

**Национальный исследовательский университет
«Высшая школа экономики»**

КОНЦЕПЦИЯ

**Проекта Data Culture по формированию у студентов НИУ ВШЭ
компетенций по Data Science посредством интеграции в учебные
планы образовательных программ бакалавриата
соответствующих элементов**

Москва 2017

Факультет компьютерных наук

Проект Data Culture (далее – Проект) направлен на усиление образовательных программ бакалавриата НИУ ВШЭ (далее – ОП) в части формирования знаний и навыков студентов по современным подходам к сбору больших объемов данных, их анализу статистическими методами в применении к практическим задачам, возникающим в различных предметных областях, а также развитие у студентов понимания текущих возможностей и ограничений методов машинного обучения и искусственного интеллекта.

Актуальность и значимость Проекта

Анализ данных – область математики и информатики, занимающаяся построением и исследованием математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных данных. Эта область включает в себя процессы сбора, структурирования и моделирования данных, их исследования и фильтрации, с целью извлечения полезной информации и принятия решений. Математические методы построения моделей на основе данных объединяют в науку, называемую машинным обучением. Совокупность наук, направленных на методы и технологии работы с данными, называют Data Science.

В связи со стремительным развитием информационных технологий за последние десятилетия данные естественным образом стали накапливаться в цифровом виде, и область анализа данных стала активно развиваться. К настоящему моменту накопленные данные достигли таких объемов, что появилась возможность извлекать из них пользу: находить в данных скрытые зависимости, с их помощью прогнозировать новые результаты и делать рекомендации, позволяющие оптимизировать различные процессы и затраты ресурсов. Методы анализа данных и машинного обучения активно развиваются и находят свое применение не только в экономике, физике, но и в социальных науках, журналистике, лингвистике, юриспруденции, политологии и гуманитарных науках – практически повсеместно.

С помощью методов машинного обучения и анализа данных могут быть решены, например, задачи обнаружения спама, оптимизации закупок и маркетинговых кампаний в интернет-коммерции, распознавания людей на фотографии, создания новых лекарств и предсказания эффектов от них, конструирования беспилотных автомобилей, персонализированных рекомендаций фильмов в социальных сетях или товаров в онлайн-магазинах, распознавания подлинности произведений искусства, прогнозирования динамики цен финансовых инструментов, выявления случаев мошенничества с кредитными картами и страховыми продуктами, формирования по результатам социологических опросов группы общественных проблем, вызывающих схожую реакцию у общества, и многие другие.

Вследствие бурного роста Data Science, специалистов, обладающих лишь компетенциями по анализу данных, становится недостаточно: на данный момент понимание возможностей и ограничений машинного обучения требуется не только от специалистов в области Data Science, но и от заказчиков, менеджеров, специалистов в предметной области – ведь именно они формулируют задачу, собирают и предоставляют данные, выбирают критерии качества решения. Если эти шаги выполнены некорректно, то и задача вряд ли

будет решена. Кроме того, данные в различных областях обладают особенными характеристиками, знание которых позволяет правильно интерпретировать получаемые результаты. Таким образом, возникает потребность в обучении основам анализа данных и программирования специалистов из областей, не связанных напрямую с математикой, информатикой и компьютерными науками. *В этой связи актуальным видится наращивание на имеющуюся в ОП НИУ ВШЭ базу составляющей, позволяющей формировать у студентов всех направлений подготовки компетенции, связанные с Data Science, что позволит им в перспективе быстро и эффективно интегрироваться в решение профессиональных задач на стыке предметных полей и компьютерных технологий, которые сегодня являются передовыми, но уже в ближайшей перспективе станут привычной практикой.*

Несмотря на то, что в России аналогичных масштабных проектов, охватывающих целые университеты и различные направления подготовки, в данный момент не существует, достройка программ в этом направлении так или иначе ведется. Более того, в отдельных университетах развиваются междисциплинарные программы вне направлений, связанных с компьютерными науками, с мощным блоком по анализу данных.

В качестве примера можно привести область Digital Humanities – цифровых гуманитарных исследований. В российском образовательном пространстве данное направление представлено всего несколькими центрами: в виде магистерских программ в Сибирском федеральном университете, Томском Государственном университете и Пермском университете, а также частично в РАНХиГС, в рамках программы Public History.

