

Enabling DevOps for Machine Learning with Azure Pipelines



Luca Milan

@fewbit

luca.milan@gmail.com

Alberto Dallagiacoma

@albertodall

alberto@albertodallagiacoma.it

www.albertodallagiacoma.it

Organizer & sponsors



About us

- Luca Milan
 - Technical Manager @ Yoox Net-A-Porter Group
 - Automation Addicted <3 Code Junkie
- Alberto Dallagiacoma
 - Software Architect R&D @ iSolutions (<https://labs.isolutions.it>)
 - Sviluppatore su .NET dalla versione 1.0
 - TFS dalla versione 2010 (e ora Azure DevOps)
 - PowerShell and F# fan!

Cosa vedrete oggi ?

Come realizzare una strategia di
«**Continuous Deployment**» per il rilascio
di modelli ML applicando i principi
DevOps ed Agile

Luca & Alberto

Agenda

- Machine Learning
- ML.NET Framework
- CI/CD per Machine Learning
- Azure DevOps Pipelines
- Demo: Vediamo il codice in azione 😊



Introduzione

La Sfida

Applicazioni (+) Intelligenti

misura in scala
di exabyte



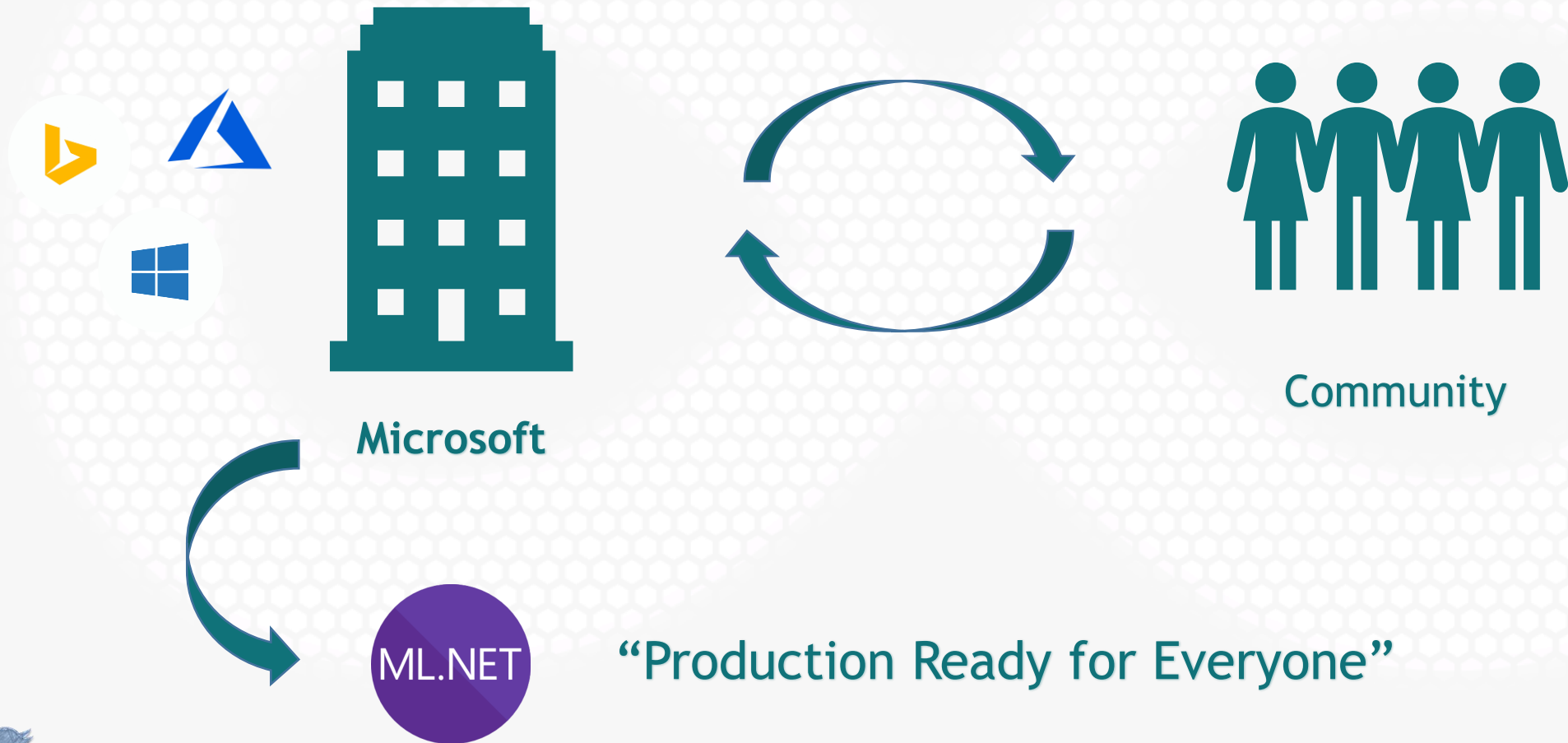
“Big Data”



“Modelli di ML”



Un'opportunità per tanti



Le peculiarità dei progetti di ML

E' collegato alla "Scienza dei Dati", servono "Specialisti" con conoscenze matematiche e statistiche approfondite

Possiede metodologie di sviluppo e strumenti dedicati non sempre allineati con gli approcci "DevOps"

Gli step del ciclo di vita di un progetto di ML sono profondamente differenti da quelli del software tradizionale

I modelli di ML vengono distribuiti come servizi a sé stanti (Black Box) e non sono "integrabili" nel codice delle applicazioni



