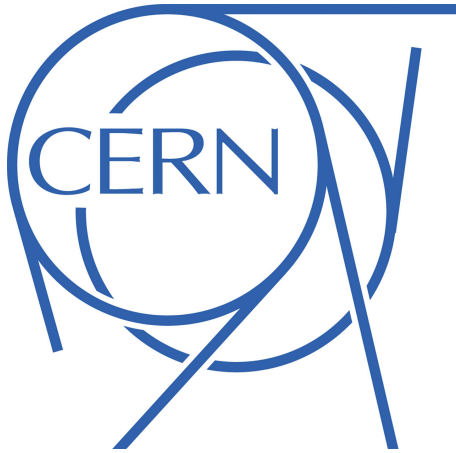




Analysis of heavy-flavor particles in ALICE with Run 3 software

Summer Student Project Report
September 6, 2021

Aman Upadhyay
Supervisors:
Dr. Vit Kucera & Dr. Nima Zardoshti

CERN Summer Student 2021
European Organization for Nuclear Research (CERN)

Acknowledgments

I am thankful to CERN for giving me this fantastic opportunity to participate in the Summer Student Program with an exciting project and people from all around the world. I display special gratitude to my supervisors, Dr. Vit Kucera and Dr. Nima Zardoshti. Their contribution to stimulating suggestions and encouragement helped me learn and explore and let me enjoy my virtual summer at CERN. I would also thank the summer student program coordinators for helping us through this program, organizing the CERN lecture series, and multiple social gatherings.

Contents

1	Introduction	1
1.1	Online–Offline Project (O^2)	1
1.2	Rectangular Cut Optimization (RCO)	3
1.2.1	Implementation	4
2	Analysis	5
2.1	D^0 decay	5
2.2	Analysis using significance	6
2.3	Analysis using ML RCO:	9
2.3.1	ROC curve for RCO	10
2.3.2	Decision Trees	11
2.3.3	PDERS	11
2.3.4	Results	12
2.3.5	BDT and PDERS configuration and plots	13
2.3.6	Comparison between methods	15
2.4	Conclusion	15
	References	17

Chapter 1

Introduction

ALICE stands for A Large Ion Collider Experiment, it is one of the four main detector experiments at CERN's Large Hadron Collider (LHC). The ALICE detector is situated at the interaction point 2 (IP2) of the LHC ring.

One of the aims of ALICE is to study the properties of Quark–Gluon Plasma (QGP). QGP is a state of matter in which quarks and gluons are in a deconfined state. The universe existed in the QGP state just after the Big Bang, and studying the interaction inside the QGP medium can offer insights into the Universe's early stages. ALICE is optimized to study the heavy-ion Pb–Pb nuclei collisions. High energy densities reached during Pb–Pb nuclei collisions create the QGP inside the LHC.

One way to probe the properties of the QGP is to use heavy-flavor particles. Hadrons that contain charm or beauty quarks or anti-quarks are known as heavy-flavored particles, and an example are D^0 mesons which contain a charm quark and up anti-quark. ALICE aims to study these heavy-flavored particles to provide essential insights into the mechanism of charm-quark or beauty-quark interactions inside the QGP. Heavy-flavor particles are created very early in the process, during the collision itself, and they are too heavy to form thermally in the QGP after the collision. They interact with the constituents of the QGP through strong interactions before hadronizing. Thus they are an excellent probe of the QGP medium.

1.1 Online–Offline Project (O^2)

The O^2 project is the ALICE data analysis model for the upcoming Run 3 and Run 4. It is expected that in Run 3, during Pb–Pb collisions, the data throughput from the detector will be around 3.5 TB/s, O^2 is designed to cope with a large amount of data by using a flat table structure and a compact data model. It will also provide a modular framework where smaller building blocks can be attached to perform tasks.

The O^2 computing model is designed to reduce the volume of read-out data. It stores the data into a collection of flat tables used by the analysis frameworks for input/output (I/O). In the

