

Документ подписан простой электронной подписью  
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной и научной работе

Дата подписания: 28.09.2023 12:42:09

Уникальный идентификатор документа:

d7a26b9e8ca85e98ec3de2eb454b4659d061f249

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
высшего образования

«Национальный исследовательский технологический университет «МИСИС»

## Рабочая программа дисциплины (модуля)

# Основы машинного обучения

Закреплена за подразделением

Кафедра АСУ

Направление подготовки

45.04.02 ЛИНГВИСТИКА

Профиль

Цифровая лингвистика и локализация

Квалификация

Магистр

Форма обучения

очная

Общая трудоемкость

3 ЗЕТ

Часов по учебному плану

108

Формы контроля в семестрах:

в том числе:

экзамен 3

аудиторные занятия

26

самостоятельная работа

46

часов на контроль

36

### Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		Итого	
Неделя	14			
Вид занятий	УП	РП	УП	РП
Лекции	13	13	13	13
Практические	13	13	13	13
Итого ауд.	26	26	26	26
Контактная работа	26	26	26	26
Сам. работа	46	46	46	46
Часы на контроль	36	36	36	36
Итого	108	108	108	108

Программу составил(и):

доц., Конов И.С.

Рабочая программа

**Основы машинного обучения**

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - магистратура Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСИС» по направлению подготовки 45.04.02 ЛИНГВИСТИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

45.04.02 ЛИНГВИСТИКА, 45.04.02-МЛГ-23-3.plx Цифровая лингвистика и локализация, утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

45.04.02 ЛИНГВИСТИКА, Цифровая лингвистика и локализация, утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

Рабочая программа одобрена на заседании

**Кафедра АСУ**

Протокол от 23.06.2020 г., №10

Руководитель подразделения Тёмкин И.О.

**1. ЦЕЛИ ОСВОЕНИЯ**

1.1	-Целью дисциплины является получение знаний, умений и навыков по обработке, анализу и обобщению информации в сфере машинного обучения (machine learning )
1.2	

**2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Блок ОП:		Б1.В
<b>2.1</b>	<b>Требования к предварительной подготовке обучающегося:</b>	
2.1.1	Контроль качества переводческих проектов	
2.1.2	Основы программирования	
2.1.3	Основы скорочтения	
2.1.4	Способы быстрого запоминания	
2.1.5	Нормативная база в области лингвистики	
2.1.6	Профильные интерфейсы лингвиста	
2.1.7	Современные подходы к управлению командами	
2.1.8	Форматирование и верстка	
<b>2.2</b>	<b>Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:</b>	
2.2.1	NLP-аналитика	
2.2.2	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.3	Производственная практика (педагогическая)	
2.2.4	Редактирование медиаконтента	
2.2.5	Субтитрование и транскрибирование	

**3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ**

<b>ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода</b>	
<b>Знать:</b>	
ПК-2-31 основные методики автоматизации лингвистических проектов.	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Знать:</b>	
ПК-1-31 основные подходы к разработке цифровые лингвистические продукты с использование языков программирования.	
<b>ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода</b>	
<b>Уметь:</b>	
ПК-2-У1 внедрять и использовать основные методики автоматизации лингвистических проектов.	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Уметь:</b>	
ПК-1-У1 внедрять основные подходы к разработке цифровые лингвистические продукты с использование языков программирования.	
<b>ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода</b>	
<b>Владеть:</b>	
ПК-2-В1 управления втоматизации лингвистических проектов.	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Владеть:</b>	
ПК-1-В1 использование инструментами внедрения и разработки ML проектов.	

**4. СТРУКТУРА И СОДЕРЖАНИЕ**

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Формируемые индикаторы компетенций	Литература и эл. ресурсы	Примечание	КМ	Выполняемые работы
	<b>Раздел 1. Раздел 1 Машинное обучение и ее модель (ML). Data science, Анализ данных.</b>							
1.1	Введение. Data science /Лек/	3	4	ПК-1-31	Л1.2Л2.1 Л2.2			
	<b>Раздел 2. Раздел 2 Математические модели, обработка данных, Python и библиотеки</b>							
2.1	Данные. Алгоритмы. Метрики. Оптимизация ML /Лек/	3	4	ПК-1-31	Л1.2			
2.2	Python. Освоение библиотеки SciKit-Learn /Пр/	3	10	ПК-1-У1 ПК-2-У1	Л2.1 Л2.2 Э3 Э4			
2.3	Изучение метрик ML /Пр/	3	2	ПК-1-У1 ПК-2-У1	Л2.1 Л2.2 Э1 Э3 Э4			
2.4	Методы оптимизации ML /Пр/	3	1	ПК-2-У1	Л2.1 Л2.2 Э1 Э3 Э4			
2.5	Подготовка к занятиям, подготовка и оформление отчетов. /Ср/	3	42	ПК-2-У1	Э1 Э2 Э3			
	<b>Раздел 3. Раздел 3 Кластеризация и визуализация данных</b>							
3.1	Кластеризация и визуализация данных K-means, иерархическая кластеризация, метрики качества кластеризации. /Лек/	3	5		Л1.1			
3.2	Самостоятельное изучение учебного материала, оформление отчета по лабораторной работе /Ср/	3	4	ПК-2-В1 ПК-1-В1	Э1 Э2 Э3		КМ1	

