



Национальный исследовательский ядерный университет
«МИФИ»

Кафедра физики элементарных частиц №40



Научная исследовательская работа студента на тему:

Применение методов машинного обучения для лучшего выделения процессов рассеяния векторных бозонов

Работа
студента 3-ого курса
Савельева Константина
Михайловича
ИЯФит

г. Москва 2020

Введение

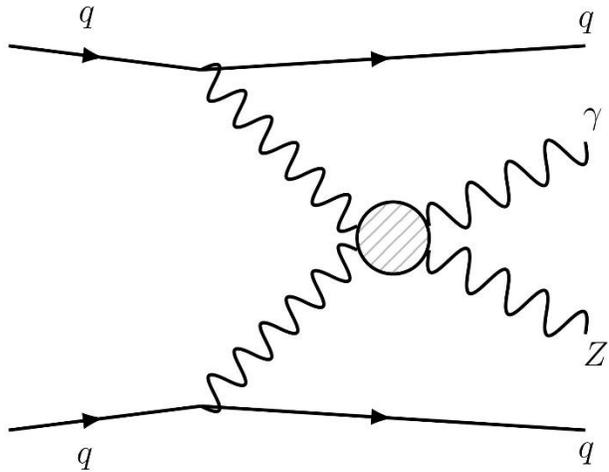


Диаграмма процесса рассеяния векторных бозонов

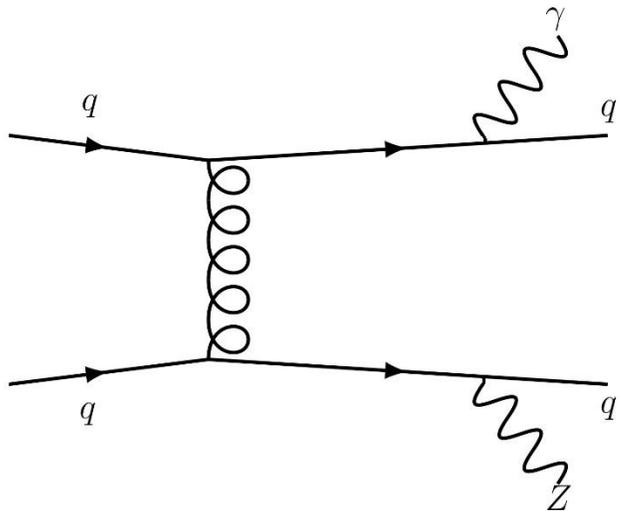


Диаграмма процесса КХД образования состояния $Z\gamma jj$

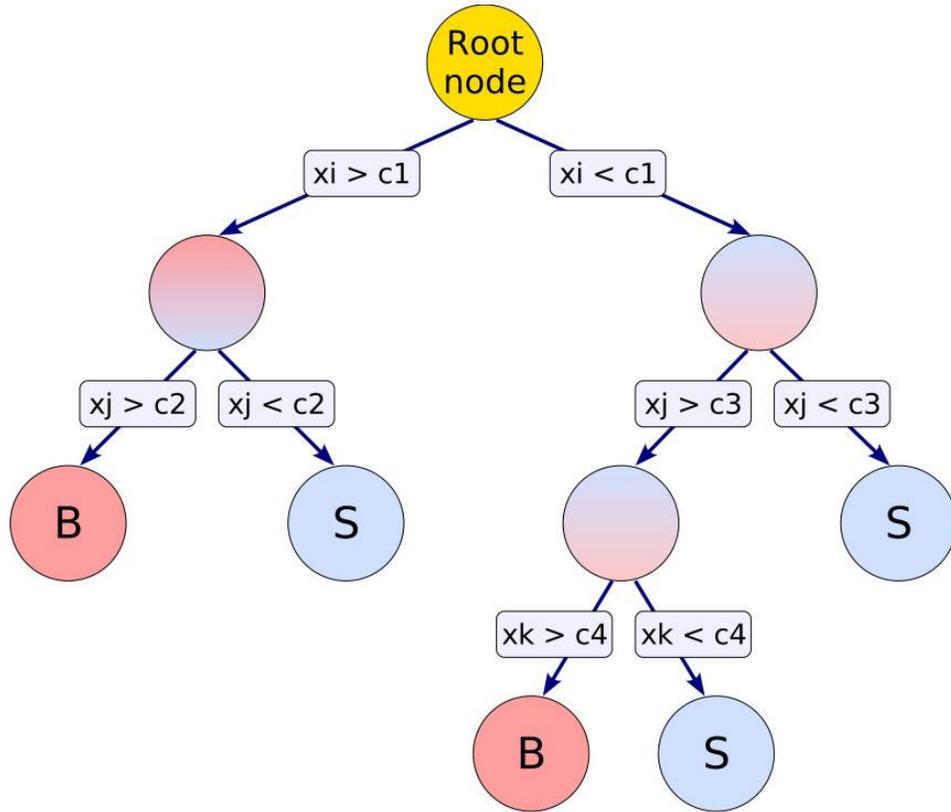
В работе изучались процессы рассеяния векторных бозонов, являющиеся интересными с точки зрения поиска так называемой «новой физики» из-за их высокой чувствительности к отклонениям параметров от Стандартной модели.

