universidade de aveiro

Milestone 2: Elaboration

ML/AI model evaluation platform

Plataforma de avaliação de modelos ML/AI

Course: "Projeto em Informática"
Supervisor: Prof. Mário Luís Pinto Antunes

2023-03-14

Team:

Diogo Oliveira Magalhães	102470
Rafael Fernandes Gonçalves	102534
Leonardo Almeida	102536
Emanuel Gaspar Marques	102565
Pedro Henrique Figueiredo Rodrigues	102778

Context

The need to handle **large amounts of data** and advances in processing technologies have led to the mass development of intelligent systems.

Machine Learning (ML) algorithms are often applied to optimise **multiple real-life scenarios**, leading to cost saving and increased productivity.

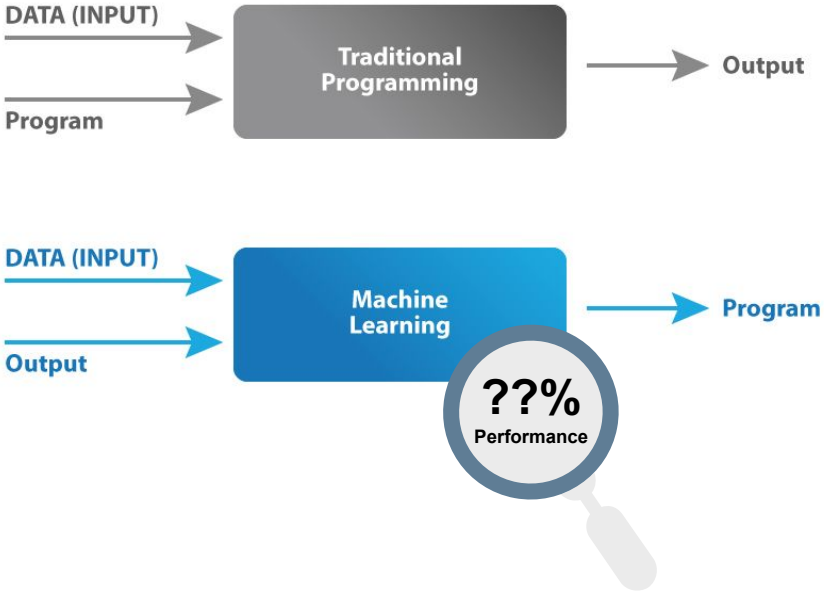
ML/AI examples

ChatGPT

GitHub Copilot

That's why promoting a better ML/AI education is so important!

Our product



MepML

Concept:
ML/AI model evaluation system, based on performance metrics provided by teachers.

Adapted from: Advani, V. (2020). *What is Machine Learning? How Machine Learning Works and future of it?* [online]. GreatLearning. https://www.mygreatlearning.com/blog/what-is-machine-learning

Requirements gathering

Talk to domain experts

Prof. Mário Antunes
Stakeholder

The teacher told us what he wanted from the project, especially from a **teacher's point of view**.

Tópicos de Aprendizagem Automática

Thanks to Prof. Petia Georgieva, we'll use our own TAA lessons as a source of continuous feedback on **students' needs**.

Requirements gathering

Persona and a sample scenario (Teacher)

Name	Daniel Pereira
Age	42 years
Sex	Male
Job title	Associate Professor
Institution	University of Aveiro
Civil status	Married
Location	Ílhavo, Aveiro, Portugal

Scenario: Daniel uploads a new exercise — Daniel opens the web application and logs in. On his dashboard, he sees a button to create a new exercise. He clicks on it, triggering a form to appear. First, Daniel enters a title and the guidelines for the exercise. Then, he uploads 2 files: a training dataset and a test dataset, whose last column contains the expected values. He also specifies some rules: deadline, maximum number of submissions (if any), visibility (public or restricted to a group) and, last but not least, the performance metrics to be used in the automatic evaluation of the students' solutions. Finally, he submits the form and the exercise is created.

The underlined excerpts hint at the actions that were **relevant for identifying the requirements**.

