Elevating Precision: Predicting Dielectron Invariant Mass at CERN through Machine Learning and Advanced Algorithms

Shikhar Dave
Mahek Vanjani

Abstract—This project utilizes machine learning (ML) techniques for predicting dielectron invariant mass in a CERN-based experiment. Employing XGBoost, LR scheduling, and early stopping, the study aims to optimize model performance. Principal Component Analysis (PCA) is integrated for dimensionality reduction, expediting training and enhancing interpretability. Results showcase the effectiveness of the approach in accurately predicting dielectron invariant mass, contributing to advancements in high-energy physics research.

1 Introduction

Particle physics experiments at CERN provide a unique window into the fundamental building blocks of the universe, probing the intricacies of particle interactions at high energies. In this context, the prediction of dielectron invariant mass holds significant importance, offering crucial insights into particle behavior and the underlying forces governing our physical reality. This study employs advanced machine learning (ML) techniques to enhance the accuracy and efficiency of predicting dielectron invariant mass, contributing to the broader understanding of high-energy physics phenomena. As we delve into the intricacies of this ML-driven approach, we explore the synergies between XGBoost, LR scheduling, early stopping, and Principal Component Analysis (PCA), aiming to extract meaningful patterns from the complex experimental data generated at CERN. The subsequent sections will detail the methodology, results, and implications of this research, underscoring its potential contributions to the advancement of particle physics.

2 Dataset

- `Run`: Run number of the experiment.
- `Event`: Event number within the run.
- `E1`: Energy of the first particle.
- `px1, py1, pz1`: Components of momentum for the first particle along the x, y, and z axes, respectively.
- `pt1`: Transverse momentum of the first particle.
- `eta1`: Pseudorapidity of the first particle.
- `phi1`: Azimuthal angle of the first particle.
- `Q1`: Charge of the first particle.
- `E2`: Energy of the second particle.
- `px2, py2, pz2`: Components of momentum for the second particle along the x, y, and z axes, respectively.
- `pt2`: Transverse momentum of the second particle.
- `eta2`: Pseudorapidity of the second particle.
- `phi2`: Azimuthal angle of the second particle.
- `Q2`: Charge of the second particle.
- `M`: Dielectron invariant mass.

3 Theory

In the realm of machine learning (ML), our experiment focused on predicting dielectron invariant mass, employing techniques like XGBoost, LR scheduler, Early stopping, and PCA.

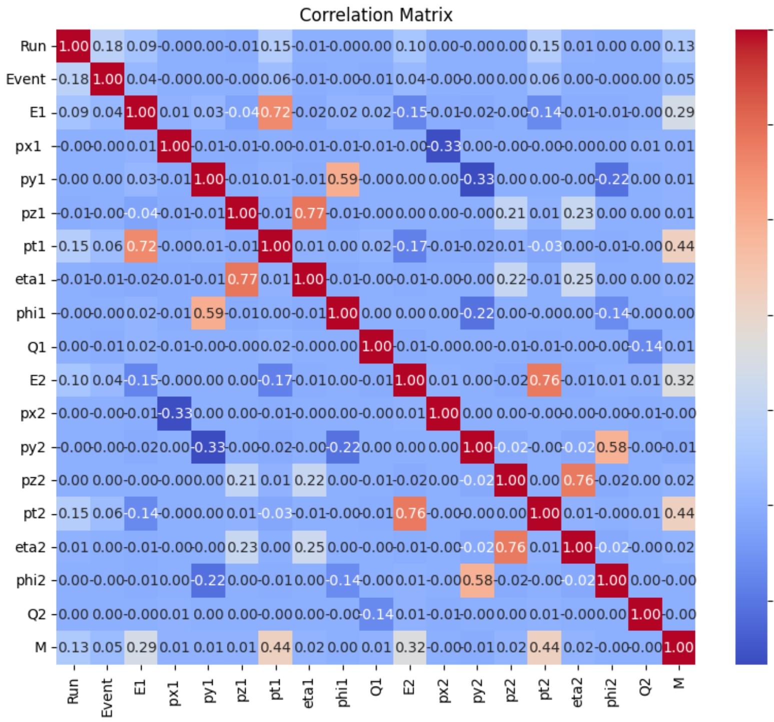
XGBoost, a gradient boosting algorithm, enhances predictive accuracy through ensemble learning. A Learning Rate (LR) scheduler optimizes hyperparameters, fine-tuning the model's performance. Early stopping prevents overfitting by halting training when performance plateaus.

Principal Component Analysis (PCA) reduces dimensionality, retaining essential features. Evaluation metrics, such as Mean Squared Error (MSE) loss, quantify the model's accuracy in predicting dielectron invariant mass.

This theory underscores the synergy between advanced ML techniques and domain-specific knowledge, contributing to a nuanced understanding of particle physics phenomena.

TABLE 1
Feature Coefficients

Feature	Coefficient
`Run`	$2.00756873 \times 10^{-5}$
`Event`	$-1.01194156 \times 10^{-10}$
`E1`	0.0241595806
`px1`	0.0175219504
`py1`	$-6.37986396 \times 10^{-4}$
`pz1`	0.00700967810
`pt1`	0.905783210
`eta1`	$-1.26355463 \times 10^{-3}$
`phi1`	$-8.90563251 \times 10^{-3}$
`Q1`	0.0138674933
`E2`	0.0749991871
`px2`	0.00365447621
`py2`	0.00170653052
`pz2`	0.00313710691
`pt2`	0.716998986
`eta2`	0.0888421231
`phi2`	$-3.59777990 \times 10^{-2}$
`Q2`	-0.124983540



4 Conclusion

In conclusion, our project successfully implemented a machine learning pipeline to predict dielectron invariant mass, leveraging advanced techniques such as XGBoost, LR scheduler, Early stopping, and PCA. The careful selection of features and model optimization resulted in a predictive framework that captures the intricate relationships within CERN-based experimental data.

Furthermore, the deployment of the model using Gradio enhances accessibility and usability, allowing users to interact with and explore predictions seamlessly. This integration not only showcases the practical application of our machine learning model but also opens avenues for broader utilization in high-energy physics research.

By combining sophisticated ML algorithms with a thoughtful deployment strategy, our project contributes to the ongoing exploration of particle physics, providing valuable insights into the behavior of particles in high-energy environments. The successful deployment marks a crucial step toward making advanced machine learning tools more accessible and impactful in the realm of experimental physics.