TECHNISCHE HOCHSCHULE
OSTWESTFALEN-LIPPE
UNIVERSITY OF
APPLIED SCIENCES
AND ARTS

Evaluating the Performance of GPU Classification Based on Graphics Memory Type

A Comparative Study of Three Algorithm

10 February 2023 | Lemgo

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Introduction

- GPU Classification
- Machine learning
- Three algorithms
  - SVM
  - RFC
  - DTC

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Motivation

- Personal Experience – Product recommendations
- Ever demanding technology world
- Increasing number of products being launched
- Difficult to keep track
- Processing large amount of data

Why ML?

- Availability of multiple solutions
- Readely available libraries
- Easy to understand language
- Experties in data mining

Focus of the study

- Supervised learning
- Classification
- Multi-label Classification

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Multi-class classification

- Predict one of multiple possible classes
- More than 2 classes
- Ex: Plant species
- One-vs-all
- One-vs-one (softmax)

Classifiers

- Support vector machine
- Random forest classifier
- Decision Tree

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Support Vector Machine

- Decision boundary between two classes of objects
- Handle non-linearly separable data
- High dimensional spaces and cases where the number of features is greater than the number of samples
- Distances of new data points to the hyperplane in the feature space
- Low training error and high testing error

$$(N, K) \rightarrow \sum_{i=1}^{n} \alpha_i \quad y_i \quad K(x_i, x) + b \rightarrow f(x)$$

Where, N – Training data, K – Kernal function

Algorithm 3: Support vector machine

Input: Data set N
Output: Find differnent classes of objects

```
Function SVC (X, y, k):
    V ← φ;
    for x ∈ X do
        x_k = k(X);
        v = v − x_k;
        V ← V ∪ v;
    end
    return V;
```

Decision Tree Classifier

- Test condition on feature
- The branches
- The leaves
- Overfitting

Algorithm 2: Decision tree

Input: Examples: E, Attributes: A, Parent Examples: PE
Output: Decision Tree: T

Function `DecisionTreeClassifier(`E, A, PE`):`
  Result: Decision Tree: T
  if $|E| = 0$ **then**
    return *pluralityValue(PE)*
  end
  if $|A| = 0$ **then**
    return *pluralityValue(E)*
  end
  if $\forall e \in E, e$ *classifies the same* **then**
    return *the classification*
  end
  $A' = \text{argmax}_{a \in A}(importance(a, E))$;
  $T = newTree(root = A')$;
  for $v \in A'$ **do**
    $exs = \{e \in E | e.A' = v\}$;
    $subtree = decisionTreeLearning(exs, A - A', E)$;
    $T.addSubtreeAsBranch(subtree, label = (A', v))$;
  end
  return T

Random Vector Classifier

- In 2001, Leo Breiman of the University of California proposed the Random Forest
- Collection of tree-structured classifiers
- Identically distributed independent random vectors
- Each tree casting a unit vote for the most popular class at input

$$(N, n, e) \rightarrow \frac{1}{n}\sum_{i=1}^{n} f_n(x, e) \rightarrow f(x)$$

Where, N – Training data, e - Entropy,
n – Number of decision tree

Algorithm 1: Random forest

Input: Data set N
Output: Find differnent classes of objects

Function `RandomForestClassifier(`X, y, n, e`):`
  $DF \leftarrow \phi$;
  for $x \in X$ **do**
    $x_i \leftarrow BTS(X)$;
    $f \subset F$;
    $d_i \leftarrow f_n(x, f, e)$;
    $DF \leftarrow DF \cup d_i$;
  end
  return DF;

Performance Matrics

- Confusion Matrix
  - Accuracy
  - Precision
  - Recall
  - F1
  - Execution Time
  - racy
- Cross-validation – StratifiedKfold
  - Accuracy
  - Execution Time

Why StratifiedKfold?

