ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Замышляева А. А.
Пользователь: zamyshliaevaaa
Дата подписания: 16.11.2025

А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.О.33 Технологии и системы обработки больших данных
для направления 01.03.02 Прикладная математика и информатика
уровень Бакалавриат
форма обучения очная
кафедра-разработчик Центр ОП топ-уровня в сфере ИИ "ВиртУм"

Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 01.03.02 Прикладная математика и информатика, утверждённым приказом Минобрнауки от 10.01.2018 № 9

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

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Замышляева А. А.
Пользователь: zamyshliaevaaa
Дата подписания: 16.11.2025

А. А. Замышляева

Разработчик программы,
д.физ.-мат.н., проф., профессор

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Соколинский Л. Б.
Пользователь: leonid.sokolinsky
Дата подписания: 13.11.2025

Л. Б. Соколинский

Челябинск

1. Цели и задачи дисциплины

Цель дисциплины - освоение студентами ключевых технологий, методов и принципов организации систем обработки больших данных. Задачи освоения дисциплины: общее представление и понимание выбора той или иной технологии обработки больших данных для решения поставленной задачи; умение организовывать распределенное хранилище и параллельную обработку на базе современных технологий больших данных; получение практических навыков разработки и отладки прикладных решений с элементами ИИ с применением различных технологий обработки данных. Профессиональная роль: инженер по данным.

Краткое содержание дисциплины

Изложены наиболее важные понятия, методы и технологии обработки больших данных. В курс входят следующие разделы: введение в большие данные; архитектура систем обработки больших данных; хранение и управление данными (DataLake); обработка больших данных (MapReduce); анализ и визуализация данных; машинное обучение и большие данные; безопасность и этика работы с большими данными.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-1 [LC-5] Способен применять и (или) проектировать различные инструменты и инженерные практики промышленной разработки систем ИИ, развертывания и сопровождения моделей машинного обучения в продуктивной среде	Умеет: - [И-1, ПУ] использовать языки программирования и интерфейсы для взаимодействия с big-data-платформами Имеет практический опыт: - [И-1, ПУ] анализа и обработки больших массивов данных с помощью стандартных инструментальных средств
ПК-10 [BD-4] Способен применять различные модели и (или) технологии обработки данных	Знает: - [И-1, ПУ] принципы модели MapReduce для параллельной обработки больших данных Умеет: - [И-1, ПУ] организовывать распределенное хранилище и параллельную обработку на базе современных технологий (Hadoop, Spark) больших данных Имеет практический опыт: - [И-2, ПУ] Разработки и отладки прикладных решений с элементами ИИ с применением различных технологий обработки данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.О.22 Машинное обучение	1.О.38 Базы данных NoSQL, 1.О.35 Основы распределенных и облачных вычислений

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

Дисциплина	Требования
1.О.22 Машинное обучение	Знает: -[И-2, ПУ] различные архитектуры ранжированного поиска (одно-двух-трехстадийное ранжирование), -[И-3, ПУ] базовые метрики качества обучения без учителя (silhouette score adjusted rand index), -[И-1, ПУ] категории задач автоматического машинного обучения, -[И-1, ПУ] основные методы статистического машинного обучения Умеет: -[И-2, ПУ] выбирать и адаптировать оптимальный алгоритм машинного обучения для конкретной задачи, -[И-2, ПУ] применять методы байесовской классификации и ансамблевые методы МО (бэггинг, бустинг, стэкинг моделей), а также производных от них (случайные леса, градиентный бустинг на деревьях), -[И-1, ПУ] проводить одномерный и многомерный анализ признаков, в том числе с использованием средств визуализации[И-4, ПУ] применять стандартные методы отбора признаков и выбирать оптимальное подмножество признаков, -[И-1, ПУ] использовать инструменты очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных, -[И-1, ПУ] анализировать специфику задачи с учётом современных трендов (например, использование AutoML для обработки больших данных, интеграция с MLOps), выбирать подходящие AutoML-инструменты Имеет практический опыт: -[И-3, ПУ] тестирования моделей перед развертыванием, оценки качества моделей машинного обучения, -[И-2, ПУ] использования инструментов оценки качества моделей ранжирования и сравнения ранжирующих моделей между собой; применения методов обучения типа pairwise и listwise; использования различных архитектур ранжированного поиска (одно-двух-трехстадийное ранжирование), -[И-3, ПУ] использования методов понижения размерности и подбора оптимальной размерности в зависимости от необходимой доли объяснённой дисперсии, -[И-3, ПУ] использования готовых инструментов для оценки качества кластеризации и других моделей без учителя, -[И-1, ПУ] адекватного выбора методов статистического машинного обучения с учётом особенностей данных и задачи

