Online software news

Emmanuel Le Guirriec

CPPM Astro meeting

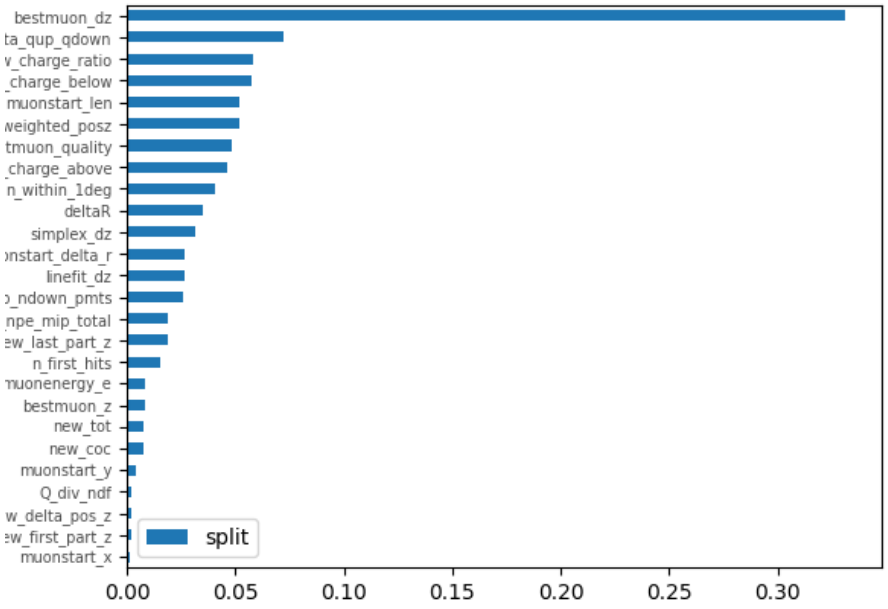
30 January 2024

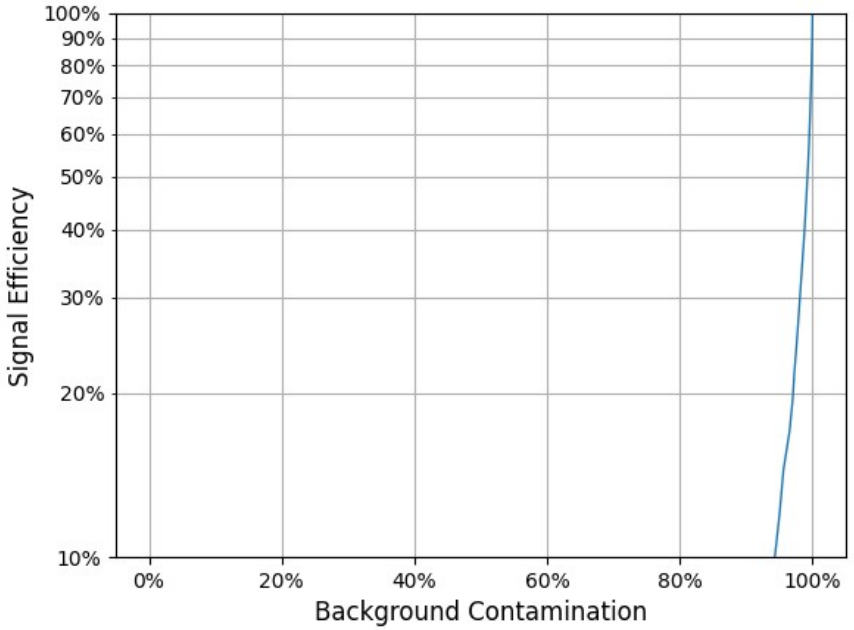
Online Classification

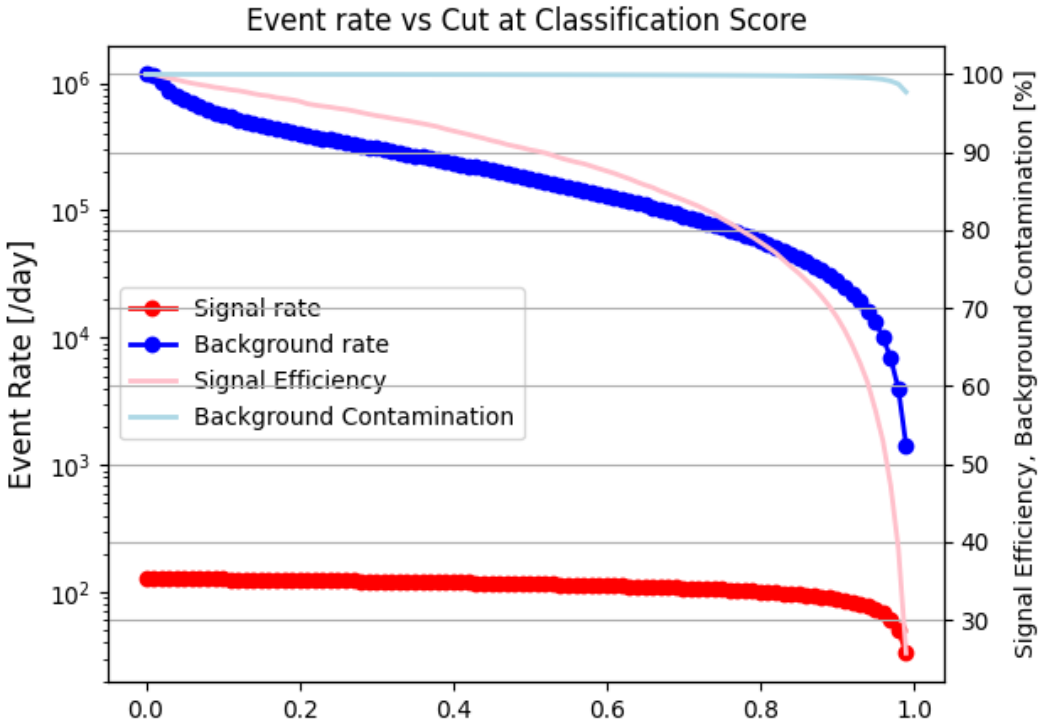
- **Xgboost version: 1.7.2 → 2.0.2**
  - Very fast
  - But gives different results
    - MAP (Mean Average Precision) metric gives probability between 0.4 and 0.6
    - Another metric is used for evaluation: Binary classification error rate
      - Error = #(wrong cases)/#(all cases)

