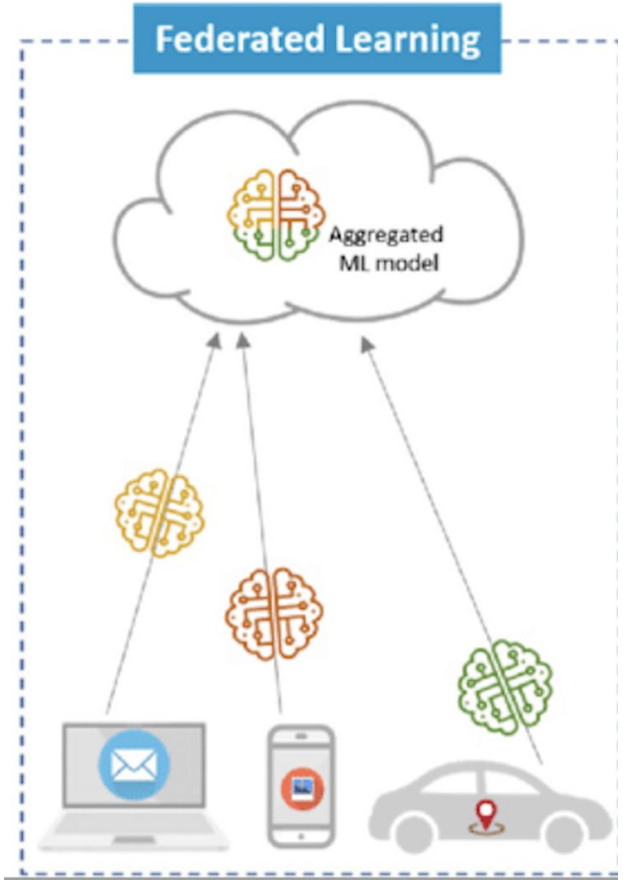
Federated Learning for Medical Institutions

Advisor: Magdalini Eirinaki
By Arselan Alvi, Poojitha Vaddey, Vatsa Patel, Yuan Wan

