



Perancangan Model Prediksi O₂ Excess Berbasis *Machine Learning* pada Gas-fired Boiler PT XYZ

Aqil Daulah As Shobri¹, Hasvienda Mohammad Ridlwan², Rivon Tridesman³

¹Program Studi Sarjana Terapan Teknologi Rekayasa Konversi Energi, Jurusan Teknik Mesin, Politeknik Negeri Jakarta, Kampus UI Depok, 16424

²Jurusan Teknik Mesin, Politeknik Negeri Jakarta, Kampus UI Depok, 16424

³PT XYZ, Bontang, Kalimantan Timur, 75324

Corresponding author *E-mail address*: hasvienda.ridlwan@mesin.pnj.ac.id

Abstrak

Jumlah kandungan oksigen dalam gas buang boiler (O₂ excess) dijadikan indikasi efisiensi boiler PT XYZ. Tidak terdapatnya O₂ analyzer pengganti dapat menjadi masalah ketika O₂ analyzer tersebut mati atau rusak sehingga tidak dapat memonitor O₂ excess. Pengukuran secara manual menggunakan portable O₂ analyzer dapat dilakukan, tetapi memiliki risiko bahaya yang tinggi. Penambahan perangkat O₂ analyzer yang serupa dapat dilakukan tetapi memerlukan biaya tambahan yang cukup besar. Alternatif solusi menggunakan model prediksi berbasis machine learning dengan memanfaatkan data histori proses boiler, tetapi terdapat tantangan dalam implementasi model machine learning pada sistem database plant information PT XYZ yaitu menyesuaikan dengan keterbatasannya. Pada penelitian ini dilakukan perancangan model prediksi O₂ excess berbasis machine learning menggunakan standard kerangka CRISP – DM. Tahap business understanding menentukan standard kelayakan model prediksi yaitu nilai Mean Absolute Error (MAE) kurang dari 0,1 %O₂ dan waktu pembelajaran kurang dari 0,8 s. Setelah tahap data preparation didapatkan 13 input pilihan dari 17 variabel. Tahap pemodelan dilakukan menggunakan tiga algoritma yaitu Random Forest, Extreme Gradient Boosting, dan Light Gradient Boosting dengan mengeliminasi satu-persatu feature yang tidak penting. Setelah melalui tahap evaluasi, model XGBoost 6 input dengan nilai MAE 0.0682 %O₂ dan waktu pembelajaran 0.6835s siap diimplementasikan.

Kata-kata kunci: O₂ excess, Model Prediksi Berbasis Machine Learning, Data Histori Proses, CRISP – DM

Abstract

The oxygen content in boiler exhaust gas (O₂ excess) indicates boiler efficiency at PT XYZ. Without a backup O₂ analyzer, monitoring becomes impossible if the current analyzer fails, posing safety risks with manual measurement. While adding another O₂ analyzer is an option, it comes with high costs. A machine learning-based prediction model, using historical boiler process data, offers an alternative solution. However, implementing this model within PT XYZ's plant information system is challenging due to system limitations. This study designs a machine learning-based O₂ excess prediction model using the CRISP-DM framework. The business understanding phase set the feasibility standards with a Mean Absolute Error (MAE) below 0.1 %O₂ and a learning time under 0.8 seconds. After data preparation, 13 out of 17 variables were selected. The modeling phase used three algorithms: Random Forest, Extreme Gradient Boosting, and Light Gradient Boosting, by eliminating less important features. After evaluation, the XGBoost model with 6 inputs, an MAE of 0.0682 %O₂, and a learning time of 0.6835 seconds is ready for implementation.

Keywords: O₂ excess, Machine Learning-Based Prediction Model, Historical Process Data, CRISP – DM

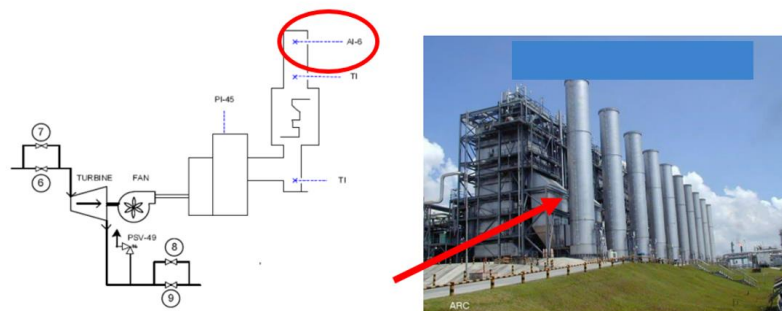
1. PENDAHULUAN

PT XYZ merupakan perusahaan yang bergerak pada produksi LNG di Indonesia. Salah satu peralatan boiler merupakan unit proses yang sangat penting di PT XYZ karena berfungsi menyediakan kebutuhan *High Pressure Steam* (HPS) untuk pembangkit listrik, penggerak kompresor, dan sebagai pemanas pada proses produksi LNG. Sebuah boiler dirancang untuk menyerap sebanyak mungkin energi panas yang dilepaskan dari proses pembakaran. Produksi steam yang optimal pada boiler dipengaruhi oleh nilai efisiensi boiler karena merepresentasikan efektifitas pembakaran pada burner dan perpindahan panas ke *Boiler Feed Water* (BFW). Salah satu parameter boiler adalah nilai kandungan oksigen pada gas buang. Pengendalian kandungan oksigen sangat penting untuk menjaga proses pembakaran pada boiler tetap stabil dan aman. Acuan untuk menentukan performa boiler tentu tidak terlepas dari efisiensi proses pembakaran yang terjadi di dalam boiler. Efisiensi pembakaran boiler secara umum menjelaskan kemampuan sebuah burner untuk membakar keseluruhan bahan bakar yang masuk ke dalam boiler.