Come Programmare l'Improgrammabile

Definizione

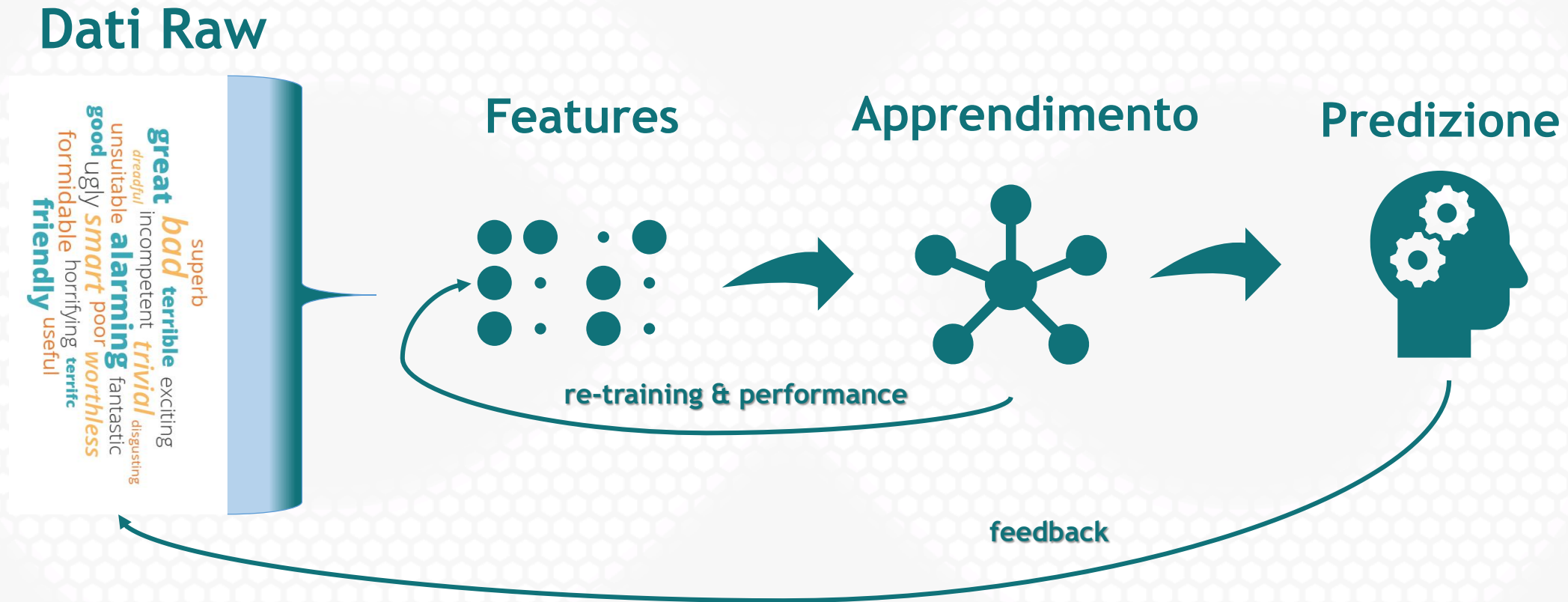
«Una macchina apprende con l'esperienza se la sua performance a svolgere un compito migliora nel corso del tempo dopo averlo svolto più volte»

Tom Michael Mitchell

Il comportamento dell'algoritmo non è pre-programmato, ma è «appreso» dai dati



Machine Learning «End to End»

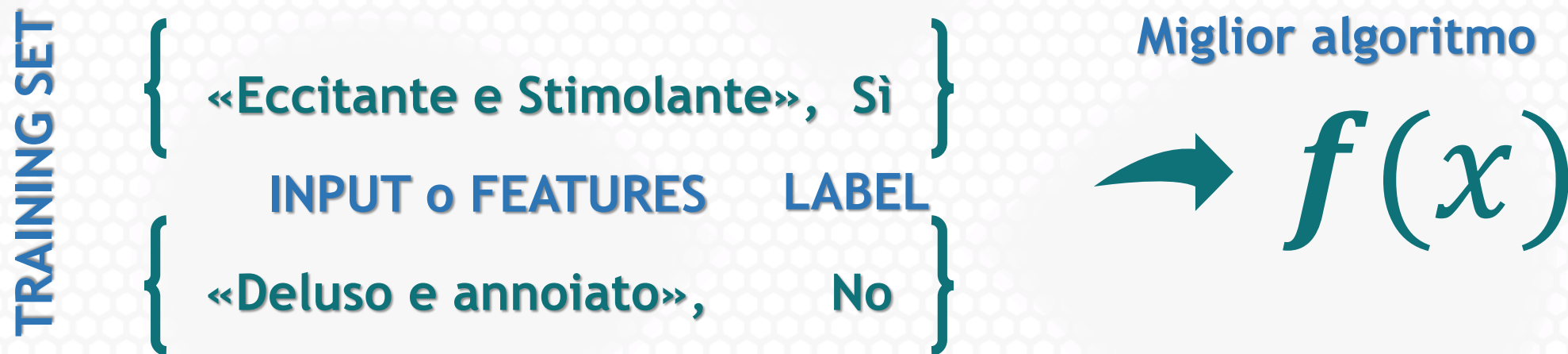


Feature: rappresentazione numerica del dato di input



Come funziona la Classificazione ?

Fase Apprendimento



Fase Predizione

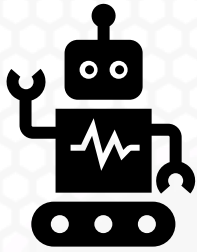
$$f\left(\left\{ \begin{array}{ll} \text{«Bravi, mi ispirate»}, & ? \end{array} \right\}\right) = \text{Sì}$$

Modello (Features) = Label OUTPUT





C# Machine Learning senza Python/R :)



Framework for
Machine Learning



.NET Standard



Open
Source



Cross
Platform

Come usare ML.NET ?

Api & Tools

API

C# / F#
(by code)



CLI

Auto-ML engine
(by tool)



UI

VS Model Builder
(by wizard)

Package Manager

.NET CLI

PackageReference

Paket CLI

```
PM> Install-Package Microsoft.ML -Version 1.3.1
```



#DOH19

ML.NET Extensions

MLNET

<https://www.nuget.org/profiles/MLNET>



Microsoft.ML



by: Microsoft dotnetframework MLNET

↓ 374.527 total downloads | ⌚ last updated 3 months ago | 📦 Latest version: 1.3.1 | 🔗 ML.NET ML Machine L...

ML.NET is a cross-platform open-source machine learning framework which makes machine learning accessible to .NET developers.



Microsoft.ML.CpuMath



by: Microsoft dotnetframework MLNET

↓ 330.205 total downloads | ⌚ last updated 3 months ago | 📦 Latest version: 1.3.1 | 🔗 ML.NET ML Machine L...

Microsoft.ML.CpuMath contains optimized math routines for ML.NET.



