

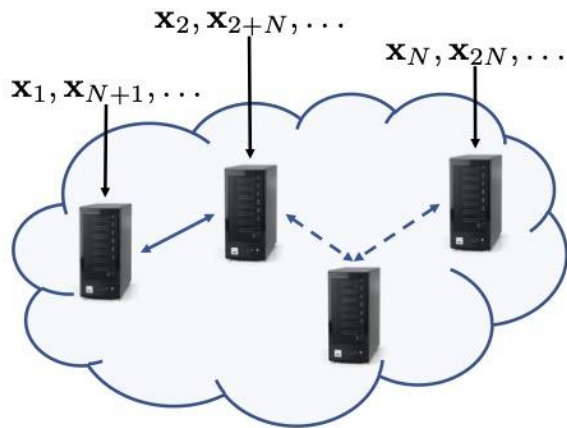
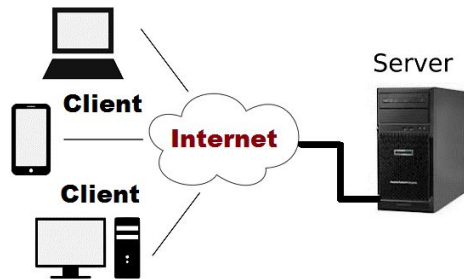
Real-time, Robust, and Reliable (R^3) Machine Learning over Wireless Networks

Team: Joshua Menezes, Nihal Abdul Muneer, Ayaan Qayyum,
Madhav Subramaniam, Hasan Ali

Prof. Anand Sarwate, Prof. Waheed Bajwa

Real-time, Robust, Reliable ML

- Learning 2 Help
 - Client - Server architecture
- Collaborative Feature Extraction
 - Distributed Network



L2H



Joshua Menezes
Rutgers ECE (G)



Madhav
Subramaniyam
Rutgers SAS (UG)

The Team

CFE



Nihal Abdul Muneer
Rutgers ECE (G)



Ayaan Qayyum
Columbia EE (G)



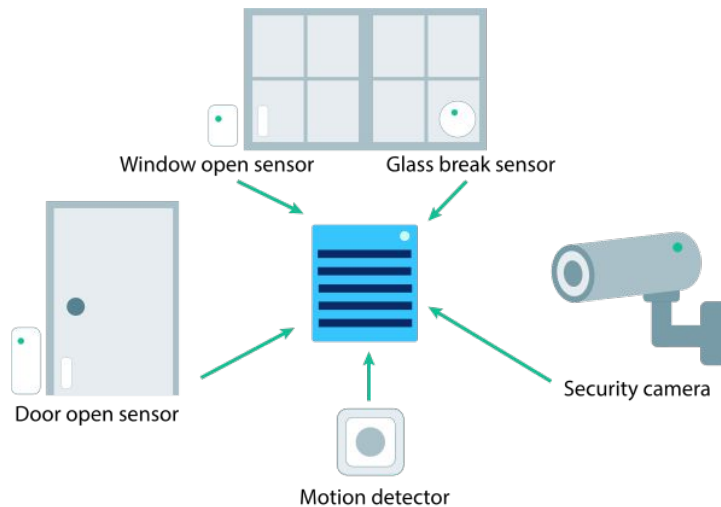
Hasan Ali
Rutgers SAS (UG)



Collaborative Feature Extraction

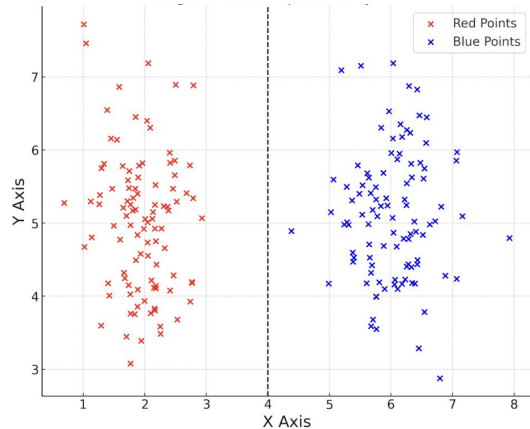
Motivation

- A lot of data science problems have data from many different sources. Some data is redundant.
- Today, lots of data is generated and streams from many computers (formally called “nodes”)
- **Our Work:** Help models by reducing redundancy through **Principal Component Analysis**
 - (Formally: data dimensionality reduction across distributed networks)

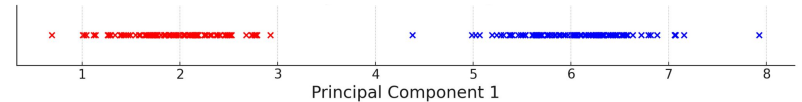


What is Principal Component Analysis?