Для вычисления сечения процесса рассеяния векторных бозонов с достаточной точностью необходимо эффективно отделять сигнальные события от фоновых. Классические одномерные фиксированные отборы не дают достаточной значимости, определяемой формулой

$$\sigma = \frac{S}{\sqrt{S + B}}$$

Поэтому в работе исследовалось применение алгоритмов машинного обучения к разделению событий.

Алгоритм Boosted Decision Trees (BDT)



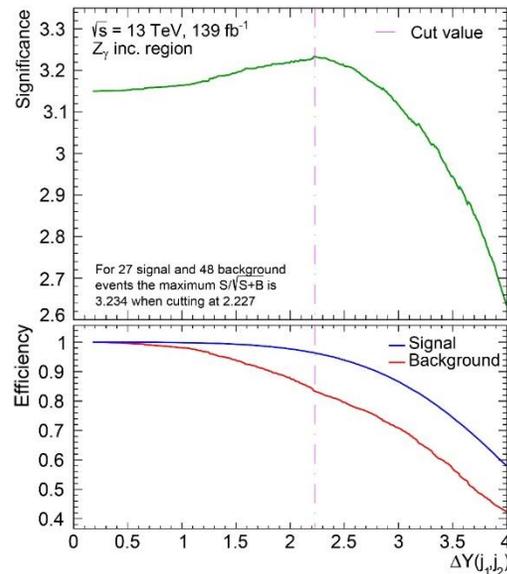
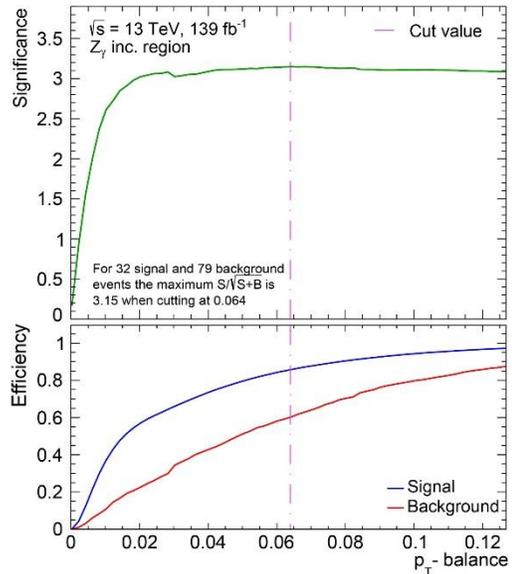
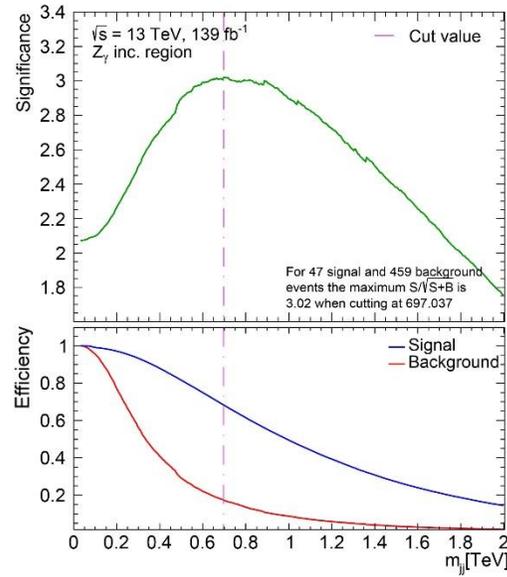
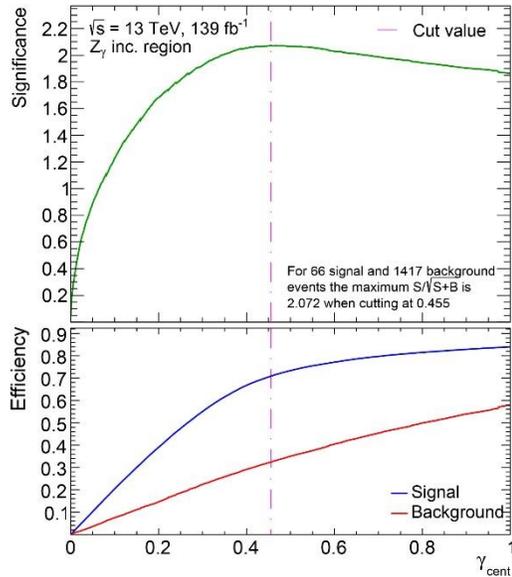
Работа метода заключается в создании леса бинарных деревьев решений.

