

OPTIMASI MODEL KLASIFIKASI PESAN PENIPUAN DIGITAL BERBAHASA INDONESIA MENGGUNAKAN XGBOOST DENGAN BAYESIAN OPTIMIZATION DAN SMOTE

Muhammad Ardiansyah Sembiring¹, Mustika Fitri Larasati Sibuea², Anjani³

Sistem Informasi, Universitas Royal

e-mail: adinmantap88@gmail.com¹

Abstract: Digital fraud through short messages has become a national problem in Indonesia; the Indonesia Anti-Scam Centre recorded 637,055 reports with losses of IDR 9.1 trillion up to July 2026. Scam messages are hard to filter with keywords because perpetrators keep changing wording and phone numbers, while official promotional messages share similar linguistic patterns. This study aims to optimize a three-class classifier (normal, fraud, promotion) for Indonesian messages by combining XGBoost, SMOTE, and Bayesian Optimization based on the Tree-structured Parzen Estimator (TPE), and to compare TPE with Grid Search and Random Search under an equal budget of 25 trials. Messages were normalized (URLs, phone numbers, amounts, and slang) and represented with combined word and character n-gram TF-IDF features. On 2,320 messages, XGBoost + SMOTE + TPE achieved a macro F1 of 0.9555, an accuracy of 96.34%, and a fraud-class recall of 0.9507, slightly higher than default XGBoost (F1 0.9531) and Grid Search (F1 0.9529), with a much lighter model that trained 3.2 times faster. Grid Search showed overfitting to the validation folds. The ablation study showed that SMOTE did not help under moderate class imbalance (F1 without SMOTE 0.9581), and Multinomial Naïve Bayes remained the best model (F1 0.9693). Most errors occurred between fraud and promotional messages.

Keywords: Digital Fraud, Text Classification, XGBoost, Bayesian Optimization, SMOTE

Abstrak: Penipuan digital melalui pesan singkat telah menjadi masalah nasional; Indonesia Anti-Scam Centre mencatat 637.055 laporan dengan kerugian Rp 9,1 triliun hingga Juli 2026. Pesan penipuan sulit disaring dengan kata kunci karena pelaku terus mengubah variasi kata dan nomor telepon, sementara pesan promosi resmi memiliki pola bahasa yang mirip. Penelitian ini bertujuan mengoptimasi model klasifikasi tiga kelas (normal, penipuan, promo) untuk pesan berbahasa Indonesia dengan menggabungkan XGBoost, SMOTE, dan Bayesian Optimization berbasis Tree-structured Parzen Estimator (TPE), serta membandingkan TPE dengan Grid Search dan Random Search pada anggaran yang sama, yaitu 25 percobaan. Pesan dinormalisasi (tautan, nomor telepon, nominal, dan kata tidak baku) lalu direpresentasikan dengan fitur TF-IDF gabungan n-gram kata dan karakter. Pada 2.320 pesan, XGBoost + SMOTE + TPE mencapai F1-macro 0,9555, akurasi 96,34%, dan recall kelas penipuan 0,9507, sedikit lebih tinggi daripada XGBoost default (F1 0,9531) dan Grid Search (F1 0,9529), dengan model yang jauh lebih ringan dan waktu pelatihan 3,2 kali lebih cepat. Grid Search menunjukkan gejala overfitting terhadap fold validasi. Uji ablasi menunjukkan SMOTE tidak membantu pada ketidakseimbangan kelas moderat (F1 tanpa SMOTE 0,9581), dan Multinomial Naïve Bayes tetap menjadi model terbaik (F1 0,9693). Sebagian besar kesalahan terjadi antara pesan penipuan dan promo.

Kata Kunci: Penipuan Digital, Klasifikasi Teks, XGBoost, Bayesian Optimization, SMOTE.

