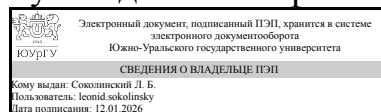


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления



Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.02 Подготовка данных для машинного обучения
для направления 02.03.02 Фундаментальная информатика и информационные технологии

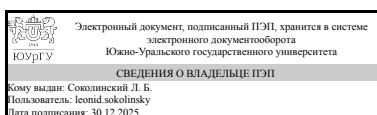
уровень Бакалавриат

форма обучения очная

кафедра-разработчик Системное программирование

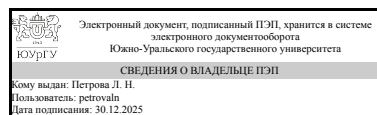
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 808

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



Л. Б. Соколинский

Разработчик программы,
старший преподаватель



Л. Н. Петрова

1. Цели и задачи дисциплины

Целью дисциплины является формирование базовых представлений, знаний и умений в области подготовки данных для машинного обучения. Основные задачи дисциплины: ознакомить студента с основными методами и подходами сбора и подготовки данных для машинного обучения, дать описание основных характеристик наборов данных, показать способы предварительной обработки данных.

Краткое содержание дисциплины

Изложение наиболее важных понятий, определений и методов работы с данными при подготовке датасетов для машинного обучения. В курс входят следующие разделы: математические основы, основы работы с изображениями и видео, основы работы с текстовыми данными, основы работы с аудио данными.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-4 Способен участвовать в организации подготовительных мероприятий по реализации проектов, а также участвовать в реализации и сопровождении проекта	Знает: базовые подходы к сбору, разметке и предварительной подготовке данных для моделей машинного обучения Умеет: Умеет осуществлять сбор исходной информации с использованием платформ данных (облачных и внутрикорпоративных) Имеет практический опыт: создания собственных наборов данных для моделей машинного обучения при решении задач с учетом особенностей решаемой задачи

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Не предусмотрены

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., 54,5 ч. контактной работы

Вид учебной работы	Всего	Распределение по семестрам
--------------------	-------	----------------------------

	часов	в часах
		Номер семестра
		6
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	48	48
Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	53,5	53,5
Подготовка собственного набора данных изображений на платформа Robooflow.	20	20
Подготовка собственного набора текстовых данных.	33,5	33.5
Консультации и промежуточная аттестация	6,5	6,5
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.зачет

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Математические основы	6	2	4	0
2	Основы работы с изображениями и видео	14	4	10	0
3	Основы работы с текстовыми данными	20	8	12	0
4	Основы работы с аудио данными	8	2	6	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Математические и статистические характеристики наборов данных	2
2-3	2	Представление изображений и видео в компьютере и связь таких представлений с машинным обучением. Операции над данными в рамках предварительной подготовки для машинного обучения. Разметка изображений.	4
4-6	3	Представление текста в компьютере и связь таких представлений с машинным обучением. Операции над текстовыми данными в рамках предварительной подготовки для машинного обучения. Особенности сбора и обработки текстовых данных для машинного обучения.	6
7	3	Особенности аугментации текстовых данных. Разметка текстов.	2
8	4	Представление аудио в компьютере и связь таких представлений с машинным обучением. Операции над аудио данными в рамках предварительной подготовки для машинного обучения. Разметка аудио.	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1-2	1	Статистические характеристики наборов данных. Работа с табличными данными, текстами и изображениями.	4

3-5	2	Представление изображений в компьютере. Рассмотрение современных библиотек работы с изображениями Pillow, OpenCV, Albumentation. Аугментация.	6
6-7	2	Представление видео данных. Кадрирование и аугментация данных.	4
8-10	3	Представление текстов в компьютере. Мешок слов, one-hot-encoding, векторное представление, embeddings. Методы работы с текстами sklearn и tensorflow.	6
11-13	3	Парсинг текстов. Подготовка текстовых данных для решения различных задач: классификации и распознавания именованных сущностей.	6
14-16	4	Представление аудио в компьютере. Рассмотрение современной библиотеки работы с аудио Librosa. Подготовка аудио данных с помощью методов tensorflow. Аугментация.	6

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Подготовка собственного набора данных изображений на платформа Robooflow.	Методические указания раздел "Подготовка собственного набора данных изображений на платформа Robooflow." https://app.roboflow.com	6	20
Подготовка собственного набора текстовых данных.	Методические указания раздел "Подготовка собственного набора текстовых данных." https://appen.com	6	33,5

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 баллов	Учитывается в ПА
1	6	Текущий контроль	Тест 1	5	5	Компьютерный тест состоит из 5 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 10 мин. Стоимость одного вопроса - 1 балл. 5 баллов: задание полностью выполнено без ошибок 1-4 баллов: задание выполнено частично или	дифференцированный зачет

