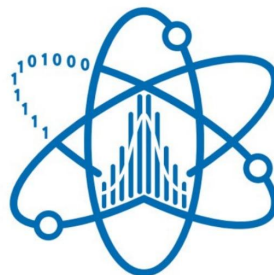


НЕР Анализ данных

Томский Государственный Университет

14 июля 2025



Направления

Повторение анализа на ATLAS Open Data [<https://arxiv.org/abs/1612.03636>]:

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года ($\sqrt{s} = 13$ ТэВ)

Повторение анализа:

WVZ в pp столкновениях - методология улучшения точности анализа данных в эксперименте ATLAS (Никита Плетнев)

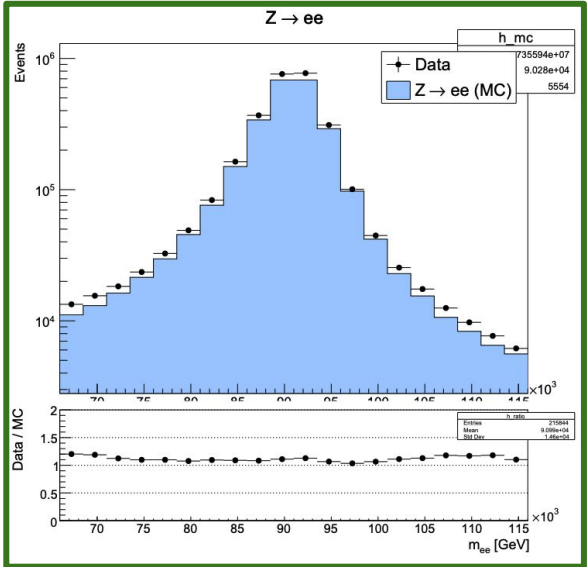
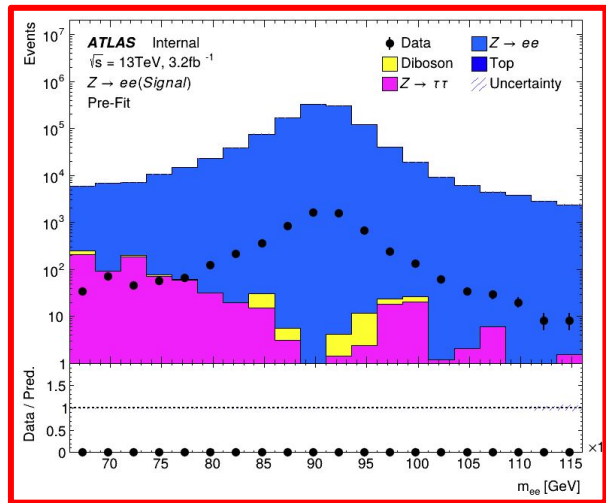
Оценка эффективности алгоритмов машинного обучения для задач классификации в анализах данных физики высоких энергий:

Измерение дифференциальных сечений топ-кварка, производимого в ассоциации с W бозоном на 13 TeV с использованием детектора ATLAS (Неда Фироз)

Рождение и дифференциальные распределения J/Ψ, ω и φ мезонов в протон-протонных столкновениях на Большом адронном коллайдере (выпускная квалификационная работа Ильи Полищука под руководством Ирины и Вадима), Илья готовит статью.

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года

- Настроена **конфигурация** TRExFitter
- **Проверена полная цепочка анализа на небольшой статистике** — процедура **отрабатывает корректно**:
- Подготавливаем гистограммы на **полной статистике** с корректным включением весов: **нормировка по сечению, pile-up, светимость**
- **Следующие шаги**: включение систематик в анализ и их пошаговая отладка (Дмитрий), настройка фита (Мария)



