

An AutoML-driven Antenna Performance Prediction Model in the Autonomous Driving Radar Manufacturing Process

So-Hyang Bak¹ and Kwanghoon Pio Kim^{1*}

¹Data and Process Engineering Laboratory
Division of AI Computer Science and Engineering
KYONGGI UNIVERSITY

154-42 Kwangkyosan-ro Youngtong-gu Suwon-si Gyeonggi-do, 16227, South Korea
[e-mail: {sa7291, kwang}@kgu.ac.kr]

*Corresponding author: Kwanghoon Pio Kim

*Received April 11, 2023; revised September 14, 2023; accepted November 21, 2023;
published December 31, 2023*

Abstract

This paper proposes an antenna performance prediction model in the autonomous driving radar manufacturing process. Our research work is based upon a challenge dataset, Driving Radar Manufacturing Process Dataset, and a typical AutoML machine learning workflow engine, Pycaret open-source Python library. Note that the dataset contains the total 70 data-items, out of which 54 used as input features and 16 used as output features, and the dataset is properly built into resolving the multi-output regression problem. During the data regression analysis and preprocessing phase, we identified several input features having similar correlations and so detached some of those input features, which may become a serious cause of the multicollinearity problem that affect the overall model performance. In the training phase, we train each of output-feature regression models by using the AutoML approach. Next, we selected the top 5 models showing the higher performances in the AutoML result reports and applied the ensemble method so as for the selected models' performances to be improved. In performing the experimental performance evaluation of the regression prediction model, we particularly used two metrics, MAE and RMSE, and the results of which were 0.6928 and 1.2065, respectively. Additionally, we carried out a series of experiments to verify the proposed model's performance by comparing with other existing models' performances. In conclusion, we enhance accuracy for safer autonomous vehicles, reduces manufacturing costs through AutoML-Pycaret and machine learning ensembled model, and prevents the production of faulty radar systems, conserving resources. Ultimately, the proposed model holds significant promise not only for antenna performance but also for improving manufacturing quality and advancing radar systems in autonomous vehicles.

Keywords: Multi-output Regression, Antenna Performance Prediction, AutoML

A preliminary version of this paper appeared in ICONI 2022, Dec. 11-13, Jeju, Republic of Korea. This research was supported by the ATC+ R&D Project of KEIT (Korea Evaluations Institute of Industrial Technology) from the Ministry of Trade, Industry and Energy in 2020 (Grant No. 20009228, Development of Deep Learning-based Predictive Production Quality and Field Safety Management Technology to Support Level-4 Smart Factories in the Textile Industries).

1. Introduction

In recent years, the rapid advancement of technology and active research into artificial intelligence and the Internet of Things (IoT) has led to an increase in Research and development in the manufacturing industry [1][2]. These efforts aim to build smart factories and predict and improve manufacturing processes more efficiently and with higher quality, using machine learning and deep learning technologies. In the manufacturing process, lots of various data is generated, such as images, log data, and videos. Among them, research is active to improve efficiency through classification and regression prediction based on deep learning and machine learning using sensor-based process log data generated during the manufacturing process [3].

For instance, there have been several cases in which a machine learning-based method has been used to diagnose and classify defects in steel plate [4], and Research is being conducted to diagnose fabric defects or to improve the quality of textile processes efficiently using data generated during the textile process [5][6].

In particular, one significant area in which machine learning is being applied is the development of self-driving vehicles, including electric cars, robo-taxis, and self-driving robots [7]. In such vehicles, radar is an essential sensor used to measure critical factors such as distance to other vehicles, relative speed, and direction. Consequently, various studies are being conducted to improve the quality and performance of radar in the manufacturing process [8][9][10][11].

Therefore, in this paper, we propose an Antenna Performance Prediction Model of Autonomous Driving Radar Manufacturing Process using Automated Machine Learning (AutoML). Our study uses a Driving Radar Manufacturing Process Dataset that predicts antenna performance to improve the process quality of autonomous driving radar, with 56 input-features and 14 output-features, making it a multi-output regression problem. During the data regression analysis and preprocessing phase, we identified several input features having similar correlations and so detached some of those input features, which may become a serious cause of the multicollinearity problem that affect the overall model performance. In the training phase, we train each of output-feature regression models by using the AutoML-Pycaret approach which is an open-source machine learning library based on Python. To solve the multi-output regression problem, we selected the top 5 models showing the higher performances based upon RMSE in the AutoML-Pycaret result reports and applied the ensemble method so as for the selected models' performances to be improved. In our experimental performance evaluation, we used MAE and RMSE as performance evaluation metrics for the regression prediction model, and achieved results of 0.6928 and 1.2065, respectively.

The next of this paper is organized as follows: The following section, we describe the conceptual background of the radar manufacturing process, which is the data domain used in this study, and AutoML, the method used in the proposed model. Subsequent sections explain the Analysis and Pre-processing of the Radar Manufacturing Process Dataset and the Antenna Performance Prediction Model of Autonomous Driving Radar Manufacturing Process we developed. Finally, we compare the experimental results of the model proposed in this paper with other deep learning and machine learning models such as XGBoost, Random Forest, and TabNet to show the performance excellence of the proposed model. The last section then presents a summary of conclusions and future work.

