RADemics

Model Evaluation Metrics for Performance Assessment in Decision Support Systems

Mr. Robin Stanley Nadar, Dr. Anshul Dube, Prof. V. Prasathkumar
THAKUR SHYAMNARAYAN ENGINEERING COLLEGE, IFTM UNIVERSITY, JANSONS INSTITUTE OF TECHNOLOGY

Model Evaluation Metrics for Performance Assessment in Decision Support Systems

[1]Robin Stanley Nadar, Assistant Professor, Department of Electronics & Computer Engineering, Thakur Shyamnarayan Engineering College, Kandivali East, Mumbai-400101. robinnadar@tsecmumbai.in

[2]Anshul Dube, Assistant Professor, Department of Mathematics, IFTM University, Lodhipur Rajput, Moradabad, Uttar Pradesh -244102. anshuldubey8063@gmail.com

[3]V. Prasathkumar, Assistant Professor, Department of Computer Science and Engineering, Jansons Institute of Technology, Karumathampatti, Coimbatore - 641 -659. prasathkumartvb@gmail.com

Abstract

The growing complexity of intelligent decision support systems necessitates rigorous evaluation frameworks capable of assessing predictive accuracy, decision quality, operational efficiency, and model reliability. This book chapter, Model Evaluation Metrics for Performance Assessment in Decision Support Systems, presents a comprehensive examination of quantitative and qualitative methods for evaluating intelligent models across diverse decision-making environments. The discussion encompasses classification and regression metrics, ranking and recommendation measures, residual analysis, hit rate, coverage, and utility-based evaluation of competing decision alternatives. Particular emphasis is placed on benchmarking advanced models against baseline decision systems, examining cross-validation strategies, and assessing robustness, generalization, fairness, explainability, and computational efficiency. The chapter also explores statistical validation, uncertainty estimation, sensitivity analysis, and continuous performance monitoring to address challenges associated with data variability, model instability, and changing operational conditions. Integrating predictive assessment with decision-oriented evaluation provides a structured foundation for determining whether computational models deliver reliable, interpretable, and practically valuable outcomes. The proposed evaluation perspective supports informed model selection, transparent performance comparison, and responsible deployment across healthcare, manufacturing, finance, and other application domains. By connecting technical performance indicators with operational objectives, this chapter establishes a comprehensive conceptual foundation for strengthening the effectiveness and trustworthiness of intelligent decision support systems.

Keywords: Decision Support Systems, Model Evaluation Metrics, Predictive Performance, Cross-Validation, Decision Quality, Model Reliability

Introduction

The advancement of artificial intelligence, machine learning, and predictive analytics has transformed the development of intelligent decision support systems across diverse application domains [1]. These systems assist decision-makers in interpreting complex information,

identifying meaningful patterns, forecasting future outcomes, and selecting appropriate actions under uncertain conditions [2]. Their effectiveness depends not merely on computational sophistication but also on the reliability and practical relevance of their outputs [3]. Model evaluation therefore represents a fundamental component of intelligent system development, providing structured methods for examining predictive accuracy, decision consistency, operational efficiency, and generalization capability [4]. A comprehensive evaluation framework supports informed model selection and strengthens confidence in computationally assisted decision-making [5].

The increasing diversity of decision support applications demands evaluation metrics suited to different analytical objectives and operational requirements [6]. Classification models require measures such as accuracy, precision, recall, sensitivity, specificity, and F1-score, while regression models depend on prediction error measures and residual diagnostics [7]. Recommendation systems require ranking quality, hit rate, and coverage assessments to determine the relevance and breadth of suggested alternatives [8]. Decision-oriented applications also require utility-based measures that account for operational costs, expected benefits, risk exposure, and competing stakeholder priorities [9]. Selecting metrics according to the underlying decision problem prevents misleading interpretations and provides a more meaningful assessment of model performance within practical decision-making environments [10].

Reliable model assessment requires more than measuring predictive performance on a single dataset [11]. Variations in data distribution, class imbalance, measurement errors, and environmental conditions can influence evaluation outcomes and reduce the reliability of reported results [12]. Benchmarking against baseline decision systems establishes reference performance levels for determining the practical contribution of advanced models [13]. Cross-validation, independent testing, statistical significance assessment, and sensitivity analysis provide complementary mechanisms for examining predictive stability and generalization [14]. Residual analysis further reveals systematic errors, nonlinear relationships, and unexplained variability. Integrating these techniques creates a stronger methodological foundation for distinguishing genuine model improvements from performance differences caused by sampling variation or inappropriate experimental design [15].