

Data Readiness Checklist

Purpose and relevance

- ☐ The prediction or task target is clearly defined.
- ☐ Data represents the intended users, time period and environment.
- ☐ Inputs will be available when the system is actually used.

Access and governance

- ☐ The data owner is identified.
- ☐ Collection and use have an appropriate legal and organizational basis.
- ☐ Consent, licensing and retention conditions are documented.
- ☐ Sensitive data is identified and protected.

Quality

- ☐ Missing values and invalid records are measured.
- ☐ Duplicate records and inconsistent formats are handled.
- ☐ Labels have a clear definition and quality review.
- ☐ Class imbalance and rare but important cases are understood.
- ☐ Potential proxy variables and historical bias are reviewed.

Evaluation readiness

- ☐ Training, validation and test boundaries prevent leakage.
- ☐ Time-based data uses an appropriate chronological split.
- ☐ Important groups and operating conditions appear in test data.
- ☐ A future data-quality monitoring owner is assigned.

Decision

- Ready / conditionally ready / not ready:
- Required improvements:
- Owner and review date: