

Modelo de Predicción con Machine Learning

Demanda de servicio de bicicletas en
Seúl Korea

Proto
type

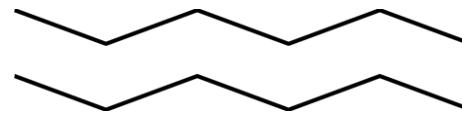
BRIEF



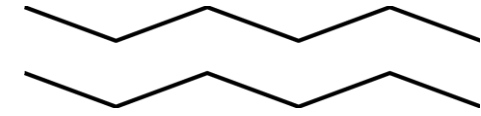
• • • DATASET



La presente *dataset* recopila las condiciones climatológicas (*Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall*) de la ciudad de Seúl, Korea, y el uso de las bicicletas públicas rentadas en diferentes puntos de la ciudad



• • • • OBJETIVO



Objetivo:

Crear un modelo predictivo de la demanda del servicio



ANÁLISIS

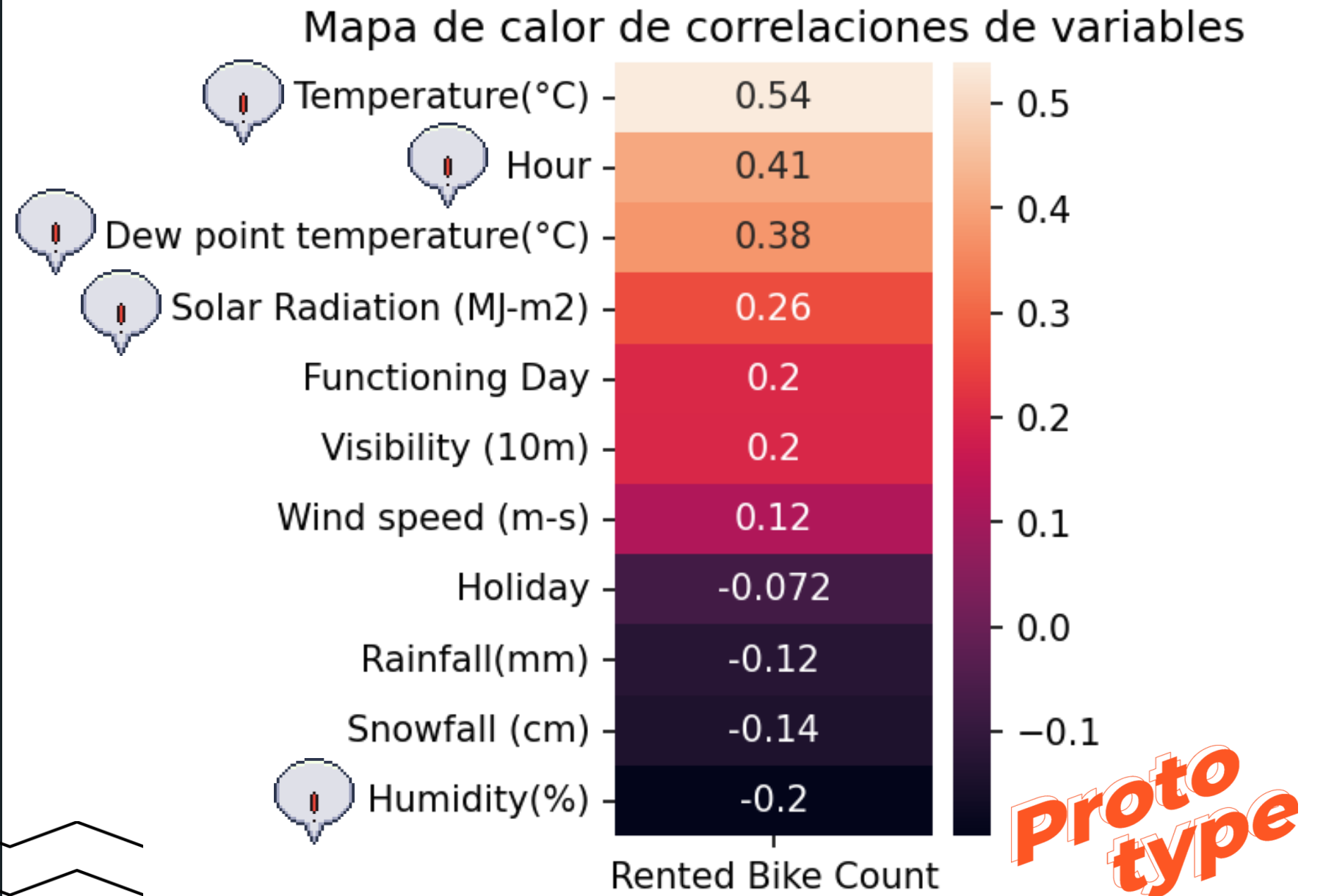


Variables

Variables con baja correlación simple de Pearson

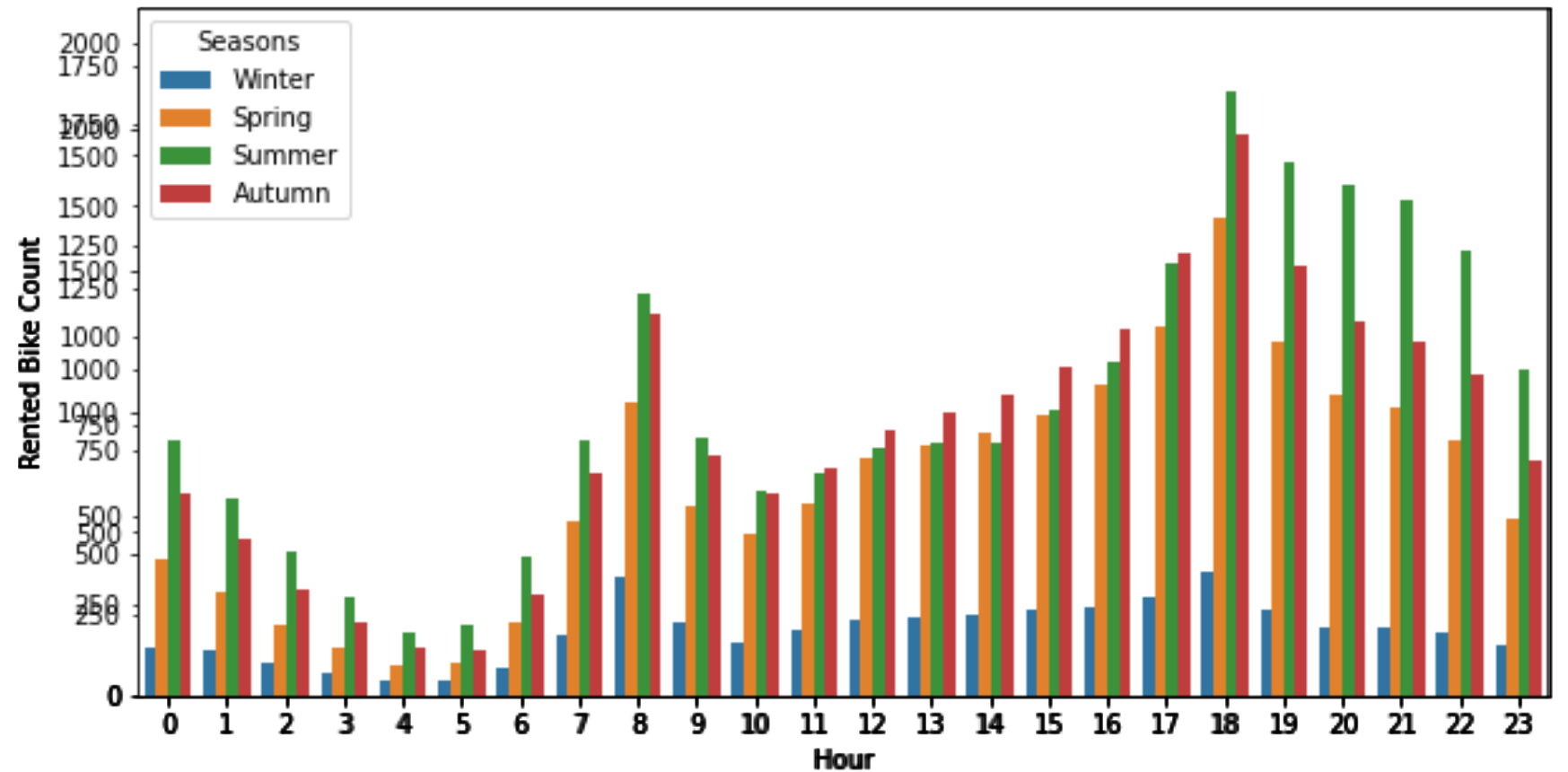
Sin embargo destacan:

- Temperatura ▲
- Hora del día ▲
- Punto de rocío ▲
- Radiación solar ▲
- Humedad ▼



