

XAI: Model-agnostic methods

Introduction

In this third session, we will explore model-agnostic methods in XAI, focusing on Partial Dependence Plots (PDP). We will use two datasets: a familiar one for predicting bike rentals and a new one for house prices. We will examine how variables such as temperature, humidity, and house characteristics influence predictions of random forest models. Additionally, version control will be ensured using Git with backup on GitHub, including a detailed report with comments and answers.

Discussion points

One dimensional Partial Dependence Plot.

First, we will analyze the dependency between the bike rentals and some of the variables. We have created some graphics, which are shown in Figure 1, about the relationship between the different variables:

- The PDP for days from 2011 shows how the predicted number of bike rentals changes as the number of days since 2011 increases, while holding the other variables constant. The plot shows that the predicted number of bike rentals increases gradually as the number of days since 2011 increases. This suggests that there is a long-term trend in bike rental numbers, with the number of rentals increasing steadily over time.
- The PDP for humidity shows how the predicted number of bike rentals changes as the humidity changes, while holding the other variables constant. The plot shows that the predicted number of bike rentals decreases as the humidity increases. This means that people are less likely to rent bikes when it is humid.
- The PDP for wind speed shows how the predicted number of bike rentals changes as the wind speed changes, while holding the other variables constant. The plot shows that the predicted number of bike rentals decreases as the wind speed increases. This means that people are less likely to rent bikes when it is windy.
- The PDP for temperature shows how the predicted number of bike rentals changes as the temperature changes, while holding the other variables constant. The plot shows that the predicted number of bike rentals increases gradually as the

temperature increases. This means that people are more likely to rent bikes when the weather is warm.

Bidimensional Partial Dependency Plot

Using the 2D Partial Dependency Plot shown in Figure 2, we can discern the correlation between bike rentals and both temperature and humidity. When temperatures fall and humidity is high, there is a decline in bike rentals, likely due to unfavorable weather conditions such as extreme cold or heat. Conversely, in mild temperatures ranging from 12 to 25 degrees, and low humidity levels between with 40% of humidity as a maximum, the ideal conditions for bike rentals emerge, as evidenced by the intense red zones on the plot.

PDP to explain the price of a house

Finally, we will apply the same techniques as in the first discussion point, but now we are using a different dataset. Now, our objective is to predict the price of a house through the number of bedrooms, floors, bathrooms and the sqft_living.

As we can see in Figure 3, we have visualized a PDP for each one of these variables, giving as the following results:

- The partial dependence plot for the number of bedrooms shows that the predicted price of a house increases when there are 1, 2 or five bedrooms, meaning that this is the optimal number of bedrooms a house selling.
- The partial dependence plot for the number of bathrooms shows that the predicted price of a house increases gradually as the number of bathrooms increases. This means that houses with more bathrooms are generally more expensive than houses with fewer bathrooms. This is likely because more bathrooms are considered a luxury feature and can make the house more appealing to buyers.
- The partial dependence plot for the soft living floor area shows that the predicted price of a house increases as the soft living floor area increases. This means that houses with more soft living floor area are generally more expensive than houses with less soft living floor area. This is likely because soft living floor area is considered a desirable feature and can make the house more appealing to buyers.
- The partial dependence plot for the total square footage shows that the predicted price of a house increases as the total square footage increases. This means that larger houses are generally more expensive than smaller houses. This is because

larger houses provide more living space and make the house more appealing to buyers.

Deployment

This task has been deployed using Git orders. From the creation of the local repository until the upload of the repository, we have use a Linux Kernel to run the Git orders. After that, we created the repository in Github Desktop and started using Github Desktop, an easy application to work with Git Orders and Github projects.

The repository is https://github.com/Parri0627/Parri0627-EDM_XAI3_Deployment.

Conclusions

As a conclusion, we can affirm that in the first dataset, there are more bike rentals as time passes, humidity and wind speed decrease, and temperature rises. We have observed the particular relationship between temperature and humidity. On the other hand, we have noticed that as the number of bathrooms and soft living floor area increases, the price also increases. Additionally, to achieve the highest price, it is necessary to have more than 2 floors and 1, 2, or 5 bedrooms.

