

INFORME DE INVESTIGACIÓN

ERRORES DE COMUNICACIÓN
EN DISPOSITIVOS IOT



**CAMPUS
DIGITAL FP**

2024 - 2025

**Curso de especialización
Inteligencia Artificial y Big Data**



Introducción

Investigación I

Investigación II

Conclusiones

¿Qué Spherag?



Hay una forma mejor
de utilizar el agua
+ inteligente



¿Qué es un error?

Un error es una publicación de **FirstConnect**

Imei	TimeStamp	Data	TEnergy	Batt
fcdfe9c...	16276...	Configured	16	3.9
fcdfe9c...	16283...	FirstConnect	1	4.4



Objetivos

1. Confirmar la existencia de una relación entre las características estáticas de los dispositivos y los errores de comunicación
 - H_0 (Hipótesis nula): No existe una relación significativa
 - H_1 (Hipótesis alternativa): Existe una relación significativa
2. Identificar patrones y anomalías a través del análisis temporal de los errores de comunicación

Dataset

1. Unificación de características y registros del dispositivo
2. Columnas de características: Internal Name, DID, LID , Firmware, Latitude, Longitude, TimeZone
3. Columnas personalizadas:
 - a. Age
 - b. NumFistConnect
 - c. CatFirstConnectEqual
 - d. CatFirstConnectExtreme

Exploración de datos

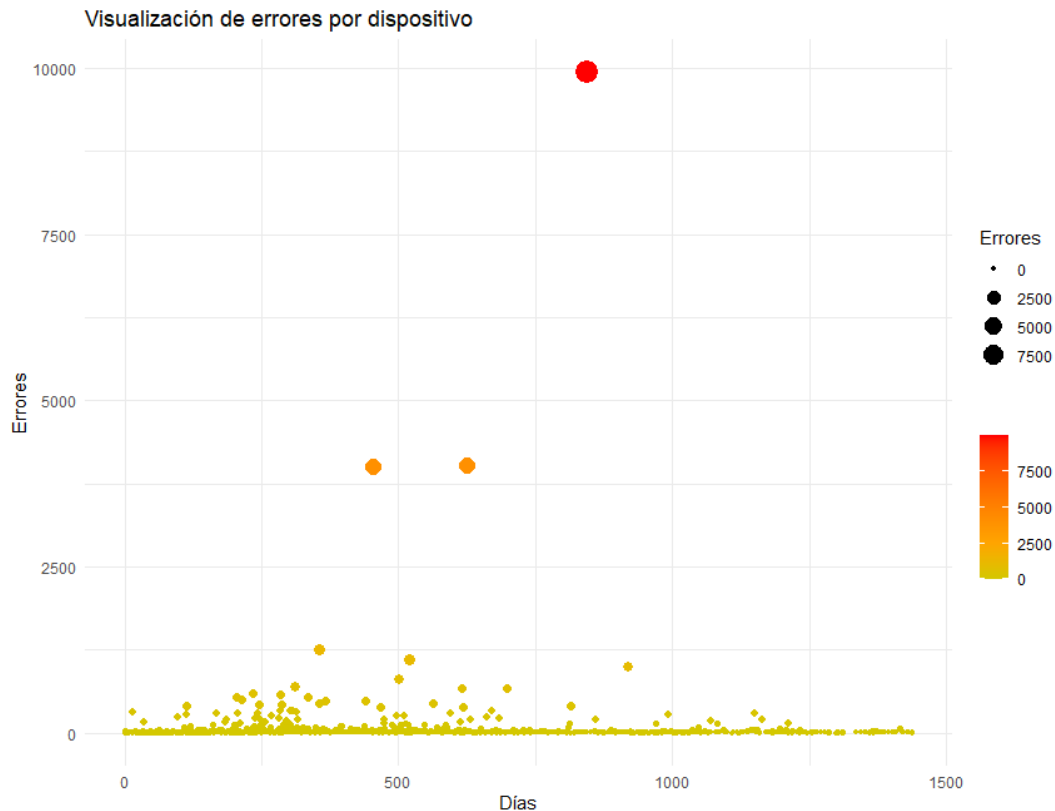
3 dispositivos anómalos

Características en común:

