Michele Usuelli Microsoft Enterprise Services

Road to agile data science in 12 steps

Target audience

Are you a data scientist?

Are you working with data scientists?

Philosophy of the 12 steps



AGILE FOR DATA SCIENCE WHY/WHAT WITH EXAMPLES ON HOW 12-WEEKS' AGILE CYCLE



Understand why, where and how data science can help

Objectives of the path

Ĵ

Design and deliver solutions



Choose the tools and techniques



Have a checklist about key aspects

Key aspects

Broader context: people, value, scope

Delivery methodology: rhythm, communication

Toolbox: software and techniques



Steps of the path

BROADER CONTEXT	DELIVERY METHODOLOGY	TOOLBOX
1 Proceed in small steps	5 Shape backlog items	9 Adopt tools for communication
2 Establish roles and collaboration	6 Define user story lifecycle	10 Adopt tools for coding
3 Mitigate gaps	7 Shape user stories	11 Adopt tools for deployment
4 Establish delivery rhythm	8 Have fully-fledged user stories	12 Define testing strategy

Broader context

Steps 1-4

1 Proceed in small steps





Questions about data science

What are Data Science and Artificial Intelligence?

Is there science feasible?

Is data science the right solution?

What is machine learning?

What is a predictive model?

What techniques to use?

What tools to use?

Questions about data science projects

People

- Who will benefit?
- Who will act upon the results?
- Who can provide feedback?

Data

- Available?
- Reliable?
- Traceable?

Outcome

- Accuracy?
- Technical feasibility?
- Success assessment?

Small steps to answer the questions





2 Establish roles and collaboration



Agile Manifesto



Individuals and interactions over processes and tools



Working software over comprehensive documentation



Customer collaboration over contract negotiation



Responding to change over following a plan



"Business people and developers must work together daily throughout the project"

Principles of team collaboration

FIT BROADER PROGRAMME APPROACH UNDERSTAND INDIVIDUALS RATHER THAN ROLES ESTABLISH KEY PLAYERS AND COLLABORATION

Data scientist activities within the team

Product owner	identify personas and actions	
	explain how insights change the process	
Scrum master	establish delivery rhythm	
	communicate progress	
Tech team	define data to onboard	
	define product deployment	



Data scientist activities outside the team

Stakeholder	understand company strategyexplain value
Solution user	identify actionunderstand usage
Business analyst	explain solutionunderstand impact

3 Mitigate gaps

Common gaps

strategy vs use cases

- understand strategic goal
- define tactical goal
- define analytical objective

solution vs value

- understand target value
- define required information
- define how to produce the information

science vs engineering

- understand deployment requirements
- define activities
- define team collaboration

Solution vs value: example of value explanation

Completeness

- More data
- Complex patterns

Performance

- Measure
 - accuracy
- Improve accuracy

Transparency

- Explain logic
- Explain results

4 Establish delivery rhythm



targeted checkpoints

time-boxed activities

smaller transaction

Delivery lifecycle



Data science

Architecture

Delivery methodology

Steps 5-8

5 Adopt tools for communication

Scrum's pillars of empiricism



TRANSPARENCY

INSPECTION

ADOPTION

Key concepts to agree on





WHY/WHAT IS ARTIFICIAL INTELLIGENCE

MACHINE LEARNING VS OTHER APPROACHES



WHAT IS A PREDICTIVE MODEL

Different vehicles... Depending on the distance



Visualization of results



Operator

Actions Explanation



Analyst

Insights Accuracy



Stakeholder

KPIs

Trends

"Distance" for Al techniques



Data

Size Complexity 0

Knowledge

Domain

Data

Why/what is artificial intelligence



Completeness

Use more data Identify relevant information Dig deeper into the data



Transparency

Measure forecasting error Explain how the forecast is produced Explain how the data is used



What is machine learning



Tools for communication

Documentation on the wiki

- Analytical flow
- Modelling choices
- Insight-driven choices

Dashboards

- Insights
- Error measurement
- Action-related KPI

Reports

- Explanation of techniques
- Interpretation of results

Example: analytical flow



Toolbox

Steps 9-12

6 Shape backlog items

Data science aspects in backlog items epics

 1.actionable results
2.reduced error
3.reduced cost features

 1.analytical objective
2.data dependencies

user stories

1.data

- requirements
- 2.insights to make choices
- 3.modelling

tasks

1.time-boxed discovery2.pre-booked collaboration

7 Define user story lifecycle

Contextualise TDSP process



UNDERSTAND TEAM COLLABORATION **DEFINE ACTORS**

ALIGN TO OBJECTIVES



Solution modelling

• Business discovery

- Analytical discovery
- Technical discovery

Analytical delivery

- Data wrangling
- Analytical discovery
- predictive modelling

Technical delivery

- Wrangling deployment
- Modelling deployment
- Integration and testing

8 Shape user stories

Why user stories

- Persona and goal
- Clarity on outcome
- Focus and simplicity





Attributes of user stories

Definition

- Key user
- Purpose
- Outcome

Acceptance criteria

- Qualitative rather than quantitative
- Technical quality
- Communication

Sizing

- Depending on the category
- Time-boxed tasks

Types of user stories - example

Solution modelling

- Define objective
- High-level choices

Data wrangling

- Define data preparation
- Get the data

Analytical discovery

- Insight-driven choices
- Modelling draft

Data modelling

- Full data modelling
- Code ready for production







Communication-based



Qualitative, not quantitative



9 Adopt tools for communication



Communication

- Input from subject-matter experts
- Output to subject-matter experts
- Value communication to the stakeholders





10 Adopt tools for experimentation







Data relevance



Technique experimentation, benchmark and optimization



11 Adopt tools for coding



Coding tools

Coding practices

• Code structuring

Naming convention

Documentation

- Code documentation
- Technical documentation

Tools on Azure

- Machine Learning SDK
- Azure DevOps



12 Adopt tools for deployment



Example of tooling for deployment

Machine Learning Service

Azure DevOps pipelines

MLOPS reference architecture

Tests to consider

Unit tests

Data tests

Integration tests

Path to testing process

01

Create separate testing user stories

02

Include testing as part of the acceptance criteria 03

Design tests as per the acceptance criteria

Path to testing tooling - Python

