Prof PMedia

DATA IS A TEAM SPORT

Roles & Responsibilities of a Data Team

Decision Maker	Domain Expert	Data Analyst	Data Engineer	ML Engineer	Evaluation Engineer
Set business objectives and success metrics Approve project progression through milestones Bridge technical and business perspectives	Provide in-depth understanding of problem space Ensure solutions meet real-world needs Assist with interpreting and applying results	Perform exploratory data analysis to derive insights Communicate trends through statistical summaries and visualizations Skills in programming, statistics, data wrangling	Build and maintain data infrastructure and pipelines Extract, transform, load data from diverse sources Expertise in databases, distributed systems, ETL	Select, train, evaluate, and tune machine learning models Specialized knowledge in ML theory and algorithms Leverage math, stats, and software engineering skills	Design model performance tests aligned to success metrics Monitor production systems for drift, degradation, and bias Focus on model reliability, accuracy, and compliance

WHAT IS A DEVELOPMENT FRAMEWORK GOOD FOR?

Cross Industry Standard Process for Data Mining (CRISP-DM)

Structured Approach	Colaboration	Iteration	Predictability	Flexibiliy	Focus
Defines a clear, methodical sequence of steps to guide the end-to-end data value creation.	Clearly defines roles & responsibility to coordinate work across teams with diverse skills.	Allows folding back to previous steps when needed to refine the solution.	Following consistent standard methodology improves predictability.	The methodology can be tailored to fit different project needs and constraints.	Business goals anchor the process to ensure work stays aligned with delivering value.

CRISP-DM

"THE PROCESS"

Business Goal
Define metrics for success.

Feedback: Data doesn't meet business objectives.

Assess Data
Explore data quality, derive insights, identify issues.

Prepare Data
Construct training dataset by cleaning and transforming data.

Feedback: May need to consider adding additional features to improve model performance

Modeling
Select, train, tune and validate performance.

Feedback: Retrain the model on newer data when performance degrades.

Evaluation
Test model with holdout test data

No Go
fails to meet success metrics

Go
meets success criteria

Deployment
Integrate into business processes and monitor for drift

OUR DATA SOURCES