

# След Космического Луча

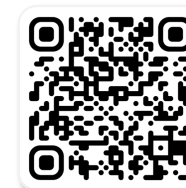
Решение от *Flush-Royal Team*



**Жамков Никита:** ML-инженер,  
техническая часть проекта и  
обработка данных



**Тимофеев Данила:** лидер,  
помощь в работе с данными

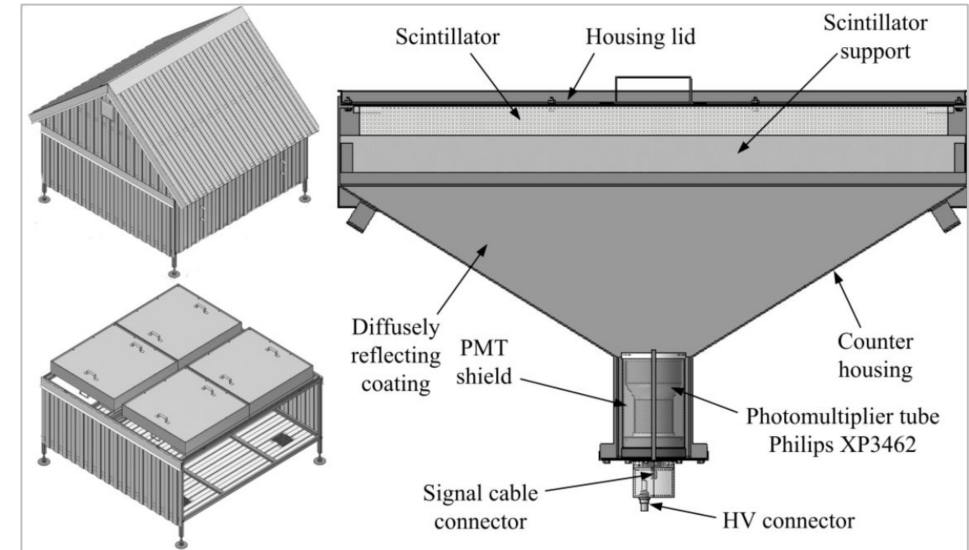
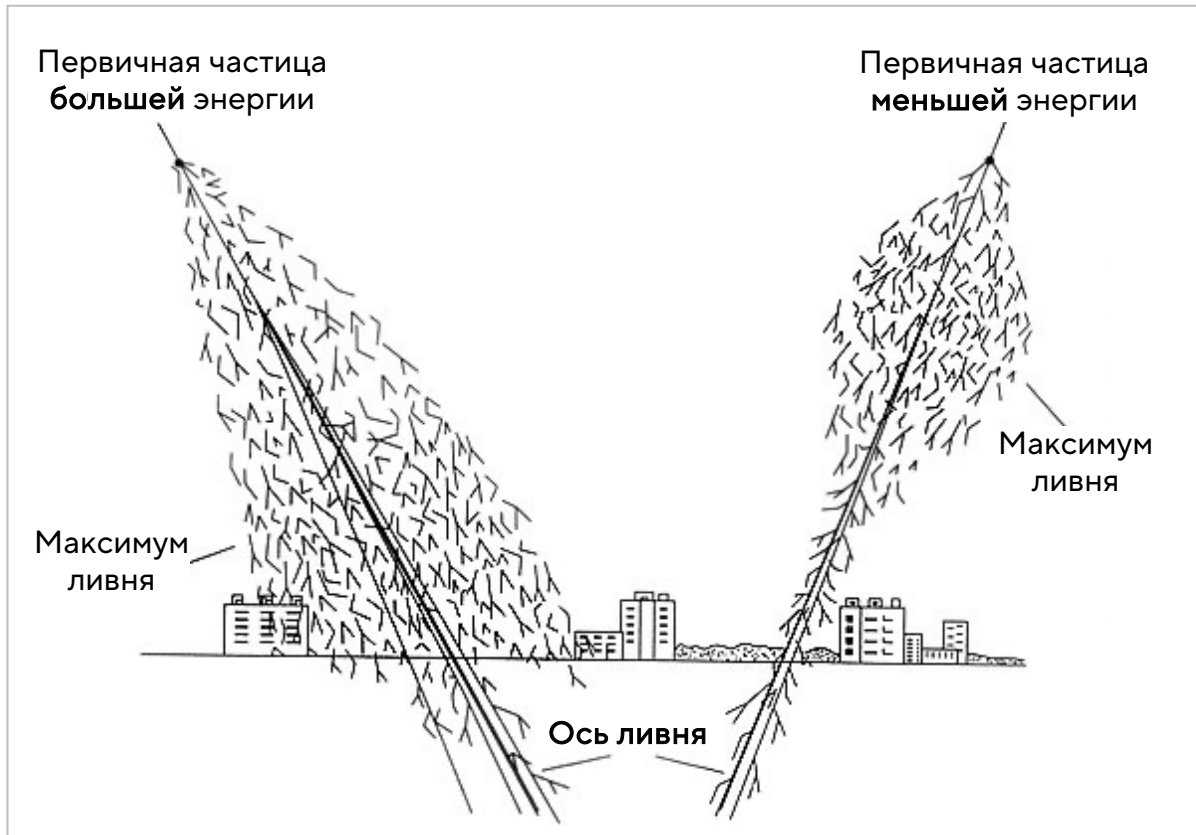


