Deployment: Model-agnostic methods

Alexandru Neamtu & Borja García & Javier López

One dimensional Partial Dependence Plot 2

Introduction 2

Process 2

Results Interpretation 2

Conclusion 3

Bidimensional Partial Dependency Plot 4

Introduction 4

Process 4

Results Interpretation 4

Conclusion 5

PDP to explain the price of a house 5

Introduction 5

Process 5

Results Interpretation 6

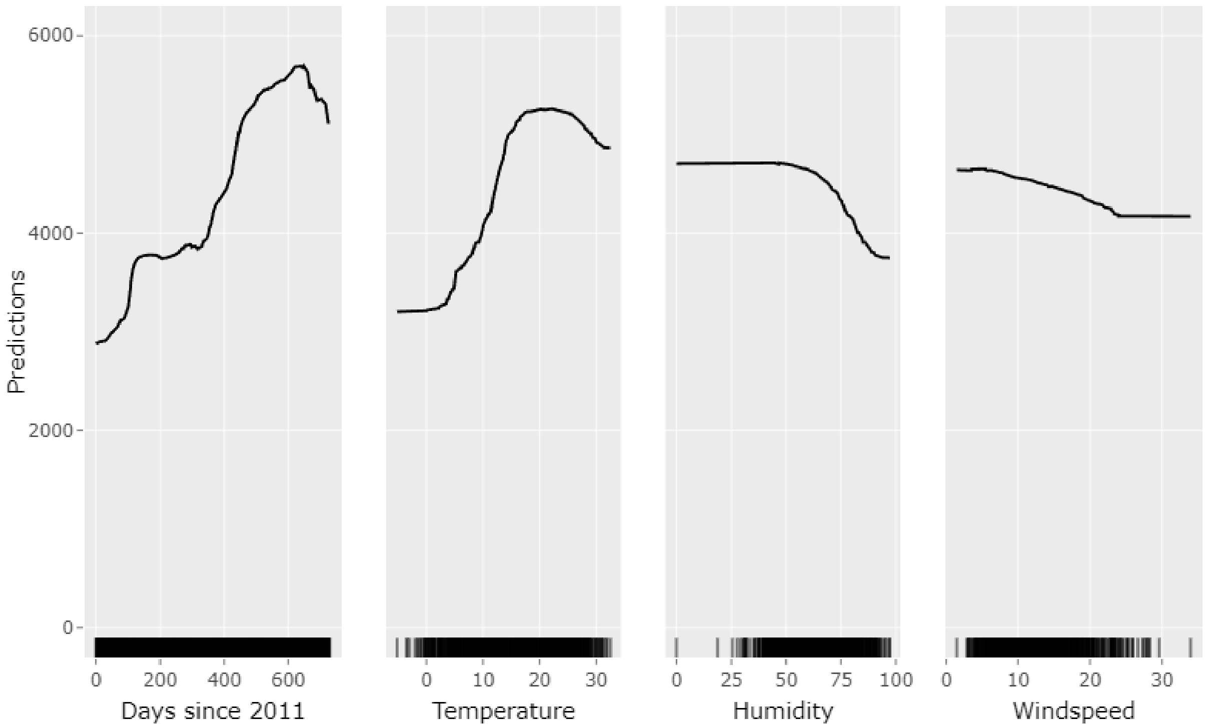
Conclusion 7

One dimensional Partial Dependence Plot

Introduction
In the analysis we will explore how different features influence the prediction of bike rental counts using a random forest model. To do this, we will use Partial Dependence Plots (PDP), which allow us to visualize the marginal effect of a feature on the predicted outcome while keeping all other features constant.

Process
First, we will fit a random forest model to predict the number of bike rentals (cnt). Then, we will use partial dependence plots to analyze the relationships the model has learned between bike rental counts and four specific features: Days since 2011, Temperature, Humidity and Wind speed.

Results Interpretation



This analysis focuses on how different variables affect bicycle rental forecasts. For temperature, it is observed that the number of bikes rented increases with temperature up to about 20 degrees Celsius, but then decreases as the temperature rises above 25 degrees Celsius. As for humidity, it remains constant up to 50%, but then decreases in proportion to the increase in humidity. Wind speed also influences bicycle rental forecasts, gradually decreasing until a wind speed of approximately 23 km/h is reached, from which the forecasts are constant. Finally, in terms of days since 2011, there is a general upward trend in bicycle rentals as time passes, although recent forecasts indicate a decrease.