Requirements gathering

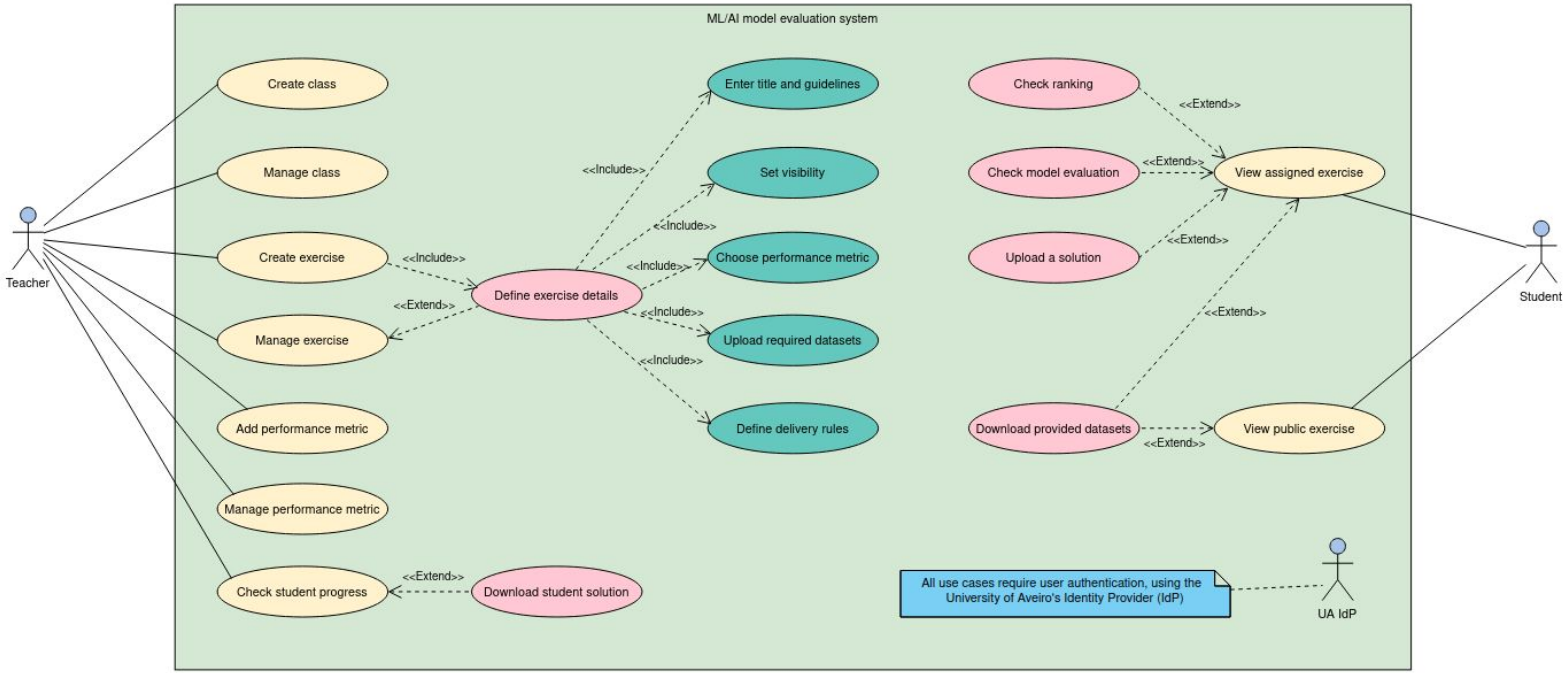
Persona and a sample scenario (Student)

Name	Joana Silva
Age	21 years
Sex	Female
Job title	MSc student - Data Science
Institution	University of Aveiro
Civil status	Single
Location	Sátão, Viseu, Portugal

Scenario: Joana submits a solution to the exercise — Joana is notified that a new exercise was assigned to her. So, she logs into the web application, goes to the exercises page and selects an exercise. Joana notes she's able to download 2 files: a training dataset and an incomplete test dataset, i.e., without the predicted values. After solving the problem, she uploads a solution file and submits it. The platform evaluates the solution and, a few seconds later, displays the results.

The underlined excerpts hint at the actions that were **relevant for identifying the requirements**.

Actors + Use Cases



CRUD operations = **C**REATE + manage (**R**EAD, **U**PDATE and **D**ELETE)

Main Functional requirements

- Student and professor user type
- UA Identity Provider integration

Professors should be able to:

- Group students into classes
- Add new metrics
- Create exercises
- View students' results
- Access students' code

Students should be able to:

- See public exercises
- See assigned exercises
- Download exercise related content
- Upload code and results
- View results table

