

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

**УТВЕРЖДЕНО**  
**Заместитель директора**

**Ю.О. Соболев**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Машинное обучение в задачах продуктовой разработки
<b>по направлению:</b>	Бизнес-информатика
<b>профиль подготовки:</b>	Управление цифровым продуктом
	центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
	центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 12 всего, в том числе:

лекции: 0 час.

семинары: 12 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 78 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

О.А. Культепина, старший методист

В.С. Акульшин, методист

П.И. Ахтямов, преподаватель

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и  
онлайн-образования "Пуск" 05.03.2024

## Аннотация

В процессе изучения курса «Машинное обучение в задачах продуктовой разработки» студенты познакомятся с основными концепциями задачи ML. Слушатели разберут цикл постановки ML-задач и познакомятся с концепцией исследовательского анализа данных, рассмотрят классические задачи решения задач машинного обучения и узнают, как простые модели позволяют узнавать неявные закономерности в данных. Также будут разобраны продвинутое задачи — распознавание изображений и анализ текстов. В практической части студенты разберут, как работают такие сервисы как Кинопоиск и FindFace.

### 1. Цели и задачи

#### Цель дисциплины

- формирование знаний, умений и навыков использования машинного обучения в задачах продуктовой разработки.

#### Задачи дисциплины

- формирование навыка оценивания базовые методы ML в разработке IT-продукта с позиции руководителя разработки;
- формирование базы для принятия решений по выбору между классическими методами ML и продвинутыми методами ML;
- формирование навыка анализа необходимого количества данных для обучения конкретной ML-модели.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен учитывать конкретные условия выполняемых задач и разрабатывать инновационные решения при управлении проектами и процессами в сфере ИКТ	ОПК-2.1 Выполняет оценку условий развития проекта в области ИКТ
	ОПК-2.2 Владеет теоретической базой управления инновационными проектами и процессами в сфере ИКТ
	ОПК-2.3 Умеет применять теоретический инструментарий на практике в ходе разработки инновационных IT-решений
ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.1 Владеет методами стратегического планирования и прогнозирования в профессиональной деятельности
	ОПК-3.2 Самостоятельно выбирает и обосновывает выбор современных методов и программного инструментария сбора, обработки и анализа данных
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
	ОПК-3.4 Способен самостоятельно приобретать, развивать и применять математические знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- тенденции использования ML в разработке IT-продуктов;
- разницу между обучением с учителем и без;
- ключевые библиотеки современного машинного обучения;
- регрессию, классификацию, кластеризацию, ансамблирование (базово);
- этапы развития ML, основные инструменты и библиотеки ML;
- кейсы использования обработки временных рядов через марковские процессы принятия решений.

уметь:

- категоризировать методы ml и определять ситуации из эффективного использования;
- формулировать задачи для ml-исследователей;
- определять базовые методы ml, необходимые использования в своих продуктах;
- извлекать и анализировать инсайты из данных с использованием инструментария машинного обучения;
- выбирать инструменты для построения моделей машинного обучения.

владеть:

- навыками подготовки задания для ML-исследователя.

#### 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	Машинное обучение: база		3		19
2	Классический ML		3		19
3	Продвинутые методы ML		3		20
4	Практические методы ML		3		20
Итого часов			12		78
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

###### 1. Машинное обучение: база

История ML. Постановка задачи ML. Циклы ML. Окружение ML-разработчика. Библиотеки для ML.

###### 2. Классический ML

Регрессия. Постановка задачи, регуляризация, метрики качества. Классификация. Постановка задачи, метрики качества. Решение задач классификации через задачу регрессии. Градиентный спуск. Ансамблирование моделей: bagging, boosting, random forest. Обучение без учителя. Кластеризация. PCA, tSNE, UMAP.

###### 3. Продвинутые методы ML

Классические нейронные сети. Computer Vision: основные типы задач, основные метрики качества

NLP: рекуррентные нейросети, механизм памяти, машинный перевод. Рекомендательные системы.

#### 4. Практические методы ML

Рекомендательные системы. Обработка временных рядов.

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

Занятия проводятся с применением мультимедийных технологий. Проводятся на платформе Skillfactory.

### 6. Перечень рекомендуемой литературы

#### Основная литература

1. Машинное обучение: новый искусственный интеллект [Текст]/Э. Алпайдин, -М., Изд. группа "Точка", 2017

#### Дополнительная литература

Рекомендуемая литература для самостоятельного изучения

1. П. Флах. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных Machine Learning and Knowledge Discovery in Databases / Peter A. Flach; Tijl Bie; Nello Cristianini. Springer Berlin Heidelberg. 2012

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

1. Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук ; пер. с англ. А.Б. Огурцова. - Москва : ДМК Пресс, 2018. - 250 с. - ISBN 978-5-97060-508-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028135> (дата обращения: 05.02.2022). – Режим доступа: по подписке.

2. Кацов, И. Машинное обучение для бизнеса и маркетинга : практическое руководство / И. Кацов. - Санкт-Петербург : Питер, 2019. - 512 с. - (Серия «IT для бизнеса»). - ISBN 978-5-4461-0926-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1783938> (дата обращения: 05.02.2022). – Режим доступа: по подписке.

