

XGBoost parameter selection Report - 西条

Prefecture code	38
Station code	38206050
Station name	西条
Target item	Ox(ppm)
Number of data points in the train set	15752
Number of data points in the test set	6751
Number of features used	140
Model	XGBoost
Objective	reg:squarederror
Booster	None
Parameter Grid (tested) n_estimators	[900, 1000, 1100]
Parameter Grid (tested) max_depth	[4, 5, 6]
Parameter Grid (tested) learning_rate	[0.04, 0.045, 0.05]
Best Parameter (found) learning_rate	0.045
Best Parameter (found) max_depth	5
Best Parameter (found) n_estimators	1000
Predictions mean	0.3183234
Predictions std	0.16849819
Real mean	0.31998179756845135
Real std	0.17555137121840067
Ljung-Box residuals autocorrelation, Prob(Q)	9.06391681326723e-22
Residuals skew	-0.0010966741945963815
Residuals kurtosis	8.088388521932865

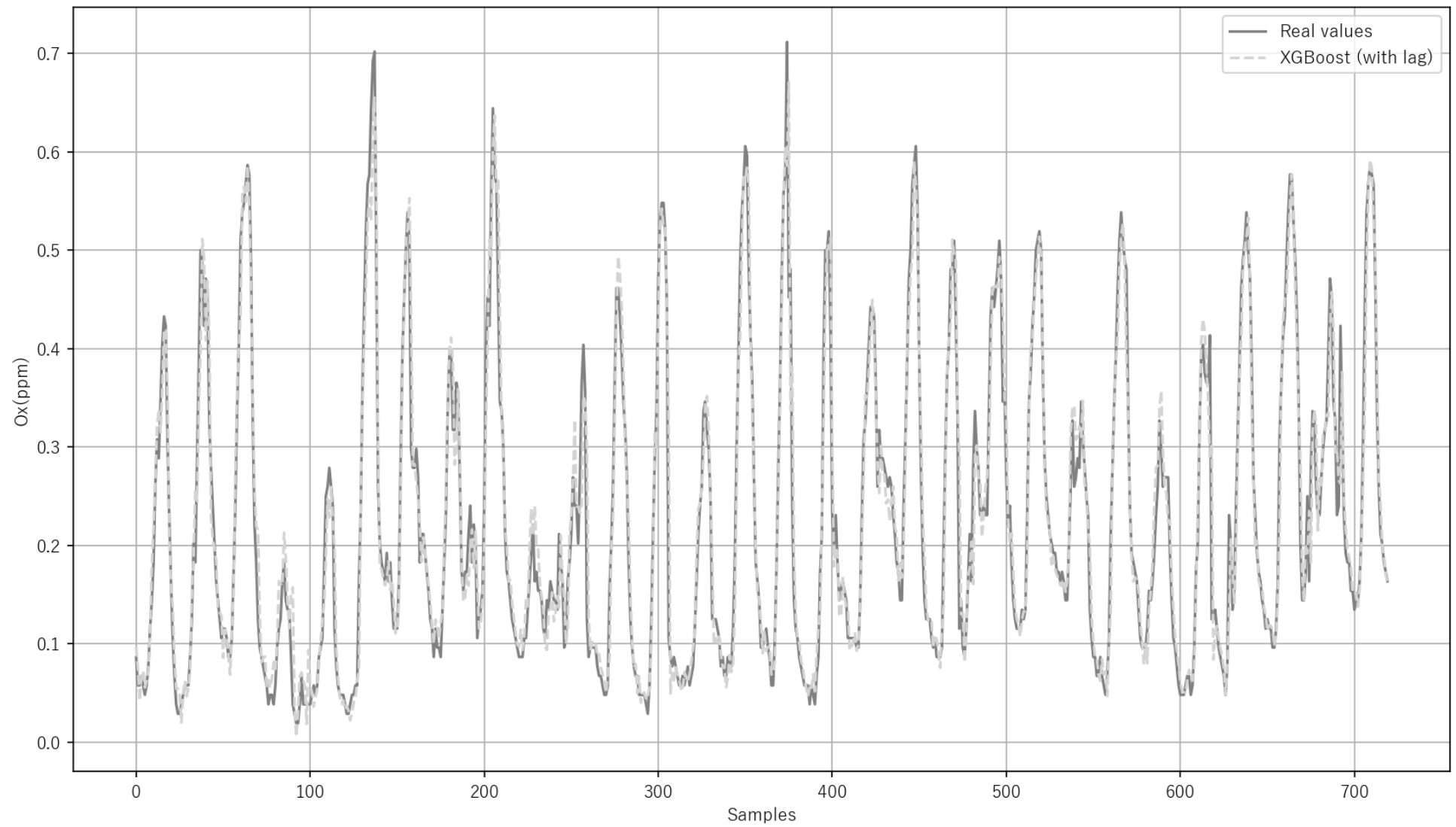
Features used for prediction

NO(ppm)	NO2(ppm)	U	V	Ox(ppm)_lag1
Ox(ppm)_lag2	Ox(ppm)_lag3	Ox(ppm)_lag4	Ox(ppm)_lag5	Ox(ppm)_lag6
Ox(ppm)_lag7	Ox(ppm)_lag8	Ox(ppm)_lag9	Ox(ppm)_lag10	Ox(ppm)_lag11
Ox(ppm)_lag12	Ox(ppm)_lag13	Ox(ppm)_lag14	Ox(ppm)_lag15	Ox(ppm)_lag16
Ox(ppm)_lag17	Ox(ppm)_lag18	Ox(ppm)_lag19	Ox(ppm)_lag20	Ox(ppm)_lag21
Ox(ppm)_lag22	Ox(ppm)_lag23	NO(ppm)_lag1	NO(ppm)_lag2	NO(ppm)_lag3
