

Prediksi Kelulusan Mahasiswa Program Studi Teknologi Informasi Universitas Darma Persada Dengan Menggunakan Algoritma Support Vector Machine dan Decision Tree

Laily Badria¹, Linda Nur Afifa^{2*}

¹ Program Studi Teknologi Informasi Fakultas Teknik, Universitas Darma Persada

² Dosen Program Studi Teknologi Informasi Fakultas Teknik, Universitas Darma Persada

Jl. Taman Malaka Selatan No.22, Pondok Kelapa, Duren Sawit, DKI Jakarta, Indonesia 13450

*Koresponden : afycena@gmail.com

Abstrak

Penelitian ini mengembangkan sistem prediksi kelulusan yang digunakan untuk mengetahui hasil prediksi dengan mudah dan tepat dengan menerapkan metode klasifikasi. Lokasi penelitian ini adalah Program Studi Teknologi Informasi Universitas Darma Persada dengan menerapkan algoritma Support Vector Machine. Dalam melakukan prediksi terdiri dari beberapa kriteria yang sudah ditetapkan seperti persentase kehadiran, riwayat tagihan mahasiswa semester 1 hingga semester 4 serta Indeks Prestasi Semester (IPS) 1 hingga 4. Dengan menerapkan metodologi yang mencakup proses mengumpulkan data, pemodelan, evaluasi serta mengimplementasi model prediksi. Penelitian ini mendapatkan hasil bahwa algoritma SVM memberikan kinerja yang baik dalam kelulusan mahasiswa dengan tingkat akurasi yang tinggi yaitu 97%. Dengan menerapkan model prediksi ini, dapat digunakan pihak universitas untuk peningkatan keberhasilan akademik mahasiswa untuk mengurangi tingkat Drop Out mahasiswa.

Kata Kunci: Prediksi Kelulusan, Support Vector Machine, Decision Tree, Data Mining, Confusion Matrix

Abstract

This research develops for a graduation prediction system that is used to find out the prediction results easily and precisely by applying the classification method. The location of this research is the Darma Persada University, Majoring Information Technology by applying the Support Vector Machine algorithm. In making prediction, it consists of several predetermined criteria such as attendance percentage, student administration history from semester 1 to 4, Semester Achievement Index (IPS) 1 to 4. By applying a methodology that includes the process of collecting data, modeling, evaluating and implementing prediction models. This study found that the SVM algorithm provides good performance in student graduation with a high accuracy rate of 97%. By implementing this prediction model, it can be used by universities to improve students' academic success to reduce student dropout rates.

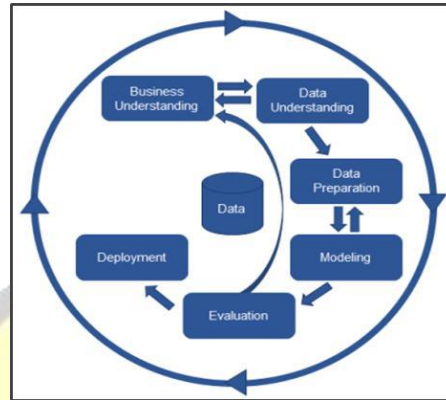
Keywords: Graduation Prediction, Support Vector Machine, Decision Tree, Data Mining, Confusion Matrix

1. Pendahuluan

Suatu perguruan tinggi bertanggung jawab atas kualitas mutu pendidikan yang diselenggarakan, agar menghasilkan mahasiswa yang memiliki kemampuan, kreativitas dan memiliki daya saing yang baik. Mutu pendidikan tersebut dapat dilihat akreditasinya melalui Badan Akreditasi Nasional Perguruan Tinggi atau BAN-PT [1]. Kelulusan mahasiswa yang tepat waktu merupakan aspek penting bagi keberlangsungan perguruan tinggi. Ketidakberhasilan akademik tepat waktu dapat mengakibatkan beban perguruan tinggi akibat jumlah data yang harus dikelola [2]. Kriteria yang digunakan dalam penelitian ini mengacu pada Peraturan Rektor Tentang Penyelenggaraan Pendidikan Universitas Darma Persada, dimana dalam peraturan tersebut terdapat ketentuan mahasiswa yang gagal studi atau Drop Out, seperti tidak dapat menyelesaikan 20 satuan kredit semester (SKS) dalam 2 semester pada tahun pertama untuk program Diploma dan Sarjana, tidak mencapai Indeks Prestasi Kumulatif (IPK) < 2,00 dalam 2 semester pada tahun pertama, tidak aktif selama 4 semester berturut-turut, dan menyelesaikan seluruh persyaratan administrasi. Maka dari itu, penelitian ini mengacu dari beberapa kriteria seperti, persentase kehadiran, riwayat tagihan mahasiswa semester 1 hingga semester 4 serta Indeks Prestasi Semester (IPS) 1 hingga 4 di Program Studi Teknologi Informasi Universitas Darma Persada. Salah satu solusi yang efektif adalah melakukan prediksi kelulusan mahasiswa secara dini dengan menggunakan teknik data mining pada aplikasi khususnya metode klasifikasi, yang bertujuan untuk memberi perlakuan khusus pada mahasiswa yang berpotensi Drop Out.

