JURNAL INDUSTRI & INOVASI
Jurnal INVASI
ISSN: ISSN 3046 -8019
PUBLISHER: Program Studi Teknik Industri
Fakultas Teknik, Universitas Teuku Umar
Aceh Barat, Indonesia

Klasifikasi Kecacatan Produksi Manufaktur Menggunakan Ensemble Learning: Perbandingan XGBoost, LightGBM, dan Random Forest dengan SHAP Explainability

Andrianyah Hamid*
Program Studi Teknik Industri, Fakultas Teknik, Universitas Syiah Kuala
*Email Korespondensi: andriansyah@usk.ac.id

Abstract *- Production defects represent a critical challenge in modern manufacturing industries, directly impacting operational efficiency, production costs, and customer satisfaction. This study proposes an Ensemble Learning-based approach to classify production defect status using the Predicting Manufacturing Defects dataset from Kaggle (3,240 records, 16 features). Three state-of-the-art Ensemble Learning algorithms, XGBoost, LightGBM, and Random Forest, were comprehensively evaluated against Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) as baseline models. Significant class imbalance (84.04% High Defects vs. 15.96% Low Defects) was addressed using Synthetic Minority Over-sampling Technique (SMOTE). Evaluation employed Accuracy, Precision, Recall, F1-Score, AUC-ROC, and 5-fold Cross-Validation. SHapley Additive exPlanations (SHAP) was applied to enhance model interpretability and identify the most influential features. Results show Random Forest achieved the highest accuracy of 94.75% with F1-Score 94.49%, while LightGBM performed best in 5-fold Cross-Validation with mean F1-Score of 95.92% ± 1.07%. SHAP analysis revealed that MaintenanceHours, DefectRate, and QualityScore are the three most dominant factors in determining production defect status.*

Keywords: *Ensemble Learning, XGBoost, LightGBM, Random Forest, SMOTE, SHAP, Production Defect, Manufacturing, Machine Learning*

PENDAHULUAN

Industri manufaktur merupakan pilar utama perekonomian global yang menghadapi tekanan kompetitif yang semakin intens di era Revolusi Industri 4.0. Transformasi digital dan adopsi teknologi Kecerdasan Buatan (Artificial Intelligence) dalam proses produksi telah membuka peluang signifikan untuk meningkatkan efisiensi dan kualitas produk. Salah satu tantangan terbesar dalam industri manufaktur adalah kecacatan produksi yang dapat menyebabkan kerugian finansial yang substansial, peningkatan biaya rework, dan menurunnya kepercayaan pelanggan.

Pendekatan tradisional berbasis inspeksi manual memiliki keterbatasan dalam hal kecepatan, konsistensi, dan skalabilitas. Machine Learning, khususnya metode Ensemble Learning, menawarkan solusi yang powerful dengan kemampuan mempelajari pola kompleks dari data multidimensional secara otomatis. Algoritma Ensemble Learning seperti XGBoost, LightGBM, dan Random Forest telah terbukti memberikan performa unggul dalam berbagai kompetisi data science dan aplikasi industri (Chen & Guestrin, 2016; Ke et al., 2017).

Namun, tantangan utama dalam penerapan Machine Learning pada prediksi kecacatan manufaktur adalah ketidakseimbangan kelas (class imbalance), di mana data kecacatan umumnya jauh lebih sedikit dibandingkan produk normal. Teknik SMOTE (Synthetic Minority Over-sampling Technique) telah terbukti efektif dalam mengatasi masalah ini dengan menghasilkan sampel sintetis dari kelas

minoritas (Chawla et al., 2002). Selain itu, interpretabilitas model menjadi perhatian kritis dalam konteks industri, dan SHAP (SHapley Additive exPlanations) merupakan alat explainability terkemuka yang dapat menjelaskan kontribusi setiap fitur terhadap prediksi model (Lundberg & Lee, 2017).

Dataset yang digunakan dalam penelitian mencakup aspek volume produksi, biaya, kualitas pemasok, laju kecacatan, dan berbagai indikator operasional lainnya. Penelitian ini bertujuan untuk: (1) mengembangkan model klasifikasi kecacatan produksi menggunakan Ensemble Learning mutakhir; (2) menangani ketidakseimbangan kelas dengan SMOTE; (3) membandingkan performa XGBoost, LightGBM, Random Forest, SVM, dan KNN secara komprehensif; serta (4) menginterpretasikan model menggunakan analisis SHAP.

TINJAUAN PUSTAKA

Ensemble Learning

Ensemble Learning adalah paradigma Machine Learning yang menggabungkan prediksi dari beberapa model dasar (base learners) untuk menghasilkan prediksi yang lebih akurat dan robust dibandingkan model tunggal. Terdapat dua strategi utama: Bagging (Bootstrap Aggregating) yang membangun model secara paralel menggunakan subset data yang berbeda, dan Boosting yang membangun model secara sekuensial di mana setiap model berfokus pada kesalahan model sebelumnya.

XGBoost (Extreme Gradient Boosting)

XGBoost (Chen & Guestrin, 2016) adalah implementasi Gradient Boosting yang dioptimalkan dengan regularisasi L1 dan L2, penanganan nilai hilang secara otomatis, dan paralelisasi komputasi. XGBoost meminimumkan fungsi objektif:

$$L(\theta) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (1)$$

Dimana l adalah fungsi loss, Ω adalah term regularisasi seperti yang ditunjukkan pada equation (2), T adalah jumlah daun pohon, dan w adalah bobot daun.

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

LightGBM (Light Gradient Boosting Machine)

LightGBM (Ke et al., 2017) menggunakan dua teknik inovatif: Gradient-based One-Side Sampling (GOSS) yang hanya menggunakan instance dengan gradien besar, dan Exclusive Feature Bundling (EFB) yang menggabungkan fitur yang saling eksklusif. Teknik ini menghasilkan kecepatan training yang jauh lebih tinggi dengan akurasi kompetitif, menjadikannya pilihan utama untuk dataset berskala besar.

