Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

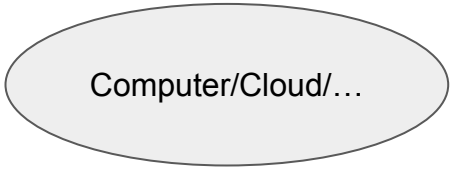
ISMB 2022 - Tutorial Session

Federated Learning in Biomedicine

Applying federated learning on geographically distributed datasets

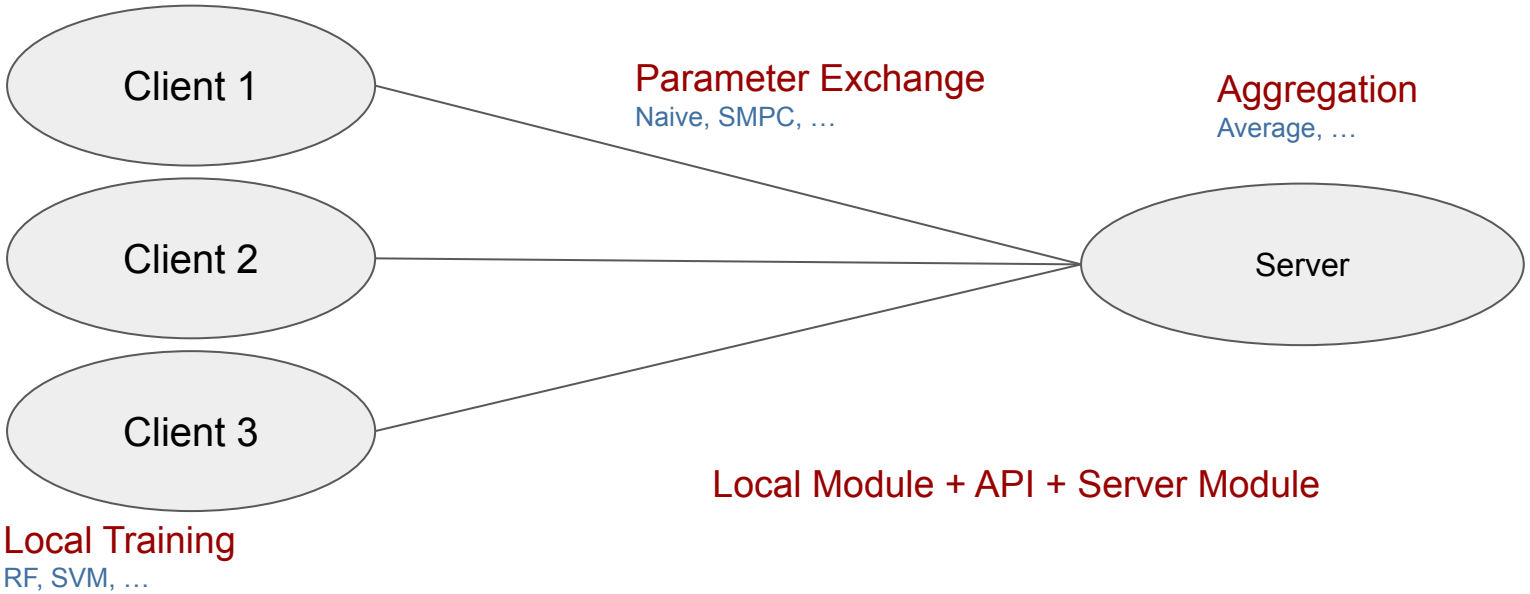
1. Application of Centralized Machine Learning
2. Application of Federated Machine Learning
   a. Concepts
3. Federated Workflows
4. Hands on Exercise
   a. Run a federated Workflow