2. Related Works

2.1 Theoretical background

Autonomous Driving Radar: Radar is an electromagnetic wave-based sensor [12]. It emits electromagnetic waves and extracts information such as distance from nearby objects, relative speed, and direction based on the reflected signal. Radar was developed to acquire distance information from airborne objects to ground observation posts via long-range wireless antennas. Over time, as its applications expanded, radar technology advanced to obtain information on the speed, size, and even shape of moving objects. Radars equipped with antennas are capable of transmitting and receiving electromagnetic wave signals, and to obtain distance information on stationary or moving targets, they must receive the reflected waves from the emitted electronic board [13]. Recently, with the development of radar technology, radar not only measures short, medium, and long distances depending on frequency, but various products such as indoor and 4D imaging radars have appeared and are being applied to various ADAS technologies like Self-driving car, robotics, smart cruise control, and emergency automatic control device [14][15]. As these sensors are applied to various technologies, there are many problems that can occur if there is a defect in the data product. This is because, for example, a radar is a defective product and cannot be recognized during autonomous driving, which can lead to a safety accident, such as causing great loss of life. Therefore, many companies go through the performance evaluation process to sort out defected products in the radar manufacturing process and strive to improve the quality of radar products and reduce product defects. Although the method and evaluation process for evaluating antenna performance may vary depending on the antenna manufacturing process, in general, the evaluation of antenna performance involves the detection of defects by assessing the RLC values of each Surface Mount Technology (SMT) component through the test points on the PCB. To ensure proper module operation, fundamental functions are verified by applying input power. Following the management of various components, an evaluation process is conducted to assess the impact on product performance by considering process and component characteristics.

Multicollinearity [16]: This means a strong correlation between independent variables. When performing regression analysis, since the precondition of regression analysis is that the explanatory variables are independent of each other, the assumption that the explanatory variables are independent of each other must be established to make the estimation of the regression coefficient more accurate. In the process of regressive analysis, if there is a correlation between explanatory variables and independence is doubtful, the presence or absence of multicollinearity should be determined. In order to solve this multicollinearity, you can remove independent variables with high correlation, transform variables, or used new observations. Alternatively, there is a method of selecting variables with high explanatory power using the principal component analysis method. Many Studies have been conducted to improve the performance of the model by resolving this multicollinearity problem [17][18].

Automated Machine Learning (AutoML) [19]: This is a rapidly growing field in artificial intelligence and machine learning that focuses on creating tools and techniques that can automate the process of building and deploying machine learning models. There are a few key components of an AutoML system. First, it must be able to automatically preprocess data, selecting and transforming the features of the dataset that are most relevant to the problem at hand. This includes handling missing data, encoding categorical features, and scaling

continuous features. Next, an AutoML system can automatically select the appropriate algorithm or set of algorithms to use for the given problem. This may involve comparing the performance of multiple algorithms on a given dataset, or even combining multiple algorithms to create an ensemble model. Finally, AutoML can also improve the quality of machine learning models. By automating the process of model selection and hyperparameter tuning, it can potentially identify better performing models than a human expert would be able to find manually. The dataset used in this study is autonomous antenna manufacturing process data for antenna performance prediction. Since AutoML has the advantage of maximizing efficiency and productivity by automatically applying various model algorithms in the machine learning development process, there are many cases of using AutoML to solve various classification and regression prediction problems, as well as research and development. In this research, the dataset we used is structured tabular data and includes multiple input features and multiple output features. Therefore, AutoML functionalities that enable hyperparameter tuning and the application of various machine learning algorithms can be used to develop a high-performance antenna performance prediction model. Additionally, we choose AutoML to easily and quickly implement ensemble techniques which combine the predictions of multiple models to enhance predictive accuracy and model robustness as utilized in this paper. Also, there are various libraries such as AutoKeras, AutoGLon, H2O, and Pycaret to use AutoML, and this study uses the Pycaret library [20].

2.2 Previous study

As previously mentioned, the presence of defects in the autonomous driving radar system can render recognition during autonomous driving impossible. This can lead to safety accidents causing severe casualties, as well as significant time and financial losses of manufacturing process. Consequently, several studies have been conducted to evaluate and predict the performance of radar antennas [21][22][23]. In recent, research and development to predict and analyze product defects through correlation analysis between manufacturing process data and product performance using artificial intelligence technologies such as deep learning and machine learning and maximize yield to reduce product disposal costs due to defects. For instance, Boya Liu et al. [24] proposed Bayesian network diagnosis method based on multi-source information fusion technology to solve the fault diagnosis problems during the radar manufacturing process. In addition, there is prior study to predict antenna performance defects using deep learning although it is not an antenna used in the autonomous driving radar manufacturing process. They proposed a method for predicting antenna performance and defect location using DNN, a basic deep neural network deep learning structure [25].

3. An Antenna Performance Prediction Model of Autonomous Driving Radar Manufacturing Process

In this section, we describe the overall procedure to predict the antenna performance, and the details of the proposed antenna performance prediction model in the autonomous driving radar manufacturing process. We utilized AutoML to implement the proposed model, specifically choosing the AutoML-PyCaret library from available AutoML libraries like AutoKeras, AutoGLon, H2O. Because Pycaret is not only easy to use, but also, we can use it to compare the performance of multiple models, as well as perform hyperparameter tuning and blending multiple models.

