NARODOWE CENTRUM NAUKI

Second Generation HLT Tracking Algorithm for Long-lived Particles Reconstruction in Upgraded LHCb Experiment

Sabin Hashmi - AGH UST, Krakow

On behalf of LHCb Collaboration

XXVIII - Cracow Epiphany Conference

10-14 January 2022

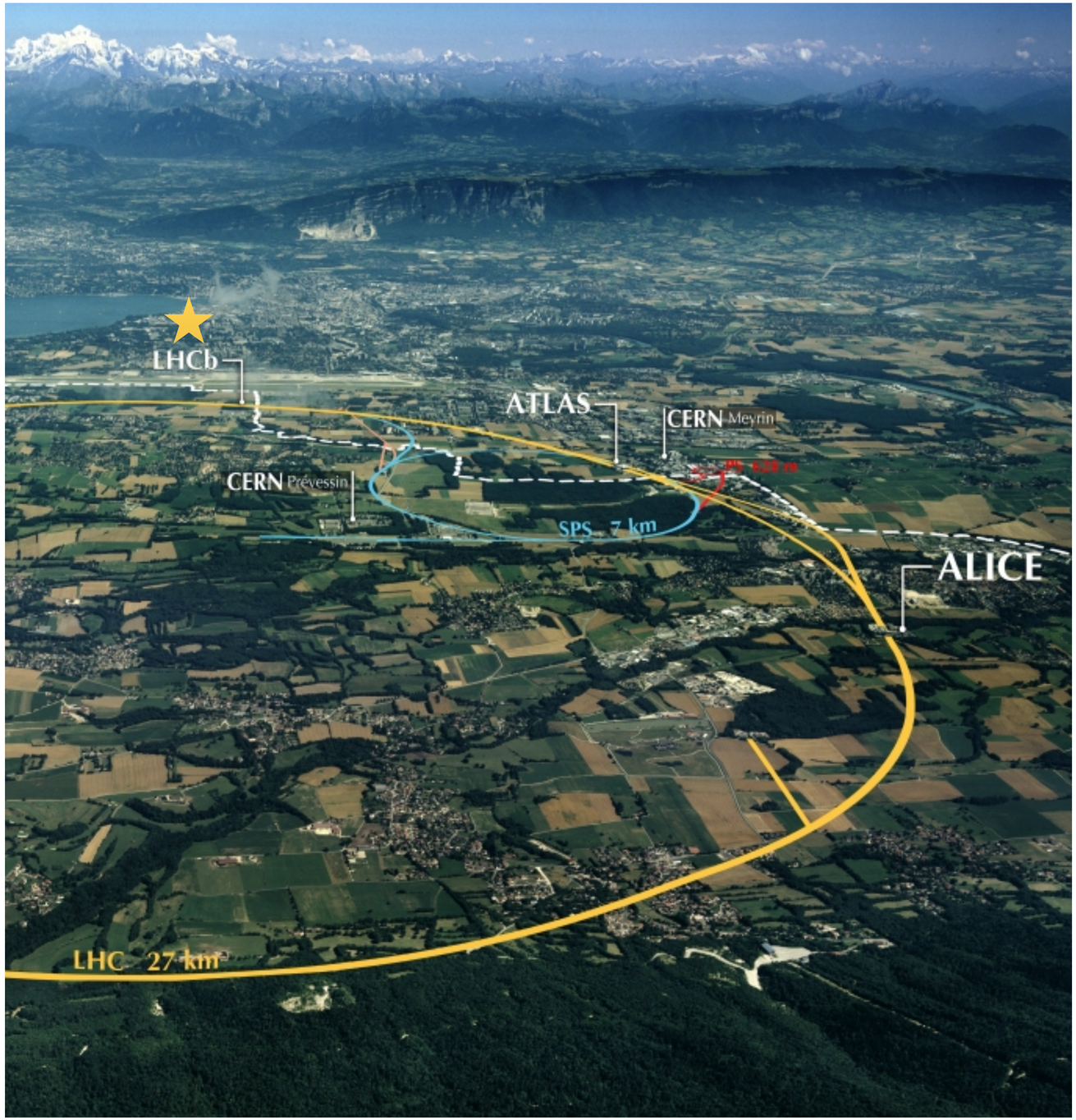
AGH

RESEARCH UNIVERSITY EXCELLENCE INITIATIVE

LHCb

Outline

1. Overview of Research
2. LHCb Upgrade - 1
3. High-Level Trigger Systems
4. Machine Learning in HEP
5. Track Segment Classifier

