

**УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
УЧЕБНОЙ ДИСЦИПЛИНЫ**

**ТЕХНОЛОГИИ СБОРА И ПОДГОТОВКИ ДАННЫХ В
АГРОИНЖЕНЕРИИ**

Направление подготовки: 35.03.06 Агроинженерия

Профили подготовки: Искусственный интеллект

Квалификация (степень) выпускника: Бакалавр

1. Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина направлена на формирование следующих компетенций выпускника:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
УК-11. Способен планировать и организовывать свою деятельность в цифровом пространстве с учетом правовых и этических норм взаимодействия человека и искусственного интеллекта и требований информационной безопасности	ИУК-11.1. Выбирает современные технологии и системы искусственного интеллекта для решения задач в профессиональной деятельности ИУК-11.2. Использует технологии сбора, обработки, интерпретации, анализа и обмена информацией с учетом требований информационной безопасности ИУК-11.3. Применяет и адаптирует правовые и этические нормы и национальные и международные стандарты в области искусственного интеллекта и смежных областях для решения задач в профессиональной деятельности в условиях изменения социально-экономических условий
ПК-17. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ИПК-17.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях ИПК-17.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения
ПК-18. Способен выполнять анализ больших данных	ИПК-18.1. Использует знания о вариантах использования больших данных, определениях, словарях и эталонной архитектуре больших данных для эффективного извлечения, хранения, подготовки больших данных ИПК-18.2. Выполняет обработку, удаленную, распределенную и объединенную аналитику, описание и управление качеством и достоверностью, использует результаты анализа больших данных

2. Разделы дисциплины и вид занятий

№ п/п	Раздел дисциплины	Лекции	ЛЗ	СРС	Всего
1.	Технологии сбора и подготовки данных в агроинженерии	5	5	24	34
2.	Трансформация и визуализация данных	6	6	22	34
3.	Очистка и предобработка данных	6	6	24	36
Всего		17	17	70	104

ЛЗ – лабораторные занятия

СРС – самостоятельная работа студента

3. Образовательные технологии

Объем аудиторных занятий всего – 34 часа, в том числе лекции – 17 часов, лабораторные работы – 17 часов.

Удельный вес занятий, проводимых в интерактивной форме, – 100% от объема аудиторных занятий.

Семестр	Вид занятия	Наименование темы	Используемые интерактивные образовательные технологии	Количество часов
3	Лекция	Технологии сбора и подготовки данных в агроинженерии	Лекция-визуализация	5
3	Лекция	Трансформация и визуализация данных	Лекция-визуализация	6
3	Лекция	Очистка и предобработка данных	Лекция-визуализация	6
3	ЛЗ	Лабораторный практикум «Сбор и подготовка агроинженерных данных для анализа»	Лабораторная работа	17
Итого:				34

4 Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

4.1 Перечень оценочных средств

Компетенции	Индикаторы достижения компетенций	Оценочные средства
УК-11. Способен планировать и организовывать свою деятельность в цифровом пространстве с учетом правовых и этических норм взаимодействия человека и искусственного интеллекта и требований информационной безопасности	ИУК-11.1. Выбирает современные технологии и системы искусственного интеллекта для решения задач в профессиональной деятельности ИУК-11.2. Использует технологии сбора, обработки, интерпретации, анализа и обмена информацией с учетом требований информационной безопасности	1. Задания для выполнения лабораторных работ. 2. Задания для самостоятельной работы. 3. Вопросы к зачету.
ПК-17. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ИПК-17.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях ИПК-17.2. Выполняет подготовку и разметку структурированных и неструктурированных	1. Задания для выполнения лабораторных работ. 2. Задания для самостоятельной работы. 3. Вопросы к зачету.

Компетенции	Индикаторы достижения компетенций	Оценочные средства
	данных для машинного обучения	
ПК-18. Способен выполнять анализ больших данных	<p>ИПК-18.1. Использует знания о вариантах использования больших данных, определениях, словарях и эталонной архитектуре больших данных для эффективного извлечения, хранения, подготовки больших данных</p> <p>ИПК-18.2. Выполняет обработку, удаленную, распределенную и объединенную аналитику, описание и управление качеством и достоверностью, использует результаты анализа больших данных</p>	<p>1. Задания для выполнения лабораторных работ.</p> <p>2. Задания для самостоятельной работы.</p> <p>3. Вопросы к зачету.</p>

4.2 Задания для самостоятельной работы:

По итогам самостоятельной работы студент готовит отчет, включающий в себя ответы на вопросы и решение заданий, предполагавшихся к выполнению в ходе самостоятельной работы. Отчет сдается преподавателю в электронной форме.

