Development of AutoML based multifunctional complex dyeing sensor for energy saving

Jeong-In Lee, Jin-Soo Han, Wan-Ki Park
Energy ICT Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
{jilee, hanjinsoo, wkpark}@etri.re.kr

***Abstract*—Since the working process in the dyeing process is performed at high temperatures and high pressure, real-time measurement is difficult. Therefore, for real-time measurement of the dyeing process, this pH, conductivity, and chromaticity sensor was additionally installed, and a correlation and prediction model with the exhaustion rate that can determine the degree of dyeing completion was implemented based on Automated Machine Learning (AutoML) regression, and Extra tree with excellent performance indicators It was predicted using regressor, and the possibility of energy saving and process optimization was confirmed.**

***Keywords*—Dyeing, Automated Machine Learning, Sensor, Potential of Hydrogen, Conductivity, Choromatic**

I. INTRODUCTION

Since most dyeing factories operate dyeing machines in a high-temperature and high-pressure environment, it was difficult to check the dyeing status of the fabric in real time because the dyeing machine fabric inlet could not be opened in real time due to the risk of accidents caused by burns or pressure.

Therefore, the time and temperature required for the dyeing process may vary depending on the process know-how of the operator.

In this paper, a model of the multifunctional complex sensor that monitors the dyeing process status information in real time and digitizes the dyeing process is described. The multifunctional complex dyeing sensor model was developed based on Automated Machine Learning (AutoML), and the model implementation process and results are explained.

II. MODEL OF MULTIFUNCTIONAL COMPLEX DYEING SENSOR

The multifunctional complex dyeing model measures the dye solution's three sensor values (pH, conductivity, and chromaticity) and analyzes the correlation with the exhaustion rate of dyeing. The structure for calculating the real-time exhaustion rate using three sensor values was introduced in [1], [2].

Multifunctional complex staining models can apply various algorithms based on time series data. However, it is necessary to develop a learning model through processes such as data preprocessing, segmentation, feature extraction, and hyperparameter tuning, which are repeatedly performed for each algorithm. Figure 1 shows a typical machine-learning workflow.

In this paper, an optimized learning model was developed by applying Automated Machine Learning (AutoML) to develop an optimized learning model for a multifunctional complex dye sensor.

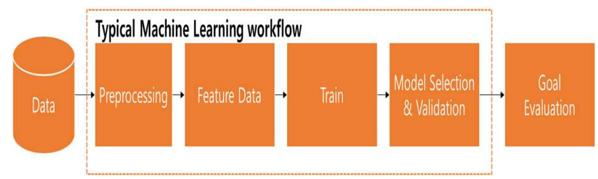

Figure 1. Typical Machine Learning Flow

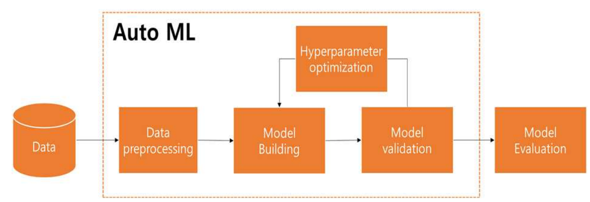

Figure 2 Automated Machine Learning Flow

Figure 2 shows the AutoML workflow, and it automates repetitive tasks of general machine learning to perform complex and repetitive tasks efficiently. AutoML is a Combined Algorithm Selection and Hyperparameter (CASH) optimization problem that explores the structure of various models when given a data Network Architecture Search, finds hyperparameters of a learning model, and automates the most appropriate model and variables [3].

For the multifunctional complex dye sensor model, an optimized algorithm was developed by limiting it to a regression analysis model specialized for time series data using the Pycaret library based on the Python scikit-learn package.

A. Settingup a model

The AutoML model environment was set up for learning using the setup( ) module. Table 1 shows the learning data setting information of the multifunctional complex dye sensor.

Table 1. AutoML setting up Environment

Description	Value
Session id	123
Target	Target
Target type	Regression
Original data shape	(1413, 4)
Transformed data shape	(1413, 4)
Transformed train set shape	(989, 4)
Transformed test set shape	(424, 4)
Numeric features	3
Preprocess	True
Imputation type	simple
Categorical imputation	mode
Fold Generator	KFold
Fold Number	10
CPU Jobs	-1
Use GPU	False
Log Experiment	False
Experiment Name	reg-default-name
USI	2562

The multifunctional complex dye sensor was limited to various regression analysis models specialized for time series and was trained, and Table 2 shows a list of regression analyses applicable to the multifunctional complex dye sensor model.

