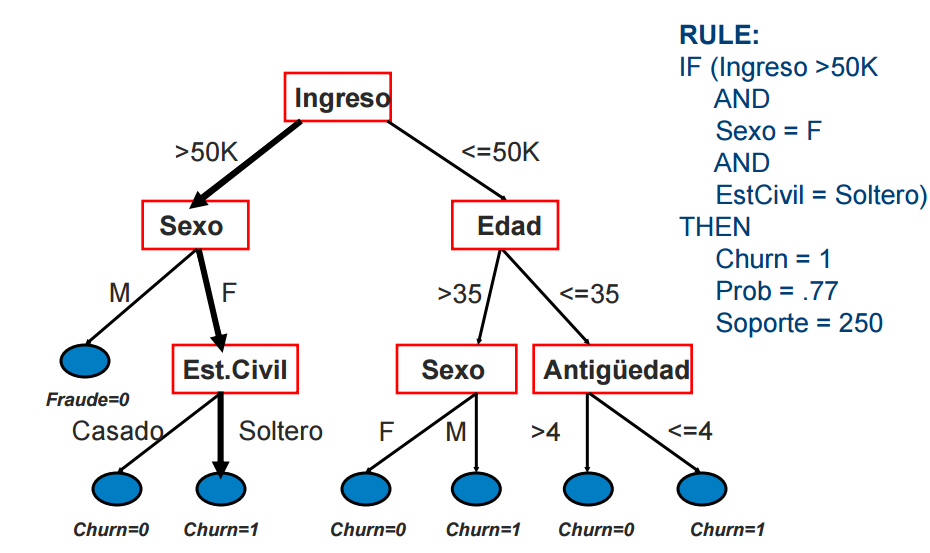
* **Árbol de decisión (Decision Tree)**

Decision Tree es un algoritmo de aprendizaje automático no paramétrico supervisado que se utiliza tanto para problemas de clasificación como de regresión y nos ofrece una representación gráfica de la estructura en forma de árbol con todas las soluciones posibles y así ayudar a tomar una decisión basada en ciertas condiciones.

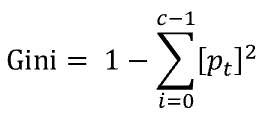
Los árboles de decisión son utilizados en diversos ámbitos de las ciencias que funciona a través de un conjunto de datos y se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

La construcción del árbol está dada por la decisión sobre cómo dividir impacta fuertemente la precisión del árbol. Dividir los nodos en función de las variables de entrada disponibles y así seleccionar la variable de entrada que resulta en el mejor conjunto de datos homogéneo.

Generalmente se utiliza CART, como árbol de clasificación y regresión que utiliza el índice de Gini (medida de impureza) y el índice de ganancia de información para construir árboles.



El índice de Gini es una medida de la pureza o impureza del nodo que hace referencia a la frecuencia con la que una variable elegida al azar se clasificará erróneamente.

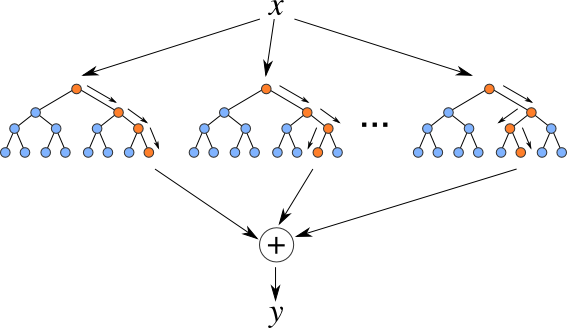


Donde C, representa el número de clases o etiquetas de la variable respuesta.

* **Bosques aleatorios ('Random forest')**

Los bosques aleatorios son una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de bagging que construye una larga colección de árboles no correlacionados y luego los promedia.

