

Perbandingan Metode Klasifikasi Naïve Bayes Dan Support Vector Machine Pada Predikat Kelulusan Mahasiswa

Lingga Kurnia Ramadhani^{1*}, Bajeng Nurul Widyaningrum²

¹*Bisnis Digital, Universitas Ivet, Indonesia*

²*Rekam Medis & Informasi Kesehatan, Politeknik Bina Trada, Indonesia*

*Corresponding Author: linggakurnia@ivet.ac.id

Abstract

College is a place for students to obtain education and universities are expected to provide quality education. Every college must expect students who pass will be in accordance with the number of incoming students. In this study, it is intended to classify using data from students who graduated with graduation predicate. Researchers tried to compare 2 algorithms that are quite popular in the classification process, namely Naïve Bayes and Support Vector Machine. In previous studies, Naïve Bayes tends to produce better performance in terms of accuracy compared to Support Vector Machine. After testing using student graduation data, it turns out that Naïve Bayes has a greater accuracy of 96.52% with an error rate (Missclassification Rate) of 0.03%, while the Support Vector Machine method has an accuracy of 86.93% with an error rate (Missclassification Rate) of 0.13%.

Keywords: classification, *Naïve Bayes*, *Support Vector Machine*, students

Abstrak

Perguruan tinggi adalah tempat bagi mahasiswa dalam memperoleh pendidikan dan perguruan tinggi diharapkan dapat menyelenggarakan pendidikan yang berkualitas. Setiap perguruan tinggi pasti mengharap mahasiswa yang lulus akan sesuai jumlahnya dengan mahasiswa yang masuk. Dalam penelitian ini bermaksud untuk melakukan klasifikasi dengan menggunakan data dari mahasiswa yang lulus dengan predikat kelulusan. Peneliti mencoba membandingkan 2 algoritma yang cukup populer dalam proses klasifikasi yaitu *Naïve Bayes* dan *Support Vector Machine*. Pada penelitian - penelitian sebelumnya *Naïve Bayes* cenderung menghasilkan performa yang lebih baik dari akurasi dibandingkan dengan *Support Vector Machine*. Setelah dilakukan pengujian dengan menggunakan data kelulusan mahasiswa ternyata *Naïve Bayes* memiliki akurasi lebih besar yaitu 96.52% dengan tingkat eror (*Missclassification Rate*) 0.03%, sedangkan metode *Support Vector Machine* memiliki akurasi 86.93% dengan tingkat eror (*Missclassification Rate*) 0.13%.

Kata Kunci: klasifikasi, *Naïve Bayes*, *Support Vector Machine*, mahasiswa

I. INTRODUCTION

Jenjang pendidikan setelah SMA dan tempat pendidikan untuk mahasiswa adalah perguruan tinggi. Kualitas pendidikan bagi mahasiswa dalam sebuah perguruan tinggi diharapkan dapat menghasilkan sumber daya manusia yang memiliki ilmu, lebih kreatif, dan mampu bersaing. Di Indonesia dalam mengukur kualitas perguruan tinggi dilihat dari akreditasi yang di tentukan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Terdapat 9 kriteria dalam BAN-PT yang digunakan untuk mengukur kualitas perguruan tinggi dengan menggunakan instrumen 4.0 dan mahasiswa masuk dalam kriteria ke tiga. Kriteria mahasiswa berisi semua yang diperoleh seperti layanan dari perguruan tinggi termasuk dengan program studi yang ada. Karena aset terpenting dari perguruan tinggi adalah mahasiswa, maka penting untuk lebih diperhatikan mengenai tingkat kelulusannya. Karena masa studi mahasiswa

mempengaruhi nilai dari sebuah institusi pendidikan dan kualitas dari sebuah perguruan tinggi dapat dilihat dari keberhasilan mahasiswa.