Non-functional requirements

Availability
Accessible
Responsive

Security
Authentication
Authorization

Maintainability
Modularity
Documentation

Usability
Intuitive
Good performance

State-of-the-Art

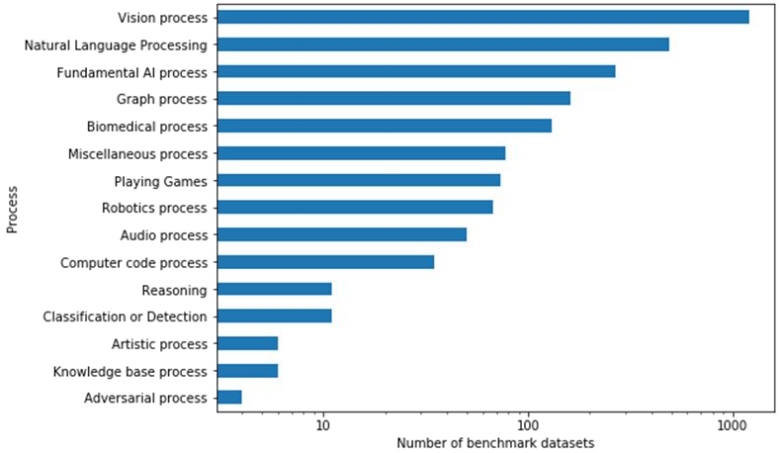
Technics, Technologies, Frameworks and Related Work

Related Work

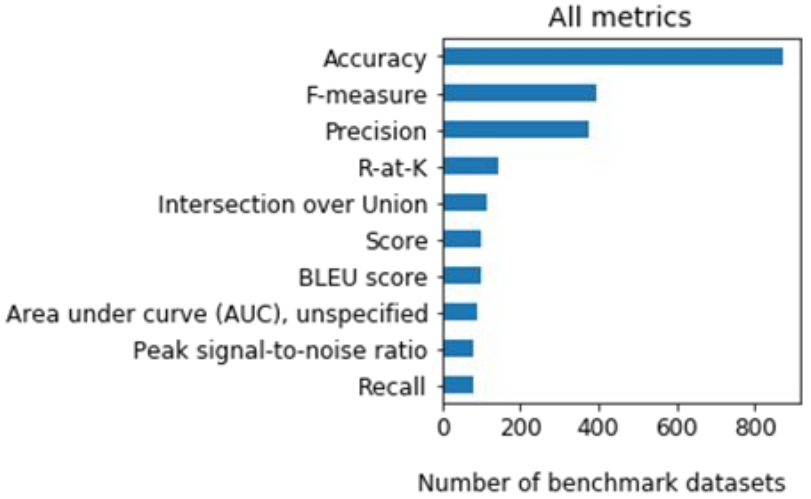
	Kaggle	CodaLab	Weights and Biases	Machine Hack	MepML
Professor can create exercises and students can join them	✓	✓	✗	✗	✓
Define exercise visibility	✓	✓	✗	✗	✓
Exercise Leaderboard	✓	✓	✗	✓	✓
Define exercise deadline	✓	✓	✗	✓	✓
Define exercise maximum number of tries	✗	✓	✗	✗	✓
Add/remove/import students to a restricted group	✗	✗	✗	✗	✓
Add new and reuse Metrics	✗	✗	✗	✗	✓
Use UA IdP to authenticate students and professors	✗	✗	✗	✗	✓
Check assigned exercises	✓	✓	✓	✓	✓
Download exercise related content	✓	✓	✓	✓	✓
View Students Code and results	✓	✗	✗	✗	✓
Students and professor user types	✗	✗	✗	✗	✓

Most Used metrics

We will take advantage of some Python libraries, such as **Scikit-learn**.