Microsoft.ML.DataView



by: Microsoft dotnetframework MLNET

↓ 195.407 total downloads | ⌚ last updated 3 months ago | 📦 Latest version: 1.3.1 | 🔗 ML.NET ML Machine L...

Contains the IDataView system which is a set of interfaces and components that provide efficient, compositional processing of schematized data for machine learning and advanced analytics applications.



Microsoft.ML.OnnxRuntime



by: Microsoft dotnetframework Microsoft AI Infra and Tools team
MLNET

↓ 117.835 total downloads | ⌚ last updated 13 days ago | 📦 Latest version: 0.5.1 | 🔗 ONNX ONNX Runtime ...



34

Packages

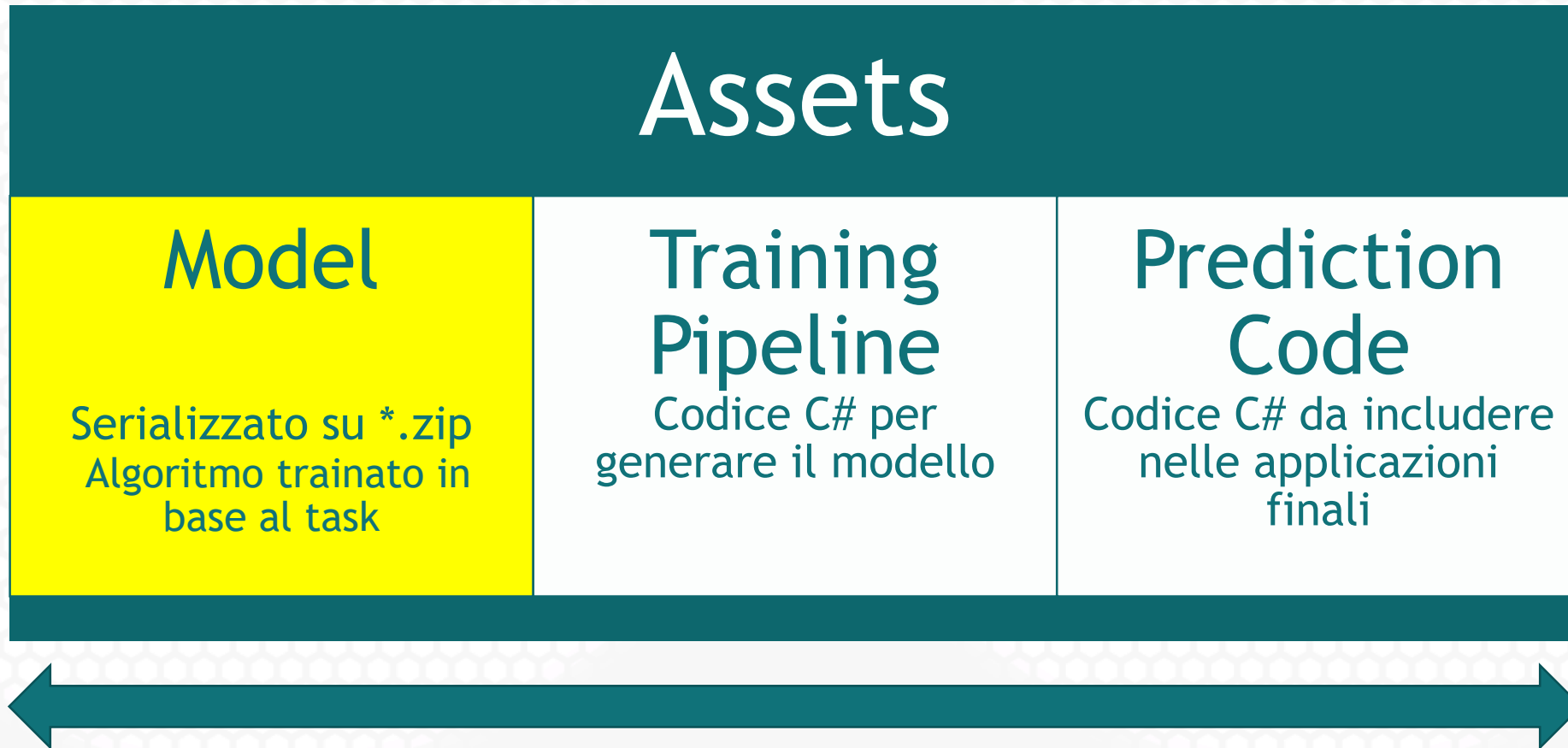
1.645.635

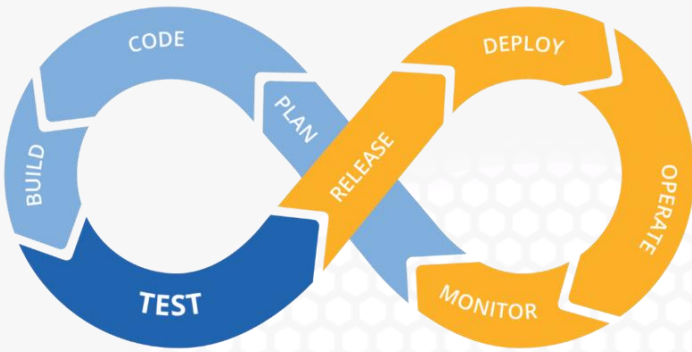
Total downloads of
packages



#DOH19

Cosa produce ML.NET ?