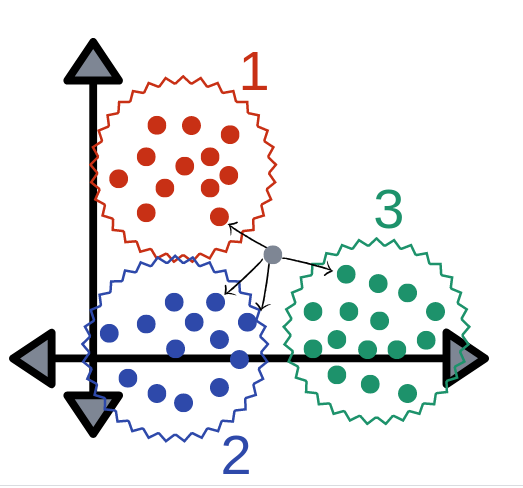
El algoritmo para inducir un random forest fue desarrollado por Leo Breiman y Adele Cutler, el método combina la idea de bagging y la selección aleatoria de atributos, para construir una colección de árboles de decisión con variación controlada.



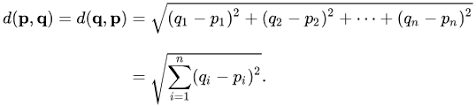
En general, son una colección de árboles de decisión entrenados cada uno con un subconjunto de los datos disponibles y que se ocupan sólo de un subconjunto de las características. La decisión final se toma como una media de las decisiones individuales de los árboles las cuales están basadas en los criterios anteriormente mencionados para el árbol de decisión.

* **KNN (K-Neirest Neigboth)**

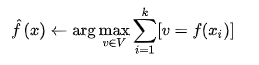
Es un método de clasificación supervisado no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C a partir de la información proporcionada por el conjunto de prototipos. En el proceso de aprendizaje no se hace ninguna suposición acerca de la distribución de las variables predictoras.



Al implementar KNN, el primer paso es transformar los puntos de datos en vectores de características, o su valor matemático. Luego, el algoritmo funciona al encontrar la distancia entre los valores matemáticos de estos puntos. La forma más común de encontrar esta distancia es la distancia euclidiana, como se muestra a continuación.



KNN ejecuta esta fórmula para calcular la distancia entre cada punto de datos y los datos de prueba. Luego encuentra la probabilidad de que estos puntos sean similares a los datos de la prueba y los clasifica en función de qué puntos comparten las probabilidades más altas.



* **Regresión Logística (Logistic regression):**

Los modelos de regresión logística pretenden conocer la relación entre una variable dependiente cualitativa dicotómica (regresión logística binaria) o con más de dos categorías (regresión logística multinomial) y entre variables explicativas independientes o también llamadas covariables, las cuales pueden ser de naturaleza cuantitativa o cualitativa.

En las variables cualitativas dicotómicas, la ausencia de la característica toma el valor 0 y su presencia toma el valor 1, la claridad en la codificación es muy relevante a la hora de la modelación ya que cualquier otra codificación provocaría modificaciones en la interpretación del modelo.

Para las variables multinomiales como en nuestro caso (con más de una categoría) se debe realizar una transformación, para poderlas incluir en el modelo. Esto consiste en crear variables dummy o ficticias, de tal forma que una se tomaría como referencia y cada una de las variables creadas entraría en el modelo de forma individual. Si la variable multinomial posee n categorías, se generarían n-1 variables ficticias.

La ecuación para un modelo de regresión logística es:

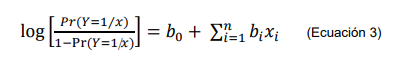


Al dividir la ecuación 1 por el complementario, se obtiene la siguiente expresión



La parte izquierda de la ecuación 2 es conocida como la odds y es una expresión matemática más sencilla de manejar.

Si se aplica el logaritmo natural a la ecuación 2, el término de la derecha se linealiza, obteniendo una expresión matemáticamente más sencilla de manejar:



El logit es el logaritmo natural de la odds de la variable dependiente, es decir la expresión que se encuentra al lado izquierdo de la ecuación 3, y la expresión al lado derecho es la expresión de una recta La ecuación 3 se puede simplificar de la siguiente forma:



Finalmente, la ecuación 4 representa de manera general un modelo logístico.

