

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по безопасности и общим вопросам

Дата подписания: 30.09.2023 16:41:18

Уникальный программный ключ:

d7a26b9e8ca85e98ac3de2ab454b4659d961f749

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский технологический университет «МИСиС»

Рабочая программа дисциплины (модуля)

Языки программирования для работы с большими данными

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 **7 ЗЕТ**
Часов по учебному плану 252
в том числе: Формы контроля в семестрах:
экзамен 2
зачет 1
аудиторные занятия 72
самостоятельная работа 144
часов на контроль 36

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	1 (1.1)		2 (1.2)		Итого	
	УП	РП	УП	РП		
Неделя	18		18			
Вид занятий	УП	РП	УП	РП	УП	РП
Лекции	8	8	4	4	12	12
Практические	28	28	32	32	60	60
В том числе инт.			32		32	
Итого ауд.	36	36	36	36	72	72
Контактная работа	36	36	36	36	72	72
Сам. работа	108	108	36	36	144	144
Часы на контроль			36	36	36	36
Итого	144	144	108	108	252	252

Программу составил(и):
ассистент, Воробьев Д.А.

Рабочая программа

Языки программирования для работы с большими данными

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 Нежурина М.И.

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Целью дисциплины «Языки программирования для работы с большими данными» является: формирование у студентов представления о вычислениях, обеспечивающих основу для хранения Больших Данных, их обработки, поиска и модификации, а также методах и средствах организации таких вычислений, что обеспечивает:
1.2	– возможность оценок существующих технологий систем хранения данных в постоянно усложняющемся ИТ-окружении;
1.3	– принятие осмысленных решений о сопровождении и/или модификации таких технологий;
1.4	– выполнение процедур по сопровождению и расширению вычислительных средств, обеспечивающих работу с Большими Данными.
1.5	Обучающиеся изучат основные конструкции языков программирования Python и R, ознакомятся с тенденциями развития современных языков программирования.
1.6	Для достижения указанной цели в процессе изучения дисциплины решаются следующие задачи:
1.7	– формирование представления о вычислении и основных структурах управления вычислениями;
1.8	– формирование представления о структурах данных во взаимосвязи со структурами Больших Данных;
1.9	– получение представления об объектно-ориентированной парадигме программирования и её использовании при работе с Большими данными;
1.10	– ознакомление с концепцией функционального программирования, лежащей в основе современных ИТ-технологий;
1.11	– ознакомление с методами разработки интерфейса пользователя, включая графический вывод результатов обработки данных;
1.12	– получение умений разработки программ на изучаемых языках.

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В
2.1	Требования к предварительной подготовке обучающегося:	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Интеллектуальный анализ данных	
2.2.2	Научно-исследовательская работа. Исследовательский проект	
2.2.3	Практика машинного обучения. Материаловедение	
2.2.4	Прикладные области анализа больших данных. Материаловедение	
2.2.5	Научно-исследовательская работа	
2.2.6	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.7	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 Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, нейронные сети, планирование экспериментов, карты контроля качества
ПК-4-31 Современные методы и инструментальные средства анализа больших данных
Уметь:
ПК-4-У1 Программировать на языках высокого уровня, ориентированных на работу с большими данными: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных
Владеть:
ПК-4-В1 Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ

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. 1 семестр								
1.1	Базовый Синтаксис языка программирования Python /Лек/	1	4	ПК-4-31 ПК-4-32	Л1.1 Л1.2 Э1 Э2 Э3 Э4 Э5 Э6			
1.2	Базовый Синтаксис языка программирования Python /Пр/	1	12	ПК-4-У1	Л1.1 Л1.2 Э3			Р1,Р2,Р3
1.3	Выполнение домашнего задания №1. /Ср/	1	16	ПК-4-32 ПК-4-У1	Л1.1 Л1.2Л2.1 Л2.2 Л2.3 Э3			
1.4	Выполнение домашнего задания №2. /Ср/	1	16	ПК-4-32 ПК-4-У1	Л1.1 Л1.2Л2.1 Л2.3 Э3			
1.5	Выполнение домашнего задания №3. /Ср/	1	16	ПК-4-32 ПК-4-У1	Л1.1 Л1.2Л2.1 Л2.3 Э3			
1.6	Углубленное изучение языка программирования Python /Лек/	1	4	ПК-4-32	Л1.1 Л1.2			
1.7	Углубленное изучение языка программирования Python /Пр/	1	16	ПК-4-У1	Л1.1 Л1.2 Э3		КМ1	Р4,Р5
1.8	Выполнение домашнего задания №4. /Ср/	1	24	ПК-4-У1	Л1.1 Л1.2Л2.1 Л2.3 Э3			
1.9	Выполнение домашнего задания №5. /Ср/	1	36	ПК-4-У1	Л1.1 Л1.2Л2.1 Л2.3 Э3			
Раздел 2. 2 семестр								
2.1	Вводная лекция-семинар. /Лек/	2	4	ПК-4-31 ПК-4-32	Л1.1 Л1.2			
2.2	Связные списки. Pandas. /Пр/	2	8	ПК-4-32 ПК-4-У1	Л1.1 Л1.2 Э3			Р6
2.3	Классификация и кластеризация. /Пр/	2	8	ПК-4-32 ПК-4-У1	Л1.1 Л1.2 Э3			Р7
2.4	Машинное обучение. /Пр/	2	8	ПК-4-32 ПК-4-У1	Л1.1 Л1.2Л2.4 Э3			Р8
2.5	Ансамблевые методы. /Пр/	2	8	ПК-4-32 ПК-4-У1	Л1.1 Л1.2Л2.4 Э3			Р9
2.6	Индивидуальный учебный проект. /Ср/	2	36	ПК-4-У1 ПК-4-В1	Л1.1 Л1.2Л2.4 Э3			

