

**PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU
DENGAN METODE *RANDOM FOREST*
BERDASARKAN ALGORITMA *K-MEANS* PADA
PASCASARJANA UNIVERSITAS SRIWIJAYA**



OLEH :

M. HAFIZH SYTAR

09012682327018

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2025**

**PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU
DENGAN METODE *RANDOM FOREST*
BERDASARKAN ALGORITMA *K-MEANS* PADA
PASCASARJANA UNIVERSITAS SRIWIJAYA**

TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Magister



OLEH :

M. HAFIZH SYTAR

09012682327018

PROGRAM STUDI MAGISTER ILMU KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2025

LEMBAR PENGESAHAN

Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode *Random Forest* Berdasarkan Algoritma *K-Means* pada Pascasarjana Universitas Sriwijaya

TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister

OLEH:

**M. Hafizh Sytar
09012682327018**

Palembang, 26 Juni 2025

Pembimbing



**Prof. Dr. Ermatita, M.Kom
NIP 196709132006042001**

**Mengetahui,
Ketua Jurusan Magister Teknik Informatika**



**Dr. Firdaus, M.Kom
NIP 197801212008121003**

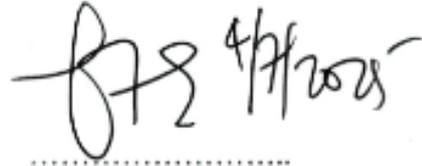
HALAMAN PERSETUJUAN

Pada hari Kamis, tanggal 26 Juni 2025 telah dilaksanakan ujian sidang tesis oleh Magister Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : M. Hafizh Sytar
NIM : 09012682327018
Judul : Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode *Random Forest* Berdasarkan Algoritma *K-Means* pada Pascasarjana Universitas Sriwijaya

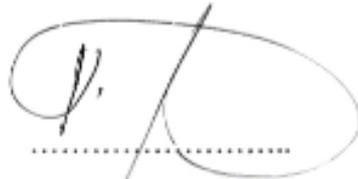
1. Ketua Penguji

Prof. Dr. Ir. Bambang Tutuko, M.T
NIP 196001121989031002



2. Penguji I

Dr. Ali Ibrahim, S.Kom., M.T
NIP 198407212019031004



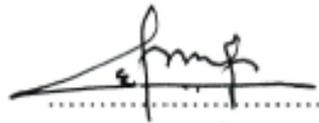
3. Penguji II

M. Qurhanul Rizqie, S.Kom., M.T, Ph.D
NIP 198712032022031006

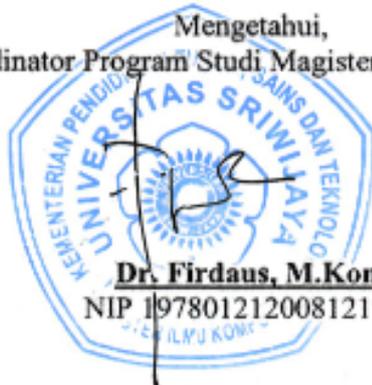


4. Pembimbing

Prof. Dr. Ermatita, M.Kom
NIP 196709132006042001



Mengetahui,
Koordinator Program Studi Magister Ilmu Komputer



Dr. Firdaus, M.Kom
NIP 197801212008121003

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : M. Hafizh Sytar
NIM : 09012681327018
Program Studi : Magister Ilmu Komputer
Judul Tesis : Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode *Random Forest* Berdasarkan Klasifikasi Algoritma *K-Means* pada Pascasarjana Universitas Sriwijaya.

Hasil Pengecekan Software iThenticate Turnitin : **19%**

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

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 26 Juni 2025



M. Hafizh Sytar
NIM. 09012682327018

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis ini dengan judul "PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU DENGAN METODE *RANDOM FOREST* BERDASARKAN ALGORITMA *K-MEANS* PADA PASCASARJANA UNIVERSITAS SRIWIJAYA" sebagai salah satu syarat untuk memperoleh gelar Magister Ilmu Komputer pada Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulisan tesis ini tidak akan terwujud tanpa bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, dengan kerendahan hati, izinkan penulis menyampaikan rasa terima kasih yang tulus kepada:

