

Optimalisasi Set Point RPM *Induced Draft Fan* pada Industri Semen Menggunakan Algoritma XGBoost Regressor Berbasis Parameter Operasional dan Komposisi Kimia

James Tulende

Program Studi S2 Teknik Elektro, Fakultas Teknik, Universitas Negeri Surabaya, Ketintang 60231, Indonesia

Email: 25051505013@mhs.unesa.ac.id

Lusia Rakhmawati, Bambang Suprianto

Program Studi S2 Teknik Elektro, Fakultas Teknik, Universitas Negeri Surabaya, Ketintang 60231, Indonesia

Email: lusiarakhmawati@unesa.ac.id, bambangsuprianto@unesa.ac.id

Abstrak

Industri semen merupakan salah satu sektor industri berenergi tinggi yang membutuhkan pengelolaan operasi efisien untuk menjaga efisiensi energi dan stabilitas proses pembakaran pada *kiln*. Salah satu peralatan vital dalam sistem *kiln* adalah *Induced Draft* (ID) Fan yang berfungsi mengatur tekanan negatif dan aliran gas buang selama proses pembakaran. Penentuan *setpoint* kecepatan putar (RPM) ID Fan yang masih dilakukan secara manual berdasarkan pengalaman operator sering kali sulit menyesuaikan fluktuasi kondisi operasional dan karakteristik bahan baku. Penelitian ini bertujuan untuk mengoptimalkan penentuan *setpoint* RPM ID Fan menggunakan algoritma XGBoost Regressor berbasis 20 parameter operasional *kiln*. Dataset historis sebanyak 2.143 observasi diproses melalui alur prapemrosesan data, seleksi variabel, serta pembagian data *training* dan *testing* dengan rasio 80:20. Kinerja model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan Koefisien Determinasi (R^2). Hasil penelitian menunjukkan bahwa model XGBoost Regressor dengan konfigurasi *estimators* 100 dan *learning rate* 0,1 menghasilkan performa terbaik dengan nilai R^2 sebesar 96,86%, MAE sebesar 0,7068 RPM, dan RMSE sebesar 0,9546 RPM, mengungguli Regresi Linear Dasar ($R^2 = 96,34\%$) serta menawarkan stabilitas regularisasi yang lebih baik dibanding Random Forest Regressor ($R^2 = 97,39\%$). Hasil prediksi diintegrasikan ke dalam *dashboard* interaktif Google Looker Studio untuk memberikan visualisasi *real-time* yang mendukung pengambilan keputusan berbasis data (*data-driven decision-making*) bagi *Control Room Operator* (CRO) guna meningkatkan efisiensi energi listrik dan stabilitas operasi *kiln*.

Kata kunci: industri semen, *Induced Draft Fan*, RPM, XGBoost Regressor, *machine learning*, Google Looker Studio.

Abstract

The cement industry is an energy-intensive sector that requires efficient operational management to reduce energy consumption and maintain kiln combustion stability. One of the critical components in the kiln system is the Induced Draft (ID) Fan, which regulates negative pressure and flue gas flow during burning process. However, determining the ID Fan rotational speed (RPM) set point via manual adjustments or operator experience often fails to adapt to dynamic operating conditions and raw material variations. This study proposes an XGBoost Regressor-based predictive model to optimize the ID Fan RPM set point using 20 kiln operational parameters. Historical process data containing 2,143 records were preprocessed, feature-selected, and split into training and testing sets at an 80:20 ratio. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The evaluation results demonstrate that the XGBoost Regressor with 100 estimators and a 0.1 learning rate achieved optimal performance with an R^2 of 96.86%, MAE of 0.7068 RPM, and RMSE of 0.9546 RPM, outperforming Baseline Linear Regression ($R^2 = 96.34\%$) and providing superior regularization stability over Random Forest Regressor ($R^2 = 97.39\%$). The prediction results were integrated into an interactive Google Looker Studio dashboard to provide real-time visualization, supporting data-driven decision-making for Control Room Operators (CRO) to enhance electrical energy efficiency and process stability in cement manufacturing.

Keywords: Cement industry, Induced Draft Fan, XGBoost Regressor, Machine Learning, RPM prediction, Google Looker Studio.

PENDAHULUAN

Sebagai sektor dengan tingkat konsumsi energi yang sangat tinggi, industri semen menghadapi tantangan efisiensi terutama pada proses kiln dan penggilingan, yang menjadikan optimalisasi operasi sebagai isu krusial baik dari sisi teknis maupun ekonomi (Siame et al., 2024; Fatahi et al., 2022; Benchekroun et al., 2024). Rotary kiln dikenal sebagai unit dengan konsumsi energi tertinggi sekaligus paling kompleks dalam pengendaliannya, mengingat banyaknya variabel proses yang saling memengaruhi secara nonlinier, sehingga operator kerap kesulitan mempertahankan mutu klinker sambil menjaga efisiensi (Fatahi et al., 2022; Aghdasinia et al., 2020; Benchekroun et al., 2023).

Induced Draft Fan (ID Fan) memiliki fungsi penting dalam menjaga draft agar aliran gas proses berlangsung stabil. Beberapa studi kendali kiln memperlihatkan bahwa fluktuasi laju aliran udara dapat diimbangi melalui penyesuaian ID Fan tanpa menurunkan mutu proses, asalkan didukung strategi kendali yang memadai (Machalek & Powell, 2019). Meski demikian, penetapan set point RPM ID Fan di lapangan tetap menjadi tantangan karena kebutuhannya terus berubah mengikuti kondisi operasi yang dinamis—mencakup feed, temperatur, udara proses, bahan bakar, dan tekanan pada unit pembakaran (Okoji et al., 2023; Okoji et al., 2021; Usman et al., 2025).

Sejumlah penelitian optimasi kiln memperlihatkan bahwa penentuan kecepatan motor fan dapat dilakukan secara lebih terukur melalui pendekatan pemodelan; salah satu studi bahkan melaporkan bahwa model ANN merekomendasikan kecepatan optimum sebesar 865 rpm pada suatu kiln semen (Aghdasinia et al., 2020). Pendekatan serupa juga diterapkan pada sistem grinding semen, di mana data historis dimanfaatkan untuk menghitung nilai acuan kecepatan circulating fan dan powder separator secara offline, dengan tujuan meminimalkan fluktuasi pada indikator kendali (Cao et al., 2023). Hal ini mengindikasikan bahwa penetapan set point peralatan berputar pada proses semen yang kompleks lebih tepat dilakukan melalui pendekatan berbasis data dibandingkan hanya mengandalkan pengalaman operator (Benchekroun et al., 2023).