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	Защита Домашнего задания №5.	ПК-4-В1;ПК-4-У1;ПК-4-32;ПК-4-31	
КМ2	Экзамен	ПК-4-31;ПК-4-32;ПК-4-У1;ПК-4-В1	<p>Вопросы, которые могут быть заданы при защите индивидуального проекта:</p> <ol style="list-style-type: none"> 1. Почему был выбран именно такой метод классификации/кластеризации? 2. Почему данная визуализация данных наиболее подходящая? 3. Как проводилась предобработка данных? 4. Как проводилась очистка данных? 5. Какой бизнес-профит может быть получен от данной модели? 6. В каких случаях требуется нормализация. Какая бывает. Как сделать? 7. Как сделать label encoding? 8. Как сделать небинарную классификацию для бинарного метода 9. Какие подходы и библиотеки используются для линейного программирования? 10. Какие виды связанных списков вы знаете? Приведите примеры использования. 11. Какие библиотеки для визуализации вы знаете? 12. Как узнать статус выполнения расчета для долгих циклов? 13. Как выполнить моделирование методом Монте-Карло на Питоне? 14. Что делать при переобучении модели? 15. Где и как используются регулярные выражения? 16. Виды гиперпараметров и подходы для их получения 17. Как легко реализовать когортный анализ на Питоне? 18. Зачем нужен Python Global Interpreter Lock и как он работает? 19. Назовите основные классы Collections

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	Домашняя работа №1. Общий синтаксис	ПК-4-У1	Пример задачи: Напишите код для проверки года на високосность
P2	Домашняя работа №2. Хеши и коллекции	ПК-4-У1	Пример задачи: Напишите функцию, которая будет складывать цифры внутри числа
P3	Домашняя работа №3. Работа с файлами	ПК-4-У1	Пример задачи: Посчитать количество определенных слов в файле
P4	Домашняя работа №4. ООП, декораторы	ПК-4-У1	Пример задачи: Напишите программу с классом Student, в котором есть три атрибута: name, groupNumber и age. По умолчанию name = Ivan, age = 18, groupNumber = 10A. Необходимо создать пять методов: getName, getAge, getGroupNumber, setNameAge, setGroupNumber. Метод getName нужен для получения данных об имени конкретного студента, метод getAge нужен для получения данных о возрасте конкретного студента, метод setGroupNumber нужен для получения данных о номере группы конкретного студента. Метод setNameAge позволяет изменить данные атрибутов, установленных по умолчанию, метод setGroupNumber позволяет изменить номер группы, установленный по умолчанию. В программе необходимо создать пять экземпляров класса Student, установить им разные имена, возраст и номер группы.
P5	Домашняя работа №5. API, токены, боты	ПК-4-У1	Разработать 3 телеграмм-бота (калькулятор, проверка доступности ресурса и анализатор текста)

P6	Практическая работа №1. Связные списки. Pandas.	ПК-4-У1	<p>Пример задачи: Связные списки</p> <ol style="list-style-type: none"> 1) Напишите функцию, которая разворачивает связный список. На вход она принимает головную ноду, а на выход отдаёт хвостовую ноду исходного списка, но только теперь если попробовать пройти по ней, она будет новой головной. 2) Напишите функцию, которая будет на вход получать голову несортированного связного списка, а на выход возвращать голову отсортированного. <p>PANDAS</p> <ol style="list-style-type: none"> 3) Выясните, сколько в среднем выходит полный счёт по пятницам на ланч у курящих мужчин (датасет tips) 4) Загрузите iris и посмотрите разбросы по всем числовым значениям для всех трёх классов цветов (датасет iris) 5) Создайте новый столбец с плотностью населения для предоставленного фрейма. Переименуйте названия первого и третьего столбца на русский язык
P7	Практическая работа №2. Классификация и кластеризация.	ПК-4-У1	<p>Пример задачи: Классификация:</p> <ol style="list-style-type: none"> 1) Креативная часть: найти параметры и комбинации их, которые будут увеличивать ассигасу (датасет Спортсмены). 2) Техническая часть: сделать под это всё удобные обёртки. Например, это функция, которая просто принимает список фичей, а на выходе показывает результат фита. <p>Кластеризация:</p> <ol style="list-style-type: none"> 3) Креативная часть: Реализовать на питоне домашнее задание по дисциплине "Специальные главы математики. Часть 2" 4) Техническая часть: сделать под это всё удобные обёртки. Например, это функция, которая просто принимает список фичей, а на выходе показывает результат фита.
P8	Практическая работа №3. Машинное обучение.	ПК-4-У1	Пример задачи: Используя RandomForestClassifier и еще 2 метода машинного обучения предсказать будущий ABC класс на основе любых фичей (Датасет Продажи)
P9	Практическая работа №4. Ансамбли.	ПК-4-У1	Используя XGBOOST предсказать размер чаевых в датасете TIPS
P10	Индивидуальный учебный проект	ПК-4-У1; ПК-4-В1	<p>Провести полноценное исследование данных.</p> <p>Задачи:</p> <ol style="list-style-type: none"> 1. Изучить контекст датасета и придумать бизнес-цель. 2. Выполнить постановку задач на исследование (минимум 2 задачи, одна под классификацию/регрессию, и одна под кластеризацию (в крайнем случае делаем допущение что у нас нет Y – для особых датасетов)). 3. Провести оценку датасета (pandas/np/statsmodel...). 4. Провести подготовку данных. 5. Выполнить исследований (в том числе с промежуточными итогами, визуализациями). 6. Итоговые визуализации и выводы.
5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)			
Экзамен проводится в виде защиты Индивидуального учебного проекта.			