CD4ML

“DevOps is a culture, a movement, a philosophy—A way to bring together the best of software development and IT operations”

Continuous «Delivery»

«Portare in produzione qualsiasi cambiamento del software in maniera rapida, sicura, affidabile e sostenibile»



Continuous Delivery per «ML»

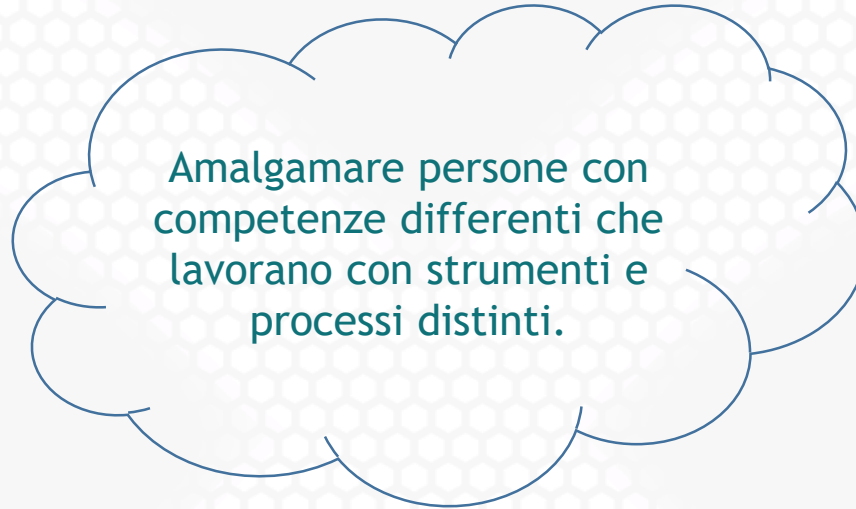


Team cross-funzionali
producono artefatti di ML
tramite un processo
automatizzato e
riproducibile basato su cicli
di sviluppo rapidi e
incrementali

Obiettivi Organizzativi



Data



Operations



Development

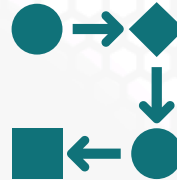
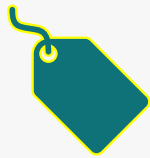


Ciclo di vita di un progetto di ML

Analisi

Apprendimento *

Distribuzione *



Training
Pipelines



Trained
Models



Prediction



Training / Test
Dataset

validation

feedback



#DOH19

Distribuzione del Modello

- Modello come «**allegato**»

Il modello è distribuito insieme all'applicazione ad esempio all'interno container docker.

- Modello come «**servizio**»

Il modello è distribuito come servizio e accessibile all'applicazione tramite chiamate REST o RPC.

- Modello come «**pacchetto**»

Il modello è distribuito in maniera indipendente e l'applicazione a runtime carica ed utilizza il modello.



Approccio GIT-centrico (1)