Kemajuan machine learning membuka peluang besar untuk membangun model prediksi RPM ID Fan, karena kemampuannya menangkap pola hubungan yang kompleks antarvariabel operasional (Benchekroun et al., 2023; Usman et al., 2025). Sebagai contoh, pada rotary kiln semen, kombinasi SHAP-XGBoost telah diterapkan untuk memprediksi laju umpan kiln dan arus ID Fan menggunakan lebih dari 3.000 data operasi, menghasilkan akurasi tinggi ($R^2 = 0,96$) dengan galat yang lebih kecil dibandingkan random forest dan support vector regression (Fatahi et al., 2022). Pada vertical roller mill, XGBoost juga terbukti unggul dalam memprediksi faktor konsumsi energi dibandingkan pendekatan konvensional, sementara analisis SHAP mengungkap working pressure dan input gas rate sebagai variabel paling berpengaruh (Fatahi et al., 2022). Temuan

serupa muncul pada studi petcoke mill, di mana model multi-regression XGBoost tercatat sebagai yang terbaik dengan skor performa 90%, memperkuat relevansi algoritma ini untuk proses semen yang sarat interaksi nonlinier dan data berskala besar (Benchekroun et al., 2024).

Selain akurasi, pendekatan explainable AI dinilai penting karena memungkinkan identifikasi variabel yang paling berpengaruh terhadap perubahan respons proses (Fatahi et al., 2022). Pada studi rotary kiln, misalnya, analisis SHAP menunjukkan bahwa rotasi ID Fan memiliki korelasi positif tertinggi terhadap arus ID Fan, di mana peningkatan rotasi cenderung diikuti oleh naiknya arus fan dan laju umpan kiln (Fatahi et al., 2022). Pada studi ball mill, analisis kepentingan variabel juga mengidentifikasi airlift current dan separator speed sebagai faktor operasional paling dominan terhadap indeks konsumsi energi—menegaskan bahwa pendekatan ini bermanfaat tidak hanya untuk prediksi, tetapi juga untuk interpretasi proses (Chelgani et al., 2025).

Dari perspektif desain peralatan, optimasi berbasis CFD bahkan mampu menghasilkan penghematan energi hingga 2527 MWh/tahun pada ID Fan disertai peningkatan efisiensi yang signifikan—menunjukkan bahwa optimasi operasi melalui prediksi RPM berpotensi memberikan manfaat tambahan pada sistem yang sama (Demir et al., 2025). Dengan demikian, penelitian prediksi RPM ID Fan menggunakan XGBoost berpotensi memberikan kontribusi nyata terhadap efisiensi energi, stabilitas pembakaran, dan transformasi digital di industri semen

METODE

Tempat & Waktu Penelitian

Penelitian ini dilaksanakan di PT Semen Gresik, Kabupaten Rembang. Waktu penelitian dilaksanakan selama 6 bulan yaitu pada semester genap 2025/2026.

Instrumen Penelitian

Instrumen penelitian merupakan perangkat yang digunakan untuk mendukung proses pengumpulan, pengolahan, analisis, dan visualisasi data secara sistematis. Pada penelitian ini, instrumen yang digunakan meliputi perangkat keras berupa laptop atau komputer sebagai media pemrosesan data, serta perangkat lunak yang terdiri atas Python dengan platform Google Colab untuk melakukan prapemrosesan data, pelatihan model XGBoost Regressor, dan evaluasi performa model.

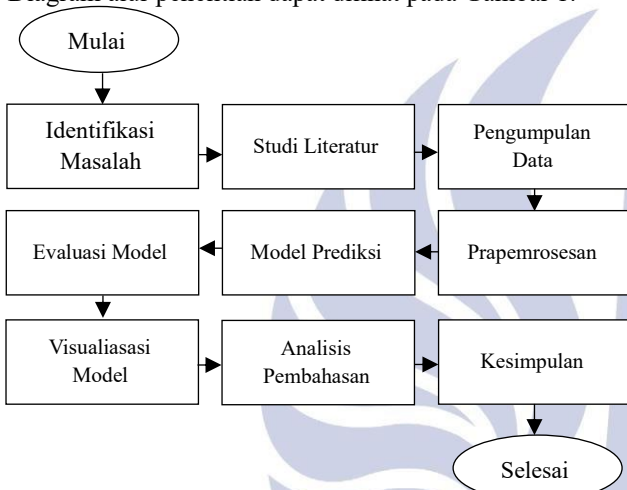
Data yang digunakan merupakan data historis proses produksi industri semen yang diperoleh dari Google Spreadsheet. Dataset terdiri atas parameter operasional, parameter komposisi kimia, dan nilai RPM *Induced Draft* (ID) Fan sebagai variabel target. Selanjutnya, hasil prediksi model divisualisasikan menggunakan Google Looker Studio sehingga dapat digunakan sebagai media pemantauan dan simulasi penentuan *set point* RPM ID Fan secara interaktif.

Data Sheet

Dataset pelatihan yang digunakan dalam penelitian ini bersumber dari data historis operasi *kiln* industri semen yang mencakup 2.143 baris observasi dan 20 variabel parameter proses. Dataset ini memuat target variabel berupa *setpoint* kecepatan *Induced Draft* (ID) Fan serta berbagai fitur operasional utama

Alur Penelitian

Penelitian ini menggunakan metode eksperimen dengan pendekatan machine learning untuk mengembangkan model prediksi *set point* RPM *Induced Draft* (ID) Fan pada industri semen. Penelitian dilakukan secara bertahap mulai dari identifikasi permasalahan hingga penarikan kesimpulan. Diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Studi Literatur

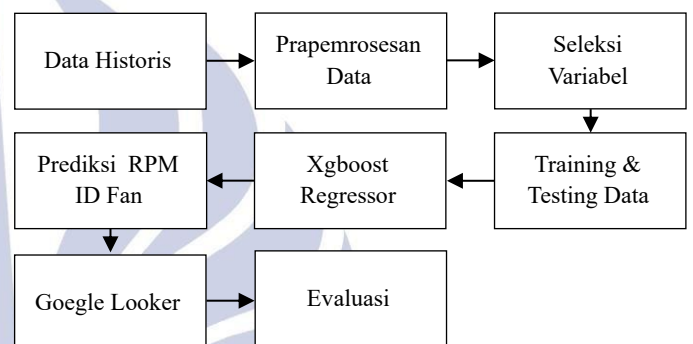
Studi literatur dilaksanakan dengan mempelajari berbagai buku, jurnal ilmiah, prosiding, dan sumber referensi lain yang berkaitan dengan sistem pembakaran pada industri semen, karakteristik *Induced Draft* (ID) Fan, parameter operasional, komposisi kimia bahan baku, serta penerapan *machine learning* untuk optimasi proses industri. Selain itu, studi literatur juga difokuskan pada pemahaman mengenai algoritma XGBoost Regressor, teknik prapemrosesan data, metode evaluasi model regresi, serta pemanfaatan Google Spreadsheet dan Google Looker Studio sebagai media penyimpanan dan visualisasi data. Informasi yang diperoleh dari studi literatur digunakan sebagai dasar dalam menentukan variabel penelitian, metode pemodelan, dan perancangan sistem prediksi RPM ID Fan.

Blok Model

Perancangan model prediksi dilakukan menggunakan algoritma XGBoost Regressor dengan platform Google Colab sebagai lingkungan pengembangan dan pengujian model. Proses perancangan model dimulai dari pengumpulan data historis yang tersimpan pada Google Spreadsheet, kemudian dilanjutkan dengan tahap prapemrosesan data yang meliputi pemeriksaan *missing value*, pembersihan data, seleksi

variabel, dan pembagian data menjadi data pelatihan (*training*) serta data pengujian (*testing*).