Самостоятельная работа по теме «Технологии сбора и подготовки данных в агроинженерии»

Проведите информационный поиск и подготовьте аналитическую записку по вопросам:

1. Классификация технологий сбора данных.
2. Применение БПЛА.
3. Коммуникационные технологии Agro IoT.
4. Хранилища данных.
5. Принципы анализа данных.
6. Структурированные и неструктурированные данные.
7. Подготовка данных к анализу.
8. Применение технологии KDD для анализа агротехнических данных.
9. Применение технологии Data Mining для анализа агротехнических данных.

Напишите эссе, раскрыв в нем функциональные возможности инструментов сбора и анализа данных: MS EXCEL, POWER BI, ORANGE DATA MINING, Python и его библиотеки.

Самостоятельная работа по теме «Трансформация и визуализация данных»

Разработайте интеллект-карту, отражающую сущность процессов трансформации и визуализации данных:

1. Трансформация упорядоченных данных.
2. Группировка данных.
3. Слияние данных.
4. Квантование.
5. Нормализация и кодирование данных.
6. Введение в визуализацию.
7. Визуализаторы общего назначения. OLAP-анализ.
8. Визуализаторы для анализа и оценки качества моделей.
9. Визуализаторы, применяемые для интерпретации результатов анализа.

Самостоятельная работа по теме «Очистка и предобработка данных»

Разработайте интеллект-карту, отражающую сущность процессов очистки и предобработки данных:

1. Технологии и методы оценки качества данных.
2. Очистка и предобработка.
3. Фильтрация данных.
4. Обработка дубликатов и противоречий.
5. Выявление аномальных значений.
6. Восстановление пропущенных значений.
7. Сокращение числа признаков.
8. Сокращение числа значений признаков и записей. Сэмплинг.

Лабораторный практикум по дисциплине «Технологии сбора и подготовки данных в агроинженерии»

Лабораторный практикум включает три лабораторные работы:

1. Технологии сбора и подготовки данных в агроинженерии.
2. Трансформация и визуализация данных.
3. Очистка и предобработка данных.

Лабораторные работы обеспечивают формирование практических навыков применения современных технологий сбора и подготовки данных в агроинженерии, включая большие данные. В качестве инструментальных средств рекомендуется применение пакетов обработки данных Data Mining. Перед студентом ставится задача сбора и подготовки данных для последующего анализа:

1. Импорт данных из Excel и CSV.
2. Верификации данных на простейшие ошибки.
3. Устранение дублей.
4. Выявление фиктивных и ошибочных данных.
5. Анализ редких значений.
6. Обнаружение пропусков в данных.
7. Устранение выбросов и аномалий данных.

Вопросы для зачета

1. Применение БПЛА для сбора данных.
2. Коммуникационные технологии Agro IoT
3. Хранилища данных.
4. Принципы анализа данных.
5. Структурированные и неструктурированные данные.
6. Подготовка данных к анализу.
7. Применение технологии KDD для анализа агротехнических данных.

8. Применение технологии Data Mining для анализа агротехнических данных.
9. Инструменты сбора и анализа данных: POWER BI.
10. Инструменты сбора и анализа данных: MS EXCEL.
11. Инструменты сбора и анализа данных: ORANGE DATA MINING, Python и его библиотеки.
12. Инструменты сбора и анализа данных: Python и его библиотеки.
13. Понятие трансформации данных.
14. Трансформация упорядоченных данных.
15. Группировка данных.
16. Слияние данных.
17. Квантование.
18. Нормализация и кодирование данных.
19. Введение в визуализацию.
20. Визуализаторы общего назначения.
21. OLAP-анализ.
22. Визуализаторы для анализа и оценки качества моделей.
23. Визуализаторы, применяемые для интерпретации результатов анализа.
24. Оценка качества данных.
25. Технологии и методы оценки качества данных.
26. Очистка и предобработка.
27. Фильтрация данных.
28. Обработка дубликатов и противоречий.
29. Выявление аномальных значений.
30. Восстановление пропущенных значений.
31. Введение в сокращение размерности.
32. Сокращение числа признаков.
33. Сокращение числа значений признаков и записей
34. Сэмплинг.
35. Структурированные и неструктурированные данные.
36. Анализ больших данных.

