Machine Learning for High Energy Physics in Search for Higgs Boson Particle

Mr. Fakruddin Mohammed
fam281@g.harvard.edu

Motivation

I was in Zurich working for UBS during the period of 2018-2020. When my family visited me in Zurich during the summer of July/Aug 2019, I took them for 1 day trip to Geneva to visit the CERN laboratory, the birth place of "God particle". The trip was amazing, it takes one through the journey of atomic particle to subatomic. There were lots of short video clips, practical demonstrations and exhibitions etc.
What caught my eyes was a poster with title "Machine Learning for High Energy Physics". The aim of the poster was to explore the new ways to improve the cross-fertilization of the two fields: astrophysics community (dark matter and galaxy zoo challenges) and neurobiology (connectomics and decoding the human brain).Therefore, I decided to explore what is particle physics and how can the machine learning and deep learning can help solve the particle physics problems. The summary of this short article was the result of my visit to CERN laboratory.

Problem Statement

After the visit to the CERN laboratory, I started exploring the fundamentals of particle physics, its fundamental equation of particles etc, even though I didn't understand much as it has been a decade since I looked into physics closely but it was very interesting research. As part of this exploration I came across a small simulated dataset at Kaggle website (2014). The goal of the Higgs Boson Machine Learning Challenge is to explore the potential of advanced machine learning methods to improve the discovery significance of the experiment. The good thing about this challenge is no knowledge of particle physics is required. Therefore, I decided to use this dataset to explore how the Machine learning algorithms can be used for Higgs Boson research.

Technology

The following software stack is used to build the device.

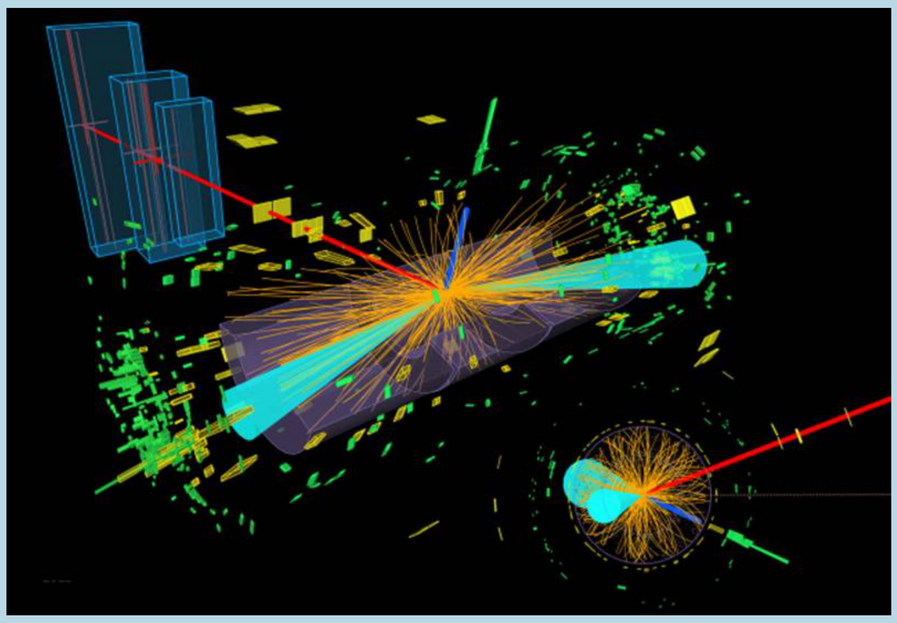
- Python 3.6 (for Exploratory Analysis)
- R for Statistics
- Matplotlib/Plotly
- SciKit (Machine Learning)

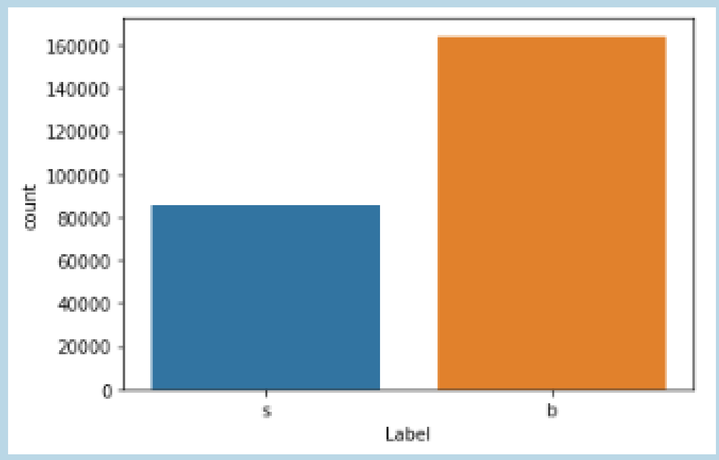
Algorithms

The following algorithms are explored for the classification of Signal from the Background Noise.