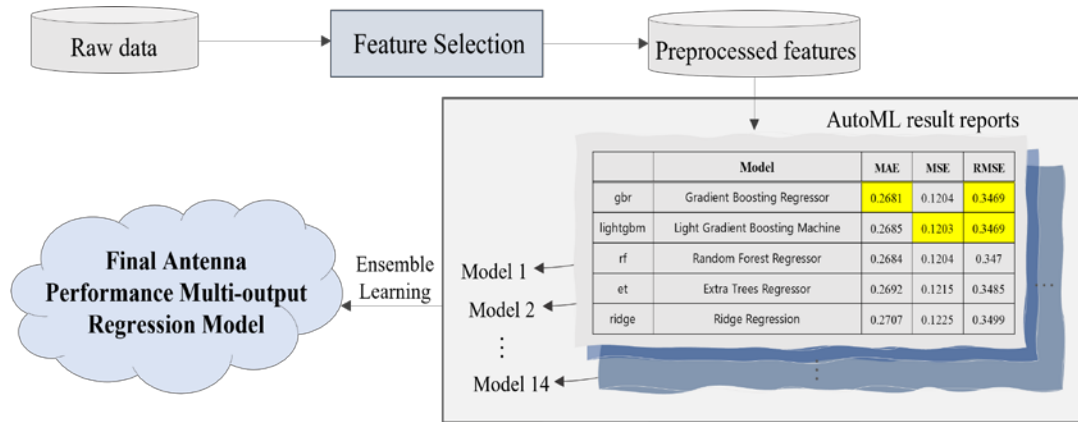


Fig. 1. Overall model development framework

Our study is based upon a dataset generated from the autonomous driving radar manufacturing process, which comprises 56 input features and 14 output features, thus constituting a multi-output regression problem. To carry out this study, it generally consists of a data analysis and preprocessing phase, an antenna performance prediction model training phase, and a performance evaluation and comparison phase to verify the performance of the model.

First, during the data analysis and preprocessing phases, we identified input features with high correlations and consequently removed some of them to mitigate the potential multicollinearity issues that could adversely affect the overall model performance. Second, in the training phase, (1) we individually trained regression models for each of the 14 output features using AutoML-PyCaret. Upon completion of this training process, we obtained 14 distinct result reports. (2) From these reports, we selectively blended the top-performing models, which were determined based on the root mean square error (RMSE). This careful selection ensured that the resulting ensemble models consistently delivered the highest prediction accuracy for each output feature. (3) While predicting each output feature in isolation yielded impressive results, our ultimate objective was to develop a robust multi-output regression prediction model capable of simultaneously forecasting all 14 output features. To achieve this, we perform ensemble learning with the 5 models that frequently show the highest performance based on RMSE among each of the 14 model outputs. Later, we conduct an evaluation and comparison experiment of the proposed model through experiments to verify the performance of the model in section 4.

Fig. 1 is the overall development process for the antenna performance multi-output regression prediction model of autonomous driving radar manufacturing process.

3.1 Driving Radar Manufacturing Process Dataset

The dataset used for this study is the Manufacturing Process Data [26] generated from the autonomous driving radar process provided by Dacon.io. Since the dataset we used in this study is a challenge e dataset provided by a company, data description such as the data's collection date and evaluation metrics are not disclosed, but it is a reliable, standardized dataset. The dataset contains the total 70 data-items, out of which 54 used as input features and 16 used as output features, and the dataset is properly built and well-fitted into resolving the multi-output regression problem. And the total number of the dataset is 39,607.

The 56 input-features are composed of various values such as pressing amount, weight and area of heat dissipation material, screw insertion length during the autonomous driving radar manufacturing process. In **Table 1**, shows the information on detailed input-features.

Table 1. Input-features which are independent variables used for learning.

| | Type | Input-features |
|----|--|--|
| 1 | Pressing amount when fastening PCB | X_01, X_02, X_05, X_06 |
| 2 | Heat dissipation material weight | X_03, X_10, X_11 |
| 3 | Whether or not the inspection passed | X_04, X_23, X_47, X_48 |
| 4 | Heat dissipation material area | X_07, X_08, X_09 |
| 5 | Coordinates relative to connector location | X_12 |
| 6 | Difference height of each antenna pad | X_13 |
| 7 | Antenna pad location | X_14, X_15, X_16, X_17, X_18 |
| 8 | Screw insertion depth | X_19, X_20, X_21, X_22 |
| 9 | Connector pin dimensions | X_24, X_25, X_26, X_27, X_28, X_29 |
| 10 | Screw insertion depth | X_30, X_31, X_32, X_33 |
| 11 | Revolutions per minute for screw fastening | X_34, X_35, X_36, X_37 |
| 12 | Housing PCB seat dimensions | X_38, X_39, X_40 |
| 13 | Radome Dimensions | X_41, X_42, X_43, X_44 |
| 14 | Antenna Part Radome Tilt | X_45 |
| 15 | Sealant Bond Requirement | X_46 |
| 16 | Waiting time before Cal injection | X_49 |
| 17 | RF Part SMT Lead Amount | X_50, X_51, X_52, X_53, X_54, X_55, X_56 |