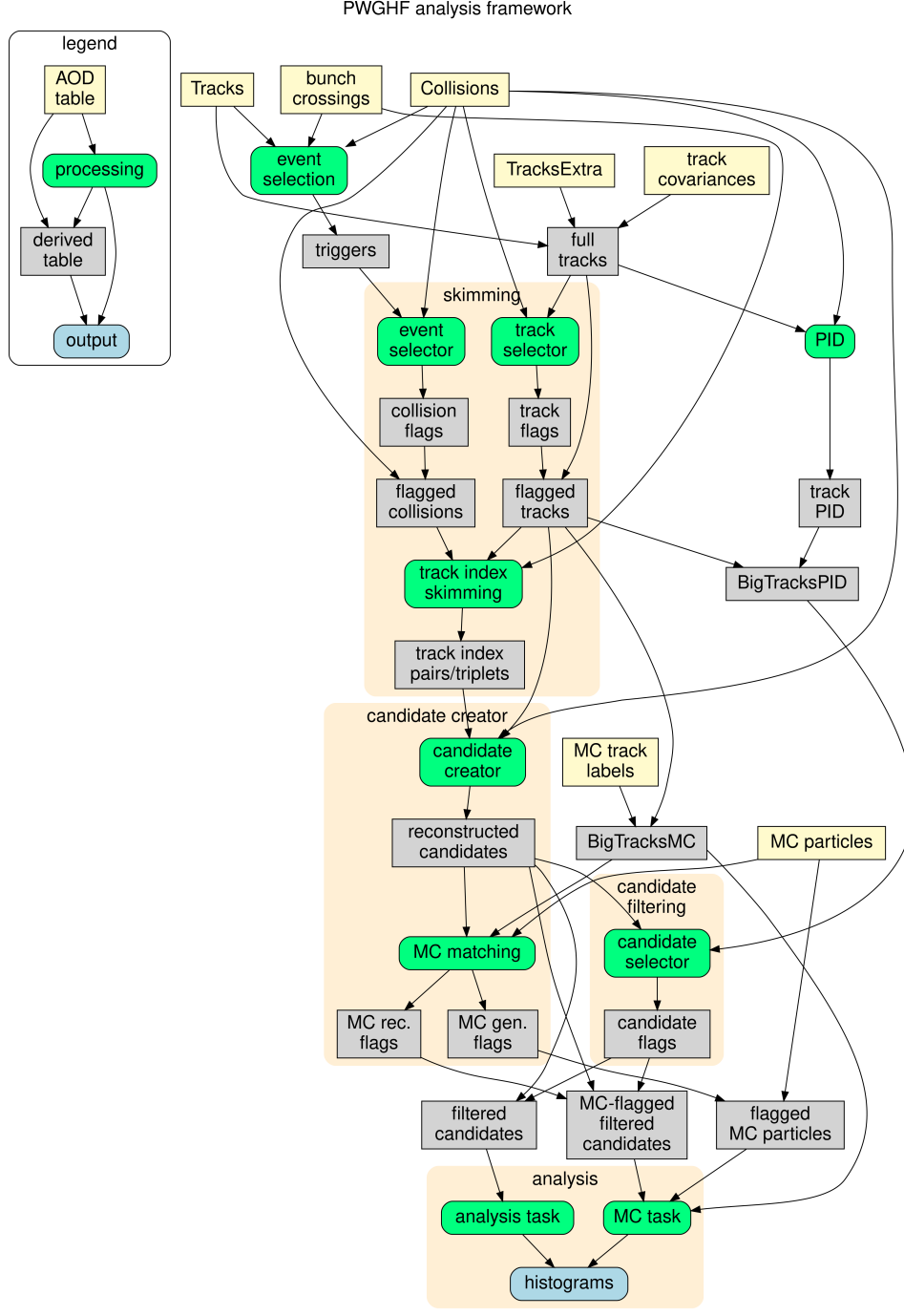
O^2 table data model, each row corresponds to an entry. For example, a row in the tracks table corresponds to a track. It can have various kinds of columns that can be used as per their requirements. Some examples are :

- **Static:** These store values like position, momentum, or flags.
- **Index:** These have references to other tables. We can use this to connect two rows in different tables.
- **Expression:** These contain calculations using values from static columns.
- **Dynamic:** These contain functions that operate on other columns (including expressions). These are calculated on the fly.

We can also manipulate these tables by joining tables to make a bigger table, grouping iteration over tracks in the same collisions, or declarative filtering based on the column content without explicit loops.

The O^2 analysis framework has a modular workflow structure defined by joining smaller workflows using the tables for I/O and JSON for configuration. The flowchart shown in Fig 1.1 is of the building blocks of the heavy-flavor analysis framework. In Pb-Pb collisions, thousands of particles are created per event. We reconstruct the heavy-flavor particles through their decays to more stable particles. However, with so many particles, we have a large number of background candidates, which are also reconstructed. The heavy-flavor workflow is designed to reconstruct these decays and suppress the background while maintaining high signal efficiency. Starting from the top of the workflow in Fig. 1.1 we use a series of filtering steps to apply selection criteria on both the daughter track properties and the decay properties to remove background candidates. First, some pre-selection cuts (invariant mass and p_T of the potential candidates) are made at the candidate creation block, which are loose cuts on the selection variables. This is done before the computationally expensive secondary-vertex reconstruction to reduce the number of reconstructed background candidates.

The candidate creator block finds and stores a table with all the potential candidates with their tracks and other candidate properties, for which not all the information needed will be stored to calculate them on the fly. The next step is the candidate filtering block. Here we make some more precise cuts. The final cuts on these potential candidates are made to suppress the background. Optimizing all the steps is essential for the analysis. My project is related to the candidate filtering block.

Figure 1.1: A flowchart of the O^2 heavy-flavor analysis framework

1.2 Rectangular Cut Optimization (RCO)

We have used Toolkit for Multivariate Analysis (TMVA) for RCO on our data. TMVA is a machine learning environment in ROOT(a C++ based framework for data analysis). The principle of RCO is to maximize background rejection (r_B) at a given signal efficiency (ϵ_S)

1.2.1 Implementation

While classifying the data into signal and background, the classification can go wrong in the following two ways (i.e. there can be two classes of errors):

1. We can include data in the signal which is actually background (false positive).
2. We can leave data out of signal which is actually signal (false negative).

We want to minimize these two types of errors during classification.

To understand background rejection (r_B) and signal efficiency (ϵ_S), we first define a few more terms.

- S = The amount of data that is classified as the signal from marked signal data.
- S_* = The amount of data classified as the signal but is from the marked background data.
- B_* = The amount of data that is classified as background but is from marked signal data.
- B = The amount of data classified as background from marked background data.

The term signal efficiency (ϵ_S) is defined as the fraction of the desired events that you actually get (i.e. it goes down as the false-negative rate goes up).

$$\epsilon_S = \frac{S}{S + B_*} \quad (1.1)$$

The false-negative ratio is given by:

$$\text{False negative ratio} = \frac{B_*}{B_* + S} \quad (1.2)$$

The term background rejection (r_B) is defined as the fraction of events that don't belong in your sample and are excluded from your sample (i.e. it goes down as the false positive rate goes up).

$$r_B = \frac{B}{B + S_*} \quad (1.3)$$

The false-positive ratio is given by:

$$\text{False positive ratio} = \frac{S_*}{S_* + B} \quad (1.4)$$

Hence, RCO maximizes the background rejection at given signal efficiency and scans over the full range of the latter quantity. This minimizes the false positive and false negative ratios.