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

По дисциплине промежуточная аттестация предусмотрена в форме зачета в 1 семестре и экзамена во 2 семестре.
Шкала оценивания - 100-балльная, распределение баллов по видам деятельности:

1 семестр:

1. Домашняя работа №1 - оценка по шкале "выполнено/не выполнено".
2. Домашняя работа №2 - 20 .
3. Домашняя работа №3 - 20 .
4. Домашняя работа №4 - 30.
5. Домашняя работа №5 - 30 .

По итогам контроля знаний по сумме набранных баллов студенту выставляется оценка:

«зачет» – от 61 до 100 баллов;
«незачет» – от 0 до 60 баллов.

2 семестр:

1. Практическое задание 1 - 10 баллов,
2. Практическое задание 2 - 25 баллов,
3. Практическое задание 3 - 10 баллов,
4. Практическое задание 4 - 5 баллов,
5. Учебный проект - 50 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	Сузи Р. А.	Язык программирования Python: учебное пособие	Электронная библиотека	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ) Бином. Лаборатория знаний, 2007
Л1.2	Сорокин А. А.	Объектно-ориентированное программирование: учебное пособие: курс лекций	Электронная библиотека	Ставрополь: Северо-Кавказский Федеральный университет (СКФУ), 2014

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Буйначев С. К., Боклаг Н. Ю.	Основы программирования на языке Python: учебное пособие	Электронная библиотека	Екатеринбург: Издательство Уральского университета, 2014
Л2.2	Мейер Б.	Объектно-ориентированное программирование и программная инженерия: учебное пособие	Электронная библиотека	Москва: Национальный Открытый Университет «ИНТУИТ», 2016
Л2.3	Северенс Ч.	Введение в программирование на Python	Электронная библиотека	Москва: Национальный Открытый Университет «ИНТУИТ», 2016
Л2.4	Балджи А. С., Хрипунова М. Б., Александрова И. А.	Математика на Python: учебно-методическое пособие	Электронная библиотека	Москва: Прометей, 2018

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Онлайн-курс «Программирование на Python»	https://stepik.org/course/67/promo
Э2	Школа программиста	http://acmp.ru/
Э3	Jupyter Notebook	http://jupyter.org
Э4	Школа бэкенд-разработки (Яндекс)	https://habr.com/ru/company/yandex/blog/498856/

Э5	Mockingbird — онлайн-конструктор фроненда	https://gomockingbird.com/home
Э6	Программирование на Python (Computer Science Center) — все курсы по Python	https://compscicenter.ru/courses/python/

6.3 Перечень программного обеспечения

П.1	Microsoft Office
П.2	R Studio
П.3	Python
П.4	Anaconda
П.5	LMS Canvas

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Веб-сервис для хостинга IT-проектов и их совместной разработки: https://github.com/
-----	---

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-103	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

Методические указания к практическим заданиям:

В каждом варианте практического задания нужна разработка документации к реализуемым функциям, которые описывают, что и как делает эта функция.

Должно быть правильное форматирование кода в соответствии с codestyle-ом.

Методические указания к учебному проекту:

Итоговая практическая работа по дисциплине - проект "Реализация ООП-проекта на языке Python", который обучающийся должен защитить в конце 2 семестра обучения по дисциплине. У проекта есть две глобальные цели: сделать что-то работающее, что более-менее походит на продукт, и попрактиковаться в изученных темах, применить знания для реальной задачи.

Общие требования к проекту:

- "Хороший" код (codestyle, документация, разделение на модули)
- Использование классов, следование принципам ООП
- Связь с вебом в любом виде: собственное веб-приложение, использование чьего-то API (VK Api, Telegram API), краулинг и т.п.

- Понимание того, как работает проект: продуманная архитектура, понимание, как работает код.

Приветствуется:

- Использование какой-либо системы контроля версии (git)
- Нужные вам нестандартные модули (от requests до keras)
- Различные фреймворки (django, tornado и т.д.)