Другой пример сферы, по которой появляются различные образовательные инициативы и проекты – финтех (область, совмещающая в себе макроэкономику, финансы, управление бизнес-процессами, а также анализ больших массивов данных, разработку математических моделей и предиктивную аналитику). В России образовательные инициативы, связанные с финтехом, представлены в виде части курсов магистерских программ лишь в нескольких университетах: НИУ ВШЭ, МФТИ, МГУ, университете Иннополис и НГУ. В 2017 г. на ФКН НИУ ВШЭ открылась магистерская программа «Финансовые технологии и анализ данных».

В ведущих зарубежных университетах мира наблюдается тенденция по разработке курсов и проектов, связанных с анализом данных, однако, в настоящее время эта тенденция также не является массовой. Отдельные курсы в разных формах и объемах, образовательные программы (в основном, это программы магистратуры) и рабочие или научные группы представлены, например, в следующих университетах и профессиональных сообществах:

- MIT (США): онлайн курс Data Analysis for Social Scientists
- Стэнфорд (США): Data Science трек для студентов направления Political Science
- Columbia University New York – Data Science Institute, в составе которого такие центры, как Data, Media & Society, Financial & Business Analytics, Smart Cities, Computational Social Science и др.
- Tilburg and Eindhoven: joint Bachelor program on Data Science, в которой помимо технической подготовки изучаются правовые, социальные, этические и предпринимательские аспекты, которые относятся к Data Science
- Университет Амстердама (Нидерланды): Master's Psychology: Behavioural Data Science

(track), Radboud University (Нидерланды): Master's programme in Behavioural Science (Research)

- Section of Science & Technology Law: Big Data в Американской ассоциации юристов (American Bar Association)

В связи с возрастающей ценностью и необходимостью навыков в области Data Science в разных отраслях, становится ясно, что со временем как подобных проектов в различных областях, так и практик «достройки» имеющихся ОП элементами, формирующими соответствующие компетенции, будет становиться на рынке образовательных услуг все больше.

Основным экспертным центром Проекта выступает факультет компьютерных наук НИУ ВШЭ (ФКН), на площадках факультета бизнеса и менеджмента и МИЭМ используются ресурсы Школы Бизнес-информатики и МИЭМ соответственно. С момента создания в 2014 году факультета компьютерных наук его коллектив вовлечен в работу по развитию в НИУ ВШЭ направлений, связанных с программированием, анализом данных, машинным обучением. При этом и для ФКН этот проект позволит нарастить кадровую базу и компетенции по работе в неспецифичных для компьютерных наук предметных полях, что усилит его конкурентоспособность.

Особенность Data Science заключается в том, что востребованность и квалификация специалиста зависят не только от его академической подготовки, но и от наличия опыта решения реальных задач, знания практических особенностей используемых методов. В настоящее время практикующие профессионалы в области программирования и анализа данных, как правило, работают в компаниях и не занимаются научной и преподавательской деятельностью.

Факультету компьютерных наук удастся активно привлекать таких специалистов к созданию и включению их курсов в образовательный процесс, развивать сотрудничество с ведущими компаниями в области анализа данных. Также на факультете существуют научно-исследовательские лаборатории, в которых накапливается опыт работы в предметных областях. Все это дает основания предполагать успешное развитие проекта Data Culture со стороны факультета компьютерных наук, который в том числе уже имеет опыт взаимодействия в этом пространстве с другими факультетами НИУ ВШЭ.

Среди основных результатов работы факультета в направлении развития Data Science на общеуниверситетском уровне выделяются следующие.

1. С 2015 года ОП «Прикладная математика и информатика» факультета компьютерных наук реализует в НИУ ВШЭ майнор «Интеллектуальный анализ данных», включающий 4 семестровых курса: «Введение в программирование», «Введение в анализ данных», «Современные методы машинного обучения» и «Прикладные задачи анализа данных».