PENDAHULUAN

Penipuan digital (*scam*) melalui SMS dan aplikasi pesan instan seperti WhatsApp menjadi salah satu ancaman keamanan siber yang paling dirasakan masyarakat Indonesia. Indonesia Anti-Scam Centre (IASC) yang dibentuk Otoritas Jasa Keuangan mencatat 637.055 laporan transaksi penipuan sejak 22 November 2024 hingga 31 Juli 2026 dengan total kerugian Rp9,1 triliun; 607.210 rekening telah diblokir dan 145.503 nomor telepon teridentifikasi terkait penipuan (Periskop.id, 2026). Pada periode Januari–Juli 2026, Kementerian Komunikasi dan Digital juga menerima laporan sekitar 30.000 nomor yang digunakan untuk *scamming* (Kompas.com, 2026).

Modus penipuan terus berkembang: undian dan hadiah palsu, pesan "mama minta pulsa", tautan resi paket berisi berkas APK berbahaya, pemberitahuan palsu perubahan tarif bank, permintaan kode OTP, pinjaman daring ilegal, lowongan kerja fiktif, hingga promosi judi daring. Di sisi lain, pengguna juga menerima pesan promosi resmi dari operator seluler, bank, dan *e-commerce* yang sama-sama memuat kata "selamat", nominal uang, dan tautan. Kemiripan ini membuat penyaringan berbasis kata kunci atau daftar hitam nomor mudah ditembus, karena pelaku cukup mengganti nomor dan variasi kata.

Klasifikasi pesan spam dengan *machine learning* telah banyak diteliti. Almeida et al. (2011) memperkenalkan korpus SMS Spam Collection dan menunjukkan kuatnya SVM linear, sedangkan kajian Abdulhamid et al. (2017) mencatat bahwa penyaringan SMS spam didominasi Naïve Bayes, SVM, dan metode ensambel. Untuk bahasa Indonesia, Wibisono (2018) mempublikasikan dataset SMS berlabel normal, penipuan, dan promo. Penelitian yang dilakukan oleh Herwanto et al. (2021) membandingkan Multinomial Naïve Bayes dan SVM dengan akurasi terbaik 95%, sedangkan Amin et al. (2024) menerapkan IndoBERT dan Multilingual

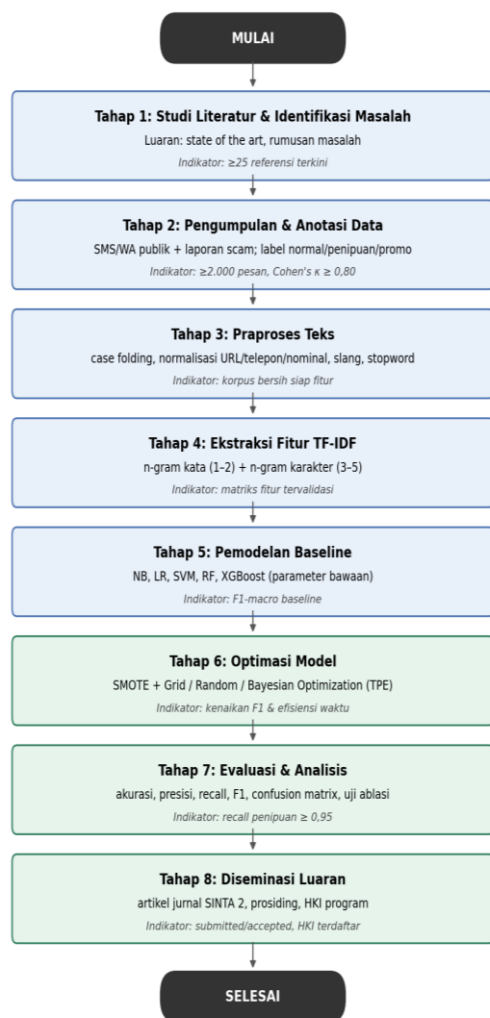
BERT dengan akurasi masing-masing 98% dan 95%, namun dengan kebutuhan komputasi yang jauh lebih besar.

