

**Project Report – CERN electron collision prediction**

# EP21BTECH110026

# Sangeet

# MA21BTECH11005

# Devashish Chaudhari

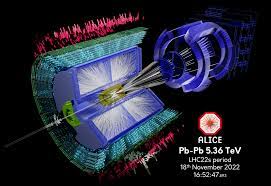
Friday, 26th April, 2024

Abstract –

We will use data by CERN (European Organization for Nuclear Research) to predict collision between electron and positron by means of ML modeling and ANNs (Artificial Neural Networks).

Introduction -

Now a days we are using Machine learning in various fields of physics to analyze and predicts what is going to happen next. Same kind of prediction and analysis we are going to do in this project which is Electron collision prediction. There are different type of training methods for ML, Supervised learning, Unsupervised Learning, Reinforcement learning.



What we are going to do?

1. Cleaning Dataset for regression and
2. Statistics and Analysis of the data
3. ML and ANN

About the Dataset?

The Data what we are using in this project contains 10^8 direction events in the invariant mass range 2-110 GeV for the use in education and project.

The collision energy is transformed into matter. There are few process involved in ML to train computational formula to do a certain task effectively.

Some information about Electron-Positron annihilation:

It occurs when an electron and a positron collide. These all cases satisfy the conservation laws (i.e. Electric charge (Net charge is zero), Linear Momentum, Angular Momentum, Total energy, Matter).

Lower energy: This annihilation occurs when electron and positron collide at lower energies, and this results into energetic photon.

There are the few possibilities of final state of annihilation which includes the creation of two or more gamma photons. Conservation of energy and momentum restricts the creation of only single photon. Commonly it is observed that two gamma photons are formed, each having energy equals to rest energy of electron or proton.

Frame of Reference: The system has no net linear Momentum before the annihilation but after collision, gamma photon emitted in the opposite direction, similar for the others create which ensures the charge parity. There is also a possibility of creation of large number of photon, but the probability reduces with each additional gamma photon. There is also one possibility which produces neutrino-antinutrino pair and there is no conservation law that forbids it, but its probability is way less than annihilation into the photon.

High Energy:

For the high energies, there may result into B mesons and W and Z bosons. In this case electron or positron or either one of them should have high kinetic energy. The excess energy is converted into relative velocities.

The energy near the mass of carriers of weak forces, the bosons, strength of weak forces become comparable to EM forces, it leads to much easy production of particles like neutrino which interact weakly with different matter. In the experiments performed the heaviest particle which is produced by this annihilation in particle accelerators are W+ and W- pairs. The heaviest single charges particle is Z-Boson.

This phenomenon serves as the basis of Positron emission tomography(PET) and positron annihilation spectroscopy(PAS). This method is used to nuclear transition, measure Fermi surface and band structure in metals by Angular correlation of electron positron annihilation radiation, study Crystal defects and a lot more.

Implementation –

We have 9 features in our dataset.

1) ***Run*** : The run number of the event.

2) ***Event*** : The event number.

3) ***E1, E2*** : The total energy of the electron (GeV) for electrons 1 and 2.

4) ***px1, py1, pz1, px2, py2, pz2***: The components of the momentum of the electron 1 and 2 (GeV).

5) ***pt1, pt2***: The transverse momentum of the electron 1 and 2 (GeV).

6) ***eta1, eta2***: The pseudorapidity of the electron 1 and 2.

7) ***phi1, phi2***: The phi angle of the electron 1 and 2 (rad).

