ACTIVE LEARNING

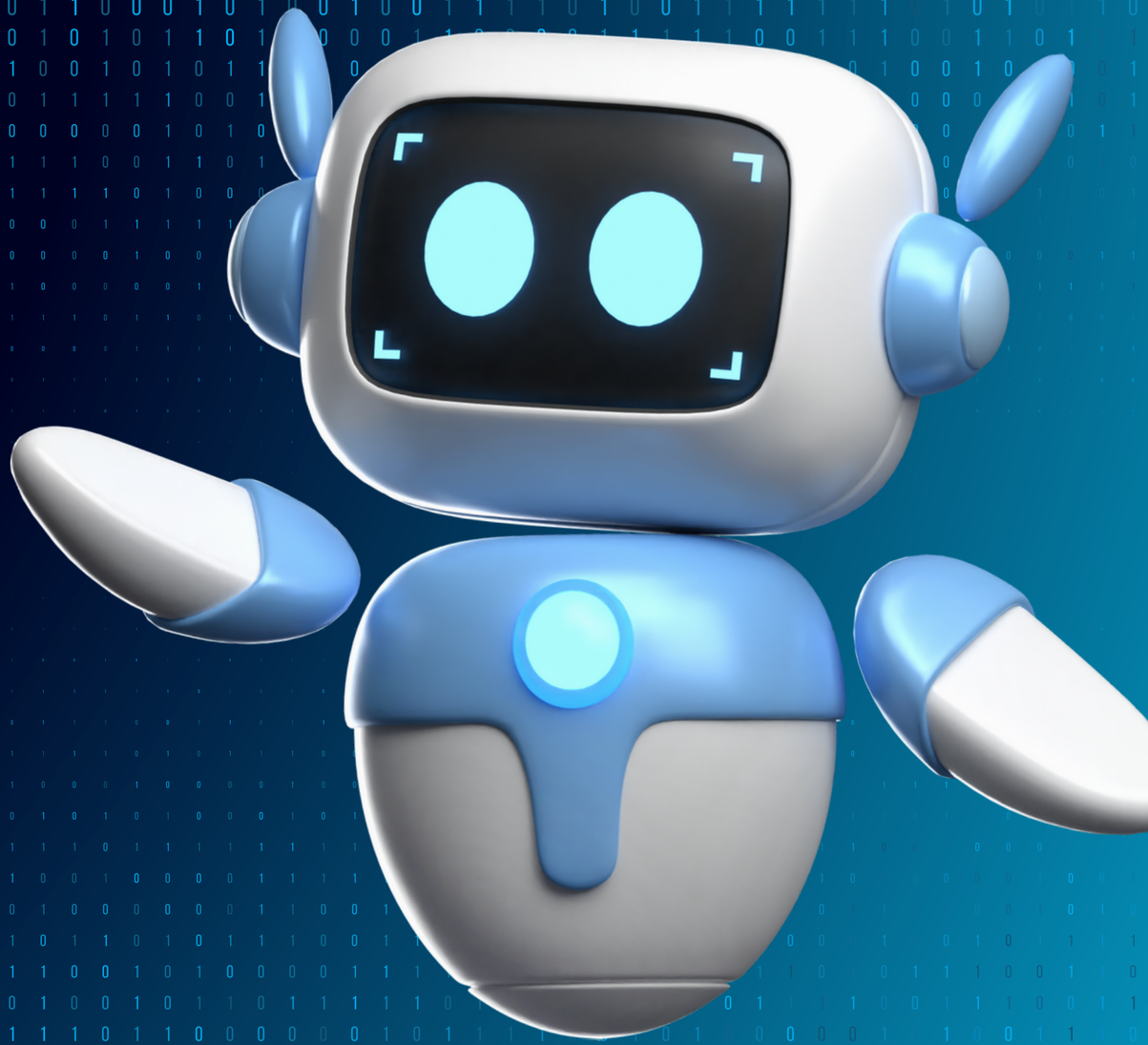


Table of content

Introduction

- AL Definition
- AL Scenarios
- AL Strategies

- Methodology
- Experimental Results

Deployment

GUI

Introduction

Definition

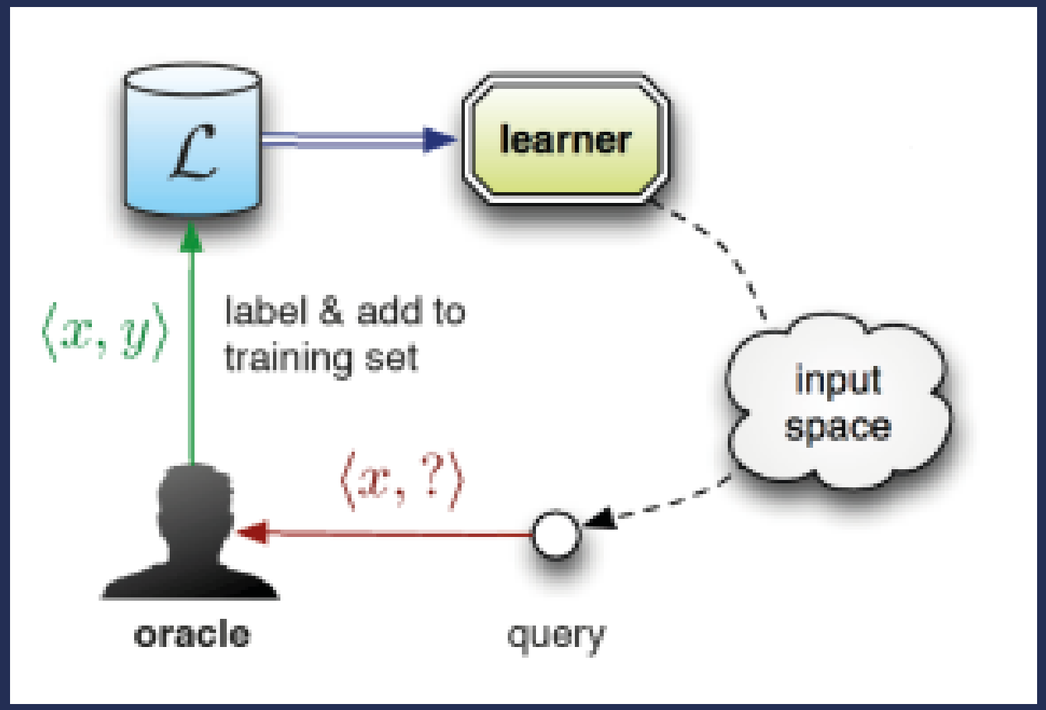
Active Learning is a machine learning approach where the algorithm selects the most informative data points, then queries an oracle (or some other information source) to label them with the goal of reducing the total amount of data needed to learn the model.

