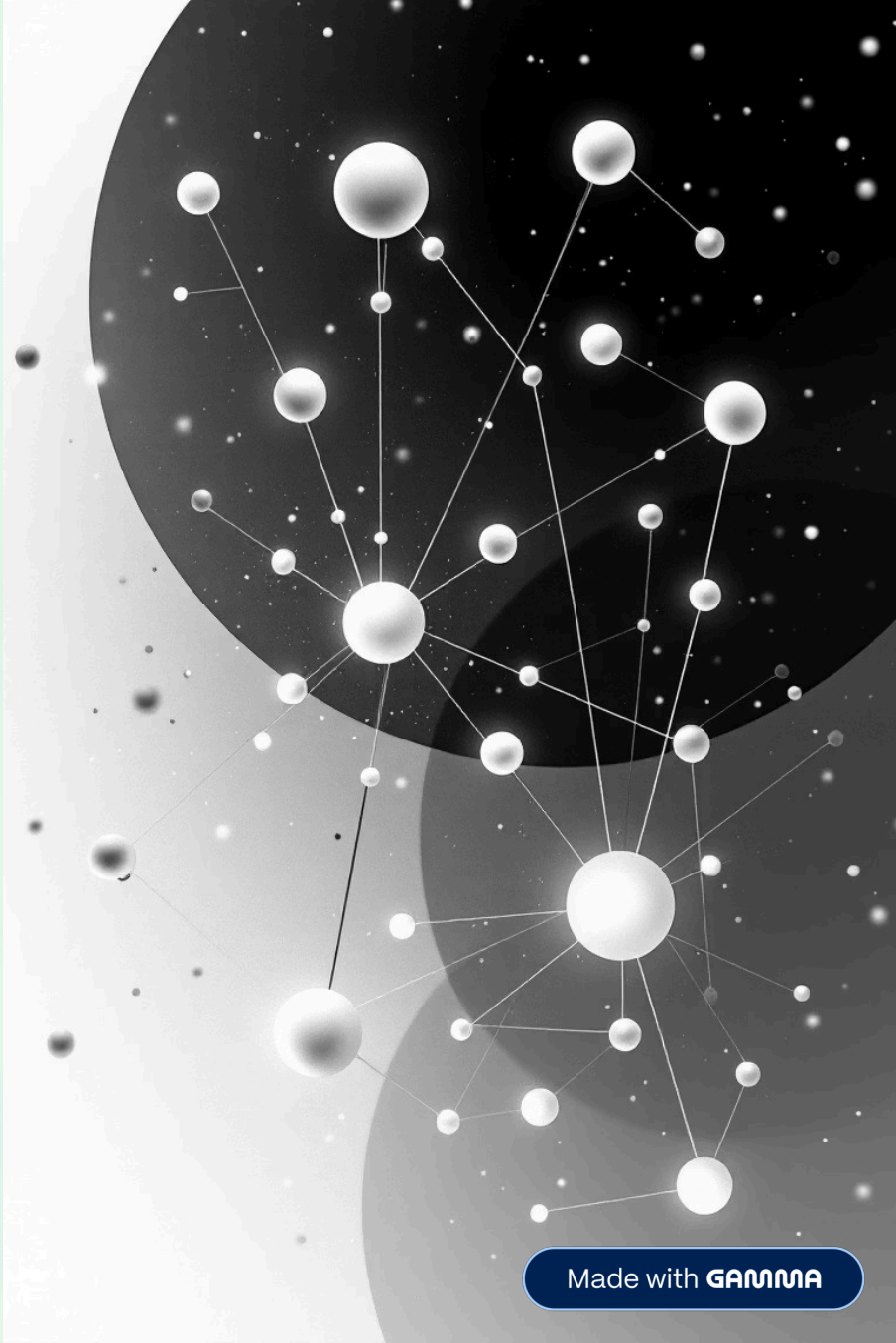
