**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Cv. Utility Project Solution

# Komparasi Skenario Fitur dan Model Machine Learning untuk Klasifikasi Risiko Perizinan DPMPTSP Kota Banjarmasin

**Raisya Alifa Khansa*[1], Windarsyah[2], Finki Dona Marleny[3]**
[1,2,3]S1 Informatika, Universitas Muhammadiyah Banjarmasin, Banjarmasin, Indonesia
Email: [1],*raisya_alifa_2255201110031@umbjm.ac.id, [2]windarsyah@umbjm.ac.id, [3]finkidona@umbjm.ac.id
Email Penulis Korespondensi: [1]raisya_alifa_2255201110031@umbjm.ac.id

**Info Artikel**

**Kata Kunci:**
klasifikasi risiko perizinan
machine learning
XGBoost
OSS-RBA
feature importance
PP No. 5/2021

## Abstrak

Tingginya beban kerja verifikasi manual perizinan di DPMPTSP Kota Banjarmasin mendorong kebutuhan terhadap sistem klasifikasi otomatis berbasis data historis OSS-RBA. Penelitian ini membandingkan enam algoritma klasifikasi *machine learning — Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, Support Vector Machine,* dan *XGBoost —* melalui tiga percobaan dengan skenario fitur yang berbeda. Percobaan 1 menggunakan empat parameter resmi PP No. 5/2021, Percobaan 2 memperluas fitur dengan menambahkan variabel KL/Sektor Pembina, Jenis Perusahaan, dan Jenis Proyek, serta Percobaan 3 melakukan optimasi *hyperparameter* menggunakan *Grid Search* pada model terbaik. *Dataset* terdiri dari 1.100 data perizinan dengan empat kelas risiko, dievaluasi menggunakan *10-Fold Stratified Cross Validation* dan pengujian pada data uji terpisah (80:20). Model terbaik yang dihasilkan adalah *XGBoost* dengan parameter diperluas dan hyperparameter teroptimasi, mencapai akurasi 84,55% dan *F1 Macro* 0,7728. Analisis *feature importance* pada model final mengungkap bahwa KL/Sektor Pembina berkontribusi 48,94% terhadap prediksi — jauh melampaui seluruh parameter resmi PP seperti Jumlah Investasi (3,98%) dan Skala Usaha (2,94%) — mengindikasikan bahwa regulasi yang berlaku belum mengoptimalkan seluruh variabel yang relevan secara empiris dalam penentuan tingkat risiko perizinan berusaha. Model *XGBoost* teroptimasi yang dihasilkan berpotensi diimplementasikan sebagai sistem rekomendasi klasifikasi risiko otomatis di lingkungan DPMPTSP Kota Banjarmasin, guna mendukung verifikasi data perizinan secara lebih efisien, konsisten, dan terstruktur.

**Keywords:**
licensing risk classification,
machine learning
XGBoost
OSS-RBA
feature importance
Government Regulation No. 5/2021

## Abstract

The high manual verification workload in business licensing at DPMPTSP Banjarmasin City necessitates an automated classification system based on OSS-RBA historical data. This study compares six machine learning classification algorithms — Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, Support Vector Machine, and XGBoost — across three experiments with different feature scenarios. Experiment 1 uses four official parameters from Government Regulation No. 5/2021, Experiment 2 expands the feature set with KL/Sektor Pembina, Company Type, and Project Type variables, and Experiment 3 applies Grid Search hyperparameter optimization to the best-performing model. The dataset comprises 1,100 licensing records with four risk classes, evaluated using 10-Fold Stratified Cross Validation and a held-out test set (80:20 split). The best model produced is XGBoost with expanded parameters and optimized hyperparameters, achieving 84.55% accuracy and F1 Macro 0.7728. Feature importance analysis on the final model reveals that KL/Sektor Pembina contributes 48.94% to predictions — far exceeding all official Government Regulation parameters such as Investment Amount (3.98%) and Business Scale (2.94%) — indicating that the existing regulation has not fully optimized all empirically relevant variables in determining business licensing risk levels. The optimized XGBoost model produced in this study has the potential to be implemented as an automated risk classification recommendation system within DPMPTSP Banjarmasin City, supporting business licensing data verification in a more efficient, consistent, and structured manner.




Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

## 1. PENDAHULUAN

Pertumbuhan realisasi investasi di Kota Banjarmasin dalam beberapa tahun terakhir berdampak langsung pada meningkatnya volume permohonan perizinan usaha yang harus diproses oleh Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kota Banjarmasin. Sebagai instansi yang bertanggung jawab atas pengelolaan perizinan berusaha di tingkat daerah, DPMPTSP mengelola data permohonan melalui platform *Online Single Submission Risk-Based Approach* (OSS-RBA) yang diamanatkan oleh Peraturan Pemerintah Nomor 5 Tahun 2021 tentang Penyelenggaraan Perizinan Berusaha Berbasis Risiko [1]. Regulasi ini mengubah paradigma perizinan secara mendasar, dari pendekatan berbasis kepatuhan administratif menjadi berbasis penilaian risiko kegiatan usaha. Dalam sistem OSS-RBA, setiap kegiatan usaha diklasifikasikan ke dalam empat tingkat risiko: Rendah, Menengah Rendah, Menengah Tinggi, dan Tinggi, berdasarkan empat parameter yang telah ditetapkan regulasi, yaitu Klasifikasi Baku Lapangan Usaha Indonesia (KBLI), skala usaha, nilai investasi, dan jumlah tenaga kerja [2]. Klasifikasi ini bersifat krusial karena secara langsung menentukan jenis dokumen legalitas yang diperlukan, intensitas pengawasan, serta apakah suatu permohonan cukup diproses otomatis oleh sistem atau harus melalui verifikasi dan persetujuan instansi teknis terkait [3].

