

# Inteligencia Artificial

Juan Pablo Restrepo Uribe

Ing. Biomedico - MSc. Automatización y Control Industrial

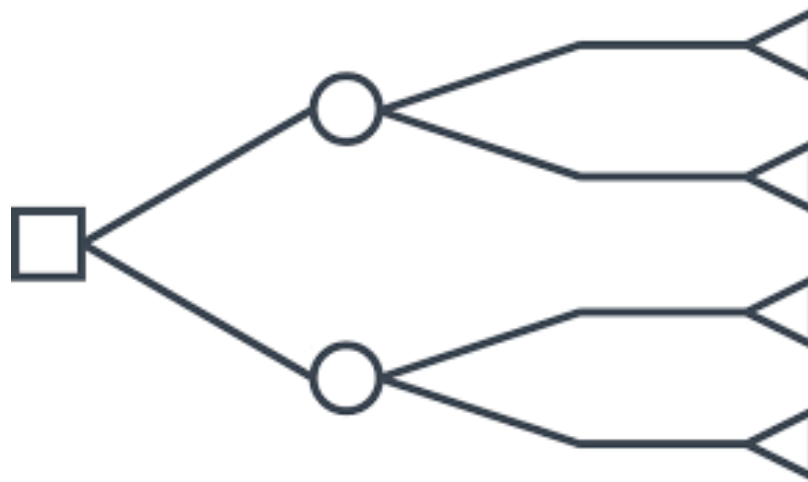
[jprestrepo@correo.iue.edu.co](mailto:jprestrepo@correo.iue.edu.co)

2023

Institución Universitaria de Envigado

## Árbol de decisión

Es un modelo de predicción utilizado en diversos ámbitos que van desde la **inteligencia artificial** hasta la **Economía**. Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.



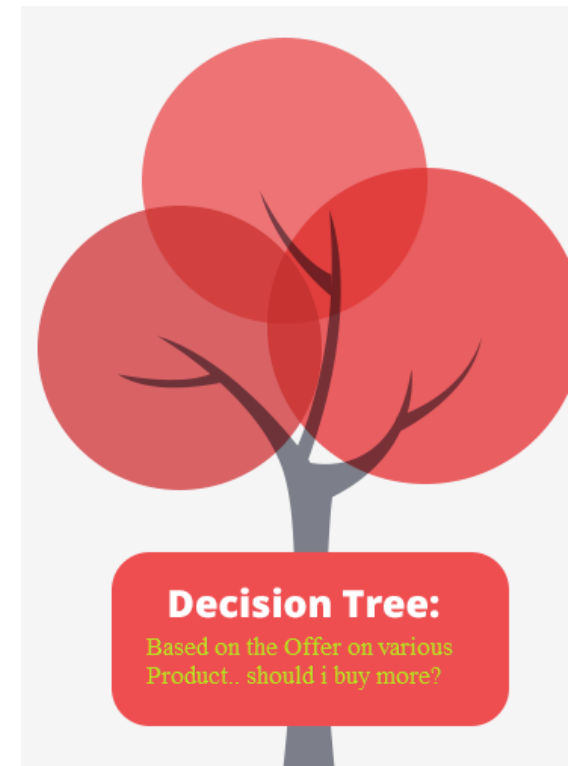
# Árbol de decisión

Son un método de aprendizaje supervisado no paramétrico que se utiliza para la clasificación y la regresión.

