

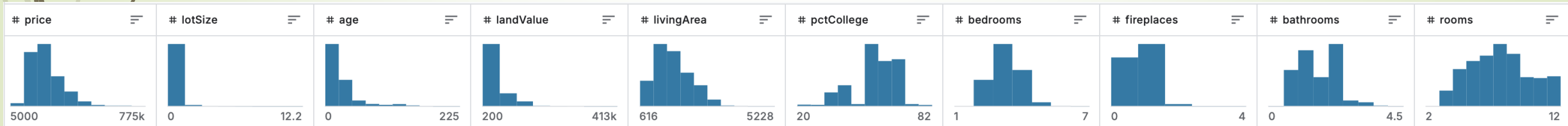


# Black Box vs Modelos Interpretables

José Luis Mezquita Jiménez, alumno del máster en Inteligencia Artificial

# Saratoga Houses

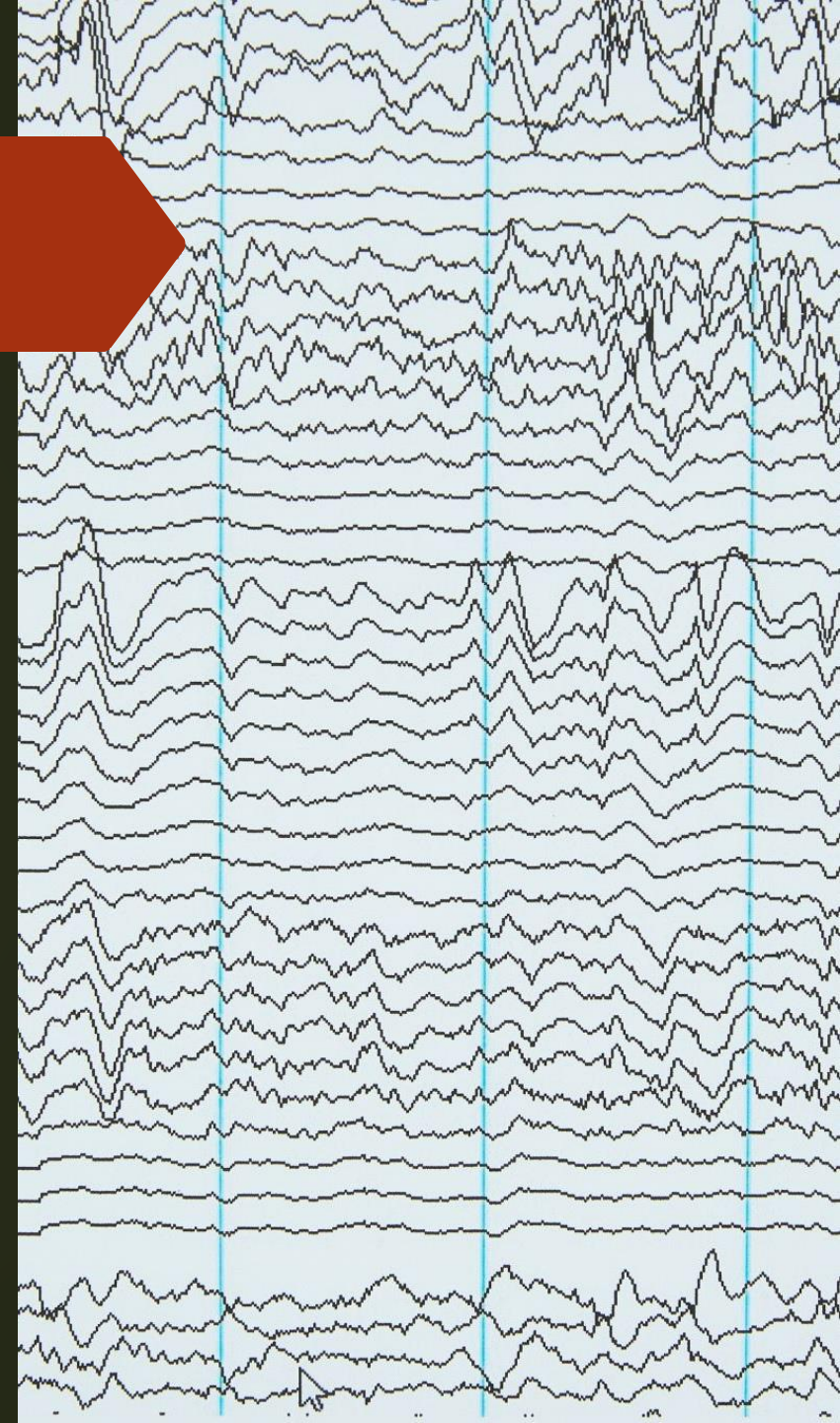
- El **dataset Saratoga Houses** contiene información sobre diferentes propiedades inmobiliarias en la zona de Saratoga, con características tanto numéricas como categóricas que describen aspectos como el tamaño del terreno, el área habitable, el número de habitaciones y baños, la edad de la casa, el tipo de calefacción, la existencia de aire acondicionado, entre otros. El objetivo principal suele ser predecir el valor o precio de las casas a partir de estas características.