- Modelo: Atlas_2S
- Lote: 7

Test de Welch: p-value
0.07246, insuficiente

Test de Wilcoxon-Mann-Whitney: p-value 1,
insuficiente

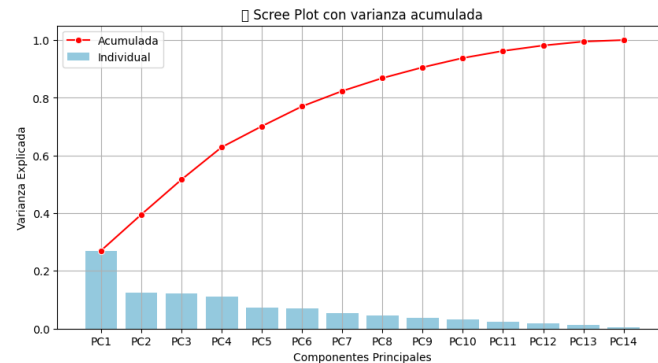
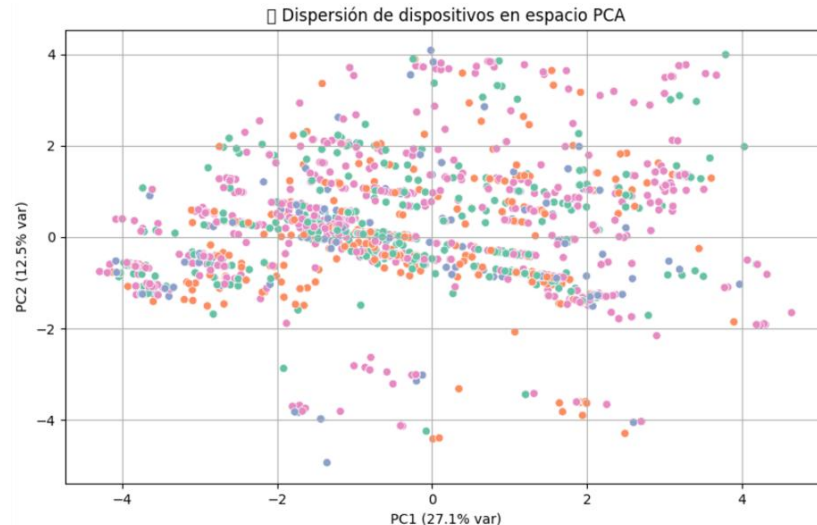


PCA

Se aplicó Análisis de Componentes Principales (PCA) para reducir la dimensionalidad del conjunto de datos

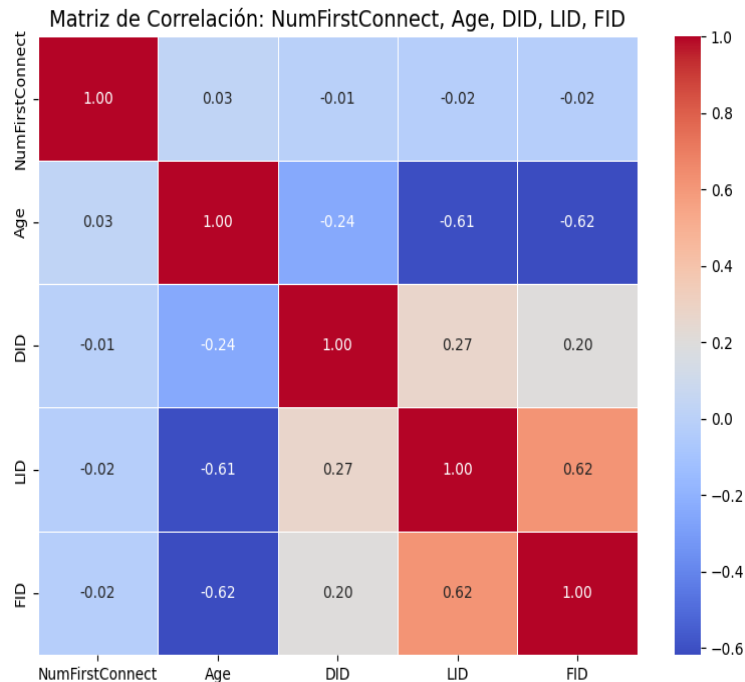
- **No se identifican grupos claramente diferenciados**
- **Varianza acumulada con alta dispersión**

El PCA no aportó evidencia suficiente para aceptar la hipótesis alternativa (H1)

