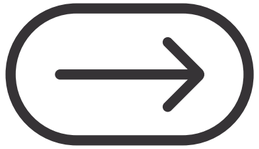
PREDICTING PEDIATRIC APPENDICITIS

Using AWS Cloud Based Machine Learning

DATS 6450 Cloud Computing - Group3
December 1, 2025

Presented By: Abirham Getie, Haeyeon Jeong, Yonathan Shimelis



Quick Agenda

01 **Project Summary**

02 **AWS Architecture + Team Roles**

03 **Demo Workflow**

04 **Results & Takeaways**

Project Summary

Goal

- **Build an end-to-end AWS ML pipeline to predict pediatric appendicitis.**

Main AWS Services

- **S3**: Shared central storage for datasets, model results, and visuals.
- **EC2 with Shared AMI** (Jupyter): Common environment for preprocessing and model training
- **IAM Policies**: Enable cross-account read/write to S3 for all team members.

Pipeline

- **Data Load → Preprocess → Model → Evaluate → Visualize**

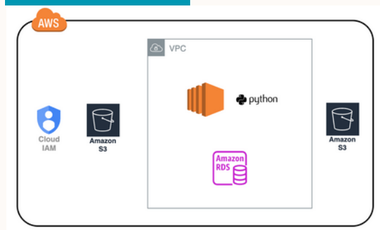
Data Source

- **Dataset: Regensburg Pediatric Appendicitis** (**UCI ID 938**)
  - 782 rows | 53 features | Target = Diagnosis (Appendicitis / No)
  - Variables: WBC, CRP, Appendix Diameter, Ultrasound findings

