**Национальный Исследовательский Университет**

**«Высшая школа экономики»**

Факультет компьютерных наук

Концепция

Совместной магистерской программы

**«Статистическая теория обучения»**

**«Statistical Learning Theory»**

Направление подготовки

**01.04.02 Прикладная математика и информатика**

Квалификация: Магистр

Москва 2018

# Общая характеристика

Совместная магистерская программа «Статистическая теория обучения» / «Statistical Learning Theory» (далее – Программа) является основной образовательной программой, реализуемой в соответствии с образовательным стандартом ВО ФГАОУ ВПО НИУ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» по направлению подготовки 01.04.02 Прикладная математика и информатика (квалификация: магистр).

Программа «Статистическая теория обучения» / «Statistical Learning Theory» является исследовательско-прикладной с теоретическими дисциплинами и практико-ориентированными дисциплинами по выбору.

Форма обучения - очная.

Бюджетные места - 20.

Программа реализуется на английском языке.

Нормативный срок освоения – 2 года при очной форме обучения, общая трудоемкость составляет 120 зачетных единиц.

Академический руководитель программы – доктор физико-математических наук, профессор-исследователь кафедры технологий моделирования сложных систем, Спокойный Владимир Григорьевич.
Телефон:+7(495) 772-9590 доб. 22867 email: vspokoiny@hse.ru

Менеджер проекта – кандидат физико-математических наук, старший преподаватель департамента больших данных и информационного поиска, Панов Максим Евгеньевич.
Телефон: +7(495) 772-9590 email: mpanov@hse.ru

Магистерская программа осуществляется совместно с Сколковским институтом науки и технологий. По результатам освоения магистерской программы выдается два диплома.

# Цель и задачи программы

Магистерская программа «Статистическая теория обучения» / «Statistical Learning Theory» направлена на подготовку исследователей в области современной теории статистического обучения — основного раздела теории машинного обучения.

Задачей магистерской программы является знакомство студентов с наиболее актуальными результатами в области статистической теории обучения и ее приложений, необходимым математическим аппаратом, а также развитие у студентов навыков проведения самостоятельных научных исследований в данной области.

Задачами программы являются:

* Развить у студентов навыки разработчика и исследователя в области статистической теории обучения до уровня, соответствующего самым высоким профессиональным требованиям, предъявляемым ведущими отечественными и зарубежными компаниями и научно-образовательными центрами. Выпускники должны иметь широкий профессиональный кругозор и владеть методологическими подходами в данной области, обладать достаточными знаниями и навыками для разработки, а также анализа эффективности новых методов и подходов для решения задач машинного обучения.
* Развить у студентов лидерские качества и навыки применять полученные фундаментальные и прикладные знания для выбора актуальной задачи (которая может быть решена в рамках концепции статистической теории обучения), разработки алгоритмов решения данной задачи, организации процесса разработки, в том числе, с привлечением узконаправленных специалистов в разных областях.
* Научить студентов разрабатывать математические и имитационные модели для решения задач машинного обучения, использовать стандартные компьютерные средства для исследования эффективности их работы, пользоваться профессиональными поисковыми системами, представлять полученные результаты в форме публикаций и докладов, делать экспертные оценки.

# Место Программы в образовательной концепции Факультета компьютерных наук.

Программа «Статистическая теория обучения» образуется на базе программы «Математические методы оптимизации и стохастики» ФКН ВШЭ и программы анализа данных Сколтеха.

Программа рассчитана на подготовку исследователей и разработчиков новых методов на стыке нескольких современных математических дисциплин: статистики, оптимизации, машинного обучения.

