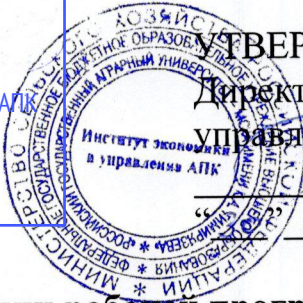


Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Хоружий Людмила Ивановна  
Должность: Директор института экономики и управления АПК  
Дата подписания: 21.03.2024 11:55:40  
Уникальный программный ключ:  
1e90b132d9b04dce67585160b015dddf2cb1e6a9



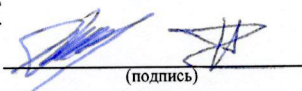
**УТВЕРЖДАЮ:**  
Директор Института экономики и  
управления АПК  
Л.И. Хоружий  
2023 г.

**Лист актуализации рабочей программы дисциплины  
«Б1.В.19 Инжиниринг данных»**

Направление: 09.03.02 «Информационные системы и технологии»  
Направленность: Компьютерные науки и интеллектуальный анализ данных  
(Computer Science & Data Mining)  
Форма обучения заочная  
Год начала подготовки: 2022  
Курс 3,4  
Семестр 6,7


- 1. В рабочую программу не вносятся изменения. Программа актуализирована для 2023 г. начала подготовки.
- 2. Программа будет распространена при организации учебного процесса на направленность (профиль): Компьютерные науки и интеллектуальный анализ данных.

Разработчик (и): Демичев В.В., канд. экон. наук, доцент  
Невзоров А.С., ассистент  
(ФИО, ученая степень, ученое звание)


  
(подпись)

«28» августа 2023 г.

Рабочая программа пересмотрена и одобрена на заседании кафедры статистики и кибернетики протокол № 11 от «28» августа 2023 г.

И.о. заведующего кафедрой статистики и кибернетики  А.В. Уколова

**Лист актуализации принят на хранение:**

И.о. заведующего кафедрой статистики и кибернетики  А.В. Уколова



**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ**  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –**  
**МСХА имени К.А. ТИМИРЯЗЕВА»**  
(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)

Институт экономики и управления АПК  
Кафедра статистики и кибернетики

УТВЕРЖДАЮ  
Директор института экономики  
и управления АПК  
Л.И. Хоружий  
«*Л.И. Хоружий*» 2022 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Б1.В.19 Инжиниринг данных**

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»  
Направленность: Компьютерные науки и интеллектуальный анализ данных  
(Computer Science & Data Mining)

Курс 3,4  
Семестр 6, 7

Форма обучения заочная


Год начала подготовки 2022

Москва, 2022

Разработчик (и): Харитонов А.Е., к.э.н.  
(ФИО, ученая степень, ученое звание)

Ульянкин А.Е., ассистент

(ФИО, ученая степень, ученое звание)

 «26» 08 2022 г.

«26» 08 2022 г.

Рецензент: Коломеева Е.С., к.э.н., доцент

(ФИО, ученая степень, ученое звание)


  
(подпись)  
«26» 08 2022 г.

Программа составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.02 «Информационные системы и технологии» и учебного плана

Программа обсуждена на заседании кафедры статистики и кибернетики протокол № 11 от 26 августа 2022 г.

И.о. зав. кафедрой Уколова А.В., к.э.н., доцент


(ФИО, ученая степень, ученое звание)

  
(подпись)  
«26» 08 2022 г.

**Согласовано:**

Председатель учебно-методической комиссии института экономики и управления АПК  
Корольков А.Ф., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
(подпись)  
«26» 08 2022 г.

И.о. заведующего выпускающей кафедрой  
статистики и кибернетики

Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
(подпись)  
«26» 08 2022 г.

