FIRST MEETING FOR START STUDENTS

ALEJANDRO SAN JUAN & JUAN CARLOS MARQUEZ

TUESDAY, JUNE 18TH, 2024

CONTENT

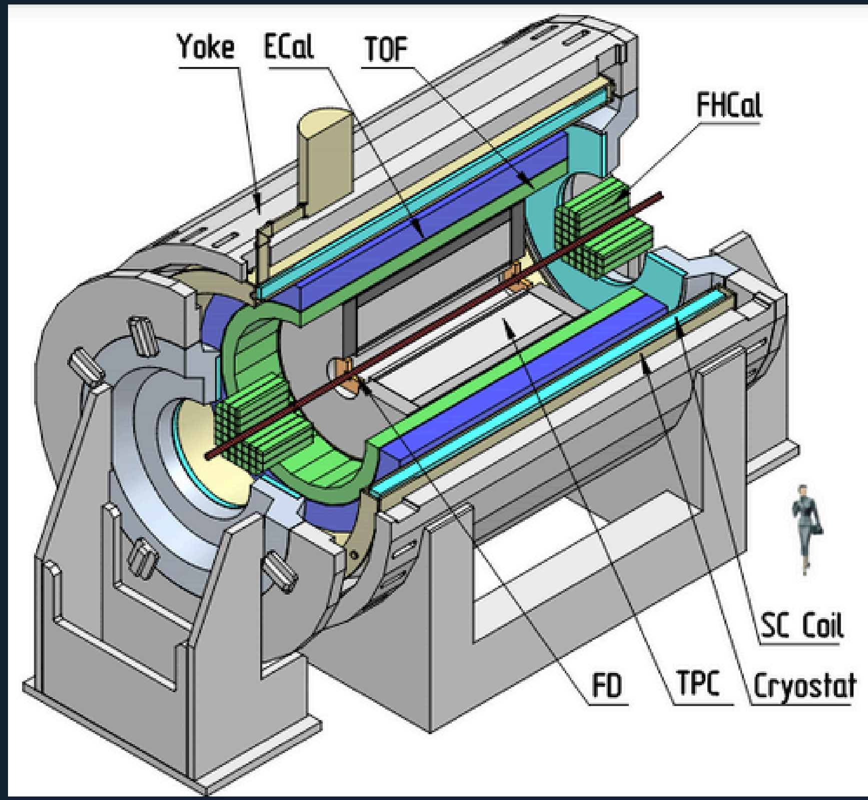
- PERSPECTIVES OF MULTISTRANGE HYPERON STUDY AT NICA/MPD FROM REALISTIC MONTE CARLO SIMULATION
- MACHINE LEARNING APPLICATION FOR PARTICLE IDENTIFICATION IN MPD
- MY OWN TASKS
- INSTALLATION OF THE MPDROOT
- DOUBTS

1 — PERSPECTIVES OF MULTISTRANGE HYPERON STUDY AT NICA/MPD FROM REALISTIC MONTE CARLO SIMULATION

HOW IS THE INFORMATION OBTAINED?

- EVENT AND DATA SET GENERATION
- DETECTOR PERFORMANCE
- TRACK AND VERTEX RECONSTRUCTION
- PARTICLE IDENTIFICATION

One of the main tasks of the NICA/MPD physics program is a study of strangeness production in nuclear collisions. In this paper the MPD detector performance for measurements of multistrange hyperons in Au +Au collisions at NICA energies is presented based on the analysis of realistically simulated data samples.



1 EVENT AND DATA SET GENERATION

SPECIFICATIONS

Processed data sets were produced by two event generators: UrQMD and PHSD.

- The UrQMD consists of 40 thousand central (0-3.0 fm) gold + gold events at SNN =9 GeV.
- The PHSD event sample contains 8 million minimum bias Au + Au events at SNN= 11 GeV.

Used GEANT3 transport package.

