

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Вятский государственный университет»

УТВЕРЖДАЮ
Начальник Управления
дополнительного образования
Курагина К. А. Курагина
« 29 » апреля 2022 г.

р/з № 03-04-2022-0565-0966

РАБОЧАЯ ПРОГРАММА

учебной дисциплины (модуля)
«Data Science: аналитика больших данных»

дополнительной профессиональной программы –
программы повышения квалификации
«Data Science: аналитика больших данных»

Киров, 2022

Рабочая программа составлена в соответствии с требованиями дополнительной профессиональной программы «Data Science: аналитика больших данных»

Рабочая программа разработана:

Лучинина Марина Ивановна, зав. кафедрой социальной работы и молодежной политики ВятГУ.

Митягина Екатерина Владимировна, д.соц.н., профессор кафедры социальной работы и молодежной политики.

Ананин Павел Валерьевич, старший преподаватель кафедры социальной работы и молодежной политики

© Вятский государственный университет, 2022

© Лучинина М.И., 2022

© Митягина Е.В., 2022

© Ананин П.В., 2022

1. РАБОЧАЯ УЧЕБНАЯ ПРОГРАММА

1.1 Пояснительная записка

Актуальность и значение дополнительной профессиональной программы – программы повышения квалификации «Data Science: аналитика больших данных» определяются необходимостью расширения возможностей для эффективной реализации научно-исследовательского потенциала слушателей.

Цели и задачи учебной дисциплины

Цель учебной дисциплины	Повышение профессиональной компетентности в аналитике больших данных у студентов и профессорско-преподавательского состава, реализующих основные образовательные программы по социально-гуманитарным направлениям подготовки.
Задачи учебной дисциплины	<ol style="list-style-type: none">1. Формирование новых и совершенствование имеющихся профессиональных компетенций по анализу больших данных с использованием методологической и технологической инфраструктуры.2. Формирование новых и совершенствование имеющихся профессиональных по управлению этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных.

Компетенции слушателя, формируемые в результате освоения учебной дисциплины / модуля

В результате освоения учебной дисциплины (модуля) слушатель должен демонстрировать следующие результаты обучения:

Виды деятельности	Профессиональные компетенции	Практический опыт	Умения	Знания
Деятельность консультативная и работы в области компьютерных технологий	ПК-1 Анализ больших данных с использованием методологической и технологической инфраструктуры	Владение навыками проведения аналитического исследования с применением технологий больших данных	Умение планировать и организовать аналитические работы с использованием технологий больших данных	Знания по методам выявления, формирования и согласования требований к результатам аналитических работ с применением

				технологий больших данных
Деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность	ПК-2 Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных	Владение навыками разработки, согласования и управления реализацией рабочего проекта методической и технологической инфраструктуры больших данных	Умение анализировать потребности заинтересованных лиц и/или организаций в исследовании больших данных	Знание об управлении получением, хранением, передачей, обработкой больших данных

1.2 Содержание учебной дисциплины (модуля)

Объем учебной дисциплины и виды учебной работы

Форма обучения	Общий объем (трудоемкость) Часов	В том числе аудиторная контактная работа обучающихся с преподавателем, час					Самостоятельная работа, час	Форма промежуточной аттестации
		Всего	Лекции	Практические (семинарские) занятия	Лабораторные занятия	Консультации		
Очно-заочная с применением ДОТ	70	36	14	22	-	-	34	Зачет в форме защиты проектов

Тематический план

№ п/п	Основные разделы и темы учебной дисциплины	Часы		Самостоятельная работа
		Лекции	Практические (семинарские занятия)	
1.	Раздел 1. Введение в BIG DATA	4	6	14
2.	Тема 1.1. Введение в BIG DATA: инструменты, кейсы, технологии	2	2	4
3.	Тема 1.2. Источники открытых данных	-	2	4