Demanda por hora

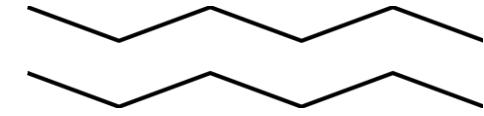
- Demanda por radiación solar
- Demanda por visibilidad
- Demanda por estación del año



Modelos de regresión



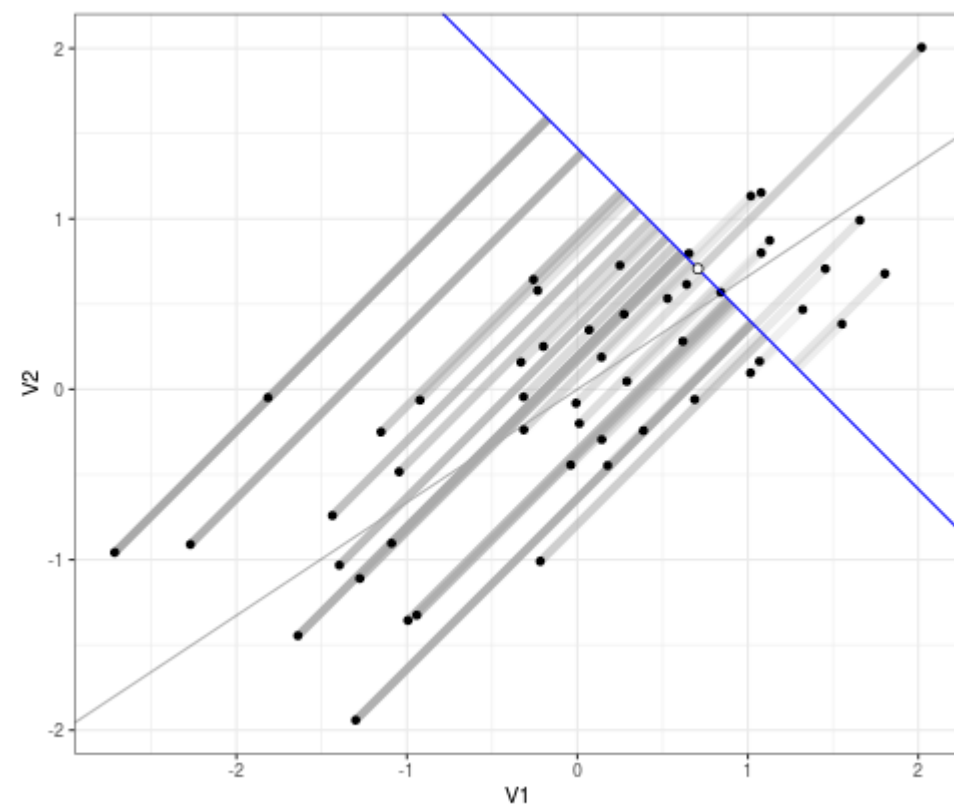
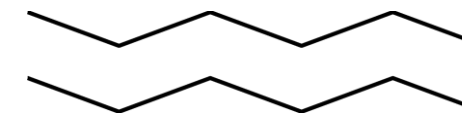
■ ■ ■ ■ **MODELOS**



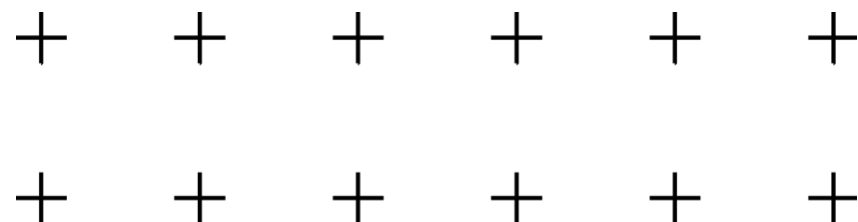
Los modelos utilizados son los siguientes:

+	+	+	+	+	+
+	+	+	+	+	+

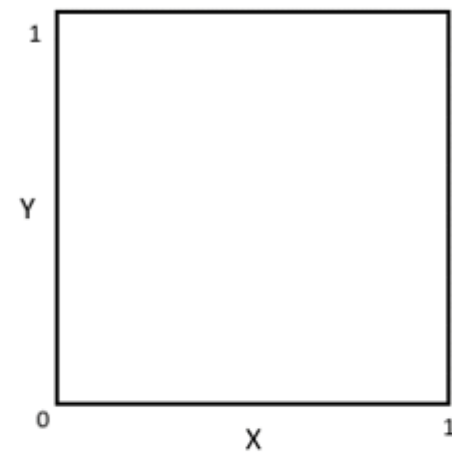
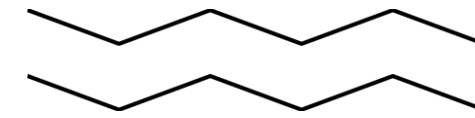
■ ■ ■ ■ MODELOS



