Model-Agnostic Interpretation with Partial Dependence Plots:

PDP Analysis on Bike Rentals and House Prices

Germán Mallo Faure and Josep Ferrer García
EDM (Evaluation, Deployment and Monitoring of Models)
Degree in Data Science
Escuela Técnica Superior de Ingeniería Informática
Universitat Politècnica de València, Spain
{gmalfau, jfergar1}@etsinf.upv.es
https://germanmallo.com

May 2025

Contents

1 Introduction

This report focuses on the application of **Partial Dependence Plots (PDPs)** as a model-agnostic tool to explain black-box machine learning models. We use Random Forests to predict outcomes in two different scenarios:

- Daily rental counts in the Capital Bikeshare dataset.
- House prices in the `kc_house_data.csv` dataset.

We analyze the marginal influence of selected features on model predictions using one-dimensional and two-dimensional PDPs, and interpret the relationships discovered by the model.

2 One-Dimensional PDP - Bike Rentals

We fitted a Random Forest regression model to predict daily bike rental counts (`cnt`) from the Capital Bikeshare dataset. We plotted PDPs for the variables `days_since_2011`, `temp`, `hum`, and `windspeed`.

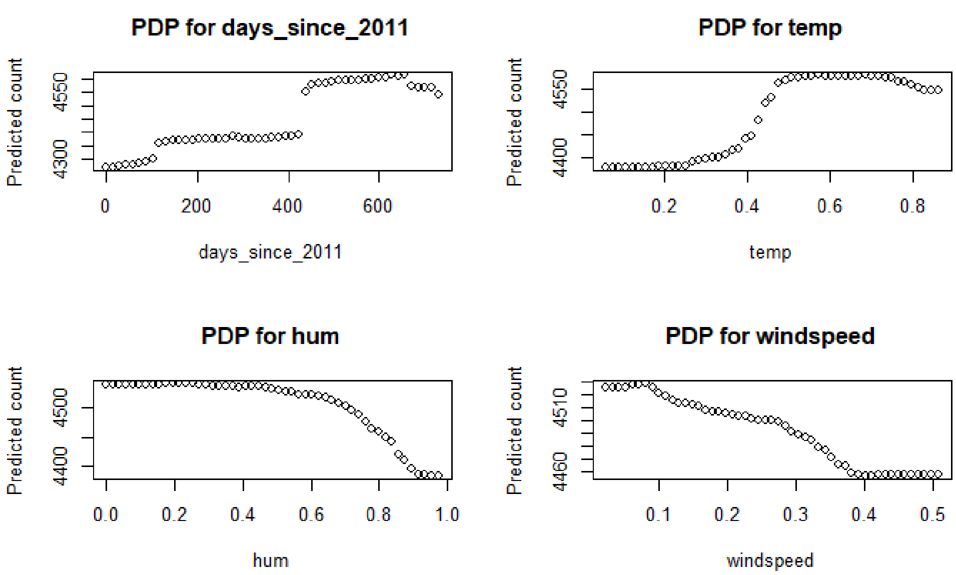

Figure 1: One-Dimensional PDPs for bike rentals: days since 2011, temperature, humidity, and wind speed

Interpretation

- **Days Since 2011 (days_since_2011):** The plot shows an upward trend in predicted rentals over time. As more days pass since the start of 2011, the model predicts more rentals, indicating a growing popularity of the service.
- **Temperature (temp):** There is a strong positive correlation between temperature and rental count up to approximately 0.6 (normalized), after which the effect stabilizes or slightly declines.

- **Humidity (hum):** High humidity negatively affects rental counts. The drop becomes steeper beyond a humidity level of 0.7.
- **Windspeed:** The model predicts fewer rentals as wind speed increases, with the most significant drop between 0.1 and 0.3.

3 Two-Dimensional PDP - Bike Rentals

We generated a 2D Partial Dependence Plot using humidity and temperature to analyze their joint impact on the number of bikes rented.

