

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
радиотехники и компьютерных
технологий**

Д.А. Гаврилов

Рабочая программа дисциплины (модуля)

по дисциплине:	Разработка цифровых продуктов на основе искусственного интеллекта в промышленности
по направлению:	Прикладные математика и физика
профиль подготовки:	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра физико-технической информатики
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Программу составил: И.В. Максимов, ассистент

Программа обсуждена на заседании кафедры физико-технической информатики 10.06.2021

Аннотация

В курсе рассматриваются инструменты, методы, методологии, необходимые для разработки цифровых продуктов на основе ИИ в промышленности.

На лекциях и семинарах разбираются реальные промышленные задачи и кейсы. Обучающиеся знакомятся с типами Data Science (DS) задач, основами языка программирования Python, основами контроля версий кода (git). Изучаются основные библиотеки для работы с данными (pandas, numpy, matplotlib), методологии и инструменты для ведения DS проектов, библиотеки машинного обучения, проектирование и разработка интерфейсов для моделей машинного обучения, web API.

В течение всего курса обучающиеся будут работать в группах над реальными задачами, защита которых пройдет в конце курса обучения.

1. Цели и задачи

Цель дисциплины

Целью дисциплины является повышение общей осведомленности о современных подходах по разработке и управлению проектами и продуктами на основе ИИ. При этом курс направлен на получение практических знаний для решения реальных промышленных задач, актуальных на производствах Росатома. Также курс направлен на формирование ключевых компетенций в областях:

- Применение технологии ИИ в промышленных задачах
- Ведение проектов по созданию решений и продуктов на основе ИИ с точки зрения Data science специалистов
- Технологии ИИ, требуемые для решения реальных промышленных задач

Задачи дисциплины

- Сформировать представление о процессе проектирования и разработки продуктов на основе ИИ
- Сформировать представление о процессе ведения проектов по разработке продуктов на основе ИИ
- Сформировать представление о командных принципах работы и инструментах разработки
- Сформировать понимание спектра задач, решаемых ИИ в промышленности
- Сформировать понимание базовых навыков и технологий, требуемых для решения задач с помощью ИИ
- Получить навыки ведения проектов, которые основаны на применении ИИ
- Получить навыки применения ИИ для создания цифровых продуктов для промышленных приложений

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук и использовать их в	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки

естественных наук, и использовать их в профессиональной деятельности	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- Направления и кейсы применения ИИ в промышленности
- Типы Data Science (DS) задач.
- Модели машинного обучения и бейзлайновые модели для различных задач.
- Виды и модальности данных в задачах машинного обучения: структурированные таблицы, тексты, изображения, звук и тд. Признаки.
- DS метрики для различных типов DS задач.
- Типовую структуру DS-проекта.
- Основы языка программирования Python. Типы данных, циклы, условия, функции, классы.
- Принципы сборки python окружения, установки необходимые пакеты, запуска jupyter-lab.
- Основы контроля версий кода (git).
- Основные библиотеки для работы с данными (pandas, numpy, scipy).
- Основные библиотеки для визуализации данных (matplotlib, seaborn).
- Основные методологии ведения DS-проектов (CRISP-DM, TDSP, LeanDS).
- Основные этапы реализации DS-проектов и критерии перехода между этапами.
- Принципы формирования и заполнения артефактов: User story map, AI canvas, чек-лист по проекту, карточка датасета.
- Типовую структуру отчета о разведочном анализе данных, а также цели и задачи его проведения.
- Принципы декомпозиции бизнес- и DS-гипотез.
- Инструменты task tracking, kanban на примере trello, принципы формирование доски проекта.
- Основы процесса моделирования (feature extraction, train/test split, cross validation, hyper parameters tuning, auttml).
- Библиотеку sklearn и другие библиотеки с методами и моделями машинного обучения.
- Мониторинг моделей. Инструменты визуализации и создания дашбордов (Grafana, Tableau).
- Разработка приложений с помощью фреймворка Flask.
- Принципы и средства создания интерфейсов для моделей машинного обучения, web API.