Принцип их построения заключается в том, что для каждого узла дерева определяется переменная и ограничение по ней, которые обеспечивают максимальное разделение сигнала и фона в дочернем узле.

Для того, чтобы исключить влияние флуктуаций в данных, применяемых для обучения классификатора, используется бустинг. Он заключается в создании леса деревьев решений, в котором при создании каждого последующего дерева алгоритм переоценивает веса событий таким образом, чтобы уделить больше внимания неверно классифицированным событиям.

Результаты работы

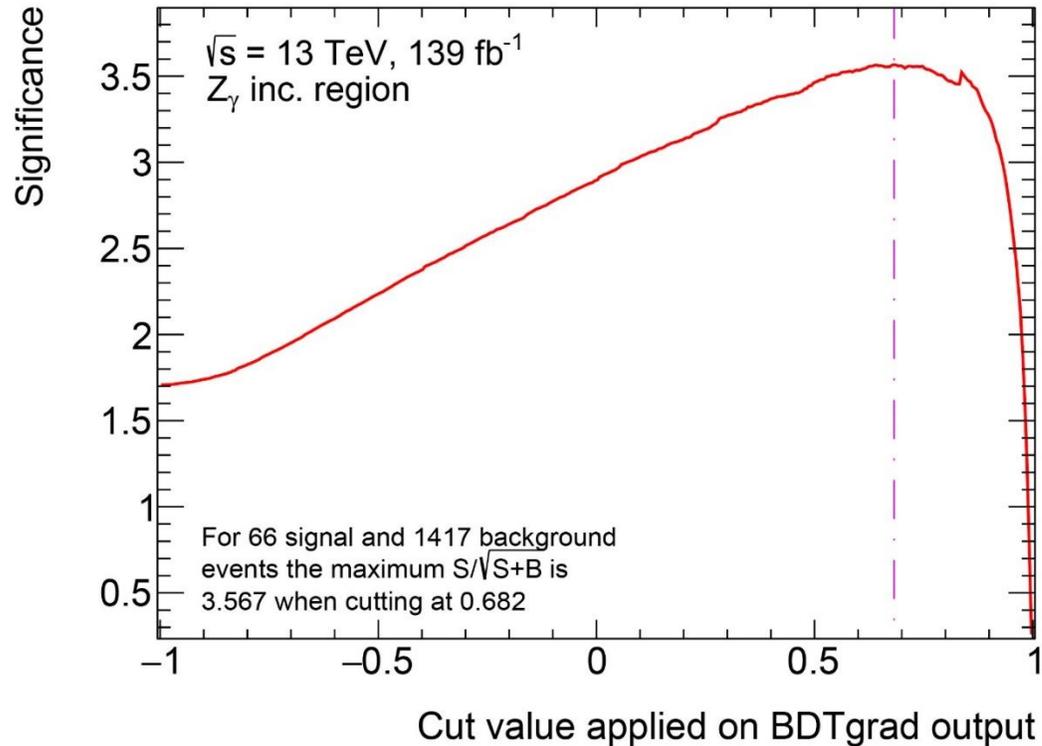
Для сравнения эффективностей фиксированных одномерных отборов и алгоритма BDT была произведена оптимизация значимости для четырёх переменных, наиболее эффективно показавших себя в разделении сигнала и фона.



