

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Хоружий Яковлев Иванович

Должность: Директор института экономики и управления АПК

Дата подписания: 2023-10-19 17:56

Уникальный электронный ключ:

1e90b132d90c41e7585160b015dddf2cb1e4



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 управ-
ления АПК


Я.И. Хоружий
« 19 » октября 2022 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ Б1.В.ДВ.04.02 Парсинг и предобработка данных

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»

Направленность:

Большие данные и машинное обучение (Machine Learning & Big Data)

Курс 3

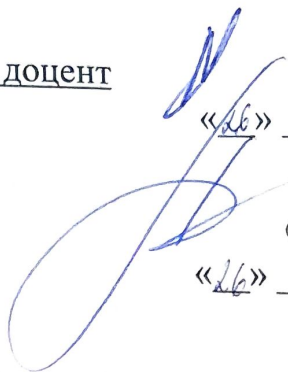
Семестр 6

Форма обучения очная

Год начала подготовки 2022

Москва, 2022

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


«26» 08 2022 г.

Рецензент: Коломеева Е.С., к.э.н.
(ФИО, ученая степень, ученое звание)

(подпись)

«26» 08 2022 г.

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

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)

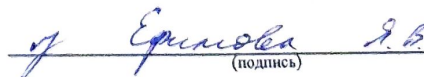
«26» 08 2022 г.

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


(подпись)

СОДЕРЖАНИЕ

АННОТАЦИЯ	4
1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ	5
3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	5
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	5
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ	6
ПО СЕМЕСТРАМ	6
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	10
4.3 ЛЕКЦИИ /ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	11
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	12
6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	13
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ.....	13
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ.....	14
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	15
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	15
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	15
7.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ, РЕКОМЕНДАЦИИ И ДРУГИЕ МАТЕРИАЛЫ К ЗАНЯТИЯМ	16
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	16
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ.....	16
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....	17
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ..	18
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	18

Аннотация

рабочей программы учебной дисциплины

Б1.В.ДВ.04.02 «Парсинг и предобработка данных»

для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности **Большие данные и машинное обучение (Machine Learning & Big Data)** и

Цель освоения дисциплины: Целью дисциплины «Парсинг и предобработка данных» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области проведения аналитического исследования с применением технологий больших данных, а также осуществлять деловую коммуникацию в устной и письменной формах.

Место дисциплины в учебном плане: включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) 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): УК-4 (УК-4.2, УК-4.3), ПКос-9 (ПКос-9.1, ПКос-9.2, ПКос-9.3).

Краткое содержание дисциплины:

Понятие парсинга данных. Веб-скрейпинга. Способы парсинга данных. Библиотеки для парсинга данных для языков программирования R и Python. Возможности парсинга данных. Сферы применения парсинга данных. Использование парсинга данных в АПК. Предварительная обработка данных. Структурирование данных. Преобразование типов данных. Заполнение пропусков, сэмплинг. Квантование. Конечные классы. Разбиение на множества. Редактирование выбросов. Сглаживание. Поиск дубликатов и противоречий во входных данных. Группировка. Дополнение данных. Замена. Кросс-таблица. Объединение. Параметры полей. Разгруппировка. Свёртка столбцов. Скользящее окно. Слияние. Соединение. Сортировка. Фильтр строк. Корреляционный анализ. Настройки нормализации. Разбиение на множества. Настройка линейной регрессии. Детальные настройки. Отбор факторов и защита от переобучения. F-тест. Применение анализа данных к преобразованным данным.

Общая трудоемкость дисциплины составляет: 3 зачетные единицы (108 часов).

Промежуточный контроль: зачет с оценкой.

1. Цель освоения дисциплины

Целью дисциплины «Парсинг и предобработка данных» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области проведения аналитического исследования с применением технологий больших данных, а также осуществлять деловую коммуникацию в устной и письменной формах.

2. Место дисциплины в учебном процессе

Дисциплина «Парсинг и предобработка данных» включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) учебного плана. Дисциплина «Парсинг и предобработка данных» реализуется в соответствии с требованиями ФГОС ВО, профессионального стандарта, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Парсинг и предобработка данных» являются «Введение в компьютерные науки на иностранном языке», «Математический анализ», «Математическая статистика», «Теория вероятностей», «Алгоритмизация и программирование», «Основы науки о данных (Data Science)», «Анализ экономических данных с использованием современных информационных технологий на иностранном языке», «Анализ экономических данных с использованием современных информационных технологий», «Хранилища и системы интеллектуального анализа данных на иностранном языке», «Хранилища и системы интеллектуального анализа данных».

