*May 16th, 2025*

**XAI Report: Model-Agnostic**

**Partial Dependency Plots (PDP)**

**Evaluation, Deployment and monitoring of models**

Data Science – UPV

*Ciobanu Borinschi, Luminita*

*Elena Navarro, Javier*

*Amores Giner, Pau*

**1. Introduction**

Partial Dependence Plots (PDP) are model-agnostic tools used to interpret machine learning models by visualizing the average effect of one or more features on the model's predictions. Unlike feature importance metrics, PDPs provide a more intuitive view of how each feature influences the predicted outcome by holding all other features constant. This is particularly useful for complex models like random forests, where the relationships between features and outcomes can be highly non-linear and difficult to interpret. PDPs can help identify both direct and indirect effects of features, revealing complex interactions that might not be immediately apparent.

In this report, two case studies are presented:

1. Bike rental prediction using the day.csv dataset.
2. House price prediction using the kc\_house\_data.csv dataset.

**2. Metodology**

**2.1 Bike Rental Analysis**

For the bike rental analysis, a random forest model was trained using the following features:

* **Year (yr)** - Coded as 0 for 2011 and 1 for 2012. This binary feature captures the impact of time on bike rental trends, including potential economic recovery and increased environmental awareness.
* **Temperature (temp)** - Normalized temperature in Celsius, capturing the influence of weather conditions on outdoor activities. This feature is likely to have a strong positive influence as people prefer to cycle in warmer weather.
* **Humidity (hum)** - Normalized humidity level, reflecting comfort levels for cyclists. High humidity can reduce physical comfort and discourage outdoor activities.
* **Wind Speed (windspeed)** - Normalized wind speed, accounting for physical resistance to cycling. High wind speeds can increase physical effort and reduce the attractiveness of cycling.

The goal was to predict the total bike rental count (cnt). The script used for this analysis is **bike\_rental\_pdp.R**, which generated the following PDPs:

* **PDP\_yr.png**
* **PDP\_temp.png**
* **PDP\_hum.png**
* **PDP\_windspeed.png**

**2.2 House Price Analysis**

For the house price analysis, a random forest model was trained using the following features:

* **Number of Bedrooms (bedrooms)** - Total number of bedrooms, representing the size and capacity of the house. More bedrooms can increase perceived value, but only up to a point.
* **Number of Bathrooms (bathrooms)** - Total number of bathrooms, indicating the comfort level and convenience for residents. This can be a major factor in house pricing, especially in high-income areas.
* **Living Area (sqft\_living)** - Total square footage of the living space, a critical factor in determining house value. This is often the most important predictor, as larger homes typically command higher prices.
* **Number of Floors (floors)** - Total number of floors in the house, reflecting the architectural complexity and potential for multi-generational living. However, this feature may have a more complex relationship with price, influenced by regional preferences and lot size.

The target variable was the house price (price). The script used for this analysis is **house\_price\_pdp.R**, which generated the following PDPs:

* **PDP\_bedrooms\_house.png**
* **PDP\_bathrooms\_house.png**
* **PDP\_sqft\_living\_house.png**
* **PDP\_floors\_house.png**

**3. Results**

**3.1 Bike Rental**

Imagen que contiene Diagrama

El contenido generado por IA puede ser incorrecto.

**PDP\_yr**: The plot shows a clear upward trend in bike rentals from 2011 to 2012. This increase likely reflects broader social trends, such as growing environmental awareness, improved cycling infrastructure, and shifting social attitudes toward sustainable transportation. Additionally, economic recovery post-2010 might have boosted disposable income, making recreational activities like cycling more accessible. This trend could also be influenced by increased urbanization and city planning efforts to reduce traffic congestion and pollution. In practice, this insight could guide urban planners and local governments in promoting sustainable transportation through improved bike lanes and public awareness campaigns.

Diagrama

El contenido generado por IA puede ser incorrecto.

**PDP\_temp**: This plot reveals a strong positive relationship between temperature and bike rentals. Warmer weather is naturally more conducive to outdoor activities, leading to higher rental volumes. The steep increase suggests a threshold effect, where small temperature increases significantly boost demand, possibly reflecting seasonal tourism, outdoor festivals, and recreational patterns. However, this relationship might vary by region, as extreme heat can also discourage cycling. For businesses, this insight can inform targeted marketing strategies for peak seasons or the design of cooling infrastructure to extend cycling viability.

