



КГЭУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)

УТВЕРЖДАЮ

Директор института цифровых
технологий и экономики

Э.И. Беляев

29 ноября 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.04 Сбор и верификация данных

Направление подготовки

09.03.03 Прикладная информатика

Направленность (профиль)

Прикладной искусственный интеллект



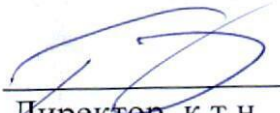
Квалификация

Бакалавр

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч. степень, уч. звание	ФИО разработчика
УрФУ АБДМВ	Доцент, к.п.н.	Новиков М.Ю.
КГЭУ ИТИС	Доцент, к.т.н., доцент	Хамитов Р.М.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	ИТИС	27.11.23	11	 И.о. зав. кафедрой, к.ф.-м.н., доц., Соловьев С.А.
Согласована	Учебно-методический совет ИЦТЭ	27.11.23	3	 Директор, к.т.н., Беляев Э.И.
Одобрена	Ученый совет ИЦТЭ	28.11.23	3	 Директор, к.т.н., Беляев Э.И.

Рецензия на рабочую программу и оценочные материалы по дисциплине « Б1.В.04 Сбор и верификация данных»

Содержание РПД и ОМ соответствует требованиям федерального государственного стандарта высшего образования по направлению подготовки 09.03.03 «Прикладная информатика» и учебному плану.

РПД и ОМ соответствуют требованиям, предъявляемым к структуре, содержанию РПД и ОМ по дисциплине, а именно:

1. Перечень формируемых компетенций, которыми должен овладеть обучающийся в результате освоения дисциплины, соответствует ФГОС ВО и профстандарту, будущей профессиональной деятельности выпускника.

2. Структура и содержание дисциплины соответствует учебному плану.

3. РПД содержит информацию об учебно-методическом, информационном и материально-техническом обеспечении дисциплины; об особенностях организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов и методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

4. Показатели и критерии оценивания компетенций в ОМ, а также шкалы оценивания обеспечивают возможность проведения всесторонней оценки результаты обучения, уровней сформированности компетенций.

5. Контрольные задания и иные материалы оценки результатов освоения разработаны на основе принципов оценивания: валидности, определённости, однозначности, надёжности, а также соответствуют требованиям к составу и взаимосвязи оценочных средств, полноте по количественному составу оценочных средств и позволяют объективно оценить результаты обучения, уровни сформированности компетенций.

6. Направленность РПД и ОМ по дисциплине соответствует целям ОП по направлению 09.03.03 «Прикладная информатика», профстандартам.

Заключение. На основании проведенной экспертизы можно сделать заключение, что РПД и ОМ по дисциплине соответствует требованиям ФГОС ВО, профессионального стандарта, современным требованиям рынка труда и рекомендуются для использования в учебном процессе.

Рецензент:

Генеральный директор
ООО "ЛПТСИСТЕМС"



Г.А. Фатыхова

Фатыхова Г.А.

25.11.2023

1. Цель, задачи и планируемые результаты обучения по дисциплине

(Цель и задачи освоения дисциплины, соответствующие цели ОП)

Целью освоения дисциплины Сбор и верификация данных является формирование у студентов компетенций в области сбора, подготовки и разметки данных для решения задач машинного обучения. От качества подобранных данных и их предварительной обработки зависит решение многих задач искусственного интеллекта. Студент в ходе обучения научится понимать какие процессы являются источниками данных, как данные создаются, как данные можно обрабатывать (алгоритмы и методы), как проводить предварительную подготовку данных, в каких форматах можно сохранять данные, как производить разметку и визуализацию данных для дальнейшего использования в задачах машинного обучения.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения;	ПК-1.1 Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей ПК-1.2 Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей
ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта;	ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения
ПК-10. Способен решать задачи профессиональной деятельности в проектном формате для достижения заданной цели и создания уникального продукта, услуги или результата с заданным качеством в условиях ограниченности ресурсов (временных, финансовых, человеческих, информационных), осознавая свою роль и ответственность в проекте	ПК-10.1 Решает задачи профессиональной деятельности в проектном формате для достижения заданной цели и создания уникального продукта, услуги или результата с заданным качеством ПК-10.2 Определяет возможные ограничения ресурсов (временных, финансовых, информационных и человеческих) в проектной деятельности ПК-10.3 Определяет свою роль и ответственность в команде при реализации проекта

2. Место дисциплины в структуре ОП

Предшествующие дисциплины (модули), практики, НИР, др. Информационные технологии, Программирование, Компьютерные сети, Информационная безопасность, Объектно-ориентированное программирование, Базы данных, Проектирование информационных систем.

Последующие дисциплины (модули), практики, НИР, др. Подсистемы хранения и извлечения данных, Системы принятия решений на основе искусственного интеллекта, Разработка систем анализа больших данных, Нереляционные базы данных, Рекомендательные системы, Пакетная и потоковая обработка данных, Прикладные и наукоемкие задачи искусственного интеллекта, Инструменты решения задач искусственного интеллекта, Платформы бизнес-приложений и языки программирования, Платформы бизнес-анализа, Приложения искусственного интеллекта, Производственная практика (преддипломная), Выполнение и защита выпускной квалификационной работы.

