



DS-ПОТОК

# Анкета на DS-поток

<https://forms.gle/gBYKmUxUuQwQF5Sv9>

- **Дедлайн 25 июня 23:59**
- Форма доступна только с аккаунта @phystech.edu;
- Форма записывает адрес почты;
- Копия ответов отправляется на вашу почту;
- Ответ можно менять до дедлайна.
- Советуем подать сразу, чтобы не забыть потом :)

Чем **DS-поток**  
отличаются от других мест,  
где обучают Data Science?



Большинство студентов:

“ХОТИМ ИЗУЧАТЬ ТО, ЧТО ВСЕ ИСПОЛЬЗУЮТ!”

“ХОТИМ ИЗУЧАТЬ ТО, ЧТО КРУТО!”

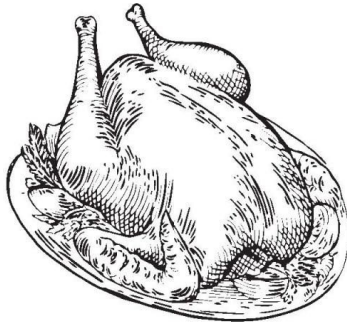
“ХОТИМ ИЗУЧАТЬ ТО, ЧТО СПРАШИВАЮТ НА СОБЕСЕДОВАНИЯХ!”



Образовательные программы  
обычно делятся на две категории



**Чисто теоретические**

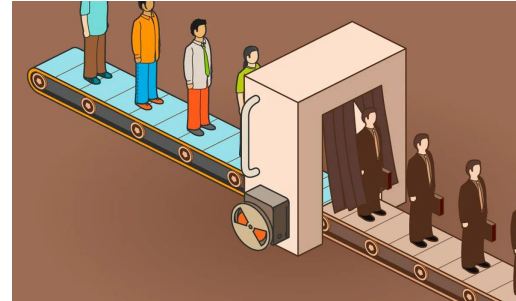


Студент напичкан знаниями  
как фаршированный гусь.

*Аналогия взята из книги Лагутина  
"Наглядная математическая статистика".*



**Чисто прикладное под то,  
что востребовано здесь и сейчас**



Часто это удовлетворяет пожеланиям студента

А что на **DS-потоке**?



# DS-поток

- В меру глубокое обучение теории в непосредственной связке с практикой.
- Обучение разным подходам в анализе данных.
- Обучение широкому кругу задач и методов их решения.
- Практические задачи на каждую тему.
- Высокие нагрузки!!!

## Следствия

- ★ Студент на DS-потоке учится не только на известных задачах, но и учится широко мыслить, искать нестандартные методы решения.
- ★ Выпускник DS-потока имеет прочный фундамент, который не ломается со временем.
- ★ Выпускник DS-потока способен вынести большую нагрузку.

# План обучения



Семестр	DS-поток	Основной поток ПМИ
5	Математическая статистика	Математическая статистика
	Машинное обучение	Машинное обучение
	Практика	Практика по мат. статистике
	Основы прикладной статистики	Курс по выбору x 2
	Курс по выбору	
6	Дискр. случ. процессы и временные ряды	Случайные процессы
	Прикладная статистика и анализ данных	Курс по выбору x 3
	Практика	
	Курс по выбору	
Кафедра АД	Курс ШАД	Методы прикладной статистики
7	Байесовский подход в анализе данных	Курс по выбору x 2
	Практика	
8	Прикладные задачи машинного обучения	Курс по выбору x 2
	Практика	

Полужирным выделены предметы, по которым экзамен.

# 5 семестр

## Основные темы

- Точечные и интервальные оценки параметров, робастные оценки, байесовский подход. Теория наилучших оценок.
- Проверка статистических гипотез, в т.ч. современные подходы к проверке гипотез..
- Линейные модели регрессии и классификации, кросс-валидация.
- Случайный лес, градиентный бустинг.
- Нейронные сети и их особенности, практика на PyTorch.
- Сверточные нейронные сети. Задачи компьютерного зрения: сегментация изображений, детекция объектов на изображениях.

## Особенности

- В этом семестре собрана вся необходимая база, которой должен владеть лучший специалист в области анализа данных.
- Некоторые темы достаточно теоретические, но все они необходимы для грамотного применения методов на практике.
- Один из немногих курсов статистики, в котором теория тесно переплетена с практикой и с машинным обучением.
- Уже в конце первого семестра происходит погружение в более специализированную область - компьютерное зрение.

Осенью 2020 было:

- **28\*** домашних заданий по 1 неделе;
- Более **100** вопросов в боте;
- **8.7** — средняя итоговая оценка студентов.