Meskipun OSS-RBA telah mengotomatisasi sebagian besar proses pengajuan, beban kerja verifikasi manual oleh petugas DPMPTSP Kota Banjarmasin tetap signifikan, khususnya untuk permohonan berkategori risiko Menengah Tinggi dan Tinggi yang memerlukan penelaahan lebih mendalam sebelum izin dapat diterbitkan. Penentuan tingkat risiko yang melibatkan parameter multidimensi, mulai dari ribuan kode KBLI hingga nilai investasi yang sangat variatif antarsektor yang berpotensi menimbulkan inefisiensi waktu pemrosesan dan subjektivitas penilaian apabila masih bertumpu pada verifikasi manual [4]. Di sisi lain, data historis perizinan yang telah terkumpul dalam sistem OSS DPMPTSP Kota Banjarmasin belum dimanfaatkan secara optimal sebagai basis pengetahuan untuk mendukung proses pengambilan keputusan [5]. Dari 1.100 data perizinan proyek investasi yang tercatat, distribusi kelas risiko menunjukkan ketidakseimbangan yang nyata: kelas Rendah mendominasi dengan 572 data (52%), sementara kelas Menengah Rendah hanya berjumlah 107 data (9,7%). Kondisi ini mencerminkan kompleksitas pola perizinan di lapangan dan menjadi tantangan tersendiri dalam pengembangan sistem klasifikasi otomatis yang akurat dan adil antar kelas. Urgensi otomatisasi ini semakin mendesak mengingat pertumbuhan permohonan perizinan di Kota Banjarmasin yang terus meningkat seiring realisasi investasi daerah, sementara kapasitas verifikasi manual petugas DPMPTSP tidak bertumbuh proporsional. Tanpa dukungan sistem klasifikasi berbasis data, risiko penumpukan permohonan dan inkonsistensi penilaian antarpetugas akan terus menjadi hambatan dalam mewujudkan pelayanan perizinan yang cepat dan terstandar.

Pendekatan machine learning telah terbukti efektif dalam menangani permasalahan klasifikasi pada domain pemerintahan dan bisnis yang melibatkan data multidimensi berskala besar. Algoritma seperti *Decision Tree, Random Forest, K-Nearest Neighbor (*KNN*), Support Vector Machine (*SVM*), Naive Bayes, dan Extreme Gradient Boosting (XGBoost)* telah banyak digunakan untuk tugas klasifikasi dengan karakteristik data yang beragam [6]. Di antara algoritma tersebut, *XGBoost* secara konsisten menunjukkan keunggulan pada data tabular berstruktur, terutama dalam kondisi distribusi kelas tidak seimbang dan fitur kategorik multidimensi [7]. Algoritma *Random Forest* juga menunjukkan performa yang kompetitif dalam klasifikasi berbasis data tidak seimbang, di mana optimasi hyperparameter dan rekayasa fitur terbukti mampu meningkatkan akurasi dan kemampuan generalisasi model secara signifikan [8]. Studi perbandingan pada konteks data bisnis menunjukkan bahwa *XGBoost* menghasilkan akurasi lebih tinggi dibandingkan Random Forest dalam menangani data dengan distribusi kelas tidak merata [9]. Performa *XGBoost* dapat lebih ditingkatkan melalui optimasi hyperparameter menggunakan *Grid Search* dengan validasi silang berlapis, yang terbukti meningkatkan akurasi dan stabilitas model secara konsisten [10].

Terdapat celah penelitian yang menjadi landasan untuk studi ini. Pertama, penelitian Erkamim et al. yang membandingkan *Random Forest* dan *XGBoost* pada klasifikasi performa UMKM hanya menguji satu skenario fitur tanpa membandingkan dampak penambahan atau pengurangan variabel terhadap performa model, sehingga tidak memberikan landasan tentang seberapa besar kontribusi masing-masing variabel terhadap akurasi prediksi [9]. Kedua, Dachi dan Sitompul dalam penelitian klasifikasi keputusan kredit menggunakan *XGBoost* dan *Random Forest* juga hanya menguji variabel yang sudah ditetapkan dalam prosedur standar lembaga keuangan tanpa mengevaluasi apakah terdapat variabel di luar prosedur tersebut yang memiliki daya prediktif lebih tinggi [11]. Ketiga, studi Anggoro dan Mukti yang membandingkan *Grid Search* dan *Random Search* untuk *tuning XGBoost* berfokus sepenuhnya pada optimasi hyperparameter tanpa menyentuh pertanyaan apakah skenario fitur yang digunakan sudah optimal [12]. Ketiga studi tersebut secara kolektif menunjukkan bahwa penelitian terdahulu belum menjawab pertanyaan mendasar: apakah parameter yang ditetapkan oleh regulasi atau prosedur resmi sudah cukup representatif secara statistik, atau justru terdapat variabel kontekstual di luar regulasi yang lebih prediktif.

Berdasarkan celah tersebut, penelitian ini bertujuan membangun dan membandingkan model klasifikasi *machine learning* untuk penentuan tingkat risiko perizinan proyek investasi menggunakan 1.100 data historis OSS DPMPTSP Kota Banjarmasin. Pengujian dilakukan melalui tiga tahap percobaan: (1) enam algoritma klasifikasi diuji menggunakan parameter resmi PP No. 5/2021, (2) pengujian diulang dengan parameter diperluas yang mencakup variabel di luar

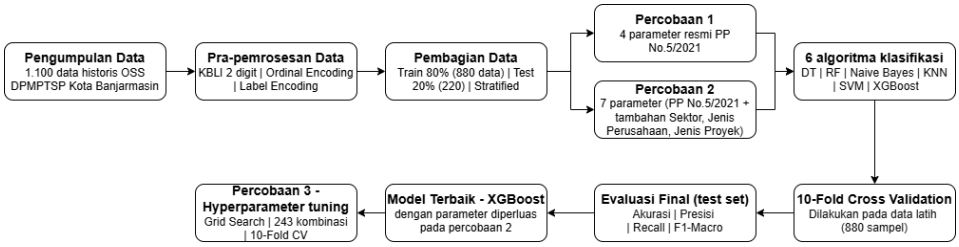

Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

regulasi, dan (3) optimasi *hyperparameter* dilakukan pada model dengan performa terbaik menggunakan *Grid Search*. Hipotesis yang diajukan adalah bahwa penambahan variabel di luar parameter resmi PP akan menghasilkan peningkatan akurasi klasifikasi yang signifikan. Komparasi model dan skenario fitur ini dilakukan bukan sekadar sebagai eksplorasi statistik, melainkan diarahkan untuk menemukan model klasifikasi terbaik yang secara praktis dapat mengurangi beban verifikasi manual petugas DPMPTSP Kota Banjarmasin dan meningkatkan konsistensi penentuan tingkat risiko perizinan dalam melakukan usaha.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan Eksperimental kuantitatif dengan menerapkan teknik *supervised learning* untuk klasifikasi tingkat risiko perizinan. Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan data hingga evaluasi model, sebagaimana ditunjukkan pada diagram alur pada Gambar 1.