# Modelos Interpretables

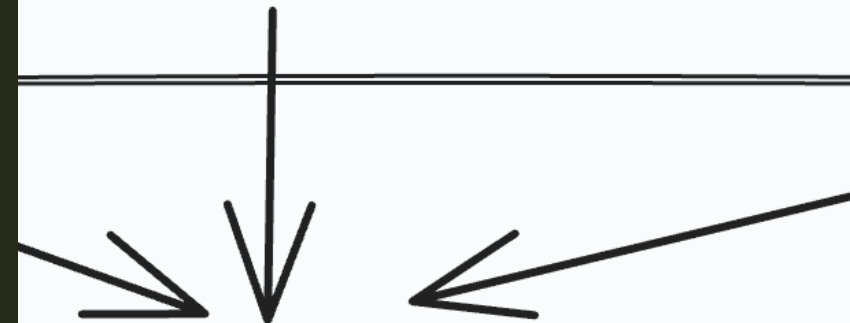
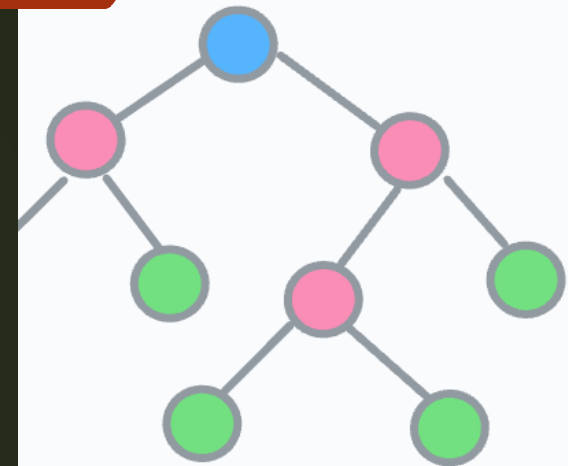
- Se han comparado 3 algoritmos explicables y estos son los resultados:
- **Regresión Lineal** presenta el mejor desempeño general entre los tres modelos, con el mayor valor de  $R^2$  (**0.6355**), lo que indica que explica aproximadamente el 63.5% de la variabilidad en los datos. Además, tiene los valores más bajos en MAE (43,941.50) y RMSE (64,195.57), lo que significa que sus predicciones son en promedio más cercanas a los valores reales y tienen menos errores grandes comparado con los otros modelos.
- **K-Vecinos más cercanos (KNN)** muestra un rendimiento intermedio, con un  $R^2$  de **0.5920**. Aunque es menos preciso que la regresión lineal, sigue siendo competitivo, especialmente porque KNN puede capturar relaciones no lineales que un modelo lineal no puede. Sin embargo, sus errores (MAE y RMSE) son mayores que los de regresión lineal, indicando menos precisión en promedio.
- **Árbol de Decisión (DT)** tiene el peor desempeño en este conjunto, con un  $R^2$  de **0.5598** y errores más altos (MAE: 46,299.48, RMSE: 70,544.17). Esto puede deberse a que los árboles de decisión tienden a sobreajustar con datos pequeños o variables ruidosas, o que el modelo no fue optimizado completamente.



