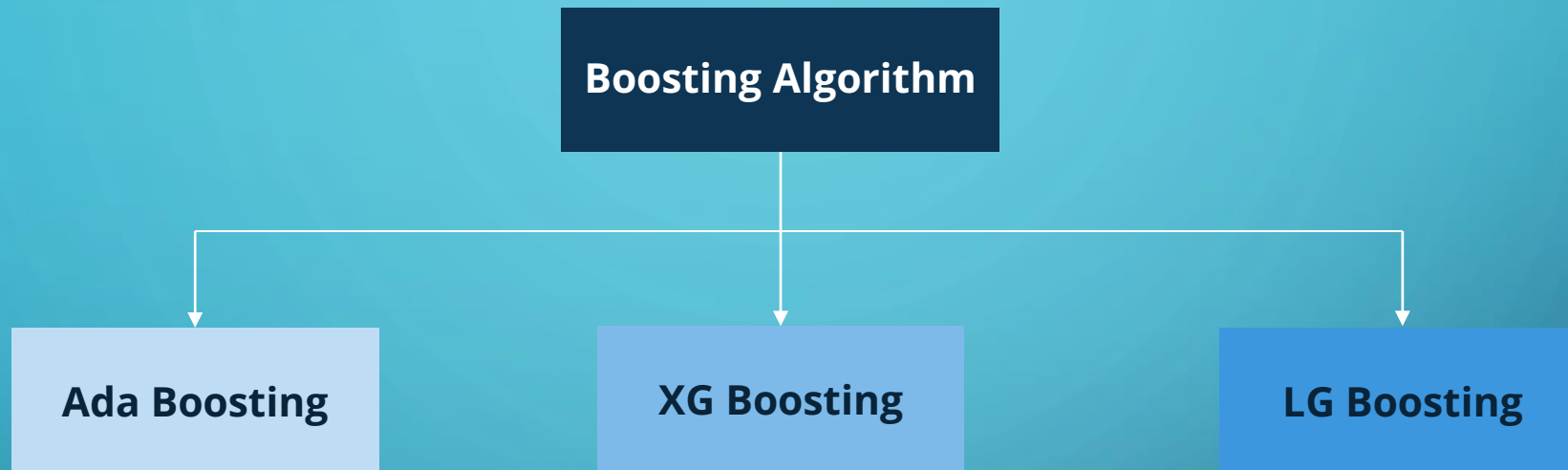


The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit board.

# **BOOSTING ALGORITHM**

# TYPES OF BOOSTING ALGORITHM



# ADA BOOSTING

# ADA BOOSTING

## Explanation

### 1. Combination of Simple Models:

- AdaBoost combines multiple simple models (weak learners) to create a stronger predictive model.

### 2. Sequential Training:

- It trains these simple models one after another, each time focusing more on the errors made by the previous models.