Заведующий отделом комплектования ЦНБ

 Ермилова Я.В.  
(подпись)

## СОДЕРЖАНИЕ

<b>АННОТАЦИЯ .....</b>	<b>4</b>
<b>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....</b>	<b>5</b>
<b>2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ .....</b>	<b>5</b>
<b>3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....</b>	<b>5</b>
<b>4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....</b>	<b>5</b>
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ .....	5
ПО СЕМЕСТРАМ .....	5
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	9
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ .....	10
<b>5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....</b>	<b>11</b>
<b>6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>11</b>
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ.....	11
<b>7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....</b>	<b>15</b>
7.1    ОСНОВНАЯ ЛИТЕРАТУРА.....	15
7.2    ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	15
<b>8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>16</b>
<b>9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ.....</b>	<b>16</b>
<b>10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....</b>	<b>16</b>
<b>11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....</b>	<b>17</b>
<b>12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ .....</b>	<b>18</b>

## Аннотация

**рабочей программы по дисциплине Б1.В.19 «Инжиниринг данных» для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»**

**Цель освоения дисциплины:** Цель освоения дисциплины - формирование у студентов необходимой теоретической базы и практических навыков подготовки данных для моделей машинного обучения. В процессе обучения рассматриваются особенности работы с данными в различных форматах на языке Python. Подробно изучаются методы очистки данных и соответствующие библиотеки на Python.

**Место дисциплины в учебном плане:** дисциплина включена в часть, формируемую участниками образовательных отношений учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

**Требования к результатам освоения дисциплины:** в результате освоения дисциплины формируются следующие компетенции (индикаторы): ПКос-3 (ПКос-3.2; ПКос-3.3); ПКос-7 (ПКос-7.2; ПКос-7.3); ПКос-9 (ПКос-9.2; ПКос-9.3).

**Краткое содержание дисциплины:**

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.

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. Общедоступные платформы для хранения данных. Подход Data-Centric AI.

Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame.

**Общая трудоемкость дисциплины составляет:** 3 зачетные единицы (108 часов, в том числе практической подготовки 2 часа).

**Промежуточный контроль:** экзамен.

## **1. Цель освоения дисциплины**

Цель освоения дисциплины - формирование у студентов необходимой теоретической базы и практических навыков подготовки данных для моделей машинного обучения. В процессе обучения рассматриваются особенности работы с данными в различных форматах на языке Python. Подробно изучаются методы очистки данных и соответствующие библиотеки на Python.

## **2. Место дисциплины в учебном процессе**

Дисциплина «Инжиниринг данных» включена в часть, формируемую участниками образовательных отношений учебного плана. Дисциплина «Инжиниринг данных» реализуется в соответствии с требованиями ФГОС ВО, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Дисциплина «Инжиниринг данных» является основополагающей для изучения следующих дисциплин: «Большие данные», «Разработка программного продукта машинного обучения», «Python для искусственного интеллекта», «Разведочный анализ данных», «Многомерные статистические методы» и др.

Рабочая программа дисциплины «Инжиниринг данных» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

## **3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы**

Изучение данной учебной дисциплины направлено на формирование у обучающихся компетенций, представленных в таблице 1.

## **4. Структура и содержание дисциплины**

### **4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам**

Общая трудоёмкость дисциплины составляет 3 зач.ед. (108 часов, в том числе практической подготовки 2 часа), их распределение по видам работ семестрам представлено в таблице 2.

Таблица 1

## Требования к результатам освоения учебной дисциплины

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
1	ПКос-3	Способность выполнять работы по обеспечению функционирования баз данных и обеспечению их информационной безопасности	ПКос-3.2 Уметь осуществлять защиту базы данных, оперативное резервирование базы данных	-	осуществлять защиту базы данных, оперативное резервирование базы данных	-
			ПКос-3.3 Владеть навыками сопровождения базы данных в оперативном и обеспечивающем режимах	-	-	навыками сопровождения базы данных в оперативном и обеспечивающем режимах
2.	ПКос-7	Способность использовать компоненты системных программных продуктов: компиляторы, загрузчики, сборщики и системные утилиты	ПКос-7.2 Уметь использовать компоненты системных программных продуктов для решения профессиональных задач	-	использовать компоненты системных программных продуктов для решения профессиональных задач	-
			ПКос-7.3 Владеть навыками выбора и конфигурирования компонентов системных программных продуктов	-	-	навыками выбора и конфигурирования компонентов системных программных продуктов
3.	ПКос-9	Способен осуществлять разработку, отладку и рефакторинг кода программного обеспечения, интеграцию программных модулей и компонент, в том числе	ПКос-9.2 Уметь использовать методы и приемы формализации и алгоритмизации поставленных задач; использовать программные продукты для графического отображения алгоритмов; применять стандартные алгоритмы в со-	-	использовать методы и приемы формализации и алгоритмизации поставленных задач; использовать программные продукты для графического отображения алгоритмов; применять стан-	-

