



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG



Source: <https://www.rmit.edu.au/study-with-us/biomedical-sciences>

ISMB 2022 - Tutorial Session

Federated Learning in Biomedicine

Applying federated learning on geographically distributed datasets

1. Application of Centralized Machine Learning
2. Application of Federated Machine Learning
3. Federated Workflows with FeatureCloud
4. Hands on Exercise
 - Assemble a federated workflow
 - Run a federated workflow with other participants

Centralized Machine Learning

Computer/Cloud/...

Cross Validation

Normalization

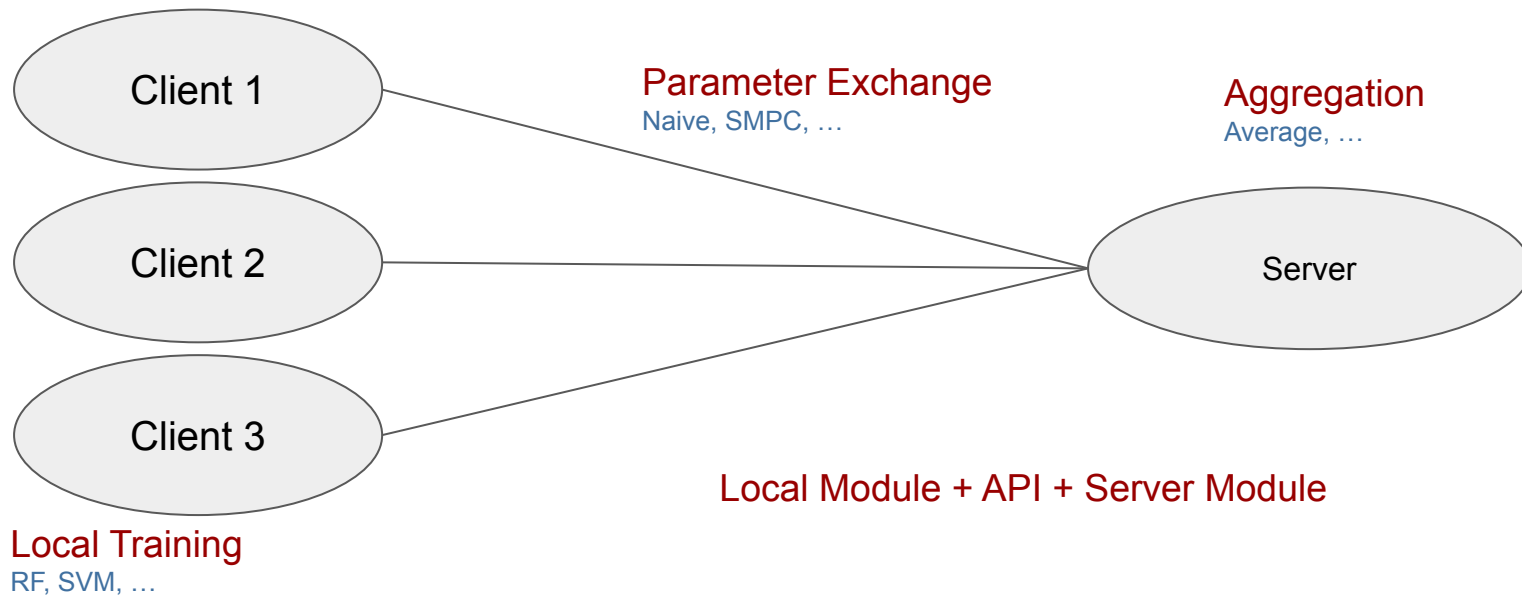
Model Training

Evaluation

```
X, y = load_data()
pipe = Pipeline([('scaler', StandardScaler()), ('clf', SVC())])

cross_validate(X, y, cv=5, scoring='roc_auc')
```

