

# ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБРАБОТКИ ДАННЫХ ЭКСПЕРИМЕНТА ALICE

---

Сергей Солохин, группа Б02-910

13 октября 2020 года

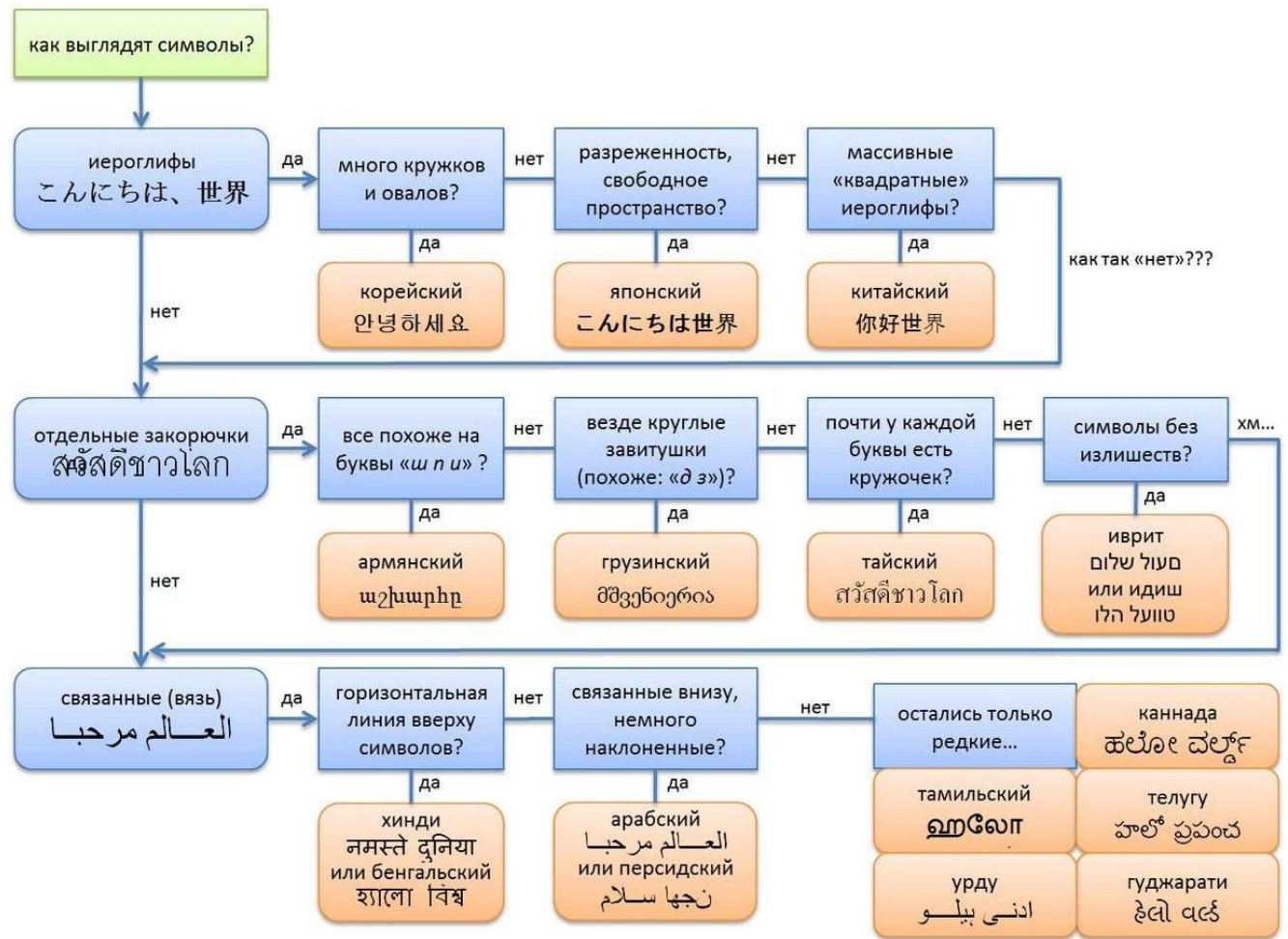
Для семинара ЛФВ МФТИ

# Дерево решений (Decision tree)

Часто применяется для задач классификации (например, справа - дерево для определения неизвестного языка по его письменности)

Совокупность узлов и рёбер. Узел содержит некоторое условие, обычно логическое. Различные результаты проверки условия в узле соответствуют различным рёбрам.

