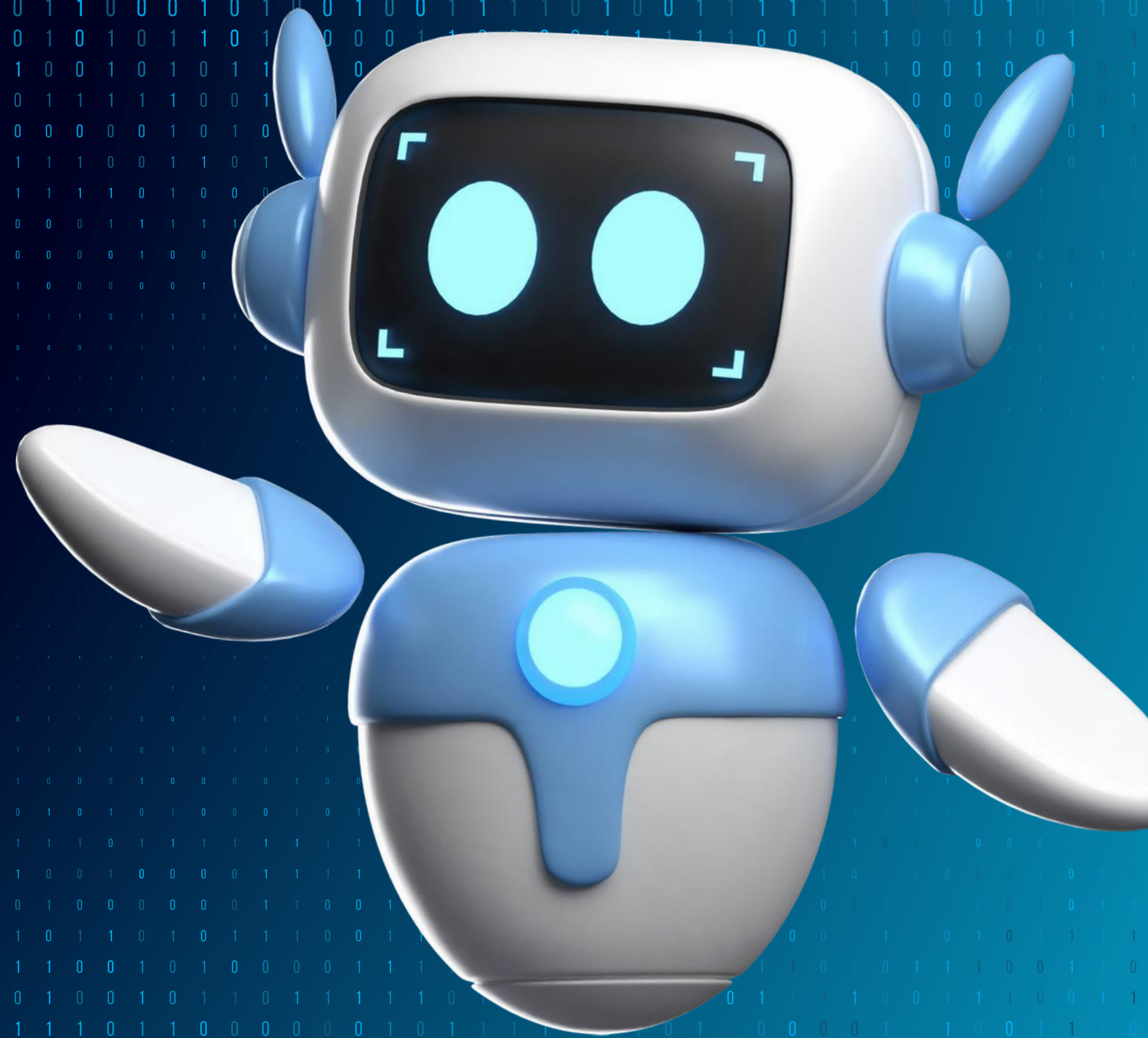


# ACTIVE LEARNING



# Introduction

# Definition

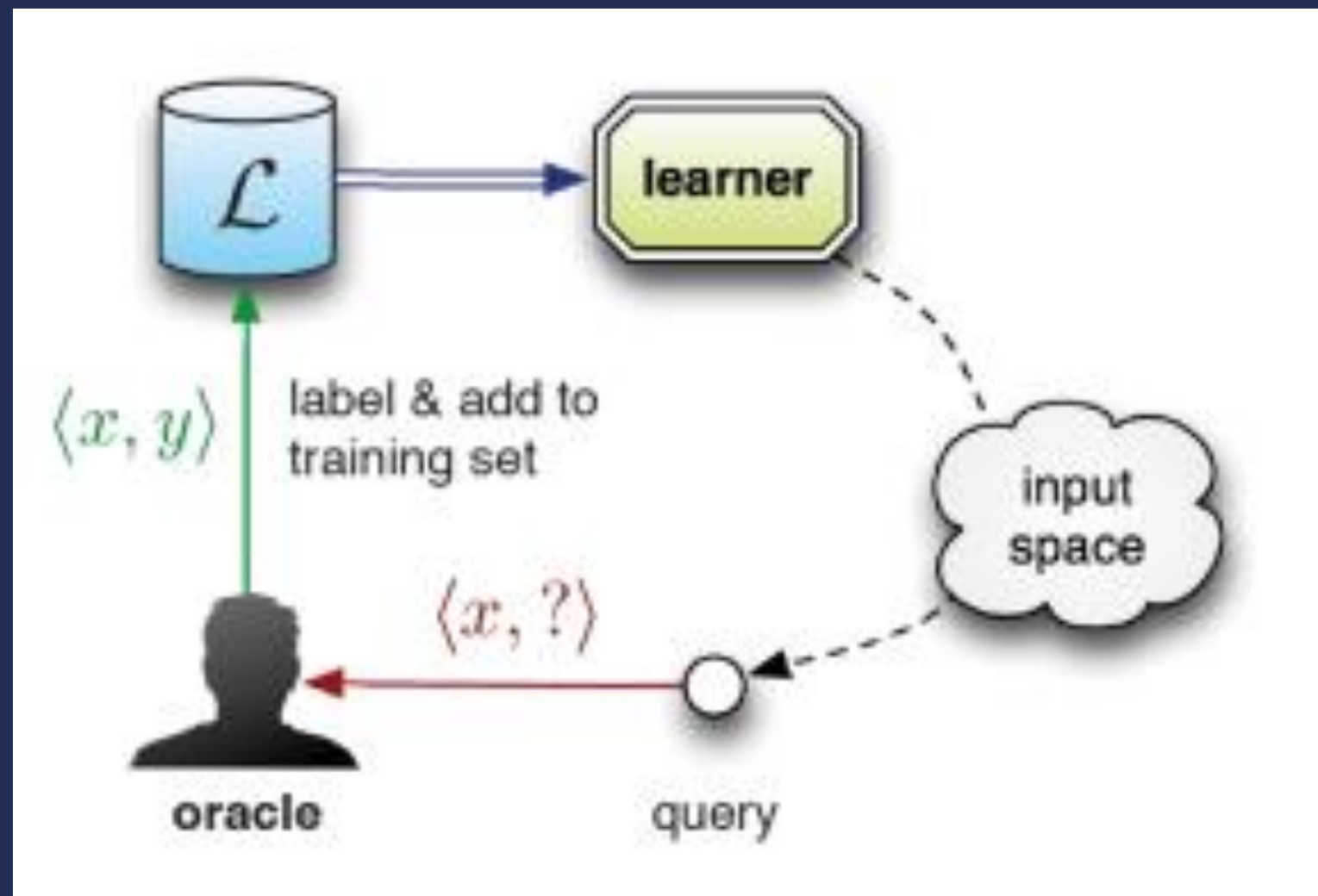
**Active Learning is a machine learning approach where the algorithm selects the most informative data points, then queries an oracle (or some other information source) to label them with the goal of reducing the total amount of data needed to learn the model.**