Sebagian besar penelitian tersebut masih memakai parameter bawaan atau penyetelan manual, belum menangani ketidakseimbangan kelas secara eksplisit, dan jarang membandingkan efisiensi metode optimasi *hyperparameter*. Padahal, kinerja algoritma *gradient boosting* seperti XGBoost (Chen & Guestrin, 2016) sangat bergantung pada konfigurasi *hyperparameter* (Probst et al., 2019), dan pada data pesan penipuan jumlah kelas minoritas umumnya lebih sedikit sehingga model cenderung bias ke kelas mayoritas (Chawla et al., 2002).

Berdasarkan kondisi tersebut, penelitian ini bertujuan mengoptimasi model klasifikasi pesan penipuan digital berbahasa Indonesia dengan menggabungkan XGBoost, SMOTE, dan *Bayesian Optimization* berbasis *Tree-structured Parzen Estimator* (TPE) (Bergstra et al., 2011; Snoek et al., 2012), serta membandingkan TPE dengan *Grid Search* dan *Random Search* pada anggaran komputasi yang sama. Kontribusi penelitian ini adalah: (1) klasifikasi tiga kelas yang secara eksplisit memisahkan pesan penipuan dari promo resmi; (2) integrasi SMOTE di dalam *pipeline* validasi silang bersama optimasi TPE; dan (3) analisis efektivitas, efisiensi, dan kepentingan *hyperparameter* dari ketiga metode optimasi.

METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental. Tahapan penelitian ditunjukkan pada Gambar 1, dimulai dari studi literatur hingga diseminasi hasil.



Gambar 1 Tahapan Penelitian

Dataset

Dataset terdiri atas 2.320 pesan singkat berbahasa Indonesia dengan tiga label, yaitu normal (1.084 pesan), penipuan (710 pesan), dan promo (526 pesan). Struktur data mengikuti dataset SMS berbahasa Indonesia yang dipublikasikan oleh Wibisono (2018), yaitu kolom teks dan label. Pesan **penipuan** adalah pesan yang bertujuan memperoleh uang, data pribadi, atau akses perangkat secara curang; pesan **promo** adalah penawaran komersial dari pihak resmi; sisanya **normal**. Data dibagi secara berstrata menjadi 80% data latih (1.856 pesan) dan 20% data uji (464 pesan).

Praproses Teks

Praproses meliputi: (1) *case folding*; (2) normalisasi entitas dengan ekspresi

reguler, yaitu tautan menjadi *tokenurl*, nomor telepon menjadi *tokentelp*, persentase menjadi *tokenpersen*, nominal rupiah menjadi *tokenuang*, dan angka lain menjadi *tokenangka*; (3) penghapusan tanda baca; (4) normalisasi kata tidak baku dengan kamus (misalnya "yg" → "yang", "tlg" → "tolong"); dan (5) penghapusan *stopword* ringan. *Stemming* tidak dilakukan karena imbuhan seperti "memenangkan" dan "mendapatkan" justru menjadi penanda pesan penipuan.

Ekstraksi Fitur

Teks direpresentasikan dengan pembobotan TF-IDF pada persamaan (1) dengan normalisasi L2.

$$w(t,d) = (1 + \log tf) \times [\log((1+N)/(1+df)) + 1] \quad (1)$$

Dua jenis fitur digabung, yaitu n-gram kata (1–2) untuk menangkap frasa seperti "klik tautan" dan n-gram karakter (3–5) untuk menangkap variasi ejaan dan salah ketik, sehingga diperoleh 4.794 fitur. Kosakata dibentuk hanya dari data latih.

