Analysis of the Most Dominant Causing Factors of Divorce in 34 Provinces in Indonesia Using the XGBoost Algorithm

Seila Amalia[1], Riski Melanton Banjarnahor[2], Risca Octaviyani Hutapea[3], Gabriel Fernando Sitorus[4], Gracia Domini[5], Arnita[6]
[1,2,3,4,5,6]Program Studi Statistika, Univesitas Negeri Medan, Indonesia
Email: seilaamaliaa21@gmail.com; riskymelanton@gmail.com; riscaoctaviyanihutapea@gmail.com; gabrielfernandositorus@gmail.com; graciadms2604@gmail.com; arnita@unimed.ac.id

ABSTRAK

Perceraian di Indonesia menunjukkan tren peningkatan yang signifikan dan memiliki dampak sosial yang luas. Penelitian ini bertujuan untuk mengidentifikasi faktor penyebab perceraian paling dominan di 34 provinsi di Indonesia menggunakan algoritma machine learning Extreme Gradient Boosting (XGBoost). Data kuantitatif sekunder dari Badan Pusat Statistik tahun 2024 dianalisis dengan proses pra-pemrosesan, pembagian data, pelatihan model, dan evaluasi performa. Hasil penelitian menunjukkan bahwa perselisihan dan pertengkaran terus-menerus merupakan faktor utama penyebab perceraian, diikuti oleh penyalahgunaan zat (madat) dan kawin paksa. Model XGBoost yang dikembangkan mencapai akurasi 75% dalam mengklasifikasikan tingkat risiko perceraian. Temuan ini memberikan wawasan baru dalam memahami faktor sosial yang mempengaruhi perceraian dan dapat menjadi dasar dalam merancang strategi pencegahan yang lebih efektif.

Keyword: Faktor Penyebab Perceraian; XGBoost; Machine Learning; Risiko Perceraian; Analisis Data Sosial

ABSTRACT

Divorce in Indonesia shows a significant increasing trend and has a broad social impact. This study aims to identify the most dominant causes of divorce in 34 provinces in Indonesia using the Extreme Gradient Boosting (XGBoost) machine learning algorithm. Secondary quantitative data from the Central Bureau of Statistics in 2024 were analyzed by pre-processing, data sharing, model training, and performance evaluation. The results showed that constant disputes and quarrels were the main causes of divorce, followed by substance abuse and forced marriage. The developed XGBoost model achieved 75% accuracy in classifying the level of divorce risk. These findings provide new insights into understanding the social factors that influence divorce and can be the basis for designing more effective prevention strategies.

Keyword: Divorce Causes; XGBoost; Machine Learning; Divorce Risk; Social Data Analysis

Corresponding Author:
Seila Amalia,
Univesitas Negeri Medan,
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan,
Kabupaten Deli Serdang, Sumatera Utara 20221, Indonesia
Email: seilaamaliaa21@gmail.com

1. INTRODUCTION

Perceraian merupakan salah satu fenomena sosial yang semakin meningkat di Indonesia dalam beberapa dekade terakhir. Data dari Badan Pusat Statistik (BPS) dan Badan Peradilan Agama Mahkamah Agung menunjukkan peningkatan signifikan jumlah kasus perceraian, yang mencapai 588.266 kasus pada tahun 2018 (Syafriani Manna et al., 2021). Fenomena ini menimbulkan berbagai dampak negatif, tidak hanya bagi pasangan suami istri, tetapi juga terutama bagi anak-anak yang terdampak broken home, yang dapat memperlihatkan perilaku menyimpang, antisocial, hingga tindakan kriminal di masa depan (Hidayati, 2021). Oleh karena itu, penting untuk mengidentifikasi faktor-faktor penyebab utama perceraian guna mendukung upaya pencegahan dan intervensi yang lebih efektif.

Perceraian (*divorce*) secara umum merupakan perpisahan resmi antara pasangan suami-istri yang berketetapan untuk tidak menjalankan tugas dan kewajiban sebagai suami istri dan tidak lagi hidup serumah bersama (Olson & DeFrain, 2003). Perceraian sering kali terjadi akibat perbedaan prinsip yang tidak dapat dipersatukan lagi, di mana masing-masing pasangan mempertahankan keinginan sendiri tanpa berusaha mengalah demi keutuhan keluarga (Dariyo, 2004). Ketidakmampuan mengakui kekurangan diri dan pasangan dapat menyebabkan masalah kecil membesar dan akhirnya berujung pada perceraian (Syafriani Manna et al., 2021). Data menunjukkan bahwa kasus perceraian paling banyak terjadi pada usia pernikahan muda, yaitu antara 2 hingga 10 tahun (Syafriani Manna et al., 2021).

