



I-DAIR

I-DAIR CODEX

Codeless AI/ML platform for
Public Health Researchers

2023 - Steve Cygu & Oliver Deak

I-DAIR CODEX: Codeless Experience AI/ML Software



- Reduce the complexity and the learning curve of using ML/AI
- Reduce the need for specialist Data Scientist know-how
- Speed up the process of ML/AI adoption

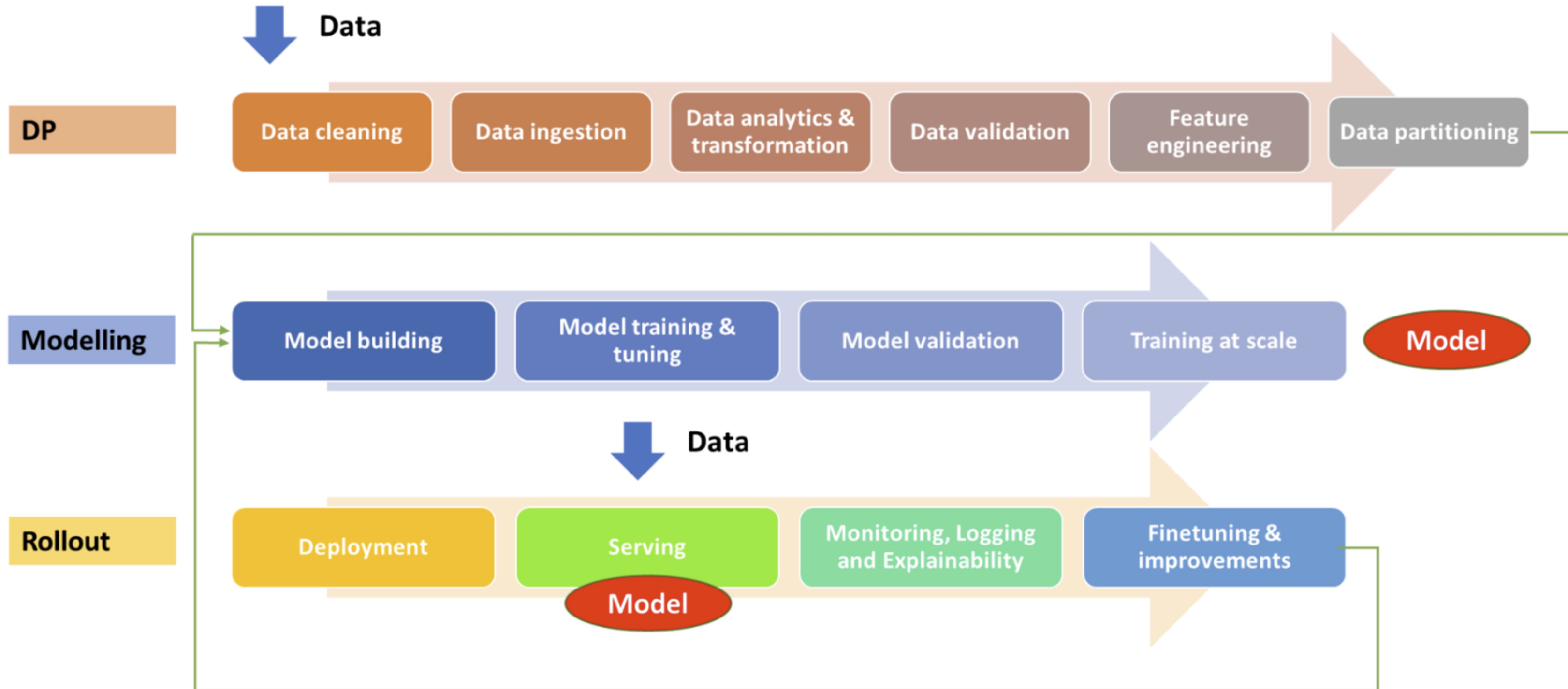




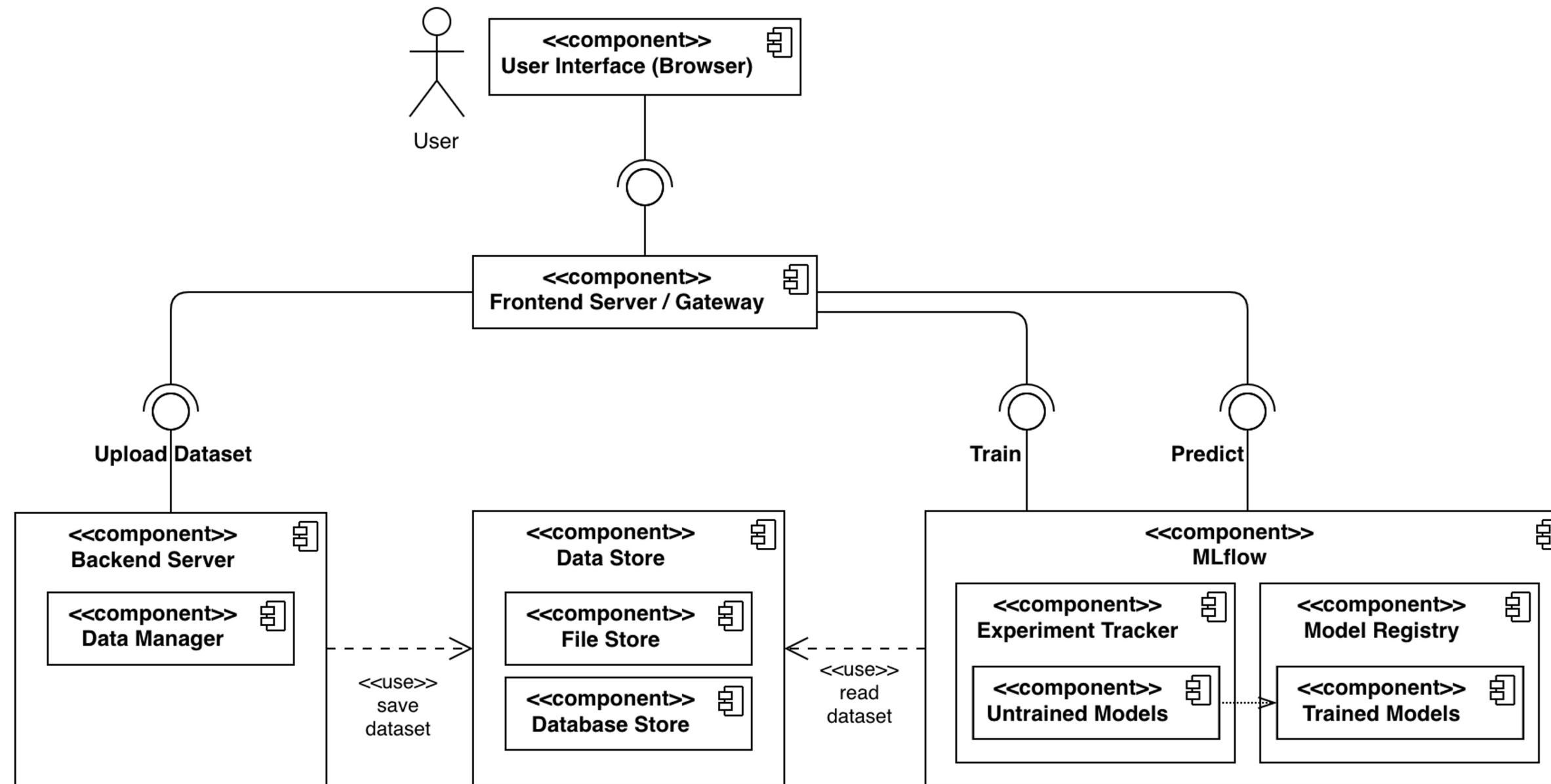
I-DAIR CODEX Overview

- Tailored ML/AI analytics geared towards easily gaining insights from structured data sources such as health and demographic surveillance sites (HDSS)
- A federated learning approach where the researchers can train and deploy their model using their own data which remains local, with an option of further sharing between collaborators
- Full automation of the entire AI/ML pipeline, allowing users to iterate through the whole data exploration, model training, deployment and prediction process much faster
- Built-in collaboration capabilities which allows researchers to invite and add collaborators to a specific project, hence fostering interdisciplinary collaboration and allows experts from diverse fields to work together seamlessly
- Support for iterative exploration and experimentation, enabling researchers to refine their methodologies and adapt to evolving research needs efficiently.

End-to-End ML Platform (CODEX) - ML Lifecycle



Architecture: Logical Component Diagram



Web based UI
(React JS)

Webserver & Reverse
Proxy (NGINX)

Web based API's
(REST)

Factory Floor
Components
(can be run on edge
node or cloud)

End-to-End ML Platform (I-DAIR CODEX)



I-DAIR

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

Upload a new dataset

Study name

Study name *

Country

Country *

Data set:

UPLOAD

SUBMIT

I-DAIR

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

Go back

Variables

tumorsize

carbonlevels

painlevel

remission

lungcapacity

I-DAIR

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

Filename	Study name	Country	Upload date	
