

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

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

| | |
|----------------------------|---|
| | Рабочая программа дисциплины (модуля) |
| по дисциплине: | Машинное обучение на практике |
| по направлению: | Информатика и вычислительная техника |
| профиль подготовки: | Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования |
| курс: | 1 |
| квалификация: | магистр |

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

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 1

Программу составил: В.В. Кантор, преподаватель

Программа обсуждена на заседании кафедры алгоритмов и технологий программирования 20.02.2025

Аннотация

Курс для студентов и профессионалов, которые хотят овладеть навыками машинного обучения и применять их на практике. На курсе будут подробно рассмотрены основные методы машинного обучения, приведены примеры их использования в различных приложениях и представлены практические задания. Особое внимание уделено выбору признаков, регуляризации моделей, работе с несбалансированными данными и управлению ошибками. Машинное обучение находит все большее применение в сельском хозяйстве, в бизнесе, по отраслям, в рекламе и т.д., позволяя улучшить производительность и эффективность различных процессов.

1. Цели и задачи

Цель дисциплины

- обучить студентов особенностям работы с алгоритмами машинного обучения в условиях современных объёмов данных.

Задачи дисциплины

- приобретение студентами навыков адаптации известных алгоритмов к большим объемам данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

| Код и наименование компетенции | Индикаторы достижения компетенции |
|---|---|
| УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели | УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов |
| | УК-3.2 Учитывает в своей социальной и профессиональной деятельности интересы, особенности поведения и мнения (включая критические) людей, с которыми работает/взаимодействует, в том числе посредством корректировки своих действий |
| | УК-3.3 Способен предвидеть результаты (последствия) как личных, так и коллективных действий |
| | УК-3.4 Способен планировать командную работу, распределять поручения членам команды, организовывать обсуждение разных идей и мнений |
| ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи | ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности |
| | ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость |
| | ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации |
| | ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения |
| | ОПК-3.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем |
| | ОПК-3.3 Способен использовать исследовательские методы при решении новых задач, применяя знания из различных областей науки (техники) |

| | |
|---|--|
| ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения | ОПК-3.4 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений |
| | ОПК-3.5 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий |
| | ОПК-3.6 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте |
| | ОПК-3.7 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач |
| ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии | ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения |
| | ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности |
| | ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий |
| | ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками |

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- современные алгоритмы машинного обучения и их адаптации к анализу больших объемов данных.

уметь:

- применять алгоритмы и подходы машинного обучения в условиях современного строения и объёмов данных.

владеть:

- фреймворками и алгоритмами распределенного машинного обучения в условиях больших данных.

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 | 3 | 3 | | 3 |
| 2 | Рекомендательные системы и поиск | 3 | 3 | | 3 |
| 3 | Ценообразование на основе данных и прогнозирование спроса | 3 | 3 | | 3 |
| 4 | Скоринг | 3 | 3 | | 3 |
| 5 | Лидогенерация | 3 | 3 | | 3 |
| 6 | Приоритизация инвестиций и оптимизация расходов производства | 3 | 3 | | 3 |
| 7 | Телеком | 3 | 3 | | 3 |
| 8 | Банки и FinTech | 3 | 3 | | 3 |
| 9 | Промышленность и сельское хозяйство | 2 | 2 | | 2 |
| 10 | РайдТех | 2 | 2 | | 2 |
| 11 | Рекламные технологии | 2 | 2 | | 2 |
| Итого часов | | 30 | 30 | | 30 |
| 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. Введение

Напоминание основ машинного обучения, обзор применений машинного обучения во взаимодействии бизнеса с клиентом и в оптимизации расходов бизнеса. Обзор отраслей, наиболее активно использующих машинное обучение. Примеры других сфер бизнеса и использования наук о данных в них.

2. Рекомендательные системы и поиск

Базовые решения на основе популярностей и коллаборативной фильтрации, различие постановок задачи, гибридные методы. Нейросетевые подходы к рекомендациям и поиску. Метрики вовлеченности, точности и разнообразия. Валидация на исторических данных и при внедрении алгоритмов.

3. Ценообразование на основе данных и прогнозирование спроса

Базовая идея оценки эластичности спроса по цене, различие постановок задачи ценообразования в случае ограниченного и динамически меняющегося предложения, обзор решений с помощью машинного обучения, роль прогнозирования спроса в ценообразовании и логистике, различие между спросом на материальные и цифровые товары и услуги.