В первый год на майнор записалось максимальное число слушателей (300 чел.) – студентов других образовательных программ. В 2016/2017 уч. г. на майнор были записаны 152 студента, что, с учетом изменения процедуры выбора майнора, является максимально

возможным числом слушателей (информация в таблице взята из LMS). Среди них – студенты следующих факультетов:

Факультет	Кол-во слушателей
Факультет экономических наук	36
Факультет Бизнеса и Менеджмента	31
<i>Программная инженерия (ФКН)</i>	23
МИЭМ	27
Факультет гуманитарных наук	11
Факультет мировой экономики и мировой политики	8
Факультет социальных наук	7
Факультет коммуникаций, медиа и дизайна	4
Факультет математики	3
Факультет права	2
Департамент иностранных языков	1

2. Факультетом компьютерных наук по запросу других факультетов были реализованы следующие курсы по программированию и анализу данных:

- одномодульный курс «Машинное обучение» на ОП бакалавриата «Совместная программа по экономике НИУ ВШЭ и РЭШ»;
- семестровые курсы «Поиск, поисковые системы и базы данных в Интернете» и «Открытые данные» на магистерских программах факультета коммуникаций, медиа и дизайна;
- семестровый курс «Программирование для лингвистов» на ОП бакалавриата «Фундаментальная и прикладная лингвистика»;
- а также общеуниверситетские факультативы по программированию и анализу данных.

Кроме того, к моменту формирования концепции проекта, на ФКН поступали запросы по проведению и разработке новых курсов, среди которых:

- «Машинное обучение», факультет экономических наук, ОП «Экономика»;
- «Анализ данных», ф-т МЭиМП, ОП «Международные отношения»;
- годовой курс «Программирование», факультет математики, ОП «Математика» и «Совместный бакалавриат ВШЭ и ЦПМ»;
- «Машинное обучение», факультет социальных наук, ОП «Социология».

3. С начала 2016 года на факультете компьютерных наук реализуются образовательные программы дополнительного образования очно-заочной формы обучения. К моменту формирования концепции Проекта количество выпускников программ ДПО ФКН составляло 130 человек, продолжали проходить обучение 100 слушателей, причем 36 из них являлись выпускниками НИУ ВШЭ. Среди пользующихся наибольшим спросом

выделялись курсы «Введение в машинное обучение и майнинг данных», «Машинное обучение и майнинг данных», «Text Mining», «Python как первый язык программирования», а также программа «Практический анализ данных и машинное обучение», состоящая из нескольких курсов.

Таким образом, как внутренняя, так и внешняя по отношению к НИУ ВШЭ среда также формирует определенный запрос на усиление предоставляемых образовательных продуктов элементами, формирующими компетенции по Data Science.

Цели и задачи Проекта

Цели

- Внедрение современных технических и технологических методов и подходов в построении образовательных программ нового поколения как стандарта высшей школы, соответствующего современным вызовам и перспективам развития не только российского образования, но и в целом вектора развития образования в мире;
- Расширение спектра компетенций специалистов каждой программы и адаптация студентов Университета к работе в современном мире, где навыки работы с данными являются неотъемлемой и обязательной составляющей как в академической, так и в бизнес-среде;
- Повышение привлекательности ОП НИУ ВШЭ в глазах абитуриентов, в том числе, поступающих на коммерческие места, благодаря включению в учебные планы современных и востребованных курсов;
- Повышение конкурентоспособности выпускников университета не только на российском, но и на международном рынке, с учетом перспектив усиления влияния технологической среды во всех предметных группах;
- Укрепление позиции прогрессивного и инновационного молодого Университета, предугадывающего мировые тренды, где помимо крепкой фундаментальной базы даются актуальные технологические знания, позволяющие выпускникам успешно влиться в мировое научное и бизнес-сообщество;
- Закрепление лидирующей позиции НИУ ВШЭ среди российских университетов по обучению анализу данных в прикладных задачах из различных областей;
- Развитие и укрепление связи между академической средой и специалистами из различных отраслей, а также создание условий для междисциплинарных проектов как внутри университета, так и с индустрией;
- Расширение возможностей для развития сотрудников ППС (включая повышение квалификации преподавателей в области наук о данных), получение опыта и дополнительная возможность трудоустройства для студентов, аспирантов и выпускников.

