

MACHINE LEARNING IMMERSION

ANDRÉ OMAR CHÁVEZ PANDURO

« Divide las dificultades que examinas en tantas partes como sea posible , para su mejor solución»



EXPOSITOR

André Omar Chávez Panduro
UNMSM

MSc in Data Science Candidate
Promotion “Erwin Kraenau Espinal”
Universidad Ricard Palma



**Senior Data
Scientist**



**Customer
Intelligence Analyst**



Data Analyst



AGENDA

- Selección de Drivers.
- Selección Paramétrica.
- Selección mediante Machine Learning.

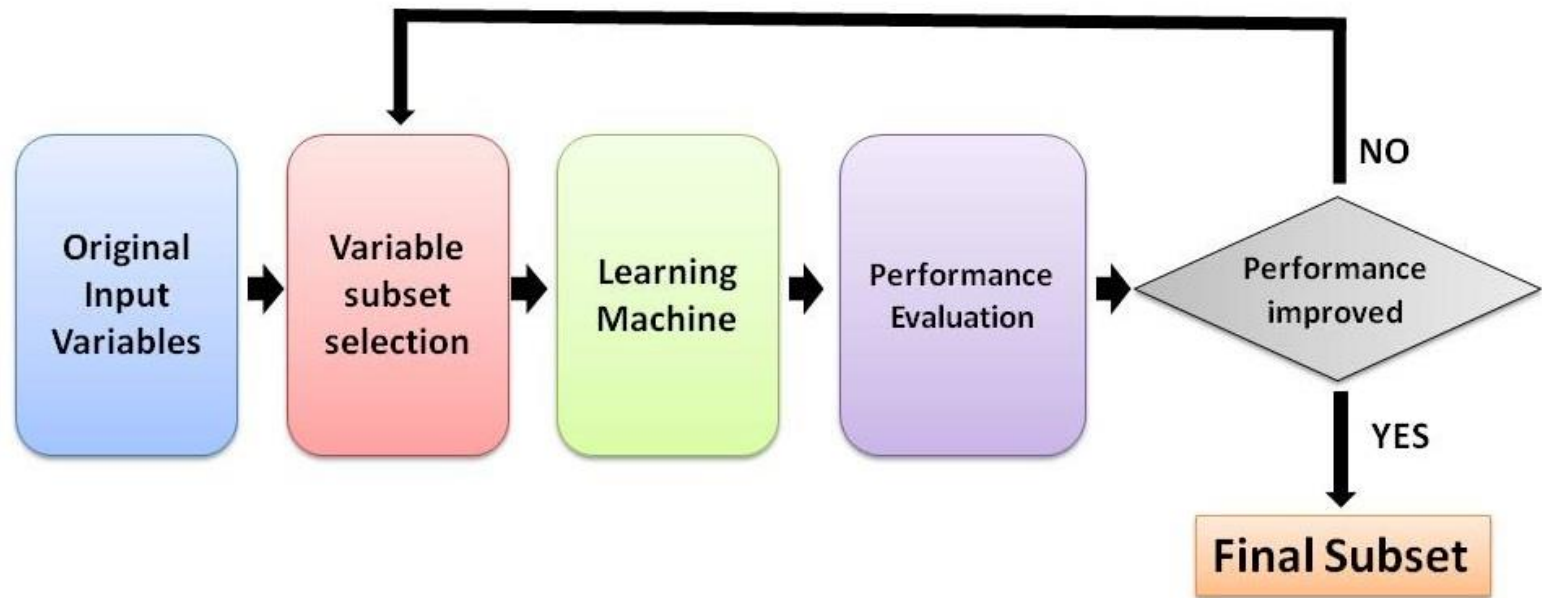


PROBLEMÁTICA : NÚMERO DE VARIABLES

- El número de V.I. o features en cantidades poco manejables es un problema con el que tenemos que lidiar.
- Podemos elegir algún método de reducción de dimensiones (ACP o ACM) o elegir las variables «**más importantes**».



MÉTODOS DE SELECCIÓN DE VARIABLES (FEATURE SELECTION)



MÉTODOS PARAMÉTRICOS DE SELECCIÓN DE VARIABLES



MÉTODOS ITERATIVOS DE SELECCIÓN DE VARIABLES

➤ Método Forward

En cada paso se selecciona la variable que más contribuye a la separación de los grupos. El proceso se detiene si ninguna variable separa los grupos significativamente más de lo que ya estaban.

➤ Método Backward

Se incluyen todas las variables y en cada paso se elimina la que menos contribuye a la separación de los grupos. El proceso se detiene cuando la exclusión de cualquiera de las variables hace disminuir significativamente la separación entre los grupos.