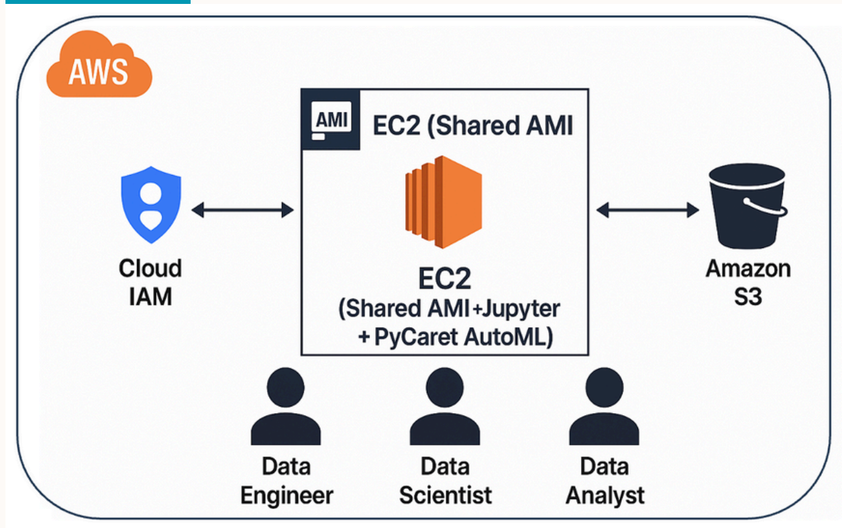
AWS Architecture + Team Roles

Services Used

Original



Improved



Used: S3, EC2 (**Shared AMI** + Jupyter +**PyCaret AutoML**), IAM Policy

Removed: RDS

Role Summary

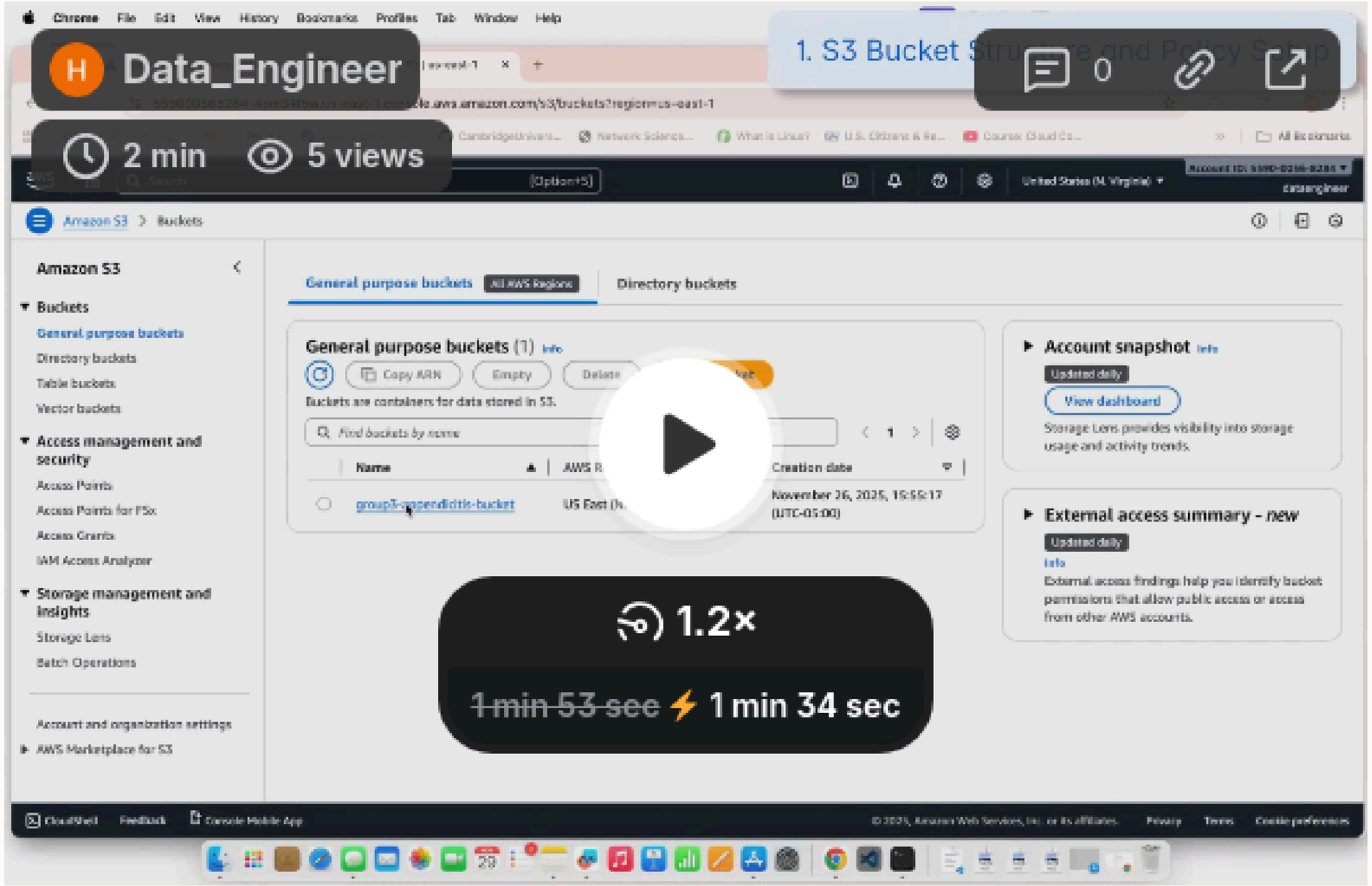
Role	Main Tasks
Data Engineer	• Creates shared AMI • Sets IAM policy • Uploads raw dataset to S3
Data Scientist	• Read the data from S3 and preprocesses data • Trains ML models and saves results to S3
Data Analyst	• Reads model results from S3 • Visualizes performance • Uploads visuals back to S3

https://www.loom.com/share/7fae3d4e12274e4289ddfe67eee854a3

Demo Workflow

01 Data Engineer

- Check 3 bucket structure + Policy
- Launch EC2 instance from the shared AMI
- Open Jupyter Notebook and run the code
- Confirm collected dataset uploaded to S3



