Ezequiel Shinzato

Análisis Predictivo

Competencia Kaggle

Ezequiel Shinzato		0.60894	12

Dificultades

- Tamaño del dataset
- Tiempo
- Recursos
- Inexperiencia

Análisis exploratorio

Generalidades

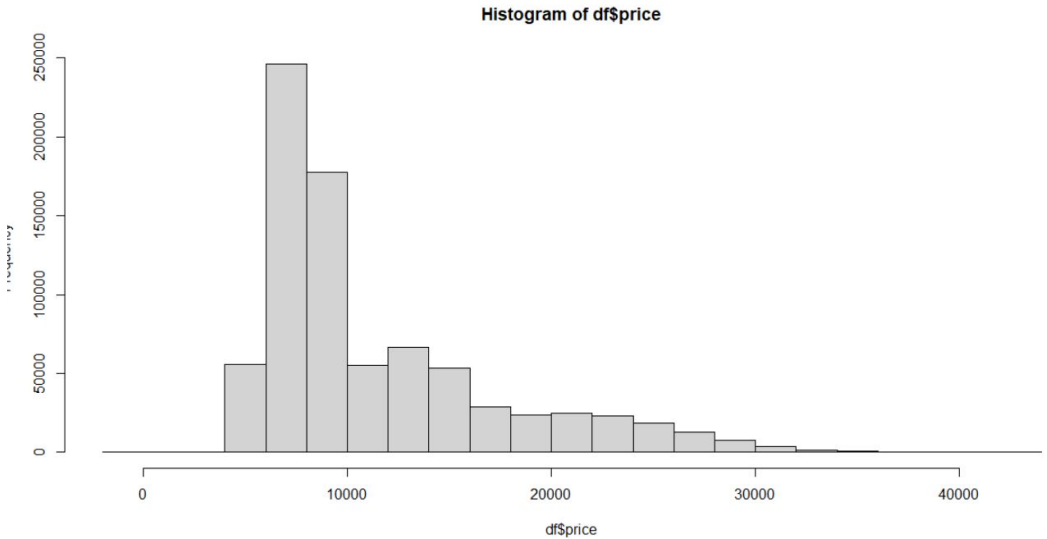
Base_train = 800.000 registros, 17 variables (-id)

Base_val = 200.000 registros, 16 variables (-id)

- No hay NAs en los datasets

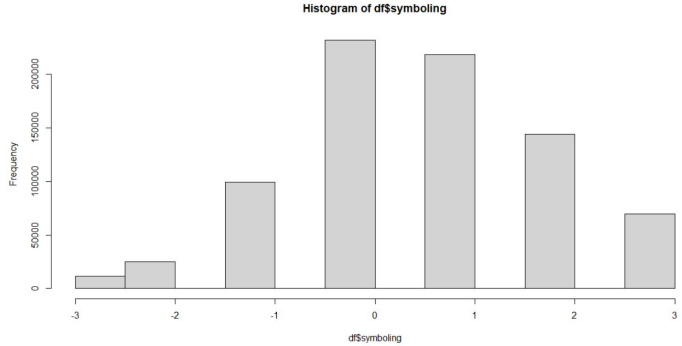
Price

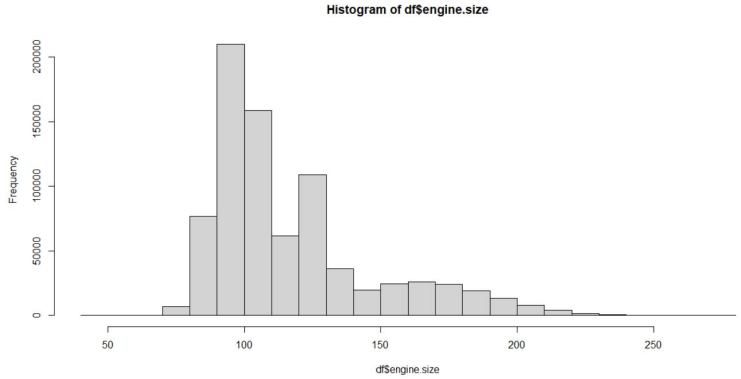
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-824.3  7461.9  9142.2 11440.9 13976.7 43392.4
```

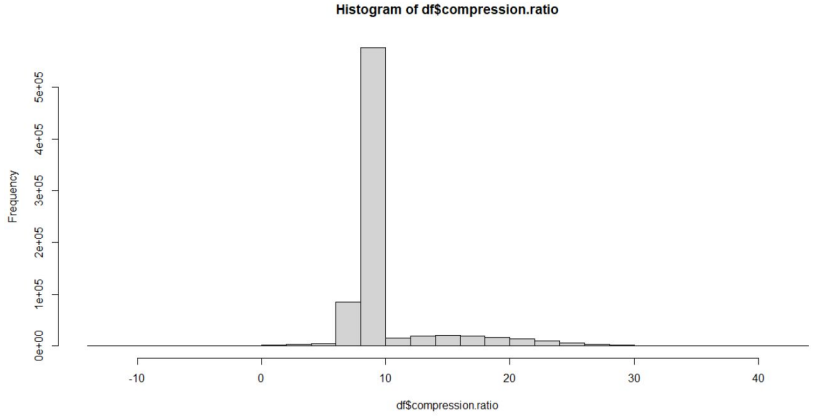


Test de KS → No es una distribución normal

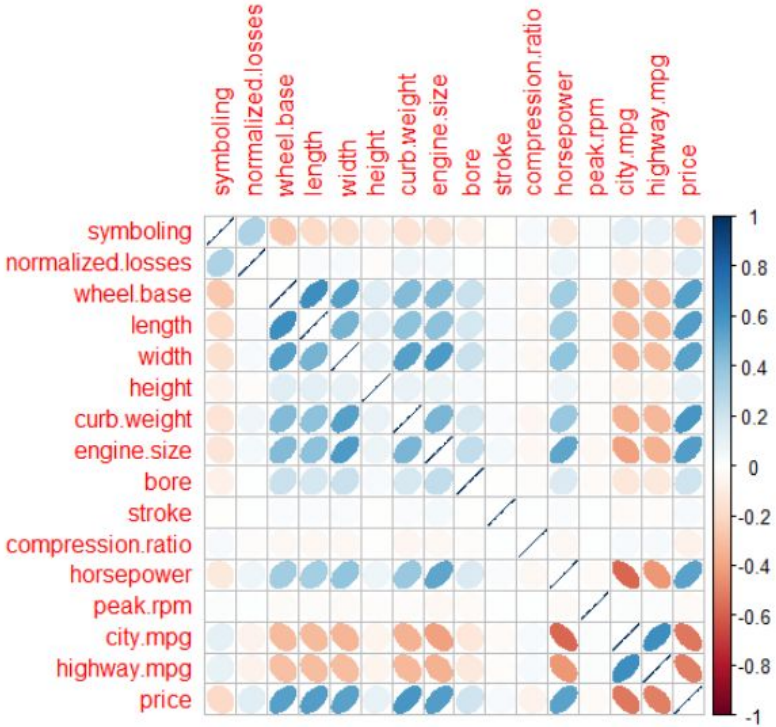
Otras variables







Correlación



Predicción

Modelos utilizados

- Redes Neuronales (cambiando cantidad de neuronas en la capa oculta)
- Support Vector Machine (cambiando parámetros kernel, gamma, degree y cost)
- Random Forest (cambiando parámetros ntree y mtry)

Otros algoritmos utilizados

- Kmeans
- Rmse (librería hydroGOF)
- Librería UnivariateML