Задачи:

- создание информационной среды, формирующей лояльное отношение к внедрению курсов по Data Science в образовательные программы и позволяющей оценить возможности и ограничения, которые несут в себе современные методы машинного обучения и искусственного интеллекта;
- в краткосрочной перспективе (2017/2018 учебный год): гибкое включение в учебные планы части ОП в качестве обязательных или элективных курсов по направлению Data Science;
- в среднесрочной перспективе (с 2018 года и далее): создание условий для наращивания возможностей студентов всех образовательных программ бакалавриата НИУ ВШЭ по обучению анализу данных; обеспечение возможности для каждого студента вне зависимости от направления подготовки изучить курсы по data skills и получить базовые знания и практические навыки в области Data Science в применении к изучаемому им направлению;
- развитие взаимодействия между подразделениями НИУ ВШЭ, с целью организации совместных исследовательских и бизнес проектов в смежных областях.

Для успешного выполнения поставленных задач в проект необходимо вовлечение экспертов по Data Science, специалистов, занимающихся анализом данных в различных отраслях, а также участия руководства НИУ ВШЭ, академических руководителей программ, сотрудников лабораторий НИУ ВШЭ, ППС и партнеров образовательных программ соответствующих направлений. Такое сотрудничество позволит обеспечить современное преподавание курсов по Data Science в различных направлениях и разработку актуальных учебных курсов.

Общая концепция Проекта

Создание условий по освоению студентами навыков в сфере Data Science (или формирование Data Culture) реализуется за счет «достройки» учебных планов (в том числе на уровне индивидуальных учебных планов студентов) и предполагает соответствующий фокус при реализации отдельных их элементов:

- отдельных обязательных дисциплин,
- дисциплин по выбору,
- факультативов,
- майноров¹,
- миноров (при наличии),
- научно-исследовательских, проектно-исследовательских и др. семинаров,
- проектов² и т.д.

¹ Продвинутый уровень владения Data Culture может быть достигнут за счет изучения майнора «Интеллектуальный анализ данных».

² В рамках Проекта при участии экспертов и менторов возможно расширение спектра

Процесс корректировки учебных планов координируется Дирекцией основных образовательных программ, содержательное наполнение элементов учебного плана координируется ФКН³ во взаимодействии с Академическими руководителями программ.

Для студентов всех ОП предлагается линейка курсов, которая может содержать как обязательные, так и вариативные блоки. Обязательным является освоение 1-2 дисциплин (включая тематический НИС по желанию ОП) общим объемом 4-12 кредитов. Их формат и объем освоения фиксируется учебным планом ОП. Наличие дополнительных курсов или их блоков регулируется ОП. К 2018-2019 планируется разработать вариативные траектории для повышения уровня соответствующей подготовки (в том числе в рамках групп программ).

Для студентов каждой ОП при координации ФКН предлагаются рекомендации по возможному наращиванию компетенций по Data Science в дополнение к основной образовательной программе (программами ДПО, курсами MOOC и т.д.). В связи с большим разнообразием и широким спектром ОП НИУ ВШЭ, студенты которых дифференцированы по базовым компетенциям в сфере математики и информатики, предлагаемые решения адаптируются под блоки программ (для студентов с сопоставимыми наборами знаний и навыков в этой области)⁴. Для каждого блока программ предлагается система курсов Data Culture в определенной вилке «сквозного уровня продвинутости» (описание уровней см. в табл. 1). Эти системы курсов определяются спецификой предметных областей и для блоков программ могут иметь общие обязательные курсы и общие элективы для наращивания уровней подготовки.

Таблица 1

Описание уровней курсов по формированию Data Culture у студентов НИУ ВШЭ

Уровень	Компетенции	Применение навыков
Начальный	- Понимание возможностей и ограничений машинного обучения - Понимание применений анализа данных и машинного обучения в предметной области, соответствующей образовательной программе - Навыки постановки задач машинного обучения	- Формулирование задач для профессионалов - Управление проектами в области Data Science
Базовый	- Понимание основных типов и источников данных - Проведение простой аналитики на данных (вычисление статистик, визуализация) - Обучение простых моделей с помощью библиотек для анализа данных	- Первичный анализ данных - Решение стандартных задач, не требующих подготовки данных или доработки типовых моделей

предлагаемых сегодня задач и методов решения.

³ за исключением факультета Бизнеса и Менеджмента и МИЭМ НИУ ВШЭ

⁴ Тем не менее, в целом границы треков проницаемы и по желанию студентов они могут обучаться и не только со своей группой программ.