Regresión lineal y polinomial

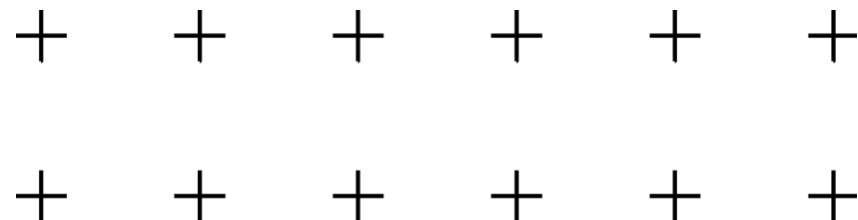


■ ■ ■ ■ **MODELOS**

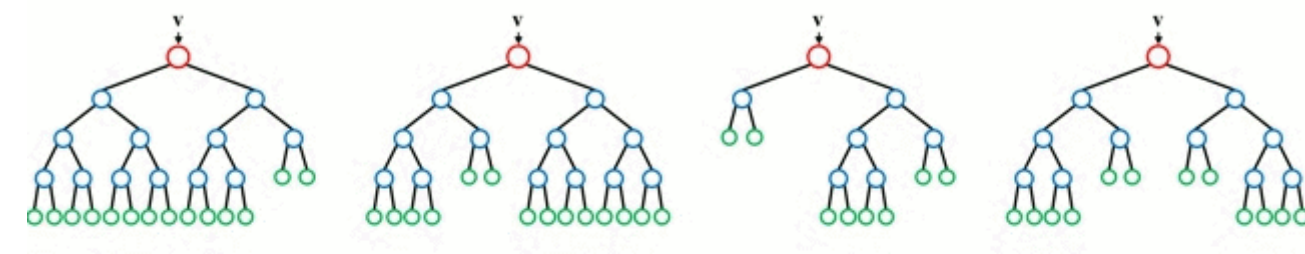
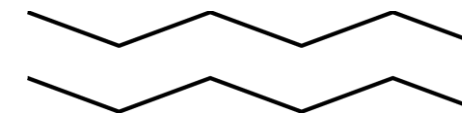


Árboles de Decisión

For more tutorials: annalysin.wordpress.com



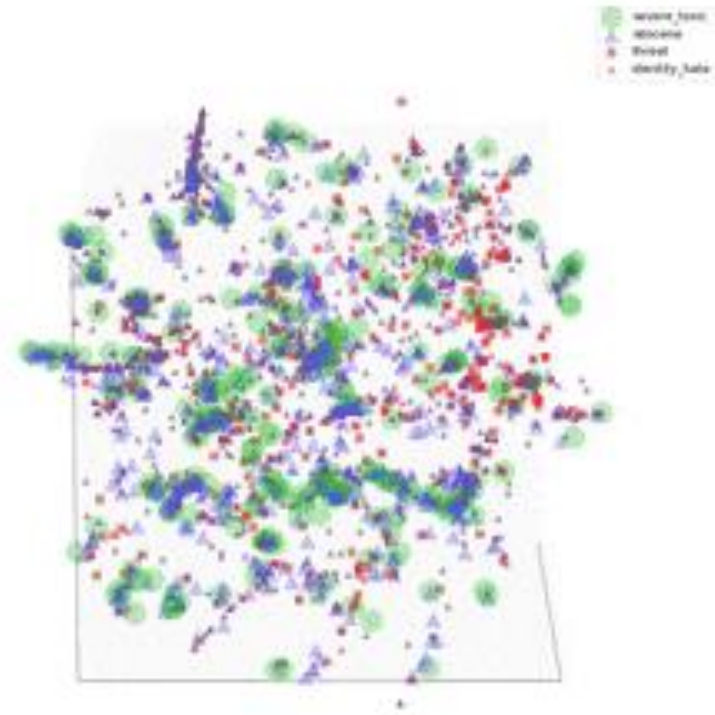
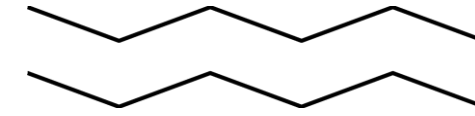
MODELOS