2 DETECTOR PERFORMANCE

TRACK AND VERTEX RECONSTRUCTION

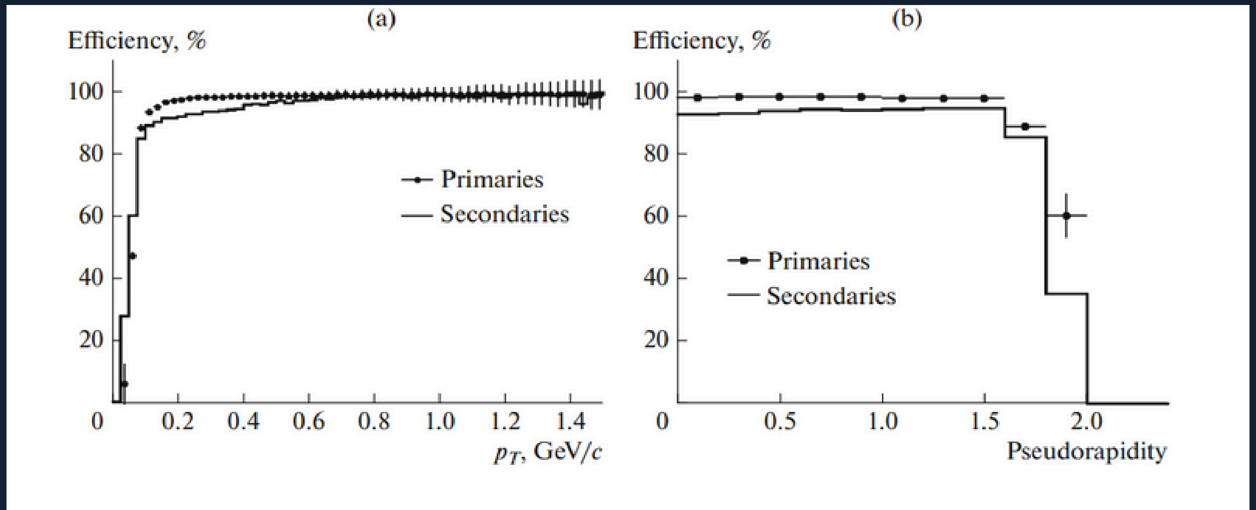


Fig. 2. Track reconstruction efficiency for particles with the number of measured points in the TPC (hits) greater than 14: (a) as a function of p_T for $|\eta| < 1.3$; (b) as a function of $|\eta|$ for $p_T > 0.1$ GeV/c. Symbols and lines present primary and secondary particles, respectively. Secondary particles were produced within 50 cm from the interaction point.