- Logistic Regression
- Decision Trees
- KNN (K Nearest Neighbourhoods)
- LDA (Latent Discriminant Analysis)
- Gaussian Naive Bayes
- SVM (Support Vector Machine)
- Random Forest Classifier

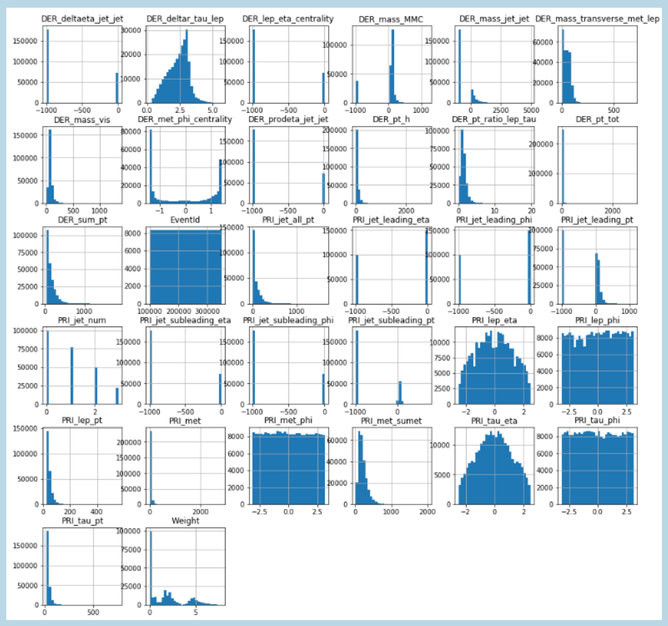
Architecture

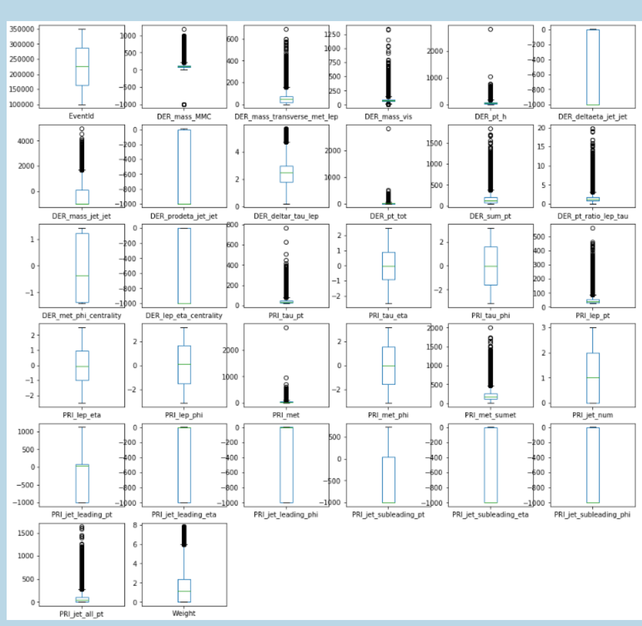




Data Flow, Current Status and Next Steps

End to End Data Flow The training data downloads from [2] has set of 250000 events. Each event has an ID column, 30 feature columns, a weight column and a label column. The distribution of output label columns are shown above.....
The input data distributions, box plots are shown below and the model accuracies from number of machine learning models are tabulated.....





Model	Train Acc.	Test Acc.
Logistic	0.9615	0.9613
Decision Trees	1.0000	1.0000
KNN	0.9321	0.9026
LDA	0.8963	0.8970
Gaussian NB	0.9754	0.9765
SVM	0.9579	0.9578
Random Forest Classifier	0.9615	0.9613

Current Status Using the data set containing a mixture of simulated signal and background events, built from simulated events provided by the ATLAS collaboration at CERN, number of machine learning algorithms are implemented to classify the signal from the background noise.

Next Steps Enhance the model to further improve the model accuracy and implement the optimization algorithm specified in the Machine Learning challenge paper.

References

[1] The Higgs Machine Learning Challenge: `https://higgsml.lal.in2p3.fr/`, this website provides full description of the challenge along with technical details. (May-Sep 2014).

[2] Abha Eli Phoboo Sylvie Brunet, ATLAS sees Higgs boson decay to fermions, The ATLAS experiment at CERN has found evidence for the Higgs boson decaying to two tau particles: `https://home.cern/news/news/experiments/atlas-sees-higgs-boson-decay-fermions/`, this website provides full description of the challenge along with technical details. (20th Nov 2013).

[3] Higgs Boson Machine Learning Challenge (Use the ATLAS experiment to identify the Higgs boson) `https://www.kaggle.com/c/higgs-boson/data/`, this website hosts the data for the machine learning challenge (2014).

YouTube URLs Github Repository

- Github Repository: `https://github.com/mohammed-fakruddin/MachineLearningForParticlePhysicsHigBoson`