**Программа предварительная. Из-за общего изменения порядка предметов программа может быть изменена.**

# 6 семестр

## Основные темы

- Рекуррентные нейронные сети. Задачи обработки естественного языка: эмбединги, языковые модели, машинный перевод.
- Методы понижения размерности.
- Кластеризация.
- Анализ зависимостей.
- АВ-тестирование и методы понижения дисперсии.
- Анализ причинно-следственных связей.
- Детектирование аномалий.
- Методы анализа и прогнозирования временных рядов.
- Основы случайных процессов. Последовательный анализ.

## Особенности

- В этом семестре продолжается основной практический материал, которым должен овладеть лучший специалист в области анализа данных.
- Многие темы достаточно практические, но не забываем и про теорию.
- Часть изучаемых методов хоть и не являются широко применяемыми в настоящее время, но способны дополнить построенный фундамент и расширить кругозор. Тем самым слушателям будет проще разрабатывать что-то новое.

Весной 2022 было:

- **38\*** домашних заданий по 1 неделе;
- Более **100** вопросов в боте;
- **8.2** — средняя итоговая оценка студентов (2021).

**Программа предварительная. Из-за общего изменения порядка предметов программа может быть изменена.**

# 7-8 семестры

## Основные темы

- Байесовский подход к оценке параметров и проверке гипотез. Байесовское АВ-тестирование. Выбор модели по Байесу.
- ЕМ-алгоритм в общем виде и его различные применения, вариационный байесовский вывод. Стох. вариационный вывод, разработка моделей в Руго.
- Байесовские нейронные сети, вариационные автоэнкодеры. Генеративно-состязательные сети.
- Распознавание и синтез речи. Сиамские сети. Различные хаки в нейронных сетях.
- Обучение с подкреплением.
- Методы обучения ранжированию и рекомендательные системы.
- Работа с большими данными.
- И другие темы...

## Особенности

- 7 семестр полностью посвящен байесовскому подходу — альтернативному традиционному подходу. Несмотря на то, что он более сложный, в современных задачах часто показывает результаты лучше частотного. Активное исследование байесовского подхода началось относительно недавно.
- 8 семестр содержит различные прикладные задачи машинного обучения, которые во многом покрывают то, с чем придется работать современному специалисту, а также помогают выбрать набор тем для дальнейшего более подробного изучения.
- Большинство тем практические, и позволяют в полной мере применить полученные знания из прошлых семестров.

**+ Работа в командах над проектными заданиями, Kaggle-соревнованиями, а также разбор научных статей.**

За два семестра 2021/2022 было:

- **22** домашних задания по 1+ неделе;
- Более **100** вопросов в боте;
- Студенты суммарно разобрали более **100** научных статей и выполнили **18** проектов.
- **8.5** — средняя итоговая оценка студентов.

# Стало **страшно** от этих чисел?

- Даже если вы сейчас не в топе, если вы готовы учиться, все получится!
- Если что-то не будет получаться, можно научиться, выполняя дорешки после разборов заданий и получая за это баллы.
- А еще у нас есть **КОТИКИ**!



# DS-поток vs. ШАД / кафедра анализа данных

- Иногда возникает ошибочное мнение, что это две взаимоисключающих альтернативы, и что прохождение того и другого несет невыносимую нагрузку.
- DS-поток выстроит четкую и прочную базу как теоретическую, так и практическую, а также расскажет о разных задачах. ШАД / кафедра АД значительно расширят ваши знания по конкретным направлениям.
- Однако, есть и пересечения, которые в основном они касаются ШАДа. В настоящее время в ШАДе курсы машинного обучения и статистики можно закрыть прохождением курсов DS-потoka.
- **Напоминаем, что кафедра анализа данных рекомендует своим студентам пойти на DS-поток.**

# Возьмут ли меня на **DS-поток**, смогу ли я там учиться?

- Если вы готовы большим нагрузкам, готовы разбираться в деталях, тратить достаточное количество времени — да!
- Если вам интересна как теория, так и практика.
- Даже если вы считаете себя недостаточно способными, но готовы много трудиться, приходите к нам.
- Если у вас не очень оценки, но по домашним заданиям и вопросам в боте видно ваши старания, у вас достаточно много шансов.



P.S. Кто-то может сказать, что на DS-потоке зверская нагрузка. Возможно, так было раньше. Сейчас мы провели работу над сокращением количества факультетских курсов для DS-потока без сокращения программы самого DS-потока.

**Ждем вас на DS-потоке!**