Objective

Learning process optimization by strategically selecting which data points to label.

Active learning scenarios

How Data Passed to Model (Nature of Data)



- **Pool-Based Active Learning**
- **Membership query syntheses**
- **stream based sampling**

Active Learning Strategies

Low Confidence Based Sampling

Margin Based Sampling

Entropy Based Sampling

Query by Commit

Methodology

- **Data Sets**
  - Balanced Datasets
    - Iris
    - digits
  - Imbalanced Datasets
    - BMI

- Used AL scenarios
  - **Pool-Based Active Learning**

- Appling 4 AL Sampling Strategies
- Applying Random Sampling
- Compare performance using Evaluation plotting

- BMI unbalancing dataset preprocessing using oversampling to make data nearly balanced and increase performance .

Experimental Results

Before active Learning

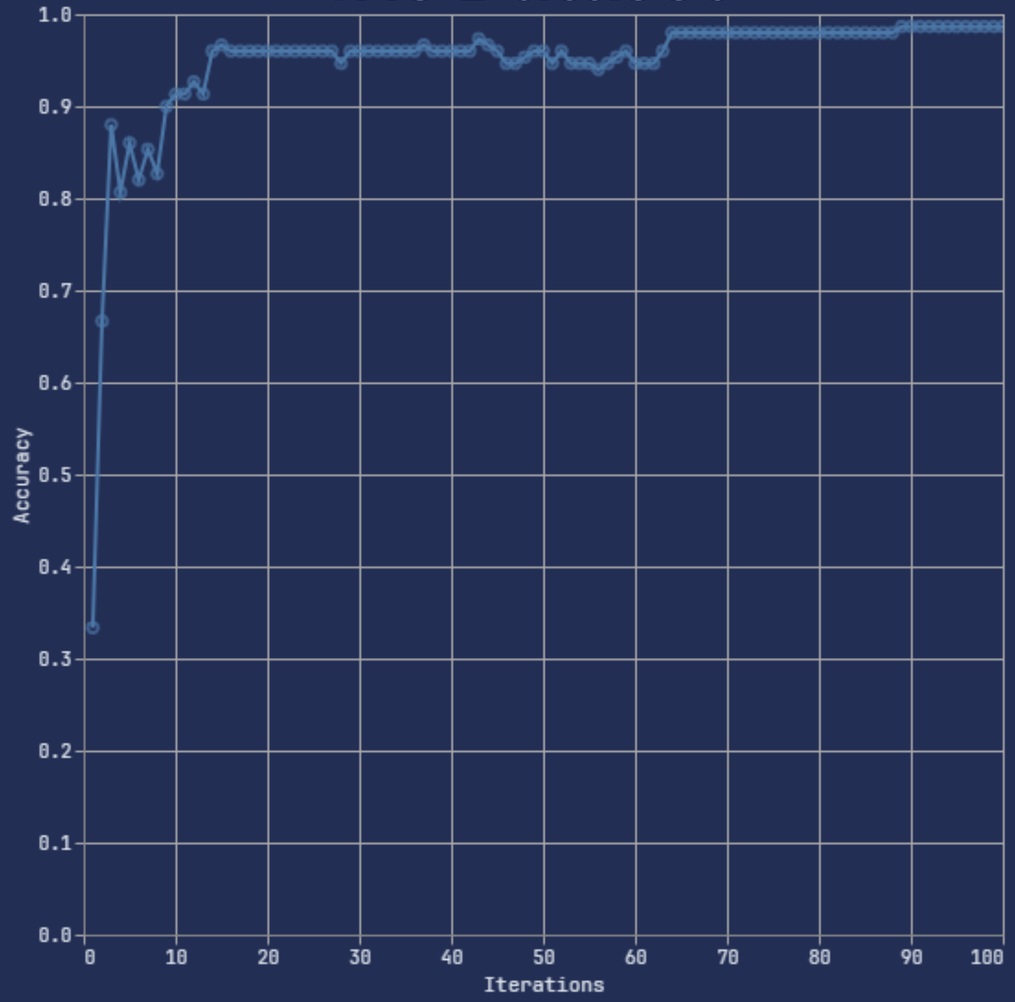
	Iris	Digits
Accuracy	%98	%93
F1-Score	%98	%93