**5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ****5.1. Контрольные мероприятия (контрольная работа, тест, коллоквиум, экзамен и т.п), вопросы для самостоятельной подготовки**

Код КМ	Контрольное мероприятие	Проверяемые индикаторы компетенций	Вопросы для подготовки
КМ1	промежуточные вопросы	ПК-1-У1;ПК-2-У1	<p>Что такое машинное обучение?</p> <p>Какие основные задачи решаются с помощью машинного обучения?</p> <p>В чем разница между обучением с учителем, без учителя и обучением с подкреплением?</p> <p>Как работает метод ближайших соседей?</p> <p>Опишите принцип работы линейной регрессии.</p> <p>Объясните, что такое логистическая регрессия и для каких задач она используется.</p> <p>Дайте определение понятию “дерево решений” и объясните, как оно работает.</p> <p>Что такое случайный лес и как он работает?</p> <p>Что представляет собой метод k-средних и для чего он используется?</p> <p>В каких случаях используется метод опорных векторов?</p>

**5.2. Перечень работ, выполняемых по дисциплине (модулю, практике, НИР) - эссе, рефераты, практические и расчетно-графические работы, курсовые работы, проекты и др.**

По дисциплине предусмотрено выполнение 8 лабораторных работ (ОПК-4-У1, УК-2-У1, ПК-6-У1, ОПК-4-В1):

Лабораторная работа № 1 "Форматы представления данных"

Лабораторная работа № 2 "Python. Загрузка данных"

Лабораторная работа № 3 "Python. Освоение библиотеки SciKit-Learn. Модели классификации"

Лабораторная работа № 4 "Python. Освоение библиотеки SciKit-Learn. Модели регрессии"

Лабораторная работа № 5 "Python. Освоение библиотеки SciKit-Learn. Методы кластеризации"

Лабораторная работа № 6 "Изучение метрик ML"

Лабораторная работа № 7 "Методы оптимизации в ML"

Лабораторная работа № 8 «Работа с окружениями»

По каждой работе обучающиеся оформляют и защищают отчет

Для выполнения работ обучающиеся самостоятельно осваивают соответствующий учебный материал

**5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)**

Пример экзаменационного билета:

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ "МИСиС"

Кафедра инженерной кибернетики

Экзаменационный билет № 9

По дисциплине МАШИННОЕ ОБУЧЕНИЕ

1. Как использовать стандартный модуль csv и функцию csv.reader () в Python для загрузки данных?

2. Построить и охарактеризовать алгоритм линейной регрессии с квадратичной функцией потерь

Зав.кафедрой \_\_\_\_\_

Список вопросов, выносимых на экзамен

Теоретический раздел

1. Градиентный бустинг. Построение градиентного бустинга
2. Метод ближайших соседей
3. Обучение без учителя. Кластеризация с помощью алгоритма K-means
4. Метод понижения размерностей с помощью PCA.
5. Алгоритмы umap и t-SNE
6. Что такое MapReduce и в каких областях применяется?
7. Обучение с учителем: постановка задачи. Задача классификации и регрессии
8. Лес решающих деревьев (Random forest), метод построения random forest
9. Решающие деревья. Построение решающих деревьев.
10. Метрики качества для задач классификации.
11. Метрики качества для задач регрессии.
12. Кросс-валидация. Виды кросс-валидаций.
13. Проблема переобучения. Регуляризация вида L1 и L2
14. Линейные модели в задачах классификации. Функционалы ошибок. Логистическая регрессия. Градиентный спуск. Стохастический градиентный спуск.

15. Линейные модели в задачах регрессии. Аналитическое решение. Функция ошибки. Градиентный спуск.

Стохастический градиентный спуск.

16. Как использовать стандартный модуль csv и функцию csv.reader () в Python для загрузки данных?