Jumlah kandungan oksigen dalam gas buang boiler (*O₂ excess*) dijadikan indikasi efisiensi boiler karena ketika *O₂ excess* terlalu banyak mengindikasikan kelebihan udara pembakaran yang dapat menyebabkan kerugian panas karena terbawa oleh gas buang melalui cerobong sehingga dapat menurunkan efisiensi boiler, sedangkan jika *O₂ excess* terlalu rendah yang mengindikasikan kelebihan bahan bakar dapat menyebabkan pembakaran tidak sempurna sehingga dapat menurunkan efisiensi boiler [1], selain itu pembakaran yang tidak sempurna dapat menyebabkan emisi gas CO yang membahayakan [2]. Pada boiler PT XYZ, pengukuran *O₂ excess* dilakukan menggunakan *O₂ analyzer* yang dipasang pada cerobong boiler.

Dalam operasional PT XYZ, kandungan *O₂ excess* sebesar 2-5% dijadikan patokan untuk menjamin pembakaran sempurna, tidak menimbulkan emisi, dan meminimalkan kerugian panas yang terbawa oleh gas buang melalui cerobong boiler sehingga pengukuran *O₂ excess* perlu dilakukan secara terus menerus (*continuous*). *O₂ excess* tidak hanya penting untuk dimonitor tetapi juga sebagai basis penting yang merepresentasikan pembakaran sempurna dalam operasi boiler [3]. Selain itu pengukuran *O₂ excess* digunakan untuk melakukan evaluasi mingguan kinerja boiler PT XYZ.

Tidak terdapatnya *O₂ analyzer* pengganti dapat menjadi masalah ketika *O₂ analyzer* tersebut mati atau rusak sehingga tidak dapat memonitor *O₂ excess*. Terdapat beberapa alternatif solusi yang dapat dilakukan, pertama yaitu melakukan pengukuran secara manual menggunakan *portable O₂ analyzer*, tetapi pekerjaan tersebut memiliki risiko bahaya yang tinggi karena bekerja di ketinggian dan temperatur tinggi saat boiler aktif sebagaimana ditunjukkan pada gambar 1. Mitigasi kedua adalah dengan penambahan perangkat *O₂ analyzer* yang serupa, hal ini dapat dilakukan tetapi memerlukan biaya tambahan yang cukup besar untuk pengadaan *O₂ analyzer* baru, pembangunan fasilitas, dan pemeliharaan peralatan ke depannya sehingga tidak menjadi prioritas saat ini, selain itu juga terdapat faktor *delivery time* yang cukup lama.



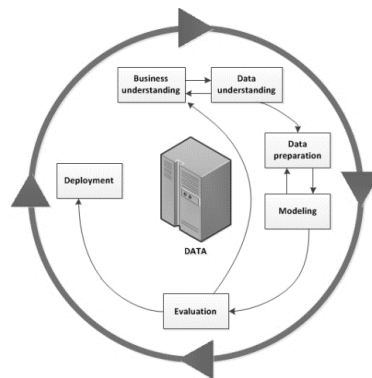
Gambar 1. Titik Pengukuran *O₂ Excess* di Boiler PT XYZ

Dalam beberapa tahun terakhir, metode pengukuran proses industri menggunakan model prediksi berbasis *machine learning* dengan memanfaatkan data historis proses telah menjadi salah satu bidang yang mengalami perkembangan pesat sebagaimana penelitian [1], [4], [5], [6]. *Machine learning* merupakan salah satu bagian kecedasan buatan yang berfokus pada pembelajaran pada sisi proses komputasi untuk dapat menjalankan sebuah tugas sehingga dapat memproses data multivariasi berdimensi tinggi dan mengekstrak hubungan tersembunyi di dalam data dalam lingkungan yang kompleks dan dinamis [7]. Kemajuan dalam pengolahan data, peningkatan kapasitas penyimpanan, dan algoritma yang semakin canggih menjadikan teknologi prediksi berbasis *machine learning* mampu memberikan wawasan yang lebih akurat dan dapat diandalkan dalam berbagai sektor [8]. Berdasarkan hal tersebut terdapat alternatif solusi yang dapat digunakan sebagai perangkat pengganti pengukuran *O₂ analyzer* yang disebutkan sebelumnya yaitu dengan menggunakan model prediksi

O₂ excess memanfaatkan data pengukuran lainnya [9]. Namun, terdapat tantangan dalam implementasi secara *close loop* (*real time*) model prediksi berbasis *machine learning* pada sistem database plant information PT Badak NGL yang bersifat *rigid*. Penelitian [7], [10], [11] menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP – DM) dalam melakukan perancangan dan implementasi model prediksi berbasis *machine learning*. Oleh karena itu penelitian ini bertujuan untuk merancang dan melakukan implementasi model prediksi O₂ excess berbasis *machine learning* pada boiler PT XYZ menggunakan metode CRISP – DM dengan memanfaatkan data histori proses.