**Петров Александр:** физик,  
литературный обзор

работа над теоретической частью проекта, создание  
презентации

# Введение

Широкий атмосферный ливень (ШАЛ) – возникает при взаимодействии потока субатомных частиц с земной атмосферой, в результате чего поток «рассеивается» – **первичные** частицы взаимодействуют с атомами в воздухе и образуют огромное количество **вторичных** частиц



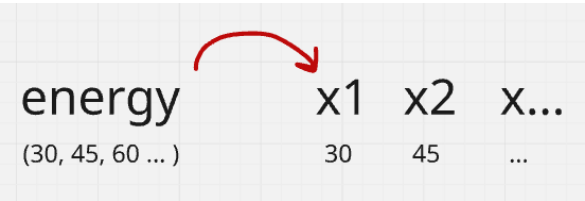
Регистрация ливней происходит на базе **НОЦ НЕВОД** при помощи станций-детекторов, разбитых на кластеры

При прохождении частицы через **сцинтиллятор**, образуется квант света – фотон, который улавливается **фотоэлектронным умножителем** и преобразуется в электрический сигнал, который несет информацию о вторичной и первичной частице

# Мощность и возраст ливня

*EdepStNE[9][4]* – 9 кластеров по 4 станции в каждой.  
*NEL, NHADR, NMU* – количество электронов, адронов и мюонов, полученных на уровне наблюдения

