

Tidymodels

¿Qué es modelar?



"Modelar es el proceso de desarrollo de una herramienta matematica que genera una prediccion precisa"

Latin R

Tidy hace referencia a ordenar, su objetivo es facilitar el ciclo de trabajo en el aprendizaje automático.

¿Qué es modelar?

1. Predecir el valor de una variable resultante en base a valores conocidos de las variables predictoras. Aquí no interesa tanto entender cómo es que las variables interactúan entre sí, o por qué lo hacen. Mientras las predicciones sean acertadas, o se acerquen lo suficiente, el modelo cumple su cometido. Los modelos predictivos se emplean en una enorme variedad de aplicaciones: inversión en bolsa, prevención de fraude, publicidad online, fijación de primas en seguros de riesgo, etc.
2. Explicar la relación entre una variable dependiente y todas las demás (las explicativas), buscando determinar si la relación es significativa. Los modelos explicativos son los que se favorecen en investigación académica, ya que ayudan a entender el fenómeno modelado.

Conceptos

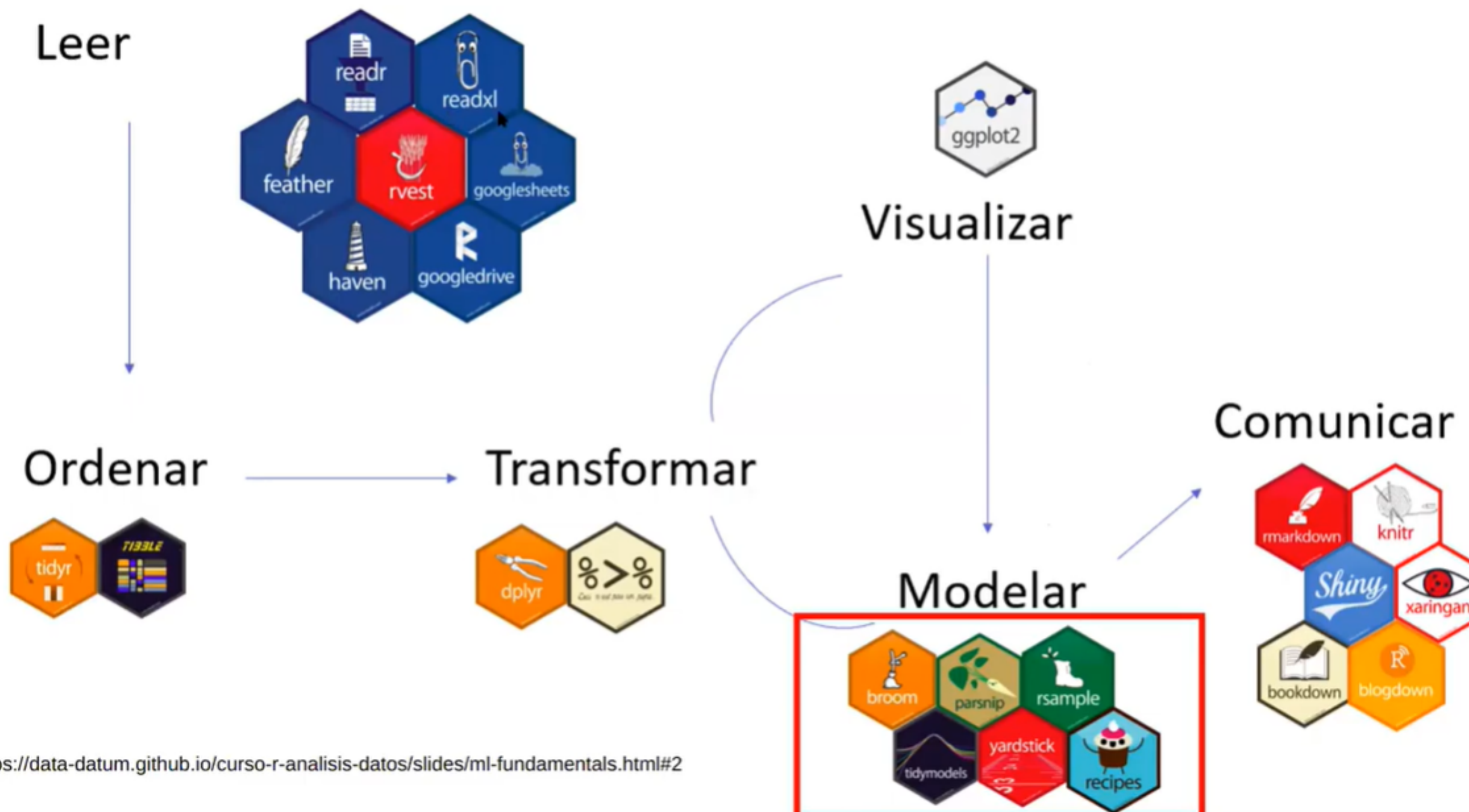
1. Muestra, observacion, punto se refiere a la unidad de analisis
2. Set de entrenamiento: datos utilizados para el modelado
3. Set de prueba: datos utilizados para medir el desempeño del modelo
4. Atributos, predictores, variables independientes: son los datos de entrada para la ecuacion de prediccion
5. Salida, variable de respuesta, variable dependiente: es la variable a predecir



Tipos de Aprendizaje

1. Aprendizaje supervisado: Regresion y clasificacion
2. Aprendizaje no supervisado: clustering y reduccion de dimensionalidad
3. Aprendizaje por refuerzo: utilizado para la robotica





Tidymodels

Es un grupo de paquetes centrados en las tareas de modelización de los datos, la modelización consta de varios pasos y para cada uno de ellos se utiliza una librería diferente.



Correlacion y Regresion Lineal

La correlacion lineal y la regresion lineal simple son métodos estadísticos que estudian la relación lineal existente entre dos variables. La correlacion se utiliza previamente a haber controlado las variables y las cuantifica, mientras que la regresion es un modelo para medir la relacion entre variables.

En la correlacion existe una fuerza de asociacion:

0: asociación nula.

0.1: asociación pequeña.

0.3: asociación mediana.

0.5: asociación moderada.

0.7: asociación alta.

0.9: asociación muy alta.

Tambien puede ser positiva o negativa.



Correlacion no es causalidad



A los datos!



Gracias!

WWW.INSTRUMENTA.COM.AR | INFO@INSTRUMENTA.COM.AR | ☎ 351 800 20 40

AV. FIGUEROA ALCORTA 482 | CENTRO | X5000KFR | CÓRDOBA