2. METODE PENELITIAN

Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan standard kerangka kerja yang digunakan dalam proses *data mining* [12], [13] dan merupakan kerangka kerja yang telah terbukti (*proven*) dalam skala industri [14]. Kerangka tersebut pertama kali dikembangkan oleh empat perusahaan yaitu *Integral Solution*, *NCR*, *Daimler Chrysler*, dan *OHRA* [13]. Terdapat 6 tahapan dalam CRISP – DM yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment* sebagaimana ditunjukkan pada Gambar 2.6 [14]. Berdasarkan Keputusan Menteri Ketenagakerjaan No. 299 Tahun 2020, Tahapan tersebut dijadikan fungsi utama dalam pemetaan Standard Kompetensi Kerja Indonesia (SKKNI) bidang *data science*.



Gambar 2. Tahapan pada CRISP – DM [14]

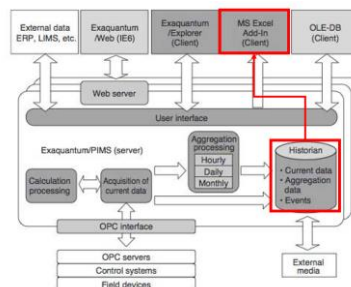
Berdasarkan gambar 2, berikut ini merupakan penjelasan terkait fungsi dan bentuk kegiatan dalam setiap tahapan CRISP – DM yang dilakukan pada penelitian kali ini.

1. Business Understanding

Tahap *business understanding* merupakan tahap CRIPS – DM dalam menganalisis kebutuhan organisasi yang meliputi kegiatan untuk menentukan tujuan penelitian, menentukan kebutuhan data, dan standard kelayakan model prediksi O₂ excess.

2. Data Understanding

Tahap *data understanding* merupakan tahap CRIPS – DM dalam menganalisis kebutuhan data, melakukan pengumpulan data, dan identifikasi data yang dikumpulkan. pengumpulan data menggunakan metode *data crawling* melalui MS Excell Add-in pada komputer yang terhubung ke sistem *database plant information* PT XYZ sebagaimana ditunjukkan pada Gambar berikut.



Gambar 3. Ilustrasi Cara Kerja MS Excell Add-in Dalam Mengakses Data Histori Pengukuran Proses pada Sistem Database Plant Information PT XYZ [15]

3. Data Preparation

Tahap *data preparation* merupakan tahap CRISP – DM dalam aktivitas yang dibutuhkan untuk membangun dataset yang akan dalam proses pemodelan [16] meliputi pembersihan data, penambahan dan pemilihan variabel input, dan pemisahan data latih dan data uji. Pemilihan variabel menggunakan analisis korelasi antar variabel menggunakan metode *Pearson Correlation Coefficient* (PCC). Hasil dari metode PCC Metode PCC dapat menjelaskan seberapa kuat korelasi antar dua variabel dengan persamaan dan kriteria yang dijelaskan berikut.

$$PCC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

Dengan,

x = Variabel Pertama

y = Variabel Kedua

Kemudian, kriteria korelasi yang dapat ditunjukkan pada tabel 1 berikut.

Tabel 1. Kriteria Korelasi Antar Dua Variabel Berdasarkan Nilai PCC

No.	Skala Nilai	Korelasi
1.	$0 < PCC < 0.19$	Korelasi sangat rendah
2.	$0.19 < PCC < 0.39$	Korelasi rendah
3.	$0.39 < PCC < 0.59$	Korelasi moderat
4.	$0.59 < PCC < 0.79$	Korelasi tinggi
5.	$0.79 < PCC < 1.0$	Korelasi sangat tinggi

Kemudian untuk mengetahui tingkat kepentingan variabel dilakukan metode *random forest feature importance*.

4. Modelling

Tahap *modelling* dilakukan dengan melakukan analisis pemilihan model prediksi berdasarkan dataset yang tersedia dan pelatihan model menggunakan dataset yang telah melalui tahap persiapan data sebelumnya.

5. Evaluation

Tahap *evaluation* merupakan tahap CRISP – DM dalam mengembangkan model yang meliputi kegiatan evaluasi hasil pemodelan dan review proses pemodelan. Nilai kelayakan model prediksi O2 excess ditentukan pada tahap pemahaman bisnis berupa nilai minimum *Mean Absoluter Error* (MAE) yang diperbolehkan, perhitungan nilai MAE dapat menggunakan persamaan berikut.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i| \quad (2)$$

Selain itu kelayakan model prediksi ditentukan dengan minimum waktu pembelajaran yang ditetapkan. Perhitungan waktu pembelajaran model dapat menggunakan persamaan berikut.

$$waktu = time(end) - time(start) \quad (3)$$

6. Deployment

Tahap *deployment* merupakan tahap CRISP – DM dalam menggunakan model yang dihasilkan yang meliputi kegiatan analisis metode implementasi dan implementasi model. Tahap *deployment* tidak dilakukan karena berfokus pada perancangan model yang siap untuk diimplementasikan.