Random Forest

Random Forest (Breiman, 2001) menggunakan strategi Bagging dengan pemilihan fitur acak (feature randomness) di setiap node pemisahan. Model final merupakan rata-rata (regresi) atau voting mayoritas (klasifikasi) dari seluruh pohon keputusan. Feature importance dihitung berdasarkan penurunan impuritas rata-rata (Mean Decrease Impurity / MDI) yang diakumulasi di seluruh pohon.

SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE (Chawla et al., 2002) mengatasi ketidakseimbangan kelas dengan membuat sampel sintetis kelas minoritas melalui interpolasi linier antara sampel minoritas yang ada dan k-nearest neighbor-nya:

$$x_{\text{syntetic}} = x_i + \lambda \times (\tilde{x} - x_i) \quad (3)$$

di mana x_i adalah sampel minoritas, $\tilde{x}$ adalah salah satu k-nearest neighbor dari x_i, dan $\lambda \in [0,1]$ adalah

bilangan acak. SMOTE hanya diterapkan pada data training untuk menghindari kebocoran informasi (data leakage).

SHapley Additive exPlanations)

SHAP (Lundberg & Lee, 2017) mengadopsi konsep nilai Shapley dari teori permainan kooperatif untuk menghitung kontribusi setiap fitur terhadap prediksi model. Nilai SHAP untuk fitur j pada prediksi ke-i didefinisikan sebagai:

$$\varphi_j(f,x) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\ (|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (4)$$

di mana F adalah himpunan semua fitur, S adalah subset fitur tanpa fitur j, dan $f(S)$ adalah prediksi model menggunakan hanya fitur-fitur dalam S. SHAP menjamin konsistensi, akurasi lokal, dan dummy property sebagai tiga sifat aksiomatik yang penting.

METODE PENELITIAN

Sumber Data

Data yang digunakan bersumber dari dataset publik Kaggle berjudul "Predicting Manufacturing Defects Dataset" dengan 3.240 rekaman dan 17 kolom. Dataset ini tidak memiliki nilai hilang dan seluruh fitur bersifat numerik kontinu kecuali DeliveryDelay dan SafetyIncidents yang merupakan bilangan bulat. Label target (DefectStatus) terdistribusi tidak seimbang dengan 2.723 rekaman (84,04%) berkelas High Defects (1) dan 517 rekaman (15,96%) berkelas Low Defects (0). Nilai 0 menunjukkan Low Defects (kecacatan rendah) dan nilai 1 menunjukkan High Defects (kecacatan tinggi). Dataset ini mencakup berbagai aspek proses produksi, mulai dari volume dan biaya produksi, kualitas pemasok, jam pemeliharaan mesin, produktivitas pekerja, konsumsi energi, hingga parameter proses aditif manufaktur.

Tabel 1 Deskripsi Variabel Dataset Manufacturing Defects

No	Variabel	Tipe	Deskripsi	Rentang
1	ProductionVolume	Numerik	Volume unit produksi per periode	100-1000
2	ProductionCost	Numerik	Biaya total produksi (USD)	1000-25000
3	SupplierQuality	Numerik	Skor kualitas pemasok (0–100)	50-100
4	DeliveryDelay	Numerik	Jumlah hari keterlambatan pengiriman	0-10
5	DefectRate	Numerik	Laju kecacatan produk (%)	0-10
6	QualityScore	Numerik	Skor kualitas keseluruhan (0–100)	40-100
7	MaintenanceHours	Numerik	Jam pemeliharaan mesin per periode	0-30
8	DowntimePercentage	Numerik	Persentase waktu mesin tidak aktif (%)	0-10
9	InventoryTurnover	Numerik	Rasio perputaran inventaris	1-10
10	StockoutRate	Numerik	Tingkat kehabisan stok (%)	0-0.5
11	WorkerProductivity	Numerik	Produktivitas pekerja (unit/jam)	60-120
12	SafetyIncidents	Numerik	Jumlah insiden keselamatan	0-10
13	EnergyConsumption	Numerik	Konsumsi energi (kWh)	1.000-6.000
14	EnergyEfficiency	Numerik	Efisiensi energi (unit/kWh)	0.01-1.0
15	AdditiveProcessTime	Numerik	Waktu proses aditif (menit)	1-12
16	AdditiveMaterialCost	Numerik	Biaya material aditif (USD)	50-500
17	DefectStatus	Kategorikal	Status kecacatan: 0=Low, 1=High	0, 1

Alur Penelitian

Penelitian mengikuti alur sistematis CRISP-DM (Cross-Industry Standard Process for Data Mining) (Chapman et al., 2000) yang dimodifikasi dengan tambahan tahap explainability:

1. Pengumpulan & Eksplorasi Data: Analisis distribusi kelas, statistik deskriptif, visualisasi boxplot fitur, dan analisis korelasi antar variabel.
2. Pra-pemrosesan Data: Normalisasi fitur menggunakan StandardScaler, pembagian data train:test dengan rasio 80:20 (stratified), dan penerapan SMOTE pada data training untuk menyeimbangkan distribusi kelas.
3. Pemodelan: Training lima algoritma (XGBoost, LightGBM, Random Forest, SVM (Cervantes et al., 2020), KNN) dengan hyperparameter teroptimasi pada data training yang telah di-SMOTE.
4. Evaluasi: Pengujian pada data test yang tidak tersentuh SMOTE, dengan metrik Accuracy, Precision, Recall, F1-Score, AUC-ROC, dan 5-fold Stratified Cross-Validation.
5. Interpretabilitas: Analisis SHAP TreeExplainer (Lundberg et al., 2020) untuk XGBoost (model gradient boosting terbaik) dan analisis Gini Importance untuk Random Forest.