3. Буйначев, С.К. Основы программирования на языке Python : учебное пособие / С.К. Буйначев, Н.Ю. Боклаг ; Министерство образования и науки Российской Федерации, Уральский федеральный университет им. первого Президента России Б. Н. Ельцина. – Екатеринбург : Издательство Уральского университета, 2014. – 92 с. : табл., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=275962> – Библиогр. в кн. – ISBN 978-5-7996-1198-9. – Текст : электронный

4. Воронцов К.В. Машинное обучение Школа Анализа данных Яндекс. МФТИ. национальный открытый университет Интуит, 2015 – Режим доступа: свободный – URL: <https://www.intuit.ru/studies/courses/13844/1241/info>

Сайт <https://www.bigdataschool.ru/>

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

Для части занятий потребуется Zoom. Google Drive для доступа к материалам курса. Необходимо наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

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

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Бизнес-информатика
<b>профиль подготовки:</b>	Управление цифровым продуктом центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

**Разработчики:**

О.А. Культепина, старший методист  
В.С. Акулышин, методист  
П.И. Ахтямов, преподаватель

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен учитывать конкретные условия выполняемых задач и разрабатывать инновационные решения при управлении проектами и процессами в сфере ИКТ	ОПК-2.1 Выполняет оценку условий развития проекта в области ИКТ
	ОПК-2.2 Владеет теоретической базой управления инновационными проектами и процессами в сфере ИКТ
	ОПК-2.3 Умеет применять теоретический инструментарий на практике в ходе разработки инновационных IT-решений
ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.1 Владеет методами стратегического планирования и прогнозирования в профессиональной деятельности
	ОПК-3.2 Самостоятельно выбирает и обосновывает выбор современных методов и программного инструментария сбора, обработки и анализа данных
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
	ОПК-3.4 Способен самостоятельно приобретать, развивать и применять математические знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение в задачах продуктовой разработки» обучающийся должен:

### знать:

- тенденции использования ML в разработке IT-продуктов;
- разницу между обучением с учителем и без;
- ключевые библиотеки современного машинного обучения;
- регрессию, классификацию, кластеризацию, ансамблирование (базово);
- этапы развития ML, основные инструменты и библиотеки ML;
- кейсы использования обработки временных рядов через марковские процессы принятия решений.

### уметь:

- категоризировать методы ml и определять ситуации из эффективного использования;
- формулировать задачи для ml-исследователей;
- определять базовые методы ml, необходимые использования в своих продуктах;
- извлекать и анализировать инсайты из данных с использованием инструментария машинного обучения;
- выбирать инструменты для построения моделей машинного обучения.

### владеть:

- навыками подготовки задания для ML-исследователя.

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

Во время текущего контроля студент проходит тестирования с автоматической проверкой на платформе.

Примеры вопросов:

Вопрос 1. Какая из перечисленных техник является методом обучения без учителя?

А. Логистическая регрессия

В. Градиентный бустинг

- C. Кластеризация
- D. Рекуррентные нейронные сети
- E. Машины опорных векторов (SVM)

Вопрос 2. Какой из следующих методов не является продвинутым методом машинного обучения (ML)?

- A. Глубокие нейронные сети
- B. Решающие деревья
- C. Методы ансамблирования, такие как случайный лес (Random Forest)
- D. Градиентный бустинг
- E. Механизмы внимания в NLP

Вопрос 3. Какая библиотека не используется для разработки ML-моделей?

- A. TensorFlow
- B. PyTorch
- C. Scikit-learn
- D. Pandas
- E. Microsoft Excel

Вопрос 4. Что из перечисленного не относится к этапам разработки модели машинного обучения?

- A. Сбор данных
- B. Подготовка данных
- C. Обучение модели
- D. Проектирование пользовательского интерфейса
- E. Оценка модели

Вопрос 5. Какой метод оценки задач ML является эффективным для определения приоритетов в разработке продукта?

- A. Регрессионный анализ
- B. Анализ временных рядов
- C. Рекомендательные системы
- D. Принцип работы с естественным языком (NLP)
- E. Кластеризация

Во время занятий могут проходить интерактивные обсуждения в чатах курса.

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

Примеры заданий для дифференцированного зачета:

1. Может ли в методе ближайших соседей при  $X = 2$  получиться лучший результат, чем при  $X = 1$ ? Отказы от классификации тоже считать ошибками.
2. Покажите, что с ростом размерности пространства признаков при равномерном распределении точек в кубе  $[0; 1]$  вероятность попасть в куб  $[0; 0, 99]$  стремится к нулю. Это одна из иллюстраций проклятия размерностей (dimension curse). Попробуйте придумать или найти еще какую-нибудь иллюстрацию к этому явлению и кратко изложить. В чем по-вашему суть проклятия размерности и какое это имеет значение для задач машинного обучения?
3. Покажите асимптотическую эквивалентность энтропийного и статистического критериев информативности.
4. Какая стратегия поведения в листьях решающего дерева приводит к меньшей вероятности ошибки: отвечать тот класс, который преобладает в листе, или отвечать случайно с тем же распределением классов, что и в листе?
5. Покажите, что регуляризатор в задаче линейной классификации имеет вероятностный смысл априорного распределения параметров моделей. Какие распределения задают 1-регуляризатор и 2-регуляризатор?



## Критерии оценивания

Оценка отлично (10 баллов) - 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 баллов) - 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 баллов) - выставляется студенту, показавшему всесторонние, 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 баллов) - 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 баллов) - 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 баллов) - 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 балла) - 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 балла) - 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 балла) - 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 балл) - 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 формате выполнения итогового задания. На выполнение отводится не более 60 минут.