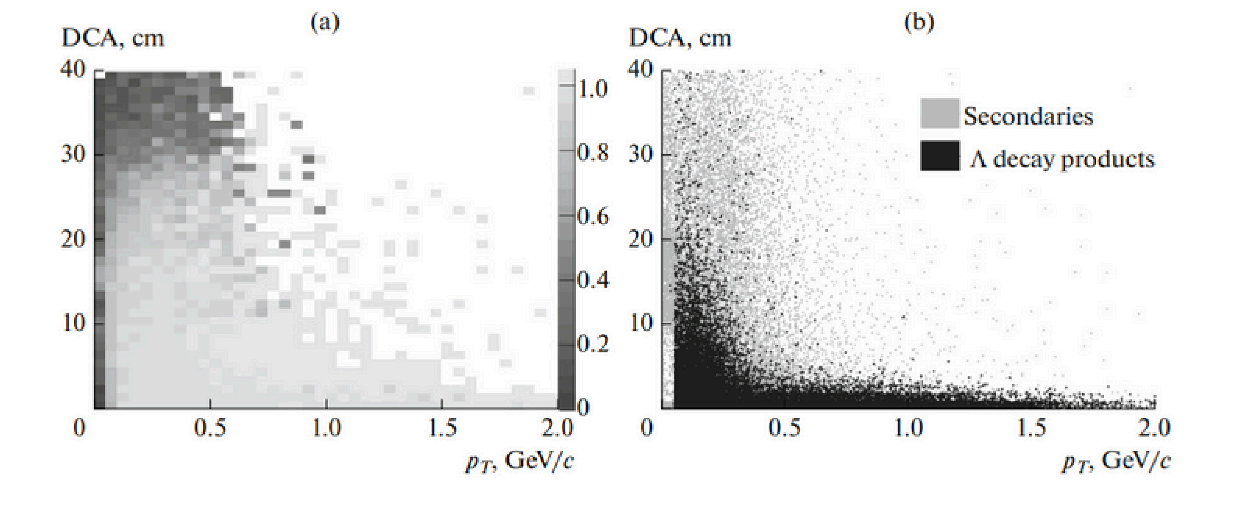
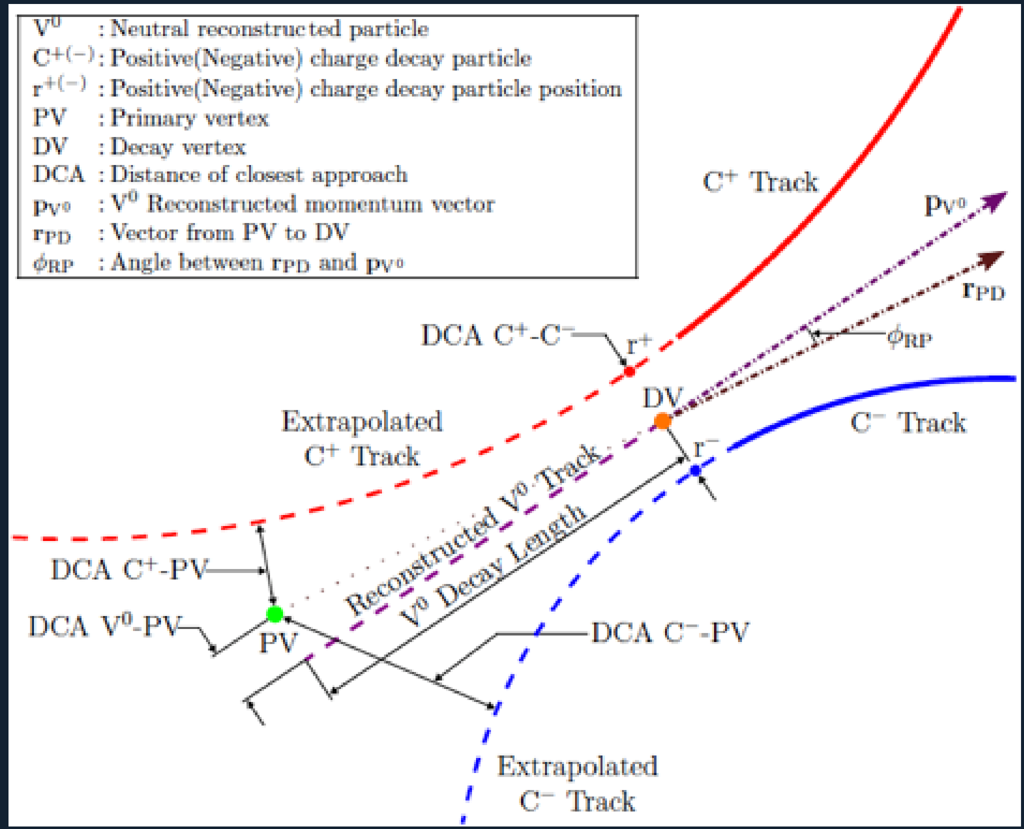


Fig. 3. (a) Secondary track reconstruction efficiency vs. DCA and p_T; (b) DCA vs. p_T for all secondary particles (grey dots) and Λ hyperon decay products (black dots).

We note that the efficiency starts to decrease for Pt below 0.6 GeV/c. Such behavior can be understood from the graph of Pt vs DCA, for large DCA values the efficiency begins to decrease.



2 — DETECTOR PERFORMANCE

PARTICLE IDENTIFICATION

Particle identification in the MPD experiment will be achieved by using the information about the energy loss in the TPC gas and the time-of-flight from the TOF detector.