Переменная	Ограничение
$\zeta(\gamma)$	< 0.455
m_{jj}	$> 697 \text{ GeV}$
$p_T - balance$	< 0.064
$\Delta Y(j_1, i_2)$	> 2.227

Значения ограничений, обеспечивающих максимальную значимость

Результаты работы



При применении классификатора BDT к исходному набору данных была оптимизирована значимость. Виден прирост значимости на 15%

$$\text{Max}(\sigma_{\text{class}}) \quad 3.57 \pm 0.06$$

$$\text{Max}(\sigma_{\text{fix}}) \quad 3.23 \pm 0.06$$

Результаты применения классификатора и фиксированных отборов

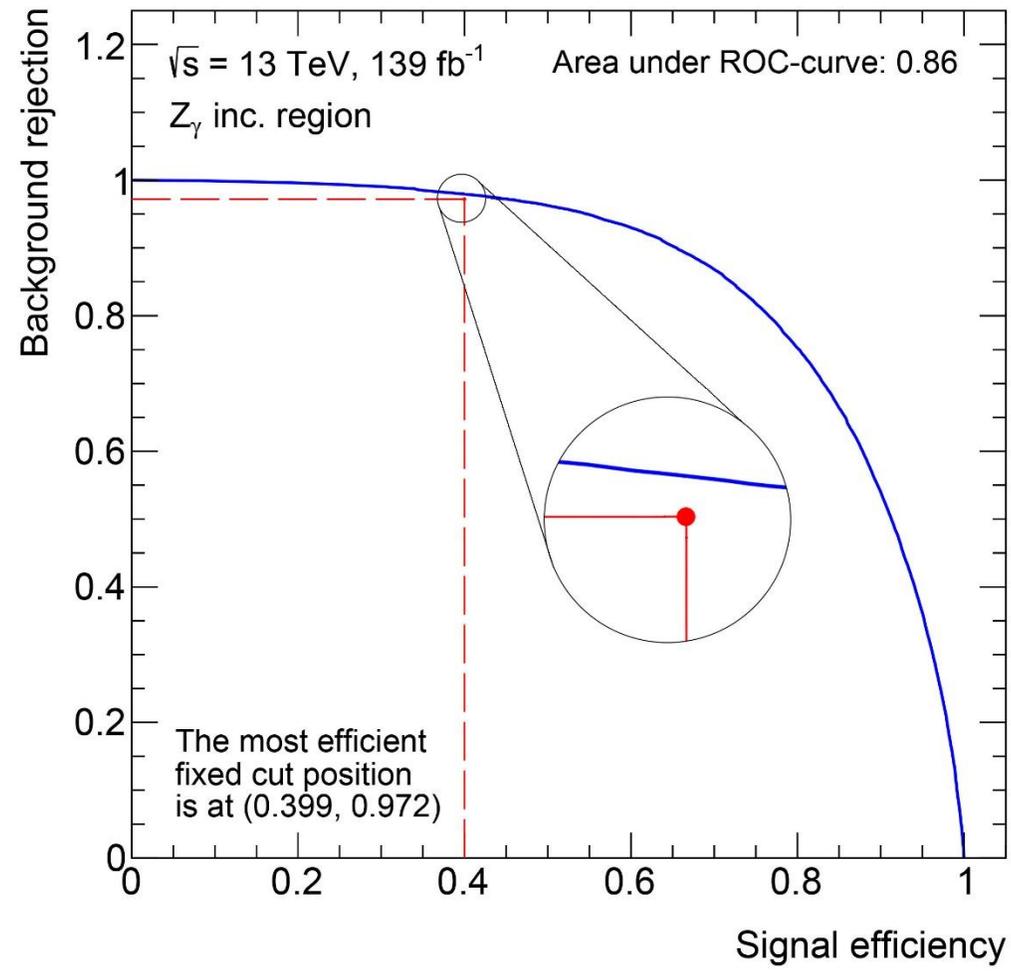
Заключение

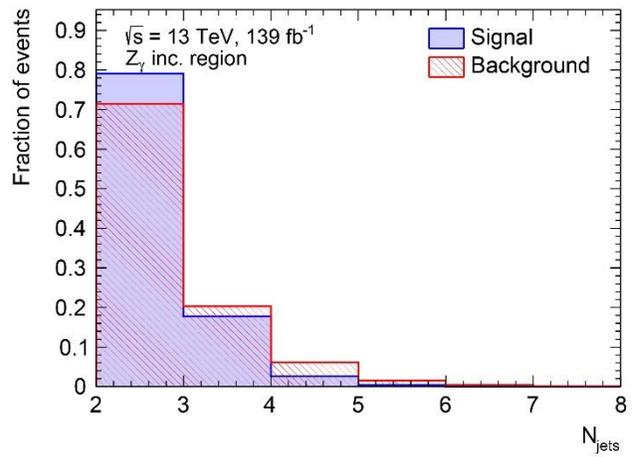
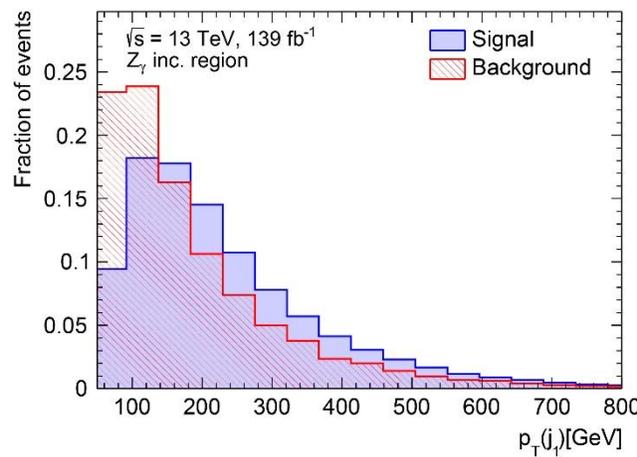
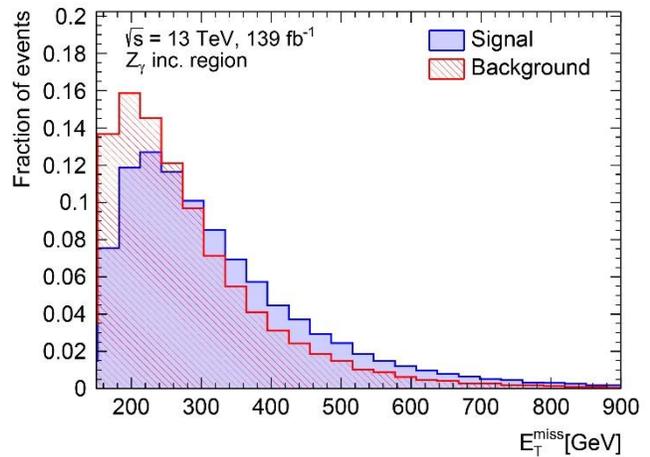
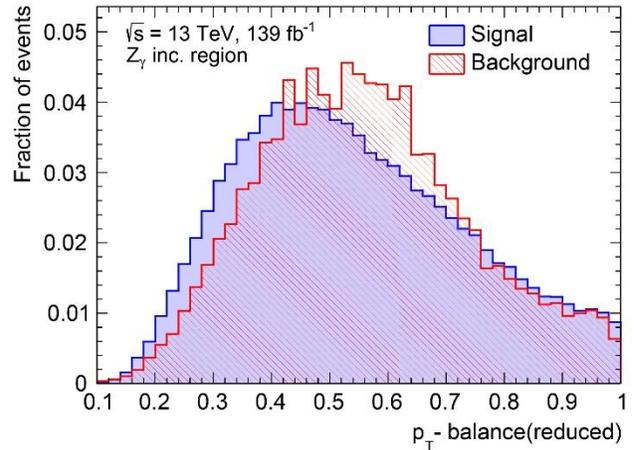
