

Introduction

Machine learning systems increasingly influence high-impact decisions across finance, healthcare, marketing, risk management, and compliance. In these contexts, evaluation metrics are not merely technical artifacts; they function as decision enablers, governance instruments, and accountability mechanisms.

Among these metrics, the Area Under the Receiver Operating Characteristic Curve (AUC) has achieved near-universal adoption. It is widely reported, frequently optimized, and often treated—explicitly or implicitly—as a proxy for model quality and business performance.

This document challenges that practice.

Central thesis. AUC is a mathematically valid ranking diagnostic that is routinely misapplied as a decision metric. This mismatch introduces systematic risk into model selection, deployment, and monitoring processes.

The purpose of this document is not to explain how to compute AUC, nor to advocate for or against its use in isolation. Instead, it provides a rigorous, decision-oriented analysis of:

- what AUC formally measures,
- what it explicitly ignores by construction,
- why these omissions matter in real systems,
- and how AUC can be correctly embedded within a broader evaluation framework.

The document is written for:

- senior data scientists and machine learning engineers,
- analytics leaders and technical managers,
- decision-makers responsible for approving or governing ML systems.

It assumes familiarity with binary classification, statistical evaluation, and production constraints. Readers seeking introductory explanations or metric tutorials are intentionally not the target audience.

Throughout the document, emphasis is placed on:

- real-world failure modes,
- lifecycle misalignments,
- and decision-level consequences.

Mathematical formalism is used where it adds precision; practical implications are emphasized where they affect outcomes.

The ultimate objective is to reframe evaluation from a reporting activity into a decision-support discipline—one in which metrics serve objectives, rather than replace them.