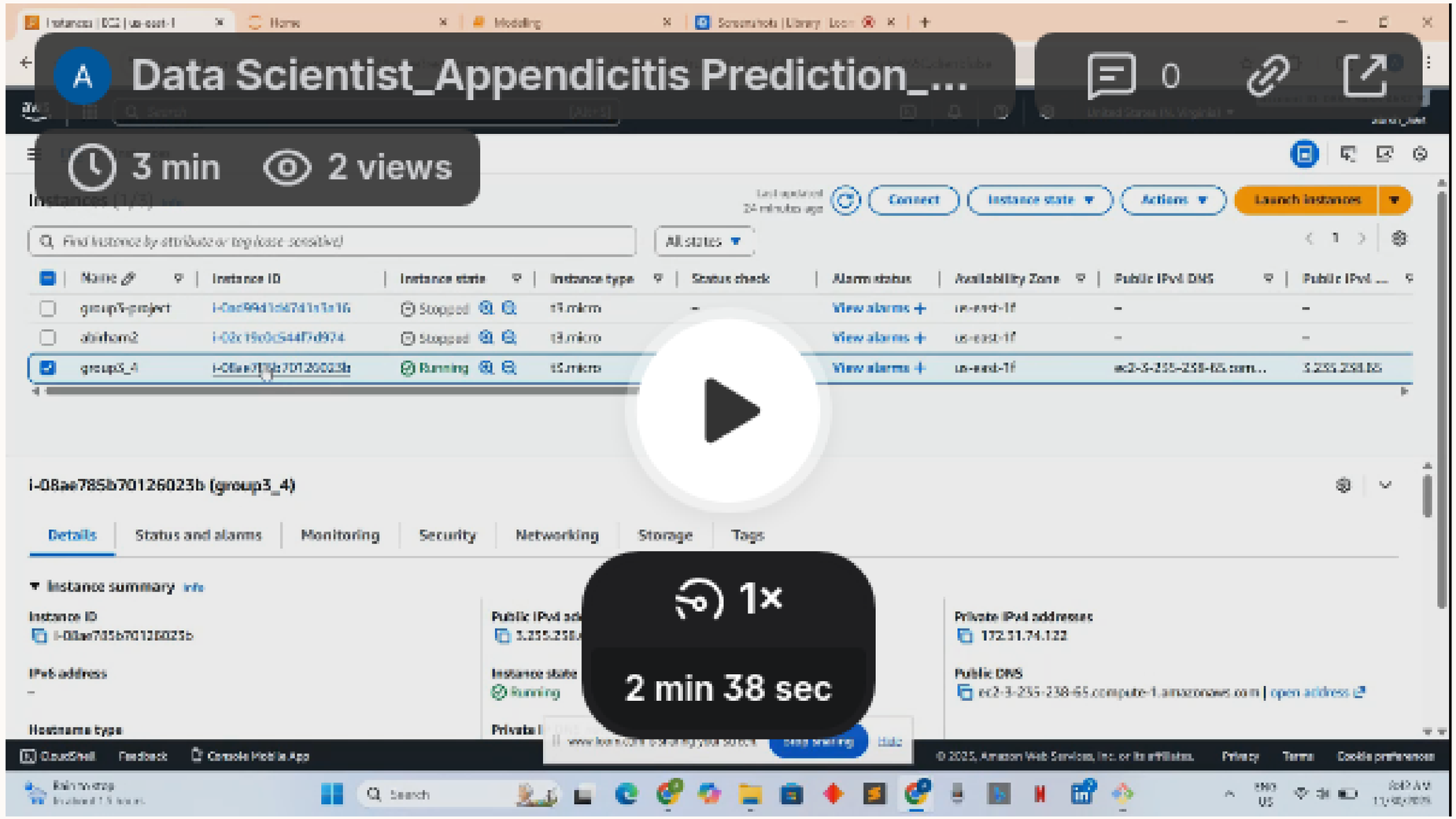
Demo Workflow

https://www.loom.com/share/6609e95ea75b40918c3afa6e61185093

02

Data Scientist

- Use Jupyter Notebook to load data from S3
- Train and evaluate ML models on EC2
- Used AutoML to identify best performing model
- Save model results to S3