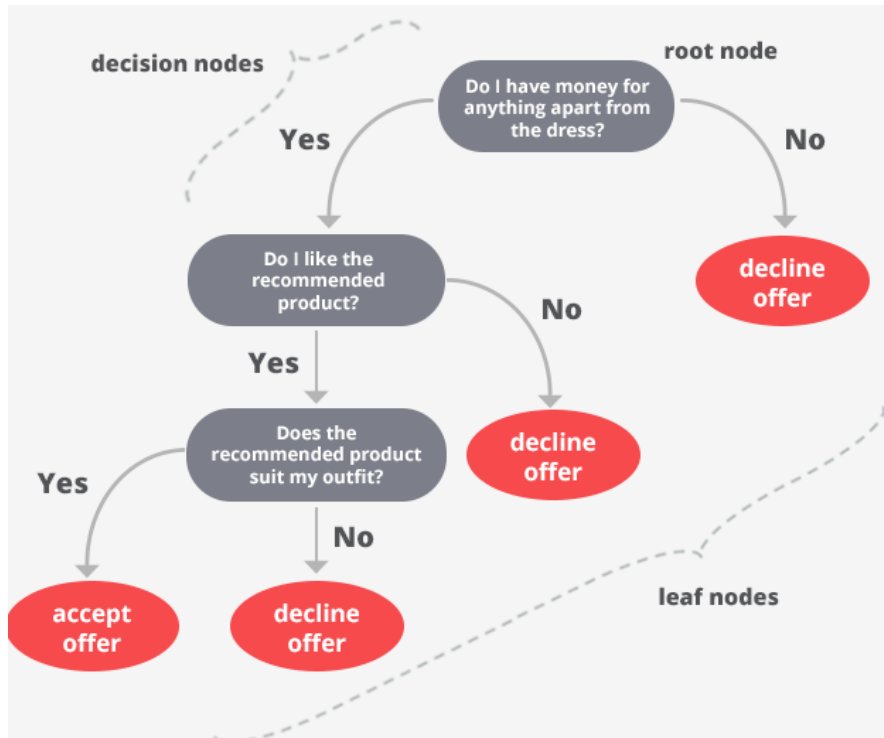
El objetivo es crear un modelo que prediga el valor de una variable de destino mediante el aprendizaje de reglas de decisión simples deducidas de las características de los datos. Un árbol puede verse como una aproximación constante por partes.

## Ventajas de los árboles de decisión

- Fácil de entender y de interpretar.
- Requiere poca preparación de datos.
- Capaz de manejar datos numéricos y categóricos. Sin embargo, la implementación de scikit-learn no admite variables categóricas por ahora.
- Capaz de manejar problemas de múltiples salidas.
- Utiliza un modelo de caja blanca.
- Posibilidad de validar un modelo mediante pruebas estadísticas.



## Desventajas de los árboles de decisión

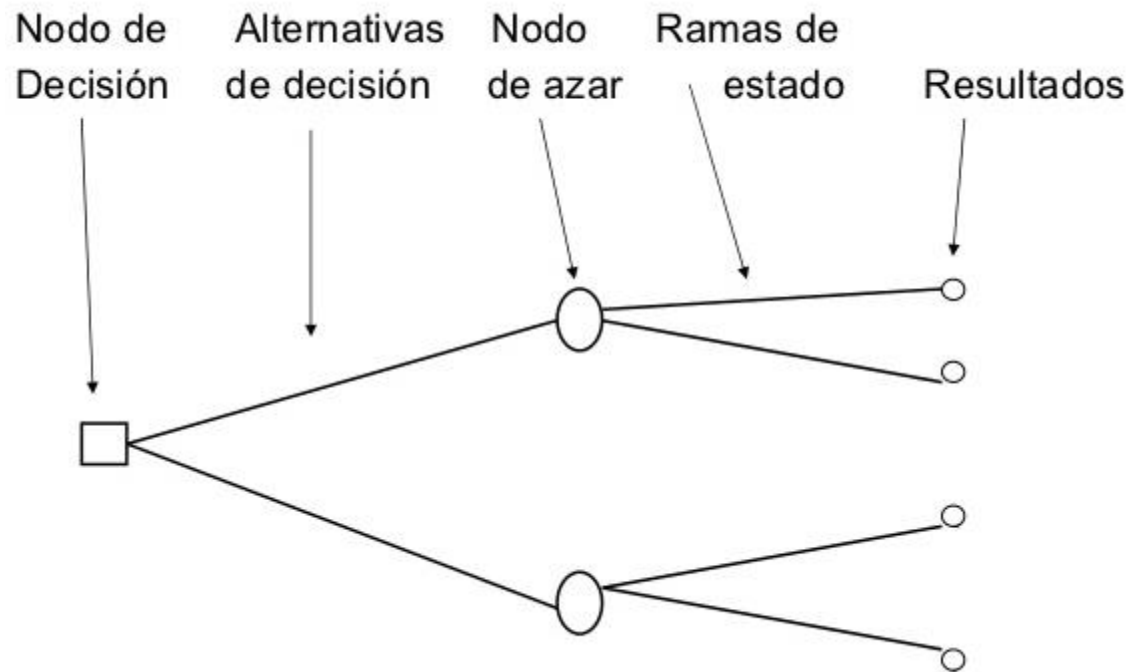


- Los aprendices de árboles de decisión pueden crear árboles demasiado complejos que no generalizan bien los datos (sobreajuste).
- Los árboles de decisión pueden ser inestables.
- Las predicciones de los árboles de decisión no son uniformes ni continuas.
- Los aprendices de árboles de decisión crean árboles sesgados si dominan algunas clases.