- Class imbalance
- Model evaluation
- Better performance

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

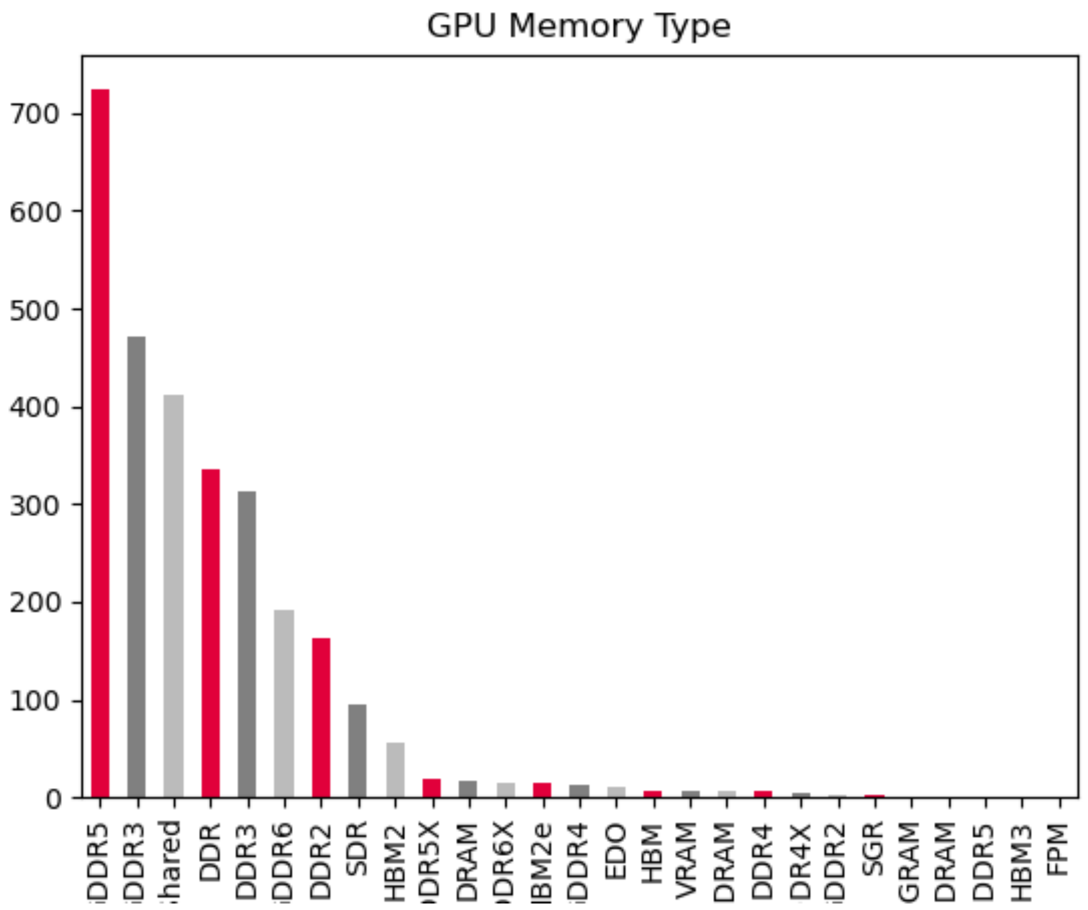
Data Preperation

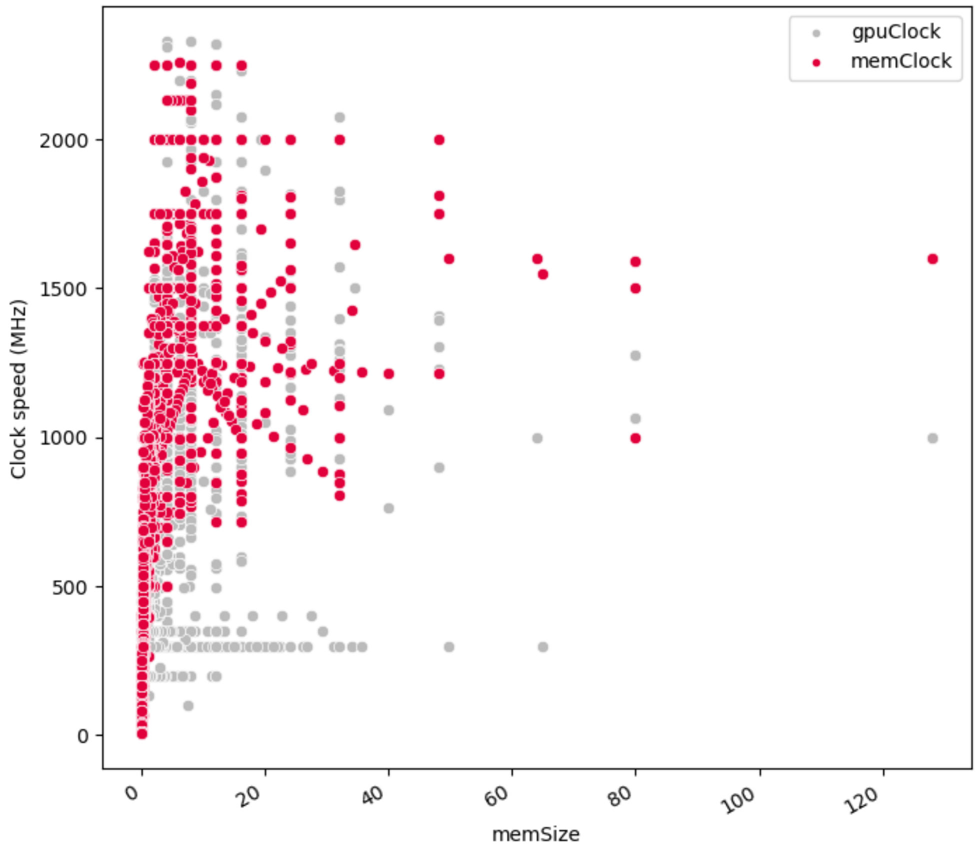
- Kaggle dataset – Data Mining UNDIP
- 2889 samples and 16 features
- 5 features considored
- DV – Memory Type
- Interpolation
- Dataset split: 80% - Training, 20% - Test
- Pearson's corelation test

	memType	memSize	gpuClock	memClock	memBusWidth
0	GDDR6	8.0	1925	2250.0	128.0
1	GDDR6	4.0	300	1500.0	64.0
2	GDDR6	4.0	300	1500.0	64.0
3	GDDR6	4.0	300	1500.0	64.0
4	GDDR6	8.0	300	1500.0	128.0

Dataset snapshot

Dataset visualisation





Algorithm Training

Models were trained using selected features:

- Memory Size
- GPU Clock
- Memory Clock
- Memory Bus Width

Evaluation

Confusion matrix

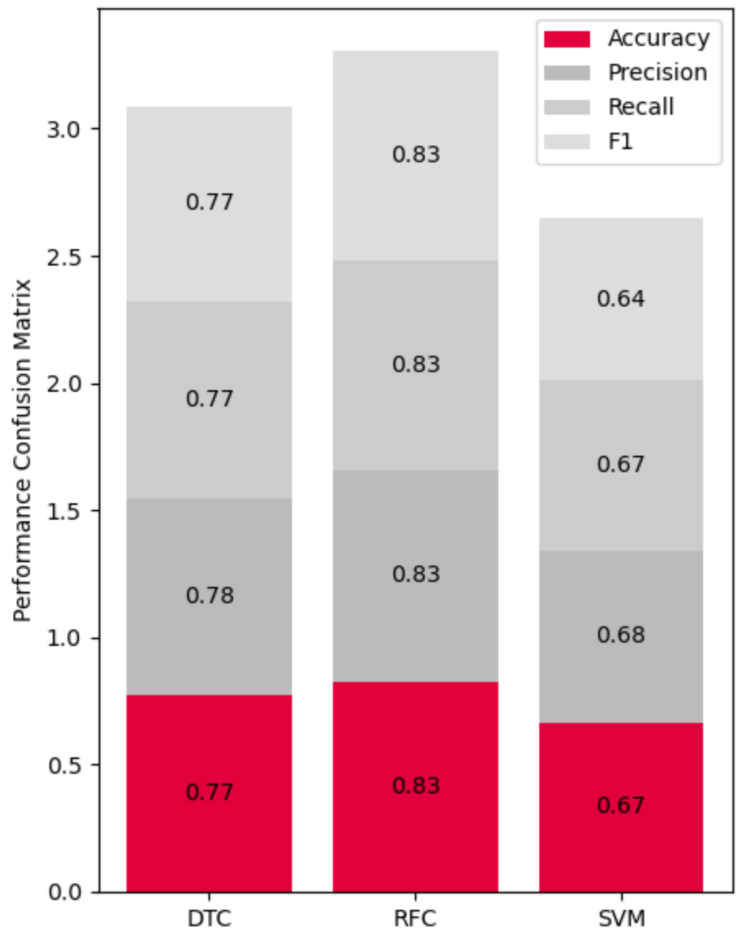
- Accuracy
- Precision
- Recall
- F1

Cross-validation

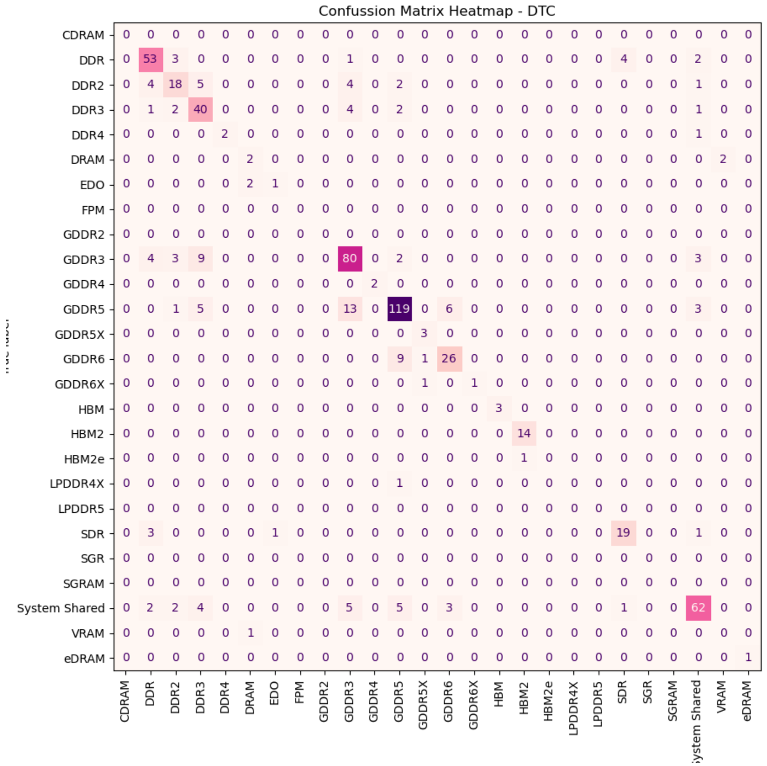
- Over 1200 tests – to avoid OS scheduling effect
- Accuracy
- Execution Time (Train and Predict)

Confusion matrix results

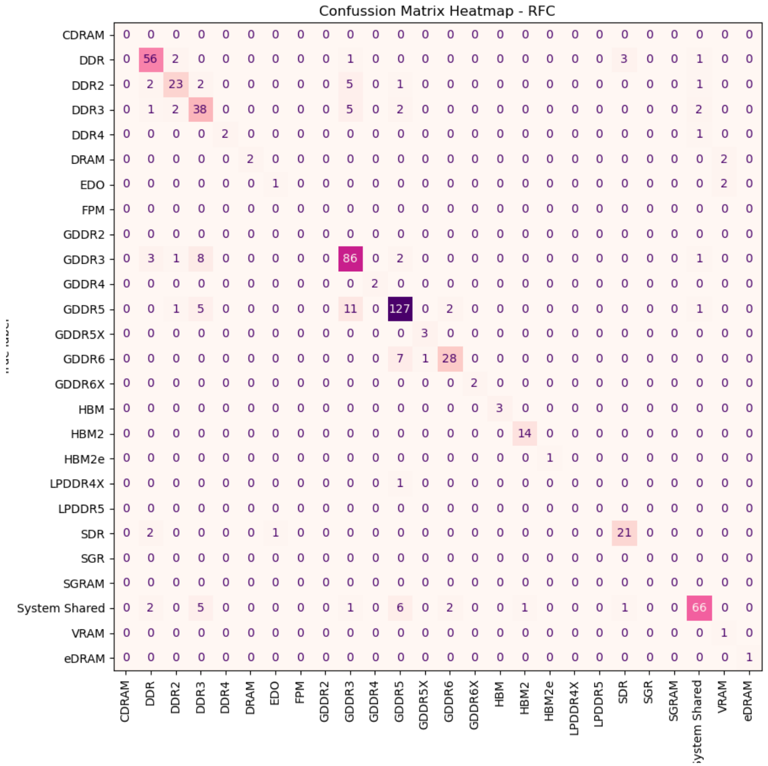
- Overall, RFC performed best
- DTC performed well
- SVM performed poorly



