



MEMBANGUN TOOL PREDIKSI YANG KOMPREHENSIF GUNA MENINGKATKAN AKURASI PRODUK AKHIR PENCAMPURAN KUALITAS BATUBARA MENERAPKAN PERBANDINGAN KOLABORATIF PADA ALGORITMA REGRESI HINGGA NEURAL NET

¹⁾Yoga P. Adiwiguna ²⁾Hario Purbaseno ³⁾Hadi Syuhara ⁴⁾Freddy J. Pribadi

⁵⁾Akhmad Sarbani ⁶⁾Almira D. Kusuma

^{1),2),3),4),5),6)}Adaro Indonesia

Artikel masuk : 11-11-2022 , Artikel diterima : 28-11-2022

Kata kunci: Kualitas Batubara, Neural Network, Pencampuran Batubara, Predictive Analytics, Regresi

Keywords: Mining Strategic Plan, Reserve Optimization, Water Quality Standards

ABSTRAK

Prediksi nilai akhir pencampuran kualitas batubara sebagian besar ditentukan hanya dengan menghitung nilai rata-rata terbobot antara kualitas dan kuantitas batubara dari masing-masing konstituen, teknik perhitungan ini mengabaikan kondisi dimana variasi hasil akhir tidak hanya ditentukan oleh faktor material tetapi juga dipengaruhi oleh faktor manusia, lingkungan, mesin, dan faktor metode, yang masing-masing berkontribusi terhadap total variasi. Kami mencatat bahwa masih ada beberapa kasus di mana perbedaan antara nilai prediksi dan nilai aktual berada di luar ketetapan ISO (± 71 kcal/kg), yang mengakibatkan kerugian yang cukup besar tidak hanya dari segi rangkaian operasional tetapi juga berdampak pada aspek keuangan karena menyebabkan munculnya berbagai biaya tambahan. Kami melihat bahwa hubungan antara nilai kalori akhir hasil analisa laboratorium dan kontributor variansnya dapat didekati sebagai masalah regresi. Memanfaatkan beberapa algoritma regresi akan membantu menghasilkan metode prediksi yang lebih baik daripada metode rata-rata terbobot sederhana. Dengan ribuan kumpulan data dan 60 fitur independen, kami secara bertahap menerapkan algoritma regresi sederhana seperti regresi linier hingga algoritma regresi lanjutan seperti Random Forest Regressor, XGBoost Regressor, dan Artificial Neural Networks (ANN). Untuk mendapatkan hasil yang optimal, pada setiap implementasi algoritma regresi lanjutan diterapkan metode hyperparameter tuning dengan iterasi hingga 1000 kali. Mempertimbangkan metrik evaluasi regresi, validasi model terbaik ditentukan dari nilai Mean Absolute Error (MAE) dan Mean Absolute Percentage Error (MAPE) terkecil. Langkah terakhir adalah memilih dan menyimpan model yang terbaik sebagai pickle string untuk membuat Django application, kemudian agar memungkinkan aksesibilitas yang luas, dilakukan deployment aplikasi di Heroku. Dengan melibatkan kontributor variasi tambahan, yaitu faktor lingkungan (curah hujan & jam hujan), faktor metode (dasar pemilihan analisis laboratorium), dan faktor material (water spray dan dust suppressant treatment), kami memperoleh metrik evaluasi regresi terbaik saat menggunakan XGBoost di mana nilai MAE (37,5) dan MAPE (0.0082).

Doi : <https://doi.org/10.36986/impj.v4i2.79>

ABSTRACT

The final prediction of coal quality blending value is mostly determined only by calculating the weighted average value between the coal quality and quantity of each constituent, this technique ignores conditions where the variation of final result is not determined only by material factors but also affected by human, environment, machine, and method factors, each of which contributes to the total variation. We note that there are still several cases where the difference between predicted value and actual value are outlying the ISO determination (± 71 kcal/kg), resulting in considerable losses not only in terms of operational sequence but also from the financial aspect due to the emergence of various additional costs. We see that the relationship between the laboratory results of calorific value and its variance contributors can be approached as a regression problem. Utilizing several regression algorithms will help generate a better prediction method than a simple weighted average method. With thousand dataset records

and 60 features, we gradually implement simple regression algorithm (Linear Regression) to advanced regression algorithms (Random Forest Regressor, XGBoost Regressor), and Artificial Neural Networks (ANN). To obtain optimal results, hyperparameter tuning method is applied to each advanced regression algorithm with iterations up to 1000 times. Considering the regression evaluation metrics, the best model validation is determined from the smallest value of Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The last step is selecting and saving the best-trained model as a pickle string for creating Django applications and for enabling extensive accessibility, this application is deployed on Heroku. By involving additional variation contributors, namely Environmental factors (rainfall & rain hour), Method factors (selection basis of a laboratory analysis), and Material factors (water spray and dust suppressant treatment), we obtain the best regression evaluation metrics when using XGBoost where the value of MAE (37.5) and MAPE (0.0082).