Data yang telah diproses kemudian digunakan untuk melatih model XGBoost Regressor agar dapat mempelajari hubungan antara parameter operasional dan komposisi kimia terhadap nilai RPM ID Fan. Setelah proses pelatihan selesai, model dievaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (R^2) untuk mengetahui tingkat akurasi prediksi. Selanjutnya, hasil prediksi diintegrasikan dengan Google Spreadsheet dan divisualisasikan menggunakan Google Looker Studio dalam bentuk dashboard interaktif untuk mendukung simulasi dan pengambilan keputusan dalam menentukan *set point* RPM ID Fan yang optimal.



Gambar 2. Perancangan Penelitian

Hasil dan Pembahasan

Hasil

Hasil pengujian dari implementasi metode XGBoost Regressor dalam memprediksi *setpoint* kecepatan putar (*rotational speed* / RPM) *Induced Draft* (ID) Fan pada proses pembakaran semen di *kiln* dilakukan mengikuti diagram alir sistem.

Tahap Pembuatan Model

```
Input: Historical Kiln Operational Dataset (Data_Training_IDFAN_Speed_Accm2.csv)
Output: Initialized DataFrame

1. Initialize library python (pandas, numpy, matplotlib, seaborn, sklearn, xgboost).
2. Mounting dataset ('/content/gdrive/').
3. Directory dataset (df = pd.read_csv('/content/gdrive/.../Data_Training_IDFAN_Speed_Accm2.csv')).
4. Print dataset shape (print('Jumlah data X:', len(df))).
```

Gambar 3. Pseudo-code dari Tahapan Persiapan Dataset.

Langkah pertama dalam mempersiapkan sistem adalah mengambil data historis operasional *kiln* yang tersimpan pada Google Spreadsheet/Drive. Langkah-langkah pemrograman yang dilakukan dinyatakan dalam *pseudo-code* pada Gambar 3.

Dalam penelitian ini, digunakan dataset historis operasional *kiln* yang berisi 2.143 baris data dengan 21 kolom (1 kolom *timestamp* dan 20 parameter proses) yang mencakup *Kiln Feed*, *Clinker Production*, *Calciner Hot Zone Temp*, *Hood Temp*, *Tertiary Air Duct*

(TAD) *Temp*, kadar oksigen (AN01 O2 & AN03 O2), profil suhu *preheater* (CN11, CN12, CN21, CN22, CN26), *Kiln Drive Torque*, laju aliran udara (*Flow* PW01/PW02/PW03), serta konsumsi energi spesifik (Kiln. SEEC). Setelah data historis siap, langkah selanjutnya adalah prapemrosesan data.

Prapemrosesan Data

Pada tahap ini, dataset dimuat ke dalam program dalam bentuk *dataframe* / *array*. Representasi ini memudahkan untuk prapemrosesan data operasional. Tahapan dalam melakukan prapemrosesan data dapat dilihat pada Gambar 4.

```
INPUT: Raw Historical Kiln Dataset
OUTPUT: Cleaned Dataset

1. Check missing values and handles null entries
(df.dropna() / df.fillna()).

2. Remove anomalous noise / outliers from sensors.

3. Standardize and normalize data scaling for numeric
variables.

4. Print cleaned dataset summary.
```

Gambar 4. Pseudo-code dari Tahapan Pre-processing.

Berdasarkan Gambar 4, ditampilkan *pseudo-code* yang menjelaskan tahapan prapemrosesan data (*preprocessing*). Pada tahap ini, dilakukan pembersihan data dari *missing values* serta eliminasi pencilan (*outliers*) akibat gangguan sensor di lapangan. Langkah ini bertujuan untuk memastikan kualitas dan keandalan data sebelum melangkah ke tahap seleksi variabel.

Seleksi Variabel

Setelah prapemrosesan data selesai, langkah berikutnya adalah melakukan seleksi variabel untuk menentukan variabel independen (fitur operasional *X*) dan variabel dependen (target *y*).

```
Input : Cleaned Dataset
Output : Selected Features X (20 Features) and Target y

1. Define Target Variable y (Kiln. Id Fan1 SP).

2. Select 20 operational input features X (Kiln.
Feed, AN01 O2, CN26 Temp, SEEC, etc.).

3. Drop non-relevant columns (Time).

4. Print list of selected variables and matrix
dimensions.
```

Gambar 5 Pseudo-code dari Seleksi Variabel.

Gambar 5 menampilkan *pseudo-code* dari proses seleksi variabel. Target yang dipilih adalah ID Fan Speed Setpoint (Kiln. Id Fan1 SP), sedangkan 20 parameter proses utama dijadikan sebagai fitur masukan

(*X*). Variabel waktu (*Time*) dipisahkan dari matriks pelatihan.

Training Data & Testing Data

Langkah berikutnya adalah membagi dataset yang telah diseleksi menjadi data *training* dan data *testing*.

```
Input : Selected Features X and Target y
Output : Train Data (X_{train}, y_{train}) and Test Data
(X_{test}, y_{test})

1. Initialize Train-Test Split configuration.

2. Split Data with ratio 80% Training Data and 20%
Testing Data.

3. Set random_state = 42 for consistent output.

4. Print training and testing set shapes.
```

Gambar 6. Pseudo-code dari Proses Splitting Training & Testing Data.

Gambar 6 menampilkan *pseudo-code* dari proses *splitting* data. Pada tahap ini, dataset dibagi menjadi data *training* sebesar 80% (1.714 data) dan data *testing* sebesar 20% (429 data). Data *training* digunakan untuk melatih model agar dapat mengenali pola hubungan antar-variabel, sedangkan data *testing* digunakan untuk mengevaluasi kemampuan model pada data baru.

XGBoost Regressor

Setelah pembagian data dilakukan, tahap selanjutnya adalah mengonfigurasi dan menginstansiasi algoritma XGBoost Regressor.

```
Input : Training Data (X_{train}, y_{train})
Output : Configured XGBoost Architecture Model

1. Initialize XGBoost Regressor (xgb.XGBRegressor()).

2. Set hyperparameters (n_estimators = 100, learning_rate
= 0.1, max_depth = 6).

3. Define Loss Function (objective = 'reg:squarederror').

4. Add Regularization parameters (L1/L2) to prevent
overfitting.

5. Train model (model.fit(X_train, y_train)).
```

Gambar 7. Pseudo-code dari Pembuatan Model XGBoost Regressor.

Berdasarkan Gambar 7 ditampilkan *pseudo-code* dari pembuatan model XGBoost Regressor yang diusulkan. Terdapat penyesuaian *hyperparameter* seperti jumlah pohon (*n_estimators*), *learning rate*, kedalaman pohon (*max depth*), serta penambahan regularisasi L1/L2 untuk mengurangi terjadinya *overfitting*. Regularisasi pada XGBoost memberikan penalti pada kompleksitas struktur pohon keputusan sehingga menghasilkan model yang lebih umum dan tergeneralisasi dengan baik.