4.	Тема 1.3. Технологии сбора данных открытых платформ VK API	2	2	6
5.	Раздел 2. Основы машинного обучения и текстовой аналитики	6	10	10
6.	Тема 2.1. Цифровое моделирование личности на основе машинного обучения	2	2	2
7.	Тема 2.2. Введение в машинное обучение	2	4	4
8.	Тема 2.3. Текстовая аналитика на платформе PolyAnalyst	2	4	4
9.	Раздел 3. Сетевой анализ	4	6	10
10.	Тема 3.1. SNA: анализ сетевого взаимодействия между субъектами	2	2	6
11.	Тема 3.2. Проектирование дизайна исследования	2	4	4
ИТОГО		14	22	34

Матрица соотнесения разделов / тем учебной дисциплины / модуля и формируемых в них компетенций

РАЗДЕЛЫ / ТЕМЫ УЧЕБНОЙ ДИСЦИПЛИНЫ	Количество часов	Компетенции		
		ПК-1	ПК-2	Общее количество компетенций
Раздел 1. Введение в BIG DATA	24			
Тема 1.1. Введение в BIG DATA: инструменты, кейсы, технологии	8	+	+	2
Тема 1.2. Источники открытых данных	6	+	+	2
Тема 1.3. Технологии сбора данных открытых платформ VK API	10	+	+	2
Раздел 2. Основы машинного обучения и текстовой аналитики	26			
Тема 2.1. Цифровое моделирование личности на основе машинного обучения	6	+	+	2
Тема 2.2. Введение в машинное обучение	10	+	+	2
Тема 2.3. Текстовая аналитика на платформе PolyAnalyst	10	+	+	2
Раздел 3. Сетевой анализ	20			

Тема 3.1. SNA: анализ сетевого взаимодействия между субъектами	10	+	+	2
Тема 3.2. Проектирование дизайна исследования	10	+	+	2
ИТОГО	70			

Краткое содержание учебной дисциплины

Раздел 1. Введение в BIG DATA

Тема 1.1. Введение в BIG DATA: инструменты, кейсы, технологии

Будут представлены различные основания определения концепта “Big Data”, показана ретроспектива развития исследований, основанных на анализе больших данных. Отдельно будут обозначены тезисы существующей дискуссии относительно возможностей и ограничений анализа больших данных в социальных науках.

Тема 1.2. Источники открытых данных

Лекции познакомят слушателей с различными источниками и способами сбора открытых данных: методами работы с API (программный интерфейс приложения), парсингом открытых данных сайтов и порталов.

Тема 1.3. Технологии сбора данных открытых платформ VK API

Будут рассмотрены преимущества социальной сети «вконтакте». Понятие «идентификатор пользователя», «access token». Приведены примеры запроса к API. Рассмотрены основные инструменты обработки результатов сбора данных.

Раздел 2. Основы машинного обучения и текстовой аналитики

Тема 2.1. Цифровое моделирование личности на основе машинного обучения

Лекция познакомит слушателей с основами машинного обучения, историей развития технологии и яркими примерами применения алгоритмов в прикладных сферах.

Тема 2.2. Введение в машинное обучение

В рамках лекции будут подробно разобраны основные теоретические понятия, а также обзор наиболее популярных инструментов для анализа данных и машинного обучения – библиотек Pandas и Scikit-learn.

Тема 2.3. Текстовая аналитика на платформе PolyAnalyst

Будут представлены возможности лингвистического анализа текстов: классификация текстов, выделение ключевых лиц и событий, контент-анализ, определение содержания и тональности текстов.

Раздел 3. Сетевой анализ

Тема 3.1. SNA: анализ сетевого взаимодействия между субъектами

Будет представлена методология анализа взаимодействия субъектов (людей, организаций) между собой в контексте различных социальных явлений. Будут проведены различия между структурным анализом, изучающим порядок взаимодействия между субъектами, позиционным анализом, показывающим наиболее влиятельных лиц в сети, а также динамическим анализом, который позволяет проанализировать, как изменяется взаимодействие между ними.

