Report XAI3

Piotr Franc and Ole Magnus Lærum

May 2025

1 Introduction

This report explores the interpretability of a random forest model using model-agnostic methods.

In the first part of the report we will try to explain a random forest model fit to data from the bike sharing dataset using Partial Dependence Plots (1 Dimensional and 2 Dimensional), in order to learn what correlations exist in the dataset and how the model uses the features to predict the number of bikes rented.

In the second part we'll try to interpret how the following features affect the price of a house:

- bedrooms
- bathrooms
- sqft_living
- sqft_lot
- floors
- yr_built

To do that, we'll also use PDPs, as well as Individual Conditional Expectation Plots to delve into the interactions further.

2 Bike Rentals Interpretation

In this section we will analyze the following features:

- days_since_2011
- temp
- hum
- windspeed

Disclaimer: The data at hand is correlated, which might distort the PDP graphs and their interpretability. For example, days since 2011 are correlated with season, which in turn dictates temperature. PDP graphs assume no correlation, and this means that they take into the account points that do not exist in the real data.

2.1 Days since 2011

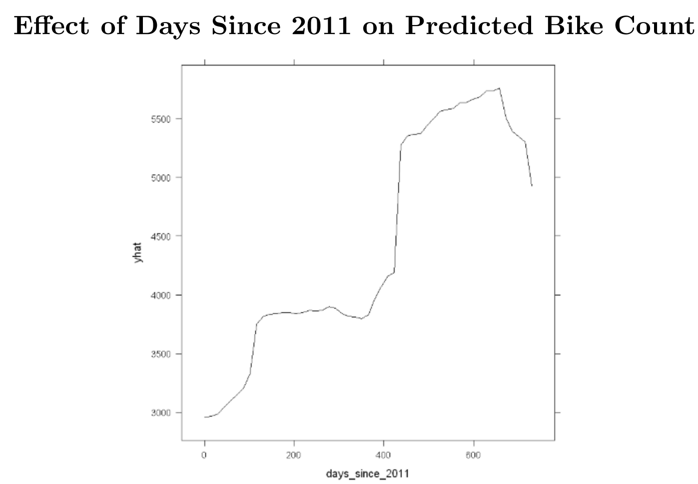

Figure 1: Partial dependence plot for `days_since_2011`.

Figure 1 shows that the days_since_2011 doesn't affect the bike count linearly - there are clearly three steps visible - one after around 100 days, and second after 500 days, and a drop after around 700 days. This means that most likely some events happened which changed the amount of bikes rented. For example:

- New stations were opened or closed
- More bikes were bought which increased the availability
- The pricing changed

2.2 Temperature

Effect of Temperature on Predicted Bike Count

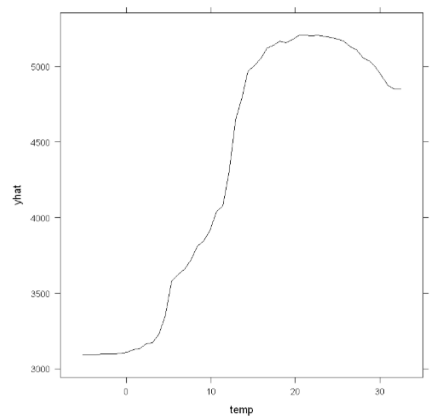

Figure 2: Partial dependence plot for `temperature`.

Figure 2 shows the marginal effect of temperature on bike rentals. This one is pretty obvious - when it's cold, not a lot of people want to ride a bike, and when the temperature is perfect (around 20 degrees) the most people are going cycling. At the end there is a drop - hot temperature deters cyclers.

2.3 Humidity

Effect of Humidity on Predicted Bike Count

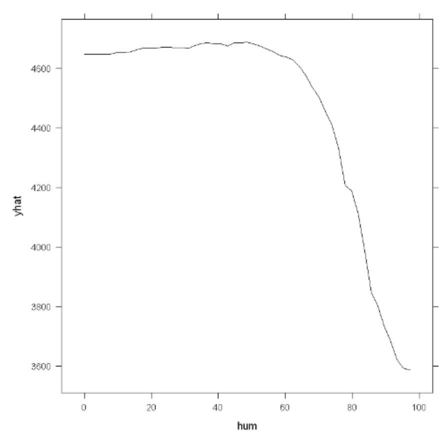

Figure 3: Partial dependence plot for `humidity`.

The marginal effect of humidity are as expected - up to some point it doesn't change anything, but after a certain point (60%) there is a sharp drop, as it starts being very unpleasant.

2.4 Windspeed

Effect of Windspeed on Predicted Bike Count

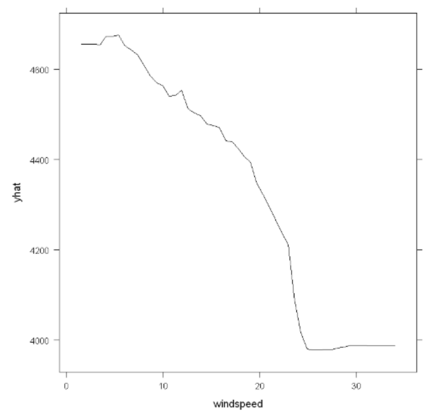

Figure 4: Partial dependence plot for `windspeed`.