Model Klasifikasi

Lima model dengan parameter bawaan digunakan sebagai pembanding (*baseline*), yaitu Multinomial Naïve Bayes, Logistic Regression, Linear SVM, Random Forest, dan XGBoost, yang diimplementasikan dengan scikit-learn (Pedregosa et al., 2011). XGBoost meminimalkan fungsi objektif teregularisasi pada persamaan (2) dan (3), dengan T jumlah daun dan w bobot daun (Chen & Guestrin, 2016).

$$L = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k) \quad (2)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

Optimasi Hyperparameter

Optimasi dilakukan pada *pipeline* SMOTE → XGBoost. SMOTE membentuk sampel sintesis kelas minoritas dengan persamaan (4), dengan $\lambda \sim U(0,1)$ dan x_{zi} salah satu dari $k = 5$ tetangga terdekat (Chawla et al., 2002; Lemaître et al., 2017). SMOTE hanya diterapkan pada *fold* latih agar tidak terjadi kebocoran data.

$$x_{\text{baru}} = x_i + \lambda (x_{zi} - x_i) \quad (4)$$

Tiga metode optimasi dibandingkan dengan anggaran 25 percobaan dan fungsi objektif rata-rata F1-*macro* validasi silang 3-fold: (a) *Grid Search* pada kisi 108 kombinasi (diambil 25 kombinasi acak agar anggaran setara); (b) *Random Search* (Bergstra & Bengio, 2012); dan (c) TPE dengan Optuna (Akiba et al., 2019). TPE memodelkan distribusi *hyperparameter* yang menghasilkan skor baik, $l(x)$, dan skor buruk, $g(x)$, lalu memilih kandidat yang memaksimalkan rasio pada persamaan (5) (Bergstra et al., 2011).

$$x^* = \operatorname{argmax} l(x) / g(x) \quad (5)$$

Tabel 1 Ruang Pencarian Hyperparameter

Parameter	Random & TPE	Grid
n_estimators	100–500	100; 200; 400
max_depth	3–10	3; 6; 9
learning_rate	0,01–0,3	0,03; 0,1; 0,3
subsample	0,6–1,0	0,7; 1,0
colsample_bytree	0,3–1,0	0,5; 1,0
min_child_weight	1–7	1
reg_lambda	0,1–10	1

Teknik Evaluasi

Model dievaluasi pada data uji yang tidak digunakan selama optimasi dengan akurasi, presisi (P), *recall* (R), dan F1-*score* rata-rata *macro* pada persamaan (6). Karena meloloskan pesan penipuan lebih merugikan daripada salah menandai pesan normal, *recall* kelas penipuan dijadikan indikator tambahan dengan target $\geq 0,95$. Evaluasi pada data tidak seimbang tidak cukup hanya menggunakan akurasi (Saito & Rehmsmeier, 2015). Analisis pendukung meliputi *confusion matrix*, kurva konvergensi, uji ablasi SMOTE, dan kepentingan *hyperparameter* (fANOVA).

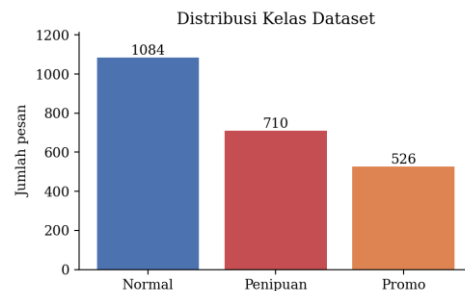
$$F1 = 2PR / (P + R) \quad (6)$$

HASIL DAN PEMBAHASAN

Data dan Praproses

Distribusi kelas ditunjukkan pada Gambar 2. Rasio kelas mayoritas terhadap

minoritas sebesar 2,06 : 1 menunjukkan ketidakseimbangan tingkat moderat. Contoh hasil praproses ditunjukkan pada Tabel 2; nomor telepon dan angka berubah menjadi token umum, sementara salah ketik seperti "ubungi" tetap dipertahankan karena akan ditangkap oleh fitur n-gram karakter.