Тема 3.2. Проектирование дизайна исследования

Наука как вид деятельности. Big Data – междисциплинарное поле исследований. Принципы цифровых исследований. Этапы цифрового исследования. Источники данных. Этапы работы с данными. Методы цифровой аналитики.

2. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

2.1. Методические рекомендации для преподавателя

Организация учебного процесса предусматривает применение инновационных форм учебных занятий, развивающих у обучающихся навыки проведения аналитического исследования с применением технологий больших данных, разработки, согласования и управления реализацией рабочего проекта методической и технологической инфраструктуры больших данных.

2.2. Методические указания для слушателей

Успешное освоение учебной дисциплины предполагает активное, творческое участие обучающегося на всех этапах ее освоения путем планомерной, повседневной работы. Обучающийся обязан посещать лекции и семинарские занятия, получать консультации преподавателя и выполнять самостоятельную работу.

Выбор методов и средств обучения, образовательных технологий осуществляется преподавателем исходя из необходимости достижения обучающимися планируемых результатов освоения дисциплины, а также с

учетом индивидуальных возможностей обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья.

Изучение дисциплины следует начинать с проработки настоящей рабочей программы, методических указаний и разработок, указанных в программе, особое внимание уделить целям, задачам, структуре и содержанию дисциплины.

Главной задачей каждой лекции является раскрытие сущности темы и анализ ее основных положений. Содержание лекций определяется настоящей рабочей программой дисциплины.

Лекции – это систематическое устное изложение учебного материала. На них обучающийся получает основной объем информации по каждой конкретной теме. Лекции носят проблемный характер и нацелены на освещение наиболее трудных и дискуссионных вопросов, кроме того, они способствуют формированию у обучающихся навыков самостоятельной работы с научной литературой.

Целью практических занятий является проверка уровня понимания обучающимися вопросов, рассмотренных на лекциях и в учебной литературе, степени и качества усвоения материала; применение теоретических знаний в реальной практике решения задач; восполнение пробелов в пройденной теоретической части курса и оказания помощи в его освоении.

Практические занятия направлены на совершенствование индивидуальных навыков решения теоретических и прикладных задач, выработку навыков интеллектуальной работы, а также ведения дискуссий. Конкретные пропорции разных видов работы в группе, а также способы их оценки определяются преподавателем, ведущим занятия.

На практических занятиях обучающиеся обсуждают дискуссионные вопросы, отвечают на вопросы тестов, закрепляя приобретенные знания, выполняют практические задания, анализируют кейсы и т.п. Для успешного проведения практического занятия обучающемуся следует тщательно подготовиться.

Основной формой подготовки обучающихся к практическим занятиям является самостоятельная работа с учебно-методическими материалами, научной литературой, статистическими данными и т.п.

Изучив конкретную тему, обучающийся может определить, насколько хорошо он в ней разобрался. Если какие-то моменты остались непонятными, целесообразно составить список вопросов и на занятии задать их преподавателю. Практические занятия предоставляют студенту возможность творчески раскрыться, проявить инициативу и развить навыки публичного ведения дискуссий и общения, сформировать определенные навыки и умения и т.п.

Самостоятельная работа слушателей включает в себя выполнение различного рода заданий (изучение учебной и научной литературы, материалов лекций, систематизацию прочитанного материала, подготовку контрольной работы, решение задач и т.п.), которые ориентированы на более глубокое усвоение материала изучаемой дисциплины. По каждой теме

учебной дисциплины преподаватель предлагает обучающимся перечень заданий для самостоятельной работы. Самостоятельная работа по учебной дисциплине может осуществляться в различных формах: подготовка докладов; анализ кейсов, написание рефератов; выполнение творческих заданий по теме изучения. Другие.