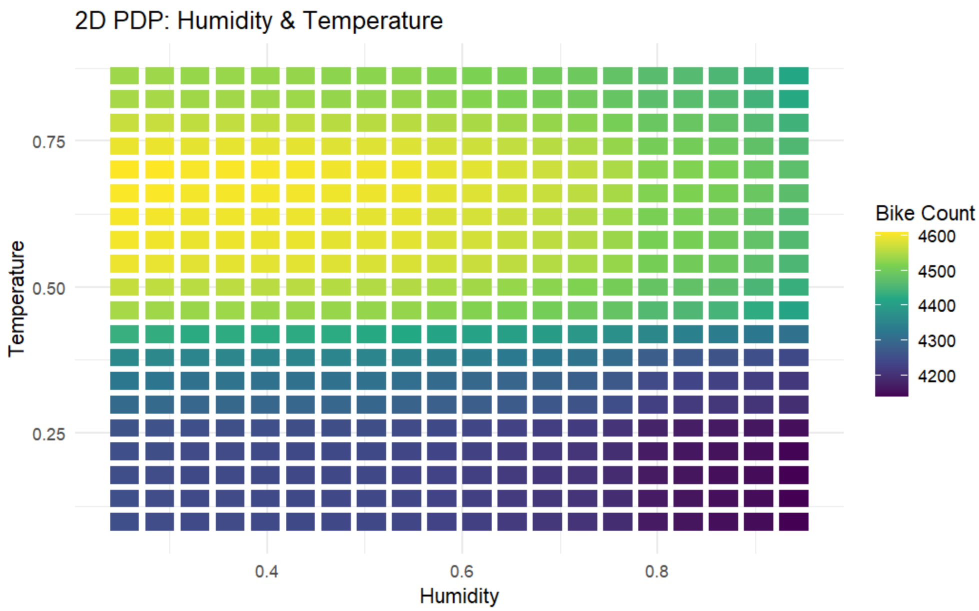

Figure 2: 2D PDP: Interaction between temperature and humidity on bike rentals

Interpretation

The plot reveals that:

- The highest rental counts are observed at moderate to high temperatures (0.6–0.8) combined with low to medium humidity (0.3–0.6).
- Low temperatures consistently yield fewer rentals regardless of humidity.
- Even at high temperatures, excessive humidity (¿0.8) reduces the number of rentals.

This interaction shows that optimal biking conditions occur in warm and not-too-humid days.

4 One-Dimensional PDP - House Prices

Using the `kc_house_data.csv` dataset, we trained a Random Forest model to predict house prices. We then created PDPs for the features: `bedrooms`, `bathrooms`, `sqft_living`, and `floors`.

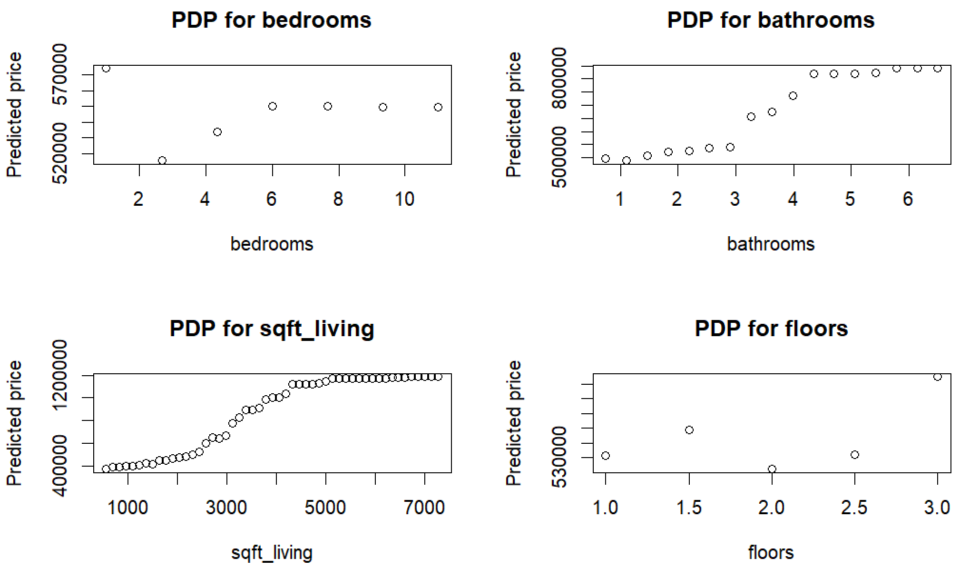

Figure 3: One-Dimensional PDPs for house prices

Interpretation

- **Bedrooms:** An unexpected finding: adding more bedrooms does not increase the predicted price. Prices even slightly decrease, possibly due to redundancy or correlation with older/less expensive homes.

- **Bathrooms:** A positive relationship is observed; more bathrooms lead to higher house prices, especially from 3 bathrooms onwards.

- **Living Area (sqft_living):** One of the strongest predictors. Price increases with size up to around 5000 sqft, then plateaus.

- **Floors:** The relationship is non-linear and unclear. Houses with 1 or 3 floors have higher predicted prices than those with 2 floors, possibly reflecting confounding factors in the dataset.

5 Conclusion

The use of Partial Dependence Plots allows us to understand how specific variables affect the predictions of complex, non-linear models such as Random Forests. In both the bike rental and housing domains, PDPs have helped reveal the marginal effects of features and their potential interactions.

By visualizing these effects, we gain valuable insights into model behavior that can inform decision-making, model trust, and feature selection.