The matching efficiency is plotted as a function of the total momentum. The efficiency is defined as a fraction of tracks having produced a Monte Carlo point in the TOF and matched with any TOF hit.

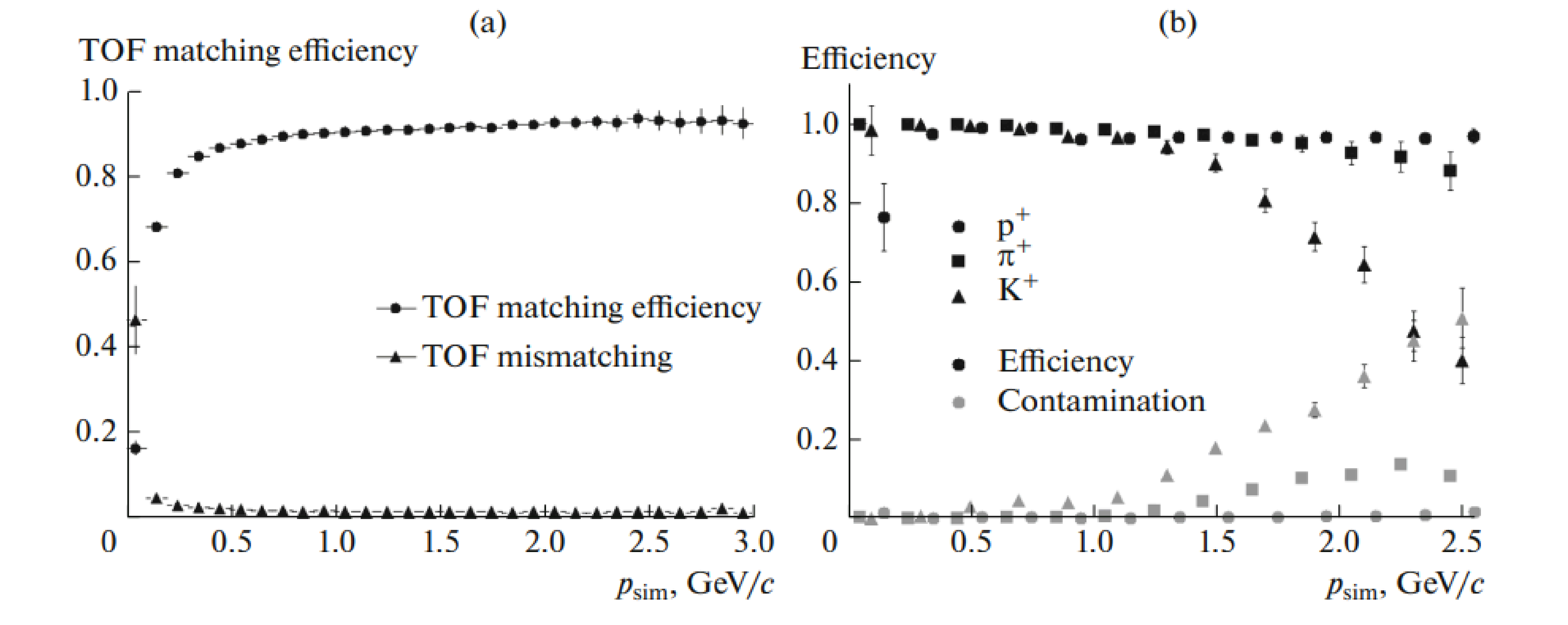


Fig. 7. (a) TOF matching efficiency; (b) Particle identification efficiency for positively charged hadrons (dark symbols) and a fraction of wrongly identified species (grey symbols) in Au + Au collisions at $\sqrt{s_{NN}} = 9$ GeV.

With the chosen set of cuts, charged kaons can be identified up to p=1.7 GeV/c with an approximately 80% efficiency and 20% contamination at the PID limit. Making the selection criteria for kaons tighter, the achieved contamination level can be decreased further resulting in a lower value for the PID efficiency.