Gráfico

El contenido generado por IA puede ser incorrecto.

**PDP\_hum**: The plot shows a moderate negative relationship between humidity and bike rentals. High humidity levels can make outdoor activities uncomfortable due to increased perspiration and reduced physical comfort, leading to lower rental counts. The less pronounced slope indicates that while humidity matters, it is not as critical as temperature. This effect may also depend on local climate tolerance and cultural factors. For example, cities in tropical regions may need to invest in shaded bike lanes or promote early morning or evening rides to mitigate the impact of humidity.

Diagrama

El contenido generado por IA puede ser incorrecto.

**PDP\_windspeed**: This plot indicates a generally negative impact of wind speed on bike rentals. High wind resistance increases physical exertion and reduces the perceived convenience of cycling, potentially discouraging users. However, the relatively flat curve suggests that cyclists may be more tolerant to moderate wind speeds, possibly reflecting experienced riders or more aerodynamic bike designs. This insight could inform the design of wind-protected cycling paths or promotional efforts targeting more experienced cyclists.

**3.2 House Price**

Imagen que contiene Gráfico

El contenido generado por IA puede ser incorrecto.

**PDP\_bedrooms**: The relationship between the number of bedrooms and house price is generally positive, as more bedrooms typically indicate a larger house and potentially more privacy and flexibility for families. However, the trend is not perfectly linear, likely due to variations in house design, location, and luxury features that can significantly alter value. In some cases, too many bedrooms without corresponding living space can actually reduce a home's appeal, highlighting the need for balanced architectural design.

Gráfico

El contenido generado por IA puede ser incorrecto.

**PDP\_bathrooms**: A positive correlation is observed between the number of bathrooms and house price. Homes with more bathrooms tend to offer greater convenience, privacy, and comfort, all of which increase market appeal and resale value. This is particularly important in markets where multi-family living is common or in regions with high population density, where the ability to accommodate larger families can significantly boost property value.

Gráfico

El contenido generado por IA puede ser incorrecto.

**PDP\_sqft\_living**: This is the most influential variable, showing a strong positive relationship with house price. Larger living spaces significantly drive up house prices, reflecting their desirability in the real estate market. This correlation is often amplified in urban areas where space is at a premium, highlighting the importance of location in real estate pricing. Developers can use this insight to prioritize large living spaces in new constructions to maximize profitability.

Diagrama

El contenido generado por IA puede ser incorrecto.

**PDP\_floors**: The plot shows a moderate positive relationship, indicating that multi-story homes tend to have higher prices. This is likely because multi-story homes can maximize living space on smaller plots, offering more versatility in high-density areas. However, the effect is less consistent, reflecting the diverse architectural styles and preferences across regions. Builders might consider this when designing homes for different market segments.

**4. Conclusions, Recommendations, and Limitations**

The PDPs generated for both case studies highlight the power of these plots in uncovering complex relationships within machine learning models. The bike rental analysis demonstrated that temperature is the most critical factor in rental demand, highlighting the strong seasonal component of this market. In contrast, the house price analysis confirmed that living area (sqft\_living) is the primary driver of property value, reflecting the importance of space in residential real estate.

However, it is important to note that PDPs have limitations. They assume feature independence, which can lead to misleading interpretations when features are highly correlated. Additionally, PDPs do not capture complex feature interactions unless explicitly modeled. Future analyses could benefit from using more advanced interpretability techniques, such as SHAP values or ALE plots, to provide a more comprehensive understanding of model behavior.

Overall, these analyses illustrate the importance of model interpretability in machine learning, emphasizing the role of PDPs in providing transparency and actionable insights for decision-making.

**Recommendations:**

* **Improving Model Accuracy:** Consider incorporating additional features, such as location data for house prices or seasonal variables for bike rentals, to capture more complex interactions and reduce model bias.
* **Hyperparameter Tuning:** Optimizing hyperparameters, such as the number of trees or maximum depth in the random forest, could improve the accuracy and interpretability of the models.
* **Advanced Interpretability Methods:** Future analyses could benefit from more advanced techniques, such as SHAP values or ALE plots, which can capture complex, non-linear feature interactions and provide more precise attributions.
* **Business Applications:** Use these insights to design targeted marketing campaigns for bike rentals during peak seasons or prioritize high-demand housing features for new real estate developments.