Gambar 2 Distribusi Kelas Dataset

Tabel 2 Contoh Hasil Praproses

Teks Asli	Hasil Praproses
HORE ANDA DAPAT DISKON PERPANJANG 15GB DARI XL CEK DI APLIKASI RESMI	hore anda dapat diskon perpanjang tokenangka gb xl cek aplikasi resmi
saya petugas BPJS anda terlibat kasus ubungi 08388007845	saya petugas bpjs anda terlibat kasus ubungi tokentelp

Kinerja Model Baseline

Kinerja lima model dengan parameter bawaan pada data uji ditunjukkan pada Tabel 3. Seluruh model mencapai F1-*macro* di atas 0,95. Multinomial Naïve Bayes memperoleh F1-*macro* tertinggi, Linear SVM memiliki *recall* penipuan tertinggi, sedangkan XGBoost bawaan berada di urutan terakhir sehingga menjadi kandidat yang paling membutuhkan optimasi.

Tabel 3 Kinerja Model Baseline

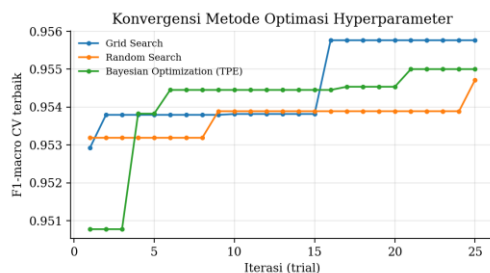
Model	Akurasi	F1- macro	Recall Penipuan
MNB	0,9741	0,9693	0,9648
LR	0,9698	0,9635	0,9789
SVM	0,9698	0,9634	0,9859
RF	0,9655	0,9581	0,9648
XGB	0,9612	0,9531	0,9577

Hasil Optimasi Hyperparameter

Perbandingan ketiga metode optimasi ditunjukkan pada Tabel 4 dan kurva konvergensi pada Gambar 3.

Tabel 4 Perbandingan Metode Optimasi

Metode	F1 CV	F1 Uji	Recall Penipuan	Waktu Latih (s)
XGB default	0,9504	0,9531	0,9577	5,9
Grid	0,9558	0,9529	0,9577	49,7
Random	0,9547	0,9554	0,9577	57,1
TPE	0,9550	0,9555	0,9507	15,4



Gambar 3 Kurva Konvergensi Metode Optimasi

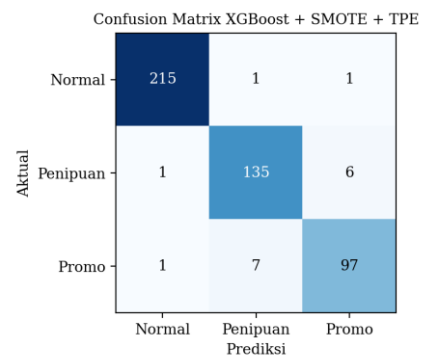
Ketiga metode optimasi meningkatkan F1 validasi silang dibanding XGBoost bawaan (0,9504), tetapi kenaikannya kurang dari satu poin persentase. *Grid Search* memperoleh F1 validasi tertinggi (0,9558) namun F1 uji terendah (0,9529); konfigurasi terbaiknya berupa pohon dalam ($\text{max_depth} = 9$) yang cenderung terlalu menyesuaikan diri dengan *fold* validasi. TPE menghasilkan F1 uji tertinggi (0,9555) dengan konfigurasi $\text{n_estimators} = 450$, $\text{max_depth} = 3$, $\text{learning_rate} = 0,021$, $\text{subsample} = 0,866$, $\text{colsample_bytree} = 0,592$, $\text{min_child_weight} = 2$, dan $\text{reg_lambda} = 0,606$. Model dangkal ini dilatih 3,2 kali lebih cepat daripada hasil *Grid Search*.

Pada Gambar 3, TPE sempat tertinggal pada tiga percobaan awal karena masih dalam fase eksplorasi acak, lalu melonjak pada percobaan ke-4 dan terus membaik hingga percobaan ke-21, sedangkan *Random Search* baru menemukan konfigurasi terbaiknya pada percobaan terakhir. Waktu optimasi ketiga metode relatif setara (858–1059 detik) karena biaya didominasi pelatihan model pada setiap percobaan. Selisih F1 uji antarmetode kurang dari 0,003 atau sekitar satu pesan dari 464 pesan uji, sehingga belum dapat dinyatakan signifikan.

