



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

 Ю.В. Торкунова

«24» ноября 2021 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Инжиниринг данных

Направление подготовки	09.04.01 Информатика и вычислительная техника
Направленность (профиль)	Инженерия искусственного интеллекта
Квалификация	Магистр
Форма обучения	Очная

г. Казань, 2021

Перечень сведений о рабочей программе	Учетные данные
Образовательная программа Инженерия искусственного интеллекта	Код ОП 09.04.01
Направление подготовки Информатика и вычислительная техника	Код направления и уровня подготовки 09.04.01

Программа составлена автором:

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимир ович	Кандидат технических наук	Доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

Программа оформлена в соответствие с ПОЛОЖЕНИЕМ О ПОРЯДКЕ РАЗРАБОТКИ И УТВЕРЖДЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ – ПРОГРАММ БАКАЛАВРИАТА, ПРОГРАММ СПЕЦИАЛИТЕТА И ПРОГРАММ МАГИСТРАТУРЫ В КГЭУ

Рекомендовано учебно-методическим советом Института цифровых технологий и экономики ФГБОУ ВО «КГЭУ»

Протокол № 4 от 24.11.2021 г.

1. Цель, задачи и планируемые результаты обучения по дисциплине

Целью освоения дисциплины является формирование у студентов теоретических и практических знаний о методах сбора, извлечения и обработки данных, методах построения систем хранения данных, обучении технологии функционирования платформ распределенной обработки больших наборов данных.

Задачами дисциплины являются:

- формирование знаний и навыков работы с данными, представленными в различных форматах;
- обучение методам подготовки данных для систем машинного обучения;
- формирование знаний в области систем хранения данных.

Компетенции, формируемые у обучающихся, запланированные результаты обучения по дисциплине, соотнесенные с индикаторами достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине (знать, уметь, владеть)
ОПК-3. Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1. Анализирует профессиональную информацию, обосновывает выводы и рекомендации по решению профессиональных задач	ОПК-3.1. 3-1. Знает принципы, методы и средства получения и подготовки данных для анализа ОПК-3.1. 3-2. Знает платформы и форматы хранения данных ОПК-3.1. У-1. Умеет работать с данными в различных форматах
	ОПК-3.2. Составляет научные доклады, публикации, аналитические обзоры в сфере профессиональной деятельности	ОПК-3.2. 3-1. Знает методические подходы для подготовки аналитических материалов ОПК-3.2. У-1. Умеет использовать средства представления и визуализации данных ОПК-3.2. В-1. Владеет навыками подготовки докладов и публикаций в сфере профессиональной деятельности

2. Место дисциплины в структуре ОПОП

Дисциплина *Инжиниринг данных* относится к обязательной части учебного плана по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Код компетенции	Предшествующие дисциплины (модули), практики, НИР, др.	Последующие дисциплины (модули), практики, НИР, др.
ОПК-3		Спортивный анализ данных Производственная практика (преддипломная практика) Выполнение и защита выпускной квалификационной работы

Для освоения дисциплины обучающийся должен:

Знать:

- системы счисления и кодирование информации;
- системы управления базами данных;
- алгоритмизацию и программирование.

Уметь:

- использовать инструментальные среды разработки;
- реализовывать алгоритмы решения задач с помощью программирования;
- использовать средства представления и визуализации данных.

Владеть:

- языком программирования Python;
- языком манипулирования данными стандарта SQL.

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы (ЗЕ), всего 108 часов, из которых 26 часов составляет контактная работа обучающегося с преподавателем (занятия лекционного типа 8 часов, практические занятия 16 часов, контроль самостоятельной работы (КСР) - 2 часа), самостоятельная работа обучающегося 82 часа.

Вид учебной работы	Всего часов	Семестр
		1
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	108	108
КОНТАКТНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ С ПРЕПОДАВАТЕЛЕМ, в том числе:	26	26
Лекции (Лек)	8	8
Практические (семинарские) занятия (Пр)	16	16
Контроль самостоятельной работы и иная контактная работа (КСР)	2	2
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ(СРС)	82	82

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Семестр	Распределение трудоемкости (в часах) по видам учебной работы, включая СРС								Формируемые результаты обучения (знания, умения, навыки)	Литература	Формы текущего контроля успеваемости	Формы промежуточной аттестации	Максимальное количество баллов по балльно - рейтинговой системе
		Занятия лекционного типа	Занятия практического / семинарского типа	Лабораторные работы	Групповые консультации	Самостоятельная работа студента, в т.ч.	Контроль самостоятельной работы (КСР)	подготовка к промежуточной аттестации	Сдача зачета / экзамена	Итого				
Раздел 1. Работа с данными в Python	1	4	8			20				32	ОПК-3.1. 3-1 ОПК-3.1. 3-2 ОПК-3.1. У-1 Л2	ПЗ		40
Раздел 2. Подготовка данных для систем машинного обучения	1	2	8			20				30	ОПК-3.1. 3-1 ОПК-3.1. 3-2 ОПК-3.1. У-1 ОПК-3.2. 3-1 ОПК-3.2. У-1 Л2	ПЗ		40
Раздел 3. Параллельная и распределенная обработка данных	1	2				42	2			46	ОПК-3.1. 3-1 ОПК-3.1. 3-2 ОПК-3.1. У-1 ОПК-3.2. 3-1 ОПК-3.2. У-1 ОПК-3.2. В-1 Л1	ПЗ		20
ИТОГО		8	16			82	2			108				100

3.3. Тематический план лекционных занятий

№ п/п	Темы лекционных занятий	Трудоемкость, час.
1.	Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами. Общедоступные платформы для хранения данных. Подход Data-Centric AI. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний	4
2.	Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.	2
3.	Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, HDFS. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame.	2
Всего		8

3.4. Тематический план практических занятий

№ п/п	Темы практических работ	Трудоемкость, час.
1.	Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML	4
2.	Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами.	4
3.	Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей.	4
4.	Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.	4
Всего		16

3.5. Тематический план лабораторных работ

Данный вид работы не предусмотрен учебным планом.

3.6. Самостоятельная работа студента

Номер раздела дисциплины	Вид СРС	Содержание СРС	Трудоемкость, час.
1	Изучение теоретического материала, подготовка к выполнению практических заданий	Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний.	20
2	Изучение теоретического	Сбор данных и формирование набора данных для систем машинного обучения.	20

Номер раздела дисциплины	Вид СРС	Содержание СРС	Трудоемкость, час.
	материала, подготовка к выполнению практических заданий	Загрузка данных из интернет и социальных сетей. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных. Общедоступные платформы для хранения данных. Подход Data-Centric AI.	
3	Изучение теоретического материала, подготовка к выполнению практических заданий	Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, HDFS. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame.	42
Всего			82

4. Образовательные технологии

В процессе обучения используются:

- дистанционные курсы, размещенные на площадке LMS Moodle, URL: <http://lms.kgeu.ru/>;
- электронные образовательные ресурсы (ЭОР), размещенные в личных кабинетах студентов Электронного университета КГЭУ, URL: <http://e.kgeu.ru/>

5. Оценка результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля успеваемости, проводимого по балльно-рейтинговой системе (БРС), и промежуточной аттестации. Текущий контроль успеваемости осуществляется в течение семестра, включает выполнение практических заданий.

Обобщенные критерии и шкала оценивания уровня сформированности компетенции (индикатора достижения компетенции) по итогам освоения дисциплины:

Планируемые результаты обучения	Обобщенные критерии и шкала оценивания результатов обучения			
	неудовлетворительно	удовлетворительно	хорошо	отлично
	не зачтено	зачтено		
Полнота знаний	<i>Уровень знаний ниже минимальных требований, имеют место грубые ошибки</i>	<i>Минимально допустимый уровень знаний, имеет место много негрубых ошибок</i>	<i>Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок</i>	<i>Уровень знаний в объеме, соответствующем программе подготовки, без ошибок</i>
Наличие умений	<i>При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки</i>	<i>Продemonстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме</i>	<i>Продemonстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами</i>	<i>Продemonстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме</i>
Наличие навыков (владение опытом)	<i>При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки</i>	<i>Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами</i>	<i>Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами</i>	<i>Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов</i>
Характеристика сформированности компетенции (индикатора достижения компетенции)	<i>Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач</i>	<i>Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству практических задач</i>	<i>Сформированность компетенции в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных практических (профессиональных) задач</i>	<i>Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных практических (профессиональных) задач</i>
Уровень сформированности компетенции (индикатора достижения компетенции)	Низкий	Ниже среднего	Средний	Высокий