уметь:

Применять на практике технологии ИИ для решения реальных промышленных задач, а также вести процесс создания продуктов и решения задач с точки зрения DS-специалистов. Также обучающиеся должны уметь выбирать ставить DS-задачи, выбирать DS-метрики, собирать python окружения для конкретных DS-проектов, пользоваться библиотеками для работы с данными и библиотеками с методами и моделями машинного обучения, заполнять User story map, AI canvas, чек-лист по проекту, проводить разведочный анализ данных, декомпонировать бизнес- и DS-гипотезы, пользоваться инструментами трэкинга задач, создавать приложения и интерфейсы для моделей машинного обучения.

владеть:

- Основными принципами ведения DS-проектов на основе популярных методологий.
- Основными технологиями, инструментами, фреймворками и библиотеками машинного обучения и анализа данных, необходимых для решения промышленных DS-задач.
- Средствами создания приложений, интерфейсов для моделей машинного обучения, web API и мониторинга метрик.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в ИИ и направления применения ИИ в промышленности	4	4		4
2	Бизнес анализ в проектах с ИИ	4	2		2
3	Разведочный анализ данных (EDA)	2	4		4
4	Создание моделей машинного обучения	3	3		3
5	Создание продукта	2	2		2
Итого часов		15	15		15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 7 (Осенний)

1. Введение в ИИ и направления применения ИИ в промышленности

Актуальность DS и ИИ. Кейсы. Стратегия. Типы Data Science (DS) задач. DS задачи: классификация, регрессия, кластеризация, recsys, поиск аномалий. Модели. Виды и модальности данных в задачах машинного обучения: структурированные таблицы, тексты, изображения, звук и тд. Признаки. DS метрики. Основы языка программирования Python. Типы данных, циклы, условия, функции, классы. Основы контроля версий кода (git). Типовая структура DS-проекта. Основные библиотеки для работы с данными (pandas, numpy, scipy). Основные библиотеки для визуализации данных (matplotlib, seaborn). Основные методологии ведения DS-проектов (CRISP-DM, TDSP, LeanDS). Основные этапы реализации DS-проектов. Критерии перехода между этапами.

2. Бизнес анализ в проектах с ИИ

Принципы формирования и заполнения артефактов: User story map, AI canvas, чек-лист по проекту. Предпроектная проработка кейса и гипотез: метрики, связи бизнес и DS метрик, данные, валидация.

3. Разведочный анализ данных (EDA)

Мотивация проводить EDA, структура, примеры. Формирование бэклога из дата/метод гипотез. Декомпозиция гипотез. Обзор инструментов task tracking, введение в kanban на примере trello, формирование доски.

4. Создание моделей машинного обучения

Основы процесса моделирования (feature extraction, train/test split, cross validation, hyperparameters tuning, automl). Библиотека sklearn. Базовые (baseline) модели.

5. Создание продукта

Создание интерфейса для моделей машинного обучения. Web API.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная меловой доской или White Board, компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

1. Искусственный интеллект: Современный подход [Текст] : [учеб. пособие для вузов] / С. Рассел, П. Норвиг ; [пер. с англ. под ред. К. А. Птицына] .— 2-е изд. — М. : Вильямс, 2007 .— 1408 с.
2. Введение в искусственный интеллект [Текст] : конспект лекций : [для студентов, аспирантов вузов] / Д.В.Смолин .— М. : Физматлит, 2004 .— 208 с.

Дополнительная литература

1. Hastie, T., Tibshirani, R., Friedman, J. The elements of statistical learning: Data Mining, Inference, and Prediction. – Springer, 2009. – 745 pp.
2. Bishop C. M. Pattern Recognition and Machine Learning. – Springer-Verlag New York Inc., 2006.
3. Azevedo, A. and Santos, M. F. (2008); KDD, SEMMA and CRISP-DM: a parallel overview. In Proceedings of the IADIS European Conference on Data Mining 2008, pp 182–185.
4. Introduction to Machine Learning with Python: A Guide for Data Scientists, Sarah Guido, Andreas C. Mueller, O'Reilly Media