Chapter 2

Analysis

Project Objectives:

- To reproduce a heavy-flavor Run 2 selection criteria analysis within the new Run 3 O² framework.
- Using ML techniques to optimize the selection criteria.

2.1 D⁰ decay

ALICE detects heavy-flavor particles through their daughter particles after their decays. Using the reconstructed daughter tracks, daughter particles are detected. A series of selection criteria are used to determine which daughter particles may have decayed from a heavy-flavored hadron.

Fig. 2.1 shows a schematic diagram of a D⁰ to π^+ K⁻ decay (we also measure $\overline{D}^0$). Due to the short lifetime and zero charge of the D⁰, we only detect the π^+ K⁻ daughters. The primary vertex is where the collision happened, and the charm quark was formed, it hadronises very soon to a D⁰ meson and then after traveling some distance decays at the secondary vertex to a π and a K. In this case, after we have detected the decay daughters, the secondary vertex is reconstructed by propagating the daughter tracks backward. The closest distance between the primary vertex and the daughter tracks is their respective impact parameters.

Important selection variables :

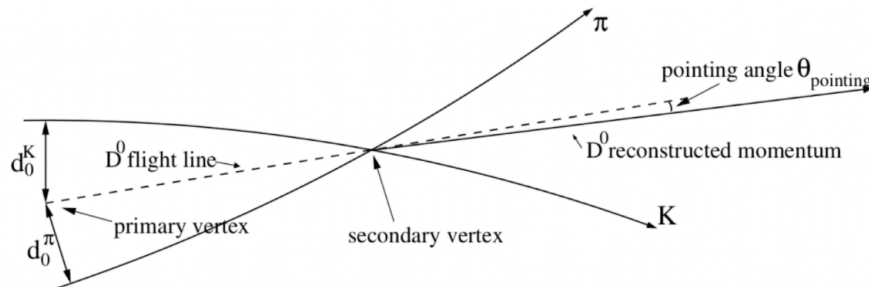


Figure 2.1: a schematic diagram of a D⁰ to π^+ K⁻ decay

We can extract multiple variables from the above configuration that will help us discriminate between signal and background by applying selection criteria to them.

The variables used for applying selection criteria:

1. p_T : Transverse Momentum of the D^0 meson.
2. **Product of Impact Parameters**: Product of impact parameters of both the daughters.
3. **CPA**: Cosine of pointing angle, the angle between D^0 momentum direction, and the line joining primary and secondary vertex.
4. **CPAXY**: Cosine of pointing angle in the X-Y plane.
5. **Decay Length**: Distance between primary and secondary vertex.
6. **Decay Length Normalised**: Decay length normalized to its uncertainty.
7. **Decay Length XY Normalized**: Decay length in X-Y plane normalized to its uncertainty.
8. **Impact Parameter 0**: Impact parameter of positive daughter.
9. **Impact Parameter 1**: Impact parameter of negative daughter.
10. **Inv. Mass**: Invariant Mass of candidate.
11. p_T **Prong 0**: Transverse Momentum of the positive daughter.
12. p_T **Prong 1**: Transverse Momentum of the negative daughter.
13. **Impact Parameter Normalized 0**: Normalized impact parameter of positive daughter to its uncertainty.
14. **Impact Parameter Normalized 1**: Normalized impact parameter of negative daughter to its uncertainty.
15. **Cos θ^*** : Cosine of the angle between the K^- and the boost direction in the D^0 rest frame.

2.2 Analysis using significance

I have converted Run 2 format to Run 3 flat table format using O² software. I did not apply any selection criteria in the candidate filtering block in the first step. I can extract histograms of selection variables per p_T bin for signal and background using the Analysis block. Using these histograms, I get the optimal selection criteria that are used in the candidate filtering block.

We use a metric known as significance to optimize selection criteria:

$$\text{Significance} = \frac{S}{\sqrt{S + S_*}} \quad (2.1)$$

The formula is such that as the signal goes up, the significance goes up, but our significance goes down as the background goes up. The idea is we want to maximize the signal with

respect to the background. Signal and background histograms for a variable represent variable values for signal and background candidates, respectively. I can calculate the significance of all possible selection criteria on our selection variables for each p_T bin using signal and background histogram. I optimized the significance as a function of selection criteria values of each variable per p_T bin.

There are four types of selection criteria I have used for various variables in our analysis. They are:

1. **Greater than x :** All candidates with variable value (x) greater than a selection criteria value are expected to be signal.
2. **Less than x :** All candidates with variable value (x) less than a certain selection value are expected to be signal.
3. **Greater than $\text{abs}(x)$:** All candidates with variable value (x), with $|x|$ greater than a certain selection value, are expected to be signal.
4. **Less than $\text{abs}(x)$:** All candidates with variable value (x), with $|x|$ less than a certain selection value, are expected to be signal.

We decided the nature of selection criteria applied to each variable based on the distribution of signal and background. CPA, CPAXY, Decay length, Decay length XY, Decay Length XY Normalized, p_T prong 0, and p_T prong 1 use greater than selection criteria, while Product of Impact parameters uses less than selection criteria. Impact Parameter Normalized 0 and Impact Parameter Normalized 1 use $\text{abs}(x)$ greater than selection criteria and $\text{Cos } \theta^*$, Inv. Mass, Impact Parameter 0 and Impact Parameter 1 use $\text{abs}(x)$ less than selection criteria. I plot significance plots for our analysis; in a significance plot, I plot signal and background histogram for a variable and a p_T bin with the significance curve, which I calculated using the eq. 2.1, significance curve is plotted in green with a significance axis on the right-hand side of the plot. Some significance plots of variables with signal and background histograms are shown in Fig 2.2, 2.3, 2.4, and 2.5.

