

# MADFOOD

R E S T A U R A N T S

G U I D E



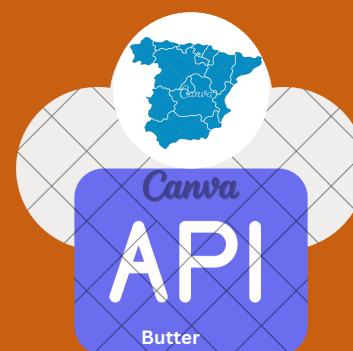
# ABOUT US

Desarrollar un modelo de Machine Learning que prediga el rating esperado de un restaurante nuevo en Madrid, considerando características como ubicación, tipo de cocina y rango de precios.

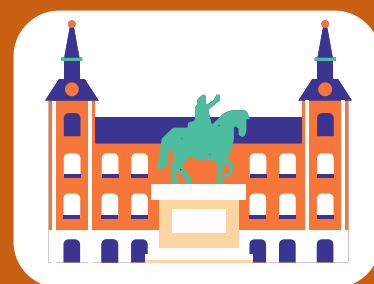
# FUENTES



Google  
Maps



Open Street  
Maps



Panel de indicadores  
de distritos y barrios  
de Madrid



Información geoespacial  
de barrios municipales  
de Madrid

## VARIABLES UTILIZADAS:


- CATEGORÍA DE COCINA (E.G., ITALIANA, JAPONESA)
- PRECIO
- UBICACIÓN GEOGRÁFICA (LATITUD Y LONGITUD)
- DENSIDAD DE RESTAURANTES EN LA ZONA
- INDICADORES SOCIODEMOGRÁFICOS DEL BARRIO (E.G., RENTA MEDIA, EDAD PROMEDIO)



# LIMPIEZA

- Nulls (rating, Precio, otros, horas, otras).
- Sitios que no eran restaurantes ('beauty\_salon', 'clothing\_store', 'doctor', 'shopping\_mall', 'physiotherapist')
- kpi anteriores.

# TRANSFORMACION

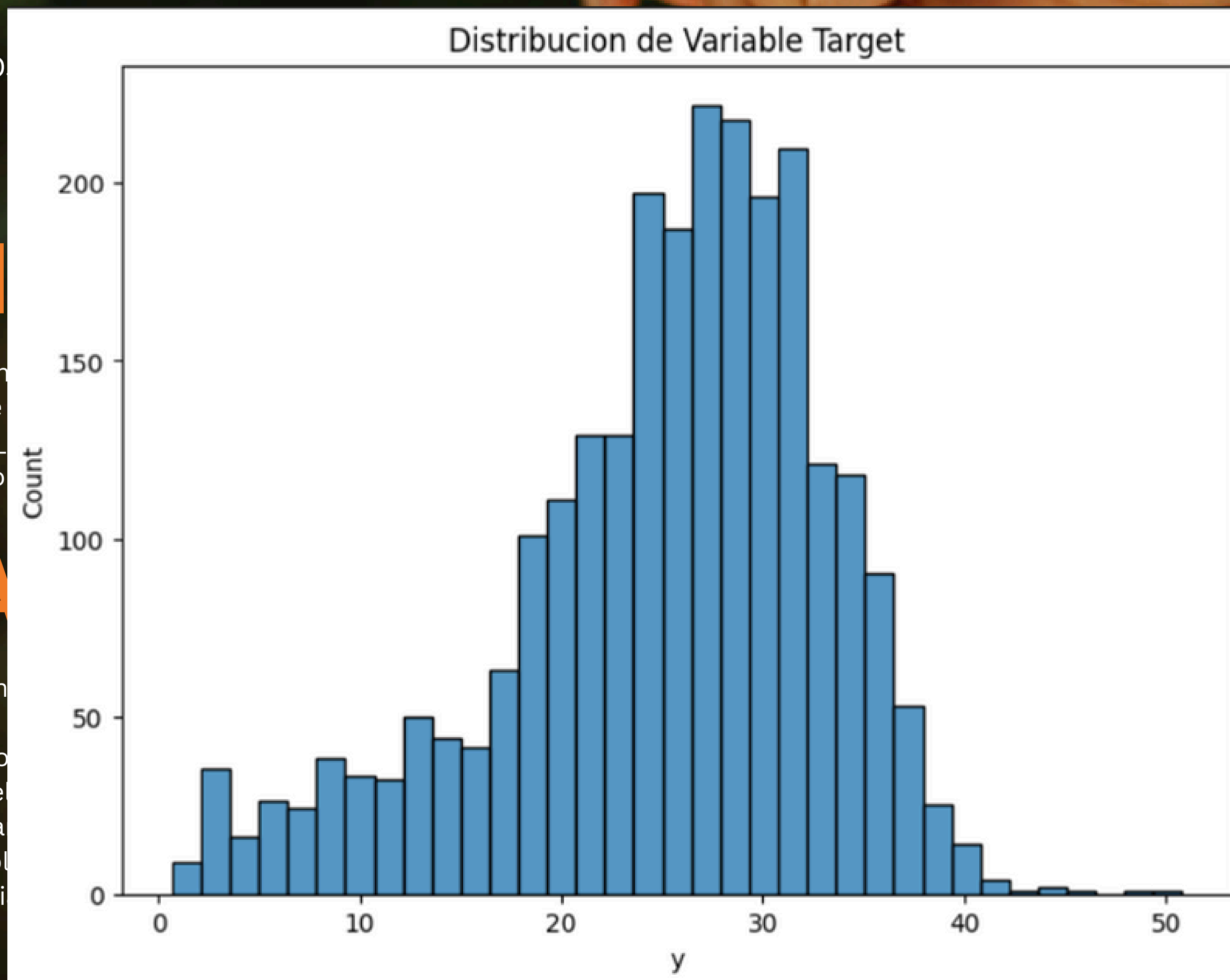
- Onehot Encoder (Tipo de cocina, Nombre distrito, Nombre de barrio).
  - Label Encoder (tipo Cocina).
  - Multi Label Binarize (types)
  - Log variables .
  - Nueva variable Y.
- 

