

WG-4

Analysis techniques

Machine learning in Alice and Atlas

Alexandre Suaide

write a program using ROOT libraries from CERN to read a data table and plot an histogram

Certainly! Here's an example program that uses ROOT libraries from CERN to read a data table from a text file and plot a histogram. This program assumes that you have ROOT installed and set up properly.

```
cpp
#include <iostream>
#include <fstream>
#include <vector>
#include <string>
#include <tree.h>
#include <TH1F.h>
#include <TCanvas.h>

using namespace std;

int main() {
    // Open the input data file
    ifstream inputFile("data.txt");
    if (!inputFile.is_open()) {
        cout << "Error opening file!" << endl;
        return 1;
    }

    // Create a ROOT file to store the histogram
    TFile outputFile("output.root", "RECREATE");

    // Create a ROOT histogram
    TH1F histogram = new TH1F("histogram", "Data Histogram", 100, 0, 1000);

    // Read data from the input file and fill the histogram
    double value;
    while (inputFile >> value) {
        histogram->Fill(value);
    }

    // Close the input file
    inputFile.close();

    // Create a canvas to display the histogram
    TCanvas canvas = new TCanvas("canvas", "Histogram Canvas");

    // Draw the histogram on the canvas
    histogram->Draw();

    // Save the canvas with the histogram to the output file
    outputFile->cd();
    canvas->Write();

    // Close the output file
    outputFile.close();

    // Cleanup
    delete histogram;
    delete canvas;

    return 0;
}
```

To use this program, save it in a file called "program.cpp" and create a text file called "data.txt" containing the data you want to plot, with each value on a separate line.

To compile and run the program, you'll need to have ROOT installed and set up properly. You can compile the program with the following command:

```
bash
g++ `root-config --cflags` `root-config --libs` program.cpp -o program
```

Then, run the compiled program:

```
bash
./program
```

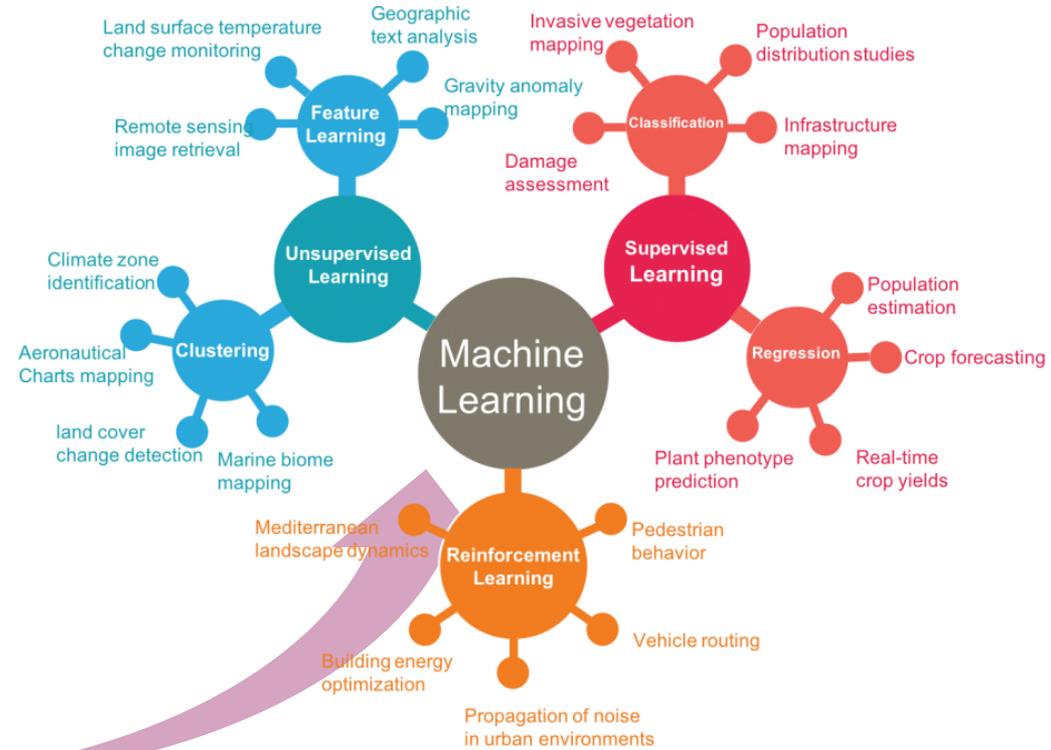
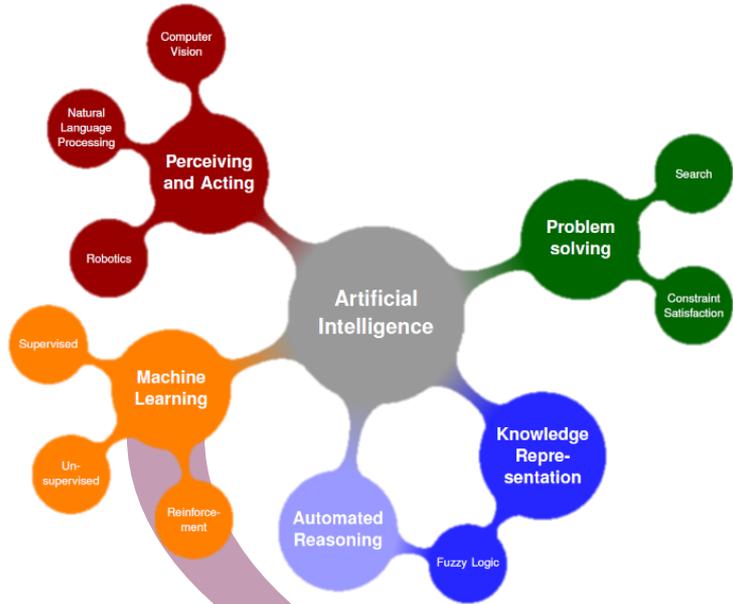
Regenerate response