Уровни оценки компетенций:

- базовый 55-69 баллов,
- повышенный 70-100 баллов.

Преподаватель проводит систематический контроль знаний студентов, ориентируясь на перечень вопросов для проведения зачета.

Критерии оценивания ответа студента на зачете

Ответ на зачете оценивается исходя из 40 баллов (максимум). Билет содержит теоретический вопрос и практическое задание, преподаватель может задавать дополнительные вопросы. Полный ответ на основной вопрос оценивается максимум в 20 баллов, предполагает свободное изложение (не чтение) всего необходимого материала, ответы студента на уточняющие вопросы, если они есть. Правильный ответ на дополнительный вопрос оценивается максимум в 5 баллов. Правильное выполнение практического задания оценивается в 20 баллов.

Критерии оценки лабораторных работ и самостоятельной работы студента (от 0 до 10 баллов):

- **9-10 баллов** выставляется студенту, если работа выполнена самостоятельно и полностью верно; представлен отчет, содержащий результаты выполнения заданий работы и ответы на контрольные вопросы; студент анализирует результаты, полученные в ходе выполнения работы, делает выводы.
- **7-8 баллов** выставляется студенту, если работа выполнена самостоятельно, в целом правильно, но имеются некоторые неточности в выполнении заданий или ответах на контрольные вопросы; представлен отчет, содержащий результаты выполнения заданий и ответы на контрольные вопросы; студент анализирует результаты, полученные в ходе выполнения работы, делает выводы.
- **5-6 баллов** выставляется студенту, если работа выполнена самостоятельно, в целом правильно, но имеются некоторые неточности в выполнении заданий или ответах на контрольные вопросы; представлен отчет, содержащий результаты выполнения заданий лабораторной работы и ответы на контрольные вопросы; студент испытывает затруднения при проведении анализа результатов, полученных в ходе выполнения лабораторной работы, и формулировке выводов.
- **3-4 балла** выставляется студенту, если студент не до конца справился с заданием, не совсем верно ответил на контрольные вопросы, однако оформил отчет по результатам работы.
- **1-2 балла** выставляется студенту, если студент не до конца справился с заданием, не совсем верно ответил на контрольные вопросы, не оформил отчет по результатам работы.
- **0 баллов** выставляется студенту, если студент не справился с заданием, неверно ответил на представленные вопросы.

4.3 Шкала и критерии оценивания компетенций на различных этапах их формирования

Шкала оценивания компетенций:

Оценка в 100-балльной шкале	Оценка в 5-ти балльной шкале	Уровень сформированности компетенций
0-54 баллов	неудовлетворительно (не зачтено)	Недостаточный
55-69 баллов	удовлетворительно (зачтено)	Пороговый уровень
70-85 баллов	хорошо (зачтено)	Продвинутый уровень
86-100 баллов	отлично (зачтено)	Высокий уровень

Критерии оценивания компетенций:

Индикаторы достижения компетенций	Критерии оценивания компетенций		
	Пороговый уровень	Продвинутый уровень	Высокий уровень
УК-11. Способен планировать и организовывать свою деятельность в цифровом пространстве с учетом правовых и этических норм взаимодействия человека и искусственного интеллекта и требований	Не знает текущее состояние информационного общества и роль искусственного интеллекта в его развитии; классификацию информационных систем и систем искусственного интеллекта, функциональность	Знает текущее состояние информационного общества и роль искусственного интеллекта в его развитии; классификацию информационных систем и систем искусственного интеллекта, функциональность	Демонстрирует глубокое знание и понимание: текущего состояния информационного общества и роль искусственного интеллекта в его развитии; классификации информационных систем и систем искусственного

