DATA COLLECTION

ITESM

PREDICTION TASK



DECISIONS



VALUE PROPOSITION

siderúrgicas,

cargas de trabajo

Reducir costos operativos

por consumo de energía y

emisiones de CO2 en plantas

predicciones de consumo para

facilitar la optimización

mediante



La información recopilada proviene de DAEWOO Steel Co. Ltd. en Gwangyang, Corea del Sur. La información sobre el consumo eléctrico se almacena en un sistema en la nube.

DATA SOURCES



Dataset UCI Steel Industry Energy Consumption (35,040 registros, 2018)

Formato: CSV

Obtenido de:

https://archive.ics.uci.edu/datas et/851/steel+industry+energy+co nsumption

Regresión supervisada de series

temporales multivariadas cuyo objetivo es la predicción del consumo de energía mediante la variable objetivo Usage_kWh con horizonte de 1 hora a 24 horas

Mediante la predicción y modificación de cargas de trabajo, se pretende reducir el costo energético necesario



Estrategia de Validación: Train/Validation/Test Split: 60% / 15% / 25% (respetando orden temporal)

Cross-validation: Time Series Split con 5 folds

IMPACT SIMULATION

Backtesting: Rolling window de 7 días para validar estabilidad

Conjunto de Test:

Últimos 2 meses del dataset (8,760 registros)

Incluye diferentes estacionalidades y tipos de carga

Sin data leakage del conjunto de entrenamiento

MAKING PREDICTIONS



- -Predicción de consumo próximas 24 horas
- -Simulación "what-if": cambio de tipo de carga
- -Recomendaciones: sugerencias de optimización

Integración:

Sprint 1-2: Demo local (Docker + FastAPI)

Sprint 3: Aplicación web con interfaz conversacional

Frontend: Streamlit o Gradio

Backend: FastAPI + Ollama (Llama 3.2)

Deployment: Cloud Run (GCP)

BUILDING MODELS



Baseline Clásicos:

- -XGBoost con hyperparameter tuning.
- -Linear Regression Recursos: CPU/GPU, RAM 16GB

Modelos inovadores:

- Chronos-T5-Small (Amazon Science, 250M parámetros)
- TimesFM (Google Research, 200M parámetros).

Recursos: PyTorch, TensorFlow GPU VRAM min 24GB, RAM 64GB

FEATURES



Raw Data:

- ·Potencia reactiva (lagging/leading) - kVarh
- Factor de potencia (lagging/leading) - ratio
- •Emisiones CO2 tCO2
- •Tipo de carga categórica (Light/Medium/Maximum)
- •NSM (Number of Seconds from Midnight) - temporal
- •day_of_week categórica (Lunes-Domingo)

Features Ingenierizadas:

- •is_weekend binaria
- ·hour temporal
- cyclical_hour_sin,
- encoding cyclical_hour_cos cíclico
- •lag_features consumo en t-1, t-2, t-24 (autoregresivos)
- •rolling_mean_24h media móvil

				<pre>•transformations : load_type × hour, power_factor</pre>
	MONITORING Métricas: • Data Drift: Distribuciones de features (KS test, PSI)			
	Target Drift: Distribución de Usage_kWh real			
	Model Performance: RMSE/MAE calculados en ventanas móviles de 7 días			
	•Prediction Drift: Distribución de predicciones del modelo			
	Alertas:			
	•Drift severo en cualquier feature (PSI > 0.2) → Email a equipo MLOps			
	•Degradación de performance (RMSE > 0.25) → Trigger reentrenamiento automático			
	•Anomalías en distribución de target → Investigación manual			



OWNML.CO