3 QUESTIONS

Task 1. Primary vertex determination and Particle Track reconstruction, optimization of cuts in η, pT, number of hits on TPC.

Task 2. Particle identification determination of spectra using information about the energy losses (dE/dx) in the TPC and the Time-of-flight from the TOF detector.

Should both of us perform both tasks?

Are assigned tasks 1 and 2 similar to those mentioned above?

Should our graphs be like the ones shown above?

2 MACHINE LEARNING APPLICATION FOR PARTICLE IDENTIFICATION IN MPD

INTRODUCCTION

KEY CONCEPTS

MACHINE LEARNING

Machine learning in the case of the MPD experiment, machine learning methods, such as gradient boosting on decision trees (CatBoost), are used to enhance particle identification in regions where conventional methods fail, processing features from experimental data to correctly classify particle types.

CATEGORICAL BOOSTING (CATBOOST)

CatBoost in the MPD experiment, is employed to address the particle identification problem, using subdetector data to correctly classify particle types across different momentum ranges, demonstrating greater efficiency than conventional methods under certain conditions.

PARTICLE IDENTIFICATION (PID)

Particle Identification is the process of determining the nature and characteristics of individual particles detected in a particle physics experiment, using information obtained from specific detectors such as the Time-Projection Chamber (TPC) and Time-of-Flight (TOF) detector.

PARTICLE IDENTIFICATION IN MPD EXPERIMENT

WHAT IS UNDERSTOOD?

Particle identification (PID) in the MPD experiment relies on the TPC and TOF subdetectors, whose details are provided in specific technical reports. These subdetectors enable achieving the required identification by reconstructing trajectories and measuring properties such as momentum, charge, energy loss, squared mass, number of TPC hits, pseudorapidity, distance of closest approach, and vertex coordinates for six particle species: proton, positive and negative kaons, positive and negative pions, and antiproton.

The PID problem was approached as a multiclass classification task in machine learning, where a CatBoost classifier trained on three Monte Carlo datasets (prod01 with real distribution, prod04 and prod05 with uniform distribution) was used. These datasets were generated simulating Bismuth and Bismuth Bi + Bi collisions at 9.2 GeV, expected to be the first collision systems in MPD. Testing data was generated under the same conditions as prod05, and CatBoost parameters were tuned using the Tree-Structured Parzen Estimator algorithm in Optuna, following developer recommendations.

This study demonstrates that the CatBoost algorithm is effective in improving accuracy in particle classification in the MPD, excelling in identification at critical momentum ranges where conventional methods show limitations.

OUTCOMES

EXPLORAMOS LA SITUACIÓN ACTUAL DE LA COMPAÑIA

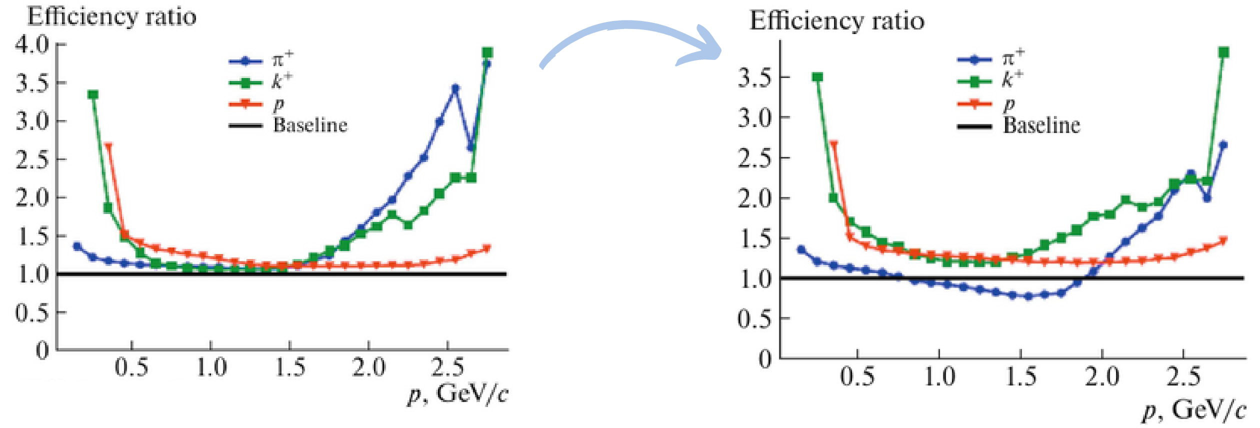
This section demonstrates the feasibility of using CatBoost for particle identification (PID) in the MPD experiment. Two metrics were used to evaluate model performance: E^s (efficiency) and C^s (contamination) for each particle species s. Efficiency E^s indicates how often the classifier provides correct answers, while contamination C^s describes the proportion of incorrectly classified tracks among those identified as type s.