Number of benchmark datasets per higher level process



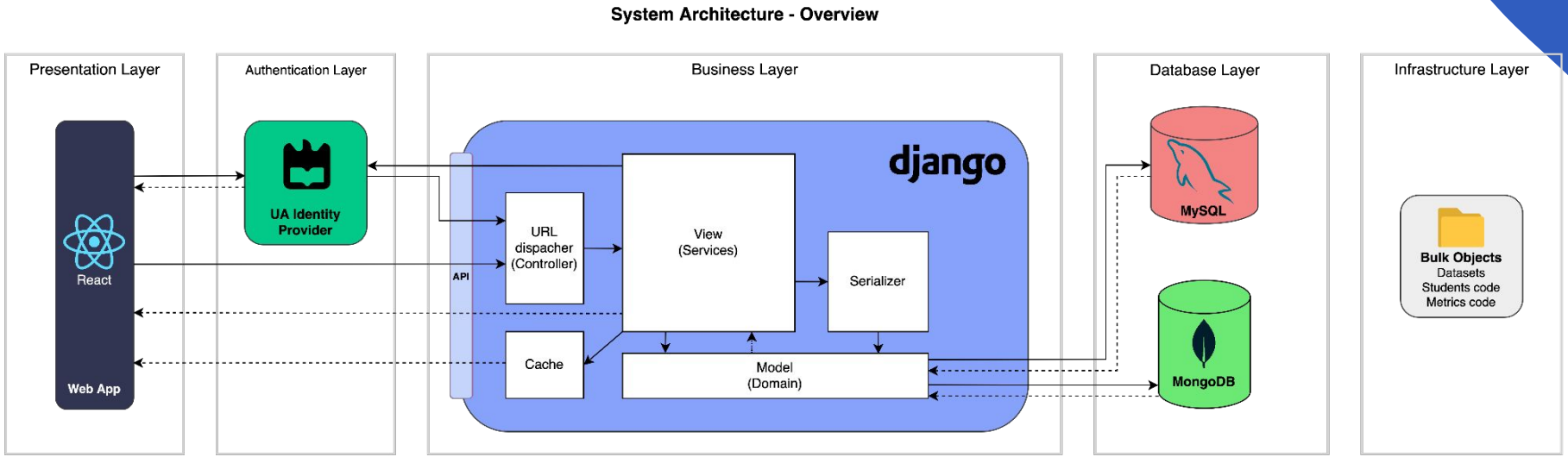
Top 10 most frequently reported performance metrics

Source: Blagec, K., Dorffner, G., Moradi, M., & Samwald, M. (2020). *A critical analysis of metrics used for measuring progress in artificial intelligence*. doi:10.48550/ARXIV.2008.02577