**Gambar 1.** Diagram Alur Penelitian

### 2.1 Pengumpulan Data

Data yang digunakan merupakan data sekunder berupa rekaman historis perizinan proyek investasi yang dikelola oleh DPMPTSP Kota Banjarmasin melalui platform OSS-RBA. Dataset terdiri dari 1.100 baris data dengan 25 atribut, mencakup periode pengajuan perizinan yang telah tersertifikasi tingkat risikonya. Variabel target adalah tingkat risiko proyek dengan empat kelas: Rendah, Menengah Rendah, Menengah Tinggi, dan Tinggi, sesuai dengan klasifikasi yang ditetapkan dalam PP No. 5/2021 [1]. Distribusi kelas target disajikan pada Tabel 1.

**Tabel 1.** Distribusi Kelas Target

| Kelas Risiko | Jumlah Data | Persentase |
|---|---|---|
| Rendah | 572 | 52,00% |
| Menengah Tinggi | 287 | 26,09% |
| Tinggi | 134 | 12,18% |
| Menengah Rendah | 107 | 9,73% |
| **Total** | **1.100** | **100%** |

### 2.2 Rancangan Percobaan

Penelitian ini membandingkan dua skenario fitur melalui tiga tahap percobaan. Percobaan 1 menggunakan empat parameter resmi PP No. 5/2021 [1]. Percobaan 2 memperluas fitur dengan menambahkan tiga variabel di luar regulasi: KL/Sektor Pembina, Jenis Perusahaan, dan Jenis Proyek. Percobaan 3 melakukan optimasi hyperparameter pada model terbaik hasil Percobaan 2. Seluruh percobaan menggunakan pembagian data, nilai random state, dan prosedur evaluasi yang identik agar hasil antarskenario dapat dibandingkan secara *apple-to-apple*. Rincian variabel pada masing-masing skenario disajikan pada Tabel 2.

**Tabel 2.** Skenario Fitur Percobaan

| No | Variabel | Percobaan 1 | Percobaan 2 | Sumber |
|---|---|---|---|---|
| 1 | KBLI | ✔ | ✔ | PP No. 5/2021 |
| 2 | Skala Usaha | ✔ | ✔ | PP No. 5/2021 |
| 3 | Jumlah Investasi | ✔ | ✔ | PP No. 5/2021 |
| 4 | Jumlah Tenaga Kerja (TKI) | ✔ | ✔ | PP No. 5/2021 |
| 5 | KL/Sektor Pembina | ✗ | ✔ | Di luar PP |
| 6 | Jenis Perusahaan | ✗ | ✔ | Di luar PP |
| 7 | Jenis Proyek | ✗ | ✔ | Di luar PP |


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

### 2.3 Pra-pemrosesan Data

Tiga tahap pra-pemrosesan diterapkan sebelum pelatihan model. Pertama, variabel KBLI dipangkas menjadi dua digit pertama untuk merepresentasikan kategori besar lapangan usaha, sehingga mengurangi dimensi dan risiko overfitting. Kedua, variabel Skala Usaha dikodekan menggunakan *Ordinal Encoding* sesuai hierarki yang ditetapkan dalam Undang-Undang Cipta Kerja, yakni Usaha Mikro=1, Usaha Kecil=2, Usaha Menengah=3, dan Usaha Besar=4. Pengkodean ordinal dipilih karena mempertahankan relasi urutan antar kategori yang secara konseptual bermakna, berbeda dengan *Label Encoding* yang menghasilkan urutan arbitrer [13]. Ketiga, variabel kategorik lainnya (KBLI, KL/Sektor Pembina, Jenis Perusahaan, Jenis Proyek) dikodekan menggunakan *Label Encoding*. Ketidakseimbangan kelas ditangani menggunakan parameter *class_weight='balanced'* pada setiap model yang mendukungnya, sehingga model diberi bobot proporsional terhadap frekuensi kelas selama pelatihan tanpa mengubah distribusi data asli [14].

### 2.4 Pembagian Data

Dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan metode *Stratified Train-Test Split* dengan *random_state=42* untuk memastikan reprodusibilitas hasil. Stratifikasi diterapkan agar proporsi setiap kelas pada data latih dan data uji tetap konsisten dengan distribusi kelas pada dataset asli [15]. Pembagian ini menghasilkan 880 data latih dan 220 data uji.

### 2.5 Model yang Digunakan

Enam algoritma klasifikasi diuji pada Percobaan 1 dan 2, yaitu *Decision Tree* (DT), *Random Forest* (RF), *Naive Bayes* (NB), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (*XGBoost*). Keenam algoritma dipilih untuk mewakili pendekatan yang beragam: *Decision Tree* sebagai model berbasis aturan yang interpretabel [16], *Random Forest* dan *XGBoost* sebagai model *ensemble* [9], *Naive Bayes* sebagai model probabilistik, serta *K-Nearest Neighbor* dan *Support Vector Machine* sebagai model berbasis jarak dan margin. Khusus untuk *K-Nearest Neighbor* dan *Support Vector Machine*, *StandardScaler* diterapkan sebelum pelatihan menggunakan *Pipeline* untuk mencegah dominasi fitur berskala besar terhadap fitur berskala kecil [6].

### 2.6. Evaluasi Model

Evaluasi dilakukan dalam dua tahap. Pertama, *10-Fold Stratified Cross Validation* diterapkan pada data latih untuk mengukur stabilitas dan generalitas model [15]. Data latih dibagi menjadi 10 bagian secara bergantian, sembilan bagian untuk pelatihan dan satu bagian untuk validasi, sehingga setiap sampel berkesempatan menjadi data validasi tepat satu kali. Kedua, model terbaik dievaluasi secara final pada data uji yang sepenuhnya terpisah dari proses pelatihan dan validasi.