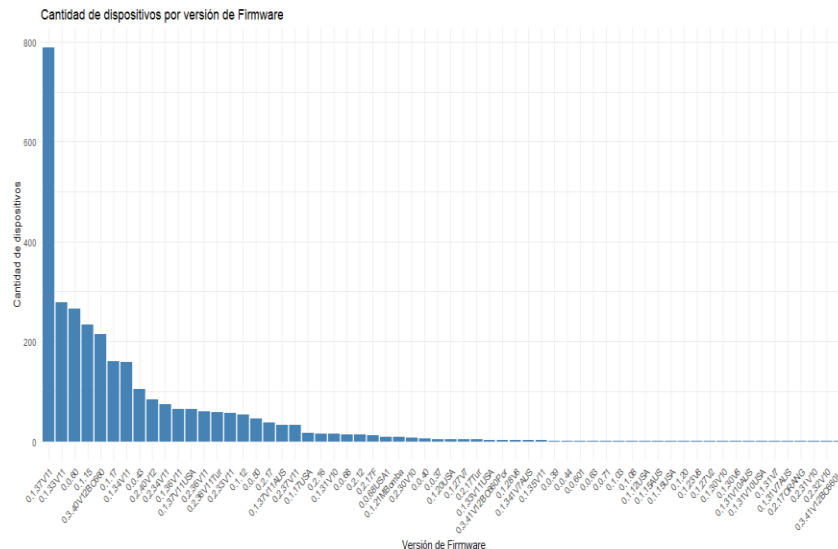


Correlación de variables

Variable	Chi²	Cramér's V
Firmware	Si 5.234e-72	Moderada
TimeZone	Si 4.252e-12	Débil
Modem	Si 1.573e-42	Débil
Country	Si 3.831e-13	Débil
Operator	Si 2.698e-27	Débil
InternalName	Si 4.573e-17	Débil
Longitude	Si 4.402e-11	Nula o muy débil
Latitude	Si 1.311e-04	Nula o muy débil



Ninguna
correlación
significativa

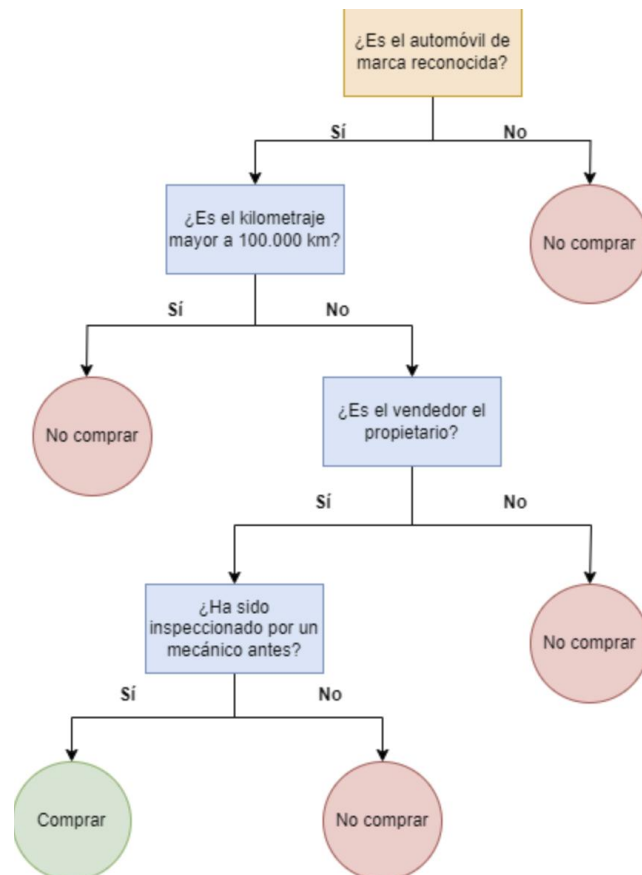


Árboles de decisión

Se entrenaron árboles de decisión para predecir errores de conexión

- El modelo para *CatFirstConnectExtreme* alcanzó un 80% de precisión global, pero falló al identificar correctamente la clase minoritaria
- El modelo para *CatFirstConnectEqual* mejoró el equilibrio, pero redujo la precisión general
- El regresor sobre *NumFirstConnect* mostró errores elevados y baja capacidad explicativa