Data compression technique that finds lines of best fit in your data



Original 2-Dimensional Data



New 1-Dimensional Data

Streaming PCA

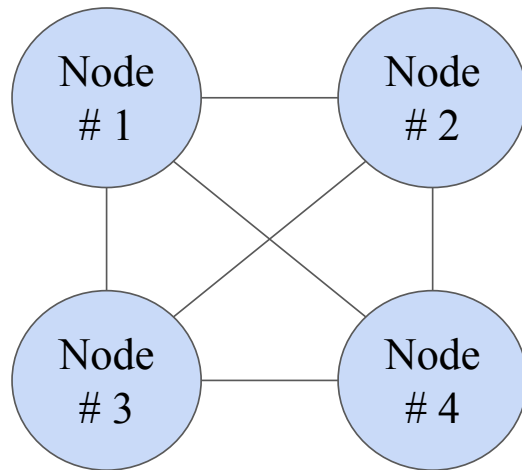
Krasulina's and Oja's method estimate the most significant principal components. Both are:

- **Streaming algorithms**
 - They process one data sample at a time
 - Make a better estimate with new data
- **Single-node algorithms**
 - Computers cannot work together
- **Designed for online learning**
 - Data that comes in continuously (sensors, logs, live feeds)

Note: Oja's method is an alternative to Krasulina. Both are very similar.

Our Work

- Extended Krasulina's method to work on **multiple** computer "nodes"
 - Nodes get help from their neighbors to improve their estimate
 - Ideal for sending and receiving a lot of data (*high-rate data streaming*)
- Tested how network parameters (*bandwidth, packet size*) effect convergence



Four computer nodes
connected in mesh

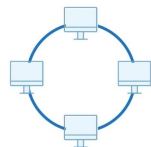
Results

	Top 1 Principal Component	Top K Principal Components
Mesh	D & DM-Krasulina ✓	DiSK ✓
Partial Graph	C-DiEGO ✓	C-DiSK

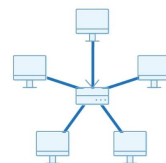
Tested on mesh, star, and other network topologies with wireless testbed



