

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 29.08.2023 14:30:23

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)

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	4 ЗЕТ	
Часов по учебному плану	144	Формы контроля в семестрах: экзамен 2
в том числе:		
аудиторные занятия	52	
самостоятельная работа	56	
часов на контроль	36	

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	2 (1.2)		Итого	
	УП	РП	УП	РП
Неделя	18			
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	56	56	56	56
Часы на контроль	36	36	36	36
Итого	144	144	144	144

Программу составил(и):

к.э.н., доц., Жукова Л.В.

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 05.03.2022 г. № 95 о.в.)

Составлена на основании учебного плана:

09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ, 09.04.02-МИСТ-23-2.plx Экосистема больших данных для цифровой трансформации, утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ, Экосистема больших данных для цифровой трансформации, утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

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	Архитектуры систем хранения данных	
2.1.2	Информационные технологии в офисной деятельности	
2.1.3	Новые направления и технологии современных СУБД	
2.1.4	Практика моделирования бизнес-процессов	
2.1.5	Специальные главы математики. Часть 1	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Научно-исследовательская работа. Исследовательский проект	
2.2.2	Практика машинного обучения. Дизайн новых материалов	
2.2.3	Практика машинного обучения. Материаловедение	
2.2.4	Прикладные области анализа больших данных. Дизайн новых материалов	
2.2.5	Прикладные области анализа больших данных. Материаловедение	
2.2.6	Научно-исследовательская работа	
2.2.7	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.8	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	Введение в интеллектуальный анализ данных: Big data, типы данных, структура и формат данных. Источники данных /Лек/	2	1	ПК-4-31 ПК-4-32	Л1.1Л2.1 Л2.2 Э1 Э2 Э8			Р3
1.2	Решение прикладных задач с помощью Data Science: базовые методики и модели. Инструменты анализа /Лек/	2	2	ОПК-8-У2 ОПК-8-В1	Л1.1Л2.2 Э1 Э6 Э8			
1.3	Предобработка данных: заполнение пропусков, валидация, нормализация, выявление выбросов и аномалий /Пр/	2	4	ОПК-8-У2 ОПК-8-В1	Л1.1Л2.2 Э1 Э6 Э8			Р3
	Раздел 2. Постановка и решение регрессионных задач							
2.1	Линейная и нелинейная регрессия /Лек/	2	1	ПК-4-У1 ПК-4-У2	Л2.2 Э1 Э2 Э4			
2.2	Построение модели линейной и нелинейной регрессии /Пр/	2	4	ОПК-8-У1 ОПК-8-У2				
2.3	Нейросетевая регрессия /Лек/	2	1	ПК-4-У1 ПК-4-У2	Л2.1 Л2.2 Э1 Э3 Э4			
2.4	Построение модели нейросетевой регрессии /Пр/	2	2	ОПК-8-У1 ОПК-8-У2	Л2.1 Л2.2 Э1 Э3 Э4			
2.5	Проверка качества и адекватности модели регрессии /Пр/	2	2	ОПК-8-У1 ОПК-8-У2	Л2.1 Л2.2 Э1 Э3 Э4 Э6			Р3
	Раздел 3. Методы классификации							
3.1	Статистические методы: Логистическая регрессия /Лек/	2	1	ПК-4-У1 ПК-4-У2	Л2.2 Э1 Э3 Э4			
3.2	Построение модели логистической регрессии /Пр/	2	2	ОПК-8-У1 ОПК-8-У2	Л2.2 Э1 Э3 Э4			
3.3	Методы машинного обучения: метод опорных векторов, Байесовский классификатор /Лек/	2	2	ПК-4-У1 ПК-4-32	Л2.1 Л2.2 Э1 Э3 Э4			
3.4	Построение модели классификации методами машинного обучения /Пр/	2	2	ОПК-8-У2 ОПК-8-В1 ПК-4-У2	Л2.1 Л2.2 Э1 Э3 Э4			
3.5	Домашнее задание 1 /Ср/	2	20	ОПК-8-У2 ОПК-8-В1 ПК-4-У2	Л2.1 Л2.2 Э1 Э3 Э4			Р1
3.6	Алгоритмы классификации: деревья решений и случайный лес /Лек/	2	2	ПК-4-32	Л2.1 Л2.2 Э1			Р3
3.7	Построение модели классификации /Пр/	2	2	ОПК-8-У2 ОПК-8-В1	Л2.1 Л2.2 Э1 Э3 Э4			