### 3. Error Correction:

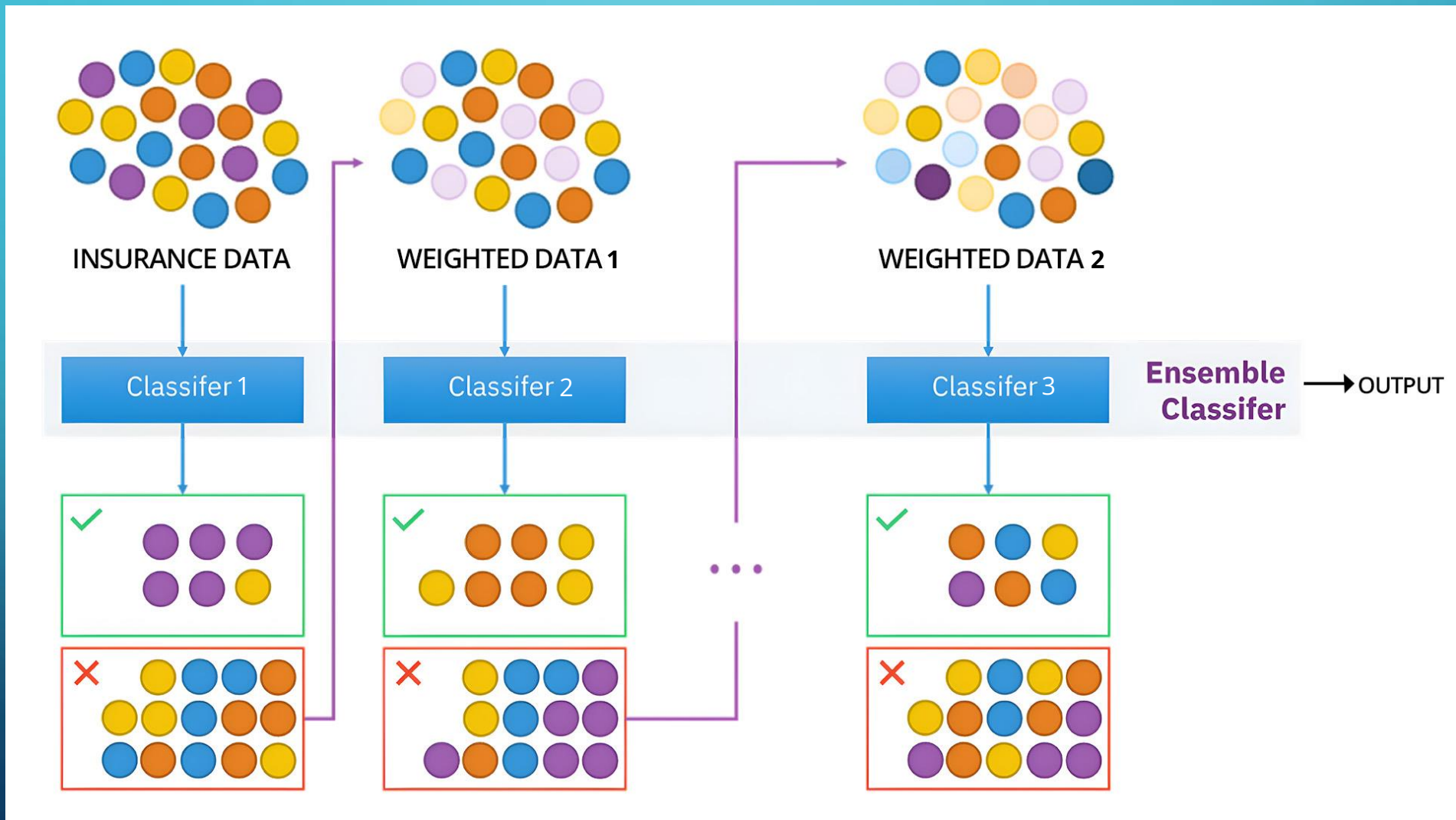
- By giving more weight to previously mis-predicted instances, the algorithm improves on hard-to-predict cases.

### 4. Final Model:

- All trained models are combined into a final model, which is more accurate than any individual simple model, making AdaBoost effective for various prediction tasks.

# ADA BOOSTING

## Schematic Diagram



# ADA BOOSTING

## Advantages

### 1. Improved Accuracy:

- By combining multiple weak learners, AdaBoost can produce a strong model with improved predictive performance.

### 2. Flexibility:

- Can be used with various base learners (like decision trees) and is versatile for both classification and regression tasks.

### 3. Adaptability:

- Automatically adjusts the weights of weak learners based on their performance, focusing on difficult-to-predict instances.

### 4. Reduction of Overfitting:

- Tends to have lower risk of overfitting compared to other ensemble methods.

# ADA BOOSTING

## Disadvantages

### 1. Sensitive to Noisy Data:

- AdaBoost can be sensitive to outliers and noisy data because it tries to fit the difficult instances which may include noise.

### 2. Computational Complexity:

- Can be computationally intensive, especially with a large number of weak learners or complex base estimators.

### 3. Dependency on Base Learner:

- The performance heavily depends on the choice of the base learner and its parameters



# ADA BOOSTING

## Applications

### 1. House Price Prediction:

- Predicting the prices of houses based on various features (like in the provided example).

### 2. Financial Forecasting:

- Used to predict stock prices, market trends, and other financial metrics.

### 3. Medical Diagnosis:

- Can be used to predict the progression of diseases or patient outcomes based on medical data.

### 4. Environmental Modeling:

- Useful for predicting environmental changes like pollution levels, weather forecasting, etc.

### 5. Sales Forecasting:

- Helps in predicting future sales based on historical data and trends.



# XG BOOSTING

# XG BOOSTING

## Explanation

### 1. Initialization:

- Start by assigning equal weights to all training samples, meaning each sample has the same importance initially.

### 2. Train Weak Learner:

- Train a weak learner (e.g., a decision stump) on the weighted training data.
- Evaluate its performance and calculate its error rate.