Random forest

+ + + + + +
+ + + + + +

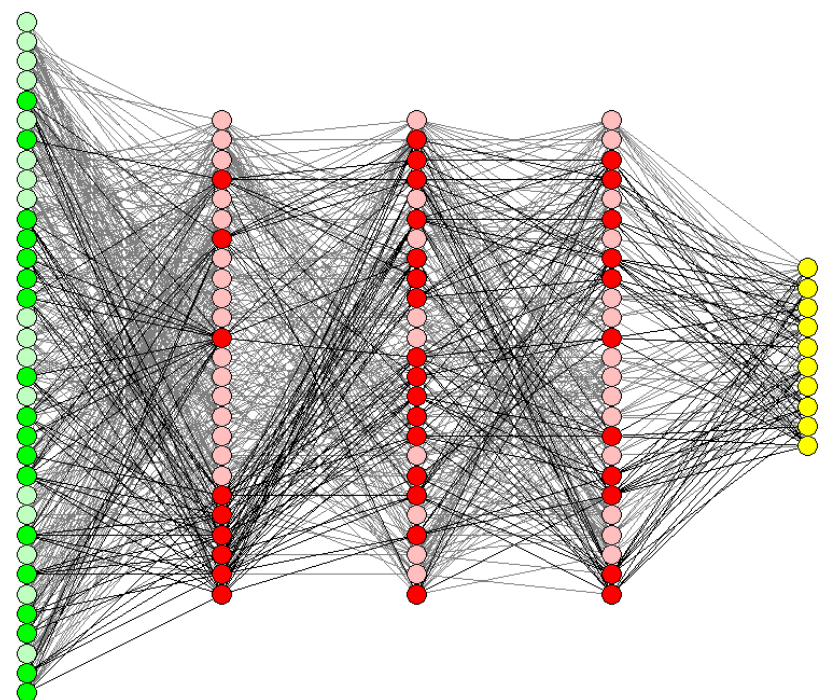
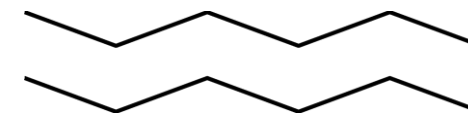
■ ■ ■ ■ **MODELOS**



Marco de refuerzo de Gradiente
(LightGBM Regressor)

+ + + + + +
+ + + + + +

■ ■ ■ ■ **MODELOS**



Redes Neuronales

+	+	+	+	+	+
+	+	+	+	+	+

■ ■ ■ ■ SCORE

A continuación se muestran los diferentes scores de validación cruzada de cada método (R^2)