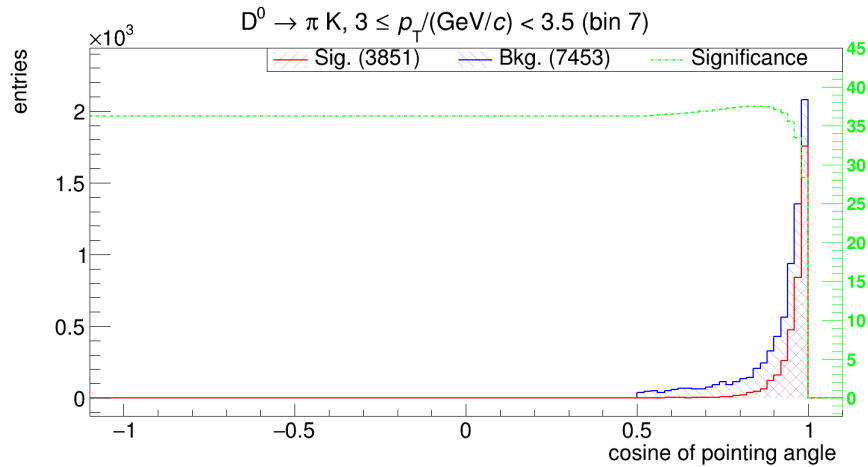
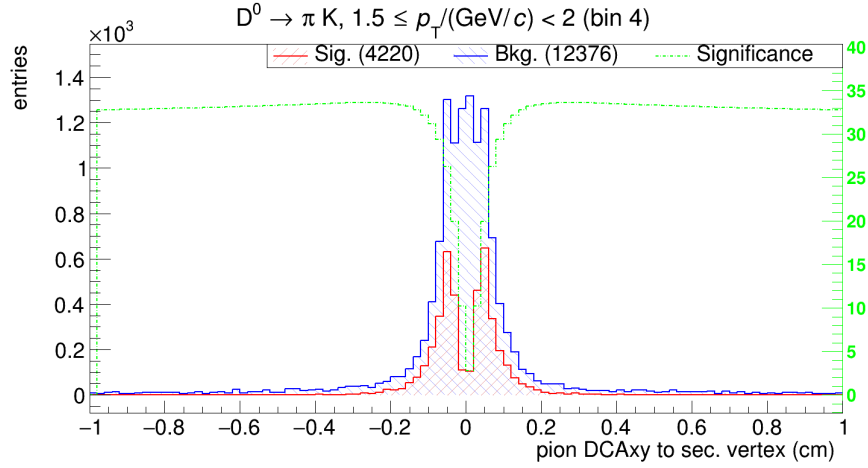
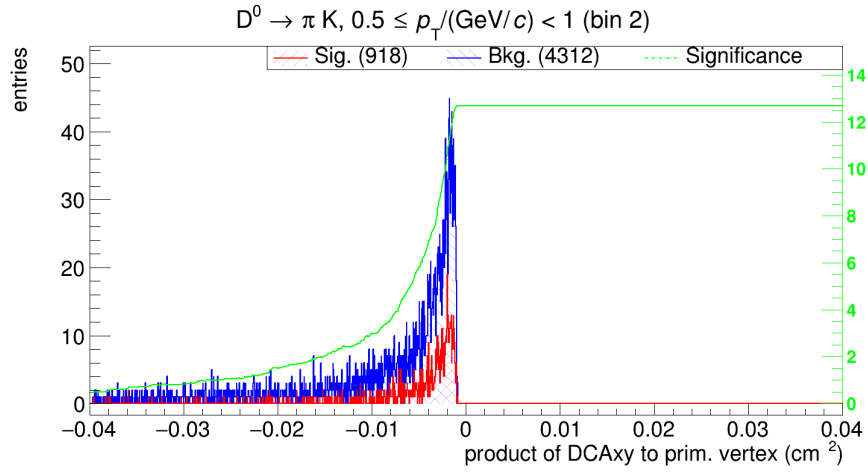
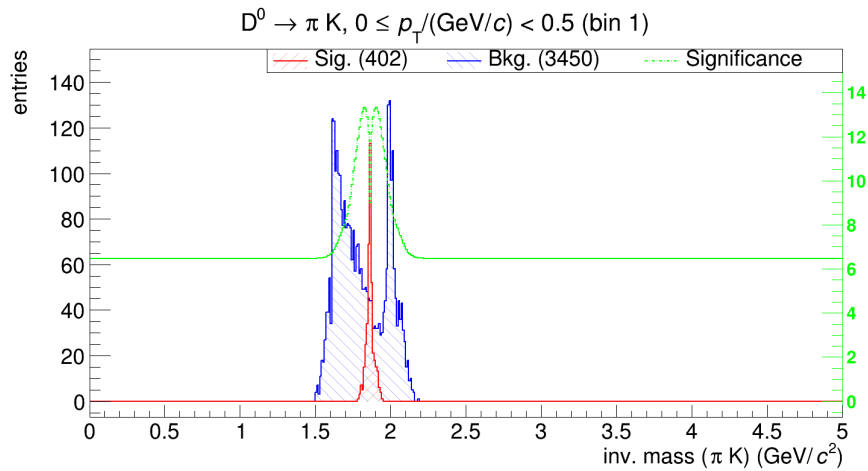


Figure 2.2: Significance plot for Cosine of pointing angle 7th p_T bin.