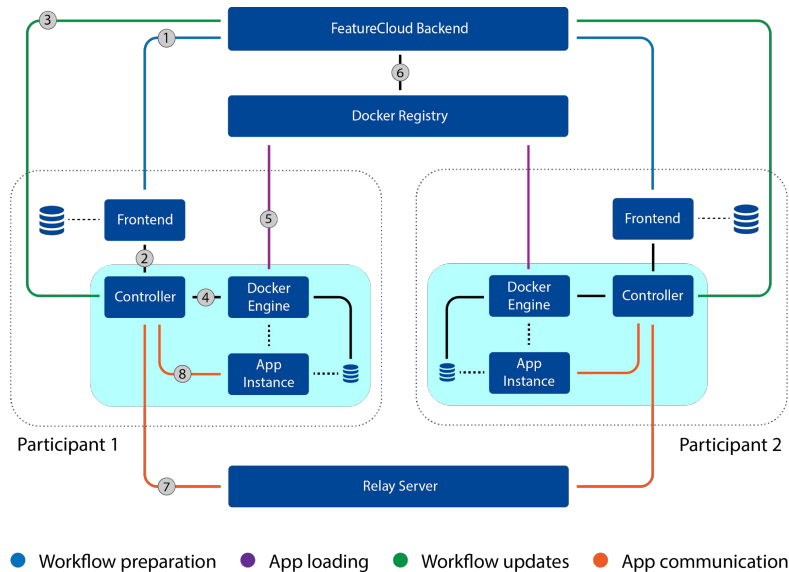
Federated Machine Learning



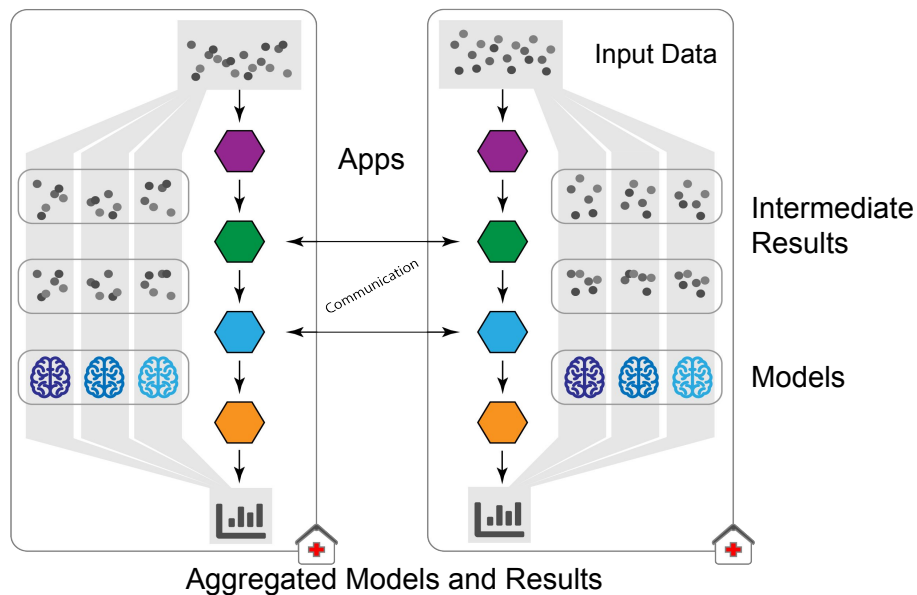
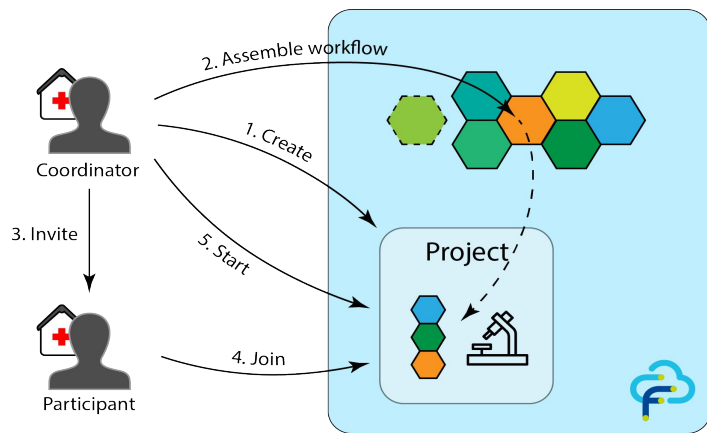
Complex Architectures

