

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по безопасности и общим вопросам

Дата подписания: 30.01.2023 16:41:18

Уникальный программный ключ:

d7a26b9e8ca85e98bc3de2ab454b4659d061f749

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский технологический университет «МИСиС»

Рабочая программа дисциплины (модуля) Интеллектуальный анализ данных

Закреплена за подразделением Кафедра магистерская школа информационных бизнес систем
Направление подготовки 09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
Профиль Экосистема больших данных для цифровой трансформации

Квалификация **Магистр**
Форма обучения **очная**
Общая трудоемкость **5 ЗЕТ**
Часов по учебному плану 180
в том числе: Формы контроля в семестрах:
экзамен 3
аудиторные занятия 52
самостоятельная работа 74
часов на контроль 54

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		Итого	
	УП	РП	УП	РП
Неделя	19			
Вид занятий	УП	РП	УП	РП
Лекции	16	16	16	16
Практические	36	36	36	36
В том числе инт.	36		36	
Итого ауд.	52	52	52	52
Контактная работа	52	52	52	52
Сам. работа	74	74	74	74
Часы на контроль	54	54	54	54
Итого	180	180	180	180

Программу составил(и):

асс., Чукреев К.Н.

Рабочая программа

Интеллектуальный анализ данных

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 09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ (приказ от 05.03.2020 г. № 95 о.в.)

Составлена на основании учебного плана:

09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ, 09.04.02-МИСТ-22-2.plx Экосистема больших данных для цифровой трансформации, утвержденного Ученым советом ФГАОУ ВО НИТУ "МИСиС" в составе соответствующей ОПОП ВО 22.09.2022, протокол № 8-22

Утверждена в составе ОПОП ВО:

09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ, Экосистема больших данных для цифровой трансформации, утвержденной Ученым советом ФГАОУ ВО НИТУ "МИСиС" 22.09.2022, протокол № 8-22

Рабочая программа одобрена на заседании

Кафедра магистерская школа информационных бизнес систем

Протокол от 24.06.2022 г., №10

Руководитель подразделения М.И. Нежурина

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Целью дисциплины «Интеллектуальный анализ данных» является формирование у студентов исчерпывающего практико-ориентированного представления об областях применения методов машинного обучения и анализа данных. Для достижения указанной цели в процессе изучения дисциплины решаются следующие задачи:
1.2	формирование понимания задач машинного обучения, мотивации к их решению и практических приложений этих задач;
1.3	получение базовых практических навыков постановки и решения задач анализа данных как инженерных задач;
1.4	ознакомление с актуальными задачи и некоторыми последними достижениями в области анализа данных.

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В.ДВ.02
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Научно-исследовательская работа. Учебный проект	
2.1.2	Программные платформы и технологии больших данных	
2.1.3	Управление проектами в современной компании	
2.1.4	Языки программирования для работы с большими данными	
2.1.5	Информационные технологии в офисной деятельности	
2.1.6	Новые направления и технологии современных СУБД	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Научно-исследовательская работа	
2.2.2	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.3	Производственная проектная практика	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ПК-4: Способен выполнять работы по сбору, обработке и анализу больших данных в междисциплинарных областях	
Знать:	
ПК-4-32	Современные методы и инструментальные средства анализа больших данных
ПК-4-31	Современный опыт использования анализа больших данных
Уметь:	
ПК-4-У2	Проводить анализ больших данных
ПК-4-У1	Проводить сравнительный анализ методов и инструментальных средств анализа данных
ПК-4-У3	Решать проблемы переобучения и недообучения алгоритма
ОПК-8: Способен осуществлять эффективное управление разработкой программных средств и проектов, демонстрировать практические навыки для решения сложных задач, выполнения сложного проектирования, а также проведения комплексных исследований, знание экономических, организационных и управленческих вопросов, таких как: управление проектами, рисками и изменениями	
Уметь:	
ОПК-8-У2	Решать задачи машинного обучения
ОПК-8-У1	Анализировать исходные данные и формулировать постановки задач машинного обучения
Владеть:	
ОПК-8-В1	Навыками реализации методов машинного обучения

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
	Раздел 1. Анализ данных и постановка задачи машинного обучения							