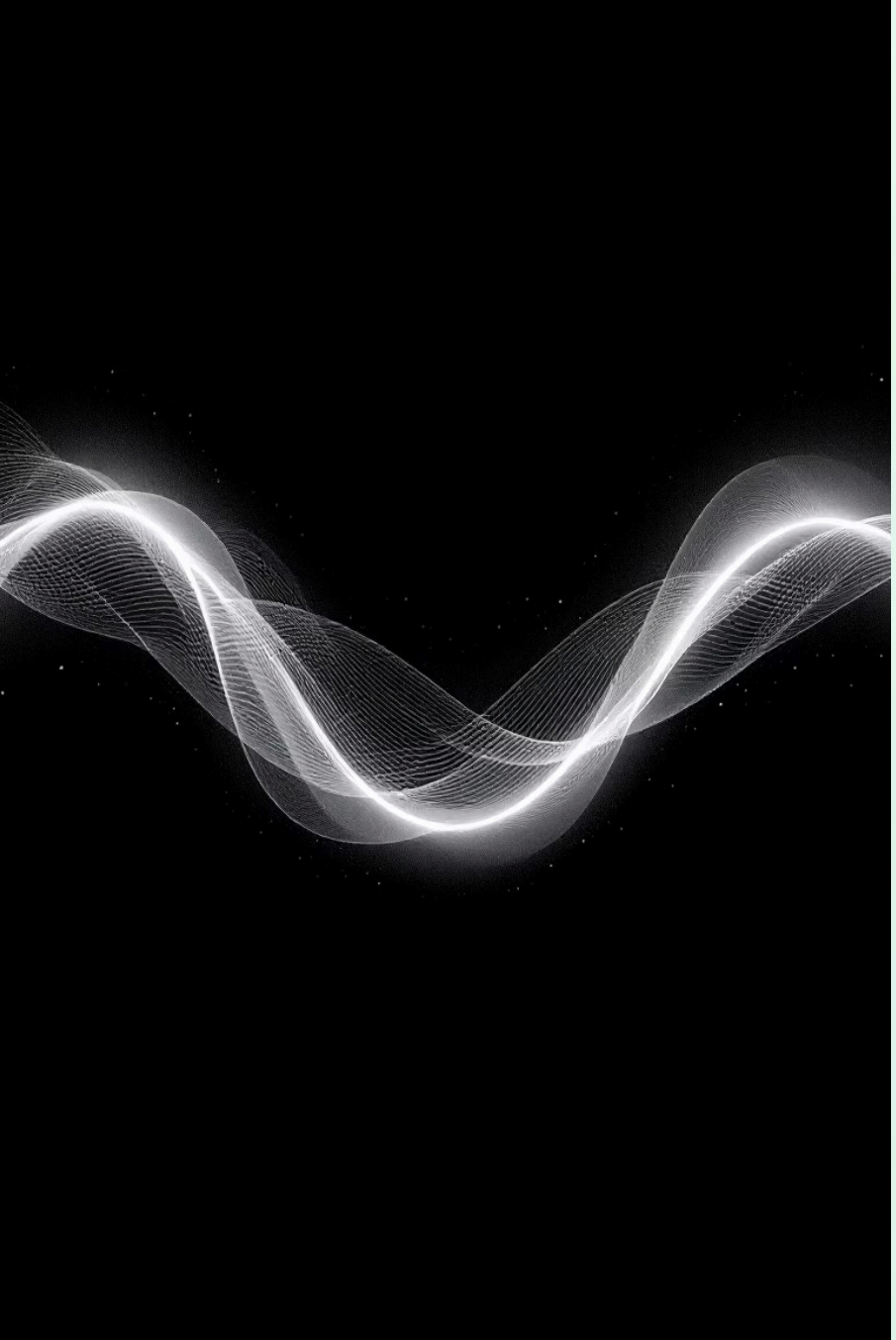
Logistic Regression