Шкала оценки результатов обучения по дисциплине:

Код индикатора достижения компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции (индикатора достижения компетенции)			
		Высокий	Средний	Ниже среднего	Низкий
		Шкала оценивания			
		отлично	хорошо	удовлетворительно	неудовлетворительно
		зачтено			не зачтено
Код компетенции: ОПК-3					
ОПК-3.1	знать:				
	принципы, методы и средства получения и подготовки данных для анализа	в полном объеме основные принципы, методы и средства получения и подготовки данных для анализа	основные принципы, методы и средства получения и подготовки данных для анализа, но имеются замечания	общие принципы, методы и средства получения и подготовки данных для анализа, но имеются замечания	уровень знаний ниже минимального требования, допускает грубые ошибки
	платформы и форматы хранения данных	все основные типы платформ хранения и обработки данных	основные типы платформ хранения и обработки данных	некоторые типы платформ хранения и обработки данных	уровень знаний ниже минимального требования, допускает грубые ошибки
	уметь:				
	работать с данными в различных форматах	самостоятельно выбирать и использовать инструменты для работать с данными в различных форматах	использовать стандартные инструменты для работать с данными в различных форматах	использовать стандартные инструменты для работать с данными в различных форматах, но допускает ошибки	не сформировано умение работать с данными в различных форматах

ОПК-3.2	знать:				
	методические подходы для подготовки аналитических материалов	знает принципы и методы подготовки аналитических материалов	знает основные принципы подготовки аналитических материалов	знает базовые принципы подготовки аналитических материалов	уровень знаний ниже минимального требования, допускает грубые ошибки
	уметь:				
	использовать средства представления и визуализации данных	самостоятельно выбирать и использовать инструменты представления и визуализации данных	использовать стандартные инструменты представления и визуализации данных	использовать стандартные инструменты представления и визуализации данных, но допускает ошибки	не сформировано умение использовать средства представления и визуализации данных
	владеть:				
	навыками подготовки докладов и публикаций в сфере профессиональной деятельности	демонстрирует уверенные умения подготовки докладов и публикаций в сфере профессиональной деятельности	демонстрирует основные умения подготовки докладов и публикаций в сфере профессиональной деятельности	демонстрирует базовые умения подготовки докладов и публикаций в сфере профессиональной деятельности	не владеет навыками подготовки докладов и публикаций в сфере профессиональной деятельности

Оценочные материалы для проведения текущего контроля успеваемости 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-разработчике в бумажном и электронном виде.*

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

Основная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
1	Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова	Реляционные базы данных и системы NoSQL	учебное пособие	Благовещенск : АмГУ	2018	https://e.lanbook.com/book/156492	
2	Харрисон, М.	Как устроен Python. Гид для разработчиков, программистов и интересующихся	справочное издание	СПб. : Питер	2019	https://ibooks.ru/reading.php?productid=359217	

Дополнительная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
1	Хахаев, И. А.	Практикум по алгоритмизации и программированию на Python	учебное пособие	М. : Национальный Открытый Университет "ИНТУИТ"	2016	https://e.lanbook.com/book/100377	
2	С. А. Мартишин, В. Л. Симонов, М. В. Храпченко	Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем	учебное пособие	Москва : ИД"ФОРУМ": ИНФРА-М	2019	https://ibooks.ru/reading.php?productid=361187	

6.2. Информационное обеспечение

6.2.1. 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	Энциклопедии, словари, справочники	http://www.rubricon.com
5	Портал "Открытое образование"	http://npod.ru
6	Единое окно доступа к образовательным ресурсам	http://window.edu.ru
7	Научная электронная библиотека	http://elibrary.ru
8	Онлайн-курс “Прикладное программирование на языке Python”.	https://openedu.ru/course/urfu/PYAP
9	Real Python Tutorials	https://realpython.com
10	Jake VanderPlas. Python Data Science Handbook	https://jakevdp.github.io/PythonDataScienceHandbook
11	Jeroen Janssens. Data Science at the Command Line	https://www.datascienceatthecommandline.com/
12	Andrew Ng. MLOps: From Model-centric to Data-centric AI	https://www.deeplearning.ai/wp-content/uploads/2021/06/MLOps-From-Model-centric-to-Data-centric-AI.pdf
13	Spark SQL, DataFrames and Datasets Guide	https://spark.apache.org/docs/latest/sql-programming-guide.html
14	Гугл Академия	https://scholar.google.ru/