hdp_data-1674473624973.csv	Cancer prediction	Kenya	23/01/2023 02:33:45	<div><div>Download</div><div>View</div></div>
hdp_data-1671177803889-1672325057870-1674138822016.csv	sss	ssss	19/01/2023 05:33:42	<div><div>Download</div><div>View</div></div>

Rows per page: 10 1–2 of 2

I-DAIR

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

Go back

Variables

tumorsize

tumorsize

Real number (R)

Distinct	7432	Minimum	31.616328
Distinct (%)	100.0 (%)	Maximum	121.550653
Missing	0	Zeros	0
Missing (%)	0.0 (%)	Zeros (%)	0.0 (%)
Infinite	0	Negative	0
Infinite (%)	0.0 (%)	Negative (%)	0.0 (%)
Mean	70.818934	Memory size	58.2 KiB

Quantile statistics

Minimum	31.616328
5-th percentile	51.707013

Descriptive statistics

Standard deviation	12.208752
Coefficient of variation (CV)	0.172394

frequency

tumorsize

End-to-End ML Platform (CODEX)



Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

1

Select dataset

2

Model selection

3

training

4

training result

Dataset

hdp_data-1674473624973...

Training 70 %

Validation 30 %

☐ With shuffle

☒ Without shuffle

Target

remission

NEXT

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

1

Select dataset

2

Model selection

3

training

4

training result

List of classification models

☐ logistic regression

☒ support vector machines (SVM)

HyperParameters

Number of iterations5

Max iterations6

BACK

NEXT

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

1

Select dataset

2

Model selection

3

training

4

training result

Training in progress....

support vector machines (SVM)

20%

BACK

NEXT

Upload Dataset

List of uploaded datasets

Train Models

Data partitioning strategy

1

Select dataset

2

Model selection

3

training

4

training result

Model name	accuracy	precision	recall	f1_score	permutation_test_score	permutation_test_pvalue	
logistic regression	0.698	0.000	0.000	0.000	0.698	1.000	BEST PERFORMANT
support vector machines (SVM)	0.624	0.131	0.043	0.065	0.501	0.089	BEST PERFORMANT

Rows per page: 10 1-2 of 2



i-dair.org