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 54,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		6
Общая трудоёмкость дисциплины	108	108
Аудиторные занятия:	48	48
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	16	16
Лабораторные работы (ЛР)	0	0
Самостоятельная работа (СРС)	53,75	53,75
Подготовка к зачету	10,75	10.75
Студенческий научный семинар "Большие данные и искусственный интеллект"	43	43
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	Введение в большие данные	6	2	4	0
2	Архитектуры систем обработки больших данных	2	2	0	0
3	Хранение, управление и обработка больших данных	14	10	4	0
4	Анализ и визуализация больших данных	10	6	4	0
5	Машинное обучение и большие данные	8	4	4	0
6	Обработка потоковых данных	6	6	0	0
7	Безопасность и этика работы с большими данными	2	2	0	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение в большие данные. Понятие больших данных. Источники больших данных. Проблемы обработки больших данных.	2
2	2	Архитектуры систем обработки больших данных. Традиционные архитектуры vs. архитектуры больших данных. Принципы распределённых систем. Hadoop и его экосистема.	2
3	3	Хранение и управление большими данными. Структурированные и неструктурированные данные. Хранение структурированных данных. Хранилища данных (Data Warehousing).	2
4	3	Хранение неструктурированных данных, озера данных (DataLake). Архитектура DataLake. Основные преимущества использования DataLake.	2

		Сравнение DataLake и Data Warehouse.	
5	3	Обработка больших данных. Парадигмы обработки данных: пакетная и потоковая.	2
6	3	Методы и технологии пакетной обработки данных (MapReduce, Apache Spark).	2
7	3	ETL-процессы и инструменты.	2
8	4	Анализ и визуализация данных. Методы анализа больших данных. Выполнения SQL-запросов на больших объемах данных (Presto).	2
9	4	Анализ данных в DataLake (Apache Hive).	2
10	4	Инструменты для визуализации данных (Tableau, Power BI).	2
11	5	Машинное обучение и большие данные. Применение машинного обучения для анализа больших данных. Библиотеки и инструменты для анализа данных (scikit-learn, TensorFlow).	2
12	5	Алгоритмы масштабируемого машинного обучения.	2
13	6	Обработка потоковых данных. Определение и примеры потоковых данных, проблемы обработки потоковых данных. Архитектура системы управления потоками данных (Stream Processing Systems), примеры популярных систем (Apache Kafka, Apache Flink, Apache Storm).	2
14	6	Методы и алгоритмы обработки потоковых данных. Получение репрезентативной выборки. Фильтрация потоков.	2
15	6	Подсчет различных элементов в потоке, алгоритм Флажоле-Мартена.	2
16	7	Безопасность и этика работы с большими данными. Защита данных и конфиденциальности. Этические аспекты обработки данных. Законодательство в области больших данных (GDPR и др.).	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Предобработка текстовых данных	4
2	3	Работа с Apache Spark	4
3	4	Создание визуализаций данных	4
4	5	Построение модели машинного обучения	4

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 к зачету	Основная литература. Дополнительная литература.	6	10,75
Студенческий научный семинар "Большие данные и искусственный интеллект"	1. Yazdani, S., Singh, A., Saxena, N. et al. Generative AI in depth: A survey of recent advances, model variants, and real-world applications. J Big Data 12, 230 (2025). https://doi.org/10.1186/s40537-025-01247-x 2. Abdalla, H.B. A brief survey on big data:	6	43

	technologies, terminologies and data-intensive applications. J Big Data 9, 107 (2022). https://doi.org/10.1186/s40537-022-00659-3 3. Saeed, N., Ashour, M. & Mashaly, M. Comprehensive review of federated learning challenges: a data preparation viewpoint. J Big Data 12, 153 (2025). https://doi.org/10.1186/s40537-025-01195-6		

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	Учи-тыва-ется в ПА
1	6	Текущий контроль	Контрольный опрос 1	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
2	6	Текущий контроль	Контрольный опрос 2	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
3	6	Текущий контроль	Контрольный опрос 3	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста.	зачет