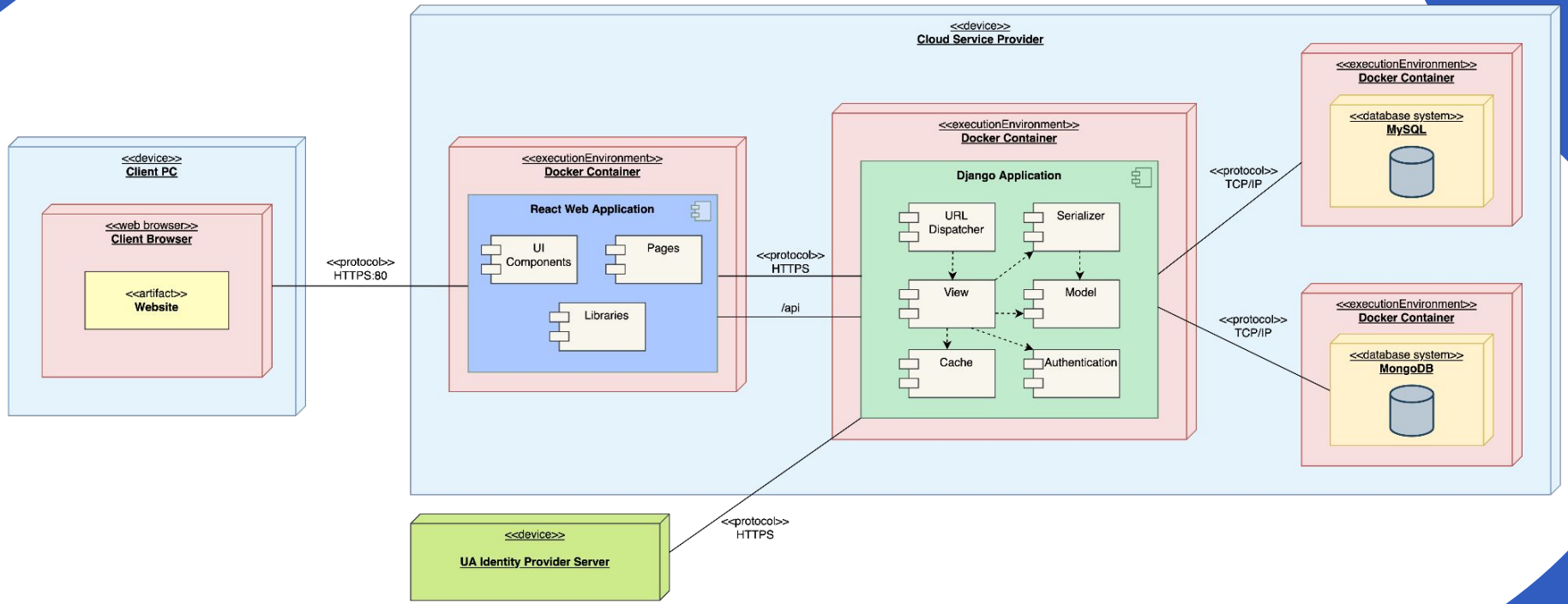
System Architecture

System architecture, Technologies, Deployment Diagram, Domain model and mock-ups

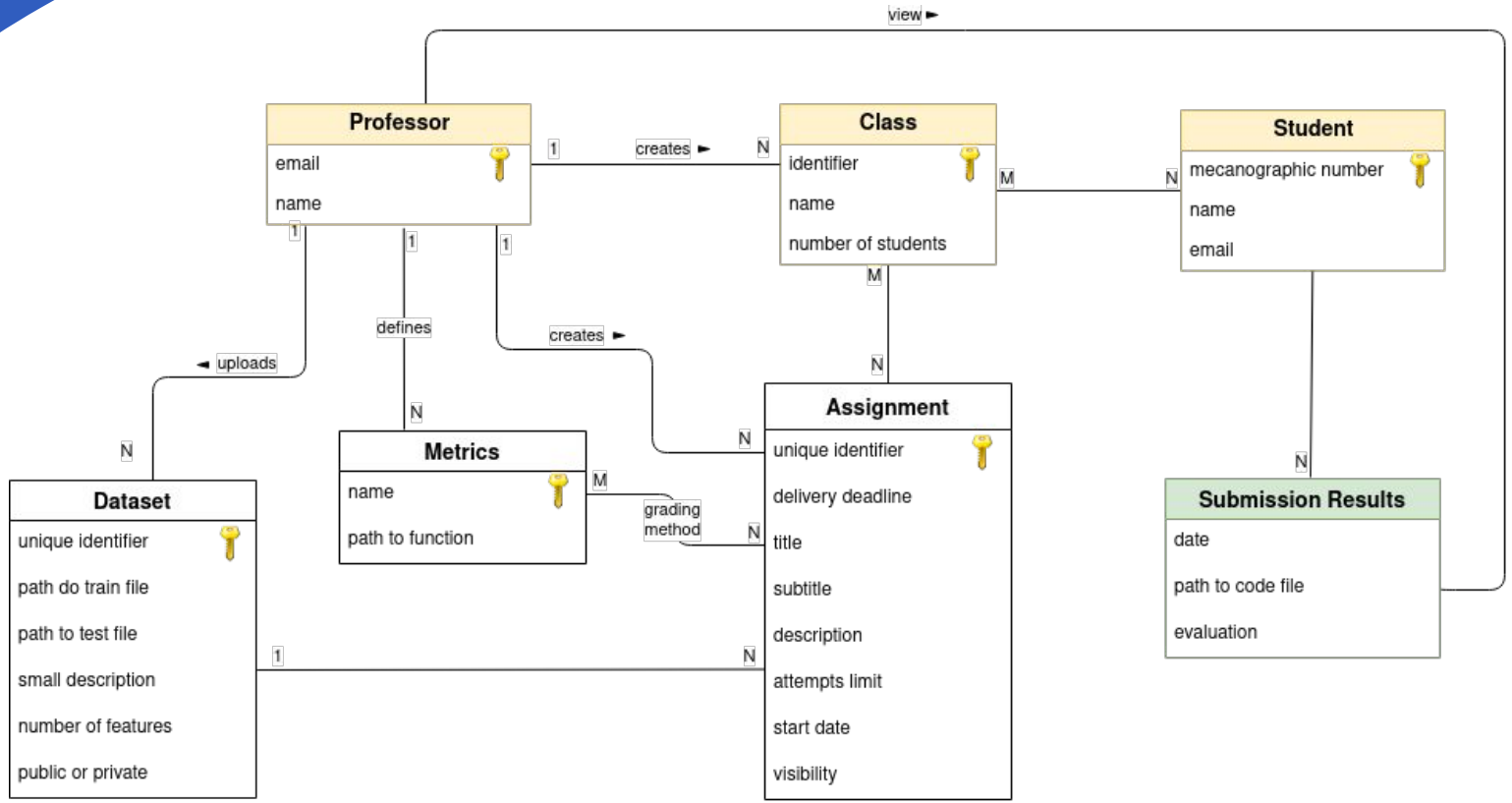