Каждому решению можно сопоставить определённый вес (характеристику важности решения)



# Расширяемое дерево (Boosted decision tree)

- Совокупность многих деревьев решений ("лес");
- Наиболее популярный алгоритм построения - AdaBoost, строящий каждое следующее дерево на основании данных, неверно классифицированных предыдущим деревом;
- Простой для понимания метод, требующий знаний лишь основ теоретической информатики;

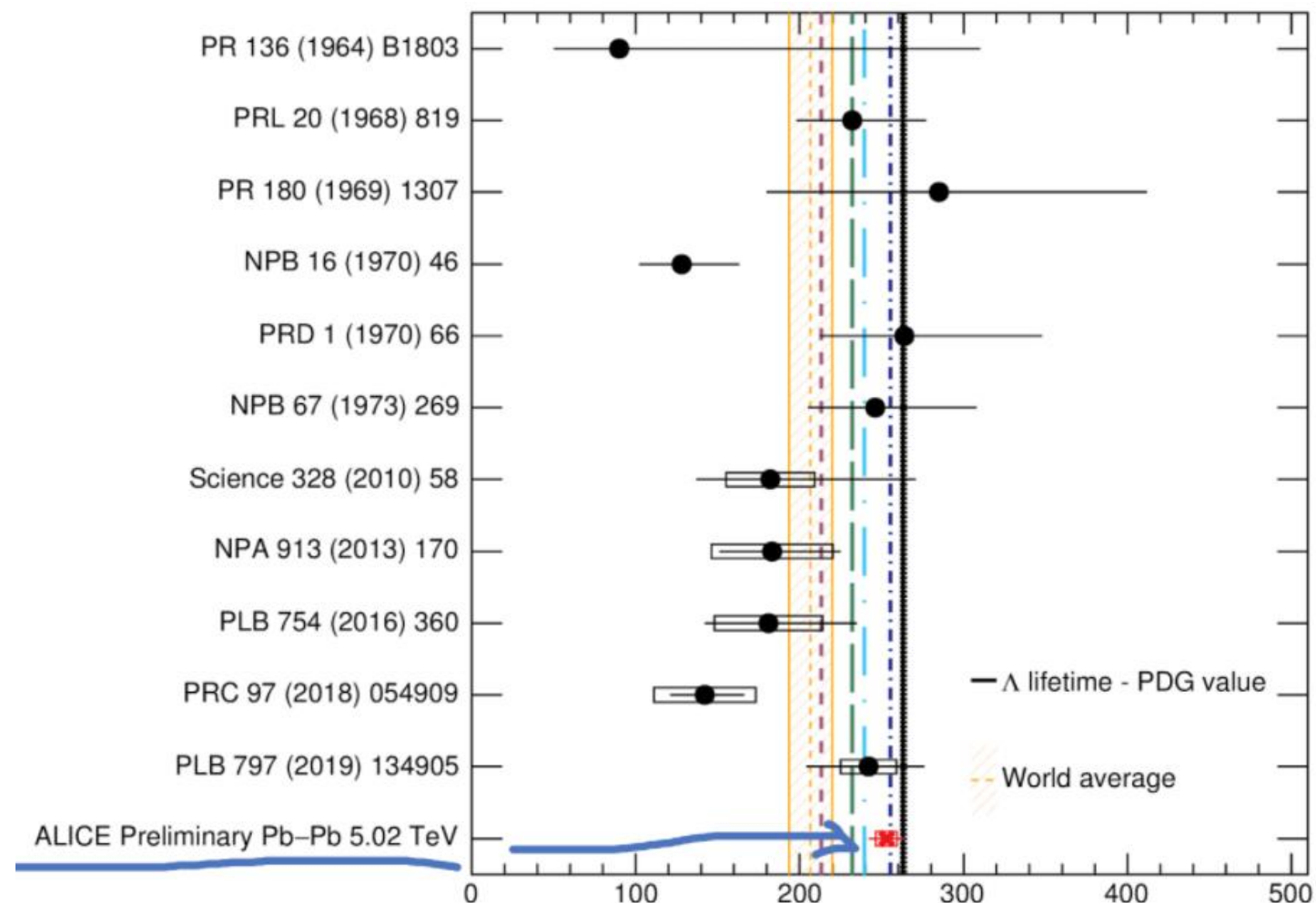
# Проблемы деревьев решений

- Деревья подвержены переобучению - необходимо найти оптимальные число входных параметров и число деревьев в лесу;
- Необходимо критически относиться к включаемым в дерево узлам, так как часть их может не влиять на результат.