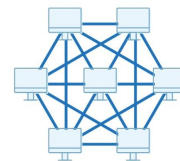
Bus



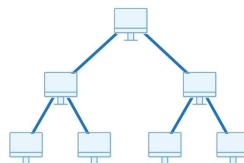
Ring



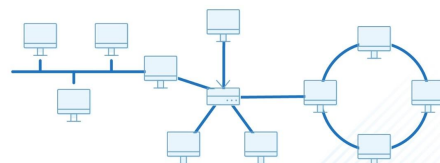
Star



Mesh

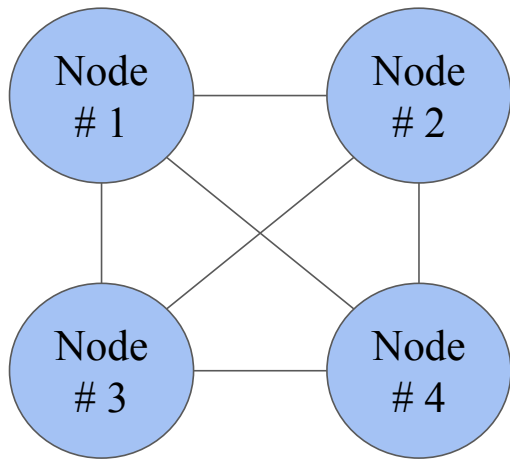


Tree

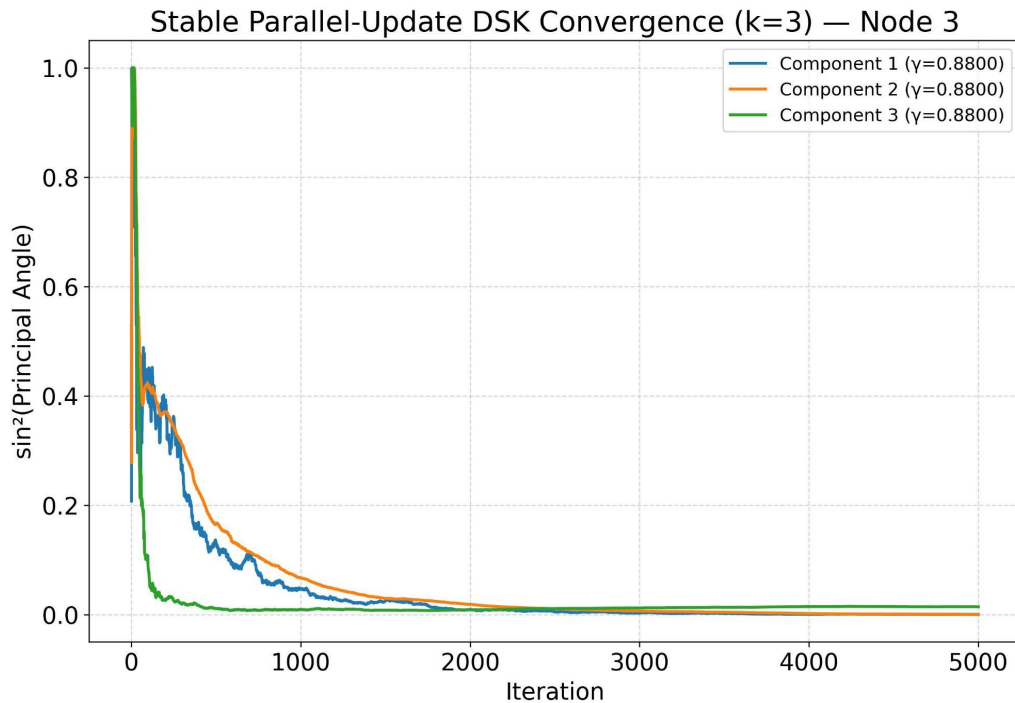


Hybrid

Results



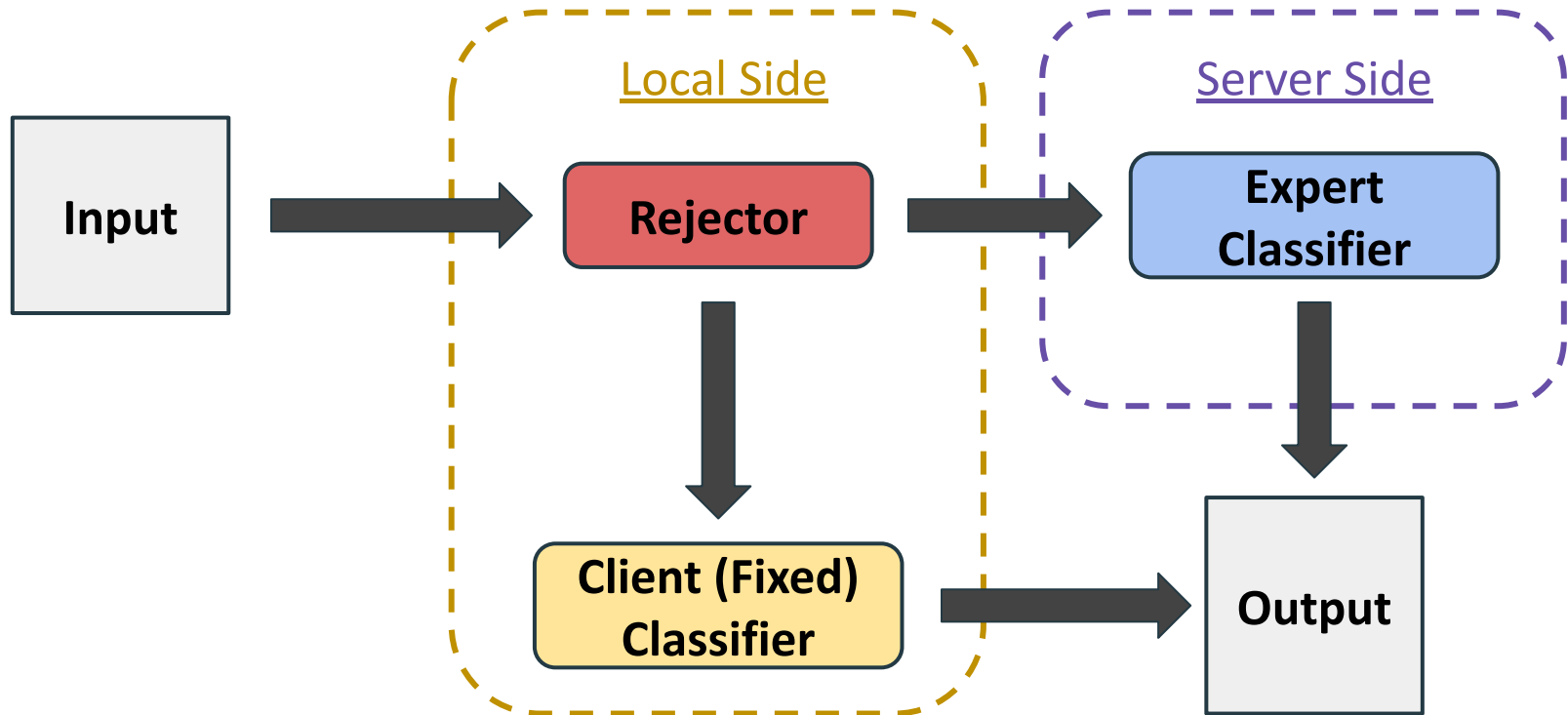