6.2.2. Профессиональные базы данных

№ 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
4	Applied Science & Technology Source. EBSCO publishing	http://search.ebscohost.com	http://search.ebscohost.com
5	Wiley Online Library	http://onlinelibrary.wiley.com/	http://onlinelibrary.wiley.com/

6.2.3. Информационно-справочные системы

№ п/п	Наименование информационно-справочных систем	Адрес	Режим доступа
1	Научная электронная библиотека	http://elibrary.ru	http://elibrary.ru

			ru
2	Российская государственная библиотека	http://www.rsl.ru	http://www.rsl.ru
3	Международная реферативная база данных научных изданий zbMATH	http://www.zbmath.org	http://www.zbmath.org
4	Международная реферативная база данных научных изданий Springerlink	http:// link.springer.com	http:// link.springer.com
5	Образовательный портал	http://www.ucheba.com	http://www.ucheba.com

6.2.4. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Способ распространения (лицензионное/свободно)	Реквизиты подтверждающих документов
1	Windows 7 Профессиональная (Pro)	Пользовательская операционная система	№2011.25486 от 28.11.2011
2	Visual Studio Express	Инструмент создания Web приложений	https://visualstudio.microsoft.com/ru/vs/express/
3	Браузер Chrome	Система поиска информации в сети интернет (включая русскоязычный интернет).	https://www.google.com/intl/ru/chrome/
4	Браузер Firefox	Свободный веб-браузер	https://www.mozilla.org/ru/firefox/new/
5	OpenOffice	Пакет офисных приложений. Одним из первых стал поддерживать новый открытый формат OpenDocument. Официально поддерживается на платформах Linux	https://www.openoffice.org/ru/download/index.html
6	Adobe Acrobat	Пакет программ	https://get.adobe.com/ru/reader/
7	LMS Moodle	Это современное программное обеспечение	https://download.moodle.org/releases/latest/

7. Материально-техническое обеспечение дисциплины

№ п/п	Вид учебной работы	Наименование специальных помещений и помещений для СРС	Оснащенность специальных помещений и помещений для СРС
1	Лекционные занятия	Учебная аудитория для проведения занятий лекционного типа В-103	180 посадочных мест, доска аудиторная, акустическая система, проектор, усилитель-микшер для систем громкой связи, экран, микрофон, миникомпьютер, монитор, подключение к сети "Интернет", доступ в электронную информационно-образовательную среду

№ п/ п	Вид учебной работы	Наименование специальных помещений и помещений для СРС	Оснащенность специальных помещений и помещений для СРС
2	Практические занятия	Учебная лаборатория В-617	44 посадочных места (20 по центру - 24 по краю), доска ученическая, моноблок (10 шт.), подключение к сети «Интернет», доступ в электронную информационно-образовательную среду
		Лаборатория В-619	46 посадочных мест (24 по центру + 22 по краю), доска ученическая; моноблок (12 шт.), подключение к сети «Интернет», доступ в электронную информационно-образовательную среду
3	Самостоятельн ая работа обучающегося	Компьютерный класс с выходом в Интернет В-600а	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран), видеокамеры, программное обеспечение
		Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, мультимедийный проектор, экран, программное обеспечение

8. 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](http://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.