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
		взаимодействующих с внешней средой, средствами выбранных языков программирования	ответственных областях; применять выбранные языки программирования для написания программного кода; использовать выбранную среду программирования; применять инструментарий для создания и актуализации исходных текстов программ; выявлять ошибки в программном коде, интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; применять методы и приемы отладки программного кода		дартные алгоритмы в соответствующих областях; применять выбранные языки программирования для написания программного кода; использовать выбранную среду программирования; применять инструментарий для создания и актуализации исходных текстов программ; выявлять ошибки в программном коде, интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; применять методы и приемы отладки программного кода	
			ПКос-9.3 Владеть навыками составления формализованных описаний решений поставленных задач в соответствии с требованиями технического задания; разработки алгоритмов решения поставленных задач в соответствии с требованиями технического задания или других принятых	-	-	навыками составления формализованных описаний решений поставленных задач в соответствии с требованиями технического задания; разработки алгоритмов решения поставленных задач в соответствии с требо-



№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
			в организации нормативных документов; создания программного кода в соответствии с техническим заданием (готовыми спецификациями); оптимизации программного кода с использованием специализированных программных средств; анализа и проверки исходного программного кода; отладки программного кода на уровне программных модулей и межмодульных взаимодействий и взаимодействий с окружением			ваниями технического задания или других принятых в организации нормативных документов; создания программного кода в соответствии с техническим заданием (готовыми спецификациями); оптимизации программного кода с использованием специализированных программных средств; анализа и проверки исходного программного кода; отладки программного кода на уровне программных модулей и межмодульных взаимодействий и взаимодействий с окружением

Таблица 2

## Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость		
	час. всего/*	В т.ч. по семестрам	
		№ 6	№7
<b>Общая трудоёмкость</b> дисциплины по учебному плану	<b>108/2</b>	<b>36</b>	<b>72/2</b>
<b>1. Контактная работа:</b>	<b>14,4/2</b>	<b>2</b>	<b>12,4/2</b>
<b>Аудиторная работа</b>	<b>14,4/2</b>	<b>2</b>	<b>12,4/2</b>
<i>практические занятия (ПЗ)</i>	8/2		8/2
<i>Лекции (Л)</i>	6	2	4
<i>контактная работа на промежуточном контроле (КРА)</i>	0,4		0,4
<b>2. Самостоятельная работа (СРС)</b>	<b>93,6</b>	<b>34</b>	<b>59,6</b>
<i>самостоятельное изучение разделов, самоподготовка (про- работка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)</i>	85	34	51
<i>Подготовка к экзамену</i>	8,6		8,6
Вид промежуточного контроля:	Экзамен		

## 4.2 Содержание дисциплины

Таблица 3

## Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа			Внеауди- торная работа СР
		Л	ПЗ/*	ПКР	
Раздел 1. Работа с данными в Python	40/2	2	4/2		32
Раздел 2. Подготовка данных для систем машинного обучения	35	2	2		31
Раздел 3. Параллельная и распределенная обработка данных	34,6	2	2		30,6
Контактная работа на промежуточном контроле (КРА)	0,4			0,4	
<b>Итого по дисциплине</b>	<b>108/4</b>	<b>6</b>	<b>8/2</b>	<b>0,4</b>	<b>93,6</b>

**Раздел 1. Работа с данными в Python**

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.

**Раздел 2. Подготовка данных для систем машинного обучения**

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. Общедоступные платформы для хранения данных. Подход Data-Centric AI.

### Раздел 3. Параллельная и распределенная обработка данных

Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame.

## 4.3 Практические занятия

Таблица 4

### Содержание практических занятий и контрольные мероприятия

№ п/п	№ раздела	№ и название практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов
1	Раздел 3. Параллельная и распределенная обработка данных	Практическая работа № 1. Работа с данными в Apache Spark.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита работы	4
		Практическая работа № 2. Использование SQL в Apache Spark.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита работы	4

Таблица 5

### Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Раздел 1. Работа с данными в Python	Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Форматы хранения больших данных и работа с ними: Parquet, Avro. ПКос-3 (ПКос-3.2; ПКос-3.3); ПКос-7 (ПКос-7.2; ПКос-7.3); ПКос-9 (ПКос-9.2; ПКос-9.3).
2.	Раздел 2. Подготовка данных для систем машинного обучения	Разметка данных. Общедоступные платформы для хранения данных. Подход Data-Centric AI. ПКос-3 (ПКос-3.2; ПКос-3.3); ПКос-7 (ПКос-7.2; ПКос-7.3); ПКос-9 (ПКос-9.2; ПКос-9.3).
3	Раздел 3. Параллельная и распределенная обработка данных	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. ПКос-3 (ПКос-3.2; ПКос-3.3); ПКос-7 (ПКос-7.2; ПКос-7.3); ПКос-9 (ПКос-9.2; ПКос-9.3).