## Elementos de los árboles de decisión

- **Nodo:** se puede definir como el momento en el que se ha de tomar una decisión de entre varias posibles, lo que va haciendo que a medida que aumenta el número de nodos aumente el número de posibles finales a los que puede llegar el individuo.
- **Vectores:** de números serían la solución final a la que se llega en función de las diversas posibilidades que se tienen, dan las utilidades en esa solución.
- **Flechas:** son las uniones entre un nodo y otro y representan cada acción distinta.
- **Etiquetas:** se encuentran en cada nodo y cada flecha y dan nombre a cada acción.

## Elementos de los árboles de decisión



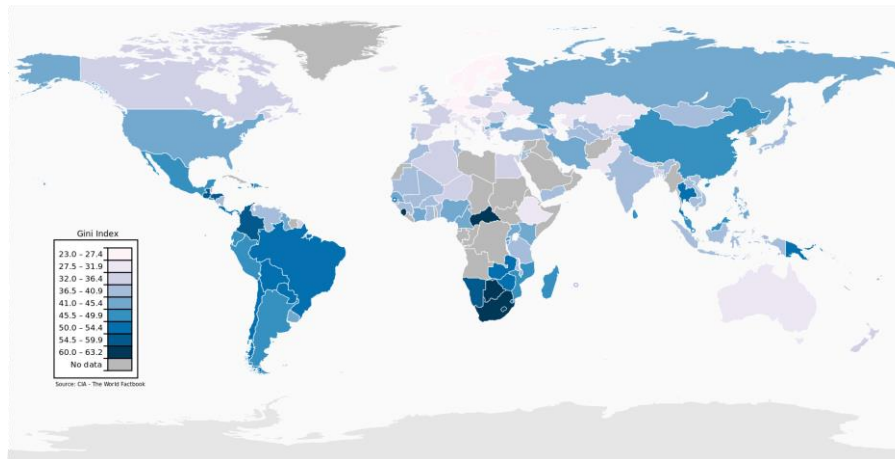
## Reglas de los arboles de decisión

- Al comienzo del juego se da un nodo inicial que no es apuntado por ninguna flecha, es el único del juego con esta característica.
- El resto de los nodos del juego son apuntados por una única flecha.
- De esto se deduce que hay un único camino para llegar del nodo inicial a cada uno de los nodos del juego. No hay varias formas de llegar a la misma solución final, las decisiones son excluyentes.



# Elementos importantes de los arboles de decisión (criterio - medida de impureza)

Función para medir la calidad de una división.