Algorithm	MAE	MAPE = 100MARE	R_Squared
Artificial Neural Network	813.3	0.1734	0.2910
Cubist	40.3	0.0087	0.9771
Decision Tree Regressor	57.7	0.0125	0.9513
Multivariate Adaptive Regression Spline	46.6	0.0101	0.9656
Random Forest Regressor	40.4	0.0088	0.9765
Simple Linear Regression	66.2	0.0147	0.9305
Support Vector Regressor	147.1	0.0343	0.5759
XGBoost Regressor	37.5	0.0082	0.9705

Even with more variance contributors, this prediction model can offer better process capability and MAE.

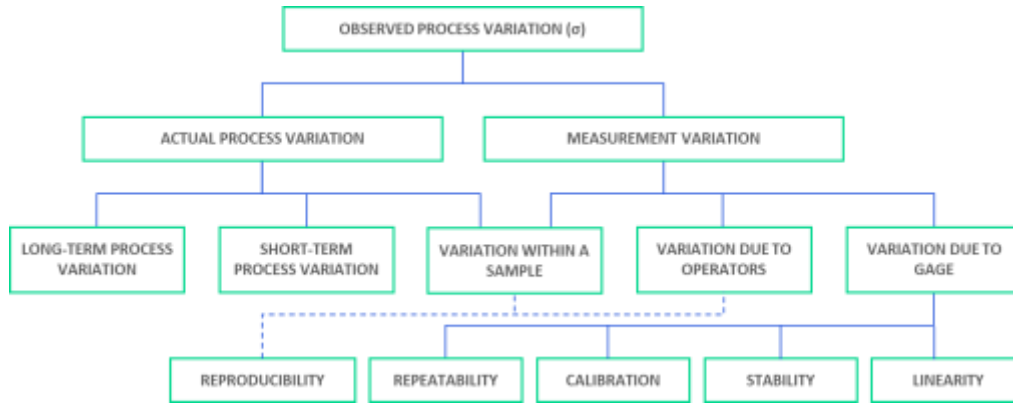
Preventive efforts to minimize the out-of-specification cases of finished products due to accuracy can be carried out as early as possible, thereby minimizing operational (Divert Barge, Rehandling Stock) and financial (Retest Document, Buyer Penalty) losses. This collaborative application will later be utilized by cross-functional teams (Coal Quality, Coal Mining Operation, Coal Dispatcher, Coal Processing Plant, Coal Laboratory) throughout daily coal blending process.

PENDAHULUAN

Metode prediksi/proyeksi nilai akhir *calorific value* pada proses pencampuran batubara umumnya dilakukan dengan melakukan kalkulasi rerata terbobot antara berat batubara

(tonnage) dan parameter kualitas (*calorific value*). Metode ini dilakukan untuk membangun model matematis proses pencampuran batubara agar dapat mendekati keadaan sesungguhnya di lapangan, selain itu metode ini memberikan gambaran cepat hasil akhir *calorific value* batubara karena formulasinya cukup mudah dipahami dan dapat diimplementasikan pada berbagai aplikasi sederhana yang memiliki kemampuan kalkulasi dan komputasi. Di lain sisi, metode ini mengabaikan kondisi di mana nilai akhir *calorific value* sebenarnya dipengaruhi tidak hanya dari faktor besar kecilnya komposisi dan jenis batubara yang digunakan, tetapi lebih jauh ada banyak faktor tambahan lain yang memiliki kontribusi terhadap hasil nilai akhir tersebut (Gambar 1). Banyak panduan yang fokus kepada studi alat

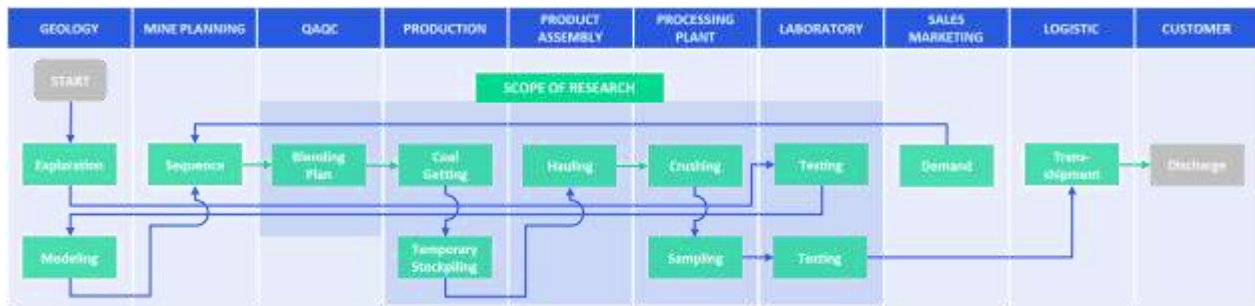
ukur, tetapi sumber variasi lain tidak dapat diabaikan, terutama variasi proses yang sebenarnya menjadi subjek utama yang harus ditangani (Barrentine, 2003).



Gambar 1. Possible sources of process variation (Barrentine, 2003)

Mengacu referensi ini kami sadar bahwa pendekatan rerata terbobot hanya melibatkan faktor besar kecilnya kuantitas batubara (*Material*) yang digunakan, tetapi belum mengikutsertakan faktor manusia (*Man*), mesin (*Machine*), metode (*Method*), lingkungan (*Environment*), maupun efek interaksi gabungan antar faktor yang mungkin saja timbul pada nilai akhir *calorific value* pencampuran batubara.