1. Orang Tua Tercinta Ayah Dr. Sutarjo, S.Pd., M.M. dan Ibu Desy Katarina, S.E., dan Adik saya Humaira Sytar, S.E., yang tiada henti memberikan kasih sayang, doa, dukungan moral, dan materi yang tak terhingga. Setiap langkah dalam perjalanan pendidikan ini adalah buah dari pengorbanan dan cinta kalian. Terima kasih atas kepercayaan yang selalu kalian berikan.
2. Kekasih Hati Saya Amanda Karmelita, A.Md. Li, Terima kasih telah menjadi bagian hidup penulis serta memberi pengertian, doa, kesabaran, dukungan, dan motivasi yang tak pernah putus selama proses penulisan tesis ini. Kehadiranmu menjadi penyeimbang dan penyemangat, sehingga penulis dapat menyelesaikan tesis ini.
3. Dosen Pembimbing: Ibu Prof. Dr. Ermatita, M.Kom., yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memberikan masukan yang sangat berharga sejak awal hingga akhir penyusunan tesis ini. Ilmu dan kebijaksanaan Ibu sangat membantu penulis dalam menyelesaikan penelitian ini.
4. Dosen Penguji Bapak Prof. Dr. Ir. Bambang Tutuko, M.T., Bapak Dr. Ali Ibrahim, S.Kom., M.T., dan Bapak M. Qurhanul Rizqie, S.Kom., M.T, Ph.D., yang telah memberikan kritik, saran, dan pertanyaan konstruktif demi kesempurnaan tesis ini.
5. Seluruh Dosen dan Staf Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer Universitas Sriwijaya, terima kasih atas ilmu yang telah dibagikan selama masa perkuliahan dan pelayanan administrasi yang selalu responsif.
6. Keluarga Besar Seluruh anggota keluarga yang selalu menjadi sumber kekuatan, inspirasi, dan kehangatan. Dukungan dan pengertian dari kalian

sangat berarti dalam setiap suka dan duka selama proses penyusunan tesis ini.

7. Sahabat dan Teman-teman Seperjuangan: Yanda Sonang Herlian, Fero Adjie Persanda, Fhajar Rindang Buana, Egga Sulaeman Fadhilah, M. Ade Triofani serta teman temanku yang lain yang selalu ada untuk berbagi cerita, canda tawa, serta dukungan moral di setiap momen sulit maupun senang. Diskusi dan semangat dari kalian sangat memotivasi.
8. Pihak-pihak lain yang Tidak Dapat Penulis Sebutkan Satu Per Satu, terima kasih atas segala bantuan, doa, dan dukungan yang telah diberikan, baik secara langsung maupun tidak langsung.
9. Terakhir, namun tidak kalah pentingnya, saya ingin berterima kasih kepada diri saya sendiri, saya ingin berterima kasih kepada diri saya sendiri karena telah percaya pada diri saya. Saya ingin berterima kasih kepada diri saya sendiri karena telah melakukan semua kerja keras ini. Saya ingin berterima kasih kepada diri saya sendiri karena tidak pernah libur. Saya ingin berterima kasih kepada diri saya sendiri karena tidak pernah menyerah. Saya ingin berterima kasih kepada diri saya sendiri karena selalu menjadi pemberi, dan mencoba memberi lebih dari yang saya terima. Saya ingin berterima kasih kepada diri saya sendiri karena mencoba berbuat lebih banyak kebaikan daripada kesalahan. Saya ingin berterima kasih kepada diri saya sendiri karena selalu menjadi diri saya sendiri setiap saat.

Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna. Oleh karena itu, segala kritik dan saran yang membangun akan senantiasa penulis terima dengan lapang dada demi perbaikan di masa yang akan datang. Semoga tesis ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan pihak-pihak yang membutuhkan.

Palembang, 26 Juni 2025



M. Hafizh Sytar

PREDICTION OF ON-TIME STUDENT GRADUATION WITH RANDOM FOREST METHOD BASED ON K-MEANS ALGORITHM AT POSTGRADUATE UNIVERSITY OF SRIWIJAYA

M. Hafizh Sytar · Ermatita

Abstract

On-time graduation is a crucial indicator for assessing the quality of higher education institutions. At the Postgraduate School of Sriwijaya University, delayed graduations remain a serious concern. This research aims to predict on-time student graduation by combining the random forest method with the K-means algorithm. The K-means algorithm is used to cluster students based on specific parameters or scopes, such as GPA, IPS, number of SKS, semester, age, and employment status. The results of this clustering are then used as input to build a predictive model using the random forest method. This research employs a data mining approach, involving stages of data collection, preprocessing, clustering, modeling, and evaluation using a confusion matrix and ROC curve. Experimental results show that this combined method provides high predictive *accuracy* in determining on-time student graduation, achieving an *accuracy* of 97.93%. This research is expected to serve as a basis for strategic decision-making by the university to enhance educational efficiency and improve the on-time graduation rate.