Berbagai penelitian di Indonesia telah mencoba mengungkap penyebab perceraian, yang meliputi konflik interpersonal, masalah ekonomi, serta faktor sosial dan budaya (Putri dan Sari, 2020; Rahman et al., 2019). Namun, mayoritas studi tersebut masih menggunakan metode statistik konvensional yang kurang mampu mengakomodasi kompleksitas dan interaksi antar variabel secara mendalam. Hal ini menimbulkan kebutuhan untuk pendekatan analisis yang lebih maju dan adaptif.

Seiring kemajuan teknologi dan ketersediaan data besar, machine learning menjadi alternatif yang efektif dalam memecahkan masalah analisis data kompleks. Machine learning bertujuan memanfaatkan data masa lalu untuk membuat keputusan di masa depan dengan metode klasifikasi, regresi, dan clustering (Roihan et al., 2019; Jo, 2021). Salah satu algoritma machine learning yang populer adalah Extreme Gradient Boosting (XGBoost), sebuah algoritma ensemble learning yang menggabungkan banyak pohon keputusan secara berurutan dengan fokus pada contoh yang salah klasifikasi sebelumnya (Chen & Guestrin, 2016; Dangeti, 2017).

XGBoost dikenal memiliki keunggulan dalam akurasi tinggi, efisiensi dalam menangani dataset besar dan kompleks, serta kemampuannya memberikan interpretasi feature importance yang berguna untuk mengidentifikasi faktor dominan dalam analisis sosial (Lundberg et al., 2017; Dava Maulana et al., 2023). Algoritma ini telah diakui secara luas, termasuk dalam kompetisi machine learning Kaggle, dengan banyak solusi pemenang yang menggunakan XGBoost (W. Zhang et al., 2021; D. Zhang et al., 2018).

Penerapan XGBoost dalam konteks analisis faktor penyebab perceraian di Indonesia masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan melakukan analisis komparatif faktor penyebab perceraian paling dominan di 34 provinsi di Indonesia menggunakan algoritma XGBoost. Pemilihan 34 provinsi sebagai unit analisis penting karena perbedaan kondisi sosial, budaya, dan ekonomi antar provinsi dapat mempengaruhi pola perceraian secara berbeda. Dengan demikian, hasil penelitian diharapkan memberikan gambaran yang lebih mendalam dan kontekstual untuk tiap wilayah, serta menjadi referensi bagi pembuat kebijakan dalam merancang strategi pencegahan perceraian yang lebih terfokus dan berbasis data.

2. RESEARCH METHOD

A. Sumber Data

Penelitian ini menggunakan data kuantitatif sekunder yang diperoleh dari dataset resmi terkait jumlah perceraian di 34 provinsi Indonesia tahun 2024, dikategorikan menurut faktor penyebab perceraian. Sumber data berasal dari instansi pemerintah yaitu BPS (Badan Pusat Statistik) https://bps.go.id/

B. Variabel Penelitian

Tabel 1. Variabel Penelitian

Variabel Dependen (Y)	Variabel Independen (X)
Jumlah Kasus Perceraian (Per Provinsi)	Ekonomi
	Perselisihan
	Kekerasan Dalam Rumah Tangga (KDRT)
	Campur Tangan Pihak Ketiga
	Masalah Moral
	Tidak Bertanggung Jawab
	Dan Lain-Lain

C. Teknik Analisis

Teknik analisis dalam penelitian ini menggunakan pendekatan algoritma *Extreme Gradient Boosting* (XGBoost), sebuah metode pembelajaran mesin berbasis pohon keputusan yang dikenal memiliki performa tinggi dalam menangani data tabular.

Langkah Analisis :

1) Pra-Pemrosesan Data

Data dibaca menggunakan pandas, dicek untuk missing value, outlier, dan dikonversi ke bentuk numerik jika perlu.

2) Pembagian Data

Analisis dilanjutkan dengan membagi dataset menjadi dua bagian, yaitu data latih dan data uji, misalnya dengan proporsi 80:20, untuk memastikan model dapat dievaluasi secara objektif.

3) Pelatihan Model

Model XGBoost kemudian dilatih dengan menggunakan variabel-variabel penyebab perceraian sebagai fitur input, sementara jumlah kasus perceraian pada masing-masing provinsi dijadikan sebagai variabel target.

