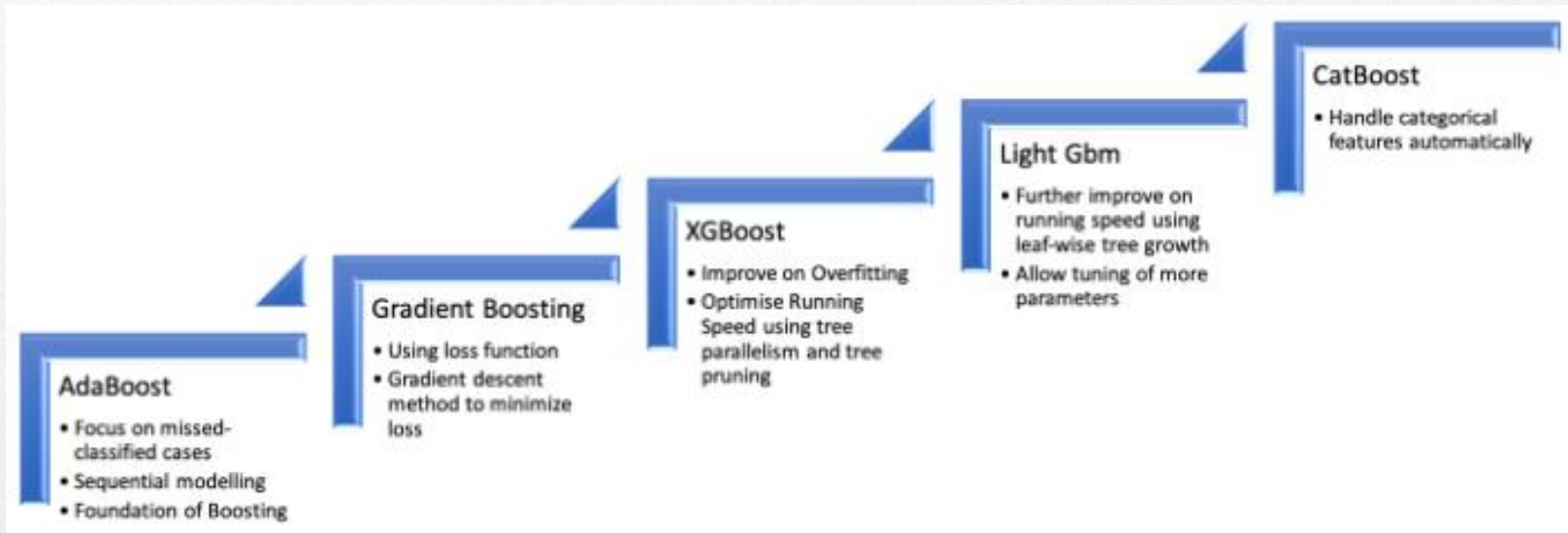


Algoritmos de Boosting

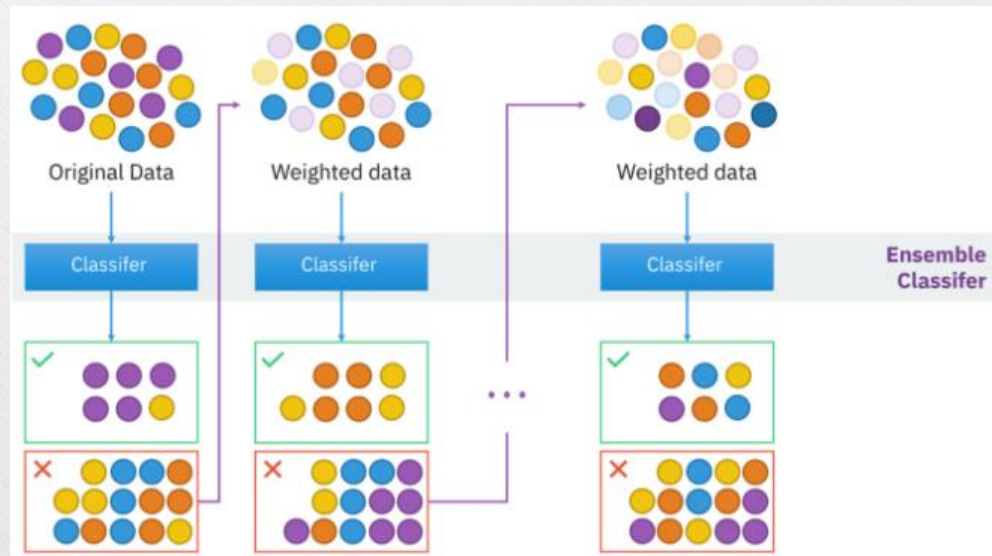
Evolución del Boosting



<https://towardsdatascience.com/how-to-select-between-boosting-algorithm-e8d1b15924f7>

