



Mango Datathon – Demo Stacking LightGBM + CatBoost

Aquesta demo mostra el **model final** que feu servir:

- LightGBM per a **features numèriques**
- CatBoost per a **features categòriques**
- Un **meta-model LightGBM** que combina les dues prediccions

També fem un **train/validation intern** per veure mètriques reals.

› Paràmetres (fixats al codi)

1 Dades

Train (95339, 31)

Test (2250, 33)

	ID	id_season	aggregated_family	family		ID	id_season	aggregated_family	family
0	1	86	Dresses and jumpsuits	Dresses	0	90	90	Skirts and shorts	Shorts
1	1	86	Dresses and jumpsuits	Dresses	1	16	90	Jackets and Blazers	Jacket
2	1	86	Dresses and jumpsuits	Dresses	2	65	90	Sweaters and Cardigans	Cardig
3	1	86	Dresses and jumpsuits	Dresses	3	138	90	Dresses and jumpsuits	Dresse
4	1	86	Dresses and jumpsuits	Dresses	4	166	90	Swim and intimate	Swimw

2 Mètriques de regressió (validació interna)

MAE valid

5088.9337

RMSE valid

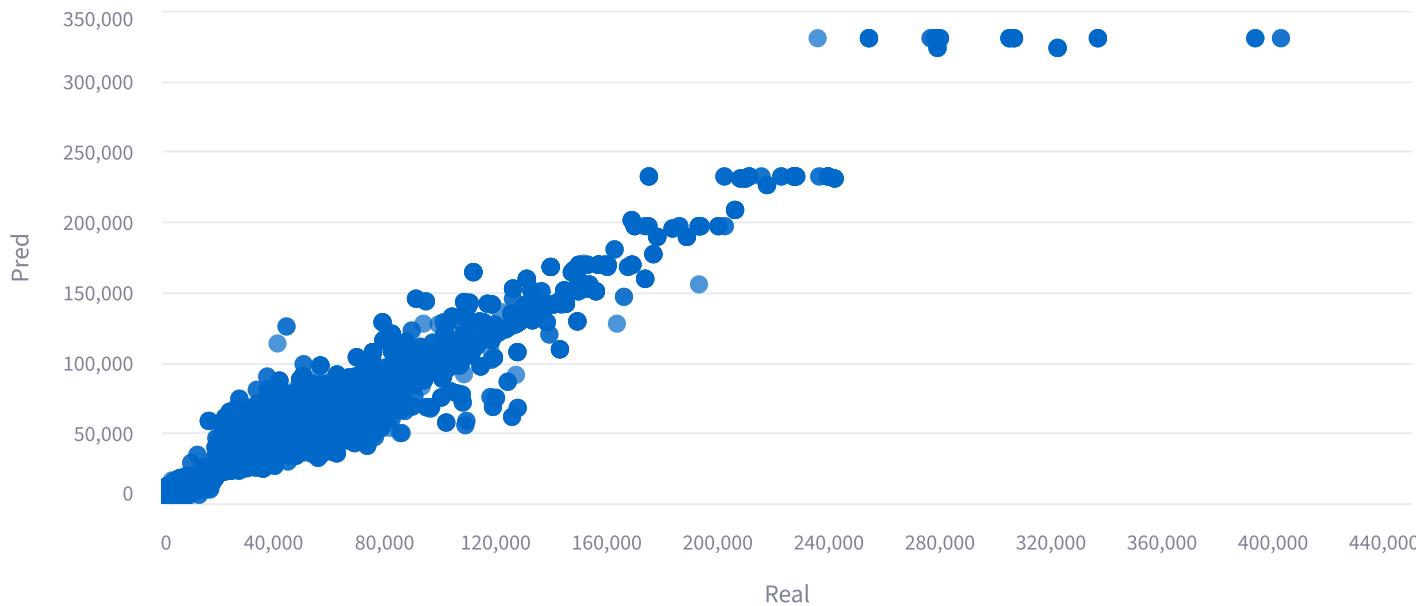
8781.8789

R² valid

0.9343

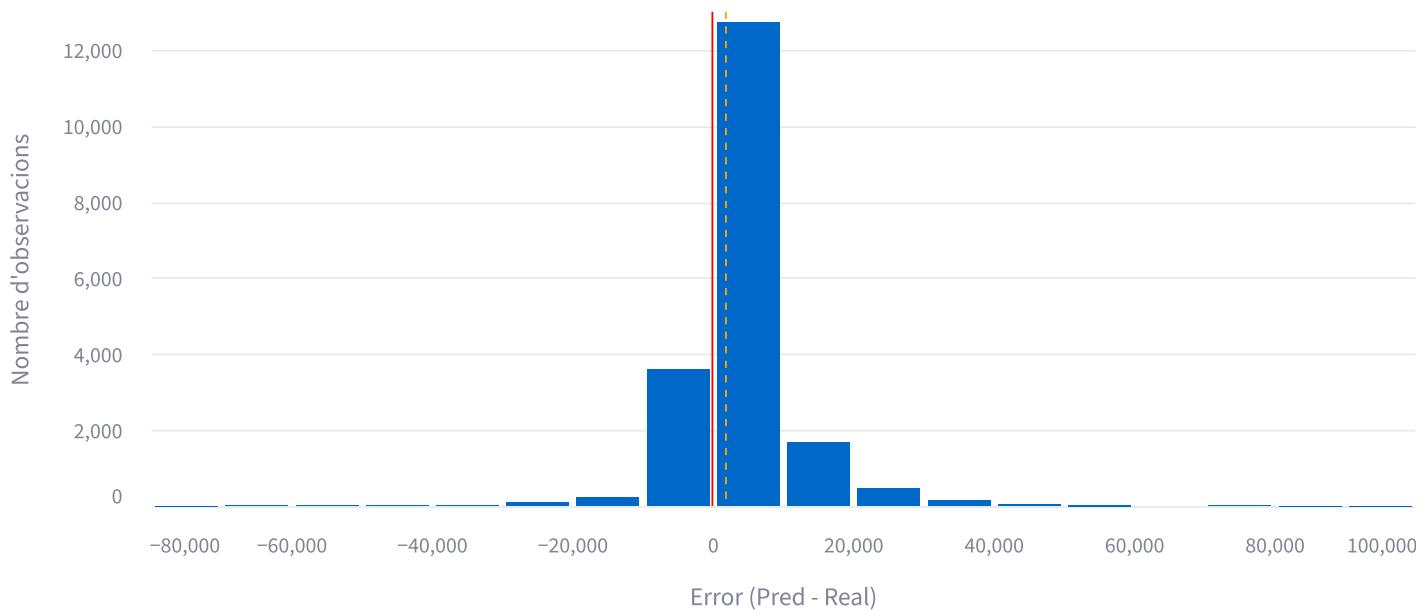
Mètriques del meta-model LightGBM sobre el conjunt de validació interna.

4 3 Gràfica Real vs Predicció (validació interna)



Idealment els punts haurien de caure a prop de la diagonal Real=Pred.

Distribució dels errors (Pred - Real)



La gràfica mostra la distribució dels errors Pred - Real. La línia vermella és l'error 0 (predicció perfecta) i la línia taronja és la mediana de l'error: si està per sobre de 0, el model tendeix a sobreestimar; si està per sota, a infraestimar.