КГУУ

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

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
для проведения текущего контроля успеваемости и промежуточной
аттестации студентов по итогам освоения дисциплины

Инжиниринг данных

Направление подготовки 09.04.01 Информатика и вычислительная техника

Направленность (профиль) Инженерия искусственного интеллекта

Квалификация Магистр

Форма обучения Очная

Составлено автором:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Влаимирович	кандидат технических наук	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

Оценочные материалы оформлены в соответствии с ПОЛОЖЕНИЕМ О ПОРЯДКЕ
РАЗРАБОТКИ И УТВЕРЖДЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ – ПРОГРАММ
БАКАЛАВРИАТА, ПРОГРАММ СПЕЦИАЛИТЕТА И ПРОГРАММ МАГИСТРАТУРЫ В
КГУУ

1. Цель и задачи текущего контроля и промежуточной аттестации студентов по дисциплине «Инжиниринг данных»

Цель текущего контроля - систематическая проверка степени освоения программы дисциплины «Инжиниринг данных», уровня сформированности знаний, умений, навыков, компетенций на текущих занятиях

Задачи текущего контроля:

1. определение индивидуального учебного рейтинга студентов;
2. своевременное выполнение корректирующих действий по содержанию и организации процесса обучения; обнаружение и устранение пробелов в усвоении учебной дисциплины;
3. подготовки к промежуточной аттестации.

В течение семестра при изучении дисциплины реализуется комплексная система поэтапного оценивания уровня освоения – балльно-рейтинговая система. За каждый вид учебных действий студенты получают определенное количество баллов. В течение семестра студент может набрать до 60-ти баллов.

Цель промежуточной аттестации - проверка степени усвоения студентами учебного материала за время изучения дисциплины, уровня сформированности компетенций после завершения изучения дисциплины. Аттестация проходит в форме зачета.

Задачи промежуточной аттестации:

1. определение уровня усвоения учебной дисциплины;
2. определение уровня сформированности компетенций.

2. Основное содержание текущего контроля и промежуточной аттестации студентов

В результате изучения дисциплины «Инжиниринг данных» формируются следующие компетенции или их составляющие:

2.1. Основное содержание текущего контроля

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Содержание оценочных заданий для выявления сформированности компетенций у студентов по завершении модуля/освоения дисциплины		
		Базовый уровень	Продвинутый уровень	Высокий уровень
1	2	3	4	5
ОПК-3	ОПК-3.1. Анализирует профессиональную информацию, обосновывает выводы и рекомендации по решению профессиональных задач ОПК-3.2. Составляет научные доклады, публикации, аналитические обзоры в сфере профессиональной деятельности	Решение практического задания	Решение практического задания, написание отчета	Решение практического задания, написание отчета, ответы на вопросы

2.2. Основное содержание промежуточной аттестации студентов

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Содержание оценочных заданий для выявления сформированности компетенций у студентов по завершении модуля/освоения дисциплины		
		Базовый уровень	Продвинутый уровень	Высокий уровень
1	2	3	4	5
ОПК-3	ОПК-3.1. Анализирует профессиональную информацию, обосновывает выводы и рекомендации по решению профессиональных задач ОПК-3.2. Составляет научные доклады, публикаций, аналитические обзоры в сфере профессиональной деятельности	Решение кейса	Решение кейса	Решение кейса

3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

п/п	Наименование компоненты	Критерии оценки	Максимальное число баллов
1	Практические занятия - 4 занятия	выполнение задания 8 баллов оформление выводов 4 балла ответы на вопросы 3 балла	15 баллов за каждое занятие 60 баллов
2	Решение кейса	определен способ преобразования данных в соответствии с решаемой задачей 5 баллов исходные данные преобразованы в набор данных, пригодных для анализа данных 10 баллов проведена разметка и очистка данных 10 баллов проведена подготовка данных 15 баллов	40 баллов
		ИТОГО	100 баллов

Содержание практических заданий

Практическое задание 1. Библиотеки для работы с данными

Написать код, снабжая его комментариями.

1. Открыть набор данных о футболистах data_sf.csv с помощью функции read_csv модуля pandas и применить к нему метод .info(). Ответить на вопросы:

1. Сколько всего колонок в наборе данных о футболистах?
2. Сколько непустых элементов в колонке Club?
3. Какой тип данных у колонки Wage?
4. Какой тип данных у колонки Position?
5. Сколько колонок имеют тип object?
6. Какая колонка будет последней в наборе данных?