Untuk itu dibutuhkan pendekatan komprehensif yang melibatkan keseluruhan faktor yang dapat dipetakan dari proses hulu-hilir rantai pasokan batubara agar mampu diidentifikasi apa saja potensi kontributor variasinya (Gambar 2), karena setiap proses penambangan memiliki model rantai suplai yang berbeda-beda menyesuaikan infrastruktur pendukungnya.



Gambar 2. Swimlane proses rantai pasokan batubara

Melibatkan lebih banyak faktor yang berpengaruh dalam penentuan nilai akhir *calorific value* berimplikasi pada banyaknya data yang dibutuhkan dalam proses persiapan hingga proses pengujian, sehingga objektif untuk menghasilkan akurasi nilai akhir *calorific value* batubara yang tinggi harus dibarengi dengan *enabler-tool* yang mumpuni dalam proses eksekusinya agar akurasi yang diperoleh dari hasil kalkulasi nanti menawarkan performa

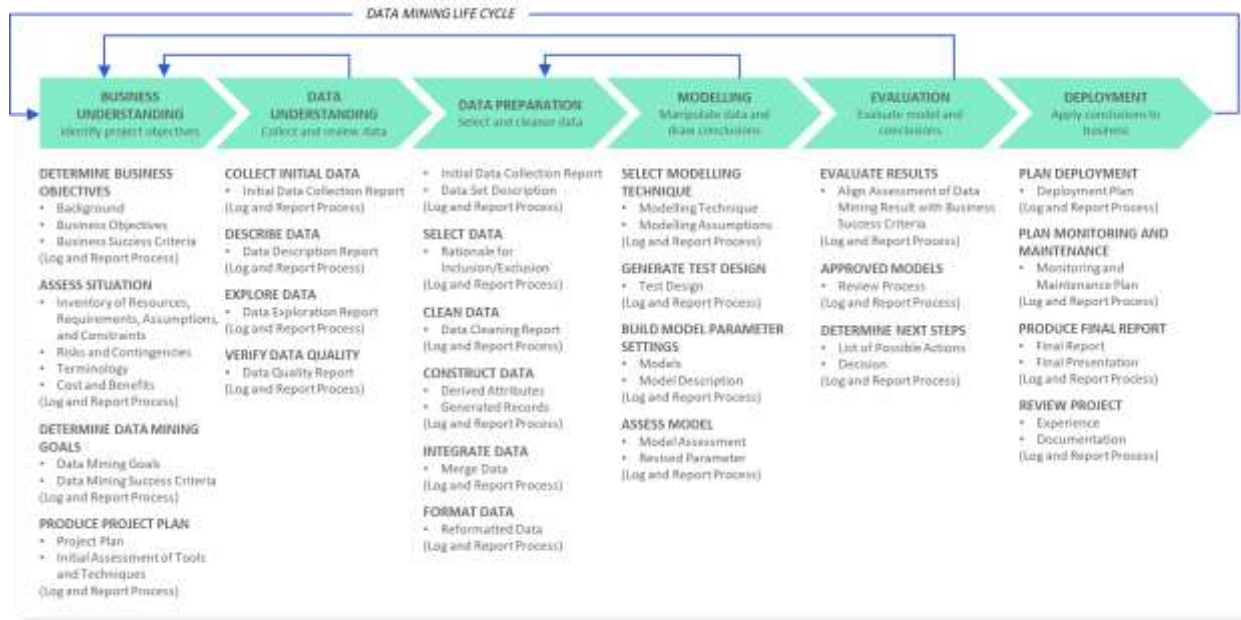
yang lebih tinggi maupun *lead time* yang dihasilkan minimal sama cepat dengan metode sebelumnya atau bahkan bisa dipercepat dengan kemampuan komputasi yang ada. Dengan adanya tool kalkulasi nilai akhir *calorific value* dengan performa prediksi yang lebih baik, maka tata kelola sekuen operasional tambang akan menjadi ramping dan efisien karena dapat meminimalisir upaya *adjustment* material, meminimalisir upaya *divert barge due to*

accuracy problem maupun *pinalti* yang berdampak secara finansial akibat *out-of-specification*.

METODOLOGI PENELITIAN

Mempertemukan antara data apa dimiliki dan data apa yang seharusnya diperlukan, kemudian menggunakan pemodelan yang tepat dari data tersebut dalam upaya pengambilan keputusan