Практический раздел

1. Запрограммировать на Python без использования внешних библиотек градиентный спуск для случая одномерной линейной регрессии.
2. Запрограммировать на Python логистическую регрессию без использования внешних библиотек.
3. Запрограммировать регуляризатор L1 и L2 с помощью библиотеки scikit-learn. Показать, что при регуляризации L1 некоторые параметры модели устремляются к нулю.
4. Применить к данным кросс-валидацию, используя библиотеку scikit-learn
5. Натренировать регрессионную модель и оценить ее качество (например, MSE) с помощью библиотеки scikit-learn
6. Натренировать модель классификации и оценить ее качество (например, Accuracy) с помощью библиотеки scikit-learn
7. Обучить решающее дерево с помощью scikit-learn. Показать, что с увеличением глубины дерево переобучается.
8. Натренировать модель random forest. Оценить качество полученной модели
9. Запрограммировать функцию факториала итеративным и рекурсивным способами
10. Натренировать модель градиентного бустинга с помощью библиотеки catboost. Показать, что при увеличении числа базовых алгоритмов, градиентный бустинг переобучается.
11. Обучить метод ближайших соседей для задачи классификации с помощью библиотеки scikit-learn. Проанализировать, как будет меняться качество алгоритма при изменении количества соседей.
12. Запрограммировать K-means с помощью scikit-learn и визуализировать кластеры
13. Запрограммировать PCA с помощью scikit-learn. Проверить, как влияет PCA после обучения модели.
14. Запрограммировать с использованием scikit-learn и визуализировать
15. Рассказать и показать основы работы с Nadoop
16. Построить и охарактеризовать алгоритм линейной регрессии с квадратичной функцией потерь

**5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)**

Требования к оцениванию: экзамен

Порядок, определяющий процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций, определен в Положении «О текущем контроле успеваемости и промежуточной аттестации обучающихся, текущем контроле посещения обучающимися аудиторных занятий в НИТУ «МИСиС» П 239.09-18, выпуск 2».

Система оценивания результатов освоения дисциплины: балльно-рейтинговая.

Критерии оценивания

--письменного отчета по лабораторной работе

«Отлично»

Отчет представлен в установленные сроки и в полном объеме.

«Хорошо»

Отчет представлен в установленные сроки в полном объеме.

или

Отчет представлен не в установленные сроки в полном объеме.

«Удовлетворительно»

Отчет представлен не в установленные сроки, но в полном объеме.

или

Отчет представлен в установленные сроки в достаточном объеме, имеются замечания по содержанию отдельных пунктов.

«Неудовлетворительно»

Отчет не представлен.

или

Отчет представлен в неполном объеме (отсутствуют отчетные материалы по отдельным пунктам индивидуального задания).

или

-- на защите отчетов по лабораторным работам

«Отлично»:

Обучающийся при ответах демонстрирует системность и глубину знаний.

Обучающийся владеет научной терминологией в области машинного обучения и data science, стилистически грамотно, логически правильно и исчерпывающе освещает поставленные вопросы.

Дает полные и аргументированные ответы на дополнительные вопросы.

«Хорошо»:

Обучающийся при ответах демонстрирует достаточную полноту знаний, при наличии лишь несущественных неточностей в освещении отдельных вопросов.

Обучающийся владеет научной терминологией в области машинного обучения и data science, стилистически грамотно, логически правильно и достаточно полно (пропуская или неточно излагая отдельные существенные детали) освещает вопросы.

При ответах на дополнительные вопросы недостаточно полно раскрывает сущность вопроса, допускает незначительные ошибки, но исправляется при наводящих вопросах.

«Удовлетворительно»:

Обучающийся при ответах демонстрирует достаточные знания по основным вопросам коллоквиума, но допускает при этом неточности.

Обучающийся в достаточной мере использует научную терминологию, в основном структурированно и содержательно излагает сущность вопросов, допуская при этом незначительные ошибки, которые при наводящих вопросах может исправить.

**6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ****6.1. Рекомендуемая литература****6.1.1. Основная литература**

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Козлова Т. В., Саркисов К. А., Козлов А. Н., Волков Д. В.	Студент в среде e-Learning: учебно-методический комплекс	Электронная библиотека	Москва: Евразийский открытый институт, 2011
Л1.2	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Пабlishер, 2017

**6.1.2. Дополнительная литература**

	Авторы, составители	Заглавие	Библиотека	Издательство, год
--	---------------------	----------	------------	-------------------

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Сузи Р. А.	Язык программирования Python: учебное пособие	Электронная библиотека	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)  Бином. Лаборатория знаний, 2007
Л2.2	Шелудько В. М.	Язык программирования высокого уровня Python: функции, структуры данных, дополнительные модули: учебное пособие	Электронная библиотека	Ростов-на-Дону, Таганрог: Южный федеральный университет, 2017