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Для очной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			5
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	5	180	180
КОНТАКТНАЯ РАБОТА*	-	95	95
АУДИТОРНАЯ РАБОТА	1,33	48	48
Лекции	0,44	16	16
Практические (семинарские) занятия	-	-	-
Лабораторные работы	0,89	32	32
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	3,67	132	132
Проработка учебного материала	0,67	24	24
Курсовой проект	2	72	72
Курсовая работа	-	-	
Подготовка к промежуточной аттестации	1	36	36
Промежуточная аттестация:			Э
			КП

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Сбор и верификация данных.	18	4	8		6	ТК1	ПК-1.3, ПК-1.У ПК-6.3, ПК-6.У
Раздел 2. Очистка данных	18	4	8		6	ТК2 ТК3	ПК-1.3, ПК-1.У, ПК-6.3, ПК-6.У
Раздел 3. Маркировка данных	18	4	8		6		ПК-1.3, ПК-1.У ПК-6.3, ПК-6.У
Раздел 4. Проверка и визуализация данных	18	4	8		6		ПК-1.3, ПК-1.У ПК-6.3, ПК-6.
Курсовой проект	72				72	ОМкп	ПК-10.В
Экзамен	36				36	ОМ	ПК-1.3, ПК-1.У, ПК-10.У ПК-6.3, ПК-6.У, ПК-10.В
Итого за 5 семестр	180	16	32		132		
ИТОГО	180	16	32		132		

3.3. Содержание дисциплины

Раздел 1. Сбор и верификация данных.

1.1 Основные принципы сбора и первичной обработки данных. Сигналы, Общие сведения о сигналах, Классификация сигналов, Первичные эмпирические сведения о социальных и экономических фактах, Парсинг (скраппинг) сайтов

1.2 Создание данных. Дискретизация и квантование, Первичный анализ данных с Pandas, Базовые операции в Pandas

1.3 Математические модели, методы и алгоритмы сбора и верификации наукометрических данных. Применение функций к ячейкам, столбцам и строкам, Фильтрация кадров данных, Объединение кадров данных, Изучение данных, Сортировка, Группировка данных.

Раздел 2. Очистка данных.

2.1 Поиск и устранение ошибок в данных. Заполнение недостающих данных. Поиск ошибок в данных. Очистка данных. Заполнение недостающих данных в pandas DataFrame.

2.2 Форматы данных. Источники данных поддерживаемые в Pandas. Чтение из SQL баз данных. Запись в базу.

2.3 Качество данных. Определения Измерения качества данных. Контроль качества данных. Оптимальное использование качества данных

Раздел 3. Маркировка данных

Разметка данных. Сценарии применения размеченных данных в задачах компьютерного зрения и обработки естественного языка. Семантическая сегментация изображений в CVAT. Разметка именованных сущностей в Label

Studioю. Сценарии применения размеченных данных в задачах компьютерного зрения и обработки естественного языка

Раздел 4. Проверка и визуализация данных

Обработка данных. Закономерности и аномалии. Визуализация данных. Обзор набора данных. Seaborn. Plotly. Пример визуального анализа данных

3.4. Тематический план практических занятий

«Данный вид работы не предусмотрен учебным планом».

3.5. Тематический план лабораторных работ

Раздел 1. Сбор и верификация данных.

1.1 Парсинг данных.

1.2 Математические модели, методы и алгоритмы сбора и верификации наукометрических данных.

Раздел 2. Очистка данных.

2.1 Поиск и устранение ошибок в данных в pandas DataFrame.

2.2 Контроль качества данных.

Раздел 3. Маркировка данных

3.1 Разметка данных.

3.2 Сценарии применения размеченных данных в задачах компьютерного зрения и обработки естественного языка.

Раздел 4. Проверка и визуализация данных

4.1 Обработка данных в pandas. Закономерности и аномалии.

4.2 Визуализация данных.

3.6. Курсовой проект /курсовая работа

Примерные темы курсовых проектов.

1. Сбор и обработка данных о вакансиях.

2. Сбор и обработка данных интернет-магазина

3. Сбор и обработка данных о продаже недвижимости.

4. Сбор и обработка данных от аналого-цифровых преобразователей.

5. Сбор и обработка данных с аналоговых приборов.

6. Сбор и обработка данных системы видеонаблюдения.

7. Сбор и обработка данных стационарных датчиков и приборов.

8. Сбор и обработка данных передвижных датчиков и приборов.

9. Сбор и обработка данных о преступности и безопасности общества.

10. Сбор и обработка данных с устройств интернета вещей IoT.

11. Сбор и агрегация данных с корпоративных информационных систем.

12. Сбор и обработка данных с датчиков смартфона.

13. Сбор и обработка данных с носимой электроники.

14. Сбор и обработка данных и файлов логирования операционной системы.

15. Сбор и обработка данных социальных сетей и интернет-форумов.

16. Сбор и анализ данных клиентских отзывов и рейтингов.

17. Сбор и обработка данных медицинских приборов и датчиков.

18. Сбор и анализ данных финансовых рынков и бирж.
19. Сбор и обработка данных о погоде и климате.
20. Сбор и анализ данных трафика и поведения пользователей в сети Интернет.
21. Сбор и обработка данных о продукции и производственных процессах.
22. Сбор и анализ данных о транспортных потоках и движении транспорта.
23. Сбор и обработка данных о состоянии окружающей среды и экологических параметрах.
24. Сбор и анализ данных об энергопотреблении и энергосистемах.
25. Сбор и обработка данных о потреблении товаров и услуг потребителями.
26. Сбор и анализ данных о поведении животных и их окружении.

4. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-1	ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей	знать:				
		признаки классификации систем искусственного интеллекта	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		уметь:				
	Идентифицировать задачи систем искусственного интеллекта	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения	
	ПК-1.2. Собирает исходную информацию	знать:				
		Знает методы сбора и обобщения	Уровень знаний в объеме,	Уровень знаний в объеме,	Минимально допустим	Уровень знаний ниже

	<p>ю и формирует требования к решению задач с использованием методов искусственного интеллекта</p>	<p>информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников знаний, а также формирования требований к системе искусственного интеллекта</p>	<p>соответствующем программе подготовки, без ошибок</p>	<p>соответствующем программе, имеет место несколько негрубых ошибок</p>	<p>ый уровень знаний, имеет место много негрубых ошибок</p>	<p>минимальных требований, имеют место грубые ошибки</p>
<p>уметь:</p>						
		<p>осуществлять сбор и обобщение информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной области, документированных источников знаний, а также формировать требования к системе искусственного интеллекта</p>	<p>Сформированы умения</p>	<p>Частично сформированы умения</p>	<p>Сформированы базовые умения</p>	<p>Не сформированы базовые умения</p>
		<p>осуществлять сбор исходной информации с использованием платформ данных</p>	<p>Сформированы умения</p>	<p>Частично сформированы умения</p>	<p>Сформированы базовые умения</p>	<p>Не сформированы базовые умения</p>

		(облачных и внутрикорпоративных)				
		владеть:				
		навыками сбора исходной информации и формировать требования к решению задач с использованием методов искусственного интеллекта	Продемонстрированы навыки	Сформированы базовые навыки	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки
ПК-6	ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях	знать:				
		Источники и методы поиска данных в открытых источниках, специализированных библиотеках и репозиториях	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		уметь:				
	осуществлять поиск данных в открытых источниках, специализированных библиотеках и репозиториях	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения	
	ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения	знать:				
		инструменты и методы подготовки и разметки структурированных и неструктурированных данных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
уметь:						
осуществлять подготовку и	Сформированы	Частично сформированы	Сформированы	Не сформированы		

		разметку структурированных и неструктурированных данных для машинного обучения	умения	базовые умения	базовые умения	базовые умения	
ПК-10	ПК 10.1 Решает задачи профессиональной деятельности и в проектном формате для достижения заданной цели и создания уникального продукта, услуги или результата с заданным качеством	владеть:					
		Навыками решения задач профессиональной деятельности в проектном формате	Продемонстрированы навыки	Сформированы базовые навыки запросов	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки, допущены значительные ошибки.	
	ПК-10.2 Определяет возможные ограничения ресурсов (временных, финансовых, информационных и человеческих) в проектной деятельности	уметь:					
		определять возможные ограничения ресурсов (временных, финансовых, информационных и человеческих) в проектной деятельности	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения	
	ПК-10.3 Определяет свою роль и ответственность в команде при реализации проекта	уметь:					
		определять свою роль и ответственность в команде при реализации проекта	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения	

Оценочные материалы для проведения текущего контроля и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины.

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика.

5. Учебно-методическое и информационное обеспечение дисциплины

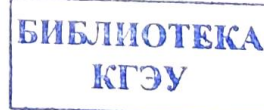
5.1. Учебно-методическое обеспечение

5.1.1. Основная литература

1. Никитина, Т. П. Программирование. Основы Python для инженеров / Т. П. Никитина, Л. В. Королев. — Санкт-Петербург : Лань, 2023. — 156 с. — ISBN 978-5-507-45284-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/302720>
2. Наместников, А. М. Базы данных. Практический курс : учебное пособие : в 2 частях / А. М. Наместников. — Ульяновск : УлГТУ, 2017 — Часть 1 : Объектно-реляционные базы данных на примере PostgreSQL 9.5 — 2017. — 113 с. — ISBN 978-5-9795-1743-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165100>.
3. Дадян, Эдуард Григорьевич. Данные: хранение и обработка : учебник / Э. Г. Дадян. - Москва : ИНФРА-М, 2021. - 205 с. - URL: <http://new.ibooks.ru/bookshelf/377999>. - ISBN 978-5-16-016447-2 (print). - ISBN 978-5-16-107936-2 (online). - Текст : электронный.
4. Волк, В. К. Базы данных. Проектирование, программирование, управление и администрирование / В. К. Волк. — 4-е изд., стер. — Санкт-Петербург : Лань, 2023. — 244 с. — ISBN 978-5-507-47243-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/346439>.
5. Тарасов С.В. СУБД для программиста. Базы данных изнутри / С.В. Тарасов. - Москва : СОЛОН-ПРЕСС, 2018. - 320 с. - ISBN 978-2-7466-7383-0. - URL: <https://ibooks.ru/bookshelf/344900/reading>. - Текст: электронный.
6. Цехановский, В. В. Управление данными : учебник / В. В. Цехановский, В. Д. Чертовской. — Санкт-Петербург : Лань, 2022. — 432 с. — ISBN 978-5-8114-1853-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/212084>.