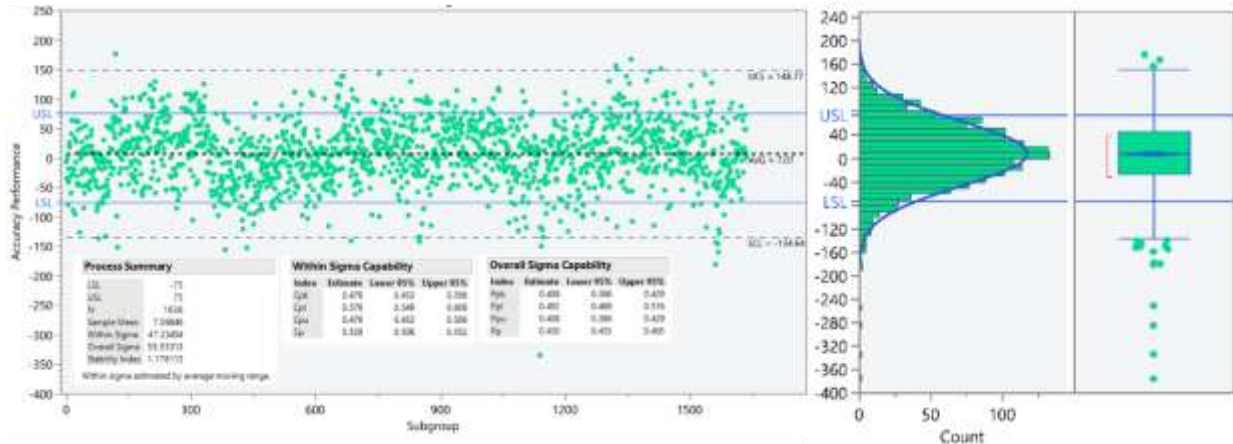
harus didukung dengan kerangka standar yang telah terbukti menyuguhkan hasil terbaik. Kami menggunakan *CRISP-DM Framework* (Gambar 3) untuk menuntun proses persiapan penelitian hingga pengujian dan pemantauan karena kerangka ini telah menjadi metodologi yang paling umum digunakan dalam bidang *Data Mining, Analytics, dan Data Science Projects* (Chapman, 2000).



Penentuan objektif penelitian didasari oleh *key performance indicator* pada bisnis unit QAQC yang bertanggung jawab untuk memastikan akurasi produk akhir pencampuran batubara sesuai dengan kebutuhan masing-masing pelanggan. Upaya pemantauan dilakukan dengan mengukur selisih nilai prediksi dan nilai aktual parameter *calorific value* berada pada rentang ± 75 kcal/kg pada masing-masing pemuatan barge, jika terdapat selisih lebih tinggi/rendah dari rentang tersebut maka akan diklasifikasikan ke dalam kasus *out-of-specification due to accuracy*. Kasus ini akan mengakibatkan rangkaian ketidaksesuaian dari pengaturan material pengganti hingga yang terparah menyebabkan penalti dan *rejection*.

ASESMEN PROSES DAN SITUASI

Selisih nilai prediksi dan nilai aktual selama 2 tahun terakhir dikumpulkan untuk mengukur *process capability* sebelum dilakukannya perbaikan, diperoleh 1636 data di mana nilai rata-ratanya sebesar 7.1 kcal/kg dengan standar deviasi 55.6 kcal/kg, kemudian nilai C_p / derajat kepresisian sebesar 0.529 dan C_{pk} / derajat keakurasian sebesar 0.479. Dari nilai tersebut menunjukkan bahwa *process capability* saat ini masih belum mumpuni untuk memenuhi batas spesifikasi yang telah ditetapkan (Gambar 4).

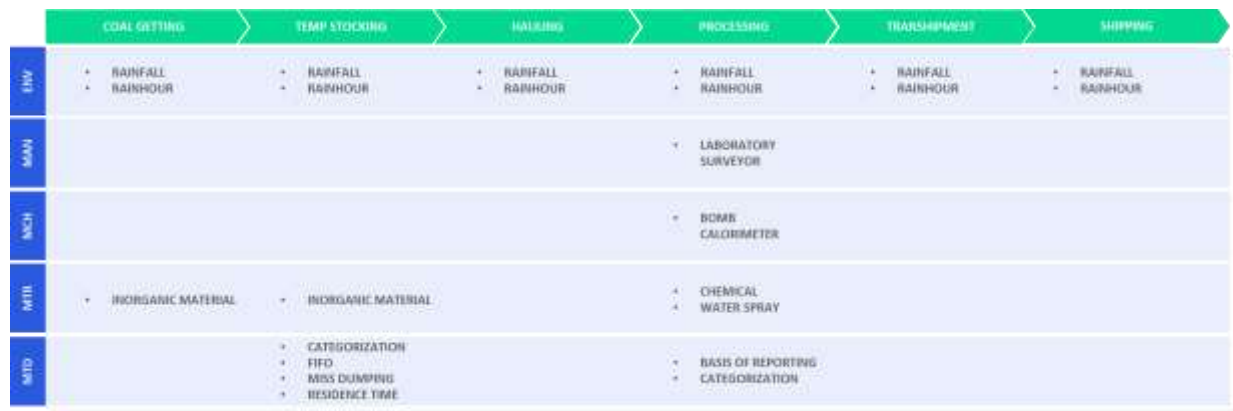


Gambar 4. Control Chart dan Process Capability untuk akurasi blending pada produk akhir

MEMBANGUN DAN MENINGTEGRASIKAN DATA

Melihat hasil performa akurasi dan presisi nilai *calorific value* yang berlangsung, besaran total variasi dari rentang spesifikasi masih terlalu lebar dan harus diidentifikasi faktor-faktor penyebabnya. Untuk dapat memetakan apa saja penyebab dan di proses mana saja variasi ini berkontribusi, kami melakukan *brainstorming* dengan para *subject matter expert* di masing-

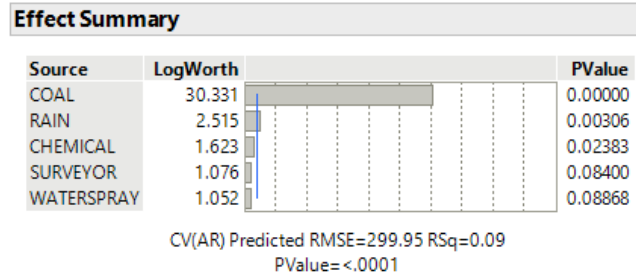
masing proses untuk dapat mengidentifikasi faktor apa saja yang berpotensi mempengaruhi besaran variasi sekaligus memastikan tersedianya data pendukung tersebut agar sesuai dengan format penyusunan data penelitian. Kontributor-kontributor yang berpotensi mempengaruhi nilai variasi ini (Gambar 5) terdiri dari gabungan jenis data numerik dan diskrit (Tabel 1) yang nanti akan diidentifikasi besaran pengaruhnya ke dalam *effect plot* (Gambar 6).



