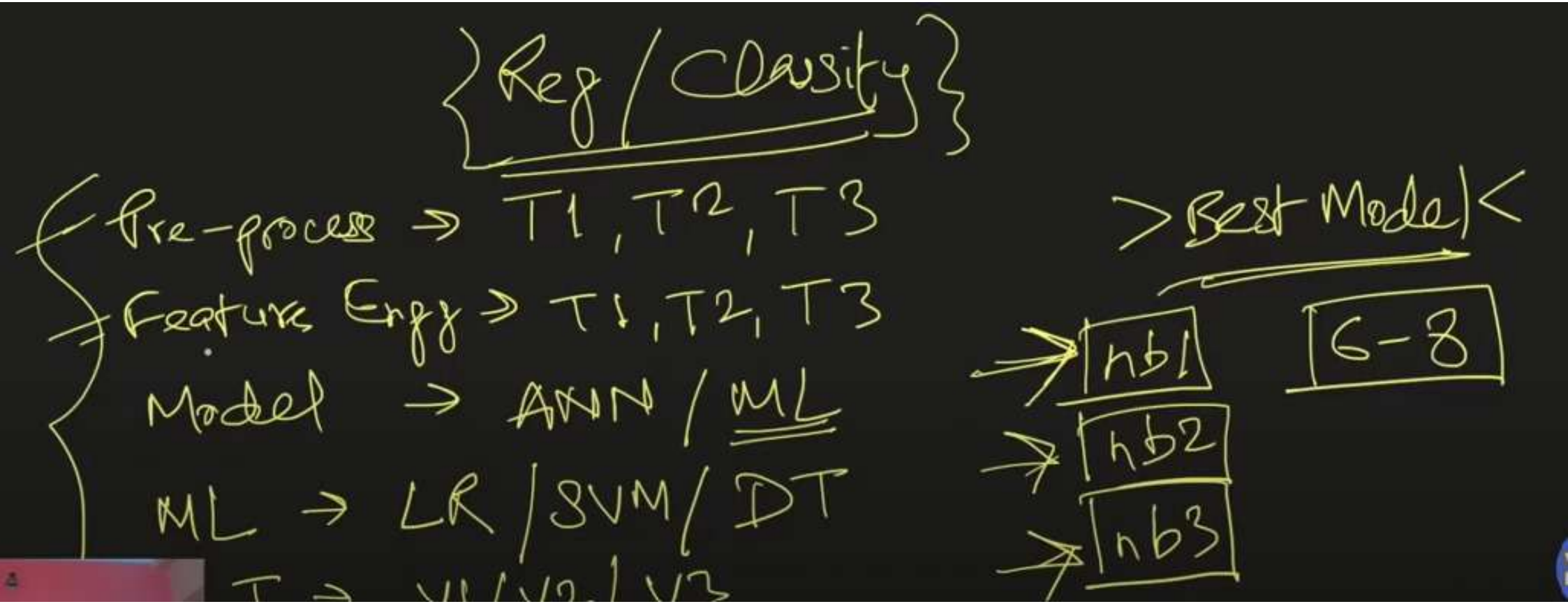