# 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**

- Applying 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

Results on Balanced Datasets

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

After Active Learning

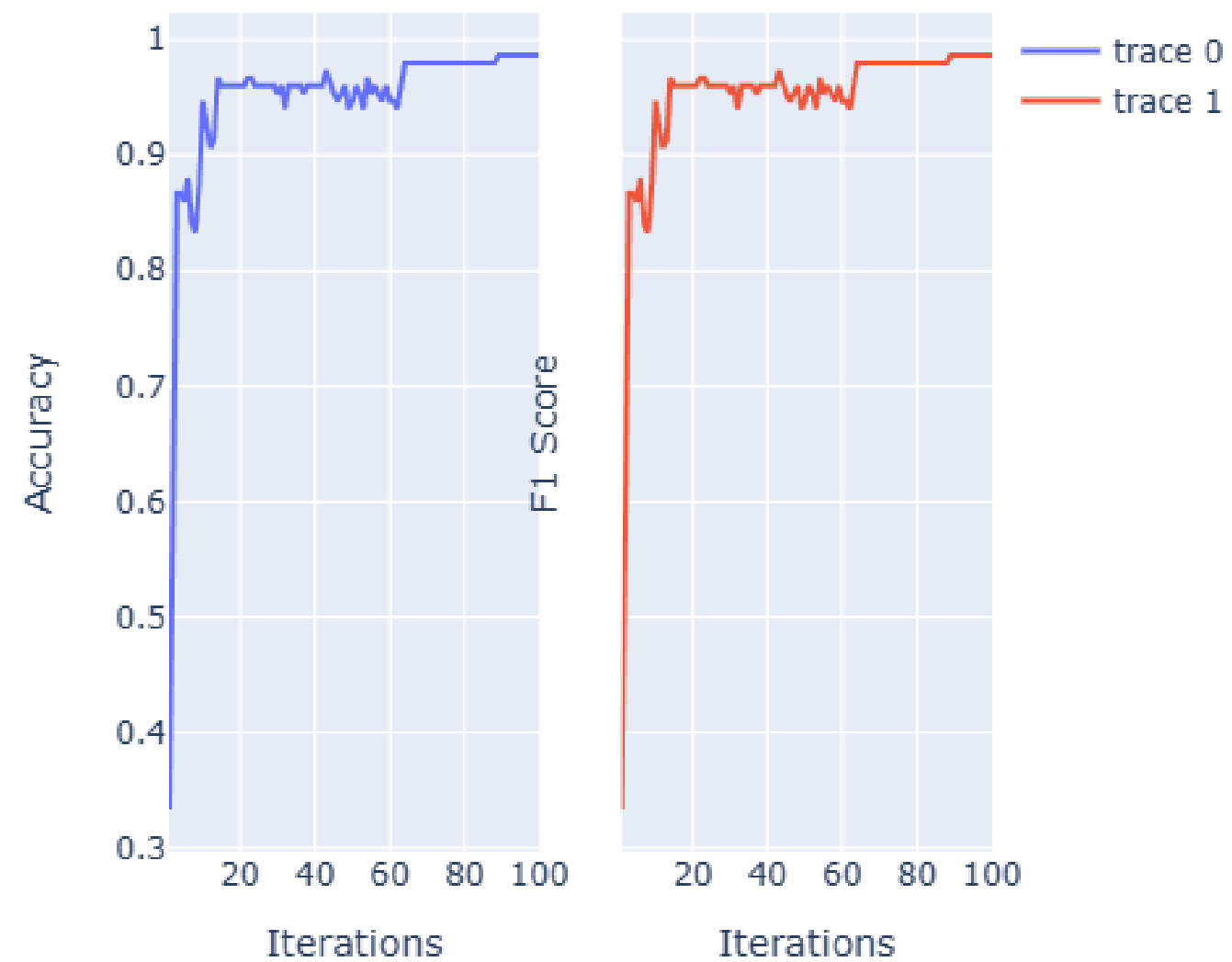
# Results on Balanced Datasets

Iris

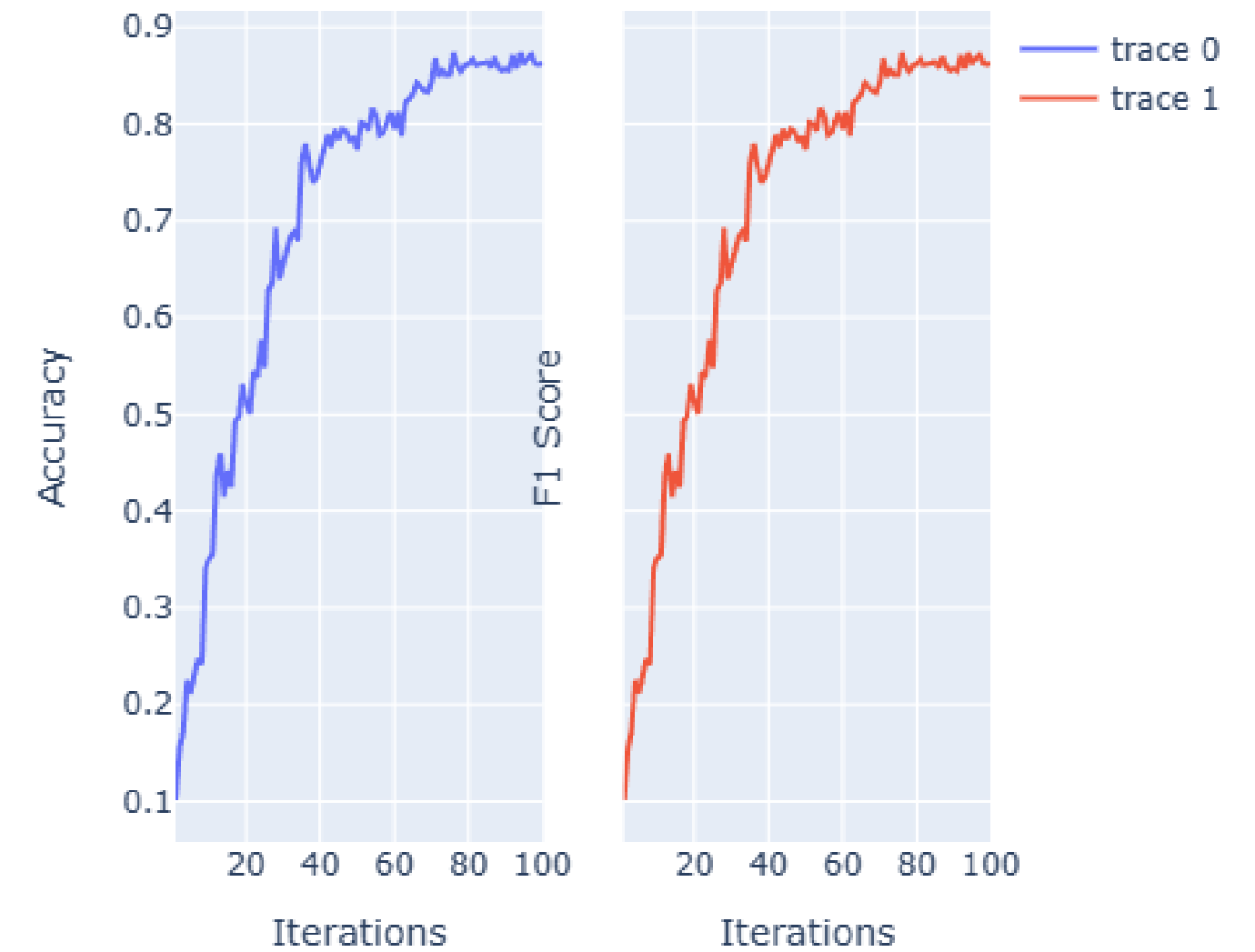
Random Sampling