2. Metodologi

Metodologi yang diterapkan pada penelitian ini adalah CRISP-DM atau *Cross-Industry Standard Process for Data Mining*.



Gambar 2. Proses CRISP-DM [3]

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang digunakan untuk mengarahkan setiap proses data mining, dimulai dari pemahaman bisnis hingga deployment. Proses pertama adalah pemahaman bisnis, yang berfokus pada identifikasi tujuan bisnis untuk mencegah mahasiswa gagal studi di Universitas Darma Persada. Dilanjutkan dengan pemahaman data, di mana data mahasiswa dikumpulkan dari portal akademik universitas. Tahap persiapan data melibatkan pembersihan dan penyeimbangan data menggunakan teknik SMOTE. Pemodelan kemudian dilakukan dengan melatih model SVM, diikuti evaluasi model menggunakan matriks confusion untuk mengukur kinerja. Akhirnya, tahap deployment melibatkan implementasi model ke dalam lingkungan produksi untuk digunakan oleh pihak universitas.

3. Landasan Teori

1. Data Mining:

Rangkaian proses penting untuk memperoleh informasi atau pengetahuan yang dari suatu kumpulan dataset yang kompleks. Data mining memiliki tujuan utama yaitu menemukan pola, keterhubungan, atau informasi yang mungkin tidak tampak secara langsung dalam data, sehingga mendapatkan manfaat dari wawasan yang telah diberikan secara mendalam. Pada proses data mining mencakup penerapan dari berbagai teknik statistik, matematis, dan kecerdasan buatan dengan tujuan untuk menganalisis data secara sistematis dan otomatis. Hasil yang diperoleh dari data mining dapat digunakan sebagai dukungan pengambilan keputusan, mengenali tren pasar, meningkatkan efisiensi operasional serta merumuskan strategi bisnis [3].

2. CRISP-DM:

CRISP-DM atau Cross-Industry Standard Process for Data Mining, yang mencakup kerangka kerja dari data mining dengan tujuan memecahkan masalah dalam berbagai kondisi bisnis ataupun untuk keperluan penelitian [4].

3. Python:

Merupakan bahasa pemrograman yang populer di kalangan pemula hingga para ahli peneliti di dunia. Python diciptakan oleh Guido van Rossum dan diumumkan pada tahun 1991. Kini, melalui laman <https://www.python.org/> dan tersedia di beberapa platform python dapat diunduh secara gratis. Python merupakan salah satu bahasa pemrograman yang didukung oleh platform Jupyter Notebook. Jupyter ialah sebuah platform yang digunakan sebagai eksekusi kode langsung melalui peranti lunak atau web. Jupyter juga salah satu alat untuk analisis data yang banyak dipakai oleh para ilmuwan data [5].

4. Unified Modelling Language (UML):

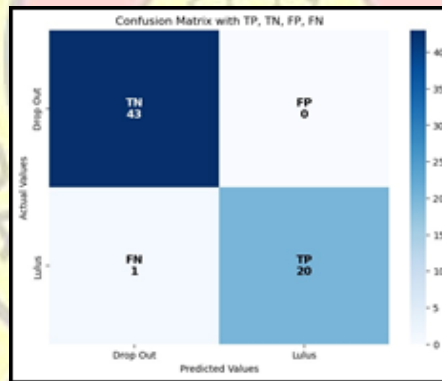
Sebuah Bahasa berbasis grafik serta gambar yang dimanfaatkan untuk memvisualisasikan serta mencatat sistem perangkat lunak yang dikembangkan berbasis Object-Oriented. UML bukan hanya bahasa pemrograman visual, namun berinteraksi dengan Bahasa pemrograman mencakup JAVA, c++, Visual Basic, serta database object-oriented [6].

5. Use Case Diagram:
 Dalam diagram ini menunjukkan hubungan antara sistem dan aktor. Jenis hubungan yang dilakukan antara pengguna sistem dengan sistem yang dapat digambarkan melalui use case diagram.
6. Activity Diagram:
 Dirancang untuk mendefinisikan alur dari beragam aktivitas dalam suatu sistem yang sedang dibangun, mulai dari tahap awal lalu melakukan proses keputusan yang akan terjadi kemudian melakukan tahap akhir dari alur tersebut. Dalam diagram ini juga menyatakan tahap paralel yang bisa saja terjadi pada langkah-langkah eksekusi. Activity Diagram merupakan bentuk state diagram tertentu yang merupakan suatu action dan transisi yang berasal dari alur sebelumnya.