Universitas Ivet merupakan salah satu perguruan tinggi swasta di Indonesia yang memiliki data mahasiswa dan data kelulusan mahasiswa. Data tersebut menghasilkan informasi seperti jumlah kelulusan setiap tahun, profil, dan hasil akademik mahasiswa. Adanya informasi mahasiswa lulus dengan predikat sangat memuaskan dan dengan pujian akan digunakan dalam proses klasifikasi.

Penelitian mengenai klasifikasi telah banyak dilakukan dengan menggunakan algoritma klasifikasi. Algoritma tersebut antara lain algoritma *Naïve Bayes* yang merupakan salah satu algoritma terpopuler yang digunakan untuk klasifikasi dan menghasilkan akurasi yang akurat. Selain itu algoritma *Support Vector Machine* juga cukup populer dalam proses klasifikasi.

Berdasarkan paparan tersebut, peneliti akan bermaksud membandingkan kinerja dari algoritma *Naïve Bayes* dan *Support Vector Machine* dalam melakukan klasifikasi pada predikat kelulusan mahasiswa.

II. LITERATURE REVIEW

Penelitian yang dilakukan oleh Sri Wahyuni tahun 2021 dengan judul “Penenarapan Algoritma *Naïve Bayes* pada Prediksi Kelulusan Mahasiswa (Studi Kasus : Institut Medika Drg. Suherman)”. Penelitian ini membahas penggunaan Algoritma *Naïve Bayes* digunakan untuk prediksi kelulusan yang diterapkan dalam bentuk aplikasi. Aplikasi tersebut digunakan untuk membantu institusi dalam menentukan langkah tepat bagi mahasiswa yang tidak lulus ke depannya.

Selanjutnya penelitian oleh Windy dkk tahun 2019 mengenai “Implementasi Algoritma *Naïve Bayes* untuk Memprediksi Lama Masa Studi dan Predikat Kelulusan Mahasiswa”. Penelitian membahas tentang aplikasi yang menggunakan *Naïve Bayes* sebagai proses prediksi masa studi beserta predikat kelulusannya. Algoritma *Naïve Bayes* dalam prediksi masa studi akurasi rata-rata yang didapatkan sebesar 77,19 % dan akurasi rata-rata prediksi predikat kelulusan mencapai 87,65 %.

Kemudian penelitian oleh Aan Rohanah dkk tahun 2021 dengan tema “Perbandingan *Naïve Bayes* dan *Support Vector Machine* untuk Klasifikasi Ulasan Pelanggan Indihome”. Pada penelitian ini membahas bagaimana melakukan klasifikasi dengan objek ulasan pelanggan Indihome dari media sosial Twitter dengan algoritma *Naïve Bayes* dan *Support Vector Machine*. Hasil klasifikasi dengan membandingkan 2 algoritma untuk pengguna IndiHome pada media sosial Twitter adalah SVM dengan nilai *accuracy* sebesar 82,11%, *Precision* senilai 76,44%, *recall* sebesar 88,01%, dan AUC 0,909.

Terakhir penelitian oleh Herwati et al tahun 2016 dengan “*Drop Out Estimation Students Based On The Study Period: Comparison Between Naïve Bayes And Support Vector Machines Algorithm Methods*”. Dalam penelitian menghasilkan bahwa *Naïve Bayes* mempunyai akurasi lebih baik untuk memprediksi dengan menggunakan data mahasiswa yang mengundurkan diri yaitu persentase sebesar 80,67% dan dengan error 19,33%.

III. RESEARCH METHOD

Penelitian yang dilakukan akan menggunakan data lulusan mahasiswa Universitas Ivet periode Oktober 2022. Atribut yang dibutuhkan predikat, jenis kelamin, IPK, dan lama kelulusan. Penelitian ini menggunakan *dataset* lulusan mahasiswa periode Oktober 2022 yang telah memenuhi syarat atribut yang dibutuhkan. Landasan teori yang digunakan sebagai pedoman dasar diperkuat dengan mengumpulkan beberapa literatur dan referensi sebagai studi pustaka.