Digits Dataset

Accuracy vs F1 Score



Accuracy vs F1 Score



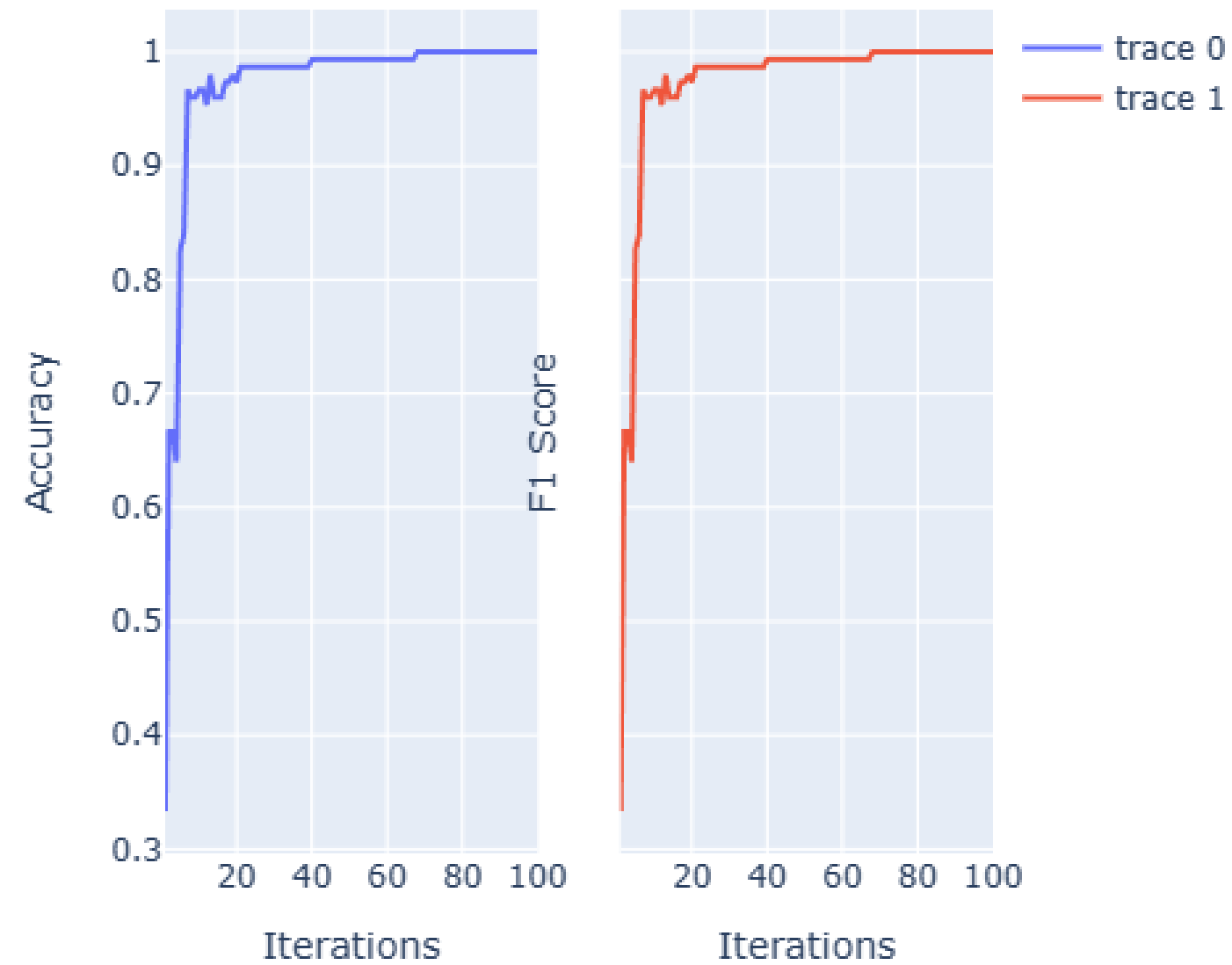
# Results on Balanced Datasets

## Query By Committee

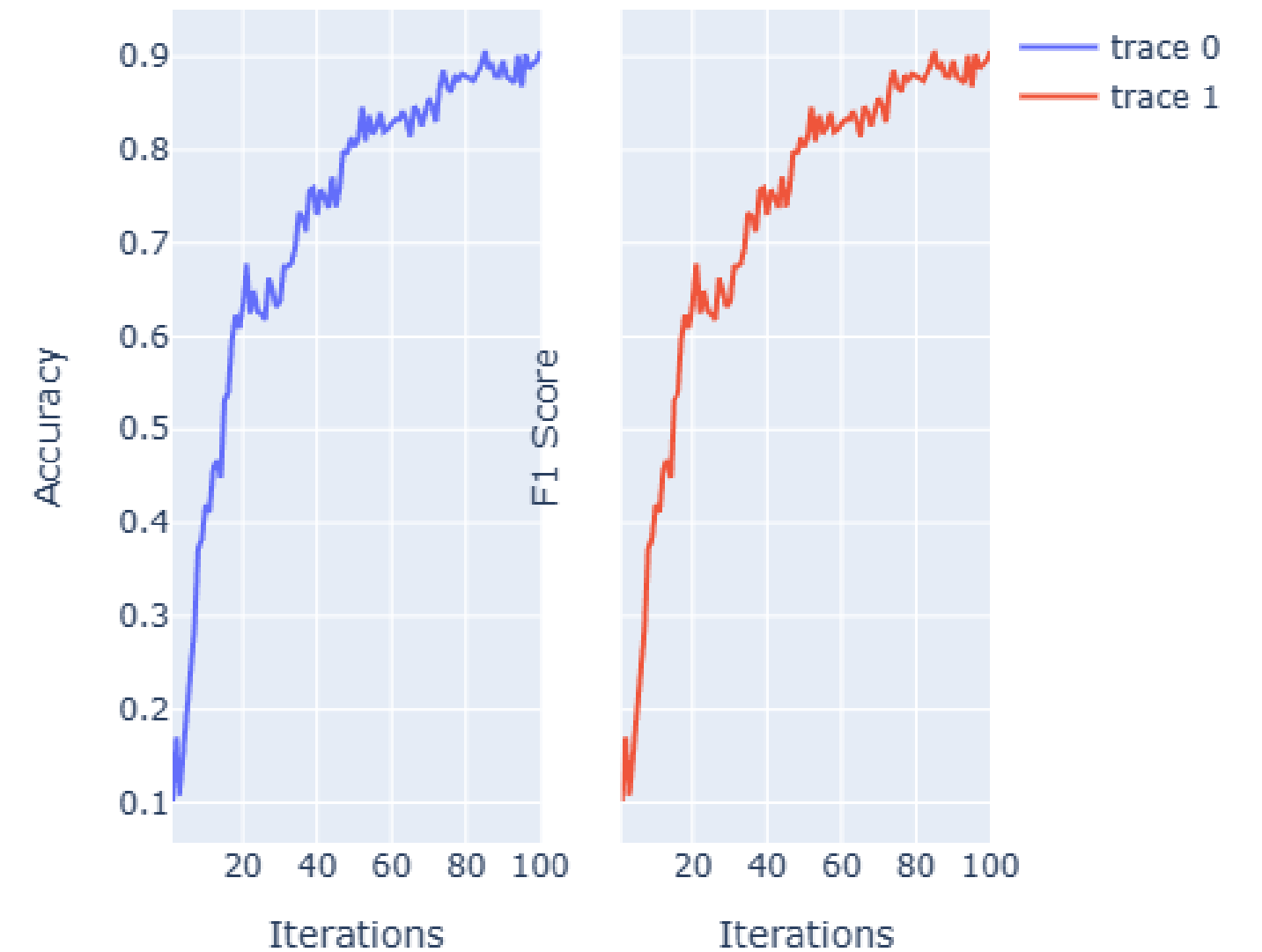
### Iris

### Digits Dataset

Accuracy vs F1 Score



Accuracy vs F1 Score



# Results on Balanced Datasets

## Uncertainty Sampling with Least Confident

### Iris



### Digits