3. HASIL DAN PEMBAHASAN

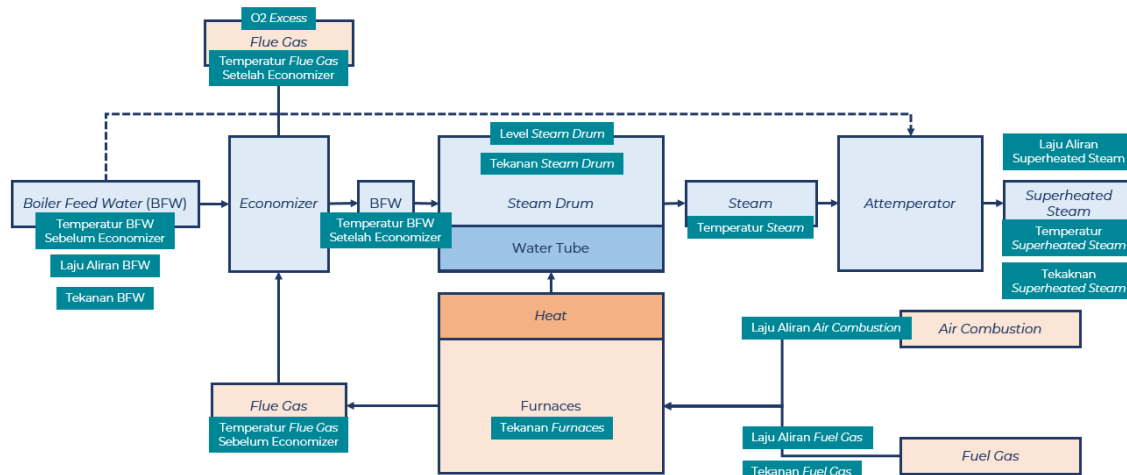
Berdasarkan metodologi yang dijelaskan sebelumnya, pada penelitian ini dilakukan 5 tahapan yang dapat dijelaskan pada hasil dan pembahasan berikut.

1. Business Understanding

Perancangan model prediksi O₂ excess menggunakan pendekatan berbasis *machine learning* merupakan metode yang efektif untuk merancang model prediksi yang optimal dan implementasi model prediksi tersebut pada sistem *database plant information* PT XYZ. Berdasarkan hasil observasi O₂ analyzer tersebut memiliki akurasi 0,1% pembacaan O₂, maka kebutuhan perancangan model prediksi berdasarkan spesifikasi O₂ analyzer tersebut adalah nilai *mean absolute error* (MAE) pada saat pengujian kurang dari 0,1% O₂. Kebutuhan data yang digunakan pada penelitian ini adalah data historis variabel proses yang tersedia pada sistem database PT XYZ. Selain itu, kebutuhan model *machine learning* yang cepat dalam pembelajaran juga dibutuhkan, karena akuisisi data pengukuran proses di PT XYZ dilakukan setiap detik, maka diperlukan model yang memiliki waktu pembelajaran kurang dari 1 detik.

2. Data Understanding

Berdasarkan studi literatur pada buku manual operasi boiler PT XYZ dapat digambarkan *block diagram* proses yang terdapat pada boiler. Secara umum proses pada boiler terbagi menjadi dua yaitu sistem pembentukan panas dan proses penghantaran panas.



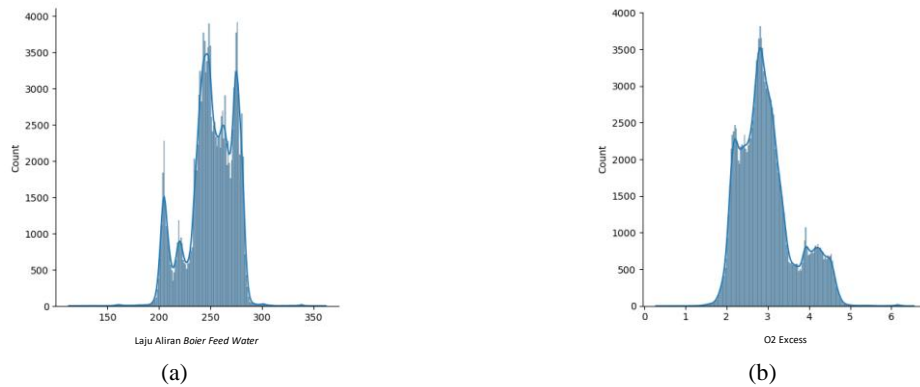
Gambar 4. Block Diagram Proses Boiler Modul II PT XYZ

Berdasarkan Gambar 4, terdapat 17 variabel proses pada boiler modul II PT XYZ yaitu data proses boiler yang dijadikan patokan operasi boiler dalam sistem pembentukan panas adalah laju aliran *air combustion*, laju aliran *fuel gas*, tekanan *fuel gas*, tekanan *furnaces*, temperatur *flue gas* sebelum *economizer*, dan temperatur *flue gas* setelah *economizer*. Kemudian, pada proses penghantaran panas yaitu laju aliran *boiler feed water* (BFW), tekanan *boiler feed water*, tekanan *boiler steam drum*, temperatur BFW sebelum *economizer*, temperatur BFW setelah *economizer*, level *boiler steam drum*, temperatur *superheated steam*, temperatur *superheated steam* setelah *attemperator*, tekanan *superheated steam*, dan laju aliran *superheated steam*.

