



# PERBANDINGAN METODE *RANDOM SURVIVAL FOREST* DAN *XGBOOST-AFT* DALAM MEMPREDIKSI KETAHANAN LAGU PADA SPOTIFY *WEEKLY CHART* INDONESIA

**SALSA RIFDA ERIRA**



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



## PERNYATAAN MENGENAI SKRIPSI, SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Metode *Random Survival Forest* dan *XGBoost-AFT* dalam Memprediksi Ketahanan Lagu pada *Spotify Weekly Chart* Indonesia.” 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 skripsi ini.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan ChatGPT dan Claude untuk mencari ide topik penelitian, mempercepat penelusuran referensi atau jurnal yang relevan, serta membantu dalam memperbaiki kesalahan sintaks pemrograman. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi karya tugas akhir ini.

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

Bogor, Juli 2026

Salsa Rifda Erira  
G1401221094



## ABSTRAK

SALSA RIFDA ERIRA. Perbandingan Metode *Random Survival Forest* dan XGBoost-AFT dalam Memprediksi Ketahanan Lagu pada *Spotify Weekly Chart* Indonesia. Dibimbing oleh INDAHWATI dan SACHNAZ DESTA OKTARINA

Tangga musik digital merupakan indikator penting dalam mengukur kesuksesan lagu di era streaming. Penelitian ini bertujuan untuk membandingkan algoritma *Random Survival Forest* (RSF) dan *Extreme Gradient Boosting–Accelerated Failure Time* (XGBoost-AFT) dalam memprediksi ketahanan lagu pada *Spotify Weekly Chart* Indonesia tahun 2020-2024 dengan 1193 amatan. Kejadian didefinisikan sebagai lagu keluar dari tangga musik, sementara lagu yang masih bertahan hingga akhir penelitian diperlakukan sebagai data tersensor. Validasi model dilakukan menggunakan *repeated stratified 5-fold cross validation* dengan metrik *concordance index* (C-index), *C-index Inverse Probability of Censoring Weighting* (C-Index IPCW), dan *mean time-dependent Area Under Curve* (*mean time-dependent AUC*). Hasil penelitian menunjukkan bahwa model XGBoost-AFT memiliki performa yang sedikit lebih baik dibandingkan RSF dengan nilai C-index, C-index IPCW, dan *mean time-dependent AUC* masing-masing sebesar 0,6440; 0,6435; dan 0,6932. Analisis peubah penting menggunakan *feature importance* dan SHAP menunjukkan bahwa genre dan bahasa lagu merupakan peubah paling penting, diikuti oleh *internal competition*, *danceability*, keberadaan music video (MV), *speechiness*, dan *valence*. Selain itu, karakteristik yang diasosiasikan dengan ketahanan lebih panjang antara lain nilai *internal competition* yang rendah, *acousticness* tinggi, *danceability* moderat, adanya MV, serta *speechiness* dan *valence* yang relatif rendah, dengan durasi lagu sekitar tiga menit.

Kata kunci: daya tahan, ketahanan lagu, *random survival forest*, spotify, xgboost-aft



## ABSTRACT

SALSA RIFDA ERIRA. Comparison of Random Survival Forest and XGBoost-AFT Methods in Predicting Song Longevity on The Spotify Weekly Chart Indonesia. Supervised by INDAHWATI and SACHNAZ DESTA OKTARINA

Digital music charts are an important indicator of song success in the streaming era. This study compares Random Survival Forest (RSF) and Extreme Gradient Boosting–Accelerated Failure Time (XGBoost-AFT) algorithms in predicting song longevity on the Spotify Weekly Chart Indonesia from 2020 to 2024, using 1,193 observations. The event is defined as a song exiting the chart, while songs that remain until the end of the observation period are treated as censored data. Model validation uses repeated stratified 5-fold cross-validation with evaluation metrics including the concordance index (C-index), C-index Inverse Probability of Censoring Weighting (C-index IPCW), and mean time-dependent Area Under the Curve (mean time-dependent AUC). Results show that XGBoost-AFT performs slightly better than RSF, with C-index, C-index IPCW, and mean time-dependent AUC values of 0.6440, 0.6435, and 0.6932, respectively. Feature importance and SHAP analysis show that genre and song language are the most influential variables, followed by internal competition, danceability, music video (MV) availability, speechiness, and valence. Songs with lower internal competition, higher acousticness, moderate danceability, MV availability, and lower speechiness and valence, with durations around three minutes, tend to have longer chart longevity.

**Keywords:** random survival forest, spotify, survival, song longevity, xgboost-aft



@Hak cipta milik IPB University

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.