- Communication
- Different Operating Systems
- Aggregation server

Example: FeatureCloud



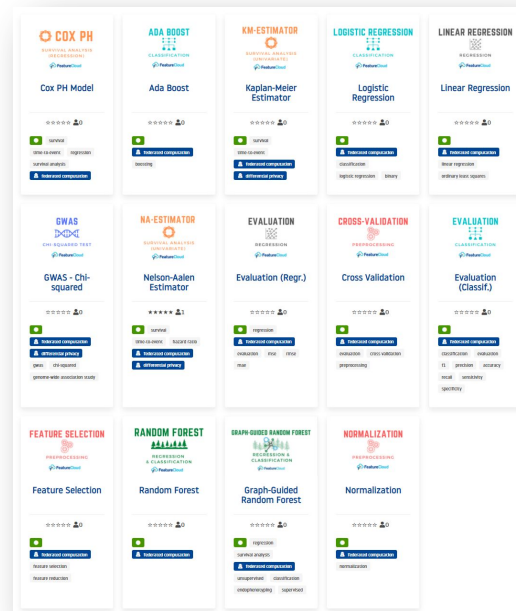
FeatureCloud for running federated workflows



FeatureCloud for running federated workflows

Ready-to-use apps by the FC community

- Cross-Validation
- Preprocessing
- Models
 - Classification
 - Regression
 - Survival
- Evaluation
 - Classification
 - Regression
 - Survival



FeatureCloud for running federated workflows

Local adjustment of certain app parameters

- Filenames
- Format
- Model parameters
- Column names
- Privacy adjustments
- ...

Config files
(yaml)

```
fc_survival_svm:
  privacy:
    enable_smpc: True # SMPC enhances privacy in a trade-off for a longer runtime, by only sending masked
    min_samples: 3 # opt out when a split of data contains less than min_samples; can not be set lower than 1
  input:
    train: "train.csv"
    test: "test.csv"
  output:
    model: "model.pickle"
    meta: "meta.yaml"
    pred: "pred.csv"
    train: "train.csv" # optional, default: fc_survival_svm.input.train; filename name for a copy of the
    test: "test.csv" # optional, default: fc_survival_svm.input.train; filename name for a copy of the
  format:
    sep: "," # separator used in csv files
    label_survival_time: "tte" # label for the time to event column
    label_event: "event" # label for the event column
    event_truth_value: True # optional, default=True; value of an entry in the event column when an event
  split:
    mode: directory # directory if cross validation was used before, else file
    dir: "cv" # cv if cross validation app was used before, else .
  svm: # only set these at coordinator; will be overwritten otherwise
    alpha: 1 # regularization parameter
    fit_intercept: False # whether to fit an intercept or not
    max_iterations: 1000 # maximum number of iterations

fc_survival_evaluation:
  privacy:
    min_concordant_pairs: 3 # minimum: 3; threshold of concordant pairs for participation
  input:
    y_test: "test.csv"
    y_pred: "pred.csv" # could be the same as y_test if predictions were appended to test data
  format:
    sep: ","
    label_survival_time: "ttest"
```


Hands-on Task

Diagnose classification of breast cancer (malignant vs. benign)

Cross Validation

- Perform a 5-fold cross validation

Normalization

- Apply a variance normalization on the data

Logistic
Regression

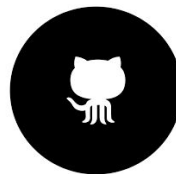
Train a logistic regression model

Evaluation
(Classif.)

Evaluate the logistic regression model

Hands-on Task (~25 minutes)

1. Find your breakout room
2. Create a FeatureCloud Account + Site
3. Choose a person as the coordinator
4. Run Docker and the FeatureCloud controller
5. Coordinator:
 - a. Create a project
 - b. Assemble workflow
 - c. Invite participants
6. Coordinator + participants
 - a. Take a look at the data
 - b. Create config files
 - c. Choose input files
7. Coordinator:
 - a. Start workflow