1.1	Постановка задачи машинного обучения. Линейная алгебра для задач машинного обучения. Методы оптимизации /Лек/	3	2	ПК-4-31 ПК-4-32	Л1.3 Л1.2Л2.3 Л2.4 Э1 Э2 Э8			
1.2	Матрично-векторное дифференцирование. Применение методов линейной алгебры в задачах машинного обучения. Оптимизационная задача классификации, логит-функция, логистическая регрессия. Ансамблевые методы и их применение /Пр/	3	4	ПК-4-У2 ПК-4-У3 ОПК-8-У2	Л1.2Л2.2 Э6 Э8			
1.3	Изучение источников. Выполнение домашнего задания 1. /Ср/	3	12	ОПК-8-У2 ОПК-8-В1	Л1.2Л1.3 Л2.4 Э1 Э6 Э8			Р1
1.4	Метрические алгоритмы (алгоритм k ближайших соседей), их применение. Использование ядерной функции для изменение метрического пространства /Пр/	3	2	ПК-4-У2 ОПК-8-У2	Л2.4Л2.5 Э1 Э6 Э8			
1.5	Выполнение домашнего задания 2. /Ср/	3	12	ПК-4-У2 ПК-4-У3	Л2.5 Э6 Э8			Р2
	Раздел 2. Алгоритмы машинного обучения							
2.1	Простые методы машинного обучения /Лек/	3	2	ПК-4-32	Л2.3 Л2.4 Э1 Э3 Э4			
2.2	Семинар. Простые методы машинного обучения /Пр/	3	2	ОПК-8-У1 ОПК-8-У2	Л2.3 Л2.4 Э1 Э3 Э4 Э6			
2.3	Изучение источников. Соревнование по машинному обучению №1. /Ср/	3	16	ПК-4-31 ПК-4-32 ПК-4-У1 ПК-4-У2 ПК-4-У3 ОПК-8-У1 ОПК-8-У2 ОПК-8-В1	Л2.3 Л2.4 Э1 Э3 Э4 Э5			Р3
2.4	Семинар. Линейные модели /Пр/	3	2	ПК-4-У1 ПК-4-У2	Л2.4 Э1 Э3 Э4 Э6			
2.5	Решающие деревья /Лек/	3	2	ПК-4-32	Л2.4 Э1			
2.6	Семинар. Решающие деревья /Пр/	3	2	ПК-4-У1 ПК-4-У2	Л2.4 Э6			
	Раздел 3. Особенности применения методов машинного обучения							
3.1	Семинар. Оценочные метрики /Пр/	3	4	ПК-4-У2 ПК-4-У3	Л2.4 Э1 Э6			
3.2	Обучение без учителя /Лек/	3	2	ПК-4-32	Л2.4 Э1			
3.3	Семинар. Обучение без учителя /Пр/	3	4	ПК-4-У2 ОПК-8-У2	Л2.3 Л2.4 Э1 Э6			
3.4	Валидация модели и работа с признаками /Лек/	3	2	ПК-4-31	Л2.4 Э1			

3.5	Семинар. Валидация модели и работа с признаками /Пр/	3	4	ПК-4-У2	Л2.3 Л2.4 Э1 Э6			
3.6	Изучение источников. Соревнование по машинному обучению №2. /Ср/	3	16	ПК-4-У3 ОПК -8-В1	Л2.3 Л2.4 Э1 Э6 Э7			Р4
Раздел 4. Применение методов машинного обучения в индустрии. Глубинное обучение								
4.1	Нейронные сети в машинном обучении /Лек/	3	2	ПК-4-32	Л1.1Л2.4			
4.2	Нейронные сети в машинном обучении. Сверточные нейронные сети. /Пр/	3	4	ПК-4-У2 ОПК -8-У1	Л1.1 Э3 Э6			
4.3	А/В тестирование в машинном обучении /Лек/	3	4	ПК-4-31	Л2.4 Э1			
4.4	Семинар. А/В тестирование в машинном обучении /Пр/	3	4	ПК-4-У2	Л2.4 Э1 Э6			
4.5	Изучение источников, подготовка к итоговому тесту. /Ср/	3	18	ПК-4-31	Л1.3 Л1.1 Л1.2Л2.2 Л2.3 Л2.4 Л2.5 Э1 Э3 Э4 Э6			
4.6	Итоговый тест /Пр/	3	4	ПК-4-31 ПК-4-32			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	Итоговый тест	ПК-4-31;ПК-4-32	<p>Тест состоит из теоритических вопросов преимущественно открытого типа.</p> <p>Темы тестовых вопросов:</p> <ol style="list-style-type: none"> 1. Формула Байеса, условная вероятность 2. Мультиклассовая классификация 3. Регрессия с регуляризацией 4. Метрики классификации, регрессии 5. Решающие деревья: ансамбли, градиентный бустинг 6. Метрики классификации и регрессии. 7. Обучение без учителя: теория, алгоритмы 8. Разделение выборок для обучения, валидация 9. Работа с признаками 10. Нейронные сети: от основ до сверточных сетей 11. Регуляризация в обучении 12. Метрики качества 13. KDE метод 14. Метод максимального правдоподобия 15. Простые методы машинного обучения: ближайшие соседи, линейные модели 16. Прототипирование задачи по ML на Python