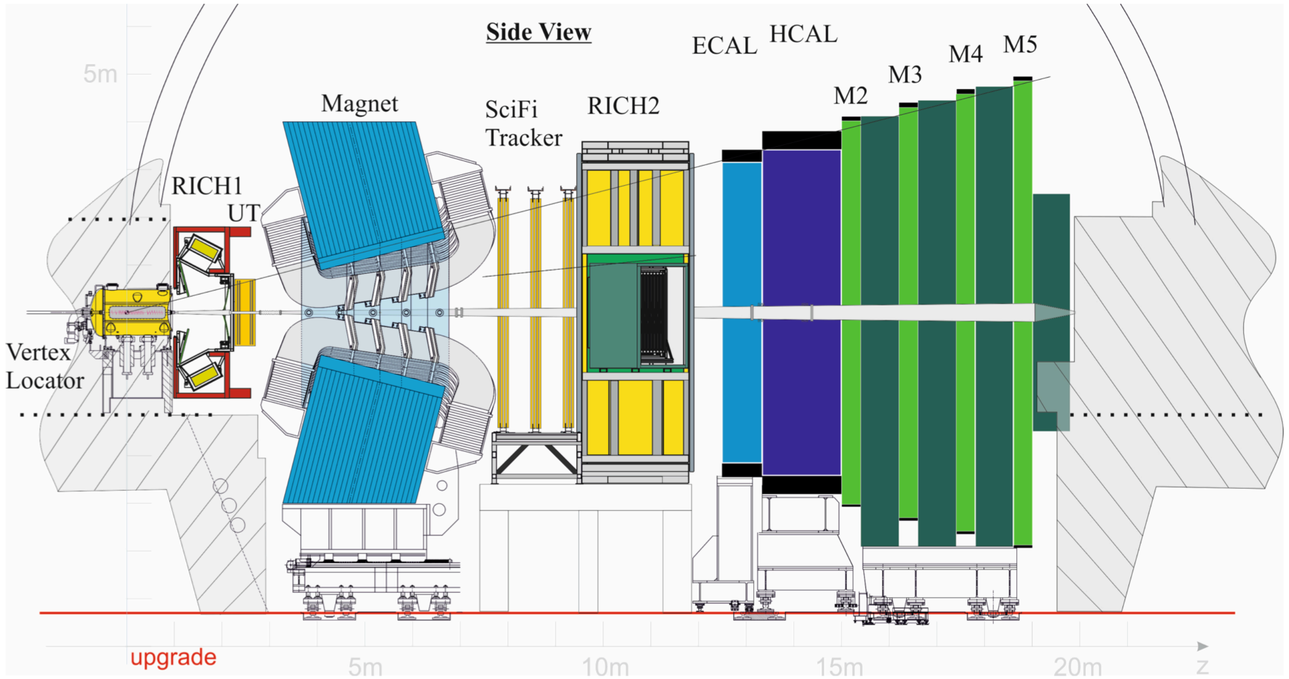


LHCb Physics and Overview of Research

- Mystery of Matter and Antimatter Distribution.
- Study about Rare Decays of Heavy Flavour Physics (Beauty and Charm-Quarks).
- Proton-Proton Crossing rate: Once in every 25ns (40MHz)
- Older Generation Readout: 1MHz
- General Purpose Forward Detector.
- Run 3 / Upgrade 1 in 2022.
- Storage of 10GB of Filtered Data in a second.
- Sophisticated Trigger System utilizing AI
- No Storage of RAW Data!

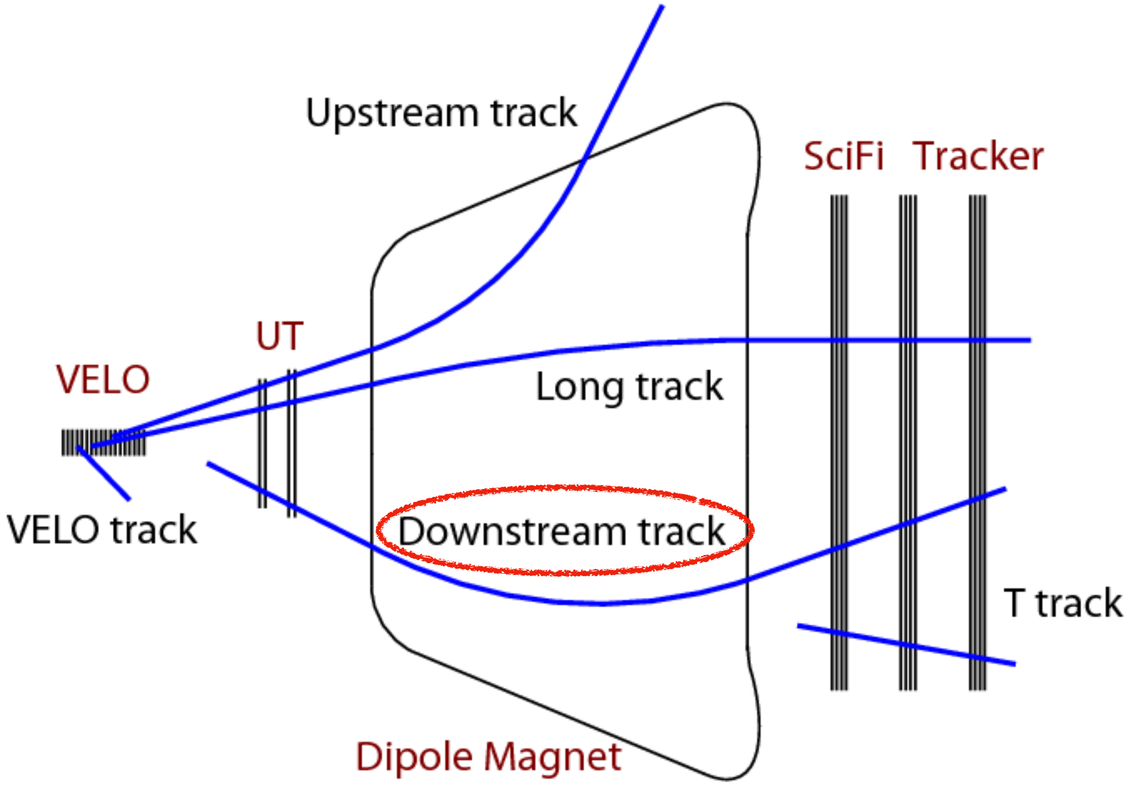
LHCb Upgrade - 1

- Run 1 and 2 provide high precision measurements using 8 $[fb^{-1}]$
- In 2022, Run 3 is expected to collect data at 50 $[fb^{-1}]$
- Replacement of Readout-Electronics
- Transition to Software Trigger System utilizing AI from Hardware Trigger System
- Overcoming Hardware Trigger (L0) event rate of 1MHz to 30Mhz by using only Software Trigger



