Logistic Regression

Metric	Value
Precision	0.8773
Recall	0.8438
F-measure	0.8602
Accuracy	0.8969

Naive Bayes

Metric	Value
Precision	0.6288
Recall	0.9792
F-measure	0.7658
Accuracy	0.775

Why high recall but lower precision?

1. Many "spam-indicator" words (e.g., *free*, *money*, *credit*) sometimes appear in legitimate emails; Naive Bayes treats each word's evidence independently, so several weak spam signals can outweigh a few strong ham signals, resulting in false positives.
2. Words like *'free*' and *'credit*' are correlated; ignoring that correlation exaggerates the joint likelihood of spam, inflating recall at the expense of precision.
3. Logistic regression sacrifices a little recall in exchange for far fewer false positives, yielding much better precision and overall Accuracy.

Decision Tree

Metric	Value
Precision	0.8676
Recall	0.9219
F-measure	0.8939
Accuracy	0.9178

A high **recall (0.9219)** indicates that the tree effectively captures most spam, while **precision (0.8676)** reflects a modest false-positive rate, typical for ID3's greedy splits, which maximize local information gain without global optimization.
