

**KLASIFIKASI KATEGORI WAKTU KELULUSAN MAHASISWA
MENGGUNAKAN DATA AKADEMIK SEBAGAI UPAYA
PERINGATAN DINI BAGI MAHASISWA AKTIF MENGGUNAKAN
ALGORITMA DECISION TREE, NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE
STUDI KASUS : JURUSAN SISTEM INFORMASI UNIVERSITAS
SRIWIJAYA**

SKRIPSI

Program Studi Sistem Informasi

Jenjang Sarjana



Oleh

Nur Hanna

09031281722080

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2022

HALAMAN PENGESAHAN

**KLASIFIKASI KATEGORI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN DATA
AKADEMIK SEBAGAI UPAYA PERINGATAN DINI BAGI MAHASISWA AKTIF
MENGGUNAKAN ALGORITMA DECISION TREE, NAÏVE BAYES DAN
SUPPORT VECTOR MACHINE**

STUDI KASUS : JURUSAN SISTEM INFORMASI UNIVERSITAS SRIWIJAYA

Sebagai salah satu syarat untuk penyelesaian
studi di Program Studi Sistem Informasi S1

Oleh

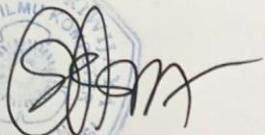
Nur Hanna

09031281722080

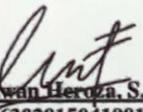
Palembang, Agustus 2022

Pembimbing

Mengetahui,
Ketua Jurusan,



Endang Lestari Ruskan, S.Kom, M.T.
NIP 19781116042001


Rahmat Iwan Herza, S.T.,M.T.
NIP 198706302015041001

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Nur Hanna

NIM : 09031281722080

Program Studi : Sistem Informasi

Judul Skripsi : Klasifikasi Kategori Waktu Kelulusan Mahasiswa menggunakan Data Akademik sebagai Upaya Peringatahn Dini Bagi Mahasiswa Aktif Menggunakan Algoritma Decision Tree, Naïve Bayes dan Support Vector Machine. Studi Kasus : Jurusan Sistem Informasi Universitas Sriwijaya

Hasil Pengecekan Software iThenticate/Turnitin : 8 %

Menyatakan bahwa laporan skripsi Saya merupakan hasil karya sendiri dan bukan penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka Saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan berlaku.

Demikian pernyataan ini Saya buat dengan sebenarnya dan tidak ada unsur paksaan dari siapapun.



NIM 09031281722080

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Senin

Tanggal : 01 Agustus 2022

Nama : Nur Hanna

NIM : 09031281722080

Judul Skripsi : Klasifikasi Kategori Waktu Kelulusan Mahasiswa menggunakan Data Akademik sebagai Upaya Peringatan Dini Bagi Mahasiswa Aktif Menggunakan Algoritma Decision Tree, Naïve Bayes dan Support Vector Machine. Studi Kasus : Jurusan Sistem Informasi Universitas Sriwijaya

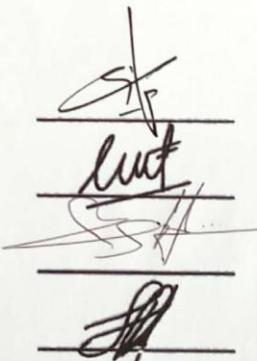
Tim Penguji :

1. Ketua : Ari Wedhasmara, M.TI.

2. Pembimbing : Rahmat Izwan Heroza, M.T

3. Penguji I : Yadi Utama, M.Kom.

4. Penguji II : Bayu Wijaya Putra, M.Kom.



Mengetahui,

Ketua Jurusan Sistem Informasi,



Endang Lestari Ruskan, S.Kom, M.T.

NIP 19781116042001

HALAMAN PERSEMBAHAN

MOTTO

**“Raihlah masa depanmu sendiri, lakukan usaha untuk mencapainya,
Tidak ada orang yang akan melakukannya untukmu”**

Skripsi ini saya persembahkan kepada :

1. Nur Hanna
2. Orang tua
3. Dosen Pebimbing & Penguji
4. Jurusan Sistem Informasi
5. Universitas Sriwijaya

**KLASIFIKASI KATEGORI WAKTU KELULUSAN MAHASISWA
MENGGUNAKAN DATA AKADEMIK SEBAGAI UPAYA
PERINGATAN DINI BAGI MAHASISWA AKTIF MENGGUNAKAN
ALGORITMA DECISION TREE, NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE
STUDI KASUS : JURUSAN SISTEM INFORMASI UNIVERSITAS
SRIWIJAYA**

