

Perbandingan Algoritma *Machine Learning* SVM dan Decision Tree untuk Prediksi Keaktifan Mahasiswa

Slamet Wiyono

Jurusan Teknik Informatika
Politeknik Harapan Bersama
Tegal, Indonesia
slamet2wiyono@gmail.com

Taufiq Abidin

Jurusan Teknik Informatika
Politeknik Harapan Bersama
Tegal, Indonesia
N3k4ther.otr@gmail.com

Abstract — Salah satu unsur yang menjadi penilaian akreditasi sebuah program studi adalah ketepatan waktu lulusan mahasiswa. Adanya mahasiswa yang tidak aktif tentu akan mempengaruhi ketepatan waktu kelulusan. Prediksi keaktifan mahasiswa diperlukan untuk mencegah siswa yang berpotensi tidak aktif. Penelitian ini bertujuan untuk membandingkan algoritma *machine learning* SVM dan *Decision Tree* dalam melakukan prediksi keaktifan mahasiswa. Proses prediksi dilakukan dengan langkah; pengumpulan data, pre-prosesing data, pembentukan model, perbandingan model, dan evaluasi. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki keakuratan lebih baik dari algoritma Decision Tree dalam memprediksi keaktifan mahasiswa. Algoritma SVM memiliki tingkat akurasi 95% dan algoritma Decision Tree memiliki tingkat akurasi 93%.

Keywords – Perbandingan Algoritma; SVM; Decision Tree

I. PENDAHULUAN

Peningkatan akreditasi sebuah program studi pada perguruan tinggi selalu diusahakan untuk menaikkan kualitas. Ketepatan waktu lulus mahasiswa merupakan salah satu elemen yang digunakan untuk penilaian akreditasi [1]. Semakin banyak mahasiswa yang lulus tepat waktu maka semakin baik nilai akreditasi. Keberadaan mahasiswa non-aktif tentunya akan berpengaruh terhadap ketepatan waktu lulus. Semakin banyak mahasiswa non-aktif maka akan semakin banyak pula mahasiswa yang lulus tidak tepat waktu. Dengan demikian, Prediksi keaktifan mahasiswa diperlukan untuk mencegah mahasiswa berpotensi tidak aktif.

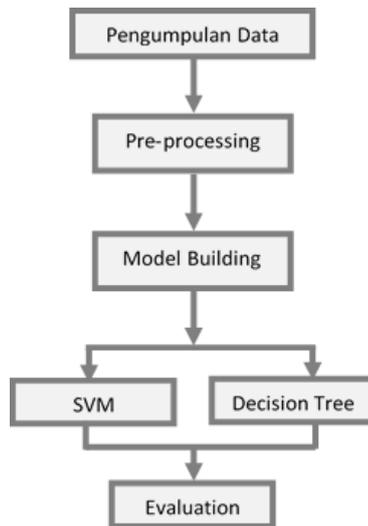
Beberapa penelitian yang berkaitan dengan prediksi keaktifan mahasiswa telah dilakukan. Beberapa penelitian menggunakan algoritma Data Mining. Algoritma Data Mining digunakan untuk pembuatan sistem penganalisa keaktifan mahasiswa (SPAS) [2], untuk memprediksi keaktifan mahasiswa (rendah, rata-rata, baik, dan sangat baik) menggunakan data pendidikan [3]. Penelitian lain dengan menerapkan algoritma Decision Tree seperti;

analisis untuk memprediksi ketepatan waktu studi mahasiswa 4 tahun [4], prediksi drop-out mahasiswa dari perguruan tinggi berdasarkan IPK [5]. Penelitian lain untuk memprediksi keaktifan mahasiswa pada saat awal mengikuti sebuah kursus [6], korelasi kegiatan luar akademik dengan keaktifan perkuliahan [7], dan analisa prediksi keaktifan mahasiswa dengan membandingkan fitur-fitur yang mempengaruhi [8]. Selain algoritma Data Mining, penelitian lain menggunakan metode Fuzzy untuk memprediksi keaktifan mahasiswa. Metode Fuzzy Support System digunakan untuk evaluasi kinerja siswa di laboratorium [9], dan aplikasi logika fuzzy untuk evaluasi keaktifan akademik siswa [10].

Penelitian-penelitian lain yang berkaitan dengan prediksi keaktifan mahasiswa juga dilakukan dengan membandingkan beberapa algoritma prediksi, diantaranya yaitu dengan membandingkan algoritma Bayesian dan Decision Tree [11], algoritma Apriori dan K-Means [12]. Namun demikian, belum ada penelitian untuk membandingkan algoritma SVM dan Decision Tree untuk memprediksi keaktifan mahasiswa.