Conclusion

These plots help us better understand how each of these features influences the predicted number of bike rentals, providing clearer insights into the relationships learned by the model.

Bidimensional Partial Dependency Plot

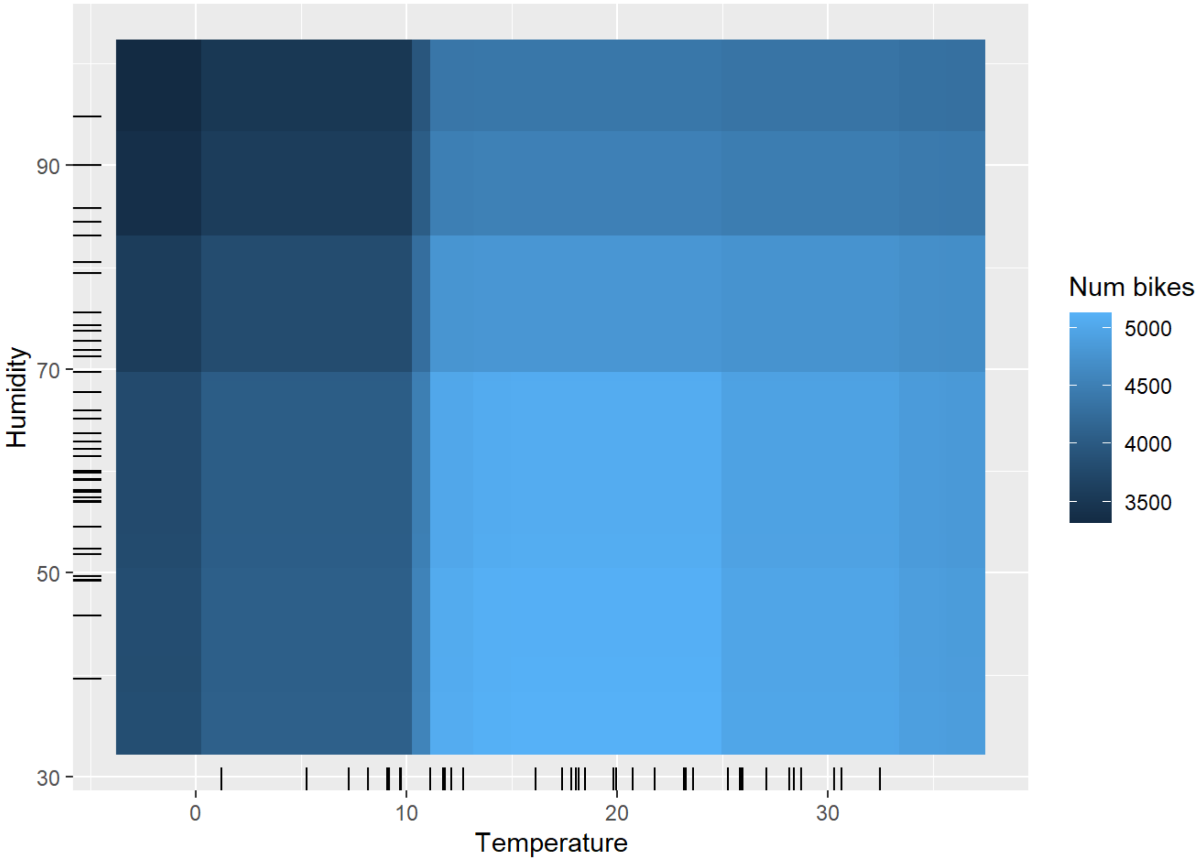
Introduction

In this approach, we will now generate a 2D Partial Dependence Plot (PDP) to visualize how two features, humidity and temperature, jointly influence the prediction of bike rental counts. This type of plot helps us understand the combined effect of these features on the predicted outcome.

Process

To create the 2D Partial Dependence Plot we will use our random forest model to predict the number of bike rentals (cnt). Due to the large size of the dataset, we will extract a random subset of samples before generating the data for the Partial Dependence Plot. We will visualize the density distribution of both humidity and temperature alongside the 2D plot using geom_tile(), ensuring that the width and height are set to avoid gaps in the plot.

Results Interpretation



The graph shows that the highest number of bicycles is rented when the temperature ranges between 15 and 20 degrees and the relative humidity between 0 and 70%. On the other hand, it can be observed that the number of bicycles rented decreases as the humidity is higher and the temperature is lower.

Conclusion
This approach allows us to interpret how changes in both humidity and temperature together affect the number of bikes rented, providing a more comprehensive understanding of the relationships learned by the model.