$$E^s = \frac{N^s_{\text{corr}}}{N^s_{\text{true}}}, \qquad C^s = \frac{N^s_{\text{incorr}}}{N^s_{\text{corr}} + N^s_{\text{incorr}}},$$

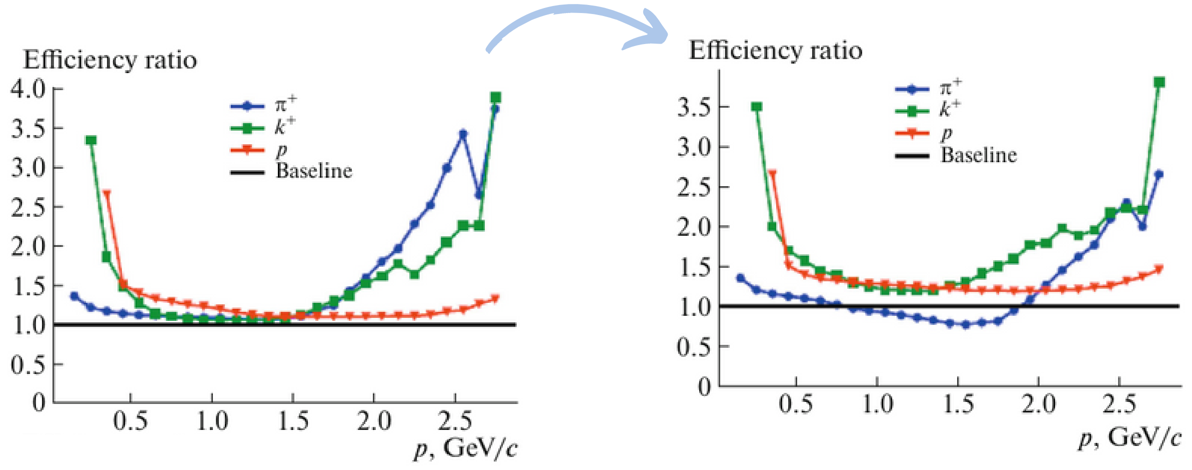
Notably, protons outnumber positive pions sevenfold when p>1.5p > 1.5p>1.5 GeV/c, affecting classification dynamics. Overall classifier efficiencies on test data are approximately 96.48%, 96.71%, and 95.99%, indicating robust performance, particularly in distinguishing between protons and pions across different production scenarios.

Similar trends are observed for negatively charged pions, where identification efficiency decreases as production shifts from prod01 to prod05, albeit with significantly lower contamination. Conversely, antiproton contamination exceeds that of protons, highlighting challenges in distinguishing these particle types.

The PID performance is depicted in Figure 1, showing how particle identification efficiency varies across different productions from prod01 to prod05 for middle and high momentum values. Proton efficiencies increase by an average of 10%, whereas the efficiency and contamination of positively charged pions decrease by approximately 20%.



CatBoost classifiers were compared with the n-sigma method currently used at MPD, as shown in Figure 2, providing a clearer assessment of the obtained results and confirming CatBoost's effectiveness in enhancing particle identification compared to conventional methods.