4. Скоринг

Классическая задача оценки вероятности дефолта, особенности постановки и валидации, популярные подходы, скоринг мошенников, скоринг склонности клиента к произвольному действию (на примере оттока).

5. Лидогенерация

Построение рекламных сегментов на основе правил, расширение и сужение сегментов с помощью look-alike моделей, positive-unlabeled learning, таргетирование рекламы с помощью прогнозов вероятности целевого действия.

6. Приоритизация инвестиций и оптимизация расходов производства

Детектирование аномалий (примеры: утечки электроэнергии, фродовые продажи, предиктивное обслуживание). Оптимизация маркетинговых расходов. Оптимизация работы персонала и процессов в компании. Управление мотивацией и численностью. Business process mining. Автоматизация работы с помощью deep learning. Чат-боты и LLM. Автоматизация с помощью Computer Vision.

7. Телеком

Планирование развития сети, блокировка спама, продвижение услуг и удержание клиентов в B2C и B2B сегментах. Ритейл и e-commerce: рекомендации, прогноз спроса, ценообразование, управление персоналом (Workforce management).

8. Банки и FinTech

Процесс выдачи кредитов (связь рисков, лимитов и ставок, типы кредитных продуктов и особенности задач), применение машинного обучения для продвижения финансовых услуг и работы с лояльностью клиентов, машинное обучение в управлении личными финансами и инвестициями.

9. Промышленность и сельское хозяйство

Оптимизация расходов при производстве, предиктивные ремонты и предиктивное устранение дефектов производства, автоматизация работы персонала.

10. РайдТех

Ценообразование в такси, кикшеринге и доставке, прогноз ожидаемого времени прибытия, применение машинного обучения для улучшения пользовательского опыта.

11. Рекламные технологии

Инфраструктура для показа digital рекламы и оптимизация открутки рекламы на своем и на внешнем инвентаре, реклама на телевидении, телемаркетинг и наружная реклама - оценка эффективности и оптимизация инвестиций. Медиа: персональные рекомендации контента в видео- и аудио- стриминге, рекомендации текстового контента, трудности оценки инвестиций в контент, рекламные механики на основе computer vision и генеративных моделей.

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

Для практических занятий:

Компьютерный класс. Каждый компьютер должен иметь выход в интернет и ПО для подключения к удалённым серверам.

6.Перечень рекомендуемой литературы

Основная литература

Фонд кафедры:

1. Методы и средства вычислений с объектами. Аппликативные вычислительные системы [Текст] : [учеб. пособие для вузов] / В. Э. Вольфенгаген ; Ин-т актуального образования <ЮрИнфоР-МГУ>, Каф. перспективных компьютерных исследований и информационных технологий. — М. : JurInfoR, 2004. — 789 с. — (Компьютерные науки и информационные технологии). - 2000 экз. - ISBN 5-89158-100-0 (в пер.) .

2. Комбинаторная логика в программировании. Вычисления с объектами в примерах и задачах [Текст] : [учеб. пособие для вузов] / В. Э. Вольфенгаген ; НОУ Ин-т Актуального образования "ЮрИнфоР-МГУ, Каф. перспективных компьт. исслед. и информ. технологий. — 3-е изд., доп. и перераб. — М. : Ин-т "ЮрИнфоР-МГУ, 2008. — 384 с.

Дополнительная литература

Фонд кафедры:

1. Параллельное программирование многопоточных систем с разделяемой памятью [Текст] : учеб. пособие для вузов / А. Г. Тормасов. — М. : Физматкнига, 2014. — 208 с.

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

<https://www.coursera.org/specializations/big-data-engineering> - специализация из 5 курсов, посвящённая тематике обработки больших данных.

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

- Компьютерный класс. Каждый компьютер должен иметь выход в интернет и ПО для подключения к удалённым серверам.
- Удалённый Hadoop-кластер. На кластере должен быть развернута последняя версия Cloudera Manager, в который нужно встроить такие сервисы: HDFS, YARN, Spark2 on YARN.