Computer node
connectivity
configuration



Distributed Streaming Krasulina (DiSK) algorithm
convergence with 5,000 samples, 5 data dimensions, with
four nodes, in a complete graph topology

Learning to Help (L2H)

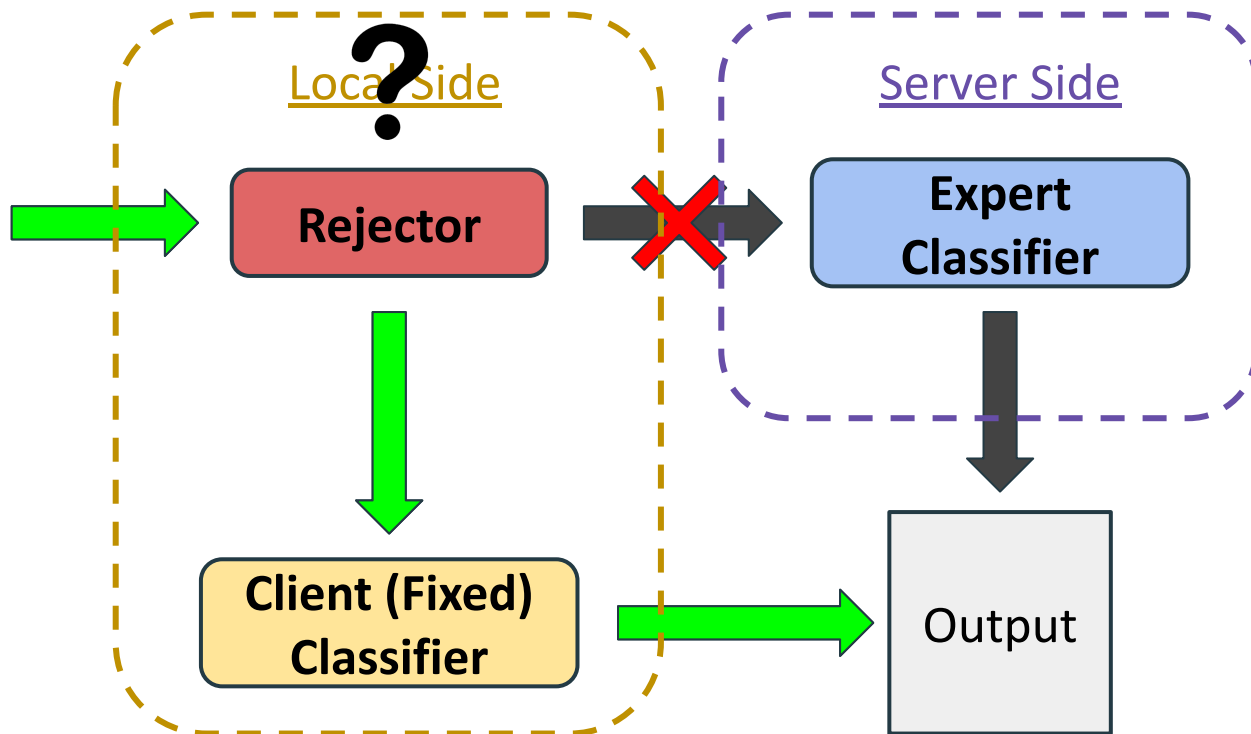
What is L2H?