5 Explorador d'un producte (conjunt de validació)

Selecciona un ID:

11628

ID: 11628

Production real (train): 32252.0000

Predictió meta-model: 32156.4977

Error (pred - real): -95.5023

Features originals d'aquest producte:

```
▼ {  
    "ID" : 11628  
    "id_season" : 87  
    "aggregated_family" : "T-shirts"  
    "family" : "T-shirt"  
    "category" : "Tops"  
    "fabric" : "CIRCULAR"  
    "color_name" : "FUCSIA"  
    "color_rgb" : "229,53,129"  
    "image_embedding" :  
        "0.060357273,-0.088701785,0.9783404,-0.77027303,0.1638521,-0.0738851,-0.5627994,-0.1517.  
        4,0.0029378831,0.23160104,-0.47735882,0.21234353,-0.045946345,-0.25345677,-0.23606795,-1  
    "length_type" : "Standard"  
    "silhouette_type" : "Regular"  
    "waist_type" : NaN  
    "neck_lapel_type" : "V-Neck"  
    "sleeve_length_type" : "Short"  
    "heel_shape_type" : NaN  
    "toecap_type" : NaN  
    "woven_structure" : "Light"  
    "knit_structure" : NaN  
    "print_type" : "Sin Estampado"  
    "archetype" : "Authentic"  
    "moment" : "TIME OFF"  
    "life_cycle_length" : 10  
    "num_stores" : 537  
    "num_sizes" : 7  
    "has_plus_sizes" : false  
    "price" : 19.99  
    "year" : 2023  
    "num_week_iso" : 22  
    "Production" : 32252  
    "phase_in_days" : 217  
    "phase_out_days" : 238  
}
```