4. Hasil dan Pembahasan

Mengembangkan aplikasi prediksi kelulusan mahasiswa berbasis website menggunakan streamlit dan berhasil memprediksi kelulusan mahasiswa dengan data baru menggunakan algoritma Support Vector Machine. Hasil evaluasi akurasi algoritma menggunakan Confusion Masebelum menggunakan teknik SMOTE 96% dan sesudah menggunakan teknik SMOTE menjadi 97%. Kinerja performa model SVM mampu mengatasi data yang tidak linear dan dapat menghindari overfitting serta penerapan model SVM yang dapat membantu pihak Universitas Darma Persada dalam mengidentifikasi mahasiswa yang berpotensi Drop Out dan dapat mengambil tindakan lebih lanjut. Berikut ini merupakan tampilan antarmuka dari aplikasi prediksi kelulusan mahasiswa berbasis website menggunakan streamlit. Penelitian ini menerapkan metode Support Vector Machine dan Decision Tree dalam studi kasus ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi kelulusan mahasiswa berdasarkan data kehadiran, tagihan, dan IPS. Perbandingan kedua metode ini memberikan panduan penting mengenai kemampuan masing-masing dalam memprediksi kelulusan. Studi kasus ini diharapkan dapat mengembangkan sistem yang akurat untuk memonitor dan mengantisipasi kelulusan mahasiswa, serta berkontribusi pada kemajuan teknologi informasi. Berikut ini merupakan hasil dari confusion matrix berdasarkan metode SVM dan struktur pohon berdasarkan metode Decision Tree.

Di bawah ini merupakan Confusion Matrix digunakan untuk mengukur performa model SVM.



Gambar 2. True Tabel Confusion Matrix

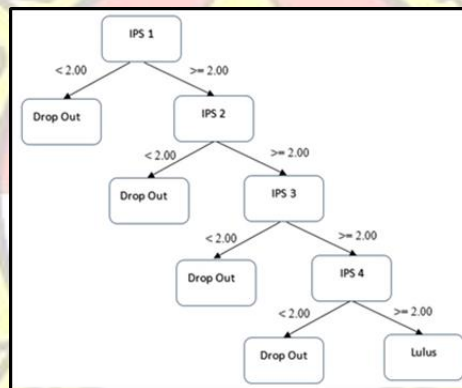
Confusion Matrix digunakan untuk mengukur performa model klasifikasi dengan membandingkan prediksi model terhadap data aktual. Pada kasus ini, Confusion Matrix menunjukkan hasil prediksi SVM yang sangat akurat dengan sedikit kesalahan klasifikasi, khususnya dalam memprediksi kelulusan atau Drop Out mahasiswa. Tabel 1 di bawah ini adalah confusion matrix yang menunjukkan empat kombinasi nilai.

Tabel 1. Confusion Matrix

| | | Actual Values | |
|------------------|----------|---------------|----------|
| | | Drop Out | Lulus |
| Predicted Values | Drop Out | 43 TN | 0 FN |
| | Lulus | 1 FP | 20 TP |

Tabel 1 merupakan nilai prediksi dari sistem (Predicted Values) dan nilai aktual dari data (Actual Values). Kombinasi ini terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Metode Decision Tree digunakan untuk memprediksi status kelulusan mahasiswa berdasarkan atribut seperti IPS, tagihan, dan kehadiran. Pada studi kasus ini, atribut dengan Gain Ratio tertinggi, seperti IPS semester 1, digunakan sebagai root node untuk membagi dataset menjadi dua cabang: satu untuk $IPS \geq 2.0$ dan satu lagi untuk $IPS < 2.0$. Proses ini berlanjut hingga mencapai keputusan akhir berdasarkan atribut yang paling relevan. Berikut merupakan struktur pohon berdasarkan metode Decision Tree.

Dibawah ini merupakan struktur pohon berdasarkan metode Decision Tree.

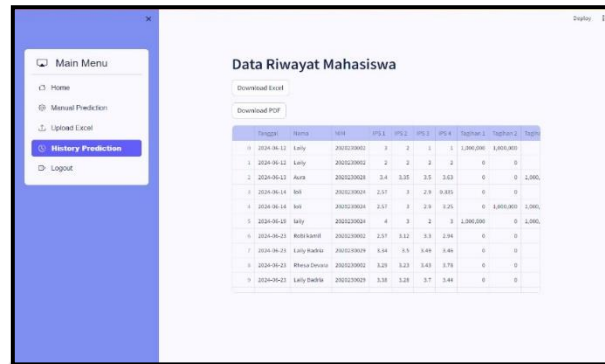


Gambar 1. Pohon Keputusan Studi Kasus

Gambar 3 di bawah ini adalah tampilan menu prediksi kelulusan mahasiswa pada website prediksi kelulusan.