II. METODE PENELITIAN

Penelitian ini menggunakan dua algoritma *machine learning*, yaitu algoritma SVM dan Decision Tree. Proses pembuatan model prediksi menggunakan bantuan R Studio dengan menggunakan package Caret. Proses pembuatan model dilakukan melalui beberapa tahap, yaitu; pengumpulan data, pre-prosesing data, pembentukan model, perbandingan model, dan evaluasi [13]. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur penelitian

Bahan penelitian yang digunakan yaitu menggunakan data akademik mahasiswa program studi Teknik Informatika Politeknik Harapan Bersama. Setelah data terkumpul maka dilakukan pre-prosesing data dengan memilih data-data yang relevan dan membuang data yang dianggap tidak relevan dengan tujuan penelitian. Data yang digunakan yaitu nilai IPS, IPK, kota asal, asal sekolah, jurusan saat sekolah, pekerjaan orang tua, dan keaktifan mahasiswa pada satu semester kedepan. Preprosesing data juga dilakukan dengan membagi data menjadi dua bagian, yaitu data training dan data testing dengan rasio perbandingan 75:25. Setelah dilakukan pembagian data menjadi data training dan data testing, maka dilakukan pembentukan model dengan menggunakan data training. Setelah model terbentuk maka dilakukan pengujian model menggunakan data testing. Hasil pengujian berupa tingkat keakuratan model dalam melakukan prediksi. Model yang terbentuk dari algoritma SVM dan Decision Tree kemudian dibandingkan atau dievaluasi. Hasil evaluasi berupa simpulan algoritma mana yang paling baik untuk melakukan prediksi berdasarkan tingkat keakuratan prediksi.

III. HASIL DAN PEMBAHASAN

Dataset yang digunakan sebanyak 1530 baris dengan 7 variabel. 6 variabel pertama digunakan untuk memprediksi variabel ke-7. Perbandingan data training dan testing yaitu 75:25. Model yang terbentuk dari data training menghasilkan model yang belum diuji. Adapun hasil model yang terbentuk ditampilkan pada Tabel I. Model yang terbentuk kemudian diuji untuk mengetahui tingkat akurasi model dalam melakukan prediksi. Pengujian ditampilkan pada Tabel II untuk algoritma SVM dan Tabel III untuk algoritma Decision Tree.

TABEL I
MODEL YANG TERBENTUK

Algorithm	Result	Accuracy
SVM	value C = 1	95.09%
Decision Tree	cp = 0.6689113	95.65%

Algoritma SVM menghasilkan model dengan akurasi terbaik jika nilai C = 1 dengan akurasi 95,09% dan algoritma Decision Tree menggunakan cp = 0.6689113 dengan akurasi 95,65%. Sebelum dilakukan testing model, algoritma Decision Tree memiliki akurasi yang lebih baik dari algoritma SVM.

TABEL II
CONFUSION MATRIX SVM

Prediction	Reference	
	active	non-active
active	311	13
non-active	5	53

TABEL III
CONFUSION MATRIX DECISION TREE

Prediction	Reference	
	active	non-active
active	308	18
non-active	4	48

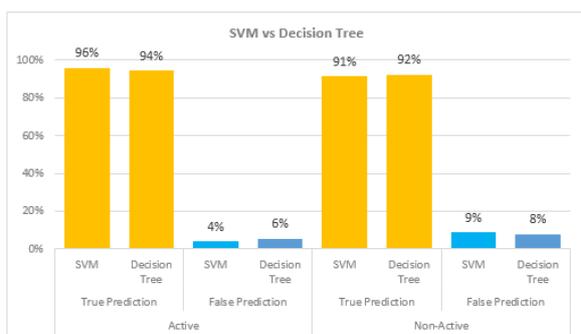
Tabel II dan Tabel III menunjukkan bahwa algoritma SVM dapat melakukan prediksi mahasiswa aktif dengan tepat sebanyak 311 sedangkan algoritma Decision Tree sebanyak 308. Adapun dalam memprediksi mahasiswa non-aktif, algoritma SVM dapat melakukan prediksi dengan tepat sebanyak 53 sedangkan algoritma Decision Tree sebanyak 48. Dengan demikian bisa diketahui bahwa algoritma SVM lebih baik hasil prediksinya dibandingkan algoritma Decision Tree. Namun demikian apabila melihat hasil akurasi sebelum dilakukan testing (Tabel I), algoritma Decision Tree memiliki akurasi lebih baik dari algoritma SVM.