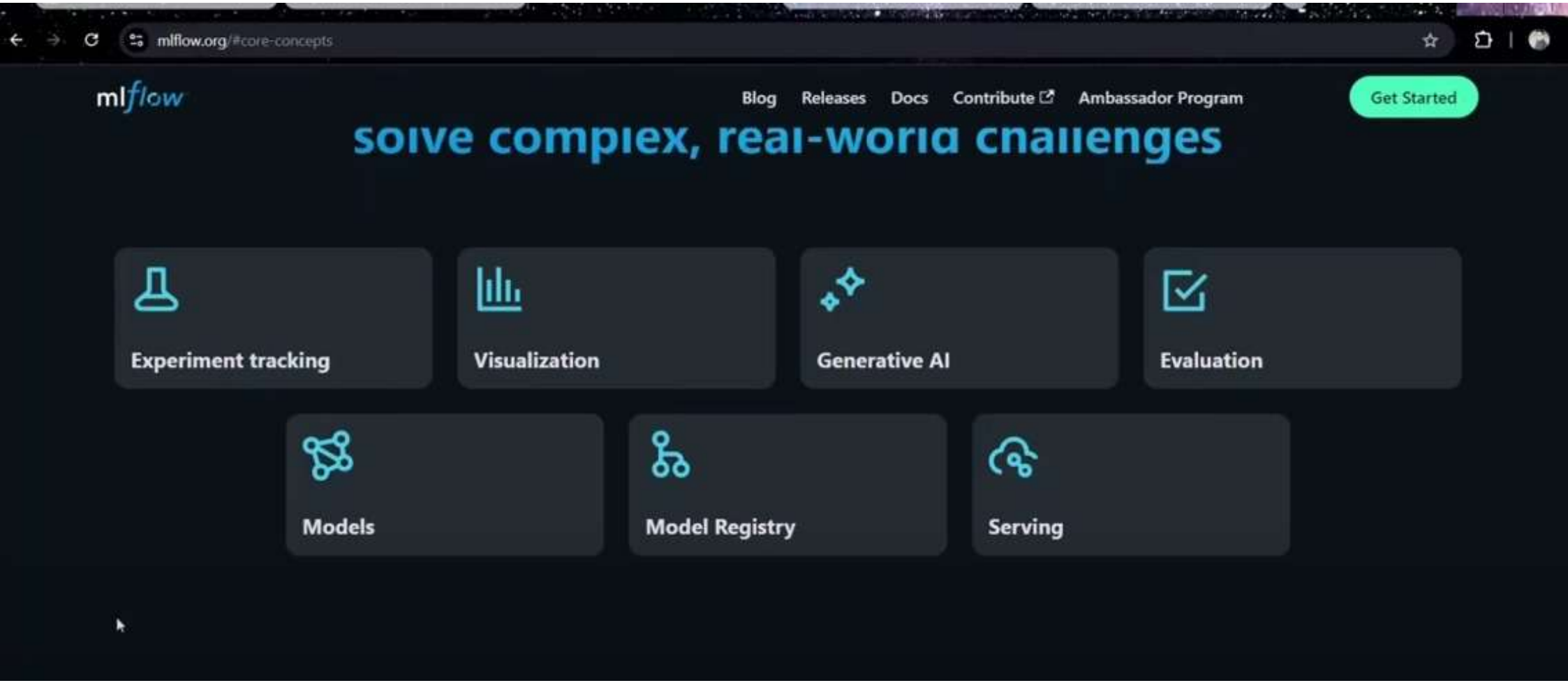
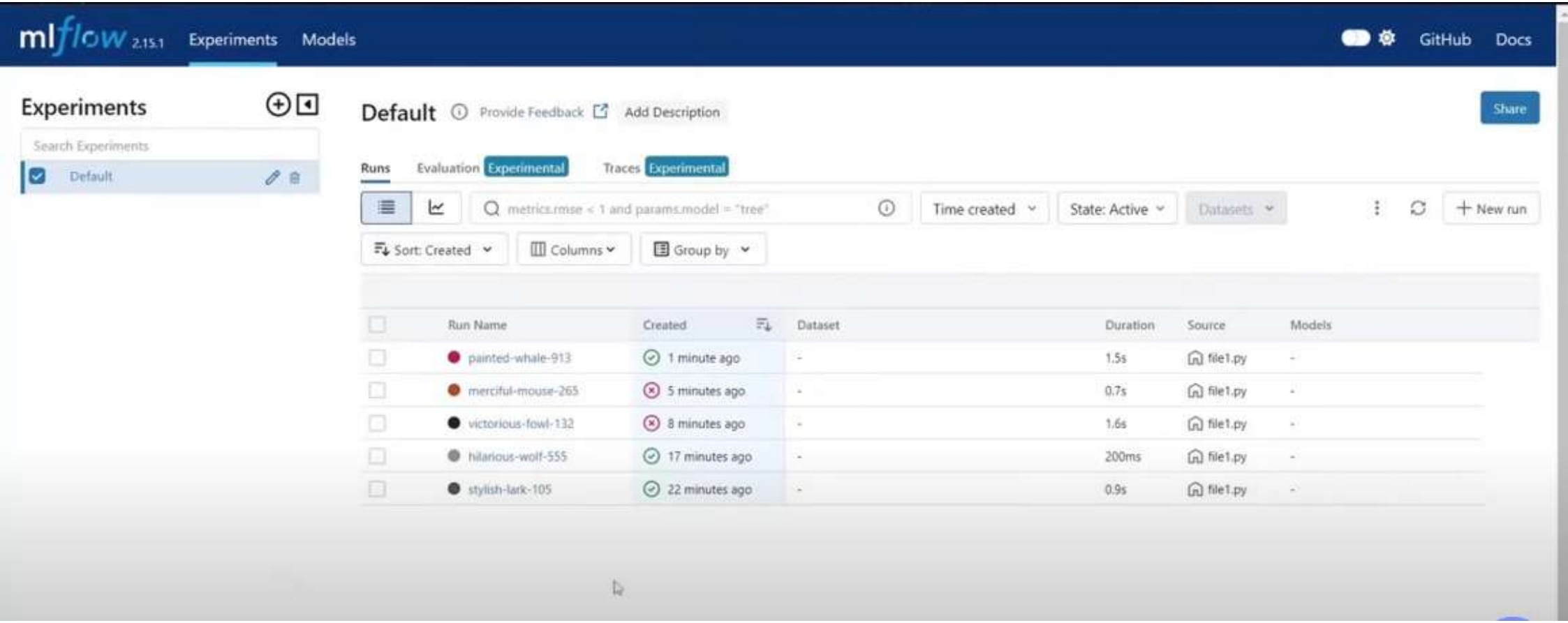