Индикаторы достижения компетенций	Критерии оценивания компетенций		
	Пороговый уровень	Продвинутый уровень	Высокий уровень
информационной безопасности	<p>программного обеспечения для решения задач профессиональной деятельности; современное состояние ИКТ в мире и перспективы их развития; знает основные методы оценки экономической эффективности применяемого программного и аппаратного обеспечения.</p> <p>Не умеет анализировать сущность и значение искусственного интеллекта в развитии современного информационного общества; выбирать необходимые инструментальные средства анализа для решения поставленных задач; формировать и использовать критерии оценки эффективности применения программного и аппаратного обеспечения в профессиональной деятельности</p>	<p>программного обеспечения для решения задач профессиональной деятельности; современное состояние ИКТ в мире и перспективы их развития; знает основные методы оценки экономической эффективности применяемого программного и аппаратного обеспечения.</p> <p>Умеет анализировать сущность и значение искусственного интеллекта в развитии современного информационного общества; выбирать необходимые инструментальные средства анализа для решения поставленных задач; формировать и использовать критерии оценки эффективности применения программного и аппаратного обеспечения в профессиональной деятельности</p>	<p>интеллекта, функциональности программного обеспечения для решения задач профессиональной деятельности; современного состояния ИКТ в мире и перспективы их развития; основных методов оценки экономической эффективности применяемого программного и аппаратного обеспечения.</p> <p>Полностью верно и самостоятельно: анализирует сущность и значение искусственного интеллекта в развитии современного информационного общества; выбирает необходимые инструментальные средства анализа для решения поставленных задач; формирует и использует критерии оценки эффективности применения программного и аппаратного обеспечения в профессиональной деятельности</p>
ПК-17. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	<p>Не знает методы поиска данных; методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных; методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок</p> <p>Не умеет отделять достоверные источники данных от</p>	<p>Знает методы поиска данных; методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных; методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок</p> <p>Умеет отделять достоверные источники данных от</p>	<p>Демонстрирует глубокое знание и понимание: методов поиска данных; методов редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных; методов планирования вычислительного эксперимента, формирования обучающей и контрольной выборок</p> <p>Полностью верно и</p>

Индикаторы достижения компетенций	Критерии оценивания компетенций		
	Пороговый уровень	Продвинутый уровень	Высокий уровень
	сомнительных, осуществлять критический отбор данных, проверять их на целостность и непротиворечивость; выявлять и исключать из массива данных ошибочные данные и выброс; выделять входные и выходные переменные с целью использования предиктивных моделей	сомнительных, осуществлять критический отбор данных, проверять их на целостность и непротиворечивость; выявлять и исключать из массива данных ошибочные данные и выброс; выделять входные и выходные переменные с целью использования предиктивных моделей	самостоятельно: отделяет достоверные источники данных от сомнительных, осуществляет критический отбор данных, проверяет их на целостность и непротиворечивость; выявляет и исключает из массива данных ошибочные данные и выброс; выделяет входные и выходные переменные с целью использования предиктивных моделей
ПК-18 Способен выполнять анализ больших данных	Не знает общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных; принципы и методы анализа больших данных Не умеет использовать программное обеспечение для очистки и валидации наборов больших данных; описывать и управлять качеством и достоверностью больших данных; применять программное обеспечение для анализа больших данных; применять программные и технические средства визуализации больших данных и результатов их анализа	Знает общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных; принципы и методы анализа больших данных Умеет использовать программное обеспечение для очистки и валидации наборов больших данных; описывать и управлять качеством и достоверностью больших данных; применять программное обеспечение для анализа больших данных; применять программные и технические средства визуализации больших данных и результатов их анализа	Демонстрирует глубокое знание и понимание: общедоступных репозиториях и специализированных библиотек, содержащих наборы больших данных; принципов и методов анализа больших данных Полностью верно и самостоятельно: использует программное обеспечение для очистки и валидации наборов больших данных; описывает и управляет качеством и достоверностью больших данных; применяет программное обеспечение для анализа больших данных; применяет программные и технические средства визуализации больших данных и результатов их анализа

5 Учебно-методическое и информационное обеспечение дисциплины

5.1 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

1. Никитина, Т. П. Программирование. Основы Python / Т. П. Никитина, Л. В. Королев. – Санкт-Петербург: Лань, 2023. – 156 с. – ISBN 978-5-507-45283-5. – Текст:

- электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/302714> (дата обращения: 05.06.2023).
2. Титов, А. Н. Python. Обработка данных: учебно-методическое пособие / А.Н. Титов, Р.Ф. Тагиева. – Казань: Изд-во КНИТУ, 2022. - 104 с. - ISBN 978-5-7882-3171-6. – Текст: электронный. – URL: <https://znanium.com/catalog/product/2069264> (дата обращения: 05.06.2023).
 3. Маккинни, У. Python и анализ данных : практическое пособие / У. Маккинни ; пер. с англ. А.А. Слинкина. – 2-е изд. – Москва: ДМК Пресс, 2020. – 540 с. - Текст: электронный. – URL: <https://znanium.com/catalog/product/2012523> (дата обращения: 13.06.2023).
 4. Васильев, А.Н. Числовые расчеты в Excel: Учебное пособие. – СПб.: Издательство «Лань», 2022. – 608 с.: – URL: <https://reader.lanbook.com/book/212198#3> (дата обращения: 13.06.2023)э

Дополнительная литература:

1. Замятин, А. В. Введение в интеллектуальный анализ данных: учебное пособие / А. В. Замятин. - Томск: Издательство Томского государственного университета, 2016. - 120 с. - ISBN 978-5-94621-531-2. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1663560> (дата обращения: 13.06.2023)
2. Жуков, Р. А. Язык программирования Python: практикум: учебное пособие / Р.А. Жуков. – Москва: ИНФРА-М, 2023. – 216 с.+ Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/textbook_5cb5ca35aaa7f5.89424805. - ISBN 978-5-16-016971-2. – Текст: электронный. – URL: <https://znanium.com/catalog/product/1915716> (дата обращения: 05.06.2023).

5.2 Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для выполнения лабораторных работ.

5.3 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

1. Информационно-справочная система «КонсультантПлюс. – URL: <http://www.consultant.ru/>
2. Программирование на Python. URL: <https://stepik.org/course/67/syllabus> (дата доступа 04.06.2023).
3. Python – обучающий курс от Сергея Балакирева. URL: <https://stepik.org/course/100707/promo> (дата доступа 04.06.2023)
4. Основы статистики – URL: <https://stepik.org/course/76/syllabus> (дата доступа 04.06.2023)
5. Научная электронная библиотека eLIBRARY.RU – URL: <http://elibrary.ru>
6. ЭБС ЛАНЬ – URL: <https://e.lanbook.com/>,
7. ЭБС Znanium.com – URL: <https://znanium.com/>

5.4 Перечень информационных технологий, используемых при изучении дисциплины, включая программное обеспечение

1. ОС семейства Microsoft Windows.
2. MS Office 365.
3. Браузер.

4. Язык программирования Python – URL: <https://www.python.org/>
5. Среда программирования на языке *Python*, например, *JupyterLab* URL: <https://jupyter.org/>.
6. NumPy – пакет для научных вычислений с Python. – URL: <https://numpy.org/>
7. Программное обеспечение с открытым исходным кодом для математики, науки и техники – URL: <https://scipy.org/>
8. Библиотека на языке программирования Python для визуализации данных двумерной и трёхмерной графикой Matplotlib – URL: <https://matplotlib.org/>
9. Инструмент для анализа и обработки данных с открытым исходным кодом Pandas – URL: <https://pandas.pydata.org/>
10. Anaconda – платформа для быстрой разработки и развертывания безопасных решений Python – URL: <https://www.anaconda.com/>
11. Google Colab – URL: https://colab.research.google.com/#scrollTo=5fCEDCU_qrC0
12. API глубокого обучения Keras – URL: <https://keras.io/>
13. Комплексная платформа машинного обучения TensorFlow – URL: <https://www.tensorflow.org/>
14. Машинное обучение с открытым исходным кодом и визуализация данных Orange Data Mining – URL: <https://orangedatamining.com/>
15. Data Analytics Platform KNIME – URL: <https://www.knime.com/knime-analytics-platform>
16. Некоммерческий проект с открытым исходным кодом Project Jupyter – URL: <https://jupyter.org/>