***gini***

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk})$$

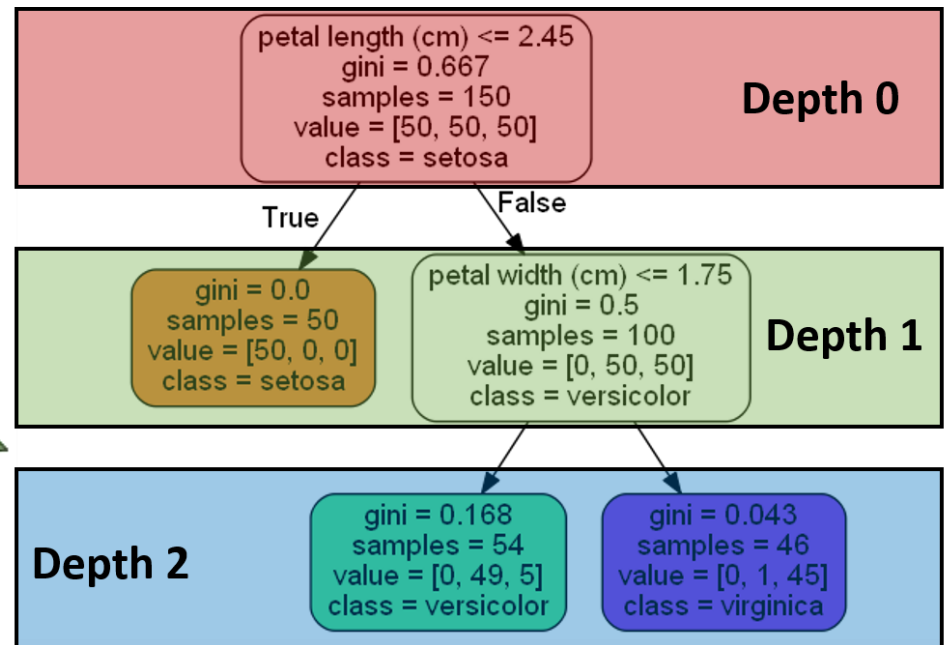
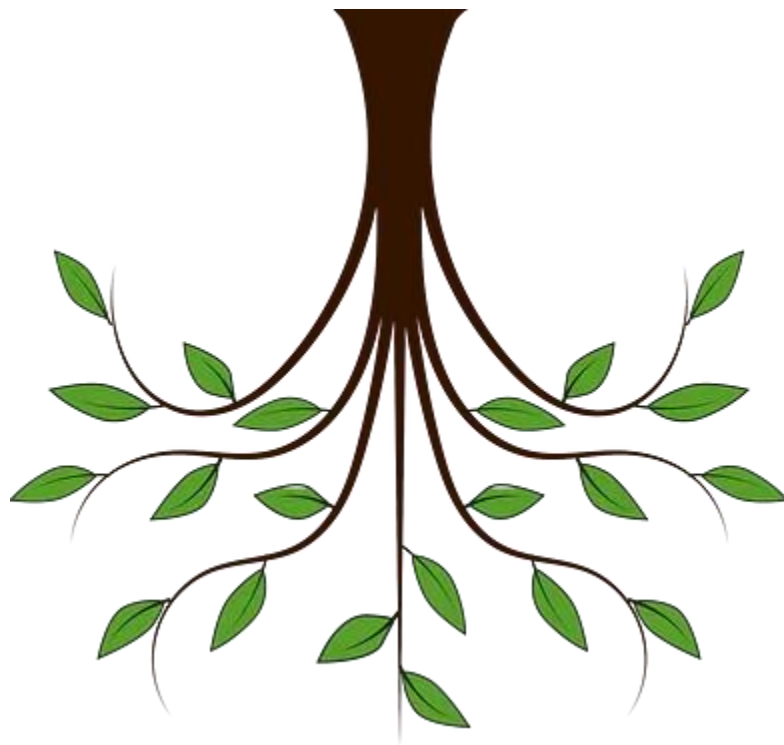
***entropía***

$$LL(D, T) = -\frac{1}{n} \sum_{(x_i, y_i) \in D} \sum_k I(y_i = k) \log(T_k(x_i))$$

***log\_loss***

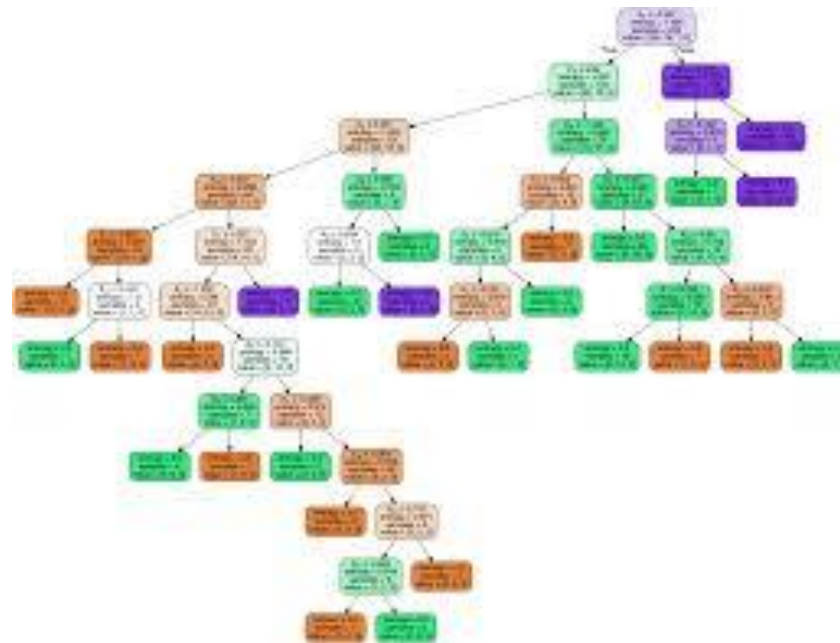
$$H(Q_m) = -\sum_k p_{mk} \log(p_{mk})$$