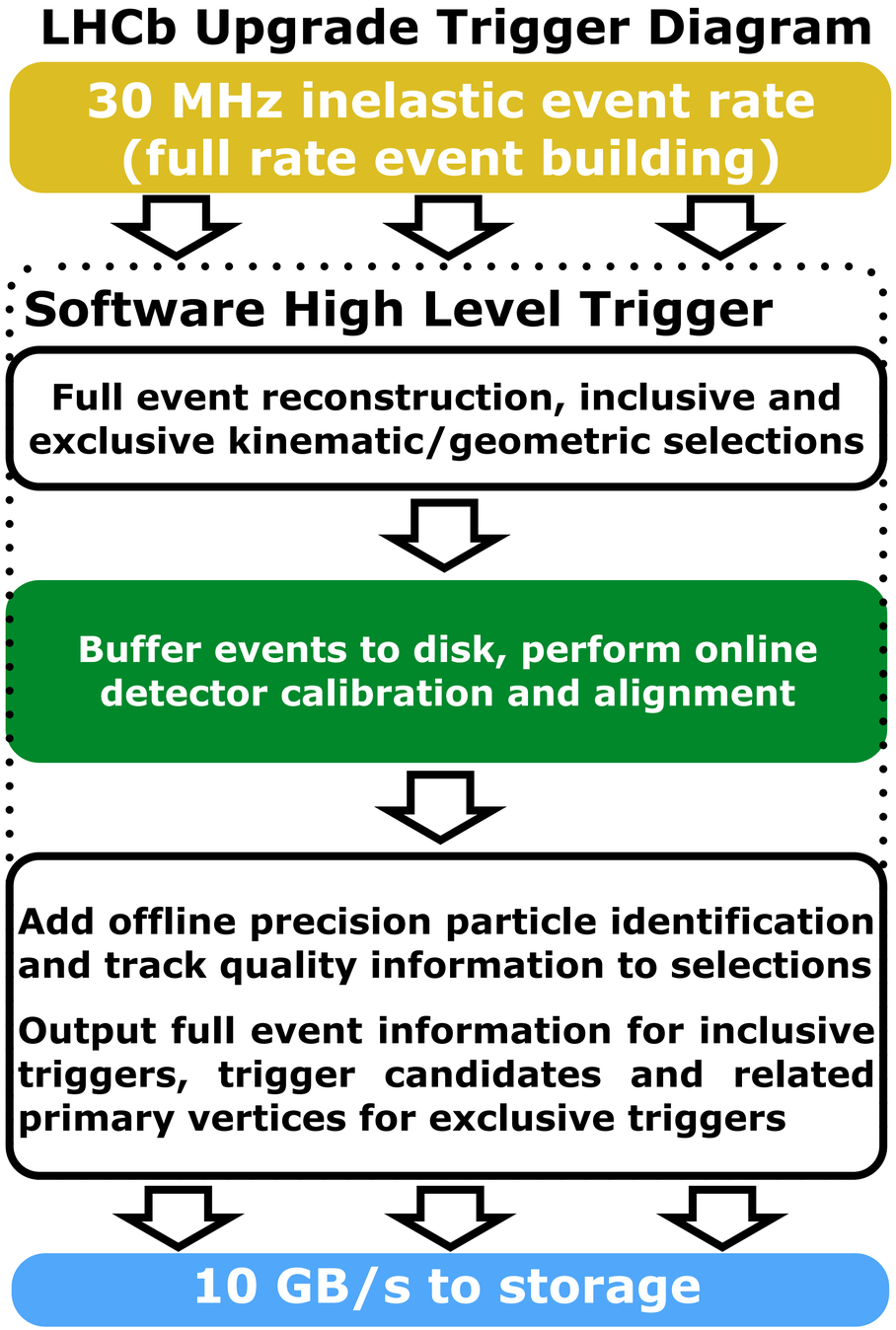
LHCb Tracking System

- Main Components
  - Vertex Locator (VELO): Precise measure of track coordinates and vertices
  - Dipole Magnets: Provide Strong Magnetic Field
  - UT and SciFi: Provide track coordinates to measure the momentum of particles.
- Long-Lived Particles decay after VELO.
- Downstream Tracks signals: UT and SciFi Tracker, no hits in VELO

