Understanding the AI Development Workflow

A formal report prepared for PLP Academy

Author: Amanuel Alemu Zewdu

Date: October 31, 2025

This report articulates a comprehensive development workflow for artificial intelligence projects, illustrating concepts with two applied examples: student dropout prediction and hospital readmission prediction. The report follows a structured approach consistent with CRISP-DM and emphasizes ethical, technical and regulatory considerations.

Part 1: Short Answer Questions. The initial phase defines the problem, objectives and stakeholders. For example, a student dropout prediction task seeks to identify at-risk students to enable timely intervention. Relevant objectives include identification of high-risk students, enabling advisor intervention and improving retention metrics. Key performance indicators are chosen according to the problem context; an appropriate KPI is the model F1-score for imbalanced classification contexts.

Data collection and preprocessing include selection of data sources (e.g. academic records, LMS logs), detection and mitigation of biases (for instance socioeconomic bias), and preprocessing steps such as missing-value imputation, normalization and categorical encoding. Model selection is justified by problem properties; models such as Random Forest are appropriate for tabular data and provide feature importance metrics. Evaluation metrics such as F1-score and ROC-AUC are essential for imbalanced tasks. Deployment challenges include scalability and ongoing monitoring for concept drift.

Part 2: Case Study — Hospital Readmission Prediction

Problem Scope: The objective is to predict risk of readmission within thirty days of discharge. Stakeholders include hospital management, clinicians and patients. Data sources include Electronic Health Records, laboratory results and demographic data. Ethical concerns emphasize patient privacy and fairness across demographic groups.

Data Strategy: A robust pipeline ingests EHR records, performs imputation for missing values, normalizes continuous features and encodes categorical variables. Feature engineering may include number of previous admissions, average length of stay and aggregated comorbidity indices. The model should be evaluated on a held-out test set and validated with cross-validation when feasible.

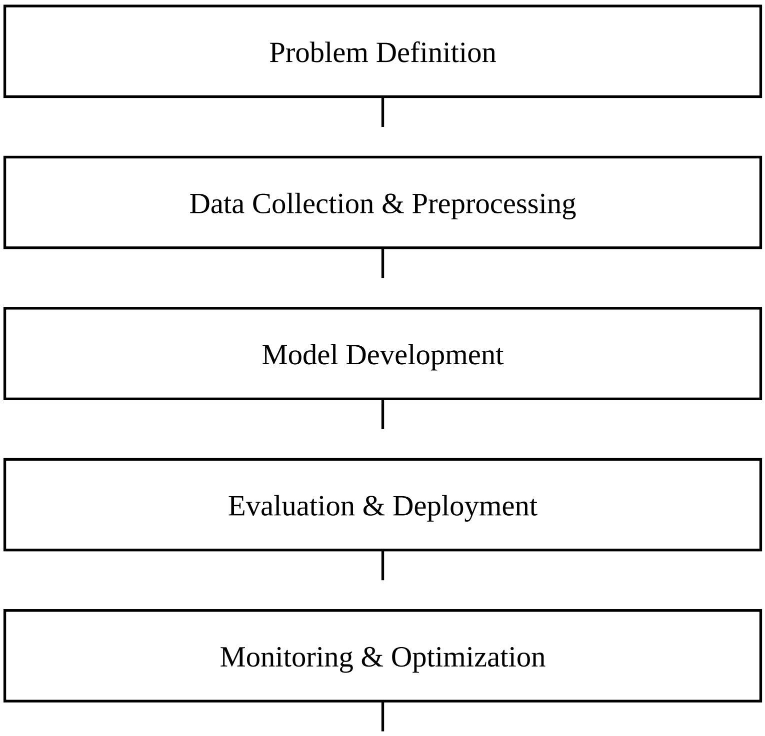
Model Development: For interpretable clinical decision support, logistic regression may be chosen. A hypothetical confusion matrix for 100 test samples: True Positives = 35, False Negatives = 15, False Positives = 10, True Negatives = 40. From these values, precision = 0.78 and recall = 0.70. Post-deployment considerations include monitoring for concept drift and periodic recalibration.

Part 3: Critical Thinking — Ethics, Bias, and Trade-offs

Biased training data may lead to systematic errors, disadvantaging particular patient groups and producing inequitable clinical outcomes. Mitigation strategies include reweighting, resampling and performing fairness audits. Model selection requires balancing interpretability and predictive performance; in healthcare, interpretability often has heightened importance due to clinical accountability.

Part 4: Reflection and Workflow Diagram

Reflection: The most challenging aspect is ensuring data quality and mitigating biases, as these aspects profoundly influence model validity. With additional resources, broader and more diverse datasets and automated experimentation infrastructure would be pursued to enhance reliability and reproducibility.



References

- Wirth, R. & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining.

- European Commission (2021). Ethics Guidelines for Trustworthy AI.

- Suresh, H. & Guttag, J. (2019). A Framework for Understanding Sources of Harm throughout the Machine Learni