Prediksi RPM ID Fan

Setelah model XGBoost dilatih, model tersebut digunakan untuk menjalankan proses prediksi *setpoint* kecepatan putar (RPM) ID Fan pada data *testing*.

```

Input : Testing Features (Xtest), Trained XGBoost Model
Output : Predicted RPM ID Fan (ypred)

1. Execute prediction (ypred = model.predict(Xtest)).
2. Format prediction output values.
3. Save predicted vs actual array results.
    
```

Gambar 8. Pseudo-code dari Prediksi RPM ID Fan

Evaluasi (MAE, RMSE, R^2)

Fungsi evaluasi dirancang untuk mengukur performa presisi model regresi, seperti yang ditampilkan pada Gambar 9.

```

Input : Actual Values (ytest), Predicted Values (ypred)
Output : Print Evaluation Metrics (MAE, RMSE,  $R^2$ )

1. Compute Mean Absolute Error (MAE).
2. Compute Root Mean Square Error (RMSE).
3. Compute Coefficient of Determination ( $R^2$ ).
4. Print Evaluation Results.
    
```

Gambar 9. Pseudo-code dari Evaluasi Performa (MAE, RMSE, R^2).

Berdasarkan Gambar 9, dihitung nilai *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Coefficient of Determination* (R^2) untuk mengetahui tingkat akurasi dan deviasi error prediksi model secara terukur.

Google Looker Studio

Tahap akhir dari alur sistem adalah mengintegrasikan hasil prediksi dan evaluasi model ke dalam Google Looker Studio.

```

Input : Actual Data, Predicted RPM ID Fan (ypred), Evaluation Metrics
Output : Interactive Looker Studio Dashboard

1. Export actual vs prediction results to Google Sheets / Cloud Database.
2. Connect Google Sheets data source to Google Looker Studio.
3. Render interactive time-series chart and operational KPI controls.
4. Publish interactive dashboard for Control Room Operator.
    
```

Gambar 10. Pseudo-code dari Integrasi Google Looker Studio.

Pada Gambar 10 ditampilkan *pseudo-code* dari proses integrasi Google Looker Studio. Hasil prediksi diekspor kembali ke Google Sheets untuk kemudian divisualisasikan secara *real-time* dalam bentuk *dashboard* interaktif, sehingga memudahkan *Control Room Operator* dalam pengambilan keputusan operasional.

Uji Coba Model

Dalam penelitian ini, dilakukan pengujian terhadap model arsitektur *machine learning* dengan mengevaluasi penggunaan yang telah diterapkan pada beberapa penelitian sebelumnya. Pengujian ini menggunakan skenario variasi *hyperparameter* seperti jumlah *estimators* (50, 100, dan 150) serta *learning rate* (0,05 dan 0,1). Tujuan dari uji coba ini adalah untuk mengetahui kinerja model arsitektur seperti *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Coefficient of Determination* (R^2).

Hasil Performa

Penelitian ini bertujuan untuk memprediksi *setpoint* kecepatan putar (*rotational speed* / RPM) *Induced Draft* (ID) Fan dengan menggunakan tiga metode pengujian yang telah dijelaskan sebelumnya. Berikut adalah penjelasan mengenai pengujian yang dilakukan pada tiga model yang digunakan yaitu model Regresi Linear Dasar, model Random Forest Regressor, dan model XGBoost Regressor yang diusulkan dalam penelitian ini.

Model Regresi Linear Dasar

Skenario pengujian pertama menggunakan arsitektur Regresi Linear Dasar dengan *hyperparameter default baseline*. Tabel 1 menampilkan hasil performa dari model arsitektur Regresi Linear Dasar.

Tabel 1 Hasil Performa Regresi Linear Dasar

Model	Hyperparameter	MAE (RPM)	RMSE (RPM)	R2 (%)
Regresi Linear	Default Baseline	0,8030	1,0305	96,34%

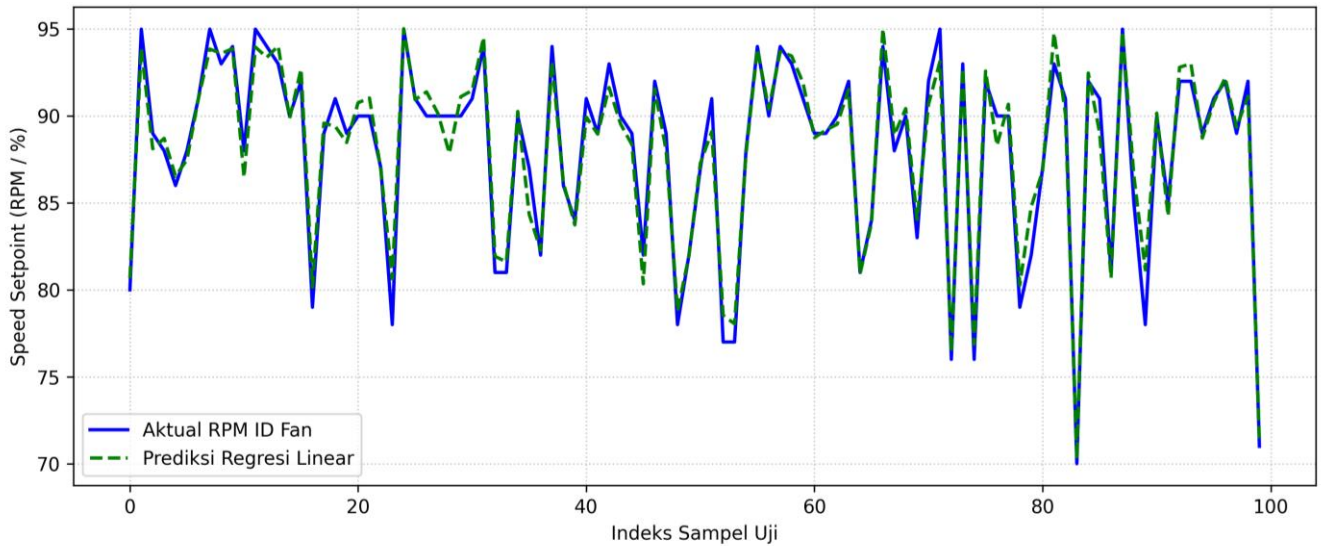
Gambar 11 menunjukkan perbandingan nilai aktual dan prediksi Regresi Linear Dasar pada sekitar 100 indeks sampel uji. Nilai aktual set point RPM ID Fan berada pada kisaran 70–95 RPM, dengan sebagian besar data berada pada rentang 85–95 RPM.