Дисциплина «Парсинг и предобработка данных» является основополагающей для изучения следующих дисциплин: «Администрирование информационных систем», «Многомерные статистические методы», «Методы искусственного интеллекта», «Большие данные».

Рабочая программа дисциплины «Парсинг и предобработка 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

Образовательные результаты освоения дисциплины обучающимся, представлены 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), 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.			УК-4.2 Уметь: применять на практике деловую коммуникацию в устной и письменной формах, методы и навыки делового общения на русском и иностранном языках		применять на практике деловую коммуникацию в сфере работы с парсингом и предобработкой данных в устной и письменной формах на русском и иностранном языках	
2.			УК-4.3 Владеть: навыками чтения и перевода текстов на иностранном языке в профессиональном общении; навыками деловых коммуникаций в устной и письменной форме на русском и иностранном языках; методикой составления суждения в межличностном деловом общении на русском и иностранном языках			навыками деловых коммуникаций в области парсинга данных в устной и письменной форме на русском и иностранном языках
5.			ПКос-9.1 Знать: предметную область	типы данных, источники и методы автоматиче-		

		применением технологий больших данных	анализа, типы больших данных, источники и методы извлечения информации, теоретические и прикладные основы анализа, технологии хранения и обработки, современные методы и инструментальные средства анализа больших данных	ского извлечения информации, технологии хранения и обработки, современные методы и инструментальные средства предобработки больших данных		
б.			ПКос-9.2 Уметь: оценивать соответствие наборов данных задачам анализа больших данных; использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников; разрабатывать и оценивать модели больших данных; автоматизировать процесс анализа больших данных; визуализировать результаты анализа больших		использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников.	

			данных			
7.			ПКос-9.3 Иметь навыки: выбора источников данных, оценки соответствия набора данных предметной области и задачам аналитических работ; получения и фильтрации, извлечения, проверки, очистки, агрегации и разработки представления больших объемов данных из гетерогенных источников			получения и фильтрации, извлечения, проверки, очистки, агрегации и разработки представления больших объемов данных из гетерогенных источников

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость, 6 семестр час. всего/*
Общая трудоёмкость дисциплины по учебному плану	108
1. Контактная работа:	50,35
Аудиторная работа	50,35
<i>в том числе:</i>	
<i>лекции (Л)</i>	16
<i>практические занятия (ПЗ)</i>	34/4
<i>контактная работа на промежуточном контроле (КРА)</i>	0,35
2. Самостоятельная работа (СРС)	57,65
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям и т.д.)</i>	48,65
<i>Подготовка к зачету (контроль)</i>	9
Вид промежуточного контроля:	зачет с оценкой

* в том числе практическая подготовка.

4.2 Содержание дисциплины

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнёно)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ всего /*	ПКР	
Тема 1 Технология парсинга данных	44/2	8	16/2		20
Тема 2 Предобработка и консолидация данных	34,65/2	4	10/2		20,65
Тема 3 Обработка данных	29	4	8		17
Контактная работа на промежуточном контроле (КРА)	0,35			0,35	
Всего за 6 семестр	108	16	34/4	0,35	57,65
Итого по дисциплине	108	16	34/4	0,35	57,65

* в том числе практическая подготовка

Тема 1 Технология парсинга данных

Понятие парсинга данных. Веб-скрейпинга. Способы парсинга данных. Библиотеки для парсинга данных для языков программирования R и Python. Возможности парсинга данных. Сферы применения парсинга данных. Использование парсинга данных в АПК.

Тема 2 Предобработка и консолидация данных

Предварительная обработка данных. Структурирование данных. Преобразование типов данных. Заполнение пропусков, сэмплинг. Квантование. Конечные классы. Разбиение на множества. Редактирование выбросов. Сглаживание. Поиск дубликатов и противоречий во входных данных.

Тема 3 Обработка данных

Группировка. Дополнение данных. Замена. Кросс-таблица. Объединение. Параметры полей. Разгруппировка. Свёртка столбцов. Скользящее окно. Слияние. Соединение. Сортировка. Фильтр строк. Корреляционный анализ. Настройки нормализации. Разбиение на множества. Настройка линейной регрессии. Детальные настройки. Отбор факторов и защита от переобучения. F-тест. Применение анализа данных к преобразованным данным.