Keywords : Student Graduation, Prediction, Random forest, K-means, Data mining, Postgraduate, Universitas Sriwijaya

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU DENGAN METODE *RANDOM FOREST* BERDASARKAN ALGORITMA *K-MEANS* PADA PASCASARJANA UNIVERSITAS SRIWIJAYA

M. Hafizh Sytar · Ermatita

Abstrak

Kelulusan mahasiswa tepat waktu menjadi indikator penting dalam menilai kualitas institusi pendidikan tinggi. Di Pascasarjana Universitas Sriwijaya, masalah keterlambatan kelulusan masih menjadi perhatian serius. Penelitian ini bertujuan untuk memprediksi kelulusan tepat waktu mahasiswa dengan menggabungkan metode *Random Forest* dan algoritma *K-means*. Algoritma *K-means* digunakan untuk mengelompokkan mahasiswa berdasarkan batasan atau *scope* seperti IPK, IPS, jumlah SKS, semester, umur, dan status pekerjaan. Hasil pengelompokan kemudian digunakan sebagai input untuk membangun model prediksi menggunakan metode *Random Forest*. Penelitian ini dilakukan dengan pendekatan *data mining* melalui tahapan pengumpulan data, praproses, pengelompokan, modeling, serta evaluasi menggunakan *confusion matrix* dan *ROC curve*. Hasil eksperimen menunjukkan bahwa kombinasi metode ini mampu memberikan akurasi prediksi yang tinggi dalam menentukan kelulusan tepat waktu mahasiswa yang menghasilkan akurasi sebesar 97.93%. Penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan strategis pihak universitas guna meningkatkan efisiensi pendidikan dan rasio kelulusan tepat waktu.

Kata Kunci : Kelulusan Mahasiswa, Prediksi, *Random forest*, *K-means*, *Data mining*, *Pascasarjana*, Universitas Sriwijaya

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
LEMBAR PERNYATAAN	iv
KATA PENGANTAR	iv
Abstract	vii
Abstrak	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR SINGKATAN	xiv
DAFTAR LAMPIRAN	xv
BAB I . PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II . TINJAUAN PUSTAKA	7
2.1 Kerangka Teori	7
2.1.1 <i>Data mining</i>	7
2.1.2 <i>Clustering</i> (Pengelompokan)	11
2.1.3 Algoritma <i>K-means</i>	12
2.1.4 Algoritma <i>Random forest</i>	13
2.1.5 Kelulusan Mahasiswa	16
2.1.6 Prediksi Kelulusan	16
2.1.7 Evaluasi dan Validasi Hasil	17
2.1.8 <i>Confusion matrix</i>	17
2.2 Kerangka Pikir	20
BAB III . METODOLOGI PENELITIAN	23

3.1	Kerangka Kerja Penelitian.....	23
3.2	Operasional Variabel.....	25
3.3	Metode Analisis Data.....	25
3.3.1	Tahap Pengumpulan Data.....	27
3.3.2	Tahap Pengolahan Awal Data.....	27
3.3.3	Pengelompokan <i>K-means</i>	28
3.3.4	Model Prediksi (<i>Random forest</i>).....	31
3.3.5	Evaluasi dan Pengolahan hasil.....	35
BAB IV . HASIL DAN ANALISA		37
4.1	Pengumpulan Data	37
4.2	Pengolahan Awal Data	38
4.3	Pengelompokan (<i>K-means</i>)	39
4.4	Model Prediksi (<i>Random forest</i>)	47
4.5	Perhitungan Manual	52
4.5.1	Perhitungan Manual <i>K-means</i>	52
4.5.2	Perhitungan Manual <i>Random forest</i>	59
4.6	Implementasi dengan Menggunakan Python	62
4.7	Evaluasi dan Pengolahan Hasil	75
4.8	Pemilihan Model Terbaik.....	79
BAB V. KESIMPULAN DAN SARAN.....		82
5.1	Kesimpulan.....	82
5.2	Saran.....	82
DAFTAR PUSTAKA		84

