

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Новосибирский национальный исследовательский государственный университет»
(Новосибирский государственный университет, НГУ)
**Структурное подразделение Новосибирского государственного университета –
Специализированный учебно-научный центр Университета (СУНЦ НГУ)**
Министерство науки и высшего образования Российской Федерации

СОГЛАСОВАНО Заместитель директора по УР  (Петровская О.В.) 23 ноября 2023 г.	УТВЕРЖДЕНО На заседании ученого совета СУНЦ НГУ Протокол № 48 от 23 ноября 2023 г.	УТВЕРЖДАЮ Директор СУНЦ НГУ  (Некрасова Л.А.) 23 ноября 2023 г.
---	--	--

РАБОЧАЯ ПРОГРАММА

курса внеурочной деятельности «Машинное обучение»

Заведующий лабораторией инженерного конструирования

Якушкин Сергей Владимирович



Новосибирск 2023

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Спецкурс «Машинное обучение» предоставляет ученикам 9-11 классов начальные знания в области машинного обучения и искусственного интеллекта. Курс направлен на получение учащимися как теоретических основ, так и освоение практических навыков, необходимых для разработки алгоритмов машинного обучения. Учащиеся изучат принципы работы различных алгоритмов машинного обучения, применяемых для решения задач классификации, регрессии, кластеризации, а также получат опыт применения изучаемых алгоритмов на практике в ходе работы над индивидуальным проектом. Наличие опыта программирования на Python является преимуществом для успешного прохождения курса.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ

По итогам прохождения специального курса учащиеся:

1. Имеют краткие сведения об истории машинного обучения и его актуальном состоянии.
2. Владели теоретическим материалом об основных методах машинного обучения.
3. Обладают пониманием математических понятий, теорий, методов в объёме, достаточном для понимания устройства существующих моделей машинного обучения и создания собственных.
4. Ориентируются в библиотеках и фреймворках на Python, используемых в машинном обучении.

В ходе спецкурса учащиеся должны выбрать тему проекта и разработать решение индивидуально или в группах.

По результатам работы и по посещению будет выставлен зачёт. Зачёт недифференцированный (зачёт / незачёт).

Объём спецкурса – 48 академических часов.

Планируемое время проведения спецкурса – I-II семестр 2023-24 учебного года (сентябрь 2023 – май 2024 г.).

СОДЕРЖАНИЕ СПЕЦКУРСА

Занятие 1 (2 часа)

Типы машинного обучения: машинное обучение с учителем, машинное обучение без учителя, машинное обучение с подкреплением. Различия. Дизайн проекта по машинному обучению (основные этапы): постановка задачи, составление датасета, обработка данных, обучение модели, тестирование модели, подсчёт метрик. Рассматривается на примере проектов прошлых лет.

Занятие 2 (2 часа)

Обработка данных. Табличные данные. Чтение и запись в файлы формата csv. Признаковое описание объектов. Работа с библиотекой pandas.

Занятие 3 (2 часа)

Разделение данных на выборки: работа с библиотекой scikit-learn. (Для работы необходимо познакомиться с платформой kaggle и загрузить наборы данных.)

Занятие 4 (2 часа)

Нормализация и проверка признаков на корреляцию. Нормализаторы “StandardScaler”, “MinMaxScaler”, преобразование в числовые данные – энкодинг.

Занятие 5 (2 часа)

Типы задач. Классификация. Линейные модели: Метод ближайших соседей, опорных векторов, линейный дискриминант Фишера.

Занятие 6 (2 часа)

Классификация: Деревья решений, случайный лес. Бинарное признаковое описание объектов.

Занятие 7 (2 часа)

Кластеризация: Метод ближайших соседей. Расстояния (евклидово, косинусное, манхэттенское) для разных типов объектов.

Занятие 8 (2 часа)

Классификация и регрессия: Логистическая регрессия. Метод максимального правдоподобия.

Занятие 9 (2 часа)

Многомерная линейная регрессия: Метод наименьших квадратов.

Занятие 10 (2 часа)

Матрица ошибок. Метрики: чувствительность и специфичность, точность и полнота, ROC-кривая и AUC, F-1 мера.