Metrik evaluasi yang digunakan meliputi akurasi, presisi, *recall*, dan *F1-score* dengan rata-rata makro (*macro average*) agar setiap kelas diperlakukan setara tanpa dipengaruhi dominasi kelas mayoritas. Persamaan masing-masing metrik didefinisikan sebagai berikut:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$
$$Precision = TP / (TP + FP) \quad (2)$$
$$Recall = TP / (TP + FN) \quad (3)$$
$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$
$$F1\text{-}Macro = (1/K) \times \Sigma\, F1k,\ k=1,...,K \quad (5)$$

di mana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, FN adalah *False Negative*, dan K adalah jumlah kelas (dalam penelitian ini K=4) [6].

### 2.7 Optimasi *Hyperparameter* (Percobaan 3)

Pada Percobaan 3, optimasi *hyperparameter XGBoost* dilakukan menggunakan *Grid Search Cross Validation* [10] dengan ruang pencarian sebagaimana ditampilkan pada Tabel 3. Metrik yang dioptimalkan adalah *F1-Macro* karena lebih representatif untuk data dengan distribusi kelas tidak seimbang dibandingkan akurasi [14]. Kombinasi total yang dievaluasi adalah $3^5 = 243$ kombinasi, masing-masing divalidasi menggunakan *10-Fold Stratified Cross Validation* sehingga total sebanyak 2.430 proses pelatihan dijalankan untuk menemukan kombinasi hyperparameter optimal.

**Tabel 3.** Ruang Pencarian *Hyperparameter Grid Search*

| Hyperparameter | Nilai yang Diuji |
|---|---|
| *n_estimators* | 100, 200, 300 |
| *max_depth* | 3, 4, 6 |
| *learning_rate* | 0.01, 0.05, 0.1 |


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

| | |
|---|---|
| *subsample* | 0.7, 0.8, 1.0 |
| *colsample_bytree* | 0.7, 0.8, 1.0 |

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari tiga percobaan yang dilakukan secara berurutan. Pembahasan mencakup hasil *cross validation* pada data latih, evaluasi final pada data uji, analisis *confusion matrix*, analisis *feature importance*, dan hasil optimasi *hyperparameter*. Seluruh percobaan menggunakan metrik yang identik sehingga perbandingan antarskenario dan antarmodel dapat dilakukan secara valid.

### 3.1 Hasil *Cross Validation* — Percobaan 1 (Parameter PP No. 5/2021)

Tahap pertama menggunakan empat parameter resmi PP No. 5/2021 sebagai fitur, yakni KBLI (2 digit), Skala Usaha, Jumlah Investasi, dan Jumlah Tenaga Kerja. Evaluasi stabilitas model dilakukan menggunakan *10-Fold Stratified Cross Validation* pada data latih (880 sampel). Hasil disajikan pada Tabel 4 [6].

**Tabel 4.** Hasil *Cross Validation* (*10-Fold*) pada *Train Set* — Percobaan 1

| Model | Akurasi | F1 Macro | Presisi Macro | Recall Macro |
|---|---|---|---|---|
| *XGBoost* | 0,7886 | 0,6596 | 0,6935 | 0,6496 |
| *KNN* | 0,7489 | 0,5815 | 0,6666 | 0,5715 |
| *Random Forest* | 0,7443 | 0,6115 | 0,6486 | 0,6056 |
| *Decision Tree* | 0,6023 | 0,5445 | 0,5848 | 0,5959 |
| *Naive Bayes* | 0,5182 | 0,1857 | 0,2153 | 0,2540 |
| *SVM* | 0,3159 | 0,3140 | 0,3659 | 0,4116 |

*XGBoost* menunjukkan kinerja *cross validation* terbaik dengan akurasi 0,7886 dan *F1 Macro* 0,6596. Hal ini mengonfirmasi bahwa *XGBoos*t paling stabil dalam menangani variasi data latih yang mencakup empat kelas risiko dengan distribusi tidak seimbang [7]. KNN dan *Random Forest* berada di posisi kedua dan ketiga dengan akurasi masing-masing 0,7489 dan 0,7443. SVM menunjukkan performa paling rendah (0,3159) karena sensitivitasnya terhadap perbedaan skala fitur — Jumlah Investasi (jutaan hingga miliaran) berbeda sangat jauh dengan TKI (satuan) meskipun *StandardScaler t*elah diterapkan. *Naive Bayes* juga *underperform* (0,5182) karena asumsi independensi antar fitur tidak terpenuhi pada data perizinan yang memiliki korelasi struktural antara KBLI dan parameter lainnya [6].

### 3.2 Hasil *Cross Validation* — Percobaan 2 (Parameter Diperluas)

Percobaan 2 memperluas fitur dengan menambahkan tiga variabel di luar PP No. 5/2021, yaitu KL/Sektor Pembina, Jenis Perusahaan, dan Jenis Proyek. Struktur evaluasi identik dengan Percobaan 1. Hasil *cross validation* disajikan pada Tabel 5 [9].

**Tabel 5.** Hasil *Cross Validation* (*10-Fold*) pada *Train Set* — Percobaan 2

| Model | Akurasi | F1 Macro | Presisi Macro | Recall Macro |
|---|---|---|---|---|
| *XGBoost* | 0,8727 | 0,7977 | 0,8168 | 0,7958 |
| *Random Forest* | 0,8602 | 0,7758 | 0,8023 | 0,7656 |
| *Decision Tree* | 0,8148 | 0,7477 | 0,7598 | 0,7908 |
| *KNN* | 0,7875 | 0,6708 | 0,7058 | 0,6555 |
| *SVM* | 0,7602 | 0,6723 | 0,6980 | 0,6727 |
| *Naive Bayes* | 0,5182 | 0,1857 | 0,2153 | 0,2540 |

Penambahan tiga variabel menghasilkan lonjakan performa yang signifikan di seluruh model. *XGBoost* kembali memimpin dengan akurasi 0,8727 dan *F1 Macro* 0,7977 — meningkat 8,41 poin persentase dibandingkan Percobaan 1. *Random Forest* mengikuti dengan akurasi 0,8602, sementara *Decision Tree* mengalami peningkatan paling dramatis dari 0,6023 menjadi 0,8148 (+21,25 poin). Fakta bahwa semua model berbasis pohon meningkat signifikan mengindikasikan bahwa variabel tambahan membawa sinyal kategorik yang kuat dan mudah dipartisi oleh struktur pohon keputusan [16].