GitHub

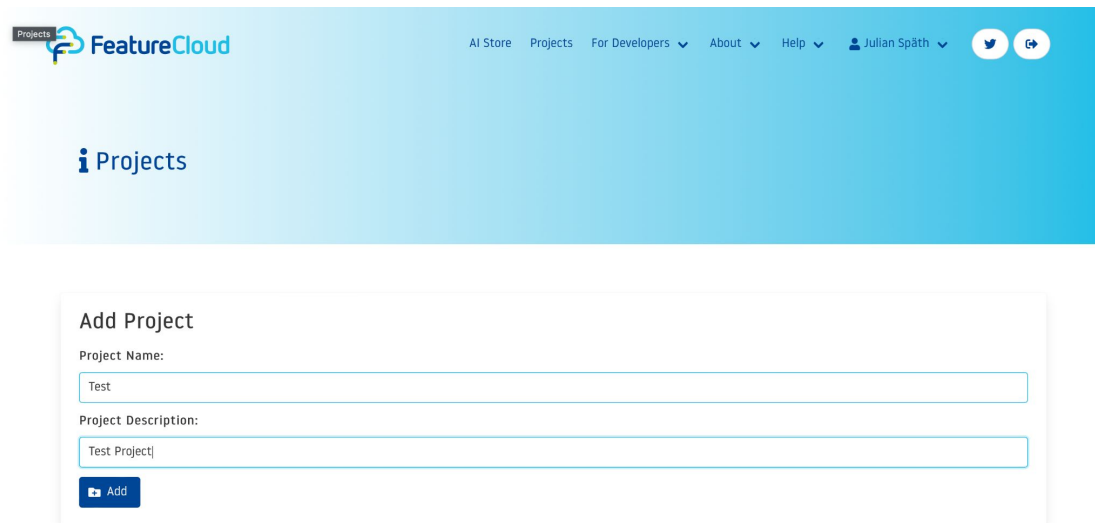
<https://github.com/FeatureCloud/ismb-tutorial-2022>



featurecloud.ai

Hands-on Task Guide

1. Create a project



The screenshot shows the FeatureCloud web interface. At the top, there is a navigation bar with the FeatureCloud logo, a 'Projects' dropdown, and links for 'AI Store', 'Projects', 'For Developers', 'About', 'Help', and a user profile for 'Julian Späth'. Social media icons for Twitter and GitHub are also present. Below the navigation bar, the main heading is 'i Projects'. The 'Add Project' form is displayed with two input fields: 'Project Name' containing the text 'Test' and 'Project Description' containing the text 'Test Project'. An 'Add' button is located at the bottom of the form.

FeatureCloud

AI Store Projects For Developers About Help Julian Späth

i Projects

Add Project

Project Name:
Test

Project Description:
Test Project

Add

Hands-on Task Guide

2. Assemble Workflow

[← Back](#) [Clone project](#) [Mark as workflow template](#)

PROJECT
Test Project

Name

Test Project

Description

Test

Available applications			Workflow	
Application	Description	Type	Application	
Cox PH Model	A Cox porportional hazard model for survival regression (Developed by: Lis Arend)	Analysis →	Cross Validation	↑ ↓ ✖
Cross Validation	A cross validation app, that creates different splits.	Pre-Processing →	Normalization	↑ ↓ ✖
Evaluation (Classif.)	An app, computing various metrics to evaluate the performance of a classification model	Evaluation →	Logistic Regression	↑ ↓ ✖
			Evaluation (Classif.)	↑ ↓ ✖

Hands-on Task Guide

3. Invite Participants

[← Back](#) [Clone project](#) [Mark as workflow template](#)

PROJECT

Test Project

STATUS
Ready

Details

Workflow

Invitations

Token	Site	Action
8313f534-2c93-4740-b9fc-e59527ca5e46	Julian Spaeth (Munich/Germany)	
3a41bd05-e3bf-42b2-9003-2a3837b25234		Copy Delete
e11c1cfa-5cd1-4868-82ac-8760460f017d		Copy Delete
60fa6bf2-33d5-47cc-a929-78a011f9f117		Copy Delete

[+ Token](#)

Hands-on Task Guide

4. Start Workflow, Choose Data

PROJECT
Test Project

STATUS
Prepare

Details

Workflow

No.	Application	Status <small>Site Name / Message / App State / Progress / Elapsed Time</small>	View/Download
1	Cross Validation	Not started	
2	Normalization	Not started	
3	Logistic Regression	Not started	
4	Evaluation (Classif.)	Not started	

Stop workflow

Please select the dataset in the section below

Select dataset

Add file ...

Continue Without Upload

Hands-on Task Guide

5. Enjoy the workflow :)

← Back

Clone project

Mark as workflow template

PROJECT

Test Project

STATUS

Running

ELAPSED TIME

00:02:27

Details

Workflow

No.	Application	Status	View/Download
Site Name / Message / App State / Progress / Elapsed Time			
1	Cross Validation	<div>⚙ Running</div> <div>Julian Spaeth (Munich/Germany) Running application <div>Running</div> 00:00:01</div>	▶ Starting application
2	Normalization	⏸ Not started	
3	Logistic Regression	⏸ Not started	
4	Evaluation (Classif.)	⏸ Not started	

Stop workflow

ISMB 2022 - Tutorial - Federated Learning in Biomedicine

Thank you!

Please, fill out the [Feedback](#) form.