Evaluasi Per Kelas

Tabel 5 Kinerja Per Kelas XGBoost + SMOTE + TPE

Kelas	Presisi	Recall	F1
Normal	0,9908	0,9908	0,9908
Penipuan	0,9441	0,9507	0,9474
Promo	0,9327	0,9238	0,9282
Macro	0,9558	0,9551	0,9555



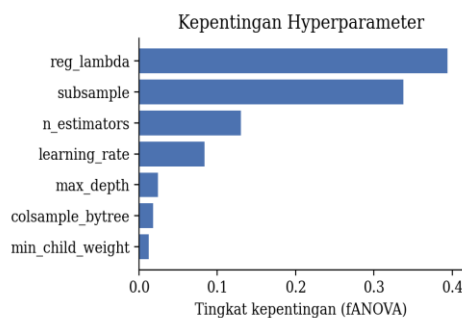
Gambar 4 Confusion Matrix XGBoost + SMOTE + TPE

Kelas normal diklasifikasikan hampir sempurna (F1 0,9908). Dari 17 kesalahan, 13 di antaranya (76,47%) terjadi antara kelas penipuan dan promo: 6 pesan penipuan diprediksi sebagai promo dan 7 pesan promo diprediksi sebagai penipuan (Gambar 4). Hal ini mengonfirmasi bahwa tantangan utama adalah membedakan pesan penipuan dari promosi resmi yang sama-sama memuat kata "selamat", nominal, dan ajakan mengklik. *Recall* kelas penipuan sebesar 0,9507 telah memenuhi target minimal 0,95.

Uji Ablasi SMOTE dan Kepentingan Hyperparameter

Ketika SMOTE dihapus dengan parameter TPE yang sama, F1-macro uji

naik dari 0,9555 menjadi 0,9581 dan *recall* penipuan naik dari 0,9507 menjadi 0,9648. Pada ketidakseimbangan moderat, interpolasi SMOTE pada ruang TF-IDF yang sangat jarang (*sparse*) diduga menghasilkan vektor sintetis yang tidak mewakili pesan nyata, terutama untuk kelas penipuan dan promo yang saling berdekatan. Karena itu, *oversampling* sebaiknya tidak diterapkan otomatis tanpa uji ablasi; alternatifnya adalah pembobotan kelas (*class weight*).



Gambar 5 Kepentingan Hyperparameter (fANOVA)

Analisis fANOVA (Gambar 5) menunjukkan bahwa *reg_lambda* (0,39) dan *subsample* (0,34) merupakan *hyperparameter* paling berpengaruh, diikuti *n_estimators* (0,13) dan *learning_rate* (0,08). Temuan ini sejalan dengan Bergstra & Bengio (2012) serta Probst et al. (2019) bahwa hanya sebagian kecil *hyperparameter* yang menentukan kinerja, sehingga ruang pencarian dapat dipersempit pada penelitian lanjutan.

Pembahasan

Walaupun optimasi meningkatkan XGBoost, Multinomial Naïve Bayes dengan parameter bawaan tetap memperoleh *F1-macro* tertinggi (0,9693) dengan waktu pelatihan hampir nol. Temuan ini konsisten dengan Almeida et al. (2011) dan Herwanto et al. (2021) yang menunjukkan kuatnya model linear untuk teks pendek berdimensi tinggi. Pada fitur TF-IDF yang jarang, batas keputusan antarkelas cenderung linear sehingga keunggulan XGBoost dalam memodelkan interaksi fitur tidak banyak termanfaatkan. Akurasi XGBoost teroptimasi (96,34%) sebanding dengan penelitian Herwanto et

al. (2021) dan berada di bawah IndoBERT (Amin et al., 2024), namun dengan kebutuhan komputasi yang jauh lebih rendah sehingga lebih layak diterapkan pada perangkat bergerak.