7. Есть ли в наборе данных колонка с названием 'Surname'?
2. Конвертировать набор данных об абитуриентах `br11er25_ab.json`. Ответить на вопросы:
 1. Сколько всего колонок в наборе данных?
 3. Какой тип данных у колонки `Numbre`?
 4. Какой тип данных у колонки `Home`?
 5. Сколько колонок имеют типы `object`?
 6. Какая колонка будет последней в наборе данных?
3. Открыть файл о калорийности продуктов `kaloriy.html`. Ответить на вопросы:
 1. Сколько всего колонок в наборе данных?
 3. Какой тип данных у колонки `Product`?
 4. Какой тип данных у колонки `Day`?
 5. Сколько колонок имеют типы `object`?
 6. Какая колонка будет последней в наборе данных?

Практическое задание 2. Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами

Написать код, снабжая его комментариями.

1. Подключиться к БД `dbRtf`. Выполнить запрос к таблице `Price` и получить все данные
- Ответить на вопросы:

1. Сколько всего колонок в наборе данных?
 2. Сколько непустых элементов в колонке `Old`?
 3. Какой тип данных у колонки `Data`?
 4. Какое последнее значение в третьей колонке?
2. Открыть файл, содержащий наборы изображений. Изображения имеют размерность (3000, 3000, 1). Значение пикселей лежат в диапазоне 0...255. Просмотреть первые 10 изображений. Преобразовать в набор данных
 3. Открыть файл, содержащий наборы изображений. Изображения имеют размерность (300, 300, 3). Значение пикселей лежат в диапазоне 0...255. Просмотреть первые 10 изображений. Преобразовать в набор данных
 4. Повторить аналогичные действия с получением набора данных из звуковых файлов.

Практическое задание 3. Сбор данных и формирование набора данных

Написать код, снабжая его комментариями.

1. Открыть набор данных о футболистах `data_sf.csv` с помощью функции `read_csv` модуля `pandas`.
2. Для ответа на дальнейшие задачи нужно применить к набору данных о футболистах метод `.describe()` (возможно, с какими-то параметрами – параметры изучить самостоятельно).
 1. Каково среднее значение возраста футболистов в наборе данных?
 2. Каков минимальный возраст футболиста?
 3. Какова максимальная заработная плата за год (`Wage`) у футболистов?
 4. Какова медианная заработная плата за год (`Wage`) у футболистов?
 5. Каково минимальное значение параметра `'Penalties'`?
 6. Какое значение у первого (25%) квартиля параметра `ShortPassing`?
 7. Какова самая частая национальность (`Nationality`) футболистов?
 8. Сколько разных клубов в наборе данных о футболистах?
 9. Сколько раз встречается самая частая позиция `'GK'` в наборе данных?

10. Каков максимальный возраст футболиста?
3. Для ответа на дальнейшие задачи нужно применить статистические методы
1. Чему равен средний возраст (Age), футболистов в наборе данных, округлённый до целого?
 2. Каково количество непустых строк в колонке Composure (Хладнокровие) набора данных о футболистах?
 3. Каково в наборе данных о футболистах стандартное отклонение параметра коротких пасов (ShortPassing), округлённое до второго знака после запятой?
 4. Какова сумма заработных плат за год (Wage) в наборе данных о футболистах?
 5. Какова минимальная стоимость футболиста (Value) в наборе данных о футболистах?
4. Для ответа на дальнейшие задачи нужно применить методы фильтрации
1. Какова средняя скорость (SprintSpeed) футболистов, зарплата (Wage) которых выше среднего?
 2. Какова средняя скорость (SprintSpeed) футболистов, зарплата (Wage) которых ниже среднего?
 3. Какую позицию (Position) занимает футболист с самой высокой зарплатой (Wage)?
 4. Сколько пенальти (Penalties) забили бразильские (Nationality, Brazil) футболисты за период, данные о котором представлены в датасете?
 5. Укажите средний возраст (Age) игроков, у которых точность удара головой (HeadingAccuracy) > 50.
 6. Укажите возраст (Age) самого молодого игрока, у которого хладнокровие (Composure) и реакция (Reactions) превышают 90% от максимального значения, представленного в датасете.
 7. Определите, насколько средняя реакция (Reactions) самых взрослых игроков (т.е. игроков, чей возраст (Age) равен максимальному) больше средней реакции самых молодых игроков.
 8. Из какой страны (Nationality) происходит больше всего игроков, чья стоимость (Value) превышает среднее значение?
 9. Определите, во сколько раз средняя зарплата (Wage) голкипера (Position, GK) с максимальным значением показателя "Рефлексy" (GKReflexes) выше средней зарплатy голкипера с максимальным значением показателя "Владение мячом" (GKHandling).
 10. Определите, во сколько раз средняя сила удара (ShotPower) самых агрессивных игроков (игроков с максимальным значением показателя "Агрессивность" (Aggression)) выше средней силы удара игроков с минимальной агрессией. Ответ округлите до сотых.