К выполнению заданий для самостоятельной работы предъявляются следующие требования: задания должны исполняться самостоятельно либо группой и представляться в установленный срок, а также соответствовать установленным требованиям по оформлению.

Регулярно рекомендуется отводить время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам.

Результатом самостоятельной работы должно стать формирование у обучающегося определенных знаний, умений, навыков, компетенций.

При проведении промежуточной аттестации обучающегося учитываются результаты текущей аттестации в течение периода обучения.

Процедура оценивания результатов освоения учебной дисциплины осуществляется на основе действующего Положения об организации текущего контроля успеваемости и промежуточной аттестации обучающихся ВятГУ.

Для приобретения требуемых компетенций, хороших знаний и высокой оценки по дисциплине обучающимся необходимо выполнять все виды работ своевременно в течение всего периода обучения.

3. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Основная литература

1. Иванова В., Перерва А. Путь аналитика. Практическое руководство IT-специалиста. 2-е изд. — СПб.: Питер, 2015. — 304 с.: ил.

2. Ын Анналин, Су Кеннет, Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с.: ил. — (Серия «Библиотека программиста»).

3. Домингос, Педро, Верховный алгоритм: как машинное обучение изменит наш мир / Педро Домингос ; пер. с англ. В. Горохова ; [науч. ред. А. Сбоев, А. Серенко]. — М. : Манн, Иванов и Фербер, 2016. — 336 с.

4. Рассел, Стюарт, Норвиг, Питер, . Искусственный интеллект: современный подход. 2-е изд. : Пер. с англ. — М. : ООО «И.Д. Вильямс», 2016. — 1408 с.

Дополнительная литература

1. Чатфилд Т., Критическое мышление: Анализируй, сомневайся, формируй свое мнение / Том Чатфилд; Пер. с англ. — М. : Альпина Паблишер, 2019. — 328 с., ил.

2. Черняк, Леонид Большие Данные -- новая теория и практика / Открытые системы. СУБД. -- М.: Открытые системы, 2011. -- № 10.
3. Грас Дж. Data Science. Наука о данных с нуля: Пер. с. Англ. – СПб.: БХВ-Петербург, 2017. – 336 с.

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

1. Портал дистанционного обучения ВятГУ.
2. Раздел официального сайта ВятГУ, содержащий описание образовательной программы.
3. Аналитическая записка «Правовое регулирование молодежной политики в современных зарубежных государствах: сравнительно-правовой анализ» – <https://fadm.gov.ru/documents/download/545/>.
4. Виртуальное методическое объединение библиотек и организаций, работающих с молодежью «Методобъединение» – <http://vmo.rgub.ru/>.
5. ЭБС «Научная электронная библиотека eLIBRARY» (<http://elibrary.ru/defaultx.asp>).
6. ЭБС «Издательства Лань» (<http://e.lanbook.com/>).
7. ЭБС «Университетская библиотека online» (www.biblioclub.ru).
8. Внутренняя электронно-библиотечная система ВятГУ (<http://lib.vyatsu.ru/>).
9. ЭБС «ЮРАЙТ» (<https://urait.ru>)

Описание материально-технической базы, необходимой для осуществления образовательного процесса

Перечень специализированных аудиторий (лабораторий)

Вид занятий	Назначение аудитории
Практика, лекция, семинар	Учебная аудитория
Самостоятельная работа	Читальные залы библиотеки

Перечень специализированного оборудования

Перечень используемого оборудования
Мультимедиа-проектор с экраном настенным
Ноутбук (персональный компьютер)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по учебной дисциплине

№ п.п	Наименование ПО	Краткая характеристика назначения ПО	Производитель ПО и/или поставщик ПО
1	Программная система с модулями	Программный комплекс для проверки текстов на предмет	ЗАО «Анти-Плагиат»