-0.715

Regresión lineal

0.508

Árbol de decisión

0.524

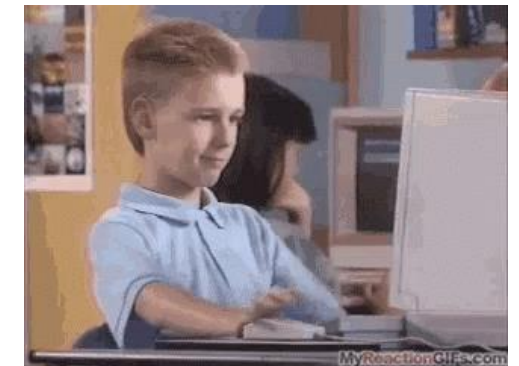
Random Forest

0.658

Light GBMr

0.831

Neural Network

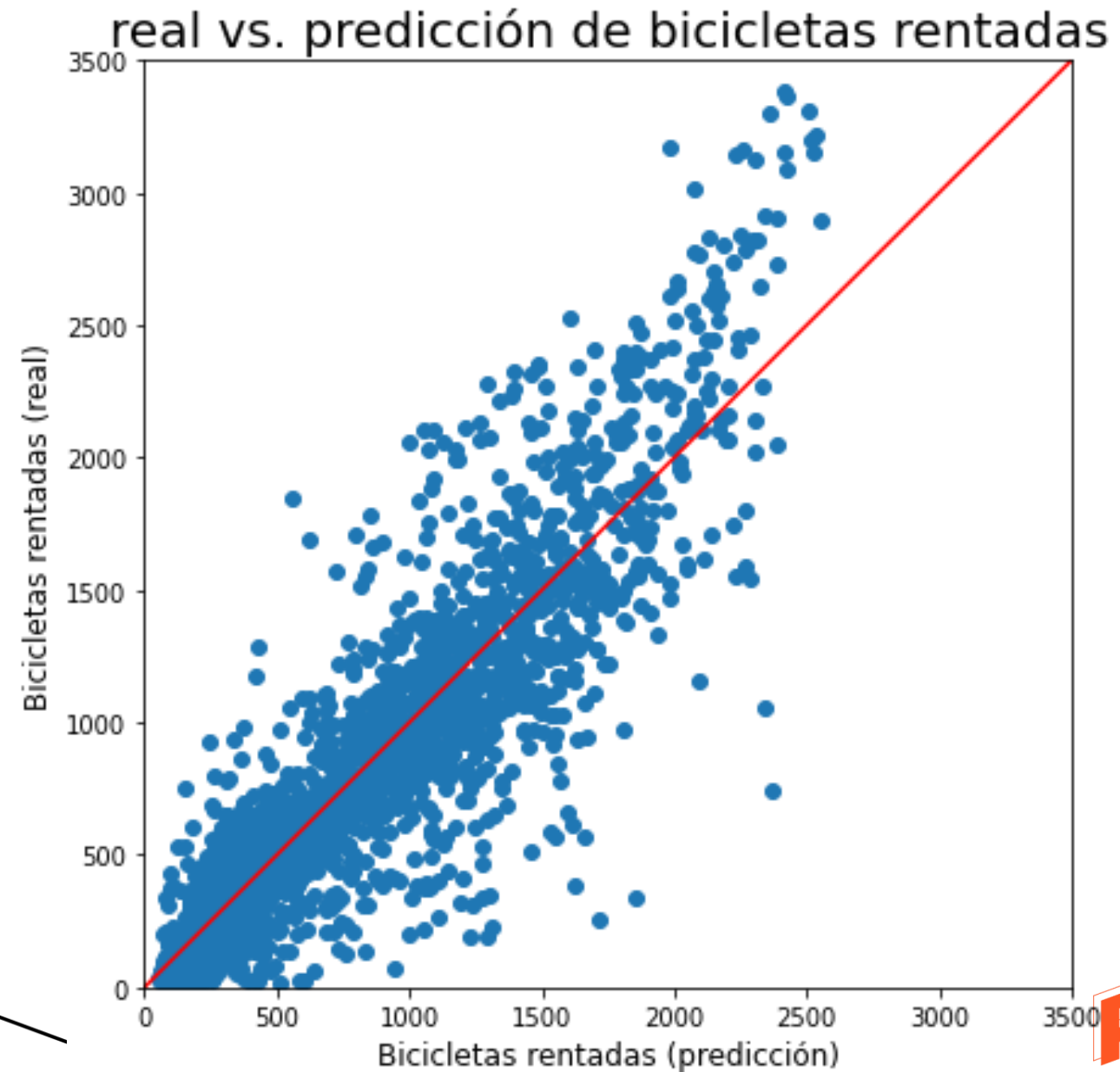


PREDICCIÓN



Comparación

El modelo de predicción comparándolo con el real es bastante competente



Proto
type

+ + + + +
+ + + + +

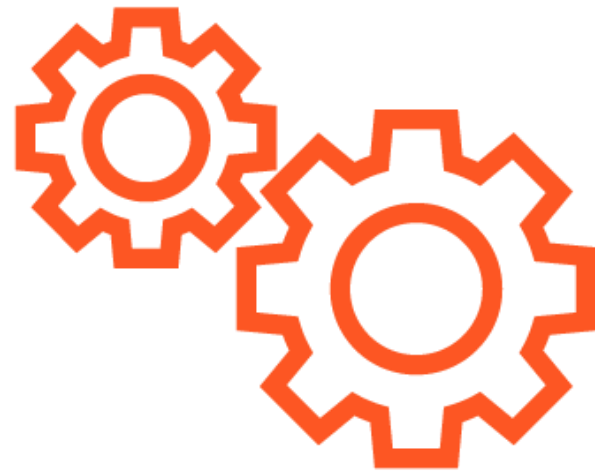
Modelo de predicción



INSIGHTS



■ ■ ■ ■ Insights



- Es sumamente importante la recopilación de datos para cualquier empresa que desee desarrollar una ventaja competitiva
- Las aplicaciones de la ciencia de datos son inmensas y urge que muchas empresas medianas y grandes diseñen a los datasets de sus ERP con las características necesarias para responder preguntas estratégicas
- El actual dataset proviene de un registro anual del 2017, quizá pueda ser más valioso actualizar la data para actualizar el modelo