### 3. Update Weights:

- Increase the weights of the misclassified samples to make them more important.
- Decrease the weights of the correctly classified samples to make them less important.

### 4. Combine Weak Learners:

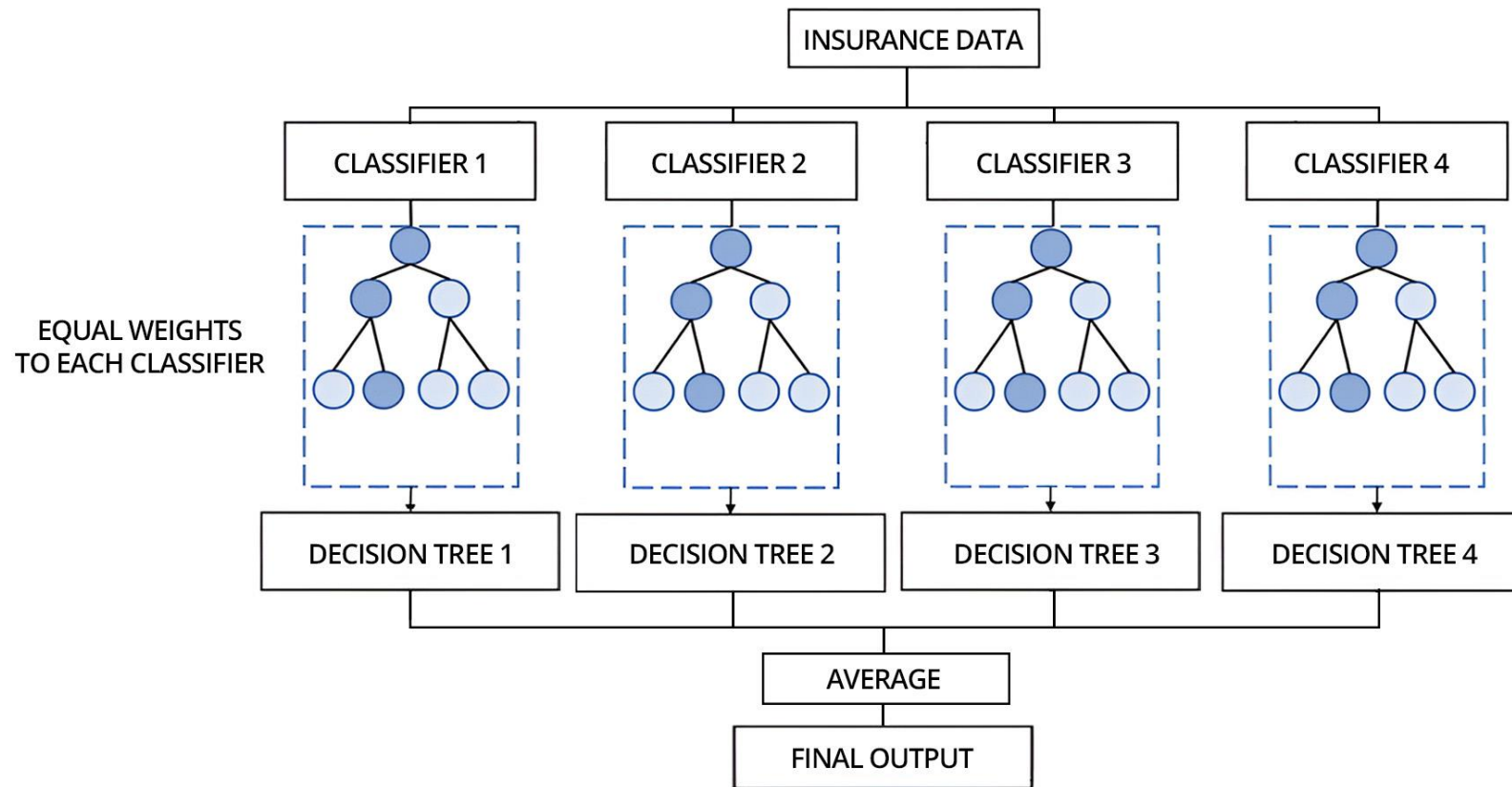
- Assign a weight to each weak learner based on its accuracy.
- Combine the weak learners' predictions to form a strong classifier, where each learner's influence is proportional to its accuracy.