# Целевая аудитория

Обучение в магистратуре рассчитано на подготовленных проходить выпускников бакалавриата Факультета компьютерных наук и других факультетов НИУ ВШЭ, математических факультетов иных российских и зарубежных вузов, а также специалисты, получившие образование по соответствующим специальностям и успешно справившихся со вступительными испытаниями. Целевой аудиторией являются абитуриенты с сильной математической подготовкой, нацеленные на построение академической карьеры в области машинного обучения, а также в областях, требующих сильной подготовки в области анализа данных

# Содержание программы

Программа включает в себя один адаптационный курс «Математика для наук о данных», базовые курсы, обязательные для всех студентов, и курсы по выбору. Курсы компонуются по блочному принципу так, чтобы в один день все курсы читались либо в Сколтехе, либо в ВШЭ. Возможно проведение отдельных курсов в ИППИ РАН и НМУ. Кроме того, каждый студент программы участвует в работе как минимум одного исследовательского семинара (рабочей группы), в рамках которого ведется его исследовательская работа. Участие в работе семинара подразумевает обзоры современной литературы, представление своих задач и полученных результатов.

Дополнительно к базовым и курсам по выбору планируется интенсивная программа привлечения ведущих мировых ученых для чтения спецкурсов по теории машинного обучения и смежным дисциплинам математики.

В рамках программы планируется организация миниконференций с приглашением зарубежных ученых и с возможностью представления полученных результатов в форме докладов или постеров. Студенты программы получат возможность участия в ежегодных школах-конференциях по информационным технологиям и по современным методам оптимизации. Планируется программа стажировок студентов в ведущих европейских центрах: Берлин, Гейдельберг, Париж, Лувен-ля-Нев, Гренобль, Хайфа и других.

Все студенты подготавливают курсовую работу в конце первого года и защищают магистерскую диссертацию в конце всего срока обучения.

# Конкурентоспособность программы

**Описание конкурентов:**

В Москве по данному или близкому направлению существует несколько магистерских программ. Среди них можно отметить направление «Анализ данных» в Школе анализа данных компании Яндекс, программу «Науки о данных» на факультете компьютерных наук ВШЭ, программа «Анализ данных» ФУПМ МФТИ. Также стоит упомянуть такие магистерские программы, как «Интеллектуальный анализ больших данных» и «Статистический анализ и прогнозирование рисков» ВМК МГУ, а также «Математические и компьютерные методы анализа» МехМата МГУ.

Также специалистов в области машинного обучения готовят в Компьютерном центре при ПОМИ РАН в Санкт-Петербурге, на магистерской программе «Управление большими данными» в университете Иннополис, а также по программе «Master Programmes in Applied Mathematics and Stochastics» в Новосибирском государственном университете.

Магистерские программы по машинному обучению существуют во многих ведущих мировых научных центрах.

**Конкурентоспособность программы:**

По сравнению со всеми схожими программами, программа “Статистическая теория обучения” выделяется существенно более сильным уровнем подготовки в следующих областях: математическая статистика, выпуклая оптимизация.

Данные области являются ключевыми при разработке новых методов статистического обучения, а также для понимания данной области знания. Это также составляет ключевое отличие от программы подготовки по направлению “Анализ данных” в Школе анализа данных Яндекс. В ШАД дается сильнее подготовка в области дискретной оптимизации, а также существенно большее внимание уделяется развитию навыков программирования.

Программа рассчитана в первую очередь на подготовку исследователей в современной теории машинного обучения и основное предполагаемое направление развития карьеры выпускников – академическое. Уровень предложенной программы “Статистическая теория обучения” во многом соответствует уровню подготовки первых лет на аналогичных Ph.D. программах зарубежных ВУЗов. Это обстоятельство дает выпускникам программы конкурентное преимущество при поступлении на Ph.D. программы зарубежных университетов. Также стоит учесть, что необходимые навыки программирования могут быть освоены благодаря большому количеству доступных онлайн-курсов, в то время, как аналогов самых продвинутых курсов программы в общем доступе фактически нет.

# Актуальность и значимость

Разработки в области машинного обучения имеют большую историю и, в том или ином виде, начались более полувека назад. Очередной бум в этой области начался в середине первой декады XXI-века и продолжается до сих пор. Сейчас разработки в области машинного обучения являются основным двигателем в областях информационного поиска, индивидуальной и диагностической медицины, финансах, разработки интеллектуальных транспортных систем, разработки рекомендательных сервисов и многих других. Экспоненциальный рост массивов доступных данных и столь же значительный рост сложности решаемых задач привели к насущной необходимости разработки новых алгоритмов.