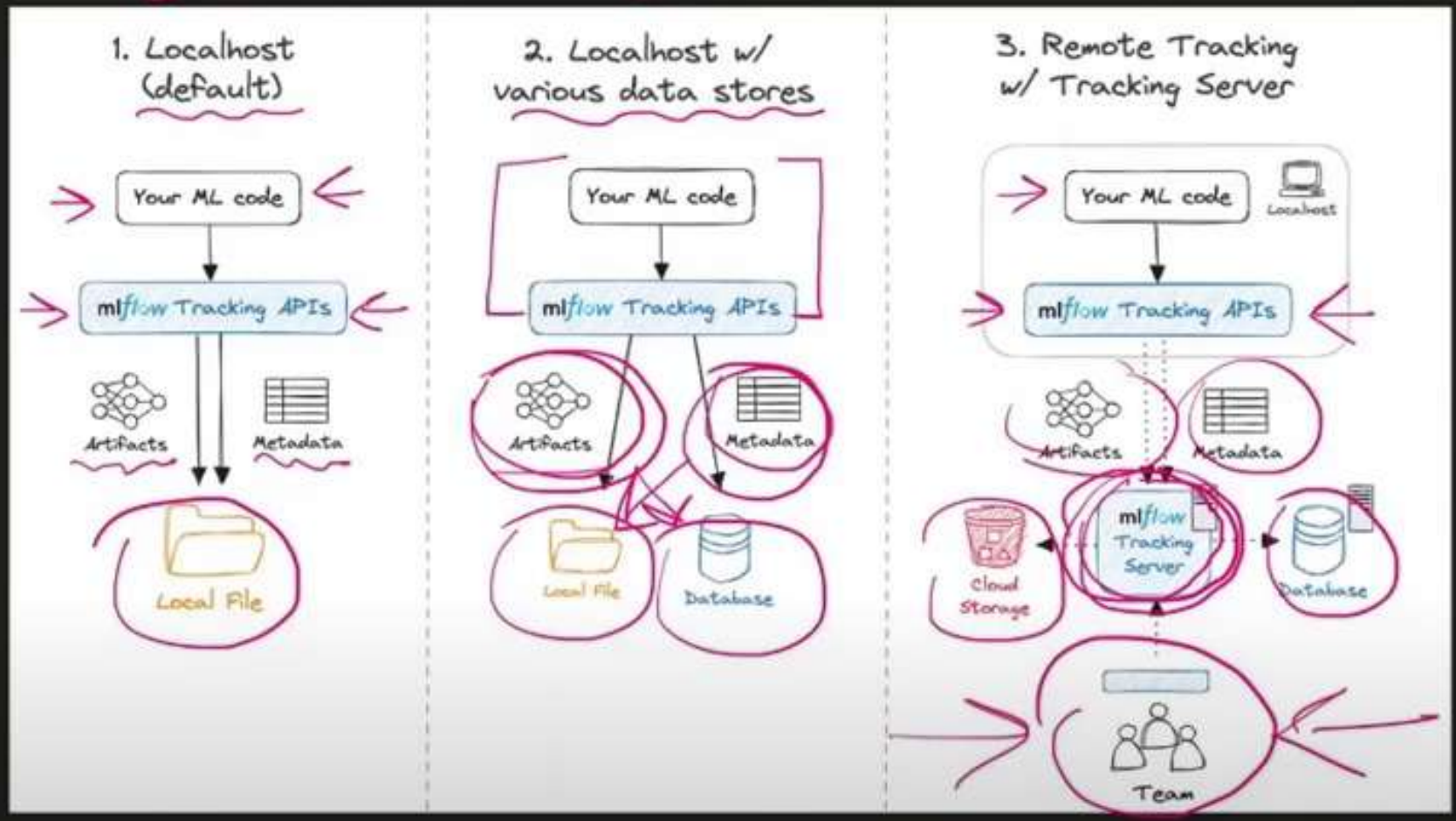
AGENDA