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>
2. https://miro.com/app/board/o9J_IVgyZjg=/
3. <https://www.the-modeling-agency.com/crisp-dm.pdf>
4. ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/CRISP_DM.pdf
5. <https://guides.github.com>
6. <https://www.edx.org/course/introduction-to-computer-science-and-programming-7>
<https://realpython.com/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся предполагается использование таких программных средств, как Jupyter notebook, Jupyter lab, открытые библиотеки и фреймворки python, в которых реализованы модели и методы машинного обучения (pandas, numpy, matplotlib, sklearn и др.), Trello, языка программирования python 3.4+.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Успешное освоение курса требует напряженной самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой.

Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, учебной и научной литературе), подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- подготовку к практическим занятиям, регулярное выполнение индивидуальных и групповых домашних заданий.

Промежуточный контроль знаний проводится в виде опросов по теории. Кроме этого в ходе освоения курса студент должен выполнить групповой проект, содержащий несколько взаимосвязанных заданий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра физико-технической информатики
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет

Разработчик: И.В. Максимов, ассистент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

2. Показатели оценивания компетенций

В результате изучения дисциплины «Разработка цифровых продуктов на основе искусственного интеллекта в промышленности» обучающийся должен:

знать:

- Направления и кейсы применения ИИ в промышленности
- Типы Data Science (DS) задач.
- Модели машинного обучения и бейзлайновые модели для различных задач.
- Виды и модальности данных в задачах машинного обучения: структурированные таблицы, тексты, изображения, звук и тд. Признаки.
- DS метрики для различных типов DS задач.
- Типовую структуру DS-проекта.
- Основы языка программирования Python. Типы данных, циклы, условия, функции, классы.
- Принципы сборки python окружения, установки необходимые пакеты, запуска jupyter-lab.
- Основы контроля версий кода (git).
- Основные библиотеки для работы с данными (pandas, numpy, scipy).
- Основные библиотеки для визуализации данных (matplotlib, seaborn).
- Основные методологии ведения DS-проектов (CRISP-DM, TDSP, LeanDS).
- Основные этапы реализации DS-проектов и критерии перехода между этапами.
- Принципы формирования и заполнения артефактов: User story map, AI canvas, чек-лист по проекту, карточка датасета.
- Типовую структуру отчета о разведочном анализе данных, а также цели и задачи его проведения.
- Принципы декомпозиции бизнес- и DS-гипотез.
- Инструменты task tracking, kanban на примере trello, принципы формирование доски проекта.
- Основы процесса моделирования (feature extraction, train/test split, cross validation, hyper parameters tuning, automl).
- Библиотеку sklearn и другие библиотеки с методами и моделями машинного обучения.
- Мониторинг моделей. Инструменты визуализации и создания дашбордов (Grafana, Tableau).
- Разработка приложений с помощью фреймворка Flask.
- Принципы и средства создания интерфейсов для моделей машинного обучения, web API.

уметь:

Применять на практике технологии ИИ для решения реальных промышленных задач, а также вести процесс создания продуктов и решения задач с точки зрения DS-специалистов. Также обучающиеся должны уметь выбирать ставить DS-задачи, выбирать DS-метрики, собирать python окружения для конкретных DS-проектов, пользоваться библиотеками для работы с данными и библиотеками с методами и моделями машинного обучения, заполнять User story map, AI canvas, чек-лист по проекту, проводить разведочный анализ данных, декомпозировать бизнес- и DS-гипотезы, пользоваться инструментами трэкинга задач, создавать приложения и интерфейсы для моделей машинного обучения.

владеть:

- Основными принципами ведения DS-проектов на основе популярных методологий.
- Основными технологиями, инструментами, фреймворками и библиотеками машинного обучения и анализа данных, необходимых для решения промышленных DS-задач.
- Средствами создания приложений, интерфейсов для моделей машинного обучения, web API и мониторинга метрик.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

С целью контроля освоения обучающимися учебного материала проводится короткий (10-15 минут) опрос по материалу прошедших занятий в устной или письменной форме.