After Active Learning

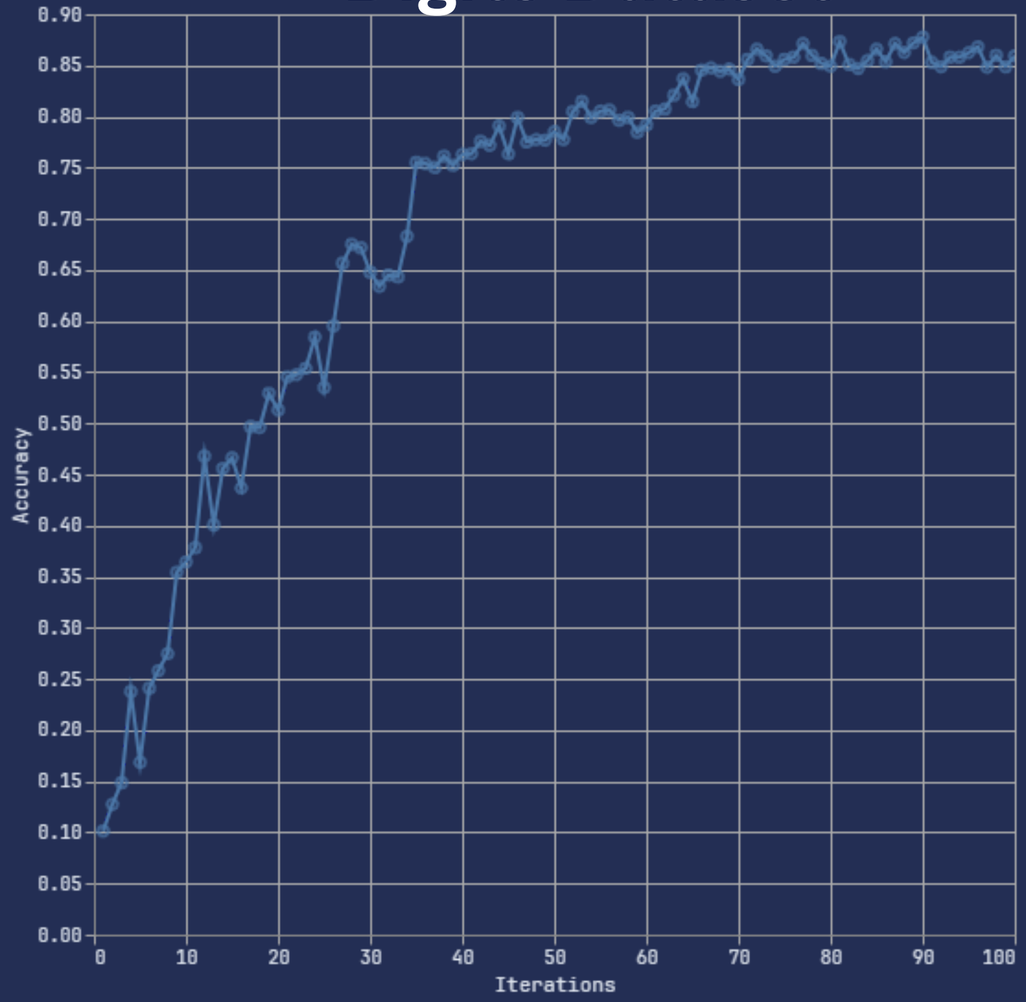
Results on Balanced Datasets

Random Sampling

Iris Dataset

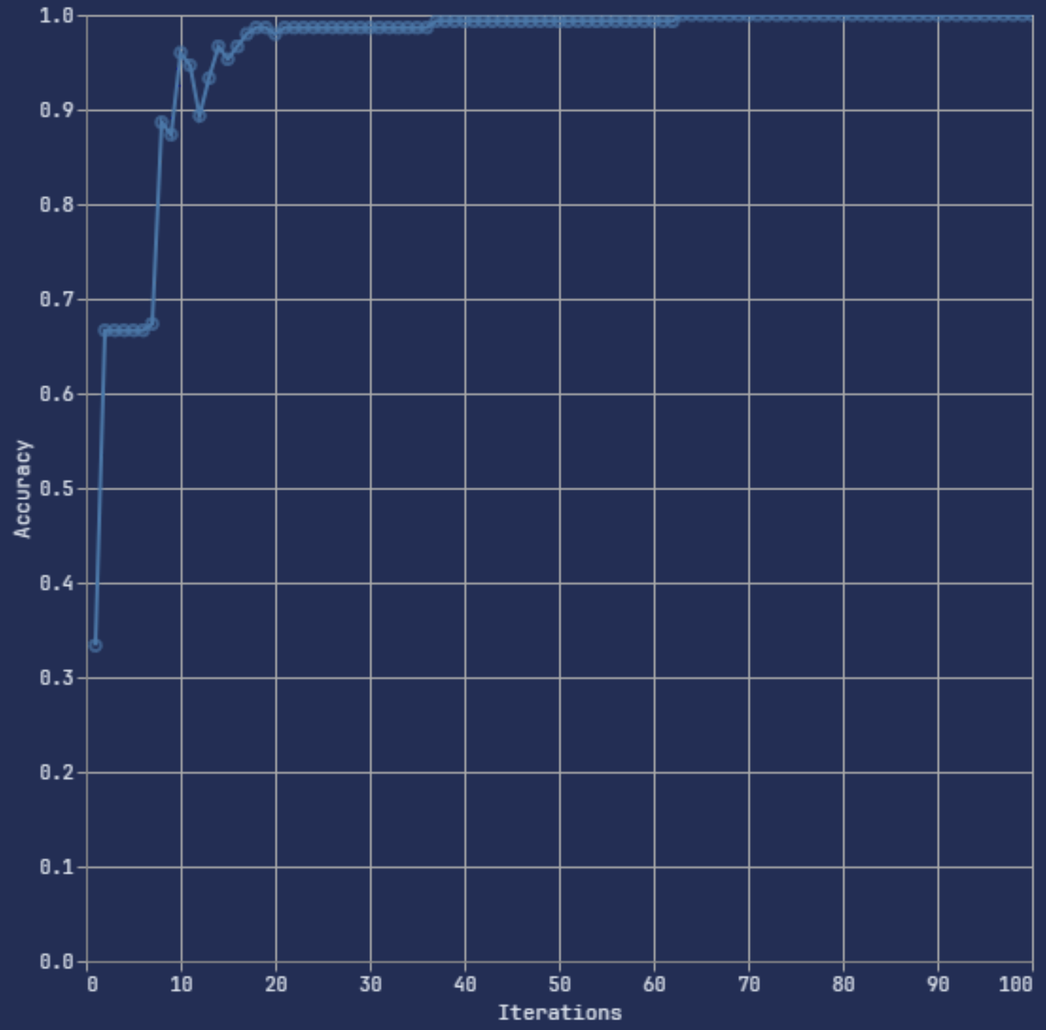


Digits Dataset

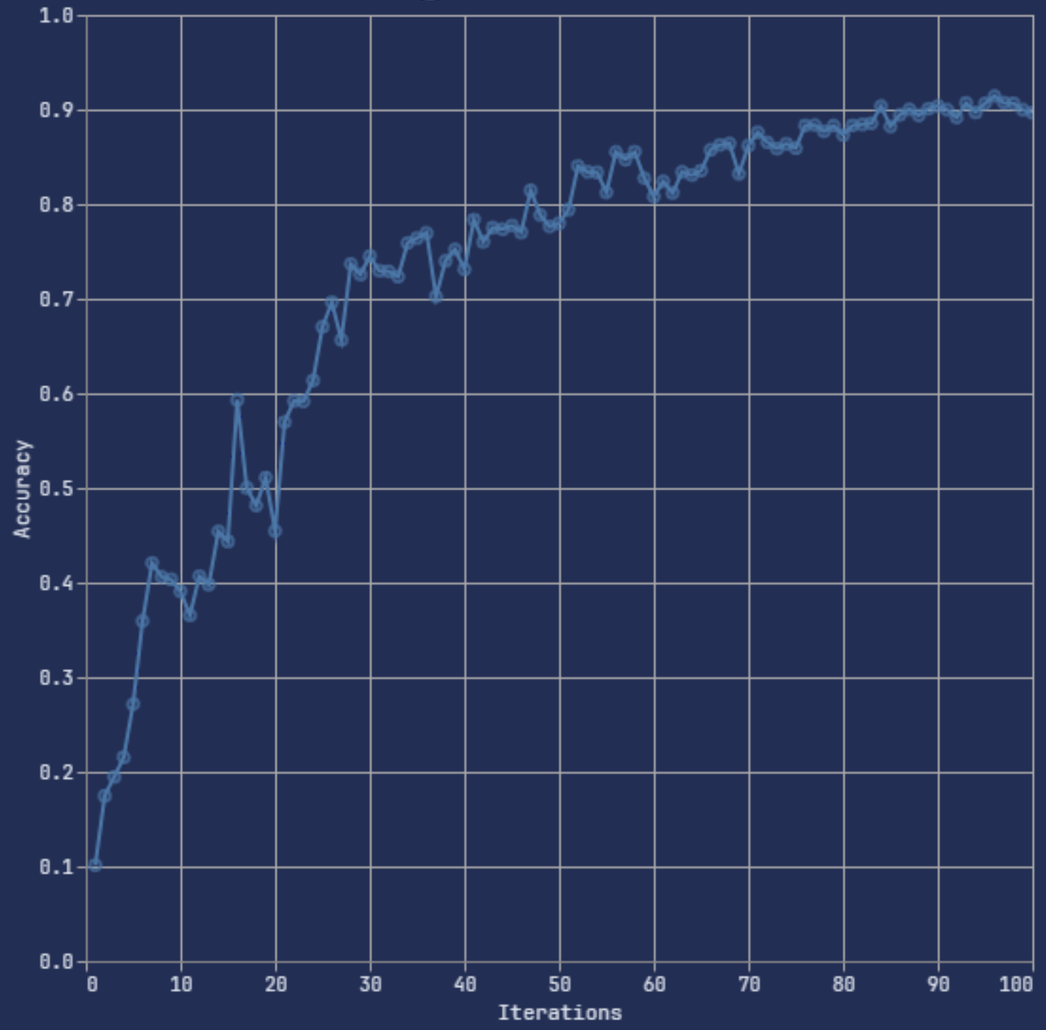


Query By Committee

Iris Dataset

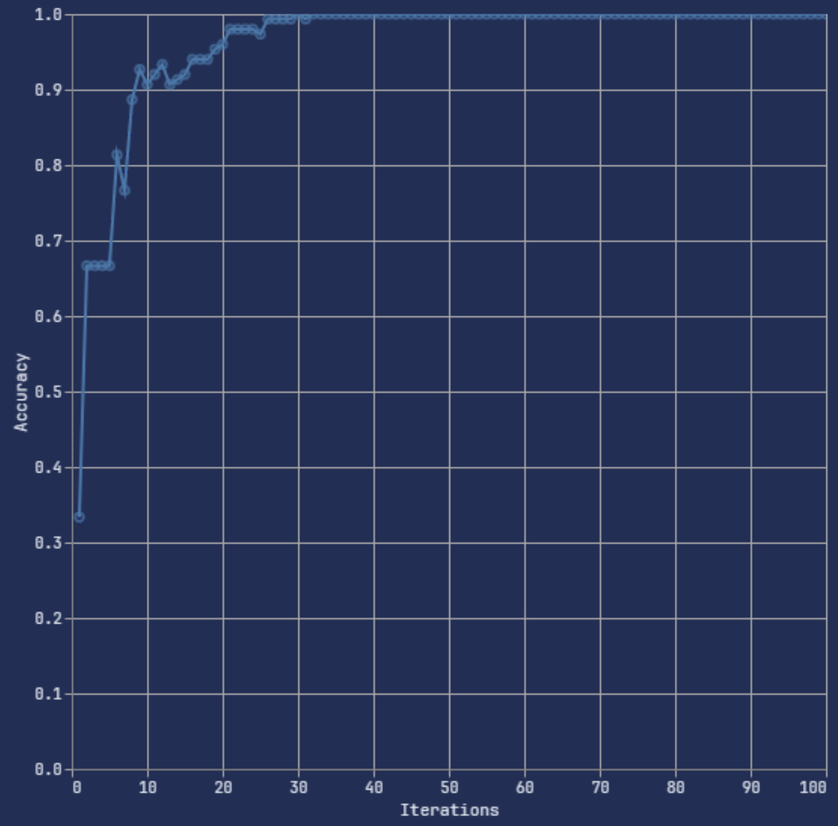


Digits Dataset