### 3.3 Evaluasi Kinerja Model pada Test Set


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Cv. Utility Project Solution

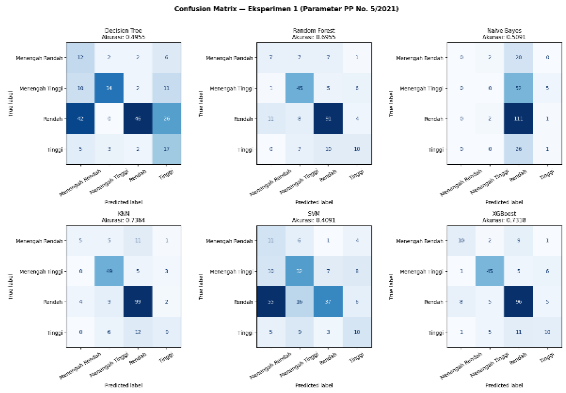
Evaluasi final dilakukan pada data uji (220 sampel) yang sepenuhnya terpisah dari proses pelatihan dan validasi. Hasil ini merepresentasikan kemampuan generalisasi nyata setiap model terhadap data yang belum pernah dilihat sebelumnya.

**a. Percobaan 1 — Parameter PP No. 5/2021**

Hasil evaluasi pada *test set* Percobaan 1 disajikan pada Tabel 6 [6].

**Tabel 6.** Hasil Evaluasi pada Test Set — Percobaan 1

| Model | Akurasi | F1 Macro | Presisi Macro | Recall Macro |
|---|---|---|---|---|
| KNN | 0,7364 | 0,5876 | 0,6613 | 0,5722 |
| *XGBoost* | 0,7318 | 0,6227 | 0,6344 | 0,6141 |
| *Random Forest* | 0,6955 | 0,5714 | 0,5804 | 0,5691 |
| *Naive Bayes* | 0,5091 | 0,1865 | 0,1685 | 0,2527 |
| *Decision Tree* | 0,4955 | 0,4793 | 0,5534 | 0,5438 |
| SVM | 0,4091 | 0,3918 | 0,4429 | 0,4391 |




**Gambar 2.** *Confusion Matrix* semua model — Percobaan 1

Pada *test set*, KNN meraih akurasi tertinggi (0,7364) meski pada *cross validation* berada di posisi kedua. *XGBoost* menyusul dengan akurasi 0,7318 namun unggul dalam *F1 Macro* (0,6227 vs 0,5876), yang menunjukkan keseimbangan prediksi yang lebih baik antar kelas [15]. *Decision Tree* dan SVM menunjukkan penurunan tajam dari *cross validation* ke *test set* — *Decision Tree* turun dari 0,6023 menjadi 0,4955 dan SVM dari 0,3159 menjadi 0,4091 — mengindikasikan kedua model mengalami *overfitting* atau ketidakstabilan yang tinggi pada parameter PP semata. Kelas Menengah Rendah secara konsisten menjadi kelas dengan performa terlemah di seluruh model, dengan *F1-score* tertinggi hanya 0,42. Kondisi ini didorong oleh jumlah data yang paling sedikit (107 dari 1.100 sampel) dan kemiripan karakteristik fitur dengan kelas Menengah Tinggi [14].

**b. Percobaan 2 — Parameter Diperluas**

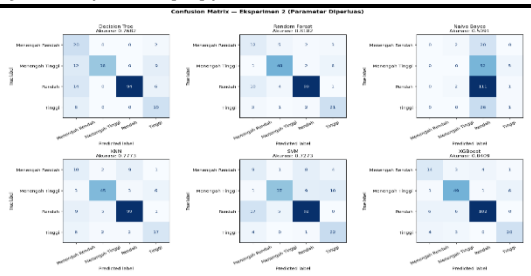
Hasil evaluasi pada test set Percobaan 2 disajikan pada Tabel 7 [9].

**Tabel 7.** Hasil Evaluasi pada Test Set — Percobaan 2

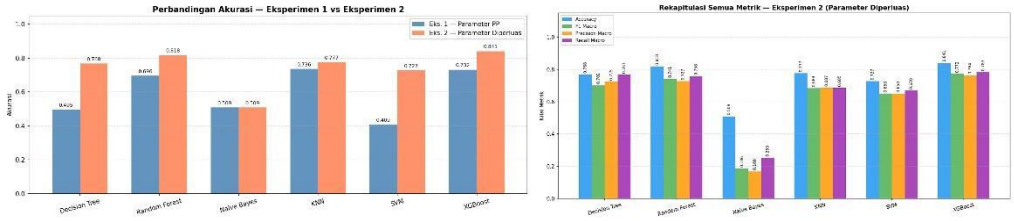
| Model | Akurasi | F1 Macro | Presisi Macro | Recall Macro |
|---|---|---|---|---|
| *XGBoost* | 0,8409 | 0,7725 | 0,7643 | 0,7829 |
| *Random Forest* | 0,8182 | 0,7421 | 0,7287 | 0,7606 |
| *KNN* | 0,7773 | 0,6950 | 0,6924 | 0,6997 |
| *Decision Tree* | 0,7682 | 0,7019 | 0,7245 | 0,7672 |
| *SVM* | 0,7273 | 0,6334 | 0,6338 | 0,6494 |
| *Naive Bayes* | 0,5091 | 0,1865 | 0,1685 | 0,2527 |


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT*
Penerbit : Cv. Utility Project Solution



**Gambar 3.** *Confusion Matrix* semua model — Percobaan 2



**Gambar 4.** Perbandingan akurasi dan Rekapitulasi semua metrik pada Percobaan 1 dan Percobaan 2

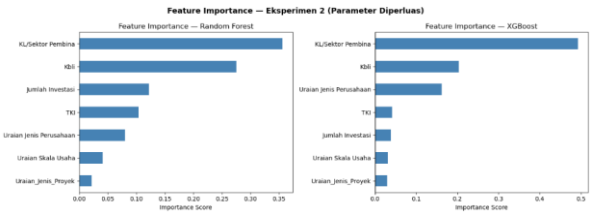
*XGBoost* kembali menjadi model terbaik dengan akurasi 0,8409 dan *F1 Macro* 0,7725, meningkat 10,91 poin persentase dari Percobaan 1. Peningkatan terbesar dialami SVM (+31,82 poin, dari 0,4091 menjadi 0,7273) dan *Decision Tree* (+27,27 poin, dari 0,4955 menjadi 0,7682). Lonjakan ekstrem pada SVM dan *Decision Tree* menegaskan bahwa performa buruk keduanya di Percobaan 1 bukan disebabkan oleh kelemahan mendasar algoritmanya, melainkan oleh keterbatasan fitur yang tersedia [6]. Satu-satunya model yang tidak terpengaruh penambahan fitur adalah *Naive Bayes* yang tetap stagnan di akurasi 0,5091, karena kegagalannya bersumber dari asumsi matematis yang tidak terpenuhi, bukan dari jumlah fitur [7].