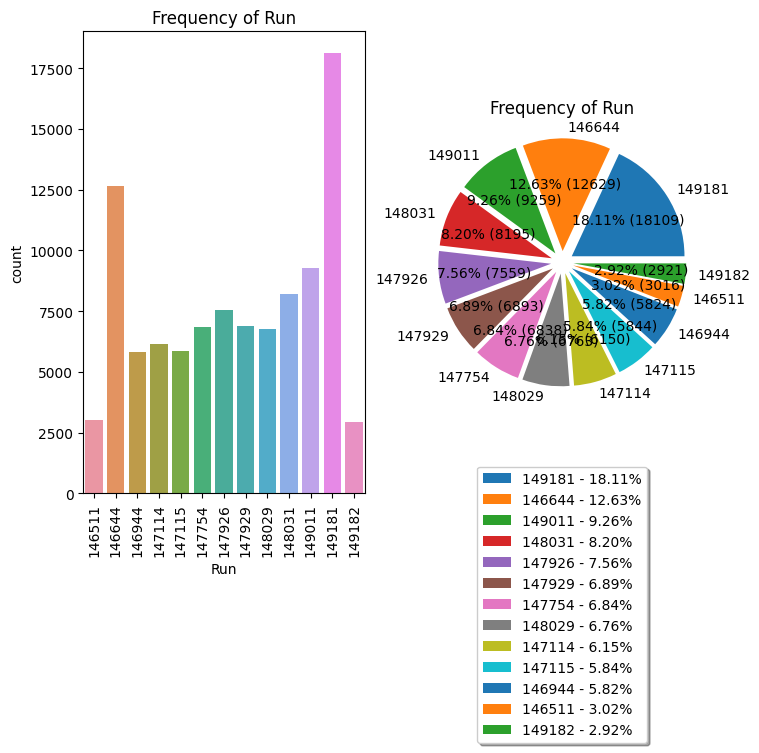
8) ***Q1, Q2***: The charge of the electron 1 and 2.

9) ***M***: The invariant mass of two electrons (GeV).

EDA (elementary data analysis) –

First, we make frequency plots of three columns, ‘Run’, ‘Q1’ and ‘Q2’.

