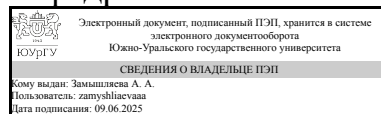


УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.М0.01 Информационный поиск, анализ и предобработка данных
для направления 02.04.02 Фундаментальная информатика и информационные технологии

уровень Магистратура

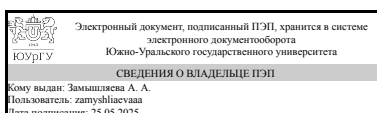
магистерская программа Разработка интеллектуальных систем

форма обучения очная

кафедра-разработчик Прикладная математика и программирование

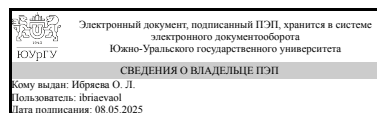
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 811

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



А. А. Замышляева

Разработчик программы,
к.физ.-мат.н., доцент



О. Л. Ибряева

1. Цели и задачи дисциплины

Цель изучения курса - узнать основные приемы сбора, предобработки, анализа набора данных, а также устройство систем информационного поиска. Задачи курса - научиться работать с библиотеками Python для анализа и визуализации данных, узнать об основных методах, применяемых при создании поисковых систем, принципах решения соответствующих задач обработки текстов на естественном языке, а также основы спектрального анализа сигналов.

Краткое содержание дисциплины

В курсе изучаются приемы сбора, преобработки и анализа данных. Большое внимание в курсе уделено временным рядам, рассматриваются стационарные и нестационарные ряды и задача их прогнозирования. Отличительной особенностью курса является изучение state-of-the-art методов преобразования временных рядов в изображения, набирающее все большую популярность, т.к. такие изображения могут потом служить входными данными для сверточных нейронных сетей.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-1 Способен разрабатывать и применять алгоритмы анализа данных, машинного обучения и нейросетевых технологий при создании интеллектуальных информационных систем	Знает: современные методологии описания, сбора, анализа и разметки данных Умеет: разрабатывать и применять алгоритмы анализа данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Оптимизационные задачи в машинном обучении, Нечеткие модели и их приложения в системах искусственного интеллекта, Современные нейросетевые технологии, Многомерный статистический анализ, Математические модели принятия решений в условиях неопределенности, Методы и технологии машинного обучения, Многокритериальная оптимизация и исследование операций

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Нет

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 38,25 ч.
контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		1
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	32	32
Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0
Лабораторные работы (ЛР)	16	16
<i>Самостоятельная работа (СРС)</i>	69,75	69,75
Изучить встроенные структуры данных, функции и файлы в Python	29,75	29,75
Изучить проектирование архитектур поисковых систем, как устроена поисковая машина Google, Yandex	20	20
Изучить пример выполнения разведочного анализа данных на языке Python для набора данных Ирисы Фишера	20	20
Консультации и промежуточная аттестация	6,25	6,25
Вид контроля (зачет, диф.зачет, экзамен)	-	зачет

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Язык Python, анализ и предобработка данных	12	6	0	6
2	Временные ряды	12	6	0	6
3	Информационный поиск	8	4	0	4

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Основы numpy и pandas. Загрузка, сохранение таблиц. Работа с колонками и фильтрами. Исследование данных, ф-ии: loc, describe, isna и т.д. Визуализация данных. Гистограмма, scatterplot.	2
2	1	Набор данных. Характеристики набора данных. Типы данных. Аугментация. Качество данных. Стандарт CRISP-DM.	2
3	1	Необходимость предобработки. Обработка пропущенных значений. Масштабирование. Коррелированные и категориальные переменные. Выбросы. Дисбаланс классов.	2
4	2	Стационарные временные ряды. Автоковариационная, автокорреляционная, частная автокорреляционная функции. Белый шум. Модели MA, AR, ARMA. Прогнозирование для модели ARMA.	2
5	2	TS ряды. Тренд, сезонность. DS ряды. Дифференцирование ряда. Интегрированные ряды. Модель ARIMA.	2
6	2	Преобразование Фурье. Спектрограмма. Преобразование Гильберта. ННТ,	2

		RP, GAF, MTF.	
7	3	Определение, модель информационного поиска. Булев поиск. Матрица «термин-документ». Обратный индекс. Выполнение запросов.	2
8	3	Предобработка текста. Токенизация. Стеemming, лемматизация. Текстовое ранжирование. Частота термина. Обратная документная частота. Вес tf-idf.	2

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	1	Основы Python	2
2	1	Разведочный анализ данных	4
3	2	Временные ряды. Arima	2
4	2	Преобразование временных рядов	2
5	2	Динамическое искажение времени	2
6	3	Система информационного поиска	4

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Изучить встроенные структуры данных, функции и файлы в Python	Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131721 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей. С. 71-104	1	29,75
Изучить проектирование архитектур поисковых систем, как устроена поисковая машина Google, Yandex	Пшехотская, Е. А. Информационный поиск : учебное пособие / Е. А. Пшехотская, Ю. Н. Филиппович. — Москва : Московский Политех, 2023. — 108 с. — ISBN 978-5-2760-2830-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/482786 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей. С. 58-68.	1	20
Изучить пример выполнения разведочного анализа данных на языке Python для набора данных Ирисы Фишера	Демидова, Л. А. Разведочный анализ данных. Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2022 — Часть 1 — 2022. —	1	20