# Elementos importantes de los arboles de decisión (max\_Depth - profundidad máxima)



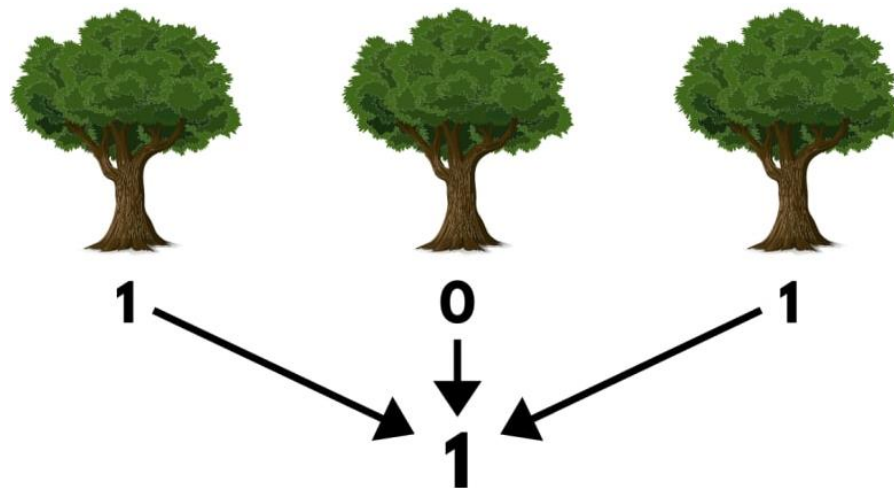
# Elementos importantes de los arboles de decisión (max\_features)