						1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	
4	6	Текущий контроль	Контрольный опрос 4	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
5	6	Текущий контроль	Контрольный опрос 5	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
6	6	Текущий контроль	Контрольный опрос 6	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
7	6	Текущий контроль	Контрольный опрос 7	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 5 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	зачет
8	6	Лабораторная работа	Практическое задание 1	1	5	5 баллов: Полная реализация, анализ результатов, ответы на контрольные вопросы. 3-4 балла: Реализация без анализа или пропущены ответы на контрольные вопросы. 0-2 балла: Не выполнены ключевые задачи.	зачет

9	6	Лабораторная работа	Практическое задание 2	1	5	5 баллов: Полная реализация, анализ результатов, ответы на контрольные вопросы. 3-4 балла: Реализация без анализа или пропущены ответы на контрольные вопросы. 0-2 балла: Не выполнены ключевые задачи.	зачет
10	6	Лабораторная работа	Практическое задание 3	1	5	5 баллов: Полная реализация, анализ результатов, ответы на контрольные вопросы. 3-4 балла: Реализация без анализа или пропущены ответы на контрольные вопросы. 0-2 балла: Не выполнены ключевые задачи.	зачет
11	6	Лабораторная работа	Практическое задание 4	1	5	5 баллов: Полная реализация, анализ результатов, ответы на контрольные вопросы. 3-4 балла: Реализация без анализа или пропущены ответы на контрольные вопросы. 0-2 балла: Не выполнены ключевые задачи.	зачет
12	6	Проме-жуточная аттестация	Итоговый тест	-	21	Тест содержит 21 равнозначный вопрос и рассчитан на 45 мин. Количество баллов равно количеству правильных ответов на вопросы теста.	зачет
13	6	Бонус	Участие в студенческом научном семинаре "Большие данные и искусственный интеллект"	-	15	За выступление с докладом начисляется 10 баллов. За участие в семинаре начисляется 1 балл.	зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля следующим образом: • Зачтено: Величина рейтинга обучающегося по дисциплине 60…100 %. • Не зачтено: Величина рейтинга обучающегося по дисциплине 0…59 %. Промежуточная аттестация проводится в форме тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 21 вопрос. На выполнение теста дается 45 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	10	11	12	13
ПК-1	Умеет: - [И-1, ПУ] использовать языки программирования и интерфейсы для взаимодействия с big-data-платформами	+			+	+	+	+	+	+	+	+	+	
ПК-1	Имеет практический опыт: - [И-1, ПУ] анализа и обработки больших массивов данных с помощью стандартных инструментальных средств		+		+	+	+	+	+	+	+	+	+	
ПК-10	Знает: - [И-1, ПУ] принципы модели MapReduce для параллельной обработки больших данных			+									+	
ПК-10	Умеет: - [И-1, ПУ] организовывать распределенное хранилище и параллельную обработку на базе современных технологий (Hadoop, Spark) больших данных									+			+	
ПК-10	Имеет практический опыт: - [И-2, ПУ] Разработки и отладки прикладных решений с элементами ИИ с применением различных технологий обработки 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:

Не предусмотрена

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

1. Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика / Юж.-Урал. гос. ун-т; ЮУрГУ. - Челябинск : Издательство ЮУрГУ, 2012-. -. URL: http://vestnik.susu.ac.ru/

г) методические указания для студентов по освоению дисциплины:

1. Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. — Санкт-Петербург : НИУ ИТМО, 2019. — 50 с. — Текст : электронный // Лань : электронно-библиотечная система.

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. — Санкт-Петербург : НИУ ИТМО, 2019. — 50 с. — Текст : электронный // Лань : электронно-библиотечная система.

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание

1	Основная литература	ЭБС издательства Лань	Системы искусственного интеллекта и большие данные. Раздел «Большие данные» : учебное пособие / А. А. Парамонов, И. А. Юрченков, Б. А. Крынецкий, И. В. Есипов. — Москва : РТУ МИРЭА, 2025. — 147 с. — ISBN 978-5-7339-2499-1. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/493550
2	Основная литература	ЭБС издательства Лань	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/93571
3	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. https://e.lanbook.com/book/131721

Перечень используемого программного обеспечения:

1. Microsoft-Office(бессрочно)
2. -Python(бессрочно)

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
Лекции	484 (3)	Проектор, компьютер
Экзамен	110 (3г)	Компьютерный класс
Лабораторные занятия	110 (3г)	Компьютерный класс