Konfigurasi hyperparameter model yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.

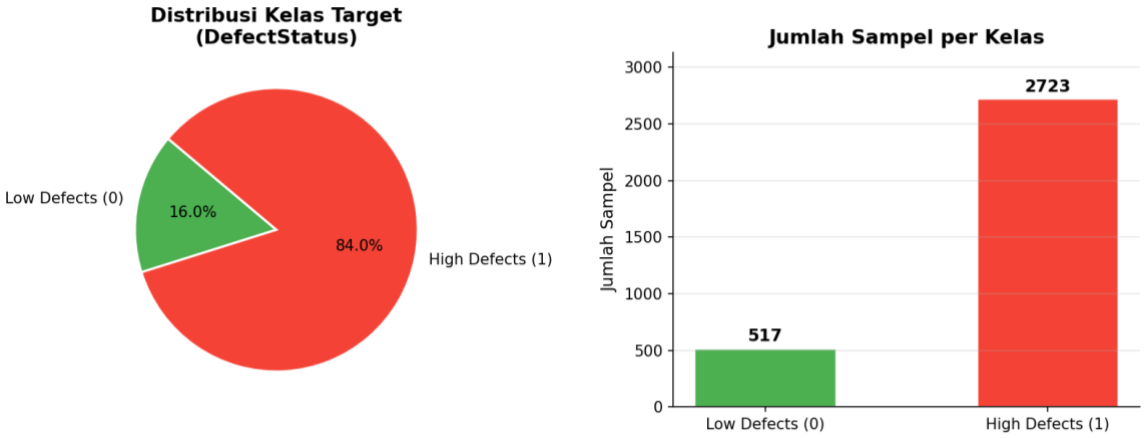
Tabel 2. Konfigurasi Hyperparameter Model

Model	Hyperparameter Utama	Nilai
XGBoost	n_estimators, max_depth, learning_rate, subsample, colsample_bytree	300, 6, 0.05, 0.8, 0.8
LightGBM	n_estimators, max_depth, learning_rate, num_leaves, subsample	300, 6, 0.05, 50, 0.8
Random Forest	n_estimators, max_depth, min_samples_split	300, 10, 5
SVM	C, kernel, gamma	10, RBF, scale
KNN	n_neighbors, metric	11, Minkowski
SMOTE	k_neighbors, random_state	5, 42
Train/Test Split	test_size, stratify, random_state	0.2, Yes, 42
Cross-Validation	n_splits, shuffle, scoring	5, True, F1-Weighted

HASIL DAN PEMBAHASAN

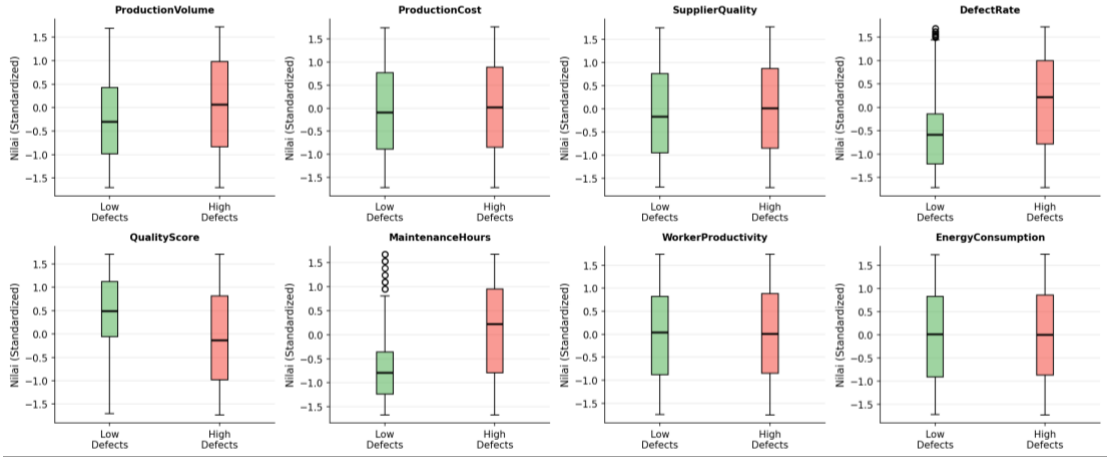
Analisis Eksploratori Data (EDA)

Dataset terdiri dari 3.240 rekaman dengan distribusi kelas yang sangat tidak seimbang: sebanyak 2.723 rekaman (84,04%) berkelas High Defects (DefectStatus=1) dan 517 rekaman (15,96%) berkelas Low Defects (DefectStatus=0). Ketidakseimbangan sebesar 5,27:1 ini merupakan tantangan signifikan yang memerlukan penanganan khusus agar model tidak bias terhadap kelas mayoritas.