	для обнаружения текстовых заимствований в учебных и научных работах «Антиплагиат.ВУЗ»	заимствования из Интернет-источников, в коллекции диссертация и авторефератов Российской государственной библиотеки (РГБ) и коллекции нормативно-правовой документации LEXPRO	
2	Microsoft Office 365 Student Advantage	Набор веб-сервисов, предоставляющий доступ к различным программам и услугам на основе платформы Microsoft Office, электронной почте бизнес-класса, функционалу для общения и управления документами	ООО «Рубикон»
3	Office Professional Plus 2013 Russian OLP NL Academic.	Пакет приложений для работы с различными типами документов: текстами, электронными таблицами, базами данных, презентациями	ООО «СофтЛайн» (Москва)
4	Windows 7 Professional and Professional K	Операционная система	ООО «Рубикон»
5	Kaspersky Endpoint Security для бизнеса	Антивирусное программное обеспечение	ООО «Рубикон»
6	Информационная система КонсультантПлюс	Справочно-правовая система по законодательству Российской Федерации	ООО «КонсультантКиров»
7	Электронный периодический справочник «Система ГАРАНТ»	Справочно-правовая система по законодательству Российской Федерации	ООО «Гарант-Сервис»
8	Security Essentials (Защитник Windows)	Защита в режиме реального времени от шпионского программного обеспечения, вирусов.	Microsoft
9	PolyAnalyst	Система проведения анализа структурированных данных, и интеллектуального текстового анализа.	ООО «Компания «Мегасьютер Интеллидженс»
10	Google Chrome	Веб-браузер	Свободное использование
11	Gephi	Пакет программного обеспечения для анализа и визуализации сетей с открытым исходным кодом, написанный на Java на платформе NetBeans.	Свободное использование

4. МАТЕРИАЛЫ, УСТАНОВЛИВАЮЩИЕ СОДЕРЖАНИЕ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ (ТКУ) И САМОСТОЯТЕЛЬНОЙ РАБОТЫ СЛУШАТЕЛЕЙ

Формы ТКУ:

- собеседование;
- тест;
- реферат.

Формы самостоятельной работы:

- конспектирование;
- реферирование литературы;
- аннотирование книг, статей;
- работа с лекционным материалом: проработка конспекта лекций, работа на полях конспекта с терминами, дополнение конспекта материалами из рекомендованной литературы;
- участие в работе семинара: подготовка сообщений, докладов, заданий.

5. МАТЕРИАЛЫ, УСТАНОВЛИВАЮЩИЕ СОДЕРЖАНИЕ И ПОРЯДОК ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНЫХ АТТЕСТАЦИЙ

К сдаче зачета в форме защиты проекта допускаются все слушатели, проходящие обучение на данной ДПП, вне зависимости от результатов текущего контроля успеваемости и посещаемости занятий, при этом, результаты текущего контроля успеваемости могут быть использованы преподавателем при оценке уровня усвоения обучающимися знаний, приобретения умений, навыков и сформированности компетенций в результате изучения учебной дисциплины.

Промежуточная аттестация проводится в форме защиты проекта по совокупности выполненных заданий по самостоятельной работе.

Зачет принимается преподавателями, проводившими лекции по данной учебной дисциплине, в форме защиты проекта.

Методические рекомендации по подготовке и проведению промежуточной аттестации

Промежуточная аттестация проводится в целях повышения эффективности обучения, определения уровня профессиональной подготовки обучающихся и контролем за обеспечением выполнения стандартов обучения.

Перечень примерных заданий к зачету в форме защиты проекта

- 1. Опишите идею своего исследовательского или прикладного проекта. Пункты, обязательные к отражению:**

- Какова проблема?
На решение какой задачи или проблемы нацелен проект?
Кто может быть стейкхолдером данного проекта?
Какова цель анализа?
Какая начальная гипотеза/исследовательский вопрос?
Какие задачи нужно выполнить для достижения Цели?
Какие данные можно использовать? Какой источник данных? Что в них должно быть отражено?
Каковы ожидаемые результаты?