Index of figures

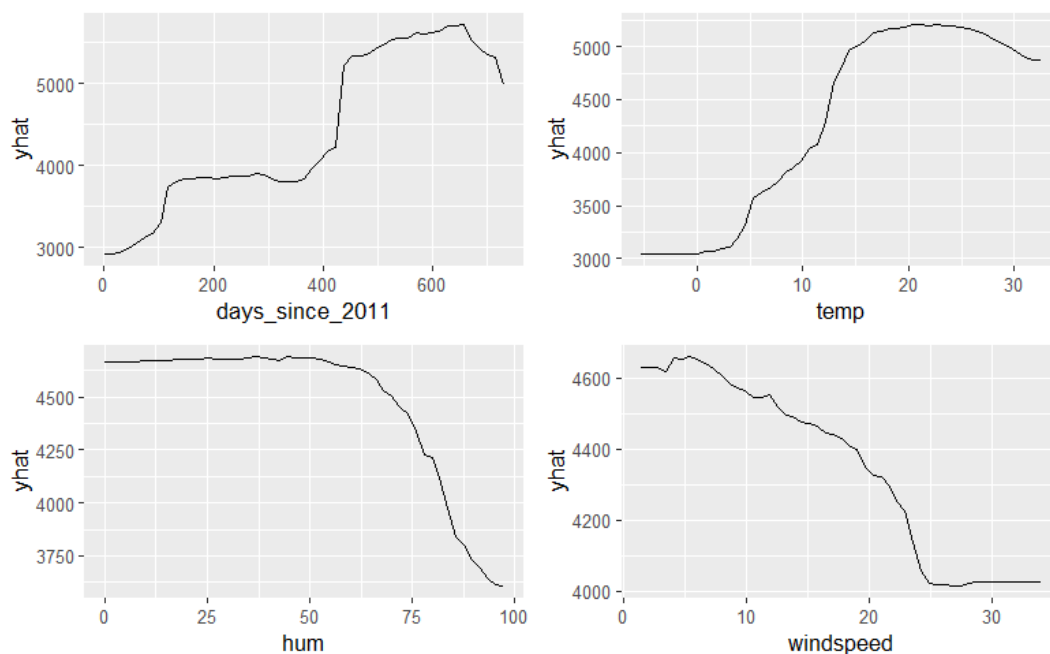


Figure 1: PDPs of the bike rentals variables

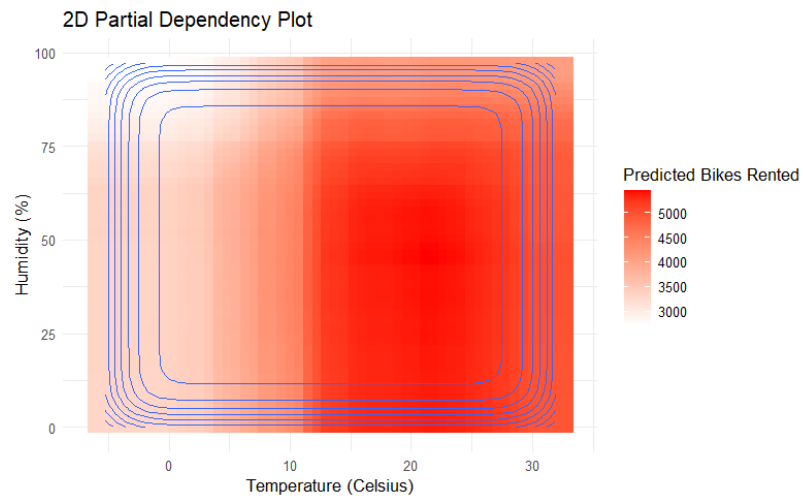


Figure 2: Bidimensional Partial Dependency Plot

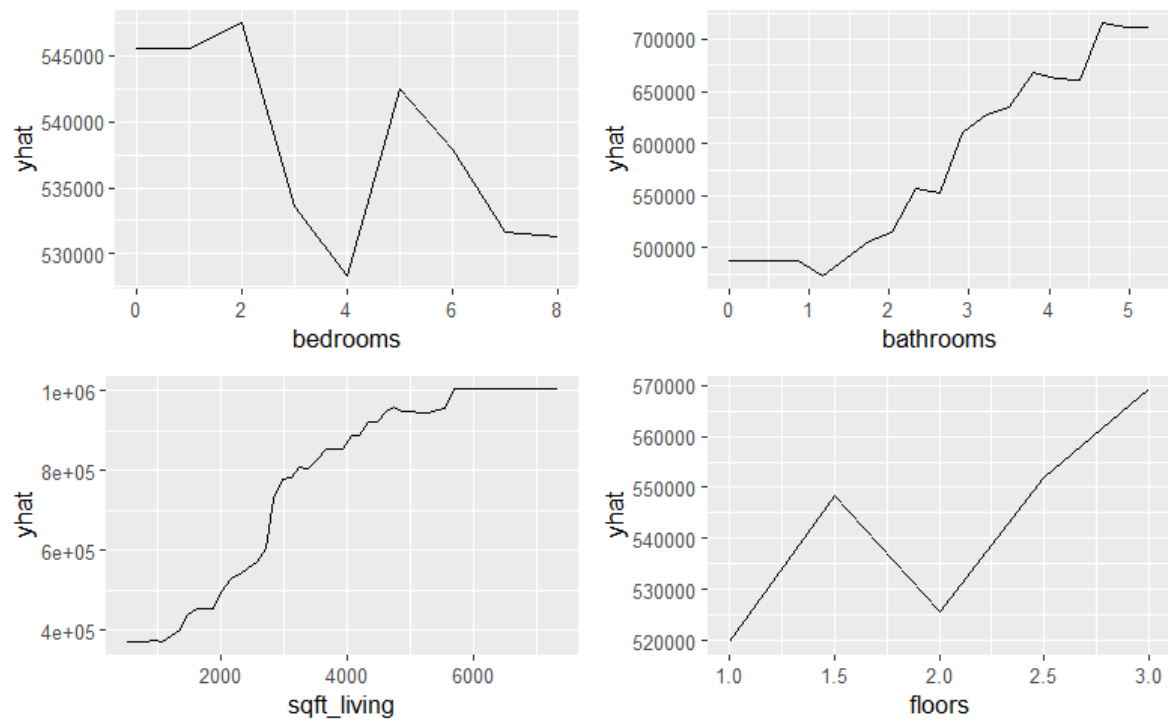


Figure 3: PDP's of the house Price variables

Pablo Parrilla Cañadas
 Claudia Martínez García
 3CD1