error = \frac{3 + 3}{136 + 33 + 3 + 3} = 0,034286$$

Nilai akurasi tersebut tergolong juga sangat tinggi dengan kategori nilai error yang sangat rendah. Sedangkan grafik ROCnya dapat dilihat pada gambar berikut ini.



Sumber: Penulis

Gambar 9. Grafik ROC untuk KNN (K=4)

Dari grafik ROC tersebut diperoleh nilai AUC adalah sebesar 0,972. Kemudian dilakukan pengujian dengan beberapa nilai untuk mendapatkan nilai ambangnya sebagai berikut:

kNN	Homework	MidTest	FinalTest
1 Passed	60	60	60
2 Failed	59	59	59

Sumber: Penulis

Gambar 10. Hasil Pengujian KNN (K=4) pada Orange Data Mining

Dari pengujian nampak bahwa nilai ambang kelulusan adalah pada nilai 59 dan 60 yang juga cukup dekat dengan nilai kelulusan yang diharapkan.

Hasil penelitian ini dapat dilihat dalam bentuk tabel berikut dimana beberapa poin perbandingan disandingkan.

Tabel 5. Perbandingan SVM dengan KNN (K=4)

No	Parameter	SVM	KNN (K=4)
1	Nilai AUC	0,966	0,972
2	Akurasi	0,954286	0,965714
3	Error	0,045714	0,034286
4	Nilai ambang kelulusan	58-59	59-60

Sumber: Penulis

4. SIMPULAN

Dari perbandingan di tabel 5, nilai AUC pada metodologi KNN lebih besar dari SVM, menunjukkan bahwa luas di bawah kurva ROC dari KNN besar yang mengartikan bahwa KNN lebih baik dari SVM. Demikian pulan dengan akurasi, dimana nilai KNN dengan jumlah *cluster* 4 lebih baik dari SVM. Untuk nilai *error*nya, KNN juga lebih baik dimana hal tersebut ditunjukkan dengan nilai error yang rendah yang berarti lebih kecil terjadi kesalahan penghitungan datanya. Sedangkan pada poin terakhir nilai dari metodologi SVM lebih baik dari KNN. Hal tersebut ditandai dengan nilai ambang lulus-tidak lulus dari SVM lebih mendekati harapan, yaitu nilai kelulusan sama dengan 56 dan di bawah nilai tersebut mahasiswa dinyatakan tidak lulus.

Dari hasil perbandingan terhadap empat poin penilaian, yaitu nilai AUC, akurasi, *error* dan nilai ambang kelulusan maka kesimpulan dapat diperoleh. Untuk tiga poin pertama dapat kita lihat bahwa untuk kasus pada penelitian ini KNN lebih unggul dari SVM. Namun untuk nilai ambang kelulusan SVM lebih baik dari KNN. Hal tersebut menunjukkan bahwa untuk kasus dalam penelitian ini, dimana terdapat beberapa data *noise*, KNN lebih baik dalam penilaian namun SVM memberikan hasil yang lebih sesuai dengan harapan.

DAFTAR PUSTAKA

- [1] M. Kubat, "An Introduction to Machine Learning," *An Introduction to Machine Learning*. Springer International Publishing., pp. 1–348, 2017.
- [2] A. Hertzmann and D. Fleet, "Machine Learning and Data Mining," 2012, p. 134.
- [3] E. Alpaydin, "Introduction to Machine Learning, Second Edition,"

- no. September. The Massachusetts Institute of Technology Press., 2010.
- [4] A. M. Ismail, "Cara Kerja Algoritma k-Nearest Neighbour." , 2018. [Online]. Available: <https://medium.com/bee-solution-partners/cara-kerja-algoritma-k-nearest-neighbor-k-nn-389297de543e>
- [5] N. Deng, Y. Tian, and C. Zang, "Support vector machines (SVM): Optimization Based Theory, Algorithms and Extension." CRC Press, Taylor & Francis Group, 2013.
- [6] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 1, no. 1. p. 19, 2017.
- [7] 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," *Jurnal Tekno Insentif*, vol. 13, no. 1. pp. 16–25, 2019.
- [8] R. Arifin, "Memahami ROC dan AUC," 2019. [Online]. Available: <https://medium.com/@arifinrio95/memahami-roc-dan-auc-2e0e4f3638bf>.
- [9] C. R. Kothari, *Research Methodology, Methods & Techniques*, Second Rev. New Age International (P) Ltd., 2004.
- [10] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *Jurnal Informatika*, vol. 6, no. 2. pp. 226–235, 2019.
- [11] S. Aulia, S. Hadiyoso, and D. N. Ramadan, "Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 3, no. 1. p. 75.