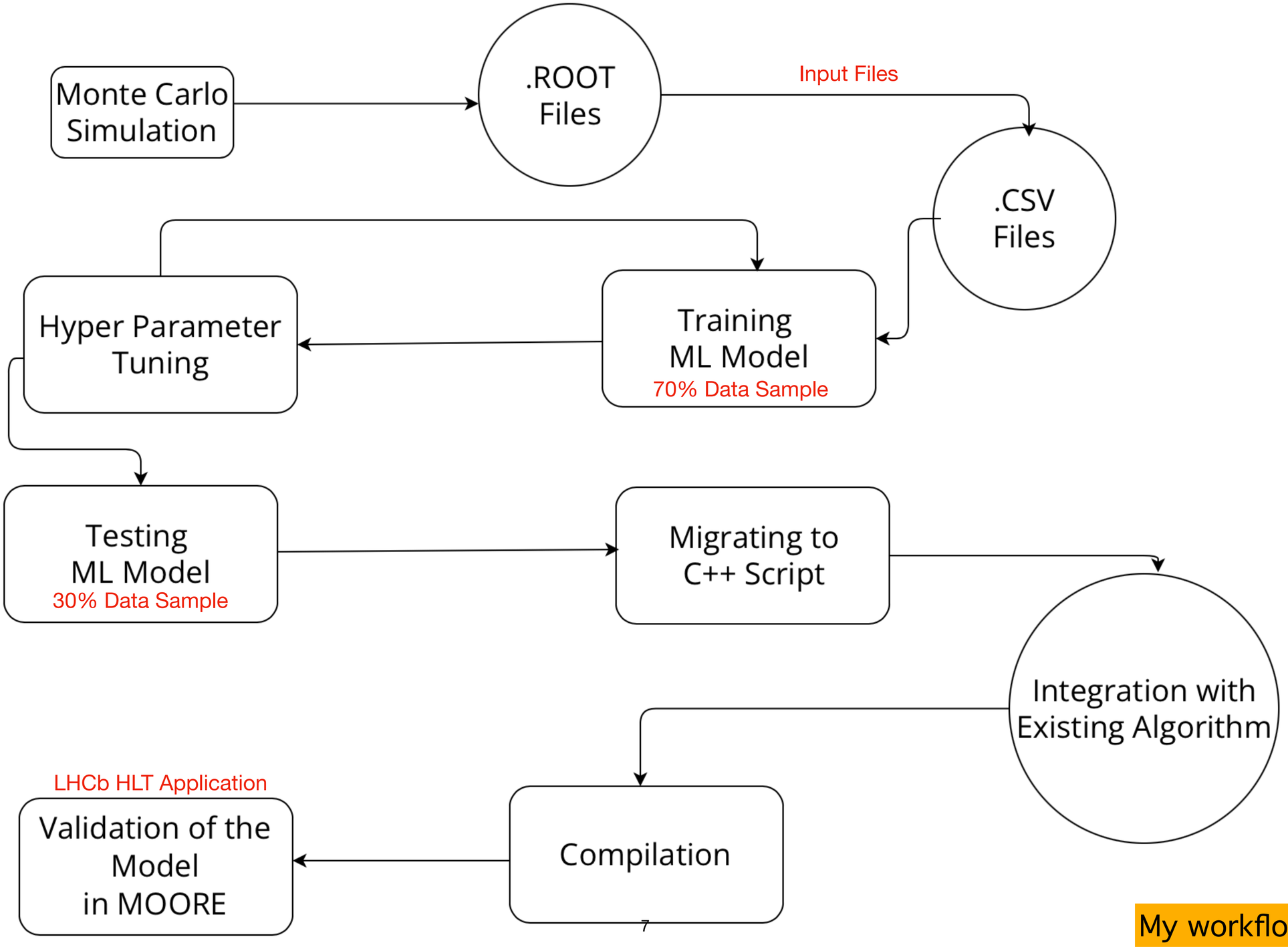


High Level Trigger (HLT)

- Trigger System works as a filter to accept/ reject events.
- Traditional HLT
  - L0 (Hardware Level Trigger)
  - HLT (Software Level Trigger)
- Upgraded HLT
  - HLT-1 : Partial Reconstruction
  - HLT -2 : Full event reconstruction
- L0 (1MHz) was a bottleneck, that was completely replaced by a software trigger system.
- To storage tape - Increment from 0.6GB/s to 10GB/s



Enhanced Workpipline



My workflow in a Nutshell

Input Data and Distributions

1. chi2PerDoF : *Chi2 per Degree of Freedom*
2. nLHCbID: *Number of Hits*
3. p: *Momentum of the Track*
4. phi : Azimuthal Angle
5. x: *x co-ordinates*
6. y: y *coordinates*
7. r: *Position Vector*
8. tx : *slope-x*
9. ty : *slope-y*
10. pseudoRapidity : *Spatial Co-Ordinate*

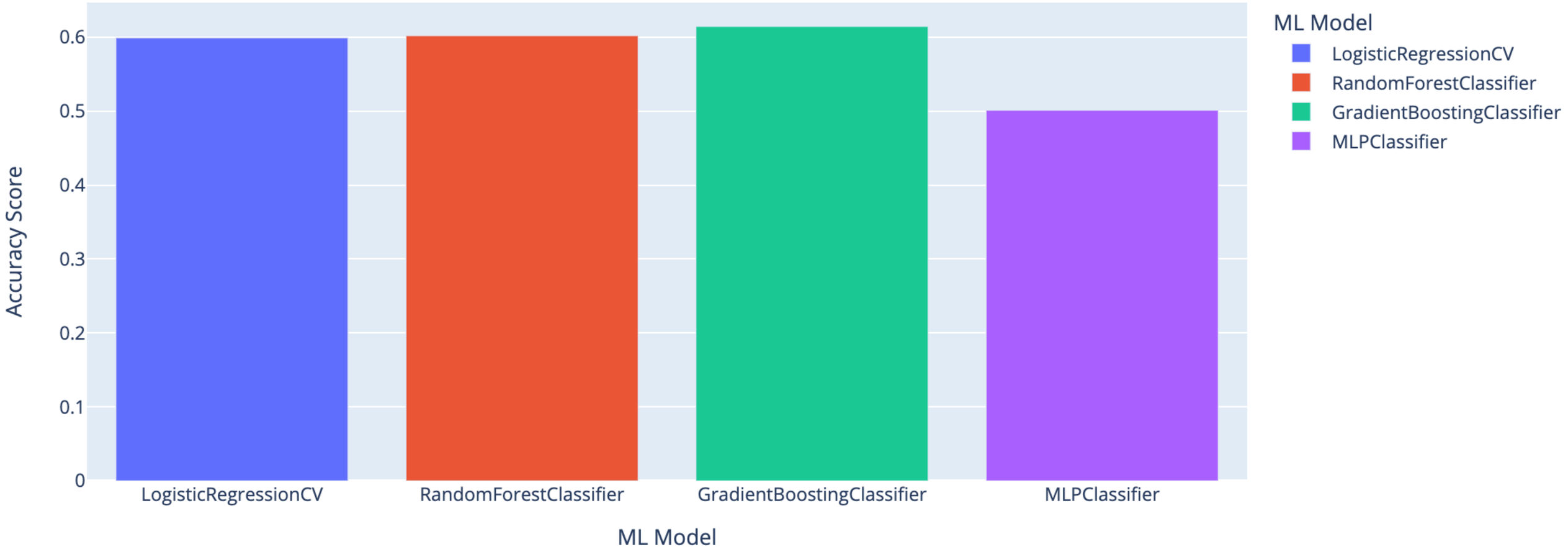
$$Decay : \Lambda_b^0 \rightarrow \Lambda^0(\rightarrow p\pi^-)\mu^+\mu^-$$

Output : Real Track/Ghost Track

Algorithm Design

- Downstream tracks are crucial to studying long-lived particle decay.
- Due to the intensity of signals, it is challenging to reconstruct the DS tracks.
- By the end, Downstream hits have to filter out from a large pool of hits and tracks.
- The majority of tracks are not DS
- ML-based classifier filters out the minority of Real DS Tracks(Signal) from the Ghost Tracks (Noise)