Figure 2.3: Significance plot for positive daughter impact parameter 4th p_T bin.Figure 2.4: Significance plot for the product of impact parameter 2nd p_T bin.Figure 2.5: Significance plot for Inv. Mass 1st p_T bin.

We can get the table of selection criteria values (table 2.2) for each variable per p_T bin using

the significance curve. We can then compare this table with the table of selection criteria values (table 2.1) from the Run 2 analysis.

Table 2.1: Run 2 selection criteria values

pT bins	mass	decay length	Cos theta *	pT prong 1	Pt prong 0	DCA prong 1	DCA prong 0	IPP	CPA	CPAXY	decay length XY Norm
1	0.4	0.035	0.8	0.5	0.5	0.1	0.1	-5.00E-05	0.8	0	0
2	0.4	0.035	0.8	0.5	0.5	0.1	0.1	-5.00E-05	0.8	0	0
3	0.4	0.03	0.8	0.4	0.4	0.1	0.1	-2.50E-04	0.8	0	0
4	0.4	0.03	0.8	0.4	0.4	0.1	0.1	-2.50E-04	0.8	0	0
5	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-2.00E-04	0.9	0	0
6	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-2.00E-04	0.9	0	0
7	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-1.20E-04	0.85	0	0
8	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-1.20E-04	0.85	0	0
9	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
10	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
11	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
12	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
13	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
14	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-8.00E-05	0.85	0	0
15	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-7.00E-05	0.85	0	0
16	0.4	0.03	0.8	0.7	0.7	0.1	0.1	-7.00E-05	0.85	0	0
17	0.4	0.03	0.9	0.7	0.7	0.1	0.1	-5.00E-05	0.85	0	0
18	0.4	0.03	0.9	0.7	0.7	0.1	0.1	-5.00E-05	0.85	0	0
19	0.4	0.03	0.9	0.7	0.7	0.1	0.1	-5.00E-05	0.85	0	0
20	0.4	0.03	1	0.7	0.7	0.1	0.1	1.00E-04	0.85	0	0
21	0.4	0.03	1	0.7	0.7	0.1	0.1	1.00E-02	0.85	0	0
22	0.4	0.03	1	0.7	0.7	0.1	0.1	1.00E-02	0.85	0	0
23	0.4	0.03	1	0.7	0.7	0.1	0.1	1.00E-02	0.85	0	0
24	0.4	0.03	1	0.7	0.7	0.1	0.1	1.00E-02	0.85	0	0
25	0.4	0.03	1	0.6	0.6	0.1	0.1	1.00E-02	0.8	0	0

Table 2.2: Run 2 selection criteria values using O²

pT bins	mass	decay length	Cos θ *	pT prong 1	Pt prong 0	DCA prong 1	DCA prong 0	IPP	CPA	CPAXY	decay length XY Norm
1	0.078	0.035	0.55	0.5	0.5	0.17	0.17	-0.00094	0.53	-0.99	1.165
2	0.058	0.035	1.03	0.5	0.5	0.21	0.21	-0.0009	0.53	-0.99	1.175
3	0.098	0.035	0.97	0.5	0.5	0.21	0.21	-0.00094	0.69	0.65	1.275
4	0.058	0.045	0.87	0.5	0.5	0.29	0.29	-0.001	0.67	0.75	1.325
5	0.038	0.035	1.01	0.5	0.5	0.25	0.25	-0.00014	0.79	0.87	1.425
6	0.038	0.035	1.01	0.5	0.5	0.21	0.21	-6.00E-05	0.85	0.93	1.465
7	0.038	0.035	1.01	0.5	0.5	0.21	0.21	-2.00E-05	0.83	0.93	1.465
8	0.058	0.015	1.01	0.5	0.5	0.27	0.27	-2.00E-05	0.87	0.93	1.465
9	0.078	0.005	0.91	0.5	0.5	0.21	0.21	-2.00E-05	0.87	0.95	1.045
10	0.118	0.015	0.89	0.5	0.5	0.23	0.23	-2.00E-05	0.91	0.95	1.105
11	0.138	0.005	0.91	0.5	0.5	0.27	0.27	-2.00E-05	0.87	0.95	0.735
12	0.198	0.005	0.93	1.5	1.5	0.33	0.33	-2.00E-05	0.83	0.95	0.145
13	0.218	0.005	0.93	1.5	1.5	0.25	0.25	-2.00E-05	0.75	0.95	0.065
14	0.258	0.005	0.95	0.5	0.5	0.27	0.27	-2.00E-05	0.75	0.95	0.165
15	0.258	0.005	0.77	1.5	1.5	0.27	0.27	-2.00E-05	0.63	0.97	0.495
16	0.278	0.005	0.95	0.5	0.5	0.25	0.25	-2.00E-05	0.75	0.97	0.095
17	0.318	0.005	0.97	1.5	1.5	0.23	0.23	-2.00E-05	0.89	0.95	0.145
18	0.318	0.005	1.07	0.5	0.5	0.33	0.33	-2.00E-05	0.93	0.97	0.725
19	0.298	0.005	0.97	1.5	1.5	0.41	0.41	-2.00E-05	0.77	0.99	0.935
20	0.358	0.025	1.09	0.5	0.5	0.31	0.31	-2.00E-05	0.97	0.97	1.055
21	1.018	0.035	1.03	0.5	0.5	0.31	0.31	-2.00E-05	0.99	0.99	2
22	0.358	0.035	1.09	0.5	0.5	0.33	0.33	-2.00E-05	0.97	0.99	1.855
23	0.278	0.035	0.99	3.5	3.5	0.21	0.21	-2.00E-05	0.99	0.83	2
24	1.578	0.155	1.03	2.5	2.5	0.11	0.11	-2.00E-05	0.99	0.99	2
25	3.718	2	1.01	51.5	51.5	0.17	0.17	-0.0023	0.99	0.99	2

2.3 Analysis using ML RCO:

I will be switching to use Multivariate classification methods based on machine learning techniques to discriminate between signal and background in our analysis. Our analysis depends on multiple variables. Hence using each selection criteria independently on variables is not ideal, and we need multivariate analysis to get optimal selection criteria for our analysis. We use Receiver Operating Characteristic (ROC) curve to judge the performance of our models.