Gambar 2. Form Prediksi

Di bawah ini terdapat Gambar 4 yang merupakan menu yang menampilkan hasil data riwayat mahasiswa yang telah diprediksi.



| Tanggal | Nama | NIM | PS1 | PS2 | PS3 | PS4 | Tugas 1 | Tugas 2 | Tugas 3 |
|------------|-------------|------------|------|------|------|-------|-----------|-----------|---------|
| 2024-06-12 | Lily | 2020230002 | 3 | 2 | 3 | 3 | 1,000,000 | 1,000,000 | 0 |
| 2024-06-12 | Lily | 2020230002 | 2 | 2 | 3 | 2 | 0 | 0 | 0 |
| 2024-06-12 | Aure | 2020230003 | 2.4 | 0.25 | 3.5 | 3.03 | 0 | 0 | 0 |
| 2024-06-14 | Siti | 2020230004 | 2.57 | 3 | 2.8 | 3.835 | 0 | 0 | 0 |
| 2024-06-14 | Siti | 2020230004 | 2.57 | 3 | 2.8 | 3.25 | 0 | 1,000,000 | 0 |
| 2024-06-18 | Lily | 2020230004 | 4 | 3 | 2 | 3 | 1,000,000 | 0 | 0 |
| 2024-06-23 | RubiKahar | 2020230002 | 2.57 | 3.12 | 3.3 | 2.94 | 0 | 0 | 0 |
| 2024-06-21 | LilyBakha | 2020230009 | 3.34 | 3.5 | 3.49 | 3.49 | 0 | 0 | 0 |
| 2024-06-23 | RhesaDevano | 2020230002 | 3.29 | 3.23 | 3.42 | 3.79 | 0 | 0 | 0 |
| 2024-06-23 | LilyBakha | 2020230003 | 3.38 | 3.28 | 3.17 | 3.44 | 0 | 0 | 0 |

Gambar 4. Data riwayat mahasiswa

5. Kesimpulan

Berdasarkan penelitian ini disimpulkan bahwa, dapat menerapkan metode Support Vector Machine (SVM) secara efisien dalam melakukan klasifikasi prediksi potensi Drop Out pada mahasiswa Program Studi Teknologi Informasi di Universitas Darma Persada. Model Support Vector Machine dapat mengidentifikasi mahasiswa yang berpotensi Drop Out secara dini berdasarkan kriteria yang sudah ditetapkan yaitu Indeks Prestasi Mahasiswa Semester 1 hingga 4, riwayat tagihan mahasiswa, serta kehadiran mahasiswa. Model SVM dalam penelitian ini mampu menunjukkan akurasi yang tinggi dalam melakukan klasifikasi mahasiswa yang berpotensi Drop Out. Merancang sistem website yang mudah digunakan, fungsional dan bermanfaat yang dapat mengidentifikasi mahasiswa yang berpotensi Drop Out secara dini. Pengembangan sistem website yang berfokus pada kriteria yang sudah ditetapkan seperti, yaitu Indeks Prestasi Mahasiswa Semester 1 hingga 4, riwayat tagihan mahasiswa, serta kehadiran mahasiswa. Dengan menerapkansistem website ini, diharapkan pihak universitas, khususnya pimpinan fakultas atau dekan dan dosen.

Daftar Pustaka

- [1] N. M. A. Mahar, Vihi Atina, and Nugroho Arif Sudibyo, "Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes" *J.Manaj. Inform. dan Sist. Inf.*, vol. 6, no. 2, 2023, doi: 10.36595/misi.v6i2.875.
- [2] A. Wibowo and A. Rohman, "Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 12, no. 2, 2022, doi: 10.36448/expert.v12i2.2810.
- [3] I. M. D. M. Rahayu, P. W., Sudipa, I. G. I., Suryani, A., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M. N., Slamet, I., Harlina, S., & Sanjaya, *Buku Ajar Data Mining*. Bandung: PT. Sonpedia Publishing Indonesia., 2024.
- [4] K. Pradnyana, G. A., & Agustini, *Konsep Dasar Data Mining*. 2022.
- [5] B. Raharjo, *Pembelajaran Mesin (Machine Learning)*. Semarang: Yayasan Prima Agus Teknik, 2021.
- [6] S. Sumirat, L. P., Cahyono, D., Kristyawan, Y., & Kacung, *Dasar-dasar Rekayasa Perangkat Lunak*, 1st ed. Madza Media, 2023.