4.3 Лекции /практические занятия

Таблица 4

Содержание лекции /практические занятия и контрольные мероприятия

№ п/п	№ раздела	№ и название практических занятий	Формируемые компетенции и (индикаторы)	Вид контрольного мероприятия	Кол-во Часов/ из них практическая подготовка
		Лекция 1. Общее понятие парсинга данных	ПКос-9.1 УК-4.2		2
		Лекция 2. Технология парсинга данных	ПКос-9.1 УК-4.2		4
		Лекция 3. Примеры использования парсинга данных	ПКос-9.1 УК-4.2		2
		Практическая работа № 1. Общие принципы парсинга данных	ПКос-9.2 ПКос-9.3 УК-4.2 УК-4.3	Вопросы для обсуждения	2
		Практическая работа № 2. Парсинг данных из различных источников	ПКос-9.2 ПКос-9.3 УК-4.2	Защита работы	14/2
		Лекция 4 Способы предобработки и консолидации данных	ПКос-9.1 УК-4.2		4
		Практическая работа № 3. Приведение данных к структурированному виду	ПКос-9.2 ПКос-9.3 УК-4.2 УК-4.3	Защита работы	10/2
		Лекция 5 Способы обработки данных	ПКос-9.1 УК-4.2		4

№ п/п	№ раздела	№ и название практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во Часов/ из них практическая подготовка
		Практическая работа № 4. Сортировка, объединение и разделение наборов данных	ПКос-9.2 ПКос-9.3 УК-4.2 УК-4.3	Защита работы	4
		Практическая работа № 5. Предварительный анализ данных	ПКос-9.2 ПКос-9.3 УК-4.2 УК-4.3	Защита работы	4

Таблица 5

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Тема 1 Технология парсинга данных	Библиотеки для парсинга данных для языков программирования R и Python. Возможности парсинга данных. Сферы применения парсинга данных. (УК-4.2, УК-4.3, ПКос-9.1, ПКос-9.2)
2.	Тема 2 Предобработка и консолидация данных	Заполнение пропусков, сэмплинг. Квантование. Конечные классы. Разбиение на множества. (УК-4.2, УК-4.3, ПКос-9.1, ПКос-9.2, ПКос-9.3)
3	Тема 3 Обработка данных	Сортировка. Фильтр строк. Корреляционный анализ. Настройки нормализации. (УК-4.2, УК-4.3, , , ПКос-9.1, ПКос-9.2, ПКос-9.3)

5. Образовательные технологии

Таблица 6

Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий
1	Лекция 2. Примеры использования парсинга данных	Л
2	Практическая работа № 4. Сортировка, объединение и разделение наборов данных	ПЗ

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

Вопросы к зачету с оценкой

1. Что такое парсинг?
2. Какой инструмент для парсинга является лучшим?
3. Пакеты python для парсинга данных.
4. Пакеты языка R для парсинга данных.
5. Парсинг данных из социальных сетей.
6. Парсинг данных из различных источников.
7. Для чего используется парсинг?
8. Отличия Data Mining и парсинга?
9. Как избежать блокировки при парсинге сайта?
10. Можно ли в процессе парсинга решать капчу (captcha)?
11. В чем разница между парсингом сайтов и их сканированием?
12. Что такое robots.txt
13. Возможно ли парсить данные на сайтах, требующих авторизации?
14. Как извлечь контент с динамических веб-страниц?
15. Может ли парсер скачивать файлы с веб-сайтов?

Практическая работа № 1. Общие принципы парсинга данных

Вопросы для обсуждения:

1. Возможности парсинга данных.
2. Где может применяться парсинг данных.
3. Источники парсинга данных.
4. Недостатки парсинга данных.
5. Достоинства парсинга данных.

Практическая работа № 2. Парсинг данных из различных источников

Реализовать алгоритм парсинга данных с различных сайтов с использованием специализированных пакетов r, python, а также облачных сервисов. Сравнить результаты, выбрать лучший инструмент для парсинга. Подготовить отчет с выводами.