3.8	Оценка качества классификации, ROC кривые, ошибки классификации /Пр/	2	2	ПК-4-32	Л2.1 Л2.2 Э1 Э3 Э4			Р3
Раздел 4. Обучение без учителя: кластеризация								
4.1	Методы кластеризации: иерархические и неиерархические алгоритмы /Лек/	2	2	ПК-4-32	Л2.2 Э1			
4.2	Построение кластеров данных /Пр/	2	4	ОПК-8-У2 ПК-4-У2	Л2.1 Л2.2 Э1 Э6			Р3
4.3	Проверка качества и адекватности результатов кластеризации. Инструмент статистических гипотез ANOVA/MANOVA для проверки адекватности модели /Пр/	2	4	ОПК-8-У2 ПК-4-У2	Л2.1 Л2.2 Э1 Э6			
Раздел 5. Анализ временных рядов								
5.1	Введение в анализ временных рядов. Модели динамики на основе декомпозиции. АРПСС модели /Лек/	2	2	ПК-4-31 ПК-4-32				
5.2	Построение модели динамики ряда данных /Пр/	2	2	ОПК-8-У2				Р3
5.3	Прогнозирование и предсказание. Интервальные прогнозы /Пр/	2	2	ОПК-8-У1 ОПК-8-У2 ПК-4-У1 ПК-4-У2 ПК-4-У3				
5.4	Домашнее задание 2 /Ср/	2	20	ОПК-8-У1 ОПК-8-У2 ПК-4-У3 ПК-4-У1 ПК-4-У2				Р2
Раздел 6. Применение методов машинного обучения в индустрии. Глубинное обучение								
6.1	Нейронные сети в машинном обучении /Лек/	2	2	ПК-4-31 ПК-4-32	Л2.2			
6.2	Применение нейронных сетей в задачах обработки текста, видео, аудио изображений /Пр/	2	2	ОПК-8-У2	Л1.2 Э3 Э6			Р3
6.3	Подготовка к итоговому тестированию /Ср/	2	16	ПК-4-31 ПК-4-32				
6.4	Итоговый тест /Пр/	2	2	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	Тест состоит из теоретических и практических вопросов преимущественно открытого типа.

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> <ol style="list-style-type: none"> 1. Этапы интеллектуального анализа данных 2. BIG DATA – особенности обработки, подготовки к анализу, особенности математического анализа и моделирования 3. Типы данных, источники данных и их достоверность. Подходы к проверке качества входных данных 4. Методы заполнения пропусков в данных, нормализация, валидация данных. 5. Выбросы и аномалии в данных и в моделях: способы выявления 6. Линейная и нелинейная регрессия с количественными и качественными переменными, нейросетевая регрессия: построение, метрики качества и проверка адекватности моделей 7. Классификация: модели логистической регрессии, метод опорных векторов, байесовский классификатор. Метрики качества классификации 8. Классификация: деревья решений и случайный лес (random forest) 9. Кластеризация методами математической статистики и методами машинного обучения, метрики качества кластеризации, интерпретация результатов 10. Временные ряды, основные подходы к моделированию и прогнозированию. Цели, задачи и виды прогнозов: прогнозирование и предсказание 11. Глубинное обучение. Нейросети. Архитектура нейросетей. 12. Обработка неструктурированных данных, текстовой информации, видео и аудио информации.
-----	---------	---	---

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 теме регрессионного анализа, классификации и группировке данных, построению нейросетевой модели
P2	Домашнее задание 2	ОПК-8-У2;ОПК-8-В1;ПК-4-У2;ПК-4-У1;ПК-4-У3	Решение задач по построению моделей динамики
P3	Работа на семинаре (1-7)	ОПК-8-У2;ОПК-8-В1;ПК-4-У2	Выполнение тестовых заданий с теоретическими и практическими вопросами

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

Экзамен проводится в письменном виде. Продолжительность экзамена 1 астрономический час.

Экзаменационное задание содержит 2 части:

- 1 - теоретическая часть, которая состоит из 15 тестовых вопросов с открытой, закрытой формами ответов,
- 2 - практическое задание на полученных от преподавателя данных или ссылках на данные, находящиеся в открытых источниках информации.

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. Работа на семинарах - максимум 25 баллов.
2. Домашняя работа 1 максимум 20 баллов, при решении задач баллы снимаются за неверную формальную постановку, неверный ход решения, решение без использования рекомендуемого ПО.
3. Домашняя работа 2 максимум 15 баллов, при решении задач баллы снимаются за неверную формальную постановку, неверный ход решения, решение без использования рекомендуемого ПО.
4. Экзамен - максимум 40 баллов.

Критерии оценивания экзамена:

Ответы на тестовые вопросы оцениваются в 20 баллов максимум, по 1 или 2 баллов за верный ответ на каждый из вопросов.

Задача максимально оценивается в 20 баллов, при этом:

- Правильное решение (возможны несущественные недочеты) - 16– 20 баллов,
- Сущности определены правильно, но есть ошибки в применении ММ - 11 – 15 баллов,
- Определены не все сущности. Суть методов и элементы определены верно - 6 –10 баллов,
- Сущности неправильные, методы выбраны не верно, но есть верные общие принципы - 1 – 5 баллов,
- Проблемы с пониманием сути задачи вообще - 0 баллов.

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	Герасимов А. Н., Громов Е. И., Скрипниченко Ю. С.	Эконометрика: продвинутый уровень: учебное пособие	Электронная библиотека	Ставрополь: Ставропольский государственный аграрный университет (СтГАУ), 2016
Л1.2	Келлехер Д., Тирни Б.	Наука о данных: базовый курс: учебное пособие	Электронная библиотека	Москва: Альпина Паблишер, 2020

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	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
Л2.2	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017

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	Компьютерный класс	рабочее место преподавателя, оборудованное персональным компьютером, пакет лицензионных программ MS Office; проектор; экран; маркерная доска; комплект учебной мебели
М-102	Учебная аудитория	Комплект учебной мебели на 15 рабочих мест, ноутбуки с подключением к сети «Интернет» и доступом в электронную информационно-образовательную среду университета
М-105	Компьютерный 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.