Windspeed has an inverse relationship with bike rentals. As wind speed increases, fewer people choose to rent bikes. Beyond approximately 25 km/h, further increases have only a marginal impact.

2.5 Temperature and Humidity

Effect of Temperature and Humidity on Predicted Bike Count

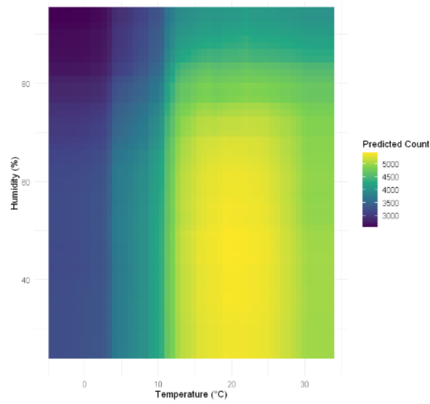

Figure 5: Partial dependence plot for temperature and humidity.

The following graph shows the marginal effect of temperature and humidity on predicted bike count - we can see that the ideal conditions for bike rentals are low humidity and high temperature. This mirrors the compound effect of two PDP graphs from before.

3 Price of houses

The second part focuses on another dataset. Here we will analyze the effect on price of the following features:

- Number of Bedrooms
- Number of Bathrooms
- Square feet of living space
- Number of Floors

The model also uses yr_built and sqft_lot for its predictions.

Disclaimer: The data in this case is probably even more correlated than one in Bike Sharing Dataset.

3.1 Plots

Effect of Bathrooms on Predicted Price

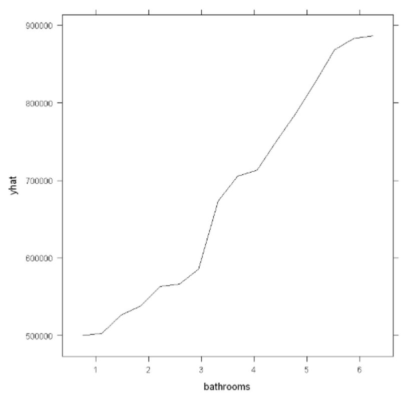

Figure 6: Partial dependence plot for `bathroom`.

Effect of Sqft_Living on Predicted House Price

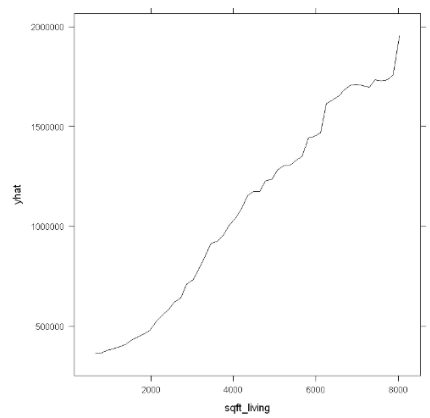

Figure 7: Partial dependence plot for `sqft_living`.

Effect of Floors on Predicted House Price

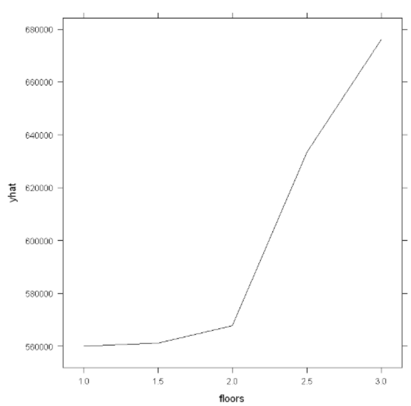

Figure 8: Partial dependence plot for `floors`.

Effect of Bedrooms on Predicted House Price

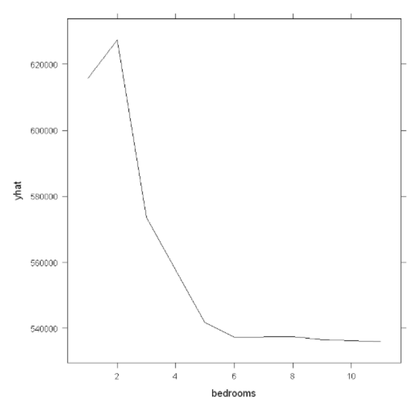

Figure 9: Partial dependence plot for `bedrooms`.

3.2 Interpretation of results

The results for bathrooms, living space and floors are as expected - the more the better. However, for bedroom the trend reverses - the price sharply declines until 6 bedrooms where it doesn't have any further bearing on the price.

On one hand, this seems counterintuitive - when thinking of houses with many bedrooms, we think that they are worth more than those with less bedrooms. However, this isn't what PDP graph conveys. The trend on PDP graph is averaged over "what does the model predict if the feature xc changes while all other features stay the same". This means, that we actually compare two same houses, but one with 2 bedrooms, and the other with 4. In such cases, bedrooms either have to be smaller, or there is less space for other places in the house. There are other examples that can explain this trend - the model picked up those subtleties and adjusted to make better predictions.

4 Summary

Partial Dependence Plots provided great insight into the marginal effects of features in question. In the Bike Renting dataset the predictions aligned closely with intuition, while in house pricing dataset the model uncovered some interesting correlations in the data. This shows that PDP's are a useful tool in explaining black box models, especially when compared to previous methods, such as linear models, which didn't pick up any subtleties in the data.