Dataset



# Results on Balanced Datasets

## Uncertainty Sampling with Margin Sampling

### Iris



### Digits Dataset





# Results on Balanced Datasets

## Uncertainty Sampling with Entropy

### Iris



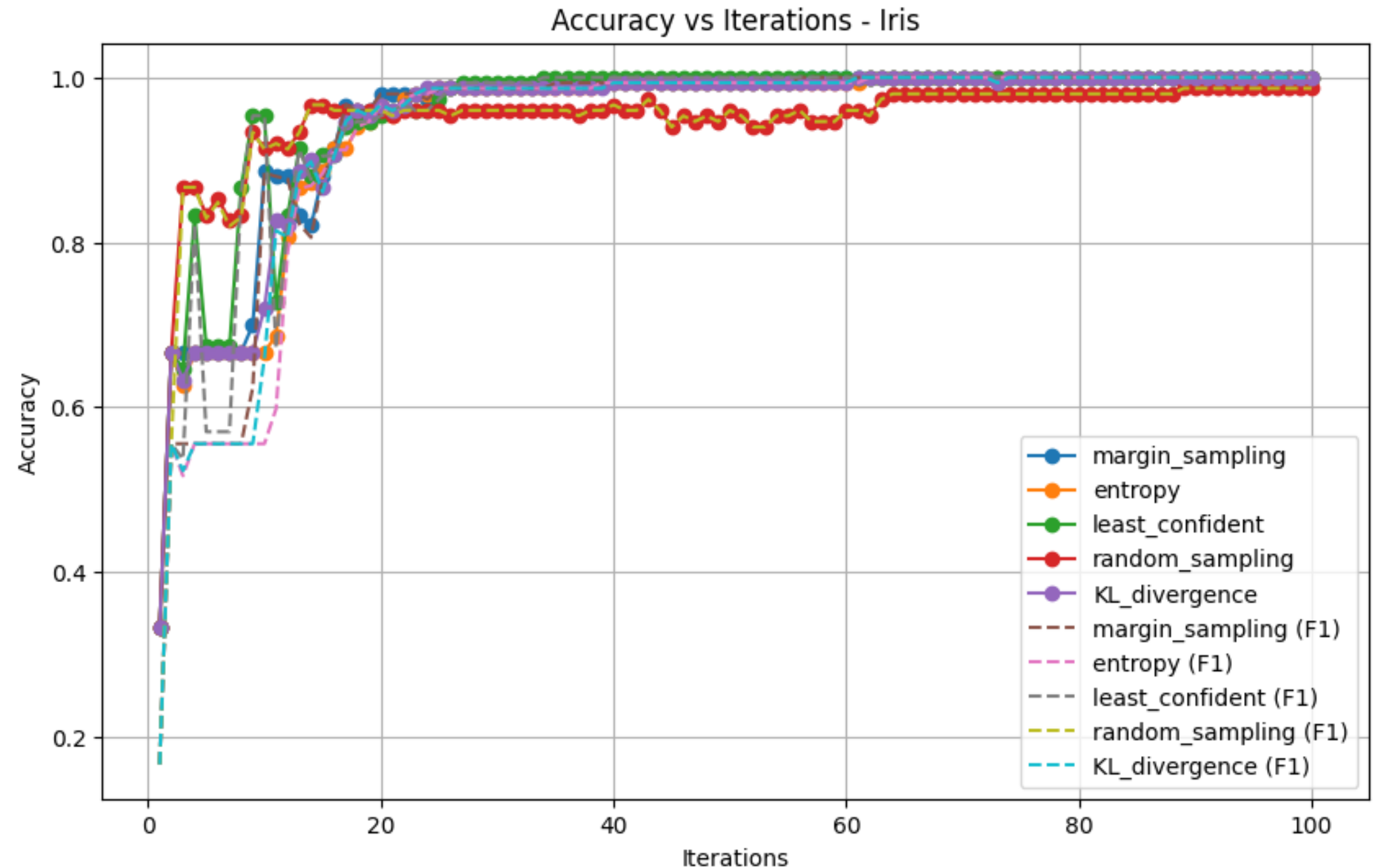
### Digits Dataset



# Iris Dataset

---

- Accuracies for dataset: Iris
- Method: margin\_sampling, Accuracy: 1.0000
- Method: entropy, Accuracy: 1.0000
- Method: least\_confident, Accuracy: 1.0000
- Method: random\_sampling, Accuracy: 0.9867
- Method: KL\_divergence, Accuracy: 1.0000
- F1-Scores for dataset: Iris
- Method: margin\_sampling, F1: 1.0000
- Method: entropy, F1: 1.0000
- Method: least\_confident, F1: 1.0000
- Method: random\_sampling, F1: 0.9867
- Method: KL\_divergence, F1: 1.0000