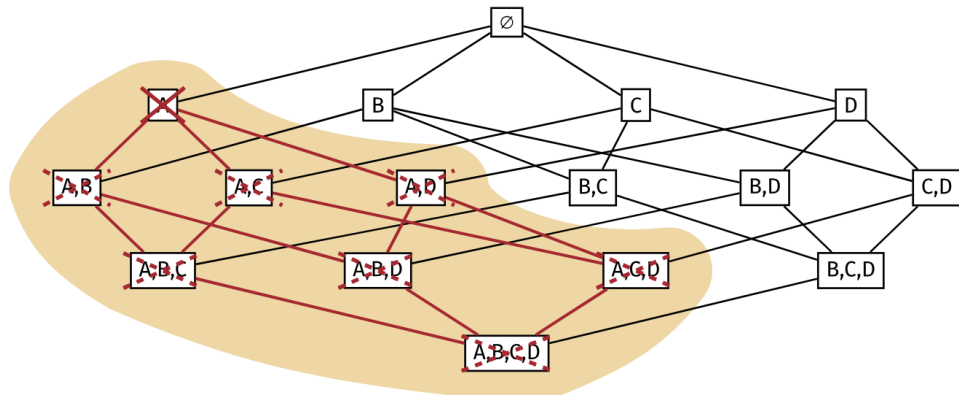
Ningún modelo permite aceptar la hipótesis alternativa H1



Reglas de asociación

Algoritmo **Apriori**, identifica los subconjuntos frecuentes más grandes para generar reglas de asociación y detectar tendencias

Las reglas obtenidas no alcanzaron un índice suficiente de relevancia, con un máximo de 0.65 frente al umbral requerido de 0.8



Reglas con mayor confianza:

1. Microprocesador=default Operator=Vodafone
541 ==> CatFirstConnect=muy bajo 351

conf:(0.65)

No se acepta la hipótesis H1

Dataset

Errores diarios

Nombre de columna	Tipo de dato	Descripción
Fecha	Fecha	Fecha
Errores	Numérico	Cantidad de errores registrada en esa fecha
Dispositivos	Numérico	Núm dispositivos publicaron ese día

45 millones de filas!

Errores mensuales

Nombre de columna	Tipo de dato	Descripción
Mes	Fecha	Fecha
CFC	Numérico	Cantidad de errores registrada ese mes
Dislmei	Numérico	Cantidad de publicaciones realizadas ese mes

