



***SURVIVAL MACHINE LEARNING* BERBASIS POHON UNTUK PREDIKSI LAMA STUDI MAHASISWA MAGISTER DI IPB UNIVERSITY**

AHMAD SYAUQI



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA KELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “*Survival Machine Learning* Berbasis Pohon untuk Prediksi Lama Studi Mahasiswa Magister di IPB University” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Ahmad Syauqi
G1501221019



RINGKASAN

AHMAD SYAUQLI. *Survival Machine Learning* Berbasis Pohon untuk Prediksi Lama Studi Mahasiswa Magister di IPB University. Dibimbing oleh ANWAR FITRIANTO dan HARI WIJAYANTO.

Lama studi merupakan salah satu indikator penting dalam evaluasi keberhasilan penyelenggaraan pendidikan tinggi. Informasi mengenai faktor-faktor yang berkaitan dengan lama studi mahasiswa dapat dimanfaatkan untuk mendukung proses monitoring akademik serta membantu mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan penyelesaian studi. Analisis lama studi memiliki karakteristik khusus karena sebagian mahasiswa belum menyelesaikan studinya pada saat pengamatan berakhir sehingga menghasilkan data tersensor. Kondisi tersebut menyebabkan metode *survival analysis* menjadi pendekatan yang sesuai untuk digunakan. Seiring berkembangnya metode analisis data, pendekatan *survival machine learning* menawarkan kemampuan yang lebih baik dalam menangkap hubungan nonlinier dan interaksi antarpeubah dibandingkan metode *survival* konvensional.

Berdasarkan latar belakang tersebut, tujuan penelitian ini adalah membandingkan performa model *survival tree* (ST), *extremely randomized survival tree* (EST), *random survival forest* (RSF), dan *gradient boosting survival* (GBS) dalam memprediksi lama studi mahasiswa program magister IPB University; mengidentifikasi peubah-peubah yang berkontribusi terhadap prediksi lama studi berdasarkan model terbaik; serta membentuk stratifikasi risiko mahasiswa sebagai dasar monitoring akademik.

Penelitian menggunakan data mahasiswa program magister IPB University angkatan 2020–2022 yang diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru IPB University. Data terdiri atas 3.417 mahasiswa dengan waktu pengamatan maksimum 42 bulan. Peubah yang digunakan meliputi karakteristik demografis, akademik, sosial, dan ekonomi mahasiswa. Evaluasi performa model dilakukan menggunakan *Concordance Index* (C-index) dan *Integrated Brier Score* (IBS) melalui skema *random split* sebanyak sepuluh kali. Model terbaik kemudian diinterpretasikan menggunakan *feature importance* dan *shapley additive explanations* (SHAP). Selanjutnya, dilakukan pembentukan stratifikasi risiko berdasarkan peluang keterlambatan studi pada bulan ke-24 yang diprediksi oleh model terbaik.

Hasil analisis deskriptif menunjukkan bahwa median lama studi mahasiswa program magister IPB University adalah 29 bulan. Analisis Kaplan-Meier menunjukkan adanya perbedaan pola lama studi berdasarkan beberapa karakteristik mahasiswa, terutama fakultas, angkatan, jenis kelamin, asal perguruan tinggi, status perkawinan, dan wilayah asal. Selain itu, terlihat adanya pola peningkatan peluang kelulusan yang cukup tajam pada beberapa periode yang berdekatan dengan akhir semester akademik, yang menunjukkan bahwa penyelesaian studi cenderung terkonsentrasi pada periode-periode tertentu.

Hasil perbandingan model menunjukkan bahwa seluruh model mampu digunakan untuk memprediksi lama studi mahasiswa. Namun demikian, model GBS memberikan performa terbaik dengan nilai rata-rata C-index sebesar 0,659 dan IBS sebesar 0,194. Hasil tersebut menunjukkan bahwa GBS memiliki



kemampuan diskriminasi dan akurasi prediksi yang lebih baik dibandingkan model lainnya pada data penelitian ini.