### 3.4 Analisis *Feature Importance*

Analisis feature importance dilakukan pada model *Random Forest* dan *XGBoost* dari Percobaan 2 untuk mengidentifikasi variabel yang paling berkontribusi terhadap prediksi tingkat risiko. Hasil disajikan pada Tabel 8 [12].

**Tabel 8.** *Feature Importance* — Percobaan 2

| Fitur | Random Forest | XGBoost | Asal |
|---|---|---|---|
| KL/Sektor Pembina | 35,96% | 48,94% | Di luar PP |
| KBLI | 26,98% | 20,68% | PP No. 5/2021 |
| Jenis Perusahaan | 7,62% | 16,16% | Di luar PP |
| Jumlah Investasi | 12,52% | 3,98% | PP No. 5/2021 |
| TKI | 10,43% | 4,17% | PP No. 5/2021 |
| Jenis Proyek | 2,43% | 3,13% | Di luar PP |
| Skala Usaha | 4,06% | 2,94% | PP No. 5/2021 |



**Gambar 5.** *Feature Importance Random Forest* dan *XGBoost* pada Percobaan 2

KL/Sektor Pembina menempati posisi pertama dengan selisih yang sangat jauh: 35,96% pada *Random Forest* dan 48,94% pada *XGBoost*. Angka ini berarti hampir separuh kemampuan prediksi *XGBoost* bersumber dari satu variabel yang tidak tercantum dalam PP No. 5/2021. KL/Sektor Pembina merepresentasikan kementerian atau lembaga yang mengawasi kegiatan usaha sesuai KBLI-nya, dan setiap kementerian mengawasi sektor dengan profil risiko operasional,


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

lingkungan, dan sosial yang secara inheren berbeda [4]. KBLI (2 digit) berada di posisi kedua (*Random Forest*: 26,98%, *XGBoost*: 20,68%), yang masuk akal karena KBLI adalah basis klasifikasi utama dalam OSS-RBA.

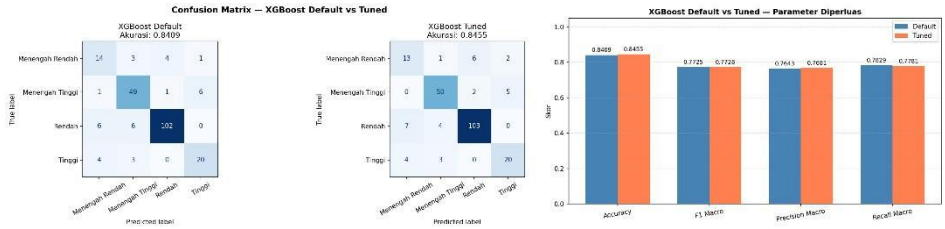
Temuan yang kontras adalah posisi Jumlah Investasi dan Skala Usaha — dua dari empat parameter resmi PP — yang justru berada di urutan bawah. Jumlah Investasi hanya berkontribusi 3,98% pada XGBoost dan Skala Usaha 2,94%, lebih rendah bahkan dari Jenis Perusahaan (16,16%) yang bukan merupakan parameter resmi. Kondisi ini dapat dijelaskan oleh distribusi nilai Jumlah Investasi yang sangat tumpang tindih antarkelas risiko — usaha berskala besar dengan investasi tinggi tidak selalu berkategori risiko Tinggi, bergantung pada sektor dan jenis kegiatannya [9] Temuan ini secara empiris mempertegas bahwa parameter resmi PP No. 5/2021 belum mengoptimalkan seluruh dimensi variabel yang relevan dalam penentuan tingkat risiko perizinan.

### 3.5 Hasil *Hyperparameter Tuning* — Percobaan 3

Percobaan 3 melakukan optimasi *hyperparameter XGBoost* dengan parameter diperluas menggunakan *Grid Search* dan *10-Fold Stratified Cross Validation*. Dari 243 kombinasi yang dievaluasi (2.430 proses pelatihan total), ditemukan kombinasi optimal: *learning_rate=0,1; max_depth=4; n_estimators=200; subsample=1,0; colsample_bytree=1,0*. Perbandingan performa sebelum dan sesudah *tuning* disajikan pada Tabel 9 [9].

**Tabel 9.** Perbandingan *XGBoost Default* vs *Tuned* pada *Test Set*

| Metrik | XGBoost Default | XGBoost Tuned | Perubahan |
|---|---|---|---|
| Akurasi | 0,8409 | 0,8455 | +0,0046 |
| F1 Macro | 0,7725 | 0,7728 | +0,0003 |
| Presisi Macro | 0,7643 | 0,7681 | +0,0038 |
| Recall Macro | 0,7829 | 0,7781 | -0,0048 |




**Gambar 6.** *Confusion Matrix* dan komparasi metriks *XGBoost Default vs Tuned*

*Tuning* menghasilkan peningkatan yang kecil namun konsisten pada akurasi (+0,0046) dan presisi (+0,0038). *F1 Macro* hanya naik 0,0003 poin, sementara *Recall Macro* sedikit turun (-0,0048) , mengindikasikan adanya *trade-off* antara presisi dan *recall* akibat perubahan *hyperparameter* [10]. Nilai *max_depth=4* yang terpilih lebih dangkal dari *default* (6), menunjukkan bahwa pohon yang lebih sederhana sudah mampu menangkap pola data dan sekaligus mengurangi risiko *overfitting*. Peningkatan kecil dari proses *tuning* ini bukan indikasi kegagalan, melainkan justru mengkonfirmasi bahwa model *XGBoost* default sudah sangat kokoh pada data perizinan yang bersifat tabular dan kategorik multidimensi ini [7].