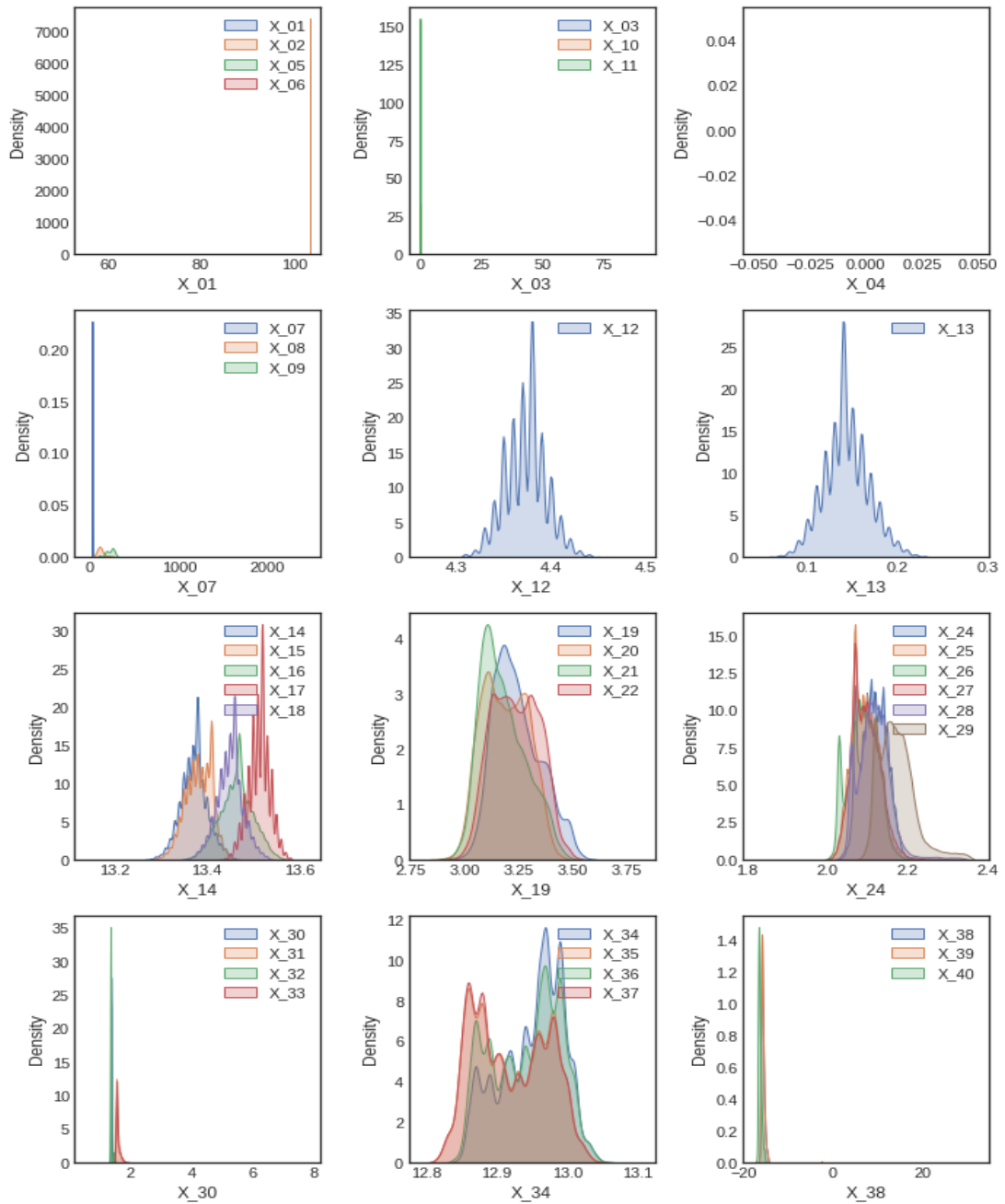
The 14 output-features are the antenna performance values of the radar sensor in the autonomous driving radar manufacturing process. We intend to predict these 14 antenna performance values in this study. **Table 2** shows detailed information about the 14 output-features we need to predict.

Table 2. Output-features which are dependent variables for predicting.

| | Type | Output-features |
|---|--|--|
| 1 | Antenna gain average according to the angle | Y_01, Y_05, Y_07, Y_11 |
| 2 | Antenna gain deviation | Y_02, Y_03 |
| 3 | Average signal-to-noise ratio | Y_04 |
| 4 | Signal to noise ratio according to the angle | Y_06, Y_08, Y_09, Y_10, Y_12, Y_13, Y_14 |

3.2 Data Analysis and Preprocessing

As a result of the analysis, there is a group of similarly correlated features, which can cause multicollinearity problems that can negatively affect performance due to high correlations between input features.



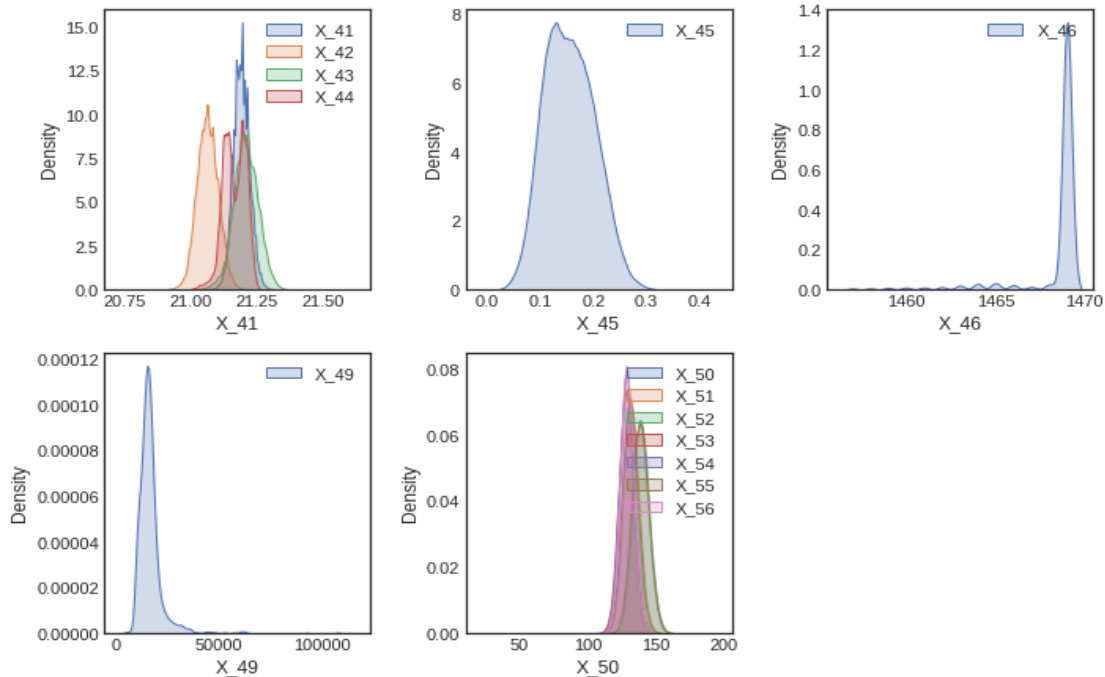


Fig. 2. Data Analysis - kernel density estimation for input-feature

Fig. 2 shows the kernel density estimation (KDE) [27] between input-features with similar distributions. In order to data preprocessing, first, the input-feature values of X_04, X_23, X_47, and X_58, which indicate whether the test passed, all have a value of 1, so we removed the features. And since the values of X_03, X_10, and X_11 input-features, which means the weight of the heat dissipating material, mostly have a value of 0, they are measured incorrectly, so we removed the features as well. In addition, we remove some input-features that show high correlation between groups with similar distributions and proceed with training with a total of 34 preprocessed input-features.

