<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DESCRIPTION</th>
<th>USE CASES</th>
<th>ADVANTAGES</th>
<th>DISADVANTAGES</th>
</tr>
</thead>
</table>
| Linear Regression          | A simple algorithm that models a linear relationship between inputs and a    | 1. Stock price prediction  
2. Predicting housing prices  
3. Predicting customer lifetime value | 1. Explainable method  
2. Interpretable results by its output coefficients  
3. Faster to train than other machine learning models | 1. Assumes linearity between inputs and output  
2. Sensitive to outliers  
3. Can underfit with small, high-dimensional data |
| Logistic Regression        | A simple algorithm that models a linear relationship between inputs and a    | 1. Credit risk score prediction  
2. Customer churn prediction | 1. Interpretable and explainable  
2. Less prone to overfitting when using regularization  
3. Applicable for multi-class predictions | 1. Assumes linearity between inputs and outputs  
2. Can overfit with small, high-dimensional data  
3. All the predictors are kept in the final model  
4. Doesn't perform feature selection |
| Ridge Regression           | Part of the regression family — it penalizes features that have low        | 1. Predictive maintenance for automobiles  
2. Sales revenue prediction | 1. Less prone to overfitting  
2. Best suited where data suffer from multicollinearity  
3. Explainable & interpretable | 1. Can lead to poor interpretability as it can keep highly correlated variables |
| Lasso Regression           | features that have low predictive outcomes by shrinking their coefficients  | 1. Predicting housing prices  
2. Predicting clinical outcomes based on health data | 1. Less prone to overfitting  
2. Can handle high-dimensional data  
3. No need for feature selection | 1. Can lead to poor interpretability as it can keep highly correlated variables |
| Decision Tree              | Decision Tree models make decision rules on the features to produce         | 1. Customer churn prediction  
2. Credit score modeling  
3. Disease prediction | 1. Explainable and interpretable  
2. Can handle missing values | 1. Prone to overfitting  
2. Sensitive to outliers |
| Random Forests             | An ensemble learning method that combines the output of multiple decision   | 1. Credit score modeling  
2. Predicting housing prices | 1. Reduces overfitting  
2. Higher accuracy compared to other models | 1. Training complexity can be high  
2. Not very interpretable |
| Gradient Boosting Regression | Gradient Boosting Regression employs boosting to make predictive models from  | 1. Predicting car emissions  
2. Predicting ride hailing fare amount | 1. Better accuracy compared to other regression models  
2. It can handle non-linearity  
3. It can handle non-linear relationships | 1. Sensitive to outliers and can therefore cause overfitting  
2. Computationally expensive and has high complexity |
| XGBoost                    | Gradient Boosting algorithm that is efficient & flexible. Can be used for    | 1. Churn prediction  
2. Claims processing in insurance | 1. Provides accurate results  
2. Captures non-linear relationships | 1. Hyper-parameter tuning can be complex  
2. Does not perform well on sparse datasets |
| LightGBM Regressor         | A gradient boosting framework that is designed to be more efficient than    | 1. Predicting flight time for airlines  
2. Predicting cholesterol levels based on health data | 1. Can handle large amounts of data  
2. Computational efficient & fast training speed  
3. Low memory usage | 1. Can overfit due to leaf-wise splitting and high sensitivity  
2. Hyper-parameter tuning can be complex |
| K-Means                    | K-Means is the most widely used clustering approach—it determines K clusters  | 1. Customer segmentation  
2. Recommendation systems | 1. Scales to large datasets  
2. Simple to implement and interpret  
3. Results in tight clusters | 1. Requires the expected number of clusters from the beginning  
2. Has trouble with varying cluster sizes and densities |
| Hierarchical Clustering    | A “bottom-up” approach where each data point is treated as its own cluster   | 1. Fraud detection  
2. Document clustering based on similarity | 1. There is no need to specify the number of clusters  
2. The resulting dendrogram is informative | 1. Doesn't always result in the best clustering  
2. Not suitable for large datasets due to high complexity |
| Gaussian Mixture Models    | A probabilistic model for modeling normally distributed clusters within a     | 1. Customer segmentation  
2. Recommendation systems | 1. Computes a probability for an observation belonging to a cluster  
2. Can identify overlapping clusters  
3. More accurate results compared to K-means | 1. Requires complex tuning  
2. Requires setting the number of expected mixture components or clusters |
| Apriori algorithm          | Rule-based approach that identifies the most frequent itemsets in a given    | 1. Product placements  
2. Recommendation engines  
3. Promotional optimization | 1. Results are intuitive and interpretable  
2. Exhaustive approach as it finds all rules based on the confidence and support | 1. Generates many uninteresting itemsets  
2. Computationally and memory intensive  
3. Results in many overlapping item sets |