Занятие 11 (2 часа)

Подготовка к защите проекта. Постановка цели, определение задач проекта, подбор датасета, алгоритмов обучения.

Занятие 12 (2 часа)

Основы нейросетей. Устройство перцептрона, типы функции активации. Градиентный спуск. Обратное распространение ошибки.

Занятие 13 (2 часа)

Свёрточные нейронные сети: слой свёртки и слой подвыборки. Практика: расчёт свёртки векторов. Классы Conv2D и MaxPooling2D.

Занятие 14 (2 часа)

Градиентный бустинг, catboost.

Занятие 15 (2 часа)

Введение в распознавание изображений (computer vision). Знакомство с наиболее популярными датасетами, используемых в задаче распознавания изображений: база ImageNet, MNIST, CIFAR. Практика: классификация изображений на примере MNIST.

Занятие 16 (2 часа)

Фреймворки для машинного обучения. Pytorch, Keras. Сходства и различия.

Занятие 17 (2 часа)

Кодирование текстовых данных, векторизация. Изучение векторизаторов “one-hot-vectorizer”, “word2vec”. Библиотека векторов RusVectors.

Занятие 18 (2 часа)

Анализ текстовых данных: библиотека SpaCy. Использование дополнительных признаков для корректирования вывода нейросети. Модель FastText.

Занятие 19 (2 часа)

Выбор модели и подбор оптимальных параметров. Шаг обучения, количество эпох обучения. Библиотека sklearn.model_selection для оптимизации гиперпараметров и валидации модели.

Занятие 20 (2 часа)

Рекуррентные нейронные сети. Метод обучения Backpropagation Through Time (BPTT).

Занятие 21 (2 часа)

Векторные представления графов. Векторизация node2vec. Обобщённый автокодировщик на графах GraphEDM.

Занятие 22 (2 часа)

Обучение нейронных сетей, основанных на архитектуре Transformer: Внимание (многомерное, иерархическое, Query–Key–Value, self-attention). Статья Attention Is All You Need.

Занятие 23 (2 часа)

Кодировщик и декодировщик.

Занятие 24 (2 часа)

Подготовка к защите проекта. Подсчёт метрик, измерение качества полученных моделей, подведение итогов курса. Подготовка презентации для защиты проекта.

ТЕМАТИЧЕСКОЕ ПЛАНИРОВАНИЕ

№ п/п	Наименование разделов и тем программы	Кол-во часов	Воспитательный компонент
1	<p>Типы машинного обучения: машинное обучение с учителем, машинное обучение без учителя, машинное обучение с подкреплением. Различия. Дизайн проекта по машинному обучению (основные этапы): постановка задачи, составление датасета, обработка данных, обучение модели, тестирование модели, подсчёт метрик. Рассматривается на примере проектов прошлых лет.</p>	2	<p>Развитие и поддержка одаренности обучающихся и обеспечение участия в олимпиадах и конкурсах</p> <p>Установление доверительных отношений между руководителем объединений и обучающимися и между обучающимися непосредственно через беседы, дискуссии.</p>
2	<p>Обработка данных. Табличные данные. Чтение и запись в файлы формата csv. Признаковое описание объектов. Работа с библиотекой pandas.</p>	2	<p>Формирование ответственного отношения к своему здоровью и безопасности.</p> <p>Готовность обучающихся к саморазвитию,</p>
3	<p>Разделение данных на выборки: работа с библиотекой scikit-learn. (Для работы необходимо познакомиться с платформой kaggle и загрузить наборы данных.)</p>	2	<p>самостоятельности и личному самоопределению. Формирование способности к оценке своих возможностей. Наличие мотивации к целенаправленной социально значимой деятельности.</p>
4	<p>Нормализация и проверка признаков на корреляцию. Нормализаторы “StandartScaler”, “MinMaxScaler”, преобразование в числовые данные – энкодинг.</p>	2	
5	<p>Типы задач. Классификация. Линейные модели: Метод ближайших соседей, опорных векторов, линейный дискриминант Фишера.</p>	2	
6	<p>Классификация: Деревья решений, случайный лес. Бинарное признаковое описание объектов.</p>	2	