Berdasarkan analisis tersebut dilakukan pengumpulan data menggunakan metode *data crawling* melalui komputer yang terhubung pada sistem *database plant information* PT XYZ. Data yang diperoleh berasal dari salah satu boiler modul II PT XYZ dengan waktu pengambilan dari tahun 1 Januari 2023 sampai dengan 20 Juli 2024. Terdapat 163008 jumlah sampel yang berhasil dikumpulkan.

3. Data Preparation

Berdasarkan tahap *data understanding*, Pada kondisi aktual tidak semua data memiliki persebaran yang terdistribusi normal, maka pembersihan data dilakukan dengan analisis batas *Interquartile* (IQR) karena analisis statistik yang baik untuk data yang tidak terdistribusi normal adalah dengan menggunakan nilai IQR dimana pembersihan data dilakukan dengan mengeliminasi nilai yang diluar *range* nilai yang telah ditetapkan pada setiap parameter. Kemudian, dilakukan *feature engineering* dengan menambahkan variabel *Air to Fuel Ratio* (AFR) yaitu membagi data laju aliran *air combustion* dengan laju aliran *fuel gas*.



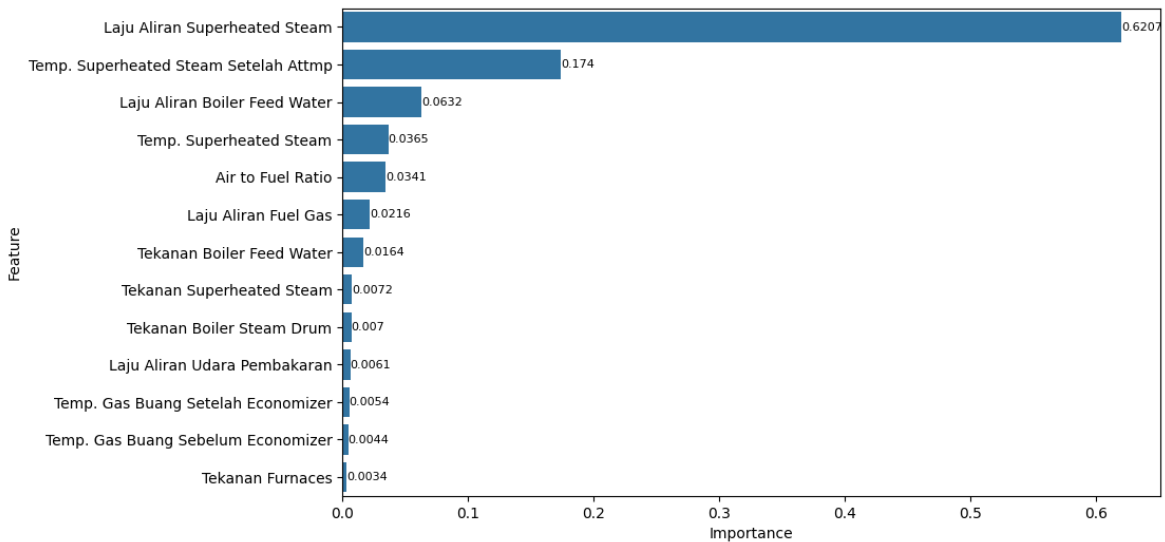
Gambar 5. Grafik distribusi tidak normal (a). Variabel Laju Aliran *Boiler Feed Water* (b). Variabel *O2 Excess*

Berdasarkan data yang telah dilakukan pembersihan didapatkan nilai PCC setiap variabel input terhadap *O2 excess* sebagai Tabel 3 berikut.

Tabel 2. Parameter Input Model Prediksi *O2 Excess* Berdasarkan Hasil Perhitungan PCC

No.	Deskripsi Parameter	PCC	Korelasi
1.	<i>Air to Fuel Ratio (AFR)</i>	0,87	Sangat Tinggi
2.	Laju Aliran <i>Air Combustion</i>	-0,77	Tinggi
3.	Laju Aliran <i>Fuel Gas</i>	-0,87	Sangat Tinggi
4.	Tekanan <i>Fuel Gas</i>	0,17	Sangat Rendah
5.	Tekanan <i>Furnaces</i>	-0,78	Tinggi
6.	Temperatur <i>Flue Gas</i> Sebelum <i>Economizer</i>	-0,85	Sangat Tinggi
7.	Temperatur <i>Flue Gas</i> Setelah <i>Economizer</i>	-0,76	Tinggi
8.	Laju Aliran <i>Boiler Feed Water (BFW)</i>	-0,89	Sangat Tinggi
9.	Tekanan <i>Boiler Feed Water (BFW)</i>	-0,86	Sangat Tinggi
10.	Temperatur BFW Sebelum <i>Economizer</i>	-0,32	Rendah
11.	Temperatur BFW Setelah <i>Economizer</i>	-0,01	Sangat Rendah
12.	Tekanan <i>Boiler Steam Drum</i>	-0,83	Sangat Tinggi
13.	Level <i>Boiler Steam Drum</i>	-0,06	Sangat Rendah
14.	Temperatur <i>Superheated Steam</i>	-0,63	Tinggi
15.	Temperatur <i>Superheated Steam</i> Setelah <i>Attemperator</i>	-0,71	Tinggi
16.	Tekanan <i>Superheated Steam</i>	-0,46	Moderate
17.	Laju Aliran <i>Superheated Steam</i>	-0,89	Sangat Tinggi

Berdasarkan hasil PCC, variabel yang memiliki korelasi rendah dan sangat rendah dapat dieliminasi. Kemudian pada tahap ini dilakukan analisis kepentingan variabel input yang telah dipilih berdasarkan metode PCC menggunakan *random forest feature importance*.