4) Evaluasi Model

Setelah proses pelatihan selesai, kinerja model dievaluasi menggunakan metrik seperti Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2) guna mengukur akurasi prediksi.

5) Interpretasi Hasil

Langkah selanjutnya adalah mengidentifikasi faktor penyebab perceraian yang paling berpengaruh melalui analisis feature importance dari model XGBoost. Nilai kepentingan (importance score) ini memberikan gambaran sejauh mana setiap variabel input berkontribusi dalam memprediksi jumlah perceraian. Hasil analisis ini selanjutnya digunakan untuk menentukan faktor penyebab perceraian yang paling dominan secara nasional, serta dapat dikembangkan lebih lanjut untuk melihat pola-pola spesifik per provinsi.

3. RESULTS AND DISCUSSION

A. Deskripsi Data

Tabel 2. Deskripsi Data

Faktor Penyebab	Count	Mean	Std Dev	Min	25%	Median	75%	Max
Zina	38	31.34	111.54	1	2	6	31	695
Mabuk	38	60.76	87.57	3	15	39.5	64.75	517
Madat	38	18.11	13.04	1	9.5	18	20.5	75
Judi	38	84.98	150.80	1	15.5	33	85	819
MeninggalkanSalah satu Pihak	38	919.61	1380.62	76	290.75	490.5	920	7019
Dihukum Penjara	38	39.24	42.21	3	10.25	25	45	177
Poligami	38	27.32	35.77	1	6.25	22.5	27.75	180
Kekerasan Dalam Rumah Tangga	38	213.03	324.31	11	57.5	122.5	241	1985
Cacat Badan	38	8.76	10.46	1	2.25	6.5	9	52
Perselisihan dan Pertengkaran Terus Menerus	38	7386.03	10679.39	276	1535.75	4060	7386	51122
Kawin Paksa	38	14.76	28.76	1	3	14.5	15	181
Murtad	38	30.26	37.57	1	7	18.5	30	164
Ekonomi	38	2947.00	8026.93	1	64.5	230.5	1605.25	33264
Total	38	1178118	20279.4	545	2174	5428.5	11781	88985

Data yang dianalisis mencakup 38 provinsi dengan 13 faktor penyebab perceraian. Dari hasil statistik deskriptif, terlihat adanya variasi yang signifikan antar provinsi pada setiap faktor. Faktor Perselisihan dan Pertengkaran Terus Menerus memiliki nilai rata-rata tertinggi, menunjukkan prevalensi yang cukup besar secara nasional. Sementara itu, faktor seperti Cacat Badan menunjukkan nilai rata-rata terendah.

Sebagian faktor memiliki nilai maksimum yang jauh lebih tinggi dibandingkan nilai median dan kuartil atas, mengindikasikan adanya beberapa provinsi dengan jumlah kasus sangat tinggi, yang menjadi outlier dalam data. Variasi ini memperlihatkan kompleksitas pola penyebab perceraian yang akan dianalisis lebih lanjut dengan metode machine learning.

B. Hasil Training Model XGBoost

Tabel 3. Hasil Evaluasi Model (Classification Report)

Kelas	Precision	Recall	F1-Score	Support
Rendah	0.60	1.00	0.75	3
Sedang	0.00	0.00	0.00	2
Tinggi	1.00	1.00	1.00	3
Accuracy			**0.75**	8
Macro Avg	0.53	0.67	0.58	8
Weighted Avg	0.60	0.75	0.66	8

Model XGBoost yang dikembangkan untuk mengklasifikasikan tingkat faktor penyebab perceraian menunjukkan hasil yang beragam antar kelas. Pada kelas Rendah, model berhasil mendeteksi seluruh data dengan tingkat recall sempurna (1.00), namun precision yang sebesar 0.60 mengindikasikan adanya beberapa kesalahan prediksi di mana data dari kelas lain salah diklasifikasikan sebagai kelas Rendah. Sebaliknya, model kurang mampu mengenali kelas Sedang, yang terlihat dari nilai precision, recall, dan F1-score yang semua mendekati nol. Hal ini kemungkinan disebabkan oleh jumlah data kelas Sedang yang lebih sedikit dibanding kelas lain, sehingga model kurang optimal dalam belajar pola pada kelas ini. Untuk kelas Tinggi, model menunjukkan performa yang sangat baik dengan nilai precision, recall, dan F1-score mencapai 1.00, menandakan semua data kelas ini berhasil diprediksi dengan akurat.