Interpretasi model menggunakan *feature importance* dan SHAP menunjukkan bahwa usia saat masuk dan IPK S1 merupakan faktor yang paling dominan dalam prediksi lama studi mahasiswa. Selain itu, fakultas, angkatan, serta interaksi antara status perkawinan dan IPK S1 juga memberikan kontribusi penting terhadap prediksi model. Analisis SHAP menunjukkan bahwa pada output model untuk horizon 24 bulan, mahasiswa yang memasuki program magister pada usia yang lebih muda dan memiliki IPK S1 yang lebih tinggi cenderung memiliki peluang yang lebih besar untuk menyelesaikan studi dalam waktu ideal. Hasil SHAP *dependence plot* juga menunjukkan bahwa hubungan usia dan IPK S1 terhadap peluang penyelesaian studi bersifat nonlinier.

Berdasarkan model GBS, mahasiswa berhasil dikelompokkan ke dalam tiga kelompok risiko, yaitu *low risk*, *medium risk*, dan *high risk*. Ketiga kelompok tersebut memiliki pola penyelesaian studi yang berbeda secara signifikan. Kelompok *low risk* memiliki rata-rata usia 22,86 tahun, rata-rata IPK S1 3,62, dan persentase kelulusan sebesar 78,5%, sedangkan kelompok *high risk* memiliki rata-rata usia 29,01 tahun, rata-rata IPK S1 3,12, dan persentase kelulusan sebesar 44,9%. Hasil tersebut menunjukkan bahwa model mampu membedakan mahasiswa berdasarkan tingkat risiko keterlambatan studi dan menghasilkan kelompok-kelompok yang memiliki karakteristik yang berbeda secara konsisten.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan *survival machine learning* berbasis pohon, khususnya GBS, mampu memberikan performa prediksi yang baik untuk data lama studi mahasiswa magister. Selain menghasilkan prediksi yang lebih baik dibanding model lain, model juga mampu mengidentifikasi faktor-faktor penting yang berkaitan dengan lama studi serta membentuk stratifikasi risiko mahasiswa yang berpotensi dimanfaatkan sebagai dasar pengembangan monitoring akademik berbasis data.

Kata kunci: lama studi, mahasiswa magister, stratifikasi risiko, *survival machine learning*



SUMMARY

AHMAD SYAUQI. Tree-Based Survival Machine Learning for Predicting Study Duration of Master's Students at IPB University. Supervised by ANWAR FITRIANTO and HARI WIJAYANTO.

Study duration is an important indicator for evaluating the effectiveness of higher education programs. Information regarding factors associated with students' study duration can support academic monitoring and help identify students who are at risk of delayed graduation. Study duration data possess unique characteristics because some students have not completed their studies when the observation period ends, resulting in censored observations. Consequently, survival analysis provides an appropriate analytical framework for studying such data. Along with recent developments in data analytics, survival machine learning methods offer greater flexibility in capturing nonlinear relationships and complex interactions among variables compared with conventional survival models.

Based on this background, the objectives of this study were to compare the performance of survival tree (ST), extremely randomized survival tree (EST), random survival forest (RSF), and gradient boosting survival (GBS) models in predicting the study duration of master's students at IPB University; to identify factors influencing study duration prediction based on the best-performing model; and to develop student risk stratification as a basis for academic monitoring.

This study utilized data from master's students enrolled at IPB University between 2020 and 2022. The dataset consisted of 3,417 students with a maximum observation period of 42 months. The variables included demographic, academic, social, and economic characteristics of students. Model performance was evaluated using the Concordance Index (C-index) and Integrated Brier Score (IBS) through ten random train-test splits. The best-performing model was subsequently interpreted using feature importance measures and shapley additive explanations (SHAP). Risk stratification was then conducted based on the predicted probability of delayed graduation beyond 24 months.

Descriptive analysis showed that the median study duration of master's students at IPB University was 29 months. Kaplan-Meier analysis revealed differences in study duration patterns across several student characteristics, particularly faculty, admission cohort, gender, undergraduate institution type, marital status, and region of origin. In addition, substantial inclines in probability of graduated were observed near the end of several academic semesters, indicating that study completion tended to cluster around specific periods.