$$w_{\sigma} = \frac{\int L dt \sigma}{\eta \sum_i w_i}$$

светимость: 3045.834 pb⁻¹

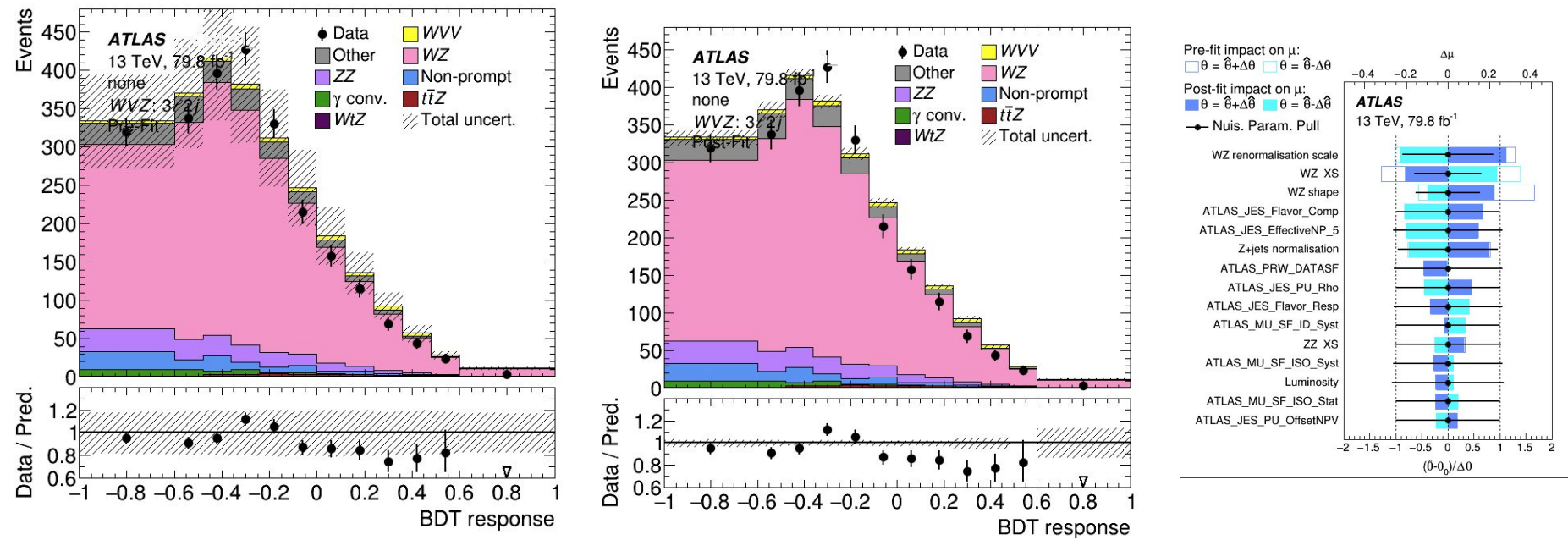
Форма MC соответствует данным, но:

- **МС сильно выше данных по норме** во всём диапазоне

Полный запуск анализа BDT и TrexFitter выполнен

Результаты полностью совпадают с последней итерацией, полученной Олесей

План: обновить обучение BDT и систематически исследовать влияние параметров на точность анализа и систематические неопределённости (Никита Плетнев)



Алгоритмы машинного обучения для задач классификации

Цель: оценить эффективность современных методов машинного обучения (Transformers, GAN, RNN, GNN, Diffusion Models и др.) для задач классификации в физике частиц (Неда Фироз)

- Анализ проводится на основе открытых данных **ATLAS OpenData** и репозитория **HEPData**.
- Рассматривается процесс **производства топ-кварка в ассоциации с W-бозоном** при энергии **13 ТэВ**, зарегистрированный детектором ATLAS.

Measurement of differential cross-sections of a single top quark produced in association with a W boson at $\sqrt{s} = 13$ TeV with ATLAS

The differential cross-section for the production of a W boson in association with a top quark is measured for several particle-level observables. The measurements are performed using 36.1 fb^{-1} of pp collision data collected with the ATLAS detector at the LHC in 2015 and 2016. Differential cross-sections are measured in a fiducial phase space defined by the presence of two charged leptons and exactly one jet matched to a b -hadron, and are normalised with the fiducial cross-section. Results are found to be in good agreement with predictions from several Monte Carlo event generators.