* **Bayes ingenuo (Naive Bayes)**

Es una aplicación simplificada de la estadística bayesiana en que, se parte de unos valores iniciales de confianza en una clasificación y, durante el entrenamiento se van ajustando.

En ellos se asume que las variables predictoras son independientes entre sí y que la presencia de una cierta característica en un conjunto de datos no está en absoluto relacionada con la presencia de cualquier otra característica.

Proporcionan una manera fácil de construir modelos con un comportamiento muy bueno debido a su simplicidad. Lo consiguen proporcionando una forma de calcular la probabilidad ‘posterior’ de que ocurra un cierto evento A, dadas algunas probabilidades de eventos ‘anteriores’.

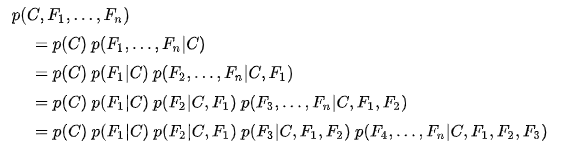
En general el modelo de probabilidad para un clasificador se define como:



Donde C es una variable independiente que está condicionada por varias variables independientes F que se podría resumir en la siguiente expresión:



Donde el numerador es equivalente a la probabilidad compuesta aplicando la definición de probabilidad condicional:



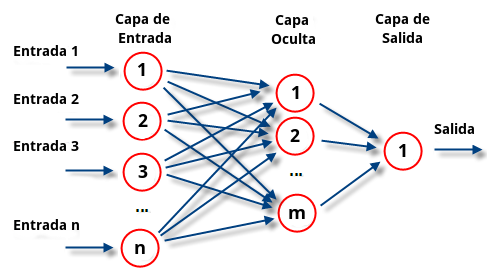
En este punto el supuesto “naive” de independencia condicional entra y asume que cada variable F es independiente cuando están condicionadas a C, luego se reescribe la probabilidad compuesta tal que la probabilidad final está dada por:



* **Red Neuronal con perceptrón multicapa (MLP)**

Un perceptrón multicapa es una clase de alimentación directa a la red neuronal artificial. Un MLP consiste en, al menos, tres capas de nodos: una capa de entrada, una capa oculta y una capa de salida. A excepción de los nodos de entrada, cada nodo es una neurona que utiliza una función no lineal de activación. MLP utiliza un aprendizaje supervisado técnica llamada retropropagación para la formación. Sus múltiples capas y la activación no lineal distinguen la red MLP de un lineal perceptrón, también se puede distinguir los datos que no es [linealmente separables](https://es.qwe.wiki/wiki/Linear_separability).

Las redes con perceptrones multicapa a veces se denominan coloquialmente como redes "vainilla", especialmente cuando tienen una sola capa oculta (al ser más básica).



Sin pérdida de generalidad, el MLPClassifier puntualmente se entrena de forma iterativa, ya que en cada paso se calculan las derivadas parciales de la función de pérdida con respecto a los parámetros del modelo para actualizar los parámetros. El aprendizaje ocurre en el perceptrón cuando se cambian los pesos de las conexiones, llevándose a cabo un proceso que denominan “retropropagación”.

Se representa el error en un nodo j en un punto n de los datos como:



Donde y es el valor producido por el perceptrón dado un valor d como valor objetivo, cuando se hacen las correcciones de pesos en los nodos basados en esta definición se está minimizando el error en toda la salida.



El cual acompañado de un descenso de gradiente, nos da el cambio en cada peso de la siguiente manera:



Siendo este, el peso de la salida de la neurona anterior en una determinada observación que a su vez contiene un parámetro de ritmo de aprendizaje para que los pesos converjan a una respuesta suficientemente rápido sin producir oscilaciones.

La derivada a ser calculada para la función de activación se describe así:



Y para los cambios en pesos de un nodo oculto, la derivada puede definirse como:



Finalmente, para cambiar los pesos de una capa oculta, se debe primero cambiar los pesos de la capa de salida de forma acorde a la derivada de la función de activación.