Introducción

Investigación I

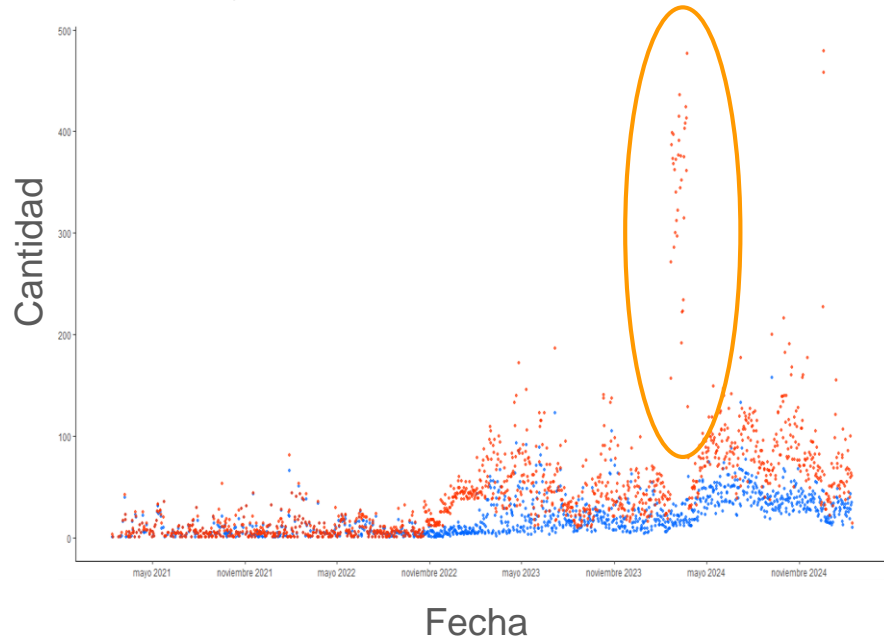
Investigación II

Conclusiones

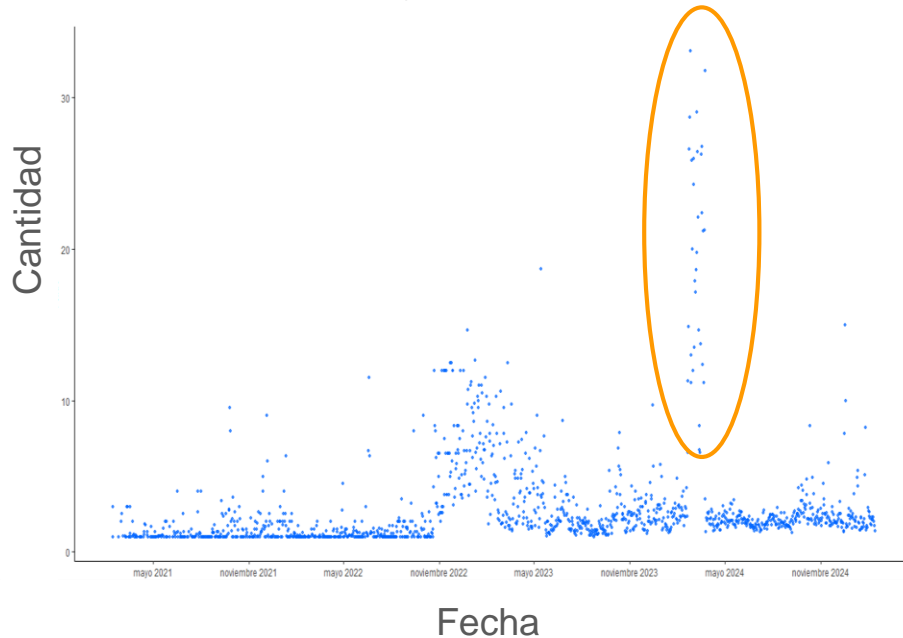
Exploración de datos

Presencia de **anomalías**

Errores y dispositivos

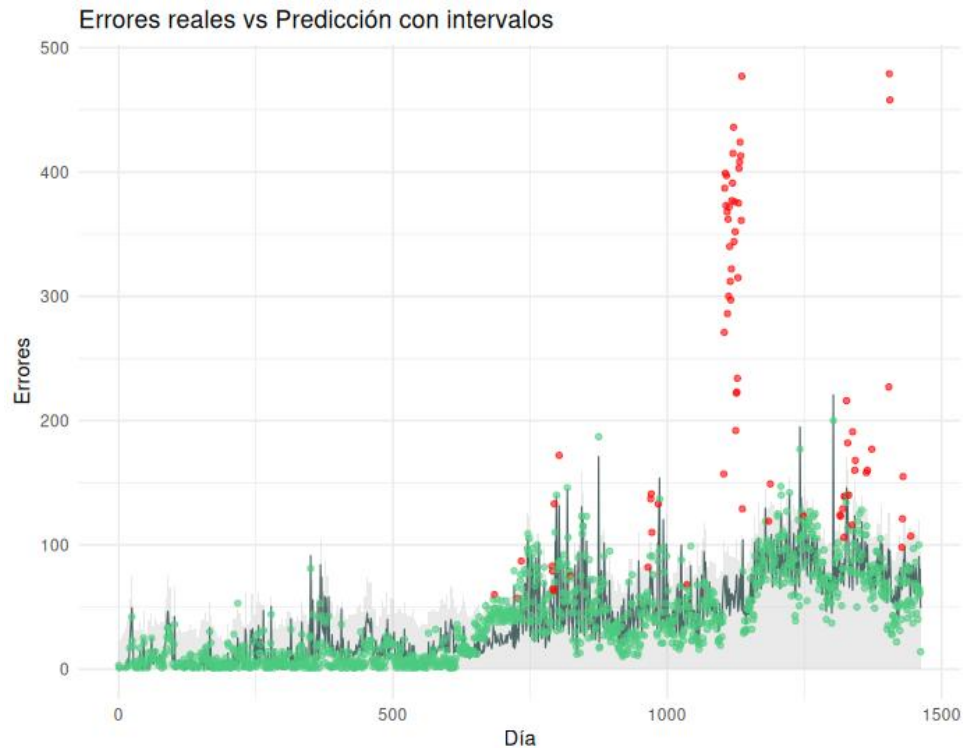
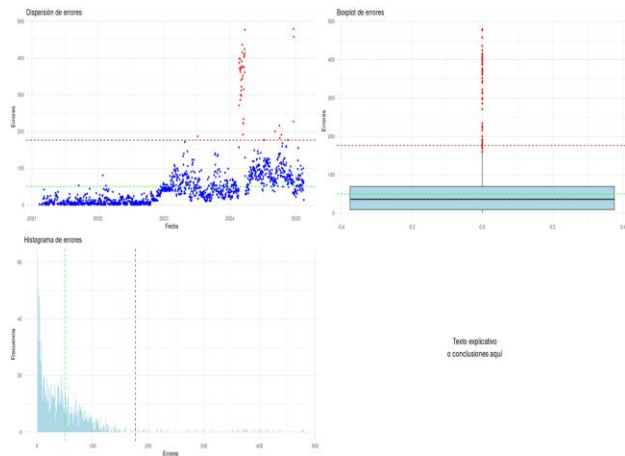