CONCLUSION

The conclusion of the study highlights that CatBoost demonstrates promising efficiency in particle identification in the MPD experiment, outperforming the currently used n-sigma method in several cases. Specifically, CatBoost showed better results in extreme momentum ranges ($p < 0.7$ GeV/c and $p > 1.5$ GeV/c), where the n-sigma method has significant limitations or lacks efficiency. This finding suggests that machine learning methods like CatBoost are suitable and offer substantial improvements in particle identification accuracy compared to traditional approaches used in particle physics. Furthermore, the study proposes continued exploration of additional particle characteristics to enhance model stability and performance across various experimental conditions of the MPD detector.

3 — OUR OWN TASKS

TASK 1

Primary vertex determination and Particle Track reconstruction, optimization of cuts in η, pT, number of hits on TPC.

I

GET TRACK RECONSTRUCTION EFFICIENCY AS A FUNCTION OF THE TRANSVERSE MOMENTUM (PT) OF TRACK, FOR PRIMARY AND SECONDARIES.

Objetivo: 40% captación de mercado

II

GET RELATIVE TRANSVERSE MOMENTUM RESOLUTION FOR PRIMARY TRACKS FOR THE PHASE SPACE (Η AND PT) AVAILABLE FOR TPC.

Objetivo: 7% de crecimiento en CTR

III

COMPARISON OF RESULTS WITH DEFAULT MAGNETIC FIELD.

I’m not sure about this part of the task.

TASK 2

Particle identification determination of spectra using information about the energy losses (dE/dx) in the TPC and the Time-of-flight from the TOF detector.

I

GET THE MOMENTUM DEPENDENCE OF THE AVERAGE ENERGY LOSS AND THE MASS SQUARED RESOLUTION FOR CHARGED PARTICLES.

Objetivo: 40% captación de mercado

II

GET PARTICLE IDENTIFICATION EFFICIENCY AS A FUNCTION OF MOMENTUM WITH TPC.

Objetivo: 7% de crecimiento en CTR

III

COMPARISON OF RESULTS WITH DEFAULT MAGNETIC FIELD.

I'm not sure about this part of the task.

4 – INSTALLATION OF THE MPDROOT

I CANNOT INSTALL THE NEW VERSION OF MPDROOT

ON THE CLUSTER

Could not connect to server.

```
ssh: connect to host ncx103.jinr.ru port 22: Connection timed out
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ ssh user@ncx103.jinr.ru
ssh: connect to host ncx103.jinr.ru port 22: Connection timed out
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ ssh user@ncx103.jinr.ru
ssh: connect to host ncx103.jinr.ru port 22: Connection timed out
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$
```

LOCALLY USING CVMFS

Could not read from remote repository.

```
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ toolbox enter a9-nica-dev
carlos@toolbox:~$  git clone -b dev --recursive git@git.jinr.ru:nica/mpdroot.git
Cloning into 'mpdroot'...
The authenticity of host 'git.jinr.ru (159.93.223.249)' can't be established.
ED25519 key fingerprint is SHA256:JVZWtGU6WQW6IXN7/Zp6zJ914lJmPsTnFbtlH56VMNA.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'git.jinr.ru' (ED25519) to the list of known hosts.
git@git.jinr.ru: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
fatal: Could not read from remote repository.

Please make sure you have the correct access rights
and the repository exists.
carlos@toolbox:~$ git clone -b dev --recursive git@git.jinr.ru:nica/mpdroot.git
Cloning into 'mpdroot'...
git@git.jinr.ru: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
fatal: Could not read from remote repository.

Please make sure you have the correct access rights
and the repository exists.
```