### 5. Iterate:

- Repeat the process for a specified number of iterations or until the error is minimized, each time focusing more on the samples that were misclassified in previous iterations

# XG BOOSTING

## Schematic Diagram



# XG BOOSTING

## Advantages

### 1. High Performance:

- XG Boost is known for its high efficiency and speed. It is optimized for performance can handle large datasets effectively.

### 2. Regularization:

- The algorithm includes L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting, which is not available in other boosting algorithms.

### 3. Parallel Processing:

- XG Boost can use parallel processing to speed up the computation, making it much faster compared to other gradient boosting algorithms.

### 3. Flexibility:

- It supports both classification and regression problems and can handle missing values automatically.

### 4. Handling Missing Values:

- XG Boost has a built-in method to handle missing data efficiently.

### 5. Tree Pruning:

- XG Boost uses a more sophisticated tree pruning algorithm, which gives it an advantage over other boosting algorithms.

# XG BOOSTING

## Disadvantages

### 1. Complexity:

- XG Boost has many hyperparameters to tune, which can be complex and time-consuming.

### 2. Computational Cost:

- While it is efficient, training large datasets with XG Boost can still be computationally expensive.

### 3. Memory Usage:

- It can use a significant amount of memory, especially with large datasets and deep trees.

### 4. Interpretability:

- Like many ensemble methods, XG Boost models can be difficult to interpret compared to simpler models like linear regression or decision trees.

# XG BOOSTING

## Applications

### 1. Finance:

- Predicting stock prices, credit scoring, and fraud detection.

### 2. Marketing:

- Customer segmentation, churn prediction, and targeted marketing.

### 3. Healthcare:

- Disease prediction, patient risk scoring, and medical diagnosis.

### 4. E-commerce:

- Recommendation systems, customer lifetime value prediction, and sales forecasting.

### 5. Energy:

- Predicting energy consumption and optimizing energy usage.

### 6. Sports Analytics:

- Predicting player performance, game outcomes, and optimizing team strategies.

# LG BOOSTING



# LG BOOSTING

## Explanation

### 1. Gradient Boosting Framework:

- LightGBM builds an ensemble of decision trees sequentially, where each new tree corrects the errors made by the previous ones, following the principles of gradient boosting.

### 2. Leaf-wise Growth:

- Instead of traditional level-wise growth, LightGBM grows trees leaf-wise. This means it focuses on the leaves with the highest loss reduction, making the training process more efficient and reducing loss more quickly.

### 3. Histogram-based Approach:

- LightGBM speeds up training by using a histogram-based algorithm. It buckets continuous features into discrete bins, which reduces the complexity and memory usage during the training process.