MÉTODOS ITERATIVOS DE SELECCIÓN DE VARIABLES : STEPWISE

Método Stepwise

F mínimo para entrar

F máximo para salir

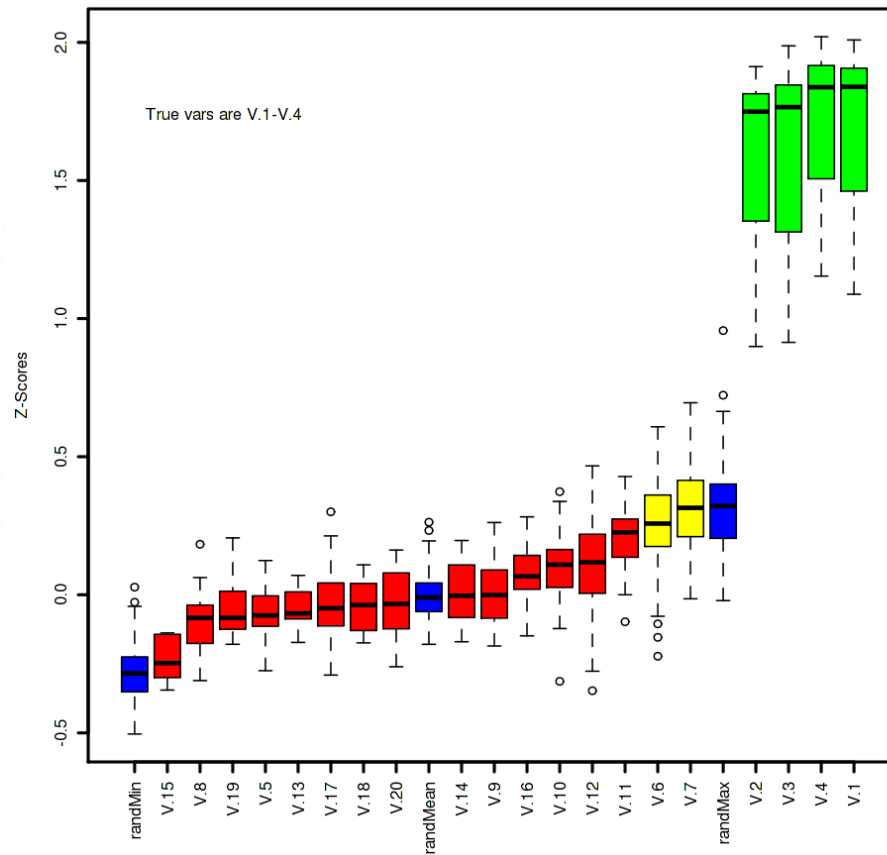
Donde (F entrada > F salida).

- **Nivel de tolerancia:** Medida del grado de asociación lineal entre las variables clasificadoras.

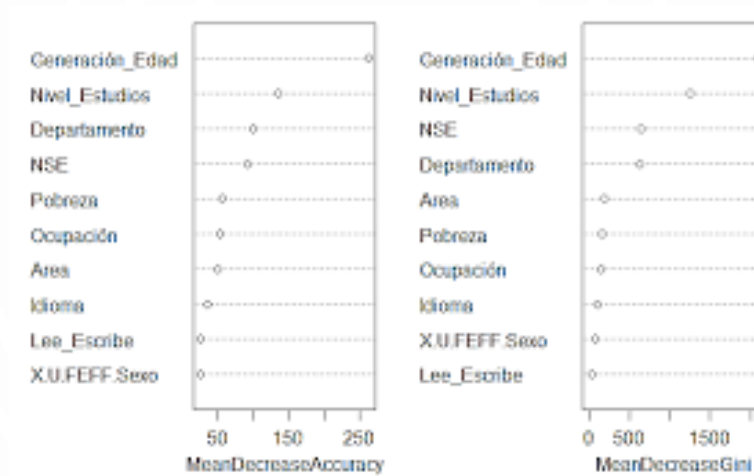
$$1 - r_i^2$$

Si la tolerancia de la variable i es muy pequeña, significa que dicha variable está muy correlacionada con el resto, lo que puede provocar problemas en la estimación. Generalmente, se fija un nivel mínimo de tolerancia.

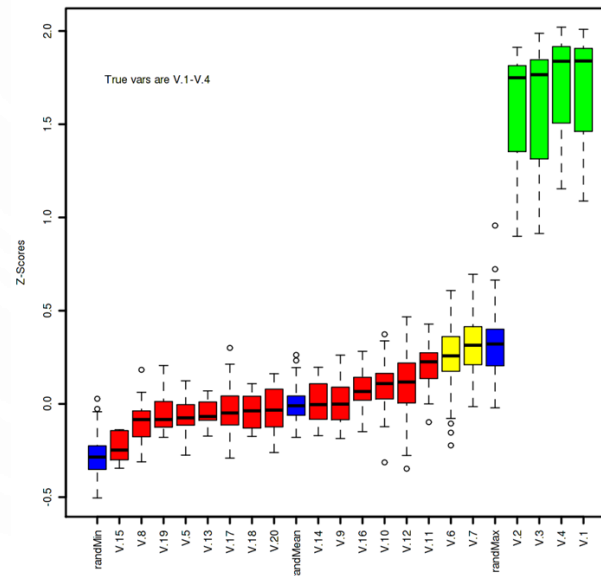
MÉTODOS MACHINE LEARNING DE SELECCIÓN DE VARIABLES



Selección Random Forest



Método Boruta



**GRACIAS
POR SU ATENCIÓN**