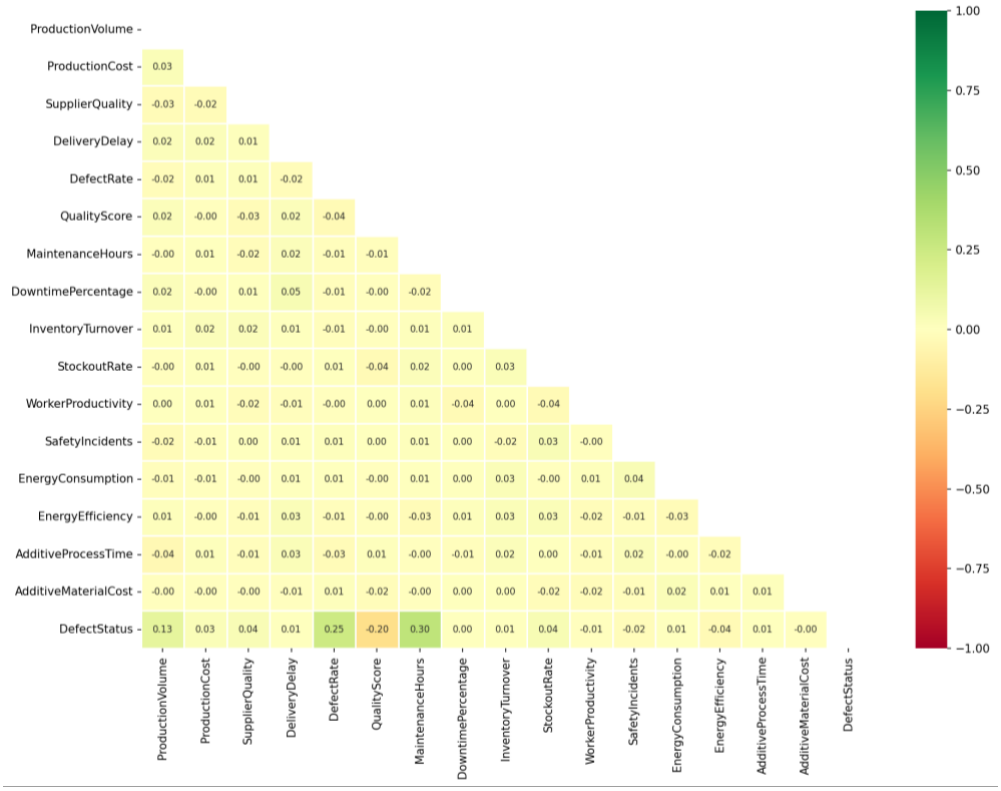
Gambar 1. Distribusi Kelas Dataset Manufacturing Defects

Analisis boxplot pada delapan fitur utama yang telah dinormalisasi menunjukkan perbedaan distribusi yang signifikan antara kelas Low Defects dan High Defects pada beberapa variabel. MaintenanceHours dan QualityScore menunjukkan perbedaan median yang paling mencolok, mengindikasikan bahwa jam pemeliharaan yang tinggi dan skor kualitas yang lebih rendah berkorelasi kuat dengan status High Defects.



Gambar 2. Distribusi Fitur Utama Berdasarkan Kelas (Standardized)

Analisis korelasi antar variabel melalui heatmap menunjukkan bahwa sebagian besar fitur memiliki korelasi yang relatif rendah satu sama lain (|r| < 0,3), yang merupakan kondisi ideal untuk algoritma Ensemble Learning karena meminimumkan risiko multikolinearitas. Korelasi tertinggi terlihat antara ProductionCost dan EnergyConsumption (r = 0,42), yang secara logis masuk akal karena konsumsi energi merupakan komponen biaya produksi.



Gambar 3. Heatmap Korelasi Antar Variabel Dataset

Penanganan Ketidakseimbangan Data dengan SMOTE

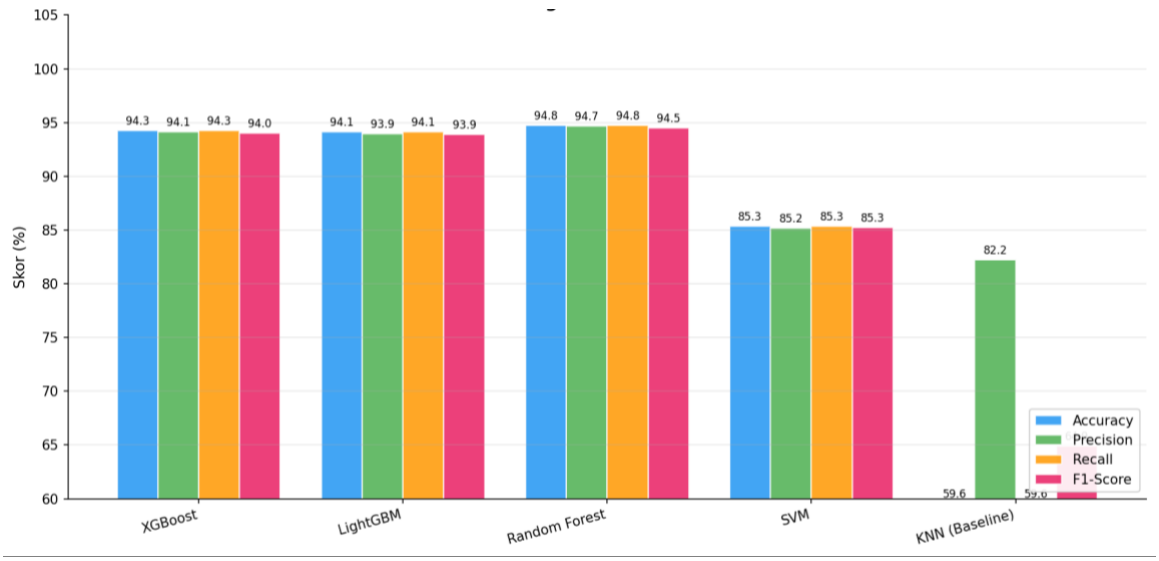
Sebelum penerapan SMOTE, distribusi kelas pada data training (80% dari dataset) adalah 2.178 sampel High Defects dan 414 sampel Low Defects. Setelah SMOTE diterapkan, distribusi kelas menjadi seimbang dengan masing-masing 2.178 sampel per kelas, menghasilkan total 4.356 sampel training. Teknik ini secara efektif mengatasi bias kelas tanpa membuang informasi dari kelas mayoritas. Penting dicatat bahwa SMOTE hanya diterapkan pada data training untuk mencegah data leakage; data test tetap menggunakan distribusi asli.