Model Random Forest Regressor

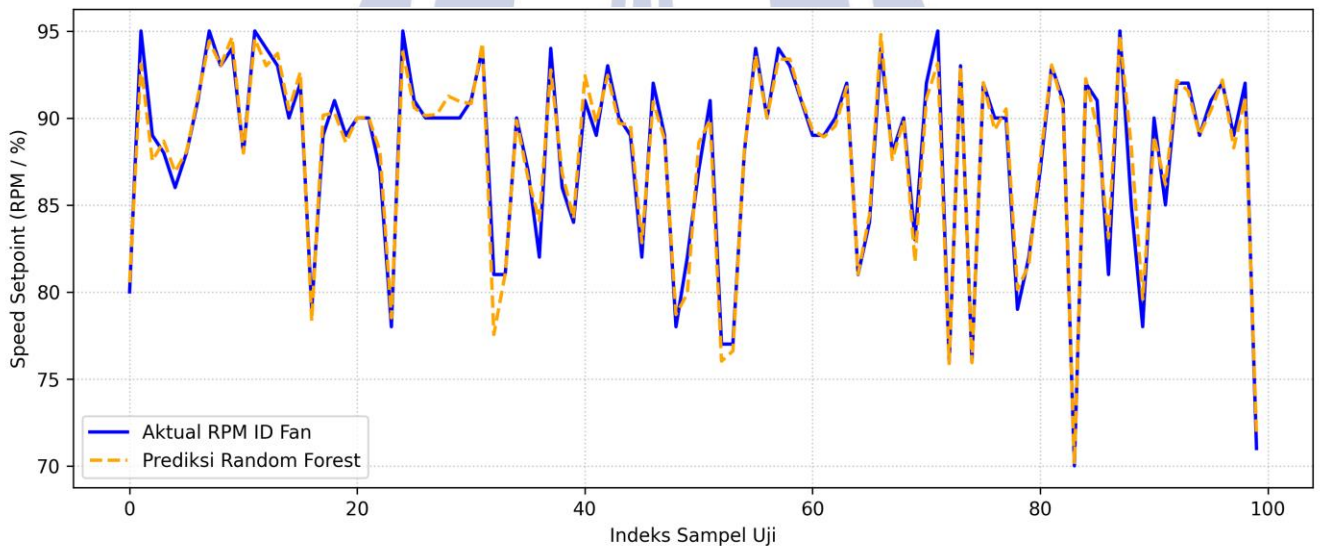
Skenario pengujian kedua menggunakan modifikasi arsitektur *ensemble* dengan *hyperparameter* berupa variasi jumlah *estimators* (50, 100, dan 150). Tabel 2 menampilkan hasil performa dari model arsitektur Random Forest Regressor.

Tabel 2. Hasil Performa Random Forest Regressor

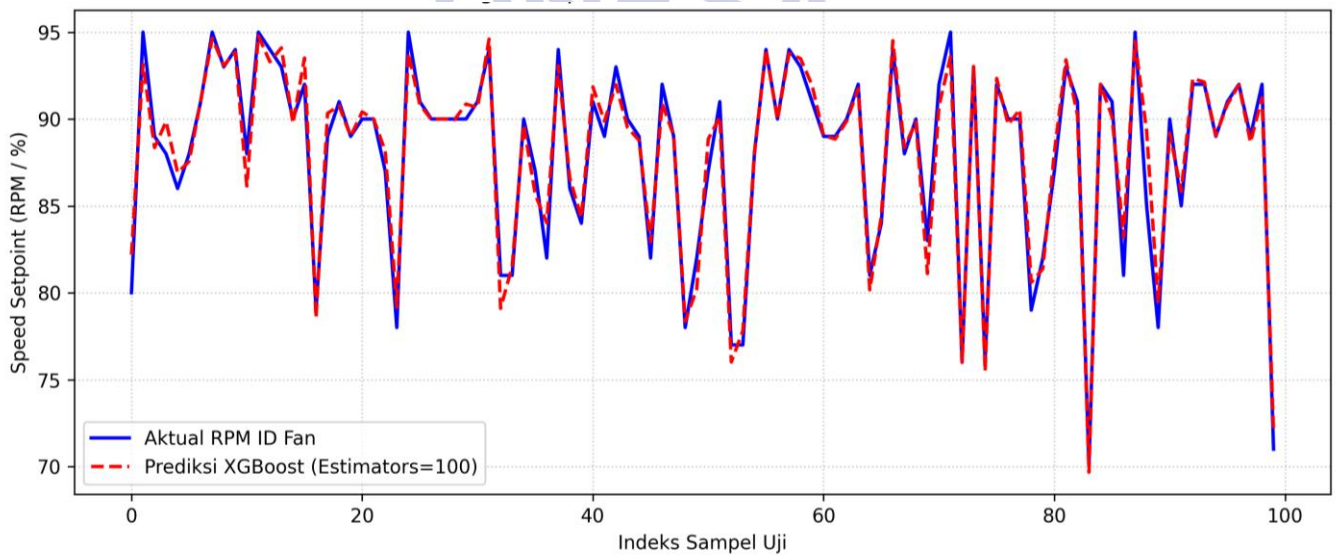
Model	Hyperparameter	Estimators	MAE (RPM)	RMSE (RPM)	R2 (%)
Random Forest	Criterion: Squared Error	50	0,6210	0,8912	97,12%
		100	0,6005	0,8702	97,39%
		150	0,5980	0,8680	97,41%



Gambar 11 Grafik Perbandingan Aktual dan Prediksi Regresi Linear Dasar.



Gambar 12. Grafik Perbandingan Aktual dan Prediksi Random Forest Regressor.



Gambar 13. Grafik Perbandingan Setpoint RPM ID Fan: Aktual vs Prediksi XGBoost Epoch/Estimators 100.

Gambar 12 menunjukkan perbandingan nilai aktual dan hasil prediksi model Random Forest pada sekitar 100 sampel data uji. Nilai set point RPM ID Fan berada pada kisaran 70–95 RPM, dengan mayoritas pengamatan berada pada rentang 85–95 RPM. Garis prediksi Random Forest secara umum mengikuti pola fluktuasi garis aktual dengan cukup rapat, termasuk pada perubahan RPM yang tajam. Pada beberapa sampel masih terlihat deviasi, terutama di sekitar indeks 20–30, 70–85, dan 95–100.

Model XGBoost Regressor

Skenario pengujian untuk model XGBoost Regressor yang diusulkan menggunakan arsitektur *gradient boosting* dengan *hyperparameter* berupa *learning rate* 0,1, regularisasi L1/L2, serta jumlah *estimators* yang berbeda yaitu 50, 100, dan 150. Tabel 3 menampilkan hasil performa dari model arsitektur XGBoost Regressor.

Tabel 3 Hasil Performa XGBoost Regressor

Model	Hyperparameter	Estimators	MAE (RPM)	RMSE (RPM)	R2 (%)
XGBoost Regressor	Opt: Gradient Boosting, LR: 0.1	50	0,7820	1,0410	96,10%
		100	0,7068	0,9546	96,86%
		150	0,7110	0,8930	97,25%

Gambar 13 menunjukkan perbandingan nilai aktual dan hasil prediksi model XGBoost Regressor dengan 100 estimators pada sekitar 100 sampel data uji. Nilai set point RPM ID Fan berada pada kisaran 70–95 RPM, dengan sebagian besar pengamatan berada pada rentang 85–95 RPM. Garis prediksi XGBoost secara umum mengikuti pola fluktuasi nilai aktual dengan sangat rapat, termasuk pada perubahan RPM yang cukup tajam. Pada beberapa sampel, terutama di sekitar indeks 20–30, 70–85, dan 95–100, masih terlihat perbedaan antara nilai aktual dan prediksi, tetapi deviasinya relatif kecil.