Secara keseluruhan, model mencapai akurasi 75%, namun metrik rata-rata menunjukkan adanya ketidakseimbangan performa antar kelas. Oleh karena itu, perlu dilakukan perbaikan pada model, misalnya dengan penyesuaian parameter, teknik sampling, atau penambahan data agar model dapat mengenali semua kelas dengan lebih baik.

Tabel 4. Feature Importance (Gain, Cover, Frequency)

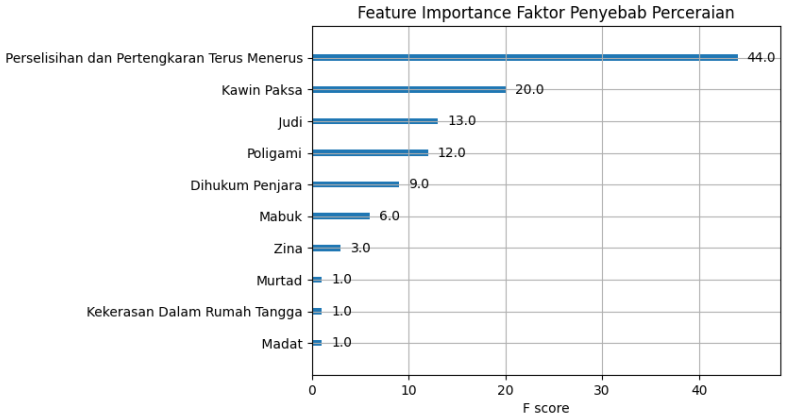
No	Fitur	Gain	Cover	Frequency
0	Perselisihan dan Pertengkaran Terus Menerus	2.458609	6.910818	44.0
1	Madat	1.872527	7.555555	1.0
2	Kawin Paksa	0.211123	2.798191	20.0
3	Judi	0.137297	2.249601	13.0
4	Kekerasan Dalam Rumah Tangga	0.112593	2.021325	1.0
5	Mabuk	0.021770	2.124622	6.0
6	Murtad	0.015241	2.091818	1.0
7	Dihukum Penjara	0.005880	2.033306	9.0
8	Zina	0.000338	2.016885	3.0
9	Poligami	0.000266	2.005066	12.0

Berdasarkan hasil analisis menggunakan algoritma XGBoost, faktor-faktor penyebab perceraian diidentifikasi dengan tingkat pengaruh yang berbeda-beda terhadap klasifikasi tingkat risiko perceraian. Faktor Perselisihan dan Pertengkaran Terus Menerus menempati posisi paling dominan dengan nilai Gain tertinggi sebesar 2,46, menunjukkan bahwa variabel ini memberikan kontribusi terbesar dalam meningkatkan akurasi model. Selain itu, faktor ini juga paling sering digunakan dalam proses pemisahan data (Frequency = 44), yang memperkuat perannya sebagai faktor utama penyebab perceraian. Faktor kedua yang cukup signifikan adalah Madat dengan nilai Gain 1,87, meskipun frekuensinya rendah (1 kali digunakan), menunjukkan bahwa meskipun jarang muncul dalam pemisahan data, dampaknya sangat besar bila dipertimbangkan.

Beberapa faktor lain seperti Kawin Paksa, Judi, dan Kekerasan Dalam Rumah Tangga juga berkontribusi cukup signifikan, dengan nilai Gain masing-masing 0,21, 0,14, dan 0,11 serta frekuensi penggunaan yang variatif. Faktor-faktor seperti Mabuk, Murtad, Dihukum Penjara, Zina, dan Poligami menunjukkan kontribusi yang relatif kecil berdasarkan nilai Gain dan frekuensi penggunaannya dalam model. Hal ini menunjukkan bahwa meskipun faktor-faktor tersebut ada dalam data, pengaruhnya terhadap prediksi risiko perceraian lebih rendah dibandingkan faktor-faktor utama.

Secara keseluruhan, hasil ini mengindikasikan bahwa konflik interpersonal yang berkelanjutan (perselisihan dan pertengkaran) merupakan penyebab utama perceraian yang paling dominan, diikuti oleh faktor penyalahgunaan zat (madat) dan beberapa faktor sosial lain seperti kawin paksa dan kekerasan dalam rumah tangga. Temuan ini dapat menjadi dasar bagi upaya pencegahan dan intervensi yang lebih terfokus pada faktor-faktor tersebut.