# LIM

- Nulls (rating)
- Sitios que
- 'shopping\_
- kpi anterior

# TRA

- Onehot En
- de barrio).
- Label Enco
- Multi Label
- Modelo pa
- Log variabl
- Nueva vari





# FEATURE ENGINEERING

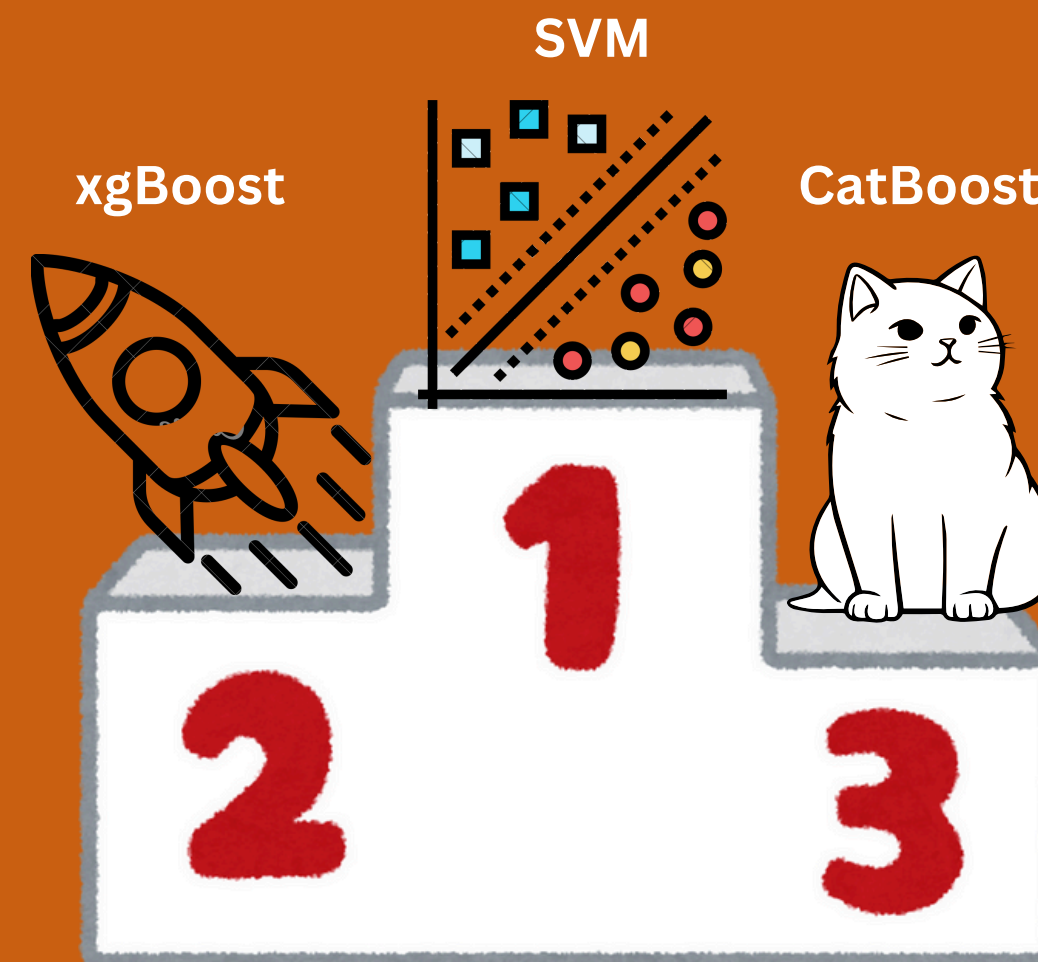
- Cálculo de la densidad de restaurantes en un radio determinado.
- Creación de variables que representan el tipo de comida que tiene el restaurante con Multinomial Bayesian.
- Análisis de correlación entre variables para identificar multicolinealidad y reducir redundancias.