LOCALLY USING ALIBUILD

Could not read from remote repository.

```
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ wget -N https://git.jinr.ru/nica/nicadist/-/
raw/master/scripts/nica-init.sh --no-check-certificate
--2024-06-17 22:59:18--  https://git.jinr.ru/nica/nicadist/-/raw/master/scripts/nica-init.sh
Resolviendo git.jinr.ru (git.jinr.ru)... 159.93.223.249
Conectando con git.jinr.ru (git.jinr.ru)[159.93.223.249]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 12805 (13K) [text/plain]
Guardando como: ‘nica-init.sh’

nica-init.sh                                   100%[======================================
========================================================================>]  12.50K  --.
-KB/s    en 0s

Falta la cabecera de fecha de la última modificación -- marcas de tiempo desactivadas.
2024-06-17 22:59:20 (95.3 MB/s) - ‘nica-init.sh’ guardado [12805/12805]

carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ chmod +x nica-init.sh && ./nica-init.sh
Installing toolbox on Ubuntu 22.04
[sudo] contraseña para carlos:
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
curl ya está en su versión más reciente (7.81.0-1ubuntu1.16).
0 actualizados, 0 nuevos se instalarán, 0 para eliminar y 111 no actualizados.
flatpak ya está en su versión más reciente (1.12.7-1).
curl ya está en su versión más reciente (7.81.0-1ubuntu1.16).
podman ya está en su versión más reciente (3.4.4+ds1-1ubuntu1.22.04.2).
0 actualizados, 0 nuevos se instalarán, 0 para eliminar y 111 no actualizados.
--2024-06-17 22:59:38--  http://old.kali.org/kali/pool/main/g/golang-github-containers-toolbox/
podman-toolbox_0.0.99.2-2_amd64.deb
Resolviendo old.kali.org (old.kali.org)... 54.39.49.227
Conectando con old.kali.org (old.kali.org)[54.39.49.227]:80... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 1421872 (1.4M) [application/octet-stream]
Guardando como: ‘podman-toolbox.deb’

podman-toolbox.deb                             100%[======================================
========================================================================>]   1.36M  2.1
1MB/s    en 0.6s

2024-06-17 22:59:39 (2.11 MB/s) - ‘podman-toolbox.deb’ guardado [1421872/1421872]

Se DESACTUALIZARÁN los siguientes paquetes:
  podman-toolbox
0 actualizados, 0 nuevos se instalarán, 1 desactualizados, 0 para eliminar y 111 no actualizado
s.
E: Se instalaron versiones anteriores de algunos paquetes y se utilizó -y sin --allow-downgrade
s.
carlos@carlos-VivoBook-ASUSLaptop-X513EA-K513EA:~$ toolbox enter a9-nica-dev
carlos@toolbox:~$ git clone -b dev --recursive git@git.jinr.ru:nica/mpdroot.git
Cloning into 'mpdroot'...
git@git.jinr.ru: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
fatal: Could not read from remote repository.

Please make sure you have the correct access rights
and the repository exists.
carlos@toolbox:~$ exit
logout
```

5 — DOUBTS

RUTA DE ACCIÓN

¿QUÉ VAMOS A HACER PARA CONSEGUIR NUESTRO OBJETIVO?

ACCIÓN NO. 1

TÍTULAR

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus ex lacus, imperdiet sed mauris vel, maximus varius diam. Nunc interdum imperdiet dui, id aliquet massa rhoncus vel. Proin sed aliquam nulla.

ACCIÓN NO. 2

TÍTULAR

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus ex lacus, imperdiet sed mauris vel, maximus varius diam. Nunc interdum imperdiet dui, id aliquet massa rhoncus vel. Proin sed aliquam nulla.

OUR OWN TASKS

TÍTULAR

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus ex lacus, imperdiet sed mauris vel, maximus varius diam. Nunc interdum imperdiet dui, id aliquet massa rhoncus vel. Proin sed aliquam nulla.

INSTALLATION OF THE MPDROOT

TÍTULAR

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus ex lacus, imperdiet sed mauris vel, maximus varius diam. Nunc interdum imperdiet dui, id aliquet massa rhoncus vel. Proin sed aliquam nulla.

THANK YOU

ALEJANDRO SAN JUAN & JUAN CARLOS MARQUEZ

alejandrosanjuan59@gmail.com & juan.marquezra@gmail.com