5.1.2. Дополнительная литература

1. Масленникова О.Е. Основы искусственного интеллекта. Учебное пособие / О.Е. Масленникова, И.В. Гаврилова. - Москва : Флинта, 2019. - 283 с. - ISBN 978-5-9765-1602-1. - URL: <https://ibooks.ru/bookshelf/337972>
2. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 228 с. — ISBN 978-5-507-46441-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/310199>
3. Селянкин, В. В. Компьютерное зрение. Анализ и обработка изображений / В. В. Селянкин. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 152 с. — ISBN 978-5-507-45583-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/>



5.2. Информационное обеспечение

5.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	https://e.lanbook.com/
2	Электронно-библиотечная система «ibooks.ru»	https://ibooks.ru/
3	Электронно-библиотечная система «book.ru»	https://www.book.ru/
4	Портал «Открытое образование»	https://npoed.ru
5	Российская национальная библиотека	https://nlr.ru/
6	КиберЛенинка	https://cyberleninka.ru
7	Техническая библиотека	https://techlibrary.ru
8	Национальная электронная библиотека (НЭБ)	https://rusneb.ru/
9	Веб-сервис для хостинга IT-проектов и их совместной разработки Github	http://www.github.ru

5.2.2. Профессиональные базы данных / Информационно-справочные системы

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	Официальный интернет-портал правовой информации	http://pravo.gov.ru	http://pravo.gov.ru
2	Справочная правовая система «Консультант Плюс»	http://consultant.ru	http://consultant.ru
3	Справочно-правовая система по законодательству РФ	http://garant.ru	http://garant.ru

5.2.3. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Описание	Реквизиты подтверждающих документов
1	Операционная система Microsoft Windows 10	Пользовательская операционная система	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
2	Microsoft Office 2019	Пакет офисных приложений	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно

3	LMS Moodle	ПО для эффективного онлайн- взаимодействия преподавателя и студента	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
4	Браузер Chrome	Система поиска информации в сети интернет	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
5	СУБД PostgreSQL	Система управления базами данных https://www.postgresql.org/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
6	pgAdmin для PostgreSQL	Система администрирования и разработки pgAdmin для PostgreSQL https://www.pgadmin.org/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
7	DBeaver Community	Free Universal Database Tool https://dbeaver.io/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.

6. Материально-техническое обеспечение дисциплины

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
Лекции	Учебная аудитория для проведения занятий лекционного типа	Специализированная учебная мебель, технические средства обучения, служащие для представления учебной информации большой аудитории (мультимедийный проектор, компьютер (ноутбук), экран), демонстрационное оборудование, учебно-наглядные пособия
Лабораторные работы	Учебная лаборатория программной инженерии, ауд. В-608	Специализированное лабораторное оборудование по профилю лаборатории программной инженерии, специализированная учебная мебель на 50 посадочных мест, 24 компьютера с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-610	Специализированная учебная мебель на 42 посадочных места, 17 компьютеров с

		возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория информационной безопасности, ауд. В-615	Специализированное лабораторное оборудование по профилю лаборатории информационной безопасности, специализированная учебная мебель на 35 посадочных мест, 15 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-617	Специализированная учебная мебель на 24 посадочных места, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-619	Специализированная учебная мебель на 26 посадочных мест, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-621	Специализированная учебная мебель на 35 посадочных мест, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория реинжиниринга и управления бизнес-процессами, ауд. В-623	Специализированное лабораторное оборудование по профилю лаборатории реинжиниринга и управления бизнес-процессами, специализированная учебная мебель на 34 посадочных места, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения

		(мультимедийный проектор, ноутбук, экран), видеокамеры, необходимое лицензионное программное обеспечение
Самостоятельная работа	Компьютерный класс с возможностью выхода в Интернет и обеспечением доступа в ЭИОС В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультиме- дийный проектор, компьютер (ноутбук), экран), видеокамеры, программное обеспечение
	Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, экран, мультимедийный проектор, программное обеспечение

7. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www//kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную

консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;

- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;

- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;

- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;

- обеспечивается необходимый уровень освещенности помещений;

- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

8. Методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

Методическое обеспечение процесса воспитания обучающихся выступает одним из определяющих факторов высокого качества образования. Преподаватель вуза, демонстрируя высокий профессионализм, эрудицию, четкую гражданскую позицию, самодисциплину, творческий подход в решении профессиональных задач, в ходе образовательного процесса способствует формированию гармоничной личности.

При реализации дисциплины преподаватель может использовать следующие методы воспитательной работы:

- методы формирования сознания личности (беседа, диспут, внушение, инструктаж, контроль, объяснение, пример, самоконтроль, рассказ, совет, убеждение и др.);

- методы организации деятельности и формирования опыта поведения (задание, общественное мнение, педагогическое требование, поручение, приучение, создание воспитывающих ситуаций, тренинг, упражнение, и др.);

- методы мотивации деятельности и поведения (одобрение, поощрение социальной активности, порицание, создание ситуаций успеха, создание ситуаций для эмоционально-нравственных переживаний, соревнование и др.)