3.3 Antenna Performance Prediction Model using AutoML

In this study, we train the antenna performance prediction model using AutoML-Pycaret library, which automatically trains all machine learning models and drives optimized performance. We check AutoML-Pycaret result reports for Y_01 through Y_14. Then, we blend the models with the top 5 models that show the highest performance of AutoML-Pycaret based upon RMSE.

Table 3 indicates the results for Y_01 among the 14 AutoML-Pycaret result reports. Afterwards, in order to further improve the performance of the model and predict the 14 output-feature simultaneously, we conduct ensemble learning with the 5 models that frequently show the highest performance based on RMSE among each of the 14 model outputs. In other words, we developed the final ensemble model by performing hyperparameter tuning on random forest, LGBM, gradient boosting regressor, ridge, and lasso models, which most frequently showed high performance based upon RMSE in 14 AutoML-Pycaret result reports and performing ensemble learning using them.

Table 3. AutoML-Pycaret result report for output-feature Y_01

| | Model | MAE | MSE | RMSE | R2 | RMSLE | MAPE | TT (Sec) |
|----------|---------------------------------|---------|--------|---------|----------|--------|--------|----------|
| gbr | Gradient Boosting Regressor | 0.2681 | 0.1204 | 0.3469 | 0.0466 | 0.1526 | 0.2819 | 12.147 |
| lightgbm | Light Gradient Boosting Machine | 0.2685 | 0.1203 | 0.3469 | 0.0469 | 0.1525 | 0.2789 | 0.121 |
| rf | Random Forest Regressor | 0.2684 | 0.1204 | 0.347 | 0.046 | 0.1526 | 0.2779 | 5.506 |
| et | Extra Trees Regressor | 0.2692 | 0.1215 | 0.3485 | 0.0377 | 0.1534 | 0.2834 | 3.132 |
| ridge | Ridge Regression | 0.2707 | 0.1225 | 0.3499 | 0.0298 | 0.1541 | 0.2863 | 0.018 |
| br | Bayesian Ridge | 0.271 | 0.1228 | 0.3504 | 0.0274 | 0.1543 | 0.2867 | 0.092 |
| omp | Orthogonal Matching Pursuit | 0.2721 | 0.1236 | 0.3516 | 0.0206 | 0.1548 | 0.2874 | 0.012 |
| ada | AdaBoost Regressor | 0.2736 | 0.1244 | 0.3527 | 0.0144 | 0.1557 | 0.2889 | 3.667 |
| lr | Linear Regression | 0.2744 | 0.1254 | 0.3541 | 0.0066 | 0.1558 | 0.2887 | 0.043 |
| en | Elastic Net | 0.2747 | 0.1261 | 0.355 | 0.0015 | 0.1562 | 0.29 | 0.018 |
| lasso | Lasso Regression | 0.2747 | 0.126 | 0.355 | 0.0016 | 0.1562 | 0.29 | 0.015 |
| llar | Lasso Least Angle Regression | 0.2749 | 0.1263 | 0.3553 | -0.0003 | 0.1564 | 0.2906 | 0.012 |
| dummy | Dummy Regressor | 0.2749 | 0.1263 | 0.3553 | -0.0003 | 0.1564 | 0.2906 | 0.003 |
| xgboost | Extreme Gradient Boosting | 0.2781 | 0.1282 | 0.358 | -0.0153 | 0.157 | 0.2831 | 0.391 |
| knn | K Neighbors Regressor | 0.2981 | 0.147 | 0.3833 | -0.1642 | 0.1677 | 0.3058 | 0.33 |
| huber | Huber Regressor | 0.3285 | 0.2198 | 0.4683 | -0.74 | 0.1802 | 0.3262 | 0.638 |
| dt | Decision Tree Regressor | 0.393 | 0.2524 | 0.5023 | -0.9997 | 0.2199 | 0.365 | 0.765 |
| par | Passive Aggressive Regressor | 0.3976 | 0.2615 | 0.5026 | -1.0748 | 0.2169 | 0.363 | 0.068 |
| lar | Least Angle Regression | 98.6234 | 157247 | 131.376 | -1203781 | 1.0429 | 85.567 | 0.021 |

4. Experiments and Results

4.1 Experiment environments

The proposed antenna performance prediction model of autonomous driving radar manufacturing process is implemented in the experiment environment of operating system and hardware as shown in [Table 4](#).