System Architecture



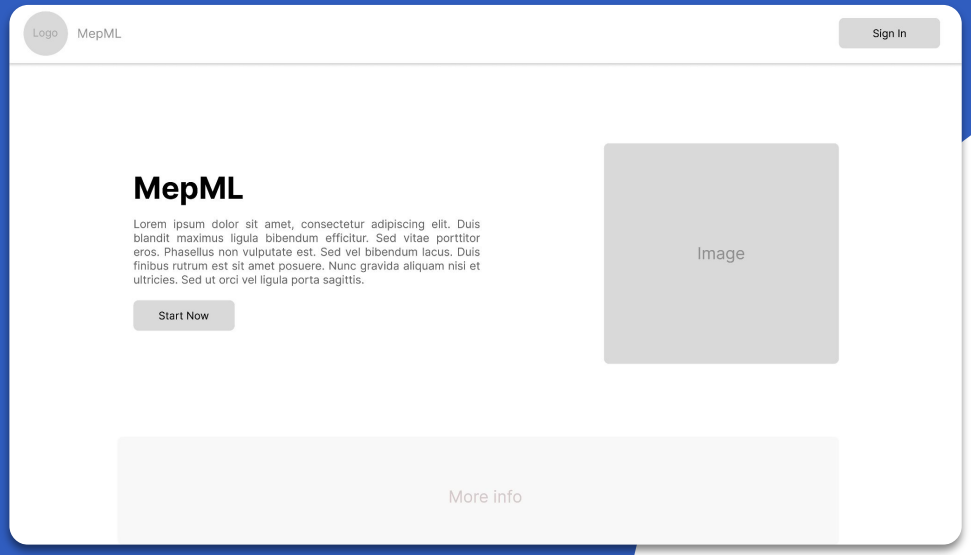
Deployment Diagram



Domain Model

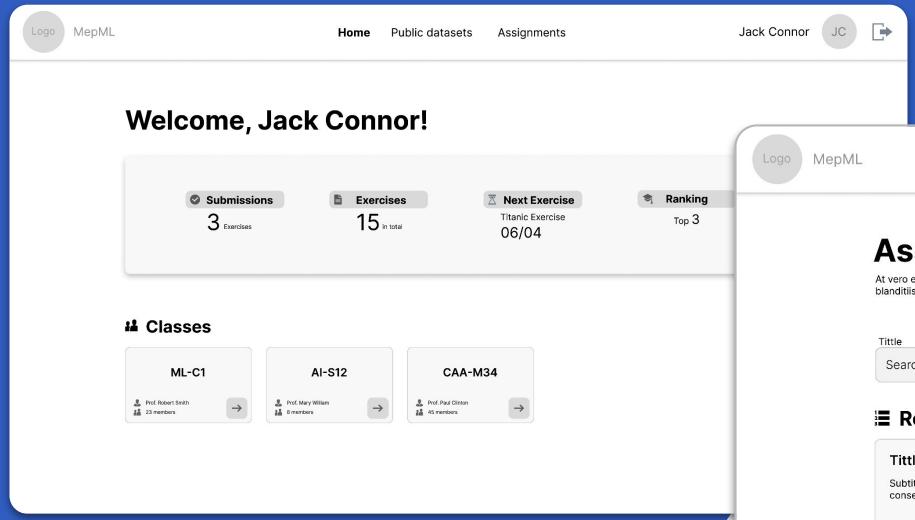


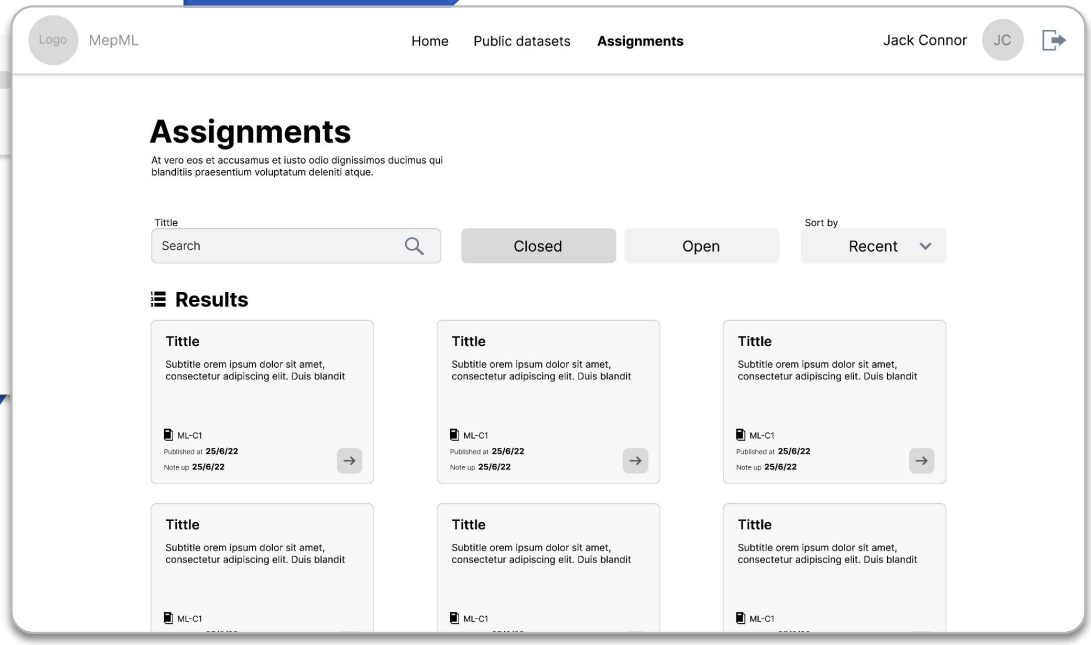
Mock-up