Практическая работа № 3. Приведение данных к структурированному виду

По результатам парсинга данных настроить автоматическое приведение данных к структурированному виду. Заполнить пропущенные значения в наборах данных. Для каждого столбца исходного набора данных пользователь может выбрать наиболее подходящий метод заполнения пропусков. Настроить

программу для автоматической корректировки выбросов и экстремальных значений в наборах данных. Для каждого поля исходного набора данных критерии определения выбросов и экстремальных значений задаются пользователем с помощью указания допустимого стандартного отклонения или интерквартильного размаха. Осуществить разбиение данных на обучающую и тестовую выборки. По итогам выполнения работы предоставить отчет с выводами.

Практическая работа № 4. Сортировка, объединение и разделение наборов данных

По данным прошлой работы провести группировку данных, разгруппировку, замену данных, слияние и свертку столбцов. Построить кросс-таблицу. Настроить параметры полей. Осуществить сортировку и фильтрацию строк. По итогам выполнения работы предоставить отчет с выводами. Провести анализ данных основе корреляционного анализа. Построить матрицу парных коэффициентов корреляции. Определить взаимосвязи между признаками. Проверить исходные данные на наличие автокорреляции. По итогам выполнения работы предоставить отчет с выводами.

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

Текущий контроль знаний, умений и навыков проводится в форме тестирования и теоретическими вопросами. Оценка работ проводится по стобалльной шкале. Индивидуальные задачи, выполняемые каждым студентом на практике оцениваются по итогам защиты по аналогичной шкале. Ликвидация студентами текущих задолженностей производится также в форме выполнения индивидуальной задачи по соответствующей теме и дальнейшей ее защиты преподавателю кафедры с оценкой по стобалльной шкале.

Для получения зачета с оценкой необходимо набрать более 60%. Вид промежуточного контроля по данному направлению – зачет с оценкой.

Градация оценок:

0 – 60% - «неудовлетворительно»;

60 – 75 – «удовлетворительно»;

75 – 85 – «хорошо»;

85 – 100 – «отлично»

Формы контроля: тестовый контроль, индивидуальное собеседование, защита выполнения практического задания по индивидуальному варианту. В итоговую сумму баллов входят результаты всех контролируемых видов вашей деятельности – посещение занятий, выполнение заданий, прохождение тестов, активность на лабораторных занятиях и т.п.

В итоговый рейтинг входит: 30% - результат выполнения контрольных мероприятий (тест, самостоятельные работы и др.), 60% - баллы за сданные индивидуальные работы и 10% - посещение занятий.

При изучении каждого раздела дисциплины проводится промежуточный контроль знаний с целью проверки и коррекции хода освоения теоретического материала и практических умений и навыков.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1 Основная литература

1. Волосова, А. В. Технологии искусственного интеллекта в ULS-системах : учебное пособие для вузов / А. В. Волосова. — Санкт-Петербург : Лань, 2022. — 308 с. — ISBN 978-5-8114-8839-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/208568>
2. Митина, О. А. Технологии организации, обработки и хранения статистических данных : учебное пособие / О. А. Митина, И. А. Юрченков. — Москва : РТУ МИРЭА, 2019. — 163 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/171511> (дата обращения: 27.11.2022). — Режим доступа: для авториз. пользователей.
3. Прокопенко, Н. Ю. Аналитические информационные системы поддержки принятия решений : учебное пособие / Н. Ю. Прокопенко. — Нижний Новгород : ННГАСУ, 2020. — 142 с. — ISBN 978-5-528-00395-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/164866> (дата обращения: 27.11.2022). — Режим доступа: для авториз. пользователей.
4. Скляр, А. Я. Технология хранения и интерактивных обработки данных : учебное пособие / А. Я. Скляр. — Москва : РТУ МИРЭА, 2020. — 69 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/163914> (дата обращения: 27.11.2022). — Режим доступа: для авториз. пользователей.
5. Data Science / Francesco Palumbo, Angela Montanari, Maurizio Vichi. Springer International Publishing AG, 2017 – Текст : электронный // Springer: электронно-библиотечная система. URL: <https://link.springer.com/book/10.1007/978-3-319-55723-6#editorsandaffiliations> (дата обращения: 27.11.2022).

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: 27.11.2022).
2. Wickman, H. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / H. Wickman, G. Grolemund. - Beijing ; Boston ; Sebastopol : O'REILLY, 2017.
3. New Advances in Statistics and Data Science / Ding-Geng, Chen Zhezhen, Jin Gang, Li Yi, Li Aiyi, Liu Yichuan, Zhao. Springer International Publishing AG, 2017 – Текст : электронный // Springer: электронно-библиотечная

система. URL: <https://link.springer.com/book/10.1007/978-3-319-69416-0#editorsandaffiliations> (дата обращения: 27.11.2022).