Table 4. Experiment environment

| | | |
|-------------------|------------------|------------------------------------|
| Operating System | Ubuntu 20.04 LTS | |
| Computer Hardware | CPU | AMD Ryzen 7 5800X 8-Core Processor |
| | GPU | NVIDIA GeForce RTX 3070 |
| | RAM | 32 GB |

| | |
|----------|--------------------------------|
| Language | Python 3.7 |
| CUDA | CUDA 11.2 cuDNN 8.1 |
| Library | Scikit-learn, Pytorch, Pycaret |

We use Ubuntu 20.04 LTS, AMD Ryzen 7 5800X 8-Core Processor, RAM 32GB, and NVIDIA GeForce RTX 3070. As for software, we use Scikit-learn, Pytorch, and Pycaret library for the proposed model implementation and comparison experiments. Scikit-learn Library is a Python library for machine learning. It provides a range of supervised and unsupervised learning algorithms for tasks such as classification, regression, clustering, and dimensionality reduction, as well as tools for model selection, preprocessing, and evaluation. We use Scikit-learn to implement the proposed model and machine learning models such as Decision Tree [28], Random Forest [29], LGBM [30], Linear Regression [31], XGBoost [32], and to evaluate and compare the model performances. Also, we use the Pytorch library to implement the TabNet [33] deep-learning model for comparison with our proposed model. We installed CUDA, including CUDA 11.2 and cuDNN 8.1, to accelerate GPU for fast learning of deep learning models. Lastly, the model we proposed uses AutoML to train machine learning models, and select 5 models with the best performance for ensemble learning. To implement this, the Pycaret library, which provides AutoML function, is used.

4.2 Evaluation Metrics

In this study, we implement the model that predicts antenna performance in Autonomous Driving Radar Manufacturing Process. The model training is based on learning from the performance values (y_i -output features) obtained through the evaluation process and the log values (x_i -input features) generated during antenna manufacturing. In other words, the model is a multiple regression problem that obtains 14 output features, such as the average and deviation of antenna gain and signal-to-noise ratio, which are factors in predicting antenna performance. Therefore, mean absolute error (MAE) and root mean square error (RMSE) performance indicators used in multiple regression analysis are used [34].

In a regression problem, MAE is a metric that measures the average absolute difference between the predicted values y_i and the actual values \hat{y}_i . It gives equal weight to all errors, regardless of their magnitude or direction. MAE is calculated by taking the absolute difference between each predicted and actual value, summing those differences, and dividing the sum by the total number of observations. The formula for MAE is presented (1) as below.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (1)$$

Root Mean Square Error (RMSE) is similar to MAE in that it measures the difference between predicted values y_i and actual values \hat{y}_i , but RMSE takes into account the squared values of these differences which gives more weight to large errors. And since it is sensitive to both small and large errors, RMSE provides a comprehensive view of the model's overall performance and can be useful metric for comparing different models. RMSE measures the square root of MSE metric. MSE quantifies the squared errors between predicted values y_i and actual values \hat{y}_i for each data point. Since it squares the errors, it transforms them into positive numbers and is highly sensitive to larger errors. Equation (2) presents the formula of MSE and (3) shows the formula for RMSE which is calculated by taking the square root of the average of the squared differences between predicted and actual values.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

4.3 Experiment results

In this paper, we develop prediction models for each of the 14 output-features by boosting the model using AutoML-Pycaret. we then developed a final multi-output regression prediction model by ensemble learning with the five machine learning models that showed the highest performance based on RMSE metric from AutoML-Pycaret result reports.

We can see the MAE and RMSE performance results of the one-output regression model that predicts 14 output-features one by one using AutoML-Pycaret in [Table 5](#) below. Regarding the output-feature models of Y_04, Y_05, Y_06, Y_10, and Y_11, both the MAE and RMSE results indicated high values in the range of 0.5 to 1.0, but the rest of the prediction models showed results of 0.5 to 0.7 or less.

Table 5. Results of MAE and RMSE performance for each one output-feature model using AutoML-Pycaret

| | MAE | RMSE | | MAE | RMSE |
|------|---------------|---------------|------|---------------|---------------|
| Y_01 | 0.2605 | 0.3372 | Y_08 | 0.4850 | 0.6244 |
| Y_02 | 0.2877 | 0.3678 | Y_09 | 0.4836 | 0.6214 |
| Y_03 | 0.2706 | 0.3454 | Y_10 | 0.5945 | 0.7970 |
| Y_04 | 2.0173 | 2.5590 | Y_11 | 0.6077 | 0.7699 |
| Y_05 | 1.8509 | 2.3174 | Y_12 | 0.4689 | 0.6028 |
| Y_06 | 0.6527 | 1.2464 | Y_13 | 0.4830 | 0.6213 |
| Y_07 | 0.3147 | 0.3948 | Y_14 | 0.4843 | 0.6216 |

However, when multi-output regression was finally performed to predict 14 output-features, the MAE result was 0.6928 and the RMSE result was 1.2065, which appeared to be relatively high. In order to prove the performance results of the proposed model in this study, we perform a performance comparison with Decision Tree, Random Forest, LGBM, Linear Regression, and XGBoost, which are machine learning models commonly used in multi-output regression models. In addition, we compare performance with the model proposed in this paper by using the TabNet deep learning model, which shows high excellence in tabular data. The models that were comparatively tested are shown in [Table 6](#) below and show relatively high performance compared to the other existing models.

Table 6. Results of MAE and RMSE Performance comparison between conventional models and the proposed model

| | Decision Tree | Random Forest | LGBM | Linear Regression | XGBoost | TabNet | Ours |
|------|---------------|---------------|--------|-------------------|---------|--------|---------------|
| MAE | 0.9831 | 0.6969 | 0.6994 | 0.7058 | 0.7118 | 0.8753 | 0.6928 |
| RMSE | 1.7293 | 1.2164 | 1.2196 | 1.2392 | 1.2409 | 1.4650 | 1.2065 |

5. Conclusion and Future Research Works

This paper proposed an antenna performance prediction model in the autonomous driving radar manufacturing process. The proposed model has achieved the MAE of 0.69 and RMSE of 1.19, which are better than the baseline models. Our approach has demonstrated the effectiveness of using AutoML-Pycaret which is an open source typical AutoML machine learning workflow engine based on Python. We developed our final multi-output regression ensemble model by performing hyperparameter tuning on random forest, LGBM, gradient boosting regressor, ridge, and lasso models, which most frequently showed high performance based upon RMSE performance metric in AutoML-Pycaret results reports and performing ensemble learning using them.