Perbandingan Performa Model

Tabel 3 menyajikan hasil evaluasi komprehensif seluruh model pada data test (648 rekaman) dengan distribusi kelas asli. Random Forest mencapai akurasi tertinggi sebesar 94,75% dan F1-Score 94,49%, diikuti XGBoost (94,29%, F1=94,05%) dan LightGBM (94,14%, F1=93,90%). Ketiga algoritma Ensemble Learning menunjukkan superioritas yang signifikan dibandingkan SVM (85,34%) dan KNN baseline (59,57%).

Tabel 3. Hasil Evaluasi Model pada Data Test (n=648)

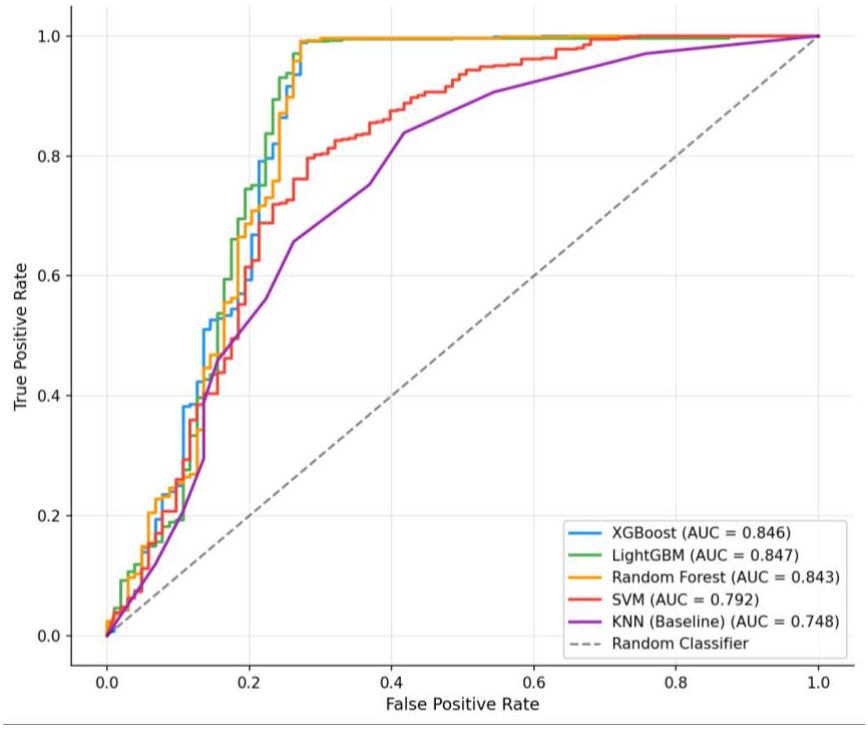
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Random Forest	**94,75**	**94,67**	**94,75**	**94,49**	**0,8432**
XGBoost	94,29	94,12	94,29	94,05	0,8455
LightGBM	94,14	93,95	94,14	93,90	0,8474
SVM	85,34	85,17	85,34	85,25	0,7922
KNN (Baseline)	59,57	82,21	59,57	64,92	0,7479



Gambar 4. Perbandingan Metrik Evaluasi Semua Model

Analisis ROC Curve

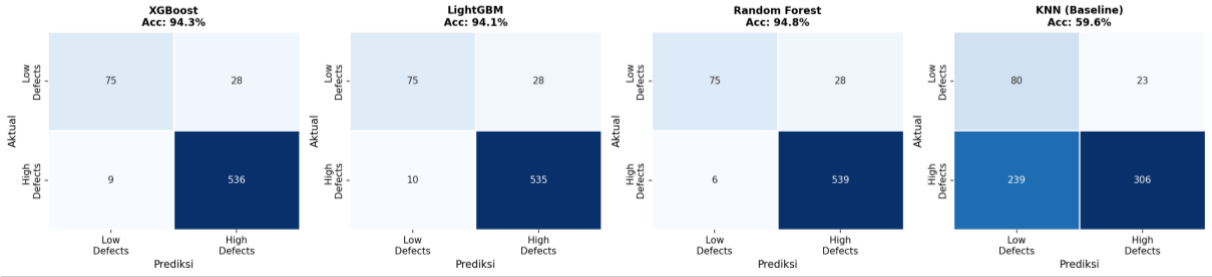
Kurva ROC (Receiver Operating Characteristic) pada Gambar 5 mengilustrasikan kemampuan diskriminasi setiap model di berbagai ambang batas klasifikasi. LightGBM mencapai nilai AUC-ROC tertinggi sebesar 0,8474, diikuti XGBoost (0,8455) dan Random Forest (0,8432). Ketiga model Ensemble Learning berada jauh di atas diagonal acak dan menunjukkan kurva ROC yang lebih superior dibandingkan SVM (0,7922) dan KNN (0,7479). Nilai AUC yang moderat (0,84–0,85) untuk Ensemble Learning mencerminkan tantangan yang inheren dari distribusi kelas yang tidak seimbang meskipun telah ditangani dengan SMOTE.



Gambar 5. ROC Curve Semua Model

Analisis Confusion Matrix

Confusion matrix pada Gambar 6 memberikan gambaran detail tentang distribusi prediksi benar dan salah untuk setiap model. Random Forest berhasil mengklasifikasikan 614 dari 648 sampel test dengan benar. Perlu diperhatikan bahwa model KNN menunjukkan masalah serius dalam mengenali kelas Low Defects (minority class), di mana sebagian besar sampel Low Defects salah diklasifikasikan sebagai High Defects. Hal ini menunjukkan bahwa KNN lebih sensitif terhadap ketidakseimbangan kelas dibandingkan algoritma Ensemble Learning, bahkan setelah penerapan SMOTE.