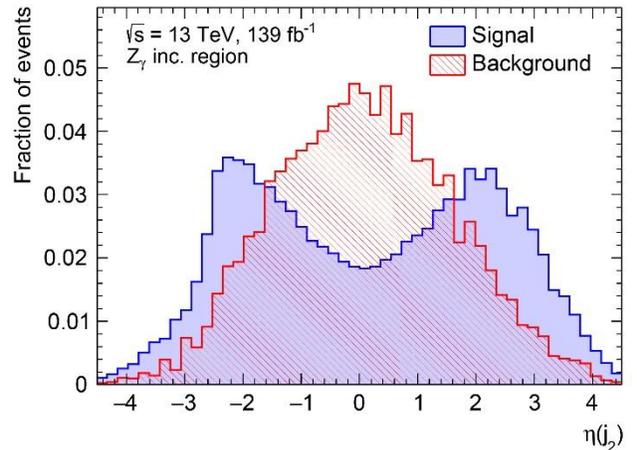
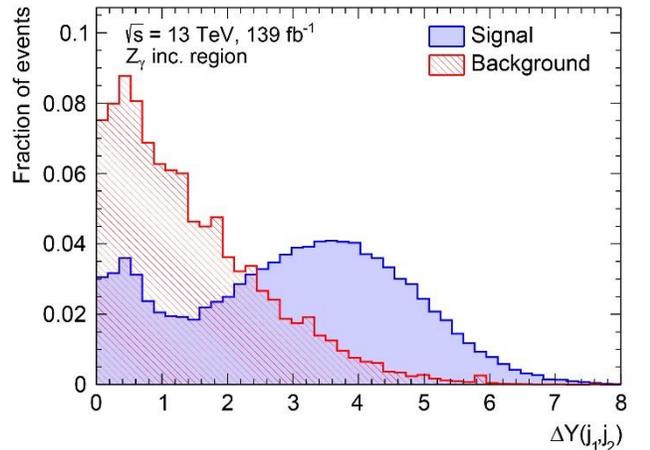
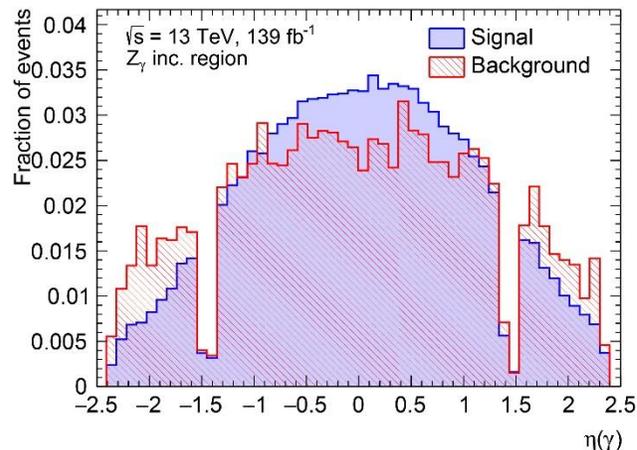
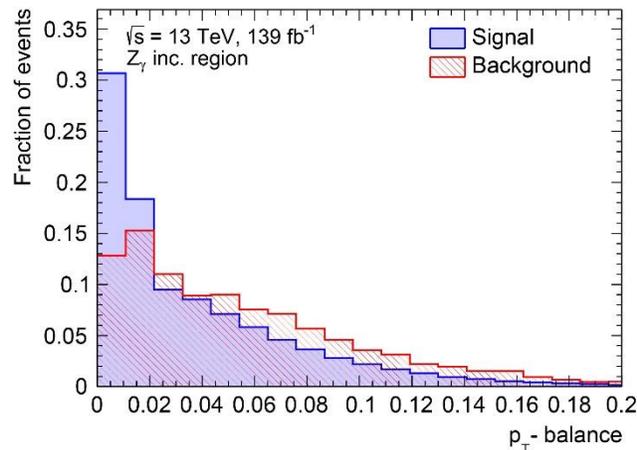
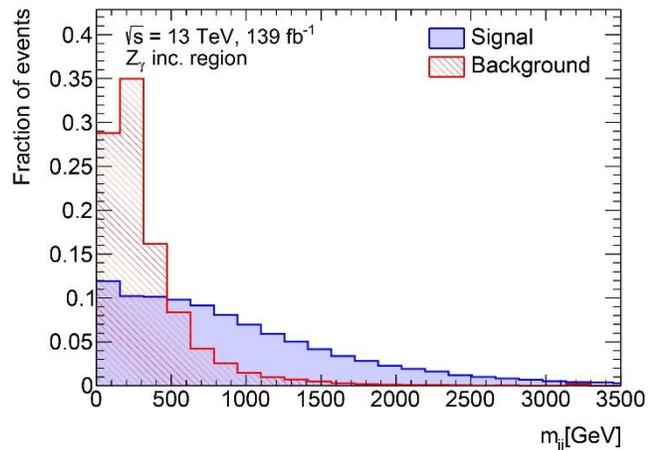
- Было произведено сравнение значимостей при применении фиксированных одномерных отборов и алгоритма BDT
- Выявлен прирост в значимости на 15% при применении классификатора BDT