C. Feature Importance



Gambar 1. Feature Importance

Grafik feature importance menunjukkan bahwa faktor Perselisihan dan Pertengkaran Terus Menerus merupakan penyebab perceraian yang paling dominan, ditandai dengan frekuensi penggunaan tertinggi dalam model. Faktor ini memiliki pengaruh signifikan dalam membedakan tingkat risiko perceraian. Selain itu, faktor seperti Kawin Paksa, Judi, dan Poligami juga berkontribusi cukup besar, meskipun frekuensinya lebih rendah. Faktor-faktor lainnya memiliki peran yang lebih kecil dalam prediksi model. Hasil ini menegaskan pentingnya mengatasi konflik interpersonal sebagai langkah utama dalam upaya pencegahan perceraian.

4. CONCLUSION

Penelitian ini berhasil mengidentifikasi faktor-faktor utama penyebab perceraian di Indonesia menggunakan pendekatan machine learning, khususnya algoritma Extreme Gradient Boosting (XGBoost), berdasarkan data dari 34 provinsi. Model yang dibangun mampu mengklasifikasikan tingkat risiko perceraian dengan akurasi 75%. Hasil analisis menunjukkan bahwa faktor perselisihan dan pertengkaran terus-menerus merupakan penyebab perceraian yang paling dominan, diikuti oleh penyalahgunaan zat (madat) dan kawin paksa. Faktor-faktor ini memiliki kontribusi terbesar dalam memengaruhi jumlah perceraian menurut hasil feature importance dari model XGBoost.

Penelitian ini memberikan kontribusi penting dengan menerapkan metode machine learning pada isu sosial di Indonesia, yang sebelumnya jarang dilakukan. Temuan ini diharapkan dapat menjadi dasar bagi pembuat kebijakan untuk menyusun strategi pencegahan perceraian yang lebih terarah

REFERENCES

Badan Pusat Statistik. (2024). *Jumlah perceraian menurut provinsi dan faktor penyebab perceraian (perkara)*. BPS Indonesia. https://bps.go.id/

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Dangeti, P. (2017). *Statistics for machine learning: Build supervised, unsupervised, and reinforcement learning models using both Python and R*. Packt Publishing Ltd.

Dariyo, A. (2004). *Memahami psikologi perceraian dalam kehidupan keluarga* (Vol. 2, Issue 2).

Dava Maulana, M., Hadiana, A. I., & Umbara, R. (2023). Algoritma XGBoost untuk klasifikasi kualitas air minum. *Jurnal Mahasiswa Teknik Informatika*, 7(5). https://www.kaggle.com/datasets/adityak

Fitri, I. S. (2022). Faktor penyebab tingginya angka cerai gugat di Pengadilan Agama Bandung. *Al-Ahwal Al-Syakhsiyyah: Jurnal Hukum Keluarga dan Peradilan Islam*, 3(1), 101–116. https://doi.org/10.15575/as.v3i1.17547

Hidayati, L. (2021). Fenomena tingginya angka perceraian di Indonesia antara pandemi dan solusi. *Jurnal Psikologi Sosial*, 3(1), 71–87.

Jihaan Fauziyah Rahman, & Fahmi Fatwa Rosyadi Satria Hamdani. (2024). Faktor-faktor penyebab tingginya angka perceraian di Kabupaten Purwakarta tahun 2021–2023. *Jurnal Riset Hukum Keluarga Islam*, 4(2), 97–104. https://doi.org/10.29313/jrhki.v4i2.5192

Jo, T. (2021). *Machine learning foundations*. Springer Nature Switzerland AG.

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://github.com/slundberg/shap

Olson, D. H., & DeFrain, J. (2003). *Marriage and family: Intimacy, diversity, and strengths* (4th ed.). McGraw-Hill.

Pengadilan Agama Bangkinang. (2024). *Faktor penyebab perceraian di Kabupaten Kampar*. Direktorat Jenderal Badan Peradilan Agama. https://badilag.mahkamahagung.go.id

Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). Pemanfaatan machine learning dalam berbagai bidang: Review paper. *Indonesian Journal on Computer and Information Technology*, 5(1), 1–12.

Syafriani Manna, N., Doriza, S., & Oktaviani, M. (2021). Cerai gugat: Telaah penyebab perceraian pada keluarga di Indonesia. *Sosiologi Hukum*, 6(1), 1–15. https://doi.org/10.36722/sh.v6i1.443

Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a XGBoost algorithm for feature importance evaluation. *Energies*, 10(8), 1118. https://doi.org/10.3390/en10081118