Objetivos do WG-4

- Desenvolver técnicas de análise e simulações envolvendo ML
 - Incorporar essas técnicas ao dia-a-dia das análises no Alice e Atlas
 - Grande interface com demais grupos
 - Grande envolvimento dos estudantes do projeto
- Infraestrutura computacional do projeto

Atividades realizadas nos últimos meses

- Mapeamento das análises que utilizam técnicas de inteligência artificial
- Avaliação e planejamento da infraestrutura necessária para desenvolvimento dessas técnicas
- Necessidade de formação na área
- Atualizações e manutenção do cluster sampa durante o primeiro ano de projeto



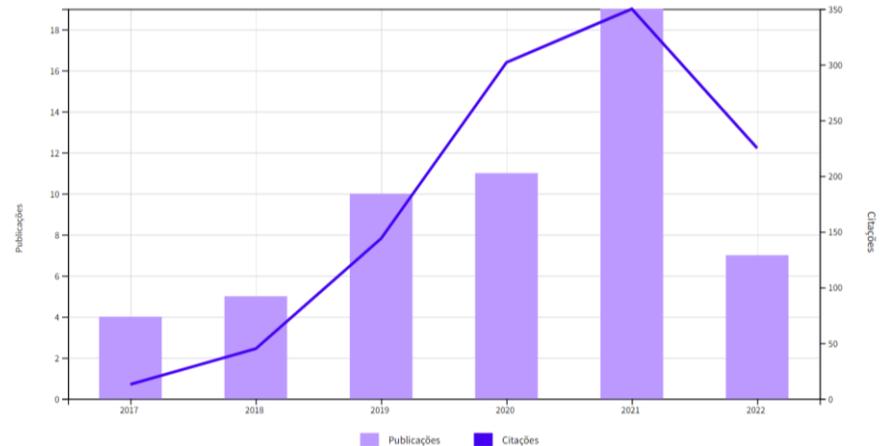
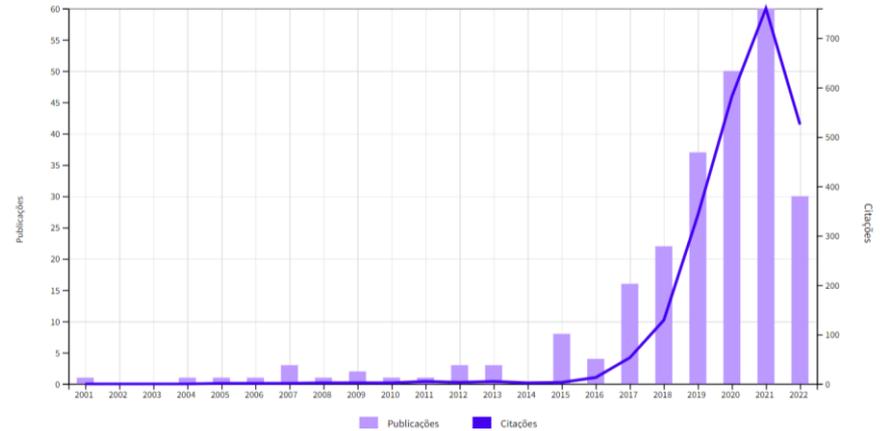


Aprendizado de máquina em HEP

Número de artigos e citações envolvendo “Machine learning” e “LHC” tem aumentado consideravelmente nos últimos anos