# Modelos de caja negra

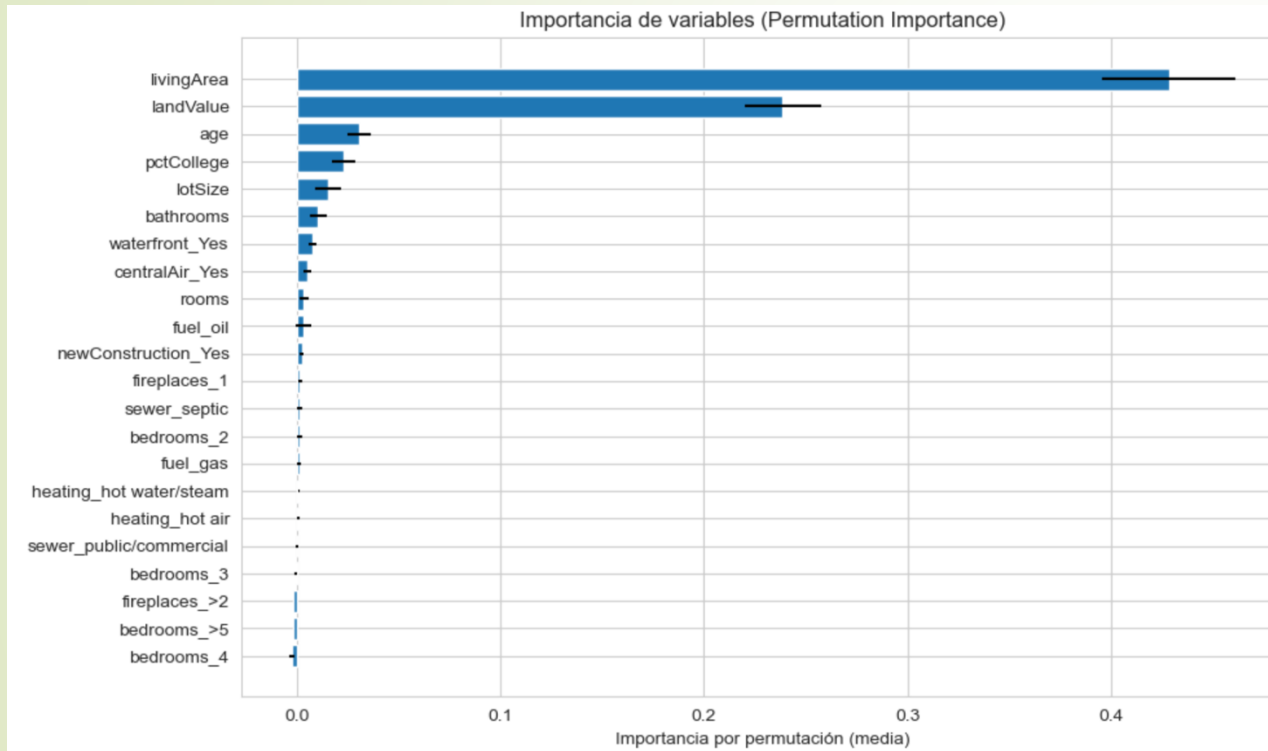
- Se ha utilizado un modelo de Random Forest con varias mejoras, aplicando como método agnóstico la técnica de Permutation Importance.
- Para ello, se entrenaron dos modelos:
  1. El primer modelo se entrenó usando las mismas columnas y datos que el modelo interpretable. Posteriormente, se aplicó Permutation Importance para identificar cuáles variables eran las más relevantes para el modelo.
  2. El segundo modelo se entrenó con el mismo algoritmo, pero excluyendo las columnas que en el primer modelo mostraron resultados negativos (es decir, variables que podrían estar confundiendo al modelo).

