Improving Linear Regression Performance for Interpretable Control: A Partial Dependence Plot Approach (September 2025)

Kenta Oshima / Independent Researcher / kenta.oshima.shr@gmail.com

Abstract— **This paper addresses the challenge of improving the performance of interpretable linear models for real-time process control in industrial and chemical systems. While black-box models such as Random Forest and XGBoost achieve high predictive accuracy, their lack of interpretability and limited suitability for real-time applications make them difficult to integrate into production environments. To overcome this limitation, we propose a novel feature engineering approach that leverages Partial Dependence Plots (PDP) to capture the complex nonlinear relationships learned by black-box models and convert them into features appropriate for linear regression. We evaluate the proposed methodology using the well-known publicly available wine quality dataset, which provides a strong benchmark due to its multivariate and nonlinear characteristics that are analogous to those observed in industrial processes. Our results show that PDP-based feature transformation substantially improves the predictive accuracy of linear models, achieving an R^2 score comparable to that of black-box models. This study proposes a practical solution for building high-performance yet interpretable models, showing strong potential for real-time deployment in process control and monitoring.**

Index Terms— **Partial Dependence Plots (PDP), Linear regression, Feature engineering, Real-time control, Industrial process**

I. Introduction

In modern industrial processes, the demand for real-time control and predictive maintenance has led to the widespread adoption of data-driven models for quality monitoring and process optimization. While complex, non-linear models such as Random Forest (RF) and XGBoost have demonstrated high predictive accuracy for multivariate datasets, they are often considered "black-box" models due to their inherent lack of transparency and interpretability [1, 2]. Their lack of interpretability presents a significant challenge for their deployment in critical production environments, where understanding the relationship between input features and model outputs is essential for diagnostics, quality assurance, and regulatory compliance [1, 3, 4]. Furthermore, the computational complexity of these models can make them unsuitable for low-latency, real-time control applications where rapid and reliable predictions are paramount.

On the other hand, linear regression models offer a high degree of interpretability, as the impact of each feature on the final prediction is directly represented by its coefficient. This makes them ideal for implementation in control systems where the logic must be both transparent and simple. However, linear models are inherently limited in their ability to capture and model complex, non-linear relationships present in real-world data, often resulting in lower predictive performance compared to their black-box counterparts.

This study proposes a novel feature engineering approach to bridge the gap between the predictive power of black-box models and the interpretability of linear models. Our method utilizes Partial Dependence Plots (PDPs), a model-agnostic technique for visualizing the marginal effect of one or two features on the predicted outcome of a black-box model. By fitting simple, interpretable functions to these PDPs, we can effectively extract the non-linear relationships learned by a complex model (e.g., RF or XGBoost) and transform them into new features. These newly engineered features can then be used to significantly improve the performance of a linear regression model, enabling it to capture non-linear patterns without sacrificing interpretability.

To validate this approach, we use the publicly available Wine Quality dataset [5]. This dataset is a suitable benchmark for our methodology as its multi-variate and non-linear characteristics are analogous to those of real-world industrial processes. The objective of this paper is to demonstrate that by leveraging insights from black-box models, we can develop a simple, interpretable linear model that achieves predictive accuracy comparable to that of complex models, making it a viable and practical solution for real-time process control and monitoring.

II. Methodology

A. Dataset and Experimental Environment

This study utilizes the publicly available Wine Quality dataset, a widely recognized benchmark for regression tasks. The dataset comprises 11 physicochemical features (e.g., alcohol content, pH, fixed acidity) and a quality score, which serves as the target variable. The multivariate and non-linear relationships within this dataset make it an ideal benchmark for evaluating our methodology, since these characteristics are closely analogous to those found in industrial processes. All data preprocessing, model training, and analysis were performed using Python with the following key libraries: pandas for data manipulation, scikit-learn for machine learning models, and matplotlib for data visualization. The complete code and detailed procedures are available in the accompanying Jupyter Notebook on our GitHub repository at [https://github.com/kentaoshimashr-maker/PDPs-study].

B. Proposed Partial Dependence Plot-based Feature Engineering

To enhance the performance of a simple, interpretable linear model, we propose a novel feature engineering approach informed by insights from black-box models. Our methodology, as implemented in the Jupyter Notebook, follows a three-step process:

1. **Black-Box Model Training**: We first train high-performance black-box models, including a Random Forest Regressor, an XGBoost Regressor, and a Multi-layer Perceptron (MLP) Regressor, on the original Wine Quality dataset. These models are chosen for their proven ability to capture complex, non-linear feature interactions, which a traditional linear model cannot.
2. **Partial Dependence Plot (PDP) Generation**: For each feature, we generate a Partial Dependence Plot using the trained black-box model. A PDP illustrates the marginal effect of a feature on the predicted outcome, providing a visual representation of the non-linear relationship the black-box model has learned. This step effectively makes the black-box model's learned behavior interpretable.
3. **Feature Transformation**: We analyze the shape of each PDP and identify a corresponding simple, interpretable function (e.g., polynomial, logarithmic, or piecewise linear) that best approximates the relationship. This function is then used to transform the original feature into a new, engineered feature. For instance, if the PDP for 'alcohol' shows a sigmoidal relationship, we apply a sigmoid function to the 'alcohol' feature to create a new one.