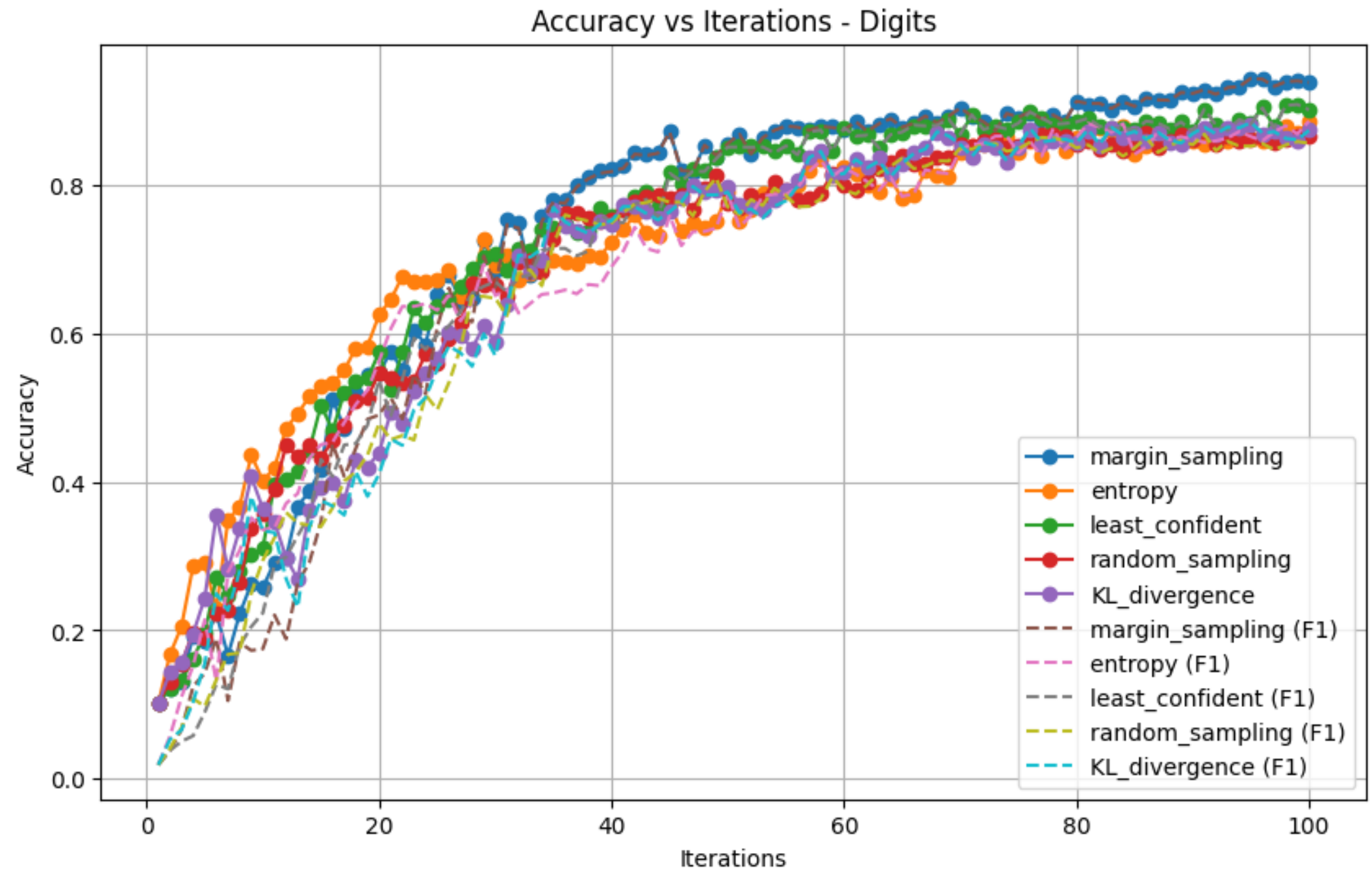
# Digits Dataset

---

- Accuracies for dataset: Digits
- Method: margin\_sampling, Accuracy: 0.9382
- Method: entropy, Accuracy: 0.8859
- Method: least\_confident, Accuracy: 0.9015
- Method: random\_sampling, Accuracy: 0.8653
- Method: KL\_divergence, Accuracy: 0.8742

## F1-Scores for dataset: Digits

- Method: margin\_sampling, F1: 0.9386
- Method: entropy, F1: 0.8866
- Method: least\_confident, F1: 0.9017
- Method: random\_sampling, F1: 0.8566
- Method: KL\_divergence, F1: 0.8738