5 December 2017

Contact: [Top conveners](#) internal

[Eur. Phys. J. C 78 \(2018\) 186](#)

e-print [arXiv:1712.01602](#) - pdf from [arXiv](#) - [Physics Briefing](#)

[Inspire record](#)

[Data points](#)

[Figures](#) | [Tables](#) | [Auxiliary Material](#)

Алгоритмы машинного обучения для задач классификации

Все необходимые данные размещены на кластере

Необходимо подготовить **ntuples** для обучения BDT

Отсутствуют ключевые переменные, необходимые для расчёта входных переменных BDT:

- Пропущенная поперечная энергия (MET)
- Флаг или значение b-tagging

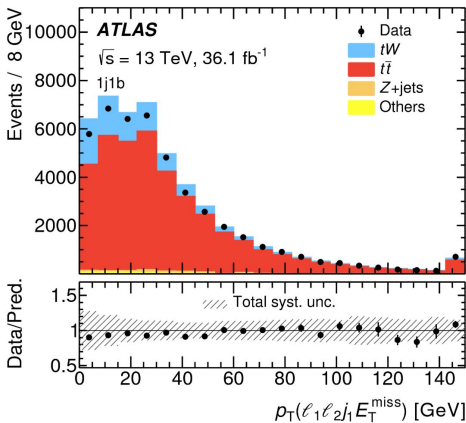
Следующие шаги:

- Добавить MET и b-tagging в ntuple-файлы *(уже выполнено)*
- Запуск генерации ntuples

Параллельно:

- Сгенерировать тестовые ntuples
- Построить гистограммы всех входных переменных BDT
- Провести тестовое обучение BDT до полной генерации ntuples
 - Неда уже подготовила код для тренировки BDT и начала тестирование
- Проверить критерии отбора событий и оценить статистику выборки

Variable	$S [10^{-2}]$
$p_T(\ell_1 \ell_2 E_T^{\text{miss}} b)$	4.1
$\Delta p_T(\ell_1 \ell_2 b, E_T^{\text{miss}})$	2.5
ΣE_T	2.3
$\eta(\ell_1 \ell_2 E_T^{\text{miss}} b)$	1.3
$\Delta p_T(\ell_1 \ell_2, E_T^{\text{miss}})$	1.1
$p_T(\ell_1 \ell_2 b)$	1.0
$C(\ell_1 \ell_2)$	0.9
$m(\ell_2, b)$	0.2
$m(\ell_1, b)$	0.1
BDT response	8.1



Thank you!

Данные Open Data Atlas 2015-2016

Open Data Atlas (<https://opendata.cern.ch/record/80020>)

Формат данных: 33 fb-1 DAOD PhysLight format

- Данные, собранные детектором ATLAS в рамках экспериментов на Большом адронном коллайдере за 2015-2016 год
- 7 групп симулированных физических процессов, включающих номинальные процессы (электрослабые бозоны, Хиггс, QCD-джеты, топ-кварки) и вариации с различными параметрами этих процессов необходимые для оценки систематических ошибок измерений.

Related datasets

Run 2 2015 [proton-proton](#) collision data
[ATLAS DAOD_PHYSLITE format Run 2 2015 proton-proton collision data](#)

Run 2 2016 [proton-proton](#) collision data
[ATLAS DAOD_PHYSLITE format Run 2 2016 proton-proton collision data](#)

MC simulation electroweak boson nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation electroweak boson nominal samples](#)

MC simulation exotic signal samples
[ATLAS DAOD_PHYSLITE format MC simulation exotic signal samples](#)

MC simulation Higgs nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs nominal samples](#)

MC simulation Higgs systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs systematic variation samples](#)

MC simulation QCD jet nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet nominal samples](#)

MC simulation QCD jet systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet systematic variation samples](#)

MC simulation SUSY signal samples
[ATLAS DAOD_PHYSLITE format MC simulation SUSY signal samples](#)

MC simulation top nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation top nominal samples](#)

MC simulation top systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation top systematic variation samples](#)

~9 миллиардов событий (реальные данные и симуляции) при
общем размере ~66.87 ТБ.