# Машинное обучение в БАК ЦЕРН

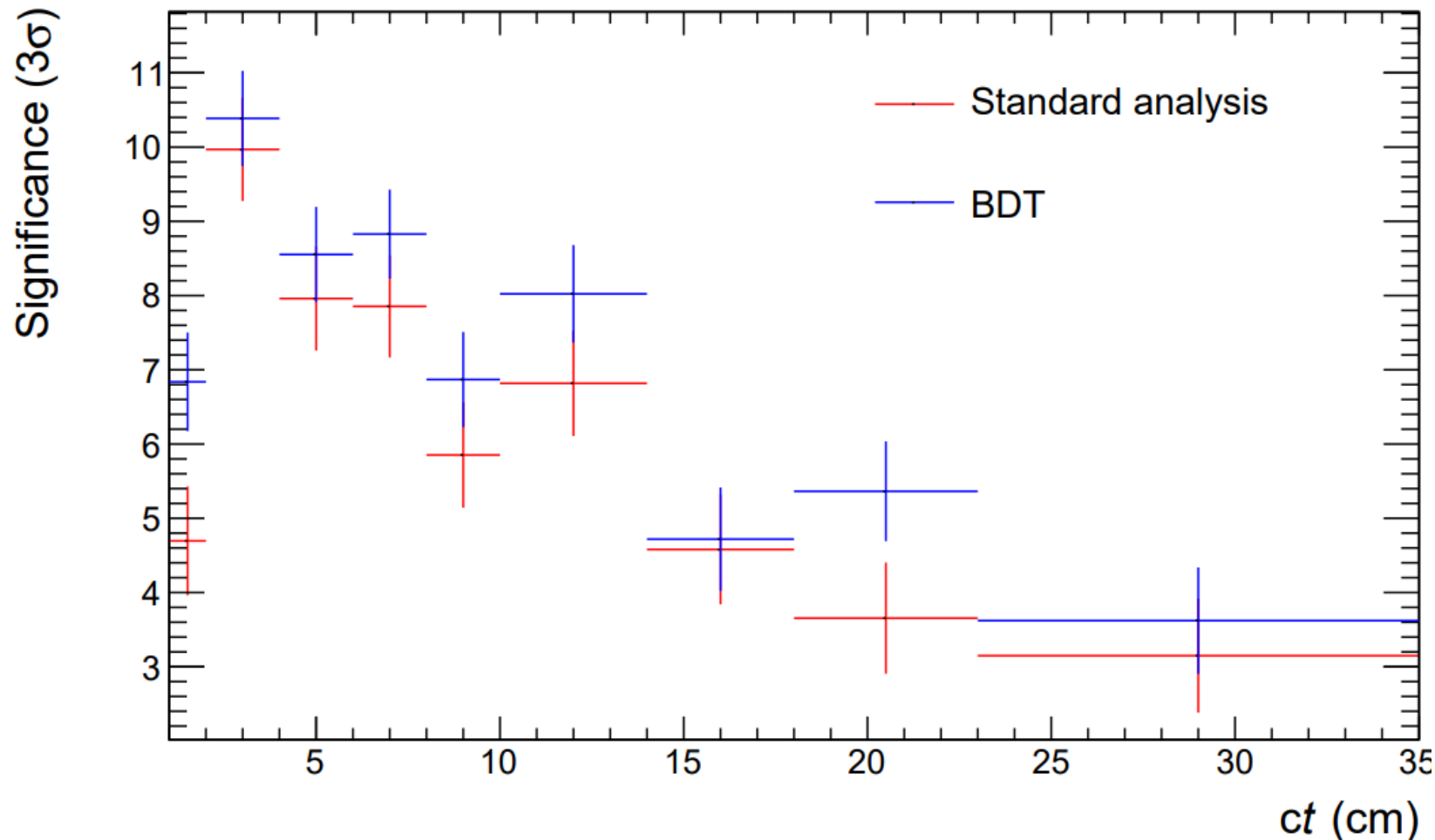
В настоящее время методы машинного обучения используются в эксперименте LHCb (Гущин, 2018 год)

Справа - результаты применения машинного обучения для оценки времени жизни гипертритона (в пикосекундах; physics week 22.09.2020; Francesco Mazzaschi, ICHEP 2020, Prague, 31/07/2020)



# Сравнение ТОЧНОСТИ МЕТОДОВ

Справа сравнение параметра Significance (отношение числа Зарегистрированных сигналов К общему числу событий), рассчитанного стандартными Методами и с помощью BDT (Hypertriton production ..., ALICE 2019)



# Машинное обучение в эксперименте ALICE

- Поиск низкоэнергетичных фотонов в калориметре (Tijmen Schaapherder's Bachelor Thesis, 2018);
- Есть данные для обучения деревьев для поиска гипертритона;
- Точность методов можно проверить с помощью данных по  $\Sigma_0$  в столкновениях  $p$ - $p$  при 7 ТэВ в качестве тестовых данных.

# Способы реализации

- Встроенные средства фреймворка *ROOT* (*TMVA*, приоритетное средство);
- Библиотеки *Python* (*scikit-learn*, *PyTorch*, *Keras*, *Tensorflow*, *PyBrain* и прочие);
- Библиотеки *MATLAB*.



# Дальнейшее применение

- Поиск  $\sigma_0$  - гиперона в столкновениях  $Pb-Pb$  при 5.02 ТэВ как первое применение BDT, как продолжение диплома А.В.Семичева, МИФИ, каф. 40, 27.06.2019

# СПАСИБО ЗА ВНИМАНИЕ!

---

По любым вопросам пишите по адресу [solokhin.sa@phystech.edu](mailto:solokhin.sa@phystech.edu)