

# COMPORTAMIENTO DE LOS SERVICIOS DE INTERNACIÓN DE LOS HOSPITALES DE LA PROVINCIA DE BUENOS AIRES

**clusterAI**

Order Jonathan  
Universidad Tecnológica Nacional  
FRBA

Order Nicolás  
Universidad Tecnológica Nacional  
FRBA

Descalzo Agustina  
Universidad Tecnológica Nacional  
FRBA

## Introducción

**OBJETIVO:** Analizar el rendimiento de los 2149 centros de salud públicos existentes en las 12 regiones sanitarias de la Provincia de Buenos Aires, para luego poder estimar mediante técnicas de regresión el porcentaje de ocupación futuro de los mismos, siendo Random Forest el más acertado de los modelos estudiados..



**Buenos Aires**  
Provincia

**INPUTS:** De cada hospital se debe conocer: egresos, promedio de camas disponibles, días de pacientes internados, % de ocupación, giro de camas, días de estadía y defunciones.

## Exploratory data analysis

Preprocesamiento



Análisis  
población  
versus  
camas  
disponibles

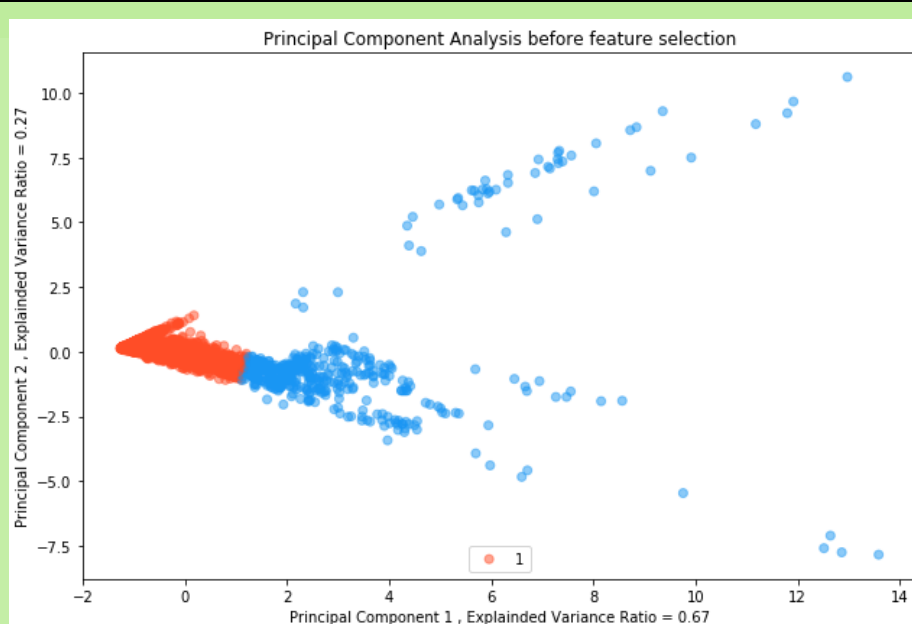
Feature Engineering



Análisis  
correlación  
entra  
variables de  
input del  
modelo

## Método paso a paso

1



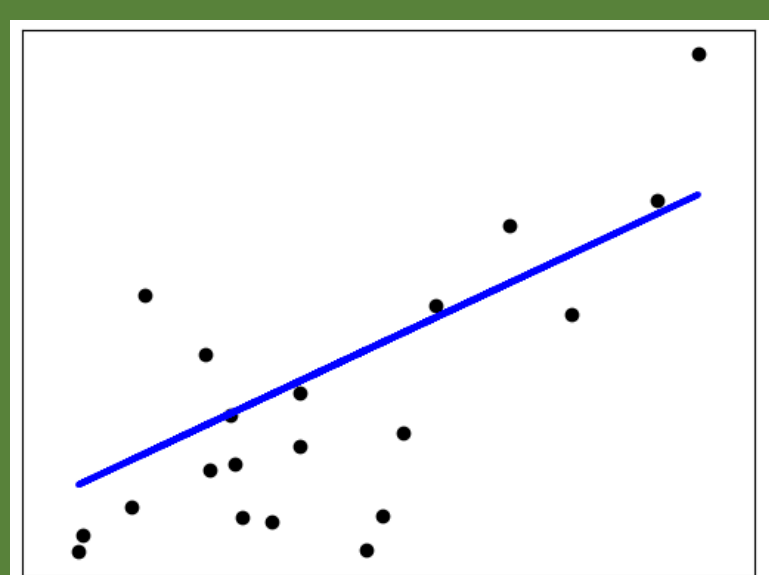
**PCA + K-Means:** Selección de variables más explicativas del modelo + formación de 2 clusters  
2 Variables explican el 94% de la variabilidad