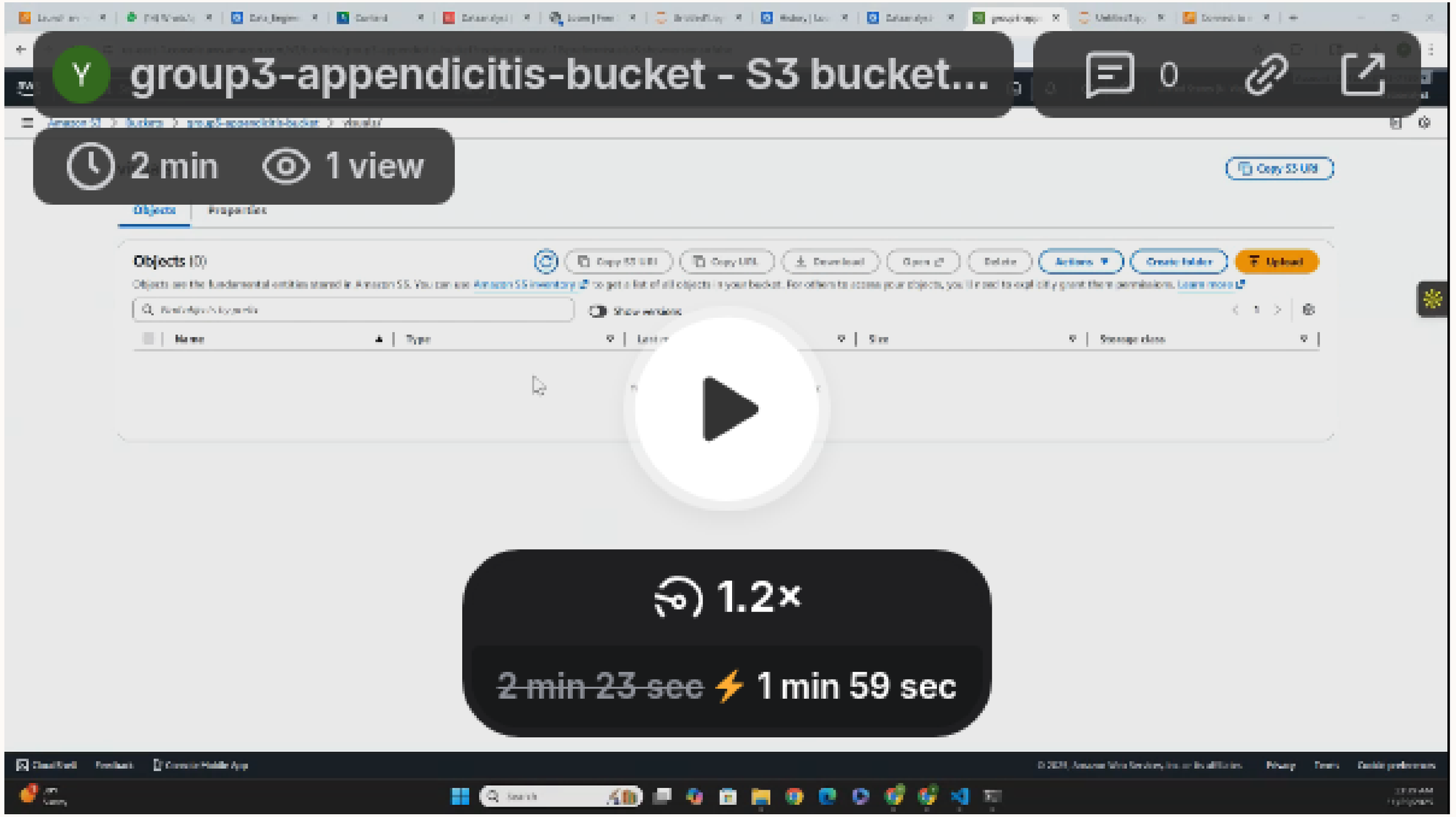
Demo Workflow

https://www.loom.com/share/eccb860892134ec197c660359b8c1a60

03

Data Analyst

- Access model results csv from S3
- Create performance charts and comparison plots
- Upload visuals to S3



Results & Takeaways

Outputs

Folders in our S3 bucket include:

- data/ → cleaned datasets
- models/ → stored models
- results/ → model performance CSV
- visuals/ → plots and charts

Outcomes:

- Fully shared AWS workflow across 3 accounts
- End-to-end automation of data cleaning and upload
- Collaboration through S3 instead of local files

Modeling takeaways

- Key features for model prediction include appendix diameter, white blood cell count, and C-reactive protein measurements
- Used AutoML to evaluate a broad range of classification models and the search identified HistogramGradientBoostingClassifier as the best-performing model with a CV F1 Macro of 0.95

Conclusion

- Demonstrated how S3 and EC2 enable team-based machine learning.
- Achieved secure, reproducible workflow without extra services like RDS.
- Future step: deploy model for real-time prediction using AWS Lambda.

Thank You

Group 3: Abirham Getie, Haeyeon Jeong, Yonathan Shimelis