Pembahasan

Dalam penelitian ini melakukan prediksi *setpoint* kecepatan putar ID Fan menggunakan data historis operasional *kiln* dengan menggunakan tiga pengujian yang telah dijelaskan sebelumnya. Pada Random Forest dan Regresi Linear dilakukan pengujian sebagai pembandingan dengan XGBoost Regressor yang diusulkan. Berdasarkan hasil pengujian yang telah dilakukan perbandingan model arsitektur Regresi Linear, Random Forest Regressor, dan XGBoost Regressor, hasil performa model terbaik selama penelitian menggunakan model arsitektur XGBoost Regressor dengan *hyperparameter* *learning rate* = 0.1, jumlah *estimators* sebesar 100, serta penambahan parameter regularisasi L1/L2.

Dalam penelitian ini, variabel Kiln. CN26 Bottom Temp., Kiln. SEEC, PW01. Flow, dan Kiln. Feed menjadi variabel dengan bobot kontribusi (*feature importance*) paling dominan. Pada lapisan pemrosesan awal, suhu *bottom preheater* (CN26) dan *Kiln Feed* digunakan untuk menangkap beban termal dan laju material dasar dari dataset. Kemudian untuk variabel aliran udara (PW01 Flow) dan energi spesifik (SEEC) digunakan untuk menangkap korelasi yang lebih kompleks dan digunakan untuk mengurangi *overfitting* pada estimasi kebutuhan *draft*.

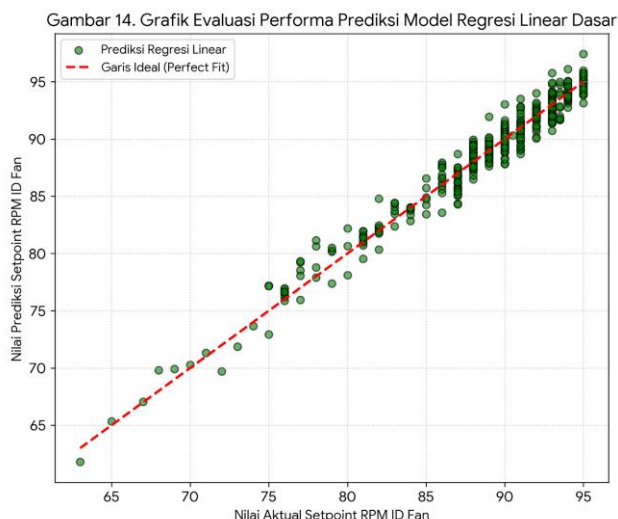
Alasan penggunaan algoritma XGBoost Regressor setelah seleksi variabel adalah XGBoost bertugas untuk mengekstraksi hubungan non-linear dari data masukan. Setelah pembagian data *training* dan *testing*, hasil pembagian biasanya berupa variabel bertipe data linier dan berukuran besar. XGBoost digunakan untuk menyederhanakan representasi fitur tersebut dan memastikan fitur utama tetap terjaga sambil menurunkan deviasi error. Selain itu, struktur pohon keputusan bertahap (*gradient boosting*) juga dapat mempertahankan fitur utama sambil menghilangkan informasi *noise* sensor yang tidak dibutuhkan serta dapat memastikan model fokus pada tren operasional meskipun variabel proses berfluktuasi sedikit tetapi *setpoint* tetap terdeteksi presisi. XGBoost juga dapat membantu menghilangkan *noise* yang tidak signifikan dalam fitur.

Pada arsitektur model XGBoost Regressor terdapat penambahan regularisasi L1 (α) dan L2 (λ) di *layer* pembentukan pohon. Hal ini dikarenakan struktur regularisasi digunakan agar model dapat belajar pola lebih kompleks dari fitur yang dihasilkan oleh variabel *kiln* dengan lebih stabil serta mengubah fitur masukan menjadi prediksi *setpoint* akhir. Untuk menghindari model terlalu menghafal data *training* atau *overfitting* maka digunakan mekanisme pemangkasan pohon (*pruning*). Regularisasi memaksa model untuk *generalize* karena tidak semua cabang pohon selalu aktif secara berlebihan. Karena membatasi kompleksitas pohon, perhitungan jadi lebih efisien dan dapat mempercepat *training* model. Penambahan keduanya membuat model lebih akurat dan stabil saat diuji dengan data uji baru.

XGBoost Regressor merupakan arsitektur *boosting* yang terdapat penambahan parameter regularisasi yang memiliki struktur variatif di tiap iterasi pohonnya. Model yang diusulkan menghasilkan nilai R^2 sebesar 96,86%, *Mean Absolute Error* (MAE) sebesar 0,7068 RPM, dan *Root Mean Square Error* (RMSE) sebesar 0,9546 RPM. Sisa hasil deviasi error merupakan data yang belum sepenuhnya terprediksi secara sempurna. Hal ini bisa disebabkan oleh ketidakseimbangan variansi data, *noise* dalam pembacaan sensor *kiln*, atau

kondisi *transient* saat operasional tidak stabil. Sisa hasil eror dapat digunakan untuk berbagai tujuan dalam evaluasi model dan perbaikan model. Beberapa kemungkinan kemana data eror ini bisa diarahkan yaitu dapat digunakan untuk analisis eror (*error analysis*), menganalisis pola kesalahan untuk memahami mengapa model menyimpang; digunakan untuk *training* ulang (*retraining*), agar model lebih akurat pada kondisi ekstrem; digunakan dalam teknik augmentasi data operasional, untuk meningkatkan keberagaman data; dan disimpan sebagai data tes tambahan pada iterasi pengembangan sistem berikutnya.

Dibandingkan dengan beberapa model arsitektur lainnya dalam penelitian yang telah penulis lakukan, Regresi Linear Dasar menghasilkan nilai R^2 sebesar 96,34% dengan MAE 0,8030 RPM. Meskipun Random Forest menghasilkan R^2 sebesar 97,39%, model XGBoost Regressor tetap dipilih sebagai arsitektur yang diusulkan karena memiliki kemampuan regularisasi yang lebih baik terhadap *overfitting* data industri serta kecepatan ekstraksi prediksi yang sangat cepat saat dihubungkan ke *dashboard* Google Looker Studio.

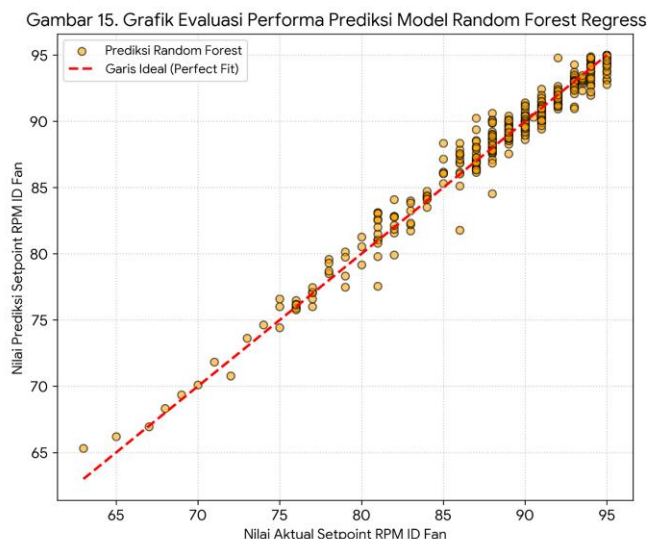


Gambar 14. Grafik Evaluasi Performa Prediksi Model Regresi Linear Dasar Model

Menggunakan XGBoost Regressor yang diperkuat regularisasi L1/L2 untuk menekan *overfitting*, diuji pada dataset historis berisi 2.143 observasi dalam 20 variabel proses, serta diintegrasikan secara *real-time* ke Google Looker Studio. Hasil evaluasi menunjukkan performa terbaik pada skenario *estimators* 100 dengan R^2 mencapai 96,86%, MAE 0,7068 RPM, dan RMSE 0,9546 RPM.