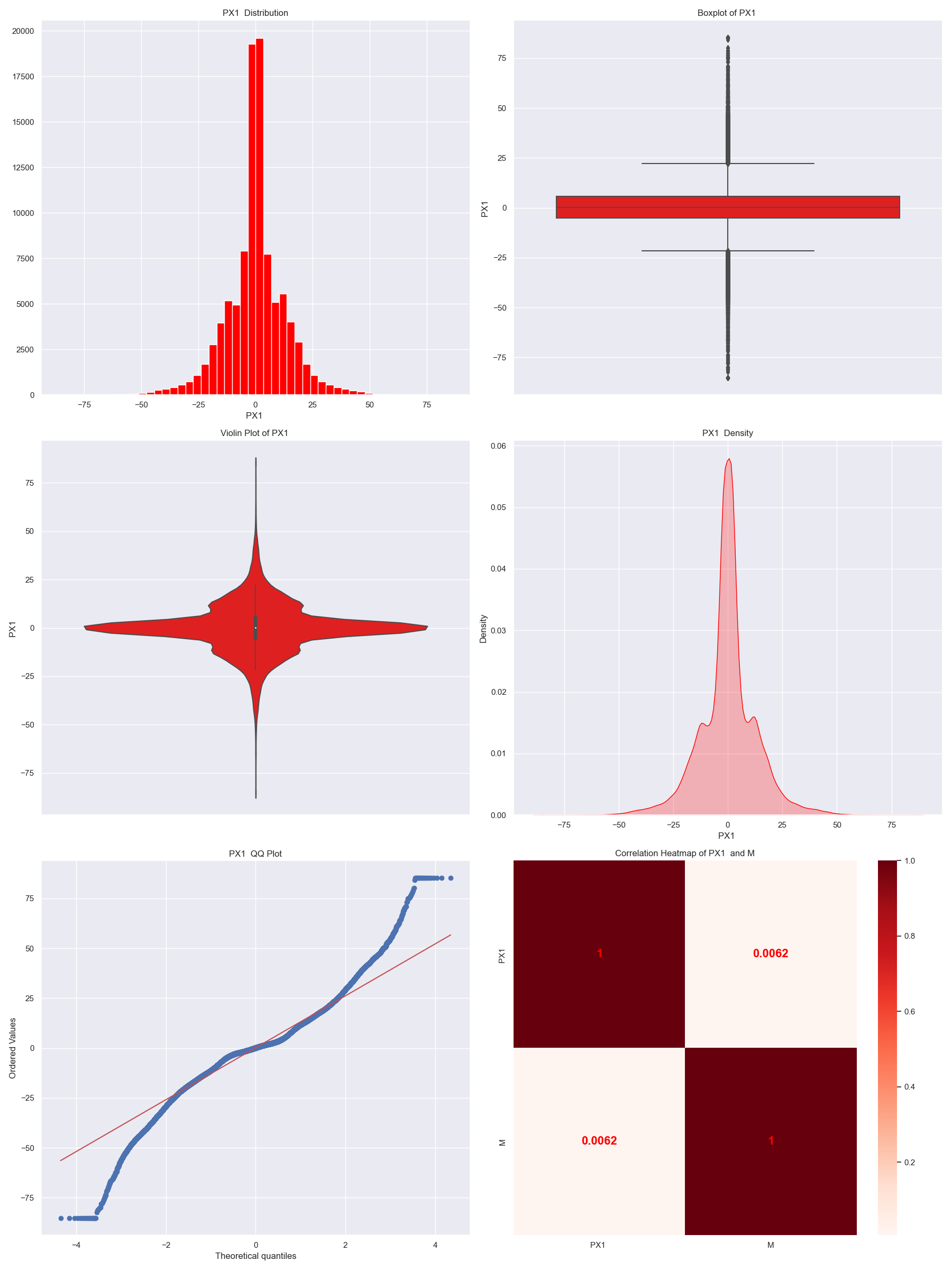
e.g. - For ‘Run’ feature



Then, for each features in ['Event','E1','px1 ','py1','pz1','pt1','eta1','phi1','E2','px2','py2','pz2','pt2','eta2','phi2','M'], we output mean, standard deviation, quantiles, skewness, and kurtosis. We also plot 6 graphs, which are distribution (a histogram), boxplot, violin plot, density plot, QQ plot and Correlation Heatmap for each of them.

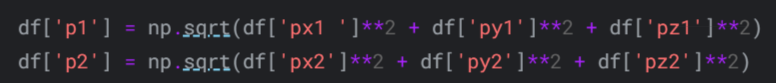
e.g. –

for ‘px1’ the plots are



Feature extraction & Correlation –

We define new features such as



Scaling –

This makes it easier for the algorithm to converge faster and leads to more accurate predictions. Also, it prevents the different scales of features from influencing the algorithm.

We do this by using the below function:

A screen shot of a computer code

Description automatically generated

ML modeling –

We use different models such as

A screen shot of a computer program

Description automatically generated

Also, we find out the errors in prediction for each of these models, which come out to be

A screenshot of a computer screen

Description automatically generated

From this we conclude that CatBoost performs the best.

ANN (Artificial Neural Network) –

The training and validation loss comes out to be,

A graph showing the loss of a training

Description automatically generated

We also plot graph between Actual and Predicted values, which comes out to be

A graph showing a line of blue dots

Description automatically generated with medium confidence

Source code –

Github - https://github.com/DevashishChaudhari/DSA-final-project

Conclusion –

In this project we demonstrated the ML and data analysis applications for predicting the electron collision by using the data set. We have used different ML models in this project such as Linear Regression, Decision Tree, XG boost, CAT boost, etc. Out of them CAT boost gives the least error. Finally, by using ANNs, we get practical vs predicted graph which shows that predicted values are nearly matching with practical values, which implies that this model can be used in various field of collision system and energies.

This project shows the potential of ML, ANNs in predicting the results without actually performing the experiment with a good accuracy. This can be further improved by training with more data set.