КМ2	Экзамен	ПК-4-31;ПК-4-32;ОПК-8-У1;ОПК-8-У2;ПК-4-У1;ПК-4-У3	<p>Темы теоретических вопросов для подготовки к экзамену:</p> <p>ПК-4-32</p> <ol style="list-style-type: none"> 1. Формула Байеса, условная вероятность 2. Мультиклассовая классификация 3. Регрессия с регуляризацией 4. Метрики классификации, регрессии 5. Решающие деревья: ансамбли, градиентный бустинг 6. Метрики классификации и регрессии. 7. Обучение без учителя: теория, алгоритмы 8. Разделение выборок для обучения, валидация 9. Работа с признаками 10. Нейронные сети: от основ до сверточных сетей 11. Регуляризация в обучении 12. Метрики качества <p>ПК-4-31</p> <ol style="list-style-type: none"> 13. KDE метод 14. Метод максимального правдоподобия 15. Простые методы машинного обучения: ближайшие соседи, линейные модели 16. Прототипирование задачи по ML на Python <p>Темы для практических заданий:</p> <p>ПК-4-У1</p> <ol style="list-style-type: none"> 1. МНК-оценки параметров регрессии 2. Доверительные интервалы и оценка значимости коэффициентов регрессии 3. Оценка адекватности уравнения по остаточной дисперсии <p>ОПК-8-У2</p> <ol style="list-style-type: none"> 4. Информационная матрица и матрица ошибок 5. Оценка дисперсий коэффициентов 6. Система нормальных уравнений <p>ПК-4-У3</p> <ol style="list-style-type: none"> 7. Вычисление по функции потерь 8. Остатки, их график 9. Оценка однородности дисперсии <p>ОПК-8-У1</p> <ol style="list-style-type: none"> 10. Постановка задачи машинного обучения. 11. Количество наблюдаемых данных для объяснения регрессии
-----	---------	---	--

5.2. Перечень работ, выполняемых по дисциплине (Курсовая работа, Курсовой проект, РГР, Реферат, ЛР, ПР и т.п.)

Код работы	Название работы	Проверяемые индикаторы компетенций	Содержание работы
P1	Домашнее задание 1	ОПК-8-У2;ОПК-8-В1;ПК-4-У2	Решение задач по темам оптимизации, матрично-векторного дифференцирования и линейной алгебры
P2	Домашнее задание 2	ОПК-8-У2;ОПК-8-В1;ПК-4-У2	Решение задач по основным свойствам распределений, на Формулу Байесса и метод KDE
P3	Соревнование №1	ОПК-8-У1;ОПК-8-У2;ПК-4-У1;ПК-4-У2;ПК-4-У3	Соревнование по машинному обучению "Предсказание, является ли строка фамилией"
P4	Соревнование №2	ОПК-8-В1;ПК-4-У3	Соревнование по машинному обучению "Распознавание изображений"

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

Экзамен проводится в письменном виде. Продолжительность экзамена 1 астрономический час. В билете 2 вопроса: теоретический вопрос + практическое задание

Пример экзаменационного билета:

1. Постановка задачи машинного обучения. Метод ближайших соседей.
2. Регуляризация в обучении. L1 (Ridge) и L2 (Lasso) регуляризация.

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

По дисциплине промежуточная аттестация предусмотрена в форме экзамена в 3 семестре.