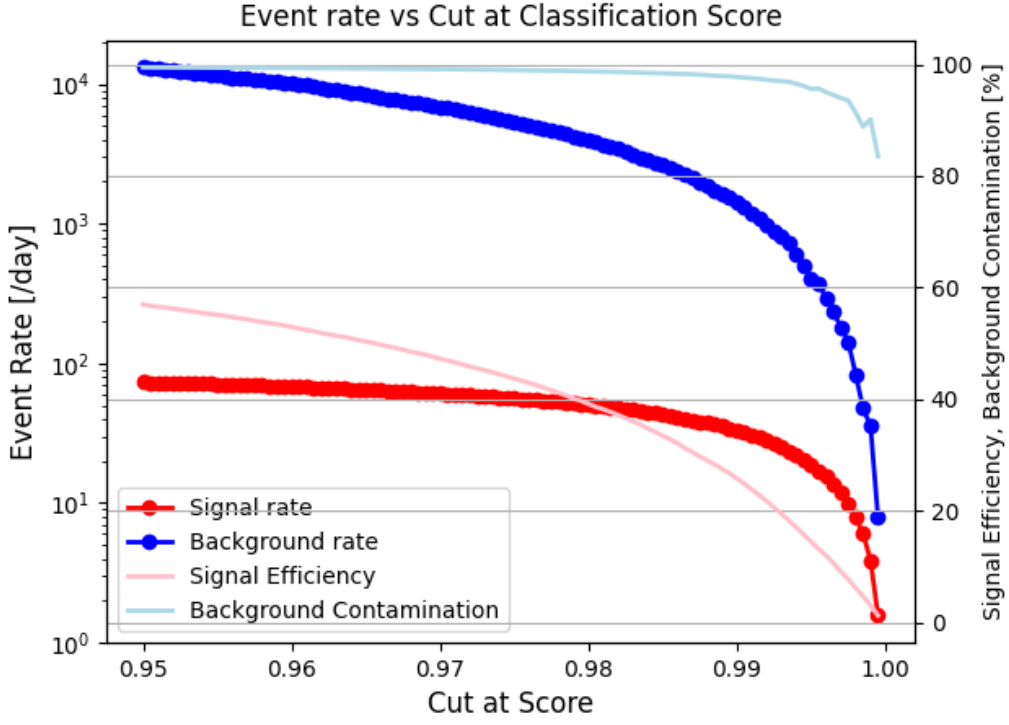
- **Model ready for production if ...**

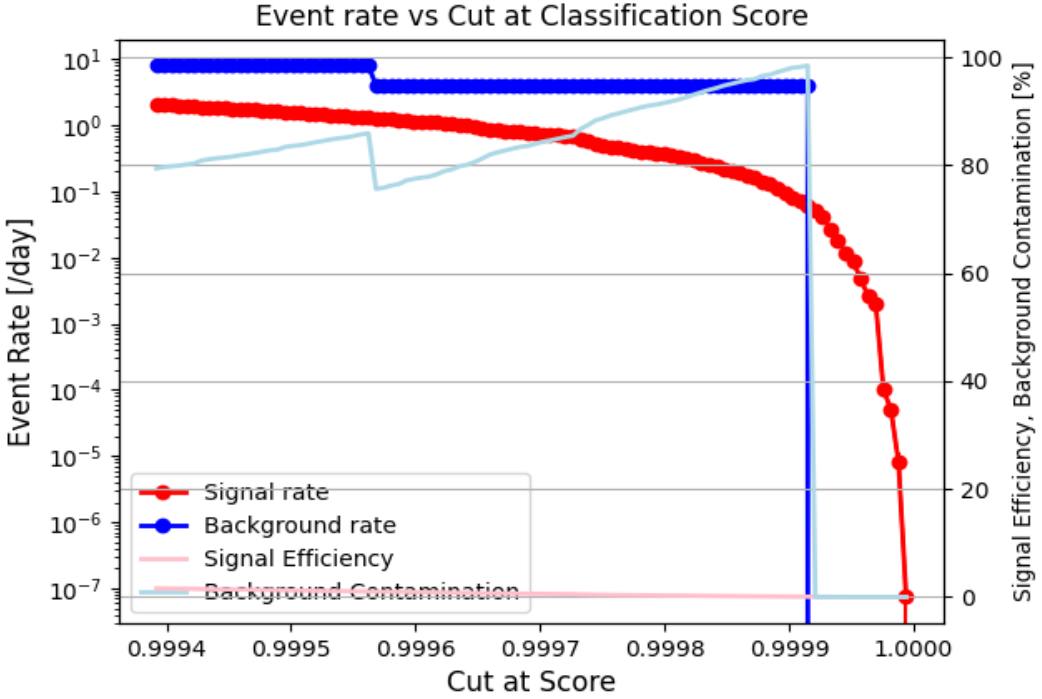
Feature importance split and SignalEff_vs_BkgCon

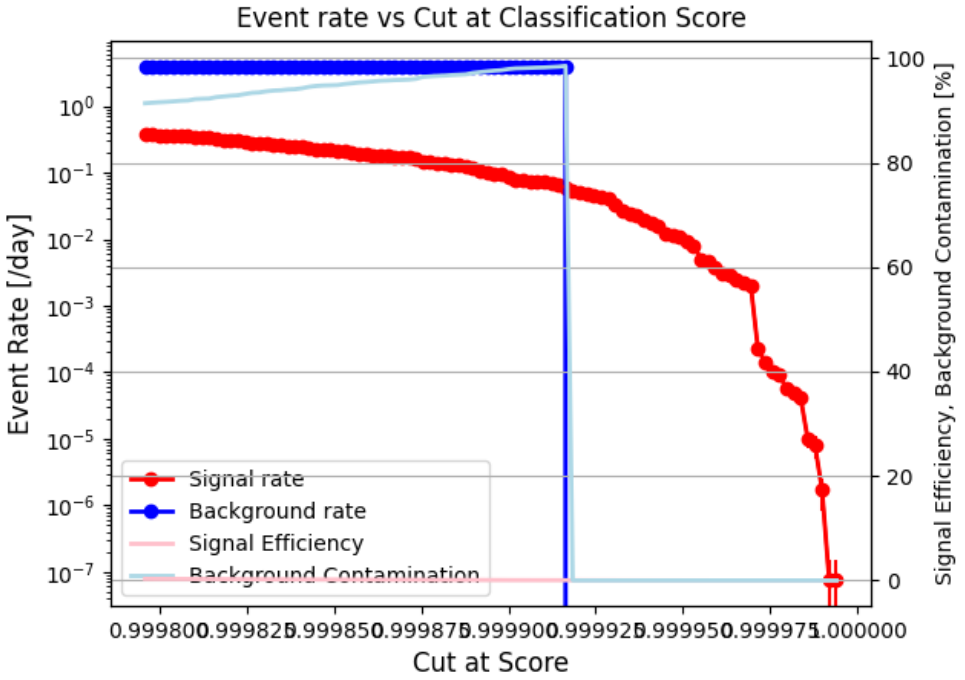





Event rate vs Cut at Classification Score
Event Rate [/day]
Signal Efficiency, Background Contamination [%]
Signal rate
Background rate
Signal Efficiency
Background Contamination


Event rate vs Cut at Classification Score
Event Rate [/day]
Signal Efficiency, Background Contamination [%]
Signal rate
Background rate
Signal Efficiency
Background Contamination
Cut at Score


Event rate vs Cut at Classification Score
Event Rate [/day]
Signal Efficiency, Background Contamination [%]
Signal rate
Background rate
Signal Efficiency
Background Contamination
Cut at Score


Event rate vs Cut at Classification Score
Event Rate [/day]
Signal Efficiency, Background Contamination [%]
Signal rate
Background rate
Signal Efficiency
Background Contamination
Cut at Score

New server for online production: antorcamm3

- Will be dedicated to production only (RTA)
- Antorcamm1 also for RTA
- Antorcamm2 dedicated to MM and monitoring
- Antorcadaq9 for development
- Installation of Alma linux 9 this week

Snakemake pipeline to prepare ML datasets

- **Git project:**

  https://git.km3net.de/online_multimessenger/orca_ml_datasets_preparation/

- **Ok for small production but not suited to mass production**
  - To many data to be downloaded
  - Use xrootd protocol to stream root input files and produce h5 files
- **Hard to know what are the type of available files and where they are stored. Same for detx files.**
  - Documentation (aanet, dst,..)?
  - Maybe some code to be rewritten … but not this week