Centralized Machine Learning



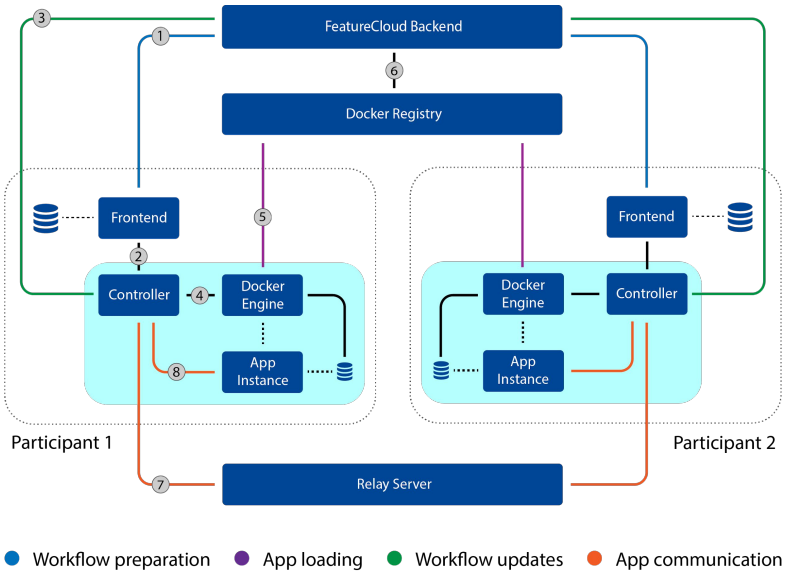

```
X, y = load_data()
pipe = Pipeline([('scaler', StandardScaler()), ('clf', SVC())])

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

Federated Machine Learning

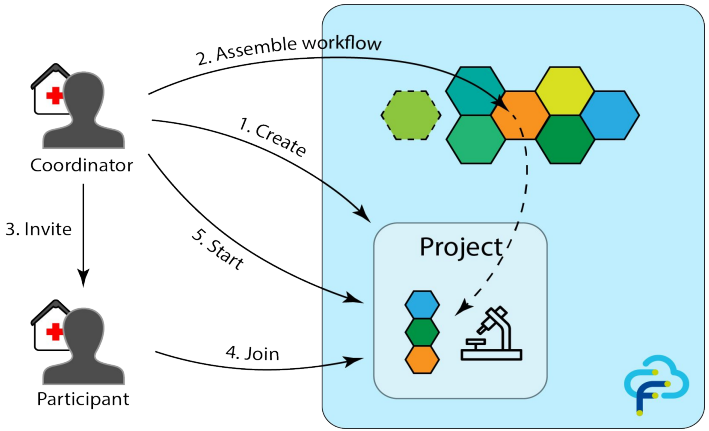


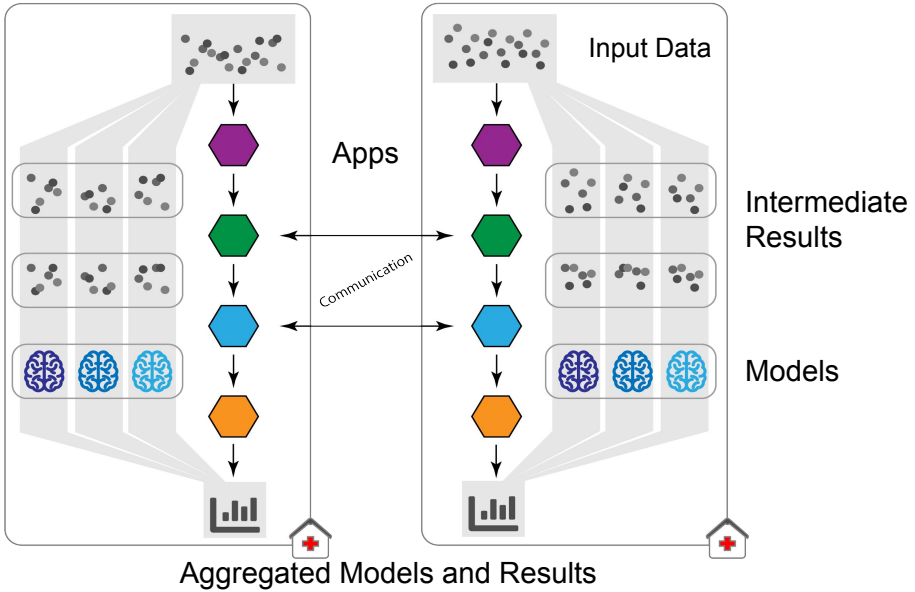
Complex Architectures

Example: FeatureCloud



FeatureCloud for running federated workflows

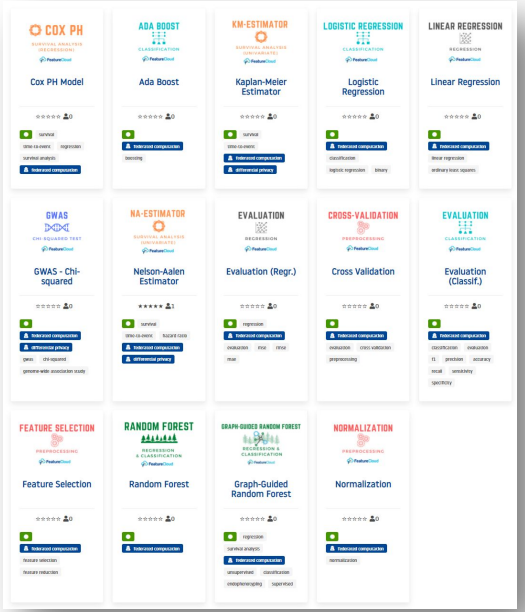




FeatureCloud for running federated workflows

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)

```yaml
fc_survival_svm:
  privacy:
    enable_smpc: True  # SMPC enhances privacy in a trade-off for a longer runtime, by only sending maske
    min_samples: 3  # opt out when a split of data contains less than min_samples; can not be set lower t
  input:
    train: "train.csv"
    test: "test.csv"
  output:
    model: "model.pickle"
    meta: "meta.yml"
    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 t
  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 even
  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: "tte"
```

Hands-on Task

Diagnose classification of breast cancer (malignent vs. benign)

- Perform a 5-fold cross validation
- Apply a variance normalization on the data
- Train a logistic regression model
- Evaluate the logistic regression model

Hands-on Task

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

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

featurecloud.ai

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

ISMB 2022 - Tutorial - Federated Learning in Biomedicine

Thank you!