NO(ppm)_lag4	NO(ppm)_lag5	NO(ppm)_lag6	NO(ppm)_lag7	NO(ppm)_lag8
NO(ppm)_lag9	NO(ppm)_lag10	NO(ppm)_lag11	NO(ppm)_lag12	NO(ppm)_lag13
NO(ppm)_lag14	NO(ppm)_lag15	NO(ppm)_lag16	NO(ppm)_lag17	NO(ppm)_lag18
NO(ppm)_lag19	NO(ppm)_lag20	NO(ppm)_lag21	NO(ppm)_lag22	NO(ppm)_lag23
NO2(ppm)_lag1	NO2(ppm)_lag2	NO2(ppm)_lag3	NO2(ppm)_lag4	NO2(ppm)_lag5
NO2(ppm)_lag6	NO2(ppm)_lag7	NO2(ppm)_lag8	NO2(ppm)_lag9	NO2(ppm)_lag10
NO2(ppm)_lag11	NO2(ppm)_lag12	NO2(ppm)_lag13	NO2(ppm)_lag14	NO2(ppm)_lag15
NO2(ppm)_lag16	NO2(ppm)_lag17	NO2(ppm)_lag18	NO2(ppm)_lag19	NO2(ppm)_lag20
NO2(ppm)_lag21	NO2(ppm)_lag22	NO2(ppm)_lag23	U_lag1	U_lag2
U_lag3	U_lag4	U_lag5	U_lag6	U_lag7
U_lag8	U_lag9	U_lag10	U_lag11	U_lag12
U_lag13	U_lag14	U_lag15	U_lag16	U_lag17
U_lag18	U_lag19	U_lag20	U_lag21	U_lag22
U_lag23	V_lag1	V_lag2	V_lag3	V_lag4
V_lag5	V_lag6	V_lag7	V_lag8	V_lag9
V_lag10	V_lag11	V_lag12	V_lag13	V_lag14
V_lag15	V_lag16	V_lag17	V_lag18	V_lag19
V_lag20	V_lag21	V_lag22	V_lag23	Ox(ppm)_roll_mean_3
Ox(ppm)_roll_std_6	NO(ppm)_roll_mean_3	NO(ppm)_roll_std_6	NO2(ppm)_roll_mean_3	NO2(ppm)_roll_std_6
U_roll_mean_3	U_roll_std_6	V_roll_mean_3	V_roll_std_6	Ox(ppm)_diff_1
Ox(ppm)_diff_2	Ox(ppm)_diff_3	NO(ppm)_diff_3	NO2(ppm)_diff_3	U_diff_3
V_diff_3	hour_sin	hour_cos	dayofweek	is_weekend