Уровень	Компетенции	Применение навыков
Продвинутый	- Знание основных методов и моделей машинного обучения - Понимание полного цикла решения задачи анализа данных: подготовка данных, разработка признаков, выбор метрики качества, выбор и обучение модели, валидация модели и т.д. - Владение методами анализа данных, специфичными для конкретной предметной области (анализ социальных сетей, анализ текстов и т.д.)	Самостоятельное решение задач анализа данных, типичных для конкретной предметной области
Профессиональный	- Владение машинным обучением с теоретической и практической стороны - Глубокое знание специфичных разделов анализа данных: прикладная статистика, анализ текстов, обработка сигналов, анализ изображений, нейросетевые методы и глубинное обучение и т.д.	- Работа на позициях Data Scientist в профильных компаниях - Решение задач анализа данных из любых областей при поддержке специалистов по предметной области
Экспертный	- Навыки решения сложных и нестандартных задач анализа данных - Знание последних научных результатов в конкретной области машинного обучения	Научно-исследовательская деятельность, разработка новых методов анализа данных

Распределение программ по уровням выглядит следующим образом. Зеленым отмечен уровень, на который выйдут все студенты данной программы, желтым — когорта студентов (в рамках обязательных занятий).

Факультет	Блоки ОП	Начальный	Базовый	Продвинутый	Профессиональный	Экспертный
ФГН	История					
	История искусств					
	Филология					
	Культурология					
	Философия					
	ИЯ и МК					

Факультет	Блоки ОП	Началь- ный	Базо- вый	Продвинутый	Профес- сиональный	Эксперт- ный
Факультет права	Юриспруденция					
	Юриспруденция: частное право					
ФМЭиМП	Востоковедение					
	Программа двух дипломов НИУ ВШЭ и Лондонского университета «Международные отношения»					
Факультет коммуни- каций, медиа и дизайна	Дизайн					
	и Мода					
ФМЭиМП	Международные отношения					
ФМЭиМП	Мировая экономика					
ФСН	Психология			Выпускники специализации «Когнитивная психология и психофизиология»		
ФСН	Социология			Выпускники специализации «Прикладные методы социологических исследований»		
ФСН	Политология					
ФСН	Государственное и муниципальное управление					

Факультет	Блоки ОП	Началь- ный	Базо- вый	Продвинутый	Профес- сиональный	Эксперт- ный
ФКМД	Медиаком.					
	Журналистика					
	Реклама и связи с общественностью					
ФЭН	Экономика					
	Совместный бакалавриат ВШЭ и РЭШ					
ФЭН	Экономика и статистика					
МИЭФ	Программа двух дипломов по экономике НИУ ВШЭ и Лондонского университета					
Факультет математики	Математика				Включение курсов ПМИ в индивидуальные учебные планы	
	Совместный бакалавриат ВШЭ и ЦПМ					
Факультет физики	Физика					
МИЭМ	Все ОП					
ФБиМ	Все ОП					
ФКН	Программная инженерия					
ФГН	Фундаментальная и прикладная лингвистика				Магистерская программа «Компьютерная лингвистика»	Аспирантская школа и дополнит. курсы по современным направлениям анализа данных

Факультет	Блоки ОП	Началь- ный	Базо- вый	Продвину- тый	Профес- сиональный	Эксперт- ный
ФКН	Прикладная математика и информатика				Выпускники специализации «Машинное обучение и приложения»	Магистратура («Науки о данных»), аспирантская школа по компьют. наукам

Дорожная карта Проекта

2016-2017
Ключевые мероприятия 1. Анализ ситуации по Data Culture в НИУ ВШЭ Доля дисциплин Data Culture (ИТ навыки, анализ данных, машинное обучение) в РУПах: 14% в целом по НИУ ВШЭ (Москва) 2. Разработка концепции Проекта по интеграции курсов Data Science 3. Разработка линеек курсов для образовательных программ и их блоков (гуманитарные, экономические, социальные и т.д.) 4. Формирование информационного поля, позиционирование Проекта 5. Подготовка и поиск кадров для участия в проекте, обучение сотрудников 6. Включение Проекта в PR материалы в рамках приемной кампании
2017-2018
Целевая модель Интеграция Проекта на части ОП (ориентировочно — 12) бакалавриата Целевой ориентир: более половины (около 50-55% образовательных программ участвуют в проекте Data Culture, с учетом ФКН, фБиМ, МИЭМ)
Ключевые мероприятия 1. Гибкая интеграция курсов по Data Science с дифференциацией по программам. С учетом текущего математического бэкграунда, заинтересованности руководства программ предполагается два пути: - на части программ — построение полной линейки, начиная с 1 курса, для студентов