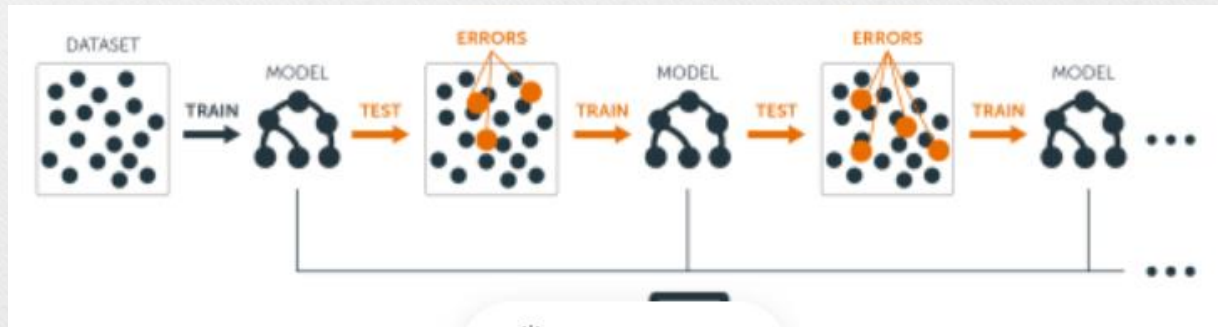
ADABOOST

AdaBoost es la forma abreviada de Boosting Adaptativo. Se llama adaptativo porque el algoritmo aprovecha los casos clasificados incorrectamente del modelo anterior y crea una nueva muestra de datos ponderada en la que los casos clasificados incorrectamente tienen más ponderación. De esa manera, el modelo recién agregado debería ser más adaptable para resolver el "error" del modelo anterior.



GRADIENT BOOSTING

- En comparación con AdaBoost, Gradient Boosting no penaliza los casos clasificados perdidos, sino que utiliza la función de pérdida en su lugar.
- La función de pérdida puede ser el error promedio medio para la regresión o la pérdida logarítmica para los problemas de clasificación. Además, el algoritmo de Gradient Boosting utiliza el método de descenso de gradiente para minimizar continuamente la función de pérdida para encontrar el punto óptimo.
- El método Gradient Boosting teóricamente funciona mejor que AdaBoost. Sin embargo, es más propenso a problemas de sobreajuste y el tiempo de ejecución es más largo.



- https://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf

XGBOOST

- XGBoost significa Extreme Gradient Boosting. Es una versión mejorada del método de Gradient Boosting.
- En primer lugar, mejora el sobreajuste mediante el uso de la regularización.
- En segundo lugar, mejora la velocidad del tiempo de ejecución al optimizar la clasificación mediante la ejecución en paralelo.
- Por último, utiliza la profundidad máxima del árbol de decisión como parámetro para podar el árbol, lo que reduce significativamente el tiempo de ejecución.

LIGHT GBM

- Light Gbm mejora aún más el tiempo de ejecución del programa al hacer que la carga de trabajo informática sea "liviana". Sin embargo, todavía puede mantener el mismo nivel o un nivel superior de rendimiento del modelo en comparación con otros algoritmos.
- Light Gbm optimiza la velocidad y la precisión del tiempo de ejecución principalmente de dos maneras:
 - Adopta el algoritmo basado en histograma, dividiendo las variables continuas en diferentes cubos (en lugar de ordenarlas individualmente). Esto mejora mucho el tiempo de ejecución.
 - Utiliza el método de crecimiento de árboles por hojas en lugar del método de crecimiento de árboles por niveles (utilizado por la mayoría de los otros métodos basados en árboles de decisión). Permite que la parte de la hoja con mayor pérdida continúe creciendo (corte de mejor ajuste) y, por lo tanto, minimiza la función de pérdida general



CATBOOST

- CatBoost significa Categorical Boost. Tiene la gran característica de manejar automáticamente variables categóricas sin necesidad de convertirlas en numéricas.
- CatBoost es el más reciente de los 5 algoritmos de boosting pero muy cerca de Light Gbm. Funciona mejor cuando hay más variables categóricas.
- <https://catboost.ai/en/docs/>