Oleh

Nur Hanna 09031281722080

ABSTRAK

Kelulusan tepat waktu merupakan salah satu diantara indikator keberhasilan mahasiswa dan kampus. Mahasiswa diharapkan lulus tepat waktu dalam waktu 4 tahun atau kurang. Dalam praktiknya secara nyata, mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun, hal tersebut perlu dievaluasi, namun belum ada data yang valid untuk menentukan penyebab angka kelulusan yang tidak tepat waktu. Memahami mahasiswa yang lulus tepat waktu, sangat penting bagi pendidik atau kampus untuk upaya peringatan dini yang dapat mengarah pada keberhasilan mahasiswa dalam jangka panjang agar kelulusan tepat waktu dapat ditingkatkan. Menggunakan data akademik akan dilakukan klasifikasi untuk mengetahui kategori masa studi mahasiswa. Algoritma yang digunakan adalah Decision Tree, Naïve Bayes dan Support Vector Machine (SVM). Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah sebuah standar yang sudah dikembangkan yang digunakan untuk membantu proses analisis dari suatu industri sebagai strategi penyelesaian masalah untuk perusahaan atau departemen penelitian.

Kata kunci : Data mining, Prediksi, Naïve bayes, Decision Tree, Support Vector Machine, Waktu Kelulusan Mahasiswa

**CLASSIFICATION OF STUDENT GRADUATION TIME
CATEGORIES USING ACADEMIC DATA AS EARLY WARNING
FOR ACTIVE STUDENTS USING DECISION TREE, NAVE BAYES
AND SUPPORT VECTOR MACHINE ALGORITHM
CASE STUDY : DEPARTMENT OF INFORMATION SYSTEMS
UNIVERSITY SRIWIJAYA**

By

Nur Hanna 09031281722

ABSTRACT

Graduating on time is one of the indicators of student and campus success. Students are expected to graduate on time in 4 years or less. In real practice, students are not always able to complete undergraduate education within four years, this needs to be evaluated, but there is no valid data to determine the cause of the graduation rate that is not on time. Understanding students who graduate on time is very important for educators or educators. campus for early warning efforts that can lead to student success in the long term so that timely graduation can be improved. Using academic data, classification will be carried out to determine the category of student study period. The algorithms used are Decision Tree, Naïve Bayes and Support Vector Machine (SVM). Cross-Industry Standard Process for Data Mining (CRISP-DM) is a standard that has been developed that is used to assist the analysis process of an industry as a problem solving strategy for companies or research departments.

Keywords: Data mining, prediction, Nave Bayes, Decision Tree, Support Vector Machine, Student Graduation Time

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas segala limpahan berkah, rahmat dan hidayah-Nya sehingga penulis dapat melaksanakan dan menyelesaikan Tugas Akhir ini yang berjudul “Klasifikasi Kategori Waktu Kelulusan Mahasiswa menggunakan Data Akademik sebagai Upaya Peringatan Dini Bagi Mahasiswa Aktif Menggunakan Algoritma Decision Tree, Naïve Bayes dan Support Vector Machine. Studi Kasus : Jurusan Sistem Informasi Universitas Sriwijaya”. Tugas Akhir ini disusun sebagai salah satu syarat untuk penyelesaian studi di Program Studi Sistem Informasi Reguler Universitas Sriwijaya.

Dalam proses penyelesaian Tugas Akhir ini penulis dibantu oleh berbagai pihak. Bantuan tersebut berupa bimbingan, pengarahan, nasehat dan pemikiran dalam menyelesaikan Tugas Akhir ini. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada :

1. Ibu Endang Lestari Ruskan, M.T sebagai Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Rahmat Izwan Heroza, M.T selaku Dosen Pembimbing Tugas Akhir yang telah bersedia meluangkan waktunya untuk memberikan arahan beserta saran dalam penyusunan Tugas Akhir ini
3. Bapak Ari Wedhasmara, M.TI., Bapak Yadi Utama, M.Kom., dan Bapak Bayu Wijaya Putra, M.Kom. selaku Dosen Pengudi
4. Dosen Sistem Informasi Universitas Sriwijaya yang telah membagikan ilmu pengetahuan selama penulis mengikuti perkuliahan di Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya..