C. Experimental Setup

To evaluate the effectiveness of our proposed approach, we conducted a comparative analysis of model performance using three distinct configurations:

- **Baseline Linear Model**: A linear regression model trained on the original dataset without any feature transformation.
- **Transformed Linear Model**: A linear regression model trained on the dataset augmented with the new, PDP-based engineered features.
- **Black-Box Models**: The original Random Forest, XGBoost, and MLP models, serving as a performance benchmark for the transformed linear model.

Model performance was evaluated using two standard metrics: R-squared (R^2), which measures the proportion of variance in the dependent variable that is predictable from the independent variables, and Root Mean Squared Error (RMSE), which indicates the average magnitude of the errors. All models were trained and evaluated using a train-test split approach, to ensure the robustness and generalizability of the results.

III. Results

We present the results of our experimental analysis comparing the performance of the baseline linear model, the black-box models, and the linear model enhanced with PDP-based feature engineering. Model performance was evaluated using R^2 and RMSE.

The initial performance of all models on the original dataset is summarized in Table 1. As expected, the non-linear black-box models such as Random Forest and XGBoost achieved higher predictive accuracy than the baseline linear regression model, which is limited in its capacity to capture complex relationships. Particularly, the Random Forest model demonstrated the highest R^2 score of 0.4628, confirming its strong ability to capture the underlying patterns within the data.

Our main findings concerning the proposed PDP-based feature engineering approach are presented in Table 2. The results indicate that the effectiveness of the feature transformation highly depends on the black-box model used to generate the PDPs. When the features were transformed based on the Random Forest model's PDPs, the linear model's performance significantly improved, with the R^2 score increasing from 0.3171 to 0.3891. This indicates that the method successfully extracted valuable non-linear information from the Random Forest model and made it accessible to the interpretable linear model, outperforming the original XGBoost model (R^2: 0.3431).

In contrast, feature transformations based on the PDPs from the XGBoost and MLP models resulted in a decline in performance for the linear model. The R^2 score for the XGBoost-transformed model dropped sharply to 0.0518, while the MLP-transformed model yielded a negative R^2 of -2.1545. These differing behaviors are visually represented by their respective PDPs, as shown in Figure 1, illustrating the varying shapes and complexities of the relationships learned by these models.

TABLE 1
PERFORMANCE COMPARISON OF ORIGINAL DATASET

Model	R2 Score	RMSE
Liner Regression (Baseline)	0.3171	0.6165
Random Forest	0.4628	0.5468
XGBoost	0.3431	0.6046
MLP (Neural Network)	0.2983	0.6249

TABLE 2
PERFORMANCE COMPARISON OF PDP-TRANSFORMED DATASET

Model	R2 Score	RMSE
LR with RF PDP-transformed features	0.3891	0.5831
LR with XGBoost PDP-transformed features	0.0518	0.7264
LR with MLP PDP-transformed features	-2.1545	1.3249

IV. Discussion

Our analysis of the experimental results reveals key insights into the effectiveness and limitations of the proposed feature engineering method. The primary finding is that the success of the PDP-based transformation depends on two key factors: the predictive accuracy of the black-box model and the interpretability of its learned relationships.

First, the notable performance of the linear model with Random Forest-transformed features demonstrates that our approach can effectively extract valuable non-linear information from a powerful black-box model and translate it into an interpretable feature representation. This not only significantly improves the linear model's performance but also allows it to outperform the original XGBoost model, confirming the method's practical utility for developing high-performance yet transparent systems for real-time process control.

However, the results from XGBoost and MLP-based transformations highlight the method's limitations. As illustrated in Figure 1, the PDPs from the XGBoost model exhibited a highly irregular and complex shape. This suggests that the model, while achieving high native accuracy, learned patterns that were too intricate to be captured by a simple, interpretable function. This indicates that the model's complex decision-making process is not well-represented by a simplified PDP, making it difficult to translate into a linear model. The result is a substantial degradation in performance when using the transformed features.

Conversely, the MLP model's PDPs were smoother and more amenable to fitting, as seen in Figure 1. This indicates that the model learned a more generalizable, less complex relationship. However, since the MLP's initial performance was low—comparable to that of the baseline linear model—the information it extracted was not of sufficient quality to improve the transformed linear model. This leads to our second key finding: the PDP-based transformation is only as good as the black-box model's ability to learn meaningful and high-quality relationships, not just simple or noisy ones.