DAFTAR GAMBAR

	Halaman
Gambar 2.1. <i>Data mining</i>	8
Gambar 2.2. Taksonomi Peranan <i>Data mining</i>	11
Gambar 2.3. Klastering dalam <i>K-means</i>	13
Gambar 2.4. <i>Random forest</i>	15
Gambar 2.5. Grafik ROC.....	19
Gambar 3.1. Kerangka Penelitian.....	23
Gambar 3.2. Flowchart Pengolahan Data.....	26
Gambar 3.3. Tahapan <i>K-means</i>	29
Gambar 3.4. Tahapan <i>Random forest</i>	32
Gambar 4.1. Proses model <i>k-means</i>	51
Gambar 4.2. Tampilan data hasil <i>clustering</i>	52
Gambar 4.3. Visualisasi data <i>clustering</i>	53
Gambar 4.4. <i>Cluster model</i>	54
Gambar 4.5. Hasil <i>centroid</i>	55
Gambar 4.6. Hasil plot <i>cluster model</i>	55
Gambar 4.7. Model <i>random forest</i>	59
Gambar 4.8. <i>Set role</i>	61
Gambar 4.9. Hasil <i>apply model</i> dari metode <i>random forest</i>	61
Gambar 4.10. <i>Plot apply model</i>	62
Gambar 4.11. Contoh data untuk <i>centroid</i>	63
Gambar 4.12. Hasil <i>load Dataset</i>	73
Gambar 4.13. Tampilan hasil perubahan nama kolom	75
Gambar 4.14. Tampilan data X dan Y setelah pemisahan kolom	76
Gambar 4.15. Tampilan hasil label <i>encoder</i>	76
Gambar 4.16. Tampilan hasil penanganan nilai NaN.....	77
Gambar 4.17. Hasil <i>cluster k-means</i>	78
Gambar 4.18. Hasil <i>accuracy random forest</i>	80
Gambar 4.19. Hasil prediksi <i>random forest</i>	82
Gambar 4.20. Hasil <i>accuracy model random forest</i>	85
Gambar 4.21. Hasil <i>precision random forest</i>	86

Gambar 4.22. Hasil <i>recall random forest</i>	87
Gambar 4.23. Hasil AUC (<i>optimistic</i>)	88
Gambar 4.24. Hasil AUC	88
Gambar 4.25. Hasil AUC (<i>pessimistic</i>).....	88
Gambar 4.26. Hasil <i>performance vector</i>	89

DAFTAR TABEL

	Halaman
Tabel 3.1 <i>Dataset mahasiswa</i>	28
Tabel 3.2 <i>Preprocessing Data</i>	28
Tabel 3.3 Data hasil <i>clustering</i>	30
Tabel 3.4 Hasil <i>decision tree</i> 1.....	32
Tabel 3.5 Hasil <i>decision tree</i> 2.....	33
Tabel 3.6 Hasil <i>decision tree</i> 3.....	34
Tabel 3.7 Prediksi keseluruhan.....	34
Tabel 3.8 Evaluasi Akhir.....	35
Tabel 4.1 <i>Dataset mahasiswa</i>	38
Tabel 4.2 <i>Dataset penelitian</i>	57
Tabel 4.3 <i>Centroid awal</i>	63
Tabel 4.4 Data awal.....	64
Tabel 4.5 Tabel jarak <i>Euclidean distance</i>	68
Tabel 4.6 Hasil <i>clustering</i>	83
Tabel 4.7 Hasil prediksi data.....	84
Tabel 4.8 Data yang diberi.....	86

DAFTAR SINGKATAN

IPK	= Indeks Prestasi Kumulatif
IPS	= Indeks Prestasi Semester
SKS	= Sistem Kredit Semester
CRISP-DM	= <i>Cross Industry Standard Process for Data Mining</i>
CART	= <i>Classification and Regression Trees</i>
OOB	= <i>Out Of Bag</i>
TP	= <i>True Positif</i>
FP	= <i>False Positif</i>
TN	= <i>True Negatif</i>
FN	= <i>False Negatif</i>
kNN	= <i>K-Nearest Neighbor</i>
ROC	= <i>Receiver Operating Characteristic</i>
AUC	= <i>Area Under the Curve</i>