Машинное обучение - это область исследования, связанная с разработкой средств и алгоритмов для анализа больших массивов данных с последующим использованием результатов для адаптивного управления и принятия решений. Теоретические основы машинного обучения лежат на пересечении математики и информатики.

Большая часть математической составляющей машинного обучения объединена в направление «Статистическая теория обучения», которая и стала основой для данной магистерской программы.

Разработка новых методов и алгоритмов машинного обучения зачастую невозможна без глубокого знания таких областей, как теория вероятностей, математическая статистика, оптимизация и комбинаторика, а также без глубокого знания современных алгоритмов. Решение сложных задач в анализе данных, возникающих во многих приложениях, требует активной работы на стыке всех этих дисциплин. Цель данной программы – подготовить высококвалифицированных специалистов, владеющих основными методами и подходами, используемые при разработке теории и алгоритмов машинного обучения, имеющими глубокое понимание математики, на которой данные методы и подходы основываются, а также привлечь студентов к активной исследовательской работе в данной области уже на уровне магистратуры. Одна из важных особенностей данной программы – ее научно-образовательная направленность. Студенты программы будут привлекаться к научной работе в одной из областей современной теории машинного обучения с самого начала магистратуры в рамках научно-исследовательских семинаров.

Программа «Статистическая теория обучения» призвана, помимо прочего, дать студентам необходимые знания и навыки и вовлечь их в проведение самостоятельных научных исследований в области современного машинного обучения. Это делает данную магистерскую программу не только значимой, но и актуальной.

# Обоснование потребности в магистрах

Статистическая теория обучения является основным теоретическим разделом машинного обучения. Эта область является одним из наиболее активно развивающихся разделов Computer Science, крайне востребованным и в других дисциплинах (медицина, физика, экономика и финансы, транспортное планирование и многие другие). Все это вкупе делает выпускников магистерской программы крайне востребованными на академическом рынке.

Специалисты в области статистического обучения также крайне востребованы во всех областях, где возникают задачи анализа данных и предсказательного моделирования.

 Среди российских компаний в выпускниках программы могут быть заинтересованы:

* Крупные интернет-компании (Яндекс, Мэйл.ру, Озон и другие),
* Телеком-компании (Мегафон, Билайн, МТС),
* Банки и финансовые организации (Сбербанк-Технологии, Тиньков Кредитные Системы, Альфа-Банк и другие),
* Крупные и средние компании в области ритейла (Эльдорадо, Техносила, Перекресток и другие),
* Другие компании и стартапы, в которых возникает необходимость решать задачи предсказательного моделирования.

Среди зарубежных компаний специалисты в области статистического обучения крайне востребованы в таких компаниях, как

* Google, Amazon.com, Netflix, Facebook и другие интернет-компании,
* WorldQuant и другие частные компании в области алгоритмического трейдинга,
* Huawei, Alcatel и другие разработчики телекоммуникационного оборудования.

Востребованность выпускников данной магистерской программы подтверждается как рейтингами специальностей, так и наличием большого количества открытых вакансий с высоким уровнем заработной платы.

При этом общее число ежегодно подготовленных специалистов в данной области в России существенно ниже спроса, который продолжает расти. Это во многом объясняется высокими требованиями к математической подготовке специалистов.

# Требования к выпускникам магистерской программы

Программа ориентирована на подготовку специалистов в области статистической теории обучения. Выпускники образовательной программы должны обладать навыками разработчика и исследователя в данной области и соответствовать самым высоким профессиональным требованиям, предъявляемым ведущими отечественными и зарубежными компаниями и научно-образовательными центрами. Выпускники должны иметь широкий профессиональный кругозор и владеть методологическими подходами в данной области, обладать достаточными знаниями и навыками для разработки, а также анализа эффективности новых методов и подходов для решения задач машинного обучения.