4. Точилкина, Т. Е. Хранилища данных и средства бизнес-аналитики : учебное пособие / Т. Е. Точилкина, А. А. Громова. — Москва : Финансовый университет, 2017. — 161 с. — ISBN 978-5-7942-1387-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/208367> (дата обращения: 27.11.2022). — Режим доступа: для авториз. пользователей.

7.3 Методические указания, рекомендации и другие материалы к занятиям

1. Харитоновна, А.Е. Парсинг и предобработка данных: методические указания / А.Е. Харитоновна. – М.: РГАУ-МСХА им. К.А. Тимирязева, 2016. – 25 с.

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

1. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)
2. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)
3. Анаконда. URL: <https://www.anaconda.com/distribution/> (открытый доступ)
4. Официальный сайт Росстата. URL: <https://rosstat.gov.ru/> (открытый доступ)
5. Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru> (открытый доступ)
6. Bureau of Economic Analysis. URL: <http://www.bea.gov> (открытый доступ)
7. Московская международная валютная биржа. <http://www.micex.ru> (открытый доступ)
8. Официальный сайт Всемирного банка . URL: <http://www.worldbank.org> (открытый доступ)
9. Официальный сайт Министерства финансов РФ. URL: <http://www.minfin.gov.ru> (открытый доступ)
10. Официальный сайт Национального бюро экономических исследований США. URL: <http://www.nber.org> (открытый доступ)

9. Перечень программного обеспечения и информационных справочных систем

Таблица 9

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Тема 1 Технология парсинга данных Тема 2 Предобработка и консолидация данных	R	расчётная	r-project	2022

	Тема 3 Обработка данных				
2	Тема 1 Технология парсинга данных Тема 2 Предобработка и консолидация данных Тема 3 Обработка данных	RStudio	расчётная	r-project	2022
3	Тема 1 Технология парсинга данных Тема 2 Предобработка и консолидация данных Тема 3 Обработка данных	Anaconda	расчётная	Anaconda Enterprise	2022

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
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации</i>	<ol style="list-style-type: none"> 1. Экран с электроприводом 1 шт. (Инв. №558771/2) 2. Проектор 1 шт. (без инв. №) – приобретался не за счет средств вуза 3. Вандалоустойчивый шкаф 1 шт. (Инв.№558850/7) 4. Системный блок с монитором 1 шт. (Инв. №558777/9) 5. Стенд «Сергеев Сергей Степанович 1910-1999» 1 шт. (Инв.№591013/25) 6. Огнетушитель порошковый 1 шт. (Инв. №559527) 7. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528) 8. Жалюзи 2шт. (Инв. №1107-221225, Инв. №1107-221225) 9. Лавка 20 шт. 10. Стол аудиторный 20 шт. 11. Стол для преподавателя 1 шт. 12. Стул 2 шт. 13. Доска маркерная 1 шт. 14. Трибуна напольная 1 шт. (без инв. №)
<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</i>	<ol style="list-style-type: none"> 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) 2. Монитор 10 шт. (без инв. №) - приобретались не за счет средств вуза 3. Шкаф 2 шт. (Инв.№594166, Инв.№594167) 4. Тумба 1 шт. (Инв.№594168) 5. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528) 6. Огнетушитель порошковый 1 шт. (Инв. №559527) 7. Жалюзи 1 шт. (Инв.№551557)

	8. Доска магнитно-маркерная 1 шт. 9. Стол 5 шт. 10. Стол компьютерный 12 шт. 11. Стул офисный 21 шт. 12. Сейф 1 шт. (без Инв.№).
<i>учебная аудитория для проведения занятий лекционного типа, 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</i>	1. Трибуна напольная 1 шт. (Инв.№ 599205) 2. Шкаф для документов 3 шт. (Инв.№593633, Инв.№593634, Инв.№559548/18) 3. Вешалка напольная 2 шт. (Инв.№1107-333144, Инв.№1107-333144) 4. Жалюзи 1 шт. (Инв.№591110) 5. Доска магнитно-маркерная 1 шт. 6. Стол 15 шт. 7. Скамейка 14 шт. 8. Стол эрго 1 шт. 9. Стул 2 шт.
<i>Центральная научная библиотека имени Н.И. Железнова</i>	Читальные залы библиотеки
<i>Студенческое общежитие</i>	Комната для самоподготовки