© 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.*

# **PERBANDINGAN METODE *RANDOM SURVIVAL FOREST* DAN XGBOOST-AFT DALAM MEMPREDIKSI KETAHANAN LAGU PADA *SPOTIFY WEEKLY CHART* INDONESIA**

**SALSA RIFDA ERIRA**

Skripsi  
sebagai salah satu syarat untuk memperoleh gelar  
Sarjana pada  
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA 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.

Penguji pada Ujian Skripsi:

1 Aulia Rizki Firdawanti, S.Stat, M.Si





Judul Skripsi : Perbandingan Metode *Random Survival Forest* dan XGBoost-AFT dalam Memprediksi Ketahanan Lagu pada Spotify *Weekly Chart* Indonesia

Nama : Salsa Rifda Erira  
NIM : G1401221094

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Ir. Indahwati, M. Si.

---

Pembimbing 2:

Sachnaz Desta Oktarina, M. Agr, Ph.D

---

Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si, M.Si

NIP 197804112005011002

---

Tanggal Ujian:  
7 Juli 2026

Tanggal Lulus:





## PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanaahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Desember 2025 sampai bulan Juni 2026 ini ialah pemodelan ketahanan lagu pada Tangga Musik, dengan judul “Perbandingan Metode *Random Survival Forest* dan *XGBoost-AFT* dalam Memprediksi Ketahanan Lagu pada *Spotify Weekly Chart* Indonesia”. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah berperan sejak penulis menempuh studi sampai proses penulisan karya ilmiah ini berakhir, di antaranya:

1. Ibu Dr. Indahwati, M.Si. dan Ibu Sachnaz Desta Oktarina, M.Agr., Ph.D. selaku dosen pembimbing skripsi yang dengan penuh kesabaran telah memberikan arahan, masukan, serta ilmu yang berharga dalam setiap langkah penulis menyelesaikan karya ilmiah ini.
2. Ibu Rahmawati dan Bapak Erisman (Alm), orang tua penulis, yang senantiasa menjadi sumber doa, inspirasi, serta kekuatan, dan telah dengan tulus mendukung serta memfasilitasi perjalanan pendidikan penulis hingga mampu mencapai tahap perguruan tinggi dan menyelesaikan karya ilmiah ini.
3. Saudari Desy Qonita Erira dan Saudara Hanif Irfan, selaku kakak kandung dan kakak ipar penulis, yang tidak hanya menjadi keluarga, tetapi juga tempat pulang kedua di perantauan, serta senantiasa memberikan dukungan, kehangatan, dan semangat bagi penulis.
4. Ibu Aulia Rizki Firdawanti, S.Stat, M.Si selaku penguji skripsi yang memberikan banyak saran dan masukan untuk membuat karya tulis ini semakin sempurna
5. Bapak Dr. Bagus Sartono, S.Si, M.Si dan Ir. Aam Alamudi, M.Si sebagai moderator pada tahapan kolokium dan seminar hasil penulis atas saran dan perbaikan yang diberikan terhadap penelitian penulis.
6. Seluruh dosen Statistika yang telah dengan tulus berbagi ilmu, pengalaman, dan bimbingan selama empat tahun terakhir. Setiap pengetahuan yang diberikan menjadi bekal berharga bagi penulis dalam menempuh perjalanan akademik hingga menyelesaikan karya ilmiah ini.
7. Delia Fitri Audina, Salsabila Fayiza, Raihanna Asma Amani, dan Sandra Berliana Putri sebagai teman terdekat penulis selama menempuh jenjang perguruan tinggi, yang senantiasa hadir dalam suka dan duka, memberikan dukungan, tawa, serta menjadi bagian penting dalam perjalanan dan kenangan berharga penulis.
8. Sembilan member ZB1 yang secara tak langsung menjadi inspirasi penulis dalam memilih topik penelitian ini