Receiver Operating Characteristic (ROC) curve: Plot of background rejection Vs. Signal efficiency. For good selection criteria, the area under the curve should be as close to one as

possible. As for the best curve, we want our signal efficiency and background rejection to be 1. This curve is called the receiver operating characteristic curve (ROC curve), and hence the area under the curve is known as the ROC-integ value.

Different ML models perform differently depending on the variable distributions and type of the problem. RCO is used to classify signal and background using rectangular selection criteria only. Other ML models like BDT can apply non-linear selection criteria, which may be useful in some problems. I have used different ML models to solve our problem, and Fig. 2.6 shows the ROC plot for these events.

2.3.1 ROC curve for RCO

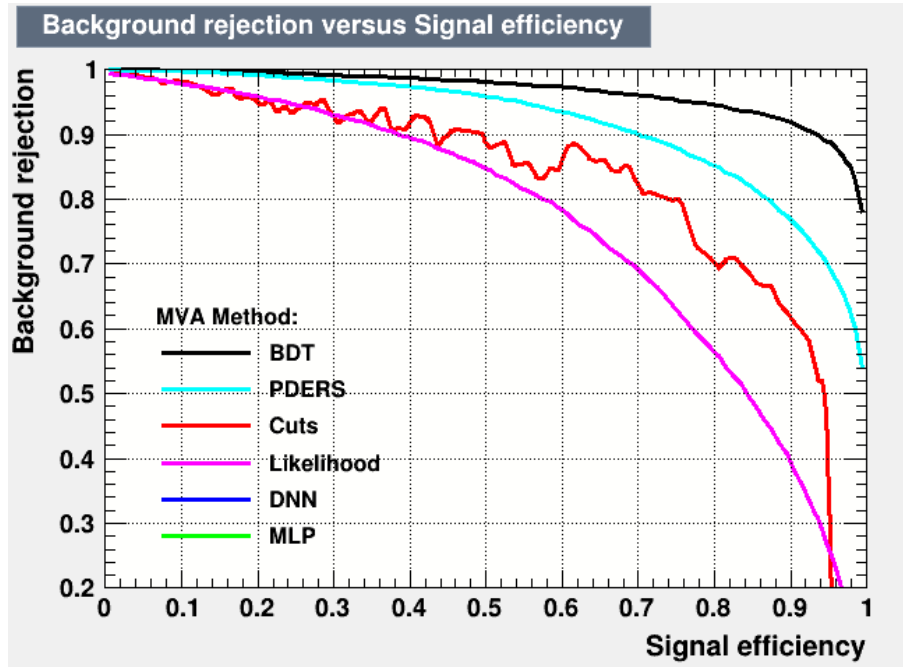


Figure 2.6: ROC plot for our analysis.

The ROC-Integ values for those techniques calculated from Fig. 2.6 are tabulated in table 2.3.

Table 2.3: ROC-integ values

Sl.No	Classification Method	ROC-Integ Value
1	BDT	0.966
2	PDERS	0.916
3	RCO	0.819
4	Likelihood	0.760
5	DNN	0.513
6	MLP	0.498

We note that BDT and PDERS methods work better than the others and even better than RCO (cuts). A brief description of BDT and PDERS is given below.

2.3.2 Decision Trees

A decision tree is defined as a tree whose internal nodes are tests (on input patterns, in our case, potential D^0 candidates) and whose leaf nodes are categories (of patterns). We show an example in Fig. 2.7. A decision tree assigns a class number (or output) to an input pattern by filtering the pattern down through the tests in the tree. Each test has mutually exclusive and exhaustive outcomes [1].

In a decision tree, repeated left/right (yes/no) decisions are taken on one single variable at a time until a stop criterion is fulfilled. The phase space is split this way into many regions that are classified as signal or background, depending on the majority of training events that end up in the final leaf node [2] [3].

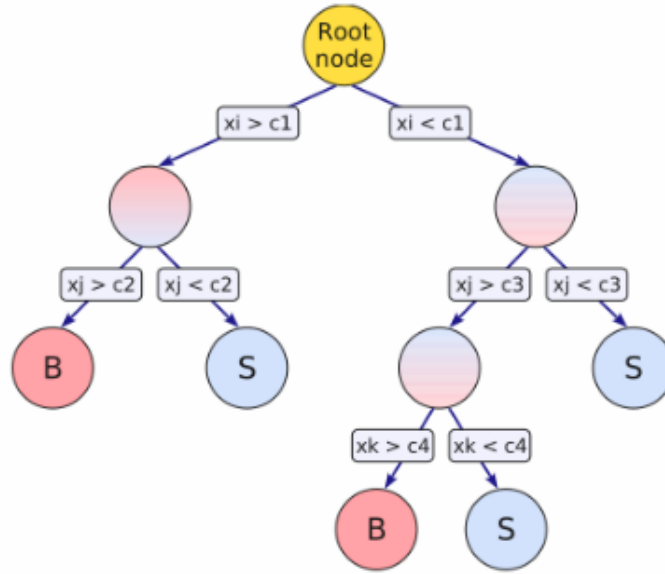


Figure 2.7: A decision tree [2].