Before active Learning

# BMI (Unbalanced Dataset)

## Accuracies

Method: margin\_sampling, Accuracy: 0.8100

Method: entropy, Accuracy: 0.7920

Method: least\_confident, Accuracy: 0.7440

Method: random\_sampling, Accuracy: 0.7780

Method: KL\_divergence, Accuracy: 0.7800

## F1

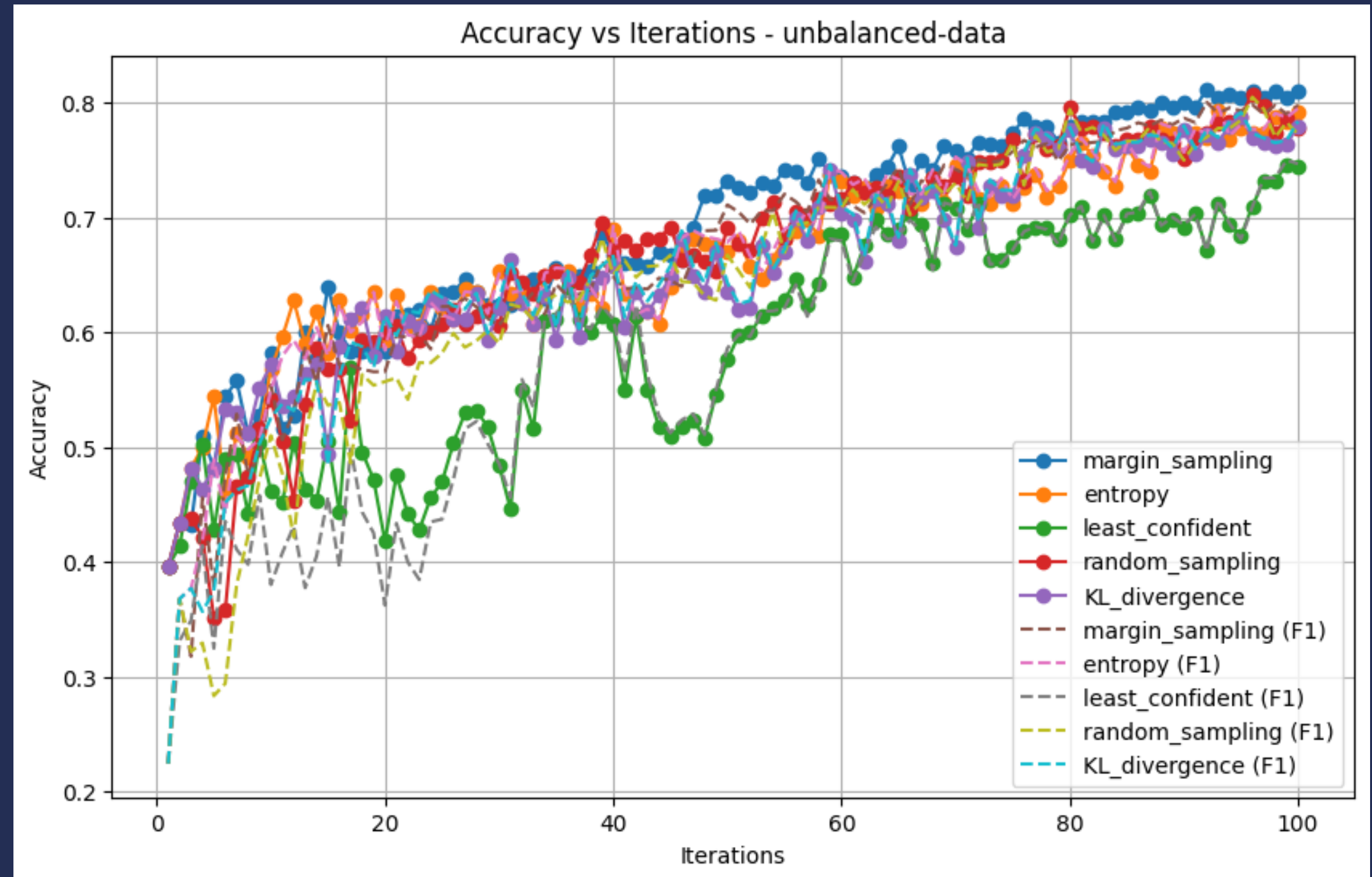
Method: margin\_sampling, F1: 0.7979

Method: entropy, F1: 0.7951

Method: least\_confident, F1: 0.7470

Method: random\_sampling, F1: 0.7728

Method: KL\_divergence, F1: 0.7839



After Active Learning



# BMI (Unbalanced Dataset)

## Accuracies

Method: margin\_sampling, Accuracy: 0.8939

Method: entropy, Accuracy: 0.8013

Method: least\_confident, Accuracy: 0.7736

Method: random\_sampling, Accuracy: 0.8232