2

**TRAIN/TEST:** 90-10%  
**RANDOM STATE = 10**  
(réplica de pruebas)

3

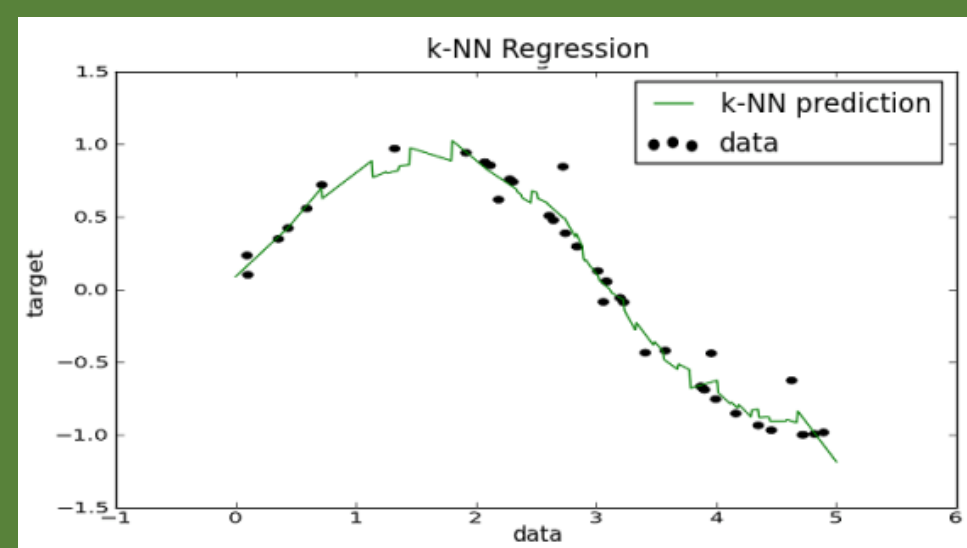
**REGRESIÓN:** Métodos estadísticos para estimar relaciones entre variables



**LINEAL**

Minimiza la suma residual de cuadrados entre las respuestas observadas y las predichas por la aproximación lineal.

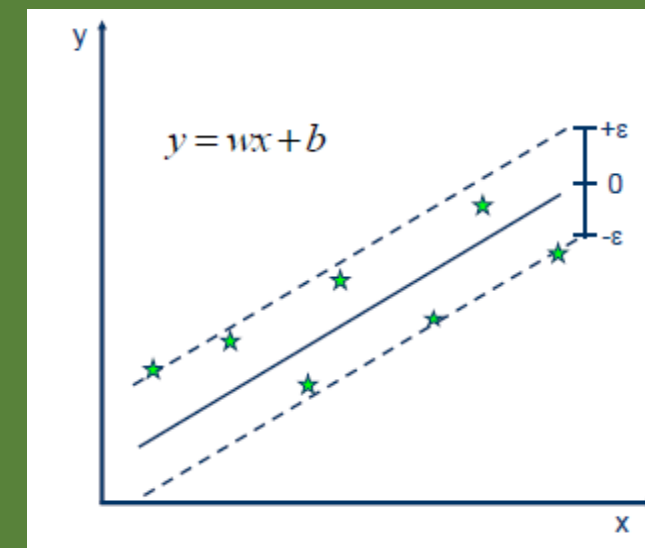
**MAE = 20.97**  
**R² = 12%**



**KNN**

Algoritmo que predice el objetivo numérico basándose en una medida de similitud de los datos más cercanos (por ejemplo, funciones de distancia).

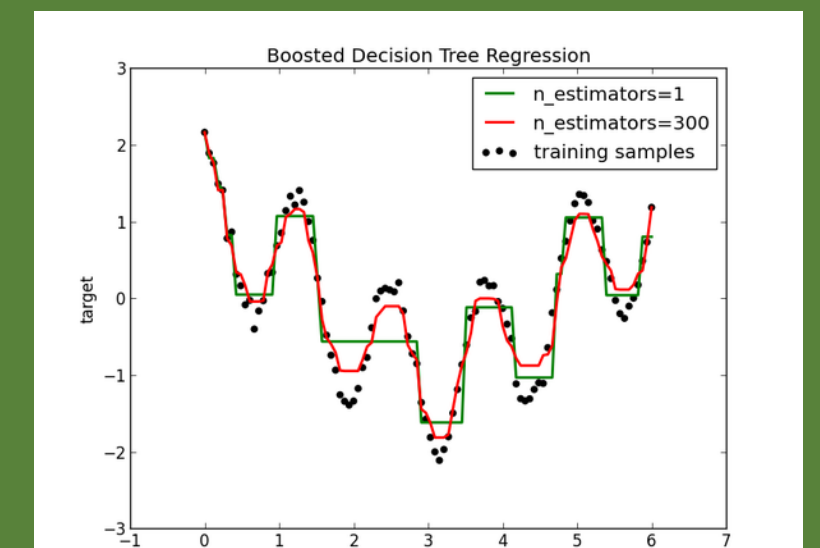
**MAE = 12.8**  
**R² = 67%**



**SVR**

Minimiza el error con un un margen de tolerancia (épsilon), buscando el hiperplano que maximiza el margen.

**Kernel lineal | Kernel RBF**  
**MAE = 20.82    MAE = 13.66**  
**R² = 13%        R² = 62%**



**Random forest**

Estimador que se ajusta a una serie de clasificadores de árboles de decisión y utiliza el promedio para mejorar la precisión predictiva y el over fitting

**MAE = 10.97**  
**R² = 76%**

## CONCLUSIONES

- Random forest** es el modelo más adecuado.
- Los modelos **lineales** no tuvieron buena performance por la naturaleza de los datos.
- Fuerte correlación: días de estadía - camas disponibles, inversamente relacionadas con % de ocupación
- Útil para asignar recursos y planear infraestructura en partidos densos con baja disponibilidad de camas
- Alta correlación entre egresos y defunciones
- La Matanza** es el partido de mayor población con baja disponibilidad de camas, al revés de La Plata/Hurlingham