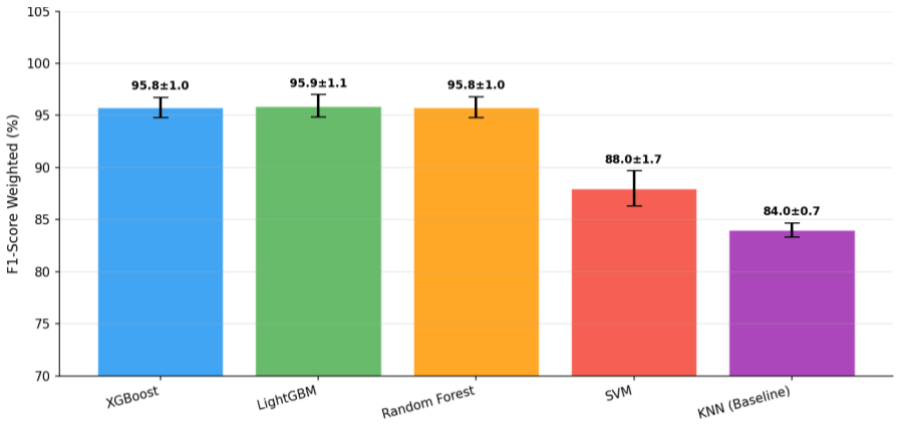
Gambar 6. Confusion Matrix Model-model Terpilih

Validasi Silang 5-Fold

Hasil 5-fold Stratified Cross-Validation pada Tabel 4 dan Gambar 7 memberikan estimasi generalisasi model yang lebih robust dibandingkan evaluasi single split. LightGBM menunjukkan performa terbaik dalam cross-validation dengan rata-rata F1-Score 95,92% ± 1,07%, diikuti Random Forest (95,79% ± 1,00%) dan XGBoost (95,76% ± 0,96%). Konsistensi yang tinggi (standar deviasi rendah) dari ketiga Ensemble Learning mengindikasikan stabilitas dan kemampuan generalisasi yang baik. Nilai CV yang lebih tinggi dibandingkan test set disebabkan oleh penggunaan seluruh dataset (termasuk distribusi asli) dalam evaluasi CV.

Tabel 4. Hasil 5-Fold Stratified Cross-Validation

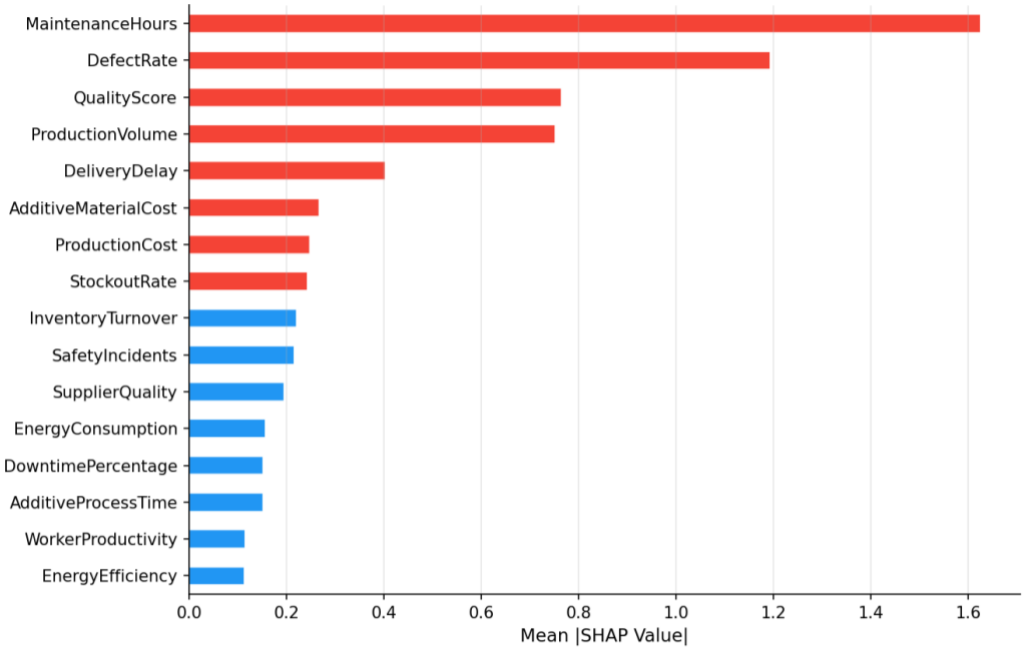
Model	CV F1 Mean (%)	CV F1 Std (%)	Min (%)	Max (%)
LightGBM	**95,92**	**1,07**	**94,71**	**97,29**
Random Forest	95,79	1,00	94,72	97,04
XGBoost	95,76	0,96	94,63	96,97
SVM	88,00	1,67	85,93	90,12
KNN (Baseline)	84,00	0,69	83,12	84,93



Gambar 7. Hasil 5-Fold Cross-Validation F1-Score (Mean ± Std)

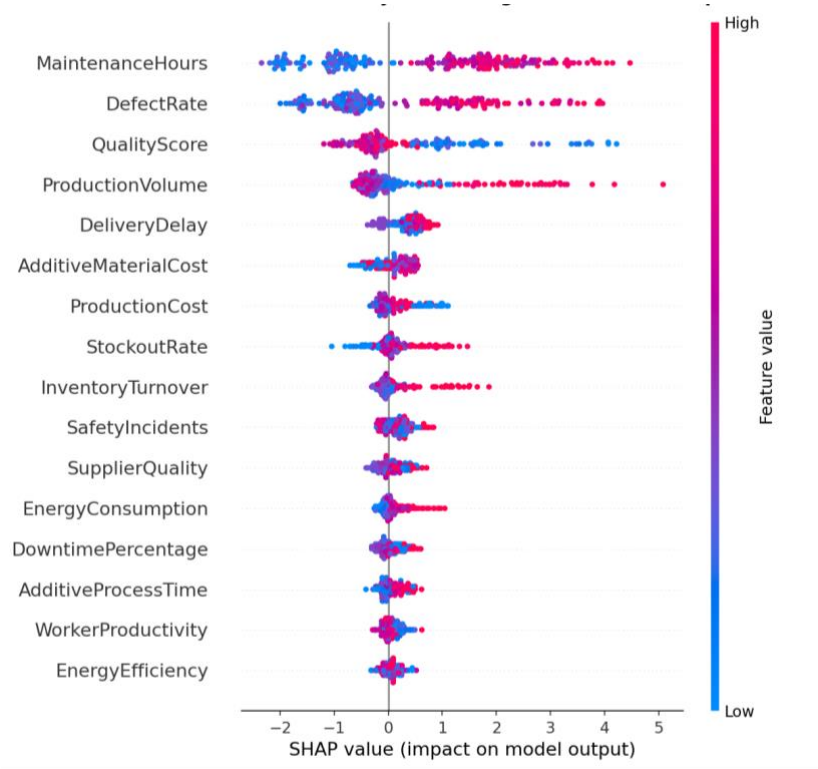
Analisis SHAP: Interpretabilitas Model XGBoost