Relación errores y dispositivos



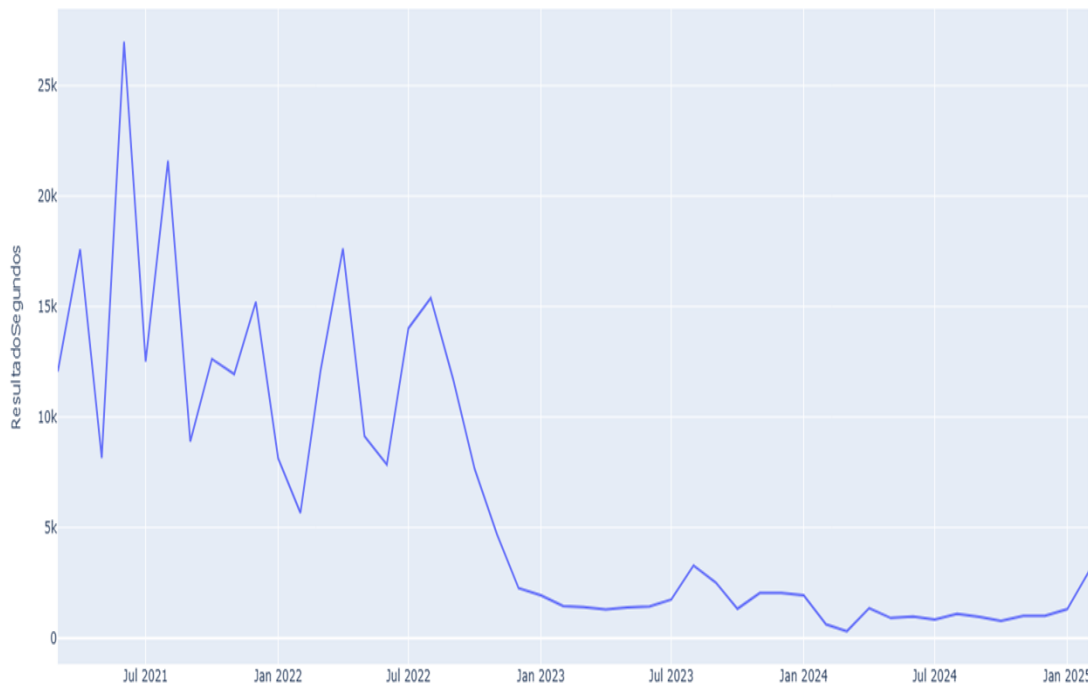
Clasificación de errores diarios

- r-squared 0.83
- p-value < 2.2e-16



Tiempo medio entre errores

Se analizó el MTBF mensual para evaluar la estabilidad.



- Inicialmente alto
- Septiembre de 2022 disminuye hasta 1000 segundos
- En febrero y marzo de 2024 cae más (anómalo)

Conclusiones

- No existe relación significativa entre características del dispositivo y cantidad de errores
- Se pueden identificar anomalías en la comunicación de dispositivos en tiempo real
- Desde finales de 2022 ha aumentado la frecuencia de errores

Líneas de investigación futuras

- Motivos de las anomalías
- Otros modelos de dispositivos
- Modelos de prestaciones