Metode *Split Validation* merupakan metode yang akan digunakan dalam penelitian untuk menganalisis data. Pertama yang perlu dilakukan adalah data yang terkumpul diseleksi dan data dibersihkan dari yang diperlukan. Kemudian menghapus atribut yang tidak diperlukan, selanjutnya data disimpan dalam format *xlsx*. Data yang telah memenuhi kriteria sesuai kebutuhan akan digunakan sebagai data untuk dilakukan pengujian menggunakan metode *Naïve Bayes* dan *Support Vector Machine*. Kedua metode tersebut digunakan untuk melakukan prediksi predikat kelulusan dengan teknik *split validation*. Sebagai

metode evaluasi digunakan algoritma *Naïve Bayes* dan *Support Vector Machine* dengan pengukuran tingkat *accuracy* dan *missclassification rate*. Sebagai acuan dapat dilihat dari hasil evaluasi untuk menentukan algoritma mana yang direkomendasikan pada penelitian ini.

IV. RESULTS AND DISCUSSION

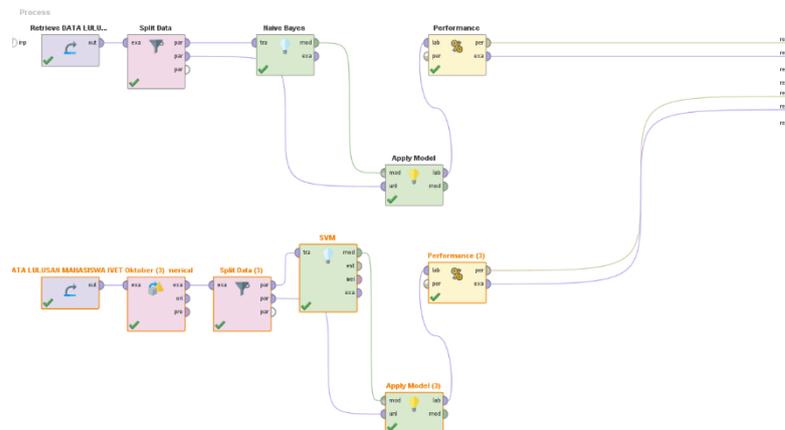
Tools *Rapid Miner Studio 9.1* digunakan peneliti dalam melakukan pengujian. Data yang telah dilakukan proses preprosesing digunakan dalam proses pengujian dengan penerapan *Naïve Bayes* dan *Support Vector Machine*. Proses pengujian menggunakan data yang telah tersimpan dalam *Microsoft Excel* dengan format *xlsx*. Total data yang digunakan sebanyak 579 dan 4 variabel dimana pembagian training dan testing sebesar 2 banding 8. Data training yang digunakan adalah 115 data dan testing sebanyak 464 data. Pertama data pada file excel di *import* ke *Rapid Miner*, selanjutnya indikator ditentukan ke masing – masing atribut dan karena sebagai penentu maka atribut diubah menjadi label.

Data mahasiswa yang digunakan adalah data yang memenuhi 4 atribut yaitu predikat, jenis kelamin, IPK, dan lama kelulusan seperti pada tabel 1 di bawah ini :

Tabel 1. Data Mahasiswa

No	Predikat	Jenis Kelamin	IPK	Lama Kelulusan
1	Sangat Memuaskan	L	3.610	Tidak Tepat Waktu
2	Sangat Memuaskan	P	3.450	Tepat Waktu
3	Sangat Memuaskan	L	3.300	Tepat Waktu
4	Dengan Pujian	P	3.750	Tepat Waktu
5	Sangat Memuaskan	L	3.280	Tepat Waktu
6	Dengan Pujian	P	3.550	Tepat Waktu
7	Sangat Memuaskan	L	3.440	Tepat Waktu
Total			579	

Pada Gambar 1 dapat di lihat proses perbandingan klasifikasi antara algoritma *Naïve Bayes* dan *Support Vector Machine* menggunakan tools *Rapid Miner Studio 9.1* dengan data mahasiswa.