# ENFOQUE EN LA METODOLOGÍA

Modelos evaluados:

- Regresión lineal
- Ridge
- Lasso
- ElasticNet
- Random Forest
- XGBoost
- SVM.
- Catboost.
- Gradient Boosting
- Red Neuronal



Delicious Grilled Chicken



Sweet Prawn Sauce



Mix Platter Sushi



# PIPELINE



## ESCALADO



Trabajamos con metodos de escaldo:

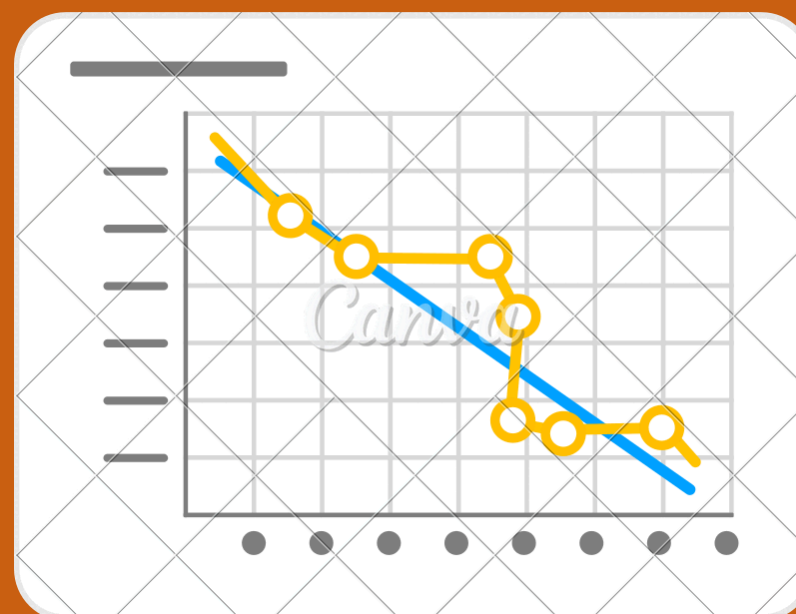
- PolynomialFeatures
- MinMaxScaler
- StandardScaler

## FEATURE SELECTION



Para seleccionar variables:

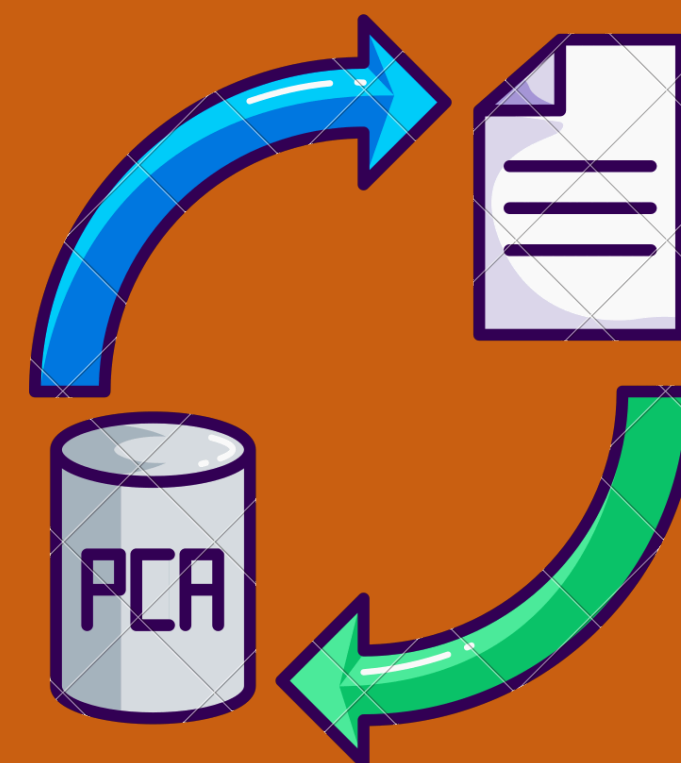
- SelectKBest.
- RFE



## NO SUPERVISADO



- PCA.





# RESULTADOS

Modelo	MAE	MAPE	MSE	R2
Ridge	5.0601	0.3520	41.2772	0.3150
RandomForest	5.0509	0.3465	40.9684	0.3201
GradientBoosting	5.0836	0.3542	42.5015	0.2947
XBoost	5.2837	0.3487	44.7541	0.2573
<b>SVM</b>	<b>4.6534</b>	<b>0.3266</b>	<b>36.9569</b>	<b>0.3867</b>
CatBoost	4.9561	0.3348	40.9501	0.3205




# MEJORAS

## Limitaciones:

- Posible sesgo en los datos debido a la fuente principal (Google Maps), que puede no representar todos los restaurantes existentes.
- La calidad del rating puede estar influenciada por factores subjetivos no capturados en las variables utilizadas.
- El modelo no considera temporalidad; las preferencias de los consumidores pueden cambiar con el tiempo.

## Posibles mejoras:

- Incorporar datos de reseñas de usuarios para capturar aspectos cualitativos.
  - Actualizar periódicamente el modelo para adaptarse a cambios en las tendencias gastronómicas.
  - Explorar más modelos de aprendizaje profundo para capturar relaciones más complejas entre variables.
- 

# THANK YOU.