1. Codice

- ✓ Applicazione
- ✓ Training Pipeline
- ✓ CI/CD Pipelines
- ✓ Provisioning

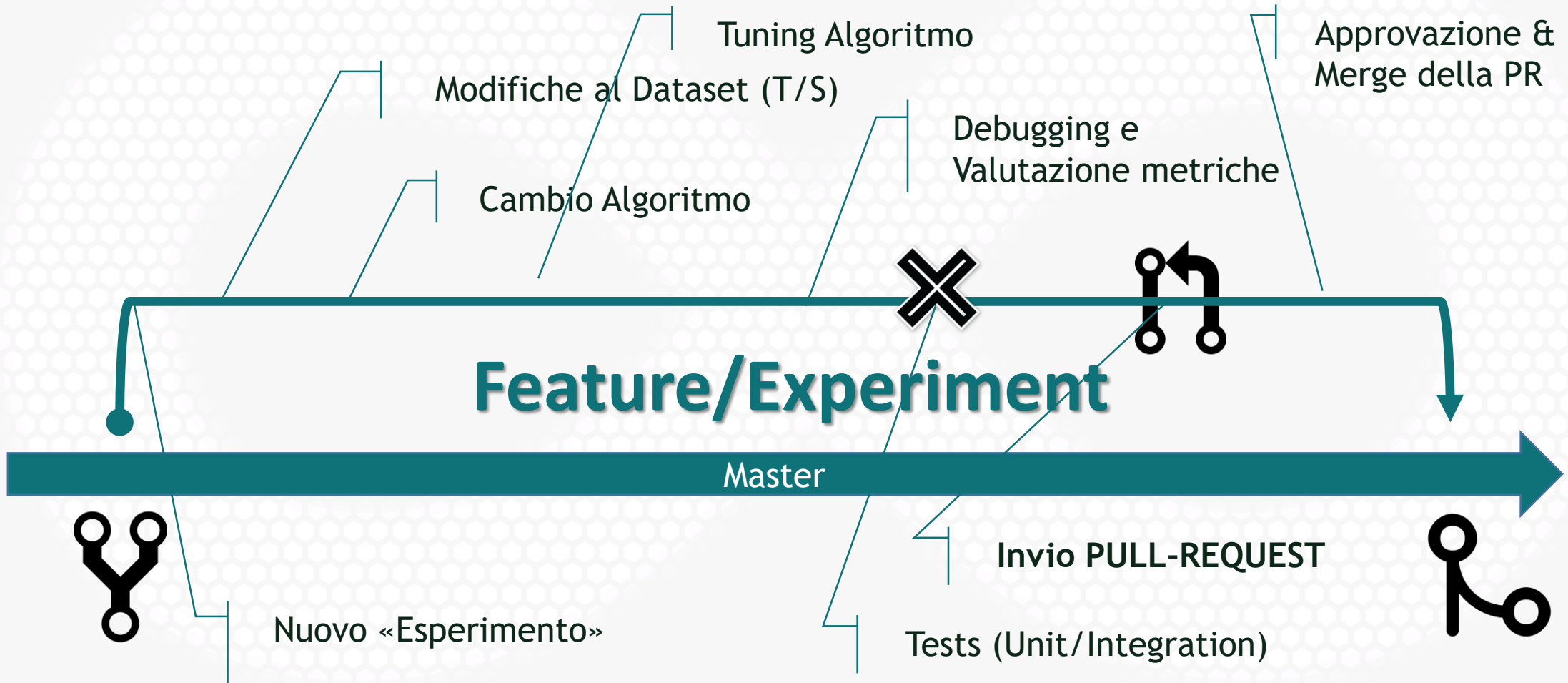
2. Dati

- ✓ Dataset Training
- ✓ Dataset Test

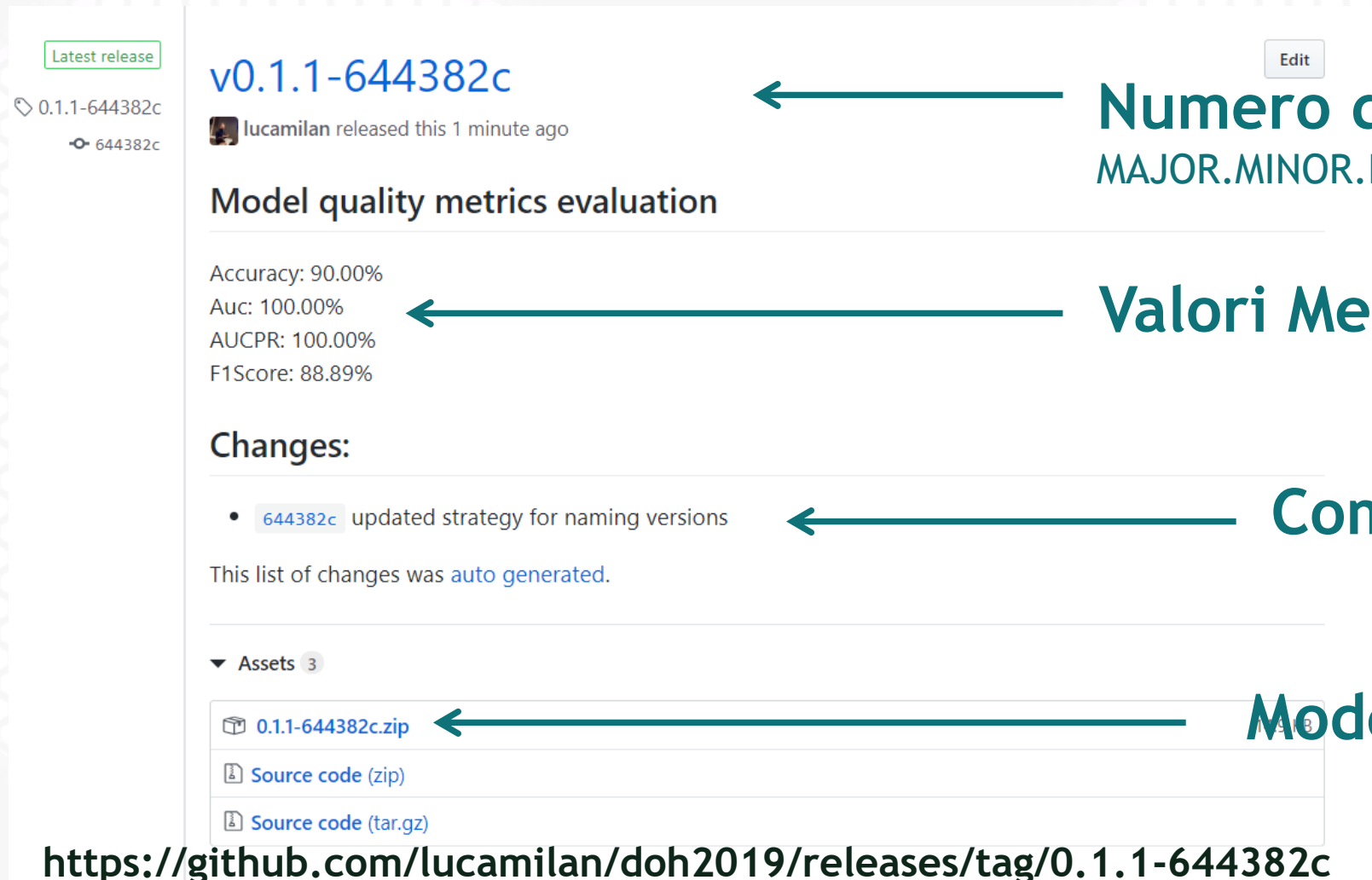
3. Modelli

- ✓ Release
- ✓ Versioning (SemVer)

Approccio GIT-centrico (2)



Approccio GIT-centrico (3)



The screenshot shows a GitHub release page for the repository 'doh2019'. The release is titled 'v0.1.1-644382c' and was released by 'lucamilan' 1 minute ago. The release description is 'Model quality metrics evaluation'. The metrics listed are: Accuracy: 90.00%, Auc: 100.00%, AUCPR: 100.00%, and F1Score: 88.89%. The changes section shows a single commit: '644382c updated strategy for naming versions'. The assets section lists three files: '0.1.1-644382c.zip', 'Source code (zip)', and 'Source code (tar.gz)'. Annotations with arrows point to specific parts of the page: 'Numero di Versione' points to the version number 'v0.1.1-644382c'; 'Valori Metriche' points to the metrics section; 'Commits' points to the commit list; and 'Modello di ML' points to the '0.1.1-644382c.zip' asset.

Latest release

0.1.1-644382c

644382c

v0.1.1-644382c

lucamilan released this 1 minute ago

Numero di Versione
MAJOR.MINOR.PATCH-SHACOMMIT

Model quality metrics evaluation

Accuracy: 90.00%
Auc: 100.00%
AUCPR: 100.00%
F1Score: 88.89%

Valori Metriche

Changes:

- 644382c updated strategy for naming versions

This list of changes was [auto generated](#).

Commits

Assets 3

- 0.1.1-644382c.zip
- Source code (zip)
- Source code (tar.gz)

