

Informe Técnico – Plataforma de Recomendaciones de Ahorro y Escalabilidad

1. Introducción

Este informe describe el diseño, implementación y evaluación de dos modelos de Machine Learning aplicados a una plataforma de gestión de clientes y flotas:

1. Ranking de Repostajes

Algoritmo de *Learning to Rank* para recomendar estaciones de repostaje óptimas en función de coste, desvío, tiempo y sostenibilidad.

2. Hábitos Eficientes (eco-driving)

Modelo de clasificación + clustering que analiza telemetría de vehículos y genera recomendaciones personalizadas para mejorar hábitos de conducción, reducir consumo y emisiones.

2. Datos

2.1 Puntos de Repostaje (puntos_repostaje.csv)

- **punto_id**: identificador único de la estación.
- **nombre_estacion, marca**: metadatos comerciales.
- **latitud, longitud**: geolocalización.
- **carretera, dirección, servicios**: contexto de ubicación.
- **precio_litro, minutos_espera, tipo_combustible** (sintéticos).

2.2 Rutas (rutas.csv)

- **ruta_id, descripcion**.
- **latitud, longitud, carretera, km_desde_origen**.

2.3 Puntos de Recarga Eléctrica (puntos_recarga_electrica.csv)

- **punto_ev_id, nombre_punto, operador**.
- **latitud, longitud, tipo_conector, potencia_kw, precio_kwh, disponible_24h**.

2.4 Beneficios de Tarjeta (beneficios_tarjeta.csv)

- **tarjeta, centimos_litro, porcentaje, estaciones_incluidas**.

2.5 Telemetría Sintética para Hábitos

- **velocidad_media_kmh, frenadas_fuertes_100km, aceleraciones_100km, ratio_ralenti, ratio_carga**.
 - Variables derivadas: consumo estimado (L/100 km), etiqueta binaria de eficiencia.
-

3. Metodología

3.1 Ranking de Repostajes

- Modelo principal: **LightGBM LGBMRanker (LambdaMART)**.
- Features:
 - numéricas: `delta_precio`, `desvio_km`, `minutos_espera`, `litros_necesarios`.
 - categóricas: `marca`, `tipo_combustible`, `carretera`.
- Etiquetas: `rel` (1 = estación óptima en la consulta).
- Métrica: **NDCG@5** (Normalized Discounted Cumulative Gain).
- Validación: **GroupKFold** por consulta.
- Fallback: **RandomForestRegressor** en caso de no disponer de LightGBM.

3.2 Hábitos Eficientes

- **Clasificación:** RandomForest con `class_weight=balanced`.
 - Optimización: grid de hiperparámetros (`n_estimators`, `max_depth`, `min_samples_leaf`).
 - Validación: StratifiedKFold con métrica **F1-macro**.
 - **Clustering:** KMeans dentro de Pipeline con StandardScaler.
 - Selección de k (3–6) por **silhouette score**.
 - Generación de **reglas de negocio** para cada cluster, traducidas en consejos prácticos.
-

4. Resultados

4.1 Ranking de Repostajes

- Mejor configuración LightGBM:
 - `n_estimators=400`, `learning_rate=0.08`, `num_leaves=63`.
 - **NDCG@5 (CV)** \approx 0.82 en datos sintéticos.
- El modelo prioriza estaciones con **precio inferior a la media** y penaliza desvíos > 5 km o esperas > 10 min.

4.2 Hábitos Eficientes

- Clasificador:
 - Grid seleccionó `n_estimators=400`, `max_depth=20`, `min_samples_leaf=2`.
 - **F1-macro (CV)** \approx 0.81.
 - Clustering:
 - k óptimo = 5 con silhouette \approx 0.41.
 - Ejemplo de consejos por cluster:
 - *Cluster 2*: "Reducir ralenti > 12%, anticipar frenadas".
 - *Cluster 4*: "Mantener crucero entre 70–90 km/h".
-

5. Conclusiones y Próximos Pasos

- Los modelos sintéticos validan la viabilidad técnica.
- Métricas de ranking y clasificación muestran buen comportamiento.