Выпускники должны быть способны проявлять лидерские качества и применять полученные фундаментальные и прикладные знания для выбора актуальной задачи (которая может быть решена в рамках концепции статистической теории обучения), разработки алгоритмов решения данной задачи, организации процесса разработки, в том числе, с привлечением узконаправленных специалистов в разных областях.

Выпускники смогут разрабатывать математические и имитационные модели для решения задач машинного обучения, использовать стандартные компьютерные средства для исследования эффективности их работы, пользоваться профессиональными поисковыми системами, представлять полученные результаты в форме публикаций и докладов, делать экспертные оценки.

#  Требования к абитуриентам и условия поступления

Магистерская программа «Статистическая теория обучения» в соответствии с образовательным стандартом НИУ ВШЭ является «до-опытной», т.е. не предъявляющей к поступающим требования наличия опыта практической работы.

Поступающие на магистерскую программу «Статистическая теория обучения» должны иметь высшее образование (степень бакалавра и/или специалиста), подтвержденное документом государственного образца.

Для успешного обучения по Программе необходимо на момент поступления владеть английским языком в достаточной степени, чтобы осваивать большие объемы научной, технической и учебной литературы, а также усваивать лекции, читаемые на английском языке.

Для успешного обучения по Программе необходимо на момент поступления обладать достаточными знаниями в области:

* Математического анализа,
* Обыкновенных дифференциальных уравнений и уравнения в частных производных,
* Теории вероятности,
* Линейной алгебры,
* Дискретной математики,
* Программирования (C++/Python).

Прием на магистерскую программу осуществляется путем конкурсного отбора на основе результатов вступительных испытаний. Конкурсный отбор проводится Приемной комиссией на основании заключения экзаменационной комиссии, создаваемой на факультете компьютерных наук.

Для подтверждения необходимого уровня владения английским языком, кандидаты представляют международный сертификат экзамена по английскому языку САЕ (Certificate of Advanced English); ВЕС Higher (Business English Certificate); IELTS с оценкой не ниже 6, TOEFL IBT (Internet Based) от 80 баллов (рассчитывается от 0 до 120) или PBT (Paper Based) – от 500 баллов (рассчитывается от 310 до 677). В случае отсутствия международного сертификата кандидат сдает квалификационный экзамен по английскому языку.

#  Структура учебного плана.

Формат программы предусматривает прочтение в рамках учебного плана магистратуры ряда обязательных курсов, а также ряда курсов по выбору студента.

В число читаемых курсов входят базовые курсы и продвинутые по оптимизации, стохастике, статистике и машинному обучению.

Вариативная часть содержит ряд специальных продвинутых курсов по отдельным разделам статистической теории обучения, а также ряд курсов, ориентированных на приложения теории статистического обучения к решению прикладных задач.

Все курсы программы читаются в первые полтора года обучения. Наибольшая интенсивность приходится на 3-4 модули первого года обучения и 1-2 модули второго года обучения.

В рамках программы предполагается участие студентов в научно-исследовательском семинаре.

# Концепция научно-исследовательского семинара.

Предлагается использовать концепцию рабочей группы (научного семинара). Каждый студент программы «Статистическая теория обучения» активно в работе одной или нескольких групп. Каждая группа фокусируется на одном из направлений статистической теории обучения. На данный момент действуют следующие научные группы:

* Задачи классификации и кластеризации (clustering, network clustering, multi-class learning, semi-supervised learning)
* Структурные методы в теории машинного обучения (Structural Learning),
* Глубинное обучение (Deep Learning),
* Современная оптимизация в машинном обучении,
* Задачи выбора модели (Model Selection),
* Байесовские методы в теории машинного обучения (Bayesian methods in statistical Machine Learning)
* Методы теории оптимального транспорта в машинном обучении.

Каждая группа имеет проводит свой научно-исследовательский семинар, в рамках которого разбираются новые результаты и актуальные задачи в соответствующей области. Студентам предлагается сделать разбор научных статей по тематике семинара и доклад. Также на семинарах проходят обсуждения научных работ студентов. Предполагается, что темы выпускных квалификационных работ студентов связаны с темой хотя бы одного семинара (его они выбирают в качестве обязательного).