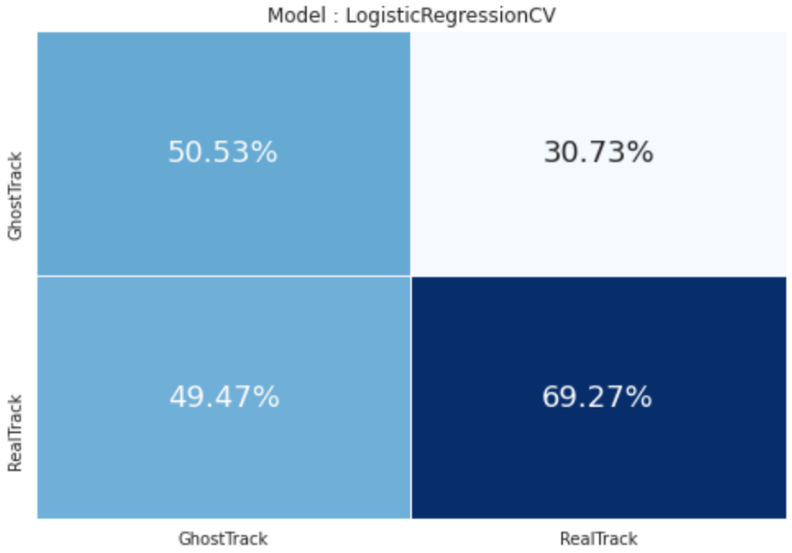
Baseline Model Performance

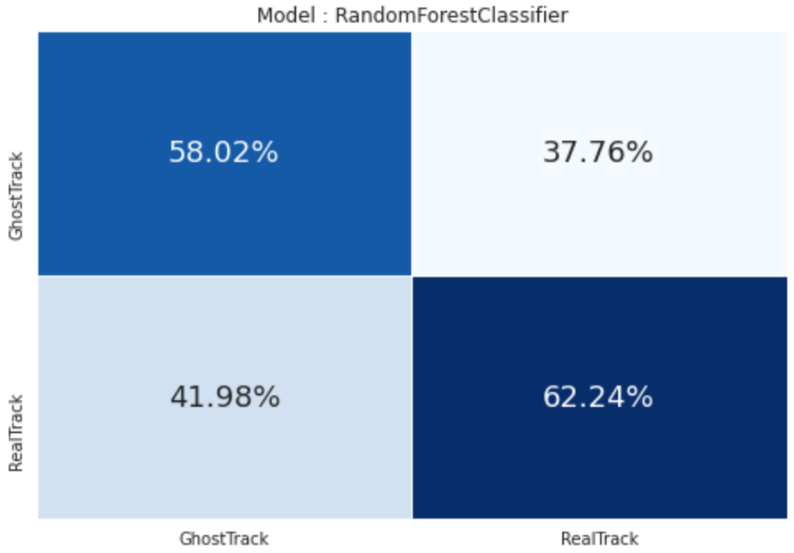


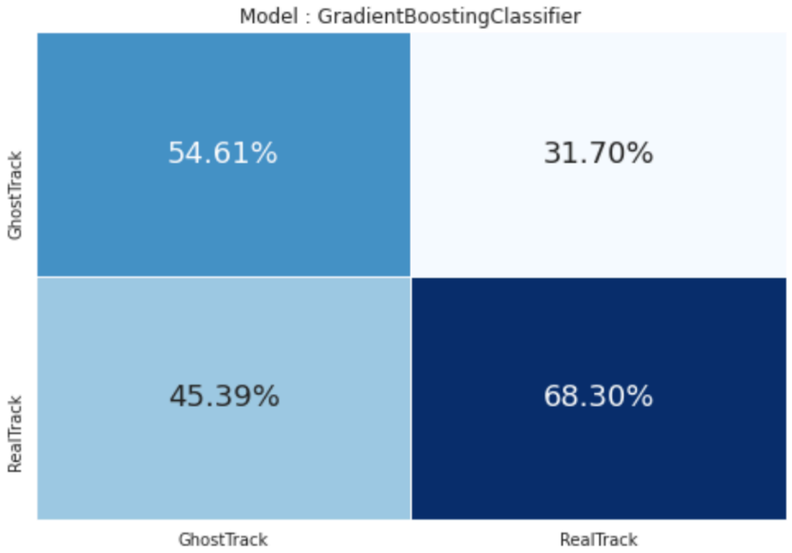
Four Most Popular ML Models and their Performance.
Accuracy Score = TP + TN / (TP+FP+TN+FN)

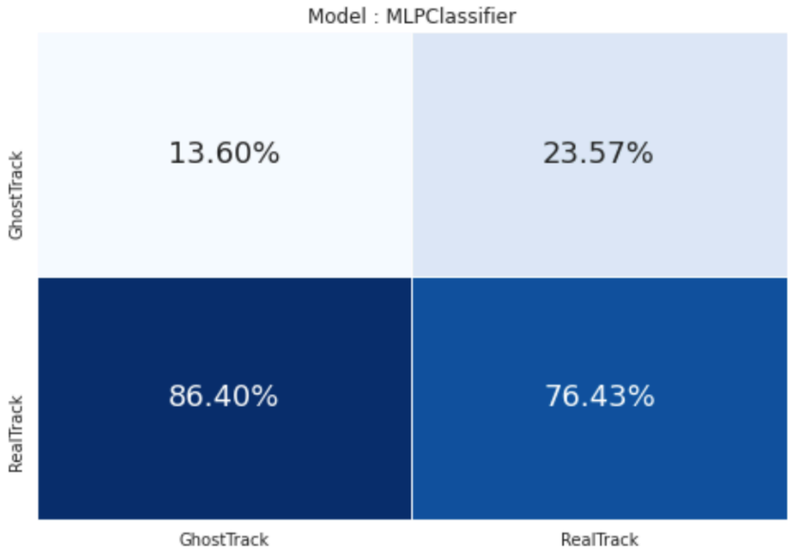
LHCb Simulation Preliminary

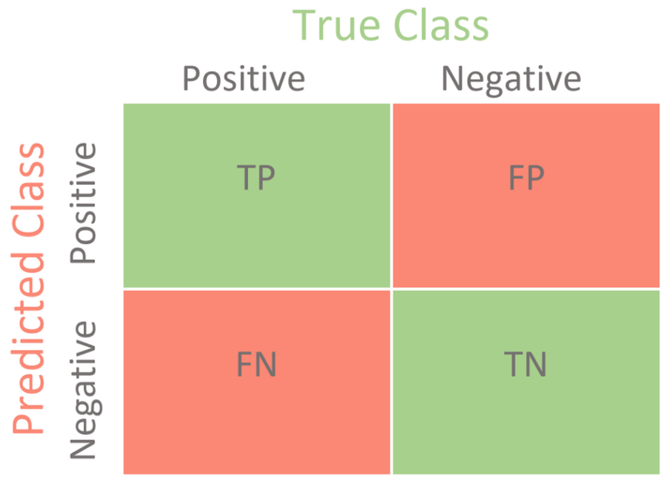
Baseline Model Performance - Confusion Matrix