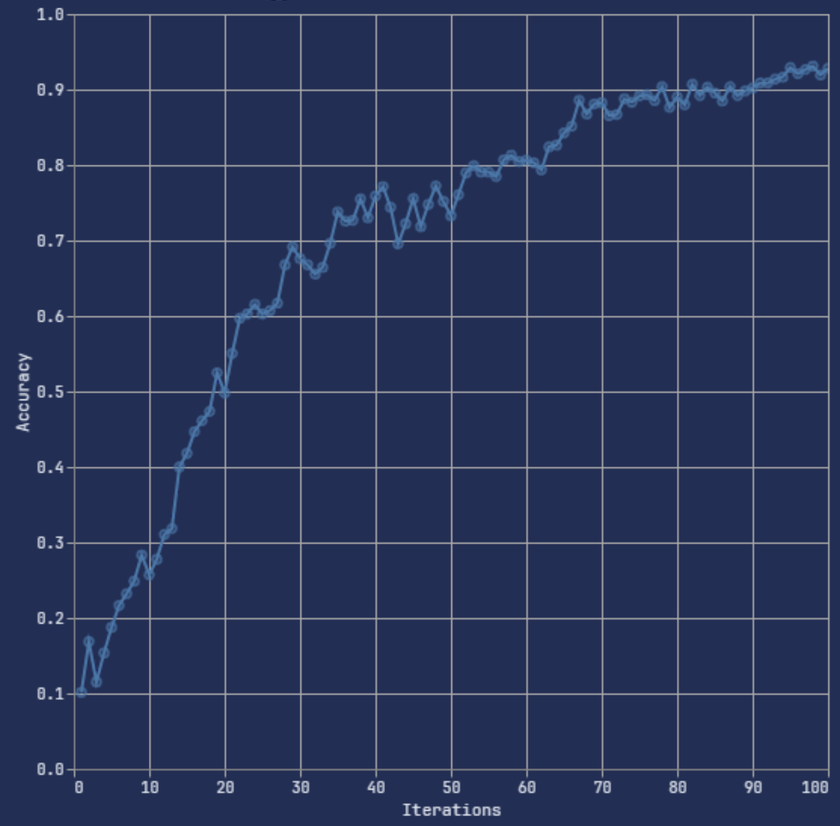


Uncertainty Sampling with Least Confident

Iris Dataset

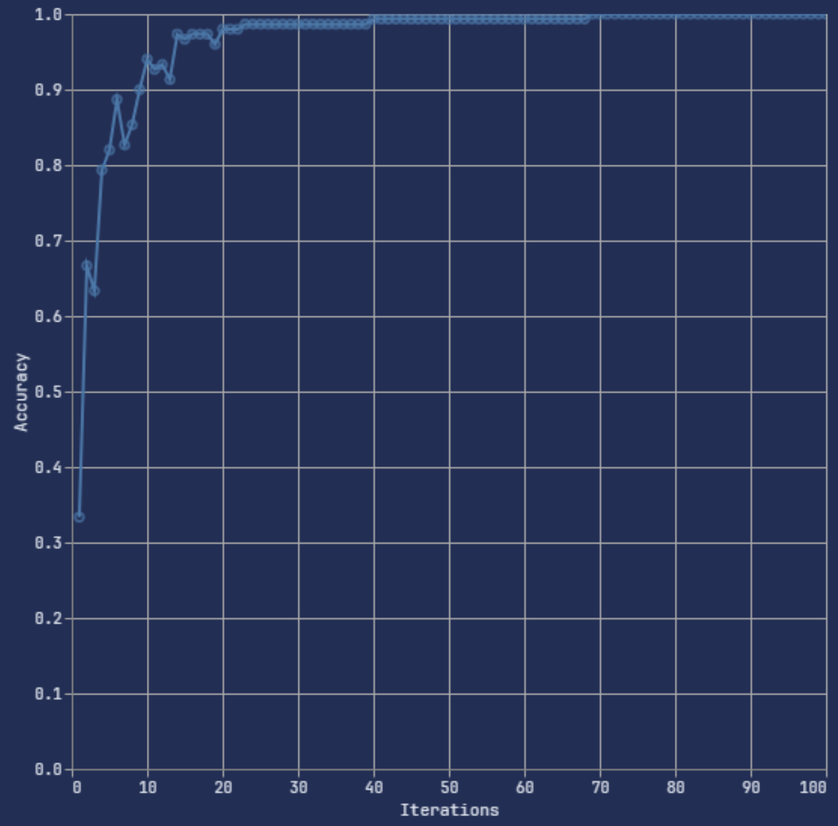


Digits Dataset

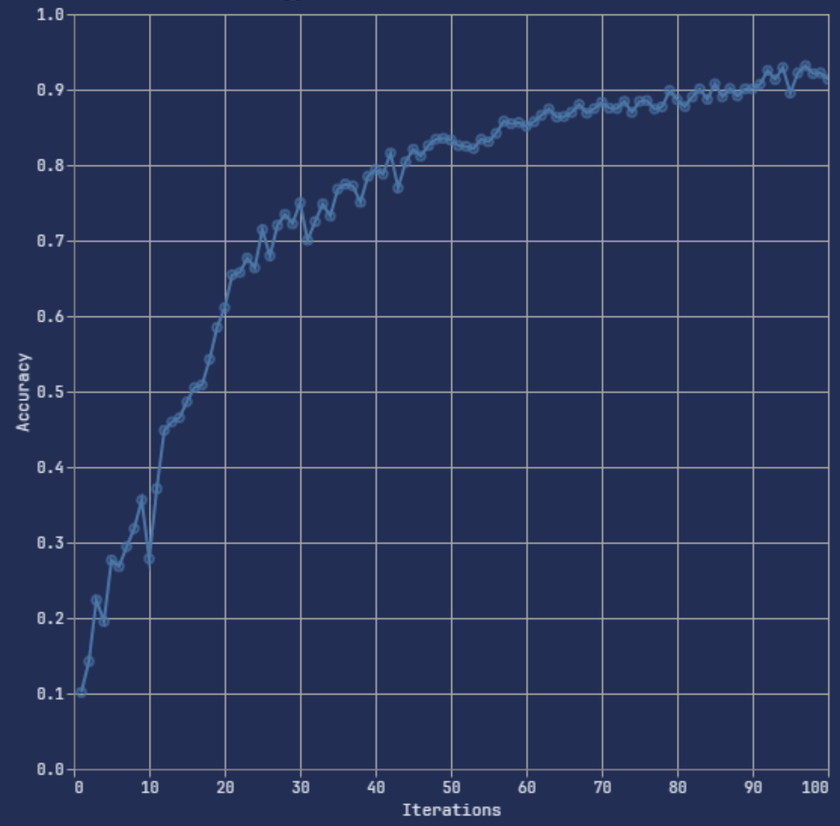


Uncertainty Sampling with Margin Sampling

Iris Dataset

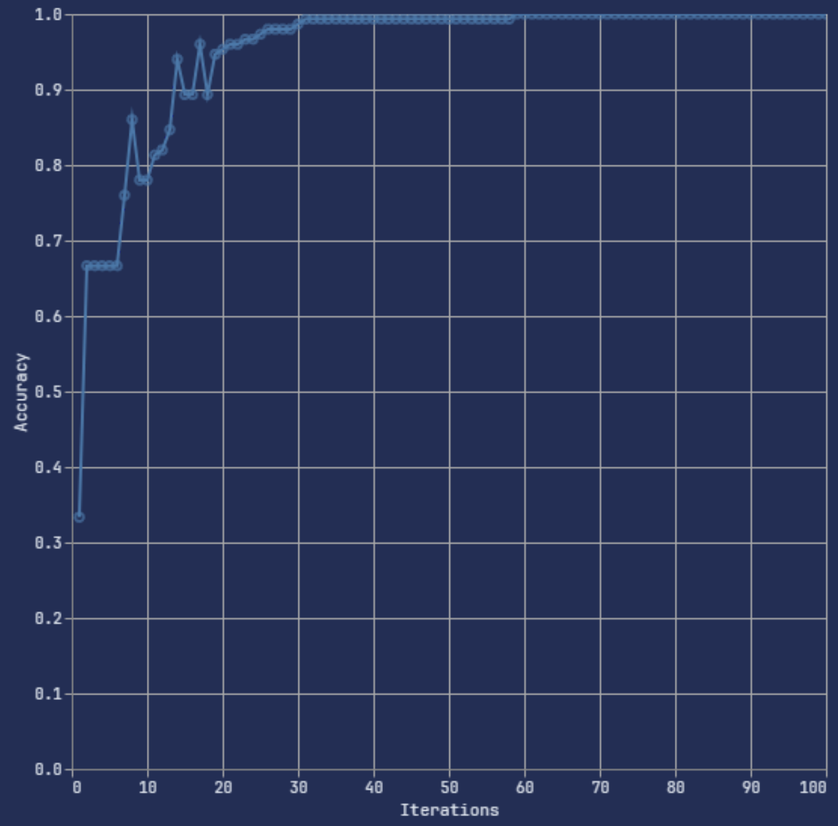


Digits Dataset

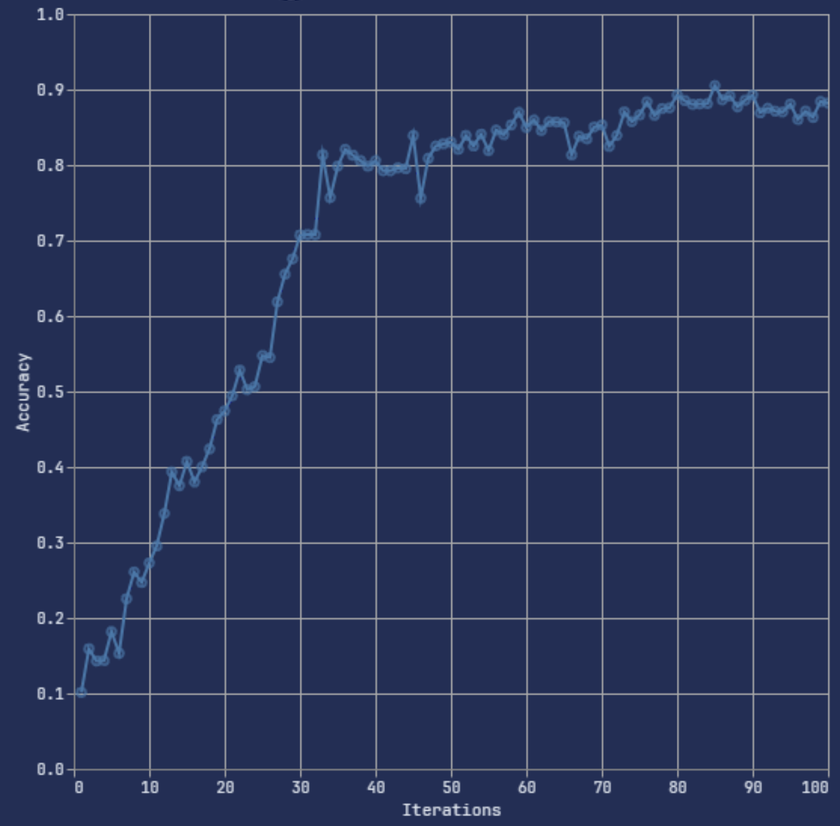


