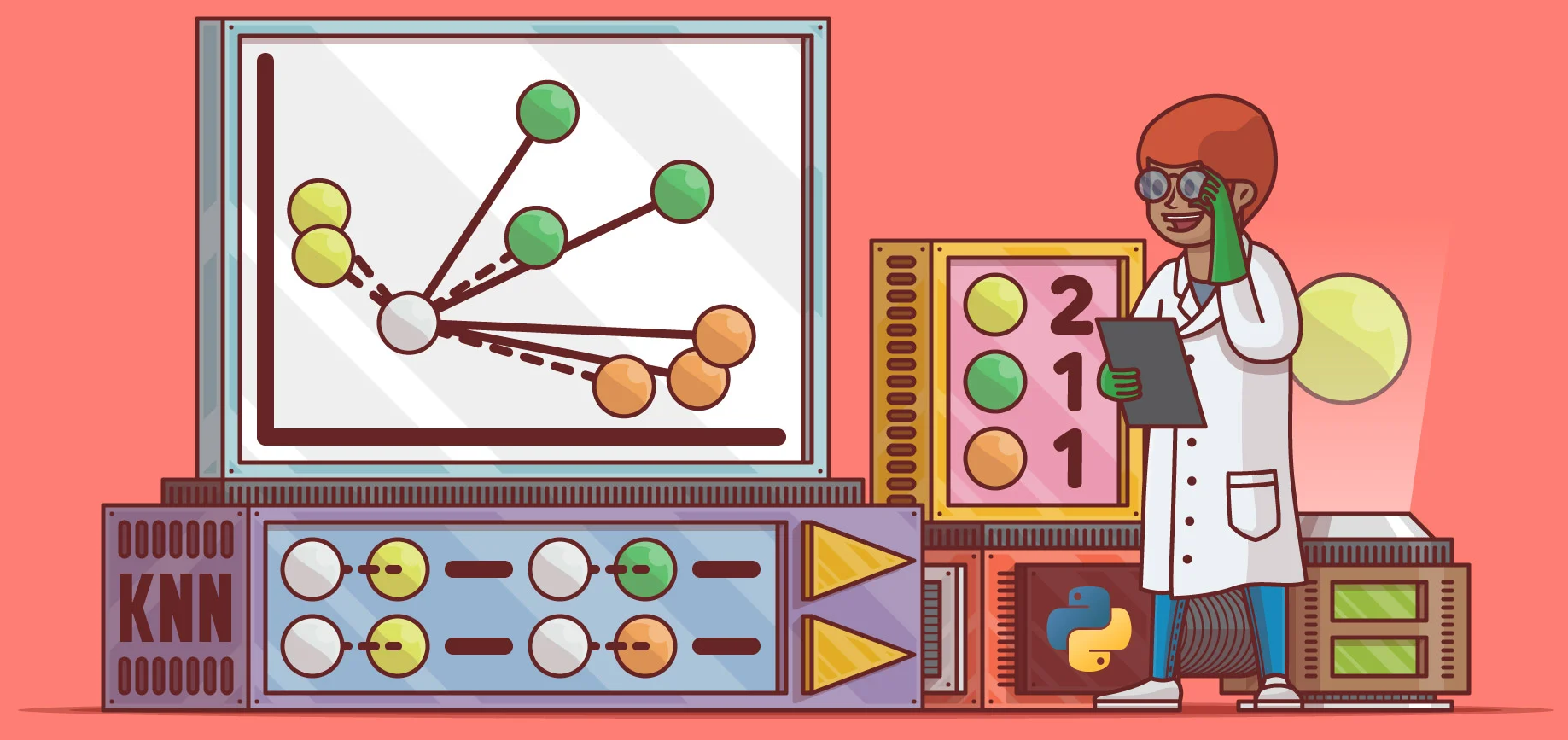
*Algoritmos de Clasificación I I*

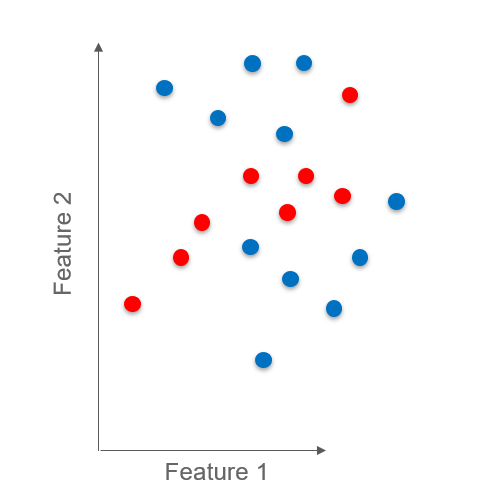
**RECAP: KNN (Vecinos Cercanos):**

Puede usarse para clasificar nuevas muestras (valores discretos) o para predecir (regresión, valores continuos).

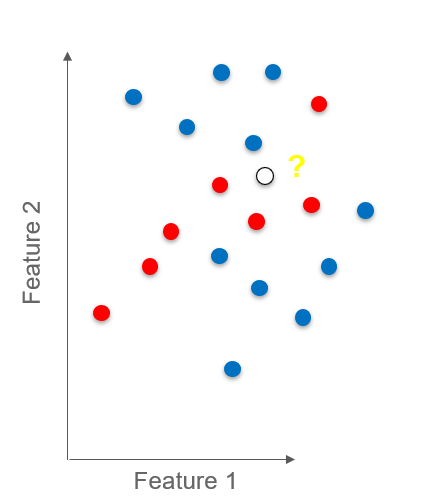
Sirve esencialmente para clasificar valores, buscando los puntos de datos “más similares” (por cercanía).



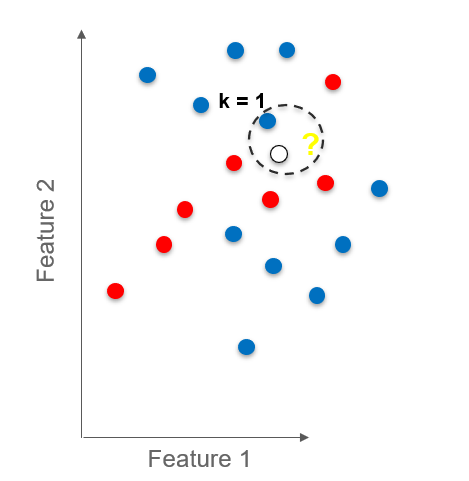
Entonces, supongamos el siguiente escenario: Tenemos un Dataset con 2 Features, en el cual cada instancia puede pertenecer a una de dos clases: “Rojo” o “Azul”.