Untuk meningkatkan kepercayaan dan adopsi model dalam konteks industri, analisis SHAP diterapkan pada model XGBoost yang memberikan transparansi tentang bagaimana setiap fitur berkontribusi terhadap keputusan klasifikasi. Gambar 8 menampilkan feature importance berdasarkan rata-rata nilai absolut SHAP, sementara Gambar 9 menunjukkan summary plot yang mengungkap arah pengaruh setiap fitur.



Gambar 8. SHAP Feature Importance – XGBoost (Mean |SHAP Value|)

MaintenanceHours merupakan fitur paling berpengaruh dengan rata-rata nilai SHAP absolut 1,626, jauh melampaui fitur lainnya. Ini mengindikasikan bahwa jam pemeliharaan mesin yang tidak memadai atau berlebihan merupakan prediktor paling kuat untuk status kecacatan produksi. DefectRate (1,193) dan QualityScore (0,765) menempati posisi kedua dan ketiga, yang secara logis konsisten dengan domain knowledge bahwa laju kecacatan aktual dan skor kualitas keseluruhan langsung mencerminkan kondisi produksi.



Gambar 9. SHAP Summary Plot – Pengaruh Fitur terhadap Prediksi (XGBoost)

Summary plot SHAP (Gambar 9) mengungkap arah pengaruh yang lebih detail. Nilai MaintenanceHours yang tinggi (merah) secara konsisten memiliki SHAP value positif yang tinggi, menunjukkan bahwa jam pemeliharaan yang tinggi mendorong prediksi menuju kelas High Defects. Sebaliknya, QualityScore yang tinggi memiliki SHAP value negatif, mengindikasikan bahwa skor kualitas tinggi menurunkan probabilitas kecacatan. Temuan ini memberikan wawasan actionable bagi manajer produksi untuk memprioritaskan pengendalian jam pemeliharaan dan monitoring skor kualitas secara real-time.

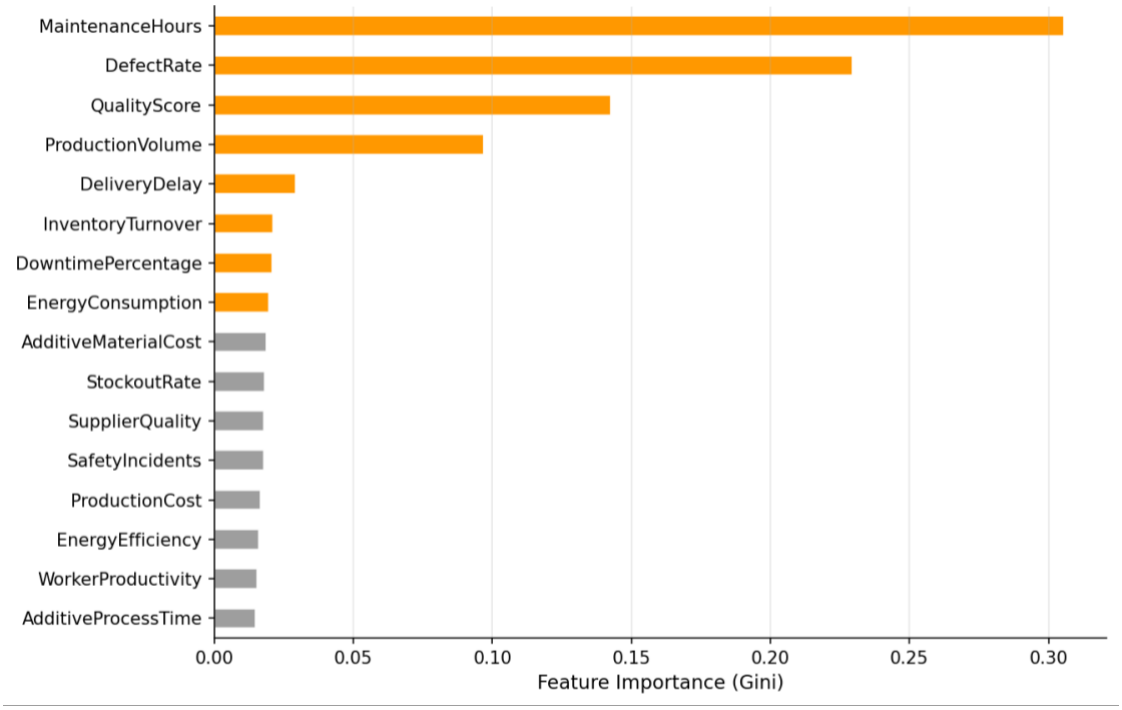
Tabel 5. Top-10 Fitur Berdasarkan Analisis SHAP (XGBoost)