Modello di ML

<https://github.com/lucamilan/doh2019/releases/tag/0.1.1-644382c>

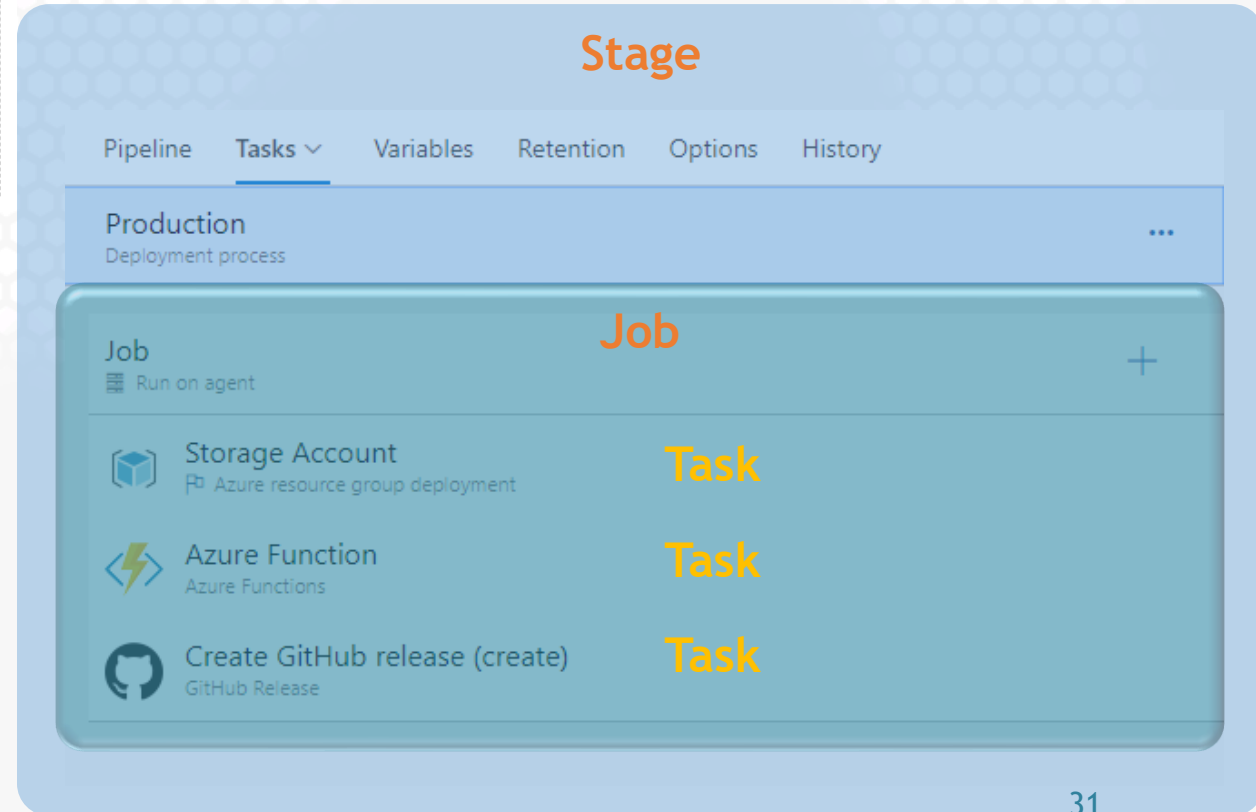
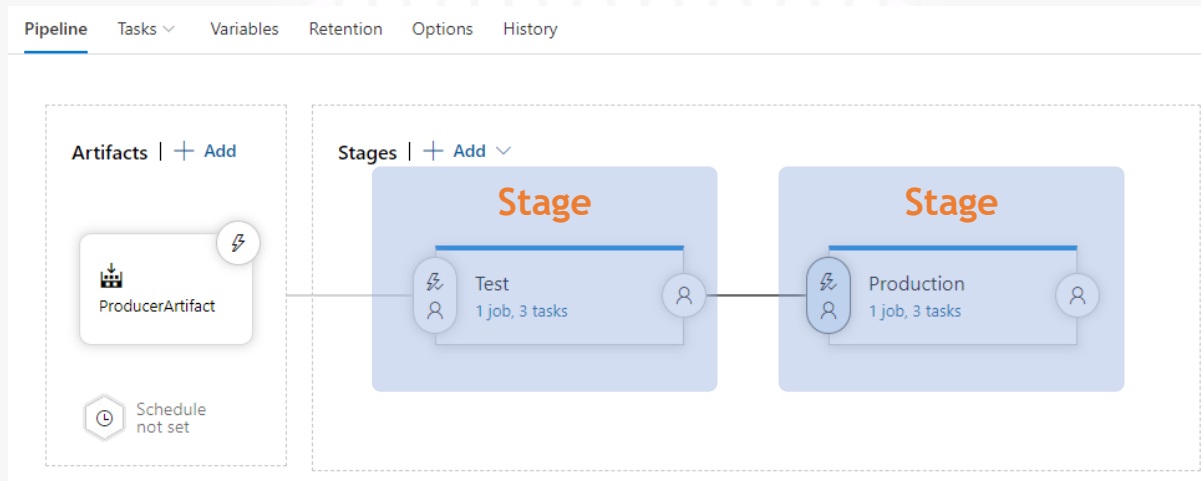


Azure DevOps Pipelines

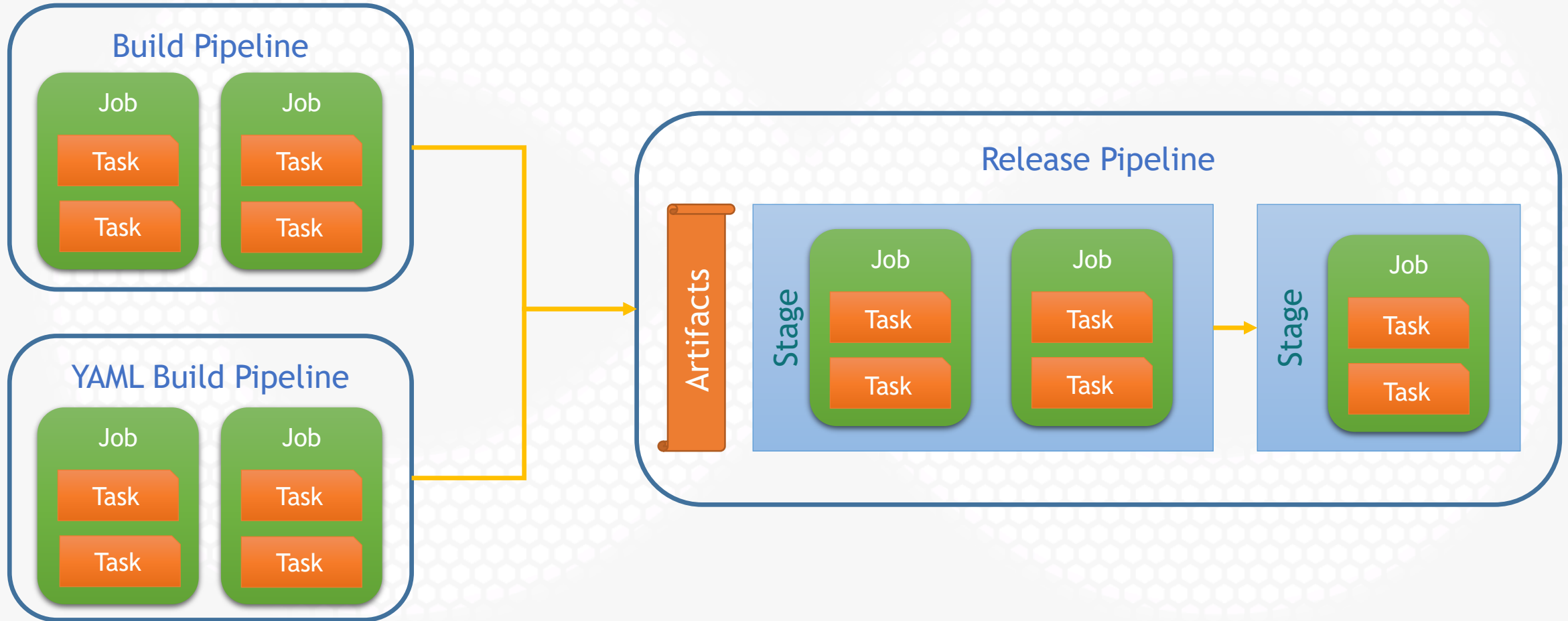
“Classic” Pipeline (Build)