PDP to explain the price of a house

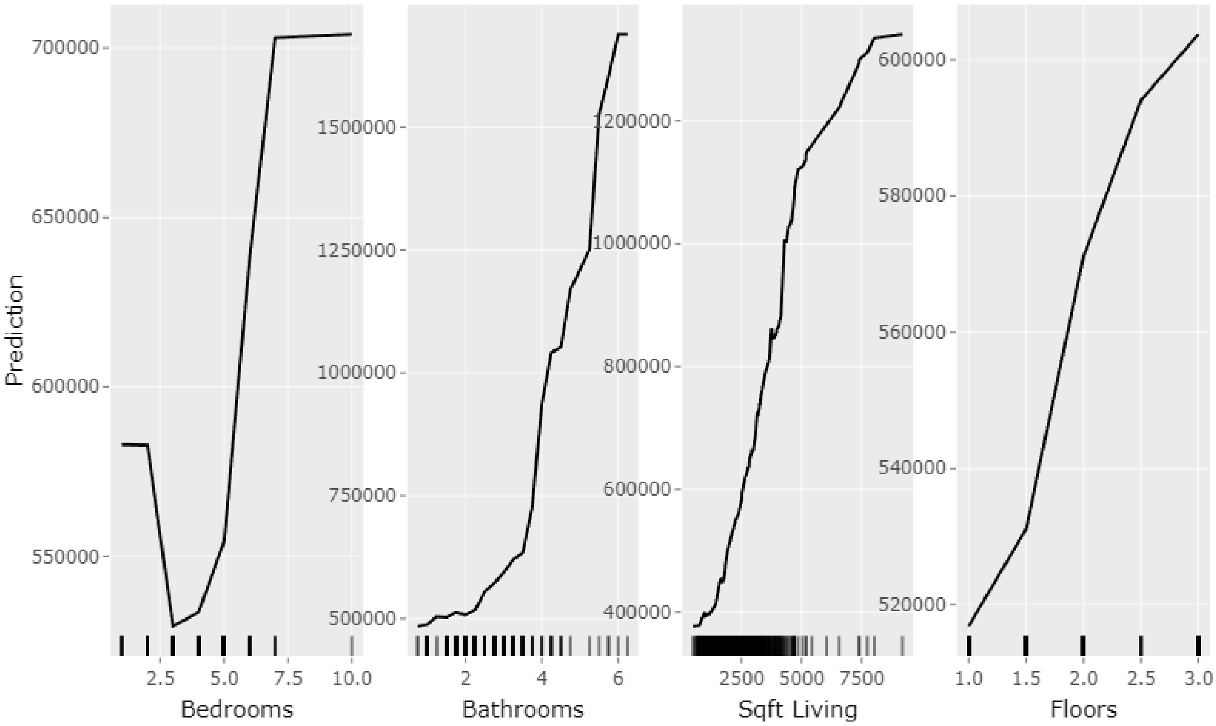
Introduction
Finally, we will explore how different features influence the prediction of house prices using a random forest model. To do this, we will use Partial Dependence Plots (PDP), which allow us to visualize the marginal effect of a feature on the predicted outcome while keeping all other features constant.

Process
First, we will fit a random forest model to predict the price of a house using the kc_house_data.csv dataset. The features we will consider for prediction are Bedrooms, Bathrooms, Square footage of the living space (sqft_living), Square footage of the lot (sqft_lot), Number of floors and Year built (yr_built).

Given the size of the dataset, we will extract a random sample before generating the data for the Partial Dependence Plot to ensure computational efficiency.

Using partial dependence plots, we will analyze the relationships the model has learned between the house prices and four specific features: Bedrooms, Bathrooms, Square footage of the living space and Number of floors.

Results Interpretation



For bedrooms, there is an interesting trend where the price of the house generally increases as the number of bedrooms increases. However, there are exceptions: houses with one or two bedrooms tend to be more expensive, with three-bedroom houses being the least expensive. Beyond three bedrooms, the price increases with the number of bedrooms.

Regarding bathrooms, there is a positive correlation with house prices. As the number of bathrooms increases, the price of the house also increases.

For livable square footage, there is a positive correlation with house prices. As the livable square footage increases, the price of the home increases as well.

Finally, the number of floors shows a positive correlation with house prices. As the number of floors increases, the price of the house also increases.

Conclusion

These plots will help us better understand how each of these features influences the predicted house prices, providing clearer insights into the relationships learned by the model. This understanding can aid in making informed decisions in real estate by highlighting key factors that affect house prices.