**Métricas consideradas**

**Log Loss (Pérdida logarítmica)**

La pérdida logarítmica mide el rendimiento de un modelo de clasificación donde la entrada de predicción es un valor de probabilidad entre 0 y 1. El objetivo de nuestros modelos de aprendizaje automático es minimizar este valor. Un modelo perfecto tendría una pérdida logarítmica de 0. La pérdida logarítmica aumenta a medida que la probabilidad prevista difiere de la etiqueta real.

Visualmente los posibles valores de la Log Loss dada una observación verdadera estarían dado por el siguiente gráfico:

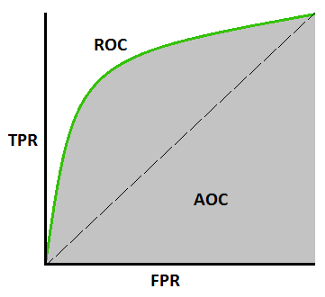


A medida que la probabilidad pronosticada se acerca a 1, la pérdida logarítmica disminuye lentamente. no obstante, a medida que disminuye la probabilidad pronosticada, la pérdida logarítmica aumenta rápidamente. La pérdida de registros penaliza ambos tipos de errores, pero especialmente aquellas predicciones que son seguras e incorrectas.

**AUC**

La curva AUC - ROC es una medida de rendimiento para problemas de clasificación en varios ajustes de umbrales. ROC es una curva de probabilidad y AUC representa el grado o medida de separabilidad. Indica cuánto es capaz de distinguir el modelo entre clases.

Cuanto mayor sea el AUC, mejor será el modelo para predecir 0s como 0s y 1s como 1s. Por analogía, cuanto mayor sea el AUC, mejor será el modelo para distinguir entre clientes que si realmente pagaran, lo harán oportunamente o no pagaran correctamente.



Un AUC cerca del 1, significa que tiene una buena medida de separabilidad. Un modelo pobre tiene AUC cerca del 0, lo que significa que tiene la peor medida de separabilidad. De hecho, significa que está correspondiendo el resultado. Está prediciendo 0s como 1s y 1s como 0s. Y cuando AUC es 0.5, significa que el modelo no tiene capacidad de separación de clases.

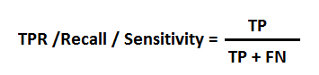
**Definición de términos utilizados en las curvas AUC y ROC:**

Esta curva representa dos parámetros:

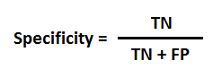
* Tasa de verdaderos positivos
* Tasa de falsos positivos

## **TPR (Tasa de verdaderos positivos) / recuperación / sensibilidad:**

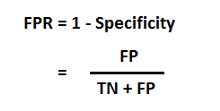
Tasa de verdaderos positivos (TPR) es sinónimo de exhaustividad y, por lo tanto, se define de la siguiente manera:



## **Especificidad:**



## **FPR (Tasa de falsos positivos)** se define de la siguiente manera:



**Pred y Support:**

Hacen referencia al número de ocurrencias de cada clase en las nuevas etiquetas de Y (y\_pred)

**F1 - Score**

la puntuación F1 se define como la media armónica entre precisión y recuperación, pues se utiliza como medida estadística para calificar el rendimiento. En otras palabras, un puntaje F1 (de 0 a 9, siendo 0 el más bajo y 9 el más alto) es una media del rendimiento de un individuo, basado en dos factores, es decir, precisión y memoria.

Considera tanto la [precisión](https://en.wikipedia.org/wiki/Precision_(information_retrieval)) p como la [recuperación](https://en.wikipedia.org/wiki/Recall_(information_retrieval)) r de la prueba para calcular la puntuación: p es el número de resultados positivos correctos dividido por el número de todos los resultados positivos devueltos por el clasificador, yr es el número de resultados positivos correctos dividido por el número de todas las muestras relevantes (todas las muestras que deberían haberse identificado como positivas)

.