### 3.6 Rekap Perbandingan Lintas Percobaan

Tabel 10 merangkum perkembangan performa *XGBoost* sebagai model terbaik di ketiga percobaan.

**Tabel 10.** Rekap Performa *XGBoost* Lintas Percobaan (*Test Set*)

| Percobaan | Skenario | Akurasi | F1 Macro |
|---|---|---|---|
| Percobaan 1 | Parameter PP No. 5/2021 | 0,7318 | 0,6227 |
| Percobaan 2 | Parameter diperluas | 0,8409 | 0,7725 |
| Percobaan 3 | Parameter diperluas + tuning | 0,8455 | 0,7728 |

Peningkatan akurasi terbesar terjadi dari Percobaan 1 ke Percobaan 2 (+10,91 poin), jauh melampaui peningkatan dari *tuning* (+0,46 poin). Pola ini menggarisbawahi bahwa pemilihan dan rekayasa fitur memiliki dampak yang lebih besar terhadap performa model dibandingkan optimasi *hyperparameter*, sebagaimana dikemukakan dalam berbagai studi komparatif algoritma *machine learning* [9], [16].

### 3.7 Pembahasan


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

Dari keseluruhan percobaan yang dilakukan, model *XGBoost* dengan parameter diperluas dan optimasi *hyperparameter* (Percobaan 3) terbukti menghasilkan performa klasifikasi terbaik dengan akurasi 84,55% dan *F1 Macro* 0,7728 pada data uji. Hasil ini merupakan puncak dari perjalanan tiga tahap percobaan yang secara sistematis membuktikan bahwa kualitas fitur dan pemilihan algoritma yang tepat memberikan dampak yang jauh lebih besar dibandingkan optimasi parameter semata.

Keunggulan *XGBoost* pada penelitian ini dapat dijelaskan melalui dua faktor utama. Pertama, mekanisme gradient boosting yang secara iteratif membangun pohon baru untuk memperbaiki kesalahan pohon sebelumnya terbukti sangat efektif pada data perizinan yang memiliki distribusi kelas tidak seimbang, kelas Rendah mendominasi 52% sementara Menengah Rendah hanya 9,73% [7]. Kedua, hasil *Grid Search* menemukan kombinasi *hyperparameter* optimal dengan *max_depth=4*, yang menghasilkan pohon yang lebih dangkal dan generalis dibandingkan default, sehingga mengurangi risiko *overfitting* pada dataset berukuran sedang (1.100 sampel) [9]. Kombinasi kedua faktor ini menghasilkan model yang tidak hanya akurat secara agregat, tetapi juga lebih seimbang dalam memprediksi kelas minoritas, *F1-score* kelas Menengah Rendah meningkat dari 0,45 pada Percobaan 1 menjadi 0,62 pada Percobaan 3.

Temuan terpenting yang dihasilkan model final ini adalah konfirmasi empiris melalui analisis *feature importance* bahwa KL/Sektor Pembina berkontribusi 48,94% terhadap prediksi *XGBoost*, hampir separuh kemampuan model bersumber dari satu variabel yang tidak diatur PP No. 5/2021. Angka ini jauh melampaui seluruh parameter resmi PP secara individual: KBLI (20,68%), Jumlah Investasi (3,98%), TKI (4,17%), dan Skala Usaha (2,94%) [4]. Pola ini mengungkap bahwa kategorisasi sektoral berdasarkan kementerian pengawas mampu menangkap dimensi risiko yang tidak dapat direpresentasikan oleh parameter finansial maupun skala usaha secara absolut. Setiap kementerian mengawasi sektor dengan karakteristik risiko operasional, lingkungan, dan sosial yang secara inheren berbeda, informasi yang secara implisit tersimpan dalam variabel KL/Sektor Pembina namun tidak tercantum dalam regulasi sebagai parameter penentu risiko resmi.

Perbandingan lintas percobaan semakin memperkuat temuan ini. Peningkatan akurasi dari penambahan fitur jauh lebih besar dibandingkan kontribusi *tuning hyperparameter*, menegaskan bahwa rekayasa fitur adalah faktor dominan dalam membangun model klasifikasi perizinan yang efektif [16]. Model-model yang tampak gagal di Percobaan 1 seperti SVM (40,91%) dan *Decision Tree* (49,55%) justru pulih signifikan di Percobaan 3 menjadi 72,73% dan 76,82% membuktikan bahwa keterbatasan performa tersebut bukan berasal dari kelemahan mendasar algoritmanya, melainkan dari keterbatasan informasi yang tersedia dalam parameter PP semata.

Dari perspektif implementasi, model *XGBoost* final ini berpotensi diintegrasikan sebagai mesin rekomendasi klasifikasi risiko dalam sistem verifikasi DPMPTSP, membantu petugas dalam mengidentifikasi tingkat risiko permohonan secara lebih cepat dan konsisten. Dari perspektif kebijakan, temuan dominansi KL/Sektor Pembina memberikan landasan empiris untuk mendorong evaluasi kelengkapan variabel dalam sistem OSS-RBA, khususnya dengan mempertimbangkan integrasi informasi sektoral berbasis kementerian pengawas sebagai dimensi penentuan risiko yang lebih eksplisit dalam regulasi perizinan berusaha di Indonesia [1], [4].

## 4. KESIMPULAN

Penelitian ini berhasil membangun dan membandingkan model klasifikasi *machine learning* untuk penentuan tingkat risiko perizinan proyek investasi menggunakan 1.100 data historis OSS DPMPTSP Kota Banjarmasin. Model terbaik yang dihasilkan adalah *XGBoost* dengan parameter diperluas dan *hyperparameter* teroptimasi, mencapai akurasi 84,55% dan *F1 Macro* 0,7728. Hasil ini secara estimatif berarti dari setiap 100 permohonan yang masuk, sekitar 85 permohonan berpotensi langsung terklasifikasi tanpa memerlukan penelaahan manual mendalam, sehingga berpotensi memangkas beban kerja verifikasi dan meminimalkan inkonsistensi penilaian antarpetugas DPMPTSP Kota Banjarmasin.