Uma fração considerável desses artigos está relacionada a jatos

- ~ 1/3 das publicações e ~ 1/2 das citações em 2021

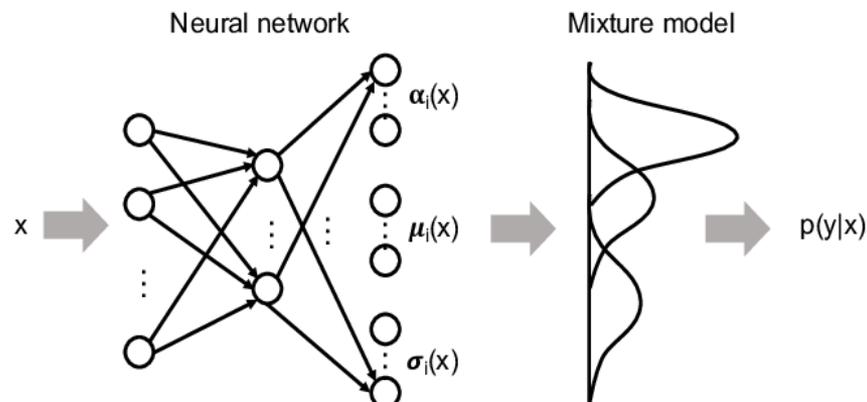




High mass measurement of tau lepton pairs

Caio Daumann

- Estimar grandezas cinemáticas de eventos com dois taus no estado final
 - Neutrinos não medidos no estado final
- Uso de Mixture Density Networks
 - Rede neural artificial cujo objetivo é aprender como fornecer na saída distribuições de probabilidade



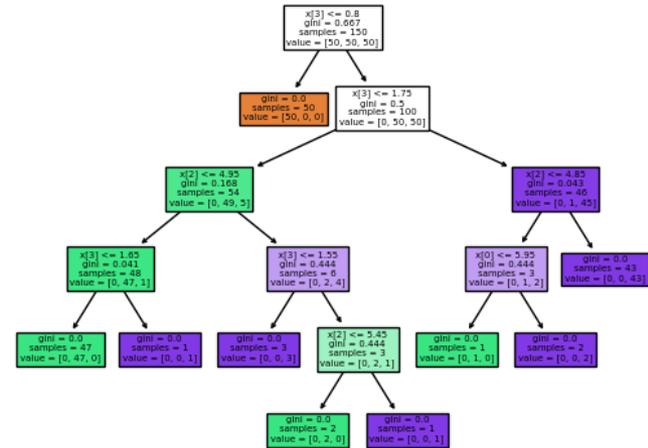


BDT para identificação de hádrons exóticos – X(3872)

Leopoldo

- Decisions trees são métodos recursivos de tomada de decisões baseadas em cortes nos parâmetros de entrada
- BDT (Boosted Decision Tree) é um método de combinar várias árvores mais fracas em um classificador mais forte

Decision tree trained on all the iris features





Uso de redes neurais para identificação de hipernúcleos no ALICE

Maria Paula

- Tradicionalmente reconstrução de hipernúcleos são feitas com cortes unidimensionais em parâmetros na geometria do decaimento
- Redes neurais podem ser usadas para otimizar a seleção de candidatos a hipernúcleos

