



DS-ПОТОК

Анкета на DS-поток

<https://forms.yandex.ru/cloud/6671a96884227c10810e4e04/>

- **Дедлайн 22 июня 23:59**
- Форма доступна только с аккаунта @phystech.edu;
- Форма записывает адрес почты;
- Копия ответов отправляется на вашу почту;
- Результаты после 15 июля;
- Отказаться можно до 1 августа;
- 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!”

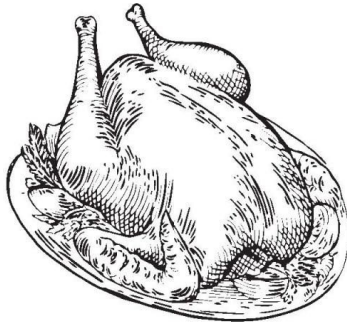
“ХОТИМ ИЗУЧАТЬ ТО, ЧТО СПРАШИВАЮТ НА СОБЕСЕДОВАНИЯХ!”



Образовательные программы обычно делятся на две категории



Чисто теоретические

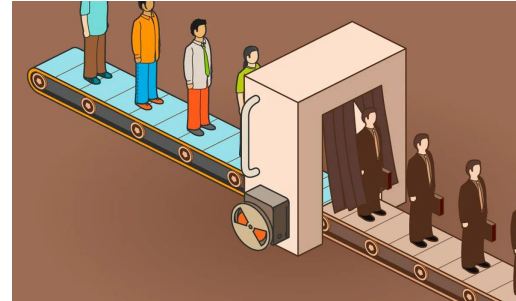


Студент напичкан знаниями
как фаршированный гусь.

*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 сломается со временем.
- ★ Выпускник DS-потока способен вынести большую нагрузку.
- ★ Студенты и выпускники DS-потока успешно проходят собеседования в топовые компании.

DS-поток это про ML или про аналитику?

И то и другое!

- *А если я хочу только Computer Vision?*
- Для этого есть много курсов в интернете :)
- Хороший специалист - тот, у которого достаточно развиты все скилы, но при этом есть преобладающие.
- Мы учим не только продвинутому ML и продвинутой аналитике, но и совмещать эти направления. Например, как самому построить сложную ML-модель и прикрутить к ней продвинутую аналитику, или же как провести крутую аналитику с использованием нейросетей.
- После прохождения DS-потока вы сможете строить сложные решения на пересечении разных областей и предлагать нестандартные подходы.

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 х 2
	Курс по выбору	
6	Дискр. случ. процессы и временные ряды	Случайные процессы
	Глубокое обучение и его приложения	Вычислительная математика
	Прикладная статистика и анализ данных	Параллельные и распределенные вычисления
	Практика	Курс по выбору
Кафедра АД	Курс ШАД	Методы прикладной статистики
7	Байесовский подход в анализе данных	Курс по выбору х 2
	Практика	
8	Прикладные задачи машинного обучения	Курс по выбору х 2
	Практика	

Студенты DS-потока дополнительно могут выбирать курсы по выбору.

Семестр	DS-поток	Основной поток ИВТ
5	Математическая статистика	Математическая статистика
	Машинное обучение	Машинное обучение
	<i>Практика</i>	<i>Практика по мат. статистике</i>
	Основы прикладной статистики	Курс по выбору x 2
	Курс по выбору	
6	Дискр. случ. процессы и временные ряды	Машинное обучение, ч. 2
	Глубокое обучение и его приложения	Вычислительная математика
	Прикладная статистика и анализ данных	Современные компьютерные сети
	<i>Практика</i>	Курс по выбору
Кафедра АД	Курс ШАД	Методы прикладной статистики
7	Байесовский подход в анализе данных	Курс по выбору x 2
	<i>Практика</i>	
8	Прикладные задачи машинного обучения	Курс по выбору x 2
	<i>Практика</i>	

Студенты DS-потока дополнительно могут выбирать курсы по выбору.