In this work, the enhanced accuracy achieved through this research leads to improved safety and reliability for autonomous vehicles on the road. Furthermore, the utilization of AutoML-Pycaret and the optimization of a variety of machine learning models present a significant opportunity to reduce manufacturing costs. Precise prediction of antenna performance plays a pivotal role in preventing the production of faulty radar systems, thereby conserving resources, and minimizing waste in the manufacturing process. As the future works, we plan to explore additional machine learning models and deep learning model to further improve the performance of our prediction model. We will also investigate the impact of different data preprocessing techniques and domain knowledge to enhance the quality of our predictions. Our work has important implications for the automotive industry, as accurate prediction of antenna performance is critical for ensuring the safety and reliability of autonomous vehicles. Conclusively, we expect that the antenna performance prediction model using the AutoML approach and proposed in this paper ought to make a meaningful contribution in terms of not only predicting the antenna performance but also improving the manufacturing process quality in the manufacturing process automation industry.

Acknowledgement

This research was supported by the ATC+ R&D Project of KEIT (Korea Evaluations Institute of Industrial Technology) from the Ministry of Trade, Industry and Energy in 2020 (Grant No. 20009228, Development of Deep Learning-based Predictive Production Quality and Field Safety Management Technology to Support Level-4 Smart Factories in the Textile Industries).

References

- [1] J. Lee, *Industrial AI*, Springer Singapore, Singapore, 2020.
- [2] Meesun Kim, Hyun Ahn, and Kwanghoon Pio Kim, "Process-Aware Internet of Things: A Conceptual Extension of the Internet of Things Framework and Architecture," *KSII Transactions on Internet and Information Systems (TIIS)*, vol.10, no.8, pp.4008-4022, 2016. [Article \(CrossRef Link\)](#)
- [3] J. Wang, C. Liu, M. Zhu, P. Guo and Y. Hu, "Sensor Data Based System-Level Anomaly Prediction for Smart Manufacturing," in *Proc. of 2018 IEEE International Congress on Big Data (BigData Congress)*, San Francisco, CA, USA, pp. 158-165, 2018. [Article \(CrossRef Link\)](#)
- [4] Yang Tian, Mengyu Fu, Fang wu, "Steel Plate Fault Diagnosis on the Basis of Support Vector Machines," *Neurocomputing*, vol.151, no.1, pp. 296-303, March 2015. [Article \(CrossRef Link\)](#)
- [5] A. Song, Y. Han, H. Hu and J. Li, "A Novel Texture Sensor for Fabric Texture Measurement and Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 7, pp. 1739-1747, July 2014. [Article \(CrossRef Link\)](#)
- [6] Chao Li, Jun Li, Yafei Li, Lingmin He, Xiaokang Fu, and Jingjing Chen, "Fabric Defect Detection in Textile Manufacturing: A Survey of the State of the Art," *Secur. Commun. Networks*, vol. 2021, pp.1-13, May 2021. [Article \(CrossRef Link\)](#)
- [7] Jianjun Ni, Yinan Chen, Yan Chen, Jinxiu Zhu, Deena Ali, and Weidong Cao, "A Survey on Theories and Applications for Self-Driving Cars Based on Deep Learning Methods," *Applied Sciences*, vol.10, no.8, March 2020. [Article \(CrossRef Link\)](#)
- [8] Z. Wang, Y. Wu and Q. Niu, "Multi-Sensor Fusion in Automated Driving: A Survey," *IEEE Access*, vol. 8, pp. 2847-2868, 2020. [Article \(CrossRef Link\)](#)
- [9] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. McCullough and A. Mouzakitis, "A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829-846, April 2018. [Article \(CrossRef Link\)](#)
- [10] G. Hakobyan and B. Yang, "High-Performance Automotive Radar: A Review of Signal Processing Algorithms and Modulation Schemes," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 32-44, Sept. 2019. [Article \(CrossRef Link\)](#)
- [11] J. Wang, Y. Lan, S. Lu and L. Cheng, "An Efficient Complex Event Processing Algorithm based on Multipattern Sharing for Massive Manufacturing Event Streams," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 3, pp. 1385-1402, 2019. [Article \(CrossRef Link\)](#)
- [12] Merrill Skolnik, *Radar Handbook*, Third Edition, USA: McGraw-Hill, 2008,
- [13] W. Menzel and A. Moebius, "Antenna Concepts for Millimeter-Wave Automotive Radar Sensors," *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2372-2379, July 2012. [Article \(CrossRef Link\)](#)
- [14] Ziebinski, A., Cupek, R., Erdogan, H., Waechter, S., "A Survey of ADAS Technologies for the Future Perspective of Sensor Fusion," in *Proc. of International Conference on Computational Collective Intelligence (ICCCI 2016)*, pp. 135–146, 2016. [Article \(CrossRef Link\)](#)
- [15] A. P. Sligar, "Machine Learning-Based Radar Perception for Autonomous Vehicles Using Full Physics Simulation," *IEEE Access*, vol. 8, pp. 51470-51476, 2020. [Article \(CrossRef Link\)](#)
- [16] Aylin Alin, "Multicollinearity," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 3, pp. 370-374, 2010. [Article \(CrossRef Link\)](#)
- [17] Il-Gyo Chong, Chi-Hyuck Jun, "Performance of some variable selection methods when multicollinearity is present," *Chemometrics and Intelligent Laboratory Systems*, vol. 78. Issues. 1-2, pp. 103-112, 2005. [Article \(CrossRef Link\)](#)
- [18] Lee CJ, Park C-S, Kim JS, Baek J-G., "A Study on Improving Classification Performance for Manufacturing Process Data with Multicollinearity and Imbalanced Distribution," *Journal of Korean Institute of Industrial Engineers*, Vol. 41, no. 1, pp. 25-33, 2015. [Article \(CrossRef Link\)](#)
- [19] Xin He, Kaiyong Zhao, Xiaowen Chu, "AutoML: A survey of the state-of-the-art," *Knowledge-Based Systems*, Vol. 212, no. 1, 2021. [Article \(CrossRef Link\)](#)
- [20] Moez Ali, "PyCaret: An open source, low-code machine learning library in Python," 2020.04. [Online]. Available: <https://www.pycaret.org>