5. Orangtua, yang memberikan doa dan dukungan pendidikan kepada penulis
6. Staff Administrasi Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Teman yang sudah memberikan dukungan, bantuan dan saran yang bermanfaat selama masa perkuliahan

Penulis menyadari didalam penulisan proposal ini masih terdapat banyak kekurangan dalam segi susunan dan tata cara penulisan. Oleh karena itu, penulis mengharapkan masukan dan kritikan yang bersifat membangun guna penyempurnaan skripsi ini. *Wassalamu 'alaikum warahmatullahi wabarakatuh.*

Palembang, Agustus 2022

Penulis

Nur Hanna

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERSEMBAHAN	v
ABSTRAK.....	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xii
BAB I	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
BAB II	5
2.1 Profil Jurusan Sistem Informasi (Studi Kasus)	5
2.2 Kelulusan Mahasiswa.....	6
2.3 Data Mining.....	7
2.3.1 Pengertian Data Mining	7
2.3.2 Tahapan Data Mining.....	7
2.3.3 Tugas Data Mining	9
2.4 Data Classification.....	10
2.5 Decision Tree	11
2.5.1 Pengertian Decision Tree	11
2.5.2 Langkah-langkah Decision Tree	11
2.6 Naïve Bayes.....	14
2.7 Support Vector Machine	16
2.8 CRISP-DM	17
BAB III.....	20
3.1 Studi Kasus Penelitian	20
3.2 Material Penelitian.....	20
3.2.1 Jenis Data	20
3.2.2 Sumber Data	20

3.3 Metode Penelitian	20
3.3.1 <i>Business Understanding</i> (Fase Pemahaman Bisnis).....	21
3.3.2 <i>Data Understanding</i> (Fase Pemahaman Data)	21
3.3.3 <i>Data Preparation</i> (Fase Persiapan Data)	22
3.3.3.1 Train Test Split	23
3.3.4 <i>Modeling</i>	23
3.3.5 <i>Evaluation</i>	25
3.3.6 Deployment	25
BAB IV.....	26
4.1 <i>Business Understanding</i> (Fase Pemahaman Bisnis).....	26
4.2 <i>Data Understanding</i> (Fase Pemahaman Data)	26
4. 3 Data Preparation (Fase Persiapan Data)	29
4.3.1 Train Test Split	39
4.4 Modeling.....	40
4.4.1 Decision Tree(C4.5).....	40
4.4.2 Naïve Bayes.....	46
4.4.3 Support Vector Machine.....	51
4.5 Evaluation	56
BAB V.....	58
5.1 Kesimpulan.....	58
5.2 Saran	58
DAFTAR PUSTAKA	59
LAMPIRAN	lxii

DAFTAR GAMBAR

Gambar 1 Fase CRISP-DM	17
Gambar 2 Grafik Perbandingan Jumlah Mahasiswa.....	27
Gambar 3 Perbandingan Waktu Kelulusan Mahasiswa.....	28
Gambar 4 Confusion matrix Decision Tree Percobaan ke 1.....	42
Gambar 5 Confusion matrix Decision Tree Percobaan ke 2.....	43
Gambar 6 Confusion matrix Naïve Bayes Percobaan ke 1.....	48
Gambar 7 Confusion matrix Naïve Bayes Percobaan ke 2.....	48
Gambar 8 Confusion matrix Support Vector Machine Percobaan ke 1.....	53
Gambar 9 Confusion matrix Support Vector Machine Percobaan ke 2.....	53

DAFTAR TABEL

Tabel 1 Dataset Sebelum Proses Preparation.....	29
Tabel 2 Konversi Dataset ke Numerik.....	30
Tabel 3 Dataset Data Akademik Mahasiswa	34
Tabel 4 Dataset yang Akan digunakan untuk membuat model	38

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kelulusan tepat waktu merupakan salah satu diantara indikator keberhasilan mahasiswa dan kampus. Kualitas kinerja program studi dapat ditingkatkan dilihat dari pencapaian status akreditasi program studi, dimana salah satu unsur instrumen penilaian terkait profil lulusan (Gunawan et al., 2019). Salah satu elemen evaluasi yang mempengaruhi akreditasi BAN-PT (Badan Akreditasi Nasional Perguruan Tinggi) adalah profil lulusan. Profil lulusan ada dua, salah satunya yaitu persentase kelulusan tepat waktu (Salim et al., 2020).

Mahasiswa diharapkan lulus tepat waktu dalam waktu 4 tahun atau kurang. Dalam praktiknya secara nyata, mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun (Wirawan et al., 2019). Hal tersebut perlu dievaluasi, namun belum ada data yang valid untuk menentukan penyebab angka kelulusan yang tidak tepat waktu (Wibowo et al., 2018).