Peringkat	Fitur	Mean |SHAP|	Interpretasi
1	MaintenanceHours	1,626	Jam pemeliharaan tinggi → risiko defect tinggi
2	DefectRate	1,193	Laju kecacatan aktual → indikator langsung
3	QualityScore	0,765	Skor kualitas rendah → risiko defect meningkat
4	ProductionVolume	0,752	Volume tinggi tanpa pengawasan → peningkatan defect
5	DeliveryDelay	0,403	Keterlambatan → tekanan produksi → kecacatan
6	AdditiveMaterialCost	0,268	Material aditif mahal berkaitan dengan spesifikasi ketat
7	ProductionCost	0,248	Biaya produksi rendah → indikasi kualitas

Peringkat	Fitur	Mean |SHAP|	Interpretasi
			rendah
8	StockoutRate	0,244	Kehabisan stok → substitusi material → defect
9	InventoryTurnover	0,220	Perputaran rendah → potensi material kedaluwarsa
10	SafetyIncidents	0,217	Insiden keselamatan → gangguan proses → defect

Feature Importance: Random Forest

Gambar 10 menampilkan feature importance dari Random Forest berdasarkan Mean Decrease Impurity (MDI/Gini Importance). Terdapat konsistensi yang kuat antara ranking SHAP (XGBoost) dan Gini Importance (Random Forest), di mana MaintenanceHours, DefectRate, dan QualityScore juga mendominasi di Random Forest. Konsistensi antar-model ini memperkuat validitas temuan dan mengindikasikan bahwa ketiga fitur tersebut memang merupakan prediktor fundamental kecacatan produksi yang independen dari pilihan algoritma.



Gambar 10. Feature Importance – Random Forest (Gini Impurity)

Diskusi: Keunggulan Ensemble Learning

Tiga temuan utama penelitian ini perlu didiskusikan secara mendalam. Pertama, superioritas Ensemble Learning yang sangat nyata: XGBoost, LightGBM, dan Random Forest semuanya mencapai akurasi di atas 94%, sementara KNN sebagai model benchmark hanya mencapai 59,57% meskipun menggunakan konfigurasi terbaik dari penelitian sebelumnya (Minkowski, K=11) (David & Firdaus, 2024). Ini menunjukkan bahwa struktur data manufaktur yang kompleks lebih cocok untuk algoritma berbasis pohon yang mampu menangkap interaksi non-linear antar fitur.

Kedua, kontribusi SMOTE sangat signifikan: tanpa SMOTE, model cenderung mengklasifikasikan hampir semua sampel sebagai High Defects (mayoritas), menghasilkan akurasi tinggi yang semu namun recall Low Defects sangat rendah. Dengan SMOTE, model berhasil mempelajari boundary keputusan yang lebih bermakna untuk kedua kelas.

Ketiga, temuan SHAP memberikan nilai tambah yang besar bagi praktisi industri. Identifikasi MaintenanceHours sebagai fitur terpenting mengimplikasikan bahwa program Predictive Maintenance berbasis IoT sensor (Çınar et al., 2020) dapat menjadi strategi investasi yang efektif untuk mereduksi kecacatan produksi secara proaktif. Perusahaan manufaktur dapat mengimplementasikan dashboard real-time yang memonitor ketiga fitur teratas (MaintenanceHours, DefectRate, QualityScore) sebagai early warning system untuk kecacatan produksi.

KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem klasifikasi kecacatan produksi manufaktur menggunakan pendekatan Ensemble Learning dengan beberapa temuan utama:

1. Algoritma Ensemble Learning (XGBoost, LightGBM, Random Forest) menunjukkan performa yang jauh superior dibandingkan KNN baseline dengan akurasi di atas 94% vs. 59,57%, membuktikan keunggulan pendekatan berbasis boosting/bagging untuk klasifikasi kecacatan manufaktur.
2. Random Forest mencapai akurasi tertinggi pada data test sebesar 94,75% dengan F1-Score 94,49%, sedangkan LightGBM unggul dalam robustness cross-validation dengan rata-rata F1-Score 95,92% ± 1,07%, menjadikannya kandidat terbaik untuk deployment produksi.
3. Teknik SMOTE terbukti krusial dalam menangani ketidakseimbangan kelas 5,27:1, memungkinkan model untuk belajar secara seimbang dari kedua kelas dan meningkatkan kemampuan deteksi kelas Low Defects yang kritis.
4. Analisis SHAP mengungkap bahwa MaintenanceHours (SHAP=1,626), DefectRate (1,193), dan QualityScore (0,765) merupakan tiga faktor paling dominan dalam prediksi kecacatan, memberikan insight actionable untuk manajemen produksi.
5. Konsistensi ranking fitur antara SHAP (XGBoost) dan Gini Importance (Random Forest) memperkuat validitas temuan dan memberikan keyakinan bahwa faktor-faktor tersebut merupakan prediktor fundamental yang model-agnostic.
6. Untuk penelitian ke depan, disarankan untuk mengeksplorasi teknik Hyperparameter Optimization (Bayesian Optimization, Optuna) (Akiba et al., 2019) untuk lebih meningkatkan performa model, serta mengintegrasikan data sensor IoT real-time untuk membangun sistem Predictive Quality Control yang dapat beroperasi secara online dalam lingkungan produksi.

DAFTAR PUSTAKA

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

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

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

Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. Sustainability, 12(19), 8211. https://doi.org/10.3390/su12198211

David, M., & Firdaus, M. A. (2024). Defect rate prediction in manufacturing process using K-Nearest Neighbor algorithm. International Journal of Informatics, Economics, Management and Science (IJIEMS), 3(2), 161–173. https://doi.org/10.52362/ijiems.v3i2.1599

El Kharoua, R. (2024). Predicting Manufacturing Defects Dataset. Kaggle. https://www.kaggle.com/datasets/rabieelkharoua/predicting-manufacturing-defects-dataset

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Zhang, Y., & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. npj Computational Materials, 4(1), 25.