**Перечень задач для самостоятельного выполнения (контрольные работы)**

№ п/п	Название раздела, темы	№ и название лекций/практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия
1	Раздел 1. Работа с данными в Python	Контрольная работа № 1. Библиотеки для работы с данными в Python: numpy, pandas.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита контрольной работы
		Контрольная работа № 2. Работа с текстовыми файлами разных форматов в Python: CSV, JSON, HTML.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита контрольной работы
		Контрольная работа № 3. Работа с базами данных в Python.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита контрольной работы
2	Раздел 2. Подготовка данных для систем машинного обучения	Контрольная работа № 4. Работа с файлами для хранения больших данных в Python.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита контрольной работы
		Контрольная работа № 5. Создание собственных наборов данных в Python. Очистка и подготовка данных.	ПКос-3.2 ПКос-3.3 ПКос-7.2 ПКос-7.3 ПКос-9.2 ПКос-9.3	Защита контрольной работы

**5. Образовательные технологии****Применение активных и интерактивных образовательных технологий**

№ п/п	Тема и форма занятия		Наименование используемых активных и интерактивных образовательных технологий
1	Раздел 1. Работа с данными в Python	Л	Лекция с ошибками
2	Раздел 2. Подготовка данных для систем машинного обучения	Л	Лекция с ошибками
3	Раздел 3. Параллельная и распределенная обработка данных	Л	Лекция с ошибками

**6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины**

**6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности**

## Практические работы

### **Контрольная работа № 1. Библиотеки для работы с данными в Python: numpy, pandas.**

**Цель:** освоение методики загрузки данных в среду разработки.

**Требуется:** загрузить таблицу данных в формате Excel из исходного файла разными способами, оценить вариацию показателей, путем расчета советующих коэффициентов, построить интервальные ряды распределения.

### **Контрольная работа № 2. Работа с текстовыми файлами разных форматов в Python: CSV, JSON, HTML.**

**Цель:** освоение вариантов обработки текстовых файлов.

**Требуется:** загрузить в среду разработки данные из исходного файла в различных форматах: CSV, JSON, HTML, оценить вариацию показателей, путем расчета советующих коэффициентов, построить интервальные ряды распределения.

### **Контрольная работа № 3. Работа с базами данных в Python.**

**Цель:** освоение основ работы с базами данных в Python.

**Требуется:** создать базу данных, внести в нее исходные данные.

### **Контрольная работа № 4. Работа с файлами для хранения больших данных в Python.**

**Цель:** освоение принципов загрузки и анализа больших данных.

**Требуется:** оценить файл с исходными данными, подобрать способ загрузки в среду разработки, оценить скорость и качество загрузки файла разными способами, провести точечную и интервальную оценку генеральной средней по выборке из исходных данных.

### **Контрольная работа № 5. Создание собственных наборов данных в Python. Очистка и подготовка данных.**

**Цель:** освоение критериев качества наборов данных и способов их подготовки.

**Требуется:** составить собственный набор данных, используя различные открытые источники данных (сайт Росстата, ЕМИСС и др), провести очистку данных (удалить дублирования, резко выделяющиеся значения, пустые строки и столбцы и т.д.)

### **Контрольная работа № 6. Работа с данными в Apache Spark.**

**Цель:** освоение навыков работы с Apache Spark.

**Требуется:** загрузить среду, создать базу данных, внести в нее информацию.

### **Контрольная работа № 7. Использование SQL в Apache Spark.**

**Цель:** освоение навыков использования SQL в Apache Spark.

**Требуется:** загрузить базу данных, прописать код для выполнения запросов для извлечения конкретных данных.

## **Вопросы к защите практических работ**

### **Раздел 1. Работа с данными в Python**

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.