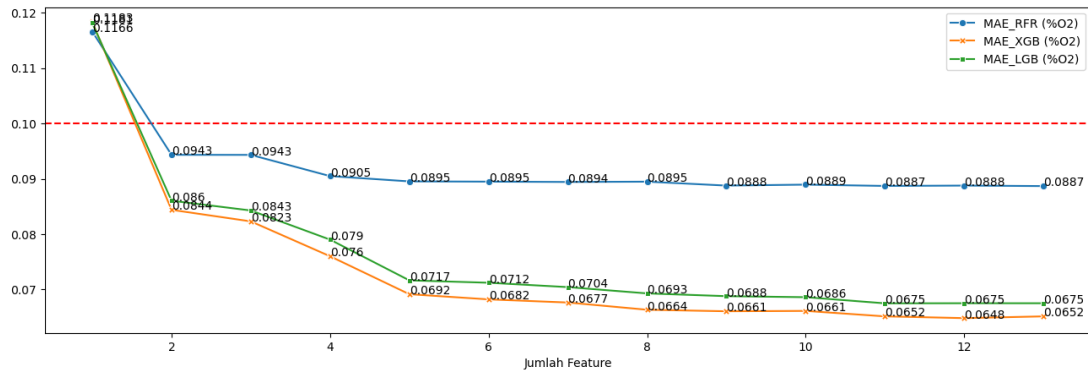
Gambar 6. Hasil *Random Forest Feature Importance*

4. Modelling

Perancangan model prediksi berbasis data merupakan bagian dari konsep *machine learning* jenis *supervised learning* tipe regresi. Terdapat ribuan model yang dapat digunakan dalam perancangan model prediksi, model yang sangat bagus dalam menangani data yang tidak terdistribusi normal adalah model prediksi berbasis *decision tree*, tetapi model tersebut cenderung *overfitting* terhadap data latih sehingga memiliki *error* yang cukup tinggi ketika pengujian model [17]. Solusi yang paling umum dan mudah untuk mengatasinya adalah dengan menggunakan teknik *ensemble*. Teknik tersebut memungkinkan beberapa model *decision tree* untuk meningkatkan performa dan mengurangi *overfitting*. Beberapa algoritma model prediksi menggunakan teknik *ensemble* berbasis *decision tree* yang populer digunakan adalah *Random Forest* (RF), *Gradient Boosting Machine* (GBM), *Extreme Gradient Boosting* (XGBoost) dan *Light Gradient Boosting* (LightGBM) [18]. *Random Forest* (RF) merupakan algoritma *machine learning* berbasis *decision tree* yang menggunakan teknik *ensemble bagging*, algoritma ini didasarkan pada penggunaan beberapa *decision tree* untuk menghasilkan prediksi yang lebih umum (*general*) sehingga dapat mengurangi *overfitting* [18]. *Gradient Boosting Machine* (GBM) menggunakan teknik *ensemble boosting*, tidak seperti teknik *ensemble bagging*, teknik GBM melakukan pelatihan *weak learner* secara independen satu sama lain dengan mengembangkan *learner* selanjutnya melalui pembelajaran dari *weak learner* sebelumnya sehingga dapat mengurangi *error* hasil prediksi dengan nilai aktual [18]. *Extreme Gradient Boosting* (XGBoost) merupakan algoritma *machine learning* yang menggunakan teknik *ensemble boosting*, tidak seperti GBM, XGBoost mempunyai solusi lebih cepat pada dataset yang besar karena dapat melakukan komputasi secara bersamaan [18]. *Light Gradient Boosting* (LightGBM) merupakan algoritma *machine learning* lainnya yang menggunakan teknik *ensemble boosting*. LightGBM merupakan algoritma yang dikembangkan oleh Microsoft untuk membangun model *machine learning* dengan performa tinggi pada dataset yang besar sehingga mampu melakukan proses pembelajaran lebih cepat dan menggunakan memori yang lebih sedikit. Pada penelitian sebelumnya telah dikembangkan model prediksi O₂ excess menggunakan teknik *ensemble* berbasis *decision tree* yaitu menggunakan algoritma *Random Forest* (RF) [5] dan *Extreme Gradient Boosting* (XGBoost) [4] sehingga pada penelitian ini dilakukan percobaan dengan menggunakan tiga algoritma *machine learning* yaitu RF, XGBoost, dan LightGBM dengan melakukan evaluasi kelayakan model sesuai yang dijelaskan pada tahap *business understanding* dan waktu pembelajaran untuk mendapatkan waktu pembelajaran yang cepat.

5. Evaluation

Melalui pengujian yang dilakukan diperoleh nilai MAE setiap jumlah variabel input berdasarkan *feature importance* berikut.