The model comparison results demonstrated that all four survival machine learning models were capable of predicting study duration. However, the GBS model achieved the best overall performance, with an average C-index of 0.659 and an average IBS of 0.194. These results indicate that GBS provided superior discrimination ability and prediction accuracy compared with the other models evaluated in this study.

Model interpretation using feature importance and SHAP revealed that age at admission and undergraduate GPA were the most influential factors in predicting study duration. Faculty, admission cohort, and the interaction between marital status and GPA also contributed substantially to model predictions. SHAP



analysis indicated that students entering the master's program at a younger age and with higher undergraduate GPAs were more likely to complete their studies within the ideal study period. Furthermore, SHAP dependence plots showed that the relationships between age, GPA, and study completion probability were nonlinear.

Based on the GBS model, students were successfully classified into three risk groups: low risk, medium risk, and high risk. These groups exhibited significantly different study completion patterns. The low risk group had an average age of 22.86 years, an average GPA of 3.62, and a graduation rate of 78.5%, whereas the high risk group had an average age of 29.01 years, an average GPA of 3.12, and a graduation rate of 44.9%. These findings demonstrate that the model was able to distinguish students according to their risk of delayed graduation and to produce risk groups with consistently different characteristics.

Overall, this study demonstrates that tree-based survival machine learning, particularly GBS, provides an effective approach for predicting the study duration of master's students. Beyond prediction, the model was able to identify important factors associated with study duration and generate student risk stratification that may support the development of data-driven academic monitoring systems.

Keywords: master's students, risk stratification, study duration, survival machine learning



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



SURVIVAL MACHINE LEARNING BERBASIS POHON UNTUK PREDIKSI LAMA STUDI MAHASISWA MAGISTER DI IPB UNIVERSITY

AHMAD SYAUQI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Sains pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



Tim Penguji pada Ujian Tesis:
Dr. Ir. Indahwati, M.Si.

@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

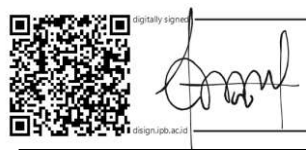
Perpustakaan IPB University

Judul Tesis : *Survival Machine Learning* Berbasis Pohon untuk
Prediksi Lama Studi Mahasiswa Magister di IPB
University

Nama : Ahmad Syauqi
NIM : G1501221019

Disetujui oleh

Pembimbing 1:
Dr. Anwar Fitrianto, S.Si., M.Sc.



Pembimbing 2:
Prof. Dr. Ir. Hari Wijayanto, M.Si.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP. 197503151999031004



Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP. 196607021993021001



Tanggal Ujian :
14 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanallahu wa ta'ala atas segala karunia-Nya sehingga tesis dengan judul “*Survival Machine Learning* Berbasis Pohon untuk Prediksi Lama Studi Mahasiswa Magister di IPB University” berhasil diselesaikan. Tesis ini penulis susun sebagai salah satu syarat untuk menyelesaikan studi Magister Sains pada Program Studi Statistika dan Sains Data. Penulisan tesis ini dapat terselesaikan tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak sehingga penulis menyampaikan ucapan terima kasih kepada:

1. Dr. Anwar Fitrianto, S.Si., M.Sc. dan Prof. Dr. Ir. Hari Wijayanto, M.Si. selaku komisi pembimbing yang senantiasa bersedia meluangkan waktunya untuk memberikan bimbingan, masukan, dukungan, dan arahan kepada penulis;
2. Ibu, Bapak, Istri, dan Kakak penulis yang senantiasa mendoakan, mendukung, dan memotivasi penulis;
3. Seluruh dosen pengajar Program Studi Statistika dan Sains Data atas ilmu serta motivasi yang telah diberikan selama penulis berkuliah;
4. Seluruh staf Program Studi Statistika dan Sains Data atas segala bantuan dalam keperluan administrasi selama penulis berkuliah;
5. Rekan-rekan mahasiswa Pascasarjana Program Studi Statistika dan Sains Data atas dukungan dan kerja sama yang baik.

