Elements of Machine Learning and Data Science

Part III: AI / Empirical Analysis & Performance Evaluation — Exam Notes (Living Document)

Emir Pisirici

January 29, 2026

Exam likelihood: High (overall AI / Empirical & Evaluation part)

This document is structured to match the lecture topics exactly and is designed for adding **exam-style notes**, **common traps**, and **visual summaries**.

Contents

1 Data Quality and Preprocessing

1.1 Introduction

1.2 Missing Values

1.3 Outliers

1.4 Transformation & Normalization

1.5 Reduction

2 Responsible Data Science

2.1 Part I: Introduction to RDS

2.2 Part II: Confidentiality

2.2.1 Risks

2.2.2 Encryption

2.2.3 Anonymization

2.2.4 Quasi-identifiers

2.2.5 K-Anonymity

2.2.6 L-Diversity

2.2.7 T-Closeness

2.3 Part III: Fairness

2.3.1 Fairness measures

2.3.2 Itemsets / association rules revisited

2.3.3 Effect (rule / outcome)

2.3.4 Making decision trees fair

3 Evaluation

3.1 Key Questions for Evaluation

3.1.1 How good is an ML model?

3.1.2 How good could an ML model be?

4 Performance Optimization

4.1 Hyperparameter Optimization

4.2 Model Selection