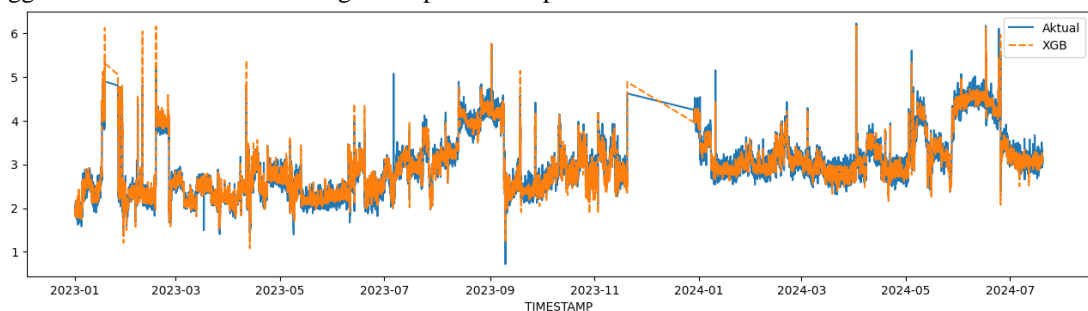
Gambar 7. Hasil Pengujian MAE Model Prediksi O2 Excess

Berikut merupakan grafik hasil pengujian waktu pembelajaran model setiap jumlah *feature importance* menggunakan model XGBoost dan LightGBM.

Tabel 3. Hasil Pengujian Waktu Pembelajaran Model

No.	Jumlah Feature Penting	RF (s)		XGBoost (s)		LightGBM (s)	
		MAE (%O2)	Train Time (s)	MAE (%O2)	Train Time (s)	MAE (%O2)	Train Time (s)
1.	1	0.116581	10.927757	0.118304	0.395772	0.118141	0.569666
2.	2	0.094327	17.181498	0.084396	0.584659	0.086041	0.603649
3.	3	0.094319	22.980611	0.082312	0.475724	0.08428	0.49671
4.	4	0.090485	30.901999	0.076013	0.550681	0.079015	0.489713
5.	5	0.089533	34.725769	0.069165	0.583661	0.071661	0.541686
6.	6	0.089487	43.444718	0.068228	0.682579	0.071231	0.583658
7.	7	0.089428	53.446892	0.067661	0.832489	0.070443	0.746568
8.	8	0.089481	55.942437	0.066356	0.914445	0.069307	0.746562
9.	9	0.088772	59.139549	0.06608	0.787543	0.068808	0.731574
10.	10	0.088947	68.630025	0.066146	0.899473	0.068622	0.750563
11.	11	0.08871	74.245774	0.06516	0.897453	0.067516	0.837511
12.	12	0.088772	81.044792	0.064844	0.988423	0.067533	0.794536
13.	13	0.088685	88.380717	0.06516	1.054389	0.067531	0.894482

Berdasarkan hasil pengujian tersebut terdapat 14 model prediksi O2 excess yang sesuai dengan standard kelayakan model prediksi yang dijelaskan sebelumnya. Kemudian, dari 16 model prediksi O2 excess tersebut nilai MAE terendah adalah XGBoost dengan 6 input, kemudian LightGBM dengan 10 input, dan LightGBM dengan 9 input, dari tiga model prediksi O2 excess tersebut direkomendasikan menggunakan XGBoost dengan 6 input penting yaitu laju aliran *superheated steam*, temperatur *superheated steam* setelah *attemperator*, laju aliran *boiler feed water*, temperatur *superheated steam*, *air to fuel ratio*, dan laju aliran *fuel gas* karena memiliki waktu pembelajaran yang relatif lebih rendah. Gambar 6 menunjukkan hasil prediksi O2 excess menggunakan model XGBoost dengan 6 input terhadap data O2 excess aktual.



Gambar 8. Hasil XGBoost

4. KESIMPULAN

Perancangan model prediksi O₂ *excess* menggunakan standard kerangka CRISP – DM berhasil dilakukan. Tahap *business understanding* menentukan tujuan penelitian yaitu merancang model prediksi O₂ *excess* menggunakan data histori proses yang tersedia pada sistem *database information plant* PT XYZ dengan standard kelayakan model prediksi yaitu nilai *Mean Absoulute Error* (MAE) kurang dari 0,1 % O₂ dan waktu pembelajaran kurang dari 0,8 s. Tahap *data understanding* menetapkan 17 variabel proses yang didasarkan pada proses boiler PT XYZ dikumpulkan menggunakan metode *data crawling* melalui komputer yang terhubung sistem *database information plant* PT XYZ dan diperoleh 163008 jumlah sampel dari 1 Januari 2023 sampai dengan 20 Juli 2024. Tahap *data preparation* didapatkan 13 input pilihan dari 17 input yang tersedia. Tahap pemodelan (*modelling*) dilakukan menggunakan tiga algoritma pilihan yaitu *Random Forest*, *Extreme Gradient Boosting*, dan *Light Gradient Boosting* dengan mengeliminasi satu-persatu *feature* yang tidak penting. Setelah melalui tahap evaluasi, terdapat 14 model prediksi yang memenuhi standard kelayakan, model XGBoost dengan 6 input penting yaitu laju aliran *superheated steam*, temperatur *superheated steam* setelah *attemperator*, laju aliran *boiler feed water*, temperatur *superheated steam*, *air to fuel ratio*, dan laju aliran *fuel gas* dengan nilai MAE 0.0682 % O₂ dan waktu pembelajaran 0.6835s direkomendasikan untuk diimplementasikan.