The screenshot displays the 'Classic' Pipeline (Build) interface in Azure DevOps. The top navigation bar includes tabs for 'Tasks', 'Variables', 'Triggers', 'Options', 'Retention', and 'History', along with a 'Save & queue' button. The main area is titled 'Build Pipeline' and shows a pipeline named 'Build pipeline'. The pipeline consists of a 'Get sources' task (labeled 'Job') and a 'Build and test' job (labeled 'Job'). The 'Build and test' job contains three tasks: 'Use .Net Core sdk 2.x', 'dotnet restore', and 'dotnet build'. Each task is labeled 'Task'.

“Classic” Pipeline (Release)



“Classic” Pipeline (Build + Release)



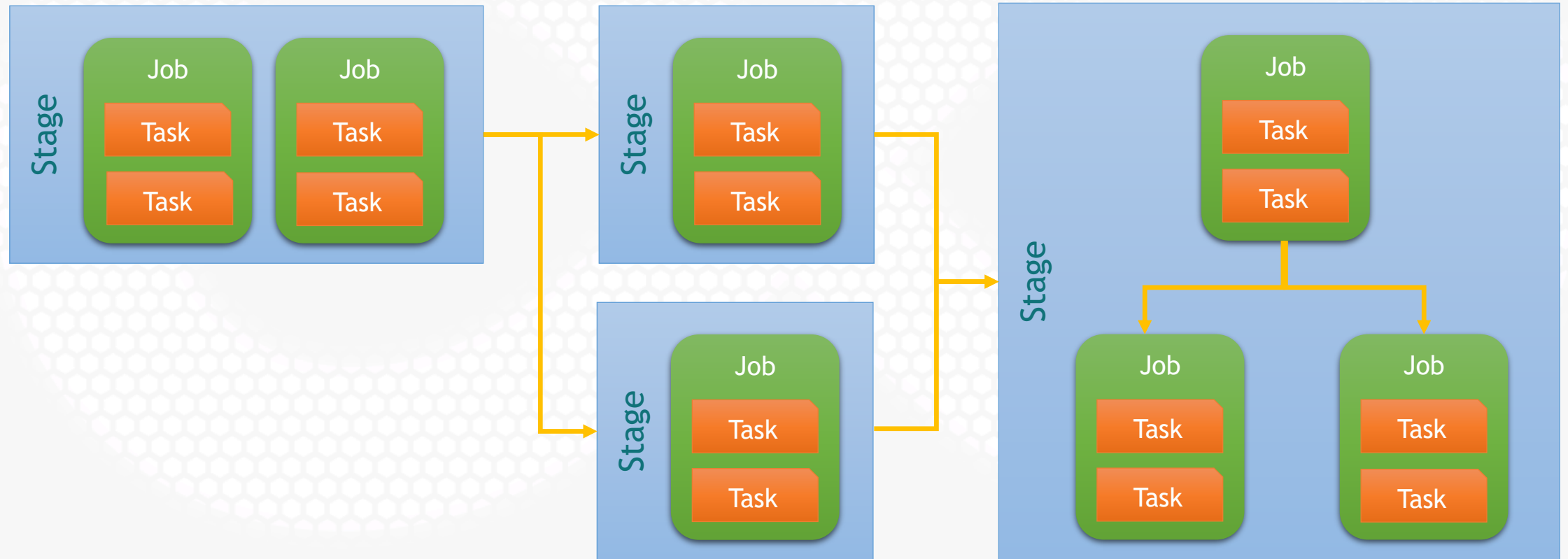
Introducing YAML Pipelines

- Unificazione di *Build* e *Release* in un'unica pipeline
 - Attualmente in preview.
- Fa parte del repository
 - Può evolvere insieme al codice.
 - Può essere composta da uno o più files `.yaml`.
 - E' possibile applicare alla pipeline le stesse *branching policies* del codice.
 - Anche una pipeline può essere soggetta a *Pull Request*.
- Non e' possibile convertire in formato YAML una pipeline esistente
 - Export dei singoli task + edit manuale.
 - Piccolo supporto da parte dell'editor visuale.