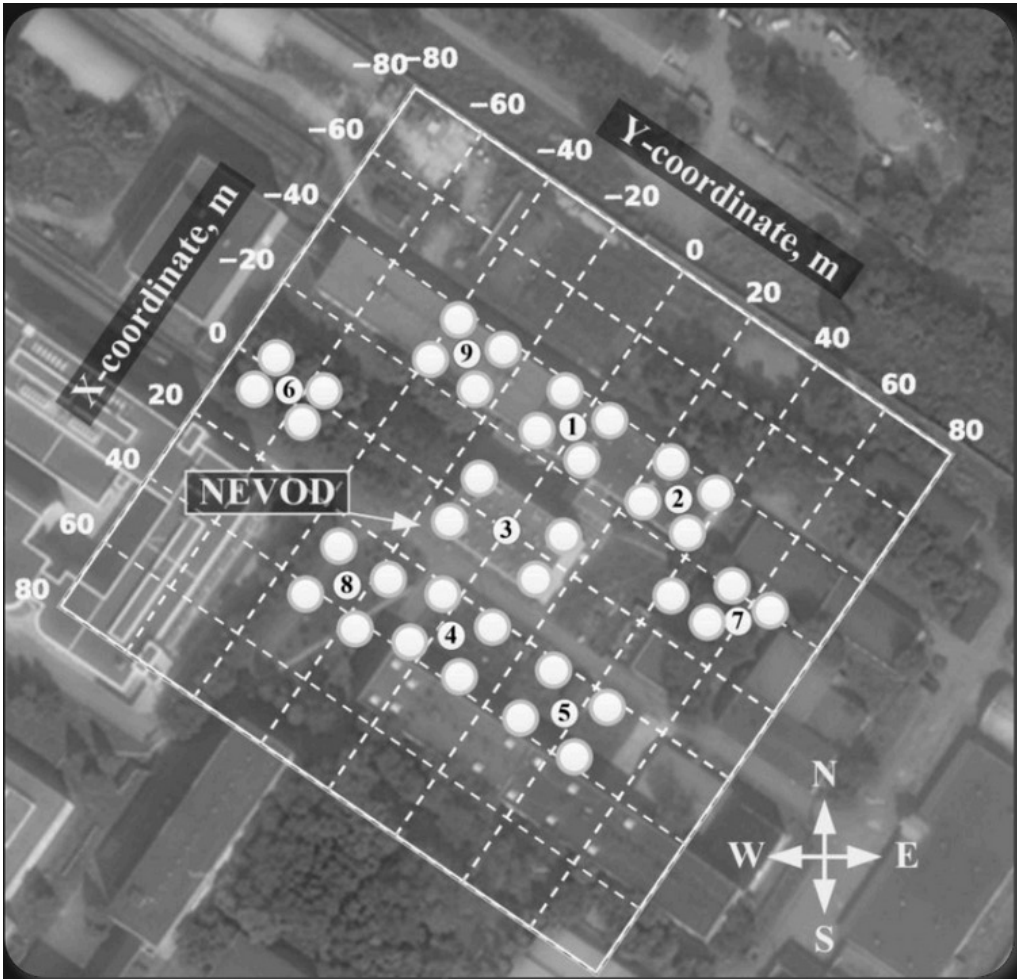
Проверена гипотеза: простое разложение значений энергии каждой станции на отдельные признаки и применение количества вторичных частиц дало впечатляющий результат для предсказания **мощности** и **возраста** ливня.



CatboostRegresso:  
depth = 3, lr = 0.05

Точность Catboost		
	MAE	R2
Power	0.029	0.995
Age	0.020	0.828

Оптимизация гиперпараметров модели не ведет к значительному улучшению точности (разница в тысячных). Было решено сфокусироваться на выделении признаков.