$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

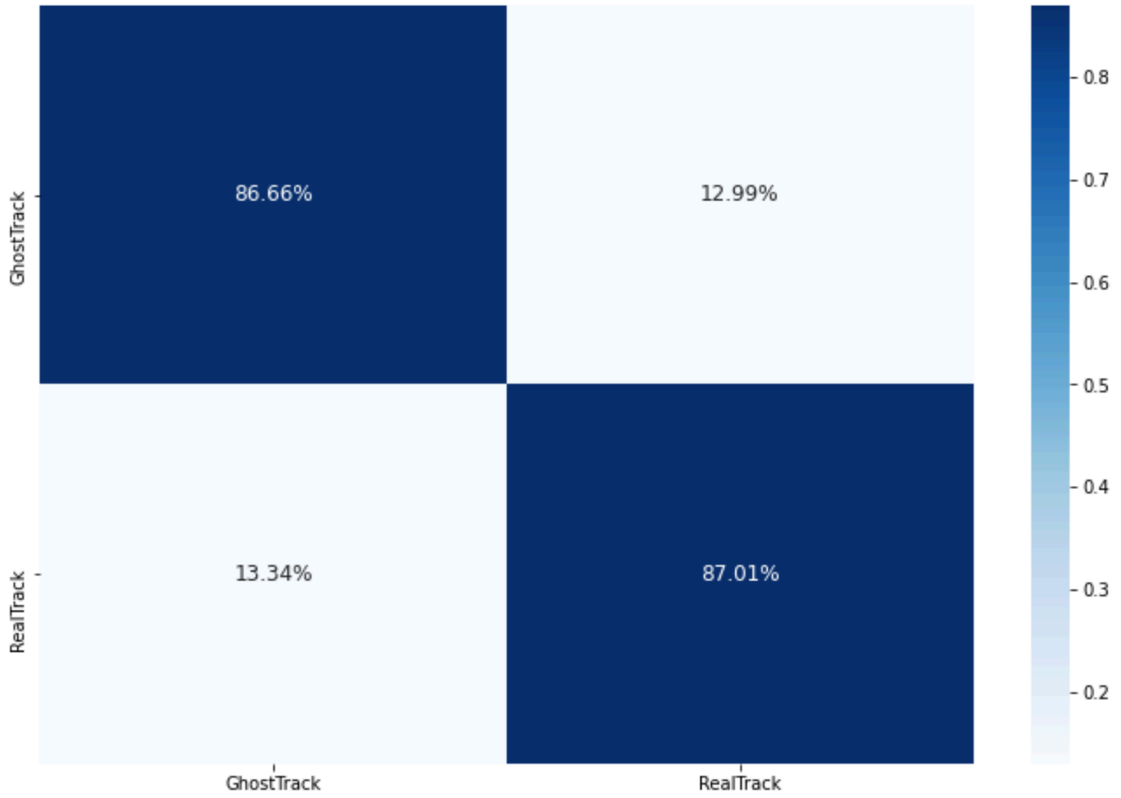
$$F1Score = \frac{2 \times precision \times \ recall}{precision + recall}$$