Production ntuples

Доступ к системам ATLAS:

- Успешно настроен доступ к инфраструктуре ATLAS с inter-ноды кластера ТГУ
- Отлажена система аутентификации и взаимодействия с основными сервисами ATLAS (Rucio, CVMFS и др.)
- Завершено расширение доступа на все вычислительные ноды — полная интеграция анализа в распределённую среду

Распределённые вычисления:

- Настроена система распределённых заданий HTCondor на кластере ТГУ
- Проведено тестирование отправки задач и мониторинга, адаптирован интерфейс фреймворка под многонодовую архитектуру

Предпродакшн и тестирование:

- Запущен предварительный продакшн ntuples на моделированных данных (MC)
 - Цель: протестировать стабильность цепочки обработки и поведение фреймворка в реальных условиях
-

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов

Реализован фреймворк для преобразования данных из формата DAOD PhysLight в удобный и читаемый формат ntuples, пригодный для последующего физического анализа

- Репозиторий: <https://git.hep.tsu.ru/mdidenko/framework/>
- Построен на базе **Athena AnalysisBase v25.2.45**

Выходной ntuple: дерево **analysis**, с сохранением информации по событиям, электронам, мюонам и джетам

- Базовая преселекция
- Подключение триггеров 2015 года (электронные и мюонные цепочки, полное соответствие данным)
- Реализация отбора объектов (электроны, мюоны, джеты)
- Удаление перекрытий (OR) между электронами, мюонами и джетами

Все параметры управления заданы в **конфигурационном YAML-файле**

- Возможности адаптации под конкретные задачи:
 - включение/отключение систематик (**runSystematics: False**)
 - переопределение контейнеров, ID, изоляций и триггеров
 - гибкая настройка выходных переменных и структуры ntuple

Фреймворк может использоваться для различных задач ATLAS без изменения кода — достаточно скорректировать конфигурацию

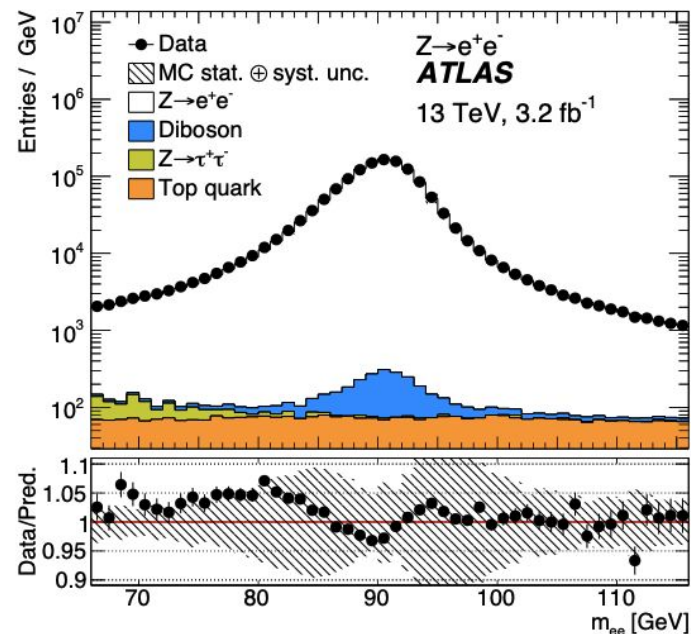
Продакшн на данных Run 2

Готовится первый продакшн на реальных данных 2015 года

- Данные: 3.2 fb^{-1} ($\sim 9 \text{ TB}$)
- MC: $Z \rightarrow ll$ (электроны и мюоны)
- Ожидаемая длительность: 2 недели

Планируемые первые результаты анализа

- Предварительные распределения отношения данных к MC (data/MC)
- Переменная: инвариантная масса Z-бозона
- Включить систематические неопределенности
- Ожидаемое завершение подготовки распределений — через 2 недели



<https://arxiv.org/abs/1612.03636>