Gambar 1 Proses Klasifikasi

Prediksi dengan *Naïve Bayes*

Berikut merupakan hasil pengujian yang dilakukan dengan metode *Naive Bayes* dengan *split validation* dapat dilihat pada Gambar 3 :

Tabel 2 Pengujian *Naive Bayes*

	True Sangat Memuaskan	True Dengan Pujian	
Pred. Sangat Memuaskan	83	4	95.40%
Pred. Dengan Pujian	0	28	100%
Class Recall	100%	87.50%	

Berdasarkan tabel 2 dapat dilihat bahwa pengujian menggunakan *Naive Bayes* menghasilkan tingkat akurasi sebesar 96.52% dan tingkat error (*Missclassification Rate*) 0.03%.

Prediksi dengan *Support Vector Machine*

Dari pengujian dengan menggunakan metode *Support Vector Machine* dengan *split validation* menghasilkan seperti pada tabel 3 berikut :

Tabel 3 Pengujian *Support Vector Machine*

	True Sangat Memuaskan	True Dengan Pujian	
Pred. Sangat Memuaskan	68	0	100%
Pred. Dengan Pujian	15	32	68.09%
Class Recall	81.93%	100%	

Pada tabel 3 merupakan hasil pengujian dengan menggunakan *Support Vector Machine* menghasilkan tingkat akurasi 86.96% dan tingkat error (*Missclassification Rate*) 0.13%.

V. Conclusion

Berdasarkan hasil eksperimen yang dilakukan, model klasifikasi yang dihasilkan dari metode *Naive Bayes* memiliki performa yang lebih tinggi. Metode *Naive Bayes* memiliki akurasi lebih besar yaitu 96.52% dengan tingkat error (*Missclassification Rate*) 0.03%, sedangkan metode *Support Vector Machine* memiliki akurasi 86.93% dengan tingkat error (*Missclassification Rate*) 0.13%.

REFERENCES

- [1] Z. Lin and Z. L. Wang, "Personal Home Page :," *Mater. Res.*, vol. 6, no. 2, pp. 1–60, 2009.
- [2] A. Rohanah, D. L. Rianti, and B. N. Sari, "Perbandingan *Naive Bayes* dan *Support Vector Machine* untuk Klasifikasi Ulasan Pelangan Indihome," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 23, 2021, doi: 10.30998/string.v6i1.9232.
- [3] M. R. Maarif, "Perbandingan *Naive Bayes Classifier* dan *Support Vector Machine* untuk Klasifikasi Judul Artikel," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 2, pp. 90–93, 2016, doi: 10.14421/jiska.2016.12-05.
- [4] S. Megira, Kusri, and E. T. Luthfi, "PERBANDINGAN KINERJA *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* UNTUK PREDIKSI HERREGISTRASI | Jurnal Sistem Informasi Komputer dan Teknologi Informasi (SISKOMTI)," *J. Siskomti*, vol. 3, no. 2, pp. 1–17, 2020, [Online]. Available: <https://www.ejournal.lembahdempo.ac.id/index.php/STMIK-SISKOMTI/article/view/119>.
- [5] Y. M. Firdaus, "Penerapan Metode *Naive Bayes Classifier* Untuk Mengklasifikasi Tingkat Prestasi Akademik Santri Pondok

- Pesantren Mahasiswa (Ppm) Baitul Jannah Malang,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 3, no. 1, pp. 327–336, 2019, [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/download/1398/1252>.
- [6] F. E. Purwiantono, “Implementasi Algoritma Naïve Bayes Untuk Memprediksi Lama Masa,” vol. 10, no. April, pp. 18–28, 2019.
- [7] Harwati, R. I. Virdyanawaty, and A. Mansur, “Drop out Estimation Students based on the Study Period: Comparison between Naïve Bayes and Support Vector Machines Algorithm Methods,” *IOP Conf. Ser. Mater. Sci. Eng.,* vol. 105, no. 1, 2016, doi: 10.1088/1757-899X/105/1/012039.