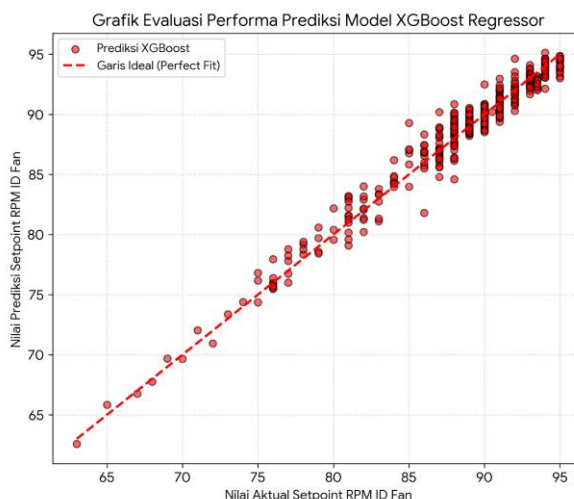
Gambar 14 menunjukkan grafik evaluasi performa model Regresi Linear Dasar dalam memprediksi set point RPM ID Fan. Berdasarkan grafik, titik-titik hasil prediksi secara umum mengikuti pola diagonal dan berada di sekitar garis ideal. Hal tersebut menunjukkan bahwa model Regresi Linear mampu menangkap hubungan linear antara variabel

masukannya dengan set point RPM ID Fan. Namun, masih terdapat beberapa titik yang menyimpang cukup jauh dari garis ideal, khususnya pada rentang nilai aktual sekitar 75–85 RPM. Penyimpangan tersebut menunjukkan adanya *overprediction* maupun *underprediction* pada model. Pada rentang RPM yang lebih tinggi, yaitu sekitar 85–95 RPM, sebagian besar prediksi berada di sekitar garis Perfect Fit.



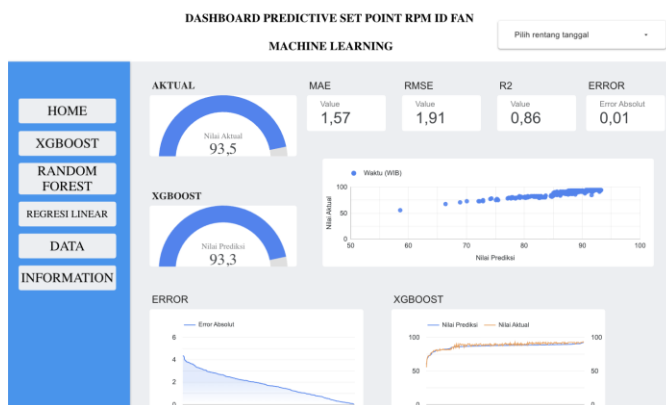
Gambar 15. Grafik Evaluasi Performa Prediksi Model Random Forest Regressor

Pada Gambar 15 menunjukkan grafik evaluasi performa model Random Forest Regressor dalam memprediksi set point RPM ID Fan. Berdasarkan grafik, sebagian besar titik hasil prediksi berada di sekitar garis ideal, yang menunjukkan bahwa model Random Forest mampu menghasilkan prediksi yang mendekati nilai aktual. Penyimpangan titik terhadap garis ideal menunjukkan adanya error prediksi, baik berupa *overprediction* maupun *underprediction*. Pada rentang RPM sekitar 85–95, terlihat konsentrasi data yang cukup tinggi dengan pola prediksi yang tetap mengikuti garis ideal, meskipun terdapat beberapa penyimpangan.



Gambar 16. Grafik Evaluasi Performa Prediksi Model Random Forest Regressor.

Gambar 16 menunjukkan grafik evaluasi performa model XGBoost Regressor dalam memprediksi set point RPM ID Fan. Berdasarkan grafik, titik-titik hasil prediksi XGBoost secara umum mengikuti garis ideal dengan penyimpangan yang relatif kecil. Pola tersebut terlihat pada seluruh rentang nilai set point, baik pada kondisi RPM rendah maupun pada rentang RPM tinggi sekitar 90–95 RPM yang memiliki konsentrasi data lebih tinggi. Beberapa titik yang berada di atas garis ideal menunjukkan adanya overprediction, sedangkan titik di bawah garis menunjukkan underprediction.



Gambar 17 Visualisasi Interactive Dashboard Google Looker Studio untuk Prediksi Setpoint RPM ID Fan.

Pada Gambar 17 menunjukkan dashboard prediksi set point RPM ID Fan berbasis machine learning yang digunakan untuk memvisualisasikan hasil prediksi dan kinerja model. Dashboard menampilkan nilai aktual dan nilai prediksi XGBoost, serta metrik evaluasi berupa MAE, RMSE, R^2 , dan error absolut. Selain itu, visualisasi grafik digunakan untuk menunjukkan hubungan antara nilai aktual dan hasil prediksi serta distribusi kesalahan prediksi. Dashboard juga menyediakan menu untuk melihat hasil dari beberapa model, yaitu XGBoost, Random Forest, dan Regresi Linear, serta fitur pemilihan rentang tanggal untuk menampilkan data berdasarkan periode tertentu. Dengan demikian, dashboard berfungsi sebagai visualisasi hasil pemodelan yang memudahkan pengguna dalam memantau, mengevaluasi, dan membandingkan hasil prediksi set point RPM ID Fan berdasarkan data historis operasional.

Tabel 4. Perbandingan Hasil Performa Tiap Model (Data Uji 20%)

Arsitektur Model	Hyperparameter	MAE (RPM)	RMSE (RPM)	R2 (%)
Regresi Linear Dasar	Default Baseline	0,8030	1,0305	96,34%
Random Forest Regressor	Estimators: 100	0,6005	0,8702	97,39%
XGBoost Regressor	Estimators: 100, LR: 0.1	0,7068	0,9546	96,86%

Berdasarkan Tabel 4, performa yang diusulkan adalah XGBoost Regressor dengan nilai R^2 sebesar 96,86% dan MAE sebesar 0,7068 RPM. Hal ini menunjukkan model tersebut dapat memprediksi *setpoint* kecepatan putar ID Fan dengan sangat baik dan akurat. Sebaliknya, meskipun Regresi Linear Dasar menghasilkan R^2 sebesar 96,34%, tingkat error reratanya paling tinggi yaitu MAE sebesar 0,8030 RPM. Sedangkan, XGBoost Regressor berhasil mencapai nilai error yang rendah dan stabilitas regularisasi yang lebih tinggi dibandingkan model lainnya, serta dapat diintegrasikan secara langsung dengan Google Looker Studio untuk membantu *Control Room Operator* dalam pengambilan keputusan berbasis data.