Практическое задание 4. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.

Написать код, снабжая его комментариями.

1. Открыть файл log.csv. Ознакомиться с данными.
2. Посмотрите все уникальные значения в sample.csv
3. Посмотрите, сколько непустых значений в колонке City
4. Файл log.csv, в колонке user_id есть записи, которые содержат технические ошибки. Укажите, что записано в поле user_id в строчках с ошибкой
5. Создайте новый датафрейм sample2, в который будут входить только записи о людях в возрасте меньше 30 лет

6. Создайте новый датафрейм `log_win`, в который будут входить только записи, где пользователь выиграл. Посчитайте, сколько таких записей, и сохраните в переменной `win_count`
7. Создайте новый датафрейм `sample2`, в который будут входить только записи о рабочих младше 30 лет
8. Найдите записи, где в городах нет буквы "о", и сохраните в переменную `sample4`. Не забудьте про NaN и параметр `na`
9. Сохраните в переменную `new_log` датафрейм, из которого удалены записи с ошибкой в поле `user_id`
10. С помощью `apply` и лямбда-функции увеличьте возраст во всех записях на 1 год и сохраните в `sample2`. В переменной `sample2` должен содержаться весь датафрейм `sample`
11. С помощью `apply` и `lambda`-функции замените все буквы в поле `City` на маленькие и сохраните в `sample2`. Вам может понадобиться функция `s.lower()`
12. Напишите функцию `profession_code`, которая на вход получает строку, а на выход возвращает:
 - 0 — если на вход поступила строка "Рабочий"
 - 1 — если на вход поступила строка "Менеджер"
 - 2 — в любом другом случае
13. Напишите функцию `age_category`, которая на вход получает число, а на выход отдаёт:
 - "молодой" — если возраст меньше 23
 - "средний" — если возраст от 23 до 35
 - "зрелый" — если возраст больше 35

Вопросы для подготовки к зачету по курсу «Инжиниринг данных»

1. Библиотека `pandas` в Python.
2. Работа с данными в формате CSV в Python.
3. Работа с данными в формате JSON в Python.
4. Работа с данными в формате HTML в Python.
5. Работа с изображениями в Python.
6. Работа с видео в Python.
7. Работа с аудио в Python.
8. Работа с Parquet в Python.
9. Работа с графами знаний в Python.
10. Этапы и инструменты создания наборов данных для машинного обучения.
11. Загрузка данных с Web-сайтов.
12. Загрузка данных из социальных сетей.
13. Методы и инструменты подготовки данных.
14. Методы и инструменты очистки данных.
15. Разметка данных.
16. Общедоступные платформы для хранения данных.
17. Архитектура центров обработки данных.
18. Кластеры для параллельных и распределенных вычислений.
19. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop.
20. Распределенная файловая система HDFS.
21. Распределенная обработка данных в Apache Spark.
22. Работа с данными с использованием Apache Spark DataFrame.
23. Источники данных для Apache Spark DataFrame.
24. Обработка данных в Apache Spark DataFrame.
25. Использование SQL в Apache Spark DataFrame.

Пример типового кейса для проведения промежуточной аттестации

1. Исходные данные представлены в файле Excel. Открыть файл и просмотреть структуру данных.
2. Сформировать набор данных. Набор данных должен быть представлен в виде датафрейма. Имена столбцов не должны содержать пробелов. Типы данных должны соответствовать значениям.
3. Проверить данные на наличие ошибок и пустых значений. Просмотреть ошибки. Преобразовать, либо удалить данные в зависимости от ошибок и пропусков. Подумайте, что следует сделать, если данные чрезвычайно разреженные по определенным признакам.
4. Выявить выбросы и удалить соответствующие записи.
5. Построить два частотных распределения, два графика и две диаграммы box-whiskies. По каждому рисунку сделать выводы.