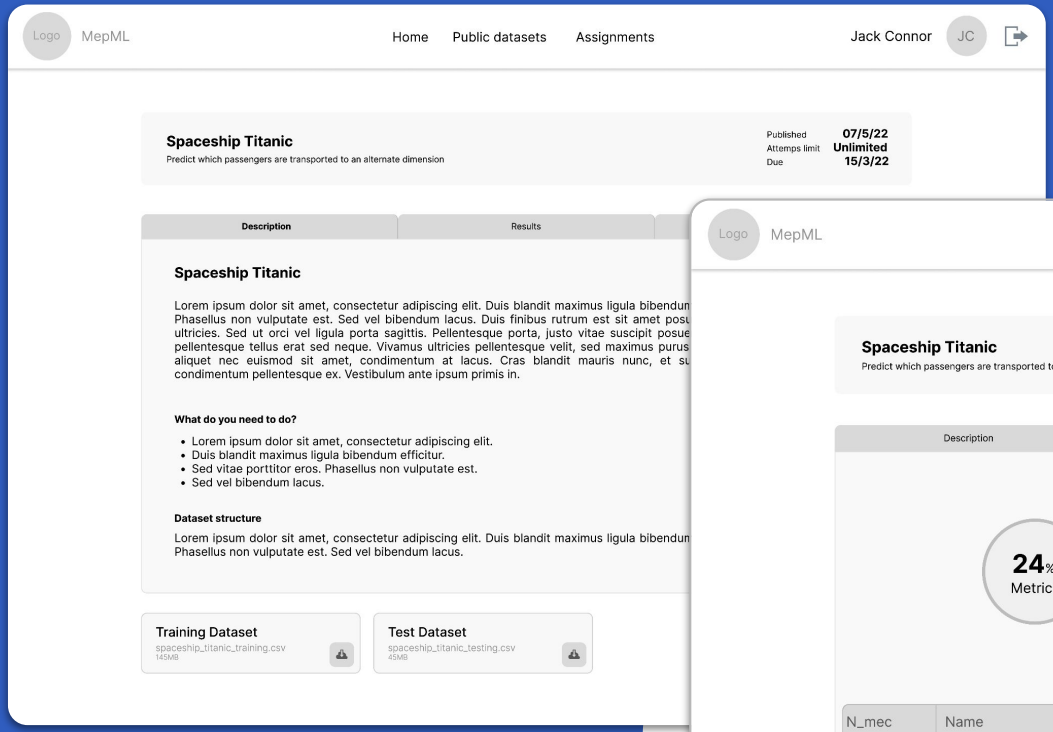
- Mockups are critical stages in the software development process
- Get a clear vision of the future product
- Receive early feedback
- Evaluate user interaction