Gambar 5. Identifikasi faktor-faktor yang berkontribusi terhadap nilai *calorific value*

Tabel 1. Contoh data yang digunakan dalam penelitian

PRODUCT	CV(AR)	STOCKPILE	KROM	TPS100	TPS200	TPSHICA	TPS300	TPSA	PB600	PB700	TSC33	TSC22	TSCA	TSN100	TSN200	RF	RH	SVY	WTSP	CHM
E4900	4838	3702	1564	1065	276	0	828	145	387	0	283	135	588	1687	0	14.50	13.50	Geoservice	90	391
E4700	4672	9872	217	523	139	0	432	0	120	0	1376	0	982	1372	0	0.50	1.55	Sucofindo	52	225
E4900	4851	1994	1600	1589	1768	0	542	0	2430	0	1912	0	2498	0	969	0.00	0.00	Geoservice	151	659
E4900	4951	2240	200	3432	144	3896	270	0	2601	0	0	142	2301	0	2.00	3.73	Sucofindo	147	643	
E4900	4859	5515	1534	1446	0	1852	1773	270	0	0	878	2230	1452	0	2.00	3.73	Sucofindo	69	303	
E4700	4790	4385	200	1513	0	2499	0	0	0	2065	1234	685	682	275	963	0.00	0.00	Sucofindo	92	406
E4700	4832	4413	400	1610	0	2815	0	0	0	1883	417	856	1424	706	1832	0.00	0.00	Sucofindo	122	527



Gambar 6. Effect Plot masing-masing parameter yang berkontribusi terhadap variasi nilai calorific value

MEMILIH TEKNIK PEMODELAN DATA

Teknik regresi dipilih karena serangkaian faktor-faktor kontribusi / variabel prediktor (x_1 hingga x_n) mengakibatkan perubahan untuk setiap nilai calorific value / variabel respon (y) sebagai kelipatan dari perubahan setiap variabel prediktor (Nisbet, 2018). Kemudian pemilihan dan penentuan algoritma regresi apa yang paling tepat untuk jenis data yang kita miliki sangatlah penting, agar peneliti tidak kehilangan informasi utama dari data yang dimiliki, karena sangat tidak praktis untuk bereksperimen dengan

banyak teknik pada dataset tertentu hanya untuk menemukan algoritma terbaik (Doan, 2015). Penentuan dan pemilihan algoritma regresi terbaik didasarkan pada hasil uji statistik gabungan menggunakan SAR metric (Squared Error, Accuracy, ROC) Persamaan (1) dan Root Mean Squared Error (RMSE) Persamaan (2) pada kategori algoritma General Linear Model, Tree-based Model, dan State-of-the-Art Model pada synthetic data dan real-world data (Doan, 2015)

$$SAR = [Accuracy + AUC + (1 - RMSE)/3] \tag{1}$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \tag{2}$$

Dari metrik gabungan tersebut diperoleh algoritma terbaik dari kategori Tree-based Model adalah Cubist, Conditional D.T, Random Forest, sedangkan dari katategori State-of-the-Art Model adalah MARS, Neuron Net, SVR (Tabel 2 dan 3). Setiap algoritma ini akan kami gunakan

sebagai model pengujian akurasi prediksi pada dataset yang dimiliki menggunakan performance metrics typology chart (Gambar 7) berupa Mean Absolute Error (MAE) Persamaan (3), dan Mean Absolute Relative Error (MARE) Persamaan (4) secara bersamaan (Butchkarev, 2018).

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_j| \tag{3}$$

$$MARE = \frac{1}{n} \sum_j \frac{|e_j|}{|A_j|} \tag{4}$$

Tabel 2. RMSE by multiple regression models

Tree Models	RMSE	Other Models	RMSE
CART	0.9291	Neuron Net	0.9699
Conditional D.T.	0.9166	SVR	0.9714
Model Tree	0.9308	KNN	0.9692
Rule Based System	0.9166	PLS	0.9669
Bagging CART	0.9251	Ridge	0.5731
Random Forest	0.9216	LARS	0.9668
Gradient Boost	0.9439	Elastic Net	0.9668
Cubist	0.9025	MARS	0.9626

Tabel 3. Spearman's ranking correlation

Tree Models	SRC	Other Models	SRC
CART	0.7721	Neuron Net	0.8104
Conditional D.T.	0.8856	SVR	0.7246
Model Tree	0.6438	KNN	0.4849
Rule Based System	0.6438	PLS	0.5619
Bagging CART	0.8374	Ridge	0.5731
Random Forest	0.8977	LARS	0.5790
Gradient Boost	0.3470	Elastic Net	0.5790
Cubist	0.8807	MARS	0.8385