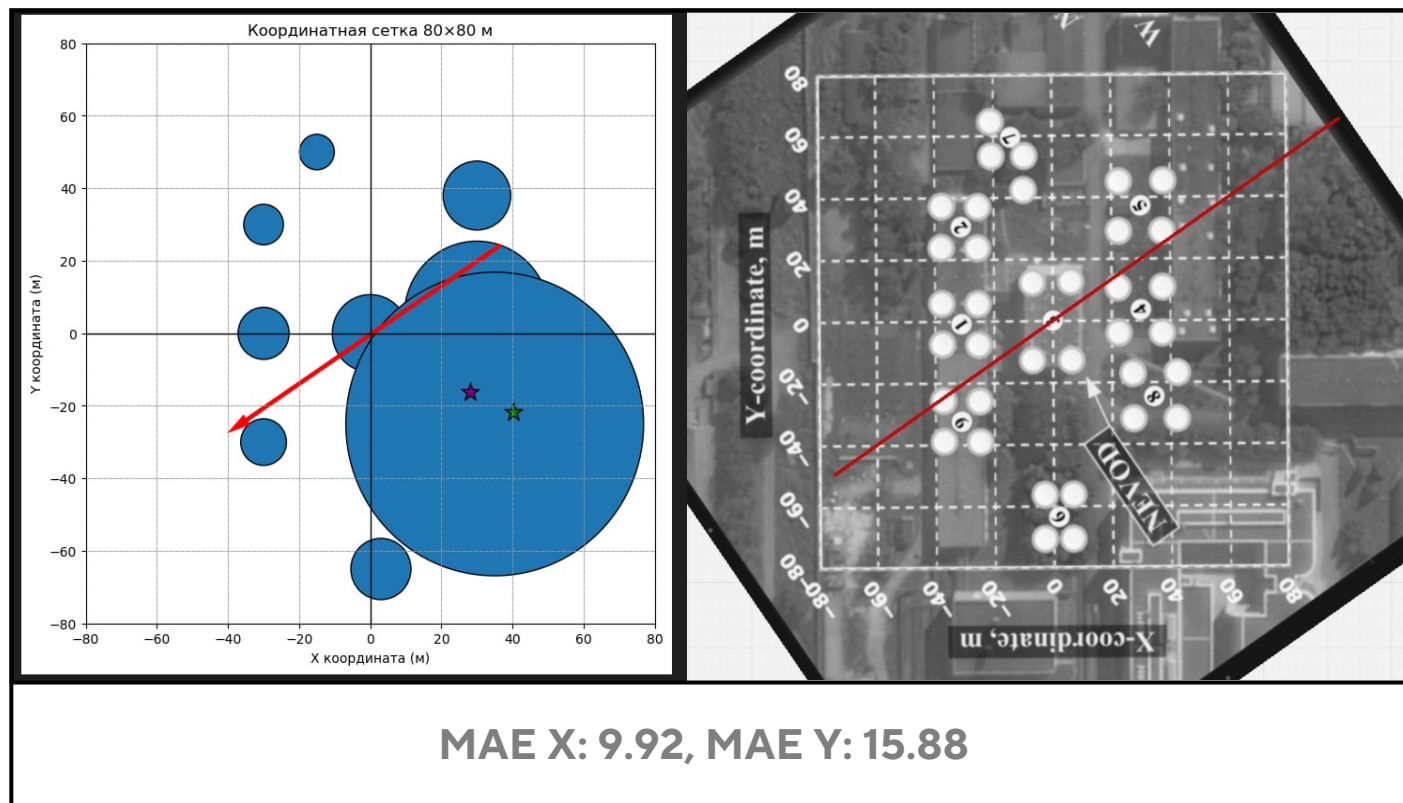
# Определение оси ШАЛ

Для определения оси ШАЛ было предложено проверить **две гипотезы**:

1. Ось можно определить по средневзвешенной энергии между кластерами
2. Ось можно определить, если найти скрытую зависимость между полученными сведениями о ШАЛ

Для того, чтобы визуализировать результаты проверки первой гипотезы, была **смоделирована карта центра** с кластерами, где 3 кластер – центральный.

Чем больше радиус фигуры – тем больше энергии было захвачено сенсорами в кластере



- ↖ Стрелка указывает на север
- ★ Предсказание первой модели
- ★ Настоящая ось



# Определение оси ШАЛ



Для модели машинного обучения нужны признаки, которые теоретически смогли помочь модели найти зависимости.

Анализируя данные стало известно, что **некоторая информация отсутствует** (результаты нулевые). К примеру:

- *nTrackSMX* - число треков (5X) в каждом из СМ ДЕКОР;
  - *nTrackSMY* - число треков (5Y) в каждом из СМ ДЕКОР;
  - *nTrackSM* - число треков (5X + 5Y) в каждом из СМ ДЕКОР;
  - *muDCR* - число треков (5 сработавших плоскостей в X проекции) в супермодулях (СМ) ДЕКОР (для операторского отбора: 3 сработавших СМ, 5 треков в X);
  - *muSM* - число треков (5 сработавших плоскостей в X и Y проекциях) в СМ ДЕКОР (для программного отбора: 3 сработавших СМ, хотя бы 1 трек в каждом);
- и т.п.

Определение оси (X,Y), Catboost		
	MAE	R2
X	5.952	0.854
Y	7.277	0.914

Модель Catboost показала **лучший результат**. Среднее отклонение по X ~6 метров, по Y ~7 метров.

Принято решение использовать решение модели МО, в связи с лучшим результатом по точности. Стоит учесть, что **метод взвешенной суммы** является более интерпретируемым.

