

Problem description:

The LHC collides bunches of protons every 50 nanoseconds within each of its four experiments, each crossing producing a random number of proton-proton collisions (with a Poisson expectation between 10 and 35, depending on the LHC parameters) called events⁸. Two colliding protons produce a small firework in which part of the kinetic energy of the protons is converted into new particles. Most of the resulting particles are very unstable and decay quickly into a cascade of lighter particles.

Based on these properties, the properties of the decayed parent particle are inferred, and the inference chain is continued until reaching the heaviest primary particles. Given the elusive nature of neutrinos, their minuscule mass, and the way they oscillate between flavors, one could very well imagine that the mass of leptons comes from an entirely different mechanism. Hence the importance of measuring as precisely as possible the coupling of the Higgs to tau arises.

Particle colliders enable us to probe the fundamental nature of matter by observing exotic particles produced by high-energy collisions. Because the experimental measurements from these collisions are necessarily incomplete and imprecise, machine learning algorithms play a major role in the analysis of experimental data. The high-energy physics community typically relies on standardized machine learning software packages for this analysis and devotes substantial effort towards improving statistical power by hand-crafting high-level features derived from the raw collider measurements.

With the given dataset, we must classify whether the given event was a signal or a background noise in the process of decay for Higgs particle acceleration.

Metric:

Precision score will be evaluation metric for this problem to reduce false positive rate ie if signal is present it should not be ignored.

$$\text{Precision} = \text{TruePositives} / (\text{TruePositives} + \text{FalsePositives})$$