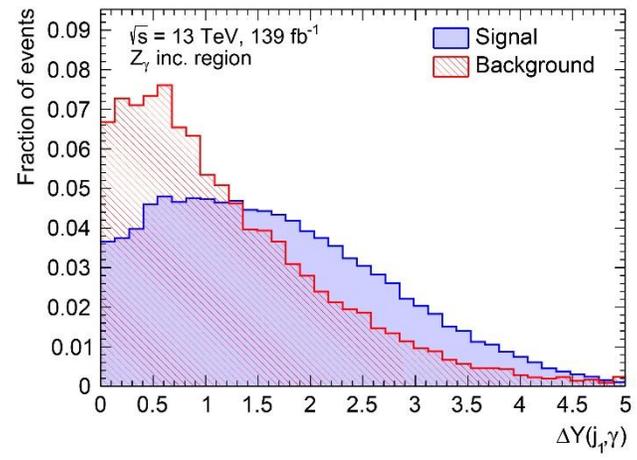
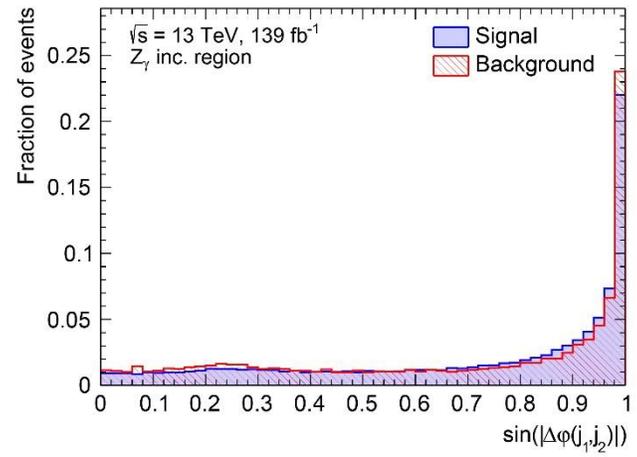
В дальнейшем планируется:

- Определение наиболее оптимальных значений настроек классификатора
- Поиск наилучших переменных для обучения
- Исследование других алгоритмов машинного обучения, а также их настройка

Спасибо за внимание!







Переменная	Ограничение
E_T^{miss}	$> 120 \text{ GeV}$
E_T^γ	$> 150 \text{ GeV}$
Число фотонов	$N_\gamma = 1$
Число струй	$N_{jets} \geq 2$
Число лептонов	$N_e = 0, N_\mu = 0$
$ \Delta\phi(\gamma, \vec{p}_T^{miss}) $	> 0.4
$ \Delta\phi(j_1, \vec{p}_T^{miss}) $	> 0.3
$ \Delta\phi(j_2, \vec{p}_T^{miss}) $	> 0.3

Количество деревьев	850
Максимальная глубина	3
Тренировочные данные	50%

	Сигнал	Фон
До применения отборов	66	1417
Классификатор	30	42
Фиксированные отборы	26	40