2.3.3 PDERS

In this method, we discriminate a test event by counting the number of signal and background training events in the vicinity of the testing event. The idea is to choose an area where the test point will exist, which we will call “the vicinity”. Then we discriminate the test point depending on the number of training samples present in the vicinity.

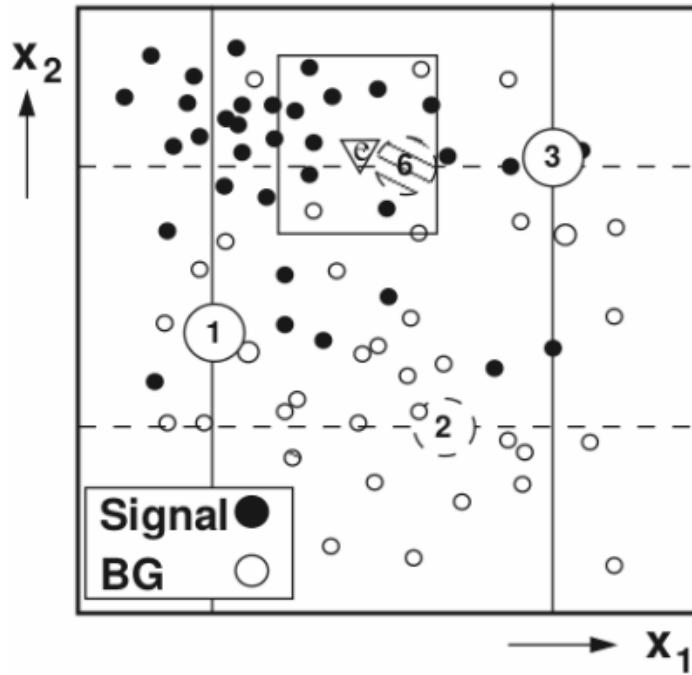


Figure 2.8: PDE-RS Example [4].

Fig. 2.8 signal and background events in the $x_1 - x_2$ plane. The open triangle depicts the event to be classified. The box around the triangle shows the region where events are considered to calculate the discriminant [4].

To implement nearest neighbor estimation for PDERS, we need to fix two parameters first:

1. Volume of the vicinity.
2. A function, which we will use to determine the weight of the training event according to their distance from the test event, is called a kernel function.

2.3.4 Results

Selection criteria values obtained using RCO for signal efficiency (0.9) are tabulated in table 2.4. Corresponding background efficiency is (0.38).

Table 2.4: ROC selection criteria values

Variable				
-0.401	<	p_T	$\leq$	45.835
-0.009	<	Product of impact parameter	$\leq$	0.0006
0.819	<	Cosine of Pointing angle	$\leq$	1.020
0.392	<	Cosine of Pointing angle XY	$\leq$	1.337
-0.0562	<	Decay length	$\leq$	7.606
-4.67	<	Decay length Norm	$\leq$	505.5
1.425	<	Decay length XY Norm	$\leq$	265.45
-0.1	<	Impact Parameter Prong 0	$\leq$	0.0991
-0.094	<	Impact Parameter Prong 1	$\leq$	0.108
1.736	<	Inv. Mass	$\leq$	2.0595
0.571	<	p_T Prong 0	$\leq$	39.222
0.530	<	p_T Prong 1	$\leq$	36.648
-408.6	<	Impact Parameter Norm Prong 0	$\leq$	60.867
-69.35	<	Impact Parameter Norm Prong 0	$\leq$	124.112
-1.012	<	Cos theta *	$\leq$	0.849

2.3.5 BDT and PDERS configuration and plots

For BDT, 1000 trees were trained of maximum depth 3 using AdaBoost [5] algorithm to decrease the variance of the model. The separation criteria used was Gini index. Response plot is shown in Fig. 2.9 and significance plot in Fig. 2.10.

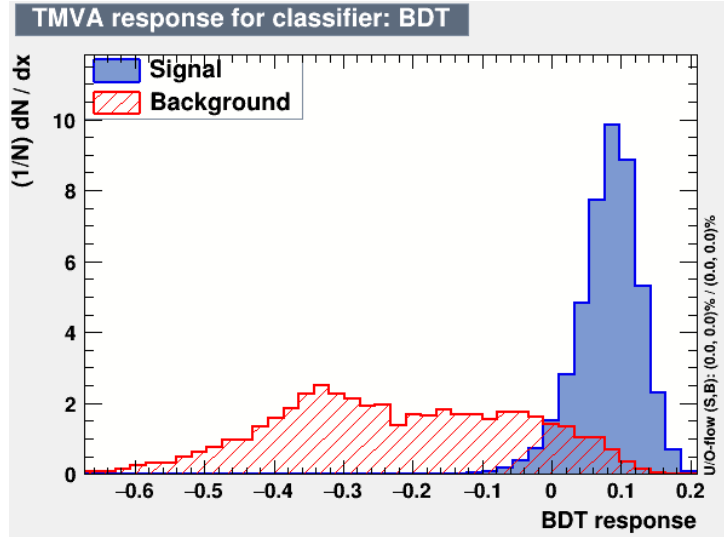


Figure 2.9: BDT response plot.