Point Distance, D	Normalization, N					Aggregation, G
	N1 = 1 Unitary	N2 = A_j^{-c} By Actual Values	N3 = $(A_j - \bar{A})^{-c}$ By Variability of Actual Values	N4 = $(A_j - P_j)^{-c}$ By Sum of Actual and Predicted Values	N5 = $[\max(A_j, P_j)]^{-c}$ By Max (or Min) of Actual and Predicted Values	
Error (magnitude of error) $D1 = A_i - P_i$	ME (MBE, bias)	MNB $c=1$ MPE = 100MNB		FB $c=1$		G1 Mean
						G2 Median
						G3 Geometric Mean
	MD					G4 Sum
Absolute error $D2 = A_i - P_i $	MAE (MAD)	MARE $c=1$ MAPE = 100MARE	MRAE $c=1$	FAE $c=1$ sMAPE = 100FAE		G1 Mean
	MdAE	MdAPE $c=1$	MdRAE $c=1$	sMdAPE $c=1$		G2 Median
	GMAE		GMRAE $c=1$			G3 Geometric Mean
	SAD		RAE $c=1$	CM $c=1$	WHD $c=1$	G4 Sum
Squared error $D3 = (A_i - P_i)^2$	MSE	MSPE $c=2$				G1 Mean
		MdsPE $c=2$				G2 Median
	GRMSE					G3 Geometric Mean
	SSE	NCS D $c=1$	RSE $c=2$	SquD $c=1$ DivD $c=2$	VSD $c=1$	G4 Sum
Log equation error $D4 = \ln(P_i/A_i)$						G1 Mean
	MdLAR					G2 Median
						G3 Geometric Mean
		KLD $c=1$				G4 Sum
Absolute Log quotient error $D5 = \ln(P_i/A_i) $	MNAFE					G1 Mean
	MdSA					G2 Median
						G3 Geometric Mean
						G4 Sum

Gambar 7. Performance metrics (error measures) typology chart (Butchkarev, 2018)

MEMBUAT DESAIN UJI DAN MENILAI MODEL

Setiap proyek *Machine Learning* (ML) harus memiliki tujuan dan metrik keberhasilan yang jelas. Metrik keberhasilan dapat berupa metrik bisnis dan atau teknis. Mengevaluasi metrik bisnis seringkali hanya dapat dievaluasi setelah model ML diproduksi. Di sisi lain, mengevaluasi metrik teknis akan lebih mudah dan dapat dilakukan selama fase pengembangan. Kami ingin mencapai metrik teknis terbaik karena ini

adalah sesuatu yang dapat dioptimalkan. *Hyperparameter Tuning* adalah proses mencari kumpulan *hyperparameter* terbaik dari kumpulan kandidat model ML. Tujuan penyetalan *hyperparameter* hanyalah untuk mendapatkan skor evaluasi maksimum pada set validasi tanpa menyebabkan masalah *overfitting* (Owen, 2022). Kami mengimplementasikan metode *Hyperparameter Tuning* pada setiap algoritma yang dipilih (Gambar 8) dan mengevaluasi akurasi prediksi terbaik pada masing-masing algoritma (Gambar 9).

Hyperparameter Tuning

```

from sklearn.model_selection import RandomizedSearchCV

#Randomized Search CV
n_estimators = [int(x) for x in np.linspace(start=100, stop=1200, num=12)]

#Various learning rate parameter
learning_rate = ['0.05', '0.1', '0.2', '0.3', '0.5', '0.6']

#Maximum number of levels in tress
max_depth = [int(x) for x in np.linspace(5, 30, num=6)]

#Subsample parameter value
subsample = [0.7, 0.6, 0.8]

#Minimum child weight
min_child_weight = [3, 4, 5, 6, 7]

#Create the random grid
random_grid = {'n_estimators': n_estimators,
               'learning_rate': learning_rate,
               'max_depth': max_depth,
               'subsample': subsample,
               'min_child_weight': min_child_weight}

regressor = xgb.XGBRegressor()
xg_random = RandomizedSearchCV(estimator = regressor,
                              param_distributions = random_grid,
                              scoring = 'neg_mean_squared_error',
                              n_iter = 100,
                              cv = 5,
                              verbose = 2,
                              random_state = 42,
                              n_jobs = 1)

xg_random.fit(x_train,y_train)

gpu_id=None, grow_policy=None,
importance_type=None,
interaction_constraints=None,
learning_rate=None, max_bin=None,
...
predictor=None, random_state=None,
reg_alpha=None, reg_lambda=None, ...),
n_iter=100, n_jobs=1,
param_distributions={'learning_rate': ['0.05', '0.1', '0.2',
                                       '0.3', '0.5', '0.6'],
                    'max_depth': [5, 10, 15, 20, 25, 30],
                    'min_child_weight': [3, 4, 5, 6, 7],
                    'n_estimators': [100, 200, 300, 400,
                                      500, 600, 700, 800,
                                      900, 1000, 1100,
                                      1200],
                    'subsample': [0.7, 0.6, 0.8]},
random_state=42, scoring='neg_mean_squared_error',
verbose=2)