Semoga penelitian ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Ahmad Syauqi



DAFTAR ISI

DAFTAR TABEL	xv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xv
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
II TINJAUAN PUSTAKA	4
2.1 Analisis Survival dan Lama Studi Mahasiswa	4
2.2 Analisis Survival	6
2.3 Evaluasi, Interpretabilitas, dan Stratifikasi Risiko	9
III METODE	13
3.1 Sumber dan Data Penelitian	13
3.2 Peubah Penelitian	13
3.3 Tahapan Analisis	14
IV HASIL DAN PEMBAHASAN	17
4.1 Data Empiris	17
4.2 Evaluasi Model dan Prediksi Lama Studi	24
4.3 Interpretasi Model GBS	27
4.4 Stratifikasi Risiko Mahasiswa Berdasarkan Model GBS	33
4.5 Ilustrasi Prediksi Individual Menggunakan Model GBS	38
4.6 Implikasi Akademik dan Manajerial Hasil Penelitian	40
V SIMPULAN DAN SARAN	41
5.1 Simpulan	41
5.2 Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN	45
RIWAYAT HIDUP	47

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Peubah penelitian	13
2	Karakteristik mahasiswa magister IPB angkatan 2020-2022	17
3	Hasil uji log-rank berdasarkan karakteristik mahasiswa	21
4	Performa keempat model berdasarkan C-index dan IBS	25
5	Jumlah mahasiswa pada setiap kelompok risiko	34
6	Persentase mahasiswa yang telah menyelesaikan studi sebelum 42 bulan pada setiap kelompok risiko	35
7	Rata-rata umur dan IPK S1 berdasarkan kelompok risiko	35
8	Karakteristik dan hasil prediksi tiga mahasiswa ilustratif menggunakan model GBS	38

DAFTAR GAMBAR

1	Diagram alir penelitian	14
2	Distribusi lama studi mahasiswa magister IPB angkatan 2020-2022	18
3	Sebaran mahasiswa magister IPB angkatan 2020-2022 per fakultas	18
4	Median lama studi mahasiswa magister berdasarkan fakultas	19
5	Distribusi umur mahasiswa magister IPB angkatan 2020-2022	19
6	Kurva peluang kelulusan Kaplan-Meier mahasiswa magister IPB angkatan 2020-2022	20
7	Kurva peluang kelulusan Kaplan-Meier per fakultas	22
8	Kurva peluang kelulusan Kaplan-Meier per angkatan	23
9	Kurva peluang kelulusan Kaplan-Meier per asal perguruan tinggi	23
10	Kurva peluang kelulusan Kaplan-Meier berdasarkan (a) wilayah asal, (b) jenis kelamin, (c) status perkawinan, dan (d) akreditasi asal perguruan tinggi	24
11	Scatter plot hasil evaluasi model berdasarkan IBS dan C-index	26
12	Feature importance model GBS	27
13	Nilai SHAP terhadap peluang kelulusan mahasiswa dalam waktu ideal	29
14	Persebaran nilai SHAP pada setiap umur	31
15	Persebaran nilai SHAP pada IPK S1	32
16	Kurva peluang kelulusan Kaplan-Meier pada setiap kelompok risiko	34
17	Proporsi kelompok risiko mahasiswa berdasarkan (a) angkatan, (b) fakultas, (c) asal perguruan tinggi sarjana, dan (d) status perkawinan	36
18	Proporsi kelompok risiko mahasiswa berdasarkan (a) asal wilayah, (b) akreditasi Perguruan Tinggi asal, (c) jenis kelamin	37
19	Ilustrasi prediksi individual menggunakan model GBS	39

DAFTAR LAMPIRAN

1	Hyperparameter yang digunakan pada setiap model	45
2	Hasil evaluasi model berdasar C-index dan IBS pada setiap ulangan	46