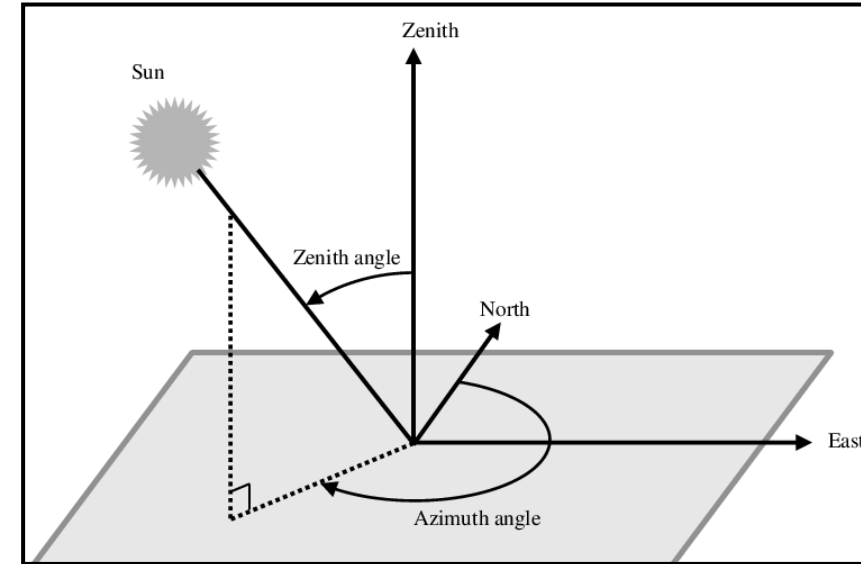
# Определение направления ШАЛ

Направление ШАЛ задается двумя углами: **зенит** (отклонение от нормали к плоскости) и **азимут** (отклонение от направления на север).

**Зенит** находится в диапазоне 0-50 градусов, в то время как **азимут** в диапазоне 0-360.

В попытках натренировать модель предсказывать азимут, метрики измерения ошибки будут давать ложные показания.

К примеру, мы предсказали угол в 355 градусов по азимуту, в то время как правильный ответ был 5 градусов. Ошибка в данном случае – 10 градусов, а не 350, как посчитали бы стандартные метрики.