LHCb Simulation Preliminary

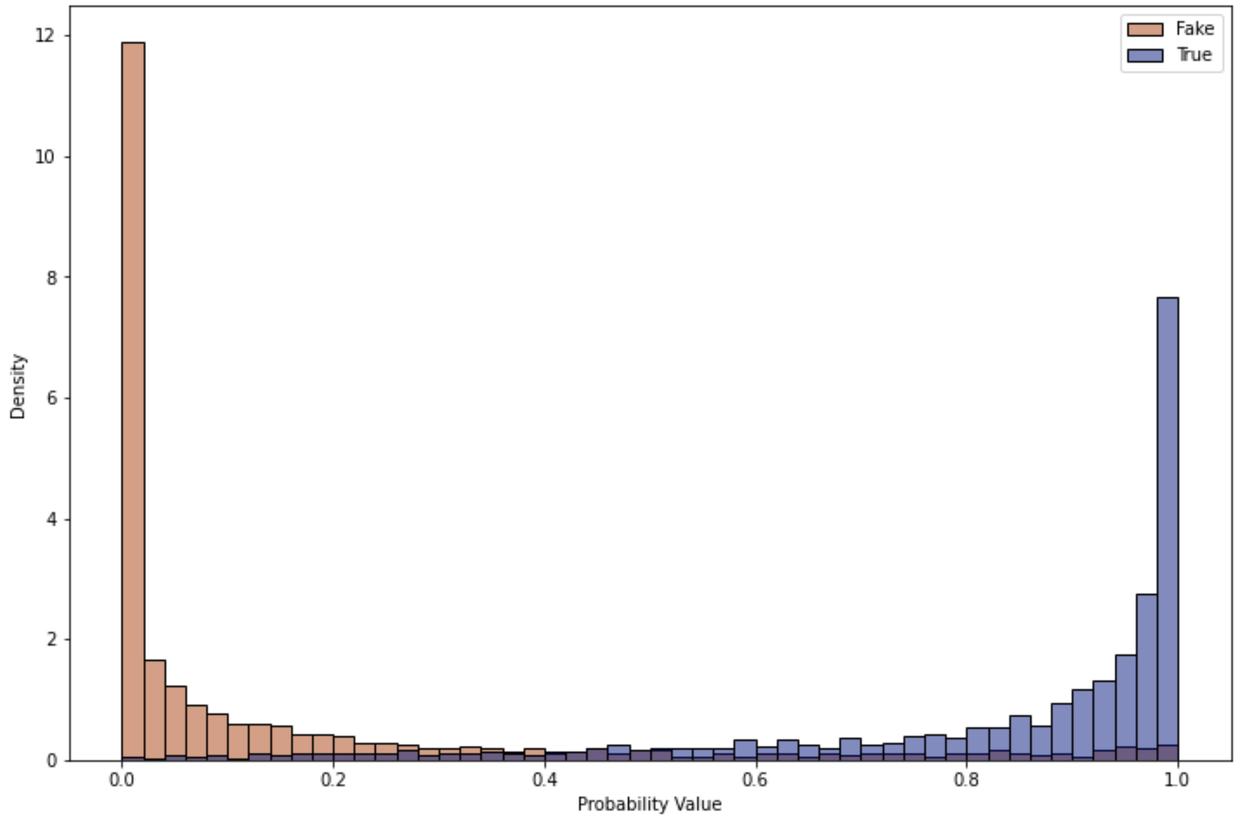
Track Segment Classifier

1. **Final Machine Learning Algorithm Used: CatBoost Classifier**
2. **Metrics of Evaluation:**
    1. **F1 Score,**
    2. **Area Under Curve,**
    3. **Accuracy Score**
3. **Catboost based classifier works better as compared to the other models evaluated to identify ghost tracks.**

Track Segment Classifier : Initial Results



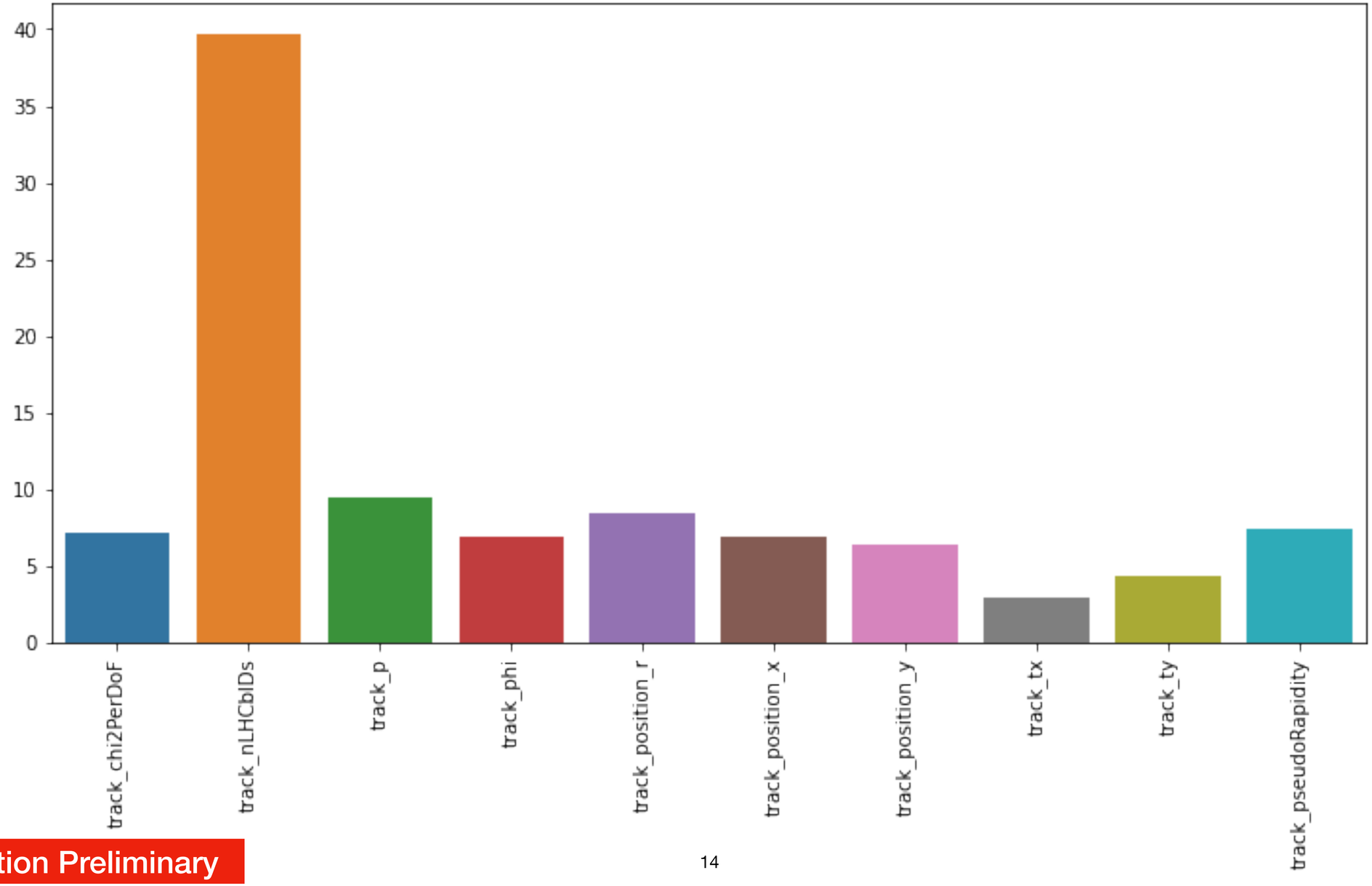
Track Segment Classifier :
Confusion Matrix