What is L2H?



Ex Task:
Determine if
this is a dog

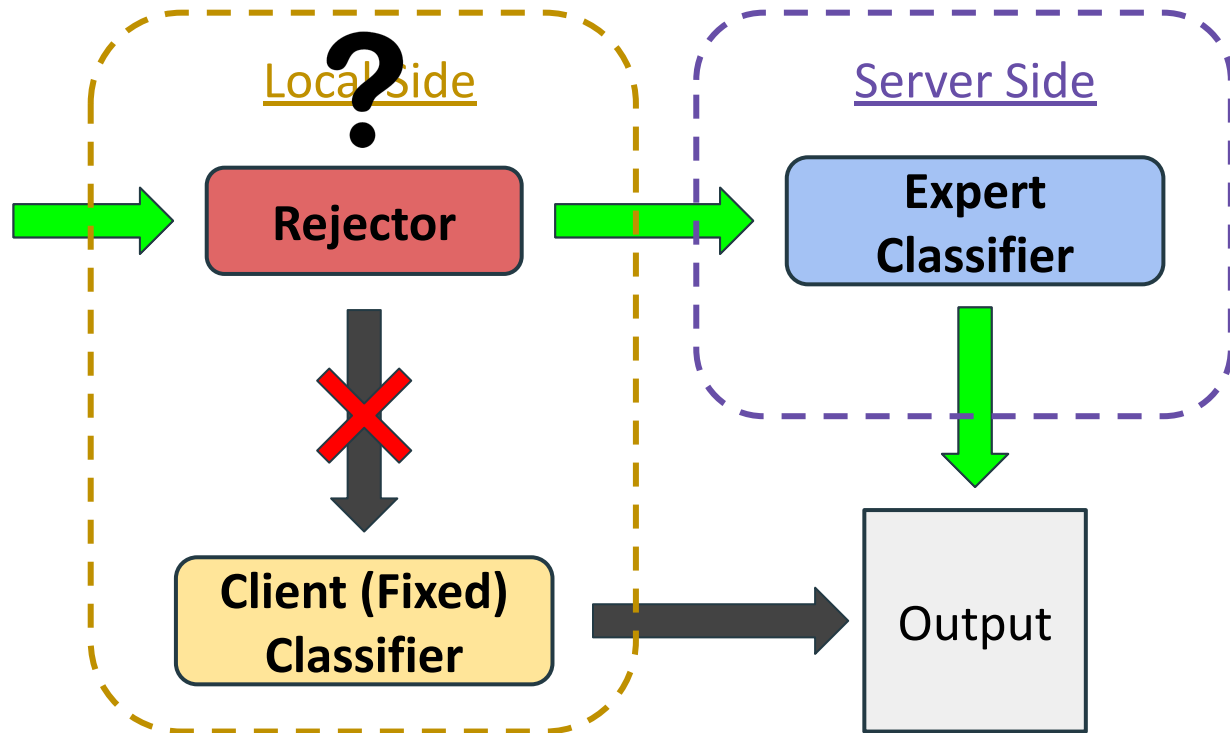


What is L2H?