11. Методические рекомендации обучающимся по освоению дисциплины

Все виды учебных работ должны быть выполнены точно в сроки, предусмотренные программой обучения. Если студент не выполнил какое-либо из учебных заданий по неуважительной причине (пропустил тестовый контроль, не выполнили домашнего задания, выполнил работу не по своему варианту и т.п.), то за данный вид учебной работы баллы рейтинга не начисляются, а подготовленные позже положенного срока работы оцениваются с понижающим коэффициентом. Если же невыполнение учебных работ произошло по уважительной причине, то следует представить преподавателю подтверждающий документ, и защитить пропущенные занятия в часы, отведенные для еженедельных консультаций.

Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан выполнить самостоятельно индивидуальную работу, выполняемую на занятиях по своему варианту.

12. Методические рекомендации преподавателям по организации обучения по дисциплине

Курс должен давать не абстрактно-формальные, а прикладные знания. Данная цель может быть реализована только при условии соблюдения в учебных планах преемственности учебных дисциплин. Базовые знания для изучения дисциплины дают такие предметы, как экономическая теория, информатика.

Преподаватель должен указывать, в какой последовательности следует изучать материал дисциплины, обращать внимание на особенности изучения отдельных тем и разделов, помогать отбирать наиболее важные и необходимые сведения из учебных пособий, а также давать объяснения вопросам программы курса, которые обычно вызывают затруднения. При этом преподавателю необходимо учитывать следующие моменты:


1. Не следует перегружать студентов творческими заданиями.

курса, которые обычно вызывают затруднения. При этом преподавателю необходимо учитывать следующие моменты:

1. Не следует перегружать студентов творческими заданиями.
2. Чередовать творческую работу на занятиях с заданиями во внеаудиторное время.
3. Давать студентам четкий инструктаж по выполнению самостоятельных заданий: цель задания; условия выполнения; объем; сроки; требования к оформлению.
4. Осуществлять текущий учет и контроль за самостоятельной работой.
5. Давать оценку обобщать уровень усвоения навыков самостоятельной, творческой работы.

Программу разработал (и):

Харитонов А.Е., к.э.н., доцент



(подпись)

РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.ДВ.04.02 «Парсинг и предобработка данных»
ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии»,
направленность «Большие данные и машинное обучение (Machine Learning & Big
Data)»

(квалификация выпускника – бакалавр)

Колосова Елена Сергеевна, доцент кафедры финансов ФГБОУ ВО г. Москвы «РГАУ-МСХА имени К.А. Тимирязева», кандидатом экономических наук (далее по тексту рецензент), проведена рецензия рабочей программы дисциплины «Парсинг и предобработка данных» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Большие данные и машинное обучение (Machine Learning & Big Data)» (бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчик – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и кибернетики).

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

1. Предъявленная рабочая программа дисциплины «Парсинг и предобработка 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 по выбору части, формируемой участниками образовательных отношений учебного цикла – Б1.В.ДВ.

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 «Парсинг и предобработка данных» закреплено **2 компетенции (5 индикаторов)**. Дисциплина «Парсинг и предобработка данных» и представленная Программа способна реализовать их в объявленных требованиях.

5. Общая трудоёмкость дисциплины «Парсинг и предобработка данных» составляет 3 зачётных единицы (108 часов/из них практическая подготовка 4 ч.).

6. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Парсинг и предобработка 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.

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

8. Программа дисциплины «Парсинг и предобработка данных» предполагает 2 часа занятий в интерактивной форме.

9. 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».

10. Представленные и описанные в Программе формы *текущей* оценки знаний (опрос, как в форме обсуждения отдельных вопросов, так и участие в деловых играх), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета с оценкой, что соответствует статусу дисциплины, как дисциплины по выбору части, формируемой участниками образовательных отношений учебного цикла – Б1.В.ДВ ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

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

11. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 5 источников (базовый учебник), дополнительной литературой – 4 наименования, Интернет-ресурсы – 10 источников и соответствует требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

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

13. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Парсинг и предобработка данных».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Парсинг и предобработка данных» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «**Большие данные и машинное обучение (Machine Learning & Big Data)**» (квалификация выпускника – бакалавр), разработанная Харитоновой А. Е., к.э.н., доцентом 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 _____ « 26 » _____ 2021 г.

(подпись)