Memahami mahasiswa yang lulus tepat waktu, sangat penting bagi pendidik atau kampus untuk mengidentifikasi faktor-faktor yang relevan dan berkontribusi yang dapat mengarah pada keberhasilan mahasiswa dalam jangka panjang agar kelulusan tepat waktu dapat ditingkatkan (Gopalakrishnan et al., 2018). Hal itu juga dapat digunakan oleh pihak kampus untuk membantu mengambil keputusan sebagai sistem peringatan dini target kelulusan mahasiswa (Wibowo et al., 2018) (D. Y. Putri et al., 2019). Mengetahui faktor

yang mempengaruhi kelulusan mahasiswa dapat meningkatkan kinerja jurusan dengan mendeteksi siswa yang kemungkinan besar tidak dapat lulus tepat waktu lebih awal, sehingga membantu para siswa tersebut meningkatkan prestasi belajarnya, dan dapat lulus tepat waktu(Mohammad Suhaimi et al., 2019).

Beberapa penelitian menunjukkan bahwa faktor akademik merupakan hal terpenting untuk dipertimbangkan dalam metode klasifikasi untuk tujuan akademik(Supianto et al., 2018). Berdasarkan catatan akademik sebelumnya, peneliti akan melakukan klasifikasi untuk mengetahui masa studi mahasiswa. Variabel target terdiri dari 2 yaitu tepat waktu dan tidak tepat waktu, tepat waktu yaitu lulus dalam waktu 4 tahun atau kurang dan tidak tepat waktu yaitu lulus dalam waktu lebih dari 4 tahun masa studi. Algoritma yang digunakan adalah Decision Tree, Naïve Bayes dan Support Vector Machine (SVM). Decision Tree merupakan salah satu algoritma yang paling terkenal karena proses perumusannya yang relatif cepat, mudah dipahami dan sederhana (Supianto et al., 2018). Naive Bayes adalah teknik pemodelan probabilistik yang relatif sederhana dan telah menunjukkan hasil yang bagus terlepas dari asumsi kemandirian fitur-fiturnya dalam membuat prediksi kelas keluaran (Previde et al., 2019). Hasil penelitian menunjukkan Support Vector Machine (SVM) mencapai akurasi yang lebih tinggi karena SVM bermanfaat untuk masalah klasifikasi biner, yang dalam konteks ini, tepat waktu atau tidak dan SVM bekerja sangat bagus dengan kumpulan data kecil(Mohammad Suhaimi et al., 2019). Data akademik mahasiswa yang digunakan adalah NIM, IPS(indeks prestasi semester) dari semester 1-4, jumlah SKS(sistem kredit semester) dari semester 1-4 dan lama waktu studi. Menurut penelitian sebelumnya mengatakan IPS rendah selama tahun pertama dan kedua masa studi, memungkinkan mereka

tidak dapat menyelesaikan studi sesuai rencana, alasannya mereka harus menunda skripsi ke semester depan yang kemungkinan besar akan memperpanjang waktu studi mereka (Mohammad Suhaimi et al., 2019) dan berdasarkan penelitian Supianto et al. atribut SKS berperan utama dalam pengambilan keputusan kelulusan tepat waktu dengan metode klasifikasi (Supianto et al., 2018)

1.2 Rumusan Masalah

Bagaimana menerapkan data mining dalam melakukan klasifikasi data akademik untuk menentukan kategori waktu kelulusan(tepat waktu/tidak) mahasiswa ?

1.3 Tujuan Penelitian

Menerapkan data mining untuk melakukan prediksi kategori waktu kelulusan mahasiswa (tepat waktu / tidak) berdasarkan data akademik mahasiswa.

1.4 Manfaat Penelitian

1. Mengetahui mahasiswa yang diperkirakan akan lulus tidak tepat waktu
2. Dapat dijadikan acuan dan bimbingan bagi mahasiswa agar lulus tepat waktu
3. Mengurangi resiko terjadinya drop out
4. Membantu praktisi merancang dan menerapkan intervensi pencegahan dini kepada mahasiswa yang beresiko terlambat lulus.