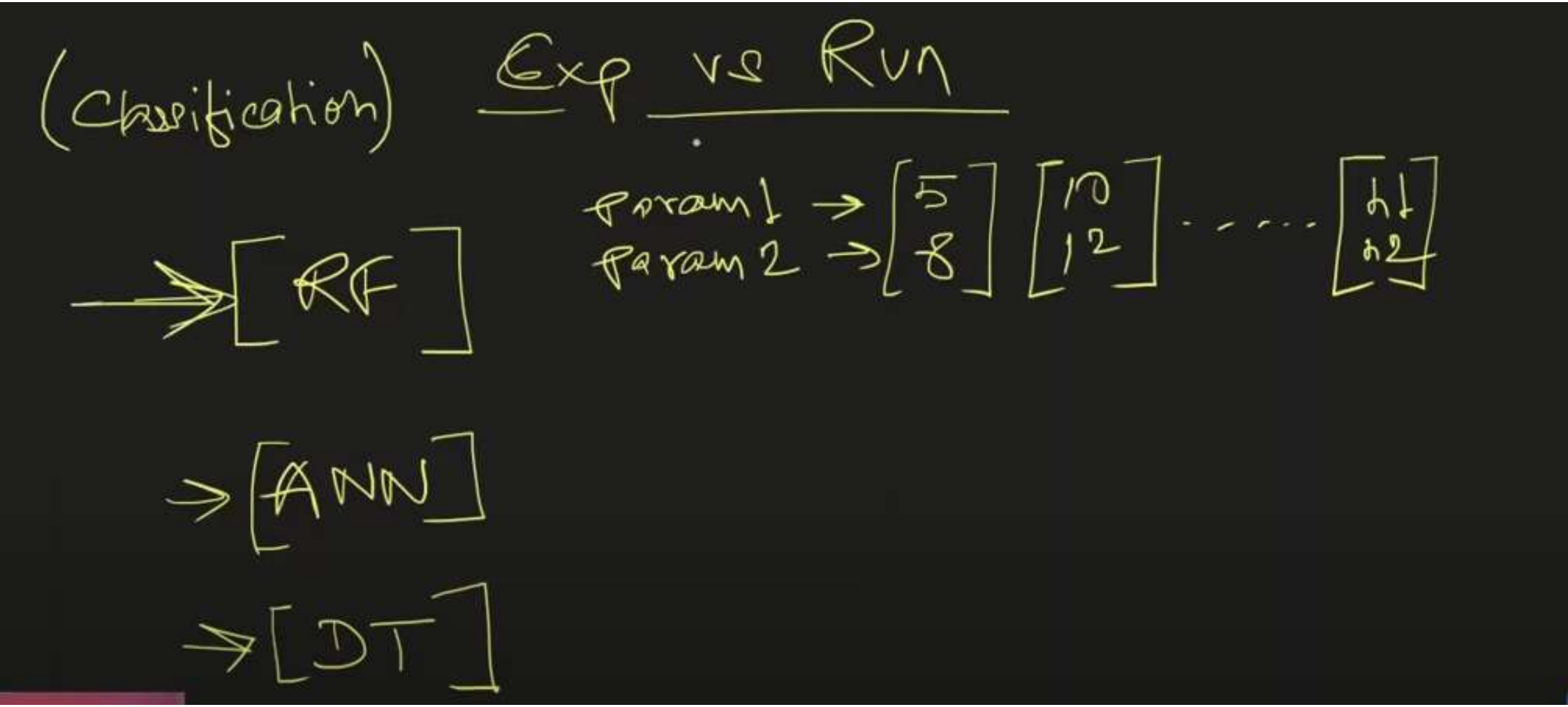
1. Experiments overview.
2. DVC vs MLflow.
3. MLflow for Data Science.
4. Code demonstration.
5. MLflow server architecture (AWS vs Dagshub)
6. Model Registry.