Семинары ведутся ведущими преподавателями программы.

К участию в работе семинара также регулярно приглашаются внешние эксперты по соответствующим направлениям.

Среди указанных семинаров, семинар «Deep Learning» в связи с популярностью среди студентов других магистерских программ получил статус факультетского.

# Обеспеченность профессорско-преподавательским составом

В реализации программы будут участвовать:

* Кабатянский Григорий Анатольевич, д.ф.-м.н.
* Нестеров Юрий Евгеньевич, д.ф.-м.н., проф.
* Беломестный Денис Витальевич, к.ф.-м.н., профессор
* Спокойный Владимир Григорьевич, к.ф.-м.н., профессор
* Бернштейн Александр Владимирович, д.ф.-м.н., проф.
* Оселедец Иван Валерьевич, д.ф.-м.н.
* Соболевский Андрей Николаевич, д.ф.-м.н.
* Чертков Михаил, к.ф.-м.н., PhD
* Наумов Алексей Александрович, к.ф.-м.н., PhD
* Лефкимматис Стамматис, PhD
* Бурнаев Евгений Владимирович, к.ф.-м.н., доцент
* Гасников Александр Владимирович, к.ф.-м.н., доцент
* Грязина Елена Николаевна, к.ф.-м.н.
* Шпирко Сергей Валерьевич, к.ф.-м.н.
* Максимов Юрий Владимирович, к.ф.-м.н.
* Панов Максим Евгеньевич, к.ф.-м.н.
* Беляев Михаил Геннадьевич, к.ф.-м.н.
* Фролов А.А., к.ф.-м.н.
* Лемпицкий В.С., к.ф.-м.н.

# Характеристика кадрового потенциала, собственного и привлекаемого со стороны, а также имеющихся и требуемых для реализации ОП ресурсов

Все запланированные курсы ведутся сотрудниками ФКН ВШЭ и Сколтеха.

Существенная часть курсов читается ведущими мировыми исследователями, в том числе:

Со стороны Высшей школы экономики:

* Владимир Григорьевич Спокойный (WIAS, Германия),
* Юрий Евгеньевич Нестеров (CORE, Бельгия),
* Денис Витальевич Беломестный (Duisburg-Essen University, Германия),

Со стороны Сколтеха:

* Михаил Чертков (Los Alamos National Laboratory).

Также предусмотрено проведение курсов ведущими российскими учеными в области машинного обучения, оптимизации, статистики и стохастики: Дмитрий Петрович Ветров (ФКН ВШЭ и ШАД Яндекс), Григорий Анатольевич Кабятянский (Сколтех и ИППИ РАН), Андрей Николаевич Соболевский (ФКН ВШЭ, НМУ и ИППИ РАН), Иван Валерьевич Оселедец (Сколтех и ИВМ РАН), Виктор Сергеевич Лемпицкий (Сколтех и ШАД Яндекс).

Планируется продолжить существующую практику приглашения ведущих зарубежных исследователей для чтения мини-курсов.

# Дальнейшие перспективы развития программы.

Предполагается активное развитие и популяризация настоящей программы с целью привлечения наиболее сильных и перспективных студентов высших учебных заведений России и мира. Для достижения поставленной задачи предполагается проведение следующих мероприятий:

* Реклама магистерской программы на организуемых сотрудниками программы мини-курсах, свободных для посещения для студентов других университетов,
* Реклама магистерской программы на организуемой сотрудниками ФКН ВШЭ летней школе «Традиционная молодежная Школа «Управление, информация и оптимизация»».
* Продвижение магистерской программы в рамках маркетинговых активностей ВШЭ по привлечению абитуриентов,
* Использование каналов привлечения Сколтеха.

Для дальнейшего развития магистерской программы предполагается активное взаимодействие с международными научно-образовательными центрами в области машинного обучения и смежных направлениях. В том числе, в рамках программы двойных дипломов (Университет Фурье, Франция), а также приглашения ведущих мировых исследователей для чтения мини-курсов (Stanford, Georgia Tech, ETH Zurich и другие).