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat ditarik kesimpulan sebagai berikut.

1. Seluruh model *baseline* dengan fitur TF-IDF gabungan *n-gram* kata dan karakter mencapai *F1-macro* di atas 0,95; Multinomial Naïve Bayes tertinggi (0,9693) dan XGBoost bawaan terendah (0,9531).
2. Optimasi XGBoost dengan SMOTE dan TPE meningkatkan *F1-macro* dari 0,9531 menjadi 0,9555 dan akurasi dari 96,12% menjadi 96,34%, dengan *recall* penipuan 0,9507.
3. Dengan anggaran yang sama, TPE menghasilkan *F1* uji tertinggi dan model paling ringan, sedangkan *Grid Search* menunjukkan gejala *overfitting* terhadap data validasi.
4. SMOTE tidak bermanfaat pada ketidakseimbangan moderat (*F1* tanpa SMOTE 0,9581), dan kesalahan didominasi kebingungan antara pesan penipuan dan promo.

Penelitian selanjutnya disarankan memvalidasi model pada data pesan penipuan nyata yang lebih besar dengan beberapa *random seed* dan uji signifikansi, mengganti SMOTE dengan pembobotan kelas, menerapkan optimasi Bayesian pada model linear dan *ensemble*, serta membandingkannya dengan IndoBERT.

DAFTAR PUSTAKA

- Abdulhamid, S. M., Shuaib, M., Osho, O., Ismaila, I., & Alhassan, J. K. (2017). A Review on Mobile SMS Spam Filtering Techniques. *IEEE Access*, 5, 15650–15666. <https://doi.org/10.1109/ACCESS.2017.2666785>

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. <https://doi.org/10.1145/3292500.3330701>
- Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011). Contributions to the Study of SMS Spam Filtering: New Collection and Results. *Proceedings of the 11th ACM Symposium on Document Engineering*, 259–262. <https://doi.org/10.1145/2034691.2034742>
- Amin, M. B. M., Hakim, G., Maulana, M. T., Alwan, M. F., Anggraheni, H. S., Naufal, M. J., & Yudistira, N. (2024). Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model BERT. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(6), 1291–1302. <https://doi.org/10.25126/jtiik.2024118121>
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *Advances in Neural Information Processing Systems*, 24, 2546–2554.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- 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>
- Herwanto, Chusna, N. L., & Arif, M. S. (2021). Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(4), 1316–1325. <https://doi.org/10.30865/mib.v5i4.3119>
- Kompas.com. (2026, Juli 23). Komdigi Terima Laporan 30.000 Nomor Scamming Sepanjang Januari-Juli 2026. <https://nasional.kompas.com/read/2026/07/23/17250811/komdigi-terima-laporan-30000-nomor-scamming-sepanjang-januari-juli-2026>
- Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, 18(17), 1–5.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Periskop.id. (2026, September 1). OJK Blokir Rp724,1 Miliar Dana Scam, 607 Ribu Rekening Dibekukan. <https://periskop.id/keuangan/20260901/ojk-blokir-rp7241-miliar-dana-scam-607-ribu-rekening-dibekukan>
- Probst, P., Boulesteix, A.-L., & Bischl, B. (2019). Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *Journal of Machine Learning Research*, 20(53), 1–32.
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural*

Information Processing Systems, 25,
2951–2959.

Wibisono, Y. (2018, Agustus 5). Dataset
Klasifikasi Bahasa Indonesia (SMS
Spam) & Klasifikasi Teks dengan

Scikit-Learn.

[https://yudiwbs.wordpress.com/2018/
08/05/dataset-klasifikasi-bahasa-
indonesia-sms-spam-klasifikasi-teks-
dengan-scikit-learn/](https://yudiwbs.wordpress.com/2018/08/05/dataset-klasifikasi-bahasa-indonesia-sms-spam-klasifikasi-teks-dengan-scikit-learn/)