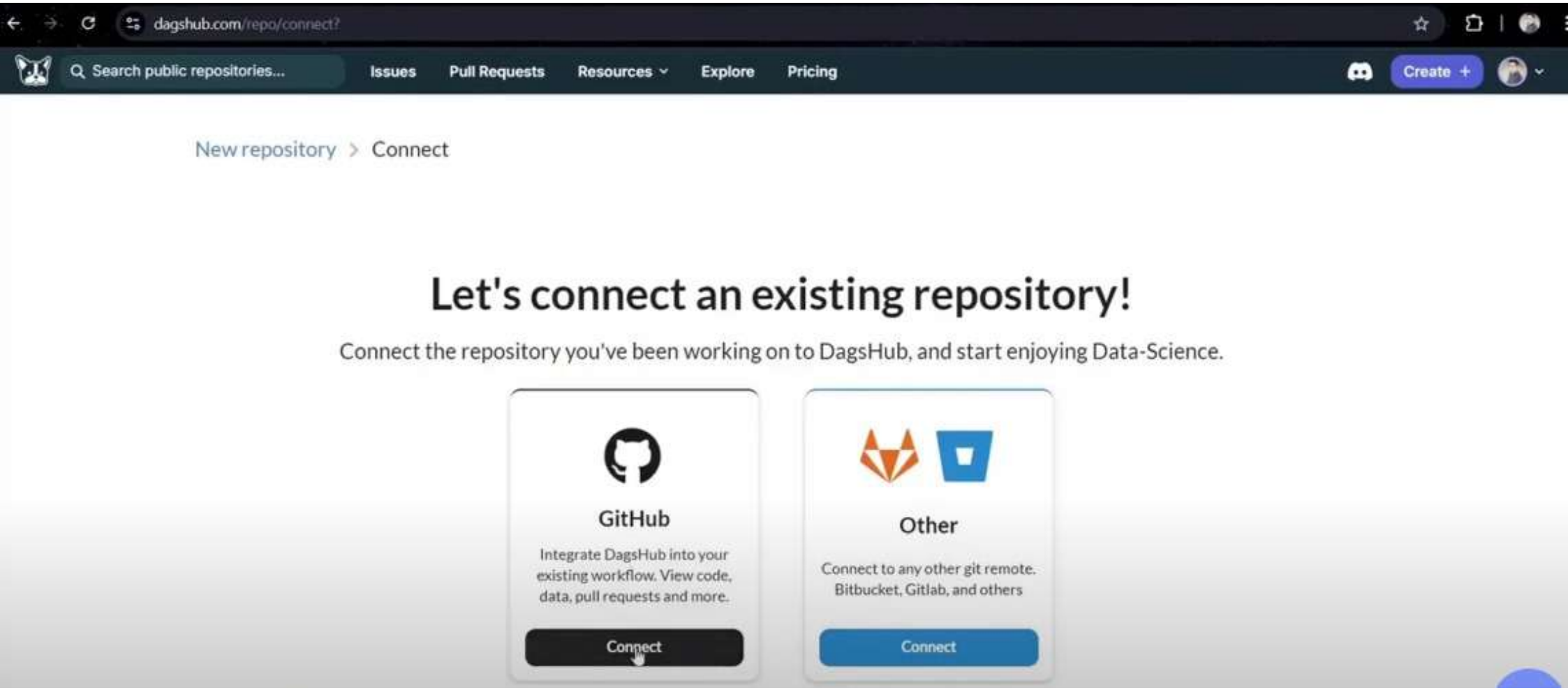

Reg / Classify
Pre-process → T1, T2, T3
Feature Engg → T1, T2, T3
Model → ANN / ML
ML → LR / SVM / DT
> Best Model <
nb1
nb2
nb3
6-8


mlflow.org/#core-concepts
mlflow
Blog
Releases
Docs
Contribute
Ambassador Program
Get Started
solve complex, real-world challenges
Experiment tracking
Visualization
Generative AI
Evaluation
Models
Model Registry
Serving


mlflow 2.15.1
Experiments
Models
GitHub
Docs
Experiments
Search Experiments
Default
Default
Provide Feedback
Add Description
Share
Runs
Evaluation Experimental
Traces Experimental
metrics.rmse < 1 and params.model = "tree"
Time created
State: Active
Datasets
New run
Sort: Created
Columns
Group by
Run Name
Created
Dataset
Duration
Source
Models
painted-whale-913
1 minute ago
1.5s
file1.py
merciful-mouse-265
5 minutes ago
0.7s
file1.py
victorious-fowl-132
8 minutes ago
1.6s
file1.py
hilarious-wolf-555
17 minutes ago
200ms
file1.py
stylish-lark-105
22 minutes ago
0.9s
file1.py


1. Localhost (default)
Your ML code
mlflow Tracking APIs
Artifacts
Metadata
Local File
2. Localhost w/ various data stores
Your ML code
mlflow Tracking APIs
Artifacts
Metadata
Local File
Database
3. Remote Tracking w/ Tracking Server
Your ML code
Localhost
mlflow Tracking APIs
Artifacts
Metadata
mlflow Tracking Server
Cloud Storage
Database
Team
AWS Dagshub


(Classification)
Exp vs Run
param1 → [5] [10] ..... [n1]
param2 → [8] [12] ..... [n2]
[RF]
[ANN]
[DT]


dagshub.com/repo/connect?
Search public repositories...