Track Segment Classifier :
Probability Distribution

LHCb Simulation Preliminary

Model Feature Importance



LHCb Simulation Preliminary

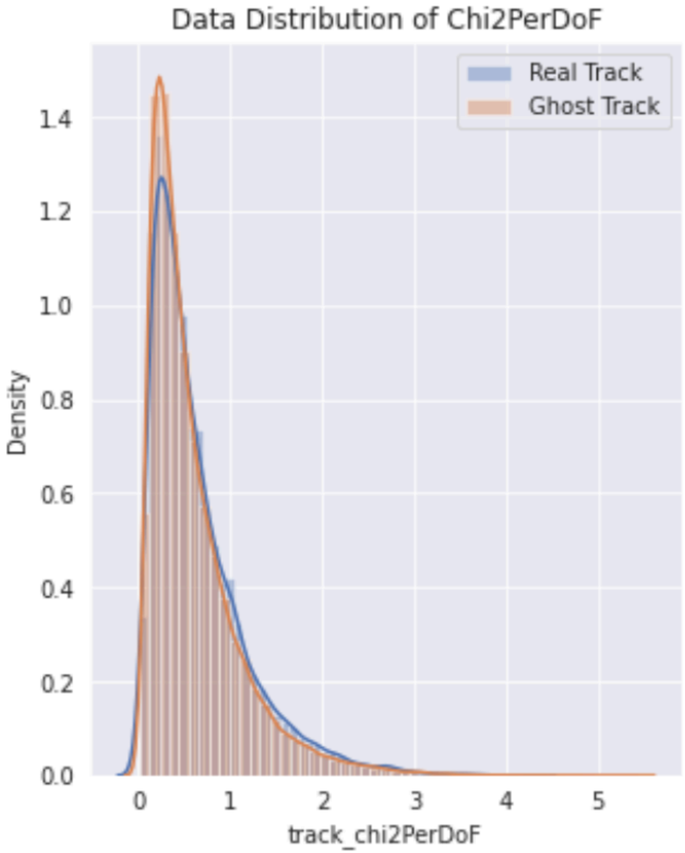
Conclusion

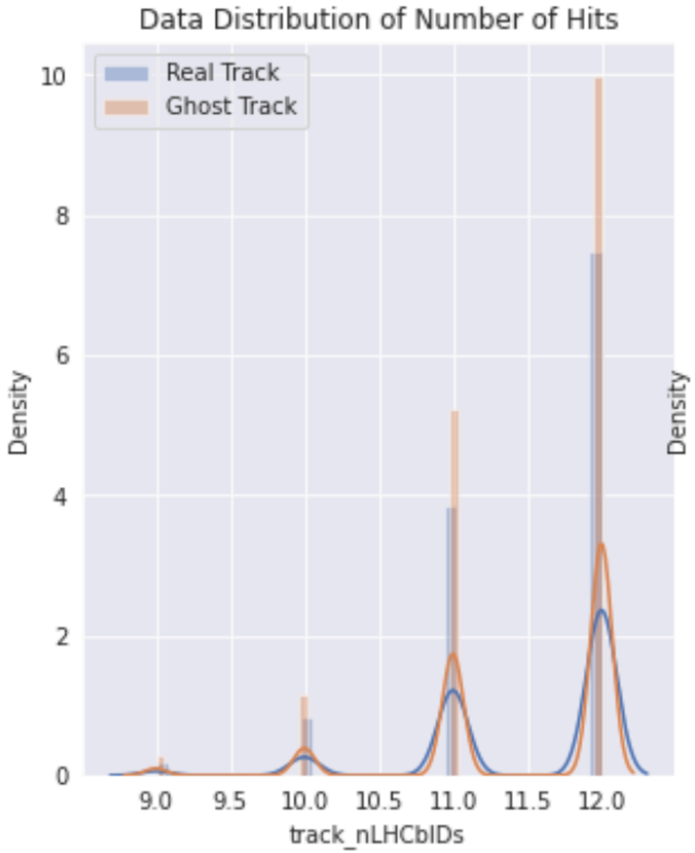
- On Run 3, a Software Trigger System will be implemented utilizing ML components.
- Downstream tracking algorithm similar to the one commissioned for Run2.
- Machine Learning filtering can improve the purity of the sample.
- CatBoost Based Machine Learning Model shows promising initial results.

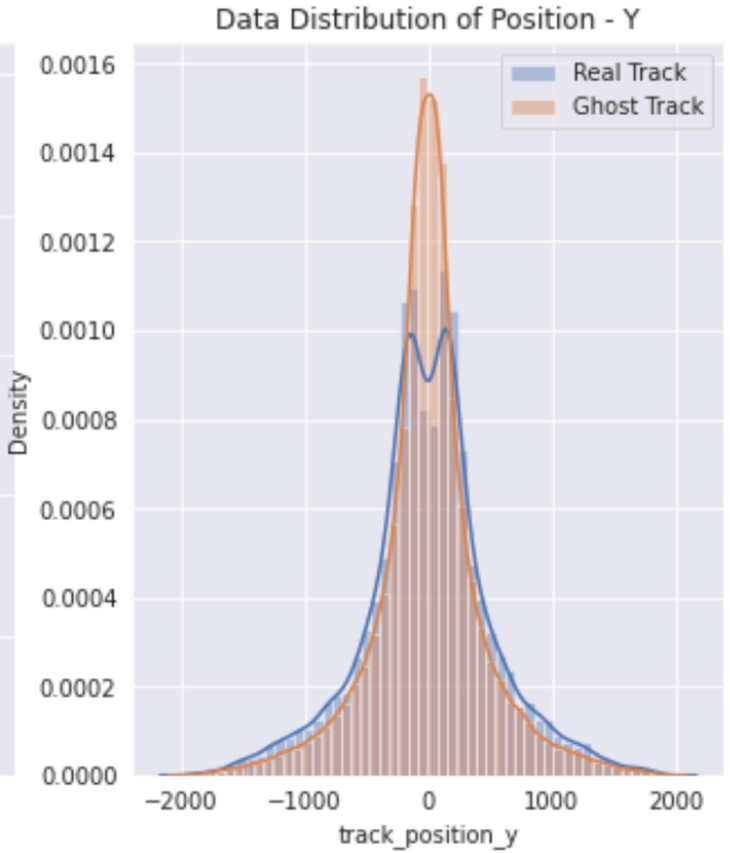
Thank you!

Sabin Hashmi
Doctoral Researcher
Department of Physics and Applied Computer Science
AGH- University of Science and Technology
✉ sabinhashmi@gmail.com

Data Distribution







LHCb Simulation Preliminary