Uncertainty Sampling with Entropy

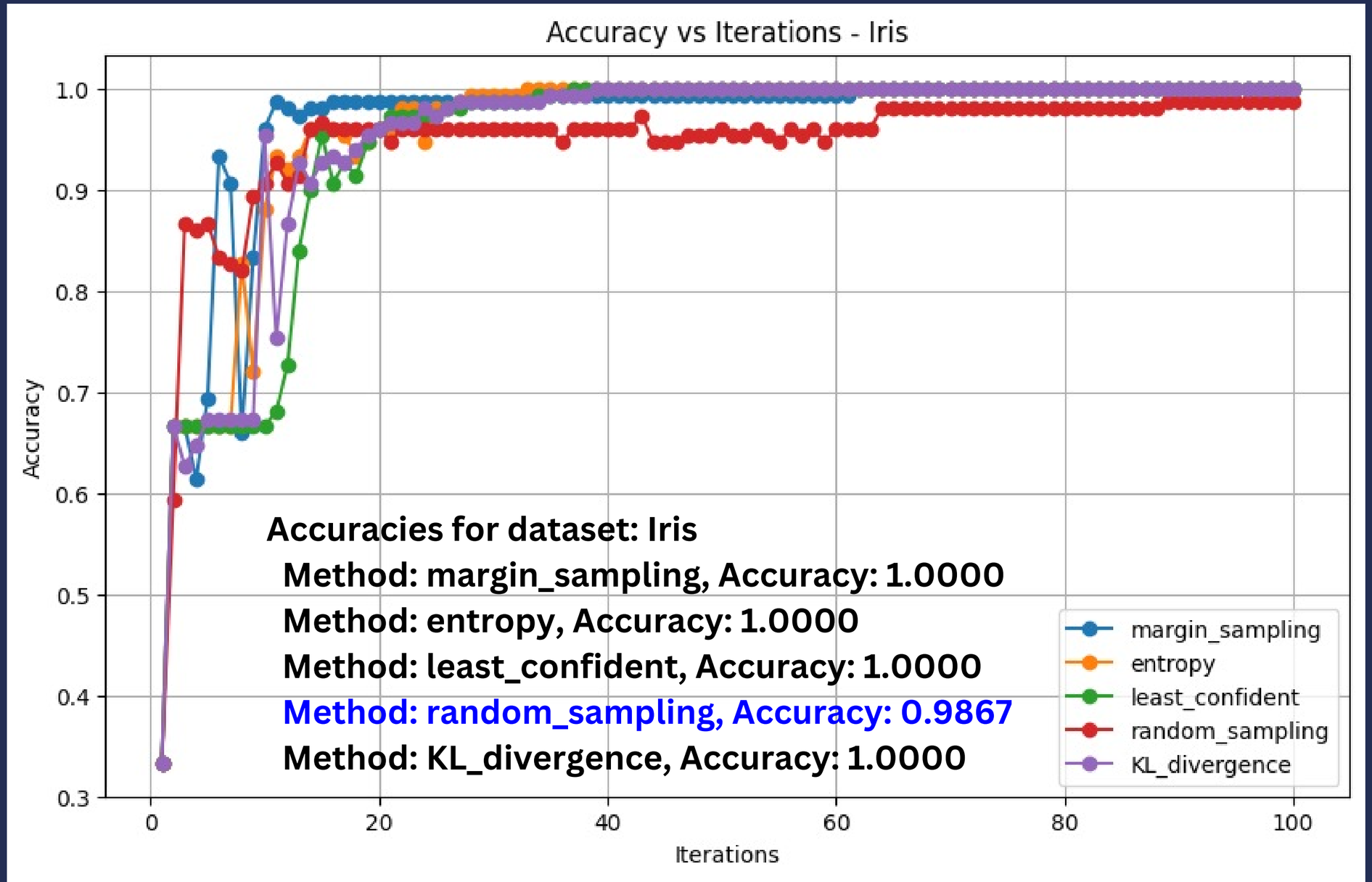
Iris Dataset



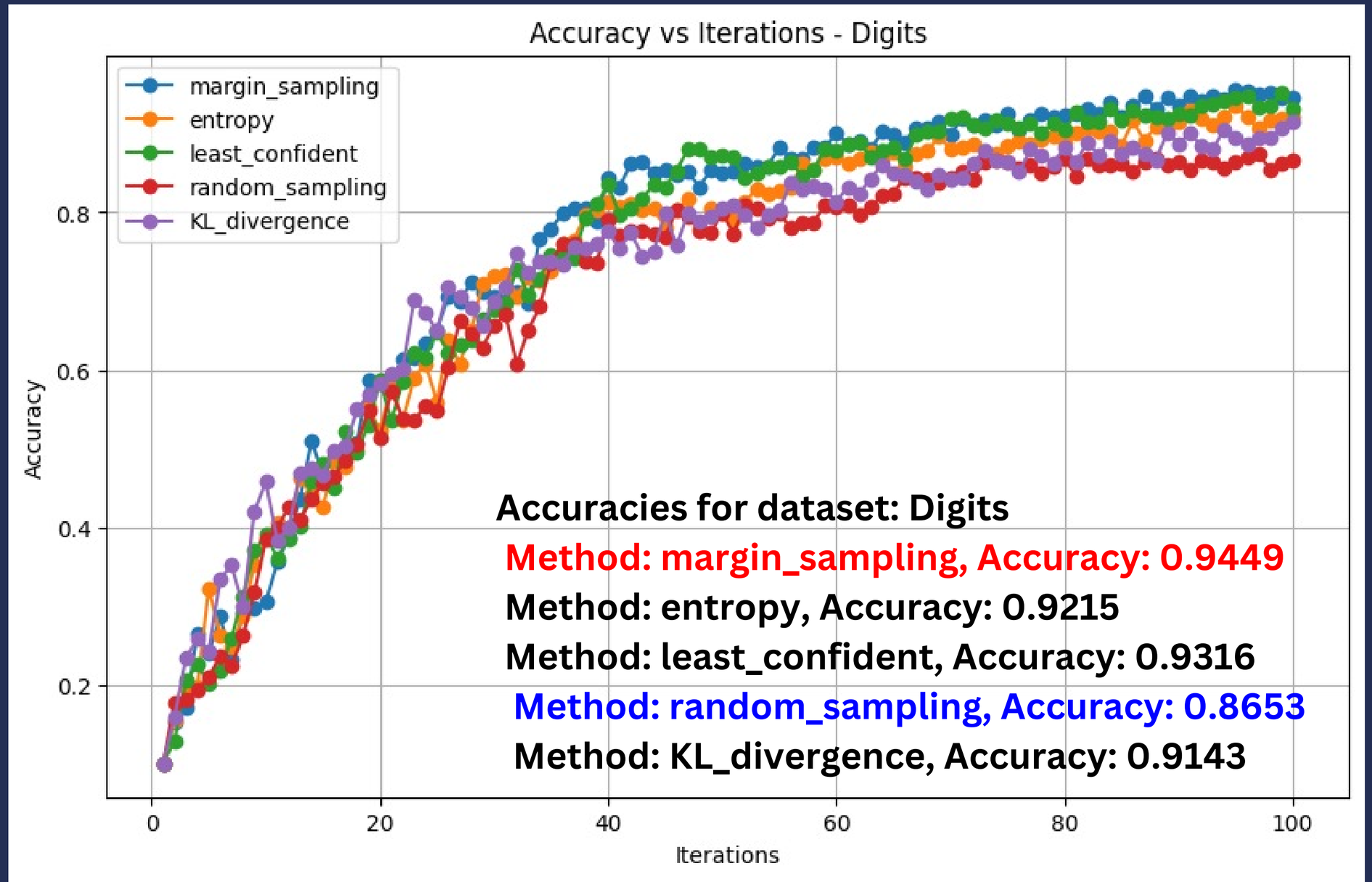
Digits Dataset



Iris Dataset

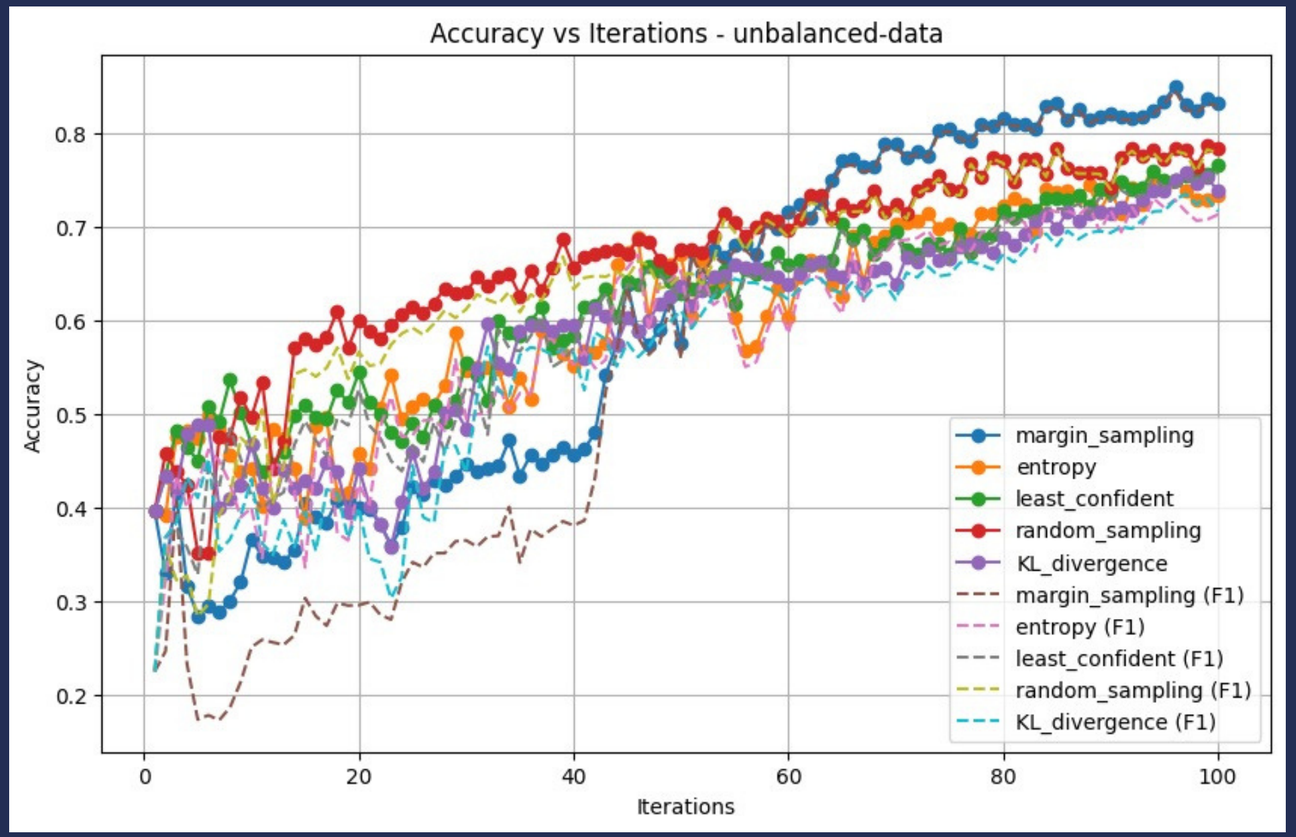


Digits Dataset



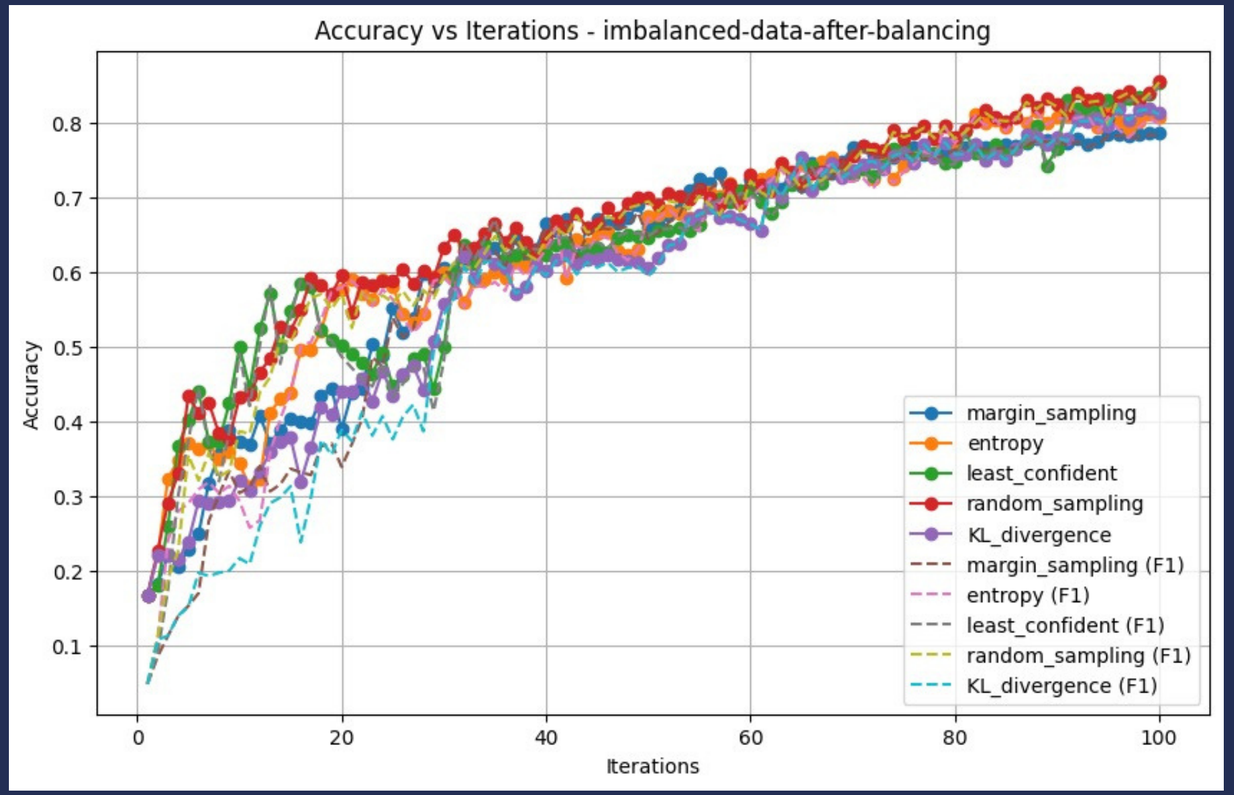
BMI unbalanced Dataset

Accuracies for a dataset: unbalanced-data
Method: margin_sampling, Accuracy: 0.8060
Method: entropy, Accuracy: 0.7840
Method: least_confident, Accuracy: 0.8040
Method: random_sampling, Accuracy: 0.7700
Method: KL_divergence, Accuracy: 0.7860



BMI unbalanced Dataset

Evaluating dataset: imbalanced-data-after-balancing
Accuracies for a dataset: imbalanced-data-after-balancing
Method: margin_sampling, Accuracy: 0.6978
Method: entropy, Accuracy: 0.8308
Method: least_confident, Accuracy: 0.8165
Method: random_sampling, Accuracy: 0.8241
Method: KL_divergence, Accuracy: 0.7769



Deployment

Select Active Learning Method

- Random Sampling
- Query By Committee
- Uncertainty Sampling with Least Confident
- Uncertainty Sampling with Margin Sampling
- Uncertainty Sampling with Entropy

Select Dataset

- Iris
- Digits

Plot

Run

Plot

Run

Thank you

Dr. Waleed

TA.Eman

- Abdelrahman Mostafa Mohamed 20200827
- Mohamed Hisham Abdelzaher 20200483
- Youssef Mohamed Abdelhay xxxxxxxx
- Abdelrahman Amin ID