Dada una nueva instancia, de la cual no sabemos cuál es su clase, vamos a recurrir a sus vecinos cercanos para clasificarla. **La pregunta sería entonces, ¿La clasificamos como rojo o como azul?**

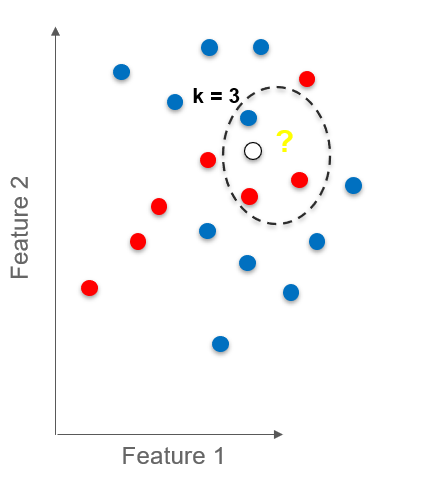


Si tomamos K=1, solo miraremos al vecino más cercano siendo K es el nro de vecinos.



Si elegimos otro valor de k, por ejemplo k > 1, nuestra clasificación cambiará significativamente.

Por ejemplo, con k = 3 tenemos dos vecinos Rojos y uno Azul. Por lo tanto en base a este escenario, la clasificación será: Rojo.