om Forest M



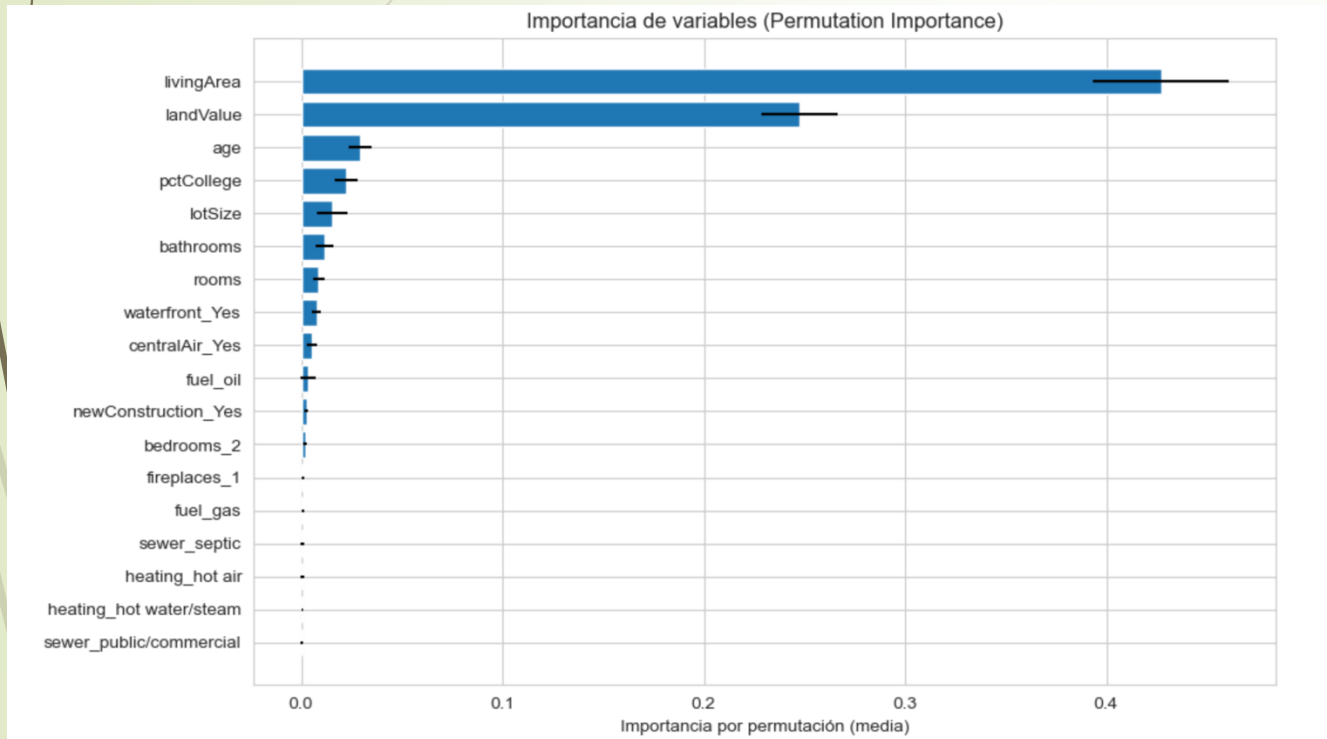
Prediction

# Resultados Modelo Agnóstico (Modelo 1)



Feature	Importance Mean	Importance Std
livingArea	0.4285	0.0328
landValue	0.2389	0.0185
age	0.0308	0.0057
pctCollege	0.0229	0.0057
lotSize	0.0154	0.0061
bathrooms	0.0104	0.0041
waterfront_Yes	0.0076	0.0018
centralAir_Yes	0.0054	0.0021
rooms	0.0037	0.0024
fuel_oil	0.0033	0.0036
newConstruction_Yes	0.0027	0.0010
fireplaces_1	0.0018	0.0007
sewer_septic	0.0016	0.0010
bedrooms_2	0.0014	0.0011
fuel_gas	0.0013	0.0009
heating_hot water/steam	0.0011	0.0004
heating_hot air	0.0008	0.0007
sewer_public/commercial	0.0002	0.0007
bedrooms_3	-0.0002	0.0007
fireplaces_>2	-0.0017	0.0003
bedrooms_>5	-0.0018	0.0001
bedrooms_4	-0.0022	0.0014

