

Modelo ML	Condiciones	Ventajas	Desventajas	Escala- miento de Variables	Datos Ausentes	Outliers	Apropiado para	Tipo de aprendizaje	Ejemplo de uso
Clasificador baSiiano ingenuo - Naïve BaSi	Variables son independientes	<ul style="list-style-type: none">Buen desempeño con las variables categóricas.Converge rápidamente: menos tiempo de entrenamientoBueno en casos de conjuntos de datos relativamente grandesBuen en conjuntos de datos con múltiples variables	<ul style="list-style-type: none">Variables correlacionadas afectan el desempeño	No	Puede manejar datos ausentes (ignora los datos faltantes)	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónClasificación Multiclass	Supervisado	<ul style="list-style-type: none">Análisis Sentimental (Sentiment Analysis)Categorización de documentosFiltro de email no deseados (spam)
Máquinas de Soporte Vectorial - Support Vector Machine (SVM)	Ninguna	<ul style="list-style-type: none">Bueno para conjuntos de datos con más variables que observaciones.Buen desempeñoBuen modelo disponible para usar en una variedad de escenariosPuede aproximar funciones no lineales complejas	<ul style="list-style-type: none">Requiere largo tiempo de entrenamientoAjustes son requeridos para determinar cual kernel es óptima para SVMs no lineales	Si	Sensible	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Predicciones de bolsa de valoresDeterminación del valor en riesgo
Regresión Lineal	<ul style="list-style-type: none">Relación lineal entre las variables y el target.Variables están normalmente distribuidasHomocedasticidad	<ul style="list-style-type: none">InterpretablePoco ajuste necesario	<ul style="list-style-type: none">Variables correlacionadas afectan el desempeñoRequire bastante ingeniería de variables	Si	Sensible	Sensible	Regresión	Supervisado	<ul style="list-style-type: none">Predicción de ventasPredicción de precios de casas
Regresión Logística	<ul style="list-style-type: none">Relación lineal entre las variables y los logits de las probabilidadesVariables están normalmente distribuidas.Homocedasticidad	<ul style="list-style-type: none">InterpretablePoco ajuste necesario	<ul style="list-style-type: none">Variables correlacionadas afectan el desempeñoRequire bastante ingeniería de variables	Si	Sensible	Potentially Sensible	Clasificación	Supervisado	<ul style="list-style-type: none">Evaluación de riesgoPrevención de fraude.
Árboles de clasificación y regresión	Ninguna	<ul style="list-style-type: none">InterpretableImportancia de las variables disponibleNo requiere mucha ingeniería de variables	<ul style="list-style-type: none">No predice una respuesta continua (para regresión)No predice más allá del rango de valores presentes en el set de entrenamientoSobre ajusta los datos	No	Algunas implementaciones no necesitan que se imputen los datos faltantes. La implementación de Scikit-learn sí requiere la imputación.	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Evaluación de riesgoPrevención de fraude
Bosques aleatorios - Random Forests	Ninguna	<ul style="list-style-type: none">InterpretableImportancia de las variables disponibleNo requiere mucha ingeniería de variables(En principio) No sobreajustaBuen desempeño / precisiónRobusto al ruido de las variablesPoco (o casi nada) ajuste de los parámetros requeridoApto para la mayoría de problemas de aprendizaje automático	<ul style="list-style-type: none">No predice una respuesta continua (para regresión)No predice más allá del rango de valores presentes en el set de entrenamientoSesgado hacia variables categóricas con múltiples etiquetasSesgado en problemas con múltiples clases hacia aquellas que son más frecuentes	No	Algunas implementaciones no necesitan que se imputen los datos faltantes. La implementación de Scikit-learn sí requiere la imputación.	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Evaluación de riesgo crediticioPredicción de falla de piezas mecánicas (industria automovilística).Evaluación de la probabilidad de desarrollar una enfermedad crónica (sector salud)Predicción del número de ‘compañeros’ en medios sociales.
Árboles de de Potenciación de gradiente - Gradient Boosted Trees	Ninguna	<ul style="list-style-type: none">Muy buen desempeñoApto para la mayoría de problemas de aprendizaje automáticoPuede aproximar la mayoría de funciones no lineales	<ul style="list-style-type: none">Tiende a sobre ajustarNecesita ajuste de parámetros	No	Algunas implementaciones no necesitan que se imputen los datos faltantes (e.j. xgboost) La implementación de Scikit-learn sí requiere imputación.	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Igual que los Bosques Aleatorios
K-vecinos cercanos - K-nearest neighbours	Ninguna	<ul style="list-style-type: none">Buen desempeño	<ul style="list-style-type: none">Lento en producir las prediccionesSusceptible a espacios de variables de alta dimensionalidad (muchas variables)	Si	Sensible	Robusto a los outliers	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Expresión de genesInteracciones entre proteínasExtracción de contenido (de páginas web por ejemplo)
Clasificador Boosting Adaptivo - AdaBoost	Ninguna	<ul style="list-style-type: none">No necesariamente sobre ajusta los datosMenos parámetros para ajustar	<ul style="list-style-type: none">Puede ser sensitivo al ruido y a los outliers	No	Puede manejarlos	Sensible	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Igual que los Bosques Aleatorios, sin embargo AdaBoost es menos popular, que xgboost y lightGBMs, implementaciones más populares de las máquinas de potenciación de gradiente.
Redes Neuronales - Neural Networks	Ninguna	<ul style="list-style-type: none">Puede aproximar cualquier funciónMuy buen desempeño	<ul style="list-style-type: none">Tiempos largos de entrenamientoNecesita del ajuste de múltiples parámetros, incluyendo la arquitectura de la red neuronalTiende a sobre ajustar los datosModelo poco interpretable	Si	Sensible	Puede manejar outliers pero el desempeño se afecta en la presencia de muchos .	<ul style="list-style-type: none">ClasificaciónRegresión	Supervisado	<ul style="list-style-type: none">Análisis de imágenesPredicciones (Forecasting)Análisis de Texto
Algoritmo de	Conjuntos de	Puede aproximar	Algunas implementaciones	Si	Algunas implementaciones	Sensible	Supervisado	Si	Supervisado