Семестр	DS-поток	Основной поток ПМФ
5	Математическая статистика	Математическая статистика
	Машинное обучение	Вычислительная математика
	Практика	Курс по выбору
	Основы прикладной статистики	
6	Дискр. случ. процессы и временные ряды	Случайные процессы
	Глубокое обучение и его приложения	Квантовая механика, ч. 2
	Прикладная статистика и анализ данных	Курс по выбору x 2
	Практика	
Кафедра АД	Курс ШАД	Методы прикладной статистики
7	Байесовский подход в анализе данных	Машинное обучение
	Практика	Курс по выбору x 2
8	Прикладные задачи машинного обучения	Курс по выбору x 2
	Практика	

Студенты DS-потока дополнительно могут выбирать курсы по выбору.

5 семестр

Основные темы

- Точечные и интервальные оценки параметров, робастные оценки. Теория наилучших оценок.
- Проверка статистических гипотез, в т.ч. современные подходы к проверке гипотез.
- Анализ зависимостей признаков.
- Линейные модели регрессии и классификации, валидация моделей.
- Случайный лес, градиентный бустинг, включая подробный анализ современных пакетов.
- Нейронные сети, их особенности и хаки, 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.

В среднем за семестр:

- **35-40*** домашних заданий по 1 неделе;
- **100** вопросов в боте;
- **8-8.5** — средняя итоговая оценка студентов.

Программа предварительная. Мы всегда работаем над поддержанием программ современной и согласованной.

6 семестр

Основные темы

- Рекуррентные нейронные сети. Задачи обработки естественного языка: эмбединги, языковые модели, машинный перевод.
- Современные трансформерные архитектуры в NLP и CV. Техники обучения больших языковых моделей.
- Современные методы АВ-тестирования, методы понижения дисперсии, работа с ratio-метриками.
- Causal inference: анализ причинно-следственных связей или как провести “АВ без АВ”.
- Методы анализа и прогнозирования временных рядов: классические и современные. Немного основ случайных процессов. Последовательный анализ.
- Байесовский подход к оценке параметров и проверке гипотез. Байесовское АВ-тестирование. Методы MCMC для байесовских моделей.

Особенности

- В этом семестре продолжается основной практический материал, которым должен овладеть лучший специалист в области анализа данных.
- Темы семестра можно условно разделить на две группы: последовательные данные и прикладная статистика, причем они логично дополняют друг друга.
- 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.

В среднем за семестр:

- **30-35*** домашних заданий по 1 неделе;
- **2** больших творческих домашних заданий;
- **100** вопросов в боте;
- **8.2-8.7** — средняя итоговая оценка студентов.

Программа предварительная. Мы всегда работаем над поддержанием программ современной и согласованной.

7-8 семестры

Основные темы

- ЕМ-алгоритм в общем виде и его различные применения, вариационный байесовский вывод. Стохастический вариационный вывод, разработка моделей в Руго. Байесовские нейронные сети.
- Генеративное моделирование: вариационные автоэнкодеры, генеративно-состязательные и диффузионные сети. Теория диффузии.
- Распознавание и синтез речи. Сиамские сети. Различные хаки в нейронных сетях.
- Обучение с подкреплением: классические и современные методы и их приложения.
- Методы обучения ранжированию и рекомендательные системы.
- **Нестандартные темы, возникающие на пересечении ранее изученных направлений DS.**

Особенности

- 7 семестр во многом 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. Третий семестр также занимает генеративное моделирование, которое логично дополняет байесовский подход.
- 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. Часть семестра занимает знакомство с нестандартными темами, например, как совместить computer vision и методы причинно-следственного анализа.
- 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.

В среднем за два семестра:

- **22-25** домашних заданий по 1+ неделе;
- **70** вопросов в боте;
- Студенты суммарно разбирают **50** научных статей и выполняют **15** командных проектов.
- **8.3-8.8** — 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, статистики, АВ-тестирования можно заменить при прохождении курсов DS-потока.
- ШАД - отдельная от МФТИ образовательная программа, в то время как кафедра АД в основном состоит из курсов ШАДа, согласовано с программами МФТИ.
- **Напоминаем, что кафедра анализа данных рекомендует своим студентам пойти на DS-поток.**

Бонусы при отборе на кафедру АД для студентов DS-потока

- Если оценки ОТЛ или ХОР за Введение в АД и по DS-поток, то техническое собеседование засчитывается.
- Если попал в топ-5 в рейтинге до ДЗ на DS-поток, то контекст и техническое собеседование засчитываются.

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.



Наличие работы во время учебы в большинстве случаев сказывается отрицательно.



Ждем вас на DS-потоке!