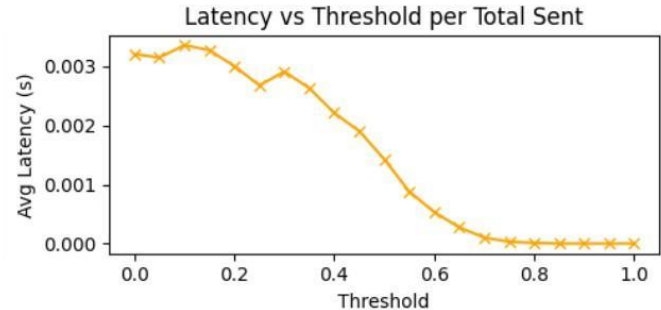
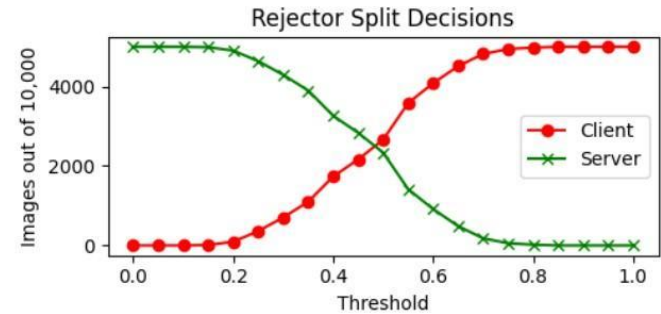
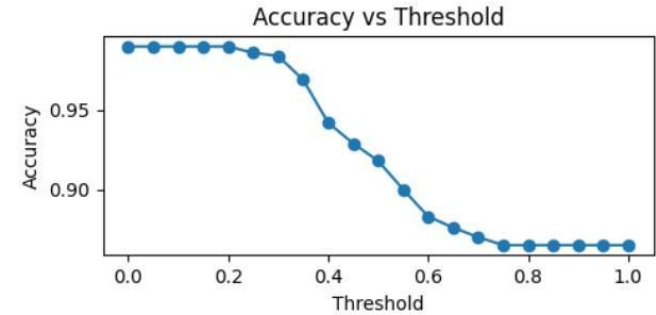
Ex Task:

Determine if
this is a dog



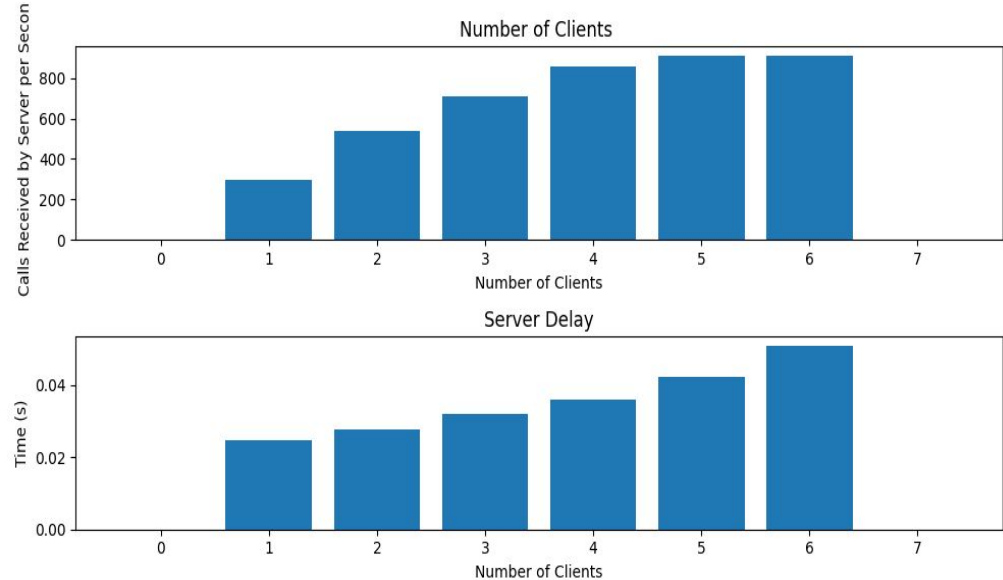
Results - System

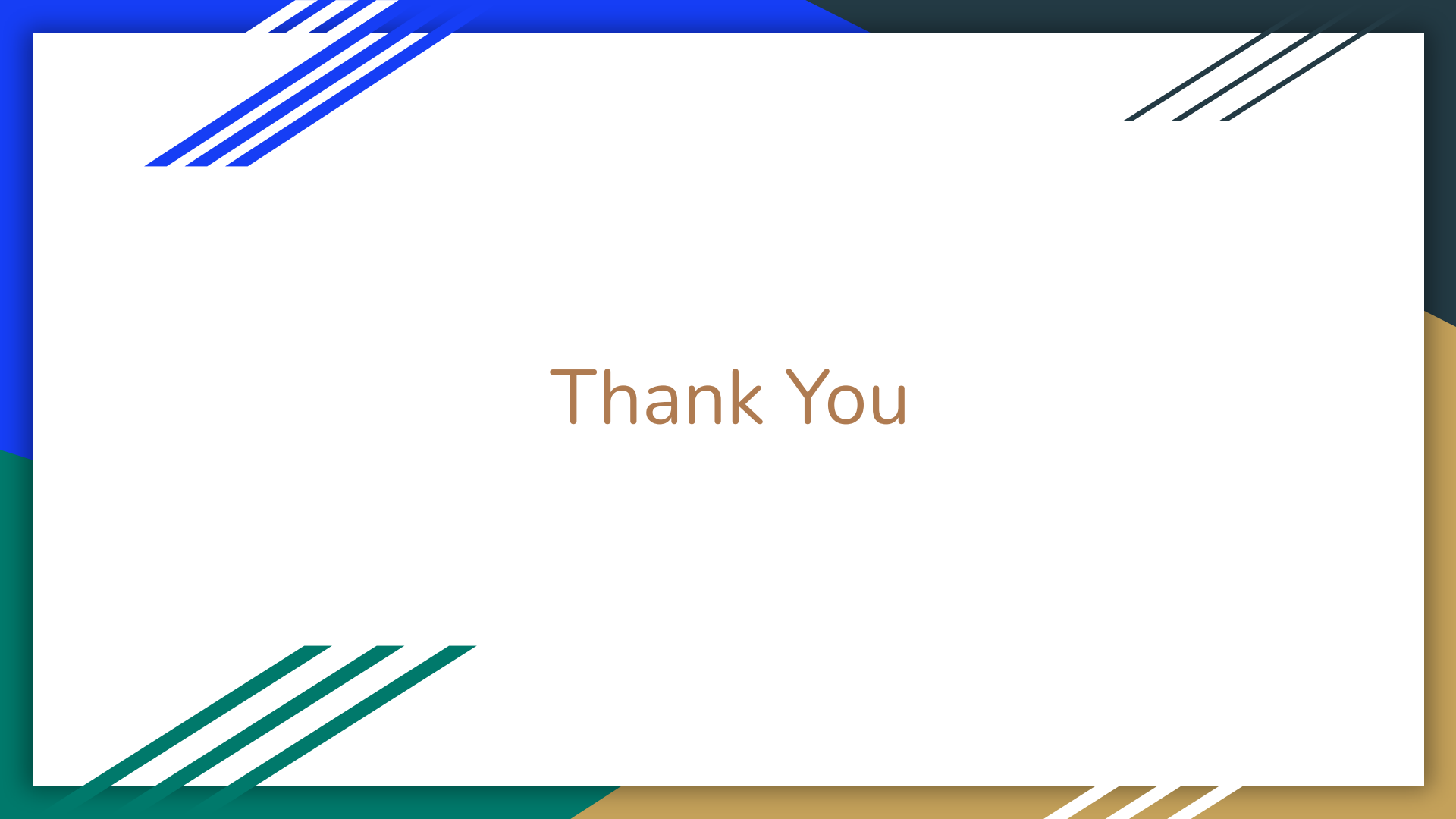
- Implemented the L2H system on a testbed
- Ran experiments
 - Tested different datasets
 - Tested different models
 - System Latency
 - System Accuracy



Results - Research

- Multiple clients per node
- Expert control over clients
- Drafted research paper





Thank You