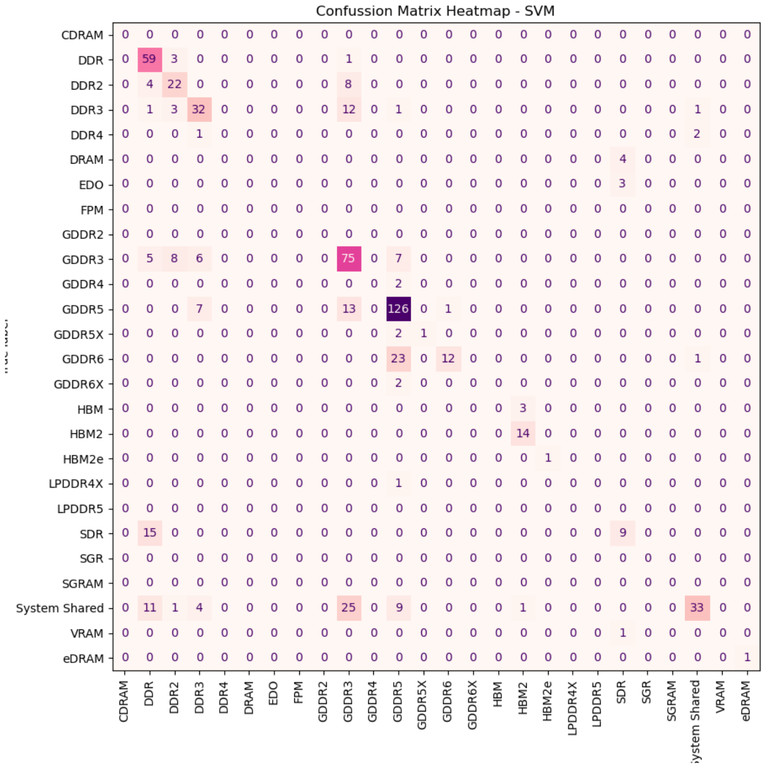
Confusion Matrix Heatmap



DTC



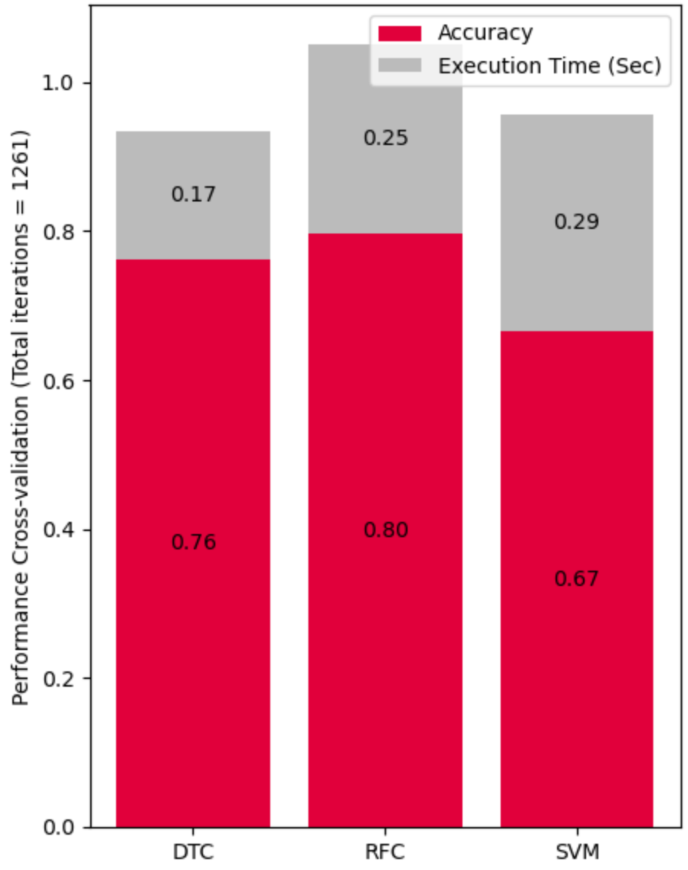
RFC



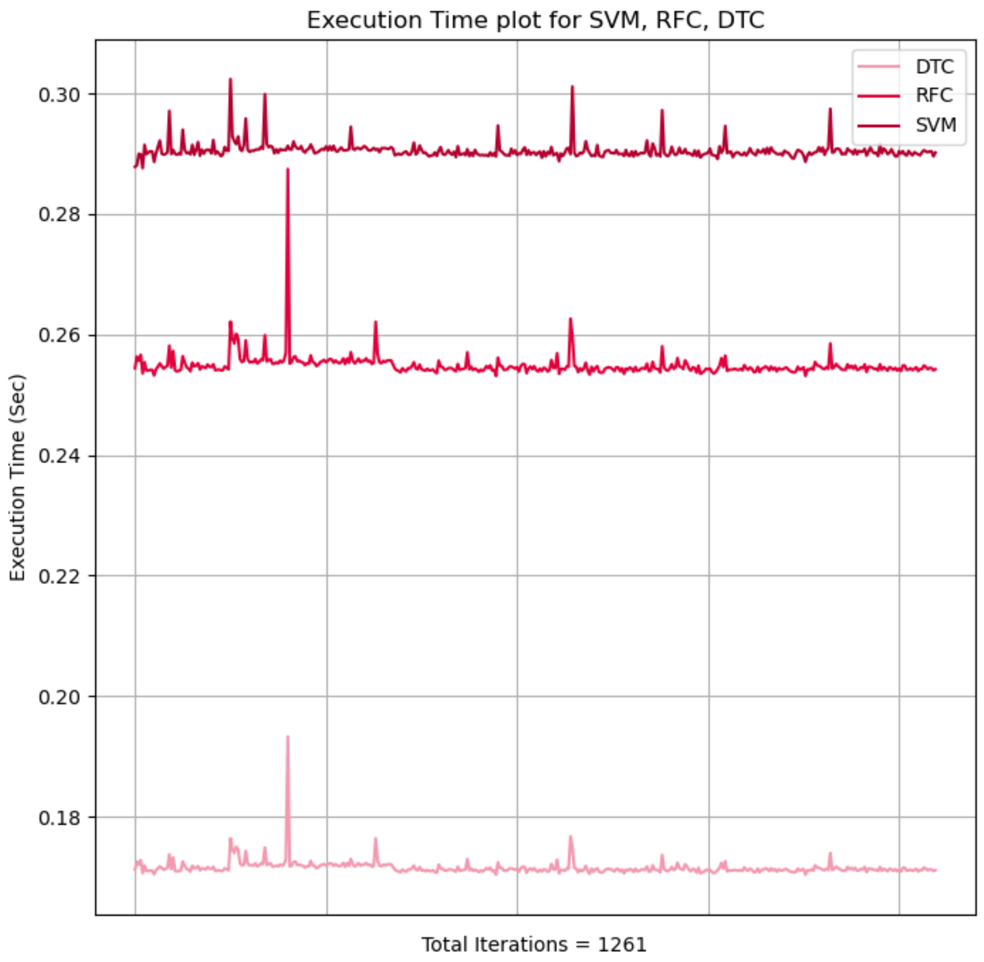
SVM

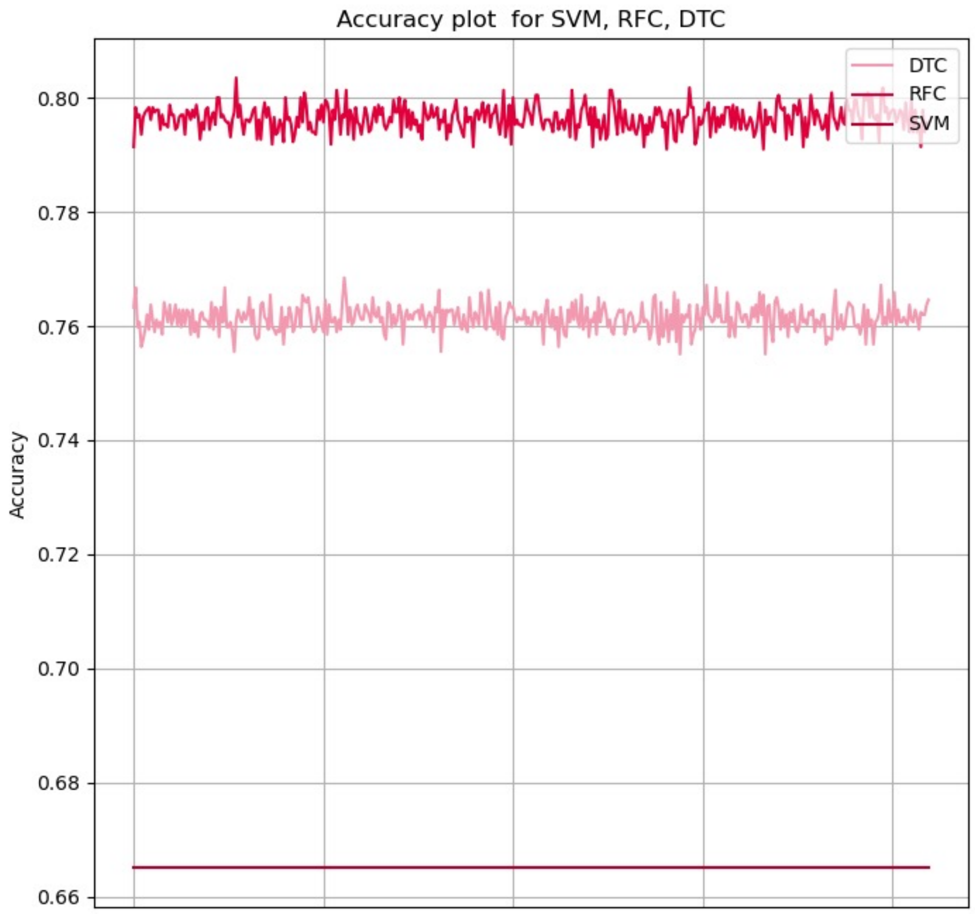
Cross-validation results

- Results are average of over 1200 tests
- In execution time, DTC performed the best
- RFC performed well
- SVM took highest time



Performance consistancy





Combined results

RFC

- Accurate
- Fast
- Precise

DTC

- Less accurate
- Fastest
- Less precise

SVM

- Least accurate
- Slowest
- Least precise

Classifier	Accuracy (CV)	Accuracy (CM)	Precision	Recall	F1	Execution Time
DTC	0.77	0.76	0.78	0.77	0.77	0.17
RFC	0.83	0.80	0.83	0.83	0.83	0.25
SVM	0.67	0.67	0.68	0.67	0.64	0.29

Overview

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Summary and Outlook

- RFC has highest acuracy, followed by DTC and SVM
- Evaluation based on Accuracy, Precisin, Recall, F1 score and Execution time.
- No parameter optimization
- Utilizes ML to effectively classify GPU based on memory type
- Solution can be extended to other technology products

https://github.com/AkshayChikhalkar/CPU-Classification-RFC-DTC-SVM

Summary

- Introduction
- Motivation
- State of the Art
- Solutions
- Experiments and Results
- Summary and Outlook

Questions

TECHNISCHE HOCHSCHULE
OSTWESTFALEN-LIPPE
UNIVERSITY OF
APPLIED SCIENCES
AND ARTS

Thank you!