Table 2. The list of all models available for AutoML training

Name	Reference
Linear Regression	sklearn.linear_model._base.LinearRegression
Lasso Regression	sklearn.linear_model._coordinate_descent.Lasso
Ridge Regression	sklearn.linear_model._ridge.Ridge
Elastic Net	sklearn.linear_model._coordinate_descent.Elast...
Least Angle Regression	sklearn.linear_model._least_angle.Lars
Lasso Least Angle Regression	sklearn.linear_model._least_angle.LassoLars
Orthogonal Matching Pursuit	sklearn.linear_model._omp.OrthogonalMatchingPu...
Bayesian Ridge	sklearn.linear_model._bayes.BayesianRidge
Automatic Relevance Determination	sklearn.linear_model._bayes.ARDRegression
Passive Aggressive Regressor	sklearn.linear_model._passive_aggressive.Passi...
Random Sample Consensus	sklearn.linear_model._ransac.RANSACRegressor
TheilSen Regressor	sklearn.linear_model._theil_sen.TheilSenRegressor
Huber Regressor	sklearn.linear_model._huber.HuberRegressor
Kernel Ridge	sklearn.kernel_ridge.KernelRidge
Support Vector Regression	sklearn.svm._classes.SVR
K Neighbors Regressor	sklearn.neighbors._regression.KNeighborsRegressor
Decision Tree Regressor	sklearn.tree._classes.DecisionTreeRegressor
Random Forest Regressor	sklearn.ensemble._forest.RandomForestRegressor
Extra Trees Regressor	sklearn.ensemble._forest.ExtraTreesRegressor
AdaBoost Regressor	sklearn.ensemble._weight_boosting.AdaBoostRegr...
Gradient Boosting Regressor	sklearn.ensemble._gb.GradientBoostingRegressor
MLP Regressor	sklearn.neural_network._multilayer_perceptron....
Light Gradient Boosting Machine	lightgbm.sklearn.LGBMRegressor
Dummy Regressor	sklearn.dummy.DummyRegressor

B. Implementing a model,

Using AutoML, the code that recommends the learning model optimized for the data set is executed. The model with the best performance can be derived by comparing the MAE, MSE, RMSE, R2, RMSLE, and MAPE scores of each learning model result. Figure 3 shows that the model optimized for the multi-functional complex dye sensor learning model is the Extra Trees Regressor.

The Extra Tree Regressor model uses the entire training set for the training data of each decision tree and uses a subset of features to randomly split the nodes of the tree to reduce variance instead of increasing bias by increasing diversity and the computational speed is fast.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	15.3031	1344.1371	36.0923	0.9394	0.2700	0.1505	0.5500
rf	Random Forest Regressor	19.2495	1778.6029	42.0789	0.9199	0.3019	0.1886	0.5460
dt	Decision Tree Regressor	17.1882	2196.7022	46.4194	0.9014	0.3458	0.1658	0.5200
lightgbm	Light Gradient Boosting Machine	27.5061	2377.7848	48.5460	0.8923	0.3471	0.2606	0.5320
gbr	Gradient Boosting Regressor	35.1464	3020.6549	54.8802	0.8636	0.3752	0.3091	0.5360
knn	K Neighbors Regressor	40.5270	4474.6514	66.6618	0.7975	0.4254	0.3473	0.5280
ada	AdaBoost Regressor	64.6472	6488.1633	80.3955	0.7073	0.5202	0.5509	0.5420
lasso	Lasso Regression	75.1719	9914.0321	99.5507	0.5554	0.5850	0.6497	0.9720
br	Bayesian Ridge	75.1848	9914.1321	99.5511	0.5553	0.5849	0.6501	0.5200
en	Elastic Net	75.1733	9914.0843	99.5509	0.5553	0.5850	0.6497	0.5440
omp	Orthogonal Matching Pursuit	78.4596	10355.8787	101.7326	0.5347	0.6066	0.6984	0.5300
ridge	Ridge Regression	76.2409	11019.4232	104.3980	0.5083	0.5865	0.6525	0.9780
llar	Lasso Least Angle Regression	81.0107	11149.0860	105.5548	0.4999	0.6233	0.7543	0.5240
lr	Linear Regression	76.3498	11247.2377	105.2854	0.4986	0.5866	0.6528	1.1880
lar	Least Angle Regression	76.3498	11247.2377	105.2854	0.4986	0.5866	0.6528	0.5220
dummy	Dummy Regressor	127.2945	22398.9332	149.6120	-0.0038	0.8036	1.1525	0.5220
huber	Huber Regressor	78.5906	24708.2091	137.9016	-0.0748	0.5913	0.6657	0.5220
par	Passive Aggressive Regressor	157.7295	43370.4820	194.3007	-0.9000	0.9964	1.4283	0.5220