Issues
Pull Requests
Resources
Explore
Pricing
Create +
New repository > Connect
Let's connect an existing repository!
Connect the repository you've been working on to DagsHub, and start enjoying Data-Science.
GitHub
Integrate DagsHub into your existing workflow. View code, data, pull requests and more.
Connect
Other
Connect to any other git remote. Bitbucket, Gitlab, and others
Connect


mlflow 2.15.1
Experiments
Models
GitHub
Docs
breast-cancer-rf-hp ›
Comparing 13 Runs from 1 Experiment
Visualizations
Parallel Coordinates Plot
Scatter Plot
Box Plot
Contour Plot
Parameters:
max_depth ×
n_estimators ×
Metrics:
accuracy ×
Clear All
max_depth
n_estimators
accuracy
None
30
20
10
100.00000
80.00000
60.00000
40.00000
20.00000
10.00000
0.95824
0.95500
0.95000
0.94500
0.94000
0.93500
0.93187
0.955
0.95
0.945
0.94
0.935
VIKAS DAS.

Experiment vs Run

Exp is like experimenting using a DL model, but run means that using that DL model, we are changing parameters, data, etc., to compare performance and get the optimal solution. So, an experiment has multiple runs.

That's why we use MLflow.

If we go for experiment tracking with AWS, then we have to first create IAM roles or users for different users for collaborative experimentation and tracking, then create an EC2 remote server to host this experimentation and tracking and also to store metadata, and finally create an S3 bucket to store artifacts.

VISIT IT --> https://chatgpt.com/share/6840b0d9-2afc-8007-aa39-e34bd2159768