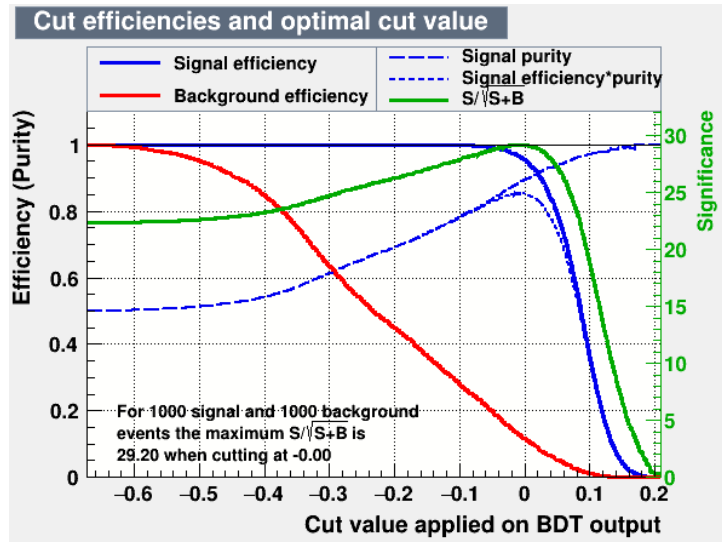


Figure 2.10: BDT significance plot.

PDERS algorithm was used with Gaussian kernel estimator with adaptive volume range. Response plot is shown in Fig. 2.11 and significance plot in Fig. 2.12.

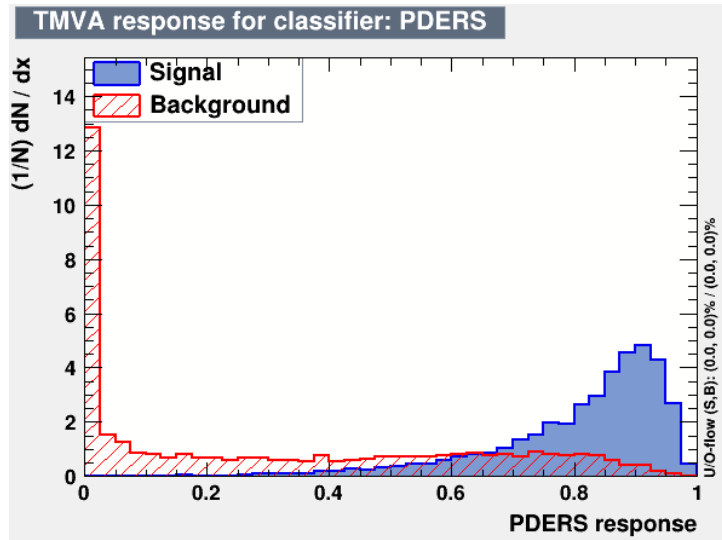


Figure 2.11: PDERS response plot.

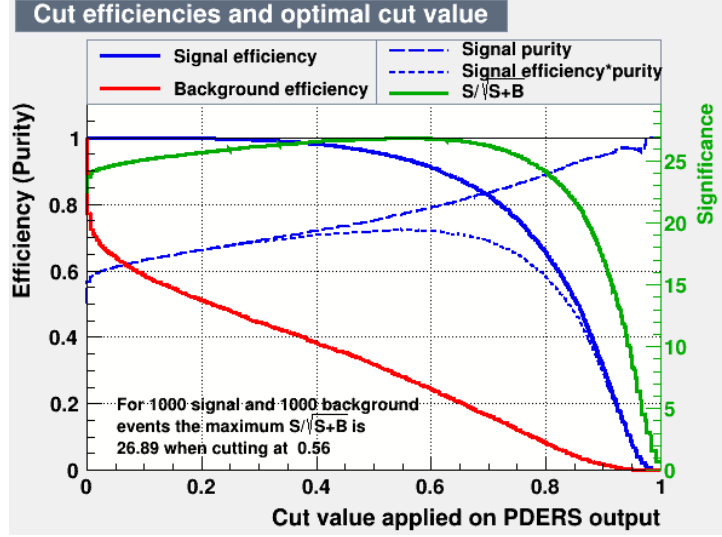


Figure 2.12: PDERS significance plot.

2.3.6 Comparison between methods

I have tried to distinguish between signal and background using four methods. They are:

1. Non-ML (Optimizing significance for each variable independently for each p_T bin.
2. Rectangular cut optimization (RCO)
3. Boosted decision trees (BDT)
4. Projective likelihood estimator - range search (PDERS)

I calculated the significance and signal efficiency for each method using 1000 signal and background events. The results are tabulated in table 2.5. In the case of RCO requested signal efficiency was (0.95).

Table 2.5: ROC selection criteria values

Method	Significance	Signal efficiency
Non-ML	22.24	0.94
RCO	24.61	0.961
BDT	29.91	0.95
PDERS	26.96	0.936

2.4 Conclusion

In the project, I reproduced the heavy-flavor Run 2 selection criteria analysis using the new Run 3 O^2 framework. I used the significance curve and optimized it as a function of selection criteria value for individual variables to get optimal selection criteria values. The selection criteria values found using O^2 were comparable to the Run 2 selection criteria values as shown in table 2.1 and 2.2. I also performed a multivariate analysis of our problem. I used RCO, which

gave us a ROC-integ value of 0.819, which is very good, especially considering RCO can only distinguish data that are not clustered and not overlapping and was trained relatively faster than other models. I found BDT and PDERS methods for classifying signal and background to be effective for our analysis. They performed better than ROC with RCO-integ values of 0.996 and 0.916, respectively. This insight can be investigated in future projects.

References

- [1] N.J Nilsson. *Introduction to machine learning*. 2005.
- [2] *TMVA users guide*. CERN. 2017. URL: <https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf> (visited on 03/03/2020).
- [3] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN: 9780387310732.
- [4] T Carli and B Koblitz. “A Multi-variate Discrimination Technique Based on Range-Searching”. In: *Nuclear Instruments and Methods in Physics* 501(2-3):576-588 (2002), pp. 251–257.
- [5] Robert Tibshirani Jerome Friedman Trevor Hastie. *The Elements of Statistical Learning*. 2001.