Máximo de características a considerar en cada división

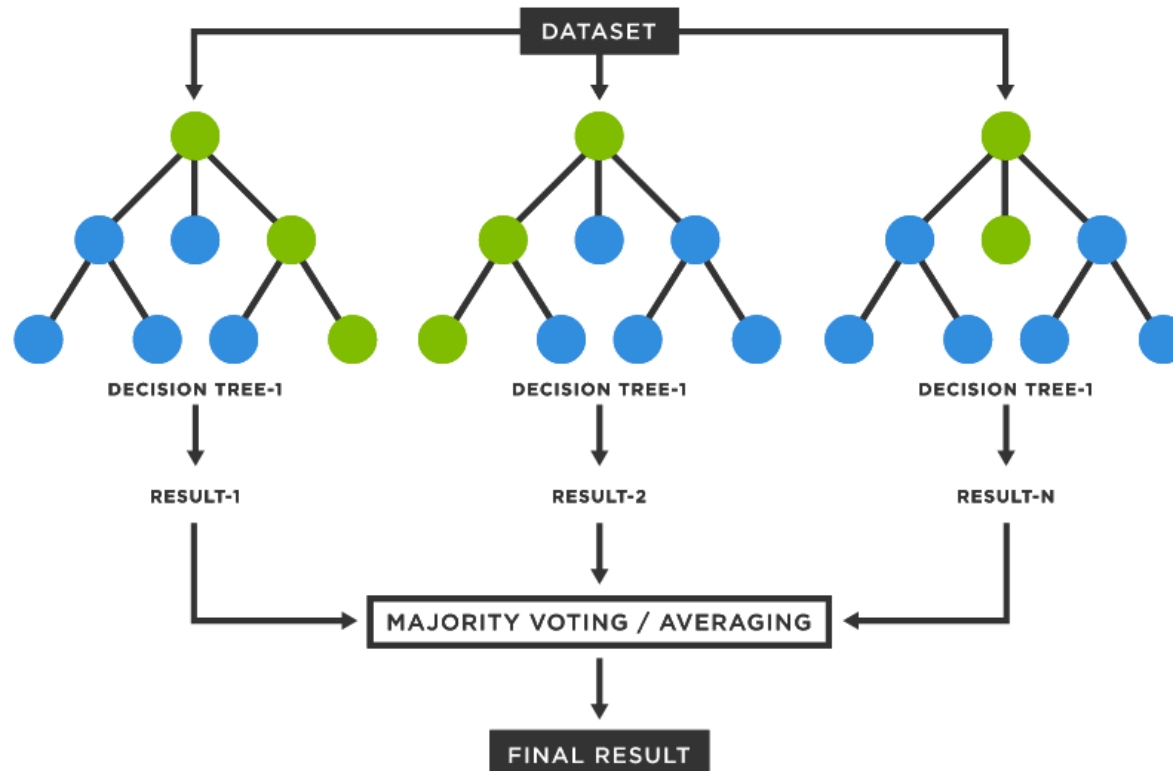


## Random Forest Classifier

Método de aprendizaje conjunto para la clasificación, la regresión y otras tareas que operan mediante la construcción de una multitud de árboles de decisión en el momento del entrenamiento. Para las tareas de clasificación, la salida del bosque aleatorio es la clase seleccionada por la mayoría de los árboles. Para tareas de regresión, se devuelve la predicción media o promedio de los árboles individuales.



# Random Forest Classifier

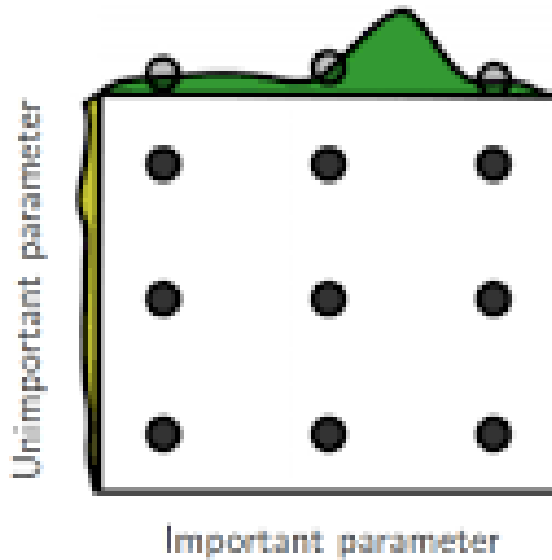


## Random Forest Classifier

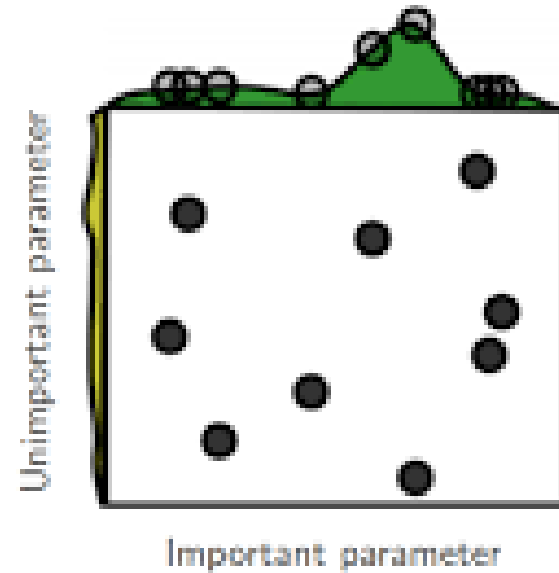
Un bosque aleatorio es un grupo de árboles de decisión. Sin embargo, hay algunas diferencias entre los dos. Un árbol de decisiones tiende a crear reglas que utiliza para tomar decisiones. Un bosque aleatorio elegirá características al azar y hará observaciones, construirá un bosque de árboles de decisión y luego promediará los resultados.