Групповые/индивидуальные задания для текущего контроля в рамках курса:

1. Бизнес анализ в рамках кейса
2. Разведочный анализ данных (EDA) для кейса
3. Создание моделей машинного обучения
4. Создание продукта

Текущий контроль включает в себя презентацию результатов бизнес-анализа и разведочного анализа данных, а также вопросы по инструментам для работы DS-специалистов, презентацию результатов работы моделей машинного обучения, API разрабатываемого продукта, а также вопросы по процессу моделирования и разработки решения.

За каждое задание студент получает оценку в соответствии с таблицей критерия оценивания.

Задания студенты выполняют в группе, оценивается вклад каждого члена группы в общий результат.

Задания выполняются в рамках курса на основе кейса (темы), выбор которого происходит в начале курса.

Список кейсов с наборами данных предоставляется в начале курса.

Кейс представляет собой набор данных и формулировку бизнес-проблемы для проработки.

Результаты оценки заданий могут быть учтены в результирующей оценке по курсу.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Перечень контрольных вопросов для сдачи дифференцированного зачета:

1. Введение в ИИ и направления применения ИИ в промышленности

- Типы Data Science (DS) задач. Особенности DS-задач: классификация, регрессия, кластеризация, recsys, поиск аномалий.
- Виды и модальности данных в задачах машинного обучения: структурированные таблицы, тексты, изображения, звук и тд. Признаки.
- DS метрики для разных типов задач.
- Основы контроля версий кода (git). Основные команды.
- Основные методологии ведения DS-проектов (CRISP-DM, TDSP, LeanDS). Основные этапы реализации DS-проектов. Критерии перехода между этапами.

2. Бизнес анализ в проектах с ИИ

- Принципы формирования и заполнения артефакта User story map.
- Принципы формирования и заполнения артефакта AI canvas.
- Предпроектная проработка кейса и гипотез: метрики, связи бизнес- и DS-метрик, данные, валидация.

3. Разведочный анализ данных (EDA)

- DS-гипотезы. Формулирование гипотез. Декомпозиция гипотез. Формирование бэклога из дата/метод гипотез.
- Мотивация проводить EDA, структура, основные разделы.

4. Создание моделей машинного обучения

- Основы процесса моделирования: предварительная обработка данных
- Основы процесса моделирования: train/test split, cross validation
- Основы процесса моделирования: hyperparameters tuning
- Baseline модели для различных DS-задач

5. Создание продукта

- Создание интерфейса для моделей машинного обучения.
- Web API.

6. Практические вопросы:

- Артефакт AI canvas для командного проекта (кейса): принципы заполнения, основные неопределенности, выводы.
- Тип DS-задачи и базовая модель (baseline) в командном проекте (кейсе): причина выбора, результат, выводы, следующие шаги.
- EDA для командного проекта (кейса): основные инсайты, сгенерированные гипотезы, выводы.
- Бизнес-метрика, DS-метрика, декомпозиция бизнес-метрики в DS-метрику в командном проекте (кейсе)
- Бэклог DS-гипотез для командного проекта (кейса): результаты проверки, принцип формирования, отчеты по DS-гипотезам, выводы.
- Подготовка данных в командном проекте (кейсе): основные этапы, используемые методы, выводы, демонстрация финального датасета.

- Приложение для демонстрации результатов решения задачи с помощью модели машинного обучения.

Критерии оценивания

Оценка отлично (10) выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9) выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8) выставляется студенту, показавшему систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7) выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6) выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5) выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4) выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3) выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2) выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1) выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Дифференцированный зачет проводится в устной форме.

При проведении дифференцированного зачета обучающемуся предоставляется 20 минут на подготовку. Опрос обучающегося проводится в течение 30 минут.

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины.

Дифференцированный зачет может проводиться по итогам сдачи заданий и путем организации демо-дня, где учащиеся представляют свои проекты (решенные кейсы).

При проведении дифференцированного зачета группе обучающихся может предоставляться до 30 минут на презентацию проекта с секцией вопросов от преподавателей и аудитории.