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan

Bogor, Juli 2026

Salsa Rifda Erira



## DAFTAR ISI

|                                                        |    |
|--------------------------------------------------------|----|
| DAFTAR TABEL                                           | x  |
| DAFTAR GAMBAR                                          | x  |
| DAFTAR LAMPIRAN                                        | x  |
| PENDAHULUAN                                            | 1  |
| 1.1 Latar Belakang                                     | 1  |
| 1.2 Rumusan Masalah                                    | 2  |
| 1.3 Tujuan                                             | 3  |
| TINJAUAN PUSTAKA                                       | 4  |
| 2.1 Tangga Musik <i>Spotify Weekly Chart</i> Indonesia | 4  |
| 2.2 Analisis Daya Tahan                                | 4  |
| 2.3 <i>Kaplan-Meier</i>                                | 5  |
| 2.4 <i>Accelareted Failure Time</i> (AFT)              | 6  |
| 2.5 <i>Random Survival Forest</i> (RSF)                | 7  |
| 2.6 <i>eXtreme Gradient Boosting</i> (XGBoost) - AFT   | 8  |
| 2.7 <i>Repeated Stratified 5-fold Cross-Validation</i> | 9  |
| 2.8 Evaluasi Model Survival                            | 9  |
| 2.9 <i>Feature Importance</i>                          | 10 |
| 2.10 SHAP (Shapley Additive exPlanation)               | 10 |
| METODE                                                 | 12 |
| 3.1 Bahan dan Data                                     | 12 |
| 3.2 Prosedur Analisis Data                             | 14 |
| HASIL DAN PEMBAHASAN                                   | 16 |
| 4.1 Praproses Data                                     | 16 |
| 4.2 Eksplorasi Data                                    | 16 |
| 4.3 Pemodelan AFT                                      | 20 |
| 4.4 Pemodelan <i>Random Survival Forest</i>            | 22 |
| 4.5 Pemodelan XGBoost-AFT                              | 22 |
| 4.6 Perbandingan dan Evaluasi Model                    | 23 |
| 4.7 Performa Model dalam Prediksi                      | 24 |
| 4.8 Performa model dalam prediksi                      | 28 |
| SIMPULAN DAN SARAN                                     | 31 |
| 5.1 Simpulan                                           | 31 |

|     |                |    |
|-----|----------------|----|
| 5.2 | Saran          | 31 |
|     | DAFTAR PUSTAKA | 32 |
|     | LAMPIRAN       | 35 |
|     | RIWAYAT HIDUP  | 37 |





## DAFTAR TABEL

|   |                                                                                          |    |
|---|------------------------------------------------------------------------------------------|----|
| 1 | Peubah Penelitian                                                                        | 13 |
| 2 | <i>Hyperparameter Random Survival Forest</i> yang digunakan                              | 14 |
| 3 | <i>Hyperparameter XGBoost-AFT</i> yang digunakan                                         | 15 |
| 4 | Nilai C-Index, AIC, dan <i>Kolmogorov-Smirnov</i> (KS) Eksponensial (1) setiap model AFT | 21 |
| 5 | Kombinasi <i>hyperparameter</i> RSF terbaik                                              | 22 |
| 6 | Kombinasi <i>hyperparameter</i> XGBoost-AFT terbaik                                      | 23 |
| 7 | Perbandingan performa model RSF dan XGBoost-AFT                                          | 24 |
| 8 | <i>Wilcoxon test</i> pada validasi RSF dan XGBoost-AFT                                   | 24 |
| 9 | Perbandingan prediksi dan nilai aktual (minggu) XGBoost-AFT                              | 29 |

## DAFTAR GAMBAR

|    |                                                                                                                                                                            |    |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1  | Bagan alir penelitian                                                                                                                                                      | 15 |
| 2  | <i>Pie chart</i> proporsi banyak kejadian lagu keluar dari Spotify <i>Weekly Chart</i> Indonesia tahun 2020-2024                                                           | 17 |
| 3  | (a) <i>Kaplan-Meier survival_week</i> , (b) histogram <i>survival_week</i>                                                                                                 | 17 |
| 4  | (a) <i>Bar chart</i> lagu masuk tangga musik per tahun, (b) <i>bar chart</i> jumlah artis unik per tahun                                                                   | 18 |
| 5  | (a) <i>Line chart</i> tren bahasa lagu per tahun di tangga musik, (b) Kaplan-Meier kategori bahasa                                                                         | 18 |
| 6  | (a) <i>Bar chart</i> kategori genre (b) <i>Kaplan-Meier</i> kategori genre                                                                                                 | 19 |
| 7  | <i>Pie Chart</i> proporsi (a) <i>artist_popularity</i> , (c) <i>company</i> (e) MV dan Kaplan-Maier Katagori untuk b) <i>artist_popularity</i> , d) <i>Company</i> , f) MV | 20 |
| 8  | <i>Cox-Snell Residual</i> model AFT (a) Weibull, (b) Log-Normal, dan (c) Log-Logistik                                                                                      | 20 |
| 9  | Boxplot 50 nilai metrik validasi RSF vs XGBoost-AFT a) <i>C-Index</i> , b) <i>C-Index IPCW</i> , c) <i>Mean Time-Dependent AUC</i>                                         | 24 |
| 10 | <i>Feature Importance</i> model XGBoost-AFT terbaik                                                                                                                        | 25 |
| 11 | SHAP model XGBoost-AFT terbaik                                                                                                                                             | 26 |

## DAFTAR LAMPIRAN

|   |                                                                                                              |    |
|---|--------------------------------------------------------------------------------------------------------------|----|
| 1 | Perhitungan Jumlah Kombinasi Random Search Optimal dengan Tingkat Kepercayaan 95% (Bergstra dan Bengio 2012) | 36 |
| 2 | Ilustrasi Pohon XGBoost-AFT pada pohon ke-1, 50, dan 100                                                     | 36 |