# Búsqueda exhaustiva (RandomizedSearchCV)

Grid Layout



Random Layout

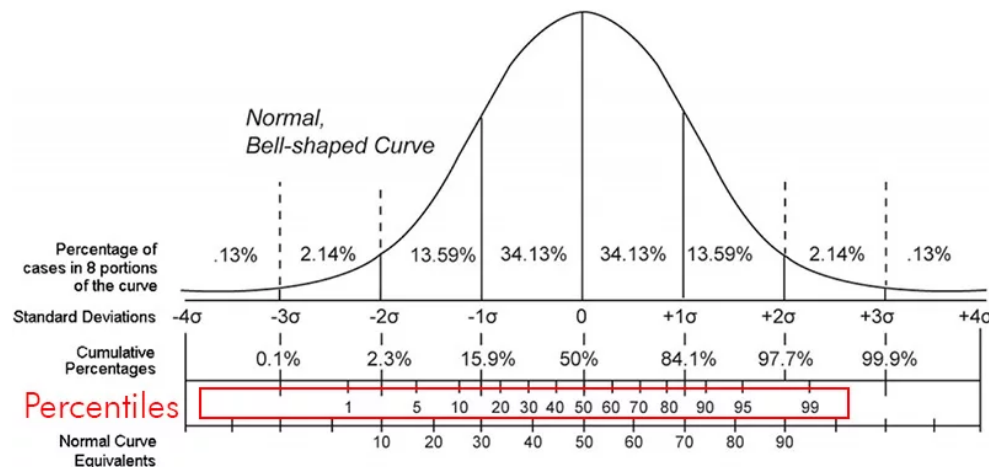


# Selección de características (SelectPercentile)

Selecciona características de acuerdo a la puntuación de los percentiles mas altas.

Para regresión: `r_regression`, `f_regression`, `mutual_info_regression`

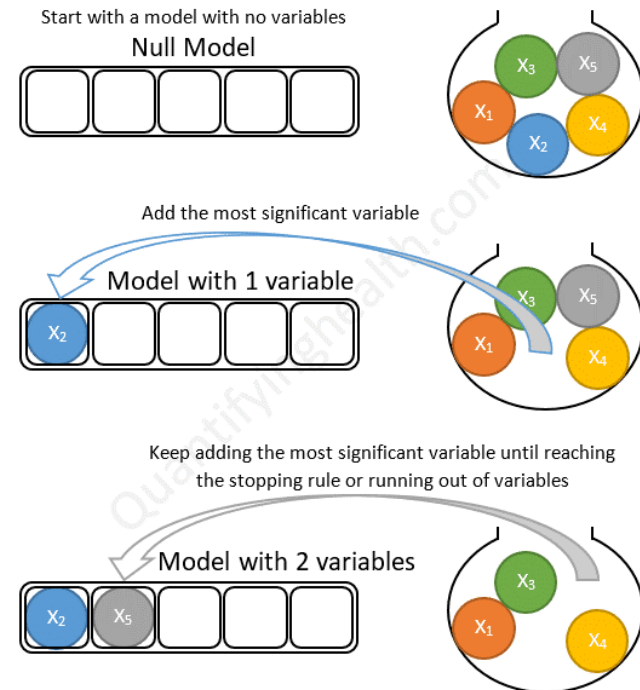
Para clasificación: `chi2`, `f_classif`, `mutual_info_classif`





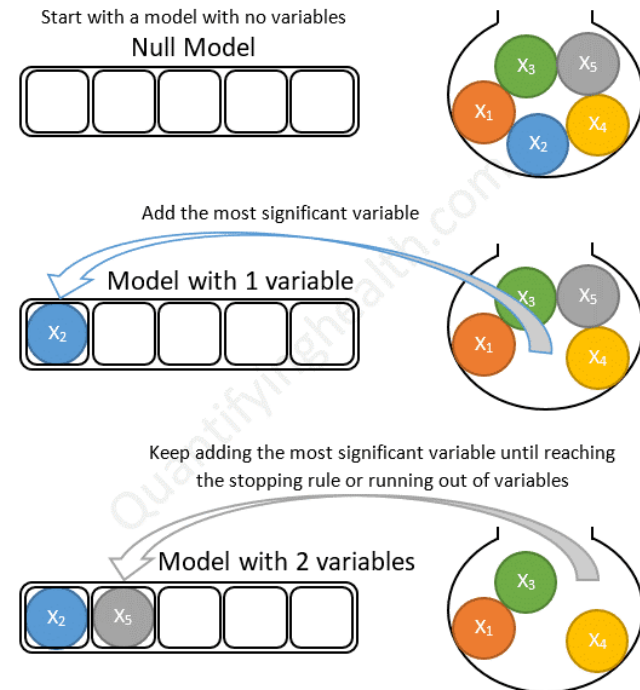
# Selección de características (Selección hacia adelante)

Procedimiento de selección de variables por pasos en el que las variables se introducen secuencialmente en el modelo. La primera variable que se considerará introducir en la ecuación será la que tenga mayor correlación, positiva o negativa, con la variable dependiente.



# Selección de características (Selección hacia adelante)

Procedimiento de selección de variables por pasos en el que las variables se introducen secuencialmente en el modelo. La primera variable que se considerará introducir en la ecuación será la que tenga mayor correlación, positiva o negativa, con la variable dependiente.



# Pipeline