DAFTAR LAMPIRAN

Lampiran 1. Halaman Dataset

Lampiran 2. Halaman Jurnal

Lampiran 3. Halaman Pengecekan Turnitin

Lampiran 4. Halaman Pengecekan Similarity

Lampiran 5. Halaman Asistensi Sidang Komprehensif

Lampiran 6. Halaman Surat Rekomendasi Melaksanakan Ujian Tesis

Lampiran 7. Halaman Surat Keputusan Pengangkatan Pembimbing Tesis

Lampiran 8. Halaman Hasil *USEPT*

BAB I. PENDAHULUAN

1.1 Latar Belakang

Pendidikan yang berkualitas pada suatu negara berawal dari kualitas pendidikan masing-masing daerah. Kualitas pendidikan masing-masing daerah di negara Indonesia belum cukup merata dan belum maksimal (Nugraha & Hairani, 2018). Kualitas pendidikan menjadi kurang bagus bisa terjadi karena terletak di daerah yang jaraknya jauh dari pusat kota (Khasanah dkk., 2022). Pendidikan tinggi memainkan peran penting dalam pengembangan sumber daya manusia yang berkualitas. Tingkat kelulusan mahasiswa tepat waktu menjadi salah satu indikator keberhasilan institusi pendidikan tinggi. Namun, pada kenyataannya, banyak mahasiswa yang mengalami keterlambatan dalam menyelesaikan studi mereka. Di Pascasarjana Universitas Sriwijaya, masalah kelulusan tepat waktu ini juga menjadi perhatian utama.

Faktor-faktor yang mempengaruhi kelulusan mahasiswa mencakup berbagai aspek seperti latar belakang akademik, kondisi sosial ekonomi, keterlibatan dalam kegiatan kampus dan dukungan finansial. Dukungan finansial seperti beasiswa, bantuan biaya pendidikan, dan kemampuan ekonomi keluarga, merupakan faktor penting yang dapat mempengaruhi kelulusan tepat waktu. Mahasiswa yang menghadapi kesulitan finansial seringkali harus bekerja paruh waktu untuk membantu dalam mendanai pendidikan mereka yang tentunya dapat mengurangi waktu dan energi yang didedikasikan dalam menyelesaikan studi. Oleh karena itu, diperlukan metode yang efektif untuk memprediksi kelulusan mahasiswa tepat waktu guna mengidentifikasi dan mengatasi masalah-masalah yang dapat menghambat proses pendidikan mereka.

Metode prediksi yang akurat dapat membantu pihak universitas dalam mengambil langkah-langkah preventif dan korektif untuk meningkatkan tingkat kelulusan tepat waktu. Dalam penelitian ini berfokus pada penggunaan metode *Random forest* untuk memprediksi kelulusan mahasiswa tepat waktu. *Random forest* dikenal sebagai salah satu algoritma machine learning yang memiliki akurasi tinggi dan kemampuan generalisasi yang baik. Algoritma ini bekerja

dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat dan stabil (Tsai & Hung, 2014).

Algoritma *Random forest*, berbasis pohon keputusan, telah diaplikasikan dalam konteks pendidikan untuk memprediksi keberhasilan akademik. Breiman (2001) memperkenalkan *Random forest* sebagai teknik yang mampu mengatasi overfitting dan menghasilkan prediksi yang lebih akurat pada data yang kompleks. Kemudian, Kotsiantis dkk. (2006) mengembangkan *Random forest* sebagai model yang efektif dalam prediksi kelulusan mahasiswa karena kemampuannya dalam memproses berbagai faktor akademik yang saling terkait. Keunggulan lain dari algoritma ini adalah kemampuan dalam menangani *Dataset* yang tidak seimbang dalam mengidentifikasi variabel kritis melalui fitur *feature importance* yang memungkinkan pihak universitas membuat kebijakan berbasis data yang lebih efektif. Penggunaan *Random forest*, tidak hanya memberi prediksi yang akurat tetapi juga membantu memahami faktor yang mempengaruhi kelulusan mahasiswa tepat waktu sehingga memungkinkan intervensi dini dan perencanaan strategis yang lebih baik oleh pihak universitas.

Selain itu, penelitian ini juga menggunakan algoritma *K-means*. Algoritma *K-means* menggunakan metode klusterisasi untuk mengelompokkan data dengan cara mengidentifikasi titik pusat kluster (*centroid*) yang paling dekat dengan data. Algoritma *clustering*, seperti *K-means*, telah digunakan untuk mengelompokkan data mahasiswa dengan tujuan menemukan pola dalam data pendidikan. Han dkk., (2012) menyatakan bahwa *K-means* efektif untuk mengelompokkan data dalam bidang pendidikan karena kemampuannya dalam mengidentifikasi pola tersembunyi. Dalam studi serupa, Romero dan Ventura (2013) menjelaskan bahwa *clustering* membantu institusi pendidikan mengelompokkan mahasiswa berdasarkan atribut tertentu, seperti tingkat IPK dan partisipasi di kelas, yang berpengaruh pada performa akademik mahasiswa.