При реализации дисциплины преподаватель должен учитывать следующие направления воспитательной деятельности:

Гражданское и патриотическое воспитание:

- формирование у обучающихся целостного мировоззрения, российской идентичности, уважения к своей семье, обществу, государству, принятым в семье и обществе духовно-нравственным и социокультурным ценностям, к национальному, культурному и историческому наследию, формирование стремления к его сохранению и развитию;

- формирование у обучающихся активной гражданской позиции, основанной на традиционных культурных, духовных и нравственных ценностях российского общества, для повышения способности ответственно реализовывать свои конституционные права и обязанности;

- развитие правовой и политической культуры обучающихся, расширение конструктивного участия в принятии решений, затрагивающих их права и интересы, в том числе в различных формах самоорганизации, самоуправления, общественно-значимой деятельности;

- формирование мотивов, нравственных и смысловых установок личности, позволяющих противостоять экстремизму, ксенофобии, дискриминации по социальным, религиозным, расовым, национальным признакам, межэтнической и межконфессиональной нетерпимости, другим негативным социальным явлениям.

Духовно-нравственное воспитание:

- воспитание чувства достоинства, чести и честности, совестливости, уважения к родителям, учителям, людям старшего поколения;

- формирование принципов коллективизма и солидарности, духа милосердия и сострадания, привычки заботиться о людях, находящихся в трудной жизненной ситуации;

- формирование солидарности и чувства социальной ответственности по отношению к людям с ограниченными возможностями здоровья, преодоление психологических барьеров по отношению к людям с ограниченными возможностями;

- формирование эмоционально насыщенного и духовно возвышенного отношения к миру, способности и умения передавать другим свой эстетический опыт.

Культурно-просветительское воспитание:

- формирование эстетической картины мира;

- формирование уважения к культурным ценностям родного города, края, страны;

- повышение познавательной активности обучающихся.

Научно-образовательное воспитание:

- формирование у обучающихся научного мировоззрения;

- формирование умения получать знания;

- формирование навыков анализа и синтеза информации, в том числе в профессиональной области.

Вносимые изменения и утверждения на новый учебный год

№ п/п	№ раздела внесения изменений	Дата внесения изменений	Содержание изменений	«Согласовано» Зав. каф. реализующей дисциплину	«Согласовано» председатель УМК института (факультета), в состав которого входит выпускающая
1	2	3	4	5	6
1					
2					
3					

*Приложение к рабочей
программе дисциплины*



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУУ»)**

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
по дисциплине**

Б1.В.04 Сбор и верификация данных

(Наименование дисциплины в соответствии с учебным планом)

Направление подготовки **09.03.03 Прикладная информатика**

(Код и наименование направления подготовки)

Квалификация **Бакалавр**

(Бакалавр / Магистр)

г. Казань, 2023

2. Оценочные материалы текущего контроля и промежуточной аттестации

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-1	ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей	знать:				
		признаки классификации систем искусственного интеллекта	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		уметь:				
		Идентифицировать задачи систем искусственного интеллекта	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
ПК-1	ПК-1.2. Собирает исходную информацию и формирует требования к решению задач с использованием методов искусственного интеллекта	знать:				
		Знает методы сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки

		знаний, а также формирования требований к системе искусственного интеллекта				
		уметь:				
		осуществлять сбор и обобщение информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной области, документированных источников знаний, а также формировать требования к системе искусственного интеллекта	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		осуществлять сбор исходной информации с использованием платформ данных (облачных и внутрикорпоративных)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		владеть:				

		навыками сбора исходной информации и формировать требования к решению задач с использованием методов искусственного интеллекта	Продемонстрированы навыки	Сформированы базовые навыки	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки
ПК-6	ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях	знать:				
		Источники и методы поиска данных в открытых источниках, специализированных библиотеках и репозиториях	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		уметь:				
	осуществлять поиск данных в открытых источниках, специализированных библиотеках и репозиториях	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения	
	ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения	знать:				
		инструменты и методы подготовки и разметки структурированных и неструктурированных данных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
уметь:						
осуществлять подготовку и разметку структурированных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые		

		нных и неструктурированных данных для машинного обучения				умения
ПК-10	ПК 10.1 Решает задачи профессиональной деятельности и в проектном формате для достижения заданной цели и создания уникального продукта, услуги или результата с заданным качеством	Навыками решения задач профессиональной деятельности в проектном формате	Продемонстрированы навыки	Сформированы базовые навыки запросов	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки, допущены значительные ошибки.
	ПК-10.2 Определяет возможные ограничения ресурсов (временных, финансовых, информационных и человеческих) в проектной деятельности	определять возможные ограничения ресурсов (временных, финансовых, информационных и человеческих) в проектной деятельности	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения
	ПК-10.3 Определяет свою роль и ответственность в команде при реализации проекта	определять свою роль и ответственность в команде при реализации проекта	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения

Оценка «отлично» выставляется за выполнение лабораторных работ в семестре; тестовых заданий; глубокое понимание методов сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных

источников знания, а также формирования требований к системе искусственного интеллекта; знание инструментов и методов подготовки и разметки структурированных и неструктурированных данных

Оценка «хорошо» выставляется за выполнение лабораторных работ в семестре; тестовых заданий; понимание глубокое понимание методов сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников знания, а также формирования требований к системе искусственного интеллекта; знание инструментов и методов подготовки и разметки структурированных и неструктурированных данных;

Оценка «удовлетворительно» выставляется за выполнение лабораторных работ в семестре и тестовых заданий;

Оценка «неудовлетворительно» выставляется за слабое и неполное выполнение лабораторных работ в семестре и тестовых заданий.

3. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Описание оценочного средства
Курсовой проект (КП), курсовая работа (КР)	Конечный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий. Позволяет оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитических, исследовательских навыков, навыков практического и творческого мышления. Может выполняться в индивидуальном порядке или группой обучающихся	Темы проектов
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов испытаний, измерений, эксперимента. Оформление отчета, защита результатов лабораторной работы по отчету	Перечень заданий и вопросов для защиты лабораторной работы, перечень требований к отчету
Тест (Тест)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося	Комплект тестовых заданий

4. Перечень контрольных заданий или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины

Пример задания

Для текущего контроля ТК1:

Проверяемая компетенция: ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения; ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей; ПК-1.2. Собирает исходную информацию и формирует требования к решению задач с использованием методов искусственного интеллекта;

ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта; ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях; ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения.

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 2 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 1. Сбор и верификация данных.

1.1 Парсинг данных.

1.2 Математические модели, методы и алгоритмы сбора и верификации наукометрических данных.

Последовательность парсинга, сбора и верификации наукометрических данных

1. Убедитесь, что источники достоверные. Проверьте источники наукометрических данных и базы цитирования, чтобы удостовериться в их авторитетности.
2. Установите Beautiful Sou, при помощи команды «`pip install beautifulsoup4`». Импортируйте BeautifulSoup и requests в вашем скрипте Python.
3. Используйте библиотеку requests для получения HTML-кода страницы. Как представлены данные после получения HTML-кода страницы?
4. Используйте методы BeautifulSoup для поиска и извлечения данных из

HTML. Какие методы применялись для поиска извлечения данных?
Как теперь представлены данные?

5. Анализируйте извлеченные данные и применяйте необходимые математические модели, методы и алгоритмы. Какие математические модели применялись и почему?

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 8 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 6 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки - 4 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 8 баллов.

Максимальное количество баллов за все задания раздела – 16 баллов.

Для текущего контроля ТК2:

Проверяемая компетенция: ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения; ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей; ПК-1.2. Собирает исходную информацию и формирует требования к решению задач с использованием методов искусственного интеллекта;

ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта; ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях; ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения.

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 2 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 2. Очистка данных.

2.1 Поиск и устранение ошибок в данных в pandas DataFrame.

2.2 Контроль качества данных.

Устранение ошибок в данных вакансий

1. Установите библиотеку pandas, при помощи команды «pip install pandas». Импортируйте pandas в свой скрипт Python.

2. Прочитайте файл с информацией о вакансиях. Ознакомление с данными, выведите первые несколько строк DataFrame, чтобы понять структуру данных. Проанализируйте, какие ошибки есть в данных? Какие данные отсутствуют? Есть ли дублирующиеся данные? Данные в каждом столбце соответствуют содержимому?

3. Устраните ошибку в pandas DataFrame. При необходимости удалите столбцы, которые не подходят по содержимому или содержат не применимые данные.

4. Проверьте, имеют ли данные теперь всю необходимую и имеют ли ошибки? Данные приведены к единообразному формату?

5. Сохраните DataFrame в файл CSV. Сравните изначальный файл и файл после очистки данных. Почему было необходимо очистить данные в изначальном файле?

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 8 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 6 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки - 4 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 8 баллов.

Максимальное количество баллов за все задания раздела – 16 баллов.

Для текущего контроля ТКЗ:

Проверяемая компетенция: ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные

методы и инструментальные средства решения; ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей; ПК-1.2. Собирает исходную информацию и формирует требования к решению задач с использованием методов искусственного интеллекта;

ПК-6. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта; ПК-6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях; ПК-6.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 2 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 3. Маркировка данных

3.1 Разметка данных.

3.2 Сценарии применения размеченных данных в задачах компьютерного зрения и обработки естественного языка.

Темы лабораторных работ:

1. Подготовка текстового набора данных к обработке нейронной сетью.
2. Представление текста в формате one hot encoding"
3. Представление текста плотными векторами (embeddings)
4. Как подготовить собственный набор изображений для обучения нейронной сети в Keras".

Раздел 4. Проверка и визуализация данных

4.1 Обработка данных в pandas. Закономерности и аномалии.

4.2 Визуализация данных.

Статистика:

1. Установите библиотеки matplotlib и pandas. Импортируйте эти библиотеки.

2. Прочитайте данные вакансий. Ознакомьтесь с данными. Проведите очистку данных, удалите дубликаты и нулевые значения. Данные готовы для анализа и визуализации?

3. Используйте методы pandas для анализа данных, получив статистическую информацию о числовых столбцах. Какие выводы можно сделать исходя из полученных данных?

4. Используйте методы pandas и проведите анализ данных для выявления закономерностей и аномалий. Какие выводы можно сделать исходя из полученных данных?

5. Используйте matplotlib для создания графиков и визуализации данных. В результате, можно ли подтвердить выводы, полученные статистически? На графиках можно ли отследить аномалии и закономерности в данных?

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 8 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 6 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки – 4 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 8 баллов.

Максимальное количество баллов за все задания раздела – 32 балла.

Курсовой проект

Каждому студенту выдается тема курсового проекта. Всего 25 тем. Тематика курсового проекта утверждается на заседании кафедры.

Целью курсового проектирования формирование практических навыков сбора и верификации данных.