# Resultados Modelo Agnóstico (Modelo 2)



Feature	Importance Mean	Importance Std
livingArea	0.4274	0.0338
landValue	0.2475	0.0189
age	0.0294	0.0057
pctCollege	0.0226	0.0056
lotSize	0.0154	0.0075
bathrooms	0.0117	0.0042
rooms	0.0086	0.0028
waterfront_Yes	0.0076	0.0021
centralAir_Yes	0.0055	0.0025
fuel_oil	0.0036	0.0039
newConstruction_Yes	0.0026	0.0011
bedrooms_2	0.0019	0.0009
fireplaces_1	0.0010	0.0008
fuel_gas	0.0009	0.0006
sewer_septic	0.0007	0.0010
heating_hot air	0.0006	0.0007
heating_hot water/steam	0.0005	0.0004
sewer_public/commercial	0.0004	0.0007



# Condiciones para eliminar columnas

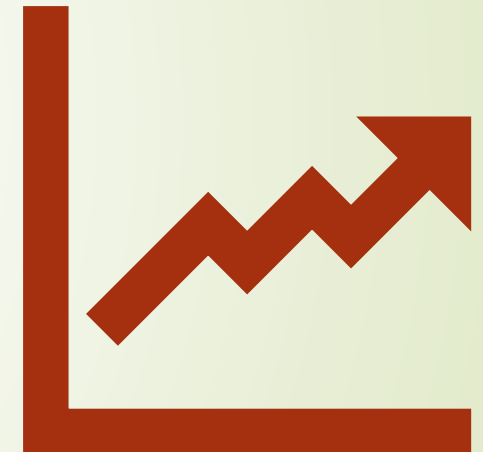
Para eliminar las columnas se han tenido en cuenta las siguientes condiciones:

- Valores grandes positivos  $\rightarrow$  esa variable es muy útil para el modelo.
- Valores cercanos a 0  $\rightarrow$  esa variable da igual.
- Valores negativos  $\rightarrow$  esa variable puede estar confundiendo al modelo. Por lo tanto, podemos eliminarla.



# Comparativa de modelos:

- **Random Forest:** Mejora el  $R^2$  al 0.6586 frente al 0.6355 de la regresión lineal, y disminuye MAE y RMSE (40,776 vs 43,942 y 62,124 vs 64,196 respectivamente), mostrando mayor capacidad para captar relaciones no lineales y complejas entre variables. Además, el método de Permutation Importance facilita interpretar la importancia de cada variable, algo que a veces puede ser más difícil en modelos más complejos.
- **Regresión Lineal:** Buen desempeño general, con buena interpretabilidad, pero limitado para relaciones no lineales o interacciones complejas. Sus errores son algo mayores y explica un poco menos de varianza.
- **KNN:** Captura relaciones no lineales, pero su desempeño es inferior al de Random Forest y regresión lineal en este caso, con  $R^2 = 0.5920$  y errores más altos, lo que indica que probablemente no logra generalizar tan bien o que el dataset no favorece mucho este tipo de modelo.
- **Árbol de Decisión:** Peor desempeño, con  $R^2 = 0.5598$  y errores más altos, posiblemente por sobreajuste o falta de optimización.





# Conclusiones

- El modelo Random Forest, complementado con Permutation Importance, proporciona un **mejor equilibrio entre precisión y explicabilidad** en comparación con los modelos clásicos evaluados. Este enfoque no solo alcanza métricas superiores, sino que también identifica de forma clara las variables que más influyen en la predicción. Aunque la regresión lineal destaca por su simplicidad e interpretabilidad, sacrifica algo de precisión, mientras que otros modelos no alcanzan su nivel de desempeño.