A powerful supervised machine learning algorithm for binary classification problems





What is Logistic Regression?

Classification Algorithm

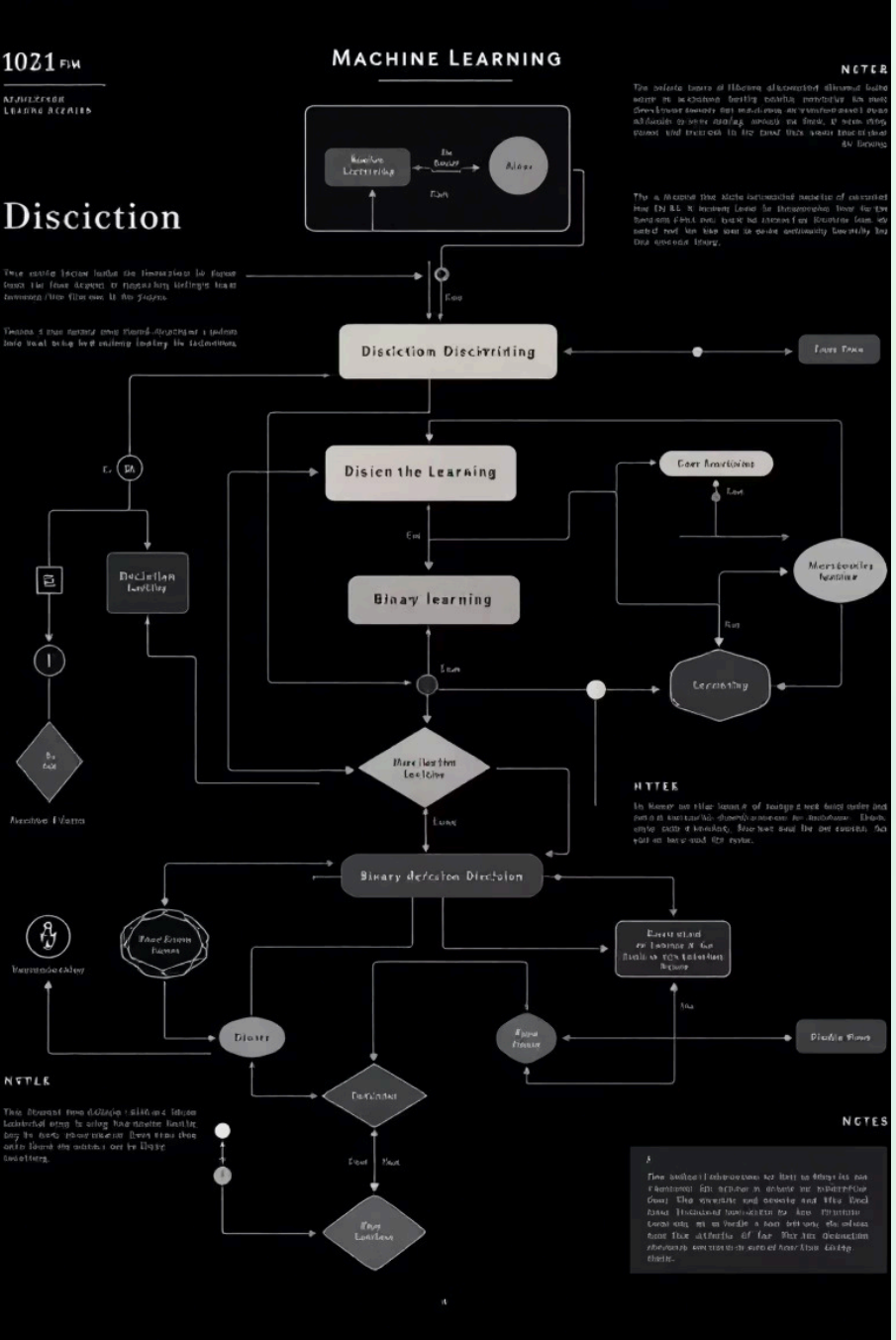
Supervised learning method designed for binary outcomes like Yes/No, 0/1, Pass/Fail.

Probability Prediction

Predicts the likelihood of an event occurring, not just the category.

Sigmoid Function

Uses the logistic function to transform outputs between 0 and 1.



Key Properties

Data Requirements

- Dependent variable must be binary categorical
- Produces binary output (0 or 1)
- No linearity assumption needed

Core Mechanisms

- Maximum Likelihood Estimation (MLE) determines category
- Odds Ratio measures predictor-outcome relationship
- Sigmoid function ensures 0-1 output range

Understanding the Confusion Matrix

True Positive (TP)

Correctly predicted positive cases

True Negative (TN)