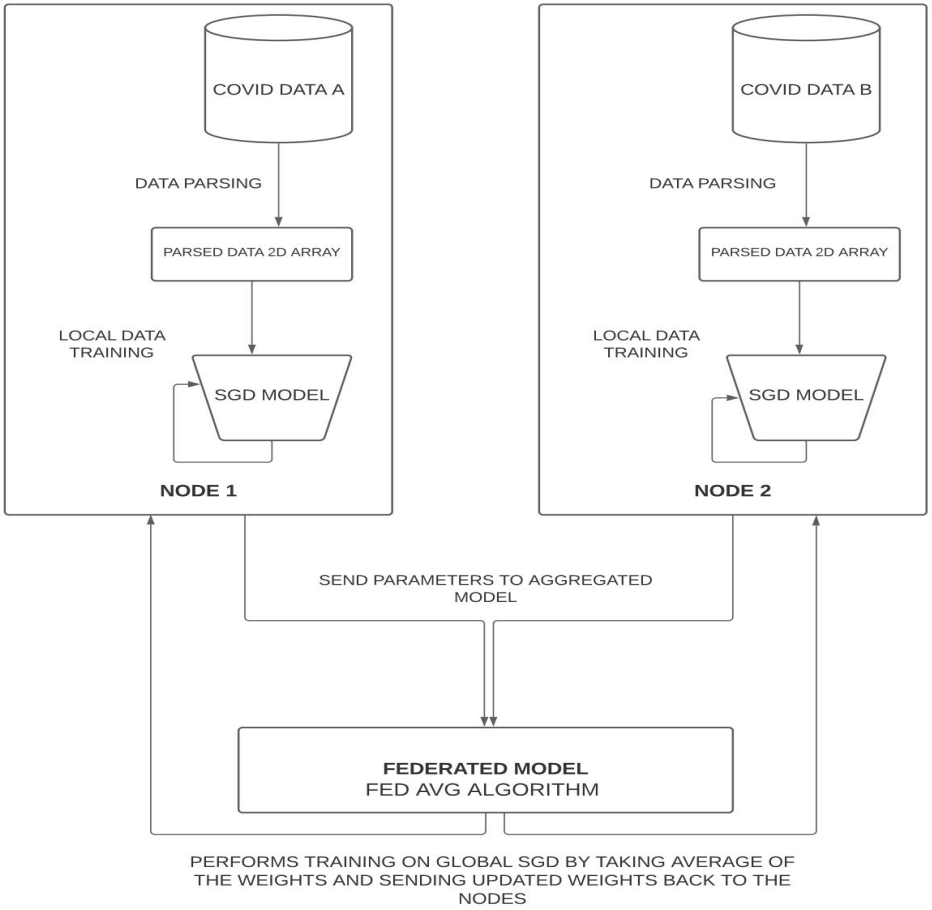
About

- Why Federated Learning?
  - Federated learning allows us to compute accurate predictions without looking at the data.
  - Provides us with the ability to seek to imagine what we could build using the data which is not accessible and still respect the data privacy.
- How are we applying the concept?
  - A software prediction tool utilized by multiple healthcare organizations to achieve predictions on their respective datasets without sharing their data.



Model Architecture

- Federated instance will not be able to look at the dataset.
    - Data is trained locally for each node and only weights are transferred back and forth during each iteration between each institution and the model.
- Predictions are based on the average calculated upon weights.

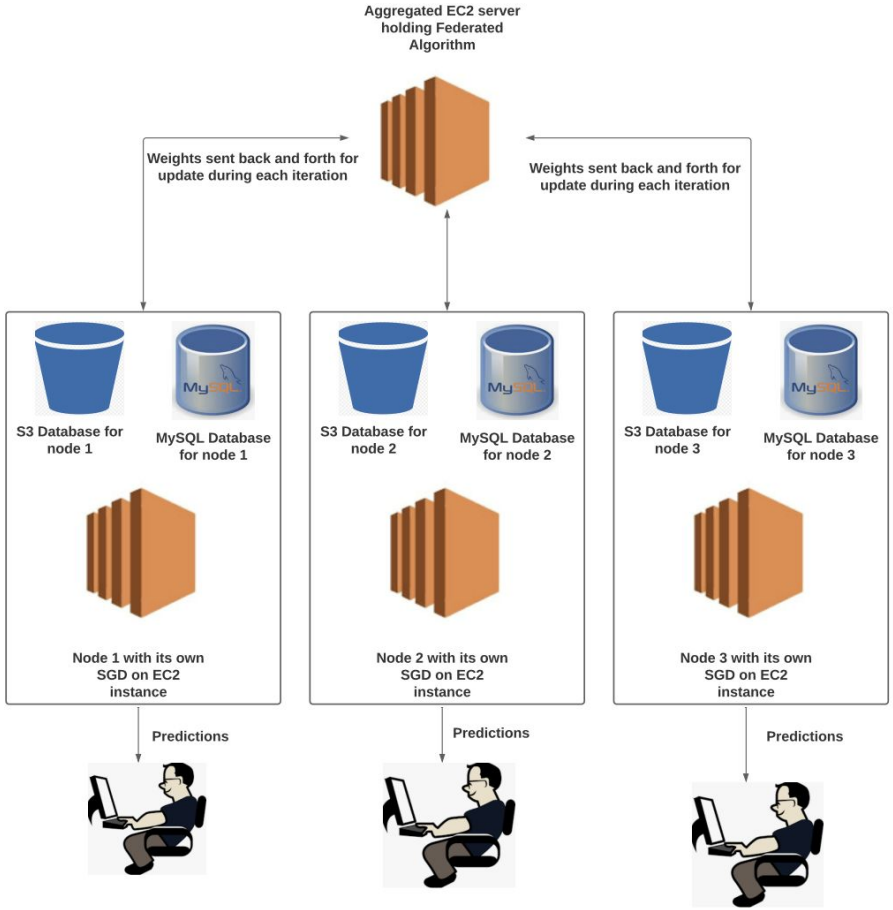


Datasets

- Datasets
  - COVID-19 datasets retrieved from healthgov.
  - ~3000 records per dataset
  - Institution 1: Bay area counties COVID-19 hospitalizations
  - Institution 2: Southern California counties COVID-19 hospitalizations
- Dataset Features:
  - Bed occupancy
  - ICU bed occupancy
  - Suspected COVID
  - Confirmed COVID
  - Total pending hospitalizations

