



Entregable 1: Modelización de la energía consumida en el equipo de Auxiliares

Ander Aramburu Siso (CAF I+D)



Entregable 1 – Equipo de Auxiliares

- Objetivo: Modelización de la energía consumida en auxiliares en función de los tiempos de funcionamiento de los componentes de HVAC

Entregable 1 – Equipo de Auxiliares

Datos de interés I

- Los equipos consumidores que cuelgan de la rama AC del convertidor de auxiliares son principalmente los compresores de producción de aire y los HVAC tanto de sala como de cabina.
- Un mes de datos (Septiembre-2019) agregado por evento (interestation/stop/maintenance)

Entregable 1 – Equipo de Auxiliares

Datos de interés II

- Dataframe (df)
 - date_time: Fecha en el siguiente formato YYYY-MM-DD HH:MM:SS
 - ut: Unidad de tren
 - eaux_train_ac: energía consumida por la rama de AC [kWh]
 - event: Tipo de evento (interestation/stop/maintenance)
 - event_duration: duración del evento [s]
 - suma_ton_compresores: suma de ton de los compresores del tren [s]
 - suma_hvac_ton_compresores: suma de ton de los compresores de los HVAC sala [s]
 - suma_hvac_ton_heater: suma de ton de los calefactores de los HVAC sala [s]
 - suma_ton_comp_cab: suma de ton de los compresores del HVAC cabina [s]
 - suma_ton_heater_cab: suma de ton de los calefactores del HVAC cabina [s]

Entregable 1 – Equipo de Auxiliares

Datos de interés II

- **Modelo matemático**

$$y \sim \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_n \cdot x_n$$

Siendo,

- y : variable independiente o respuesta que se va a modelar
- $x_{1..n}$: variables independientes o explicativas
- $\beta_{1..n}$: coeficientes de la regresión que indican el peso de cada una de las variables independientes
- β_0 : coeficiente de la regresión que introduce un sesgo al sistema. Es una magnitud común a todos los sujetos.

Para el caso concreto de estudio los términos de la ecuación son los siguientes:

- y : energía de auxiliares AC por evento [kWh]
- $x_{1..n}$: tiempos de funcionamiento de los distintos componentes de los equipos de auxiliares [s]
- $\beta_{1..n}$: Potencias medias estimadas por el modelo

Entregable 1 – Equipo de Auxiliares

- 1) Cargar los datos “Entregable1.RData” (**0.5 puntos**)
- 2) Filtrar los datos (**0.5 puntos**)
 - Eliminar eventos de más de 600 segundos
 - Eliminar eventos de mantenimiento (“maintenance”)
- 3) Comparativa de consumos (**1 punto**)
 - ¿Existe diferencia significativa en términos de consumo (eaux_train_ac) entre el tipo de evento “interstation” vs “stop”? Justifica tu respuesta.
 - ¿Existe diferencia significativa en términos de potencia (kW) entre el tipo de evento “interstation” vs “stop”? Justifica tu respuesta.
- 4) Generar un **modelo de regresión lineal del consumo (eaux_train_ac) con la primera quincena de datos**. Para ello, modelizar el consumo de auxiliares con los datos filtrados (del punto 3). (**2 puntos**)

NOTA: No considerar en el modelo ni el efecto del tiempo ni la ut

Entregable 1 – Equipo de Auxiliares

5) Revisión del modelo de regresión generado (**3 puntos**)

- ¿El modelo generado es explicativo? Justifica la respuesta
- ¿Tiene sentido trabajar con un modelo reducido o por estratos? Justifica la respuesta
- Enumerar las variables que son estadísticamente significativas
- Explicar el significado de cada coeficiente de regresión
- Estimar el consumo en KW para cada coeficiente regresión
- Dibujar la distribución de los residuos. ¿Se observa alguna anomalía? En el caso de que exista, ¿se puede explicar? Justifica tu respuesta.

6) Validar el modelo de regresión generado en el punto 4) **con los datos de la segunda quincena de septiembre. (1 punto)**

- Usar RMSE como métrica de evaluación. ¿El modelo de regresión generaliza bien? ¿El error obtenido en el entrenamiento y validación son similares?

7) Generar un modelo predictivo de ML más complejo con el subset de entrenamiento y validarla en el subset de validación definidos anteriormente. (**2 puntos**)

- Comparar las nuevas métricas de RMSE en entrenamiento y validación del nuevo modelo generado de ML con el modelo de regresión lineal del punto 4). ¿Se puede mejorar la capacidad predictiva del consumo en comparación con el modelo líneal del punto 4)? ¿Qué ocurre con la explicabilidad de la variables en el nuevo modelo?
- Seleccionar el modelo final a desplegar. Justifica la respuesta.

Entregable: Código R/Python **comentando cada paso (*.R/*.py/*.ipynb/Markdown)**

Deadline: 02/11/2025

A thin, dark grey diagonal line extending from the bottom left towards the top right.

ROLLING STOCK
SIGNALLING
SERVICES
EQUIPMENT & COMPONENTS
TRANSPORT SYSTEMS