### **Раздел 2. Подготовка данных для систем машинного обучения**

1. Методы и инструменты подготовки данных.
2. Методы и инструменты очистки данных.
3. Разметка данных.
4. Общедоступные платформы для хранения данных.
5. Архитектура центров обработки данных.
6. Кластеры для параллельных и распределенных вычислений.

### **Раздел 3. Параллельная и распределенная обработка данных**

1. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop.
2. Распределенная файловая система HDFS.
3. Распределенная обработка данных в Apache Spark.
4. Работа с данными с использованием Apache Spark DataFrame.
5. Источники данных для Apache Spark DataFrame.
6. Обработка данных в Apache Spark DataFrame.
7. Использование SQL в Apache Spark DataFrame.

### **Примерный перечень вопросов к экзамену**

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.

## 6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания

Текущий контроль знаний, умений и навыков проводится в форме контрольной работа с расчетными задачами, а также предусмотрена защита индивидуальных задач, выполняемых каждым студентом на практических занятиях. Ликвидация студентами текущих задолженностей производится также в форме выполнения индивидуальной задачи по соответствующей теме и дальнейшей ее защиты преподавателю кафедры.

Максимальная оценка за выполнение каждой работы составляет 10 баллов. Задержка выполнения индивидуального практического задания на одну неделю штрафуется одним баллом. Критерии оценки выполненных работ приведены в таблице:

Оценка	Критерии оценивания
«10» (отлично)	оценку <b>«отлично»</b> заслуживает студент, выполнивший практическую работу полностью, в работе корректно применены статистические методы, нет ошибок в расчетах, сделаны глубокие выводы. Студент дал полные ответы на все заданные вопросы по работе. Недостатков по оформлению работы не имеется.
«8-9» (хорошо)	оценку <b>«хорошо»</b> заслуживает студент, выполнивший практическую работу полностью, имеются недочеты в применении статистических методов, проведенном анализе и полученных выводах. Студент дал верные ответы на все заданные вопросы по работе. Недостатков по оформлению работы не имеется.
«6-7» (удовлетворительно)	оценку <b>«удовлетворительно»</b> заслуживает студент, выполнивший практическую работу полностью, но имеются существенные недочеты в применении отдельных статистических методов и полученных выводах. Студент дал верные ответы не на все вопросы. По оформлению работы имеются недостатки.
«1-5» (неудовлетворительно)	оценку <b>«неудовлетворительно»</b> заслуживает студент, выполнивший практическую работу не по своему варианту или с грубейшими нарушениями применения методов и последовательности анализа.

Итоговая оценка учитывает результаты рейтинговой системы контроля знаний (вклад 50%), результаты экзамена (вклад 50%). Баллы, заработанные студентами за выполнение лекций с ошибками, могут дополнительно учитываться в рейтинге. Критерии выставления оценок по системе:

- 0-49 % от максимального количества баллов – «неудовлетворительно»;
- 50 -69 % – «удовлетворительно»;
- 70-84 % – «хорошо»;
- свыше 85% – «отлично».

## **7. Учебно-методическое и информационное обеспечение дисциплины**

### **7.1 Основная литература**

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100> (дата обращения: 24.09.2022).
2. Маркин, А. В. Программирование на SQL в 2 ч. Часть 1 : учебник и практикум для вузов / А. В. Маркин. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 403 с. — (Высшее образование). — ISBN 978-5-534-12256-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/491238> (дата обращения: 24.09.2022).
3. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2 : учебник и практикум для вузов / А. В. Маркин. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 340 с. — (Высшее образование). — ISBN 978-5-534-12258-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490104> (дата обращения: 24.09.2022).
4. Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — Москва : Издательство Юрайт, 2022. — 286 с. — (Высшее образование). — ISBN 978-5-534-14350-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/496893> (дата обращения: 24.09.2022).

### **7.2 Дополнительная литература**

1. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2022. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490020> (дата обращения: 27.12.2022).



2. Воронов, М. В. Системы искусственного интеллекта : учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. — Москва : Издательство Юрайт, 2022. — 256 с. — (Высшее образование). — ISBN 978-5-534-14916-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/485440> (дата обращения: 27.12.2022).
3. Нестеров, С. А. Базы данных : учебник и практикум для среднего профессионального образования / С. А. Нестеров. — Москва : Издательство Юрайт, 2022. — 230 с. — (Профессиональное образование). — ISBN 978-5-534-11629-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/495981> (дата обращения: 24.09.2022).