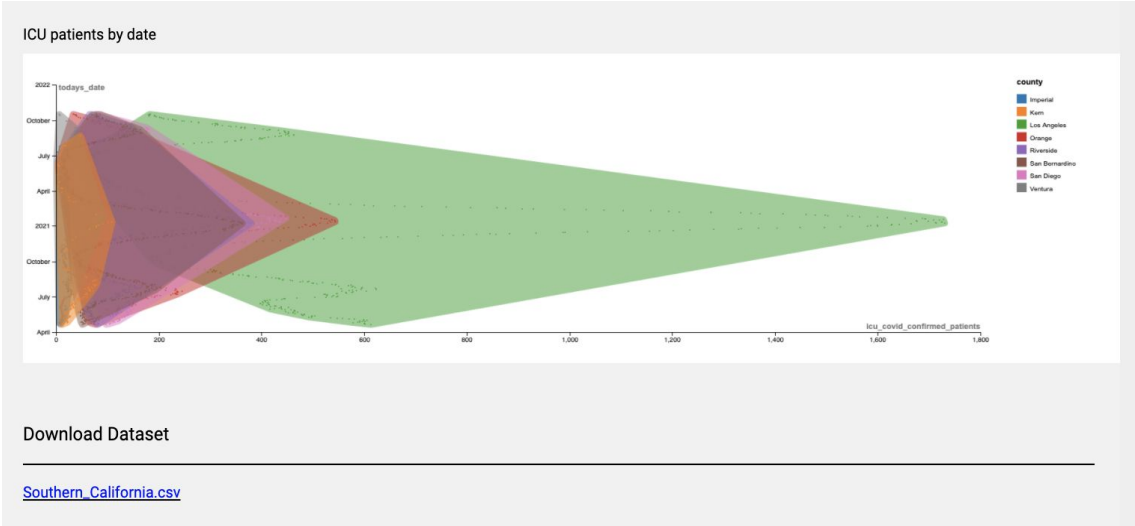
System Architecture

- A unique S3 server to hold the dataset and MySQL database for a particular institution.
    - Dataset: COVID-19 records
    - MySQL:
        - Credentials
        - Prediction jobs
        - Graph coordinates
        - Error Analysis
- EC2 instance for the Federated model and each institution.



Features - Home / Dashboard

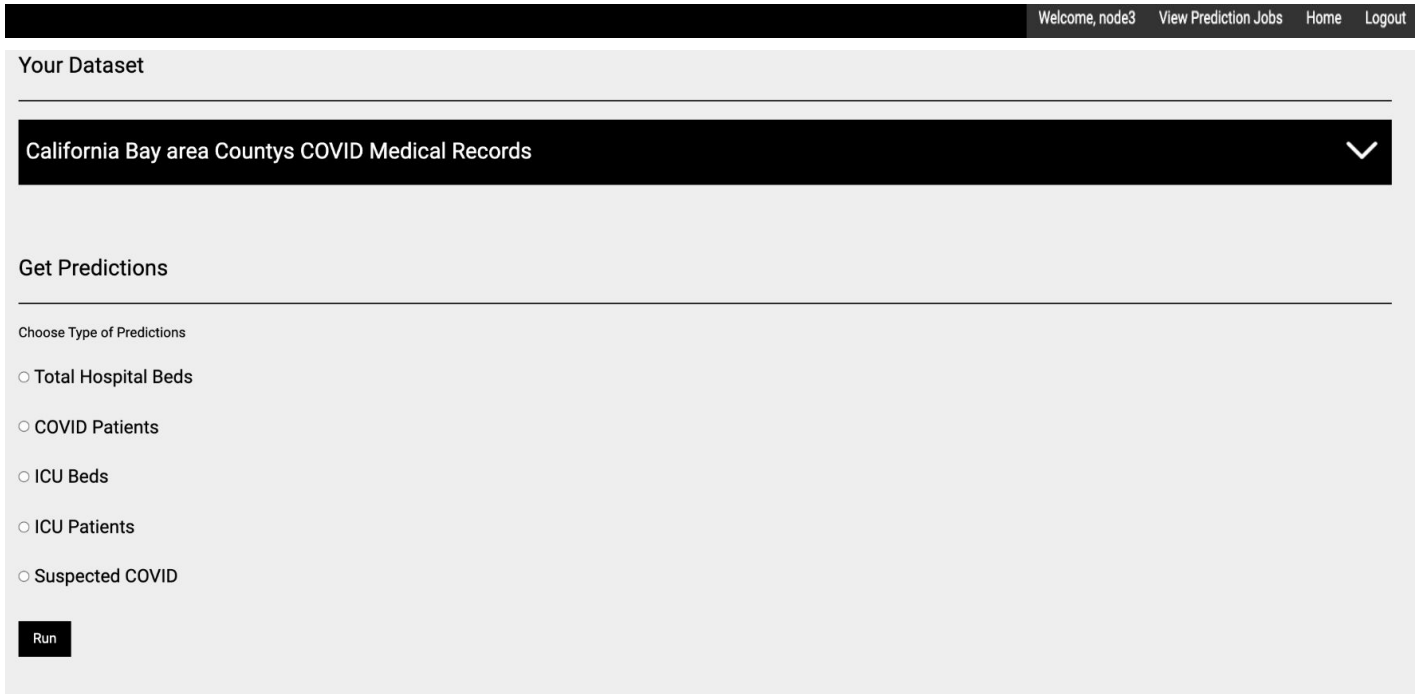
- Secure Authentication
  - Institutions are assigned credentials which they can use to can log into their server.
- Dataset Visualization