```

Gambar 8. Penerapan *Hyperparameter Tuning* pada algoritma yang dipilih

Simple Linear

```

#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))

MAE : 66.20155365631
MAPE : 0.01467971472198625

```

Random Forest

```

#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))

MAE : 40.40316448557762
MAPE : 0.008774392396391554

```


Artificial Neural Network

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 813.2523631155972
MAPE : 0.17344774801333784

MARS

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 46.59937935938692
MAPE : 0.01017871717252874

Cubist

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 40.29576617822765
MAPE : 0.008747977507567072

Support Vector Regressor

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 147.06976243972034
MAPE : 0.03434994458863547

Decision Tree Regressor

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 57.69763058696823
MAPE : 0.012541465565915871

XGBoost

```
#Regression evaluation metrics
from sklearn import metrics
print('MAE :',metrics.mean_absolute_error(y_test, prediction))
print('MAPE :',metrics.mean_absolute_percentage_error(y_test, prediction))
```

MAE : 37.535070794207726
MAPE : 0.008171991484559752

Gambar 9. Menguji masing-masing algoritma prediksi untuk memodelkan data yang dimiliki

Dari pengujian performa akurasi dari masing-masing algoritma, kami memperoleh hasil terbaik saat mengimplementasikan prediksi menggunakan algoritma *XGBoost Regressor* (*Conditional Decision Tree*) dengan nilai MAE

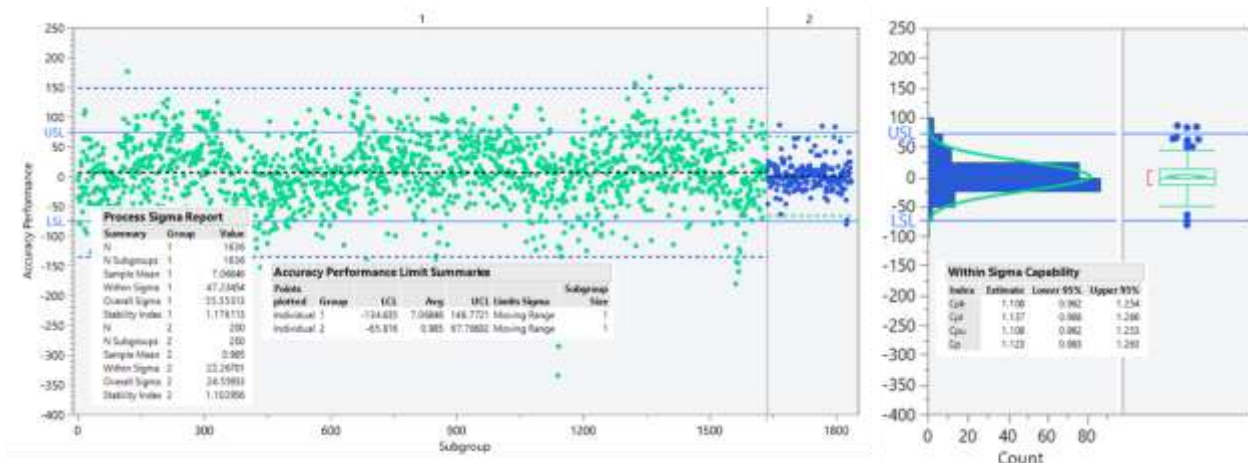
37.5, MAPE 0.0082, R_Squared 0.9705 (Tabel 4). Model pembelajaran ini akan diujikan pada 200 dataset baru untuk melihat sejauh apa performa akurasi yang dihasilkan di luar fase desain.

Tabel 4. Hasil *performance metrics* untuk masing-masing algoritma yang dipilih

Algorithm	MAE	MAPE = 100MARE	R_Squared
Artificial Neural Network	813.3	0.1734	0.2910
Cubist	40.3	0.0087	0.9771
Decision Tree Regressor	57.7	0.0125	0.9513
Multivariate Adaptive Regression Spline	46.6	0.0101	0.9656
Random Forest Regressor	40.4	0.0088	0.9765
Simple Linear Regression	66.2	0.0147	0.9305
Support Vector Regressor	147.1	0.0343	0.5759
XGBoost Regressor	37.5	0.0082	0.9705

Setelah pengujian pada dataset baru tersebut, kami mengevaluasi performa akurasi ke dalam *control chart* fase kedua (Gambar 9), yang menunjukkan perbaikan pada nilai rata-rata sebesar 0.9 kcal/kg dengan standar deviasi 24.6 kcal/kg, kemudian nilai C_p / derajat kepresisian algoritma prediksi

sebesar 1.123 dan C_{pk} / derajat keakurasian sebesar 1.108. Dari nilai tersebut menunjukkan adanya peningkatan *process capability* pada akurasi blending kualitas batubara yang signifikan setelah penerapan



Gambar 9. *Control Chart* dan *Process Capability* untuk akurasi blending setelah menerapkan algoritma prediksi

Kami menilai bahwa metode prediksi nilai akhir pencampuran kualitas batubara menggunakan algoritma prediksi ini sangat handal dibandingkan dengan hanya menggunakan metode rerata terbobot yang belum banyak mempertimbangkan variabel prediktor di luar faktor material. Sehingga para pelaku maupun profesional yang menggeluti bidang ini dapat segera beralih menggunakan metode yang lebih komprehensif, karena setiap peningkatan akurasi nilai pencampuran kualitas batubara akan langsung berimplikasi pada tata kelola rantai

pasokan yang memberikan keuntungan tambahan dari sisi keuangan dan kepuasan pelanggan.