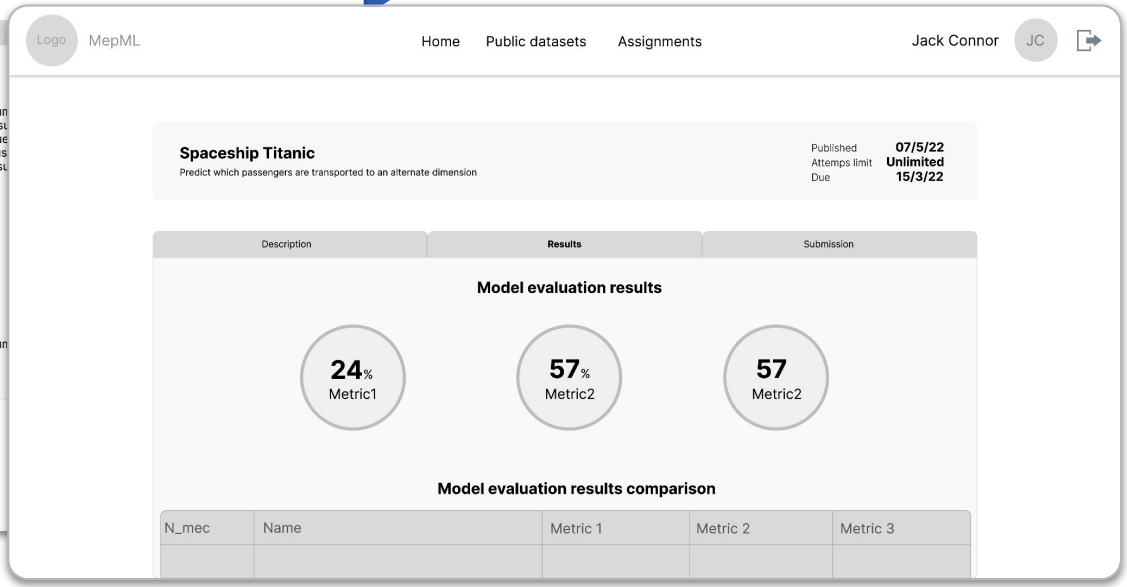
Mock-up


Logo
MepML
Home
Public datasets
Assignments
Jack Connor
JC
Welcome, Jack Connor!
Submissions
3
Exercises
Exercises
15
In total
Next Exercise
Titanic Exercise
06/04
Ranking
Top 3
Classes
ML-C1
AI-S12
CAA-M34


Logo
MepML
Home
Public datasets
Assignments
Jack Connor
JC
Assignments
At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditis praesentium voluptatum deleniti atque.
Tittle
Search
Closed
Open
Sort by
Recent
Results
Tittle
Subtitle orem ipsum dolor sit amet, consectetur adipiscing elit. Duis blandit
ML-C1
Published at 25/6/22
Note up 25/6/22

Mock-up


Logo
MepML
Home
Public datasets
Assignments
Jack Connor
JC
Spaceship Titanic
Predict which passengers are transported to an alternate dimension
Published 07/5/22
Attemps limit Unlimited
Due 15/3/22
Description
Results
Spaceship Titanic
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis blandit maximus ligula bibendun
Phasellus non vulputate est. Sed vel bibendum lacus. Duis finibus rutrum est sit amet posu
ultricies. Sed ut orci vel ligula porta sagittis. Pellentesque porta, justo vitae suscipit posue
pellentesque tellus erat sed neque. Vivamus ultricies pellentesque velit, sed maximus purus
aliquet nec euismod sit amet, condimentum at lacus. Cras blandit mauris nunc, et su
condimentum pellentesque ex. Vestibulum ante ipsum primis in.
What do you need to do?
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Duis blandit maximus ligula bibendum efficitur.
Sed vitae porttitor eros. Phasellus non vulputate est.
Sed vel bibendum lacus.
Dataset structure
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis blandit maximus ligula bibendum
Phasellus non vulputate est. Sed vel bibendum lacus.
Training Dataset
spaceship_titanic_training.csv
145MB
Test Dataset
spaceship_titanic_testing.csv
45MB


Logo
MepML
Home
Public datasets
Assignments
Jack Connor
JC
Spaceship Titanic
Predict which passengers are transported to an alternate dimension
Published 07/5/22
Attemps limit Unlimited
Due 15/3/22
Description
Results
Submission
Model evaluation results
24% Metric1
57% Metric2
57 Metric2
Model evaluation results comparison
N_mec
Name
Metric 1
Metric 2
Metric 3

Resources

- https://slidesgo.com/theme/retato-slideshow#position-38&results-1357
- https://storyset.com
- https://www.pexels.com/
- https://www.utm.mx/~caff/doc/OpenUPWeb/index.htm
- https://www.kaggle.com
- https://wandb.ai/site
- https://machinehack.com/
- https://codalab.lisn.upsaclay.fr
- https://online.visual-paradigm.com/pt/
- https://www.akendi.com/blog/scenarios-user-stories-and-use-casesoh-my/
- https://www.analyticsinsight.net/top-5-model-evaluation-metrics-for-machine-learning-projects/
- https://blog.idexlab.com/state-of-the-art-example
- https://towardsdatascience.com/4-data-science-competition-platforms-other-than-kaggle-6d1795ff46a

Relevant paper:

Alex Serban, Koen van der Blom, Holger Hoos, and Joost Visser. 2020. **Adoption and Effects of Software Engineering Best Practices in Machine Learning**. In *Proceedings of the 14th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, ACM . DOI:https://doi.org/10.1145/3382494.3410681

AI research has much to improve, hence the need for our platform.