Berdasarkan tiga percobaan yang dilakukan, diperoleh tiga simpulan utama. Pertama, *XGBoost* terbukti konsisten menjadi model terbaik di seluruh skenario, dari akurasi 73,18% pada parameter PP semata hingga 84,55% setelah penambahan fitur dan tuning, didukung oleh *mekanisme* gradient boosting yang efektif menangani distribusi kelas tidak seimbang. Kedua, penambahan variabel di luar PP No. 5/2021 memberikan dampak peningkatan akurasi yang jauh lebih besar (+10,91 poin) dibandingkan optimasi *hyperparameter* (+0,46 poin), membuktikan bahwa pemilihan fitur lebih determinan dibandingkan konfigurasi algoritma dalam konteks data perizinan ini. Ketiga, analisis *feature importance* mengungkap bahwa KL/Sektor Pembina yang merupakan variabel di luar PP menjadi prediktor paling dominan dengan kontribusi 48,94% pada *XGBoost*, jauh melampaui parameter resmi seperti Jumlah Investasi (3,98%) dan Skala Usaha (2,94%), mengkonfirmasi secara empiris bahwa regulasi yang berlaku belum mengoptimalkan seluruh variabel yang relevan dalam penentuan tingkat risiko perizinan dalam melakukan usaha.


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
**Sains dan Ilmu Terapan**
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

Temuan ini memberikan landasan *data-driven* bagi pemangku kebijakan untuk mempertimbangkan evaluasi dan perluasan variabel dalam sistem OSS-RBA, khususnya dengan mengintegrasikan informasi sektoral berbasis kementerian pengawas sebagai dimensi penentuan risiko yang lebih eksplisit..

## UCAPAN TERIMAKASIH

Puji syukur kehadirat Allah SWT atas segala rahmat-Nya sehingga penelitian ini dapat rampung dengan baik. Penulis menyampaikan terima kasih yang tulus kepada kedua orang tua atas doa dan dukungan tanpa henti. Terima kasih sebesar-besarnya juga ditujukan kepada Bapak Windarsyah, M.Kom. selaku Dosen Pembimbing I dan Ibu Finki Dona Marleny, M.Kom. selaku Kaprodi S1 Informatika Universitas Muhammadiyah Banjarmasin & Dosen Pembimbing II atas segala arahan dan dedikasinya. Apresiasi setinggi-tingginya penulis sampaikan kepada seluruh dosen S1 Informatika Informatika Universitas Muhammadiyah, pihak DPMPTSP Kota Banjarmasin atas bantuan data penelitian, serta sahabat dan kerabat yang selalu memberikan dukungan moral. Semoga Allah SWT membalas semua kebaikan tersebut dengan berkah yang melimpah..

## REFERENCES

[1] Pemerintah Republik Indonesia, Peraturan Pemerintah Nomor 5 Tahun 2021 tentang Penyelenggaraan Perizinan Berusaha Berbasis Risiko. Jakarta: Sekretariat Negara, 2021.
[2] A. F. Cahya, H. Warsono, and K. Kismartini, "Advocating the Use of Risk-Based Online Single Submission (OSS) to MSMEs in Sukabumi District," *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 10, no. 1, pp. 246–251, 2024.
[3] L. Lestari and Zulkarnaini, "Pelaksanaan E-Government Melalui Online Single Submission Risk Based Approach (OSS RBA) Di DPMPTSP Kabupaten Indragiri Hulu," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 8, pp. 276–286, 2023.
[4] M. Lontoh, I. Pangkey, and A. R. Dilapanga, "Implementation of OSS-RBA Licensing at the Department of Capital Investment and One-Door Integrated Services, North Minahasa District," *International Journal of Information Technology and Education*, vol. 3, no. 1, pp. 127–139, 2023.
[5] R. Hendrawan, C. Andersen, and T. S. Dewi, "Evaluasi Program Pelayanan Perizinan Online melalui Online Single Submission (OSS) di DPMPTSP Kabupaten Sintang," *Jurnal Pemerintahan dan Kebijakan (JPK)*, 2022.
[6] Y. Amelia, "Perbandingan Metode Machine Learning untuk Mendeteksi Penyakit Jantung," *IDEALIS: Indonesia Journal Information System*, vol. 6, no. 2, pp. 220–225, 2023.
[7] R. Harahap et al., "Perbandingan Kinerja Algoritma Random Forest dan XGBoost dalam Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien," *BETRIK: Jurnal Teknologi Terapan*, 2023.
[8] R. A. Nurhafiz, F. D. Marleny, and A. A. Ningrum, "Optimasi Algoritma Random Forest dalam Mengukur Kepuasan Peserta Pelatihan Guru pada Lembaga HAFECS," *Jurnal Media Informatika (JUMIN)*, vol. 7, no. 1, pp. 302–311, Jan.–Feb. 2026.
[9] M. Erkamim, S. Suswadi, M. Subarkah, and E. Widarti, "Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 127–134, 2023.
[10] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search pada Klasifikasi Penyakit Diabetes," *SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 2, 2022.
[11] D. A. Anggoro and S. S. Mukti, "Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 6, pp. 198–207, 2021.
[12] J. M. A. S. Dachi and P. Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam*, vol. 2, no. 2, pp. 87–103, 2023.
[13] N. A. Pramudhyta and M. S. Rohman, "Perbandingan Optimasi Metode Grid Search dan Random Search dalam Algoritma XGBoost untuk Klasifikasi Stunting," *Jurnal MEDIA Informatika BUDIDARMA*, vol. 8, no. 1, pp. 19–27, 2024.
[14] S. E. Herni Yulianti, O. Soesanto, and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBoost) pada Klasifikasi Nasabah Kartu Kredit," *Journal of Mathematics Theory and Applications*, vol. 4, no. 1, pp. 21–26, 2022.
[15] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter," *Seminar Nasional Official Statistics*, 2022.


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**

**Jurnal Kolaborasi**
# Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 226-236
*ISSN 2964- 5581 (media online)*
*https://utilityprojectsolution.org/ejournal/index.php/JuKSIT* Penerbit : Cv. Utility Project Solution

[16] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 61–67, 20


Submitted: **22/05/2026**; Accepted: **21/06/2026**; Published:**22/07/2026**