- Multiple graphs provided to visualise data.
- Users can choose to download and view their dataset in csv format.

Features - Submit Predictions

- Submit Prediction job based on the dataset feature.



Features- View Prediction Status

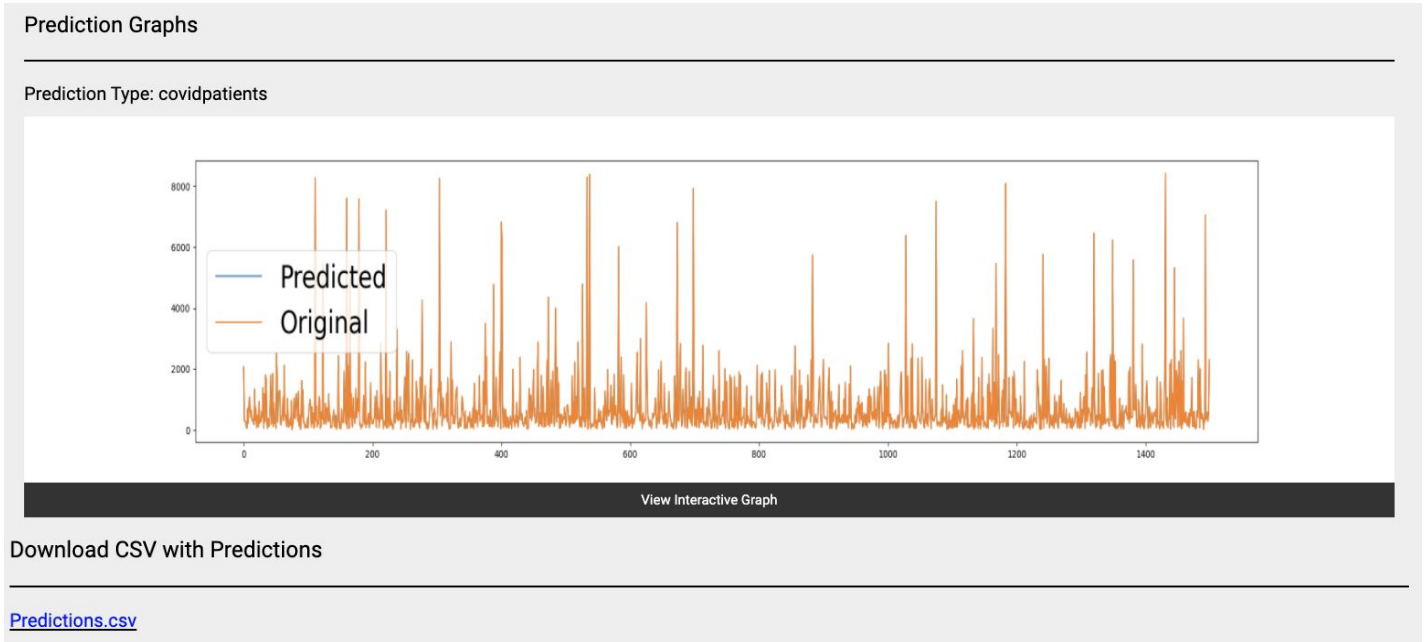
- View live status and history of all the prediction jobs with time analysis.