2. Задание по API

- Получить token для использования VK API (https://oauth.vk.com/authorize?client_id=5440699&display=page&redirect_uri=vk.com/callback&response_type=token&v=5.81&scope=groups,wall)
- Получить список подписчиков сообщества VK (пример - <https://vk.com/tsupython>), используя метод <https://vk.com/dev/groups.getMembers> с fields включающим city
- Построить запрос в виде URL с параметрами как показано здесь https://vk.com/dev/api_requests. Данный запрос должен обращаться к методу <https://vk.com/dev/wall.get> и выгружать первые 100 постов с любой страницы вконтакте.
- Отправить в качестве ответа на данное задание скриншот результата

3. Обучение модели прогнозирования.

- Скачайте архив с данными clock.ru/Y2PPG. Распакуйте содержимое архива.
- Запустите в браузере приложение по обучению прогнозных моделей
- clock.ru/Xpd8r
- Создайте на этапе подготовки к обучению 2 класса "Мужчины" и "Женщины". Загрузите в каждый класс изображения из папки "train" (данные скачены в п.1). В класс "Мужчины" загрузите изображения из папки "train/man", в класс "Женщины" из папки "train/woman"
- Внимательно просмотрите все изображения, понимание на каком фотографическом материале обучается ваша модель понадобится в последующих шагах для объяснения причин ошибок в прогнозе.
- Обучите модель
- Проверьте точность распознавания моделью пола людей на фотографиях из папки "test" (данные скачены в п.1), для каждого примера фотографии составьте краткий отчет и прикрепите его в ответ на это задание. Если модель даёт ошибочный прогноз или прогноз не со 100% точностью, дайте своё объяснение этому.
- Шаблон отчета:
test (3).jpg - мужчина 100% уверенность

- test (4).jpg - мужчина 90% уверенность, причины "неуверенности" модели (см. п.4)
- test (5).jpg -
- test (6).jpg -
- test (7).jpg -
- test (8).jpg -
- test (9).jpg -
- test (91).jpg -
- test (92).jpg -
- test (95).jpg -

4. Произвести разметку текста, следуя следующим указаниям:

- Изучить классификатор, который находится на странице «код-лист» в файле Domashnee_zadanie_klassif.

А	В
Образовательные направления	Описание
Общие программы	
Базовые программы	Базовые общие программы: доначальные, дошкольные, элементарные начальные, средние и т. д.
Грамотность и навыки счета	Простая и функциональная грамотность, навыки счета.
Развитие личности	Развитие личных навыков, т.е. бихевиористских способностей, умственных навыков, самоорганизации, программ жизненной ориентации.
Образование (педагогика)	
Подготовка учителей и педагогическая наука	Подготовка учителей для дошкольных учреждений, детских садов, начальных школ по профессиональным предметам, практическим предм
Языки	Иностранные языки: живые или "мертвые" языки и литература на этих языках;
Гуманитарные науки и искусство	Религия и теология; Другие гуманитарные науки: устный и письменный перевод, лингвистика, сравнительное литературоведение, история, а
Социальные и поведенческие науки	Экономика, история экономики, политология, социология, демография, антропология (за исключением физической антропологии), этнологи

- Перенести все данные из выгрузки на страницу «разметка» в файл Domashnee_zadanie_klassif.
- Добавить столбец «Образовательное направление»:

link	Образовательное направление
.com/club103017799	
.com/club50874595	
.com/paralleli_roman	
.com/club139138754	
.com/club22302701	
.com/clubromanian	
.com/english_express	
.com/globalcenter34	

- Проставить к каждому сообществу соответствующую образовательную категорию (обязательно проверив контент сообщества по ссылке в выгрузке).

5. Проанализировать графы (Вы можете взять один из трёх предлагаемых графов <https://moodle.ido.tsu.ru/mod/resource/view.php?id=36836>); визуально представить граф: выделить цветом кластеры и ЛОМов; загрузить скриншот