Method: KL\_divergence, Accuracy: 0.7786

## F1

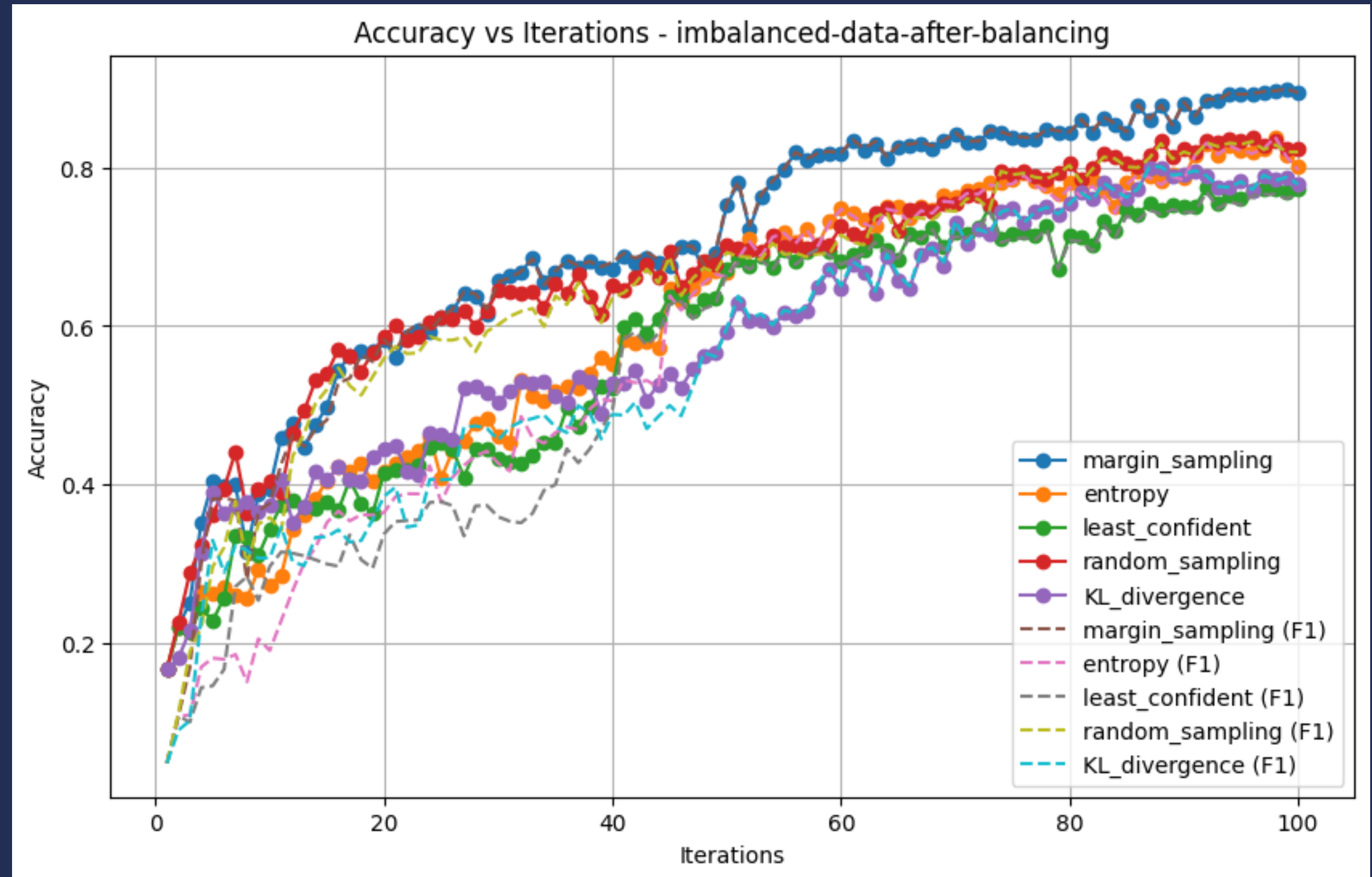
Method: margin\_sampling, F1: 0.8947

Method: entropy, F1: 0.7972

Method: least\_confident, F1: 0.7668

Method: random\_sampling, F1: 0.8197

Method: KL\_divergence, F1: 0.7789



# Deployment



Hello World

Bye World

Select Active Learning Method

☒ Random Sampling

☒ Query By Committee

☒ Uncertainty Sampling with Least Confident

☒ Uncertainty Sampling with Margin Sampling

☒ Uncertainty Sampling with Entropy

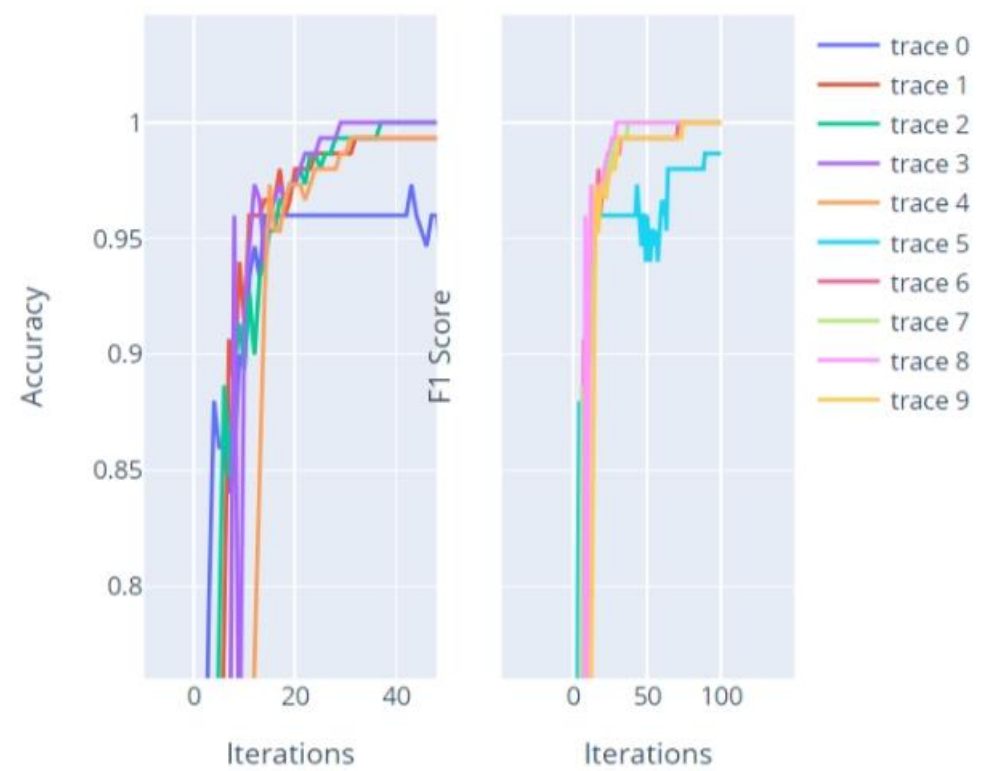
Select Dataset

☒ Iris

☐ Digits

Plot

Accuracy vs F1 Score



Run

# Thank you

- Abdelrahman Mostafa Mohamed 20200827
- Mohamed Hisham Abdelzaher 20200483
- Youssef Mohamed Abdelhay 20200669
- Abdelrahman Amin 20200311