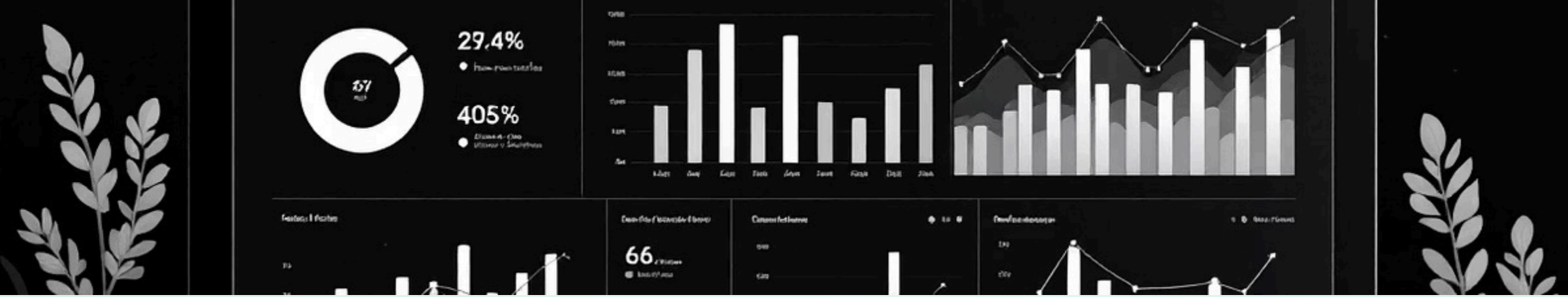
Correctly predicted negative cases

False Positive (FP)

Type I Error: Predicted positive when actually negative

False Negative (FN)

Type II Error: Predicted negative when actually positive



Evaluation Metrics

01

Accuracy

Overall correctness: (TP + TN) / Total predictions

02

Precision

How many predicted positives are actually positive: TP / (TP + FP)

03

Recall (Sensitivity)

How many actual positives were caught: TP / (TP + FN)

04

F1 Score

Harmonic mean balancing precision and recall

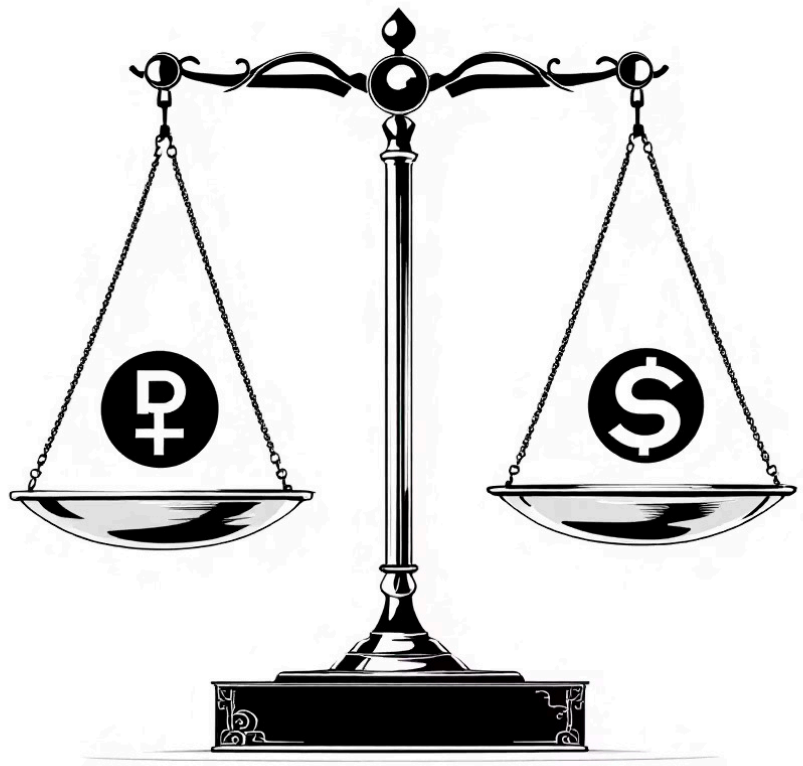
F1 Score

Balancing Precision and Recall

> "How smartly your model catches positives without making too many mistakes."

The F1 Score measures how well your model balances being right (precision) and catching all real cases (recall).

Higher F1 = Better balanced performance



Encoding Categorical Data

Two essential techniques for converting categorical variables into numerical format

Label Encoding

Assigns sequential numbers (0, 1, 2...) to categories. Best for binary or ordinal data.

One-Hot Encoding

Creates separate binary columns for each category. Ideal for 3+ unordered categories.

Label Encoding

When to Use

- Binary categories (2 options)
- Ordinal data with natural order

How It Works

Assigns sequential numbers to each category

Examples

Original	Encoded
Yes / No	1 / 0
Male / Female	1 / 0
Low / Medium / High	0 / 1 / 2

One-Hot Encoding

When to Use

- 3 or more categories
- No natural ordering
- Nominal data

How It Works

Creates separate binary columns for each unique category

Example: Colors

Color	Red	Blue	Green
Red	1	0	0
Blue	0	1	0
Green	0	0	1

Key Takeaways

Binary Classification

Logistic regression excels at predicting binary outcomes using probability

Evaluation Matters

Use confusion matrix and F1 score to assess model performance

Proper Encoding

Choose label or one-hot encoding based on your data structure