Identifikasi faktor-faktor yang mempengaruhi keberhasilan akademik mahasiswa, termasuk latar belakang akademik, kehadiran, partisipasi dalam kelas, dan jumlah SKS yang diambil. Penelitian ini menunjukkan bahwa faktor-faktor tersebut merupakan variabel yang dapat dimasukkan ke dalam model prediksi untuk memperkirakan kelulusan tepat waktu. Penelitian lebih lanjut oleh Dekker

dkk., (2009) juga menunjukkan bahwa faktor-faktor akademik dan demografi memberikan kontribusi signifikan pada prediksi kelulusan mahasiswa. Penggabungan metode *Random forest* dan *K-means* ini memungkinkan analisis yang lebih mendalam dan pembuatan model prediksi yang lebih kuat sehingga dapat membantu pihak universitas dalam merancang program dan kebijakan yang lebih tepat sasaran. Selain itu, telah dilakukan beberapa penelitian terdahulu terkait prediksi kelulusan mahasiswa pada program sarjana, namun kajian spesifik untuk Pascasarjana Universitas Sriwijaya masih terbatas. Oleh karena itu, penelitian ini memiliki urgensi tinggi dalam mengisi kekosongan literatur khusus pada tingkat pascasarjana.

Penelitian ini tidak hanya bertujuan untuk memprediksi kelulusan mahasiswa tepat waktu tetapi juga untuk mengidentifikasi variabel-variabel yang paling berpengaruh terhadap kelulusan tersebut. Dengan demikian, hasil penelitian ini dapat digunakan sebagai dasar untuk pengambilan keputusan strategis oleh pihak universitas dalam upaya meningkatkan kualitas pendidikan dan layanan akademik. Faktor-faktor seperti IPK, kehadiran, keterlibatan dalam kegiatan ekstrakurikuler, dan dukungan finansial akan dianalisis untuk menentukan pengaruhnya terhadap kelulusan mahasiswa. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode prediksi yang lebih canggih di bidang pendidikan tinggi. Penggunaan algoritma machine learning seperti *Random forest* dan *K-means* dalam konteks ini merupakan langkah inovatif yang dapat membuka peluang untuk penelitian lebih lanjut dan pengembangan aplikasi yang lebih luas. Hasil penelitian ini juga diharapkan dapat diterapkan di universitas lain yang menghadapi masalah serupa, sehingga memberikan dampak positif yang lebih luas.

Penelitian ini memiliki relevansi yang tinggi dalam konteks peningkatan kualitas pendidikan tinggi. Prediksi kelulusan mahasiswa tepat waktu dengan metode *Random forest* dan pengelompokan algoritma *K-means* diharapkan dapat menjadi alat yang efektif bagi Universitas Sriwijaya dalam mengidentifikasi masalah-masalah yang menghambat kelulusan mahasiswa dan merancang strategi yang lebih baik untuk memastikan bahwa lebih banyak mahasiswa dapat lulus tepat waktu. Berdasarkan latarbelakang tersebut penelitian ini berjudul **“Prediksi**

Kelulusan Mahasiswa Tepat Waktu dengan Metode *Random forest* Berdasarkan Algoritma *K-means* pada Pascasarjana Universitas Sriwijaya”.

1.2 Perumusan Masalah

Berdasarkan latar belakang penelitian yang telah diuraikan, dapat dirumuskan beberapa permasalahan utama sebagai berikut:

1. Bagaimana tahapan melakukan klustering faktor-faktor yang mempengaruhi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya menggunakan metode *Random forest* dan algoritma *K-means*?
2. Bagaimana melakukan prediksi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya menggunakan metode *Random forest* yang telah dikelompokkan menggunakan algoritma *K-means*?

1.3 Batasan Masalah

Adapun Batasan masalah pada penelitian ini, ialah sebagai berikut :

1. Masalah ini berfokus pada kemampuan metode *Random forest* dalam menghasilkan prediksi yang akurat mengenai kelulusan mahasiswa. Tingkat akurasi yang tinggi akan menunjukkan efektivitas metode ini dalam konteks yang diteliti.
2. Mengidentifikasi faktor batasan atau *scope* seperti IPK, IPS, jumlah SKS, semester, umur, dan status pekerjaan. Yang mempengaruhi kelulusan mahasiswa dengan menggunakan algoritma *K-means*, sehingga dapat memperoleh pemahaman.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini ialah sebagai berikut :

1. Mengetahui klustering faktor-faktor yang mempengaruhi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya menggunakan metode *Random forest* dan algoritma *K-means*.
2. Memprediksi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya menggunakan metode *Random forest* dan algoritma *K-means*.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini ialah sebagai berikut :

1. Memberikan panduan langkah klustering menggunakan metode *Random forest* dan algoritma *K-means* sehingga berguna bagi akademisi dan praktisi dalam meningkatkan tingkat kelulusan tepat waktu.
2. Mengembangkan model prediksi yang lebih akurat dan efisien dalam menentukan faktor yang mempengaruhi kelulusan tepat waktu.
3. Memberikan wawasan baru mengenai faktor yang mempengaruhi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya sehingga dapat membantu pihak universitas dalam merancang kebijakan yang lebih efektif.