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

Успешное освоение курса требует большой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- проработку учебного материала;
- подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- доказательство отдельных утверждений, свойств.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

| | |
|----------------------------|---|
| по направлению: | Информатика и вычислительная техника |
| профиль подготовки: | Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования |
| курс: | <u>1</u> |
| квалификация: | магистр |

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

Разработчик: В.В. Кантор, преподаватель

1. Компетенции, формируемые в процессе изучения дисциплины

| Код и наименование компетенции | Индикаторы достижения компетенции |
|--|--|
| УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели | УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов |
| | УК-3.2 Учитывает в своей социальной и профессиональной деятельности интересы, особенности поведения и мнения (включая критические) людей, с которыми работает/взаимодействует, в том числе посредством корректировки своих действий |
| | УК-3.3 Способен предвидеть результаты (последствия) как личных, так и коллективных действий |
| | УК-3.4 Способен планировать командную работу, распределять поручения членам команды, организовывать обсуждение разных идей и мнений |
| ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи | ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности |
| | ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость |
| | ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации |
| ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения | ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения |
| | ОПК-3.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем |
| | ОПК-3.3 Способен использовать исследовательские методы при решении новых задач, применяя знания из различных областей науки (техники) |
| | ОПК-3.4 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений |
| | ОПК-3.5 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий |
| | ОПК-3.6 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте |
| | ОПК-3.7 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач |
| | ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения |

| | |
|---|---|
| ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии | ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности |
| | ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий |
| | ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками |

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение на практике» обучающийся должен:

знать:

- современные алгоритмы машинного обучения и их адаптации к анализу больших объемов данных.

уметь:

- применять алгоритмы и подходы машинного обучения в условиях современного строения и объёмов данных.

владеть:

- фреймворками и алгоритмами распределенного машинного обучения в условиях больших данных.

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

1. Какие способы построения рекомендательных систем Вы знаете? Кратко опишите суть каждого из них.
2. Обработка естественного языка: примеры задач, этапы решения задач естественного языка.
3. Обработка естественного языка: применение регулярных, контекстно-свободных и контекстно-зависимых грамматик.

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

1. Приоритизация инвестиций и оптимизация расходов производства.
2. Детектирование аномалий (примеры: утечки электроэнергии, фродовые продажи, предиктивное обслуживание).
3. Оптимизация маркетинговых расходов.
4. Оптимизация работы персонала и процессов в компании. Управление мотивацией и численностью. Business process mining.
5. Автоматизация работы с помощью deep learning. Чат-боты и LLM. Автоматизация с помощью Computer Vision.
6. Инфраструктура для показа digital рекламы и оптимизация открутки рекламы. Планирование развития сети, блокировка спама, продвижение услуг и удержание клиентов в B2C и B2B сегментах. Ритейл и e-commerce: рекомендации, прогноз спроса, ценообразование, управление персоналом.
7. Построение рекламных сегментов на основе правил, расширение и сужение сегментов с помощью look-alike моделей, positive-unlabeled learnin.
8. Роль прогнозирования спроса в ценообразовании и логистике, различие между спросом на материальные и цифровые товары и услуги.
9. Базовые решения на основе популярностей и коллаборативной фильтрации, различие постановок задачи, гибридные методы.
10. Примеры других сфер бизнеса и использования наук о данных в них.

11. Оптимизация взаимодействия с клиентом.

Билет №1

1. Инфраструктура для показа digital рекламы и оптимизация открутки рекламы.
2. Оптимизация маркетинговых расходов.

Билет №2

1. Оптимизация работы персонала и процессов в компании.
2. Оптимизация взаимодействия с клиентом.

Критерии оценивания

Оценка "Отлично" (10) 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" (9) выставляется студенту, показавшему свободное оперирование знаниями учебной программы дисциплины, выполнение заданий творческого характера.

Оценка "Отлично" (8) выставляется студенту, показавшему владение программным учебным материалом с наличием несущественных ошибок в действиях, самостоятельно исправляемых учащимся.

Оценка "Хорошо" (7) 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" (6) выставляется студенту в случае если он осознает воспроизведение программного учебного материала, в том числе и различной степени сложности, с несущественными ошибками, затруднения в применении отдельных навыков.

Оценка "Хорошо" (5) выставляется студенту если теоретическое содержание освоено не полностью, некоторые практические навыки сформированы недостаточно, в некоторых случаях были допущены ошибки.

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 владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

Оценка "Удовлетворительно" (3) выставляется студенту в случае большого количества недочётов и неправильных ответов, а также пассивной работе в ходе занятий, многие учебные задания не выполнены.

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

Оценка "Неудовлетворительно" (1) выставляется студенту, который не освоил теоретическое и практическое содержание курса, все выполненные учебные задания содержат грубые ошибки.

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

При проведении зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.