	107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/310970 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей. С. 69 - 88		
--	--	--	--

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления бал
1	1	Текущий контроль	Тест 1. Набор данных	2	2	Тест содержит 2 вопроса, количество баллов равно кол
2	1	Текущий контроль	Тест 2. Введение в Python	2	2	Тест содержит 2 вопроса, количество баллов равно кол
3	1	Текущий контроль	Тест 3. Предобработка данных	3	3	Тест содержит 3 вопроса, количество баллов равно кол
4	1	Текущий контроль	Тест 4. Временные ряды	3	3	Тест содержит 3 вопроса, количество баллов равно кол
5	1	Текущий контроль	Тест 5. ARIMA	2	2	Тест содержит 2 вопроса, количество баллов равно кол
6	1	Текущий контроль	Тест 6. Информационный поиск	2	2	Тест содержит 2 вопроса, количество баллов равно кол
7	1	Текущий контроль	Тест 7. Текстовая обработка	3	3	Тест содержит 3 вопроса, количество баллов равно кол
8	1	Текущий контроль	Итоговый тест	20	20	Тест содержит 20 вопросов, количество баллов равно к
9	1	Текущий контроль	Лабораторная работа 1. Основы Python	5	5	В ноутбуке https://colab.research.google.com/drive/1ZafOAJeVkYCYZFW49_gmzsVS3qCjn8?usp=sharing содержится в ноутбуке https://colab.research.google.com/drive/1eLP6K5zZ1LfgL . Количество баллов равно количеству выполненных зад
10	1	Текущий контроль	Лабораторная работа 2. Разведочный анализ данных	15	15	Задания и баллы за них указаны в приложенном файле.
11	1	Текущий контроль	Лабораторная работа 3. Временные ряды	15	15	Задания и баллы за них указаны в приложенном файле.
12	1	Текущий контроль	Лабораторная работа 4.	10	10	Задание в приложенном файле. Загружен набор данных балл. Переведены сигналы в изображения RP, SDP, MT

			Преобразование временных рядов			регулировка параметров изображений с целью получить изображения, соответствующие разным видам движения
13	1	Текущий контроль	Лабораторная работа 5. DTW	8	8	Для использованных в лабораторной работе 4 сигналов метрику с приложенным сигналом и определите его тип построен - 1 балл, найдена DTW с 6 сигналами - 6 баллов
14	1	Текущий контроль	Лабораторная работа 6. Система информационного поиска	10	10	Задания и баллы за них указаны в приложенном файле.
15	1	Промежуточная аттестация	Тест	-	20	Тест содержит 20 вопросов, количество баллов равно количеству правильных ответов

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	Тест, 20 вопросов, за каждый правильный ответ 1 балл и далее используется вес 5, так что максимальный балл - 100.	В соответствии с пп. 2.5, 2.6 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ПК-1	Знает: современные методологии описания, сбора, анализа и разметки данных	+			+			+	+	+	+				+	+
ПК-1	Умеет: разрабатывать и применять алгоритмы анализа данных		+	+	+	+		+	+	+	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

Не предусмотрена

б) дополнительная литература:

1. Седжвик Р. Программирование на языке Python : учеб. курс ; пер. с англ. / Р. Седжвик, К. Уэйн, Р. Дондеро. - М.; СПб. : Диалектика, 2019. - 736 с. : ил.

2. Бендат Д. С. Прикладной анализ случайных данных / Пер. с англ. В. Е. Привольского, А. И. Кочубинского; Под ред. И. Н. Коваленко. - М. : Мир, 1989. - 540 с. : ил.

в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Краткие основы Python

из них: учебно-методическое обеспечение самостоятельной работы студента:

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	ЭБС издательства Лань	Демидова, Л. А. Разведочный анализ данных. Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2022 — Часть 1 — 2022. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/310970 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей.
2	Основная литература	ЭБС издательства Лань	Демидова, Л. А. Разведочный анализ данных. Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2023 — Часть 2 — 2023. — 92 с. — ISBN 978-5-7339-1933-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/382691 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей.
3	Основная литература	ЭБС издательства Лань	Кизбикенов, К. О. Прогнозирование и временные ряды : учебное пособие / К. О. Кизбикенов. — Барнаул : АлтГПУ, 2017. — 115 с. — ISBN 978-5-88210-869-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/112174 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей.
4	Основная литература	ЭБС издательства Лань	Пшехотская, Е. А. Информационный поиск : учебное пособие / Е. А. Пшехотская, Ю. Н. Филиппович. — Москва : Московский Политех, 2023. — 108 с. — ISBN 978-5-2760-2830-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/482786 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей.
5	Дополнительная литература	ЭБС издательства Лань	Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131721 (дата обращения: 08.05.2025). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. -Python(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

1. -Информационные ресурсы ФГУ ФИПС(бессрочно)

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Зачет	333 (36)	компьютерная техника
Лабораторные занятия	333 (36)	компьютерная техника, проектор
Лекции	336 (36)	Проектор