## **8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

1. Язык программирования Python. URL: <https://www.python.org/> (открытый доступ)
2. Анаконда. URL: <https://www.anaconda.com/distribution/> (открытый доступ)
3. SQL Online IDE: <https://sqliteonline.com/> (открытый доступ)
4. Официальный сайт Росстата. URL: <http://www.gks.ru/> (открытый доступ)
5. Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru> (открытый доступ)

## **9. Перечень программного обеспечения**

Таблица 9

### **Перечень программного обеспечения**

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Раздел 1. Работа с данными в Python	Anaconda, MS Office	расчётная	Anaconda Enterprise, Microsoft	2019, 2016
2	Раздел 2. Подготовка данных для систем машинного обучения				
3	Раздел 3. Параллельная и распределенная обработка данных				

## **10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Таблица 10

### **Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями**

Наименование специальных помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы
1	2

<p><i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (2й учебный корпус, 102 ауд.)</i></p>	<ol style="list-style-type: none"> <li>1. Экран с электроприводом 1 шт. (Инв. №558771/2)</li> <li>2. Просектор 1 шт. (без инв. №) – приобретался не за счет средств вуза</li> <li>3. Вандалоустойчивый шкаф 1 шт. (Инв.№558850/7)</li> <li>4. Системный блок iP-4 541 3200 MHz/1024 Mb/ 80 Gb / DVD-R с монитором 1 шт. (Инв. №558777/9)</li> <li>5. Стенд «Сергеев Сергей Степанович 1910-1999» 1 шт. (Инв.№591013/25)</li> <li>6. Огнетушитель порошковый 1 шт. (Инв. №559527)</li> <li>7. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528)</li> <li>8. Жалюзи 2шт. (Инв. №1107-221225, Инв. №1107-221225)</li> <li>9. Лавка 20 шт.</li> <li>10. Стол аудиторный 20 шт.</li> <li>11. Стол для преподавателя 1 шт.</li> <li>12. Стул 2 шт.</li> <li>13. Доска маркерная 1 шт.</li> <li>14. Трибуна напольная 1 шт. (без инв. №)</li> </ol>
<p><i>учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы (2й учебный корпус, 302 ауд.)</i></p>	<ol style="list-style-type: none"> <li>1. Системный блок Intel Core Intel Core i3-2100/4096Mb/500Gb/DVD-RW 10 шт. (Инв.№601997, Инв.№601998, Инв.№601999, Инв.№602000, Инв.№602001, Инв.№602002, Инв.№602003, Инв.№602004, Инв.№602005, Инв.№602006)</li> <li>2. Монитор 10 шт. (без инв. №) - приобретались не за счет средств вуза</li> <li>3. Шкаф 2 шт. (Инв.№594166, Инв.№594167)</li> <li>4. Тумба 1 шт. (Инв.№594168)</li> <li>5. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528)</li> <li>6. Огнетушитель порошковый 1 шт. (Инв. №559527)</li> <li>7. Жалюзи 1 шт. (Инв.№551557)</li> <li>8. Доска магнитно-маркерная 1 шт.</li> <li>9. Стол 5 шт.</li> <li>10. Стол компьютерный 12 шт.</li> <li>11. Стул офисный 21 шт.</li> <li>12. Сейф 1 шт. (без Инв.№).</li> </ol>
<p>Студенческое общежитие</p>	<p>Комнаты для самоподготовки</p>
<p>ЦНБ имени Н.И. Железнова</p>	<p>Читальный зал</p>

## 11. Методические рекомендации студентам по освоению дисциплины

Приступая к изучению дисциплины «Инжиниринг данных», студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке РГАУ-МСХА им. К.А. Тимирязева, получить в библиотеке рекомендованные учебники и учебно-методические пособия, завести новую тетрадь для работы с первоисточниками.

Предполагается, что студент выполняет практическое задание в аудитории, дома оформляет и готовится по теоретическим вопросам к защите работы на следующем занятии.

### Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан самостоятельно выполнить сообщение (презентацию), рассмотренную на практическом занятии и подготовиться по контрольным вопросам к защите работы в рамках часов консультаций.

## **12. Методические рекомендации преподавателям по организации обучения по дисциплине**

Комплексное освоение студентами учебной дисциплины «Инжиниринг данных» предполагает изучение рекомендуемой учебно-методической литературы, подготовку к практическим занятиям, самостоятельную работу при выполнении практических заданий, домашних заданий.