- [21] Doyle, Keith B., "Antenna performance predictions of a radio telescope subject to thermal perturbations," *Optical Modeling and Performance Predictions IV*, SPIE, vol. 7427, pp. 129-137, 2009. [Article \(CrossRef Link\)](#)
- [22] Bathker, D. A., Veruttipong, W., Otoshi, T. Y., & Cramer, P. W., "Beam-waveguide antenna performance predictions with comparisons to experimental results," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 6, pp. 1274-1285, 1992. [Article \(CrossRef Link\)](#)
- [23] Xiaodong Hou, Jiangping Yang, Yu Zhang, Zengjun Bi, "Study on prediction method of fault channel numbers in radar antenna array," in *Proc. of 2017 Prognostics and System Health Management Conference (PHM-Harbin)*, pp. 1-6, July 2017. [Article \(CrossRef Link\)](#)
- [24] Boya Liu, Xiaowen Bi, Lijuan Gu, Jie Wei, Baozhong Liu, "Application of a Bayesian Network Based on Multi-Source Information Fusion in the Fault Diagnosis of a Radar Receiver," *Sensors*, vol. 22, no. 17, August 2022. [Article \(CrossRef Link\)](#)
- [25] Kaijing Chen, Wendi Wang, Xiaohui Chen, Huarui Yin, "Deep Learning Based Antenna Array Fault Detection," in *Proc. of 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, pp. 1-5, 2019. [Article \(CrossRef Link\)](#)
- [26] "Antenna Performance Prediction AI Contest for Autonomous Driving Sensors," *dacon.io*, 2022.
- [27] Yen-Chi Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics & Epidemiology*, vol. 1, no. 1, pp. 161-187, 2017. [Article \(CrossRef Link\)](#)
- [28] Song YY, Lu Y., "Decision tree methods: applications for classification and prediction," *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130-135, 2015. [Article \(CrossRef Link\)](#)
- [29] Biau, G., Scornet, E., "A random forest guided tour," *TEST*, vol. 25, pp. 197-227, 2016. [Article \(CrossRef Link\)](#)
- [30] Aziz, R.M., Baluch, M.F., Patel, S. et al., "LGBM: a machine learning approach for Ethereum fraud detection," *International Journal of Information Technology*, vol. 14, pp. 3321-3331, 2022. [Article \(CrossRef Link\)](#)
- [31] Gülden Kaya Uyanık, Neşe Güler, "A Study on Multiple Linear Regression Analysis," *Procedia - Social and Behavioral Sciences*, vol. 106, pp. 234-240, 2013. [Article \(CrossRef Link\)](#)
- [32] Tianqi Chen, Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. of KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*, pp. 785-794, August 2016. [Article \(CrossRef Link\)](#)
- [33] Sercan Ö. Arik, Tomas Pfister, "TabNet: Attentive Interpretable Tabular Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 6679-6687, May 2021. [Article \(CrossRef Link\)](#)
- [34] Cort J. Willmott, Kenji Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79-82, 2005. [Article \(CrossRef Link\)](#)



So-Hyang Bak received her B.S. degree from the Division of Computer Science and Engineering, Kyonggi University, South Korea, in 2021. She has received an M.S. degree from the School of Department of Computer Science, Kyonggi University, Suwon, South Korea. She has worked as a researcher at the Data and Process Engineering Research Laboratory, Kyonggi University. Her research interests include data analysis, artificial intelligence, process mining, deep learning, machine learning, auto-ML, and smart factory.



Kwanghoon Pio Kim received the B.S. degree in computer science from Kyonggi University, South Korea, in 1984, the M.S. degree in computer science from Chungang University, South Korea, in 1986, and the M.S. and Ph.D. degrees in computer science from the University of Colorado Boulder, USA, in 1994 and 1998, respectively. He had been the Dean of the Computerization and Informatics Institute, Kyonggi University, from 2017 to 2021. Since 2007, he has been the Founder and the Director in charge of the Contents Convergence Software Research Institute, Kyonggi University, where he has been leading and fulfilling a multi-million dollar research project that will be continuously supported and funded by the National Research Foundation of Korea, from 2020 to 2029, ever since the institute was officially designated as the National Science and Engineering Research Institute by the Ministry of Education, South Korea, in 2020. He had worked as a Researcher and a Developer at Aztek Engineering, American Educational Products Inc., and IBM, USA, and a Research Member of AI Computer Science and Engineering and the Founder and the Supervisor of the Data and Process Engineering Research Laboratory, Kyonggi University, South Korea. His research interests include groupware, workflow and business process management system, process warehousing and mining, predictive process monitoring, workflow-supported social networks discovery and analysis, process-aware Internet of Things, predictive process monitoring, process deep-learning.