Comparison of Machine Learning Algorithms

Model	Description	Advantages	Disadvantages	Areas of Application
KNN	A non-parametric algorithm that uses proximity to classify data points	• Simple • Easy to understand • Good for small datasets	• Sensitive to outliers • Computationally expensive	• Image recognition • Recommender systems • Ecology
Naive Bayes	A probabilistic algorithm that applies Bayes' theorem to classify data points	• Fast • Works well with high-dimensional data • Handles missing data	• Assumes independence of features • Not suitable for complex data	• Email spam filtering • Text classification • Sentiment analysis
Linear Regression	A linear approach to modeling the relationship between a dependent variable and one or more independent variables	• Simple • Easy to interpret • Good for continuous variables	• Assumes linear relationship • Sensitive to outliers	• Sales forecasting • Stock prices • Trend analysis
Logistic Regression	A statistical model used to analyze and predict binary outcomes	• Simple • Fast • Good for binary outcomes	• Assumes linear relationship • Not suitable for complex data	• Medical diagnosis • Credit scoring • Marketing
Decision Trees	A hierarchical model that partitions data into subsets based on rules or decisions	• Easy to interpret • Can handle both numerical and categorical data	• Can easily overfit • Not suitable for linear data	• Fraud detection • Customer segmentation • Quality control
Random Forest	A collection of decision trees that makes predictions by aggregating the outputs of each tree	• Robust • Handles missing data • Good for complex data	• Can be slow in prediction • Difficult to interpret	• Image recognition • Medical diagnosis • Financial forecasting
Gradient Boosting	An ensemble technique that trains models sequentially, with each subsequent model attempting to correct the errors of the previous model	• Good for high-dimensional data • Can handle non-linear relationships	• Computationally expensive • Sensitive to outliers	• Click-through rate prediction • Fraud detection • Medical diagnosis
SVM	A model that maps data to a high-dimensional feature space to separate classes with a hyperplane	• Effective for high-dimensional data • Can handle non-linear data	• Can be sensitive to kernel choice • Computationally expensive	• Text classification • Image recognition • Bioinformatics
Neural Networks	A set of algorithms designed to recognize patterns by simulating the function of the human brain	• Effective for complex data • Can handle non-linear data • Good for image and speech recognition	• Requires large amounts of data • Computationally expensive • Can be difficult to interpret	• Image recognition • Speech recognition • NLP