Шкала оценивания - 100-балльная, распределение баллов по видам деятельности:

1. Домашняя работа 1 максимум 15 баллов, при решении задач баллы снимаются за неверную формальную постановку, неверный ход решения, решение без использования рекомендуемого ПО.
2. Домашняя работа 2 максимум 15 баллов, при решении задач баллы снимаются за неверную формальную постановку, неверный ход решения, решение без использования рекомендуемого ПО.
3. Теоретический тест - максимум 10 баллов.
4. Соревнование №1 по машинному обучению - максимум 15 баллов.
5. Соревнование №2 по машинному обучению - максимум 15 баллов.

Критерии оценивания экзамена:

Общее количество баллов за экзамен – максимум 30. В каждом билете 2 вопроса по 15 баллов максимум.

Шкала перевода в 5-балльную систему оценки:

более 85 - отлично,

70-84 - хорошо,

61-69 - удовлетворительно,

0-60 - неудовлетворительно

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

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Яхьяева Г. Э.	Основы теории нейронных сетей	Электронная библиотека	Москва: Национальный Открытый Университет «ИНТУИТ», 2016
Л1.2	Смирнов А. П.	Методы оптимизации. Алгоритмические основы задач оптимизации: курс лекций	Библиотека МИСиС	М.: Изд-во МИСиС, 2014

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Романиков А. Н.	Линейная алгебра: учебное пособие	Электронная библиотека	Москва: Московский государственный университет экономики, статистики и информатики, 2007
Л2.2	Казанская О. В., Юн С. Г., Альсова О. К.	Модели и методы оптимизации: учебное пособие	Электронная библиотека	Новосибирск: Новосибирский государственный технический университет, 2012
Л2.3	Мыльников Л. А., Краузе Б., Кютц М., Баде К., Шмидт И. А.	Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография	Электронная библиотека	Москва: Библио-Глобус, 2017
Л2.4	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017
Л2.5	Мельниченко А. С.	Математическая статистика и анализ данных (N 3431): учеб. пособие	Электронная библиотека	М.: [МИСиС], 2018

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Онлайн-курс "Машинное обучение и анализ данных"	https://ru.coursera.org/specializations/machine-learning-data-analysis
Э2	Курс по математике для Data Science	https://skillfactory.ru/math-stat-for-ds
Э3	Pattern Recognition and Machine Learning, Springer, 2006. Cristopher Bishop.	http://statweb.stanford.edu/~tibs/ElemStatLearn/

Э4	Machine Learning: a Probabilistic Perspective. Kevin Patrick Murphy	http://www.cs.ubc.ca/~murphyk/MLbook/
Э5	Machine Learning in Python	http://scikit-learn.org
Э6	Jupyter Notebook	http://jupyter.org
Э7	Платформа для соревнований по машинному обучению	http://kaggle.com
Э8	Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных	http://www.machinelearning.ru/

6.3 Перечень программного обеспечения

П.1	Microsoft Office
П.2	Python
П.3	Anaconda
П.4	LMS Canvas

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Реферативная математическая база данных: https://zbmath.org/
-----	--

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
М-105	Мультимедийный тренинговый учебный класс:	рабочее место преподавателя, оборудованное персональным компьютером, пакет лицензионных программ MS Office; проектор; экран; маркерная доска; комплект учебной мебели
М-102	Аудитория для самостоятельной работы студентов и курсового проектирования:	Комплект учебной мебели на 12 рабочих мест, ноутбуки с подключением к сети «Интернет» и доступом в электронную информационно-образовательную среду университета
М-105	Мультимедийный тренинговый учебный класс:	рабочее место преподавателя, оборудованное персональным компьютером, пакет лицензионных программ MS Office; проектор; экран; маркерная доска; комплект учебной мебели

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

Традиционным и наиболее распространённым вариантом построения структуры занятия является урок, включающий три основных этапа: просмотр презентаций и конспектов лекций, получение и отработка практических навыков, подведение итогов и рефлексия.

Лекционные занятия включают в себя разбор теоретического материала и фундаментальных математических основ интеллектуального анализа данных.

Для интерактивной работы на практических занятиях и для выполнения домашних заданий используется Jupyter Notebook - инструмент для представления данных и их анализа.

Основным видом текущего контроля в дисциплине по отработке практических умений и навыков являются соревнования по машинному обучению, которые проводятся на платформе Kaggle — системе организации конкурсов по исследованию данных, которая также является социальной сетью специалистов по обработке данных и машинному обучению.