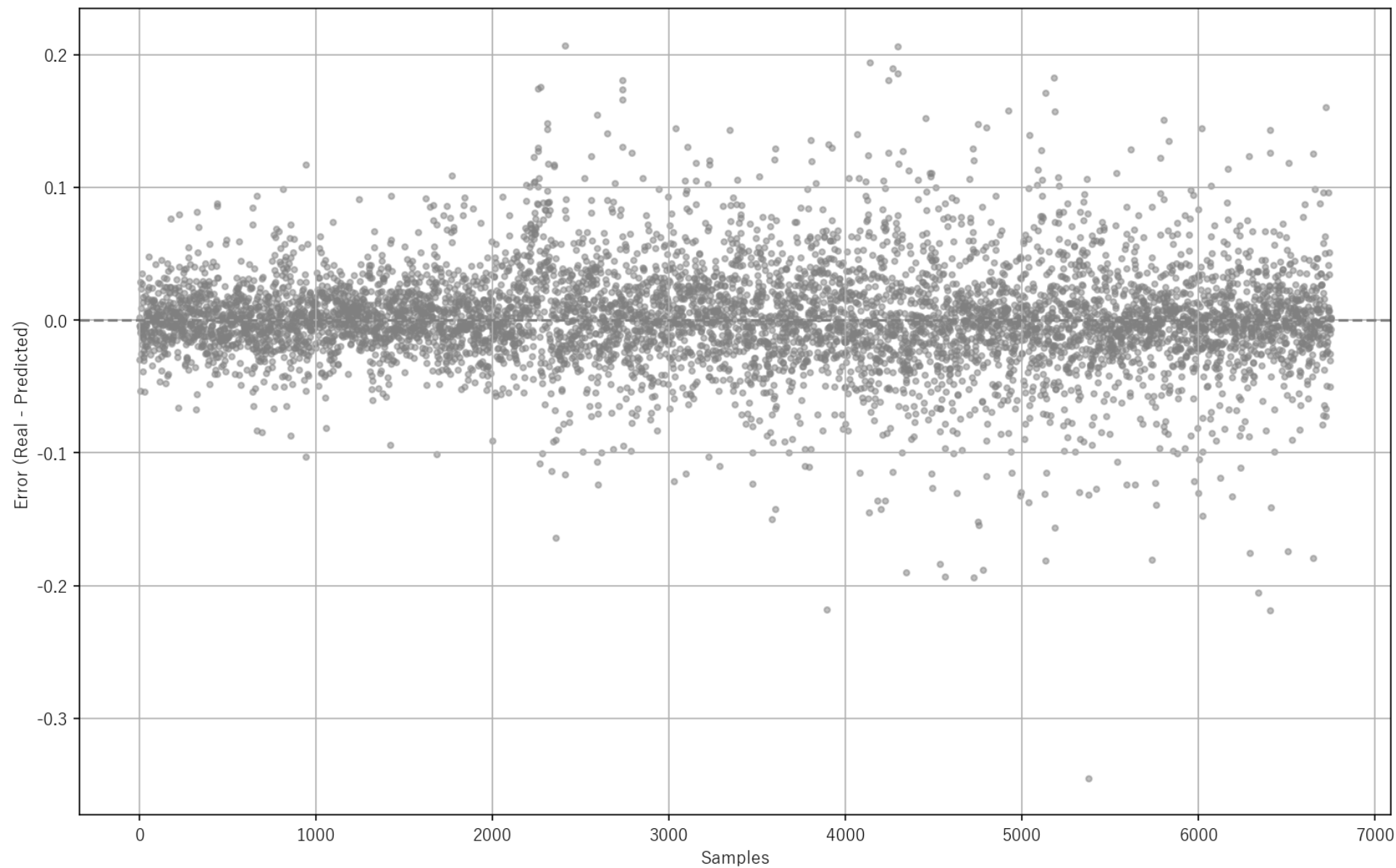
Model accuracy

Target	R ²	MAE	RMSE
Ox(ppm)	0.9548	0.0256	0.0373

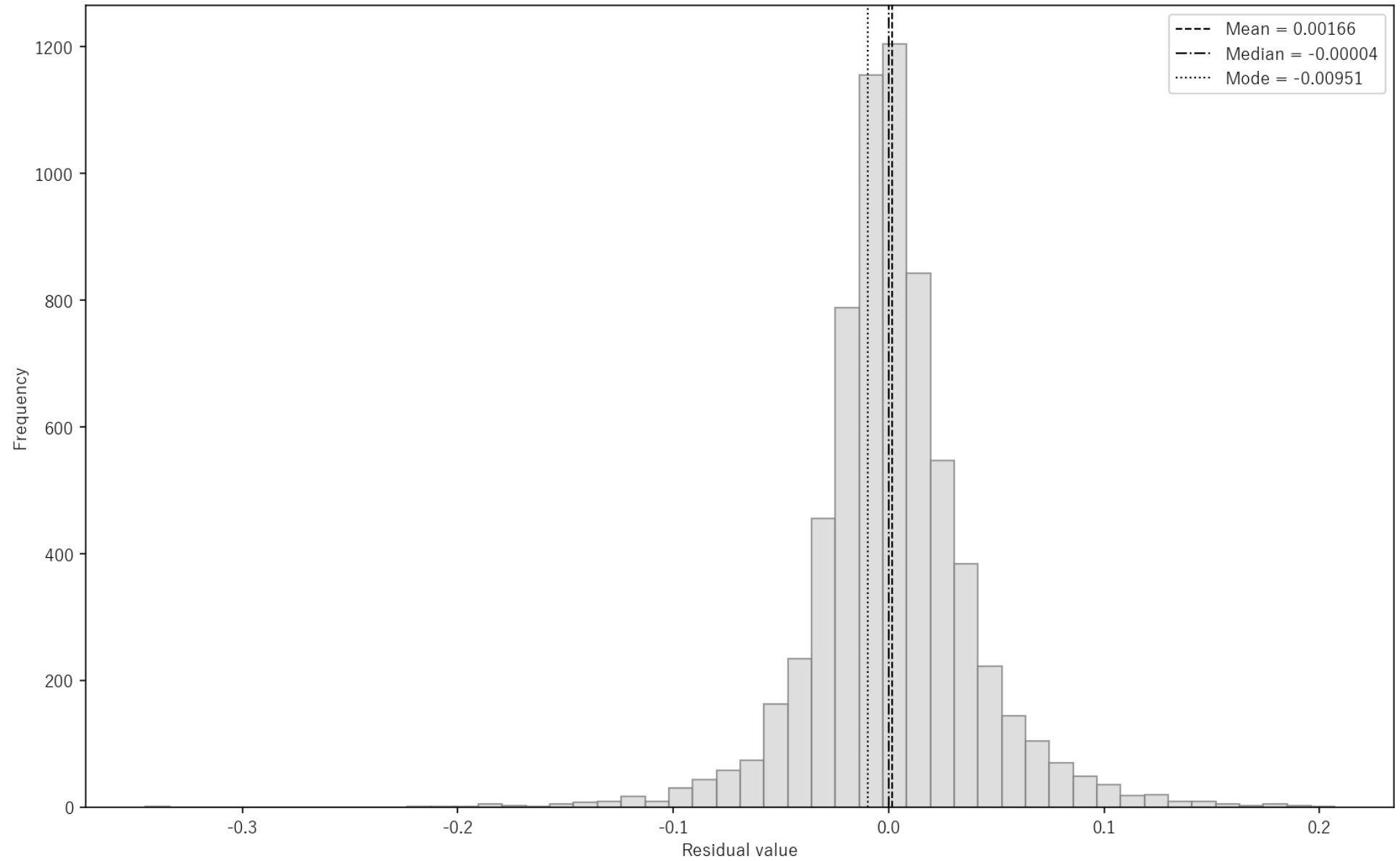
Regression with lag
 $R^2: 0.95477$



Distribution of Residual Errors



Histogram of Residuals – Distribution & Central Tendency



Top 20 Feature Importances (XGBoost)