KESIMPULAN

Beberapa poin yang dapat disimpulkan dari penelitian ini antara lain:

1. Pemilihan algoritma regresi terbaik dari *Tree-based Model* adalah (1) *Cubist*, (2) *Conditional Decision Tree*, (3) *Random Forest*, sedangkan dari kategori *State-of-the-*

Art Model adalah (1) *MARS*, (2) *Neuron Net*, (3) *SVR* didasari nilai *SAR metric* (*Squared Error, Accuracy, ROC*) dan nilai *Root Mean Squared Error* (RMSE) terkecil.

2. Pemilihan *performance metrics* yang sesuai dengan obyektif penelitian merujuk pada *typology chart* point distance kategori absolute error adalah *Mean Absolute Error* (MAE), dan *Mean Absolute Relative Error* (MARE).
3. Diperoleh *performance metrics* terbaik pada saat mengimplemtasikan algoritma *XGBoost Regressor* (*Conditional Decision Tree*) dengan nilai MAE 37.5 dan MARE 0.0082.
4. Terjadi peningkatan *process capability* pada akurasi blending kualitas batubara dari rata-rata 7.1 kcal/kg menjadi 0.91 kcal/kg, standar deviasi dari 55.6 kcal/kg menjadi 24.6 kcal/kg, C_p / derajat kepresisian dari 0.529 menjadi 1.123 dan C_{pk} / derajat keakurasian dari 0.479 menjadi 1.108.

UCAPAN TERIMA KASIH

Pada kesempatan ini, kami sampaikan terima kasih sebesar-besarnya kepada PERHAPI karena telah menyelenggarakan TPT XXXI PERHAPI 2022. Acara ini sangat mendukung pengembangan ilmu pengetahuan dan memacu kemunculan terobosan baru di bidang pertambangan. Tidak lupa juga kami ucapkan terima kasih kepada seluruh pihak yang telah mendukung penulisan hingga penerbitan makalah ini, semoga Tuhan Yang Maha Esa menganugerahkan balasan yang setimpal atas kebaikan Bapak/Ibu sekalian.

DAFTAR PUSTAKA

Artificial Neural Network – Rectified Linear Unit (ReLU), data diperoleh melalui situs internet:

<https://keras.io/api/layers/activations/>.

Diakses pada tanggal 10 Juni 2022.

Barrentine, L. (2003): Possible Sources of Process Variation, *Concept for R&R Studies*, ASQ Quality Press, Wisconsin, USA, 1 – 2.

Botchkarev, A. (2018): Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology, *Interdisciplinary Journal of Information, Knowledge, and Management*, 17 – 18.

Chapman, P. (2000): Step by Step Data Mining Guide, *CRISP-DM 1.0*, CRISP-DM Consortium: NCR Systems Engineering Copenhagen, USA, 12 - 64.

Doan, T. (2015): Selecting Machine Learning Algorithms using Regression Models, *IEEE 15th International Conference on Data Mining Workshops*, 1 - 6.

Effect Summary Report in Fitting Linear Models, data diperoleh melalui situs internet:

[https://www.jmp.com/support/help/en/16.2/index.shtml#page/jmp/effect-](https://www.jmp.com/support/help/en/16.2/index.shtml#page/jmp/effect-summary.shtml)

[summary.shtml](https://www.jmp.com/support/help/en/16.2/index.shtml#page/jmp/effect-summary.shtml). Diakses pada tanggal 20

Mei 2022.

International Organization for Standardization. (2009). *Solid mineral fuels – Determination of gross calorific value by bomb calorimetric method and calculation of net calorific value* (ISO Standard No. 1928:2009).

Multivariate Adaptive Regression Splines (MARS), data diperoleh melalui situs internet: <https://github.com/scikit-learn-contrib/py-earth>. Diakses pada tanggal 9

Agustus 2022.

Nisbet, R. (2018): Numerical Prediction, *Handbook of Statistical Analysis and Data Mining Application*, Elsevier, London Wall, United Kingdom, 187 - 188.

Owen, L (2022): Implementing Random Search, *Hyperparameter Tuning with Python*, Packt Publishing, Birmingham, UK, 121 – 123.

Process Capability Analysis, data diperoleh melalui situs internet: https://www.jmp.com/en_gb/learning-library/topics/quality-and-process/process-capability-analysis.html. Diakses pada tanggal 9 Agustus 2022.

Quinlan's Cubist Regression Model, data diperoleh melalui situs internet: <https://pypi.org/project/cubist/>. Diakses pada tanggal 9 Agustus 2022.

Random Forest Regressor, data diperoleh melalui situs internet: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. Diakses pada tanggal 9 Agustus 2022.

Support Vector Regression, data diperoleh melalui situs internet: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>. Diakses pada tanggal 9 Agustus 2022.

Variables Control Charts – I/MR Charts, data diperoleh melalui situs internet: https://www.jmp.com/en_us/learning-library/topics/quality-and-process/variables-control-charts-i-mr-charts.html. Diakses pada tanggal 9 Agustus 2022.