Figure 3. list of the Top performing models

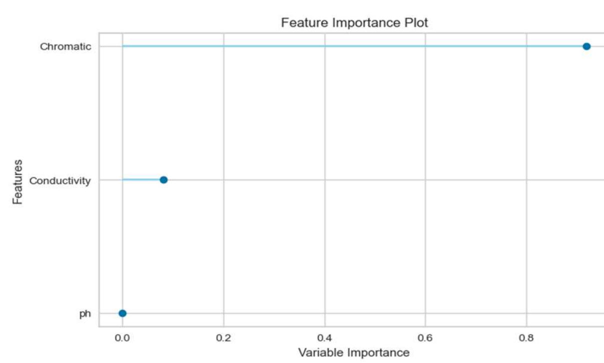

Figure 4. Feature Importance

Figure 4 shows the priority order of importance of the three sensor values used as input variables of the multifunctional complex dye sensor model, and it can be seen that the chromaticity value has the best influence in the process of predicting the target value.

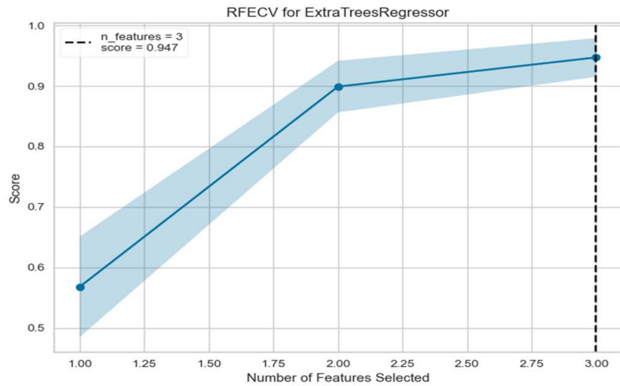

Figure 5. Feature selection using RFECV
(Recursive Feature Elimination with Cross Validation)

Figure 5 shows Recursive Feature Elimination with Cross Validation (RFECV) for feature selection using AutoML, and the highest performance is obtained when a learning model is created by selecting all pH, conductivity, and chromaticity values as independent variables of a multifunctional complex dye sensor model.

C. Training process

Figure 6 shows the learning curve of the multifunctional complex dye sensor model. It can be seen that both learning performance and verification performance improve with the number of times of learning.

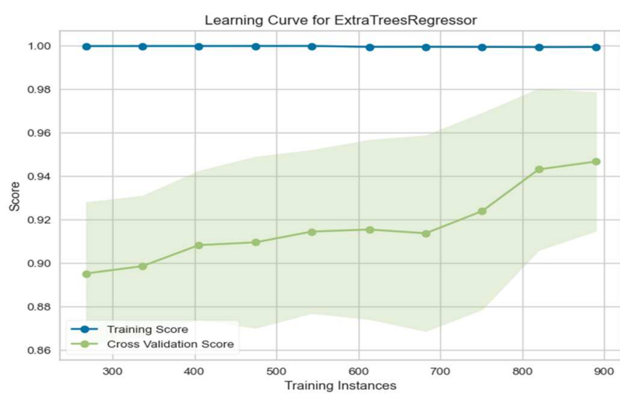

Figure 6. Learning Curve for the best performing model

Figure 7 shows the verification curve of the Extra Tree Regressor selected as a multifunctional complex dye sensor model. Looking at max-depth, as the depth of the tree increases, the training score and the test score increase, so it can be seen that a tree depth limit of less than depth 5 levels is necessary.