Predictions

No.	Time Started	Job Status	Total Time (minutes)	Job Link
1	2021-11-12 13:13:42	completed	2.102271318435669	View Job
2	2021-11-11 14:43:00	completed	1.5407821226119995	View Job
3	2021-11-10 17:58:33	completed	2.0247301959991457	View Job
4	2021-11-10 12:53:16	completed	1.5584645652770996	View Job
5	2021-11-09 22:30:29	completed	1.583590998649597	View Job
6	2021-11-09 21:37:24	completed	1.7111522626876832	View Job
7	2021-11-09 21:33:57	completed	1.5355055952072143	View Job
8	2021-11-09 21:01:29	completed	23.157845816612245	View Job
9	2021-11-09 19:35:44	completed	14.224087944030762	View Job
10	2021-11-09 19:19:06	completed	0.643025336265564	View Job
11	2021-11-09 17:51:41	completed	0.5025140571594239	View Job

Features - View Predictions

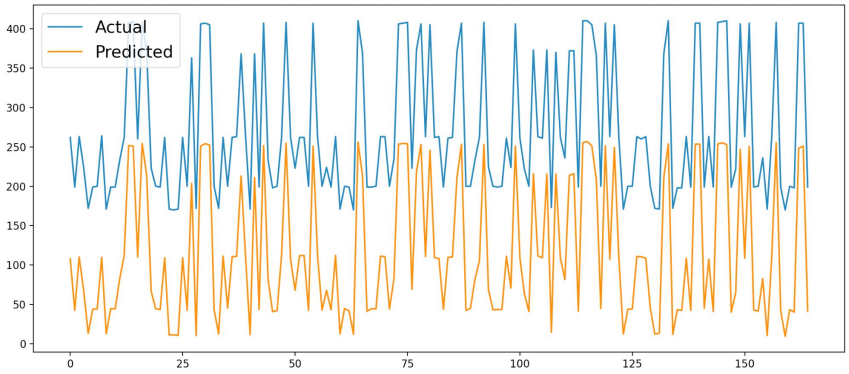
- View multiple accuracy graphs, data loss analysis, download predictions in csv file and perform interactive analysis.



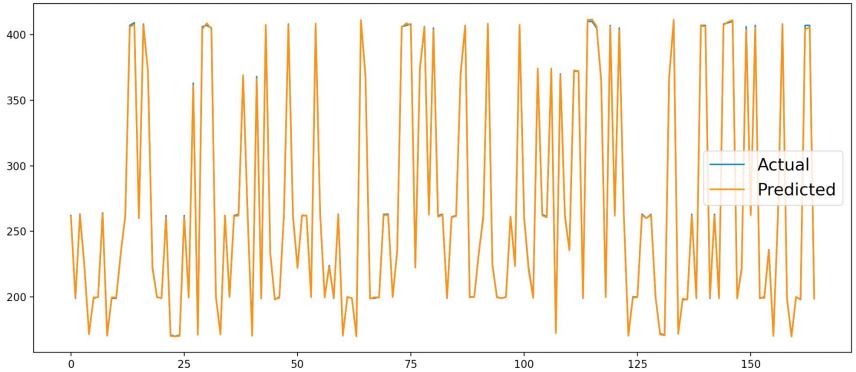
DEMO

Analysis and Results

Prediction error rate achieved by Centralized vs Federated model.