1.5 Batasan Masalah

Penelitian dilakukan menggunakan data akademik jurusan sistem informasi universitas sriwijaya 9 tahun terakhir dari angkatan tahun 2008 sampai tahun 2016. Algoritma yang digunakan adalah Decision Tree, Support Vector Machine(SVM) dan Naïve Bayes.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2015). Data Mining. In *Cancer Letters* (Vol. 14, Issue 3).
[https://doi.org/10.1016/0304-3835\(81\)90152-X](https://doi.org/10.1016/0304-3835(81)90152-X)
- Burman, I., & Som, S. (2019). Predicting Students Academic Performance Using Support Vector Machine. *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 756–759.
<https://doi.org/10.1109/AICAI.2019.8701260>
- Chapman, P. (1999). The CRISP-DM User Guide. *The CRISP-DM User Guide*, 14.
- Gopalakrishnan, A., Kased, R., Yang, H., Love, M. B., Graterol, C., & Shada, A. (2018). A multifaceted data mining approach to understanding what factors lead college students to persist and graduate. *Proceedings of Computing Conference 2017, 2018-Janua*(July), 372–381.
<https://doi.org/10.1109/SAI.2017.8252128>
- Gunawan, Hanes, & Catherine. (2019). Information Systems Students' Study Performance Prediction Using Data Mining Approach. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*. <https://doi.org/10.1109/ICIC47613.2019.8985718>
- Kusrini, E. taufiq luthfi. (2009). *Algoritma Data Mining*. Andi.
- Mohammad Suhaimi, N., Abdul-Rahman, S., Mutalib, S., Abdul Hamid, N. H., & Hamid, A. (2019). Review on Predicting Students' Graduation Time Using Machine Learning Algorithms. *International Journal of Modern Education and Computer Science*, 11(7), 1–13.

<https://doi.org/10.5815/ijmecs.2019.07.01>

Ojha, T., Heileman, G. L., Martinez-Ramon, M., & Slim, A. (2017).

Prediction of graduation delay based on student performance.

Proceedings of the International Joint Conference on Neural Networks,

2017-May, 3454–3460. <https://doi.org/10.1109/IJCNN.2017.7966290>

Previde, P., Graterol, C., Love, M. B., & Yang, H. (2019). A Data Mining

Approach to Understanding Curriculum-Level Factors That Help

Students Persist and Graduate. *Proceedings - Frontiers in Education*

Conference, FIE, 2019-Octob, 1–9.

<https://doi.org/10.1109/FIE43999.2019.9028488>

Pujianto, U., Azizah, E. N., & Damayanti, A. S. (2018). Naive Bayes using to

predict students' academic performance at faculty of literature.

Proceeding - 2017 5th International Conference on Electrical,

Electronics and Information Engineering: Smart Innovations for

Bridging Future Technologies, ICEEIE 2017, 2018-Janua, 163–169.

<https://doi.org/10.1109/ICEEIE.2017.8328782>

Putri, D. Y., Andreswari, R., & Hasibuan, M. A. (2019). Analysis of Students

Graduation Target Based on Academic Data Record Using C4.5

Algorithm Case Study: Information Systems Students of Telkom

University. *2018 6th International Conference on Cyber and IT Service*

Management, CITSM 2018, Citsm, 1–6.

<https://doi.org/10.1109/CITSM.2018.8674366>

Salim, A. P., Laksitowening, K. A., & Asror, I. (2020). Time Series Prediction

on College Graduation Using KNN Algorithm. *2020 8th International*

Conference on Information and Communication Technology (ICoICT).

Samuel, Y. T., Hutapea, J. J., & Jonathan, B. (2019). Predicting the timeliness of student graduation using decision tree c4.5 algorithm in universitas advent Indonesia. *Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019*, 276–280. <https://doi.org/10.1109/ICTS.2019.8850948>

Supianto, A. A., Julisar Dwitama, A., & Hafis, M. (2018). Decision Tree Usage for Student Graduation Classification: A Comparative Case Study in Faculty of Computer Science Brawijaya University. *3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018 - Proceedings*, 308–311.

<https://doi.org/10.1109/SIET.2018.8693158>

Wibowo, S., Andreswari, R., & Hasibuan, M. A. (2018). Analysis and design of decision support system dashboard for predicting student graduation time. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2018-Octob*, 684–689.

<https://doi.org/10.1109/EECSI.2018.8752876>

Wirawan, C., Khudzaeva, E., Hasibuan, T. H., Karjono, & Lubis, Y. H. K. (2019). Application of Data mining to Prediction of Timeliness Graduation of Students (A Case Study). *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*, 18–21. <https://doi.org/10.1109/CITSM47753.2019.8965425>