***RAMDOM FOREST – BOSQUES ALEATORIOS***

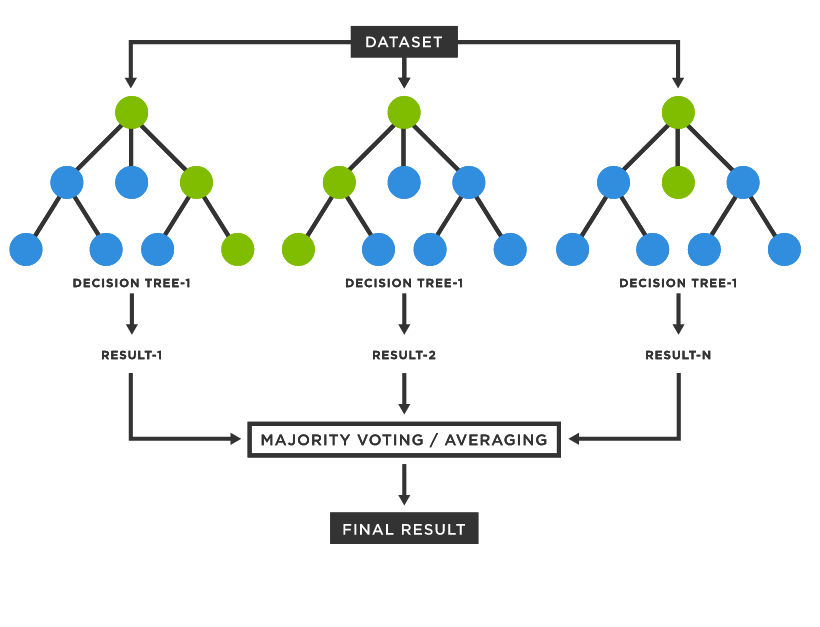
Random Forest es un tipo de Ensamble en Machine Learning en donde combinaremos diversos árboles de decisión. Pero entonces, ¿qué son los métodos de Ensamble en ML?

Los métodos de ensamble de modelos o métodos combinados intentan ayudar a mejorar el rendimiento de los modelos de Machine Learning. Este es un proceso mediante el cual se construyen estratégicamente varios modelos de Machine Learning para resolver un problema particular.

Por otro lado, resulta importante mencionar, que Random Forest, al igual que el árbol de decisión, son modelos de aprendizaje supervisado comúnmente utilizados en problemas clasificación (aunque también puede usarse para problemas de regresión).

***¿Cómo surge Random Forest?***

Uno de los problemas que aparecía con la creación de un árbol de decisión es que si le damos la profundidad suficiente, el árbol tiende a “memorizar” las soluciones en vez de generalizar el aprendizaje. Es decir, a padecer de **overfitting**. La solución para evitar esto es la de crear muchos árboles y que trabajen en conjunto.



**RECAP! – Regresión Logística**

Técnica de aprendizaje automático que proviene del campo de la estadística. A pesar de su nombre no es un algoritmo, sino que es un método para problemas de clasificación, en los que se obtienen un valor binario entre 0 y 1.

*Ejemplo:*

Un problema de clasificación **es identificar si una operación dada es fraudulenta o no**, asociándole una etiqueta “fraude” a unos registros y “no fraude” a otros. 👌

Entonces, la Regresión Logística describe y estima la relación entre una variable binaria dependiente y las variables independientes.

Si la curva va a **infinito positivo** la predicción se convertirá en 1, y si la curva pasa el **infinito negativo**, la predicción se convertirá en 0.

Si la salida de la función Sigmoide es **mayor que 0.5**, podemos clasificar el resultado como 1 o SI, y si es **menor que 0.5** podemos clasificarlo como 0 o NO.

Por su parte **si el resultado es 0.75**, podemos decir en términos de probabilidad como, hay un 75% de probabilidades de que el paciente sufra cáncer.