In summary, for our method to be successful, it is essential to use a black-box model that not only achieves high predictive accuracy but also learns relationships that are good candidates for simplification. The Random Forest model proved ideal in this regard, offering a balance between predictive power and a relationship structure that could be effectively simplified and leveraged by a linear model. This study provides a practical blueprint for integrating powerful insights from black-box models into interpretable linear systems.

While this study demonstrates the potential of PDP-based feature engineering, its findings provide several important avenues for future work First, our experiments were conducted on a single dataset, and the generalizability of our findings needs to be validated across a wider range of industrial and chemical process data. Future work will involve applying this methodology to other datasets, such as those related to chemical reactions or manufacturing defect rates, to confirm its robustness.

Second, the process of manually identifying and fitting functions to the PDPs can be labor-intensive. Future research could focus on developing an automated algorithm that analyzes the shape of a PDP and automatically selects the best-fitting interpretable function. This would significantly enhance the practicality and scalability of our proposed method. Finally, exploring the integration of other Explainable AI (XAI) techniques, such as SHAP or LIME, could provide a more comprehensive framework for building high-performance and fully transparent models.

V. Conclusion

This study presents a novel and practical approach for enhancing the performance of interpretable linear models by leveraging insights from black-box models. We demonstrate that by using Partial Dependence Plots (PDP) to transform features, a simple linear regression model can achieve predictive accuracy comparable to that of complex models such as Random Forest and XGBoost. Our results reveal that the success of this method is highly dependent on both the predictive power of the black-box model and the nature of the relationships it learns. Specifically, transforming features based on a high-performing and smoothly behaving model such as Random Forest yields significant performance gains, enabling the linear model to surpass the original XGBoost model.

The proposed methodology offers a viable solution for building high-performance yet transparent models that can be effectively deployed in critical environments such as real-time process control. This work bridges the critical gap between model performance and interpretability, providing a pathway to integrate the predictive power of black-box models into systems where transparency and simplicity are paramount.

While this study demonstrates promising results, its

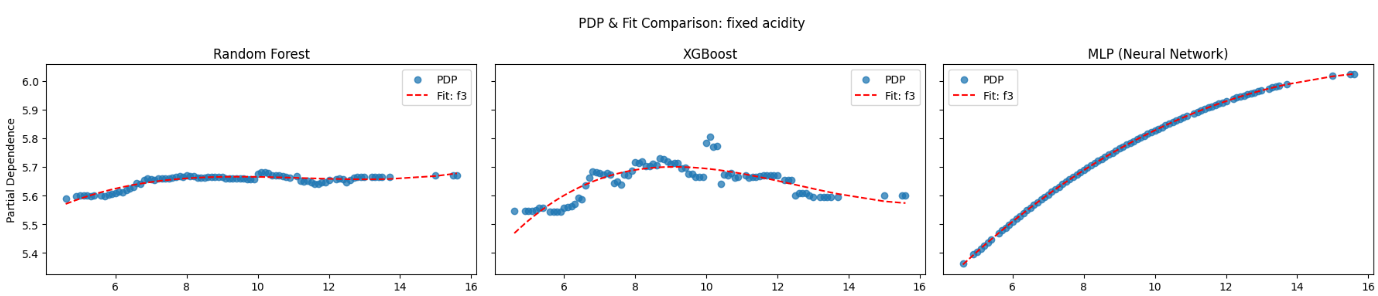

Figure 1: Partial Dependence Plots for 'fixed acidity' from Random Forest, XGBoost, and MLP models

validation is limited to a single dataset, and further investigations are necessary to confirm the approach's generalizability across diverse industrial and chemical processes. Additionally, although the current automated feature transformation considers only linear, quadratic, cubic, and logarithmic functions, exploring other functional forms—such as piecewise linear, sigmoid, or spline functions—could capture more complex non-linear behaviors more effectively. Advancing automated strategies for selecting the most suitable transformation functions will be essential to improving the method's scalability, flexibility, and predictive performance.

Future work will involve applying the proposed methodology to other datasets from various industrial applications, such as chemical reactions and manufacturing defect analysis, to assess robustness and expand applicability. Moreover, integrating automated algorithms for analyzing PDP shapes and fitting interpretable functions will enhance efficiency and objectivity. Finally, exploring complementary Explainable AI (XAI) methods like SHAP and LIME may provide a more comprehensive framework for developing high-performance yet fully transparent models.

REFERENCES

[1] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[2] Chen, T., & Guestrin, C. "XGBoost: A Scalable Tree Boosting System". KDD '16, 2016.

[3] P. J. Werbos, "Backpropagation: Past and future," in Proceedings of the IEEE International Conference on Neural Networks, 1988.

[4] Rudin, C. et al. "Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges", 2021.

[5] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," Decision Support Systems, vol. 47, no. 4, pp. 547-553, 2009.