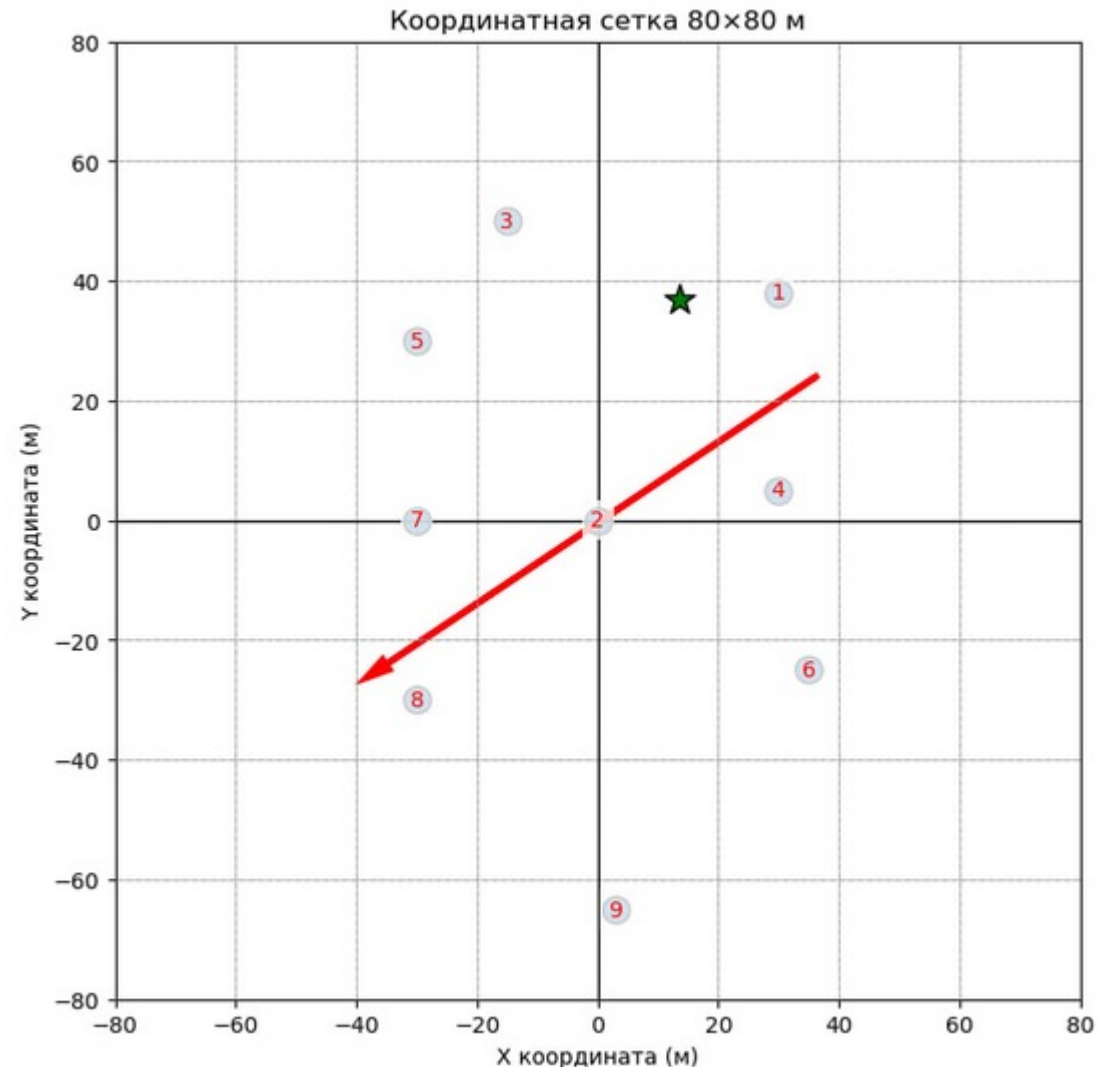
Было принято решение переводить угол азимута в радианы, а потом в синус угла, с целью предотвращения данной неточности

# Определение направления ШАЛ

**Гипотеза:** основываясь на времени детекции станций в каждом кластере можно выяснить путь передвижения ливня.

Для проверки этой гипотезы было решено воспользоваться визуализацией. Каждому кластеру назначен порядковый номер – порядок получения сигнала.

**Вывод:** далеко не всегда возможно интерпретировать полученные сведения, ввиду неконкретного тренда движения ливня.



# Результаты



Наша модель с **высокой точностью** предсказывает целевые переменные: Power, age, X, Y.

Средняя точность с допустимой ошибкой для: Zenith (зенит)

**Высокая ошибка** на Azimuth (азимут). Вызвана недостаточным количеством признаков, влияющих на данную переменную.

Точность Catboost		
	MAE	R2
Power	0.029	0.995
Age	0.020	0.828

Определение направления		
	MAE	R2
Zenith	4.454	0.746
Azimuth	0.526	0.234

Определение оси (X,Y), Catboost		
	MAE	R2
X	5.952	0.854
Y	7.277	0.914



1. Продолжение анализа признаков из датасета. Выявление наиболее эффективных.
2. Реализовать переход от синуса азимута к градусной мере.
3. Создание **новой модели** визуализации по временному признаку для определения направления ШАЛ.
4. Изучение **большого объема** литературы по ШАЛ для лучшего распознавания зенита направления ШАЛ и азимута.
5. Применение проверенных методик по определению **азимута** на практике с целью выявить наиболее эффективные методы для имплементации.
6. Использование ансамбля (блендинг, стэкинг, бэггинг) моделей для улучшения точности предсказания каждой целевой переменной.
7. Тюнинг гиперпараметров моделей.

**Спасибо за внимание!**  
**Мы готовы ответить на ваши вопросы!**