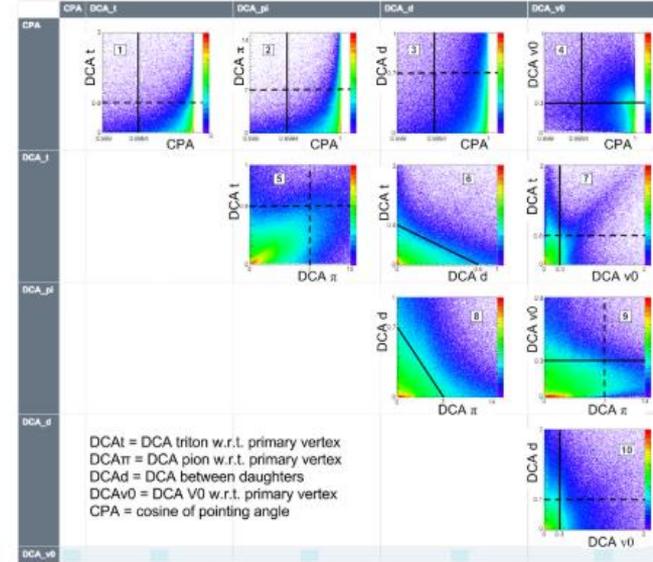


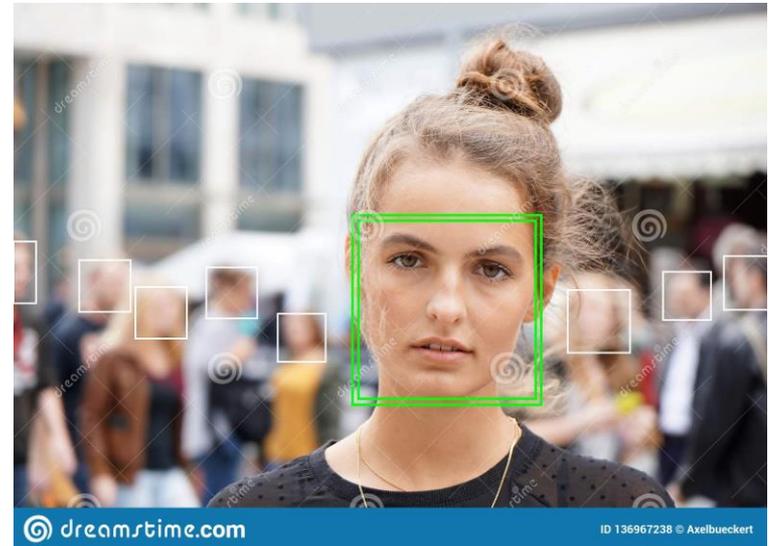
Figura 3.8: Two-dimensional distributions of the geometric variables of $\Lambda b \rightarrow {}^3\text{H} + \pi$ decays injected in the Monte Carlo. The cuts adopted on data, represented by solid and dashed lines in the figure, were extracted after inspection of the Monte Carlo distributions.



Identificação de jatos usando técnicas de classificação de imagens com deep learning

Jhoão

- Algoritmos para identificação e reconstrução de jatos são usualmente baseados em métodos estatísticos
 - kT, anti-kT, cone, etc.
- A ideia desse trabalho é identificar jatos evento por evento usando técnicas de deep learning no reconhecimento e classificação de imagens





Avaliação geral

- Há várias análises promissoras usando técnicas sofisticadas de inteligência artificial
- Contudo
 - Infraestrutura computacional pouco adequada
 - Formação mais específica na área de inteligência artificial
 - Hoje as ferramentas são caixas pretas para a maioria dos alunos



Ações para o próximo ano

- Montagem de um workbench computacional para inteligência artificial
- Duas opções
 - Compra de um servidor com algumas GPU's para processamento de dados usando IA
 - Servidor + 2-4 GPUS de fornecedor tradicional (HP, SGI, etc)
 - Estimado em R\$ 200-300 k
 - Atualização de uma máquina do cluster (jarvis) e compra de GPUS de prateleira
 - Estimado em R\$ 30-40 k



Ações para o próximo ano

- Formação de pessoal
 - Proposta de uma disciplina de pós-graduação chamada “Aprendizado de máquina e inteligência artificial em física”
 - Oferecimento no primeiro semestre de 2023
 - Alexandre Suaide e Tiago Fiorini



Infraestrutura Cluster SAMPA

- Configuração atual
 - 2408 cores
 - 18.7 kHS06
 - 0.85 TB de armazenamento
 - 2 GPU Tesla K20
- Uso
 - Processamento em GRID do ALICE e ATLAS
 - Análise de dados e simulações
 - Processamento em geral
 - Desenvolvimento





Cronograma de upgrades

	Requested Processing Power (kHS06)	Extra Compute Modules for the SAMPA cluster	Processing Power per Compute Module (kHS06)
2022	20,7	4	0.62
2023	23,8	6	0.62
2024	27,4	5	0.75
2025	31,5	6	0.75
2026	36,3	6	0.90
TOTAL		27	

Table 2: Expect grown of the processing power of the computer cluster assuming 15% per year increase in the demand and 20% evolution every 2 years of the processing power of the compute modules

	Requested Disk Storage (PB)	Number of Extra Disks for the SAMPA Cluster	Single Disk Capacity (TB)
2022	2,12	91	14
2023	2,44	23	14
2024	2,80	23	16
2025	3,22	27	16
2026	3,71	27	18
TOTAL		191	

Table 3: Expect grown of the storage power of the computer cluster assuming 15% per year increase in the demand and 20% evolution every 2 years of single-disk capacity



Atividades sendo realizadas

- Instalação do storage recentemente adquirido
- Atualização do software do cluster para ALMA 9, como recomendado pelo CERN
- Atualização de link para 40 Gbps usando redNESP (Research and Education Network at São Paulo) – antiga ANSP
 - Em instalação no “CCE”
 - Compra de transdutores
 - ~ R\$ 5-10 k