REFERENSI

- [1] Z. Tang, H. Zhang, dan H. Yang, “Artificial neural networks model for predicting oxygen content in flue gas of power plant,” *Proc. 29th Chinese Control Decis. Conf. CCDC 2017*, hal. 1379–1382, 2017, doi: 10.1109/CCDC.2017.7978732.
- [2] Pittara, “Keracunan Karbon Monoksida,” ALODOKTER. Diakses: 21 Juli 2024. [Daring]. Tersedia pada: <https://www.alodokter.com/keracunan-karbon-monoksida>
- [3] L. Ma, X. Zuo, dan Y. Wang, “Boiler oxygen optimization based on double-hidden-layer bp neural network and improved pso algorithm,” *1st Int. Conf. Ind. Artif. Intell. IAI 2019*, hal. 1–5, 2019, doi: 10.1109/ICIAI.2019.8850735.
- [4] N. Kurniawan, Eko & Effendy, Nazrul & Arif, Agus & Dwianto, Kenny & Muddin, “Soft Sensor for the Prediction of Oxygen Content in Boiler Flue Gas Using Neural Networks and Extreme Gradient Boosting,” *Neural Comput Applic*, vol. 35, hal. 345–352, 2022, doi: <https://doi.org/10.1007/s00521-022-07771-8>.
- [5] N. Effendy, E. D. Kurniawan, K. Dwianto, A. Arif, dan N. Muddin, “The prediction of the oxygen content of the flue gas in a gasfired boiler system using neural networks and random forest,” *IAES Int. J. Artif. Intell.*, vol. 11, no. 3, hal. 923–929, 2022, doi: 10.11591/ijai.v11.i3.pp923-929.
- [6] S. N. Sembodo, N. Effendy, K. Dwianto, dan N. Muddin, “Radial basis network estimator of oxygen content in the flue gas of debutanizer reboiler,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, hal. 3044–3050, 2022, doi: 10.11591/ijece.v12i3.pp3044-3050.
- [7] N. Hafidhoh, A. P. Atmaja, G. N. Syaifuddiin, dan I. B. Sumafta, “Machine Learning untuk Prediksi Kegagalan Mesin dalam Predictive Maintenance System,” vol. 15, no. May, 2024, doi: 10.14710/jmasif.15.1.63641.
- [8] F. Provost dan T. Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. New York: O’Reilly, 2013.
- [9] L. Fortuna, S. Graziani, A. Rizzo, dan M. G. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*. Springer, 2007.
- [10] A. M. M. Fattah, A. Voutama, N. Heryana, dan N. Sulistiyowati, “Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, hal. 1669, 2022, doi: 10.30865/jurikom.v9i5.5021.
- [11] Y. Yudiana, A. Yulia Agustina, dan dan Nur Khofifah, “Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan,” *Indones. J. Islam. Econ. Bus.*, vol. 8, no. 1, hal. 01–20, 2023, [Daring]. Tersedia pada: <http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb>
- [12] A. Khumaidi, “DATA MINING FOR PREDICTING THE AMOUNT OF COFFEE PRODUCTION USING CRISP-DM METHOD,” vol. 17, no. 1, hal. 1–8, 2020.
- [13] A. Pambudi, Z. Abidin, dan Permata, “PENERAPAN CRISP - DM MENGGUNAKAN MLR K - FOLD PADA DATA SAHAM PT . TELKOM INDONESIA (PERSERO) TBK (TLKM) (STUDI KASUS : BURSA EFEK,” vol. 4, no. March, hal. 1–14, 2023, doi: 10.33365/jdmsi.v4i1.2462.
- [14] International Business Machines Corporation, *IBM SPSS Modeler CRISP-DM Guide*. International

- Business Machines Corporation, 2021. [Daring]. Tersedia pada: https://www.ibm.com/docs/it/SS3RA7_18.3.0/pdf/ModelerCRISPDM.pdf
- [15] Yokogawa, *Technical Information PIMS*. 2007. [Daring]. Tersedia pada: <https://www.astbearings.com/bearing-materials.html>
- [16] M. Sholeh, E. K. Nurnawati, dan U. Lestari, “Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, hal. 10–21, 2023, doi: 10.14421/jiska.2023.8.1.10-21.
- [17] H. Tatsat, S. Puri, dan B. Lookabaugh, “Supervised Learning: Models and Concepts,” in *Machine Learning and Data Science Blueprints for Finance by Hariom Tatsat, Sahil Puri, Brad Lookabaugh*, O’Reilly Media, 2020, hal. 429. [Daring]. Tersedia pada: <https://www.oreilly.com/library/view/machine-learning-and/9781492073048/ch04.html>
- [18] V. Gül, “Machine Learning Part-4 (Random Forest-GBM-XGBoost-LightGBM-CatBoost),” Medium. [Daring]. Tersedia pada: <https://medium.com/@veribilimi35/machine-learning-part-4-random-forest-gbm-xgboost-lightgbm-catboost-10e4e69eef33>