7	Кластеризация: Метод ближайших соседей. Расстояния (евклидово, косинусное, манхэттенское) для разных типов объектов.	2
8	Классификация и регрессия: Логистическая регрессия. Метод максимального правдоподобия.	2
9	Многомерная линейная регрессия: Метод наименьших квадратов.	2
10	Матрица ошибок. Метрики: чувствительность и специфичность, точность и полнота, ROC-кривая и AUC, F-1 мера.	2
11	Подготовка к защите проекта. Постановка цели, определение задач проекта, подбор датасета, алгоритмов обучения.	2
12	Основы нейросетей. Устройство перцептрона, типы функции активации. Градиентный спуск. Обратное распространение ошибки.	2
13	Свёрточные нейросети: слой свёртки и слой подвыборки. Практика: расчёт свёртки векторов. Классы Conv2D и MaxPooling2D.	2
14	Градиентный бустинг, catboost.	2

15	Введение в распознавание изображений (computer vision). Знакомство с наиболее популярными датасетами, используемых в задаче распознавания изображений: база ImageNet, MNIST, CIFAR. Практика: классификация изображений на примере MNIST.	2
16	Фреймворки для машинного обучения. Pytorch, Keras. Сходства и различия.	2
17	Кодирование текстовых данных, векторизация. Изучение векторизаторов “one-hot-vectorizer”, “word2vec”. Библиотека векторов RusVectores.	2
18	Анализ текстовых данных: библиотека SpaCy. Использование дополнительных признаков для корректирования вывода нейросети.	2
19	Выбор модели и подбор оптимальных параметров. Шаг обучения, количество эпох обучения. Библиотека sklearn.model_selection для оптимизации гиперпараметров и валидации модели.	2
20	Рекуррентные нейронные сети. Метод обучения Backpropagation Through Time (BPTT).	2
21	Векторные представления графов. Векторизация node2vec. Обобщённый автокодировщик на графах GraphEDM.	2
22	Обучение нейронных сетей, основанных на архитектуре Transformer: Внимание (многомерное, иерархическое, Query–Key–Value, self-attention). Статья Attention Is All You Need.	2

23	Кодировщик и декодировщик.	2
24	Подготовка к защите проекта. Подсчёт метрик, измерение качества полученных моделей, подведение итогов курса. Подготовка презентации для защиты проекта.	2
	Всего	48

Приложение 1.

УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

1. Франсуа Шолле “Глубокое обучение на Python” – Издательский дом «Питер», 2018.
2. Rosebrock, Adrian “Deep Learning for Computer Vision with Python” – Соединенные Штаты Америки, PyImageSearch, 2017.
3. Ian Goodfellow and Yoshua Bengio and Aaron Courville “Глубокое обучение” – MIT Press, 2016.
URL: deeplearningbook.org
4. Hobson Lane, Cole Howard, Hannes Max Napke “NLP in action: Understanding, analyzing, and generating text with Python”
5. М. Минский, С. Пейперт Перцептроны = Perceptrons. — М.: «Мир», 1971.
6. Ф. Вассерман Нейрокомпьютерная техника: Теория и практика. — М.: «Мир», 1992.
7. Саймон Хайкин Нейронные сети: полный курс = Neural Networks: A Comprehensive Foundation. — 2-е изд. — М.: «Вильямс», 2006.
8. Машинное обучение (курс лекций, К.В.Воронцов)
URL: https://machinelearning.ru/wiki/index.php?title=Машинное_обучение%28курс_лекций%2C_К.В.Воронцов%29 (дата обращения: 17.10.2023)
9. Сергей Николенко, Артур Кадури, Екатерина Архангельская. Глубокое обучение. Погружение в мир нейронных сетей. — «Питер», 2018.
10. Маккини, У. Python и анализ данных : первичная обработка данных с применением pandas, NumPy и IPython; пер. с англ. А. А. Слинкина. – 2-е изд. – Москва : ДМК

Пресс, 2020.

URL: <https://biblioclub.ru/index.php?page=book&id=686719> (дата обращения:
17.10.2023).