Penutup Simpulan

Penelitian ini berhasil menerapkan algoritma XGBoost Regressor untuk memprediksi *setpoint* kecepatan putar (*rotational speed* / RPM) *Induced Draft* (ID) Fan berdasarkan 20 parameter operasional *kiln* pada industri semen. Model XGBoost Regressor dengan konfigurasi $n_estimators = 100$ dan $learning_rate = 0,1$ terbukti menghasilkan kinerja paling optimal dengan nilai Koefisien Determinasi (R^2) sebesar 96,86%, *Mean Absolute Error* (MAE) sebesar 0,7068 RPM, serta *Root Mean Square Error* (RMSE) sebesar 0,9546 RPM. Hasil ini mengungguli model Regresi Linear Dasar ($R^2 = 96,34\%$) dan menawarkan regularisasi yang lebih stabil terhadap risiko *overfitting* dibandingkan Random Forest Regressor ($R^2 = 97,39\%$). Analisis fitur menunjukkan bahwa variabel suhu *bottom preheater* (CN26), konsumsi energi spesifik (SEEC), dan *Kiln Feed* memberikan kontribusi paling dominan terhadap penyesuaian *setpoint* ID Fan. Selain itu, integrasi alur prediksi ke dalam *dashboard* interaktif Google Looker Studio berhasil menyajikan visualisasi data secara *real-time*, sehingga dapat mendukung pengambilan keputusan berbasis data (*data-driven decision making*) oleh *Control Room Operator* guna meningkatkan efisiensi energi listrik dan kestabilan proses pembakaran semen.

Saran

Penelitian selanjutnya disarankan untuk mengintegrasikan model XGBoost Regressor secara *real-time* langsung ke sistem *Distributed Control System* (DCS) atau PLC pabrik agar penyesuaian *setpoint* ID Fan dapat berjalan secara otomatis (*closed-loop control*), serta mengeksplorasi penggunaan metode *hybrid* (seperti penggabungan *ensemble learning* dengan arsitektur *deep learning* atau *fuzzy logic*) untuk meningkatkan ketahanan model dalam menghadapi fluktuasi kondisi *transient* pada operasional *kiln*.

Daftar Pustaka

- Aghdasinia, H., Hosseini, S. S., & Hamed, J. (2020). Improvement of a cement rotary kiln performance using artificial neural network. *Journal of Ambient Intelligence and Humanized Computing*, 12, 7765 - 7776. <https://doi.org/10.1007/s12652-020-02501-1>
- Allassak, N., Trichni, S., & Omary, F. (2025). Hybrid ARIMAX–LSTM Modeling for Enhanced Vibration Prediction in Rotating Machinery: Application to a Cement Mill Fan. *Engineering, Technology & Applied Science Research*. <https://doi.org/10.48084/etasr.11845>
- Benchekroun, M. T., Zaki, S., Aboussaleh, M., Belrhiti, H., & Diassana, F. (2024). Development of a kiln petcoke mill predictive model based on a multi-regression XGBoost algorithm. *The International Journal of Advanced Manufacturing Technology*, 1-14. <https://doi.org/10.1007/s00170-023-12689-z>
- Benchekroun, M. T., Zaki, S., & Aboussaleh, M. (2023). Kiln predictive modelization for performance optimization. *The International Journal of Advanced Manufacturing Technology*, 127, 1333 - 1339. <https://doi.org/10.1007/s00170-023-11563-2>
- Bolbot, I., & Slovikovskyi, O. (2025). Intelligent Optimization Of Gas Burner Operating Modes For Cement Drying Units. *Lighting Engineering & Power Engineering*. <https://doi.org/10.33042/2079-424x.2025.64.1.02>
- Cao, G., Zhang, Q., Lu, Z., & Liu, Z. (2023). Optimization Study of Setting Reference Values based on Combined Cement Grinding System. *2023 China Automation Congress (CAC)*, 2202-2207. <https://doi.org/10.1109/cac59555.2023.10450494>
- Chelgani, S. C., Fatahi, R., Pournazari, A., & Nasiri, H. (2025). Modeling energy consumption indexes of an industrial cement ball mill for sustainable production. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-025-03232-z>
- Demir, F., Özer, S., Demir, U., Körükçü, K., Oduncu, H., & Ekin, M. Ş. (2025). Design Optimization and Field Validation of Industrial Fans with CFD for Cement Production: Performance, Energy Savings, and Environmental Benefits. *Sustainability*. <https://doi.org/10.3390/su172210279>
- Fatahi, R., Nasiri, H., Homafar, A., Khosravi, R., Siavoshi, H., & Chelgani, C. S. (2022). Modeling operational cement rotary kiln variables with explainable artificial intelligence methods – a “conscious lab” development. *Particulate Science and Technology*, 41, 715 - 724. <https://doi.org/10.1080/02726351.2022.2135470>
- Fatahi, R., Nasiri, H., Dadfar, E., & Chelgani, C. S. (2022). Modeling of energy consumption factors for an industrial cement vertical roller mill by SHAP-XGBoost: a “conscious lab” approach. *Scientific Reports*, 12. <https://doi.org/10.1038/s41598-022-11429-9>
- Machalek, D., & Powell, K. M. (2019). Model predictive control of a rotary kiln for fast electric demand response. *Minerals Engineering*, 144, 106021. <https://doi.org/10.1016/j.mineng.2019.106021>
- Okoji, A., Anozie, A., Omoleye, J., Taiwo, A., & Babatunde, D. (2023). Evaluation of adaptive neuro-fuzzy inference system-genetic algorithm in the prediction and optimization of NOx emission in cement precalcining kiln. *Environmental Science and Pollution Research*, 30, 54835-54845. <https://doi.org/10.1007/s11356-023-26282-0>
- Okoji, A., Anozie, A., Omoleye, J., Taiwo, A., & Osulale, F. (2021). Energetic assessment of a precalcining rotary kiln in a cement plant using process simulator and neural networks. *Alexandria Engineering Journal*. <https://doi.org/10.1016/j.aej.2021.10.010>
- Okoji, A., Anozie, A., & Omoleye, J. (2022). Evaluating the thermodynamic efficiency of the cement grate clinker cooler process using artificial neural networks and ANFIS. *Ain Shams Engineering Journal*. <https://doi.org/10.1016/j.asej.2022.101704>
- Shin, D., Lee, J., Son, J., Yun, Y., Song, Y.-A., & Song, J. (2024). Intelligent Combustion Control in Waste-to-Energy Facilities: Enhancing Efficiency and Reducing Emissions Using AI and IoT. *Energies*. <https://doi.org/10.3390/en17184634>
- Siame, M. C., Zvarivadza, T., Edjeou, W., Simate, I. N., & Lusambo, E. (2024). Integrating Process Re-Engineering Models in Cement Production to Improve Energy Efficiency. *Applied Sciences*. <https://doi.org/10.3390/app14198850>
- Usman, M., Ahmad, I., Kano, M., Ahmad, F., Ahsan, M., & Sikander, U. (2025). Machine Learning-Based Optimization of Thermodynamic Performance of Combustion Processes in a Cement Manufacturing Plant. *ACS Engineering Au*. <https://doi.org/10.1021/acsengineeringau.5c00050>