### 4. Efficient Handling of Large Datasets:

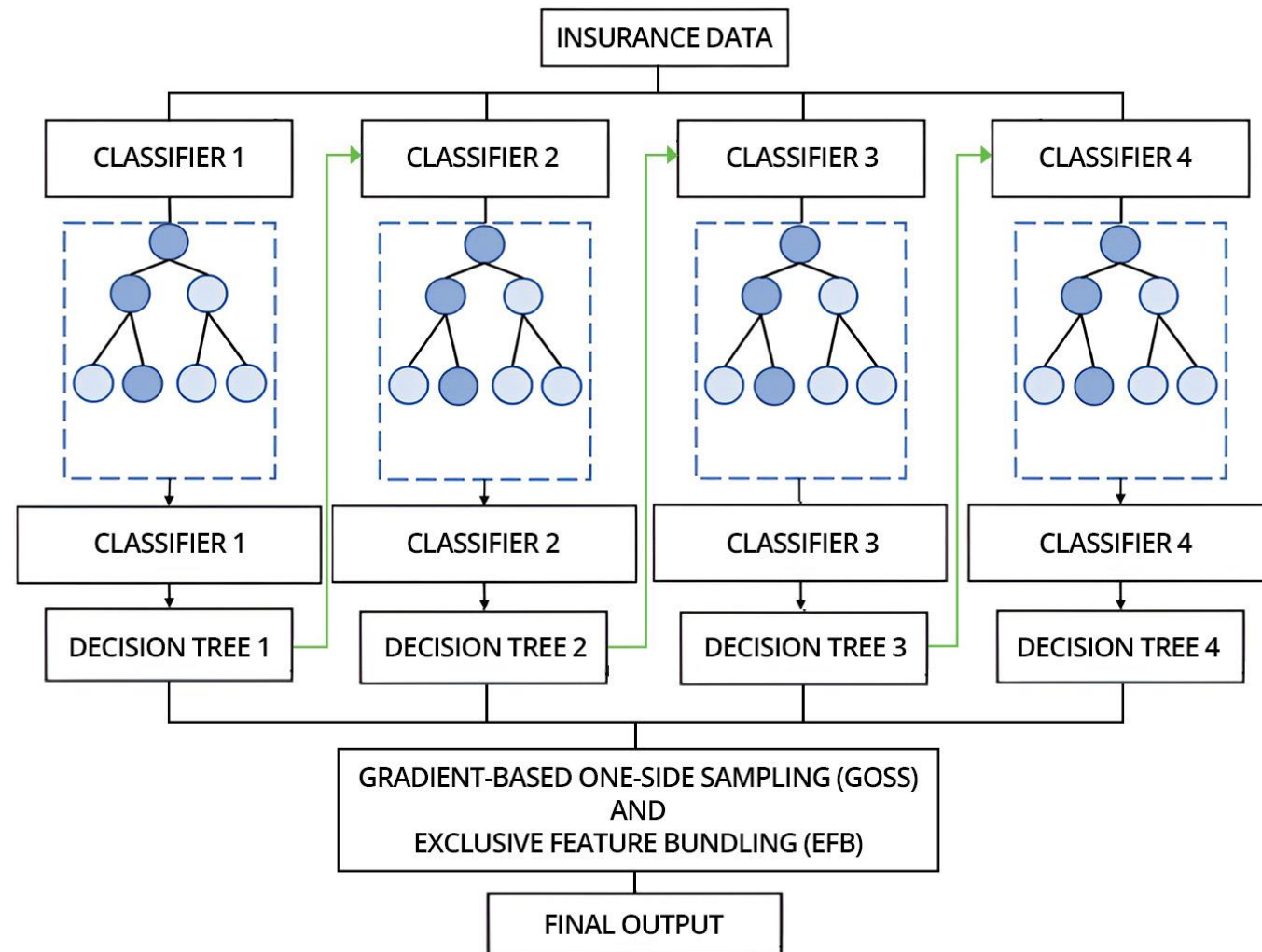
- LightGBM is designed to handle large datasets efficiently. It supports distributed training across multiple machines and includes techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to optimize performance and reduce computational load.

### 5. Versatile and Regularized:

- LightGBM supports a wide range of tasks (regression, classification, ranking) and includes regularization techniques (L1, L2) to prevent overfitting. It offers extensive hyperparameter tuning options to fine-tune model performance.

# LG BOOSTING

## Schematic Diagram



# LG BOOSTING

## Advantages

### 1. Efficiency:

- LightGBM is designed to be efficient both in terms of speed and memory usage. It handles large datasets and high-dimensional data very well.

### 2. Accuracy:

- It often provides better accuracy compared to other boosting algorithms due to its leaf-wise growth strategy.

### 3. Flexibility:

- It supports various objective functions, including regression, classification, and ranking.

### 4. Scalability:

- LightGBM can be easily parallelized and distributed, making it suitable for large-scale data processing.

### 5. Handling of Large Data:

- LightGBM can handle large datasets and is optimized for performance.

# LG BOOSTING

## Disadvantages

### 1. Complexity:

- Tuning the hyperparameters can be complex and requires experience to get the best performance.

### 2. Memory Consumption:

- Although LightGBM is more memory-efficient than some other algorithms, it can still consume a significant amount of memory for very large datasets.

### 3. Sensitivity to Hyperparameters:

- Performance can be sensitive to the choice of hyperparameters, and improper tuning can lead to suboptimal results.

### 4. Less Interpretability:

- As with other tree-based ensemble methods, the resulting models can be less interpretable compared to simpler models like linear regression.

# LG BOOSTING

## Applications

### 1. Finance:

- LightGBM can be used for credit scoring, fraud detection, and risk assessment.

### 2. Marketing:

- Useful for customer segmentation, predicting customer churn, and recommendation systems.

### 3. Healthcare:

- It is applied in predicting patient outcomes, disease diagnosis, and personalized treatment recommendations.

### 4. E-commerce:

- Personalizing product recommendations, predicting sales, and customer lifetime value estimation.

### 5. Real Estate:

- Can be used to predict housing prices, rental prices, and property values.

### 6. Energy:

- Forecasting energy consumption and optimizing energy production.

The background is a blue gradient with faint, concentric circular patterns. In the corners, there are white line-art elements resembling circuit boards or neural network connections, with lines and small circles.

**THE END**