### 6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	ЭОИС Canvas	<a href="https://lms.misis.ru/login/canvas">https://lms.misis.ru/login/canvas</a>
Э2	ГОСТ 7.32-2017 МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ. Система стандартов по информации, библиотечному и издательскому делу – СИБИБД. ОТЧЕТ О НАУЧНО-ИС-СЛЕДОВАТЕЛЬСКОЙ РАБОТЕ. Структура и правила оформления.	<a href="http://docs.cntd.ru/document/1200157208">http://docs.cntd.ru/document/1200157208</a>
Э3	ГОСТ Р 2.105-2019 Единая система конструкторской документации. Общие требования к текстовым документам	<a href="https://allgosts.ru/01/110/gost_r_2.105-2019.pdf">https://allgosts.ru/01/110/gost_r_2.105-2019.pdf</a>
Э4	Машинное обучение электронный ресурс	<a href="https://en.wikipedia.org/wiki/Machine_learning">https://en.wikipedia.org/wiki/Machine_learning</a>
Э5	Машинное обучение с Python	<a href="https://coderlessons.com/tutorials/python-technologies/uznaite-mashinnoe-obuchenie-s-python/mashinnoe-obuchenie-s-python-kratkoe-rukovodstvo">https://coderlessons.com/tutorials/python-technologies/uznaite-mashinnoe-obuchenie-s-python/mashinnoe-obuchenie-s-python-kratkoe-rukovodstvo</a>
Э6	Hadoop — MapReduce	<a href="https://coderlessons.com/tutorials/bolshie-dannye-i-analitika/uchitsia-hadoop/hadoop-mapreduce">https://coderlessons.com/tutorials/bolshie-dannye-i-analitika/uchitsia-hadoop/hadoop-mapreduce</a>

### 6.3 Перечень программного обеспечения

П.1	LMS Canvas
П.2	MS Teams
П.3	Python
П.4	Hadoop

### 6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Полнотекстовые российские научные журналы и статьи:
И.2	— Научная электронная библиотека eLIBRARY <a href="https://elibrary.ru/">https://elibrary.ru/</a>
И.3	— Полнотекстовые деловые публикации информагентств и прессы по 53 отраслям <a href="https://polpred.com/news">https://polpred.com/news</a>
И.4	Иностранные базы данных (доступ с IP адресов МИСиС):
И.5	— аналитическая база (индексы цитирования) Web of Science <a href="https://apps.webofknowledge.com">https://apps.webofknowledge.com</a>
И.6	— аналитическая база (индексы цитирования) Scopus <a href="https://www.scopus.com/">https://www.scopus.com/</a>
И.7	— наукометрическая система InCites <a href="https://apps.webofknowledge.com">https://apps.webofknowledge.com</a>
И.8	— научные журналы издательства Elsevier <a href="https://www.sciencedirect.com/">https://www.sciencedirect.com/</a>

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
Любой корпус Мультимедийная	Учебная аудитория для проведения занятий лекционного типа и/или для проведения практических занятий:	комплект учебной мебели до 36 мест для обучающихся, мультимедийное оборудование, магнитно-маркерная доска, рабочее место преподавателя, ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus
Читальный зал электронных ресурсов		комплект учебной мебели на 55 мест для обучающихся, 50 ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus.



Б-904а	Компьютерный класс	20 стационарных компьютеров (core i5-3470 8gb RAM), пакет лицензионных программ MS Office, демонстрационное оборудование: доска, проектор мультимедийный, экран, колонки, комплект учебной мебели
Б-902	Учебная аудитория	12 стационарных компьютеров (2 x core i5-3470 8gb RAM, 10 x ryzen5 2400g 32gb RAM), пакет лицензионных программ MS Office, демонстрационное оборудование: доска, проектор мультимедийный, комплект учебной мебели на 19 мест

### 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

В соответствии с расписанием занятий студенты самостоятельно прорабатывают материал выданный перед очередной лекцией (читают литературу, проводят поисковые исследования в сети Internet). Перед выполнением лабораторных работ студенты знакомятся с работой и самостоятельно прорабатывают необходимый материал. После выполнения работ студенты самостоятельно проводят анализ проделанной работы и готовят отчёты по ним.

Лекции читаются в аудиториях с мультимедийным оборудованием с использованием электронных презентаций, представляющих собой опорный иллюстрированный конспект по соответствующей теме. Кроме лекционных материалов преподаватель может рекомендовать к изучению материалы, которые учащийся самостоятельно может получить из перечня профессиональных баз данных и информационных справочных систем (см. соотв. раздел).

Лабораторные работы (ЛР) проводятся в специализированных классах (лабораториях) кафедры инженерной кибернетики. По каждой ЛР проводится защита работы, в ходе которой учащийся демонстрирует полученные результаты, и отвечает на поставленные вопросы.