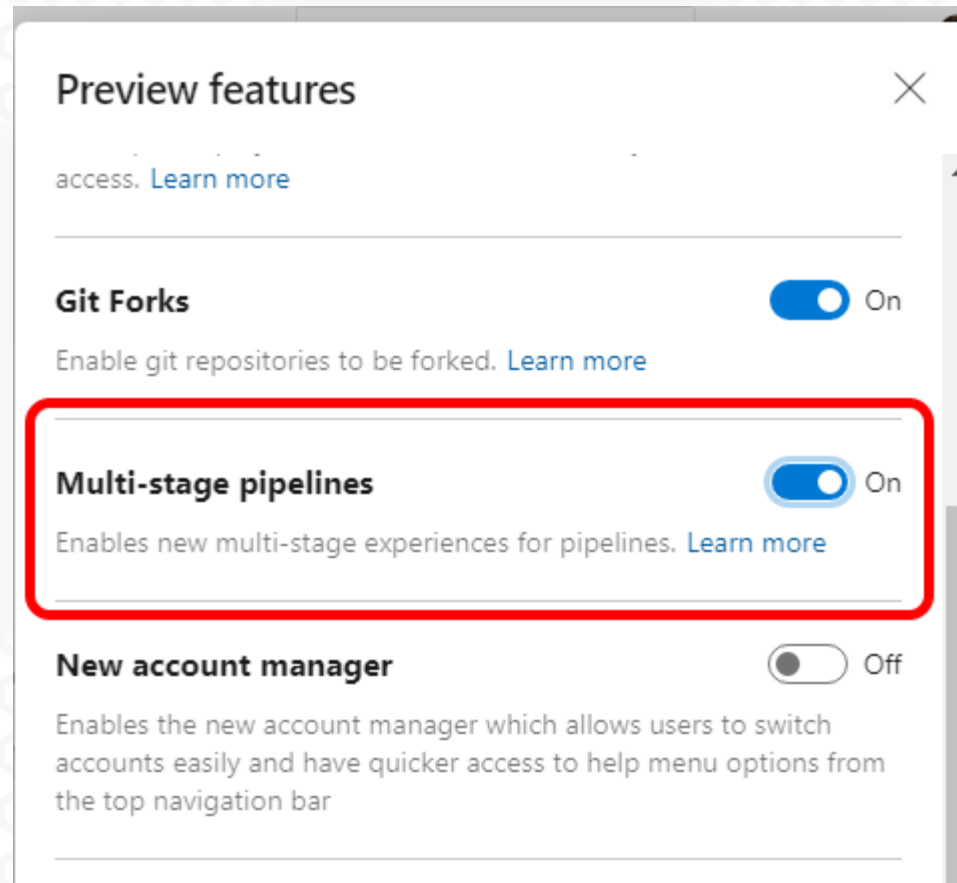


YAML Pipeline

Multi-Stage YAML Pipeline



Abilitare le Multi-Stage Pipelines



Azure Pipelines Tips And Tricks

- Checkout del branch
 - Di default, ogni job esegue il checkout del branch.
 - Può impattare sui tempi di esecuzione in caso di codebase grandi.
 - checkout: none
 - Publish Pipeline Artifacts
 - Artefatti prodotti dalla fase di build disponibili ad altri job/stages della pipeline.
- Don't Repeat Yourself (DRY)
 - Dove possibile, riutilizzare parti di pipeline tramite template.
- Variabili
 - Definire i parametri «globali» della pipeline in un template a parte.
 - Singolo punto dove applicare modifiche.
 - Definire a livello di job le variabili utilizzate da uno specifico job.

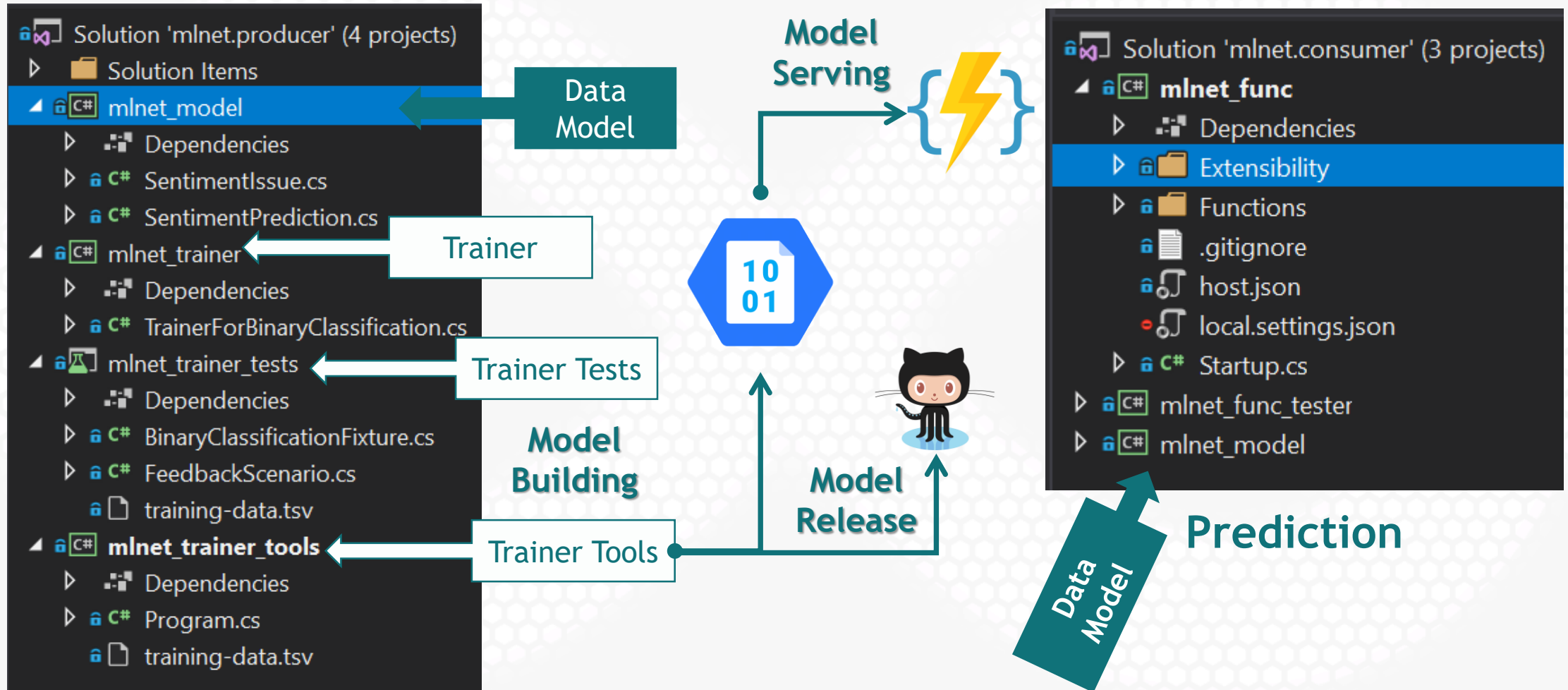


DEMO



Show me the code! 😊

Organizzazione della Solution .NET Core



Training



#DOH19



THANK YOU!



#DOH19