						выполнено с ошибками 0 баллов: задание не выполнено	
2	6	Текущий контроль	Практическая 1	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
3	6	Текущий контроль	Практическая 2	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
4	6	Текущий контроль	Практическая 3	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
5	6	Текущий контроль	Практическая 4	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
6	6	Текущий контроль	Практическая 5	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
7	6	Текущий контроль	Практическая 6	5	5	5 заданий, каждое задание 1 балл. 0 баллов: задание не выполнено	дифференцированный зачет
8	6	Промежуточная аттестация	Тест 2	-	5	Компьютерный тест состоит из 5 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 10 мин. Стоимость одного вопроса - 1 балл. 5 баллов: задание полностью выполнено без ошибок 1-4 баллов: задание выполнено частично или выполнено с ошибками 0 баллов: задание не выполнено	дифференцированный зачет
9	6	Промежуточная аттестация	зачет	-	15	Компьютерный тест состоит из 15 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 30 минут. Стоимость одного вопроса - 1 балл. 15 баллов: задание полностью выполнено без ошибок 1-14 баллов: задание выполнено частично или выполнено с ошибками 0 баллов: задание не выполнено	дифференцированный зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
дифференцированный зачет	При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. № 25-13/09). Процедура прохождения промежуточной аттестации осуществляется согласно Положению о текущем контроле успеваемости и промежуточной аттестации (приказ ректора от 27.02.2024 № 33-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля следующим образом: • Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %. • Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. • Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 %. • Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Если студент согласен с оценкой, полученной по результатам текущего контроля, то он может в день, предшествующий промежуточной аттестации дать свое согласие на автомат в личном кабинете. В случае явки студента на промежуточную аттестацию, давшего свое согласие на автомат в личном кабинете, студент имеет право пройти мероприятия текущего контроля по дисциплине на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Снижение оценки в этом случае запрещено. Если студент не дал согласия в личном кабинете, то он может согласиться с оценкой лично на промежуточной аттестации в день ее проведения. Если студент не согласен с оценкой, то он имеет право пройти контрольно-рейтинговые мероприятия на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Фиксация результатов учебной деятельности по дисциплине проводится в день промежуточной аттестации на основе согласия студента, данного им в личном кабинете. При отсутствии согласия в журнале дисциплины фиксация результатов происходит при личном присутствии студента. Если студент не дал согласие в личном кабинете и не явился на промежуточную аттестацию – ему выставляется «неявка». Промежуточная аттестация проводится в форме тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 15 вопросов, на выполнение теста дается 30 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации.	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
ПК-4	Знает: базовые подходы к сбору, разметке и предварительной подготовке данных для моделей машинного обучения	+	+	+	+	+	+	+	+	+
ПК-4	Умеет: Умеет осуществлять сбор исходной информации с использованием платформ данных (облачных и внутрикорпоративных)	+	+	+	+	+	+	+	+	+
ПК-4	Имеет практический опыт: создания собственных наборов данных для моделей машинного обучения при решении задач с учетом особенностей решаемой задачи	+	+	+	+	+	+	+	+	+

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:*

Не предусмотрена

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

Не предусмотрены

г) *методические указания для студентов по освоению дисциплины:*

1. методические указания

из них: *учебно-методическое обеспечение самостоятельной работы студента:*

1. методические указания

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	ЭБС издательства Лань	Бизли, Д. Python. Книга рецептов / Д. Бизли, Б. К. Джонс ; перевод с английского Б. В. Уварова. — Москва : ДМК Пресс, 2019. — 646 с. — ISBN 978-5-97060-751-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131723 (дата обращения: 29.12.2025). — Режим доступа: для авториз. пользователей.
2	Основная литература	ЭБС издательства Лань	Ганегедара, Т. Обработка естественного языка с TensorFlow : руководство / Т. Ганегедара ; перевод с английского В. С. Яценкова. — Москва : ДМК Пресс, 2020. — 382 с. — ISBN 978-5-97060-756-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/140584 (дата обращения: 29.12.2025). — Режим доступа: для авториз. пользователей.
3	Основная литература	ЭБС издательства Лань	Антонио, Д. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / Д. Антонио, П. Суджит ;

			перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2018. — 294 с. — ISBN 978-5-97060-573-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/111438 (дата обращения: 29.12.2025). — Режим доступа: для авториз. пользователей.
4	Дополнительная литература	ЭБС издательства Лань	Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва : ДМК Пресс, 2018. — 418 с. — ISBN 978-5-97060-481-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/116122 (дата обращения: 29.12.2025). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

Нет

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Зачет	110 (3Г)	Компьютерный класс
Практические занятия и семинары	110 (3Г)	Проектор, компьютерный класс
Лекции	110 (3Г)	Проектор, компьютерный класс