1.6 Sistematika Penulisan

Adapun sistematika penulisan proposal tesis ialah sebagai berikut :

BAB I Pendahuluan

Bab Pendahuluan berisi tentang gambaran secara singkat mengenai isi tesis ini, bab ini terdiri dari: Latar Belakang, Perumusan Masalah, Tujuan dan Manfaat, Batasan Masalah, dan Sistematika Penulisan

BAB II Tinjauan Pustaka

Pada bab ini dibahas mengenai pustaka-pustaka yang berkaitan dengan penelitian baik yang bersumber dari buku, artikel, jurnal, dan karya ilmiah untuk mengetahui hubungan antara penelitian yang akan dilakukan dengan studi penelitian sebelumnya.

BAB III Metodologi Penelitian

Bab ini mengandung uraian tentang alasan dan metode penelitian di laboratorium, pemilihan lokasi untuk penelitian di lapangan, dan studi literatur; metode pengambilan sampel; variabel dan data yang digunakan, metode analisis data, dan metode penyajian data.

BAB IV Hasil dan Analisa

Bab ini berisi tentang hasil pengujian yang telah dilakukan, data yang diambil, serta analisa yang dilakukan di dalam penelitian.

BAB V Kesimpulan dan Saran

Bab ini berisi tentang kesimpulan yang diperoleh dari penelitian dan saran untuk penelitian berikutnya.

DAFTAR PUSTAKA

- Alvianatinova, V., Ali, I., Rahaningsih, N., & Bahtiar, A. (2024). Penerapan Algoritma *K-means Clustering* dalam Pengelompokan Data Penjualan Supermarket Berdasarkan Cabang (Branch). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1529–1535.
- Attenberg, J., & Ertekin, Ş. (2013). Class Imbalance and Active Learning. In H. He & Y. Ma (Eds.), *Imbalanced Learning* (1st ed., pp. 101–149). Wiley. <https://doi.org/10.1002/9781118646106.ch6>
- Banjarsari, M. A., Budiman, H. I., & Farmadi, A. (2015). Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 2(2), 50–64. <http://dx.doi.org/10.20527/klik.v2i2.26>
- Blaxter, L., Hughes, C., & Tight, M. (2010). *How to research* (4th ed). McGraw-Hill/Open University Press.
- Breiman, L. (2001). *Random forest*. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., & Cutler, A. (2001). Package “*Random forest*.” *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Dawson, C. (2009). *Introduction to Research Methods: A practical guide for anyone undertaking a research project*. Little Brown Book Group.
- Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *Educational Data mining*, 41–50.
- Fithri, D. L., & Darmanto, E. (2021). Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes. *Prosiding SNATIF Ke-1*, 319–324.
- Genç, G. (2016). Attributions To Success And Failure In English Language Learning: The Effects Of Gender, Age And Perceived Success. *European Journal of Education Studies*, 2(12), 26–43. <https://doi.org/10.5281/ZENODO.192396>

- Gorunescu, F. (2011). *Data mining: Concepts, models and techniques*. Springer-Verlag.
- Gupta, A., & Gupta, G. (2019). Comparative Study of *Random forest* and Neural Network for Prediction in Direct Marketing. In H. Malik, S. Srivastava, Y. R. Sood, & A. Ahmad (Eds.), *Applications of Artificial Intelligence Techniques in Engineering* (Vol. 697, pp. 401–410). Springer Singapore. https://doi.org/10.1007/978-981-13-1822-1_37
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed). Elsevier/Morgan Kaufmann.
- Jain, A. K. (2010). Data clustering: 50 years beyond *K-means*. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jatmiko, Y. A., Padmadisastra, S., & Chadidjah, A. (2019). Analisis Perbandingan Kinerja Cart Konvensional, Bagging dan *Random forest* pada Klasifikasi Objek: Hasil dari Dua Simulasi. *MEDIA STATISTIKA*, 12(1), 1. <https://doi.org/10.14710/medstat.12.1.1-12>
- Jin, X., & Han, J. (2016). *K-means Clustering*. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data mining* (pp. 1–3). Springer US. https://doi.org/10.1007/978-1-4899-7502-7_431-1
- Kamagi, D. H., & Hansun, S. (2021). Implementasi *Data mining* dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal ULTIMATICS*, 6(1), 15–20. <https://doi.org/10.31937/ti.v6i1.327>
- Khasanah, N., Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). Prediksi Kelulusan Mahasiswa dengan Metode Naive Bayes. *Technologia : Jurnal Ilmiah*, 13(3), 207. <https://doi.org/10.31602/tji.v13i3.7312>
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (2nd ed). New Age International.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of *classification* and combining techniques. *Artificial Intelligence Review*, 26(3), 159–190. <https://doi.org/10.1007/s10462-007-9052-3>

- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: An introduction to data mining* (Second edition). Wiley.
- Liaw, A., & Wiener, M. (2002). *Classification and Regression by RandomForest*. *R News*, 2(3), 18–22.
- MacQueen, J. B. (1967). Some Methods for *classification* and Analysis of Multivariate Observations. *Proceedings of 5-Th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
- Maimon, O., & Rokach, L. (2009). Introduction to Knowledge Discovery and *Data mining*. In O. Maimon & L. Rokach (Eds.), *Data mining and Knowledge Discovery Handbook* (pp. 1–15). Springer US. https://doi.org/10.1007/978-0-387-09823-4_1
- Marlina, H., Elmayati, Zulus, A., & Wijaya, H. O. L. (2023). Penerapan Algoritma *Random forest* dalam Klasifikasi Penjurusan di SMA Negeri Tugumulyo. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 4(2), 139–143.
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., & Hasbi, I. (2021). *Data mining dan Aplikasinya*. Penerbit Widina Bhakti Persada.
- Nisbet, R., Elder, J. F., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press/Elsevier.
- Nugraha, G. S., & Hairani, H. (2018). Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode *K-means*. *Jurnal MATRIK*, 17(2), 13–23. <https://doi.org/10.30812/matrik.v17i2.84>
- Nugroho, Y. S. (2019). Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014*. Seminar Nasional Aplikasi Sains dan Teknologi IST AKPRIND 2014, Yogyakarta.
- Orsenigo, C., & Vercellis, C. (2009). Multicategory *classification* via discrete support vector machines. *Computational Management Science*, 6(1), 101–114. <https://doi.org/10.1007/s10287-008-0068-1>
- Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting University Student Graduation Using Academic Performance and Machine Learning:

- A Systematic Literature Review. *IEEE Access*, 12, 23451–23465.
<https://doi.org/10.1109/ACCESS.2024.3361479>
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan *Data mining* Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 7(1), 59–64.
- Romero, C., & Ventura, S. (2013). *Data mining* in education. *WIREs Data mining and Knowledge Discovery*, 3(1), 12–27.
<https://doi.org/10.1002/widm.1075>
- Saifudin, A. (2022). *Level Data dan Algoritma untuk Penanganan Ketidakseimbangan Kelas*. PT Mediatama Digital Cendekia.
- 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. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 276–280.
<https://doi.org/10.1109/ICTS.2019.8850948>
- Siswoyo, D. (2007). *Ilmu Pendidikan*. Universitas Negeri Yogyakarta.
- Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Sumartini, & Disman. (2018). Analisis Faktor-Faktor yang Mempengaruhi Penyelesaian Studi Tepat Waktu serta Implikasinya terhadap Kualitas Lulusan. *Indonesian Journal of Economic Education*, 1(1), 43–54.
- Syam, S., Tokoro, Y., Judijanto, L., Garonga, M., & Sinaga, F. M. (2024). *Data mining (Teori dan Penerapannya dalam Berbagai Bidang)*. PT Sonpedia Publishing Indonesia.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (Second Ed). Pearson Addison Wesley.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed). Morgan Kaufmann.
- Wu, J. (2009). A Novel Artificial Neural Network Ensemble Model Based on K-Nearest Neighbor Nonparametric Estimation of Regression Function and Its Application for Rainfall Forecasting. *2009 International Joint Conference on Computational Sciences and Optimization*, 44–48.
<https://doi.org/10.1109/CSO.2009.307>

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of *Clustering* Algorithms. *Annals of Data Science*, 2(2), 165–193. <https://doi.org/10.1007/s40745-015-0040-1>