В процессе курсового проектирования обучающийся должен:

- проанализировать предметную область в контексте задания, сформулировать требования к системе; описать источники данных, структуру и состав реквизитов;
- организовать процесс сбора и верификация данных; при необходимости выполнить очистку данных;
- выполнить маркировку, проверку и визуализацию данных

Примеры тем:

1. Сбор и обработка данных о вакансиях.
2. Сбор и обработка данных интернет-магазина
3. Сбор и обработка данных о продаже недвижимости.
4. Сбор и обработка данных от аналого-цифровых преобразователей.
5. Сбор и обработка данных с аналоговых приборов.
6. Сбор и обработка данных системы видеонаблюдения.
7. Сбор и обработка данных стационарных датчиков и приборов.
8. Сбор и обработка данных передвижных датчиков и приборов.
9. Сбор и обработка данных о преступности и безопасности общества.
10. Сбор и обработка данных с устройств интернета вещей IoT.
11. Сбор и агрегация данных с корпоративных информационных систем.
12. Сбор и обработка данных с датчиков смартфона.
13. Сбор и обработка данных с носимой электроники.
14. Сбор и обработка данных и файлов логирования операционной системы.
15. Сбор и обработка данных социальных сетей и интернет-форумов.
16. Сбор и анализ данных клиентских отзывов и рейтингов.
17. Сбор и обработка данных медицинских приборов и датчиков.
18. Сбор и анализ данных финансовых рынков и бирж.
19. Сбор и обработка данных о погоде и климате.
20. Сбор и анализ данных трафика и поведения пользователей в сети Интернет.
21. Сбор и обработка данных о продукции и производственных процессах.
22. Сбор и анализ данных о транспортных потоках и движении транспорта.
23. Сбор и обработка данных о состоянии окружающей среды и экологических параметрах.
24. Сбор и анализ данных об энергопотреблении и энергосистемах.
25. Сбор и обработка данных о потреблении товаров и услуг потребителями.
26. Сбор и анализ данных о поведении животных и их окружении.

Уровень освоения- высокий: 13-15 баллов. Поставленная задача реализована полностью в виде готового проекта информационной системы.

Учтены все показатели предметной области. Присутствуют все отчеты. Разработанные документы, модели и схемы не содержат ошибок.

Уровень освоения- средний: 9-12 баллов. Поставленная задача реализована полностью в виде готового проекта информационной системы. Учтены не все показатели предметной области. Присутствуют все отчеты. Разработанные документы, модели и схемы содержат незначительные ошибки.

Уровень освоения- ниже среднего: 6-8 баллов. Поставленная задача реализована в виде готового проекта информационной системы не полностью. Учтены не все показатели предметной области. Присутствуют не все отчеты. Разработанные документы, модели и схемы содержат ошибки.

Уровень освоения- низкий менее 6 баллов. Поставленная задача реализована частично, отчетов недостаточно, не учтены критичные для реализации показатели предметной области. Отсутствуют важные документы, модели и схемы.

Контроль проводится в три этапа.

I этап. Анализ предметной области и планирование проекта информационной системы.

На этом этапе студенты обязаны представить черновой вариант Части I пояснительной записки, касающейся:

- определение целей и задач курсового проекта;
- изучение литературы и источников по выбранной теме сбора и обработки данных;
- сбор и анализ существующих методов и инструментов для сбора и обработки данных, выбор наиболее подходящих для проекта;
- определение источников данных для сбора информации.

II этап. На этом этапе студенты обязаны представить отредактированный вариант Части II пояснительной записки, касающейся:

- выбрать программные и технические средства для реализации информационной системы;
- определить информационную базу задачи и, соответственно, структуру базы данных проектируемой информационной системы;
- разработка методики сбора, очистки и верификации данных из выбранных источников;
- сбор данных с помощью выбранных методик.

III этап. Заключительный: Завершение выполнения курсового проекта.

- маркировка и визуализация собранных данных с использованием соответствующих инструментов и программ.
- Анализ полученных результатов и выводы по проведенной работе.
- Написание отчета о курсовом проекте с описанием всех этапов работы, полученных данных и результатов анализа.
- Защита курсового проекта перед научным руководителем или комиссией

При оценке этапов КР учитываются следующие критерии:

I этап

выполнен во полном объеме – 15 баллов

выполнен с недочетами – 10 баллов

II этап (допускается, если I этап выполнен с ≥ 10 баллов)

выполнен в полном объеме – 15 баллов

выполнен с недочетами – 10 баллов

III (допускается, если II этап выполнен с ≥ 10 баллов)

этап выполнен в полном объеме – 25 баллов

выполнен с недочетами – 15 баллов

Максимальное количество баллов – 55

Для промежуточной аттестации:

Тест

Вопрос	Варианты ответа
Что включает в себя верификация наукометрических данных?	a. Обучение моделей машинного обучения
	b. Проверка, подтверждение и обеспечение достоверности данных
	c. Создание новых публикаций
	d. Работа с аппаратным обеспечением
Какой метод pandas используется для удаления дубликатов из DataFrame?	a. <code>delete_duplicates()</code>
	b. <code>drop_duplicates()</code>
	c. <code>clean_duplicates()</code>
	d. <code>remove_duplicates()</code>
Какой метод pandas используется для вычисления корреляции между числовыми столбцами в DataFrame?	a. <code>correlation()</code>
	b. <code>corr_matrix()</code>
	c. <code>correlation_matrix()</code>
	d. <code>corr()</code>
Верификация наукометрических данных включает в себя проверку этических аспектов?	a. Да
	b. Нет
Какой метод pandas используется для проверки полноты данных в DataFrame?	a. <code>check_completeness()</code>
	b. <code>validate_data()</code>
	c. <code>check_null()</code>
	a. <code>isnull()</code>
Визуализация данных с использованием matplotlib включает в себя создание трехмерных графиков?	a. Да
	b. Нет
Может ли BeautifulSoup использоваться для парсинга XML данных?	a. Да
	b. Нет
Сопоставьте методы BeautifulSoup с задачами парсинга данных, выполнив данные методы в Python	<code>find()</code>
	<code>find_all()</code>
	<code>select('.your_class')</code>
	<code>find('img').get('src')</code>
Как создать scatter plot в matplotlib для двух столбцов 'X' и 'Y'?	a. <code>scatter(df['X'], df['Y'])</code>
	b. <code>plot.scatter(df['X'], df['Y'])</code>
	c. <code>plot(df['X'], df['Y'], kind='scatter')</code>
	d. <code>scatter_plot(df['X'], df['Y'])</code>
Какой метод pandas используется для вычисления корреляции между числовыми столбцами в DataFrame?	a. <code>correlation()</code>
	b. <code>corr_matrix()</code>
	c. <code>correlation_matrix()</code>
	d. <code>corr()</code>

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика и содержит 50 тестовых вопросов на каждую компетенцию, из них 20% - закрытого типа, 80% - открытого типа.

Курсовой проект

Оценочные материалы

Сначала студент выступает с докладом и после этого комиссией задаются вопросы, которые зависят от нескольких нюансов:

1. Какая оценка выставлена студенту за дипломный проект рецензентом. Если получена «5» и также отлично прочитан доклад, тогда комиссия не будет слишком строга к обучающемуся.
 2. Насколько интересно прочитан доклад.
 3. Насколько быстро и уверенно отвечает студент. Здесь сразу видно, писал ли он отчет о проекте самостоятельно или нет.
 4. Как зарекомендовал себя обучающийся на протяжении всего учебного процесса.
 5. Как и по каким требованиям оформлен отчет о проекте.
- Задаются вопросы по теоретической и практической части. Количество вопросов, как правило, не более 5 шт. Сложность вопросов зависит от содержания выступления.

Пример вопросов по теории:

В чём заключается проект и какие могут быть результаты;

Что обучающийся проанализировал во время написания теоретической части;

Какая основная идея проекта;

Какие проблемы не удалось решить;

Какие перспективы могут быть;

Какие методы использованы, чтобы решить основную проблему;

Выбор и обоснование метода решения

Каково назначение и область применения проекта

Что даёт практическое использование;

Можно ли обойтись на практике без конкретных моделей, вариантов или методов;

Как и каким образом использовались программные средства.

Критерии оценки курсового проекта

1 Текст программы - 10 баллов

2 Блок-схемы моделей и алгоритмов - 7 баллов

3 Экранные формы - 6 баллов

4 Правильно оформленная документация - 7 баллов

5 Выступление и ответы на вопросы-10 баллов

Шкала оценивания:

Курсовой проект грамотно оформлен, теоретический раздел характеризуется логичным и последовательным изложением построения алгоритма, с соответствующими выводами о применяемых методах и обоснованными предложениями по структуризации программы, блок-схемы всех модулей оформлены понятно и наглядно представляют суть алгоритмов. Курсовой проект должен иметь положительный отзыв научного руководителя.

При ее защите студент четко отвечает на поставленные вопросы, грамотно обосновывает выбранный им способ решения задачи 45 баллов.

Курсовой проект в целом имеет положительный отзыв научного руководителя, но содержит ряд незначительных замечаний, как по теоретической части, так и по оформлению блок-схем и кодов программ. При ее защите студент без особых затруднений отвечает на поставленные вопросы, демонстрирует степень самостоятельности при её разработке – 30-34 баллов.

Код программы отлажен, представлены необходимые расчёты, однако и к теоретическому разделу отчета о курсовом проекте, и к алгоритмической структуре проекта имеются существенные замечания: текст характеризуется непоследовательностью в изложения материала, блок-схемы выполнены с нарушением требований, структуризация программы не полная. В отзыве научного руководителя имеются серьезные замечания по содержанию проекта и методах алгоритмизации. При ее защите студент не дает полного аргументированного ответа на заданные вопросы, проявляет неуверенность, демонстрирующую, чаще всего, несамостоятельность при выполнении проекта – 25-29 баллов.

В курсовом проекте поставленная задача не решена или решение и соответствующее оформление пояснительной записки не отвечает требованиям, изложенным в данных методических указаниях по выполнению курсового проекта. В отчете о курсовом проекте нет выводов, либо они носят декларативный характер. В отзыве научного руководителя имеются серьезные замечания. При защите курсового проекта студент затрудняется отвечать на поставленные, при ответе допускает существенные ошибки. Либо на защиту представлена чужая курсовой проект, написанный и уже защищенный в другом вузе или на другой кафедре 0-24 балла.

Максимальное количество баллов за защиту курсового проекта – 40.