На первом занятии преподаватель закрепляет за каждым студентом номер варианта для выполнения индивидуальных работ (как правило, номер варианта соответствует порядковому номеру студента в журнале преподавателя). По каждой индивидуальной работе должна быть поставлена оценка по факту ее защиты. Защиту рекомендуется проводить на следующем после получения задания занятии. Преподаватель обязан проверить соответствие выполненного задания исходным данным варианта студента. Таким образом, исключается вероятность плагиата.

В рамках курса предусмотрены формы работы студентами в малых группах, ориентированные на развитие навыков взаимодействия у студентов при решении профессиональных задач.

Преподаватель должен стимулировать студентов к занятию научно-исследовательской работой, изучению научной эконометрической литературы, в т.ч. отечественной и зарубежной периодики.

Студент может провести собственное статистическое наблюдение за социально-экономическими явлениями, представляющими его научный интерес, построить статистическую модель, сделать прогноз. В случае надлежащего качества, его работа может быть заслушана на научном кружке кафедры или на студенческой научной конференции. По решению кафедры, студенты, занявшие призовые места на научных студенческих конференциях, может быть награжден дополнительными баллами к своему личному рейтингу.

**Программу разработал (и):**

Харитонов А.Е., к.э.н., доцент

Ульянкин А.Е., ассистент



## РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.19 «Инжиниринг данных»  
ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии»,  
направленность «Компьютерные науки и интеллектуальный анализ данных  
(Computer Science & Data Mining)» (квалификация выпускника – бакалавр)

Коломеевой Еленой Сергеевной, доцентом кафедры финансов ФГБОУ ВО г. Москвы «РГАУ-МСХА имени К.А. Тимирязева», кандидатом экономических наук (далее по тексту рецензент), проведено рецензирование рабочей программы дисциплины «Инжиниринг данных» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)» (уровень обучения - бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчики – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и кибернетики и Ульянов Александр Евгеньевич, ассистент кафедры статистики и кибернетики).

Рассмотрев представленные на рецензирование материалы, рецензент пришел к следующим выводам:

1. Предъявленная рабочая программа дисциплины «Инжиниринг данных» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.03.02 «Информационные системы и технологии». Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

2. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к дисциплинам части, формируемой участниками образовательных отношений учебного цикла учебного цикла — Б1.В.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

4. В соответствии с Программой за дисциплиной «Инжиниринг данных» закреплено 3 компетенции (6 индикаторов). Дисциплина «Инжиниринг данных» и представленная Программа способна реализовать их в объявленных требованиях.

5. Результаты обучения, представленные в Программе в категориях знать, уметь, владеть соответствуют специфике и содержанию дисциплины и демонстрируют возможность получения заявленных результатов.

6. Общая трудоёмкость дисциплины «Инжиниринг данных» составляет 3 зачётные единицы (108 часов, в том числе практической подготовки 2 часа).

7. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Инжиниринг данных» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии» и возможность дублирования в содержании отсутствует.

8. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

9. Программа дисциплины «Инжиниринг данных» предполагает проведение занятий в интерактивной форме.

10. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

11. Представленные и описанные в Программе формы текущей оценки знаний (защита работ), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме экзамена в 5 семестре, что соответствует статусу дисциплины, как

дисциплины части, формируемой участниками образовательных отношений учебного цикла – Б1.В ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

12. Формы оценки знаний, представленные в Программе, соответствуют специфике дисциплины и требованиям к выпускникам.

13. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 4 источника (базовый учебник), дополнительной литературой – 3 наименования Интернет-ресурсы – 6 источников и соответствует требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

14. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Инжиниринг данных» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

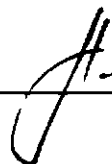
15. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Инжиниринг данных».

#### **ОБЩИЕ ВЫВОДЫ**

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины ««Инжиниринг данных» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность **«Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»** (квалификация выпускника – бакалавр), разработанная Харитоновой А. Е., к.э.н., доцентом кафедры статистики и кибернетики и Ульяновым А.Е. ассистентом кафедры статистики и кибернетики, соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

**Рецензент: Коломеева Е.С., доцент кафедры финансов ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», кандидат экономических наук**

(подпись)



«24» августа 2022 г.