Hasil pengujian (Tabel II dan Tabel III) belum terlihat tingkat keberhasilannya karena belum dibandingkan dengan jumlah kegagalan prediksi. Oleh karena itu perlu dilakukan persentase keberhasilan prediksi dan kegagalan prediksi. Tabel

IV menampilkan perbandingan prosentase keberhasilan dan kegagalan prediksi kedua algoritma.

TABLE IV
PROSENTASE KEBERHASILAN DAN KEGAGALAN PREDIKSI

Prediction		SVM	Decision Tree
Active	TRUE	96%	94%
	FALSE	4%	6%
Non-Active	TRUE	91%	92%
	FALSE	9%	8%



Gambar 2. Grafik perbandingan prediksi SVM dan Decision Tree

Tabel IV menunjukkan bahwa algoritma SVM dapat memprediksi mahasiswa aktif dengan tepat sebesar 96% sedangkan algoritma Decision Tree sebesar 94%. Namun dalam memprediksi mahasiswa non-aktif, algoritma Decision dapat memprediksi dengan tepat sebesar 92% sedangkan algoritma SVM 91%. Gambar 2 menunjukkan bahwa algoritma SVM lebih baik dalam memprediksi mahasiswa aktif namun tidak lebih baik dari algoritma Decision Tree dalam memprediksi mahasiswa non-aktif.

Hasil yang diperoleh dari Tabel IV masih menggabungkan prosentase keberhasilan dan kegagalan dalam memprediksi, belum memperlihatkan akurasi prediksi secara menyeluruh. Tabel V memperlihatkan prosentase akurasi prediksi antara algoritma SVM dan Decision Tree secara menyeluruh.

TABEL V
AKURASI PREDIKSI KESELURUHAN

SVM	Decision Tree
95%	93%

Tabel V menampilkan hasil akurasi keseluruhan antara algoritma SVM dan Decision Tree. Algoritma SVM secara umum memiliki tingkat akurasi prediksi yang lebih baik dibandingkan dengan algoritma Decision Tree.

IV. KESIMPULAN

Algoritma SVM dapat memprediksi keaktifan mahasiswa dengan baik menggunakan model $C = 1$ dan algoritma Decision Tree menggunakan $cp = 0.6689113$. Secara umum, algoritma SVM lebih akurat dalam melakukan prediksi keaktifan mahasiswa dibandingkan dengan algoritma Decision Tree.

UCAPAN TERIMA KASIH

Terima kasih kepada Kementerian Riset, Teknologi, dan Pendidikan Tinggi (RISTEKDIKTI) Republik Indonesia yang telah membiayai penelitian ini.

REFERENCES

- [1] BAN-PT, "Buku I Naskah Akademik Akreditasi Institusi Perguruan Tinggi." 2011.
- [2] C. L. Sa, D. H. b. A. Ibrahim, E. D. Hossain, and M. bin Hossin, "Student performance analysis system (SPAS)," in *The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, 2014, pp. 1–6.
- [3] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using Educational Data Mining," in *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 2016, pp. 91–95.
- [4] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, 2017.
- [5] M. N. Quadri and N. V. Kalyankar, "Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques," *Glob. J. Comput. Sci. Technol.*, vol. 10, no. 2, pp. 2–5, 2010.
- [6] A. Vihavainen, "Predicting Students' Performance in an Introductory Programming Course Using Data from Students' Own Programming Process," in *2013 IEEE 13th International Conference on Advanced Learning Technologies*, 2013, pp. 498–499.
- [7] R. Conijn, A. Van den Beemt, and P. Cuijpers, "Predicting student performance in a blended MOOC," *J. Comput. Assist. Learn.*, no. March, pp. 1–14, 2018.
- [8] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting Student Performance using Advanced Learning Analytics," *Proc. 26th Int. Conf. World Wide Web Companion*, pp. 415–421, 2017.
- [9] Z. Yıldız and A. F. Baba, "Evaluation of student performance in laboratory applications using fuzzy decision support system model," in *2014 IEEE Global Engineering Education Conference (EDUCON)*, 2014, pp. 1023–1027.
- [10] P. Nagar, "Application of Fuzzy Logic for Evaluation of

- [11] Academic Performance of Students of Computer.” *Iran. Conf. Fuzzy Syst.*, vol. 3, no. X, pp. 260–267, 2013.
- [12] A. U. Khasanah and Harwati, “A Comparative Study to Predict Student’s Performance Using Educational Data Mining Techniques,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 215, p. 012036, Jun. 2017.
- [13] G. S. Gowri, R. Thulasiram, and M. A. Baburao, “Educational Data Mining Application for Estimating Students Performance in Weka Environment,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 263, no. 3, 2017.
- [13] B. Lantz, *Machine Learning with R*, 2nd ed. Birmingham-Mumbai: Packt Publishing Ltd., 2015.