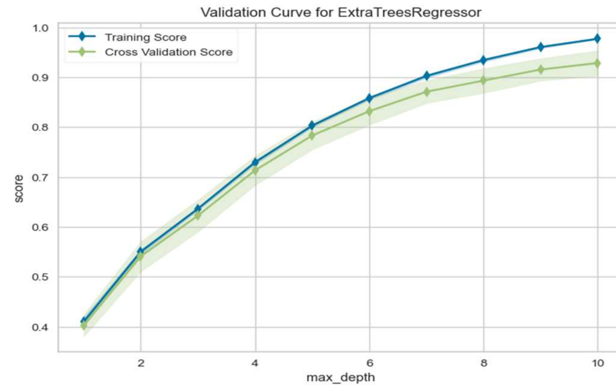

Figure 7. Validation Curve for the Best-performing model

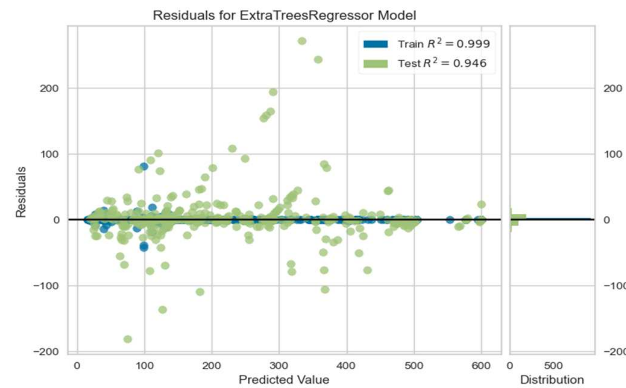

Figure 8. Residual for the best-performing model

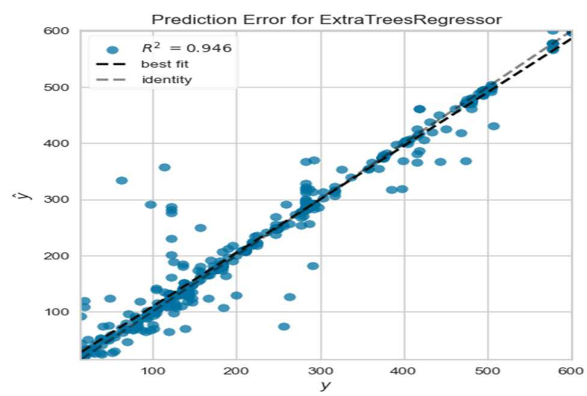

Figure 9. Prediction Error for the best-performing model

Figure 8 is the residual distribution of the multifunctional complex dye sensor model, and both train and test results are similarly distributed.

Figure 9 shows the prediction error of the extra tree regressor model. The R-squared score of the extra tree regressor is a model with an explanatory power of about 94% accuracy. The accuracy is higher if the model has more correct predictions.

III. RESULT

The multifunctional complex dye sensor learning model was implemented with the Extra Tree regressor derived based on AutoML, and the model fit is high through the distribution of residuals, and it can be seen that there is not much difference between the predicted result and the actual value.

Figure 10 and Figure 11 show the prediction results using performance indicators and actual data of the multifunctional complex dye sensor model.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Extra Trees Regressor	1.2254	21.5955	4.6471	0.9981	0.0182	0.0053

	no	concentration	configTemp	currentTemp	sensorTemp	Chromatic	ph	Conductivity	Target	prediction_label
0	463	2.1	80.099998	81	37	1942	4.2	463	32.099998	32.099998
1	463	2.1	82.500000	83	39	1939	4.2	482	34.400002	34.400002
2	463	2.1	84.900002	85	40	1939	4.2	484	36.799999	36.799999
3	463	2.1	87.300003	87	42	1937	4.2	484	39.099998	40.900000
4	463	2.1	89.800003	90	43	1937	4.2	484	42.700001	40.900000
...	...	...	...	...	...	...	...	...	...	...
75	463	2.1	120.000000	120	61	559	4.3	513	405.000000	405.000000
76	463	2.1	120.000000	119	61	562	4.3	512	401.399994	401.399994
77	463	2.1	120.000000	119	61	565	4.3	512	401.399994	401.399994
78	463	2.1	120.000000	119	62	566	4.3	512	401.399994	401.399994
79	463	2.1	120.000000	119	61	566	4.3	513	401.399994	401.399994

Figure 10. Prediction results for test dataset

Using the multifunctional complex dyeing sensor, the dyeing process can be predicted using real-time information about the dyeing solution, energy can be saved, and the possibility of process optimization was also confirmed.

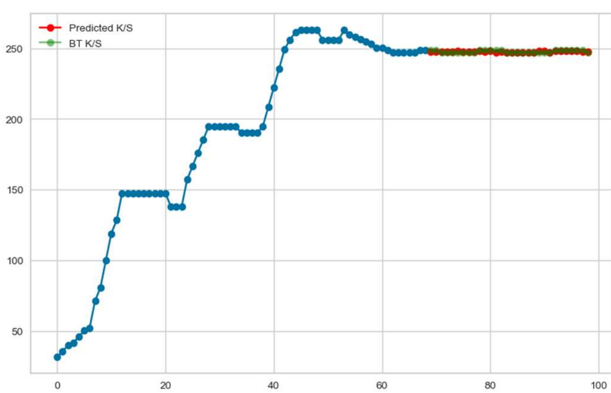

Figure 11. Best-performing model predictions

ACKNOWLEDGMENT

This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korea government (No. 20202000000010, Development and demonstration of artificial intelligence composite sensor to expand energy management systems).

REFERENCES

[1] Jeongin Lee, Wanki Park, and Sangha Kim. "Development of a composite sensor the exhaustion rate measurement model for the the exhaustion rate measurement of a dyeing factory." Journal of the Institute of Electrical and Electronics Engineers 26.2 (2022): 219-225.

[2] Jeongin Lee, Wanki Park, and Sangha Kim. "Data Analysis for Energy Savings in Dyeing and Processing Plants." Korean Society of Electronics Engineers Conference (2021): 976-977.

[3] Conrad, Felix, et al. "Benchmarking AutoML for regression tasks on small tabular data in materials design." Scientific Reports 12.1 (2022): 19350.