- нового набора
 - на части программ — частичное достраивание ОП курсами Data Culture за счет добавления и/или расширения пула курсов по выбору или MOOC
2. Вариативная интеграция курсов ДПО по Data Science, разработка общеуниверситетских курсов по анализу данных от вендоров (SAS, SAP и др.)

2018-2019

Целевая модель

- 100% участие образовательных программ бакалавриата НИУ ВШЭ (Москва) в Проекте
- Для каждой из программ выстроена и представлена к реализации одна из линеек, в зависимости от релевантного уровня:
 - 1) Начальный
 - 2) Базовый
 - 3) Продвинутый
 - 4) Профессиональный
 - 5) Экспертный
- Формирование совместных научных групп, включающих сотрудников ФКН и других факультетов НИУ ВШЭ, с целью проведения исследований по направлениям, связанным с тематикой непрофильных факультетов и анализом данных;
- Развитие совместных коммерческих проектов по кросс-дисциплинарным темам (например, экономические предсказания с использованием продвинутых методов машинного обучения).

Ключевые мероприятия

1. Интеграция Проекта в полном масштабе:
 - встраивание линеек курсов соответствующего уровня на программах, не охваченных в 2017-2018гг.,
 - достраивание линеек курсов для программ, охваченных ранее частично, с наличием обязательного компонента Data Culture
2. Развитие межфакультетских и кроссдисциплинарных проектов

Кадровый состав Проекта

В деятельность Проекта предполагается вовлечение преподавательского состава

- из индустрии: участники сообществ по анализу данных (например, Open Data Science), участники тематических мероприятий по анализу данных, проводимых в IT-компаниях (например, Data&Science, DataFest, Moscow Data Science meetup), сотрудники компаний;
- из академической среды: ППС ФКН и других факультетов НИУ ВШЭ, сотрудники департамента математики факультета экономических наук и общеуниверситетской кафедры высшей математики, студенты магистратуры, аспиранты и выпускники ФКН и других факультетов НИУ ВШЭ, выпускники Школы анализа данных Яндекса, партнеры образовательных программ, сотрудники лабораторий НИУ ВШЭ.

Команда проекта:

- **Ольга Подольская** – руководитель Проекта со стороны факультета компьютерных наук;
- **Евгений Соколов**, заместитель руководителя департамента больших данных и информационного поиска, соруководитель специализации «Машинное обучение и приложения» на ПМИ ФКН, преподаватель на майноре «Интеллектуальный анализ данных», соавтор курса «Введение в машинное обучение» и специализации «Машинное обучение и анализ данных» на Coursera, руководитель исследовательской группы по анализу данных в Яндексе – ведущий эксперт;
- **Юлия Лежнина** (заместитель проректора) – координация работы по обновлению образовательных программ бакалавриата;
- **Алексей Соболевский** (заместитель директора Дирекции по связям с общественностью и информационным ресурсам НИУ ВШЭ) – координация межфакультетского взаимодействия и PR Проекта;
- **Алёна Северцева** (менеджер факультета компьютерных наук) – менеджер проекта Data Culture.

Также в качестве экспертов в проекте участвуют и другие специалисты по Data Science, в том числе

- **Дмитрий Ветров**, профессор-исследователь ФКН, руководитель специализации «Машинное обучение и приложения» на ОП «Прикладная математика и информатика» ФКН, основатель и руководитель научной группы байесовских методов, разработчик уникальных для России курсов по байесовским методам и графическим моделям, которые читаются в Школе анализа данных Яндекса, Сколтехе, факультете ВМК МГУ;
- **Михаил Левин**, директор по машинному интеллекту Яндекс.Маркета, руководитель специализации больших данных Школы анализа данных, соавтор специализации Data Structures and Algorithms от НИУ ВШЭ и UCSD на Coursera.