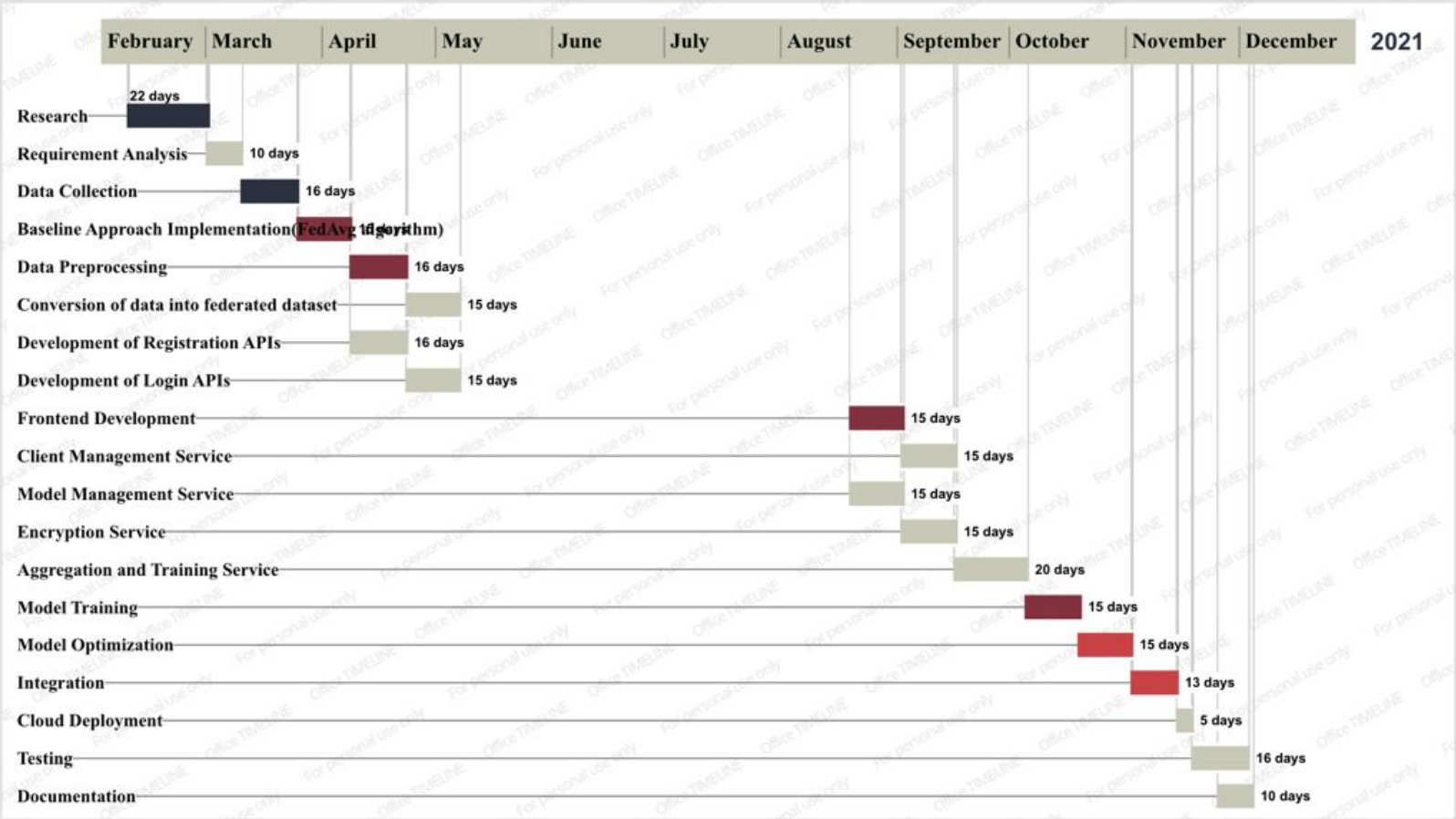
Centralized



Federated

Federated model not only maintains privacy but also provides accurate predictions compared to the Centralized approach

Project Timeline



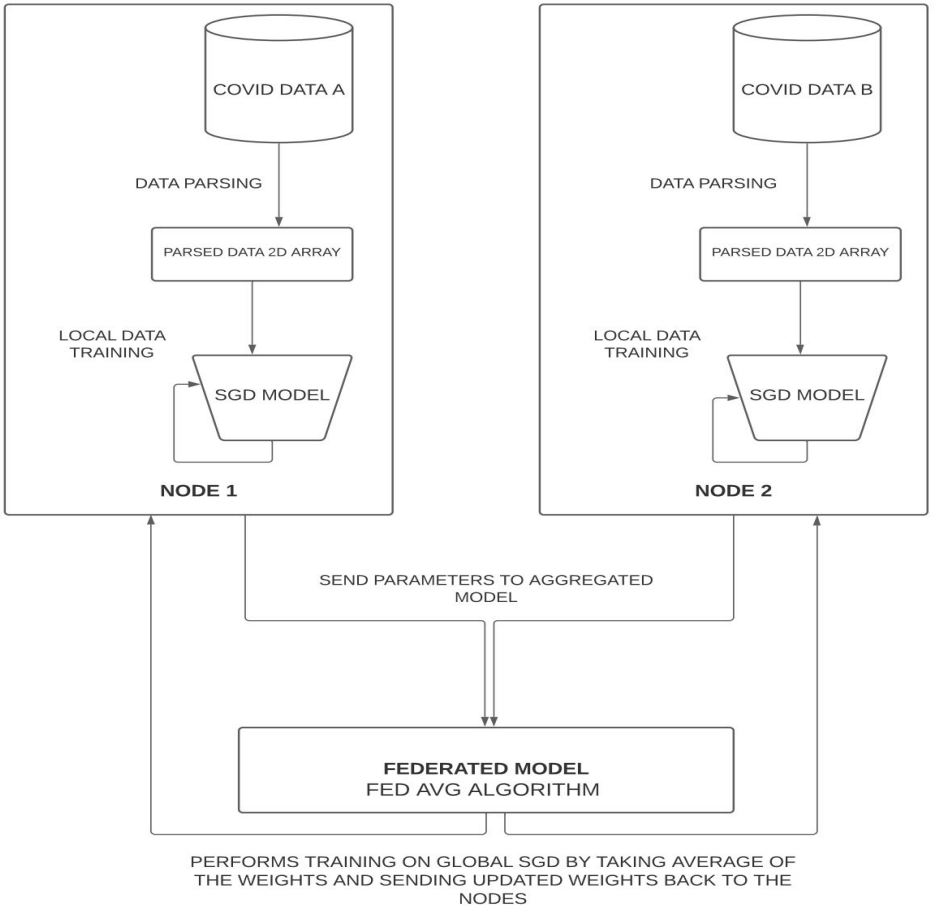
Tools and technologies

- **Python 3**
  - Building Model
- **Flask Framework and Jinja2**
  - Tool Integration
- **HTML/CSS, javascript and bootstrap**
  - Frontend
- **MySQL Databases**
  - To hold server credentials, predictions.
- **AWS Services**
  - Deployment
- **COVID-19 Datasets retrieved from healthcare.gov**

Thank you

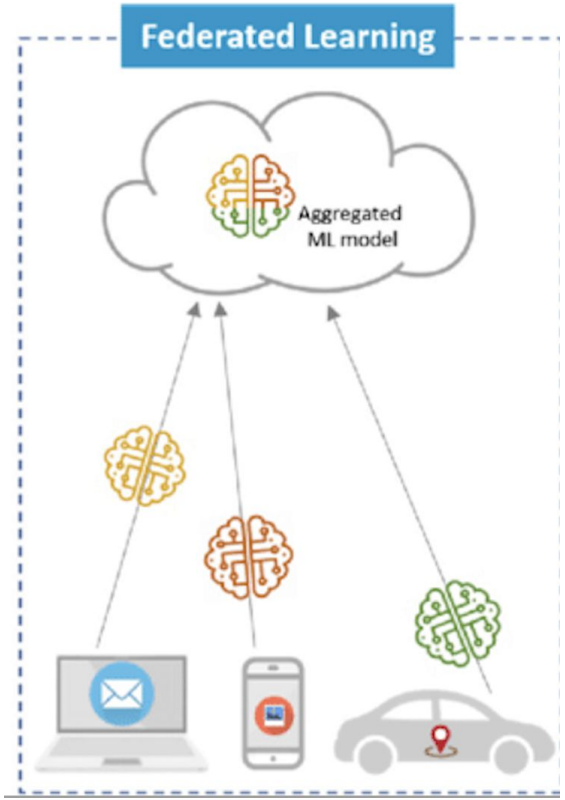
Backup

Model setup and Dataset



About

- A software prediction tool which is utilized by multiple healthcare organizations to achieve predictions on their respective datasets and also maintaining data privacy.
- Why Federated Learning?
    - For most healthcare organizations, data sharing is risky because of privacy issues.
    - By utilizing the concept of Federated Learning, multiple organizations can achieve accurate predictions which are computed by an aggregated model without looking at the data, hence maintaining privacy.



Datasets

- Datasets
  - COVID-19 datasets retrieved from healthgov.
  - ~3000 records per dataset
  - Institution 1: Bay area counties COVID-19 hospitalizations
  - Institution 2: Southern California counties COVID-19 hospitalizations
- Dataset Features:
  - Bed occupancy
  - ICU bed occupancy
  - Suspected COVID
  - Confirmed COVID
  - Total pending hospitalizations

Model Architecture

- Federated instance will not be able to look at the dataset.
    - Data is trained locally for each node and only weights are transferred back and forth during each iteration between each institution and the model.
- Predictions are based on the average calculated upon weights.